UNIVERSITÉ SAVOIE MONT BLANC

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ SAVOIE MONT BLANC

Spécialité : **Physique Subatomique et Astroparticules**

Arrêté ministériel : 25 Mai 2016

Présentée par

Mathis DUBAU

Thèse dirigée par **Olivier ARNAEZ**

préparée au sein du **Laboratoire d'Annecy de Physique des Particules**
dans **l'École Doctorale de Physique de Grenoble**

Étude multivariée des états de polarisation dans les événements WZ de diffusion de bosons vecteurs électrofaibles dans les données d'ATLAS au LHC

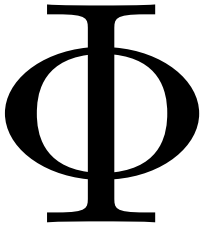
Multivariate study of polarisation states in electroweak vector boson scattering WZ events using ATLAS data at the LHC

Thèse soutenue publiquement le **28 Septembre 2026**,
devant le jury composé de :

M. Nicolas CHANON
Directeur de recherche, CNRS, Rapporteur
M. Nikola MAKOVEC
Directeur de recherche, CNRS, Rapporteur
M^{me} Corinne GOY
Directrice de recherche, CNRS, Examinatrice
M^{me} Rodanthi NICOLAIDOU
Ingénieur – Chercheur, CEA, Examinatrice
M. Gilles MAURIN
Professeur des universités, Université Savoie Mont-Blanc, Président du jury

« Un chercheur qui cherche, on en trouve, mais un chercheur qui trouve, on en cherche. »

Mon grand-père, Michel



Summary

English version

The cornerstone of the Standard Model, the theory widely accepted in particle physics, is the spontaneous breaking of symmetries, which predicts the appearance of a new scalar boson, the Higgs boson, discovered in 2012 at the LHC by the ATLAS and CMS experiments, and generates the mass of the vector bosons of the electroweak sector: the W and Z bosons. Owing to this particular feature, these bosons are the only elementary particles to possess three polarisation states, including a longitudinal one in which they behave at high energy like the Goldstone bosons of the theory prior to symmetry breaking. The study of VV events produced through electroweak processes therefore provides a unique way to probe the fundamental structure of the Standard Model. This thesis presents the first measurement of the longitudinal polarisation states of one or both bosons in WZjj-EW processes, using the data recorded by the ATLAS experiment during the full Run-2 of the LHC, corresponding to an integrated luminosity of $140\,\mathrm{fb}^{-1}$ at a centre-of-mass energy of $13\,\mathrm{TeV}$, combined with the first part of Run-3, corresponding to $164\,\mathrm{fb}^{-1}$ at $13.6\,\mathrm{TeV}$. Deep neural networks are trained to separate WZjj-EW events from the background and then to disentangle the polarisation states. Their outputs define a two-dimensional discriminant plane, in which a clustering algorithm builds an optimised Voronoi binning that maximises the sensitivity of a binned maximum-likelihood fit. The fraction of events in which both bosons are longitudinally polarised is measured to be $f_{00} = 0.000^{+0.053}_{-0.000}$, for a Standard Model prediction of 0.090 including the NLO electroweak corrections, and the fraction in which the W boson is longitudinally polarised regardless of the polarisation of the Z boson to be $f_{0\mathrm{X}} = 0.007^{+0.094}_{-0.007}$, for a prediction of 0.276, a 2.5σ departure from the Standard Model. Upper limits at $95\,\%$ confidence level are set at 0.12 (0.21 expected) and 0.19 (0.46 expected), respectively. By contrast, the fraction in which the Z boson is longitudinally polarised regardless of the polarisation of the W boson is measured to be $f_{\mathrm{X}0} = 0.324^{+0.092}_{-0.085}$, for a prediction of 0.275, with an observed (expected) significance of 4.0 (2.9) standard deviations, providing the first evidence for longitudinally polarised Z-boson states in WZjj-EW events. The signal strength of the process is found to be $\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}} = 1.06 \pm 0.18$, and its fiducial cross-section per lepton flavour, $\sigma(\mathrm{WZjj\text{-}EW} \to \ell'\nu_\ell\ell\ell\mathrm{jj})$ where ℓ and ℓ' are each either an electron or a muon, is reported at $13.6\,\mathrm{TeV}$ for the first time, $0.374 \pm 0.077\,(\mathrm{stat.}) \pm 0.053\,(\mathrm{syst.}) \pm 0.009\,(\mathrm{lumi.})\,\mathrm{fb}$. These measurements are statistically limited, and will benefit from the High-Luminosity LHC and from the replacement of the current tracking detector with the new Inner Tracker (ITk), extending the absolute pseudorapidity coverage from 2.5 to 4.0. In this context, this thesis also presents a first calibration and identification of electrons in this forward region with the ITk, improving the background rejection at fixed identification efficiency and preparing future analyses with Run-4 data.

Version française

La clé de voûte du Modèle Standard, théorie faisant consensus en physique des particules, est la brisure spontanée de symétries, qui prédit l'apparition d'un nouveau boson scalaire, le boson de Higgs, découvert en 2012 au LHC par les expériences ATLAS et CMS, et engendre la masse des bosons vecteurs du secteur électrofaible : les bosons W et Z. Du fait de cette particularité, ces bosons sont les seules particules élémentaires à posséder trois états de polarisation, dont un état longitudinal, dans lequel ils se comportent à haute énergie comme les bosons de Goldstone de la théorie avant la brisure de symétrie. Étudier les événements VV produits par des processus électrofaibles offre donc un moyen inédit de sonder la structure même du Modèle Standard. Cette thèse présente la première mesure des états de polarisation longitudinale de l'un ou des deux bosons dans les processus WZjj-EW, à partir des données enregistrées par l'expérience ATLAS durant la totalité du Run-2 du LHC, correspondant à une luminosité intégrée de $140\,\mathrm{fb}^{-1}$ à une énergie dans le centre de masse de $13\,\mathrm{TeV}$, combinées à la première partie du Run-3, correspondant à $164\,\mathrm{fb}^{-1}$ à $13{,}6\,\mathrm{TeV}$. Des réseaux de neurones profonds sont entraînés à séparer les événements WZjj-EW du bruit de fond puis à dissocier les états de polarisation. Leurs sorties définissent un plan discriminant à deux dimensions, dans lequel un algorithme de partitionnement construit une segmentation de Voronoï optimisée qui maximise la sensibilité d'un ajustement par maximum de vraisemblance sur histogramme. La fraction d'événements où les deux bosons sont polarisés longitudinalement est mesurée à $f_{00} = 0{,}000^{+0{,}053}_{-0{,}000}$, pour une prédiction du Modèle Standard de $0{,}090$ incluant les corrections électrofaibles au deuxième ordre (NLO), et celle où le boson W est polarisé longitudinalement indépendamment de la polarisation du boson Z à $f_{0\mathrm{X}} = 0{,}007^{+0{,}094}_{-0{,}007}$, pour une prédiction de $0{,}276$, soit un écart de $2{,}5\sigma$ au Modèle Standard. Des limites supérieures à $95\,\%$ de niveau de confiance sont posées à $0{,}12$ ($0{,}21$ attendue) et $0{,}19$ ($0{,}46$ attendue), respectivement. En revanche, la fraction où le boson Z est polarisé longitudinalement indépendamment de la polarisation du boson W est mesurée à $f_{\mathrm{X}0} = 0{,}324^{+0{,}092}_{-0{,}085}$, pour une prédiction de $0{,}275$, avec une significance observée (attendue) de $4{,}0$ ($2{,}9$) déviations standards, apportant la première mise en évidence d'états où le boson Z est polarisé longitudinalement dans les événements WZjj-EW. La force du signal du processus vaut $\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}} = 1{,}06 \pm 0{,}18$, et sa section efficace fiducielle par saveur de lepton, $\sigma(\mathrm{WZjj\text{-}EW} \to \ell'\nu_\ell\ell\ell\mathrm{jj})$ où ℓ et ℓ' désignent chacun un électron ou un muon, est rapportée pour la première fois à $13{,}6\,\mathrm{TeV}$, $0{,}374 \pm 0{,}077\,(\mathrm{stat.}) \pm 0{,}053\,(\mathrm{syst.}) \pm 0{,}009\,(\mathrm{lumi.})\,\mathrm{fb}$. Ces mesures sont limitées par la statistique et bénéficieront du LHC à haute luminosité ainsi que du remplacement du trajectographe actuel par le nouvel Inner Tracker (ITk), qui étendra la couverture en pseudorapidité absolue de $2{,}5$ à $4{,}0$. Dans ce contexte, cette thèse présente également une première calibration et identification des électrons dans cette région vers l'avant avec l'ITk, améliorant la réjection du bruit de fond à efficacité d'identification fixée et préparant les futures analyses avec les données du Run-4.

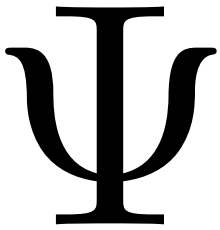


Table of Contents

Summary i
English version i
Version française ii

Introduction 1
English version 1
Summary of the thesis and personal contribution 2
Version française 5
Résumé de la thèse et contribution personnelle 6

1 The Standard Model 10
1.1 Formalism 11
1.1.1 Units 11
1.1.2 Special relativity 11
1.1.3 Field theory 11
1.1.4 Quantum field theory 12
1.2 Quantum electrodynamics 13
1.3 Quantum chromodynamics 15
1.4 Weak interaction 17
1.5 Brout–Englert–Higgs mechanism 19
1.6 Glashow–Weinberg–Salam theory 20
1.7 Higgs boson 22
1.8 Neutrino 24
1.9 Standard Model 25
1.10 Beyond the Standard Model 27
1.11 Vector boson polarisation 28

2 The Large Hadron Collider 32
2.1 Facility 33
2.2 Luminosity 35
2.3 Pile-up 36
2.4 Timeline 37
2.5 Collider physics 38

3 The ATLAS Experiment 42
3.1 Coordinate system 43
3.2 Inner Detector 44
3.2.1 Pixel Detector 44
3.2.2 Semi-Conductor Tracker 46
3.2.3 Transition Radiation Tracker 47
3.3 Calorimeters 47
3.3.1 Electromagnetic Calorimeter 48
3.3.2 Hadronic Calorimeter 51
3.3.2.1 Tile Calorimeter 54
3.3.2.2 Hadronic End-cap Calorimeter 55

3.3.3 Forward Calorimeter 56
3.4 Muon Spectrometer 57
3.4.1 Monitored Drift Tube Chamber 57
3.4.2 Cathode-Strip Chamber 59
3.4.3 Resistive Plate Chamber 60
3.4.4 Thin Gap Chamber 61
3.4.5 MicroMegas 61
3.5 Magnet System 62
3.6 Performance 63
3.6.1 Tracking 63
3.6.2 Energy 63
3.7 Trigger system 64
3.8 Luminosity measurement 65
3.9 Phase-II 66
3.9.1 Muon Spectrometer update 67
3.9.1.1 Thin Gap Chamber 67
3.9.1.2 Small-diameter Monitored Drift Tube 67
3.9.1.3 Barrel Inner Resistive Plate Chamber 67
3.9.1.4 High-η tagger 67
3.9.2 Inner Tracker 68
3.9.2.1 Strip Detector 68
3.9.2.2 Pixel Detector 70
3.9.3 High-Granularity Timing Detector 70

4 Machine Learning 72
4.1 Neural network 73
4.1.1 Neuron 73
4.1.2 Network 73
4.1.3 Loss function 74
4.1.4 Gradient descent 75
4.1.4.1 Stochastic gradient descent 75
4.1.4.2 Adaptive moment estimation 76
4.1.5 Back-propagation 77
4.1.6 Activation function 78
4.1.7 Training 79
4.1.7.1 Overfitting 80
4.1.7.2 k-folds 80
4.1.7.3 Imbalanced dataset 82
4.1.7.4 Metrics 82
4.1.8 Regularisation 83
4.1.9 Architecture 83
4.2 Boosted decision tree 84
4.2.1 Decision tree 84
4.2.2 Boosting 85
4.2.3 Application 86
4.3 Clustering 86
4.3.1 k-means 86
4.3.2 Voronoi diagram 87
4.4 Variable ranking and pruning 87
4.5 Bayesian optimisation 89
4.6 Monte Carlo 92

5 Electron in ATLAS 98
5.1 Reconstruction 98
5.1.1 Cluster formation 99
5.1.2 Tracking 101
5.1.3 Track–cluster matching 104
5.1.4 Photon conversion reconstruction 104
5.1.5 Super-cluster reconstruction 105
5.1.6 Ambiguity resolver 106
5.2 Calibration 107

5.3 Identification 111
5.3.1 The likelihood discriminant 111
5.3.2 Pile-up dependence 112
5.3.3 Input variables 112
5.3.4 Binning 113
5.3.5 High-E_{T} electrons 113
5.3.6 Working points 113
5.3.7 Efficiency measurements and data/MC corrections 116
5.3.8 Charge identification 116
5.3.9 DNN-based identification 117
5.4 Isolation 117
5.4.1 Calorimeter-based isolation 118
5.4.2 Track-based isolation 118
5.4.3 Operating points 119
5.4.4 Efficiency measurements and scale factors 120
5.5 Photon identification and isolation 120
5.6 Forward electrons 121
5.6.1 Cluster moments 122
5.6.2 Cut-based identification 122
5.6.3 Likelihood-based identification 123
5.6.4 Efficiency measurements and scale factors 124
5.7 Forward electrons in Phase-II 124
5.7.1 p_{T}-calibration 124
5.7.2 Identification 129
5.7.2.1 p_{T}-decorrelation 129
5.7.3 DNN output 132
5.7.4 Extrapolation at high p_{T} 138
5.7.5 Pile-up dependence correction 140
5.7.6 Robustness of the p_{T}-decorrelation 141
5.7.7 Outlook 141

6 Polarisation in vector boson scattering 144
6.1 Historical overview of diboson measurements 144
6.2 Vector boson scattering topology 149
6.2.1 Object reconstruction 150
6.2.1.1 Muons 150
6.2.1.1.1 Reconstruction 150
6.2.1.1.2 Identification 151
6.2.1.1.3 Isolation 152
6.2.1.1.4 Momentum calibration 152
6.2.1.1.5 Efficiency measurements and data/MC corrections 153
6.2.1.1.6 Run-3 developments 154
6.2.1.2 Jets 154
6.2.1.2.1 Hadrons in the detector 154
6.2.1.2.2 Particle-flow reconstruction 155
6.2.1.2.3 Jet clustering 157
6.2.1.2.4 Energy calibration 158
6.2.1.2.5 Pile-up jet suppression 161
6.2.1.2.6 Flavour tagging 162
6.2.1.2.7 From Run-2 to Run-3 162
6.2.1.3 Missing transverse momentum 163
6.2.1.4 Tau leptons 164
6.2.1.5 Overlap removal 165
6.2.2 Event reconstruction 165
6.2.2.1 Longitudinal neutrino momentum 166
6.2.3 Discriminating variables 166
6.2.3.1 Jet variables 167
6.2.3.2 Bosonic variables 167
6.2.3.3 Leptonic variables 168
6.2.4 Reference frames and polarisation observables 169
6.3 Polarisation density matrix 169

6.3.1 Single boson system 171
6.3.2 Diboson system 177
6.3.3 Production of bosons 178
6.4 Narrow-width approximation and gauge invariance 179
6.5 Monte Carlo generation of polarised samples 182
6.5.1 Leading-order generation 182
6.5.1.1 Additional jets 183
6.5.1.2 Automated one-loop 183
6.5.1.3 Limitations of leading order 183
6.5.2 NLO QCD corrections 184
6.5.3 NLO EW corrections 184
6.5.4 NLO EW calculation procedure 186
6.5.5 Impact of NLO predictions 187
6.5.6 State-of-the-art computation 188
6.6 Analysis setup 188
6.6.1 Fiducial phase-space 188
6.6.2 Object and event selection 189
6.6.2.1 Object selection 190
6.6.2.1.1 Muon selection 190
6.6.2.1.2 Electron selection 190
6.6.2.1.3 Jet selection 190
6.6.2.1.4 Missing transverse momentum selection 190
6.6.2.2 Overlap removal 191
6.6.2.3 Event selection 192
6.6.3 Processes involved 193
6.6.3.1 Signal definition 193
6.6.3.2 Background definition 196
6.6.4 Standard Model predictions and expected yields 196
6.7 Multivariate analysis 198
6.7.1 Generalities 199
6.7.2 Signal versus background DNN 201
6.7.3 Polarisation DNN 201
6.7.4 Input features 202
6.7.5 Evaluation of performance 204
6.7.6 Sanity checks 208
6.7.7 Discriminant plane and binning 209

7 Polarisation fraction measurement 214
7.1 Parametrisation 214
7.2 Systematic uncertainties 216
7.2.1 Experimental systematic uncertainties 217
7.2.1.1 Scale-factor versus four-momentum uncertainties 217
7.2.1.2 Electrons 218
7.2.1.3 Flavour tagging 218
7.2.1.4 Jets 218
7.2.1.5 Muons 219
7.2.1.6 Missing transverse momentum 219
7.2.1.7 Pile-up 219
7.2.1.8 Luminosity 219
7.2.2 Theoretical uncertainties 219
7.2.2.1 Misidentified leptons 227
7.2.2.2 Cross-section uncertainties 227
7.2.2.3 PDF uncertainties 227
7.2.2.4 QCD scale uncertainties 228
7.2.2.5 Modelling uncertainties 228
7.2.2.6 NLO EW correction uncertainties 229
7.2.3 Pruning of systematic uncertainties 230
7.3 Results 230
7.3.1 Goodness-of-fit 230
7.3.2 Post-fit distributions 231
7.3.3 Post-fit yields 236

7.3.4 Normalisation factors 236
7.3.5 Correlation matrix 236
7.3.6 2D likelihood scans 242
7.3.7 Nuisance parameters ranking 242
7.3.8 Measurements 245
7.3.9 Impact of systematics 246
7.3.10 Fiducial cross-section 247
7.3.11 Upper limits 247
7.3.12 Consistency between polarisation states 251
7.3.13 Cross-checks 251
7.4 Summary 252

Conclusion **254**

Appendices **256**

A Vector boson scattering cross-section calculation **258**
A.1 $\mathrm{W^+W^-} \to \mathrm{W^+W^-}$ 261
A.2 $\mathrm{W^\pm W^\pm} \to \mathrm{W^\pm W^\pm}$ 261
A.3 $\mathrm{W^\pm Z} \to \mathrm{W^\pm Z}$ 262
A.4 $\mathrm{ZZ} \to \mathrm{ZZ}$ 262
A.5 $\mathrm{W^\pm W^\mp} \to \mathrm{ZZ}$ 263
A.6 Higgs mass constraint from unitarity 263

B Polarisation formalism for joint WZ system **266**
B.1 Joint spin density matrix 266
B.1.1 Unpolarised and single-boson polarisations 266
B.1.2 Joint spin correlations 267
B.1.3 Joint spin density matrix decomposition 268
B.2 Physical polarisation fractions 270
B.3 Decay density matrix 271
B.4 Joint angular distribution 271

C Polarised vector boson Feynman diagrams **276**
C.1 Charged current 276
C.2 Neutral current 276
C.3 Vector boson decay 277
C.4 Decay widths and branching fractions 279

D One-loop leading-logarithm corrections **280**

E Data-driven estimation of the misidentified-lepton background **284**
E.1 The matrix method 284
E.1.1 Principle 284
E.1.2 Determination of the fake factors 286
E.1.3 $|\eta|$-dependent correction 287
E.1.4 Fake factors by fake-lepton origin 287
E.1.5 Closure test and systematic uncertainties 287
E.2 The data/MC scale factor method 288
E.2.1 Principle 288
E.2.2 Control regions 289
E.2.3 Determination of the correction factors 289
E.2.4 Application 290

F List of Monte Carlo samples **292**
F.1 Signal and WZjj-EW modelling 292
F.1.1 The WZjj-EW full samples 293
F.1.2 Polarised WZjj-EW samples 293
F.1.2.1 Sherpa 3.0 (nominal) 293
F.1.2.2 MadGraph (alternative) 294
F.1.2.3 Accuracy and limitations 294

F.2 Background processes 296

G 3D template 298

H Bayesian optimisation of the network hyperparameters 300

I Input variable distributions 306

J Correlation matrices 320

K Variable ranking 324
K.1 Variable ranking for the full set of variables 324
K.2 Variable ranking for the pruned set of variables 329
K.3 ROC-curve comparisons: fold stability and variable pruning 335

L Generator effects on DNN distributions 338
L.1 Signal versus background DNN 338
L.2 Polarisation DNN 341
L.2.1 Joint polarisation 00 DNN 341
L.2.2 Single polarisation $0\mathrm{X}$ DNN 343
L.2.3 Single polarisation $\mathrm{X}0$ DNN 345

M Binning optimisation 348
M.1 Significance summary 348

N Voronoi binning in depth 350
N.1 Joint polarisation: the 00 plane 351
N.2 Single polarisation: the $0\mathrm{X}$ plane 353
N.3 Single polarisation: the $\mathrm{X}0$ plane 355

O Statistical methods for the template fit 358
O.1 Likelihood-based inference 358
O.2 Statistical tests and significance 360
O.3 The binned likelihood model of the fit 361
O.4 Parameter estimation 362
O.5 Significance of the measurement 362
O.6 Confidence intervals and limits 363
O.7 The χ^2 distribution and goodness-of-fit 364
O.8 Mapping to the results of the main text 365

P Theoretical systematics representation 366

Q Pruning of the systematic uncertainties 378

R Post-fit nuisance-parameter pulls 386

S Profile-likelihood scans of the polarisation fits 390

T Injection tests 392
T.1 Closure of the polarisation-fraction measurement 392
T.1.1 Nominal fractions 392
T.1.2 Varied fractions 392
T.1.3 Transverse–transverse component increased 393
T.1.4 Reduction of WZjj-QCD contribution 393
T.2 Effect of the component with τ leptons 394

U Decorrelation of the QCD scale uncertainties for WZjj-QCD 396

V Change of fit parametrisation for normalisation factors 398

W Impact of object systematic uncertainties on the measurement 400

X Measurement of the doubly transverse polarisation state 402

Y Impact of the polarisation states' composition in fitted templates **404**

Z Scotogenic models and their parameter-space exploration **406**
Z.1 The scotogenic mechanism 406
Z.1.1 Lagrangian and mass spectrum 407
Z.1.2 Radiative neutrino masses and the Casas–Ibarra parametrisation 407
Z.1.3 Constraints and numerical tools 408
Z.2 Parameter-space exploration by active learning 408
Z.2.1 Oracle scoring 409
Z.2.2 Discriminator and query strategy 409
Z.2.3 Toy-model validation 410
Z.2.4 Performance and outlook 410

Bibliography **412**

Acronyms **436**

Acknowledgements **444**

α Introduction

English version

Particle physics is the study of the fundamental constituents of matter and their interactions. The field, developed over the last century, has led to the Standard Model (SM), a theory built over decades of discoveries, experiments, and theoretical developments. Extremely predictive, it has been tested in a wide range of experiments and has been confirmed to an extraordinary degree of precision. The most recent and spectacular confirmation was the discovery of the Higgs boson in 2012 [1, 2], which was the last missing piece of the SM. The discovery of this particle was a major milestone in particle physics, confirming the mechanism of spontaneous symmetry breaking (SSB) and the generation of mass for fundamental particles.

No experimental result (beyond neutrino oscillations) has ever been found to contradict the SM, and it is widely accepted as the best description of the fundamental particles and their interactions to this day. However, the SM is not the final theory (if such a thing exists) and it is known to be incomplete. It does not, for example, provide a candidate for dark matter (DM), it does not explain the matter–antimatter asymmetry in the Universe, and it does not include gravity. Therefore, beyond the Standard Model (BSM) theories have been proposed to address these shortcomings, and many of them predict new particles or interactions that could be observed in experiments. The search for BSM physics is one of the main goals of current and future particle physics experiments.

Apart from new particles, the study of the properties of known particles predicted by the SM is also a powerful probe for new physics by pushing the theory to its limits. A critical example is the study of electroweak (EW) symmetry breaking, a mechanism through which the Higgs field acquires a vacuum expectation value (VEV) giving a physical scalar boson, the Higgs, conferring mass on the fermions and the weak gauge bosons at the same time. This central phenomenon therefore provides a wide range of properties to test, which are tightly constrained by the theory.

A particularly interesting property, only recently accessible experimentally, concerns the vector bosons of the EW sector. These bosons, namely the Z and W, are the only massive vector bosons, a property they acquire through EW SSB. This has a direct impact on the number of accessible polarisation states. The additional degree of freedom results in a third polarisation state, unique to these two bosons, unlike the gluon or the photon. Bosons in this state behave like their corresponding Goldstone bosons of the theory prior to SSB, and exhibit a characteristic energy dependence as stated by the Goldstone boson equivalence theorem (GBET) [3, 4]. This represents a unique probe of the theory in the regime where the effects of symmetry breaking are most directly manifest. In addition, the scattering of longitudinally polarised vector bosons is central to the internal consistency of the SM, as without a Higgs boson, the amplitude of such a process would grow as a function of the centre-of-mass (CM) energy and violate unitarity. Only the presence of a scalar boson can cancel this growth.

To be produced and potentially observed, this process requires high-energy incoming particles that must independently radiate two vector bosons, which subsequently scatter. As they are massive bosons, they decay, either hadronically or leptonically, before reaching any active material. The event and thus the polarisation states have to be inferred from the properties of the resulting decay products, among which a few observables are highly sensitive to this phenomenon.

However, processes of this kind are exceedingly rare and have only been observed recently, at the CERN Large Hadron Collider (LHC) [5–7], a circular proton–proton (pp) collider where collisions occur and are recorded by four experiments along the ring. Among them, only the multipurpose detectors, the ATLAS and CMS experiments [8, 9], recorded those events with sufficient acceptance and luminosity during the second period of data taking, Run-2, from 2015 to 2018. The cross-sections measured are

among the smallest ever recorded at a collider experiment. These processes first had to be disentangled from the considerable background sharing the same final state.

In the meantime, the polarisation of pairs of vector bosons produced inclusively, thus without scattering, was also measured for the first time [10, 11], paving the way for this thesis. Now that both of these measurements have been achieved, and with the larger dataset provided by the first part of the third data-taking period of the LHC, Run-3, the polarisation states can be measured in vector boson scattering (VBS) events for the first time. Among the three possible channels, ZZ, WZ, and WW, the WZ one is targeted as it provides a good compromise between cross-section and final-state purity. The fully leptonic decay is used, giving three charged leptons and missing transverse momentum (MET) in the final state, accompanied by the two forward jets characteristic of the EW production, denoted WZjj-EW. The dominant irreducible background is the strong production of the same final state.

The same separation of signal and background has to be performed to isolate the signal events from the large background contamination. In addition, a polarisation separation has to be developed, following the one done in the inclusive measurements but taking into account the very different topology of the two processes. For this purpose, machine learning (ML) techniques using deep neural networks (DNNs) are used to build a template that maximises the polarisation sensitivity. A template fit then measures both the polarisation fraction and the cross-section of this process for both data-taking periods. Other advanced ML methods are also used, such as Bayesian optimisation (BO) and clustering algorithms, to provide the tools best suited to the measurement.

Three separate measurements are performed: one with both bosons longitudinally polarised, the most challenging one due to the low expected yields, and two in which a single boson is longitudinally polarised regardless of the polarisation of the other, the W in one case and the Z in the other. The latter two are more sensitive and represent the main objective of this thesis.

However, even with the largest dataset ever recorded in particle physics, these measurements are only a first step towards the full measurement that will be carried out with the High-Luminosity LHC (HL-LHC) [12], extending the dataset by a factor of ten. One of the main upgrades of the ATLAS detector for this new era will be the replacement of the current Inner Detector (ID) by the Inner Tracker (ITk) [13, 14], extending the coverage of track reconstruction in the forward regions from $|\eta| = 2.5$ to $|\eta| = 4.0$. This will greatly benefit such measurements by enhancing the reconstruction of forward jets, pile-up rejection, and the identification of electrons in the forward region. In this context, this thesis also presents the calibration and identification of electrons in the forward region with the ITk, carried out within the e/γ Combined Performance group.

The structure of this thesis follows the path from theory to measurement. Chapter 1 introduces the SM in detail, the particles it predicts, its strengths, and its limitations, and gives a first account of the polarisation of vector bosons. Chapter 2 describes the LHC and how it delivers collisions to the experiments, together with the physics of colliders and of hadron colliders in particular. Chapter 3 gives an extensive description of the ATLAS detector and of its sub-detectors, and of how they together record the data produced by the collisions. Chapter 4 provides the ML tools required to follow the analysis. The reconstructed objects are described in Chapter 5 for electrons and photons, which also presents a first preliminary study of electron identification in the forward region with the ITk, and in Chapter 6 for muons, jets, MET, and τ leptons. The latter also gives a historical overview of diboson measurements, the theoretical framework, the event selection, and the construction of the template used to extract the polarisation fractions. Finally, Chapter 7 details the statistical framework and presents the results.

As the results obtained are the first of their kind, they are another milestone in the study of the EW sector and will pave the way for future measurements at the HL-LHC. The ITk will provide a unique opportunity to enhance the sensitivity of such measurements and will be a key element in the future of particle physics.

Summary of the thesis and personal contribution

The objective of this thesis is to measure the polarisation of vector bosons in VBS events, in particular in the WZ channel with a longitudinal polarisation state. Such a measurement is performed here for the first time in this channel, and is among the first of its kind in any channel. Only accessible with the large dataset provided by the LHC and its pp collisions, it is a unique opportunity to probe the SM and its EW sector prior to SSB. The measurement uses the full dataset collected by the ATLAS experiment during the second data-taking period of the LHC, from 2015 to 2018 at $\sqrt{s} = 13\,\mathrm{TeV}$, corresponding to an integrated luminosity of $140\,\mathrm{fb}^{-1}$, combined with the first part of the third data-taking period, from 2022 to 2024 at $\sqrt{s} = 13.6\,\mathrm{TeV}$, corresponding to $164\,\mathrm{fb}^{-1}$. A template fit to the data is performed, where the templates are obtained from ML techniques using DNNs. Three different polarisation states are considered: both bosons longitudinally polarised, and a single boson longitudinally polarised regardless of the polarisation

of the other, the W in one case and the Z in the other. The results obtained are a first step towards the full exploitation of the HL-LHC and its upgraded ATLAS detector, in particular the ITk, which will provide a unique opportunity to enhance the sensitivity of such measurements and will be a key element in the future of particle physics. In this context, this thesis is divided into seven chapters, ordered logically, each providing the background and tools needed to understand the measurement.

Chapter 1. The SM is introduced in detail following the historical development of the field through theories, discoveries, and experiments. Each particle predicted by this framework is described with its properties and interactions. The EW sector is detailed, in particular the mechanism of SSB and the generation of mass for the vector bosons. The polarisation of vector bosons is introduced and its importance in the study of the SM is highlighted. Limitations of the SM are also discussed, motivating the search for BSM physics, for which the study of the polarisation of vector bosons in VBS events is a powerful probe.

Chapter 2. The source of the data used in this thesis, the LHC and its pp collisions, is detailed. The physics of colliders is introduced, in particular that of hadron colliders, and the different components of the LHC are described. The data-taking periods are detailed, as well as the different experiments along the ring and their role in the study of particle physics.

Chapter 3. The ATLAS detector is described in detail, with its different sub-detectors and their role in the reconstruction of the events produced by the pp collisions. The Data AcQuisition (DAQ) system is also detailed, as well as the trigger system and its role in selecting the events of interest. The initial configuration of the experiment is described, as well as the upgrades performed during the first, second, and third data-taking periods, but also the future upgrades planned for the HL-LHC.

Chapter 4. The ML techniques used in this thesis are introduced, with a focus on neural networks (NNs), describing their operation, their various features, as well as the different metrics used to evaluate the performance of the models. Boosted decision trees (BDTs) are also introduced as the approach used in previous measurements. BO is also introduced as a method to optimise the hyperparameters of the models. A presentation of clustering algorithms is also provided, as they are used in the construction of the template for the measurement. Finally, the Monte Carlo (MC) generators used to produce the simulated samples are introduced.

Chapter 5. The electron and photon workflow within ATLAS is detailed, with the different algorithms used to identify, reconstruct, calibrate, and isolate these objects. The first part is focused on the central regions where the ID provides tracks for charged particles, while the second part is focused on the forward regions where no tracking is available. The current status of electron identification in this challenging region is presented. It is followed by the first preliminary study of the identification (and calibration) of electrons in the forward region with the ITk, as part of my work in the e/γ Combined Performance group. I provided a first transverse-momentum calibration for electrons in this region and working points (WPs) to set the identification efficiency with a p_{T}-decorrelation to make them flat in p_{T}. This work was cross-checked with different pile-up conditions. This workflow was implemented in Athena, the ATLAS software framework, providing the first preliminary forward-electron identification with the ITk.

Chapter 6. The VBS process is introduced in detail, with first a historical overview of the previous measurements performed in this channel, and the theoretical framework used to describe it. The different final states are described, with a focus on the WZ channel, and the different backgrounds that can mimic the signal are detailed. The event selection is presented, as well as the different variables used to discriminate between signal and background. The construction of the template for the measurement, which I developed, is detailed, with the use of ML techniques to enhance the polarisation sensitivity. A novel binning method is also introduced to optimise the sensitivity of the measurement using clustering and yielding Voronoi bins.

Chapter 7. The statistical framework, which I developed to perform the measurement, is detailed, with the different methods used to extract the polarisation fraction and the cross-section of the process. The results obtained are presented, with a discussion on their implications and their comparison with the theoretical predictions. The details of all systematic uncertainties are also provided, with a discussion on their impact on the measurement and the methods used to evaluate them.

Appendix Z. A class of scotogenic BSM models is introduced, which addresses some of the limitations of the SM, in particular the absence of a DM candidate and the origin of neutrino masses. The models are briefly described, with their particle content and interactions, and the constraints that they have to satisfy are detailed. The high-dimensional parameter space of these models is explored with an

active learning (AL) strategy built on a DNN, which is then compared to a more traditional Markov chain Monte Carlo (MCMC) approach. The AL methodology is detailed, with the different steps of the scan, and its performance is then presented. During an internship on supersymmetric models, I reused the AL strategy, following the scoring method developed in [15]. In collaboration with theoretical physicists, we adapted the method to their models and provided a new way of scanning high-dimensional parameter spaces in the context of BSM theories. The work was published in [16].

Version française

La physique des particules est l'étude des constituants fondamentaux de la matière et de leurs interactions. Ce domaine, développé au cours du siècle dernier, a conduit à l'élaboration du Modèle Standard, une théorie bâtie au fil de décennies de découvertes, d'expériences et de développements théoriques. Extrêmement prédictif, il a été éprouvé dans un large éventail d'expériences et confirmé avec un degré de précision extraordinaire. La confirmation la plus récente et la plus spectaculaire en a été la découverte du boson de Higgs en 2012 [1, 2], dernière pièce manquante du Modèle Standard. La découverte de cette particule a constitué une étape majeure de la physique des particules, confirmant le mécanisme de brisure spontanée de symétrie et la génération de la masse des particules fondamentales.

Aucun résultat expérimental (hormis les oscillations de neutrinos) n'a jamais contredit le Modèle Standard, et celui-ci est aujourd'hui largement admis comme la meilleure description des particules fondamentales et de leurs interactions. Cependant, le Modèle Standard n'est pas la théorie ultime (si tant est qu'une telle théorie existe) et il est connu pour être incomplet. Il ne fournit pas, par exemple, de candidat à la matière noire, il n'explique pas l'asymétrie matière-antimatière de l'Univers et il n'inclut pas la gravitation. Des théories au-delà du Modèle Standard ont donc été proposées pour remédier à ces limitations, et nombre d'entre elles prédisent de nouvelles particules ou interactions susceptibles d'être observées expérimentalement. La recherche de physique au-delà du Modèle Standard constitue l'un des principaux objectifs des expériences de physique des particules actuelles et futures.

Outre la recherche de nouvelles particules, l'étude des propriétés des particules connues prédites par le Modèle Standard constitue également une sonde puissante de la nouvelle physique, en poussant la théorie dans ses retranchements. Un exemple essentiel est l'étude de la brisure de la symétrie électrofaible, mécanisme par lequel le champ de Higgs acquiert une valeur moyenne dans le vide non nulle, donnant naissance à un boson scalaire physique, le boson de Higgs, tout en conférant une masse aux fermions et aux bosons de jauge faibles. Ce phénomène central offre ainsi un large éventail de propriétés à tester, très fortement contraintes par la théorie.

Une propriété particulièrement intéressante, accessible expérimentalement seulement depuis peu, concerne les bosons vecteurs du secteur électrofaible. Ces bosons, à savoir le Z et le W, sont les seuls bosons vecteurs massifs, propriété qu'ils acquièrent par la brisure spontanée de la symétrie électrofaible. Cela a une conséquence directe sur le nombre d'états de polarisation accessibles. Ce degré de liberté supplémentaire se traduit par un troisième état de polarisation, propre à ces deux bosons, contrairement au gluon ou au photon. Les bosons dans cet état se comportent comme les bosons de Goldstone correspondants de la théorie avant la brisure spontanée de symétrie, et présentent une dépendance caractéristique en énergie, comme l'énonce le théorème d'équivalence des bosons de Goldstone [3, 4]. Cela constitue une sonde unique de la théorie dans le régime où les effets de la brisure de symétrie se manifestent le plus directement. De plus, la diffusion de bosons vecteurs polarisés longitudinalement est essentielle à la cohérence interne du Modèle Standard : en l'absence de boson de Higgs, l'amplitude d'un tel processus croîtrait avec l'énergie dans le centre de masse et violerait l'unitarité. Seule la présence d'un boson scalaire permet d'annuler cette croissance.

Pour être produit et éventuellement observé, ce processus requiert des particules incidentes de haute énergie rayonnant indépendamment deux bosons vecteurs, lesquels diffusent ensuite l'un sur l'autre. Étant massifs, ces bosons se désintègrent, hadroniquement ou leptoniquement, avant d'atteindre toute matière active. L'événement, et donc les états de polarisation, doivent par conséquent être déduits des propriétés des produits de désintégration, parmi lesquelles quelques observables sont fortement sensibles à ce phénomène.

Ce type de processus est toutefois très rare et n'a été observé que récemment, auprès du Grand collisionneur de hadrons (LHC) [5–7] du CERN, un collisionneur circulaire proton-proton où les collisions se produisent et sont enregistrées par quatre expériences réparties le long de l'anneau. Parmi elles, seuls les détecteurs polyvalents que sont les expériences ATLAS et CMS [8, 9] ont enregistré ces événements avec une acceptance et une luminosité suffisantes, durant la deuxième période de prise de données, le Run-2, de 2015 à 2018. Les sections efficaces obtenues comptent parmi les plus faibles jamais mesurées auprès d'un collisionneur. Ces processus ont d'abord dû être extraits du bruit de fond considérable partageant le même état final.

Entre-temps, la polarisation de paires de bosons vecteurs produites de manière inclusive, c'est-à-dire sans diffusion, a également été mesurée pour la première fois [10, 11], ouvrant la voie à cette thèse. Maintenant que ces deux mesures ont été réalisées, et grâce au jeu de données plus important fourni par la première partie de la troisième période de prise de données du LHC, le Run-3, les états de polarisation peuvent être mesurés dans des événements de diffusion de bosons vecteurs pour la première fois. Parmi les trois canaux possibles, ZZ, WZ et WW, celui en WZ est retenu car il offre un bon compromis entre

section efficace et pureté de l'état final. La désintégration purement leptonique est utilisée, conduisant à trois leptons chargés et à une impulsion transverse manquante dans l'état final, accompagnés des deux jets vers l'avant caractéristiques de la production électrofaible, notée WZjj-EW. Le bruit de fond irréductible dominant est la production du même état final par interaction forte.

La même séparation du signal et du bruit de fond doit être effectuée afin d'isoler les événements de signal de l'importante contamination par le bruit de fond. Il faut en outre développer une séparation des polarisations, à l'image de celle réalisée dans les mesures inclusives, mais en tenant compte de la topologie très différente des deux processus. À cette fin, des techniques d'apprentissage automatique reposant sur des réseaux de neurones profonds sont employées pour construire un *template* maximisant la sensibilité à la polarisation, en vue d'un ajustement de *templates* permettant de mesurer à la fois la fraction de polarisation et la section efficace de ce processus, pour les deux périodes de prise de données. D'autres méthodes avancées d'apprentissage automatique sont également utilisées, telles que l'optimisation bayésienne et des algorithmes de partitionnement (*clustering*), afin de disposer des outils les mieux adaptés à la mesure.

Trois mesures distinctes sont réalisées : la première avec les deux bosons polarisés longitudinalement, la plus difficile en raison du très faible nombre d'événements attendus, et deux autres où un seul boson est polarisé longitudinalement, indépendamment de la polarisation de l'autre, le W dans un cas et le Z dans l'autre. Ces deux dernières sont plus sensibles et constituent l'objectif principal de cette thèse.

Cependant, même avec le plus grand jeu de données jamais enregistré en physique des particules, ces mesures ne constituent qu'une première étape vers la mesure complète qui sera réalisée auprès du LHC à haute luminosité (HL-LHC) [12], lequel multipliera par dix la taille du jeu de données. L'une des principales améliorations du détecteur ATLAS pour cette nouvelle ère sera le remplacement du détecteur interne actuel par le nouveau trajectographe interne, l'ITk (*Inner Tracker*) [13, 14], étendant la couverture de la reconstruction des traces dans les régions vers l'avant de $|\eta| = 2{,}5$ à $|\eta| = 4{,}0$. Cela bénéficiera grandement à de telles mesures en améliorant la reconstruction des jets vers l'avant, le rejet de l'empilement et l'identification des électrons dans la région vers l'avant. Dans ce contexte, cette thèse présente également la calibration et l'identification des électrons dans la région vers l'avant avec l'ITk, réalisées au sein du groupe Combined Performance e/γ.

La structure de cette thèse suit le cheminement de la théorie vers la mesure. Le chapitre 1 introduit en détail le Modèle Standard, les particules qu'il prédit, ses forces mais aussi ses limites, et donne un premier aperçu de la polarisation des bosons vecteurs. Le chapitre 2 décrit le LHC et la manière dont il délivre les collisions aux expériences, ainsi que la physique des collisionneurs et en particulier des collisionneurs hadroniques. Le chapitre 3 présente une description détaillée du détecteur ATLAS et de ses sous-détecteurs, et de la façon dont ils enregistrent ensemble les données produites par les collisions. Le chapitre 4 fournit les outils d'apprentissage automatique nécessaires à la compréhension de l'analyse. Les objets reconstruits sont décrits dans le chapitre 5 pour les électrons et les photons, qui présente également une première étude préliminaire de l'identification des électrons dans la région vers l'avant avec l'ITk, et dans le chapitre 6 pour les muons, les jets, l'impulsion transverse manquante et les leptons τ. Ce dernier donne aussi un aperçu historique des mesures de dibosons, le cadre théorique, la sélection des événements et la construction du *template* utilisé pour extraire les fractions de polarisation. Enfin, le chapitre 7 détaille le cadre statistique et présente les résultats.

Les résultats obtenus étant les premiers de ce type, ils constituent une nouvelle étape marquante dans l'étude du secteur électrofaible et ouvriront la voie à de futures mesures auprès du HL-LHC. L'ITk offrira une occasion unique d'améliorer la sensibilité de telles mesures et sera un élément clé de l'avenir de la physique des particules.

Résumé de la thèse et contribution personnelle

L'objectif de cette thèse est de mesurer la polarisation des bosons vecteurs dans des événements de diffusion de bosons vecteurs, en particulier dans le canal WZ avec un état de polarisation longitudinal. Une telle mesure est réalisée ici pour la première fois dans ce canal, et fait partie des toutes premières de ce type, quel que soit le canal. Accessible uniquement grâce au grand jeu de données fourni par le LHC et ses collisions proton-proton, elle offre une occasion unique de sonder le Modèle Standard et son secteur électrofaible avant la brisure spontanée de symétrie. La mesure utilise l'intégralité du jeu de données collecté par l'expérience ATLAS durant la deuxième période de prise de données du LHC, de 2015 à 2018 à $\sqrt{s} = 13\,\mathrm{TeV}$, correspondant à une luminosité intégrée de $140\,\mathrm{fb}^{-1}$, combiné à la première partie de la troisième période de prise de données, de 2022 à 2024 à $\sqrt{s} = 13{,}6\,\mathrm{TeV}$, correspondant à $164\,\mathrm{fb}^{-1}$. Un ajustement de *templates* aux données est réalisé, les *templates* étant obtenus à partir de techniques d'apprentissage automatique reposant sur des réseaux de neurones profonds. Trois états de polarisation différents sont considérés : les deux bosons polarisés longitudinalement, et un seul boson

polarisé longitudinalement indépendamment de la polarisation de l'autre, le W dans un cas et le Z dans l'autre. Les résultats obtenus constituent une première étape vers la pleine exploitation du HL-LHC et de son détecteur ATLAS amélioré, en particulier de l'ITk, qui offrira une occasion unique d'améliorer la sensibilité de telles mesures et sera un élément clé de l'avenir de la physique des particules. Dans ce contexte, cette thèse est divisée en sept chapitres, chacun fournissant, dans un ordre logique, le contexte et les outils nécessaires à la compréhension de la mesure réalisée.

Chapitre 1. Le Modèle Standard est introduit en détail en suivant le développement historique du domaine à travers les théories, les découvertes et les expériences. Chaque particule prédite par ce cadre est décrite avec ses propriétés et ses interactions. Le secteur électrofaible est détaillé, en particulier le mécanisme de brisure spontanée de symétrie et la génération de la masse des bosons vecteurs. La polarisation des bosons vecteurs est introduite et son importance dans l'étude du Modèle Standard est soulignée. Les limites du Modèle Standard sont également discutées, motivant la recherche de physique au-delà du Modèle Standard, pour laquelle l'étude de la polarisation des bosons vecteurs dans les événements de diffusion de bosons vecteurs constitue une sonde puissante.

Chapitre 2. La source des données utilisées dans cette thèse, le LHC et ses collisions proton-proton, est détaillée. La physique des collisionneurs est introduite, en particulier celle des collisionneurs hadroniques, et les différents composants du LHC sont décrits. Les périodes de prise de données sont détaillées, ainsi que les différentes expériences réparties le long de l'anneau et leur rôle dans l'étude de la physique des particules.

Chapitre 3. Le détecteur ATLAS est décrit en détail, avec ses différents sous-détecteurs et leur rôle dans la reconstruction des événements produits par les collisions proton-proton. Le système d'acquisition de données est également détaillé, ainsi que le système de déclenchement et son rôle dans la sélection des événements d'intérêt. La configuration initiale de l'expérience est décrite, de même que les améliorations apportées durant les première, deuxième et troisième périodes de prise de données, mais aussi les améliorations futures prévues pour le HL-LHC.

Chapitre 4. Les techniques d'apprentissage automatique utilisées dans cette thèse sont introduites, en mettant l'accent sur les réseaux de neurones, en décrivant leur fonctionnement, leurs différentes caractéristiques, ainsi que les différentes métriques utilisées pour évaluer les performances des modèles. Les arbres de décision boostés sont également présentés, en tant que méthode employée dans les mesures précédentes. L'optimisation bayésienne est aussi introduite comme méthode d'optimisation des hyperparamètres des modèles. Une présentation des algorithmes de partitionnement (*clustering*) est également fournie, ceux-ci étant utilisés dans la construction du *template* pour la mesure. Enfin, les générateurs Monte-Carlo utilisés pour produire les échantillons simulés sont introduits.

Chapitre 5. La chaîne de traitement des électrons et des photons au sein d'ATLAS est détaillée, avec les différents algorithmes utilisés pour identifier, reconstruire, calibrer et isoler ces objets. La première partie se concentre sur les régions centrales, où le détecteur interne fournit les traces des particules chargées, tandis que la seconde porte sur les régions vers l'avant, où aucune information de trajectographie n'est disponible. L'état actuel de l'identification des électrons dans cette région difficile est présenté. Vient ensuite la première étude préliminaire de l'identification (et de la calibration) des électrons dans la région vers l'avant avec l'ITk, réalisée dans le cadre de mon travail au sein du groupe Combined Performance e/γ. J'ai ainsi fourni une première calibration en impulsion transverse pour les électrons de cette région, ainsi que des points de travail définissant l'efficacité d'identification, décorrélée en p_{T} afin de la rendre plate en p_{T}. Ce travail a été vérifié pour différentes conditions d'empilement. Cette chaîne de traitement a été implémentée dans Athena, le cadre logiciel d'ATLAS, fournissant la première identification préliminaire des électrons vers l'avant avec l'ITk.

Chapitre 6. Le processus de diffusion de bosons vecteurs est introduit en détail, avec d'abord un aperçu historique des mesures précédentes réalisées dans ce canal, puis le cadre théorique utilisé pour le décrire. Les différents états finaux sont décrits, avec un accent sur le canal WZ, et les différents bruits de fond susceptibles d'imiter le signal sont détaillés. La sélection des événements est présentée, ainsi que les différentes variables utilisées pour discriminer le signal du bruit de fond. La construction du *template* pour la mesure, que j'ai développé, est détaillée, avec l'utilisation de techniques d'apprentissage automatique pour renforcer la sensibilité à la polarisation. Une nouvelle méthode de *binning* est également introduite pour optimiser la sensibilité de la mesure, reposant sur un partitionnement et produisant des cellules de Voronoï.

Chapitre 7. Le cadre statistique, que j'ai développé pour réaliser la mesure, est détaillé, avec les différentes méthodes utilisées pour extraire la fraction de polarisation et la section efficace du processus. Les résultats obtenus sont présentés, avec une discussion sur leurs implications et leur comparaison avec les prédictions théoriques. Le détail de toutes les incertitudes systématiques est également fourni, avec une discussion sur leur impact sur la mesure et sur les méthodes utilisées pour les évaluer.

Annexe Z. Une classe de modèles scotogéniques au-delà du Modèle Standard est introduite, répondant à certaines des limitations du Modèle Standard, en particulier l'absence de candidat à la matière noire et l'origine des masses des neutrinos. Les modèles sont brièvement décrits, avec leur contenu en particules et leurs interactions, et les contraintes qu'ils doivent satisfaire sont détaillées. L'espace des paramètres de grande dimension de ces modèles est exploré à l'aide d'une stratégie d'apprentissage actif reposant sur un réseau de neurones profond, qui est ensuite comparée à une approche plus traditionnelle par méthode de Monte-Carlo par chaînes de Markov. La méthodologie d'apprentissage actif est détaillée, avec les différentes étapes du balayage, puis sa performance est présentée. Lors d'un stage sur les modèles supersymétriques, j'ai réutilisé la stratégie d'apprentissage actif, en suivant la méthode de score développée dans [15]. En collaboration avec des physiciens théoriciens, nous avons adapté la méthode à leurs modèles et proposé une nouvelle manière de balayer les espaces de paramètres de grande dimension dans le contexte des théories au-delà du Modèle Standard. Ce travail a été publié dans [16].

1 The Standard Model

Physics is the branch of science focused on the nature and properties of matter and energy. One of the most fundamental quests in human history has been to understand the structure of matter's fundamental constituents. Driven by human curiosity to understand the world, this quest can be traced back to several centuries before the Common Era. Various civilisations around the globe have developed philosophical theories about what matter is composed of. In particular, among the Greeks, Democritus named these fundamental constituents *atomos*, meaning uncuttable in ancient Greek. The first milestones came only in the XVIII century, with research on various elements that culminated in the Mendeleev periodic table in 1871 [17].

In 1897, the electron "e" was observed by Thomson in cathode-ray experiments[Nobel] [18], paving the way for subatomic physics. Later, in 1900, Planck introduced the idea that energy could only take discrete levels to explain the blackbody radiation problem, thus creating the concept of quantisation[Nobel] [19]. Rutherford demonstrated in 1911 that atoms are mostly vacuum via the Geiger–Marsden experiment, consisting of shooting α particles into gold (Au) foils [20]. The atom was described as a nucleus with electrons orbiting around it. Based on this, Bohr developed his own model in 1913, with electrons orbiting in quantised orbits[Nobel] [21]. De Broglie suggested in 1924 that electrons have wave-like properties in addition to their original particle properties, coining the wave–particle duality[Nobel] [22].

Modern quantum physics emerged in 1925, following the introduction of matrix mechanics by Heisenberg[Nobel], Born[Nobel], and Jordan [23–25], who built upon the previous work of de Broglie. This development provided a mathematical description of quantised energy levels, furthering Bohr's model. Only one year later, Schrödinger invented wave mechanics with his famous equation[Nobel] [26]:

$$\mathrm{i}\hbar \frac{\partial}{\partial t}\Psi(\vec{r},t) = \left[-\frac{\hbar^2}{2m}\nabla^2 + V(\vec{r},t)\right]\Psi(\vec{r},t)\,, \tag{1.1}$$

generalising de Broglie's idea. Dirac later demonstrated that both approaches were equivalent.

In parallel with the advancement made in particle physics, Einstein developed in 1905 the theory of special relativity based on the principle that the laws of physics are the same in all inertial reference frames [27]. It also postulates that the speed of light c is constant, as hinted by Maxwell's equations in 1865 [28]. A consequence of this is the relation linking the energy and the mass, $E = mc^2$. Minkowski, in 1908, set up the framework of space-time with a four-vector [29].

Unifying the previously developed quantum theory and special relativity was appealing. This led to the Klein–Gordon equation [30, 31]:

$$\left(\Box + \frac{m^2c^2}{\hbar^2}\right)\phi = 0\,, \tag{1.2}$$

developed in 1926. However, this equation has negative energy solutions and can lead to negative probability densities. Instead, Dirac in 1928 used a four-component column vector, known as a *Dirac spinor*, to achieve the unification[Nobel] [32]. This results in the Dirac equation:

$$(\mathrm{i}\hbar\gamma^\mu\partial_\mu - mc)\,\psi = 0\,, \tag{1.3}$$

leading also to negative solutions. By treating the wave function ψ as a field, a quantum field operator, as in Jordan and Fock's work [33], negative solutions can be seen as antiparticles, thereby predicting them.

[Nobel] Indicates a work that was awarded the Nobel Prize.

This Dirac equation turns out to be the first step of the quantum electrodynamics (QED) theory, which in turn was the first cornerstone of the Standard Model (SM) theory. The SM is the most precise theory developed in particle physics, uniting three of the four fundamental interactions: electromagnetism, the weak interaction, and the strong interaction. Highly predictive and extensively tested through various experiments, it is the result of a century of breakthroughs in many related fields, as described previously. It is a succession of theories interlinked with discoveries to this day. However, this is not the final theory (if something like this exists) as several questions remain unsolved, therefore hinting at some beyond the Standard Model (BSM) theory.

This chapter will present the mathematical framework of the SM and its construction, culminating in its most recent experimental confirmation through the discovery of the Higgs boson. It will also incorporate some hints on how to probe the limits of this theory, on which the main part of this thesis will be based.

1.1 Formalism

1.1.1 Units

Natural units are a convenient choice for expressing equations and quantities at the scale of particle physics. Therefore, we define:

$$\hbar = c = 1\,, \tag{1.4}$$

where $\hbar$ is the reduced Planck constant given by $\hbar = \frac{h}{2\pi}$ with $h = 6.626\,070\,15 \cdot 10^{-34}\,\mathrm{J}\cdot\mathrm{s}$ and c the speed of light in vacuum of $299\,792\,458\,\mathrm{m}\cdot\mathrm{s}^{-1}$. We use this convention for the rest of this thesis. This allows us to express the mass of particles in GeV instead of GeV/c^2.

Fundamental values used throughout this thesis are taken from the Particle Data Group (PDG) [34].

1.1.2 Special relativity

Using Einstein notation we can define the four-vectors $x^\mu = (t, \vec{x})$ following the relation:

$$x^\mu = g^{\mu\nu} x_\nu\,, \tag{1.5}$$

where $(\mu, \nu) \in \{0, 1, 2, 3\}$ and the Minkowski metric tensor $g^{\mu\nu} = \mathrm{diag}(+, -, -, -)$.

In special relativity, each particle can be characterised by its four-momentum vector $p^\mu = (E, \vec{p})$ relating its energy E and momentum. The invariant mass m is defined by the inner product of p^μ:

$$m^2 = p^\mu p_\mu = E^2 - \vec{p}^{\,2}\,. \tag{1.6}$$

Lorentz invariance ensures that scalar quantities constructed from four-vectors, such as the invariant mass, remain unchanged under rotation and Lorentz boosts. The Lorentz factor characterises a boost along any spatial direction:

$$\gamma = \frac{1}{\sqrt{1-\beta^2}}\,, \tag{1.7}$$

with $\beta = \frac{v}{c}$. The four-gradient is defined as $\partial_\mu = (\partial_t, \vec{\nabla})$.

1.1.3 Field theory

Until Lagrange's work in the late XVIII century, the description of physics was done with classical equations using the position of an object described by its coordinates r and time t. Lagrange developed a new formalism based on a function called the *Lagrangian* [35]:

$$L(r, \dot{r}) = T(r, \dot{r}) - V(r, \dot{r})\,, \tag{1.8}$$

involving energies rather than forces, kinetic T and potential V, both depending on the coordinates r and their time derivative, known as velocity $\dot{r}$. Based on the action principle, the laws of classical physics can be derived from the action S. Any system evolving between an initial state a and a final state b will do so in a way that minimises it. It is expressed as the integral of the Lagrangian over time:

$$S = \int_a^b L(r, \dot{r})\mathrm{d}\,t\,, \tag{1.9}$$

passing from the differential approach formulated by Newton to an integral one.

A variation of the action δS is given by:

$$\delta S = \int_a^b \left(\frac{\partial L}{\partial r} \delta r + \frac{\partial L}{\partial \dot{r}} \delta \dot{r} \right) \mathrm{d}\, t \,. \tag{1.10}$$

The pathway has to satisfy $\delta S = 0$ implying:

$$\frac{\partial L}{\partial r} - \frac{\mathrm{d}}{\mathrm{d}\, t} \left(\frac{\partial L}{\partial \dot{r}} \right) = 0 \,, \tag{1.11}$$

which is known as the *Euler–Lagrange equation* [36] and describes the motion of the system. It allows us to go from the Lagrangian to the equation of motion.

This formalism can be extended to fields more suitable for studying non-point-like objects, such as the wave function in Equation 1.2. The Lagrangian is then a Lagrangian density $\mathcal{L}(\psi, \partial_\mu \psi)$ depending on the field ψ and its four-gradient. From now on, the distinction between Lagrangian and Lagrangian density will be omitted for simplicity. Using this, we get the action:

$$S(\psi) = \int_a^b \mathcal{L}(\psi, \partial_\mu \psi) \mathrm{d}^4\, x \,, \tag{1.12}$$

which is integrated over space and time. Similarly, we can derive the Euler–Lagrange equation as:

$$\frac{\partial \mathcal{L}}{\partial \psi} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \psi)} \right) = 0 \,. \tag{1.13}$$

1.1.4 Quantum field theory

When unifying quantum mechanics with special relativity, the most suitable choice for defining the particles is quantum fields instead of the wave function, resulting in the development of quantum field theory (QFT). They consist of excitations of a quantum field, whose states live in a space called the Fock space [33], with $|0\rangle$ defining the vacuum, where the creation and annihilation operators, $\hat{a}^\dagger(\vec{p})$ and $\hat{a}(\vec{p})$, will create a particle with a momentum $\vec{p}$ and an energy following Equation 1.6. We can express the field as a Fourier decomposition in the four-momentum space. In the case of the Klein–Gordon field, Equation 1.2, it is written as:

$$\phi(\vec{x}, t) = \int \left(\hat{a}(\vec{p}) \mathrm{e}^{-\mathrm{i}p\cdot x} + \hat{a}^\dagger(\vec{p}) \mathrm{e}^{\mathrm{i}p\cdot x} \right) \frac{\mathrm{d}^3\, p}{(2\pi)^{3/2} \sqrt{2E}} \,, \tag{1.14}$$

where the differential element and its denominator cover the phase-space to ensure E is positive.

The Lagrangian in quantum field theory can be decomposed into two distinct parts, the free part describing the propagation of a field without external contributions, and the interaction part describing the interactions between these fields. To quantify the probability of the whole system to evolve from an initial state $|i\rangle$ to a final state $|f\rangle$, we use the transition amplitude $\mathcal{S}_{fi} = \langle f|\mathcal{S}|i\rangle$ which is a complex number contained in the $\mathcal{S}$-matrix. The latter is expressed as:

$$\mathcal{S} = T \exp\left(-\mathrm{i} \int \mathcal{H}_{\mathrm{int}}(x)\, \mathrm{d}^4\, x \right) \,, \tag{1.15}$$

where T is the time ordering operator ensuring causality between the interactions and $\mathcal{H}_{\mathrm{int}}$ the Hamiltonian density of the interaction part, derived from a Legendre transform of the initial Lagrangian. This equation can be expanded as a Taylor series, producing an infinite sum of terms also referred to as the *Dyson series* [37]. This is a perturbative description of interacting quantum fields. The first term of this expansion is known as leading order (LO), followed by the next-to-leading order (NLO) and so on. For practicality, the Feynman amplitude $\mathcal{M}_{fi}$ is used instead of $\mathcal{S}_{fi}$ following:

$$\mathcal{S}_{fi} = \underbrace{\delta_{fi}}_{\text{no interaction}} + \underbrace{\mathrm{i}(2\pi)^4 \delta^{(4)}(p_f - p_i)}_{\text{four-momentum conservation}} \prod_{i=1}^{n_i} \frac{1}{\sqrt{2E_i V}} \prod_{f=1}^{n_f} \frac{1}{\sqrt{2E_f V}} \mathcal{M}_{fi} \,, \tag{1.16}$$

denoting the non-interaction term, the four-dimensional (4D) Dirac distribution $\delta^{(4)}$ ensuring four-momentum conservation between the initial and final states and a product over every entering and exiting field in the initial and final state, respectively. The volume V stands for the total phase-space volume accessible during the process. The Feynman amplitudes are commonly represented by Feynman diagrams, where a single diagram is equivalent to $\mathrm{i}\mathcal{M}_{fi}$. More details on the computation of Feynman diagrams are given in [38].

Finally, given the latter $\mathcal{M}_{fi}$, one can access key parameters of particles such as the decay rates Γ, but most importantly in the context of this thesis, the cross-section σ. It represents the probability that two initial particles a and b collide to produce a final state with n_f particles. It is given by:

$$\sigma_{a+b\to 1+\cdots+n_f} = \frac{(2\pi)^4}{F}\int |\mathcal{M}_{fi}|^2 \delta^{(4)}\left(p_a + p_b - \sum_{f=1}^{n_f} p_f\right) \prod_{f=1}^{n_f} \frac{\mathrm{d}^3 p_f}{(2\pi)^3 2E_f}\,, \tag{1.17}$$

using Fermi's golden rule. The product of the differential element and its denominator is known as Lorentz-invariant phase-space (LIPS), ensuring the Lorentz invariance of the integral. The term F is the flux factor for initial particles:

$$F = 4\sqrt{(p_a \cdot p_b)^2 - (m_a m_b)^2}\,, \tag{1.18}$$

also Lorentz invariant, thus making the cross-section Lorentz invariant too.

1.2 Quantum electrodynamics

Another major milestone in the early XX century was the explanation of spin. Observations such as the Zeeman effect [39], hyperfine structure, or the remarkable Stern and Gerlach experiment in 1922 [40] hinted at new properties of particles. First interpreted as an orbital angular momentum, spin was later understood to be a purely quantum property with no equivalent in classical mechanics. Pauli introduced in 1927 a set of matrices acting as operators on the Schrödinger wave function [41]:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_y = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \qquad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{1.19}$$

named after him. This applies to particles with a half-integer spin-$\frac{1}{2}$. According to Pauli's exclusion principle, such particles follow Fermi–Dirac statistics and are therefore *fermions*, such as the electron or the muon "μ". The latter was discovered in 1936 by Anderson while studying cosmic rays in a cloud chamber [42], presenting the same quantum properties as the electron but with a higher mass.

Dirac, when trying to unify special relativity and quantum mechanics, started from the Schrödinger equation, changing the original Hamiltonian by a set of operators α and β to have first-order space and time derivatives, such that:

$$E\Psi(\vec{r},t) = (\vec{\alpha}\cdot\vec{p} + \beta m)\,\Psi(\vec{r},t)\,. \tag{1.20}$$

The Klein–Gordon equation given in Equation 1.2 is obtained simply by applying Equation 1.6 to a wave function. The second derivative in time appears in the d'Alembert operator defined as $\Box = \partial^\mu \partial_\mu$. Regarding spin, this applies to a scalar field with spin-0 that follows Bose–Einstein statistics. Therefore, if the Dirac equation is to represent relativistic particles, it must also satisfy it. By expanding Equation 1.20 and identifying α and β with Equation 1.2, one can identify their properties given by:

- their squares must be equal to the identity matrix of dimension D, $\mathbb{I}_D$
- they must be mutually anti-commuting, therefore implying that they are matrices
- they must have a zero trace
- their eigenvalues must be ± 1, which forces the matrices to have even dimension
- they must be Hermitian.

Because there are only three mutually anti-commuting 2×2 traceless matrices, for example, the Pauli spin matrices, the lowest-dimension objects that can represent α and β are 4×4 matrices. These Dirac matrices are given, in the chiral representation, by:

$$\gamma^0 = \begin{pmatrix} 0 & \mathbb{I}_2 \\ \mathbb{I}_2 & 0 \end{pmatrix}, \qquad \gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix}, \tag{1.21}$$

including the Pauli matrices and completed by anti-commutation relations:

$$\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}\mathbb{I}_4\,, \tag{1.22}$$

forming a Clifford algebra. Consequently, the Dirac Hamiltonian of Equation 1.20 is a 4×4 matrix of operators that must act on a four-component wave function, known as a Dirac spinor ψ. The latter includes the two spin states, up $\uparrow$ and down $\downarrow$, as well as the positive energy solutions and negative energy solutions

identified as antiparticles. By requiring the first-order space and time derivative, the spin naturally arises in the equations, giving us the Dirac equation:

$$\left(\mathrm{i}\partial\!\!\!/ - m\right)\psi = 0\,, \tag{1.23}$$

where we use the Feynman notation with $\partial\!\!\!/ = \gamma^\mu\partial_\mu$. This equation follows, through the Euler–Lagrange equation of Equation 1.13, from the Dirac Lagrangian describing a free propagating massive spin-$\frac{1}{2}$ field:

$$\mathcal{L}_\mathrm{D} = \overline{\psi}(\mathrm{i}\partial\!\!\!/ - m)\psi\,, \tag{1.24}$$

with $\overline{\psi} = \psi^\dagger\gamma^0$ the adjoint.

Weyl, in 1929, based on Dirac's work, proposed a simplified equation for massless particles [43]:

$$\overline{\sigma}^\mu\partial_\mu\psi_\mathrm{L} = 0\,, \quad \sigma^\mu\partial_\mu\psi_\mathrm{R} = 0\,, \tag{1.25}$$

where $\sigma^\mu = (\mathbb{I}_2, \sigma^i)$ and $\overline{\sigma}^\mu = (\mathbb{I}_2, -\sigma^i)$. The wave functions ψ_L and ψ_R are defined as Weyl spinors, each being a two-component spinor (one for each spin state), respectively left-handed (LH) and right-handed (RH).

A convenient decomposition of the Dirac spinor can therefore be achieved through $\gamma^5 = \mathrm{i}\gamma^0\gamma^1\gamma^2\gamma^3$:

$$\gamma^5 = \begin{pmatrix} -\mathbb{I}_2 & 0 \\ 0 & \mathbb{I}_2 \end{pmatrix}. \tag{1.26}$$

We define the chirality projectors:

$$P_\mathrm{R/L} = \frac{1 \pm \gamma^5}{2}\,, \tag{1.27}$$

which give the LH and RH components of the Dirac spinors as Weyl ones:

$$\psi_\mathrm{L} = P_\mathrm{L}\psi\,, \quad \psi_\mathrm{R} = P_\mathrm{R}\psi \quad \Rightarrow \psi = \psi_\mathrm{L} + \psi_\mathrm{R}\,. \tag{1.28}$$

The Dirac spinors are thus a sum of both components.

The Poincaré group consists of the Lorentz group $\mathrm{SO}(1,3)$ plus translations in space-time. By decomposing the Dirac spinors into Weyl spinors, we do the following:

$$\mathrm{SO}(1,3) \cong \mathrm{SL}(2,\mathbb{C}) \cong \mathrm{SU}(2)_\mathrm{L} \oplus \mathrm{SU}(2)_\mathrm{R} \quad \text{(locally)}\,. \tag{1.29}$$

Massive representations necessarily couple LH and RH Weyl components, which is why a Dirac spinor contains both chiralities written as a direct sum of two Weyl spinors $\psi = (\frac{1}{2}, 0) \oplus (0, \frac{1}{2})$ equivalent to:

$$\psi = \begin{pmatrix} \psi_\mathrm{L} \\ \psi_\mathrm{R} \end{pmatrix}. \tag{1.30}$$

Indeed, the term:

$$\overline{\psi}_\mathrm{L} m \psi_\mathrm{R} + \overline{\psi}_\mathrm{R} m \psi_\mathrm{L}\,, \tag{1.31}$$

must have both LH and RH components for it to be Poincaré invariant.

A gauge theory is a type of field theory in which the Lagrangian does not change under local transformations. The simplest ones arise from the Abelian (that is, commutative operator) group $\mathrm{U}(1)$. Such symmetries are given by the unitary operator $U = \mathrm{e}^{\mathrm{i}Qe\theta}$ with θ a constant ($\partial_\mu\theta = 0$) and Q the generator of the group that can be considered as $\mathbb{I}_1$.

Therefore, the Dirac Lagrangian is globally invariant under $\mathrm{U}(1)$ transformations. Applying the gauge principle, we pass from a global to a local invariance, meaning now $\theta(x)$ is a function of space-time coordinates. The field thus transforms as:

$$\psi(x) \to \mathrm{e}^{\mathrm{i}Qe\theta(x)}\psi(x)\,. \tag{1.32}$$

Adding such a transformation breaks the invariance of the Dirac Lagrangian due to the derivative term. This can be resolved by introducing the covariant derivative to replace the original one:

$$D_\mu = \partial_\mu + \mathrm{i}QeA_\mu\,, \tag{1.33}$$

where e is a constant. This includes the field A_μ characterised via the same quantum number as the generator Q. It transforms under the symmetry introduced before as:

$$A_\mu(x) \to A_\mu(x) - \partial_\mu\theta(x)\,. \tag{1.34}$$

Since it transforms as a vector, having a representation as $(\frac{1}{2}, \frac{1}{2})$, this makes it a spin-$1$ gauge field. The introduction of this gauge field restores the local gauge invariance of the Lagrangian by cancelling out the derivative of $\theta(x)$.

We can make this field dynamic by adding to the Lagrangian a kinetic term:

$$\mathcal{L}_{\text{kin}} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} \,, \tag{1.35}$$

based on the field strength:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \,. \tag{1.36}$$

This gives us the following Lagrangian:

$$\mathcal{L}_{\text{QED}} = \overline{\psi}(\mathrm{i}\not{\partial} - m)\psi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} - Qe\overline{\psi}\gamma^\mu \psi A_\mu \,. \tag{1.37}$$

We can identify the gauge field A_μ as the photon "γ" boson, carrier of the electromagnetic (EM) force, with Equation 1.35 encoding Maxwell's equations. This boson was experimentally observed in 1923 by Compton [44] through the scattering of X-rays on electrons. The constant e is therefore the elementary charge, $1.602\,176 \cdot 10^{-19}\,\mathrm{C}$, and Q the fermion charge in units of e. The photon must remain massless, as a term like $\frac{1}{2} m_\gamma^2 A^\mu A_\mu$ would break the gauge invariance. Also, the photon is neutral under EM interaction, directly derived from the Abelian nature of the $\mathrm{U}(1)$ group.

The last term of Equation 1.37 can be seen as an interaction term involving the fermion field and the photon, resulting in the Feynman diagram given in Figure 1.1.

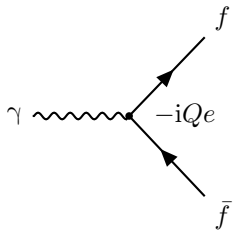


Figure 1.1: Feynman diagram representing the interaction between two fermions and the photon in QED.

At the same time, no self-interaction terms for the photons are present.

The Lagrangian in Equation 1.37 therefore encodes the full QED theory with interactions and propagations of the fermions and photons. When computing the cross-section with QED Feynman diagrams, we can see that it will become proportional to the square of the amplitude modulus as seen in Equation 1.17. Therefore, it is dependent on e^2. The strength of the EM interaction is set by the fine-structure constant $\alpha = \frac{e^2}{4\pi} \simeq \frac{1}{137}$.

1.3 Quantum chromodynamics

In 1932, Heisenberg proposed a new quantum number, the isospin [45], derived from the spin to explain that the proton "p" and the neutron "n" can both be treated as part of the same particle: the nucleon. Indeed, they both share nearly identical properties apart from their electrical charge, and they are treated the same way by the strong interaction. Therefore, he formed an $\mathrm{SU}(2)_{\text{iso}}$ group. The proton and the neutron are both associated with a doublet. In the late 1940s, several new particles were discovered, falling into two categories:

- The mesons: lighter than the proton and neutron, with integer spin, and unstable. They were originally considered as mediators of the strong force, from which they derive their name. The first one was the pion "π" in 1947 [46].
- The baryons: with half-integer spin, heavier than a proton, therefore giving their name to this category. Protons and neutrons are part of this category.

At first, it was possible to categorise them following the model of Heisenberg, for example, the pions were associated with a triplet of $\mathrm{SU}(2)_{\text{iso}}$. However, the "particle zoo" kept extending, and the model could not explain it. For example, four kaon mesons "K" were not possible. A new "strange" number was needed. The model evolved towards an $\mathrm{SU}(3)_{\text{s}}$ splitting the previous classification into a triplet of up "u", down "d", and strange "s", called *quarks* and part of the fermion family. This model, also referred to as the *Eightfold*

Way, was developed by Gell-Mann in 1964 [47]. This model worked remarkably well, finally offering a classification of the hadron (particles composed of quarks) zoo, as well as predicting particles that were later found.

The discovery of baryons Δ^-, Δ^{++} [48], and Ω^- [49] of spin-$\frac{3}{2}$ raised an issue. Indeed, they are composed of identical fermions in the same spin state, which is forbidden by Pauli's exclusion principle. Therefore, an underlying effect was needed to resolve this. The colours were proposed [50]. Each quark is associated with a colour, red, green, and blue, to form, by analogy with additive light, a white particle. This was motivated by the fact that no free quarks were observed, but always as bound states. The introduction of colour is known as quantum chromodynamics (QCD). This phenomenon is referred to as *colour confinement*. For baryons, three different colours are needed: qqq, while for mesons, the association of colour and anticolour is needed: $q\bar{q}$.

Considering a quark as a triplet of colour in $\mathrm{SU}(3)_\mathrm{C}$ we have $q_\mathrm{f} = (q_\mathrm{f}^\mathrm{red}, q_\mathrm{f}^\mathrm{green}, q_\mathrm{f}^\mathrm{blue})$ with f the flavour. We can thus write the QCD Lagrangian:

$$\mathcal{L}_\mathrm{QCD} = \sum_\mathrm{f} \mathrm{i}\overline{q}_\mathrm{f} D_\mu \gamma^\mu q_\mathrm{f} - \frac{1}{4}\sum_{a=1}^{8} G^{\mu\nu,a} G^a_{\mu\nu}\,, \tag{1.38}$$

written here for massless quarks. The first term is similar to Equation 1.37, encoding the free propagation of quarks and their interaction with the gauge fields, within the covariant derivative. Similarly to the development done in Section 1.2, it is introduced to restore the invariance under such a transformation:

$$q_\mathrm{f} \to \mathrm{e}^{\mathrm{i}g_\mathrm{S}\theta^a(x)T^a} q_\mathrm{f}\,, \tag{1.39}$$

where T^a are the generators of this group, defined as $T^a = \frac{1}{2}\lambda^a$ with λ^a the Gell-Mann matrices developed in his quark model. The resulting covariant derivative is:

$$D_\mu = \partial_\mu + \mathrm{i}g_\mathrm{S}\sum_{a=1}^{8} G^a_\mu T^a\,, \tag{1.40}$$

with G^a_μ corresponding to the eight gauge fields associated with the gluon "g" (as they "glue" the quarks together), each associated with a generator of $\mathrm{SU}(3)_\mathrm{C}$ and discovered in 1979 at Deutsches Elektronen-Synchrotron (DESY) via the $\mathrm{e}^+\mathrm{e}^-$ Positron-Elektron-Tandem-Ring-Anlage (PETRA) collider and its four experiments Two Arm Spectrometer SOlenoid (TASSO) [51], MARK J [52], JADE [53], and PLUTO [54].

The second term in Equation 1.38 corresponds to the gluons' kinetic term, where the field strength is:

$$G^{\mu\nu,a} = \partial^\mu G^{\nu,a} - \partial^\nu G^{\mu,a} - g_\mathrm{S}\sum_{b,c=1}^{8} f^{abc} G^{\mu,b} G^{\nu,c}\,, \tag{1.41}$$

where f^{abc} are the structure constants of the $\mathrm{SU}(3)$ algebra. They arise from the non-Abelian (non-commutative) nature of $\mathrm{SU}(3)_\mathrm{C}$ where we have:

$$[T^a, T^b] = \mathrm{i}f^{abc}T^c\,. \tag{1.42}$$

This last term, not present in QED, makes the field strength $G^{\mu\nu,a}$ quadratic in the gauge field. Thus, the last term of Equation 1.38 will produce cubic and quartic interaction terms between the gluons themselves. This contrasts with the photon, which does not couple to itself since it is electrically neutral. The gluons, on the other hand, are colour charged, therefore interacting strongly with each other.

Similarly to the fine-structure constant, the strong interaction parameter g_S can be expressed as $\alpha_\mathrm{S} = \frac{g_\mathrm{S}^2}{4\pi}$ known as the *strong coupling constant*. It will also appear in every interaction term linking quarks and gluons as photons do for fermions.

In a perturbative theory, as expressed by the infinite-sum decomposition of Equation 1.15, higher orders involve loops. A renormalisation is needed to compute the integrals at such an order. Indeed, loops are divergent. It will consist of a re-parametrisation of the Lagrangian. This is mainly because, experimentally, the observed parameters already include all their contributions. Therefore, renormalisation ensures that the theory remains predictive at all energy scales. This can be done only if the theory is locally gauge invariant.

Due to the self-interaction of the gluons, the fermionic loops and the bosonic loops will interfere negatively, resulting in a total effect lower than a diagram of a lower order. With renormalisation, this will be manifested by the running of the strong coupling constant α_S that will vary with the momentum transfer Q during interactions. It is given by:

$$\alpha_\mathrm{S}(Q^2) = \frac{\alpha_\mathrm{S}(\mu_0^2)}{1 + B\alpha_\mathrm{S}(\mu_0^2)\ln\left(\frac{Q^2}{\mu_0^2}\right)}\,, \tag{1.43}$$

where μ_0^2 is a given reference energy scale and $B = \frac{11n_c - 2n_f}{12\pi}$, with n_c the number of colours (three) and n_f the number of quark flavours. As predicted, the running constant α_S decreases with increasing energy. However, it rapidly increases at low Q^2. In this domain, it is no longer possible to use perturbation theory, and calculations become intractable. At larger Q^2, α_S is small enough for perturbation theory to be used. However, higher orders (NLO, next-to-next-to-leading order (NNLO), and beyond) have a non-negligible effect. This decrease in the coupling constant is known as *asymptotic freedom* [55, 56]. Quarks can be treated as nearly free at high Q^2, leading to a state of matter known as quark–gluon plasma (QGP) where quarks and gluons are no longer confined and can thus move freely. In the very first instants, our Universe was fully in this state. The transition between the two domains happened around $\Lambda_{QCD} \simeq 200\,\mathrm{MeV}$.

Similarly, the QED coupling constant also runs with renormalisation. However, with no bosonic loops, the constant slightly increases with Q^2, but perturbation theory is still feasible at such energy scales. It is given by:

$$\alpha(Q^2) = \frac{\alpha(\mu_0^2)}{1 - \alpha(\mu_0^2)\frac{1}{3\pi}\ln\left(\frac{Q^2}{\mu_0^2}\right)}\,. \tag{1.44}$$

1.4 Weak interaction

Radioactivity was discovered in 1896 by Becquerel (giving its name to the unit of activity) in uranium [57] and extended by Curie to thorium, discovering in the meantime polonium [58] and radium [59]. Radioactivity consists of a spontaneous emission of particles: α (helium nucleus), β, and γ. The α-decay led to the discovery of the proton by Rutherford in 1919 [60]. Neutrons, on the other hand, were discovered in 1932 by Chadwick [61], resolving the problem of proton–electron-only models failing to describe nature.

In the case of the β-decay, it was observed that a neutron was emitting an electron and turning into a proton. Therefore, the energy spectrum of this electron should have been a narrow distribution since the electron is supposed to carry the energy difference between the initial and final nuclear state. However, a broad spectrum was observed, hinting at a misunderstanding of this phenomenon. To save energy conservation, Pauli proposed in 1930, in his famous letter, a new neutral particle of spin $\frac{1}{2}$ and very small mass, emitted together with the electron and carrying away the missing energy. Fermi developed his theory in 1933 to explain the β-decay problem, building on this idea with an extremely light particle, the neutrino "ν", which interacts only weakly with matter (no observation of it was possible at the time) [62]. This was the first attempt to describe this new interaction called weak.

The electron neutrino "ν_e" was discovered in 1956 by Cowan and Reines [63] while studying the inverse β-decay of a proton into a neutron via a nuclear reactor following: $\bar{\nu}_e + p \to n + e^+$. The muon neutrino "$\nu_\mu$" was discovered in 1962 [64] at Brookhaven National Laboratory (BNL) while studying the decay of pions $\pi^+ \to \mu^+ + \nu_\mu$ and the subsequent decay of the muon $\mu^+ \to e^+ + \nu_e + \bar{\nu}_\mu$. No electron was observed in the first interaction, proving that the neutrino associated with the muon is different from the electron one.

A parity transformation P changes the signs of spatial coordinates. Therefore, under such a transformation, vectors change sign, while pseudo-vectors such as the magnetic field or the spin do not. In 1957, the Wu experiment [65] showed that under an external magnetic field, the electrons produced in the β-decay of cobalt were preferentially emitted in the direction opposite to the magnetic field. If parity were conserved, electrons should be emitted in both directions with no distinction. This experiment showed that parity is violated by the weak interaction.

To formalise this phenomenon, a vertex mixing a vector V and an axial vector A is needed. The chirality projector operator defined in Equation 1.27 is the perfect example. The LH coupling is denoted as *V–A* while the RH one is *V+A*. An experiment showed that the structure *V–A* is the natural one. Therefore, the weak interaction only couples LH chiral particles and RH chiral antiparticles.

As a consequence, the charge symmetry C is also violated since under a charge conjugation transformation, particles are replaced by their antiparticles, while the helicity is unchanged: an LH neutrino is mapped onto an LH antineutrino, which is not observed. Therefore, the charge-conjugated interaction is not an allowed weak interaction.

In 1954, Yang and Mills extended the concept of gauge theory from an Abelian group, such as QED, to a non-Abelian group using the group $SU(2)$. The generalisation of this is called a *Yang–Mills theory* [66], of which QCD with its $SU(3)_C$ group (Section 1.3) is an example.

The bosons of this theory couple charged leptons to neutrinos of the same generation and up-type quarks to down-type quarks. Given the specificity of the weak interaction regarding chirality, it seems

natural to build the weak theory based on the $\mathrm{SU}(2)_\mathrm{L}$, resulting in three gauge bosons. The weak isospin doublet, by analogy with the strong isospin, is defined as:

$$\begin{pmatrix} \mathrm{u} \\ \mathrm{d} \end{pmatrix}_\mathrm{L} , \quad \begin{pmatrix} \nu_\mathrm{e} \\ \mathrm{e} \end{pmatrix}_\mathrm{L} . \tag{1.45}$$

The RH fields, on the other hand, are singlets of this group.

The local gauge invariance is given by:

$$\varphi_\mathrm{L}(x) \to \mathrm{e}^{\mathrm{i}g_\mathrm{L}\theta^i(x)T^i}\varphi_\mathrm{L}(x) , \tag{1.46}$$

where $\varphi_\mathrm{L}(x)$ is a doublet and the T^i are the generators of this group, defined as $T^i = \frac{1}{2}\sigma^i$ with σ^i the Pauli matrices given in Equation 1.19. To restore the local invariance from the Dirac Lagrangian $\mathcal{L}_\mathrm{D}$ (Equation 1.24), written here for a massless doublet, we introduce the covariant derivative:

$$D_\mu = \partial_\mu + \mathrm{i}g_\mathrm{L}W_\mu^i T^i , \tag{1.47}$$

giving three gauge fields, given the structure of the group and denoted as W_μ^i. The field strength is:

$$W_{\mu\nu}^i = \partial_\mu W_\nu^i - \partial_\nu W_\mu^i - g_\mathrm{L}\varepsilon^{ijk}W_\mu^j W_\nu^k , \tag{1.48}$$

where ε^{ijk} is the Levi-Civita tensor defining the structure constant of the group.

The full Lagrangian is therefore:

$$\mathcal{L}_\mathrm{L} = \overline{\varphi}_\mathrm{L}\,\mathrm{i}\partial\!\!\!/\,\varphi_\mathrm{L} - \frac{1}{4}W_{\mu\nu}^i W_i^{\mu\nu} - g_\mathrm{L}\overline{\varphi}_\mathrm{L}\gamma^\mu T^i \varphi_\mathrm{L} W_\mu^i . \tag{1.49}$$

The first problem arises with this Lagrangian. Indeed, no mass term for the gauge bosons is present as it would break the local gauge invariance, similarly to the photon in QED. However, experiments show that the weak interaction is short-ranged, which implies that the gauge bosons must be massive.

We can still define the boson "$\mathrm{W}^\pm$" of this theory as:

$$\mathrm{W}_\mu^\pm = \frac{1}{\sqrt{2}}\left(W_\mu^1 \mp \mathrm{i}W_\mu^2\right) . \tag{1.50}$$

We could naively identify the remaining gauge boson W_μ^3 as the boson "Z^0", although experiments show that this boson couples both LH and RH particles. Therefore, we need a more complex theory that lets us introduce the Z properly and also the mass of the weak bosons while keeping the gauge invariance.

It was also observed that certain hadronic decays via weak interaction were suppressed compared to others. For example, the $\mathrm{K}^- \to \mu^-\bar{\nu}_\mu$ decay was less probable than the $\pi^- \to \mu^-\bar{\nu}_\mu$ one. An explanation was proposed by Cabibbo in 1963 [67], stating that the weak interaction eigenstates were not the same as the mass eigenstates and therefore were mixed by an angle, the Cabibbo angle:

$$\begin{pmatrix} \mathrm{d}' \\ \mathrm{s}' \end{pmatrix}_\mathrm{L} = \begin{pmatrix} \cos\theta_\mathrm{C} & \sin\theta_\mathrm{C} \\ -\sin\theta_\mathrm{C} & \cos\theta_\mathrm{C} \end{pmatrix} \begin{pmatrix} \mathrm{d} \\ \mathrm{s} \end{pmatrix}_\mathrm{m} . \tag{1.51}$$

Flavour-changing processes via the weak interaction, such as $\mathrm{K}^0 \to \mu^+\mu^-$, were also observed but are highly suppressed. Glashow, Iliopoulos, and Maiani in 1970 proposed the existence of a fourth quark, the charm "c", to explain this suppression via loop diagrams. This is known as the *GIM mechanism* [68]. The charm quark was later discovered in 1974 via the J/ψ mesons[9] [69, 70], in $\mathrm{e}^+\mathrm{e}^-$ collisions at Stanford Linear Accelerator Center (SLAC) and in proton–beryllium collisions at BNL.

As explained before, the weak interaction violates both the parity and the charge symmetries. This implies that the combined symmetry charge-parity (CP) is not necessarily conserved. In 1964, Cronin and Fitch observed the decay of the long-lived neutral kaon K_L^0 into two pions $\pi^+\pi^-$ [71], a CP-forbidden process. This was the first observation of CP violation[9].

By extending the quark mixing to three generations, Kobayashi and Maskawa in 1973 [72] showed that a complex phase appears in the mixing matrix, allowing for CP violation. This matrix is now known as the Cabibbo–Kobayashi–Maskawa (CKM) matrix[9] V_CKM and is given by:

$$\begin{pmatrix} \mathrm{d}' \\ \mathrm{s}' \\ \mathrm{b}' \end{pmatrix} = V_\mathrm{CKM} \begin{pmatrix} \mathrm{d} \\ \mathrm{s} \\ \mathrm{b} \end{pmatrix} = \begin{pmatrix} V_\mathrm{ud} & V_\mathrm{us} & V_\mathrm{ub} \\ V_\mathrm{cd} & V_\mathrm{cs} & V_\mathrm{cb} \\ V_\mathrm{td} & V_\mathrm{ts} & V_\mathrm{tb} \end{pmatrix} \begin{pmatrix} \mathrm{d} \\ \mathrm{s} \\ \mathrm{b} \end{pmatrix} . \tag{1.52}$$

The parametrisation can be done in several ways, but the standard one is given by three mixing angles θ_{12}, θ_{23}, θ_{13}, and a complex phase δ:

$$V_{\mathrm{CKM}} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}\mathrm{e}^{-\mathrm{i}\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}\mathrm{e}^{\mathrm{i}\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}\mathrm{e}^{\mathrm{i}\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}\mathrm{e}^{\mathrm{i}\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}\mathrm{e}^{\mathrm{i}\delta} & c_{23}c_{13} \end{pmatrix}, \tag{1.53}$$

where $c_{ij} = \cos\theta_{ij}$ and $s_{ij} = \sin\theta_{ij}$. The presence of this complex phase δ is the source of CP violation in the quark sector.

This groundbreaking work predicted the existence of the third generation of quarks, hence postulating the top "t" and bottom "b" quarks. In 1977, the bottom quark b was discovered in a fixed-target experiment at Fermi National Accelerator Laboratory (Fermilab), in proton–nucleus collisions [73], followed only in 1995 by the top quark t at the Tevatron collider via the Collider Detector at Fermilab (CDF) [74] and DØ [75]. This set the number of quark flavours n_{f}, defined in Section 1.3, at six.

1.5 Brout–Englert–Higgs mechanism

The Lagrangian of a gauge theory must remain invariant under local gauge transformations. However, this symmetry can be spontaneously broken if the minimal energy state of this theory, and not the Lagrangian itself, violates the gauge invariance. This is referred to as spontaneous symmetry breaking (SSB).

In 1960, the Nambu–Goldstone theorem[theorem] [76] stated that for every continuous symmetry that is spontaneously broken, there must exist a corresponding massless boson called a *Goldstone boson*.

In 1964, Higgs [77] and Brout and Englert [78] independently demonstrated that it is possible for a gauge theory to generate masses for bosons by combining the Goldstone bosons and the massless gauge bosons[boson].

A visual representation of this mechanism can be achieved by considering a self-interacting scalar field ϕ and a gauge field B_μ. The resulting Lagrangian is:

$$\mathcal{L}_{\mathrm{H}} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + (D_\mu\phi)^\dagger(D^\mu\phi) - V(\phi)\,, \tag{1.54}$$

where the covariant derivative is defined as $D_\mu = \partial_\mu + \mathrm{i}g_{\mathrm{H}}B_\mu$. This Lagrangian is invariant under the local $\mathrm{U}(1)$ transformation:

$$\phi(x) \to \mathrm{e}^{\mathrm{i}g_{\mathrm{H}}\theta(x)}\phi(x)\,, \quad B_\mu(x) \to B_\mu(x) - \partial_\mu\theta(x)\,. \tag{1.55}$$

We can choose the form of the potential $V(\phi)$ as:

$$V(\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2\,, \tag{1.56}$$

with $\lambda > 0$ for stability. If $\mu^2 > 0$, the potential has a minimum at $\phi = 0$ and the gauge symmetry is preserved. However, if $\mu^2 < 0$, the potential takes the shape of a sombrero with a local maximum at $\phi = 0$ and a circle of minima at:

$$|\phi| = \frac{v}{\sqrt{2}} \quad \text{with} \quad v = \sqrt{\frac{-\mu^2}{\lambda}}\,. \tag{1.57}$$

We refer to this value v as the vacuum expectation value (VEV) of the field. The ground state of the theory is therefore no longer invariant under the gauge transformation, leading to SSB of the $\mathrm{U}(1)$ symmetry. Indeed, the potential is invariant under rotation in the complex plane, but choosing a particular ground state in the circle of minima breaks this symmetry. This is represented in Figure 1.2 when passing from position A to position B.

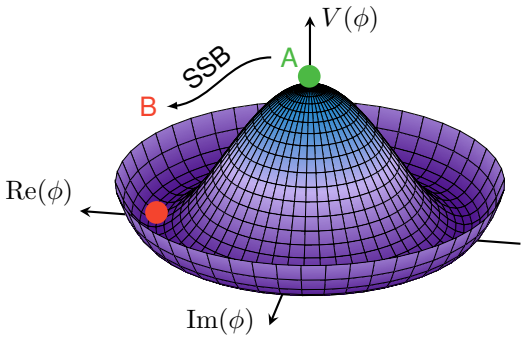


Figure 1.2: Representation of the potential $V(\phi)$.

If we choose the ground state as the point B in Figure 1.2 located at $(\mathrm{Re}\,\phi, \mathrm{Im}\,\phi) = (\frac{v}{\sqrt{2}}, 0)$ (this choice is arbitrary as every minimum is equivalent by rotation), we can perform a perturbative expansion of the field around this point to represent the quantum fluctuations:

$$\phi(x) = \frac{1}{\sqrt{2}}\left(v + \phi_1(x) + \mathrm{i}\phi_2(x)\right) . \tag{1.58}$$

We can then rewrite the potential as:

$$V(\phi) = -\frac{\lambda v^4}{4} + \lambda v^2\phi_1^2 + \lambda v\phi_1^3 + \frac{\lambda\phi_1^4}{4} + \frac{\lambda\phi_2^4}{4} + \lambda v\phi_1\phi_2^2 + \frac{\lambda\phi_1^2\phi_2^2}{2} . \tag{1.59}$$

By comparison with the Klein–Gordon Lagrangian given in Equation 1.2, we can identify the mass of the field ϕ_1 as $m_{\phi_1} = \sqrt{2\lambda v^2}$ while ϕ_2 remains massless.

We can thus expand the product of the covariant derivative in Equation 1.54 as:

$$(D_\mu\phi)^\dagger(D^\mu\phi) = \frac{1}{2}(\partial_\mu\phi_1)^2 + \frac{1}{2}(\partial_\mu\phi_2)^2 + g_\mathrm{H} v B^\mu\partial_\mu\phi_2 + \frac{1}{2}g_\mathrm{H}^2 v^2 B_\mu B^\mu + \dots , \tag{1.60}$$

where the dots represent cubic and quartic interaction terms. The gauge field B_μ has acquired a mass term $m_B = g_\mathrm{H} v$. Therefore, via the SSB, the gauge boson has acquired a mass, while the expansion of the potential yields a massive scalar ϕ_1 and a massless Goldstone mode ϕ_2.

The latter is involved in an interaction term with the gauge boson in the third term of Equation 1.60 corresponding to the following Feynman diagram:

$$= \mathrm{i}g_\mathrm{H} v(-\mathrm{i}k^\mu) = m_B k^\mu . \tag{1.61}$$

This interaction has no physical meaning and therefore has to be removed.

1.6 Glashow–Weinberg–Salam theory

Given the failure of the simple Yang–Mills theory based on $\mathrm{SU}(2)_\mathrm{L}$ to describe the weak interaction, it seems natural to start from a higher symmetry group. In 1961, Glashow proposed to extend the gauge group to $\mathrm{SU}(2)_\mathrm{L} \times \mathrm{U}(1)$ [79]. However, he could not generate the masses of the weak bosons. In 1967, Weinberg [80] and Salam [81] independently proposed to use the Brout–Englert–Higgs (BEH) mechanism in the $\mathrm{SU}(2)_\mathrm{L} \times \mathrm{U}(1)_\mathrm{Y}$ gauge theory to generate the masses of the weak bosons. This is known as the Glashow–Weinberg–Salam (GWS) theory. The core of this theory is to unify the EM and weak interactions into a single electroweak (EW) interaction.

To achieve this, we start again via a scalar field ϕ from the $\mathrm{SU}(2)_\mathrm{L}$ representing the weak interaction through the gauge fields W_μ^i and the $\mathrm{U}(1)_\mathrm{Y}$ representing the hypercharge interaction via the gauge field B_μ.

The Lagrangian is therefore:

$$\mathcal{L}_\mathrm{GWS} = (D_\mu\phi)^\dagger(D^\mu\phi) - \mu^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2 - \frac{1}{4}W_{\mu\nu}^i W_i^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} , \tag{1.62}$$

where $B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$ and the covariant derivative is defined as:

$$D_\mu = \partial_\mu + \mathrm{i}g_\mathrm{L} W_\mu^i T^i + \mathrm{i}g_\mathrm{Y}\frac{Y}{2}B_\mu . \tag{1.63}$$

The transformation of the scalar field under the local gauge group is:

$$\phi(x) \to \mathrm{e}^{\mathrm{i}g_\mathrm{L}\theta^i(x)T^i}\mathrm{e}^{\mathrm{i}g_\mathrm{Y}Y\beta(x)/2}\phi(x) . \tag{1.64}$$

It is chosen as a doublet of $\mathrm{SU}(2)_\mathrm{L}$ with a hypercharge $Y = 1$ under $\mathrm{U}(1)_\mathrm{Y}$.

Choosing again $\mu^2 < 0$, the potential will have a minimum at the VEV following Equation 1.57. We can choose the ground state as:

$$\langle\phi\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v\end{pmatrix} . \tag{1.65}$$

A transformation using $\theta^1 = \theta^2 = 0$ and $g_\mathrm{L}\theta^3 = g_\mathrm{Y}\beta$ thus leaves this ground state invariant. The three broken generators give rise to three Goldstone bosons, which are absorbed by the gauge bosons acquiring a mass via the BEH mechanism. This combination results in the following operator:

$$Q = T^3 + \frac{Y}{2} = \begin{pmatrix}\frac{1}{2} & 0\\ 0 & -\frac{1}{2}\end{pmatrix} + \begin{pmatrix}\frac{1}{2} & 0\\ 0 & \frac{1}{2}\end{pmatrix} = \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix} . \tag{1.66}$$

This operator corresponds to the electric charge operator. Therefore, the unbroken symmetry is the EM one. The SSB of the EW interaction has therefore led to the EM interaction following $\mathrm{SU}(2)_\mathrm{L} \times \mathrm{U}(1)_\mathrm{Y} \to \mathrm{U}(1)_\mathrm{EM}$.

It results in the scalar field being charged under the EW interaction as:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} . \tag{1.67}$$

The remaining three operators of the broken symmetries, T^1, T^2, and $T^3 - \frac{Y}{2}$, are each associated with a Goldstone boson.

The mass terms arise from the covariant derivative product, where the relevant terms are:

$$(D_\mu\phi)^\dagger(D^\mu\phi) \supset \frac{1}{2}\begin{pmatrix} 0 & v \end{pmatrix}\left(g_\mathrm{L}W_\mu^i T^i + g_\mathrm{Y}\frac{Y}{2}B_\mu\right)\left(g_\mathrm{L}W^{\mu,j}T^j + g_\mathrm{Y}\frac{Y}{2}B^\mu\right)\begin{pmatrix} 0 \\ v \end{pmatrix} . \tag{1.68}$$

Expanding this expression leads to:

$$(D_\mu\phi)^\dagger(D^\mu\phi) \supset \frac{1}{2}\frac{v^2}{4}\left[g_\mathrm{L}^2(W_\mu^1 + \mathrm{i}W_\mu^2)(W^{\mu,1} - \mathrm{i}W^{\mu,2}) + (g_\mathrm{L}W_\mu^3 - g_\mathrm{Y}B_\mu)(g_\mathrm{L}W^{\mu,3} - g_\mathrm{Y}B^\mu)\right] . \tag{1.69}$$

We can therefore identify the resulting gauge bosons:

$$\mathrm{W}_\mu^\pm = \frac{1}{\sqrt{2}}\left(W_\mu^1 \mp \mathrm{i}W_\mu^2\right) \text{ with the mass } m_\mathrm{W} = g_\mathrm{L}\frac{v}{2} , \tag{1.70}$$

$$\mathrm{Z}_\mu = \frac{1}{\sqrt{g_\mathrm{L}^2 + g_\mathrm{Y}^2}}\left(g_\mathrm{L}W_\mu^3 - g_\mathrm{Y}B_\mu\right) \text{ with the mass } m_\mathrm{Z} = \sqrt{g_\mathrm{L}^2 + g_\mathrm{Y}^2}\frac{v}{2} , \tag{1.71}$$

$$A_\mu = \frac{1}{\sqrt{g_\mathrm{L}^2 + g_\mathrm{Y}^2}}\left(g_\mathrm{Y}W_\mu^3 + g_\mathrm{L}B_\mu\right) \text{ with the mass } m_A = 0 , \tag{1.72}$$

where A_μ remains massless and is identified as the photon field, analogously to what was done in QED in Section 1.2. While the mass term of the W is straightforward, the Z and A_μ fields arise from the diagonalisation of the mass matrix to retrieve their eigenstates:

$$\begin{aligned} \frac{1}{2}\frac{v^2}{4}(g_\mathrm{L}W_\mu^3 - g_\mathrm{Y}B_\mu)(g_\mathrm{L}W^{\mu,3} - g_\mathrm{Y}B^\mu) &= \frac{v^2}{8}\begin{pmatrix} W_\mu^3 & B_\mu \end{pmatrix}\begin{pmatrix} g_\mathrm{L}^2 & -g_\mathrm{L}g_\mathrm{Y} \\ -g_\mathrm{L}g_\mathrm{Y} & g_\mathrm{Y}^2 \end{pmatrix}\begin{pmatrix} W^{\mu,3} \\ B^\mu \end{pmatrix} \\ &= \begin{pmatrix} A_\mu & \mathrm{Z}_\mu \end{pmatrix}\begin{pmatrix} \frac{1}{2}m_A^2 & 0 \\ 0 & \frac{1}{2}m_\mathrm{Z}^2 \end{pmatrix}\begin{pmatrix} A^\mu \\ \mathrm{Z}^\mu \end{pmatrix} . \end{aligned} \tag{1.73}$$

The gauge fields W_μ^3 and B_μ are therefore mixed via the weak mixing angle θ_W also called the *Weinberg angle*:

$$\begin{pmatrix} A_\mu \\ \mathrm{Z}_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_\mathrm{W} & \sin\theta_\mathrm{W} \\ -\sin\theta_\mathrm{W} & \cos\theta_\mathrm{W} \end{pmatrix}\begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix} , \tag{1.74}$$

where $\tan\theta_\mathrm{W} = \frac{g_\mathrm{Y}}{g_\mathrm{L}}$.

The GWS theory thus successfully unifies the EM and weak interactions while generating the masses of the weak bosons via the Brout–Englert–Higgs mechanism.

However, for several years, it was unclear whether such theories were renormalisable at the quantum level. This issue was resolved by 't Hooft, who proved that non-Abelian gauge theories with SSB are renormalisable [82]. Subsequently, together with Veltman, he developed the systematic renormalisation and regularisation framework required for practical loop calculations in the EW theory [83]. This work laid the foundation for precise theoretical predictions within the GWS framework, enabling rigorous comparisons with experimental data and solidifying the theory's status as a cornerstone of modern particle physics.

This remarkable theory predicted the existence of new bosons, leading to the discovery of the W [84, 85] and Z [86, 87] bosons with the UA1 [88] and UA2 [89] experiments in 1983.

Another proof of the powerful predictions of the GWS theory is given in Figure 1.3, which presents the Z lineshape via the process $\mathrm{e}^+\mathrm{e}^- \to \mathrm{hadrons}$ [90]. The coloured dots from various experiments at different centre-of-mass (CM) energies $\sqrt{s}$ represent the measured cross-section of this process. The solid line is the theoretical prediction from the GWS theory, in excellent agreement with the data. The peak at $\sqrt{s} \simeq 91\,\mathrm{GeV}$ corresponds to the mass of the Z boson, which is produced resonantly at this energy.

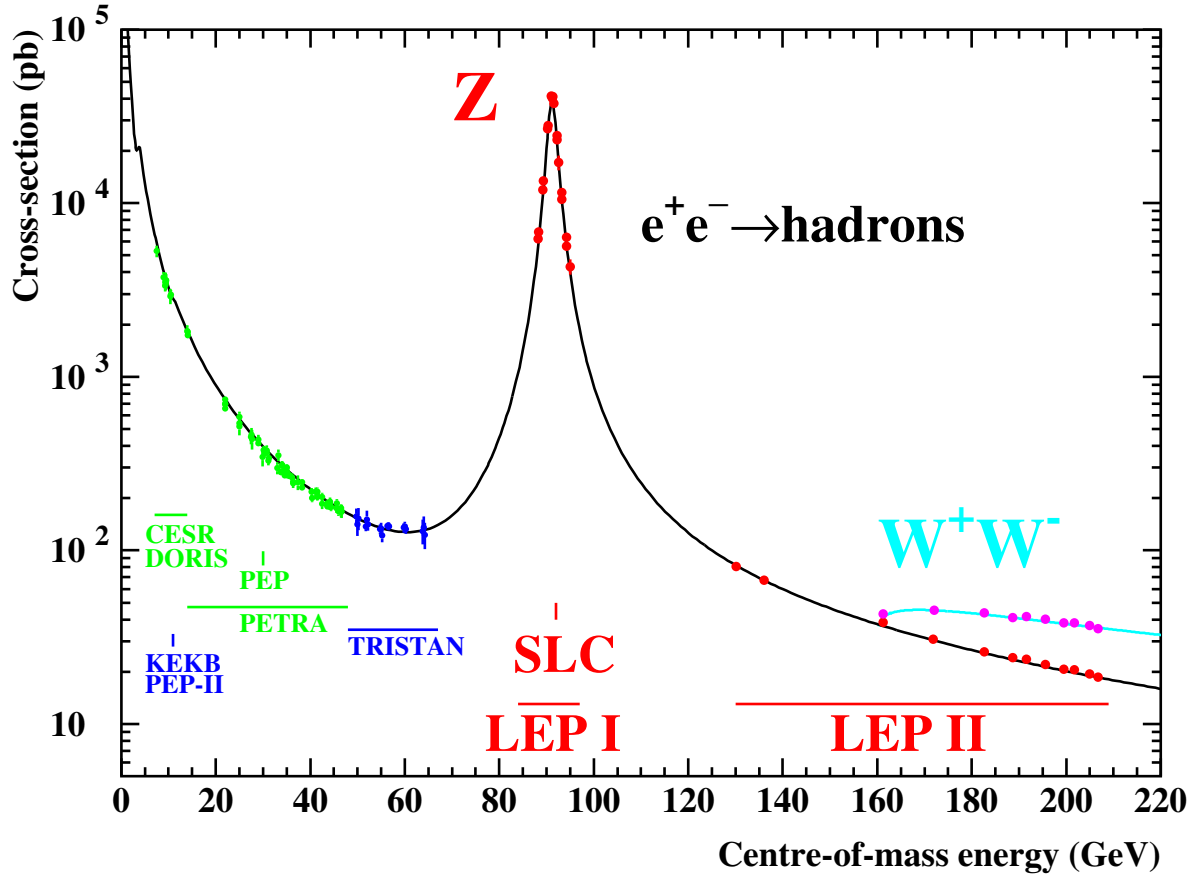


Figure 1.3: Measurement of the cross-section of the process $\mathrm{e^+e^-} \to \mathrm{hadrons}$ as a function of the CM energy $\sqrt{s}$ showing the Z resonance peak [90].

However, a final problem remains. Indeed, fermions are also observed with a mass in nature. Introducing a term like:

$$\mathcal{L}_\mathrm{m} = -m_f \overline{\psi}\psi = -m_f \left(\overline{\psi}_\mathrm{L}\psi_\mathrm{R} + \overline{\psi}_\mathrm{R}\psi_\mathrm{L}\right) , \tag{1.75}$$

would break the gauge invariance since the LH and RH fields lie in different representations of the gauge group $\mathrm{SU(2)_L}$ and possess different hypercharges under $\mathrm{U(1)_Y}$.

A perturbative expansion around the VEV would also involve terms linking gauge bosons to Goldstone bosons as in Equation 1.61.

Therefore, to address these issues, we need to introduce one final boson: the Higgs boson.

1.7 Higgs boson

If we pursue the perturbative expansion of the scalar field around the VEV done in Section 1.5, we can introduce the unitary gauge defined as:

$$\theta(x) = \frac{-\phi_2(x)}{g_\mathrm{H} v} \quad \Rightarrow \quad B_\mu \to B_\mu + \frac{1}{g_\mathrm{H} v}\partial_\mu \phi_2 \, , \tag{1.76}$$

leading to:

$$\phi(x) \to \mathrm{e}^{\mathrm{i}g_\mathrm{H}\theta(x)}\phi(x) = \mathrm{e}^{-\mathrm{i}\phi_2(x)/v}\phi(x) \, , \tag{1.77}$$

where ϕ is the scalar field defined in Equation 1.54. Following its definition in Equation 1.58, we can rewrite it at first order in the fields as:

$$\phi(x) = \frac{1}{\sqrt{2}}\left[v + \phi_1(x)\right]\mathrm{e}^{\mathrm{i}\phi_2(x)/v} \, . \tag{1.78}$$

Applying the unitary gauge transformation leads to:

$$\phi(x) \to \frac{1}{\sqrt{2}}\left[v + \phi_1(x)\right] = \frac{1}{\sqrt{2}}\left[v + h(x)\right] \, . \tag{1.79}$$

Therefore, the Goldstone boson ϕ_2 has been removed from the theory, absorbed by the gauge boson. The remaining field $h(x)$ is a massive scalar boson called the *Higgs boson* "H". From Equation 1.54, a term $-\lambda v^2 h^2$ appears, leading to the mass of the Higgs boson as $m_{\mathrm{H}} = \sqrt{2\lambda v^2}$. Other terms of self-interaction and interaction with the gauge bosons also appear.

If we now consider the full GWS theory, a perturbative expansion of the scalar field around the VEV defined in Equation 1.65 leads to:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1(x) + \mathrm{i}\phi_2(x) \\ v + h(x) + \mathrm{i}\phi_3(x) \end{pmatrix} , \tag{1.80}$$

where ϕ_1, ϕ_2, and ϕ_3 are the three Goldstone bosons associated with the broken symmetries. The Higgs boson is represented by the field $h(x)$. The unitary gauge transformation can be parametrised, at first order in the fields, as:

$$\begin{aligned} \theta^1(x) &= \frac{-2\phi_2(x)}{g_{\mathrm{L}} v} , \quad \theta^2(x) = \frac{-2\phi_1(x)}{g_{\mathrm{L}} v} , \\ \theta^3(x) &= \frac{2g_{\mathrm{L}}\phi_3(x)}{(g_{\mathrm{L}}^2 + g_{\mathrm{Y}}^2)v} , \quad \beta(x) = \frac{-2g_{\mathrm{Y}}\phi_3(x)}{(g_{\mathrm{L}}^2 + g_{\mathrm{Y}}^2)v} , \\ \Rightarrow \quad \phi(x) &\to \mathrm{e}^{\mathrm{i}g_{\mathrm{L}}\theta^i(x)T^i} \mathrm{e}^{\mathrm{i}g_{\mathrm{Y}}Y\beta(x)/2}\phi(x) , \end{aligned} \tag{1.81}$$

which is equivalent to a transformation $U(x)$ under the local gauge group $\mathrm{SU(2)_L}$, combined with a $\mathrm{U(1)_Y}$ transformation of parameter $\beta(x)$:

$$U(x) = \mathrm{e}^{\mathrm{i}g_{\mathrm{L}}\theta^i(x)T^i} . \tag{1.82}$$

This is a product of a transformation of the gauge fields as:

$$W_\mu^i T^i \to U W_\mu^i T^i U^\dagger + \frac{\mathrm{i}}{g_{\mathrm{L}}}(\partial_\mu U)U^\dagger , \quad B_\mu \to B_\mu - \partial_\mu \beta(x) , \tag{1.83}$$

leading also to the removal of the three Goldstone bosons from the theory, absorbed by the physical gauge bosons $\mathrm{W}^\pm$ and Z while the photon remains massless since $\mathrm{U(1)_{EM}}$ is unbroken. The choice of $\beta(x)$ ensures that the photon field $A_\mu = \sin\theta_{\mathrm{W}} W_\mu^3 + \cos\theta_{\mathrm{W}} B_\mu$ is left unchanged.

Expanding the Lagrangian with the unitary gauge applied leads to:

$$\mathcal{L}_{\mathrm{GWS}} \supset \frac{1}{2}\partial_\mu h \partial^\mu h - \lambda v^2 h^2 + \left[m_{\mathrm{W}}^2 \mathrm{W}_\mu^+ \mathrm{W}^{\mu,-} + \frac{1}{2} m_{\mathrm{Z}}^2 \mathrm{Z}_\mu \mathrm{Z}^\mu \right] \left(1 + \frac{h}{v}\right)^2 + \dots , \tag{1.84}$$

where we see the cubic and quartic interaction terms between the Higgs boson and the weak bosons appearing. The self-interaction terms of the Higgs boson are not shown here for clarity. We can also see clearly that the photon does not interact with the Higgs boson since it arises from the Q operator that remains unbroken.

Regarding the fermion masses, we define a doublet φ_{L} and a singlet φ_{R} as done for the weak interaction in Section 1.4. A term $\overline{\varphi}_{\mathrm{L}}\phi$ is therefore invariant under $\mathrm{SU(2)_L}$. This implies that a product like $\overline{\varphi}_{\mathrm{L}}\phi\varphi_{\mathrm{R}}$ ($[\overline{\varphi}_{\mathrm{L}}\phi\varphi_{\mathrm{R}}]^\dagger = \overline{\varphi}_{\mathrm{R}}\phi^\dagger\varphi_{\mathrm{L}}$) is also invariant. The same holds with the conjugate doublet $\mathrm{i}\sigma^2\phi^*$, which transforms like ϕ under $\mathrm{SU(2)_L}$ but carries the opposite hypercharge. Thus, the terms:

$$\mathcal{L}_{\mathrm{Yukawa}} = -\left(y_d\, \overline{\varphi}_{\mathrm{L}}\phi\, d_{\mathrm{R}} + y_u\, \overline{\varphi}_{\mathrm{L}} \left(\mathrm{i}\sigma^2\phi^*\right) u_{\mathrm{R}} + \mathrm{h.c.} \right) , \tag{1.85}$$

can be added to the GWS Lagrangian without breaking the gauge invariance, where d_{R} and u_{R} are the RH singlets of the lower and upper members of the doublet φ_{L}. The terms y_d and y_u are called the *Yukawa couplings* [91]. Since the VEV of ϕ lies in its lower component, $\langle\phi\rangle = (0, \frac{v}{\sqrt{2}})$, the y_d term gives mass to the lower members of the doublets (the down-type quarks and the charged leptons), while $\mathrm{i}\sigma^2\langle\phi\rangle^* = (\frac{v}{\sqrt{2}}, 0)$ lets the y_u term give mass to the upper members (the up-type quarks). For a fermion f with Yukawa coupling y_f (y_d or y_u), after the SSB, and the application of the unitary gauge, these terms in the GWS Lagrangian become:

$$\mathcal{L}_{\mathrm{Yukawa}} = -m_f \overline{f} f - \frac{m_f}{v} h \overline{f} f . \tag{1.86}$$

The fermion mass is therefore defined as $m_f = y_f \frac{v}{\sqrt{2}}$, and a coupling term between the fermions and the Higgs arises.

This theory was remarkably successful, confirmed by the discovery of the Higgs boson at the Large Hadron Collider (LHC) in 2012 by the ATLAS [1], visualised in Figure 1.4, and CMS [2] experiments.

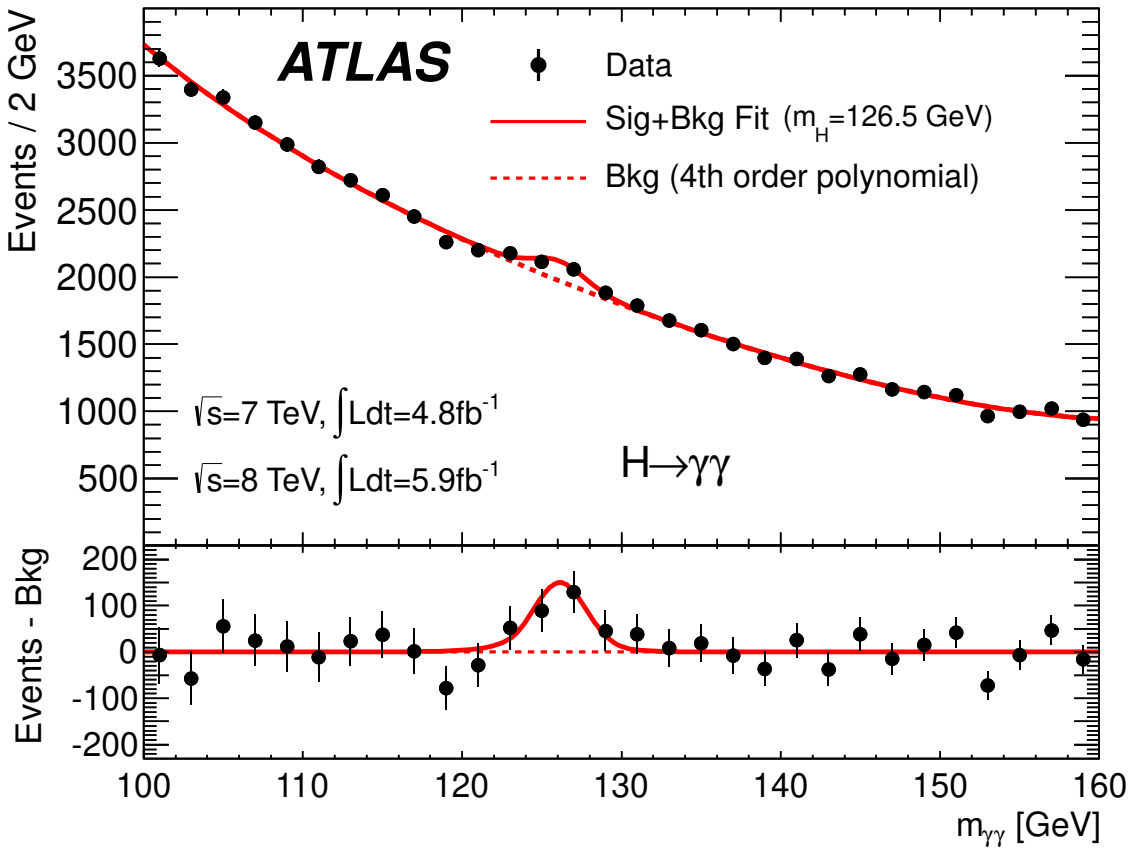


Figure 1.4: Invariant mass distribution of diphoton candidates [1].

1.8 Neutrino

The hint of a third quark generation arising from the CP violation in 1964 was also verified in the lepton sector. Indeed, in 1975, the tau lepton "τ" was discovered via the SLAC e^+e^- Stanford Positron–Electron Asymmetric Ring (SPEAR) collider and the Lawrence Berkeley Laboratory (LBL) magnetic detector [92]. This heavier lepton therefore opened the way for the third neutrino "ν_τ". That there are exactly three light families was established by the measurement of the Z lineshape [90] shown in Figure 1.3. The width of that resonance also counts the light neutrino species the Z can decay into, since every additional family would widen it, which puts the number of light neutrino families at three. The tau neutrino itself was discovered only in 2000 [93], equalising the number of leptons and quarks. When we introduced the neutrinos, we stated that they were very light. Indeed, until now, no mass term has been added for them. They could be considered massless. However, due to the observation of their oscillation between flavours by the Super-Kamiokande [94] and Sudbury Neutrino Observatory (SNO) [95] experiments, they must possess a mass. This results in a difference between the interacting eigenstates and the mass eigenstates as predicted by Pontecorvo in 1968 [96]. The Pontecorvo–Maki–Nakagawa–Sakata (PMNS) matrix [97] translates one basis to another, similarly to the CKM one:

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = U_{\text{PMNS}} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} . \tag{1.87}$$

If the neutrinos are Dirac particles (that is, their antiparticles are distinct particles), the matrix follows the same parametrisation as the CKM matrix given in Equation 1.53. However, if they are Majorana particles [98] (that is, their antiparticles are themselves), two extra complex phases are required inside a Majorana phase matrix. The full parametrisation is therefore:

$$U_{\text{PMNS}} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{i\alpha_1/2} & 0 \\ 0 & 0 & e^{i\alpha_2/2} \end{pmatrix} , \tag{1.88}$$

where $c_{ij} = \cos\theta_{ij}$ and $s_{ij} = \sin\theta_{ij}$. The angles θ_{12}, θ_{23}, θ_{13}, and the phase δ are equivalent to the ones in the CKM matrix while α_1 and α_2 are the two extra Majorana phases.

Also, since RH leptons do not interact with the W boson, they are not charged under this group and are set as singlets. We could treat the RH neutrinos as singlets too, as it was not explicitly mentioned before. However, it would be irrelevant since they are singlets under the other groups, $\mathrm{SU}(3)_\mathrm{C}$ and $\mathrm{U}(1)_\mathrm{Y}$, and therefore do not interact with them, nor with the Z boson.

1.9 Standard Model

The SM [34, 99] is the current theory used in particle physics. It is built as the culmination of an intricate exchange between theory prediction and experimental observation over the past century, benefiting from mathematical developments.

All the pillars of this theory and the main discoveries were detailed in this chapter. In this section, we will provide a summary of the current state of this model.

The SM is a gauge theory based on the local gauge group:

$$\mathrm{SU}(3)_\mathrm{C} \times \mathrm{SU}(2)_\mathrm{L} \times \mathrm{U}(1)_\mathrm{Y} \,, \tag{1.89}$$

where the $\mathrm{SU}(3)_\mathrm{C}$ group describes the strong interaction via the exchange of gluons g between quarks. The $\mathrm{SU}(2)_\mathrm{L}$ group describes the weak interaction while the $\mathrm{U}(1)_\mathrm{Y}$ group describes the hypercharge interaction. The latter two groups are unified into the EW interaction through the Glashow–Weinberg–Salam theory and the Brout–Englert–Higgs mechanism. The SM group is then:

$$\mathrm{SU}(3)_\mathrm{C} \times \mathrm{SU}(2)_\mathrm{L} \times \mathrm{U}(1)_\mathrm{Y} \xrightarrow{\mathrm{BEH}} \mathrm{SU}(3)_\mathrm{C} \times \mathrm{U}(1)_\mathrm{EM} \,. \tag{1.90}$$

The resulting EM interaction is mediated by the massless photon γ while the weak interaction is mediated by the massive $\mathrm{W}^\pm$ and Z bosons. A scalar boson also remains after the SSB: the Higgs boson H, which confers masses on the weak bosons and the fermions via their Yukawa couplings.

The fermions are divided into three generations following a mass hierarchy. We distinguish two types of fermions: the quarks and the leptons. The quarks interact via the strong, weak, and EM interactions. They exist only in confinement inside hadrons, bound by gluons through their colour charge. The leptons, on the other hand, do not interact via the strong interaction but only via the weak and EM interactions (except for the neutrinos).

All those particles form the building blocks of matter and can be classified into the SM summarised in Figure 1.5.

The resulting Lagrangian of the SM is:

$$\begin{aligned}
\mathcal{L}_\mathrm{SM} = \; & \underbrace{-\frac{1}{4} B_{\mu\nu} B^{\mu\nu} - \frac{1}{4} W^i_{\mu\nu} W_i^{\mu\nu} - \frac{1}{4} G^a_{\mu\nu} G_a^{\mu\nu}}_{\text{Kinetic energies and self-interactions of the gauge bosons}} \\
& + \underbrace{\overline{\varphi}_\mathrm{L} \gamma^\mu \left(\mathrm{i}\partial_\mu - g_\mathrm{Y} \frac{Y}{2} B_\mu - g_\mathrm{L} \frac{\sigma^i}{2} W^i_\mu \right) \varphi_\mathrm{L} + \overline{\varphi}_\mathrm{R} \gamma^\mu \left(\mathrm{i}\partial_\mu - g_\mathrm{Y} \frac{Y}{2} B_\mu \right) \varphi_\mathrm{R}}_{\text{Kinetic energies and electroweak interactions of the fermions}} \\
& + \underbrace{\left| \left(\mathrm{i}\partial_\mu - g_\mathrm{L} \frac{\sigma^i}{2} W^i_\mu - g_\mathrm{Y} \frac{Y}{2} B_\mu \right) \phi \right|^2 - \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2}_{\text{Electroweak boson and Higgs interaction, masses, and couplings}} \\
& - \underbrace{\left(y_d \, \overline{\varphi}_\mathrm{L} \phi \, d_\mathrm{R} + y_u \, \overline{\varphi}_\mathrm{L} \left(\mathrm{i}\sigma^2 \phi^* \right) u_\mathrm{R} + \mathrm{h.c.} \right)}_{\text{Yukawa interactions between fermions and Higgs field}} \\
& - \underbrace{\frac{1}{2} g_\mathrm{S} \overline{q} \gamma^\mu \lambda^a G^a_\mu q}_{\text{Strong interactions between quarks and gluons}} \,.
\end{aligned} \tag{1.91}$$

The number of free parameters of the SM is 19 if we consider the neutrinos as massless, 26 to 28 if we consider them as massive, depending on their nature (Dirac or Majorana). Those parameters are not predicted by the theory and have to be measured experimentally. They are summarised in Table 1.A.

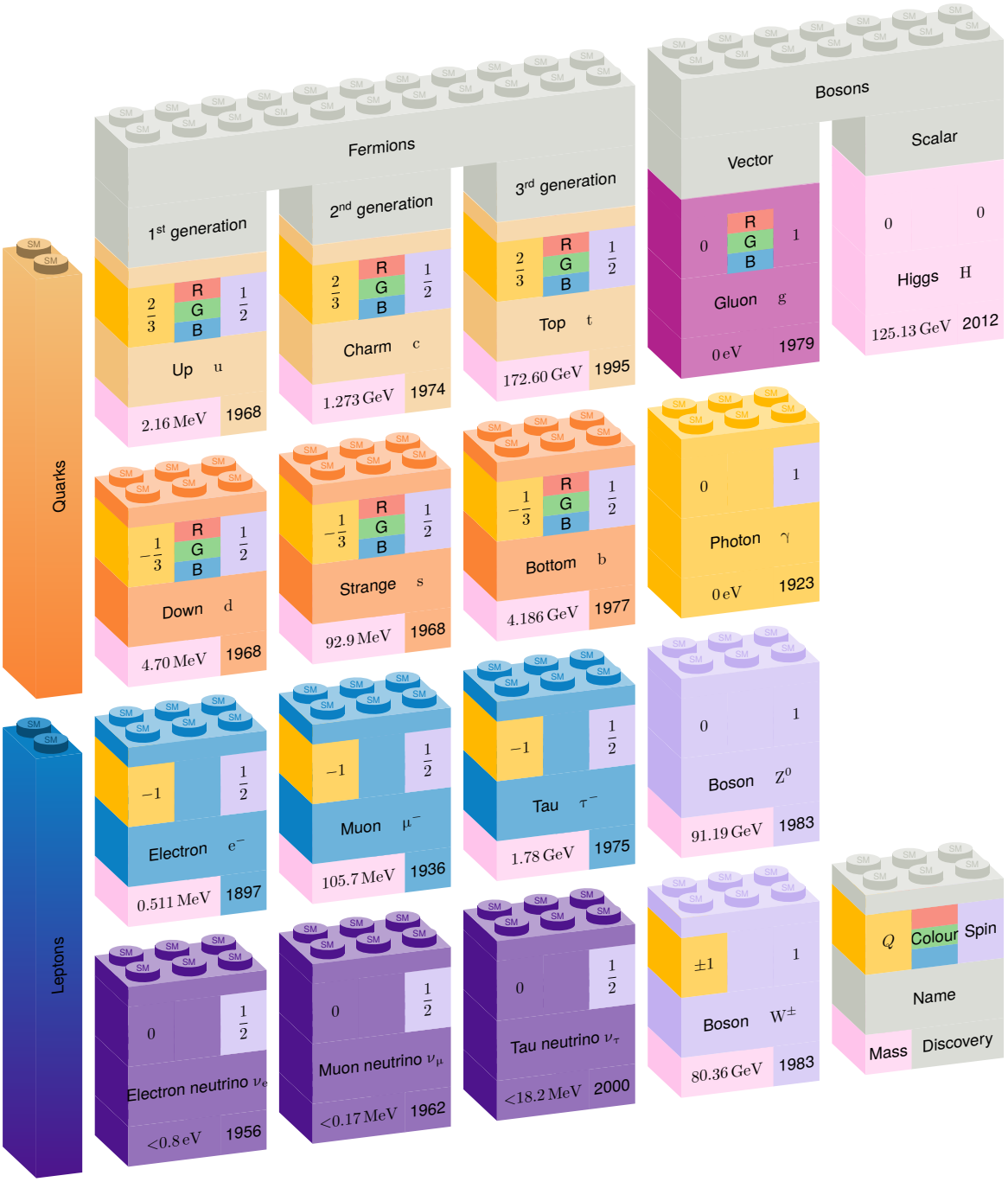


Figure 1.5: Particles predicted by the SM. The correlation between the particles and their interactions is represented by the colours of the corresponding bosons.

Table 1.A: Free parameters of the SM and their measured values [34].

Parameter type	Number	Original free parameter	Measured	Value
Gauge couplings	3	g_{S}		1.218
		g_{L}		0.652
		g_{Y}		0.357
Higgs sector	2	μ^2	m_{H}	125.13 GeV
		λ	v	246 GeV
Quark masses	6	m_{u}		2.16 MeV
		m_{c}		1.273 GeV
		m_{t}		172.60 GeV
		m_{d}		4.70 MeV
		m_{s}		92.9 MeV
		m_{b}		4.186 GeV
Lepton masses	3	m_{e}		0.511 MeV
		m_{μ}		105.7 MeV
		m_{τ}		1.78 GeV
CKM matrix	4	θ_{12}		13.0°
		θ_{23}		2.4°
		θ_{13}		0.2°
		δ		66.1°
QCD vacuum angle	1	θ_{QCD}		~ 0
Neutrino masses	3	$m_{\nu_1}, m_{\nu_2}, m_{\nu_3}$	Δm^2_{21}	$7.60 \cdot 10^{-5}\,\mathrm{eV}^2$
			$\lvert\Delta m^2_{32}\rvert$	$2.445 \cdot 10^{-3}\,\mathrm{eV}^2$
			$\sum_i m_{\nu_i}$	$< 0.12\,\mathrm{eV}$
PMNS matrix	4–6	θ_{12}	$\sin^2(2\theta_{12})$	0.851
		θ_{13}	$\sin^2(2\theta_{13})$	0.085
		θ_{23}	$\sin^2(\theta_{23})$	0.533
		δ		$\sim 218^\circ$
		α_1		unknown
		α_2		unknown

The SM has been highly successful in describing the fundamental particles and their interactions. Extensively tested experimentally, it has shown remarkable agreement with observations, including the discovery of the Higgs boson in 2012. To this day, no significant deviations from the SM predictions have been observed, therefore confirming its validity as the prevailing theory in particle physics. However, it is known that the SM is not a complete theory of fundamental interactions, thus pointing to physics beyond the SM.

1.10 Beyond the Standard Model

Despite its success and robustness, the SM has several limitations and open questions that motivate the search for theories beyond it. Some of the main reasons to explore physics beyond the SM are:

Dark matter. The SM does not provide a viable candidate for dark matter (DM), which constitutes about $27\,\%$ of the total energy density of the Universe. Ordinary baryonic matter accounts for only about $5\,\%$, while the remaining $\sim 68\,\%$ is attributed to dark energy [100]. DM is inferred from astrophysical

and cosmological observations, such as galaxy rotation curves, gravitational lensing, and the cosmic microwave background, but its particle nature remains unknown.

Neutrino masses. The observation of neutrino oscillations implies that neutrinos have mass, which is not accounted for in the original SM. This suggests the need for new physics to explain the origin of neutrino masses.

Matter–antimatter asymmetry. The SM cannot fully explain the observed matter–antimatter asymmetry in the Universe. Additional sources of CP violation beyond those present in the SM are required to account for this asymmetry.

Hierarchy problem. The large difference between the EW scale and the Planck scale ($10^{19}\,\mathrm{GeV}$) raises questions about the stability of the Higgs boson mass against quantum corrections. This problem suggests the existence of new physics at higher energy scales.

Unification of forces. The SM does not provide a unified description of all fundamental forces. Gravity is not unified with the three other fundamental forces. Also, the coupling constants do not converge at a single point at very high energy. The search for a theory that unifies all forces, such as string theory or grand unified theory (GUT) models, is an ongoing area of research.

Naturalness. The SM contains several parameters that appear to be fine-tuned, raising questions about the naturalness of the theory. This has led to the exploration of theories that address this issue, such as supersymmetry.

All these open questions and limitations motivate the search for BSM theories that can provide a more complete understanding of fundamental physics. Experimental efforts are ongoing, searching for hints of new physics. One way is to study longitudinal weak boson scattering at high energies, which is sensitive to new physics effects.

1.11 Vector boson polarisation

The gauge vector bosons W_μ and B_μ both possess two degrees of freedom. After the SSB and the acquisition of mass, those bosons become the physical $\mathrm{W}^\pm$, Z, and A_μ bosons. The photon A_μ remains massless and thus still possesses only two degrees of freedom. However, the massive $\mathrm{W}^\pm$ and Z bosons now possess three degrees of freedom. These degrees of freedom are manifested as polarisation states.

The polarisation describes the spin state of the vector boson with respect to a chosen spin-quantisation axis. In the helicity basis, this axis is chosen to be the direction of motion of the boson, and the helicity is defined as:

$$h = \frac{\vec{S}\cdot\vec{p}}{|\vec{p}|}\,. \tag{1.92}$$

The corresponding polarisation states are therefore not Lorentz invariant in general, since their definition depends on the chosen reference frame and spin-quantisation axis. In the latter basis, the polarisation states are labelled by the helicity eigenvalue $\lambda = -1, 0, +1$.

The helicity is Lorentz invariant for massless particles, but not for massive particles, since a suitable boost can reverse the direction of the momentum. In the ultra-relativistic regime, helicity becomes approximately stable under ordinary boosts because reversing the momentum direction would require a boost velocity very close to the speed of light.

The vector bosons have spin 1. Therefore, the projection of their spin onto a chosen quantisation axis can take three values $m = -1, 0, +1$, corresponding to the three possible spin-projection eigenvalues. For a given momentum mode and polarisation state λ, the vector field can be written as:

$$B^\mu(x) \propto \varepsilon^\mu_\lambda(p)\,\mathrm{e}^{-\mathrm{i}p\cdot x}\,, \tag{1.93}$$

where $\varepsilon^\mu_\lambda(p)$ is the polarisation four-vector.

Starting from the Proca Lagrangian [101] and its corresponding equation of motion:

$$\mathcal{L}_{\mathrm{Proca}} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}m^2 B_\mu B^\mu \quad \Rightarrow \partial_\nu\left(\partial^\nu B^\mu - \partial^\mu B^\nu\right) + m^2 B^\mu = 0\,, \tag{1.94}$$

we can derive the Lorenz condition by taking the four-divergence of the equation of motion:

$$m^2 \partial_\mu B^\mu = 0\,. \tag{1.95}$$

For a massive vector boson ($m \neq 0$) this leads to the Lorenz condition:

$$\partial_\mu B^\mu = 0 \Leftrightarrow \varepsilon^\mu p_\mu = 0\,. \tag{1.96}$$

This condition reduces the number of independent components of the four-vector B^μ from four to three, corresponding to the three polarisation states of a massive vector boson.

For a massless vector boson, Equation 1.94 becomes the Maxwell equations:

$$\mathcal{L}_{\text{Maxwell}} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad \Rightarrow \partial_\nu\left(\partial^\nu A^\mu - \partial^\mu A^\nu\right) = 0\,. \tag{1.97}$$

This allows us to choose a gauge such as the Lorenz gauge $\partial_\mu A^\mu = 0$ or the Coulomb gauge $\nabla \cdot \vec{A} = 0$. The gauge freedom, together with a gauge choice and the residual freedom left after this choice, removes the unphysical components of A^μ. A massless vector boson therefore has only two physical transverse polarisation states.

If we consider a vector boson V that propagates along the z-axis with a four-momentum $p^\mu = (E, 0, 0, p_z)$, up to phase and sign conventions, the two transverse polarisation states can be represented as:

$$\varepsilon^\mu_\pm(p) = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ \mp 1\\ -\mathrm{i}\\ 0\end{pmatrix}\,. \tag{1.98}$$

These correspond to the two circular transverse polarisation states, conventionally labelled as RH and LH, or equivalently $\lambda = +1$ and $\lambda = -1$. If we consider the spin-1 operator $\vec{S}$ along the z-axis:

$$S_z = \begin{pmatrix}0 & -\mathrm{i} & 0\\ \mathrm{i} & 0 & 0\\ 0 & 0 & 0\end{pmatrix}\,, \tag{1.99}$$

which consists of a generalisation of the Pauli matrices defined in Equation 1.19 that formed the spin operators for spin-$\frac{1}{2}$ particles as $S_i = \frac{1}{2}\sigma_i$. We can verify that the two transverse polarisation states are eigenstates of the spin operator:

$$S_z\vec{\varepsilon}_\pm = \pm\vec{\varepsilon}_\pm\,, \tag{1.100}$$

where $\vec{\varepsilon}_\pm$ is the spatial part of the polarisation four-vector $\varepsilon^\mu_\pm$. The two transverse polarisation states correspond to the two possible spin orientations of the vector boson along its direction of motion. As both circular polarisation states are transverse to the direction of motion, they are collectively denoted by T. In this notation, the transverse category corresponds to the sum of the LH and RH contributions.

For a massive vector boson, the longitudinal polarisation state has its spatial component along the direction of motion. Nevertheless, as stated in Equation 1.96, as a four-vector it satisfies the transversality condition:

$$\varepsilon^\mu_0 p_\mu = 0\,. \tag{1.101}$$

This leads to the following expression for the longitudinal polarisation state:

$$\varepsilon^\mu_0(p) = \frac{1}{m_V}\begin{pmatrix}p_z\\ 0\\ 0\\ E\end{pmatrix}\,, \tag{1.102}$$

where m_V is the mass of the vector boson, with this form arising from the covariant normalisation:

$$\varepsilon^\mu_\lambda \varepsilon^*_{\lambda'\mu} = -\delta_{\lambda\lambda'}\,. \tag{1.103}$$

A representation of the polarisation four-vectors in the helicity basis for a vector boson propagating along the z-axis is shown in Figure 1.6 as well as the spin orientation of the polarisation states with respect to the chosen spin-quantisation axis. The longitudinal polarisation state corresponds to the zero eigenvalue of the spin operator, hence it does not have a spin orientation. The two transverse polarisation states are orthogonal to the direction of motion and correspond to the two possible spin orientations of the vector boson along its direction of motion.

As stated at the beginning of this section, the polarisation is not Lorentz invariant. Therefore, the polarisation states can be different in different reference frames. For instance, consider the rest frame of the vector boson, where:

$$p^{\text{rest}}_\mu = (m_V, 0, 0, 0)\,. \tag{1.104}$$

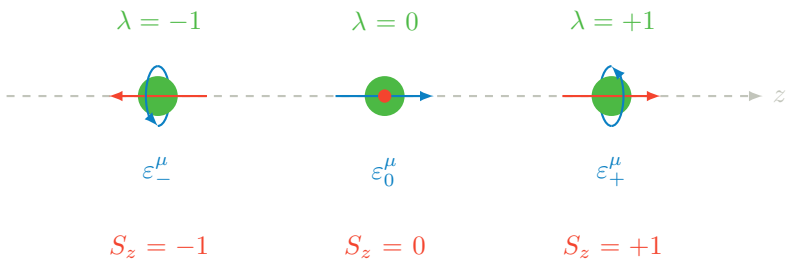


Figure 1.6: Scheme showing polarisation four-vectors for a vector boson propagating along the z-axis. The two transverse polarisation states $\varepsilon^\mu_\pm$ are orthogonal to the direction of motion, while the longitudinal polarisation state ε^μ_0 has its spatial component along the direction of motion. The spin orientation of the polarisation states is defined with respect to the chosen spin-quantisation axis, which in this case is the direction of motion. For the longitudinal state there is no spin orientation as it corresponds to the zero eigenvalue of the spin operator.

To transform from the moving frame to the boson rest frame, one applies a boost along the z-axis with velocity $-\beta$. Following Equation 1.7, we get the Lorentz factor $\gamma = \frac{E}{m_V}$ and the boost velocity $\beta = \frac{p_z}{E}$. The boost matrix is then:

$$\Lambda^\mu{}_\nu = \begin{pmatrix} \gamma & 0 & 0 & -\gamma\beta \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\gamma\beta & 0 & 0 & \gamma \end{pmatrix} . \tag{1.105}$$

Applying this boost to the longitudinal polarisation state in the moving frame, we get:

$$\Lambda^\mu{}_\nu \varepsilon^\nu_0(p) = \frac{1}{m_V} \begin{pmatrix} \gamma p_z - \gamma\beta E \\ 0 \\ 0 \\ -\gamma\beta p_z + \gamma E \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \varepsilon^\mu_0(p^{\text{rest}}) . \tag{1.106}$$

Meanwhile, the transverse polarisation states remain unchanged:

$$\Lambda^\mu{}_\nu \varepsilon^\nu_\pm(p) = \varepsilon^\mu_\pm(p) . \tag{1.107}$$

Therefore, the choice of reference frame affects the explicit form of the polarisation states. The choice of spin-quantisation axis and reference frame defines the polarisation basis. Boosts along this axis leave the transverse polarisation vectors unchanged, whereas the longitudinal polarisation vector transforms covariantly between its moving-frame and rest-frame expressions.

In the ultra-relativistic limit, the difference between the longitudinal and transverse polarisation states becomes large. Indeed, only the longitudinal polarisation state depends on the mass of the vector boson. In the computation of matrix elements (MEs) involving vector bosons, the high-energy behaviour of the longitudinal polarisation vector is especially important:

$$\varepsilon^\mu_0(p) = \frac{p^\mu}{m_V} + \mathcal{O}\left(\frac{m_V}{E}\right) . \tag{1.108}$$

As a result, individual contributions to amplitudes involving longitudinal vector bosons can contain apparent powers of $\frac{E}{m_V}$. In a gauge-invariant calculation, however, cancellations between diagrams are essential and ensure the correct high-energy behaviour.

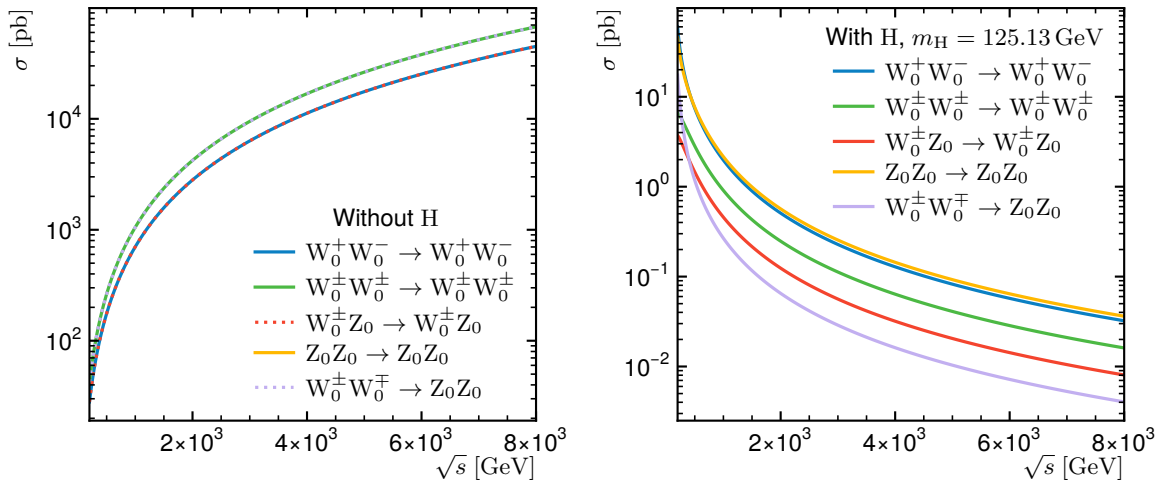
This high-energy behaviour is the origin of the Goldstone boson equivalence theorem (GBET) [102]. At energies much larger than the vector boson mass, scattering amplitudes involving external longitudinal vector bosons are related, up to corrections of order $\frac{m_V}{E}$, to the corresponding amplitudes involving the associated Goldstone bosons. In this sense, a high-energy longitudinally polarised vector boson behaves like its corresponding Goldstone mode as the mass becomes negligible ($|\vec{p}| \simeq E \gg m_V$).

The longitudinal polarisation state therefore can be seen as the term $\partial_\mu\phi_2$ in Equation 1.60 leading to the Feynman diagram in Equation 1.61. This term arises from the Goldstone boson ϕ_2 absorbed by the gauge boson during the SSB via the unitary gauge transformation of Equation 1.76.

A key process to study vector boson polarisation is vector boson scattering (VBS). In particular, the scattering of longitudinally polarised vector bosons is directly connected to the EW spontaneous symmetry breaking mechanism. Without the Higgs boson contribution, the high-energy cancellations are incomplete and the longitudinal VBS amplitude grows with energy, eventually violating perturbative unitarity. The

inclusion of the Higgs boson exchange diagrams restores the required cancellations and preserves the consistency of the EW theory.

This behaviour is illustrated in Figure 1.7, which compares the longitudinal VBS cross-section as a function of the CM energy $\sqrt{s}$ with and without the Higgs boson contribution. The details of the calculation can be found in Appendix A, where the GBET is used at LO.



(a) Without Higgs boson

(b) With Higgs boson

Figure 1.7: Comparison of the scattering cross-section of longitudinally polarised vector bosons as a function of the CM energy $\sqrt{s}$ with and without the Higgs boson contribution.

Therefore, studying the polarisation states of vector bosons, more specifically in VBS processes, provides a powerful probe of the EW symmetry breaking mechanism and the role of the Higgs boson in preserving unitarity at high energies. Any significant deviation from the SM prediction could indicate new physics in the EW symmetry breaking sector or anomalous interactions between the Higgs boson and EW gauge bosons.

2 The Large Hadron Collider

The European Organization for Nuclear Research (CERN) (originally *Conseil Européen pour la Recherche Nucléaire*), was founded in 1954 in Meyrin, close to Genève, Switzerland, and to the French border (Figure 2.1).

This laboratory was made to bring Europe back to a competitive level after the Second World War. To face brain drain to the United States of America and the fact that particle physics required resources beyond those of individual European nations, cooperation was needed and CERN was born [103].



Figure 2.1: On 17th May 1954, the first shovel of earth was dug on the Meyrin site.

Probing the structure of matter at the smallest scales is made possible through the use of particle accelerators. With such apparatuses, we can accelerate particles and collide them with either a fixed target or other moving particles. New particles are then created that we need to detect.

The main purpose of CERN was and is to deliver accelerators and experiments built around them to observe the particles produced. From the first accelerator to the present day, many have followed one another, each time increasing the size and energy of the collisions and deepening our understanding of particle physics.

The first one built was the Synchro-Cyclotron [104, 105] in 1957, then the Proton Synchrotron (PS) [106–108] in 1959 which with the Gargamelle [109] bubble chamber discovered the neutral currents (NCs) [110, 111], the Super Proton Synchrotron (SPS) [112] in 1976 which led to the discovery of the W [84, 85] and Z [86, 87] bosons with the UA1 [88] and UA2 [89] experiments in 1983. In 1985, a ring $26.7\,\mathrm{km}$ long, $3.8\,\mathrm{m}$ wide, and approximately $100\,\mathrm{m}$ deep was dug to build the Large Electron–Positron Collider (LEP) [113–115] to carry out $\mathrm{e^+e^-}$ collisions, starting operation in 1989. With the four experiments Apparatus for LEP PHysics (ALEPH) [116], DEtector with Lepton, Photon and Hadron Identification (DELPHI) [117], Omni-Purpose Apparatus at LEP (OPAL) [118], and L3 [119] around the ring, they determined from the width of the resonance that there are three light neutrino families, and obtained a precise measurement of the Z boson lineshape [90] (Figure 1.3 in Section 1.6).

The LEP was removed in 2001 to make space for the LHC [5–7].

2.1 Facility

The LHC is currently the largest and most powerful particle collider ever built. With a length of $26.7\,\mathrm{km}$ and reusing the tunnel of LEP, it now collides protons of the same sign in two beams rotating in opposite directions. Due to a lack of space, both of them share the same yoke and are curved by superconducting magnets made of niobium–titanium cables, cooled down to $1.9\,\mathrm{K}$ with superfluid helium and producing a magnetic field of $8.33\,\mathrm{T}$ [120]. Figure 2.2 shows a three-dimensional (3D) cutaway view of an LHC dipole magnet in the tunnel, illustrating the two beam pipes inside the shared yoke.

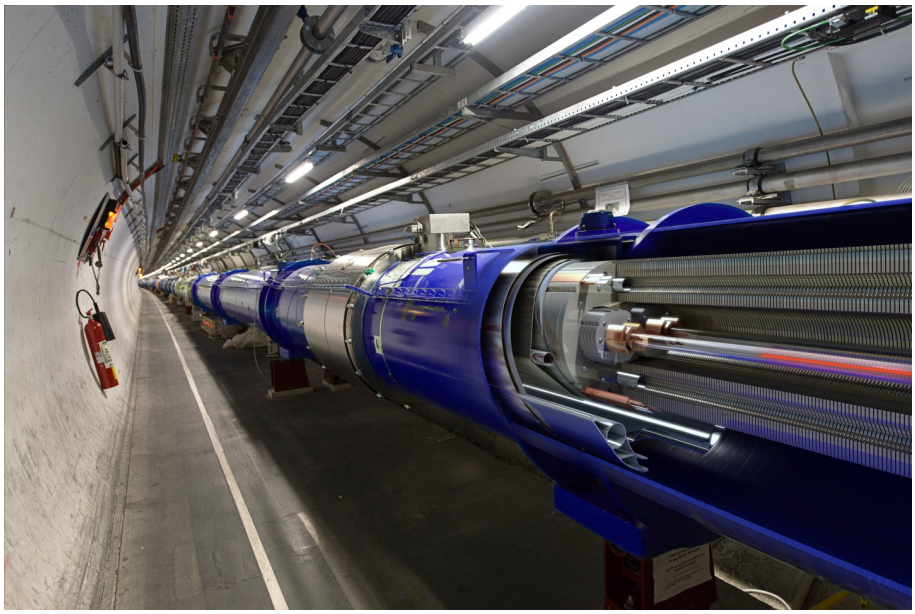


Figure 2.2: 3D cut of the LHC dipole magnet in the tunnel showing the two beams.

The magnets of the LHC are unable to bring the proton from its initial energy up to the final beam energy. To counter this, most of the past CERN accelerators are used in sequence as shown in Figure 2.3. Until 2018, the source was a bottle of dihydrogen H_2, ionised using a Duoplasmatron [121] to form protons H^+. Then the second LINear ACcelerator (LINAC), LINAC-2 (built in 1978 [122] and upgraded over the years to meet the requirements of the new accelerators, particularly the LHC [123]) accelerated them up to $50\,\mathrm{MeV}$. Starting from 2022, the LINAC-4 [124] took over and changed the injection system. The source is now a bottle of hydrogen (H) that is injected into a plasma chamber. An externally applied electric field heats the atoms and forms a plasma confined by magnetic fields. The radio frequency (RF) field applied excites a resonance of the atoms that can attract an electron, thus creating an H^- ion. The LINAC-4 then accelerates these ions to $160\,\mathrm{MeV}$. For both LINAC-2 and LINAC-4, the next step is injection into the Proton Synchrotron Booster (PSB). For LINAC-4, the ions are stripped of their two electrons upon injection, leaving only protons. This allows more particles to accumulate in the synchrotron, simplifies the injection process, reduces beam loss at injection, and results in a higher-brightness beam (that is, the beam contains more particles, is more tightly focused, and has lower divergence).

The PSB is made of four distinct circular rings that increase the energy of the beams to $1.4\,\mathrm{GeV}$ and $2\,\mathrm{GeV}$, respectively, for LINAC-2 and LINAC-4. After this initial boost, the beams enter the PS, which brings them to $26\,\mathrm{GeV}$, and then the SPS where they are accelerated to $450\,\mathrm{GeV}$. The beam is then split into two and injected into its final accelerator: the LHC. They will be accelerated up to $7\,\mathrm{TeV}$ for the nominal energy [120].

This process is not linear, and the beams are not continuous. The injection is organised in units called *bunches*. There are various injection schemes to adjust different beam parameters. Each time, a defined number of bunches is injected into an intermediate accelerator, following the chain described above. At each stage of the chain, a set number of bunches is injected into the accelerator until it is filled. Once a ring is full, its beam is accelerated to the nominal energy for that ring and then transferred to the next accelerator, where the bunches are stacked until that machine is filled as well. This process is repeated until the SPS is filled. Its beam is then injected clockwise or anticlockwise into the LHC: this represents one step described in Figure 2.4(a). Around ten injections are required to fill the LHC as seen in Figure 2.4(b).

The LHC is organised as described in Figure 2.5. It is not fully circular and is made of eight straight sections of $545\,\mathrm{m}$, while the arcs are $2.45\,\mathrm{km}$ long. The injection takes place in Octants 2 and 8. The

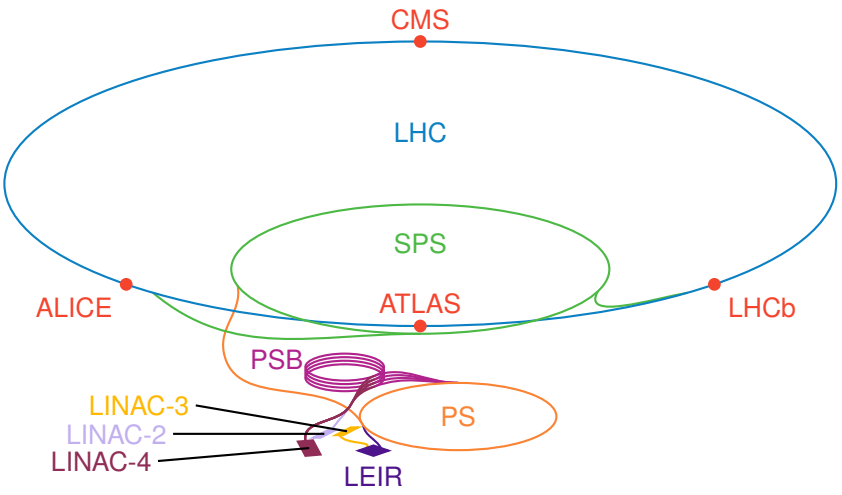
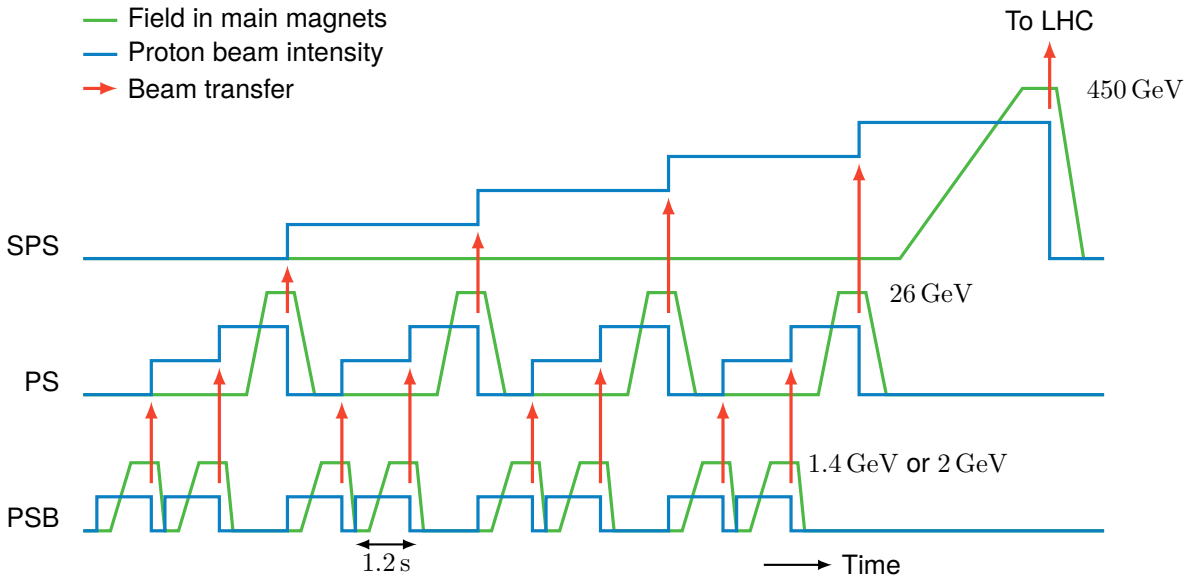
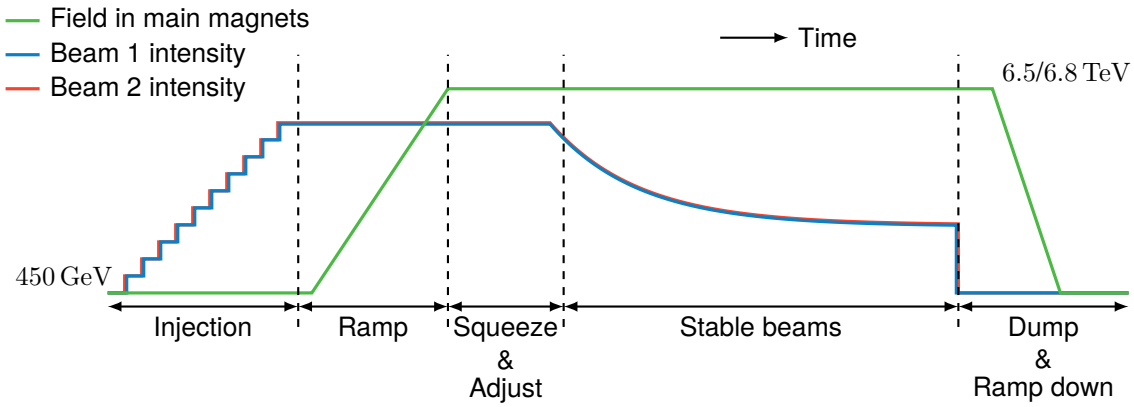


Figure 2.3: The CERN accelerator complex.



(a) Filling scheme throughout the CERN accelerator complex



(b) Steps of a full run of the LHC

Figure 2.4: Injection scheme of every accelerator involved in the chain and detail of the run steps.

acceleration will be done by RF cavities of $400\,\mathrm{MHz}$ located in Octant 4. They bring the protons from their initial energy of $450\,\mathrm{GeV}$ close to their nominal energy, which corresponds to the ramp phase in Figure 2.4(b). Magnets for cleaning the beams and re-collimating them are situated in Octants 3 and 7. They intervene during the squeeze and adjust phase in Figure 2.4(b). The whole process takes approximately two hours.

Each bunch is made of around $1.15 \cdot 10^{11}$ protons and spaced every $25\,\mathrm{ns}$, which represents a frequency of $40\,\mathrm{MHz}$. Since the LHC is $26.7\,\mathrm{km}$ long and the protons are accelerated to nearly the speed of light thanks to the RF cavities, there is theoretically space for 3564 bunches. However, the nominal value is only 2808. Indeed, the kicker magnet for deflecting the beam, located in Octant 6, takes a few microseconds to ramp up. Meanwhile, no proton should pass through this device. In fact, one bunch takes less than $90\,\mu\mathrm{s}$ to do a full turn. The absorber of the dump is made of a $7\,\mathrm{m}$-long segmented carbon cylinder of $700\,\mathrm{mm}$ diameter, contained in a steel cylinder. This is water-cooled and surrounded by about $750\,\mathrm{t}$ of concrete and iron (Fe) shielding. Such a strong absorber is needed. The total energy of the beams, considering the 2808 bunches made of $1.15 \cdot 10^{11}$ protons at $7\,\mathrm{TeV}$ each, represents $362\,\mathrm{MJ}$. This is enough energy to bring $1.5\,\mathrm{t}$ of gold from $25\,^{\circ}\mathrm{C}$ to its melting point.

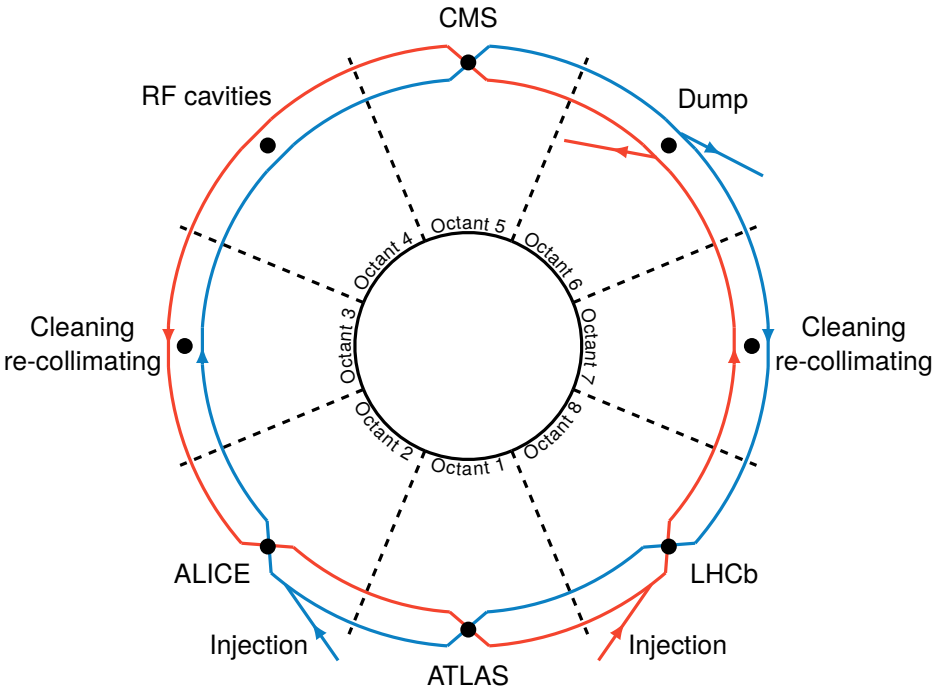


Figure 2.5: Layout of the LHC.

Once the beams are stable, collisions can start at every interaction point (IP). There are four of them and each of them hosts an experiment: A Toroidal LHC ApparatuS (ATLAS) in Octant 1, A Large Ion Collider Experiment (ALICE) in Octant 2, Compact Muon Solenoid (CMS) in Octant 5, and Large Hadron Collider beauty experiment (LHCb) in Octant 8. ATLAS [8] and CMS [9] are multipurpose detectors maximising the full exploitation of the proton–proton (pp) collisions. ALICE [125] is specialised in heavy-ion collisions (LINAC-3 and Low Energy Ion Ring (LEIR) in Figure 2.3 form the injection chain). Thanks to collisions of lead (Pb) ions (Pb–Pb), it can create QGP [126] as predicted by QCD in Section 1.3 and study its properties. LHCb [127] is an asymmetrical detector in the shape of a spectrometer to fully study the decays of b-hadrons, highly correlated with CP violation measurement. This allows us to extend the limits of the SM and probe new physics as mentioned in Section 1.10.

During the collisions, a key parameter to monitor is the luminosity.

2.2 Luminosity

We saw that from the early stages of CERN to the present day, the accelerators were always built to accelerate the particles to higher energy than their predecessors to increase the discovery reach of higher-energy phenomena. This energy is called the *centre-of-mass energy* $\sqrt{s}$. These parameters

determine what kind of physics processes can be observed. The energy scale of the LHC was necessary to reach the Higgs mass and production. The production rate of a given physical process is defined by:

$$\frac{\mathrm{d}\,N}{\mathrm{d}\,t} = \sigma \cdot L\,. \tag{2.1}$$

An important implication of this is that statistical analyses of rare processes require a large amount of data, which can be achieved by increasing the luminosity.

The cross-section σ represents all the physics involved in the process, such as the probability of interaction and of decay. It is a function of $\sqrt{s}$. The instantaneous luminosity, L, depends purely on the proton beams and describes the number of collisions that take place at the IP.

It can be written as follows:

$$L = \frac{N_\mathrm{b}^2 n_\mathrm{b} f_\mathrm{rev} \gamma_\mathrm{r}}{4\pi\epsilon_\mathrm{n}\beta^*} F\,, \tag{2.2}$$

where the parameters are the following:

- N_b the number of particles per bunch
- n_b the number of bunches per beam
- $f_\mathrm{rev} = \frac{\beta c}{2\pi R}$ the revolution frequency, with $\beta c \simeq c$ and $2\pi R = 26.7\,\mathrm{km}$ we get $f_\mathrm{rev} = 11.245\,\mathrm{kHz}$
- $\gamma_\mathrm{r} = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}$ the relativistic factor
- ϵ_n the normalised transverse beam emittance which dictates the proton density in the beam. It should be as low as possible
- β^* the β function at the IP. It represents the envelope of the beam. The squeezing steps in Figure 2.4(b) have to make it as small as possible
- F the geometric luminosity reduction factor due to the crossing angle and defined in Equation 2.3.

The factor F is made of θ_c, the full crossing angle at the IP, σ_z, the root mean square (RMS) bunch length, and σ^*, the transverse RMS. It is expressed as:

$$F = \frac{1}{\sqrt{1 + \left(\frac{\theta_\mathrm{c}\sigma_z}{2\sigma^*}\right)^2}}\,. \tag{2.3}$$

Since the cross-section is always constant in time for a fixed proton energy, we can integrate Equation 2.2 over the running period of the collider to obtain the integrated luminosity $\mathcal{L}$. We can thus get the event yield of the process with the corresponding cross-section.

The instantaneous luminosity L thus depends exclusively on multiple factors related to the beam's shape. During collisions in the stable beams step in Figure 2.4(b), the bunches and beams are "damaged" by the collisions, which affects the instantaneous luminosity. Several ways exist to maintain a constant L through all the collisions. This is called *levelling* and is performed with the insertion magnets around the IPs, by changing the crossing angle θ_c, the envelope β^*, or the overlap of two bunches at the IP.

When the bunch loses too much intensity because of the protons destroyed in the collisions, we dump the beam as described before. This is done when it is more efficient, luminosity-wise, to restart the injection chain and the steps in Figure 2.4(b). The collision steps take between ten and twelve hours.

A run corresponds to the entire process represented in Figure 2.4(b) from injection to dump. A run number is associated with it. This number started with the first ever data taking of the LHC and has never been reset since. The event number, meanwhile, is reset at every run and counts the recorded events. The luminosity has to be measured and recorded all the time. In this way, a run is divided into luminosity blocks representing around one minute of data taking, that is a few tens of thousands of events, over which the luminosity is assumed to be constant.

2.3 Pile-up

Another key parameter to take into account, apart from the luminosity, is the pile-up μ. In particle colliders and particularly in hadron ones, collisions do not happen one by one. At every bunch crossing, for one collision of interest, several others can occur in the vicinity of the first one due to additional hard-scattered pp collisions. The mean number of interactions at each bunch crossing is given by [128]:

$$\langle\mu\rangle = \frac{L \cdot \sigma_\mathrm{inelastic}}{n_\mathrm{b} f_\mathrm{rev}}\,. \tag{2.4}$$

The pp inelastic scattering cross-section is $\sigma_{\text{inelastic}}$[†], L is the instantaneous luminosity, n_{b} the number of bunches per beam, and f_{rev} the revolution frequency.

As seen in Equation 2.4, an increase in luminosity also leads to an increase in the pile-up. While having more luminosity helps to accumulate more statistics for a given analysis, more pile-up makes it harder to identify and reconstruct the event of interest. Hence, compromises need to be made. Levelling techniques such as β^* levelling, which starts a fill with a larger β^* and lowers it as the beams degrade, have become the main solution to limit the pile-up while keeping a high luminosity.

2.4 Timeline

The first beam was injected into the LHC on 10 September 2008. Unfortunately, due to a failure in an electrical connection, a magnet quench damaged over fifty dipole magnets, leading to a delay of fourteen months. The first data-taking period started in 2010, and since then, a tight schedule has been followed as represented in Figure 2.6.

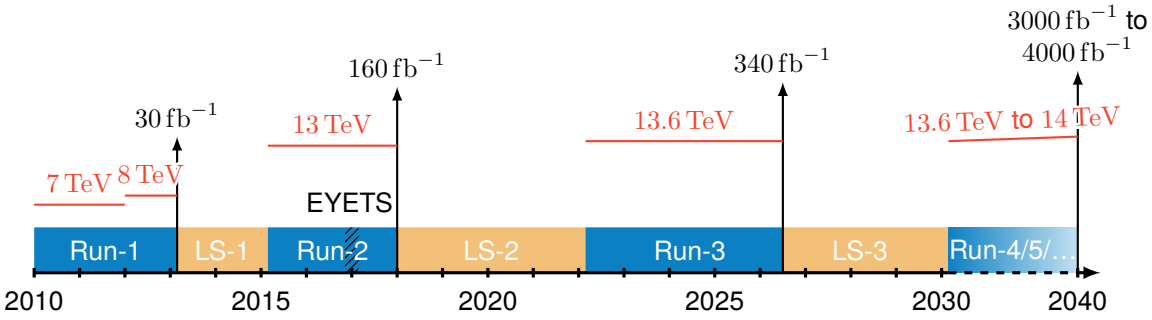


Figure 2.6: Timeline.

The data taking works in runs. The first one, Run-1, was from 2010 to 2013. Each year starts with commissioning, first without beams and then with beams. After this, the proton physics starts for several months, a few weeks or months are usually allocated to heavy-ion physics, and the year is concluded by a technical stop. During the first two years, the centre-of-mass energy $\sqrt{s}$ was $7\,\text{TeV}$ and slightly increased to $8\,\text{TeV}$ during the last year of data taking. This is far from the nominal value of $14\,\text{TeV}$, but it was chosen to ensure safe and stable operation to avoid any new incident. At the end of Run-1, the total luminosity delivered by the LHC reached $30\,\text{fb}^{-1}$. Following the first run, the first Long Shutdown (LS-1) provided two years of complete halt of data taking, focused on maintenance and upgrade. During this period, the magnet splices were repaired, and the collimation scheme was upgraded to increase the $\sqrt{s}$ to reach the nominal values.

Run-2, from 2015 to 2018, was operated at $\sqrt{s} = 13\,\text{TeV}$ and reached $160\,\text{fb}^{-1}$. An Extended Year-End Technical Stop (EYETS) was done at the end of 2016 to perform powering tests, magnet exchange, and dump replacement. The LS-2 from 2018 to 2022 saw the installation of the new LINAC-4 [124], an upgrade of the injection beam energy in the PSB to reduce the beam emittance, and new cryogenics plants that separate the cooling circuits of the superconducting RF cavities from those of the superconducting magnets. This LS was longer than expected due to the COVID-19 pandemic.

Run-3, from July 2022 to June 2026, raised $\sqrt{s}$ to $13.6\,\text{TeV}$ thus delivering $340\,\text{fb}^{-1}$ (Table 2.A).

The LS-3 will bring several important modifications, bringing the LHC to the new High-Luminosity LHC (HL-LHC) [12]. It will therefore operate at a higher luminosity and thus pile-up. The experiments will undergo massive updates to cope with these new challenging conditions. A first step towards this new era was the Phase-I upgrade, implemented during LS-2.

To reach the desired luminosity, the beam intensity has to be increased substantially. The beams must also stay tightly focused at the IPs to maximise the collision probability. New technologies are needed to achieve these goals. More powerful magnets and new RF cavities will be installed during the LS-3. Made of niobium–tin, these magnets will be able to generate a magnetic field of $12\,\text{T}$. Therefore, they can squeeze the beams more effectively before colliding, increasing the instantaneous luminosity. However, the smaller β^* and the higher bunch intensity also require a larger crossing angle at the IP, which reduces the luminosity. To compensate for this effect, RF crab cavities will be used to tilt the proton bunches before they meet, maximising their overlap. A total of sixteen crab cavities (some of them were already installed during Run-3) will be installed on both sides of the ATLAS and CMS experiments.

[†]The value of $\sigma_{\text{inelastic}}$ is not precisely known. A reference value of $80\,\text{mb}$, adopted by all LHC experiments for pp collisions at $\sqrt{s} = 13\,\text{TeV}$, is used.

A summary table showing the various key parameters to monitor the condition of the beam through the year and runs is given in Table 2.A.

Table 2.A: Parameters of the LHC through the runs [5, 12, 34]. The reduction for the beam envelope is done via levelling.

Parameter			Nominal	Run-1	Run-2	Run-3	HL-LHC
Beam energy		[TeV]	7.0	4.0 (3.5)	6.5	6.8	7.0
Crossing angle	θ_c	[µrad]	285	290	$320 \rightarrow 260$	$320 \rightarrow 270$	500
Bunch length	σ_z	[cm]	7.55	9.4	8	9.7	7.6
Normalised transverse beam emittance	ϵ_n	[µm]	3.75	2.5	2.2	2.5	2.5
Beam envelope	β^*	[m]	0.55	0.6	$0.3 \rightarrow 0.25$	$1.2 \rightarrow 0.3$	$1 \rightarrow 0.15$
Number of particles per bunch	N_b	$[\times 10^{11}]$	1.15	1.6	1.1	1.6	2.2
Number of bunches	n_b		2808	1380	2544	2452	2748
Luminosity lifetime	τ_L	[h]	15	6	8.4	12	7.5
Peak luminosity	L	$[10^{34}\mathrm{cm}^{-2} \cdot \mathrm{s}^{-1}]$	1.0	0.77	2.1	2.0	5
Delivered luminosity	$\mathcal{L}$	$[\mathrm{fb}^{-1}]$	—	23.3 (6.1)	160	340	250/yr
Mean pile-up	$\langle\mu\rangle$		19	18	35	55	200

2.5 Collider physics

Throughout the history of particle accelerators, we can identify two categories of them: circular and linear. In both cases, two beams of the same energy colliding head-on reach a CM energy of:

$$\sqrt{s} = 2E_{\mathrm{beam}}\,, \tag{2.5}$$

whereas a beam of energy E_{beam} hitting a fixed target of mass m only reaches $\sqrt{s} \simeq \sqrt{2E_{\mathrm{beam}}m}$, which is why colliders are used to explore the highest energies.

The linear collider is simpler to design, build, and operate. The beam is used only once and collides either with another beam or a fixed target. One experiment is built at the IP and no beam dump is required. On the other hand, the circular collider is much more complex to operate. The beams are accelerated and used multiple times, requiring injection, acceleration, collision, and dump systems. Several experiments can be built around the ring at different IPs.

The beam also loses energy at each turn because of synchrotron radiation. It evolves as:

$$E_{\mathrm{rad}} \propto \frac{E_{\mathrm{beam}}^4}{m_{\mathrm{beam}}^4 R}\,, \tag{2.6}$$

where R is the radius of the ring. This loss is therefore significant for light particles such as electrons and positrons, making circular colliders for these particles even more challenging. For protons, which have a ratio $\frac{m_p}{m_e} \simeq 1836$, the loss is strongly reduced and becomes negligible for heavier particles, which is why circular colliders remain the preferred choice.

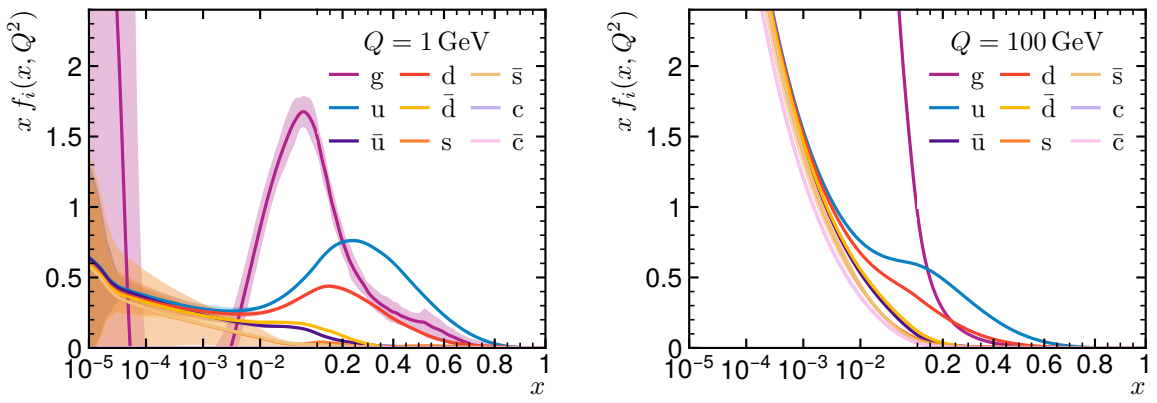
This comes with a drawback. Protons are composite particles made of quarks and gluons as explained in Section 1.3. Thus, when two protons collide, it is actually their constituents that interact. The CM energy of the collision involved is therefore not known precisely and can vary from one event to another. This is why luminosity is a key parameter to monitor in hadron colliders.

Due to colour confinement explained in Section 1.3, quarks and gluons cannot be free. A proton is formed of three valence quarks (two up quarks and one down quark) held together by gluons. Short-lived quark–antiquark pairs and gluons are constantly created and annihilated inside the proton. They are referred to as *sea quarks*.

Since the coupling constant of the strong interaction α_S is small at high energy scales such as those of the collider, we can consider the protons as a sum of free partons [129] (quarks and gluons) each carrying a fraction x of the proton momentum. The probability of finding a parton with a fraction x of the proton momentum at a scale Q^2 is given by the parton distribution function (PDF) $f_i(x, Q^2)$, where i represents the type of parton (for example an up quark, a down quark, or a gluon). It represents the probability density function of the parton inside the proton. These functions are not calculable from first principles and have to be determined experimentally, extrapolating them from measurements of deep inelastic scattering experiments, colliding electrons and protons.

As probability distributions, the PDFs are normalised such that:

$$\int_0^1 \sum_i x f_i(x, Q^2)\,\mathrm{d}\,x = 1\,, \tag{2.7}$$



(a) **(b)**

Figure 2.7: PDFs multiplied by the momentum fraction x at different energy scales. The uncertainties are also represented. Computed via LHAPDF [133] using the `NNPDF3.0` set [134].

which means that the sum of the momentum fractions of all partons inside the proton equals the total momentum of the proton.

Also, due to the sea quarks, normalisation for each parton is not constant, but the valence quarks have to respect:

$$\int_0^1 \left[f_{\mathrm{u}}(x,Q^2) - f_{\bar{\mathrm{u}}}(x,Q^2)\right] \mathrm{d}\,x = 2\,, \tag{2.8}$$

$$\int_0^1 \left[f_{\mathrm{d}}(x,Q^2) - f_{\bar{\mathrm{d}}}(x,Q^2)\right] \mathrm{d}\,x = 1\,. \tag{2.9}$$

The PDF depends on the energy scale Q^2 due to the phenomenon of parton shower (PS) and gluon radiation. As the energy scale increases, the probability of finding gluons and sea quarks inside the proton also increases. This is described by the Dokshitzer–Gribov–Lipatov–Altarelli–Parisi (DGLAP) equations [130–132], which govern the evolution of the PDF with respect to Q^2.

Figure 2.7 shows the PDF multiplied by the momentum fraction x at two different energy scales. We can see how the distribution of partons changes with Q.

At high x values, the valence quarks dominate, while at low x values, the gluons take over. At low scales, the valence quarks carry roughly half of the proton momentum and the gluons most of the remainder. The fraction carried by the sea quarks is smallest at low Q^2 and grows with Q^2 through gluon splitting, while the gluons carry close to half of the proton momentum at LHC scales.

During a pp collision, we find:

- hard-scattering processes: interactions between partons with high momentum transfer
- soft processes: interactions with low momentum transfer, often involving multi-parton interactions
- initial-state radiation (ISR) and final-state radiation (FSR): emission of gluons or photons before or after the hard scattering
- beam remnants: leftover parts of the protons that did not participate in the hard scattering.

The hard-scattering processes are the ones of interest, as they can lead to the production of new particles.

The cross-section of a given process X in pp collisions can be calculated using the PDF with the cross-section defined in Equation 1.17 as follows:

$$\sigma_{\mathrm{pp}\to X} = \sum_{i,j} \int_0^1 \int_0^1 f_i(x_1,Q^2) f_j(x_2,Q^2) \hat{\sigma}_{ij\to X}(\hat{s})\,\mathrm{d}\,x_1\,\mathrm{d}\,x_2\,, \tag{2.10}$$

where:

- i and j are the types of partons involved in the interaction
- x_1 and x_2 are the momentum fractions of the partons from each proton
- $\hat{\sigma}_{ij\to X}(\hat{s})$ is the partonic cross-section for the subprocess $ij \to X$ at the partonic CM energy $\hat{s} = x_1 x_2 s$, with s being the square of the pp CM energy.

Another important aspect of hadronic collisions is the phenomenon of *hadronisation*. Quarks and gluons produced in high-energy collisions cannot exist as free particles due to the property of colour confinement in QCD, as explained before in Section 1.3. The quarks and gluons generated in the initial hard-scattering process or the ones not involved in it (the underlying event) will undergo a process called hadronisation or fragmentation, where they combine to form colour-neutral hadrons such as mesons and baryons. They will radiate gluons and create quark–antiquark pairs, leading to a cascade of particles of progressively lower energy that eventually hadronise into stable or long-lived hadrons that can be detected experimentally. This collimated spray of hadrons is referred to as a *jet* "j" [135, 136].

The top quark is a notable exception to this behaviour. Due to its large mass and extremely short lifetime, it decays into a bottom quark and a W boson before hadronisation can occur.

The bottom quark, on the other hand, hadronises promptly into b-hadrons whose longer lifetime lets them decay at a vertex displaced from the primary vertex, which is a key feature for identifying bottom quarks in collider experiments. This is known as b-*tagging*.

To conclude this comparison, lepton colliders offer a well-defined initial state and cleaner events, well suited to precision measurements, mostly of EW processes, while hadron colliders reach the highest energies.

The future of particle colliders, after the HL-LHC, could see the development of new linear colliders such as the International Linear Collider (ILC) or the Compact Linear Collider (CLIC) for electron–positron collisions, or even new circular colliders like the Future Circular Collider (FCC) for leptons in its first phase and hadrons in its second phase. These projects aim to push the energy and luminosity frontiers further, enabling the exploration of new BSM physics. Hadron colliders provide precision measurements thanks to the large collision statistics they deliver.

3 The ATLAS Experiment

The ATLAS experiment [137–139] is a general-purpose detector, designed to have the broadest field of observation of the physics phenomena occurring in high-energy pp collisions. It is located in a cavern $100\,\mathrm{m}$ below ground near the main CERN site at the LHC IP-1.

The detector itself is cylindrical, measuring $46\,\mathrm{m}$ in length and $25\,\mathrm{m}$ in diameter for a total weight of approximately $7000\,\mathrm{t}$, making it the largest detector ever built for a collider. ATLAS is built in a concentric-layer structure (like an onion) to maximise the solid angle around the IP. Fully forward–backward symmetric, it comprises multiple sub-detectors arranged in layers parallel to the beam line that form the barrel, and is completed by two end-caps, one at each end. We can distinguish three main parts as illustrated in Figure 3.1: the **Inner Detector (ID)**, the **Calorimeters**, and the **Muon Spectrometer (MS)**. Particles produced in a collision will move outward and pass through these different detectors, each serving a specific purpose. The ID provides tracking of outgoing charged particles, and the Calorimeters measure the energy of electrons, photons, and hadrons. At the same time, the MS handles tracking and momentum measurement of the weakly interacting muons passing through the previous parts. Each sub-detector is designed to detect specific types of particles and measure some of their properties. Only by making them work in synergy can the complete reconstruction of an event be achieved. The whole is paired with a **Magnet System** to bend the trajectory of charged particles to determine their momentum and sign.

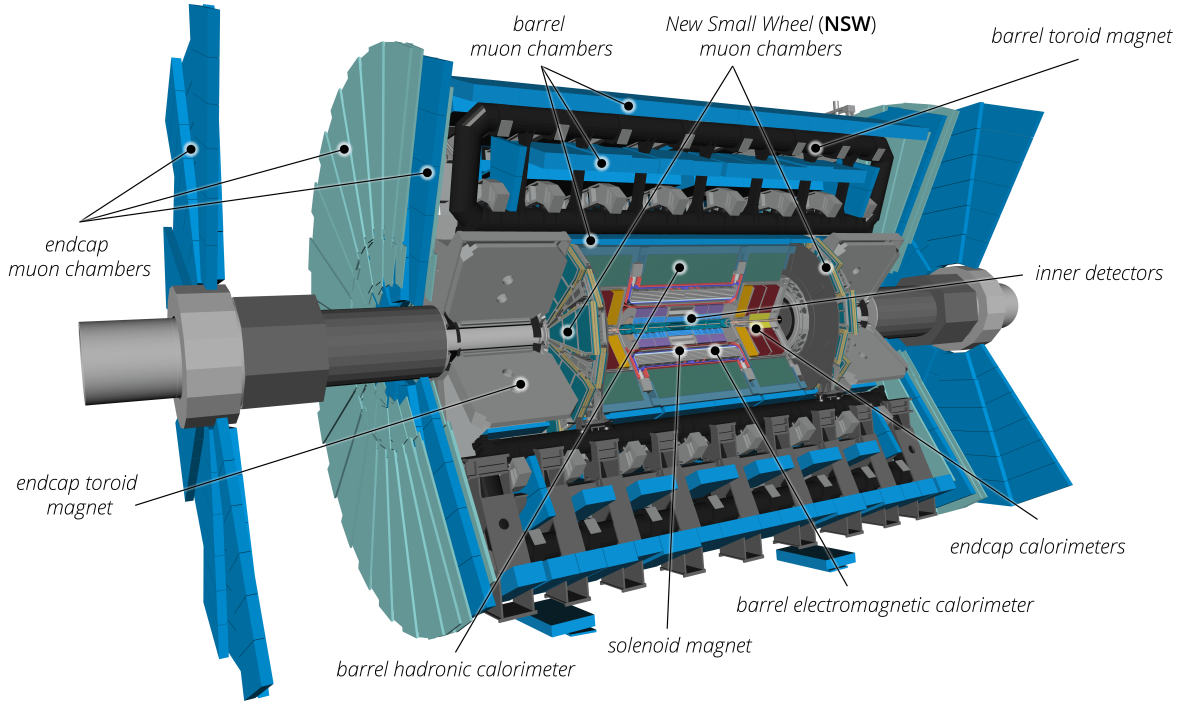


Figure 3.1: Scheme of the ATLAS experiment in its Run-3 configuration [138].

3.1 Coordinate system

A Cartesian right-handed coordinate system is defined to study collisions inside the experiment, with the origin set at the IP. The z-axis is oriented along the direction of the anticlockwise circulating beam, x points towards the centre of the LHC and y points upwards, shown in Figure 3.2(a). Since the (x, y) plane is perpendicular to the beam, it is referred to as *transverse* T, whereas the z-axis is labelled as *longitudinal* L. Side-A of the detector corresponds to positive z values, while Side-C corresponds to negative z values.

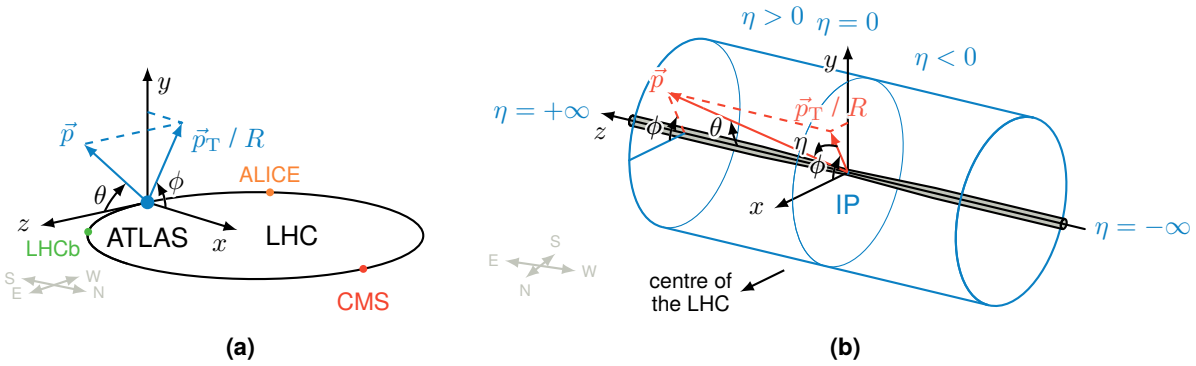


Figure 3.2: ATLAS coordinate system within the LHC (a), and ATLAS internal coordinates (b).

Because of the detector's geometry, a spherical coordinate system is used in ATLAS. The azimuthal angle ϕ is measured around the beam axis, while the polar angle θ is defined with respect to the z-axis. The radial direction is R.

Following the formalism introduced in Section 1.1, we can define additional key quantities used in collider physics.

The rapidity is defined using the longitudinal momentum p_z:

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right) . \tag{3.1}$$

Differences in rapidity are Lorentz invariant under boosts along the longitudinal z-axis, and rapidity transforms additively under such boosts.

Most particles can be considered massless at the energy scale of the LHC collisions. Indeed, the mass of particles of a few $\mathrm{GeV/MeV}$ is far lower than their energy: $m \ll |\vec{p}|$, so one can make the approximation that $E \simeq |\vec{p}|$. In this way, we can redefine the rapidity to match this approximation. This new quantity, called *pseudorapidity* η, is defined by:

$$\eta = \frac{1}{2} \ln \left(\frac{|\vec{p}| + p_z}{|\vec{p}| - p_z} \right) . \tag{3.2}$$

In the context of the approximation, we therefore have $y \simeq \eta$. It can be rewritten using the polar angle θ as shown in Equation 3.3. The link between pseudorapidity and the polar angle can be seen as:

$$\eta = -\ln \left(\tan \frac{\theta}{2} \right) . \tag{3.3}$$

This quantity shares the properties of the rapidity defined previously regarding Lorentz invariance and additivity under boosts. The distance between two objects ΔR is defined as:

$$\Delta R = \sqrt{(\Delta y)^2 + (\Delta \phi)^2} , \tag{3.4}$$

using $\Delta\phi$ and Δy, respectively, the difference in ϕ and y. In the low mass approximation, we can replace Δy by $\Delta\eta$, giving us a purely angular quantity.

The quantity is also Lorentz invariant, a property particularly useful in the context of pp collisions, where the partons carry different fractions of the longitudinal momentum, resulting in final states generally boosted along the beam axis. More details about the parton model are given in Section 2.5.

Finally, two important quantities are defined in the transverse plane, perpendicular to the beam axis, in which R is the radial coordinate (Figures 3.2(a) and 3.2(b)): the transverse momentum $p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2}$, and the transverse energy E_{T} defined by:

$$E_{\mathrm{T}} = \sqrt{m^2 + p_{\mathrm{T}}^2} = \frac{E}{\cosh y} . \tag{3.5}$$

Any particle can then be perfectly described at each instant with the polar coordinates $(p_{\mathrm{T}}, \theta, \phi)$ (or (R, θ, ϕ)), but in practice it is preferred to use η instead of θ, giving us $(p_{\mathrm{T}}, \eta, \phi)$. Both coordinate systems are shown in the geometry of the ATLAS detector in Figure 3.2(b).

3.2 Inner Detector

The ID [140, 141] is the closest sub-detector to the beam and the IP. Its role is to provide high-precision measurements of the trajectory of charged particles (labelled as tracks) originating from the pp collisions. Doing so, it must also distinguish primary vertices (the original points where the collisions happen) and secondary vertices (auxiliary points of interaction displaced from the primary vertex due to particles' decay).

To achieve those goals, the ID must be made of minimal material to avoid additional IPs and minimise the energy loss. It must also withstand the heavy radiation due to its proximity to the IP, which could damage the electronic circuits and thus deteriorate the data taking.

The ID, designed to meet these demanding requirements, consists of three sub-systems from the beam pipe outward: the Pixel Detector, the Semi-Conductor Tracker (SCT), and the Transition Radiation Tracker (TRT). They form a cylinder made of a barrel and two end-caps visible in Figure 3.3(a) for a total coverage in $|\eta|$ up to 2.5.

Every charged particle crossing a detector layer leaves a signal, referred to as a *hit*. By using reconstruction algorithms to connect these hits, we can reconstruct the tracks of the particles produced in the collision. Different technologies are employed in each sub-system to meet specific requirements.

3.2.1 Pixel Detector

The Pixel Detector [142] is the closest to the beam: its role will mostly be to reconstruct the primary and secondary vertices.

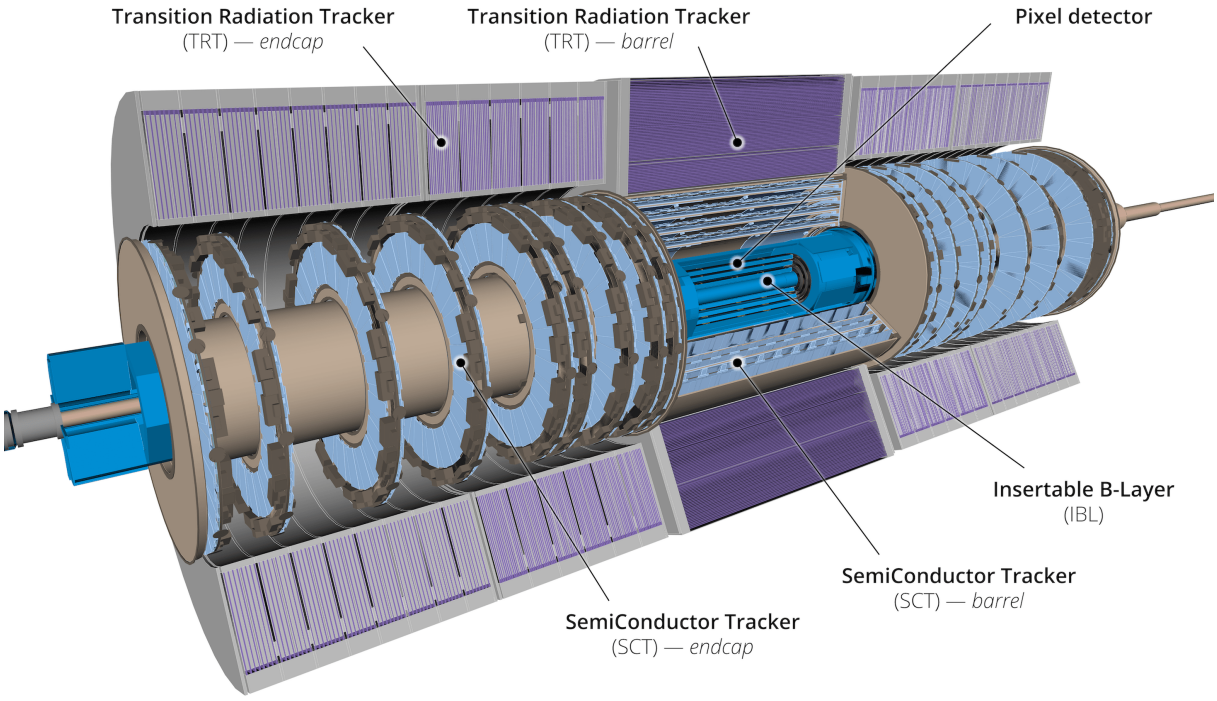
The pixel sensor technology relies on silicon (Si) and its properties. This material is a semiconductor with a filled valence band and an empty conduction band at zero temperature. No free carriers are present in its ground state, but the crystal can be doped, that is, by introducing impurities. Atoms in the 15th periodic table column donate extra electrons and make the crystal n-type, with electrons as free carriers. Conversely, atoms in the 13th column accept electrons, leaving holes as free carriers, and make the crystal p-type. The sensors are then composed of a junction between an n-doped region and a p-doped region, leading to a difference in charge density between the two sides. A central region is formed without free carriers, called the *depletion region* and encapsulated in an electric field due to the difference in charge at each end. An external electric field is applied to enhance the effect. To reduce the noise coming from thermal excitation, the Si is cooled down to about $-7\,^{\circ}\mathrm{C}$.

When an incoming charged particle passes through the sensor, electrons in the crystal get excited, moving from the valence band to the conduction band. This results in electron–hole pairs that drift in opposite directions from the depletion region towards the p-layer and n-layer, generating a current recorded by the electronic readout.

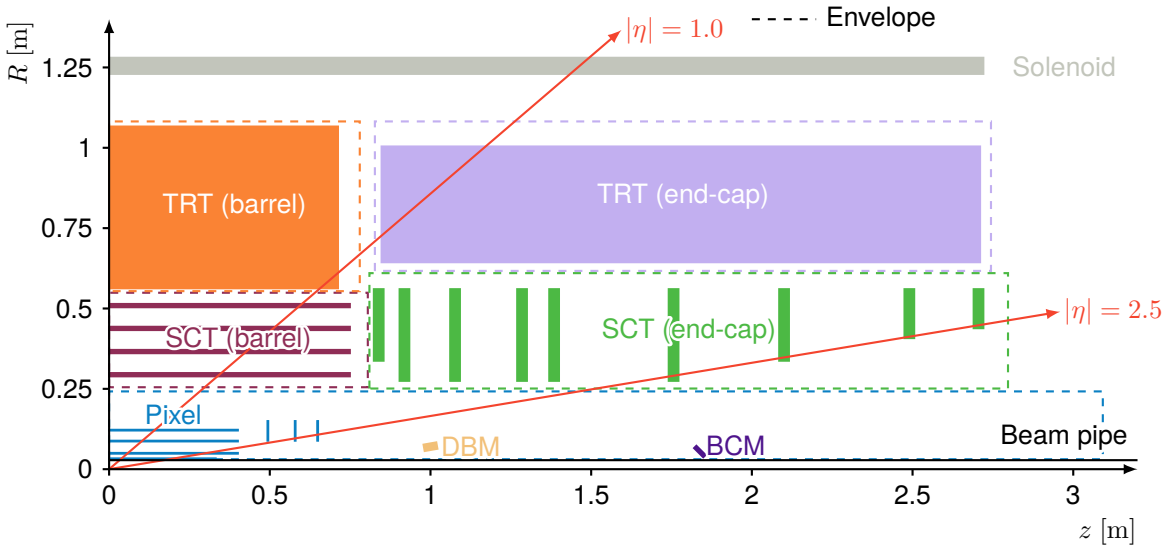
Some important features of Si sensors have to be taken into account, dictating the current design of the Pixel Detector. The sensors used in ATLAS are called n-in-n, meaning that the connection between the sensor and the front-end electronics is done with a highly doped n-type (also referred to as n^{+}) region called a pixel distributed like a grid on the bulk, which is lightly n-doped. At the same time, the backplane is heavily p-doped (also referred to as p^{+}). The depletion region grows from the backplane to the pixel. This turns out to be efficient against irradiation. If the energy transferred to a Si atom is larger than about $25\,\mathrm{eV}$, defects in the crystal, such as vacancies or interstitials, appear, leading to increased leakage current. In the high-radiation environment of the Pixel Detector, the following effects occur:

- Electrons or holes can be trapped by defects and no longer reach the electrodes: the signal is reduced. Since electrons move faster than holes in Si, collecting electrons with n^{+} pixels makes the detector more robust against trapping
- The sensor can undergo type inversion, where the n-type bulk can change to p-type. The depletion will then grow from the pixel side inward, and we still collect electrons at the pixels.

Thanks to the n-in-n design, the pixels can still work when facing defects caused by radiation. To make it even more robust and increase the lifetime of the sensor, impurities of oxygen are added to the bulk to make the crystal less likely to have defects and slow down a potential type inversion. On the other hand, the drawback of this design is its cost, which is high due to the double-sided process required. A scheme showing how an n-in-n pixel sensor works to detect a particle is shown in Figure 3.4(a).

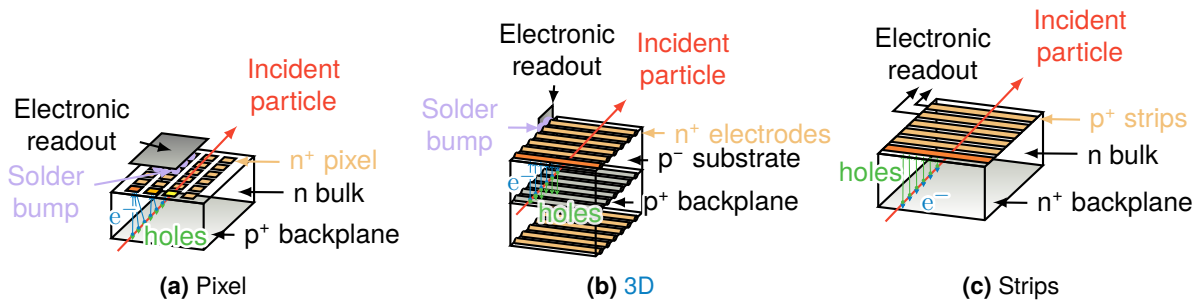


(a)



(b)

Figure 3.3: Inner Detector in 3D [138] (a) and its layout (b).



(a) Pixel **(b)** 3D **(c)** Strips

Figure 3.4: Scheme of different types of Si-based sensors used in the ID.

The Pixel Detector is made of 92 million pixels (therefore 92 million electronic channels) distributed in four concentric layers in the barrel with 1736 sensor modules and three disks in each end-cap with 288 modules. The innermost layer in the barrel is called Insertable b-Layer (IBL) [143] and was installed during the LS-1 to improve the tracking performance by reducing the distance between the IP and the first layer and thus improving the b-tagging. The b quark has a longer lifetime than other quarks, allowing it to travel a greater distance before decaying. The distinction between the primary vertex and this secondary vertex is essential to distinguish the b events from the rest. This new layer uses planar Si sensors in the central region, while in the forward region, 3D sensors are installed. These sensors are still based on Si, but with several layers of n^+ and p^+ column electrodes penetrating the lightly p-doped bulk, so that the depletion region grows laterally between the columns as represented in Figure 3.4(b). When a particle passes through and excites the medium, electrons (and holes) will now have much less distance to travel, making this sensor quicker than the two-dimensional (2D) sensor described above. Despite its name, it still gives a 2D reconstruction of the events in the (R, ϕ) plane. The information on z is obtained by combining the different measurement points (or hits) in each layer. The use of these new technologies is chosen to withstand the heavy radiation encountered in the forward region and the vicinity of the beam.

The distribution of the pixel modules is done along the (R, ϕ) plane and the z-axis. The minimum pixel size is $250\,\mu\mathrm{m}$ thick and $50\,\mu\mathrm{m} \times 250\,\mu\mathrm{m}$ in the IBL and $50\,\mu\mathrm{m} \times 400\,\mu\mathrm{m}$ in the outer layers. The achieved intrinsic spatial resolution is $8\,\mu\mathrm{m}$ in the (R, ϕ) plane within the IBL, slightly degrading to $10\,\mu\mathrm{m}$ in the outer layers and the end-caps. The resolution along the z-axis is $40\,\mu\mathrm{m}$ in the IBL, degrading to $115\,\mu\mathrm{m}$ in the z direction for the outer layers and R direction for the end-cap. Finally, as seen in Figure 3.3(b), the Pixel Detector provides between three and four hits for the track reconstruction up to $|\eta| < 2.5$.

3.2.2 Semi-Conductor Tracker

The SCT is the tracking sub-system surrounding the Pixel Detector, covering a significantly larger area ($61\,\mathrm{m}^2$ compared to the $\sim 1.9\,\mathrm{m}^2$ of the Pixel Detector, including the IBL). Using the same pixel technology was not feasible for such large surfaces due to cost, and because the radiation environment at larger radii is less severe. Si was still used for the SCT but in the form of microstrips. Microstrips are a simpler technology than pixels. Instead of a grid of pixels, the sensors use parallel, heavily p-doped strips implanted on a lightly n-doped bulk, forming a p-in-n sensor as represented in Figure 3.4(c). The depletion region extends from the strips into the bulk, and charge carriers (holes) are collected at the p^+ implants. This technology is more vulnerable to type inversion after irradiation, but at the larger distance from the IP, the radiation levels are lower, so this remains manageable up to 2026. It does not require double-sided processing, making it less costly than pixels. The cooling is identical to pixels. Since strips provide only one-dimensional (1D) position measurements, a 2D reconstruction is obtained by combining two layers of strips rotated by a stereo angle of $40\,\mathrm{mrad}$. The SCT consists of four concentric barrel layers, each composed of double-layer Si microstrip modules, while the end-caps consist of nine disks per side.

It has 4088 two-sided modules, resulting in over six million readout strips (therefore six million channels). The barrel modules are rectangular with 768 strips ordered in a pitch of $80\,\mu\mathrm{m}$. The end-cap modules are trapezoidal with a strip pitch ranging from $56.9\,\mu\mathrm{m}$ to $94.2\,\mu\mathrm{m}$. The module thickness is $285 \pm 15\,\mu\mathrm{m}$. In the barrel, the strips run parallel to the beam (z-axis), while in the end-caps, they are oriented radially. The achieved resolution in the barrel is $17\,\mu\mathrm{m}$ in the (R, ϕ) plane and $580\,\mu\mathrm{m}$ along the z direction and in the R for the end-caps. Finally, as seen in Figure 3.3(b), the SCT provides between four and nine hits for the track reconstruction up to $|\eta| < 2.5$.

3.2.3 Transition Radiation Tracker

The last sub-system of the ID is the TRT [144, 145] occupying a volume of $12\,\mathrm{m}^3$ to cover a pseudorapidity range of $|\eta| < 2.0$. It acts both as a drift chamber and as a transition radiation (TR) detector. It is composed of $50\,000$ drift tubes, also called *straws*, in the barrel, each measuring $144\,\mathrm{cm}$ long and $250\,000$ straws in both end-caps measuring $37\,\mathrm{cm}$ long. Each straw has a diameter of $4\,\mathrm{mm}$ and forms the cathode operating at $1530\,\mathrm{V}$, while the anode inside is a $31\,\mu\mathrm{m}$ tungsten (W) wire coated with $0.5\,\mu\mathrm{m}$ to $0.7\,\mu\mathrm{m}$ of gold, creating an electric field. The straw is filled with a gas mixture of $70\,\%$ of xenon (Xe), $27\,\%$ of carbon dioxide (CO_2), and $3\,\%$ of dioxygen (O_2). When a particle passes through, it will ionise the gas mixture, thanks to the applied electric field. Electrons will move from the straws by electron avalanche to the central anode wire in less than $48\,\mathrm{ns}$, producing a signal. The position of the hit is then determined using the drift time of the electrons to reach the wire. The CO_2 helps lower the electrons' drift time.

In addition to these methods of measurement, TR is also used, giving its name to this sub-system. All the straws are placed in a medium called a *radiator* composed of polypropylene fibres in the barrel and foils in the end-caps. When a charged ultra-relativistic particle crosses the boundary between materials with two distinct electric permittivities ε, it has a chance to emit TR consisting of photons in the X-ray range. They will be absorbed by the photoelectric effect by Xe atoms, producing electrons that the anode will collect. The O_2 helps by absorbing ultraviolet (UV) photons created in an avalanche to prevent secondary ones. The intensity of the X-ray photons produced depends on the Lorentz factor $\gamma = \frac{E}{m}$. Since electrons reach much higher γ than pions at the same energy, the TRT can distinguish between them in combination with the calorimeter measurement on the energy. This allows effective discrimination between electrons and pions that have a similar range of energy but very different masses: $0.511\,\mathrm{MeV}$ for the electron and $139.57\,\mathrm{MeV}$ for $\pi^{\pm}$ [34]. A scheme showing the two techniques used by the TRT is shown in Figure 3.5.

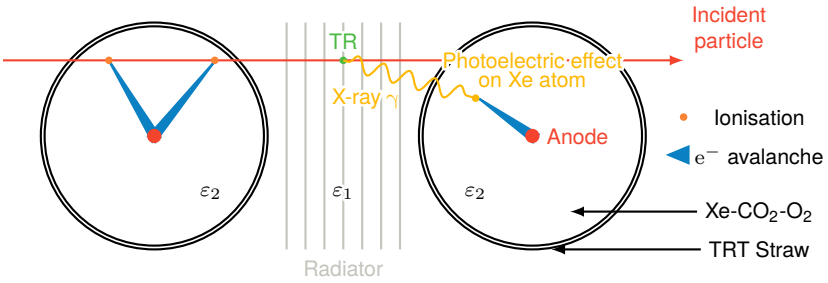


Figure 3.5: Scheme of the TRT.

Due to leaks of the gas mixture during Run-1, replacing part of the Xe with argon (Ar) was considered to reduce the cost. It was shown that the TR-based particle identification performance was significantly reduced, but the tracking properties were preserved [146]. Since the number of leaks was expected to grow with increasing LHC activity, the Xe was slowly replaced during Run-2 and the LS-2. Now the barrel is exclusively made of Ar with an additional end-cap wheel for Run-3.

In the barrel, the straws are oriented in 73 layers parallel to the z-axis, while in the end-caps, they are arranged in 160 radial layers. The achieved spatial resolution in the (R, ϕ) plane is $130\,\mu\mathrm{m}$ in both the barrel and end-cap regions. It is not possible to have information on the z-axis (therefore η) with this sub-system. Finally, the TRT provides a great number of hits, typically around 36, but around 22 in the transition region between $0.8 < |\eta| < 1.0$.

3.3 Calorimeters

Right after the ID and the detection of charged particles lie the calorimeters [147] responsible for the energy measurement, extending the identification to neutral particles and disentangling ambiguities between EM and hadronic (HAD) ones. Muons and neutrinos interact with matter very weakly: they are not targeted by this sub-detector. The neutrinos are associated with missing transverse momentum (MET), while muons benefit from the MS.

The measurement of the energy by a calorimeter is a destructive process. Highly dense materials are used to initiate a cascade, also called a *shower*, generating secondary particles with progressively lower energy that are fully absorbed by the calorimeter. The deposited energy is converted into a measurable quantity, and by summing across the whole shower, it is possible to determine the energy of the initial particle.

Two major techniques of calorimetry exist. The first one is the homogeneous calorimeter, consisting of a single material that both absorbs and generates the signal. The second one is a sampling calorimeter, which consists of alternating layers of absorber and an active layer to collect the deposited energy and produce a signal. The advantage of the first technique is that the entire energy deposited is collected: there is no energy loss, unlike in a sampling calorimeter, where the energy deposited in the absorber is not measured. This results in a better energy resolution for a homogeneous calorimeter.

An example of such a calorimeter is the EM calorimeter of CMS [148] that uses lead tungstate, which combines high density with scintillating properties in a compact medium. The drawback of this technology is its cost for producing a large number of these crystals, but also the degradation when facing heavy radiation.

The ATLAS calorimeter [149] thus uses a sampling calorimeter through three different sub-systems consisting of the EM Calorimeter measuring the energy of electrons and photons, the HAD Calorimeter focusing on hadrons, and lastly the Forward Calorimeter (FCal) to measure the high-η particles, either EM or HAD. The design of each of these calorimeters is made to contain the showers generated by incoming particles. The constituents, the size, and the geometry have to be adapted depending on the type of particles and their interaction with the detector. A representation of each calorimeter and its different technologies is shown in Figure 3.6(a) and will be explained below.

3.3.1 Electromagnetic Calorimeter

At high energy, such as the one occurring in the pp collisions, the electron and photon interact with the detector almost exclusively via bremsstrahlung and pair production, respectively. They initiate an EM shower when encountering the absorber, creating secondary particles that will undergo the same process. The number of these particles and the size of the shower produced will increase, developing the shower, until the average energy of secondary particles falls below a certain threshold called the *critical energy*, E_c. Below this energy, the main interaction process with matter comes from collision with the atoms and molecules of the medium. Electrons lose energy via ionisation and thermal excitation, while photons undergo the photoelectric effect and Compton scattering.

The main parameter to describe an EM shower is the radiation length X_0 that depends on the characteristics of the material. It can be written as [150]:

$$X_0 = \frac{1}{4\alpha r_\mathrm{e}^2 N_\mathrm{A}} \cdot \frac{A}{Z^2(L_\mathrm{rad} - f(Z)) + ZL'_\mathrm{rad}}\,, \tag{3.6}$$

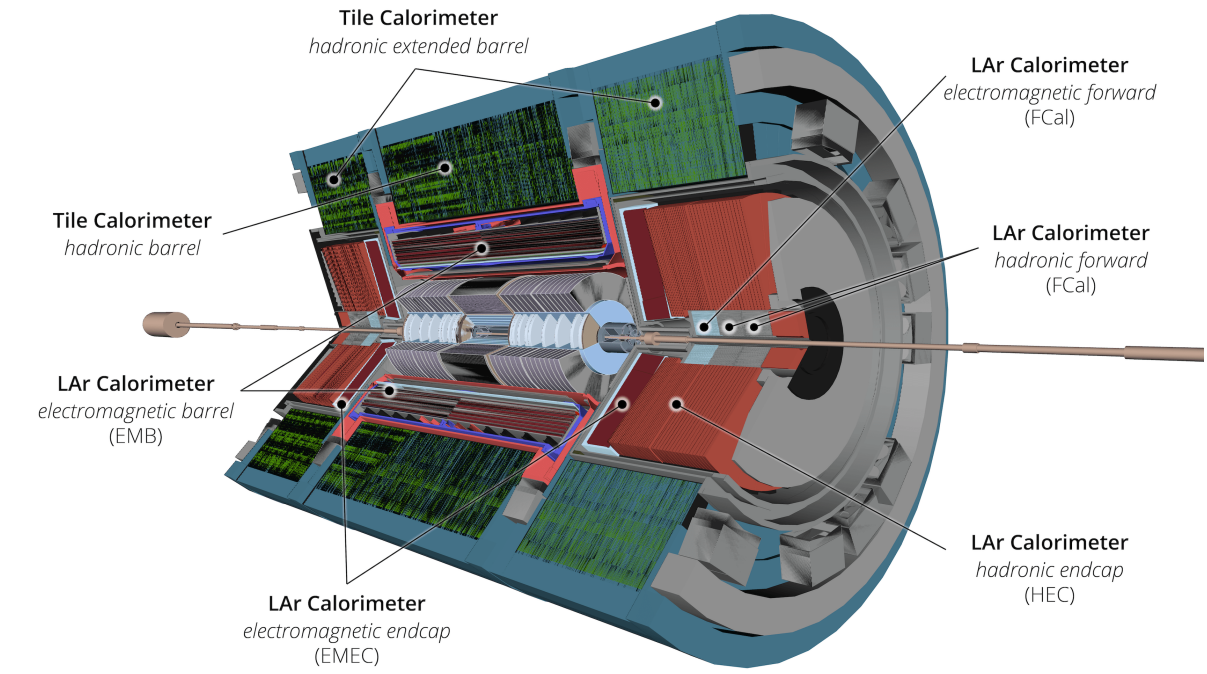
where the parameters are the following:

- α the fine-structure constant
- the classical electron radius defined as $r_\mathrm{e} = \frac{e^2}{4\pi\varepsilon_0 m_\mathrm{e} c^2}$ with e the elementary charge, ε_0 the vacuum permittivity, and m_e the mass of the electron
- N_A the Avogadro number
- A the atomic weight in $\mathrm{g \cdot mol^{-1}}$ and Z the atomic number
- the factor $\frac{1}{4\alpha r_\mathrm{e}^2 N_\mathrm{A}}$ is equivalent to $716.408\,\mathrm{g \cdot cm^{-2}}$ [34]
- L_rad and L'_rad are given in Table 3.A
- $f(Z)$ is an infinite sum, but can be represented for elements up to uranium by:

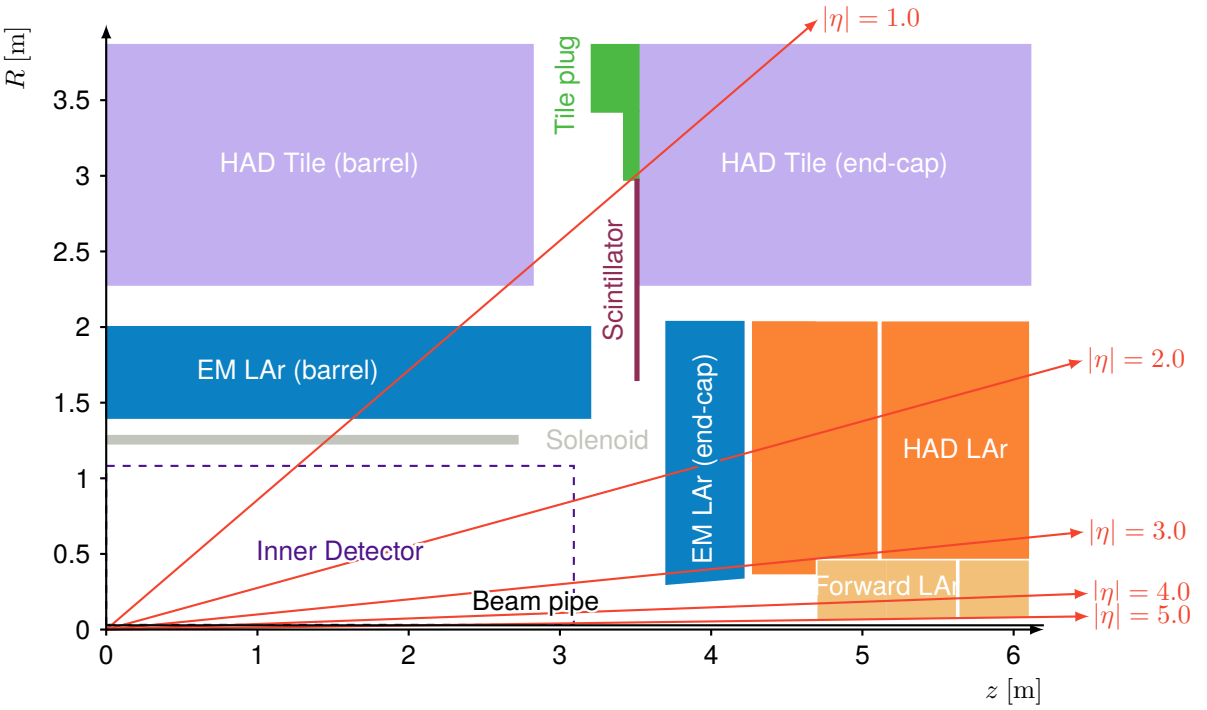
$$f(Z) = (\alpha Z)^2 \left((1 + (\alpha Z)^2)^{-1} + 0.20206 - 0.0369(\alpha Z)^2 + 0.0083(\alpha Z)^4 - 0.002(\alpha Z)^6\right)\,. \tag{3.7}$$

Table 3.A: L_rad and L'_rad used for the radiation length in an element.

Element	Z	L_rad	L'_rad
H	1	5.31	6.144
He	2	4.79	5.621
Li	3	4.74	5.805
Be	4	4.71	5.924
Others	> 4	$\ln\left(184.15Z^{-1/3}\right)$	$\ln\left(1194Z^{-2/3}\right)$



(a)



(b)

Figure 3.6: Calorimeters in 3D [138] (a) and their layout (b).

Table 3.B: Parameters of matter interaction for different elements [34].

Element	Z	E_c [MeV]	X_0 [cm]	R_M [cm]	λ_{int} [cm]
LAr	18	32.84	14.00	9.043	85.77
Fe	26	21.68	1.757	1.719	16.77
Cu	29	19.42	1.436	1.568	15.32
W	74	7.97	0.3504	0.9327	9.946
Pb	82	7.43	0.5612	1.602	17.59

The radiation length represents the average distance x that an electron has to travel through a material to reduce its energy to e^{-1} of its original energy E_0 (that is, by a factor e) via bremsstrahlung, that is, $\langle E(x)\rangle = E_0\, e^{-x/X_0}$. Similarly for a photon beam of initial intensity I_0, the average distance is given by $x = \frac{9}{7}X_0$, that is, $\langle I(x)\rangle = I_0\, e^{-7x/9X_0}$. This shows that the energy range over which a shower develops is similar for electrons and photons. When expressed in terms of X_0, we can parametrise the EM shower regardless of the material used.

In the literature, there exist two definitions regarding the critical energy. The first one says that it corresponds to the energy where losses by ionisation and bremsstrahlung are equal, giving us:

$$\left.\frac{dE}{dx}\right|_{\text{ionisation}} = \left.\frac{dE}{dx}\right|_{\text{bremsstrahlung}} . \tag{3.8}$$

The second one, given in [151], defines it as the energy at which the ionisation loss per X_0 equals the electron energy, giving us Equation 3.9. That is similar to the first definition under the approximation of the bremsstrahlung losses:

$$\left.\frac{dE}{dx}\right|_{\text{ionisation}} = \frac{E}{X_0} \simeq \left.\frac{dE}{dx}\right|_{\text{bremsstrahlung}} . \tag{3.9}$$

In the context of the second version, we can write E_c for solids (gases) as:

$$E_c = \frac{610\,(710)\,\text{MeV}}{Z + 1.24\,(0.92)} . \tag{3.10}$$

This definition has been found to describe transverse EM shower development more accurately.

Now that we have described how electrons and photons interact with matter, we can focus on the geometry that characterises the EM shower. The maximum of the shower t_{max} represents the depth where the largest number of secondary particles is produced. It is given by:

$$t_{max} \simeq \ln\frac{E_0}{E_c} + \begin{cases} -0.5 & \text{for an e-induced shower} \\ +0.5 & \text{for a } \gamma\text{-induced shower} \end{cases} . \tag{3.11}$$

At this depth, the shower particles have reached the critical energy that stops the showering process, and the remaining secondary particles are absorbed. The depth t_{max} is expressed in units of the radiation length X_0.

The calorimeter thickness required to contain $95\,\%$ of the shower energy using units of X_0 is approximated by:

$$t_{95\,\%} \simeq t_{max} + 0.08Z + 9.6 . \tag{3.12}$$

Meanwhile, the transverse size of the EM shower is described using the Molière radius as shown in [152]:

$$R_M = \frac{21\,\text{MeV}\cdot X_0}{E_c} . \tag{3.13}$$

It represents the dispersion of the electron's scattering from the shower axis after travelling a distance equal to X_0. It does not depend on the initial energy of the particle, contrary to the depth formula, but only on the material. Some $95\,\%$ of the shower is contained within $2R_M$.

The ATLAS EM calorimeter [153, 154] is a cylindrical detector spread in two parts: the ElectroMagnetic Barrel (EMB) section and the ElectroMagnetic End-cap (EMEC) section.

The barrel is divided into two half-barrels centred around the z-axis, covering the regions up to $|\eta| < 1.475$. Its inner radius is $1.4\,\text{m}$ while the outer radius is $2\,\text{m}$ for a total length of $3.2\,\text{m}$ for each half. The end-caps consist of a $51\,\text{cm}$ deep wheel delimited externally with a radius of $2.034\,\text{m}$ and internally by a cone with an aperture of $4.7°$ and with a smaller radius of $30.2\,\text{cm}$. They cover the regions

$1.375 < |\eta| < 3.2$. The transition areas in between these two parts are called *crack regions* and are located in $1.37 < |\eta| < 1.52$. This can be seen in Figure 3.6(b), which shows the EMB and EMEC calorimeter coverage.

The entire EM calorimeter is a sampling one using liquid argon (LAr) as the active medium and lead for the absorber. It has an accordion geometry to provide full coverage in ϕ without any cracks as seen in Figure 3.7(a), providing as well very uniform performance in terms of linearity and resolution as a function of ϕ. In the barrel, the accordion waves are axial along ϕ while in the end-caps, they are parallel to the radial direction running axially. The LAr gap of $4.2\,\mathrm{mm}$ is kept constant in the barrel but increases with radius in the end-caps. The absorbers have various thicknesses to limit the decrease of the sampling fractions as $|\eta|$ increases. In the barrel it is $1.53\,\mathrm{mm}$ for $|\eta| < 0.8$ and goes down to $1.13\,\mathrm{mm}$ for $|\eta| > 0.8$. In the end-caps the plates of lead are $1.7\,\mathrm{mm}$ for $|\eta| < 2.5$ and $2.2\,\mathrm{mm}$ for $|\eta| > 2.5$.

The ionisation charges produced by the shower are collected by copper (Cu) electrodes in the middle of the LAr gap for a total drift time of about $450\,\mathrm{ns}$. The readout electronics are located at each end of the calorimeter. The electrodes are positioned in the middle of the gap by honeycomb spacers.

Both the barrel and the end-caps are encapsulated inside their own cryostat to maintain the LAr at its functional temperature of $89\,\mathrm{K}$. The corresponding amount of matter before the calorimeter (ID, solenoid cryostat, but also mechanical structure) is equivalent to a minimum of 1.7 radiation length units and goes up to around 5. Regarding the barrel, it has at least $22\,X_0$ increasing from 22 to 30 between $|\eta| = 0$ and $|\eta| = 0.8$ and from 24 to 33 between $|\eta| = 0.8$ and $|\eta| = 1.3$. Meanwhile, the end-cap's total is greater than $24\,X_0$. The thickness increases from 24 to $38\,X_0$ as $|\eta|$ increases from 1.475 to 2.5 (outer wheel) and from 26 to $36\,X_0$ as $|\eta|$ increases from 2.5 to 3.2 (inner wheel).

The parameters in Table 3.B show the key parameters discussed before for the elements used inside the EM calorimeter. For an incident electron with an initial energy of $100\,\mathrm{GeV}$, the maximum of the shower will be reached in $9\,X_0$ ($5.1\,\mathrm{cm}$) of lead, while $95\,\%$ of the longitudinal shower is contained within $25\,X_0$ ($14.1\,\mathrm{cm}$). The shower will also spread transversely, given its Molière radius, and $95\,\%$ of the area will be contained within $2R_\mathrm{M} = 3.2\,\mathrm{cm}$. The structure of the EM calorimeter is thus well designed to fully contain a shower for electrons or photons in the GeV range as lead is a strong absorber.

The EM calorimeter is subdivided into $173\,312$ elementary cells ($109\,568$ in the barrel and $63\,744$ in the end-cap) along η and ϕ spread over four distinct layers, each having a specific purpose:

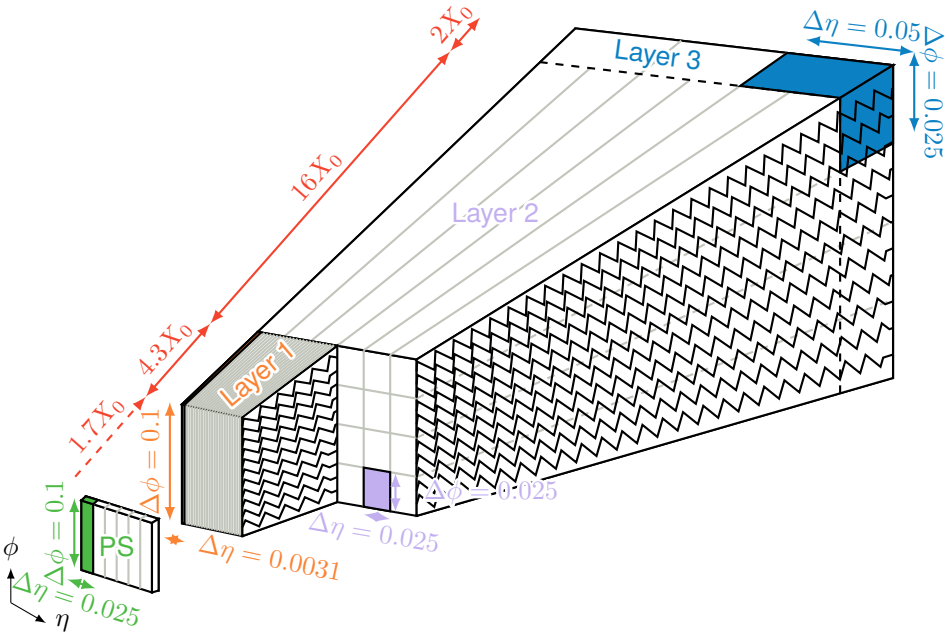
- the innermost one is the pre-sampler (PS), a very thin $11\,\mathrm{mm}$ LAr layer placed distinctly from the rest of the calorimeter, just behind the solenoid magnet and in front of the first accordion layer within the cryostat, to correct for the energy losses in the $1.7\,X_0$ material before the calorimeter
- the Layer 1 (L1) or strips have the finest granularity in $\Delta\eta$ to distinguish between photon and π^0 decay
- the Layer 2 (L2) is the largest one and contains most of the energy deposit
- the Layer 3 (L3) is designed to estimate the tail energy and the leaks into the following HAD calorimeter.

A representation of these elementary cells for the barrel is shown in Figure 3.7(a) while their respective granularity for both barrel and end-caps can be seen in Table 3.C. Until Run-2, the elementary cells were grouped to form a Trigger Tower. They were composed in terms of $n_\eta \times n_\phi$ elementary cells corresponding to 4×1 in the pre-sampler, 32×1 in Layer 1, 4×4 in Layer 2, and 2×4 for the last layer. This corresponds to a total of 60 cells, and the deposited energy was summed across all layers, giving approximately a granularity in $\Delta\eta \times \Delta\phi$ of 0.1×0.1 for a single Trigger Tower. The LS-2 brought the ATLAS Phase-I Upgrade, preparing for the HL-LHC luminosity. The Trigger Towers were replaced by Super Cells [154, 155], by changing the electronic readout and stopping the sum across all the layers. Now the distribution is still 4×1 in the pre-sampler giving a granularity of 0.1×0.1, but for the Layer 1 and 2 it is lowered to 0.025×0.1 by grouping the elementary cells in 8×1 and 1×4, respectively. The last layer is grouped in 2×4, giving the same granularity as the pre-sampler. This change of electronics brings a better granularity in the two middle layers and splits the energy measurement across the four distinct layers.

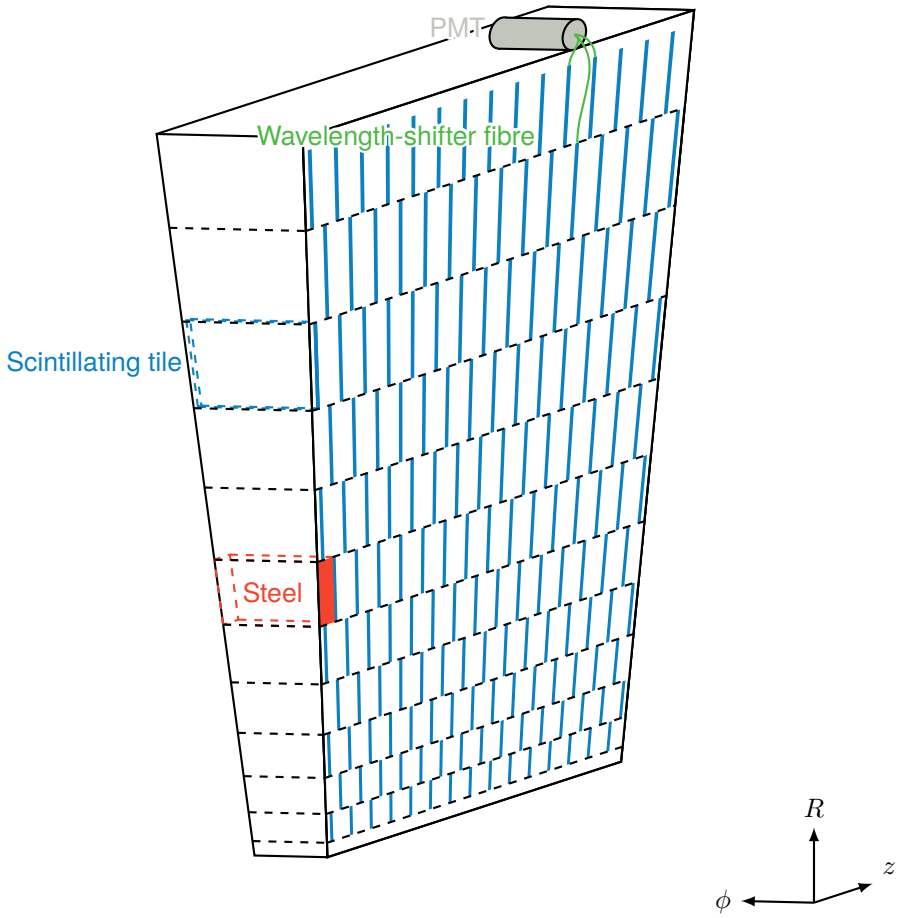
3.3.2 Hadronic Calorimeter

The principles of the HAD calorimeter and thus the cascade process are similar to the EM one, but more complex due to the wide range of secondary particles that have a variety of subsequent interactions with the nuclei of the absorber.

An inelastic collision of an incoming hadron and a nucleus initiates the shower. These collisions will generate a large number of secondary particles that will undergo further inelastic collisions until they fall below the pion (lightest hadrons, $139.6\,\mathrm{MeV}$ for $\pi^\pm$ and $135.0\,\mathrm{MeV}$ for π^0) production threshold. The particles produced go through a sequential decay before being absorbed:



(a)



(b)

Figure 3.7: Layout of the EM barrel calorimeter (a) and Tile Calorimeter (b).

Table 3.C: EM calorimeter configuration.

(a) Barrel configuration

Layer	$\lvert\eta\rvert$ range	Granularity $\Delta\eta \times \Delta\phi$
PS	< 1.52	0.025×0.1
Layer 1	< 1.4	0.0031×0.1
	1.4 to 1.475	0.025×0.025
Layer 2	< 1.4	0.025×0.025
	1.4 to 1.475	0.075×0.025
Layer 3	< 1.35	0.05×0.025

(b) End-cap configuration

Layer	$\lvert\eta\rvert$ range	Granularity $\Delta\eta \times \Delta\phi$
PS	1.5 to 1.8	0.025×0.1
Layer 1	1.375 to 1.425	0.05×0.1
	1.425 to 1.5	0.025×0.1
	1.5 to 1.8	0.0031×0.1
	1.8 to 2.0	0.0041×0.1
	2.0 to 2.4	0.00625×0.1
	2.4 to 2.5	0.025×0.1
	2.5 to 3.2	0.1×0.1
Layer 2	1.375 to 1.425	0.05×0.025
	1.425 to 2.5	0.025×0.025
	2.5 to 3.2	0.1×0.1
Layer 3	1.5 to 2.5	0.05×0.025

- $\pi^0 \to \gamma\gamma \to$ EM shower
- fission fragments $\to \beta$-decay, γ decay
- nuclear spallation: individual nucleons knocked out of nucleus, de-excitation
- neutron capture $\to$ nucleus* $\to$ fission.

A part of the shower is therefore composed of particles interacting electromagnetically. The fraction of it, called f_{em}, is given by [156]:

$$f_{\text{em}} = 1 - \left(\frac{E}{E_0}\right)^{m-1}, \tag{3.14}$$

with E the incident energy of the hadrons, E_0 a scale energy corresponding to a cutoff where there is no further hadronic production, and m a constant ranging from 0.80 to 0.85.

The other fraction of the shower is then hadronic. The fraction is simply given by:

$$f_{\text{had}} = 1 - f_{\text{em}}\,. \tag{3.15}$$

The hadronic fraction is of order 0.5 (0.3) for a 100 (1000) GeV shower. The EM fraction is the largest part of the total shower [147].

The hadronic part of the cascade will deposit energy via several mechanisms:

- ionisation, between $40\,\%$ and $60\,\%$, decreasing with increasing A, the largest contribution ($70\,\%$) coming from spallation protons
- generation of neutrons, between $10\,\%$ and $15\,\%$, decreasing with decreasing A
- generation of photons from fission, accounting for a few percent ($\sim 3\,\%$)
- nuclear break-up and recoil of nuclear fragments and target, accounting for between $30\,\%$ and $40\,\%$, decreasing with decreasing A.

The last point is an important feature of a hadronic shower. Due to the nuclear binding energy, some of the energy does not contribute to the detector signal: this is referred to as *invisible energy*. This energy, released during nuclear reactions, accounts for up to $40\,\%$ of the total hadronic energy of the shower, as mentioned before.

Consequently, the signal produced in an HAD calorimeter is smaller for an incoming hadron than for an electron or photon at the same energy scale. By defining the EM signal as S_{e} and the hadronic signal as S_{h}, the ratio $\frac{S_{\text{e}}}{S_{\text{h}}}$ is typically greater than one. The objective of compensation is to bring this ratio closer to unity.

Several techniques exist to improve compensation, including:

- using high-Z absorbers, which increase invisible energy losses but also produce more secondary neutrons. In hydrogen-rich active media, these neutrons can be detected through recoil protons
- carefully adjusting the absorber/active medium ratio to balance the EM and HAD responses

- applying software compensation, for example, by segmenting the calorimeter longitudinally and assigning smaller weights to regions with a strong EM component
- exploiting the different time profiles of EM and HAD showers (with delayed neutrons), where precision timing layers can help discriminate and reweight the signals
- employing dual-readout techniques, in which Čerenkov and scintillation light are measured simultaneously to determine the EM fraction event by event [157].

In any of these cases, a precise knowledge of the geometry of the hadronic shower development is needed. Due to the various complications mentioned earlier, the use of Monte Carlo (MC) programs to simulate the hadronic cascade development is needed.

The equivalent of the radiation length X_0 is the nuclear interaction length defined in [158]:

$$\lambda_{\text{int}} = \frac{A}{\rho N_{\text{A}} \cdot \sigma_{nA(\text{inelastic})}}\,, \tag{3.16}$$

where A, N_{A}, ρ are the atomic weight, the Avogadro number, and the density of the material (in $\text{g} \cdot \text{cm}^{-3}$), respectively. The inelastic cross-section (in cm^2) on the nucleus of atomic weight A is given by $\sigma_{nA(\text{inelastic})}$.

In Table 3.B, the nuclear interaction length is given for different elements. It is very much larger than the radiation length X_0. For lead, the λ_{int} is thirty times larger than X_0.

The maximum of the shower expressed in terms of nuclear interaction length is given by:

$$t_{\text{max}} \simeq 0.2 \ln \frac{E}{1\,\text{GeV}} + 0.7\,. \tag{3.17}$$

The calorimeter thickness required to contain $95\,\%$ of the shower energy is approximated in [158]:

$$t_{95\,\%} \simeq t_{\text{max}} + 2.5\lambda_a\,, \tag{3.18}$$

using units of λ_{int} where λ_a describes the exponential decay of the cascade beyond t_{max} and varies with the energy as $E^{0.13}$ for an energy E in GeV.

Regarding the transverse size of the hadronic shower, $95\,\%$ of it is contained within a size of approximately λ_{int}.

Following the heavy constraint inherent in the hadron shower, the ATLAS HAD calorimeter consists of two parts: the Tile Calorimeter and the Hadronic End-cap Calorimeter (HEC), each designed to optimise the measurement of hadronic showers in their respective pseudorapidity regions, ensuring full coverage and precise energy reconstruction across the detector.

3.3.2.1 Tile Calorimeter

The Tile Calorimeter [159] is the central HAD calorimeter of ATLAS, providing crucial measurements of jet energies, MET, and hadronic τ decays. It is a barrel, $5.8\,\text{m}$ long, covering the region $|\eta| < 1.0$, and is completed with two extended barrels, $2.6\,\text{m}$ long, increasing the coverage to the $0.8 < |\eta| < 1.7$ region, as seen in Figure 3.6(b). Its inner radius is $2.28\,\text{m}$ and its outer radius $4.25\,\text{m}$.

The barrel and extended barrels are divided azimuthally into 64 modules and segmented in depth into three layers, approximately 1.5, 4.1, and $1.8\,\lambda_{\text{int}}$ thick for the barrel, and 1.5, 2.6, and $3.3\,\lambda_{\text{int}}$ for the extended barrel. The total detector thickness is $9.7\,\lambda_{\text{int}}$ at $|\eta| = 0$.

It is a sampling calorimeter using steel as the absorber and scintillating tiles as the active medium.

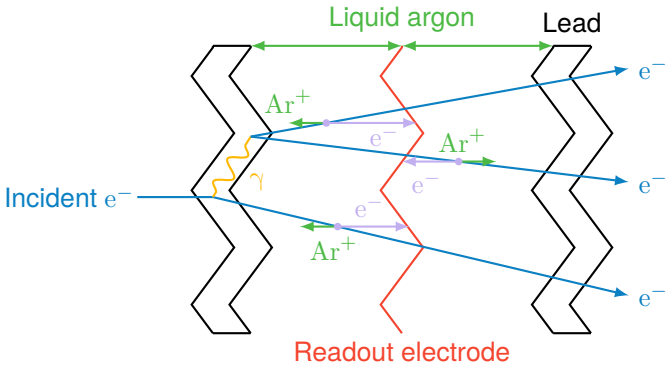
Each scintillating tile is made of polystyrene doped with fluor. They are $3\,\text{mm}$ thick, with lengths ranging from $97\,\text{mm}$ to $187\,\text{mm}$ radially and $200\,\text{mm}$ to $400\,\text{mm}$ azimuthally. When an incident particle ionises the medium, it produces UV scintillation light, which is converted to visible light by the wavelength-shifting fluor. Each side of the scintillating tiles is read out by optical fibres into PhotoMultiplier Tubes (PMTs) located on the outer edge of the detector. This provides an approximately projective geometry in pseudorapidity, giving a granularity of 0.1×0.1 in $\Delta\eta \times \Delta\phi$, slightly degrading to 0.2×0.1 in the last layer. The total number of scintillating tiles is over $460\,000$, providing 9852 channels. This segmentation provides sufficient resolution for jet reconstruction and MET measurements.

A representation of a Tile module is given in Figure 3.7(b) while a scheme showing how the energy is collected is represented in Figure 3.8(b).

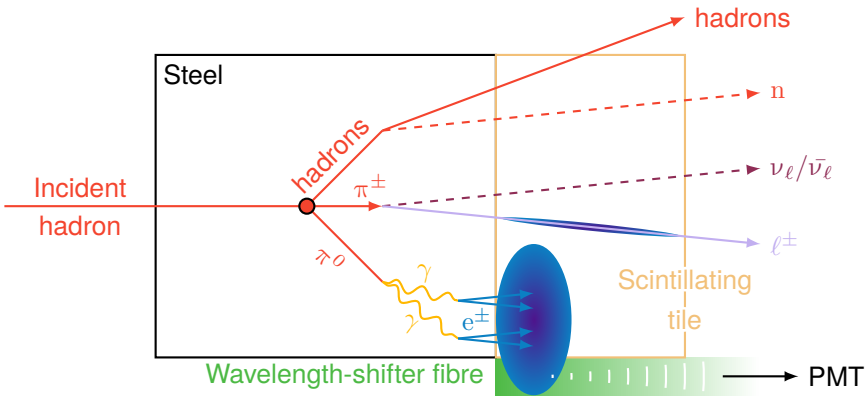
In the gap region between the barrel and the end-cap lies the Intermediate Tile Calorimeter (ITC). It is designed to maximise the volume of active material in this transition region while leaving room for the services and cables. It consists of the plug calorimeter in the region $0.8 < |\eta| < 1.0$, and, due to severe space constraints, only scintillator is present between $1.0 < |\eta| < 1.6$. In $1.0 < |\eta| < 1.2$ these are called *gap scintillators*, while in $1.2 < |\eta| < 1.6$ they are referred to as *crack scintillators*. The plug and

gap scintillators provide hadronic shower sampling, while the crack scintillator plays an important role in sampling EM showers in a region where the normal sampling is compromised by the dead material of the cryostat walls and ID cables.

The steel used in the Tile Calorimeter is a soft, low-carbon steel of the type Fe360 [159]. It is mainly composed of iron while the various other constituents make up less than $2\,\%$. It is safe to approximate it as iron for the following calculations. For a hadronic shower initiated by a $100\,\mathrm{GeV}$ hadron, the maximum is reached at $1.6\,\lambda_{\mathrm{int}}$ ($27.2\,\mathrm{cm}$) while $95\,\%$ of the longitudinal shower is contained within $6.2\,\lambda_{\mathrm{int}}$ ($104.0\,\mathrm{cm}$). Of the transverse size, $95\,\%$ is contained within $1\,\lambda_{\mathrm{int}}$ ($16.77\,\mathrm{cm}$). The structure of the Tile Calorimeter is well designed to fully contain hadronic showers in the GeV range. Steel is an effective and inexpensive absorber that also helps with partial compensation: its relatively high Z produces secondary neutrons, some of which are detected through interactions with the hydrogen-rich scintillating tiles.



(a)



(b)

Figure 3.8: Scheme showing the measurement of the energy of incoming electrons or photons (a) and hadrons (b).

3.3.2.2 Hadronic End-cap Calorimeter

The HEC consists of two independent wheels per end-cap, which share the same cryostat directly behind the EMEC. They cover the region $1.5 < |\eta| < 3.2$. The wheels are cylindrical, with thicknesses of $81.65\,\mathrm{cm}$ and $96.10\,\mathrm{cm}$ for the first and second wheel, respectively. They both share an outer radius of $2.03\,\mathrm{m}$. To reduce the drop in material density at the transition between the end-cap and the FCal (around $|\eta| = 3.1$), the HEC extends out to $|\eta| = 3.2$, overlapping with the FCal.

The technology used is similar to the EM calorimeter, using Cu as absorber and LAr for the active medium, forming a sampling calorimeter. The front wheels are made of 24 Cu plates, $25\,\mathrm{mm}$ thick, plus a $12.5\,\mathrm{mm}$-thick front plate, while the back wheels have 16 Cu plates, $50\,\mathrm{mm}$ thick, and a $25\,\mathrm{mm}$-thick front plate. The gap between the plates is always $8.5\,\mathrm{mm}$. The inner radius is thus $37.2\,\mathrm{cm}$ for the first nine plates and $47.5\,\mathrm{cm}$ for the remaining plates. Three electrodes collect the driving electrons, forming four separate LAr drift zones of $1.8\,\mathrm{mm}$ width. The space is maintained using a honeycomb sheet.

The 5632 readout channels are defined by pads etched on the central foil in each gap. The arrangement of these pads provides a semi-pointing geometry. The granularity is 0.1×0.1 in $\Delta\eta \times \Delta\phi$ in the region $|\eta| < 2.5$, increasing to 0.2×0.2 above.

The first wheel has a nuclear interaction length of $4.2\,\lambda_{\text{int}}$ while the second has $5.5\,\lambda_{\text{int}}$. In terms of radiation length, the HEC amounts to $\sim 103\,X_0$. Based on the calculations of a hadronic shower initiated by a $100\,\text{GeV}$ hadron detailed in Section 3.3.2.1, the HEC is well designed to contain such a shower.

An important feature of the HEC is its ability to detect muons and measure radiative energy loss.

3.3.3 Forward Calorimeter

The FCal [160] covers the region $3.1 < |\eta| < 4.9$, enabling energy measurement of particles at high pseudorapidity. It comprises both an EM (FCal-1) and two hadronic sampling calorimeters (FCal-2 and 3) using again LAr for the active medium. The absorber is Cu for the EM module, while tungsten is used for the hadronic ones to reduce the lateral spread of hadronic showers, R_{M} in Table 3.B. The modules are cylindrical wheels $444\,\text{mm}$ thick with a coaxial trough through which the beam passes. Their outer radius is $449\,\text{mm}$.

They are located inside the same cryostat, resulting in a hermetic design minimising energy losses in cracks between them and limiting the background reaching the muon system (a plug of Cu alloy lies behind the last HAD calorimeter, increasing the shielding). They are located at $4.7\,\text{m}$ from the IP, and have an inner radius going from $72.3\,\text{mm}$ to $85.7\,\text{mm}$, therefore exposing them to high particle fluxes, dictating their design. Very small LAr gaps were needed to avoid the ion build-up problem while having the highest density possible. These small gaps also lead to a faster signal. This was achieved by stacking plates of Cu together for the FCal-1 and drilling $12\,260$ holes inside to place the electrode (1008 readout channels). These electrodes are coaxial Cu rods and tubes separated by a radiation-hard plastic fibre spiralling around the rod for a width of $0.269\,\text{mm}$. The rest is filled with LAr. The FCal-2 and 3 are optimised for a high absorption length by maximising the amount of tungsten in the modules. They are made of two Cu end-plates spanned with the same electrode structure but using tungsten for the rods. A representation of these electrodes is given in Figure 3.9. The number of rods is reduced to $10\,200$ and 8224 (500 and 254 readout channels) while increasing the LAr gap to $0.376\,\text{mm}$ and $0.508\,\text{mm}$, respectively, for the second and third module. The electron drift time, thanks to this design, is reduced to $61\,\text{ns}$.

The granularity as a function of the pseudorapidity and the radiation length, as well as the nuclear interaction length, is given in Table 3.D. Based on the calculations of a hadronic shower initiated by a $100\,\text{GeV}$ hadron detailed in Section 3.3.2.1, the FCal is well designed to contain such a shower. For the EM shower initiated by a $100\,\text{GeV}$ electron and for Cu the maximum of the shower will be reached after $8\,X_0$ ($11.6\,\text{cm}$), while $95\,\%$ of the longitudinal shower is contained within $20\,X_0$ ($28.7\,\text{cm}$). The shower will also spread transversely, given its Molière radius, and $95\,\%$ of the area will be contained within $2R_{\text{M}} = 3.1\,\text{cm}$. The structure of the FCal is thus well designed to fully contain a shower for forward electrons or photons in the GeV range.

Table 3.D: FCal configuration.

Layer	X_0	λ_{int}	$\lvert\eta\rvert$ range	Granularity $\Delta x \times \Delta y$ [cm]
FCal-1	27.6	2.66	3.15 to 4.30	3.0×2.6
			3.10 to 3.15	~ 4 times finer
			4.30 to 4.83	~ 4 times finer
FCal-2	91.3	3.68	3.24 to 4.50	3.3×4.2
			3.20 to 3.24	~ 4 times finer
			4.50 to 4.81	~ 4 times finer
FCal-3	89.2	3.60	3.32 to 4.60	5.4×4.7
			3.29 to 3.32	~ 4 times finer
			4.60 to 4.75	~ 4 times finer

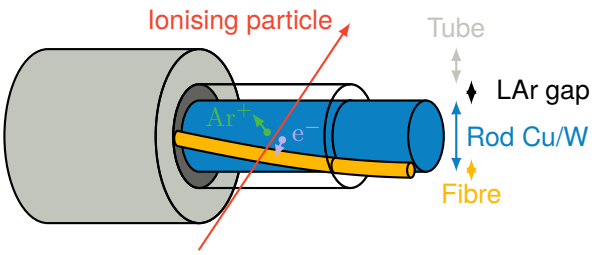


Figure 3.9: Scheme showing an electrode and the measurement of the energy.

3.4 Muon Spectrometer

The muon is significantly heavier than the electron ($m_\mathrm{e} = 0.511\,\mathrm{MeV}$ versus $m_\mu = 105.66\,\mathrm{MeV}$), which strongly suppresses energy loss via bremsstrahlung. Similarly, as explained in Section 3.3.1, the critical energy is defined as the threshold where the loss by bremsstrahlung is equal to ionisation. It is expressed for solids (gases) in [161]:

$$E_\mathrm{c} = \frac{5700\,(7980)\,\mathrm{GeV}}{(Z + 1.47\,(2.03))^{0.838\,(0.879)}}\,. \tag{3.19}$$

The order of magnitude for materials such as iron is in the hundreds of GeV. Therefore, they do not initiate a shower below this threshold. Meanwhile, the main loss of energy above this threshold is heavily reduced. The muons are referred to as minimum ionising particles (MIPs): their energy loss is close to the minimum. As stated in [161] for an incident $1\,\mathrm{TeV}$ muon on $3\,\mathrm{m}$ of iron the most probable loss is only $8\,\mathrm{GeV}$. On the other hand, as they are leptons, they do not interact strongly with the detector material.

For this reason, the muons can easily pass through the ID and the Calorimeter. That is why at the outermost part of the ATLAS experiment lies the MS [162], responsible for providing both tracking and momentum measurement. It is divided into small (S) and large (L) sectors, alternating around the circumference to ensure complete azimuthal coverage. Each sector is built from overlapping chambers, each using different techniques to serve various purposes. The different sub-systems can be seen in Figure 3.10(a), while the large and small sectors can be seen in Figure 3.10(b).

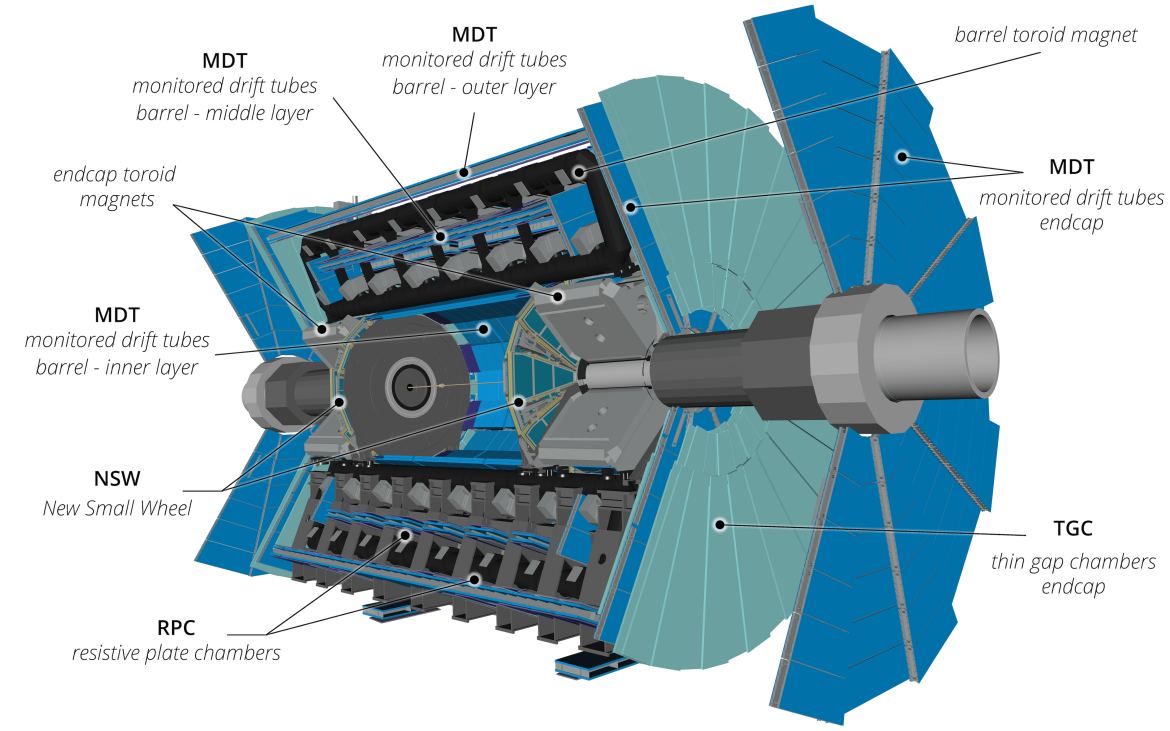
3.4.1 Monitored Drift Tube Chamber

The main component of the MS is the Monitored Drift Tube (MDT). It is made of a drift tube with a diameter of $29.97\,\mathrm{mm}$ and filled with $93\,\%$ of Ar and $7\,\%$ of CO_2. A wire of $50\,\mu\mathrm{m}$ made of gold-plated tungsten and rhenium ($97\,\%$ and $3\,\%$, respectively) is located in its centre to collect the electrons produced by ionisation. The advantage of using this cylindrical geometry is the accuracy. It depends weakly on the angle of incidence of the track onto the chamber plane, as the coordinate of the track is determined by the radius of the circle around the wire to which the track is tangential. The coordinates of the hits are then obtained in the bending plane (R, z) with a precision of $83\,\mu\mathrm{m}$. The drawback is the drift time that can go up to $700\,\mathrm{ns}$. Their length varies from $0.8\,\mathrm{m}$ to $6.5\,\mathrm{m}$. A scheme showing a tube and how it identifies muons is given in Figure 3.11(a).

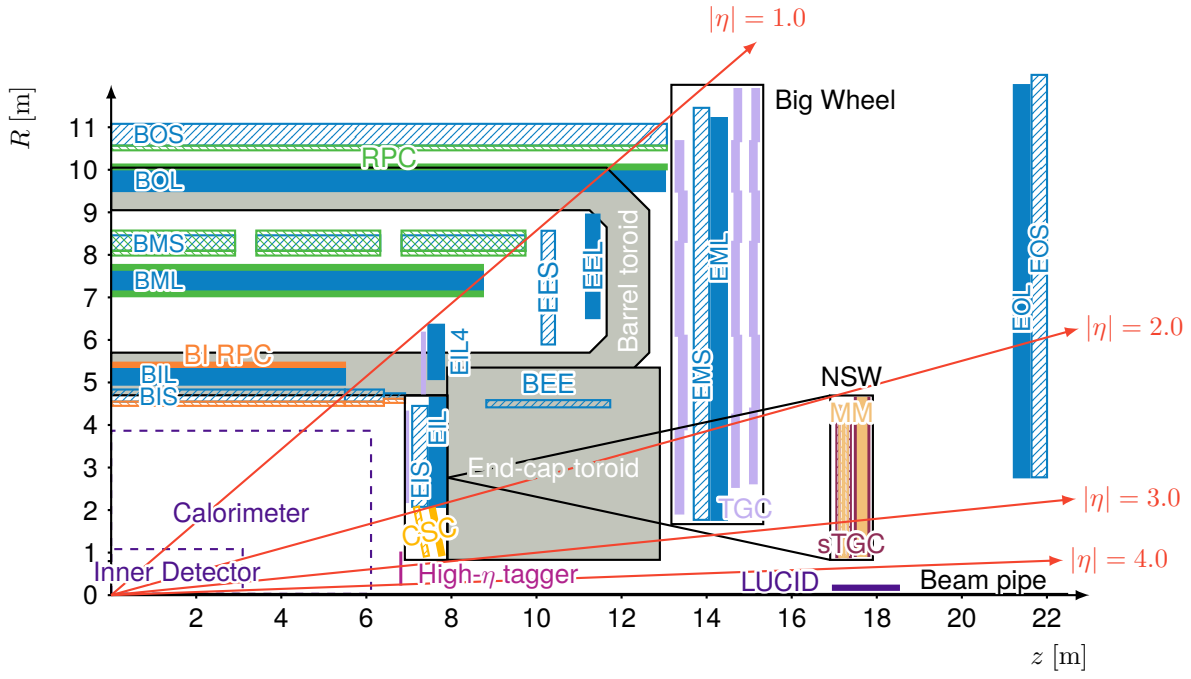
These chambers are arranged either in concentric circles in the barrel (B) or in a wheel in the end-cap (E). We distinguish the inner, middle, and outer chamber layers (I, M, and O, respectively) that are at $4.8\,\mathrm{m}$, $7.7\,\mathrm{m}$, and $10.4\,\mathrm{m}$ for the barrel and $7.4\,\mathrm{m}$, $10.8\,\mathrm{m}$, $14.1\,\mathrm{m}$, and $21.7\,\mathrm{m}$ for the end-cap wheels. This results in a coverage of $|\eta| < 2.7$.

The Barrel End-cap Extra (BEE) and the eighth Barrel Inner Small chamber (BIS.8) (the last small rectangle in Figure 3.10(b)) are special chambers located in the castellations of the end-cap toroid cryostats. They are used in the measurement of tracks passing from the barrel to the end-cap. These chambers consist of a single multi-layer of respectively four and three tube layers, while the rest consist of two groups of tube layers, called multi-layers, separated by a mechanical spacer in aluminium (Al). In the innermost layer of the muon detector, each multi-layer consists of four tube layers to enhance the pattern-recognition performance, while in the middle and outer layers, each multi-layer consists of three tube layers only. By combining the two multi-layers, the resolution goes down to $35\,\mu\mathrm{m}$ ($30\,\mu\mathrm{m}$) for three (four) tubes.

As visible in Figure 3.10(b), the inner end-cap wheel does not project fully into the end-cap outer wheel. To allow for momentum measurement in this region, an intermediate ring of chambers, the Extra End-cap Small (EES) and Extra End-cap Large (EEL) chambers, has been introduced with an offset of about $3\,\mathrm{m}$ to $3.6\,\mathrm{m}$ with respect to the corresponding inner end-cap wheel chambers.

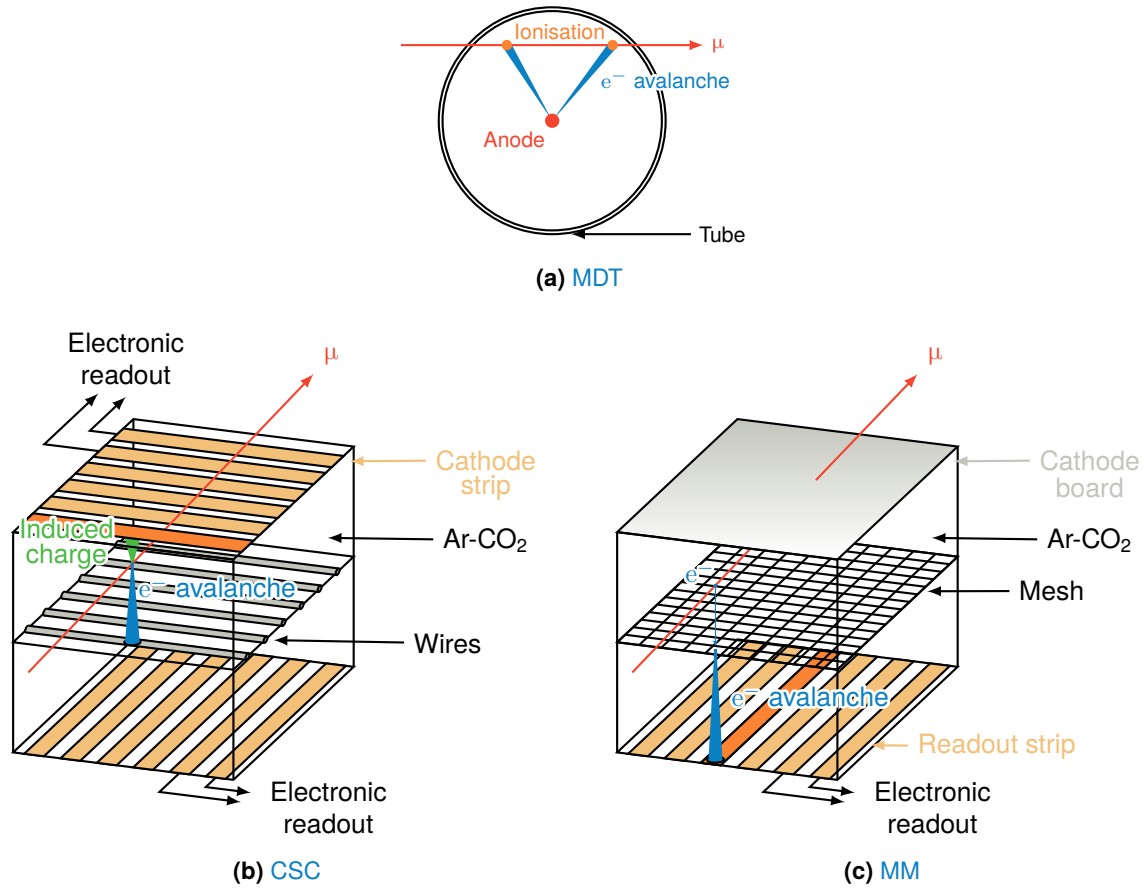


(a)



(b)

Figure 3.10: Muon Spectrometer in 3D [138] (a) and its layout (b). The hatched sub-systems are in the small sectors, while the filled ones are in the large sectors.



(a) MDT **(b)** CSC **(c)** MM

Figure 3.11: Scheme of different types of muon tracking sensors.

This arrangement of the various layers provides three hits for track reconstruction.

In total 1171 chambers are used for the spectrometer for a total of $354\,240$ tubes.

During the LS-2, the BIS.7 and BIS.8 were replaced by small-diameter Monitored Drift Tubes (sMDTs) [163] by merging them into the same module: the BIS78 to free space around it for future upgrades. They kept the same architecture of a bigger and smaller module. The term small refers to the change of the diameter of the tube going down to $15\,\mathrm{mm}$, reducing as well the drift time to $175\,\mathrm{ns}$. These tubes provide a coarser resolution of $106\,\mu\mathrm{m}$ but outperform the initial MDT under the high background rate expected in the latest LHC run.

The innermost wheel in the end-cap copes with more stringent conditions since most of the produced particles are found in the forward region, relatively close to the IP. The End-cap Inner Small (EIS) and End-cap Inner Large (EIL) cover only up to $|\eta| < 2.0$ and are replaced by the Cathode-Strip Chamber (CSC) above this.

3.4.2 Cathode-Strip Chamber

The CSC is located at $7.4\,\mathrm{m}$ from the IP. It consists of two disks of eight chambers (large and small), each containing four cathode-strip planes giving four independent measurements in η and ϕ for each track. It covers the region $2.0 < |\eta| < 2.7$ as visible in Figure 3.10(b).

The chambers used are multiwire proportional chambers[8] with the anode wires oriented in the radial direction with a diameter of $30\,\mu\mathrm{m}$. They are made of the same material as the wire in the MDT tube. Both cathodes, made of Cu, are segmented, one with the strips perpendicular to the wires (providing the precision coordinate) and the other parallel to the wires, providing the transverse coordinate. The position of the track is obtained by interpolation between the charges induced on neighbouring cathode strips. A representation of how the sensor works is shown in Figure 3.11(b).

In the bending plane, the strip widths are $1.519\,\mathrm{mm}$ and $1.602\,\mathrm{mm}$, respectively, for the large and small chambers. The inter-strip gap is $0.25\,\mathrm{mm}$ resulting in a readout pitch of $5.31\,\mathrm{mm}$ and $5.56\,\mathrm{mm}$, respectively. This allows a resolution of $60\,\mu\mathrm{m}$ per plane. The total number of wires is 250 and 402 for the small and large chambers for 192 channels.

In the non-bending direction, the cathode segmentation is coarser, $21\,\mathrm{mm}$ and $12.92\,\mathrm{mm}$ for large and small chambers, respectively. This leads to a resolution of $5\,\mathrm{mm}$. There are 48 strips per plane.

Thanks to this small design, the electron drift times are less than $40\,\mathrm{ns}$, resulting in a timing resolution of about $7\,\mathrm{ns}$ per plane.

The chamber is filled with a mixture of $80\,\%$ of Ar and $20\,\%$ of CO_2. This results in low neutron sensitivity because of the small gas volume and the absence of hydrogen in the gas mixture.

The CSC thus combines high spatial, time, and double track resolution with low neutron sensitivity in this demanding region of the MS.

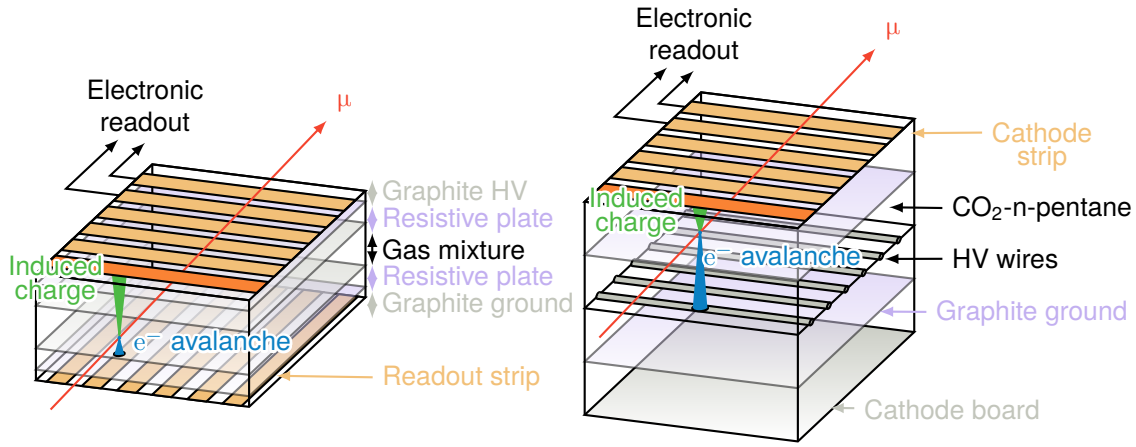
In combination with the precision tracking, the MS also uses triggering sub-systems providing fast information on tracks traversing the detector, allowing the recognition of their multiplicity and approximate energy range. They must also help with bunch-crossing identification, give fast and coarse tracking information, second coordinate measurement, and robustness towards random hits due to n/γ background.

3.4.3 Resistive Plate Chamber

The MDT chambers work in conjunction with the Resistive Plate Chamber (RPC) detectors in the barrel region. The RPC system consists of three concentric cylindrical layers around the beam axis. The first two RPC layers (the innermost being the first) are located around the Barrel Middle Large (BML) and Barrel Middle Small (BMS) chambers, while the third one is positioned on top of the Barrel Outer Large (BOL) and below the Barrel Outer Small (BOS). They cover the region $|\eta| < 1.05$.

The distance between the inner and outer RPC layers allows the trigger to select high-momentum muons in the range $9\,\mathrm{GeV}$ to $35\,\mathrm{GeV}$ (high-p_T trigger), while the two inner layers provide the low-p_T trigger in the range $6\,\mathrm{GeV}$ to $9\,\mathrm{GeV}$.

Each RPC is a gaseous parallel-plate detector consisting of two independent gas volumes of $2\,\mathrm{mm}$ height, where the avalanches develop. The correct gap spacing is maintained by insulating spacers. The induced signals are read out by two orthogonal sets of strips with a width of $25\,\mathrm{mm}$ to $35\,\mathrm{mm}$ and a thickness of about $17\,\mu\mathrm{m}$, with a time resolution of $1.5\,\mathrm{ns}$. A graphite layer at high voltage (HV) and a ground plate generate an electric field to produce an avalanche, which a resistive plate then collects. Figure 3.12(a) shows how these sensors work. The resolution in η is $3.7\,\mathrm{mm}$ to $7.3\,\mathrm{mm}$ and in ϕ is $3.7\,\mathrm{mm}$ to $7.3\,\mathrm{mm}$ for clusters with one strip or with two strips. In total, 606 chambers are used for $373\,000$ channels.



(a) RPC **(b)** TGC

Figure 3.12: Scheme of different types of muon triggering sensors.

The gas mixture used is $C_2H_2F_4$/iso-C_4H_{10}/SF_6 ($94.7\,\%/5\,\%/0.3\,\%$), which provides low operating voltage due to the low SF_6 concentration, is non-flammable and cost-effective, and ensures stable avalanche operation.

The two orthogonal readout layers of each RPC measure both η and ϕ. This redundancy in track reconstruction suppresses fake tracks from noise hits and significantly improves the trigger efficiency even in the presence of small chamber inefficiencies.

3.4.4 Thin Gap Chamber

Like the Resistive Plate Chamber, the Thin Gap Chamber (TGC) provides the trigger functionality, but in the end-cap region. It also measures the azimuthal coordinate, complementing the MDT measurement in the bending (radial) direction. A good time resolution (with an efficiency of at least $99\,\%$) is required to correctly identify the bunch crossing, along with fine granularity to ensure a sharp p_{T} threshold for the triggering muon. To achieve these requirements, the TGCs are designed as multiwire proportional chambers.

Each chamber consists of wire planes (anodes) with a gold-coated tungsten wire pitch of $1.8\,\mathrm{mm}$ and a wire diameter of $50\,\mu\mathrm{m}$, placed between Cu-coated cathode planes. The wire-to-cathode distance is $1.4\,\mathrm{mm}$. The bending (radial) coordinate is measured by the wire planes, while the azimuthal coordinate is obtained from radial strips on the cathode. The electric field produced by the HV wires and the graphite at ground produces electron avalanches, inducing a signal read by the cathode strip. This is represented in Figure 3.12(b). The strip pitch varies from $14.6\,\mathrm{mm}$ to $49.1\,\mathrm{mm}$. The chambers are built as either doublets or triplets (two or three wire planes), separated by a $20\,\mathrm{mm}$-thick honeycomb paper spacer for mechanical stiffness. Two of the Cu cathode layers in the triplets and doublets are segmented into readout strips to measure the azimuthal coordinate. The gas mixture used is $55\,\%$ CO_2 and $45\,\%$ n-pentane. The time resolution is $4\,\mathrm{ns}$ while the spatial resolution is from $2\,\mathrm{mm}$ to $6\,\mathrm{mm}$ in the radial direction and $3\,\mathrm{mm}$ to $7\,\mathrm{mm}$ in the azimuthal direction.

Each TGC forms part of a circular disk. On the Big Wheel, three layers of TGC are installed, one triplet and two doublets, corresponding to seven wire planes in total. This configuration provides coverage in the region $1.05 < |\eta| < 2.4$. The triplet layer helps to reject false coincidences from background hits, which are more frequent in the end-cap region than in the barrel. In the Small Wheel, only one doublet is installed, segmented radially into two non-overlapping regions in front of the EIS/EIL and EIL4 chambers, covering the region $1.05 < |\eta| < 1.92$.

In total, 3588 TGC chambers are installed, totalling about $318\,000$ channels.

The Small Wheel was replaced during LS-2, as the original system showed degraded performance due to increased background from higher luminosity. The upgraded New Small Wheel (NSW) [164] introduced two new detector technologies, improving precision tracking and triggering capabilities. Each of them is made from eight large and small sectors as before. One sector consists of two MicroMegas (MM) detectors sandwiched between two small-strip Thin Gap Chambers (sTGCs).

The chamber of the sTGC uses the same geometry, electric field, and gas mixture as the original TGC. The key difference is the strip pitch, now $3.2\,\mathrm{mm}$ ($2.7\,\mathrm{mm}$ strip width with a $0.5\,\mathrm{mm}$ gap), which is much smaller than the previous one, giving it its name. The signal is collected within $25\,\mathrm{ns}$ and the spatial resolution of $75\,\mu\mathrm{m}$ is much finer than that of the original TGC.

3.4.5 MicroMegas

The MM is the other component of the NSW responsible for tracking. It must provide good position resolution independently of the particle incident angle, high efficiency even under strong background rates, and good two-track separation.

The MM (also known as a micro-mesh gaseous structure) is a thin, wireless gaseous particle detector. It consists of a planar (drift) electrode, a gas gap of $5\,\mathrm{mm}$ filled with $93\,\%$ Ar and $7\,\%$ CO_2 acting as the conversion and drift region, and a thin metallic mesh with a pitch of 325 lines per inch (about 12.8 lines per mm), positioned $128\,\mu\mathrm{m}$ above the readout electrode, creating the amplification region. The entire assembly lies on top of a $0.5\,\mathrm{mm}$ printed circuit board (PCB).

Charged particles traversing the drift space ionise the gas, and the liberated electrons drift towards the mesh. With an electric field in the amplification region about 66 times stronger than that in the drift region, the mesh is transparent to more than $95\,\%$ of the electrons. The avalanche then develops in the thin amplification region, immediately above the Cu readout electrode, which is $17\,\mu\mathrm{m}$ thick, with a strip pitch of $0.425\,\mathrm{mm}$ to $0.445\,\mathrm{mm}$ and a strip width of about $300\,\mu\mathrm{m}$. Figure 3.11(c) shows its principle.

The electron drift in the conversion gap is relatively slow, on the order of tens of nanoseconds, whereas the amplification process occurs within a fraction of a nanosecond, producing a fast electron pulse on the readout strip. The ions created in the avalanche move in the opposite direction, back towards the amplification mesh. Due to their relatively low drift velocity, it takes them roughly $100\,\mathrm{ns}$ to reach the mesh. This rapid evacuation of positive ions makes the MM particularly well suited for operation at very high particle fluxes.

Sparks can occur when the total number of electrons in an avalanche reaches a few 10^7 (Raether limit [165]). Such ionisation levels can easily be reached by low-energy α-decays or slowly moving charged debris from neutron interactions in the detector. To mitigate this, a protection scheme was developed by

adding a layer of resistive strips ($20\,\mu\mathrm{m}$ thick) on top of a thin insulating layer ($64\,\mu\mathrm{m}$ thick), directly above the readout electrode.

Each sector in the MM contains two MM units, each composed of four detection layers.

The achieved position resolution in the (R, ϕ) plane is about $90\,\mu\mathrm{m}$. The very fine segmentation of the readout strips, together with the good time resolution, complements the sTGC-based trigger, adding robustness and redundancy.

3.5 Magnet System

A charged particle moving through a magnetic field will have its trajectory curved. The specific radius of this curvature depends on the particle momentum and the strength of the applied field. The transverse component of the momentum is given by:

$$p_{\mathrm{T}} = q \cdot R \cdot B\,, \tag{3.20}$$

where the magnetic field is B, the charge of the particle is q, and R is the bending radius.

The Magnet System of ATLAS [166] thus comprises two distinct magnets visible in Figure 3.13, each serving a different purpose and allowing measurement of the momentum given the curvature of the reconstructed tracks.

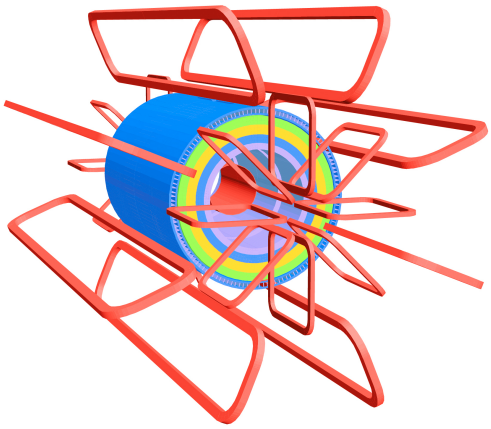


Figure 3.13: Scheme of the ATLAS magnets in red and the barrel calorimeter [137].

The first one and the innermost is a solenoid magnet [167] encompassing the whole ID (about $30\,\mathrm{m}^3$) and aligned with respect to the beam axis. It is a superconducting coil composed of aluminium, Cu, and niobium–titanium. Its cross-section is $30\,\mathrm{mm} \times 4.25\,\mathrm{mm}$, its total length $10\,\mathrm{km}$, and it carries a current of $7.73\,\mathrm{kA}$. It is cooled down to $4.5\,\mathrm{K}$ by placing it with the cryostat of the EM calorimeter. The intensity of the magnetic field produced is $2\,\mathrm{T}$. The solenoid itself has an outer radius of $1.28\,\mathrm{m}$ and an inner one of $1.23\,\mathrm{m}$. Its length is $5.8\,\mathrm{m}$, for only $0.66\,X_0$.

The second one is a toroidal magnet system [168, 169] located outside the calorimeter system as visible in Figure 3.10(a). It consists of a barrel toroid and two end-caps responsible for bending the muon coming out of the inner part of the detector. Unlike most other 4π-symmetric detectors, which typically rely on a large solenoid for muon measurement, ATLAS uses this toroidal geometry. The choice of a toroidal geometry allows ATLAS to achieve a high-precision transverse-momentum measurement while avoiding a massive iron return yoke, thereby reducing the amount of material that muons must traverse and minimising multiple scattering. The barrel one is composed of eight superconducting coils arranged radially and symmetrically around the beam axis. The coil is of the same material as the solenoid, but with a cross-section of $57\,\mathrm{mm} \times 12\,\mathrm{mm}$ and a total length of $56\,\mathrm{km}$, and it carries a current of $20.5\,\mathrm{kA}$. They are all housed in their own cryostat to be cooled down to $4.5\,\mathrm{K}$. The resulting bending field is $0.5\,\mathrm{T}$. Each toroid itself has an outer radius of $10.05\,\mathrm{m}$ and an inner one of $4.7\,\mathrm{m}$. Its length is $25.3\,\mathrm{m}$. The two end-caps also use superconducting coils of the same materials, with a cross-section of $41\,\mathrm{mm} \times 12\,\mathrm{mm}$ and a total length of $13\,\mathrm{km}$, and they carry a current of $20.5\,\mathrm{kA}$. They are all housed in their own cryostat to be cooled down to $4.5\,\mathrm{K}$. The resulting bending field is approximately $1\,\mathrm{T}$. Each toroid itself has an outer radius of $5.35\,\mathrm{m}$ and an inner one of $0.825\,\mathrm{m}$. Its length is $5\,\mathrm{m}$.

3.6 Performance

Having gone through all the components of the ATLAS experiment, we now turn to the performance of each main part [170].

3.6.1 Tracking

The transverse momentum resolution in solenoids is given as:

$$\frac{\sigma_{p_\mathrm{T}}}{p_\mathrm{T}} = a_\mathrm{M} \times p_\mathrm{T} \oplus b_\mathrm{MS}\,, \tag{3.21}$$

where a_M is the term related to the measurement error of the track, b_MS corresponds to the multiple scattering effect, and p_T in GeV.

The nominal resolution of the tracking resulting from the ID is given by:

$$\frac{\sigma_{p_\mathrm{T}}}{p_\mathrm{T}} = 0.05\,\% \times p_\mathrm{T} \oplus 1\,\%\,. \tag{3.22}$$

When including the MS, we need to add a term c_cal to Equation 3.21 to take into account the energy loss in the calorimeter. This term is inversely proportional to the transverse momentum. The nominal resolution for the muon is $10\,\%$ at a p_T of $1\,\mathrm{TeV}$.

The impact parameters are central to the tracking performance. We distinguish the transverse impact parameter d_0 and the longitudinal one z_0. The first one is the distance of closest approach of the track to the beam axis in the transverse plane, while z_0 is the longitudinal position of the track at the point of closest approach to the beam axis. A visualisation of these two parameters is given in Figure 5.3 (Section 5.1.2). Their resolution is given by:

$$\sigma_{d_0} = a_\mathrm{M} \oplus \frac{b_\mathrm{MS}}{p_\mathrm{T} \times \sqrt{\sin\theta}}\,, \tag{3.23}$$

$$\sigma_{z_0} = a_\mathrm{M} \oplus \frac{b_\mathrm{MS}}{p_\mathrm{T} \times \sin^{3/2}\theta}\,, \tag{3.24}$$

where a_M is the term related to the measurement error of the track, b_MS corresponds to the multiple scattering effect, p_T in GeV, and θ is the polar angle. The resolution of the transverse impact parameter is about $10\,\mu\mathrm{m}$ to $15\,\mu\mathrm{m}$ and the longitudinal one is about $100\,\mu\mathrm{m}$. The resolution will degrade at large $|\eta|$ because of the reduced number of hits in the ID and the shorter track length in the solenoid, which reduces the lever arm for the momentum measurement. At low p_T, the resolution is dominated by multiple scattering, while at high p_T, it is dominated by the measurement error. The resolution of z_0 is worse than d_0 because the latter is measured in the transverse plane, where a better resolution is achieved due to the geometry of the detector and the magnetic field configuration. Also, the Pixel Detector and the SCT have a much finer granularity in (R, ϕ) than along the z-axis. The significance, defined as the ratio of the impact parameter to its uncertainty, $\frac{|d_0|}{\sigma_{d_0}}$ and $\frac{|z_0 \sin\theta|}{\sigma_{z_0 \sin\theta}}$, is usually used instead of the relative precision, since the aim is to make these values as small as possible.

The use of these variables is made explicit in the track reconstruction (Section 5.1.2).

3.6.2 Energy

The energy resolution for a calorimeter is:

$$\frac{\sigma_E}{E} = \frac{a_\mathrm{S}}{\sqrt{E}} \oplus \frac{b_\mathrm{N}}{E} \oplus c_\mathrm{C}\,, \tag{3.25}$$

where a_S is the stochastic term, b_N the noise term, c_C the constant term, and E in GeV.

The stochastic term affects the resolution by describing the statistical fluctuations in the shower development. For EM calorimeters, the main contribution arises from the sampling fluctuations due to event-by-event variations in the total energy deposit. It is influenced by the sampling fractions and the sampling frequency (number of active layers traversed by the shower along its longitudinal development). For HAD calorimeters, it is more complex. In addition to the effect of the EM case, we have to take into account the event-by-event variations in the EM part of the shower and the invisible energy. The noise term accounts for the electronic noise of the readout circuits, generally negligible in the energy range involved, and for the pile-up noise. The constant term comes from factors such as detector non-uniformity, alignment, and uncertainties in electronic calibration. The constant term becomes the limiting factor for resolution at very high energies.

The resolution for every calorimeter is given in Table 3.E.

Table 3.E: Parameters of the energy resolution $\frac{\sigma_E}{E}$ for the calorimeters [137].

Calorimeter	a_{S} [%]	b_{N} [GeV]	c_{C} [%]
EM	10.1 ± 0.4	negligible	0.2 ± 0.1
HEC	70.6 ± 1.5	negligible	5.8 ± 0.2
Tile	56.4 ± 0.4	negligible	5.5 ± 0.1
FCal-EM	28.5 ± 1.0	negligible	3.5 ± 0.1
FCal-HAD	94.2 ± 1.6	negligible	7.5 ± 0.4
EM & Tile	52.0 ± 1.0	1.6 ± 0.1	3.0 ± 0.1

3.7 Trigger system

The collisions at the IP occur at a frequency of $40\,\mathrm{MHz}$, meaning a bunch crossing every $25\,\mathrm{ns}$. Depending on the pile-up, there are n simultaneous pp collisions, all detected and recorded as an event. The average size of an event is between $1\,\mathrm{MB}$ and $2\,\mathrm{MB}$, thus the output would be around $40\,\mathrm{TB}\cdot\mathrm{s}^{-1}$ to $80\,\mathrm{TB}\cdot\mathrm{s}^{-1}$. Processing and storing this volume of data is not feasible. Therefore, a fast decision must be made to determine which event has to be kept or not for further analysis.

The ATLAS experiment operates a trigger system managing the Data AcQuisition (DAQ), the two together being referred to as the Trigger and Data AcQuisition (TDAQ) system, to handle the event selection and the readout of the signal sent by the roughly one hundred million channels that compose the detector and were listed in the previous sections. This system reduces the throughput to $1\,\mathrm{kHz}$ in Run-2 and $3\,\mathrm{kHz}$ in Run-3 thanks to two different sub-systems: the Level 1 (L1) trigger and the High-Level Trigger (HLT).

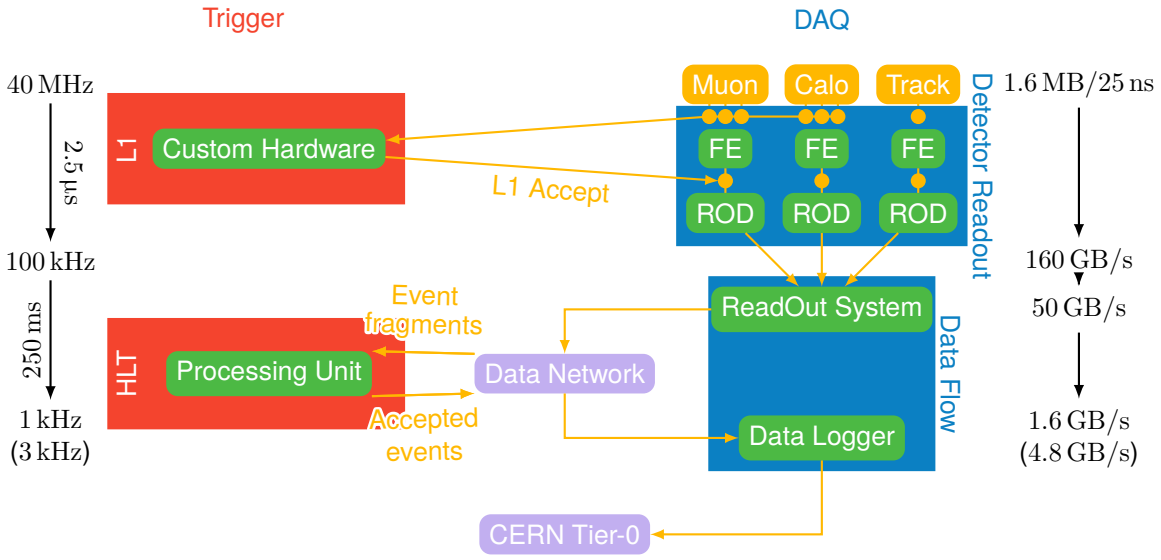


Figure 3.14: Trigger system in ATLAS. The output rates are given as Run-2 (Run-3).

The L1 [155, 171–173] has to identify basic signatures of interesting physics with a high efficiency. It uses custom application-specific integrated circuits (ASICs) and field-programmable gate arrays (FPGAs) to perform fast computations on low-granularity information coming from specific sub-detectors.

- The trigger chamber for the muon system, such as the RPC, TGC, and sTGC, as described in Section 3.4 to trigger on high p_{T}. This is referred to as the Level 1 Muon (L1Muon).
- The Trigger Tower or Super Cell in the EM calorimeter for respectively Run-2 and Run-3 as described in Section 3.3.1 to trigger on high-energy clusters as signatures of electrons, photons, taus, jets, or MET. This is referred to as the Level 1 Calorimeter (L1Calo).

It also considers the number of physics objects exceeding a certain p_{T} threshold, through the Central Trigger Processor (CTP), or topological quantities, through the Level 1 Topological (L1Topo), such as invariant masses and angular distances. Based on this varied information, the L1 decides within $2.5\,\mu\mathrm{s}$ reducing the event rate from $40\,\mathrm{MHz}$ down to $100\,\mathrm{kHz}$. The DAQ records every signal using the full granularity from all the sub-systems' readout electronics via the ReadOut Driver (ROD). It processes only

channels with active signals, formatting them and applying the different calibrations, and passes them to the ReadOut System (ROS) for buffering.

If the decision is made to keep the event, it goes to the next part of the trigger system, the HLT [155, 174]. It receives the full event data as well as the region of interests (RoIs) in the detector (η, ϕ) plane given by the L1 trigger. The HLT is a computer farm outside the cavern where the experiments lie, made of $40\,000$ $(60\,000)$ central processing units (CPUs) for Run-2 (Run-3). Their higher-level algorithms are applied to the data to perform partial event reconstruction, to reduce the event rate down to the final output of $1\,\mathrm{kHz}$ corresponding to $1.6\,\mathrm{GB}\cdot\mathrm{s}^{-1}$ for Run-2, and to increase it to $3\,\mathrm{kHz}$ for Run-3 using multithreading and more core units. The selection is done with a mean processing time of about $250\,\mathrm{ms}$ per event. The selected events are then sent to be stored for offline reconstruction. The data are transferred from the ROS to the CERN Tier-0 data-processing centre. They are first fully reconstructed and later distributed to the various computing centres of the LHC worldwide grid [175].

Here, the bottleneck is the speed at which the data can be written on disk and computed. Therefore, the limitation comes from the offline resources. The output of the HLT, already higher than expected, cannot be increased without altering the physics performance. The increase of pile-up through the different runs means more events are accepted by the L1 and a higher HLT output. To maintain the same rate without reducing or increasing the physics threshold, therefore reducing physics performance, both systems have to improve their rejection rate. This is done with new technologies such as the Super Cell for L1Calo or the NSW for the L1Muon, or implementation of more offline resources, resulting in a higher output.

The initial TDAQ system of ATLAS in Run-1 differed substantially from the one described here. It included a second trigger level, called Level 2 (L2), which is not described here, see [174]. A complete representation of the current TDAQ system is shown in Figure 3.14.

3.8 Luminosity measurement

As explained in Section 2.2, the determination of luminosity is a key factor in LHC operation and in particular in ATLAS physics analyses. If we rewrite Equation 2.4 by considering the luminosity for a single bunch, we obtain:

$$L_{\mathrm{b}} = \frac{\mu \cdot f_{\mathrm{rev}}}{\sigma_{\mathrm{inelastic}}} = \frac{\varepsilon_{\mathrm{eff}} \cdot \mu \cdot f_{\mathrm{rev}}}{\varepsilon_{\mathrm{eff}} \cdot \sigma_{\mathrm{inelastic}}} = \frac{\mu_{\mathrm{visible}} \cdot f_{\mathrm{rev}}}{\sigma_{\mathrm{visible}}}\,, \tag{3.26}$$

where $\varepsilon_{\mathrm{eff}}$ is the product of the acceptance and efficiency of the luminosity detector. This leads to μ_{visible}, the visible (detected) number of collisions per bunch crossing, and $\sigma_{\mathrm{visible}}$, the corresponding visible (effective) cross-section.

The parameter μ_{visible} depends on the luminosity monitoring method. It can be estimated using different algorithms, such as event-counting (number of interactions per crossing), hit-counting (counting the fraction of channels hit in a given bunch crossing), track- or vertex-counting, and flux-counting algorithms (flow of particles through specific sensors).

The quantity $\sigma_{\mathrm{visible}}$ depends on the luminosity calibration technique (minimum-bias measurement). Three main approaches exist for the absolute luminosity calibration:

- Determination from the rate of a well-known reference process, such as W or Z production
- Elastic scattering in the forward region using the optical theorem, although this is extremely delicate at the LHC due to low statistics, demanding beam conditions, and challenging backgrounds
- Determination via the beam parameters using the van der Meer (vdM) method [176]: convolving the beam profiles obtained from beam-separation scans at low μ. This method is performed during special vdM runs, where the beams are scanned in discrete steps by varying their separation in the transverse plane, σ^* in Equation 2.3.

In ATLAS, several sub-detectors provide luminosity measurements. The ID offers track-counting (from the mean number of reconstructed tracks) with bunch-by-bunch capability (limited by TDAQ), the Tile Calorimeter provides luminosity via PMT currents, while the LAr calorimeter (both ElectroMagnetic End-cap and Forward Calorimeter) does so via its gap currents. In addition, ATLAS employs a dedicated detector exclusively for luminosity measurement: LUminosity Cherenkov Integrating Detector (LUCID). Having multiple independent detectors is essential to assess the stability and systematic uncertainties of the luminosity measurement over time.

LUCID is the only detector primarily dedicated to online luminosity monitoring. Its main purpose is to detect inelastic pp interactions in the forward direction, to measure the integrated luminosity, and to provide real-time monitoring of the instantaneous luminosity and beam conditions. Its nominal design is described in [137], but it was completely rebuilt during the LS-1 shutdown, resulting in the LUCID-2 configuration [177].

The current detector is composed of two arms located at about $17\,\mathrm{m}$ from the IP, installed between the muon Big Wheels. Each arm contains four groups of quartz Čerenkov counters, each coupled to a radiation-hard PMT for a total length of $1.55\,\mathrm{m}$. They are visible in Figure 3.10(b). Charged particles with momentum above the Čerenkov threshold produce Čerenkov light in the quartz radiator, which is collected and converted to an electrical signal by the PMT. This process can be seen in Figure 3.15. The full LUCID system comprises 160 channels in total. The luminosity can be determined either by event counting or by hit counting. LUCID-2 remains in operation for Run-3, with new PMT prototypes of LUCID-3 [178], the Phase-II luminometer, installed further from the beam line, and the current PMTs replaced with new ones due to degradation from radiation.

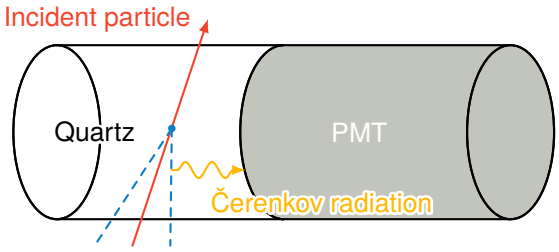


Figure 3.15: Scheme of the LUCID detector.

Diamond can serve as an alternative sensor material to the Si used in the Pixel and SCT detectors. Its wide band gap provides a high displacement threshold, ensuring minimal leakage current even after radiation exposure. This allows the sensor to operate at room temperature without active cooling. Additionally, the lower dielectric constant of diamond reduces the sensor's capacitance, lowering electronic noise. The main drawback is that the large band gap results in a smaller signal generated by incident particles compared to Si. Despite its high cost and limited availability, diamond's exceptional material properties make it highly suitable for environments with intense radiation.

The Beam Conditions Monitor (BCM) [179] is suspended from the beam pipe support that also holds the Pixel Detector. It is located only $55\,\mathrm{mm}$ from the beam line and $20\,\mathrm{mm}$ from the pipe, at a distance of $183.8\,\mathrm{cm}$ from the IP. It is visible in Figure 3.3(b). A diamond bulk a few hundred $\mu\mathrm{m}$ thick is placed between two electrodes, creating an electric field that moves the electrons produced by ionising incident particles. Each sensor covers an area of $8\,\mathrm{mm} \times 8\,\mathrm{mm}$ and is arranged around the beam pipe in a cross. The principle is similar to the Si detector of Section 3.2, and its timing resolution is on the order of ns. The BCM was designed for fast monitoring of beam conditions to protect the detector from damage. It served as the primary luminometer during Run-1 but underwent performance degradation in later runs and is now used as a cross-check for LUCID in vdM scans.

During the IBL [143] implementation, diamond sensors were also added in the form of the Diamond Beam Monitor (DBM) [180]. These sensors monitor instantaneous luminosity and the bunch-by-bunch position of the beam spot. Their design is similar to Pixel sensors, with each diamond pixel connected to the readout electronics via a solder bump (see Figure 3.4(a)). Each sensor contains $26\,880$ pixels arranged in 80 columns with a $250\,\mu\mathrm{m}$ pitch and 336 rows with a $50\,\mu\mathrm{m}$ pitch, covering an active area of $20\,\mathrm{mm} \times 16.8\,\mathrm{mm}$. This detector's area is nearly two orders of magnitude larger than that of the previous diamond detector, allowing high-precision particle tracking. The sensors are located $1\,\mathrm{m}$ from the IP, at an angle of $4°$, and $7\,\mathrm{cm}$ from the beam pipe. They are also visible in Figure 3.3(b). Diamond sensors are arranged in groups of three stacked along the beam axis, with four groups on each side of the detector, positioned symmetrically around the pipe.

ATLAS therefore computes μ_{visible} using signals from several detectors, primarily LUCID. The absolute calibration is obtained during dedicated vdM scans. By combining these measurements with the calibration, the luminosity per bunch crossing can be determined as described in Equation 3.26. The use of multiple, independent luminometers provides redundancy, allowing cross-checks and ensuring the most precise measurement possible.

3.9 Phase-II

After Run-3, as explained in Section 2.4, the LHC and more particularly ATLAS will undergo a massive upgrade to cope with the higher luminosity and pile-up expected in this high-luminosity (HL) phase.

Among the many upgrades, we can distinguish two categories: updates of pre-existing detectors and installations of new systems.

Regarding the first category, we have the electronics upgrades on the LAr [181] and Tile Calorimeter [182], but also on the MS [163]. This concerns both on- and off-detector electronics to provide a continuous readout at a frequency of $40\,\mathrm{MHz}$, as well as bringing a finer trigger segmentation, already begun with the Super Cells [154]. This comes with a rework of the TDAQ system [183, 184]. The new Level 0 (L0) trigger level will read the data output, giving a response at $1\,\mathrm{MHz}$ with a latency of $10\,\mu\mathrm{s}$. These new systems, therefore, exploit the full detector granularity from calorimeters and muon systems. After this L0 trigger comes the DAQ using unified back-end electronics with a Front-End Link eXchange (FELIX) card. The event filter will bring the output down to $10\,\mathrm{kHz}$, extending the tracking range and improving the muon efficiency. The use of graphics processing units (GPUs), machine learning (ML), and neural networks (NNs) will be an important part of the online reconstruction.

In addition to this, the ATLAS experiment will see several new sub-systems installed. New muon chambers will be incorporated, replacing current ones to improve the trigger efficiency, momentum resolution, and reduce the fake rate. The most important change will be the complete replacement of the ID by the Inner Tracker (ITk) [13, 14], improving the coverage from $|\eta| < 2.5$ to 4. It will be accompanied by the High-Granularity Timing Detector (HGTD) [185], providing precision time measurement and improving the pile-up separation and the bunch-by-bunch luminosity measurement.

In addition to this, the luminometer LUCID-3 [178] will be finalised, completing the installation of the additional prototypes tested during Run-3.

3.9.1 Muon Spectrometer update

The upgrade to the MS [163] brings changes to the TGC detector, the replacement of the MDT by the sMDT, as well as the installation of an RPC in the innermost barrel layer. Another system will also be added: the high-η tagger.

3.9.1.1 Thin Gap Chamber

The question of replacing the TGC in the Big Wheel was raised, but it was found to be unnecessary since they can sustain the expected radiation dose and still perform well. Only the EIL4 chambers will be replaced. They were not originally designed to be part of the trigger system and so are not optimised to work for triggering. They are only double layers, making it harder to trigger if one of the layers malfunctions. Therefore, they will be replaced by triple-layer TGCs with higher granularity while maintaining the same chamber envelopes to avoid affecting the other systems around. The technology used will be similar to the one detailed in Section 3.4.4.

3.9.1.2 Small-diameter Monitored Drift Tube

The current MDT of the innermost layer of the Muon Barrel Spectrometer will be replaced by sMDT. As explained in Section 3.4.1, they are the same technology as the initial one but with the diameter of the tube reduced by half, bringing it down to $15\,\mathrm{mm}$. This implementation started with the change of BIS78 only during LS-2 as a pre-test. This will be extended to the whole layer to make space for the insertion of the Barrel Inner (BI) RPC.

3.9.1.3 Barrel Inner Resistive Plate Chamber

A layer of BI RPC will be installed in the spectrometer to add redundancy to the system and recover most of the current geometrical acceptance holes. The trigger system will then be more efficient and will increase the acceptance. At the same time, it will lose the requirement on the number of hits in the outer layers by adding a requirement on the coincidence with this new layer. This new RPC chamber will have three sensitive gas gaps, with their size decreased to $1\,\mathrm{mm}$ as well as the electrode thickness reduced to $1\,\mathrm{mm}$. They can be seen in Figure 3.10(b). Again, the BIS78 was a first step towards the full installation planned.

3.9.1.4 High-η tagger

A proposal for a high-η tagger is being studied to extend the muon coverage in the forward region. It will help identify (tag) the muons that penetrate the end-cap calorimeter. They will be identified by matching a track from the ITk to a segment reconstructed in the tagger. The p_{T} would then be determined by the ITk. The coverage would then be similar to the tracker ($|\eta| < 4.0$).

This new tagger would be located at $6.8\,\mathrm{m}$ from the IP as shown in Figure 3.10(b). It would be a wheel with an inner radius of $25\,\mathrm{cm}$ and outer radius of $90\,\mathrm{cm}$ and of $5\,\mathrm{cm}$ thickness. The requirement for the spatial resolution is $0.5\,\mathrm{mm}$.

Diverse technologies are considered for this new system. The most likely candidate is a micro-pattern gas detector (MPGD), with different technologies such as Resistive Pixelated MM (similar to the one used in the NSW), micro-pixel chamber (μ-PIC), micro-resistive WELL (μ-RWELL), and fast-timing MM. A Si pixel option is also considered.

A very high timing resolution of a few tens of ps may be useful to reject hits due to cavern background (the diffuse flux of low-energy neutrons and photons filling the experimental cavern, produced by the interactions of collision debris with the detector, the shielding, and the beam-line elements) based on precision timing, and fast timing detectors are therefore also being studied. Technologies may also be combined for separate optimisation of, for example, the inner and outer regions.

A delay in the development of this system has reduced the priority of the high-η tagger. It is more likely that it will be installed only after LS-3 and not used for Run-4.

3.9.2 Inner Tracker

The current ID will be replaced by the ITk (Figure 3.16), as it was not originally designed to withstand the challenging conditions of the HL-LHC.

Firstly, it was not built to be part of the trigger system. Incorporating a fast, hardware-based track-finding capability would greatly enhance processing speed and reduce costs, thereby improving the overall physics performance. The ITk is not part of the hardware trigger either: in the Phase-II TDAQ system, tracks are only reconstructed in the software event filter [184]. Secondly, the current tracker has suffered significant radiation damage and will be unable to operate in the high-radiation environment expected with a pile-up of around 200. Thirdly, since it was not designed to record the large number of hits expected at such pile-up levels, severe bandwidth saturation is anticipated. Lastly, the increased number of hits, without a corresponding improvement in granularity, will lead to confusion in pattern recognition and track finding, reducing efficiency. For all these reasons, the current ID is outdated, and its replacement with the ITk is essential.

Due to the importance of analyses involving charged particles with high p_{T}, the idea of extending the coverage in the forward region was raised. It was therefore decided to enhance the region covered by the ITk to provide tracking up to $|\eta| < 4.0$.

In the central region, the sensors are arranged in cylinders around the beam axis with five pixel layers in the innermost part, followed by two short-strip layers and then two long-strip stereo module pairs with an angle of $52\,\mathrm{mrad}$ in the barrel and $40\,\mathrm{mrad}$ in the end-caps. The forward regions are covered by six strip disks and multiple pixel rings, resulting in one or more hits depending on the η position. This configuration provides up to 13 hits for $|\eta| < 2.7$, combining both strips and pixels, decreasing to at least nine hits at higher $|\eta|$.

The total amount of material corresponds to roughly $0.3\,X_0$ for $0 < |\eta| < 1$, peaking at almost $2\,X_0$ between $3.0 < |\eta| < 4.0$ due to the support structure. Compared to the current ID, this represents a $30\,\%$ reduction for $|\eta| < 4.0$.

The system is cooled down to $-40\,^{\circ}\mathrm{C}$ using CO_2.

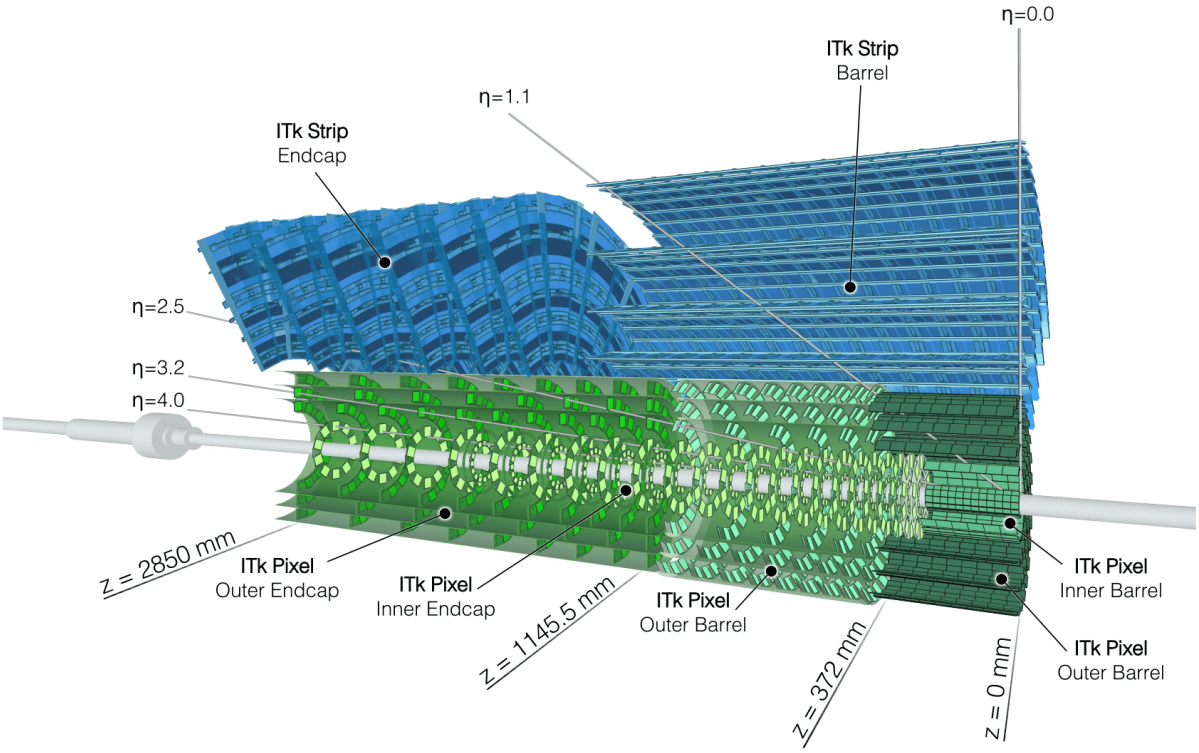
3.9.2.1 Strip Detector

The ITk Strip Detector [14] consists of a four-layer barrel section and one end-cap on each side with six disks. It covers the region $|\eta| < 2.7$, extending to $1.4\,\mathrm{m}$ from the IP in the barrel and up to $2.85\,\mathrm{m}$ in the end-caps. The two inner barrel layers are equipped with short strips of $24.1\,\mathrm{mm}$ length, while the two outer layers have longer strips of $48.2\,\mathrm{mm}$. All barrel strips have a pitch of $75.5\,\mu\mathrm{m}$. The barrel sensors have an active area of $96.64\,\mathrm{mm} \times 96.669\,\mathrm{mm}$ with 1280 readout strips.

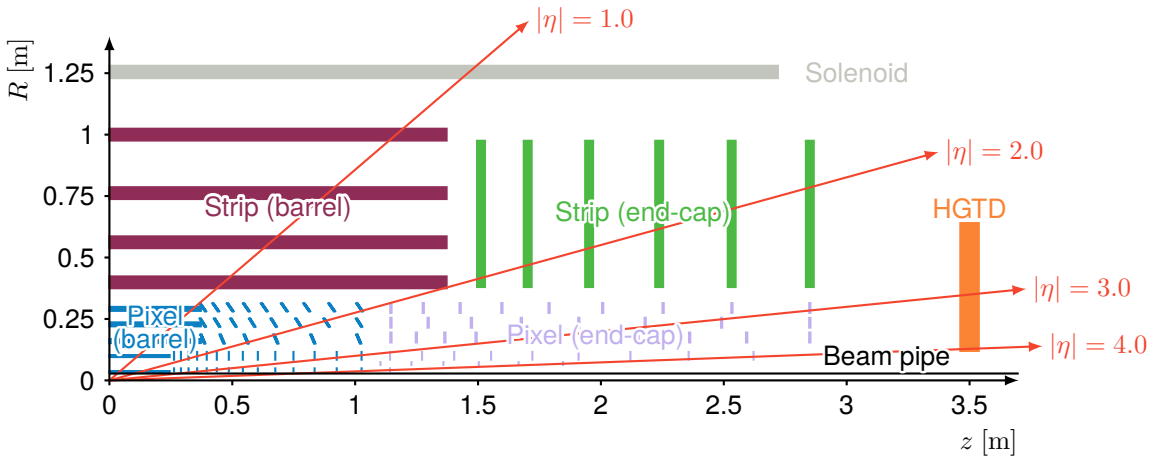
Stereo angles are implemented in both the barrel and end-cap systems, providing a second coordinate measurement similar to the current SCT, as explained in Section 3.2.2.

The strips are made with n-type implants in a p-type float-zone Si bulk. This type of sensor collects electrons and does not undergo radiation-induced type inversion. The target thickness is between $300\,\mu\mathrm{m}$ and $320\,\mu\mathrm{m}$. Choosing this n-in-p technology over the p-in-n used in the SCT ensures a large signal even after irradiation. A representation of the strip used in ITk is given in Figure 3.17(a).

In total, about $18\,000$ strip modules are used, providing approximately 60 million channels and covering a surface of $165.25\,\mathrm{m}^2$.

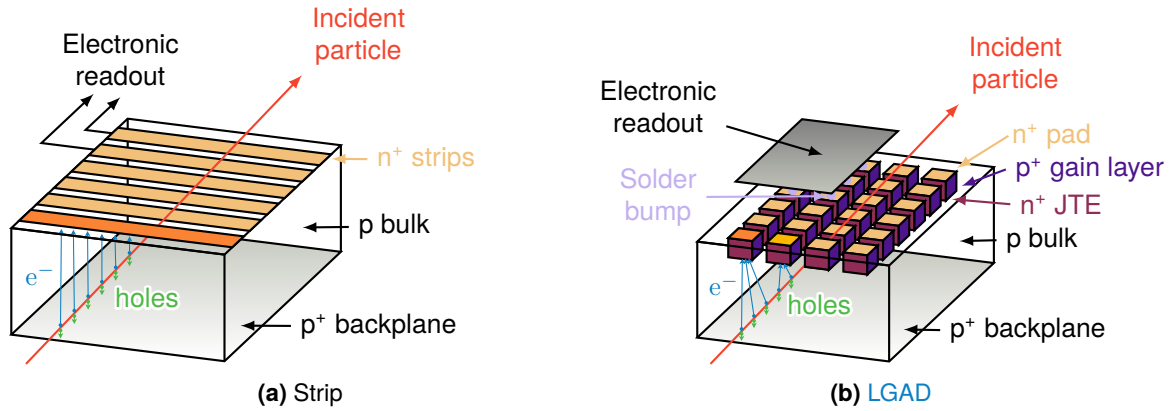


(a)



(b)

Figure 3.16: ITk in 3D [138] (a) and its layout (b).



(a) Strip **(b)** LGAD

Figure 3.17: Scheme of different types of Si-based sensors used in the ITk (a) and the HGTD (b).

3.9.2.2 Pixel Detector

The ITk Pixel Detector [13] layout consists of five barrel layers and four end-cap ring layers, covering the region $|\eta| < 4.0$, extending to $1.1\,\mathrm{m}$ from the IP in the barrel and $2.85\,\mathrm{m}$ in the end-caps.

Tilting the sensors in the end-caps allows for full-angle coverage with fewer pixels. This corresponds to a $20\,\%$ reduction in the total Si area, reducing both cost and multiple scattering. Moreover, smaller pixel clusters per particle crossing improve tracking precision.

Two distinct technologies will be used. First, 3D pixel sensors, similar to those present in the IBL (see Figure 3.4(b)), will be installed on the innermost layer. They will have a size of $25\,\mu\mathrm{m} \times 100\,\mu\mathrm{m}$ in the barrel and $50\,\mu\mathrm{m} \times 50\,\mu\mathrm{m}$ in the end-caps. Second, the remaining layers will use planar sensors of $50\,\mu\mathrm{m} \times 50\,\mu\mathrm{m}$ across all regions.

The planar pixels used in the ITk are similar to those in the current ID. The main differences are the reduced pixel size, improving intrinsic resolution and two-track separation, and the upgraded readout chip design. The analogue front-end operates at a lower threshold to compensate for charge loss due to radiation damage. The readout architecture handles higher hit density and event rates, and the radiation tolerance has been significantly increased. Like the strips, the sensors use n-in-p technology, making them radiation-hard and less susceptible to type inversion.

About 9200 pixel modules will cover about $13\,\mathrm{m}^2$, increasing the number of channels by two orders of magnitude to about five billion.

3.9.3 High-Granularity Timing Detector

In the expected pile-up conditions, an average interaction density of roughly $1.8\,\mathrm{vertex} \cdot \mathrm{mm}^{-1}$ is foreseen. Moreover, the resolution on the impact parameter of a track worsens as $|\eta|$ increases. Consequently, several vertices may be merged, degrading the performance in forward particle reconstruction. Introducing an additional dimension, time, in the measurement provides a new handle to mitigate the effects of the increased interaction density per mm. Assigning a time to each track with sufficiently small time resolution would greatly enhance the discrimination power of ATLAS, by a factor of about six for the envisaged time resolution of $30\,\mathrm{ps}$.

To provide this timing information, a new detector, the HGTD [185], will be installed during LS-3. It will cover the region $2.4 < |\eta| < 4.0$. It lies at the very end of the ITk end-caps, at a distance of $3.5\,\mathrm{m}$ from the IP, forming a wheel of $75\,\mathrm{mm}$ thickness with an inner radius of $120\,\mathrm{mm}$ and an outer radius of $640\,\mathrm{mm}$.

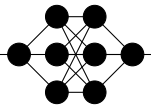
The sensors used are planar Si called low-gain avalanche detectors (LGADs). Each pad consists of a p-type multiplication layer located just beneath the n^+ implantation. It is surrounded by a junction termination extension (JTE), also n^+, which shapes the electric field at the edges. The bulk consists first of a high-density p-type layer followed by a low-density p-type layer. The structure is sandwiched by a metal plate where the electrons are collected on top. This results in an n-in-p sensor that remains radiation-hard. The key feature of the LGAD is the highly doped p-layer located underneath the readout, which provides internal gain that reduces noise amplification. The sensor is presented in Figure 3.17(b). The pad dimensions are $1.3\,\mathrm{mm} \times 1.3\,\mathrm{mm}$ (within $2\,\mathrm{cm} \times 4\,\mathrm{cm}$ sensors) and only $50\,\mu\mathrm{m}$ thick, making them suitable for fast timing response as required.

Two layers of sensors are mounted back-to-back on the wheel, providing at least two hits per track. The whole system is operated at $-30\,^\circ\mathrm{C}$.

In total, 8032 modules are used, corresponding to about 3.6 million channels, forming a total active area of $6.3\,\mathrm{m}^2$.

The HGTD can also provide instantaneous luminosity measurements, giving a fast signal as a function of the number of hits per bunch crossing. The high granularity also ensures low occupancy.

4 Machine Learning



The field of ML has become one of the most transformative areas in modern computational science, bridging statistics, computer science, and applied mathematics to enable computers to learn from data. Its origins trace back to the 1950s, when the concept of automating pattern recognition and decision-making began to take shape. The term ML was formally introduced in 1959 [186] in the context of self-improving computer programs. This development traces its philosophical roots to Alan Turing's seminal question, *"Can machines think?"* [187], which laid the foundation for the field of artificial intelligence (AI), of which ML is a major branch.

In its early decades, the field advanced through research in pattern recognition and probabilistic modelling, disciplines focused on learning directly from empirical data using methods such as linear regression (LR), Bayesian inference, and nearest-neighbour classification. By the 1980s, with the growth of computational power and the increasing availability of data, algorithmic frameworks such as decision trees (DTs), support vector machines, and ensemble methods emerged. These developments formalised learning as an optimisation problem defined over empirical risk.

A key milestone in this evolution was the development of NNs. The concept was first introduced in 1943 [188] as a mathematical model inspired by biological neural processes. An artificial neuron combines multiple inputs to produce an output, mimicking the behaviour of a biological neuron. Building on this idea, Hebb's postulate of neural plasticity in 1949 [189] proposed that connections between neurons are strengthened through simultaneous activation, establishing the theoretical basis for learning in neural systems. One of the first artificial NNs, the perceptron [190], implemented this principle by linearly combining inputs and applying a threshold function to make decisions, using adjustable connection strengths, or *weights*. The introduction of the back-propagation algorithm in 1986 [191] enabled efficient training of multi-layer networks and laid the groundwork for what would become *deep learning*.

Although early NNs faced significant challenges, such as limited computational power and the vanishing gradient problem, the field experienced a resurgence in the 2000s. This revival was driven by the availability of large-scale datasets, advances in optimisation techniques, and the emergence of powerful graphics processing units, enabling the training of deep architectures. These deep neural networks (DNNs) revolutionised numerous fields, from computer vision and natural language processing to high-energy physics, where they are now integral tools for data analysis and event classification.

Modern machine learning can thus be viewed as a synthesis of decades of interdisciplinary progress. It combines the theory of statistics with the expressive capacity of high-dimensional neural architectures. It provides a powerful framework for constructing models that can infer complex patterns, generalise to unseen data, perform categorisation, and approximate underlying generative processes.

This chapter follows this historical description and conceptual foundation to introduce the formalism, mathematical tools, and computational mechanisms of NNs. It explores the architecture and training of modern NNs, emphasising their ability to model non-linear dependencies in high-dimensional feature spaces. Moreover, it discusses practical aspects such as data preprocessing, regularisation, and optimisation strategies. By connecting theoretical concepts to experimental analyses, the chapter aims to provide both an intuitive and quantitative understanding of how ML techniques are shaping contemporary research in high-energy physics.

4.1 Neural network

4.1.1 Neuron

NNs take inspiration from biological neurons (brain cells), imitating their design. Therefore, the basic constituents of such networks are the neurons, also called *nodes*. An illustration of a single neuron is depicted in Figure 4.1.

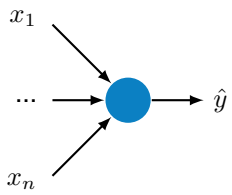


Figure 4.1: Scheme of a single neuron.

In this model, the neuron takes n inputs (denoted x) and produces an output $\hat{y}$. For this, it linearly sums all the inputs and incorporates weights w to represent their importance. Finally, we introduce a constant offset to this sum, called *bias* b. This allows the neuron to produce an output even with zero input. In addition to this, the entire sum (including the bias) is passed through an activation function ϕ, allowing it to rescale the output but also act as a threshold determining whether or not to produce an output. This gives us:

$$\hat{y} = \phi\left(\sum_{i=1}^{n} w_i \cdot x_i + b\right) . \tag{4.1}$$

The bias thus affects this activation. More details on this activation function will be given later.

4.1.2 Network

With the basic constituents of NNs established, we can now consider the larger structure. We can combine them by forming layers and stacking them as illustrated in Figure 4.2.

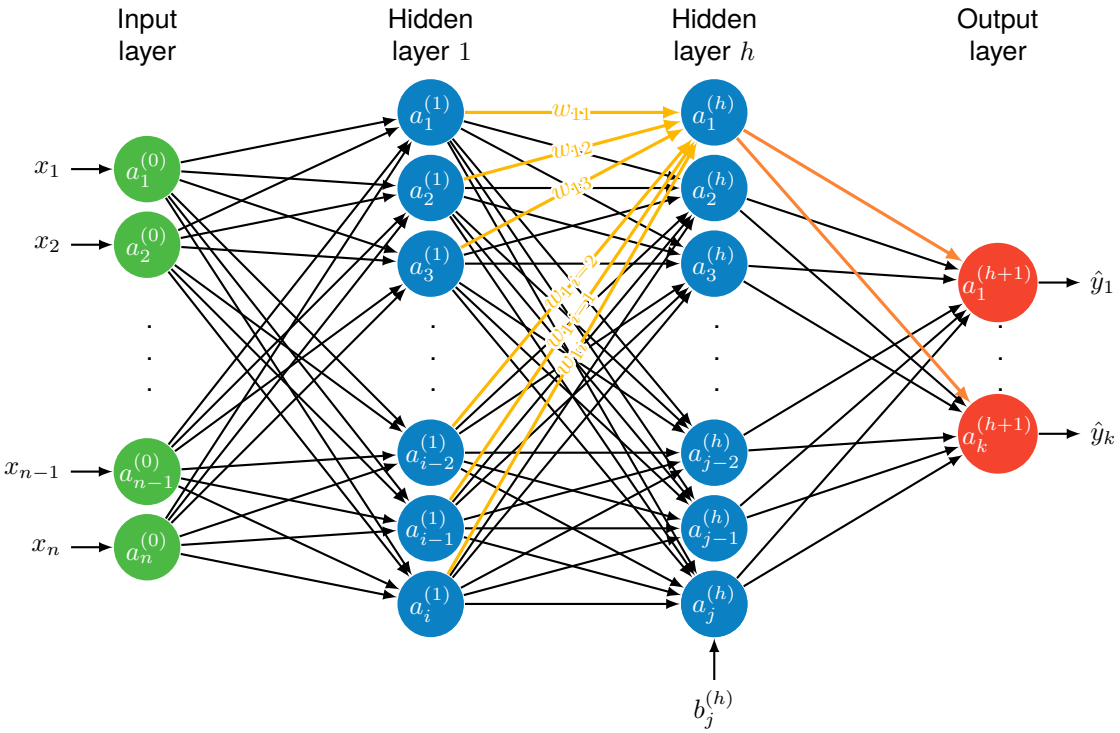


Figure 4.2: Scheme of an NN.

We can distinguish three distinct types of layers. The input layer labelled (0) receives the data that the NN will process. These input data can be of various types: images, numbers, integers, or booleans. Next

are the hidden layers numbered from (1) to the total number of them (h) (this number can be zero). The values flowing through the hidden layers are generally not directly observable or interpretable by the user. Finally, after the hidden layers, the output layer produces the final results $\hat{y}$. The number of nodes per layer (represented here by n, i, j, and k) can be equal or different and is generally chosen based on the complexity of the problem or computational constraints.

By applying Equation 4.1 to this network, we obtain:

$$a_k^{(l)} = \phi^{(l)} \left(\sum_{i=1}^{n^{(l-1)}} w_{ki}^{(l)} \, a_i^{(l-1)} + b_k^{(l)} \right) = \phi^{(l)}(z_k^{(l)}) \,. \tag{4.2}$$

Here, $a_k^{(l)}$ denotes the output of neuron k in layer l, computed using the activation function $\phi^{(l)}$. It sums over all neurons i in the previous layer $(l-1)$, multiplying each output $a_i^{(l-1)}$ by the corresponding weight $w_{ki}^{(l)}$, and then adds the bias $b_k^{(l)}$ associated with neuron k.

It can be expressed in matrix and vector form as:

$$\begin{aligned} \begin{pmatrix} a_1^{(l)} \\ a_2^{(l)} \\ \vdots \\ a_m^{(l)} \end{pmatrix} &= \phi^{(l)} \left(\begin{pmatrix} w_{11}^{(l)} & w_{12}^{(l)} & \cdots & w_{1n}^{(l)} \\ w_{21}^{(l)} & w_{22}^{(l)} & \cdots & w_{2n}^{(l)} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m1}^{(l)} & w_{m2}^{(l)} & \cdots & w_{mn}^{(l)} \end{pmatrix} \begin{pmatrix} a_1^{(l-1)} \\ a_2^{(l-1)} \\ \vdots \\ a_n^{(l-1)} \end{pmatrix} + \begin{pmatrix} b_1^{(l)} \\ b_2^{(l)} \\ \vdots \\ b_m^{(l)} \end{pmatrix} \right) \\ \boldsymbol{a}^{(l)} &= \phi^{(l)} \left(\mathbf{W}^{(l)} \boldsymbol{a}^{(l-1)} + \boldsymbol{b}^{(l)} \right) = \phi^{(l)}(\boldsymbol{z}^{(l)}) \,, \end{aligned} \tag{4.3}$$

where bold characters denote the matrix and vector elements.

The input layer, represented by the $a^{(0)}$ values, consists simply of the input parameters x, while the output $\hat{y}$ is equal to $a^{(h+1)}$.

The idea behind this kind of structure is that with enough hidden layers (the depth of a network) and enough neurons (its width), the NN can approximate any function. Indeed, Equation 4.2 can be viewed as a linear function. Therefore, by stacking them as the network structure does, each layer adds a piecewise-linear region that contributes to approximating the target function. This is known as the *Universal Approximation Theorem* [192]. It was later shown [193] that the activation function must be non-polynomial to avoid the network being a simple linear function. It is important to note that the theorem is existential: it guarantees the existence of a network that can approximate the function, but does not provide a method to find the optimal weights.

To conclude, a whole NN can be viewed as a function f depending on various parameters taking input $\boldsymbol{x}$, resulting in output $\hat{\boldsymbol{y}}$ as stated by:

$$\hat{\boldsymbol{y}} = f(\boldsymbol{x}) \,. \tag{4.4}$$

We have seen that NNs can approximate a function, but we must evaluate how well they do it. For this, we use a loss function.

4.1.3 Loss function

Loss functions, L, evaluate the mismatch between the predicted value $\hat{\boldsymbol{y}}$ of the NN and the true value $\boldsymbol{y}$, referred to as the *label*. Different choices are available and will depend on the problem the NN will face.

For example, the simplest one is the mean square error (MSE) given by:

$$L_{\text{MSE}} = \frac{1}{D} \sum_{d=1}^{D} (\boldsymbol{y}_d - \hat{\boldsymbol{y}}_d)^2 \,. \tag{4.5}$$

It corresponds to the sum, over the number D of elements in the dataset, of the squared difference between the predicted value and the label, averaged over the number of entries in the dataset D. The sum over the number of outputs is implicit, given by the vector notation. The squaring in this function removes information on the direction of the deviation.

In the literature, one can find the terms loss functions and cost functions. To be precise, the loss function is actually each of the terms of the sum in Equation 4.5 while the cost function is the overall quantity. Later on, those terms will be used interchangeably.

This function works well for regression. For a classification problem, better options are available. In those situations, we deal with classes. It means that the NN has to differentiate between two or more objects, each associated with a specific label. Here we can have either one output providing two or more

labels, multiple outputs providing a single value for each class (multi-class categorisation) or even multiple outputs producing at the same time different values for each class to identify (multi-label).

In the case of a binary classification, that is with two classes and one output, either 0 or 1, a natural choice for the loss function is the binary cross entropy (BCE) defined as:

$$L_{\mathrm{BCE}} = \frac{-1}{D}\sum_{d=1}^{D}\left[(1-\boldsymbol{y}_d)\ln\left(1-\hat{\boldsymbol{y}}_d\right)+\boldsymbol{y}_d\ln\hat{\boldsymbol{y}}_d\right]. \tag{4.6}$$

In the case of multiple classes, two choices can be made regarding the output. For example, one single output can have a specific integer for each class, but in this case resemblance between two classes can be misidentified by the NN due to the non-linearity in this output. Alternatively, we can one-hot encode the label. Instead of having one single output, we choose as many outputs as classes, and each will be either 0 or 1. The latter choice is more robust when some classes are similar. In this case, the function L_{BCE} can be adapted to become the categorical cross entropy (CCE) (coming from the categorical distribution) and given by:

$$L_{\mathrm{CCE}} = \frac{-1}{D}\sum_{d=1}^{D}\sum_{c=1}^{C}\boldsymbol{y}_{d,c}\cdot\ln\left(\hat{\boldsymbol{y}}_{d,c}\right). \tag{4.7}$$

The sum runs over the classes $c = 1,\ldots,C$.

The loss function, therefore, heavily depends on the problem the NN has to answer. Many choices exist and custom functions are also possible, but they all provide information on how well the NN performs. They map the phase-space defined by the problem with minima and extrema. To optimise the parameters of the NN to enhance its performance, we therefore need to minimise this loss function. This is done via gradient descent by finding minima corresponding to the best solution of the problem.

4.1.4 Gradient descent

Let us consider a loss L depending on all the parameters Ω of the network (weights and biases). To minimise this loss, we must first compute the derivatives of the loss with respect to the parameters:

$$\frac{\partial L}{\partial \Omega} = \begin{bmatrix} \frac{\partial L}{\partial \Omega_0} \\ \vdots \\ \frac{\partial L}{\partial \Omega_N} \end{bmatrix}. \tag{4.8}$$

We can update each parameter following the rule:

$$\Omega \leftarrow \Omega - \alpha\cdot\frac{\partial L}{\partial \Omega}, \tag{4.9}$$

with α a positive scalar determining the magnitude of change.

The first step of the gradient descent is therefore to compute the gradient of the loss function. This gives us the uphill direction of the function. In the second step, we move by a small distance α downhill (via the negative sign). This parameter is called the *learning rate*. We repeat this process, and once the minimum is reached, the surface must be flat, resulting in a vanishing gradient, stopping the parameters from evolving.

Due to the large number of parameters an NN can have, it is computationally challenging to find the global minimum. We can find a minimum, but without certainty that it corresponds to the global one. The final destination of this descent is determined by the starting point. For example, in Figure 4.3, the orange points show a gradient descent with a low learning rate. It goes down smoothly, but gets stuck in a local minimum. Meanwhile, in yellow, the high learning rate makes it pass over the minima but ends up bouncing on the side of the loss function. Algorithms tend to address this issue to optimise the descent by modifying the parameter update.

4.1.4.1 Stochastic gradient descent

The stochastic gradient descent (SGD) [194] attempts to remedy this problem by adding noise to the gradient at each step. The solution still moves downhill on average, but at any given iteration, the direction chosen is not necessarily in the steepest downhill direction. Indeed, it might not be downhill at all. This algorithm has the possibility of moving temporarily uphill and hence jumping from one “valley” of the loss function to another.

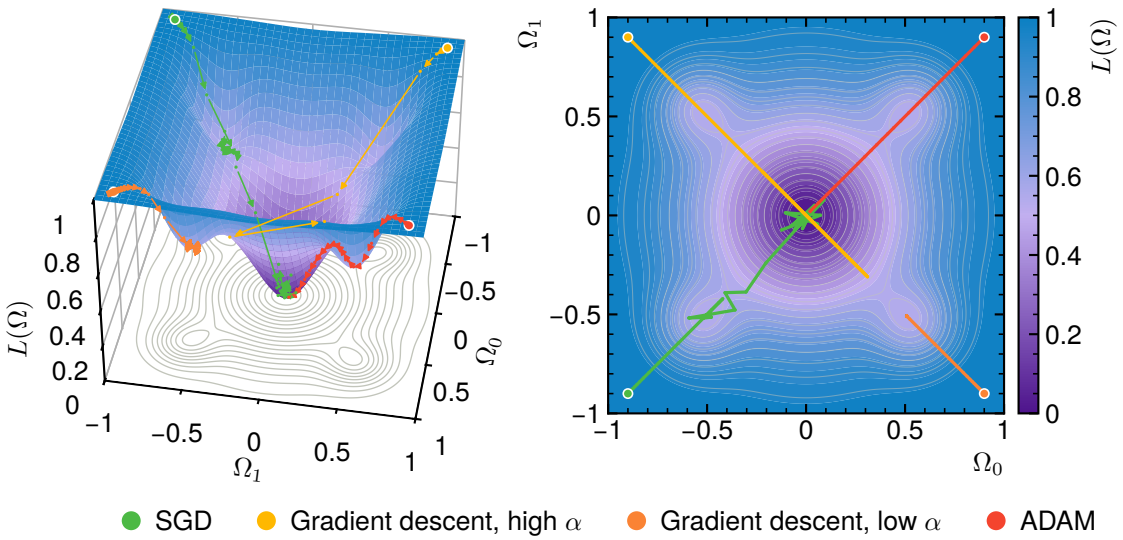


Figure 4.3: Comparison of various gradient descent algorithms.

This is visible in Figure 4.3 where the green points use an SGD method. It first goes to the local minimum but succeeds in escaping it via the noise induced by the batch.

At each iteration, the algorithm takes a random batch of the dataset and computes the gradient from it. The update rule for the parameters at this iteration t is given by:

$$\Omega_{t+1} \leftarrow \Omega_t - \alpha \cdot \sum_{i \in \mathcal{B}_t} \frac{\partial\, l_i(\Omega_t)}{\partial\, \Omega}\,, \tag{4.10}$$

where $\mathcal{B}_t$ contains all the indices of the input and output pairs in the current batch, and l_i is the loss due to the i-th pair.

This process is done for every batch until the full dataset is seen: this is referred to as an *epoch*.

This technique, despite adding noise, still improves the fit to a subset of the dataset at each iteration. The updates tend to be sensible even if they are not optimal. Moreover, since it sees the entire dataset, the contribution of each batch is equal. It is also less computationally expensive to compute the gradient on a batch rather than the whole dataset. In principle, it can escape local minima by moving uphill. Finally, this algorithm does not properly converge, but when near the global minimum, all the data points are well described. Consequently, the gradient will be small, whatever batch is chosen.

A learning rate scheduler is often used alongside it to add robustness. It makes the learning rate evolve through the epochs. In the early stages, a bigger value is chosen to explore the parameter space and jump from valley to valley to find a sensible region. At the later stages, we start decreasing these values to be finer in the adjustments and stay within the valley.

4.1.4.2 Adaptive moment estimation

Gradient descent with a fixed step size makes large changes to parameters associated with larger gradients and vice versa for small gradients. When the gradient is much steeper in one direction than another, it is difficult to choose a learning rate that makes good progress in both directions and is stable.

A solution is to normalise the gradient to move a fixed distance (dictated by the learning rate) in each direction. For this, we compute the gradient and the pointwise squared gradient. This will result in good progress in both directions, but will not converge unless it lands exactly at the minimum. Otherwise, it will bounce back and forth around the minimum.

The ADAptive Moment estimation (ADAM) [195] takes this idea and adds momentum [196] to it. The momentum m and the pointwise squared gradient v are defined as:

$$m_{t+1} \leftarrow \beta \cdot m_t + (1-\beta) \sum_{i \in \mathcal{B}_t} \frac{\partial\, l_i(\Omega_t)}{\partial\, \Omega}\,, \tag{4.11}$$

$$v_{t+1} \leftarrow \gamma \cdot v_t + (1-\gamma) \left(\sum_{i \in \mathcal{B}_t} \frac{\partial\, l_i(\Omega_t)}{\partial\, \Omega} \right)^2\,, \tag{4.12}$$

with $\beta, \gamma \in [0, 1)$, the momentum coefficients that control the degree to which the gradient is smoothed over time. The use of the learning rate as well as the batch principle is still present.

This recursive formulation of the momentum calculation implies that the gradient step is an infinite weighted sum of all the previous gradients, where the weights get smaller as we move back in time. The effective learning rate increases if all these gradients are aligned over multiple iterations, but decreases if the gradient direction repeatedly changes as the terms in the sum cancel out. The overall effect is a smoother trajectory, reducing the oscillation behaviour in valleys.

At the start of the procedure, all the previous measurements are effectively zero, resulting in unrealistically small estimates. Therefore, we modify it using:

$$\tilde{m}_{t+1} \leftarrow \frac{m_{t+1}}{1 - \beta^{t+1}}\,, \tag{4.13}$$

$$\tilde{v}_{t+1} \leftarrow \frac{v_{t+1}}{1 - \gamma^{t+1}}\,. \tag{4.14}$$

Since β and γ are in the range $[0, 1)$, the terms with exponents $t + 1$ become smaller with each time step, making the denominators close to one.

We update the parameter as follows:

$$\Omega_{t+1} \leftarrow \Omega_t - \alpha \cdot \frac{\tilde{m}_{t+1}}{\sqrt{\tilde{v}_{t+1}} + \epsilon}\,, \tag{4.15}$$

where ϵ is a small constant to prevent division by zero.

The positive root of v is used to normalise the gradient itself, so all that remains is the sign in each coordinate direction. The result is that the algorithm moves a fixed distance α along each coordinate, where the direction is determined by whichever way is downhill. ADAM also has the advantage of being less sensitive to the initial learning rate, so it does not need complex learning rate schedules. In Figure 4.3, the red points show an ADAM method. It can be seen that the momentum gain in the descent allows it to overcome the local minimum and reach the global one smoothly.

Having seen how to efficiently minimise the loss function, we now need to correctly update all the parameters throughout all the layers. For this, we use an algorithm called *back-propagation*.

4.1.5 Back-propagation

The goal of the back-propagation algorithm [191] is to compute the derivative term in Equations 4.10 and 4.11 with respect to the weights w and the biases b for every node in every layer. Any change to one of the NN parameters will cause a cascade of changes through the network. Indeed, a modification of one node will change its value in the following layer and so on until it changes the output, and consequently the loss. To know how a change will affect the loss, we need to know each subsequent change in each layer. For example, in Figure 4.2, a change in the parameters of $a_1^{(h)}$ will affect the outputs (orange arrows) and the loss. We need to first look at the previous connection these units (yellow arrows) have, and so on, until going back to the input layer.

For a single sample $i \in \mathcal{B}_t$, where $\mathcal{B}_t$ is the current batch, the derivative of the loss l_i with respect to a given weight or bias in layer l can be written as:

$$\begin{aligned}
\frac{\partial\, l_i(\Omega_t)}{\partial\, w_{kj}^{(l)}} &= \frac{\partial\, z_k^{(l)}}{\partial\, w_{kj}^{(l)}}\overbrace{a_j^{(l-1)}} \frac{\partial\, a_k^{(l)}}{\partial\, z_k^{(l)}} \frac{\partial\, l_i}{\partial\, a_k^{(l)}}\,, \\
\frac{\partial\, l_i(\Omega_t)}{\partial\, b_k^{(l)}} &= \underbrace{\frac{\partial\, z_k^{(l)}}{\partial\, b_k^{(l)}}}_{1} \underbrace{\frac{\partial\, a_k^{(l)}}{\partial\, z_k^{(l)}}}_{\phi^{(l)\prime}(z_k^{(l)})} \underbrace{\frac{\partial\, l_i}{\partial\, a_k^{(l)}}}_{\propto\, \boldsymbol{y} - \hat{\boldsymbol{y}}}\,.
\end{aligned} \tag{4.16}$$

To retrieve the full loss of the NN, we do as in Section 4.1.4.1 with a sum over every element of the batch. The index k corresponds to the neuron k in layer (l) connected to the neuron j in the layer $(l-1)$.

The first term is straightforward to compute with Equation 4.2. It is simply $a_j^{(l-1)}$ or one, respectively, when differentiated with respect to the weight and the bias. This implies that a change in the weight will have a stronger effect on $z_k^{(l)}$ when the previous neuron is more active. The second term, the derivative with respect to $z_k^{(l)}$, will be the derivative of the activation function $\phi^{(l)}$. In addition to the properties of this function given before, it must also be continuous and differentiable. Finally, the last derivative with respect

to $a_k^{(l)}$ will depend on the choice of the loss function. The various examples of loss functions presented in Section 4.1.3 are easily differentiable. It will often be proportional to the difference between the prediction of the NN and the truth value. This means that when the actual output is far from the truth value, even small changes to the activation can make a large difference in the cost. Therefore, Equation 4.16 tells us how a change of one particular weight in the last layer will affect the loss of the NN.

To get to the previous layer, we use:

$$\frac{\partial\, l_i(\Omega_t)}{\partial\, a_k^{(l-1)}} = \sum_{j=1}^{n^{(l)}} \underbrace{\frac{\partial\, z_j^{(l)}}{\partial\, a_k^{(l-1)}}}_{w_{jk}^{(l)}} \frac{\partial\, a_j^{(l)}}{\partial\, z_j^{(l)}} \frac{\partial\, l_i}{\partial\, a_j^{(l)}}\,, \tag{4.17}$$

where the new term with the derivative with respect to $a_k^{(l-1)}$ is simply the weight $w_{jk}^{(l)}$. When the value of a neuron in the second-to-last layer is changed, the effect on $z^{(l)}$ is proportional to the weight. To have the derivative of $a_k^{(l-1)}$, we must sum over all the nodes j of the layer (l).

By extending the chain rules to the weight of the layer $(l-1)$ and benefiting from Equation 4.17, we get:

$$\begin{aligned}
\frac{\partial\, l_i(\Omega_t)}{\partial\, w_{kj}^{(l-1)}} &= \overbrace{\frac{\partial\, z_k^{(l-1)}}{\partial\, w_{kj}^{(l-1)}}}^{a_j^{(l-2)}} \frac{\partial\, a_k^{(l-1)}}{\partial\, z_k^{(l-1)}} \frac{\partial\, l_i(\Omega_t)}{\partial\, a_k^{(l-1)}}\,, \\
\frac{\partial\, l_i(\Omega_t)}{\partial\, b_k^{(l-1)}} &= \frac{\partial\, z_k^{(l-1)}}{\partial\, b_k^{(l-1)}} \frac{\partial\, a_k^{(l-1)}}{\partial\, z_k^{(l-1)}} \frac{\partial\, l_i(\Omega_t)}{\partial\, a_k^{(l-1)}}\,.
\end{aligned} \tag{4.18}$$

This is where the back-propagation algorithm is particularly efficient. Instead of computing the derivative each time for each weight, the software needs to first compute the derivative for every weight in the last layer. Saving them in memory allows it to reuse them for the second-to-last layer and so on until the algorithm passes through the whole network. The only operations are addition and multiplication, with the weight matrix $\mathbf{W}$ being the most demanding.

This recursive formulation illustrates the efficiency of the back-propagation algorithm. Rather than recomputing all derivatives independently, the algorithm computes the derivative of the loss for each parameter in the output layer first and then reuses them to calculate the gradients for each preceding layer. Each step only involves matrix multiplications and element-wise operations, making the process highly efficient and scalable.

The initialisation of the weights at the creation of the NN can be changed by the user. The default choice is called the *Glorot Uniform* or *Xavier Uniform* [197]:

$$\begin{aligned}
w &= \mathcal{U}\left(-\sqrt{\frac{6}{n_{\text{in}}+n_{\text{out}}}}, \sqrt{\frac{6}{n_{\text{in}}+n_{\text{out}}}}\right)\,, \\
b &= 0\,,
\end{aligned} \tag{4.19}$$

where the weights are drawn from a uniform distribution, with n_{in} and n_{out} being the number of input and output nodes of the layer. The biases are initialised to zero.

The last missing piece to complete an NN is the activation function.

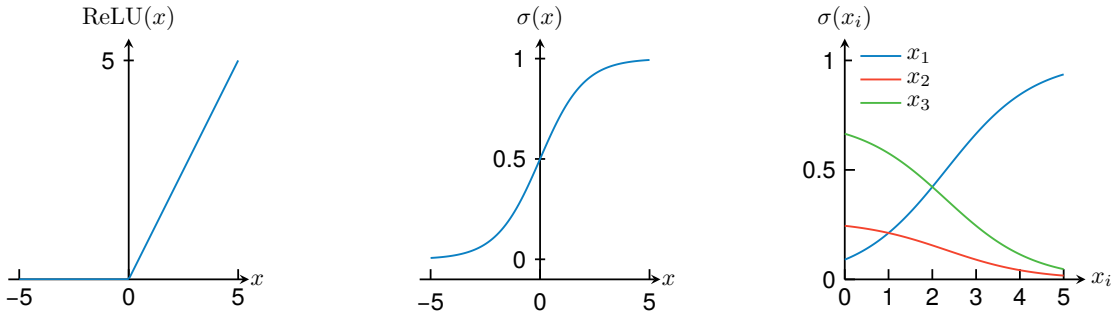
4.1.6 Activation function

The activation function ϕ is present in every layer and must follow various properties explained previously:

- non-linearity: this allows the NN to approximate functions following the Universal Approximation Theorem
- scaling: depending on the choice of the activation, this allows the values inside the network or even the output to be clipped within a certain range
- differentiability: this is needed to perform back-propagation.

In addition to these major properties, depending on the choices, it can help to reduce the computational complexity for the derivative, or, with a monotonic function, it helps the gradient to converge more easily. In Figure 4.4, several examples of activation functions are presented.

$$\mathrm{ReLU}(x) = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases} \qquad \sigma(x) = \frac{1}{1+\mathrm{e}^{-x}} \qquad \sigma(x_i) = \frac{\mathrm{e}^{x_i}}{\sum_{c=1}^{C} \mathrm{e}^{x_c}}$$



(a) ReLU **(b)** Sigmoid **(c)** Softmax

Figure 4.4: Examples of activation functions.

An overall good choice is the rectified linear unit (ReLU) [198] activation function defined in Figure 4.4(a). It rescales any input to $[0, +\infty)$. It is differentiable everywhere except at $x = 0$. It has the advantage of being computationally inexpensive, but also helps to reduce the vanishing gradient problem, which occurs when the gradient shrinks towards zero, consequently slowing down the learning. The disadvantage is that it can lead to "dead neurons", meaning that some neurons can get stuck at zero. This function is generally a good choice for the hidden layers.

The sigmoid function defined in Figure 4.4(b) is a strong choice to use for an output layer with only one output. It maps this output to represent a probability by scaling it between zero and one, making it the preferred choice in binary classification. The disadvantage is that it can lead to vanishing gradients. Using it only on the output layer reduces the impact of this issue. A generalisation of this function for multi-class classification is the softmax defined in Figure 4.4(c). It takes into account the number of classes C and scales each output between zero and one, making each of them the probability of having one class. It presents the same drawbacks as the sigmoid.

With all the elements that compose an NN now in place, we turn to how it is used in practice.

4.1.7 Training

Consider a perfectly defined NN with its set of hyperparameters. This corresponds to every parameter the user can modify:

- number of layers
- number of neurons per layer
- activation functions
- weight initialisation
- batch size
- gradient optimiser (momentum, batch)
- learning rate and scheduler.

We want to train it to make its predictions as accurate as possible. For that, we need a dataset made of input parameters, each set associated with a truth value for the network to predict. We split the dataset first into two parts: the training dataset and the validation dataset. The first split corresponds to the data where the NN is trained, while the second part is used to monitor the training on events the network is not trained on. Because the validation set steers choices such as early stopping, it is not fully independent. When possible, the final performance is quoted on a third, test dataset that the network has never seen. The training/validation split is illustrated in Figure 4.7. A usual split is $80\,\%$ for training and $20\,\%$ for validation.

Since the inputs can be of various types and differ by orders of magnitude, rescaling is often necessary. It improves the gradient stability and speeds up convergence. Different choices exist, but a simple one is the Min-Max normalisation given by:

$$x_i' = \frac{x_i - \min x_i}{\max x_i - \min x_i} \,. \tag{4.20}$$

This type of normalisation has the advantage of not distorting the shape of the input data and only scaling it down to $[0, 1]$. It is a good range when paired with the ReLU activation function. This scaler is first computed on the training dataset to avoid any leak from the statistics of the validation dataset. To evaluate the performance, it is later applied to the validation dataset.

As described in Section 4.1.4.1, the training dataset is subdivided into batches. We iterate over every batch and feed the NN (step Ⅰ in Figure 4.7). It gives us a prediction that allows us to compute the loss, steps Ⅱ and Ⅲ. Based on this, we back-propagate through the network to update the various weights, step Ⅳ. Once the full training dataset has been seen, an epoch is complete, and we start over again, step Ⅴ.

4.1.7.1 Overfitting

This process is done until we reach the global minimum or what seems to be it. For this, we monitor the loss per epoch. With a good setup of hyperparameters, both the losses of the training and validation datasets should decrease until they reach an optimum represented in Figure 4.5. This should correspond to the global minimum, meaning no further decrease of the loss. It can also correspond to a local minimum represented by a stagnation of the loss. In either case, the training should stop beyond this optimum point.

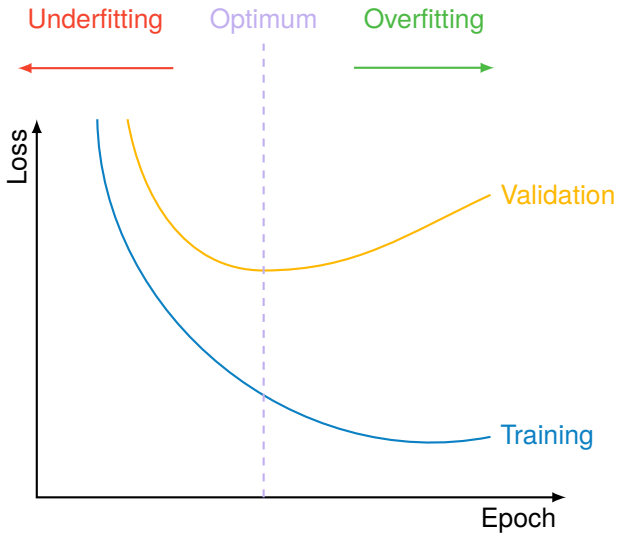


Figure 4.5: Evolution of loss during training and overfitting representation.

If the training is continued, we fall into overfitting. The NN is trained to predict only the training dataset, forgetting the validation dataset, increasing the loss as seen in Figure 4.5. The over-trained NN is effective only on the data it was trained on, unable to adapt to unseen data, making it useless.

Choosing the number of epochs to train the NN is also a hyperparameter. To make it easier, an early stopper algorithm [199] can monitor the loss and cut short the training whenever the loss stops evolving for a certain number of epochs. After it stops, it restores the weights before the stagnation. This is a robust way to prevent overfitting.

4.1.7.2 k-folds

Another technique that can indirectly prevent overfitting bias is to use a k-fold split of the dataset. Previously, we mentioned splitting the dataset with $80\,\%$ for training and $20\,\%$ for validation. We therefore lose this $20\,\%$ to improve the NN, which can be a huge loss when statistics are limited in the dataset. To remedy this, the k-fold splitting uses k different classifiers, each time with a different preparation of the splitting. In Figure 4.6, a five-fold procedure is shown with five classifiers and an $80\,\%$–$20\,\%$ split. At the end, every entry of the dataset has been used at least once for both training and validation. We can also use all the predictions made by the various NNs on the validation dataset to have a whole dataset prediction with no overfitting bias since none of them were used for training in their corresponding splits. The drawback of this method is that it requires multiple NNs, making it more complex to implement.

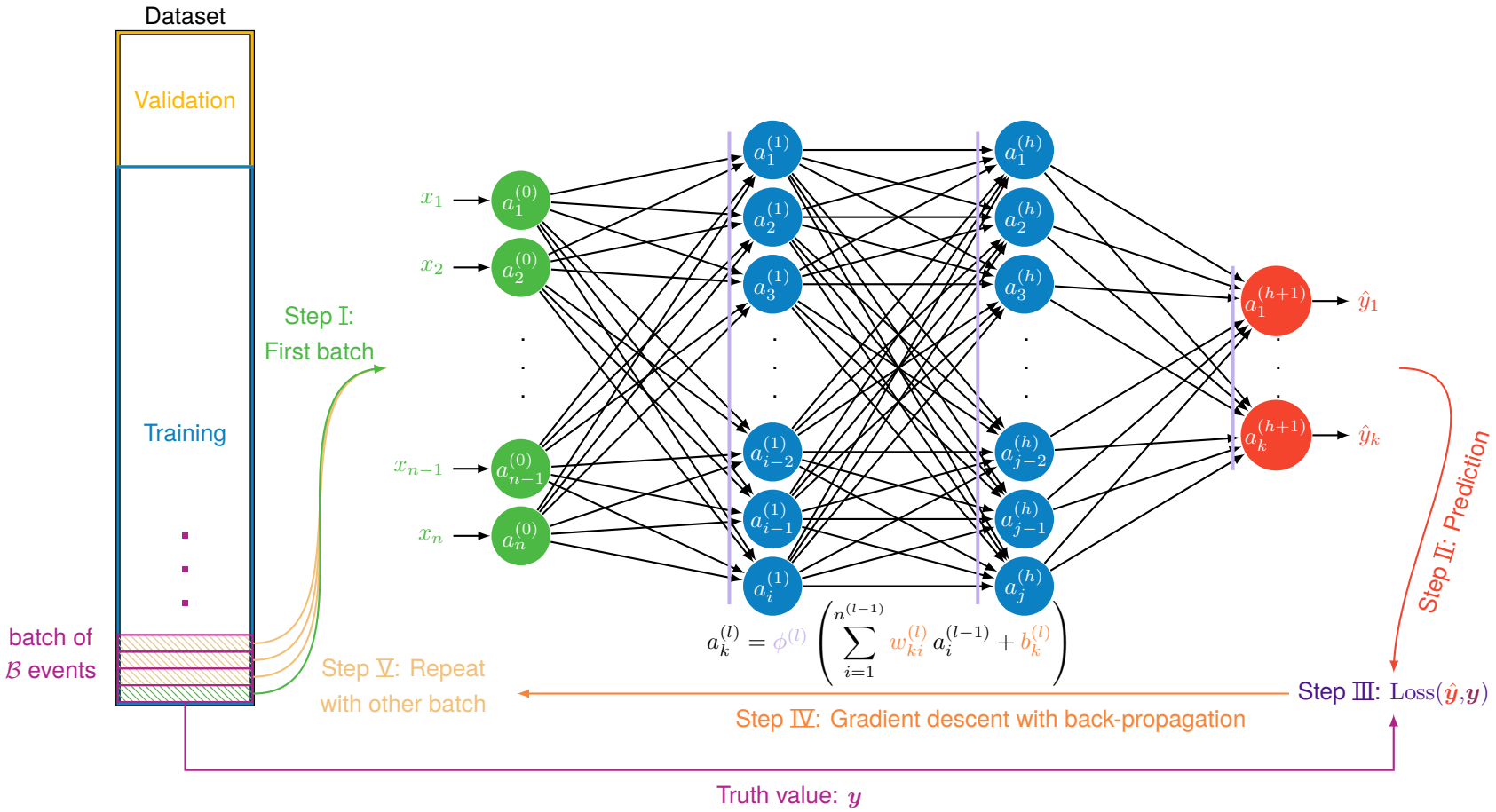


Figure 4.7: Training step of an NN.

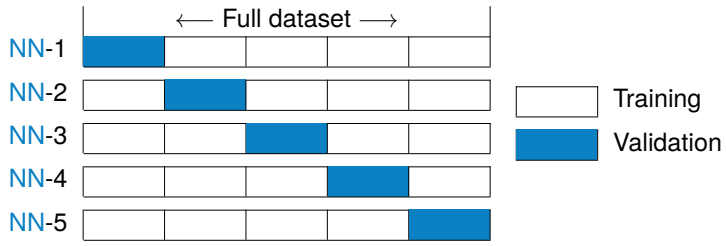


Figure 4.6: Representation of a k-fold splitting.

4.1.7.3 Imbalanced dataset

It can happen that the dataset on which the NN will be trained will be imbalanced, meaning one class is overrepresented or underrepresented and thus not reflecting reality. There are multiple ways to deal with this case. The loss can be reweighted either with class weights, one weight per class, or with sample weights, one weight per sample of the dataset. The loss used is thus given by:

$$L_{\text{weighted}} = \frac{\sum_{i \in \mathcal{B}_t} \omega_i \cdot l_i}{\sum_{i \in \mathcal{B}_t} \omega_i} , \tag{4.21}$$

where the sample or class weight is described by ω.

Another technique is to modify the loss function using focal loss instead. The term "focal" means that the loss will penalise errors on samples from the underrepresented class more heavily, while reducing the impact of errors on the overrepresented ones. This technique is somewhat harder to implement since the penalisation factor requires some fine-tuning of this new hyperparameter.

4.1.7.4 Metrics

During the training, various metrics are monitored. Among the most popular are the accuracy, Equation 4.22, the recall, Equation 4.23, and the F1 score, Equation 4.24. They are based on the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) defined in the confusion matrix represented in Table 4.A. They allow us to evaluate the performance of the NN. They are monitored during and after the training.

Another key metric to monitor is the receiver operating characteristic (ROC) plot and its corresponding area under curve (AUC), which is shown in Figure 4.8. It represents the signal efficiency versus the background rejection. It shows different thresholds for which the NN can discriminate a certain amount of signal while rejecting a certain quantity of background. A random classifier being no better than tossing a coin will be characterised by the dashed line and an AUC of 0.5. The better the discriminant, the closer it should get to the upper-right corner $(1, 1)$, resulting in an AUC closer to one.

Table 4.A: Confusion matrix.

		Truth	
		Positive	**Negative**
Predicted	**Positive**	TP	FP
	Negative	FN	TN

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}} , \tag{4.22}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} , \tag{4.23}$$

$$\text{F1} = \frac{\text{TP}}{\text{TP} + \frac{1}{2}(\text{FP} + \text{FN})} . \tag{4.24}$$

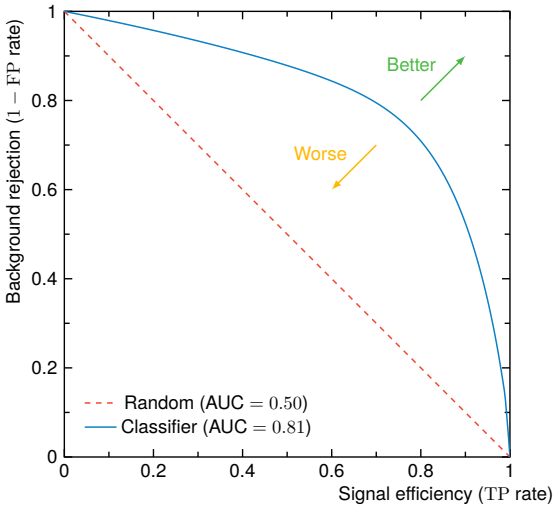


Figure 4.8: Example of an ROC curve and corresponding AUC.

4.1.8 Regularisation

One last key step in the NN structure is its regularisation. These procedures help to prevent overfitting, improve the generalisation, and stabilise the training.

One that acts directly on the loss is the weight regularisation, which adds a penalty to the loss proportional to the weights of the NN, giving the ℓ_1 and ℓ_2 regularisation, respectively:

$$L_{\text{total}} = L_{\text{data}} + \lambda \sum_i |\Omega_i| \,, \tag{4.25}$$

$$L_{\text{total}} = L_{\text{data}} + \lambda \sum_i \Omega_i^2 \,. \tag{4.26}$$

The first one encourages sparse weights, while the second one encourages smaller weights without setting them to 0. It is also possible to combine both of them. The parameter λ dictates the degree of penalisation.

Adding the term $-\alpha\lambda \mathbf{W}_t$ for the weights w only to Equation 4.10 or Equation 4.15 is called *weight decay* [200]. For the SGD case only, it is equivalent to ℓ_2 regularisation. This regularisation term will act directly on the weight and not on the loss. It is often preferred over the two other techniques for its finer control independent of the loss derivative. This improves the convergence stability for a deeper network.

Another one acts on the forward pass of the NN. We add dropout layers [201] that will randomly set a fraction (defined by the user) of $a_k^{(l)}$ to 0 during the training only. This helps the network to learn redundant representations and prevent it from relying too heavily on some specific nodes.

Other layers, such as batch normalisation [202], help too. At each batch injection, they normalise the node to have zero mean and unit variance.

The previously discussed early stopper and learning rate scheduling are part of regularisation techniques. The architecture also plays a role, for example, by having a smaller model, reducing the number of parameters.

A final technique is called *label smoothing* [203], which softens the target labels by a small percentage to prevent overconfidence and smooth the decision boundaries.

4.1.9 Architecture

This section has covered all the constituents of NNs. By varying the hyperparameters, many applications become possible, making NNs highly versatile tools whose use has grown steadily across domains. The following non-exhaustive list presents some important NN architectures and their applications:

- DNN: it corresponds to a simple NN with many layers, making it deep. It is used for regression, classification, or feature extraction tasks
- convolutional neural network (CNN) [204]: based on convolutional layers for grid-like data. It is used for image classification or video analysis
- recurrent neural network (RNN) [205]: designed for sequential data, it maintains a hidden state that carries information across time steps. It is used for time series prediction, language modelling, and speech recognition
- Transformer [206]: unlike RNNs, it is designed for sequential data like text and, with its attention mechanism, can process in parallel different parts of the text, learning relationships between so-called *tokens* (words) in a sequence. It is used for text generation and translation. It is also the predominant component of the large language models (LLMs) widely used nowadays
- AutoEncoder (AE) [207]: it transforms the input data from their original dimension space to a reduced space called *latent space*, a compact representation that captures the main features of the data. The network learns correlations and specific features in this reduced space. It is used for dimensionality reduction, data compression, or denoising
- generative adversarial network (GAN) [208]: it combines two distinct NNs, the generator that has to create fake data, while the discriminator detects real versus fake data. They are trained simultaneously, with one trying to fool the other. It is used for image and video generation
- graph neural network (GNN) [209]: instead of using a sequential structure, it operates on graph-structured data (nodes and edges). It aggregates feature information from neighbouring nodes to learn node or graph representations. It is used, for example, in molecular property prediction but also in social network analysis.

Complete details of every architecture, as well as further explanation of the formalism introduced in this section, are available in [210].

This variety of NN architectures, each facing different problems, makes them a robust choice when facing ML tasks.

The first mention of NNs in the context of high-energy physics can be traced back to 1988 [211, 212] where recurrent neural networks were used for track reconstruction. Meanwhile, the first use for an analysis was in 1992 [213], done by DELPHI, where a feed-forward NN was applied to study the decay of the Z into b and c. Later, in 1997, hardware-based NNs [214] were used, one of the first times in an experiment, in the H1 experiment at the Hadron-Elektron-Ringanlage (HERA) collider at DESY. The same technologies were considered for the prototype in the CMS L1 trigger. Despite being promising, and despite the physics community's eagerness to try these new technologies, adoption had a slow start, as other techniques were still outperforming them.

Although NNs were conceived long ago, major upgrades in the following decades have made them the first choice in terms of ML algorithms. NN usage has grown steadily, driven by their impressive discriminating power, the evolution of hardware and graphics processing units, and the availability of large training datasets. This shift is visible in research, and in particular in high-energy physics. Here, NNs have started to take over older technologies such as boosted decision trees (BDTs) that were the predominant choices for physics analysis until a few years ago. The first CNN using images of jets left in the calorimeter was used in 2016 [215], while in 2018, a GNN was used for track reconstruction [216].

Further examples will be given in the next chapters, on how NNs are used more and more in the various stages of ATLAS physics analysis, from reconstruction to final analysis. A key ML tool used prior to NNs is the BDT, which is still widely used and will be described in the next section.

The NNs used in this thesis are built with Keras [217] on TensorFlow [218], using NumPy [219] for array manipulation and Matplotlib [220] for plots.

4.2 Boosted decision tree

The BDT is a subcategory of DTs paired with boosting methods. We will first see the principles of DTs, then how these methods benefit from boosting.

4.2.1 Decision tree

A DT [221] is another ML tool mostly used for classification (regression is possible but not explained here). This type of discriminant is cut-based: it applies a sequence of binary splits on the events forming decision nodes. At each of them, the splitting is done based on the value of a given variable as represented in Figure 4.9(a). The starting point of the tree is called the *root node*, while a terminal node is referred to as a *leaf*. A DT is characterised by its depth, the maximal number of nodes from the root to leaves, and its total number of nodes and leaves. The whole set of leaves partitions the full phase-space as shown in Figure 4.9(b).

Generally, the classification is binary, therefore differentiating the signal, S, from the background, B. The DT gives either the signal purity, defined as:

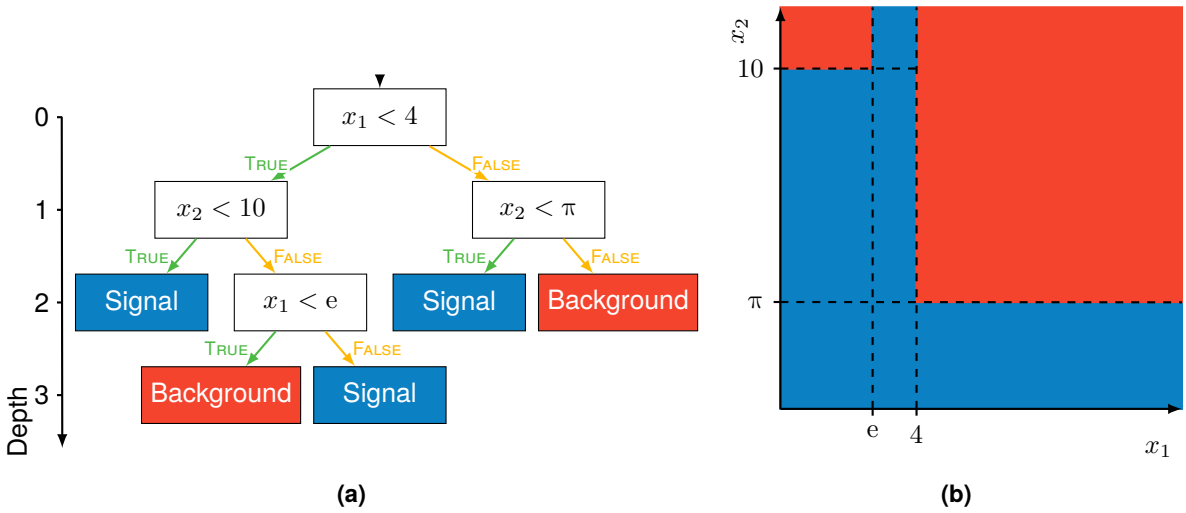
$$P = \frac{S}{S+B}\,, \tag{4.27}$$

with a value close to one for signal and zero for background, or another continuous discriminant value, correlated with the signal probability, where higher values correspond to more signal-like events.

Before training, event weights are typically initialised equally. The training dataset is balanced so that signal and background have similar overall contributions, avoiding bias in the discriminant. Also, no preprocessing like scaling down the variables is needed since every split is done on one variable at a time.

Then we need to define a metric to judge the goodness of the split as well as a terminal condition. Regarding the terminal condition, it can be chosen according to various criteria:

- on the minimum size this leaf should have for statistical significance
- insufficient improvement from further splitting
- perfect classification with every event falling within one single class
- maximal tree depth (computing concerns).



(a) (b)

Figure 4.9: Scheme of a DT (a) and the corresponding phase-space representation following its cuts (b).

To decide a separation criterion, we have to minimise an impurity function. The impurity of a node is maximal for an equal mix of signal and background, and minimised when the node contains only signal or only background events.

This function can be either the cross entropy (CE), similar to the loss function used in NN and given by:

$$I_{\mathrm{CE}} = -\sum_{i\in\{S,B\}} p_i \ln p_i \,, \tag{4.28}$$

where:

$$p_S = \frac{S}{S+B}\,, \qquad p_B = \frac{B}{S+B}\,, \tag{4.29}$$

or the Gini index given by:

$$I_{\mathrm{Gini}} = 1 - \sum_{i\in\{S,B\}} p_i^2 = \frac{2S\cdot B}{(S+B)^2}\,. \tag{4.30}$$

The last one performs similarly to the cross entropy one but is a more popular choice.

The separation $\mathcal{S}$ represents the reduction in impurity achieved by a split, weighted by the number of events ω in the parent and child (left and right, respectively) nodes and giving us:

$$\mathcal{S} = \omega_{\mathrm{parent}}\cdot I_{\mathrm{parent}} - \omega_{\mathrm{left}}\cdot I_{\mathrm{left}} - \omega_{\mathrm{right}}\cdot I_{\mathrm{right}}\,. \tag{4.31}$$

The final value of the separation of the DT is the sum of $\mathcal{S}$ over all the nodes.

The DT has the advantages, as we have seen, of needing minimal or even no data preprocessing. The final results are straightforward to interpret. This method is not a black box: the decision process is transparent, with no hidden intermediate steps like those found in hidden layers in NNs. However, small changes from one training to another can induce large variations in the structure of the tree. It is also prone to overfitting. One way to compensate for this is by using boosting techniques.

4.2.2 Boosting

Boosting is a general method used in ML. It arises from the fact that it is hard to make a very good discriminant but relatively easy to make simple, error-prone ones. The goal of boosting is to combine such weak classifiers into a new, more stable one with a smaller error. In the case of BDT, several trees are successively built, referred to as a *forest*. The output of one tree is used to weight the training dataset of events, which is then given as an input for the training of the next DT. This procedure is repeated for K trees, chosen before the training. The forest's trees are weak classifiers as they have smaller hyperparameter values, such as the number of nodes and depth. Therefore, their discrimination is not sophisticated. At the end of the K steps, an average classifier is built from the combination of the whole forest, forming a more complex and powerful DT.

The task of classification can be viewed as a minimisation problem. The objective is to minimise a loss function similarly to the NN case. The BDT can be viewed as a sum over every DT:

$$\mathrm{BDT}(\boldsymbol{x}) = \sum_{k=1}^{K} \beta_k \cdot \mathrm{DT}(\boldsymbol{x}, \Omega_k)\,, \tag{4.32}$$

depending on their sets of parameters Ω, and associated with a combination of weights β.

At every step of the boosting procedure, we increment the sum over the K steps by one. We therefore want to optimise Ω_k to minimise the loss. Similarly, as explained in Section 4.1.4, the gradient will be used. Indeed, it gives the direction and magnitude of the main changes in a function. By following this direction, we get to the minimum (local or global).

A common choice for binary classification with BDT is the binomial log-likelihood (BLL) given by:

$$L_{\mathrm{BLL}} = \ln\left(1 + \mathrm{e}^{-2y\hat{y}}\right) , \tag{4.33}$$

with the label encoded as $y = \pm 1$.

In the case of BDT, the steps are the following:

- a newly trained DT is fixed
- we compute its gradient of the loss, referred to as *residuals*
- the new DT is trained to learn this residual contribution
- it is added to the sequence in Equation 4.32 with a weight given by the learning rate.

This whole procedure is called *gradient boosting* and is represented by Figure 4.10. Other techniques exist for DT, such as Bootstrap AGGregatING (BAGGING) or random forest, but will not be treated here.

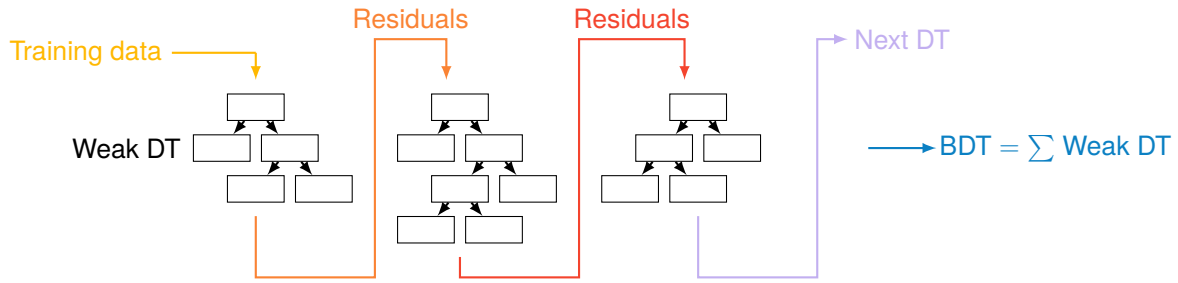


Figure 4.10: Scheme of a boosting procedure for BDT.

4.2.3 Application

DTs are a relatively old technology, developed in the 1980s. They underwent several upgrades, making them more robust. As they are relatively easy to implement in code and straightforward to interpret, they were a strong choice at the beginning of the XXI century for high-energy physics research, leading, for example, to the first observation of single top quark production [222, 223] at the Fermilab Tevatron Collider by the DØ and CDF collaborations. They remain a key algorithm class to this day and are still widely used in analyses of the LHC experiments, such as ATLAS. However, a few years ago, the widespread adoption of NNs in every field led to a turning point in the choice of ML algorithm in favour of the latter.

The BDTs used in this thesis are built with Toolkit for MultiVariate data Analysis (TMVA) [224] based on the ROOT data-analysis framework [225].

4.3 Clustering

The two previous ML methods presented are known as *supervised learning*. It means that for each vector of input x there is an associated label giving its true output class. Often, this is not accessible in the dataset, and we fall into *unsupervised learning*. The machine has to learn by itself by finding patterns, correlations, or similarities between the data. One major constituent of this type of ML is the clustering algorithm.

A cluster is not precisely defined as a general notion and will depend on which type of clustering algorithm we speak of. In this thesis, we will focus only on the k-means algorithm.

4.3.1 k-means

The k-means clustering algorithm [226, 227] aims to partition a dataset into k clusters. Each element $\boldsymbol{x}$ will belong to one of the clusters C with the nearest mean $\boldsymbol{\mu}$:

$$\boldsymbol{\mu} = \frac{1}{|C|} \sum_{\boldsymbol{x} \in C} \boldsymbol{x} \, . \tag{4.34}$$

This is computed for one cluster and called the *cluster centre* or *cluster centroid*. The algorithm minimises a criterion known as the *inertia* or the within-cluster sum-of-squares (equivalent to squared Euclidean distances) defined as:

$$\text{Inertia} = \sum_{i=1}^{k} \sum_{\boldsymbol{x} \in C_i} \|\boldsymbol{x} - \boldsymbol{\mu}_i\|^2 . \tag{4.35}$$

Throughout this thesis the distance metric is the Euclidean one, the ℓ_2 norm between two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, summed over their components j:

$$d(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|_2 = \sqrt{\sum_j (x_j - y_j)^2} , \tag{4.36}$$

so that the inertia of Equation 4.35 is a sum of its squares.

This metric can be viewed as a measure of how internally coherent the clusters are. The number of clusters has to be specified beforehand and has to be lower than or equal to the number of elements D. Also, the centres are not necessarily points from the dataset, even if they lie within the same phase-space.

The algorithm itself works in two steps:

- Assignment step: each element of the dataset is assigned to the nearest centroid following a Euclidean distance
- Update step: each centroid is recomputed for every element assigned to the cluster by taking the mean value of every entry inside it.

These steps are repeated until no changes in the centroids are observed. The algorithm will always converge with enough time, but not necessarily to the global minimum.

This algorithm is simple to implement. However, it suffers from some drawbacks. It assumes the clusters to be isotropic and of similar size. Other clustering algorithms also tend to face this assumption. It is also computationally heavy with a large dataset, as it needs to recompute the distances between all points and centroids at each step. Batch methods can be used in this case.

The k-means algorithms used in this thesis are built with scikit-learn [228].

4.3.2 Voronoi diagram

A Voronoi diagram [229, 230] partitions a phase-space into regions, known as *Voronoi cells*. Each cell is associated with a unique point, called a *seed* (or *site*). A Voronoi cell contains all points in the space that are closer to its corresponding seed than to any other, according to a chosen distance metric, here the Euclidean distance of Equation 4.36. Another metric, such as the Manhattan distance, would move the boundaries of the cells and partition the same seeds differently.

A simple illustration can be built from a map, for instance, the main site of CERN. Considering several points of interest as seeds, such as IP-1, Proton Synchrotron, LINAC-2, the Tier-0 data centre, Restaurants 1 and 2, and Building 40, we can determine, for any location on the map, which seed it is closest to. The resulting partitioning is shown in Figure 4.11(a), where each cell, outlined in white, corresponds to one seed (white cross).

Voronoi diagrams are intimately related to clustering algorithms such as k-means: at each iteration, the assignment step is precisely the Voronoi partition induced by the current centroids. In the analogy above, every data point $\boldsymbol{x}$ is an individual standing somewhere on the map, and every centroid a point of interest, as illustrated in Figure 4.11(b). The two steps of the algorithm are detailed in Section 4.3.1.

The Voronoi diagrams used in this thesis are built with SciPy [231].

4.4 Variable ranking and pruning

The input variables to an ML model are often numerous, and not all of them contribute equally to its performance. Some may be redundant, carry little discriminating power, or even be detrimental to the training. Retaining too many inputs increases the number of free parameters, and therefore the risk of overfitting (Figure 4.5), the sensitivity to statistical fluctuations, and the training time. It is thus important to rank the candidate variables by their relevance and to keep only a compact subset that preserves the discriminating power while limiting the complexity of the model.

There is no unique way to rank variables, and different definitions of importance can yield different orderings, each with its own strengths and blind spots. To mitigate the bias introduced by any single definition, three complementary methods are used and their results cross-checked. The first two, SHapley



(a) (b)

Figure 4.11: Map of CERN partitioned into Voronoi cells around a set of points of interest (a), and the same partition seen as the outcome of a clustering algorithm (b).

Additive exPlanations (SHAP) values and permutation importance, are model-agnostic and apply to any ML model, including NNs and BDTs. The third is specific to differentiable models such as NNs. Throughout, the model is the function f introduced in Equation 4.4, mapping an input vector $\boldsymbol{x}$ to the prediction $\hat{\boldsymbol{y}}$.

SHAP values [232]. A method borrowed from game theory that shares the model prediction out among the input features according to how much each one contributes. Write $f(S)$ for the prediction the model would make using only the subset S of the features. Switching on one further feature j then changes the prediction by its *marginal contribution*

$$\Delta_j(S) = f(S \cup \{j\}) - f(S)\,, \tag{4.37}$$

where $S \cup \{j\}$ is the set S with feature j added. This change depends on the context: a variable can look decisive when few others are present, yet turn out redundant once its correlated partners are already switched on. To remove this ambiguity, one averages the marginal contribution over every order in which the features could be switched on. For a given order, S is simply the set of features coming before j. This average defines the *Shapley value* of feature j:

$$\phi_j = \big\langle\, \Delta_j(S) \,\big\rangle_{\text{orderings}}\,, \tag{4.38}$$

where $\langle\,\cdot\,\rangle_{\text{orderings}}$ denotes the average over all $|\mathcal{F}|!$ orderings of the full feature set $\mathcal{F}$. The importance of a variable is then taken as its mean absolute Shapley value over the training dataset. The construction is additive: the Shapley values of all features sum to the difference between the prediction and the value the model returns with no feature switched on, and account naturally for correlations and interactions between inputs. Its cost grows with the number of feature orderings, so in practice ϕ_j is not computed exactly but estimated.

Permutation importance [233]. A model-agnostic measure obtained by randomly shuffling the values of a single feature across the events of the validation dataset, thereby breaking its association with the target while preserving its marginal distribution. The importance is defined as the resulting increase in the loss L (Section 4.1.3), or equivalently the decrease in a performance metric such as the AUC of the ROC curve (Figure 4.8):

$$I_j = \frac{1}{K}\sum_{k=1}^{K}\Big[\, L\big(\boldsymbol{y}, f(\boldsymbol{x}^{\pi_k^{(j)}})\big) - L\big(\boldsymbol{y}, f(\boldsymbol{x})\big)\,\Big]\,, \tag{4.39}$$

averaged over K independent permutations $\pi_k^{(j)}$ to suppress statistical fluctuations. A large increase indicates that the model relies strongly on the permuted feature. The method must be used with care

when inputs are correlated, as permuting one feature in isolation can probe unphysical configurations and dilute the apparent importance of correlated variables, one of the motivations for cross-checking it against the other two methods.

Gradient-based importance [234]. A measure of the sensitivity of the model output to each input, specific to differentiable models such as NNs and obtained from the same gradients computed during back-propagation (Equation 4.16), here taken with respect to the inputs rather than the parameters. The importance of a feature is defined as the mean absolute gradient of the output with respect to that input over the normalised events (Equation 4.20) of a representative sample:

$$I_j = \frac{1}{D}\sum_{d=1}^{D}\left|\frac{\partial f(\boldsymbol{x}_d)}{\partial x_{d,j}}\right|, \tag{4.40}$$

where D is the number of events in the sample. Inputs to which the output is most sensitive receive the largest scores. For a multi-output network the gradient is taken with respect to the relevant class score. Being a direct by-product of the network gradients, this estimator is inexpensive to evaluate, but it probes only the local sensitivity around each event.

The three methods each yield an ordered list of the candidate inputs. These rankings are combined into a single consensus ordering, for instance by averaging the per-variable normalised scores, which limits the influence of the blind spots of any individual method. Starting from the least important variable, inputs are then removed iteratively and the model retrained, its performance being monitored at each step through the loss (Section 4.1.3) and the AUC (Figure 4.8) evaluated on the validation dataset (Figure 4.7). A variable is discarded only when its removal does not significantly degrade the performance, and the procedure is stopped once any further removal would incur a measurable loss. This yields a compact and robust set of inputs for a given model. This is referred to as *pruning* of the input space, and it is repeated for each model architecture under consideration, as the importance of a variable can depend on the model type and its hyperparameters.

4.5 Bayesian optimisation

As discussed in Section 4.1.7, the training of an NN involves the optimisation of a loss function over a high-dimensional parameter space. The performance of the network is highly sensitive to the choice of hyperparameters, such as learning rate, batch size, number of layers, and number of nodes per layer. Finding the optimal set of hyperparameters is a challenging, non-trivial optimisation problem, often requiring extensive experimentation and computational resources, as the loss function is typically non-convex, noisy, and expensive to evaluate. Traditional methods such as grid search or random search can be inefficient and may not scale well with the number of hyperparameters [235].

Bayesian optimisation (BO) [236] is a probabilistic, model-based strategy tailored to this regime of expensive black-box objectives. It treats the objective as an unknown function and builds a cheap-to-evaluate surrogate model, a Gaussian process (GP) here, that provides both a prediction and a calibrated estimate of its own uncertainty. This uncertainty is then exploited to decide, in a principled way, where to evaluate next, so that an optimum is located in as few evaluations as possible.

Let $f(\boldsymbol{x})$ denote the objective to be *maximised*, where $\boldsymbol{x}$ is a configuration of hyperparameters and f quantifies the resulting model performance (for instance the negative validation loss or a validation metric). Minimising a loss is equivalent to maximising its negative, so the maximisation convention adopted here involves no loss of generality and matches the illustrative example of Figure 4.12. A surrogate model approximates f from a set of past evaluations, and an acquisition function $\alpha(\boldsymbol{x})$ assigns to every candidate $\boldsymbol{x}$ a utility quantifying how promising it is to evaluate there. The acquisition function encodes the trade-off between *exploration* (sampling regions of high predictive uncertainty) and *exploitation* (sampling regions the surrogate predicts to be good). Writing the surrogate's posterior predictive mean and standard deviation at $\boldsymbol{x}$ as $\mu(\boldsymbol{x})$ and $\sigma(\boldsymbol{x})$, so that $f(\boldsymbol{x})$ is treated as the random variable $f(\boldsymbol{x}) \sim \mathcal{N}\left(\mu(\boldsymbol{x}), \sigma^2(\boldsymbol{x})\right)$, common acquisition functions include:

- the expected improvement (EI) [237, 238], defined as:

$$\mathrm{EI}(\boldsymbol{x}) = \mathbb{E}\left[\max\left(0,\, f(\boldsymbol{x}) - f_{\mathrm{best}} - \xi\right)\right], \tag{4.41}$$

where the expectation is taken over the posterior of $f(\boldsymbol{x})$, f_{best} is the best objective value observed so far, and $\xi \geq 0$ is a small constant biasing the search towards exploration. For a Gaussian posterior this admits the closed form $\mathrm{EI}(\boldsymbol{x}) = (\mu(\boldsymbol{x}) - f_{\mathrm{best}} - \xi)\,\Phi(z) + \sigma(\boldsymbol{x})\,\phi(z)$, with $z = \frac{\mu(\boldsymbol{x}) - f_{\mathrm{best}} - \xi}{\sigma(\boldsymbol{x})}$ and Φ, ϕ the standard normal cumulative and probability density function [239].

- the probability of improvement (PI) [240], defined as:

$$\mathrm{PI}(\boldsymbol{x}) = \mathbb{P}\left(f(\boldsymbol{x}) > f_{\mathrm{best}} + \xi\right) = \Phi\left(\frac{\mu(\boldsymbol{x}) - f_{\mathrm{best}} - \xi}{\sigma(\boldsymbol{x})}\right), \tag{4.42}$$

  which favours points likely to improve on the current best by at least ξ.

- the upper confidence bound (UCB) [241], defined as:

$$\mathrm{UCB}(\boldsymbol{x}) = \mu(\boldsymbol{x}) + \kappa\,\sigma(\boldsymbol{x}), \tag{4.43}$$

  where $\kappa > 0$ controls the balance between exploration and exploitation: a large κ rewards predictive uncertainty and thus promotes exploration, while a small κ favours exploitation of the predicted mean.

The next configuration to evaluate is the maximiser of the chosen acquisition function, $\boldsymbol{x}_{\mathrm{next}} = \arg\max_{\boldsymbol{x}} \alpha(\boldsymbol{x})$. The procedure can be summarised as follows:

1. Initialise the surrogate model with a small number of hyperparameter configurations and their corresponding objective evaluations
2. Fit the surrogate model to the observed data, obtaining the posterior predictive mean and uncertainty over the hyperparameter space
3. Maximise the acquisition function to select the next configuration to evaluate, balancing exploration and exploitation
4. Evaluate the objective for the selected configuration and augment the surrogate model with the new observation
5. Repeat steps 2–4 until a stopping criterion is met, such as a maximum number of iterations or convergence of the best observed value.

The surrogate is modelled as a GP, a flexible non-parametric prior over functions that yields a Gaussian posterior in closed form and therefore supplies the predictive mean and uncertainty required by the acquisition function [242]. A GP is fully specified by a mean function $m(\boldsymbol{x})$, taken to be zero in the absence of prior knowledge, and a covariance (kernel) function $k(\boldsymbol{x}, \boldsymbol{x}')$ encoding the similarity between configurations:

$$f(\boldsymbol{x}) \sim \mathcal{GP}\left(m(\boldsymbol{x}),\, k(\boldsymbol{x}, \boldsymbol{x}')\right). \tag{4.44}$$

The kernel determines the smoothness and generalisation properties of the surrogate. For a d-dimensional hyperparameter space it is taken here to be:

$$k(\boldsymbol{x}, \boldsymbol{x}') = \sigma_f^2\, k_{\mathrm{Matérn}}(r) + \sigma_n^2\, \delta(\boldsymbol{x}, \boldsymbol{x}'), \tag{4.45}$$

where:

- σ_f^2 is the signal variance, setting the overall amplitude of the function's fluctuations
- $k_{\mathrm{Matérn}}(r)$ is the Matérn kernel, expressed as a function of the scaled distance r defined below. The $\nu = \frac{5}{2}$ member of the family is used:

$$k_{\mathrm{Matérn}}(r) = \left(1 + \sqrt{5}\,r + \frac{5}{3}\,r^2\right)\exp\left(-\sqrt{5}\,r\right), \tag{4.46}$$

  which yields functions that are twice differentiable, smooth enough to be well behaved, yet flexible enough for a wide range of objectives, making it a standard choice for modelling ML hyperparameter landscapes [242]. It is the special case of the general Matérn kernel:

$$k_{\mathrm{Matérn}}(r) = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\sqrt{2\nu}\,r\right)^{\nu} K_\nu\left(\sqrt{2\nu}\,r\right), \tag{4.47}$$

  where Γ is the gamma function, K_ν the modified Bessel function of the second kind, and the smoothness parameter $\nu > 0$ controls the differentiability of the sampled functions, larger ν giving smoother functions.

- σ_n^2 is the noise variance, accounting for the inherent noise in the objective evaluations, and $\delta(\boldsymbol{x}, \boldsymbol{x}')$ is the Kronecker delta, equal to 1 if $\boldsymbol{x} = \boldsymbol{x}'$ and 0 otherwise, so that the noise contributes only to the diagonal of the covariance matrix.

The scaled distance r is taken to be the anisotropic, automatic-relevance-determination distance:

$$r = \sqrt{\sum_{i=1}^{d} \frac{(x_i - x_i')^2}{l_i^2}}\,, \tag{4.48}$$

where l_i is a separate length scale for the i-th hyperparameter and the isotropic Euclidean case is recovered for $l_i = l$ for all i. Per-dimension length scales let the surrogate assign a different sensitivity to each hyperparameter, which is particularly useful when they have different scales or units, rather than treating all of them on an equal footing. The kernel hyperparameters $\{\sigma_f^2, \sigma_n^2, l_i\}$ are obtained from the data by maximising the log marginal likelihood:

$$\ln p(\boldsymbol{y}|\mathbf{X}) = -\frac{1}{2}\boldsymbol{y}^\intercal \mathbf{K}_y^{-1}\boldsymbol{y} - \frac{1}{2}\ln|\mathbf{K}_y| - \frac{n}{2}\ln 2\pi\,, \qquad \mathbf{K}_y = \mathbf{K} + \sigma_n^2 \mathbb{I}_n\,, \tag{4.49}$$

where $\mathbf{X}$ and $\boldsymbol{y}$ collect the n observed configurations and objective values and $\mathbf{K}_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$. Given the observations, the GP posterior at a test configuration $\boldsymbol{x}$ is itself Gaussian, with mean and variance defined as:

$$\mu(\boldsymbol{x}) = \boldsymbol{k}_*^\intercal \mathbf{K}_y^{-1}\boldsymbol{y}\,, \tag{4.50}$$

$$\sigma^2(\boldsymbol{x}) = k(\boldsymbol{x}, \boldsymbol{x}) - \boldsymbol{k}_*^\intercal \mathbf{K}_y^{-1}\boldsymbol{k}_*\,, \tag{4.51}$$

where $\boldsymbol{k}_* = (k(\boldsymbol{x}, \boldsymbol{x}_1), \ldots, k(\boldsymbol{x}, \boldsymbol{x}_n))^\intercal$. These are exactly the mean prediction and uncertainty band shown in Figure 4.12, and the quantities passed to the acquisition function.

In practice, several of the hyperparameters tuned here, the number of layers and of nodes per layer, are integer-valued, and are handled by optimising over a continuous relaxation and rounding to the nearest admissible value when the objective is evaluated. GP inference requires inverting the $n \times n$ matrix $\mathbf{K}_y$ and therefore scales as $\mathcal{O}(n^3)$ in the number of evaluations, with a further loss of efficiency in very high-dimensional spaces. Both are mild limitations here, since the cost of training the network keeps the number of evaluations n small, which is precisely the regime where BO is the most advantageous.

An example of the procedure is illustrated in Figure 4.12, where the objective is the test function $f(x) = \sin(2\pi x)$ on $x \in [0, 1]$. Starting from a single observation, the surrogate approximates the objective from the available data while the acquisition function guides the choice of the next point to evaluate. As evaluations accumulate, the posterior uncertainty collapses around the sampled region and the surrogate progressively locates the maximum near $x \approx 0.25$. The surrogate is a GP with a Matérn $\nu = \frac{5}{2}$ kernel, implemented with scikit-learn [228]. For clarity the kernel hyperparameters are held fixed in this illustration. The acquisition function is the UCB with $\kappa = 2$, providing a good balance between exploration and exploitation, evaluated from the posterior mean and standard deviation of the surrogate.

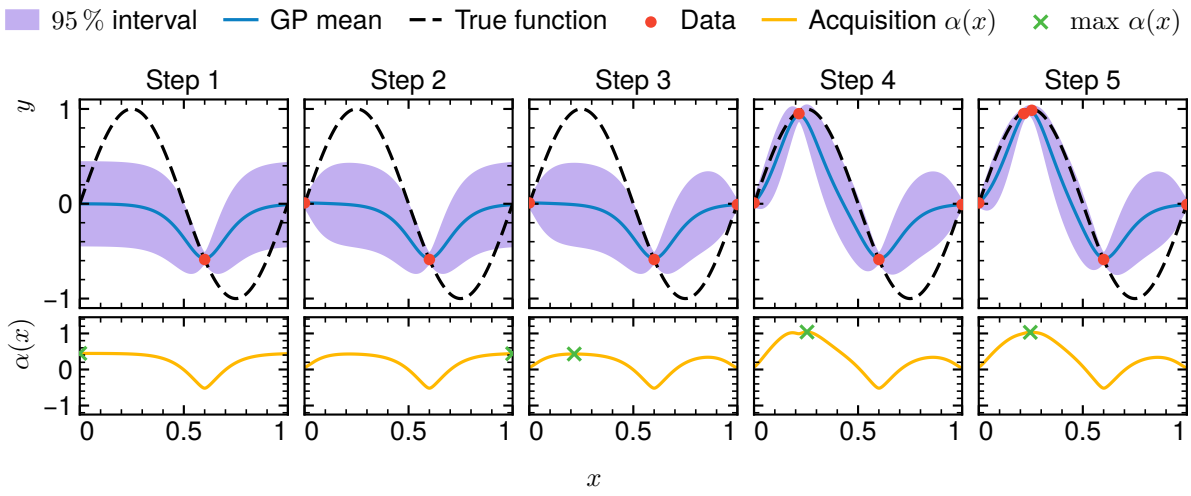


Figure 4.12: Illustration of the BO procedure on the test function $f(x) = \sin(2\pi x)$. The GP surrogate (mean prediction and 95 % confidence interval) approximates the true function from the observed data, and the UCB acquisition function $\alpha(x)$, with $\kappa = 2$, guides the selection of the next configuration to evaluate, marked by its maximum. As evaluations accumulate, the surrogate converges towards the maximum of the objective.

4.6 Monte Carlo

The MC method is a powerful algorithmic approach to compute approximated numerical values by using random processes. Its name comes from the Monte Carlo Casino in Monaco, reflecting the method's reliance on randomness and probability. Let us consider that we want to compute an expectation $\mathbb{E}[f(X)]$ where X is a random variable associated with a probability density function. By definition, we have:

$$\mathbb{E}[f(X)] = \int f(x)p(x)\mathrm{d}\,x\,. \tag{4.52}$$

An MC method will generate N random samples X_i from the distribution $p(x)$ and compute the average of $f(X_i)$:

$$\hat{\mu}_N = \frac{1}{N}\sum_{i=1}^{N} f(X_i)\,. \tag{4.53}$$

This estimator $\hat{\mu}_N$ is an unbiased estimator of $\mathbb{E}[f(X)]$, meaning that $\mathbb{E}[\hat{\mu}_N] = \mathbb{E}[f(X)]$. Moreover, by the strong law of large numbers, $\hat{\mu}_N$ converges almost surely to $\mathbb{E}[f(X)]$ as $N \to \infty$. Furthermore, the central limit theorem states that the distribution of $\hat{\mu}_N$ approaches a normal distribution with mean $\mathbb{E}[f(X)]$ and variance $\frac{\sigma^2}{N}$, where σ^2 is the variance of $f(X)$. Therefore, the error of the MC estimator decreases as $\frac{1}{\sqrt{N}}$, which is a key property of the method. The method is very robust at high dimensions, but the convergence is slow: to improve the error by a factor of 10, one needs to increase the number of samples by a factor of 100.

The most common use of MC methods is numerical integration. For instance, if we want to compute the integral I of a function $f(x)$ over a domain D:

$$I = \int_D f(x)\mathrm{d}\,x\,, \tag{4.54}$$

we can rewrite it as an expectation over a uniform random distribution over D:

$$I = \int_D f(x)\frac{1}{V}\mathrm{d}\,x = V\mathbb{E}[f(X)]\,, \tag{4.55}$$

where V is the volume of the domain D and X is a random variable uniformly distributed over D. Then, we can apply the MC method to estimate I by generating N random samples X_i uniformly distributed over D and computing:

$$\hat{I}_N = V\frac{1}{N}\sum_{i=1}^{N} f(X_i)\,. \tag{4.56}$$

Once again the error of the estimator $\hat{I}_N$ decreases as $\frac{1}{\sqrt{N}}$, and does not rely on the dimensionality of the problem, which makes it particularly useful for high-dimensional integrals.

A simple visualisation of how the MC method works for numerical integration can be done by trying to estimate the area of a quarter circle inscribed in a unit square. The area of the quarter circle is $\frac{\pi}{4}$, and the area of the unit square is 1. By generating random points uniformly distributed in the unit square and counting how many fall inside the quarter circle and thus satisfy the condition $x^2 + y^2 \leq 1$, we can estimate the area of the quarter circle as the ratio between the number of points inside the quarter circle and the total number of points, multiplied by the area of the unit square. We can then estimate π as four times the estimated area of the quarter circle. As we increase the number of random points, our estimate of π converges to the true value, demonstrating the power of the MC method for numerical integration.

When the integrand is strongly peaked, uniform sampling becomes inefficient because most samples land in regions that contribute little to the integral. *Importance sampling* addresses this by drawing samples from a chosen distribution $g(x)$ that mimics the shape of $|f(x)|$, and reweighting them by $\frac{f(X_i)}{g(X_i)}$. The variance of the estimator is minimised when $g(x) \propto |f(x)|$, in which case it would formally vanish. In practice, the optimal g is unknown, and adaptive algorithms iteratively refine a piecewise-constant approximation to $|f|$, while multi-channel integration combines several mappings tailored to the resonant or t-channel structures of the integrand. These techniques are at the heart of every modern MC generator and turn the otherwise impractical $\frac{1}{\sqrt{N}}$ convergence into a tractable computation for high-dimensional phase-spaces.

In the particular case where the direct sampling from the distribution $p(x)$ is difficult, we can use a technique called Markov chains. In Bayesian statistics, the posterior is given by:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}\,, \tag{4.57}$$

where y represents the observed data and θ represents the parameters of interest. The denominator $p(y)$, formulated as:

$$p(y) = \int p(y|\theta)p(\theta)\mathrm{d}\,\theta\,, \tag{4.58}$$

may be difficult to compute directly, especially in high-dimensional parameter spaces. The Markov condition is defined as:

$$p(X_{n+1}|X_n, X_{n-1}, \ldots, X_0) = p(X_{n+1}|X_n)\,, \tag{4.59}$$

which means that the next state X_{n+1} depends only on the current state X_n and not on the previous states. An analogous Markov assumption appears in the combinatorial Kalman filter (CKF) described in Section 5.1.2, where the prediction in layer $n+1$ uses only the state estimated at layer n, although the use there is filtering rather than generation. A Markov chain is therefore a sequence of random variables that satisfies this Markov condition. The idea is to construct a Markov chain whose stationary distribution is the target distribution $p(\theta|y)$. By simulating the Markov chain for a sufficiently long time, we can obtain samples that are approximately distributed according to $p(\theta|y)$, allowing us to estimate expectations and other properties of the posterior distribution. This approach is commonly used in Bayesian inference when direct sampling from the posterior is not feasible. This method is known as Markov chain Monte Carlo (MCMC) and is widely used in various fields, including physics, statistics, and machine learning, to perform inference and estimate parameters in complex models. Appendix Z compares an MCMC scan of BSM models with the active learning (AL) scan used there.

In the context of particle physics, a particularly relevant equation to estimate is the cross-section of a process given previously in Equation 1.17. The differential cross-section is a function on a phase-space of dimension $3n-4$, where n is the number of particles in the final state, which can range from a few to several tens. There is no analytic solution for the cross-section of a realistic hadronic process, and the dimensionality of the problem is too high for grid-based numerical integration. The MC method is therefore the standard solution to estimate the cross-section, as it converges as $\frac{1}{\sqrt{N}}$ regardless of the dimensionality of the problem. The main drawback is the statistical error, which makes an efficient importance-sampling strategy essential to keep the number of evaluations of the ME manageable.

Before detailing the factorised cross-section, two energy scales must be introduced:

- The renormalisation scale μ_R is introduced to handle the UV divergences arising in fixed-order calculations in perturbation theory, and is typically chosen to be of the order of the energy scale of the process under study. The renormalisation procedure absorbs these divergences into the redefinition of physical parameters, such as couplings and masses, and the residual dependence of physical predictions on μ_R formally vanishes as higher orders are included.
- The factorisation scale μ_F separates the short-distance, perturbatively calculable part of the cross-section from the long-distance regime, in which the QCD coupling becomes large and a perturbative treatment is no longer valid. The collinear singularities associated with QCD ISR are absorbed into the PDFs at the scale μ_F, which is also typically chosen to be of the order of the energy scale of the process. Below this scale, in the soft QCD regime, the perturbative description is replaced by non-perturbative inputs whose form depends on the context.

The factorisation theorems [243, 244] state that for hard processes with a momentum transfer $Q \gg \Lambda_\mathrm{QCD}$, the hadronic cross-section can be written as a convolution of universal PDFs with a perturbatively calculable partonic cross-section already given in Equation 2.10 where we introduce the new energy scales for $f_i(x, \mu_\mathrm{F}^2)$. The hard-scattering process is a short-distance perturbative process that can be computed order-by-order in the QCD coupling α_S, since $Q \gg \Lambda_\mathrm{QCD}$ implies $\alpha_\mathrm{S}(Q) \ll 1$. The shower and hadronisation stages, which are long-distance non-perturbative processes at the scale $\frac{1}{\Lambda_\mathrm{QCD}}$, are not part of this convolution. They are applied event-by-event to the partonic final state produced by the ME, and can be modelled sequentially without quantum interferences between stages.

The PDFs themselves are delivered through the LHAPDF [133] interface, which provides a common access layer to all modern PDF sets [134, 245–247]. An example is given in Figure 2.7. For the PDF we can distinguish two flavour schemes to treat the heavy quarks:

- The 5-flavour number scheme (5FNS), where the b quark is treated as a massless parton and included in the PDFs. This scheme is more accurate for processes with a hard scale $Q \gg m_\mathrm{b}$, where the b quark can be considered as a parton in the proton.
- The 4-flavour number scheme (4FNS), where the b quark is treated as a massive particle and not included in the PDFs. This scheme is more accurate for processes with a hard scale $Q \sim m_\mathrm{b}$, where the mass of the b quark cannot be neglected.

At LO, commonly named the Born level, the cross-section is the squared tree-level ME times the parton luminosity:

$$\mathrm{d}\,\sigma^{\mathrm{LO}} = \sum_{i,j} f_i f_j \frac{|\mathcal{M}_n|^2}{F \cdot S}\,\mathrm{d}\,\Phi_n\,, \tag{4.60}$$

where $F = 2\hat{s}$ is the flux factor for massless initial partons, S is the symmetry factor accounting for identical particles in the final state, $\mathrm{d}\,\Phi_n$ is the n-body LIPS, and $|\mathcal{M}_n|^2$ is the squared ME summed over final state and averaged over initial state colours and spins. The LO calculation is the simplest and the fastest to compute. It allows one to retrieve the dominant production channels and their relative ordering, gives the correct qualitative kinematic structure, and reproduces the symmetries of the process exactly. On the other hand, it is not sufficient to capture the full picture and translate into a realistic simulation. The renormalisation and factorisation scales are not constrained at LO, and the dependence on these scales is large. Moreover, LO does not include any contribution from higher-order corrections, which can be significant. In particular, LO cannot populate the high-jet-multiplicity tails of the phase-space, leading to biased estimates of the cross-section in those regions and distorted kinematic distributions.

The NLO cross-section can be expressed schematically as:

$$\sigma^{\mathrm{NLO}} = \int \mathrm{d}\,\Phi_n \left[B_n + V_n + I_n\right] + \int \mathrm{d}\,\Phi_{n+1} \left[R_{n+1} - S_{n+1}\right]\,, \tag{4.61}$$

where each term is integrated over the appropriate phase-space, with the flux, symmetry, and PDF factors implicit:

- B_n is the Born contribution, equal to the LO cross-section
- V_n is the virtual contribution, arising from the interference of the one-loop amplitude with the Born amplitude: $V_n \propto 2\,\mathrm{Re}\left(\mathcal{M}_n^{(1)\,*}\mathcal{M}_n^{(0)}\right)$. It contains infrared $\frac{1}{\varepsilon}$ poles in dimensional regularisation
- $I_n = \int_{+1} S_{n+1}$ is the analytic integration of the local subtraction term over the unresolved one-particle phase-space. Added to V_n, it cancels the explicit $\frac{1}{\varepsilon}$ poles and leaves a finite n-body integrand
- $R_{n+1} \propto \left|\mathcal{M}_{n+1}^{(0)}\right|^2$ is the real contribution, arising from tree-level diagrams with one additional parton in the final state
- S_{n+1} is the local subtraction term that mimics R_{n+1} in the soft and collinear limits, regularising the real contribution so that $R_{n+1} - S_{n+1}$ is integrable in 4D.

The addition of NLO corrections improves the description of the process: it includes contributions from higher-order diagrams, accounts for the effects of virtual particles, and significantly reduces the dependence on the renormalisation and factorisation scales, leading to a more stable prediction. However, the computation of NLO corrections is significantly more complex and computationally intensive than LO, due to the evaluation of one-loop amplitudes and the implementation of the subtraction machinery.

Since the virtual contribution, once combined with the integrated subtraction term, is finite but not bounded from below, and the same holds for $R_{n+1} - S_{n+1}$, the MC integrand can take negative values, producing events with negative weights. These events are fully part of the NLO prediction and must not be discarded, as they are necessary to obtain a finite and accurate cross-section. When taking into account all weights, positive or negative, the full MC dataset reproduces the expected cross-section and kinematic distributions at NLO accuracy.

The local subtraction term S_{n+1} is not unique. Two systematic and process-independent constructions are in common use. In the *Catani–Seymour (CS)* dipole scheme [248], the counterterm is written as a sum of *dipoles*:

$$S_{n+1} = \sum_{\mathrm{pairs}(i,j)} \sum_{k \neq i,j} \mathcal{D}_{ij,k}\,, \tag{4.62}$$

each describing the collinear splitting of an emitter pair (i, j) in the presence of a spectator parton k. Every dipole reproduces the exact soft and collinear behaviour of $|\mathcal{M}_{n+1}|^2$ in the corresponding limit and factorises into a reduced n-parton Born ME times a universal singular kernel that depends only on the emitter, the spectator, and the splitting variables. The spectator absorbs the recoil through an exact on-shell momentum mapping that conserves total four-momentum, so that the singular $(n+1)$-parton configuration is mapped onto a Born-like n-parton one. The analytic integral of each dipole over the one-parton unresolved phase-space, $I_n = \int_{+1} S_{n+1}$, produces the explicit $\frac{1}{\varepsilon}$ poles that cancel those of the virtual contribution. Because the kernels are universal, the scheme is fully automated in modern generators. It also admits extensions to massive emitters and to photon radiation in QED.

The *Frixione–Kunszt–Signer (FKS)* scheme [249, 250] achieves the same cancellation with a different organisation. The real-emission phase-space is partitioned into a small number of disjoint sectors, each containing at most one soft and one collinear singularity, selected by partition functions. Within each sector the singular limits are parametrised by the scaled energy and emission angle of the radiated parton and isolated using plus-distributions, so that the subtraction is carried out analytically through the plus-prescription. The two schemes are equivalent in their results but differ in the number and form of the subtraction terms. FKS typically generates fewer terms.

Going beyond NLO to NNLO and higher orders allows for an even more accurate description, but the evaluation of two-loop amplitudes and the corresponding subtraction schemes is far more involved, and NNLO predictions are available only for a limited set of processes.

The hard-scattering process consists of a sum over all Feynman diagrams contributing to the process at a given order in perturbation theory with n partons. This procedure works well for well-separated particles, but in the collinear or soft limit the denominator of an internal propagator becomes small and gives rise to logarithmic enhancements that spoil the fixed-order expansion. The parton shower algorithms are designed to resum the leading logarithmic contributions from soft and collinear emissions to all orders in perturbation theory. They use a probabilistic approach to simulate the emission of additional partons, starting from the hard scattering and evolving down to lower energy scales. The factorisation of the squared amplitudes in the collinear limit is given by:

$$|\mathcal{M}_{n+1}|^2 \xrightarrow{j \parallel k} \frac{8\pi\alpha_{\mathrm{S}}}{s_{jk}} P_{i\to jk}(z) |\mathcal{M}_n|^2 \,, \tag{4.63}$$

where $s_{jk} = (p_j + p_k)^2 = 2\, p_j \cdot p_k$ is the invariant mass squared of the two collinear partons and $P_{i\to jk}(z)$ is the Altarelli–Parisi splitting function that describes the probability for parton i to split into partons j and k with momentum fraction z. The splittings cover all relevant cases: $\mathrm{g} \to \mathrm{gg}$, $q \to q\mathrm{g}$, and $\mathrm{g} \to q\bar{q}$. The splitting functions are universal in the sense that they depend only on the type of parton and the momentum fraction, and not on the specific process. This universality allows the parton shower to be applied independently of the hard scattering. For ISR, the evolution is run backwards from the hard scale towards the parton scale. For FSR, it is run forwards from the hard scale towards Λ_{QCD}.

The full chain for an event generated with MC, hard scattering, parton-shower evolution including ISR, FSR, and multi-parton interactions (MPI), hadronisation, and decay, can be seen as a Markov chain:

$$P(\text{event}) = P(\text{hard scattering}) \otimes P(\text{ISR, FSR, MPI}) \otimes P(\text{hadronisation}) \otimes P(\text{decay}) \,, \tag{4.64}$$

where each stage depends only on the partonic configuration produced by the previous stage.

Once the showering process and the NLO computations are done in parallel, a matching procedure [251–253] must be applied to avoid double counting between the real emission in the NLO calculation and the first emission generated by the parton shower.

The NLO matching only provides NLO accuracy for the lowest multiplicity. If the process of interest requires final states with three, four, or more jets described at ME accuracy, the ME calculations at several multiplicities must be combined with the parton shower, in a procedure known as merging [254, 255]. Events at multiplicities n, $n+1$, …, $n_{\max}$ are generated separately, showered, and the partons are clustered with a jet algorithm. For each event, the ME partons are matched to the showered jets: if all ME partons are matched, the event is kept, otherwise it is rejected.

Underlying events and multi-parton interactions are modelled using phenomenological models that simulate the secondary parton–parton interactions occurring in a single pp collision. The model includes the impact-parameter profile of the proton, an interleaved evolution in transverse momentum, and a colour reconnection mechanism that determines how the colour lines of the different interactions are rearranged before hadronisation.

After all the hard partonic activity, the proton remnants must be accounted for, since they can hadronise and contribute to the final state. These beam remnants form a colour-charged, longitudinally fast, low-p_{T} ensemble that has to be hadronised in the proton-fragmentation region. This ensures conservation of colour charge and baryon number, and can lead to the production of additional particles in the final state, mostly at large rapidity.

The last stage of the event generation is the hadronisation process, which models the transition from partons to hadrons when the energy scale of the interaction becomes comparable to Λ_{QCD}. This process is non-perturbative and cannot be calculated from first principles, but it can be modelled using phenomenological prescriptions. The Lund string model [256] assumes that the colour field between a $q\bar{q}$ pair forms a flux tube of approximately constant transverse cross-section, giving a linear potential of tension. As the partons separate, the energy stored in the string grows linearly with the distance until it becomes energetically favourable to break the string by spontaneous $q\bar{q}$ pair creation, fragmenting it into hadrons. The cluster model [257], instead, forcibly splits each gluon at the end of the shower into a $q\bar{q}$

pair, so that the natural objects emerging from the shower are colour-singlet clusters that subsequently decay into hadrons.

The hadronisation leaves a set of stable and unstable hadrons. The generator then walks down the decay tree, decaying each unstable hadron until only PDG-stable particles remain or particles with a lifetime long enough to be observed in the detector. Each decay is weighted by the corresponding branching fraction (BR), and the kinematics of the decay products is determined by the ME of the decay. QED FSR from charged particles, including photon emission off final-state leptons, is important for the precise modelling of leptonic final states.

The generated events are then handed off to the detector simulation, which encodes the full event history including truth-level parents and decay chains. The detector simulation models the response of the experimental apparatus to the particles produced in the collision, including the interactions of the particles with the detector material and the electronic response of the readout.

Pile-up is included by overlaying additional minimum-bias events, on top of the hard-scattering event during the digitisation step. The distribution of the number of interactions per bunch crossing $\langle\mu\rangle$ is matched to the one observed in data, and both in-time pile-up (extra interactions in the same bunch crossing) and out-of-time pile-up (interactions in neighbouring bunch crossings) are simulated.

A visualisation of the full chain of interactions resulting from a pp collision simulated with an MC generator is shown in Figure 4.13.

The theoretical uncertainty on MC predictions has several standard components:

- The dependence on the unphysical scales μ_R and μ_F is probed by the seven-point variation, in which each scale is varied independently by factors of $\frac{1}{2}$ and 2 with the constraint $\frac{1}{2} \leq \frac{\mu_\mathrm{R}}{\mu_\mathrm{F}} \leq 2$
- PDF uncertainties are propagated via the Hessian eigenvectors or replica members of the PDF set, combined with a variation of $\alpha_\mathrm{S}(m_\mathrm{Z})$ around its central value of 0.118
- Parton shower uncertainties are estimated through the dedicated weight variations provided by the generator.

The simulated events are then analysed with the same reconstruction software as the data, allowing for direct comparisons between theory and experiment and for the extraction of physical parameters. As explained throughout this section, the production of simulated events in particle physics is a complex multi-stage process, each stage relying on dedicated algorithms and tunes. Several frameworks provide one or more of these stages. The ones relevant to this thesis are:

- MadGraph (MadGraph5_aMC@NLO [258]) for the hard scattering at LO and NLO. It is worth clarifying the name, which denotes the combination of two parts: MadGraph5, the ME generator that builds tree-level (and, through MadLoop [259], one-loop) amplitudes automatically from a model file and integrates them with MadEvent, and aMC@NLO [258], the module that *automates* the matching of the NLO ME to the parton shower. The matching itself follows the MC@NLO method [251], originally introduced as a separate, process-by-process tool, of which aMC@NLO ("automated MC@NLO") is the general implementation. The real-emission singularities are subtracted with FKS via MadFKS [250].
- POWHEG [260] for the hard scattering at NLO matched to the parton shower with the alternative POWHEG method [261], which generates the hardest emission first with positive weights.
- Pythia [262, 263] for the parton shower, hadronisation via the Lund string model, and the underlying event. MadSpin [264] is used to perform spin-correlated decays of heavy resonances, and Photos [265] for QED FSR.
- Sherpa [266, 267] for full event generation, including ME (Comix [268]), MC@NLO-type matching and multijet merging, its own dipole-based shower [269], hadronisation, and underlying event. One-loop amplitudes are obtained through interfaces to providers such as OpenLoops [270] or Recola [271].
- Matrix element plus parton shower (MEPS) merging [253–255], which is not a generator but the prescription by which ME calculations with different numbers of additional partons are combined with the parton shower without double counting, at LO accuracy (MEPS@LO) or at NLO accuracy (MEPS@NLO). It is the scheme Sherpa applies to the multijet samples of this thesis.
- Geant4 [272] for the detector simulation, combined with the ATLAS Simulation Infrastructure [273] for the digitisation and reconstruction of the simulated events.
- EvtGen [274] for the decays of heavy-flavour hadrons.

Having several frameworks available allows cross-checks and modelling comparisons between alternative ME, matching, shower, and hadronisation choices, which enter directly as systematic uncertainties in the analysis.

Since analyses at high-energy colliders are typically performed blinded to the data in the signal region (SR), MC predictions are the core of every step: they are used to design the analysis strategy, to estimate the expected sensitivity, to train multivariate classifiers, and to interpret the results. A faithful description of what happened in the collision is therefore crucial to extract meaningful physics results from the data.

With the increase of computing power and the development of new algorithms, the use of MC methods in particle physics has become increasingly sophisticated, allowing for more accurate and detailed simulations of the complex processes that occur in high-energy collisions. This goes hand in hand with the increase of the amount of data collected by the experiments, as discussed in Section 3.7. The combination of both is particularly beneficial for the use of NNs, which require large training samples, as discussed in Section 4.1.

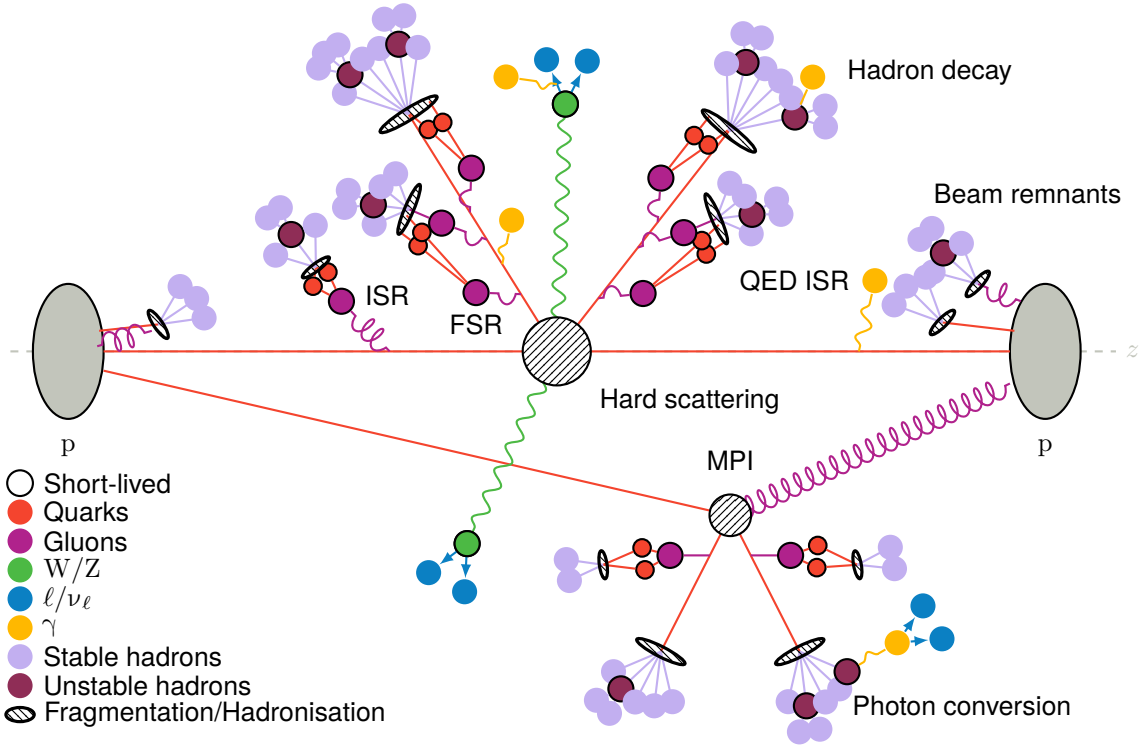
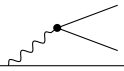


Figure 4.13: Scheme of the interactions in a pp collision as simulated by an MC generator.

5 Electron in ATLAS



We presented in Chapter 3 the ATLAS detector and how its various sub-detectors turn the particles produced in pp collisions into electrical signals. After an event passes the trigger, it is sent for offline reconstruction to associate each low-level object with a physics object usable for analysis. The combination of the various hits in the ID and MS by charged particles leads to *tracks* giving the trajectory of incoming particles. The energy deposits in the EM and HAD calorimeters are grouped to form *calorimeter clusters*. The combination of those two types of low-level objects allows a unique identification of every particle of the SM.

A well-thought-out detector design has to be complemented by a solid reconstruction strategy to fully exploit the potential of the detector. The identification, reconstruction, and calibration of physics objects define the basis of any physics analysis.

In this chapter, we will focus on one of the most fundamental particles, central to many analyses: the electron. Because electrons alone give only a partial picture, we will also discuss the reconstruction of photons, which are intrinsically linked to electrons. A scheme showing how both these particles are identified via the combination of tracks and calorimetric clusters is shown in Figure 5.1.

We will present the reconstruction and calibration of electrons and photons in the barrel region of the detector. The end-cap region is more complex due to the lack of track coverage, making the reconstruction of electrons harder and the disentanglement from photons more difficult. We will see how it is done currently in this region and the core of this chapter will be to study the potential gain with the introduction of the ITk to expand the tracking coverage and thus the electron reconstruction in the end-cap region.

The work related to physics objects is assigned to various Combined Performance groups in ATLAS, which are responsible for the reconstruction and calibration of the various physics objects. This chapter is based on the work I have contributed to in the e/γ Combined Performance group.

5.1 Reconstruction

As detailed in Section 3.3.1, the EM calorimeter is well designed to fully contain a shower initiated by an incoming photon or electron. Therefore, the reconstruction for these particles starts by defining clusters from energy deposits in the calorimeter. The difference between an electron and a photon will only arise from the presence or absence of a track. Since the electron is charged, it will leave tracks. However, due to interactions in the ID, photons can undergo pair production, leaving tracks displaced from the IP. About $20\,\%$ of photons convert at low $|\eta|$ and up to $65\,\%$ at $|\eta| \simeq 2.3$ due to the varying amount of material traversed. We then differentiate the unconverted photon (purely reconstructed with a calorimeter cluster) from the converted photon (reconstructed with a calorimeter cluster and one or two tracks). Both cases are represented in Figure 5.1.

The starting point for the reconstruction is to define clusters from the energy deposits in topologically connected EM calorimeter cells (commonly named topo-clusters). Then tracks from the ID and potential conversion vertices are matched with the topo-cluster energy. The nature of the objects is thus defined at this step depending on the presence of tracks (electron), conversion vertices (converted photon) or not (photon).

The next step will be to construct super-clusters [275], derived from matching track and topo-clusters. They are more complex objects than topo-clusters as they are not only defined from the energy deposits in the calorimeter but also from the tracks and conversion vertices. They are designed with an adaptive size to incorporate additional clusters from bremsstrahlung photons into the original electron cluster or

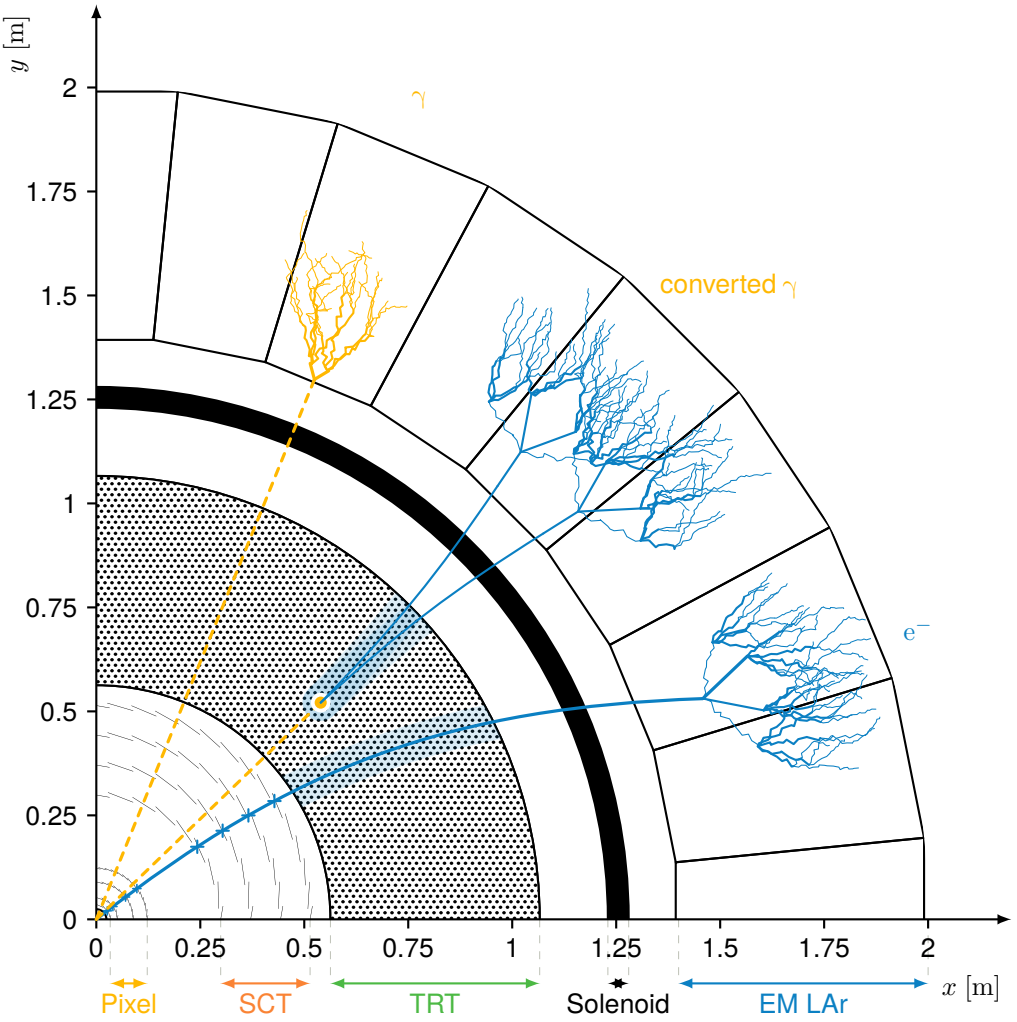


Figure 5.1: Electron and photon reconstruction in the ATLAS detector.

to merge electron–positron clusters from photon conversions (PCs). Once they are formed, tracks and conversion vertices are matched with them again.

These steps are done in parallel for both electrons and photons. A single topo-cluster can be associated with both an electron and a photon super-cluster. In this case both an electron and a photon are reconstructed but flagged as ambiguous. The final physics objects used in the analysis are then constructed: the ambiguous objects are classified, the electron and photon energies are calibrated, and finally all the necessary information for identification and isolation is computed. Each of these steps is represented in Figure 5.2. Note that for the particular case of the forward electron, above the ID coverage, the reconstruction is done only with the calorimeter information and thus only the cluster formation and final physics object construction steps are performed. In this region with $|\eta| > 2.5$, the electron is reconstructed as a photon and then classified as an electron using a dedicated identification algorithm.

We will now go through all these various steps of the reconstruction in more detail [276–278]. We will start with the formation of topo-clusters, then we will see how tracks are reconstructed and matched with the clusters and finally how super-clusters are formed and the final physics objects are constructed.

5.1.1 Cluster formation

The starting point of electron and photon reconstruction is the topo-clusters [279] that correspond to variable-size clusters of energy deposits connected topologically through adjacent cells of the EM calorimeter. The central parameter that governs the formation of topo-clusters is the significance of a given cell, defined as:

$$\zeta_{\text{cell}}^{\text{EM}} = \frac{E_{\text{cell}}^{\text{EM}}}{\sigma_{\text{noise,cell}}^{\text{EM}}}, \tag{5.1}$$

where $E_{\text{cell}}^{\text{EM}}$ corresponds to the energy deposit in the cell while $\sigma_{\text{noise,cell}}^{\text{EM}}$ is the expected noise from both the pile-up and the readout electronics. The topo-cluster reconstruction algorithm uses thresholds over

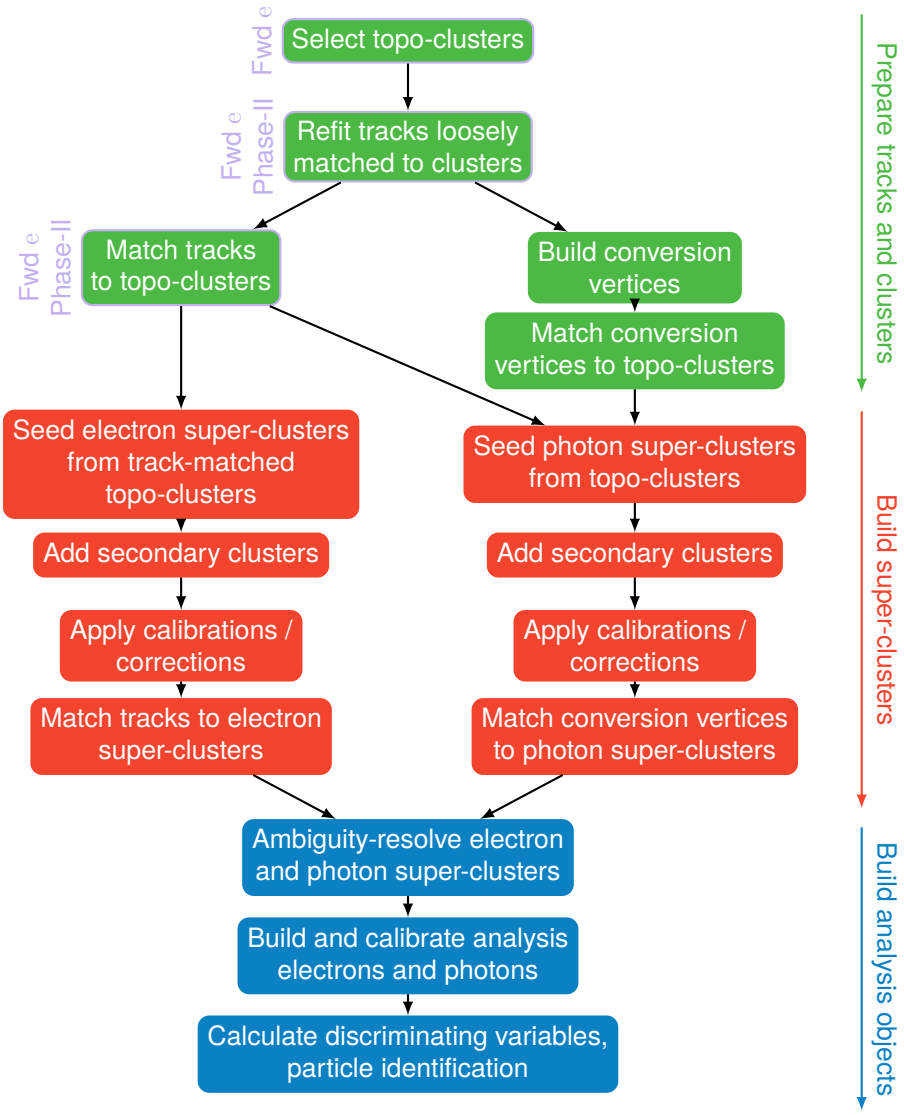


Figure 5.2: Electron and photon reconstruction steps.

this parameter while scanning the full calorimeter layout to select cells and grow clusters around them following the "4-2-0" scheme [280] defined in this order:

1. Formation of proto-clusters where $\left|\zeta_{\text{cell}}^{\text{EM}}\right| \geq 4$ in the EM calorimeter without taking into account the pre-sampler and Layer 1. This is done to avoid the formation of noise clusters.
2. The neighbouring cells around the initial proto-clusters with $\left|\zeta_{\text{cell}}^{\text{EM}}\right| \geq 2$ are added. They are then used as seed cells for the next iteration, each collecting again its neighbours matching this threshold. In the case where the condition is met for two distinct clusters, they are merged together resulting in a single cluster.
3. A perimeter is formed by taking all nearest-neighbour cells with $\left|\zeta_{\text{cell}}^{\text{EM}}\right| \geq 0$ in the cluster.
4. Topo-clusters with two or more local maxima (that is, cells with $E_{\text{cell}}^{\text{EM}} > 500\,\text{MeV}$ with at least four neighbours and when none of them have a larger signal) are split into different topo-clusters.

The algorithm uses the absolute value of the significance to avoid biasing the cluster energy upwards.

As this algorithm is used for the reconstruction of other particles such as hadrons, the topo-clusters may include cells from both the EM and HAD calorimeters. To avoid this and retrieve only EM showers initiated by photons or electrons, we define the fraction of EM component of the topo-clusters following:

$$f_{\text{EM}} = \frac{E^{\text{L1}} + E^{\text{L2}} + E^{\text{L3}}}{E^{\text{cluster}}}\,, \tag{5.2}$$

with $E^{\text{L}i}$ the cluster energy in the i-th layer of the EM calorimeter and E^{cluster} the total cluster energy. This fraction has to be greater than 0.5 in order for a cluster to be considered as originating from an EM

shower. In this case the cells from the HAD calorimeter are removed. This allows one to reject around $60\,\%$ of pile-up clusters without affecting the efficiency for selecting true electron topo-clusters.

In the pre-sampler and scintillator region $1.37 < |\eta| < 1.63$, the energy of the pre-sampler and the scintillator visible in Figure 3.6(b) is also taken into account in the cluster formation.

After this selection, only the topo-clusters with an EM energy larger than $400\,\mathrm{MeV}$ are kept.

5.1.2 Tracking

A charged particle traversing the ID leaves hits in the various layers of the pixel, SCT, and TRT detectors. These hits are then combined to form tracks, which represent the reconstructed trajectories of the charged particles [281, 282].

Each track follows a helical trajectory in the solenoid magnetic field. A complete representation of the track at the point of closest approach to the beam axis, known as the *perigee*, is given by five parameters: the transverse and longitudinal impact parameters d_0 and z_0, respectively, the azimuthal angle ϕ and the polar angle θ, and the signed curvature $\frac{q}{p}$ where q is the charge of the particle and p is its total momentum. All these parameters are represented in Figure 5.3 following the coordinate system used in ATLAS defined in Section 3.1.

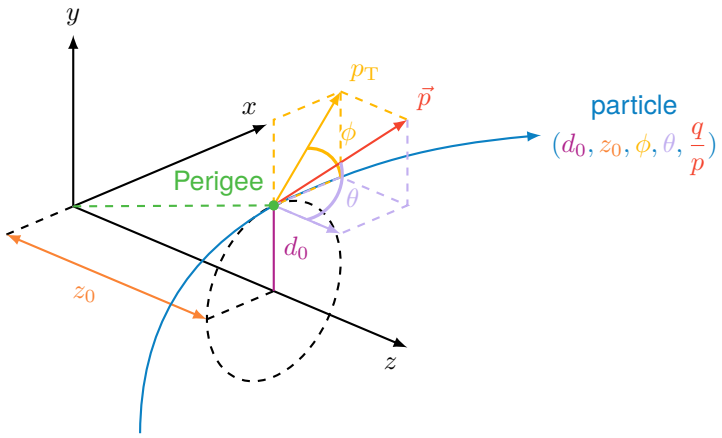


Figure 5.3: Track parameters at the perigee.

In addition to these five parameters, a 5×5 covariance matrix describes the uncertainties on and correlations between the track parameters. The resolution of the track parameters is crucial for the performance of the detector, as it directly impacts the precision of the momentum measurement and the ability to distinguish between different particles, as discussed in Section 3.6.1.

To predict where a track will intersect the next layer of the detector, the track parameters must be propagated through the (generally inhomogeneous) magnetic field. The equation of motion for a charged particle in a magnetic field, parametrised along the arc length s, is given by:

$$\frac{\mathrm{d}^2\,\vec{r}}{\mathrm{d}\,s^2} = \frac{q}{p}\left(\frac{\mathrm{d}\,\vec{r}}{\mathrm{d}\,s} \times \vec{B}(\vec{r})\right) , \tag{5.3}$$

where $\vec{r}$ is the position of the particle, q is its charge, p its momentum and $\vec{B}(\vec{r})$ is the magnetic field. This equation can be reformulated as a system of coupled second-order differential equations along the z-axis and is solved numerically using a fourth-order Runge–Kutta method [283, 284]. Material effects such as multiple scattering and energy loss are also taken into account during propagation.

Seeds are the starting point of the track reconstruction. They are defined as a combination of three space points in the pixel and SCT detectors that are compatible with a helical trajectory originating from near the interaction region. Loose requirements on the impact parameters d_0 and z_0 are applied to reject obviously incompatible combinations and limit the seed multiplicity. A fourth space point is then sought, either inward or outward from the seed, to confirm it and form a track candidate. This approach is known as *local seeding*. Other methods exist, such as global pattern recognition using a Hough transform [285] to find track candidates in the space of track parameters.

Once the seeds are formed, search roads are constructed. They correspond to portions of the detector where compatible hits are expected given the seed's estimated trajectory and its uncertainties. Within these search roads, a CKF [286, 287] is applied to extend seeds into full track candidates. This algorithm starts from the seed parameters and propagates them to the next detector layer. The predicted position

and its uncertainty define a window within which compatible hits are sought. If a single hit is found, the track parameters are updated accordingly. This procedure is represented in Figure 5.4 where $q_{k|k-1}$ represents the predicted track parameters at layer k given the measurements up to layer $k-1$, while $q_{k|k}$ represents the updated track parameters after including the measurement at layer k. This is repeated iteratively until the outermost Si layer is reached. If multiple hits fall within the window, the filter branches: each compatible hit spawns a separate track hypothesis. At each subsequent layer, each hypothesis is independently propagated and updated, and branching continues until the outermost Si layer is reached. The branching is controlled by χ^2 cuts to prevent an exponential growth of the number of track candidates. Hypotheses with poor fit quality are pruned. The full process is run both inward and outward from the seed to ensure a better smoothing of the trajectory. More details on the χ^2 distribution are provided in Section O.7.

Seeds that fail the standard track-finding step are not discarded immediately. If a seed lies within a calorimeter RoI, the track-finding is re-attempted with relaxed criteria that allow additional kinks in the trajectory. This is particularly important for electron reconstruction, as electrons are likely to emit bremsstrahlung photons that cause sudden changes in track curvature due to the corresponding momentum loss.

After all track candidates are formed, an ambiguity resolution step is applied to reduce duplicate tracks and suppress fakes. Each track candidate is assigned a score based on several criteria: the number of precision hits (with pixel hits receiving higher weight than SCT hits), the fit χ^2, the number of holes, the number of shared hits with other candidates, and the track momentum. A hole is defined as the absence of a hit on an active sensor that the track is expected to have traversed, and is indicative of either detector inefficiency or an incorrect hit-to-track association. Tracks are then processed in descending score order. When a high-scoring track claims a hit, that hit becomes unavailable for subsequent candidates, which are accordingly penalised or rejected. Quality selections on the minimum number of hits, maximum number of holes, and minimum momentum are also applied.

The surviving candidates are then passed to a global χ^2 fitter [288] to refine the track parameters and their covariance matrix. This approach is mathematically equivalent to the CKF, but instead of processing the measurements sequentially, it minimises a χ^2 function constructed from all measurements simultaneously. Material effects are incorporated by adding scattering angle parameters at each material layer as additional degrees of freedom in the χ^2. Measurements that contribute excessively to the χ^2 can be flagged as outliers and down-weighted or removed from the fit, preventing a single poorly measured or incorrectly assigned hit from biasing the entire track.

Additionally, tracks with Si hits that are loosely matched to fixed-size EM clusters in the calorimeter are refitted using a Gaussian sum filter (GSF) [289, 290], a dedicated algorithm designed to handle the non-Gaussian energy losses from bremsstrahlung. This procedure, described in Section 5.1.3, significantly improves the track parameter estimation for electron candidates.

Once the Si-only track is established, an extension into the TRT is attempted as visible in Figure 5.4. A road is constructed around the extrapolated track and a CKF is applied to collect compatible drift circles. Each drift circle defines a radius around the anode wire without indicating on which side the particle passed (left–right ambiguity). This ambiguity is resolved using the direction of the extrapolated Si track. The track is then refitted using the full ID information. The inclusion of the TRT improves the momentum resolution, since the sagitta (s in Figure 5.4) is measured over a longer lever arm. It also provides particle identification through the fraction of high-threshold TRT hits, which allows distinguishing electrons from pions. If the TRT extension worsens the fit χ^2, it is rejected and the Si-only track is retained. Extensions with too few leading-edge hits are also rejected.

A second tracking pass recovers tracks that the first inside-out pass failed to reconstruct. This outside-in pass starts from TRT segments found in calorimeter RoIs, then builds seeds in the Si detectors using only two space points to allow for particles produced at larger radii, such as those from PCs. These seeds are extended into tracks using the same workflow as the primary pass: the combinatorial CKF, ambiguity resolution, global χ^2 fit, and extension into the TRT. This second pass operates exclusively on hits not already claimed by the primary pass, and is essential for recovering tracks from PCs occurring in the detector material.

The tracking performance is quoted for particles with a lifetime larger than $30\,\mathrm{ps}$, which corresponds to a decay length of about $1\,\mathrm{cm}$. This means that tracks from particles decaying before this threshold, such as τ leptons or b and c hadrons, are not reconstructed. However, tracks from particles decaying after this threshold, such as $\mathrm{K_S}$ mesons or Λ baryons, are reconstructed and can be used for secondary vertex reconstruction and b-tagging.

The reconstructed tracks are subsequently used to identify the primary vertices of the pp collisions. With each pile-up collision producing its own primary vertex, correctly identifying these vertices is essential for associating tracks and physics objects to the correct interaction, suppressing pile-up contributions, and reconstructing secondary vertices from heavy-flavour (HF) hadron decays used in b-tagging.

The primary vertex reconstruction in ATLAS [291] proceeds in two stages: vertex seeding and vertex fitting. The vertex seeding identifies positions along the beam axis that are likely to correspond to pp interaction vertices. In the current ATLAS software, a Gaussian Smoother algorithm is used: each track is modelled as a 2D Gaussian probability distribution in the (d_0, z_0) plane, with a shape determined by the track's covariance matrix. The sum of these distributions along the beam axis forms a track density function, whose local maxima provide the vertex seed positions.

Starting from each seed, tracks within a longitudinal window are collected, and an adaptive vertex fit is performed using a CKF approach. The beam spot, the measured luminous region of the colliding beams, whose transverse size is governed by the beam optics parameter σ^* introduced in Equation 2.3, is used as an initial constraint on the transverse vertex position. Since σ^* is typically of the order of tens of μm, this provides a very tight prior on the transverse coordinates of the vertex, significantly improving the resolution. A key feature of the algorithm, known as the adaptive multi-vertex finder (AMVF) [292], is that tracks can be associated with multiple vertex candidates with a weight. When a track is shared between vertices, all linked vertices are refitted simultaneously, ensuring a globally consistent solution. Outlier tracks with poor χ^2 contributions are iteratively removed. A vertex candidate is accepted if it contains at least two tracks, of which at least one is not shared with any other vertex, and if it is well separated from previously reconstructed vertices. The compatible tracks are then removed from the seed pool and the procedure is repeated until no viable seeds remain.

Among the reconstructed primary vertices, the hard-scattering vertex is identified as the one with the largest sum of the squared transverse momenta of its associated tracks, $\sum p_{\mathrm{T}}^2$. This criterion exploits the fact that the hard-scattering interaction typically produces higher-p_{T} particles than the softer minimum-bias pile-up collisions. The hard-scattering vertex serves as the reference point for computing impact parameters and for associating physics objects to the correct interaction throughout the analysis.

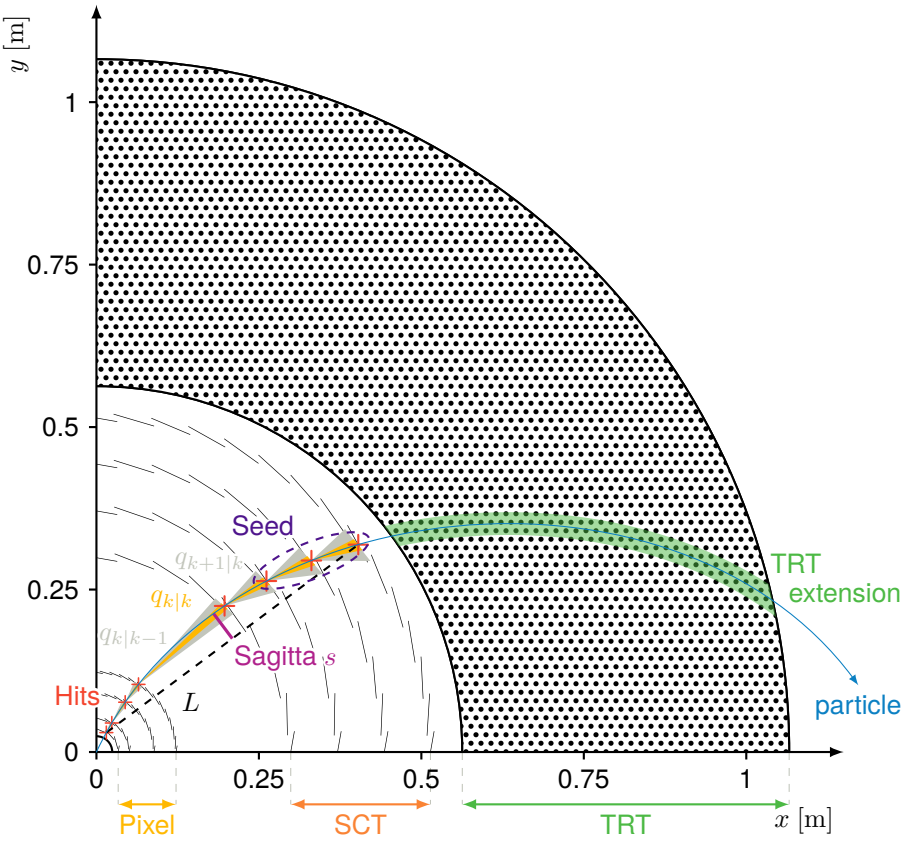


Figure 5.4: Track reconstruction workflow.

5.1.3 Track–cluster matching

The next step in the reconstruction is the matching of ID tracks to energy deposits in the EM calorimeter, which provides the initial differentiation between electrons, unconverted photons, and converted photons.

Fixed-size clusters in the EM calorimeter whose longitudinal and lateral shower profiles are compatible with an EM shower are used to define RoIs already during the tracking stage, as described in Section 5.1.2. After the global χ^2 fit, tracks with Si hits are loosely matched to these fixed-size clusters by extrapolating the track from the last Si hit or from the perigee to the calorimeter. The loose matching requires $|\Delta\eta| < 0.05$ and the azimuthal residual $q \cdot (\phi_{\text{track}} - \phi_{\text{cluster}})$ to fall within an asymmetric window, where q is the reconstructed charge of the track. The asymmetry accounts for the fact that bremsstrahlung photons emitted along the trajectory shift the cluster barycentre in ϕ relative to the extrapolated track position. The window is defined as $[-0.20, 0.05]$ when extrapolating from the last Si hit using the track energy, and $[-0.10, 0.05]$ when extrapolating from the perigee using the cluster energy. The wider window in the former case reflects the fact that the track direction at the last Si hit has not been corrected for upstream bremsstrahlung losses, leading to a potentially larger azimuthal displacement.

Tracks passing this loose match are then refitted using a GSF. The highly non-Gaussian energy-loss distribution from bremsstrahlung invalidates the Gaussian assumption underlying the standard global χ^2 fit. The GSF addresses this by approximating the energy-loss distribution as a weighted mixture of Gaussian components. At each material layer, the filter forks into multiple branches, each with its own weight, energy-loss hypothesis, and trajectory. After propagation to the next measurement surface, branches whose predictions are incompatible with the actual measurement are pruned. This procedure correctly handles the asymmetric energy-loss tails and significantly improves the track parameter estimation for electron candidates.

The refitted tracks are then matched to EM topo-clusters by extrapolating the track from the perigee to the second layer of the EM calorimeter. To improve the matching for electron candidates with significant bremsstrahlung losses, the extrapolation is performed twice: once using the measured track momentum, and once rescaling the momentum magnitude to match the cluster energy. A track is considered matched if, with either momentum hypothesis, the residuals satisfy $|\Delta\eta| < 0.05$ and $-0.10 < q \cdot (\phi_{\text{track}} - \phi_{\text{cluster}}) < 0.05$. The asymmetric ϕ window reflects the same bremsstrahlung-induced shift as in the loose matching step. This track–cluster matching forms the basis for the subsequent electron and photon identification.

When multiple tracks are matched to the same cluster, a ranking procedure selects the best candidate. Tracks with pixel hits are preferred over tracks with only SCT hits, owing to the finer granularity and better spatial resolution of the Pixel Detector. Additional weight is given to hits in the innermost pixel layer. Among tracks of the same category, the one with the smallest ΔR to the cluster in the second calorimeter layer is preferred, unless the ΔR differences are smaller than 0.01, in which case the track with more pixel hits is chosen. The highest-ranked track defines the baseline reconstructed electron properties.

5.1.4 Photon conversion reconstruction

The tracks loosely matched to the fixed-size clusters in the EM calorimeter are also used to reconstruct PC vertices. Two types of PCs are distinguished: single-track conversions, where only one of the two tracks from the $\mathrm{e}^+\mathrm{e}^-$ pair is reconstructed, and double-track conversions, where both tracks are reconstructed.

Single-track conversion vertices correspond to asymmetric conversions where one of the two tracks carries very little momentum and fails to be reconstructed, or to cases where a highly energetic photon produces a very collimated $\mathrm{e}^+\mathrm{e}^-$ pair that cannot be resolved into two separate tracks. Single-track conversions occur predominantly at large radii, in particular within the TRT, where the coarser spatial resolution compared to the Si detectors makes it more likely for a track to be missed or for two closely spaced tracks to be merged. To suppress misidentification of prompt electrons as single-track conversions, single-track conversion vertices are reconstructed only when the track has no hit in the innermost pixel layer, which indicates that the conversion occurred at a radius beyond this layer.

A prompt particle is one that originates from the primary vertex, as opposed to a secondary particle produced in a decay or interaction within the detector material. In the context of PC reconstruction, prompt electrons are those that are produced directly in the hard-scattering interaction at the primary vertex, rather than being the result of a photon converting into an $\mathrm{e}^+\mathrm{e}^-$ pair within the detector material. A schematic representation of a prompt and non-prompt electron is shown in Figure 5.5.

Double-track conversion vertices are reconstructed by pairing oppositely charged tracks that originate from a common vertex. Geometric constraints are applied to the candidate pairs, requiring consistency in the distance of closest approach between the two tracks, their opening angle, and their invariant mass, which must be compatible with the decay of a massless particle. Depending on the presence of Si hits on

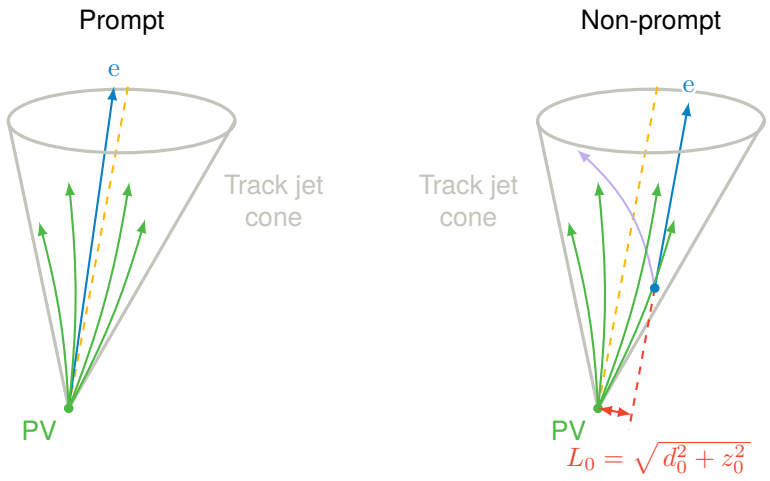


Figure 5.5: Schematic representation of a prompt electron (left) and a non-prompt electron (right). The prompt electron originates from the primary vertex, while the non-prompt electron is produced in a secondary interaction.

the associated tracks, double-track conversion vertices are classified as double-Si, Si–TRT, or TRT–TRT, referring to whether both, one, or neither of the tracks has Si hits.

The conversion vertices are then matched to EM topo-clusters by extrapolating the vertex position to the second layer of the EM calorimeter and comparing it to the topo-cluster barycentre in η and ϕ. For tracks with Si hits, the matching requires $|\Delta\eta| < 0.05$ and $|\Delta\phi| < 0.05$. For tracks without Si hits, the requirements are loosened to account for the poorer track extrapolation: in the TRT barrel, $|\Delta\eta| < 0.35$ and $|\Delta\phi| < 0.025$, while in the end-cap tighter criteria of $|\Delta\eta| < 0.2$ and $|\Delta\phi| < 0.02$ are applied.

When multiple conversion vertices are matched to a single topo-cluster, a ranking procedure selects the best candidate. Priority is given to double-track conversions with Si hits, then to other double-track conversions, and finally to single-track conversions. Among candidates of the same category, the one with the smallest ΔR to the cluster in the second calorimeter layer is preferred. The selected conversion vertex defines the baseline reconstructed photon properties.

5.1.5 Super-cluster reconstruction

The topo-clusters loosely matched to tracks are used as inputs to form super-clusters. It works in two stages. In the first stage, EM topo-clusters are tested to check whether they can be used as seed cluster candidates that will form the basis of super-clusters. In the second stage, topo-clusters near the seed candidates are identified as satellite cluster candidates, which may emerge from bremsstrahlung radiation or topo-cluster splitting. The satellite clusters are added to the seed candidates to form the final super-clusters if they satisfy the necessary selection criteria that will be detailed below.

The procedure to search for seeds starts by scanning the reconstructed topo-clusters sorted according to descending E_{T}, computed using the EM energy. An exception is made for clusters in the transition region that fail the standard selection but pass a looser selection. They are then added at the end of the list. All clusters are then tested one by one following the ranking made. To be considered as a seed for an electron super-cluster the topo-cluster must have $E_{\mathrm{T}} > 1\,\mathrm{GeV}$ and be matched to a track with at least four hits in the Si detectors. For a photon, E_{T} must be greater than $1.5\,\mathrm{GeV}$ to be considered as a seed, and no requirements are made on tracks or conversion vertices. A cluster cannot be used as a super-cluster seed if it has already been added as a satellite to another one.

Once a topo-cluster passes the seed requirements, the next steps will be to find satellite clusters. For both electrons and photons, it must fall within a window of 3×5 in units of $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$ around the seed cluster barycentre as it tends to represent secondary EM showers originating from the same initial electron or photon. For electrons, a cluster will also be considered as a satellite if it falls within an extended window of 5×12 and its best-matched track is also the best-matched track for the seed cluster. For photons with conversion vertices made only of tracks containing Si hits, a cluster is added as a satellite if its best-matched tracks belong to the conversion vertex matched to the seed cluster. These steps rely on tracking information to discriminate distant radiative photons or conversion electrons from pile-up noise or other unrelated clusters. The three different cases are represented in Figure 5.6 where the seed cluster is represented in red and the satellite clusters in blue.

The super-cluster is therefore defined as the combination of the seed cluster and its associated satellites. The remaining step is to assign calorimeter cells to a given super-cluster. Only cells from the pre-sampler and the first three LAr calorimeter layers are considered, except in the scintillator region $1.4 < |\eta| < 1.6$, where the energy measured in the scintillator (see Figure 3.6(b)) is added. To reduce the sensitivity to pile-up noise, the size of each constituent topo-cluster is constrained to a maximal width of 0.075 or 0.125 in the η direction in the barrel or end-cap region.

Due to the magnetic field in the ID being parallel to the beam line, interactions between electron or photon and the detector material generally cause the EM shower to spread in the ϕ direction, so the restriction in η still generally allows the electron or photon energy to be captured. No restriction is applied in the ϕ direction.

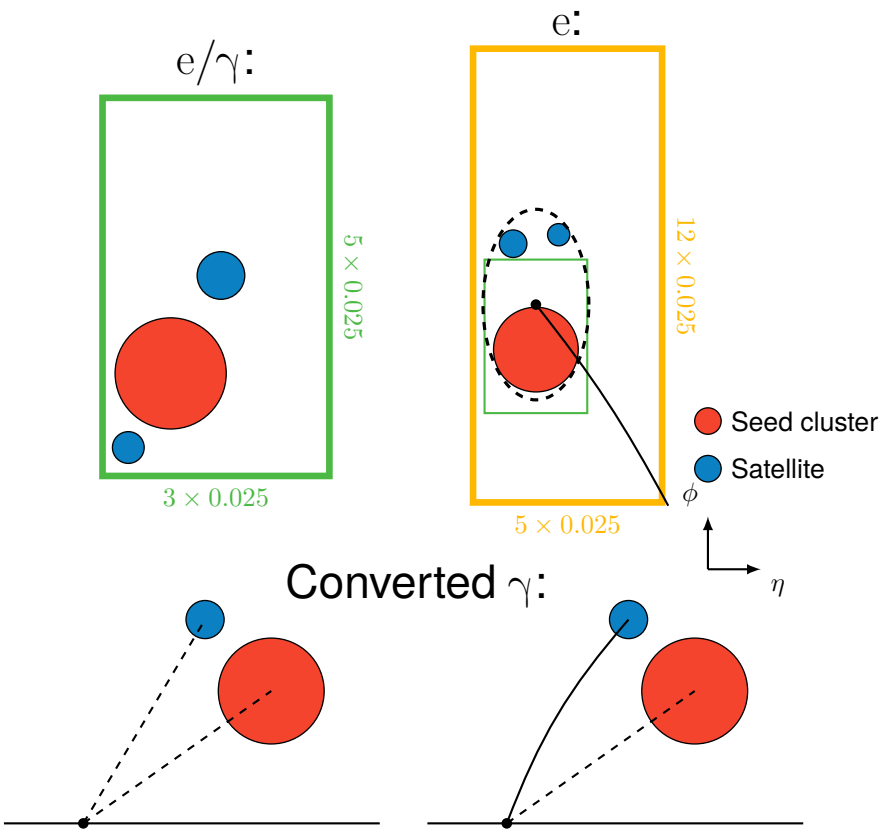


Figure 5.6: Super-cluster reconstruction for electrons and photons.

5.1.6 Ambiguity resolver

Once the super-clusters are formed, the same procedure used for topo-cluster to match tracks and conversion vertices is done. Electron and photon super-clusters are formed in parallel, so that a single topo-cluster can act as a seed for both. To resolve this, an ambiguity resolution algorithm is applied to complete the classification of the candidate. It consists of a succession of if-else cases resulting in the attribution of the object as an electron, a photon, or both.

A super-cluster associated with only a photon or an electron is not ambiguous and therefore only a photon or an electron is reconstructed.

A super-cluster is categorised as a photon if its associated electron track has no hits in the Si trackers or, if it does, has no pixel hits and is part of a matched double-Si conversion vertex.

For an ambiguous object to be categorised as an electron, it has to have four Si hits in its track and to not be part of a conversion vertex. If a conversion vertex is found, the particle can still be reconstructed as an electron if the track has a hit in the innermost layer and the matched conversion vertex is not a double-Si vertex, or, if that is the case, if at least one of its conversion tracks does not have an innermost hit. If the latter conditions fail, an electron is reconstructed if its track is part of a Si–TRT or a TRT–TRT conversion vertex whose radial distance with the innermost hit of the track is above $40\,\mathrm{mm}$.

An object will remain ambiguous, and thus classified as both an electron and a photon, if none of the previous requirements are satisfied, or if the track p_{T} is smaller than $2\,\mathrm{GeV}$, or if $\frac{E^{\mathrm{cluster}}}{p^{\mathrm{track}}} > 10$. The final decision is left to each analysis depending on its specific requirements to disentangle these objects.

A summary of the ambiguity resolution is represented in Figure 5.7 where the different cases are illustrated.

Having seen the complete reconstruction, combining information from the different sub-detectors of ATLAS into physics objects, we now turn to the next steps needed to calibrate, identify, and isolate them for use in analyses.

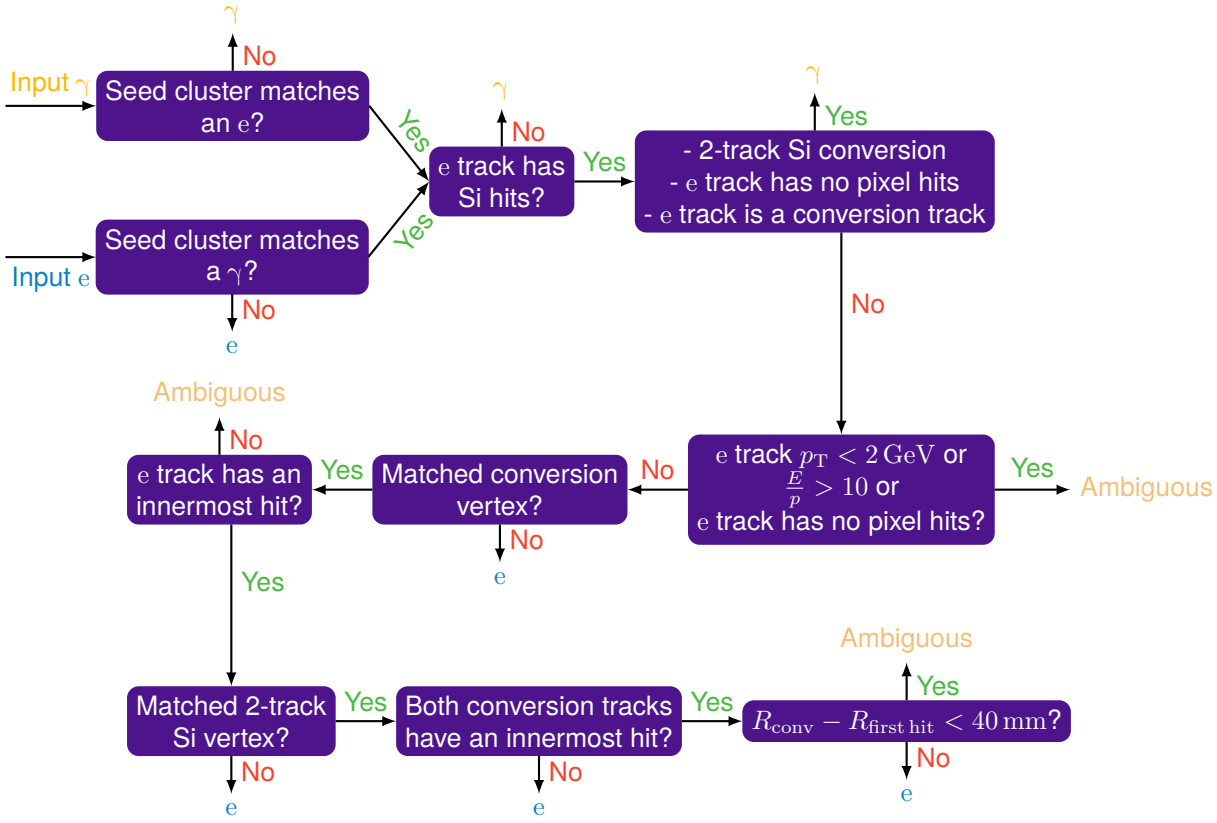


Figure 5.7: Ambiguity resolution for electron and photon super-clusters.

5.2 Calibration

During the particle's journey from its production at the IP to its destruction and energy measurement in the calorimeter, several sources of energy losses exist such as interaction with the tracker or also the fact that the ATLAS EM calorimeter is a sampling calorimeter as detailed in Section 3.3 and thus only a part of the energy is measured. Therefore, to correct this discrepancy between the true value and the measurement a powerful and robust calibration must be employed [293].

The energy measurement starts by converting the electronic signal readout produced by the LAr as explained in Section 3.3.1 and illustrated in Figure 3.8(a). This is done in each cell within the super-cluster formed during the reconstruction step. The signal is converted into a raw energy value representing the energy deposited by the incident particle in this specific cell.

The ionisation current produced in each LAr cell has a triangular pulse shape and is proportional to the energy deposited in the active medium. The signal is pre-amplified and then shaped by a bipolar filter equipped with three overlapping gain stages (low, medium, and high) to accommodate the large dynamic range of energy deposits expected in the calorimeter. Upon an L1 trigger accept, the shaped signal is sampled every $25\,\mathrm{ns}$ and digitised by an analogue-to-digital converter (ADC) using the most suitable gain. The raw cell energy $E_{\mathrm{cell}}^{\mathrm{raw}}$ is reconstructed as a linear combination of the first four digitised samples, after subtracting a pedestal contribution determined from dedicated calibration runs. The combination weights, known as optimal filter coefficients (OFCs) [294], are derived from an optimal filtering algorithm that maximises the energy and timing resolution while minimising the sensitivity to electronic noise and pile-up. The resulting amplitude is then converted from ADC counts to energy through a chain of calibration factors: an $F_{\mathrm{DAC}\to\mu\mathrm{A}}$ factor and an $F_{\mu\mathrm{A}\to\mathrm{MeV}}$ factor obtained from dedicated calibration runs in which known charge pulses are injected into the readout electronics, the gain G used during digitisation, and a

correction $\frac{M^{\text{phys}}}{M^{\text{calib}}}$ that accounts for the different cell responses to ionisation and calibration pulses. The output of this electronic chain, $E_{\text{cell}}^{\text{raw}}$, is the starting point of the offline calibration workflow described in the remainder of this section.

The full workflow of the calibration contains several steps that are represented in Figure 5.8 and will be detailed below following the same ordering. They are applied offline but require prior analysis to obtain specific parameters, and must remain consistent with the evolving conditions of the ATLAS detector and LHC beam.

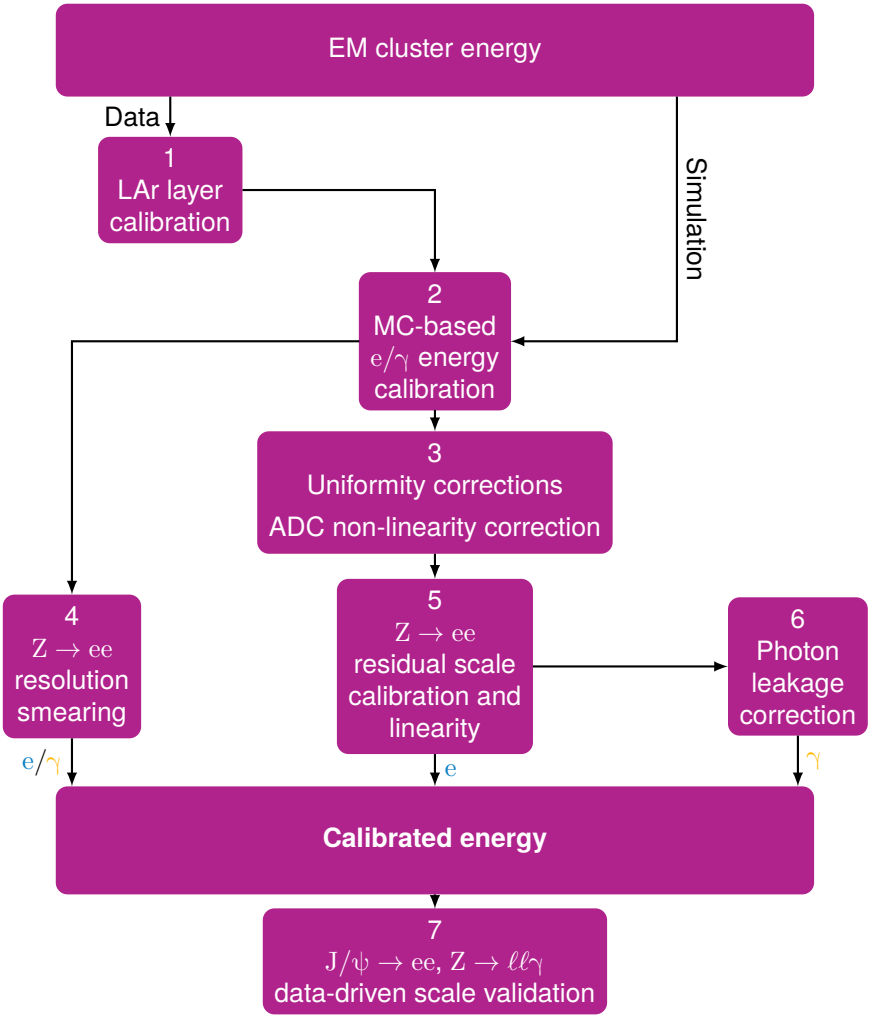


Figure 5.8: Calibration workflow for electrons and photons.

1. LAr layer calibration: the EM calorimeter is longitudinally segmented as visible in Figure 3.7(a), therefore the response of each layer is calibrated independently. The raw energy in data is calibrated to make it match the one obtained in simulation. To achieve this, several energy scale factors (SFs) are required and need to be computed: α_{PS} for the pre-sampler, α_{12} the relative responses of the first and second accordion layers (Layer 1 and Layer 2), and α_{acc} the global scale. The latter is used as an overall correction to the energy measured in all three layers (Layer 1, Layer 2, and Layer 3). Once applied to data, the overall raw energy, E^{raw}, is obtained by summing the cell energies contained in the super-cluster.

   The α_{PS} is determined using EW boson decays into muons in dedicated low-pile-up data samples, as they allow a clean measurement as muons are largely insensitive to the passive material upstream of the calorimeter. The value is determined by computing the ratio of the mean pre-sampler energy deposited by muons in data to the same quantity in simulation.

   The α_{12} is determined by comparing the energy ratio between Layer 1 and Layer 2 in data and simulation, using muons from $\text{Z} \to \mu\mu$ and $\text{W} \to \mu\nu_{\mu}$ events, where the Layer 1 and Layer 2 energy deposits along the muon track are compared between data and simulation, and electrons from $\text{Z} \to \text{ee}$ events, where the constancy of $\frac{E}{p}$ and m_{ee} as a function of $\frac{E_1}{E_2}$ is required. Any non-constant

behaviour is then corrected by applying the α_{12} factor. The value of α_{12} is obtained by combining both channels.

The α_{acc} parameter is based upon the determination of the SF α determined in the next steps of the calibration. It is given by the following relation:

$$\alpha_{\text{acc}} = \frac{\alpha}{\frac{E^{\text{raw}}_{\text{acc}}}{E^{\text{raw}}}}\,, \tag{5.4}$$

where $E^{\text{raw}}_{\text{acc}}$ is the raw accordion energy excluding the pre-sampler contribution.

2. MC-based energy calibration: a simulation-based calibration is applied to correct the raw energy values in both data and MC simulations. This correction accounts for the EM shower energy deposited outside the super-cluster and for the energy lost in inactive material upstream of the calorimeter. This allows the true energy of the incoming particle to be represented as accurately as possible. A BDT-based regression is trained in intervals of $|\eta|$ and E_{T} and independently for electrons, unconverted photons, and converted photons. The inputs are specifically designed to capture the geometry of the cluster such as the longitudinal and lateral spread of the shower, the ratio of raw energies across different layers, the η coordinate of cell edges, the periodicity of the lead absorber in each region or the radius of PCs. The BDT is trained on single-particle samples to predict the ratio $\frac{E^{\text{true}}}{E^{\text{raw}}}$ and bring it as close as possible to one, where E^{true} corresponds to the true energy of the particles. Once trained, it is applied at the reconstruction level to both data and MC simulations. The BDT thus produces a predicted correction factor that is multiplied by E^{raw} to yield the calibrated energy E.

3. Uniformity and ADC non-linearity corrections: after the MC-based calibration, residual non-uniformities in the calorimeter response remain due to imperfections in the detector construction, localised HV variations, and non-linearities in the front-end electronics.

   The dominant electronics effect arises from the conversion of ADC counts to ionisation current. In the nominal reconstruction this conversion is parametrised as a linear function of the injected charge (either from the incoming particle or artificially via the calibration board), fitted to dedicated calibration runs. The residuals of this linear fit are however not zero: they reach the percent level at low ADC counts and decrease at higher amplitudes. Since the cell energy enters all subsequent calibration steps, a non-zero residual translates into a non-linearity of the reconstructed energy as a function of E_{T}. The residuals are corrected by re-fitting the calibration data with a fifth-order polynomial in ADC counts and propagating the corresponding correction cell by cell.

   A second effect is associated with the multi-gain readout. Each LAr cell is read out simultaneously with three amplification stages and the most appropriate gain is selected on a cell-by-cell basis depending on the recorded amplitude. $\text{Z} \to \text{ee}$ events are dominated by the high-gain readout in Layer 2, while higher-energy processes have a sizeable medium-gain fraction. A relative miscalibration between medium and high gain would therefore bias the extrapolation from the Z in-situ calibration to higher-energy analyses. To probe this, dedicated low-threshold runs were recorded in 2017 and 2018 (totalling $\sim 0.3\,\text{fb}^{-1}$) in which the transition was lowered well below its nominal value, allowing the position of the Z peak to be compared between the special and standard configurations. After the ADC non-linearity correction, the residual bias is found to affect the energy response by approximately $0.1\,\%$ in the barrel and $0.4\,\%$ in the end-cap, and is propagated as a dedicated systematic uncertainty.

4. $\text{Z} \to \text{ee}$ resolution smearing: the previous steps of the calibration chain are designed to bring the reconstructed energy as close as possible to the true energy of the incoming particle. Residual discrepancies between data and simulation may nonetheless remain, originating from imperfect modelling of the detector, or from the EM layer inter-calibration itself. To address the resolution mismatch, $\text{Z} \to \text{ee}$ events are exploited: the invariant mass of the electron pair is reconstructed from the calibrated energies and compared, in bins of η, to the well-known Z lineshape, the width of which carries information on the energy resolution. Smearing parameters c are extracted from this comparison and applied to the simulation to reproduce the slightly degraded resolution observed in data, following the relation:

$$\left(\frac{\sigma(E)}{E}\right)^{\text{data}}_{i} = \left(\frac{\sigma(E)}{E}\right)^{\text{MC}}_{i} \oplus c_i\,, \tag{5.5}$$

   where i denotes the η bin and $\oplus$ indicates a quadratic sum.

For a pair of electrons falling in the η bins i and j, the miscalibration of the di-electron invariant mass resolution can be written as:

$$\left(\frac{\sigma(m_{\mathrm{ee}})}{m_{\mathrm{ee}}}\right)_{ij}^{\mathrm{data}} = \left(\frac{\sigma(m_{\mathrm{ee}})}{m_{\mathrm{ee}}}\right)_{ij}^{\mathrm{MC}} \oplus \frac{c_i \oplus c_j}{2}\,, \tag{5.6}$$

where the last term is often denoted c_{ij} and referred to as the *di-electron smearing parameter*. The c_{ij} are determined, independently in each (i,j) bin, as the values that best reproduce the observed m_{ee} distribution once the simulation has been smeared accordingly. A global fit across all bins is then performed to unfold the single-electron smearing parameters c_i from the di-electron c_{ij}, which are subsequently applied to the simulation.

The c_i are typically of the order of $1\,\%$ in the barrel and up to $2\,\%$ in the end-cap. They encapsulate the fact that the energy resolution observed in data is slightly worse than the one predicted by the simulation, an effect generally attributed to an underestimation of the passive material in the detector description.

5. $\mathrm{Z} \to \mathrm{ee}$ residual scale calibration and linearity: in parallel to the resolution smearing, the position of the reconstructed Z peak is compared between data and simulation. Any shift reveals a residual scale mis-modelling that is not captured by the preceding calibration steps. Global SFs α are extracted from this comparison and applied to data so that the reconstructed Z peak position aligns with the one in simulation. They are determined together with the resolution parameters c described above, by minimising a χ^2 between the observed m_{ee} distribution and a set of MC templates generated for different values of (α, c).

   The SFs relate the uncorrected energy in data to the simulated one, the corrected data energy being $\frac{E_i^{\mathrm{data}}}{1+\alpha_i}$:

$$E_i^{\mathrm{data}} = (1+\alpha_i)\, E_i^{\mathrm{MC}}\,, \tag{5.7}$$

   which propagates to the di-electron invariant mass as:

$$\left(m_{\mathrm{ee}}^{\mathrm{data}}\right)_{ij} \approx \left(m_{\mathrm{ee}}^{\mathrm{MC}}\right)_{ij} \left(1 + \frac{\alpha_i + \alpha_j}{2}\right)\,, \tag{5.8}$$

   where the half-sum, valid to first order in α, is denoted α_{ij} and referred to as the di-electron SF. As for the resolution, a global fit across all (i,j) bins is then performed to extract the single-electron α_i from the α_{ij}. Because the SFs enter the chain through the accordion energy that is in turn used by the MC-based calibration, the procedure is iterated a second time so that the closure SFs $\alpha^{\mathrm{closure}}$ obtained from the second iteration are statistically compatible with zero.

   The α_i are typically of the order of a percent, and the closure factors of the second iteration of the order of 10^{-4}, except in the scintillator region ($1.4 < |\eta| < 1.6$), where a fraction of the electron energy is collected by the Tile scintillators placed in the crack between the barrel and end-cap cryostats. The response of these scintillators is not described by the accordion calibration chain and is instead corrected through a dedicated SF, derived from the same $\mathrm{Z} \to \mathrm{ee}$ events but extracted independently of the α_i of the surrounding η bins. The corresponding correction is of the order of a few percent and carries a larger uncertainty than the rest of the η range, reflecting both the limited size of the scintillator sample and the imperfect modelling of the passive material in this region.

   Because the systematic uncertainties associated with the passive material and the layer calibration are themselves η- and E_{T}-dependent, $\mathrm{Z} \to \mathrm{ee}$ events are further used to probe any residual dependence of the energy scale on the electron E_{T} (the so-called *linearity fit*), providing additional constraints on the scale uncertainties that go beyond the inclusive Z measurement.

6. Photon leakage correction: the in-situ calibration steps described above are driven entirely by electrons from Z decays, and the extracted SFs are subsequently applied to photons under the assumption that the EM response of the calorimeter is identical for the two particle types. While the bulk of the energy lost outside the super-cluster is already corrected by the MC-based calibration, photon showers tend to be slightly broader laterally than electron showers of the same energy, owing to the absence of a track and the higher conversion probability in the upstream material, so that the fraction of energy contained in a fixed-size cluster is not identical for the two species. Any imperfect modelling of this difference in the simulation would therefore bias the electron-to-photon extrapolation of the calibration. To quantify this effect, a lateral-leakage observable is defined in the second accordion layer as:

$$l = \frac{E_{7\times11}^{\mathrm{L2}}}{E_{\mathrm{nom}}^{\mathrm{L2}}} - 1\,, \tag{5.9}$$

where $E^{\mathrm{L2}}_{7\times11}$ is the energy collected in a 7×11 cell window around the cluster barycentre and $E^{\mathrm{L2}}_{\mathrm{nom}}$ is the nominal cluster energy in the same layer. The data-to-MC difference of l is measured separately for electrons (using $\mathrm{Z} \to \mathrm{ee}$ events) and for photons (using radiative $\mathrm{Z} \to \ell\ell\gamma$ events), and a residual photon correction is built from their double difference:

$$\alpha_l = (l_{\mathrm{e}} - l_{\gamma})^{\mathrm{data}} - (l_{\mathrm{e}} - l_{\gamma})^{\mathrm{MC}} \,, \tag{5.10}$$

which is applied as a multiplicative correction to the photon energy. The double difference is constructed so that any common data/MC mis-modelling of the lateral shower width cancels and only the electron–photon difference remains. The values of α_l range from approximately $-0.3\,\%$ to $-0.2\,\%$ at low E_{T} down to $-0.1\,\%$ to $-0.05\,\%$ at high E_{T}, with absolute uncertainties between $0.01\,\%$ and $0.07\,\%$ depending on the E_{T} and η region.

7. Data-driven scale validation: the full calibration chain is finally validated using two independent processes that probe energy ranges and particle types that are not directly used in the in-situ extraction.

   Low-E_{T} electrons from $\mathrm{J}/\psi \to \mathrm{ee}$ decays (reconstructed in the range $7 < E_{\mathrm{T}} < 30\,\mathrm{GeV}$) provide a handle on the energy linearity well below the average $E_{\mathrm{T}} \approx 40\,\mathrm{GeV}$ of electrons from Z decays, where the Z-based calibration is necessarily extrapolated. A residual energy scale α' is measured by comparing the position of the J/ψ mass peak in data and simulation in bins of η and E_{T}, and is expected to be statistically compatible with zero if the E_{T}-dependence of the calibration is correctly modelled. The measured residuals are typically below a few per mille across the full η range, well within the systematic uncertainty band associated with the calibration model.

   The extrapolation from electrons to photons is similarly validated using radiative $\mathrm{Z} \to \ell\ell\gamma$ decays, in which a final-state photon is emitted by one of the two leptons. Because the parent Z mass is fixed, the kinematics of the $\ell\ell\gamma$ system constrain the photon energy without relying on the calorimeter measurement of the photon itself, providing a clean, in-situ probe of the photon energy scale. The residual photon scale measured this way is also found to be compatible with zero within uncertainties, both for converted and unconverted photons and across the full η acceptance.

   Beyond the simple closure check, the residual scales $\alpha'(\eta, E_{\mathrm{T}})$ extracted from $\mathrm{J}/\psi \to \mathrm{ee}$ and $\mathrm{Z} \to \ell\ell\gamma$ are also used in a global linearity fit, in which they are confronted with the prediction of the full systematic model through a χ^2 minimisation. This fit pulls and constrains the calibration nuisance parameters (NPs) and produces a post-fit covariance matrix that captures their correlations. The procedure typically reduces the energy scale uncertainty by $30\,\%$ to $50\,\%$ for electrons of $E_{\mathrm{T}} = 60\,\mathrm{GeV}$.

5.3 Identification

After the reconstruction and calibration steps, the electron and photon candidates are still contaminated by a large background of hadronic jets, which can produce EM showers in the calorimeter that mimic the signal. Three categories of background are particularly relevant: light-flavour (LF) jets that fluctuate into EM-like signatures, photons converting into electron–positron pairs in the detector material, and non-prompt electrons originating from the semileptonic decay of hadrons containing HF quarks. To suppress these contributions while retaining high efficiency for prompt electrons produced, for example, in Z- or W-boson decays, a powerful identification algorithm is employed.

The identification tools are based on a multivariate approach, developed independently for the different particle types (electron, unconverted photon, converted photon) and for different $|\eta|$ regions. The inputs to the multivariate discriminant include a wide range of variables that capture the physics properties of the particle and reflect its interaction with the detector.

5.3.1 The likelihood discriminant

Prompt electrons are selected using a likelihood (LH) discriminant. It is based on the product of n probability density functions (PDFs), P, which are formed by smoothing histograms of the n input variables via an adaptive kernel density estimation (KDE). This is done independently for signal and background, in nine bins of $|\eta|$ and seven bins of E_{T}.

For each electron candidate, two LHs are constructed:

$$L_{\mathrm{S(B)}}(\vec{x}) = \prod_{i=1}^{n} P_i^{\mathrm{S(B)}}(x_i) \,, \tag{5.11}$$

where $\vec{x}$ is the vector of input quantities, and P_i^{S} (P_i^{B}) is the signal (background) PDF for the i-th input variable. Correlations between the input quantities are neglected. The final discriminant is defined as:

$$d_L = \frac{L_{\mathrm{S}}}{L_{\mathrm{S}} + L_{\mathrm{B}}}\,, \tag{5.12}$$

and takes values between 0 and 1, with higher values indicating a higher compatibility with the signal hypothesis.

Because d_L exhibits sharp peaks at 0 and 1 for background and signal, respectively, the selection of operating points on such a distribution would require an extremely fine binning. For practical reasons, an inverse-sigmoid transformation is applied:

$$d_L' = -\tau^{-1} \ln\left(d_L^{-1} - 1\right)\,, \tag{5.13}$$

with $\tau = 15$ [224]. The transformed discriminant d_L' is no longer bounded to $[0, 1]$ and exhibits a smoother distribution on which threshold-based selections can be applied.

The advantages of using an LH discriminant over a selection-criteria-based approach (also known as *cut-based*) are numerous. First, a prompt electron may fail the cut-based selection because it does not satisfy the criterion for a single quantity, while the LH discriminant can still assign a high score if the other variables are consistent with the signal hypothesis. Second, the LH approach allows for a better combination of the input variables, taking into account their relative importance, which leads to better separation between signal and background and allows the inclusion of variables (such as R_ϕ or f_1) whose distributions overlap too much between signal and background to be effectively used in a cut-based approach. Finally, the LH discriminant provides a continuous output that can be used to define multiple working points (WPs) with different signal efficiencies, allowing for more flexibility in the analysis.

The PDFs used in the LH are derived from simulation. Two samples are used depending on the E_{T} regime: a $\mathrm{J}/\psi \to \mathrm{ee}$ sample provides the low-E_{T} range ($4.5 < E_{\mathrm{T}} < 15\,\mathrm{GeV}$), while a $\mathrm{Z} \to \mathrm{ee}$ sample is used for $E_{\mathrm{T}} > 15\,\mathrm{GeV}$.

The histograms of each input variable are smoothed using an adaptive kernel density estimation implementation available in TMVA, which avoids non-physical fluctuations arising from the limited size of the simulation samples.

Imperfect detector modelling induces residual differences between the simulated and observed distributions of several identification variables, which would bias the LH if left uncorrected. These differences typically appear either as a constant offset between data and simulation (a shift of the distribution) or as a difference in resolution, quantified by the full width at half maximum (FWHM) of the distribution. In some cases both effects are present simultaneously.

For the most affected variables, correction factors are derived in each $|\eta|$ bin using the $\mathrm{Z} \to \mathrm{ee}$ tag-and-probe method (explicitly described in Section 5.3.7). The simulated distribution of a given variable is either shifted by a constant offset or scaled by a multiplicative factor applied around its central value, such that it matches the data distribution after minimisation of a χ^2. Shift-only corrections are applied to R_η, $w_{\eta 2}$, and R_ϕ, width-scaling corrections are applied to R^{HAD}, $\Delta\eta_1$, and $\Delta\phi_{\mathrm{res}}$, while both shift and width-scaling corrections are applied to f_1 and f_3. The resulting corrected PDFs are then used both for training the LH and for comparing identification efficiencies between data and simulation.

5.3.2 Pile-up dependence

The PDFs of some discriminating variables, in particular the shower-shape variables R^{HAD} and R_η, depend on the amount of additional activity in the calorimeter due to pile-up interactions. To mitigate the associated efficiency dependence, the threshold on the transformed discriminant d_L' is parametrised as a linear function of the number of reconstructed primary vertices, N_{PV}. The linear slope is adjusted, in each bin of $|\eta|$ and E_{T}, such that the background efficiency remains approximately independent of N_{PV}. This comes at the price of a mild (at most $\sim 5\,\%$) decrease of the signal efficiency with increasing N_{PV}.

5.3.3 Input variables

The variables used as inputs to the LH discriminant are summarised in Table 5.A. They can be broadly categorised into three groups: calorimeter-based variables, which capture the shower shape and energy distribution in the calorimeter, track-based variables, which describe the properties of the associated track, and track–cluster matching variables, which quantify the consistency between the track and the calorimeter cluster. The table also indicates which background types each variable is most effective at rejecting: LF, PCs, and HF.

The calorimeter-based variables include the hadronic leakage, which measures the energy deposited in the HAD calorimeter relative to the EM cluster energy, and the shower-shape variables, which characterise the lateral and longitudinal spread of the energy deposit in the EM calorimeter. Track-based variables include the number of hits in the Si detectors and in the TRT, as well as the presence of a hit in the innermost pixel layer, which is particularly effective at rejecting PCs. Track–cluster matching variables include the distance between the track extrapolation and the cluster position in η and ϕ, as well as the ratio of cluster energy to track momentum ($\frac{E}{p}$). The latter variable is particularly powerful at rejecting hadronic jets, as genuine electrons typically have a ratio close to unity, while for hadronic jets the ratio tends to be significantly different due to the presence of neutral particles and the broader shower shape.

5.3.4 Binning

The PDFs entering the LH are built in nine regions of $|\eta|$ and seven bins of E_{T}. The $|\eta|$ bins are chosen according to the calorimeter geometry, the detector acceptance, and the variation of material in the ID, while the E_{T} bins reflect the evolution of the discriminating power of the input variables with particle energy, as well as the different rate and composition of the background. The boundaries of these bins are listed in Table 5.B.

To avoid discontinuities of the signal efficiency at bin boundaries, the threshold on the transformed discriminant d_L' is refined to 12 E_{T} bins, with the PDF values and discriminant requirements linearly interpolated between the centres of adjacent bins.

Table 5.B: Binning used for the construction of the identification LH.

Binning	Boundaries
$\|\eta\|$ bins (PDFs and discriminant)	$0.0,\ 0.6,\ 0.8,\ 1.15,\ 1.37,\ 1.52,\ 1.81,\ 2.01,\ 2.37,\ 2.47$
E_{T} bins for the PDFs [GeV]	$4.5,\ 7,\ 10,\ 15,\ 20,\ 30,\ 40,\ \infty$
E_{T} bins for the discriminant [GeV]	$4.5,\ 7,\ 10,\ 15,\ 20,\ 25,\ 30,\ 35,\ 40,\ 45,\ 80,\ 150,\ \infty$

5.3.5 High-E_{T} electrons

Two effects degrade the identification performance at high E_{T}. First, the f_3 variable loses discriminating power at large energies, because high-E_{T} electrons deposit a larger fraction of their energy in the third calorimeter layer, making them more similar to hadronic background candidates. Furthermore, f_3 is not well modelled by the simulation in this regime. For this reason, f_3 is removed from the LH for $E_{\mathrm{T}} > 80\,\mathrm{GeV}$.

Second, for $E_{\mathrm{T}} > 300\,\mathrm{GeV}$, the R^{HAD} and f_1 variables evolve in a way that reduces the signal efficiency of the LH. To mitigate this effect, for candidates with $E_{\mathrm{T}} > 150\,\mathrm{GeV}$ the threshold on d_L' for the `Tight` WP is set to the one of the `Medium` WP, and two additional cut-based requirements are added:

- the total shower width $w_{\text{S-tot}}$ in the first calorimeter layer must satisfy an η-dependent requirement
- the ratio of cluster energy to track momentum must satisfy $\frac{E}{p} < 10$, a value chosen conservatively to account for the degraded momentum resolution of tracks with p_{T} of several hundred GeV.

With these additional requirements, a good signal efficiency and background rejection are maintained for very high-E_{T} electrons.

5.3.6 Working points

Four thresholds are defined on the LH discriminant, corresponding to the so-called WPs labelled `VeryLoose`, `Loose`, `Medium`, and `Tight`, with increasing values of the threshold and decreasing signal efficiency. For an electron with $E_{\mathrm{T}} = 40\,\mathrm{GeV}$, the signal efficiencies are approximately $93\,\%$, $88\,\%$, and $80\,\%$ for `Loose`, `Medium`, and `Tight`, respectively. Compared to `Loose`, the background rejection factors for misidentified electrons from multijet production increase by factors of approximately 2.5 for `Medium` and 5 for `Tight`, over the E_{T} range of $4\,\mathrm{GeV}$ to $50\,\mathrm{GeV}$.

In addition to the LH threshold, each WP imposes requirements on the associated track:

- the `Loose`, `Medium`, and `Tight` WPs require at least two hits in the Pixel Detector and at least seven hits in the Si detectors (pixel + SCT)
- the `Medium` and `Tight` WPs additionally require a hit in the innermost pixel layer, or, if the innermost module is inoperative, in the next-to-innermost layer. This requirement is particularly effective in rejecting PCs

Table 5.A: Electron identification variables used in the LH and cut-based selections. The "Rejects" columns indicate discrimination against LF jets, PCs (γ), and HF backgrounds.

Type	Description	Name	Rejects			Usage
			LF	γ	HF	
Hadronic leakage	Ratio of E_{T} in the first layer of the HAD calorimeter to E_{T} of the EM cluster (used over the range $\lvert\eta\rvert < 0.8$ or $\lvert\eta\rvert > 1.37$)	R^{HAD1}	✓	✓	✗	LH
	Ratio of E_{T} in the HAD calorimeter to E_{T} of the EM cluster (used over the range $0.8 < \lvert\eta\rvert < 1.37$)	R^{HAD}	✓	✓	✗	LH
Third layer of EM calorimeter	Ratio of the energy in the third layer to the total energy in the EM calorimeter. This variable is only used for $E_{\mathrm{T}} < 80\,\mathrm{GeV}$, due to inefficiencies at high E_{T}, and is also removed from the LH for $\lvert\eta\rvert > 2.37$, where it is poorly modelled by the simulation.	f_3	✓	✗	✗	LH
Second layer of EM calorimeter	Lateral shower width, $\sqrt{\frac{\sum E_i \eta_i^2}{\sum E_i} - \left(\frac{\sum E_i \eta_i}{\sum E_i}\right)^2}$, 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	$w_{\eta 2}$	✓	✓	✗	LH
	Ratio of the energy in 3×3 cells over the energy in 3×7 cells centred at the electron cluster position	R_ϕ	✓	✓	✗	LH
	Ratio of the energy in 3×7 cells over the energy in 7×7 cells centred at the electron cluster position	R_η	✓	✓	✓	LH
First layer of EM calorimeter	Shower width, $\sqrt{\frac{\sum E_i (i - i_{\max})^2}{\sum E_i}}$, where i runs over all strips in a window of $\Delta\eta \times \Delta\phi \approx 0.0625 \times 0.2$, corresponding typically to 20 strips in η, and $i_{\max}$ is the index of the highest-energy strip, used for $E_{\mathrm{T}} > 150\,\mathrm{GeV}$ only	$w_{\mathrm{S\text{-}tot}}$	✓	✓	✓	Cut-based
	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	$E^{\mathrm{L1}}_{\mathrm{ratio}}$	✓	✓	✗	LH
	Ratio of the energy in the first layer to the total energy in the EM calorimeter	f_1	✓	✗	✗	LH
Track conditions	Number of hits in the innermost pixel layer	$n_{\mathrm{B\text{-}L}}$	✓	✗	✗	Cut-based
	Number of hits in the Pixel Detector	n_{pixel}	✓	✗	✗	Cut-based
	Total number of hits in the Pixel and SCT detectors	n_{Si}	✓	✗	✗	Cut-based
	Transverse impact parameter relative to the beam line	d_0	✓	✓	✗	LH
	Significance of transverse impact parameter defined as the ratio of d_0 to its uncertainty	$\left\lvert\frac{d_0}{\sigma_{d_0}}\right\rvert$	✓	✓	✗	LH
	Momentum lost by the track between the perigee and the last measurement point divided by the momentum at perigee	$\frac{\Delta p}{p}$	✓	✗	✗	LH
TRT	Likelihood probability based on TR in the TRT	$\mathcal{L}_{\mathrm{TRT}}$	✓	✗	✗	LH
Track–cluster matching	$\Delta\eta$ between the cluster position in the first layer and the extrapolated track	$\Delta\eta_1$	✓	✓	✗	LH
	$\Delta\phi$ between the cluster position in the second layer of the EM calorimeter and the momentum-rescaled track, extrapolated from the perigee, times the charge q	$\Delta\phi_{\mathrm{res}}$	✓	✓	✗	LH
	Ratio of the cluster energy to the track momentum, used for $E_{\mathrm{T}} > 150\,\mathrm{GeV}$ only	$\frac{E}{p}$	✓	✓	✗	Cut-based

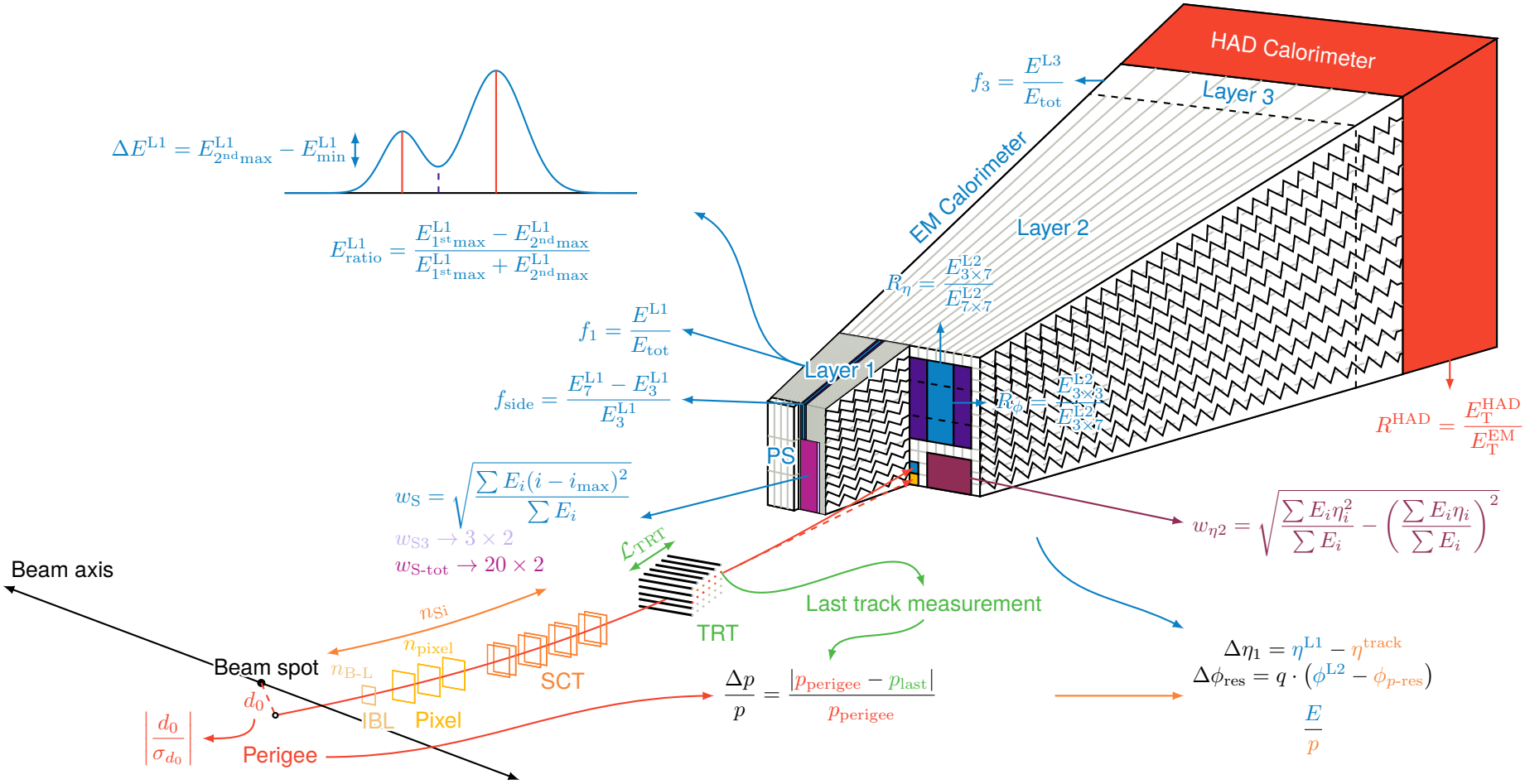


Figure 5.9: Electron and photon identification variables used.

- the `VeryLoose` WP only requires one pixel hit and does not impose any requirement on the innermost layer. It is mainly used for studies of the background
- a variant named `LooseAndBLayer` uses the `Loose` LH threshold combined with the requirement of a hit in the innermost pixel layer.

By construction, candidates satisfying a tighter WP also satisfy looser ones, which was not the case in Run-1 where different input variables were used across WPs.

In addition to the identification and tracking requirements listed above, analyses routinely impose requirements on the transverse and longitudinal impact parameters of the primary electron track, measured with respect to the reconstructed primary vertex. The standard requirements are $\left|\frac{d_0}{\sigma_{d_0}}\right| < 5$ and $|\Delta z_0 \sin\theta| < 0.5\,\mathrm{mm}$, which are effective at suppressing non-prompt electrons originating from displaced vertices, such as those from HF decays or from PCs in the detector material. These requirements are not part of the identification definition strictly speaking, but are applied on top of the chosen WP in ATLAS analyses.

The online electron identification used in the HLT follows the same LH (in Run-2) or DNN (in Run-3) approach as the offline selection, with a reduced set of inputs reflecting the central processing unit constraints of the trigger system. Variables such as $\frac{\Delta p}{p}$, which rely on the full GSF refit, are removed from the online discriminant, and the average number of interactions per bunch crossing $\langle\mu\rangle$ replaces N_{PV} in the pile-up parametrisation. The online WPs are designed to have efficiencies relative to the offline selection close to unity, with residual inefficiencies of at most a few percent at the `Tight` operating point.

5.3.7 Efficiency measurements and data/MC corrections

The simulation does not perfectly reproduce the identification efficiency measured in data, due to residual mis-modelling of tracking properties and of shower shapes in the calorimeters. Since this mis-modelling directly affects any analysis that relies on simulated samples to predict signal or background yields, correction factors, often referred to as SFs, are derived and applied to the simulation.

The correction factors are defined as the ratio of the efficiency measured in data to the one extracted from simulation, and are determined using tag-and-probe methods on two complementary resonances: $\mathrm{Z} \to \mathrm{ee}$ decays, which provide a clean high-purity sample of prompt electrons with $E_{\mathrm{T}} \gtrsim 15\,\mathrm{GeV}$, and $\mathrm{J}/\psi \to \mathrm{ee}$ decays, which extend the measurement down to $E_{\mathrm{T}} \approx 4.5\,\mathrm{GeV}$. The two measurements are combined through a χ^2 minimisation in the overlap region $15 < E_{\mathrm{T}} < 20\,\mathrm{GeV}$, and the correction factors are provided in bins of (E_{T}, η). For all WPs the correction factors are close to unity, with total uncertainties ranging from a few percent at low E_{T} to the per mille level for $E_{\mathrm{T}} = 40\,\mathrm{GeV}$.

The tag-and-probe method selects events containing a resonance decaying to two electrons and exploits one well-identified electron, the *tag*, to provide an unbiased sample of the second, the *probe*, whose efficiency for passing a given selection is then measured directly. The probe is required only to form, together with the tag, an invariant mass compatible with the resonance, so that its selection is independent of the criterion under study. The residual background under the mass peak is subtracted using the sidebands. The efficiency is the fraction of probes that pass the criterion, and the ratio of the values measured in data and simulation defines the SF.

5.3.8 Charge identification

The electric charge of an electron is determined from the curvature of the associated track reconstructed in the ID. The misidentification of the electron charge can arise either from the matching of an incorrect track to the electron candidate, typically when bremsstrahlung-induced conversions produce additional close-by tracks, or from a mis-measurement of the curvature of the primary electron track, which becomes more probable at high E_{T} and at large $|\eta|$ where the amount of traversed material is largest. The associated charge-misidentification rate ranges from a few per mille in the central region of the detector to a few percent at $|\eta| > 2$, and constitutes a non-negligible background for analyses that select same-sign or opposite-sign dilepton final states.

To suppress charge-misidentified electrons, a dedicated BDT has been trained on a simulated single-electron sample, using a set of eight input variables related to the quality of the primary electron track, the lateral development of the EM shower, and combined tracking and calorimeter quantities. The BDT parameters are optimised to maximise the rejection of wrong-charge electrons while retaining a signal efficiency of approximately $97\,\%$ for correctly reconstructed ones. The inclusion of the BDT requirement reduces the charge-misidentification rate by approximately a factor of three at high $|\eta|$, bringing it below $1\,\%$ over most of the detector acceptance. The corresponding data-to-simulation SFs are measured on $\mathrm{Z} \to \mathrm{ee}$ events using an LH-based method that extracts the misidentification probability from the number of same-sign and opposite-sign electron pairs, and are typically compatible with unity within a few percent,

with larger deviations at very high $|\eta|$ due to residual mis-modelling of the shower-shape distributions in the simulation.

5.3.9 DNN-based identification

The LH discriminant described above neglects, by construction, the correlations between its input variables. At the typical operating points used in ATLAS analyses this is a non-negligible source of lost separation power, in particular for the rejection of non-prompt electrons from HF decays, where several identification variables are simultaneously shifted away from their prompt-electron values. A dedicated DNN-based identification has been developed to exploit these correlations and replaces the LH as the baseline electron identification for Run-3 analyses [295].

The DNN is a fully connected feed-forward network with a multi-class output. In addition to the prompt electron class, four background classes are defined: LF jets, PCs, electrons from HF decays, and electrons from charge-flip processes. The multi-class structure allows the network to learn a distinct internal representation for each background type, which is the main origin of the improved HF rejection compared with the LH. The output nodes are interpreted as posterior probabilities via a softmax activation, and the final discriminant used to define the WP is built from a log-ratio of the prompt-electron probability to a weighted sum of the background probabilities.

The network takes as input essentially the same quantities as the LH (see Table 5.A), complemented by a few additional track-quality and track–cluster matching variables which were previously used only as cut-based requirements. The kinematic variables E_{T} and η are also provided as inputs, such that the network can learn an E_{T}- and η-dependent combination of the discriminating quantities without the need for an explicit binning. This removes the discontinuities at bin boundaries that are present in the LH and eliminates the need for the linear interpolation procedure described in Section 5.3.4.

Training is performed on simulated samples enriched in each of the five classes, with class weights chosen to reflect the relative populations expected in a typical analysis selection. The same shift-and-width corrections used for the LH PDFs are applied to the input distributions, ensuring a consistent data–simulation agreement for all identification variables.

At fixed signal efficiency, the DNN improves the background rejection by a factor of 2 to 5 with respect to the LH, with the largest gains obtained at low E_{T} and for candidates originating from HF decays. For an electron with $E_{\mathrm{T}} = 40\,\mathrm{GeV}$ and a signal efficiency of $80\,\%$ (corresponding to the `Tight` WP), the HF rejection is improved by approximately a factor of three, while the rejection of LF backgrounds improves by a factor of 1.5 to 2. As for the LH, four WPs (`VeryLoose`, `Loose`, `Medium`, `Tight`) are defined with the same nominal signal efficiencies, allowing a transparent replacement of the LH in existing analyses.

Data-to-simulation correction factors for the DNN are extracted using the same tag-and-probe methods as for the LH, on $\mathrm{Z} \to \mathrm{ee}$ and $\mathrm{J}/\psi \to \mathrm{ee}$ events. The correction factors are close to unity over the full kinematic range, with uncertainties of a few percent at low E_{T} decreasing to the per mille level at $E_{\mathrm{T}} = 40\,\mathrm{GeV}$, comparable to or slightly larger than those of the LH at low E_{T} due to the larger number of input variables that must be well modelled in simulation. The dominant systematic uncertainties arise from the modelling of the shower-shape variables and of the tracking quantities in the simulation, and are propagated by varying the corresponding shift-and-width corrections within their measured uncertainties.

For Run-4, further improvements to the DNN identification are under study, including the use of calorimeter cell-level information as additional input (in place of, or in addition to, the high-level shower-shape variables).

5.4 Isolation

Beyond identification, a second powerful handle to reject background is the requirement that the candidate be isolated from other activity in the event. Prompt electrons and photons produced in the decay of EW bosons are typically well separated from other final-state particles, whereas hadronic jets faking a lepton, PCs, and non-prompt electrons from HF decays are usually surrounded by substantial hadronic activity in the calorimeter and additional tracks in the ID. This complementarity with the identification-level variables provides additional discrimination that is particularly effective in busy environments, such as events containing top quarks, or for boosted topologies where several objects are reconstructed in close angular proximity.

Two complementary isolation variables are constructed: a calorimeter-based one, built from neighbouring calorimeter energy deposits, and a track-based one, built from the transverse momenta of nearby tracks. They are sensitive to different aspects of the background and are often combined in the analysis-level WPs to maximise rejection while controlling the dependence on pile-up.

5.4.1 Calorimeter-based isolation

The calorimeter-based isolation is built within a cone of radius $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ around the candidate cluster direction. A typical cone size of $\Delta R = 0.2$ is used, while a larger cone of $\Delta R = 0.4$ is also available for analyses that benefit from a wider integration range.

The raw isolation energy, $E_{\mathrm{T,raw}}^{\mathrm{isol}\Delta R}$, is defined as the scalar sum of the transverse energies of all positive-energy topological clusters whose barycentre falls within the cone. Topological clusters are 3D ensembles of calorimeter cells, seeded on cells whose deposited energy exceeds four times the expected noise level (including pile-up contributions), and grown iteratively by merging neighbouring cells satisfying a looser threshold. The use of topological clusters in place of raw cells, which was introduced in Run-2, provides a natural noise-suppression mechanism and substantially improves the agreement between data and simulation. The topological clusters are kept at the EM energy scale, with no candidate-specific calibration applied.

The raw isolation energy includes the candidate's own contribution, which must be subtracted. This is done by removing the cells contained in a rectangular window of size $\Delta\eta \times \Delta\phi = 0.125 \times 0.175$, corresponding to 5×7 cells of the second calorimeter layer, centred on the candidate. This is represented in Figure 5.10. The width of this core region is motivated by the typical lateral spread of the EM shower. Because part of the candidate's energy can leak outside of this rectangular window, an additional leakage correction $E_{\mathrm{T,leakage}}(E_{\mathrm{T}}, \eta)$ is applied. This correction is derived from simulated samples of single electrons or photons, without pile-up, by fitting the distribution of the energy leaking into the isolation cone as a function of E_{T} with a Crystal Ball function [296]. The most probable value of the fit is used as the average leakage, parametrised in ten bins of $|\eta|$. The leakage correction is typically of the order of $1\,\%$ to $5\,\%$ of the candidate energy.

The contribution from pile-up interactions and from the underlying event is estimated on an event-by-event basis using an ambient energy-density approach. For each event, positive-energy topological clusters in the full calorimeter acceptance ($|\eta| < 5$) are re-clustered using the k_t [297, 298] jet algorithm with a radius parameter $R = 0.5$ and no p_{T} threshold. This algorithm is detailed in Paragraph 6.2.1.2.3.

The area A of each resulting jet is estimated, and the transverse energy density of each jet is computed as $\rho = \frac{p_{\mathrm{T}}}{A}$ [299]. The median ρ_{median} of the distribution of jet densities is used as a robust estimator of the transverse energy density of the event, separately in the central ($|\eta| < 1.5$) and forward ($1.5 < |\eta| < 3.0$) regions. The corresponding correction is:

$$E_{\mathrm{T,pile\text{-}up}}(\eta, \Delta R) = \rho_{\mathrm{median}}(\eta) \times \left(\pi \Delta R^2 - A_{\mathrm{core}}\right) , \tag{5.14}$$

where A_{core} is the area of the rectangular core region subtracted from the cone. For a typical $\mathrm{Z} \to \mathrm{ee}$ sample at $\sqrt{s} = 13\,\mathrm{TeV}$ and $\langle\mu\rangle \approx 22$, ρ_{median} is of the order of a few GeV per unit of η-ϕ area.

The fully corrected calorimeter isolation is then:

$$E_{\mathrm{T}}^{\mathrm{cone}\Delta R} = E_{\mathrm{T,raw}}^{\mathrm{isol}\Delta R} - E_{\mathrm{T,core}} - E_{\mathrm{T,leakage}}(E_{\mathrm{T}}, \eta) - E_{\mathrm{T,pile\text{-}up}}(\eta, \Delta R) . \tag{5.15}$$

Because only positive-energy clusters enter the sum, the distribution of $E_{\mathrm{T}}^{\mathrm{cone}\Delta R}$ is biased slightly positive, reflecting the asymmetric response of the noise-suppression scheme.

5.4.2 Track-based isolation

The track-based isolation variable is constructed from charged-particle tracks reconstructed within the ID acceptance ($|\eta| < 2.5$) that satisfy a set of quality criteria calibrated on simulated $\mathrm{t\bar{t}}$ samples. The selected tracks must have $p_{\mathrm{T}} > 1\,\mathrm{GeV}$, a minimum number of hits in the Si detectors, and a maximum number of inoperable modules crossed by the trajectory.

To limit the impact of pile-up, tracks must originate from the reconstructed primary vertex of the event, defined as the vertex with the largest $\sum p_{\mathrm{T}}^2$ of its associated tracks. This is enforced through a requirement on the longitudinal impact parameter, corrected for the position of the primary vertex $|\Delta z_0 \sin\theta| < 3\,\mathrm{mm}$, which is particularly effective at suppressing soft tracks from out-of-time and in-time pile-up interactions. As a consequence, track-based isolation variables are largely insensitive to pile-up, in contrast to their calorimetric counterparts.

The contribution of the candidate's own track (and of secondary tracks generated by bremsstrahlung from the primary electron) is excluded by extrapolating all selected tracks to the second calorimeter layer and removing those falling within a window $\Delta\eta \times \Delta\phi = 0.05 \times 0.1$ around the candidate's cluster position.

The sum of the transverse momenta of the remaining tracks defines the track-based isolation $p_{\mathrm{T}}^{\mathrm{cone}\Delta R}$. Because the granularity of the tracker is much finer than that of the calorimeter, the cone radius can be

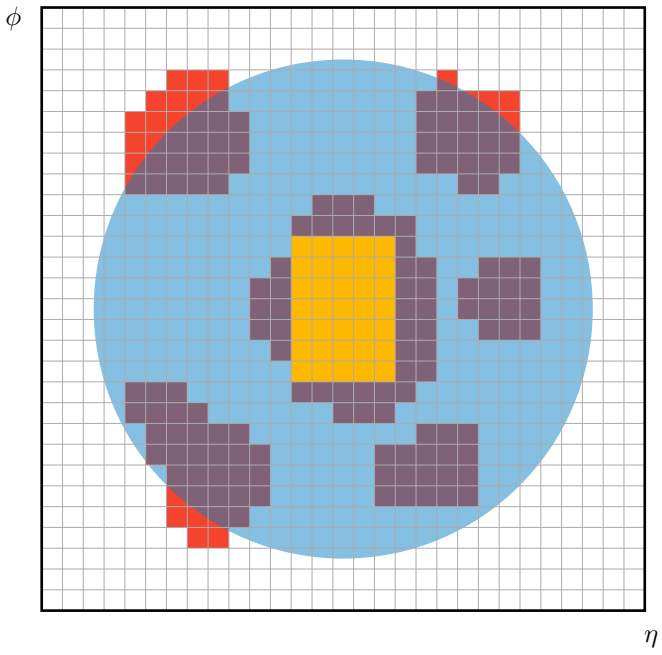


Figure 5.10: Schematic representation of the calorimeter-based isolation construction.

chosen significantly smaller than $\Delta R = 0.2$. In particular, a variable-cone version of the track isolation, $p_{\mathrm{T}}^{\mathrm{varcone}\Delta R_{\mathrm{max}}}$, is defined with a radius that shrinks with increasing candidate p_{T}:

$$\Delta R = \min\left(\frac{10\,\mathrm{GeV}}{p_{\mathrm{T}}},\ \Delta R_{\mathrm{max}}\right), \tag{5.16}$$

where ΔR_{max} is typically 0.2 or 0.4. This construction mitigates the efficiency loss observed for electrons produced in the decay of heavy resonances (top quarks or boosted $\mathrm{W/Z}$ bosons), where accompanying hadronic activity can lie within a fixed cone even for a truly prompt candidate.

5.4.3 Operating points

The isolation requirements are tailored to the needs of individual analyses: precision measurements at low p_{T} typically favour tighter operating points with strong background rejection, while high-p_{T} searches favour looser selections that preserve high signal efficiency in busy environments. Three classes of WPs are defined:

- operating points targeting a fixed isolation efficiency $\varepsilon_{\mathrm{iso}}$, uniform in both E_{T} and η (labelled `Loose` in the 2016–2017 nomenclature)
- operating points with E_{T}-dependent efficiency but uniform in η (labelled `Gradient`)
- operating points defined by fixed requirements on the isolation variables (labelled `Fix` in the 2016–2017 nomenclature, and `Loose`, `Tight`, `TightTrackOnly`, and `HighPtCaloOnly` in the final Run-2 recommendations used in Section 6.6.2).

The specific thresholds are derived in bins of (E_{T}, η) with fine E_{T} granularity between $7\,\mathrm{GeV}$ and $60\,\mathrm{GeV}$.

The main electron isolation WPs used in ATLAS analyses are summarised in Table 5.C. They combine calorimeter isolation with cone radius $\Delta R = 0.2$ and track isolation with variable cone $\Delta R_{\mathrm{max}} = 0.2$ (or $\Delta R_{\mathrm{max}} = 0.4$ for analyses with large accompanying hadronic activity). Additional calorimeter-only or track-only WPs are also provided. In the variable names, the number gives the cone size in units of 0.01 in ΔR: $E_{\mathrm{T}}^{\mathrm{cone20}}$ uses $\Delta R = 0.2$ and $p_{\mathrm{T}}^{\mathrm{varcone20}}$ uses $\Delta R_{\mathrm{max}} = 0.2$.

Recent developments introduced for Run-3 analyses have extended the isolation toolkit beyond the classical variables described above. First, particle flow (PFlow) isolation variants have been developed [300], in which both the calorimeter and track components are built from PFlow objects rather than from topological clusters and tracks treated separately. This construction improves the pile-up robustness of the calorimeter isolation by effectively removing the energy deposits already associated with charged tracks, and reduces the residual dependence on the ambient energy-density correction. Second, multivariate combinations of the classical isolation variables with additional track and vertex quantities,

Table 5.C: Definition of the main electron isolation WPs, with the final Run-2 names for the fixed-cut WPs. The rows marked with an asterisk follow the 2016–2017 nomenclature and were not retained in the final Run-2 recommendations.

Working point	$E_{\mathrm{T}}^{\mathrm{cone20}}$	$p_{\mathrm{T}}^{\mathrm{varcone20}}$	Total $\varepsilon_{\mathrm{iso}}$
`Loose (Track Only)`*	—	$\varepsilon_{\mathrm{iso}} = 99\,\%$	$99\,\%$
`Loose`*	$\varepsilon_{\mathrm{iso}} = 99\,\%$	$\varepsilon_{\mathrm{iso}} = 99\,\%$	$98\,\%$
`Gradient`	$\varepsilon_{\mathrm{iso}} = 0.1143\,\% \cdot \mathrm{GeV}^{-1} \times p_{\mathrm{T}} + 92.14\,\%$		$90\,\%$ to $99\,\%$ at $25\,\mathrm{GeV}$ to $60\,\mathrm{GeV}$
`Gradient (Loose)`*	$\varepsilon_{\mathrm{iso}} = 0.057\,\% \cdot \mathrm{GeV}^{-1} \times p_{\mathrm{T}} + 95.57\,\%$		$95\,\%$ to $99\,\%$ at $25\,\mathrm{GeV}$ to $60\,\mathrm{GeV}$
`Loose`	$\frac{E_{\mathrm{T}}^{\mathrm{cone20}}}{E_{\mathrm{T}}} < 0.20$	$\frac{p_{\mathrm{T}}^{\mathrm{varcone20}}}{E_{\mathrm{T}}} < 0.15$	—
`Tight`	$\frac{E_{\mathrm{T}}^{\mathrm{cone20}}}{E_{\mathrm{T}}} < 0.06$	$\frac{p_{\mathrm{T}}^{\mathrm{varcone20}}}{E_{\mathrm{T}}} < 0.06$	—
`TightTrackOnly`	—	$\frac{p_{\mathrm{T}}^{\mathrm{varcone20}}}{E_{\mathrm{T}}} < 0.06$	—
`HighPtCaloOnly`	$E_{\mathrm{T}}^{\mathrm{cone20}} < \max(0.015 \times p_{\mathrm{T}}, 3.5\,\mathrm{GeV})$	—	—
`Fix (Track` $\Delta R_{\mathrm{max}} = 0.4$)*	$\frac{E_{\mathrm{T}}^{\mathrm{cone20}}}{E_{\mathrm{T}}} < 0.11$	$\frac{p_{\mathrm{T}}^{\mathrm{varcone40}}}{E_{\mathrm{T}}} < 0.06$	—

known as the prompt-lepton veto (PLV) [301] and, in its most recent iteration, the prompt-lepton improved veto (PLIV) [302], have been introduced. These discriminants combine calorimeter and track isolation, impact-parameter significances, and information on nearby tracks in a BDT or a DNN, providing a substantially better rejection of non-prompt electrons from HF decays than the classical isolation WPs alone. The PLV/PLIV discriminants are now the default for Run-3 analyses requiring strong non-prompt-lepton rejection, such as multi-lepton searches and measurements in top-quark-rich final states.

5.4.4 Efficiency measurements and scale factors

As for identification, residual differences between data and simulation in the isolation distributions must be corrected to ensure accurate modelling of signal and background in physics analyses. These differences are due to imperfect modelling of the underlying event, of the lateral shower development, and of the calorimeter and tracker responses. SFs are derived as the ratio of the isolation efficiency measured in data to the one measured in simulation, in bins of (E_{T}, η).

For electrons, the efficiency is measured with a tag-and-probe method on $\mathrm{Z} \to \mathrm{ee}$ events. The probe is required to satisfy the chosen identification WP (typically `Tight`), and the efficiency is defined as the fraction of probes that additionally pass the isolation requirement. Background contamination is estimated using sideband subtraction around the Z-boson mass peak. Over most of the kinematic range, agreement between data and simulation is within $5\,\%$, and the SF uncertainties range from a few percent at low E_{T} to below $0.5\,\%$ for $E_{\mathrm{T}} > 50\,\mathrm{GeV}$. Above $E_{\mathrm{T}} = 500\,\mathrm{GeV}$, extrapolations up to $2\,\mathrm{TeV}$ are applied, with uncertainties that can reach $40\,\%$.

A further advantage of the track-based isolation variables, which is particularly relevant in the high-pile-up conditions of Run-3 and Run-4, is their robustness against additional interactions. Thanks to the primary-vertex association requirement, the data/MC agreement of the track-based isolation efficiency remains stable with increasing $\langle\mu\rangle$, while the calorimeter-based isolation exhibits a mild but non-negligible dependence that must be carefully reproduced in the simulation. For analyses carried out under HL conditions, such as the HL-LHC programme, track-based isolation is therefore expected to remain the workhorse variable, while the calorimeter isolation will rely on dedicated pile-up mitigation techniques.

5.5 Photon identification and isolation

The reconstruction and calibration steps described earlier, combined with the ambiguity resolver, provide a clean sample of photon candidates, which are nevertheless still contaminated by a significant background from hadronic jets. The dominant sources of fake photons are neutral mesons (mostly $\pi^0 \to \gamma\gamma$) produced in the fragmentation of hadronic jets, whose decay photons are too collimated to be resolved as two separate showers, and QCD jets depositing a large fraction of their energy in the EM calorimeter. A dedicated identification is therefore applied to suppress these backgrounds.

Photon identification in ATLAS [303, 304] is based on a set of rectangular requirements on shower-shape variables, computed independently for converted and unconverted photon candidates. The variables used are very similar to those entering the electron LH identification (see Table 5.A): the hadronic leakage R^{HAD}, the lateral shower-width variables $w_{\eta 2}$, R_η, and R_ϕ in the second EM calorimeter layer, and, for the `Tight` WP, the fine-granularity variables in the strip layer: the total shower width $w_{\mathrm{S\text{-}tot}}$ computed over 20 strips in η and the core shower width w_{S3} computed over the three central strips, the lateral energy fraction

f_{side} defined as the energy in the seven central strips outside the three central ones, relative to the energy in the three central strips, the ratio $E_{\text{ratio}}^{\text{L1}}$ between the energy difference of the highest and second-highest local maxima and their sum, and the energy difference ΔE^{L1} between the second local maximum and the minimum energy deposit found between the two maxima. These variables exploit the fine η-segmentation of the first calorimeter layer to resolve the two close-by photons from a π^0 decay, which would otherwise mimic a single prompt photon. Those additional variables with respect to the electron identification are also visually represented in Figure 5.9.

Two WPs are defined: `Loose`, which uses only the hadronic-leakage and middle-layer variables and reaches an efficiency of $\varepsilon_{\text{id}} \approx 99\,\%$ for $E_{\text{T}} > 25\,\text{GeV}$, and `Tight`, which adds the strip-layer variables and reaches an efficiency of $\varepsilon_{\text{id}} \approx 85\,\%$ for the same E_{T} range, with a background rejection factor of the order of 5000. Because the detector response to photons, in particular the development of EM showers in the first calorimeter layer, is imperfectly modelled by the simulation, the shower-shape distributions used by the identification are corrected on a variable-by-variable basis via fudge factors. These corrections originally consisted of a simple shift of the simulated distribution, whereas a combined shift-and-scale is now applied, which provides a substantially better description of the data.

The identification efficiency is measured in-situ using three complementary methods that cover different E_{T} ranges. Radiative $\text{Z} \to \ell\ell\gamma$ decays provide a pure sample of prompt photons ($P \approx 95\,\%$ to $99\,\%$) at low E_{T}, typically below $50\,\text{GeV}$. An extrapolation-based method uses electrons from $\text{Z} \to \text{ee}$ decays selected with the tag-and-probe technique and transforms their shower shapes into photon-like distributions via a Smirnov transform. Finally, an inclusive photon sample, in which the residual jet contamination is estimated via the ABCD method in a 2D plane of photon identification and isolation, is used to cover the high-E_{T} range up to $1.5\,\text{TeV}$. The three measurements are combined through a χ^2 minimisation, providing data-to-simulation SFs with uncertainties ranging from a few percent at low E_{T} to the per mille level for E_{T} above $100\,\text{GeV}$.

Isolation requirements complement the identification in rejecting non-prompt and jet backgrounds. Photon isolation uses the same calorimeter-based and track-based variables as for electrons, with the same core-subtraction, leakage, and pile-up correction procedures described in Section 5.4, but adopts simpler, cut-based operating points. Three WPs, `Loose`, `Tight`, and `TightCaloOnly`, are defined and summarised in Table 5.D. The `TightCaloOnly` WP, which omits any track-isolation requirement, is used in analyses where the photon tracking information may be unreliable, for example close to conversion vertices where the conversion status makes the track-isolation ambiguous.

Table 5.D: Definition of the photon isolation WPs.

Working point	Calorimeter isolation	Track isolation
`Loose`	$E_{\text{T}}^{\text{cone20}} < 0.065 \times E_{\text{T}}$	$\frac{p_{\text{T}}^{\text{cone20}}}{E_{\text{T}}} < 0.05$
`Tight`	$E_{\text{T}}^{\text{cone40}} < 0.022 \times E_{\text{T}} + 2.45\,\text{GeV}$	$\frac{p_{\text{T}}^{\text{cone20}}}{E_{\text{T}}} < 0.05$
`TightCaloOnly`	$E_{\text{T}}^{\text{cone40}} < 0.022 \times E_{\text{T}} + 2.45\,\text{GeV}$	—

Since no analogous resonance is available for photons, the corresponding efficiencies are measured by combining two complementary methods. Radiative $\text{Z} \to \ell\ell\gamma$ decays provide a clean sample of prompt photons at low E_{T}, while an inclusive photon sample extends the measurement to higher E_{T} by modelling the jet-background contamination through an isolation template fit. In the latter, the photon isolation distribution in data is fitted to a linear combination of a signal template, extracted from tight, non-isolated photons in simulation, and a background template, derived from photon candidates failing a subset of the identification criteria. The resulting efficiencies range from $70\,\%$ to $95\,\%$ depending on the WP and E_{T}, with SF uncertainties of $2\,\%$ to $5\,\%$ below $E_{\text{T}} = 25\,\text{GeV}$ and of $1\,\%$ to $2\,\%$ above.

5.6 Forward electrons

The previous sections detailed the full workflow for electrons and photons in the central region of the ATLAS detector, meaning within the track coverage of the ID: $|\eta| < 2.5$. Above, where no tracks are accessible for charged particles, we define the forward regions as $2.5 < |\eta| < 4.9$. In addition to the lack of tracking, the calorimeter coverage is also reduced in the forward region, with only the EMEC and HEC providing measurements up to $|\eta| = 3.2$, and the FCal covering the range $3.1 < |\eta| < 4.9$ with a coarser granularity than the central calorimeters. The reconstruction and identification of electrons in this region therefore rely solely on calorimetric information, and are consequently more challenging than in the central region [305]. The disentanglement of electrons and photons is impossible, and separation

from background is more difficult. A minimum $E_{\mathrm{T}} > 20\,\mathrm{GeV}$ is therefore imposed on the forward electron candidates used by the identification and by the analyses, above the $5\,\mathrm{GeV}$ at which their reconstruction starts, and the achievable background rejection is substantially lower than in the central region, with correspondingly larger efficiency uncertainties.

Figure 5.2 shows a schematic representation of the reconstruction and identification steps for electrons and photons in the central and forward regions of the ATLAS detector. For forward electrons, only the first steps with the selection of topo-clusters based on E_{T} are achievable.

The reconstruction in these regions is based on topological clusters: the cells with deposited energy significantly above the noise level are grouped in 3D in an iterative procedure, starting from seed cells, following the "4-2-0" scheme detailed in Section 5.1.1. The number of cells in the cluster is not fixed and the sum of their energies defines the energy of the cluster, with corrections made to account for energy losses in the passive material in front of the calorimeters. The direction of the forward electron candidates is defined by the barycentre of the cells belonging to the cluster. The efficiency of the cluster reconstruction is better than $99\,\%$ for $E_{\mathrm{T}} > 20\,\mathrm{GeV}$, as determined from simulation. An electron candidate in the forward region is reconstructed only if it has a transverse energy of $E_{\mathrm{T}} > 5\,\mathrm{GeV}$ and has only a small energy component in the HAD calorimeters.

5.6.1 Cluster moments

Forward electron identification exploits cluster moments, scalar shape descriptors derived from the energy-weighted distribution of cell energies within a cluster. The n-th moment of a cell variable x is defined as:

$$\langle x^n \rangle = \frac{1}{E_{\mathrm{norm}}} \sum_{\{i \,|\, E_i > 0\}} E_i \, x_i^n \,, \qquad E_{\mathrm{norm}} = \sum_{\{i \,|\, E_i > 0\}} E_i \,, \tag{5.17}$$

where the sum runs over cells with strictly positive energy only. To define longitudinal and transverse shower-shape moments, a shower axis $\hat{s}$ is needed. It is obtained from a principal-component analysis of the energy-squared-weighted covariance matrix of cell positions:

$$C_{\alpha\beta} = \frac{1}{w} \sum_{\{i \,|\, E_i > 0\}} E_i^2 \, (\alpha_i - \langle \alpha \rangle) \, (\beta_i - \langle \beta \rangle) \,, \qquad w = \sum_{\{i \,|\, E_i > 0\}} E_i^2 \,, \tag{5.18}$$

for $(\alpha, \beta) \in \{x, y, z\}$. The eigenvector of the resulting symmetric 3×3 matrix C closest to the IP-to-cluster-centre direction is taken as $\hat{s}$. If it deviates by more than $20°$, the IP-to-cluster-centre direction is used directly. Given the shower centre $\vec{c}$, the transverse and longitudinal cell distances are:

$$r_i = |(\vec{x}_i - \vec{c}) \times \hat{s}| \,, \qquad \lambda_i = (\vec{x}_i - \vec{c}) \cdot \hat{s} \,. \tag{5.19}$$

They are represented schematically in Figure 5.11 for a given cell i and the cluster barycentre $\vec{c}$, with the shower axis $\hat{s}$ defining the longitudinal direction.

The second moments $\langle r^2 \rangle$ and $\langle \lambda^2 \rangle$ quantify the lateral and longitudinal shower widths, respectively, while the shower depth $\lambda_{\mathrm{centre}}$ measures the distance of the cluster barycentre from the calorimeter front face along $\hat{s}$. They are also visually represented in Figure 5.11.

Normalised variants are formed by assigning reference distances to the two most energetic cells, every other cell keeping its own distance r_i: for the normalised lateral moment the two leading cells enter lat_2 with $r = 0$ and $\mathrm{lat}_{\mathrm{max}}$ with $r = 4\,\mathrm{cm}$, giving:

$$\mathrm{lateral} = \frac{\mathrm{lat}_2}{\mathrm{lat}_2 + \mathrm{lat}_{\mathrm{max}}} \,, \tag{5.20}$$

analogously:

$$\mathrm{longitudinal} = \frac{\mathrm{long}_2}{\mathrm{long}_2 + \mathrm{long}_{\mathrm{max}}} \,, \tag{5.21}$$

with a reference scale of $10\,\mathrm{cm}$. Both normalised moments take values in $[0, 1]$.

5.6.2 Cut-based identification

The six discriminating variables employed for forward electron identification are summarised in Table 5.E. Three reference WPs, `Loose`, `Medium`, and `Tight`, apply sequential cuts on subsets of these moments: the `Loose` selection uses $\langle \lambda^2 \rangle$, $\langle r^2 \rangle$, and $\lambda_{\mathrm{centre}}$, while the `Tight` selection further requires f_{max}, the normalised lateral moment, and the normalised longitudinal moment, achieving a rejection factor two to three times higher than `Loose` at some cost to efficiency.

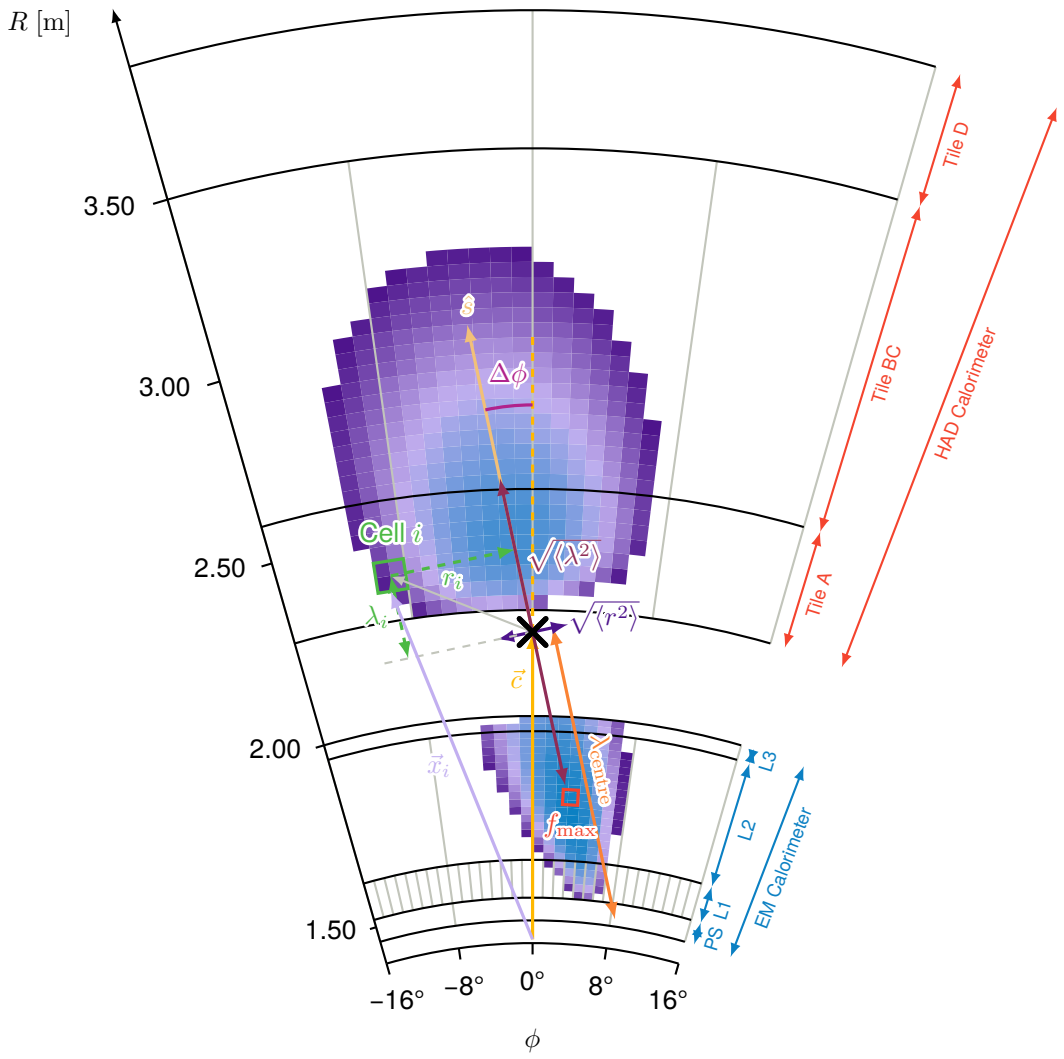


Figure 5.11: Schematic representation of geometrical moments for topo-clusters.

Table 5.E: Discriminating variables for forward electron identification, based on topological cluster moments.

Description	Symbol
Fraction of cluster energy in most energetic cell	f_{max}
Normalised longitudinal moment	longitudinal
Second longitudinal moment	$\langle\lambda^2\rangle$
Normalised lateral moment	lateral
Second transverse moment	$\langle r^2\rangle$
Shower depth along axis from calorimeter front face	λ_{centre}

Due to the harsher pile-up environment at high $|\eta|$, the cut thresholds are optimised directly on data in nine $|\eta|$ bins (six in the EMEC, $2.5 < |\eta| < 3.16$, and three in the FCal, $3.35 < |\eta| < 4.9$, excluding the inter-calorimeter transition $3.16 < |\eta| < 3.35$) and in four bins of the number of reconstructed primary vertices N_{PV} (1–3, 4–6, 7–10, 11–∞), ensuring a stable efficiency across pile-up conditions.

5.6.3 Likelihood-based identification

An LH discriminant was developed for forward electrons for Run-2, exploiting the same six cluster moments but combining them into a single score with greater discriminating power than the cut-based approach, particularly under the higher pile-up conditions of Run-2. The discriminant is similar in its construction to the one for the central region in Section 5.3.1.

5.6.4 Efficiency measurements and scale factors

The forward electron identification efficiency is measured using a $Z \to ee$ tag-and-probe technique. A central electron passing the `Tight` WP with $E_T > 25\,\text{GeV}$ and $|\eta| < 2.47$ serves as the tag. The forward cluster with $E_T > 20\,\text{GeV}$ in $2.5 < |\eta| < 4.9$ serves as the probe. The tag–probe invariant mass is fitted in the range $55 < m_{ee} < 130\,\text{GeV}$ to a Crystal Ball function convolved with a non-relativistic Breit–Wigner [306] (signal) and a Landau [307] function (background), with signal-to-background ratios of approximately seven and five in the EMEC and FCal, respectively. SFs correcting for the residual data–simulation discrepancy, originating predominantly from shower-shape mis-modelling, are derived per $|\eta|$ bin and carry a total uncertainty of $2\,\%$ to $4\,\%$ in the EMEC and $4\,\%$ to $8\,\%$ in the FCal, dominated by the systematic component from the background model and fit range.

5.7 Forward electrons in Phase-II

In the Phase-II section of Chapter 3 (Section 3.9), we detailed the major updates that will be implemented in the ATLAS detector for the HL-LHC era during the LS-3. The most relevant for this chapter is the full replacement of the ID with an all-Si tracker, the ITk, which will extend the tracking coverage up to $|\eta| = 4.0$ and provide a much finer granularity than the current detector. This will allow the reconstruction of tracks in the forward region, enabling the extension of electron identification and isolation to $|\eta| = 4.0$ and providing additional handles to reject background in this region. In addition to the ITk, the HGTD will provide high-granularity timing information for tracks and calorimeter clusters, which can be used to further suppress pile-up contributions to the isolation variables. The combination of these two major upgrades will significantly enhance the performance of electron and photon reconstruction, identification, and isolation in the forward region.

As mentioned before, the only possible steps for forward electrons in the current detector are the cluster reconstruction steps shown in Figure 5.2. With the ITk, it can be extended to extrapolate tracks to the clusters, find matches, and perform the GSF refit of the matched tracks.

The topo-cluster growth is constrained using $\zeta_{\text{cell}}^{\text{EM}}$ defined in Equation 5.1, but at large pile-up, large noise blurs the boundaries of the topo-clusters. An idea derived from the super-cluster formation was developed and named “cookie-cutting”. The cluster growth is now governed organically up to a physically motivated limit, here, the Molière radius (see Equation 3.13). The forward topo-clusters are roughly symmetric in η and ϕ, so a 0.3×0.3 cookie-cutter window is used. However, the FCal cells are not symmetric in (η, ϕ) but in (x, y) so a cone size of $\Delta R < 0.3$ is used to collect cells around the hottest cell. An additional step is thus added to the new workflow: after reading the forward topo-cluster, its size is restricted using the cookie-cutting method. Then the newly GSF-refit tracks are matched again with these newly cut clusters.

Once all these steps are completed, track variables can be used for calibration and identification.

5.7.1 p_T-calibration

Until Phase-II, the forward electron p_T reconstruction and calibration relied solely on the calorimeter measurement E and η. This is known as p_T^{calo}. With the ITk, the track momentum measurement can be used to improve the p_T reconstruction. It is labelled p_T^{track} and is purely based on the tracker information. The combination of the two measurements is done in another variable: p_T. Until now, $p_T = p_T^{\text{calo}}$.

Figure 5.12 shows the ratio of the reconstructed p_T following the three different approaches to the true p_T of the electron in simulation. The track-only measurement p_T^{track} is far from the truth, showing a poor resolution, which is not entirely surprising given the limited lever arm for tracks in the forward region. The calorimeter-only measurement p_T^{calo} and the combined measurement p_T both follow the same distribution, almost centred at one, but with a sizeable tail on each side of the peak. This suggests that the current calibration methods developed for p_T^{calo} cannot cope with the new tracker and the p_T^{track} is not transferable from the central region to the forward one. This suggests the need for a dedicated (or updated) calibration to properly account for the new information provided by the tracker and better handle the reduced amount of matter prior to the calorimeter. The calibration applied to p_T^{calo} is indeed the one derived for the Run-2 and Run-3 detector, which has not been re-derived for the Run-4 conditions (the ITk material in front of the calorimeter and $\langle\mu\rangle = 200$).

The following work is done via two sets of MC samples: a $Z \to ee$ sample and a single-particle sample with electrons generated uniformly in p_T and η. Both are simulated with the full ATLAS detector and the expected pile-up conditions for the HL-LHC era with $\langle\mu\rangle = 200$ and with a CM energy of $\sqrt{s} = 14\,\text{TeV}$. The $Z \to ee$ sample allows one to add more physical effects such as FSR and the underlying event, while the single-particle sample allows one to have a more uniform coverage of the phase-space and thus

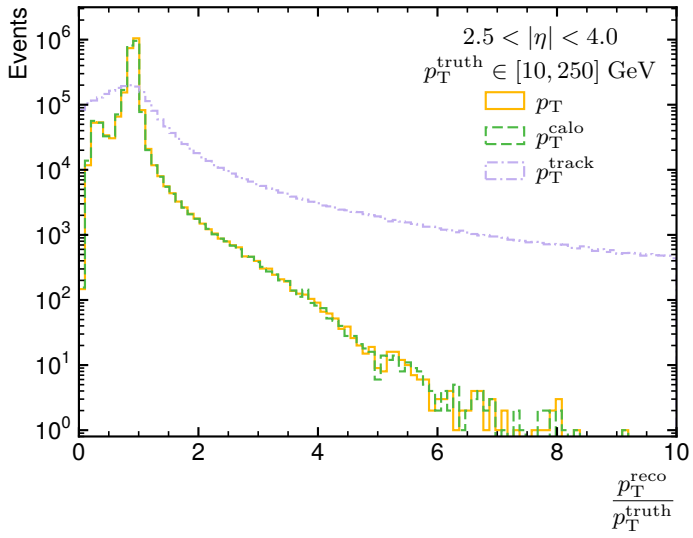


Figure 5.12: Ratio of the reconstructed p_{T} to the true p_{T} for forward electrons in simulation with $\langle\mu\rangle = 1$, comparing the calorimeter-only measurement $p_{\mathrm{T}}^{\mathrm{calo}}$, the track-only measurement $p_{\mathrm{T}}^{\mathrm{track}}$, and the combined measurement p_{T}.

better understand the behaviour of the calibration in different regions of E_{T} and η. The combination of both samples provides a broad range of physics effects and kinematic coverage to develop and test the calibration methods. The calibration DNN is trained on the $\langle\mu\rangle = 200$ samples only, split into a training and a validation set, the validation set being used for the learning rate scheduler and the early stopping.

Among the samples, a truth-matching is applied to every electron candidate with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ within $\Delta R < 0.3$ of a true electron from the generator-level information. Every electron candidate failing these requirements is discarded and not used. Those passing it are selected to form the signal dataset. We also ensure that the selected electrons originate from truth final-state electrons and are not PCs. To mimic the gathering of energy in the calorimeter cell clusters, in the case where the truth electron has suffered radiation, the truth p_{T} considered is that of the initial electron before radiation.

Once the signal dataset is formed, a given set of variables is chosen to reflect the information from the different sub-detectors and the matching between them. The variables are summarised in Table 5.F and can be grouped into three main categories: calorimeter variables, tracker variables, and track–calorimeter matching variables.

The calorimeter variables contain the six cluster moments described in Section 5.6.1, complemented with $\langle\rho^2\rangle$ which is the second moment in energy density, with $\rho_i = \frac{E_i}{V_i}$ where V_i is the volume of the cell i. In addition to the cluster moments, the η and ϕ of the cluster are also included as variables, as well as the fraction of energy deposited in each layer of the EM and HAD calorimeters to better handle the particular geometry of the calorimeter in the forward regions. We therefore define f_i^{EM} and f_i^{HAD} as follows:

$$f_i^{\mathrm{EM}} = \frac{\mathrm{EM\,end\text{-}cap}_i}{E} \quad \text{for } i \in \{1, 2, 3\}\,, \tag{5.22}$$

$$f_0^{\mathrm{HAD}} = \frac{\mathrm{HEC}_0}{E}\,, \qquad f_i^{\mathrm{HAD}} = \frac{\mathrm{HEC}_i + \mathrm{FCal}_i}{E} \quad \text{for } i \in \{1, 2, 3\}\,, \tag{5.23}$$

where E is the total energy of the cluster, $\mathrm{EM\,end\text{-}cap}_i$ is the energy deposited in the i-th layer of the EM calorimeter, HEC_i is the energy deposited in the i-th layer of the HEC and FCal_i is the energy deposited in the i-th module of the FCal (FCal-1 to FCal-3 of Section 3.3.3).

The tracker variables include the η and ϕ of the track, as well as the number of hits in the pixel and strip detectors.

Track–calorimeter matching variables are defined to quantify the consistency between the track and calorimeter measurements. They include the differences in η and ϕ between the extrapolated track impact point in the second calorimeter layer and the corresponding energy deposit, as well as a rescaled version of $\Delta\phi$ that takes into account the calorimeter energy. Additionally, $\Delta\phi^{\mathrm{last}}$ is defined as the difference in ϕ between the last measured track position and the corresponding calorimeter energy deposit, which can provide information on potential mismatches due to bremsstrahlung or other effects.

An additional variable from the HGTD timing information, t^{HGTD}, is also included to provide sensitivity to pile-up contributions, as particles originating from different interactions will have different time-of-flight signatures.

Table 5.F: Variables used for the forward electron p_{T}-calibration and identification in Phase-II.

Detector	Name	Description
HGTD	t^{HGTD}	HGTD time information
Calorimeter	η^{calo}	η from the calorimeter cluster
	ϕ^{calo}	ϕ from the calorimeter cluster
	f_i^{EM}	Fraction of the cluster energy deposited in the i-th layer of the EM calorimeter, for $i \in \{1, 2, 3\}$
	f_i^{HAD}	Fraction of the cluster energy deposited in the i-th layer of the HAD calorimeter, for $i \in \{0, 1, 2, 3\}$
	f_{max}	Fraction of the cluster energy deposited in the most energetic cell of the cluster
	$\mathrm{longitudinal}$	Shower shape in the longitudinal direction of the cluster, based on the distance of each cell to the shower core
	$\langle \lambda^2 \rangle$	Second moment in λ, where λ is the distance of each cell to the cluster centre along the shower axis
	$\mathrm{lateral}$	Lateral moment of the shower taking into account the two most energetic cells, which constitute the shower core
	$\langle r^2 \rangle$	Second moment in r, where r is the radial distance of each cell to the shower axis
	$\lambda_{\mathrm{centre}}$	Distance of the shower centre from the front of the calorimeter along the shower axis
	$\langle \rho^2 \rangle$	Second moment in energy density
Tracker	η^{track}	η from track
	ϕ^{track}	ϕ from track
	n_{pixel}	Number of pixel hits
	n_{strip}	Number of strip hits
Track–calorimeter matching	$\Delta\eta$	Difference in η between the extrapolated track impact point in the second calorimeter layer and the energy deposit in the same layer
	$\Delta\phi$	Difference in ϕ between the extrapolated track impact point in the second calorimeter layer and the energy deposit in the same layer
	$\Delta\phi^{\mathrm{rescaled}}$	Difference in ϕ between the track impact point rescaled using the calorimeter energy and the corresponding energy deposit
	$\Delta\phi^{\mathrm{last}}$	Difference in ϕ between the last measured track position and the corresponding calorimeter energy deposit

The p_{T}-calibration is done in three bins of $|\eta|$:

- $2.5 < |\eta| < 2.7$: this bin sits at the edge of the tracker acceptance and at the transition between the outer and inner wheels of the EMEC, where the granularity coarsens to 0.1×0.1 and the calorimeter response changes, so the performance is expected to be worse than further inside the end-cap. The calibration in this bin is particularly challenging because of the additional material in front of the inner wheel and the change in response across the wheel boundary
- $2.7 < |\eta| < 3.2$: this bin corresponds to the end-cap region of the calorimeter, where the performance is expected to be better than in the transition region but still affected by the reduced granularity and increased pile-up
- $3.2 < |\eta| < 4.0$: this bin corresponds to the region covered by the FCal, where the granularity is coarser and the performance is expected to be worse than in the end-cap region. The calibration in this bin is particularly challenging due to the reduced information available from the calorimeter and the increased pile-up.

We use η^{calo} to define the $|\eta|$ bins, as it is the variable used for the cluster reconstruction and thus more relevant for the calibration. The p_{T} range considered for the calibration is $15 < p_{\mathrm{T}} < 150\,\mathrm{GeV}$, which covers the typical range of interest for physics analyses involving forward electrons. The lower bound is set to ensure a reasonable signal-to-background ratio and to avoid regions where the performance is expected to be very poor, while the upper bound is governed by the very poor statistics of the samples at high p_{T} and the fact that the calibration is expected to be more stable in this region due to the reduced impact of pile-up and other effects.

The choice for the calibration method is a DNN regression, which is trained to predict the true p_{T} of the electron given the input variables described in Table 5.F. The DNN architecture and training procedure are optimised to achieve the best possible performance in terms of resolution and bias of the calibrated p_{T}. The main hyperparameters and architecture choices are summarised in Table 5.G. The same architecture and training procedure are used for the three $|\eta|$ bins, with separate models trained for each bin to account for the different performance and characteristics of the detector in these regions.

Table 5.G: Summary of DNN hyperparameters and architecture choices for the calibration network.

Category	Parameter	Value
Preprocessing	Input normalisation	Min-Max $\in [0, 1]$
	Selection	$p_{\mathrm{T}}^{\mathrm{calo}} \in [15, 150]\,\mathrm{GeV}$
	Target normalisation	$p_{\mathrm{T}}^{\mathrm{truth}}$ with $[10, 255]\,\mathrm{GeV}$ mapped to $[0, 1]$
Architecture	Number of models	3 (one per $\|\eta\|$ bin)
	Input layer	25 variables
	First dense layer	128 nodes
	Normalisation	Batch Normalisation
	Hidden dense layers	6×256 nodes
	Hidden activation	ReLU
	Kernel regulariser	ℓ_2, factor 10^{-4}
	Dropout	None
	Output layer	1 node, softplus
Training	Loss function	MSE, L_{MSE}
	Optimiser	NADAM
	Initial learning rate	10^{-3}
	Learning rate scheduler factor	0.3 (patience 5 epochs on validation loss)
	Minimum learning rate	10^{-6}
	Maximum epochs	250
	Early stopping	patience 7 epochs on validation loss
	Batch size	64
	Train / validation split	75 % / 25 % (at $\langle\mu\rangle = 200$)

Most of the parameters here were already introduced and discussed in Section 4.1 and will not be further discussed here. The only new elements are the following:

1. The batch normalisation layer is added after the first dense layer to help with the training stability and convergence, given the large number of input variables and the potential correlations between them.

It normalises the output of the first dense layer to have zero mean and unit variance, which can help to mitigate issues related to vanishing or exploding gradients and improve the overall performance of the network.

2. The output activation function is chosen to be the softplus function, defined in Figure 5.13, which is a smooth approximation of the ReLU function. The softplus activation is chosen to ensure that the output of the network is always positive, which is a desirable property for a regression task where the target variable (the true p_{T}) is strictly positive. Additionally, the softplus function has a non-zero gradient for all input values, which can help to improve the training dynamics and avoid issues related to dead neurons that can arise with the ReLU activation. It behaves like a linear function for large positive inputs and like an exponential function for large negative inputs, providing a smooth transition between these two regimes.

3. The Nesterov-accelerated ADAptive Moment estimation (NADAM) [308] optimiser is chosen for training the network, which is an extension of the ADAM optimiser that incorporates Nesterov momentum [309]. The NADAM optimiser can provide faster convergence and better performance compared to ADAM in some cases, particularly for deep networks with complex loss landscapes. It combines the benefits of both ADAM and Nesterov momentum, allowing for adaptive learning rates while also incorporating a look-ahead mechanism that can help to escape local minima and improve generalisation.

$$\mathrm{Softplus}(x) = \ln(1 + \mathrm{e}^{x})$$

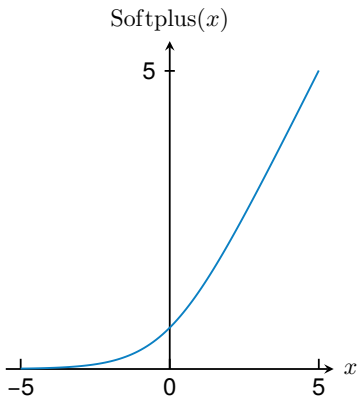


Figure 5.13: Softplus activation function.

Given the different orders of magnitude of the input variables, a Min-Max normalisation is applied to each variable to rescale them in the range $[0, 1]$. The target variable, which is the true p_{T} of the electron, is also normalised in the same way to ensure that the output of the network is in a similar range as the input variables, which can help to improve the training stability and convergence. The $p_{\mathrm{T}}^{\mathrm{min}}$ and $p_{\mathrm{T}}^{\mathrm{max}}$ used for the normalisation are set to $10\,\mathrm{GeV}$ and $255\,\mathrm{GeV}$, respectively, which are slightly wider than the actual range of p_{T} considered for the calibration to allow for some extrapolation and to ensure that the network can handle potential outliers in the data. The normalisation is defined following Equation 4.20. The use of the softplus activation function ensures that the output of the network is always positive, which is consistent with the physical interpretation of p_{T} but also does not use a hard lower bound of zero, allowing the network to learn a more flexible mapping between the input variables and the target p_{T}. To retrieve the calibrated p_{T}, labelled $p_{\mathrm{T}}^{\mathrm{DNN}}$, the output of the network is rescaled back to the original range using the inverse of the Min-Max normalisation.

The choice of the MSE loss function is motivated by the fact that it is a common choice for regression tasks. It penalises large errors more than small ones, which can help to improve the overall performance of the calibration. The use of the MSE loss encourages the network to learn a mapping that minimises the overall error in the p_{T} prediction, which is crucial for achieving a good calibration performance.

No dropout layers are used in order to prevent the DNN from forgetting important features or correlations between parameters.

The performance of the calibration is evaluated by comparing the distribution of the reconstructed p_{T} to the true p_{T} for forward electrons in simulation, on the validation set at $\langle\mu\rangle = 200$. This is presented in Figure 5.14. The same conclusion regarding $p_{\mathrm{T}}^{\mathrm{track}}$ can be drawn here, with a very poor resolution and a large bias. The $p_{\mathrm{T}}^{\mathrm{calo}}$ performs better but still does not follow the peak of the truth distribution. A clear

discrepancy is observed for the latter methods in the $3.2 < |\eta| < 4.0$ bin. The p_{T} using information from both the calorimeter and the tracker follows almost exactly $p_{\mathrm{T}}^{\mathrm{calo}}$, showing again the poor performance and very low contribution of $p_{\mathrm{T}}^{\mathrm{track}}$. The $p_{\mathrm{T}}^{\mathrm{DNN}}$, on the other hand, follows the true p_{T} distribution very closely in all $|\eta|$ bins, showing a significant improvement compared to the other methods.

This is further confirmed by looking at the distribution of the ratio of the reconstructed p_{T} to the true p_{T}, shown in Figure 5.14. In each $|\eta|$ bin, the $p_{\mathrm{T}}^{\mathrm{DNN}}$ distribution is centred on one, where the $p_{\mathrm{T}}^{\mathrm{calo}}$ and p_{T} distributions sit at 0.86 to 0.90: the calibration removes the bias of the existing methods in the whole forward region. Its width, on the other hand, is comparable to that of the calorimeter-based methods in the two inner bins and clearly narrower only in the $3.2 < |\eta| < 4.0$ bin, where the $p_{\mathrm{T}}^{\mathrm{calo}}$ and p_{T} distributions are at their widest and most biased, confirming the poor performance of these methods in the FCal region. The $p_{\mathrm{T}}^{\mathrm{track}}$ distribution is very wide and significantly biased, further confirming that it is not a viable method for p_{T} reconstruction in the forward region. The gain of the DNN calibration over the existing methods is therefore first a bias correction, everywhere, and a resolution gain where it matters most, in the FCal.

The importance of the input variables of the calibration network is summarised in Figure 5.22, together with that of the identification network, and discussed at the end of the identification section.

5.7.2 Identification

The identification of forward electrons in Phase-II will also benefit from the extended tracking coverage provided by the ITk. The presence of a track matched to the calorimeter cluster can provide a powerful handle to discriminate between genuine electrons and background, particularly in the high-pile-up environment of the HL-LHC.

In addition to the $\langle\mu\rangle = 200$ samples used for the p_{T}-calibration, the same samples (single electrons flat in η and p_{T}, and $\mathrm{Z} \to \mathrm{ee}$) are available at $\langle\mu\rangle = 1$. The same selection and truth-matching are applied to form the signal dataset for the identification. The background dataset is formed from a QCD dijet sample with both $\langle\mu\rangle = 1$ and $\langle\mu\rangle = 200$. The same selection is applied and only the candidate electrons that fail the truth-matching are selected to form the background dataset. The rest is discarded. Only the $\langle\mu\rangle = 200$ samples enter the training, split into a training and a validation set as for the calibration. The $\langle\mu\rangle = 1$ samples are never seen by the DNN and serve as an independent test set.

The same binning used for the p_{T}-calibration is used for the identification, with three bins in $|\eta|$: $2.5 < |\eta| < 2.7$, $2.7 < |\eta| < 3.2$, and $3.2 < |\eta| < 4.0$.

5.7.2.1 p_{T}-decorrelation

Figure 5.15 presents the current identification efficiencies of electrons as a function of E_{T}. It is clearly visible that a strong dependence of the efficiency on E_{T} is present. This is due both to the cuts applied and to the dependence of the underlying discriminants on E_{T}. The present work aims to establish a discriminant using variables that are less correlated with p_{T} and trained across the entire p_{T} range.

The goal is to measure distortions in shape using a single DNN to compute selection efficiencies for all p_{T} bins in a correlated manner. By ensuring that the discriminant is less dependent on p_{T}, we aim to reduce p_{T}-dependent mis-modelling as well. This technique would also allow for extrapolation to high p_{T} regions for WPs where direct measurement or treatment of data-versus-MC discrepancies is not feasible.

To achieve this, we propose training a single DNN per $|\eta|$ bin following the one defined for the calibration and over the full p_{T} range $15 < p_{\mathrm{T}} < 150\,\mathrm{GeV}$ using the newly calibrated $p_{\mathrm{T}}^{\mathrm{DNN}}$. The variables considered are the same as for the p_{T}-calibration, as summarised in Table 5.F, but with the application of a decorrelation technique to reduce the correlation of the input variables with p_{T}.

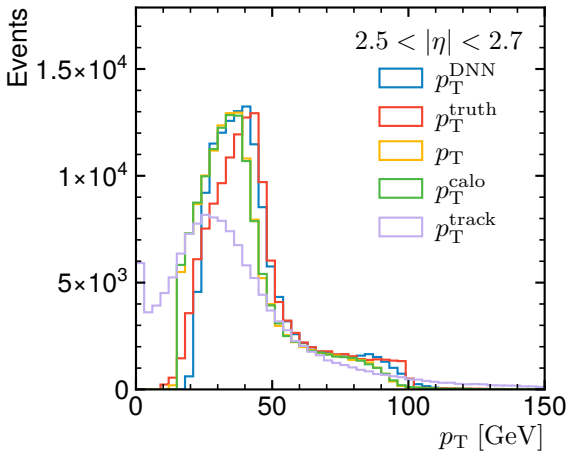
The variables that are most correlated with p_{T} are the cluster moments. To achieve decorrelation we train a simple LR model to fit each of the seven cluster moments against $p_{\mathrm{T}}^{\mathrm{DNN}}$ for the signal dataset at $\langle\mu\rangle = 200$ only. The residuals of this regression are defined as:

$$\mathrm{residual} = y - f_{\mathrm{LR}}(x)\,, \tag{5.24}$$

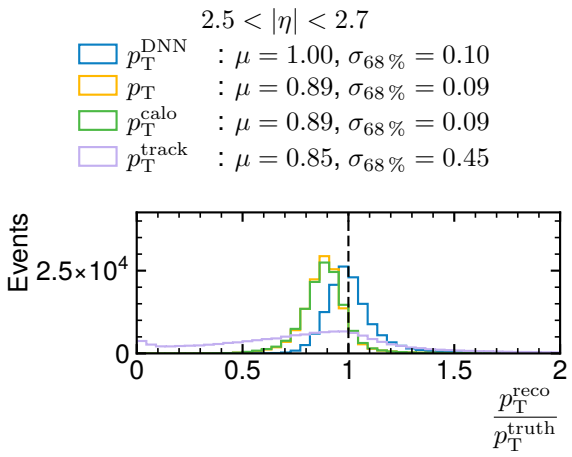
where y is the cluster moment and x is the $p_{\mathrm{T}}^{\mathrm{DNN}}$. This is visually represented in Figure 5.16.

We can thus define the uncorrelated cluster moments as the residuals for both the signal and background dataset for $\langle\mu\rangle = 200$ and $\langle\mu\rangle = 1$. The $p_{\mathrm{T}}^{\mathrm{DNN}}$ variable remains unchanged, while the cluster moments are now decorrelated with respect to it.

The correlation matrices between the cluster moments and the various p_{T} definitions for the signal dataset with $\langle\mu\rangle = 200$ are shown in Figure 5.17, with the original cluster moments on the left column and the decorrelated cluster moments on the right column. The correlation with $p_{\mathrm{T}}^{\mathrm{DNN}}$ is zero for the decorrelated cluster moments, while it can reach up to 0.6 for the original cluster moments. We can see that the correlation with respect to $p_{\mathrm{T}}^{\mathrm{truth}}$ follows the same reduction. However, slight correlations remain

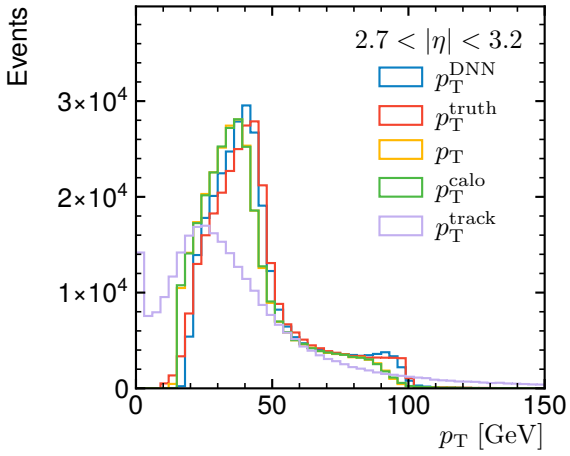


(i) p_{T} distribution

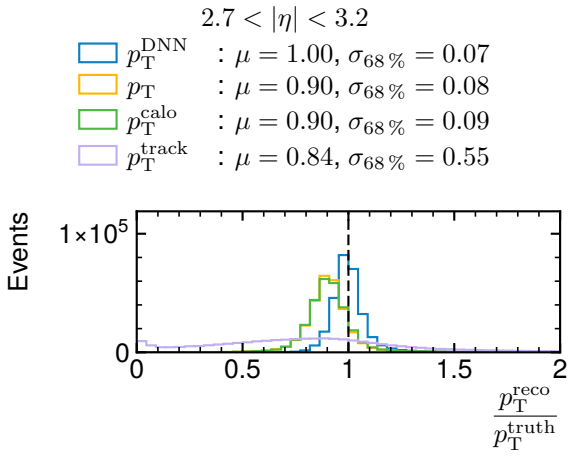


(ii) Ratio of reconstructed to true p_{T}

(a) $2.5 < |\eta| < 2.7$

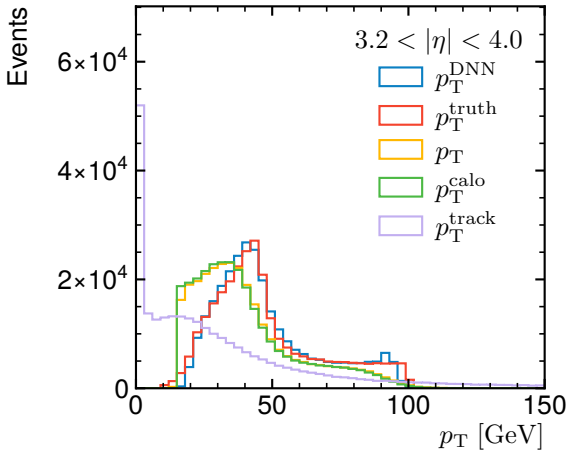


(i) p_{T} distribution

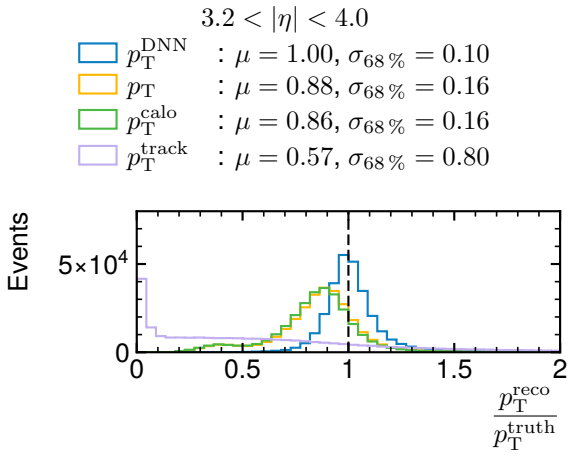


(ii) Ratio of reconstructed to true p_{T}

(b) $2.7 < |\eta| < 3.2$



(i) p_{T} distribution



(ii) Ratio of reconstructed to true p_{T}

(c) $3.2 < |\eta| < 4.0$

Figure 5.14: Distribution of the various reconstructed p_{T} (left column) and their corresponding ratio with respect to the true p_{T} (right column) for forward electrons in the validation set with $\langle\mu\rangle = 200$ in different $|\eta|$ bins. The central value μ and the standard deviation $\sigma_{68\,\%}$ of the distribution are indicated for each method.

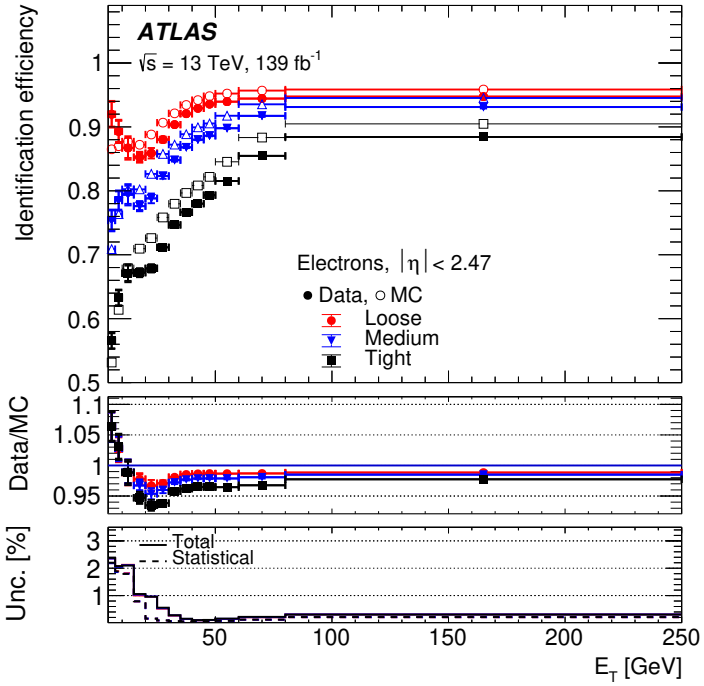


Figure 5.15: Identification efficiencies of electrons from $\mathrm{Z} \to \mathrm{ee}$ decays as a function of E_{T} for the different identification WPs [278]. The top panel shows the efficiencies in data and simulation, the middle panel shows the data-to-simulation ratio (SF), and the bottom panel shows the relative uncertainties in the SF. Full Run-2 dataset.

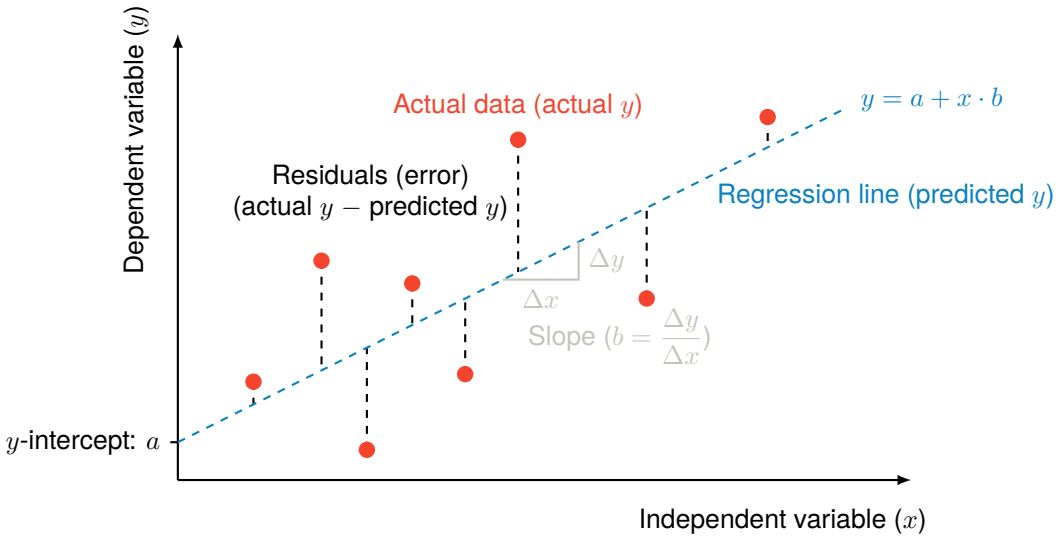


Figure 5.16: LR and residuals.

with the other reconstructed p_{T} methods, suggesting their lower performance and thus lower correlation with the true p_{T} of the electron. The correlation among the cluster moments with this method remains unchanged, leaving the same information for the DNN to learn how to discriminate between signal and background, while reducing the correlation with p_{T} and thus the potential p_{T}-dependent mis-modelling.

The decorrelation presented here is done independently per $|\eta|$ bin.

5.7.3 DNN output

The same type of DNN architecture and training procedure used for the p_{T}-calibration is used for the identification, with the hyperparameters of Table 5.H and separate models trained for each $|\eta|$ bin. The output of the network is a single node with a sigmoid activation function, which outputs a value between 0 and 1 that can be interpreted as the probability of the candidate being a genuine electron (signal) versus background. The loss was already presented and discussed in Section 4.1, while the additions with respect to that formulation were already presented above, for the calibration DNN.

Table 5.H: Summary of DNN hyperparameters and architecture choices for the identification network.

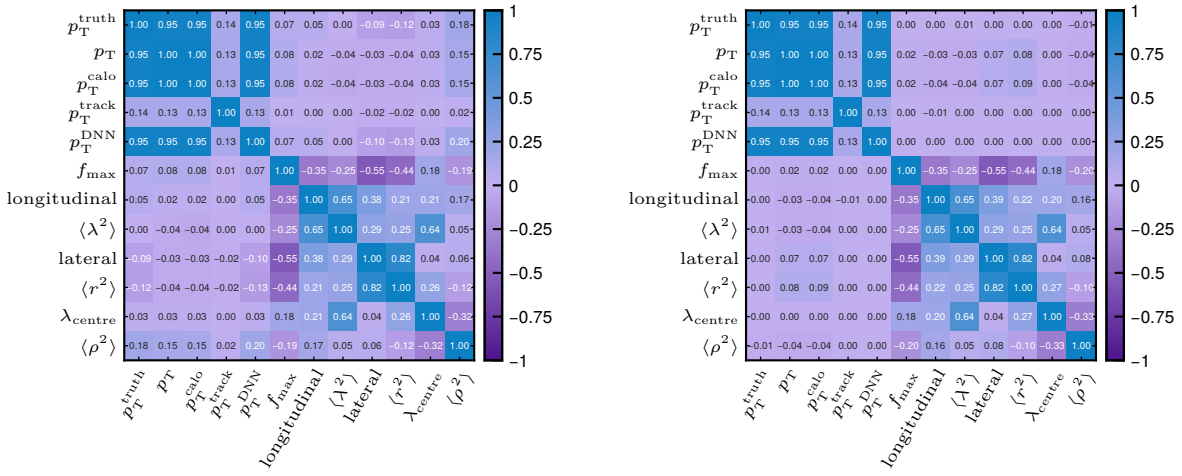
Category	Parameter	Value
Preprocessing	Input normalisation	Min-Max $\in [0, 1]$
	Scaling reference	Signal+background at $\langle\mu\rangle = 200$
Architecture	Number of models	3 (one per $\|\eta\|$ bin)
	Input layer	25 variables
	First dense layer	128 nodes
	Normalisation	Batch Normalisation
	Hidden dense layers	4×256 nodes
	Hidden activation	ReLU
	Kernel regulariser	ℓ_2, factor 10^{-4}
	Kernel constraint	max-norm $= 2$
	Dropout	0.5 (after each dense layer)
	Output layer	1 node, sigmoid
Training	Loss function	BCE, L_{BCE}
	Optimiser	NADAM
	Initial learning rate	$5 \cdot 10^{-4}$
	Learning rate scheduler factor	0.33 (patience 5 epochs on validation loss)
	Minimum learning rate	10^{-10}
	Maximum epochs	250
	Early stopping	patience 10 epochs on validation loss
	Batch size	256
	Train / validation split	$75\,\%$ / $25\,\%$ (at $\langle\mu\rangle = 200$)

A max-norm constraint of 2 on each kernel and a dropout rate of 0.5 are added to limit over-training of the classifier.

The WPs are defined to have a flat signal efficiency. For this, a threshold is defined per $p_{\mathrm{T}}^{\mathrm{DNN}}$ bin, where the $p_{\mathrm{T}}^{\mathrm{DNN}}$ is the calibrated p_{T} defined by the output of the calibration DNN. The threshold is defined on the continuous output of the identification DNN to have a flat signal efficiency across the $p_{\mathrm{T}}^{\mathrm{DNN}}$ bins. The required efficiency for the `Tight` WP is $70\,\%$, for the `Medium` WP is $80\,\%$, and for the `Loose` WP is $90\,\%$. The background rejection is evaluated as a function of $p_{\mathrm{T}}^{\mathrm{DNN}}$ to ensure that the performance is stable across the p_{T} range and to identify any potential issues with p_{T}-dependent mis-modelling.

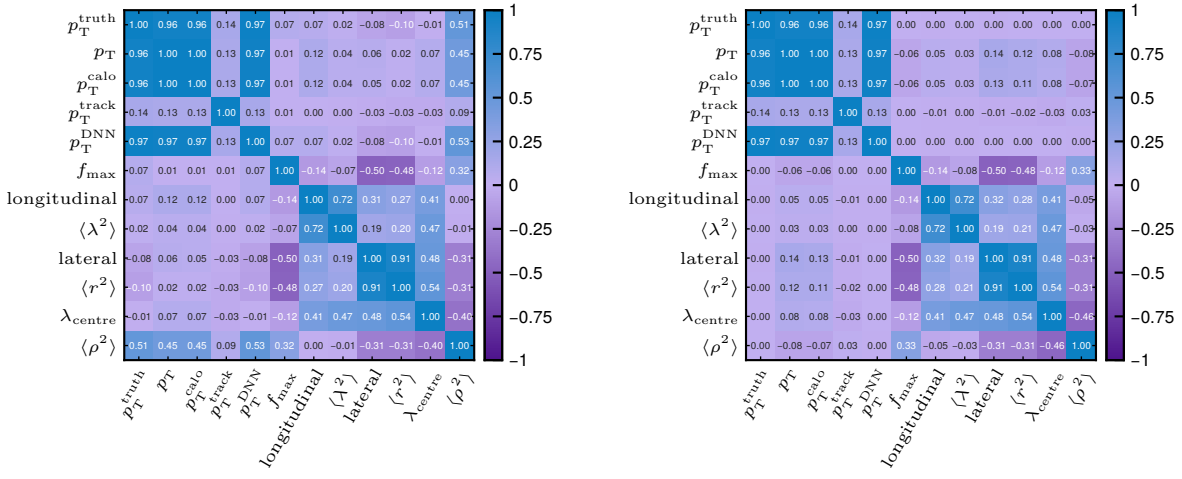
A representation of the full workflow for both interconnected DNNs is shown in Figure 5.18. The calibration DNN is trained first to define the calibrated $p_{\mathrm{T}}^{\mathrm{DNN}}$, which is then used for decorrelation of the cluster moments and to define the p_{T} bins in which a threshold is set for the definition of three WPs. The $p_{\mathrm{T}}^{\mathrm{DNN}}$ bins are also presented.

The resulting DNN output distribution for the signal and background training dataset with $\langle\mu\rangle = 200$ in different $|\eta|$ bins for each $p_{\mathrm{T}}^{\mathrm{DNN}}$ bin is shown in Figure 5.19. We can see that the signal has a larger low-score tail (misidentified signal) in the lowest $p_{\mathrm{T}}^{\mathrm{DNN}}$ bin, while in this bin the background has the smallest high-score tail. Otherwise, the DNN output remains within a few percent of efficiency when varying the $p_{\mathrm{T}}^{\mathrm{DNN}}$ bin, showing that the training across the full p_{T} range and the decorrelation technique are effective in achieving a stable performance across the p_{T} range.



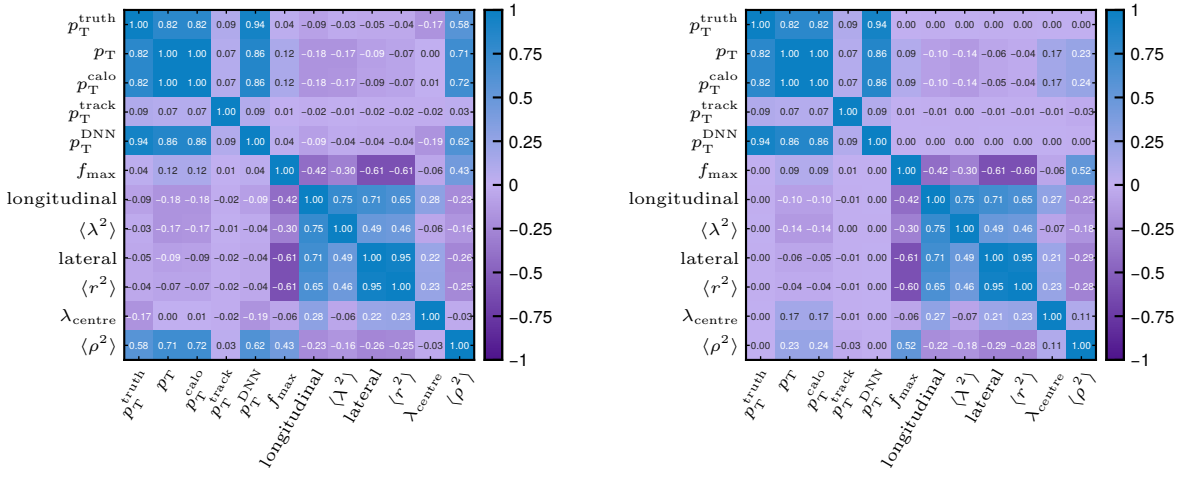
(i) Before decorrelation **(ii)** After decorrelation

(a) $2.5 < |\eta| < 2.7$



(i) Before decorrelation **(ii)** After decorrelation

(b) $2.7 < |\eta| < 3.2$



(i) Before decorrelation **(ii)** After decorrelation

(c) $3.2 < |\eta| < 4.0$

Figure 5.17: Correlation matrix between the cluster moments (left column) and the decorrelated cluster moments (right column) and the various p_{T} definitions for the signal dataset with $\langle\mu\rangle = 200$ in different $|\eta|$ bins.

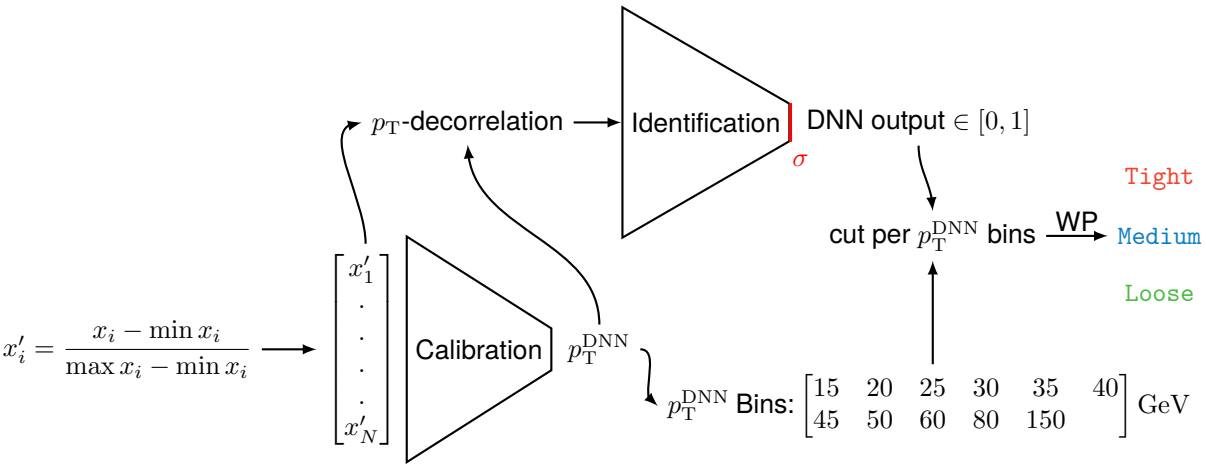


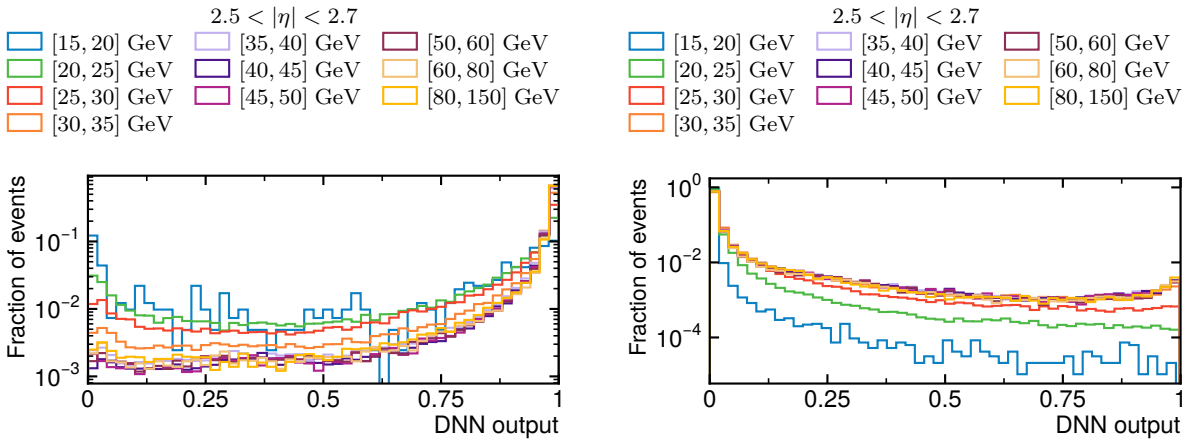
Figure 5.18: Workflow for the DNN calibration and identification.

Figure 5.20 presents the DNN output distribution for the signal and background training dataset on the left column with the corresponding cuts for the definition of the WPs and the ROC curve on the right column for each $|\eta|$ bin and in the $p_{\mathrm{T}}^{\mathrm{DNN}} \in [40, 45]\,\mathrm{GeV}$ bin, close to the Z peak. The cuts are done on the training dataset to have the required signal efficiency for each WP and then applied on the validation dataset to evaluate the background rejection. The ROC curve is also evaluated on the validation dataset. The performance of the DNN is compared to the current `Tight`, `Medium`, and `Loose` WPs for both the LH and the cut on cluster calorimeter moments method defined in Section 5.6.3 and Section 5.6.2, respectively. We can see that the DNN outperforms the other methods in all $|\eta|$ bins, with a significant improvement in background rejection for the same signal efficiency. The LH appears to be clearly inadequate for the new stringent conditions expected for Run-4. The cluster calorimeter method performs better than the LH but is still significantly worse than the DNN. However, those pre-existing methods were not designed for the new conditions expected for Run-4 and thus are not optimised for the new detector and high-pile-up environment. The DNN presented here is trained on those new conditions, allowing it to achieve better performance than the pre-existing methods. The fact that both the validation and training WPs overlap on the ROC curve indicates that no overfitting was observed during the training of the DNN. Since the validation set also drives the early stopping, the independent check is the $\langle\mu\rangle = 1$ test set discussed below.

Figure 5.21 presents the signal efficiency and the background rejection for the different WPs as a function of $p_{\mathrm{T}}^{\mathrm{DNN}}$, for each $|\eta|$ bin. The signal efficiency is evaluated on the validation dataset and is stable across the $p_{\mathrm{T}}^{\mathrm{DNN}}$ bins for all WPs as expected following our methodology. The background rejection is also evaluated on the validation dataset and is stable across the $p_{\mathrm{T}}^{\mathrm{DNN}}$ bins for all WPs as well, but with a slightly worse rejection at lower bins. A comparison with the $\langle\mu\rangle = 1$ test dataset, never seen by the DNN, is provided with the percentage difference between each corresponding WP. As expected by construction of the DNN discriminant for the $\langle\mu\rangle = 200$ sample, the efficiencies are flat. When cross-checking those at $\langle\mu\rangle = 1$, the efficiencies differ by up to $15\,\%$ over most of the range, by up to $\sim 30\,\%$ in the lowest p_{T} bin, and become lower than at $\langle\mu\rangle = 200$ at high p_{T} in the two innermost $|\eta|$ bins. A clear correlation between pile-up and efficiency is observed. However, without samples at intermediate pile-up values an interpolation to account for this effect was impossible for Run-4. Such a correction was instead tested with Run-3 samples, which cover a continuous range of pile-up, as described in Section 5.7.5. Also, not surprisingly the background rejection still exhibits a dependence on p_{T} as it was not targeted during the decorrelation phase.

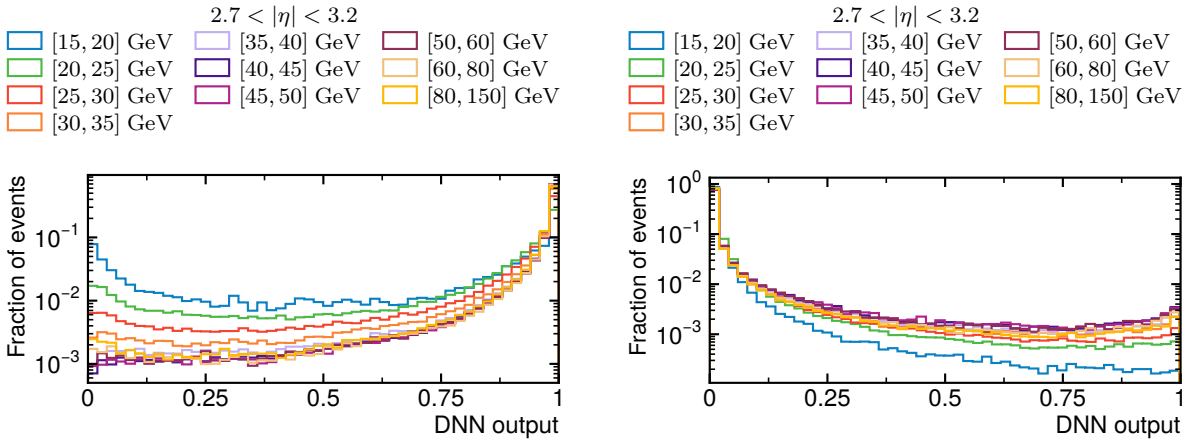
The variables driving the two networks are identified with the three ranking methods presented in Section 4.4 of Chapter 4, the SHAP values, the permutation importance, and the gradient-based importance, evaluated separately in each $|\eta|$ bin. The rank of each of the 25 input variables is given in Figure 5.22.

For the calibration, the second moment in energy density $\langle\rho^2\rangle$ ranks first in all nine configurations. It is followed by the cluster and track pseudorapidities, which locate the shower in the calorimeter, and by the other shower moments ($\langle r^2\rangle$, $\lambda_{\mathrm{centre}}$, f_{max}, $\mathrm{lateral}$): the network corrects the energy response as a function of the shower development and of its position. The layer energy fractions come next, the hadronic ones gaining importance in the FCal bin, while the azimuthal angles, the track–calorimeter matching variables, and the hit counts rank low. The calibration therefore relies mostly on the calorimeter: apart



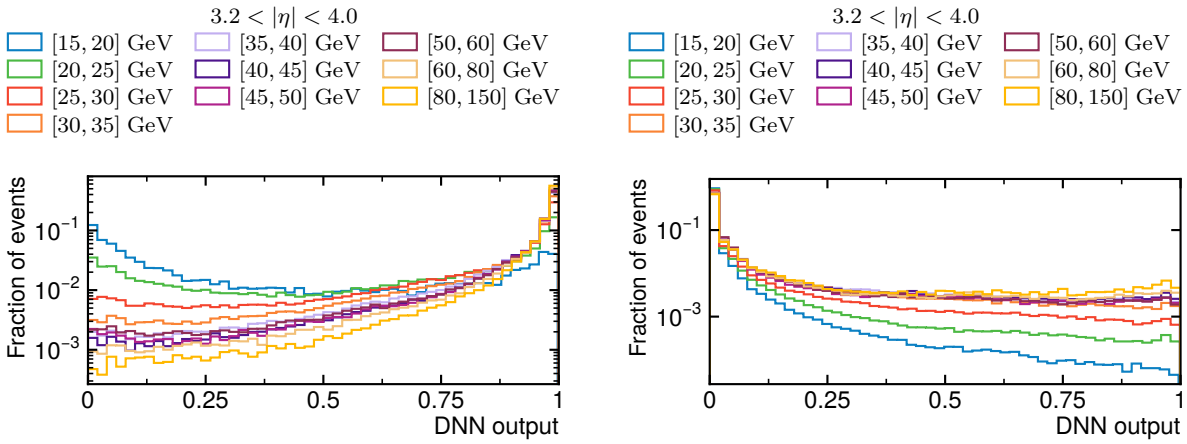
(i) Signal **(ii)** Background

(a) $2.5 < |\eta| < 2.7$



(i) Signal **(ii)** Background

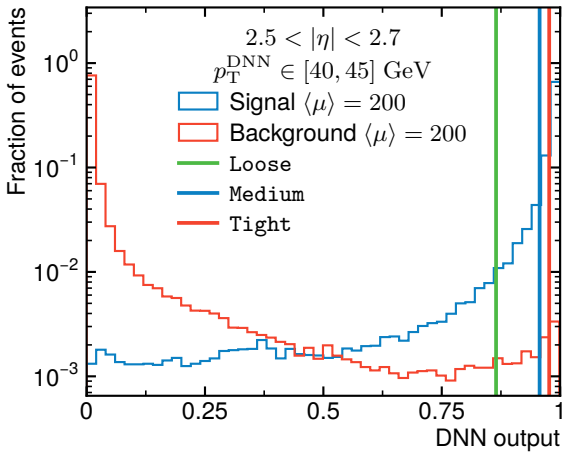
(b) $2.7 < |\eta| < 3.2$



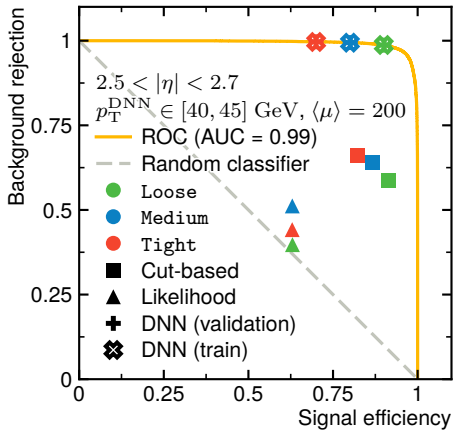
(i) Signal **(ii)** Background

(c) $3.2 < |\eta| < 4.0$

Figure 5.19: DNN output distribution for the signal (left column) and background (right column) training dataset with $\langle\mu\rangle = 200$ in different $|\eta|$ bins for each $p_{\mathrm{T}}^{\mathrm{DNN}}$ bin.

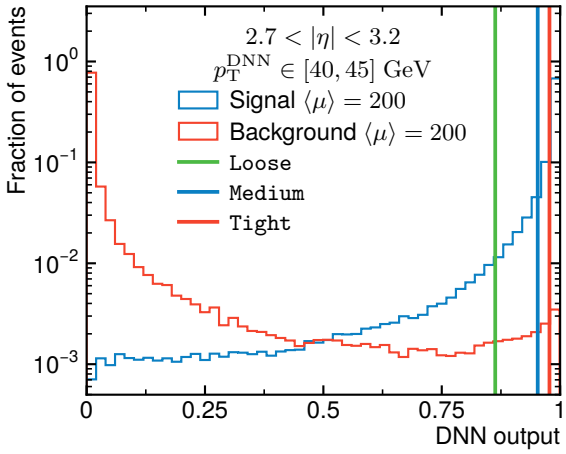


(i) DNN output distribution

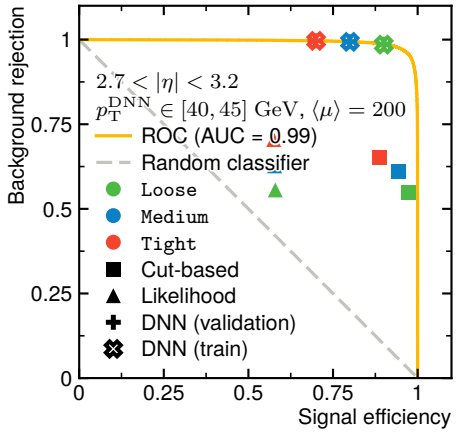


(ii) ROC curve

(a) $2.5 < |\eta| < 2.7$

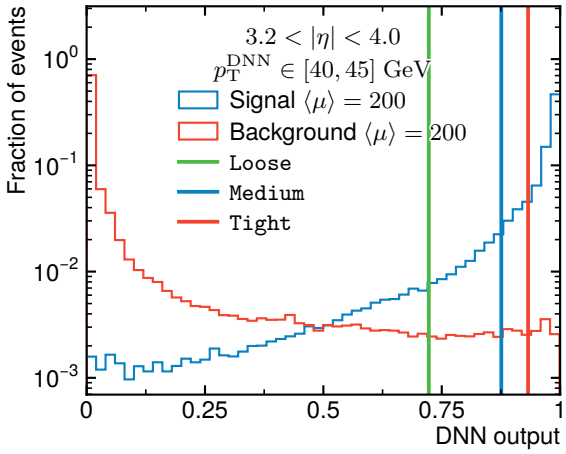


(i) DNN output distribution

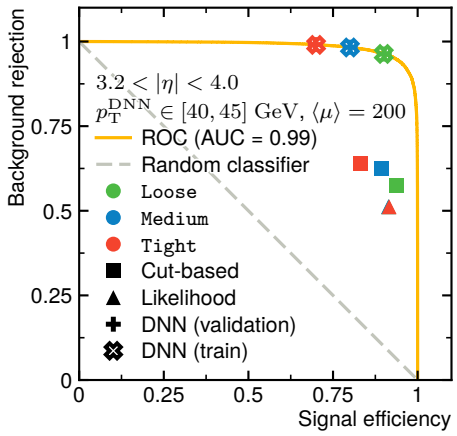


(ii) ROC curve

(b) $2.7 < |\eta| < 3.2$



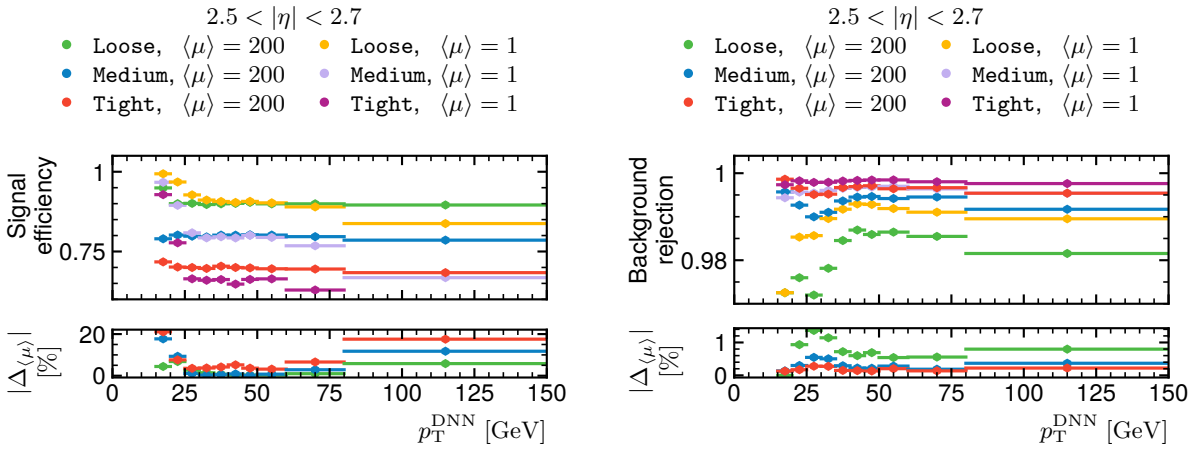
(i) DNN output distribution



(ii) ROC curve

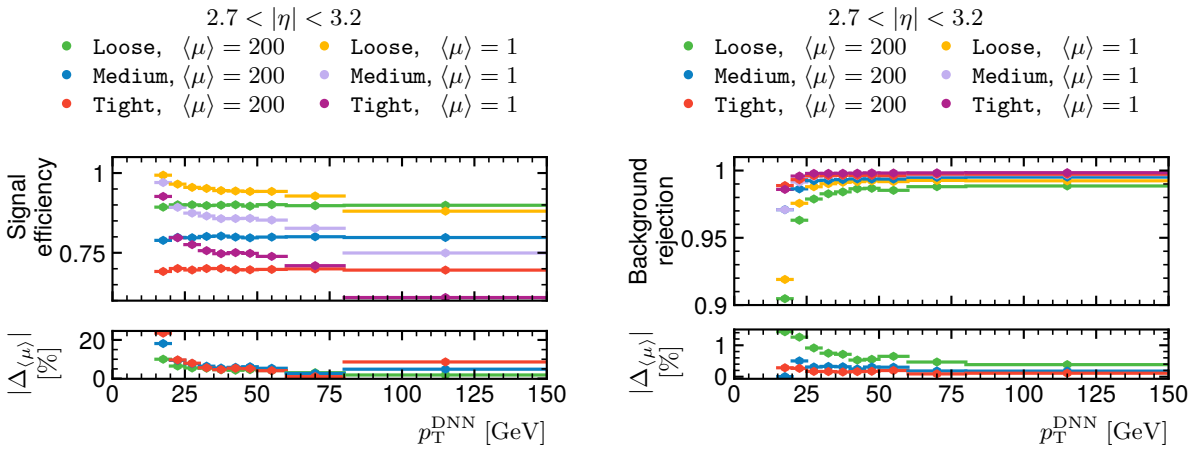
(c) $3.2 < |\eta| < 4.0$

Figure 5.20: DNN output distribution for the signal and background training dataset (left column) and corresponding ROC curve on the validation dataset (right column) with $\langle\mu\rangle = 200$ in different $|\eta|$ bins for $p_{\mathrm{T}}^{\mathrm{DNN}} \in [40, 45]\,\mathrm{GeV}$. Comparison with other WPs.



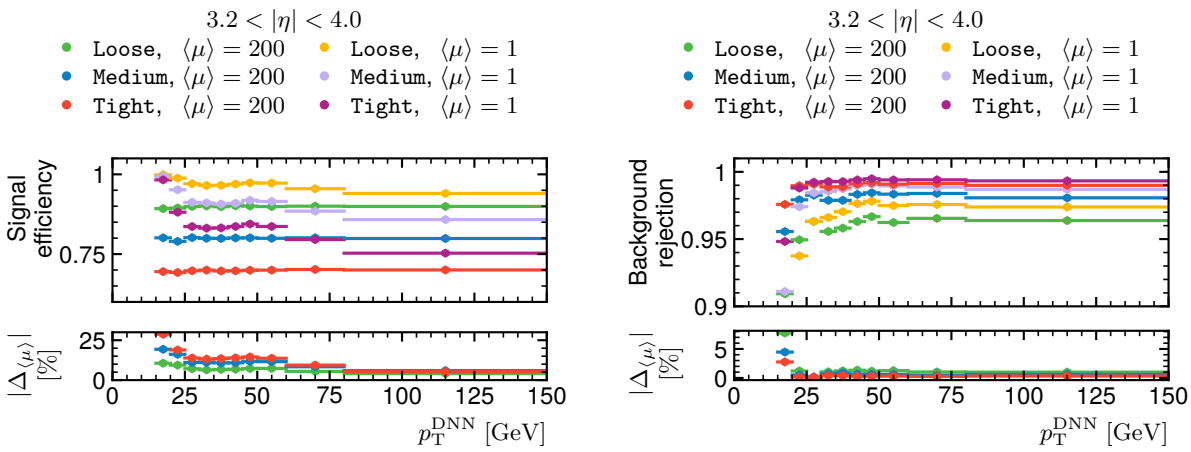
(i) Signal **(ii)** Background

(a) $2.5 < |\eta| < 2.7$



(i) Signal **(ii)** Background

(b) $2.7 < |\eta| < 3.2$



(i) Signal **(ii)** Background

(c) $3.2 < |\eta| < 4.0$

Figure 5.21: Signal efficiencies and background rejection for the DNN WPs on the validation dataset at $\langle\mu\rangle = 200$ and compared with the $\langle\mu\rangle = 1$ test dataset. The lower panel gives the absolute percentage difference of each $\langle\mu\rangle = 1$ efficiency (left column) or background rejection (right column) with respect to $\langle\mu\rangle = 200$.

from η^{track}, which duplicates η^{calo}, the tracker variables rank low, so that most of the gain over $p_{\text{T}}^{\text{calo}}$ comes from the shower shapes and not from the tracker.

For the identification, the importance is more evenly shared. The lateral moment ranks first with the SHAP and permutation methods in all three $|\eta|$ bins, and the track–calorimeter matching variable $\Delta\phi$ is among the three most important ones with the gradient-based importance in every bin, $\Delta\phi$ and $\Delta\phi^{\text{rescaled}}$ being among the best-ranked variables on average: the shower width and the track matching provided by the ITk are the two main handles against jets. The energy fractions of the last EM layers (f_2^{EM}, f_3^{EM}) matter most in the two inner bins, and $\langle r^2 \rangle$ and the hadronic fractions in the FCal bin, following the change of calorimeter technology. In both networks, the azimuthal angles, the numbers of pixel and strip hits, and the HGTD time t^{HGTD} are among the least important inputs, the latter being always among the seven last variables and the last one in eight of the eighteen configurations.

5.7.4 Extrapolation at high p_{T}

The DNN produces an output score $\mathcal{P} \in [0, 1]$ through a sigmoid output layer, where larger values indicate that the electron candidate is more likely to be a genuine electron. WP cuts $\mathcal{P} > \mathcal{P}_{\text{cut}}(p_{\text{T}})$ are derived independently in $(p_{\text{T}}, |\eta|)$ bins to target fixed signal efficiencies. The per-bin thresholds are then fitted with a smooth analytic function of p_{T} in order to extrapolate the cut into wide pseudo-bins outside the training range (namely $[5, 15]\,\text{GeV}$ and $[150, 500]\,\text{GeV}$), owing to the p_{T}-decorrelation approach. This extrapolation is needed because no simulated sample was available in these p_{T} ranges. Providing thresholds in p_{T} bins not covered by the samples was precisely the purpose of the p_{T}-decorrelation: since the discriminant depends only weakly on p_{T}, the thresholds evolve smoothly with p_{T} and can be extrapolated beyond the training range.

Because the DNN output is itself a sigmoid defined in Figure 4.4(b), a logistic function σ is a natural bounded parametrisation of $\mathcal{P}_{\text{cut}} \in [0, 1]$. The dependence is taken in $\ln(p_{\text{T}} + \varepsilon)$ with $\varepsilon = 2.5\,\text{GeV}$, which regularises the low-p_{T} behaviour and reflects the approximately logarithmic shape of the observed turn-on. Two functional forms are used.

The first two WPs, `Medium` and `Tight`, show a simple turn-on with a plateau at high p_{T}, captured by a single sigmoid:

$$\mathcal{P}_{\text{cut}}(p_{\text{T}}) = y_{\min} + (y_{\max} - y_{\min})\sigma\left(k\left[\ln(p_{\text{T}} + \varepsilon) - \ln(p_0 + \varepsilon)\right]\right) , \tag{5.25}$$

where $y_{\min}, y_{\max} \in [0, 1]$ are the low- and high-p_{T} plateaus, p_0 is the turn-on midpoint, and k controls the slope.

The last WP, `Loose`, shows a mild reduction of the cut at high p_{T}, captured by the product of two sigmoids:

$$\begin{aligned} \mathcal{P}_{\text{cut}}(p_{\text{T}}) &= y_{\min} + (y_{\max} - y_{\min})\sigma_1(p_{\text{T}})\left[1 - \alpha\,\sigma_2(p_{\text{T}})\right] , \\ \sigma_i(p_{\text{T}}) &\equiv \sigma\left(k_i\left[\ln(p_{\text{T}} + \varepsilon) - \ln(p_{i-1} + \varepsilon)\right]\right) , \end{aligned} \tag{5.26}$$

with (k_1, p_0) describing the turn-on, (k_2, p_1) the position and steepness of the high-p_{T} drop, and $\alpha \in [0, 0.8]$ its maximum relative amplitude.

The parameters are obtained from a non-linear least-squares fit with bounds $y_{\min}, y_{\max} \in [0, 1]$ and $p_{0,1} \leq 500\,\text{GeV}$. In the least-squares fit, the three lowest-p_{T} thresholds are given an uncertainty $\sigma_i = 0.5$ instead of 1.0 for the other points, that is a weight $\frac{1}{\sigma_i^2}$ four times larger, to stabilise the fit in the steep turn-on region. The threshold quoted for any bin $[x_1, x_2]$ is finally taken as the bin-averaged value of the fitted function:

$$\mathcal{P}_{\text{cut}}^{[x_1,x_2]} = \frac{1}{x_2 - x_1}\int_{x_1}^{x_2} \mathcal{P}_{\text{cut}}(p_{\text{T}})\,\mathrm{d}\,p_{\text{T}} , \tag{5.27}$$

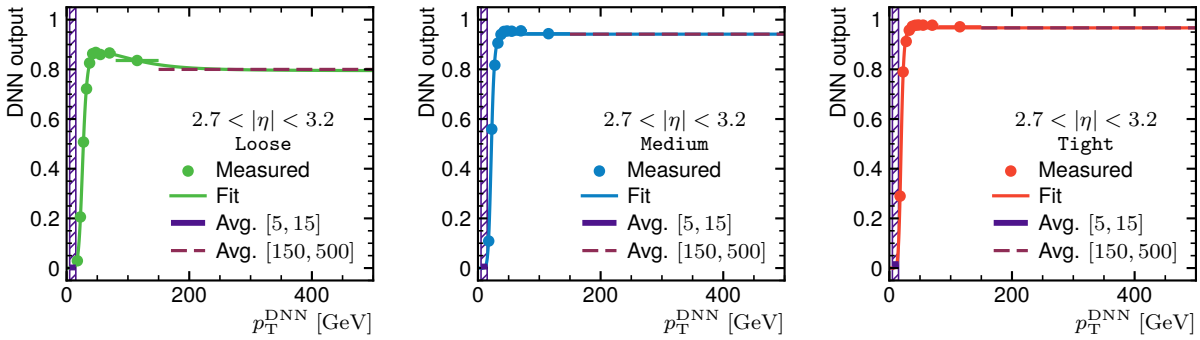
evaluated numerically by trapezoidal integration on a uniform grid.

The resulting fits are shown in Figure 5.23 for the three WPs in the $2.7 < |\eta| < 3.2$ bin, with the per-bin thresholds and the fitted function. The extrapolation to the low and high p_{T} region is shown by the averages over $[5, 15]\,\text{GeV}$ (solid, hatched) and $[150, 500]\,\text{GeV}$ (dashed). The fit captures well the turn-on and plateau behaviour of the thresholds, with a good agreement between the per-bin thresholds and the fitted function. The extrapolation to the low and high p_{T} region is also reasonable, with a smooth behaviour of the thresholds. Similar results are observed for the other $|\eta|$ bins as well.

This whole workflow (the calibration, identification, and WP definition) was implemented inside the ATLAS software Athena [310] to be available for pre-analysis studies and to be used for the physics analyses in Run-4.



Figure 5.22: Ranking of the 25 input variables of the calibration and identification DNNs in each $|\eta|$ bin, for the three ranking methods: SHAP values (S), permutation importance (P), and gradient-based importance (G). Each cell gives the rank of the variable, from 1 (most important) to 25 (least important), the three most important variables of each column being in bold. For the identification, the cluster moments are the p_{T}-decorrelated ones.



(a) Loose **(b)** Medium **(c)** Tight

Figure 5.23: Cut thresholds as a function of $p_{\mathrm{T}}^{\mathrm{DNN}}$ for the three WPs in the $2.7 < |\eta| < 3.2$ bin, with the per-bin thresholds and the fitted function. The extrapolation to the low and high p_{T} region is shown by the averages over $[5, 15]\,\mathrm{GeV}$ (solid, hatched) and $[150, 500]\,\mathrm{GeV}$ (dashed).

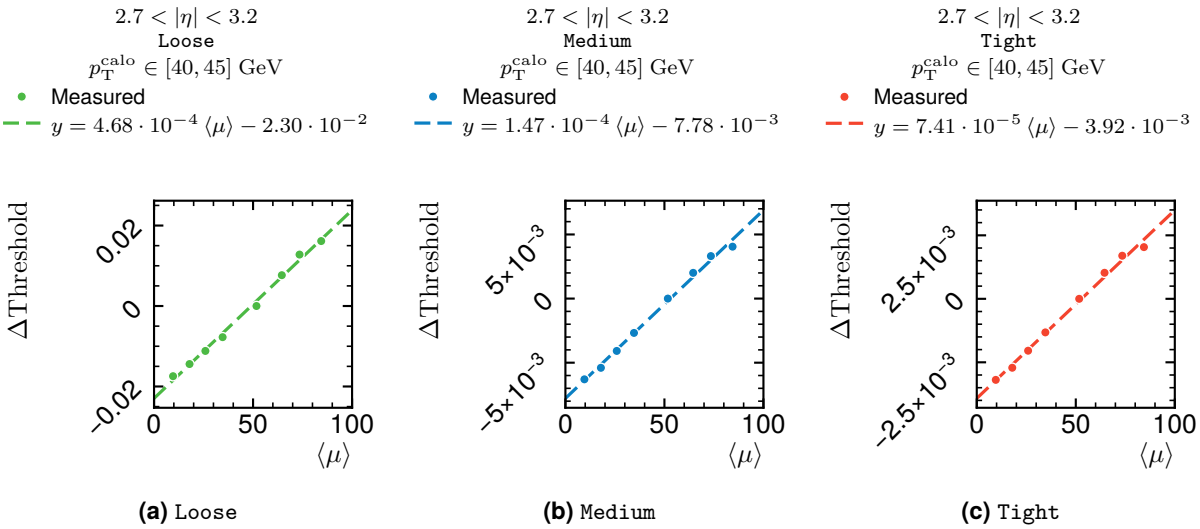
5.7.5 Pile-up dependence correction

The same DNN for identification and its p_{T}-decorrelation were carried out for Run-3, removing from the inputs every tracking-related variable and the HGTD time, neither the ITk nor the HGTD being available in Run-3. With more accessible values of pile-up, some corrections were tried. The training was done over the pile-up bin $\langle\mu\rangle \in [40, 60]$ and the performance was evaluated on its validation dataset, and on the other pile-up bins, never seen by the DNN, as test datasets. A clear dependence on pile-up is observed on the left column of Figure 5.25 where the signal efficiency is shown for the different WPs as a function of $p_{\mathrm{T}}^{\mathrm{calo}}$ (no DNN for calibration was done) for different pile-up bins.

Plotting, as a function of pile-up, the shift of the cut value needed to restore the desired efficiency in each pile-up and p_{T} bin gives Figure 5.24 for the $2.7 < |\eta| < 3.2$ bin and $p_{\mathrm{T}}^{\mathrm{calo}} \in [40, 45]\,\mathrm{GeV}$ bin. A clear dependence is observed, with a linear behaviour in pile-up. A correction was applied to the value of the cut as a function of pile-up to bring back the efficiency to the desired one. We can thus define a function:

$$\Delta\mathrm{Threshold}_i^{[x_1,x_2]}(\langle\mu\rangle) = a \cdot \langle\mu\rangle + b\,, \tag{5.28}$$

where a and b are the parameters of the linear fit, i is the index of the WP and $[x_1, x_2]$ is the $p_{\mathrm{T}}^{\mathrm{calo}}$ bin. The parameters are obtained from a linear fit and given in Figure 5.24.



(a) Loose **(b)** Medium **(c)** Tight

Figure 5.24: Shift of the cut value needed to restore the desired efficiency in each pile-up and p_{T} bin, as a function of pile-up, for the three WPs in the $2.7 < |\eta| < 3.2$ bin and $p_{\mathrm{T}}^{\mathrm{calo}} \in [40, 45]\,\mathrm{GeV}$ bin. The parameters of the linear fit are given on each plot.

Once this correction is applied to the value of the cut for each pile-up bin, the signal efficiency is re-evaluated and shown in Figure 5.25 on the right column. We can see that the pile-up dependence is significantly reduced, with a much more stable performance across the different pile-up bins. A fluctuation

of the order of $1\,\%$ to $2\,\%$ on average, up to $4\,\%$ in the lowest $p_{\mathrm{T}}^{\mathrm{calo}}$ bin, is observed for the three WPs across the different pile-up bins, which is a significant improvement compared to the $5\,\%$ to $10\,\%$ observed before the correction. This correction allows for a more stable performance of the DNN across different pile-up conditions. The parameters of the linear correction are specific to the Run-3 detector and DNN, so they cannot be transferred to Run-4, only the method can. For the HL-LHC, it requires samples at several intermediate pile-up values to fit the slope, as noted in the outlook.

5.7.6 Robustness of the p_{T}-decorrelation

A last test was performed to evaluate the robustness of the p_{T}-decorrelation technique. The identification DNN was trained with the same variables normally used for the identification in the central region and given in Table 5.A. All variables related to the calorimeter were decorrelated following our methodology. The training was carried out in the $2.0 < |\eta| < 2.47$ bin and the $\langle\mu\rangle \in [40, 60]$ bin. The WPs were defined as we did in this section. A comparison was done with the pre-existing LH, DNN, and cut-based identification methods for the same $|\eta|$ and $\langle\mu\rangle$ bin. Figure 5.26 shows that the DNN slightly outperforms the LH and the cut-based identification, but with a much smaller improvement than that observed in Run-4: the legacy methods are more robust and better optimised for these central-region conditions. On the other hand, the DNN presented in Section 5.3.9 lies on the same ROC curve, showing a performance similar to that of the new DNN presented here. The discrepancy between our WPs and the DNN in terms of signal efficiency and background rejection is due to different threshold criteria. However, this is strong evidence that using a p_{T}-decorrelation to obtain stable performance across the p_{T} range is a robust method, with no significant loss of performance observed in the central region where the training was done. This motivates using this method for the Run-4 DNN identification as well, to obtain stable performance across the p_{T} range and to extrapolate to high p_{T}, where no training sample is available.

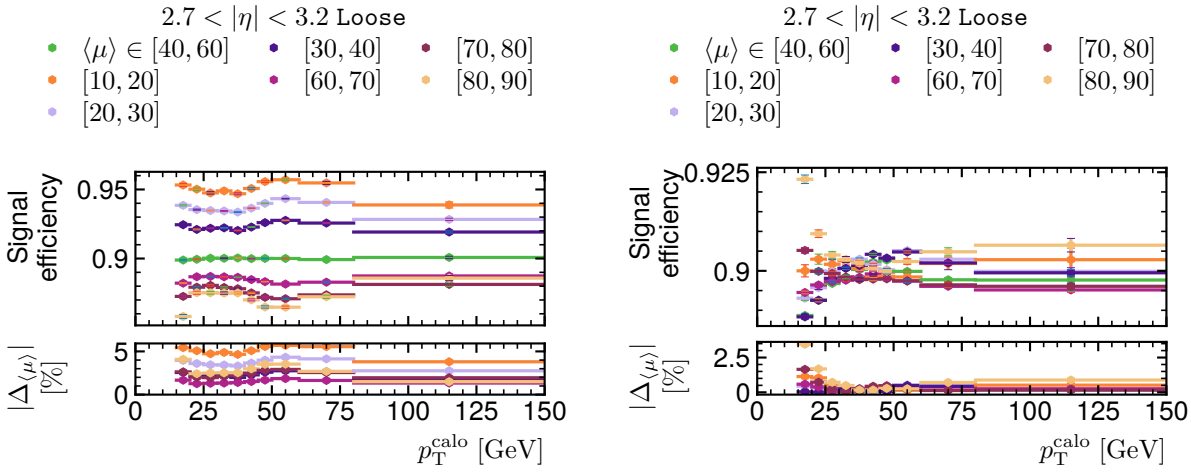
5.7.7 Outlook

The replacement of the Inner Detector by the all-Si ITk, together with the timing information from the HGTD, opens the forward region to track-based electron reconstruction for the first time in ATLAS. The work presented here exploited these new handles to develop a dedicated calibration and identification chain for forward electrons in the high-pile-up environment of the HL-LHC, extending the coverage out to $|\eta| = 4.0$.

The p_{T}-calibration was addressed with a DNN regression combining the cluster moments, the layer energy fractions, the tracker hits, the track–calorimeter matching variables, and the HGTD timing into a single calibrated estimator $p_{\mathrm{T}}^{\mathrm{DNN}}$. As summarised in Figure 5.14, the track-only measurement is unusable in the forward region owing to the very short lever arm, while the calorimeter-only and combined measurements suffer from a sizeable bias and broad tails, in particular in the FCal region. The DNN instead recovers a response centred on unity in all three $|\eta|$ bins, removing the bias, with a resolution equal to that of the calorimeter-based methods in the two inner bins and markedly better in the FCal region, where their tails are largest.

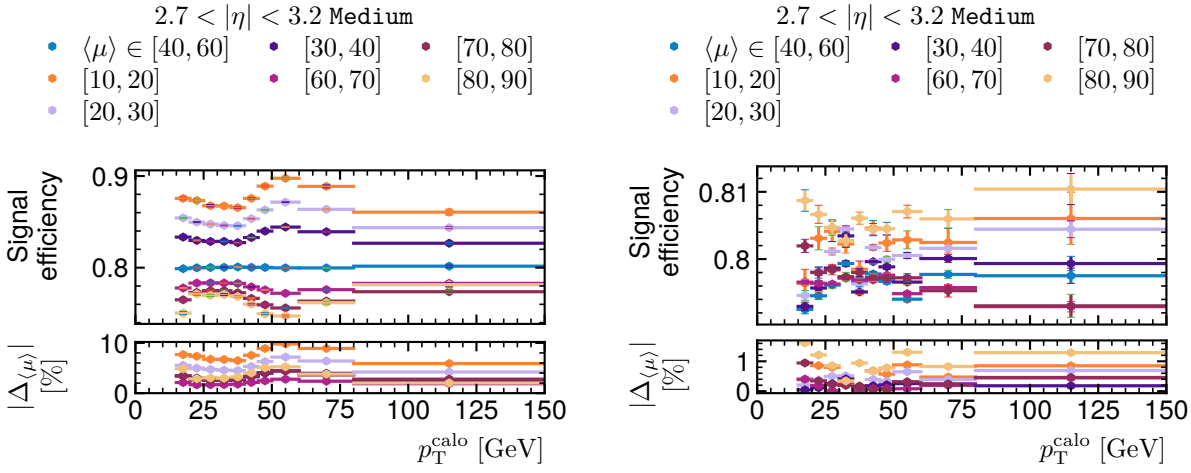
For the identification, a second DNN classifier was trained together with a p_{T}-decorrelation technique, in which the cluster moments are replaced by the residuals of an LR against $p_{\mathrm{T}}^{\mathrm{DNN}}$. This provides WPs with a flat signal efficiency across the full p_{T} range and a substantially better background rejection than the legacy LH and cluster-moment selections (Figure 5.20), which were not designed for the stringent conditions expected in Run-4. The decorrelation further enables a smooth and bounded extrapolation of the WP thresholds beyond the training range, down to $5\,\mathrm{GeV}$ and up to $500\,\mathrm{GeV}$ where no training sample is available. The complete chain, comprising the calibration, the identification, and the WP definition, was implemented in Athena and is ready to be used for the upcoming Run-4 physics analyses.

Several aspects remain open and define natural directions for future work. The Run-4 study relied on only two pile-up points ($\langle\mu\rangle = 1$ and $\langle\mu\rangle = 200$), which prevented a correction of the residual pile-up dependence of the efficiencies. The Run-3 investigation showed that this dependence is well described by a linear function of $\langle\mu\rangle$ and can be brought down to the percent level (Figure 5.25). The same procedure should be ported to Run-4 once samples with intermediate pile-up become available. Finally, the robustness test carried out in the central region (Figure 5.26), where the new DNN reproduces the performance of the existing one, confirms that the p_{T}-decorrelation does not degrade the discrimination and motivates its wider adoption. The use of these WPs in preliminary studies for Run-4 constitutes the next milestone.



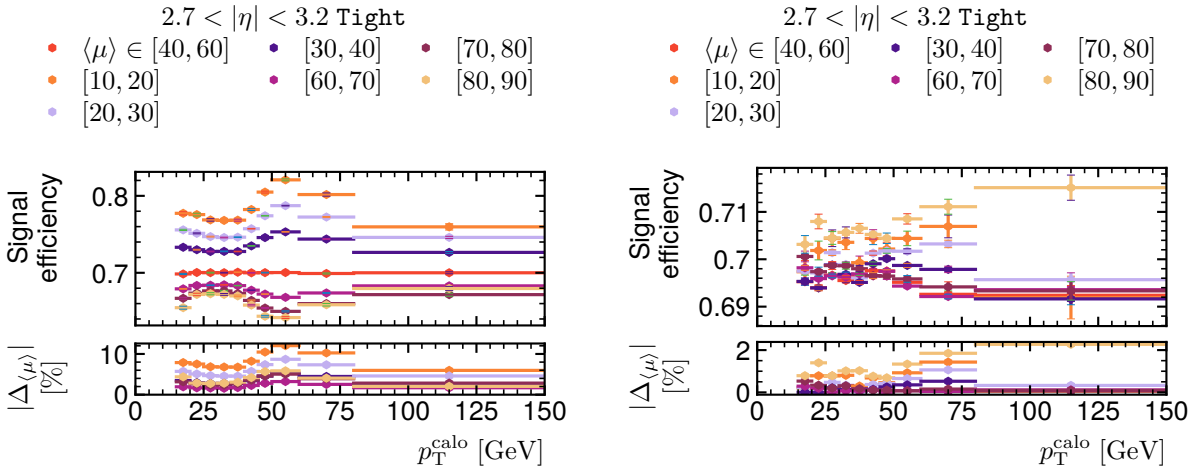
(i) Before pile-up correction **(ii)** After pile-up correction

(a) `Loose`



(i) Before pile-up correction **(ii)** After pile-up correction

(b) `Medium`



(i) Before pile-up correction **(ii)** After pile-up correction

(c) `Tight`

Figure 5.25: Signal efficiencies for the DNN WPs on the validation dataset at $\langle\mu\rangle \in [40, 60]$ and compared with the other $\langle\mu\rangle$ bins used as test datasets. The left column shows the signal efficiency before the pile-up dependence correction and the right column after. The lower panel gives the absolute percentage difference of each pile-up bin with respect to $\langle\mu\rangle \in [40, 60]$.

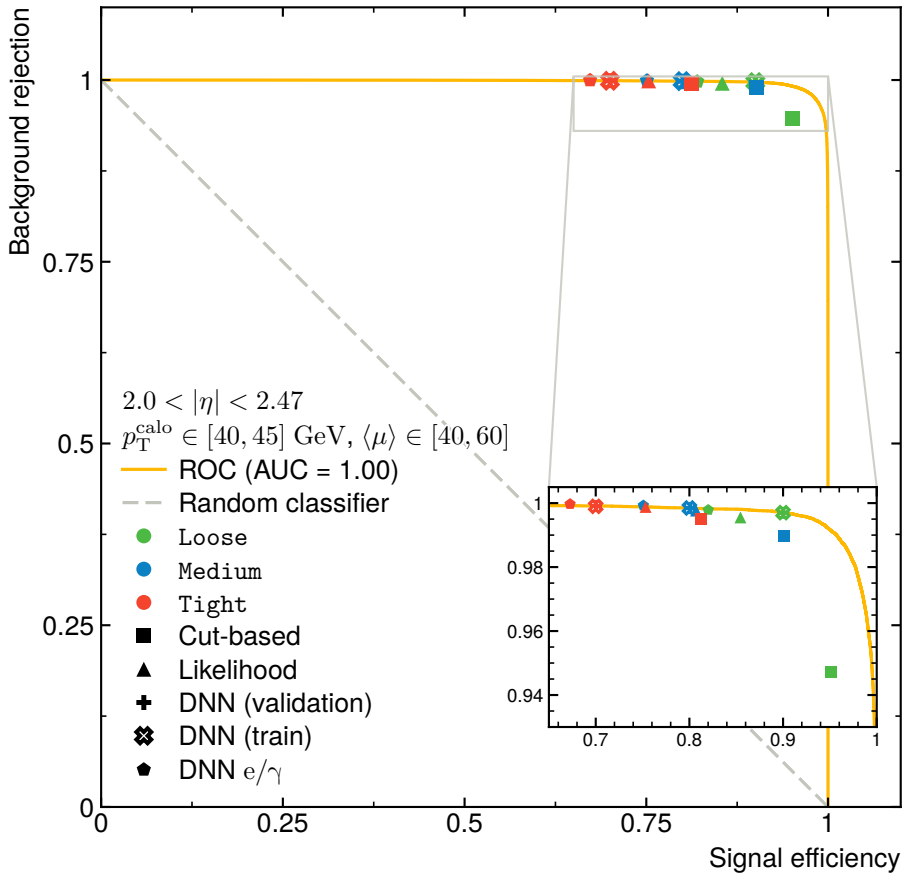


Figure 5.26: ROC curve for the DNN WPs with $\langle\mu\rangle \in [40, 60]$ in the $2.0 < |\eta| < 2.47$ and $p_{\mathrm{T}}^{\mathrm{calo}} \in [40, 45]\,\mathrm{GeV}$ bin. Comparison with other WPs.

6 Polarisation in vector boson scattering



In Chapter 1 we presented the SM and its limitations, with a particular highlight on the vector boson polarisation and how it plays a key role in probing the SSB mechanism. Indeed, the additional degree of freedom arising from the Goldstone boson is manifested by a longitudinal polarisation state, which is a unique feature of the vector boson. The particular process of VBS exhibits a scattering amplitude that grows with the CM energy $\sqrt{s}$ of the process [311]. As shown in Figure 1.7, the cross-section of these processes is dependent on the SSB mechanism and the resulting Higgs boson. This represents a unique way to test the limits of the SM.

However, such measurements are not easily achievable, and have been accessible only recently with the large dataset of Run-2 and most recently Run-3. The history of these measurements is intrinsically correlated with *inclusive* production of diboson pairs, which are the first step towards the observation of VBS. In the following sections we will review the history of these measurements to give a clear picture of the current state of the art and the challenges that remain to be tackled. We will then present the particularities of VBS events and how they can be exploited to perform measurements of the vector boson polarisation.

6.1 Historical overview of diboson measurements

The first measurement of the diboson production can be traced back to the LEP collider in $\mathrm{e^+e^-}$ collisions. The WW was first observed in 1996 by the OPAL collaboration [312] while the ZZ was first measured in 1999 by the ALEPH collaboration [313]. Results corroborating this measurement from the other LEP experiments were also obtained. The last combination for VV diboson pair production for massive vector bosons is WZ production, but it could not be measured at LEP as the initial state carries a total charge of zero while the resulting WZ system is charged. This measurement had to wait for the start of the Tevatron where it was first observed in 2007 by the CDF collaboration [314]. A measurement of the $\mathrm{p\bar{p}} \to \mathrm{WZ} + X$ cross-section was also performed by the DØ collaboration in the same year [315].

The Tevatron also provided the first hadron-collider measurements of the neutral and charged diboson channels. The WW production cross-section was measured by the DØ and CDF collaborations in 2005 [316, 317]. The ZZ channel, because of its smaller production rate due to its higher system mass, was established later: strong evidence was reported by CDF in 2008 [318], followed by the observation by DØ in the same year [319]. These measurements completed the establishment of the massive diboson processes at hadron colliders before the start of the LHC programme.

At the LHC, inclusive diboson production was rapidly remeasured at higher CM energies. The WW production cross-section was first measured at $\sqrt{s} = 7\,\mathrm{TeV}$ by CMS and ATLAS [320, 321], and later at $\sqrt{s} = 8\,\mathrm{TeV}$ by both experiments [322, 323]. Similarly, inclusive WZ production was measured by ATLAS at $\sqrt{s} = 7\,\mathrm{TeV}$ and $8\,\mathrm{TeV}$ [324, 325], and by CMS at both energies [326]. The inclusive ZZ channel was also established during Run-1, with measurements at $\sqrt{s} = 7\,\mathrm{TeV}$ and $8\,\mathrm{TeV}$ by both collaborations [327–330].

The increase in CM energy to $\sqrt{s} = 13\,\mathrm{TeV}$ during Run-2 allowed more precise inclusive measurements and the first detailed differential studies. Early Run-2 measurements of WZ and ZZ production were performed by both ATLAS and CMS [331–334]. These were followed by full Run-2 precision measurements of WW, WZ, and ZZ production [335–340]. The first Run-3 measurements at $\sqrt{s} = 13.6\,\mathrm{TeV}$ have already extended this programme, with measurements of inclusive WW, WZ, and ZZ production [341–343].

Beyond inclusive cross-section measurements, diboson production also gives direct access to the polarisation structure of the EW interaction. The first measurements of single-boson polarisation fractions in inclusive WZ production were performed by ATLAS [10], followed by a more precise CMS measurement using the full Run-2 dataset [344]. The first observation of joint-polarisation states in WZ production was reported by ATLAS [11]. In addition to the article, the Ph.D. thesis of Selem [345] presents in detail the measurement, which this thesis builds upon. In the ZZ channel, ATLAS reported evidence for pair production of longitudinally polarised vector bosons [346]. More recently, the energy dependence of the polarisation fractions and the radiation amplitude zero (RAZ) effect were studied in WZ production [347].

These remarkable measurements of boson polarisation in diboson production were possible only with the substantial increase of the dataset collected by the experiments, as well as the development of new analysis techniques to maximise the sensitivity to these observables.

The study of diboson production in association with two jets provides access for the first time to VBS topologies. The first evidence for EW same-sign WWjj production was reported by ATLAS at $\sqrt{s} = 8\,\mathrm{TeV}$ [348], followed by a CMS study in the same final state [349]. The observation of this process was then achieved during Run-2 by CMS and ATLAS [350, 351]. With the full Run-2 dataset, both experiments further refined the measurement of same-sign WWjj production, including measurements of polarised contributions [352–354]. The opposite-sign WWjj channel was observed later by CMS and ATLAS [355, 356], completing the experimental establishment of the main EW WWjj production modes.

The EW production of WZjj is particularly relevant for this thesis. It was first observed by ATLAS using partial Run-2 data [357], while CMS performed a measurement of EW WZ production in association with two jets and searched for BSM physics in the same final state [358]. A combined CMS measurement of WZjj and same-sign WWjj production followed [359]. The full Run-2 ATLAS measurement of EW WZjj production provided a more precise determination of this process [360]. Recently, the first Run-3 measurement of EW WZjj production was reported by CMS [361], extending the study of this process to a higher CM energy and a larger dataset.

The EW ZZjj channel was established later because of its smaller cross-section. Evidence for EW production of four charged leptons and two jets was reported by CMS [362], followed by the observation of EW ZZjj production by ATLAS [363]. A CMS observation using the full Run-2 dataset was subsequently reported [364].

Finally, semileptonic VBS final states provide complementary sensitivity thanks to their larger branching fractions. Searches and measurements in these channels were performed by ATLAS and CMS [365–368], completing the present experimental landscape of diboson and VBS measurements.

A complete timeline of the major measurements of diboson and VBS production discussed here is shown in Figure 6.1 for the inclusive diboson measurements and in Figure 6.2 for the VBS ones. Also, the evolution of the cross-section measurements of total inclusive diboson production as a function of the CM energy $\sqrt{s}$ is shown in Figure 6.3 for the measurements reported previously.

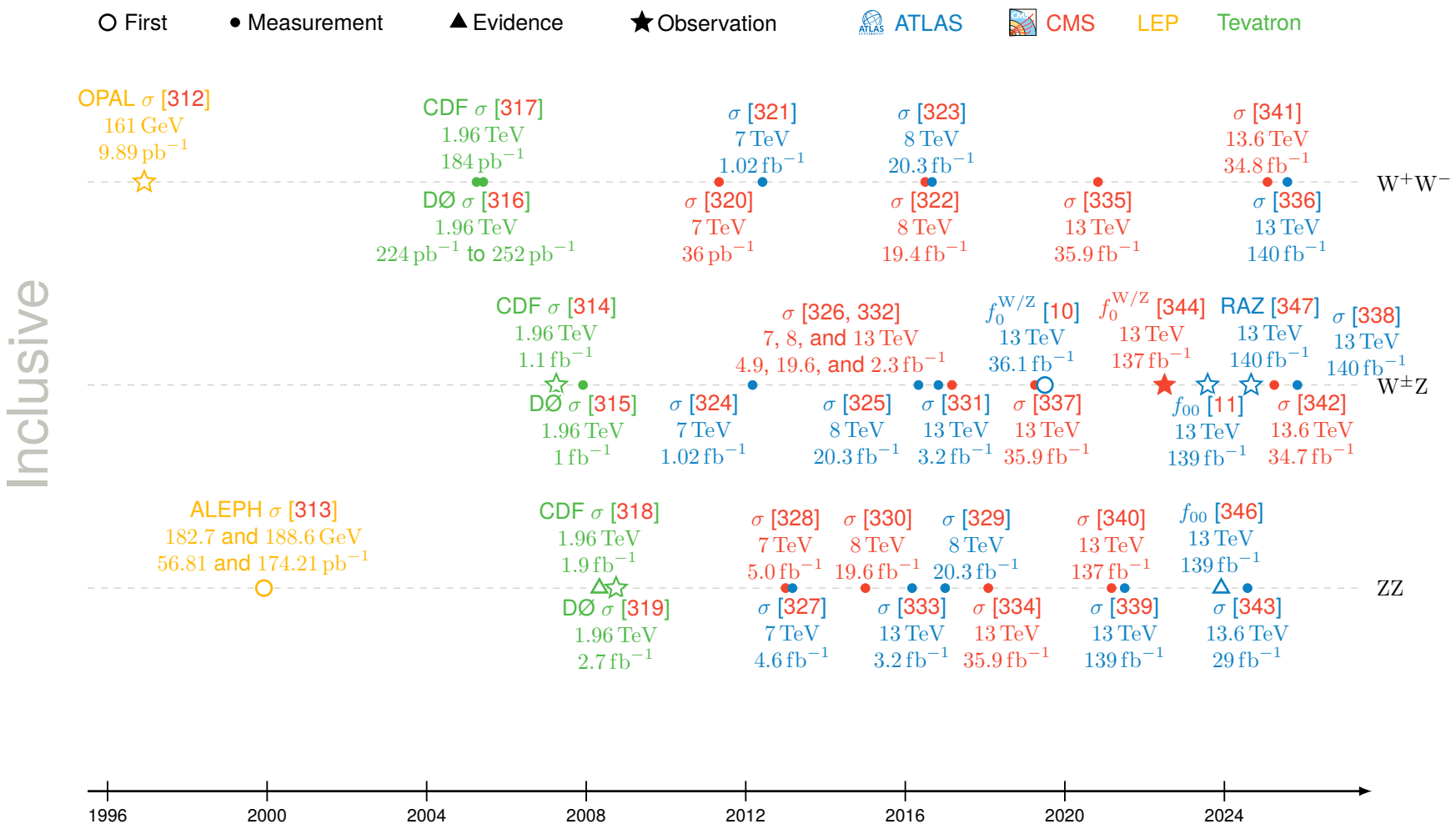


Figure 6.1: Timeline of the major inclusive diboson measurements.

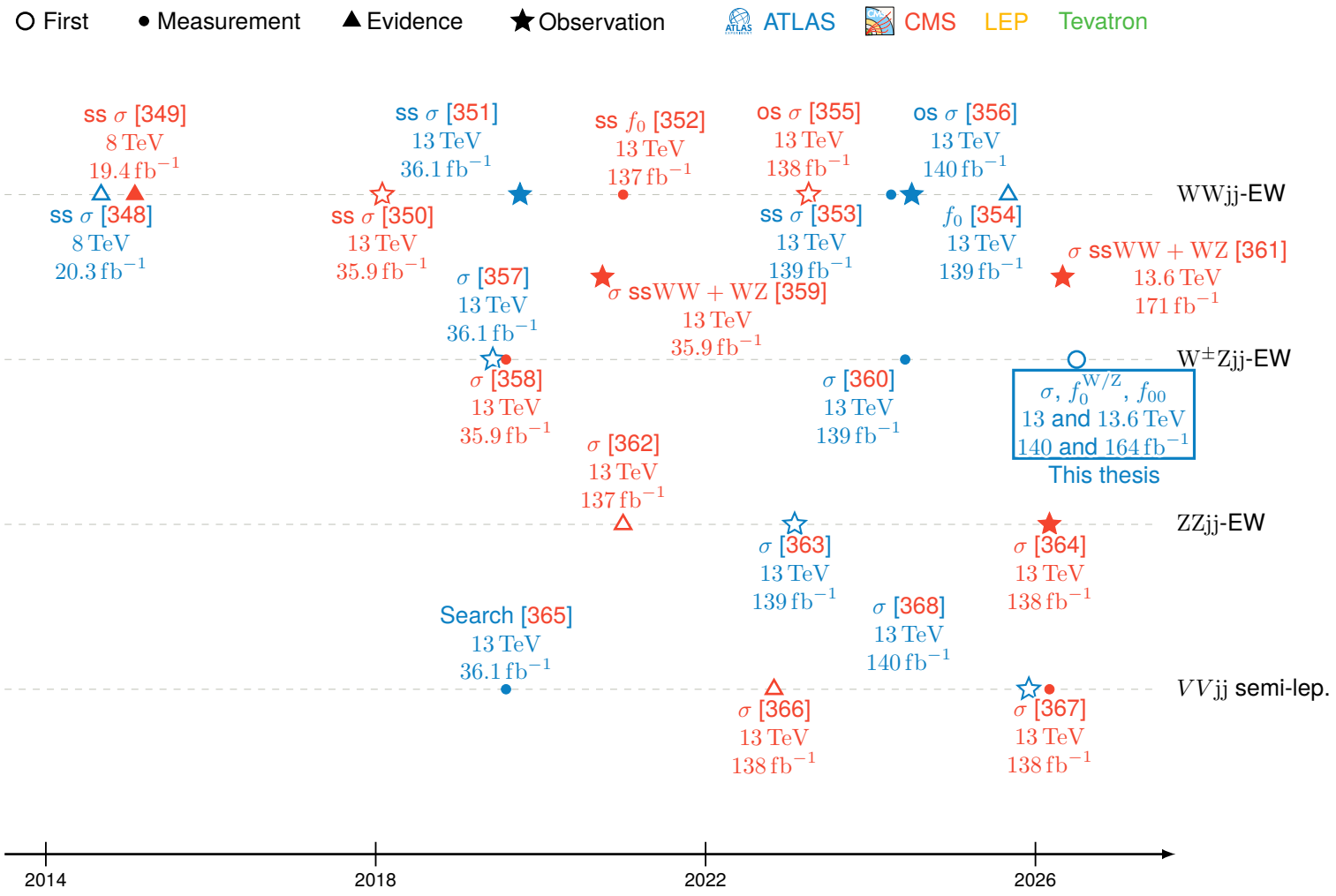


Figure 6.2: Timeline of the major VBS measurements.

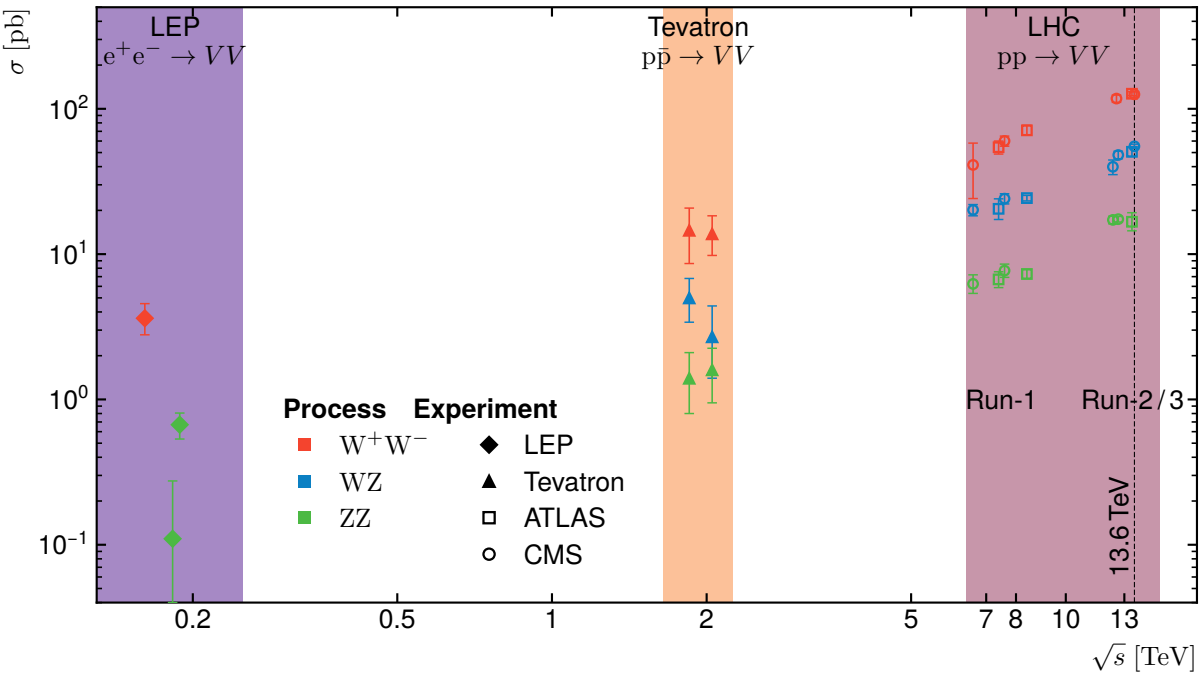


Figure 6.3: Cross-section of the diboson production as a function of the CM energy $\sqrt{s}$ from various colliders and experiments. An artificial offset is applied to the $\sqrt{s}$ of each point for better visibility when two experiments share the same $\sqrt{s}$.

The work presented in this thesis aims to extend this programme by studying the polarisation structure of EW WZjj production, denoted WZjj-EW, using a combination of Run-2 and partial Run-3 datasets. This large dataset allows for the first time a detailed study of the polarisation structure of this process, which is the next logical milestone in the experimental study of VBS, following the establishment of inclusive diboson measurements and the first observation of VBS itself.

The aforementioned channel offers a good compromise between:

- the WWjj channel, which has a larger cross-section but also a larger background contribution due to the presence of two neutrinos in the final state, which makes the reconstruction of the event more challenging
- the ZZjj channel, which has a smaller cross-section but also a smaller background contribution due to the presence of four charged leptons in the final state, which makes the reconstruction of the event easier with a clean signature left in the detector.

This provides a detailed handle on the polarisation structure of EW VBS production while remaining experimentally accessible with the current dataset.

The main challenge, however, is the background contamination. When studying the WZjj-EW process, the main background contribution comes from the QCD-induced WZjj-QCD production, which has a much larger cross-section than the EW process, as reflected in Figure 6.3: the inclusive production of WZ pairs is dominated by this QCD-induced process. Indeed, as shown in Figure 6.2, the first measurement of pure EW WZjj-EW production was achieved only in 2019, with the 2015–2016 data ($36.1\,\mathrm{fb}^{-1}$), while the inclusive WZ production was measured much earlier, as seen in Figure 6.1. However, a clear feature of VBS events is the presence of two forward jets, which can be exploited to enhance the sensitivity to the EW process and reduce the background contribution. Requiring these two additional jets reduces the inclusive WZ background contribution, but the WZjj-QCD production still remains a significant background. More details on the particular topology of VBS events will be given in the following section. In [360], the most precise measurement of the cross-section of the WZjj-EW and WZjj-QCD processes for Run-2 at $\sqrt{s} = 13\,\mathrm{TeV}$ is presented:

$$\sigma^{\mathrm{fid.,13\,TeV}}_{\mathrm{WZjj\text{-}EW}\to\ell'\nu_\ell\ell\ell\mathrm{jj}} = 0.37 \pm 0.07\,\mathrm{fb}\,, \tag{6.1}$$

$$\sigma^{\mathrm{fid.,13\,TeV}}_{\mathrm{WZjj\text{-}QCD}\to\ell'\nu_\ell\ell\ell\mathrm{jj}} = 1.09 \pm 0.14\,\mathrm{fb}\,. \tag{6.2}$$

The WZjj-QCD process is thus about three times larger than the WZjj-EW process, making the measurement of the latter particularly challenging. The order of the cross-section, fb, can directly be compared to the one obtained in Figure 6.3, where the cross-section is of the order of $\sim 10^2\,\mathrm{pb}$. A difference of

10^5 is observed between the inclusive WZ production and the WZjj processes, which explains the latter measurement being achieved only recently, owing to the small cross-section and the need for a large dataset. The measurement of the polarisation structure of the WZjj-EW process is thus a challenging task that requires first a clear separation of the EW and QCD-induced processes, and then an additional splitting of the EW process into its different polarisation states.

This thesis will therefore provide a comprehensive study of the WZjj-EW process, with a particular focus on the polarisation structure of the vector bosons. This will be achieved through a combination of advanced analysis techniques, including multivariate analysis and ML, in combination with the latest MC generators and theoretical predictions to model the signal and background processes, as well as a detailed understanding of theoretical uncertainties associated with this measurement. Such work is built upon the previous measurements of inclusive diboson production and VBS production, building on the expertise developed in these analyses to push the sensitivity to the polarisation structure of the vector bosons in VBS production.

6.2 Vector boson scattering topology

VBS events present a unique topology, with a broad set of observables arising from the many particles involved. At LO the cross-section is of order α_{EW}^6 and corresponds to:

$$\mathrm{pp} \to \mathrm{W}^{\pm}\mathrm{Zjj} \to \ell'^{\pm}\nu_{\ell}\ell^{+}\ell^{-}\mathrm{jj}\,, \tag{6.3}$$

counting two quark–boson couplings, one quartic or two triple gauge vertices for the scattering itself, and one decay vertex for each boson.

At parton level the process is seen as two incoming quarks that each radiate an EW vector boson, the two bosons then rescattering into the final WZ pair. The two quarks continue forward and form the jets known as *tagging jets*. Three classes of diagram contribute [369]:

- t- and u-channel exchange of a W, Z, γ, or H between the two radiated bosons
- s-channel exchange through the triple gauge coupling WWZ
- a quartic WWZZ vertex.

Each of these diagrams is gauge-dependent and grows with energy in a way that would violate unitarity if it were considered on its own. Only the full coherent sum is physical and well-behaved at high energy, the Higgs contribution cancelling the divergences as shown in Appendix A. The Feynman diagram of the WZjj-EW process is shown in Figure 6.4, where the interaction between the two bosons is not drawn explicitly but can be any of the three classes above.

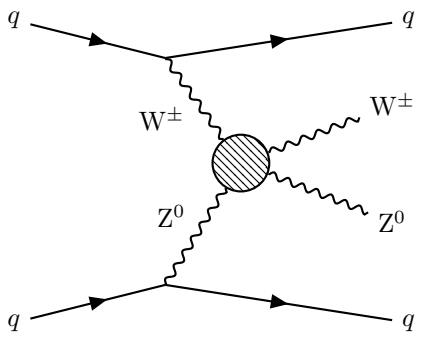


Figure 6.4: Feynman diagram of the WZjj process through VBS. The two incoming quarks radiate an EW boson that rescatters into the final WZ pair. The two outgoing quarks form the tagging jets.

The dominant background with the same final state is the QCD-induced WZjj production, which involves two strong-interaction vertices, associated with gluon exchange or QCD radiation, together with four EW vertices for the production and leptonic decay of the WZ pair. Its cross-section therefore scales as $\alpha_{\mathrm{S}}^2\,\alpha_{\mathrm{EW}}^4$ at LO, much larger than the α_{EW}^6 of the EW process. Two representative diagrams are shown in Figure 6.5.

The interference between the two processes arises from the product of their amplitudes and is therefore of order $\alpha_{\mathrm{S}}\,\alpha_{\mathrm{EW}}^5$ at LO. Although this nominal order in fact exceeds that of the pure EW process (since $\alpha_{\mathrm{S}} > \alpha_{\mathrm{EW}}$), the interference is strongly suppressed in practice, to the percent level, by the mismatch in colour structure between the two processes (colour-singlet versus colour-octet exchange) and by the different kinematic regions they populate. It is consequently expected to be small.

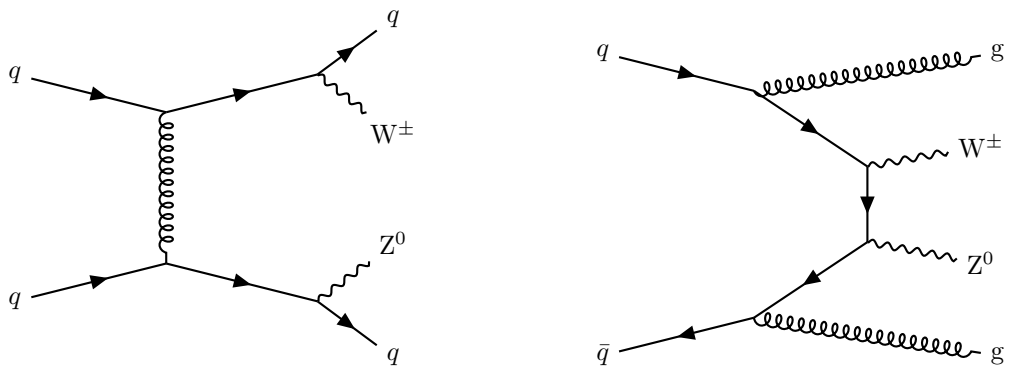


Figure 6.5: Feynman diagrams of the WZjj process through QCD-induced production. The two incoming quarks interact through a gluon exchange or QCD radiation, and produce the WZ pair through EW vertices.

6.2.1 Object reconstruction

The WZjj-EW process is characterised by a final state with three charged leptons, one neutrino, and two jets. The reconstruction of these objects is therefore crucial for the analysis of this process. The following sections will describe the reconstruction of each of these objects in the ATLAS detector except for the electron that is already described in Chapter 5.

6.2.1.1 Muons

Muons are leptons roughly 200 times heavier than electrons, and as a result they radiate far less, since bremsstrahlung is suppressed by the inverse square of the mass. They therefore traverse the calorimeters as MIPs without showering, losing only a small amount of energy through ionisation. As a result, muons penetrate matter much more easily than electrons and are the only charged particles that reach the outermost layers of the detector, where the MS, detailed in Section 3.4, is located. This set of detectors measures their trajectory and momentum within the toroidal magnetic field over $|\eta| < 2.7$, while its trigger chambers provide a dedicated muon trigger over $|\eta| < 2.4$. This results in a clean and unambiguous signature for muons, which can be identified with high efficiency and low background contamination.

6.2.1.1.1 Reconstruction Since muons are reconstructed from tracks in two independent systems [370], the ID and the MS, their reconstruction amounts to combining the two measurements, and their calibration to correcting the momentum scale and resolution of the combined track. A schematic view of the muon reconstruction is shown in Figure 6.7(b). We will not redefine here the track reconstruction itself: the concepts of seeds, combinatorial filtering, and ambiguity resolution introduced in Section 5.1.2 apply identically. In the MS, the reconstruction starts from the hits in the precision drift chambers, which are first combined within each station into local straight-line *segments* by a pattern-recognition step in the bending plane similar to the one used in the ID. Segments from different stations that are compatible with a single trajectory through the toroidal field are then fitted together to form a spectrometer track, the curvature of which provides a standalone momentum measurement using the known magnetic field and the measurement of the sagitta.

Depending on which sub-detectors contribute to the measurement, several types of muon candidates are defined:

- *Combined* muons: a track reconstructed independently in the MS is matched to an ID track, and a combined fit of all the hits of the two systems, accounting through the calorimeter material for the energy lost by the muon, provides the final track. This outside-in approach is the main category, with the best momentum resolution and purity. The re-fit of the hits of both the ID and the MS yields the final track parameters, and the momentum is taken from the combined fit. For $|\eta| > 2.5$ the track formed by the pixel and SCT can still be combined with the MS: this is referred to as a Si-associated forward muon. A complementary *inside-out* algorithm starts instead from an ID track and extrapolates it outwards, searching for compatible spectrometer hits. It recovers muons that cross regions of reduced spectrometer coverage or that have too few hits to be reconstructed standalone.

- *Segment-tagged* muons: an ID track extrapolated to the spectrometer is matched to at least one local segment, without requiring a full spectrometer track. This recovers low-p_{T} muons that do not traverse enough stations, the momentum being taken from the ID alone.
- *Calorimeter-tagged* muons: an ID track is matched to a calorimeter energy deposit compatible with an MIP. Although the least pure category, it recovers the acceptance around $|\eta| < 0.1$, where the spectrometer is only partially instrumented to allow for the passage of the services of the ID and calorimeters.
- *Extrapolated* muons: a spectrometer track loosely compatible with originating from the IP, but without a matching ID track. These extend the acceptance to $2.5 < |\eta| < 2.7$, beyond the tracker coverage, the momentum being measured by the spectrometer alone.

When several candidates share the same ID track, an overlap-removal procedure keeps the highest-quality one, combined muons taking precedence, so that each muon is reconstructed only once.

6.2.1.1.2 Identification As for electrons, the reconstructed candidates are contaminated by non-prompt muons, dominated by in-flight decays of charged pions and kaons. Such decays in the ID volume produce a kink in the trajectory, degrading the quality of the combined fit and the compatibility between the independent ID and MS momentum measurements. The identification therefore relies on a set of track-quality and compatibility variables: the number of hits in the various sub-detectors, the normalised χ^2 of the combined track fit, the charge-to-momentum significance defined as:

$$\frac{\left|\left(\frac{q}{p}\right)_{\mathrm{ID}} - \left(\frac{q}{p}\right)_{\mathrm{MS}}\right|}{\sqrt{\sigma^2\left(\frac{q}{p}\right)_{\mathrm{ID}} + \sigma^2\left(\frac{q}{p}\right)_{\mathrm{MS}}}}\,, \tag{6.4}$$

which compares the charge-signed inverse momenta measured by the two systems normalised to their combined uncertainty, and the analogous relative difference ρ between the two transverse-momentum measurements with respect to the combined (CB) one:

$$\rho = \frac{\left|p_{\mathrm{T}}^{\mathrm{MS}} - p_{\mathrm{T}}^{\mathrm{ID}}\right|}{p_{\mathrm{T}}^{\mathrm{CB}}}\,. \tag{6.5}$$

In analogy with the electron WPs, three inclusive WPs of increasing purity are defined by progressively tightening these requirements: `Loose`, `Medium`, and `Tight` [371].

The `Medium` WP is the default choice and the best compromise between efficiency and purity. Within the ID acceptance it accepts combined and inside-out muons with at least two precision stations, the $\frac{q}{p}$ significance being required to be below seven to ensure a loose agreement between the ID and MS measurements. In the central crack $|\eta| < 0.1$, where the spectrometer coverage is incomplete, candidates with a single precision station are also accepted provided they have at most one precision-hole station. Beyond the tracker coverage, in $2.5 < |\eta| < 2.7$, the acceptance is extended by extrapolated and Si-associated forward muons with at least three precision stations. This selection yields a per-mille-level fake rate and a high efficiency for prompt muons over a wide kinematic range: more than $98\,\%$ of the prompt muons it selects are combined muons.

The `Loose` WP is optimised for analyses with high muon multiplicity, maximising the efficiency at the cost of a slightly lower purity. It accepts all `Medium` muons and, in addition, recovers the central crack by including calorimeter-tagged and segment-tagged muons in $|\eta| < 0.1$. To further improve the efficiency at low p_{T}, inside-out muons with $p_{\mathrm{T}} < 7\,\mathrm{GeV}$ and a single precision station are accepted in $|\eta| < 1.3$ when they are independently confirmed by the segment-tagged reconstruction. The resulting efficiency gain with respect to `Medium` reaches about $20\,\%$ for $3 < p_{\mathrm{T}} < 5\,\mathrm{GeV}$ and $1\,\%$ to $2\,\%$ at higher p_{T}.

The `Tight` WP, conversely, maximises the purity for analyses limited by non-prompt background, at the price of a reduced efficiency. Only combined and inside-out muons with at least two precision stations are retained, the normalised χ^2 of the combined fit is required to be below eight to reject the pathological tracks induced by hadron decays in flight, and tighter $|\eta|$-dependent requirements are applied to the $\frac{q}{p}$ significance and to ρ.

Two dedicated WPs complete the menu. The *Low-*p_{T} WP targets muons below $\sim 10\,\mathrm{GeV}$ (down to $3\,\mathrm{GeV}$), where the spectrometer information is scarce and very low-p_{T} muons may not even reach its middle station. It is built from combined and inside-out muons requiring at least one precision station, and suppresses in-flight hadron decays through the momentum-balance significance (MBS):

$$\mathrm{MBS} = \frac{p_{\mathrm{T}}^{\mathrm{ID}} - p_{\mathrm{T}}^{\mathrm{MS}} - E_{\mathrm{loss}}}{\sigma(E_{\mathrm{loss}})} < 3\,, \tag{6.6}$$

together with scattering-angle significance variables: above $18\,\mathrm{GeV}$ it converges to the `Medium` requirements. A multivariate variant based on a BDT is also available. The *High-*p_{T} WP instead targets muons with $p_{\mathrm{T}} > 100\,\mathrm{GeV}$, as needed by searches in extreme phase-space regions, optimising the momentum resolution where the spectrometer sagitta measurement becomes the limiting factor towards the TeV scale. Starting from `Medium` combined muons, it requires at least three precision stations to sharpen the sagitta resolution, four in the regions where the toroidal field integral changes sign, together with an upper bound on the $\frac{q}{p}$ uncertainty: these requirements cannot be met in the barrel–end-cap transition around $1.0 < |\eta| < 1.1$.

6.2.1.1.3 Isolation Prompt muons from W and Z decays, for example, are further separated from the non-prompt muons produced inside HF jets by requiring them to be isolated. The two isolation variables are constructed exactly as for electrons in Section 5.4, with only minor muon-specific adaptations.

The calorimeter-based isolation $E_{\mathrm{T}}^{\mathrm{cone}\Delta R}$ is built within a cone of radius $\Delta R = 0.2$ around the muon direction, from the scalar sum of the transverse energies of the positive-energy topological clusters whose barycentre falls within the cone. As for electrons, the candidate's own deposit is removed through a core subtraction, and the pile-up and underlying-event contribution is subtracted with the ambient energy-density method of Section 5.4:

$$E_{\mathrm{T}}^{\mathrm{cone}\Delta R} = E_{\mathrm{T,raw}}^{\mathrm{isol}\Delta R} - E_{\mathrm{T,core}} - E_{\mathrm{T,pile\text{-}up}}(\eta, \Delta R)\,, \tag{6.7}$$

where the shower-leakage term present for electrons is negligible here, since a muon does not develop an EM shower and deposits only a small minimum-ionising signal. This variable is sensitive to the neutral component of the activity around the muon but is more pile-up dependent.

The track-based isolation $p_{\mathrm{T}}^{\mathrm{varcone}\Delta R_{\mathrm{max}}}$ is built from the scalar sum of the transverse momenta of the ID tracks falling within a cone around the muon, excluding the muon track itself. The selected tracks must have $p_{\mathrm{T}} > 1\,\mathrm{GeV}$ and originate from the primary vertex through the requirement $|z_0 \sin\theta| < 3\,\mathrm{mm}$, which makes the variable largely insensitive to pile-up at the price of being blind to the neutral component. As for electrons, a variable cone is used:

$$\Delta R = \min\left(\frac{10\,\mathrm{GeV}}{p_{\mathrm{T}}},\ \Delta R_{\mathrm{max}}\right)\,, \tag{6.8}$$

but with a larger default maximum size $\Delta R_{\mathrm{max}} = 0.3$, giving $p_{\mathrm{T}}^{\mathrm{varcone30}}$, against the 0.2 used for electrons. The shrinking cone preserves the efficiency for muons produced in boosted topologies, where a genuine prompt muon may be accompanied by nearby activity from the recoiling system.

Several WPs are defined from requirements on these variables, ranging from loose fixed-threshold selections to p_{T}-dependent ones targeting a constant efficiency across the spectrum.

Recent developments for Run-3 have extended the isolation toolkit beyond these classical variables, exactly as for electrons in Section 5.4. First, PFlow isolation variants have been introduced [300], in which the calorimeter component is built from neutral PFlow objects rather than from topological clusters: the energy already associated with reconstructed charged tracks is removed, which improves the pile-up robustness of the calorimeter isolation and reduces its residual dependence on the ambient energy-density correction. The corresponding `PflowLoose` and `PflowTight` WPs combine this PFlow calorimeter isolation with the track-based isolation and largely supersede the purely calorimetric selections. Second, the same multivariate prompt-lepton discriminants used for electrons, the PLV [301] and its improved version the PLIV [302], are also defined for muons, combining the calorimeter and track isolation, the impact-parameter significances, and information on nearby tracks in a BDT or a DNN to reach a substantially stronger rejection of non-prompt muons from HF decays than the classical isolation WPs alone. These discriminants are the default choice for Run-3 analyses requiring strong non-prompt-lepton rejection, such as multi-lepton searches and measurements in top-quark-rich final states.

6.2.1.1.4 Momentum calibration The muon momentum measured by the combined track has to be corrected for residual mis-modelling of the detector [372], exactly as the electron energy scale of Section 5.2. The corrections are applied separately to the ID and MS measurements before their combination, and are parametrised as:

$$p_{\mathrm{T}}^{\mathrm{calib}} = \frac{p_{\mathrm{T}}^{\mathrm{reco}} + s_0 + s_1 \cdot p_{\mathrm{T}}^{\mathrm{reco}}}{1 + g_0 \cdot \dfrac{r_0}{p_{\mathrm{T}}^{\mathrm{reco}}} + g_1 \cdot r_1 + g_2 \cdot r_2 \cdot p_{\mathrm{T}}^{\mathrm{reco}}}\,, \tag{6.9}$$

where the s_n terms correct the momentum scale, s_0 accounting for the modelling of the energy loss in the material upstream of the MS (and set to zero for the ID, where the energy loss is negligible) and s_1 for the magnetic field integral and the radial dimensions of the detector, while the denominator terms smear the

resolution with g_m normally distributed random variables, following the three contributions to the relative momentum resolution: the fluctuations of the energy loss ($\propto \frac{1}{p_\mathrm{T}}$, negligible and set to zero in practice), the multiple scattering (constant), and the intrinsic spatial resolution and residual misalignments ($\propto p_\mathrm{T}$). The parameters are extracted in regions of (η, ϕ) from template fits to the invariant mass distributions of $\mathrm{Z} \to \mu\mu$ and $\mathrm{J}/\psi \to \mu\mu$ events, in close analogy with the in-situ $\mathrm{Z} \to \mathrm{ee}$ calibration of the electron energy scale.

A residual charge-dependent bias remains, however, beyond the reach of these resonance fits: any radial misalignment of the detector modifies the measured sagitta, introduced in Section 5.1.2, in opposite directions for the two charges, biasing the momenta as $\frac{q}{p_\mathrm{T}} \to \frac{q}{p_\mathrm{T}} + \delta_\mathrm{sagitta}$ while leaving the dimuon invariant mass largely unchanged. The corresponding correction applied to the data reads as:

$$p_\mathrm{T}^\mathrm{corr} = \frac{p_\mathrm{T}^\mathrm{reco}}{1 - q\delta_\mathrm{sagitta}(\eta, \phi) p_\mathrm{T}^\mathrm{reco}} \,. \tag{6.10}$$

This bias is corrected in-situ by minimising the charge asymmetry of the $\mathrm{Z} \to \mu\mu$ peak position across the (η, ϕ) regions: it is reduced on average from up to $0.4\,\mathrm{TeV}^{-1}$ to $2 \cdot 10^{-4}\,\mathrm{TeV}^{-1}$ for momenta up to $450\,\mathrm{GeV}$, with a residual uncertainty at the level of $0.03\,\mathrm{TeV}^{-1}$. After the full calibration, at the Z peak the momentum scale is known with an uncertainty of $0.05\,\%$ in the barrel, rising to about $0.15\,\%$ in the forward region. The momentum resolution itself, of the order of $2\,\%$ at central rapidities and up to $\sim 3\,\%$ in the forward region for muons from Z decays, is measured with an uncertainty of $1.5\,\%$ ($2\,\%$ at the J/ψ peak).

A schematic representation of the effect of the momentum-scale and resolution corrections on the dimuon invariant mass in $\mathrm{Z} \to \mu\mu$ events is shown in Figure 6.6. The dimuon mass distribution is shifted and broadened by the mis-modelling of the detector, and the calibration restores the correct peak position and width.

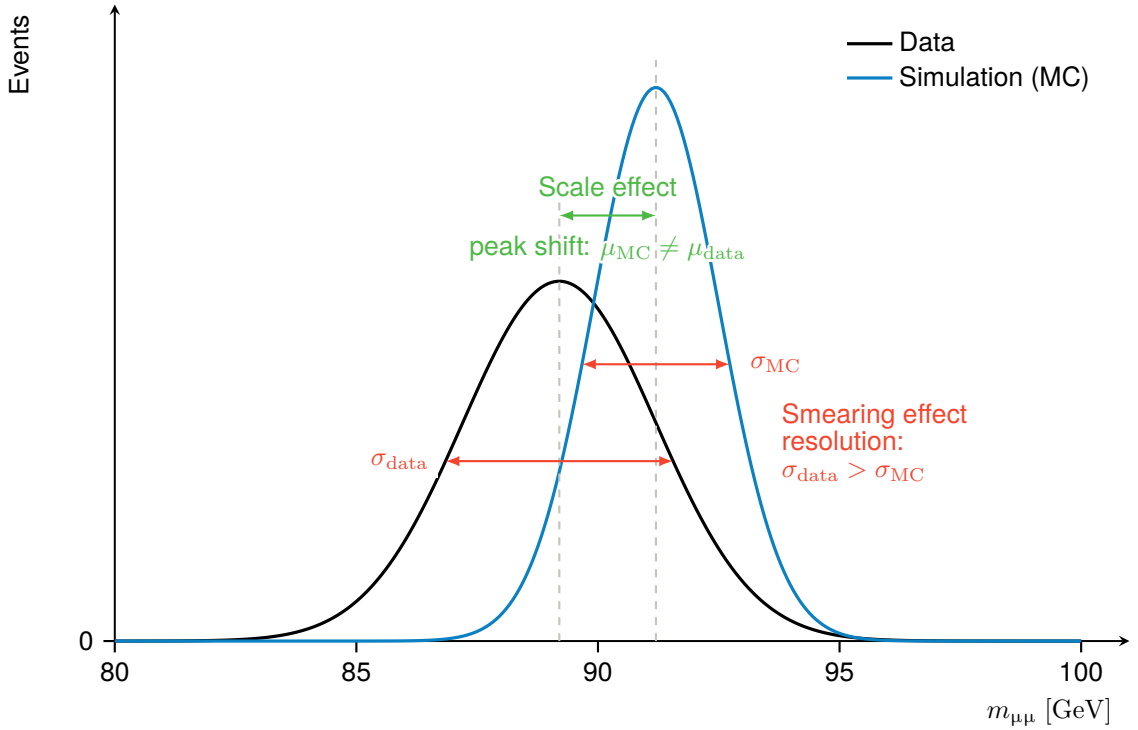


Figure 6.6: Schematic representation of the effect of the scaling and smearing corrections on $m_{\mu\mu}$ in $\mathrm{Z} \to \mu\mu$ events. Not to scale.

6.2.1.1.5 Efficiency measurements and data/MC corrections As for electrons, the simulation does not perfectly reproduce the reconstruction, identification, and isolation efficiencies measured in data, and correction factors are derived following the same tag-and-probe strategy as for electrons in Section 5.3.7. Two resonances are used in complementary kinematic ranges: $\mathrm{Z} \to \mu\mu$ provides a high-purity sample of prompt muons above $p_\mathrm{T} \approx 10\,\mathrm{GeV}$, while $\mathrm{J}/\psi \to \mu\mu$ extends the measurement down to $p_\mathrm{T} \approx 3\,\mathrm{GeV}$.

The SFs, derived in bins of (p_T, η), are found to be close to unity, with uncertainties at the per mille level for the reconstruction and identification over most of the kinematic range, significantly smaller than their electron counterparts thanks to the cleaner muon signature. The largest deviations are localised in

the poorly instrumented regions, in particular around $|\eta| < 0.1$, where the calorimeter-tagged candidates dominate. The SFs for the trigger, reconstruction, identification, track-to-vertex association (TTVA), and isolation efficiencies, together with the momentum-scale and resolution corrections of Paragraph 6.2.1.1.4, constitute the full set of muon corrections.

6.2.1.1.6 Run-3 developments The muon reconstruction described above is largely common to Run-2 and Run-3, but a few elements changed at the start of Run-3. The most important is the replacement, during LS-2, of the inner end-cap small wheels by the NSW: the MDT and CSC precision layers and the TGC trigger of the original small wheel are replaced by sTGC and MM, covering $1.3 < |\eta| < 2.7$. The MMs provide the precision tracking and the sTGCs the fast L1 trigger, together sharpening the forward momentum measurement and strongly suppressing the fake end-cap triggers induced by cavern background. A layer of inner-barrel RPCs (BIS78) similarly improves the barrel trigger acceptance, both being precursors of the full upgrades foreseen for the HL-LHC as explained in Section 3.9.

On the software side, a modified set of quality WPs with dedicated requirements in the NSW region is used, the calorimeter-tagged muons rely on a CNN tagger, and the PFlow isolation discussed above becomes the default. These changes preserve the Run-2 performance under the higher pile-up of Run-3, where the average number of interactions per bunch crossing reaches $\langle\mu\rangle \approx 60$.

6.2.1.2 Jets

Because the LHC collides protons, hadronic activity dominates the final state of most events. As explained in Section 2.5, due to the colour confinement, the quarks and gluons produced in the hard-scattering process (or underlying events) cannot be observed as free particles. As a coloured parton moves away from the IP, the potential energy stored in the colour field grows linearly with the separation until it becomes energetically favourable to create a new quark–antiquark pair from the vacuum. This *fragmentation* cascade repeats until all the available energy has been converted into colourless hadrons, which then decay further. The net result is a collimated spray of hadrons travelling along the direction of the initial parton, referred to as a jet [135, 136], whose total momentum and direction reflect those of the parton that initiated it.

Therefore, contrary to other objects such as electrons, muons, or photons, which can be identified and reconstructed as single particles, a jet is a composite object made of many hadrons. The number of hadrons in a jet can vary from a few to several hundreds, depending on the energy of the initiating parton and the details of the fragmentation process.

The reconstruction of a jet proceeds in three steps: the constituents are first built by combining the information of the ID and of the calorimeters following the PFlow approach, they are then clustered into jets with the anti-k_t algorithm, and finally, as the calorimeter response to hadrons differs significantly from the EM scale at which the cells are calibrated, a dedicated multistep calibration is applied to bring the jet energy back to the true particle-level scale.

Two additional essential components consist of the suppression of jets originating from pile-up and the identification of jets containing b-hadrons.

6.2.1.2.1 Hadrons in the detector Before describing how the constituents are built, it is worth recalling which particles actually reach the detector and how each is measured, as this is what motivates the PFlow approach. The fragmentation of a parton produces numerous hadrons, but only those living long enough to traverse a macroscopic distance are seen directly: by convention, the *truth jets* used to define the particle-level scale are clustered from all final-state particles with a laboratory lifetime such that $\tau > 30\,\mathrm{ps}$ ($c\tau > 9\,\mathrm{mm}$), excluding muons and neutrinos, which are treated as separate objects.

On average, roughly two thirds of the energy of a jet is carried by charged hadrons, predominantly charged pions, about a quarter by photons from the decay $\pi^0 \to \gamma\gamma$, and the remaining $\sim$10 % by long-lived neutral hadrons, mainly neutrons and K^0_L. These three components are measured very differently:

- Charged hadrons ($\pi^\pm$, $\mathrm{K}^\pm$, p, …) ionise the ID and leave a track before depositing their energy as a hadronic shower in the calorimeters. Their momentum is best measured by the tracker at low p_T, and they can be associated with the hard-scattering vertex.
- Photons from π^0 decays develop EM showers, well measured by the EM calorimeter, with no associated track.
- Neutral hadrons such as the neutron and the K^0_L leave no track and are visible only through their hadronic shower, the calorimeter being the sole handle on this component.

An intermediate case is that of the weakly decaying strange hadrons K^0_S ($c\tau \approx 2.7\,\mathrm{cm}$) and Λ ($c\tau \approx 7.9\,\mathrm{cm}$), which decay inside the tracking volume into charged hadrons and are reconstructed from the tracks of those decay products, emerging from a displaced secondary vertex. Shorter-lived resonances decay essentially at the IP and enter only through their stable decay products. This separation of the jet energy into a charged component measured in the tracker, an EM component, and a neutral hadronic component measured in the calorimeters is precisely what the PFlow reconstruction described below exploits.

A visual representation of the different components of a jet and how they are measured in the ATLAS detector is shown in Figure 6.7(a). The charged hadrons are reconstructed as tracks in the ID and deposit their energy in the calorimeters, while the photons from π^0 decays develop EM showers and the neutral hadrons deposit their energy through hadronic showers.

6.2.1.2.2 Particle-flow reconstruction A jet contains a multitude of particles, both charged and neutral, which can be reconstructed with different sub-detectors. As seen for example with the electron and photon in Chapter 5, the ID provides tracks for charged particles, while the calorimeters measure the energy deposited by both charged and neutral particles. Until the early Run-2 analyses, ATLAS reconstructed jets using the same topo-clusters defined in Section 5.1.1 but keeping the hadronic component. Since the full Run-2 reprocessing, the baseline jet reconstruction is based on the PFlow approach [300]: the tracking and calorimeter information are both combined to reconstruct each individual stable particle of the event, combining the best of both worlds. This results in particle-flow objects (PFOs) that are then used as input to the jet clustering algorithm. A schematic representation of the PFlow reconstruction is shown in Figure 6.8.

This approach exploits the complementarity of the two sub-detectors:

- At low energy, the momentum resolution of the ID is significantly better than the energy resolution of the calorimeters as seen in Section 3.6. It provides a superior angular resolution and recovers low-p_T charged particles that are swept out of the jet cone by the solenoid field before reaching the calorimeters. It also allows the constituents to be associated with the hard-scattering vertex, serving as the main lever to suppress pile-up.
- On the other hand, the calorimeters are more efficient at high energy, where the ID is limited by the number of hits in the tracking layers and by the bending of the tracks in the magnetic field. It is also the only sub-detector capable of measuring neutral particles, which are not reconstructed in the ID. Its coverage is also larger than the ID, allowing the reconstruction of particles at high pseudorapidity. The drawback is that the energy obtained from the calorimeters has to be subtracted from the energy of the charged particles reconstructed in the ID to avoid double counting.

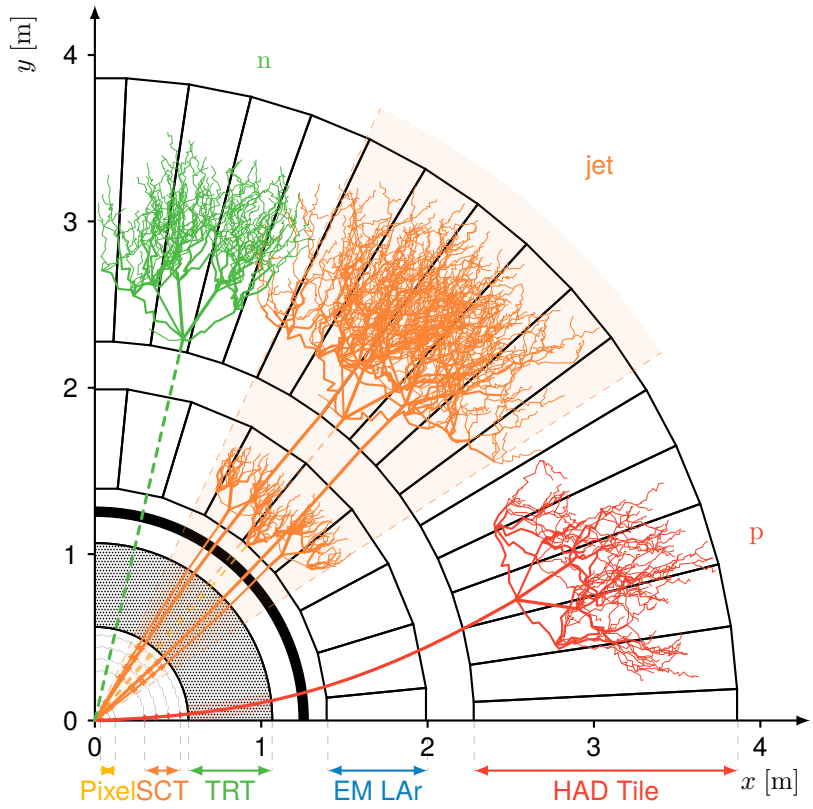
The PFlow reconstruction takes as inputs the topo-clusters of Section 5.1.1 and the tracks of Section 5.1.2. The tracks are selected with $p_\mathrm{T} > 0.5\,\mathrm{GeV}$ and $|\eta| < 2.5$, with at least nine Si hits and no missing hits in the Pixel Detector. The tracks are not removed, but the calorimeter measurement is increasingly preferred over the track as p_T grows, with the subtraction effectively off by $p_\mathrm{T} \sim 100\,\mathrm{GeV}$ to avoid mis-measurements at high energy reflecting the better energy resolution of the calorimeters. Energetic particles are in any case embedded in the dense core of the jet, where they are poorly isolated from the nearby activity and an accurate subtraction of their calorimeter energy is not possible. Tracks matched to medium-quality electrons or muons are also masked: the algorithm is designed to reconstruct hadrons.

Each track passing the selection is then matched to a topo-cluster by requiring $\frac{E^\mathrm{cluster}}{p_\mathrm{T}^\mathrm{track}} > 0.1$, and then ranking the candidate clusters according to their distance:

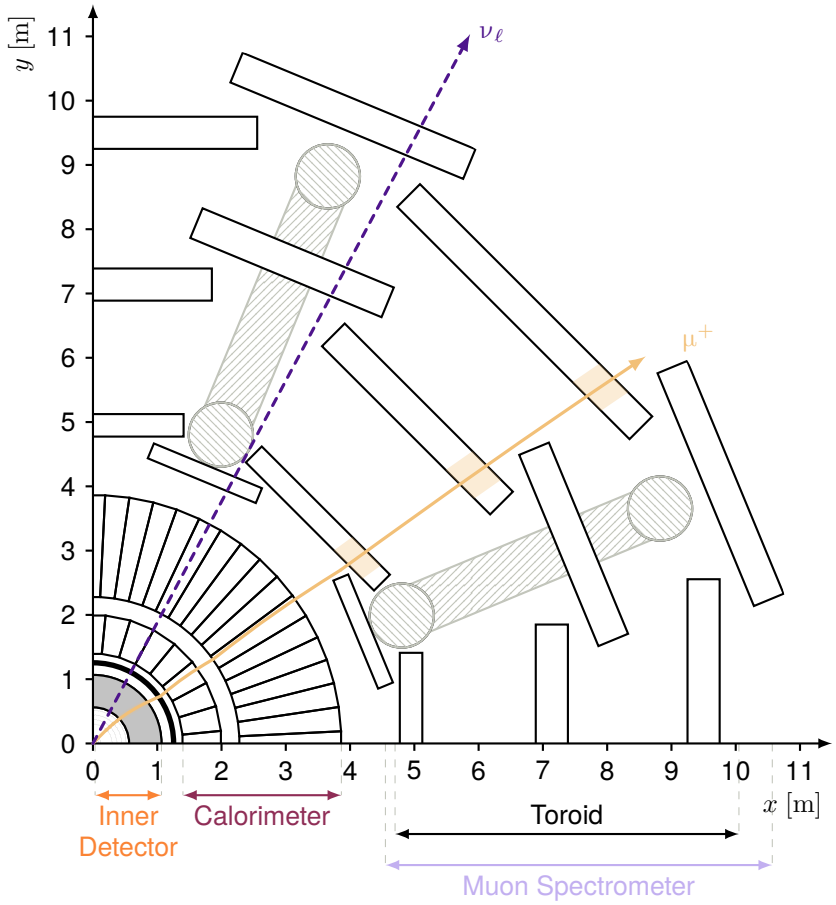
$$\Delta R = \sqrt{\left(\frac{\Delta\eta}{\sigma_\eta}\right)^2 + \left(\frac{\Delta\phi}{\sigma_\phi}\right)^2}, \tag{6.11}$$

where $\Delta\eta$ and $\Delta\phi$ are computed between the cluster barycentre and the extrapolation of the track to the Layer 2 of the EM calorimeter, while σ_η and σ_ϕ are the lateral widths of the cluster in the η and ϕ directions. The energy expected to be deposited in the calorimeter by the particle that produced the track is estimated following:

$$\left\langle E^\mathrm{expected}_\mathrm{deposit} \right\rangle = p^\mathrm{track} \cdot \left\langle \frac{E^\mathrm{cluster}_\mathrm{ref}}{p^\mathrm{track}_\mathrm{ref}} \right\rangle, \tag{6.12}$$



(a) A proton, a neutron, and a jet



(b) A muon and a neutrino

Figure 6.7: Reconstruction of physics objects in the ATLAS detector, drawn in the transverse plane: a proton, a neutron, and a jet (a), and a muon and a neutrino, the latter inferred from the missing transverse momentum (b).

where the reference ratio is derived from simulated single-pion samples without pile-up from a cone of radius $\Delta R = 0.4$ around a track. The compatibility between the matched cluster and this expectation is quantified by the following significance:

$$Z(E^{\text{cluster}}) = \frac{E^{\text{cluster}} - \left\langle E^{\text{expected}}_{\text{deposit}} \right\rangle}{\sigma(E^{\text{expected}}_{\text{deposit}})}\,, \tag{6.13}$$

where $\sigma(E^{\text{expected}}_{\text{deposit}})$ is the spread of the reference $\frac{E^{\text{cluster}}_{\text{ref}}}{p^{\text{track}}_{\text{ref}}}$ distribution. If $Z(E^{\text{cluster}}) < -1$, the shower is assumed to be split among several clusters and the additional topo-clusters are searched for in a cone of $\Delta R < 0.2$ around the track. The expected energy $\langle E^{\text{expected}}_{\text{deposit}} \rangle$ is then subtracted from the matched set of clusters, working outwards from the most energetic cell and proceeding through concentric rings of cells whose granularity matches the lateral shower profile expected for the energy and position of the particle. After the subtraction, if the energy remaining in the cluster system is smaller than $1.5\sigma(E^{\text{expected}}_{\text{deposit}})$ (compatible with the fluctuations of a single shower), the remnant is removed entirely, on the assumption that it is consistent with the residual of the shower already accounted for by the track.

Once a set of topo-clusters has been matched to a track, the cell-by-cell subtraction of the expected energy deposit is applied only if the following criterion is satisfied:

$$\frac{E^{\text{cluster}} - \langle E^{\text{expected}}_{\text{deposit}} \rangle}{\sigma(E^{\text{expected}}_{\text{deposit}})} < 33.2 \times \log_{10}\left(\frac{40\,\text{GeV}}{p^{\text{track}}_{\text{T}}}\right)\,. \tag{6.14}$$

The right-hand side decreases monotonically with the track momentum: it is positive below $p^{\text{track}}_{\text{T}} = 40\,\text{GeV}$, vanishes at $40\,\text{GeV}$, and turns increasingly negative above, so that the subtraction is progressively switched off and becomes effectively inactive by $p^{\text{track}}_{\text{T}} \sim 100\,\text{GeV}$. The left-hand side carries the complementary information: a large E^{cluster} relative to the expectation raises the significance and likewise fails the criterion, suppressing the subtraction in dense environments such as the cores of energetic jets, where the energy associated with the track cannot be cleanly disentangled from that of the surrounding particles. In both regimes the calorimeter measurement is retained in place of an unreliable subtraction. This smooth treatment replaces the hard $p^{\text{track}}_{\text{T}} < 40\,\text{GeV}$ veto of the original PFlow implementation [300] and was introduced for the Run-2 calibration [373].

The final set of PFOs consists of the selected tracks that are matched to the hard-scattering vertex, representing the charged component, and of the positive-energy topo-clusters surviving the subtraction, representing the neutral one. The two are then treated on the same footing as inputs to the jet clustering.

This combination brings two benefits. First, the angular and energy resolution of low-p_{T} jets improves significantly, the tracker measurement replacing the poorly resolved calorimeter one for the charged constituents: the jet energy resolution (JER) is improved by up to $\sim$25 % to 30 % at low p_{T} with respect to the calorimeter-only reconstruction. Second, because the charged constituents are individually associated with the hard-scattering vertex, the reconstruction is markedly more robust against pile-up, the jet response and resolution being substantially more stable as a function of the number of interactions per bunch crossing.

6.2.1.2.3 Jet clustering The PFOs are clustered into jets using the anti-k_t algorithm [374], as implemented in the `FastJet` package [298]. It belongs to the family of sequential recombination algorithms, a category of clustering algorithms addressed in Section 4.3, which iteratively merge the pair of objects closest in a generalised distance. This family is parametrised by an exponent $2n$ acting on the transverse momenta:

$$d_{ij} = \min\left(p^{2n}_{\text{T},i}, p^{2n}_{\text{T},j}\right)\frac{\Delta^2_{ij}}{R^2}\,, \qquad d_i = p^{2n}_{\text{T},i}\,, \tag{6.15}$$

where i and j run over the input objects, $\Delta^2_{ij} = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ is their separation in the rapidity–azimuth plane, d_i is the distance between object i and the beam, and R is the radius parameter. The latter controls the size of the jet: if chosen too large, the contamination from the underlying event and pile-up becomes significant, whereas if chosen too small, the jet does not capture the full shower. A typical value is $R = 0.4$, the default for most ATLAS analyses, while larger values up to $R = 1.0$ are used for boosted topologies, where the decay products of a heavy particle are collimated into a single jet.

We distinguish three distance measures depending on the value of n, which determine the order in which the objects are clustered:

- For $n = 1$, the distance is proportional to the transverse momentum squared: low-p_{T} pairs give the smallest d_{ij}, so soft objects cluster first. This is the k_t algorithm [297, 375], whose boundaries follow the soft radiation and are therefore irregular.

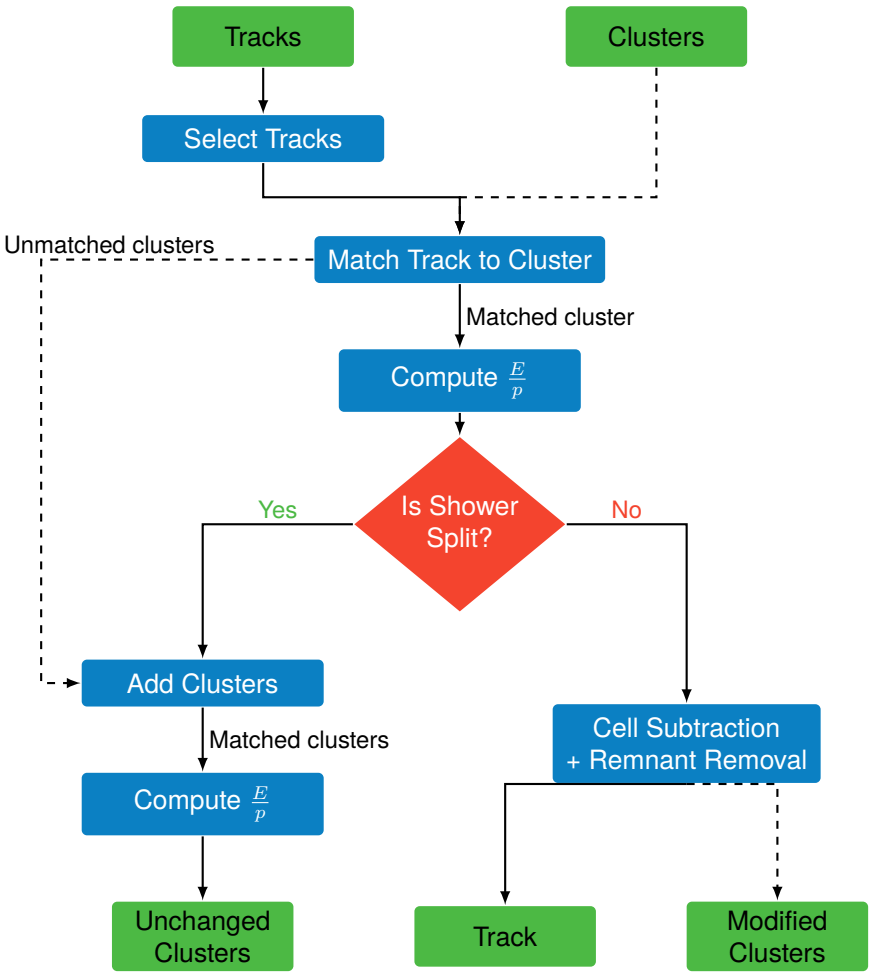


Figure 6.8: Schematic representation of the PFlow reconstruction steps.

- For $n = 0$, the distance is independent of the transverse momentum and the clustering proceeds purely by angle. This is the Cambridge/Aachen algorithm.
- For $n = -1$, the distance is inversely proportional to the transverse momentum squared: high-p_{T} objects give the smallest d_{ij}, so hard objects cluster first. This is the anti-k_t algorithm, in which a hard object sweeps up the surrounding soft radiation into a rigid, cone-like jet of radius R.

At each step, the smallest distance is identified: if it is a d_{ij}, the objects i and j are merged by summing their four-momenta. If it is a d_i, the object i is declared a jet and removed from the list. The procedure is repeated until no input objects remain. The radius R therefore fixes the angular reach of a jet: an object keeps merging while another lies within R of it, and is emitted as a jet once nothing remains within R of it. The two algorithms and the anti-k_t clustering sequence are illustrated in Figures 6.9(a) and 6.9(b).

The $n = -1$ choice makes the clustering start from the hardest objects, which accumulate the soft activity around them before the soft objects cluster among themselves. The resulting jets are regular cones of radius R centred on the hard constituents, with boundaries insensitive to soft radiation. Most importantly, the algorithm is infrared- and collinear-safe: the set of reconstructed jets is unchanged by the emission of an infinitely soft particle or by the collinear splitting of an existing one, a property required for the jet cross-sections to be free of divergences and thus calculable at all orders in perturbation theory. The same jet definition can therefore be applied identically to truth particles, to detector-level constituents, and to the partons of a fixed-order calculation, which is essential for relating the measurement to theory. Each jet is also assigned a catchment *area* A [376], which quantifies its susceptibility to soft contamination and is exploited in the pile-up subtraction.

6.2.1.2.4 Energy calibration In the same spirit as the electron and photon calibration chain presented in Section 5.2, the reconstructed jet four-momentum has to be corrected back to the particle-level scale, defined by *truth jets* clustered from the stable final-state particles of the simulation. The initial energy of the constituents is set at the EM scale, which correctly describes the response to electrons and photons, but underestimates the energy of hadronic showers due to the non-compensating nature of

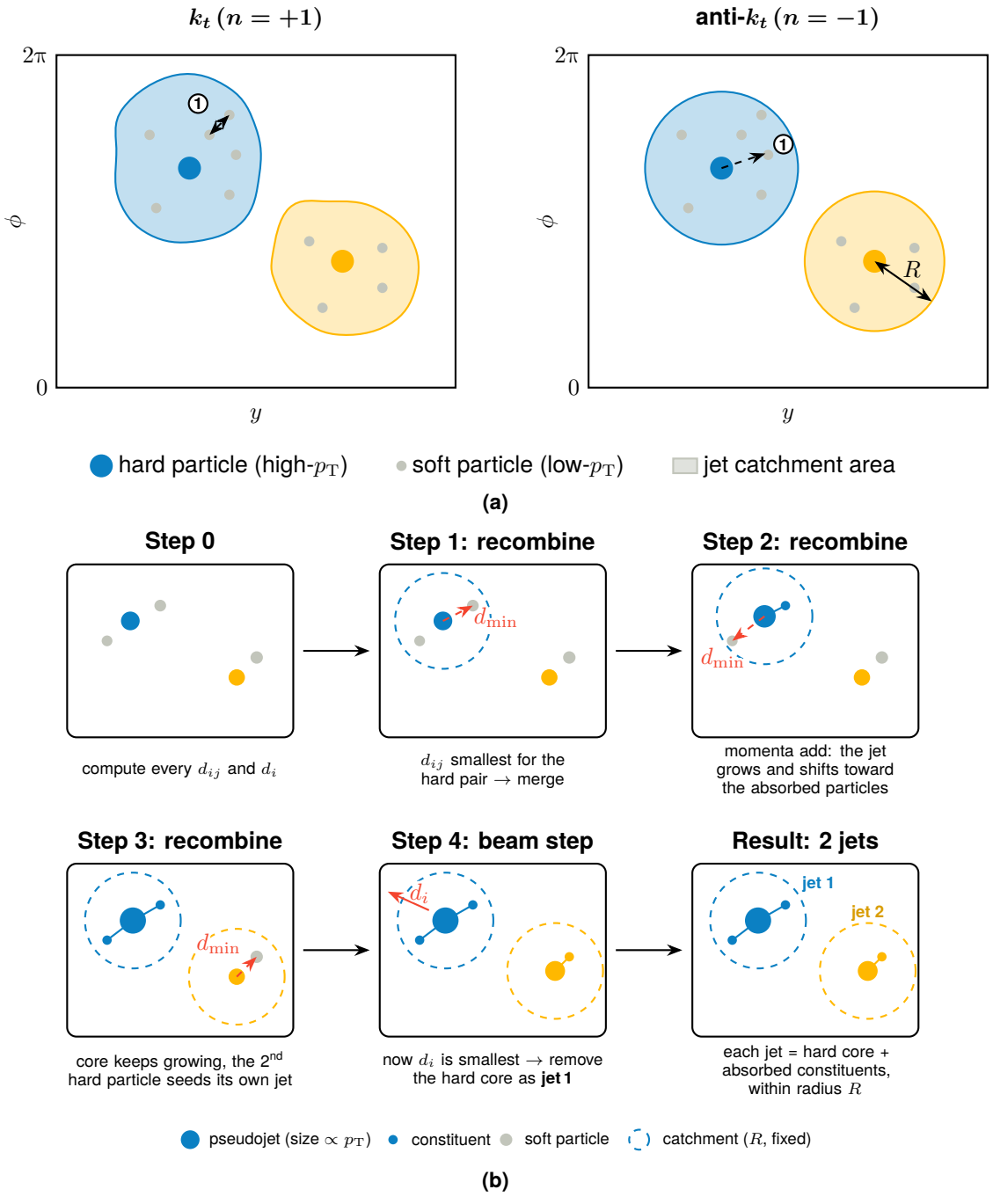


Figure 6.9: Scheme showing the difference between the k_t and anti-k_t algorithms on (a) and the clustering sequence of the anti-k_t algorithm on (b).

the calorimeters. The jet energy scale (JES) calibration restores the correct scale through the following sequence of corrections, applied to the jet p_{T}, energy and, when relevant, mass:

1. *Pile-up corrections*: the energy contributed by pile-up from the same (in-time) and nearby (out-of-time) bunch crossing is subtracted first. An area-based correction removes the product of the jet catchment area A and the median p_{T} density ρ of the event [299, 376], measured in the central region of the calorimeter. A residual correction, parametrised on simulation, then removes the remaining dependence of the jet p_{T} on the average number of interactions per bunch crossing $\langle\mu\rangle$ and on the number of reconstructed primary vertices N_{PV}.

2. *Absolute MC-based calibration*: the jet energy and direction are corrected to match those of the matched truth jet in simulated dijet events. This step accounts for the non-compensating calorimeter response, the energy lost in passive material, the energy deposited outside the cone of radius R used for the jet clustering (out-of-cone leakage) and the biases in the η reconstruction arising at the transitions between calorimeter technologies and granularities visible in Figure 3.6(b).

3. *Global sequential calibration (GSC)*: jets of identical (p_{T}, η) may still exhibit a different average response and a different resolution depending on their flavour (quark- versus gluon-initiated) and on the longitudinal and transverse development of their showers. A sequence of multiplicative corrections is applied as a function of a set of observables sensitive to these differences: the fraction of energy deposited in the first layers of the HAD and EM calorimeters, the number of tracks associated with the jet and their p_{T}-weighted width, and the number of MS segments behind the jet, the latter being sensitive to the *punch-through* of energetic showers that are not fully contained and leak behind the calorimeters. Each correction is constructed to remove the dependence of the response on the observable while leaving the mean energy unchanged, improving the resolution at fixed scale.

4. *In-situ corrections*: the residual data-to-simulation differences are measured by balancing a jet against a well-calibrated reference object. The η-intercalibration first equalises the response of forward jets ($0.8 < |\eta| < 4.5$) to that of central jets using dijet events: the absolute scale of central jets is then measured using Z+jet and γ+jet balance up to $p_{\mathrm{T}} \sim 1\,\mathrm{TeV}$, extended to higher p_{T} by balancing a high-p_{T} jet against a recoil system of well-calibrated lower-p_{T} jets in multijet events. The combined data-to-simulation response ratio is inverted and applied as a correction to data only.

The full chain is illustrated in Figure 6.10.

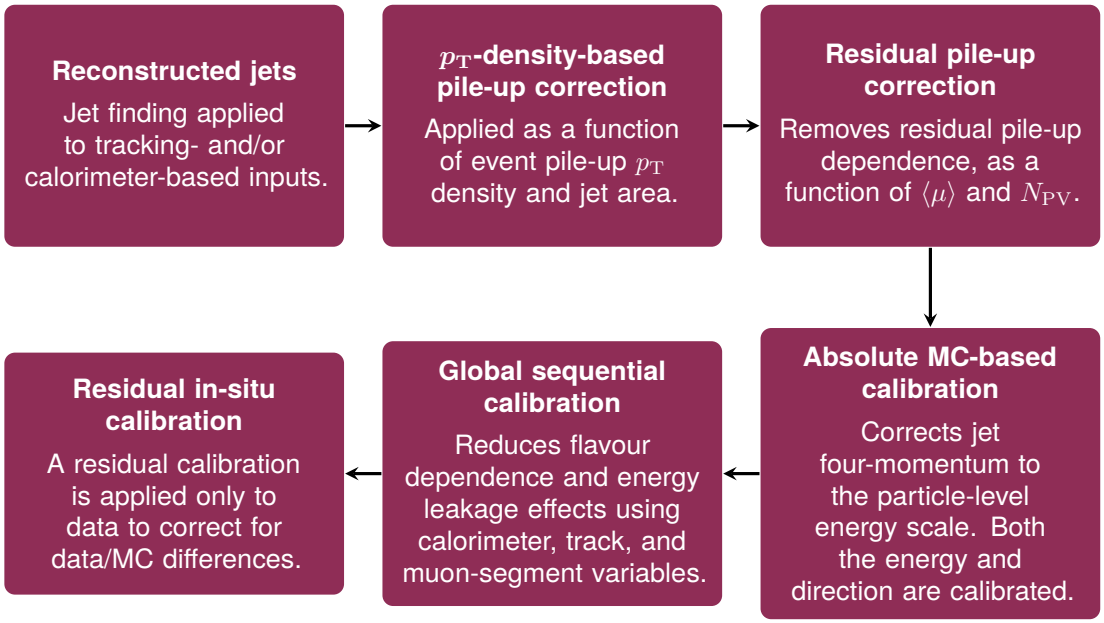


Figure 6.10: Workflow of the jet calibration chain.

This introduces about 125 sources of systematic uncertainty [373], propagated from the event selections, the MC modelling, and the limited size of the in-situ samples, which can be combined into reduced sets of NPs for use in physics analyses. Their combination amounts to about 1 % for central jets with $250 \lesssim p_{\mathrm{T}} \lesssim 2000\,\mathrm{GeV}$, growing to 5 % at $p_{\mathrm{T}} = 20\,\mathrm{GeV}$, where the pile-up subtraction and flavour response

dominate, and to $3.5\,\%$ above $2.5\,\mathrm{TeV}$, where the statistically limited multijet balance is supplemented by single-hadron response measurements. The uncertainty also grows in the forward region.

In parallel to the scale, the JER is measured and corrected. It is conveniently parametrised as:

$$\frac{\sigma(p_{\mathrm{T}})}{p_{\mathrm{T}}} = \frac{N}{p_{\mathrm{T}}} \oplus \frac{S}{\sqrt{p_{\mathrm{T}}}} \oplus C\,, \tag{6.16}$$

where the noise term N accounts for the electronics and pile-up contributions and dominates at low p_{T}, the stochastic term S reflects the Poissonian fluctuations of the shower sampling, and the constant term C captures the residual non-uniformities and dominates at high p_{T}, mirroring the structure of the calorimeter energy resolution discussed in Section 3.3. The S and C terms are measured in data from the p_{T} balance of dijet events, through the asymmetry of the two leading jets, while the noise term is determined separately with a random-cones method on zero-bias data, sensitive to the pile-up and electronic noise alone. As for the electron energy resolution discussed in Section 5.2 a smearing correction is applied to the simulation wherever its resolution is found to be better than the one measured in data.

6.2.1.2.5 Pile-up jet suppression Even after the constituent-level pile-up mitigation provided by the PFlow approach and the average subtraction of the JES chain, entire jets may originate from pile-up interactions, either from a hard pile-up jet or from the stochastic overlap of pile-up deposits. Within the tracker acceptance, these jets are suppressed using the jet vertex tagger (JVT) [377], a likelihood discriminant built from two complementary track-based variables. Both rely on the association of each track to a vertex, either the hard-scattering vertex PV_0 or one of the pile-up vertices PV_n ($n \geq 1$).

The first variable is the corrected jet-vertex fraction. Its uncorrected form, the jet-vertex fraction, is the fraction of the scalar p_{T} sum of the tracks associated to the jet that originates from the hard-scattering vertex defined as:

$$\mathrm{JVF} = \frac{\sum_k p_{\mathrm{T}}^{\mathrm{track}_k}(\mathrm{PV}_0)}{\sum_l p_{\mathrm{T}}^{\mathrm{track}_l}(\mathrm{PV}_0) + \sum_{n\geq 1}\sum_l p_{\mathrm{T}}^{\mathrm{track}_l}(\mathrm{PV}_n)}\,. \tag{6.17}$$

The drawback of Equation 6.17 is that the pile-up contribution in the denominator grows with the number of pile-up interactions, so that the hard-scatter efficiency at a fixed requirement degrades as the pile-up increases. This dependence is removed in the corrected variable, where the pile-up term is divided by its average value:

$$\mathrm{corrJVF} = \frac{\sum_k p_{\mathrm{T}}^{\mathrm{track}_k}(\mathrm{PV}_0)}{\sum_l p_{\mathrm{T}}^{\mathrm{track}_l}(\mathrm{PV}_0) + \dfrac{\sum_{n\geq 1}\sum_l p_{\mathrm{T}}^{\mathrm{track}_l}(\mathrm{PV}_n)}{k \cdot n_{\mathrm{track}}^{\mathrm{PU}}}}\,, \tag{6.18}$$

where $n_{\mathrm{track}}^{\mathrm{PU}}$ is the total number of pile-up tracks in the event and k is the measured slope of the average pile-up track p_{T} with $n_{\mathrm{track}}^{\mathrm{PU}}$. Normalising by $k \cdot n_{\mathrm{track}}^{\mathrm{PU}} \approx \langle p_{\mathrm{T}}^{\mathrm{PU}} \rangle$ makes the denominator, and hence the efficiency, stable against pile-up. A hard-scatter jet has $\mathrm{corrJVF} \to 1$ and a pile-up jet $\mathrm{corrJVF} \to 0$, while jets with no associated tracks are assigned $\mathrm{corrJVF} = -1$.

The second variable is:

$$R_{p_{\mathrm{T}}} = \frac{\sum_k p_{\mathrm{T}}^{\mathrm{track}_k}(\mathrm{PV}_0)}{p_{\mathrm{T}}^{\mathrm{jet}}}\,, \tag{6.19}$$

the scalar p_{T} sum of the tracks of the jet pointing to the hard-scattering vertex, divided by the fully calibrated jet p_{T}. It quantifies how much of the jet momentum is carried by hard-scatter tracks: for hard-scatter jets it is peaked at the average charged-momentum fraction, whereas for pile-up jets, which carry little or no hard-scatter track momentum, it is sharply peaked at zero. The two variables are combined into a single discriminant, JVT, through a clustering k-nearest-neighbour likelihood in the $(\mathrm{corrJVF}, R_{p_{\mathrm{T}}})$ plane, giving for each jet the local fraction of hard-scatter jets in its neighbourhood. A requirement on the JVT output is applied to jets with $p_{\mathrm{T}} < 60\,\mathrm{GeV}$ and $|\eta| < 2.4$. At the default WP, it retains $92\,\%$ of hard-scatter jets while reducing the residual pile-up jet rate to about $2\,\%$. Above $60\,\mathrm{GeV}$, where the pile-up contamination becomes negligible, no requirement is applied.

Beyond the tracker coverage, no vertex association is possible and a dedicated forward tagger, the forward jet vertex tagger (fJVT) [378], is used instead for jets within $2.5 < |\eta| < 4.5$. It exploits momentum conservation: a forward QCD pile-up jet originates from a single pile-up interaction and is, on average, balanced in the transverse plane by the rest of the activity from that same vertex, which falls in the central, track-covered region. For each pile-up vertex PV_i, the central jets are assigned to it through their $R_{p_{\mathrm{T}}}$,

and the corresponding per-vertex missing transverse momentum $\vec{p}_{\mathrm{T},i}^{\,\mathrm{miss}}$ is reconstructed from this central activity. A forward jet is then tested for its alignment with each of these values, and the maximum is taken as the fJVT discriminant:

$$\mathrm{fJVT}_i = \frac{\vec{p}_{\mathrm{T},i}^{\,\mathrm{miss}} \cdot \vec{p}_{\mathrm{T}}^{\,\mathrm{fj}}}{\left(p_{\mathrm{T}}^{\mathrm{fj}}\right)^2}\,, \qquad \mathrm{fJVT} = \max_i \mathrm{fJVT}_i\,, \tag{6.20}$$

where $\vec{p}_{\mathrm{T}}^{\,\mathrm{fj}}$ is the forward jet transverse momentum and the maximum runs over all pile-up vertices. The normalisation by $(p_{\mathrm{T}}^{\mathrm{fj}})^2$ reflects the fact that the missing momentum required to balance the jet scales with its p_{T}. A forward pile-up jet that exactly balances the activity of vertex i yields $\mathrm{fJVT}_i \simeq 1$. The forward jet is tagged as pile-up, and removed, when fJVT exceeds a chosen threshold: two WPs are defined, with hard-scatter efficiencies of $92\,\%$ and $85\,\%$, targeting pile-up jets typically in the $20\,\mathrm{GeV}$ to $50\,\mathrm{GeV}$ range. This suppression is of direct relevance for this thesis, as the VBS signature relies on forward tagging jets in a region where pile-up contamination is largest.

6.2.1.2.6 Flavour tagging The identification of jets originating from the hadronisation of b-quarks, referred to as b-jets, plays an important role as the top quark decays almost exclusively to Wb. The hadronisation of a b-quark produces a b-hadron carrying most of the jet momentum, with a lifetime of about $1.5\,\mathrm{ps}$ ($c\tau \approx 450\,\mu\mathrm{m}$): for a typical jet p_{T} of $50\,\mathrm{GeV}$, the b-hadron travels about $3\,\mathrm{mm}$ before decaying. As illustrated in Figure 6.11(a), its decay products therefore form a secondary vertex displaced from the primary one, and their tracks exhibit large transverse and longitudinal impact parameters d_0 and z_0, defined in Section 5.1.2, the significance of which (the impact parameter divided by its uncertainty) is the primary discriminating handle. Together with the possible presence of a tertiary vertex from the subsequent c-hadron decay, of a soft muon from the semileptonic decays of the heavy hadrons, and of a larger and harder constituent multiplicity, these features distinguish HF jets from LF and gluon jets.

These properties are exploited by multivariate taggers taking as inputs the parameters of the tracks associated with the jet. Earlier generations of ATLAS taggers proceeded in two stages, with dedicated low-level algorithms reconstructing the impact-parameter significances and the secondary-vertex properties, whose outputs were then combined by a high-level discriminant such as the DNN-based `DL1r` [379]. The current baseline, GN2 [380], instead adopts an end-to-end approach: it is a transformer[†] processing directly the unordered set of tracks associated with the jet, each represented by its kinematics and impact parameters, with the relations between tracks learned through an attention mechanism. The NN is trained on a multitask objective, simultaneously classifying the jet flavour, grouping the tracks into their vertices of origin and predicting the underlying physics process of each track. The two auxiliary tasks guide the NN towards a physically meaningful internal representation and improve the flavour separation itself. Its output probabilities p_{b}, p_{c}, p_τ, and p_{light} are combined into a single discriminant:

$$D_{\mathrm{b}} = \ln \frac{p_{\mathrm{b}}}{f_{\mathrm{c}}\, p_{\mathrm{c}} + f_\tau\, p_\tau + (1 - f_{\mathrm{c}} - f_\tau)\, p_{\mathrm{light}}}\,, \tag{6.21}$$

with $f_{\mathrm{c}} = 0.2$ and $f_\tau = 0.05$ tuneable fractions setting the relative weight of the charm and tau backgrounds (the previous-generation `DL1r`, without a τ output, uses $f_{\mathrm{c}} = 0.018$). Requirements on D_{b} define five WPs of fixed b-jet efficiency, from $65\,\%$ to $90\,\%$, in the same spirit as the identification WPs of Section 5.3.1. GN2 substantially improves the rejection of light- and charm-jets at fixed b-jet efficiency with respect to the previous algorithm.

As for the electron and photon efficiencies, the tagging efficiencies measured in simulation do not perfectly match those in data, and SFs are derived to correct for the difference. The b-jet efficiency is measured in a high-purity sample of $\mathrm{t\bar{t}}$ dilepton events, while the mis-tagging efficiencies of c- and light-jets are measured in dedicated samples, the latter exploiting jets with negative impact-parameter significances or secondary vertices reconstructed on the wrong side of the primary vertex, which arise only from resolution effects and therefore probe the light-flavour mis-tag rate directly. The resulting SFs, close to unity, and their uncertainties are applied to the simulation following the same logic as in Section 5.2.

6.2.1.2.7 From Run-2 to Run-3 The jet reconstruction and calibration described above have evolved between Run-2 and Run-3. The common thread is the generalisation of machine learning techniques across the whole workflow. The flavour-tagging baseline moved from the DNN-based `DL1r` to the transformer-based GN2 discriminant presented above. In the same spirit, the JVT likelihood is being superseded by an NN-JVT feeding the same $\mathrm{corrJVF}$ and $R_{p_{\mathrm{T}}}$ variables, together with p_{T} and η, into a small network. The

[†]The GN family is named after GN1, which was built as a GNN. GN2 keeps the name despite being a single transformer model.

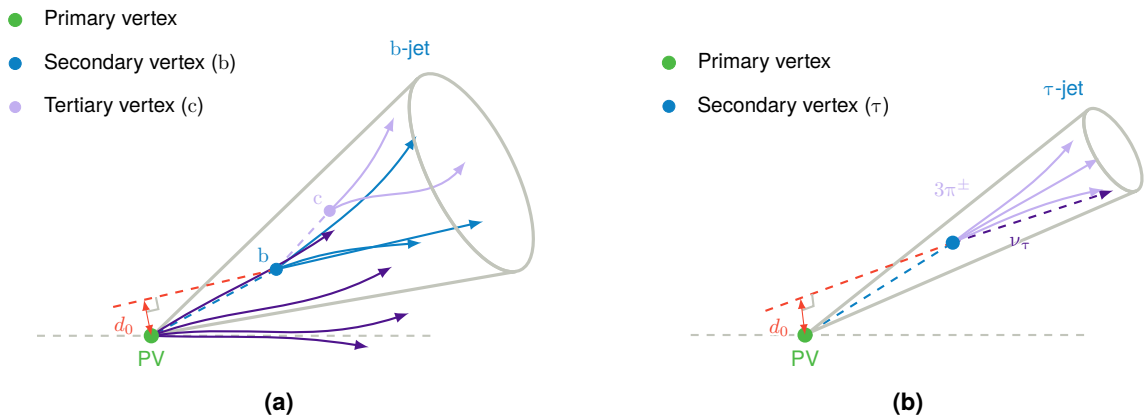


Figure 6.11: Scheme showing the topology of a b-jet on (a) and of a τ-jet on (b).

global sequential calibration step of the energy calibration is being replaced by a single transformer-based regression that exploits the correlations between the calibration observables, yielding a relative JER improvement of 15 % to 25 % and a more universal calibration of b-jets [381]. The full JES and JER calibration has been re-derived on the Run-3 dataset, with the simulation found to describe the data slightly better than in Run-2 and with comparable systematic uncertainties [382]. Together with the timing information that the HGTD will bring in Phase-II, these developments are essential to preserve the jet performance in the high-pile-up regime.

6.2.1.3 Missing transverse momentum

The remaining particles among the predicted ones from the SM that were not discussed so far are the neutrinos, which are neutral and only interact via the weak force. They therefore escape the detector without leaving any trace, and their presence can only be inferred from the momentum imbalance of the event. The missing transverse momentum $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ is defined as the negative vectorial sum of the transverse momenta of all reconstructed objects in the event, including jets, electrons, muons, and photons, together with a soft term accounting for low-p_{T} activity not associated with any of these objects [383]. Its magnitude is denoted $E_{\mathrm{T}}^{\mathrm{miss}}$, and the vector is computed as:

$$\vec{E}_{\mathrm{T}}^{\mathrm{miss}} = \underbrace{\underbrace{-\sum_{\text{selected e}} \vec{p}_{\mathrm{T}}^{\,\mathrm{e}}}_{\vec{E}_{\mathrm{T}}^{\mathrm{miss,e}}} \underbrace{-\sum_{\text{accepted }\gamma} \vec{p}_{\mathrm{T}}^{\,\gamma}}_{\vec{E}_{\mathrm{T}}^{\mathrm{miss},\gamma}} \underbrace{-\sum_{\text{selected }\mu} \vec{p}_{\mathrm{T}}^{\,\mu}}_{\vec{E}_{\mathrm{T}}^{\mathrm{miss},\mu}} \underbrace{-\sum_{\text{accepted }\tau} \vec{p}_{\mathrm{T}}^{\,\tau}}_{\vec{E}_{\mathrm{T}}^{\mathrm{miss},\tau}} \underbrace{-\sum_{\text{accepted j}} \vec{p}_{\mathrm{T}}^{\,\mathrm{j}}}_{\vec{E}_{\mathrm{T}}^{\mathrm{miss,j}}}}_{\text{hard term}} \underbrace{\underbrace{-\sum_{\text{unused tracks}} \vec{p}_{\mathrm{T}}^{\,\mathrm{track}}}_{\vec{E}_{\mathrm{T}}^{\mathrm{miss,track}}}}_{\text{soft term}} . \tag{6.22}$$

Because the same detector signal may be reconstructed as more than one object, a signal ambiguity resolution is applied when building the sum: each track and calorimeter signal is assigned to a single object following a fixed priority (e, γ, τ, μ, j), so that no momentum is counted twice. This is the same logic as the overlap removal of Section 6.2.1.5, here applied specifically to the $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction. The reconstruction of $E_{\mathrm{T}}^{\mathrm{miss}}$ is therefore sensitive to all aspects of the event reconstruction.

The hard term is built from the calibrated objects, and its performance is therefore directly related to the performance of each of them. The soft term is built from tracks associated with the hard-scattering vertex but not associated with any of the reconstructed objects, and it is therefore robust against pile-up. Two definitions of the soft term exist: the cluster soft term uses all calorimeter clusters that are not in any objects in the hard term and can thus capture neutral particles, whereas the track soft term uses only tracks associated with the hard-scattering vertex, and is therefore more robust against pile-up. The choice of the latter is the default in ATLAS analyses.

The $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution is dominated by the soft term at low $E_{\mathrm{T}}^{\mathrm{miss}}$, while at high $E_{\mathrm{T}}^{\mathrm{miss}}$ it is dominated by the hard term, and in particular by the jets. Indeed, as they are the most common objects and have the worst resolution, the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution is strongly correlated with the JER. Among jets, the forward

ones are particularly important as they are not covered by the tracker and are therefore more sensitive to pile-up. Different WPs apply different p_{T} cuts on forward jets:

- `Loose`: $p_{\mathrm{T}} > 20\,\mathrm{GeV}$
- `Tight`: $p_{\mathrm{T}} > 30\,\mathrm{GeV}$
- `Tighter`: $p_{\mathrm{T}} > 35\,\mathrm{GeV}$
- `Tenacious`: $p_{\mathrm{T}} > 40\,\mathrm{GeV}$.

The uncertainties on the $E_{\mathrm{T}}^{\mathrm{miss}}$ are propagated from those on the upstream hard-term objects, exactly as for the jets, with the soft term treated separately. The soft term is decomposed, with respect to the direction of the hard-term sum, into a longitudinal component $p_{\parallel}^{\mathrm{soft}}$, which carries both a scale and a resolution, and a transverse component $p_{\perp}^{\mathrm{soft}}$, which has zero mean by symmetry and therefore only a resolution. These are parametrised as a function of the number of primary vertices and the scalar p_{T} sum of the hard-term objects, and propagated to the $E_{\mathrm{T}}^{\mathrm{miss}}$ by varying the soft term within its uncertainty. A scheme showing the various contributions to the $E_{\mathrm{T}}^{\mathrm{miss}}$ is shown in Figure 6.12.

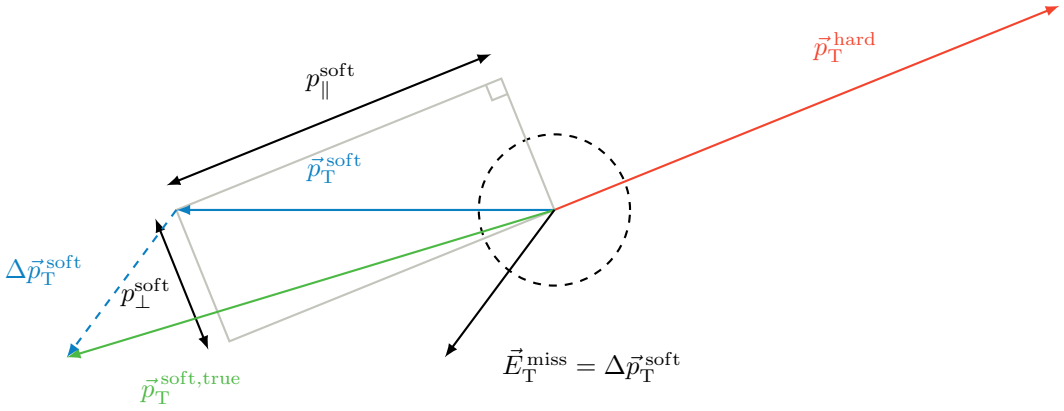


Figure 6.12: Scheme showing the various contributions to the missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}}$.

To assess how well the $E_{\mathrm{T}}^{\mathrm{miss}}$ is reconstructed and whether it originates from real sources, the $E_{\mathrm{T}}^{\mathrm{miss}}$ significance is defined as:

$$Z = \frac{E_{\mathrm{T}}^{\mathrm{miss}}}{\sqrt{\sigma_{\mathrm{L}}^2\,(1-\rho_{\mathrm{LT}}^2)}}\,, \tag{6.23}$$

where σ_{L} is the total longitudinal resolution from objects and soft term, and ρ_{LT} is the correlation between the longitudinal and transverse components of the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution. The $E_{\mathrm{T}}^{\mathrm{miss}}$ significance is therefore a measure of how likely the observed $E_{\mathrm{T}}^{\mathrm{miss}}$ is to be due to real missing transverse momentum rather than fluctuations in the reconstructed objects.

6.2.1.4 Tau leptons

The τ lepton is the heaviest of the three charged leptons, with a mass of $1.777\,\mathrm{GeV}$ that makes it the only lepton heavy enough to decay into hadrons. With a proper decay length $c\tau \approx 87\,\mu\mathrm{m}$, it decays before reaching the active layers of the detector and is therefore reconstructed exclusively from its decay products. In about $35\,\%$ of the cases the decay is leptonic, $\tau \to \ell\nu_\ell\nu_\tau$ with ℓ an electron or a muon, and the resulting light lepton is reconstructed as a prompt electron or muon, essentially indistinguishable from a directly produced one. The remaining $65\,\%$ are hadronic decays, $\tau \to \mathrm{hadrons} + \nu_\tau$, which form the experimentally distinctive signature and the object reconstructed as a τ candidate. A scheme showing the topology of a τ-jet is shown in Figure 6.11(b).

The visible hadronic decay products, denoted $\tau_{\text{had-vis}}$, consist almost exclusively of charged and neutral pions, accompanied by a single ν_τ that escapes detection and contributes to the missing transverse momentum. The decay yields either one charged pion (about $72\,\%$ of hadronic decays) or three charged pions (about $22\,\%$), in both cases frequently accompanied by one or more $\pi^0 \to \gamma\gamma$ that contribute EM energy. The experimental signature is thus a narrow, low-multiplicity jet with one or three associated tracks and a well-collimated calorimeter deposit, in contrast to the broader, higher-multiplicity quark- and gluon-initiated jets.

The reconstruction [384] starts from anti-k_t jets of radius $R = 0.4$ built from topo-clusters, which seed the $\tau_{\text{had-vis}}$ candidates with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ and $|\eta| < 2.5$, excluding the calorimeter transition region. Tracks

are associated in a core cone $\Delta R < 0.2$ around the seed axis, the surrounding annulus $0.2 < \Delta R < 0.4$ serving to characterise the isolation, and are classified according to whether they originate from the τ decay, from PCs, from pile-up, or from nearby jet activity. A dedicated calibration, the tau energy scale, then corrects the candidate four-momentum back to the particle level, accounting for the calorimeter response to the hadronic decay products, the out-of-cone energy, and the pile-up contribution.

The dominant background to $\tau_{\text{had-vis}}$ arises from quark- and gluon-initiated jets faking the signature. It is suppressed by a dedicated identification discriminant [385] based on an RNN that processes the sequences of associated tracks and calorimeter clusters, together with high-level shower-shape and lifetime variables, separately optimised for 1- and 3-pion candidates. Several WPs of increasing purity are defined as fixed-efficiency requirements on its output. As electrons can also mimic a 1-pion $\tau_{\text{had-vis}}$, a separate NN electron veto is applied. These multivariate algorithms, already deployed in Run-2 and further improved for Run-3, provide the jet and electron rejection required to use hadronic τ decays in physics analyses.

6.2.1.5 Overlap removal

The reconstruction algorithms described in the previous subsections operate largely independently of one another, so that a single energy deposit or track can give rise to several reconstructed objects. An electron, for instance, is almost always also reconstructed as a jet, since its calorimeter shower is clustered by the anti-k_t algorithm, and a muon undergoing hard bremsstrahlung may seed a nearby jet. Using all of these objects simultaneously would amount to counting the same physical particle several times. An *overlap removal* procedure is therefore applied after object selection and before the event-level quantities are built, resolving each ambiguity in favour of the object most likely to be correctly identified.

The procedure consists of a sequence of pairwise comparisons between objects that either share a track or lie close to one another in the η–ϕ plane, the angular distance ΔR being the main criterion. These comparisons encode physical priors on which interpretation is the more probable. Cleanly identified leptons and photons generally take precedence over jets: a jet overlapping a lepton within a narrow cone is interpreted as the calorimeter deposit of that lepton and discarded, whereas a lepton found inside a wider cone around a surviving jet is more likely a non-prompt lepton produced in the decay of an HF hadron and is itself removed. Ambiguities between muons and jets are resolved using the track content of the jet, which separates an isolated muon from the bremsstrahlung or punch-through of an energetic jet, and hadronically decaying τ candidates are likewise cleaned against nearby leptons and jets. The cone sizes involved are typically in the range $\Delta R \sim 0.2$–0.4.

The precise ordering of these steps, the cone sizes, and the auxiliary criteria, such as the track-multiplicity requirement, are not universal: they are optimised by each analysis according to its final state and the relative importance of the objects involved. The underlying principle is, however, common, namely that each detector signature should contribute to the event reconstruction, and in particular to the missing transverse momentum of Section 6.2.1.3, exactly once.

6.2.2 Event reconstruction

Each final-state particle is reconstructed as a detector object and is fully described by its energy E, transverse momentum p_{T}, pseudorapidity η, and azimuthal angle ϕ, written as $(E, p_{\mathrm{T}}, \eta, \phi)$. The neutrino is not measured directly. Instead, the MET $E_{\mathrm{T}}^{\mathrm{miss}}$ is reconstructed as the negative vectorial sum of the transverse momenta of all reconstructed objects in the event. From this set of quantities the kinematic structure of the process can be deduced.

The Z is reconstructed from a same-flavour opposite-charge lepton pair with an invariant mass within $\pm 10\,\mathrm{GeV}$ of the pole mass m_{Z}. The W is reconstructed from the remaining isolated lepton combined with the missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}}$ left by the undetected neutrino.

From the W-lepton and the missing transverse momentum, the transverse mass of the W is obtained as:

$$m_{\mathrm{T}}^{\mathrm{W}} = \sqrt{2\, p_{\mathrm{T}}^{\ell}\, E_{\mathrm{T}}^{\mathrm{miss}}\, (1 - \cos\Delta\phi(\ell_{\mathrm{W}}, \nu_{\ell}))}\,, \tag{6.24}$$

where $\Delta\phi(\ell_{\mathrm{W}}, \nu_{\ell})$ is the azimuthal separation between the lepton and the missing transverse momentum. It is bounded from above by m_{W} for on-shell production and exceeds it only for off-shell W. Combining the reconstructed Z and W, the transverse mass of the full WZ system is defined as:

$$m_{\mathrm{T}}^{\mathrm{WZ}} = \sqrt{\left(\sum_{\ell=1}^{3} p_{\mathrm{T}}^{\ell} + E_{\mathrm{T}}^{\mathrm{miss}}\right)^{2} - \left[\left(\sum_{\ell=1}^{3} p_{x}^{\ell} + E_{x}^{\mathrm{miss}}\right)^{2} + \left(\sum_{\ell=1}^{3} p_{y}^{\ell} + E_{y}^{\mathrm{miss}}\right)^{2}\right]}\,, \tag{6.25}$$

where the sum runs over the three charged leptons and E_x^{miss}, E_y^{miss} are the components of the missing transverse momentum. It tracks the diboson invariant-mass scale and serves as a proxy for it, since the longitudinal neutrino momentum p_z^ν cannot be fully reconstructed, as discussed next.

6.2.2.1 Longitudinal neutrino momentum

A first approach constrains the lepton–neutrino system to the W pole mass:

$$p_{\mathrm{W}}^2 = m_{\mathrm{W}}^2 = \left(p_\ell^\mu + p_\nu^\mu\right)^2 , \tag{6.26}$$

where p_{W}^μ, p_ℓ^μ, and p_ν^μ are the four-momenta of the W, the lepton, and the neutrino from the W decay, respectively. Only p_z^ν is unknown. Neglecting the lepton mass and writing out the energy and momentum components gives:

$$\frac{m_{\mathrm{W}}^2}{2} = E^\ell E^\nu - \vec{p}_{\mathrm{T}}^{\,\ell} \cdot \vec{p}_{\mathrm{T}}^{\,\nu} - p_z^\ell p_z^\nu , \tag{6.27}$$

with $\vec{p}_{\mathrm{T}}^{\,\ell} \cdot \vec{p}_{\mathrm{T}}^{\,\nu} = p_{\mathrm{T}}^\ell p_{\mathrm{T}}^\nu \cos\Delta\phi(\ell_{\mathrm{W}}, \nu_\ell)$. This can be rearranged and squared:

$$\left(\xi + p_z^\ell p_z^\nu\right)^2 = \left(E^\ell E^\nu\right)^2 = {E^\ell}^2 \left({p_{\mathrm{T}}^\nu}^2 + {p_z^\nu}^2\right) , \tag{6.28}$$

with $\xi = \frac{1}{2} m_{\mathrm{W}}^2 + \vec{p}_{\mathrm{T}}^{\,\ell} \cdot \vec{p}_{\mathrm{T}}^{\,\nu}$, leading to:

$$p_z^\nu = \frac{p_z^\ell\, \xi \pm \sqrt{\Delta}}{{p_{\mathrm{T}}^\ell}^2} , \qquad \Delta = {p_z^\ell}^2 \xi^2 - {p_{\mathrm{T}}^\ell}^2 \left({E^\ell}^2 {p_{\mathrm{T}}^\nu}^2 - \xi^2\right) . \tag{6.29}$$

A positive discriminant Δ yields two real solutions, while a negative Δ gives no real solution and corresponds to an off-shell W with a transverse mass above its pole mass. A common prescription is to take the solution of smallest absolute value when $\Delta > 0$ and the real part of the complex solution when $\Delta < 0$. This *analytical* method, used in the first polarisation measurements in inclusive WZ production [10], has poor resolution because

- p_{T} and $E_{\mathrm{T}}^{\text{miss}}$ are subject to misreconstruction
- particles other than the hard-scattering neutrino contribute to $E_{\mathrm{T}}^{\text{miss}}$, weakening the approximation $p_{\mathrm{T}}^\nu \simeq E_{\mathrm{T}}^{\text{miss}}$
- off-shell W production and the choice of solution add further misreconstruction.

A second approach, developed for the joint polarisation measurement of WZ production [11], and detailed in [345], uses a DNN to estimate p_z^ν from the p_{T} and p_z of the W-lepton, the components of $E_{\mathrm{T}}^{\text{miss}}$ along and perpendicular to the lepton direction:

$$\left(E_{\mathrm{T}}^{\text{miss}}\right)^{\parallel} = E_{\mathrm{T}}^{\text{miss}} \cos\Delta\phi(\ell_{\mathrm{W}}, \nu_\ell) , \tag{6.30}$$

$$\left(E_{\mathrm{T}}^{\text{miss}}\right)^{\perp} = E_{\mathrm{T}}^{\text{miss}} \sin\Delta\phi(\ell_{\mathrm{W}}, \nu_\ell) , \tag{6.31}$$

and the analytical p_z^ν. These inputs encode the ϕ-rotational symmetry of the detector. The loss function sums the MSE on the neutrino momentum components and on m_{W}:

$$\mathcal{L} = \mathrm{MSE}\left(p_z^\nu, p_z^{\nu,\text{true}}\right) + \mathrm{MSE}\left(m_{\mathrm{W}}, m_{\mathrm{W}}^{\text{true}}\right) , \tag{6.32}$$

which simultaneously improves the p_z^ν resolution and enforces the on-shell W condition. It recovers a reasonable estimate for events where the analytical method fails and reduces the standard deviation of the p_z^ν error by $10\,\%$, mainly through a better resolution at high momentum. The m_{W} resolution improves from $41\,\%$ to $31\,\%$. This method is the baseline for the determination of p_z^ν in the rest of this thesis, and the resulting improvement in the W four-momentum, and hence in the full WZ system, is important for the polarisation-sensitive variables introduced below.

6.2.3 Discriminating variables

The key variables fall into three sets:

Bosonic variables: characterise the bosons and the process that produced them

Leptonic variables: describe the leptons from the boson decays

Jet variables: provide additional information through the tagging jets and any extra jets.

The QCD-induced background, lacking the colour-singlet structure of the signal, populates these variables differently: its jets tend to be more central and less separated in rapidity, its rapidity gap is more actively radiated, and its bosons and leptons are produced less centrally. Each variable below exploits one of these differences, and its signal-versus-background behaviour is quoted alongside the definition.

6.2.3.1 Jet variables

The forward nature of the tagging jets follows from the propagators of the radiated bosons:

$$\text{W/Z} \sim\sim\sim\sim\sim = \frac{-\mathrm{i}}{q_V^2 - m_V^2}\left(g^{\mu\nu} - \frac{q_V^\mu q_V^\nu}{m_V^2}\right), \tag{6.33}$$

where q_V is the four-momentum of the radiated boson and m_V its mass. The factor $\frac{1}{q_V^2 - m_V^2}$ is largest at small momentum transfer $|q_V^2|$, that is for nearly collinear emission, so the cross-section is enhanced when the quark is barely deflected. The mass scale m_V in the denominator sets the typical transverse recoil, $p_\mathrm{T} \approx m_V$, which is moderate compared with the quark's longitudinal momentum. The radiation amplitude therefore peaks at small scattering angle, the quark staying close to its original direction. The energy dependence of the polarisation fractions and the RAZ effect have been studied in WZ production [347, 386].

The two tagging quarks consequently end up in opposite forward–backward hemispheres, $\eta^{\mathrm{j}_1} \cdot \eta^{\mathrm{j}_2} < 0$, with high energy but moderate p_T, large absolute rapidities $|y^\mathrm{j}|$, and a large invariant mass between them:

$$m_{\mathrm{jj}}^2 = (p_{\mathrm{j}_1} + p_{\mathrm{j}_2})^2 \approx 2\, p_\mathrm{T}^{\mathrm{j}_1} p_\mathrm{T}^{\mathrm{j}_2} \left(\cosh \Delta y_{\mathrm{jj}} - \cos \Delta\phi_{\mathrm{jj}}\right), \tag{6.34}$$

where Δy_{jj} and $\Delta\phi_{\mathrm{jj}}$ are the rapidity and azimuthal-angle differences between the jets and the approximation holds in the massless limit. The large mass is driven by the large rapidity separation, giving $m_{\mathrm{jj}} \gg m_V$ for the signal. Background jets, produced with smaller rapidity separation and/or higher p_T, give smaller m_{jj}. This variable is represented in Figure 6.13(a).

The p_T balance of the two jets is quantified by:

$$A_{p_\mathrm{T}}^{\mathrm{jj}} = \frac{p_\mathrm{T}^{\mathrm{j}_1} - p_\mathrm{T}^{\mathrm{j}_2}}{p_\mathrm{T}^{\mathrm{j}_1} + p_\mathrm{T}^{\mathrm{j}_2}}, \tag{6.35}$$

which is small for the signal, where the two radiating quarks have similar kinematics, and larger for backgrounds with a more imbalanced pair. The complementary average $\langle p_\mathrm{T}^{\mathrm{jj}} \rangle = \frac{1}{2}(p_\mathrm{T}^{\mathrm{j}_1} + p_\mathrm{T}^{\mathrm{j}_2})$ stays moderate for the signal, reflecting the moderate recoil $p_\mathrm{T} \approx m_V$, and is larger when the jets come from a harder production mechanism.

The centrality of the jets is captured by:

$$\zeta_{\mathrm{jets}} = \frac{p_\mathrm{T}^{\mathrm{j}_1} + p_\mathrm{T}^{\mathrm{j}_2}}{E^{\mathrm{j}_1} + E^{\mathrm{j}_2}}, \tag{6.36}$$

which, in the massless limit where $\frac{p_\mathrm{T}}{E} = \frac{1}{\cosh\eta}$, is the energy-weighted mean of the per-jet $\frac{p_\mathrm{T}}{E}$. Forward jets have large $|\eta|$ and hence small ζ_{jets}. A more central pair gives a larger value.

Finally, because the incoming quarks exchange no colour, the event is a colour singlet and no QCD radiation bridges the two hemispheres. The rapidity gap between the tagging jets is therefore, on average, sparsely populated with additional jets N_{jets}. Backgrounds populate the gap more, while bremsstrahlung tends to follow the colour flow along the beam, dressing the tagging jets rather than filling the central region.

6.2.3.2 Bosonic variables

In contrast to the forward jets, the WZ system is produced centrally, between the two tagging jets in rapidity. The combination of central diboson production and forward tagging jets is the clean signature of the VBS topology. This centrality is quantified by the Zeppenfeld variable of an object X:

$$z_X = \frac{y^X - \frac{1}{2}\left(y^{\mathrm{j}_1} + y^{\mathrm{j}_2}\right)}{\Delta y_{\mathrm{jj}}}, \tag{6.37}$$

the rapidity position of X within the gap, normalised to the full width of the gap, so that the two tagging jets sit at $z = \pm 0.5$. Evaluated for the bosons, and, below, for the leptons, it peaks sharply at zero for the signal and spreads towards ± 0.5 and beyond when the object is produced more forward, as in background events.

The rapidity separation between the reconstructed Z and the W-lepton:

$$A_y^{\mathrm{WZ}} = \left| y^{\mathrm{Z}} - y^{\ell_{\mathrm{W}}} \right|, \tag{6.38}$$

is small for the centrally produced signal and larger when one boson is more forward. The relative transverse momentum of the two bosons:

$$r_{21} = \frac{\min\left(p_{\mathrm{T}}^{\mathrm{W}}, p_{\mathrm{T}}^{\mathrm{Z}}\right)}{\max\left(p_{\mathrm{T}}^{\mathrm{W}}, p_{\mathrm{T}}^{\mathrm{Z}}\right)}, \tag{6.39}$$

is close to one for the signal, where the bosons recoil against the balanced tagging jets with similar p_{T}, and smaller when one boson is produced harder. The overall transverse balance of the hard final state is measured by:

$$R_{p_{\mathrm{T}}}^{\mathrm{hard}} = \frac{\left| \vec{p}_{\mathrm{T}}^{\,\ell_{\mathrm{W}}} + \vec{p}_{\mathrm{T}}^{\,\nu_\ell} + \vec{p}_{\mathrm{T}}^{\,\ell_{\mathrm{Z}_1}} + \vec{p}_{\mathrm{T}}^{\,\ell_{\mathrm{Z}_2}} + \vec{p}_{\mathrm{T}}^{\,\mathrm{j}_1} + \vec{p}_{\mathrm{T}}^{\,\mathrm{j}_2} \right|}{p_{\mathrm{T}}^{\ell_{\mathrm{W}}} + E_{\mathrm{T}}^{\mathrm{miss}} + p_{\mathrm{T}}^{\ell_{\mathrm{Z}_1}} + p_{\mathrm{T}}^{\ell_{\mathrm{Z}_2}} + p_{\mathrm{T}}^{\mathrm{j}_1} + p_{\mathrm{T}}^{\mathrm{j}_2}}, \tag{6.40}$$

the magnitude of the vectorial p_{T} sum of the six hard objects over their scalar p_{T} sum. A clean, balanced signal event gives a small value, whereas additional QCD radiation in background events unbalances the system and increases it. Lastly, the distance in the η–ϕ plane between the Z and the leading tagging jet, $\Delta R(\mathrm{j}_1, \mathrm{Z}) = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, is large for the signal, where the central Z sits far from the forward jets, and smaller when the Z is produced more forward.

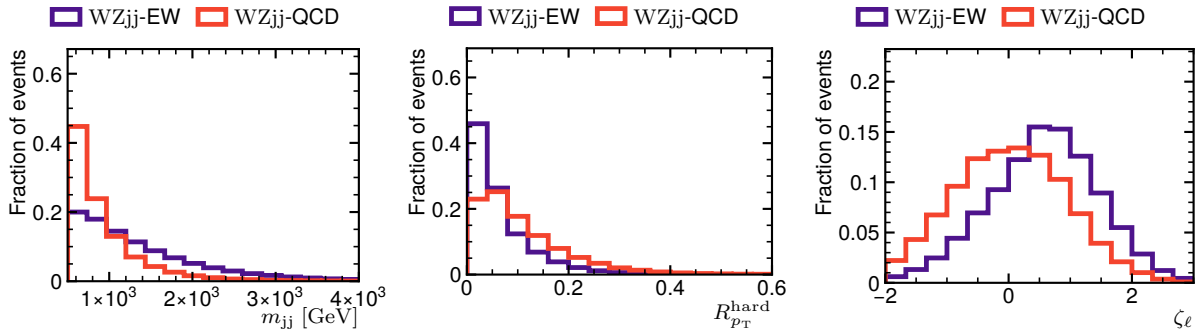
Of the bosonic variables, $R_{p_{\mathrm{T}}}^{\mathrm{hard}}$ given in Equation 6.40 is the one that separates WZjj-EW from WZjj-QCD best, and it is the one represented in Figure 6.13(b): the signal piles up at small values, where the six hard objects balance, while the background carries the longer tail left by its extra QCD radiation.

6.2.3.3 Leptonic variables

The leptons inherit the central production of their parent bosons. Their centrality relative to the tagging jets is quantified by:

$$\zeta_\ell = \min\left[\left(\min_i \eta^{\ell_i} - \min\left(\eta^{\mathrm{j}_1}, \eta^{\mathrm{j}_2}\right) \right), \left(\max\left(\eta^{\mathrm{j}_1}, \eta^{\mathrm{j}_2}\right) - \max_i \eta^{\ell_i} \right) \right], \tag{6.41}$$

the smaller of the two pseudorapidity gaps separating the leptons from the jets. It is large and positive for the signal, where the three leptons lie well inside the forward dijet interval, and becomes negative if any lepton is produced outside it. This variable is represented in Figure 6.13(c). The Zeppenfeld variable defined above is also evaluated for each lepton and for the three-lepton system, with the same expectation of a sharp peak at zero for the signal.



(a) m_{jj} **(b)** $R_{p_{\mathrm{T}}}^{\mathrm{hard}}$ **(c)** ζ_ℓ

Figure 6.13: Distributions of one representative variable from each of the three sets of discriminating variables: m_{jj} for the jets, $R_{p_{\mathrm{T}}}^{\mathrm{hard}}$ for the bosons, and ζ_ℓ for the leptons. The signal is shown in purple and the QCD-induced background in red. The complete list of discriminating variables is given in Appendix I. Run-3 SR events at reconstructed level are shown.

6.2.4 Reference frames and polarisation observables

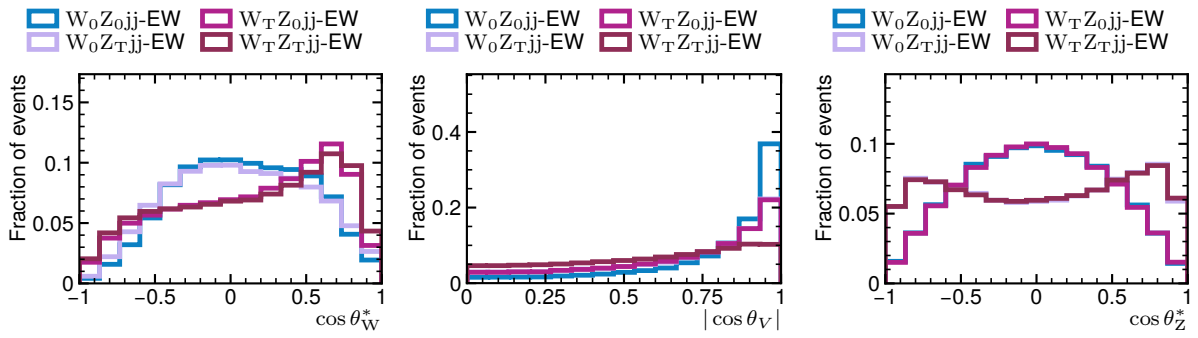
As discussed in Section 1.11, the polarisation is not Lorentz invariant and therefore depends on the chosen reference frame. Since the bosons are massive, they are observed only through their decay products, which retain information on the polarisation of their parent, expressed through angular variables. The polarisation axis must therefore be defined. Of the many possible choices, only those relevant to this thesis are discussed.

The simplest is the laboratory frame, in which the polarisation is decomposed with respect to the beam axis. This is theoretically convenient but mixes the longitudinal and transverse contributions in a way that depends on the boson boost, so it is rarely used for such measurements. The helicity frame is instead defined as the rest frame of the boson, with the polarisation axis along the boson's direction of motion. In its simplest form this direction is taken in the laboratory frame. Here, because the two bosons are analysed as a system, the direction is instead evaluated in the diboson CM frame. This defines the *modified-helicity* frame, which gives the cleanest separation of longitudinal and transverse states and is the standard choice for such measurements.

From the detector objects, the helicity frame is reconstructed in two boosts. First, the four leptons are boosted from the laboratory into the diboson CM frame using $\vec{\beta}_{\mathrm{WZ}} = \frac{\vec{p}_{\mathrm{WZ}}}{E_{\mathrm{WZ}}}$. In this frame the W and Z recoil back-to-back, and the production angle θ_V is the polar angle of the boson momentum with respect to the beam axis. Second, the decay products of each boson are boosted from the diboson frame into that boson's own rest frame using $\vec{\beta}_V = \frac{\vec{p}_V}{E_V}$. The quantisation axis of each boson is its flight direction in the diboson CM frame, carried into its rest frame.

In each boson rest frame, the polar angles θ^*_{Z} and θ^*_{W} are the angles between this quantisation axis and the decay leptons of the Z and W, respectively, and carry different angular distributions for longitudinal and transverse polarisations. In the same frames the azimuthal angles $\Delta\phi(\ell_{\mathrm{Z}_1}, \ell_{\mathrm{Z}_2})$, between the two Z-leptons, and $\Delta\phi(\ell_{\mathrm{W}}, \nu_\ell)$, between the W-lepton and the missing transverse momentum, are likewise sensitive to the polarisation states. A visualisation of the cosine of the polar angles is shown in Figure 6.14.

The polarisation states are therefore quantised in the WZ CM frame from now on. This is also known as the *modified-helicity coordinate system* [10]. The longitudinal mode is obtained by selecting the single $\lambda = 0$ term, while the transverse mode is the coherent sum of the LH and RH modes, including their interference.



(a) $\cos\theta^*_{\mathrm{W}}$ **(b)** $|\cos\theta_V|$ **(c)** $\cos\theta^*_{\mathrm{Z}}$

Figure 6.14: Distributions of the polarisation-sensitive variables $\cos\theta^*_{\mathrm{W}}$, $|\cos\theta_V|$, and $\cos\theta^*_{\mathrm{Z}}$ for the WZjj-EW signal in each polarisation state. Run-3 SR events at reconstructed level are shown.

A full picture of the WZjj-EW topology is shown in Figure 6.15, where the variables described above are illustrated: the two forward jets, the central WZ system, and the angles and kinematic variables used to study the boson polarisation. In the event drawn, the W is transversely polarised and the Z longitudinally polarised. Two additional jets are present, one from final-state gluon radiation off a tagging quark, one from ISR, giving $N_{\mathrm{jets}} = 4$, already atypical for a VBS signal, whose gap is usually sparsely populated. The definition of the centralities is also indicated.

6.3 Polarisation density matrix

As explained in Section 1.11, the helicity corresponds to the projection of the spin along the momentum direction as stated in Equation 1.92. On the other hand, the polarisation is a statistical concept that describes the mixture of helicity states across the ensemble of bosons produced in a given process. Helicity labels the basis states while the polarisation describes the probability distribution over them.

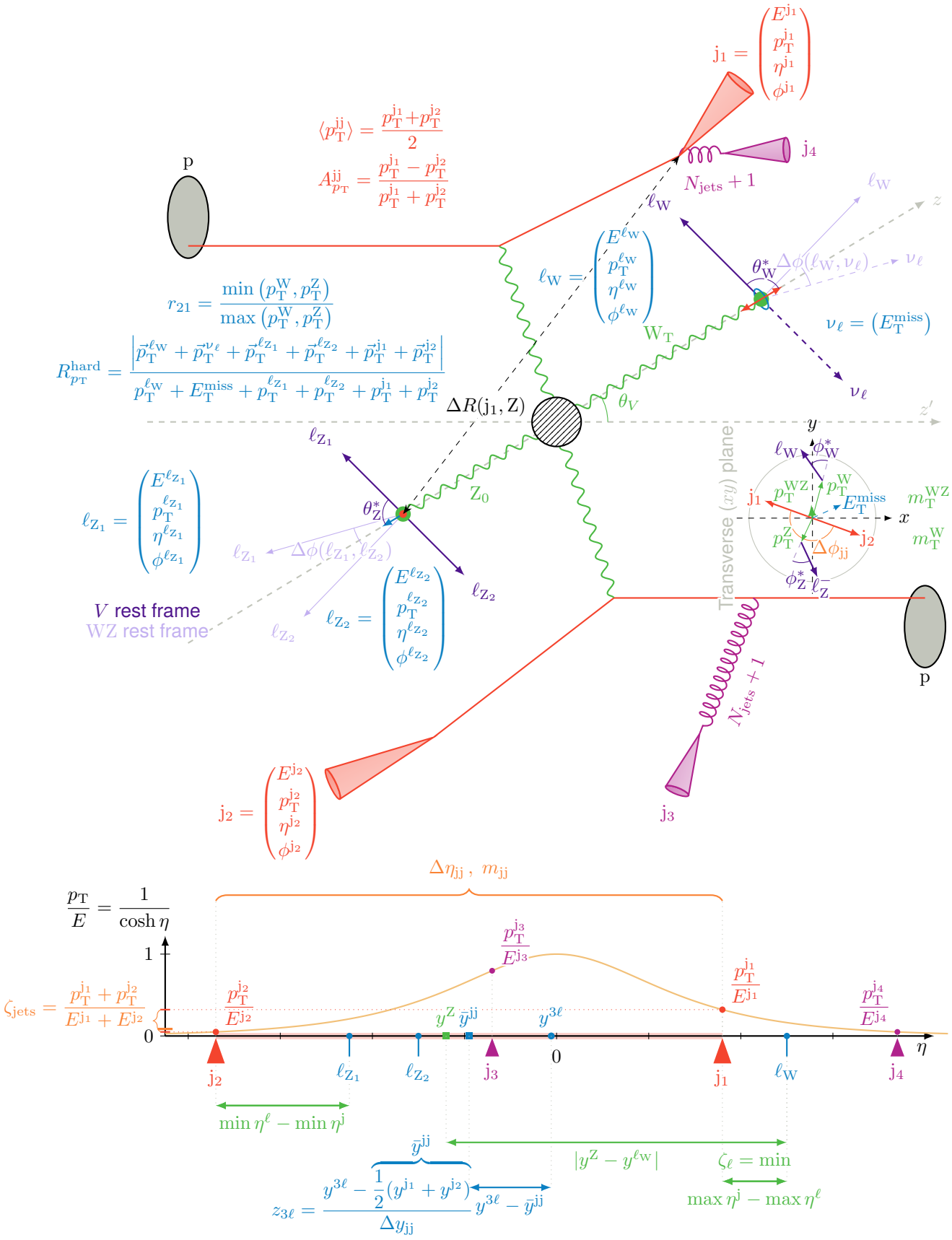


Figure 6.15: Scheme of a WZjj-EW event and the resulting topology of the final state with key parameters describing the event.

Given a pure quantum state:

$$|\psi\rangle = \sum_{\lambda} c_\lambda |\lambda\rangle \,, \tag{6.42}$$

we can define the density operator [387, 388] as:

$$\hat{\rho} = |\psi\rangle\langle\psi| = \sum_{\lambda,\lambda'} c_\lambda c^*_{\lambda'} |\lambda\rangle\langle\lambda'| = \sum_{\lambda,\lambda'} \rho_{\lambda\lambda'} |\lambda\rangle\langle\lambda'| \,, \tag{6.43}$$

where we introduce $\rho_{\lambda\lambda'} = c_\lambda c^*_{\lambda'}$ as the elements of the density matrix.

Since at hadron colliders different partonic processes, initial-state colours, and phase-space configurations contribute incoherently, the boson spin state is described by a *mixed state*. In the most general case, this mixed state is represented by:

$$\hat{\rho} = \sum_{\lambda,\lambda'=-1}^{+1} \rho_{\lambda\lambda'} |1,\lambda\rangle\langle 1,\lambda'| \,. \tag{6.44}$$

If the state is incoherent in the helicity basis, this reduces to the diagonal form:

$$\hat{\rho} = \sum_{\lambda=-1}^{+1} \Pi_\lambda |1,\lambda\rangle\langle 1,\lambda| \,, \tag{6.45}$$

where Π_λ is the probability of the boson being in the spin state $|1,\lambda\rangle$, and therefore lies in the range $[0,1]$ with the normalisation condition $\sum_\lambda \Pi_\lambda = 1$. This defines a 3×3 density matrix in the spin space, known as the spin density matrix (SDM) $\rho_{\lambda\lambda'} = \langle\lambda|\hat{\rho}|\lambda'\rangle$. This matrix satisfies the properties of a density matrix, such as being Hermitian ($\rho = \rho^\dagger$), positive ($\rho \geq 0$), and normalised ($\mathrm{Tr}(\rho) = 1$). It is thus a diagonalisable matrix, with positive eigenvalues that sum to one, which can be interpreted as the probabilities of the boson being in the corresponding eigenstates. Additionally, any observable $\hat{O}$ can be computed as the expectation value of the operator with respect to the density matrix:

$$\langle O\rangle = \mathrm{Tr}(\hat{\rho}\hat{O}) \,. \tag{6.46}$$

6.3.1 Single boson system

If we consider $\mathcal{M}^{\mathrm{prod}}_\lambda$ as the production amplitude of a vector boson V with helicity λ, since the production amplitude is contracted from the polarisation four-vector and the hard current we get:

$$\mathcal{M}^{\mathrm{prod}}_\lambda \propto \varepsilon^\mu_\lambda J_\mu \,. \tag{6.47}$$

From this we can derive the expression of the SDM elements as:

$$\rho^V_{\lambda\lambda'} = \frac{\sum_{\text{unobs.}} \mathcal{M}^{\mathrm{prod}}_\lambda \left(\mathcal{M}^{\mathrm{prod}}_{\lambda'}\right)^*}{\sum_{\lambda''}\sum_{\text{unobs.}} \left|\mathcal{M}^{\mathrm{prod}}_{\lambda''}\right|^2} \,, \tag{6.48}$$

where the sum over unobserved degrees of freedom includes the sum over initial-state spin and colours. The diagonal elements $\rho^V_{\lambda\lambda}$ represent the probability of the vector boson V being produced with helicity λ and can be written as:

$$f_\mathrm{R} \equiv \rho^V_{++} \,, \qquad f_0 \equiv \rho^V_{00} \,, \qquad f_\mathrm{L} \equiv \rho^V_{--} \,, \qquad f_\mathrm{T} \equiv f_\mathrm{R} + f_\mathrm{L} \,, \qquad f_\mathrm{R} + f_0 + f_\mathrm{L} = 1 \,, \tag{6.49}$$

and can be viewed as the fractions of bosons produced with each helicity. The f_T fraction is the total transverse fraction, while f_0 is the longitudinal fraction. Meanwhile, the off-diagonal elements $\rho^V_{\lambda\lambda'}$ with $\lambda \neq \lambda'$ represent the coherence between different helicity states.

A 3×3 Hermitian matrix with unit trace has $3^2 - 1 = 8$ independent real parameters. Therefore, the SDM of a vector boson can be decomposed into a rank-1 vector polarisation $\langle\vec{S}\rangle$ (three parameters) and a rank-2 tensor alignment $\langle T_k\rangle$ (five parameters) as [389, 390]:

$$\rho^V = \frac{1}{3}\mathbb{I}_3 + \frac{1}{2}\sum_{k=-1}^{+1} \langle S_k\rangle^* S_k + \sum_{k=-2}^{+2} \langle T_k\rangle^* T_k \,, \tag{6.50}$$

where $\mathbb{I}_3$ is the identity matrix, S_k are the spin-1 operators and T_k are the rank-2 tensor operators.

For a spin-$\frac{1}{2}$ particle, the SDM has $2^2 - 1 = 3$ independent parameters, which can be fully described by a rank-1 vector polarisation $\langle\vec{S}\rangle$. In our case of a spin-1 particle, the richer structure of $\mathrm{SU}(2)$ necessitates the inclusion of five additional parameters to account for the tensor alignment. The SDM is therefore made of two distinct components:

Vector polarisation $\langle\vec{S}\rangle$. This component describes the average spin orientation of the boson and is represented by a 3D vector. It captures the net alignment of the boson's spin along different axes. If $\langle S_z\rangle \neq 0$, it indicates a preference for the boson to be produced with a helicity of either $+1$ or -1. It vanishes, however, for an equal mixture of $\lambda = +1$ and $\lambda = -1$ states even if $\lambda = 0$ is depleted.

Tensor alignment $\langle T_k\rangle$. This component describes the anisotropies in the spin distribution that are invisible to the first-order moment $\langle\vec{S}\rangle$. It captures the correlations between different spin states and can reveal more subtle features of the boson's polarisation. A non-zero $\langle T_0\rangle$ can exist even when the vector polarisation $\langle\vec{S}\rangle$ is zero, for instance when $f_0 \neq \frac{1}{3}$ but $f_\mathrm{R} = f_\mathrm{L}$. There is no analogue for this effect in the spin-$\frac{1}{2}$ case.

To understand how this SDM decomposition connects to alternative representations found in the literature, it is instructive to look at the SDM expressed in the Cartesian basis usually found. Using Cartesian spatial indices ($i, j = x, y, z$), the density matrix is often written as:

$$\rho^V = \frac{1}{3}\mathbb{I}_3 + \frac{1}{2}\vec{p}\cdot\vec{S} + \frac{1}{\sqrt{6}}\sum_{i,j} T_{ij}(S_i S_j + S_j S_i)\,, \tag{6.51}$$

where S_i are the standard Cartesian spin-1 matrices, and T_{ij} represents the Cartesian tensor alignment components. In this convention, the notation $\vec{p}$ explicitly denotes the polarisation vector $\langle\vec{S}\rangle$ rather than a kinematic momentum. Expanding the Cartesian inner product yields:

$$\vec{p}\cdot\vec{S} = p_x S_x + p_y S_y + p_z S_z = \sum_i p_i S_i\,, \tag{6.52}$$

where $p_i \equiv \langle S_i\rangle$.

Moving from this Cartesian representation to the helicity framework requires mapping both the polarisation vector $\vec{p}$ and the spin operators S_i into a spherical basis. The spherical components p_k of the polarisation vector are defined using standard phase conventions:

$$p_{+1} = -\frac{1}{\sqrt{2}}(p_x + \mathrm{i}p_y)\,, \qquad p_0 = p_z\,, \qquad p_{-1} = \frac{1}{\sqrt{2}}(p_x - \mathrm{i}p_y)\,. \tag{6.53}$$

Because the basis vectors of the spherical representation are complex-valued, the Cartesian inner product structurally shifts to accommodate the unitary metric, transforming directly into the spherical expansion:

$$\vec{p}\cdot\vec{S} = \sum_{k=-1}^{+1}(-1)^k p_{-k} S_k = \sum_{k=-1}^{+1}\langle S_k\rangle^* S_k\,. \tag{6.54}$$

Substituting Equation 6.54 into the vector term $\frac{1}{2}\vec{p}\cdot\vec{S}$ of Equation 6.51 demonstrates its exact equivalence with the spherical formulation $\frac{1}{2}\sum_{k=-1}^{+1}\langle S_k\rangle^* S_k$ introduced in Equation 6.50, the complex conjugate being what makes the spherical expansion Hermitian.

The spin-1 operators S_k can be represented by the following matrices in the helicity basis:

$$S_x = \frac{1}{\sqrt{2}}\begin{pmatrix}0&1&0\\1&0&1\\0&1&0\end{pmatrix}\,, \qquad S_y = \frac{1}{\sqrt{2}}\begin{pmatrix}0&-\mathrm{i}&0\\\mathrm{i}&0&-\mathrm{i}\\0&\mathrm{i}&0\end{pmatrix}\,, \qquad S_z = \begin{pmatrix}1&0&0\\0&0&0\\0&0&-1\end{pmatrix}\,, \tag{6.55}$$

which can be used to form:

$$S_{+1} = -\frac{1}{\sqrt{2}}(S_x + \mathrm{i}S_y)\,, \qquad S_0 = S_z\,, \qquad S_{-1} = \frac{1}{\sqrt{2}}(S_x - \mathrm{i}S_y)\,, \tag{6.56}$$

which correspond to the ladder operators and the spin projection operator along the z-axis. From this we can derive the explicit form of the rank-2 tensor operators T_k as:

$$T_{\pm1} = \mp\frac{1}{2}\left(S_z S_\pm + S_\pm S_z\right)\,, \qquad T_0 = \frac{1}{\sqrt{6}}\left(3S_z^2 - 2\mathbb{I}_3\right)\,, \qquad T_{\pm2} = \frac{1}{2}S_\pm^2\,. \tag{6.57}$$

They correspond to the symmetric, traceless products of spin operators (analogous to spherical harmonics Y_S^m). The most important component of the tensor alignment is $\langle T_0 \rangle$, which is obtained from:

$$\vec{S}^2 = S(S+1)\mathbb{I}_3 = 2\mathbb{I}_3 \quad \Rightarrow \quad S_x^2 + S_y^2 + S_z^2 = 2\mathbb{I}_3 \quad \Rightarrow \quad T_0 = \frac{1}{\sqrt{6}}\left(3S_z^2 - 2\mathbb{I}_3\right) , \tag{6.58}$$

where the normalisation factor $\frac{1}{\sqrt{6}}$ is chosen such that $\mathrm{Tr}(T_0) = 0$ and $\mathrm{Tr}(T_0^2) = 1$. This component captures the deviation of the longitudinal polarisation fraction f_0 from the unpolarised value of $\frac{1}{3}$, and therefore is directly related to the depletion or enhancement of the longitudinal polarisation fraction. The other components of the tensor alignment capture more subtle features of the polarisation, such as the coherence between different helicity states and the anisotropies in the spin distribution. Similarly to the ladder operator we can write down the off-diagonal tensor operator as combinations:

$$A_1 = \frac{1}{2}\left(T_{+1} - T_{-1}\right) , \qquad A_2 = -\frac{\mathrm{i}}{2}\left(T_{+1} + T_{-1}\right) , \tag{6.59}$$

$$B_1 = \frac{1}{2}\left(T_{+2} + T_{-2}\right) , \qquad B_2 = -\frac{\mathrm{i}}{2}\left(T_{+2} - T_{-2}\right) . \tag{6.60}$$

Building upon this definition, Equation 6.50 can be expanded in the helicity basis where $\lambda = +1, 0, -1$:

$$\rho^V = \begin{pmatrix} \frac{1}{3} + \frac{\langle S_z \rangle}{2} + \frac{\langle T_0 \rangle}{\sqrt{6}} & \frac{\frac{\langle S_x \rangle}{2} - \langle A_1 \rangle - \mathrm{i}\left(\frac{\langle S_y \rangle}{2} - \langle A_2 \rangle\right)}{\sqrt{2}} & \langle B_1 \rangle - \mathrm{i}\langle B_2 \rangle \\ \frac{\frac{\langle S_x \rangle}{2} - \langle A_1 \rangle + \mathrm{i}\left(\frac{\langle S_y \rangle}{2} - \langle A_2 \rangle\right)}{\sqrt{2}} & \frac{1}{3} - \frac{2\langle T_0 \rangle}{\sqrt{6}} & \frac{\frac{\langle S_x \rangle}{2} + \langle A_1 \rangle - \mathrm{i}\left(\frac{\langle S_y \rangle}{2} + \langle A_2 \rangle\right)}{\sqrt{2}} \\ \langle B_1 \rangle + \mathrm{i}\langle B_2 \rangle & \frac{\frac{\langle S_x \rangle}{2} + \langle A_1 \rangle + \mathrm{i}\left(\frac{\langle S_y \rangle}{2} + \langle A_2 \rangle\right)}{\sqrt{2}} & \frac{1}{3} - \frac{\langle S_z \rangle}{2} + \frac{\langle T_0 \rangle}{\sqrt{6}} \end{pmatrix} . \tag{6.61}$$

Using Equation 6.46, we can derive the expression of the helicity fractions in terms of the vector polarisation and tensor alignment parameters as:

$$f_\mathrm{R} = \frac{1}{3} + \frac{\langle S_z \rangle}{2} + \frac{\langle T_0 \rangle}{\sqrt{6}} , \qquad f_0 = \frac{1}{3} - \frac{2\langle T_0 \rangle}{\sqrt{6}} , \qquad f_\mathrm{L} = \frac{1}{3} - \frac{\langle S_z \rangle}{2} + \frac{\langle T_0 \rangle}{\sqrt{6}} . \tag{6.62}$$

From this we can see that the vector polarisation $\langle S_z \rangle$ controls the difference between the RH and LH helicity fractions, while the tensor alignment $\langle T_0 \rangle$ controls the deviation of the longitudinal helicity fraction from the unpolarised value of $\frac{1}{3}$. In the unpolarised limit where $\langle S_z \rangle = 0$ and $\langle T_0 \rangle = 0$, we get $f_\mathrm{R} = f_0 = f_\mathrm{L} = \frac{1}{3}$, which corresponds to an equal mixture of the three helicity states. The SDM falls back as expected to $\frac{1}{3}\mathbb{I}_3$ in this limit.

Wigner, in 1927, introduced the concept of the D-matrix [391], which describes the transformation properties of quantum states under rotations. Given a set of three generators of the group $\mathrm{SU}(2)$, for example the spin operators $\vec{S} = (S_x, S_y, S_z)$, we can define a generic rotation operator as:

$$R(\alpha, \beta, \gamma) = \mathrm{e}^{-\mathrm{i}\alpha S_z}\mathrm{e}^{-\mathrm{i}\beta S_y}\mathrm{e}^{-\mathrm{i}\gamma S_z} , \tag{6.63}$$

where α, β, and γ are the Euler angles that parametrise the rotation. The Wigner D-matrix is then defined as the ME of the rotation operator in a given basis, for example the helicity basis:

$$D^S_{m'm}(\alpha, \beta, \gamma) = \langle S, m' | R(\alpha, \beta, \gamma) | S, m \rangle = \mathrm{e}^{-\mathrm{i}\alpha m'} d^S_{m'm}(\beta) \mathrm{e}^{-\mathrm{i}\gamma m} , \tag{6.64}$$

where $d^S_{m'm}(\beta)$ is the Wigner small d-matrix, which depends only on the angle β. The Wigner D-matrix describes how the quantum states transform under rotations and can be used to relate the density matrix in different reference frames [392].

The formal expression of the latter is given by:

$$\begin{aligned} d^S_{m'm}(\beta) = \sum_k & \frac{(-1)^{m'-m+k}\sqrt{(S+m')!(S-m')!(S+m)!(S-m)!}}{(S+m-k)!k!(m'-m+k)!(S-m'-k)!} \\ & \times \left(\cos\frac{\beta}{2}\right)^{2S+m-m'-2k} \left(\sin\frac{\beta}{2}\right)^{m'-m+2k} , \end{aligned} \tag{6.65}$$

where the sum over k runs over all integers for which the arguments of the four factorials in the denominator remain non-negative.

We note that the elements of the Wigner D-matrix with a second index equal to zero are proportional to the spherical harmonics Y_S^m:

$$D^S_{m0}(\alpha,\beta,0)^* = \sqrt{\frac{4\pi}{2S+1}}Y^m_S(\beta,\alpha)\,. \tag{6.66}$$

For a spin-1 boson, we can thus write the Wigner d-matrix as a function of the polar decay angle θ^*:

$$d^1(\theta^*) = \begin{pmatrix} \frac{1+\cos\theta^*}{2} & -\frac{\sin\theta^*}{\sqrt{2}} & \frac{1-\cos\theta^*}{2} \\ \frac{\sin\theta^*}{\sqrt{2}} & \cos\theta^* & -\frac{\sin\theta^*}{\sqrt{2}} \\ \frac{1-\cos\theta^*}{2} & \frac{\sin\theta^*}{\sqrt{2}} & \frac{1+\cos\theta^*}{2} \end{pmatrix}, \tag{6.67}$$

where rows (columns) are indexed by m' (m) in the order $+1, 0, -1$. The element $d^1_{\lambda m}(\theta^*)$ is used below with the row index identified with the boson helicity λ and the column index with the helicity difference m of the decay-fermion pair. The three helicity templates follow from the column $m = +1$:

$$\left|d^1_{+1,+1}\right|^2 = \frac{(1+\cos\theta^*)^2}{4}\,, \qquad \left|d^1_{0,+1}\right|^2 = \frac{\sin^2\theta^*}{2}\,, \qquad \left|d^1_{-1,+1}\right|^2 = \frac{(1-\cos\theta^*)^2}{4}\,. \tag{6.68}$$

Each integrates to $\frac{2}{3}$ over $\cos\theta^* \in [-1,1]$.

Given a boson in helicity state λ decaying to two massless fermions with helicities h_1 and h_2, the relevant spin projection of the decay-fermion pair is $m = h_1 - h_2 = +1$ or -1. The decay amplitude at angles (θ^*,ϕ^*) is related to the Wigner d-matrix as:

$$\mathcal{M}^{\text{decay}}_\lambda(\theta^*,\phi^*) = \sum_{m=\pm1} g_m\, d^1_{\lambda m}(\theta^*)\, \mathrm{e}^{\mathrm{i}\lambda\phi^*}\,, \tag{6.69}$$

where g_m are coupling weights fixed by the gauge structure of the decay vertex. The decay density matrix is built from the decay amplitude as:

$$\Gamma^V_{\lambda\lambda'}(\theta^*,\phi^*) = \mathcal{M}^{\text{decay}}_\lambda(\theta^*,\phi^*)\mathcal{M}^*_{\lambda'}(\theta^*,\phi^*)\,. \tag{6.70}$$

This object is purely kinematic: it encodes the angular pattern produced by each helicity state, and is computed once from the coupling constants of the decay vertex.

For $\mathrm{W}^+ \to \ell^+\nu_\ell$ with ℓ^+ as the angular reference, the V–A vertex restricts the lepton pair spin projection to $m = +1$ (the ℓ^+ is RH and the ν_ℓ is LH). Only the $m = +1$ column of the Wigner d-matrix contributes:

$$\Gamma^{\mathrm{W}^+}_{\lambda\lambda'}(\theta^*,\phi^*) = d^1_{\lambda,+1}(\theta^*)\left(d^1_{\lambda',+1}(\theta^*)\right)^* \mathrm{e}^{\mathrm{i}(\lambda-\lambda')\phi^*}\,. \tag{6.71}$$

The explicit matrix is therefore:

$$\Gamma^{\mathrm{W}^+}(\theta^*,\phi^*) = \begin{pmatrix} \frac{(1+\cos\theta^*)^2}{4} & \frac{\sin\theta^*(1+\cos\theta^*)}{2\sqrt{2}}\,\mathrm{e}^{+\mathrm{i}\phi^*} & \frac{\sin^2\theta^*}{4}\,\mathrm{e}^{+2\mathrm{i}\phi^*} \\ \frac{\sin\theta^*(1+\cos\theta^*)}{2\sqrt{2}}\,\mathrm{e}^{-\mathrm{i}\phi^*} & \frac{\sin^2\theta^*}{2} & \frac{\sin\theta^*(1-\cos\theta^*)}{2\sqrt{2}}\,\mathrm{e}^{+\mathrm{i}\phi^*} \\ \frac{\sin^2\theta^*}{4}\,\mathrm{e}^{-2\mathrm{i}\phi^*} & \frac{\sin\theta^*(1-\cos\theta^*)}{2\sqrt{2}}\,\mathrm{e}^{-\mathrm{i}\phi^*} & \frac{(1-\cos\theta^*)^2}{4} \end{pmatrix}. \tag{6.72}$$

The overall normalisation is chosen such that the trace of the decay density matrix equals unity at every point in the angular phase-space:

$$\mathrm{Tr}\left(\Gamma^{\mathrm{W}^+}(\theta^*,\phi^*)\right) = \frac{(1+\cos\theta^*)^2}{4} + \frac{\sin^2\theta^*}{2} + \frac{(1-\cos\theta^*)^2}{4} = 1\,. \tag{6.73}$$

The diagonal entries reproduce the templates of Equation 6.68. The off-diagonal entries carry the $\mathrm{e}^{\pm\mathrm{i}\phi^*}$ and $\mathrm{e}^{\pm2\mathrm{i}\phi^*}$ azimuthal dependence. The matrix is Hermitian and satisfies $\mathrm{Tr}\,\Gamma^{\mathrm{W}^+}(\theta^*,\phi^*) = 1$ at each angular point. The normalised angular distribution, not $\mathrm{Tr}\,\Gamma$ alone, integrates to unity over $\mathrm{d}\,\Omega$.

For $\mathrm{W}^- \to \ell^-\bar\nu_\ell$ with ℓ^- as the angular reference, the V–A coupling now restricts the lepton pair spin projection to $m = -1$ (the ℓ^- is LH and the $\bar\nu_\ell$ is RH). Only the $m = -1$ column of the Wigner d-matrix contributes:

$$\Gamma^{\mathrm{W}^-}_{\lambda\lambda'}(\theta^*,\phi^*) = d^1_{\lambda,-1}(\theta^*)\left(d^1_{\lambda',-1}(\theta^*)\right)^* \mathrm{e}^{\mathrm{i}(\lambda-\lambda')\phi^*}\,. \tag{6.74}$$

The diagonal templates are obtained from the W^+ ones by the replacement $\cos\theta^* \to -\cos\theta^*$, while the off-diagonal signs follow from the signs in the $m = -1$ column of the Wigner d-matrix. Explicitly:

$$\Gamma^{\mathrm{W}^-}(\theta^*, \phi^*) = \begin{pmatrix} \frac{(1-\cos\theta^*)^2}{4} & -\frac{\sin\theta^*(1-\cos\theta^*)}{2\sqrt{2}}\,\mathrm{e}^{+\mathrm{i}\phi^*} & \frac{\sin^2\theta^*}{4}\,\mathrm{e}^{+2\mathrm{i}\phi^*} \\ -\frac{\sin\theta^*(1-\cos\theta^*)}{2\sqrt{2}}\,\mathrm{e}^{-\mathrm{i}\phi^*} & \frac{\sin^2\theta^*}{2} & -\frac{\sin\theta^*(1+\cos\theta^*)}{2\sqrt{2}}\,\mathrm{e}^{+\mathrm{i}\phi^*} \\ \frac{\sin^2\theta^*}{4}\,\mathrm{e}^{-2\mathrm{i}\phi^*} & -\frac{\sin\theta^*(1+\cos\theta^*)}{2\sqrt{2}}\,\mathrm{e}^{-\mathrm{i}\phi^*} & \frac{(1+\cos\theta^*)^2}{4} \end{pmatrix}. \tag{6.75}$$

For $\mathrm{Z} \to \ell^+\ell^-$, both $m = +1$ (RH ℓ^- with coupling g_{R}^ℓ) and $m = -1$ (LH ℓ^- with coupling g_{L}^ℓ) contribute. With ℓ^- as the angular reference, the decay density matrix is given by:

$$\Gamma^{\mathrm{Z}}_{\lambda\lambda'}(\theta^*, \phi^*) = \frac{(g_{\mathrm{R}}^\ell)^2 d^1_{\lambda,+1}(\theta^*)\left(d^1_{\lambda',+1}(\theta^*)\right)^* + (g_{\mathrm{L}}^\ell)^2 d^1_{\lambda,-1}(\theta^*)\left(d^1_{\lambda',-1}(\theta^*)\right)^*}{(g_{\mathrm{R}}^\ell)^2 + (g_{\mathrm{L}}^\ell)^2}\,\mathrm{e}^{\mathrm{i}(\lambda-\lambda')\phi^*}. \tag{6.76}$$

We can define the coupling-asymmetry parameter A_ℓ as:

$$A_\ell = \frac{(g_{\mathrm{L}}^\ell)^2 - (g_{\mathrm{R}}^\ell)^2}{(g_{\mathrm{L}}^\ell)^2 + (g_{\mathrm{R}}^\ell)^2}. \tag{6.77}$$

Both LH and RH couplings of the Z vertex contribute. Defining the LH and RH chiral coupling constants as:

$$g_{\mathrm{L}}^\ell = -\frac{1}{2} + \sin^2\theta_{\mathrm{W}} \approx -0.269\,, \qquad g_{\mathrm{R}}^\ell = \sin^2\theta_{\mathrm{W}} \approx 0.231\,, \tag{6.78}$$

the coupling-asymmetry parameter can be rewritten as:

$$A_\ell = \frac{2 g_{\mathrm{V}}^\ell g_{\mathrm{A}}^\ell}{(g_{\mathrm{V}}^\ell)^2 + (g_{\mathrm{A}}^\ell)^2}\,, \qquad g_{\mathrm{V}}^\ell = \frac{g_{\mathrm{L}}^\ell + g_{\mathrm{R}}^\ell}{2}\,, \qquad g_{\mathrm{A}}^\ell = \frac{g_{\mathrm{L}}^\ell - g_{\mathrm{R}}^\ell}{2}\,, \tag{6.79}$$

yielding numerically:

$$g_{\mathrm{V}}^\ell \approx -0.019\,, \qquad g_{\mathrm{A}}^\ell \approx -0.250\,, \qquad A_\ell = \frac{2\times(-0.019)\times(-0.250)}{(-0.019)^2 + (-0.250)^2} \approx +0.15\,. \tag{6.80}$$

The explicit decay density matrix is thus:

$$\Gamma^{\mathrm{Z}}(\theta^*, \phi^*) = \begin{pmatrix} \frac{1+\cos^2\theta^* - 2A_\ell\cos\theta^*}{4} & \frac{\sin\theta^*(\cos\theta^* - A_\ell)}{2\sqrt{2}}\,\mathrm{e}^{+\mathrm{i}\phi^*} & \frac{\sin^2\theta^*}{4}\,\mathrm{e}^{+2\mathrm{i}\phi^*} \\ \frac{\sin\theta^*(\cos\theta^* - A_\ell)}{2\sqrt{2}}\,\mathrm{e}^{-\mathrm{i}\phi^*} & \frac{\sin^2\theta^*}{2} & -\frac{\sin\theta^*(\cos\theta^* + A_\ell)}{2\sqrt{2}}\,\mathrm{e}^{+\mathrm{i}\phi^*} \\ \frac{\sin^2\theta^*}{4}\,\mathrm{e}^{-2\mathrm{i}\phi^*} & -\frac{\sin\theta^*(\cos\theta^* + A_\ell)}{2\sqrt{2}}\,\mathrm{e}^{-\mathrm{i}\phi^*} & \frac{1+\cos^2\theta^* + 2A_\ell\cos\theta^*}{4} \end{pmatrix}. \tag{6.81}$$

We use here again the same normalisation convention as for the W decay density matrix.

Some features of this latter matrix:

- The $(2,2)$ longitudinal diagonal entry in $\frac{\sin^2\theta^*}{2}$ is independent of A_ℓ since $\left|d^1_{0,+1}\right|^2 = \left|d^1_{0,-1}\right|^2 = \frac{\sin^2\theta^*}{2}$. The longitudinal template is therefore identical for the W and Z.
- The $(1,3)$ and $(3,1)$ off-diagonal entries $\frac{\sin^2\theta^*}{4}\,\mathrm{e}^{\pm 2\mathrm{i}\phi^*}$ are also independent of A_ℓ: they involve $d^1_{+1,m} d^1_{-1,m}$ which contributes equally for both m values.
- Setting the coupling-asymmetry parameter to $A_\ell = -1$ corresponds to a pure RH coupling of the reference lepton, $g_{\mathrm{L}}^\ell = 0$, and selects the $m = +1$ column. With the present phase convention, this gives the same decay matrix as Γ^{W^+} for the reference antilepton.
- Setting the coupling-asymmetry parameter to $A_\ell = +1$ corresponds to a pure LH coupling of the reference lepton, $g_{\mathrm{R}}^\ell = 0$, and selects the $m = -1$ column. With the present phase convention, this gives the same decay matrix as Γ^{W^-} for the reference lepton ℓ^-.
- The off-diagonal $(1,2)$ entry involves $(\cos\theta^* - A_\ell)$ for the Z versus $(1+\cos\theta^*)$ for the W: the finite A_ℓ shifts the zero of the longitudinal–transverse interference template.

If ℓ^+ is taken as the angular reference instead of ℓ^-, the replacement $A_\ell \to -A_\ell$ is required (equivalently $\cos\theta^* \to -\cos\theta^*$), and the $(1,1)$–$(3,3)$ entries swap.

Once both the production density matrix ρ^V and the decay density matrix Γ^V are defined, we can derive the full angular distribution of the boson decay products by combining the SDM and the decay density matrix. The angular distribution is given by the trace of the product of the production SDM and the decay density matrix, summed over all helicity states:

$$\frac{1}{\sigma_V}\frac{\mathrm{d}\,\sigma_V}{\mathrm{d}\,\Omega_V} = \frac{3}{4\pi}\mathrm{Tr}\left[\rho^V(\Gamma^V)^\intercal\right] . \tag{6.82}$$

The prefactor $\frac{3}{4\pi}$ ensures that the full angular distribution integrates to unity over the solid angle. This expression encapsulates the full angular dependence of the decay products in terms of the boson's polarisation state and the decay dynamics. The trace operation ensures that we sum over all possible helicity states of the boson, weighted by their respective probabilities in the production process and their decay characteristics.

Inserting Equation 6.72 into Equation 6.82 and taking the trace yields the full 2D angular distribution:

$$\begin{aligned}\frac{1}{\sigma_{\mathrm{W}^\pm\to\ell^\pm\nu_\ell}}\frac{\mathrm{d}^2\,\sigma_{\mathrm{W}^\pm\to\ell^\pm\nu_\ell}}{\mathrm{d}\,\cos\theta^*\mathrm{d}\,\phi^*} = \frac{3}{16\pi}\Big[&1+\cos^2\theta^* \pm 2\langle S_z\rangle\cos\theta^* \\ &+\left(\frac{1}{3}-\frac{2}{\sqrt{6}}\langle T_0\rangle\right)\left(1-3\cos^2\theta^*\right) \\ &\pm 2\langle S_x\rangle\sin\theta^*\cos\phi^* \pm 2\langle S_y\rangle\sin\theta^*\sin\phi^* \\ &-2\langle A_1\rangle\sin 2\theta^*\cos\phi^* - 2\langle A_2\rangle\sin 2\theta^*\sin\phi^* \\ &+2\langle B_1\rangle\sin^2\theta^*\cos 2\phi^* + 2\langle B_2\rangle\sin^2\theta^*\sin 2\phi^*\Big] .\end{aligned} \tag{6.83}$$

Integrating over ϕ^* removes all off-diagonal SDM terms:

$$\frac{1}{\sigma_{\mathrm{W}^\pm\to\ell^\pm\nu_\ell}}\frac{\mathrm{d}\,\sigma_{\mathrm{W}^\pm\to\ell^\pm\nu_\ell}}{\mathrm{d}\,\cos\theta^*} = \frac{3}{8}\left[f_\mathrm{L}(1\mp\cos\theta^*)^2 + f_\mathrm{R}(1\pm\cos\theta^*)^2 + 2f_0\left(1-\cos^2\theta^*\right)\right] . \tag{6.84}$$

Inserting Equation 6.81 into Equation 6.82 and explicitly evaluating the trace yields the full 2D angular distribution for the Z boson:

$$\begin{aligned}\frac{1}{\sigma_{\mathrm{Z}\to\ell^+\ell^-}}\frac{\mathrm{d}^2\,\sigma_{\mathrm{Z}\to\ell^+\ell^-}}{\mathrm{d}\,\cos\theta^*\mathrm{d}\,\phi^*} = \frac{3}{16\pi}\Big[&1+\cos^2\theta^* - 2A_\ell\langle S_z\rangle\cos\theta^* \\ &+\left(\frac{1}{3}-\frac{2}{\sqrt{6}}\langle T_0\rangle\right)\left(1-3\cos^2\theta^*\right) \\ &-2A_\ell\langle S_x\rangle\sin\theta^*\cos\phi^* - 2A_\ell\langle S_y\rangle\sin\theta^*\sin\phi^* \\ &-2\langle A_1\rangle\sin 2\theta^*\cos\phi^* - 2\langle A_2\rangle\sin 2\theta^*\sin\phi^* \\ &+2\langle B_1\rangle\sin^2\theta^*\cos 2\phi^* + 2\langle B_2\rangle\sin^2\theta^*\sin 2\phi^*\Big] .\end{aligned} \tag{6.85}$$

Integrating over ϕ^* removes all off-diagonal SDM terms, isolating the diagonal contributions:

$$\begin{aligned}\frac{1}{\sigma_{\mathrm{Z}\to\ell^+\ell^-}}\frac{\mathrm{d}\,\sigma_{\mathrm{Z}\to\ell^+\ell^-}}{\mathrm{d}\,\cos\theta^*} = \frac{3}{8}\Big[&f_\mathrm{R}\left(1+\cos^2\theta^* - 2A_\ell\cos\theta^*\right) + f_\mathrm{L}\left(1+\cos^2\theta^* + 2A_\ell\cos\theta^*\right) \\ &+2f_0\left(1-\cos^2\theta^*\right)\Big] .\end{aligned} \tag{6.86}$$

Setting $A_\ell = -1$ recovers the 1D distribution for the W^+, while setting $A_\ell = +1$ recovers the 1D distribution for the W^-. The reduced forward–backward asymmetry (factor $A_\ell \approx 0.15$ versus 1 for the W) makes the $f_\mathrm{L}/f_\mathrm{R}$ separation harder experimentally for the Z.

We recover the expression derived in Appendix C from the Feynman amplitude calculation, confirming the consistency of the SDM formalism with the direct amplitude approach. The SDM formalism, however, provides a more general and systematic framework to describe the angular distributions, including the off-diagonal coherences and their impact on the angular patterns.

6.3.2 Diboson system

The calculation above determines the angular distribution for the decay of a single vector boson of *known* polarisation state λ. In a diboson process $\mathrm{pp} \to VV' \to 4\ell$, however, neither boson is produced in a definite polarisation eigenstate: the production amplitude $\mathcal{M}^{\mathrm{prod}}_{\lambda\lambda'}$ coherently populates all combinations of helicities λ and λ' of the two bosons simultaneously. The total amplitude for the full $2 \to 4$ process factorises in the narrow-width approximation (NWA) [393, 394] (more details are given in Section 6.4) as:

$$\mathcal{M}_{\mathrm{tot}} = \sum_{\lambda,\lambda'} \mathcal{M}^{\mathrm{prod}}_{\lambda\lambda'}\, \mathcal{M}^{\mathrm{decay}}_{\lambda}\, \mathcal{M}^{\mathrm{decay}}_{\lambda'}\,, \tag{6.87}$$

where $\mathcal{M}^{\mathrm{decay}}_{\lambda}$ is defined in Equation 6.69 and represents the decay amplitude of a boson with helicity λ into a fermion pair. Squaring the total amplitude gives:

$$|\mathcal{M}_{\mathrm{tot}}|^2 = \sum_{\lambda_V,\lambda_{V'}} \sum_{\lambda'_V,\lambda'_{V'}} \mathcal{M}^{\mathrm{prod}}_{\lambda_V\lambda_{V'}} \left(\mathcal{M}^{\mathrm{prod}}_{\lambda'_V\lambda'_{V'}}\right)^* \mathcal{M}^{\mathrm{decay},V}_{\lambda_V} \left(\mathcal{M}^{\mathrm{decay},V}_{\lambda'_V}\right)^* \mathcal{M}^{\mathrm{decay},V'}_{\lambda_{V'}} \left(\mathcal{M}^{\mathrm{decay},V'}_{\lambda'_{V'}}\right)^*. \tag{6.88}$$

From this we can identify the object $\rho^{VV'}_{\lambda_V\lambda_{V'},\lambda'_V\lambda'_{V'}}$ corresponding to the SDM of the diboson system [395]. It encodes not only the diagonal populations $\rho_{\lambda\lambda}$ of each helicity state, but also the off-diagonal coherences $\rho_{\lambda\lambda'}$ with $\lambda \neq \lambda'$, which arise whenever the production process creates a coherent superposition of polarisation states. It can be formally expressed as:

$$\rho^{VV'}_{\lambda_V\lambda_{V'},\lambda'_V\lambda'_{V'}} = \frac{\sum_{\text{unobs.}} \mathcal{M}^{\mathrm{prod}}_{\lambda_V\lambda_{V'}} \left(\mathcal{M}^{\mathrm{prod}}_{\lambda'_V\lambda'_{V'}}\right)^*}{\sum_{\lambda''_V,\lambda''_{V'}} \sum_{\text{unobs.}} \left|\mathcal{M}^{\mathrm{prod}}_{\lambda''_V\lambda''_{V'}}\right|^2}. \tag{6.89}$$

Using a WZ system, we can write the full SDM ρ^{WZ} following the decomposition used in Equation 6.50 for each boson:

$$\begin{aligned}\rho^{\mathrm{WZ}} = {} & \frac{1}{9}\mathbb{I}_3 \otimes \mathbb{I}_3 + \frac{1}{6}\sum_{k=-1}^{+1} \left(\langle S^{\mathrm{W}}_k\rangle^* S_k \otimes \mathbb{I}_3 + \langle S^{\mathrm{Z}}_k\rangle^* \mathbb{I}_3 \otimes S_k\right) + \frac{1}{3}\sum_{k=-2}^{+2} \left(\langle T^{\mathrm{W}}_k\rangle^* T_k \otimes \mathbb{I}_3 + \langle T^{\mathrm{Z}}_k\rangle^* \mathbb{I}_3 \otimes T_k\right) \\ & + \frac{1}{4}\sum_{k_1=-1}^{+1}\sum_{k_2=-1}^{+1} \langle S^{\mathrm{W}}_{k_1} S^{\mathrm{Z}}_{k_2}\rangle^* S_{k_1} \otimes S_{k_2} + \frac{1}{2}\sum_{k_1=-1}^{+1}\sum_{k_2=-2}^{+2} \langle S^{\mathrm{W}}_{k_1} T^{\mathrm{Z}}_{k_2}\rangle^* S_{k_1} \otimes T_{k_2} \\ & + \frac{1}{2}\sum_{k_1=-2}^{+2}\sum_{k_2=-1}^{+1} \langle T^{\mathrm{W}}_{k_1} S^{\mathrm{Z}}_{k_2}\rangle^* T_{k_1} \otimes S_{k_2} + \sum_{k_1=-2}^{+2}\sum_{k_2=-2}^{+2} \langle T^{\mathrm{W}}_{k_1} T^{\mathrm{Z}}_{k_2}\rangle^* T_{k_1} \otimes T_{k_2}\,.\end{aligned} \tag{6.90}$$

The full 9×9 SDM ρ^{WZ} for the two-boson system represents $9^2 - 1 = 80$ independent real parameters. We can express it in the helicity basis formed by the direct product of the two bosons' helicity bases, which is ordered as $\{|+1\rangle|+1\rangle, |+1\rangle|0\rangle, |+1\rangle|-1\rangle, |0\rangle|+1\rangle, |0\rangle|0\rangle, |0\rangle|-1\rangle, |-1\rangle|+1\rangle, |-1\rangle|0\rangle, |-1\rangle|-1\rangle\}$. The explicit notation of this SDM is $\rho^{\mathrm{WZ}}_{\lambda_{\mathrm{W}}\lambda_{\mathrm{Z}},\lambda'_{\mathrm{W}}\lambda'_{\mathrm{Z}}}$, where λ_{W} and λ_{Z} are the helicity states of the W and Z, respectively, and the primed indices correspond to the complex conjugate states in the density matrix. The diagonal entries of this matrix correspond to the joint helicity fractions of the two bosons, while the off-diagonal entries capture the interference effects and correlations between different helicity states of the two bosons. The details on the computation of this object are presented in Appendix B. From this we can group the transverse and longitudinal helicity fractions of the W and Z into the following combinations by mapping the standard helicity indices $+1$ and -1 to the physical RH (R) and LH (L) transverse polarisation configurations, respectively:

$$f_{00} \equiv \rho^{\mathrm{WZ}}_{00,00}\,, \tag{6.91}$$

$$f_{0\mathrm{T}} \equiv \rho^{\mathrm{WZ}}_{0-,0-} + \rho^{\mathrm{WZ}}_{0+,0+}\,, \tag{6.92}$$

$$f_{\mathrm{T}0} \equiv \rho^{\mathrm{WZ}}_{-0,-0} + \rho^{\mathrm{WZ}}_{+0,+0}\,, \tag{6.93}$$

$$f_{\mathrm{TT}} \equiv \sum_{i,j\in\{-,+\}} \rho^{\mathrm{WZ}}_{ij,ij}\,, \tag{6.94}$$

with the following normalisation:

$$f_{00} + f_{\mathrm{T}0} + f_{0\mathrm{T}} + f_{\mathrm{TT}} = 1\,. \tag{6.95}$$

This yields the following decomposition of each polarisation fraction in terms of the single-boson polarisation parameters:

$$f_{00} = \frac{1}{9} - \frac{2\langle T_0^{\mathrm{W}}\rangle}{3\sqrt{6}} - \frac{2\langle T_0^{\mathrm{Z}}\rangle}{3\sqrt{6}} + \frac{2}{3}\langle T_0^{\mathrm{W}}\rangle\langle T_0^{\mathrm{Z}}\rangle\,, \tag{6.96}$$

$$f_{0\mathrm{T}} = \frac{2}{9} - \frac{4\langle T_0^{\mathrm{W}}\rangle}{3\sqrt{6}} + \frac{2\langle T_0^{\mathrm{Z}}\rangle}{3\sqrt{6}} - \frac{2}{3}\langle T_0^{\mathrm{W}}\rangle\langle T_0^{\mathrm{Z}}\rangle\,, \tag{6.97}$$

$$f_{\mathrm{T}0} = \frac{2}{9} + \frac{2\langle T_0^{\mathrm{W}}\rangle}{3\sqrt{6}} - \frac{4\langle T_0^{\mathrm{Z}}\rangle}{3\sqrt{6}} - \frac{2}{3}\langle T_0^{\mathrm{W}}\rangle\langle T_0^{\mathrm{Z}}\rangle\,, \tag{6.98}$$

$$f_{\mathrm{TT}} = \frac{4}{9} + \frac{4\langle T_0^{\mathrm{W}}\rangle}{3\sqrt{6}} + \frac{4\langle T_0^{\mathrm{Z}}\rangle}{3\sqrt{6}} + \frac{2}{3}\langle T_0^{\mathrm{W}}\rangle\langle T_0^{\mathrm{Z}}\rangle\,, \tag{6.99}$$

where we give here only the explicit decomposition under the factorisation approximation. The full decomposition including the spin-correlation tensor is given in Appendix B.

Under this factorisation approximation $\rho^{\mathrm{WZ}} \approx \rho^{\mathrm{W}} \otimes \rho^{\mathrm{Z}}$, yielding $f_{00} \approx f_0^{\mathrm{W}} \cdot f_0^{\mathrm{Z}}$. In general, the spin-correlation tensor $\mathcal{C}^{\mathrm{WZ}} = \rho^{\mathrm{WZ}} - \rho^{\mathrm{W}} \otimes \rho^{\mathrm{Z}}$ is non-zero, and the factorisation approximation breaks down. The spin-correlation tensor captures the correlations between the helicity states of the two bosons, which can arise from interference effects and higher-order corrections in the production process. These correlations can lead to deviations from the simple product of the individual density matrices, and therefore can affect the angular distribution of the decay products. More emphasis on the decomposition of the two-boson SDM and the interpretation of the spin-correlation tensor is also given in Appendix B.

The full angular distribution of the decay products of both bosons in the WZ production process is given by the trace of the product of the joint SDM and the tensor product of the decay density matrices of the two bosons:

$$\frac{1}{\sigma_{\mathrm{WZ}\to\ell'^{\pm}\nu_\ell\ell\ell}}\frac{\mathrm{d}^2\,\sigma_{\mathrm{WZ}\to\ell'^{\pm}\nu_\ell\ell\ell}}{\mathrm{d}\,\Omega_{\mathrm{W}}^{*}\mathrm{d}\,\Omega_{\mathrm{Z}}^{*}} = \left(\frac{3}{4\pi}\right)^2 \mathrm{Tr}\left[\rho^{\mathrm{WZ}}\left(\Gamma^{\mathrm{W}}\otimes\Gamma^{\mathrm{Z}}\right)^{\intercal}\right]\,. \tag{6.100}$$

The explicit expression of the decay density matrix $\Gamma^{\mathrm{W}} \otimes \Gamma^{\mathrm{Z}}$ is given in Appendix B, and the explicit expression of the full angular distribution is given in Appendix B as well. We give below only the expression of the angular distribution integrated over the azimuthal angles ϕ_{W}^{*} and ϕ_{Z}^{*}, which isolates the diagonal SDM terms and groups together the transverse and longitudinal helicity fractions of the two bosons, after folding in $\cos\theta^{*}$ (the terms odd in $\cos\theta^{*}$, proportional to $f_{\mathrm{R}} - f_{\mathrm{L}}$, drop out):

$$\begin{aligned}\frac{1}{\sigma_{\mathrm{WZ}\to\ell'^{\pm}\nu_\ell\ell\ell}}\frac{\mathrm{d}^2\,\sigma_{\mathrm{WZ}\to\ell'^{\pm}\nu_\ell\ell\ell}}{\mathrm{d}\,\cos\theta_{\mathrm{W}}^{*}\mathrm{d}\,\cos\theta_{\mathrm{Z}}^{*}} = \frac{9}{64}\Big[&4f_{00}\left(1-\cos^2\theta_{\mathrm{W}}^{*}\right)\left(1-\cos^2\theta_{\mathrm{Z}}^{*}\right)\\ &+2f_{0\mathrm{T}}\left(1-\cos^2\theta_{\mathrm{W}}^{*}\right)\left(1+\cos^2\theta_{\mathrm{Z}}^{*}\right)\\ &+2f_{\mathrm{T}0}\left(1+\cos^2\theta_{\mathrm{W}}^{*}\right)\left(1-\cos^2\theta_{\mathrm{Z}}^{*}\right)\\ &+f_{\mathrm{TT}}\left(1+\cos^2\theta_{\mathrm{W}}^{*}\right)\left(1+\cos^2\theta_{\mathrm{Z}}^{*}\right)\Big]\,.\end{aligned} \tag{6.101}$$

6.3.3 Production of bosons

The formalism above is independent of the production mechanism [395]. The latter enters only through the production SDM. In the case of EW WZjj production, corresponding to the VBS topology, the produced WZ system is accompanied by two tagging jets. The relevant partonic subprocess can therefore be written schematically as:

$$qq \to qq\mathrm{WZ} \to qq\ell'\nu_\ell\ell\ell\,. \tag{6.102}$$

The full amplitude factorises at the level of helicity amplitudes into a production part and the decay amplitudes of the two vector bosons:

$$\mathcal{M}_{\mathrm{tot}} = \sum_{\lambda_{\mathrm{W}},\lambda_{\mathrm{Z}}} \mathcal{M}^{\mathrm{prod}}_{\lambda_{\mathrm{W}}\lambda_{\mathrm{Z}}}\left(qq\to qq\mathrm{W}_{\lambda_{\mathrm{W}}}\mathrm{Z}_{\lambda_{\mathrm{Z}}}\right)\mathcal{M}^{\mathrm{W}\to\ell\nu_\ell}_{\lambda_{\mathrm{W}}}\mathcal{M}^{\mathrm{Z}\to\ell\ell}_{\lambda_{\mathrm{Z}}}\,. \tag{6.103}$$

The VBS production mechanism is therefore encoded in the joint SDM:

$$\rho^{\mathrm{WZ,VBS}}_{\lambda_{\mathrm{W}}\lambda_{\mathrm{Z}},\lambda'_{\mathrm{W}}\lambda'_{\mathrm{Z}}} = \frac{\sum_{\mathrm{unobs.}}\int_{\Phi_{\mathrm{VBS}}}\mathcal{M}^{\mathrm{prod}}_{\lambda_{\mathrm{W}}\lambda_{\mathrm{Z}}}\left(\mathcal{M}^{\mathrm{prod}}_{\lambda'_{\mathrm{W}}\lambda'_{\mathrm{Z}}}\right)^{*}\mathrm{d}\,\Phi_{\mathrm{WZjj}}}{\sum_{\mathrm{unobs.}}\int_{\Phi_{\mathrm{VBS}}}\sum_{\lambda_{\mathrm{W}},\lambda_{\mathrm{Z}}}\left|\mathcal{M}^{\mathrm{prod}}_{\lambda_{\mathrm{W}}\lambda_{\mathrm{Z}}}\right|^2\mathrm{d}\,\Phi_{\mathrm{WZjj}}}\,. \tag{6.104}$$

Here Φ_{VBS} denotes the fiducial phase-space used to select the VBS topology. The decay matrices Γ^{W} and Γ^{Z} are unchanged with respect to the inclusive case, since they only describe the decay of spin-1 bosons. The difference between inclusive WZ production and WZjj-EW production is therefore entirely contained in the numerical values of the elements of ρ^{WZ}.

6.4 Narrow-width approximation and gauge invariance

The EW production of a WZ pair in association with two jets, the VBS process, is experimentally observed through its final decay and in the case of the fully leptonic channel:

$$\mathrm{pp} \to \mathrm{W}^{\pm}\mathrm{Zjj} \to \ell'^{\pm}\nu_{\ell}\ell^{+}\ell^{-}\mathrm{jj}\,. \tag{6.105}$$

A fundamental point is that the W and Z bosons never appear as external, asymptotic particles in this reaction: they are unstable intermediate resonances. Neither their spin nor their polarisation can be directly measured, but only inferred from the angular distributions of their decay products.

The physical amplitude therefore contains Breit–Wigner propagators for each intermediate boson V:

$$\mathcal{M} \propto \frac{1}{q_V^2 - m_V^2 + \mathrm{i}m_V\Gamma_V}\,, \tag{6.106}$$

and can be decomposed according to its resonance structure as:

$$\mathcal{M} = \underbrace{\mathcal{M}_{\text{doubly res.}}}_{\text{VBS}} + \mathcal{M}_{\text{singly res.}} + \mathcal{M}_{\text{non-res.}} + \mathcal{M}_{\text{interf.}}\,. \tag{6.107}$$

Only the doubly resonant part corresponds to the VBS topology in which both bosons are produced near their mass shell. A critical consequence of Equation 6.107 is that *none* of the individual helicity projections of $\mathcal{M}$ is separately gauge invariant.

To assign a well-defined polarisation state to the W or Z, one must first separate the production subprocess from the subsequent decay. This separation is made possible by the fact that both bosons are narrow resonances: their total widths $\Gamma_{\mathrm{W}} \approx 2.09\,\mathrm{GeV}$ and $\Gamma_{\mathrm{Z}} \approx 2.50\,\mathrm{GeV}$ satisfy $\frac{\Gamma_V}{m_V} \lesssim 3\,\%$. In Appendix C, we detail the computation of the total and partial widths as well as the BRs of both vector bosons. They will be used in the following without further discussion.

When the boson is produced near its mass shell, the propagator peaks sharply around $q_V^2 = m_V^2$. In the limit $\Gamma_V \to 0$ it reduces to a delta function:

$$\frac{1}{\left(q_V^2 - m_V^2\right)^2 + m_V^2\Gamma_V^2} \longrightarrow \frac{\pi}{m_V\Gamma_V}\delta\left(q_V^2 - m_V^2\right)\,, \tag{6.108}$$

and the phase-space factorises into a production part times a decay part. Integrating over the invariant mass using this delta function yields the NWA cross-section:

$$\sigma_{\mathrm{pp}\to\ell'\nu_\ell\ell\ell\mathrm{jj}} = \sigma_{\mathrm{pp}\to\mathrm{WZjj}} \times \underbrace{\frac{\Gamma(\mathrm{W}\to\ell\nu_\ell)}{\Gamma_{\mathrm{W}}}}_{\mathrm{BR}(\mathrm{W}\to\ell\nu_\ell)} \times \underbrace{\frac{\Gamma(\mathrm{Z}\to\ell^+\ell^-)}{\Gamma_{\mathrm{Z}}}}_{\mathrm{BR}(\mathrm{Z}\to\ell^+\ell^-)}\,, \tag{6.109}$$

exact up to corrections of order $\mathcal{O}(\frac{\Gamma_V}{m_V}) \simeq 2.5\,\%$ to $3.0\,\%$.

The BRs appearing in Equation 6.109 are *not* independent inputs: they are determined by the same gauge couplings that govern the production amplitude. From the SM EW vertices, the partial leptonic decay widths are:

$$\Gamma(\mathrm{W}\to\ell\nu_\ell) = \frac{G_{\mathrm{F}}m_{\mathrm{W}}^3}{6\pi\sqrt{2}} \approx 227\,\mathrm{MeV}\ (226\,\mathrm{MeV}\ \text{measured})\,, \tag{6.110}$$

$$\Gamma(\mathrm{Z}\to\ell^+\ell^-) = \frac{G_{\mathrm{F}}m_{\mathrm{Z}}^3}{3\pi\sqrt{2}}\left[(g_{\mathrm{L}}^{\ell})^2 + (g_{\mathrm{R}}^{\ell})^2\right] \approx 84\,\mathrm{MeV}\,, \tag{6.111}$$

and the measured per-generation branching fractions are:

$$\mathrm{BR}(\mathrm{W}\to\ell\nu_\ell) \approx 10.86\,\%\,, \qquad \mathrm{BR}(\mathrm{Z}\to\ell^+\ell^-) \approx 3.37\,\%\,. \tag{6.112}$$

In the three-lepton SR, requiring two same-flavour opposite-charge leptons from the Z (either $\mathrm{e}^+\mathrm{e}^-$ or $\mu^+\mu^-$) plus any charged lepton from the W, the combined leptonic branching fraction becomes:

$$\mathrm{BR}(\mathrm{W}\to\ell\nu_\ell) \times \mathrm{BR}(\mathrm{Z}\to\mathrm{e}^+\mathrm{e}^- + \mu^+\mu^-) \approx 3 \times 10.86\,\% \times 6.74\,\% \approx 2.2\,\%\,, \tag{6.113}$$

which directly sets the overall signal yield accessible to the analysis. The consistency of the coupling structure across production and decay is essential for the validity of the polarisation measurement and its interpretation in terms of the underlying physics. If the τ is not considered, as its reconstruction is more challenging given that it further decays into lighter leptons and neutrinos in the leptonic decay channel, the combined leptonic branching fraction becomes:

$$\mathrm{BR}(\mathrm{W} \to \mathrm{e}\nu_\mathrm{e} + \mu\nu_\mu) \times \mathrm{BR}(\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^- + \mu^+\mu^-) \approx 2 \times 10.86\,\% \times 6.74\,\% \approx 1.5\,\%\,. \tag{6.114}$$

For a boson produced in polarisation state λ, the differential decay rate as a function of the helicity angle θ^* in the boson rest frame is:

$$\frac{1}{\Gamma_\lambda}\frac{\mathrm{d}\,\Gamma_\lambda}{\mathrm{d}\,\cos\theta^*} = \frac{3}{2}\left|d^1_{\lambda,+1}(\theta^*)\right|^2\,, \tag{6.115}$$

where the Wigner d-ME encodes the angular shape characteristic of each helicity state. Deviations from a democratic distribution across helicity states arise entirely from the production dynamics, not from the decay: the partial decay width $\Gamma_\lambda = \Gamma(V \to \ell\ell')$ is the same for the three helicity states by rotational invariance.

Therefore, in the NWA, the polarised cross-section for each joint state (ij) is:

$$\sigma_{\mathrm{W}_i\mathrm{Z}_j\mathrm{jj\text{-}EW}\to\ell'\nu_\ell\ell\ell\mathrm{jj}} = f_{ij} \times \sigma_{\mathrm{pp}\to\mathrm{WZjj\text{-}EW}} \times \mathrm{BR}(\mathrm{W} \to \ell\,\nu_\ell) \times \mathrm{BR}(\mathrm{Z} \to \ell^+\ell^-)\,, \tag{6.116}$$

where f_{ij} is the polarisation fraction, and the BRs connect the undecayed WZjj-EW production cross-section to the observable leptonic final state.

The NWA in its strictest form places both bosons fully on-shell, discarding all off-shell effects. A more refined treatment is provided by the double pole approximation (DPA) [396], which retains only the doubly resonant diagrams of Equation 6.107 while keeping the full Breit–Wigner propagator shape for each boson. The amplitude is given by:

$$\mathcal{M}_\mathrm{DPA} = \sum_{\lambda_\mathrm{W},\lambda_\mathrm{Z}} \frac{\mathcal{M}^\mathrm{prod}_{\lambda_\mathrm{W}\lambda_\mathrm{Z}}}{(q^2_\mathrm{W} - m^2_\mathrm{W} + \mathrm{i}m_\mathrm{W}\Gamma_\mathrm{W})\,(q^2_\mathrm{Z} - m^2_\mathrm{Z} + \mathrm{i}m_\mathrm{Z}\Gamma_\mathrm{Z})}\mathcal{M}^\mathrm{decay}_{\lambda_\mathrm{W}}\mathcal{M}^\mathrm{decay}_{\lambda_\mathrm{Z}}\,. \tag{6.117}$$

Here only the doubly resonant residue is retained, but the off-shell momentum q_V ($q_V^2 \neq m_V^2$) is replaced by an on-shell projected momentum $\tilde{q}_V$ ($\tilde{q}_V^2 = m_V^2$) only in the amplitude numerator, while the Breit–Wigner denominator in Equation 6.106 keeps the original off-shell kinematics. The lineshape, and hence the leading off-shell effects, are therefore retained while gauge invariance is preserved. The on-shell mapping $q_V \to \tilde{q}_V$ is not unique and depends on which kinematic quantities are held fixed.

The DPA renders the polarisation decomposition self-consistent. Under this approximation, each helicity projection is individually gauge invariant to accuracy $\mathcal{O}(\frac{\Gamma_V}{m_V}) \approx 3\,\%$, which justifies defining polarised templates as separate, physically meaningful helicity projections and provides the theoretical foundation for the polarised MC samples. Compared to the strict NWA, the DPA, applied to the NLO corrections on top of the full LO, has an intrinsic uncertainty of $\mathcal{O}(\frac{\alpha\Gamma_V}{m_V})$ relative to LO and naturally retains off-shell propagator effects important for differential distributions away from the resonance peak. Both approximations agree in the limit $\Gamma_V \to 0$ and differ from the full off-shell result only at the percent level. For the present process the residual off-shell effects are about $1.5\,\%$ to $1.6\,\%$ as visible in Table 6.A, well inside the $\mathcal{O}(\frac{\Gamma_V}{m_V}) \approx 3\,\%$ to which the LO polarised cross-sections reproduce the full off-shell result.

These two approximations are needed to define the polarisation states of the W and Z in a gauge-invariant way. The NWA projects the resonant bosons onto their mass shell, so that the helicity states are well-defined. The DPA retains the full Breit–Wigner shape of the propagator for each boson. The necessity of these assumptions is detailed below.

Assigning a polarisation to an internal line requires factorising the amplitude into a production part, a propagator, and a decay part. The bridge between these three components is given by the completeness relation of the massive vector boson propagator, formulated from the vector boson propagator in Equation 6.33 and the normalisation of the polarisation vectors in Equation 1.103:

$$-g^{\mu\nu} + \frac{q_V^\mu q_V^\nu}{m_V^2} = \sum_\lambda \varepsilon^\mu_\lambda(q_V)\varepsilon^{*\nu}_\lambda(q_V)\,, \tag{6.118}$$

where the sum runs over the physical polarisation states $\lambda = 0, \pm 1$. On the mass shell the three physical polarisations saturate the sum. Off shell, the difference is a term proportional to $q_V^\mu q_V^\nu$, which is required for the gauge invariance off the mass shell but cancels exactly against the corresponding Goldstone boson contribution, and vanishes altogether for conserved currents such as those of massless final-state

fermions. Indeed, the term $g^{\mu\nu}$ comes from the propagation of a vector field, but a massive vector boson should only propagate the physical spin-1 components. The direction proportional to q_V^μ is not an independent degree of freedom. Thus the term in $q_V^\mu q_V^\nu$ removes the unphysical momentum-parallel component.

A massive spin-1 boson has three physical polarisation vectors $\varepsilon_\lambda^\mu(q_V)$, which satisfy the transversality condition $q_{V,\mu}\varepsilon_\lambda^\mu(q_V) = 0$ that implies that each physical polarisation is orthogonal to the boson four-momentum. Together with their normalisation, they form a complete basis in the subspace orthogonal to q_V^μ. Hence, summing over them gives the projector onto that subspace, which is exactly the left-hand side of Equation 6.118. The completeness relation is therefore a direct consequence of the transversality condition and the normalisation of the polarisation vectors. It is also a direct consequence of the gauge invariance of the theory, which ensures that only physical degrees of freedom propagate.

The propagator in Equation 6.33 can thus be rewritten using the NWA and the completeness relation of the polarisation vectors:

$$\frac{-\mathrm{i}}{q_V^2 - m_V^2 + \mathrm{i}m_V\Gamma_V}\left(g^{\mu\nu} - \frac{q_V^\mu q_V^\nu}{m_V^2}\right) = \frac{\mathrm{i}}{q_V^2 - m_V^2 + \mathrm{i}m_V\Gamma_V}\sum_{\lambda=0,\pm1}\varepsilon_\lambda^\mu(q_V)\varepsilon_\lambda^{*\nu}(q_V)\,, \tag{6.119}$$

which gives:

$$\mathcal{M} = \sum_\lambda \underbrace{\mathcal{M}^{\mu,\mathrm{prod}}\varepsilon_\lambda^\mu(q_V)}_{\mathcal{M}_\lambda^{\mathrm{prod}}}\frac{\mathrm{i}}{q_V^2 - m_V^2 + \mathrm{i}m_V\Gamma_V}\underbrace{\varepsilon_\lambda^{*\nu}(q_V)\mathcal{M}^{\nu,\mathrm{decay}}}_{\mathcal{M}_\lambda^{\mathrm{decay}}}\,. \tag{6.120}$$

Given Equation 1.108 and projecting the latter equation on the longitudinal polarisation state we get:

$$\mathcal{M}_{\lambda=0}^{\mathrm{prod}} = \varepsilon_0^\mu(q_V)\mathcal{M}^{\mu,\mathrm{prod}} = \frac{q_V^\mu}{m_V}\mathcal{M}_\mu^{\mathrm{prod}} + \mathcal{O}\left(\frac{m_V}{E}\right)\,. \tag{6.121}$$

Gauge invariance relates the contraction with the vector boson momentum to the amplitude involving the corresponding Goldstone boson. The generalisation of the Ward identity [397] ($q_V^\mu\mathcal{M}_\mu = 0$) to the non-Abelian case is given by the Slavnov–Taylor identities [398, 399], which yield:

$$q_V^\mu\mathcal{M}_\mu^{\mathrm{prod}}(V) \simeq m_V\mathcal{M}^{\mathrm{prod}}(\phi_V)\,, \tag{6.122}$$

so that:

$$\mathcal{M}_{\lambda=0}^{\mathrm{prod}} \simeq \mathcal{M}^{\mathrm{prod}}(\phi_V) + \mathcal{O}\left(\frac{m_V}{E}\right)\,. \tag{6.123}$$

This is the content of the GBET [3, 4]: at energies much larger than the vector boson mass, amplitudes involving longitudinally polarised massive vector bosons are equal, up to corrections of order $\frac{m_V}{E}$, to the corresponding amplitudes involving the associated Goldstone bosons.

This interpretation is well-defined for an on-shell vector boson, or more generally for the gauge-invariant residue of the vector boson pole. Away from the pole, $q_V^2 \neq m_V^2$, the exchanged vector boson is not a physical asymptotic state. The separation of the amplitude into transverse and longitudinal components is then not gauge invariant.

In particular, the longitudinal projection is sensitive to the q_V^μ-dependent part of the amplitude, which is tied by the Slavnov–Taylor identities to the Goldstone sector. Off shell, this contribution does not define a physical longitudinal vector boson amplitude by itself: gauge-dependent terms, Goldstone boson contributions, and non-resonant diagrams cancel only in the full amplitude. Therefore, away from the pole, a quantity such as $\mathcal{M}_{\lambda=0}^{\mathrm{prod}}$ or a longitudinal polarisation fraction cannot be interpreted as an observable independently of the prescription used to define it. A gauge-invariant definition is obtained by projecting the pole residue, as done in the NWA or DPA.

For an on-shell massive vector boson, the polarisation states are well-defined within the NWA and can be projected out in a gauge-invariant way. In WZ production, the two bosons are produced simultaneously after their interactions and can be entangled in their spin states. The DPA is needed to define the joint helicity states of the WZ system in a gauge-invariant way, and to ensure that the polarisation decomposition is self-consistent. The production amplitude can thus be projected onto the helicity basis of the W and Z, respectively, yielding two transverse states for each boson ($\lambda = \pm1$) and one longitudinal state ($\lambda = 0$) therefore giving rise to four possible joint helicity states:

$$\mathrm{W}_0\mathrm{Z}_0\,, \qquad \mathrm{W}_0\mathrm{Z}_\mathrm{T}\,, \qquad \mathrm{W}_\mathrm{T}\mathrm{Z}_0\,, \qquad \mathrm{W}_\mathrm{T}\mathrm{Z}_\mathrm{T}\,, \tag{6.124}$$

each associated with one of the four polarisation fractions defined in Equation 6.91. The full amplitude can be decomposed into the sum of the contributions from each of the four helicity combinations:

$$\mathcal{M} = \mathcal{M}_{00} + \mathcal{M}_{0\mathrm{T}} + \mathcal{M}_{\mathrm{T}0} + \mathcal{M}_{\mathrm{TT}}\,. \tag{6.125}$$

This factorisation, however, is only meaningful for the doubly resonant topologies as explained above, which means both bosons are put on-shell. The physical state also receives singly resonant and interference contributions as in Equation 6.107, for which no polarisation can be defined and which are individually gauge dependent: they cannot simply be discarded without breaking gauge invariance. The *full* unpolarised cross-section is then not simply the incoherent sum of the four polarised contributions, but also contains interference terms:

$$|\mathcal{M}|^2 = \sum_{\lambda} |\mathcal{M}_\lambda|^2 + \sum_{\lambda \neq \lambda'} 2\mathrm{Re}\left(\mathcal{M}_\lambda \mathcal{M}^*_{\lambda'}\right) . \tag{6.126}$$

The cross-term contributes to the off-diagonal elements of the per-event SDM in Equation 6.89, which encodes the interference between different helicity states of the two bosons.

A consequence is that the sum of separately generated polarised templates does not necessarily reproduce the full unpolarised prediction unless the interference terms are negligible or treated consistently. Gauge invariance also plays a key role, as one cannot select a subset of diagrams and assign polarisation states naively. Hence the individual diagrams are not gauge-invariant objects: as seen in Section 1.11 and in more detail in Appendix A, the Higgs-exchange diagram is needed to ensure the gauge invariance of the amplitude and the unitarity of the cross-section at high energy, and it contributes to the longitudinal polarisation state. The interference between the Higgs-mediated and non-Higgs-mediated diagrams is therefore an intrinsic part of the longitudinal polarisation component, and cannot be neglected without breaking gauge invariance. In the process considered the interference is found to be at the $1\,\%$ to $2\,\%$ level as visible in Table 6.A.

6.5 Monte Carlo generation of polarised samples

Built upon the formalism described in the previous section, the generation of polarised MC samples is a challenging task that is done in three steps:

1. Generation of all Feynman diagrams at the appropriate coupling order
2. Numerical evaluation of the colour- and spin-summed squared ME at each phase-space point
3. Integration over phase-space with the PDF convolution given in Equation 4.60.

This task is performed with current MC generators at LO. More details on these specificities and how to go beyond will be provided in the current section.

6.5.1 Leading-order generation

A *polarised sample* of type $(\lambda_\mathrm{W}\lambda_\mathrm{Z})$ is produced by retaining *only* the corresponding term in the sum of Equation 6.126. The event weight associated is given by:

$$w_{\lambda_\mathrm{W}\lambda_\mathrm{Z}} = \left|\mathcal{M}^{\mathrm{prod}}_{\lambda_\mathrm{W}\lambda_\mathrm{Z}}\right|^2 \cdot \left|\mathcal{M}^{\mathrm{decay}}_{\lambda_\mathrm{W}}(\Omega_\mathrm{W})\right|^2 \cdot \left|\mathcal{M}^{\mathrm{decay}}_{\lambda_\mathrm{Z}}(\Omega_\mathrm{Z})\right|^2 \cdot \mathrm{d}\,\Phi\,, \tag{6.127}$$

where $\mathrm{d}\,\Phi$ is the LIPS weight of a given event. The event weight for the interference contribution on the other hand is:

$$w_{\mathrm{interf}} = 2\,\mathrm{Re}\left(\mathcal{M}_0\,\mathcal{M}^*_\mathrm{T}\right) \cdot \mathrm{d}\,\Phi\,, \tag{6.128}$$

which can be negative.

The two state-of-the-art MC generators for polarised samples are MadGraph and Sherpa, which implement the polarisation in two complementary strategies following the formalism defined in Section 6.4 for the gauge invariance. The difference arises in how the polarisation projection is applied. This two-way technique allows cross-checks between produced samples.

MadGraph: projection *a priori*. MadGraph [400] isolates the polarised amplitude at the ME level via the DPA, before the phase-space integration. For a sample labelled $(\lambda_\mathrm{W}\lambda_\mathrm{Z})$ the spin sum over the intermediate W (Z) is truncated to the single helicity λ_W (λ_Z) by retaining only the corresponding polarisation vector $\varepsilon^\mu_{\lambda_\mathrm{W}}$ ($\varepsilon^\mu_{\lambda_\mathrm{Z}}$) in Equation 6.118. The off-shell lineshape is kept and the decays are part of the ME with the full spin-correlation information, so that the lepton angular distribution in the boson rest frame correctly reflects the helicity state. The non-resonant and singly resonant topologies (for which no polarisation is defined) are removed at generation level. Each of the four joint states $\mathrm{W}_0\mathrm{Z}_0$, $\mathrm{W}_0\mathrm{Z}_\mathrm{T}$, $\mathrm{W}_\mathrm{T}\mathrm{Z}_0$, and $\mathrm{W}_\mathrm{T}\mathrm{Z}_\mathrm{T}$ is generated as a separate sample. The interference

is obtained from additional runs with mixed polarisation states with a signed weight corresponding to the real part of the off-diagonal SDM elements. The polarisation is fixed at generation and must be identical across all samples.

SHERPA: reweighting *a posteriori*. SHERPA [401, 402] instead runs an effectively unpolarised simulation and computes the polarisation information as per-event weights. For each event, the production amplitude is recomputed with open helicity indices, giving the full per-event SDM of Equation 6.89. The spin-correlated decays are generated with the Richardson algorithm [403], and the intermediate boson masses are smeared according to the Breit–Wigner distribution of Equation 6.106 to restore the off-shell lineshape on top of the on-shell projection. The polarised weight for state $(\lambda_{\mathrm{W}}\lambda_{\mathrm{Z}})$ is the diagonal element $\rho^{\mathrm{WZ}}_{\lambda_{\mathrm{W}}\lambda_{\mathrm{Z}},\lambda_{\mathrm{W}}\lambda_{\mathrm{Z}}}$ and the interference weights are the real parts of the off-diagonal elements. The decisive practical advantage is that all four polarisation combinations, the interferences, and several reference-frame definitions are obtained simultaneously from a single run, with negative-weight interference events handled natively and the polarisation frame chosen *a posteriori*.

6.5.1.1 Additional jets

The pure Born final state $\mathrm{pp} \to \mathrm{WZjj}$ already contains two tagging jets, so only two partons are required at ME level: $\mathrm{WZjj} + 0\mathrm{j}$. To describe additional hard radiation, MEs with one extra parton ($+1\mathrm{j}$) are also generated and combined with the parton shower through multijet merging, the shower resumming the softer and collinear emissions. This results in a more realistic description of the final state, with up to three jets at ME level and additional soft and collinear radiation from the shower. The polarisation information is retained for all partonic multiplicities, so that the polarised templates are available for both the $+0\mathrm{j}$ and $+1\mathrm{j}$ merged samples. This merging is only used for the SHERPA samples, with the MEPS@LO prescription [254, 255]. The polarised MADGRAPH5_AMC@NLO samples are generated at LO for $\mathrm{WZjj} + 0\mathrm{j}$ only, without any additional parton and therefore without merging: the additional radiation comes only from the PYTHIA shower.

6.5.1.2 Automated one-loop

The virtual contribution to any NLO cross-section is the interference of the one-loop amplitude with the Born amplitude, corresponding to the term V_n of the NLO formula in Equation 4.61. For a final state of this multiplicity the number of one-loop diagrams per partonic channel reaches $\mathcal{O}(10^4)$, so the amplitudes are generated automatically by dedicated one-loop providers rather than by hand. Two ingredients are involved. First, the loop integrand is built by numerical recursion:

- off-shell current recursion in RECOLA
- open-loop recursion in OPENLOOPS
- or an equivalent construction in MADLOOP

which avoids the factorial growth of a diagram-by-diagram evaluation. Second, the resulting tensor integrals are reduced to a basis of scalar integrals. RECOLA and OPENLOOPS were designed to deliver the full SM EW corrections, not only the QCD ones.

6.5.1.3 Limitations of leading order

Both generators provide the polarised VBS signal only at LO (optionally matched to the parton shower). The NLO QCD corrections to the fractions arise predominantly from real emission rather than from virtual diagrams, and SHERPA includes these real QCD corrections through its MEPS@LO multijet merging [254, 255], but the virtual corrections and the ultra-soft and ultra-collinear emissions are still neglected, so the accuracy on the fractions themselves remains effectively LO. Neither generator provides NLO EW corrections for polarised VBS. Going beyond LO faces three genuine obstacles:

Gauge invariance. Individual helicity projections are not automatically gauge invariant at loop level. The DPA on-shell projection introduces ambiguities at $\mathcal{O}(\alpha_{\mathrm{S}})$ once virtual corrections connect diagrams that are only separately gauge invariant within the approximation.

Real-emission kinematics. The NLO real-emission process $\mathrm{pp} \to \mathrm{WZjj} + \mathrm{g}$ alters the boson recoil and makes the helicity axis, and the on-shell projection, more delicate when an additional hard parton modifies the event kinematics.

Infrared cancellations. Helicity-separated virtual and real contributions do not individually cancel infrared singularities. The subtraction must be performed on the full amplitude before any polarisation projection is applied, thus preventing a straightforward decomposition of NLO corrections into polarised components.

6.5.2 NLO QCD corrections

The NLO QCD corrections to $\mathrm{WZjj\text{-}EW}$ are of order $\mathcal{O}(\alpha_\mathrm{S}\,\alpha_\mathrm{EW}^6)$ [404–407] and consist of two separately infrared-divergent pieces that cancel in the sum:

Virtual corrections. One-loop diagrams with a single QCD gluon running in the loop, attached to the quark lines. These produce UV divergences (renormalised by the running of α_S) and infrared divergences, the latter encoded in soft and collinear poles in dimensional regularisation.

Real corrections. Tree-level emission of an additional gluon (or a quark–antiquark pair from a gluon), $\mathrm{pp} \to \ell'\nu_\ell\ell\ell\mathrm{jjj}$, which produces infrared-divergent phase-space regions when the extra parton becomes soft or collinear.

NLO QCD corrections are visible in Figure 6.16 as virtual loop corrections or real external emission.

The cancellation is organised with a local subtraction scheme, CS dipole subtraction or FKS, both described in Section 4.6. For $\mathrm{WZjj\text{-}EW}$, the QCD K-factor is relatively small:

$$K_{\mathrm{WZjj\text{-}EW}}^{\mathrm{QCD}} = \frac{\sigma^{\mathrm{NLO\,QCD}}}{\sigma^{\mathrm{LO}}} \sim 1.0\text{–}1.05\,, \tag{6.129}$$

because the dominant production mechanism is t-channel colour-singlet exchange: there is no colour flow between the two quark lines at LO, so QCD radiation off the two sides is largely independent, and the net correction is suppressed, in contrast to the s-channel processes, where coloured particles exchange colour in the hard scattering. The scale uncertainty is correspondingly reduced from LO to NLO QCD.

For the QCD-induced background $\mathrm{WZjj\text{-}QCD}$, of order $\mathcal{O}(\alpha_\mathrm{S}^3\,\alpha_\mathrm{EW}^4)$ at NLO, the corrections are much larger, $K_{\mathrm{WZjj\text{-}QCD}}^{\mathrm{QCD}} \sim 1.5\text{–}2.0$, reflecting the s-channel colour flow between the incoming quarks and the diboson system. A precise NLO prediction for this dominant background is essential to avoid mis-modelling.

A useful feature, established for same-sign $\mathrm{WWjj\text{-}EW}$ [408, 409] scattering and expected to hold here, is that the NLO QCD corrections are considerably more independent of the polarisation state than the NLO EW ones. Applying differential QCD K-factors obtained from the unpolarised off-shell calculation to the polarised NLO EW results is therefore a reasonable approximation. The exact $\mathcal{O}(\alpha_\mathrm{S}\,\alpha_\mathrm{EW}^6)$ corrections to *polarised* WZ scattering have not yet been computed.

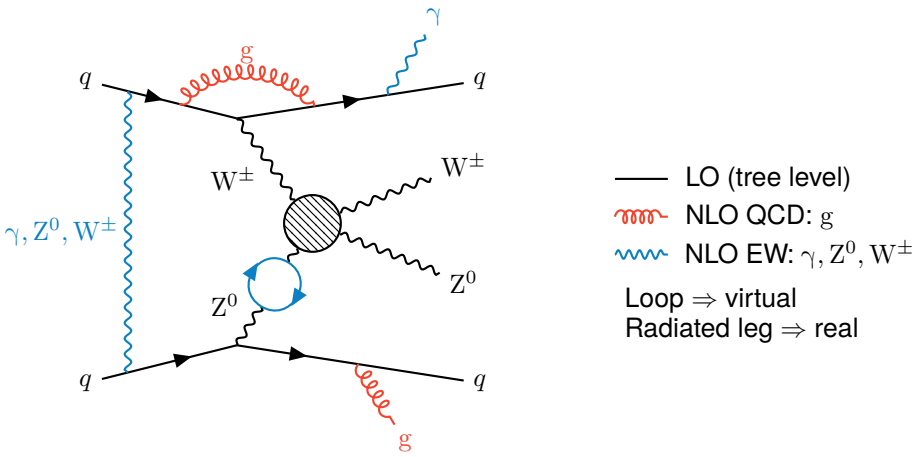


Figure 6.16: Feynman diagram of the WZjj process through VBS. The NLO effects from QCD and EW corrections are shown as virtual loop or real external emission.

6.5.3 NLO EW corrections

The NLO EW corrections to $\mathrm{WZjj\text{-}EW}$ are of order $\mathcal{O}(\alpha_\mathrm{EW}^7)$ and are dominated by EW *Sudakov logarithms* [410–412]. In the high-energy regime, $Q^2 \gg m_\mathrm{V}^2$, a particle can emit a soft or collinear gauge boson with large probability, and the integral over the virtual gauge-boson momentum then contains large

logarithms of the form $\ln\left(\frac{Q^2}{m_V^2}\right)$. This is most transparent for the eikonal integral I_{ij}^{eikonal} that contributes to the virtual term V_n of Equation 4.61, for a gauge boson of mass m_V exchanged between two hard, light-like external legs with momenta p_i and p_j:

$$I_{ij}^{\text{eikonal}} = \int \frac{\mathrm{d}^4 q}{(2\pi)^4} \frac{p_i \cdot p_j}{(q^2 - m_V^2)((p_i \cdot q)(p_j \cdot q))} \,. \tag{6.130}$$

Using a Sudakov parametrisation of the loop momentum q in terms of the external momenta p_i and p_j with the form $q = \alpha p_i + \beta p_j + q_\perp$, where $q_\perp$ is the component of q orthogonal to both p_i and p_j, the integral can be rewritten as:

$$I_{ij}^{\text{eikonal}} = \int \frac{\mathrm{d}\,\alpha}{\alpha} \frac{\mathrm{d}\,\beta}{\beta} \frac{\mathrm{d}^2 q_\perp}{\alpha\beta s_{ij} - q_\perp^2 - m_V^2} \,, \tag{6.131}$$

where $s_{ij} = 2p_i \cdot p_j$. From this we obtain two regimes:

Soft regime. $\alpha, \beta \to 0$ together, the boson energy tends to zero.

Collinear regime. $\beta \to 0$ ($\alpha \to 0$), the boson is emitted collinear to the external leg p_i (p_j).

In both cases the integral is logarithmically divergent, and the divergence is regulated by the boson mass m_V, which acts as an infrared physical cutoff. A cutoff below $\alpha, \beta \sim \frac{m_V^2}{s_{ij}}$ gives a logarithmic contribution $\ln\left(\frac{s_{ij}}{m_V^2}\right)$, which is the origin of the Sudakov logarithms. The double logarithm arises from the overlap of the soft and collinear regions, where both α and β are small. The leading double logarithm is therefore proportional to $\ln^2\left(\frac{s_{ij}}{m_V^2}\right)$, while the subleading single logarithm is proportional to $\ln\left(\frac{s_{ij}}{m_V^2}\right)$.

The full NLO EW correction can be organised, in the high-energy limit, as:

$$\delta_{\text{full}}^{\text{EW}} = a \ln^2\left(\frac{Q^2}{m_V^2}\right) + b \ln\left(\frac{Q^2}{m_V^2}\right) + c + \mathcal{O}\left(\frac{m_V^2}{Q^2}\right) \,, \tag{6.132}$$

where a and b are the coefficients of the leading double and the subleading single logarithm, and c is the finite, non-logarithmic remainder. Appendix D gives an explicit breakdown of the one-loop EW corrections to WZjj-EW following [410, 411], which determines the first two terms of Equation 6.132 exactly. The double-logarithmic coefficient a is universal: it is a sum over the external legs fixed by their gauge Casimirs alone, independent of the kinematics and of the Born coupling structure. The single-logarithmic coefficient b is more involved, encoding in addition the scattering kinematics (through the angular logarithms of δ^{SSC}) and the coupling structure of the Born amplitude (through δ^{PR}). The remainder c is process-dependent and accessible only from the full numerical calculation.

This formalism applies to any process whose Born amplitude is not mass-suppressed, and returns the logarithmically enhanced part of its one-loop EW corrections: a truncation that keeps only the energy-growing piece. For the polarised WZjj-EW process it provides a first estimate of the size of the EW corrections and, above all, of how they interplay with the polarisation states, capturing the hierarchy between the modes. Obtaining the exact result, all terms of Equation 6.132, instead requires evaluating the complete one-loop amplitude numerically. The approximation $\delta_{\text{DP}}^{\text{EW}}$ of Equation D.1 thus omits the finite remainder c and the mass-suppressed terms, and does not include the real photon radiation $\text{pp} \to \ell'\nu_\ell\ell\ell\text{jj} + \gamma$ or the photon-induced channels that complete the full NLO EW correction: all of these are supplied by the calculation of [413].

NLO EW corrections are visible in Figure 6.16 as virtual loop corrections or real external emission.

The size of EW corrections depends on the renormalisation of the EM coupling. For this charged current (CC), W-mediated process the G_μ scheme [412] is the natural choice, deriving α from the Fermi constant (more details given in Appendix C in Section C.4):

$$\alpha_{G_\mu} = \frac{\sqrt{2}}{\pi} G_{\text{F}} m_{\text{W}}^2 \left(1 - \frac{m_{\text{W}}^2}{m_{\text{Z}}^2}\right) \,, \tag{6.133}$$

which absorbs the running of α up to the EW scale. The unstable bosons are treated in the complex-mass scheme [414], $\mu_V^2 = m_V^2 - \mathrm{i} m_V \Gamma_V$, which preserves the Ward and Slavnov–Taylor identities at one loop and is essential when several propagators are simultaneously near resonance. Consistently with the DPA, the polarised results are evaluated with real couplings.

For a realistic Run-3 setup at $\sqrt{s} = 13.6\,\text{TeV}$, the integrated NLO EW correction to the unpolarised signal is:

$$\delta_{\text{unpol.}}^{\text{EW}} \approx -18.5\,\% \,, \tag{6.134}$$

as obtained in [413], of the same size as the $-16\,\%$ found in the earlier off-shell calculation at $13\,\mathrm{TeV}$ [407]. The corrections for the individual doubly polarised signals follow the Casimir hierarchy of Equation D.3: they are largest for the purely transverse mode and smallest for the purely longitudinal one. The complete set of integrated cross-sections and corrections from [413] is collected in Table 6.A.

These corrections were computed by the authors of [413], with whom the DNN-based observable of this analysis was implemented (Section 7.2.2.6). The calculation is performed with the MC program MoCaNLO [415], which obtains the tree-level and one-loop amplitudes from Recola [271] with the tensor-integral reduction of Collier [416], the unstable-boson propagators being treated in the complex-mass scheme. As an independent validation, the authors of [413] recompute the same observables with a second MC framework, BBMC, using the same amplitude input. The agreement between the two provides a stringent check of the dipole-subtraction and on-shell-projection machinery specific to the doubly polarised DPA [413].

Table 6.A: Integrated cross-sections for $\mathrm{W^+Z}$ scattering at the LHC at $\sqrt{s} = 13.6\,\mathrm{TeV}$. The polarisation is defined in the modified-helicity frame. The "full" row is the full off-shell result and the interference is the one *between polarisation states*, defined by subtracting the polarised sum from the unpolarised DPA cross-section. Two normalisation conventions are given for the fractions f. In the first one they are divided by the unpolarised DPA cross-section, so that the four doubly polarised modes and the interference add up to unity. In the second one they are divided by the sum $\sum_{\mathrm{pol.}}$ of the four doubly polarised modes alone. The two conventions differ by the $1.7\,\%$ to $1.8\,\%$ of interference. The two singly polarised modes are the sums $\mathrm{W_0Z_X} = \mathrm{W_0Z_0} + \mathrm{W_0Z_T}$ and $\mathrm{W_XZ_0} = \mathrm{W_0Z_0} + \mathrm{W_TZ_0}$. Cross-sections and EW corrections from [413].

Mode	σ^{LO} [fb]	δ^{EW} [%]	$\sigma^{\mathrm{NLO\,EW}}$ [fb]	f [%] (σ^{DPA})		f [%] ($\sum_{\mathrm{pol.}}$)	
				LO	NLO EW	LO	NLO EW
Full	0.272 08	−18.5	0.2216	101.6	101.5	103.4	103.4
Unpolarised	0.267 77	−18.5	0.2183	100.0	100.0	101.8	101.8
$\mathrm{W_0Z_0}$	0.023 045	−14.5	0.0197	8.6	9.0	8.8	9.2
$\mathrm{W_0Z_T}$	0.049 365	−17.9	0.0405	18.4	18.6	18.8	18.9
$\mathrm{W_TZ_0}$	0.049 023	−17.9	0.0402	18.3	18.4	18.6	18.8
$\mathrm{W_TZ_T}$	0.141 725	−19.6	0.1140	52.9	52.2	53.9	53.1
Interference	0.004 62	−15.0	0.0039	1.7	1.8	1.8	1.8
$\mathrm{W_0Z_X}$	0.072 410	−16.8	0.0602	27.0	27.6	27.5	28.1
$\mathrm{W_XZ_0}$	0.072 068	−16.8	0.0599	26.9	27.5	27.4	28.0

The most important feature is that the EW K-factor is *not* the same for all polarisation states:

$$K^{\mathrm{EW}}_{\mathrm{W_0Z_0}} \neq K^{\mathrm{EW}}_{\mathrm{W_0Z_T}} \approx K^{\mathrm{EW}}_{\mathrm{W_TZ_0}} \neq K^{\mathrm{EW}}_{\mathrm{W_TZ_T}}\,, \tag{6.135}$$

with the explicit ordering $\left|\delta^{\mathrm{EW}}_{00}\right| < \left|\delta^{\mathrm{EW}}_{0\mathrm{T},\mathrm{T}0}\right| < \left|\delta^{\mathrm{EW}}_{\mathrm{TT}}\right|$, of $-14.5\,\%$, $-17.9\,\%$, and $-19.6\,\%$. This is a direct consequence of the different EW Casimir coefficients of Equation D.3: the longitudinal amplitude inherits the smaller Casimir of the Goldstone/Higgs sector, the transverse amplitude the larger adjoint Casimir. The same hierarchy was found for polarised same-sign WWjj-EW scattering [408]. At the differential level the splitting grows with energy: in the tails of the transverse momentum of the tagging jet system, the longitudinal mode reaches about $-30\,\%$ at $p_{\mathrm{T}}^{\mathrm{jj}} \approx 800\,\mathrm{GeV}$ while all other modes reach about $-40\,\%$, and in the most asymmetric tagging jet configurations the corrections reach up to $-50\,\%$. However, the integrated polarisation fractions are only mildly shifted between LO and NLO EW.

6.5.4 NLO EW calculation procedure

Extending the DPA of Equation 6.117 to one loop is conceptually straightforward but requires several technical precautions, which we summarise following [408, 413].

On-shell projection at NLO. All MEs entering the calculation, virtual one-loop as well as real-emission tree-level with one extra particle, are evaluated after the numerator replacement $q_V \to \tilde{q}_V$ as done for Equation 6.117, the off-shell kinematics being retained in the LIPS measure and in the propagator denominators. The projection must be applied in the *same* reference frame across all contributions, virtual and real, to avoid spurious frame-dependent terms.

Helicity projection. Once the bosons are on-shell, the physical polarisation sum of Equation 6.118, is split into longitudinal and transverse contributions and each polarised amplitude is extracted by inserting the appropriate $\varepsilon^{\mu}_{\lambda}(\tilde{q}_V)$. This must be done consistently for the loop integrands (evaluated with $\tilde{q}_V$) and the real-emission amplitudes.

Infrared subtraction in the DPA. The soft and collinear singularities of the real-emission contribution are removed with the CS dipole subtraction of Section 4.6. The singular regions must be carefully separated into those associated with the production and with the decay of the bosons, and the dipole counterterms must be mapped onto the DPA kinematics without disrupting the on-shell projection. Photon radiation off a charged resonance contributes to both production and decay and requires additional, dedicated subtraction counterterms straddling the production–decay boundary.

Triply resonant topologies. The VBS signal overlaps with triple-boson production, which is suppressed but not removed by the selection cuts. Setting the EW-boson widths to zero everywhere except in the two target propagators, as the DPA requires for gauge invariance, leaves the propagator of a possible third s-channel boson unprotected, producing integrable-in-principle but numerically dangerous large-weight events. Rather than restoring the finite width in the DPA ME (which would reintroduce a gauge dependence and, for radiating resonances, spoil the infrared subtraction), the authors of [408, 413] multiply the real ME and the corresponding local and integrated counterterms by a gauge-invariant *fudge factor*. Close to the pole the factor effectively restores a finite-width propagator. Because it multiplies the complete gauge-invariant amplitude it introduces no gauge dependence, and applying it identically to real and subtraction terms preserves the infrared cancellation. The associated error is of order $\frac{\Gamma_V}{m_V}$ and restricted to a small region of phase-space.

QED radiation and the lepton definition. The NLO EW real corrections include photon emission, $\mathrm{pp} \to \ell'\nu_\ell\ell\ell\mathrm{jj} + \gamma$. Photons collinear to a charged lepton generate a mass singularity whose treatment depends on the observable lepton definition: with *dressed* leptons the photon is recombined with the lepton when it falls within a cone of radius $R = 0.1$, cancelling the collinear logarithm and reducing the size of the corrections. The photon-induced (γq, $\gamma\gamma$) channels enter at the same order $\mathcal{O}(\alpha^7_{\mathrm{EW}})$ and contribute only at the one-percent level to the fiducial cross-section. Because they enhance the triple-boson topologies they are regarded as part of the irreducible triple-boson background rather than as part of the VBS signal.

6.5.5 Impact of NLO predictions

Several independent arguments make NLO predictions indispensable for the polarisation measurement.

Normalisation of the signal. The polarised cross-sections enter the template fit directly through the expected yield per polarisation category as done in Equation 6.116. An LO normalisation carries a $\pm 5\,\%$ to $10\,\%$ scale uncertainty on the $\mathrm{WZjj\text{-}EW}$ rate, which propagates to all extracted fractions. The NLO QCD K-factor of Equation 6.129 reduces this scale uncertainty while shifting the central value only mildly, but the $-18.5\,\%$ NLO EW correction of Equation 6.134 changes the expected yield by a non-negligible amount.

Shape distortions in the discriminant distributions. Kinematic observables have tails that are precisely where the EW Sudakov logarithms of Equation D.4 grow. These logarithms distort the high-m_{WZ}, high-$p_{\mathrm{T}}^{\mathrm{W}}$, high-$p_{\mathrm{T}}^{\mathrm{Z}}$, and high-$m_{\mathrm{jj}}$ regions differently for each polarisation state [417]. Considering only LO effects and ignoring the NLO EW distributions would bias the extracted fractions.

Helicity-dependent corrections and the longitudinal fraction. As quantified in Equation 6.135 and Table 6.A, the EW K-factor differs between polarisation states. The doubly longitudinal fraction f_{00} is the most interesting physically, and although its integrated value is only mildly shifted, it is the most strongly reshaped relative to the transverse modes. The distributions most sensitive to it are the reconstructed decay angle $\cos\theta^*$ of Equation 6.68, the rapidity separation A_y^{WZ} (which in VBS favours large separations for the longitudinal mode, opposite to inclusive WZ production), and the tagging-jet transverse-momentum asymmetry.

Control of the QCD background. The QCD-induced $\mathrm{WZjj\text{-}QCD}$ background has a large NLO K-factor and dominates the SR. At LO its polarisation composition is predominantly transverse, since the QCD production mechanism favours helicity-conserving quark–boson vertices. NLO corrections modify both its rate and its polarisation composition, so an accurate NLO QCD prediction is required to avoid mis-modelling the template distributions.

6.5.6 State-of-the-art computation

The most up-to-date fixed-order predictions for WZ scattering are:

- NLO QCD for the unpolarised signal, matched to the parton shower at NLO QCD [405, 407]
- NLO EW for the unpolarised signal, both differential and integrated, in the full off-shell description [407, 417], the precursor large-EW-correction result for VBS having been established for same-sign WWjj-EW scattering
- LO polarised signals in the pole approximation, matched to parton shower, in PHANTOM [418] and MADGRAPH
- NLO EW for the *doubly polarised* signal in the DPA, the first such calculation for WZ scattering [413], following the same machinery developed for polarised same-sign WWjj-EW scattering [408].

In this analysis, the polarised signal templates are taken from the LO generators: SHERPA 3.0, whose MEPS@LO merging of the $+0\mathrm{j}$ and $+1\mathrm{j}$ MEs includes the real-emission part of the NLO QCD corrections, is the nominal generator, and MADGRAPH, at pure LO, is the alternative one (Section F.1.2). The NLO EW corrections are taken from the DPA calculation of [413] performed with MOCANLO. They are not applied to the nominal templates: they are propagated from parton to particle level and enter the fit as a one-sided systematic uncertainty on the template shapes (Section 7.2.2.6). The MOCANLO predictions of the polarisation fractions are also used as SM references when comparing with the measured values in Chapter 7.

6.6 Analysis setup

Building on the history of past measurements, the current state-of-the-art theoretical predictions, and the particular topology of the WZjj final state discussed above, we can now define the analysis setup to enhance the sensitivity to the longitudinal polarisation and to extract the polarisation fractions from data. The analysis setup is defined in terms of the fiducial phase-space, the event selection, and the background treatment. The following sections will detail each of these components, providing a comprehensive framework for the measurement of WZ scattering polarisation at the LHC with the ATLAS detector.

The measurement is performed with $140\,\mathrm{fb}^{-1}$ recorded at $13\,\mathrm{TeV}$ during Run-2 (2015–2018) and $164\,\mathrm{fb}^{-1}$ collected at $13.6\,\mathrm{TeV}$ during the start of Run-3 (2022–2024), for a total of $304\,\mathrm{fb}^{-1}$.

6.6.1 Fiducial phase-space

The fiducial phase-space is a set of kinematic and geometric requirements applied to the final-state particles in order to define a region of the total detector acceptance where the measurement is performed. The fiducial phase-space is designed to enhance the sensitivity to the signal while maintaining a low background contamination.

The fiducial phase-space defined for the measurement presented in this thesis is based on the combination of the one used for the inclusive WZ production measurement at $\sqrt{s} = 8\,\mathrm{TeV}$ [325] and $13\,\mathrm{TeV}$ [338] and the one used for the WZjj production measurement at $\sqrt{s} = 13\,\mathrm{TeV}$ [360]. The phase-space of the WZjj measurement is a subset of that of the inclusive WZ measurement, the former being more restrictive to account for the tagging jets in the final state. The inclusive part of the phase-space defined here is also the one used in the inclusive WZ polarisation measurement at $\sqrt{s} = 13\,\mathrm{TeV}$ [10, 11] and the early inclusive WZ cross-section measurement at $\sqrt{s} = 13\,\mathrm{TeV}$ [331].

It is defined at particle level, that is on the stable particles of the simulated event after parton shower, hadronisation, and decays, and not on the partons of the hard process. It relies on final-state prompt leptons [419] (a lepton that is not produced in the decay of a hadron or a τ or their descendants, as illustrated in Figure 5.5) associated with a W and a Z boson decay.

The fiducial phase-space is defined by the following requirements detailed in Table 6.B. These requirements reflect the particular topology of the WZjj final state already discussed in Section 6.2, such as the requirement of two jets in opposite hemispheres of the detector with a large invariant mass above $500\,\mathrm{GeV}$ and a large rapidity separation. Additional jets are also taken into account arising from ISR or FSR. Hard processes with a b-quark in the initial or final states at the ME level are not considered as part of the WZjj-EW signal.

The phase-space definition uses dressed kinematics of the final-state charged leptons. Dressed leptons are obtained from the bare leptons, that is after QED FSR, by adding the four-momenta of all photons within a cone of $\Delta R < 0.1$ around the bare-lepton direction. They are designed to be as close as possible to what is measured in the electron channel, since the ATLAS cluster reconstruction combines

Table 6.B: Phase-space requirements for the fiducial region of the WZjj polarisation measurement.

Variable	Requirement
Inclusive	
Lepton $\lvert\eta\rvert$	< 2.5
p_{T} of ℓ_{Z}	$> 15\,\mathrm{GeV}$
p_{T} of ℓ_{W}	$> 20\,\mathrm{GeV}$
m_{Z} range	$\lvert m_{\mathrm{Z}} - m_{\mathrm{Z}}^{\mathrm{PDG}}\rvert < 10\,\mathrm{GeV}$
$m_{\mathrm{T}}^{\mathrm{W}}$	$> 30\,\mathrm{GeV}$
$\Delta R(\ell_{\mathrm{Z}_1}, \ell_{\mathrm{Z}_2})$	> 0.2
$\Delta R(\ell_{\mathrm{Z}}, \ell_{\mathrm{W}})$	> 0.3
VBS	
p_{T} of two leading jets	$> 40\,\mathrm{GeV}$
$\lvert\eta\rvert$ of two leading jets	< 4.5
N_{jets}	≥ 2
$\eta^{\mathrm{j}_1} \cdot \eta^{\mathrm{j}_2}$	< 0
m_{jj}	$> 500\,\mathrm{GeV}$
$\Delta R(\mathrm{j}, \ell)$	> 0.3

the bare electron energy with that of close-by photons. Because the majority of QED FSR photons are collinear, and thus close to the EM cluster or to the track of the bare lepton, the amount of remaining QED FSR correction is reduced if dressed leptons are used. Therefore, the use of dressed leptons allows for the combination of channels with leptons radiating differently, namely electron and muon channels. These dressed leptons and the final-state neutrinos that do not originate from hadron or τ decays are associated with the W and Z boson decay products with an algorithmic approach, called "resonant shape", as used and defined for previous WZ measurements [325, 338]. This algorithm is based on the value of an estimator expressing the product of the nominal lineshapes of the W and Z boson resonances:

$$P = \left| \frac{1}{m^2_{(\ell_{\mathrm{Z}_1},\ell_{\mathrm{Z}_2})} - \left(m_{\mathrm{Z}}^{\mathrm{PDG}}\right)^2 + \mathrm{i}\Gamma_{\mathrm{Z}}^{\mathrm{PDG}} m_{\mathrm{Z}}^{\mathrm{PDG}}} \right|^2 \times \left| \frac{1}{m^2_{(\ell_{\mathrm{W}},\nu_\ell)} - \left(m_{\mathrm{W}}^{\mathrm{PDG}}\right)^2 + \mathrm{i}\Gamma_{\mathrm{W}}^{\mathrm{PDG}} m_{\mathrm{W}}^{\mathrm{PDG}}} \right|^2 , \tag{6.136}$$

where $m_{\mathrm{Z}}^{\mathrm{PDG}}$, $m_{\mathrm{W}}^{\mathrm{PDG}}$, $\Gamma_{\mathrm{Z}}^{\mathrm{PDG}}$, and $\Gamma_{\mathrm{W}}^{\mathrm{PDG}}$ are the world average values of the Z and W boson masses and widths reported by the PDG [34]. The input to the estimator is the invariant mass of all possible $(\ell_{\mathrm{Z}_1}, \ell_{\mathrm{Z}_2})$ and $(\ell_{\mathrm{W}}, \nu_\ell)$ pairs satisfying the fiducial requirements of Table 6.B. The final choice of which leptons are assigned to the W or Z bosons corresponds to the configuration exhibiting the highest value of the estimator. Using this specific association algorithm, the gauge boson kinematics can be computed using the kinematics of the associated leptons independently of any internal MC generator details. The "resonant shape" approach was compared to other assignment algorithms, and provided the best agreement, within $0.5\,\%$, with the built-in algorithm.

Particle-level jets are reconstructed from stable particles with a lifetime of $\tau > 30\,\mathrm{ps}$ ($c\tau > 9\,\mathrm{mm}$) in the simulation after parton shower, hadronisation, and decay of particles with $\tau < 30\,\mathrm{ps}$. Muons, electrons, neutrinos, and photons associated with the W and Z decays are excluded from the jet reconstruction. If the angular distance between a selected lepton and a jet is below 0.3, the jet is discarded.

The fiducial phase-space is identical for Run-2 and Run-3. The object and event selections differ.

6.6.2 Object and event selection

The object and event selection is designed to select events in the fiducial phase-space defined in Table 6.B. It consists of a set of requirements applied to the reconstructed objects in the event, such as leptons and jets, entering the fiducial phase-space. Among these, the various tools from the Combined Performance group are used to identify, calibrate, correct via the SFs, and select the objects depending on the WPs. In Chapter 5 and Section 6.2.1 we have already described such tools for the electron, muon, jet, and MET objects. The following sections will describe the object selection and event selection used in this analysis.

Adapted triggers are used to select events. In all four lepton-flavour channels, events are required to fire one of the single-lepton (electron or muon) triggers from the trigger menu of the corresponding data-taking year. In order to stay on the plateau of the trigger efficiency, the transverse momentum of the reconstructed leading lepton, matched to the fired trigger element, is required to be greater than $27\,\mathrm{GeV}$

for every year except for 2015, where the threshold is $25\,\mathrm{GeV}$ (a lower trigger threshold was possible due to the lower pile-up). The Run-3 data are selected with the single-lepton triggers of the Run-3 menu.

6.6.2.1 Object selection

6.6.2.1.1 Muon selection Three levels of lepton selection are defined in this analysis: *baseline*, Z, and W selection, summarised in Table 6.C. They are applied sequentially, each level applying tighter requirements than the previous one (the electron isolation WPs are not strictly nested). The baseline leptons use the least restrictive selection criteria in order to provide a more efficient veto of four-lepton events. Meanwhile the Z and W selections are tighter in order to suppress contamination from fake-lepton backgrounds.

The muon selection criteria are summarised in Table 6.C(a). The p_{T}-threshold increases with the level of selection, starting from $5\,\mathrm{GeV}$ for the baseline selection, to $15\,\mathrm{GeV}$ for the Z selection and finally to $20\,\mathrm{GeV}$ for the W selection. The $|\eta|$-threshold is set to 2.7 for the baseline selection and to 2.5 for the Z and W selections. This allows one to increase the rejection of ZZ background contamination by applying a veto on the events with more than three leptons, as detailed in Table 6.D. All muons within the coverage of the ID are required to pass the lepton-vertex association selection. Track parameter constraints such as the significance of the transverse impact parameter and the longitudinal impact parameter are applied to the baseline selection. The quality of the muon is required to be `Loose` for the baseline, `Medium` for Z selection, while it is required to be `Tight` for the W selection. The isolation of the muons is required to be `PflowLoose` for the baseline and Z selections, while it is required to be `PflowTight` for the W selection. The fixed-radius isolation (`FixedRad`) is used in Run-2 and the variable-radius one (`VarRad`) in Run-3. The muon–jet overlap removal is applied only for the Z and W selections following the procedure explained in Section 6.2.1.5.

6.6.2.1.2 Electron selection The electron selection criteria are summarised in Table 6.C(b). The p_{T}-threshold follows the muon selection criteria. The electron is required to be within the EM calorimeter $\left|\eta^{\mathrm{cluster}}\right| < 2.47$ and within the coverage of the ID $|\eta| < 2.5$ since the calorimeter defines the electron energy and the ID defines the electron direction. The calorimeter crack region between the barrel and end-cap calorimeters is excluded for the Z and W selections, corresponding to $1.37 < \left|\eta^{\mathrm{cluster}}\right| < 1.52$ to allow more efficient ZZ background rejection using the baseline selection. The `LooseAndBLayer`, `Medium`, and `Tight` identification WPs have a respective efficiency of $84\,\%$ to $96\,\%$, $72\,\%$ to $93\,\%$, and $68\,\%$ to $88\,\%$ for electrons with $10 < p_{\mathrm{T}} < 80\,\mathrm{GeV}$. These WPs are those of the LH identification of Section 5.3.1, which is used in both runs, the DNN-based identification of Section 5.3.9 not being used by this analysis. Similar track parameter constraint requirements as for the muons are applied to the baseline selection with a looser constraint on the significance of the transverse impact parameter. The electron isolation is required to be `Loose` for the baseline, `HighPtCaloOnly` for the Z and `Tight` for the W selections. The electron–jet overlap removal is applied only for the Z and W selections following the procedure explained in Section 6.2.1.5. A tool is used to remove ambiguities between electrons and photons following the procedure explained in Section 5.1.6 for the W selection. Within this selection, electrons are required to not pass the photon reconstruction in order to suppress Zγ background. Its inefficiency on true electrons from the W decay is found to be $\sim 5\,\%$. An additional rejection of converted photons misidentified as electrons is applied by requiring the `DFCommonAddAmbiguity` tagger to be equal to or below zero. The inefficiency of this requirement on true electrons from the W decay is found to be $\sim 0.5\,\%$, while it further rejects the Zγ background by about $35\,\%$ in the inclusive WZ selection.

6.6.2.1.3 Jet selection The jet objects are reconstructed from PFOs using the anti-k_t algorithm with a radius parameter $R = 0.4$ as detailed in Paragraph 6.2.1.2.3. The jet selection criteria are summarised in Table 6.C(c). Only jets with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\eta| < 4.5$ are considered. Jets originating from pile-up interactions are rejected with the JVT (NN-based for Run-3) at its `Tight` WP for jets with $|\eta| < 2.4$ and $p_{\mathrm{T}} < 60\,\mathrm{GeV}$, and with the fJVT at its `Loose` WP for jets with $2.5 < |\eta| < 4.5$ and $p_{\mathrm{T}} < 120\,\mathrm{GeV}$. The b-tagged jets use respectively the `DL1r` and the GN2 for Run-2 and Run-3, both WPs with a $70\,\%$ efficiency.

6.6.2.1.4 Missing transverse momentum selection The missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}}$, described in Section 6.2.1.3, is defined as the transverse momentum imbalance in the detector. It is calculated as the negative vector sum of the momentum (Equation 6.22) of user-selected calibrated objects (electrons, muons, jets), and of the soft-event contribution, which is reconstructed from tracks or calorimeter cell clusters not associated with the hard objects. In this analysis, $E_{\mathrm{T}}^{\mathrm{miss}}$ is reconstructed using calibrated electrons and muons passing the baseline selection, and calibrated jets before any selection.

Table 6.C: Lepton and jet object selection criteria for the Run-2 and Run-3 datasets. A threshold value indicates the cut applied at that level. ✔ marks a requirement applied identically across the relevant levels and a dash (—) one that is not applied. The lepton criteria are common to the two datasets except the muon isolation variant, and the jet criteria are given per dataset.

(a) Muon object selection criteria

Selection	Baseline	Z	W
p_{T} [GeV]	> 5	> 15	> 20
$\lvert\eta\rvert$	< 2.7	< 2.5	< 2.5
Quality	`Loose`	`Medium`	`Tight`
Isolation	`PflowLoose`	`PflowLoose`	`PflowTight`
Isolation variant	`FixedRad` (Run-2), `VarRad` (Run-3)		
$\left\lvert\frac{d_0}{\sigma_{d_0}}\right\rvert < 3$ (for $\lvert\eta\rvert < 2.5$)		✔	
$\lvert\Delta z_0 \sin\theta\rvert < 0.5\,\mathrm{mm}$ (for $\lvert\eta\rvert < 2.5$)		✔	
μ–jet overlap removal	—	✔	✔

(b) Electron object selection criteria

Selection	Baseline	Z	W
p_{T} [GeV]	> 5	> 15	> 20
Identification	`LooseAndBLayer`	`Medium`	`Tight`
Isolation	`Loose`	`HighPtCaloOnly`	`Tight`
Object quality		✔	
$\left\lvert\eta^{\mathrm{cluster}}\right\rvert < 2.47$, $\lvert\eta\rvert < 2.5$		✔	
$\left\lvert\frac{d_0}{\sigma_{d_0}}\right\rvert < 5$		✔	
$\lvert\Delta z_0 \sin\theta\rvert < 0.5\,\mathrm{mm}$		✔	
e–μ and e–e overlap removal		✔	
e–jet overlap removal	—	✔	✔
Exclude $1.37 < \left\lvert\eta^{\mathrm{cluster}}\right\rvert < 1.52$	—	✔	✔
Unambiguous author	—	—	✔
`DFCommonAddAmbiguity` ≤ 0	—	—	✔

(c) Jet object selection criteria

Selection	Run-2	Run-3
Anti-k_t, $R = 0.4$, PFOs		✔
p_{T} [GeV]		> 25
$\lvert\eta\rvert$		< 4.5
JVT ($\lvert\eta\rvert < 2.4$, $p_{\mathrm{T}} < 60\,\mathrm{GeV}$)	`Tight`	`Tight` (NN-based)
fJVT ($2.5 < \lvert\eta\rvert < 4.5$, $p_{\mathrm{T}} < 120\,\mathrm{GeV}$)		`Loose`
b-tagging (70 % WP)	`DL1r`	GN2

Pile-up degrades the resolution of the calorimeter-based measurement of the soft term. The `PFlow` $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction with the track-based measurement of the soft term is used in this analysis. The `Tight` jet selection WP is used.

6.6.2.2 Overlap removal

In addition to the general description of the overlap removal in Section 6.2.1.5, the exact procedure used in this analysis is done following this order:

1. Jets found within $\Delta R < 0.2$ of a Z-electron are discarded, and jets with fewer than three tracks and a distance $\Delta R < 0.2$ from the direction of a Z-muon are also rejected.

2. Baseline electrons sharing an ID track with a baseline muon are ignored, resulting in $0.4\,\%$ of the baseline electrons being removed.
3. Electrons passing the Z selection and overlapping with overlap-removal-cleaned jets within $\Delta R < 0.4$ are discarded. This selection has an efficiency of $> 99\,\%$ for Z-electrons.
4. Muons passing the Z selection and found within $\Delta R < 0.4$ of the direction of an overlap-removal-cleaned jet with at least three tracks are also discarded. This selection keeps $> 98\,\%$ of muon candidates.

6.6.2.3 Event selection

The selection of WZ candidate events is summarised in Table 6.D. Events are required to satisfy good quality conditions, achieved by rejecting corrupted and/or incomplete events. Event cleaning in both data and simulation is achieved by rejecting events with at least one jet misidentified as originating from non-collision background or detector problems, using the high-efficiency Loose jet-cleaning WP. A primary vertex with at least two associated tracks is required to be present in the event.

To continue the effort to reduce the background contamination from ZZ events, events with four or more baseline leptons are rejected via the "ZZ veto". This means rejection of events with more than three leptons passing the baseline muon or electron selections. The use of the baseline selection for the ZZ veto is motivated by the need to have a high efficiency for prompt leptons, while still rejecting events with additional leptons that could indicate a ZZ background.

A stringent selection is applied requiring exactly three leptons passing the Z-lepton selection. The Z candidate is therefore built from two of these leptons, while the remaining lepton is associated with the W candidate. The Z candidate is required to be formed by two same-flavour opposite-charge (SFOC) leptons, and if more than one pair of leptons satisfies this condition, the pair with invariant mass closest to the Z boson mass is selected. The transverse momentum threshold of the leading lepton is increased to $27\,\mathrm{GeV}$ to ensure that the lepton firing the trigger is above the trigger efficiency turn-on curve. The SFOC lepton pair is required to have an invariant mass within $10\,\mathrm{GeV}$ of the Z boson mass. Finally, the transverse mass of the W candidate, defined in Equation 6.24, is required to be above $30\,\mathrm{GeV}$.

This selection coincides with the inclusive WZ selection used in the previous measurement. On top of this selection, additional requirements are applied to define four different regions of interest for the VBS polarisation measurement. Those criteria are summarised in Table 6.D.

The main criteria to define the VBS event selection rely on the kinematics of the jets associated with the events, since we have shown in Section 6.2 the key role of the jets in this analysis. Events are therefore required to have at least two jets having a p_{T} above $40\,\mathrm{GeV}$ and an absolute pseudorapidity below 4.5. The two leading jets are required to be in opposite hemispheres of the detector, meaning that their pseudorapidities have opposite signs. The leading tagging jet is selected as the jet with the highest p_{T} in the event independently of its position. The second one is thus selected as the highest p_{T} jet in the opposite hemisphere of the detector among the remaining jets. Finally, the invariant mass of the two leading jets is required to be above $150\,\mathrm{GeV}$ in order to reduce the background contamination from triboson events.

From this VBS selection, four additional regions are defined, two control regions (CRs) and a validation region (VR) to control the background contamination and an SR to extract the polarisation fractions:

- The ZZ-CR is built on the same VBS selection, with the ZZ veto inverted: instead of rejecting the events containing a fourth baseline lepton, at least four baseline leptons are required. No additional requirement is applied either on the b-tagged jet multiplicity or on m_{jj}, so that this region covers the whole $m_{\mathrm{jj}} > 150\,\mathrm{GeV}$ phase-space. It is dominated by ZZjj-QCD events and is used to constrain the ZZ background contamination.
- The b-CR is defined by requiring at least one b-tagged jet in the event. This region is used to constrain the $\mathrm{t\bar{t}} + V$ and tZj background contamination.
- The WZjj-QCD VR is defined by requiring no b-tagged jets in the event and an invariant mass of the two leading jets below $500\,\mathrm{GeV}$. This region is used to validate the WZjj-QCD background modelling.
- The SR is defined by requiring no b-tagged jets in the event and an invariant mass of the two leading jets above $500\,\mathrm{GeV}$. This region is used to extract the polarisation fractions of the WZjj-EW signal.

The b-CR, the WZjj-QCD VR, and the SR are three mutually orthogonal sub-regions of the three-lepton VBS selection, separated by the b-tagged jet multiplicity and by m_{jj}. The ZZ-CR uses the same jet requirements but is orthogonal to the three others by its lepton multiplicity alone, since it selects precisely the four-lepton events that the ZZ veto removes from the VBS selection.

Table 6.D: Event selection of the WZjj polarisation measurement.

Type	Requirement
Inclusive	
Event cleaning	Reject LAr, Tile, and SCT corrupted events and incomplete events
Primary vertex	Hard scattering with at least two tracks
ZZ veto	Fewer than four baseline leptons
N_ℓ	Exactly three leptons passing the Z-lepton selection
Leading lepton p_T	$p_\mathrm{T} > 27\,\mathrm{GeV}$ $p_\mathrm{T} > 25\,\mathrm{GeV}$ for 2015
Z-leptons	Two SFOC leptons passing Z-lepton selection
Mass window	$\left\lvert m_{\ell\ell} - m_\mathrm{Z}^\mathrm{PDG}\right\rvert < 10\,\mathrm{GeV}$
W-lepton	Remaining lepton passing W-lepton selection
W transverse mass	$m_\mathrm{T}^\mathrm{W} > 30\,\mathrm{GeV}$
VBS	
Jet multiplicity	$N_\mathrm{jets} \geq 2$
p_T of two tagging jets	$p_\mathrm{T} > 40\,\mathrm{GeV}$
$\lvert\eta\rvert$ of two tagging jets	$\lvert\eta\rvert < 4.5$
η of two tagging jets	Opposite sign $\eta^{\mathrm{j}_1} \cdot \eta^{\mathrm{j}_2} < 0$
Invariant mass of two tagging jets	$m_\mathrm{jj} > 150\,\mathrm{GeV}$
b-*CR*	
b-tagged jet multiplicity	$N_\text{b-jet} > 0$
WZjj-*QCD* *VR*	
b-tagged jet multiplicity	$N_\text{b-jet} = 0$
Invariant mass of two tagging jets	$m_\mathrm{jj} < 500\,\mathrm{GeV}$
SR	
b-tagged jet multiplicity	$N_\text{b-jet} = 0$
Invariant mass of two tagging jets	$m_\mathrm{jj} > 500\,\mathrm{GeV}$
ZZ-*CR*	
N_ℓ	≥ 4 baseline leptons
b-tagged jets and m_jj	No additional requirement

6.6.3 Processes involved

Throughout this chapter we described in depth the WZjj process, its topology, and the different contributions to the total cross-section. Among the explanations, we mentioned various other processes that are involved and needed to capture the full picture. Many of these processes are treated as background in the analysis, but they are also of interest in their own right. This section will summarise and give the complete overview of all processes involved in the analysis, including the signal and background processes.

The polarisation measurement is performed in the fiducial phase-space defined in Table 6.B and using the object and event selection described in Section 6.6.2. The details of the MC samples used to model the signal and background processes are described in Appendix F.

The extraction of the boson polarisation is obtained from a combination of Run-2 and Run-3 datasets. This combination is possible only if the two datasets exhibit consistent behaviour in terms of polarisation composition and SR acceptance.

6.6.3.1 Signal definition

The signal is defined as the WZjj-EW processes involving exclusively EW interactions at Born level of order α_EW^6 including the boson decays. To enhance the sensitivity to the polarisation of the bosons, only situations where the W and Z bosons are produced on-shell are considered part of the signal. The tau decays are excluded from the signal definition and are treated as background. The signal definition therefore comprises only processes where the bosons decay leptonically into electrons or muons.

Interference effects between different polarisation states of the bosons as well as non-resonant effects are included in the signal definition. The NLO EW calculation of [413], collected in Table 6.A in Section 6.5.3, quantifies the two separately. The interference between polarisation states, defined there as the unpolarised DPA cross-section minus the sum of the four doubly polarised ones, amounts

to $1.7\,\%$ of the unpolarised cross-section at LO and $1.8\,\%$ at NLO EW. The off-shell contribution, the difference between the full off-shell result and the unpolarised DPA one, amounts to $1.6\,\%$ at LO and $1.5\,\%$ at NLO EW, the tight $\pm 10\,\mathrm{GeV}$ window around m_{Z} in the fiducial phase-space keeping it small.

Those computations use the same fiducial phase-space definition as the one used in this analysis, so their polarisation fractions are directly applicable. They are however performed for $\mathrm{W}^{+}\mathrm{Z}$ only, so the absolute cross-sections of Table 6.A are not to be compared with the $\mathrm{W}^{\pm}\mathrm{Z}$ yields of Table 6.E: only the fractions are, and only those normalised to the sum of the four doubly polarised modes, which is the convention of Table 6.E.

Both effects are too small to be constrained by the present data statistics. They will therefore be included as part of the measured polarisation fractions, which are consequently *effective* fractions, following the same prescription used in the previous ATLAS publication of the WZ joint polarisation measurement [11].

To also evaluate the size of these contributions from the MC samples described in Appendix F, the following procedure is used: in the case of interference effects, the difference between the on-shell WZjj-EW *polarisation-sum* sample and the sum of the four polarised samples is taken as an estimate of the interference contribution. This is the same definition as the one used for the interference row of Table 6.A, so the two numbers are directly comparable: the MC estimate is below $2\,\%$ of the total WZjj-EW contribution for both Run-2 and Run-3, in agreement with the $1.7\,\%$ to $1.8\,\%$ obtained from the NLO EW calculation [413]. It is therefore not treated as a separate component but absorbed into the measured fractions, as stated above.

This interference between polarisation states must not be confused with the interference between the WZjj-EW and WZjj-QCD productions, WZjj-INT, discussed in the next section: the latter is a distinct contribution at the level of $6\,\%$ to $7\,\%$ of the WZjj-EW yield and *is* modelled by a dedicated sample and fitted as a background.

For off-shell effects, the difference between the WZjj-EW *full* sample, with W and Z bosons decaying to electrons or muons, and the on-shell sample is taken as an estimate of the off-shell contribution. This difference is constituted by the contribution from non-resonant diagrams but also contains effects from triboson processes, which cannot be disentangled from the other off-shell contributions in this case. The size of these off-shell effects is found to be at the level of $1.5\,\%$ of the total WZjj-EW contribution, matching the $1.5\,\%$ to $1.6\,\%$ of Table 6.A, and is likewise absorbed into the effective fractions instead of being treated as a separate component.

The expected numbers of events in the fiducial phase-space and in the reconstructed phase-space (obtained after detector simulation and reconstruction) are reported in Table 6.E for the Run-2 and Run-3 datasets, together with the corresponding fiducial cross-sections, polarisation fractions f_i, and correction factors C_i of each polarisation state i. The correction factor C_i is defined as the ratio between the number of events after detector reconstruction and the number of events in the fiducial phase-space. The same quantities obtained with the Sherpa samples are listed in the same table, so that the two generators can be compared directly.

The correction factors agree between the Run-2 and Run-3 datasets to within $1.5\,\%$, which indicates that the transition from Run-2 to Run-3 does not introduce any significant variation in the polarisation pattern or in the relative acceptance between channels. A good agreement is likewise observed between the MadGraph and Sherpa predictions, both in terms of polarisation composition and of acceptance.

Computations of NLO EW corrections to the WZjj-EW process have been performed in the context of the present analysis [413]. The results of these computations are reported in Section 6.5.3. The NLO EW corrections are found to be sizeable and negative, ranging from approximately $-14\,\%$ for the $\mathrm{W_0Z_0}$ configuration to about $-20\,\%$ for the $\mathrm{W_TZ_T}$ configuration. These corrections are found to be relatively uniform across the different polarisation states, so that the integrated fractions move by less than one percentage point (Table 6.A), resulting in a flat behaviour of the NLO EW K-factors correction. The NLO EW corrections are therefore not applied to the MC samples used in this analysis, but they are instead included as a systematic uncertainty in the measurement of the polarisation fractions. A comparison between the polarisation fractions obtained from the MC samples and the ones obtained from the NLO EW computation can be made using Table 6.E and the last two columns of Table 6.A in Section 6.5.3, which are the ones normalised to the sum of the four doubly polarised modes, as the MC fractions are. The like-for-like comparison is the LO one, since neither MadGraph nor Sherpa includes the EW corrections: the calculation gives 8.8, 18.8, 18.6, and $53.9\,\%$ for $\mathrm{W_0Z_0}$, $\mathrm{W_0Z_T}$, $\mathrm{W_TZ_0}$, and $\mathrm{W_TZ_T}$, against 9.1, 19.1, 18.6, and $53.3\,\%$ for MadGraph and 8.8, 19.2, 18.7, and $53.3\,\%$ for Sherpa in the Run-2 configuration, and 9.0, 19.0, 18.5, and $53.5\,\%$ for MadGraph and 8.7, 19.2, 18.7, and $53.5\,\%$ for Sherpa in the Run-3 configuration. The three predictions agree to better than one percentage point on every fraction, that is within the spread between the two generators themselves, which confirms that the MC samples used in this analysis provide a good description of the polarisation composition of the WZjj-EW process.

Table 6.E: Fiducial cross-section (per lepton-flavour channel), polarisation fraction, fiducial and reconstructed event yields, and the associated C-factor for each polarisation state of the WZjj-EW process, shown for the MadGraph and Sherpa predictions and for the Run-2 and Run-3 configurations. The upper block lists the doubly polarised states and the lower block the singly polarised states. Uncertainties are statistical MC only.

		Run-2					Run-3				
State	**Generator**	σ [ab]	f_i [%]	$N^{\mathrm{fid.}}$	N^{reco}	C_i	σ [ab]	f_i [%]	$N^{\mathrm{fid.}}$	N^{reco}	C_i
$\mathrm{W_0Z_0}$	MadGraph	28.46 ± 0.02	9.05 ± 0.01	15.95 ± 0.01	8.60 ± 0.01	0.539 ± 0.001	31.00 ± 0.02	8.98 ± 0.01	20.55 ± 0.01	10.95 ± 0.02	0.533 ± 0.001
	Sherpa	29.31 ± 0.06	8.76 ± 0.02	16.42 ± 0.03	8.83 ± 0.03	0.537 ± 0.002	32.32 ± 0.07	8.68 ± 0.02	21.23 ± 0.04	11.23 ± 0.05	0.529 ± 0.003
$\mathrm{W_0Z_T}$	MadGraph	59.89 ± 0.04	19.05 ± 0.02	33.55 ± 0.02	18.28 ± 0.02	0.545 ± 0.001	65.66 ± 0.04	19.03 ± 0.02	43.52 ± 0.03	23.62 ± 0.04	0.543 ± 0.001
	Sherpa	64.33 ± 0.10	19.23 ± 0.03	36.05 ± 0.05	19.62 ± 0.04	0.544 ± 0.001	71.45 ± 0.11	19.19 ± 0.03	46.93 ± 0.07	25.61 ± 0.09	0.546 ± 0.002
$\mathrm{W_TZ_0}$	MadGraph	58.37 ± 0.04	18.57 ± 0.02	32.71 ± 0.02	19.04 ± 0.02	0.582 ± 0.001	63.87 ± 0.04	18.51 ± 0.01	42.33 ± 0.03	24.48 ± 0.04	0.578 ± 0.001
	Sherpa	62.51 ± 0.09	18.68 ± 0.03	35.02 ± 0.05	20.35 ± 0.04	0.581 ± 0.002	69.47 ± 0.10	18.66 ± 0.03	45.62 ± 0.07	26.40 ± 0.09	0.579 ± 0.002
$\mathrm{W_TZ_T}$	MadGraph	167.67 ± 0.13	53.33 ± 0.05	93.95 ± 0.07	54.72 ± 0.06	0.582 ± 0.001	184.60 ± 0.13	53.49 ± 0.04	122.34 ± 0.08	71.34 ± 0.11	0.583 ± 0.001
	Sherpa	178.46 ± 0.19	53.33 ± 0.07	99.99 ± 0.10	58.40 ± 0.09	0.584 ± 0.001	199.07 ± 0.20	53.47 ± 0.07	130.74 ± 0.13	76.53 ± 0.18	0.585 ± 0.002
$\mathrm{W_0Z_X}$	MadGraph	88.34 ± 0.05	28.10 ± 0.02	49.50 ± 0.03	26.88 ± 0.02	0.543 ± 0.001	96.67 ± 0.05	28.01 ± 0.02	64.07 ± 0.03	34.56 ± 0.04	0.539 ± 0.001
	Sherpa	93.64 ± 0.11	27.98 ± 0.04	52.47 ± 0.06	28.44 ± 0.05	0.542 ± 0.001	103.77 ± 0.13	27.87 ± 0.04	68.15 ± 0.08	36.84 ± 0.10	0.541 ± 0.002
$\mathrm{W_XZ_0}$	MadGraph	86.83 ± 0.05	27.62 ± 0.02	48.65 ± 0.03	27.64 ± 0.02	0.568 ± 0.001	94.87 ± 0.05	27.49 ± 0.02	62.88 ± 0.03	35.43 ± 0.04	0.563 ± 0.001
	Sherpa	91.81 ± 0.11	27.44 ± 0.04	51.44 ± 0.06	29.17 ± 0.05	0.567 ± 0.001	101.79 ± 0.12	27.34 ± 0.04	66.85 ± 0.08	37.64 ± 0.10	0.563 ± 0.002

6.6.3.2 Background definition

Several background processes may mimic the experimental signature of WZjj production and contribute to the selected event sample. Such processes can be classified into two groups: irreducible background with at least three prompt leptons in the final state and reducible background, containing at least one "fake" lepton.

The first group is modelled with MC simulation, with the normalisation of each process either constrained by the data or taken directly from simulation. The normalisations of the ZZjj-QCD, $\mathrm{t\bar{t}} + V$, and tZj backgrounds are left free in the fit and adjusted using dedicated CRs (defined in Table 6.D), each enriched in the process of interest and orthogonal to the SR. The resulting normalisation factors (NFs) are propagated to the SR. The other irreducible processes, including the dominant WZjj-QCD background, are kept to their nominal MC expectation with their associated modelling uncertainties, the WZjj-QCD modelling being validated in a dedicated VR covering a low-m_{jj} region orthogonal to the SR.

The dominant background is the QCD-induced production of the same WZjj final state, WZjj-QCD, produced via strong interactions at Born level of order $\alpha_{\mathrm{S}}^2\alpha_{\mathrm{EW}}^4$. Having an identical final state to the signal, it is irreducible. Here the two accompanying jets originate from strong radiation rather than from the t-channel quark scattering that characterises the VBS signal, yielding a more central dijet system of lower invariant mass. This is why WZjj-QCD dominates the low-m_{jj} VR used to validate its modelling.

Between the signal and the WZjj-QCD background, there is the interference between the two processes: WZjj-INT. The interference is treated as a background in this analysis. The size of the interference contribution is found to be at the level of $6\,\%$ to $7\,\%$ of the total WZjj-EW contribution.

The second dominant irreducible background comes from ZZ production when one lepton is lost. The estimation of this background is validated in a corresponding CR, enriched in ZZ events. Among them, we distinguish as for WZjj the EW and QCD production modes: ZZjj-EW and ZZjj-QCD. The ZZjj-QCD processes are produced via strong interactions at Born level of order $\alpha_{\mathrm{S}}^2\alpha_{\mathrm{EW}}^4$, while the ZZjj-EW processes are produced via EW interactions at Born level of order α_{EW}^6.

Processes with a b-quark in the initial or final state are not considered as signal but contribute to the selected event sample and are therefore treated as background: the accompanying b-jets place these events in the b-CR, where their normalisation is constrained. The main contribution comes from $\mathrm{t\bar{t}} + V$ production, where a top–antitop pair is accompanied by a W or Z boson ($V = \mathrm{W}, \mathrm{Z}$), at Born level of order $\alpha_{\mathrm{S}}^2\alpha_{\mathrm{EW}}^3$. The three prompt leptons arise from the leptonic decays of the top quarks ($\mathrm{t} \to \mathrm{Wb} \to \ell\nu_\ell\,\mathrm{b}$) and of the associated boson. A smaller contribution comes from tZj production, resulting from a t-channel exchange of a W boson between a b- and a u-quark and giving a final state with a t-quark, a Z boson, and a light-quark jet, the Z supplying the lepton pair and the leptonic top decay the third lepton.

Finally, a small irreducible contribution arises from triboson production (VVV), corresponding to the associated production of three massive gauge bosons (WWW, WWZ, WZZ, ZZZ) through purely EW interactions, where a combination of leptonic and hadronic boson decays reproduces the three-lepton-plus-jets signature. It is taken directly from simulation and stays at the percent level or below in the SR.

The second group contains contributions from $\mathrm{Z}+$ jets, $\mathrm{Z}\gamma$, and $\mathrm{t\bar{t}}$ processes. A fake lepton can be a jet misidentified as a lepton or a non-prompt lepton within a jet. In the case of electrons, converted photons reconstructed as electrons are an additional source of fakes. Such background is estimated using the data-driven matrix method technique. The technique and associated uncertainties are described in Section E.1. It will be the baseline method used to determine reducible background. A second semi-data-driven method, called the MC scale factor method, is also used to determine the amount of reducible background events independently of the matrix method. This method is presented in Section E.2. It is used as a cross-check of the background determined using the matrix method.

An example of the Feynman diagrams for the main background processes is shown in Figure 6.17. The irreducible backgrounds are shown in Figures 6.17(b) to 6.17(e), 6.17(h), and 6.17(i). The reducible backgrounds are shown in Figures 6.17(a), 6.17(f), and 6.17(g). The $\mathrm{t\bar{t}} + V$ background is shown in its $\mathrm{t\bar{t}W}$ and $\mathrm{t\bar{t}Z}$ topologies. The tZj background is shown in its t-channel topology. The VVV background is shown in its WWZ topology. The ZZjj-QCD backgrounds are shown in their t-channel topologies. The $\mathrm{Z}+$ jets background is shown in its Drell–Yan topology. The $\mathrm{Z}\gamma$ background is shown in its t-channel topology.

6.6.4 Standard Model predictions and expected yields

The data analysed in this thesis combine $140\,\mathrm{fb}^{-1}$ recorded at $13\,\mathrm{TeV}$ during Run-2 with $164\,\mathrm{fb}^{-1}$ collected at $13.6\,\mathrm{TeV}$ during the start of Run-3, a total of $304\,\mathrm{fb}^{-1}$ corresponding to about $2.4 \cdot 10^{16}$ inelastic pp

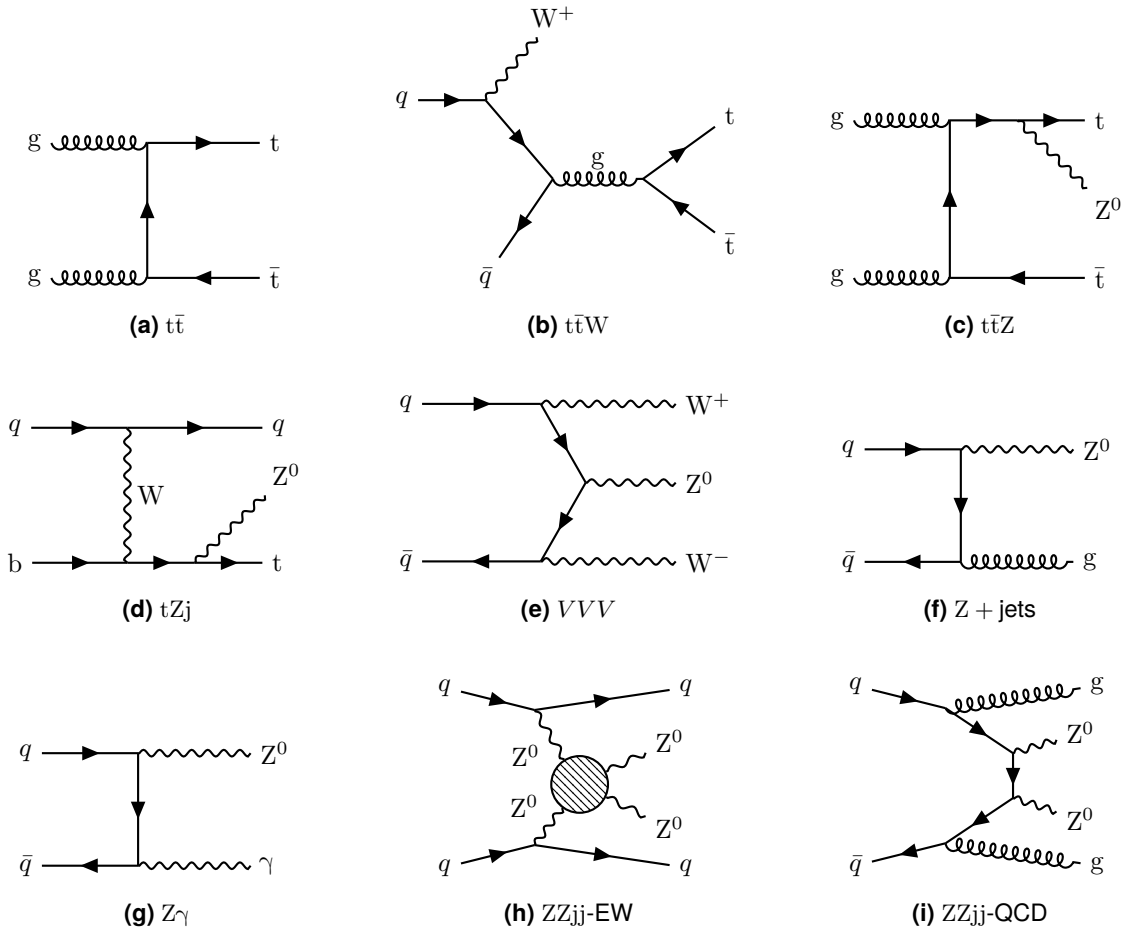


Figure 6.17: Feynman diagrams for the background processes. The irreducible backgrounds: $\mathrm{t\bar{t}} + V$ in its $\mathrm{t\bar{t}W}$ (b) and $\mathrm{t\bar{t}Z}$ (c) topologies, tZj (d), VVV (e), ZZjj-EW (h), and ZZjj-QCD (i) providing at least three prompt leptons. The reducible backgrounds: $\mathrm{t\bar{t}}$ (a), $\mathrm{Z}+\mathrm{jets}$ (f), and $\mathrm{Z}\gamma$ (g) enter through a fake or non-prompt lepton.

collisions*. The EW production of a WZ pair in association with two jets is among the rarest signatures accessible in this dataset, and its expected yield in the SR follows directly from the polarised samples of Table 6.E through the chain $N^{\mathrm{reco}} = C \times \sigma^{\mathrm{fid.}} \times \mathcal{L}$. Summed over the four joint-polarisation states and the four light-lepton channels ($\mathrm{Z} \to \mathrm{ee}, \mu\mu$ and $\mathrm{W} \to \mathrm{e}\nu_{\mathrm{e}}, \mu\nu_{\mu}$), the signal fiducial cross-section in the analysis phase-space reaches $\approx 1.3\,\mathrm{fb}$ at $13\,\mathrm{TeV}$ and $\approx 1.5\,\mathrm{fb}$ at $13.6\,\mathrm{TeV}$, so that $N^{\mathrm{fid.}} = \sigma^{\mathrm{fid.}}\mathcal{L} \approx 190$ signal events are produced within the fiducial volume during Run-2 and ≈ 245 during the partial Run-3. The reconstruction-level correction factor, $C \approx 0.57$, then brings these to $N^{\mathrm{reco}} \approx 107$ and ≈ 140 reconstructed events, respectively, close to 250 in total, roughly one signal event for every 10^{14} inelastic collisions recorded by ATLAS.

These few hundred signal events sit in a heavily background-dominated SR, whose total prediction amounts to about 800 events for Run-2 and 1060 for Run-3. The yields are dominated by the irreducible QCD-induced WZjj-QCD production (530 and 710 events, respectively), with smaller contributions from misidentified leptons, ZZjj, tZj, $\mathrm{t\bar{t}} + V$, and the triboson VVV production. As described in Section 6.6.2, the normalisations of the ZZjj-QCD, $\mathrm{t\bar{t}} + V$, and tZj processes are constrained in the dedicated CRs defined in Table 6.D, while the WZjj-QCD normalisation is kept at its MC prediction, and the reducible background from misidentified leptons is taken from the data-driven matrix method (Section E.1) and cross-checked against the MC scale factor estimate (Section E.2). The expected yields in the SR and in the two CRs are summarised in Table 6.F, separately for the Run-2 and Run-3 datasets, with all sources of uncertainty described in Section 7.2 included. At this pre-fit stage the WZjj-EW signal accounts for about $13\,\%$ of the SR and the WZjj-QCD background for about $67\,\%$.

The signal itself is largely transverse: the fully transverse $\mathrm{W_TZ_T}$ state alone accounts for about $53\,\%$ of it, the two mixed states $\mathrm{W_0Z_T}$ and $\mathrm{W_TZ_0}$ for about $19\,\%$ each, and the doubly longitudinal $\mathrm{W_0Z_0}$ state for only about $9\,\%$. This last state represents barely $1\,\%$ of the SR, close to 20 reconstructed

*Estimated from an inelastic cross-section $\sigma_{\mathrm{inelastic}} \approx 80\,\mathrm{mb}$, giving about $1.1 \cdot 10^{16}$ collisions at $13\,\mathrm{TeV}$ and $1.3 \cdot 10^{16}$ at $13.6\,\mathrm{TeV}$.

Table 6.F: Expected pre-fit event yields in the SR and in the two CRs for the Run-2 and Run-3 datasets, with all sources of uncertainty included.

Process	SR		ZZ-CR		b-CR	
	Run-2	Run-3	Run-2	Run-3	Run-2	Run-3
Data	646	953	210	235	666	766
Total pred.	800 ± 190	1060 ± 270	204 ± 11	240 ± 24	687 ± 56	751 ± 75
WZjj-EW	—	—	0.6 ± 0.1	0.8 ± 0.2	5.1 ± 1.3	3.4 ± 0.9
$\mathrm{W_0Z_0jj}$-EW	8.8 ± 0.2	11.2 ± 0.6	—	—	—	—
$\mathrm{W_0Z_Tjj}$-EW	19.6 ± 0.5	25.6 ± 1.4	—	—	—	—
$\mathrm{W_TZ_0jj}$-EW	20.3 ± 0.5	26.4 ± 1.4	—	—	—	—
$\mathrm{W_TZ_Tjj}$-EW	58.4 ± 1.8	76.5 ± 4.5	—	—	—	—
WZjj-QCD	530 ± 190	710 ± 260	6.3 ± 2.1	8.7 ± 3.0	93 ± 36	75 ± 27
Misid. leptons	46 ± 11	62 ± 19	1.7 ± 0.4	2.2 ± 0.5	155 ± 38	195 ± 61
ZZjj-QCD	41.6 ± 2.9	51.8 ± 9.4	158.4 ± 8.6	181 ± 20	10.1 ± 0.6	9.0 ± 0.9
tZj	28.5 ± 1.0	37.3 ± 3.5	$0.4 \pm < 0.1$	0.8 ± 0.1	134.1 ± 5.0	154.4 ± 9.4
$\mathrm{t\bar{t}} + V$	20.9 ± 0.7	22.9 ± 2.3	10.0 ± 0.3	11.9 ± 1.0	289 ± 11	312 ± 19
ZZjj-EW	6.7 ± 1.7	15.4 ± 3.9	22.8 ± 5.7	30.4 ± 7.7	0.3 ± 0.1	0.3 ± 0.1
WZjj-INT	6.5 ± 1.3	9.2 ± 1.9	$0.1 \pm < 0.1$	$0.1 \pm < 0.1$	0.6 ± 0.1	0.4 ± 0.1
WZjj-EW in τ	5.9 ± 0.6	7.5 ± 0.8	—	—	—	—
VVV	2.4 ± 0.6	3.7 ± 0.9	4.1 ± 1.0	4.8 ± 1.2	0.4 ± 0.1	0.4 ± 0.1

events across the two periods, or roughly one for every 10^{15} inelastic collisions. Isolating such a small and specifically polarised signal from a background several times larger is the central difficulty of this measurement: building on the event selection and background estimates established above, it further requires the complete reconstruction of the event topology, from which the polarisation fractions are ultimately extracted.

6.7 Multivariate analysis

The aim of this analysis is to measure the polarisation of the W and Z bosons in the WZjj process, in particular in the VBS topology. The longitudinal polarisation of the bosons is a key feature of the EW SSB mechanism, and its measurement is a crucial test of the SM. The polarisation being only a statistical property of the bosons, it is not possible to measure it directly via an observable, but it can be inferred from the angular distribution of the decay products of the bosons thanks to the SDM formalism developed in Section 6.3. It describes how the polarisation of the bosons is transferred to the angular distribution of their decay products.

To extract the polarisation fractions from data, a multivariate analysis is performed: the outputs of the two DNNs define a 2D discriminant plane, which is partitioned by a clustering algorithm so as to maximise the sensitivity to the longitudinal polarisation. This section presents the final template retained for the fit. The Voronoi cells are ordered and flattened into a 1D histogram, which serves as the template for the binned maximum-likelihood fit described in Chapter 7.

A promising template that was not retained, due to the lack of statistics and to its insufficient discrimination power, is a 3D template built from the DNNs output combined with a pyramidal fractional binning. It is presented in Appendix G.

Three distinct measurements of the polarisation of the W and Z bosons in the VBS production of the WZjj process are considered:

- The first is the measurement of the polarisation fractions when both bosons are longitudinally polarised: $\mathrm{W_0Z_0jj}$-EW. It is the most challenging measurement due to the small cross-section of the longitudinal polarisation, with only $1\,\%$ of the total event yield presented in Table 6.F.
- The second is the measurement of the polarisation fractions when only the W boson is longitudinally polarised: $\mathrm{W_0Z_Xjj}$-EW, where the X in the Z boson polarisation indicates that the Z boson is either transversely or longitudinally polarised. It is a less challenging measurement than the first due to the larger cross-section of the longitudinal polarisation, with about $3.5\,\%$ of the total event yield presented in Table 6.F. However, it is still a challenging measurement due to the neutrino in the W decay, which makes the W boson harder to reconstruct than the Z boson decaying into two charged leptons.

- The third is the measurement of the polarisation fractions when only the Z boson is longitudinally polarised: $\mathrm{W_X Z_0 jj}$-EW, where the X in the W boson polarisation indicates that the W boson is either transversely or longitudinally polarised. It is a less challenging measurement than the first due to the larger cross-section of the longitudinal polarisation, with about $3.6\,\%$ of the total event yield presented in Table 6.F. It is the most promising channel due to the Z boson decay into two charged leptons, which is easier to reconstruct than the neutrino decay of the W boson.

In order to maximise the sensitivity to the longitudinal polarisation and account for these three distinct measurements, a 2D template is built from the DNNs output. A first DNN common to the three measurements is trained to separate the WZjj-EW signal from the main background. Then three distinct DNNs are trained to separate the targeted polarisation from the other polarisations. The three DNNs are trained to separate respectively the $\mathrm{W_0 Z_0 jj}$-EW, $\mathrm{W_0 Z_X jj}$-EW, and $\mathrm{W_X Z_0 jj}$-EW polarisation from the other polarisations. The 2D template is built from the three DNN outputs and a clustering algorithm is applied to maximise the sensitivity to the longitudinal polarisation.

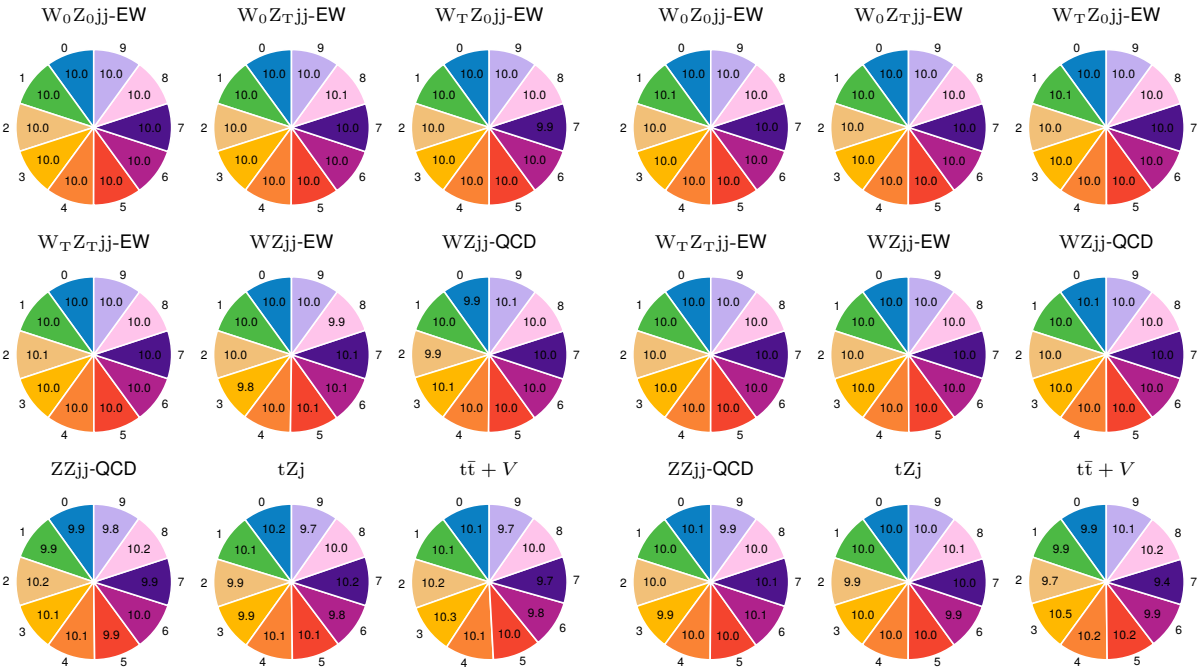
6.7.1 Generalities

Each DNN is trained on the MC samples described in Appendix F. The training is performed on the events passing the selection described in Section 6.6.2. The events are split into a training and a validation set with a ratio of $80\,\%$ for the training set and $20\,\%$ for the validation set. The training is performed using a k-fold cross-validation with $k = 5$. The fold partitioning is performed using the last digit of the event number. For the five-fold partition, the fold index is defined as:

$$i_{\text{fold}} = \left\lfloor \frac{\texttt{Event Number} \bmod 10}{2} \right\rfloor , \qquad i_{\text{fold}} \in \{0, \ldots, 4\} . \tag{6.137}$$

Consequently, fold 0 contains events whose event number ends in 0 or 1, fold 1 those ending in 2 or 3, fold 2 those ending in 4 or 5, fold 3 those ending in 6 or 7, and fold 4 those ending in 8 or 9. Since each event is assigned to exactly one fold, the training and validation subsets do not overlap. The last digit of the event number is expected to be uniformly distributed between 0 and 9. Its distribution for each sample is shown in Figure 6.18.

As a consequence, each MC event is evaluated once by the DNN that was not trained on it, and the DNN output is therefore free of over-training bias. The DNN output for each fold is then combined to build the final DNN output for the whole dataset, using only events not seen in training.



(a) Run-2 **(b)** Run-3

Figure 6.18: Event distribution across folds from the event-number last digit for Run-2 (a) and Run-3 (b).

All input features are normalised using a Min-Max normalisation in the range $[0, 1]$ fitted on the training set and applied to the validation set in order to ensure statistical independence between the two sets.

This normalisation is performed for each fold independently. This handles the large span of the input features and allows having a better convergence of the DNN during the training.

All DNNs are feed-forward fully connected NNs with a ReLU activation function for the hidden layers. The output layer has a number of nodes equal to the number of classes and thus depends on the DNN task. Between the hidden layers, a dropout layer is added to prevent overfitting.

The DNN architecture was subject to constraints due to portability to the analysis framework in C++ but also to Fortran, in which MoCaNLO is written. MoCaNLO is a fixed-order integrator, not an event generator, so the non-negligible NLO EW correction to the measured observable can only be obtained by evaluating the DNNs and the Voronoi binning inside the integration, event by event. More details are given in Section 6.5.3. More elaborate architectures could therefore not be used. However, a study comparing the present setup to an adaptation of the transformer using the GN2 [380] developed for jet tagging and presented in Paragraph 6.2.1.2.6 was carried out. No improvement was observed with respect to the present methods, suggesting no necessity to change the current DNN architecture.

The optimiser used for the training is the ADAM optimiser with a learning rate scheduler that reduces the learning rate by a factor of 0.33 if the monitored metric does not improve for 15 epochs. The training is stopped early if the monitored metric does not improve for 25 epochs and the final model is the one with the best monitored metric on the validation set. The CCE loss of the S. vs B. network is trained with label smoothing, the target of the correct class being lowered from 1 to $1-\epsilon$ and the remainder shared among the other classes, with $\epsilon = 0.05$. It regularises the training by discouraging over-confident outputs. The three polarisation networks use the BCE loss without smoothing. The hyperparameters of each DNN are optimised using a BO following the procedure described in Section 4.5. Each hyperparameter is optimised in a range of values presented in Table H.A. The hyperparameters are optimised to maximise the targeted metrics for each DNN that will be described in the following. The results of the BO optimisation are presented in Appendix H. The hyperparameters of each DNN are summarised in Table 6.G.

Table 6.G: Summary of DNN hyperparameters and architecture choices for the four networks employed in this analysis.

Category	**Parameter**	**Value**			
		S. vs B.	**Pol. 00**	**Pol. 0X**	**Pol. X0**
Preprocessing	Input normalisation	Min-Max $\in [0, 1]$			
	Number of k-folds	5			
	Train / validation split	$80\,\%$ / $20\,\%$ (per fold)			
Architecture	Number of hidden layers	3	3	4	3
	Nodes per layer	660	985	965	510
	Hidden activation	ReLU			
	Dropout rate	0.45	0.6	0.6	0.2
	Output nodes	4		1	
	Output activation	Softmax		Sigmoid	
Training	Loss function	CCE, L_{CCE}		BCE, L_{BCE}	
	Label smoothing ϵ	0.05		—	
	Optimiser	ADAM			
	Initial learning rate	10^{-4}	10^{-3}	10^{-2}	10^{-4}
	Weight decay	10^{-2}	10^{-2}	10^{-3}	10^{-5}
	Learning rate scheduler factor	0.33 (patience 15 epochs)			
	Minimum learning rate	10^{-10}			
	Early stopping	patience 25 epochs			
	Batch size	$10\,000$	320	$10\,000$	160
	Monitored metric	Weighted accuracy		Weighted AUC	

The raw events or entries of each MC sample used for the training are presented in Table 6.H. For Run-2, the number of raw events (entries) for WZjj-EW (polarised or not) is around one million while the main background, WZjj-QCD, accounts for $200\,000$ events. The total background used for the S. vs B. DNN is around $450\,000$ events. Similarly for Run-3, the number of events is of the same order of magnitude. To avoid any issue due to class imbalance, we decided to use the weights of each event during the training. Negative event weights are handled by taking their absolute value. Another reason for using the weights while still keeping a physical sense for the event is that Sherpa 3.0 incorporates polarisation effects purely via weights as explained in Section 6.5.1.

Table 6.H: Order of magnitude of the number of raw MC events for each process.

Process	Number of events	
	(Run-2)	(Run-3)
WZjj-EW (polarisation-sum)	$1.1 \cdot 10^6$	$1.3 \cdot 10^6$
$\mathrm{W_0Z_0jj}$-EW	$1.0 \cdot 10^6$	$1.2 \cdot 10^6$
$\mathrm{W_0Z_Tjj}$-EW	$1.1 \cdot 10^6$	$1.2 \cdot 10^6$
$\mathrm{W_TZ_0jj}$-EW	$1.2 \cdot 10^6$	$1.5 \cdot 10^6$
$\mathrm{W_TZ_Tjj}$-EW	$1.1 \cdot 10^6$	$1.3 \cdot 10^6$
WZjj-QCD	$2.0 \cdot 10^5$	$6.3 \cdot 10^5$
ZZjj-QCD	$5.6 \cdot 10^4$	$1.8 \cdot 10^5$
ZZjj-EW	$1.3 \cdot 10^5$	$1.0 \cdot 10^4$
tZj	$1.7 \cdot 10^5$	$6.6 \cdot 10^5$
$\mathrm{t\bar{t}} + V$	$2.1 \cdot 10^4$	$3.2 \cdot 10^4$

6.7.2 Signal versus background DNN

The first DNN is trained to separate the WZjj-EW (with the polarisation-sum sample, generated with MadGraph) signal from the main background. The main background is the WZjj-QCD (generated with MadGraph for Run-2 and Sherpa for Run-3) process, but the other processes are included: ZZjj-QCD, tZj, and $\mathrm{t\bar{t}} + V$. The fakes, VVV, and WZjj-INT are not considered due to their low contribution to the total background. Also, the ZZjj-EW process is not considered since it closely resembles the signal and confounds the DNN training, making the disentanglement of the signal and background more difficult.

It will be labelled as S. vs B. in the following.

The approach is to use a DNN with four output nodes, one for each class: WZjj-EW, WZjj-QCD, ZZjj-QCD, and the last one grouping the other processes: tZj and $\mathrm{t\bar{t}} + V$. By using a softmax activation function on the output layer, given in Figure 4.4(c), we can interpret the output of the DNN as a probability distribution over the four classes. The targeted values are therefore one-hot encoded vectors, where the correct class is represented by a 1 and the other classes by 0. The DNN is trained to minimise the CCE loss function, given in Equation 4.7, which measures the difference between the predicted probability distribution and the true one-hot encoded vector. This approach allows the DNN to learn to distinguish between the signal and background classes effectively, even in the presence of multiple background processes. A simpler approach with only two output nodes, one for the signal and one for the background, was tested, but it was found to be less effective in separating the signal from the background. The four output nodes approach allows the DNN to learn more complex decision boundaries and better capture the nuances of the different background processes.

Negative weights, present due to some samples being generated at NLO, are handled by using the absolute value of the weights and scaling the number of WZjj-EW events to the sum of the absolute values of the weights for the background. This allows one to put more emphasis on the signal events during the training.

The S. vs B. DNN is trained to maximise the weighted accuracy metric, which measures the model's overall predictive accuracy while accounting for the imbalance between the small signal class and the larger background classes, providing a more informative evaluation of performance than plain accuracy. The class with the highest score is taken as the prediction. The ROC curve and the AUC are also monitored.

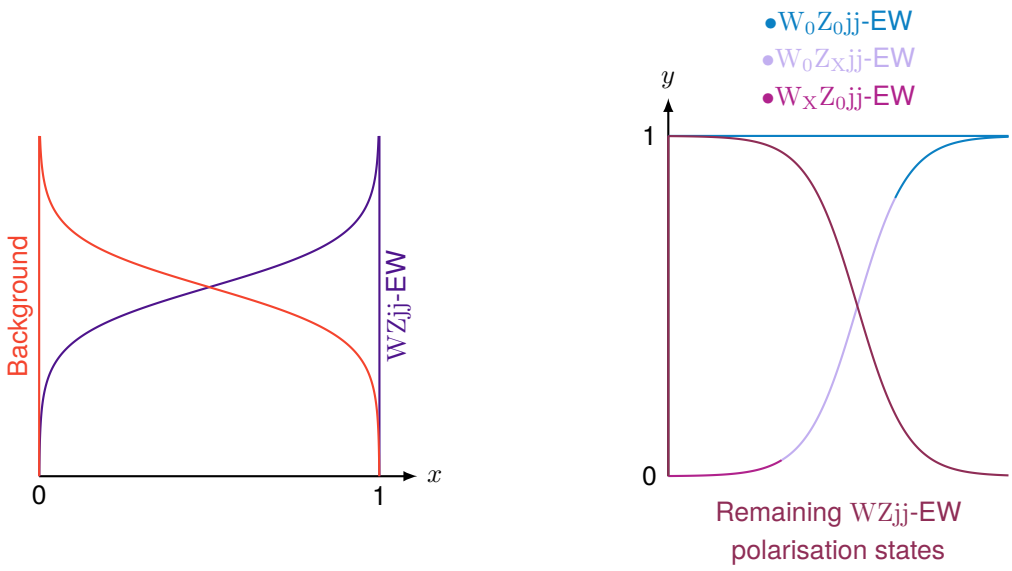
Only the output node corresponding to the WZjj-EW class is used to build the 2D template. The output of the DNN is a continuous variable in the range $[0, 1]$, where a value close to 1 indicates a high probability of being signal-like, and a value close to 0 indicates a high probability of being background-like. A scheme showing the expected output of the S. vs B. DNN is presented in Figure 6.19.

The S. vs B. DNN was first trained on pre-existing MC samples from [360] and then retrained on the MC samples for Run-3 to take into account the differences in the detector and the pile-up conditions.

6.7.3 Polarisation DNN

Three independent DNNs are trained to separate the targeted polarisation from the other polarisation states. The three DNNs are trained to separate respectively the $\mathrm{W_0Z_0jj}$-EW, $\mathrm{W_0Z_Xjj}$-EW, and $\mathrm{W_XZ_0jj}$-EW polarisation from the other polarisations (with the samples generated by MadGraph). Only the WZjj-EW events are used for the training of these DNNs.

They will be labelled as Pol. 00, Pol. $0\mathrm{X}$, and Pol. $\mathrm{X}0$ in the following.



(a) S. vs B.

(b) Pol. 00 / Pol. 0X / Pol. X0

Figure 6.19: Scheme showing the output of the DNNs used in this analysis. The S. vs B. DNN (a) is trained to separate the WZjj-EW signal from the main background, while the Pol. 00 / Pol. 0X / Pol. X0 DNNs (b) are trained to separate the targeted polarisation from the other polarisation states.

The approach is to use a DNN with a single output node with a sigmoid activation function on the output layer, given in Figure 4.4(b), to interpret the output of the DNN as a probability of being in the targeted polarisation class. The targeted values are therefore binary encoded vectors, where the correct class is represented by a 1 and the other classes by 0. The DNN is trained to minimise the BCE loss function, given in Equation 4.6, which measures the difference between the predicted probability and the true binary encoded vector. This approach allows the DNN to learn to distinguish between the targeted polarisation and the other polarisation states effectively.

No negative weights are present in the WZjj-EW samples, since they are generated at LO, but a rescaling of the MC weights is done to fix every polarisation state to match the higher number of events in the W_TZ_Tjj-EW polarisation state. This allows one to put more emphasis on the targeted polarisation events during the training.

The Pol. 00, Pol. 0X, and Pol. X0 DNNs are trained to maximise the weighted AUC metric, which measures the model's ability to distinguish between the classes while accounting for the imbalance between the smaller signal class and the larger background classes, providing a more informative evaluation of performance than plain accuracy. Unlike accuracy, it does not depend on the choice of a classification threshold, making it a robust metric for evaluating a single-class model. The ROC curve and the AUC metric are presented in Section 4.1.7.4.

Only the output node corresponding to the targeted polarisation class is used to build the 2D template. The output of the DNN is a continuous variable in the range $[0, 1]$, where a value close to 1 indicates a high probability of being the targeted polarisation, and a value close to 0 indicates a high probability of being the opposite polarisation.

A scheme showing the expected output of the Pol. 00, Pol. 0X, and Pol. X0 DNNs is presented in Figure 6.19(b). The DNN is expected to output a value close to 1 for the targeted polarisation events, and a value close to 0 for the other polarisation events. The distribution of the DNN output for the targeted polarisation and the other polarisation events is expected to be well separated, allowing for effective discrimination between the two classes.

The Pol. 00, Pol. 0X, and Pol. X0 DNNs were trained only on the MC samples for Run-2 and not retrained on the MC samples for Run-3. We saw in Table 6.E that the polarisation fractions are similar between Run-2 and Run-3, and therefore the DNNs trained on Run-2 samples are expected to perform similarly on Run-3 samples. This will be further confirmed in Section 6.7.5.

6.7.4 Input features

In Section 6.2.3 we present a comprehensive list of discriminating variables that can be used in different ways in order to retain information either on the bosons, jets, or leptons and either separate the signal

from the background or the targeted polarisation from the other polarisation states. The complete list of discriminating variables is presented in Table 6.I. They can be visualised with all the samples considered in this analysis in the DNN baseline and the fit baseline as stacked histograms to present the data and MC comparison in Appendix I. Good overall agreement is observed between data and MC for the discriminating variables used in the DNNs.

The correlation between the input features is presented in Appendix J for the dataset used for the DNN training. The S. vs B. DNN uses both signal and background while the Pol. 00, Pol. $0\mathrm{X}$, and Pol. $\mathrm{X}0$ DNNs use only the polarised samples which therefore have a common correlation matrix. Some variables are highly correlated, such as those from the decay of the bosons and their parent bosons, as expected. However, the DNN is able to handle these correlations and still learn to discriminate between different polarisation states effectively. The correlation matrix also shows that some variables are not strongly correlated with others, which can provide complementary information to the DNN for improving its performance. The first Pol. 00 DNN seems to have less correlated variables than the other two DNNs, which can be explained by the fact that it is trained to separate a more specific polarisation state, the joint longitudinally polarised state, which has more specific kinematic features than the single polarisation states. Variables on both bosons are needed, and thus some variables cannot be correlated as they are related to different bosons. On the other hand, the single polarisation DNNs are trained to separate a more general polarisation state, which can have more similar kinematic features to other polarisation states, leading to more correlated variables. This is well reflected in the correlation matrix where the single polarisation DNNs have more correlated variables than the joint polarisation DNN. Overall, the correlation matrix confirms that the selected input variables for the polarisation DNNs are not redundant and can provide useful information for the DNN to learn from.

The input features are selected based on their discriminating power and their relevance to the specific classification task. For this we trained various DNNs with different sets of input features and evaluated their performance using the metrics described above. The ranking of the input features is done using the three methods presented in Section 4.4. They can be found in Appendix K, with the ranking of the full set of variables in Section K.1, the pruned set of variables in Section K.2, and the ROC curve assessing the performance of the DNN with the full and pruned set of variables in Section K.3.

The final set of input features for each DNN is selected based on the performance metrics and the ranking of the input features. Only the overall top 25 input features across the three methods are retained for each DNN to avoid overfitting and to ensure that the model is not overly complex. Then a further pruning of the input features is done to retain only the most relevant features for each DNN by removing poorly ranked features and observing the impact on the ROC curve and the AUC metric. The final set of input features for each DNN is presented in Table 6.J.

The ranking methods agree on the most discriminating variables, namely m_{jj}, N_{jets}, and $\Delta\eta_{\mathrm{jj}}$. This is expected, as these are the defining observables of the VBS topology: the colour-singlet exchange produces two forward tagging jets separated by a large rapidity gap, and hence a large dijet invariant mass, while the absence of colour flow across the gap leaves it sparsely populated and keeps the jet multiplicity low. The centrality variables ζ_ℓ and A_y^{WZ} are likewise ranked highly by all methods, reflecting the central production of the WZ system and of its decay leptons, in contrast to the more forward and less balanced QCD background. The methods differ only among the lower-ranked variables, whose effect on the DNN is more subtle and partly absorbed through correlations. The agreement therefore gives confidence in the input set chosen for the S. vs B. DNN, which relies predominantly on the tagging-jet and topological observables.

For the polarisation DNNs, the selected variables follow the boson whose polarisation is being measured. As a massive boson is reconstructed only through its decay products, the polar decay angle $\cos\theta^*$ in the boson rest frame carries the angular distribution that separates the longitudinal and transverse states, and it is the variable on which the DNNs concentrate. The Pol. 00 DNN keeps both $\cos\theta^*_{\mathrm{W}}$ and $\cos\theta^*_{\mathrm{Z}}$, whereas each single-boson DNN keeps only the angle of its target. The same logic extends to the leptonic inputs: the W-polarisation Pol. $0\mathrm{X}$ DNN draws on the W-decay kinematics: $p_{\mathrm{T}}^{\ell_{\mathrm{W}}}$, $E_{\mathrm{T}}^{\mathrm{miss}}$ from the neutrino, and the opening angle $\Delta\phi(\ell_{\mathrm{W}},\nu_\ell)$, while the Z-polarisation Pol. $\mathrm{X}0$ DNN relies on the two Z leptons through their transverse momenta $p_{\mathrm{T}}^{\ell_{\mathrm{Z}_1}}$, $p_{\mathrm{T}}^{\ell_{\mathrm{Z}_2}}$, and pseudorapidities. Each DNN thus inherits the decay kinematics of the boson it targets, the clearest separation appearing in the polar angle.

Beyond these target-specific inputs, the boson transverse momenta $p_{\mathrm{T}}^{\mathrm{W}}$, $p_{\mathrm{T}}^{\mathrm{Z}}$, and $p_{\mathrm{T}}^{\mathrm{WZ}}$, together with $m_{\mathrm{T}}^{\mathrm{WZ}}$, are ranked highly by all the polarisation DNNs, since the polarisation fractions depend on the boson energies. The tagging-jet variables, which describe the production rather than the decay, are generally ranked lower for the polarisation DNNs. The exception is $\Delta R(\mathrm{j}_1,\mathrm{Z})$, which retains discriminating power and is selected by all three. As for the S. vs B. case, the agreement between the ranking methods gives confidence in the input variables chosen for the polarisation DNNs.

Table 6.I: All variables accessible.

Category	Symbol	Description	Unit
Event / Topology			
	N_{jets}	Multiplicity of jets	—
Tagging Jets / VBS System			
	$E^{\text{j}_{1,2}}$	Energy of the tagging jets	GeV
	$p_{\text{T}}^{\text{j}_{1,2}}$	Transverse momentum of the tagging jets	GeV
	$\eta^{\text{j}_{1,2}}$	Pseudorapidity of the tagging jets	—
	$\phi^{\text{j}_{1,2}}$	Azimuthal angle of the tagging jets	—
	m_{jj}	Invariant mass of the dijet system	GeV
	$\Delta\eta_{\text{jj}}$	Pseudorapidity difference between the two tagging jets	—
	$\Delta\phi_{\text{jj}}$	Azimuthal-angle difference between the two tagging jets	—
	$\Delta R(\text{j}_1, \text{Z})$	Separation in the η–ϕ plane between the Z and the leading tagging jet	—
	$\langle p_{\text{T}}^{\text{jj}} \rangle$	Average transverse momentum of the two tagging jets	GeV
	$A_{p_{\text{T}}}^{\text{jj}}$	Transverse-momentum asymmetry of the two tagging jets	—
	$R_{p_{\text{T}}}^{\text{hard}}$	Ratio of the vectorial to scalar p_{T} sum of the six hard objects	—
Boson / Diboson Kinematics + MET			
	$p_{\text{T}}^{\text{W}}, p_{\text{T}}^{\text{Z}}$	Transverse momentum of the W and Z bosons	GeV
	p_{T}^{WZ}	Transverse momentum of the WZ system	GeV
	r_{21}	Relative transverse momentum of the two bosons, $\frac{\text{min}}{\text{max}}$	—
	$m_{\text{T}}^{\text{W}}, m_{\text{T}}^{\text{WZ}}$	Transverse mass of the W boson and WZ system	GeV
	$E_{\text{T}}^{\text{miss}}$	MET	GeV
Lepton Kinematics			
	$p_{\text{T}}^{\ell_{\text{Z}_1,\text{Z}_2,\text{W}}}$	Transverse momentum of the three leptons (leading, subleading of Z, and the W)	GeV
	$\eta^{\ell_{\text{Z}_1,\text{Z}_2,\text{W}}}$	Pseudorapidity of the three leptons (leading, subleading of Z, and the W)	—
	$\phi^{\ell_{\text{Z}_1,\text{Z}_2,\text{W}}}$	Azimuthal angle of the three leptons (leading, subleading of Z, and the W)	—
Angular / Polarisation-Sensitive Observables			
	$\cos\theta_{\text{W}}^*, \cos\theta_{\text{Z}}^*$	Cosine of the W/Z decay polar angles	—
	$\lvert\cos\theta_V\rvert$	Absolute cosine of the boson production angle θ_V in the WZ CM frame	—
	$\lvert\cos\theta_{V\text{jj}}\rvert$	Absolute cosine of the boson production polar angle in the rest frame of the WZjj system (tagging jets included)	—
	$\phi_{\text{W}}^*, \phi_{\text{Z}}^*$	Azimuthal angle of the W and Z decays in their rest frame	—
	$\Delta\phi(\ell_{\text{Z}_1}, \ell_{\text{Z}_2})$	Azimuthal-angle difference between the two Z leptons	—
	$\Delta\phi(\ell_{\text{W}}, \nu_\ell)$	Azimuthal-angle difference between the W lepton and the neutrino	—
Centralities / Rapidity Observables			
	ζ_{jets}	Centrality of the tagging jets	—
	ζ_ℓ	Lepton centrality, smaller pseudorapidity gap between the leptons and the tagging jets	—
	A_y^{WZ}	Rapidity separation between the Z and the W lepton	—
	$z_{3\ell}$	Zeppenfeld variable of the leptons	—

6.7.5 Evaluation of performance

The performance of the DNNs, despite the various metrics, is truly evaluated using their significance in dedicated individual template fits. More details on how a template fit is performed can be found in Appendix O.

The ROC curves, the accuracy, and other useful metrics are monitored during the training, as well as the loss function for both training and validation samples to check the convergence and the absence of overfitting. However, they do not reflect the true performance of how well the DNN will perform in the final fit when combining them together to form the template.

To evaluate the performance of the DNN we perform individual fits to measure the significance of the model in performing its task: either separating signal against background or splitting polarisation states. For every fit the regions are defined as in Table 6.D and the same samples as for the DNN training are used. We carry out a binned fit in the signal and background regions on a purely Asimov dataset. However, there are differences in the evaluation of the individual performance between the two types of DNNs: S. vs B. and Pol. XX.

S. vs B. DNN. The fit is done using the SR as well as the two CRs. All the backgrounds of the main fit are used. The signal is the full WZjj-EW only. The fakes use the MC samples in the SR and ZZ-CR while in the b-CR we use the data-driven estimate (computing the data-driven estimate is time-consuming and doing it for every trial of a DNN is not manageable). The parameter of interest (POI) is $\mu_{\mathrm{WZjj\text{-}EW}}$ and the other NFs are constant and are the following: $\mu_{\mathrm{WZjj\text{-}QCD}}$, $\mu_{\mathrm{ZZjj\text{-}QCD}}$, $\mu_{\mathrm{t\bar{t}}+V}$, μ_{tZj}.

Pol. XX DNN. The fit is done using the SR only. We use the four polarised WZjj-EW samples as signal and all the backgrounds depending on the joint or single polarisation fit. A floating NF is associated with the number of WZjj-EW events: $N_{\mathrm{WZjj\text{-}EW}}$. Another NF is the fraction of polarisation of the signal label f_{XX} and associated to the signal and "background" (the non-targeted polarisation states) as:

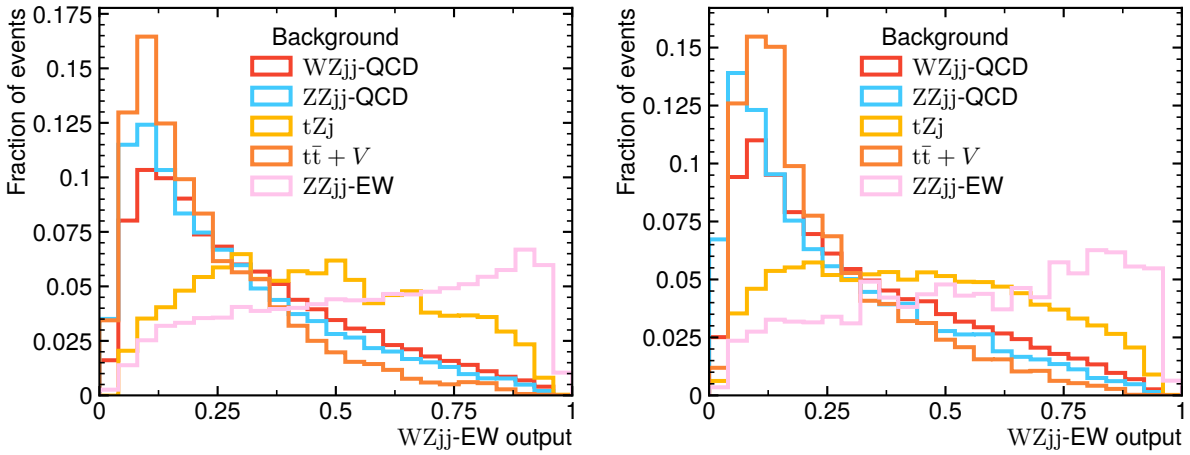
$$\mu_{\mathrm{signal}} = \frac{N_{\mathrm{WZjj\text{-}EW}} \cdot f_{\mathrm{XX}}}{N^{\mathrm{MC}}_{\mathrm{signal}}}, \tag{6.138}$$

$$\mu_{\mathrm{bkg}} = \frac{N_{\mathrm{WZjj\text{-}EW}} \cdot (1 - f_{\mathrm{XX}})}{N^{\mathrm{MC}}_{\mathrm{bkg}}}. \tag{6.139}$$

Table 6.J: Input variables used for the training of the different DNN classifiers. A mark ✔ (✘) indicates that the variable is (is not) used for a specific DNN.

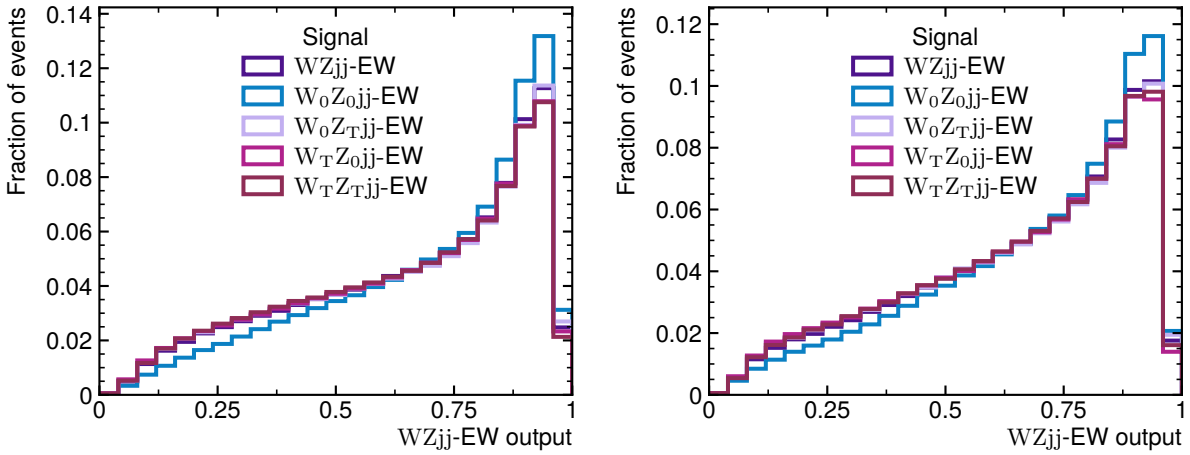
Symbol	S. vs B.	Pol. 00	Pol. $0\mathrm{X}$	Pol. $\mathrm{X}0$
Event / Topology				
N_{jets}	✔	✘	✘	✘
Tagging Jets / VBS System				
E^{j_1}	✔	✘	✘	✘
$p_{\mathrm{T}}^{\mathrm{j}_1}$	✔	✘	✘	✔
$p_{\mathrm{T}}^{\mathrm{j}_2}$	✘	✘	✘	✔
η^{j_1}	✘	✔	✔	✔
η^{j_2}	✔	✔	✔	✘
ϕ^{j_2}	✘	✘	✔	✘
m_{jj}	✔	✘	✘	✘
$\Delta\eta_{\mathrm{jj}}$	✔	✔	✔	✔
$\Delta\phi_{\mathrm{jj}}$	✔	✔	✔	✘
$\Delta R(\mathrm{j}_1, \mathrm{Z})$	✘	✔	✔	✔
$\langle p_{\mathrm{T}}^{\mathrm{jj}} \rangle$	✘	✘	✘	✔
$A_{p_{\mathrm{T}}}^{\mathrm{jj}}$	✔	✔	✔	✘
$R_{p_{\mathrm{T}}}^{\mathrm{hard}}$	✔	✘	✘	✘
Boson / Diboson Kinematics + MET				
$p_{\mathrm{T}}^{\mathrm{W}}$	✔	✔	✔	✔
$p_{\mathrm{T}}^{\mathrm{Z}}$	✘	✔	✔	✔
$p_{\mathrm{T}}^{\mathrm{WZ}}$	✔	✔	✔	✔
r_{21}	✘	✔	✘	✘
$m_{\mathrm{T}}^{\mathrm{WZ}}$	✘	✔	✔	✔
$E_{\mathrm{T}}^{\mathrm{miss}}$	✘	✘	✔	✔
Lepton Kinematics				
$p_{\mathrm{T}}^{\ell_{\mathrm{Z}_1}}$	✘	✔	✘	✔
$p_{\mathrm{T}}^{\ell_{\mathrm{Z}_2}}$	✘	✘	✘	✔
$p_{\mathrm{T}}^{\ell_{\mathrm{W}}}$	✘	✘	✔	✔
$\eta^{\ell_{\mathrm{Z}_1}}$	✔	✔	✔	✔
$\eta^{\ell_{\mathrm{Z}_2}}$	✘	✔	✘	✔
$\eta^{\ell_{\mathrm{W}}}$	✔	✔	✔	✔
Angular / Polarisation-Sensitive Observables				
$\cos\theta^*_{\mathrm{W}}$	✘	✔	✔	✘
$\cos\theta^*_{\mathrm{Z}}$	✘	✔	✘	✔
$\lvert\cos\theta_V\rvert$	✔	✔	✔	✘
$\Delta\phi(\ell_{\mathrm{Z}_1}, \ell_{\mathrm{Z}_2})$	✔	✔	✘	✘
$\Delta\phi(\ell_{\mathrm{W}}, \nu_\ell)$	✔	✘	✔	✘
Centralities / Rapidity Observables				
ζ_ℓ	✔	✘	✘	✘
A_y^{WZ}	✔	✔	✔	✔
$z_{3\ell}$	✘	✘	✘	✔

The final output distribution for each DNN is shown in Figure 6.20 for the S. vs B. DNN and in Figure 6.21 for the polarisation DNNs using the full k-fold splitting. The output distributions are shown for both Run-2 and Run-3 samples. The S. vs B. DNN output distributions show a clear separation between the signal and background contributions, with the signal peaking at high values of the output score and the background peaking at low values. The polarisation DNNs output distributions show a relatively less stringent separation as the polarisation is only seen through some specific variables on a statistical basis.



(i) Run-2 **(ii)** Run-3

(a) Background



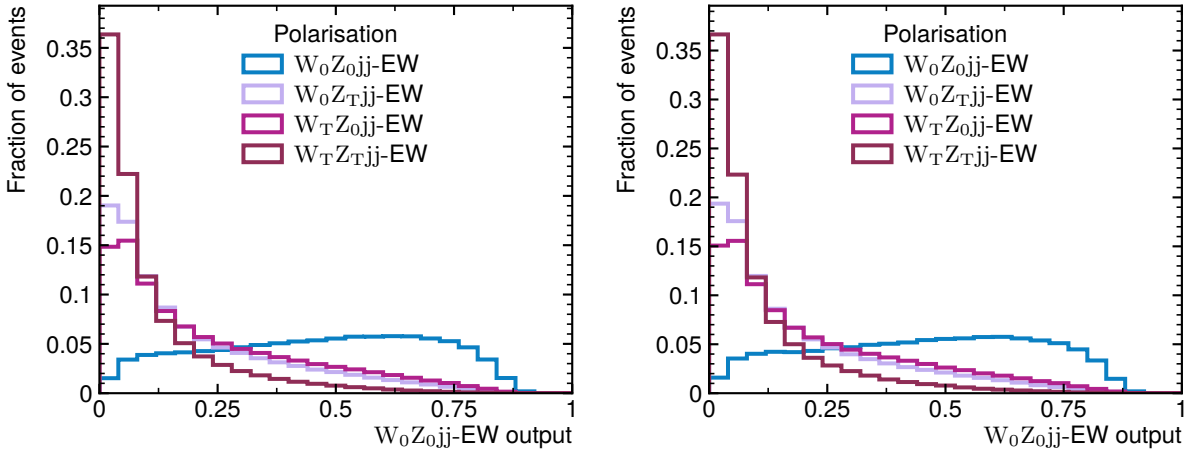
(i) Run-2 **(ii)** Run-3

(b) Signal

Figure 6.20: Output score of the S. vs B. DNN, for the background (a) and the signal contributions (b), each shown for Run-2 (i) and Run-3 (ii).

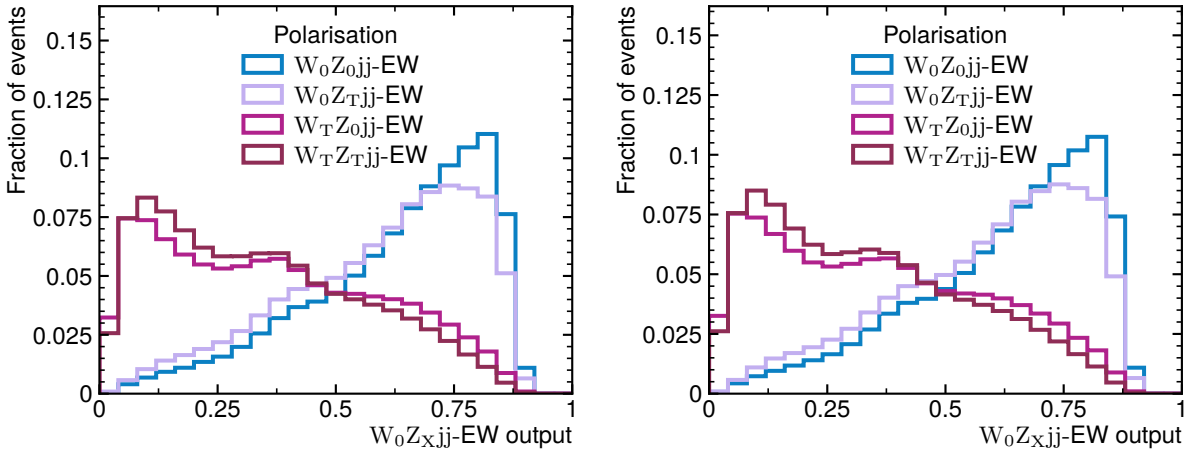
The expected significances obtained for each DNN category and for each fold are shown in Table 6.K for Run-2 and Run-3. For each DNN, the columns labelled by the fold number correspond to fits performed using the response of a single trained DNN. In these columns, the templates are built from the output of one fold only: the events belonging to the validation subset of this fold are evaluated independently, while the remaining events correspond to events that were used during the training of this particular DNN. These fold-by-fold results are therefore mainly used as stability checks of the training procedure and should not be interpreted as independent final results. This consistency between the folds can also be seen on the ROC curve in Figures K.11 to K.13.

These individual fits use the output distribution of each network alone as template, with a simplified uncertainty model, so their significances are comparative figures of merit only and are not directly comparable to those of the final Voronoi-binned fits (Section 7.3). The column labelled “All” corresponds to the final fold-combination strategy used in the analysis. In this case, the templates are obtained by combining the five validation folds. Consequently, each event entering the final template is evaluated by a DNN that was not trained on that event. This provides a template based on the full available MC statistics, while avoiding possible over-training biases in the DNN response. For the processes whose fit sample is the one used in the training (ZZjj-QCD, $t\bar{t}+V$, tZj), the template is thus built on the validation folds. The polarised WZjj-EW templates and, in Run-2, the WZjj-QCD template are instead built from Sherpa samples while the training used MadGraph, and the ZZjj-EW, VVV, and WZjj-INT processes do not enter the training at all. These events, each scored by the DNN of the fold given by its event number, were never seen by the networks and form an independent test dataset. The same holds for the Run-3 outputs of the polarisation DNNs, which reuse the Run-2 trainings.



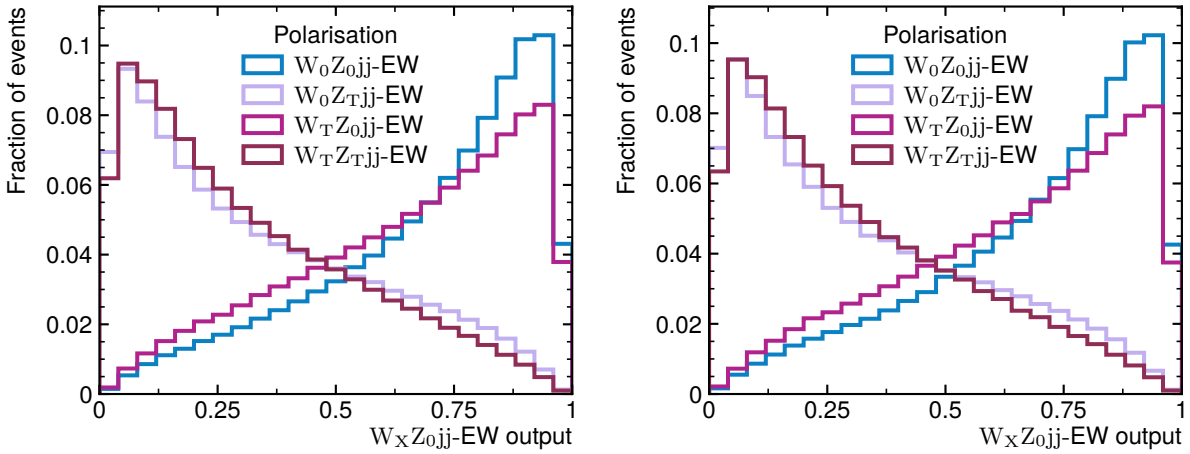
(i) Run-2 **(ii)** Run-3

(a) Pol. 00



(i) Run-2 **(ii)** Run-3

(b) Pol. $0X$



(i) Run-2 **(ii)** Run-3

(c) Pol. $X0$

Figure 6.21: Output score of the Pol. XX DNNs, for the Pol. 00 (a), Pol. $0X$ (b), and Pol. $X0$ (c) targeted polarisation states, each shown for Run-2 (i) and Run-3 (ii).

With the same individual-fit setup, the expected significance of the WZjj-EW signal against the background obtained with the BDT used in [360], to be compared with the S. vs B. rows of Table 6.K, and the expected significance of the doubly longitudinal state obtained with the previous DNN used in [11], to be compared with the Pol. 00 rows, are respectively 9.35 (8.62) and 0.50 (0.52) for Run-2 (Run-3). The new DNNs therefore show a significant improvement with respect to these previous approaches.

For the polarisation DNNs, the comparison with the previous DNN of [11] should however be interpreted with care. As shown in [413], the polarisation composition of inclusive WZ production and EW WZjj production can be significantly different. This difference explains part of the improvement observed for the polarisation DNNs, since the previous DNN was trained using inclusive WZ as signal, whereas the DNNs used in this analysis are trained on the EW WZjj process relevant for the measurement.

Table 6.K: Significance obtained for each discriminant and their folds on individual fits for Run-2 and Run-3. The column labelled "All" corresponds to the final fold-combination strategy used in the analysis.

Type	Trained on	Applied on	Fold					
			0	1	2	3	4	All
S. vs B.	Run-2	Run-2	9.77	9.73	9.73	9.77	9.63	9.68
		Run-3	9.21	9.10	9.17	9.21	9.07	9.09
	Run-3	Run-2	9.89	9.77	9.87	9.88	9.87	9.76
		Run-3	9.47	9.41	9.36	9.35	9.48	9.34
Pol. 00	Run-2	Run-2	1.66	1.65	1.65	1.65	1.65	1.63
		Run-3	1.73	1.73	1.73	1.73	1.72	1.73
Pol. $0\mathrm{X}$	Run-2	Run-2	3.41	3.39	3.44	3.43	3.42	3.37
		Run-3	3.59	3.58	3.61	3.61	3.61	3.60
Pol. $\mathrm{X}0$	Run-2	Run-2	4.42	4.40	4.39	4.44	4.41	4.33
		Run-3	4.67	4.66	4.66	4.67	4.66	4.66

These results also demonstrate the benefit of training separate S. vs B. DNNs for Run-2 and Run-3. A significant improvement in performance is observed when the DNN is trained separately for each run period, reflecting the different detector conditions and event characteristics.

For the polarisation DNNs, the improvement observed in Run-3 is mainly driven by the increase in the expected signal yield rather than by a change in the DNN response itself. Therefore, no dedicated retraining of the polarisation DNNs for Run-3 was considered necessary. Reusing the Run-2 polarisation trainings for Run-3 also avoids the need to train an additional set of 15 DNNs, corresponding to three polarisation DNN categories times five folds.

The main conclusion of this study is that the new S. vs B. DNN outperforms the previous BDT analysis in terms of significance. In addition, a dedicated training for Run-3 of this DNN led to a significant improvement in the significance compared to using the Run-2 DNN for Run-3, which is consistent with the observed differences in the input variable distributions between both datasets in Appendix I. For the polarisation DNNs, the resulting significances are consistent between the two runs and there is no need for a dedicated retraining for Run-3 as the response of the DNNs trained on Run-2 is insensitive to the differences between both datasets and they still perform well on both of them. A very slight drop in performance is observed when using the full fold, as no events used during training are evaluated. The use of the k-fold method ensures there is no over-training bias in the final template.

6.7.6 Sanity checks

Various checks are done to ensure the robustness of the DNN. The comparison between training and validation datasets is monitored during each training, and they show consistent agreement between both datasets, indicating no over-training issue, which confirms that the various precautions such as dropout, early stopping, or weight decay mentioned above are effective in preventing over-training.

Other tests are done to ensure that the DNN is not learning generator-specific features. The MadGraph and Sherpa generators are used to produce the WZjj-EW and WZjj-QCD samples, and a polarisation-sum sample for the signal is used rather than the individual polarised samples used later in the fit. These checks are described in Appendix L. The difference between the polarisation-sum WZjj-EW samples used for training in the S. vs B. DNN and the full sample is around $2\,\%$. The full sample also shows a difference of $2\,\%$ to $4\,\%$ with the per-polarisation samples, and up to $12\,\%$ with the $\mathrm{W_0Z_0jj}$-EW sample, which is expected as this polarisation state is less represented in the full WZjj-EW sample. A difference

of only $5\,\%$ for WZjj-QCD and of $1\,\%$ to $3\,\%$ for the polarised WZjj-EW samples is observed between the MadGraph and Sherpa samples, allowing us to interchange the two generators for the baseline without worrying about a potential bias. These observations are valid for both Run-2 and Run-3 datasets.

Changing the baseline generator for the DNN and the k-fold method ensure that the DNN is not learning generator-specific features and thus can be used to discriminate the polarised states in the template without worrying about a potential bias due to the difference between the polarisation-sum WZjj-EW sample used for training and the polarised ones used in the template, or due to differences between generators.

In the Run-2 panel of Figure 6.21(a), the output of the Pol. 00 DNN does not fill the full range $[0, 1]$ allowed by the sigmoid activation function. This comes from the class weights used in the training. With the absolute values of the MC weights, the $\mathrm{W_0Z_0jj}$-EW, $\mathrm{W_0Z_Tjj}$-EW, and $\mathrm{W_TZ_0jj}$-EW samples are each normalised to the yield of the $\mathrm{W_TZ_Tjj}$-EW sample, so that the four states have the same total weight. This results in the "flat" distribution of Figure 6.22. Two alternatives are also shown in this figure: normalising the $\mathrm{W_0Z_0jj}$-EW sample to the sum of the three other states (light purple), or first equalising the four states and then normalising the $\mathrm{W_0Z_0jj}$-EW sample to the sum of the three others (magenta). Both stretch the output to the full $[0, 1]$ range with a sharper peak at 1, but they also push more $\mathrm{W_TZ_Tjj}$-EW events, the main background of this measurement, to high output values, which lowers the significance of the final fit. The equal-yield weighting is therefore kept. Therefore, the final decision was to use the first reweighting procedure, resulting in a DNN output not filling the full range but with better separation between the targeted polarisation state and thus a better final fit significance. By comparison, without any reweighting and taking directly the MC weights, the DNN output behaves as a background as it is completely dominated by the other polarisations. A reweighting procedure is thus necessary.

6.7.7 Discriminant plane and binning

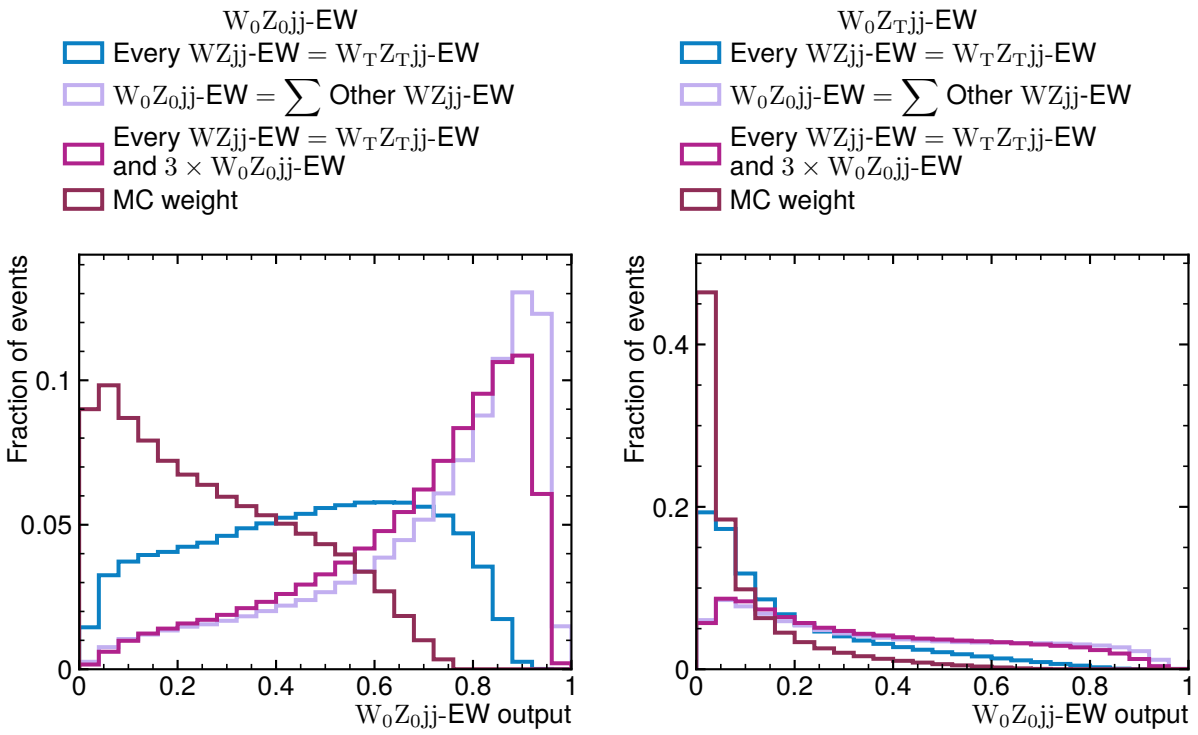
The final discriminant planes are built using the output of the S. vs B. DNN only with the WZjj-EW output (giving us the probability of having the signal) on the x-axis and the output of one of the polarisation DNNs on the y-axis (giving us the probability of having the desired polarisation state). The other outputs of the S. vs B. DNN are not used in the final fit. The expected plane and the repartition of the signal and background are schematically shown in Figure 6.23. The signal is expected to be concentrated in the top right corner of the plane, while the background is expected to be concentrated in the bottom left corner. The bottom right corner is expected to be the signal WZjj-EW but with the non-targeted polarisation states. The task presented in this section is to find the best method to accommodate this particular repartition of the plane and enhance the sensitivity for the targeted polarisation state while maintaining enough handle to discriminate the background and the remaining polarisation state in the final fit.

The study of the binning was a lengthy and complex task because the vast majority of tools are not designed to increase the significance of a signal versus background but rather to increase the ratio of signal over background or to modify the population of each bin to have a smoother distribution.

The retained method works in two steps. The first part is to form a 2D histogram showing the ratio of signal over the square root of the background ($\frac{S}{\sqrt{B}}$ without MC weights[‡]) using a fine binning (here 75×75 bins). Then we compute the k-means clusters, following Section 4.3.1, on the signal only, using the $\frac{S}{\sqrt{B}}$ value in each bin as a per-event weight. The k-means algorithm is an unsupervised ML method that partitions the data into a specified number of clusters based on their proximity in a feature space. Here the feature space is the plane spanned by the two DNN outputs (the signal-versus-background output and the polarisation output), and the objects being clustered are the signal events at their coordinates in this plane. The $\frac{S}{\sqrt{B}}$ value of the fine bin containing each event enters as its weight, so that signal-rich bins with a large $\frac{S}{\sqrt{B}}$ exert a stronger pull on the cluster centres. In this way the algorithm concentrates the centres in the regions of the discriminant plane with the highest signal significance.

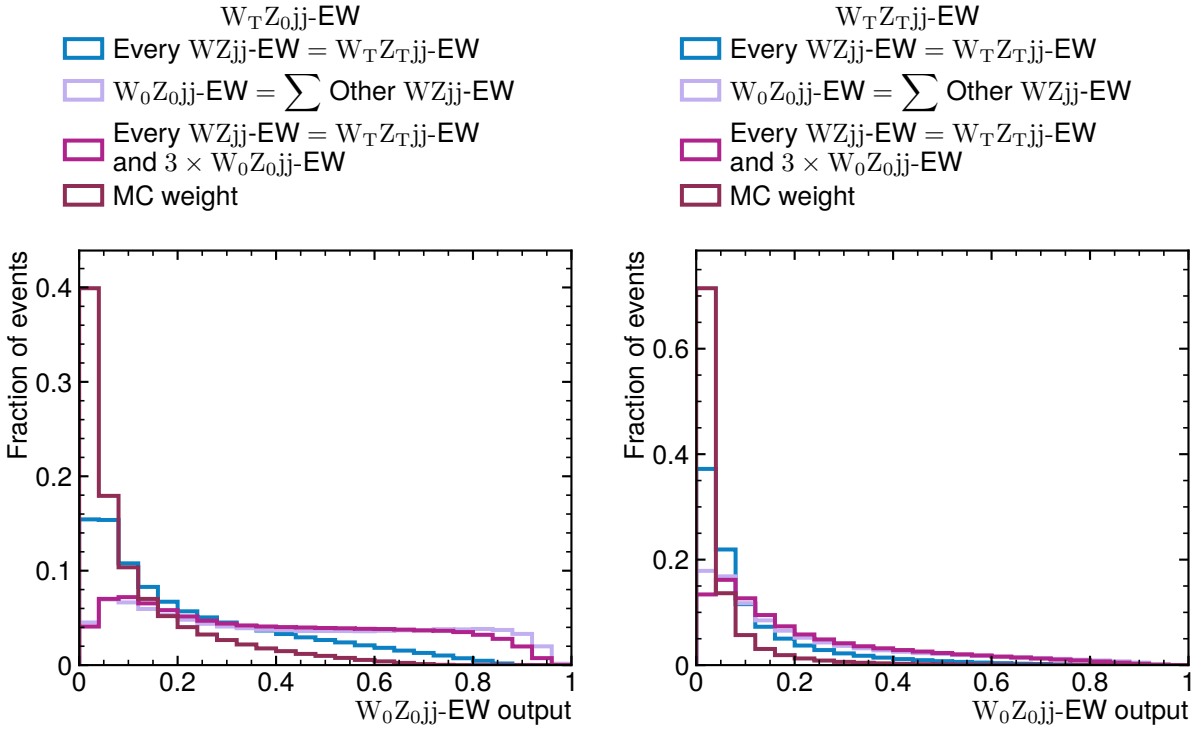
The distance metric used in the k-means algorithm is the Euclidean distance, which measures the straight-line distance between two points in the feature space. In this case, the distance is evaluated between each signal event and the cluster centres in the discriminant plane. Each event is assigned to the cluster with the nearest centre, and the centres are then updated as the weighted mean of their assigned events, using the $\frac{S}{\sqrt{B}}$ weights. This process is repeated until convergence is reached. From the resulting clusters, Voronoi regions naturally arise as seen in Section 4.3.2, which are defined as the set of points that are closer to a particular cluster centre than to any other cluster centre. These Voronoi regions can be used to define the final binning of the plane, where each bin corresponds to a specific Voronoi region.

[‡]The MC weights are homogeneously distributed across the plane, as can be seen in Figures N.1, N.3, and N.5, so applying them would not meaningfully alter the clustering.



(a) W_0Z_0jj-EW

(b) W_0Z_Tjj-EW



(c) W_TZ_0jj-EW

(d) W_TZ_Tjj-EW

Figure 6.22: Pol. 00 DNN output for the four polarised $WZjj$-EW samples in Run-2, with different reweighting procedures applied to the samples.

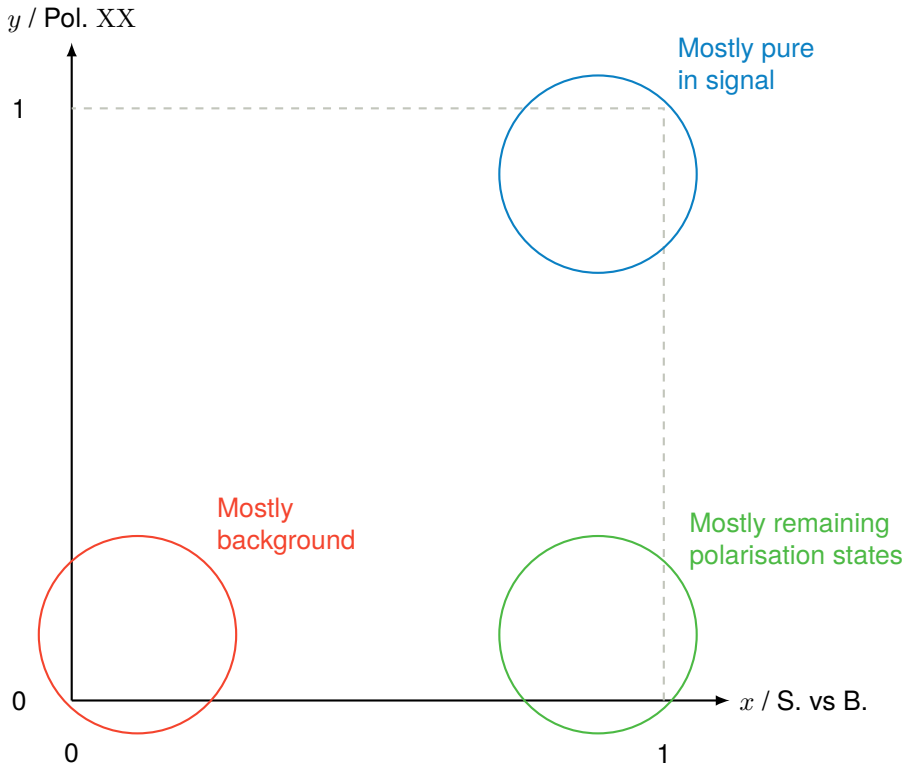


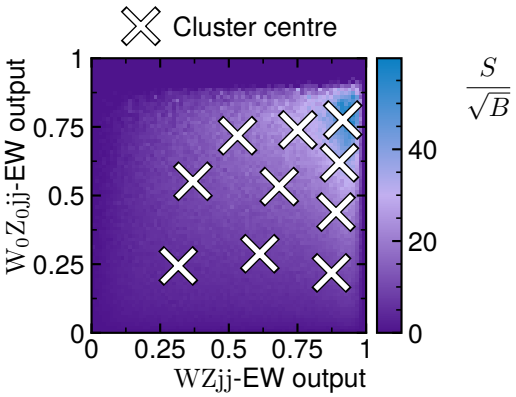
Figure 6.23: Scheme of the discriminant plane built from the outputs of the two DNNs.

The number of clusters is set to 10, presenting a good compromise between having enough bins to exploit the DNN discrimination power and keeping a reasonable number of bins to avoid large statistical uncertainty in each bin. The k-means algorithm thus groups the signal events into 10 clusters in the discriminant plane, weighted by their $\frac{S}{\sqrt{B}}$ value. The result is shown in Figure 6.24 for Run-2.

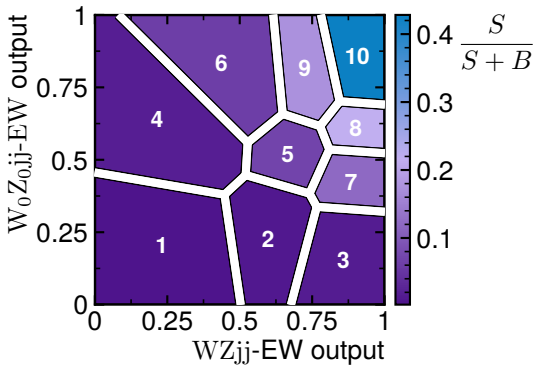
The second step then consists of applying a Voronoi binning using the cluster centres as seeds. The final Voronoi regions based on the plane are given in the right-hand panels of Figure 6.24 for Run-2. The final binning is then applied to the plane to obtain the final templates used in the individual fits. The final binning is also applied to the Run-3 dataset without re-optimisation, as the same binning is used for both Run-2 and Run-3 datasets. The binning is optimised on the Run-2 MadGraph polarised signal and on the background samples of the DNN baseline, each event being scored by the DNN of its fold. For the processes used in the training, these outputs are validation outputs, so the cell boundaries are tuned on the validation folds. The Sherpa templates of the fit are then filled into these fixed cells.

The bins are ordered by increasing $\frac{S}{S+B}$ (with MC weights) of the Voronoi regions. The first bin is the one with the lowest $\frac{S}{S+B}$, the background-dominated corner of the plane that holds most of the events, and the last bin is the one with the highest $\frac{S}{S+B}$. This ordering is done to have a better understanding of the purity of each bin and to be able to read the final templates more easily. This information is given in Figures N.2(b), N.4(b), and N.6(b).

This new method yields a clear improvement, with an increase in expected significance of a few percent over the best rectangular configuration and up to $\sim 11\,\%$ over the worst one, depending on the measured polarisation state. This number arises from the comparison of the significance obtained with this Voronoi binning and the one obtained with a more traditional rectangular binning with random cuts to improve the Poisson significance. The two binning methods were then compared on the same individual fit to ensure a fair comparison. The detailed comparison, for the three polarisation hypotheses and for both Run-2 and Run-3, is presented in Appendix M: at fixed bin multiplicity, the 10-bin Voronoi tessellation systematically matches or exceeds the best rectangular configuration. The rectangular results also depend strongly on the choice of axis ordering, with the 2×5 and 5×2 layouts differing by up to $\sim 7\,\%$ in expected significance, while the Voronoi binning naturally avoids this arbitrariness. It is particularly efficient in increasing the significance while keeping the number of bins low enough. Each bin is always populated with enough events (without weights) both in signal and background to avoid large statistical uncertainty. This Voronoi binning also tackles the reduced output range of the Pol. 00 DNN output, as the bins are not rectangular and can be adapted to the shape of the distribution. Finally, the bins with the highest indices are very pure in signal. Additional plots to help interpret this binning are given in Appendix N.

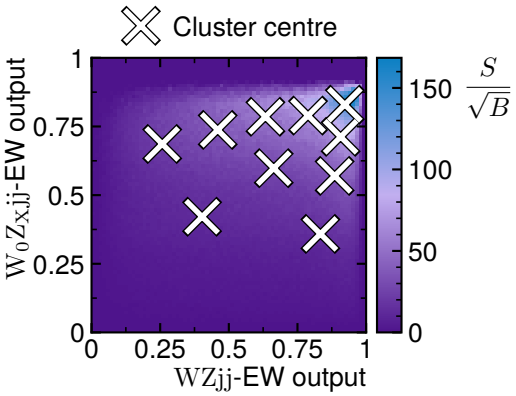


(i) $\frac{S}{\sqrt{B}}$ histogram and k-means clustering

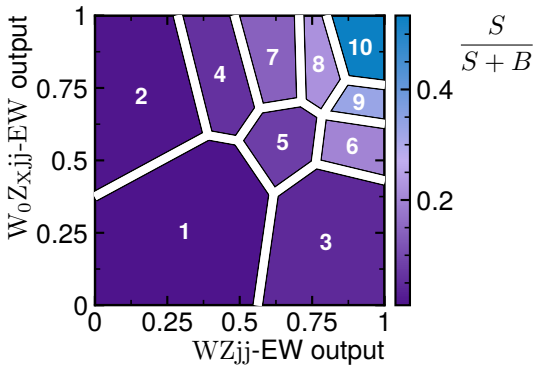


(ii) Voronoi binning, coloured by $\frac{S}{S+B}$

(a) Pol. 00

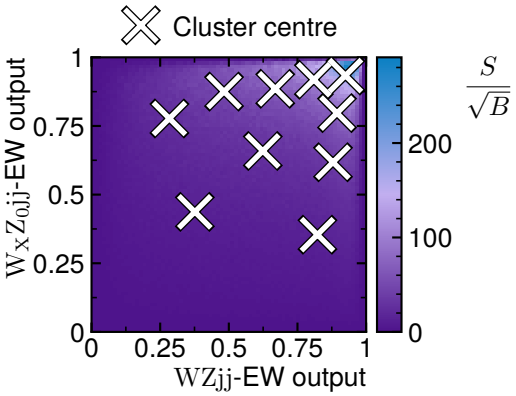


(i) $\frac{S}{\sqrt{B}}$ histogram and k-means clustering

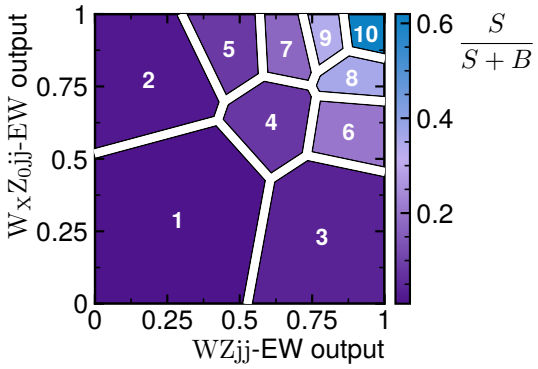


(ii) Voronoi binning, coloured by $\frac{S}{S+B}$

(b) Pol. $0X$



(i) $\frac{S}{\sqrt{B}}$ histogram and k-means clustering



(ii) Voronoi binning, coloured by $\frac{S}{S+B}$

(c) Pol. $X0$

Figure 6.24: Steps of the discriminant plane binning for Run-2, one polarisation flavour per row: Pol. 00 (a), Pol. $0X$ (b), and Pol. $X0$ (c). For each flavour, the ratio $\frac{S}{\sqrt{B}}$ histogram with the k-means clustering and the resulting final Voronoi binning are shown.

7 Polarisation fraction measurement

The polarisation is not a direct observable, but a property of the underlying physics process. It is therefore not possible to measure it directly: it is rather needed to infer it from the kinematic distributions of the final-state particles. The templates we build in Section 6.7 are designed to be sensitive to a given polarisation state, and can be used to extract the associated fractions of polarisation. Three distinct and separate measurements are performed, one for each longitudinal polarisation hypothesis: the joint polarisation $\mathrm{W_0Z_0}$jj-EW, and the two single polarisations $\mathrm{W_0Z_X}$jj-EW and $\mathrm{W_XZ_0}$jj-EW.

The template design is based on the Voronoi tessellation of the discriminant plane formed by the DNN outputs, as described in Section 6.7.7. The templates are built for each polarisation hypothesis, and for each of the three data-taking periods: Run-2, partial Run-3, and their combination.

Until now, all the studies have been performed on simulated data, MC samples, which are used to build the templates. The next step is to perform a fit of these templates to the actual data produced by the pp collisions at the LHC and collected by the ATLAS detector over the Run-2 and Run-3 periods.

The templates are then used to perform a binned maximum-likelihood fit to the data, where the polarisation fractions are treated as free parameters. The fit is performed simultaneously in all bins of the unrolled Voronoi template, in a fit combining both Run-2 and partial Run-3 data. The fit is performed using the `TRExFitter` machinery [420], which builds the likelihood with the `HistFactory` formalism [421].

This chapter aims to describe the fit procedure, the treatment of systematic uncertainties, the correlation scheme of parameters between the different data-taking periods, and the results obtained when the templates are fitted to data.

The formalism of the fit is described in Appendix O, introducing all relevant concepts and definitions for this chapter.

7.1 Parametrisation

The fits are divided into three independent regions of the total phase-space, as described in Section 6.6.2. The three regions are defined by different cuts and requirements to suppress or enrich particular signal or background processes. The three regions are defined as follows:

SR. This region is where the signal is expected to be dominant, and thus where the measurement will be performed. For this region, the fitted variable is the unrolled Voronoi binning of the 2D discriminant plane, as described in Section 6.7.7. "Unrolled" means that the plane is transformed into a 1D histogram, where each bin corresponds to a Voronoi cell following the ordering visible in the right-hand panels of Figures 6.24(a) to 6.24(c).

ZZ-CR. This region is defined to enhance the ZZ (both EW and QCD production modes) background contribution, and is used to constrain the ZZjj-QCD normalisation. A natural choice for the fitted variable in this region is the invariant mass of the two leading jets, m_{jj}, which is a good discriminant between the "softer" ZZjj-QCD background and the "harder" WZjj-EW signal with the transition from one production mode to the other happening around $m_{\mathrm{jj}} \sim 300\,\mathrm{GeV}$.

b-CR. This region is defined to enhance the $\mathrm{t\bar{t}} + V$ and tZj background contributions, meaning the background most susceptible to producing events with a b-jet which is vetoed in the SR and easier to tag. The fitted variable in this region is a dedicated BDT trained to discriminate between the $\mathrm{t\bar{t}} + V$ and tZj backgrounds only, putting the tZj events at unity and the $\mathrm{t\bar{t}} + V$ events at minus unity. The BDT was trained in [360].

The MC samples are used to build the templates for each of the fitted variables in the three regions. The list of the various samples is given in Appendix F. The fit baseline differs slightly from the one used for the training of the DNN. The changes with respect to the DNN baseline, and the distribution of the samples across the regions, are the following:

- The full WZjj-EW Sherpa sample is used (without separation of the polarisation states and including off-shell component and tau lepton decay) in the CRs as signal since it has a small contribution in these regions.
- The various polarised WZjj-EW samples are used with Sherpa compared to the DNN training where MadGraph was used. Appendix L assesses the difference between the two generators and shows that the impact on the templates is negligible. The Sherpa production was available at a later stage and is used in the fit as it includes the emission of an additional jet at LO (0 and 1 jet merged) and is thus more accurate.
- The WZjj-EW in τ decay mode is included in the fits as a distinct component in the SR, obtained with Sherpa as the difference between the full WZjj-EW sample and the polarisation-sum sample. It is treated as a background: the leptonic τ decay brings additional neutrinos, which make the event harder to reconstruct and therefore its polarisation harder to identify. The τ decay mode is not included in the CRs since it is included in the full WZjj-EW sample.
- The WZjj-QCD sample is modelled using Sherpa instead of MadGraph (in Run-2 only) as in the DNN training. The impact of this change is assessed in Appendix L and shows that the impact on the templates is small (about $5\,\%$ on the DNN output shape).
- For the misidentified-lepton background: $\mathrm{Z}+\mathrm{jets}$, $\mathrm{t\bar{t}}$, and $\mathrm{Z}\gamma$ use the matrix-method estimate in their dominant region, the SR and b-CR, while in the ZZ-CR, where their contribution is small, the MC prediction corrected with the scale factors of Section E.2 is used. Both methods are described in Appendix E.
- ZZjj-QCD uses the Sherpa sample in both the fit and the DNN baseline.
- $\mathrm{t\bar{t}}+V$ and tZj use the same samples in both the fit and the DNN baseline.
- ZZjj-EW uses Sherpa for Run-3 and MadGraph for Run-2.
- The VVV uses Sherpa for both runs.
- The WZjj-INT is modelled by MadGraph for both runs.

The inputs of the DNN are presented in Appendix I where both stacked distributions with data and the MC samples are shown, as well as only the shaped distributions of the MC samples. Both baselines are shown for comparison.

The pre-fit distributions in the two CRs are shown in Figures 7.1(a) and 7.1(b) for the Run-2 dataset, and in Figures 7.3(a) and 7.3(b) for the partial Run-3 dataset. They are common for the three polarisation measurements. The SR distribution of the unrolled Voronoi binning is shown in Figures 7.2(a) to 7.2(c) for the Run-2 dataset, and in Figures 7.4(a) to 7.4(c) for the partial Run-3 dataset. Each of these distributions corresponds to a different polarisation measurement. The uncertainty band around the SM expectation includes all systematic uncertainties at their pre-fit values.

The associated yields of all samples were already provided in Table 6.F.

The POIs of the fit are the ratio of data with respect to the MC expectation for the fiducial WZjj-EW production cross-section, labelled as $\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}}$, and the polarisation fractions f_{00}, $f_{0\mathrm{X}}$, and $f_{\mathrm{X}0}$. As we have three different f_{XX} for three different measurements, the POI definition changes from one measurement to another.

The observed data cross-section is defined as:

$$\sigma^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}} = \mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}} \cdot \frac{N^{\mathrm{fid.,MC}}_{\mathrm{WZjj\text{-}EW}}}{4\mathcal{L}}\,, \tag{7.1}$$

with $N^{\mathrm{fid.,MC}}_{\mathrm{WZjj\text{-}EW}}$ the total number of events in the fiducial phase-space for the WZjj-EW process (all polarisation states included but excluding the τ decay mode and off-shell effects) and $\mathcal{L}$ the integrated luminosity of the dataset. The factor 4 is included to account for the four decay channels, eee, $\mathrm{e\mu\mu}$, $\mathrm{\mu ee}$, and $\mathrm{\mu\mu\mu}$. We consider lepton universality and thus the same cross-section for all four channels.

The fiducial polarisation fraction is thus defined as the ratio of the number of events in a given polarisation state to the total number of WZjj-EW events in the fiducial phase-space. To account for this we can set the NF as follows:

$$\mu_{\mathrm{XX}} = \frac{\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}} \cdot 4\mathcal{L} \cdot \sigma^{\mathrm{fid.,MC}}_{\mathrm{WZjj\text{-}EW}}}{N^{\mathrm{fid.,MC}}_{\mathrm{W_X Z_X jj\text{-}EW}}} \times f^{\mathrm{fid.}}_{\mathrm{XX}}\,, \tag{7.2}$$

which is applied to the sample(s) corresponding to the polarisation state being probed, thus either $\mathrm{W_0Z_0jj}$-EW, $\mathrm{W_0Z_Xjj}$-EW, or $\mathrm{W_XZ_0jj}$-EW. We use XX in indices to refer to the polarisation state being measured.

The non-targeted polarisation states, meaning the transverse polarisation states of non-interest, have a different NF definition:

$$\mu_{\overline{\mathrm{XX}}} = \frac{\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}} \cdot 4\mathcal{L} \cdot \sigma^{\mathrm{fid.,MC}}_{\mathrm{WZjj\text{-}EW}}}{N^{\mathrm{fid.,MC}}_{\mathrm{WZjj\text{-}EW}} - N^{\mathrm{fid.,MC}}_{\mathrm{W_XZ_Xjj\text{-}EW}}} \times (1 - f^{\mathrm{fid.}}_{\mathrm{XX}})\,. \tag{7.3}$$

In the fit the POIs are thus $\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}}$ and $f^{\mathrm{fid.}}_{\mathrm{XX}}$, with the other NFs being derived from these two parameters. Only these two POIs are free to float in the fit. This parametrisation allows us to measure directly the fiducial cross-section and polarisation fraction leaving the normalisation of each polarised sample to float accordingly. The POIs are constrained to be positive, and the polarisation fractions are by design constrained to be between 0 and 1 with a nominal value corresponding to the MC expectation. We consider in this setup that the relative fractions of polarisation not being measured are fixed. The cross-section is allowed to vary between 0 and 5 times the SM expectation.

We can correlate the POIs between the two data-taking periods, because the $\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}}$ is set to be the ratio between the observed and expected cross-section, and assumed independent of the data-taking period. The polarisation fractions are also expected to be independent of the data-taking period, and thus can be correlated between the two periods. Indeed, Table 6.E shows that the polarisation fractions are expected to be the same between the two data-taking periods, and thus can be correlated. The difference arises only after a few decimals, far beyond the precision of the measurement.

Regarding the irreducible backgrounds of interest in the CRs, we constrain their normalisation using NFs μ. These samples correspond to ZZjj-QCD, $\mathrm{t\bar{t}} + V$, and tZj. The NFs are defined as the ratio of data with respect to the MC expectation for the total cross-section of the given process. The NFs are free to float in the fit, and are constrained to be positive with a nominal value corresponding to unity, that is the MC expectation.

The summary of the POIs and NFs is given in Table 7.A. Only the POIs are correlated between the two data-taking periods, but the NFs are specific to each data-taking period since different MC generators are used for the two periods.

For the other background processes, their normalisation is fixed to the MC expectation with their associated systematic uncertainties.

Table 7.A: Strength parameters used in the fit.

Process	Strength parameter	Regions	Correlated between runs
$\mathrm{W_XZ_Xjj}$-EW	$\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}}$	SR	✔
$\mathrm{W_XZ_Xjj}$-EW	f_{XX}	SR	✔
ZZjj-QCD	$\mu_{\mathrm{ZZjj\text{-}QCD}}$	SR, ZZ-CR	✘
$\mathrm{t\bar{t}} + V$	$\mu_{\mathrm{t\bar{t}}+V}$	SR, ZZ-CR, b-CR	✘
tZj	μ_{tZj}	SR, ZZ-CR, b-CR	✘

7.2 Systematic uncertainties

Every measurement is an estimation of the true value of a physical quantity and is thus affected by uncertainties. Among them we can distinguish two types of uncertainties: statistical and systematic. The statistical uncertainty is related to the finite size of the data sample: repeating the same measurement with a different data sample would lead to a different result. The only way to overcome this uncertainty is to increase the size of the data sample. The systematic uncertainty is related to the imperfect knowledge of the measurement methods and steps: using a different method for the same measurement would lead to a different result. The only way to overcome this uncertainty is to improve the knowledge of the tools and methods used in the measurement.

A systematic uncertainty therefore quantifies the possible bias in a measurement due to imperfect knowledge of the detector response, the signal and background modelling, or the analysis procedure.

Each systematic uncertainty at various stages of the analysis is associated with an NP denoted α. Among them, we distinguish two types of uncertainties:

Overall normalisation uncertainties. This is a simple uncertainty when an effect is not known to affect the shape of the distribution, but only the overall normalisation. The value of the uncertainty is

typically given in percentage, and the effect is applied to all bins of the template. The NP associated with this uncertainty is a multiplicative factor applied to the template, and is constrained by a Gaussian prior with a mean of zero and a standard deviation of one. The effect of this uncertainty is thus to scale the template up or down by a certain percentage.

Shape uncertainties. This is a more complex uncertainty when an effect is known to affect the shape of the distribution. The effect is typically given by two variations of the template, one for the upward variation, commonly denoted as "Up", and one for the downward variation, commonly denoted as "Down". The NP associated with this uncertainty is a continuous parameter that interpolates between the nominal template and the Up and Down variations. The effect of this uncertainty is thus to change the shape of the template in a continuous way, depending on the value of the NP. The interpolation is typically done using a polynomial or spline function, and is constrained by a Gaussian prior with a mean of zero and a standard deviation of one.

Two sources of systematic uncertainties are considered in the fit:

Experimental uncertainties. These are uncertainties related to the detector performance, such as the calibration of the energy scale and resolution of the jets and leptons, the efficiency of the trigger and reconstruction algorithms, the modelling of the pile-up, or the luminosity measurement. These uncertainties are typically evaluated using dedicated calibration samples and data-driven methods, and are propagated to the final templates used in the fit. Most of the sources of these experimental uncertainties were introduced in the various sections of each object and are now linked to the fit.

Theoretical uncertainties. These are uncertainties related to the modelling of the physics processes in the MC simulation, such as the choice of the PDF set, the renormalisation and factorisation scales, the parton shower, or the higher-order corrections. These uncertainties are typically evaluated by varying the relevant parameters in the MC simulation and propagating the effect to the final templates used in the fit. The theoretical uncertainties are typically larger for the signal processes. This is even more true in this measurement: in Section 6.5 we detailed the rigorous treatment of the production of MC samples for the polarised states and their limitations.

7.2.1 Experimental systematic uncertainties

The reconstruction, calibration, and identification of the physics objects are only imperfectly reproduced by the simulation. The residual differences with respect to data are measured by the dedicated ATLAS Combined Performance groups and are propagated to the analysis as experimental systematic uncertainties, affecting the shape and normalisation of the MC-based predictions. The object uncertainties are not propagated to the VVV sample, to the WZjj-EW full sample of the CRs, and to the τ component, which carry a normalisation uncertainty only. Each source enters the fit as one or more NPs that deform the templates within their measured range and are then constrained by the fit together with the parameters of interest. They cover the broad list of objects considered in this analysis: the pile-up reweighting, the electron energy and muon momentum scale and resolution, the $E_{\mathrm{T}}^{\mathrm{miss}}$ scale and resolution, the JES and JER, and the SFs that reweight the simulation to reproduce the trigger, reconstruction, identification, isolation, and flavour-tagging efficiencies measured in data.

Table 7.B summarises the resulting NPs for the Run-2 and Run-3 datasets, whose object definitions differ slightly (Tables 6.C(a) to 6.C(c)).

7.2.1.1 Scale-factor versus four-momentum uncertainties

Three classes are distinguished in the "Class" column of Table 7.B (p_4, SF, and norm. for the luminosity). *Four-momentum* uncertainties (p_4) affect an object's energy or momentum calibration (the electron energy scale and resolution, the muon combined momentum and momentum scale, the sagitta bias, the JES and JER, and the $E_{\mathrm{T}}^{\mathrm{miss}}$ soft term). They are evaluated by re-deriving the object four-momenta with the shifted or smeared calibration and recomputing every downstream quantity, so they migrate events between bins and reshape the templates. *SFs* instead reweight events to vary the data-to-simulation efficiency corrections (reconstruction, identification, isolation, trigger, flavour tagging, JVT, and pile-up reweighting). They change the yield and acceptance while leaving the kinematics of each event within a selection unchanged, although they can still affect the normalisation. As the measurement is only weakly sensitive to the electron four-momentum, its scale and resolution use the simplified single-parameter decorrelation model `1NP_v1`.

7.2.1.2 Electrons

The electron reconstruction, identification, isolation, and trigger efficiencies are assessed with tag-and-probe methods in $\mathrm{Z} \to \mathrm{ee}$ events (Section 5.3.7) [278], while the energy scale and resolution follow the dedicated e/γ energy calibration (Section 5.2) [293]. Electron identification uses the LH of Section 5.3.1 in both runs, the DNN-based identification of Section 5.3.9 not being used by this analysis, each efficiency provided as a single NP per run. The trigger is described by one NP in Run-2 and by a second in Run-3 from the updated SF treatment. As both datasets use different electron calibrations and different identification scale factors, the NPs are uncorrelated between the two datasets.

7.2.1.3 Flavour tagging

The flavour-tagging efficiencies change most between the datasets, following the move from the `DL1r` tagger in Run-2 [379] to GN2 in Run-3 [380] (both at the $70\,\%$ WP). The efficiency SFs are calibrated in bins of jet p_{T} and flavour [422–424], giving a large set of correlated sources. To make them tractable, this covariance is diagonalised into a smaller orthogonal set of NPs, an eigenvector decomposition that preserves both the total uncertainty and its bin-to-bin correlations [425]. Run-2 retains only the reduced quadrature scheme, with a single NP per jet flavour plus two extrapolation terms (the high-p_{T} extrapolation and the extrapolation from charm to τ-jets), whereas Run-3 keeps the full decomposition: ten eigenvector NPs each for the b-, c-, and LF-jet efficiencies and four high-p_{T} extrapolation variations per flavour. All are SFs, because the `DL1r` and GN2 calibrations have different covariances, so their eigenvectors differ, and none of these NPs is correlated between the datasets.

7.2.1.4 Jets

Jet uncertainties are based on the in-situ JES and JER measurements in data [373, 382]. The absolute JES is set by in-situ balance measurements (Z+jets, γ+jets, and multijet) and, at high p_{T}, by the single-hadron response, from which a large number of uncertainty sources arise (the statistical precision in each p_{T} bin, the experimental systematics, and the modelling variations) giving a set of correlated NPs too large to propagate through the analysis. It is reduced by diagonalising the covariance matrix of these sources and keeping a smaller orthogonal set, the "effective" NPs [426]: "effective" meaning that each is no longer an individual physical source but a linear combination of them, the index ordering the eigenvectors within a group by decreasing size. In the `CategoryReduction` scheme used here [373], with $\sim$30 NPs, the sources are grouped into four physically motivated categories *before* the reduction, so that the correlation structure is preserved when the result is combined with other analyses:

Detector. Detector-response effects in the in-situ measurements, such as the energy scale of the reference e, γ, and Z objects.

Modelling. Physics- and MC-modelling effects of the in-situ balance measurements: the generator difference, obtained by computing the balance with two different generators, the radiation suppression of the second jet, where the in-situ selections veto a second jet above a fraction of the reference-object p_{T} to suppress additional radiation that spoils the balance, the uncertainty being obtained by varying this veto threshold, and out-of-cone radiation, where part of the energy of the initial parton falls outside the jet cone so that the jet does not exactly balance the reference object, the uncertainty covering the modelling of this energy in the simulation.

Mixed. Sources carrying both a detector and a modelling character, which cannot be assigned unambiguously to either of the above.

Statistical. The statistical precision of the in-situ measurements themselves.

Three further four-momentum terms are added on top of these effective NPs: the b-jet response, which accounts for the different calorimeter response to the harder fragmentation and semileptonic decays of b-jets, the flavour composition and response for the unknown quark- and gluon-jet fractions of the sample and the different response to quark- and gluon-initiated jets, and the pile-up terms, which correct the offset from the average number of interactions ($\langle\mu\rangle$) and the number of primary vertices (N_{PV}) together with the residual p_{T}- and ρ-topology dependence of the jet-area subtraction.

A calibration is said to *close* when, after it is applied, the calibrated quantity recovers the true value in simulation, or is consistent between independent samples (the residual bias thus vanishes). The *non-closure* terms are the residual systematic deviations that remain once the nominal calibration has been applied, since the correction does not perfectly restore a unit response. As these are known imperfections, their size is taken as an uncertainty. The η-intercalibration non-closure terms cover, after forward jets

have been calibrated against central jets with the dijet p_{T} balance, the residual η- and energy-dependent deviations: at high E, for either sign of η, for the 2018 running conditions, and, in the pre-recommendation (PreRec) MC23 calibration, in the transition region $|\eta| \approx 0.2$. The in-situ and FCal non-closure terms are the analogous residuals of the in-situ calibration and of the forward-calorimeter region that the not-yet-finalised MC23 calibration does not fully remove, and the AtlFast3-versus-full simulation (AF3-versus-FS) term is the residual response difference between the fast and full (GEANT4-based) detector simulations. The punch-through term covers high-energy hadronic showers that are not fully contained in the calorimeter and leak into the muon system.

The JER is described by the `SimpleJER` scheme of eight NPs, a data-versus-simulation term and reduced effective NPs from the same in-situ measurements, with only the simulation smeared, complemented in Run-3 by the corresponding MC23 noise and AF3-versus-FS terms. Most jet terms are common to the two datasets, the campaign-specific MC16-versus-MC23 terms above being the exception.

Jet-vertex tagging is the only SF jet component: the cut-based JVT in Run-2 and the NN-based JVT in Run-3, with the fJVT term common to both [377].

All systematics sharing the same name in Table 7.B for JES, JER, and jet-vertex tagging are correlated between the two datasets, while those with different names are uncorrelated. Different names arise for campaign-specific terms, for example the η-intercalibration non-closure terms, which are different between the two datasets because of the different running conditions and the different MC campaigns. For Run-3, as the full period of data taking is not yet available, the systematics are PreRec and thus not finalised, and are expected to change in the future.

7.2.1.5 Muons

The muon momentum-scale, combined-measurement (CB muons), and sagitta-bias corrections are four-momentum uncertainties, while the reconstruction, isolation, TTVA, and trigger efficiencies are SFs, each efficiency split into a statistical and a systematic component. The sagitta-bias correction removes a charge-dependent bias in the measured track curvature (sagitta) arising from residual misalignments of the tracking detectors: it uses two NPs in Run-2 and four in Run-3, adding the global and p_{T}-extrapolation terms. Similarly to the electrons, the scale and resolution uncertainties are assessed with tag-and-probe methods in $\mathrm{Z} \to \mu\mu$ events [370–372]. No correlation is assumed between the two datasets, as the muon reconstruction and calibration methods differ between Run-2 and Run-3.

7.2.1.6 Missing transverse momentum

The $E_{\mathrm{T}}^{\mathrm{miss}}$ uncertainty is obtained by propagating the scale and resolution uncertainties of the reconstructed objects and by applying momentum scale and resolution uncertainties to the track-based soft term [383]. The soft-term scale and the two resolution components, parallel and perpendicular to the hard-object recoil, are used for both datasets. No correlation is assumed between the two datasets for the $E_{\mathrm{T}}^{\mathrm{miss}}$ uncertainty, as the $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction methods differ between Run-2 and Run-3.

7.2.1.7 Pile-up

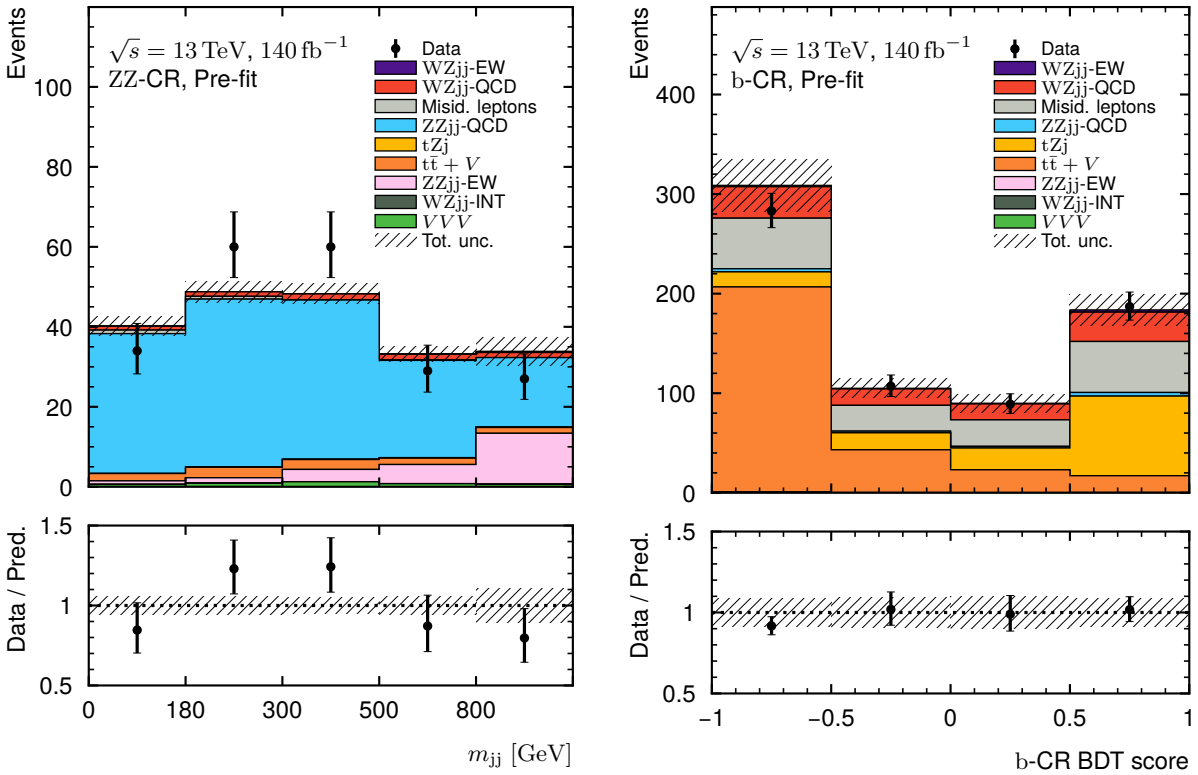
The simulation is reweighted so that its distribution of the average number of interactions per bunch crossing reproduces the data. The associated uncertainty is driven by the measurement of the inelastic pp cross-section that enters this reweighting [427], and is applied identically in both runs. Because the pile-up conditions differ substantially, no correlations are assumed between the two datasets.

7.2.1.8 Luminosity

The uncertainty on the integrated luminosity affects the normalisation of all MC-based samples. It is evaluated separately for the two datasets, following the vdM calibration with LUCID-2 (Section 3.8) of the Run-2 [128] and Run-3 [428] data, and enters as two uncorrelated normalisation NPs. The Run-2 uncertainty is $0.83\,\%$ and the Run-3 uncertainty is $1.9\,\%$. This uncertainty is applied to the signal normalisation as well as to all irreducible background contributions which are estimated purely using MC simulations.

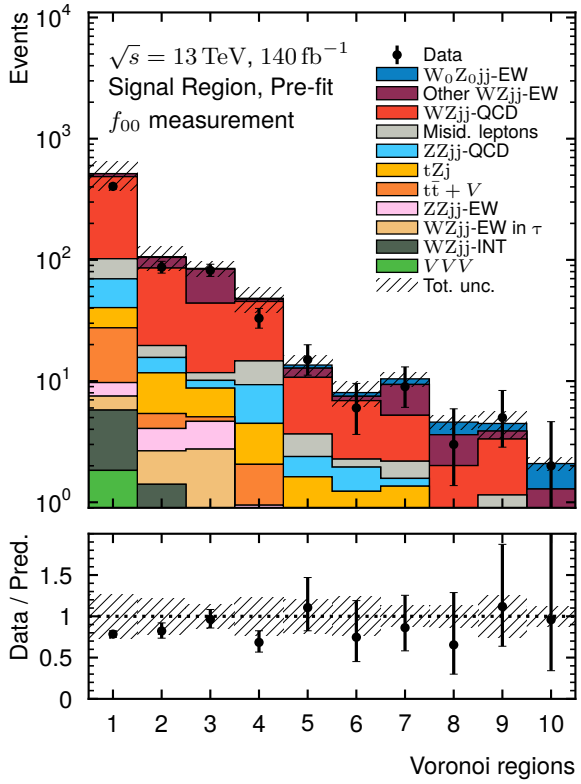
7.2.2 Theoretical uncertainties

The theoretical uncertainties account for the limited knowledge of the underlying physics processes. Various sources of uncertainty are considered, including the choice of PDF, the value of the strong

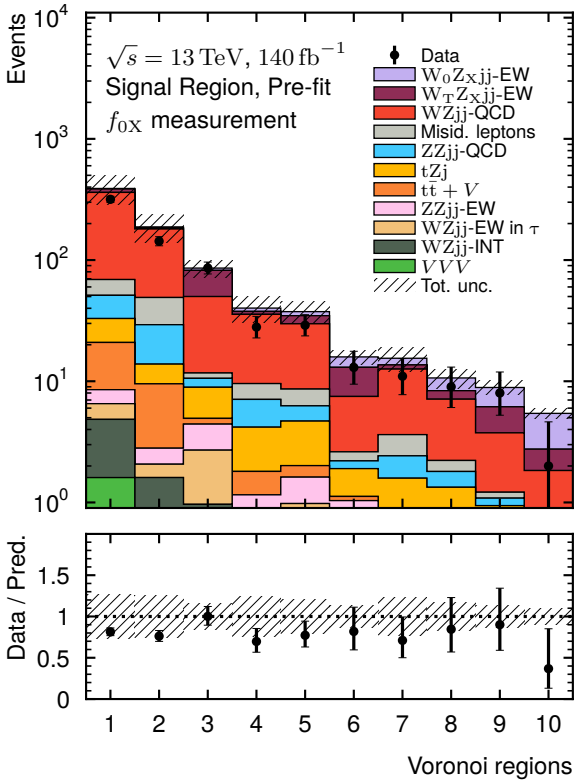


(a) ZZ-CR **(b)** b-CR

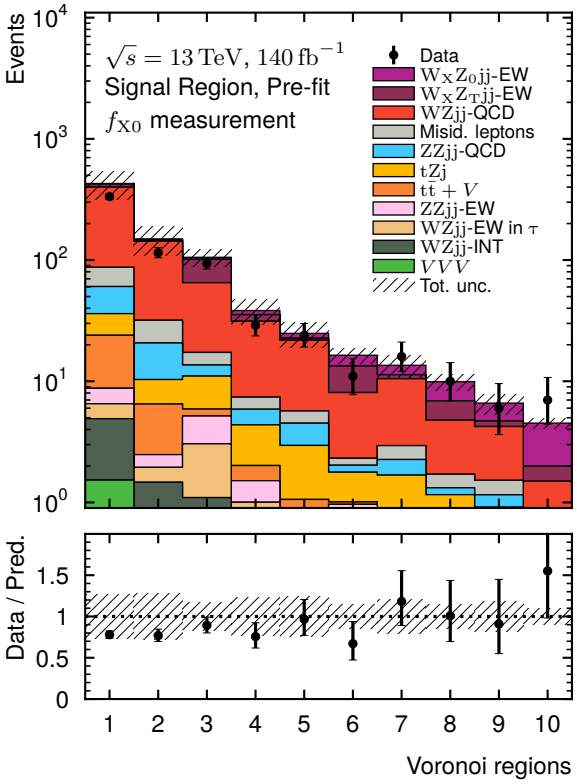
Figure 7.1: Pre-fit distributions for the Run-2 dataset. The CRs show m_{jj} in the ZZ-CR and the BDT score in the b-CR. The uncertainty band around the SM expectation includes all systematic uncertainties at their pre-fit values.



(a) f_{00} measurement

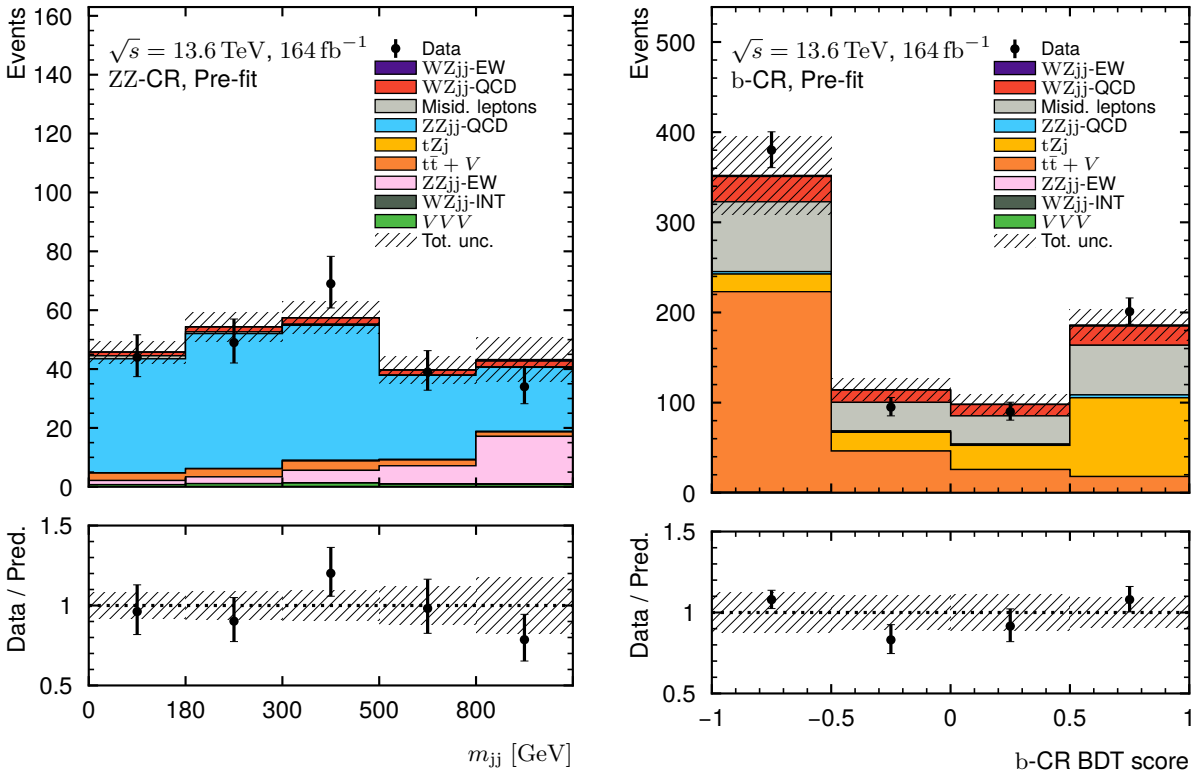


(b) f_{0X} measurement



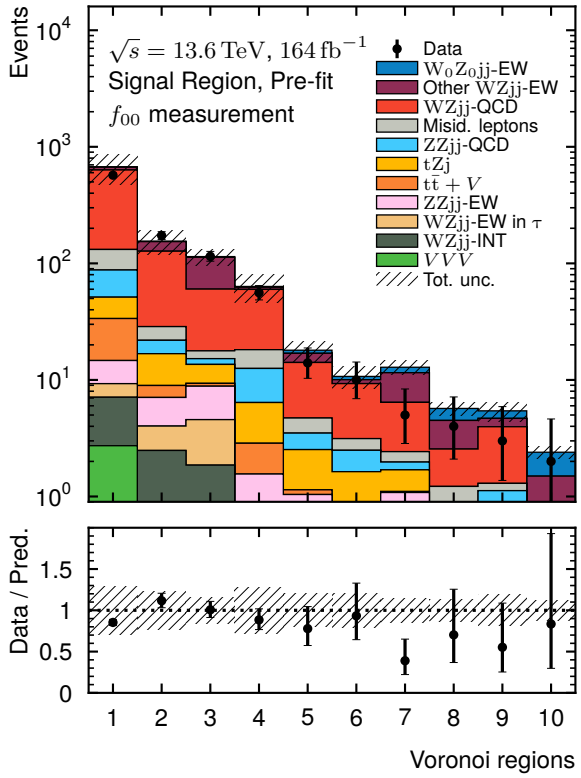
(c) f_{X0} measurement

Figure 7.2: Pre-fit distributions for the Run-2 dataset. In the SR, the Voronoi regions are shown for the f_{00} measurement (a), the f_{0X} measurement (b), and the f_{X0} measurement (c). The uncertainty band around the SM expectation includes all systematic uncertainties at their pre-fit values.

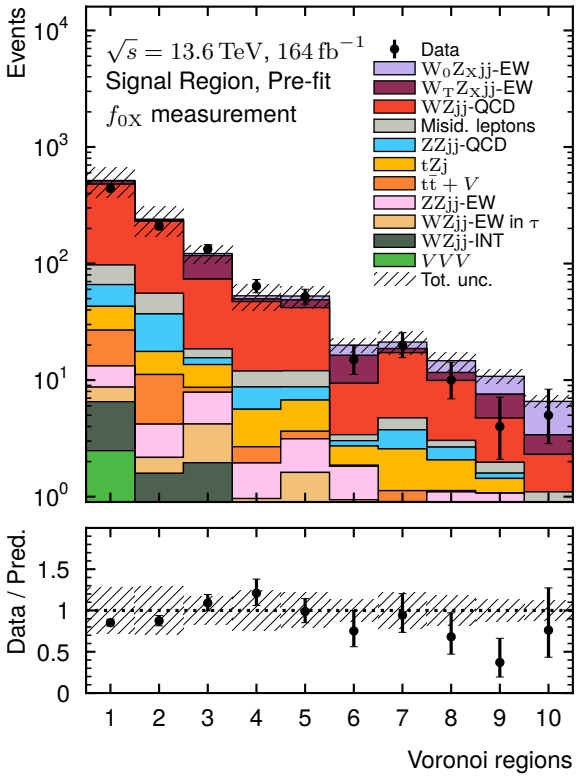


(a) ZZ-CR **(b)** b-CR

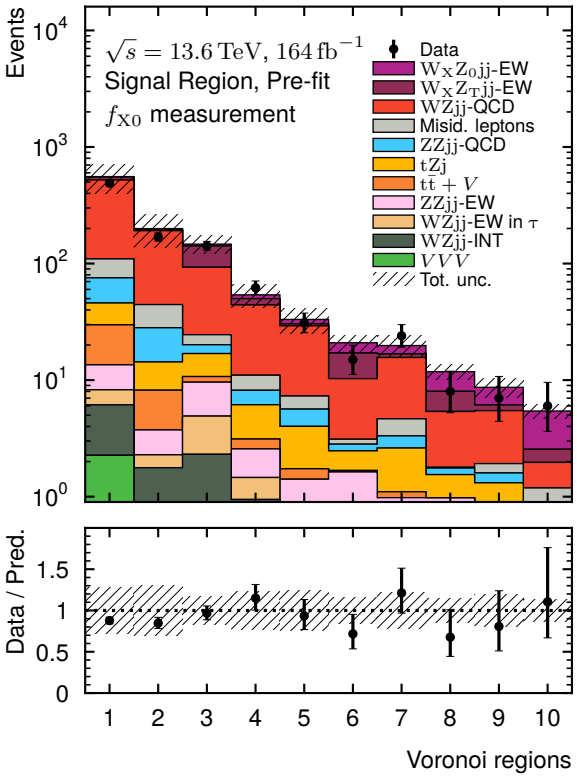
Figure 7.3: Pre-fit distributions for the Run-3 dataset. The CRs show m_{jj} in the ZZ-CR and the BDT score in the b-CR. The uncertainty band around the SM expectation includes all systematic uncertainties at their pre-fit values.



(a) f_{00} measurement



(b) f_{0X} measurement



(c) f_{X0} measurement

Figure 7.4: Pre-fit distributions for the Run-3 dataset. In the SR, the Voronoi regions are shown for the f_{00} measurement (a), the f_{0X} measurement (b), and the f_{X0} measurement (c). The uncertainty band around the SM expectation includes all systematic uncertainties at their pre-fit values.

Table 7.B: Experimental systematic uncertainties. The "Name" column gives the label used for each NP in plots (an index range for a collapsed stack), the "Class" column marks a four-momentum calibration (p_4), an efficiency SF, or a normalisation. Each parameter is marked (✔) when active in a given dataset, (✘) otherwise.

Name	Description	Class	Run-2	Run-3
Electrons (energy calibration Section 5.2, efficiencies Section 5.3.7)				
e energy scale	e energy scale	p_4	✔	✔
e energy resolution	e energy resolution	p_4	✔	✔
e ID eff.	e identification efficiency	SF	✔	✔
e iso. eff.	e isolation efficiency	SF	✔	✔
e reco. eff.	e reconstruction efficiency	SF	✔	✔
e trigger eff.	e trigger efficiency	SF	✔	✔
e trigger eff. (extra)	e trigger efficiency (additional Run-3 term)	SF	✘	✔
Flavour tagging (Paragraph 6.2.1.2.6)				
b-jet tag. eff.	b-jet tagging efficiency	SF	✔	✘
c-jet tag. eff.	c-jet tagging efficiency	SF	✔	✘
LF-jet tag. eff.	LF-jet tagging efficiency	SF	✔	✘
b-jet tag. high-p_{T} extrap.	b-jet high-p_{T} extrapolation	SF	✔	✘
b-jet tag. extrap. from c	b-jet extrapolation from c (τ jets)	SF	✔	✘
b-jet tag. eff. eigen 0–9	b-jet tagging efficiency, eigenvector var. 0–9	SF	✘	✔
c-jet tag. eff. eigen 0–9	c-jet tagging efficiency, eigenvector var. 0–9	SF	✘	✔
LF-jet tag. eff. eigen 0–9	LF-jet tagging efficiency, eigenvector var. 0–9	SF	✘	✔
b-jet tag. p_{T}-extrap. eigen 0–3	b-jet high-p_{T} extrapolation, eigenvector var. 0–3	SF	✘	✔
c-jet tag. p_{T}-extrap. eigen 0–3	c-jet high-p_{T} extrapolation, eigenvector var. 0–3	SF	✘	✔
LF-jet tag. p_{T}-extrap. eigen 0–3	LF-jet high-p_{T} extrapolation, eigenvector var. 0–3	SF	✘	✔
JES (Paragraph 6.2.1.2.4)				
JES b-jet response	b-jet response	p_4	✔	✔
JES detector NP 1–2	effective NP, detector 1–2	p_4	✔	✔
JES mixed NP 1–3	effective NP, mixed 1–3	p_4	✔	✔
JES modelling NP 1–4	effective NP, modelling 1–4	p_4	✔	✔
JES stat. NP 1–6	effective NP, statistical 1–6	p_4	✔	✔

continued on next page

Table 7.B – continued from previous page

Name	Description	Class	Run-2	Run-3
JES η-intercalibration modelling	η-intercalibration, modelling	p_4	✓	✓
JES η-intercalibration total stat.	η-intercalibration, total statistical	p_4	✓	✓
JES non-closure 2018	η-intercalibration non-closure, 2018 data (MC16)	p_4	✓	✗
JES non-closure high E	η-intercalibration non-closure, high E (MC16)	p_4	✓	✗
JES non-closure $\eta < 0$	η-intercalibration non-closure, $\eta < 0$ (MC16)	p_4	✓	✗
JES non-closure $\eta > 0$	η-intercalibration non-closure, $\eta > 0$ (MC16)	p_4	✓	✗
JES $\lvert\eta\rvert \approx 0.2$ non-closure (PreRec)	η-intercalibration non-closure, $\lvert\eta\rvert \approx 0.2$ (MC23 PreRec)	p_4	✗	✓
JES flavour composition	flavour composition	p_4	✓	✓
JES flavour response	flavour response	p_4	✓	✓
JES pile-up offset $\langle\mu\rangle$	pile-up offset, $\langle\mu\rangle$-dependent	p_4	✓	✓
JES pile-up offset N_{PV}	pile-up offset, N_{PV}-dependent	p_4	✓	✓
JES pile-up p_{T} term	pile-up p_{T} term	p_4	✓	✓
JES pile-up ρ topology	pile-up ρ-topology term	p_4	✓	✓
JES FCal non-closure (PreRec)	FCal non-closure (MC23 PreRec)	p_4	✗	✓
JES in-situ non-closure (PreRec)	in-situ non-closure (MC23 PreRec)	p_4	✗	✓
JES punch-through	punch-through (MC16)	p_4	✓	✗
JES noise (PreRec)	noise (MC23 PreRec)	p_4	✗	✓
JES vertexing algorithm (PreRec)	vertexing algorithm (MC23 PreRec)	p_4	✗	✓
JES AtlFast3 vs full sim. (PreRec)	AF3 versus FS (MC23 PreRec)	p_4	✗	✓
JER [373]				
JER data vs sim.	data versus simulation (MC16)	p_4	✓	✓
JER NP 1–7	effective NP 1–7	p_4	✓	✓
JER noise (PreRec)	noise (MC23 PreRec)	p_4	✗	✓
JER AtlFast3 vs full sim. (PreRec)	AF3 versus FS (MC23 PreRec)	p_4	✗	✓
Jet-vertex tagging (Paragraph 6.2.1.2.5) [377, 378]				
JVT eff.	JVT efficiency	SF	✓	✗
NN-JVT eff.	NN-JVT efficiency	SF	✗	✓
fJVT eff.	fJVT efficiency	SF	✓	✓

continued on next page

Table 7.B – continued from previous page

Name	Description	Class	Run-2	Run-3
Muons (Paragraphs 6.2.1.1.4 and 6.2.1.1.5)				
μ CB	μ combined momentum (CB)	p_4	✓	✓
μ momentum scale	μ momentum scale	p_4	✓	✓
μ reco. eff. stat.	μ reconstruction efficiency, statistical	SF	✓	✓
μ reco. eff. syst.	μ reconstruction efficiency, systematic	SF	✓	✓
μ iso. eff. stat.	μ isolation efficiency, statistical	SF	✓	✓
μ iso. eff. syst.	μ isolation efficiency, systematic	SF	✓	✓
μ TTVA eff. stat.	μ TTVA efficiency, statistical	SF	✓	✓
μ TTVA eff. syst.	μ TTVA efficiency, systematic	SF	✓	✓
μ trigger eff. stat.	μ trigger efficiency, statistical	SF	✓	✓
μ trigger eff. syst.	μ trigger efficiency, systematic	SF	✓	✓
μ sagitta data stat.	μ sagitta bias, data statistical	p_4	✓	✓
μ sagitta residual bias	μ sagitta bias, residual bias	p_4	✓	✓
μ sagitta global	μ sagitta bias, global	p_4	✗	✓
μ sagitta p_{T} extrap.	μ sagitta bias, p_{T} extrapolation	p_4	✗	✓
MET (Section 6.2.1.3)				
$E_{\mathrm{T}}^{\mathrm{miss}}$ soft-term scale	$E_{\mathrm{T}}^{\mathrm{miss}}$ soft-term scale	p_4	✓	✓
$E_{\mathrm{T}}^{\mathrm{miss}}$ soft-term reso. $\parallel$	$E_{\mathrm{T}}^{\mathrm{miss}}$ soft-term resolution, $\parallel$	p_4	✓	✓
$E_{\mathrm{T}}^{\mathrm{miss}}$ soft-term reso. $\perp$	$E_{\mathrm{T}}^{\mathrm{miss}}$ soft-term resolution, $\perp$	p_4	✓	✓
Pile-up (Section 7.2.1.7)				
pile-up reweighting	pile-up reweighting ($\langle\mu\rangle$)	SF	✓	✓
Luminosity (Section 3.8) [128, 428]				
Run-2 Luminosity	integrated luminosity, Run-2	norm.	✓	✗
Run-3 Luminosity	integrated luminosity, Run-3	norm.	✗	✓

coupling constant α_S, the choice of renormalisation and factorisation scales, and the modelling of parton shower. The theoretical uncertainties are evaluated by varying the relevant parameters in the MC simulation and propagating the resulting changes through the analysis. The impact of these variations on the predicted yields and kinematic distributions is then assessed, and the corresponding uncertainties are included in the final results.

As the target of these measurements is to measure the cross-section of WZjj-EW polarised states in the SR, only effects in the reconstruction efficiency and in the shape of distributions of MC predictions are included in the definition of the uncertainty bands. All considered variations of the MC predictions are therefore normalised to the same fiducial cross-section, as predicted by the Sherpa polarised MC samples for the WZjj-EW fiducial phase-space. The same treatment is done for ZZjj-QCD.

For the WZjj-QCD prediction, the variations are not normalised to the nominal cross-section: each NP changes both the shape and the normalisation. As no NF is used for this background, these NPs, dominated by the QCD scale variations, act as a constrained normalisation of WZjj-QCD in the fit (see Section 7.3.7).

7.2.2.1 Misidentified leptons

For the reducible backgrounds, the misidentified leptons, two estimates are used: the matrix method in the dominant regions (SR and the b-CR) and, where their contribution is limited (the ZZ-CR), the MC prediction corrected with the scale factors of Section E.2. Both methods are detailed in Appendix E. For the matrix method, the associated uncertainties are detailed in Section E.1.5 and consist of:

- the statistical uncertainty of the nominal histogram, taken into account as part of it
- the statistical uncertainty on the fake factors. It is labelled as "F stat."
- the subtraction of irreducible events from the CRs and SRs. It is labelled as "Irr. subtraction"
- the $|\eta|$ correction. It is labelled as "$|\eta|$ correction"
- the flavour composition of the fake factors. It is labelled as "F Flavour comp."
- the flavour composition in the SRs. It is labelled as "Flavour comp."

This results in a total uncertainty of the order of $30\,\%$. They are not correlated between the two datasets, as the fake factors are derived independently for each dataset. For the scale-factor-corrected MC prediction in the ZZ-CR, the yields are low (1.7 to 2.2 events), so an overall uncertainty on the cross-section of $\sigma_{\text{Misid. leptons}} = 25\,\%$ is applied, which is not correlated between the two datasets.

7.2.2.2 Cross-section uncertainties

The total event yields of backgrounds from ZZjj-QCD, $\mathrm{t\bar{t}} + V$, and tZj processes are estimated using data in dedicated CRs in a combined fit with the SR. Therefore, no normalisation uncertainty is attributed to these backgrounds.

The other backgrounds are attributed the following overall normalisation uncertainties, quoted as "σ_X" with X the process name:

- $\sigma_{\mathrm{WZjj\text{-}EW}} = 25\,\%$ in the CRs for the full samples and reflecting the low statistics in these regions. This is correlated between the two datasets, as the same generator is used for both datasets.
- $\sigma_{VVV} = 25\,\%$ [429], which is not correlated between the two datasets, as the Sherpa version differs between the datasets.
- $\sigma_{\mathrm{ZZjj\text{-}EW}} = 25\,\%$, to account for the potentially large impact of QCD scale variations. It is not correlated between the two datasets, as the generator is different for each dataset.
- $\sigma_{\mathrm{WZjj\text{-}EW\,in\,\tau}} = 10\,\%$, which is correlated between the two datasets, as the same generator is used for both datasets.
- $\sigma_{\mathrm{WZjj\text{-}INT}} = 20\,\%$, which is correlated between the two datasets, as the same generator is used for both datasets.

7.2.2.3 PDF uncertainties

The PDF uncertainties are taken into account for the polarised WZjj-EW samples in the SR, the WZjj-QCD samples in all regions, and the ZZjj-QCD in its dominant regions: the SR and ZZ-CR.

For the WZjj-EW samples they are computed using the built-in variations of the Sherpa MC samples. Uncertainty bands at the $68\,\%$ confidence level (CL) on MC template distributions and due to the PDF

uncertainties are computed using MC replicas or Hessian eigenvectors, depending on the baseline PDF set used in the MC generation, and according to the PDF4LHC recommendations [430]. For the WZjj-QCD and ZZjj-QCD samples, the PDF uncertainties are computed in the same way but using the built-in variations using the 100 MC replicas of the NNPDF3.0 PDF set, following the PDF4LHC recommendations [430].

The effect of variation of the α_{S} used for the determination of the PDF set is also considered, by varying α_{S} from its nominal value of 0.118 to 0.117 or 0.119. Finally, it was also checked that the variations of the distribution obtained with other PDF sets stay within the $68\,\%$ CL uncertainty band of the baseline PDF set.

The resulting Up and Down bands for these theory uncertainties for the WZjj samples are shown in Appendix P for the inputs to the template fit.

As the same PDF set is used for both ZZjj-QCD and WZjj-QCD MC simulations, the corresponding PDF uncertainty will be considered in the fit as fully correlated between the two MC samples and between the two datasets. For the WZjj-EW samples, the same assumption is made between the two datasets, but the PDF uncertainty is considered uncorrelated with the WZjj-QCD and ZZjj-QCD samples, as they are produced with different generators and settings.

7.2.2.4 QCD scale uncertainties

The QCD scale uncertainties are taken into account for the polarised WZjj-EW samples in the SR, the WZjj-QCD samples in all regions, and the ZZjj-QCD in its dominant regions: the SR and ZZ-CR.

These uncertainties are computed using the built-in variations of the MC samples. They are estimated by varying the renormalisation and factorisation scale, μ_{R} and μ_{F}, introduced in Section 4.6, independently by factors of $\frac{1}{2}$ and 2 around the nominal scale and discarding the combinations where the variations differ by a factor of four. The envelope of the largest deviations is taken as a QCD scale uncertainty band.

The resulting Up and Down bands for these theory uncertainties for the WZjj samples are shown in Appendix P for the inputs to the template fit.

As the production mechanisms of the three samples are of a different nature, their corresponding QCD scale uncertainties are considered in the fit as uncorrelated between the three MC samples. The polarised WZjj-EW samples are however considered as fully correlated between them, as they are produced with the same generator and settings. For the WZjj-EW and WZjj-QCD samples, the QCD scale uncertainties are considered correlated between the two datasets, whereas for the ZZjj-QCD samples they are considered uncorrelated.

We note that the QCD scale uncertainty is quite large for the WZjj-QCD samples, Sherpa being very generous in its estimation of the QCD scale uncertainties: in the SR the band reaches, cell by cell, $^{+46}_{-28}\,\%$ in Run-2 and $^{+45}_{-27}\,\%$ in Run-3 for the joint template, and up to $+58\,\%$ in the last cell of the f_{X0} template in Run-3 ($^{+43}_{-27}\,\%$ and $^{+42}_{-26}\,\%$ on the integral of the unrolled joint template), against at most $\pm 4\,\%$ to $5\,\%$ per cell on the polarised WZjj-EW templates, whose normalisation is fixed. This is visible on the pre-fit yields given in Table 6.F.

7.2.2.5 Modelling uncertainties

No built-in variations for the parton shower uncertainties are present in Sherpa or in the alternative MadGraph+Pythia MC samples for WZjj processes for both runs. No simulation with another parton shower like Herwig is available.

The difference in shape between the baseline sample Sherpa and the alternative MadGraph+Pythia predictions is thus used for the distributions used in the SR and CRs to estimate a global modelling uncertainty. For the Run-3 WZjj-QCD sample, the alternative prediction is obtained by reweighting the nominal Sherpa sample at truth level to the MadGraph+Pythia shape, instead of directly from the alternative sample. The difference between the two predictions is symmetrised and used to define an uncertainty band around the nominal Sherpa prediction.

The resulting Up and Down bands for these theory uncertainties for the WZjj samples are shown in Appendix P for the inputs to the template fit.

We note that for the WZjj-EW samples, the nominal Sherpa samples are produced at LO plus 0 and 1 jet versus plus 0 jets at LO for the MadGraph+Pythia samples. This makes the Sherpa samples more accurate, as they include real QCD corrections. This modelling uncertainty is therefore considered to be representative of the effect of parton shower and generator differences on the polarisation templates. The envelope of the deviation is taken as a systematic uncertainty accounting for differences in ME implementation, matching, and merging schemes, and parton shower modelling between the two generators. The good agreement between the two MC samples was already assessed in Appendix L in the output of each DNN.

For the WZjj-QCD samples the difference grows towards the signal-like (high-index) bins, reaching $33\,\%$ in Run-2 and $25\,\%$ in Run-3 in the last cells of the joint template ($24\,\%$ to $28\,\%$ for the two single-boson templates), while staying below $5\,\%$ to $10\,\%$ in the background-like (low-index) ones. The tendency for the polarised WZjj-EW samples is the opposite: the deviation is largest in the background-like cells, in the first cell of the joint template, $12\,\%$ to $15\,\%$ in Run-2 and $14\,\%$ to $18\,\%$ in Run-3 depending on the polarisation state, and in the second cell of the single-boson templates, $12\,\%$ to $21\,\%$ in Run-2 and $10\,\%$ to $24\,\%$ in Run-3, and stays at the few-percent level, up to $15\,\%$, elsewhere. This is expected as the signal populates the high-index bins and the background the low-index ones, so each generator difference shows up where the corresponding process is scarce and least well constrained.

These modelling uncertainties are correlated between the runs independently for the EW and QCD production of WZjj. For the WZjj-EW, it is also correlated between the polarisation states.

7.2.2.6 NLO EW correction uncertainties

The samples used for the WZjj-EW process are produced at LO in Sherpa. In Section 6.5.3, we show that NLO EW corrections are expected to be of the order of $-15\,\%$ to $-20\,\%$ and are therefore not negligible. The computation has been performed at NLO EW to fixed order in the coupling, using calculations from MoCaNLO [415]. These computations provide state-of-the-art predictions for VBS processes, including the full set of virtual and real EW contributions relevant for the definition of polarised final states. In particular, the corrections affect both the overall normalisation and the shapes of kinematic distributions used to extract polarisation fractions, highlighting the importance of incorporating these effects when comparing theoretical templates with experimental measurements.

We worked in collaboration with the authors of the MoCaNLO framework to implement the DNN-based observable and the Voronoi binning to their code in `Fortran`. This allowed us to obtain theoretical predictions for the exact template that we used to disentangle the different polarisation states in WZjj-EW production. The applicability of this procedure relies on the compatibility of MoCaNLO and MadGraph at parton level. This is verified at LO using parton-level distributions extracted from the MadGraph Les Houches events, for the DNN scores and for the Voronoi templates. The good agreement observed ensures that the LO-to-NLO EW difference computed with MoCaNLO can be ported onto the MadGraph-based parton-to-particle migration, meaning that no difference in the underlying-event generation or parton-level modelling between the two generators biases the correction.

To evaluate the impact on the shapes of the predicted polarised distributions due to the missing higher-order EW corrections in the reference samples, distributions obtained with MoCaNLO at LO and NLO EW are compared. These comparisons cannot be directly applied to our reference samples, as the MoCaNLO predictions are only available at parton level. The measurements are performed at detector level, so this comparison misses effects from parton to detector level.

To obtain a more realistic correction, a dedicated procedure has been developed to propagate the MoCaNLO predictions from parton level to particle level using available MC samples. The procedure uses the MadGraph samples to derive parton-to-particle-level migration matrices, which are then applied to the NLO EW corrections obtained from MoCaNLO.

This workflow lets us obtain an estimate of the NLO EW corrections at particle level. At this stage, the shape effect is only up to $\sim 5\,\%$ for the Pol. XX DNNs with a flat distribution while up to $\sim 15\,\%$ to $20\,\%$ for the S. vs B. DNN with a decrease from background- to signal-like region. The effect is more pronounced in background-like regions and smaller in signal-like regions for the Pol. XX DNNs. We then evaluate the differences of the NLO EW versus LO at particle level, and we use them to construct K-factors as a function of the DNN-based observable. These K-factors are then applied at truth particle level, as a function of the unrolled Voronoi template, as event-by-event reweighting factors to the Sherpa MC events. The reweighted Sherpa events are subsequently passed through the full analysis chain to obtain a reweighted *detector-level* distribution of the unrolled Voronoi template that incorporates the NLO EW corrections. The parton-to-particle migration has the effect of smearing and reducing the corrections at detector level.

This reweighted distribution is then used as a one-sided systematic uncertainty on the Sherpa MC template in the final fit, allowing the template to move from the LO prediction to the one incorporating the NLO EW corrections. The variation is one-sided by construction: the LO-to-NLO EW deformation of the template is fixed by the calculation and an upward variation would have no physical meaning. It is essentially a shape effect, the reweighted templates differing from the nominal ones by at most $4.4\,\%$ in normalisation and by $-12\,\%$ to $+10\,\%$ cell by cell, the overall correction of Section 6.5.3 being absorbed by the free $\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}}$. Only the down template is provided, so that the response is the nominal template for positive values of the nuisance parameter and the full correction at -1σ.

This procedure provides an estimate of the systematic bias associated with missing higher-order EW corrections in the template construction and quantifies the sensitivity of the extracted polarisation fractions

to NLO EW effects. It is observed that the parton-to-particle migration smears and reduces the NLO EW effect seen at parton level, so that the residual impact on the templates at detector level is smaller than the correction obtained from the pure parton-level MoCaNLO prediction.

The resulting Down band for these theory uncertainties for the WZjj samples is shown in Appendix P for the inputs to the template fit.

These NLO EW uncertainties are correlated between the runs and between the polarised templates of a given fit. The three polarisation measurements being separate fits with their own pair of templates, no correlation between the fits is implied.

7.2.3 Pruning of systematic uncertainties

Pruning is a procedure that removes systematic uncertainties from the fit that have a negligible impact on the final results. This is done to reduce the complexity of the fit and improve its stability. The pruning procedure is based on the evaluation of the impact of each systematic uncertainty on the final results, and those with an impact below a certain threshold are removed from the fit.

We apply pruning only to object-related uncertainties, using a threshold of $0.2\,\%$. For a given sample and region, the Up and Down variation templates are first normalised to the same yield as the nominal template, only for the purpose of the pruning decision. The bin-by-bin shape and normalisation difference between each variation and the nominal template is then evaluated. If the difference exceeds the pruning threshold in at least one bin, the corresponding NP is kept in the fit. Otherwise, the uncertainty is pruned, meaning the corresponding NP is either removed from the fit for the sample and region or kept using only the shape or normalisation effect. This procedure reduces the number of NPs entering the fit while keeping uncertainties that produce a non-negligible effect.

The results of the pruning procedure are summarised in Appendix Q. The plots presented in this appendix show which systematic uncertainties are kept or pruned for each sample and region. These plots also allow the reader to see clearly which systematics enter the fit, to which samples they are applied, and whether they are correlated between various samples or runs. A common name for an NP indicates that it is shared between various samples. The mention of “Run-2” or “Run-3” indicates that the NP is applied only to the corresponding dataset and is thus not correlated between the two datasets. This same appendix also presents the overall effects of this pruning procedure, assessing the negligible impact on the measurement and making the fit less complex by removing negligible systematics.

7.3 Results

In this section we present the results of the three polarisation measurements. The results are obtained from a combined fit of the SR and CRs for the Run-2 and Run-3 datasets. The results are presented in terms of the measured polarisation fractions, as well as the corresponding uncertainties. The results are also compared to the theoretical predictions, and the compatibility between the measurements and the predictions is assessed.

Unless stated otherwise, uncertainties quoted as $\pm\sigma$ are the parabolic Hesse errors from the covariance matrix, while those quoted as $^{+a}_{-b}$ are the Minos errors obtained by profiling the likelihood. The two agree in the Gaussian regime and differ where a parameter is non-parabolic or bounded, most visibly for the fractions pinned at the physical boundary $f = 0$: there the Hesse value is the symmetric parabolic approximation and the Minos interval the physical one.

Throughout its development the analysis was blinded to the polarisation-sensitive information in the SR: the DNN output scores, the two-dimensional discriminant plane, and the unrolled Voronoi templates were not examined in data, and no fit involving a POI was run on the SR data. The analysis strategy (the discriminant training, the Voronoi binning, the pruning, and the systematic model) was fixed on simulation and Asimov pseudo-data, and the expected sensitivities quoted below were obtained in this configuration. Single-variable distributions in the SR, including the DNN inputs, were compared with data, as they carry no polarisation information individually and are those of the published WZjj-EW cross-section measurement whose selection the SR reproduces. The control and validation regions, in which the background modelling is validated, were not blinded. The SR was unblinded only once the fit and its validation had been finalised.

7.3.1 Goodness-of-fit

The goodness-of-fit quantifies how well the post-fit model describes the data, independently of the parameter values it returns. It is evaluated here with a saturated-model test: the profiled likelihood of

the nominal model is compared to that of a reference ("saturated") model with one free parameter per bin, which by construction reproduces the data exactly. The test statistic $-2\ln(\frac{L_{\text{model}}}{L_{\text{saturated}}})$ follows a χ^2 distribution, from which a p-value is computed. A small p-value ($\lesssim 0.05$) signals tension between the model and the data, whereas a large one indicates that the observed distribution is a plausible fluctuation of the fitted model. Values close to unity, in turn, can flag an over-conservative uncertainty model. The p-values for the three polarisation fits, for the combination and the individual Run-2 and Run-3 fits, are collected in Table 7.C.

Table 7.C: Saturated-model goodness-of-fit p-values for the three polarisation fits, for the combination of Run-2 and Run-3 and for the individual per-run fits.

Measurement	Combined	Individual fits	
		Run-2	Run-3
f_{00}	0.20	0.62	0.08
$f_{0\text{X}}$	0.26	0.73	0.07
$f_{\text{X}0}$	0.32	0.56	0.22

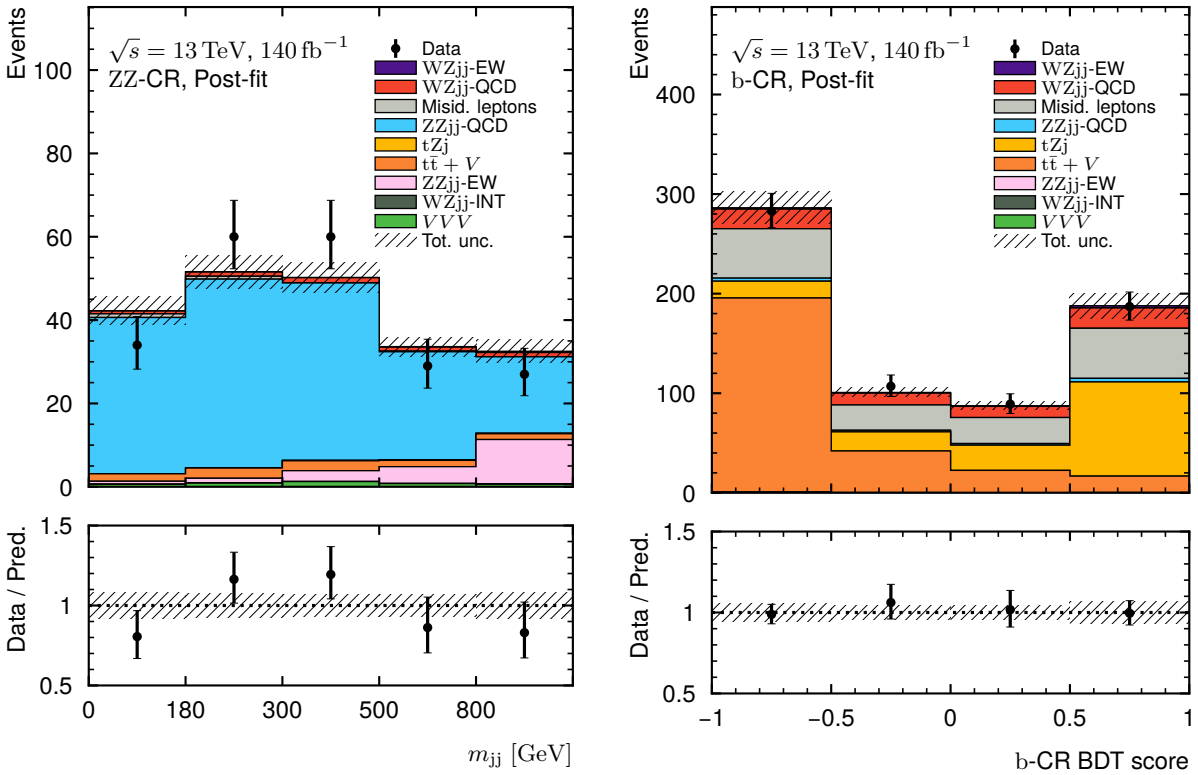
All nine p-values lie comfortably within the range expected for a well-behaved fit, confirming that the model provides a good description of the data in every configuration and consolidating the fit-quality picture drawn from the pulls, constraints, and correlations. The combined fits give p-values of 0.20, 0.26, and 0.32 for the f_{00}, $f_{0\text{X}}$, and $f_{\text{X}0}$ measurements: none is low enough to indicate mis-modelling, and none is close enough to unity to suggest an over-conservative uncertainty. The values are stable across the three polarisation configurations, as expected since the three fits share the same data, regions, and background model and differ only in the signal template being probed.

The individual-run fits are also all acceptable, with a mild pattern worth noting: the Run-3 p-values (0.08, 0.07, 0.22) sit systematically lower than the Run-2 ones (0.62, 0.73, 0.56). This reflects a somewhat poorer, though still statistically acceptable, description of the Run-3 data. In the combination the two periods average out and every p-value returns to the 0.20–0.32 range, indicating that neither run pulls the joint fit into tension.

7.3.2 Post-fit distributions

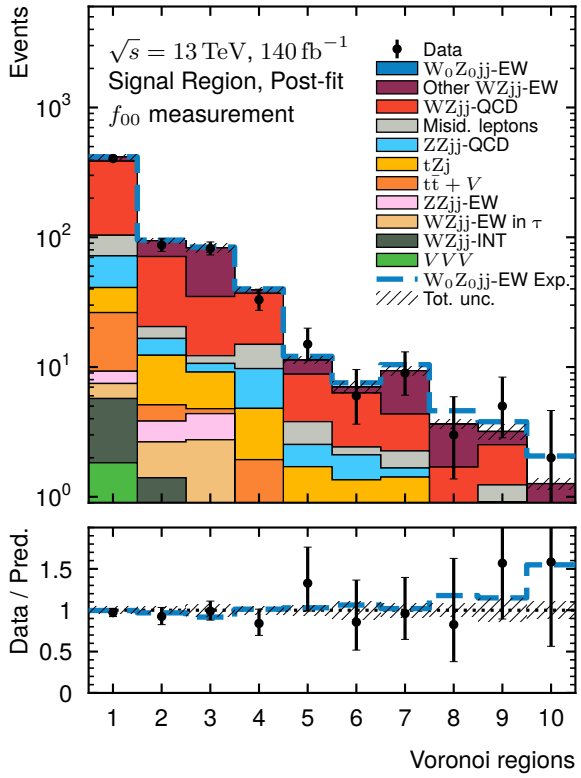
The post-fit distributions for the Run-2 and Run-3 datasets are shown in Figures 7.5 to 7.8. The post-fit distributions show good agreement between the data and the SM predictions, both in the CRs and in the SR, with some deviations observed in the high-index regions of the Voronoi template, which are dominated by the signal. These deviations are consistent with statistical fluctuations and do not indicate any significant discrepancy between the data and the SM predictions. The uncertainty bands around the SM predictions include all systematic uncertainties as obtained from the fit, and they provide a comprehensive representation of the total uncertainty in the measurements. Each Voronoi region is well populated, indicating no bias or mis-modelling of this method.

The post-fit uncertainty bands are markedly narrower than the pre-fit ones (Figures 7.1 to 7.4 versus Figures 7.5 to 7.8), the reduction being driven predominantly by the constraint of the $\text{WZ}jj$-QCD normalisation.

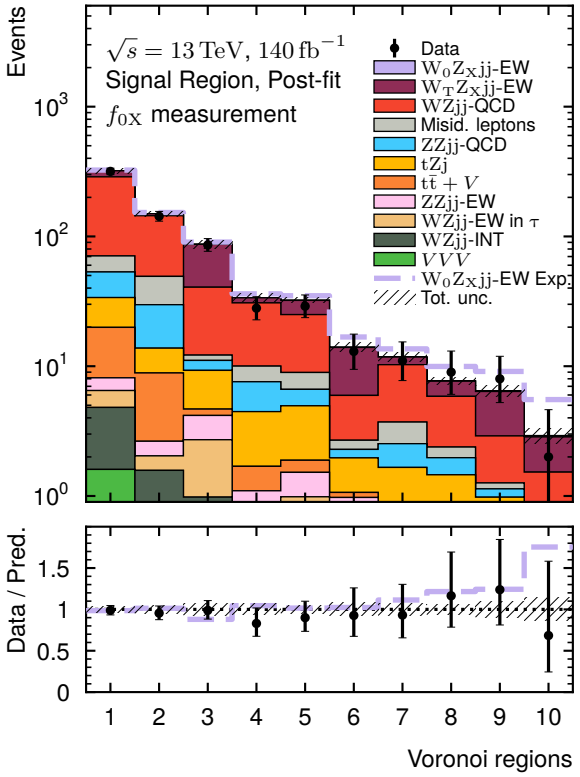


(a) ZZ-CR **(b)** b-CR

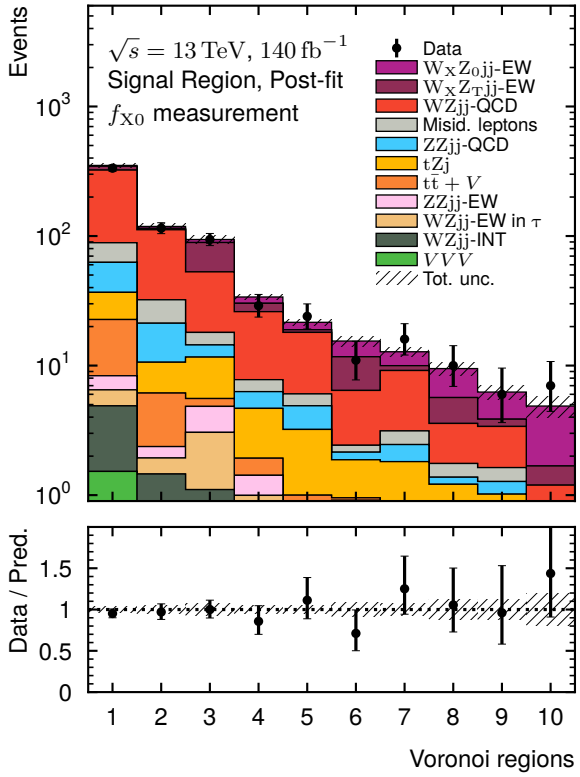
Figure 7.5: Post-fit distributions for the Run-2 dataset. The CRs show m_{jj} in the ZZ-CR and the BDT score in the b-CR. The uncertainty band around the SM expectation includes all systematic uncertainties as obtained from the fit.



(a) f_{00} measurement

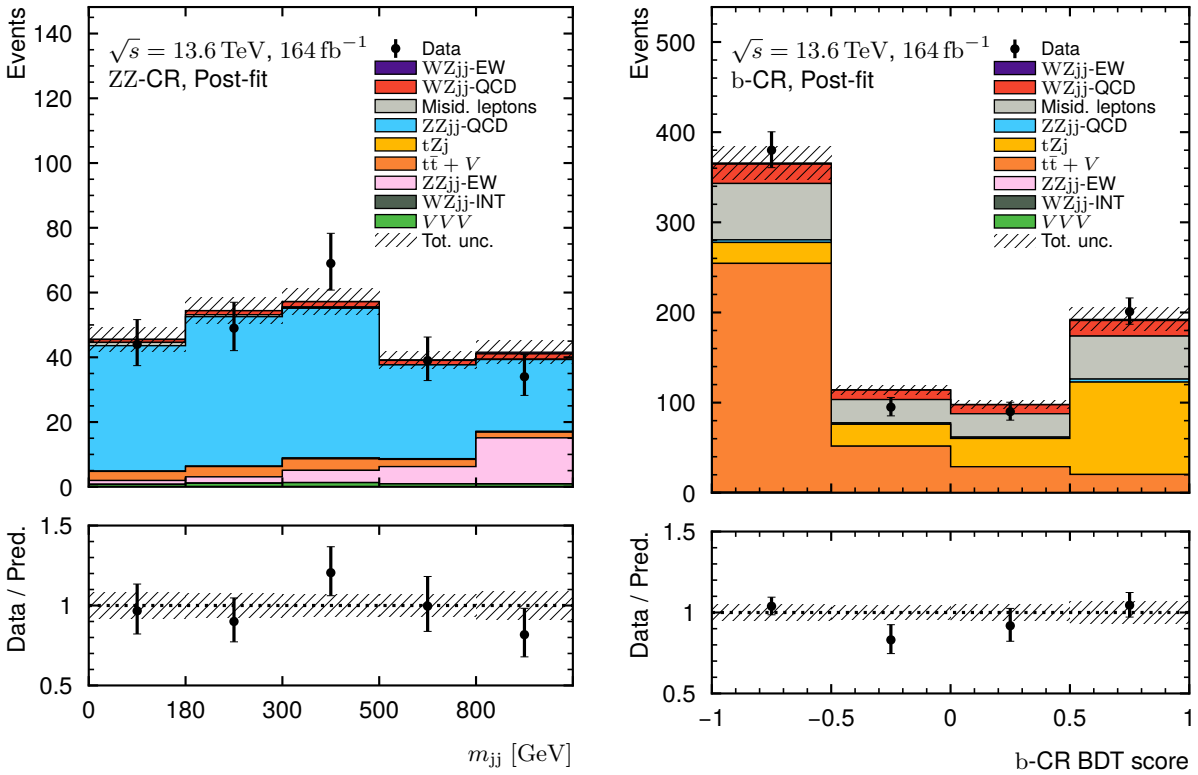


(b) f_{0X} measurement



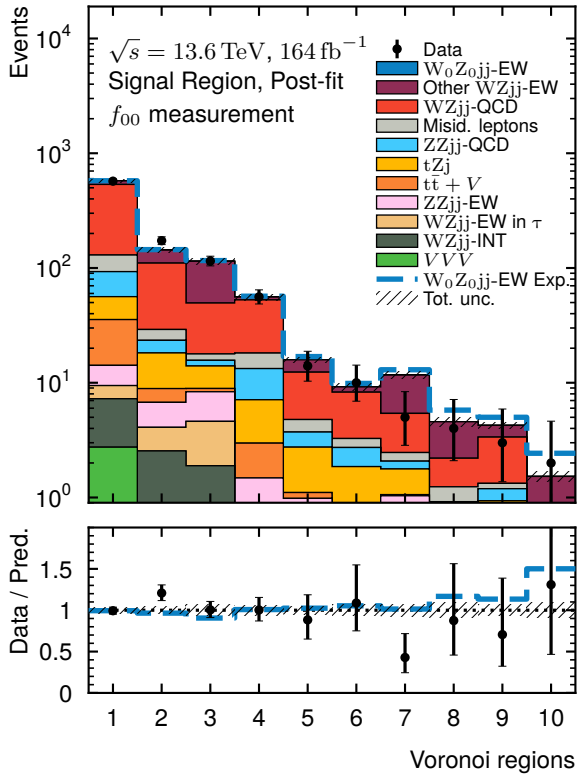
(c) f_{X0} measurement

Figure 7.6: Post-fit distributions for the Run-2 dataset. In the SR, the Voronoi regions are shown for the f_{00} measurement (a), the f_{0X} measurement (b), and the f_{X0} measurement (c). The uncertainty band around the SM expectation includes all systematic uncertainties as obtained from the fit. The dashed line in (a) and (b) is the prediction with the targeted signal at its SM (pre-fit) yield.

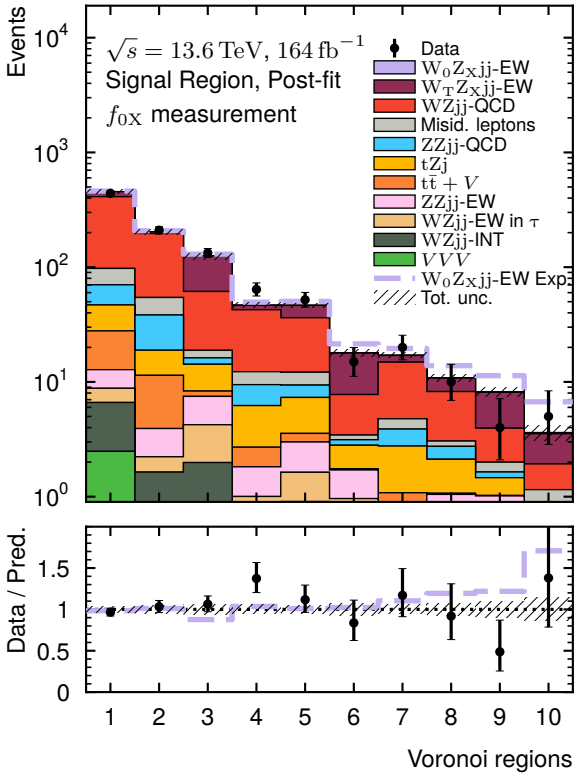


(a) ZZ-CR **(b)** b-CR

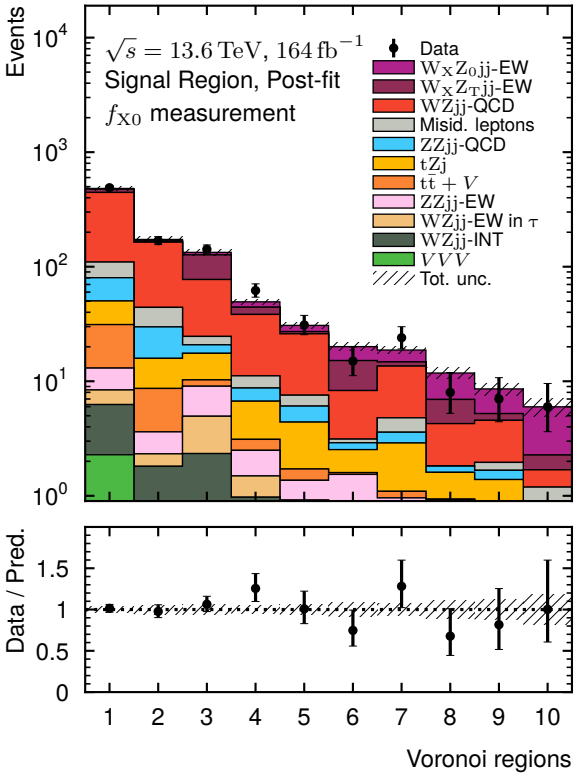
Figure 7.7: Post-fit distributions for the Run-3 dataset. The CRs show m_{jj} in the ZZ-CR and the BDT score in the b-CR. The uncertainty band around the SM expectation includes all systematic uncertainties as obtained from the fit.



(a) f_{00} measurement



(b) $f_{0\text{X}}$ measurement



(c) $f_{\text{X}0}$ measurement

Figure 7.8: Post-fit distributions for the Run-3 dataset. In the SR, the Voronoi regions are shown for the f_{00} measurement (a), the $f_{0\text{X}}$ measurement (b), and the $f_{\text{X}0}$ measurement (c). The uncertainty band around the SM expectation includes all systematic uncertainties as obtained from the fit. The dashed line in (a) and (b) is the prediction with the targeted signal at its SM (pre-fit) yield.

7.3.3 Post-fit yields

The resulting post-fit yields are summarised in Table 7.E. The post-fit yields agree well with the SM predictions in the SR and CRs, despite an initial small overprediction of over 100 events that the fit corrects. The three measurements are also consistent with each other, especially in the CRs.

The post-fit yields also show that the WZjj-QCD process is well constrained by the fit, with a reduction of the initial large uncertainty (the NP QCD scale, which is correlated between runs). Table 7.D presents the pre-fit and post-fit yields for the WZjj-QCD process in the SR as well as the ratio between the two. The results are presented with both a standalone fit, done by considering only one data-taking period, and the nominal combined fit. The post-fit yields, including the contribution of the WZjj-QCD process, are in good agreement with the data, indicating that the fit has successfully constrained this background contribution. The post-fit yields for the WZjj-QCD process are also consistent between the standalone and combined fits, indicating that the fit is robust and stable across different data-taking periods. The results are consistent between measurements and between individual runs and their combination through a common NP.

In the previous measurement of WZjj-EW [360], the observed ratio between pre- and post-fit was reported as 0.78 ± 0.08, which is consistent with the results obtained in this analysis. The same-sign WWjj-EW measurement [353] also reported a similar ratio of 0.75 ± 0.08, indicating a consistent trend across different analyses. This ratio tends to be slightly higher for Run-3 but consistent with the observation in Run-2.

The yields also determine the composition of each region. The SR is $\sim 17\,\%$ pure in WZjj-EW signal on a dominant WZjj-QCD background, while the ZZ- and b-CRs are enriched to $\gtrsim 75\,\%$ in ZZjj-QCD and in $\mathrm{t\bar{t}} + V$, tZj, and misidentified leptons, respectively, confirming that each CR constrains its intended background. In the $\mathrm{W_0Z_0jj}$-EW fit the jointly longitudinal yield is driven to zero and the EW signal is dominated by its transverse–transverse component.

Quantitatively, the fit reduces the WZjj-QCD uncertainty in the SR from $\sim 36\,\%$ pre-fit to $\sim 7\,\%$ post-fit, a factor of five, while pulling the central yield down by $\sim 25\,\%$ to $30\,\%$ in Run-2 and $\sim 13\,\%$ to $20\,\%$ in Run-3, consistently between the standalone and combined fits and across the three measurements.

7.3.4 Normalisation factors

The resulting NFs for the three measurements are summarised in Table 7.F. The NFs obtained are consistent with the SM expectation in the CRs, and agree between the three measurements and between the individual runs and their combination through their dedicated NF. The results are also consistent with the previous measurements on WZjj-EW [360] where the following NFs were reported:

- $\mu_{\mathrm{ZZjj\text{-}QCD}} = 1.08 \pm 0.11$
- $\mu_{\mathrm{tZj}} = 1.23 \pm 0.22$
- $\mu_{\mathrm{t\bar{t}}+V} = 0.90 \pm 0.13$

All NFs are compatible with unity within 1–2σ and stable across the three polarisation fits, their differences lying well below the uncertainties, which validates the common treatment of the data-constrained backgrounds. The tZj normalisation is consistently $\sim 15\,\%$ high, in line with the 1.23 ± 0.22 of [360].

7.3.5 Correlation matrix

The correlation matrix between all the fit parameters is represented in Figures 7.9 to 7.11 for the three measurements, for the combination of Run-2 and Run-3 and with a threshold of 0.15 for clarity.

Table 7.D: WZjj-QCD number of events in the SR for Run-2, Run-3 pre-fit and post-fit. Ratios between post-fit and pre-fit are also shown. The uncertainty on the ratio is the post-fit yield uncertainty relative to the pre-fit prediction, the post-fit constraint on the WZjj-QCD normalisation.

Run	Measurement	Pre-fit	Standalone post-fit	Standalone ratio	Combination post-fit	Combination ratio
Run-2	$\mathrm{W_0Z_0jj}$-EW	531.7 ± 190.3	375.9 ± 32.7	0.71 ± 0.06	390.8 ± 28.6	0.74 ± 0.05
	$\mathrm{W_0Z_Xjj}$-EW	531.7 ± 190.3	380.4 ± 33.1	0.72 ± 0.06	393.2 ± 28.7	0.74 ± 0.05
	$\mathrm{W_XZ_0jj}$-EW	531.7 ± 190.3	375.8 ± 33.7	0.71 ± 0.06	393.1 ± 29.9	0.74 ± 0.06
Run-3	$\mathrm{W_0Z_0jj}$-EW	707.9 ± 258.2	603.8 ± 44.7	0.85 ± 0.06	571.4 ± 34.9	0.81 ± 0.05
	$\mathrm{W_0Z_Xjj}$-EW	707.9 ± 258.2	602.1 ± 45.2	0.85 ± 0.06	575.3 ± 35.2	0.81 ± 0.05
	$\mathrm{W_XZ_0jj}$-EW	707.9 ± 258.2	614.2 ± 46.9	0.87 ± 0.07	572.7 ± 36.6	0.81 ± 0.05

Table 7.E: Post-fit event yields in the SR and in the two CRs for the Run-2 and Run-3 datasets, with all sources of uncertainty. The three measurements are shown.

(a) f_{00} measurement

Process	SR		ZZ-CR		b-CR	
	Run-2	Run-3	Run-2	Run-3	Run-2	Run-3
Data	646	953	210	235	666	766
Total pred.	668 ± 23	935 ± 28	210 ± 14	238 ± 15	662 ± 26	770 ± 28
WZjj-EW	—	—	0.6 ± 0.1	0.8 ± 0.2	5.0 ± 1.3	3.4 ± 0.9
W_0Z_0jj-EW	$0.0^{<0.1}_{-0.0}$	$0.0^{<0.1}_{-0.0}$	—	—	—	—
W_0Z_Tjj-EW	23.0 ± 3.7	30.9 ± 5.0	—	—	—	—
$W_TZ_0$jj-EW	23.8 ± 3.8	31.9 ± 5.1	—	—	—	—
$W_TZ_T$jj-EW	68 ± 11	93 ± 15	—	—	—	—
WZjj-QCD	391 ± 29	571 ± 35	4.9 ± 0.4	6.8 ± 0.6	63.1 ± 9.8	59.1 ± 4.9
Misid. leptons	45 ± 10	52 ± 14	1.7 ± 0.4	2.2 ± 0.5	151 ± 36	162 ± 46
ZZjj-QCD	43.9 ± 4.0	52.4 ± 5.5	170 ± 15	182 ± 16	9.9 ± 0.6	9.3 ± 0.6
tZj	32.9 ± 6.4	44.1 ± 8.7	0.5 ± 0.1	1.0 ± 0.2	156 ± 29	181 ± 34
$t\bar{t}+V$	20.0 ± 2.5	25.9 ± 3.7	9.5 ± 1.2	13.5 ± 1.8	276 ± 32	354 ± 41
ZZjj-EW	5.6 ± 1.3	13.5 ± 3.2	19.2 ± 4.6	26.5 ± 6.3	0.2 ± 0.1	0.2 ± 0.1
WZjj-INT	6.5 ± 1.3	9.4 ± 1.9	0.1 ± < 0.1	0.1 ± < 0.1	0.6 ± 0.1	0.4 ± 0.1
WZjj-EW in τ	5.9 ± 0.6	7.6 ± 0.8	—	—	—	—
VVV	2.4 ± 0.6	3.7 ± 0.9	4.1 ± 1.0	4.9 ± 1.2	0.4 ± 0.1	0.4 ± 0.1

(b) f_{0X} measurement

Process	SR		ZZ-CR		b-CR	
	Run-2	Run-3	Run-2	Run-3	Run-2	Run-3
Data	646	953	210	235	666	766
Total pred.	666 ± 23	935 ± 28	210 ± 14	239 ± 15	662 ± 26	771 ± 28
WZjj-EW	—	—	0.6 ± 0.1	0.8 ± 0.2	5.0 ± 1.3	3.4 ± 0.9
W_0Z_0jj-EW	0.2 ± 1.2	0.3 ± 1.6	—	—	—	—
W_0Z_Tjj-EW	0.5 ± 2.7	0.7 ± 3.6	—	—	—	—
$W_TZ_0$jj-EW	28.6 ± 4.7	38.3 ± 6.2	—	—	—	—
$W_TZ_T$jj-EW	82 ± 14	111 ± 18	—	—	—	—
WZjj-QCD	393 ± 29	575 ± 35	4.9 ± 0.4	6.7 ± 0.6	63.8 ± 9.7	60.1 ± 5.0
Misid. leptons	45.1 ± 9.7	54 ± 15	1.7 ± 0.4	2.2 ± 0.5	152 ± 36	168 ± 49
ZZjj-QCD	44.0 ± 4.0	52.4 ± 5.4	170 ± 15	184 ± 16	9.9 ± 0.5	9.3 ± 0.5
tZj	32.9 ± 6.4	44.2 ± 8.9	0.5 ± 0.1	0.9 ± 0.2	155 ± 30	183 ± 35
$t\bar{t}+V$	19.8 ± 2.4	25.2 ± 3.5	9.5 ± 1.2	13.2 ± 1.7	275 ± 32	346 ± 42
ZZjj-EW	5.6 ± 1.3	13.4 ± 3.2	19.1 ± 4.6	26.6 ± 6.3	0.2 ± 0.1	0.3 ± 0.1
WZjj-INT	6.5 ± 1.3	9.5 ± 1.9	0.1 ± < 0.1	0.1 ± < 0.1	0.6 ± 0.1	0.4 ± 0.1
WZjj-EW in τ	5.8 ± 0.6	7.6 ± 0.8	—	—	—	—
VVV	2.4 ± 0.6	3.7 ± 0.9	4.1 ± 1.0	4.8 ± 1.2	0.4 ± 0.1	0.4 ± 0.1

(c) f_{X0} measurement

Process	SR		ZZ-CR		b-CR	
	Run-2	Run-3	Run-2	Run-3	Run-2	Run-3
Data	646	953	210	235	666	766
Total pred.	668 ± 23	934 ± 28	210 ± 14	238 ± 15	663 ± 26	771 ± 28
WZjj-EW	—	—	0.6 ± 0.1	0.8 ± 0.2	5.0 ± 1.3	3.4 ± 0.9
W_0Z_0jj-EW	10.9 ± 3.2	14.3 ± 4.3	—	—	—	—
W_0Z_Tjj-EW	19.1 ± 4.5	25.7 ± 6.0	—	—	—	—
$W_TZ_0$jj-EW	25.1 ± 7.5	34 ± 10	—	—	—	—
$W_TZ_T$jj-EW	57 ± 14	77 ± 18	—	—	—	—
WZjj-QCD	393 ± 30	573 ± 37	5.0 ± 0.4	6.8 ± 0.6	62.2 ± 10.0	59.1 ± 5.0
Misid. leptons	45 ± 10	54 ± 15	1.7 ± 0.4	2.2 ± 0.5	152 ± 36	168 ± 48
ZZjj-QCD	44.0 ± 4.0	52.4 ± 5.4	170 ± 15	183 ± 16	9.9 ± 0.6	9.3 ± 0.5
tZj	33.3 ± 6.4	44.1 ± 8.9	0.5 ± 0.1	1.0 ± 0.2	157 ± 30	182 ± 35
$t\bar{t}+V$	19.8 ± 2.4	25.7 ± 3.6	9.5 ± 1.2	13.3 ± 1.8	275 ± 32	349 ± 42
ZZjj-EW	5.6 ± 1.3	13.5 ± 3.2	19.1 ± 4.6	26.7 ± 6.3	0.2 ± 0.1	0.2 ± 0.1
WZjj-INT	6.5 ± 1.3	9.4 ± 1.9	0.1 ± < 0.1	0.1 ± < 0.1	0.6 ± 0.1	0.4 ± 0.1
WZjj-EW in τ	5.9 ± 0.6	7.6 ± 0.8	—	—	—	—
VVV	2.4 ± 0.6	3.7 ± 0.9	4.1 ± 1.0	4.9 ± 1.2	0.4 ± 0.1	0.4 ± 0.1

Table 7.F: Post-fit NFs for the three measurements, for the individual and combined Run-2 and Run-3 fits.

Normalisation factor	f_{00}		f_{0X}		f_{X0}	
	Individual	**Combined**	**Individual**	**Combined**	**Individual**	**Combined**
Run-2						
$\mu_{\mathrm{ZZjj\text{-}QCD}}$	$1.11^{+0.11}_{-0.11}$	$1.09^{+0.11}_{-0.10}$	$1.10^{+0.11}_{-0.11}$	$1.09^{+0.11}_{-0.10}$	$1.10^{+0.11}_{-0.11}$	$1.09^{+0.11}_{-0.10}$
$\mu_{\mathrm{t\bar{t}}+V}$	$0.95^{+0.13}_{-0.12}$	$0.96^{+0.12}_{-0.12}$	$0.95^{+0.13}_{-0.12}$	$0.96^{+0.12}_{-0.12}$	$0.95^{+0.13}_{-0.12}$	$0.96^{+0.12}_{-0.12}$
μ_{tZj}	$1.15^{+0.23}_{-0.22}$	$1.17^{+0.22}_{-0.22}$	$1.16^{+0.23}_{-0.22}$	$1.17^{+0.22}_{-0.22}$	$1.18^{+0.23}_{-0.22}$	$1.18^{+0.22}_{-0.22}$
Run-3						
$\mu_{\mathrm{ZZjj\text{-}QCD}}$	$1.03^{+0.12}_{-0.11}$	$0.97^{+0.11}_{-0.10}$	$1.03^{+0.12}_{-0.11}$	$0.98^{+0.11}_{-0.10}$	$1.03^{+0.12}_{-0.11}$	$0.97^{+0.11}_{-0.10}$
$\mu_{\mathrm{t\bar{t}}+V}$	$1.18^{+0.15}_{-0.15}$	$1.10^{+0.14}_{-0.14}$	$1.16^{+0.15}_{-0.15}$	$1.08^{+0.14}_{-0.14}$	$1.17^{+0.15}_{-0.15}$	$1.09^{+0.14}_{-0.14}$
μ_{tZj}	$1.15^{+0.22}_{-0.22}$	$1.15^{+0.22}_{-0.21}$	$1.16^{+0.23}_{-0.22}$	$1.16^{+0.22}_{-0.22}$	$1.14^{+0.22}_{-0.22}$	$1.15^{+0.22}_{-0.22}$

The matrix is strongly diagonal-dominant: the vast majority of parameter pairs have $|\rho| < 0.15$ and fall below the display threshold, confirming that the NPs and POIs are largely independent and that the fit is well conditioned. The few sizeable off-diagonal entries are physically expected. The signal strength $\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}}$ is moderately anti-correlated with the WZjj-QCD QCD-scale uncertainty and correlated with the WZjj-QCD modelling term. This reflects the competition between the EW signal and the dominant QCD-induced background in the signal-like Voronoi bins. In the absence of a dedicated NF, this competition is absorbed by these large Sherpa-driven NPs. The per-run misidentified-lepton fake-factor NPs are anti-correlated with the tZj and $\mathrm{t\bar{t}} + V$ normalisations in the b-CR, as both describe the same event population. Correlations between the two data-taking periods appear only through the parameters that are correlated by construction (the POIs and the fully correlated theory NPs), as intended by the correlation scheme of Table 7.A. No pathological correlation is observed in any of the three configurations.

For f_{X0} the fraction and the signal strength are weakly anti-correlated ($\rho = -0.15$), the two parameters trading off against each other, which tilts the 2D contour of Figure 7.12(c).

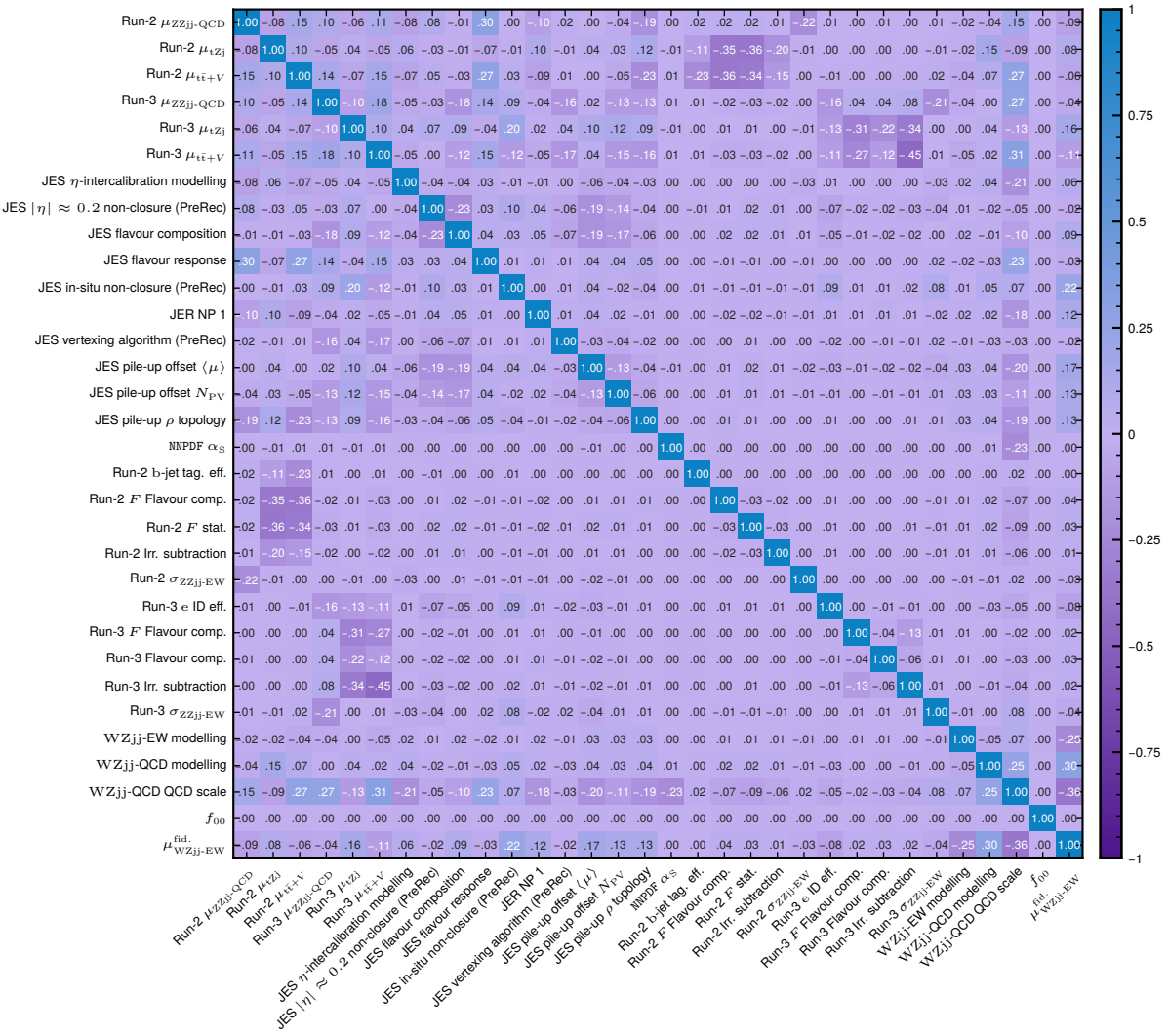


Figure 7.9: Correlation matrix for the f_{00} measurement, for the combination of Run-2 and Run-3. Only parameters with at least one correlation above a threshold of 0.15 are shown, the POIs being always kept for clarity. Each coefficient is printed to two decimals with the leading zero omitted.

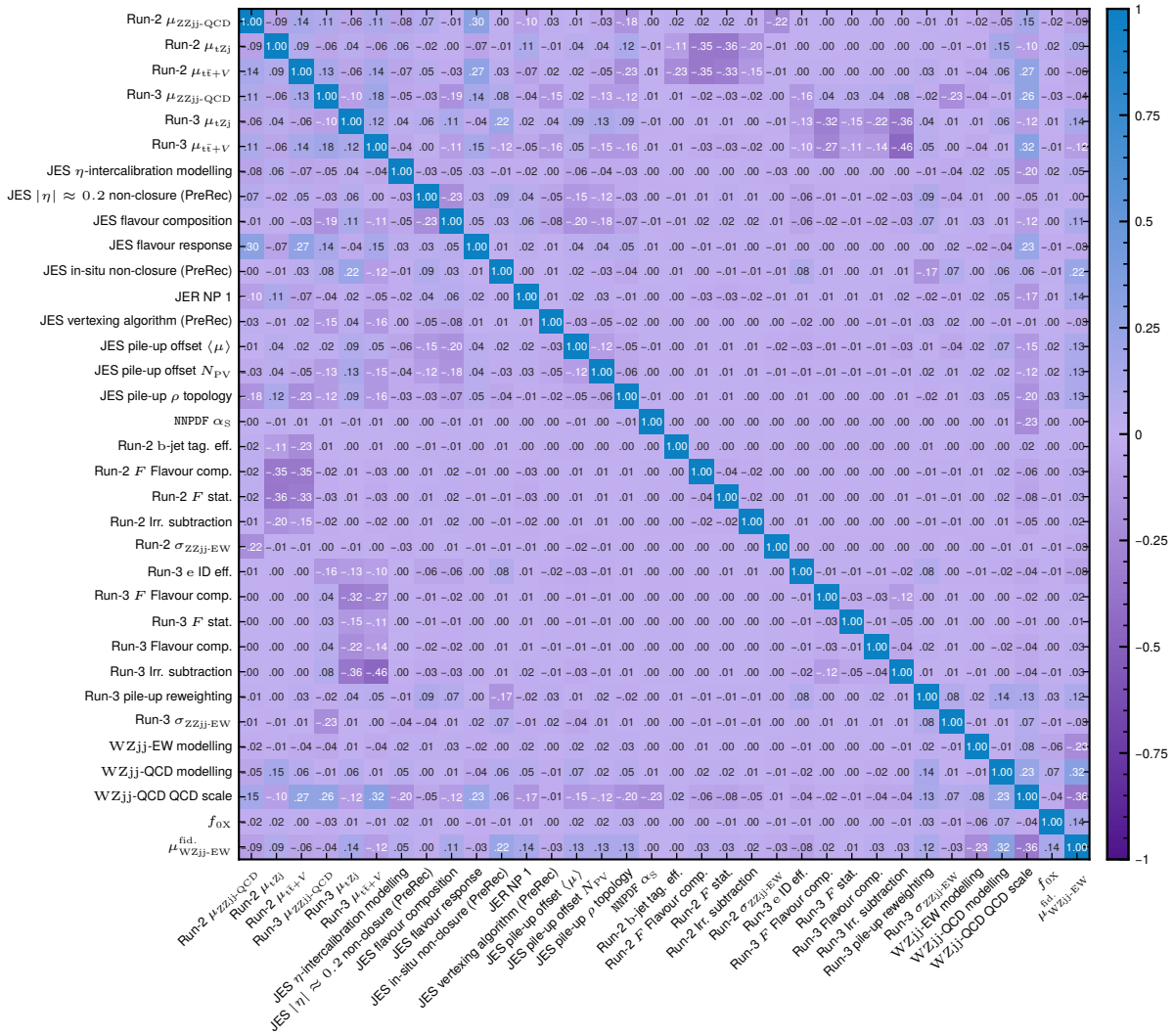


Figure 7.10: Correlation matrix for the $f_{0\mathrm{X}}$ measurement, for the combination of Run-2 and Run-3. Only parameters with at least one correlation above a threshold of 0.15 are shown, the POIs being always kept for clarity. Each coefficient is printed to two decimals with the leading zero omitted.

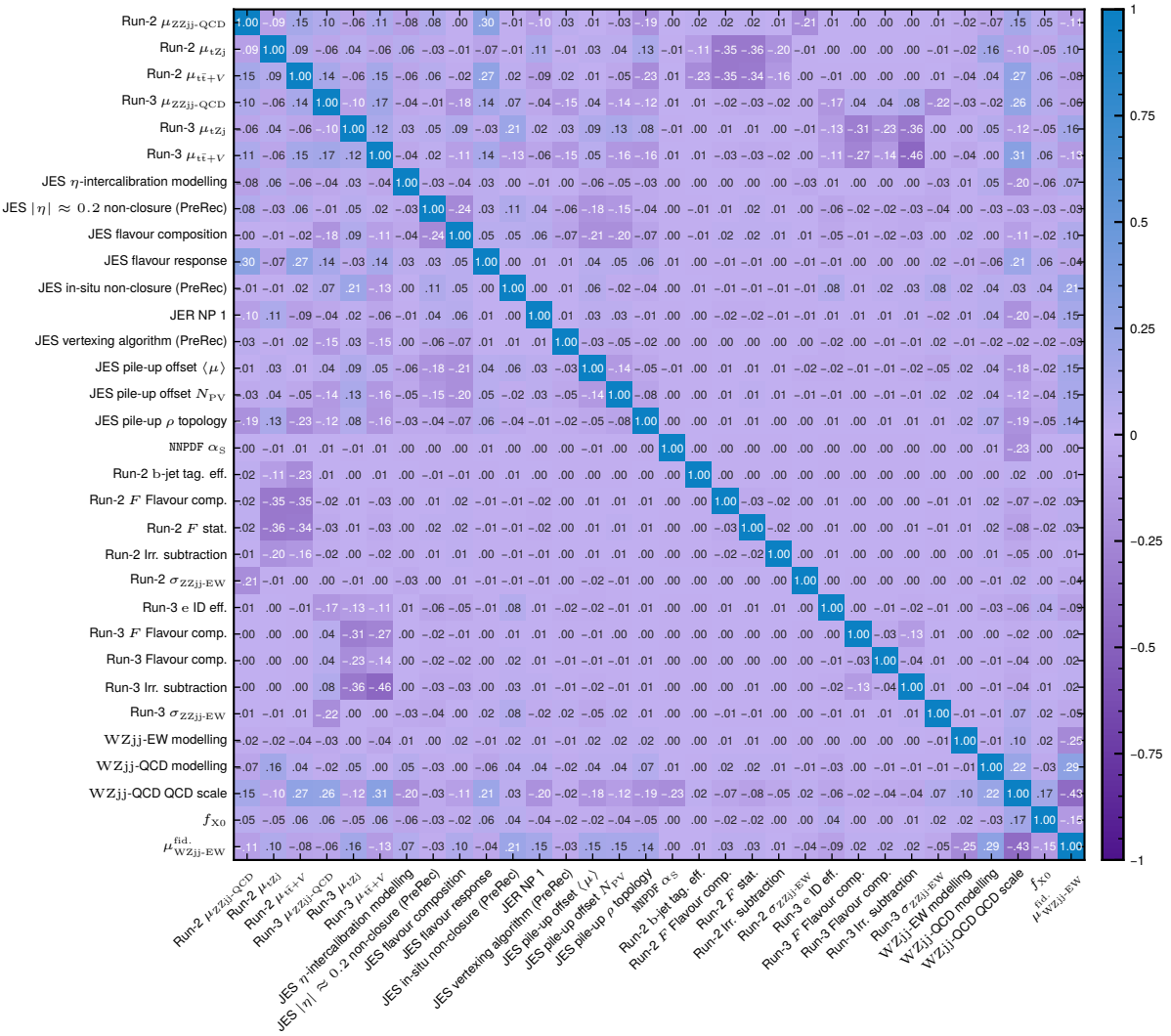


Figure 7.11: Correlation matrix for the f_{X0} measurement, for the combination of Run-2 and Run-3. Only parameters with at least one correlation above a threshold of 0.15 are shown, the POIs being always kept for clarity. Each coefficient is printed to two decimals with the leading zero omitted.

7.3.6 2D likelihood scans

The 2D likelihood scans for the three measurements are shown in Figure 7.12 for the combination of Run-2 and Run-3.

The scans are 2D in $(\mu^{\text{fid.}}_{\text{WZjj-EW}}, f_{\text{XX}})$ with both parameters left free, so the confidence regions are those of a two-parameter estimation. The 1σ, 2σ, and 3σ contours are drawn at increases of the negative-log-likelihood above its minimum, $-\Delta \ln L \equiv \text{NLL} - \text{NLL}_{\text{min}}$, of 1.15, 3.09, and 5.915, the two-degrees-of-freedom thresholds enclosing $68.3\,\%$, $95.4\,\%$, and $99.7\,\%$ probability ($\Delta\chi^2 = 2\,(-\Delta \ln L) = 2.30$, 6.18, and 11.83). These are not the one-parameter values: the 1σ intervals quoted for the individual parameters (Appendix S) come from the profiled scans at $-\Delta \ln L = 0.5$, so the on-axis extent of the 2D 1σ contour is larger than those intervals and the two are not expected to coincide. The likelihood is evaluated on a grid in the two parameters, referred to the grid minimum, and the contours are interpolated on the resulting surface. The MadGraph, Sherpa, and MoCaNLO predictions are overlaid with their $\mu^{\text{fid.}}_{\text{WZjj-EW}}$ uncertainty shown as a horizontal bar. The corresponding f_{XX} uncertainties are below the marker size.

In all three cases the $68\,\%$ and $95\,\%$ confidence contours are smooth, with a single well-defined minimum and no secondary minima, indicating a numerically stable fit. For f_{X0} the MadGraph, Sherpa, and MoCaNLO predictions overlaid on the panel lie inside the $68\,\%$ contour, for f_{00} they sit at its edge, and for $f_{0\text{X}}$ between the $95\,\%$ and $99.7\,\%$ contours, the deficit discussed in Section 7.3.12. Along the $\mu^{\text{fid.}}_{\text{WZjj-EW}}$ direction all three scans peak at unity within their uncertainty, consistent with the SM cross-section.

The orientation of the f_{X0} ellipse reflects the weak $\mu^{\text{fid.}}_{\text{WZjj-EW}}$–$f_{\text{X0}}$ anti-correlation ($\rho = -0.15$) of Section 7.3.5, and the single minima of the three scans confirm the absence of degeneracies in the two-parameter plane.

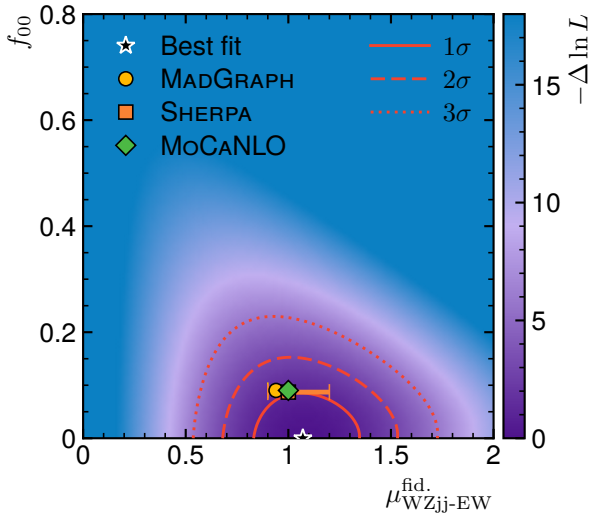
The 1D profile-likelihood scans underlying these results (Appendix S) are near-parabolic in $\mu^{\text{fid.}}_{\text{WZjj-EW}}$ and, for f_{X0}, in the fraction as well, confirming the Gaussian regime. For f_{00} and $f_{0\text{X}}$ they are truncated at the physical boundary $f = 0$, which is why only limits are quoted.

7.3.7 Nuisance parameters ranking

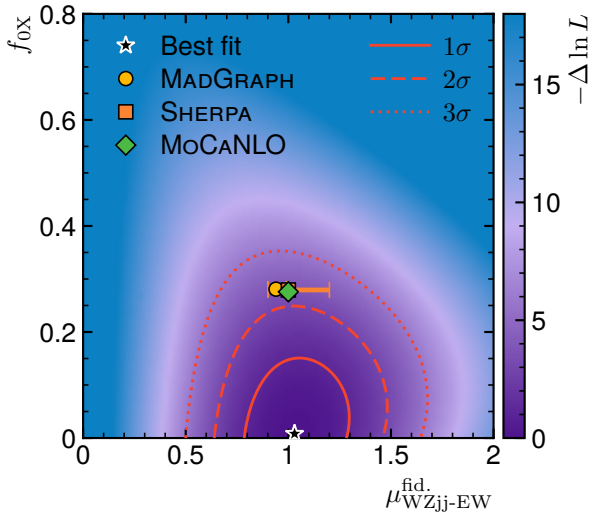
The pulls of the NPs, together with the correlations of Figures 7.9 to 7.11, characterise the stability of the fit. A mis-modelled distribution or an inaccurate template can be absorbed by a POI or an NP, biasing the result. This would show up as a large pull, an over-constraint, or a strong correlation. Over-constraints ($\hat{\sigma} < 1$) are acceptable for the loosely defined theory and modelling uncertainties, for which the fit can legitimately be more informative than the prescription they derive from, but are not expected for the object-level uncertainties provided by the ATLAS Combined Performance groups.

The pulls and constraints of the 20 highest-ranked NPs are shown in Figure 7.13 for the POIs f_{XX} and $\mu^{\text{fid.}}_{\text{WZjj-EW}}$, ordered by their post-fit impact on the POI following the covariance method of [431]. No significant pull is observed: all parameters are compatible with zero within 1σ, and their constraints are compatible with unity within 1σ except for the WZjj-QCD QCD-scale term, which is pulled to lower values and constrained, and for a mild constraint of a few jet-related terms. This is expected: with its large Sherpa uncertainty and no associated NF, this NP plays the role of a free normalisation for the dominant WZjj-QCD background and is therefore determined by the data. The ranking is dominated by the WZjj theory-modelling terms and the JES and JER, consistent with the correlation structure discussed above. The three polarisation configurations yield the same picture, confirming that the fit is robust and free of mis-modelling. For the f_{00} and $f_{0\text{X}}$ measurements, the uncertainty variation for the f_{XX} POI is naturally one-sided, as the POI is bounded at zero.

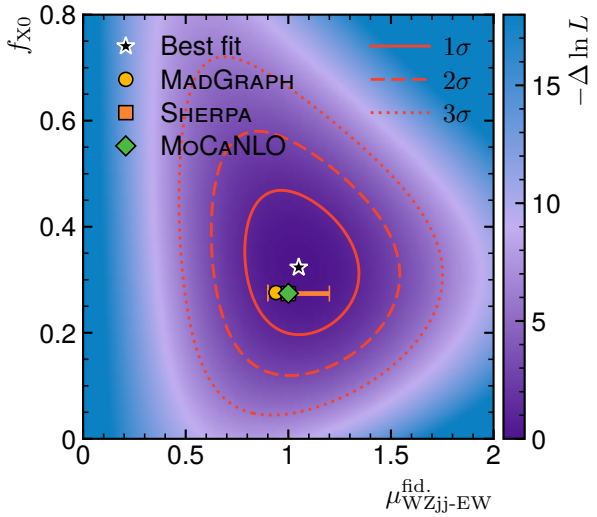
A complementary observation is that the highest-impact NPs, the WZjj-EW theory-modelling and NLO EW terms and the JES and JER, carry pulls compatible with zero: the parameters with the largest impact are not significantly pulled, so the leading templates are not mis-modelled. The largest pulls are those of the two WZjj-QCD terms, the QCD scale and the generator difference, which the data use to set the normalisation of the dominant background, as discussed in Section 7.3.3. For the $\text{W}_{\text{X}}\text{Z}_{0}\text{jj}$-EW measurement the top-ranked parameter is the NLO EW correction, which directly reshapes the signal template and therefore has the largest leverage on f_{X0}. The full set of post-fit pulls for all NPs is collected in Appendix R.



(a) f_{00} measurement

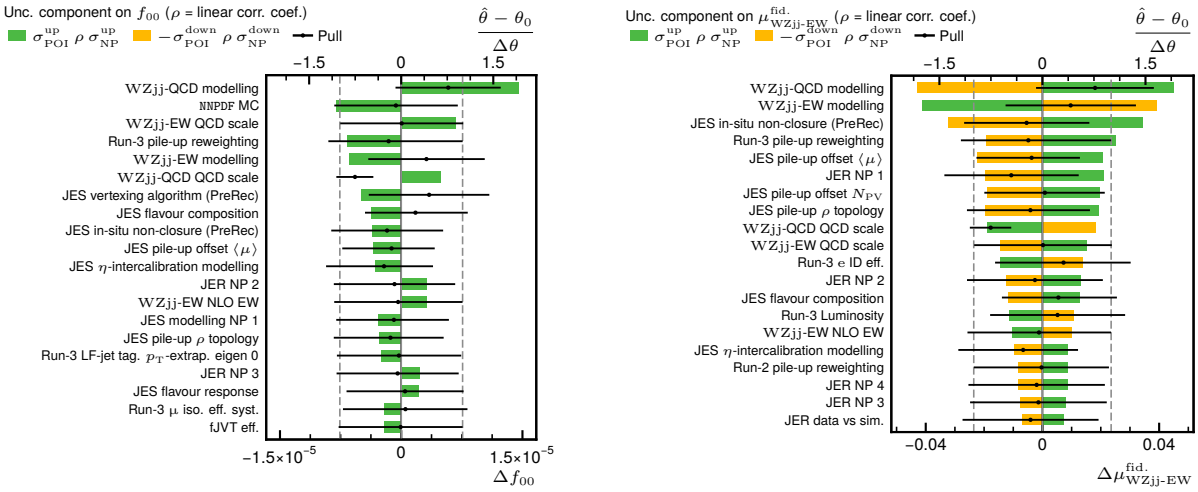


(b) f_{0X} measurement



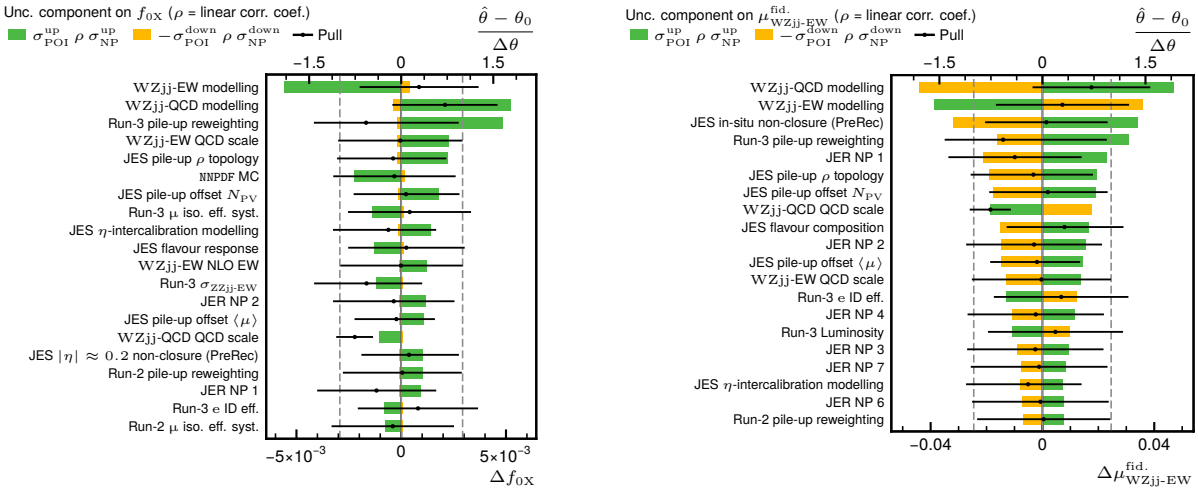
(c) f_{X0} measurement

Figure 7.12: 2D likelihood scans for the combination of Run-2 and Run-3.



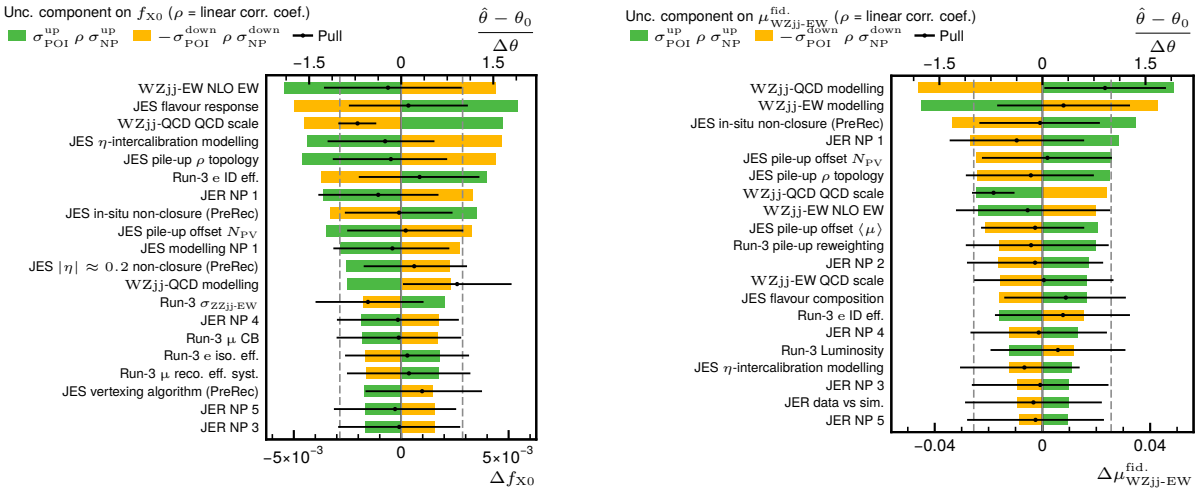
(i) f_{00} **(ii)** $\mu^{\text{fid.}}_{\text{WZjj-EW}}$

(a) f_{00} measurement



(i) $f_{0\text{X}}$ **(ii)** $\mu^{\text{fid.}}_{\text{WZjj-EW}}$

(b) $f_{0\text{X}}$ measurement



(i) f_{X0} **(ii)** $\mu^{\text{fid.}}_{\text{WZjj-EW}}$

(c) f_{X0} measurement

Figure 7.13: Ranking and pulls from the 20 most important NPs for the combination of Run-2 and Run-3, for the f_{00} (a), $f_{0\text{X}}$ (b), and f_{X0} (c) measurements. In each measurement, the ranking is shown for both POIs: f_{XX} and $\mu^{\text{fid.}}_{\text{WZjj-EW}}$.

7.3.8 Measurements

This section presents the results of the three polarisation measurements, obtained from the combined fit of the SR and CRs for the Run-2 and Run-3 datasets. The central result is the first evidence for a longitudinally polarised Z boson in WZjj-EW production: the $\mathrm{W_XZ_0}$jj-EW fit rejects the null hypothesis $f_{\mathrm{X0}} = 0$ at 4.0σ, for an expected 2.9σ. The two remaining configurations, $\mathrm{W_0Z_0}$jj-EW and $\mathrm{W_0Z_X}$jj-EW, return best-fit fractions compatible with zero, so only upper limits are set. The full pattern of significances follows directly from the data fluctuations about the SM: f_{X0} fluctuates *upward*, exceeding its expected significance, whereas f_{00} and $f_{0\mathrm{X}}$ fluctuate *downward* to the physical boundary $f = 0$ and fall short of theirs. The measured fractions, fiducial cross-sections, and signal strength are reported below with their statistical, systematic, and luminosity uncertainties (Table 7.G), compared throughout to the MadGraph, Sherpa, and MoCaNLO predictions. They are also visually summarised in Figure 7.14.

The headline result is the first evidence for a longitudinally polarised Z boson in WZjj-EW production: the combined fit yields $f_{\mathrm{X0}} = 0.324^{+0.092}_{-0.085}$, a 4.0σ rejection of the null hypothesis $f_{\mathrm{X0}} = 0$, to be compared with an expected significance of 2.9σ. The measured value is in good agreement, within one standard deviation, with the MadGraph ($f_{\mathrm{X0}} = 0.2758$), Sherpa (0.274), and MoCaNLO NLO EW (0.2745) predictions of Table 7.G. These results are also reported by the ATLAS Collaboration in [432].

For the two remaining configurations the fits return values compatible with zero and only upper limits are set. For the doubly longitudinal state $f_{00} = 0.000^{+0.053}_{-0.000}$ (0.0σ, expected 1.3σ) and for the single-longitudinal W state $f_{0\mathrm{X}} = 0.007^{+0.094}_{-0.007}$ (0.1σ, expected 2.6σ). Both central values lie below the SM predictions, the signature of a downward statistical fluctuation given the expected sensitivity. The corresponding observed $95\,\%$ CL upper limits are $f_{00} < 0.12$ and $f_{0\mathrm{X}} < 0.19$, detailed in Section 7.3.11. These constitute the first limits on the jointly longitudinal and on the W-only longitudinal WZjj-EW final states. The $f_{0\mathrm{X}}$ deficit is the more pronounced of the two: under the SM hypothesis the probability of observing $f_{0\mathrm{X}} \leq 0.007$ is $0.6\,\%$, a 2.5σ downward fluctuation, itself consistent with the slight upward fluctuation of the longitudinal-Z fraction f_{X0} obtained in the separate fit.

The signal strength is compatible with the SM, $\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}} = 1.06 \pm 0.18$, as are the combined-fit fiducial cross-sections for the two CM energies, $\sigma^{\mathrm{fid.,13\,TeV,comb.}}_{\mathrm{WZjj\text{-}EW}} = 0.353 \pm 0.061\,\mathrm{fb}$ and $\sigma^{\mathrm{fid.,13.6\,TeV,comb.}}_{\mathrm{WZjj\text{-}EW}} = 0.393 \pm 0.068\,\mathrm{fb}$, both in agreement with the MadGraph and Sherpa expectations. The NFs of the data-constrained backgrounds (Table 7.F) are consistent with unity and with the values reported in the previous WZjj-EW measurement [360]. All results are stable between the individual Run-2 and Run-3 fits and their combination, as summarised in Table 7.H.

The expected and observed significances of the three fits are collected in Table 7.I. The observed significances track the expected ones only for the combination taken as a whole: f_{X0} fluctuates upward, its 4.0σ observed significance exceeding the 2.9σ expected, driven by a central value $\sim 0.6\sigma$ above the SM. Conversely, f_{00} and $f_{0\mathrm{X}}$ fluctuate downward to the physical boundary, their observed significances being 0.0σ and 0.1σ (expected 1.3σ and 2.6σ). All three fractions nonetheless remain statistically compatible with the SM predictions within the quoted uncertainties.

Table 7.G: Measured polarisation fractions f_{00}, $f_{0\mathrm{X}}$, f_{X0}, and the fiducial signal strength $\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}}$, for WZjj-EW events. The $\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}}$ value obtained from the f_{X0} measurement is reported. The total, statistical, systematic, and luminosity uncertainties in the measurements are reported. The measurements are compared with predictions from MadGraph, Sherpa, and from NLO EW fixed-order calculations using MoCaNLO. The uncertainties in the MadGraph and Sherpa predictions include statistical, PDF, and QCD scale uncertainties, while those in the NLO EW fixed-order prediction include only QCD scale uncertainties. The calculation from MoCaNLO is done for only $\mathrm{W^+Z}$ events and therefore no inclusive WZjj-EW cross-section is provided for it.

POI	Data total (stat., syst., lumi.)	MadGraph	Sherpa	MoCaNLO
f_{00}	$0.000^{+0.053}_{-0.000}$ $(^{+0.053}_{-0.000},\ <0.001,\ <0.0001)$	0.0902 ± 0.0005	0.087 ± 0.001	0.0902 ± 0.0005
$f_{0\mathrm{X}}$	$0.007^{+0.094}_{-0.007}$ $(^{+0.093}_{-0.000},\ ^{+0.009}_{-0.007},\ ^{+0.0001}_{-0.0005})$	0.2808 ± 0.0002	0.2793 ± 0.0009	0.2758 ± 0.0008
f_{X0}	$0.324^{+0.092}_{-0.085}$ $(^{+0.091}_{-0.084},\ ^{+0.017}_{-0.018},\ ^{+0.0015}_{-0.0014})$	0.2758 ± 0.0003	0.274 ± 0.001	0.2745 ± 0.0008
$\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}}$	1.055 ± 0.181 $(\pm 0.146,\ \pm 0.107,\ \pm 0.013)$	$0.94^{+0.05}_{-0.04}$	$1.00^{+0.2}_{-0.1}$	—

Table 7.H: Measured polarisation fractions f_{00}, f_{0X}, f_{X0}, the fiducial signal strength $\mu^{\text{fid.}}_{\text{WZjj-EW}}$, and the fiducial cross-sections, for WZjj-EW events in the combined and per-run measurements. The $\mu^{\text{fid.}}_{\text{WZjj-EW}}$ value obtained from the f_{X0} measurement is reported. The total, statistical, systematic, and luminosity uncertainties in the measurements are reported.

POI	Combined	Individual fits	
		Run-2	Run-3
	total (stat., syst., lumi.)		
f_{00}	$0.000^{+0.053}_{-0.000}$ $(^{+0.053}_{-0.000},\ <0.001,\ <0.0001)$	$0.056^{+0.102}_{-0.056}$ $(^{+0.102}_{-0.055},\ ^{+0.008}_{-0.009},\ <0.0001)$	$0.000^{+0.041}_{-0.000}$ $(^{+0.041}_{-0.000},\ <0.001,\ <0.0001)$
f_{0X}	$0.007^{+0.094}_{-0.007}$ $(^{+0.093}_{-0.000},\ ^{+0.009}_{-0.007},\ ^{+0.0001}_{-0.0005})$	$0.000^{+0.114}_{-0.000}$ $(^{+0.114}_{-0.000},\ <0.001,\ <0.0001)$	$0.024^{+0.126}_{-0.024}$ $(^{+0.125}_{-0.018},\ ^{+0.016}_{-0.016},\ ^{+0.0001}_{-0.0007})$
f_{X0}	$0.324^{+0.092}_{-0.085}$ $(^{+0.091}_{-0.084},\ ^{+0.017}_{-0.018},\ ^{+0.0015}_{-0.0014})$	$0.411^{+0.148}_{-0.130}$ $(^{+0.146}_{-0.128},\ ^{+0.025}_{-0.026},\ ^{+0.0004}_{-0.0003})$	$0.270^{+0.137}_{-0.123}$ $(^{+0.134}_{-0.120},\ ^{+0.029}_{-0.029},\ ^{+0.0010}_{-0.0009})$
$\mu^{\text{fid.}}_{\text{WZjj-EW}}$	1.055 ± 0.181 $(\pm0.146,\ \pm0.107,\ \pm0.013)$	1.037 ± 0.236 $(\pm0.204,\ \pm0.117,\ \pm0.010)$	1.004 ± 0.252 $(\pm0.206,\ \pm0.143,\ \pm0.024)$
$\sigma^{\text{fid.,13 TeV}}_{\text{WZjj-EW}}$	0.353 ± 0.061 $(\pm0.049,\ \pm0.036,\ \pm0.004)$	0.347 ± 0.079 $(\pm0.068,\ \pm0.039,\ \pm0.003)$	—
$\sigma^{\text{fid.,13.6 TeV}}_{\text{WZjj-EW}}$	0.393 ± 0.068 $(\pm0.054,\ \pm0.040,\ \pm0.005)$	—	0.374 ± 0.094 $(\pm0.077,\ \pm0.053,\ \pm0.009)$

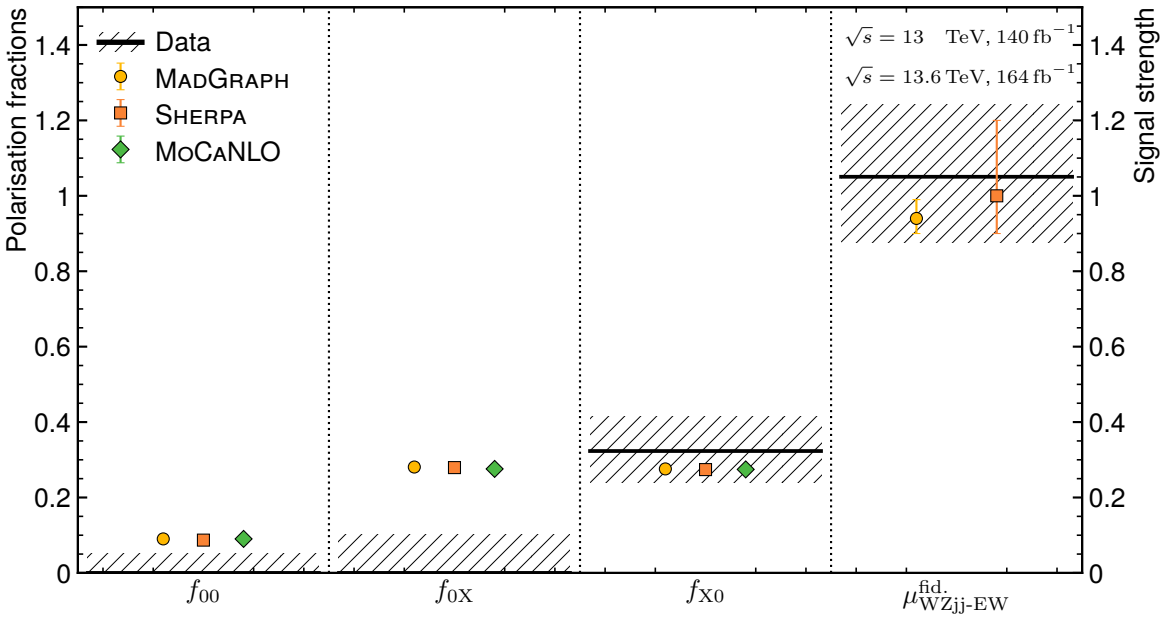


Figure 7.14: Summary of the measured polarisation fractions f_{00}, f_{0X}, and f_{X0}, and of the fiducial signal strength $\mu^{\text{fid.}}_{\text{WZjj-EW}}$ read on the right-hand axis, for the combined Run-2 and Run-3 fit, compared with the MadGraph, Sherpa, and MoCaNLO predictions. MoCaNLO provides no inclusive cross-section, so it has no marker in the signal-strength slot.

7.3.9 Impact of systematics

The impact of each group of systematics on the POIs f_{XX} and $\mu^{\text{fid.}}_{\text{WZjj-EW}}$ is shown in Tables 7.J to 7.L. The impact of each group of systematics is shown for Run-2, Run-3, and their combination. The contribution of correlated systematics between Run-2 and Run-3 is depicted in the extra lines. The Theory modelling category includes the QCD scale, the difference in generators, the uncertainty on the overall cross-section of the process, and NLO effects as presented in Section 7.2.2. The PDF category includes both PDF and α_S variations.

Tables 7.J to 7.L show that the three measurements are statistically dominated: the total uncertainty is driven by the data statistics, with the total systematic contribution a factor of a few smaller for the polarisation fractions, while for $\mu^{\text{fid.}}_{\text{WZjj-EW}}$ the systematic term reaches three quarters of the statistical one. Among the systematic sources, the JES and JER, the finite MC statistics of the templates (for the polarisation fractions) and the WZjj theory-modelling uncertainties (QCD scale and generator difference) are the largest. All remaining categories (the electron, muon, and MET calibrations, pile-up, luminosity, and the misidentified leptons estimate) contribute at the percent level, up to about $3\,\%$ for the Run-3 electron, pile-up, and luminosity terms on the cross-section, and are subdominant. The luminosity term is negligible for the polarisation fractions, as expected for a measurement of a cross-section ratio in which the dominant backgrounds are constrained. The larger relative systematic seen for $\mu^{\text{fid.}}_{\text{WZjj-EW}}$ compared

Table 7.I: Measured polarisation fractions f_{00}, $f_{0\mathrm{X}}$, and $f_{\mathrm{X}0}$, together with the EW fiducial cross-section $\sigma^{\text{fid.}}_{\text{WZjj-EW}}$ and the signal strength $\mu^{\text{fid.}}_{\text{WZjj-EW}}$, obtained from the individual Run-2 and Run-3 fits and from their combination. The quoted σ values are the corresponding significances. Sub-table (a) gives the expected results and sub-table (b) the observed results.

(a) Expected

Parameter / Measurement	Run	f_{XX} Individual	f_{XX} Combined	$\sigma^{\text{fid.}}_{\text{WZjj-EW}}$ [fb] Individual	$\mu^{\text{fid.}}_{\text{WZjj-EW}}$ Combined
f_{00}	2	$0.088^{+0.119}_{-0.088}$ 0.88σ	$0.087^{+0.078}_{-0.069}$ 1.29σ	$0.335^{+0.084}_{-0.079}$	$1.000^{+0.195}_{-0.185}$
	3	$0.087^{+0.109}_{-0.087}$ 0.96σ		$0.373^{+0.098}_{-0.093}$	
$f_{0\mathrm{X}}$	2	$0.280^{+0.177}_{-0.160}$ 1.79σ	$0.279^{+0.117}_{-0.109}$ 2.59σ	$0.335^{+0.086}_{-0.080}$	$1.000^{+0.200}_{-0.190}$
	3	$0.279^{+0.161}_{-0.146}$ 1.91σ		$0.373^{+0.103}_{-0.096}$	
$f_{\mathrm{X}0}$	2	$0.274^{+0.157}_{-0.139}$ 2.03σ	$0.274^{+0.104}_{-0.095}$ 2.91σ	$0.335^{+0.091}_{-0.086}$	$1.000^{+0.212}_{-0.202}$
	3	$0.273^{+0.147}_{-0.129}$ 2.15σ		$0.373^{+0.106}_{-0.100}$	

(b) Observed

Parameter / Measurement	Run	f_{XX} Individual	f_{XX} Combined	$\sigma^{\text{fid.}}_{\text{WZjj-EW}}$ [fb] Individual	$\mu^{\text{fid.}}_{\text{WZjj-EW}}$ Combined
f_{00}	2	$0.056^{+0.102}_{-0.056}$ 0.63σ	$0.000^{+0.053}_{-0.000}$ 0.00σ	$0.354^{+0.078}_{-0.073}$	$1.079^{+0.174}_{-0.166}$
	3	$0.000^{+0.041}_{-0.000}$ 0.00σ		$0.409^{+0.091}_{-0.086}$	
$f_{0\mathrm{X}}$	2	$0.000^{+0.114}_{-0.000}$ 0.00σ	$0.007^{+0.094}_{-0.007}$ 0.08σ	$0.334^{+0.076}_{-0.072}$	$1.028^{+0.175}_{-0.163}$
	3	$0.024^{+0.126}_{-0.024}$ 0.20σ		$0.397^{+0.092}_{-0.086}$	
$f_{\mathrm{X}0}$	2	$0.411^{+0.148}_{-0.130}$ 3.44σ	$0.324^{+0.092}_{-0.085}$ 4.00σ	$0.347^{+0.082}_{-0.077}$	$1.055^{+0.188}_{-0.178}$
	3	$0.270^{+0.137}_{-0.123}$ 2.23σ		$0.374^{+0.097}_{-0.092}$	

with the polarisation fractions reflects the residual sensitivity of the absolute normalisation to the WZjj modelling, which partially cancels in the fractions.

7.3.10 Fiducial cross-section

Beyond the polarisation fractions, the standalone Run-3 fit provides the first measurement of the integrated fiducial WZjj-EW production cross-section at $\sqrt{s} = 13.6\,\mathrm{TeV}$. Per lepton flavour:

$$\sigma^{\text{fid.,13.6 TeV}}_{\mathrm{WZjj\text{-}EW}\to\ell'\nu_\ell\ell\ell\mathrm{jj}} = 0.374 \pm 0.077\,(\text{stat.}) \pm 0.053\,(\text{syst.}) \pm 0.009\,(\text{lumi.})\,\text{fb}\,, \tag{7.4}$$

in agreement with the SHERPA prediction of $0.373^{+0.065}_{-0.041}\,\text{fb}$. This standalone Run-3 result is marginally below the combined-fit value $\sigma^{\text{fid.,13.6 TeV,comb.}}_{\mathrm{WZjj\text{-}EW}} = 0.393 \pm 0.068\,\text{fb}$ of Table 7.H. The two are statistically compatible, the combined-fit value being tightened by the common $\mu^{\text{fid.}}_{\mathrm{WZjj\text{-}EW}}$ constraint from the Run-2 data. The measurement is statistically dominated, and its central value is stable to within $10\,\%$ across the three polarisation fits, confirming that the cross-section is insensitive to the fraction being probed. The Run-2 fit gives $\sigma^{\text{fid.,13 TeV}}_{\mathrm{WZjj\text{-}EW}} = 0.347 \pm 0.079\,\text{fb}$, also consistent with the MADGRAPH and SHERPA expectations.

The cross-section measurement for Run-2 is consistent with the previous WZjj-EW measurement [360] given in Equation 6.1.

7.3.11 Upper limits

The best-fit fractions of the doubly longitudinal (f_{00}) and single-longitudinal-W ($f_{0\mathrm{X}}$) configurations sit at the physical boundary $f = 0$, so a two-sided interval is not meaningful and one-sided $95\,\%$ CL upper

Table 7.J: Uncertainty breakdown for the $\mathrm{W_0Z_0jj}$-EW measurement. The contribution of each category of systematics is shown as well as the total systematic, statistical, and total uncertainties. The results are shown for Run-2, Run-3, and their combination. The contribution of correlated systematics between Run-2 and Run-3 is depicted in the extra lines. The polarisation fraction is fitted at its boundary, so the uncertainties on the fraction are given in absolute units, those on $\sigma^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}}$ and $\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}}$ as relative ones.

Parameter		f_{00}				$\sigma^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}}$ [fb]	$\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}}$
Measurement		**Individual**		**Combined**		**Individual**	**Combined**
Run	**2**	$0.056^{+0.102}_{-0.056}$	0.63σ	$0.000^{+0.053}_{-0.000}$	0.00σ	$0.354^{+0.078}_{-0.073}$	$1.079^{+0.174}_{-0.166}$
	3	$0.000^{+0.041}_{-0.000}$	0.00σ			$0.409^{+0.091}_{-0.086}$	

Source	**Absolute (f_{00}) and relative [%] (σ, μ) uncertainty and $^{\mathrm{up}}_{\mathrm{down}}$ variations for**						**Run**	**2 3 correlated only**
WZjj-EW Theory modelling	0.0021	$^{+0.0023}_{-0.0013}$	0.0000	$^{+0.0000}_{-0.0000}$	4.9	$^{+4.8}_{-5.1}$	4.0	$^{+4.0}_{-4.2}$
	0.0000	$^{+0.0000}_{-0.0000}$			3.4	$^{+3.4}_{-3.5}$		
WZjj-QCD Theory modelling	0.0050	$^{+0.0057}_{-0.0031}$	0.0000	$^{+0.0000}_{-0.0000}$	5.4	$^{+5.6}_{-5.4}$	4.3	$^{+4.5}_{-4.4}$
	0.0000	$^{+0.0000}_{-0.0000}$			5.8	$^{+6.1}_{-5.9}$		
WZjj-INT Theory modelling	0.0001	$^{+0.0001}_{-0.0000}$	0.0000	$^{+0.0000}_{-0.0000}$	0.27	$^{+0.26}_{-0.28}$	0.35	$^{+0.35}_{-0.36}$
	0.0000	$^{+0.0000}_{-0.0000}$			0.47	$^{+0.46}_{-0.48}$		
PDFs	0.0006	$^{+0.0006}_{-0.0006}$	0.0000	$^{+0.0000}_{-0.0000}$	0.82	$^{+0.82}_{-0.84}$	0.75	$^{+0.76}_{-0.77}$
	0.0000	$^{+0.0000}_{-0.0000}$			0.78	$^{+0.79}_{-0.80}$		
Electrons	0.0012	$^{+0.0014}_{-0.0007}$	0.0000	$^{+0.0000}_{-0.0000}$	0.85	$^{+0.84}_{-0.89}$	0.44	$^{+0.44}_{-0.46}$
	0.0000	$^{+0.0000}_{-0.0000}$	0.0000	$^{+0.0000}_{-0.0000}$	2.9	$^{+2.9}_{-3.0}$	1.4	$^{+1.4}_{-1.4}$
b-tagging	0.0004	$^{+0.0005}_{-0.0003}$	0.0000	$^{+0.0000}_{-0.0000}$	0.17	$^{+0.18}_{-0.17}$	0.08	$^{+0.08}_{-0.08}$
	0.0000	$^{+0.0000}_{-0.0000}$	0.0000	$^{+0.0000}_{-0.0000}$	0.79	$^{+0.80}_{-0.78}$	0.35	$^{+0.36}_{-0.35}$
Jets	0.0036	$^{+0.0027}_{-0.0039}$	0.0000	$^{+0.0000}_{-0.0000}$	5.8	$^{+6.1}_{-5.8}$	0.12	$^{+0.12}_{-0.11}$
	0.0000	$^{+0.0000}_{-0.0000}$	0.0000	$^{+0.0000}_{-0.0000}$	8.9	$^{+9.0}_{-9.0}$	3.1	$^{+3.2}_{-3.0}$
			0.0000	$^{+0.0000}_{-0.0000}$			4.4	$^{+4.6}_{-4.5}$
MET	0.0009	$^{+0.0009}_{-0.0008}$	0.0000	$^{+0.0000}_{-0.0000}$	0.67	$^{+0.70}_{-0.66}$	0.30	$^{+0.31}_{-0.30}$
	0.0000	$^{+0.0000}_{-0.0000}$	0.0000	$^{+0.0000}_{-0.0000}$	0.51	$^{+0.53}_{-0.50}$	0.25	$^{+0.26}_{-0.24}$
Muons	0.0009	$^{+0.0005}_{-0.0010}$	0.0000	$^{+0.0000}_{-0.0000}$	0.93	$^{+0.90}_{-0.96}$	0.45	$^{+0.45}_{-0.47}$
	0.0000	$^{+0.0000}_{-0.0000}$	0.0000	$^{+0.0000}_{-0.0000}$	1.4	$^{+1.3}_{-1.4}$	0.66	$^{+0.65}_{-0.68}$
Pile-up	0.0003	$^{+0.0003}_{-0.0002}$	0.0000	$^{+0.0000}_{-0.0000}$	1.5	$^{+1.6}_{-1.5}$	0.77	$^{+0.81}_{-0.77}$
	0.0000	$^{+0.0000}_{-0.0000}$	0.0000	$^{+0.0000}_{-0.0000}$	1.5	$^{+1.7}_{-1.6}$	2.1	$^{+2.3}_{-1.8}$
Other backgrounds	0.0008	$^{+0.0007}_{-0.0008}$	0.0000	$^{+0.0000}_{-0.0000}$	1.0	$^{+1.0}_{-1.1}$	0.58	$^{+0.58}_{-0.62}$
	0.0000	$^{+0.0000}_{-0.0000}$	0.0000	$^{+0.0000}_{-0.0000}$	1.8	$^{+1.8}_{-1.9}$	0.80	$^{+0.81}_{-0.84}$
			0.0000	$^{+0.0000}_{-0.0000}$			0.43	$^{+0.42}_{-0.45}$
Misidentified leptons	0.0006	$^{+0.0004}_{-0.0007}$	0.0000	$^{+0.0000}_{-0.0000}$	1.4	$^{+1.4}_{-1.4}$	0.77	$^{+0.80}_{-0.76}$
	0.0000	$^{+0.0000}_{-0.0000}$	0.0000	$^{+0.0000}_{-0.0000}$	1.0	$^{+1.1}_{-1.0}$	0.56	$^{+0.59}_{-0.54}$
MC statistics	0.0062	$^{+0.0038}_{-0.0071}$	0.0000	$^{+0.0000}_{-0.0000}$	0.92	$^{+0.89}_{-0.96}$	0.59	$^{+0.58}_{-0.60}$
	0.0000	$^{+0.0000}_{-0.0000}$			0.84	$^{+0.84}_{-0.86}$		
All systematics	0.0093	$^{+0.0080}_{-0.0089}$	0.0000	$^{+0.0000}_{-0.0000}$	9.8	$^{+10.0}_{-9.9}$	8.7	$^{+8.9}_{-8.7}$
	0.0000	$^{+0.0000}_{-0.0000}$			12.0	$^{+12.2}_{-12.1}$		
Statistics	0.0905	$^{+0.1018}_{-0.0550}$	0.1345	$^{+0.0531}_{-0.0000}$	18.7	$^{+19.5}_{-18.0}$	13.0	$^{+13.4}_{-12.6}$
	0.0508	$^{+0.0405}_{-0.0000}$			17.9	$^{+18.6}_{-17.2}$		
Luminosity	0.0001	$^{+0.0001}_{-0.0000}$	0.0000	$^{+0.0000}_{-0.0000}$	0.97	$^{+0.94}_{-1.0}$	0.49	$^{+0.48}_{-0.50}$
	0.0000	$^{+0.0000}_{-0.0000}$	0.0000	$^{+0.0000}_{-0.0000}$	2.3	$^{+2.2}_{-2.4}$	1.0	$^{+1.0}_{-1.1}$
Total	0.0910	$^{+0.1021}_{-0.0557}$	0.1345	$^{+0.0531}_{-0.0000}$	21.1	$^{+21.9}_{-20.5}$	15.6	$^{+16.1}_{-15.3}$
	0.0508	$^{+0.0405}_{-0.0000}$			21.6	$^{+22.3}_{-21.1}$		

Table 7.K: Uncertainty breakdown for the $\mathrm{W_0Z_Xjj}$-EW measurement. The contribution of each category of systematics is shown as well as the total systematic, statistical, and total uncertainties. The results are shown for Run-2, Run-3, and their combination. The contribution of correlated systematics between Run-2 and Run-3 is depicted in the extra lines. The polarisation fraction is fitted at its boundary, so the uncertainties on the fraction are given in absolute units, those on $\sigma^{\text{fid.}}_{\text{WZjj-EW}}$ and $\mu^{\text{fid.}}_{\text{WZjj-EW}}$ as relative ones.

Parameter	f_{0X}				$\sigma^{\text{fid.}}_{\text{WZjj-EW}}$ [fb]	$\mu^{\text{fid.}}_{\text{WZjj-EW}}$
Measurement	**Individual**		**Combined**		**Individual**	**Combined**
Run	**2** $0.000^{+0.114}_{-0.000}$	0.00σ	$0.007^{+0.094}_{-0.007}$	0.08σ	$0.334^{+0.076}_{-0.072}$	$1.028^{+0.175}_{-0.163}$
	3 $0.024^{+0.126}_{-0.024}$	0.20σ			$0.397^{+0.092}_{-0.086}$	

Source	**Absolute (f_{0X}) and relative [%] (σ, μ) uncertainty and $^{\text{up}}_{\text{down}}$ variations for**						**Run**	**2** **3** **correlated only**
WZjj-EW Theory modelling	0.0000	$^{+0.0000}_{-0.0000}$	0.0099	$^{+0.0026}_{-0.0056}$	4.7	$^{+4.6}_{-4.9}$	3.8	$^{+3.8}_{-4.0}$
	0.0098	$^{+0.0039}_{-0.0089}$			3.4	$^{+3.4}_{-3.6}$		
WZjj-QCD Theory modelling	0.0000	$^{+0.0000}_{-0.0000}$	0.0084	$^{+0.0052}_{-0.0011}$	5.4	$^{+5.5}_{-5.4}$	4.6	$^{+4.9}_{-4.6}$
	0.0109	$^{+0.0094}_{-0.0051}$			6.1	$^{+6.4}_{-6.1}$		
WZjj-INT Theory modelling	0.0000	$^{+0.0000}_{-0.0000}$	0.0010	$^{+0.0000}_{-0.0006}$	0.26	$^{+0.25}_{-0.27}$	0.34	$^{+0.33}_{-0.36}$
	0.0011	$^{+0.0002}_{-0.0011}$			0.49	$^{+0.49}_{-0.52}$		
PDFs	0.0000	$^{+0.0000}_{-0.0000}$	0.0036	$^{+0.0004}_{-0.0022}$	0.66	$^{+0.66}_{-0.67}$	0.63	$^{+0.64}_{-0.64}$
	0.0029	$^{+0.0009}_{-0.0028}$			0.71	$^{+0.72}_{-0.73}$		
Electrons	0.0000	$^{+0.0000}_{-0.0000}$	0.0006	$^{+0.0003}_{-0.0002}$	0.91	$^{+0.89}_{-0.94}$	0.46	$^{+0.45}_{-0.48}$
	0.0020	$^{+0.0013}_{-0.0015}$	0.0016	$^{+0.0006}_{-0.0008}$	2.7	$^{+2.7}_{-2.9}$	1.3	$^{+1.3}_{-1.3}$
b-tagging	0.0000	$^{+0.0000}_{-0.0000}$	0.0002	$^{+0.0001}_{-0.0000}$	0.15	$^{+0.16}_{-0.15}$	0.08	$^{+0.09}_{-0.08}$
	0.0011	$^{+0.0008}_{-0.0007}$	0.0008	$^{+0.0003}_{-0.0004}$	0.92	$^{+0.96}_{-0.93}$	0.44	$^{+0.46}_{-0.44}$
Jets	0.0000	$^{+0.0000}_{-0.0000}$	0.0005	$^{+0.0003}_{-0.0000}$	6.7	$^{+6.9}_{-6.5}$	0.12	$^{+0.12}_{-0.12}$
	0.0098	$^{+0.0093}_{-0.0034}$	0.0020	$^{+0.0010}_{-0.0008}$	8.7	$^{+9.1}_{-8.8}$	3.2	$^{+3.4}_{-3.1}$
			0.0069	$^{+0.0039}_{-0.0019}$			4.6	$^{+5.0}_{-4.7}$
MET	0.0000	$^{+0.0000}_{-0.0000}$	0.0013	$^{+0.0001}_{-0.0008}$	0.75	$^{+0.78}_{-0.74}$	0.34	$^{+0.36}_{-0.34}$
	0.0019	$^{+0.0004}_{-0.0018}$	0.0014	$^{+0.0001}_{-0.0009}$	0.55	$^{+0.57}_{-0.53}$	0.31	$^{+0.33}_{-0.31}$
Muons	0.0000	$^{+0.0000}_{-0.0000}$	0.0012	$^{+0.0001}_{-0.0007}$	0.82	$^{+0.81}_{-0.86}$	0.43	$^{+0.43}_{-0.46}$
	0.0034	$^{+0.0014}_{-0.0031}$	0.0025	$^{+0.0006}_{-0.0014}$	1.3	$^{+1.3}_{-1.4}$	0.61	$^{+0.60}_{-0.64}$
Pile-up	0.0000	$^{+0.0000}_{-0.0000}$	0.0016	$^{+0.0010}_{-0.0001}$	1.3	$^{+1.3}_{-1.3}$	0.68	$^{+0.73}_{-0.68}$
	0.0068	$^{+0.0069}_{-0.0013}$	0.0046	$^{+0.0048}_{-0.0002}$	2.5	$^{+2.7}_{-2.4}$	1.7	$^{+3.0}_{-1.6}$
Other backgrounds	0.0000	$^{+0.0000}_{-0.0000}$	0.0011	$^{+0.0004}_{-0.0006}$	1.0	$^{+1.0}_{-1.1}$	0.58	$^{+0.58}_{-0.62}$
	0.0043	$^{+0.0013}_{-0.0042}$	0.0020	$^{+0.0005}_{-0.0012}$	1.6	$^{+1.6}_{-1.6}$	0.69	$^{+0.71}_{-0.71}$
			0.0011	$^{+0.0001}_{-0.0007}$			0.42	$^{+0.41}_{-0.44}$
Misidentified leptons	0.0000	$^{+0.0000}_{-0.0000}$	0.0012	$^{+0.0003}_{-0.0007}$	1.6	$^{+1.6}_{-1.5}$	0.80	$^{+0.84}_{-0.79}$
	0.0008	$^{+0.0002}_{-0.0008}$	0.0014	$^{+0.0001}_{-0.0009}$	1.2	$^{+1.3}_{-1.2}$	0.73	$^{+0.78}_{-0.71}$
MC statistics	0.0000	$^{+0.0000}_{-0.0000}$	0.0078	$^{+0.0006}_{-0.0049}$	0.63	$^{+0.64}_{-0.64}$	0.54	$^{+0.55}_{-0.56}$
	0.0099	$^{+0.0021}_{-0.0097}$			0.98	$^{+0.97}_{-1.0}$		
All systematics	0.0000	$^{+0.0000}_{-0.0000}$	0.0185	$^{+0.0088}_{-0.0071}$	10.2	$^{+10.4}_{-10.2}$	8.7	$^{+9.4}_{-8.9}$
	0.0224	$^{+0.0157}_{-0.0160}$			12.0	$^{+12.6}_{-12.2}$		
Statistics	0.2069	$^{+0.1144}_{-0.0000}$	0.1498	$^{+0.0933}_{-0.0000}$	19.5	$^{+20.2}_{-18.7}$	13.5	$^{+14.1}_{-13.1}$
	0.1278	$^{+0.1249}_{-0.0185}$			18.4	$^{+19.3}_{-17.8}$		
Luminosity	0.0000	$^{+0.0000}_{-0.0000}$	0.0001	$^{+0.0001}_{-0.0000}$	0.98	$^{+0.96}_{-1.0}$	0.50	$^{+0.49}_{-0.53}$
	0.0007	$^{+0.0001}_{-0.0007}$	0.0007	$^{+0.0000}_{-0.0005}$	2.2	$^{+2.2}_{-2.3}$	0.98	$^{+0.97}_{-1.0}$
Total	0.2069	$^{+0.1144}_{-0.0000}$	0.1509	$^{+0.0938}_{-0.0072}$	22.0	$^{+22.7}_{-21.4}$	16.1	$^{+17.0}_{-15.9}$
	0.1297	$^{+0.1259}_{-0.0244}$			22.1	$^{+23.1}_{-21.7}$		

Table 7.L: Uncertainty breakdown for the $W_X Z_0 jj$-EW measurement. The contribution of each category of systematics is shown as well as the total systematic, statistical, and total uncertainties. The results are shown for Run-2, Run-3, and their combination. The contribution of correlated systematics between Run-2 and Run-3 is depicted in the extra lines. All uncertainties are relative.

Parameter		f_{X0}				$\sigma^{\text{fid.}}_{\text{WZjj-EW}}$ [fb]		$\mu^{\text{fid.}}_{\text{WZjj-EW}}$	
Measurement		**Individual**		**Combined**		**Individual**		**Combined**	
Run	**2**	$0.411^{+0.148}_{-0.130}$	3.44σ	$0.324^{+0.092}_{-0.085}$	4.00σ	$0.347^{+0.082}_{-0.077}$		$1.055^{+0.188}_{-0.178}$	
	3	$0.270^{+0.137}_{-0.123}$	2.23σ			$0.374^{+0.097}_{-0.092}$			
Source		**Relative uncertainty [%] and $^{\text{up}}_{\text{down}}$ variations for**						**Run**	**2**
									3
								correlated only	
$WZjj$-EW Theory modelling		0.64	$^{+0.69}_{-0.62}$	1.7	$^{+1.5}_{-1.8}$	5.2	$^{+5.1}_{-5.4}$	4.9	$^{+4.7}_{-5.1}$
		2.0	$^{+2.0}_{-2.1}$			3.9	$^{+3.9}_{-4.0}$		
$WZjj$-QCD Theory modelling		2.4	$^{+2.6}_{-2.4}$	1.6	$^{+1.6}_{-1.6}$	6.1	$^{+6.2}_{-6.1}$	5.0	$^{+5.2}_{-5.0}$
		3.8	$^{+3.8}_{-3.9}$			7.5	$^{+7.8}_{-7.5}$		
$WZjj$-INT Theory modelling		0.05	$^{+0.06}_{-0.05}$	0.02	$^{+0.02}_{-0.02}$	0.28	$^{+0.28}_{-0.29}$	0.38	$^{+0.37}_{-0.39}$
		0.14	$^{+0.14}_{-0.16}$			0.54	$^{+0.53}_{-0.56}$		
PDFs		0.17	$^{+0.18}_{-0.16}$	0.29	$^{+0.28}_{-0.30}$	0.65	$^{+0.65}_{-0.66}$	0.68	$^{+0.68}_{-0.69}$
		0.77	$^{+0.76}_{-0.83}$			0.83	$^{+0.83}_{-0.84}$		
Electrons		0.54	$^{+0.58}_{-0.51}$	0.29	$^{+0.31}_{-0.29}$	0.86	$^{+0.84}_{-0.89}$	0.40	$^{+0.40}_{-0.42}$
		2.0	$^{+2.2}_{-2.0}$	1.3	$^{+1.4}_{-1.3}$	3.1	$^{+3.0}_{-3.2}$	1.6	$^{+1.5}_{-1.6}$
b-tagging		0.01	$^{+0.01}_{-0.01}$	0.08	$^{+0.09}_{-0.08}$	0.20	$^{+0.21}_{-0.20}$	0.09	$^{+0.10}_{-0.09}$
		0.74	$^{+0.73}_{-0.79}$	0.48	$^{+0.48}_{-0.50}$	0.97	$^{+1.0}_{-0.97}$	0.49	$^{+0.50}_{-0.49}$
Jets		4.3	$^{+4.3}_{-4.7}$	0.17	$^{+0.17}_{-0.18}$	7.3	$^{+7.6}_{-7.3}$	0.14	$^{+0.15}_{-0.14}$
		7.7	$^{+8.1}_{-8.0}$	1.4	$^{+1.4}_{-1.4}$	10.5	$^{+10.8}_{-10.3}$	3.2	$^{+3.4}_{-3.3}$
				3.3	$^{+3.4}_{-3.4}$			5.8	$^{+6.0}_{-5.8}$
MET		0.20	$^{+0.19}_{-0.22}$	0.06	$^{+0.06}_{-0.07}$	0.73	$^{+0.76}_{-0.72}$	0.38	$^{+0.39}_{-0.37}$
		0.48	$^{+0.52}_{-0.47}$	0.25	$^{+0.27}_{-0.25}$	0.52	$^{+0.55}_{-0.52}$	0.26	$^{+0.27}_{-0.25}$
Muons		0.21	$^{+0.22}_{-0.22}$	0.13	$^{+0.13}_{-0.14}$	0.96	$^{+0.95}_{-1.0}$	0.44	$^{+0.44}_{-0.46}$
		1.2	$^{+1.3}_{-1.3}$	0.77	$^{+0.78}_{-0.78}$	1.4	$^{+1.4}_{-1.5}$	0.76	$^{+0.75}_{-0.78}$
Pile-up		0.76	$^{+0.75}_{-0.85}$	0.32	$^{+0.31}_{-0.34}$	1.4	$^{+1.5}_{-1.4}$	0.76	$^{+0.80}_{-0.76}$
		0.74	$^{+0.88}_{-0.76}$	0.42	$^{+0.36}_{-0.47}$	1.0	$^{+1.2}_{-1.1}$	1.7	$^{+1.9}_{-1.5}$
Other backgrounds		0.64	$^{+0.70}_{-0.64}$	0.39	$^{+0.40}_{-0.40}$	1.1	$^{+1.1}_{-1.2}$	0.69	$^{+0.69}_{-0.73}$
		1.6	$^{+1.8}_{-1.6}$	0.70	$^{+0.76}_{-0.70}$	2.2	$^{+2.2}_{-2.3}$	0.99	$^{+1.0}_{-1.0}$
				0.02	$^{+0.02}_{-0.02}$			0.46	$^{+0.45}_{-0.47}$
Misidentified leptons		0.81	$^{+0.78}_{-0.89}$	0.61	$^{+0.60}_{-0.65}$	1.2	$^{+1.2}_{-1.2}$	0.77	$^{+0.80}_{-0.76}$
		0.46	$^{+0.45}_{-0.50}$	0.20	$^{+0.22}_{-0.20}$	1.3	$^{+1.4}_{-1.3}$	0.62	$^{+0.65}_{-0.60}$
MC statistics		2.8	$^{+2.9}_{-2.8}$	2.4	$^{+2.3}_{-2.5}$	1.7	$^{+1.7}_{-1.8}$	1.1	$^{+1.1}_{-1.1}$
		4.4	$^{+4.2}_{-4.8}$			1.3	$^{+1.3}_{-1.3}$		
All systematics		5.9	$^{+6.0}_{-6.2}$	5.2	$^{+5.3}_{-5.4}$	11.3	$^{+11.5}_{-11.4}$	10.2	$^{+10.4}_{-10.3}$
		10.3	$^{+10.7}_{-10.8}$			14.2	$^{+14.6}_{-14.2}$		
Statistics		32.5	$^{+35.5}_{-31.0}$	26.6	$^{+28.0}_{-25.8}$	19.7	$^{+20.4}_{-19.0}$	13.8	$^{+14.4}_{-13.3}$
		45.8	$^{+49.7}_{-44.4}$			20.5	$^{+21.3}_{-19.9}$		
Luminosity		0.09	$^{+0.10}_{-0.08}$	0.12	$^{+0.12}_{-0.12}$	0.98	$^{+0.97}_{-1.0}$	0.46	$^{+0.45}_{-0.47}$
		0.34	$^{+0.37}_{-0.33}$	0.42	$^{+0.45}_{-0.41}$	2.4	$^{+2.4}_{-2.5}$	1.1	$^{+1.1}_{-1.2}$
Total		33.0	$^{+36.0}_{-31.7}$	27.1	$^{+28.5}_{-26.3}$	22.7	$^{+23.5}_{-22.2}$	17.2	$^{+17.8}_{-16.9}$
		47.0	$^{+50.8}_{-45.7}$			25.1	$^{+25.9}_{-24.6}$		

limits are quoted instead. The observed limits are computed for the combination of Run-2 and Run-3 with the asymptotic formulae for the one-sided test statistic $\tilde{q}_{f_{\mathrm{XX}}}$ and the CL_s criterion (Section O.6). The expected median limits and the bands of Table 7.M are obtained from pseudo-experiments generated under the SM hypothesis.

The observed limits are $f_{00} < 0.12$ and $f_{0\mathrm{X}} < 0.19$, both tighter than the expected median limits of 0.21 and 0.46, respectively. This is the expected consequence of the downward fluctuation of the two fractions towards $f = 0$. Together with the 4.0σ evidence for f_{X0}, these constraints complete the polarisation picture of WZjj-EW production: a clear longitudinal-Z component alongside longitudinal-W and doubly longitudinal components that are consistent with, but below, the SM expectation. Together, they constitute the first limits on the latter two final states.

The observed and expected limits are summarised in Table 7.M. The expected median limit is given with its ±1, ±2, and $\pm3\,\sigma$ bands under the SM hypothesis.

Table 7.M: Observed and expected $95\,\%$ CL upper limits on the f_{00} and $f_{0\mathrm{X}}$ fractions, for the combination of Run-2 and Run-3. The expected median limit is given with its ±1, ±2, and $\pm3\,\sigma$ bands under the SM hypothesis.

Fraction	Expected	$\pm1\sigma$	$\pm2\sigma$	$\pm3\sigma$	Observed
f_{00}	0.21	[0.15, 0.29]	[0.08, 0.36]	[0.02, 0.44]	0.12
$f_{0\mathrm{X}}$	0.46	[0.35, 0.58]	[0.24, 0.70]	[0.13, 0.81]	0.19

7.3.12 Consistency between polarisation states

The observation of an $f_{0\mathrm{X}}$ fraction close to zero implies an increase in $\mathrm{W_TZ_0jj}$-EW and $\mathrm{W_TZ_Tjj}$-EW events. As $\mathrm{W_XZ_0jj}$-EW events are composed mainly of $\mathrm{W_TZ_0jj}$-EW and $\mathrm{W_0Z_0jj}$-EW events, the measured $f_{0\mathrm{X}}$ fraction close to zero is therefore compatible with an increased f_{X0} fraction, if the remainder goes preferentially into the $\mathrm{W_TZ_0jj}$-EW component. This is indeed what is observed. Both measurements of $f_{0\mathrm{X}}$ and f_{X0} are therefore consistent. Since each single-longitudinal fit fixes the relative composition of the non-targeted states to the SM one, the fitted fractions are not additive and, with the present statistical accuracy, the two $\mathrm{W_TZ_0jj}$-EW and $\mathrm{W_TZ_Tjj}$-EW components cannot be disentangled in any of the fits. It is therefore not possible to quantify the exact level of consistency between the two measured values of $f_{0\mathrm{X}}$ and f_{X0}. However, as in the present measurement setup a low $f_{0\mathrm{X}}$ fraction is compatible with an increased f_{X0} fraction, the lowest probability that needs to be estimated to quantify the consistency of the observation with the SM is the probability to observe a fraction $f_{0\mathrm{X}} \leq 0.007$.

The procedure developed to quantify the consistency of the observation with the SM is the following. First, we fit the data using a model where all the NPs are free and $f_{0\mathrm{X}}$ and $\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}}$ are fixed at their SM values. This is done to ensure that the NPs reflect the data. This follows the recommendations from [433] to reduce under-coverage. We then need to estimate the probability of the hypothesis: *if the SM correctly describes our data and the NP values match those obtained from the fit, what is the probability that we would observe* $f_{0\mathrm{X}} \leq 0.007$*?* For that, we create signal-plus-background toy models where S, the signal, is the SM ($f_{0\mathrm{X}}$ as in the SM and $\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}} = 1$) and the NP values are those obtained from the previous fit. We calculate the fraction of toy models that have a fitted value of $f_{0\mathrm{X}}$ lower than the one we have in the data, $f_{0\mathrm{X}} \leq 0.007$. The probability obtained, under the SM hypothesis ($f_{0\mathrm{X}}$ at its SM value and $\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}} = 1$, with the NPs at their fitted values), is $0.6\,\%$, corresponding to a one-sided deviation of 2.5σ from the SM.

Another way of estimating the compatibility of the observed $f_{0\mathrm{X}}$ value with the SM is to use the value of the log-likelihood (LL) scan of the combined $f_{0\mathrm{X}}$ fit in Figure S.2(a) for $f^{\mathrm{SM}}_{0\mathrm{X}} = 0.28$. It is $-\Delta \ln L = 3.9$, corresponding to a significance of 2.8σ.

In conclusion, the measurement of $f_{0\mathrm{X}} = 0.007$ is likely to be due to a statistical fluctuation of the data.

7.3.13 Cross-checks

A series of cross-checks was performed to validate the results of the polarisation measurement. The main cross-checks include:

- Injection of different polarisation fractions in the MC samples to verify that the fit is able to recover the injected values. The results of these tests are shown in Appendix T.
- Decorrelation of the QCD scale uncertainties for the WZjj-QCD samples to quantify the effect on the normalisation of this process per run. The results of this test are shown in Appendix U.

- Change of the fit parametrisation to use NFs instead of the $\mu^{\text{fid.}}_{\text{WZjj-EW}}$ and the polarisation fraction to test if the results are stable under a different parametrisation. The results of this test are shown in Appendix V.
- Removing the object systematics from the fit to quantify their effect on the polarisation fractions. The results of this test are shown in Appendix W.
- Measurement of the doubly transverse polarisation fraction $\mathrm{W_TZ_T}$jj-EW to verify the excess of events in the $\mathrm{W_TZ_T}$jj-EW configuration. The results of this test are shown in Appendix X.
- The impact of letting the mixed polarisation states $\mathrm{W_0Z_T}$jj-EW and $\mathrm{W_TZ_0}$jj-EW float more freely instead of fixing them to the SM values. The results of this test are shown in Appendix Y.

All these tests show that the results of the polarisation measurement are robust against variations of the fit setup and the systematic uncertainties. The excess of events in the $\mathrm{W_TZ_T}$jj-EW configuration is reflected in the measurement of the $\mathrm{W_TZ_T}$jj-EW fraction, which is consistent with the picture given by the three fits presented here, although the four fits use the same data and templates and are not independent tests.

7.4 Summary

This chapter presented the first measurement of longitudinal-polarisation fractions in WZjj-EW production, performed with a simultaneous binned maximum-likelihood fit of the unrolled Voronoi templates in the SR together with the ZZ- and b-CRs, combining the full Run-2 and the partial Run-3 datasets. Three separate fits target the jointly longitudinal (f_{00}) and the two singly longitudinal ($f_{0\mathrm{X}}$, $f_{\mathrm{X}0}$) configurations, with the fiducial signal strength and the polarisation fraction as the only free parameters of interest and a full treatment of experimental and theoretical systematic uncertainties, including a dedicated NLO EW correction derived template-by-template with MoCaNLO.

The fit is stable and statistically limited in all three configurations. The main result is 4.0σ evidence for a longitudinally polarised Z boson in VBS events, $f_{\mathrm{X}0} = 0.324^{+0.092}_{-0.085}$, together with the first upper limits on the jointly longitudinal and W-only longitudinal states, a fiducial signal strength compatible with the SM, and the first measurement of the fiducial WZjj-EW cross-section at $\sqrt{s} = 13.6\,\mathrm{TeV}$, $\sigma^{\text{fid.},13.6\,\mathrm{TeV}}_{\mathrm{WZjj\text{-}EW}\to\ell'\nu_\ell\ell\ell\mathrm{jj}} = 0.374 \pm 0.094\,\mathrm{fb}$ (per lepton flavour, where ℓ and ℓ' are each either an electron or a muon), in agreement with the SM.

These results can be set against the other polarisation measurements in diboson final states. In same-sign WWjj-EW production, where the two neutrinos leave no diboson rest frame and the helicity states are defined in the WW CM frame, CMS measured $\sigma(\mathrm{W_0W_X}\text{jj-EW}) = 1.20^{+0.56}_{-0.53}\,\mathrm{fb}$ with an observed (expected) significance of 2.3σ (3.1σ) and set the limit $\sigma(\mathrm{W_0W_0}\text{jj-EW}) < 1.17\,\mathrm{fb}$ [352], and ATLAS measured $\sigma(\mathrm{W_0W_X}\text{jj-EW}) = 0.88 \pm 0.30\,\mathrm{fb}$ at 3.3σ (4.0σ) with $\sigma(\mathrm{W_0W_0}\text{jj-EW}) < 0.45\,\mathrm{fb}$ [354]. In inclusive WZ production, whose yield is far larger in the absence of the two-jet requirement, the jointly longitudinal fraction $f_{00} = 0.067 \pm 0.010$ was measured at 7.1σ [11]. These fractions and cross-sections are not directly comparable across channels or reference frames. States with a single longitudinal boson in VBS are now seen at 2.3σ (CMS, same-sign WWjj-EW), 3.3σ (ATLAS, same-sign WWjj-EW) and 4.0σ in this measurement, the first evidence for a longitudinal Z boson in the WZjj channel, while the doubly longitudinal state remains beyond the reach of the Run-2 and partial Run-3 statistics.

These measurements establish the analysis strategy and provide the baseline for the more precise polarisation measurements expected with the full Run-3 and HL-LHC datasets, where the extended tracking coverage discussed in Section 5.7 will further sharpen the WZjj topology.



Conclusion

This thesis has presented a comprehensive study of the VBS process, from its production via pp collisions at the LHC to the final-state detection and reconstruction by the ATLAS experiment. The theoretical and experimental aspects of this interesting process have been discussed in detail, highlighting the importance of VBS as a probe of physics beyond the Standard Model. The longitudinal polarisation state unique to the vector bosons W and Z presents unprecedented properties that can be exploited to test the SM prior to the SSB mechanism.

The previous observation of these rare processes as well as the polarisation measurement in inclusive WZ production has paved the way for the presented work. Building on these results, this thesis has presented a study on the WZjj channel to first disentangle the signal from the background and then distinguish the polarisation states. Three separate measurements were carried out to measure the polarisation fractions of the W and Z bosons in the fully leptonic decay channel. In the first, both bosons are longitudinally polarised, while in the other two only one boson is longitudinally polarised regardless of the other boson's polarisation state, all at fiducial level.

All measurements have been performed using a multivariate approach based on ML techniques. DNNs have been trained to discriminate the different polarisation states and the VBS signal from the background. The resulting output distributions have been used to form a plane on which a clustering algorithm has been applied to form a Voronoi binning. The resulting templates have been fitted to data to extract the polarisation fractions.

The dataset used corresponds to the combination of the full Run-2 dataset collected by the ATLAS experiment at $\sqrt{s} = 13\,\mathrm{TeV}$ during the years 2015–2018 representing a total integrated luminosity of $140\,\mathrm{fb}^{-1}$, and the first period of Run-3 data collected during the years 2022–2024 at $\sqrt{s} = 13.6\,\mathrm{TeV}$ representing a total integrated luminosity of $164\,\mathrm{fb}^{-1}$.

The fraction of polarisation where both bosons are longitudinally polarised f_{00} has been found to be zero in data, $f_{00} = 0.000^{+0.053}_{-0.000}$, for an expected value of $f_{00} = 0.087^{+0.078}_{-0.069}$ corresponding to an expected significance of 1.3σ. An observed (expected) upper limit at $95\,\%$ CL has been set at $f_{00} < 0.12$ (0.21).

The fraction of polarisation where the W boson is longitudinally polarised regardless of the polarisation of the Z boson, $f_{0\mathrm{X}}$, has been measured to be almost zero in data with $f_{0\mathrm{X}} = 0.007^{+0.094}_{-0.007}$ for an expected value of $f_{0\mathrm{X}} = 0.279^{+0.117}_{-0.109}$ corresponding to an expected significance of 2.6σ. An observed (expected) upper limit at $95\,\%$ CL has been set at $f_{0\mathrm{X}} < 0.19$ (0.46). This represents a deviation of 2.5σ from the SM prediction. Both observed limits are tighter than the expected ones, which is the direct consequence of the downward fluctuation of the two fractions towards the physical boundary $f = 0$ and not of an improved sensitivity.

The fraction of polarisation where the Z boson is longitudinally polarised regardless of the polarisation of the W boson, $f_{\mathrm{X}0}$, has been measured to be $f_{\mathrm{X}0} = 0.324^{+0.092}_{-0.085}$ corresponding to an observed (expected) significance of 4.0σ (2.9σ). This represents the first evidence of the longitudinal polarisation of the Z boson in VBS production and in particular in the WZjj channel. The measurement presented in this thesis is published by the ATLAS Collaboration in [432]. The measured value is slightly higher than the SM prediction, $f_{\mathrm{X}0} = 0.2758$ from MadGraph, 0.274 from Sherpa, and 0.2745 from the NLO EW calculation of MoCaNLO.

This increase of the longitudinal polarisation fraction of the Z boson in VBS production is consistent with the vanishing longitudinal polarisation fraction of the W boson. An $f_{0\mathrm{X}}$ fraction close to zero implies an increase of the $\mathrm{W_TZ_0jj}$-EW and $\mathrm{W_TZ_Tjj}$-EW components, and since $\mathrm{W_XZ_0jj}$-EW events are predominantly composed of $\mathrm{W_TZ_0jj}$-EW events, together with the $\mathrm{W_0Z_0jj}$-EW ones, an increased $f_{\mathrm{X}0}$

fraction is compatible with this observation, the single-longitudinal fits fixing the relative composition of the non-targeted states so that their fractions are not additive.

The measurement of the polarisation fractions included, among all systematic uncertainties, the uncertainty on the NLO EW corrections not taken into account in the SHERPA and MADGRAPH generators. This was computed from a state-of-the-art fixed-order calculation provided by the MOCANLO framework. The impact of this uncertainty on the polarisation fractions is found to be non-negligible, being the top-ranked NP for the f_{X0} measurement. However, the effect is not reflected in the deviation of the measured values from the SM predictions for the three fits, the prediction of MOCANLO being similar to those of the MC generators.

The signal strength of the VBS WZjj process has been measured in the combined fit to be $\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}} = 1.06 \pm 0.18$, in good agreement with the SM prediction of $\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}} = 0.94$ for MADGRAPH and $\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}} = 1$ for SHERPA. The standalone fit of a single data-taking period converts the signal strength into a fiducial cross-section per lepton flavour, where ℓ and ℓ' are each either an electron or a muon, in particular for the Run-3 dataset where it is reported for the first time:

$$\sigma^{\mathrm{fid.,13.6\,TeV}}_{\mathrm{WZjj\text{-}EW}\to\ell'\nu_\ell\ell\ell\mathrm{jj}} = 0.374 \pm 0.077\,(\mathrm{stat.}) \pm 0.053\,(\mathrm{syst.}) \pm 0.009\,(\mathrm{lumi.})\,\mathrm{fb}\,,$$

for the fully leptonic decay channel, without considering τ, and in good agreement with the SHERPA prediction of $0.373^{+0.065}_{-0.041}\,\mathrm{fb}$. The Run-2 fit gives:

$$\sigma^{\mathrm{fid.,13\,TeV}}_{\mathrm{WZjj\text{-}EW}\to\ell'\nu_\ell\ell\ell\mathrm{jj}} = 0.347 \pm 0.068\,(\mathrm{stat.}) \pm 0.039\,(\mathrm{syst.}) \pm 0.003\,(\mathrm{lumi.})\,\mathrm{fb}\,,$$

fully compatible with the previous measurement [360].

These measurements are the first of their kind in this channel and represent a significant step forward in our understanding of the EW sector of the SM. The results provide valuable insights into the nature of EW SSB and the role of longitudinally polarised vector bosons in high-energy collisions. The measured polarisation fractions, in particular the one where the W boson is longitudinally polarised, remain compatible with the SM predictions, and discriminating a statistical fluctuation from a real effect will require more data, first with the full Run-3 dataset and then with the future HL-LHC [12] dataset. Indeed, these measurements are statistically limited and the uncertainty on the polarisation fractions is dominated by the amount of data collected.

This new era for the LHC will provide unprecedented opportunities to probe the EW sector and will multiply the size of the dataset by a factor of ten, allowing for more precise measurements of the polarisation fractions. ATLAS will upgrade its tracker with the implementation of the ITk [13, 14] extending the track coverage to $|\eta| < 4.0$ and improving the track reconstruction efficiency. This improvement in the forward region is of particular interest for such measurements, as it will add leverage to the reconstruction of forward jets, pile-up rejection, and the identification of electrons in the forward region.

In this context, this thesis has also presented a first calibration and identification of electrons in the forward region with the ITk. This study showed a significant improvement in the background rejection at fixed identification efficiency in the forward region over the methods currently in place. This was achieved by exploiting the new tracking capabilities of the ITk and by using ML techniques. A prior p_{T}-calibration was done to correct the electron energy in the forward region, and a new DNN-based identification was developed with a prior p_{T}-decorrelation to have WPs with a fixed efficiency across the p_{T} spectrum. Pile-up cross-checks were carried out to ensure the stability of the identification performance in high-pile-up conditions. This work resulted in the implementation of this workflow into the ATLAS reconstruction software ATHENA and the development of a new ITk-based electron identification algorithm for the forward region, which will be available for preliminary Run-4 analyses.

The hardware upgrades of the ATLAS detector, the increased luminosity, and the improved ML techniques for identification and reconstruction of the final-state objects will further sharpen these measurements in the years to come. The results presented in this thesis thus represent another major milestone towards understanding the EW sector of the SM and the search for new physics beyond it.

Appendices

Overview of the Appendices

The appendices gather material that supports the main text but would otherwise interrupt its narrative: analytic derivations, technical cross-checks, and additional distributions. They are organised in five thematic blocks.

Theoretical framework. Appendices A to D: the analytic VBS cross-section, the polarisation formalism for the joint WZ system, the associated Feynman diagrams, and the one-loop EW correction used as a systematic uncertainty.

Background and simulated samples. Appendices E and F: the data-driven estimation of the misidentified-lepton background and the full list of Monte Carlo samples used in the analysis.

Multivariate polarisation observable. Appendices G to N: construction of the 3D template, BO of the DNN hyperparameters, the input-variable distributions, their correlation matrices, variable ranking, generator effects, and the Voronoi binning of the template plane.

Statistical analysis. Appendices O to Y: the fit methodology, treatment, and pruning of systematic uncertainties, post-fit nuisance-parameter pulls, profile-likelihood scans, injection tests, and cross-checks of the polarisation fits under alternative assumptions.

Beyond the Standard Model interpretation. Appendix Z: exploration of the parameter space of scotogenic dark matter models using an active learning strategy built on the DNN toolkit developed for this measurement.

A Vector boson scattering cross-section calculation

Let us consider the VBS processes $V_0^a(p_a)V_0^b(p_b) \to V_0^c(p_c)V_0^d(p_d)$, where V_0 represents a weak vector boson longitudinally polarised, with associated four-momenta p_a to p_d.

If we consider the same notation as in Section 1.11, the longitudinal polarisation vector of a massive vector boson with momentum along the z-axis can be written as:

$$\varepsilon_0^\mu(p) = \frac{1}{m_V}(p_z, 0, 0, E)\,, \tag{A.1}$$

where we have a clear dependence on the energy E of the vector boson. This dependence makes the longitudinal polarisation vector grow with the energy of the boson, which in turn dominates the scattering amplitude at high energy. The boson in this polarisation state can therefore be seen as the Goldstone boson that was absorbed by the vector boson when becoming massive via the Higgs mechanism.

The GBET states that, in the high-energy limit ($E \gg m_V$), the scattering amplitude involving longitudinally polarised vector bosons can be approximated by the scattering amplitude of the corresponding Goldstone bosons [3, 4]. This equivalence is particularly useful because calculations involving scalar particles (the Goldstone bosons) are often simpler than those involving vector bosons. In this limit, the longitudinal polarisation vector can be approximated as:

$$\varepsilon_0^\mu(p) = \frac{p^\mu}{m_V} + \mathcal{O}\left(\frac{m_V}{E}\right)\,, \tag{A.2}$$

where the subleading terms are suppressed by powers of $\frac{m_V}{E}$.

This implies at the amplitude level:

$$\mathcal{M}\left(V_0^a V_0^b \to V_0^c V_0^d\right) = \mathcal{M}\left(\phi^a\phi^b \to \phi^c\phi^d\right) + \mathcal{O}\left(\frac{m_V}{E}\right)\,, \tag{A.3}$$

where ϕ^a denotes the Goldstone boson associated with the vector field V^a.

In Section 1.6, we have introduced the scalar doublet ϕ of the GWS theory, denoted Φ in the following to distinguish it from its Goldstone components ϕ_i:

$$\Phi = \frac{1}{\sqrt{2}}\begin{pmatrix}\phi_1 + \mathrm{i}\phi_2 \\ v + h + \mathrm{i}\phi_3\end{pmatrix}\,, \tag{A.4}$$

where ϕ_1, ϕ_2, and ϕ_3 are the three Goldstone bosons and h is the physical Higgs boson field. The Goldstone bosons associated with the W and Z bosons can be identified at tree level as:

$$\mathrm{w}^\pm = \frac{1}{\sqrt{2}}(\phi_1 \pm \mathrm{i}\phi_2)\,, \tag{A.5}$$

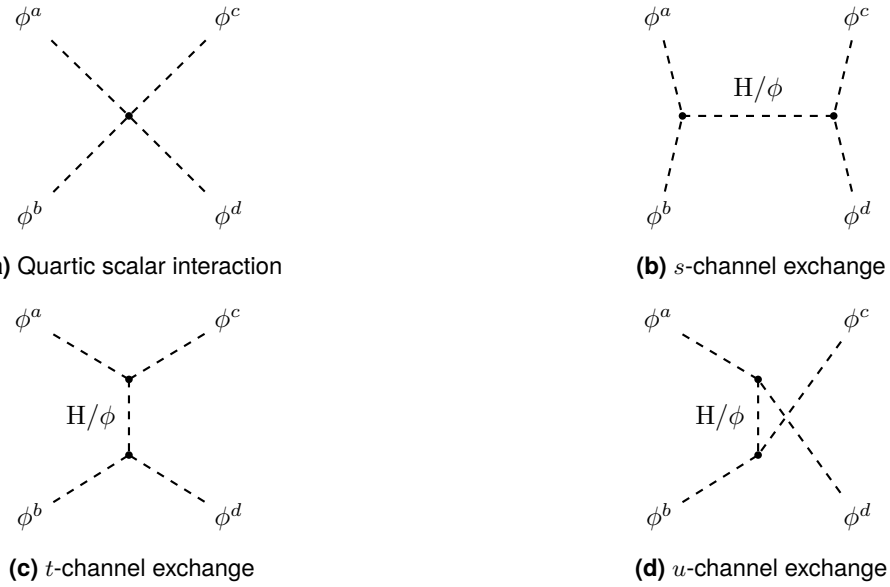
$$\mathrm{z}^0 = \phi_3\,. \tag{A.6}$$

The field w^+ is therefore the charged component ϕ^+ of the doublet introduced in Section 1.6, while its neutral component reads $\phi^0 = \frac{1}{\sqrt{2}}(v + h + \mathrm{i}\mathrm{z}^0)$, such that Φ and $\Phi^\dagger$ can be written as:

$$\Phi = \frac{1}{\sqrt{2}}\begin{pmatrix}\sqrt{2}\mathrm{w}^+ \\ v + h + \mathrm{i}\mathrm{z}^0\end{pmatrix}\,, \tag{A.7}$$

$$\Phi^\dagger = \frac{1}{\sqrt{2}}\begin{pmatrix}\sqrt{2}\mathrm{w}^- & v + h - \mathrm{i}\mathrm{z}^0\end{pmatrix}\,. \tag{A.8}$$

As a $2 \to 2$ scattering process, we have contact interactions via quartic vertices as well as s-, t-, and u-channel exchange of gauge bosons and Higgs bosons. The relevant Feynman diagrams are shown in Figure A.1.



(a) Quartic scalar interaction

(b) s-channel exchange

(c) t-channel exchange

(d) u-channel exchange

Figure A.1: Feynman diagrams contributing to the VBS process via Goldstone boson scattering.

In this context, we can define the Mandelstam variables as:

$$s = (p_a + p_b)^2 = (p_c + p_d)^2 \,, \tag{A.9}$$

$$t = (p_a - p_c)^2 = (p_b - p_d)^2 \,, \tag{A.10}$$

$$u = (p_a - p_d)^2 = (p_b - p_c)^2 \,. \tag{A.11}$$

They satisfy the relation:

$$s + t + u = \sum_{i=1}^{4} m_i^2 \,, \tag{A.12}$$

where m_i are the masses of the particles involved in the scattering process.

We can parametrise the ME $\mathcal{M}_{fi}$ as:

$$\mathcal{M}(\phi^a \phi^b \to \phi^c \phi^d) = A(s)\delta_{ab}\delta_{cd} + A(t)\delta_{ac}\delta_{bd} + A(u)\delta_{ad}\delta_{bc} \,, \tag{A.13}$$

where A is the following function:

$$A(x) = \frac{x}{v^2}\left(1 - \frac{x}{x - m_{\mathrm{H}}^2}\right) = \frac{-m_{\mathrm{H}}^2}{v^2}\frac{x}{x - m_{\mathrm{H}}^2} \,, \tag{A.14}$$

with v the VEV introduced in Section 1.6 and m_{H} the mass of the Higgs boson. The first term in the parentheses corresponds to the contribution from the quartic interaction, while the second term corresponds to the contributions from the Higgs boson exchange in the s-, t-, and u-channels. The latter corresponds to the Higgs propagator.

The Higgs boson plays a crucial role in ensuring the unitarity of the scattering amplitude at high energies [4]. Without the Higgs boson contribution, the amplitude would grow with energy, leading to a violation of unitarity. The inclusion of the Higgs boson exchange terms cancels the energy-growing terms, thus preserving unitarity in accordance with the GBET.

We will highlight this crucial role of the Higgs boson in the unitarity of the VBS process by considering the A defined above in two scenarios: with and without the Higgs boson contribution. Without the Higgs boson contribution, the function $A(x)$ simplifies to:

$$A(x) = \frac{x}{v^2} \,. \tag{A.15}$$

In this case, the scattering amplitude grows linearly with the Mandelstam variable x, leading to a violation of unitarity at high energies.

The possible combinations for a VBS process are:

- $\mathrm{W}^+\mathrm{W}^- \to \mathrm{W}^+\mathrm{W}^-$
- $\mathrm{W}^\pm\mathrm{W}^\pm \to \mathrm{W}^\pm\mathrm{W}^\pm$
- $\mathrm{W}^\pm\mathrm{Z} \to \mathrm{W}^\pm\mathrm{Z}$
- $\mathrm{ZZ} \to \mathrm{ZZ}$
- $\mathrm{W}^\pm\mathrm{W}^\mp \to \mathrm{ZZ}$

They all have the quartic interaction. The Higgs boson exchange contribution depends on the specific process.

In the CM frame, we can express the four-momenta of the incoming and outgoing vector bosons as:

$$p_a = (E, 0, 0, |\vec{p}|)\,, \tag{A.16}$$
$$p_b = (E, 0, 0, -|\vec{p}|)\,. \tag{A.17}$$

The s variable is therefore the total energy squared in the CM frame:

$$s = (p_a + p_b)^2 = 4E^2\,. \tag{A.18}$$

The outgoing four-momenta can be expressed as:

$$p_c = (E, |\vec{p}|\sin\theta, 0, |\vec{p}|\cos\theta)\,, \tag{A.19}$$
$$p_d = (E, -|\vec{p}|\sin\theta, 0, -|\vec{p}|\cos\theta)\,, \tag{A.20}$$

where θ is the scattering angle between the incoming and outgoing vector bosons.

We can now express the t and u variables in terms of the scattering angle θ:

$$t = (p_a - p_c)^2 = 2m_V^2 - 2\left(E^2 - |\vec{p}|^2\cos\theta\right)\,, \tag{A.21}$$
$$u = (p_a - p_d)^2 = 2m_V^2 - 2\left(E^2 + |\vec{p}|^2\cos\theta\right)\,, \tag{A.22}$$

where m_V is the mass of the vector boson. The t variable can be interpreted as the squared four-momentum transfer between the incoming and outgoing vector bosons, while the u variable represents the squared four-momentum transfer between the incoming vector boson and the other outgoing vector boson, known as the crossing channel.

In the ultra-relativistic limit, thus in the GBET context, where the energy of the vector bosons is much greater than their mass ($E \gg m_V$), we can approximate $|\vec{p}| \simeq E$. This leads to simplified expressions for the Mandelstam variables:

$$s \gg m_V^2\,, \tag{A.23}$$
$$t \simeq -\frac{s}{2}(1 - \cos\theta)\,, \tag{A.24}$$
$$u \simeq -\frac{s}{2}(1 + \cos\theta)\,. \tag{A.25}$$

We recall that the cross-section for a $2 \to 2$ scattering process is given in Equation 1.17. For a $2 \to 2$ scattering process, the differential cross-section in the CM frame can be expressed as:

$$\frac{\mathrm{d}\,\sigma_{a+b\to c+d}}{\mathrm{d}\,\Omega} = \frac{1}{64\pi^2 s}\frac{|\vec{p}_f|}{|\vec{p}_i|}|\mathcal{M}_{fi}|^2\,, \tag{A.26}$$
$$\sigma_{a+b\to c+d} = \frac{1}{2}\int_{-1}^{1}\frac{1}{32\pi s}|\mathcal{M}_{fi}|^2\mathrm{d}\,(\cos\theta)\,, \tag{A.27}$$

where $\mathrm{d}\,\Omega$ is the differential solid angle, s is the Mandelstam variable representing the square of the total energy in the CM frame, $\vec{p}_i$ and $\vec{p}_f$ are the momenta of the incoming and outgoing particles in the CM frame, respectively, and $\mathcal{M}_{fi}$ is the ME for the transition from the initial to the final state. The term $\frac{|\vec{p}_f|}{|\vec{p}_i|}$ is equal to 1 in the ultra-relativistic limit where the masses of the particles are negligible compared to their energies, valid in the GBET context. The factor of $\frac{1}{2}$ in the total cross-section accounts for when there are identical particles in the final state, preventing double counting.

A.1 $\mathrm{W^+W^-} \to \mathrm{W^+W^-}$

For this process, the contribution arises from the s- and t-channels. The total ME is given by:

$$\mathcal{M}_{\mathrm{W^+W^-}\to\mathrm{W^+W^-}} = A(s) + A(t)\,. \tag{A.28}$$

This leads to the following expression for the scattering amplitude:

$$\mathcal{M}_{\mathrm{W^+W^-}\to\mathrm{W^+W^-}} = \frac{-m_\mathrm{H}^2}{v^2}\left(\frac{s}{s-m_\mathrm{H}^2} + \frac{t}{t-m_\mathrm{H}^2}\right), \tag{A.29}$$

$$\mathcal{M}^{\text{no Higgs}}_{\mathrm{W^+W^-}\to\mathrm{W^+W^-}} = \frac{s+t}{v^2}\,. \tag{A.30}$$

Using Equation A.26, we can compute the differential cross-section for this process both with and without the Higgs boson contribution:

$$\sigma_{\mathrm{W^+W^-}\to\mathrm{W^+W^-}} = \int_{-1}^{1} \frac{1}{32\pi s}\left|\frac{-m_\mathrm{H}^2}{v^2}\left(\frac{s}{s-m_\mathrm{H}^2} + \frac{t}{t-m_\mathrm{H}^2}\right)\right|^2 \mathrm{d}\left(\cos\theta\right), \tag{A.31}$$

$$\sigma^{\text{no Higgs}}_{\mathrm{W^+W^-}\to\mathrm{W^+W^-}} = \int_{-1}^{1} \frac{1}{32\pi s}\left|\frac{s+t}{v^2}\right|^2 \mathrm{d}\left(\cos\theta\right), \tag{A.32}$$

which, after integration, gives:

$$\sigma_{\mathrm{W^+W^-}\to\mathrm{W^+W^-}} = \frac{m_\mathrm{H}^4}{16\pi v^4 s}\left[\left(\frac{2s-m_\mathrm{H}^2}{s-m_\mathrm{H}^2}\right)^2 + \frac{2m_\mathrm{H}^2}{s}\frac{2s-m_\mathrm{H}^2}{s-m_\mathrm{H}^2}\ln\left(\frac{m_\mathrm{H}^2}{s+m_\mathrm{H}^2}\right) + \frac{m_\mathrm{H}^2}{s+m_\mathrm{H}^2}\right], \tag{A.33}$$

$$\sigma^{\text{no Higgs}}_{\mathrm{W^+W^-}\to\mathrm{W^+W^-}} = \frac{s}{48\pi v^4}\,. \tag{A.34}$$

The result shows that without the Higgs boson contribution, the cross-section grows linearly with the CM energy squared s, leading to a violation of unitarity at high energies. In contrast, the inclusion of the Higgs boson exchange terms ensures that the cross-section remains finite and well-behaved at high energies, thus preserving unitarity in accordance with the GBET.

A.2 $\mathrm{W^\pm W^\pm} \to \mathrm{W^\pm W^\pm}$

For this process, the contribution arises from the t- and u-channels. The total ME is given by:

$$\mathcal{M}_{\mathrm{W^\pm W^\pm}\to\mathrm{W^\pm W^\pm}} = A(t) + A(u)\,. \tag{A.35}$$

This leads to the following expression for the scattering amplitude:

$$\mathcal{M}_{\mathrm{W^\pm W^\pm}\to\mathrm{W^\pm W^\pm}} = \frac{-m_\mathrm{H}^2}{v^2}\left(\frac{t}{t-m_\mathrm{H}^2} + \frac{u}{u-m_\mathrm{H}^2}\right), \tag{A.36}$$

$$\mathcal{M}^{\text{no Higgs}}_{\mathrm{W^\pm W^\pm}\to\mathrm{W^\pm W^\pm}} = \frac{t+u}{v^2} = \frac{-s}{v^2}\,. \tag{A.37}$$

Using Equation A.26, we can compute the differential cross-section for this process both with and without the Higgs boson contribution:

$$\sigma_{\mathrm{W^\pm W^\pm}\to\mathrm{W^\pm W^\pm}} = \frac{1}{2}\int_{-1}^{1} \frac{1}{32\pi s}\left|\frac{-m_\mathrm{H}^2}{v^2}\left(\frac{t}{t-m_\mathrm{H}^2} + \frac{u}{u-m_\mathrm{H}^2}\right)\right|^2 \mathrm{d}\left(\cos\theta\right), \tag{A.38}$$

$$\sigma^{\text{no Higgs}}_{\mathrm{W^\pm W^\pm}\to\mathrm{W^\pm W^\pm}} = \frac{1}{2}\int_{-1}^{1} \frac{1}{32\pi s}\left|\frac{t+u}{v^2}\right|^2 \mathrm{d}\left(\cos\theta\right), \tag{A.39}$$

which, after integration, gives:

$$\sigma_{\mathrm{W^\pm W^\pm}\to\mathrm{W^\pm W^\pm}} = \frac{m_\mathrm{H}^4}{16\pi v^4 s}\left[\frac{2s+3m_\mathrm{H}^2}{s+m_\mathrm{H}^2} + \frac{2m_\mathrm{H}^2}{s}\frac{2s+3m_\mathrm{H}^2}{s+2m_\mathrm{H}^2}\ln\left(\frac{m_\mathrm{H}^2}{s+m_\mathrm{H}^2}\right)\right], \tag{A.40}$$

$$\sigma^{\text{no Higgs}}_{\mathrm{W^\pm W^\pm}\to\mathrm{W^\pm W^\pm}} = \frac{s}{32\pi v^4}\,. \tag{A.41}$$

The result shows again that without the Higgs boson contribution, the cross-section grows linearly with the CM energy squared s, leading to a violation of unitarity at high energies. In contrast, the inclusion of the Higgs boson exchange terms ensures that the cross-section remains finite and well-behaved at high energies, thus preserving unitarity in accordance with the GBET.

A.3 $\mathrm{W^{\pm}Z \to W^{\pm}Z}$

For this process, the contribution arises only from the t-channel. The total ME is given by:

$$\mathcal{M}_{\mathrm{W^{\pm}Z \to W^{\pm}Z}} = A(t)\,. \tag{A.42}$$

This leads to the following expression for the scattering amplitude:

$$\mathcal{M}_{\mathrm{W^{\pm}Z \to W^{\pm}Z}} = \frac{-m_{\mathrm{H}}^2}{v^2}\left(\frac{t}{t-m_{\mathrm{H}}^2}\right)\,, \tag{A.43}$$

$$\mathcal{M}^{\text{no Higgs}}_{\mathrm{W^{\pm}Z \to W^{\pm}Z}} = \frac{t}{v^2}\,. \tag{A.44}$$

Using Equation A.26, we can compute the differential cross-section for this process both with and without the Higgs boson contribution:

$$\sigma_{\mathrm{W^{\pm}Z \to W^{\pm}Z}} = \int_{-1}^{1} \frac{1}{32\pi s}\left|\frac{-m_{\mathrm{H}}^2}{v^2}\left(\frac{t}{t-m_{\mathrm{H}}^2}\right)\right|^2 \mathrm{d}\left(\cos\theta\right)\,, \tag{A.45}$$

$$\sigma^{\text{no Higgs}}_{\mathrm{W^{\pm}Z \to W^{\pm}Z}} = \int_{-1}^{1} \frac{1}{32\pi s}\left|\frac{t}{v^2}\right|^2 \mathrm{d}\left(\cos\theta\right)\,, \tag{A.46}$$

which, after integration, gives:

$$\sigma_{\mathrm{W^{\pm}Z \to W^{\pm}Z}} = \frac{m_{\mathrm{H}}^4}{16\pi v^4 s}\left[\frac{s+2m_{\mathrm{H}}^2}{s+m_{\mathrm{H}}^2} + \frac{2m_{\mathrm{H}}^2}{s}\ln\left(\frac{m_{\mathrm{H}}^2}{s+m_{\mathrm{H}}^2}\right)\right]\,, \tag{A.47}$$

$$\sigma^{\text{no Higgs}}_{\mathrm{W^{\pm}Z \to W^{\pm}Z}} = \frac{s}{48\pi v^4}\,. \tag{A.48}$$

The result shows again that without the Higgs boson contribution, the cross-section grows linearly with the CM energy squared s, leading to a violation of unitarity at high energies. In contrast, the inclusion of the Higgs boson exchange terms ensures that the cross-section remains finite and well-behaved at high energies, thus preserving unitarity in accordance with the GBET.

A.4 $\mathrm{ZZ \to ZZ}$

For this process, the contribution arises from the s-, t-, and u-channels. The total ME is given by:

$$\mathcal{M}_{\mathrm{ZZ \to ZZ}} = A(s) + A(t) + A(u)\,. \tag{A.49}$$

This leads to the following expression for the scattering amplitude:

$$\mathcal{M}_{\mathrm{ZZ \to ZZ}} = \frac{-m_{\mathrm{H}}^2}{v^2}\left(\frac{s}{s-m_{\mathrm{H}}^2} + \frac{t}{t-m_{\mathrm{H}}^2} + \frac{u}{u-m_{\mathrm{H}}^2}\right)\,, \tag{A.50}$$

$$\mathcal{M}^{\text{no Higgs}}_{\mathrm{ZZ \to ZZ}} = \frac{s+t+u}{v^2} = 0\,. \tag{A.51}$$

Using Equation A.26, we can compute the differential cross-section for this process both with and without the Higgs boson contribution:

$$\sigma_{\mathrm{ZZ \to ZZ}} = \frac{1}{2}\int_{-1}^{1} \frac{1}{32\pi s}\left|\frac{-m_{\mathrm{H}}^2}{v^2}\left(\frac{s}{s-m_{\mathrm{H}}^2} + \frac{t}{t-m_{\mathrm{H}}^2} + \frac{u}{u-m_{\mathrm{H}}^2}\right)\right|^2 \mathrm{d}\left(\cos\theta\right)\,, \tag{A.52}$$

$$\sigma^{\text{no Higgs}}_{\mathrm{ZZ \to ZZ}} = \frac{1}{2}\int_{-1}^{1} \frac{1}{32\pi s}\left|\frac{s+t+u}{v^2}\right|^2 \mathrm{d}\left(\cos\theta\right)\,, \tag{A.53}$$

which, after integration, gives:

$$\sigma_{\mathrm{ZZ \to ZZ}} = \frac{m_{\mathrm{H}}^4}{32\pi v^4 s}\left[\left(\frac{3s-2m_{\mathrm{H}}^2}{s-m_{\mathrm{H}}^2}\right)^2 + \frac{2m_{\mathrm{H}}^2}{s+m_{\mathrm{H}}^2} + \frac{12m_{\mathrm{H}}^2\left(s^2+sm_{\mathrm{H}}^2-m_{\mathrm{H}}^4\right)}{s\left(s-m_{\mathrm{H}}^2\right)\left(s+2m_{\mathrm{H}}^2\right)}\ln\left(\frac{m_{\mathrm{H}}^2}{s+m_{\mathrm{H}}^2}\right)\right]\,, \tag{A.54}$$

$$\sigma^{\text{no Higgs}}_{\mathrm{ZZ \to ZZ}} = 0\,. \tag{A.55}$$

The absence of the Higgs boson contribution in this process leads to a vanishing scattering amplitude and thus a zero cross-section. In this specific case, there is no violation of unitarity since the cross-section does not grow with energy. It simply remains zero without the Higgs boson contribution. In contrast, the inclusion of the Higgs boson exchange terms ensures that the cross-section remains finite and well-behaved at high energies, thus preserving unitarity in accordance with the GBET.

A.5 $\mathrm{W^{\pm}W^{\mp}} \to \mathrm{ZZ}$

For this process, the contribution arises from the s-channel only. The total ME is given by:

$$\mathcal{M}_{\mathrm{W^{\pm}W^{\mp}\to ZZ}} = A(s)\,. \tag{A.56}$$

This leads to the following expression for the scattering amplitude:

$$\mathcal{M}_{\mathrm{W^{\pm}W^{\mp}\to ZZ}} = \frac{-m_{\mathrm{H}}^2}{v^2}\left(\frac{s}{s-m_{\mathrm{H}}^2}\right), \tag{A.57}$$

$$\mathcal{M}^{\text{no Higgs}}_{\mathrm{W^{\pm}W^{\mp}\to ZZ}} = \frac{s}{v^2}\,. \tag{A.58}$$

Using Equation A.26, we can compute the differential cross-section for this process both with and without the Higgs boson contribution:

$$\sigma_{\mathrm{W^{\pm}W^{\mp}\to ZZ}} = \frac{1}{2}\int_{-1}^{1}\frac{1}{32\pi s}\left|\frac{-m_{\mathrm{H}}^2}{v^2}\left(\frac{s}{s-m_{\mathrm{H}}^2}\right)\right|^2 \mathrm{d}\left(\cos\theta\right), \tag{A.59}$$

$$\sigma^{\text{no Higgs}}_{\mathrm{W^{\pm}W^{\mp}\to ZZ}} = \frac{1}{2}\int_{-1}^{1}\frac{1}{32\pi s}\left|\frac{s}{v^2}\right|^2 \mathrm{d}\left(\cos\theta\right), \tag{A.60}$$

which, after integration, gives:

$$\sigma_{\mathrm{W^{\pm}W^{\mp}\to ZZ}} = \frac{m_{\mathrm{H}}^4}{32\pi v^4}\frac{s}{(s-m_{\mathrm{H}}^2)^2}\,, \tag{A.61}$$

$$\sigma^{\text{no Higgs}}_{\mathrm{W^{\pm}W^{\mp}\to ZZ}} = \frac{s}{32\pi v^4}\,. \tag{A.62}$$

The result shows again that without the Higgs boson contribution, the cross-section grows linearly with the CM energy squared s, leading to a violation of unitarity at high energies. In contrast, the inclusion of the Higgs boson exchange terms ensures that the cross-section remains finite and well-behaved at high energies, thus preserving unitarity in accordance with the GBET.

A.6 Higgs mass constraint from unitarity

Unitarity implies that $\mathcal{S}^{\dagger}\mathcal{S} = 1$, where $\mathcal{S}$ is defined in Equation 1.15. This condition ensures the conservation of probability in quantum mechanical processes. The transition matrix $\mathcal{T}$ is defined as the transition amplitude between initial and final states:

$$\langle f|\mathcal{T}|i\rangle = (2\pi)^4\delta^{(4)}(p_f - p_i)\mathcal{M}_{fi}\,, \tag{A.63}$$

consistent with Equation 1.16, which carries in addition the normalisation factors of the external states. We can thus rewrite the S-matrix in terms of the transition matrix $\mathcal{T}$ as $\mathcal{S} = 1 + \mathrm{i}\mathcal{T}$. The unitarity condition can then be rewritten as:

$$-\mathrm{i}(\mathcal{T} - \mathcal{T}^{\dagger}) = \mathcal{T}^{\dagger}\mathcal{T}\,. \tag{A.64}$$

This relation connects the imaginary part of the transition amplitude to the total probability of all possible transitions from the initial state to any final state.

If we consider an elastic scattering process where the initial and final states are the same, the unitarity condition leads to:

$$-\mathrm{i}\langle i|(\mathcal{T} - \mathcal{T}^{\dagger})|i\rangle = \sum_n \langle i|\mathcal{T}^{\dagger}|n\rangle\langle n|\mathcal{T}|i\rangle \quad \Rightarrow \quad 2\,\mathrm{Im}\,\langle i|\mathcal{T}|i\rangle = \sum_n |\langle n|\mathcal{T}|i\rangle|^2\,. \tag{A.65}$$

Following this, we have the cross-section derived from Equation 1.17:

$$\sigma_{\mathrm{tot}} = \sum_n \sigma_{i\to n} = \frac{1}{4k\sqrt{s}}\sum_n \int \mathrm{d}\,\Phi_n |\langle n|\mathcal{T}|i\rangle|^2\,, \tag{A.66}$$

where $\mathrm{d}\,\Phi_n$ is the n-body LIPS and k is the magnitude of the three-momentum where the flux F in Equation 1.18 is given by $F = 4k\sqrt{s}$ in the CM frame. We can write k as:

$$k = \frac{\lambda^{1/2}(s, m_a^2, m_b^2)}{2\sqrt{s}}\,, \tag{A.67}$$

where λ is the Källén function defined as:

$$\lambda(x,y,z) = x^2 + y^2 + z^2 - 2xy - 2xz - 2yz = \left[x - (\sqrt{y} + \sqrt{z})^2\right]\left[x - (\sqrt{y} - \sqrt{z})^2\right] . \tag{A.68}$$

Using Equation A.65, we can relate the total cross-section to the imaginary part of the forward scattering amplitude:

$$\sigma_{\text{tot}}(s) = \frac{1}{2k\sqrt{s}} \operatorname{Im} \mathcal{M}(s, t = 0) . \tag{A.69}$$

In our ultra-relativistic limit, we can approximate $k \simeq \frac{\sqrt{s}}{2}$, leading to:

$$\sigma_{\text{tot}} = \frac{1}{s} \operatorname{Im} \mathcal{M} . \tag{A.70}$$

This relation shows that the total cross-section is directly proportional to the imaginary part of the forward scattering amplitude, scaled by the inverse of the CM energy squared. It is known as the optical theorem.

At tree level, the amplitudes we consider have no imaginary part below particle thresholds, and perturbative unitarity demands that the amplitude not grow faster with s than allowed by the optical theorem.

A more systematic way to enforce unitarity is to expand the amplitude in partial waves. For a scattering amplitude that depends only on $\cos\theta$, one introduces the partial-wave expansion:

$$\mathcal{M}(s, \cos\theta) = 16\pi \sum_{\ell=0}^{\infty} (2\ell + 1)\, a_\ell(s)\, P_\ell(\cos\theta) , \tag{A.71}$$

where $P_\ell(\cos\theta)$ are Legendre polynomials and the partial-wave coefficients $a_\ell(s)$ are given by:

$$a_\ell(s) = \frac{1}{32\pi} \int_{-1}^{1} \mathrm{d}\,(\cos\theta)\, P_\ell(\cos\theta)\, \mathcal{M}(s, \cos\theta) . \tag{A.72}$$

The elements of the Wigner d-matrix of Equation 6.65 with both indices equal to zero are simply the Legendre polynomials, $d^\ell_{00}(\theta) = P_\ell(\cos\theta)$.

For an elastic $2 \to 2$ scattering process, the amplitude depends on s and t (or equivalently on s and $\cos\theta$). The forward limit corresponds to $t = 0$, that is $\cos\theta = 1$.

Elastic unitarity implies that each partial wave lies on the unitarity circle. In particular, a commonly used perturbative unitarity condition at tree level is:

$$|\operatorname{Re} a_\ell(s)| \leq \frac{1}{2} , \tag{A.73}$$

and especially for the s-wave ($\ell = 0$):

$$|\operatorname{Re} a_0(s)| \leq \frac{1}{2} . \tag{A.74}$$

To illustrate how this leads to an upper limit on the Higgs mass, we apply these ideas to the high-energy longitudinal VBS amplitudes using the GBET and the Goldstone parametrisation as done before. In the ultra-relativistic limit ($s \gg m_V^2, m_{\text{H}}^2$), the amplitude for longitudinal $V_0V_0 \to V_0V_0$ scattering is dominated by the Goldstone boson amplitude, and we can rewrite Equation A.13 as:

$$\mathcal{M}(\phi^a\phi^b \to \phi^c\phi^d) \simeq -\frac{m_{\text{H}}^2}{v^2} \left(\delta_{ab}\delta_{cd} + \delta_{ac}\delta_{bd} + \delta_{ad}\delta_{bc}\right) , \tag{A.75}$$

where we have used the high-energy limit of Equation A.14:

$$A(x) = \frac{-m_{\text{H}}^2}{v^2} \frac{x}{x - m_{\text{H}}^2} \xrightarrow{s \gg m_{\text{H}}^2} -\frac{m_{\text{H}}^2}{v^2} . \tag{A.76}$$

This is the manifestation of the cancellation of energy growth by Higgs exchange discussed previously. In this limit, the amplitude becomes approximately constant with respect to s, controlled by the Higgs boson mass and VEV.

The dominant contribution to the partial-wave amplitude in this energy regime comes from the s-wave:

$$a_0(s) = \frac{1}{32\pi} \int_{-1}^{1} \mathrm{d}\,(\cos\theta)\, \mathcal{M}(s, \cos\theta) \simeq -\frac{m_{\text{H}}^2}{16\pi v^2}\, c , \tag{A.77}$$

where c is a numerical factor depending on the specific channel combination considered. Demanding that Equation A.74 be satisfied in all such channels leads to an upper bound on $\frac{m_{\mathrm{H}}^2}{v^2}$. The most stringent usual bound is obtained from the isospin $I = 0$ ($c = 1$) combination of longitudinal weak bosons, giving:

$$|\mathrm{Re}\, a_0(s)| \simeq \frac{m_{\mathrm{H}}^2}{16\pi v^2} \leq \frac{1}{2}\,, \tag{A.78}$$

or:

$$m_{\mathrm{H}}^2 \leq 8\pi v^2 \quad \Longrightarrow \quad m_{\mathrm{H}} \leq \sqrt{8\pi} v \simeq 1.2\,\mathrm{TeV}\,, \tag{A.79}$$

where $v = 246\,\mathrm{GeV}$ is the VEV.

Therefore, the Higgs boson mass must be below approximately $1.2\,\mathrm{TeV}$ to ensure perturbative unitarity of longitudinal VBS amplitudes at high energies [4]. If the Higgs boson were heavier than this bound, the scattering amplitudes would grow too large, violating unitarity and indicating the breakdown of perturbation theory. This suggests that new physics or strong dynamics would be required to restore unitarity in such a scenario.

B Polarisation formalism for joint WZ system

B.1 Joint spin density matrix

To construct the full 9×9 joint SDM [388, 389, 392, 395] ρ^{WZ} in the helicity basis ($|{+1}\rangle, |0\rangle, |{-1}\rangle \otimes |{+1}\rangle, |0\rangle, |{-1}\rangle$) without expanding into an 81-element grid, we decompose Equation 6.90 into its seven constituent sums.

First, we define two 3×3 matrix-generating operators, $\mathbf{v}(S)$ and $\mathbf{t}(T)$, which map a set of vector or tensor parameters into the standard helicity basis matrices:

$$\mathbf{v}(S) \equiv \frac{1}{2} \sum_{k=-1}^{+1} \langle S_k \rangle^* S_k = \begin{pmatrix} \frac{\langle S_z \rangle}{2} & \frac{\langle S_x \rangle - \mathrm{i}\langle S_y \rangle}{2\sqrt{2}} & 0 \\ \frac{\langle S_x \rangle + \mathrm{i}\langle S_y \rangle}{2\sqrt{2}} & 0 & \frac{\langle S_x \rangle - \mathrm{i}\langle S_y \rangle}{2\sqrt{2}} \\ 0 & \frac{\langle S_x \rangle + \mathrm{i}\langle S_y \rangle}{2\sqrt{2}} & -\frac{\langle S_z \rangle}{2} \end{pmatrix}, \tag{B.1}$$

$$\mathbf{t}(T) \equiv \sum_{k=-2}^{+2} \langle T_k \rangle^* T_k = \begin{pmatrix} \frac{\langle T_0 \rangle}{\sqrt{6}} & \frac{-\langle A_1 \rangle + \mathrm{i}\langle A_2 \rangle}{\sqrt{2}} & \langle B_1 \rangle - \mathrm{i}\langle B_2 \rangle \\ \frac{-\langle A_1 \rangle - \mathrm{i}\langle A_2 \rangle}{\sqrt{2}} & -\frac{2\langle T_0 \rangle}{\sqrt{6}} & \frac{\langle A_1 \rangle - \mathrm{i}\langle A_2 \rangle}{\sqrt{2}} \\ \langle B_1 \rangle + \mathrm{i}\langle B_2 \rangle & \frac{\langle A_1 \rangle + \mathrm{i}\langle A_2 \rangle}{\sqrt{2}} & \frac{\langle T_0 \rangle}{\sqrt{6}} \end{pmatrix}, \tag{B.2}$$

which can be seen as a decomposition of the vector and tensor parts of the SDM for a single boson given in Equation 6.61.

By utilising the tensor product $\otimes$, the state of the W boson determines the outer 3×3 grid, while the state of the Z boson populates the inner 3×3 sub-matrices. We can rewrite Equation 6.90 as a matrix that corresponds to the sum of seven distinct 9×9 blocks:

$$\rho^{\mathrm{WZ}} = \rho_{\mathrm{unpol.}} + \rho_{\mathrm{V}} + \rho_{\mathrm{T}} + \rho_{\mathrm{VV}} + \rho_{\mathrm{VT}} + \rho_{\mathrm{TV}} + \rho_{\mathrm{TT}}\,, \tag{B.3}$$

where $\rho_{\mathrm{unpol.}}$ is the unpolarised background, ρ_{V} and ρ_{T} are the single-boson vector and tensor contributions, and ρ_{VV}, ρ_{VT}, ρ_{TV}, and ρ_{TT} are the joint spin correlation blocks.

B.1.1 Unpolarised and single-boson polarisations

The unpolarised background.

$$\rho_{\mathrm{unpol.}} = \frac{1}{9}\mathbb{I}_3 \otimes \mathbb{I}_3 = \frac{1}{9}\begin{pmatrix} \mathbb{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbb{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbb{I}_3 \end{pmatrix}. \tag{B.4}$$

Single-boson vector polarisations. Let $\mathbf{v}^{\mathrm{W}} = \mathbf{v}(\langle \vec{S}^{\mathrm{W}} \rangle)$ and $\mathbf{v}^{\mathrm{Z}} = \mathbf{v}(\langle \vec{S}^{\mathrm{Z}} \rangle)$:

$$\rho_{\mathrm{V}} = \frac{1}{3}\begin{pmatrix} v^{\mathrm{W}}_{++}\mathbb{I}_3 + \mathbf{v}^{\mathrm{Z}} & v^{\mathrm{W}}_{+0}\mathbb{I}_3 & \mathbf{0} \\ v^{\mathrm{W}}_{0+}\mathbb{I}_3 & \mathbf{v}^{\mathrm{Z}} & v^{\mathrm{W}}_{0-}\mathbb{I}_3 \\ \mathbf{0} & v^{\mathrm{W}}_{-0}\mathbb{I}_3 & v^{\mathrm{W}}_{--}\mathbb{I}_3 + \mathbf{v}^{\mathrm{Z}} \end{pmatrix}, \tag{B.5}$$

where v^V_{xy} is the element (x, y) of the vector matrix for a boson V given in Equation B.1.

Single-boson tensor alignments. Similarly, let $\mathbf{t}^{\mathrm{W}} = \mathbf{t}(\langle \vec{T}^{\mathrm{W}} \rangle)$ and $\mathbf{t}^{\mathrm{Z}} = \mathbf{t}(\langle \vec{T}^{\mathrm{Z}} \rangle)$:

$$\rho_{\mathrm{T}} = \frac{1}{3} \begin{pmatrix} t^{\mathrm{W}}_{++}\mathbb{I}_3 + \mathbf{t}^{\mathrm{Z}} & t^{\mathrm{W}}_{+0}\mathbb{I}_3 & t^{\mathrm{W}}_{+-}\mathbb{I}_3 \\ t^{\mathrm{W}}_{0+}\mathbb{I}_3 & t^{\mathrm{W}}_{00}\mathbb{I}_3 + \mathbf{t}^{\mathrm{Z}} & t^{\mathrm{W}}_{0-}\mathbb{I}_3 \\ t^{\mathrm{W}}_{-+}\mathbb{I}_3 & t^{\mathrm{W}}_{-0}\mathbb{I}_3 & t^{\mathrm{W}}_{--}\mathbb{I}_3 + \mathbf{t}^{\mathrm{Z}} \end{pmatrix}, \tag{B.6}$$

where t^V_{xy} is the element (x, y) of the tensor matrix for a boson V given in Equation B.2.

B.1.2 Joint spin correlations

For the cross-correlated terms, we feed the joint expectation values directly into the inner matrices. The notation $\mathbf{v}(\langle S^{\mathrm{W}}_z \vec{S}^{\mathrm{Z}} \rangle)$ represents the 3×3 vector matrix $\mathbf{v}$ where every parameter is the joint measurement of the Z boson's vector spin coinciding with the W boson's S_z projection.

Vector–vector correlations.

$$\rho_{\mathrm{VV}} = \frac{1}{2} \begin{pmatrix} \mathbf{v}(\langle S^{\mathrm{W}}_z \vec{S}^{\mathrm{Z}} \rangle) & \frac{\mathbf{v}(\langle S^{\mathrm{W}}_x \vec{S}^{\mathrm{Z}} \rangle) - \mathrm{i}\mathbf{v}(\langle S^{\mathrm{W}}_y \vec{S}^{\mathrm{Z}} \rangle)}{\sqrt{2}} & \mathbf{0} \\ \frac{\mathbf{v}(\langle S^{\mathrm{W}}_x \vec{S}^{\mathrm{Z}} \rangle) + \mathrm{i}\mathbf{v}(\langle S^{\mathrm{W}}_y \vec{S}^{\mathrm{Z}} \rangle)}{\sqrt{2}} & \mathbf{0} & \frac{\mathbf{v}(\langle S^{\mathrm{W}}_x \vec{S}^{\mathrm{Z}} \rangle) - \mathrm{i}\mathbf{v}(\langle S^{\mathrm{W}}_y \vec{S}^{\mathrm{Z}} \rangle)}{\sqrt{2}} \\ \mathbf{0} & \frac{\mathbf{v}(\langle S^{\mathrm{W}}_x \vec{S}^{\mathrm{Z}} \rangle) + \mathrm{i}\mathbf{v}(\langle S^{\mathrm{W}}_y \vec{S}^{\mathrm{Z}} \rangle)}{\sqrt{2}} & -\mathbf{v}(\langle S^{\mathrm{W}}_z \vec{S}^{\mathrm{Z}} \rangle) \end{pmatrix}, \tag{B.7}$$

where $\vec{S}^V = (S^V_x, S^V_y, S^V_z)$ is the vector of spin operators for a boson V.

Vector–tensor correlations.

$$\rho_{\mathrm{VT}} = \frac{1}{2} \begin{pmatrix} \mathbf{t}(\langle S^{\mathrm{W}}_z \vec{T}^{\mathrm{Z}} \rangle) & \frac{\mathbf{t}(\langle S^{\mathrm{W}}_x \vec{T}^{\mathrm{Z}} \rangle) - \mathrm{i}\mathbf{t}(\langle S^{\mathrm{W}}_y \vec{T}^{\mathrm{Z}} \rangle)}{\sqrt{2}} & \mathbf{0} \\ \frac{\mathbf{t}(\langle S^{\mathrm{W}}_x \vec{T}^{\mathrm{Z}} \rangle) + \mathrm{i}\mathbf{t}(\langle S^{\mathrm{W}}_y \vec{T}^{\mathrm{Z}} \rangle)}{\sqrt{2}} & \mathbf{0} & \frac{\mathbf{t}(\langle S^{\mathrm{W}}_x \vec{T}^{\mathrm{Z}} \rangle) - \mathrm{i}\mathbf{t}(\langle S^{\mathrm{W}}_y \vec{T}^{\mathrm{Z}} \rangle)}{\sqrt{2}} \\ \mathbf{0} & \frac{\mathbf{t}(\langle S^{\mathrm{W}}_x \vec{T}^{\mathrm{Z}} \rangle) + \mathrm{i}\mathbf{t}(\langle S^{\mathrm{W}}_y \vec{T}^{\mathrm{Z}} \rangle)}{\sqrt{2}} & -\mathbf{t}(\langle S^{\mathrm{W}}_z \vec{T}^{\mathrm{Z}} \rangle) \end{pmatrix}, \tag{B.8}$$

where $\vec{T}^V = (T^V_0, A^V_1, A^V_2, B^V_1, B^V_2)$ is the vector of tensor operators for a boson V.

Tensor–vector correlations.

$$\rho_{\mathrm{TV}} = \begin{pmatrix} \frac{\mathbf{v}(\langle T^{\mathrm{W}}_0 \vec{S}^{\mathrm{Z}} \rangle)}{\sqrt{6}} & \frac{-\mathbf{v}(\langle A^{\mathrm{W}}_1 \vec{S}^{\mathrm{Z}} \rangle) + \mathrm{i}\mathbf{v}(\langle A^{\mathrm{W}}_2 \vec{S}^{\mathrm{Z}} \rangle)}{\sqrt{2}} & \mathbf{v}(\langle B^{\mathrm{W}}_1 \vec{S}^{\mathrm{Z}} \rangle) - \mathrm{i}\mathbf{v}(\langle B^{\mathrm{W}}_2 \vec{S}^{\mathrm{Z}} \rangle) \\ \frac{-\mathbf{v}(\langle A^{\mathrm{W}}_1 \vec{S}^{\mathrm{Z}} \rangle) - \mathrm{i}\mathbf{v}(\langle A^{\mathrm{W}}_2 \vec{S}^{\mathrm{Z}} \rangle)}{\sqrt{2}} & -\frac{2\mathbf{v}(\langle T^{\mathrm{W}}_0 \vec{S}^{\mathrm{Z}} \rangle)}{\sqrt{6}} & \frac{\mathbf{v}(\langle A^{\mathrm{W}}_1 \vec{S}^{\mathrm{Z}} \rangle) - \mathrm{i}\mathbf{v}(\langle A^{\mathrm{W}}_2 \vec{S}^{\mathrm{Z}} \rangle)}{\sqrt{2}} \\ \mathbf{v}(\langle B^{\mathrm{W}}_1 \vec{S}^{\mathrm{Z}} \rangle) + \mathrm{i}\mathbf{v}(\langle B^{\mathrm{W}}_2 \vec{S}^{\mathrm{Z}} \rangle) & \frac{\mathbf{v}(\langle A^{\mathrm{W}}_1 \vec{S}^{\mathrm{Z}} \rangle) + \mathrm{i}\mathbf{v}(\langle A^{\mathrm{W}}_2 \vec{S}^{\mathrm{Z}} \rangle)}{\sqrt{2}} & \frac{\mathbf{v}(\langle T^{\mathrm{W}}_0 \vec{S}^{\mathrm{Z}} \rangle)}{\sqrt{6}} \end{pmatrix}. \tag{B.9}$$

Tensor–tensor correlations.

$$\rho_{\mathrm{TT}} = \begin{pmatrix} \frac{\mathbf{t}(\langle T^{\mathrm{W}}_0 \vec{T}^{\mathrm{Z}} \rangle)}{\sqrt{6}} & \frac{-\mathbf{t}(\langle A^{\mathrm{W}}_1 \vec{T}^{\mathrm{Z}} \rangle) + \mathrm{i}\mathbf{t}(\langle A^{\mathrm{W}}_2 \vec{T}^{\mathrm{Z}} \rangle)}{\sqrt{2}} & \mathbf{t}(\langle B^{\mathrm{W}}_1 \vec{T}^{\mathrm{Z}} \rangle) - \mathrm{i}\mathbf{t}(\langle B^{\mathrm{W}}_2 \vec{T}^{\mathrm{Z}} \rangle) \\ \frac{-\mathbf{t}(\langle A^{\mathrm{W}}_1 \vec{T}^{\mathrm{Z}} \rangle) - \mathrm{i}\mathbf{t}(\langle A^{\mathrm{W}}_2 \vec{T}^{\mathrm{Z}} \rangle)}{\sqrt{2}} & -\frac{2\mathbf{t}(\langle T^{\mathrm{W}}_0 \vec{T}^{\mathrm{Z}} \rangle)}{\sqrt{6}} & \frac{\mathbf{t}(\langle A^{\mathrm{W}}_1 \vec{T}^{\mathrm{Z}} \rangle) - \mathrm{i}\mathbf{t}(\langle A^{\mathrm{W}}_2 \vec{T}^{\mathrm{Z}} \rangle)}{\sqrt{2}} \\ \mathbf{t}(\langle B^{\mathrm{W}}_1 \vec{T}^{\mathrm{Z}} \rangle) + \mathrm{i}\mathbf{t}(\langle B^{\mathrm{W}}_2 \vec{T}^{\mathrm{Z}} \rangle) & \frac{\mathbf{t}(\langle A^{\mathrm{W}}_1 \vec{T}^{\mathrm{Z}} \rangle) + \mathrm{i}\mathbf{t}(\langle A^{\mathrm{W}}_2 \vec{T}^{\mathrm{Z}} \rangle)}{\sqrt{2}} & \frac{\mathbf{t}(\langle T^{\mathrm{W}}_0 \vec{T}^{\mathrm{Z}} \rangle)}{\sqrt{6}} \end{pmatrix}. \tag{B.10}$$

In quantum mechanics, a joint measurement of two observables on a bipartite state yields the expectation value $\langle O^{\mathrm{W}} O^{\mathrm{Z}} \rangle$. If the two gauge bosons are produced completely independently without any spin or kinematic correlation, their joint state is a pure tensor product of their individual density matrices, and their joint expectation values naturally factorise: $\langle O^{\mathrm{W}} O^{\mathrm{Z}} \rangle = \langle O^{\mathrm{W}} \rangle \langle O^{\mathrm{Z}} \rangle$.

However, because the W and Z bosons are correlated by their shared production mechanism, this factorisation is incomplete. The deviation from pure independence is captured by the statistical covariance:

$$\langle O^{\mathrm{W}} O^{\mathrm{Z}} \rangle = \langle O^{\mathrm{W}} \rangle \langle O^{\mathrm{Z}} \rangle + \mathrm{cov}(O^{\mathrm{W}}, O^{\mathrm{Z}}). \tag{B.11}$$

Substituting this elementary statistical decomposition into the joint spin correlation terms allows each 9×9 block to be split linearly into a factorised (unentangled) component and a genuine correlation component:

$$\rho_{\mathrm{XY}} = \rho_{\mathrm{XY}}^{\mathrm{fact}} + \rho_{\mathrm{XY}}^{\mathrm{cov}}\,, \tag{B.12}$$

where $\mathrm{X}, \mathrm{Y} \in \{\mathrm{V}, \mathrm{T}\}$. The factorised matrices ($\rho_{\mathrm{XY}}^{\mathrm{fact}}$) are populated entirely by the cross terms $\langle O^{\mathrm{W}} \rangle \langle O^{\mathrm{Z}} \rangle$. The covariance matrices ($\rho_{\mathrm{XY}}^{\mathrm{cov}}$) have the exact same structural layout as their parent matrices, but every joint expectation value $\langle O^{\mathrm{W}} O^{\mathrm{Z}} \rangle$ is replaced directly by $\mathrm{cov}(O^{\mathrm{W}}, O^{\mathrm{Z}})$.

B.1.3 Joint spin density matrix decomposition

The true power of this decomposition emerges when we assemble the full 9×9 system. The unpolarised background, the single-boson terms, and the factorised correlation matrices exactly reconstruct the pure tensor product of the two independent single-boson density matrices. The remainder consists entirely of the true covariance matrices.

Consequently, the complete joint SDM ρ^{WZ} is elegantly split into an uncorrelated production background and the connected (non-factorisable) correlation tensor:

$$\begin{aligned} \rho^{\mathrm{WZ}} &= \rho_{\mathrm{unpol.}} + \rho_{\mathrm{V}} + \rho_{\mathrm{T}} + \sum_{\mathrm{X,Y} \in \{\mathrm{V,T}\}} \left(\rho_{\mathrm{XY}}^{\mathrm{fact}} + \rho_{\mathrm{XY}}^{\mathrm{cov}} \right) \\ &= \left(\rho^{\mathrm{W}} \otimes \rho^{\mathrm{Z}} \right) + \sum_{\mathrm{X,Y} \in \{\mathrm{V,T}\}} \rho_{\mathrm{XY}}^{\mathrm{cov}} \\ &= \rho^{\mathrm{W}} \otimes \rho^{\mathrm{Z}} + \mathcal{C}^{\mathrm{WZ}}\,, \end{aligned} \tag{B.13}$$

where ρ^{W} and ρ^{Z} are the independent 3×3 single-boson spin density matrices defined in Equation 6.50, and $\mathcal{C}^{\mathrm{WZ}}$ is the total 9×9 spin-correlation tensor containing exclusively the true statistical covariance terms.

To map this cleanly into the 9×9 matrix grid, we partition each of the 9×9 covariance matrices ($\rho_{\mathrm{XY}}^{\mathrm{cov}}$) into a 3×3 grid of inner 3×3 sub-matrices. We define $\mathbf{C}_{\mathrm{XY}}^{ij,\mathrm{cov}}$ as the specific 3×3 inner sub-matrix corresponding to the transition $|i\rangle\langle j|$ of the outer W boson.

To find the total correlation contribution mapping to the $|i\rangle\langle j|$ state of the W boson, we simply sum these four sub-blocks. We define this total inner correlation matrix as $\mathbf{\Sigma}_{\mathrm{cov}}^{ij}$:

$$\mathbf{\Sigma}_{\mathrm{cov}}^{ij} = \mathbf{C}_{\mathrm{VV}}^{ij,\mathrm{cov}} + \mathbf{C}_{\mathrm{VT}}^{ij,\mathrm{cov}} + \mathbf{C}_{\mathrm{TV}}^{ij,\mathrm{cov}} + \mathbf{C}_{\mathrm{TT}}^{ij,\mathrm{cov}}\,. \tag{B.14}$$

Physically, $\mathbf{\Sigma}_{\mathrm{cov}}^{ij}$ represents the complete set of spin correlations modifying the Z boson's state specifically when the W boson undergoes the $i \to j$ helicity transition.

Summing the factorised product and these $\mathbf{\Sigma}_{\mathrm{cov}}^{ij}$ matrices block-by-block yields the fully developed 9×9 joint SDM:

$$\rho^{\mathrm{WZ}} = \begin{pmatrix} \left[\left(\frac{1}{9} + \frac{v^{\mathrm{W}}_{++} + t^{\mathrm{W}}_{++}}{3}\right)\mathbb{I}_3 + \frac{\mathbf{v}^{\mathrm{Z}} + \mathbf{t}^{\mathrm{Z}}}{3} + (v^{\mathrm{W}}_{++} + t^{\mathrm{W}}_{++})(\mathbf{v}^{\mathrm{Z}} + \mathbf{t}^{\mathrm{Z}}) + \mathbf{\Sigma}^{++}_{\mathrm{cov}}\right] & \left[\frac{v^{\mathrm{W}}_{+0} + t^{\mathrm{W}}_{+0}}{3}\mathbb{I}_3 + (v^{\mathrm{W}}_{+0} + t^{\mathrm{W}}_{+0})(\mathbf{v}^{\mathrm{Z}} + \mathbf{t}^{\mathrm{Z}}) + \mathbf{\Sigma}^{+0}_{\mathrm{cov}}\right] & \left[\frac{t^{\mathrm{W}}_{+-}}{3}\mathbb{I}_3 + t^{\mathrm{W}}_{+-}(\mathbf{v}^{\mathrm{Z}} + \mathbf{t}^{\mathrm{Z}}) + \mathbf{\Sigma}^{+-}_{\mathrm{cov}}\right] \\ \left[\frac{v^{\mathrm{W}}_{0+} + t^{\mathrm{W}}_{0+}}{3}\mathbb{I}_3 + (v^{\mathrm{W}}_{0+} + t^{\mathrm{W}}_{0+})(\mathbf{v}^{\mathrm{Z}} + \mathbf{t}^{\mathrm{Z}}) + \mathbf{\Sigma}^{0+}_{\mathrm{cov}}\right] & \left[\left(\frac{1}{9} + \frac{t^{\mathrm{W}}_{00}}{3}\right)\mathbb{I}_3 + \frac{\mathbf{v}^{\mathrm{Z}} + \mathbf{t}^{\mathrm{Z}}}{3} + t^{\mathrm{W}}_{00}(\mathbf{v}^{\mathrm{Z}} + \mathbf{t}^{\mathrm{Z}}) + \mathbf{\Sigma}^{00}_{\mathrm{cov}}\right] & \left[\frac{v^{\mathrm{W}}_{0-} + t^{\mathrm{W}}_{0-}}{3}\mathbb{I}_3 + (v^{\mathrm{W}}_{0-} + t^{\mathrm{W}}_{0-})(\mathbf{v}^{\mathrm{Z}} + \mathbf{t}^{\mathrm{Z}}) + \mathbf{\Sigma}^{0-}_{\mathrm{cov}}\right] \\ \left[\frac{t^{\mathrm{W}}_{-+}}{3}\mathbb{I}_3 + t^{\mathrm{W}}_{-+}(\mathbf{v}^{\mathrm{Z}} + \mathbf{t}^{\mathrm{Z}}) + \mathbf{\Sigma}^{-+}_{\mathrm{cov}}\right] & \left[\frac{v^{\mathrm{W}}_{-0} + t^{\mathrm{W}}_{-0}}{3}\mathbb{I}_3 + (v^{\mathrm{W}}_{-0} + t^{\mathrm{W}}_{-0})(\mathbf{v}^{\mathrm{Z}} + \mathbf{t}^{\mathrm{Z}}) + \mathbf{\Sigma}^{-0}_{\mathrm{cov}}\right] & \left[\left(\frac{1}{9} + \frac{v^{\mathrm{W}}_{--} + t^{\mathrm{W}}_{--}}{3}\right)\mathbb{I}_3 + \frac{\mathbf{v}^{\mathrm{Z}} + \mathbf{t}^{\mathrm{Z}}}{3} + (v^{\mathrm{W}}_{--} + t^{\mathrm{W}}_{--})(\mathbf{v}^{\mathrm{Z}} + \mathbf{t}^{\mathrm{Z}}) + \mathbf{\Sigma}^{--}_{\mathrm{cov}}\right] \end{pmatrix}. \tag{B.15}$$

A version of this same matrix with every term explicitly written out in terms of the underlying spin expectation values is provided at the end of this appendix in Equation B.32.

To see how a specific index maps out, we look at the absolute first element of the 9×9 matrix (row 1, column 1). This is the $\rho_{++,++}$ entry representing both bosons simultaneously collapsing into the $+1$ helicity state:

$$\begin{aligned}\rho^{\mathrm{WZ}}_{++,++} &= \left(\frac{1}{3} + v^{\mathrm{W}}_{++} + t^{\mathrm{W}}_{++}\right)\left(\frac{1}{3} + v^{\mathrm{Z}}_{++} + t^{\mathrm{Z}}_{++}\right) + \left(\boldsymbol{\Sigma}^{++}_{\mathrm{cov}}\right)_{++} \\ &= \frac{1}{9} + \frac{1}{3}\left(\frac{\langle S^{\mathrm{W}}_z\rangle}{2} + \frac{\langle T^{\mathrm{W}}_0\rangle}{\sqrt{6}}\right) + \frac{1}{3}\left(\frac{\langle S^{\mathrm{Z}}_z\rangle}{2} + \frac{\langle T^{\mathrm{Z}}_0\rangle}{\sqrt{6}}\right) \\ &\quad + \left(\frac{\langle S^{\mathrm{W}}_z\rangle}{2} + \frac{\langle T^{\mathrm{W}}_0\rangle}{\sqrt{6}}\right)\left(\frac{\langle S^{\mathrm{Z}}_z\rangle}{2} + \frac{\langle T^{\mathrm{Z}}_0\rangle}{\sqrt{6}}\right) + \left(\boldsymbol{\Sigma}^{++}_{\mathrm{cov}}\right)_{++} .\end{aligned} \tag{B.16}$$

By grouping the individual terms, this element directly matches the product of independent density elements plus the explicit correlation contribution:

$$\rho^{\mathrm{WZ}}_{++,++} = \rho^{\mathrm{W}}_{++} \otimes \rho^{\mathrm{Z}}_{++} + \mathcal{C}^{\mathrm{WZ}}_{++,++} . \tag{B.17}$$

This confirms that the element-by-element derivation and the statistical covariance tensor approach align perfectly across the full 9×9 matrix space.

B.2 Physical polarisation fractions

The most physically meaningful observables extracted from the master joint SDM are the joint polarisation fractions $f_{\lambda_{\mathrm{W}}\lambda_{\mathrm{Z}}}$. These fractions project the abstract SDM back into the measurable combinations of experimental helicity states. Mapping the standard helicity indices $+1$ and -1 to the physical RH (R) and LH (L) transverse polarisation configurations, respectively, the key diagonal joint fractions are isolated by adapting the formalism introduced in this appendix to Equation 6.91:

$$f_{00} \equiv \rho^{\mathrm{WZ}}_{00,00} , \tag{B.18}$$

$$f_{0\mathrm{T}} \equiv \rho^{\mathrm{WZ}}_{0\mathrm{L},0\mathrm{L}} + \rho^{\mathrm{WZ}}_{0\mathrm{R},0\mathrm{R}} = \rho^{\mathrm{WZ}}_{0-,0-} + \rho^{\mathrm{WZ}}_{0+,0+} , \tag{B.19}$$

$$f_{\mathrm{T}0} \equiv \rho^{\mathrm{WZ}}_{\mathrm{L}0,\mathrm{L}0} + \rho^{\mathrm{WZ}}_{\mathrm{R}0,\mathrm{R}0} = \rho^{\mathrm{WZ}}_{-0,-0} + \rho^{\mathrm{WZ}}_{+0,+0} , \tag{B.20}$$

$$f_{\mathrm{TT}} \equiv \sum_{i,j\in\{\mathrm{L},\mathrm{R}\}} \rho^{\mathrm{WZ}}_{ij,ij} = \sum_{i,j\in\{-,+\}} \rho^{\mathrm{WZ}}_{ij,ij} . \tag{B.21}$$

Here, f_{00} represents the purely longitudinal joint state, $f_{0\mathrm{T}}$ and $f_{\mathrm{T}0}$ represent mixed longitudinal–transverse combinations, and f_{TT} denotes the completely transverse joint polarisation state. Because these fractions correspond directly to real-space cross-sections and angular distributions, they serve as the crucial bridge connecting the structural components of Equation B.15 to experimental validation.

By substituting the explicit diagonal elements from the 9×9 super-matrix into the definitions above, and crucially including the factorised quadratic terms ($\mathbf{t}^{\mathrm{W}} \otimes \mathbf{t}^{\mathrm{Z}}$) that arise from the direct tensor product $\rho^{\mathrm{W}} \otimes \rho^{\mathrm{Z}}$, we can write out the physical polarisation fractions directly.

We recall that for the single-boson matrices, the sum of the outer diagonal elements simplifies elegantly: $v_{++} + v_{--} = 0$ and $t_{++} + t_{--} = -t_{00}$. Utilising these relations, and noting that the product of the independent tensor elements expands precisely as $t^{\mathrm{W}}_{00} t^{\mathrm{Z}}_{00} = \left(-\frac{2\langle T^{\mathrm{W}}_0\rangle}{\sqrt{6}}\right)\left(-\frac{2\langle T^{\mathrm{Z}}_0\rangle}{\sqrt{6}}\right) = \frac{2}{3}\langle T^{\mathrm{W}}_0\rangle\langle T^{\mathrm{Z}}_0\rangle$, the joint polarisation fractions expand into:

$$f_{00} = \frac{1}{9} - \frac{2\langle T^{\mathrm{W}}_0\rangle}{3\sqrt{6}} - \frac{2\langle T^{\mathrm{Z}}_0\rangle}{3\sqrt{6}} + \frac{2}{3}\langle T^{\mathrm{W}}_0\rangle\langle T^{\mathrm{Z}}_0\rangle + \left(\boldsymbol{\Sigma}^{00}_{\mathrm{cov}}\right)_{00} , \tag{B.22}$$

$$f_{0\mathrm{T}} = \frac{2}{9} - \frac{4\langle T^{\mathrm{W}}_0\rangle}{3\sqrt{6}} + \frac{2\langle T^{\mathrm{Z}}_0\rangle}{3\sqrt{6}} - \frac{2}{3}\langle T^{\mathrm{W}}_0\rangle\langle T^{\mathrm{Z}}_0\rangle + \left(\boldsymbol{\Sigma}^{00}_{\mathrm{cov}}\right)_{++} + \left(\boldsymbol{\Sigma}^{00}_{\mathrm{cov}}\right)_{--} , \tag{B.23}$$

$$f_{\mathrm{T}0} = \frac{2}{9} + \frac{2\langle T^{\mathrm{W}}_0\rangle}{3\sqrt{6}} - \frac{4\langle T^{\mathrm{Z}}_0\rangle}{3\sqrt{6}} - \frac{2}{3}\langle T^{\mathrm{W}}_0\rangle\langle T^{\mathrm{Z}}_0\rangle + \left(\boldsymbol{\Sigma}^{++}_{\mathrm{cov}}\right)_{00} + \left(\boldsymbol{\Sigma}^{--}_{\mathrm{cov}}\right)_{00} , \tag{B.24}$$

$$f_{\mathrm{TT}} = \frac{4}{9} + \frac{4\langle T^{\mathrm{W}}_0\rangle}{3\sqrt{6}} + \frac{4\langle T^{\mathrm{Z}}_0\rangle}{3\sqrt{6}} + \frac{2}{3}\langle T^{\mathrm{W}}_0\rangle\langle T^{\mathrm{Z}}_0\rangle + \sum_{i,j\in\{+,-\}} \left(\boldsymbol{\Sigma}^{ii}_{\mathrm{cov}}\right)_{jj} . \tag{B.25}$$

We note a fundamental consequence of this expansion: if we sum all four fractions, $f_{00}+f_{0\mathrm{T}}+f_{\mathrm{T}0}+f_{\mathrm{TT}}$, the single-boson tensor components ($\langle T_0 \rangle$) and their cross terms ($\langle T_0^{\mathrm{W}} \rangle \langle T_0^{\mathrm{Z}} \rangle$) completely cancel out. The baseline constants mathematically sum to exactly 1 ($\frac{1}{9}+\frac{2}{9}+\frac{2}{9}+\frac{4}{9}=1$), and the correlation elements sum to their matrix trace, which is exactly 0. This perfectly preserves the quantum probability requirement that $\mathrm{Tr}(\rho^{\mathrm{WZ}})=1$. Furthermore, the independent single-boson fractions factorise perfectly: for instance, ignoring the covariance term, f_{00} reduces exactly to $\left(\frac{1}{3}-\frac{2\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\left(\frac{1}{3}-\frac{2\langle T_0^{\mathrm{Z}}\rangle}{\sqrt{6}}\right)=f_0^{\mathrm{W}} f_0^{\mathrm{Z}}$.

B.3 Decay density matrix

The individual decay density matrices [392] Γ^{W} and Γ^{Z} were already derived in Equations 6.72 and 6.81. Γ^{W} is rewritten here to include the $\pm$ signs to explicitly show the difference between the W^+ and W^- cases:

$$\Gamma^{\mathrm{W}^{\pm}}(\theta_{\mathrm{W}}^*,\phi_{\mathrm{W}}^*)=\begin{pmatrix} \frac{(1\pm\cos\theta_{\mathrm{W}}^*)^2}{4} & \pm\frac{\sin\theta_{\mathrm{W}}^*(1\pm\cos\theta_{\mathrm{W}}^*)\mathrm{e}^{\mathrm{i}\phi_{\mathrm{W}}^*}}{2\sqrt{2}} & \frac{\sin^2\theta_{\mathrm{W}}^*\mathrm{e}^{2\mathrm{i}\phi_{\mathrm{W}}^*}}{4} \\ \pm\frac{\sin\theta_{\mathrm{W}}^*(1\pm\cos\theta_{\mathrm{W}}^*)\mathrm{e}^{-\mathrm{i}\phi_{\mathrm{W}}^*}}{2\sqrt{2}} & \frac{\sin^2\theta_{\mathrm{W}}^*}{2} & \pm\frac{\sin\theta_{\mathrm{W}}^*(1\mp\cos\theta_{\mathrm{W}}^*)\mathrm{e}^{\mathrm{i}\phi_{\mathrm{W}}^*}}{2\sqrt{2}} \\ \frac{\sin^2\theta_{\mathrm{W}}^*\mathrm{e}^{-2\mathrm{i}\phi_{\mathrm{W}}^*}}{4} & \pm\frac{\sin\theta_{\mathrm{W}}^*(1\mp\cos\theta_{\mathrm{W}}^*)\mathrm{e}^{-\mathrm{i}\phi_{\mathrm{W}}^*}}{2\sqrt{2}} & \frac{(1\mp\cos\theta_{\mathrm{W}}^*)^2}{4} \end{pmatrix}. \tag{B.26}$$

The joint decay operator in Equation 6.100 factorises as the direct product $\Gamma^{\mathrm{WZ}}=\Gamma^{\mathrm{W}}\otimes\Gamma^{\mathrm{Z}}$. Following the product basis ordering $\{|+1\rangle|+1\rangle,|+1\rangle|0\rangle,\ldots,|-1\rangle|-1\rangle\}$, the diagonal elements of this 9×9 matrix correspond to the pure joint helicity decay profiles:

$$\mathrm{diag}\left(\Gamma^{\mathrm{W}}\otimes\Gamma^{\mathrm{Z}}\right)=\begin{pmatrix} \frac{(1\pm\cos\theta_{\mathrm{W}}^*)^2(1+\cos^2\theta_{\mathrm{Z}}^*-2A_\ell\cos\theta_{\mathrm{Z}}^*)}{16} \\ \frac{(1\pm\cos\theta_{\mathrm{W}}^*)^2\sin^2\theta_{\mathrm{Z}}^*}{8} \\ \frac{(1\pm\cos\theta_{\mathrm{W}}^*)^2(1+\cos^2\theta_{\mathrm{Z}}^*+2A_\ell\cos\theta_{\mathrm{Z}}^*)}{16} \\ \hline \frac{\sin^2\theta_{\mathrm{W}}^*(1+\cos^2\theta_{\mathrm{Z}}^*-2A_\ell\cos\theta_{\mathrm{Z}}^*)}{8} \\ \frac{\sin^2\theta_{\mathrm{W}}^*\sin^2\theta_{\mathrm{Z}}^*}{4} \\ \frac{\sin^2\theta_{\mathrm{W}}^*(1+\cos^2\theta_{\mathrm{Z}}^*+2A_\ell\cos\theta_{\mathrm{Z}}^*)}{8} \\ \hline \frac{(1\mp\cos\theta_{\mathrm{W}}^*)^2(1+\cos^2\theta_{\mathrm{Z}}^*-2A_\ell\cos\theta_{\mathrm{Z}}^*)}{16} \\ \frac{(1\mp\cos\theta_{\mathrm{W}}^*)^2\sin^2\theta_{\mathrm{Z}}^*}{8} \\ \frac{(1\mp\cos\theta_{\mathrm{W}}^*)^2(1+\cos^2\theta_{\mathrm{Z}}^*+2A_\ell\cos\theta_{\mathrm{Z}}^*)}{16} \end{pmatrix}. \tag{B.27}$$

The off-diagonal entries are given by $\left(\Gamma^{\mathrm{W}}\otimes\Gamma^{\mathrm{Z}}\right)_{\lambda_{\mathrm{W}}\lambda_{\mathrm{Z}},\lambda'_{\mathrm{W}}\lambda'_{\mathrm{Z}}}=\Gamma^{\mathrm{W}}_{\lambda_{\mathrm{W}}\lambda'_{\mathrm{W}}}\cdot\Gamma^{\mathrm{Z}}_{\lambda_{\mathrm{Z}}\lambda'_{\mathrm{Z}}}$, keeping track of the interference phases $\mathrm{e}^{\mathrm{i}(n_{\mathrm{W}}\phi_{\mathrm{W}}^*+n_{\mathrm{Z}}\phi_{\mathrm{Z}}^*)}$.

B.4 Joint angular distribution

Equation 6.100 can be rewritten in the more explicit form of a 4D angular distribution by substituting the full 9×9 joint SDM decomposition from Equation B.13 and the direct product decay operator $\Gamma^{\mathrm{W}}\otimes\Gamma^{\mathrm{Z}}$:

$$\frac{1}{\sigma_{\mathrm{WZ}\to\ell'^{\pm}\nu_\ell\ell\ell}}\frac{\mathrm{d}^4\,\sigma_{\mathrm{WZ}\to\ell'^{\pm}\nu_\ell\ell\ell}}{\mathrm{d}\,\cos\theta^*_{\mathrm{W}}\mathrm{d}\,\phi^*_{\mathrm{W}}\mathrm{d}\,\cos\theta^*_{\mathrm{Z}}\mathrm{d}\,\phi^*_{\mathrm{Z}}} = \left(\frac{3}{4\pi}\right)^2 \sum_{\lambda_{\mathrm{W}},\lambda'_{\mathrm{W}}}\sum_{\lambda_{\mathrm{Z}},\lambda'_{\mathrm{Z}}} \rho^{\mathrm{WZ}}_{\lambda_{\mathrm{W}}\lambda_{\mathrm{Z}},\lambda'_{\mathrm{W}}\lambda'_{\mathrm{Z}}} \times \Gamma^{\mathrm{W}^\pm}_{\lambda_{\mathrm{W}}\lambda'_{\mathrm{W}}}(\theta^*_{\mathrm{W}},\phi^*_{\mathrm{W}})\,\Gamma^{\mathrm{Z}}_{\lambda_{\mathrm{Z}}\lambda'_{\mathrm{Z}}}(\theta^*_{\mathrm{Z}},\phi^*_{\mathrm{Z}})\,. \tag{B.28}$$

Integrating Equation B.28 over both azimuthal angles ϕ^*_{W} and ϕ^*_{Z} removes all off-diagonal SDM interference terms. This isolates the nine diagonal joint helicity fractions $f_{\lambda_{\mathrm{W}}\lambda_{\mathrm{Z}}} \equiv \rho^{\mathrm{WZ}}_{\lambda_{\mathrm{W}}\lambda_{\mathrm{Z}},\lambda_{\mathrm{W}}\lambda_{\mathrm{Z}}}$ (with $\lambda_V \in \{\mathrm{L},0,\mathrm{R}\}$), yielding the pure 2D polar angular distribution:

$$\begin{aligned}\frac{1}{\sigma_{\mathrm{WZ}\to\ell'^{\pm}\nu_\ell\ell\ell}}\frac{\mathrm{d}^2\,\sigma_{\mathrm{WZ}\to\ell'^{\pm}\nu_\ell\ell\ell}}{\mathrm{d}\,\cos\theta^*_{\mathrm{W}}\mathrm{d}\,\cos\theta^*_{\mathrm{Z}}} = \frac{9}{64}\Big[& \\ & f_{\mathrm{LL}}\,(1\mp\cos\theta^*_{\mathrm{W}})^2\,(1+\cos^2\theta^*_{\mathrm{Z}}+2A_\ell\cos\theta^*_{\mathrm{Z}}) \\ & + f_{\mathrm{RR}}\,(1\pm\cos\theta^*_{\mathrm{W}})^2\,(1+\cos^2\theta^*_{\mathrm{Z}}-2A_\ell\cos\theta^*_{\mathrm{Z}}) \\ & + f_{00}\,4\,(1-\cos^2\theta^*_{\mathrm{W}})\,(1-\cos^2\theta^*_{\mathrm{Z}}) \\ & + f_{\mathrm{L0}}\,(1\mp\cos\theta^*_{\mathrm{W}})^2\,2\,(1-\cos^2\theta^*_{\mathrm{Z}}) \\ & + f_{\mathrm{R0}}\,(1\pm\cos\theta^*_{\mathrm{W}})^2\,2\,(1-\cos^2\theta^*_{\mathrm{Z}}) \\ & + f_{0\mathrm{L}}\,2\,(1-\cos^2\theta^*_{\mathrm{W}})\,(1+\cos^2\theta^*_{\mathrm{Z}}+2A_\ell\cos\theta^*_{\mathrm{Z}}) \\ & + f_{0\mathrm{R}}\,2\,(1-\cos^2\theta^*_{\mathrm{W}})\,(1+\cos^2\theta^*_{\mathrm{Z}}-2A_\ell\cos\theta^*_{\mathrm{Z}}) \\ & + f_{\mathrm{LR}}\,(1\mp\cos\theta^*_{\mathrm{W}})^2\,(1+\cos^2\theta^*_{\mathrm{Z}}-2A_\ell\cos\theta^*_{\mathrm{Z}}) \\ & + f_{\mathrm{RL}}\,(1\pm\cos\theta^*_{\mathrm{W}})^2\,(1+\cos^2\theta^*_{\mathrm{Z}}+2A_\ell\cos\theta^*_{\mathrm{Z}})\Big]\,.\end{aligned} \tag{B.29}$$

To simplify Equation B.29 into the $\{0,\mathrm{T}\}$ formalism, where the transverse states are grouped as $\mathrm{T}=\mathrm{L}+\mathrm{R}$, we must isolate the parity-conserving components. This is achieved experimentally by folding the angular distributions ($\cos\theta^* \to |\cos\theta^*|$), or equivalently by summing symmetrically over the W^+ and W^- channels.

Folding removes every term odd in either angle: the parity-violating terms linear in $\cos\theta^*$ and the doubly odd $A_\ell\cos\theta^*_{\mathrm{W}}\cos\theta^*_{\mathrm{Z}}$ term of the transverse block vanish identically. Consequently, the transverse angular polynomials collapse symmetrically: $(1\mp\cos\theta^*)^2+(1\pm\cos\theta^*)^2 \to 2(1+\cos^2\theta^*)$. Applying this symmetry and combining the respective LH and RH fractions ($f_{0\mathrm{T}}=f_{0\mathrm{L}}+f_{0\mathrm{R}}$, $f_{\mathrm{T}0}=f_{\mathrm{L}0}+f_{\mathrm{R}0}$, $f_{\mathrm{TT}}=f_{\mathrm{LL}}+f_{\mathrm{RR}}+f_{\mathrm{LR}}+f_{\mathrm{RL}}$), the distribution condenses cleanly into:

$$\begin{aligned}\frac{1}{\sigma_{\mathrm{WZ}\to\ell'^{\pm}\nu_\ell\ell\ell}}\frac{\mathrm{d}^2\,\sigma_{\mathrm{WZ}\to\ell'^{\pm}\nu_\ell\ell\ell}}{\mathrm{d}\,\cos\theta^*_{\mathrm{W}}\mathrm{d}\,\cos\theta^*_{\mathrm{Z}}} = \frac{9}{64}\Big[& 4f_{00}\,(1-\cos^2\theta^*_{\mathrm{W}})\,(1-\cos^2\theta^*_{\mathrm{Z}}) \\ & + 2f_{0\mathrm{T}}\,(1-\cos^2\theta^*_{\mathrm{W}})\,(1+\cos^2\theta^*_{\mathrm{Z}}) \\ & + 2f_{\mathrm{T}0}\,(1+\cos^2\theta^*_{\mathrm{W}})\,(1-\cos^2\theta^*_{\mathrm{Z}}) \\ & + f_{\mathrm{TT}}\,(1+\cos^2\theta^*_{\mathrm{W}})\,(1+\cos^2\theta^*_{\mathrm{Z}})\Big]\,.\end{aligned} \tag{B.30}$$

Finally, to observe how the quantum correlations explicitly modify the kinematics, we substitute the expanded definitions of the physical polarisation fractions (Equations B.22 to B.25) into Equation B.30:

$$
\begin{aligned}
\frac{1}{\sigma_{\mathrm{WZ}\to\ell'^{\pm}\nu_\ell\ell\ell}}\frac{\mathrm{d}^2\,\sigma_{\mathrm{WZ}\to\ell'^{\pm}\nu_\ell\ell\ell}}{\mathrm{d}\,\cos\theta^*_{\mathrm{W}}\mathrm{d}\,\cos\theta^*_{\mathrm{Z}}} = \frac{1}{4}
&\\
&- \frac{\sqrt{6}}{16}\langle T_0^{\mathrm{W}}\rangle\left(1-3\cos^2\theta^*_{\mathrm{W}}\right)\\
&- \frac{\sqrt{6}}{16}\langle T_0^{\mathrm{Z}}\rangle\left(1-3\cos^2\theta^*_{\mathrm{Z}}\right)\\
&+ \frac{3}{32}\langle T_0^{\mathrm{W}}\rangle\langle T_0^{\mathrm{Z}}\rangle\left(1-3\cos^2\theta^*_{\mathrm{W}}\right)\left(1-3\cos^2\theta^*_{\mathrm{Z}}\right)\\
&+ \frac{9}{64}\Big[4\left(\boldsymbol{\Sigma}^{00}_{\mathrm{cov}}\right)_{00}(1-\cos^2\theta^*_{\mathrm{W}})(1-\cos^2\theta^*_{\mathrm{Z}})\\
&\quad + 2\left(\boldsymbol{\Sigma}^{0\mathrm{T}}_{\mathrm{cov}}\right)_{0\mathrm{T}}(1-\cos^2\theta^*_{\mathrm{W}})(1+\cos^2\theta^*_{\mathrm{Z}})\\
&\quad + 2\left(\boldsymbol{\Sigma}^{\mathrm{T}0}_{\mathrm{cov}}\right)_{\mathrm{T}0}(1+\cos^2\theta^*_{\mathrm{W}})(1-\cos^2\theta^*_{\mathrm{Z}})\\
&\quad + \left(\boldsymbol{\Sigma}^{\mathrm{TT}}_{\mathrm{cov}}\right)_{\mathrm{TT}}(1+\cos^2\theta^*_{\mathrm{W}})(1+\cos^2\theta^*_{\mathrm{Z}})\Big]\,.
\end{aligned}
\tag{B.31}
$$

Here the shorthand $(\boldsymbol{\Sigma}^{0\mathrm{T}}_{\mathrm{cov}})_{0\mathrm{T}}$, $(\boldsymbol{\Sigma}^{\mathrm{T}0}_{\mathrm{cov}})_{\mathrm{T}0}$, and $(\boldsymbol{\Sigma}^{\mathrm{TT}}_{\mathrm{cov}})_{\mathrm{TT}}$ denotes the correlation sums entering Equations B.23 to B.25, that is $(\boldsymbol{\Sigma}^{00}_{\mathrm{cov}})_{++}+(\boldsymbol{\Sigma}^{00}_{\mathrm{cov}})_{--}$, $(\boldsymbol{\Sigma}^{++}_{\mathrm{cov}})_{00}+(\boldsymbol{\Sigma}^{--}_{\mathrm{cov}})_{00}$, and $\sum_{i,j\in\{+,-\}}(\boldsymbol{\Sigma}^{ii}_{\mathrm{cov}})_{jj}$, respectively. This final form separates the unpolarised production term, the factorised single-boson alignments governed by $(1-3\cos^2\theta^*)$, and the spin correlations between the two bosons, encoded in the $\boldsymbol{\Sigma}$ terms.

$$
\rho^{\mathrm{WZ}} = \begin{pmatrix}
\left(\frac{1}{3}+\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{1}{3}+\frac{\langle S_x^{Z}\rangle}{2}+\frac{\langle T_0^{Z}\rangle}{\sqrt{6}}\right)\mathcal{C}_{++,++} & \left(\frac{1}{3}+\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{\langle S_x^{Z}\rangle-2\langle A_1^{Z}\rangle-\mathrm{i}(\langle S_y^{Z}\rangle-2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{++,+0} & \left(\frac{1}{3}+\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\langle B_1^{Z}\rangle-\mathrm{i}\langle B_2^{Z}\rangle\right)\mathcal{C}_{++,+-} & \left(\frac{1}{3}+\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{\langle S_x^{Z}\rangle-2\langle A_1^{Z}\rangle+\mathrm{i}(\langle S_y^{Z}\rangle-2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{++,0+} & \left(\frac{1}{3}+\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{1}{3}-\frac{2\langle T_0^{Z}\rangle}{\sqrt{6}}\right)\mathcal{C}_{++,00} & \left(\frac{1}{3}+\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{\langle S_x^{Z}\rangle+2\langle A_1^{Z}\rangle-\mathrm{i}(\langle S_y^{Z}\rangle+2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{++,0-} & \left(\frac{1}{3}+\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\langle B_1^{Z}\rangle+\mathrm{i}\langle B_2^{Z}\rangle\right)\mathcal{C}_{++,-+} & \left(\frac{1}{3}+\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{\langle S_x^{Z}\rangle+2\langle A_1^{Z}\rangle+\mathrm{i}(\langle S_y^{Z}\rangle+2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{++,-0} & \left(\frac{1}{3}+\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{1}{3}-\frac{\langle S_x^{Z}\rangle}{2}+\frac{\langle T_0^{Z}\rangle}{\sqrt{6}}\right)\mathcal{C}_{++,--} \\
\left(\frac{\langle S_x^{\mathrm{W}}\rangle-2\langle A_1^{\mathrm{W}}\rangle-\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle-2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\frac{1}{3}+\frac{\langle S_x^{Z}\rangle}{2}+\frac{\langle T_0^{Z}\rangle}{\sqrt{6}}\right)\mathcal{C}_{+0,++} & \left(\frac{\langle S_x^{\mathrm{W}}\rangle-2\langle A_1^{\mathrm{W}}\rangle-\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle-2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\frac{\langle S_x^{Z}\rangle-2\langle A_1^{Z}\rangle-\mathrm{i}(\langle S_y^{Z}\rangle-2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{+0,+0} & \left(\frac{\langle S_x^{\mathrm{W}}\rangle-2\langle A_1^{\mathrm{W}}\rangle-\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle-2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\langle B_1^{Z}\rangle-\mathrm{i}\langle B_2^{Z}\rangle\right)\mathcal{C}_{+0,+-} & \left(\frac{\langle S_x^{\mathrm{W}}\rangle-2\langle A_1^{\mathrm{W}}\rangle-\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle-2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\frac{\langle S_x^{Z}\rangle-2\langle A_1^{Z}\rangle+\mathrm{i}(\langle S_y^{Z}\rangle-2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{+0,0+} & \left(\frac{\langle S_x^{\mathrm{W}}\rangle-2\langle A_1^{\mathrm{W}}\rangle-\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle-2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\frac{1}{3}-\frac{2\langle T_0^{Z}\rangle}{\sqrt{6}}\right)\mathcal{C}_{+0,00} & \left(\frac{\langle S_x^{\mathrm{W}}\rangle-2\langle A_1^{\mathrm{W}}\rangle-\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle-2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\frac{\langle S_x^{Z}\rangle+2\langle A_1^{Z}\rangle-\mathrm{i}(\langle S_y^{Z}\rangle+2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{+0,0-} & \left(\frac{\langle S_x^{\mathrm{W}}\rangle-2\langle A_1^{\mathrm{W}}\rangle-\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle-2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\langle B_1^{Z}\rangle+\mathrm{i}\langle B_2^{Z}\rangle\right)\mathcal{C}_{+0,-+} & \left(\frac{\langle S_x^{\mathrm{W}}\rangle-2\langle A_1^{\mathrm{W}}\rangle-\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle-2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\frac{\langle S_x^{Z}\rangle+2\langle A_1^{Z}\rangle+\mathrm{i}(\langle S_y^{Z}\rangle+2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{+0,-0} & \left(\frac{\langle S_x^{\mathrm{W}}\rangle-2\langle A_1^{\mathrm{W}}\rangle-\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle-2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\frac{1}{3}-\frac{\langle S_x^{Z}\rangle}{2}+\frac{\langle T_0^{Z}\rangle}{\sqrt{6}}\right)\mathcal{C}_{+0,--} \\
\left(\frac{1}{3}+\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{1}{3}+\frac{\langle S_x^{Z}\rangle}{2}+\frac{\langle T_0^{Z}\rangle}{\sqrt{6}}\right)\mathcal{C}_{+-,++} & \left(\frac{1}{3}+\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{\langle S_x^{Z}\rangle-2\langle A_1^{Z}\rangle-\mathrm{i}(\langle S_y^{Z}\rangle-2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{+-,+0} & \left(\frac{1}{3}+\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\langle B_1^{Z}\rangle-\mathrm{i}\langle B_2^{Z}\rangle\right)\mathcal{C}_{+-,+-} & \left(\frac{1}{3}+\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{\langle S_x^{Z}\rangle-2\langle A_1^{Z}\rangle+\mathrm{i}(\langle S_y^{Z}\rangle-2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{+-,0+} & \left(\frac{1}{3}+\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{1}{3}-\frac{2\langle T_0^{Z}\rangle}{\sqrt{6}}\right)\mathcal{C}_{+-,00} & \left(\frac{1}{3}+\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{\langle S_x^{Z}\rangle+2\langle A_1^{Z}\rangle-\mathrm{i}(\langle S_y^{Z}\rangle+2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{+-,0-} & \left(\frac{1}{3}+\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\langle B_1^{Z}\rangle+\mathrm{i}\langle B_2^{Z}\rangle\right)\mathcal{C}_{+-,-+} & \left(\frac{1}{3}+\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{\langle S_x^{Z}\rangle+2\langle A_1^{Z}\rangle+\mathrm{i}(\langle S_y^{Z}\rangle+2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{+-,-0} & \left(\frac{1}{3}+\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{1}{3}-\frac{\langle S_x^{Z}\rangle}{2}+\frac{\langle T_0^{Z}\rangle}{\sqrt{6}}\right)\mathcal{C}_{+-,--} \\
\left(\frac{\langle S_x^{\mathrm{W}}\rangle-2\langle A_1^{\mathrm{W}}\rangle+\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle-2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\frac{1}{3}+\frac{\langle S_x^{Z}\rangle}{2}+\frac{\langle T_0^{Z}\rangle}{\sqrt{6}}\right)\mathcal{C}_{0+,++} & \left(\frac{\langle S_x^{\mathrm{W}}\rangle-2\langle A_1^{\mathrm{W}}\rangle+\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle-2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\frac{\langle S_x^{Z}\rangle-2\langle A_1^{Z}\rangle-\mathrm{i}(\langle S_y^{Z}\rangle-2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{0+,+0} & \left(\frac{\langle S_x^{\mathrm{W}}\rangle-2\langle A_1^{\mathrm{W}}\rangle+\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle-2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\langle B_1^{Z}\rangle-\mathrm{i}\langle B_2^{Z}\rangle\right)\mathcal{C}_{0+,+-} & \left(\frac{\langle S_x^{\mathrm{W}}\rangle-2\langle A_1^{\mathrm{W}}\rangle+\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle-2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\frac{\langle S_x^{Z}\rangle-2\langle A_1^{Z}\rangle+\mathrm{i}(\langle S_y^{Z}\rangle-2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{0+,0+} & \left(\frac{\langle S_x^{\mathrm{W}}\rangle-2\langle A_1^{\mathrm{W}}\rangle+\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle-2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\frac{1}{3}-\frac{2\langle T_0^{Z}\rangle}{\sqrt{6}}\right)\mathcal{C}_{0+,00} & \left(\frac{\langle S_x^{\mathrm{W}}\rangle-2\langle A_1^{\mathrm{W}}\rangle+\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle-2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\frac{\langle S_x^{Z}\rangle+2\langle A_1^{Z}\rangle-\mathrm{i}(\langle S_y^{Z}\rangle+2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{0+,0-} & \left(\frac{\langle S_x^{\mathrm{W}}\rangle-2\langle A_1^{\mathrm{W}}\rangle+\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle-2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\langle B_1^{Z}\rangle+\mathrm{i}\langle B_2^{Z}\rangle\right)\mathcal{C}_{0+,-+} & \left(\frac{\langle S_x^{\mathrm{W}}\rangle-2\langle A_1^{\mathrm{W}}\rangle+\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle-2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\frac{\langle S_x^{Z}\rangle+2\langle A_1^{Z}\rangle+\mathrm{i}(\langle S_y^{Z}\rangle+2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{0+,-0} & \left(\frac{\langle S_x^{\mathrm{W}}\rangle-2\langle A_1^{\mathrm{W}}\rangle+\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle-2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\frac{1}{3}-\frac{\langle S_x^{Z}\rangle}{2}+\frac{\langle T_0^{Z}\rangle}{\sqrt{6}}\right)\mathcal{C}_{0+,--} \\
\left(\frac{1}{3}-\frac{2\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{1}{3}+\frac{\langle S_x^{Z}\rangle}{2}+\frac{\langle T_0^{Z}\rangle}{\sqrt{6}}\right)\mathcal{C}_{00,++} & \left(\frac{1}{3}-\frac{2\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{\langle S_x^{Z}\rangle-2\langle A_1^{Z}\rangle-\mathrm{i}(\langle S_y^{Z}\rangle-2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{00,+0} & \left(\frac{1}{3}-\frac{2\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\langle B_1^{Z}\rangle-\mathrm{i}\langle B_2^{Z}\rangle\right)\mathcal{C}_{00,+-} & \left(\frac{1}{3}-\frac{2\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{\langle S_x^{Z}\rangle-2\langle A_1^{Z}\rangle+\mathrm{i}(\langle S_y^{Z}\rangle-2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{00,0+} & \left(\frac{1}{3}-\frac{2\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{1}{3}-\frac{2\langle T_0^{Z}\rangle}{\sqrt{6}}\right)\mathcal{C}_{00,00} & \left(\frac{1}{3}-\frac{2\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{\langle S_x^{Z}\rangle+2\langle A_1^{Z}\rangle-\mathrm{i}(\langle S_y^{Z}\rangle+2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{00,0-} & \left(\frac{1}{3}-\frac{2\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\langle B_1^{Z}\rangle+\mathrm{i}\langle B_2^{Z}\rangle\right)\mathcal{C}_{00,-+} & \left(\frac{1}{3}-\frac{2\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{\langle S_x^{Z}\rangle+2\langle A_1^{Z}\rangle+\mathrm{i}(\langle S_y^{Z}\rangle+2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{00,-0} & \left(\frac{1}{3}-\frac{2\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{1}{3}-\frac{\langle S_x^{Z}\rangle}{2}+\frac{\langle T_0^{Z}\rangle}{\sqrt{6}}\right)\mathcal{C}_{00,--} \\
\left(\frac{\langle S_x^{\mathrm{W}}\rangle+2\langle A_1^{\mathrm{W}}\rangle-\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle+2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\frac{1}{3}+\frac{\langle S_x^{Z}\rangle}{2}+\frac{\langle T_0^{Z}\rangle}{\sqrt{6}}\right)\mathcal{C}_{0-,++} & \left(\frac{\langle S_x^{\mathrm{W}}\rangle+2\langle A_1^{\mathrm{W}}\rangle-\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle+2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\frac{\langle S_x^{Z}\rangle-2\langle A_1^{Z}\rangle-\mathrm{i}(\langle S_y^{Z}\rangle-2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{0-,+0} & \left(\frac{\langle S_x^{\mathrm{W}}\rangle+2\langle A_1^{\mathrm{W}}\rangle-\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle+2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\langle B_1^{Z}\rangle-\mathrm{i}\langle B_2^{Z}\rangle\right)\mathcal{C}_{0-,+-} & \left(\frac{\langle S_x^{\mathrm{W}}\rangle+2\langle A_1^{\mathrm{W}}\rangle-\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle+2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\frac{\langle S_x^{Z}\rangle-2\langle A_1^{Z}\rangle+\mathrm{i}(\langle S_y^{Z}\rangle-2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{0-,0+} & \left(\frac{\langle S_x^{\mathrm{W}}\rangle+2\langle A_1^{\mathrm{W}}\rangle-\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle+2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\frac{1}{3}-\frac{2\langle T_0^{Z}\rangle}{\sqrt{6}}\right)\mathcal{C}_{0-,00} & \left(\frac{\langle S_x^{\mathrm{W}}\rangle+2\langle A_1^{\mathrm{W}}\rangle-\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle+2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\frac{\langle S_x^{Z}\rangle+2\langle A_1^{Z}\rangle-\mathrm{i}(\langle S_y^{Z}\rangle+2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{0-,0-} & \left(\frac{\langle S_x^{\mathrm{W}}\rangle+2\langle A_1^{\mathrm{W}}\rangle-\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle+2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\langle B_1^{Z}\rangle+\mathrm{i}\langle B_2^{Z}\rangle\right)\mathcal{C}_{0-,-+} & \left(\frac{\langle S_x^{\mathrm{W}}\rangle+2\langle A_1^{\mathrm{W}}\rangle-\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle+2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\frac{\langle S_x^{Z}\rangle+2\langle A_1^{Z}\rangle+\mathrm{i}(\langle S_y^{Z}\rangle+2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{0-,-0} & \left(\frac{\langle S_x^{\mathrm{W}}\rangle+2\langle A_1^{\mathrm{W}}\rangle-\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle+2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\frac{1}{3}-\frac{\langle S_x^{Z}\rangle}{2}+\frac{\langle T_0^{Z}\rangle}{\sqrt{6}}\right)\mathcal{C}_{0-,--} \\
\left(\frac{1}{3}+\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{1}{3}+\frac{\langle S_x^{Z}\rangle}{2}+\frac{\langle T_0^{Z}\rangle}{\sqrt{6}}\right)\mathcal{C}_{-+,++} & \left(\frac{1}{3}+\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{\langle S_x^{Z}\rangle-2\langle A_1^{Z}\rangle-\mathrm{i}(\langle S_y^{Z}\rangle-2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{-+,+0} & \left(\frac{1}{3}+\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\langle B_1^{Z}\rangle-\mathrm{i}\langle B_2^{Z}\rangle\right)\mathcal{C}_{-+,+-} & \left(\frac{1}{3}+\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{\langle S_x^{Z}\rangle-2\langle A_1^{Z}\rangle+\mathrm{i}(\langle S_y^{Z}\rangle-2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{-+,0+} & \left(\frac{1}{3}+\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{1}{3}-\frac{2\langle T_0^{Z}\rangle}{\sqrt{6}}\right)\mathcal{C}_{-+,00} & \left(\frac{1}{3}+\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{\langle S_x^{Z}\rangle+2\langle A_1^{Z}\rangle-\mathrm{i}(\langle S_y^{Z}\rangle+2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{-+,0-} & \left(\frac{1}{3}+\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\langle B_1^{Z}\rangle+\mathrm{i}\langle B_2^{Z}\rangle\right)\mathcal{C}_{-+,-+} & \left(\frac{1}{3}+\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{\langle S_x^{Z}\rangle+2\langle A_1^{Z}\rangle+\mathrm{i}(\langle S_y^{Z}\rangle+2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{-+,-0} & \left(\frac{1}{3}+\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{1}{3}-\frac{\langle S_x^{Z}\rangle}{2}+\frac{\langle T_0^{Z}\rangle}{\sqrt{6}}\right)\mathcal{C}_{-+,--} \\
\left(\frac{\langle S_x^{\mathrm{W}}\rangle+2\langle A_1^{\mathrm{W}}\rangle+\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle+2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\frac{1}{3}+\frac{\langle S_x^{Z}\rangle}{2}+\frac{\langle T_0^{Z}\rangle}{\sqrt{6}}\right)\mathcal{C}_{-0,++} & \left(\frac{\langle S_x^{\mathrm{W}}\rangle+2\langle A_1^{\mathrm{W}}\rangle+\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle+2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\frac{\langle S_x^{Z}\rangle-2\langle A_1^{Z}\rangle-\mathrm{i}(\langle S_y^{Z}\rangle-2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{-0,+0} & \left(\frac{\langle S_x^{\mathrm{W}}\rangle+2\langle A_1^{\mathrm{W}}\rangle+\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle+2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\langle B_1^{Z}\rangle-\mathrm{i}\langle B_2^{Z}\rangle\right)\mathcal{C}_{-0,+-} & \left(\frac{\langle S_x^{\mathrm{W}}\rangle+2\langle A_1^{\mathrm{W}}\rangle+\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle+2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\frac{\langle S_x^{Z}\rangle-2\langle A_1^{Z}\rangle+\mathrm{i}(\langle S_y^{Z}\rangle-2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{-0,0+} & \left(\frac{\langle S_x^{\mathrm{W}}\rangle+2\langle A_1^{\mathrm{W}}\rangle+\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle+2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\frac{1}{3}-\frac{2\langle T_0^{Z}\rangle}{\sqrt{6}}\right)\mathcal{C}_{-0,00} & \left(\frac{\langle S_x^{\mathrm{W}}\rangle+2\langle A_1^{\mathrm{W}}\rangle+\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle+2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\frac{\langle S_x^{Z}\rangle+2\langle A_1^{Z}\rangle-\mathrm{i}(\langle S_y^{Z}\rangle+2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{-0,0-} & \left(\frac{\langle S_x^{\mathrm{W}}\rangle+2\langle A_1^{\mathrm{W}}\rangle+\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle+2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\langle B_1^{Z}\rangle+\mathrm{i}\langle B_2^{Z}\rangle\right)\mathcal{C}_{-0,-+} & \left(\frac{\langle S_x^{\mathrm{W}}\rangle+2\langle A_1^{\mathrm{W}}\rangle+\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle+2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\frac{\langle S_x^{Z}\rangle+2\langle A_1^{Z}\rangle+\mathrm{i}(\langle S_y^{Z}\rangle+2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{-0,-0} & \left(\frac{\langle S_x^{\mathrm{W}}\rangle+2\langle A_1^{\mathrm{W}}\rangle+\mathrm{i}(\langle S_y^{\mathrm{W}}\rangle+2\langle A_2^{\mathrm{W}}\rangle)}{2\sqrt{2}}\right)\times\left(\frac{1}{3}-\frac{\langle S_x^{Z}\rangle}{2}+\frac{\langle T_0^{Z}\rangle}{\sqrt{6}}\right)\mathcal{C}_{-0,--} \\
\left(\frac{1}{3}-\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{1}{3}+\frac{\langle S_x^{Z}\rangle}{2}+\frac{\langle T_0^{Z}\rangle}{\sqrt{6}}\right)\mathcal{C}_{--,++} & \left(\frac{1}{3}-\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{\langle S_x^{Z}\rangle-2\langle A_1^{Z}\rangle-\mathrm{i}(\langle S_y^{Z}\rangle-2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{--,+0} & \left(\frac{1}{3}-\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\langle B_1^{Z}\rangle-\mathrm{i}\langle B_2^{Z}\rangle\right)\mathcal{C}_{--,+-} & \left(\frac{1}{3}-\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{\langle S_x^{Z}\rangle-2\langle A_1^{Z}\rangle+\mathrm{i}(\langle S_y^{Z}\rangle-2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{--,0+} & \left(\frac{1}{3}-\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{1}{3}-\frac{2\langle T_0^{Z}\rangle}{\sqrt{6}}\right)\mathcal{C}_{--,00} & \left(\frac{1}{3}-\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{\langle S_x^{Z}\rangle+2\langle A_1^{Z}\rangle-\mathrm{i}(\langle S_y^{Z}\rangle+2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{--,0-} & \left(\frac{1}{3}-\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\langle B_1^{Z}\rangle+\mathrm{i}\langle B_2^{Z}\rangle\right)\mathcal{C}_{--,-+} & \left(\frac{1}{3}-\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{\langle S_x^{Z}\rangle+2\langle A_1^{Z}\rangle+\mathrm{i}(\langle S_y^{Z}\rangle+2\langle A_2^{Z}\rangle)}{2\sqrt{2}}\right)\mathcal{C}_{--,-0} & \left(\frac{1}{3}-\frac{\langle S_x^{\mathrm{W}}\rangle}{2}+\frac{\langle T_0^{\mathrm{W}}\rangle}{\sqrt{6}}\right)\times\left(\frac{1}{3}-\frac{\langle S_x^{Z}\rangle}{2}+\frac{\langle T_0^{Z}\rangle}{\sqrt{6}}\right)\mathcal{C}_{--,--}
\end{pmatrix}. \tag{B.32}
$$

C Polarised vector boson Feynman diagrams

In Equation 1.91 we presented the full Lagrangian of the SM. In this appendix, we will focus on the EW sector of the SM and in particular the vector boson interactions with fermions through polarisation using the Feynman rules. For interactions of the vector bosons with the Higgs boson or with themselves, we refer the reader to Appendix A in which we compute the cross-section of the VBS process in the GBET limit.

The relevant terms of the SM Lagrangian for our discussion are the following:

$$\mathcal{L}_{\mathrm{EW}} \supset \overline{\varphi}_{\mathrm{L}}\gamma^{\mu}\left(\mathrm{i}\partial_{\mu} - g_{\mathrm{Y}}\frac{Y}{2}B_{\mu} - g_{\mathrm{L}}\frac{\sigma^{i}}{2}W_{\mu}^{i}\right)\varphi_{\mathrm{L}} + \overline{\varphi}_{\mathrm{R}}\gamma^{\mu}\left(\mathrm{i}\partial_{\mu} - g_{\mathrm{Y}}\frac{Y}{2}B_{\mu}\right)\varphi_{\mathrm{R}}\,, \tag{C.1}$$

with $\varphi_{\mathrm{L}} = \begin{pmatrix}\nu_{\mathrm{L}}\\ \ell_{\mathrm{L}}\end{pmatrix}$ and $\varphi_{\mathrm{R}} = \ell_{\mathrm{R}}$ for the leptons. The interactions of the vector bosons with the fermions are given by the terms proportional to g_{Y} and g_{L}.

We can rewrite the above Lagrangian to express the interaction terms only:

$$\mathcal{L}_{\mathrm{EW}} \supset -g_{\mathrm{Y}}\overline{\varphi}_{\mathrm{L}}\gamma^{\mu}\frac{Y}{2}B_{\mu}\varphi_{\mathrm{L}} - g_{\mathrm{L}}\overline{\varphi}_{\mathrm{L}}\gamma^{\mu}\frac{\sigma^{i}}{2}W_{\mu}^{i}\varphi_{\mathrm{L}} - g_{\mathrm{Y}}\overline{\varphi}_{\mathrm{R}}\gamma^{\mu}\frac{Y}{2}B_{\mu}\varphi_{\mathrm{R}}\,. \tag{C.2}$$

C.1 Charged current

The $\mathrm{SU}(2)_{\mathrm{L}}$ part of Equation C.2:

$$\mathcal{L}_{\mathrm{EW}} \supset -g_{\mathrm{L}}\overline{\varphi}_{\mathrm{L}}\gamma^{\mu}\left(\frac{\sigma^{1}}{2}W_{\mu}^{1} + \frac{\sigma^{2}}{2}W_{\mu}^{2} + \frac{\sigma^{3}}{2}W_{\mu}^{3}\right)\varphi_{\mathrm{L}}\,, \tag{C.3}$$

can be rewritten using the physical W bosons defined in Equation 1.70. The W_{μ}^{1} and W_{μ}^{2} therefore mix the LH neutrino and charged lepton fields, and we can identify the charged-current interactions:

$$\mathcal{L}_{\mathrm{CC}} = -\frac{g_{\mathrm{L}}}{\sqrt{2}}\left(\overline{\nu}_{\mathrm{L}}\gamma^{\mu}\ell_{\mathrm{L}}\mathrm{W}_{\mu}^{+} + \overline{\ell}_{\mathrm{L}}\gamma^{\mu}\nu_{\mathrm{L}}\mathrm{W}_{\mu}^{-}\right)\,. \tag{C.4}$$

Using the chirality projectors of Equation 1.27, we can rewrite the above Lagrangian as:

$$\mathcal{L}_{\mathrm{CC}} = -\frac{g_{\mathrm{L}}}{\sqrt{2}}\left(\overline{\nu}\gamma^{\mu}P_{\mathrm{L}}\ell\mathrm{W}_{\mu}^{+} + \overline{\ell}\gamma^{\mu}P_{\mathrm{L}}\nu\mathrm{W}_{\mu}^{-}\right)\,. \tag{C.5}$$

This defines the Feynman rules for the charged-current interactions of the W bosons with the leptons.

C.2 Neutral current

The neutral part under $\mathrm{SU}(2)_{\mathrm{L}} \times \mathrm{U}(1)_{\mathrm{Y}}$ of Equation C.2:

$$\mathcal{L}_{\mathrm{EW}} \supset -g_{\mathrm{Y}}\overline{\varphi}_{\mathrm{L}}\gamma^{\mu}\frac{Y}{2}B_{\mu}\varphi_{\mathrm{L}} - g_{\mathrm{Y}}\overline{\varphi}_{\mathrm{R}}\gamma^{\mu}\frac{Y}{2}B_{\mu}\varphi_{\mathrm{R}} - g_{\mathrm{L}}\overline{\varphi}_{\mathrm{L}}\gamma^{\mu}\frac{\sigma^{3}}{2}W_{\mu}^{3}\varphi_{\mathrm{L}}\,, \tag{C.6}$$

can also be rewritten using the physical Z boson defined in Equation 1.71. The W_μ^3 and B_μ fields therefore mix the LH neutrino and charged lepton fields, and we can identify the NC interactions:

$$\mathcal{L}_{\text{NC}} = -\sqrt{g_{\text{L}}^2 + g_{\text{Y}}^2}\left(\overline{\nu}_{\text{L}}\gamma^\mu \frac{1}{2}\nu_{\text{L}} + \overline{\ell}_{\text{L}}\gamma^\mu\left(-\frac{1}{2} + \sin^2\theta_{\text{W}}\right)\ell_{\text{L}} + \overline{\ell}_{\text{R}}\gamma^\mu \sin^2\theta_{\text{W}}\ell_{\text{R}}\right)\text{Z}_\mu\,, \tag{C.7}$$

where we have used the definition of the weak mixing angle θ_{W}:

$$\sin^2\theta_{\text{W}} = \frac{g_{\text{Y}}^2}{g_{\text{L}}^2 + g_{\text{Y}}^2}\,. \tag{C.8}$$

Using the chirality projectors of Equation 1.27, we can rewrite the above Lagrangian as:

$$\mathcal{L}_{\text{NC}} = -\sqrt{g_{\text{L}}^2 + g_{\text{Y}}^2}\left(\overline{\nu}\gamma^\mu P_{\text{L}}\frac{1}{2}\nu + \overline{\ell}\gamma^\mu\left(P_{\text{L}}\left(-\frac{1}{2} + \sin^2\theta_{\text{W}}\right) + P_{\text{R}}\sin^2\theta_{\text{W}}\right)\ell\right)\text{Z}_\mu\,. \tag{C.9}$$

This defines the Feynman rules for the NC interactions of the Z boson with the leptons.

The lepton sector of $\mathcal{L}_{\text{NC}}$ can be rewritten more compactly by identifying the chiral coupling coefficients directly from the Lagrangian used in Equation 6.78 so that the $\ell\bar{\ell}\text{Z}$ interaction term reads:

$$\mathcal{L}_{\ell\text{Z}} = -\sqrt{g_{\text{L}}^2 + g_{\text{Y}}^2}\,\overline{\ell}\,\gamma^\mu\left(g_{\text{L}}^\ell P_{\text{L}} + g_{\text{R}}^\ell P_{\text{R}}\right)\ell\,\text{Z}_\mu\,, \tag{C.10}$$

by substituting the chirality projectors:

$$g_{\text{L}}^\ell P_{\text{L}} + g_{\text{R}}^\ell P_{\text{R}} = \frac{g_{\text{L}}^\ell + g_{\text{R}}^\ell}{2} - \frac{g_{\text{L}}^\ell - g_{\text{R}}^\ell}{2}\gamma_5 \equiv g_{\text{V}}^\ell - g_{\text{A}}^\ell\,\gamma_5\,. \tag{C.11}$$

This identification naturally introduces the coupling-asymmetry parameter A_ℓ, defined in Equation 6.79, which measures the relative imbalance between the LH and RH couplings of the Z boson to charged leptons. The equivalence between the NC Lagrangian and the CC Lagrangian is recovered in the limit $A_\ell \to 1$ (or equivalently $g_{\text{R}}^\ell \to 0$), which corresponds to the case where the Z boson couples only to LH leptons, as is the case for the W bosons.

C.3 Vector boson decay

Now that we have identified the Feynman rules for the interactions of the vector bosons with the leptons, we can use them to compute the ME for the decay of a vector boson V into two leptons.

We consider the helicity basis where the polarisation four-vectors are defined in Equation 1.98 and Equation 1.102 for a vector boson propagating along the z-axis. We therefore choose the z-axis along the direction of motion of the decaying boson, with the x-axis and y-axis defined in the transverse plane, chosen such that the leptons decay in the xOz plane. Finally, we consider the rest frame of the decaying boson. Under these assumptions, the four-momenta of the two leptons can be written as:

$$p_1^\mu = (E_1^*,\, p^*\sin\theta_\ell^*,\, 0,\, p^*\cos\theta_\ell^*)\,, \qquad p_2^\mu = (E_2^*,\, -p^*\sin\theta_\ell^*,\, 0,\, -p^*\cos\theta_\ell^*)\,, \tag{C.12}$$

where E^* and p^* are respectively the energy and the norm of the momentum in the rest frame of the decaying boson, and θ_ℓ^* is the decay angle of the fermion with respect to the z-axis. Given that the lepton masses are negligible compared to the mass of the decaying boson, we can consider them as massless and therefore $E_1^* = E_2^* = p^* = \frac{m_V}{2}$, with m_V the mass of the decaying boson.

The interaction between the vector boson V and the leptons is described by the CC Lagrangian $\mathcal{L}_{\text{CC}}$ for the W boson, and the NC Lagrangian $\mathcal{L}_{\text{NC}}$ for the Z boson, as defined in Equations C.5 and C.9.

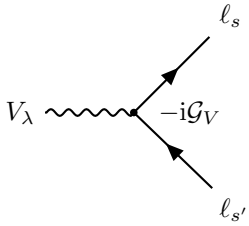


Figure C.1: Feynman diagram representing the decay of a vector boson V into two leptons. The polarisation of the vector boson is denoted by λ, and the spins of the leptons are denoted by s and s'.

The ME for a vector boson V with polarisation λ, giving two leptons of spin s and s', represented in Figure C.1, can be written in a unified form as:

$$\mathrm{i}\mathcal{M}_\lambda = -\mathrm{i}\,\mathcal{G}_V\,\varepsilon_\lambda^\mu\,\bar{u}_s(p_1)\,\gamma_\mu\,(g_\mathrm{V} - g_\mathrm{A}\,\gamma_5)\,v_{s'}(p_2)\,, \tag{C.13}$$

where the boson-dependent parameters are:

$$\mathrm{W}: \quad \mathcal{G}_\mathrm{W} = \frac{g_\mathrm{L}}{\sqrt{2}}\,, \quad g_\mathrm{V} = g_\mathrm{A} = \frac{1}{2}\,, \qquad\qquad \mathrm{Z}: \quad \mathcal{G}_\mathrm{Z} = \sqrt{g_\mathrm{L}^2 + g_\mathrm{Y}^2}\,, \quad g_\mathrm{V} = g_\mathrm{V}^\ell\,, \quad g_\mathrm{A} = g_\mathrm{A}^\ell\,. \tag{C.14}$$

Note that the W case corresponds to $g_\mathrm{V} = g_\mathrm{A} = \frac{1}{2}$, so that $g_\mathrm{V} - g_\mathrm{A}\gamma_5 = P_\mathrm{L}$, recovering the pure V−A structure of Equation C.5. The Z case with $g_\mathrm{V}^\ell \neq g_\mathrm{A}^\ell$ reflects the partial breaking of this symmetry measured by A_ℓ.

Summing over fermion spins and using the completeness relations $\sum_s u_s\bar{u}_s = \not{p} + m$ and $\sum_s v_s\bar{v}_s = \not{p} - m$:

$$\begin{aligned}
|\mathcal{M}_\lambda|^2 &= \mathcal{G}_V^2 \sum_{s,s'} \varepsilon_\lambda^\mu\,\bar{u}_s(p_1)\,\gamma_\mu(g_\mathrm{V} - g_\mathrm{A}\gamma_5)\,v_{s'}(p_2)\,\varepsilon_\lambda^{*\nu}\,\bar{v}_{s'}(p_2)\,\gamma_\nu(g_\mathrm{V} - g_\mathrm{A}\gamma_5)\,u_s(p_1) \\
&= \mathcal{G}_V^2\,\varepsilon_\lambda^\mu\,\varepsilon_\lambda^{*\nu}\,\mathrm{Tr}\left[(\not{p}_1 + m_1)\,\gamma_\mu(g_\mathrm{V} - g_\mathrm{A}\gamma_5)\,(\not{p}_2 - m_2)\,\gamma_\nu(g_\mathrm{V} - g_\mathrm{A}\gamma_5)\right]\,.
\end{aligned} \tag{C.15}$$

Neglecting the fermion masses and expanding using the γ_5 trace identities:

$$\begin{aligned}
|\mathcal{M}_\lambda|^2 &= \mathcal{G}_V^2\,\varepsilon_\lambda^\mu\,\varepsilon_\lambda^{*\nu}\Big[\big((g_\mathrm{V})^2 + (g_\mathrm{A})^2\big)\,\mathrm{Tr}(\not{p}_1\gamma_\mu\not{p}_2\gamma_\nu) - 2g_\mathrm{V}g_\mathrm{A}\,\mathrm{Tr}(\not{p}_1\gamma_\mu\not{p}_2\gamma_\nu\gamma_5)\Big] \\
&= \mathcal{G}_V^2\Bigg[4\big((g_\mathrm{V})^2 + (g_\mathrm{A})^2\big)\,\varepsilon_\lambda^\mu\,\varepsilon_\lambda^{*\nu}\,p_1^\alpha p_2^\beta(g_{\alpha\mu}g_{\beta\nu} + g_{\alpha\nu}g_{\beta\mu} - g_{\alpha\beta}g_{\mu\nu}) \\
&\qquad - 8\mathrm{i}\,g_\mathrm{V}g_\mathrm{A}\,\varepsilon_\lambda^\mu\,\varepsilon_\lambda^{*\nu}\,p_1^\alpha\,p_2^\beta\,\epsilon_{\mu\alpha\nu\beta}\Bigg]\,.
\end{aligned} \tag{C.16}$$

Collecting the Lorentz contractions:

$$\begin{aligned}
|\mathcal{M}_\lambda|^2 = 4\mathcal{G}_V^2\big((g_\mathrm{V})^2 + (g_\mathrm{A})^2\big)\Big[(\varepsilon_\lambda\cdot p_1)(\varepsilon_\lambda^*\cdot p_2) + (\varepsilon_\lambda\cdot p_2)(\varepsilon_\lambda^*\cdot p_1) - (\varepsilon_\lambda\cdot\varepsilon_\lambda^*)(p_2\cdot p_1)\Big] \\
- 8\mathrm{i}\,\mathcal{G}_V^2\,g_\mathrm{V}g_\mathrm{A}\,\varepsilon_\lambda^\mu\,\varepsilon_\lambda^{*\nu}\,p_1^\alpha\,p_2^\beta\,\epsilon_{\mu\alpha\nu\beta}\,.
\end{aligned} \tag{C.17}$$

Inserting the explicit polarisation four-vectors and four-momenta:

$$(\varepsilon_0\cdot p_1)(\varepsilon_0^*\cdot p_2) + (\varepsilon_0\cdot p_2)(\varepsilon_0^*\cdot p_1) - (\varepsilon_0\cdot\varepsilon_0^*)(p_2\cdot p_1) = 2p^{*2}\sin^2\theta_\ell^* = 2p^{*2}(1-\cos^2\theta_\ell^*)\,, \tag{C.18}$$

$$(\varepsilon_\pm\cdot p_1)(\varepsilon_\pm^*\cdot p_2) + (\varepsilon_\pm\cdot p_2)(\varepsilon_\pm^*\cdot p_1) - (\varepsilon_\pm\cdot\varepsilon_\pm^*)(p_2\cdot p_1) = p^{*2}(1+\cos^2\theta_\ell^*)\,, \tag{C.19}$$

and for the antisymmetric contractions:

$$\mathrm{i}\,\varepsilon_0^\mu\,\varepsilon_0^{*\nu}\,p_1^\alpha\,p_2^\beta\,\epsilon_{\mu\alpha\nu\beta} = 0\,, \tag{C.20}$$

$$\begin{aligned}
\mathrm{i}\,\varepsilon_\pm^\mu\,\varepsilon_\pm^{*\nu}\,p_1^\alpha\,p_2^\beta\,\epsilon_{\mu\alpha\nu\beta} &= \frac{\mathrm{i}}{2}\Big[\epsilon_{1\alpha2\beta}\,(\mp\mathrm{i})\,p_1^\alpha p_2^\beta + \epsilon_{2\alpha1\beta}\,(\pm\mathrm{i})\,p_1^\alpha p_2^\beta\Big] \\
&= \frac{\mathrm{i}}{2}\Big[-(\mp\mathrm{i})\,(p_1^0p_2^3 - p_1^3p_2^0) + (\pm\mathrm{i})\,(p_1^0p_2^3 - p_1^3p_2^0)\Big] \\
&= \frac{\mathrm{i}}{2}\Big(\mp 4\mathrm{i}\,p^{*2}\cos\theta_\ell^*\Big) = \pm\,2p^{*2}\cos\theta_\ell^*\,.
\end{aligned} \tag{C.21}$$

It then follows that:

$$|\mathcal{M}_0|^2 = 8\mathcal{G}_V^2\,\big((g_\mathrm{V})^2 + (g_\mathrm{A})^2\big)\,p^{*2}\Big(1 - \cos^2\theta_\ell^*\Big)\,, \tag{C.22}$$

$$|\mathcal{M}_\pm|^2 = 4\mathcal{G}_V^2\,\big((g_\mathrm{V})^2 + (g_\mathrm{A})^2\big)\,p^{*2}\,\big(1 + \cos^2\theta_\ell^* \mp 2A_\ell\cos\theta_\ell^*\big)\,. \tag{C.23}$$

Integrating over $\cos\theta_\ell^* \in [-1,+1]$ yields the same total rate for all three polarisation states, confirming that the total decay rate is independent of polarisation. The normalised angular distributions are:

$$\frac{1}{\sigma_{\lambda=0}}\frac{\mathrm{d}\,\sigma_{\lambda=0}}{\mathrm{d}\,\cos\theta_\ell^*} = \frac{3}{4}\Big(1 - \cos^2\theta_\ell^*\Big)\,, \tag{C.24}$$

$$\frac{1}{\sigma_{\lambda=\pm}}\frac{\mathrm{d}\,\sigma_{\lambda=\pm}}{\mathrm{d}\,\cos\theta_\ell^*} = \frac{3}{8}\,\big(1 + \cos^2\theta_\ell^* \mp 2A_\ell\cos\theta_\ell^*\big)\,. \tag{C.25}$$

For the W boson, $g_\mathrm{V} = g_\mathrm{A} = \frac{1}{2}$, so that the coupling-asymmetry parameter is $A_\ell = +1$ for the W^- (with ℓ^- as the angular reference) and $A_\ell = -1$ for the W^+ (with ℓ^+ as the angular reference). We can see that we recover exactly the expression given in Equation 6.84 for the W boson, where $\pm$ denotes the W charge (not the helicity), and the expression given in Equation 6.86 for the Z boson.

C.4 Decay widths and branching fractions

For a spin-1 boson of mass m_V decaying into two massless leptons, the LIPS two-body element in the rest frame, after integrating over the recoil momentum using the energy-momentum conservation delta function, reduces to:

$$\mathrm{d}\,\Phi_2 = \frac{\mathrm{d}\,\Omega}{32\pi^2} = \frac{2\pi\,\mathrm{d}\,\cos\theta_\ell^*}{32\pi^2} = \frac{\mathrm{d}\,\cos\theta_\ell^*}{16\pi}\,, \tag{C.26}$$

where the 2π azimuthal integral is trivial because the squared MEs of Equations C.22 and C.23 carry no ϕ dependence in the chosen frame. The partial decay width for a boson in helicity state λ is then defined by:

$$\Gamma_\lambda = \frac{1}{2m_V}\int \sum_{s,s'} |\mathcal{M}_\lambda|^2 \mathrm{d}\,\Phi_2 = \frac{1}{32\pi m_V}\int_{-1}^{+1} |\mathcal{M}_\lambda|^2 \mathrm{d}\,\cos\theta_\ell^*\,. \tag{C.27}$$

Inserting the explicit expressions from Equations C.22 and C.23, with $p^* = \frac{m_V}{2}$ for massless decay products, and using:

$$\int_{-1}^{+1}(1-\cos^2\theta_\ell^*)\mathrm{d}\,\cos\theta_\ell^* = \frac{4}{3}\,, \qquad \int_{-1}^{+1}(1+\cos^2\theta_\ell^*)\mathrm{d}\,\cos\theta_\ell^* = \frac{8}{3}\,, \qquad \int_{-1}^{+1}\cos\theta_\ell^*\,\mathrm{d}\,\cos\theta_\ell^* = 0\,, \tag{C.28}$$

one finds:

$$\int_{-1}^{+1} |\mathcal{M}_0|^2 \mathrm{d}\,\cos\theta_\ell^* = 8\mathcal{G}_V^2\left[(g_\mathrm{V})^2+(g_\mathrm{A})^2\right]\frac{m_V^2}{4}\cdot\frac{4}{3} = \frac{8m_V^2}{3}\mathcal{G}_V^2\left[(g_\mathrm{V})^2+(g_\mathrm{A})^2\right], \tag{C.29}$$

$$\int_{-1}^{+1} |\mathcal{M}_\pm|^2 \mathrm{d}\,\cos\theta_\ell^* = 4\mathcal{G}_V^2\left[(g_\mathrm{V})^2+(g_\mathrm{A})^2\right]\frac{m_V^2}{4}\cdot\frac{8}{3} = \frac{8m_V^2}{3}\mathcal{G}_V^2\left[(g_\mathrm{V})^2+(g_\mathrm{A})^2\right]. \tag{C.30}$$

The linear $\cos\theta_\ell^*$ term in $|\mathcal{M}_\pm|^2$ vanishes upon integration, as expected from parity-integrated quantities. Substituting Equations C.29 and C.30 into Equation C.27 gives the same result for all three helicity states:

$$\Gamma_\lambda = \frac{1}{32\pi m_V}\cdot\frac{8m_V^2}{3}\mathcal{G}_V^2\left[(g_\mathrm{V})^2+(g_\mathrm{A})^2\right] = \frac{m_V\mathcal{G}_V^2\left[(g_\mathrm{V})^2+(g_\mathrm{A})^2\right]}{12\pi}\,. \tag{C.31}$$

This equality, which holds for any combination of vector and axial couplings, is the explicit confirmation that the partial width $\Gamma_\lambda = \Gamma(V\to\ell\ell')$ is the same for the three helicity states by rotational invariance, as stated before Equation 6.116.

The coupling constants $\mathcal{G}_\mathrm{W} = \frac{g_\mathrm{L}}{\sqrt{2}}$ and $\mathcal{G}_\mathrm{Z} = \sqrt{g_\mathrm{L}^2+g_\mathrm{Y}^2}$ can be traded for the usual Fermi constant G_F [412], which is defined from the measured muon decay rate via the low-energy effective four-fermion interaction obtained by integrating out the W propagator:

$$\frac{G_\mathrm{F}}{\sqrt{2}} = \frac{g_\mathrm{L}^2}{8m_\mathrm{W}^2}\,. \tag{C.32}$$

For the W boson, where $g_\mathrm{V} = g_\mathrm{A} = \frac{1}{2}$, one has $(g_\mathrm{V})^2+(g_\mathrm{A})^2 = \frac{1}{2}$ and $\mathcal{G}_\mathrm{W}^2 = \frac{g_\mathrm{L}^2}{2} = \frac{4\sqrt{2}G_\mathrm{F}m_\mathrm{W}^2}{2} = 2\sqrt{2}G_\mathrm{F}m_\mathrm{W}^2$. Substituting into Equation C.31:

$$\Gamma(\mathrm{W}\to\ell\nu_\ell) = \frac{m_\mathrm{W}\cdot 2\sqrt{2}G_\mathrm{F}m_\mathrm{W}^2\cdot\frac{1}{2}}{12\pi} = \frac{\sqrt{2}G_\mathrm{F}m_\mathrm{W}^3}{12\pi} = \frac{G_\mathrm{F}m_\mathrm{W}^3}{6\pi\sqrt{2}} \approx 227\,\mathrm{MeV}\ (226\,\mathrm{MeV}\ \text{measured})\,. \tag{C.33}$$

For the Z boson, the analogous relation $\frac{G_\mathrm{F}}{\sqrt{2}} = \frac{g_\mathrm{Z}^2}{8m_\mathrm{Z}^2}$ holds at tree level, where $\mathcal{G}_\mathrm{Z} = g_\mathrm{Z}$ so $\mathcal{G}_\mathrm{Z}^2 = g_\mathrm{Z}^2 = 4\sqrt{2}G_\mathrm{F}m_\mathrm{Z}^2$. Substituting into Equation C.31 and using $(g_\mathrm{V}^\ell)^2+(g_\mathrm{A}^\ell)^2 = \frac{(g_\mathrm{L}^\ell)^2+(g_\mathrm{R}^\ell)^2}{2}$:

$$\Gamma(\mathrm{Z}\to\ell^+\ell^-) = \frac{\sqrt{2}G_\mathrm{F}m_\mathrm{Z}^3}{3\pi}\left[(g_\mathrm{V}^\ell)^2+(g_\mathrm{A}^\ell)^2\right] = \frac{G_\mathrm{F}m_\mathrm{Z}^3}{3\pi\sqrt{2}}\left[(g_\mathrm{L}^\ell)^2+(g_\mathrm{R}^\ell)^2\right] \approx 84\,\mathrm{MeV}\,. \tag{C.34}$$

The numerical suppression relative to Equation C.33 arises from two competing effects: the larger Z mass increases the rate, but the weak isospin–hypercharge combination gives $(g_\mathrm{L}^\ell)^2+(g_\mathrm{R}^\ell)^2 = \left(-\frac{1}{2}+\sin^2\theta_\mathrm{W}\right)^2+\left(\sin^2\theta_\mathrm{W}\right)^2 \approx 0.126$ for $\sin^2\theta_\mathrm{W}\approx 0.231$, which reduces the effective coupling strength below that of the W.

The leptonic branching fractions follow from comparing the partial widths of Equations C.33 and C.34 to the respective total widths. The W boson decays into three leptonic families and two hadronic doublets (u,d') and (c,s'), each with a colour factor of three, giving $\Gamma_\mathrm{W} \approx \left(3+6(1+\frac{\alpha_\mathrm{S}}{\pi})\right)\Gamma(\mathrm{W}\to\ell\nu_\ell) \approx 9\Gamma(\mathrm{W}\to\ell\nu_\ell)$ up to small QCD and EW corrections, hence $\mathrm{BR}(\mathrm{W}\to\ell\nu_\ell)\approx\frac{1}{9}\approx 11\,\%$. The total Z width receives contributions from all five kinematically accessible quark pairs, three charged-lepton pairs, and three neutrino pairs, yielding $\Gamma_\mathrm{Z}\approx 2.497\,\mathrm{GeV}$. The measured per-generation branching fractions are:

$$\mathrm{BR}(\mathrm{W}\to\ell\nu_\ell)\approx 10.86\,\%\,, \qquad \mathrm{BR}(\mathrm{Z}\to\ell^+\ell^-)\approx 3.37\,\%\,. \tag{C.35}$$

D One-loop leading-logarithm corrections

Starting from the eikonal integral of Equation 6.131, after integration over q and summation over all external legs, the complete one-loop NLO EW correction to $\mathrm{W}\mathrm{Z}\mathrm{j}\mathrm{j}$-EW can be decomposed, following the formalism of Denner–Pozzorini [410–412], as:

$$\delta^{\mathrm{EW}}_{\mathrm{DP}} = \delta^{\mathrm{LSC}} + \delta^{\mathrm{SSC}} + \delta^{\mathrm{C}} + \delta^{\mathrm{PR}} + \delta^{\mathrm{EM}} , \tag{D.1}$$

which retains every logarithmically enhanced one-loop term and neglects only the non-logarithmic remainder of $\mathcal{O}(\alpha_{\mathrm{EW}})$ and mass-suppressed $\mathcal{O}\left(\frac{m_V^2}{Q^2}\right)$ terms. Each contribution is a correction factor on the Born amplitude, fixed entirely by the gauge quantum numbers of the external legs. Before introducing them we define:

$$L = \ln\left(\frac{Q^2}{m_V^2}\right) , \qquad r_{kl} = (p_k + p_l)^2 , \tag{D.2}$$

where r_{kl} are the invariants of the external legs k and l.

We keep the dominant *genuine-EW* logarithms and quote the pure-EM and m_{Z}–m_{W} mass-gap pieces only where relevant. The EW Casimir operator of leg k is defined by:

$$C_k^{\mathrm{EW}} = \frac{1}{\sin^2\theta_{\mathrm{W}}} C_2^{\mathrm{SU(2)_L}}(k) + \frac{1}{\cos^2\theta_{\mathrm{W}}}\left(\frac{Y_k}{2}\right)^2 , \qquad C_2^{\mathrm{SU(2)_L}}(k) = T_k(T_k+1) , \tag{D.3}$$

with $C_2^{\mathrm{SU(2)_L}} = 2$ for the $\mathrm{SU}(2)_{\mathrm{L}}$ triplet (transverse gauge bosons) and $\frac{3}{4}$ for a doublet (fermions and the scalars describing longitudinal bosons). Its explicit values for the external states of $\mathrm{W}\mathrm{Z}\mathrm{j}\mathrm{j}$-EW are collected in Table D.A. A direct analogy can be drawn with the strong Casimir operators $C_{\mathrm{F}} = \left.\frac{N_{\mathrm{C}}^2-1}{2N_{\mathrm{C}}}\right|_{N_{\mathrm{C}}=3} = \frac{4}{3}$ and $C_{\mathrm{A}} = N_{\mathrm{C}} = 3$ (the fundamental and adjoint representations governing QCD radiation), but built from the $\mathrm{SU}(2)_{\mathrm{L}} \times \mathrm{U}(1)_{\mathrm{Y}}$ quantum numbers and weighted by the weak mixing angle θ_{W}. The crucial point for the polarisation is that this coefficient is different for transverse and for longitudinal bosons. The transverse W and Z are gauge fields living in the $\mathrm{SU}(2)_{\mathrm{L}}$ adjoint (isospin $T = 1$, $C_2^{\mathrm{SU(2)_L}} = 2$), whereas a longitudinal boson couples, through the GBET of Section 1.11, as the corresponding Goldstone component of the Higgs doublet (isospin $T = \frac{1}{2}$, $C_2^{\mathrm{SU(2)_L}} = \frac{3}{4}$). The smaller Casimir of the scalar doublet means the leading Sudakov double logarithms, and hence the negative NLO EW corrections, are intrinsically *smaller in magnitude for longitudinal than for transverse bosons*.

Leading soft–collinear, δ^{LSC}. The double logarithms arise from a soft and collinear gauge boson exchanged between pairs of legs. Using global $\mathrm{SU}(2)_{\mathrm{L}} \times \mathrm{U}(1)_{\mathrm{Y}}$ invariance of the S-matrix, $\sum_k I_k^V \mathcal{M}_0 = 0$, the sum over pairs collapses to a single sum over legs:

$$\delta^{\mathrm{LSC}} = -\frac{\alpha_{\mathrm{EW}}}{8\pi} \sum_k \left[C_k^{\mathrm{EW}}\, L^2 - 2\,(I_k^{\mathrm{Z}})^2\, \ln\frac{m_{\mathrm{Z}}^2}{m_{\mathrm{W}}^2}\, L \right] . \tag{D.4}$$

The first term is the leading double log, the second is a subleading single log generated by the W–Z mass splitting ($\ln\left(\frac{m_{\mathrm{Z}}^2}{m_{\mathrm{W}}^2}\right) \approx 0.26$). For external longitudinal bosons the quantum numbers of the associated Goldstone scalars are used, via the equivalence theorem, $\delta^{\mathrm{LSC}}_{\mathrm{W}_0} = \delta^{\mathrm{LSC}}_{\mathrm{w}^\pm}$ and $\delta^{\mathrm{LSC}}_{\mathrm{Z}_0} = \delta^{\mathrm{LSC}}_{\mathrm{z}^0}$ where $\mathrm{w}^\pm$ and z^0 are the charged and neutral Goldstone bosons of the Higgs doublet following the formalism introduced in Appendix A.

Subleading soft–collinear, δ^{SSC}. Isolating the angular-dependent part of the same pair exchange, where the full double logarithm expands as:

$$\ln^2\left(\frac{|r_{kl}|}{m_V^2}\right) = \ln^2\left(\frac{s}{m_V^2}\right) + 2\ln\left(\frac{s}{m_V^2}\right)\ln\left(\frac{|r_{kl}|}{s}\right) + \dots\,, \tag{D.5}$$

yields a sum over pairs that remains genuinely off-diagonal:

$$\delta^{\mathrm{SSC}}\mathcal{M}_0 = \frac{\alpha_{\mathrm{EW}}}{2\pi} L \sum_{k<l} \ln\frac{|r_{kl}|}{s} \sum_{V=\gamma,\mathrm{Z},\mathrm{W}^\pm} I_k^V I_l^{\bar{V}} \mathcal{M}_0\,. \tag{D.6}$$

The neutral-boson contributions ($V=\gamma,\mathrm{Z}$) are diagonal and proportional to $\mathcal{M}_0$. The charged-boson contributions ($V=\mathrm{W}^\pm$) act with the isospin-raising/lowering operators $I^\pm$ and therefore produce $\mathrm{SU}(2)_\mathrm{L}$-rotated Born amplitudes, which must be evaluated explicitly. This is the only term sensitive to the scattering angle through the ratios $\frac{|t|}{s}$, $\frac{|u|}{s}$.

Collinear single logarithms, δ^{C}. A gauge boson collinear to a single external leg, combined with the external field-renormalisation constants, factorises into one collinear factor per leg:

$$\delta^{\mathrm{C}} = \frac{\alpha_{\mathrm{EW}}}{4\pi} L \sum_k \hat{\delta}_k^{\mathrm{C}}\,, \tag{D.7}$$

where the coefficient $\hat{\delta}_k^{\mathrm{C}}$ depends on the particle type (Table D.B):

$$\text{chiral fermion } f_\sigma^\kappa: \quad \hat{\delta}^{\mathrm{C}} = \frac{3}{2}C_{f^\kappa}^{\mathrm{EW}} - \frac{1}{8\sin^2\theta_\mathrm{W}}\left[(1+\delta_{\kappa R})\frac{m_{f_\sigma}^2}{m_\mathrm{W}^2} + \delta_{\kappa L}\frac{m_{f_{-\sigma}}^2}{m_\mathrm{W}^2}\right], \tag{D.8}$$

$$\text{transverse } \mathrm{W}_\mathrm{T}: \quad \hat{\delta}^{\mathrm{C}} = \frac{1}{2}b_\mathrm{W}^{\mathrm{EW}}\,, \tag{D.9}$$

$$\text{transverse } \mathrm{Z}_\mathrm{T}: \quad \hat{\delta}^{\mathrm{C}} = \frac{1}{2}b_{\mathrm{ZZ}}^{\mathrm{EW}} \quad (\text{plus off-diagonal } \gamma\mathrm{Z} \text{ mixing } b_{\gamma\mathrm{Z}}^{\mathrm{EW}})\,, \tag{D.10}$$

$$\text{longitudinal } \mathrm{W}_0, \mathrm{Z}_0 \text{ (Goldstone } \Phi): \quad \hat{\delta}^{\mathrm{C}} = 2C_\Phi^{\mathrm{EW}} - \frac{N_\mathrm{C}^\mathrm{t}}{4\sin^2\theta_\mathrm{W}}\frac{m_\mathrm{t}^2}{m_\mathrm{W}^2}\,, \tag{D.11}$$

with $N_\mathrm{C}^\mathrm{t}=3$ the colour factor of the top quark. Two features are decisive for the polarisation measurement. First, transverse gauge bosons couple through the β-function coefficient b^{EW} (gauge-field renormalisation) given by:

$$b_{cd}^{\mathrm{EW}} = \underbrace{\frac{11}{3}C_{cd}^{\mathrm{EW}}(V)}_{\text{gauge bosons}} - \underbrace{\frac{1}{3}D_{cd}^{\mathrm{EW}}(\Phi)}_{\text{Higgs scalar}} - \underbrace{\frac{2}{3}\sum_{f,i} N_\mathrm{C}^f D_{cd}^{\mathrm{EW}}(f)}_{\text{fermions}}\,, \tag{D.12}$$

where $D_{cd}^{\mathrm{EW}} = \mathrm{Tr}(I_c I_d)$ is the Dynkin index of the representation, whereas longitudinal bosons couple through the scalar Casimir C_Φ^{EW}. Numerically the symmetric parts are close, $\frac{1}{2}b_\mathrm{W}^{\mathrm{EW}} = 7.08$ versus $2C_\Phi^{\mathrm{EW}} = 7.35$. Second, and more importantly, only the *longitudinal* legs receive the large top–Yukawa term $-\frac{N_\mathrm{C}^\mathrm{t}}{4\sin^2\theta_\mathrm{W}}\frac{m_\mathrm{t}^2}{m_\mathrm{W}^2} \approx -15.9$ per boson, which has no transverse counterpart. The Higgs boson receives the same collinear factor as a longitudinal gauge boson.

Parameter renormalisation, δ^{PR}. Finally, the running of the couplings between m_W and Q generates UV single logarithms, obtained by renormalising the couplings in the Born amplitude. Writing the Born in terms of the $\mathrm{U}(1)_\mathrm{Y}$ and $\mathrm{SU}(2)_\mathrm{L}$ couplings $g_\mathrm{Y} = \frac{e}{\cos\theta_\mathrm{W}}$, $g_\mathrm{L} = \frac{e}{\sin\theta_\mathrm{W}}$, with $\mathcal{M}_0 \propto g_\mathrm{Y}^{n_\mathrm{Y}} g_\mathrm{L}^{n_\mathrm{L}}$:

$$\delta^{\mathrm{PR}} = -\frac{\alpha_{\mathrm{EW}}}{8\pi} L\left(n_\mathrm{Y}\tilde{b}_B^{\mathrm{EW}} + n_\mathrm{L}\tilde{b}_W^{\mathrm{EW}}\right), \qquad \tilde{b}_B^{\mathrm{EW}} = -\frac{41}{6\cos^2\theta_\mathrm{W}}\,, \quad \tilde{b}_W^{\mathrm{EW}} = \frac{19}{6\sin^2\theta_\mathrm{W}}\,, \tag{D.13}$$

where $\tilde{b}_B^{\mathrm{EW}}, \tilde{b}_W^{\mathrm{EW}}$ are the one-loop $\mathrm{U}(1)_\mathrm{Y}$ and $\mathrm{SU}(2)_\mathrm{L}$ β-function coefficients (Table D.C). Since the WZjj-EW amplitude is dominated by the $\mathrm{SU}(2)_\mathrm{L}$ (CC) couplings, δ^{PR} is large and negative, and is essentially independent of the boson polarisations.

EM and mass-gap logarithms, δ^{EM}. The original formalism includes terms for the external legs of the process, which are not genuine-EW logarithms. They are ordinary QED soft or collinear photon emissions. They are not explicitly written down here as they are infrared and will cancel against real radiation. Also, they rely on a photon-mass regulator λ giving logarithms of the form $\ln\left(\frac{Q^2}{\lambda^2}\right)$, which is not physical until combined with real emission.

The breakdown makes explicit which terms carry the helicity dependence. In δ^{LSC} it enters through the boson Casimirs of Table D.A: replacing a transverse boson by a longitudinal one swaps $C^{\mathrm{EW}}_{\mathrm{W_T}} = 8.94 \to C^{\mathrm{EW}}_{\Phi} = 3.68$ (for the W) or $C^{\mathrm{EW}}_{\mathrm{Z_T}} = 6.94 \to 3.68$ (for the Z). Restricting to the two polarised legs, the double-log coefficient $\sum C^{\mathrm{EW}}$ orders the modes as:

$$\underbrace{7.35}_{00} < \underbrace{10.6}_{0\mathrm{T}} < \underbrace{12.6}_{\mathrm{T}0} < \underbrace{15.9}_{\mathrm{TT}}, \tag{D.14}$$

on top of the common quark- and lepton-leg contribution. In δ^{C} each longitudinal leg adds the top–Yukawa term (≈ -15.9), so 00 receives it twice, $0\mathrm{T}$ and $\mathrm{T}0$ once, and TT not at all. The angular term δ^{SSC} differs between modes through the $\mathrm{SU}(2)_{\mathrm{L}}$-rotated Born amplitudes, while δ^{PR} is essentially polarisation-independent. Combining all four contributions over the full set of external legs and the fiducial phase-space yields the integrated corrections $-14.5\,\%$ (00), $-17.9\,\%$ ($0\mathrm{T}$, $\mathrm{T}0$), and $-19.6\,\%$ (TT) of Table 6.A: the leading-logarithmic ordering set by δ^{LSC} is preserved, but its magnitude is diluted by the polarisation-blind quark legs and reshaped by the Yukawa and angular terms, which is why the full one-loop result of [413] is required for the templates rather than the logarithmic approximation alone.

Table D.A: EW quantum numbers and Casimir eigenvalue C^{EW} for the external states of WZjj-EW: the light quark tagging jet partons of the production (LH and RH), the lepton decay products of the W and Z, and the polarised bosons. Longitudinal $\mathrm{W}_0, \mathrm{Z}_0$ are represented by the Goldstone scalars $\mathrm{w}^{\pm}$, z^0 of the doublet Φ. Values from [410, 411], evaluated at $\sin^2\theta_{\mathrm{W}} = 0.2236$. The transverse Z Casimir is the ZZ component of a matrix that also mixes $\gamma\mathrm{Z}$.

State	T	$\frac{Y}{2}$	Q	$C_2^{\mathrm{SU(2)_L}}$	C^{EW} (symbolic)	C^{EW}
Quarks						
$\mathrm{u_L}, \mathrm{d_L}$	$\frac{1}{2}$	$\frac{1}{6}$	$\pm\frac{2}{3}, \mp\frac{1}{3}$	$\frac{3}{4}$	$\dfrac{\sin^2\theta_{\mathrm{W}} + 27\cos^2\theta_{\mathrm{W}}}{36\,\sin^2\theta_{\mathrm{W}}\cos^2\theta_{\mathrm{W}}}$	3.39
$\mathrm{u_R}$	0	$\frac{2}{3}$	$\frac{2}{3}$	0	$\dfrac{4}{9\,\cos^2\theta_{\mathrm{W}}}$	0.57
$\mathrm{d_R}$	0	$-\frac{1}{3}$	$-\frac{1}{3}$	0	$\dfrac{1}{9\,\cos^2\theta_{\mathrm{W}}}$	0.14
Leptons						
ν_ℓ	$\frac{1}{2}$	$-\frac{1}{2}$	0	$\frac{3}{4}$	$\dfrac{1+2\cos^2\theta_{\mathrm{W}}}{4\,\sin^2\theta_{\mathrm{W}}\cos^2\theta_{\mathrm{W}}}$	3.68
ℓ_{L}	$\frac{1}{2}$	$-\frac{1}{2}$	-1	$\frac{3}{4}$	$\dfrac{1+2\cos^2\theta_{\mathrm{W}}}{4\,\sin^2\theta_{\mathrm{W}}\cos^2\theta_{\mathrm{W}}}$	3.68
ℓ_{R}	0	-1	-1	0	$\dfrac{1}{\cos^2\theta_{\mathrm{W}}}$	1.29
Bosons						
$\mathrm{W}^{\pm}_{\mathrm{T}}$	1	0	±1	2	$\dfrac{2}{\sin^2\theta_{\mathrm{W}}}$	8.94
$\mathrm{Z_T}$	1	0	0	2	$\dfrac{2\cos^2\theta_{\mathrm{W}}}{\sin^2\theta_{\mathrm{W}}}$	6.94
$\mathrm{w}^{\pm}\,(\mathrm{W}_0),\ \mathrm{z}^0\,(\mathrm{Z}_0),\ \mathrm{H}$	$\frac{1}{2}$	$\frac{1}{2}$	$\pm1, 0$	$\frac{3}{4}$	$\dfrac{1+2\cos^2\theta_{\mathrm{W}}}{4\,\sin^2\theta_{\mathrm{W}}\cos^2\theta_{\mathrm{W}}}$	3.68

Table D.B: Genuine EW collinear single-log coefficients $\hat{\delta}^{\mathrm{C}}$ for the external legs of WZjj-EW, split into the symmetric part and the top–Yukawa part. Values from [410, 411], evaluated with $\sin^2\theta_{\mathrm{W}} = 0.2236$, $m_{\mathrm{t}} = 175\,\mathrm{GeV}$, $m_{\mathrm{W}} = 80.35\,\mathrm{GeV}$: the top-quark mass is the value used in [410, 411].

Leg	Symmetric part	Value	Yukawa part
$\mathrm{u_L}, \mathrm{d_L}$	$\frac{3}{2}C^{\mathrm{EW}}$	5.08	—
$\mathrm{u_R}$	$\frac{3}{2}C^{\mathrm{EW}}$	0.86	—
$\mathrm{d_R}$	$\frac{3}{2}C^{\mathrm{EW}}$	0.21	—
$\nu_\ell,\ \ell_{\mathrm{L}}$	$\frac{3}{2}C^{\mathrm{EW}}$	5.51	—
ℓ_{R}	$\frac{3}{2}C^{\mathrm{EW}}$	1.93	—
$\mathrm{W_T}$	$\frac{1}{2}b^{\mathrm{EW}}_{\mathrm{W}}$	7.08	—
$\mathrm{Z_T}$	$\frac{1}{2}b^{\mathrm{EW}}_{\mathrm{ZZ}}$	4.51	—
$\mathrm{W}_0, \mathrm{Z}_0$	$2C^{\mathrm{EW}}_{\Phi}$	7.35	$-\dfrac{3}{4\sin^2\theta_{\mathrm{W}}}\dfrac{m_{\mathrm{t}}^2}{m_{\mathrm{W}}^2} \approx -15.9$

Table D.C: One-loop EW β-function coefficients entering δ^{C} and δ^{PR}, in the symmetric ($\tilde{b}$) and physical bases. Values from [410, 411], evaluated at $\sin^2\theta_{\mathrm{W}} = 0.2236$.

	$\tilde{b}^{\mathrm{EW}}_{B}$	$\tilde{b}^{\mathrm{EW}}_{W}$	$b^{\mathrm{EW}}_{\gamma\gamma}$	$b^{\mathrm{EW}}_{\gamma\mathrm{Z}}$	$b^{\mathrm{EW}}_{\mathrm{ZZ}}$	$b^{\mathrm{EW}}_{\mathrm{W}}$
Symbolic	$-\dfrac{41}{6\cos^2\theta_{\mathrm{W}}}$	$\dfrac{19}{6\sin^2\theta_{\mathrm{W}}}$	$-\dfrac{11}{3}$	$-\dfrac{19+22\sin^2\theta_{\mathrm{W}}}{6\,\sin\theta_{\mathrm{W}}\cos\theta_{\mathrm{W}}}$	$\dfrac{19-38\sin^2\theta_{\mathrm{W}}-22\sin^4\theta_{\mathrm{W}}}{6\,\sin^2\theta_{\mathrm{W}}\cos^2\theta_{\mathrm{W}}}$	$\dfrac{19}{6\sin^2\theta_{\mathrm{W}}}$
Numeric	-8.80	14.16	-3.67	-9.57	9.03	14.16

E Data-driven estimation of the misidentified-lepton background

The WZjj final state is selected by requiring three charged leptons, one associated with the leptonic decay of the W boson and two SFOC leptons associated with the Z boson. Background processes in which at least one of these objects is not a prompt lepton from a W or Z decay form the *reducible* background. The misidentified, or "fake", object can be a jet reconstructed as a lepton, a non-prompt lepton produced inside a jet (typically from an HF hadron decay) or, in the case of electrons, an electron from a PC. The dominant contributions are $\mathrm{Z}+\text{jets}$ events, in which a jet fakes the W-lepton, $\mathrm{Z}\gamma$ events, in which an electron from a converted photon is misidentified as the W-lepton, and $\mathrm{t\bar{t}}$ events, which provide two prompt leptons from the two W bosons plus a third lepton from a b-hadron decay.

The probability for such objects to satisfy the signal-lepton requirements is poorly reproduced in the MC simulation. The reducible background is therefore estimated from data with the *matrix method*, described in Section E.1, which provides the baseline estimate. Its results are cross-checked with the semi-data-driven *MC scale factor method*, described in Section E.2. Both methods follow closely the techniques developed for the WZ analyses of $8\,\mathrm{TeV}$ [325] and $13\,\mathrm{TeV}$ [338] data and the inclusive WZ polarisation measurement [11].

E.1 The matrix method

E.1.1 Principle

The matrix method exploits, on a statistical basis, the difference in reconstruction, identification, and isolation quality between real and fake leptons. The three leptons of the WZ final state are ordered (i, j, k) as the W-lepton, the leading-p_{T} Z-lepton, and the trailing-p_{T} Z-lepton. Each lepton is assigned, according to its origin, to one of two truth categories:

- *Real* leptons (R): prompt leptons from a W or Z decay. Events with three real leptons comprise the WZ signal and the irreducible backgrounds (ZZ, $\mathrm{t\bar{t}}+V$, tZj).
- *Fake* leptons (F): leptons inside jets, hadrons misidentified as leptons, or electrons from PCs.

Depending on the origin of the three leptons, eight categories defining the number of events can be constructed:

- N_{RRR}: three real leptons from prompt leptons. Here the WZ signal events and backgrounds with three prompt leptons coming from ZZ, $\mathrm{t\bar{t}}+V$, and tZj are contributing. This is the irreducible background.
- N_{FRR}: fake W-lepton, two real Z-leptons. This category is dominated by $\mathrm{Z}+\text{jets}$, $\mathrm{Z}\gamma$, and $\mathrm{t\bar{t}}$ events.
- N_{RFR} and N_{RRF}: real W-lepton, one real and one fake Z-lepton (the leading and trailing Z-leptons are distinguished). These categories receive contributions from $\mathrm{Z}+\text{jets}$ and $\mathrm{Z}\gamma$ when a real lepton is mis-paired with a fake to form the Z candidate, as well as from $\mathrm{t\bar{t}}$.
- N_{FFR}, N_{FRF}, N_{RFF}: two fake leptons and one real lepton. These categories receive contributions from $\mathrm{W}+\text{jets}$, $\mathrm{Z}+\text{jets}$, $\mathrm{t\bar{t}}$, and multijet production.
- N_{FFF}: three fake leptons. This category is found in simulation to be negligible and is set to zero.

Leptons are additionally classified by their reconstruction quality as *tight* (t) or *loose* (l). Tight leptons are the signal leptons defined in Section 6.6.2 (W-lepton requirements for the first lepton, Z-lepton requirements for the other two). Loose leptons satisfy the looser "matrix-method lepton" selection of Table E.A but fail the signal selection: loose electrons pass basic object-quality, the `LooseAndBLayer` identification, and the impact-parameter requirements, but fail the isolation and/or the tighter identification of signal electrons. Loose muons pass the `Medium` identification but fail the isolation (and, for W-muons, the tighter identification). The t and l samples are by construction disjoint, so "loose" is to be understood throughout as "loose, not tight".

Table E.A: Selection of "matrix-method lepton".

Selection	Electrons	Muons
p_{T}	$> 15\,\mathrm{GeV}$	$> 15\,\mathrm{GeV}$
$\lvert\eta\rvert$	< 2.47	< 2.7
Crack region	not within $1.37 < \lvert\eta\rvert < 1.52$	—
Identification	`LooseAndBLayer`	`Medium`
$\lvert\Delta z_0 \sin\theta\rvert$	$< 0.5\,\mathrm{mm}$	$< 0.5\,\mathrm{mm}$
$\left\lvert\frac{d_0}{\sigma_{d_0}}\right\rvert$	< 5	< 3

The truth and quality classifications are related by the probabilities for a lepton to pass or fail the tight requirements:

- $\mathrm{r} = P(\mathrm{t}|\mathrm{R})$ is the probability for a real lepton to pass the tight selection, and $\bar{\mathrm{r}} = 1 - \mathrm{r} = P(\mathrm{l}|\mathrm{R})$ the probability to fail it.
- $\mathrm{f} = P(\mathrm{t}|\mathrm{F})$ is the probability for a fake lepton to pass the tight selection, and $\bar{\mathrm{f}} = 1 - \mathrm{f} = P(\mathrm{l}|\mathrm{F})$ the probability to fail it.

These probabilities depend on the lepton and are labelled with the index $(1, 2, 3)$ of the lepton in the ordering defined above. In a three-lepton final state the relation between the eight truth and eight quality categories would be an 8×8 matrix. Setting $N_{\mathrm{FFF}} = 0$ and, symmetrically, dropping the N_{lll} equation reduces the problem to the square and readily tractable 7×7 system:

$$\begin{pmatrix} N_{\mathrm{ttt}} \\ N_{\mathrm{ttl}} \\ N_{\mathrm{tlt}} \\ N_{\mathrm{ltt}} \\ N_{\mathrm{tll}} \\ N_{\mathrm{ltl}} \\ N_{\mathrm{llt}} \end{pmatrix} = \begin{pmatrix} \mathrm{r}_1\mathrm{r}_2\mathrm{r}_3 & \mathrm{r}_1\mathrm{r}_2\mathrm{f}_3 & \mathrm{r}_1\mathrm{f}_2\mathrm{r}_3 & \mathrm{f}_1\mathrm{r}_2\mathrm{r}_3 & \mathrm{r}_1\mathrm{f}_2\mathrm{f}_3 & \mathrm{f}_1\mathrm{r}_2\mathrm{f}_3 & \mathrm{f}_1\mathrm{f}_2\mathrm{r}_3 \\ \mathrm{r}_1\mathrm{r}_2\bar{\mathrm{r}}_3 & \mathrm{r}_1\mathrm{r}_2\bar{\mathrm{f}}_3 & \mathrm{r}_1\mathrm{f}_2\bar{\mathrm{r}}_3 & \mathrm{f}_1\mathrm{r}_2\bar{\mathrm{r}}_3 & \mathrm{r}_1\mathrm{f}_2\bar{\mathrm{f}}_3 & \mathrm{f}_1\mathrm{r}_2\bar{\mathrm{f}}_3 & \mathrm{f}_1\mathrm{f}_2\bar{\mathrm{r}}_3 \\ \mathrm{r}_1\bar{\mathrm{r}}_2\mathrm{r}_3 & \mathrm{r}_1\bar{\mathrm{r}}_2\mathrm{f}_3 & \mathrm{r}_1\bar{\mathrm{f}}_2\mathrm{r}_3 & \mathrm{f}_1\bar{\mathrm{r}}_2\mathrm{r}_3 & \mathrm{r}_1\bar{\mathrm{f}}_2\mathrm{f}_3 & \mathrm{f}_1\bar{\mathrm{r}}_2\mathrm{f}_3 & \mathrm{f}_1\bar{\mathrm{f}}_2\mathrm{r}_3 \\ \bar{\mathrm{r}}_1\mathrm{r}_2\mathrm{r}_3 & \bar{\mathrm{r}}_1\mathrm{r}_2\mathrm{f}_3 & \bar{\mathrm{r}}_1\mathrm{f}_2\mathrm{r}_3 & \bar{\mathrm{f}}_1\mathrm{r}_2\mathrm{r}_3 & \bar{\mathrm{r}}_1\mathrm{f}_2\mathrm{f}_3 & \bar{\mathrm{f}}_1\mathrm{r}_2\mathrm{f}_3 & \bar{\mathrm{f}}_1\mathrm{f}_2\mathrm{r}_3 \\ \mathrm{r}_1\bar{\mathrm{r}}_2\bar{\mathrm{r}}_3 & \mathrm{r}_1\bar{\mathrm{r}}_2\bar{\mathrm{f}}_3 & \mathrm{r}_1\bar{\mathrm{f}}_2\bar{\mathrm{r}}_3 & \mathrm{f}_1\bar{\mathrm{r}}_2\bar{\mathrm{r}}_3 & \mathrm{r}_1\bar{\mathrm{f}}_2\bar{\mathrm{f}}_3 & \mathrm{f}_1\bar{\mathrm{r}}_2\bar{\mathrm{f}}_3 & \mathrm{f}_1\bar{\mathrm{f}}_2\bar{\mathrm{r}}_3 \\ \bar{\mathrm{r}}_1\mathrm{r}_2\bar{\mathrm{r}}_3 & \bar{\mathrm{r}}_1\mathrm{r}_2\bar{\mathrm{f}}_3 & \bar{\mathrm{r}}_1\mathrm{f}_2\bar{\mathrm{r}}_3 & \bar{\mathrm{f}}_1\mathrm{r}_2\bar{\mathrm{r}}_3 & \bar{\mathrm{r}}_1\mathrm{f}_2\bar{\mathrm{f}}_3 & \bar{\mathrm{f}}_1\mathrm{r}_2\bar{\mathrm{f}}_3 & \bar{\mathrm{f}}_1\mathrm{f}_2\bar{\mathrm{r}}_3 \\ \bar{\mathrm{r}}_1\bar{\mathrm{r}}_2\mathrm{r}_3 & \bar{\mathrm{r}}_1\bar{\mathrm{r}}_2\mathrm{f}_3 & \bar{\mathrm{r}}_1\bar{\mathrm{f}}_2\mathrm{r}_3 & \bar{\mathrm{f}}_1\bar{\mathrm{r}}_2\mathrm{r}_3 & \bar{\mathrm{r}}_1\bar{\mathrm{f}}_2\mathrm{f}_3 & \bar{\mathrm{f}}_1\bar{\mathrm{r}}_2\mathrm{f}_3 & \bar{\mathrm{f}}_1\bar{\mathrm{f}}_2\mathrm{r}_3 \end{pmatrix} \begin{pmatrix} N_{\mathrm{RRR}} \\ N_{\mathrm{RRF}} \\ N_{\mathrm{RFR}} \\ N_{\mathrm{FRR}} \\ N_{\mathrm{RFF}} \\ N_{\mathrm{FRF}} \\ N_{\mathrm{FFR}} \end{pmatrix} . \tag{E.1}$$

The number of selected (tight, tight, tight) events is N_{ttt}, while N_{RRR} is the number of events with three real leptons: the signal and irreducible backgrounds. The reducible-background contamination of the SR is the difference between the two, defined by:

$$N_{\mathrm{fake}} = N_{\mathrm{ttt}} - \mathrm{r}_1\mathrm{r}_2\mathrm{r}_3 N_{\mathrm{RRR}} \,. \tag{E.2}$$

Applying a suitable linear combination of the second to seventh rows of Equation E.1 to the first one, and neglecting terms of order $\mathcal{O}(\mathrm{f}^3)$ in the small fake-pass probability, N_{fake} can be expressed in terms of the loose-region yields:

$$\begin{aligned} N_{\mathrm{fake}} = &\left[N_{\mathrm{ttl}} - \mathrm{r}_1\mathrm{r}_2\bar{\mathrm{r}}_3 N_{\mathrm{RRR}}\right]\frac{\mathrm{f}_3}{\bar{\mathrm{f}}_3} + \left[N_{\mathrm{tlt}} - \mathrm{r}_1\bar{\mathrm{r}}_2\mathrm{r}_3 N_{\mathrm{RRR}}\right]\frac{\mathrm{f}_2}{\bar{\mathrm{f}}_2} + \left[N_{\mathrm{ltt}} - \bar{\mathrm{r}}_1\mathrm{r}_2\mathrm{r}_3 N_{\mathrm{RRR}}\right]\frac{\mathrm{f}_1}{\bar{\mathrm{f}}_1} \\ &- \left[N_{\mathrm{tll}} - \mathrm{r}_1\bar{\mathrm{r}}_2\bar{\mathrm{r}}_3 N_{\mathrm{RRR}}\right]\frac{\mathrm{f}_2}{\bar{\mathrm{f}}_2}\frac{\mathrm{f}_3}{\bar{\mathrm{f}}_3} - \left[N_{\mathrm{ltl}} - \bar{\mathrm{r}}_1\mathrm{r}_2\bar{\mathrm{r}}_3 N_{\mathrm{RRR}}\right]\frac{\mathrm{f}_1}{\bar{\mathrm{f}}_1}\frac{\mathrm{f}_3}{\bar{\mathrm{f}}_3} - \left[N_{\mathrm{llt}} - \bar{\mathrm{r}}_1\bar{\mathrm{r}}_2\mathrm{r}_3 N_{\mathrm{RRR}}\right]\frac{\mathrm{f}_1}{\bar{\mathrm{f}}_1}\frac{\mathrm{f}_2}{\bar{\mathrm{f}}_2} \,. \end{aligned} \tag{E.3}$$

Each term of the form $\mathrm{r}_i\mathrm{r}_j\bar{\mathrm{r}}_k\, N_{\mathrm{RRR}}$ is the contribution of three-real-lepton events (signal and irreducible background) to the corresponding loose category. Writing the irreducible contribution to a category as N^{irr}_{ijk} and the purely reducible contribution as $N^{\mathrm{red}}_{ijk} = N_{ijk} - N^{\mathrm{irr}}_{ijk}$, and introducing the *fake factor*:

$$F_i = \frac{\mathrm{f}_i}{\bar{\mathrm{f}}_i} \,, \tag{E.4}$$

Equation E.3 becomes:

$$N_{\text{fake}} = N^{\text{red}}_{\text{ttl}} F_3 + N^{\text{red}}_{\text{tlt}} F_2 + N^{\text{red}}_{\text{ltt}} F_1 - N^{\text{red}}_{\text{tll}} F_2 F_3 - N^{\text{red}}_{\text{ltl}} F_1 F_3 - N^{\text{red}}_{\text{llt}} F_1 F_2 \,. \tag{E.5}$$

Since the two Z-lepton candidates obey identical identification criteria, their fake factors are equal, $F_2 = F_3 \equiv F_{\text{Z}}$, and only the W-lepton fake factor $F_1 \equiv F_{\text{W}}$ remains distinct. Equation E.5 then reduces to:

$$N_{\text{fake}} = N^{\text{red}}_{\text{ttl}} F_{\text{Z}} + N^{\text{red}}_{\text{tlt}} F_{\text{Z}} + N^{\text{red}}_{\text{ltt}} F_{\text{W}} - N^{\text{red}}_{\text{tll}} F_{\text{Z}}^2 - N^{\text{red}}_{\text{ltl}} F_{\text{W}} F_{\text{Z}} - N^{\text{red}}_{\text{llt}} F_{\text{W}} F_{\text{Z}} \,. \tag{E.6}$$

The loose-region yields N_{ijk} are taken directly from data (by applying the full WZ selection while reversing the tight requirement on the relevant lepton), and the irreducible contributions N^{irr}_{ijk} are subtracted using the MC prediction. In the WZjj selection these come almost entirely from WZjj-QCD production, which amounts to about 7 % of the loose-region yield.

The fake factors are measured as a function of the fake-lepton p_{T}. To account for this dependence, Equation E.5 is evaluated in bins i, j of p_{T}:

$$\begin{aligned} N_{\text{fake}} = &\sum_i N^{\text{red}\,i}_{\text{ttl}} F^i_{\text{Z}} + \sum_i N^{\text{red}\,i}_{\text{tlt}} F^i_{\text{Z}} + \sum_i N^{\text{red}\,i}_{\text{ltt}} F^i_{\text{W}} \\ &- \sum_i \sum_j N^{\text{red}\,ij}_{\text{tll}} F^i_{\text{Z}} F^j_{\text{Z}} - \sum_i \sum_j N^{\text{red}\,ij}_{\text{ltl}} F^i_{\text{W}} F^j_{\text{Z}} - \sum_i \sum_j N^{\text{red}\,ij}_{\text{llt}} F^i_{\text{W}} F^j_{\text{Z}} \,. \end{aligned} \tag{E.7}$$

E.1.2 Determination of the fake factors

The fake factor of Equation E.4 is the ratio of the probability that a fake lepton passes the signal selection to the probability that it fails it. For a sample of N_{ctrl} fake leptons it can be written as:

$$F = \frac{\text{f}}{\bar{\text{f}}} = \frac{N_{\text{t}}}{N_{\text{ctrl}}} \cdot \frac{N_{\text{ctrl}}}{N_{\text{l}}} = \frac{N_{\text{t}}}{N_{\text{l}}} \,, \tag{E.8}$$

where N_{t} and N_{l} are the numbers of fake leptons in the sample passing and failing the tight requirements, with $N_{\text{t}} + N_{\text{l}} = N_{\text{ctrl}}$. The fake factor is therefore measured in CRs enriched in fake leptons. Any contamination by real leptons would bias it towards larger values.

Three CRs are defined, each targeting a different fake origin: a Z + jets, a Zγ (PC), and a $\text{t}\bar{\text{t}}$ CR. Their selections are summarised in Table E.B. In each of them two Z-leptons are required, and the fake-lepton candidate ℓ_m is the highest-p_{T} matrix-method lepton that is not part of the Z candidate. The Z + jets region selects an SFOC Z-lepton pair within 15 GeV of the Z mass, with m_{T}^{W} and $E_{\text{T}}^{\text{miss}}$ inverted with respect to the signal selection to suppress WZ events. For fake electrons an $m_{\ell\ell} > 81\,\text{GeV}$ cut further rejects PCs. The Zγ region, defined only for fake electrons, requires two opposite-sign Z-muons with $55 < m_{\ell\ell} < 85\,\text{GeV}$ and a three-body mass $m_{3\ell} < 105\,\text{GeV}$, so that the radiated photon shifts the dimuon mass below the Z peak while the muon–muon–electron system reconstructs the Z. This region is statistically independent of the Z + jets region. The $\text{t}\bar{\text{t}}$ region selects an opposite-charge electron–muon Z-lepton pair and assigns as fake candidate the highest-p_{T} matrix-method lepton that forms a same-flavour, same-charge pair with a Z-lepton passing the W selection, mimicking the WZ topology while remaining pure in $\text{t}\bar{\text{t}}$.

Table E.B: Selections of the three CRs used to measure the fake factors. In every region two Z-leptons are required and the fake-lepton candidate ℓ_m is the remaining highest-p_{T} matrix-method lepton.

Z + jets	**Zγ**	**$\mathbf{t\bar{t}}$**
e^+e^- or $\mu^+\mu^-$ Z-leptons	$\mu^+\mu^-$ Z-leptons	$\text{e}^\pm\mu^\mp$ Z-leptons
$\lvert m_{\ell\ell} - m_{\text{Z}}^{\text{PDG}} \rvert < 15\,\text{GeV}$	$55 < m_{\ell\ell} < 85\,\text{GeV}$	Z-lepton of the same flavour
$m_{\text{T}}^{\text{W}} < 30\,\text{GeV}$	$m_{3\ell} < 105\,\text{GeV}$	and charge as ℓ_m,
$E_{\text{T}}^{\text{miss}} < 30\,\text{GeV}$ ($\ell_m = \text{e}$)	$m_{\text{T}}^{\text{W}} < 30\,\text{GeV}$	passing the W selection
$m_{\ell\ell} > 81\,\text{GeV}$ ($\ell_m = \text{e}$)	$E_{\text{T}}^{\text{miss}} < 30\,\text{GeV}$	

In each CR the signal and irreducible-background contributions (predominantly WZ and ZZ events, in which the third real lepton is wrongly assigned as the fake candidate) are estimated from MC and subtracted from data before the fake factor is formed. The subtracted yields are assigned global normalisation uncertainties of 5 % for WZ [10], 10 % for ZZ [339], 15 % for $\text{t}\bar{\text{t}} + V$ and tZj [434], and 25 % to 30 % for the (negligible) VBS and triboson processes. This systematic uncertainty is treated as correlated across the tight and loose selections and the p_{T} bins. The fake factors are measured in four bins of fake-lepton p_{T} in each CR, in data and in MC, for W- and Z-type leptons separately.

E.1.3 $|\eta|$-dependent correction

The fake factor also depends on the detector region in which the fake lepton is reconstructed. The statistics available in the CRs are insufficient to bin simultaneously in p_{T} and η, so a 1D correction as a function of $|\eta|$ is derived, the detector response being assumed symmetric about $\eta = 0$. When the fake factor measured differentially in η, $F_V(\eta)$, is normalised by its mean value inclusive in p_{T} and η, the resulting shape is found to be very similar in all three CRs. The normalised distributions are therefore combined with a weighted least-squares average, using the inverse squared total relative uncertainties as weights, and a parabolic function:

$$f(|\eta|) = a|\eta|^2 + c\,, \tag{E.9}$$

is fitted to the combined profile, separately for fake electrons and fake muons. The fit result provides a continuous $|\eta|$-dependent correction factor that multiplies the fake-factor weights, and the $68\,\%$ confidence interval returned by the fit is taken as its uncertainty.

E.1.4 Fake factors by fake-lepton origin

The fake factor depends on the nature and topology of the faking object. The CR fake factors F_V^{CR} are converted into fake factors that depend on the *origin* of the fake lepton, F_V^{flavour}. Three origins are distinguished: HF fakes, from non-isolated leptons inside b- or c-hadron jets, LF fakes, from jets of u, d, s quarks, and PC fakes, relevant for electrons only. The `MCTruthClassifier` tool is used to determine, in each CR and in the loose SRs, the origin of the selected loose leptons. Cases in which the origin cannot be determined (of order $1\,\%$ to $2\,\%$, up to $4.5\,\%$ for muons) are classified as LF.

Denoting by $f_{\mathrm{flavour}}^{\mathrm{CR}}$ the fraction of each fake origin in a given CR, the origin fake factors are obtained by inverting:

$$\begin{pmatrix} F_V^{\mathrm{Z+jets}} \\ F_V^{\mathrm{t\bar{t}}} \\ F_V^{\mathrm{Z}\gamma} \end{pmatrix} = \begin{pmatrix} f_{\mathrm{HF}}^{\mathrm{Z+jets}} & f_{\mathrm{LF}}^{\mathrm{Z+jets}} & f_{\mathrm{PC}}^{\mathrm{Z+jets}} \\ f_{\mathrm{HF}}^{\mathrm{t\bar{t}}} & f_{\mathrm{LF}}^{\mathrm{t\bar{t}}} & f_{\mathrm{PC}}^{\mathrm{t\bar{t}}} \\ f_{\mathrm{HF}}^{\mathrm{Z}\gamma} & f_{\mathrm{LF}}^{\mathrm{Z}\gamma} & f_{\mathrm{PC}}^{\mathrm{Z}\gamma} \end{pmatrix} \begin{pmatrix} F_V^{\mathrm{HF}} \\ F_V^{\mathrm{LF}} \\ F_V^{\mathrm{PC}} \end{pmatrix}, \tag{E.10}$$

where V stands for the W- or Z-lepton selection. The fractions f are taken from the sum of MC predictions in each CR, and the F_V^{CR} from data. The resulting compositions, evaluated with `MCTruthClassifier`, are summarised in Table E.C. The resulting origin fake factors are applied to the loose leptons of the loose SRs, the origin of each loose lepton being drawn randomly according to the p_{T}-dependent flavour composition of the loose SR, itself determined from MC in the same p_{T} binning as the fake factors. This makes the procedure differential in p_{T} and accounts for the evolution of the fake-lepton composition with p_{T}. The $|\eta|$ dependence is not binned together with p_{T}: it enters through the multiplicative correction of Section E.1.3, common to all p_{T} bins, so that the fake factor of a loose lepton is the product of a p_{T}-dependent and an $|\eta|$-dependent term, the two dependences being assumed to factorise.

E.1.5 Closure test and systematic uncertainties

The consistency of the method is verified with a closure test in which the full matrix-method procedure, fake-factor measurement in the CRs, $|\eta|$ correction, and origin decomposition, is applied to MC events, and the resulting fake-lepton estimate in the WZjj-EW SR is compared with the fake-lepton yield ($\mathrm{Z}+$jets, $\mathrm{t\bar{t}}$, $\mathrm{Z}\gamma$) predicted directly by the same MC. The method reproduces the MC truth at the percent level: summed over the four three-lepton channels it returns $51.7 \pm 1.1 \pm 6.8$ fake-lepton events against 52.3 ± 5.5 from the direct MC prediction, where the two uncertainties on the estimate are the statistical uncertainties of the loose-region data (N) and of the CRs (F). Good closure is also observed in the shapes of the discriminating observables, the S. vs B. DNN and the Pol. 00, Pol. $0\mathrm{X}$, and Pol. $\mathrm{X}0$ DNN scores, with $\frac{\chi^2}{n_{\mathrm{points}}}$ of 1.00, 0.71, 0.23, and 1.63, respectively, well within the MC statistical uncertainty.

The systematic uncertainty on the estimated fake-lepton yield receives contributions from:

- the statistical uncertainty of the data in the loose SRs (the N_{ijk} terms)
- the statistical uncertainty of the CR samples used to measure the fake factors (the F_i terms)
- the subtraction of irreducible (real-lepton) events from the CRs and the loose SRs, propagated through the method from the normalisation uncertainties quoted in Section E.1.2 together with a $30\,\%$ uncertainty on WZjj-QCD in the loose SRs
- the $|\eta|$ correction
- the flavour composition of the loose SRs
- the flavour composition of the CRs used to derive the origin fake factors

Table E.C: Origin composition (HF, LF, and PC) of the loose leptons in the three CRs and the loose SRs, for Run-2 (a) and Run-3 (b). The muon rows do not include the few percent of leptons of undetermined origin, so they need not sum to $100\,\%$. Uncertainties are MC statistical only.

(a) Run-2 ($13\,\mathrm{TeV}$, $140\,\mathrm{fb}^{-1}$)

Selection	Flavour fraction [%]		
	HF	LF	PC
Fake electrons			
$\mathrm{Z}+\mathrm{jets}$	33.9 ± 0.3	51.3 ± 0.5	14.8 ± 0.3
$\mathrm{t\bar{t}}$	84.9 ± 0.6	11.6 ± 0.3	3.5 ± 0.2
$\mathrm{Z}\gamma$	11.4 ± 2.1	22.5 ± 3.0	66.0 ± 8.1
Loose SR, W-lepton	36.2 ± 0.4	43.1 ± 0.4	20.7 ± 0.3
Loose SR, Z-lepton	64.1 ± 0.6	28.9 ± 0.7	7.0 ± 0.3
Fake muons			
$\mathrm{Z}+\mathrm{jets}$	79.2 ± 0.4	20.8 ± 0.2	—
$\mathrm{t\bar{t}}$	80.9 ± 0.3	19.1 ± 0.2	—
Loose SR, W-lepton	79.6 ± 0.7	18.8 ± 0.3	—
Loose SR, Z-lepton	83.9 ± 0.6	14.1 ± 0.7	—

(b) Run-3 ($13.6\,\mathrm{TeV}$, $164\,\mathrm{fb}^{-1}$)

Selection	Flavour fraction [%]		
	HF	LF	PC
Fake electrons			
$\mathrm{Z}+\mathrm{jets}$	42.9 ± 0.7	46.9 ± 0.8	10.2 ± 0.4
$\mathrm{t\bar{t}}$	85.5 ± 0.5	11.0 ± 0.3	3.5 ± 0.2
$\mathrm{Z}\gamma$	14.7 ± 1.3	32.6 ± 5.9	52.7 ± 4.0
Loose SR, W-lepton	41.3 ± 0.8	44.7 ± 0.9	14.0 ± 0.4
Loose SR, Z-lepton	45.5 ± 1.2	48.4 ± 1.4	6.0 ± 0.4
Fake muons			
$\mathrm{Z}+\mathrm{jets}$	84.6 ± 0.4	15.4 ± 0.2	—
$\mathrm{t\bar{t}}$	81.0 ± 0.3	18.8 ± 0.1	—
Loose SR, W-lepton	83.0 ± 1.0	16.7 ± 0.4	—
Loose SR, Z-lepton	83.1 ± 0.8	15.4 ± 1.4	—

Applied to the full Run-3 dataset ($164\,\mathrm{fb}^{-1}$ at $\sqrt{s} = 13.6\,\mathrm{TeV}$), the matrix method predicts the reducible-background yield of Table E.D in the SR, $60.9 \pm 2.8 \pm 6.4$ events summed over the four channels (the estimate used as input to the fit, from a later iteration of the same procedure, is 61.7 events, $1.3\,\%$ above this value), against 52.3 ± 5.5 from the direct MC prediction of the same three processes. The lower part of the table gives the per-source breakdown of the uncertainty, the total of each column being the quadrature sum of its six sources: the total is about $29\,\%$, dominated by the subtraction of the irreducible background and by the flavour composition of the loose SRs, while the $|\eta|$ correction contributes about $2\,\%$.

The same two methods were used in the previous Run-2 VBS WZjj measurement [360], where the reducible background in the EW SR amounted to 49 ± 11 events at $\sqrt{s} = 13\,\mathrm{TeV}$ for $140\,\mathrm{fb}^{-1}$, of comparable size to the present estimate once the higher Run-3 luminosity and CM energy are taken into account.

E.2 The data/MC scale factor method

E.2.1 Principle

The MC scale factor method is a semi-data-driven cross-check of the matrix method that has been employed in other analyses and in particular in the previous WZjj-EW measurement [360], where it was found to agree with the matrix-method estimate. Rather than predicting the fake yield from data, it takes

Table E.D: Reducible-background yield in the WZjj-EW SR from the matrix method, per three-lepton channel and summed, compared with the direct MC prediction of the three fake-lepton processes (uncertainties on the latter are MC statistical only). The two uncertainties on the matrix-method yield are the statistical uncertainties of the loose-region data (N) and of the fake factors (F).

	eee	eμμ	μee	μμμ	Total
Matrix method	$16.1 \pm 1.7 \pm 2.0$	$12.6 \pm 0.8 \pm 1.2$	$17.0 \pm 1.8 \pm 1.7$	$15.2 \pm 0.9 \pm 1.4$	$60.9 \pm 2.8 \pm 6.4$
MC ($\mathrm{t\bar{t}} + \mathrm{Z} + \mathrm{jets} + \mathrm{Z}\gamma$)	9.3 ± 1.8	20.1 ± 3.5	9.1 ± 3.0	13.8 ± 2.4	52.3 ± 5.5
Relative uncertainty on the matrix-method estimate [%]					
Stat. (N)	10.74	6.69	11.08	6.19	4.71
F stat.	12.45	9.88	10.61	10.18	10.85
Irr. subtraction	14.83	19.60	18.11	23.83	18.90
$\lvert\eta\rvert$ correction	1.93	1.91	1.90	1.93	1.92
Flavour comp. (loose SR)	23.32	6.27	16.91	4.38	13.41
F flavour comp. (CRs)	11.87	11.88	13.38	16.37	13.37
Total	34.33	26.66	32.13	31.63	29.32

the fake-lepton contribution from the MC simulation and corrects it for the mis-modelling of the probability that a fake object is reconstructed as a tight lepton, separately for each fake source.

Correction factors quantifying the data and MC difference in this fake-identification probability are extracted in dedicated CRs enriched in $\mathrm{Z} + \mathrm{jets}$, $\mathrm{t\bar{t}}$, and $\mathrm{Z}\gamma$ events (the $\mathrm{Z}\gamma$ region again being used for electrons only). Each region requires at least two Z-leptons and a third, "loose", lepton, called the fake-lepton candidate satisfying the looser selection of Table E.E, together with criteria that maximise the probability that this third lepton is of misidentification origin. The probability for the fake candidate to pass the tight W- or Z-lepton selection is measured by applying those criteria to the loose lepton, and the data/MC difference is turned into a multiplicative correction applied to the simulation.

Table E.E: Selection of the "loose leptons".

Selection	Electrons	Muons
p_{T}	$> 15\,\mathrm{GeV}$	$> 15\,\mathrm{GeV}$
$\lvert\eta\rvert$	< 2.47	< 2.7
Crack region	not within $1.37 < \lvert\eta\rvert < 1.52$	—
Identification	`Loose`	`Medium`
$\lvert\Delta z_0 \sin\theta\rvert$	$< 0.5\,\mathrm{mm}$	$< 0.5\,\mathrm{mm}$
$\left\lvert\frac{d_0}{\sigma_{d_0}}\right\rvert$	< 5	< 3

E.2.2 Control regions

The three CRs are constructed to be mutually independent and, where possible, pure in a single process. Their selections are summarised in Table E.F. In contrast to the matrix-method CRs, the loose lepton here is additionally required to pass the tight signal-lepton selection, since the method scales the MC fake yield that actually enters the SR. The $\mathrm{t\bar{t}}$ region selects an opposite-charge electron–muon Z-lepton pair and a loose third lepton forming an opposite-flavour, opposite-charge pair with a Z-lepton. It is very pure in $\mathrm{t\bar{t}}$ and shows a flat excess of the $\mathrm{t\bar{t}}$ prediction over data. The $\mathrm{Z} + \mathrm{jets}$ region is tightened with respect to its matrix-method counterpart, the Z-candidate mass is restricted to $85 < m_{\ell\ell} < 96\,\mathrm{GeV}$ and events with a three-body mass in $87 < m_{3\ell} < 93\,\mathrm{GeV}$ are vetoed, to remove PC and $\mathrm{t\bar{t}}$ contamination, reaching a $\mathrm{Z} + \mathrm{jets}$ purity of about $92\,\%$ for electrons and close to unity for muons, with a global data/MC agreement of about $20\,\%$. The $\mathrm{Z}\gamma$ region is identical to that of the matrix method.

E.2.3 Determination of the correction factors

A separate correction factor μ is derived for each fake source, lepton flavour, and lepton selection. For the W-lepton selection, for instance, five factors are extracted:

- $\mu_{\mathrm{t\bar{t}}}^{\mathrm{W\text{-}e}}$, $\mu_{\mathrm{Z+jets}}^{\mathrm{W\text{-}e}}$, $\mu_{\mathrm{Z}\gamma}^{\mathrm{W\text{-}e}}$ for fake W-electrons,
- $\mu_{\mathrm{t\bar{t}}}^{\mathrm{W\text{-}\mu}}$, $\mu_{\mathrm{Z+jets}}^{\mathrm{W\text{-}\mu}}$ for fake W-muons,

Table E.F: Selections of the three CRs used to extract the MC scale factors.

Z + jets (pure)	**Zγ**	**tt̄**
e^+e^- or $\mu^+\mu^-$ Z-leptons $85 < m_{\ell\ell} < 96\,\text{GeV}$ veto $87 < m_{3\ell} < 93\,\text{GeV}$ $m_\text{T}^\text{W} < 30\,\text{GeV}$ $E_\text{T}^\text{miss} < 30\,\text{GeV}$ ($\ell_m = e$)	$\mu^+\mu^-$ Z-leptons $55 < m_{\ell\ell} < 85\,\text{GeV}$ $m_{3\ell} < 105\,\text{GeV}$ $m_\text{T}^\text{W} < 30\,\text{GeV}$ $E_\text{T}^\text{miss} < 30\,\text{GeV}$	$e^\pm\mu^\mp$ Z-leptons loose lepton of flavour ≠ the opposite-charge Z-lepton, forming the "fake Z"
loose lepton = highest-p_T loose lepton, also required to pass the tight W/Z selection		

with the corresponding factors for the Z-lepton selection (no Zγ factor is needed for muons). Because none of the CRs is perfectly pure, the factors for a given lepton flavour are determined together in a combined fit of the three regions:

$$N_\text{data}^{X,\text{CR}} - N_\text{others}^{X,\text{CR}} = \mu_{\text{t}\bar{\text{t}}}^X N_{\text{t}\bar{\text{t}}}^{X,\text{CR}} + \mu_{\text{Z+jets}}^X N_{\text{Z+jets}}^{X,\text{CR}} + \mu_{\text{Z}\gamma}^X N_{\text{Z}\gamma}^{X,\text{CR}}\,, \tag{E.11}$$

where CR runs over the three CRs and X denotes the lepton flavour and selection of interest. $N_\text{data}^{X,\text{CR}}$ is the number of events selected in data, while $N_{\text{t}\bar{\text{t}}}$, $N_{\text{Z+jets}}$, $N_{\text{Z}\gamma}$, and N_others are the MC expectations for $\text{t}\bar{\text{t}}$, Z + jets, Zγ, and the remaining (real-lepton) processes, respectively.

Because the jet multiplicity is not perfectly described by the $\text{t}\bar{\text{t}}$ and Z + jets samples, each CR is further split into events with fewer than two jets and events with at least two jets, and two independent fits are performed. The uncertainties entering the fit are the statistical uncertainties of the data and of all MC samples, together with global normalisation uncertainties on the real-lepton processes (N_others): 5 % for WZ [10], 10 % for ZZ [339], 15 % for $\text{t}\bar{\text{t}} + V$ and tZj [434], and 25 % to 30 % for the (negligible) VBS and triboson processes. No normalisation uncertainty is assigned to the Z + jets, $\text{t}\bar{\text{t}}$, and Zγ samples, whose normalisations are the result of the fit. The factors obtained from the fits are summarised in Table E.G. The quoted uncertainties are those returned by the fit.

Table E.G: MC scale factors from the combined fit of the CRs, for Run-2 (a) and Run-3 (b). Electrons (left) and muons (right) within each.

(a) Run-2 ($13\,\text{TeV}$, $140\,\text{fb}^{-1}$)

	Electrons						Muons			
	< 2 jets			≥ 2 jets			< 2 jets		≥ 2 jets	
	Zγ	**Z + jets**	**tt̄**	**Zγ**	**Z + jets**	**tt̄**	**Z + jets**	**tt̄**	**Z + jets**	**tt̄**
W-electron	1.11 ± 0.16	0.74 ± 0.13	1.30 ± 0.3	0.2 ± 0.8	2.3 ± 0.3	1.24 ± 0.2	W-muon	0.78 ± 0.25	1.43 ± 0.20	2.9 ± 0.8	1.5 ± 0.2
Z-electron	1.50 ± 0.11	0.75 ± 0.04	1.11 ± 0.1	0.3 ± 0.4	2.0 ± 0.2	0.96 ± 0.07	Z-muon	0.85 ± 0.08	1.01 ± 0.08	2.6 ± 0.3	1.01 ± 0.07

(b) Run-3 ($13.6\,\text{TeV}$, $164\,\text{fb}^{-1}$)

	Electrons						Muons			
	< 2 jets			≥ 2 jets			< 2 jets		≥ 2 jets	
	Zγ	**Z + jets**	**tt̄**	**Zγ**	**Z + jets**	**tt̄**	**Z + jets**	**tt̄**	**Z + jets**	**tt̄**
W-electron	1.45 ± 0.10	1.2 ± 0.2	1.1 ± 0.2	0.4 ± 0.3	2.7 ± 0.7	1.6 ± 0.2	W-muon	1.1 ± 0.25	1.27 ± 0.15	1.9 ± 0.5	2.0 ± 0.13
Z-electron	1.40 ± 0.07	0.97 ± 0.04	0.89 ± 0.07	0.9 ± 0.3	1.5 ± 0.2	1.26 ± 0.06	Z-muon	0.97 ± 0.07	1.09 ± 0.06	1.3 ± 0.15	1.43 ± 0.06

E.2.4 Application

The correction factors are then applied to the MC events in the WZ SR according to the MC process ($\text{t}\bar{\text{t}}$, Z + jets, or Zγ) and to the number and flavour of fake leptons in the event. The `MCTruthClassifier` tool is used to decide whether each selected lepton is real or fake: each MC event is weighted by the correction factor corresponding to the source and flavour of its fake lepton, two factors being applied when the event contains two fake leptons. The shapes and yields of the reducible background obtained in this way agree with the matrix-method estimate within the total uncertainty of the latter, which is consequently retained as the baseline.

F List of Monte Carlo samples

This appendix presents the MC samples used to model the signal and background processes of the analysis, for both the full Run-2 (2015–2018, $\sqrt{s} = 13\,\text{TeV}$, $140\,\text{fb}^{-1}$) and the partial Run-3 (2022–2024, $\sqrt{s} = 13.6\,\text{TeV}$, $164\,\text{fb}^{-1}$) datasets. For every process the nominal and, where available, the alternative generator are listed together with the ME accuracy and multijet merging, the parton shower program, the PDF set, and the main generation settings. The general structure of an event generation chain (hard scattering, parton shower, hadronisation, underlying event, and decay) together with the LO and NLO formalism, the merging and matching prescriptions, and the meaning of the renormalisation and factorisation scales, is described in Section 4.6. Only the process-specific choices are summarised here. Pile-up is modelled by overlaying each hard-scattering event with inelastic pp collisions generated with Pythia 8.186 using the `NNPDF2.3LO` PDF set and the A3 set of tuned parameters [435]. The simulated events are reweighted so that the average number of interactions per bunch crossing $\langle\mu\rangle$ matches the data, the measured $\langle\mu\rangle$ being rescaled by 1.03 ± 0.04 to improve the data/MC agreement of the charged-particle track distributions and the visible inelastic pp cross-section. Unless stated otherwise, the generator-internal dynamical renormalisation and factorisation scales are used. For all simulated processes except those modelled with Sherpa, the decays of bottom and charm hadrons are handled by EvtGen [274]. The detector response is simulated with Geant4, and the simulated events are passed through the same reconstruction chain as the data.

The samples fall into three groups: the WZjj-EW signal, the dominant QCD-induced WZjj-QCD background together with its interference with the signal, and the remaining background processes. Background processes with at least one misidentified ("fake") lepton are not taken from the simulation directly but are estimated from MC normalised with data-driven techniques (the matrix method and the MC scale factor cross-check), as detailed in Appendix E.

F.1 Signal and WZjj-EW modelling

Several distinct WZjj samples enter the analysis, distinguished by the order of the EW coupling fixed at generation. The EW signal, WZjj-EW, is generated at order α_{EW}^{6}, the QCD-induced background, WZjj-QCD, at order α_{EW}^{4}, and the interference between the two, WZjj-INT, of order α_{EW}^{5}, is treated as a separate contribution. All three share the same fully leptonic final state $\ell'^{\pm}\nu_{\ell}\,\ell^{+}\ell^{-}\,\text{jj}$.

For the polarisation measurement, two flavours of WZjj-EW sample are used and must be carefully distinguished:

***Full* samples.** These include all polarisation states, the off-shell contributions, and the τ decays of the bosons. They provide the most complete description of the process but do not isolate a definite polarisation state.

***Polarised* samples.** Here both the W and the Z are forced on-shell and decay only to light leptons (e, μ), with the intermediate bosons projected onto definite helicity configurations. One sample is produced per joint state: W_0Z_0, $\text{W}_0\text{Z}_\text{T}$, $\text{W}_\text{T}\text{Z}_0$, and $\text{W}_\text{T}\text{Z}_\text{T}$, following the helicity decomposition of Equation 6.91, plus a *polarisation-sum* sample that retains only on-shell production without separating the helicity states.

The τ and off-shell contributions are deliberately excluded from the signal definition: τ polarisation is not readily accessible given the complexity of reconstructing τ decays, so these contributions are instead

treated as part of the background. They are evaluated by subtraction, exploiting the fact that they are present in the full sample but absent from the polarisation-sum sample. The off-shell contribution is found to be at the level of $1.5\,\%$ of the WZjj-EW yield in the phase-space of the measurement, in line with the $1.5\,\%$ to $1.6\,\%$ of the NLO EW calculation collected in Table 6.A. It is not treated as a separate component but absorbed into the effective fractions, as discussed in Section 6.6.3.1. The τ contribution amounts to about $5\,\%$ in both runs of all WZjj-EW events in the SR and is included as a background. Triboson final states with more than three leptons and/or more than one neutrino are modelled by dedicated VVV samples and counted as background. The triboson and s-channel diagrams that share the signal final state are partly contained in the on-shell signal samples, at the level of a few percent and without a way to separate them, and are suppressed in the MadGraph alternative by a generator-level m_{jj} requirement, so that neither biases the polarisation extraction.

F.1.1 The WZjj-EW full samples

The nominal full sample is generated with Sherpa, version 2.2.12 for Run-2 and 2.2.14 for Run-3, at order α_{EW}^{6} with no strong coupling (α_{S}^{0}). The LO-accurate MEs for the WZjj-EW final state with up to one additional parton, WZjj-EW $+\,0,1\mathrm{j}$, are merged with the dipole parton shower using the MEPS merging prescription at LO accuracy, MEPS@LO [252–255], based on CS dipole factorisation. In both runs, diagrams with a b-quark in either the initial or final state are excluded from the ME. This sample is the one used to extract, by subtraction against the polarisation-sum sample, both the τ and the off-shell contributions discussed above.

As an alternative, the full process is also generated at LO with MadGraph (2.6.5 for Run-2, 3.5.7 for Run-3) interfaced to Pythia 8 (8.240 and 8.308, respectively) for the parton shower, hadronisation, and underlying event. In both cases the local-recoil dipole (LRD) shower is used, which provides an improved description of deep-inelastic-scattering data and of VBS topologies. The two runs differ in their treatment of the b-quark, which is tied to the removal of the overlap with the resonant tZj process:

- For Run-2 the 5FNS is used, so that b-quarks appear in the ME diagrams. The sample is therefore split according to the presence or absence of a b-quark in the initial state, the two parts modelling the WZjj-EW signal and the tZj background, respectively, the latter being instead taken from a dedicated NLO sample.
- For Run-3 the 4FNS is used, which removes b-quarks from the ME altogether. This produces a total cross-section close to that of the 5FNS with the b-quark diagrams vetoed, while avoiding the dedicated tZj removal that the 5FNS required, and the associated integration instability.

The renormalisation and factorisation scale is set to the MadGraph default (option "-1") for Run-2 but to the well-defined choice (option "3"):

$$\mu = \frac{1}{2}\sum_{i=1}^{N_q}\sqrt{m_i^2 + p_{\mathrm{T},i}^2}\,, \tag{F.1}$$

for Run-3, the change being motivated by the difficulty of verifying the value returned by the "-1" default. The same scale choice (Equation F.1) is adopted for all Run-3 MadGraph signal samples.

F.1.2 Polarised WZjj-EW samples

The polarised samples are the heart of the measurement. Assigning a definite polarisation to the intermediate W and Z requires factorising the amplitude into production, propagator, and decay parts, which is only gauge invariant once the bosons are projected onto their mass shell through the NWA or the DPA, as detailed in Section 6.4. Two state-of-the-art generators implement this projection with two complementary strategies, which allows a direct cross-check between the produced samples.

F.1.2.1 Sherpa 3.0 (nominal)

The nominal polarised samples are generated with Sherpa 3.0.1 for both Run-2 and Run-3, considering on-shell WZjj-EW production with up to one additional jet at LO, WZjj-EW $+\,0,1\mathrm{j}$. Triboson contributions that share the signal final state are kept: the divergences they could cause are avoided by keeping the vector-boson masses finite, and the gauge dependence this introduces was checked to be small against a full off-shell Sherpa 3 prediction at the same order. Diagrams with a b-quark in either the initial or final state are not considered. The polarisation is defined in the diboson CM frame (the modified-helicity coordinate system). The on-shell projection that removes the non-resonant diagrams is performed with a spin-correlated NWA, and the polarisation information is handled through on-the-fly, per-event weights:

the event generation itself corresponds to an effectively unpolarised simulation, and each polarised configuration is recovered as the corresponding diagonal element of the per-event SDM of Equation 6.89, the interference terms being the real parts of the off-diagonal elements. This single-sample approach is the decisive practical advantage of SHERPA: all four joint polarisation states, the single-boson polarisations, their interference terms, and several reference-frame definitions are obtained *simultaneously* from one run, with negative-weight interference events handled natively. Four samples are kept, one per joint state, together with a fifth polarisation-sum sample retaining only on-shell production. The residual interference between polarisation states contained in the polarisation-sum sample is below $2\,\%$ in the fiducial region, consistent with the $1.7\,\%$ to $1.8\,\%$ of Section 6.5.3, and is neglected.

The LO-accurate MEs for up to one additional parton are matched to the CS-dipole parton shower with the MEPS@LO prescription [254, 255], so that the SHERPA 3.0 prediction retains the *real* QCD corrections, the additional hard jet of the WZjj-EW $+\,1\mathrm{j}$ ME, that a pure LO generation lacks. Those corrections matter because the NLO QCD corrections to the polarisation fractions arise predominantly from real emission rather than from virtual diagrams [401, 402]. This is precisely the difference probed by the SHERPA-versus-MADGRAPH comparison: contrasting the SHERPA 3.0 prediction, accurate to LO $0, 1\mathrm{j}$ in QCD, with the pure-LO MADGRAPH prediction isolates the impact of these real QCD corrections on the polarisation predictions. For this reason SHERPA 3.0 is taken as the nominal polarised generator and MADGRAPH as the cross-check.

F.1.2.2 MADGRAPH (alternative)

The alternative polarised samples are generated with MADGRAPH5_AMC@NLO (2.9.16 interfaced to PYTHIA 8.307p1 for Run-2, 3.5.7 interfaced to PYTHIA 8.308 for Run-3, with EVTGEN for the HF decays). In contrast to SHERPA, the polarisation projection is applied *a priori*, at the ME level via the DPA, before the phase-space integration: the four joint states $\mathrm{W_0Z_0}$, $\mathrm{W_0Z_T}$, $\mathrm{W_TZ_0}$, and $\mathrm{W_TZ_T}$ are each produced as a *separate* sample with the polarisation fixed at generation, the polarisation being defined in the diboson CM frame, and a fifth polarisation-sum sample completes the set. The generation uses the 4FNS, again to avoid the tZj overlap and the 5FNS, with the scale choice of Equation F.1 (option "3"). A generator-level cut $m_{\mathrm{jj}} > 110\,\mathrm{GeV}$ is applied to suppress the resonant $\mathrm{W^{\pm}W^{\mp}Z}$ and $\mathrm{W^{\pm}ZZ}$ triboson contributions. This threshold lies well below the minimum m_{jj} of any analysis region ($m_{\mathrm{jj}} > 150\,\mathrm{GeV}$ in the CRs) and has no impact on the measurement. As for SHERPA, the residual polarisation interference in the polarisation-sum sample is below $2\,\%$, consistent with the $1.7\,\%$ to $1.8\,\%$ of Section 6.5.3, and is neglected.

F.1.2.3 Accuracy and limitations

The multijet merging in SHERPA captures the real-emission part of the NLO QCD corrections, which dominates the effect on the polarisation fractions. The residual virtual corrections and the ultra-soft and ultra-collinear emissions are neglected, leaving a small LO-like uncertainty on the fractions themselves. Neither generator provides NLO EW corrections for the polarised process. These are sizeable and polarisation-dependent ($\delta^{\mathrm{EW}}_{\mathrm{unpol.}} \approx -18.5\,\%$, with a marked hierarchy between the longitudinal and transverse modes) and are taken from the dedicated DPA calculation discussed in Section 6.5.3, the integrated results of which are collected in Table 6.A.

The signal samples are summarised in Table F.A.

Table F.A: Signal and WZjj samples (EW, QCD-induced, and their interference) for Run-2 and Run-3.

Process	Generator (Run-2 / Run-3) + PS	ME accuracy / merging	PDF	Main settings
WZjj-EW *full* (nominal)	Sherpa 2.2.12 / 2.2.14	LO, WZjj-EW $+\,0,1\mathrm{j}$ (MEPS@LO), CS-dipole PS	`NNPDF3.0NLO`	$\mathcal{O}(\alpha_{\mathrm{EW}}^{6})$, no b in ME, default Sherpa scale, τ & off-shell extracted by subtraction
WZjj-EW *full* (alt.)	MadGraph 2.6.5+Pythia 8.240 / MadGraph 3.5.7+Pythia 8.308	LO	`NNPDF3.0NLO` (PS `NNPDF2.3LO`)	LRD shower, 5FNS (Run-2, split on b for tZj), 4FNS (Run-3), scale -1 (Run-2), 3 (Run-3)
WZjj-EW *polarised* (Sherpa, nominal)	Sherpa 3.0.1 (Run-2 & Run-3)	LO, WZjj-EW $+\,0,1\mathrm{j}$, MEPS@LO merging	`PDF4LHC21`	on-shell, diboson CM frame, per-event weights $\Rightarrow$ 4 joint states + single-boson + interf. + multi-frame from one run, e,μ decays
WZjj-EW *polarised* (MadGraph, alt.)	MadGraph 2.9.16+Pythia 8.307p1 / MadGraph 3.5.7+Pythia 8.308 (+EvtGen)	LO	`NNPDF3.0NLO` 4FNS (PS A14 [436] `NNPDF2.3LO`)	on-shell, diboson CM frame, 4FNS, scale 3, $m_{\mathrm{jj}} > 110\,\mathrm{GeV}$, 4 states generated separately + pol.-sum
WZjj-QCD (nominal)	Sherpa 2.2.12 / 2.2.16	NLO $0,1\mathrm{j}$ + LO $2,3\mathrm{j}$, MEPS@NLO (Comix, OpenLoops)	`NNPDF3.0NNLO`	$\mathcal{O}(\alpha_{\mathrm{EW}}^{4})$, default Sherpa scale, single inclusive sample[†]
WZjj-QCD (alt.)	MadGraph 2.6.5+Pythia 8.240 / MadGraph 3.5.7+Pythia 8.308	NLO $\leq 2\mathrm{j}$, FxFx-merged	`NNPDF3.0NLO` (PS `NNPDF2.3LO`)	$q_{\mathrm{cut}} = 25\,\mathrm{GeV}$, $p_{\mathrm{T}}^{\mathrm{j}} > 10\,\mathrm{GeV}$, scale -1. Run-3 used as a truth-level reweighting of the nominal Sherpa sample
WZjj-INT	MadGraph 2.6.5+Pythia 8.240 / MadGraph 3.5.7+Pythia 8.308	LO	`NNPDF3.0NLO`	$\mathcal{O}(\alpha_{\mathrm{EW}}^{5}\alpha_{\mathrm{S}})$ at $\lvert\mathcal{M}\rvert^{2}$ level, same cuts as WZjj-EW, $\sim 6\,\%$ (Run-2), $\sim 7\,\%$ (Run-3) of WZjj-EW

[†]The merging scale q_{cut} separates emissions described by the ME from those generated by the parton shower: hard, well-resolved emissions with $Q > q_{\mathrm{cut}}$ are described by the ME, while softer and more collinear emissions with $Q < q_{\mathrm{cut}}$ are generated by the parton shower.

F.2 Background processes

The dominant and irreducible background is the QCD-induced WZjj-QCD production, which shares the signal final state but is generated at order α_{EW}^{4} ($\alpha_{\mathrm{EW}}^{4}\alpha_{\mathrm{S}}^{2}$ at LO), and has a cross-section roughly three times larger than the signal. A single inclusive sample targeting the $3\ell\,1\nu_{\ell}\,2\mathrm{j}$ final state is used, with the triboson contribution retained as in the signal definition. It is generated with Sherpa (2.2.12 for Run-2, 2.2.16 for Run-3), with all diagrams containing four EW vertices computed at NLO $0,1\mathrm{j}$ + LO $2,3\mathrm{j}$ in QCD using Comix and OpenLoops, merged with the Sherpa parton shower following the MEPS@NLO prescription. The large NLO K-factor of this process (Section 6.5.3) makes an accurate NLO prediction essential to avoid mis-modelling the template distributions, and an alternative MadGraph sample (2.6.5 for Run-2, 3.5.7 for Run-3, interfaced to Pythia 8), with up to two jets at NLO merged via the FxFx scheme [437], is produced for comparison. For Run-3, the modelling uncertainty of Section 7.2.2 is derived from the nominal Sherpa sample reweighted at truth level to the MadGraph+Pythia shape, instead of directly from the alternative sample.

The interference between WZjj-EW and WZjj-QCD, of order $\mathcal{O}(\alpha_{\mathrm{EW}}^{5})$, is generated at LO with MadGraph (2.6.5 for Run-2, 3.5.7 for Run-3) using the same parametrisation and cuts as the WZjj-EW samples but requiring the squared QCD coupling order to be two at the squared ME level, that is $\mathcal{O}(\alpha_{\mathrm{S}})$ in the cross-section. It is positive and amounts to about $6\,\%$ (Run-2) and $7\,\%$ (Run-3) of the WZjj-EW contribution in the fiducial region, rising to at most $\sim 12\,\%$ for events with three or more jets with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$.

The remaining backgrounds with three or more prompt leptons are estimated from simulation:

Diboson ZZ. The $q\bar{q}$- and gg-initiated ZZ processes, including the $\mathrm{gg} \to \mathrm{H} \to \mathrm{ZZ}$ contribution, are generated with Sherpa (2.2.2 for Run-2, 2.2.16 for Run-3) at NLO $0,1\mathrm{j}$ + LO $2,3\mathrm{j}$ in QCD.

EW and QCD ZZjj. The EW ZZjj-EW (order α_{EW}^{6}) and QCD-induced ZZjj-QCD (order α_{EW}^{4}) production of two jets in association with a ZZ pair contribute when one lepton is lost. They are generated with MadGraph5_aMC@NLO 2.6.7 for the Run-2 ZZjj-EW sample, Sherpa 2.2.14 for the Run-3 one, and Sherpa (2.2.2 / 2.2.16) for ZZjj-QCD, consistently with the inclusive ZZ treatment.

Triboson VVV. The on-shell, fully leptonic WWW, WWZ, WZZ, and ZZZ processes, restricted to final states beyond the signal one ($> 3\ell$ and/or $> 1\nu$), are generated with Sherpa (2.2.2 for Run-2, 2.2.14 for Run-3) at NLO $0\mathrm{j}$ + LO $1,2\mathrm{j}$.

Top-quark backgrounds. The $\mathrm{t\bar{t}Z}$ and $\mathrm{t\bar{t}W}$ processes, collectively denoted $\mathrm{t\bar{t}} + V$, are generated at NLO in QCD with MadGraph5_aMC@NLO 2.3.3 interfaced to Pythia 8 for Run-2, with K-factors of 1.12 and 1.1, respectively, and with MadGraph5_aMC@NLO 3.5.1 interfaced to Pythia 8.309 for Run-3. $\mathrm{t\bar{t}}$ is generated with POWHEG+Pythia 8 at NLO, the dedicated tZj sample with MadGraph5_aMC@NLO (2.3.3 / 3.5.1) at NLO with MadSpin decays, and single top (t-, s-channel, and Wt) with POWHEG+Pythia 8 at NLO.

V+photon and V+jets. The $\mathrm{Z}\gamma$ and $\mathrm{W}\gamma$ processes are generated with Sherpa (2.2.4 / 2.2.14) at NLO, and $\mathrm{Z}+$ jets with POWHEG+Pythia 8 (Run-2, with Photos for the QED FSR) and Sherpa (Run-3). Overlapping events between the $\mathrm{Z}\gamma$ and $\mathrm{Z}+$ jets samples are removed.

The background samples are summarised in Table F.B.

Table F.B: Background samples for Run-2 and Run-3.

Process	Generator (Run-2 / Run-3) + PS	ME accuracy / merging	PDF	Main settings
ZZ	Sherpa 2.2.2 / 2.2.16	NLO $0,1\text{j}$ + LO $2,3\text{j}$	`NNPDF3.0NNLO`	incl. $\text{gg} \to \text{H} \to \text{ZZ}$, default Sherpa settings
ZZjj-EW	MadGraph5_aMC@NLO 2.6.7+Pythia 8.230 / Sherpa 2.2.14	LO ($\mathcal{O}(\alpha_{\text{EW}}^6)$)	`NNPDF3.0NLO` (Run-2), `NNPDF3.0NNLO` (Run-3)	A14 tune [436], PS `NNPDF2.3LO` (Run-2)
ZZjj-QCD	Sherpa 2.2.2 / 2.2.16	NLO $0,1\text{j}$ + LO $2,3\text{j}$	`NNPDF3.0NLO` (Run-2), `NNPDF3.0NNLO` (Run-3)	default Sherpa settings, $\mathcal{O}(\alpha_{\text{EW}}^4)$
VVV	Sherpa 2.2.2 / 2.2.14	NLO 0j + LO $1,2\text{j}$	`NNPDF3.0NNLO`	on-shell, fully leptonic, final states beyond signal ($> 3\ell$ and/or $> 1\nu$)
$\text{t}\bar{\text{t}} + V$	MadGraph5_aMC@NLO 2.3.3+Pythia 8.210 / MadGraph5_aMC@NLO 3.5.1+Pythia 8.309	NLO QCD	`NNPDF3.0NLO` (PS `NNPDF2.3LO`)	$K = 1.12$ ($\text{t}\bar{\text{t}}\text{Z}$), 1.1 ($\text{t}\bar{\text{t}}\text{W}$) for Run-2
$\text{t}\bar{\text{t}}$	POWHEG+Pythia 8 (Run-2 & Run-3)	NLO QCD	`NNPDF3.0NLO` (PS `NNPDF2.3LO`)	A14 tune [436], $K = 1.14$
single top (t-, s-channel, Wt)	POWHEG+Pythia 8 (Run-2 & Run-3)	NLO QCD	`NNPDF3.0NLO`	A14 tune [436], enters mainly via misidentified leptons
tZj	MadGraph5_aMC@NLO 2.3.3+Pythia 8.230 / MadGraph5_aMC@NLO 3.5.1+Pythia 8.309	NLO QCD (MadSpin decays)	`NNPDF3.0NLO` (PS `NNPDF2.3LO`)	dedicated sample, overlap with the 5FNS WZjj-EW alternative removed
$\text{Z}\gamma$, $\text{W}\gamma$	Sherpa 2.2.4 / 2.2.14	NLO	`NNPDF3.0NNLO`	overlap removal with $\text{Z}+$ jets
$\text{Z}+$ jets	POWHEG+Pythia 8 / Sherpa	NLO ($K = 1.026$ to NNLO, Run-2)	`CT10` (Run-2)	Photos QED FSR (Run-2)

G 3D template

Among the various templates tested to find the best configuration for the polarisation discrimination, a 3D template was also tested. The idea was to use the output of two DNNs similar to those used in the main analysis:

- The first DNN was trained to separate the WZjj-EW signal from the main background (S. vs B.).
- The second DNN was trained to separate the four polarisation states at the same time using four output nodes and a softmax activation function similar to that of the S. vs B. DNN.

From the output of the two DNNs a 3D template is built using the S. vs B. DNN output as the z-axis and the two polarisation outputs as the x and y axes.

The two coordinates x and y are constructed from the four outputs p_{00}, $p_{0\mathrm{T}}$, $p_{\mathrm{T}0}$, and p_{TT} of the joint-polarisation classifier as:

$$x = \frac{p_{00} + p_{0\mathrm{T}}\,(1 - p_{\mathrm{T}0})}{p_{00} + p_{0\mathrm{T}} + p_{\mathrm{T}0} + p_{\mathrm{TT}}}\,, \tag{G.1}$$

$$y = \frac{p_{00} + p_{\mathrm{T}0}\,(1 - p_{0\mathrm{T}})}{p_{00} + p_{0\mathrm{T}} + p_{\mathrm{T}0} + p_{\mathrm{TT}}}\,, \tag{G.2}$$

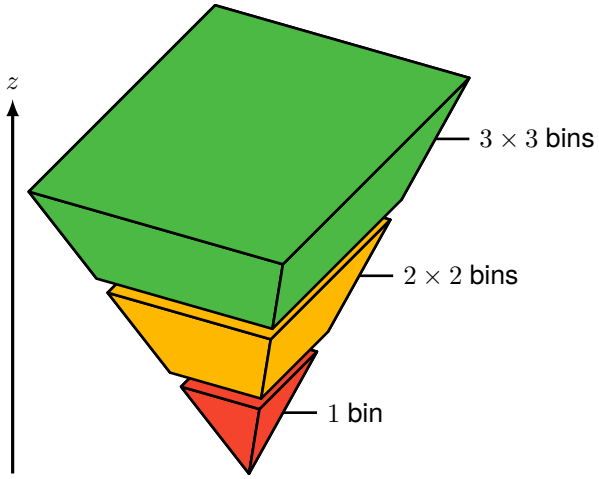
where the first (second) index denotes the polarisation state of the W (Z) boson. The resulting (x, y) plane is shown in Figure G.1(b).

The z-axis, on the other hand, is defined by the S. vs B. DNN output, which is used to separate the signal from the background.

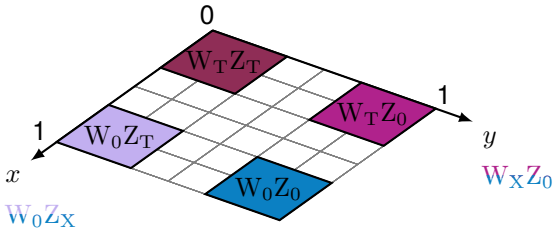
The 3D template is then binned in a pyramidal shape, as shown in Figure G.1(a), with a custom binning along the z-axis. More granularity is needed on the high values of the z-axis, where the signal is more concentrated, while a coarser binning is sufficient for the low values of the z-axis, where the background dominates.

This template therefore allows a simultaneous polarisation measurement of the four states using a single fit, which is not possible with the 2D template used in the main analysis. Three POIs are needed to measure the three polarisation fractions, the fourth being constrained by the sum of the four fractions being equal to one. This is a more complex fit than the one used in the main analysis, which only needs one POI to measure the targeted polarisation fraction.

Despite having an elegant solution to measure the four polarisation fractions simultaneously, this template does not provide enough discrimination power to achieve sufficient significance on the longitudinal polarisation. Indeed, a multi-output DNN for the polarisation separation performs less well than a single-output DNN trained to separate one polarisation from the others. The $\mathrm{W_TZ_Tjj}$-EW polarisation is the most frequent with over $50\,\%$ of the total signal, while the $\mathrm{W_0Z_0jj}$-EW polarisation is the least frequent with less than $10\,\%$ of the total signal. The multi-output DNN therefore struggles to distinguish the longitudinal polarisation from the others, which results in a poor significance for this polarisation.



(a) Pyramidal binning of the 3D discriminant space



(b) Projection onto the (x, y) plane

Figure G.1: Proposed binning scheme in the 3D space constructed from the outputs of the two DNNs. The (x, y) plane is defined by the four polarisation outputs of the joint classifier, while the remaining axis is given by the S. vs B. DNN output.

H Bayesian optimisation of the network hyperparameters

The hyperparameters of each network were tuned with BO. A GP was used as a surrogate for the dependence of a weighted validation metric on the six-dimensional search space:

- the learning rate
- the number of nodes per layer
- the number of hidden layers
- the batch size
- the dropout rate
- the weight decay

and was refined iteratively, each step proposing a new configuration that trades off exploration of the space against exploitation of the most promising regions. The search range for each parameter is summarised in Table H.A. The number of layers scanned excludes a fixed first dense layer added by the training framework, so that the hidden-layer counts reported in Table 6.G are one unit higher than the optimised value.

The procedure follows the description of Section 4.5, where the acquisition function is the UCB with $\kappa = 2$.

Table H.A: Search ranges of the hyperparameters scanned by the Bayesian optimisation.

Hyperparameter	Range	Description
Learning rate	10^{-5}–10^{-2}	Step size for gradient descent optimisation
Nodes per layer	32–1024	Number of neurons in each hidden layer
Number of layers	1–12	Depth of the neural network architecture
Batch size	100–10 000	Number of samples per training iteration
Dropout rate	0.0–0.6	Fraction of neurons randomly deactivated during training
Weight decay	10^{-6}–10^{-2}	Regularisation penalty on network weights

The result is presented for every classifier in the same two-part format. The upper panels show, for each hyperparameter, the value sampled at each iteration coloured by the metric it yields, alongside the evolution of the metric with iteration. The best configuration is ringed and its value annotated. The networks were then trained with rounded values of these optima (Table 6.G). The lower panels show the GP posterior mean and its $95\,\%$ confidence interval as a function of each hyperparameter, with the individual evaluations overlaid. Figure H.1 corresponds to the S. vs B. DNN, optimised over 100 iterations on the weighted validation accuracy, whereas Figures H.2 to H.4 show the single polarisation (Pol. $\mathrm{X}0$ and Pol. $0\mathrm{X}$) and the joint polarisation (Pol. 00) DNNs, each optimised over 50 iterations on the weighted validation AUC. In all cases the metric varies little over most of the scanned range, the flat posterior means revealing a broad plateau of near-optimal configurations, while the isolated evaluations dropping to 0.5 correspond to training runs that failed to converge.

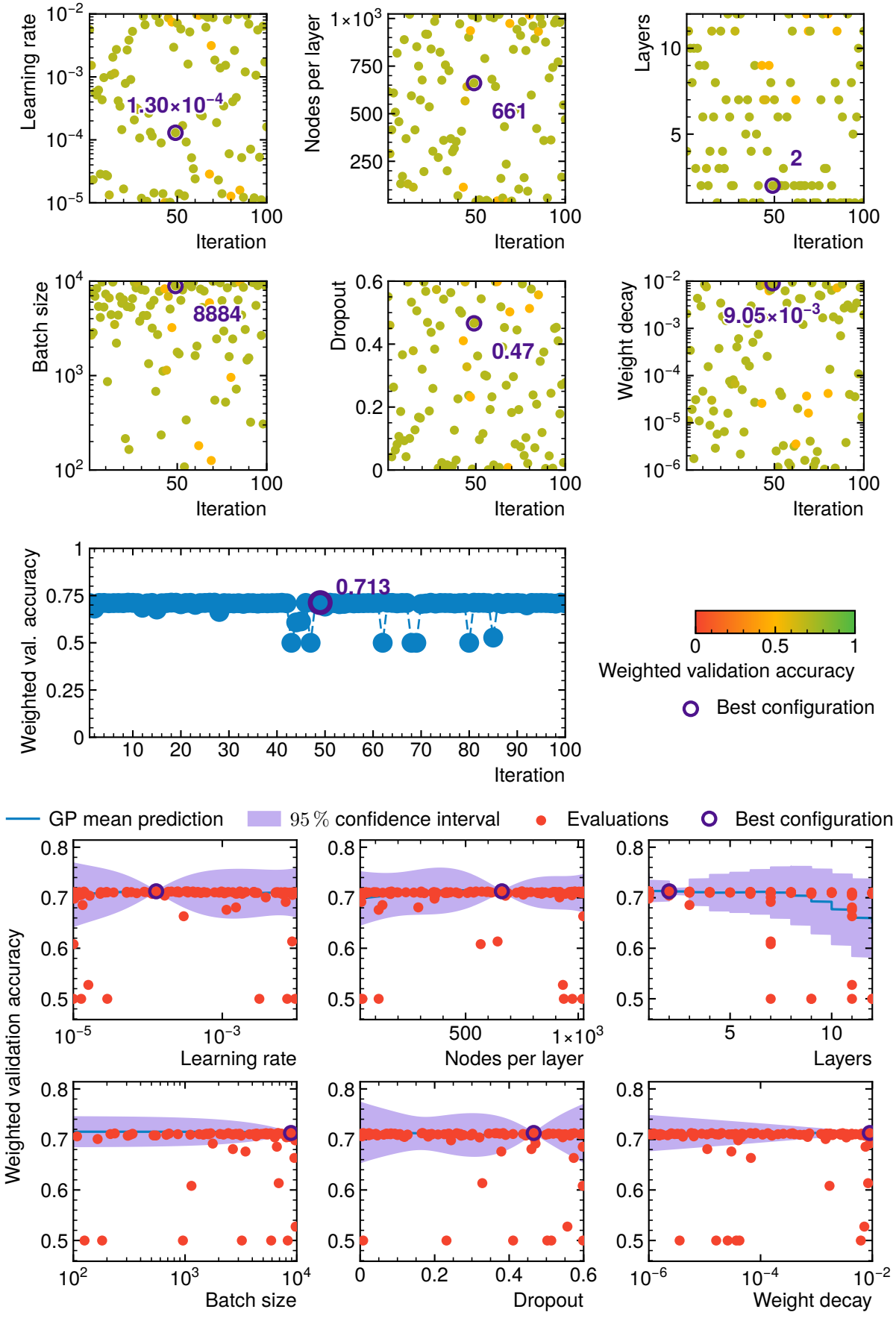


Figure H.1: Hyperparameter optimisation of the S. vs B. DNN.

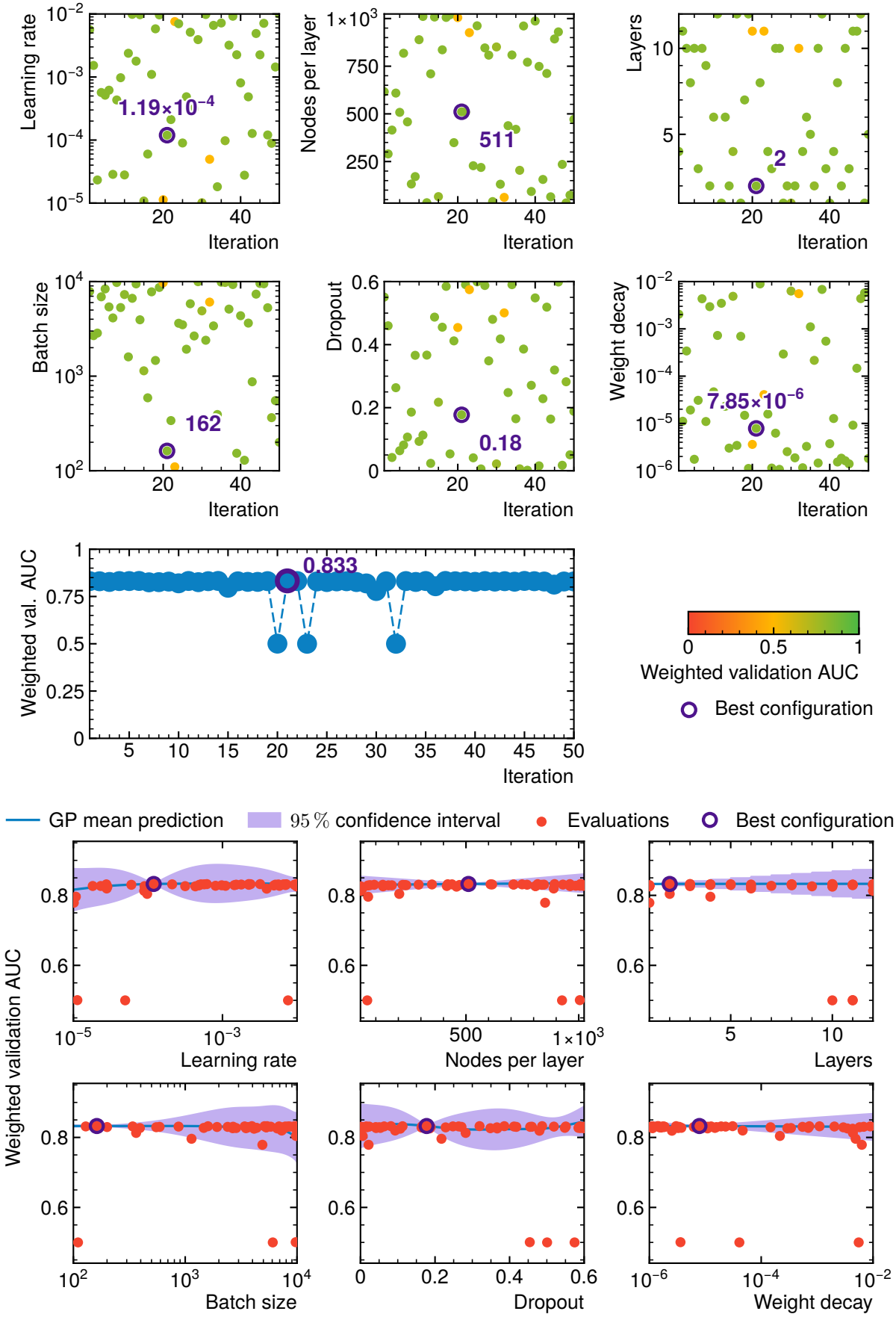


Figure H.2: Hyperparameter optimisation of the Pol. X0 DNN.

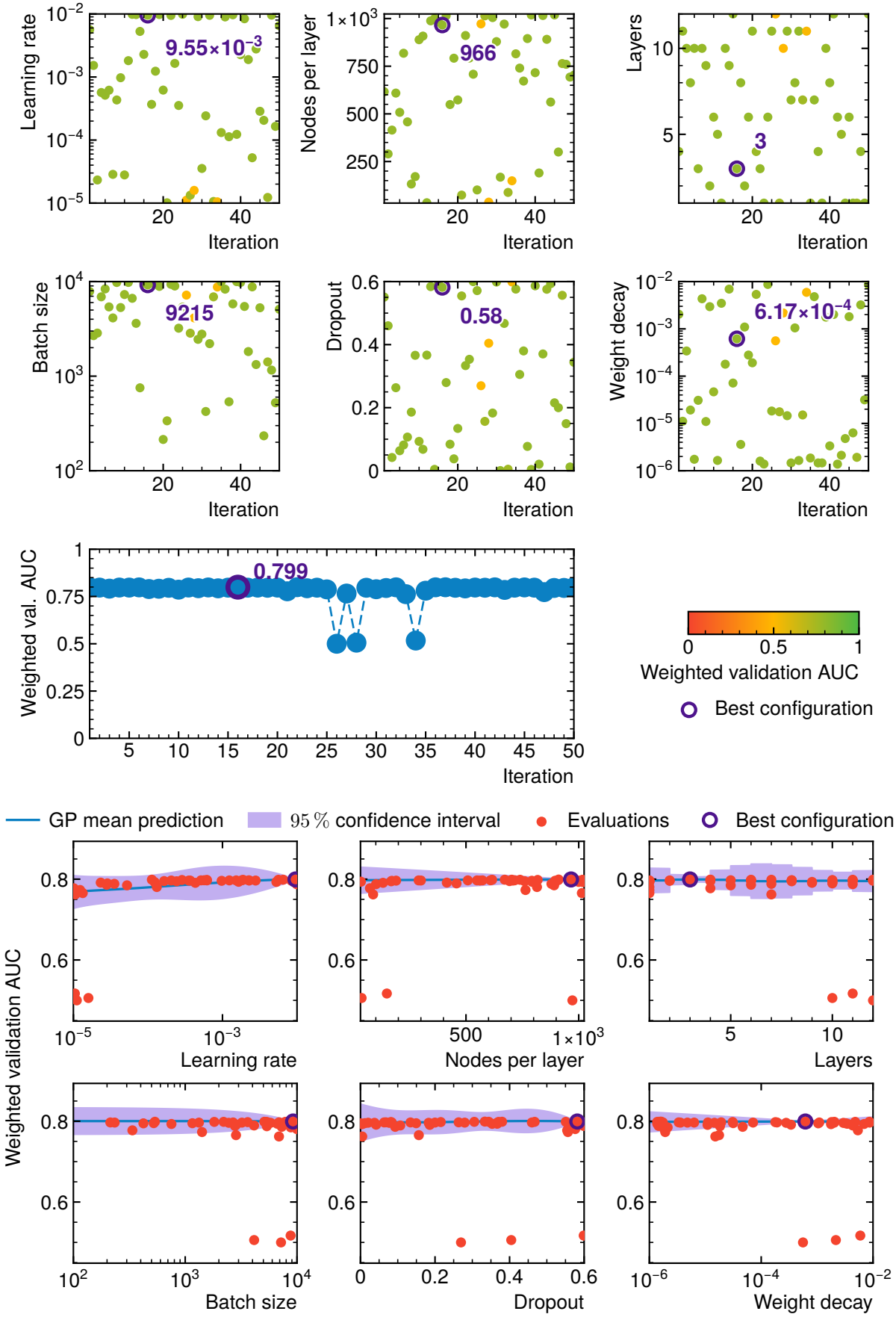


Figure H.3: Hyperparameter optimisation of the Pol. 0X DNN.

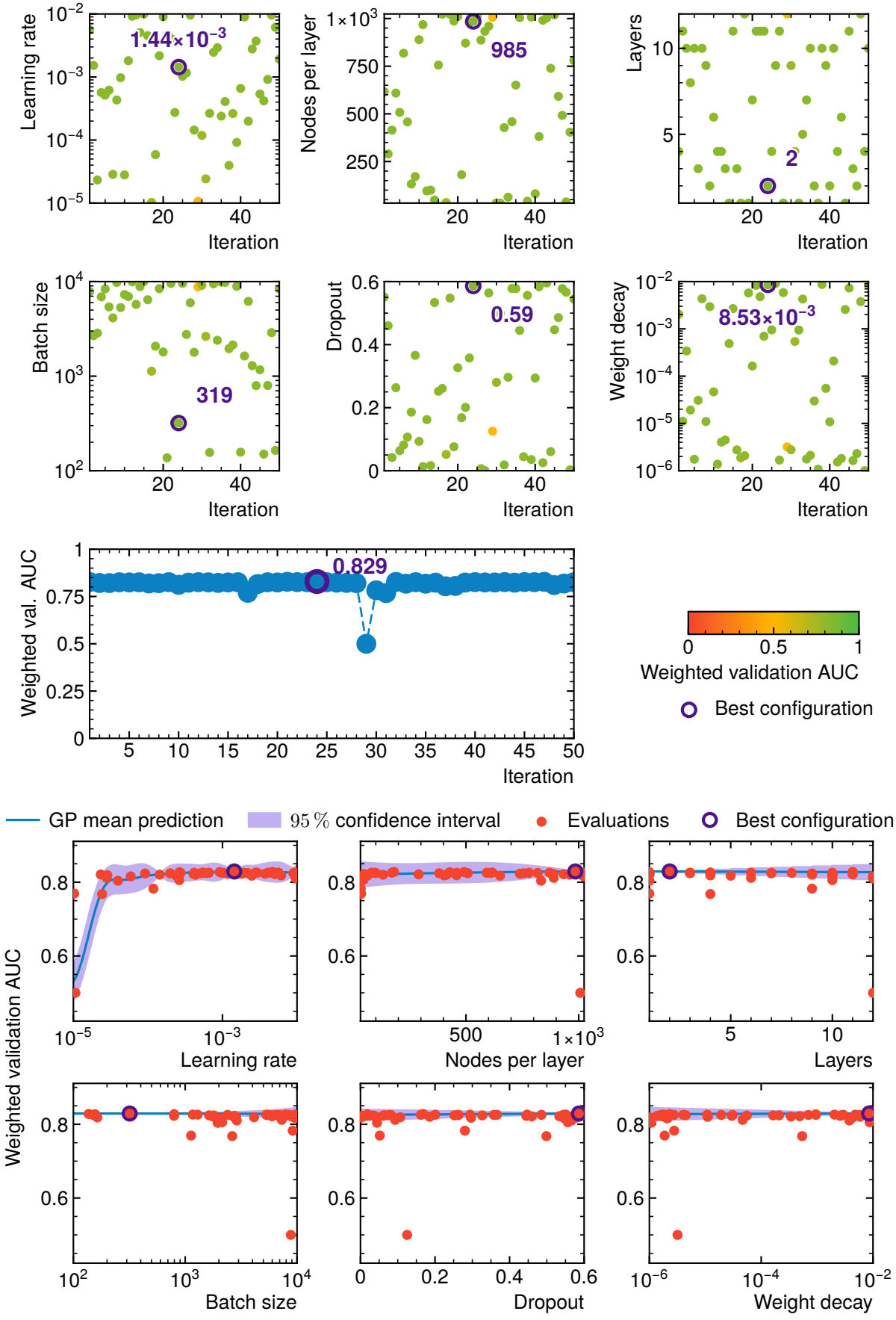


Figure H.4: Hyperparameter optimisation of the Pol. 00 DNN.

Input variable distributions

This appendix shows the full set of input-variable distributions provided to the DNNs, with data overlaid on the stacked MC prediction and the corresponding data-to-prediction ratio panel below each distribution. The variables are shown for two different baselines: the DNN training level and the fit level, for Run-2 and Run-3 separately. Each panel grid is reproduced on its own page for legibility.

The training-level distributions use MadGraph for the WZjj-EW and WZjj-QCD processes since they were already available from the past analysis [360], while the MC samples used for the measurement are generated with Sherpa for a more precise description of the processes.

The training distributions (Figures I.1 and I.2) correspond to the variables seen directly by the DNNs, after the preselection and before any pruning. They include the reconstructed kinematics of the three charged leptons and the two tagging jets, the dijet-system variables that drive the VBS topology, and the angular and centrality variables that separate the EW signal from the QCD background. All are presented in detail in Section 6.2.3. Within the statistical uncertainties of the simulation the prediction describes the data across the full range of every variable, with no localised mis-modelling that could bias the network response.

The fit baseline distributions (Figures I.3 and I.4) are shown after the full event selection used in the measurement. They are provided for completeness, so that the agreement can be inspected in the phase-space that actually enters the polarised-template fit described in Chapter 7.

The shape distributions of the MC samples are also shown for the training and fit baselines, for both Run-2 and Run-3, and for signal and background separately. These distributions are normalised to show the fraction of events per bin, so that the shape of the distributions can be compared independently of the sample normalisation. The signal distributions are shown in Figures I.5 to I.8 and the background distributions in Figures I.9 to I.12. Both baselines are also shown here for comparison.

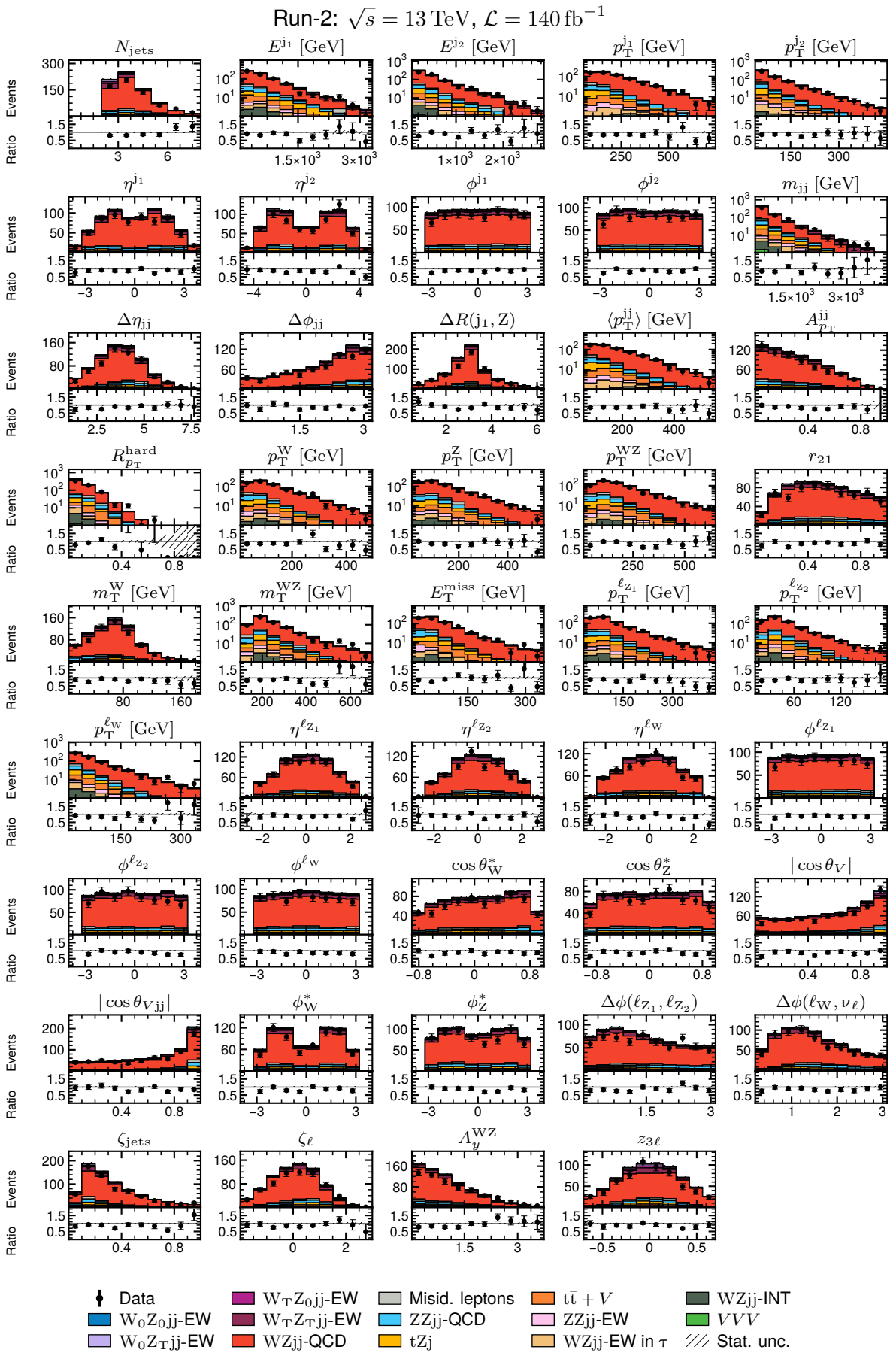


Figure I.1: Distributions of the DNN input variables at training level for Run-2. Data are overlaid on the stacked MC prediction. The lower panel of each pad shows the data-to-prediction ratio.

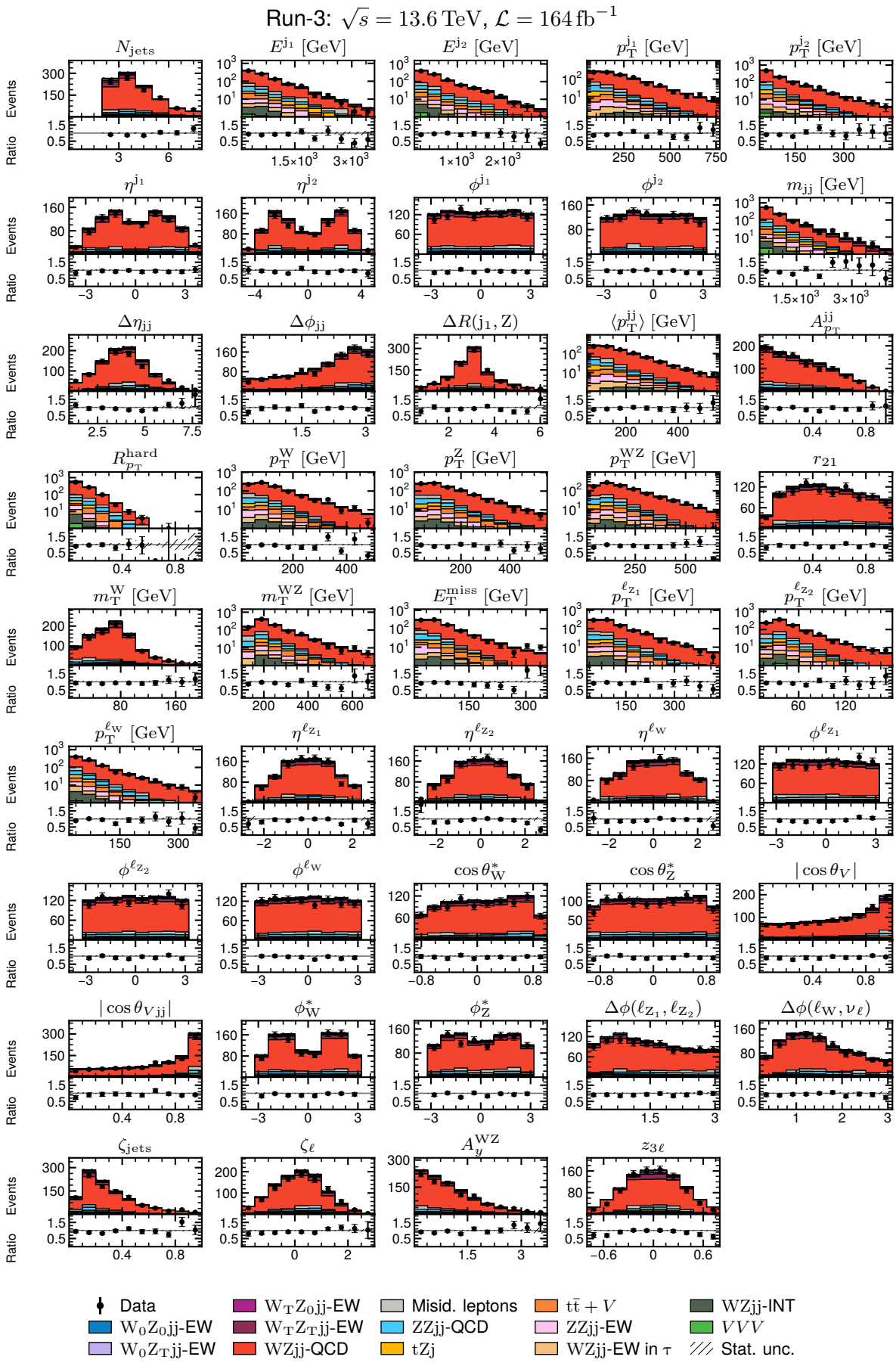


Figure I.2: Distributions of the DNN input variables at training level for Run-3. Data are overlaid on the stacked MC prediction. The lower panel of each pad shows the data-to-prediction ratio.

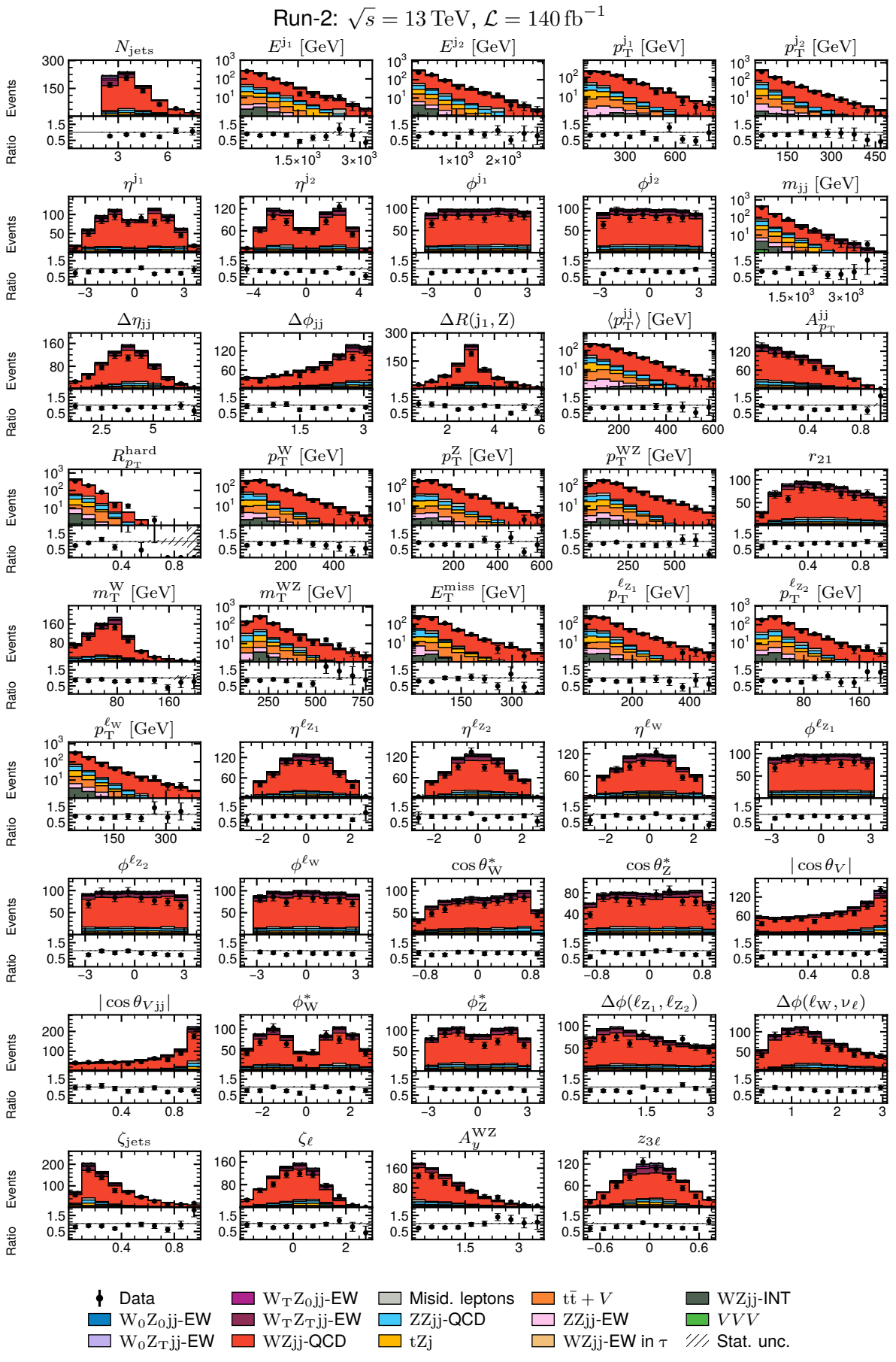


Figure I.3: Distributions of the DNN input variables with the fit baseline for Run-2. Data are overlaid on the stacked MC prediction. The lower panel of each pad shows the data-to-prediction ratio.

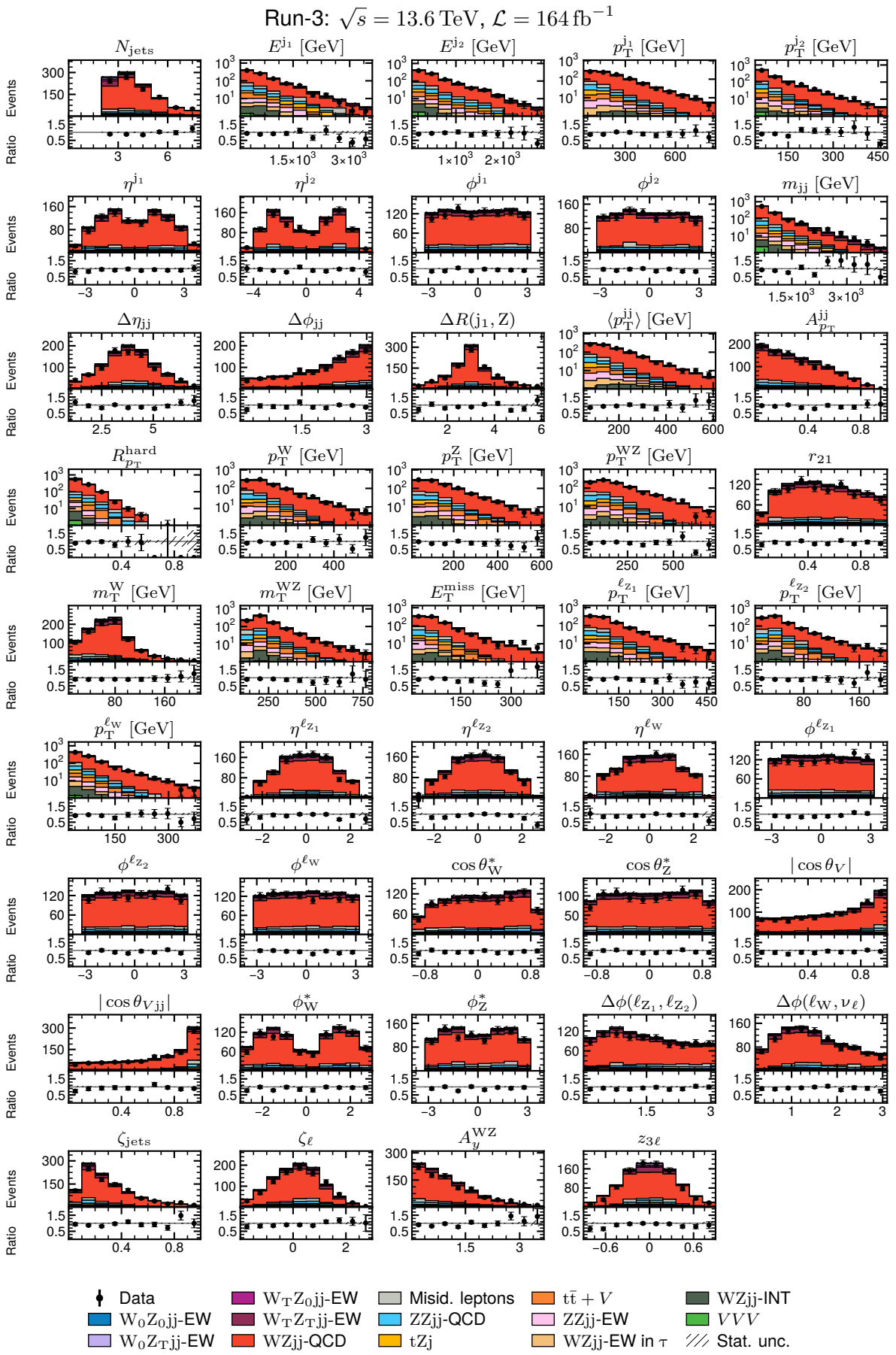


Figure I.4: Distributions of the DNN input variables with the fit baseline for Run-3. Data are overlaid on the stacked MC prediction. The lower panel of each pad shows the data-to-prediction ratio.

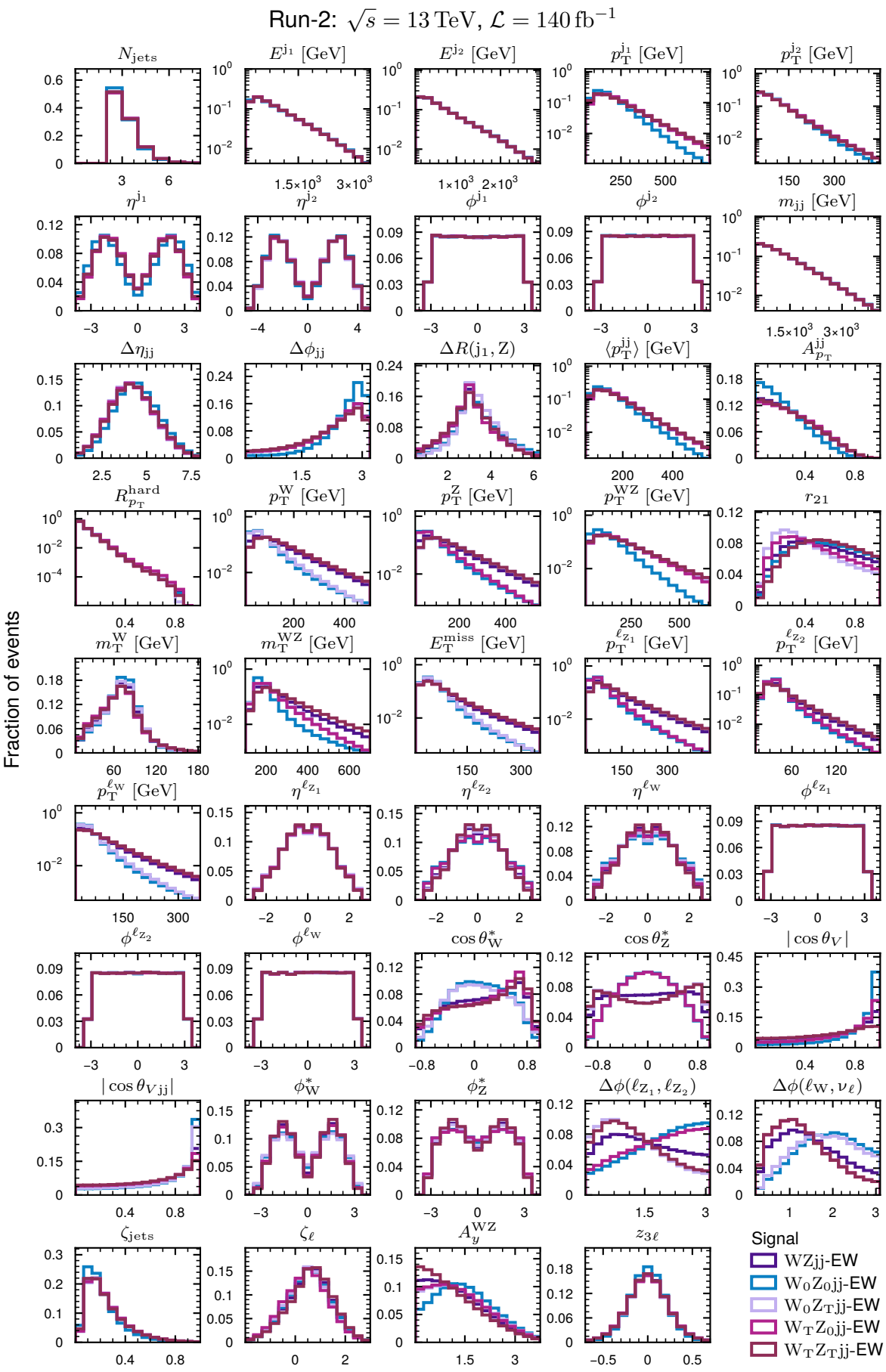


Figure I.5: Distributions of the DNN input variables at training level for Run-2. Shape distributions for signal samples.

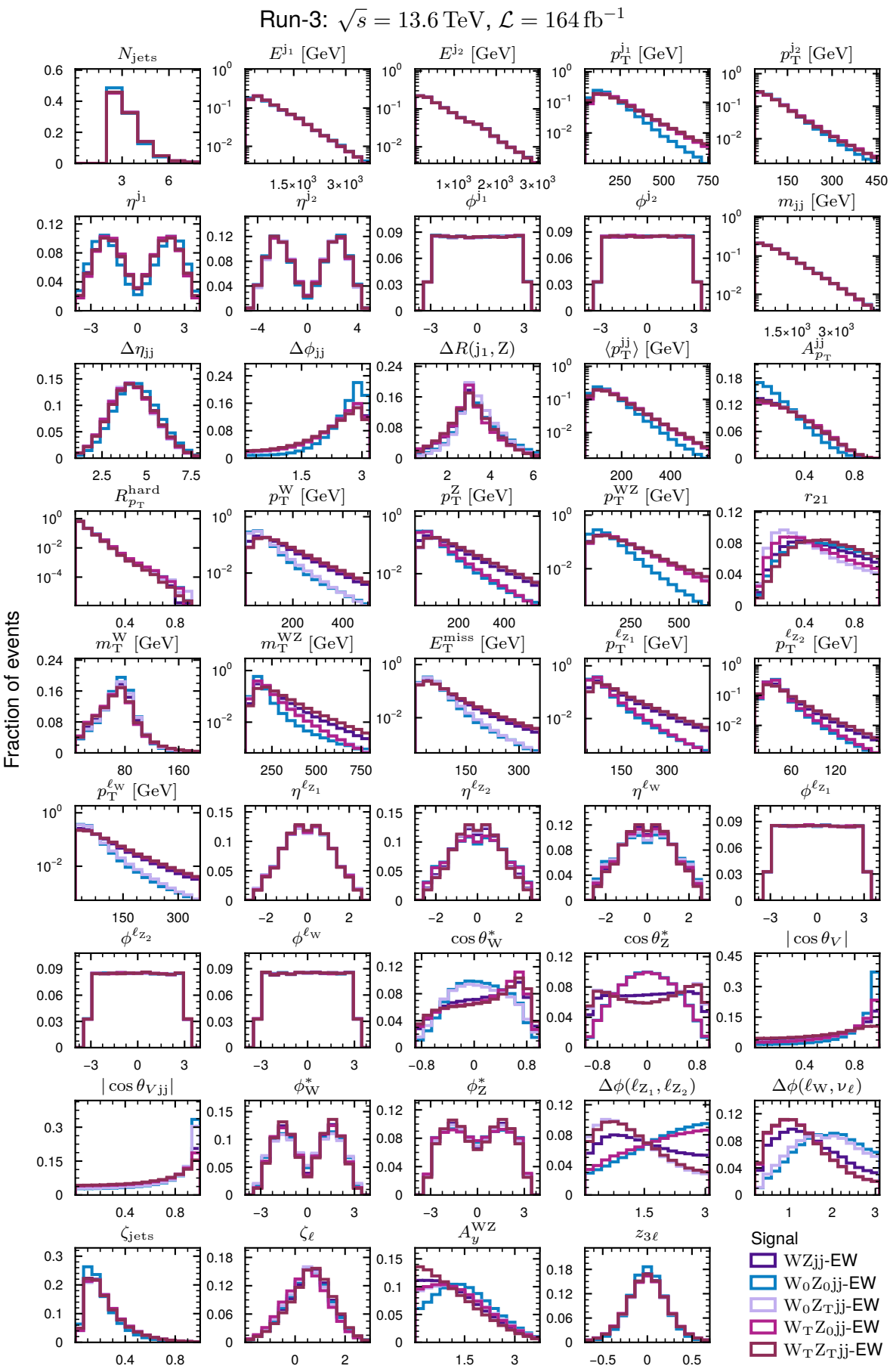


Figure I.6: Distributions of the DNN input variables at training level for Run-3. Shape distributions for signal samples.

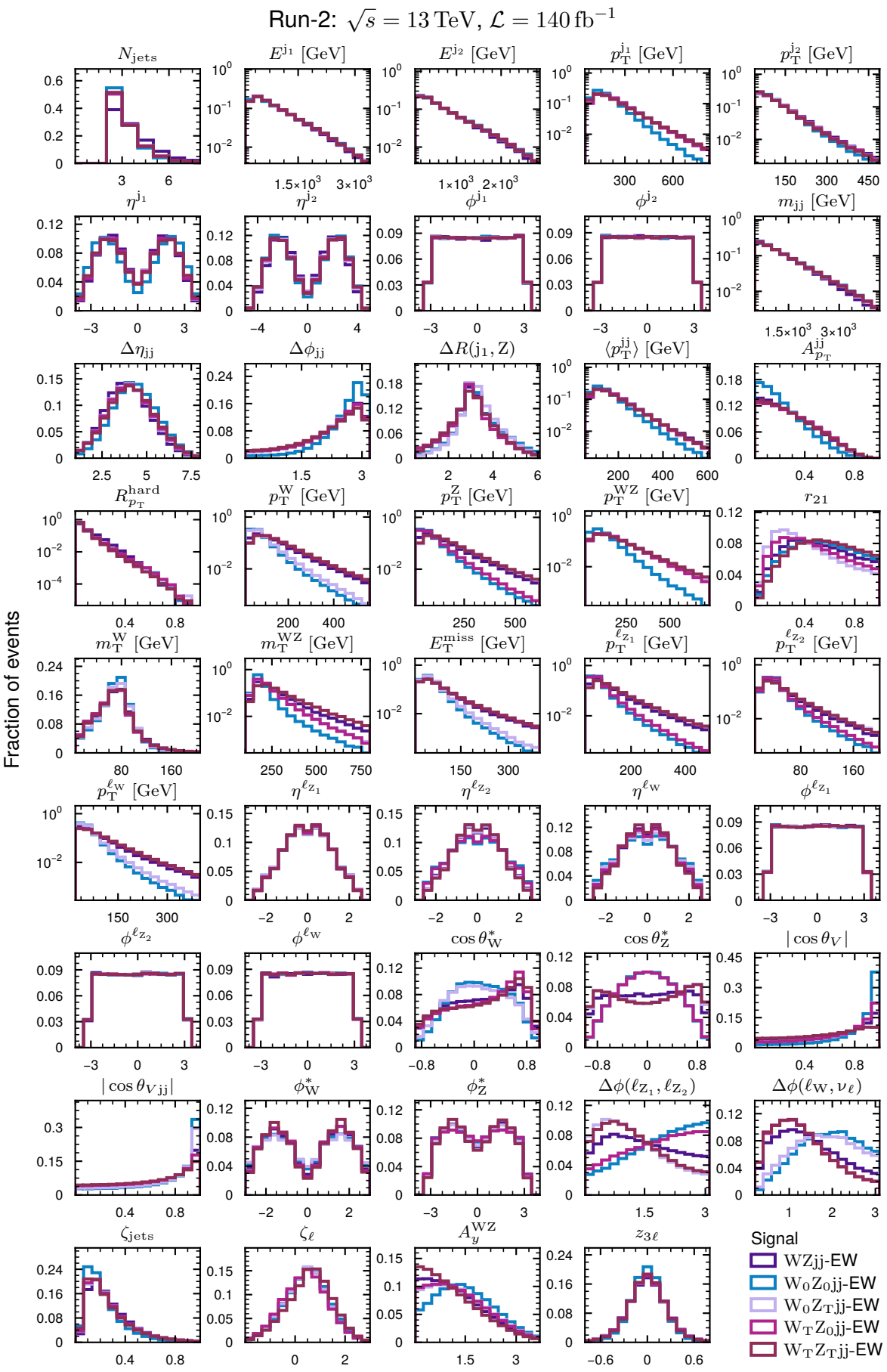


Figure I.7: Distributions of the DNN input variables with the fit baseline for Run-2. Shape distributions for signal samples.

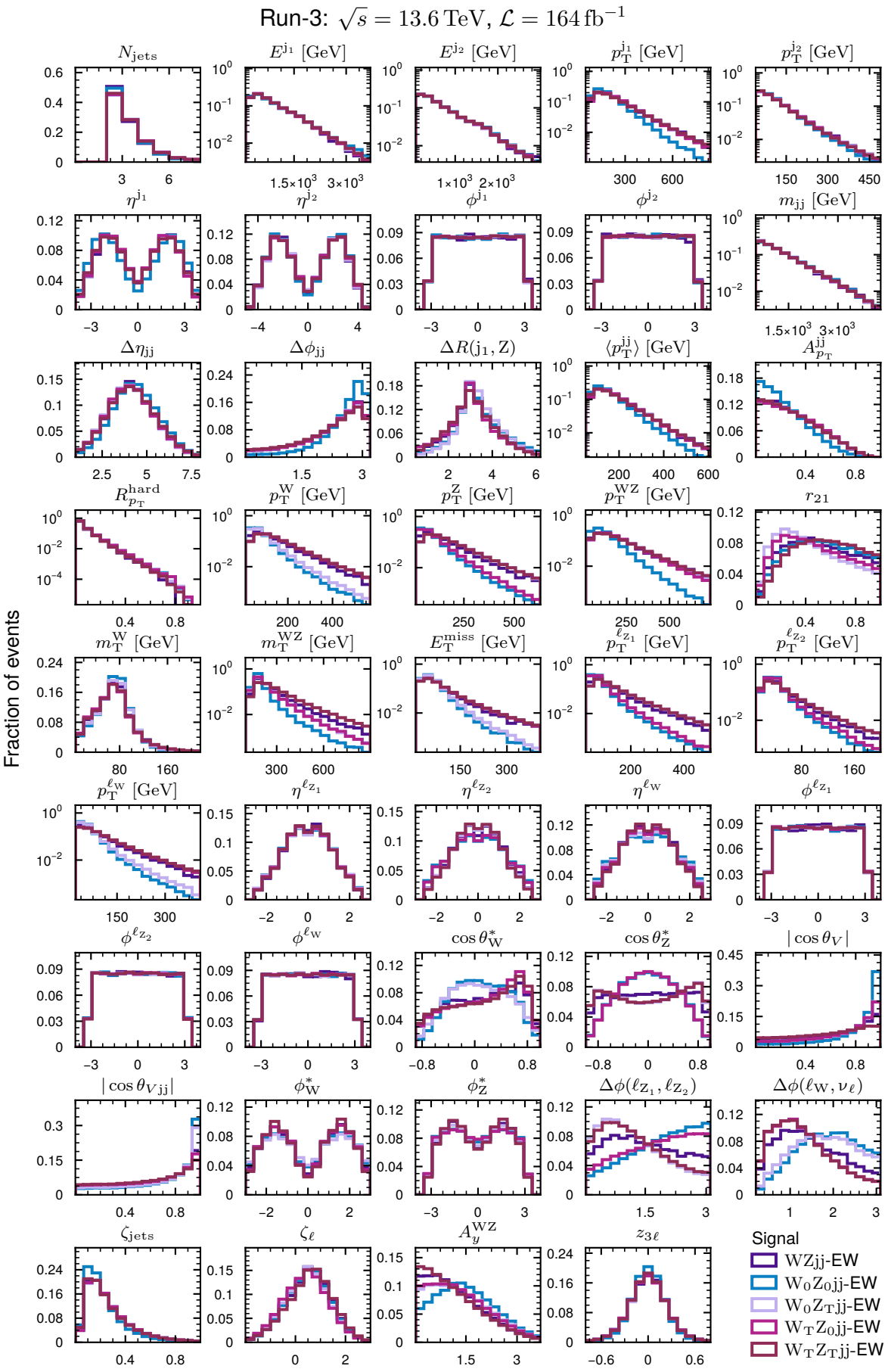


Figure I.8: Distributions of the DNN input variables with the fit baseline for Run-3. Shape distributions for signal samples.

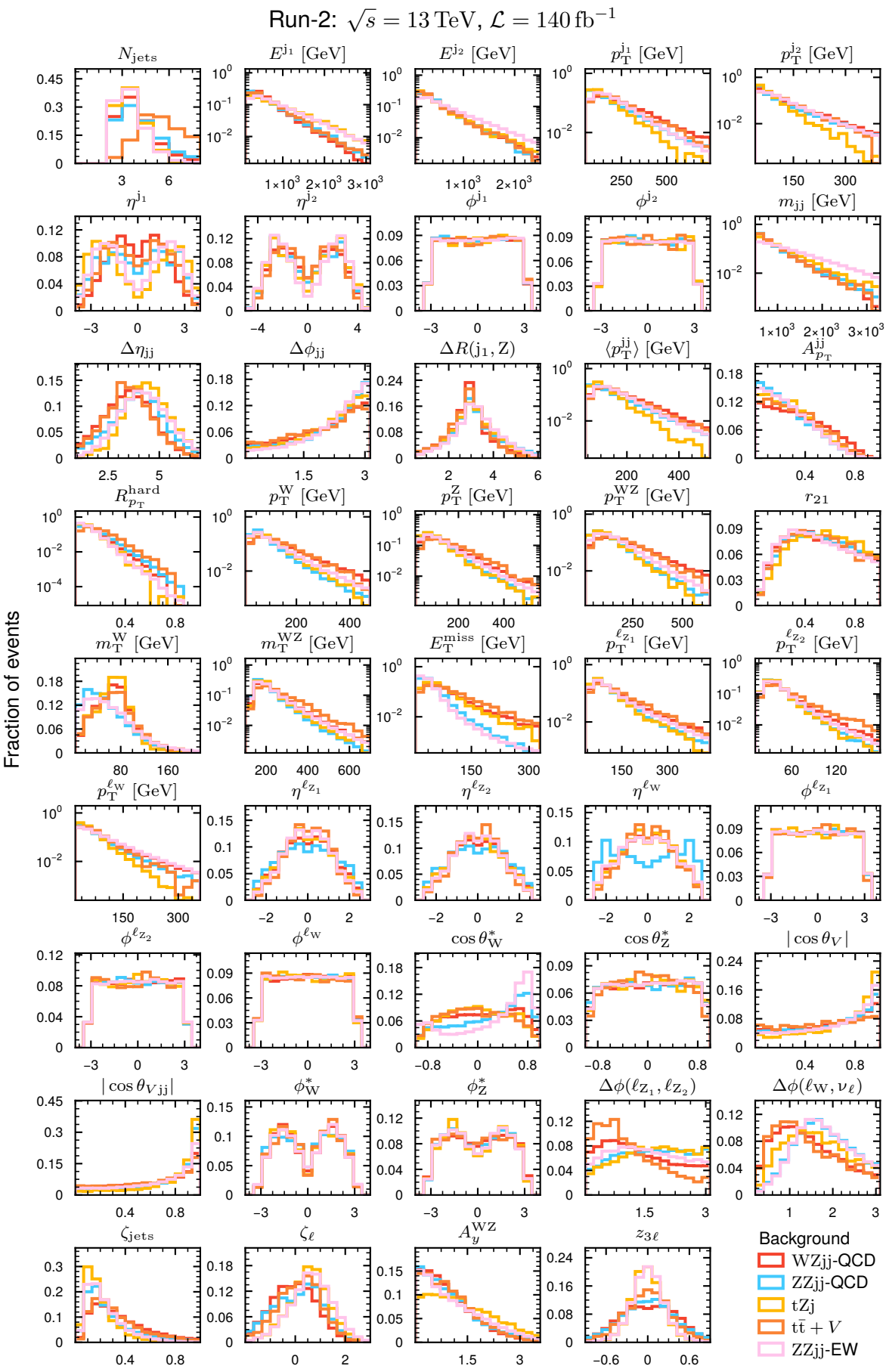


Figure I.9: Distributions of the DNN input variables at training level for Run-2. Shape distributions for background samples.

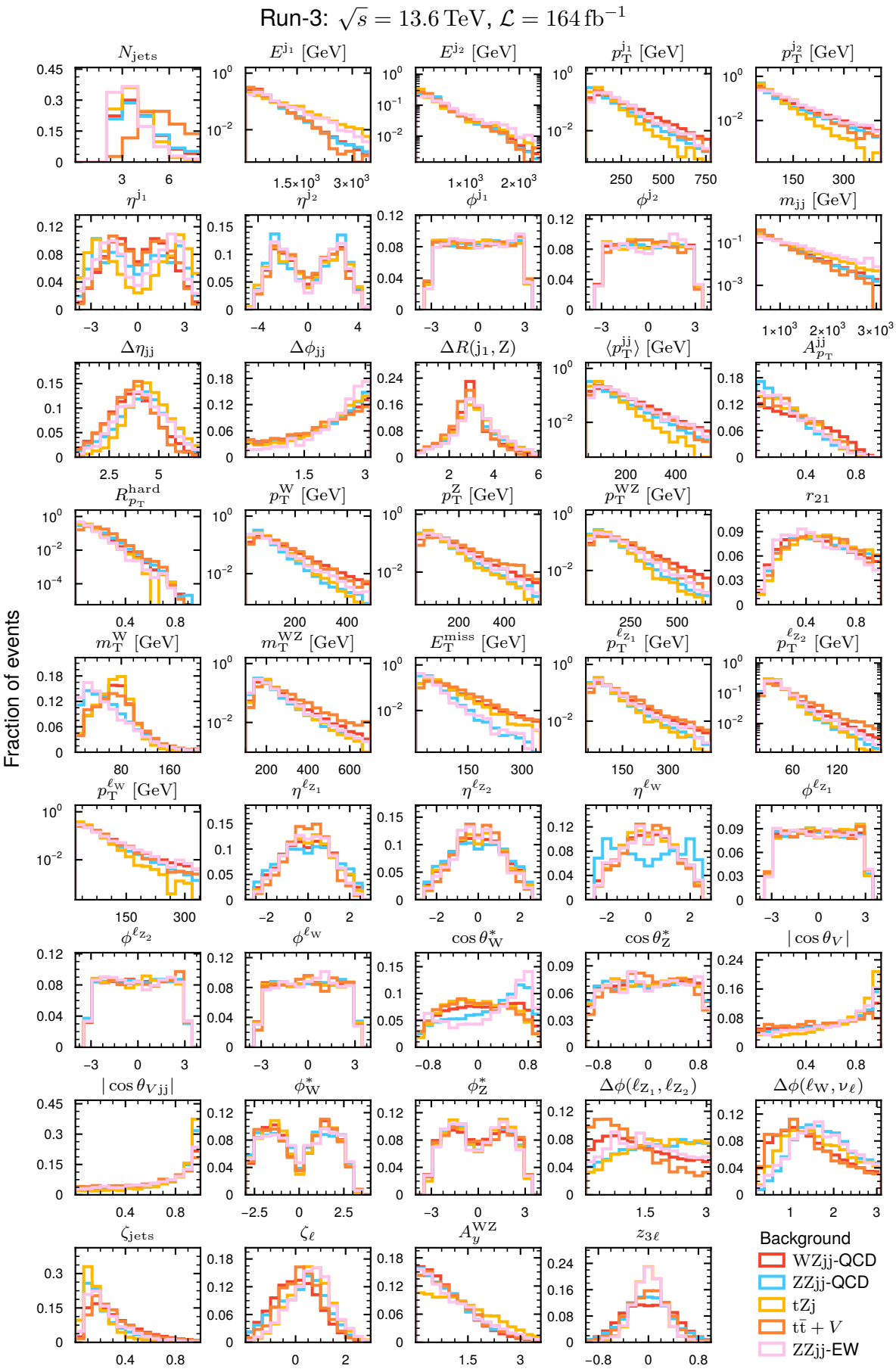


Figure I.10: Distributions of the DNN input variables at training level for Run-3. Shape distributions for background samples.

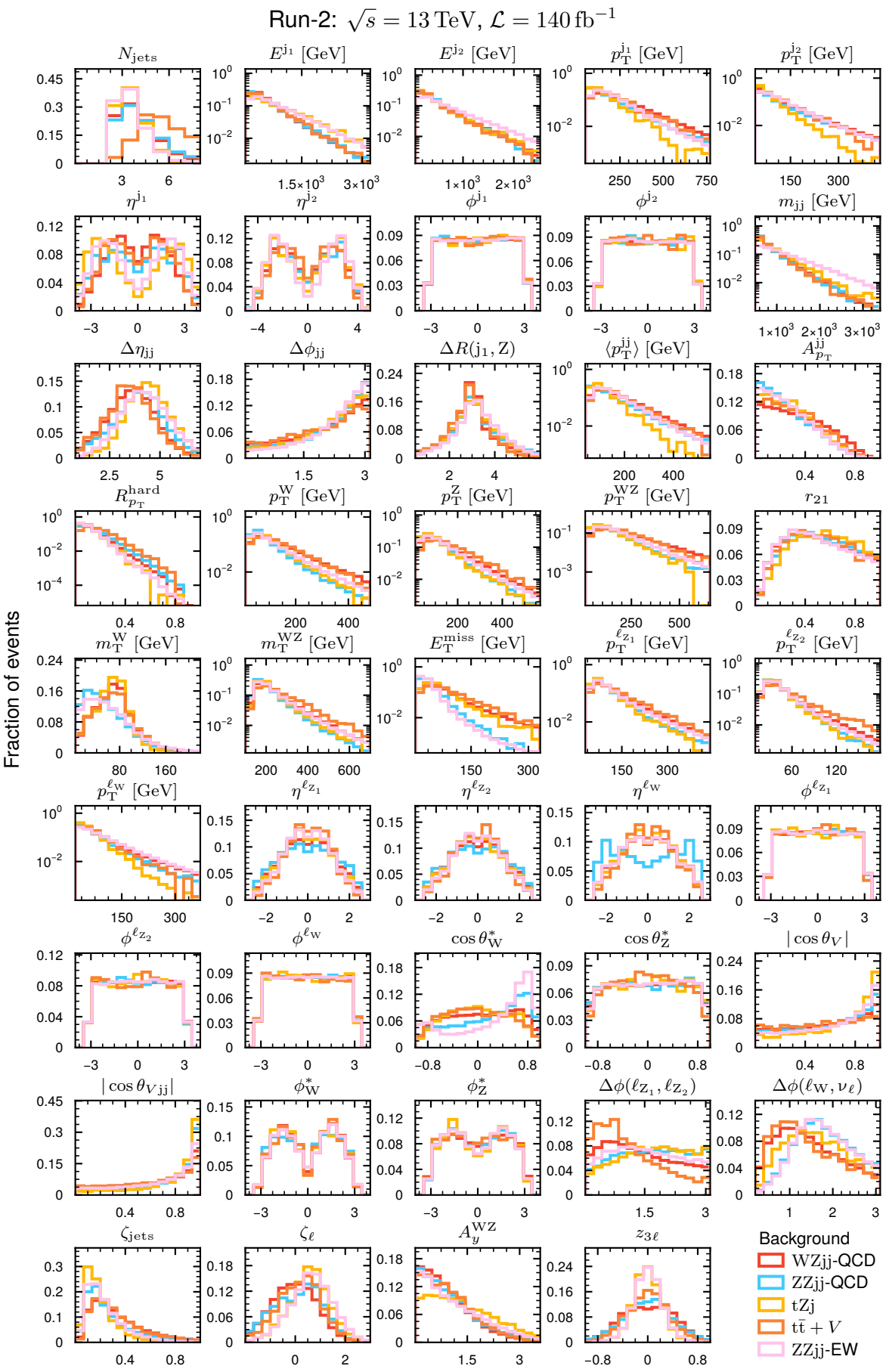


Figure I.11: Distributions of the DNN input variables with the fit baseline for Run-2. Shape distributions for background samples.

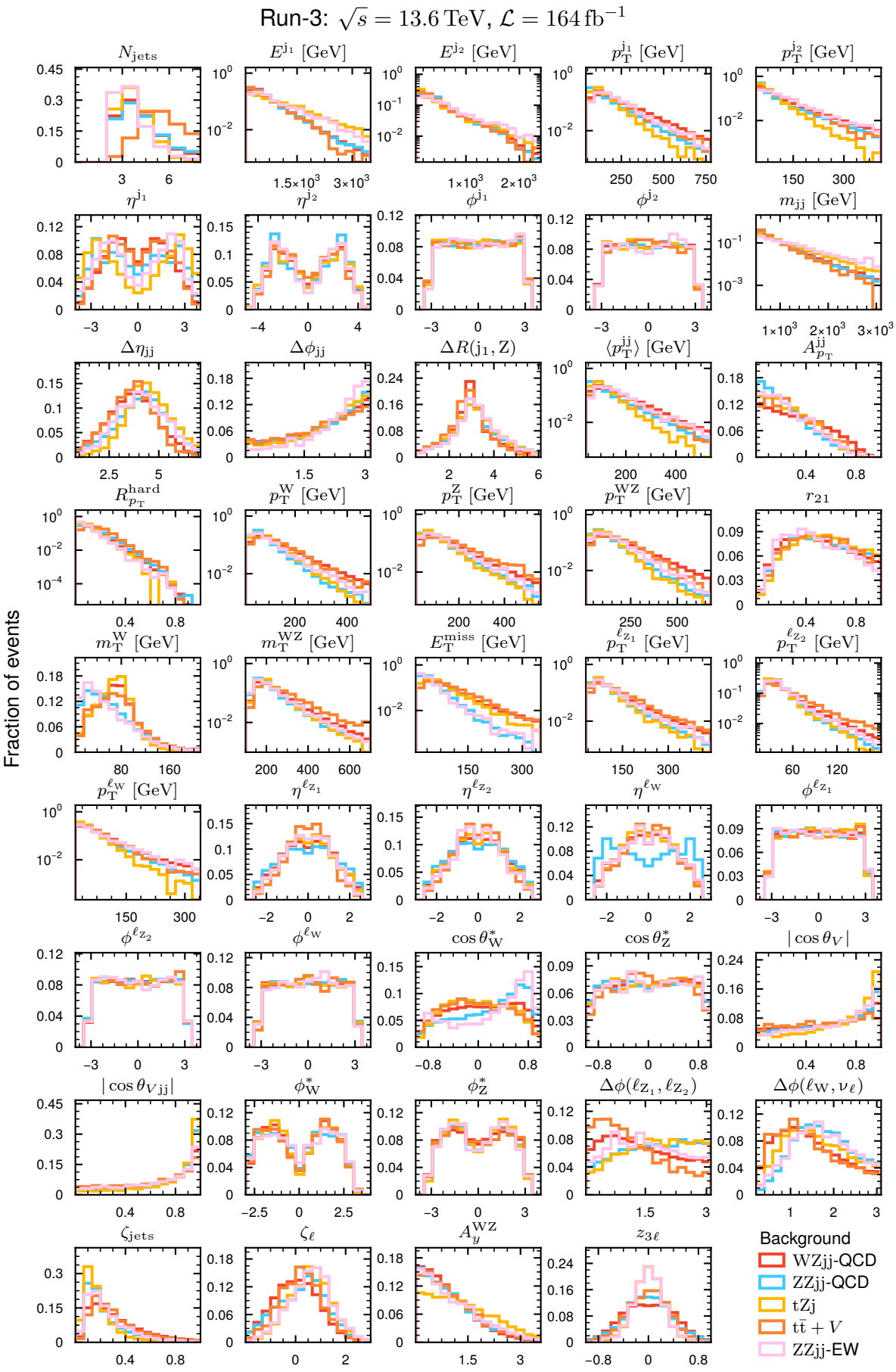


Figure I.12: Distributions of the DNN input variables with the fit baseline for Run-3. Shape distributions for background samples.

J Correlation matrices

This appendix collects the linear (Pearson) correlation matrices of the full set of candidate input variables of Table 6.I, evaluated on the training samples. The S. vs B. DNN inputs are shown for Run-2 and Run-3, and the inputs of the polarisation DNNs for Run-2 (same dataset for the three DNN configurations).

The matrices are shown in Figures J.1 to J.3.

The linear correlation coefficient ρ_{ij} between two input variables x_i and x_j is defined as the ratio of their covariance to the product of their standard deviations:

$$\rho_{ij} = \frac{\mathrm{cov}(x_i, x_j)}{\sigma_i \sigma_j} = \frac{\langle (x_i - \mu_i)(x_j - \mu_j) \rangle}{\sigma_i \sigma_j}\,, \tag{J.1}$$

where μ_i and σ_i are the mean and standard deviation of variable x_i [438]. By construction $\rho_{ij} \in [-1, +1]$, with $\rho_{ij} = 0$ for linearly uncorrelated variables and $|\rho_{ij}| = 1$ for an exact linear relation. The coefficient quantifies only the *linear* component of the dependence: two variables with $\rho_{ij} \approx 0$ may still be statistically dependent through higher-order or non-linear correlations, which the DNNs are able to exploit directly from the joint distributions. For this reason, strongly correlated variables are retained as inputs where they carry complementary information. In the matrices below each coefficient is printed to two decimals with the leading zero omitted, since $|\rho_{ij}| \leq 1$: ".55" reads $\rho_{ij} = 0.55$. Each matrix is shown on its own page.

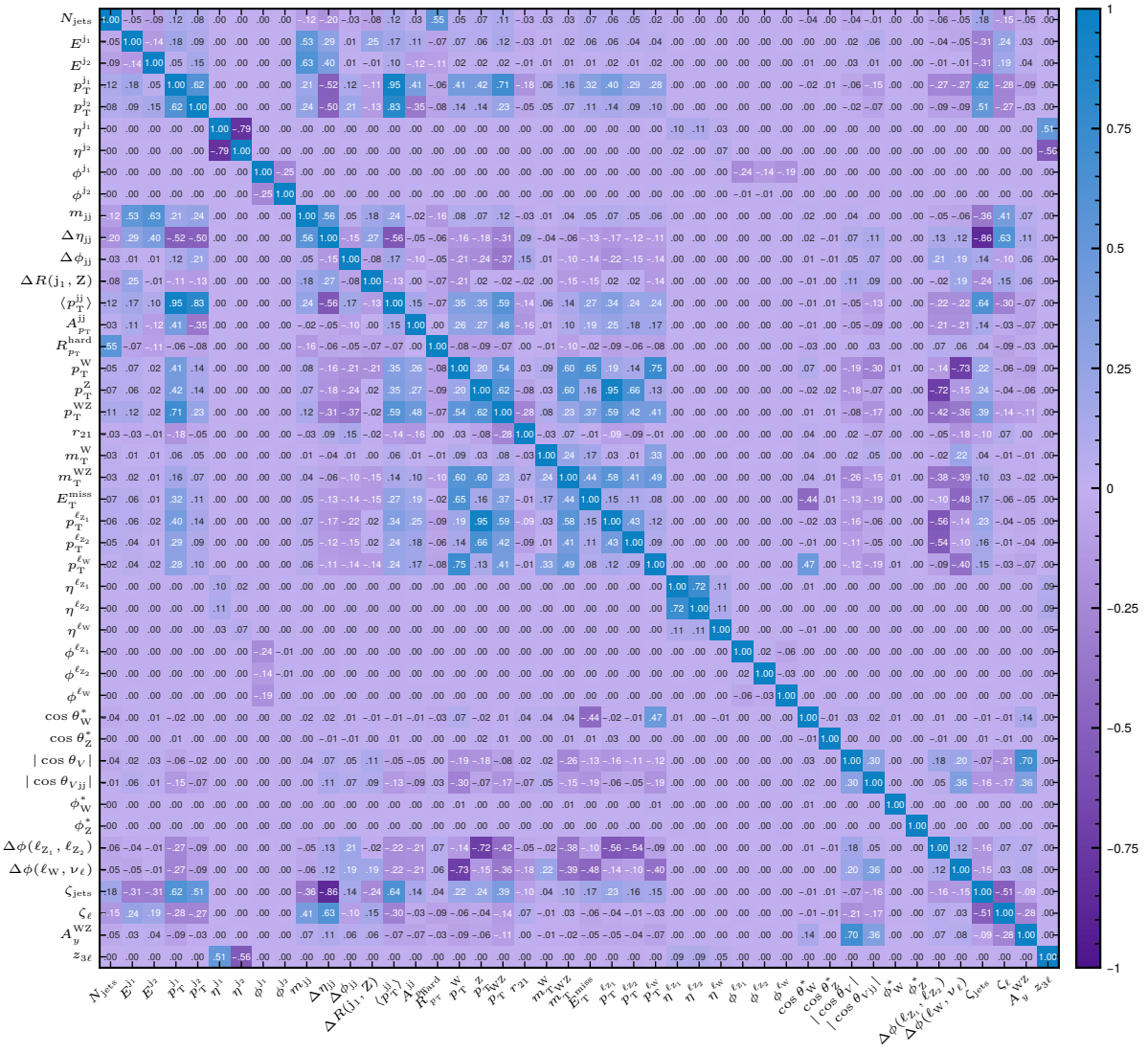


Figure J.1: Input correlation matrix for the S. vs B. DNN for Run-2.

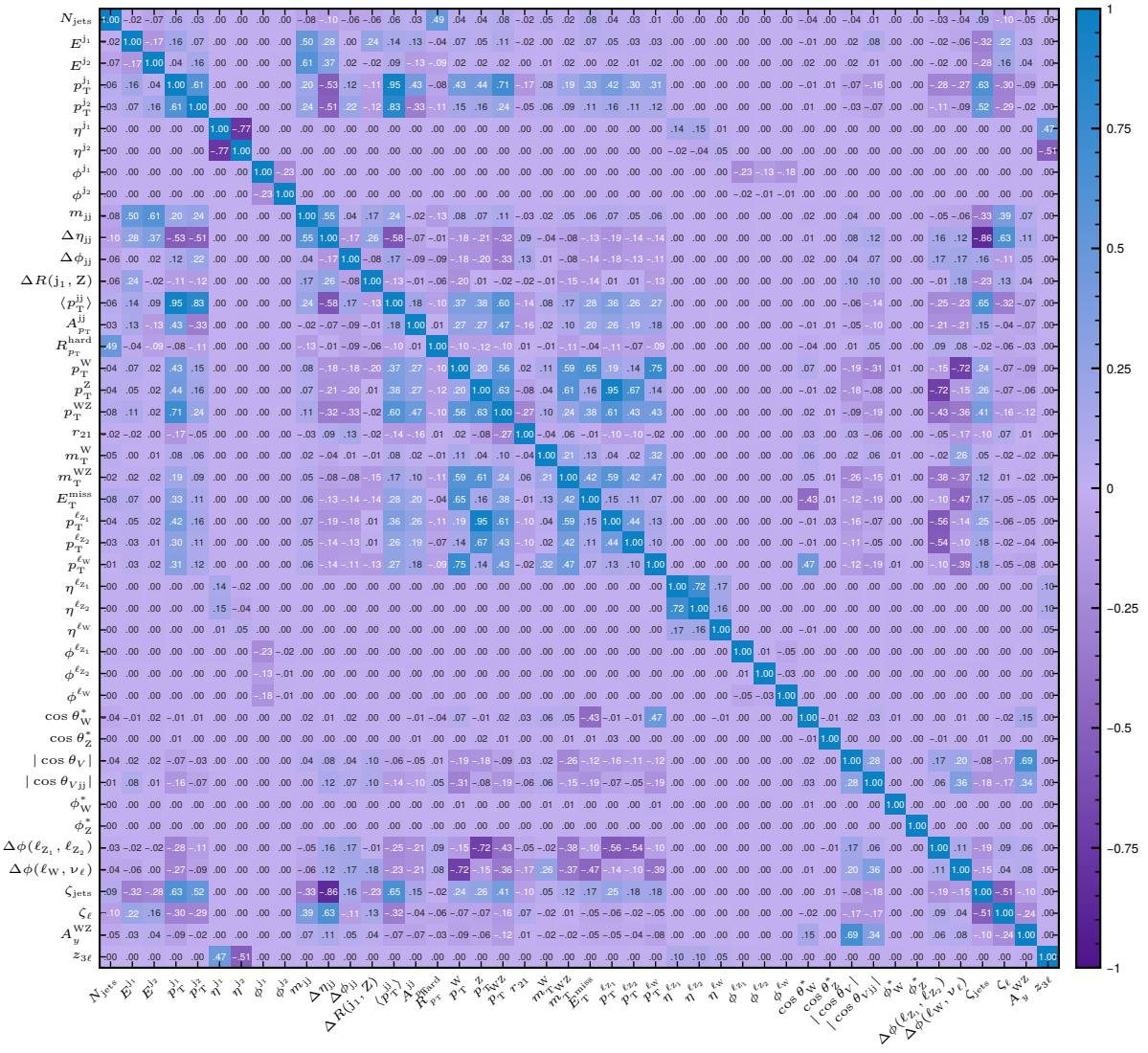


Figure J.2: Input correlation matrix for the S. vs B. DNN for Run-3.

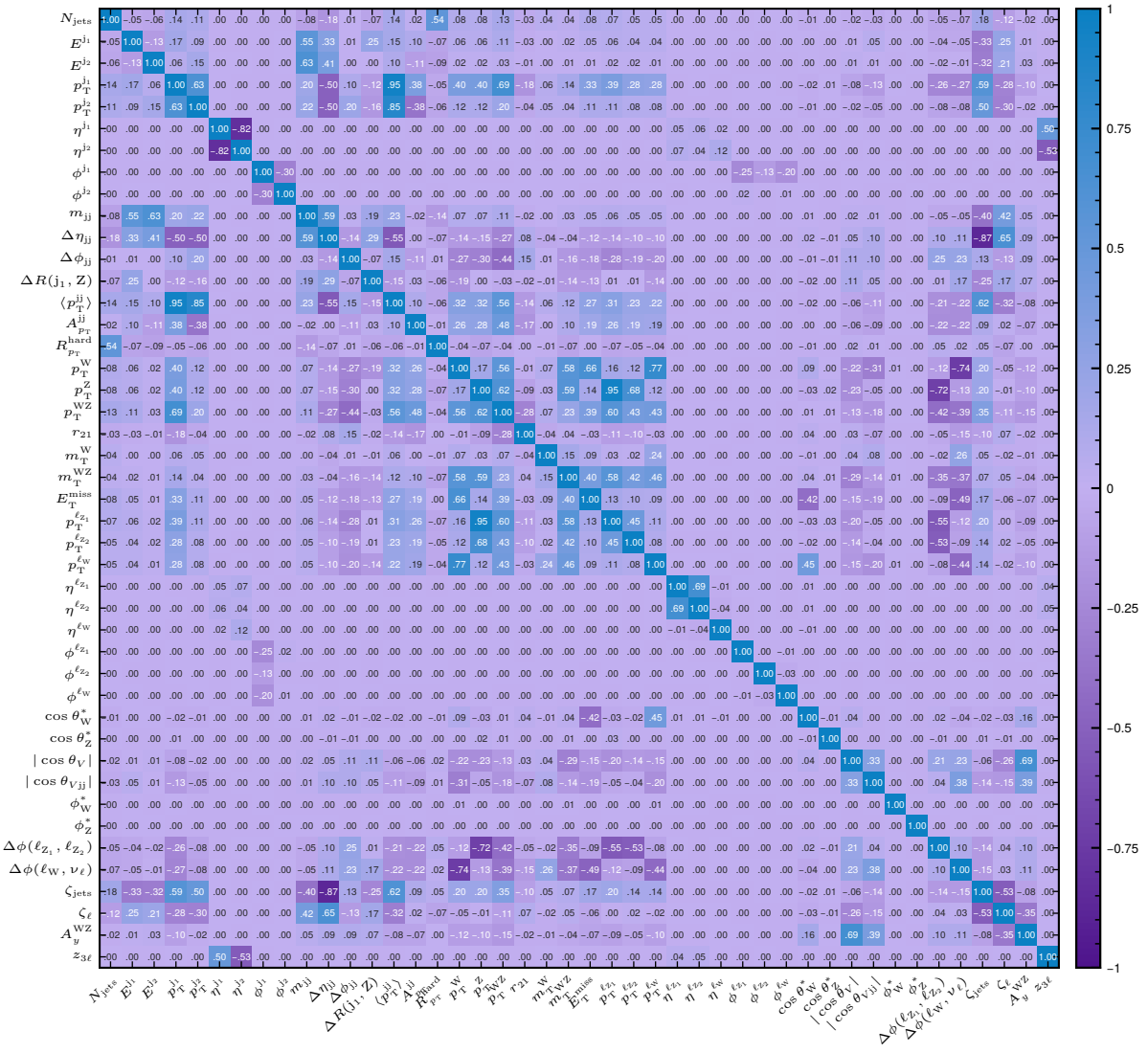


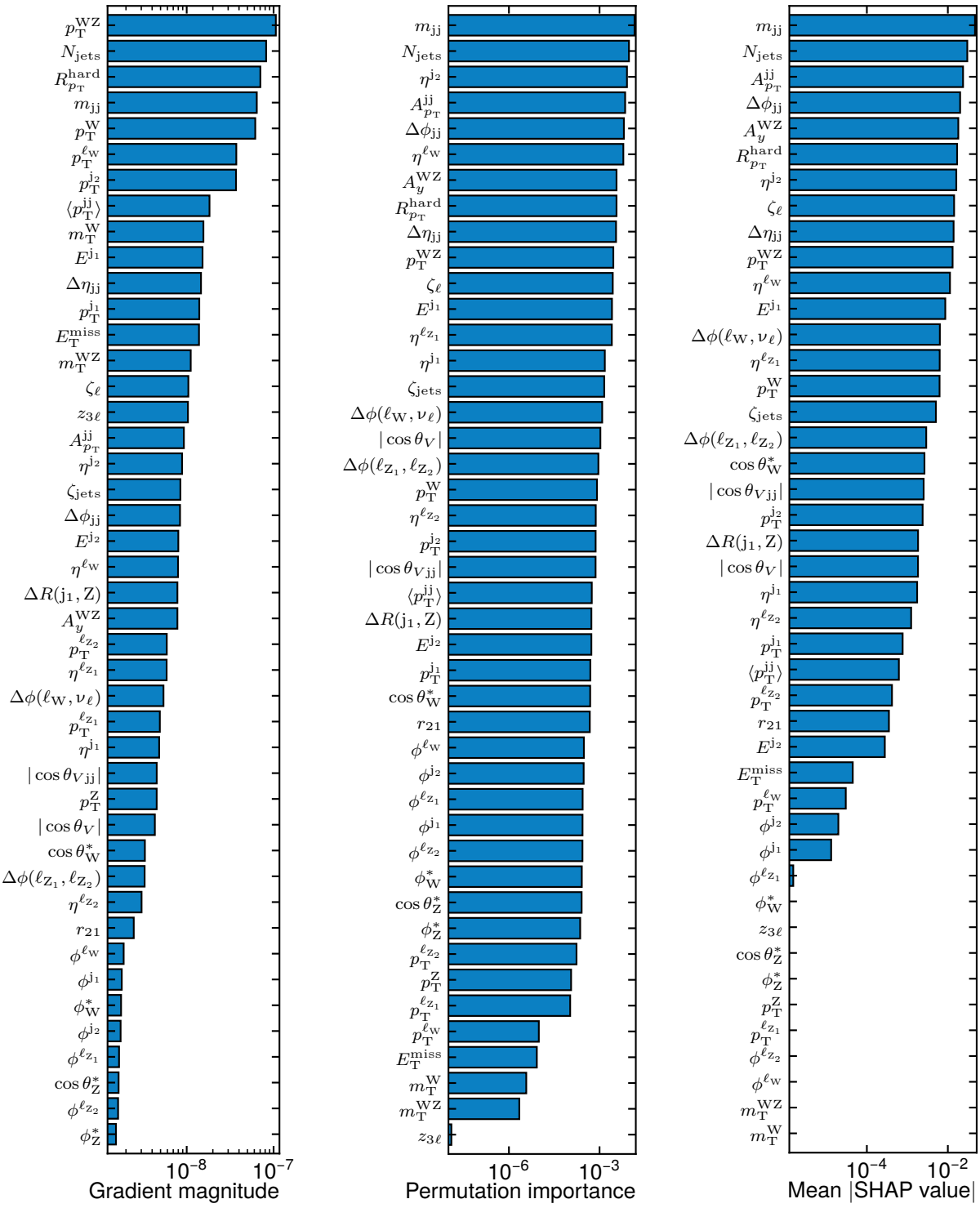
Figure J.3: Input correlation matrix for the polarisation DNNs for Run-2.

K Variable ranking

This appendix collects the supplementary material related to the input-variable ranking and to the classifier performance for the DNNs used in the analysis. For each of them, the variable importance is quantified with three complementary estimators: the gradient magnitude of the DNN output with respect to each input, the permutation importance, and the SHAP values as presented in Section 4.4. The ranking is shown first for the *full* variable set, which motivates the pruning, and then for the *pruned* configuration that is actually used in the final training, so that the stability of the ordering under the variable reduction can be assessed. The ROC curves are then given in two complementary forms: a *per-k-fold* overlay, which probes the stability of the training across the cross-validation folds, and a *pruned versus full* comparison, which quantifies the loss of separation power induced by discarding the lowest-ranked variables.

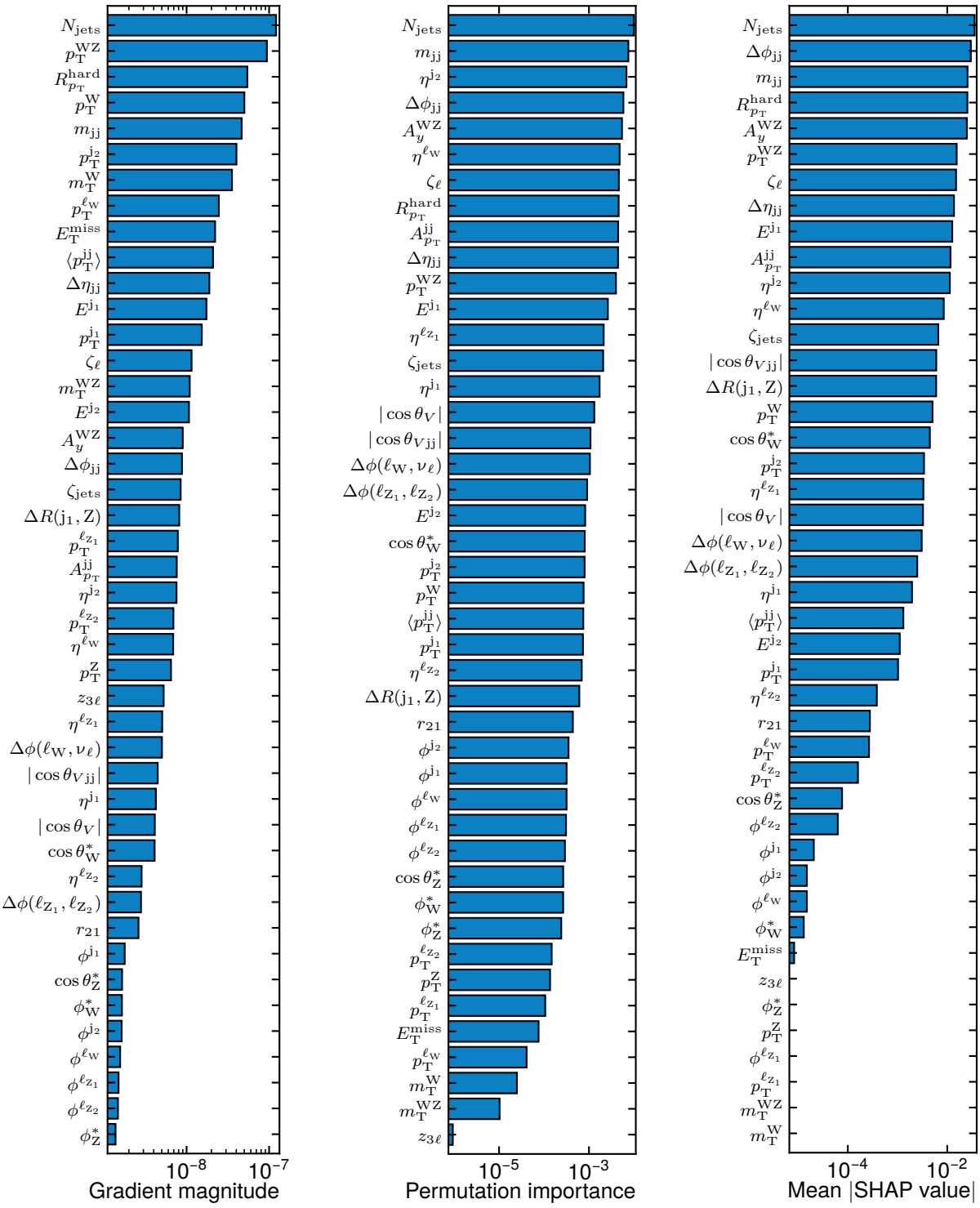
K.1 Variable ranking for the full set of variables

The variable rankings obtained with the full input set are shown in Figures K.1 to K.5, one DNN per page, with the gradient-magnitude, permutation, and SHAP estimators side by side. The three estimators probe different aspects of the DNN response and are not expected to give an identical ordering, but a variable ranked low by all three is a safe candidate for removal. The lowest-ranked variables identified here define the pruned configuration of the following section.



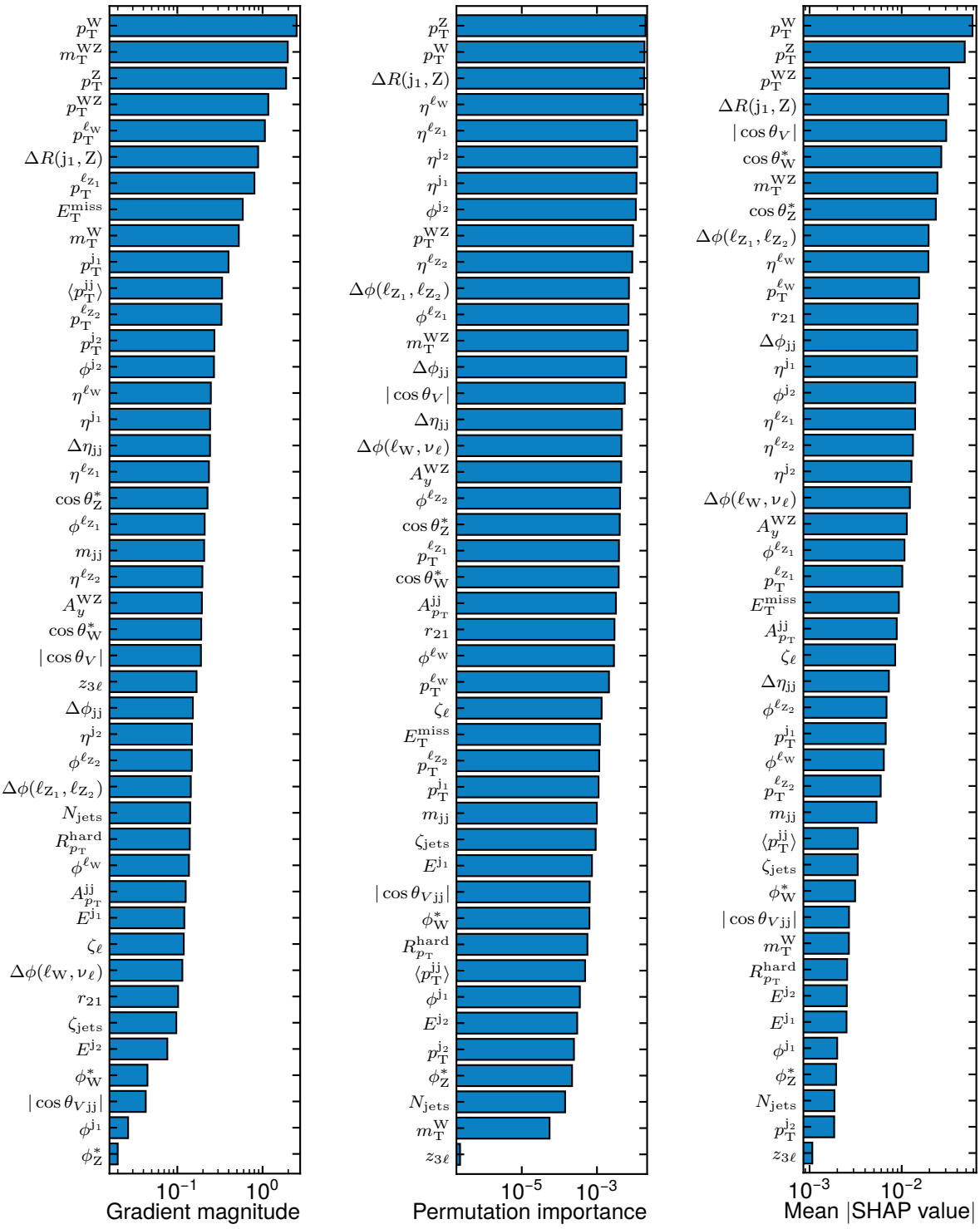
(a) Gradient magnitude **(b)** Permutation importance **(c)** SHAP value

Figure K.1: Variable ranking for the S. vs B. DNN for Run-2 with the full set of variables.



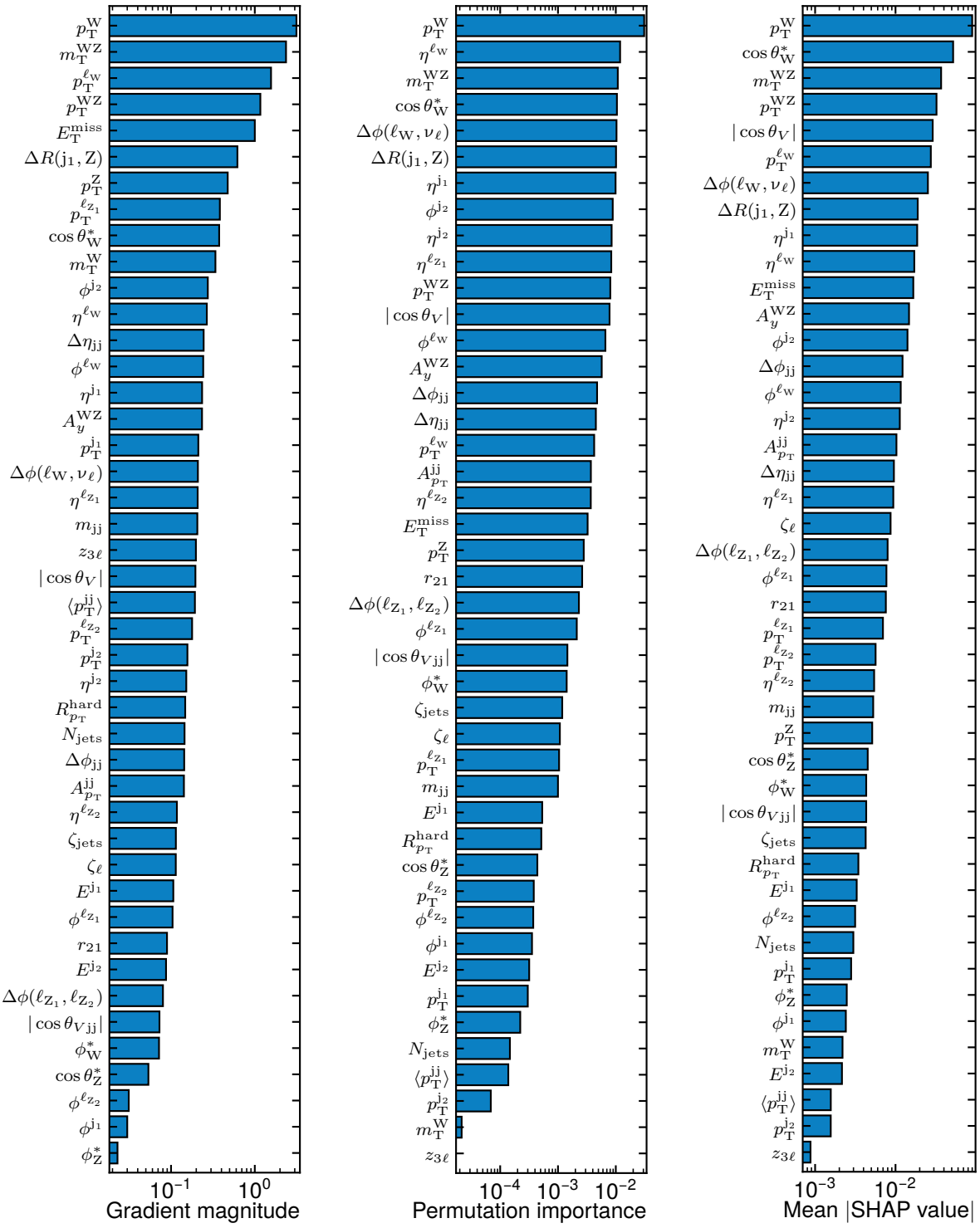
(a) Gradient magnitude **(b)** Permutation importance **(c)** SHAP value

Figure K.2: Variable ranking for the S. vs B. DNN for Run-3 with the full set of variables.



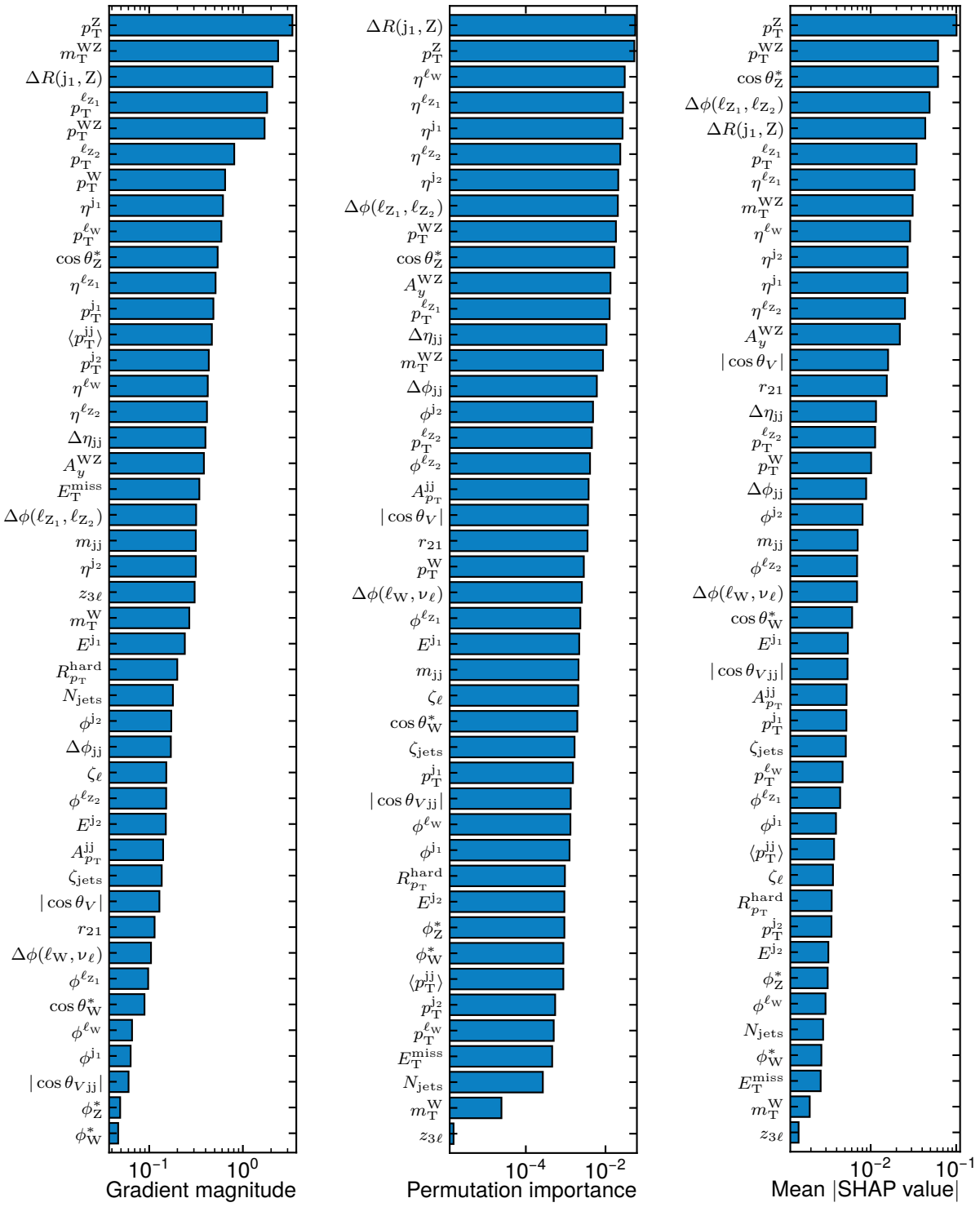
(a) Gradient magnitude **(b)** Permutation importance **(c)** SHAP value

Figure K.3: Variable ranking for the Pol. 00 DNN for Run-2 with the full set of variables.



(a) Gradient magnitude **(b)** Permutation importance **(c)** SHAP value

Figure K.4: Variable ranking for the Pol. $0\mathrm{X}$ DNN for Run-2 with the full set of variables.

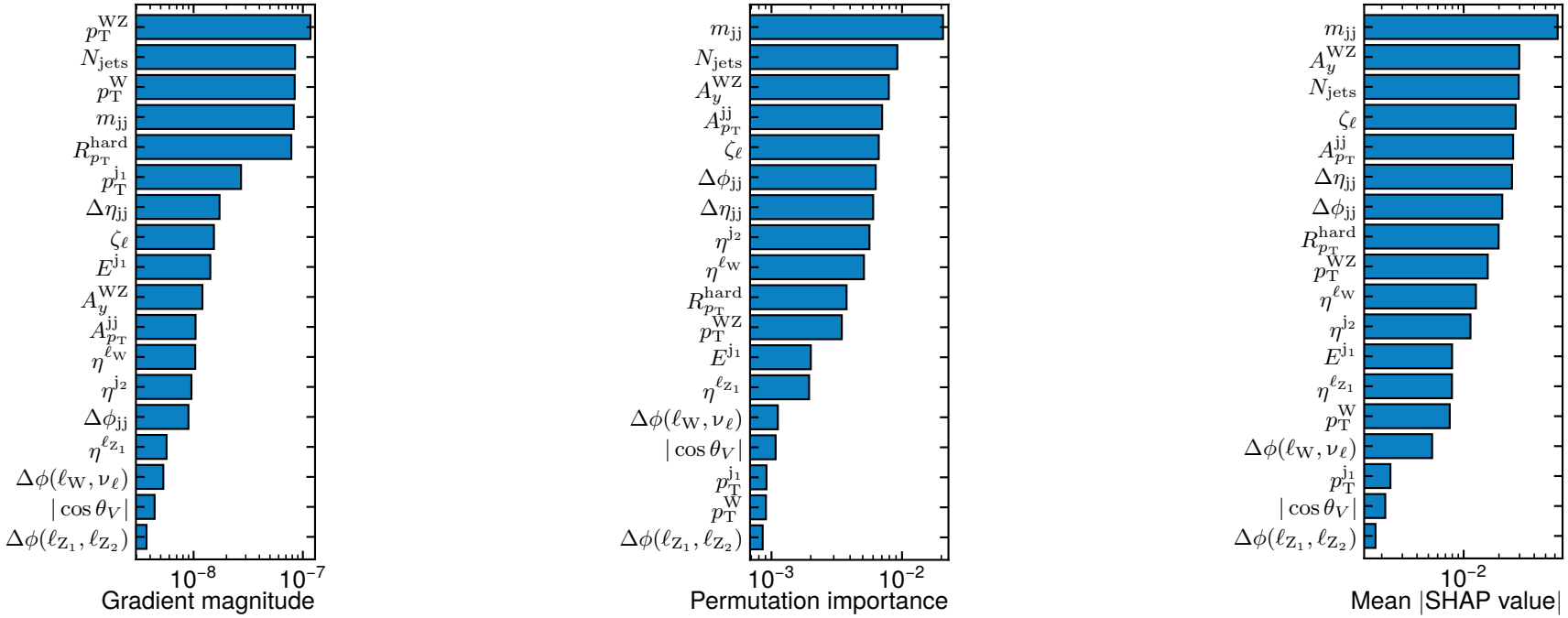


(a) Gradient magnitude **(b)** Permutation importance **(c)** SHAP value

Figure K.5: Variable ranking for the Pol. X0 DNN for Run-2 with the full set of variables.

K.2 Variable ranking for the pruned set of variables

After removing the lowest-ranked variables, the DNNs are retrained on the reduced input set and the ranking is recomputed, as shown in Figures K.6 to K.10. The orderings confirm that the most discriminating variables retain their leading positions and that no single removed variable was carrying a large fraction of the separation power, validating the pruning.

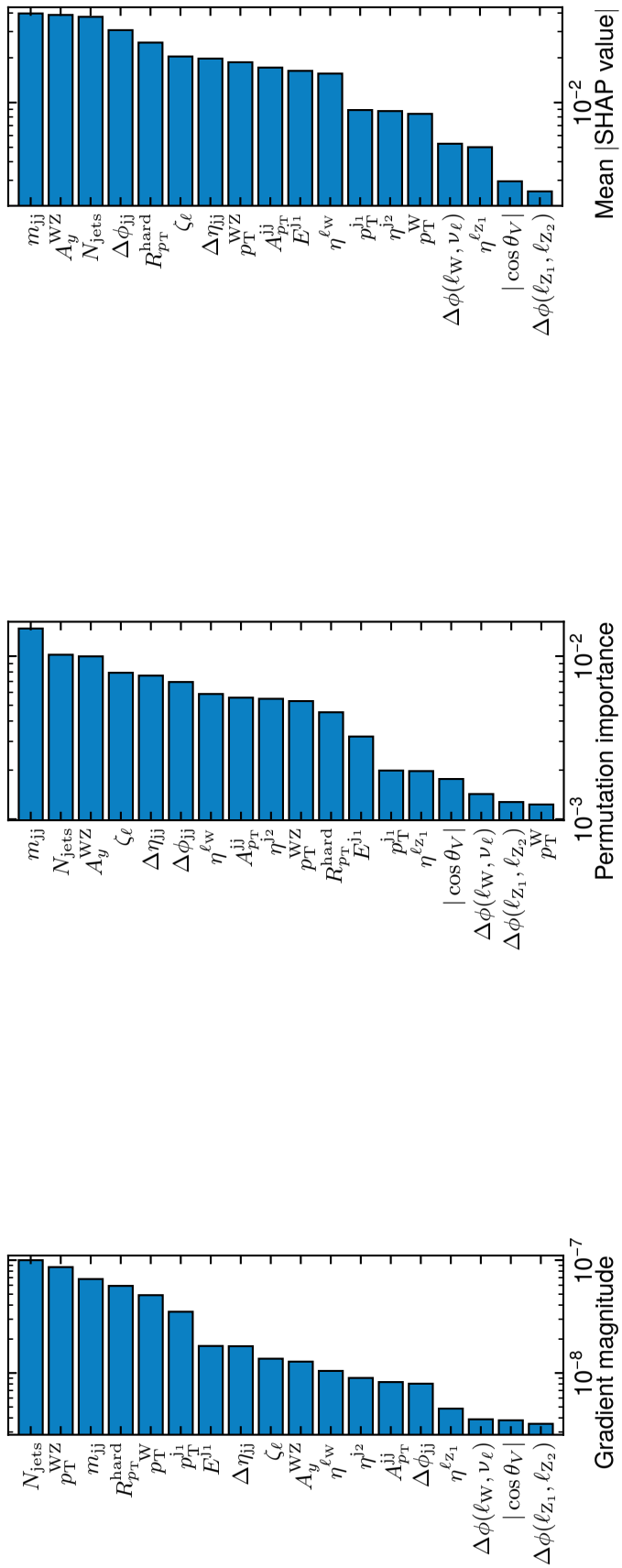


(a) Gradient magnitude

(b) Permutation importance

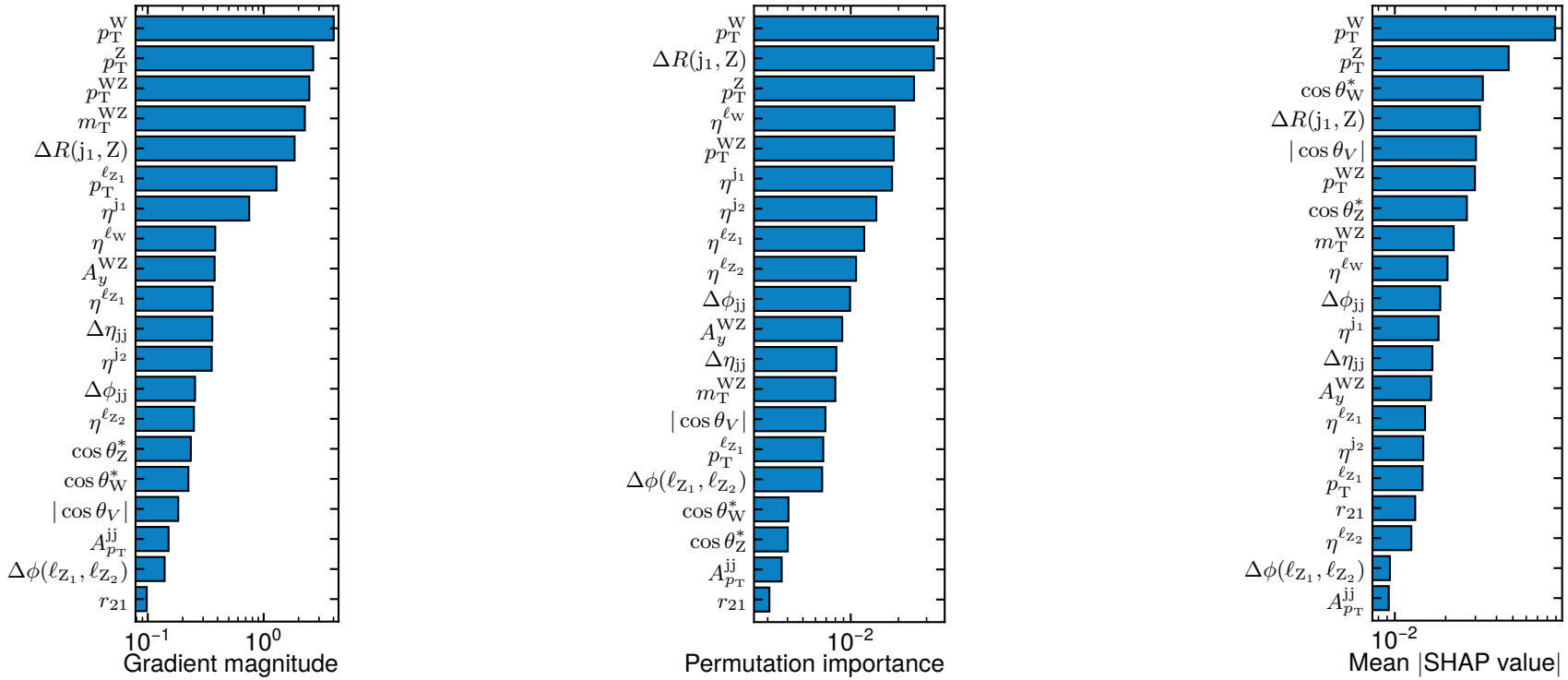
(c) SHAP value

Figure K.6: Variable ranking for the S. vs B. DNN for Run-2 with the pruned set of variables.



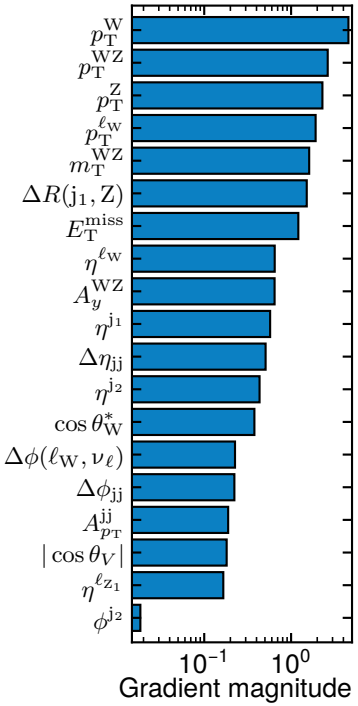
(a) Gradient magnitude **(b)** Permutation importance **(c)** SHAP value

Figure K.7: Variable ranking for the S. vs B. DNN for Run-3 with the pruned set of variables.

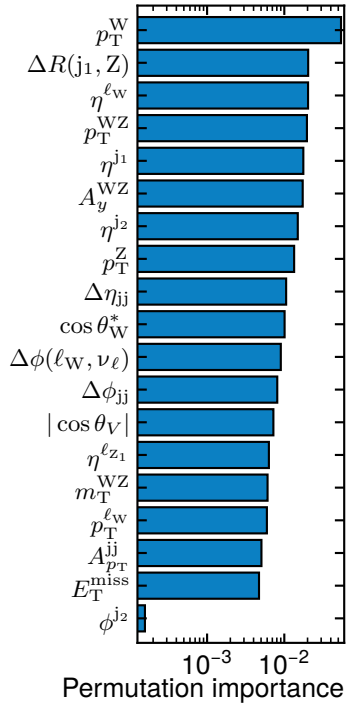


(a) Gradient magnitude **(b)** Permutation importance **(c)** SHAP value

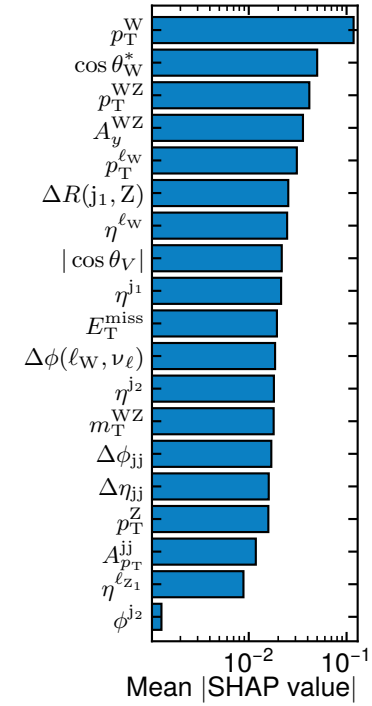
Figure K.8: Variable ranking for the Pol. 00 DNN for Run-2 with the pruned set of variables.



(a) Gradient magnitude

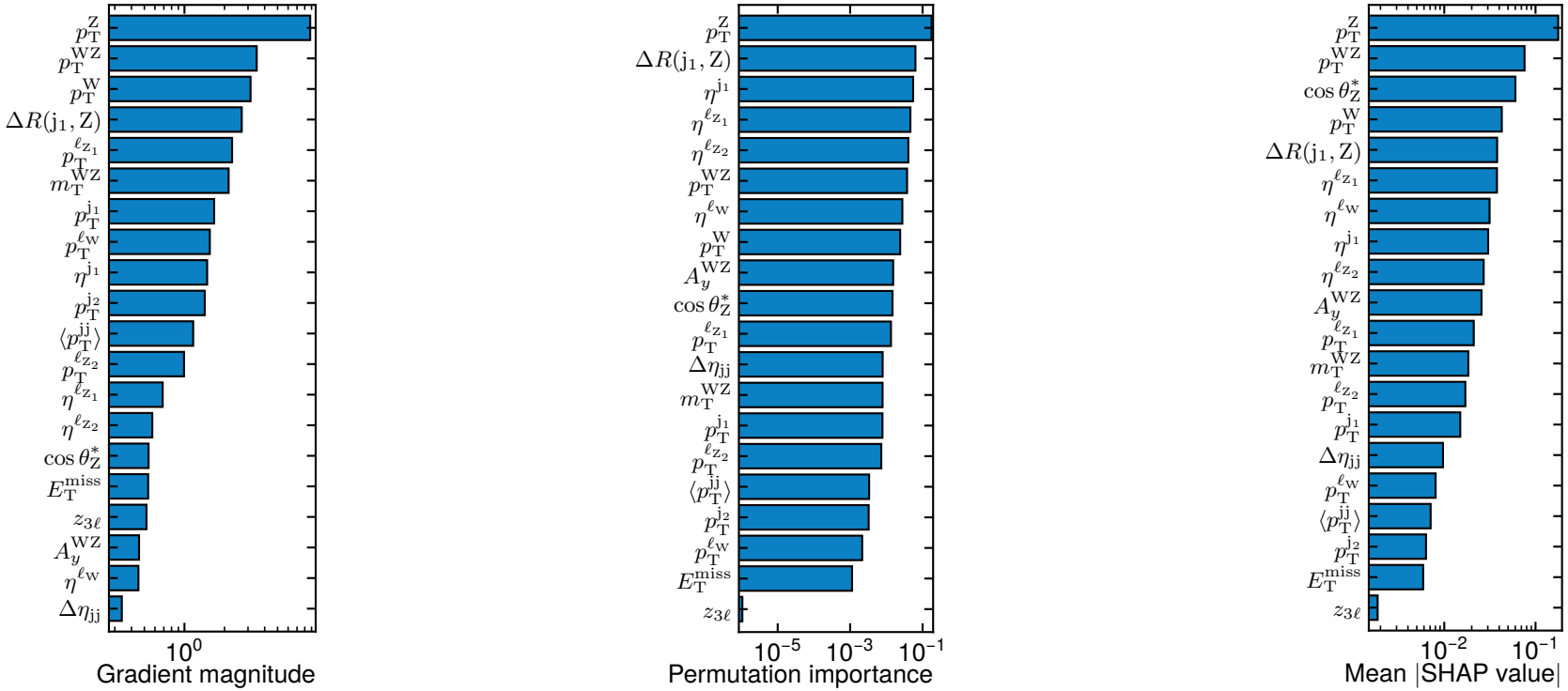


(b) Permutation importance



(c) SHAP value

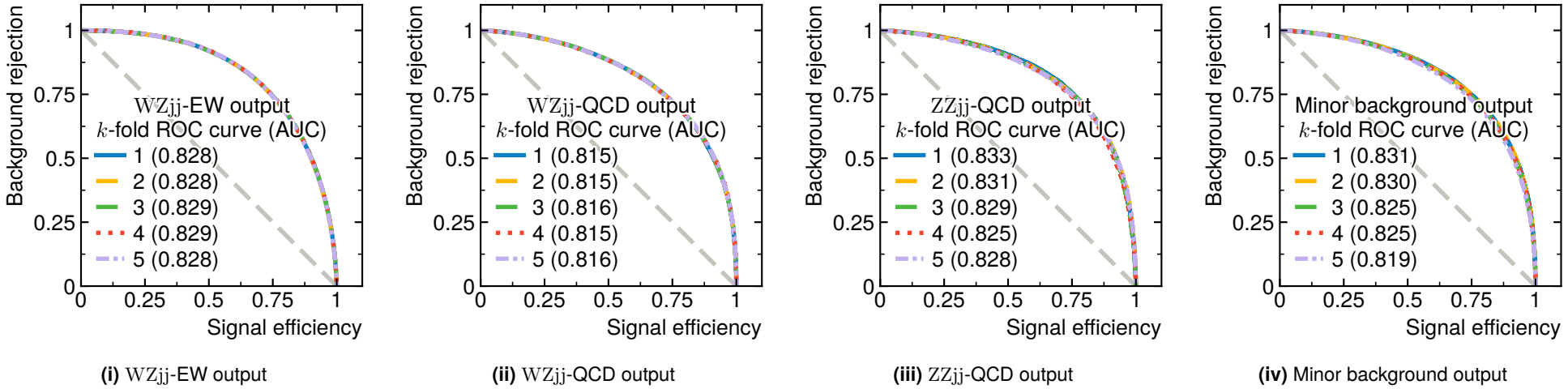
Figure K.9: Variable ranking for the Pol. $0\mathrm{X}$ DNN for Run-2 with the pruned set of variables.



(a) Gradient magnitude **(b)** Permutation importance **(c)** SHAP value

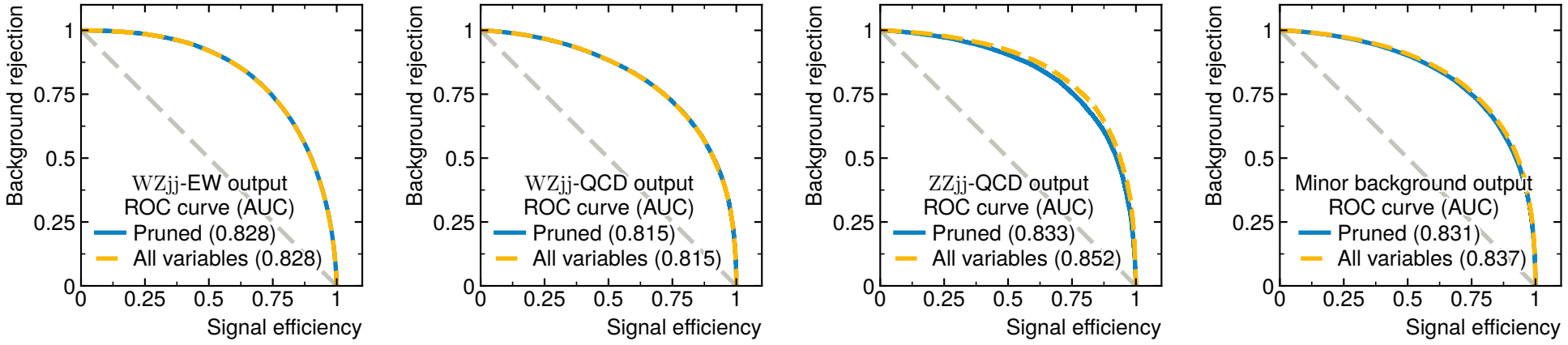
Figure K.10: Variable ranking for the Pol. X0 DNN for Run-2 with the pruned set of variables.

K.3 ROC-curve comparisons: fold stability and variable pruning



(i) WZjj-EW output **(ii)** WZjj-QCD output **(iii)** ZZjj-QCD output **(iv)** Minor background output

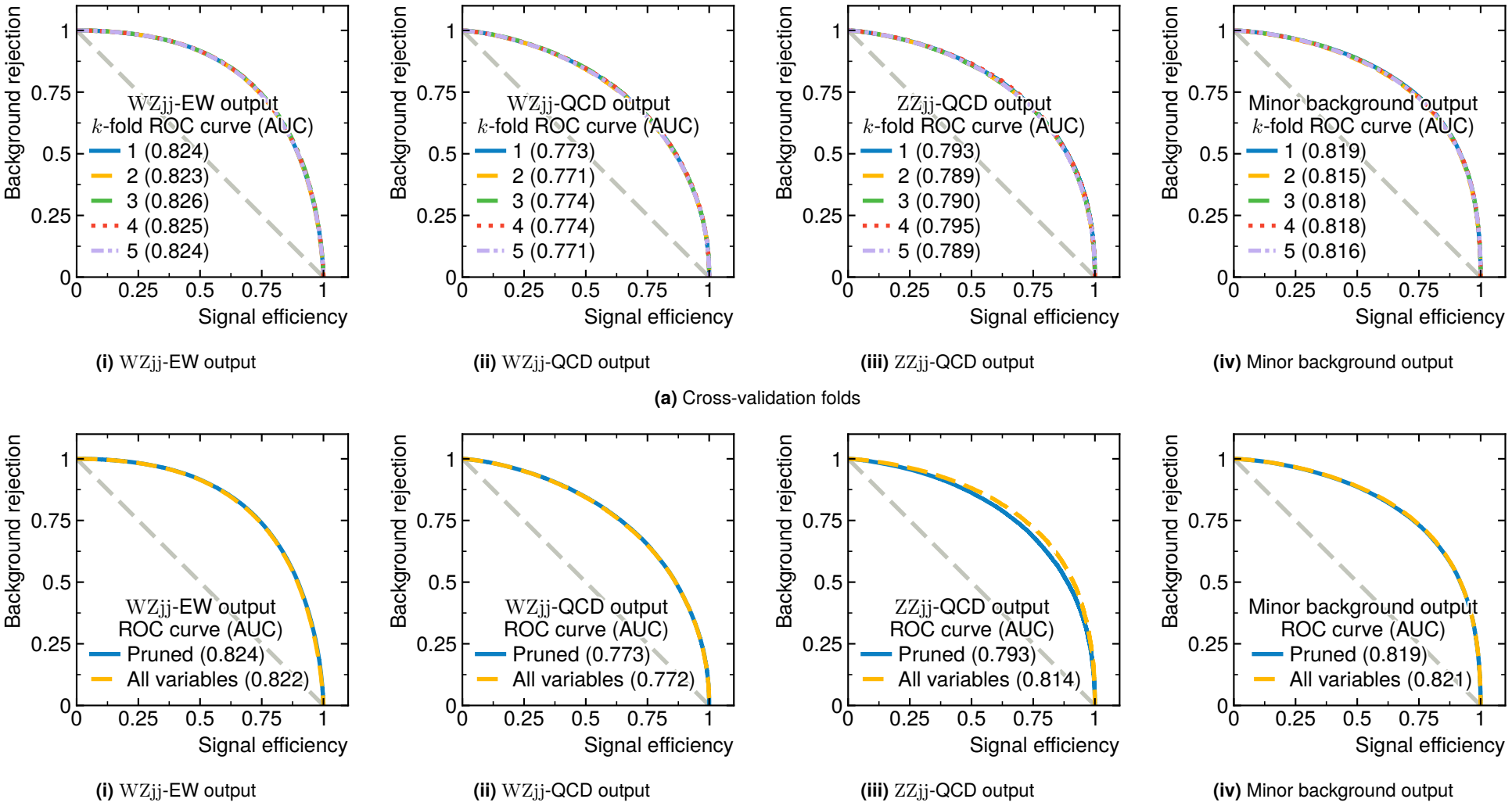
(a) Cross-validation folds



(i) WZjj-EW output **(ii)** WZjj-QCD output **(iii)** ZZjj-QCD output **(iv)** Minor background output

(b) Variable pruning

Figure K.11: ROC comparison for the S. vs B. DNN for Run-2, showing the overlay of the five cross-validation folds (a) and the comparison of the pruned variable set against the full set (b), with one panel per output node: WZjj-EW (i), WZjj-QCD (ii), ZZjj-QCD (iii), and the minor background output (iv).



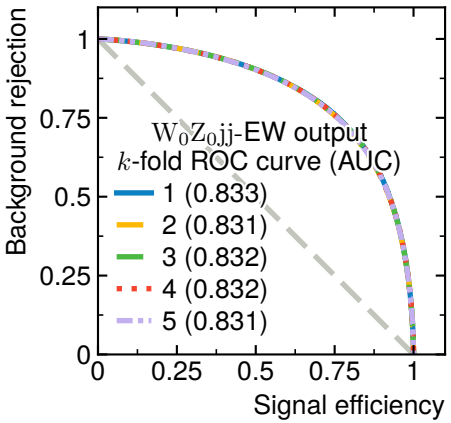
(i) WZjj-EW output **(ii)** WZjj-QCD output **(iii)** ZZjj-QCD output **(iv)** Minor background output

(a) Cross-validation folds

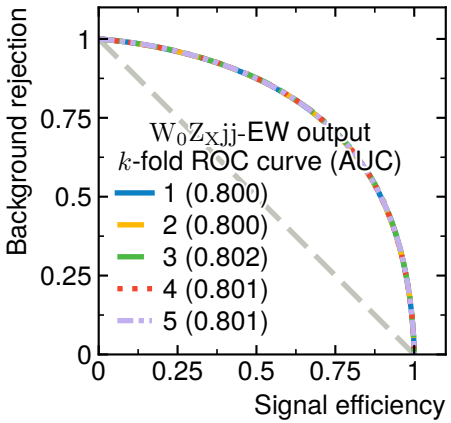
(i) WZjj-EW output **(ii)** WZjj-QCD output **(iii)** ZZjj-QCD output **(iv)** Minor background output

(b) Variable pruning

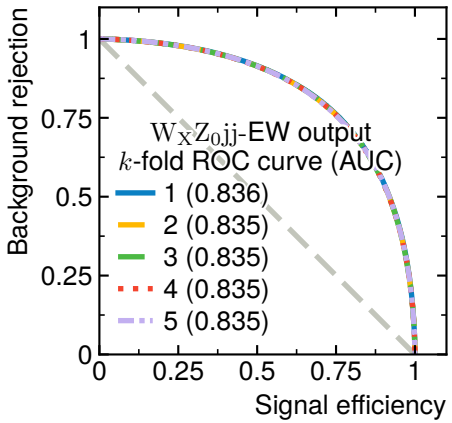
Figure K.12: ROC comparison for the S. vs B. DNN for Run-3, showing the overlay of the five cross-validation folds (a) and the comparison of the pruned variable set against the full set (b), with one panel per output node: WZjj-EW (i), WZjj-QCD (ii), ZZjj-QCD (iii), and the minor background output (iv).



(i) W_0Z_0jj-EW output

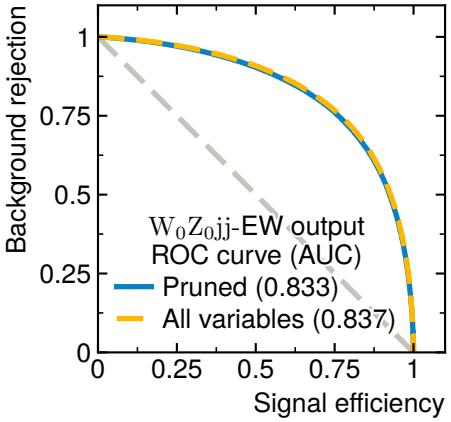


(ii) W_0Z_Xjj-EW output

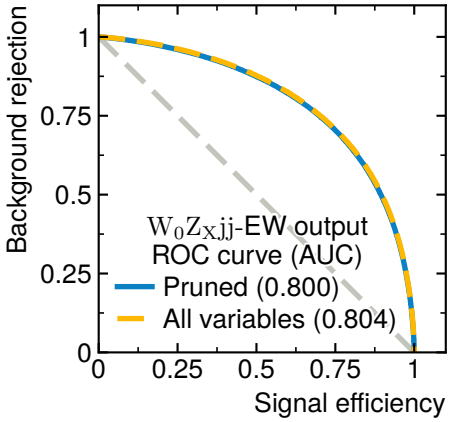


(iii) W_XZ_0jj-EW output

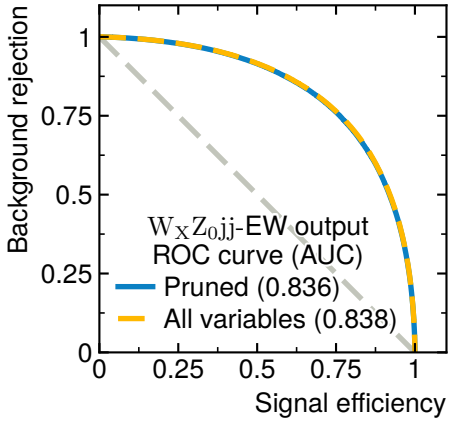
(a) Cross-validation folds



(i) W_0Z_0jj-EW output



(ii) W_0Z_Xjj-EW output



(iii) W_XZ_0jj-EW output

(b) Variable pruning

Figure K.13: ROC comparison for the three polarisation DNNs for Run-2, showing the overlay of the five cross-validation folds (a) and the comparison of the pruned variable set against the full set (b), with one panel per polarisation classifier: W_0Z_0jj-EW (i), W_0Z_Xjj-EW (ii), and W_XZ_0jj-EW (iii).

L Generator effects on DNN distributions

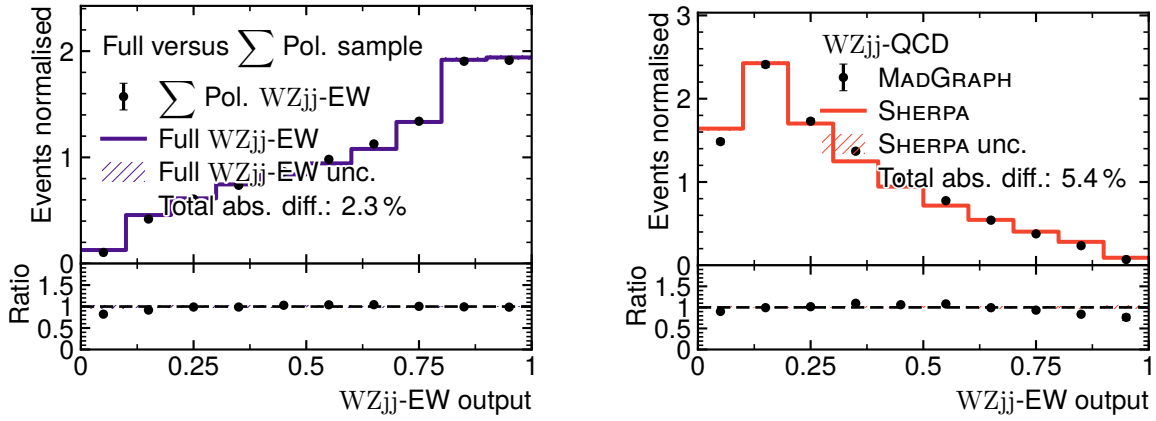
This appendix gathers the ratio comparisons of the critical samples used during training and in the fit. In every ratio panel, “Total abs. diff.” is the sum over the bins of the absolute difference of the two unit-normalised distributions.

For the S. vs B. DNN the training was performed with the polarisation-sum sample, which is cross-checked against the individual polarised samples later used in the DNNs for polarisation (joint and single). The comparison is performed for Run-2 and Run-3, for each polarisation state. In addition, the closure of the polarisation-sum sample against the full sample is shown for Run-2, as well as the comparison of the QCD background between the two generators, MadGraph (used for Run-2 training) and Sherpa (used for Run-3 training and for the fit in both runs).

For the polarisation DNNs the outputs are compared between the MadGraph (used for training) and Sherpa (used for the fit) generators, for Run-2 and Run-3.

L.1 Signal versus background DNN

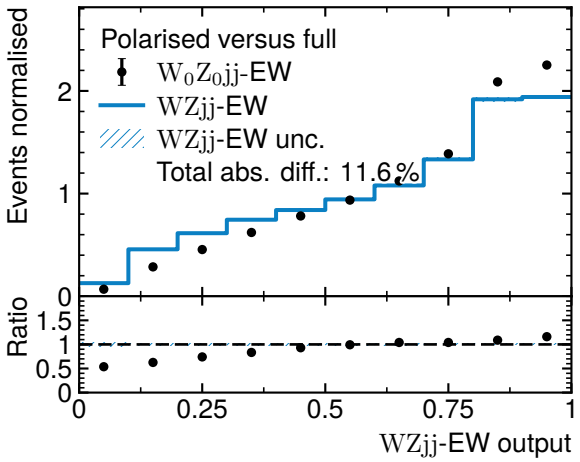
For the S. vs B. DNN the four polarised WZjj-EW templates are first compared against the full sample (Run-2) or the polarisation-sum sample (Run-3). Two further cross-checks follow: the closure of the polarisation-sum sample against the full sample, which probes any interference or normalisation mismatch, and the comparison of the WZjj-QCD background between the two generators. These comparisons are shown in Figures L.1 to L.3.



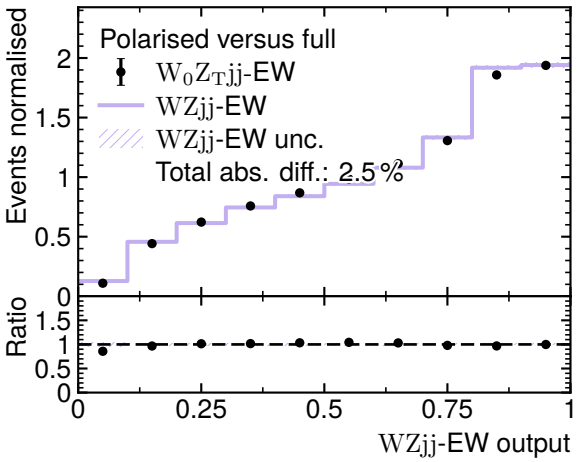
(a) Full versus sum of polarisations

(b) WZjj-QCD: MadGraph versus Sherpa

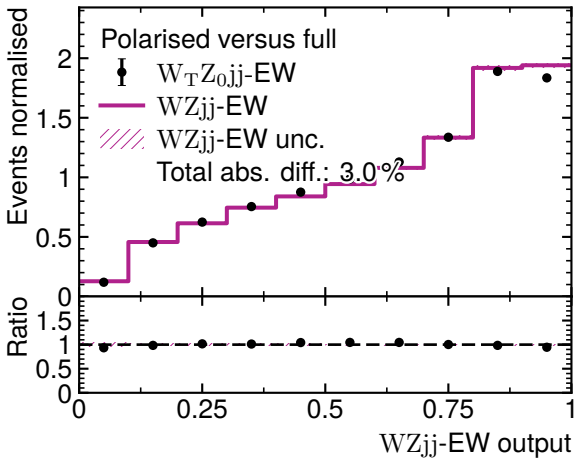
Figure L.1: Closure and generator cross-checks for S. vs B. DNN in Run-2.



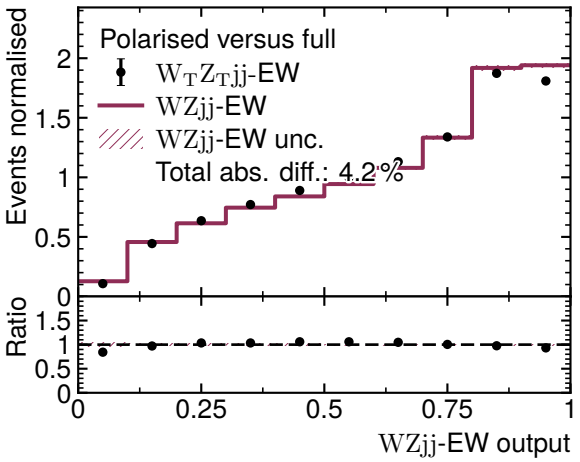
(a) W_0Z_0jj-EW



(b) W_0Z_Tjj-EW

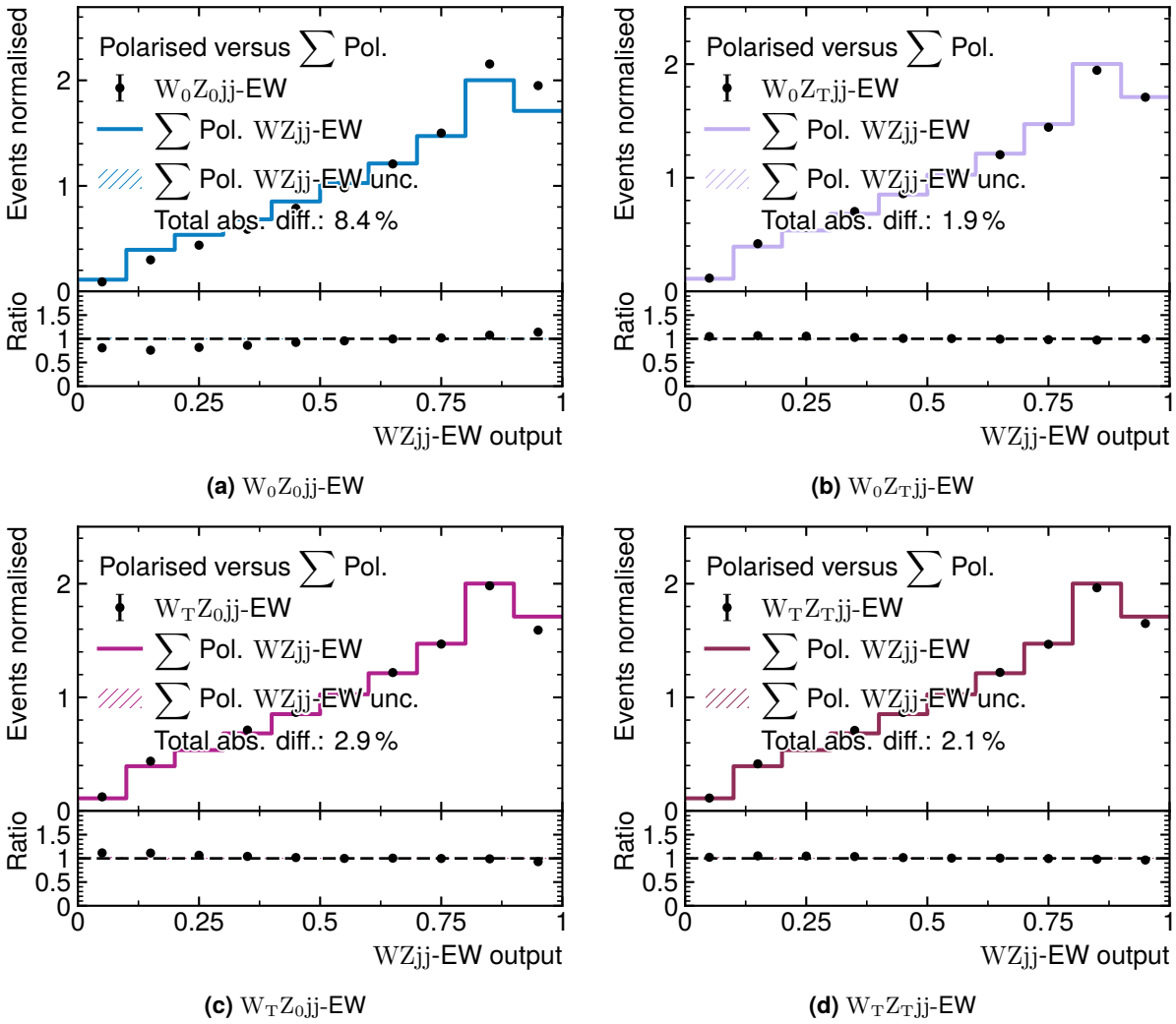


(c) W_TZ_0jj-EW



(d) W_TZ_Tjj-EW

Figure L.2: Polarised versus full for S. vs B. DNN in Run-2.



(a) W_0Z_0jj-EW **(b)** W_0Z_Tjj-EW

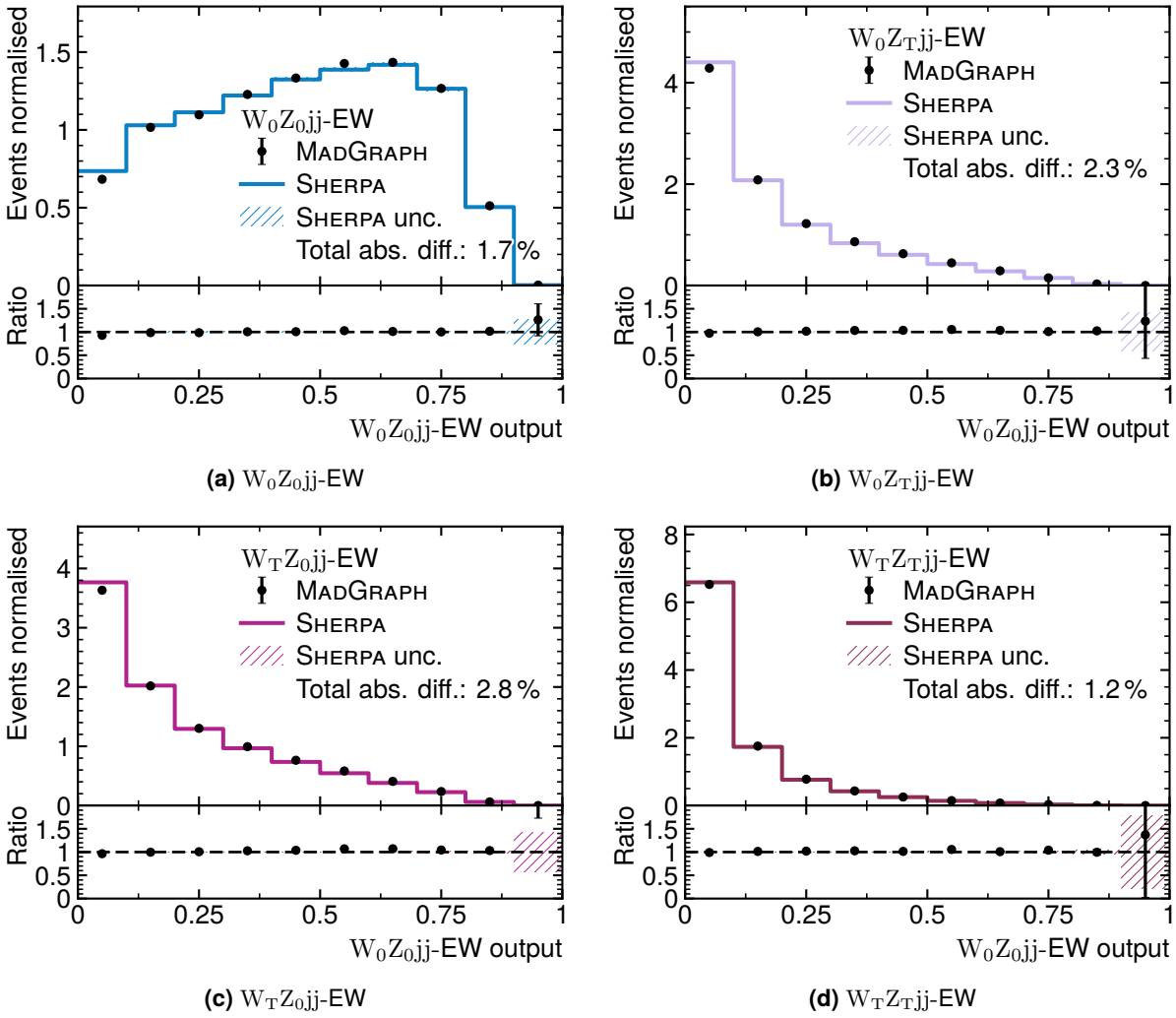
(c) W_TZ_0jj-EW **(d)** W_TZ_Tjj-EW

Figure L.3: Polarised versus the polarisation-sum sample for S. vs B. DNN in Run-3.

L.2 Polarisation DNN

For the polarisation classifiers the four polarised templates are compared between the MadGraph and Sherpa generators on each DNN output, separately for Run-2 and Run-3. Good agreement between the two generators supports the change from the training samples (MadGraph) to the fit samples (Sherpa). The comparison is performed for the joint Pol. 00 DNN and the two single polarisation Pol. $0\mathrm{X}$ and Pol. $\mathrm{X}0$ DNNs, as shown in Figures L.4 to L.9.

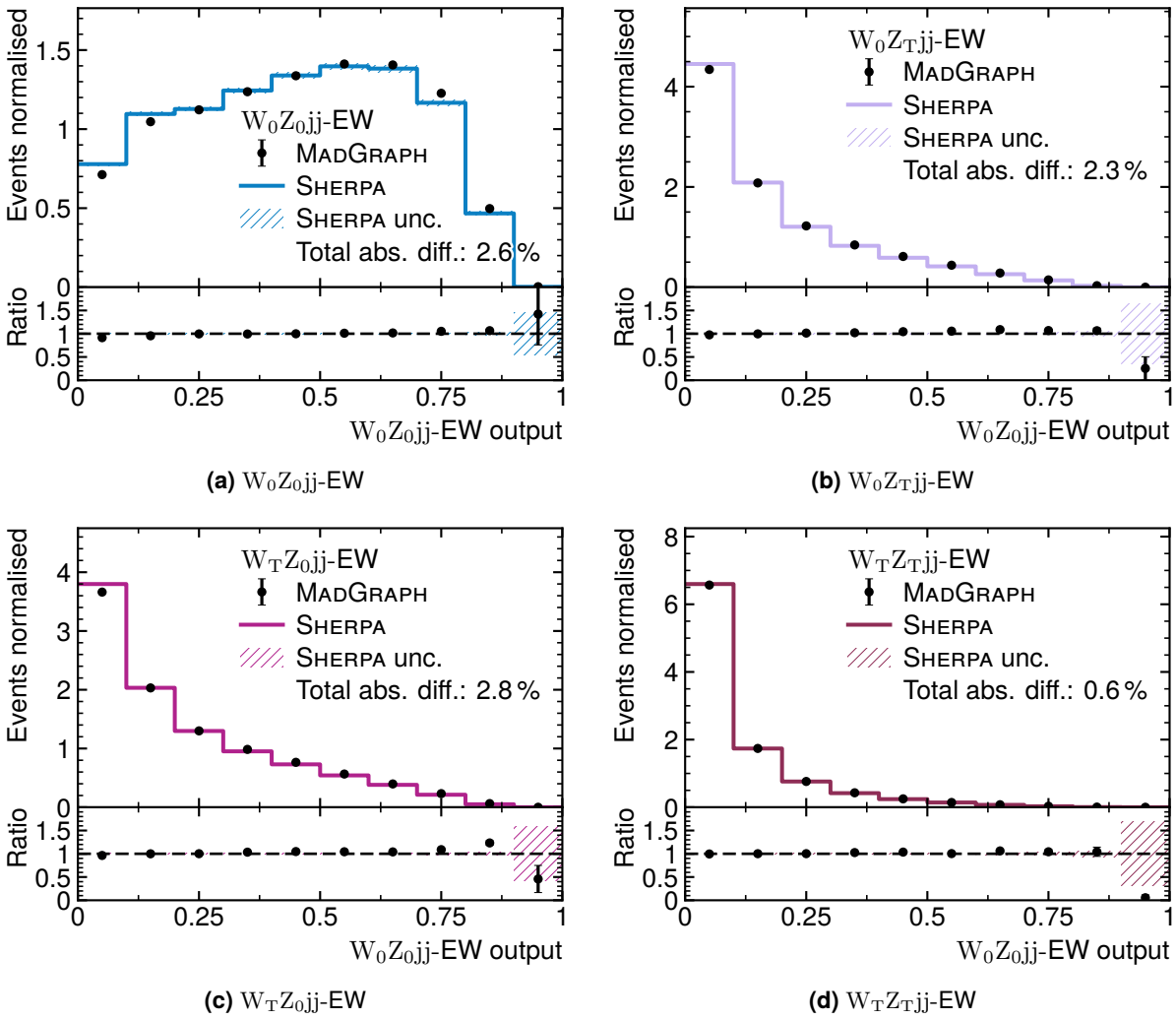
L.2.1 Joint polarisation 00 DNN



(a) $\mathrm{W_0Z_0jj}$-EW **(b)** $\mathrm{W_0Z_Tjj}$-EW

(c) $\mathrm{W_TZ_0jj}$-EW **(d)** $\mathrm{W_TZ_Tjj}$-EW

Figure L.4: Generator comparison for Pol. 00 DNN in Run-2.

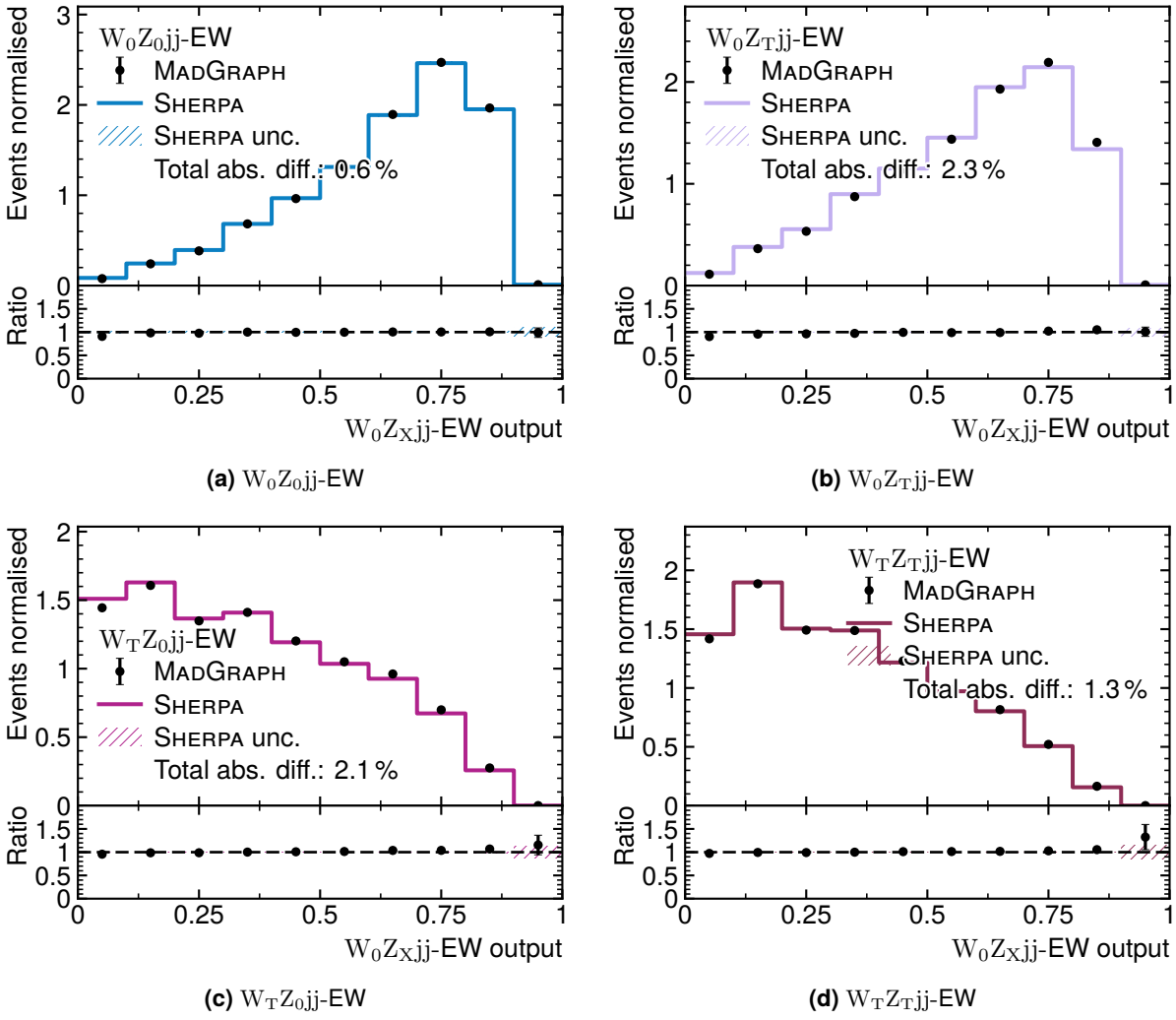


(a) W_0Z_0jj-EW **(b)** W_0Z_Tjj-EW

(c) W_TZ_0jj-EW **(d)** W_TZ_Tjj-EW

Figure L.5: Generator comparison for Pol. 00 DNN in Run-3.

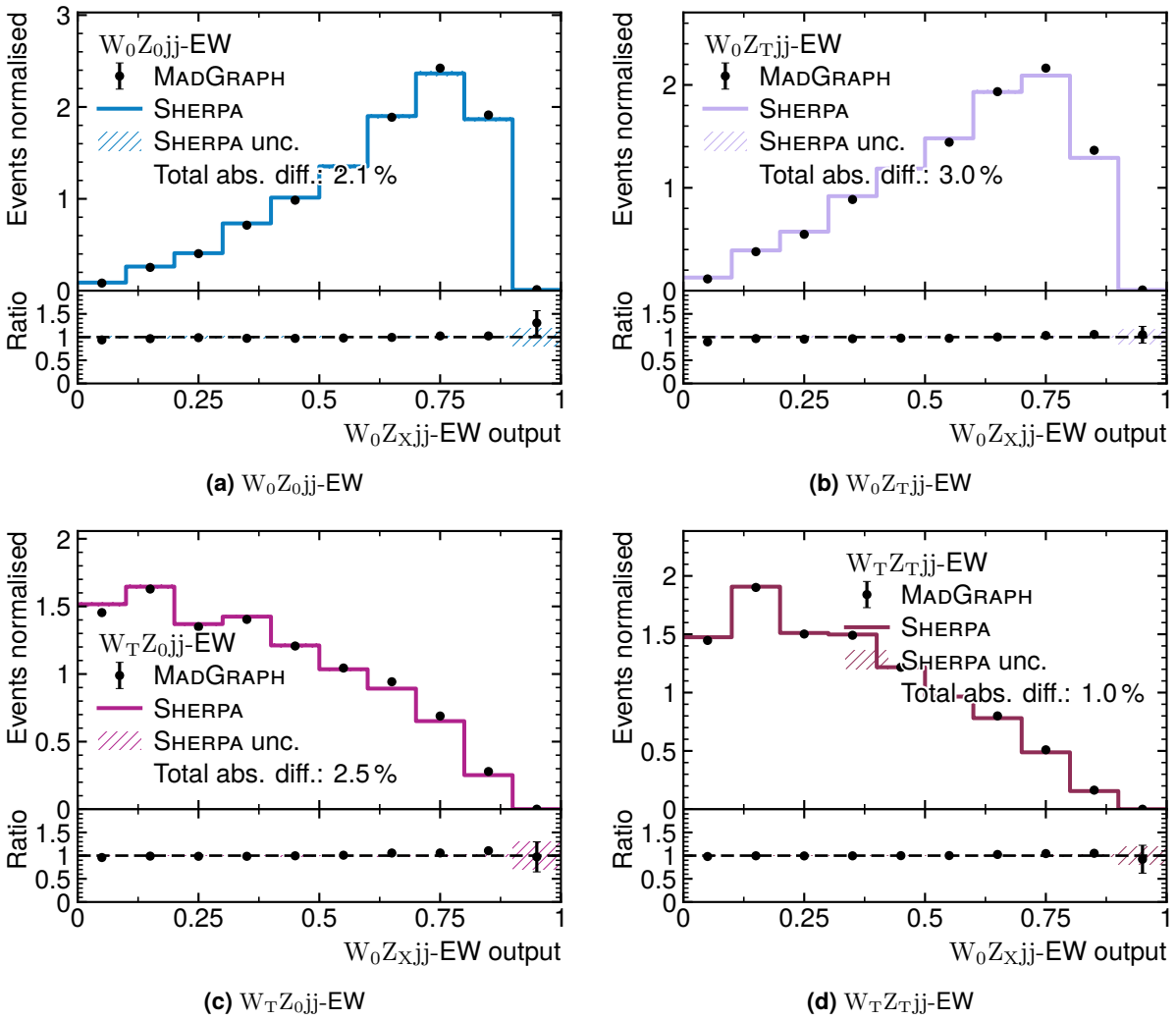
L.2.2 Single polarisation $0\mathrm{X}$ DNN



(a) $\mathrm{W_0Z_0jj}$-EW **(b)** $\mathrm{W_0Z_Tjj}$-EW

(c) $\mathrm{W_TZ_0jj}$-EW **(d)** $\mathrm{W_TZ_Tjj}$-EW

Figure L.6: Generator comparison for Pol. $0\mathrm{X}$ DNN in Run-2.

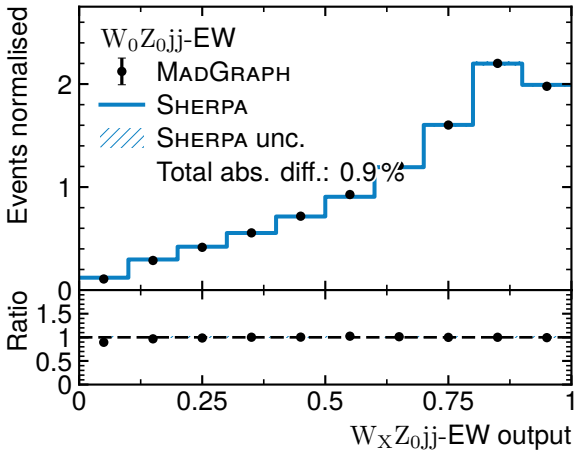


(a) $\mathrm{W_0Z_0jj}$-EW **(b)** $\mathrm{W_0Z_Tjj}$-EW

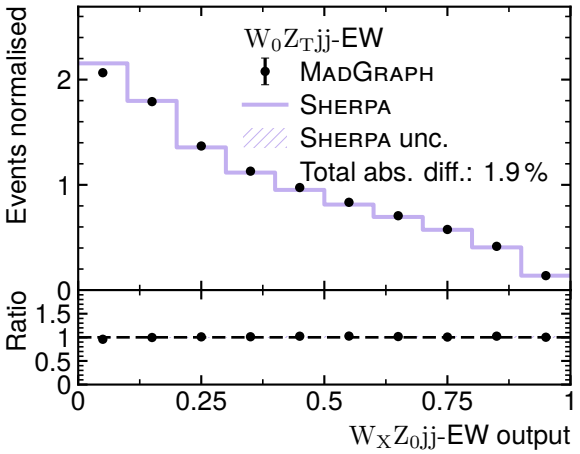
(c) $\mathrm{W_TZ_0jj}$-EW **(d)** $\mathrm{W_TZ_Tjj}$-EW

Figure L.7: Generator comparison for Pol. $0\mathrm{X}$ DNN in Run-3.

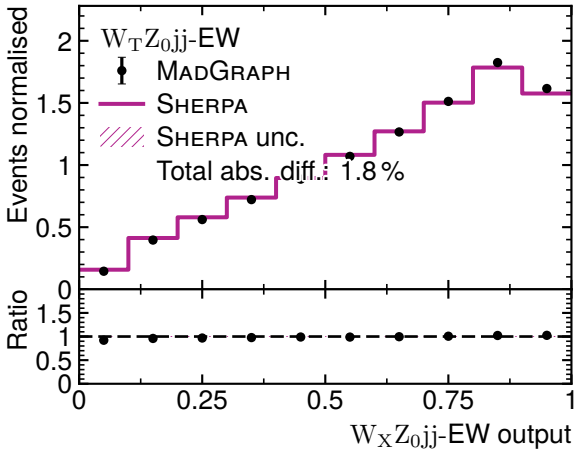
L.2.3 Single polarisation X0 DNN



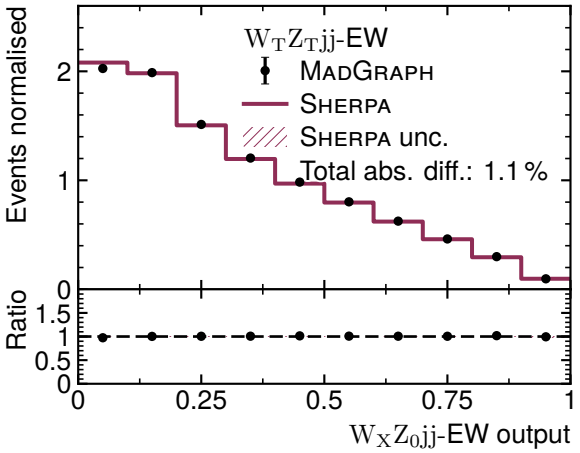
(a) W_0Z_0jj-EW



(b) W_0Z_Tjj-EW

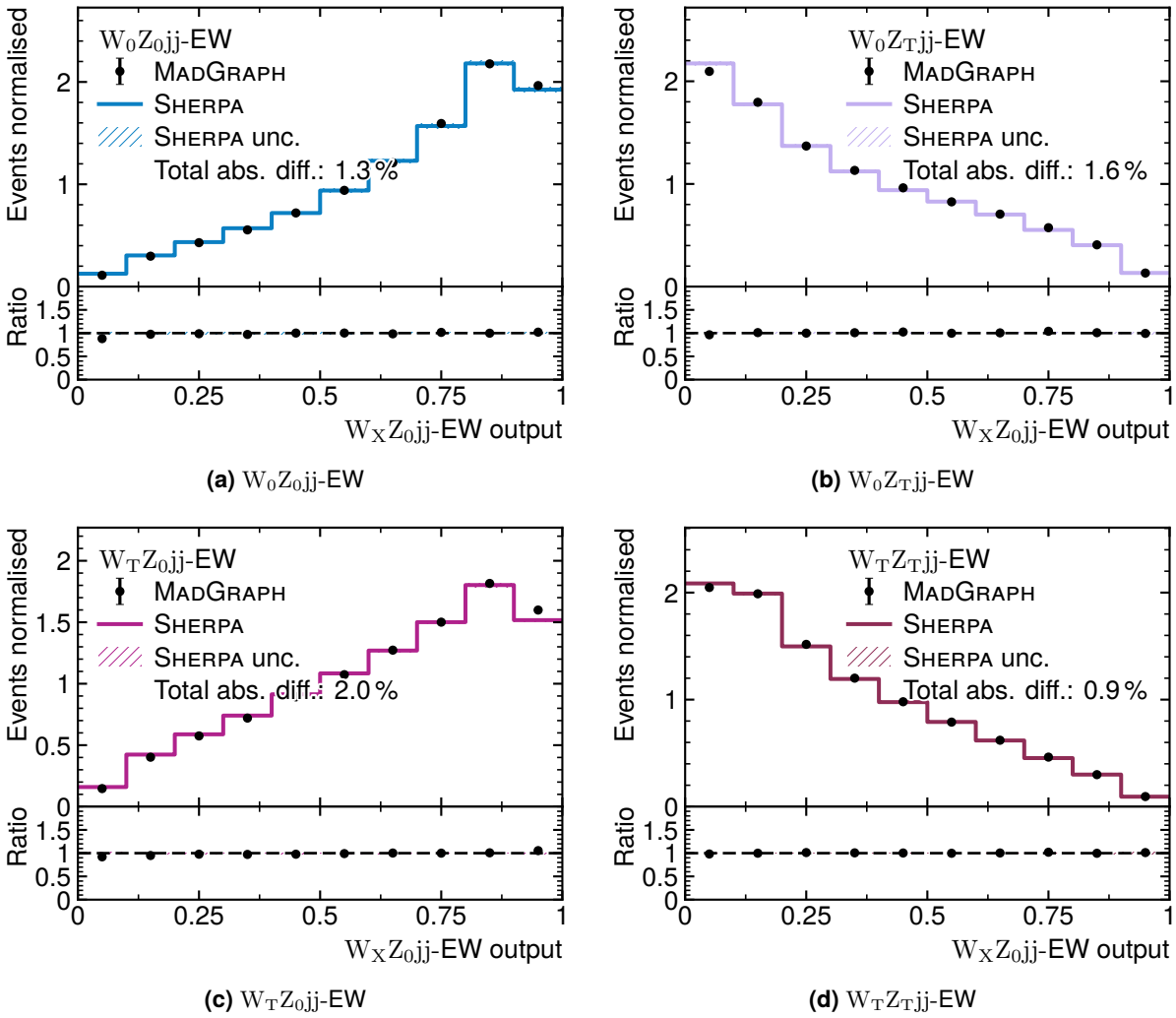


(c) W_TZ_0jj-EW



(d) W_TZ_Tjj-EW

Figure L.8: Generator comparison for Pol. X0 DNN in Run-2.



(a) $\mathrm{W_0Z_0jj}$-EW **(b)** $\mathrm{W_0Z_Tjj}$-EW

(c) $\mathrm{W_TZ_0jj}$-EW **(d)** $\mathrm{W_TZ_Tjj}$-EW

Figure L.9: Generator comparison for Pol. X0 DNN in Run-3.

M Binning optimisation

This appendix summarises the comparison between rectangular cuts and Voronoi binning for the three polarisation measurements. Three rectangular configurations are considered, 2×5, 5×2, and 3×3, denoted $x \times y$ with x the number of bins along the S. vs B. axis and y along the polarisation axis. They are compared with 9- and 10-bin Voronoi binnings. The rectangular edges are optimised by random variation at fixed bin multiplicity along each axis to maximise the expected Poisson significance:

$$Z = \sqrt{2\left((S+B)\ln\left(1+\frac{S}{B}\right) - S\right)}, \tag{M.1}$$

where S and B are the expected signal and background yields in a bin. At least one expected background event is required per bin to avoid divergences. The Voronoi binning, described in Section 6.7.7, instead relies on the local signal and background densities in the 2D discriminant plane and does not depend on this formula. No systematic uncertainties are included here, so the absolute significances are higher than those of the main analysis, but the relative comparison between configurations remains meaningful. All expected significances are collected in Table M.A, the 10-bin Voronoi binning being the baseline of the main analysis.

Two trends are common to the three hypotheses. At fixed bin multiplicity the Voronoi binning is always at least as good as the best rectangular grid: at nine bins it matches or slightly exceeds the 3×3 configuration, with relative gains ranging from $+0.0\,\%$ to $+2.4\,\%$ across the runs, and at ten bins it improves on the 2×5 grid by $+3.4\,\%$ to $+7.0\,\%$ and on the 5×2 grid by up to $\sim 11\,\%$. The pronounced spread between the 2×5 and 5×2 layouts, reaching $\sim 7\,\%$ in expected significance, reflects the sensitivity of any rigid factorised grid to the choice of axis ordering, a limitation the Voronoi binning avoids by construction.

For the joint polarisation $\mathrm{W_0Z_0}$jj-EW the improvement is modest: the 9-bin Voronoi binning essentially reproduces the 3×3 grid, while the 10-bin Voronoi binning reaches a combined significance of 1.30 against 1.27 for the best rectangular configuration. The single polarisation $\mathrm{W_0Z_X}$jj-EW shows slightly larger relative gains, the 10-bin Voronoi binning reaching 2.61 against 2.55, with the improvement amplified in Run-3 where the additional statistics populate regions of the discriminant plane that are poorly resolved by a fixed grid. The single polarisation $\mathrm{W_XZ_0}$jj-EW benefits the most: it has the highest absolute significance of the three channels, and the 10-bin Voronoi binning reaches 3.08 against 2.99, a gain large enough to be relevant for the final sensitivity.

M.1 Significance summary

Across the three hypotheses the Voronoi binning provides a small but systematic improvement over the best rectangular configuration while removing the arbitrariness of the axis ordering and the rigidity of square grids. By adapting to the local signal and background densities it makes a more efficient use of the available statistics, which is the reason the 10-bin Voronoi binning is adopted as the baseline of the main analysis. The expected significances for all configurations are summarised in Table M.A.

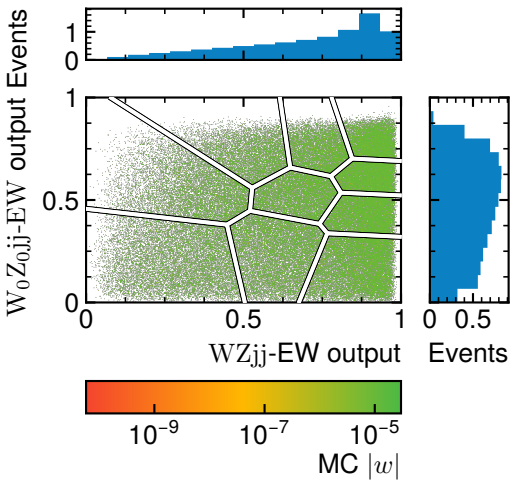
Table M.A: Expected significance for the three polarisation hypotheses, comparing rectangular cuts with Voronoi binning for Run-2, Run-3, and their combination.

		Significance		
Type	**Name**	**Run-2**	**Run-3**	**Combination**
Pol. 00	Rectangular cuts 2×5	0.89	0.90	1.25
	Rectangular cuts 5×2	0.83	0.85	1.17
	Rectangular cuts 3×3	0.91	0.91	1.27
	10 Voronoi bins	0.92	0.94	1.30
	9 Voronoi bins	0.91	0.92	1.28
Pol. 0X	Rectangular cuts 2×5	1.77	1.77	2.50
	Rectangular cuts 5×2	1.67	1.71	2.38
	Rectangular cuts 3×3	1.80	1.82	2.55
	10 Voronoi bins	1.83	1.87	2.61
	9 Voronoi bins	1.82	1.85	2.59
Pol. X0	Rectangular cuts 2×5	2.14	2.00	2.92
	Rectangular cuts 5×2	2.02	2.01	2.84
	Rectangular cuts 3×3	2.17	2.09	2.99
	10 Voronoi bins	2.24	2.14	3.08
	9 Voronoi bins	2.21	2.14	3.06

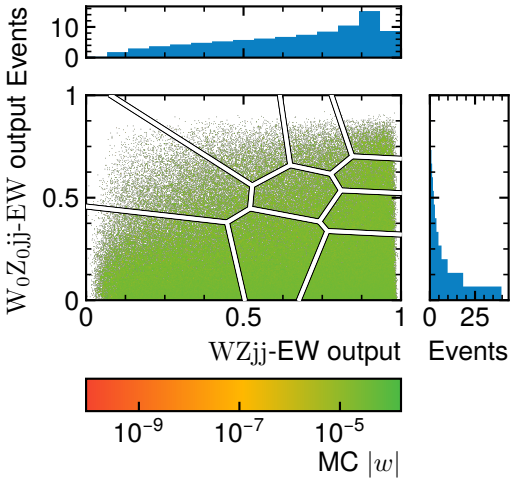
N Voronoi binning in depth

This appendix provides additional material on the Voronoi-based adaptive binning of the 2D discriminant plane, for each of the three polarisation configurations. For every configuration the per-class MC weight densities are shown first, followed by the region statistics (event yields, class fractions, and per-region significance) and the process composition of each Voronoi region, shown in Figures N.2, N.4, and N.6. The three per-class panels correspond to the classification targets in the order: targeted polarisation, remaining polarisation states, and background.

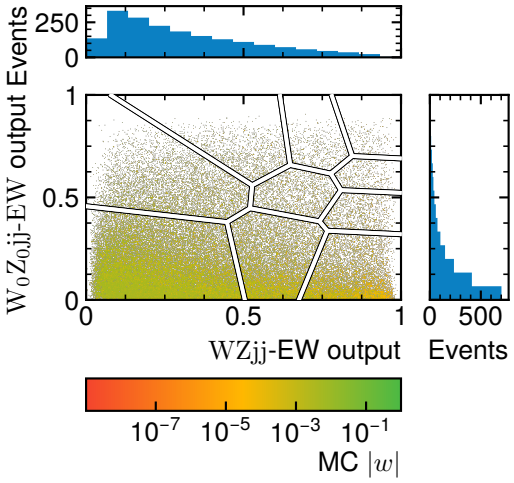
N.1 Joint polarisation: the 00 plane



(a) W_0Z_0jj-EW

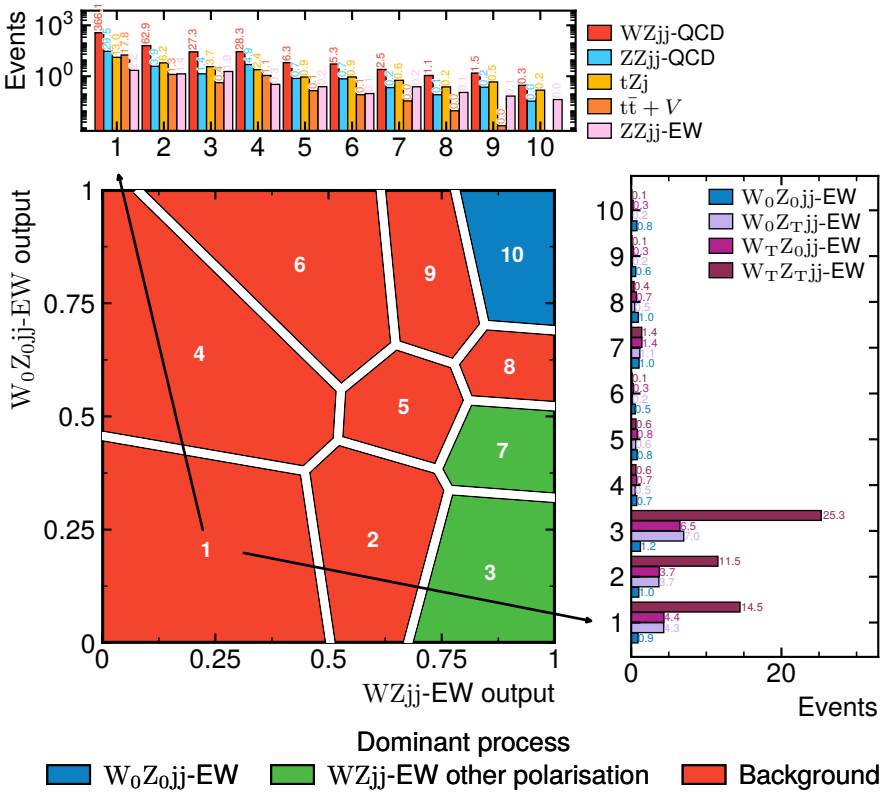


(b) Remaining polarisation states



(c) Background

Figure N.1: Per-class Voronoi densities, W_0Z_0jj-EW configuration.



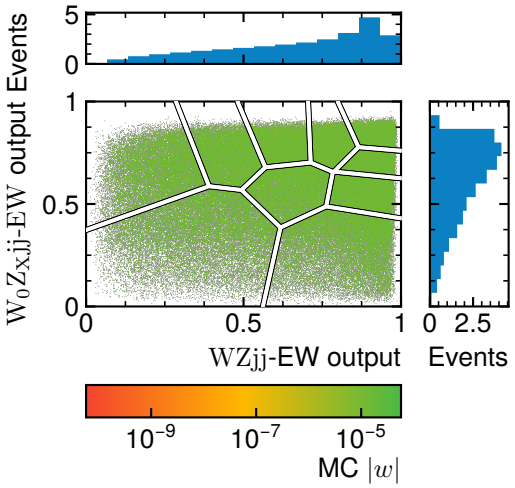
(a) The process that dominates each region

Regions	Events			Overall [%]			Raw fraction			$\frac{S}{S+B}$	$\frac{S}{\sqrt{B}}$
	W_0Z_0jj-EW	other	Bkg.	W_0Z_0jj-EW	other	Bkg.	W_0Z_0jj-EW	other	Bkg.		
1	0.9	23.2	453.0	10.6	25.2	71.3	0.108	0.246	0.596	0.002	0.04
2	1.0	19.0	78.5	11.8	20.6	12.4	0.118	0.203	0.153	0.010	0.10
3	1.2	38.8	37.0	14.4	42.1	5.8	0.143	0.392	0.116	0.016	0.14
4	0.7	1.9	41.9	8.6	2.0	6.6	0.087	0.027	0.069	0.017	0.11
5	0.8	2.1	8.5	9.2	2.2	1.3	0.092	0.030	0.022	0.070	0.24
6	0.5	0.6	7.9	6.2	0.6	1.2	0.063	0.009	0.015	0.060	0.18
7	1.0	4.0	3.7	12.2	4.3	0.6	0.121	0.055	0.014	0.120	0.38
8	1.0	1.5	1.9	11.2	1.7	0.3	0.111	0.023	0.006	0.218	0.52
9	0.6	0.5	2.3	7.0	0.6	0.4	0.069	0.009	0.007	0.173	0.35
10	0.8	0.5	0.5	8.9	0.5	0.1	0.088	0.008	0.003	0.428	0.76

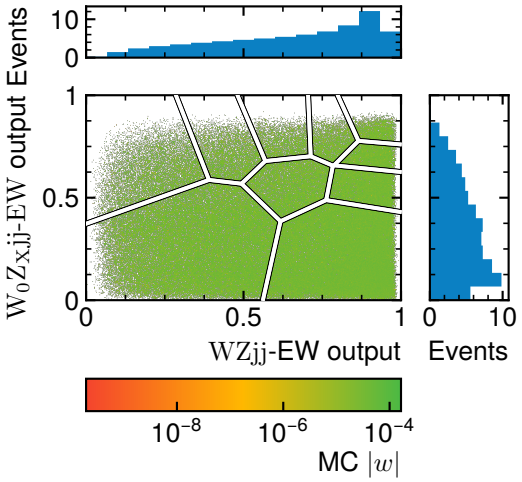
(b) The yields behind it

Figure N.2: The Voronoi regions of the W_0Z_0jj-EW configuration, for Run-2: the process that dominates each one (a), and the yields behind that (b). The region numbers are the same in the map and in the table, and each group of three columns gives the polarised signal, the other polarisations, and the background, in that order. “Overall” is the weighted fraction of each class that falls in the region and “Raw fraction” the same fraction with raw MC events, while $\frac{S}{S+B}$ is the purity of the region.

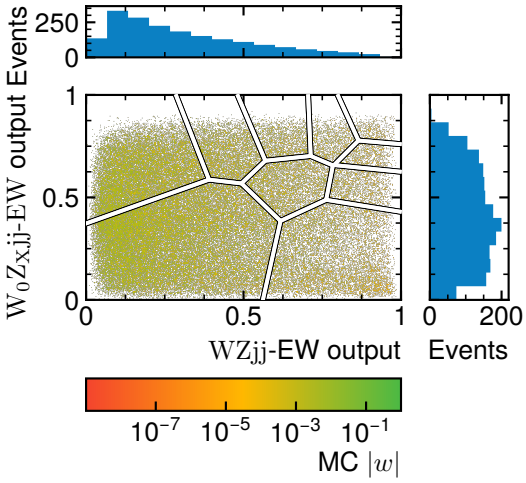
N.2 Single polarisation: the $0\mathrm{X}$ plane



(a) $\mathrm{W_0Z_Xjj}$-EW

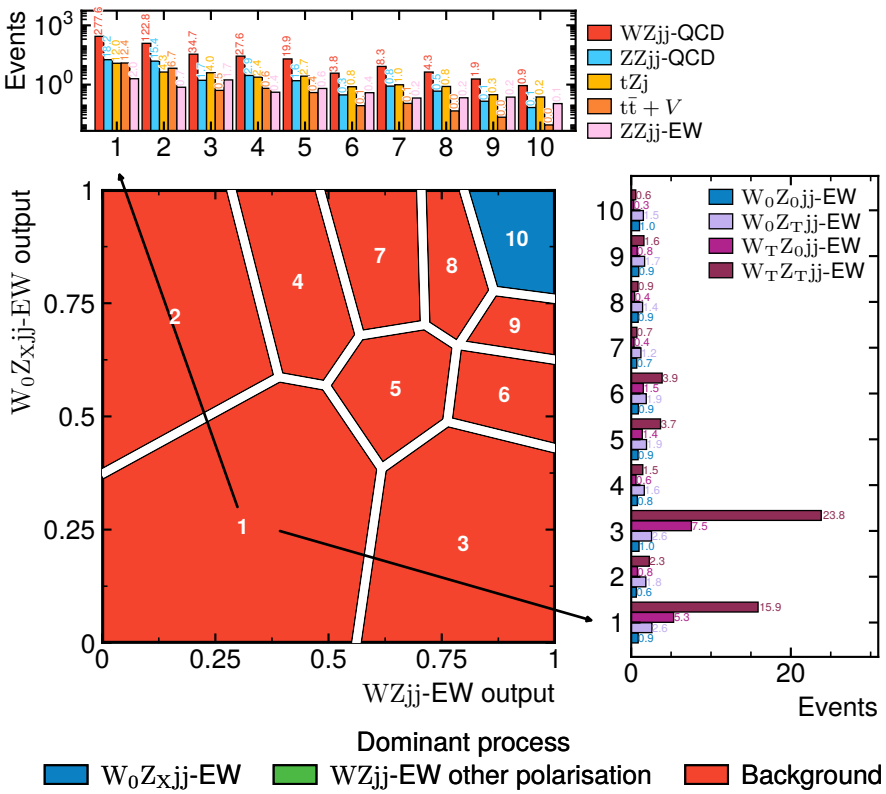


(b) $\mathrm{W_TZ_Xjj}$-EW



(c) Background

Figure N.3: Per-class Voronoi densities, $\mathrm{W_0Z_Xjj}$-EW configuration.



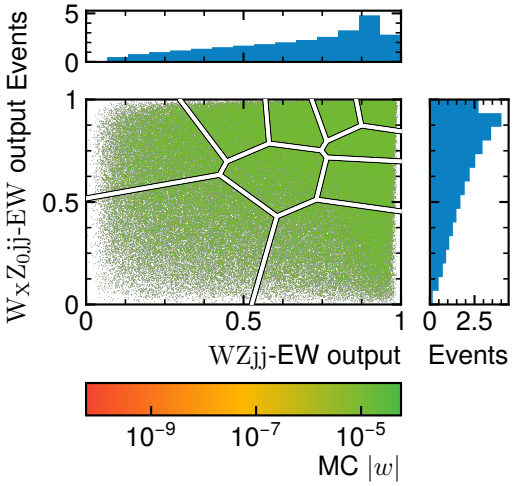
(a) The process that dominates each region

Regions	Events			Overall [%]			Raw fraction			$\frac{S}{S+B}$	$\frac{S}{\sqrt{B}}$
	W_0Z_Xjj-EW	other	Bkg.	W_0Z_Xjj-EW	other	Bkg.	W_0Z_Xjj-EW	other	Bkg.		
1	3.5	21.2	336.9	12.9	28.8	53.0	0.124	0.287	0.468	0.010	0.18
2	2.5	3.1	163.7	9.2	4.1	25.8	0.090	0.042	0.214	0.015	0.19
3	3.5	31.4	44.4	13.1	42.5	7.0	0.127	0.413	0.120	0.045	0.41
4	2.4	2.1	36.5	9.0	2.8	5.7	0.090	0.029	0.060	0.059	0.39
5	2.8	5.1	26.3	10.4	6.9	4.1	0.103	0.070	0.058	0.081	0.50
6	2.8	5.4	5.6	10.3	7.3	0.9	0.103	0.075	0.022	0.201	0.84
7	1.9	1.1	10.6	7.1	1.5	1.7	0.074	0.016	0.024	0.141	0.56
8	2.3	1.3	6.7	8.6	1.7	1.1	0.089	0.019	0.017	0.222	0.81
9	2.6	2.4	3.0	9.8	3.2	0.5	0.100	0.035	0.012	0.328	1.13
10	2.6	0.9	1.3	9.6	1.2	0.2	0.101	0.014	0.006	0.534	1.72

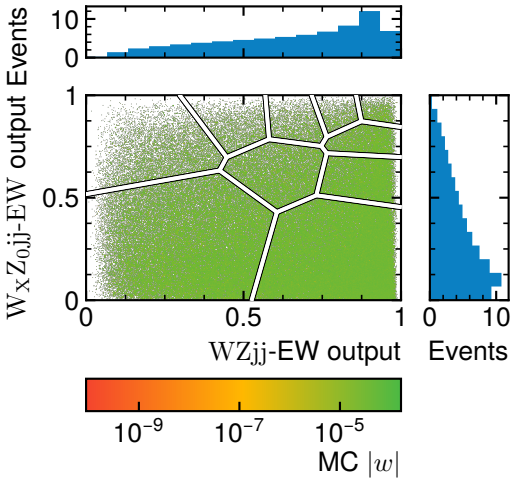
(b) The yields behind it

Figure N.4: The Voronoi regions of the W_0Z_Xjj-EW configuration, for Run-2: the process that dominates each one (a), and the yields behind that (b). The region numbers are the same in the map and in the table, and each group of three columns gives the polarised signal, the other polarisations, and the background, in that order.

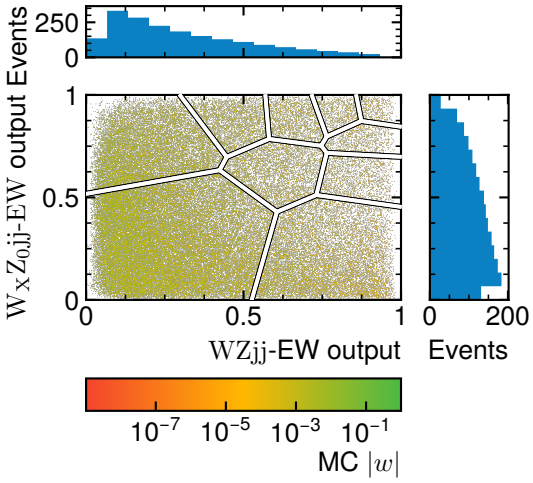
N.3 Single polarisation: the X0 plane



(a) $W_X Z_0$jj-EW

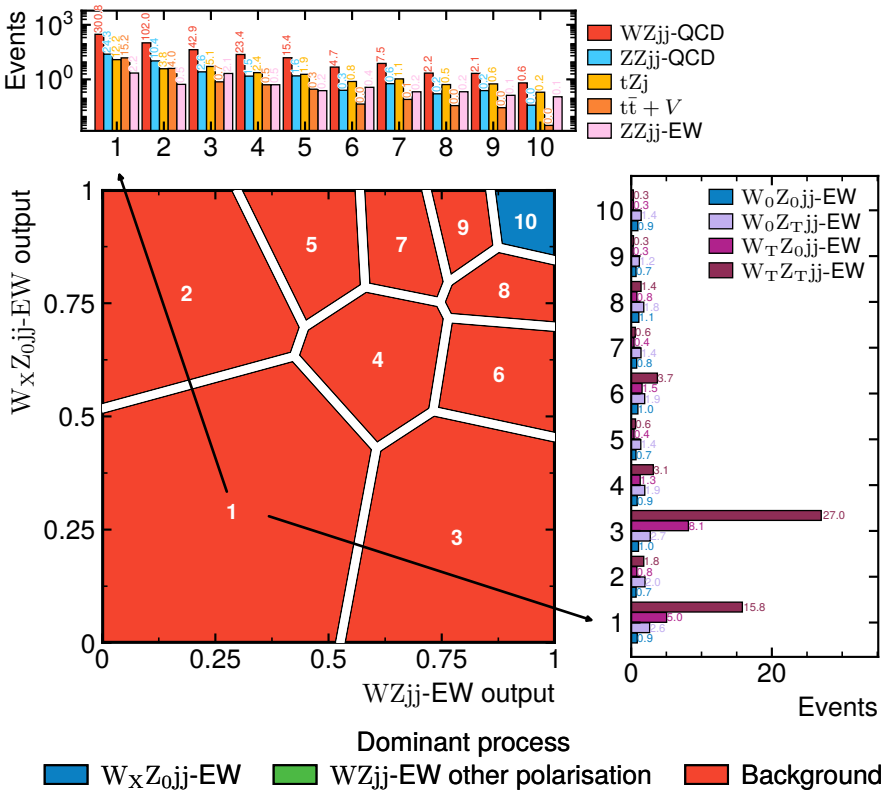


(b) $W_X Z_T$jj-EW



(c) Background

Figure N.5: Per-class Voronoi densities, $W_X Z_0$jj-EW configuration.



(a) The process that dominates each region

Regions	Events			Overall [%]			Raw fraction			$\frac{S}{S+B}$	$\frac{S}{\sqrt{B}}$
	W_XZ_0jj-EW	other	Bkg.	W_XZ_0jj-EW	other	Bkg.	W_XZ_0jj-EW	other	Bkg.		
1	3.6	20.9	375.5	12.9	28.6	59.1	0.125	0.286	0.512	0.009	0.18
2	2.7	2.4	128.5	9.8	3.3	20.2	0.097	0.033	0.172	0.020	0.24
3	3.9	35.5	56.6	14.0	48.7	8.9	0.136	0.478	0.150	0.040	0.40
4	2.9	4.3	29.3	10.4	5.9	4.6	0.103	0.060	0.054	0.079	0.49
5	2.1	0.9	21.6	7.6	1.2	3.4	0.077	0.013	0.039	0.086	0.44
6	3.0	5.2	6.7	10.7	7.1	1.0	0.107	0.073	0.021	0.200	0.86
7	2.2	0.8	9.6	8.1	1.2	1.5	0.083	0.013	0.022	0.175	0.69
8	3.0	2.1	3.2	10.8	2.8	0.5	0.110	0.031	0.012	0.359	1.29
9	1.9	0.5	3.2	6.9	0.7	0.5	0.071	0.007	0.010	0.341	0.99
10	2.4	0.5	1.0	8.8	0.6	0.2	0.092	0.007	0.006	0.620	1.99

(b) The yields behind it

Figure N.6: The Voronoi regions of the W_XZ_0jj-EW configuration, for Run-2: the process that dominates each one (a), and the yields behind that (b). The region numbers are the same in the map and in the table, and each group of three columns gives the polarised signal, the other polarisations, and the background, in that order.

O Statistical methods for the template fit

This appendix presents the statistical framework used for the measurements, from the general principles of likelihood-based inference to their application in the binned profile-likelihood fit, as implemented in the `HistFactory` [421] formalism used by `TRExFitter` [420].

The presentation of the general theory follows the statistics review from the PDG [34] and [439].

O.1 Likelihood-based inference

A measurement in particle physics never accesses a parameter directly. What is recorded is a finite and statistically fluctuating sample of events, from which the parameters of the underlying model, here the fiducial cross-section and the polarisation fractions, must be *inferred* and their uncertainties quantified. Statistics provides the principled tools for this inference. Two complementary approaches are distinguished by the meaning they assign to probability:

Frequentist. Probability is the limiting frequency of an outcome in an experiment repeated $N \to \infty$ times. A statement such as "the fraction lies in $[a, b]$ at $68\,\%$ confidence" is then a property of the *procedure*: it would bracket the true value in $68\,\%$ of repetitions.

Bayesian. Probability quantifies a degree of belief in a hypothesis. It allows a probability density to be assigned to a parameter itself, reflecting the state of knowledge about where its true value lies.

The measurements presented in this thesis are carried out in the frequentist framework. The Bayesian approach is used for Bayesian optimisation in Section 4.5 and briefly discussed in Section 4.6 for the Markov chains.

The outcome of an experiment is a set of measured values collected in a vector $\boldsymbol{x}$. A hypothesis H is a statement about the probability of the data, $P(\boldsymbol{x}|H)$. The same quantity, read as a function of the hypothesis with the data held *fixed*, is called the *likelihood* of H and written $L(H)$. This change of viewpoint is the conceptual foundation of what follows: the likelihood is not a probability density for H, it is the probability of the observed data evaluated for each candidate hypothesis. When H is specified by a set of parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_M)$, the likelihood function is $L(\boldsymbol{\theta}) = P(\boldsymbol{x}|\boldsymbol{\theta})$.

The most natural estimate of $\boldsymbol{\theta}$ is the set of values under which the observed data are the most probable. These are the maximum likelihood estimates (MLEs) $\hat{\boldsymbol{\theta}}$, defined as the maximum of L and found by solving the likelihood equations:

$$\frac{\partial \ln L}{\partial \theta_i} = 0\,, \qquad i = 1, \ldots, M\,. \tag{O.1}$$

The logarithm is used purely for practicality: it turns the products of probabilities encountered below into numerically stable sums, and since it is monotonic it is maximised by the same $\boldsymbol{\theta}$. Beyond being intuitive, the MLE is favoured for its behaviour in the large-sample, or *asymptotic*, limit, where it is consistent (it converges to the true value as data accumulate), asymptotically unbiased, and efficient, meaning its variance reaches the smallest value attainable by any unbiased estimator.

In the same limit the likelihood tends to a Gaussian in the parameters, so that the LL becomes a quadratic function:

$$\ln L(\boldsymbol{\theta}) = \ln L_{\max} - \frac{1}{2}\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right)^{\intercal} \mathbf{V}^{-1} \left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right)\,, \tag{O.2}$$

where $\ln L_{\max}$ is the value of $\ln L$ at the MLEs and $\mathbf{V}$ is the covariance matrix of the estimators. Writing $\mathbb{E}[\cdot]$ for the expectation value, the probability-weighted average of a quantity, its elements are the covariances $V_{ij} = \operatorname{cov}(\hat{\theta}_i, \hat{\theta}_j) = \mathbb{E}\big[(\hat{\theta}_i - \mathbb{E}[\hat{\theta}_i])(\hat{\theta}_j - \mathbb{E}[\hat{\theta}_j])\big]$, with the diagonal $V_{ii} = \mathrm{V}[\hat{\theta}_i]$ the variance of $\hat{\theta}_i$. The inverse $\mathbf{V}^{-1}$ is the curvature of the LL at its maximum:

$$V_{ij} = \operatorname{cov}\left(\hat{\theta}_i, \hat{\theta}_j\right) , \qquad V_{ij}^{-1} = - \left. \frac{\partial^2 \ln L}{\partial \theta_i \partial \theta_j} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} . \tag{O.3}$$

This relation has a simple reading: the more sharply the likelihood peaks around its maximum, the more the data constrain the parameters, and the smaller their uncertainty. In this regime the standard deviations are obtained directly from the likelihood itself: the interval containing s standard deviations is bounded by the values of $\boldsymbol{\theta}$ at which the LL has fallen from its maximum by $\frac{s^2}{2}$:

$$\ln L(\boldsymbol{\theta}) = \ln L_{\max} - \frac{s^2}{2} . \tag{O.4}$$

For a single parameter and $s = 1$, the $\pm 1\sigma$ interval is therefore delimited by a drop of $\frac{1}{2}$ in $\ln L$, equivalently a rise of one unit in $-2 \ln L$. The geometry of this construction is shown in Figure O.2. This prescription is independent of the number of parameters and is the practical rule by which uncertainties are read off a likelihood scan.

The results above are general. To apply them, the likelihood must be written down explicitly, and its form depends on how the data are structured.

When the data are a collection of n statistically independent events, each an independent draw from the same probability density function f, as for the selected collision events of an analysis, the joint probability factorises and the likelihood is the product of the per-event densities:

$$L(\boldsymbol{\theta}) = P(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta}) , \tag{O.5}$$

with n regarded as fixed.

Often the number of observed events n itself depends on $\boldsymbol{\theta}$, for instance when a cross-section fixes the expected event rate. Discarding this dependence would throw away the information carried by the overall yield. If n follows a Poisson distribution of mean μ, the dependence is restored by multiplying Equation O.5 by the corresponding Poisson term:

$$L(\boldsymbol{\theta}) = P(\boldsymbol{x}, n|\boldsymbol{\theta}) = P(n)P(\boldsymbol{x}|n) = \frac{\mu^n}{n!} \mathrm{e}^{-\mu} \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta}) . \tag{O.6}$$

This is the *extended* likelihood: when μ is a function of $\boldsymbol{\theta}$, including the probability of n adds information about the parameters and reduces their statistical uncertainties.

For a large sample it is convenient to group the events into N bins of a histogram, yielding a vector of counts $\boldsymbol{n} = (n_1, \ldots, n_N)$ with expectation values $\boldsymbol{\mu} = \mathbb{E}[\boldsymbol{n}]$. The likelihood is then maximised on the bin contents, the mean values $\boldsymbol{\mu}$ being expressed as functions of the parameters $\boldsymbol{\theta}$. Two situations arise, depending on the treatment of the total number of entries $n_{\text{tot}} = \sum_i n_i$:

- If n_{tot} is fixed by construction, the counts follow a *multinomial* distribution:

$$f(\boldsymbol{n}; \boldsymbol{\theta}) = \frac{n_{\text{tot}}!}{n_1! \ldots n_N!} p_1^{n_1} \ldots p_N^{n_N} , \tag{O.7}$$

  where the bin probabilities p_i are functions of the parameters $\boldsymbol{\theta}$.

- If the entries in each bin are instead independent and Poisson-distributed, with $\mu_i = \mu_{\text{tot}} p_i$, the data are described by a product of Poisson probabilities:

$$f(\boldsymbol{n}; \boldsymbol{\theta}) = \prod_{i=1}^{N} \frac{\mu_i^{n_i}}{n_i!} \mathrm{e}^{-\mu_i} , \tag{O.8}$$

  the total number of entries n_{tot} then following a Poisson distribution of mean $\mu_{\text{tot}} = \sum_i \mu_i$.

In collider physics the number of selected events is not fixed in advance but fluctuates from one dataset to the next, so the Poisson form of Equation O.8 is the natural choice. It is the starting point for the fit model of Section O.3.

O.2 Statistical tests and significance

Estimating the parameters is only the first part: one must also decide whether the data support a given hypothesis, for instance whether a signal is present at all. This is the role of a statistical test, which quantifies the compatibility between the data and a hypothesis and expresses it as a significance.

The agreement between the data and a hypothesis is condensed into a single number, a test statistic $t(\boldsymbol{x})$, constructed so that larger values correspond to stronger disagreement. The hypothesis under test, the null hypothesis H_0, typically the absence of signal, fixes the probability density function $f(t|H_0)$ of the statistic. The level of disagreement is measured by the p-*value*, the probability, assuming H_0, of obtaining a value of t at least as extreme as the one observed:

$$p = \int_{t_{\mathrm{obs}}}^{\infty} f(t|H_0)\,\mathrm{d}\,t\,, \tag{O.9}$$

where t_{obs} is the value measured on the data. The value p is *not* the probability that H_0 is true, it is the probability of the data under H_0.

A p-value is conventionally re-expressed as a significance Z, defined as the number of standard deviations at which the upper tail of a standard Gaussian encloses a probability p:

$$Z = \Phi^{-1}(1-p)\,, \tag{O.10}$$

as illustrated in Figure O.1. Here Φ is the cumulative distribution of the standard normal:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-t^2/2}\mathrm{d}\,t = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)\right]\,, \tag{O.11}$$

expressed through the error function:

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_{0}^{x} \mathrm{e}^{-t^2}\,\mathrm{d}\,t\,, \tag{O.12}$$

whose inverse, the quantile, returns the significance from the p-value:

$$\Phi^{-1}(p) = \sqrt{2}\,\mathrm{erf}^{-1}(2p-1)\,. \tag{O.13}$$

By this convention the discovery threshold used in particle physics, $Z = 5$, corresponds to $p = 2.87 \cdot 10^{-7}$, and the evidence level $Z = 3$ to $p \simeq 1.3 \cdot 10^{-3}$. The full correspondence is collected in Table O.A.

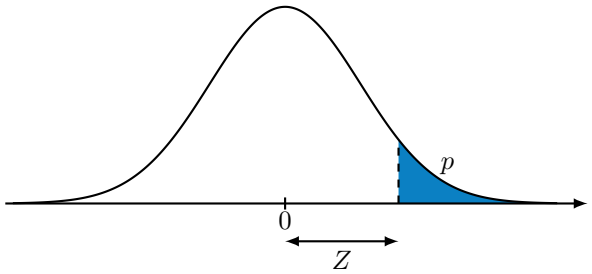


Figure O.1: Relation between the p-value and the significance Z of Equation O.10. The curve is the standard normal, the shaded area is the upper-tail probability p, and Z is the corresponding quantile.

Table O.A: One-sided p-value corresponding to a significance Z.

Z $[\sigma]$	1	2	3	4	5
p	$1.59 \cdot 10^{-1}$	$2.28 \cdot 10^{-2}$	$1.35 \cdot 10^{-3}$	$3.17 \cdot 10^{-5}$	$2.87 \cdot 10^{-7}$

It remains to choose the test statistic. The parameters are separated into the POIs ψ and the NPs $\boldsymbol{\theta}$, the latter needed for an accurate model but not of direct interest. Maximising the likelihood freely over all of them gives the unconditional maximum $L(\hat{\psi}, \hat{\boldsymbol{\theta}})$. According to the Neyman–Pearson lemma, the ratio of the likelihoods of two simple hypotheses, $\lambda = \frac{L(H_1)}{L(H_0)}$, is the most powerful discriminant between them.

For composite hypotheses containing NPs this generalises to the *profile-likelihood ratio*: for each tested value of the POIs the NPs are profiled to their conditional estimators $\hat{\hat{\boldsymbol{\theta}}}(\boldsymbol{\psi})$, giving:

$$\lambda_p(\boldsymbol{\psi}) = \frac{L\big(\boldsymbol{\psi}, \hat{\hat{\boldsymbol{\theta}}}(\boldsymbol{\psi})\big)}{L\big(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\theta}}\big)}\,, \tag{O.14}$$

with the conditional (profiled) likelihood in the numerator and the unconditional maximum in the denominator, so that $0 \leq \lambda_p \leq 1$, with $\lambda_p = 1$ when the tested value coincides with the estimate.

The usefulness of Equation O.14 rests on Wilks' theorem [440]: in the large-sample limit, and for M POIs under test, the statistic:

$$-2\ln\lambda_p(\boldsymbol{\psi}) = -2\left[\ln L\big(\boldsymbol{\psi}, \hat{\hat{\boldsymbol{\theta}}}(\boldsymbol{\psi})\big) - \ln L\big(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\theta}}\big)\right]\,, \tag{O.15}$$

follows a χ^2 distribution with M degrees of freedom (Equation O.26), independently of the profiled NPs. For a single POI it reduces to the parabola $-2\ln\lambda_p(\psi) \to \left(\frac{\psi-\hat{\psi}}{\sigma_\psi}\right)^2$, in agreement with the Gaussian regime of Equation O.2. The p-value then follows from the upper tail of the χ^2 distribution, that is by subtracting its cumulative distribution from unity, and is converted to a significance through Equation O.10. Because the distribution is known analytically, the significance can be evaluated whenever the asymptotic conditions are satisfied. Their adequacy for the model used here is confirmed with pseudo-experiments.

Two significances are quoted throughout. The *observed* significance is computed on the measured data and states whether the deviation from H_0 is large enough to reject it. The *expected* significance is the value that would be obtained, on average, were the alternative hypothesis true. It is evaluated on a representative "Asimov" dataset, built with the signal injected at its predicted values, and quantifies the sensitivity of the analysis independently of the particular fluctuation seen in the data.

O.3 The binned likelihood model of the fit

Specialising the binned Poisson likelihood of Equation O.8 to the measurement, the data are the observed bin contents $\boldsymbol{n} = (n_1, \ldots, n_{N_{\text{bin}}})$ of all fitted distributions, that is the unrolled Voronoi template in the SR and the discriminants of the ZZ- and b-CRs, for both the Run-2 and the partial Run-3 datasets, each bin treated as an independent Poisson count. The full set of parameters is $\boldsymbol{\alpha} = (\boldsymbol{\psi}, \boldsymbol{\theta})$, with the POIs $\boldsymbol{\psi} = (\mu^{\text{fid.}}_{\text{WZjj-EW}}, f_{\text{XX}})$ the two free parameters of the measurement and $\boldsymbol{\theta}$ the NPs, namely the systematic uncertainties and the free background NFs $\mu_{\text{ZZjj-QCD}}$, $\mu_{\text{t}\bar{\text{t}}+V}$, and μ_{tZj}.

The expected content of bin i, written ν_i (the μ_i of Equation O.8, renamed to reserve μ for the per-sample normalisations below), is the sum over the contributing samples s:

$$\nu_i(\boldsymbol{\alpha}) = \sum_s \mu_s(\boldsymbol{\alpha})\tau_{s,i}(\boldsymbol{\theta})\,, \tag{O.16}$$

where $\mu_s(\boldsymbol{\alpha})$ is the total normalisation of sample s, that is the product of the relevant NFs, including the derived μ_{XX} and $\mu_{\overline{\text{XX}}}$ of the main text, and of the normalisation systematics, while $\tau_{s,i}(\boldsymbol{\theta})$ is the nominal per-bin template yield modulated by the shape systematics. Each systematic source p is represented by a single nuisance parameter θ_p, whose response for μ_s or $\tau_{s,i}$ is provided at $\theta_p = 0$ (nominal) and $\theta_p = \pm 1$ (the "Up" and "Down" variations). Systematics whose bin-by-bin effect is everywhere below the pruning threshold are removed, as described in Section 7.2.3.

The likelihood is the extended binned likelihood of Equations O.6 and O.8 augmented by a constraint term per NP:

$$L(\boldsymbol{\psi}, \boldsymbol{\theta}) = \underbrace{\prod_{i=1}^{N_{\text{bin}}} \frac{\nu_i(\boldsymbol{\psi}, \boldsymbol{\theta})^{n_i} \mathrm{e}^{-\nu_i(\boldsymbol{\psi}, \boldsymbol{\theta})}}{n_i!}}_{\text{data}} \times \underbrace{\prod_p f_p(\theta_p)}_{\text{constraints}}\,. \tag{O.17}$$

The constraint terms f_p encode the auxiliary information on the NPs and take the following forms:

- A *Gaussian* constraint, $f_p(\theta_p) \propto \exp\left[-\frac{1}{2}(\theta_p - \theta_p^0)^2\right]$, for the experimental and theoretical systematics derived from an external calibration or auxiliary measurement. By convention the NPs are centred at $\theta_p^0 = 0$ with unit width, so that the post-fit value $\hat{\theta}_p$ is read directly as a "pull" in units of the input uncertainty.
- A Gaussian constraint with an *exponential* (log-normal-like) response for the normalisation systematics, which keeps the corresponding yield non-negative.

- A *Poisson* (γ) constraint for the finite size of the MC samples, implemented as one additional NP per bin in which the MC statistical uncertainty exceeds $1\,\%$, its width set by the effective number of simulated events in that bin.
- *No* constraint (a uniform term) for the POIs and for the free background NFs, which are determined by the data alone.

The POIs are bounded to their physical range, and the free NFs to non-negative values.

O.4 Parameter estimation

The parameters are estimated by maximum likelihood, the fit minimising $-2\ln L(\boldsymbol{\alpha})$ since $-2\ln L$ is both numerically convenient and asymptotically χ^2-distributed. The covariance of the estimators follows from the curvature of the LL at the minimum through Equation O.3, and the one-standard-deviation uncertainty on a single parameter is the interval over which $-2\ln L$ rises by one unit, as in Equation O.4. The near-parabolic LL scans reported in Appendix S confirm that this Gaussian regime holds for $\mu^{\text{fid.}}_{\text{WZjj-EW}}$ and, for the f_{X0} measurement, for the fraction as well.

To isolate the POIs from the NPs, the latter are *profiled* as in the ratio of Equation O.14: for a fixed $\boldsymbol{\psi}$ the conditional estimators $\hat{\hat{\boldsymbol{\theta}}}(\boldsymbol{\psi})$ define the profile likelihood:

$$L_p(\boldsymbol{\psi}) = L\big(\boldsymbol{\psi}, \hat{\hat{\boldsymbol{\theta}}}(\boldsymbol{\psi})\big)\,. \tag{O.18}$$

Profiling propagates the systematic uncertainties into the POIs: the presence of NPs enlarges the POI variance through their correlations, whereas any external or in-situ constraint, such as that provided by the two CRs on the dominant backgrounds, reduces it. The correlations among all parameters are summarised by the post-fit correlation matrix of Section 7.3.5.

Two post-fit quantities characterise the behaviour of each NP and the stability of the fit. The *pull*, $\frac{\hat{\theta}_p - \theta_p^0}{\Delta\theta_p}$, is the shift of the post-fit value from its pre-fit input θ_p^0 in units of the pre-fit width $\Delta\theta_p$, and is expected to be compatible with zero within its post-fit uncertainty. The *constraint*, $\frac{\hat{\sigma}_p}{\Delta\theta_p}$, is the ratio of the post-fit to the pre-fit uncertainty. A value below unity signals that the data constrain the parameter beyond its prior.

The contribution of an NP θ_p to the total uncertainty on the POI ψ is read directly from the post-fit covariance matrix following [431], at negligible cost relative to a per-NP re-fit. In the Gaussian regime this uncertainty component is:

$$\mathcal{C}_p = \sigma_\psi \rho_{\psi\theta_p} \hat{\sigma}_p = \operatorname{cov}(\psi, \theta_p)\,, \tag{O.19}$$

with σ_ψ the POI uncertainty, $\rho_{\psi\theta_p}$ the post-fit POI–NP correlation, and $\hat{\sigma}_p$ the post-fit uncertainty of θ_p. It quantifies the uncertainty propagated to $\hat{\psi}$ from the external 1σ constraint on θ_p. Because the constraint terms are independent, the components of distinct NPs add in quadrature to the total systematic uncertainty and may be propagated to downstream combinations [431]. This is to be distinguished from the systematic *impact* $\rho_{\psi\theta_p}\sigma_\psi$, the shift of $\hat{\psi}$ when θ_p is displaced by its post-fit uncertainty, which measures the sensitivity of the fit to θ_p rather than its contribution and does not share this additivity. The two are related through the post-fit width, $\mathcal{C}_p = \hat{\sigma}_p(\rho_{\psi\theta_p}\sigma_\psi)$, and coincide only when the parameter is unconstrained by the data ($\hat{\sigma}_p = \Delta\theta_p$). Its up and down variations, built from the asymmetric post-fit POI and NP uncertainties, are drawn as the two coloured boxes of Figure 7.13, and the NPs are ordered, as `TRExFitter` does, by their post-fit impact $\max\left(|\rho^{+}_{\psi\theta_p}\sigma_\psi|, |\rho^{-}_{\psi\theta_p}\sigma_\psi|\right)$. The black markers on the upper axis show the pulls and their post-fit constraints defined above. The complete set is collected in Appendix R.

O.5 Significance of the measurement

The workflow of Section O.2 is applied here to the null hypothesis that the targeted longitudinal fraction vanishes: $f_{\text{XX}} = 0$, which is tested for each measurement to obtain the significances of Table 7.I. Because only an excess of longitudinal signal ($\hat{f} > 0$) constitutes evidence against the null, the appropriate one-sided statistic is:

$$q_0 = \begin{cases} -2\ln\lambda_p(f_{\text{XX}} = 0) & \hat{f} \geq 0\,, \\ 0 & \hat{f} < 0\,, \end{cases} \tag{O.20}$$

with λ_p the profile-likelihood ratio of Equation O.14. By Wilks' theorem for a single POI ($M = 1$), the discovery significance is:

$$Z_0 = \sqrt{q_0}\,. \tag{O.21}$$

The *expected* significances of Table 7.I(a) are obtained without pseudo-experiments through the asymptotic approximation [439, 441]. The test statistic is evaluated on an "Asimov" dataset: the artificial dataset in which every bin is set exactly to its expected content, with the POIs fixed to the injected signal values (the predicted fractions and $\mu^{\text{fid.}}_{\text{WZjj-EW}} = 1$) and the NPs at their nominal values. Because this dataset carries no statistical fluctuation, a fit to it (the "Asimov fit") recovers the injected parameters exactly, every NP returning a null pull, and the significance with which it rejects $f_{\text{XX}} = 0$ is the median significance expected under the Standard Model hypothesis, in which the longitudinal fraction takes its predicted value. The signal is injected deliberately: a background-only dataset would contain no longitudinal component and return a vanishing expected significance by construction, so it is the injection at the predicted values that turns the number into a statement of sensitivity. For a single-bin counting experiment with S expected signal and B expected background events this reduces to the closed form:

$$\text{med}[Z_0] = \sqrt{2\big[(S+B)\ln(1+\frac{S}{B}) - S\big]} \xrightarrow[S \ll B]{} \frac{S}{\sqrt{B}}\,, \tag{O.22}$$

the ratio $\frac{S}{\sqrt{B}}$ emerging only in the small-signal limit. The full multi-bin measurement uses the generalisation of Equation O.22 to the likelihood of Equation O.17, evaluated with the same Asimov fit.

When the estimator lies at the physical boundary $\hat{f} = 0$, the regularity conditions of Wilks' theorem are violated and the asymptotic distribution of q_0 becomes a mixture of a χ^2_1 and a δ-function at zero. The one-sided definition of Equation O.20 already accounts for this, and it is the reason a fluctuation to $\hat{f} = 0$ is reported as 0.0σ. It is also why, for the f_{00} and $f_{0\text{X}}$ measurements whose central values fall on the boundary, only upper limits are quoted.

Because the significances above rest on the asymptotic formulae of Wilks' theorem and of the Asimov approximation, whose validity is not guaranteed at the modest event yields and low significances of this measurement, the asymptotic treatment is validated with a dedicated pseudo-experiment study. For each polarisation measurement, $5 \cdot 10^4$ toys are generated under the background-only and signal-plus-background hypotheses and the statistic q_0 is evaluated on each. The resulting null distribution is compared with the asymptotic χ^2_1 and used to recompute the p-value from the value observed in data. The toy distributions reproduce the χ^2_1 form and yield significances in agreement with the asymptotic ones, confirming the approximation down to the low-significance regime probed here. The conclusion is moreover stable against the construction of the null hypothesis, whether the total WZjj-EW yield is held fixed or left to vary when the longitudinal fraction is set to zero.

O.6 Confidence intervals and limits

A confidence interval at CL $1-\alpha$ is a data-dependent interval constructed, following Neyman [442], so as to contain the true parameter value in a fraction $1-\alpha$ of repeated experiments, irrespective of that value. This coverage property, not a statement about the probability of the fixed observed interval, defines the frequentist CL. Equivalently, the interval is the set of parameter values not rejected by a test of size α.

In the asymptotic regime the interval is read directly off the profile-likelihood ratio: combining the Gaussian regime of Equation O.2 with Wilks' theorem, the $1-\alpha$ confidence region for M POIs is:

$$\big\{\boldsymbol{\psi} : -2\ln\lambda_p(\boldsymbol{\psi}) \le F^{-1}_{\chi^2_M}(1-\alpha)\big\}\,, \tag{O.23}$$

where $F^{-1}_{\chi^2_M}$ is the χ^2_M quantile. The thresholds are collected in Table O.B. For a single POI the $68.3\,\%$ interval corresponds to $-2\Delta\ln L = 1$, so the endpoints of Equation O.23 coincide with the $\pm1\sigma$ points of the 1D scans of Appendix S. The 2D contours of Figure 7.12 are drawn at $-2\Delta\ln L = 2.30$, 6.18, and 11.83, the 1σ, 2σ, and 3σ contours ($68.3\,\%$, $95.4\,\%$, and $99.7\,\%$). The threshold for exactly $95\,\%$ would be 5.99. The construction is illustrated in Figure O.2.

For a fraction bounded below at zero and compatible with the null, a one-sided upper limit is set rather than a two-sided interval. The unified frequentist construction of [443] treats precisely this case, using the likelihood ratio as the ordering rule in the Neyman construction to preserve coverage and to fix, from the data, whether a one- or two-sided interval is reported for a parameter near a physical boundary.

The corresponding test statistic $\tilde{q}_{f_{\text{XX}}}$ is the one-sided version of Equation O.14, truncated so that a downward fluctuation below the tested value is not counted as evidence against it. In the asymptotic regime, and with the p-value of the signal hypothesis alone (CL_{s+b}), the observed $95\,\%$ CL upper limit would be the value of the fraction at which the one-sided scan crosses:

$$-2\Delta\ln L = \big(\Phi^{-1}(0.95)\big)^2 = 2.71\,. \tag{O.24}$$

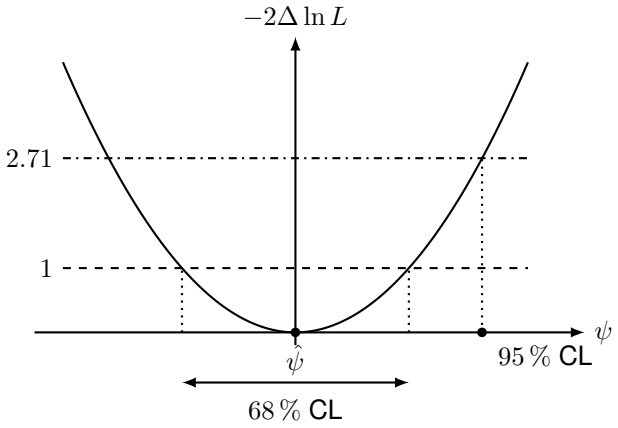


Figure O.2: Schematic profile-likelihood scan of a POI ψ. The $\pm 1\sigma$ ($68\,\%$ CL) interval is read where the curve crosses $-2\Delta \ln L = 1$, and the one-sided $95\,\%$ CL upper limit where it crosses $-2\Delta \ln L = 2.71$.

Table O.B: Threshold on $-2\Delta \ln L = \Delta\chi^2$ defining a confidence region of level $1-\alpha$ for the joint estimation of M parameters of interest, from Equation O.23.

$1-\alpha$	$M=1$	$M=2$	$M=3$
68.27 %	1.00	2.30	3.53
90 %	2.71	4.61	6.25
95 %	3.84	5.99	7.82
95.45 %	4.00	6.18	8.03
99 %	6.63	9.21	11.34
99.73 %	9.00	11.83	14.16

The limits of Section 7.3.11 are instead computed by `TRExFitter` with the asymptotic formulae and the CL_s criterion of Equation O.25, which for a best-fit value at the boundary ($1 - p_b \simeq 0.5$) corresponds to $p_{s+b} = 0.025$, that is to a crossing at $-2\Delta \ln L = 3.84$. The expected median limits and their bands are obtained from pseudo-experiments under the SM hypothesis.

Such a construction can nonetheless exclude a signal to which the analysis has little sensitivity, for instance when the data fluctuate below the background expectation. To avoid this, the modified-frequentist CL_s criterion [444] replaces the raw p-value of the signal hypothesis by:

$$\mathrm{CL}_s = \frac{p_\mu}{1 - p_b}\,, \tag{O.25}$$

with p_b the p-value of the background-only hypothesis, and excludes a value when $\mathrm{CL}_s \leq \alpha$. Since the denominator is at most unity, the criterion is conservative, its coverage exceeding the nominal level: where sensitivity vanishes the denominator shrinks in step with p_μ and exclusion is suppressed.

O.7 The χ^2 distribution and goodness-of-fit

The χ^2 distribution underlies both the interval estimation of Section O.6 and the goodness-of-fit tests below. If $z = \sum_{i=1}^{k} x_i^2$ is the sum of the squares of k independent standard-normal variables, then z follows the χ^2 distribution with k degrees of freedom:

$$f(z;k) = \frac{z^{k/2-1}\mathrm{e}^{-z/2}}{2^{k/2}\Gamma(\frac{k}{2})}\,, \qquad \mathbb{E}[z] = k\,, \qquad \mathrm{V}[z] = 2k\,, \tag{O.26}$$

with Γ the gamma function. For large k, the distribution tends to a Gaussian of mean k and variance $2k$.

It can be generalised to the case of uncorrelated measurements with different variances:

$$\chi^2(\boldsymbol{\alpha}) = \sum_{i=1}^{k} \frac{\big(x_i - \mu_i(\boldsymbol{\alpha})\big)^2}{\sigma_i^2}\,, \tag{O.27}$$

with $\mu_i(\boldsymbol{\alpha})$ the model prediction and σ_i the standard deviation of measurement x_i, so that each $\frac{x_i-\mu_i}{\sigma_i}$ is a standard-normal variable. The method of least squares estimates the parameters $\boldsymbol{\alpha}$ of a model by

minimising Equation O.27 over $\boldsymbol{\alpha}$. In the linear or asymptotic Gaussian regime, its minimum value $\chi^2_{\min}$ follows a χ^2 distribution with $k - M$ degrees of freedom, M being the number of fitted parameters.

Its central role is a consequence of Wilks' theorem: $-2\ln\lambda_p$ of Equation O.15 is asymptotically χ^2-distributed, so the thresholds of Table O.B and the significances of Section O.5 are all χ^2 quantiles. The same distribution governs the method of least squares, obtained from Equation O.17 in the Gaussian limit, for which $-2\ln L = \chi^2 + \text{const}$.

For a binned fit the agreement between the N_{bin} observed contents and the post-fit expectation can be quantified with Pearson's statistic:

$$\chi^2 = \sum_i \frac{(n_i - \hat{\nu}_i)^2}{\hat{\nu}_i}\,, \tag{O.28}$$

but the likelihood-ratio statistic against the *saturated model* [445] (one free parameter per bin, for which $\hat{\nu}_i = n_i$ and the likelihood is maximal):

$$-2\ln\lambda = 2\sum_{i=1}^{N_{\text{bin}}}\left[\hat{\nu}_i - n_i + n_i\ln\frac{n_i}{\hat{\nu}_i}\right]\,, \tag{O.29}$$

is preferred for Poisson data as it converges faster to the asymptotic χ^2. With M fitted parameters the number of degrees of freedom is $n_{\text{dof}} = N_{\text{bin}} - M$ for Poisson-distributed bins (and $N_{\text{bin}} - M - 1$ when the total normalisation is instead held fixed). The goodness-of-fit p-value is the upper-tail probability:

$$p = \int_{-2\ln\lambda}^{\infty} f(t; n_{\text{dof}})\,\mathrm{d}\,t\,, \tag{O.30}$$

and a "reasonable" fit has a reduced statistic $\frac{\chi^2}{n_{\text{dof}}} \approx 1$. A poor goodness-of-fit signals a mis-modelling but does not by itself imply large parameter uncertainties: the uncertainties of Equation O.3 reflect only how the estimates would scatter if the assumed model and input errors were correct, and do not include the systematic bias from an incorrect model. This test is the quantitative counterpart of the post-fit data/prediction agreement.

O.8 Mapping to the results of the main text

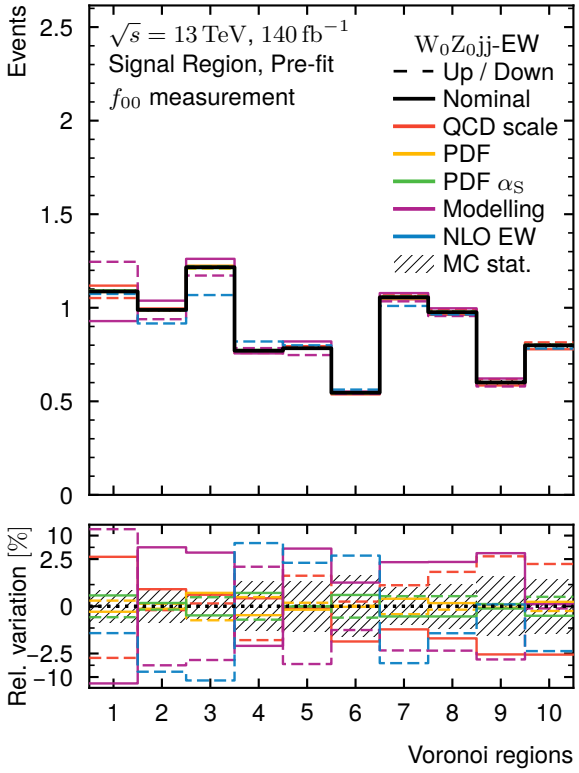
The elements above map onto the results of the main text as follows.

- The binned likelihood of Equation O.17 is minimised by `TRExFitter` to produce the post-fit distributions and yields. The free NFs constrain the dominant backgrounds through the two CRs.
- The 1D profile-likelihood scans of Appendix S give the $\pm1\sigma$ intervals via $-2\Delta\ln L = 1$. The $95\,\%$ CL upper limits are asymptotic CL_s limits (Equation O.25), their expected medians and bands being obtained with pseudo-experiments under the SM hypothesis. The 2D scans of Figure 7.12 give the 1σ, 2σ, and 3σ contours via $-2\Delta\ln L = 2.30/6.18/11.83$.
- The observed and expected significances of Table 7.I are $Z_0 = \sqrt{q_0}$ from Equations O.20 and O.21, the expected values evaluated on the signal-injected Asimov dataset (the Asimov fit) through Equation O.22. The asymptotic approximation underlying both is validated with pseudo-experiments.
- The pulls, constraints, and correlations of Appendix R, Figures 7.9 to 7.11 and the impact rankings of Figure 7.13 follow from the profiling of Equation O.18.

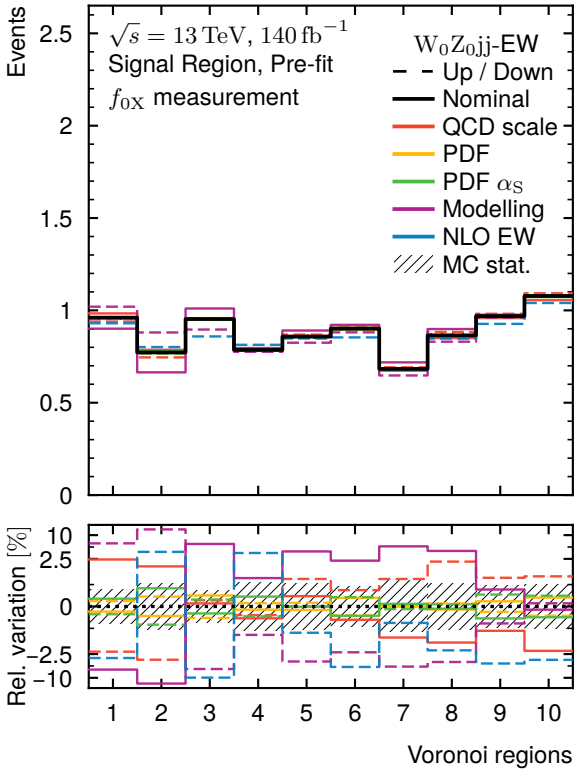
P Theoretical systematics representation

This appendix collects the theoretical modelling systematic uncertainties evaluated on the polarised WZjj EW signal templates and on the QCD-induced WZjj background. For each process the relative effect of every modelling variation on the observable entering the fit is shown for each of the four polarised templates ($\mathrm{W_0Z_0jj}$-EW, $\mathrm{W_0Z_Tjj}$-EW, $\mathrm{W_TZ_0jj}$-EW, $\mathrm{W_TZ_Tjj}$-EW) and for WZjj-QCD, in the three measurements: the joint f_{00} fit and the two single fits $f_{0\mathrm{X}}$ and $f_{\mathrm{X}0}$ (panels (a), (b), and (c) of each figure). Throughout, the subscripts 0 and T denote the longitudinal and transverse polarisation states, X a combination of both. Results are given separately for the Run-2 and Run-3 periods (one figure per period).

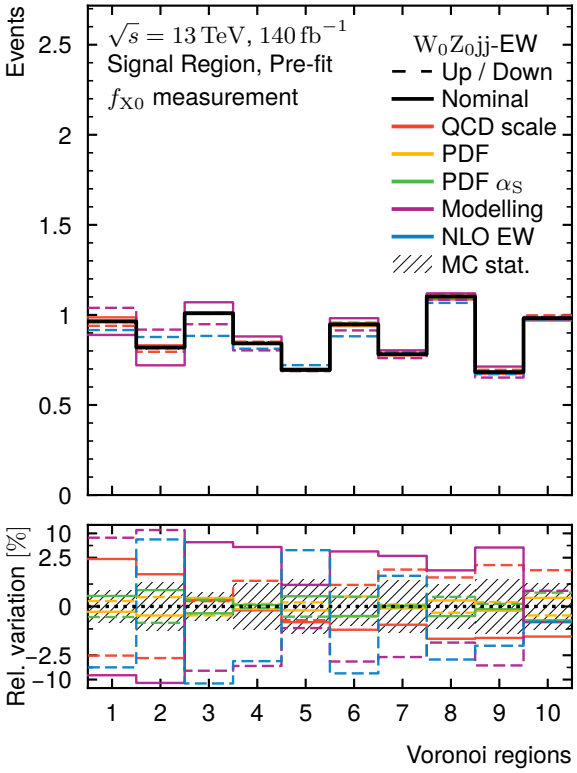
The list of theoretical modelling systematic variations considered in this analysis is given in Section 7.2.2. The effect of each variation is shown in Figures P.1, P.3, P.5, P.7, and P.9 for Run-2 and in Figures P.2, P.4, P.6, P.8, and P.10 for Run-3, each figure showing the f_{00}, $f_{0\mathrm{X}}$, and $f_{\mathrm{X}0}$ measurements. The effect is shown as a relative shift with respect to the nominal prediction as a function of the fit observable. The theoretical systematics include the variation of the QCD scale, the variation of the PDF set, the PDF α_{S}, the modelling (difference of generators), and, for the EW signal, the NLO EW corrections.



(a) f_{00} measurement

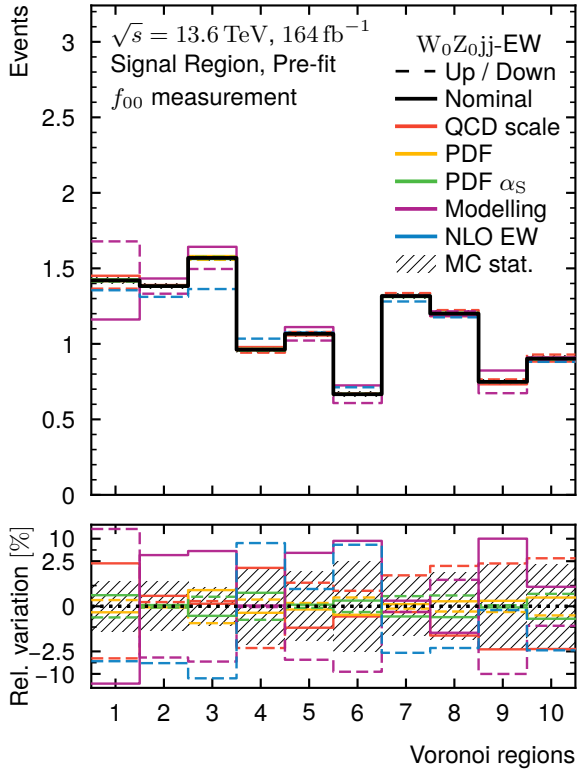


(b) f_{0X} measurement

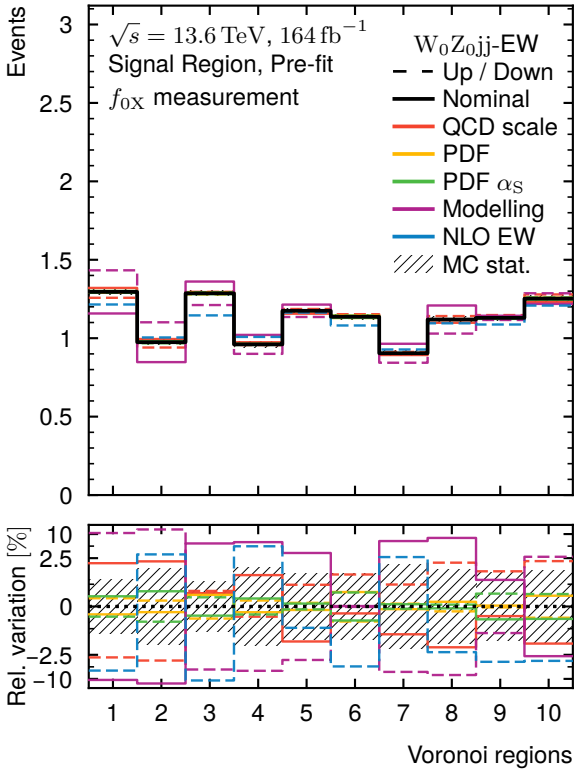


(c) f_{X0} measurement

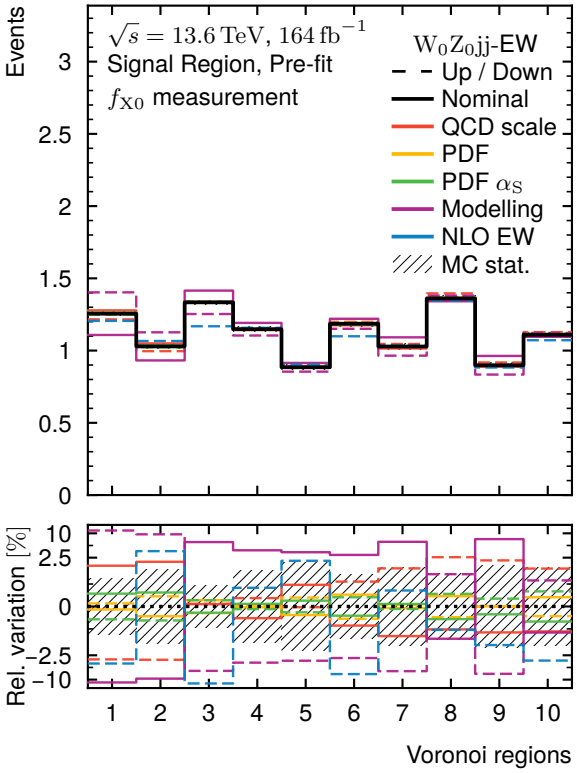
Figure P.1: Theoretical modelling systematic variations on the W_0Z_0jj-EW template, shown as relative shifts with respect to the nominal prediction as a function of the fit observable. The panels correspond to the f_{00}, f_{0X}, and f_{X0} measurements in Run-2.



(a) f_{00} measurement

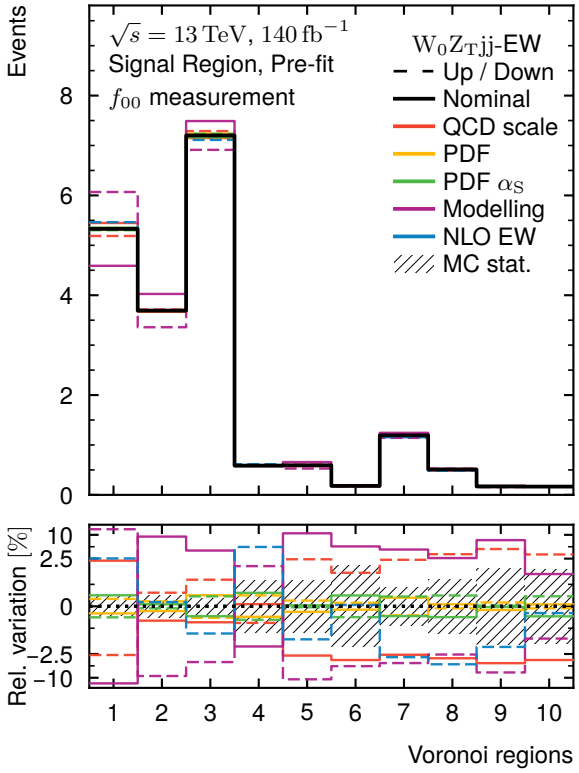


(b) f_{0X} measurement

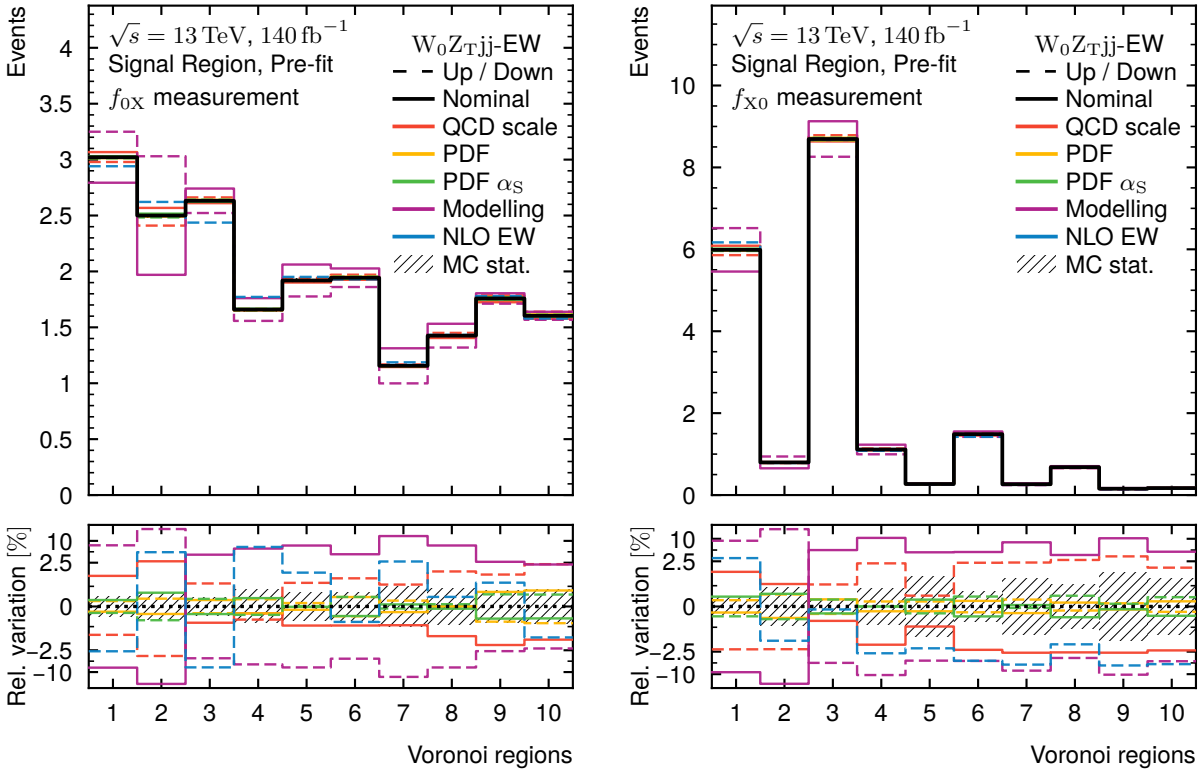


(c) f_{X0} measurement

Figure P.2: Theoretical modelling systematic variations on the W_0Z_0jj-EW template, shown as relative shifts with respect to the nominal prediction as a function of the fit observable. The panels correspond to the f_{00}, f_{0X}, and f_{X0} measurements in Run-3.



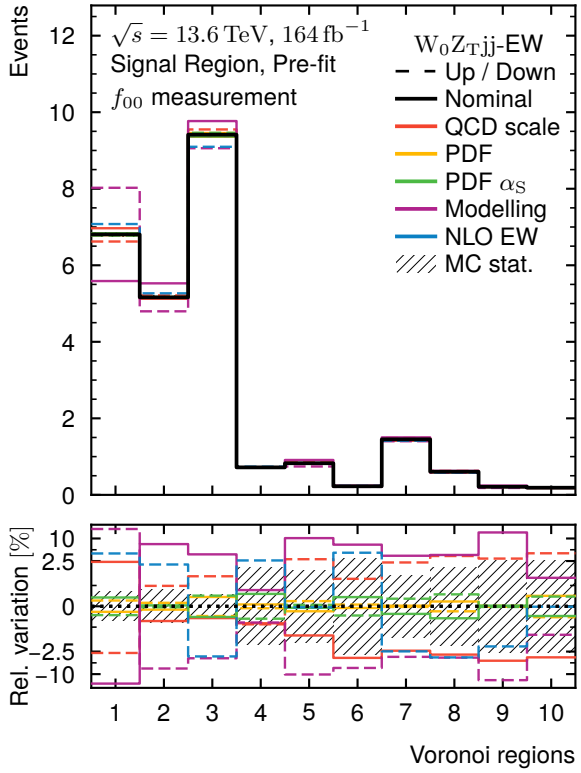
(a) f_{00} measurement



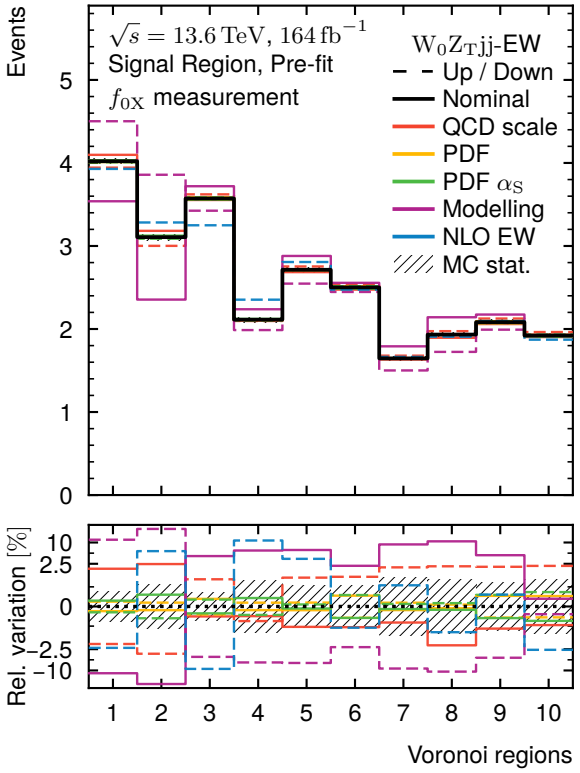
(b) f_{0X} measurement

(c) f_{X0} measurement

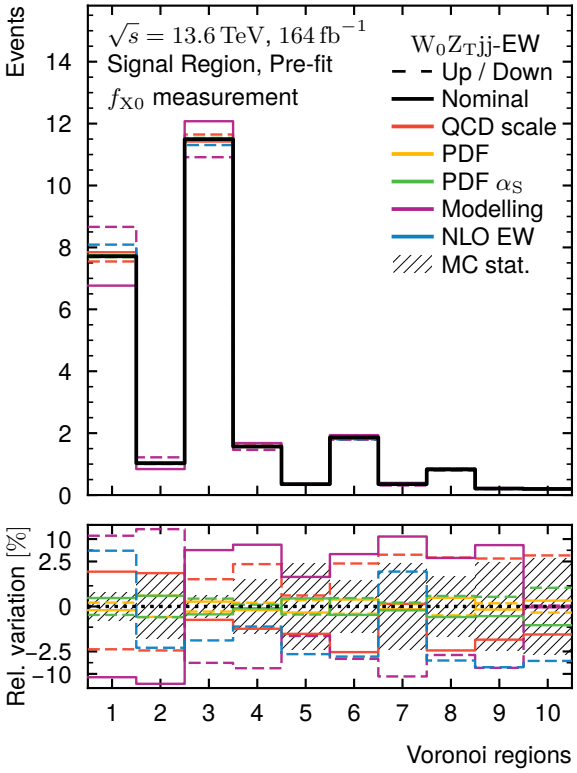
Figure P.3: Theoretical modelling systematic variations on the W_0Z_Tjj-EW template, shown as relative shifts with respect to the nominal prediction as a function of the fit observable. The panels correspond to the f_{00}, f_{0X}, and f_{X0} measurements in Run-2.



(a) f_{00} measurement

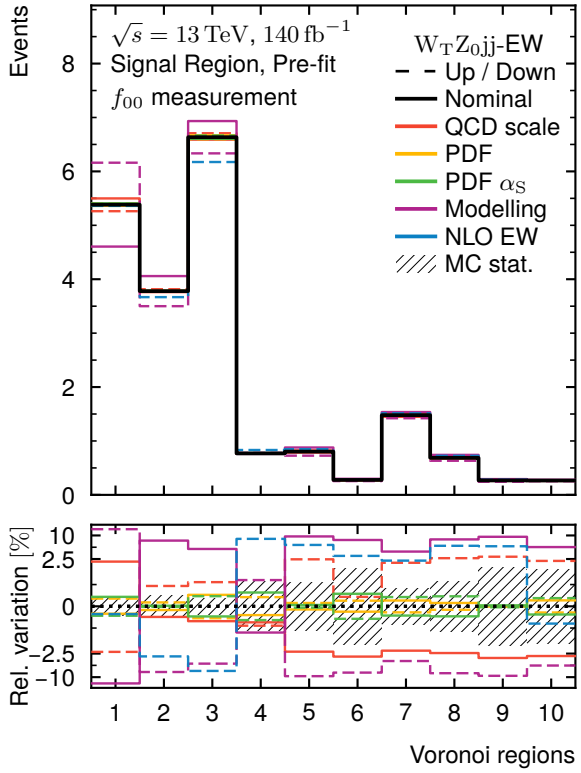


(b) f_{0X} measurement

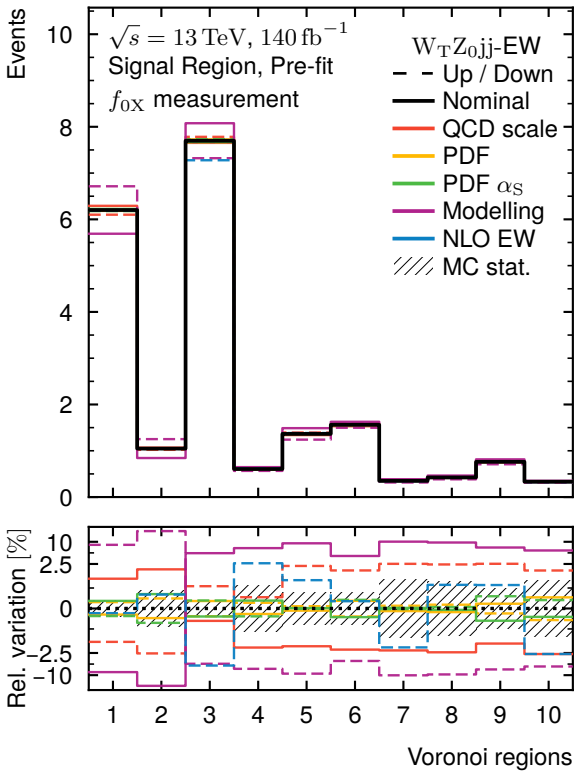


(c) f_{X0} measurement

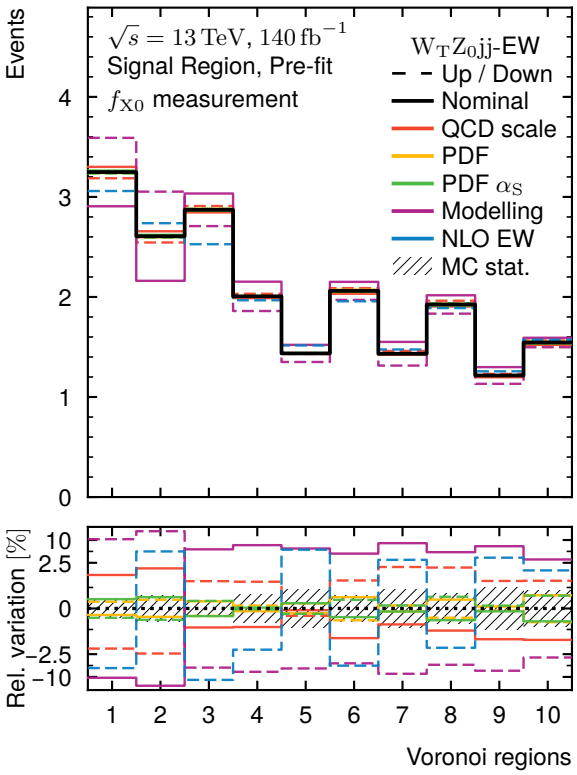
Figure P.4: Theoretical modelling systematic variations on the W_0Z_Tjj-EW template, shown as relative shifts with respect to the nominal prediction as a function of the fit observable. The panels correspond to the f_{00}, f_{0X}, and f_{X0} measurements in Run-3.



(a) f_{00} measurement

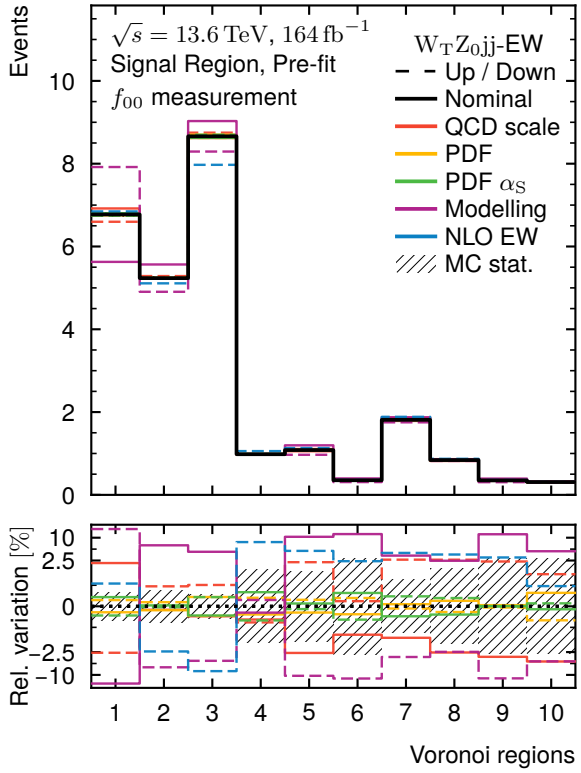


(b) f_{0X} measurement

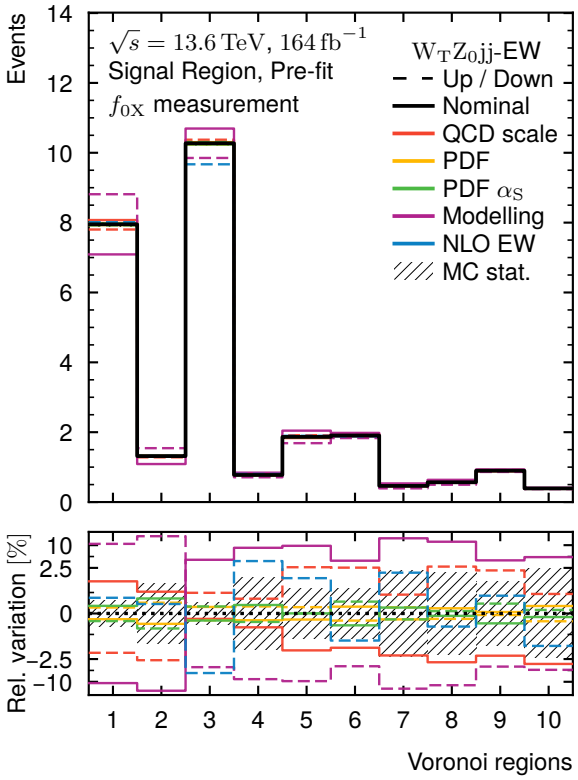


(c) f_{X0} measurement

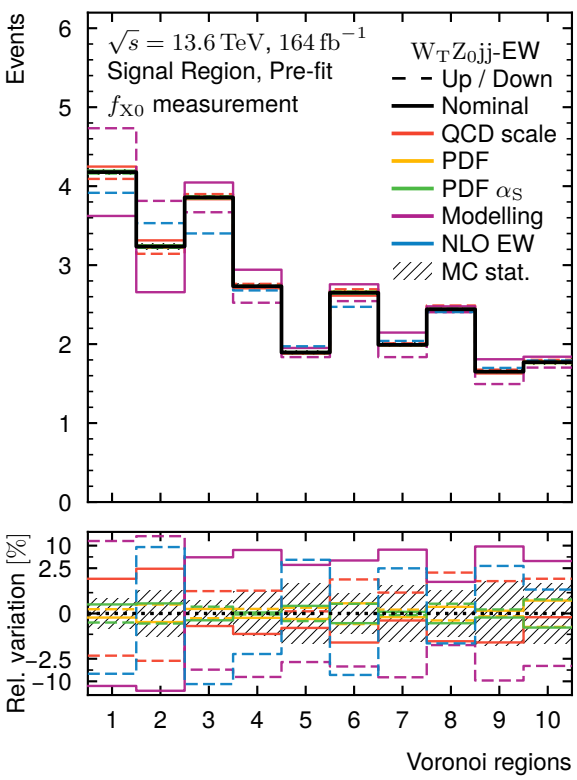
Figure P.5: Theoretical modelling systematic variations on the W_TZ_0jj-EW template, shown as relative shifts with respect to the nominal prediction as a function of the fit observable. The panels correspond to the f_{00}, f_{0X}, and f_{X0} measurements in Run-2.



(a) f_{00} measurement

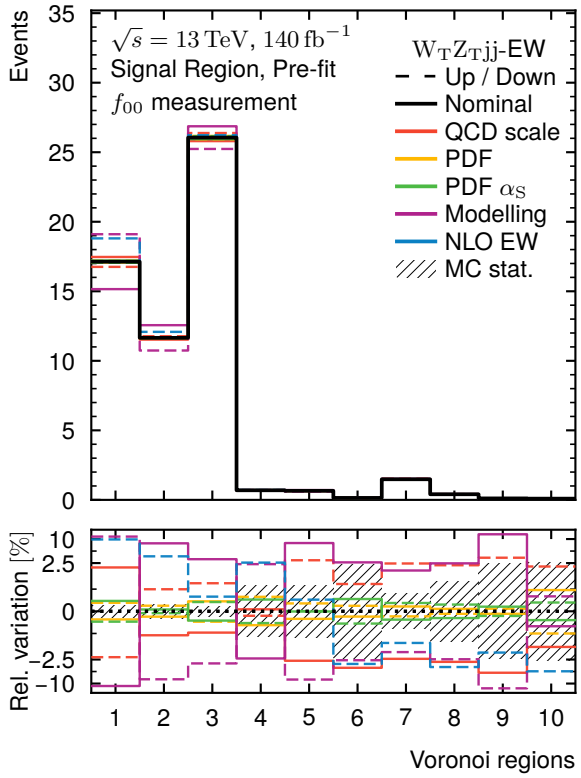


(b) $f_{0\text{X}}$ measurement

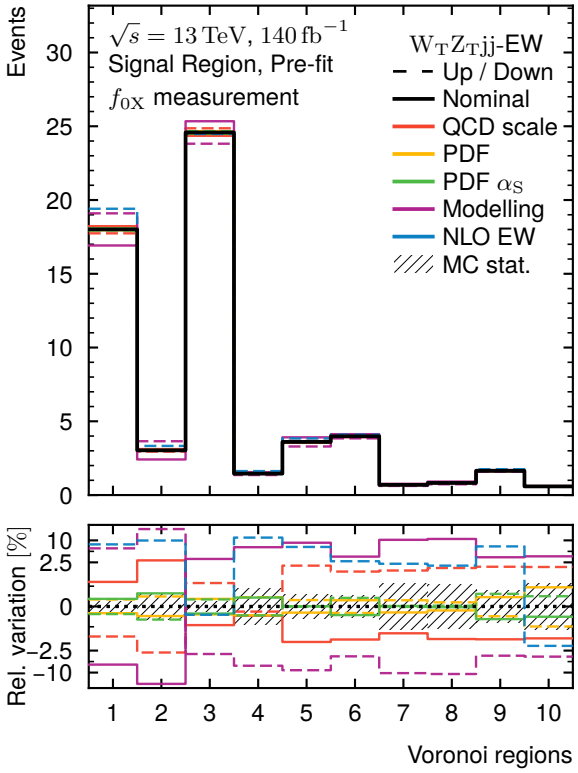


(c) $f_{\text{X}0}$ measurement

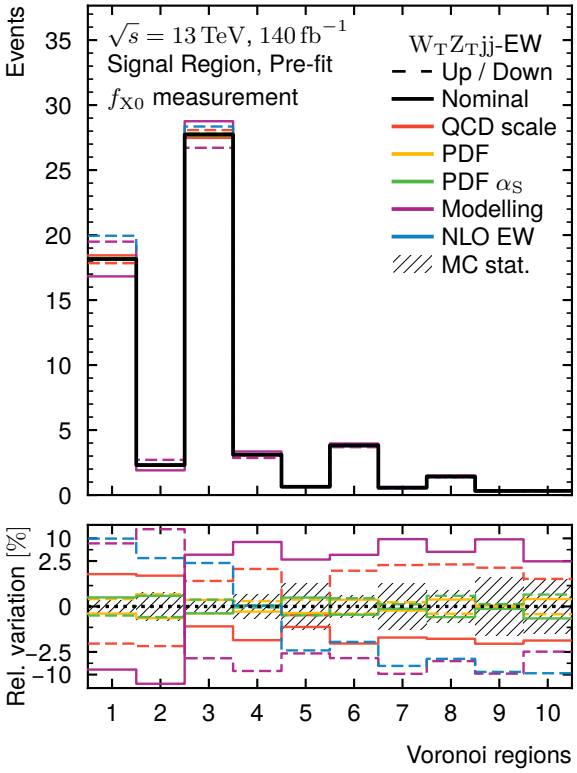
Figure P.6: Theoretical modelling systematic variations on the $\text{W}_\text{T}\text{Z}_0\text{jj}$-EW template, shown as relative shifts with respect to the nominal prediction as a function of the fit observable. The panels correspond to the f_{00}, $f_{0\text{X}}$, and $f_{\text{X}0}$ measurements in Run-3.



(a) f_{00} measurement

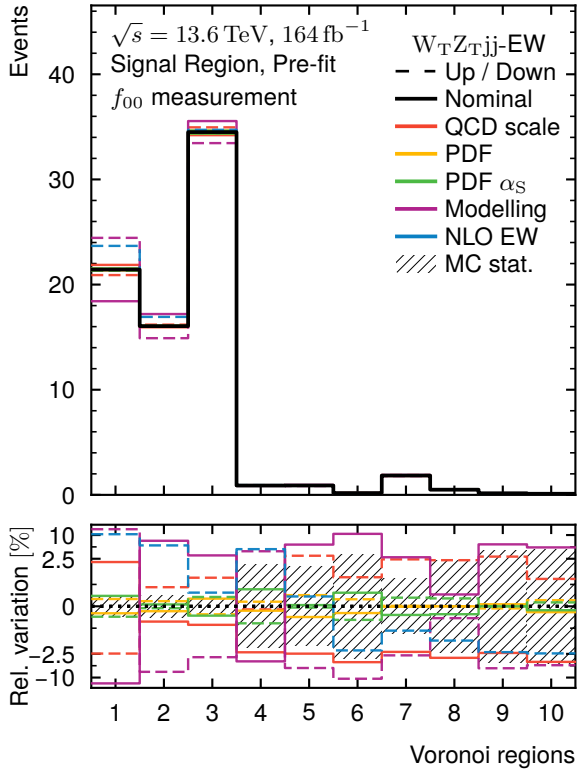


(b) $f_{0\text{X}}$ measurement

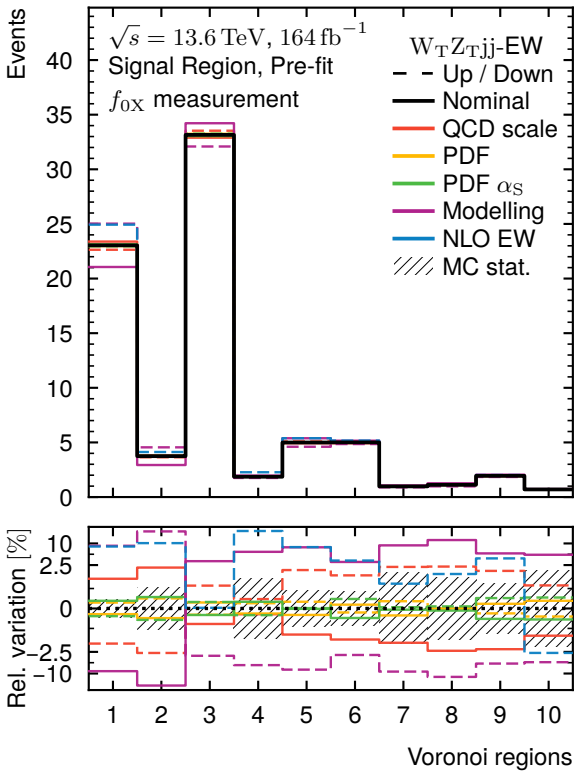


(c) $f_{\text{X}0}$ measurement

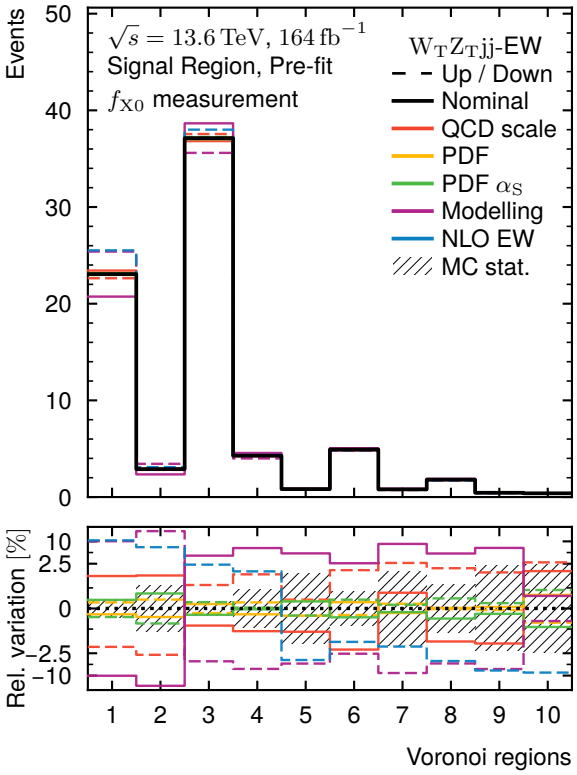
Figure P.7: Theoretical modelling systematic variations on the $\text{W}_\text{T}\text{Z}_\text{T}$jj-EW template, shown as relative shifts with respect to the nominal prediction as a function of the fit observable. The panels correspond to the f_{00}, $f_{0\text{X}}$, and $f_{\text{X}0}$ measurements in Run-2.



(a) f_{00} measurement

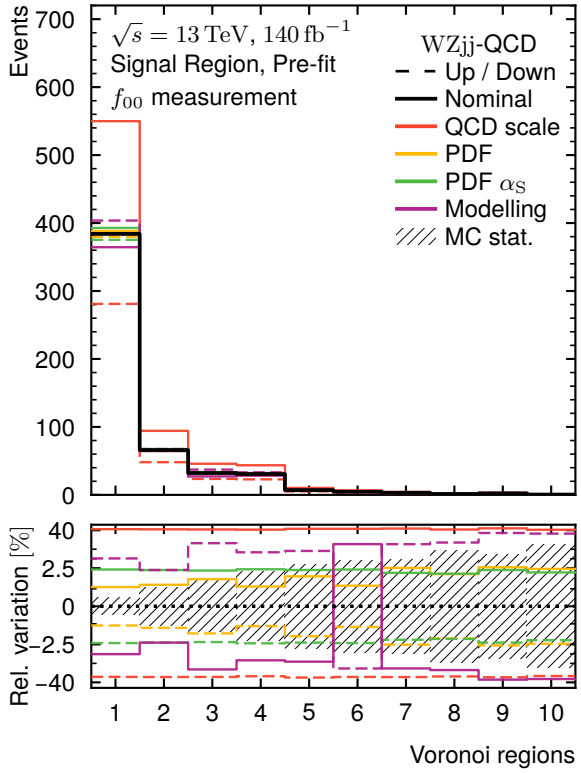


(b) $f_{0\text{X}}$ measurement

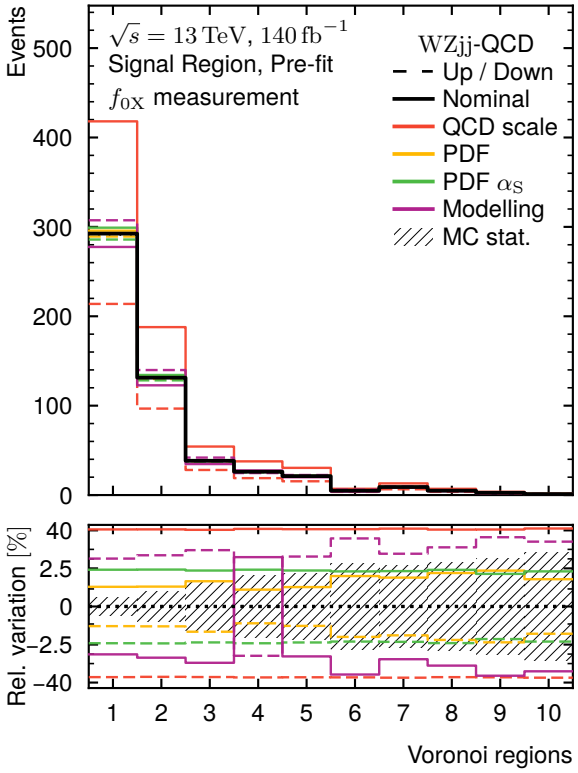


(c) $f_{\text{X}0}$ measurement

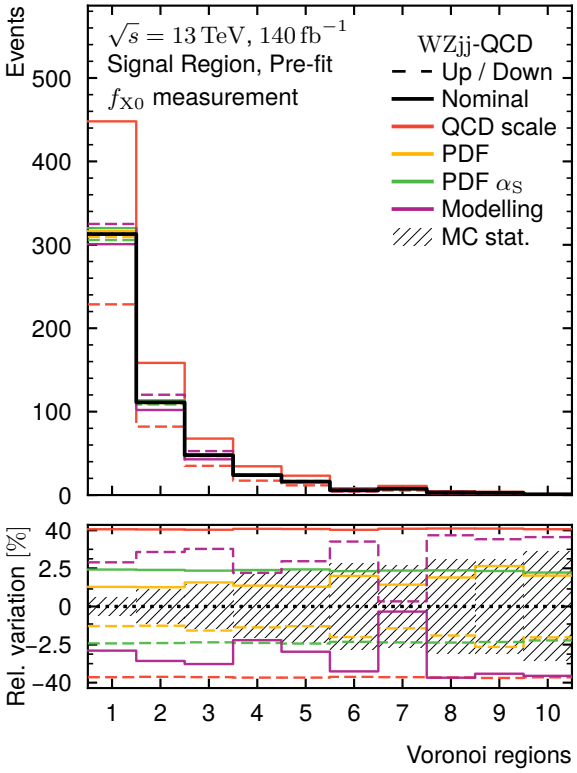
Figure P.8: Theoretical modelling systematic variations on the $\text{W}_\text{T}\text{Z}_\text{T}$jj-EW template, shown as relative shifts with respect to the nominal prediction as a function of the fit observable. The panels correspond to the f_{00}, $f_{0\text{X}}$, and $f_{\text{X}0}$ measurements in Run-3.



(a) f_{00} measurement

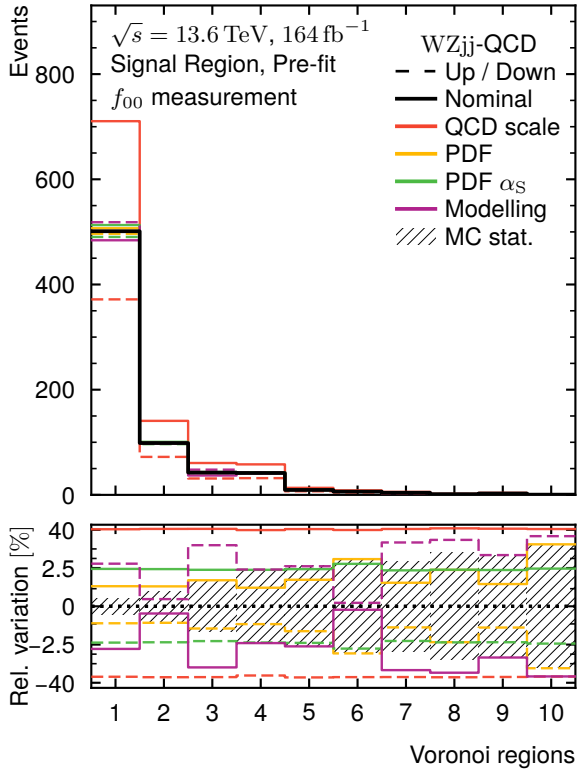


(b) f_{0X} measurement

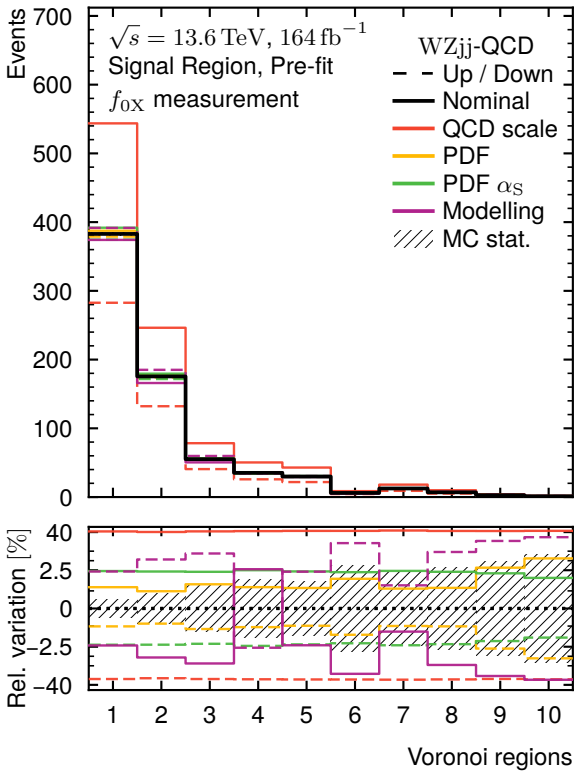


(c) f_{X0} measurement

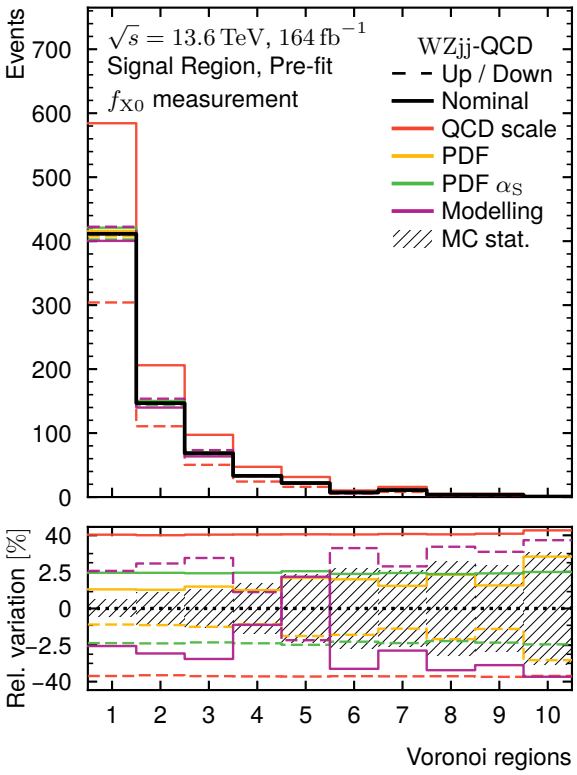
Figure P.9: Theoretical modelling systematic variations on the WZjj-QCD template, shown as relative shifts with respect to the nominal prediction as a function of the fit observable. The panels correspond to the f_{00}, f_{0X}, and f_{X0} measurements in Run-2.



(a) f_{00} measurement



(b) f_{0X} measurement



(c) f_{X0} measurement

Figure P.10: Theoretical modelling systematic variations on the WZjj-QCD template, shown as relative shifts with respect to the nominal prediction as a function of the fit observable. The panels correspond to the f_{00}, f_{0X}, and f_{X0} measurements in Run-3.

Q Pruning of the systematic uncertainties

The profile-likelihood fit includes a large number of NPs, most of which have a negligible effect in any individual region and sample. To keep the minimisation stable and to reduce its computational cost, such variations are pruned following the standard `TRExFitter` procedure. In a given region and for a given sample, an NP is dropped when both its normalisation and its shape effect fall below the configured pruning threshold of $0.2\,\%$. When only one of the two components is significant, the parameter is retained as a normalisation-only or as a shape-only variation, and variations whose shape is found to be statistically unreliable are flagged as *bad shape* and treated accordingly.

The outcome is summarised, separately for each data-taking period and for each polarisation configuration of the signal template, in Figures Q.1 to Q.6. In every plot the rows list the NPs and the columns the samples entering the SR, the ZZ-CR, and the b-CR. The colour of each cell encodes whether the variation is kept with both its shape and normalisation components, retained as normalisation-only or shape-only, dropped, or flagged as having a bad shape. The impact on the fit results is summarised in Table Q.A: the expected fractions, uncertainties, and significances obtained without any pruning are identical to those of Table 7.I(a) to within 0.001, so that the pruning only removes variations without effect on the fit.

Table Q.A: Expected polarisation fractions f_{00}, $f_{0\mathrm{X}}$, and $f_{\mathrm{X}0}$, together with the EW fiducial cross-section $\sigma^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}}$ and the signal strength $\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}}$, obtained from the individual Run-2 and Run-3 fits and from their combination. The quoted σ values are the corresponding significances. These expected results are obtained in the absence of pruning.

Parameter	Run	f_{XX}		$\sigma^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}}$ [fb]	$\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}}$
Measurement		Individual	Combined	Individual	Combined
f_{00}	2	$0.088^{+0.119}_{-0.088}$ 0.88σ	$0.087^{+0.078}_{-0.069}$ 1.29σ	$0.335^{+0.084}_{-0.079}$	$1.000^{+0.194}_{-0.185}$
	3	$0.087^{+0.109}_{-0.087}$ 0.96σ		$0.373^{+0.098}_{-0.093}$	
$f_{0\mathrm{X}}$	2	$0.280^{+0.177}_{-0.160}$ 1.79σ	$0.279^{+0.117}_{-0.109}$ 2.59σ	$0.335^{+0.085}_{-0.080}$	$1.000^{+0.199}_{-0.189}$
	3	$0.279^{+0.161}_{-0.146}$ 1.91σ		$0.373^{+0.103}_{-0.096}$	
$f_{\mathrm{X}0}$	2	$0.274^{+0.157}_{-0.139}$ 2.03σ	$0.274^{+0.104}_{-0.095}$ 2.91σ	$0.335^{+0.091}_{-0.086}$	$1.000^{+0.211}_{-0.201}$
	3	$0.273^{+0.147}_{-0.129}$ 2.15σ		$0.373^{+0.106}_{-0.100}$	

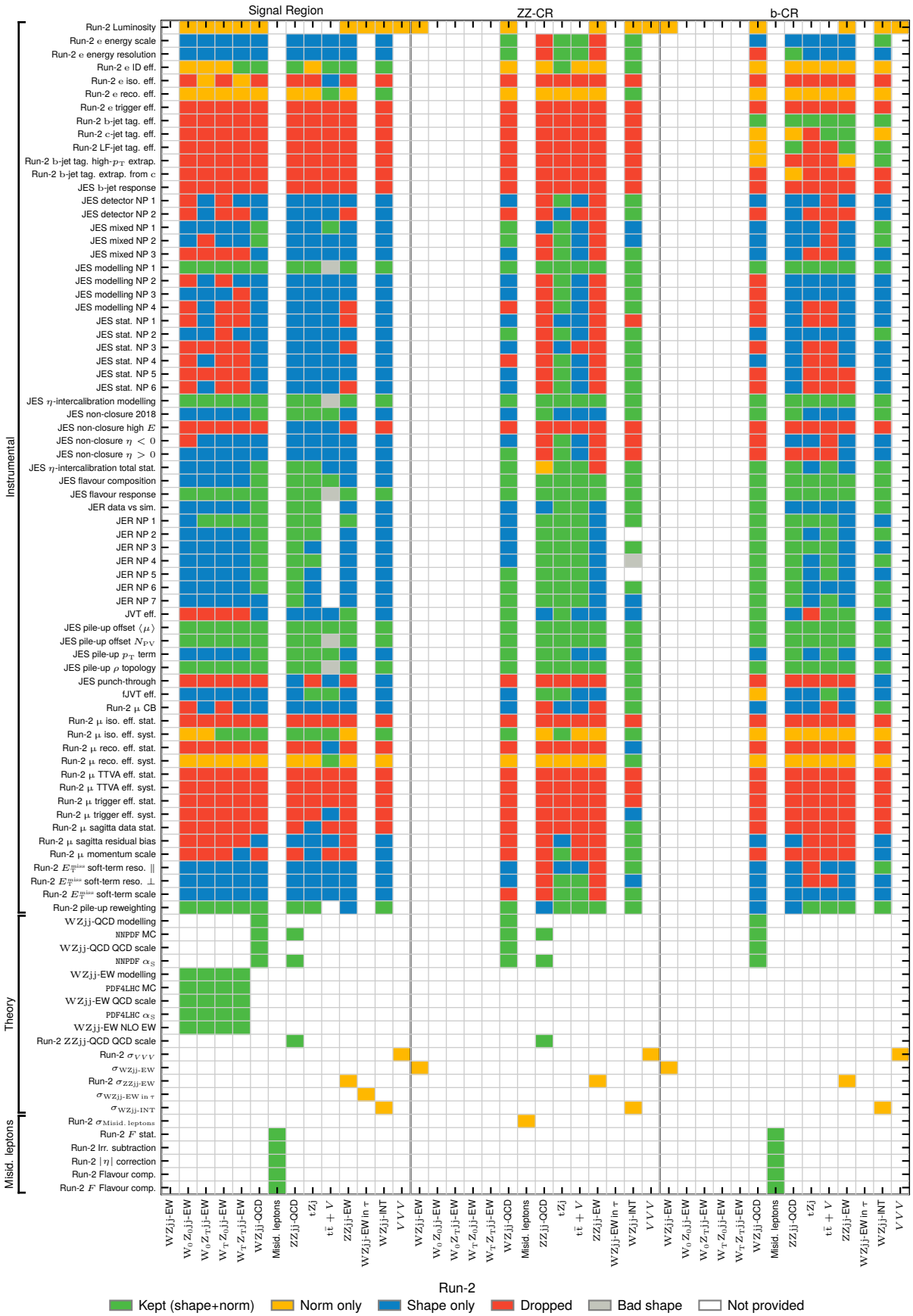


Figure Q.1: Pruning of the systematic uncertainties for the Run-2 f_{00} measurement. Each cell gives the treatment of an NP (rows) for a given sample in the SR, the ZZ-CR, and the b-CR (columns): kept with both shape and normalisation components, kept as normalisation-only or shape-only, dropped, or flagged as having a bad shape.

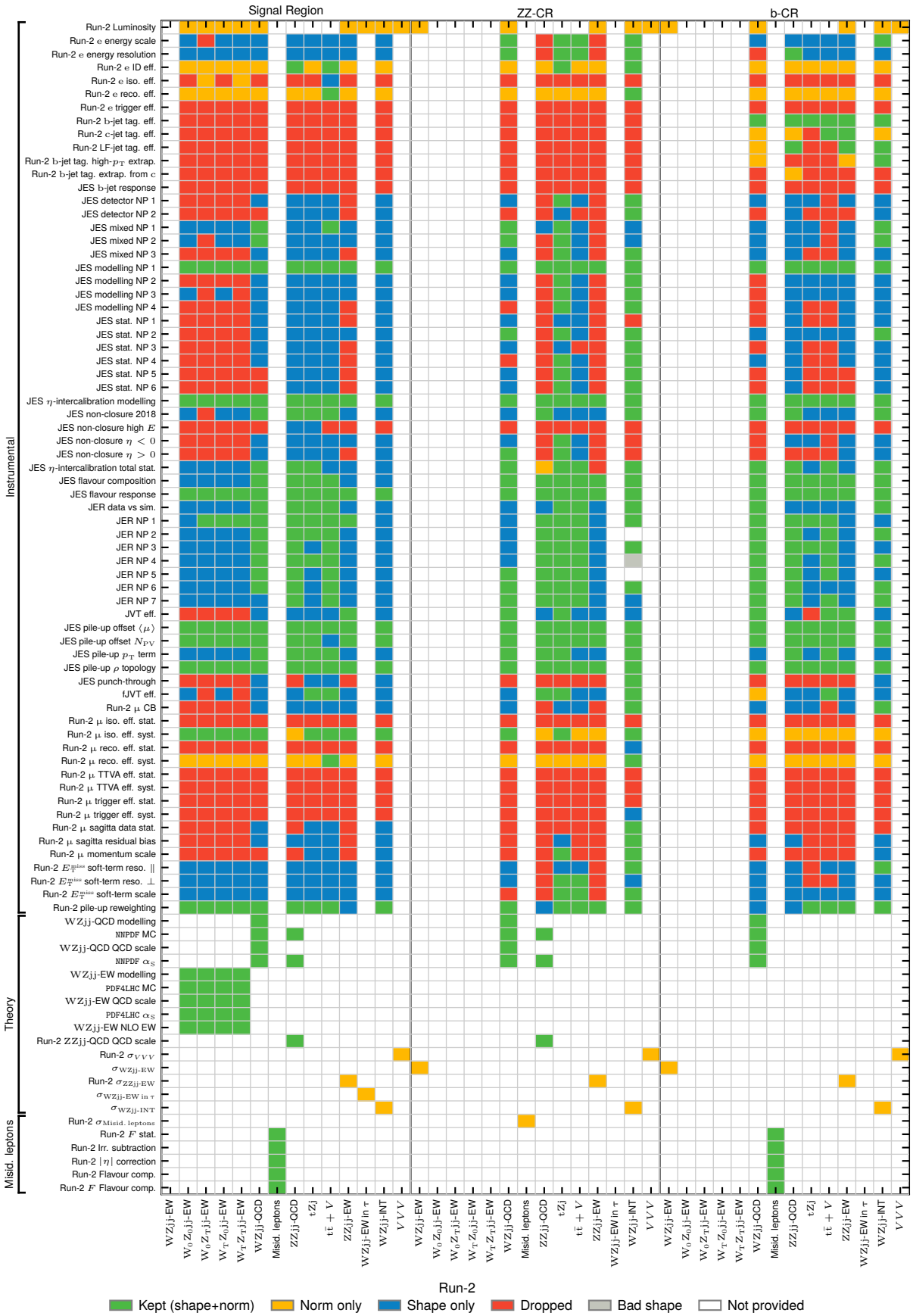


Figure Q.2: Pruning of the systematic uncertainties for the Run-2 $f_{0\mathrm{X}}$ measurement. Each cell gives the treatment of an NP (rows) for a given sample in the SR, the ZZ-CR, and the b-CR (columns): kept with both shape and normalisation components, kept as normalisation-only or shape-only, dropped, or flagged as having a bad shape.

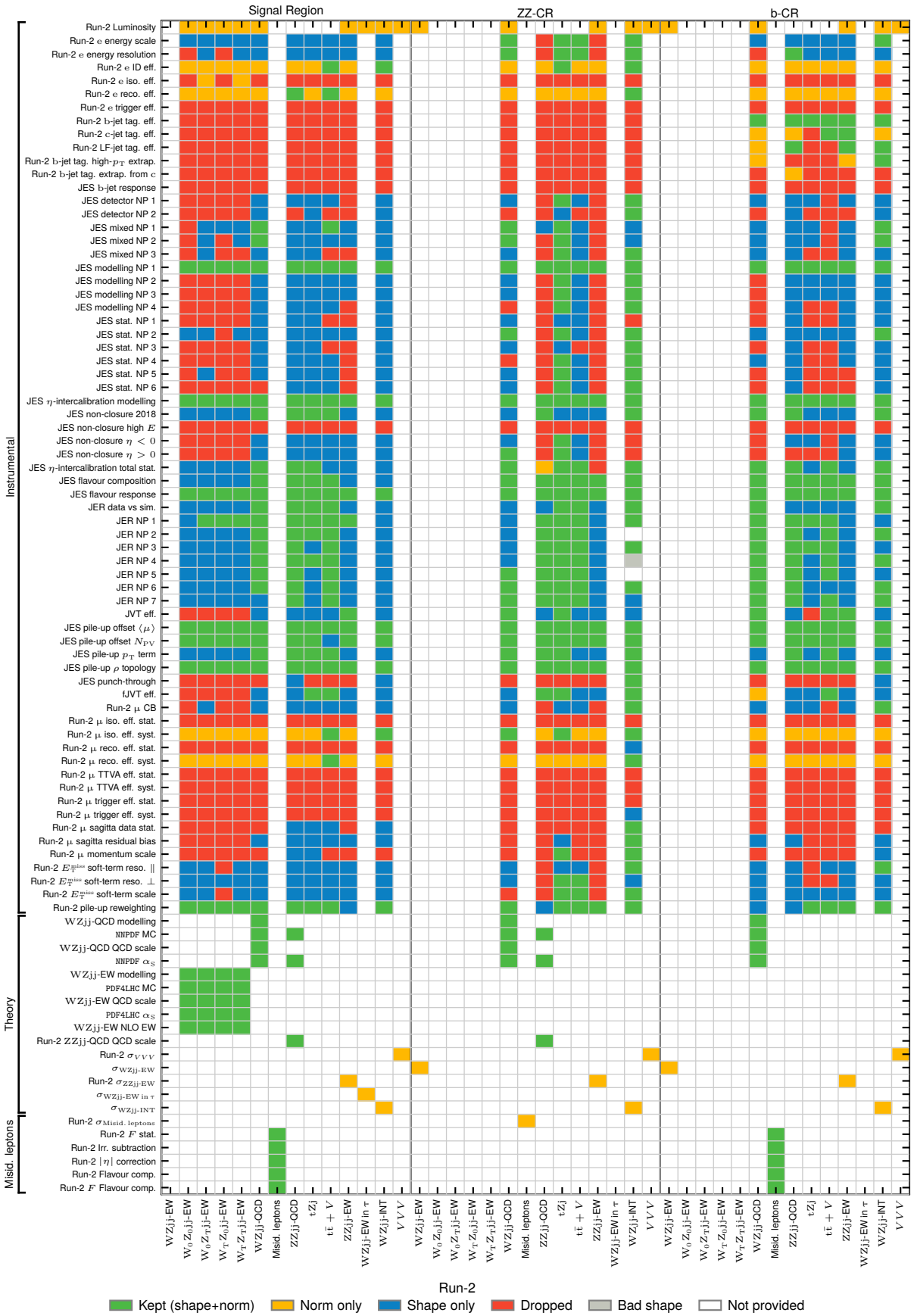


Figure Q.3: Pruning of the systematic uncertainties for the Run-2 f_{X0} measurement. Each cell gives the treatment of an NP (rows) for a given sample in the SR, the ZZ-CR, and the b-CR (columns): kept with both shape and normalisation components, kept as normalisation-only or shape-only, dropped, or flagged as having a bad shape.

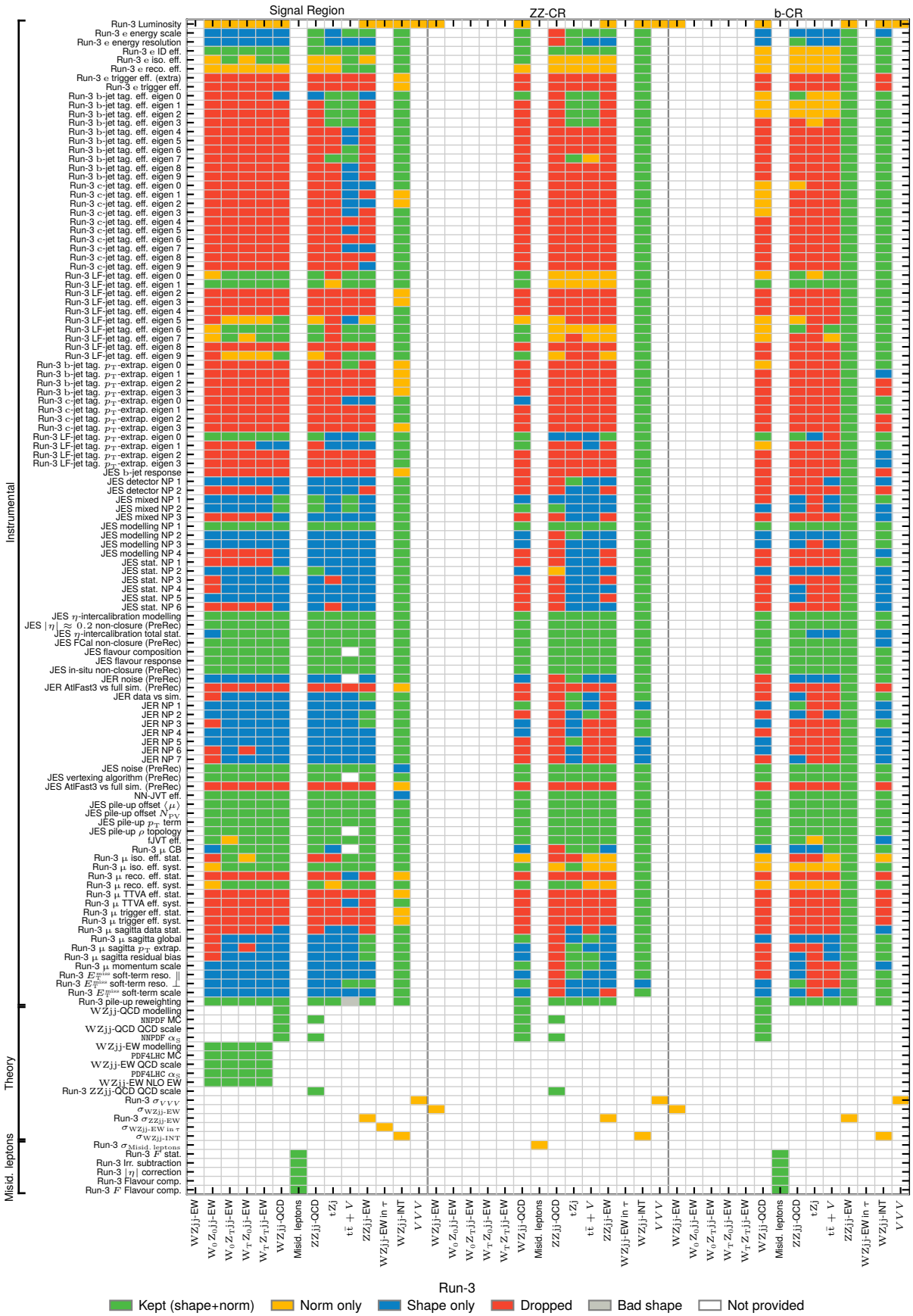


Figure Q.4: Pruning of the systematic uncertainties for the Run-3 f_{00} measurement. Each cell gives the treatment of an NP (rows) for a given sample in the SR, the ZZ-CR, and the b-CR (columns): kept with both shape and normalisation components, kept as normalisation-only or shape-only, dropped, or flagged as having a bad shape.

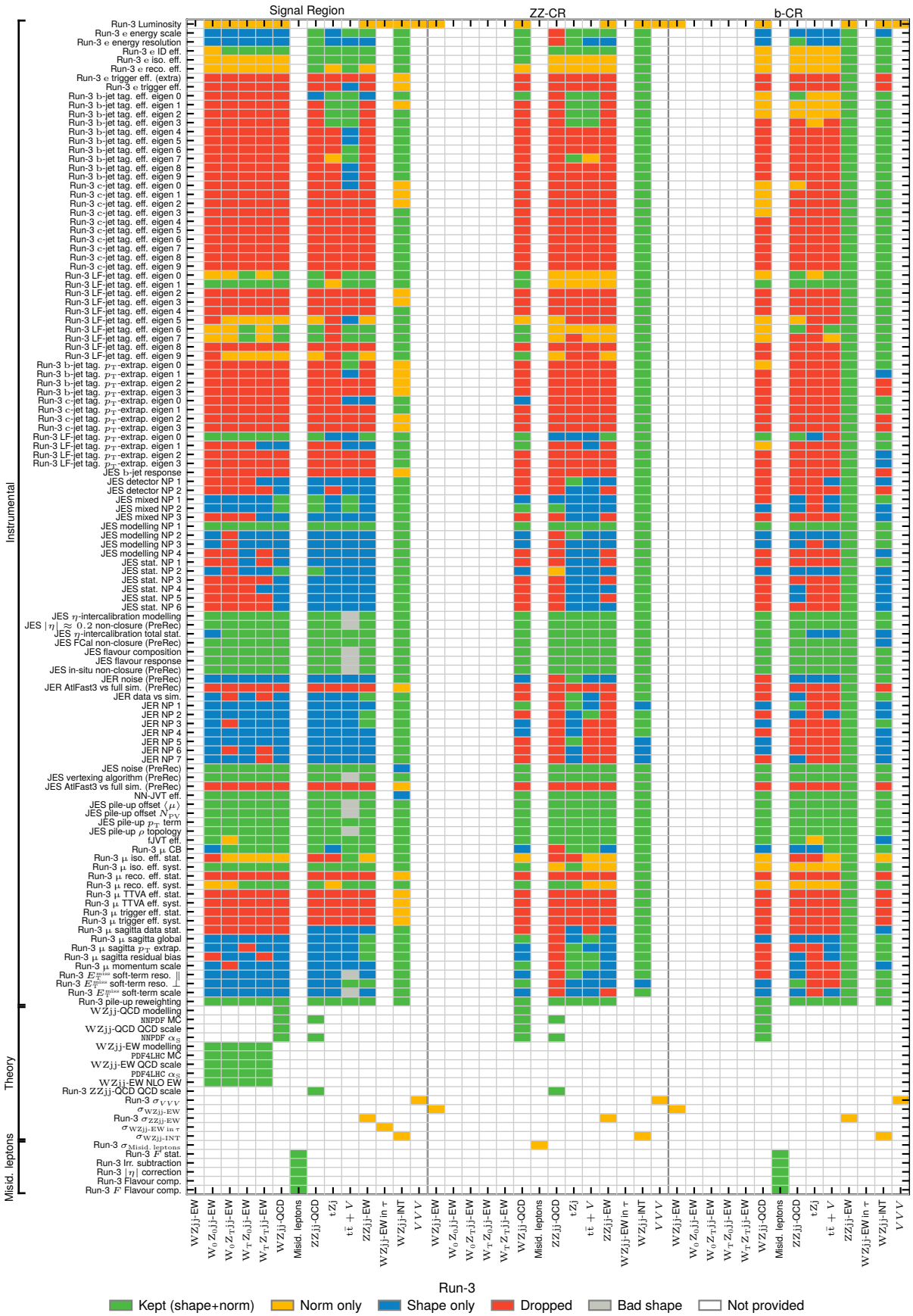


Figure Q.5: Pruning of the systematic uncertainties for the Run-3 $f_{0\mathrm{X}}$ measurement. Each cell gives the treatment of an NP (rows) for a given sample in the SR, the ZZ-CR, and the b-CR (columns): kept with both shape and normalisation components, kept as normalisation-only or shape-only, dropped, or flagged as having a bad shape.

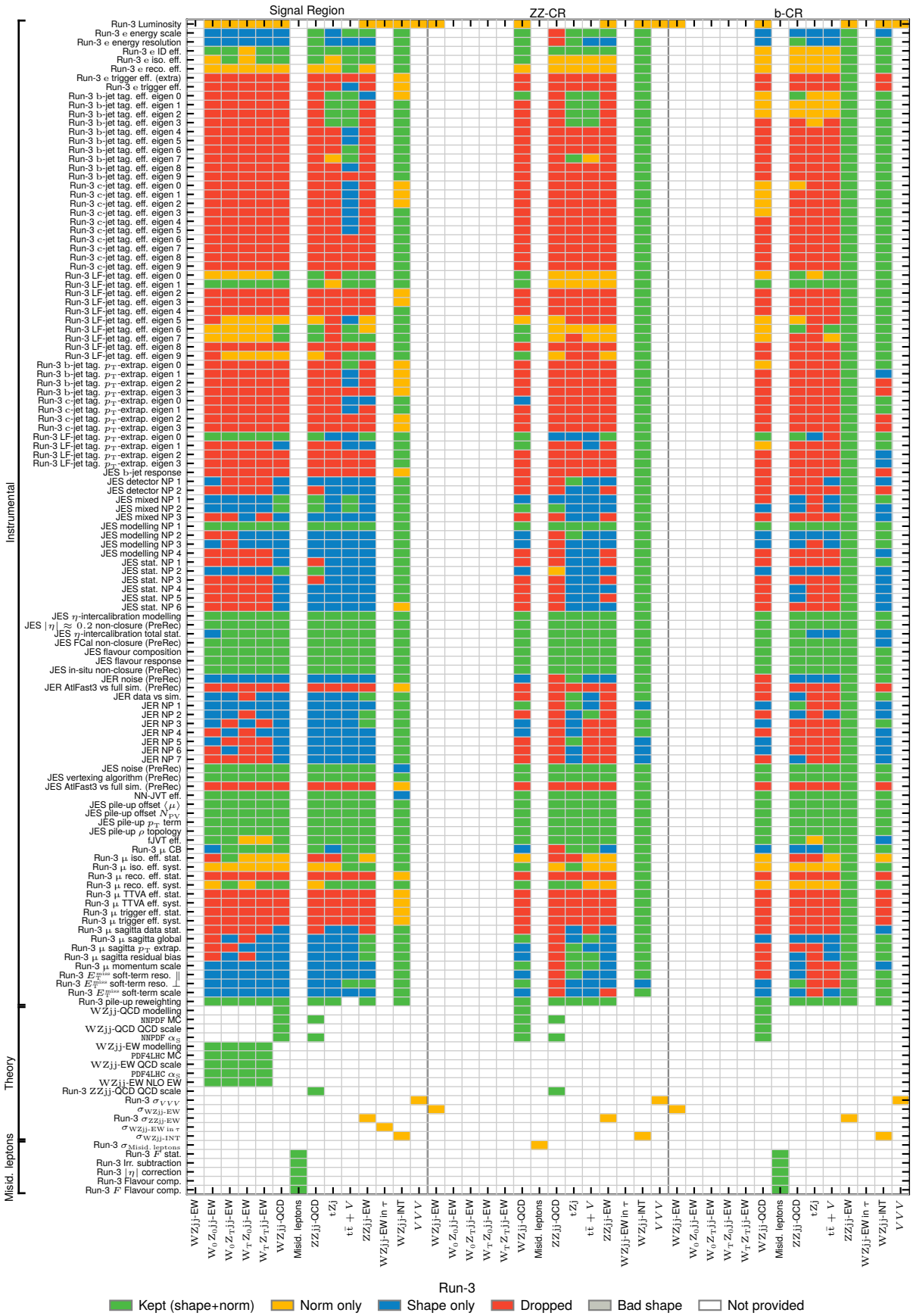


Figure Q.6: Pruning of the systematic uncertainties for the Run-3 f_{X0} measurement. Each cell gives the treatment of an NP (rows) for a given sample in the SR, the ZZ-CR, and the b-CR (columns): kept with both shape and normalisation components, kept as normalisation-only or shape-only, dropped, or flagged as having a bad shape.

R Post-fit nuisance-parameter pulls

The post-fit values of the NPs of the combined Run-2+Run-3 maximum-likelihood fit are summarised in this appendix. For each parameter, the *pull* is defined as:

$$\text{pull} = \frac{\hat{\theta} - \theta_0}{\Delta\theta}\,, \tag{R.1}$$

where $\hat{\theta}$ is the post-fit central value, θ_0 the pre-fit central value, and $\Delta\theta$ the pre-fit uncertainty of the parameter. The markers show the post-fit central values together with their post-fit uncertainties. The inner (green) band indicates the $\pm 1\sigma$ and the outer (yellow) band the $\pm 2\sigma$ interval of the pre-fit uncertainty. A marker displaced from zero signals a pull of the parameter, while an uncertainty smaller than unity indicates that the parameter has been constrained by the data. The parameters are grouped by origin following their descriptions in Section 7.2.

The fit is stable across the three configurations: the large majority of NPs are neither significantly pulled from their pre-fit values nor strongly constrained, and the small number of parameters that are mildly constrained is consistent between the three fits. The pulls for the three measurement configurations are shown in Figures R.1 to R.3.

The three configurations share the same, well-behaved pattern. Of the ~ 170 NPs the overwhelming majority are neither pulled (all central values within $\pm 1\sigma$ of their pre-fit value) nor constrained (post-fit uncertainty compatible with unity), so for these sources the data add essentially no information beyond the external calibrations. Constraints appear mainly for the small number of deliberately loosely defined theory uncertainties, the WZjj-QCD modelling and QCD-scale terms, whose large pre-fit ranges the fit legitimately narrows, and, mildly, for a few jet-related terms. The WZjj-QCD QCD-scale parameter is, in addition, pulled to lower values, absorbing the $13\,\%$ to $30\,\%$ over-prediction of this background in the absence of a dedicated NF. The residual sub-1σ pulls of the ZZjj-EW cross-section parameters of both runs and of the Run-3 pile-up reweighting parameter are statistically insignificant. Apart from that Run-3 pile-up reweighting term, no object-level parameter, whether jets, electrons, muons, $E_{\mathrm{T}}^{\mathrm{miss}}$, flavour tagging, pile-up, or luminosity, is pulled by more than half a standard deviation, and only a few jet terms are mildly constrained, confirming that the uncertainties delivered by the ATLAS Combined Performance groups are respected and that the sensitivity is not artificially inflated by an unphysical constraint of a detector systematic.

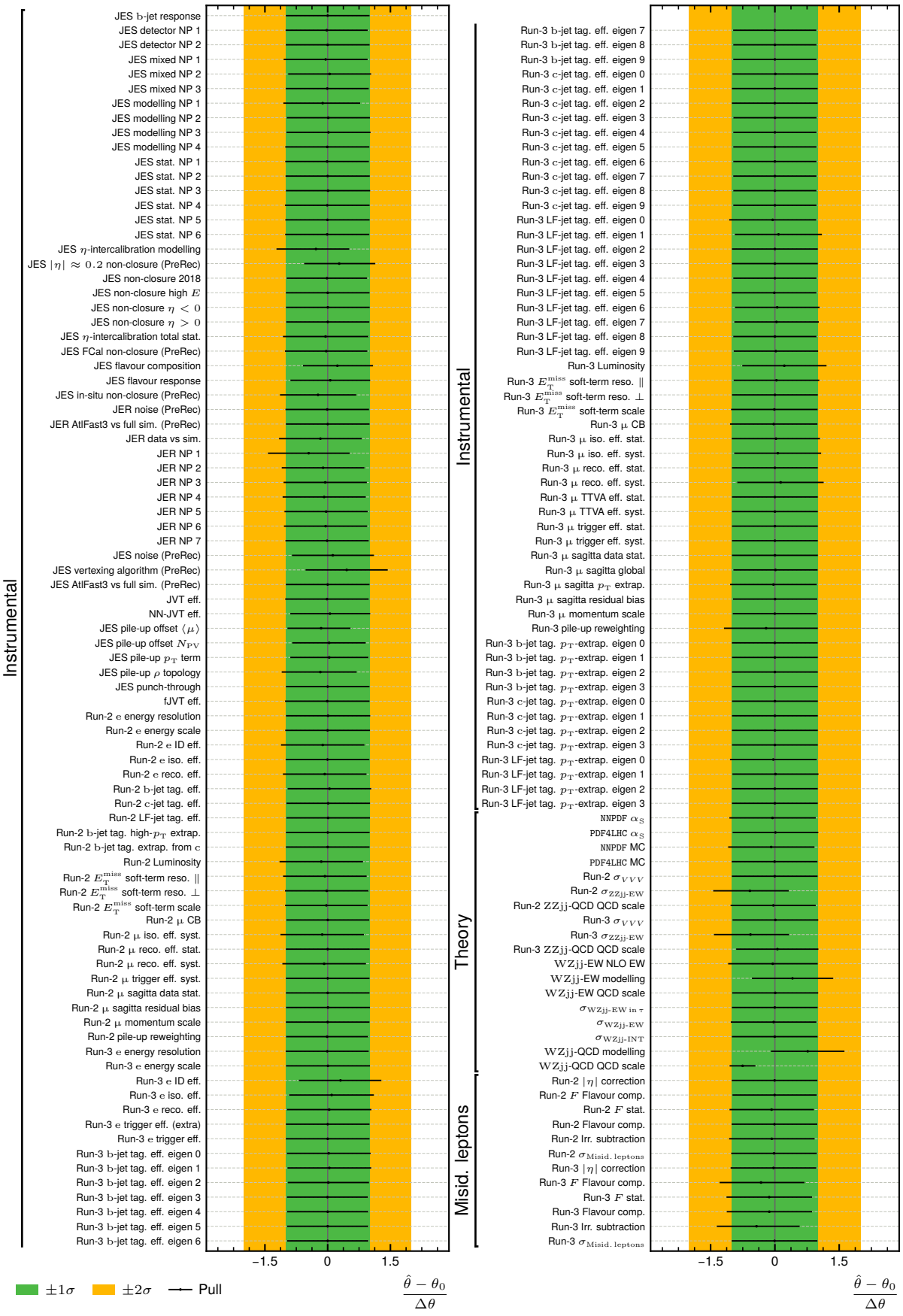


Figure R.1: Post-fit NP pulls of the combined Run-2+Run-3 fit for the f_{00} measurement. The green and yellow bands denote the $\pm1\sigma$ and $\pm2\sigma$ pre-fit intervals.

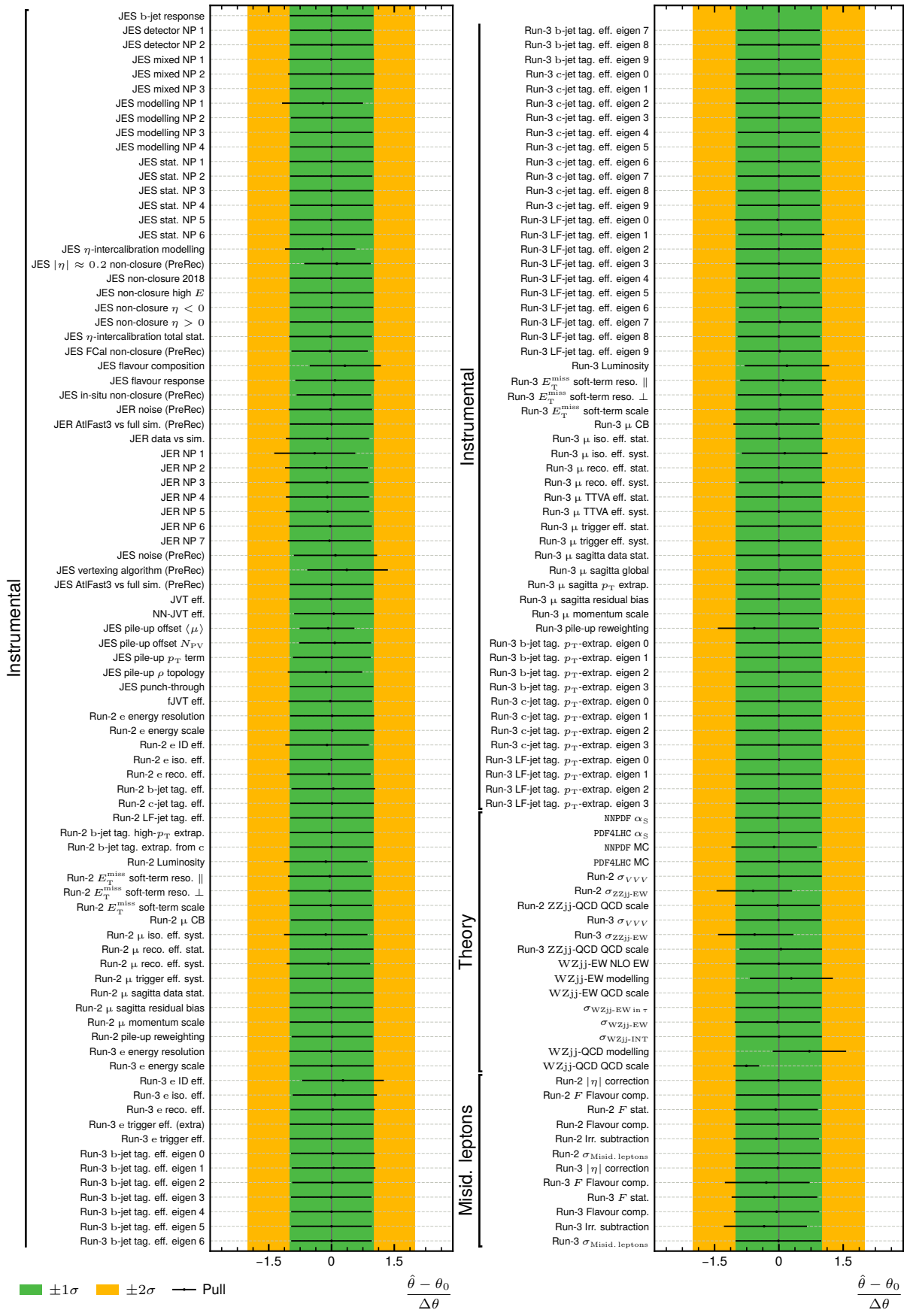


Figure R.2: Post-fit NP pulls of the combined Run-2+Run-3 fit for the $f_{0\mathrm{X}}$ measurement. The green and yellow bands denote the $\pm1\sigma$ and $\pm2\sigma$ pre-fit intervals.

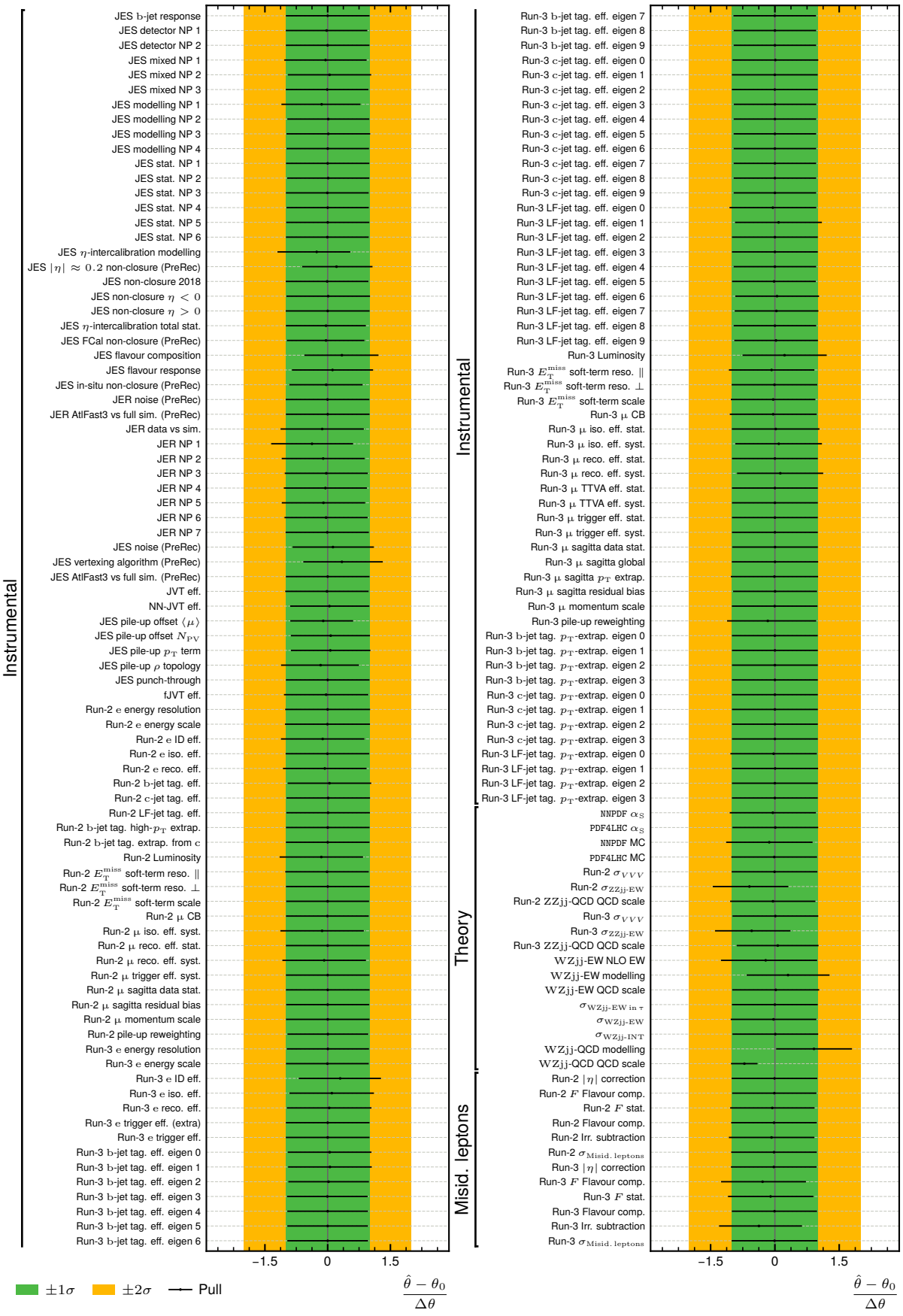


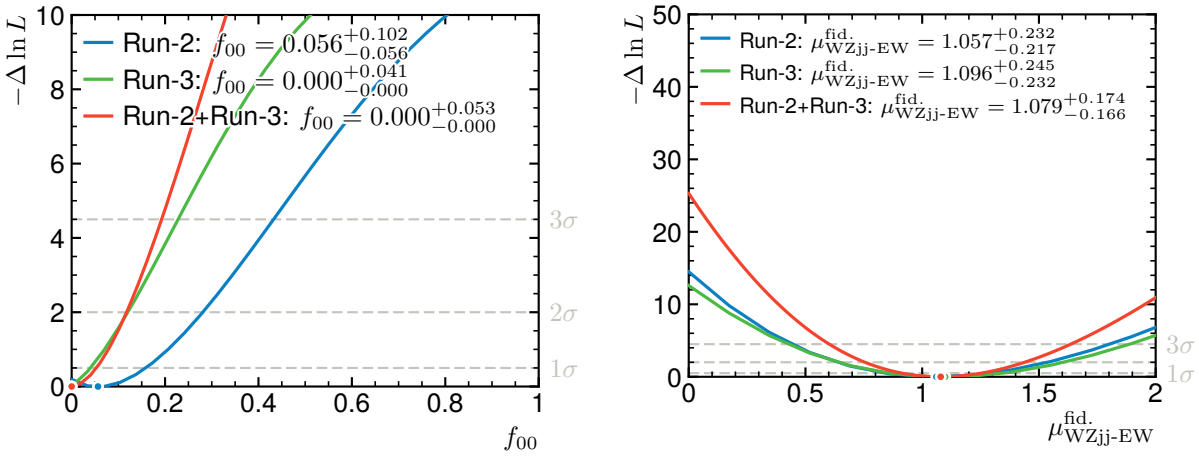
Figure R.3: Post-fit NP pulls of the combined Run-2+Run-3 fit for the f_{X0} measurement. The green and yellow bands denote the $\pm1\sigma$ and $\pm2\sigma$ pre-fit intervals.

S Profile-likelihood scans of the polarisation fits

The polarisation measurement is extracted from a set of maximum-likelihood fits in which one polarisation fraction is taken as the POI alongside the fiducial signal strength $\mu^{\text{fid.}}_{\text{WZjj-EW}}$ of EW WZjj production.

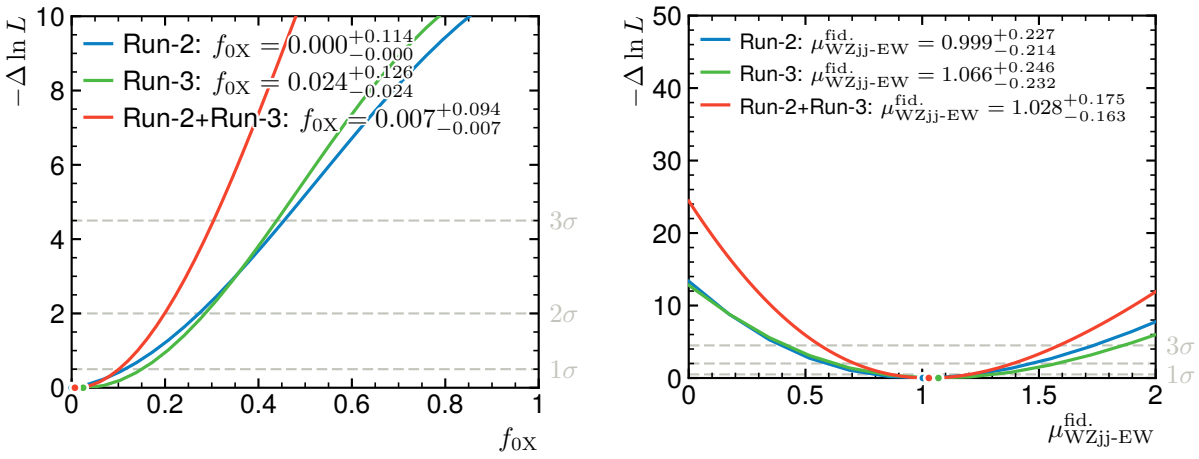
For each fit configuration, the 1D profile-likelihood scans of both POIs are collected in Figures S.1 to S.3. Each scan is performed separately on the Run-2 and Run-3 datasets and on their combination. The horizontal lines mark the 1σ, 2σ, and 3σ confidence intervals, corresponding to $-\Delta \ln L = 0.5,\ 2.0,\ 4.5$ for a single POI.

Three features recur across the measurements. First, the combined scan is in every case steeper than either single-run scan, and the resulting uncertainty follows the expected statistical scaling: for f_{X0} the per-run uncertainties of ≈ 0.13 combine to ≈ 0.09, close to the $\frac{0.13}{\sqrt{2}}$ expectation, so the combination is statistics-limited and free of tension. Second, the Run-2 and Run-3 minima are mutually compatible for all parameters of interest: for f_{X0} the two values, 0.411 and 0.270, differ by $\approx 0.8\sigma$, justifying the combined fit. Third, the $\mu^{\text{fid.}}_{\text{WZjj-EW}}$ scans are close to parabolic and minimise at $\mu^{\text{fid.}}_{\text{WZjj-EW}} \approx 1.0$–$1.1$ across all runs and measurements, consistent with the SM cross-section, while the f_{00} and $f_{0\text{X}}$ scans are one-sided, rising from the boundary $f = 0$. Only f_{X0} shows a genuine parabolic minimum away from the boundary, which is the origin of the 4.0σ evidence. The three $\mu^{\text{fid.}}_{\text{WZjj-EW}}$ determinations agree to within 0.06, well within their uncertainties, showing that the extracted signal strength is independent of which polarisation fraction is profiled.



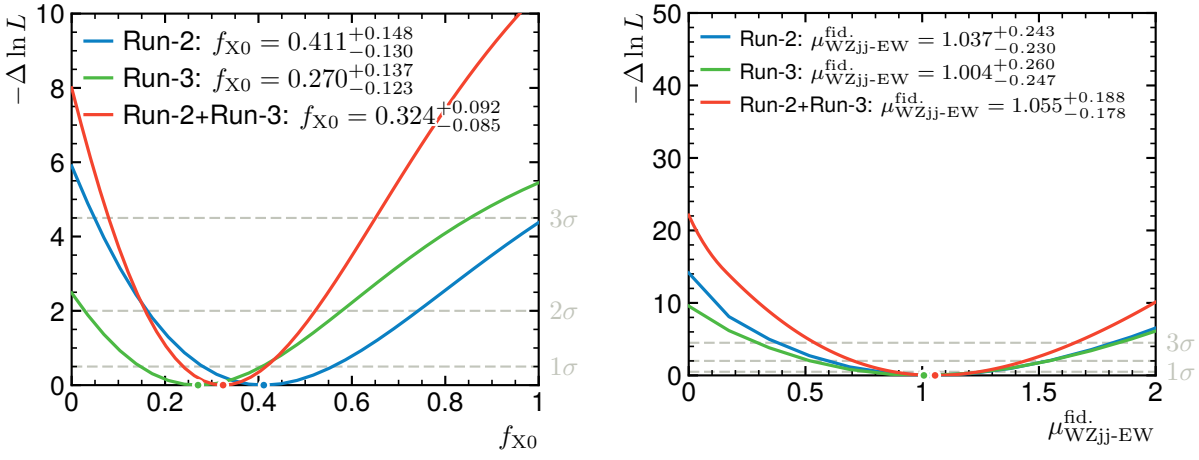
(a) f_{00} **(b)** $\mu^{\text{fid.}}_{\text{WZjj-EW}}$

Figure S.1: Profile-likelihood scans for the f_{00} measurement, for the polarisation fraction f_{00} (a) and the fiducial signal strength $\mu^{\text{fid.}}_{\text{WZjj-EW}}$ (b), shown for the Run-2, Run-3, and combined Run-2+Run-3 fits.



(a) f_{0X} **(b)** $\mu^{\text{fid.}}_{\text{WZjj-EW}}$

Figure S.2: Profile-likelihood scans for the f_{0X} measurement, for the polarisation fraction f_{0X} (a) and the fiducial signal strength $\mu^{\text{fid.}}_{\text{WZjj-EW}}$ (b), shown for the Run-2, Run-3, and combined Run-2+Run-3 fits.



(a) f_{X0} **(b)** $\mu^{\text{fid.}}_{\text{WZjj-EW}}$

Figure S.3: Profile-likelihood scans for the f_{X0} measurement, for the polarisation fraction f_{X0} (a) and the fiducial signal strength $\mu^{\text{fid.}}_{\text{WZjj-EW}}$ (b), shown for the Run-2, Run-3, and combined Run-2+Run-3 fits.

T Injection tests

Injection tests have been carried out to demonstrate the ability of the statistical treatment to accommodate deviations from the assumptions underlying the nominal prediction. Two series of tests are presented in this appendix: the first characterises the recovery of the polarisation composition, and the second assesses the impact of the signal component containing τ leptons, which is treated as a background.

T.1 Closure of the polarisation-fraction measurement

The closure of the polarisation-fraction measurement is assessed using pseudo-datasets (representing Run-3 data) fitted with the nominal model foreseen for real data. These pseudo-datasets deviate from the nominal model in several respects:

- the MadGraph samples are used for the WZjj-EW signal instead of the Sherpa ones
- in the b-CR, the sum of the "polarised" samples and the "only-τ" WZjj-EW sample replaces the full WZjj-EW sample used in the nominal model
- the MC-based prediction for fakes is used instead of the matrix-method estimate.

T.1.1 Nominal fractions

With these modifications in place, the fits for f_{X0} and f_{00} are carried out. Their results are given in Table T.A. Only a subset of the POIs, NFs, and NPs is shown.

Table T.A: Closure fits for the nominal injected fractions $f_{\mathrm{X0}} = 0.27$, $f_{00} = 0.09$.

(a) f_{00} measurement

Fit parameter	Best-fit value
f_{00}	$0.08^{+0.10}_{-0.08}$
$\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}}$	$1.00^{+0.49}_{-0.19}$
$\mu_{\mathrm{ZZjj\text{-}QCD}}$	$1.01^{+0.16}_{-0.11}$
μ_{tZj}	$1.13^{+0.29}_{-0.29}$
$\mu_{\mathrm{t\bar{t}}+V}$	$0.89^{+0.11}_{-0.11}$

(b) f_{X0} measurement

Fit parameter	Best-fit value
f_{X0}	$0.29^{+0.13}_{-0.13}$
$\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}}$	$1.01^{+0.22}_{-0.22}$
$\mu_{\mathrm{ZZjj\text{-}QCD}}$	$1.01^{+0.11}_{-0.11}$
μ_{tZj}	$1.12^{+0.29}_{-0.29}$
$\mu_{\mathrm{t\bar{t}}+V}$	$0.89^{+0.12}_{-0.12}$

The injected fractions are recovered well within the uncertainties.

T.1.2 Varied fractions

On top of the modifications above, a second test additionally alters the composition of the polarisation states, so that the injected values become $f_{\mathrm{X0}} = 0.365$ and $f_{00} = 0.18$. The results are given in Table T.B.

The fit accommodates the substantially varied fractions and recovers them well within the uncertainties. Note that the best-fit values are not exactly the injected ones, since other aspects of the pseudo-dataset

Table T.B: Closure fits for the varied injected fractions $f_{\mathrm{X0}} = 0.365$, $f_{00} = 0.18$.

(a) f_{00} measurement

Fit parameter	Best-fit value
f_{00}	$0.17^{+0.11}_{-0.10}$
$\mu^{\text{fid.}}_{\mathrm{WZjj\text{-}EW}}$	$1.00^{+0.48}_{-0.19}$
$\mu_{\mathrm{ZZjj\text{-}QCD}}$	$1.01^{+0.16}_{-0.11}$
μ_{tZj}	$1.13^{+0.29}_{-0.29}$
$\mu_{\mathrm{t\bar{t}}+V}$	$0.89^{+0.11}_{-0.11}$
$\lvert\eta\rvert$ correction	$-0.01^{+1.00}_{-1.00}$
F Flavour comp.	$-0.08^{+0.99}_{-0.99}$
F stat.	$-0.07^{+0.99}_{-0.99}$
Flavour comp.	$-0.01^{+1.00}_{-1.00}$
Irr. subtraction	$-0.02^{+1.00}_{-1.00}$
$\sigma_{\text{Misid. leptons}}$	$0.00^{+1.00}_{-1.00}$

(b) f_{X0} measurement

Fit parameter	Best-fit value
f_{X0}	$0.39^{+0.15}_{-0.14}$
$\mu^{\text{fid.}}_{\mathrm{WZjj\text{-}EW}}$	$1.01^{+0.52}_{-0.21}$
$\mu_{\mathrm{ZZjj\text{-}QCD}}$	$1.01^{+0.16}_{-0.11}$
μ_{tZj}	$1.12^{+0.29}_{-0.33}$
$\mu_{\mathrm{t\bar{t}}+V}$	$0.89^{+0.12}_{-0.12}$

are varied simultaneously, in addition to the WZjj-EW polarisation fractions. In particular, the small pulls on the misidentified-lepton-related NPs reflect the use of the MC-based estimate corrected with scale factors in the pseudo-data, whereas the matrix-method estimate is used in the fit model.

T.1.3 Transverse–transverse component increased

In a further test, the $\mathrm{W_TZ_Tjj}$-EW component is scaled up by a factor 1.5, giving $\mu^{\text{fid.}}_{\mathrm{WZjj\text{-}EW}} = 1.27$ and $f_{00} = 0.07$. The resulting best-fit values are reported in Table T.C.

Table T.C: Fit of an Asimov pseudo-dataset in which the $\mathrm{W_TZ_Tjj}$-EW component is scaled up by a factor 1.5.

Fit parameter	Best-fit value
f_{00}	$0.04^{+0.08}_{-0.04}$
$\mu^{\text{fid.}}_{\mathrm{WZjj\text{-}EW}}$	$1.28^{+0.58}_{-0.21}$
$\mu_{\mathrm{ZZjj\text{-}QCD}}$	$1.01^{+0.16}_{-0.11}$
μ_{tZj}	$1.14^{+0.29}_{-0.29}$
$\mu_{\mathrm{t\bar{t}}+V}$	$0.89^{+0.11}_{-0.11}$
$\lvert\eta\rvert$ correction	$-0.01^{+1.00}_{-1.00}$
F Flavour comp.	$-0.08^{+0.99}_{-0.99}$
F stat.	$-0.08^{+0.99}_{-0.99}$
Flavour comp.	$-0.01^{+1.00}_{-1.00}$
Irr. subtraction	$-0.03^{+1.00}_{-1.00}$
$\sigma_{\text{Misid. leptons}}$	$0.00^{+1.00}_{-1.00}$
WZjj-QCD QCD scale	$0.04^{+0.49}_{-0.31}$

Again, the fit strategy captures the deviations of the pseudo-data from the nominal fit model.

T.1.4 Reduction of WZjj-QCD contribution

In an additional test, the WZjj-QCD background is scaled down by a factor 0.75, which correspondingly affects the pull of the relevant NP, giving a best-fit value of $-0.41^{+0.44}_{-0.22}$.

T.2 Effect of the component with τ leptons

To probe the effect of a potential mis-modelling of the WZjj-EW component with τ leptons, pseudo-data have been built with twice as many such events, while the fit model retained the nominal yields and uncertainties (that is, $10\,\%$ of the original yields).

In this situation, where the injected mis-modelling lies well beyond the assigned uncertainties, the effect on the combination of Run-2 and Run-3 pseudo-data is a shift of less than $3\,\%$ in the $\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}}$ signal strength that scales the WZjj-EW yield (to be compared with its total relative uncertainty of $^{+22}_{-16}\,\%$), and a change in f_{X0} of 0.003 (a relative variation below $1\,\%$, to be compared with the $\sim 35\,\%$ relative uncertainty on this fraction). This demonstrates the immunity of the polarisation-fraction measurements to a mis-modelling of the τ component.

U Decorrelation of the QCD scale uncertainties for WZjj-QCD

In this appendix, we present a selection of the results obtained for the combination of Run-2 and Run-3 when decorrelating the QCD scale uncertainties for WZjj-QCD between the two runs. The motivation for this decorrelation is to allow the fit to better accommodate the data in the SR which is over-predicted by the MC samples for both runs. This over-prediction is observed in both Run-2 and Run-3 and is compensated by a pull of the WZjj-QCD normalisation in the fit. By decorrelating the QCD scale uncertainties, we allow the fit to have more freedom to adjust the WZjj-QCD normalisation independently for each run.

The ranking is similar to that of Section 7.3.7: the QCD scale uncertainties for WZjj-QCD are among the most impactful NPs for both the f_{XX} and cross-section POIs, and they are pulled in the fit. This is expected, as the fit needs to adjust the WZjj-QCD normalisation to better fit the data in the SR. The Run-2 term is the more strongly pulled (-0.9σ against -0.2σ), as the over-prediction of the MC samples is more pronounced in Run-2 than in Run-3. On $\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}}$ the two decorrelated terms rank just below the correlated one, each with a smaller impact, while on f_{X0} the Run-3 term becomes the top-ranked parameter with an impact (0.007) larger than that of the correlated term (0.005).

The results of the fit for the combination of Run-2 and Run-3 are summarised in Table U.A. The impact of decorrelating the QCD scale uncertainties for WZjj-QCD on the final results is almost negligible, with a slight increase in the central value of the f_{X0} POI and a slight decrease in the central value of the f_{0X} POI, while the f_{00} POI is almost unchanged. The uncertainties and significances are also almost unchanged. The two Sherpa versions used for the WZjj-QCD samples in Run-2 and Run-3 are different, with Sherpa 2.2.12 used for Run-2 and Sherpa 2.2.16 used for Run-3. However, they are equivalent from a theoretical point of view as they are both based on the same ME and parton shower algorithms, and they only differ by some bug fixes and improvements in the implementation. Also, the choice of renormalisation and factorisation scales is the same for both versions, and the missing higher-order QCD effects are expected to be similar. Therefore, the correlation of the QCD scale uncertainties for WZjj-QCD between Run-2 and Run-3 is justified.

Table U.A: Fit results and significance for each polarisation-state measurement for Run-2 and Run-3 and their combination, obtained when the QCD scale uncertainties for WZjj-QCD are decorrelated between Run-2 and Run-3.

Parameter / **Measurement**	**Run**	f_{XX} **Individual**		f_{XX} **Combined**		$\sigma^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}}$ [fb] **Individual**	$\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}}$ **Combined**
f_{00}	2	$0.056^{+0.102}_{-0.056}$	0.63σ	$0.000^{+0.057}_{-0.000}$	0.00σ	$0.354^{+0.078}_{-0.073}$	$1.064^{+0.174}_{-0.166}$
	3	$0.000^{+0.041}_{-0.000}$	0.00σ			$0.409^{+0.091}_{-0.086}$	
f_{0X}	2	$0.000^{+0.114}_{-0.000}$	0.00σ	$0.000^{+0.092}_{-0.000}$	0.00σ	$0.334^{+0.076}_{-0.072}$	$1.020^{+0.173}_{-0.162}$
	3	$0.024^{+0.126}_{-0.024}$	0.20σ			$0.397^{+0.092}_{-0.086}$	
f_{X0}	2	$0.411^{+0.148}_{-0.130}$	3.44σ	$0.342^{+0.098}_{-0.090}$	4.04σ	$0.347^{+0.082}_{-0.077}$	$1.028^{+0.186}_{-0.178}$
	3	$0.270^{+0.137}_{-0.123}$	2.23σ			$0.374^{+0.097}_{-0.092}$	

V Change of fit parametrisation for normalisation factors

In this appendix, we present results obtained for the combination of Run-2 and Run-3 when changing the fit parametrisation from the nominal one described in Section 7.1 to a new one. This new parametrisation is based on the same principles as the nominal one, but with a different choice of NFs. Instead of having a $\mu^{\text{fid.}}_{\text{WZjj-EW}}$ and a polarisation fraction f_{XX} as POIs, we define two NFs: one μ for the targeted polarisation, called μ_{XX}, and another for the non-targeted polarisation, called $\mu_{\overline{\text{XX}}}$.

The results obtained with this fit parametrisation are compatible with the nominal ones, as shown in Table V.A(a) where the fit results and the significance for each polarisation-state measurement for Run-2 and Run-3 and their combination are presented when using μ as a POI. In Table V.A(b), the same results are presented but converted back to a cross-section.

The event yields are exactly the same as the ones obtained with the nominal fit parametrisation, and are given in Table 7.E.

Table V.A: Fit results and significance for each polarisation-state measurement for Run-2 and Run-3 and their combination.

(a) Results expressed as signal strengths μ

Parameter Measurement	Run	Individual			Combined		
		μ_{XX}	$\mu_{\overline{\mathrm{XX}}}$	Z	μ_{XX}	$\mu_{\overline{\mathrm{XX}}}$	Z
μ_{00}	2	$0.673^{+1.245}_{-0.673}$	$1.094^{+0.264}_{-0.248}$	0.63σ	$0.000^{+0.659}_{-0.000}$	$1.183^{+0.191}_{-0.185}$	0.00σ
	3	$0.000^{+0.511}_{-0.000}$	$1.201^{+0.268}_{-0.254}$	0.00σ			
$\mu_{0\mathrm{X}}$	2	$0.000^{+0.418}_{-0.000}$	$1.387^{+0.316}_{-0.302}$	0.00σ	$0.026^{+0.360}_{-0.026}$	$1.418^{+0.238}_{-0.233}$	0.01σ
	3	$0.093^{+0.513}_{-0.093}$	$1.442^{+0.338}_{-0.321}$	0.20σ			
μ_{X0}	2	$1.553^{+0.589}_{-0.527}$	$0.842^{+0.312}_{-0.297}$	3.44σ	$1.246^{+0.392}_{-0.358}$	$0.983^{+0.232}_{-0.223}$	4.00σ
	3	$0.992^{+0.551}_{-0.485}$	$1.009^{+0.328}_{-0.315}$	2.23σ			

(b) Results converted from μ back to a cross-section. The combined-fit signal strengths are common to the two runs, so they are converted with each run's own prediction and carry the $\mathrm{comb.}$ superscript of Chapter 7.

Parameter Measurement	Run	Individual			Combined		
		$\sigma^{\mathrm{fid.}}_{\mathrm{XX}}$ [fb]	$\sigma^{\mathrm{fid.}}_{\overline{\mathrm{XX}}}$ [fb]	Z	$\sigma^{\mathrm{fid.,comb.}}_{\mathrm{XX}}$ [fb]	$\sigma^{\mathrm{fid.,comb.}}_{\overline{\mathrm{XX}}}$ [fb]	Z
$\sigma^{\mathrm{fid.}}_{00}$	2	$0.020^{+0.036}_{-0.020}$	$0.334^{+0.081}_{-0.076}$	0.63σ	$0.000^{+0.019}_{-0.000}$	$0.361^{+0.058}_{-0.056}$	0.00σ
	3	$0.000^{+0.017}_{-0.000}$	$0.409^{+0.091}_{-0.086}$	0.00σ	$0.000^{+0.021}_{-0.000}$	$0.403^{+0.065}_{-0.063}$	
$\sigma^{\mathrm{fid.}}_{0\mathrm{X}}$	2	$0.000^{+0.039}_{-0.000}$	$0.334^{+0.076}_{-0.073}$	0.00σ	$0.002^{+0.034}_{-0.002}$	$0.342^{+0.057}_{-0.056}$	0.01σ
	3	$0.010^{+0.053}_{-0.010}$	$0.388^{+0.091}_{-0.086}$	0.20σ	$0.003^{+0.037}_{-0.003}$	$0.381^{+0.064}_{-0.063}$	
$\sigma^{\mathrm{fid.}}_{\mathrm{X0}}$	2	$0.143^{+0.054}_{-0.048}$	$0.204^{+0.076}_{-0.072}$	3.44σ	$0.114^{+0.036}_{-0.033}$	$0.239^{+0.056}_{-0.054}$	4.00σ
	3	$0.101^{+0.056}_{-0.049}$	$0.273^{+0.089}_{-0.085}$	2.23σ	$0.127^{+0.040}_{-0.037}$	$0.266^{+0.063}_{-0.061}$	

W Impact of object systematic uncertainties on the measurement

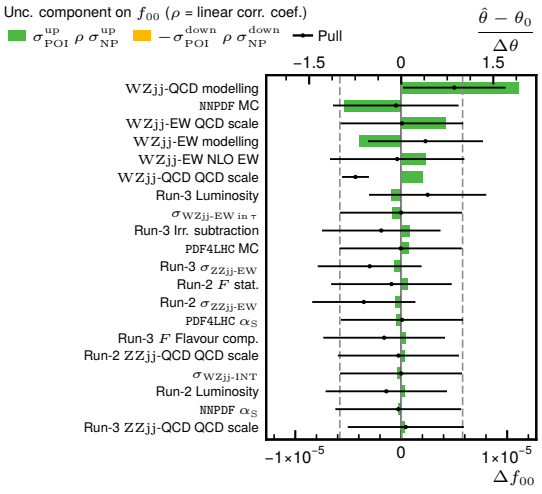
This appendix presents the impact of the object systematic uncertainties on the measurement of the polarisation states. To this end, the same nominal fit presented in Section 7.1 is performed but without the object systematic uncertainties. The results are shown in Table W.A.

This study aims to show that the impact of the object systematics on the fit and pulls is negligible for the polarisation fractions and amounts to about $11\,\%$ to $13\,\%$ of the $\mu^{\text{fid.}}_{\text{WZjj-EW}}$ uncertainty.

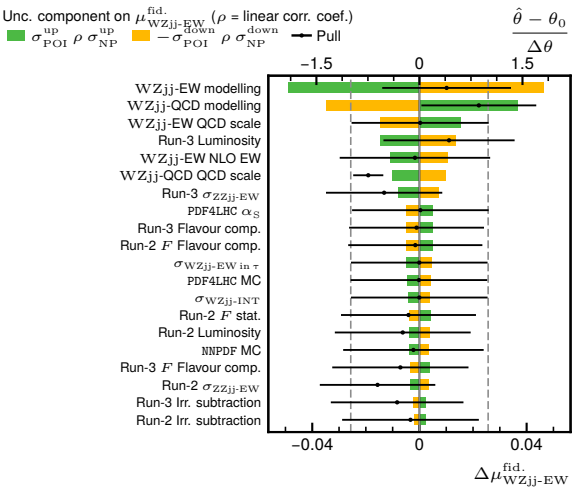
The fit confirms this: the fit does not compensate by pulling the remaining systematics to accommodate the data in the SR, the fitted fractions and their significances are closely similar to the case where the object systematics are included in the fit, and the $\mu^{\text{fid.}}_{\text{WZjj-EW}}$ uncertainty shrinks from about ± 0.16–0.19 to about ± 0.14–0.16. The corresponding ranking is shown in Figure W.1. This confirms that the results obtained in Section 7.3 are not driven by these systematics.

Table W.A: Fit results and significance for each polarisation-state measurement for Run-2 and Run-3 and their combination without the object systematic uncertainties.

Parameter / **Measurement**	**Run**	f_{XX} **Individual**	f_{XX} **Combined**	$\sigma^{\text{fid.}}_{\text{WZjj-EW}}$ [fb] **Individual**	$\mu^{\text{fid.}}_{\text{WZjj-EW}}$ **Combined**
f_{00}	2	$0.054^{+0.099}_{-0.054}$ 0.63σ	$0.000^{+0.052}_{-0.000}$ 0.00σ	$0.363^{+0.074}_{-0.069}$	$1.125^{+0.155}_{-0.147}$
	3	$0.000^{+0.038}_{-0.000}$ 0.00σ		$0.421^{+0.075}_{-0.071}$	
$f_{0\mathrm{X}}$	2	$0.000^{+0.113}_{-0.000}$ 0.00σ	$0.011^{+0.092}_{-0.011}$ 0.13σ	$0.343^{+0.072}_{-0.067}$	$1.063^{+0.154}_{-0.145}$
	3	$0.034^{+0.125}_{-0.034}$ 0.28σ		$0.397^{+0.075}_{-0.071}$	
$f_{\mathrm{X}0}$	2	$0.404^{+0.140}_{-0.126}$ 3.47σ	$0.324^{+0.088}_{-0.083}$ 4.16σ	$0.355^{+0.076}_{-0.072}$	$1.085^{+0.164}_{-0.154}$
	3	$0.255^{+0.124}_{-0.115}$ 2.30σ		$0.389^{+0.079}_{-0.075}$	

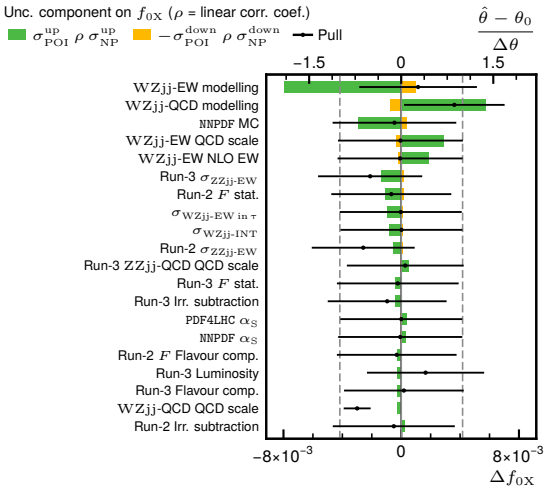


(i) f_{00}

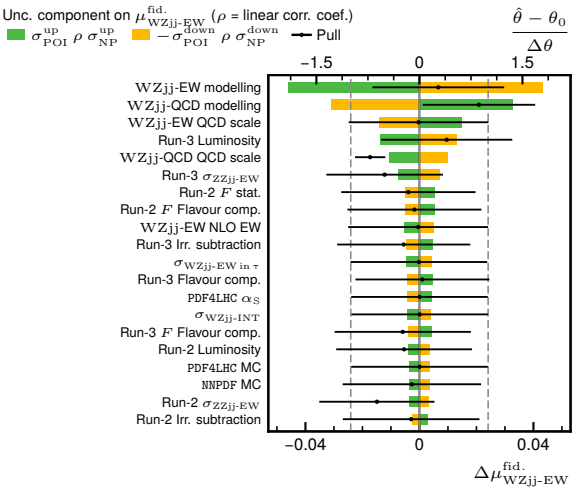


(ii) $\mu^{\text{fid.}}_{\text{WZjj-EW}}$

(a) f_{00} measurement

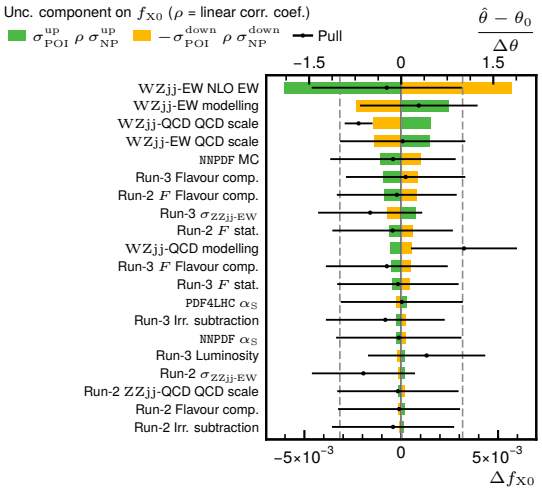


(i) $f_{0\text{X}}$

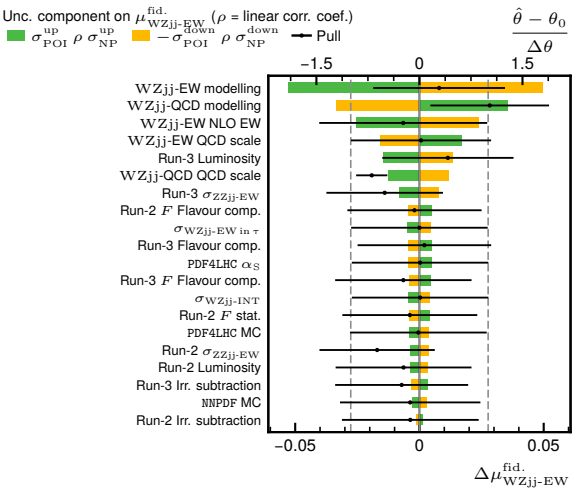


(ii) $\mu^{\text{fid.}}_{\text{WZjj-EW}}$

(b) $f_{0\text{X}}$ measurement



(i) f_{X0}



(ii) $\mu^{\text{fid.}}_{\text{WZjj-EW}}$

(c) f_{X0} measurement

Figure W.1: Ranking and pulls from the 20 most important NPs for the combination of Run-2 and Run-3, with the object systematic uncertainties removed from the fit, for the f_{00} (a), $f_{0\text{X}}$ (b), and f_{X0} (c) measurements. In each measurement, the ranking is shown for both POIs: f_{XX} and $\mu^{\text{fid.}}_{\text{WZjj-EW}}$.

X Measurement of the doubly transverse polarisation state

In this appendix, we present the results of the fit that targets the $\mathrm{W_TZ_Tjj}$-EW contribution instead of the $\mathrm{W_0Z_0jj}$-EW one, using the same template as the nominal fit. The change is made by replacing the NF f_{00} with f_{TT} in the fit model. The expected fraction is thus the one giving the $\mathrm{W_TZ_Tjj}$-EW contribution in the fiducial phase-space, while the $1 - f_{\mathrm{TT}}$ fraction is the one giving the $\mathrm{W_0Z_0jj}$-EW, $\mathrm{W_0Z_Tjj}$-EW, and $\mathrm{W_TZ_0jj}$-EW all together.

The SM prediction for the $\mathrm{W_TZ_Tjj}$-EW contribution is $f_{\mathrm{TT}}^{\mathrm{SM}} = 0.53$ in the fiducial phase-space.

The normalisation of the $\mathrm{W_TZ_Tjj}$-EW contribution is free in the fit and the non-targeted polarisation fractions are floating as the $1 - f_{\mathrm{TT}}$ fraction. The results of the fit are presented in Table X.A. The expected and observed significances are also reported.

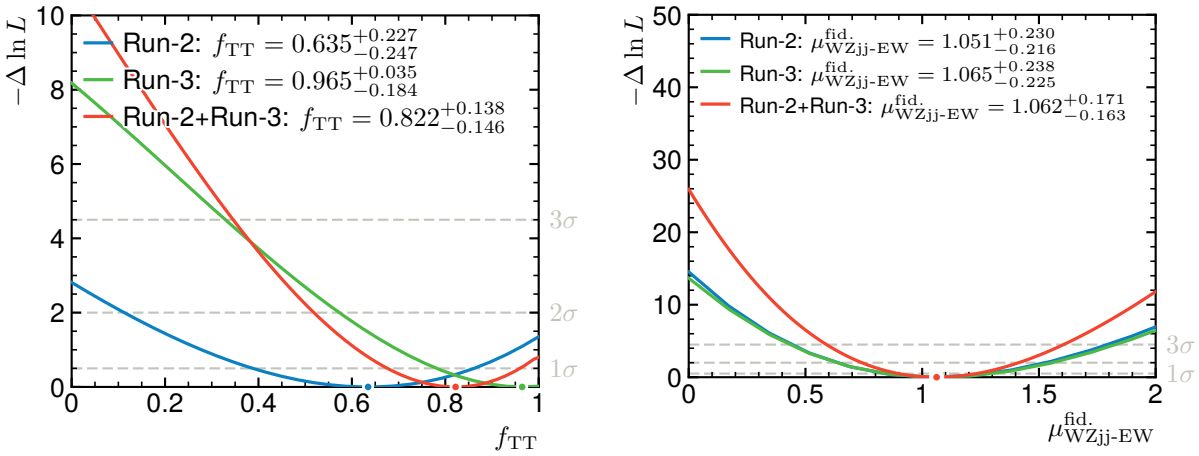
A clear excess of the $\mathrm{W_TZ_Tjj}$-EW contribution is observed in the Run-3 data, while the Run-2 data are compatible with the SM prediction. The combined fit rejects $f_{\mathrm{TT}} = 0$ at 4.7σ. With respect to the SM expectation of 0.53, the measured $f_{\mathrm{TT}} = 0.822^{+0.138}_{-0.146}$ is a 2.0σ excess, driven by the Run-3 data ($0.965^{+0.035}_{-0.184}$), the Run-2 value ($0.635^{+0.227}_{-0.247}$) being compatible with the prediction. It is the counterpart of the downward fluctuation of $f_{0\mathrm{X}}$ discussed in Section 7.3.12: a deficit of longitudinal-W events removes $\mathrm{W_0Z_0}$ and $\mathrm{W_0Z_T}$ events alike, and the remainder is shared between $\mathrm{W_TZ_0}$ and $\mathrm{W_TZ_T}$, which is why f_{X0} and f_{TT} both sit above their expectation. The four fits use the same data and templates and are not independent tests.

We note that the fitted templates are the same as the ones used in the nominal fit, which means that this fit was not specifically designed to target the $\mathrm{W_TZ_Tjj}$-EW contribution. The results of this fit are thus not optimal, and simply reflect this Run-3 excess.

The profile-likelihood scans of f_{TT} and $\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}}$ are shown in Figure X.1 for the Run-2, Run-3, and combined fits, with the same conventions as Appendix S. The Run-3 scan rises steeply below its minimum at $f_{\mathrm{TT}} = 0.965$, close to the boundary $f_{\mathrm{TT}} = 1$, while the Run-2 scan is broad and centred near the SM value, and the combined scan sits between the two.

Table X.A: Results of the fits when fitting the $\mathrm{W_TZ_Tjj}$-EW contribution with the f_{00} template.

POI	Run-2	Run-3	Combined
$\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}}$	$1.051^{+0.230}_{-0.216}$	$1.065^{+0.238}_{-0.225}$	$1.062^{+0.171}_{-0.163}$
$\sigma^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}}$ [fb]	$0.352^{+0.077}_{-0.072}$	$0.397^{+0.089}_{-0.084}$	—
f_{TT}	$0.635^{+0.227}_{-0.247}$	$0.965^{+0.035}_{-0.184}$	$0.822^{+0.138}_{-0.146}$
Z (expected)	1.79	1.91	2.58
Z (observed)	2.37	4.04	4.67



(a) f_{TT} **(b)** $\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}}$

Figure X.1: Profile-likelihood scans for the $\mathrm{W_TZ_T}$jj-EW fit, for the polarisation fraction f_{TT} (a) and the fiducial signal strength $\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}}$ (b), shown for the Run-2, Run-3, and combined Run-2+Run-3 fits.

Y Impact of the polarisation states' composition in fitted templates

In this appendix, we study the impact of the polarisation states' composition in the fitted templates. We start from the modified fit setup of Appendix V, where two NFs are used: μ_{00} for the targeted polarisation and $\mu_{\overline{00}}$ for the non-targeted one. In addition, we assign an overall $25\,\%$ uncertainty to the mixed polarisation templates, $\mathrm{W_0Z_Tjj}$-EW and $\mathrm{W_TZ_0jj}$-EW, to account for the uncertainty on their polarisation composition.

The post-fit values of μ_{00} and $\mu_{\overline{00}}$ are respectively $0.001^{+0.703}_{-0.001}$ and $1.208^{+0.242}_{-0.214}$.

The $25\,\%$ uncertainty on the mixed templates has several effects:

- The post-fit NFs shift similarly to the fixed-composition fit (Appendix V): $\mu_{00} = 0.001$ and $\mu_{\overline{00}} = 1.208$, compatible with both that result and the nominal fit.
- The new uncertainty significantly changes the systematic ranking: it becomes the leading systematic for both NFs. This is expected, as it is by far the largest uncertainty in these samples, well beyond the NLO EW one. The fit uses it to adjust the normalisation of the mixed templates while treating the other systematics as shape uncertainties, which are consequently ranked lower than in the nominal fit.
- The post-fit yields, presented in Table Y.A, of $\mathrm{W_0Z_Tjj}$-EW and $\mathrm{W_TZ_0jj}$-EW are slightly shifted with respect to the fixed-composition fit (Appendix V) and are not reproduced by scaling the pre-fit yields by the post-fit μ_{00} and $\mu_{\overline{00}}$, since the fit can now redistribute the normalisation of the mixed templates through the new $25\,\%$ freedom. The shift is of order $3\,\%$ downwards in both runs and for both templates, as the associated NP is common to the two runs and to the two templates. It is pulled by about -0.2σ, which lowers the mixed templates by about $5\,\%$, partly compensated by the slightly larger $\mu_{\overline{00}}$ which also raises the $\mathrm{W_TZ_Tjj}$-EW yield by about $2\,\%$. The uncertainties on the mixed templates grow accordingly, and the total post-fit yields are unchanged.

Table Y.A: Pre-fit and post-fit event yields in the SR and in the two CRs for the Run-2 and Run-3 datasets, with all sources of uncertainty, for the $\mathrm{W_0Z_0jj}$-EW measurement, in the fit with μ_{00} and $\mu_{\overline{00}}$ as normalisation factors and a $25\,\%$ normalisation uncertainty on the mixed polarisation templates.

(a) Pre-fit

Process	SR		ZZ-CR		b-CR	
	Run-2	Run-3	Run-2	Run-3	Run-2	Run-3
Data	646	953	210	235	666	766
Total pred.	800 ± 190	1060 ± 270	204 ± 11	240 ± 24	687 ± 56	751 ± 75
WZjj-EW	—	—	0.6 ± 0.1	0.8 ± 0.2	5.1 ± 1.3	3.4 ± 0.9
$\mathrm{W_0Z_0jj}$-EW	8.8 ± 0.2	11.2 ± 0.6	—	—	—	—
$\mathrm{W_0Z_Tjj}$-EW	19.6 ± 4.9	25.6 ± 6.5	—	—	—	—
$\mathrm{W_TZ_0jj}$-EW	20.3 ± 5.1	26.4 ± 6.7	—	—	—	—
$\mathrm{W_TZ_Tjj}$-EW	58.4 ± 1.7	76.5 ± 4.5	—	—	—	—
WZjj-QCD	530 ± 190	710 ± 260	6.3 ± 2.1	8.7 ± 3.0	93 ± 36	75 ± 27
Misid. leptons	46 ± 11	62 ± 19	1.7 ± 0.4	2.2 ± 0.5	155 ± 38	195 ± 61
ZZjj-QCD	41.6 ± 2.9	51.8 ± 9.4	158.4 ± 8.6	181 ± 20	10.1 ± 0.6	9.0 ± 0.9
tZj	28.5 ± 1.0	37.3 ± 3.5	$0.4 \pm < 0.1$	0.8 ± 0.1	134.1 ± 5.0	154.4 ± 9.4
$\mathrm{t\bar{t}} + V$	20.9 ± 0.4	22.9 ± 2.2	10.0 ± 0.3	11.9 ± 1.0	289 ± 11	312 ± 19
ZZjj-EW	6.7 ± 1.7	15.4 ± 3.9	22.8 ± 5.7	30.4 ± 7.7	0.3 ± 0.1	0.3 ± 0.1
WZjj-INT	6.5 ± 1.3	9.2 ± 1.9	$0.1 \pm < 0.1$	$0.1 \pm < 0.1$	0.6 ± 0.1	0.4 ± 0.1
WZjj-EW in τ	5.9 ± 0.6	7.5 ± 0.8	—	—	—	—
VVV	2.4 ± 0.6	3.7 ± 0.9	4.1 ± 1.0	4.8 ± 1.2	0.4 ± 0.1	0.4 ± 0.1

(b) Post-fit

Process	SR		ZZ-CR		b-CR	
	Run-2	Run-3	Run-2	Run-3	Run-2	Run-3
Data	646	953	210	235	666	766
Total pred.	668 ± 23	935 ± 28	210 ± 14	238 ± 15	662 ± 26	770 ± 28
WZjj-EW	—	—	0.6 ± 0.1	0.8 ± 0.2	5.0 ± 1.3	3.4 ± 0.9
$\mathrm{W_0Z_0jj}$-EW	$0.0^{+0.3}_{-0.0}$	$0.0^{+0.4}_{-0.0}$	—	—	—	—
$\mathrm{W_0Z_Tjj}$-EW	22.2 ± 4.9	29.9 ± 6.6	—	—	—	—
$\mathrm{W_TZ_0jj}$-EW	23.0 ± 5.1	30.9 ± 6.8	—	—	—	—
$\mathrm{W_TZ_Tjj}$-EW	70 ± 14	95 ± 18	—	—	—	—
WZjj-QCD	391 ± 29	571 ± 35	4.9 ± 0.4	6.8 ± 0.6	63.2 ± 9.8	59.1 ± 4.9
Misid. leptons	45 ± 10	52 ± 14	1.7 ± 0.4	2.2 ± 0.5	152 ± 36	162 ± 46
ZZjj-QCD	43.9 ± 4.0	52.4 ± 5.5	170 ± 15	182 ± 16	9.9 ± 0.6	9.3 ± 0.6
tZj	32.9 ± 6.4	44.1 ± 8.7	0.5 ± 0.1	1.0 ± 0.2	156 ± 29	181 ± 34
$\mathrm{t\bar{t}} + V$	20.0 ± 2.5	25.9 ± 3.7	9.5 ± 1.2	13.5 ± 1.8	276 ± 32	354 ± 41
ZZjj-EW	5.6 ± 1.3	13.5 ± 3.2	19.2 ± 4.6	26.5 ± 6.3	0.2 ± 0.1	0.2 ± 0.1
WZjj-INT	6.5 ± 1.3	9.4 ± 1.9	$0.1 \pm < 0.1$	$0.1 \pm < 0.1$	0.6 ± 0.1	0.4 ± 0.1
WZjj-EW in τ	5.9 ± 0.6	7.6 ± 0.8	—	—	—	—
VVV	2.4 ± 0.6	3.7 ± 0.9	4.1 ± 1.0	4.9 ± 1.2	0.4 ± 0.1	0.4 ± 0.1

Z Scotogenic models and their parameter-space exploration

This appendix is a companion to Chapter 1 and to the ML toolkit of Sections 4.1 and 4.6. It first introduces the *scotogenic* class of BSM models, which addresses simultaneously two of the shortcomings listed in Section 1.10:

- the absence of a DM candidate
- the origin of neutrino masses

and then details how their high-dimensional parameter space is explored with an AL strategy built on the DNN. The MCMC scan mentioned in Section 4.6 serves as its benchmark. The material condenses the study of [16], whose methodological groundwork was laid during an internship that built on the scoring methods of [15]. A complete description of the models, the constraints, and the numerical tools is given in [446], which also compares the AL scan to a more traditional MCMC approach [447]. The present appendix focuses on the AL methodology and its performance.

Z.1 The scotogenic mechanism

The scotogenic[†] idea, introduced by Ma [448], extends the SM by a dark sector odd under an exact discrete Z_2 symmetry. The lightest Z_2-odd state is stable and provides a weakly interacting massive particle (WIMP) DM candidate, while the same Z_2-odd fields running in a one-loop diagram generate Majorana neutrino masses. The two open questions of Section 1.10 are therefore linked at the level of a single mechanism: neutrinos are massless at tree level and acquire mass radiatively through the dark sector, so DM and neutrino phenomenology cannot be decoupled.

Depending on the Z_2-odd field content, several realisations of the loop topology exist [449]. Two frameworks are considered here:

"T1-2G". A scalar singlet S, an inert scalar doublet η, a fermionic Dirac doublet Ψ, and two generations of fermionic triplet $\Sigma_{1,2}$. The two triplets allow three non-zero neutrino masses. Its MCMC phenomenology was studied in [447] and serves here as a benchmark. The MCMC methodology is described in Section 4.6.

"T1-2B ext.". Two fermionic singlets $F_{1,2}$, a Dirac doublet Ψ, an inert scalar doublet η, and a complex scalar triplet Δ. It shares the fermionic sector of the singlet–doublet model of [450] but adds the triplet. The second singlet (the "ext." of the name) provides the third neutrino mass.

The Z_2-odd content of both frameworks, together with the seed "T1-2B" model, is summarised in Table Z.A. Only the SM Higgs doublet, written H here as in the scotogenic literature (it is the ϕ of Section 1.6), acquires a VEV, so Z_2 remains unbroken and the DM candidate stable.

[†] In Greek, "*scoto*" means "dark" and "*genic*" means "to produce".

Table Z.A: Z_2-odd field content of the scotogenic frameworks, with their $\mathrm{SU}(2)_\mathrm{L}$ and $\mathrm{U}(1)_\mathrm{Y}$ representations. A check mark denotes the presence of the field in the given framework.

	F_1	F_2	Ψ_1	Ψ_2	Σ_1	Σ_2	S	η	Δ
$\mathrm{SU}(2)_\mathrm{L}$	1	1	2	2	3	3	1	2	3
$\mathrm{U}(1)_\mathrm{Y}$	0	0	-1	1	0	0	0	1	0
"T1-2B"	✓	—	✓	✓	—	—	—	✓	✓
"T1-2B ext."	✓	✓	✓	✓	—	—	—	✓	✓
"T1-2G"	—	—	✓	✓	✓	✓	✓	✓	—

Z.1.1 Lagrangian and mass spectrum

The two frameworks differ only in their scalar and fermionic content. The "T1-2G" case is quoted here as representative, the "T1-2B ext." expressions following from $S \to \Delta$, the extra singlet F_2, and an additional $\kappa\eta^\dagger\Delta H$ term (full forms in [16]). The scalar potential reads:

$$\begin{aligned}-\mathcal{L}^{\text{T1-2G}}_{\text{scalar}} \supset & M_H^2|H|^2 + \lambda_H|H|^4 + \frac{1}{2}M_S^2S^2 + \frac{1}{2}\lambda_{4S}S^4 + M_\eta^2|\eta|^2 + \lambda_{4\eta}|\eta|^4 \\ & + \frac{1}{2}\lambda_S S^2|H|^2 + \frac{1}{2}\lambda_{S\eta}S^2|\eta|^2 + \lambda_\eta|\eta|^2|H|^2 + \lambda'_\eta|\eta^\dagger H|^2 \\ & + \frac{1}{2}\lambda''_\eta\left((\eta^\dagger H)^2 + \text{h.c.}\right) + \kappa\left(S\eta^\dagger H + \text{h.c.}\right)\,,\end{aligned} \tag{Z.1}$$

while the Z_2-odd fermion masses arise from:

$$-\mathcal{L}^{\text{T1-2G}}_{\text{fermion}} \supset M_\Psi\Psi_1\Psi_2 + \frac{1}{2}M_\Sigma^{ij}\mathrm{Tr}\left(\Sigma_i\Sigma_j\right) + y_1^j\Psi_1\Sigma_j H + y_2^j\Psi_2\Sigma_j\tilde{H} + \text{h.c.}\,, \tag{Z.2}$$

with $\tilde{H} = \mathrm{i}\sigma_2 H^c$ and $i,j \in \{1,2\}$. After spontaneous symmetry breaking, fields sharing the same quantum numbers mix into physical eigenstates: for "T1-2G" four neutral fermions $\chi^0_{1\ldots4}$, three charged fermions $\chi^\pm_{1,2,3}$, two neutral scalars $\phi^0_{1,2}$, one charged scalar $\phi^\pm$, and a pseudoscalar A^0. The DM candidate is the lightest neutral state.

The interaction between the dark sector and the SM LH lepton doublets L_α and RH charged lepton singlets ℓ^c_α ($\alpha = \mathrm{e},\mu,\tau$) both drives neutrino-mass generation and induces the lepton flavour violation (LFV) transitions used as constraints below:

$$-\mathcal{L}^{\text{T1-2G}}_{\text{int}} \supset g^\alpha_\Psi L_\alpha\Psi_2 S + g^\alpha_{\Sigma_j}\eta\Sigma_j L_\alpha + g^\alpha_R\ell^c_\alpha\tilde{\eta}\Psi_1 + \text{h.c.}\,, \tag{Z.3}$$

with $\tilde{\eta} = \mathrm{i}\sigma_2\eta^c$ and implicit sums over j and α. The Yukawa-type couplings are collected in a 3×3 matrix:

$$G = \begin{pmatrix} g^\mathrm{e}_\Psi & g^\mu_\Psi & g^\tau_\Psi \\ g^\mathrm{e}_1 & g^\mu_1 & g^\tau_1 \\ g^\mathrm{e}_2 & g^\mu_2 & g^\tau_2 \end{pmatrix}\,, \tag{Z.4}$$

where $g^\alpha_i \equiv g^\alpha_{\Sigma_i}$ for "T1-2G" (and $g^\alpha_i \equiv g^\alpha_{F_i}$ for "T1-2B ext.").

Z.1.2 Radiative neutrino masses and the Casas–Ibarra parametrisation

Neutrino masses are generated at one loop by the Z_2-odd scalars and fermions of Equation Z.3, as represented in Figure Z.1. The resulting Majorana mass term factorises the couplings from the loop function:

$$\mathcal{L} \supset -\frac{1}{2}\bar{\nu}^c G^\intercal M_L G\nu\,, \tag{Z.5}$$

where the 3×3 matrix M_L contains the loop integrals over the physical scalar and fermion masses and mixings (explicit entries in [16]). Rather than being scanned blindly, the couplings G are fixed from the measured neutrino data through the Casas–Ibarra parametrisation [451]:

$$G = U_L D_L^{-1/2} R D_\nu^{1/2} U^*_{\text{PMNS}}\,, \tag{Z.6}$$

with $D_L = U_L^\dagger M_L U_L$, D_ν the diagonal matrix of light-neutrino masses, R a complex orthogonal rotation depending on three complex parameters r_k ($k=1,2,3$), and U_{PMNS} the PMNS matrix of Section 1.9, including the two Majorana phases α_1,α_2. Equation Z.6 guarantees that every scanned point reproduces the oscillation data by construction, so the neutrino sector never has to be imposed as an *a posteriori* cut. The mixing angles and squared-mass differences are taken from [452], the lightest mass m_{ν_1} and the phases scanned freely within the cosmological bound on $\sum_i m_{\nu_i}$.

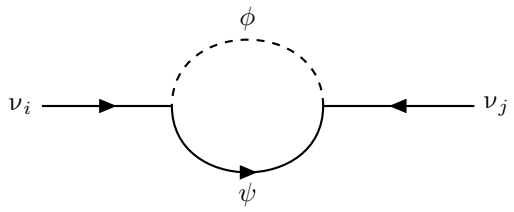


Figure Z.1: One-loop radiative generation of neutrino masses. The generic Z_2-odd fermion $\psi \in \{F, \Psi, \Sigma\}$ and scalar $\phi \in \{S, \eta, \Delta\}$ run in the loop. The external legs are the light neutrinos $\nu_{i,j}$.

Z.1.3 Constraints and numerical tools

Once the neutrino sector is fixed by Equation Z.6, the surviving parameter space is shaped mainly by the DM relic density and by the LFV bounds inherited from Equation Z.3. The headline constraints are collected in Table Z.B. The full set (the complete list of radiative and three-body LFV decays, μ–e conversion on several nuclei, leptonic electric dipole moments, and (for "T1-2G") the anomalous magnetic moments a_e, a_μ) is given in [16].

Table Z.B: Representative constraints imposed in the scan (headline subset). The relic density and Higgs mass enter as Gaussian constraints, their allowed window is widened in the DNN scan, while the LFV observables enter as one-sided upper limits.

Observable	Constraint
m_H	$125.25 \pm 3.00\,\mathrm{GeV}$
$\Omega_\mathrm{DM}h^2$	0.1200 ± 0.0042
$\mathrm{BR}(\mu \to \mathrm{e}\gamma)$	$< 4.2 \cdot 10^{-13}$
$\mathrm{BR}(\mu \to \mathrm{eee})$	$< 1.0 \cdot 10^{-12}$
$\mathrm{CR}(\mu \to \mathrm{e}, \mathrm{Au})$	$< 7.0 \cdot 10^{-13}$

The observables are computed numerically. The models are implemented in SARAH [453], which generates the SPheno [454] modules used to obtain the one-loop mass spectrum and the electroweak and LFV observables, and the CalcHEP files feeding micrOMEGAs [455] for the relic density and the spin-independent direct-detection cross-section. A parameter point is fully specified in SUSY Les Houches Accord (SLHA) format, passed to both codes, and accepted or rejected on the basis of Table Z.B.

Z.2 Parameter-space exploration by active learning

The scan of Table Z.A involves $N \sim \mathcal{O}(20)$ free parameters subject to the stringent constraints of Table Z.B, so a naive random scan is hopeless. Two strategies are compared in [16]: the MCMC technique of Section 4.6, which weights points by a likelihood and converges towards best-fit regions (detailed in [447]), and the AL DNN-based technique described here, which instead *maps the exclusion frontier* defined by the constraints. The AL approach reuses the neural network machinery of Section 4.1 and follows the strategy of [15].

The scan couples two objects:

The Oracle. The expensive but exact evaluator (SPheno and micrOMEGAs) that labels a point as "good" or "bad" from a physics perspective, that is, whether it satisfies the constraints of Table Z.B.

The discriminator. A fast DNN trained to reproduce the Oracle's decision, which then proposes new points to be labelled.

The AL cycle iterates:

1. Train the discriminator on a small labelled set of points.
2. Use the discriminator to select the most informative candidates from a large pool of unlabelled points.
3. Query the Oracle for their labels.
4. Augment the training set with the newly labelled points and retrain the discriminator.

5. Repeat until the labelling budget is exhausted.

The term “active” refers to the fact that the discriminator actively selects the points it wants to learn from, rather than passively receiving a random sample. The DNN is therefore not only a classifier but also a guide to the Oracle, which is queried only on the most informative points.

Z.2.1 Oracle scoring

To avoid a hard cut that discards points marginally outside a single bound, the Oracle assigns a continuous quality score. A candidate is a set of N model parameters $\boldsymbol{x} = [x_1, \dots, x_N]$ governing the physical properties of the model. The DNN operates on the Min-Max normalised parameters $x'_i \in [0,1]$. These are de-normalised back to their physical values x_i before being handed to the Oracle, which evaluates the M observables $\hat{\mathcal{X}}_1, \dots, \hat{\mathcal{X}}_M$ of Table Z.B through SPheno and micrOMEGAs. Each observable is then mapped to an ordinate $\hat{\mathcal{Y}}_i \in [0,1]$ from its prior distribution: a Gaussian observable gives $\hat{\mathcal{Y}}_i = 1$ at the central value and decreases away from it, whereas a one-sided limit is evaluated at a $50\,\%$ tolerance threshold. The point score and the corresponding upper-bound score are the geometric means:

$$\hat{\mathcal{Y}} = \left(\prod_{i=1}^{M} \hat{\mathcal{Y}}_i \right)^{1/M}, \qquad \mathcal{Y} = \left(\prod_{i=1}^{M} \mathcal{Y}_i \right)^{1/M}, \tag{Z.7}$$

the geometric mean being chosen so that no single observable can be completely ignored. The global score:

$$\mathcal{S} = \frac{\hat{\mathcal{Y}}}{\mathcal{Y}}, \tag{Z.8}$$

compares the point to the boundary of the allowed region: it is labelled *good* if $\mathcal{S} > 1$ and *bad* otherwise.

Z.2.2 Discriminator and query strategy

The discriminator is a feed-forward, fully connected DNN of 13 dense layers (21 inputs, one output), with ReLU activations on the hidden layers and a sigmoid on the output, trained by minimising the BCE loss of Equation 4.6 with the NADAM optimiser (a Nesterov-accelerated variant of the ADAM scheme of Section 4.1, with a learning rate of 0.01). Its output $\hat{\boldsymbol{y}} \in [0,1]$ is a continuous “goodness” estimate rather than a hard label, which is exactly what makes the query step below possible.

At each step K candidate points are proposed. Writing q for the current mis-classification rate, taken as the mean absolute difference between the predicted and true labels rounded to one decimal, define:

$$p = 2\min(q, 0.5) \in [0,1], \tag{Z.9}$$

so that $p = 1$ when the network is no better than a coin toss. A fraction pK of the candidates is then drawn uniformly (pure exploration), and the remaining $(1-p)K$ are placed near the current good points to refine the boundary. The boundary candidates are selected from a large pool of $L \gg K$ uniform points by a diversity score combining the network uncertainty and a proximity penalty. The per-point uncertainty:

$$s_i = \hat{y}_i(1 - \hat{y}_i), \tag{Z.10}$$

is maximal at $\hat{y}_i = 0.5$, that is, on the decision boundary.

Defining the squared distance $d_{ij}^2 = \frac{1}{N}\|\boldsymbol{x}'_i - \boldsymbol{x}'_j\|^2$, averaged over the N coordinates, between a pool candidate i and a selected good point j, and the kernel:

$$r(\boldsymbol{x}'_i, \boldsymbol{x}'_j) = \begin{cases} \dfrac{-a}{a + d_{ij}^2} & \text{if } d_{ij} < 0.01\,, \\ 0 & \text{otherwise}\,, \end{cases} \qquad a = 10^{-4}\,, \tag{Z.11}$$

the proximity of candidate i to the good points is $r_i = \sum_j r(\boldsymbol{x}'_i, \boldsymbol{x}'_j)$. The final selection score is:

$$S_i = (1-\alpha)\, s_i + \alpha\, r_i\, \sigma_{\text{std}}\, \frac{1}{4 r_{\min}}\,, \tag{Z.12}$$

where $r_{\min} = \min_i(r_i)$, σ_{std} is the standard deviation of the s_i, and the diversity weight α is taken small when few interesting regions remain. The highest-scoring points are appended to the training set and the discriminator retrained, closing the cycle.

The full workflow is summarised in Figure Z.2.

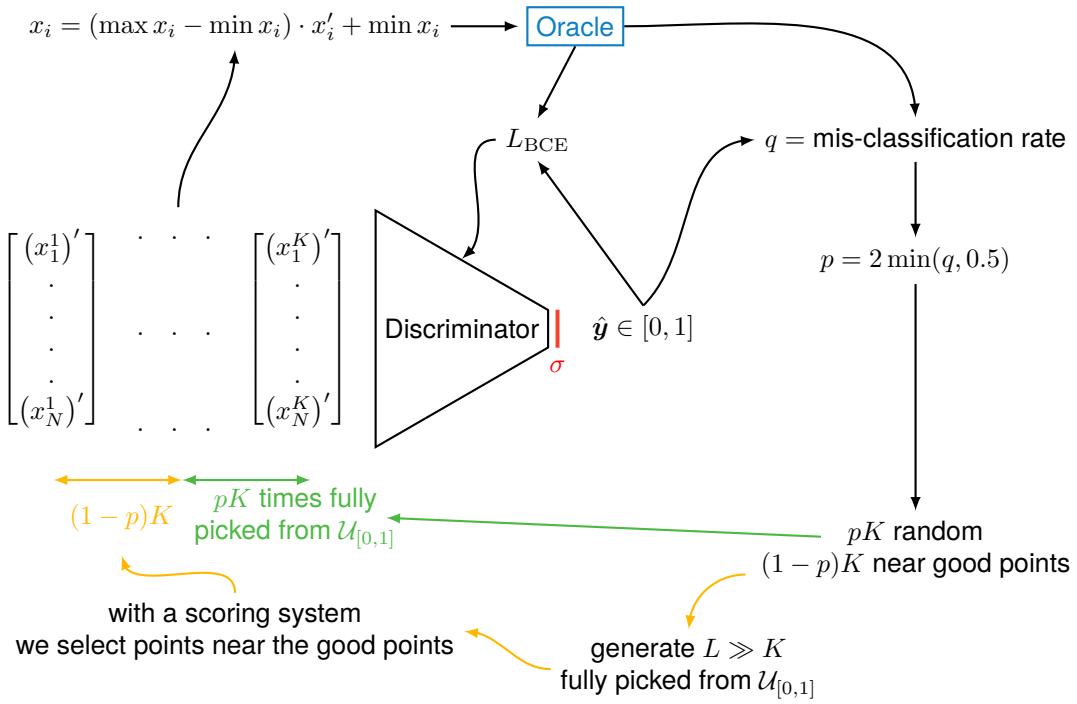


Figure Z.2: Workflow for the AL process. The input data, of size K, are both given to the DNN discriminator and the Oracle (SPheno and micrOMEGAs).

Z.2.3 Toy-model validation

The procedure is validated on a 2D toy problem in which the "physics" is a known region: a point $(x, y) \in [0,1]^2$ is good if it lies inside a circle of radius 0.5 centred on $(0.2, 0.6)$:

$$\mathcal{C} = \begin{cases} 1 & \text{if } (x-0.2)^2 + (y-0.6)^2 \leq 0.25\,, \\ 0 & \text{otherwise}\,. \end{cases} \tag{Z.13}$$

Running the AL loop with $K = 100$, $L = 5000$, $\alpha = 5 \cdot 10^{-3}$, and $I = 200$ iterations, the discriminator reproduces the circle boundary and the sampled density concentrates on it, as shown in Figure Z.3. The high-dimensional scotogenic scan proceeds identically: the AL loop maps the exclusion frontier of the Oracle, which, unlike the toy case, cannot be drawn directly.

Z.2.4 Performance and outlook

Applied to "T1-2G", the AL scan submitted $\sim 6 \cdot 10^7$ points to the Oracle over 13 weeks and retained $27\,000$ good points, an efficiency of $\sim 0.045\,\%$. The MCMC scan of [447] reached $0.29\,\%$ ($117\,000$ of $4 \cdot 10^7$ in 8 weeks), over a larger space since it also varies the neutrino-sector parameters of the Casas–Ibarra parametrisation, which the AL scan keeps fixed. The MCMC thus accepts a larger fraction and costs less per accepted point, but the two methods answer different questions: because the AL loop targets the boundary rather than best-fit regions, its DM-mass distributions are comparatively flat between well-defined edges, whereas the likelihood-driven MCMC peaks at the best-fit mass. For the fermionic candidate, both nonetheless yield compatible phenomenology: co-annihilating fermionic DM in the doublet/triplet with small mass splittings, DM masses in the TeV range, and spin-independent cross-sections close to the current direct-detection limits.

The main bottleneck is the SUSY Les Houches Accord write-out to the external codes. A natural improvement is a GAN (Section 4.1) whose generator, coupled to the trained discriminator, would propose good points directly and largely replace the random component of Equations Z.9 and Z.12, or even bypass Equation Z.6 to supply viable neutrino-sector solutions, an avenue left to future work.

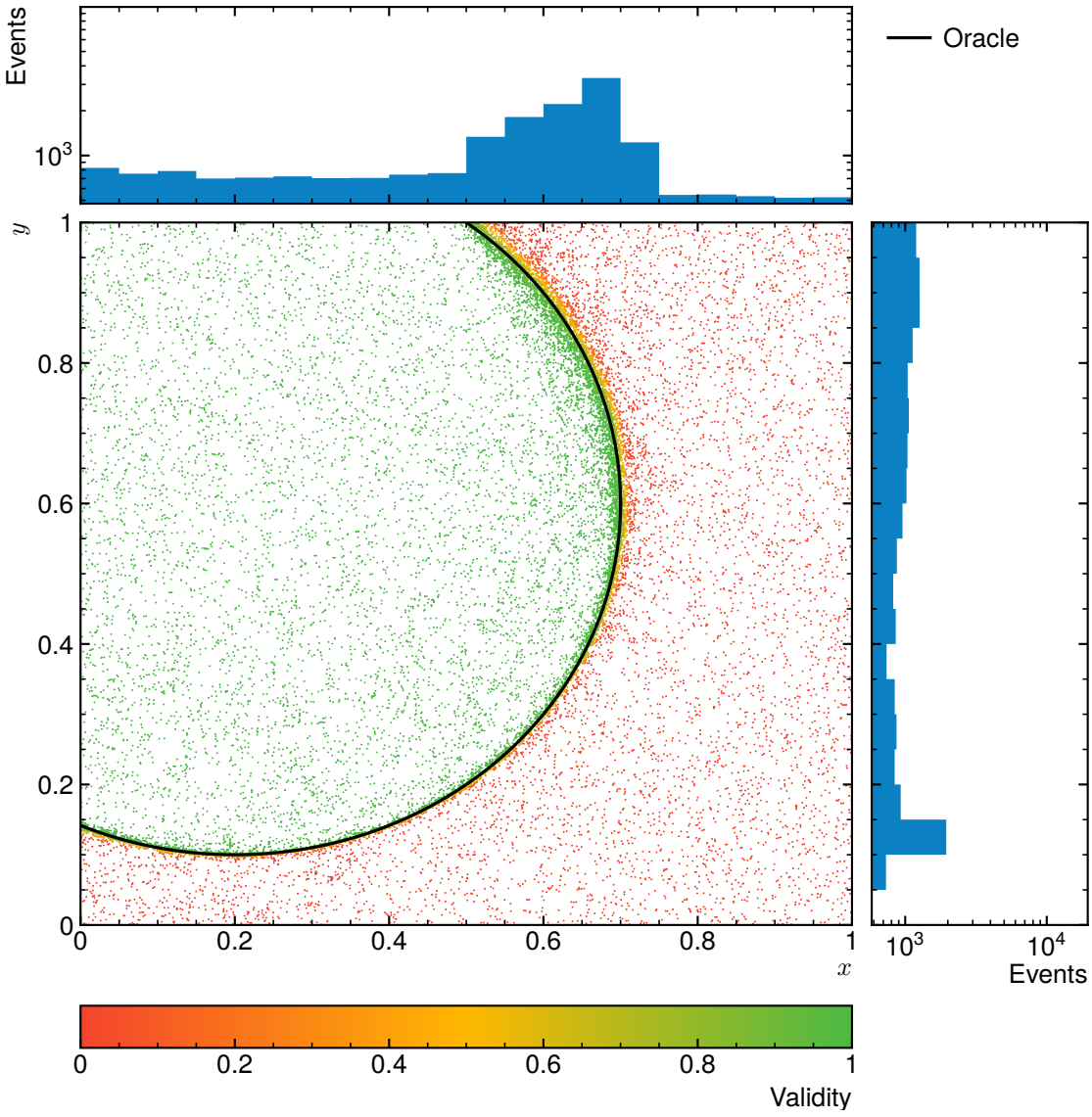
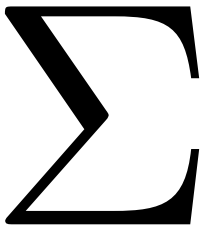


Figure Z.3: AL on the toy model: good and bad points (green/red) and the Oracle boundary (solid). The projected densities peak on the boundary, demonstrating that the AL loop maps the frontier rather than converging to an interior optimum.



Bibliography

[1] ATLAS Collaboration, *Observation of a New Particle in the Search for the Standard Model Higgs Boson with the ATLAS Detector at the LHC*, Physics Letters B **716** (2012) 1, URL: https://cds.cern.ch/record/1471031 (Cited on pages 1, 5, 23, 24).

[2] CMS Collaboration, *Observation of a New Boson at a Mass of* $125\,\mathrm{GeV}$ *with the CMS Experiment at the LHC*, Physics Letters B **716** (2012) 30, URL: https://cds.cern.ch/record/1471016 (Cited on pages 1, 5, 23).

[3] J. M. Cornwall, D. N. Levin and G. Tiktopoulos, *Derivation of gauge invariance from high-energy unitarity bounds on the $\mathcal{S}$ matrix*, Physical Review D **10** (1974) 1145 (Cited on pages 1, 5, 181, 258).

[4] B. W. Lee, C. Quigg and H. B. Thacker, *Weak interactions at very high energies: The role of the Higgs-boson mass*, Physical Review D **16** (1977) 1519 (Cited on pages 1, 5, 181, 258, 259, 265).

[5] CERN, *LHC Design Report, Vol. I: The LHC Main Ring*, CERN Yellow Reports: Monographs CERN-2004-003-V-1, CERN, 2004, URL: https://cds.cern.ch/record/782076 (Cited on pages 1, 5, 32, 38).

[6] CERN, *LHC Design Report, Vol. II: The LHC Infrastructure and General Services*, CERN Yellow Reports: Monographs CERN-2004-003-V-2, CERN, 2004, URL: https://cds.cern.ch/record/815187 (Cited on pages 1, 5, 32).

[7] CERN, *LHC Design Report, Vol. III: The LHC Injector Chain*, CERN Yellow Reports: Monographs CERN-2004-003-V-3, CERN, 2004, URL: https://cds.cern.ch/record/823808 (Cited on pages 1, 5, 32).

[8] CERN LHC Experiments Committee, *ATLAS: A general-purpose* pp *experiment at the Large Hadron Collider at CERN*, LHC Technical Proposal CERN-LHCC-94-43, LHCC-P-2, CERN, 1994, URL: https://cds.cern.ch/record/290968 (Cited on pages 1, 5, 35).

[9] CERN LHC Experiments Committee, *CMS: The Compact Muon Solenoid*, LHC Technical Proposal CERN-LHCC-94-38, LHCC-P-1, CERN, 1994, URL: https://cds.cern.ch/record/290969 (Cited on pages 1, 5, 35).

[10] ATLAS Collaboration, *Measurement of* $\mathrm{W^{\pm}Z}$ *production cross sections and gauge boson polarisation in* pp *collisions at* $\sqrt{s} = 13\,\mathrm{TeV}$ *with the ATLAS detector*, The European Physical Journal C **79** (2019) 535, URL: https://cds.cern.ch/record/2658356 (Cited on pages 2, 5, 145, 146, 166, 169, 188, 286, 290).

[11] ATLAS Collaboration, *Observation of gauge boson joint-polarisation states in* $\mathrm{W^{\pm}Z}$ *production from* pp *collisions at* $\sqrt{s} = 13\,\mathrm{TeV}$ *with the ATLAS detector*, Physics Letters B **843** (2023) 137895, URL: https://cds.cern.ch/record/2841241 (Cited on pages 2, 5, 145, 146, 166, 188, 194, 208, 252, 284).

[12] CERN, *High-Luminosity Large Hadron Collider (HL-LHC)*, CERN Yellow Reports: Monographs CERN-2020-010, CERN, 2020, URL: https://cds.cern.ch/record/2749422 (Cited on pages 2, 6, 37, 38, 255).

[13] ATLAS Collaboration, *ATLAS Inner Tracker Pixel Detector*, Technical Design Report CERN-LHCC-2017-021, ATLAS-TDR-030, CERN, 2017, URL: https://cds.cern.ch/record/2285585 (Cited on pages 2, 6, 67, 70, 255).

[14] ATLAS Collaboration, *ATLAS Inner Tracker Strip Detector*, Technical Design Report CERN-LHCC-2017-005, ATLAS-TDR-025, CERN, 2017, URL: https://cds.cern.ch/record/2257755 (Cited on pages 2, 6, 67, 68, 255).

[15] M. D. Goodsell and A. Joury, *Active learning BSM parameter spaces*, The European Physical Journal C **83** (2023) 268 (Cited on pages 4, 8, 406, 408).

[16] U. de Noyers, M. Dubau, B. Herrmann and O. Arnaez, *On the efficiency of parameter space exploration: A scotogenic case study*, 2025, arXiv: 2512.06378 [hep-ph], URL: https://arxiv.org/abs/2512.06378 (Cited on pages 4, 8, 406–408).

[17] D. I. Mendeleev, *The natural system of elements and its application to the indication of the properties of undiscovered elements*, Journal of the Russian Chemical Society **3** (1871) 25 (Cited on page 10).

[18] J. J. Thomson, *Cathode Rays*, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **44** (1897) 293 (Cited on page 10).

[19] M. Planck, *Zur Theorie des Gesetzes der Energieverteilung im Normalspectrum*, Verhandlungen der Deutschen Physikalischen Gesellschaft **2** (1900) 237 (Cited on page 10).

[20] E. Rutherford, *The Scattering of α and β Particles by Matter and the Structure of the Atom*, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **21** (1911) 669 (Cited on page 10).

[21] N. Bohr, *On the Constitution of Atoms and Molecules*, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **26** (1913) 1 (Cited on page 10).

[22] L. de Broglie, *Recherches sur la Théorie des Quanta*, Annales de Physique **10** (1925) 22 (Cited on page 10).

[23] W. Heisenberg, *Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen*, Zeitschrift für Physik **33** (1925) 879 (Cited on page 10).

[24] M. Born and P. Jordan, *Zur Quantenmechanik*, Zeitschrift für Physik **34** (1925) 858 (Cited on page 10).

[25] M. Born, W. Heisenberg and P. Jordan, *Zur Quantenmechanik. II.*, Zeitschrift für Physik **35** (1926) 557 (Cited on page 10).

[26] E. Schrödinger, *Quantisierung als Eigenwertproblem*, Annalen der Physik **384** (1926) 361 (Cited on page 10).

[27] A. Einstein, *Zur Elektrodynamik bewegter Körper*, Annalen der Physik **322** (1905) 891 (Cited on page 10).

[28] J. C. Maxwell, *A Dynamical Theory of the Electromagnetic Field*, Philosophical Transactions of the Royal Society of London **155** (1865) 459 (Cited on page 10).

[29] H. Minkowski, *Raum und Zeit*, Physikalische Zeitschrift **10** (1909) 75 (Cited on page 10).

[30] O. Klein, *Quantentheorie und fünfdimensionale Relativitätstheorie*, Zeitschrift für Physik **37** (1926) 895 (Cited on page 10).

[31] W. Gordon, *Der Comptoneffekt nach der Schrödingerschen Theorie*, Zeitschrift für Physik **40** (1926) 117 (Cited on page 10).

[32] P. A. M. Dirac, *The Quantum Theory of the Electron*, Proceedings of the Royal Society of London. Series A **117** (1928) 610 (Cited on page 10).

[33] V. Fock, *Konfigurationsraum und zweite Quantelung*, Zeitschrift für Physik **75** (1932) 622 (Cited on pages 10, 12).

[34] F. Takahashi et al., *Review of particle physics*, International Journal of Modern Physics A **41** (2026) 2630011 (Cited on pages 11, 25, 27, 38, 47, 48, 50, 189, 358).

[35] J. L. Lagrange, *Mécanique analytique*, Paris: Veuve Desaint, 1788 (Cited on page 11).

[36] L. Euler, *Methodus inveniendi lineas curvas maximi minimive proprietate gaudentes*, Lausannae et Genevae: Apud Marcum-Michaelem Bousquet & Socios, 1744 (Cited on page 12).

[37] F. J. Dyson, *The $\mathcal{S}$ Matrix in Quantum Electrodynamics*, Physical Review **75** (1949) 1736 (Cited on page 12).

[38] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory*, Addison–Wesley, 1995 (Cited on page 12).

[39] P. Zeeman, *On the Influence of Magnetism on the Nature of the Light Emitted by a Substance*, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **43** (1897) 226 (Cited on page 13).

[40] W. Gerlach and O. Stern, *Der experimentelle Nachweis der Richtungsquantelung im Magnetfeld*, Zeitschrift für Physik **9** (1922) 349 (Cited on page 13).

[41] W. Pauli, *Zur Quantenmechanik des magnetischen Elektrons*, Zeitschrift für Physik **43** (1927) 601 (Cited on page 13).

[42] S. H. Neddermeyer and C. D. Anderson, *Note on the Nature of Cosmic-Ray Particles*, Physical Review **51** (1937) 884 (Cited on page 13).

[43] H. Weyl, *Gravitation and the Electron*, Proceedings of the National Academy of Sciences **15** (1929) 323 (Cited on page 14).

[44] A. H. Compton, *A Quantum Theory of the Scattering of X-rays by Light Elements*, Physical Review **21** (1923) 483 (Cited on page 15).

[45] W. Heisenberg, *Über den Bau der Atomkerne. I*, Zeitschrift für Physik **77** (1932) 1 (Cited on page 15).

[46] C. M. G. Lattes, H. Muirhead, G. P. S. Occhialini and C. F. Powell, *Processes Involving Charged Mesons*, Nature **159** (1947) 694 (Cited on page 15).

[47] M. Gell-Mann, *A Schematic Model of Baryons and Mesons*, Physics Letters **8** (1964) 214 (Cited on page 16).

[48] H. L. Anderson, E. Fermi, E. A. Long and D. E. Nagle, *Total Cross Sections of Positive Pions in Hydrogen*, Physical Review **85** (1952) 936 (Cited on page 16).

[49] V. E. Barnes et al., *Observation of a Hyperon with Strangeness Minus Three*, Physical Review Letters **12** (1964) 204 (Cited on page 16).

[50] M. Y. Han and Y. Nambu, *Three-Triplet Model with Double* $\mathrm{SU}(3)$ *Symmetry*, Physical Review **139** (1965) B1006 (Cited on page 16).

[51] TASSO Collaboration, *Evidence for planar events in* $\mathrm{e}^{+}\mathrm{e}^{-}$ *annihilation at high energies*, Physics Letters B **86** (1979) 243 (Cited on page 16).

[52] Mark J Collaboration, *Discovery of Three-Jet Events and a Test of Quantum Chromodynamics at PETRA*, Physical Review Letters **43** (1979) 830 (Cited on page 16).

[53] JADE Collaboration, *Observation of planar three-jet events in* $\mathrm{e}^{+}\mathrm{e}^{-}$ *annihilation and evidence for gluon bremsstrahlung*, Physics Letters B **91** (1980) 142 (Cited on page 16).

[54] PLUTO Collaboration, *Evidence for gluon bremsstrahlung in* $\mathrm{e}^{+}\mathrm{e}^{-}$ *annihilations at high energies*, Physics Letters B **86** (1979) 418 (Cited on page 16).

[55] D. J. Gross and F. Wilczek, *Ultraviolet Behavior of Non-Abelian Gauge Theories*, Physical Review Letters **30** (1973) 1343 (Cited on page 17).

[56] H. D. Politzer, *Reliable Perturbative Results for Strong Interactions?*, Physical Review Letters **30** (1973) 1346 (Cited on page 17).

[57] H. Becquerel, *Sur les radiations invisibles émises par les corps phosphorescents*, Comptes Rendus de l'Académie des Sciences **122** (1896) 501 (Cited on page 17).

[58] P. Curie and M. Curie, *Sur une substance nouvelle radio-active, contenue dans la pechblende*, Comptes Rendus de l'Académie des Sciences **127** (1898) 175 (Cited on page 17).

[59] P. Curie, M. Curie and G. Bémont, *Sur une nouvelle substance fortement radio-active, contenue dans la pechblende*, Comptes Rendus de l'Académie des Sciences **127** (1898) 1215 (Cited on page 17).

[60] E. Rutherford, *Collision of* α *Particles with Light Atoms. IV. An Anomalous Effect in Nitrogen*, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **37** (1919) 581 (Cited on page 17).

[61] J. Chadwick, *The Existence of a Neutron*, Proceedings of the Royal Society of London. Series A **136** (1932) 692 (Cited on page 17).

[62] E. Fermi, *Tentativo di una Teoria dei Raggi* β, Il Nuovo Cimento **11** (1934) 1 (Cited on page 17).

[63] C. L. Cowan, F. Reines, F. B. Harrison, H. W. Kruse and A. D. McGuire, *Detection of the Free Neutrino: A Confirmation*, Science **124** (1956) 103 (Cited on page 17).

[64] G. Danby et al., *Observation of High-Energy Neutrino Reactions and the Existence of Two Kinds of Neutrinos*, Physical Review Letters **9** (1962) 36 (Cited on page 17).

[65] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes and R. P. Hudson, *Experimental Test of Parity Conservation in Beta Decay*, Physical Review **105** (1957) 1413 (Cited on page 17).

[66] C. N. Yang and R. L. Mills, *Conservation of Isotopic Spin and Isotopic Gauge Invariance*, Physical Review **96** (1954) 191 (Cited on page 17).

[67] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Physical Review Letters **10** (1963) 531 (Cited on page 18).

[68] S. L. Glashow, J. Iliopoulos and L. Maiani, *Weak Interactions with Lepton–Hadron Symmetry*, Physical Review D **2** (1970) 1285 (Cited on page 18).

[69] J. J. Aubert et al., *Experimental Observation of a Heavy Particle* J, Physical Review Letters **33** (1974) 1404 (Cited on page 18).

[70] SLAC-LBL (Mark I) Collaboration, *Discovery of a Narrow Resonance in* $\mathrm{e^+e^-}$ *Annihilation*, Physical Review Letters **33** (1974) 1406 (Cited on page 18).

[71] J. H. Christenson, J. W. Cronin, V. L. Fitch and R. Turlay, *Evidence for the* 2π *Decay of the* $\mathrm{K_2^0}$ *Meson*, Physical Review Letters **13** (1964) 138 (Cited on page 18).

[72] M. Kobayashi and T. Maskawa, *CP-Violation in the Renormalizable Theory of Weak Interaction*, Progress of Theoretical Physics **49** (1973) 652 (Cited on page 18).

[73] E288 Collaboration, *Observation of a Dimuon Resonance at* $9.5\,\mathrm{GeV}$ *in* $400\,\mathrm{GeV}$ *Proton–Nucleus Collisions*, Physical Review Letters **39** (1977) 252 (Cited on page 19).

[74] CDF Collaboration, *Observation of Top Quark Production in* $\mathrm{\overline{p}p}$ *Collisions with the Collider Detector at Fermilab*, Physical Review Letters **74** (1995) 2626 (Cited on page 19).

[75] DØ Collaboration, *Observation of the Top Quark*, Physical Review Letters **74** (1995) 2632 (Cited on page 19).

[76] J. Goldstone, *Field Theories with Superconductor Solutions*, Il Nuovo Cimento **19** (1961) 154 (Cited on page 19).

[77] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Physical Review Letters **13** (1964) 508 (Cited on page 19).

[78] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Physical Review Letters **13** (1964) 321 (Cited on page 19).

[79] S. L. Glashow, *Partial-Symmetries of Weak Interactions*, Nuclear Physics **22** (1961) 579 (Cited on page 20).

[80] S. Weinberg, *A Model of Leptons*, Physical Review Letters **19** (1967) 1264 (Cited on page 20).

[81] A. Salam, "Weak and Electromagnetic Interactions", *Elementary Particle Theory*, ed. by N. Svartholm, Stockholm: Almqvist & Wiksell, 1968, pp. 367–377 (Cited on page 20).

[82] G. 't Hooft, *Renormalizable Lagrangians for massive Yang–Mills fields*, Nuclear Physics B **35** (1971) 167 (Cited on page 21).

[83] G. 't Hooft and M. Veltman, *Regularization and renormalization of gauge fields*, Nuclear Physics B **44** (1972) 189 (Cited on page 21).

[84] UA1 Collaboration, *Experimental observation of isolated large transverse energy electrons with associated missing energy at* $\sqrt{s} = 540\,\mathrm{GeV}$, Physics Letters B **122** (1983) 103, URL: https://cds.cern.ch/record/142059 (Cited on pages 21, 32).

[85] UA2 Collaboration, *Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN* $\mathrm{\overline{p}p}$ *collider*, Physics Letters B **122** (1983) 476, URL: https://cds.cern.ch/record/142759 (Cited on pages 21, 32).

[86] UA1 Collaboration, *Experimental observation of lepton pairs of invariant mass around* $95\,\mathrm{GeV}/c^2$ *at the CERN SPS collider*, Physics Letters B **126** (1983) 398, URL: https://cds.cern.ch/record/163857 (Cited on pages 21, 32).

[87] UA2 Collaboration, *Evidence for* $\mathrm{Z^0 \to e^+e^-}$ *at the CERN* $\mathrm{\overline{p}p}$ *collider*, Physics Letters B **129** (1983) 130, URL: https://cds.cern.ch/record/146503 (Cited on pages 21, 32).

[88] CERN SPS Experiments Committee, *A 4π solid angle detector for the SPS used as a proton-antiproton collider at a centre of mass energy of $540\,\mathrm{GeV}$*, Technical Proposal CERN-SPSC-78-6, CERN-SPSC-78-06, SPSC-P-92, CERN, 1978, URL: https://cds.cern.ch/record/319371 (Cited on pages 21, 32).

[89] CERN SPS Experiments Committee, *Proposal to study antiproton–proton interactions at $540\,\mathrm{GeV}$ cm energy*, Technical Proposal CERN-SPSC-78-8, SPSC-P-93, CERN, 1978, URL: https://cds.cern.ch/record/596804 (Cited on pages 21, 32).

[90] ALEPH, DELPHI, L3, OPAL and SLD Collaborations, LEP Electroweak Working Group, SLD Electroweak and Heavy Flavour Groups, *Precision Electroweak Measurements on the Z Resonance*, Physics Reports **427** (2006) 257 (Cited on pages 21, 22, 24, 32).

[91] H. Yukawa, *On the Interaction of Elementary Particles. I*, Proceedings of the Physico-Mathematical Society of Japan. 3rd Series **17** (1935) 48 (Cited on page 23).

[92] SLAC-LBL (Mark I) Collaboration, *Evidence for Anomalous Lepton Production in $\mathrm{e^+e^-}$ Annihilation*, Physical Review Letters **35** (1975) 1489 (Cited on page 24).

[93] DONUT Collaboration, *Observation of tau neutrino interactions*, Physics Letters B **504** (2001) 218 (Cited on page 24).

[94] Super-Kamiokande Collaboration, *Evidence for Oscillation of Atmospheric Neutrinos*, Physical Review Letters **81** (1998) 1562 (Cited on page 24).

[95] SNO Collaboration, *Measurement of the Rate of $\nu_\mathrm{e} + \mathrm{d} \to \mathrm{p} + \mathrm{p} + \mathrm{e}^-$ Interactions Produced by $^8\mathrm{B}$ Solar Neutrinos at the Sudbury Neutrino Observatory*, Physical Review Letters **87** (2001) 071301 (Cited on page 24).

[96] B. Pontecorvo, *Neutrino Experiments and the Problem of Conservation of Leptonic Charge*, Soviet Physics JETP **26** (1968) 984 (Cited on page 24).

[97] Z. Maki, M. Nakagawa and S. Sakata, *Remarks on the Unified Model of Elementary Particles*, Progress of Theoretical Physics **28** (1962) 870 (Cited on page 24).

[98] E. Majorana, *Teoria simmetrica dell'elettrone e del positrone*, Il Nuovo Cimento **14** (1937) 171 (Cited on page 24).

[99] M. Thomson, *Modern Particle Physics*, Cambridge University Press, 2013 (Cited on page 25).

[100] Planck Collaboration, *Planck 2015 results. XIII. Cosmological parameters*, Astronomy & Astrophysics **594** (2016) A13 (Cited on page 27).

[101] A. Proca, *Sur la théorie ondulatoire des électrons positifs et négatifs*, Journal de Physique et le Radium **7** (1936) 347 (Cited on page 28).

[102] M. S. Chanowitz and M. K. Gaillard, *The TeV physics of strongly interacting W's and Z's*, Nuclear Physics B **261** (1985) 379 (Cited on page 30).

[103] *Convention for the establishment of a European organization for nuclear research: Paris, 1st July, 1953, as amended*, Geneva: CERN, 1971, URL: https://cds.cern.ch/record/330625 (Cited on page 32).

[104] CERN Synchro-Cyclotron Division, *The design of the CERN Synchro-cyclotron control system*, CERN Yellow Reports: Monographs CERN-56-01, CERN, 1956, URL: https://cds.cern.ch/record/212487 (Cited on page 32).

[105] CERN Synchro-Cyclotron Division, *The design of the CERN Synchro-cyclotron magnet*, CERN Yellow Reports: Monographs CERN-55-03, CERN, 1955, URL: https://cds.cern.ch/record/212135 (Cited on page 32).

[106] CERN Proton Synchrotron Division, *The CERN Proton Synchrotron, Vol. I*, CERN Yellow Reports: Monographs CERN-59-29, CERN, 1959, URL: https://cds.cern.ch/record/214352 (Cited on page 32).

[107] CERN Proton Synchrotron Division, *The CERN Proton Synchrotron, Vol. II*, CERN Yellow Reports: Monographs CERN-60-26, CERN, 1960, URL: https://cds.cern.ch/record/102347 (Cited on page 32).

[108] CERN Proton Synchrotron Division, *The CERN Proton Synchrotron, Vol. III*, CERN Yellow Reports: Monographs CERN-62-03, CERN, 1962, URL: https://cds.cern.ch/record/278715 (Cited on page 32).

[109] A. Lagarrigue, "The 'Gargamelle' project", *CERN Yellow Reports: Conference Proceedings*, CERN-65-32, Geneva: CERN, 1965, pp. 263–268, URL: https://cds.cern.ch/record/862832 (Cited on page 32).

[110] Gargamelle Collaboration, *Search for elastic muon neutrino electron scattering*, Physics Letters B **46** (1973) 121,
URL: https://cds.cern.ch/record/243640 (Cited on page 32).

[111] Gargamelle Collaboration, *Observation of neutrino-like interactions without muon or electron in the Gargamelle neutrino experiment*, Nuclear Physics B **73** (1974) 1,
URL: https://cds.cern.ch/record/203096 (Cited on page 32).

[112] CERN, *The* $300\,\mathrm{GeV}$ *programme*, Technical Design Report CERN-1050, CERN, 1972,
URL: https://cds.cern.ch/record/104068 (Cited on page 32).

[113] CERN Large Electron–Positron Division, *LEP Design Report, Vol. I: The LEP injector chain*, Technical Design Report CERN-LEP-TH-83-29, CERN-PS-DL-83-31, CERN-SPS-83-26, LAL-RT-83-09, CERN, 1983,
URL: https://cds.cern.ch/record/98881 (Cited on page 32).

[114] CERN Large Electron–Positron Division, *LEP Design Report, Vol. II: The LEP main ring*, Technical Design Report CERN-LEP-84-01, CERN, 1984,
URL: https://cds.cern.ch/record/102083 (Cited on page 32).

[115] CERN Large Electron–Positron Division, *LEP Design Report, Vol. III: LEP2*, Technical Design Report CERN-AC-96-01-LEP-2, CERN, 1996,
URL: https://cds.cern.ch/record/314187 (Cited on page 32).

[116] CERN LEP Experiments Committee, *ALEPH*, Technical Proposal CERN-LEPC-83-2, LEPC-P-1, CERN, 1983,
URL: https://cds.cern.ch/record/300680 (Cited on page 32).

[117] CERN LEP Experiments Committee, *DELPHI*, Technical Proposal CERN-LEPC-83-3, DELPHI-83-66, LEPC-P-2, CERN, 1983,
URL: https://cds.cern.ch/record/300668 (Cited on page 32).

[118] CERN LEP Experiments Committee, *OPAL*, Technical Proposal CERN-LEPC-83-4, LEPC-P-3, CERN, 1983,
URL: https://cds.cern.ch/record/300669 (Cited on page 32).

[119] CERN LEP Experiments Committee, *L3*, Technical Proposal CERN-LEPC-83-5, LEPC-P-4, CERN, 1983,
URL: https://cds.cern.ch/record/297266 (Cited on page 32).

[120] L. R. Evans and P. Bryant, *LHC Machine*, Journal of Instrumentation **3** (2008) S08001,
URL: https://cds.cern.ch/record/1129806 (Cited on page 33).

[121] O. B. Morgan, G. G. Kelley and R. C. Davis, *Technology of Intense dc Ion Beams*, Review of Scientific Instruments **38** (1967) 467 (Cited on page 33).

[122] CERN New Linac Working Group, *Project study for a new* $50\,\mathrm{MeV}$ *linear accelerator for the C.P.S.*, Technical Report CERN-MPS-LINP-73-1, CERN, 1973,
URL: https://cds.cern.ch/record/414071 (Cited on page 33).

[123] C. E. Hill, A. M. Lombardi, E. Tanke and M. Vretenar, "Present performance of the CERN proton linac", *Proceedings of the 19th International Linear Accelerator Conference*, Argonne, USA: Argonne Nat. Lab., 1998, pp. 427–429,
URL: https://cds.cern.ch/record/366818 (Cited on page 33).

[124] CERN, *Linac4 Design Report*, CERN Yellow Reports: Monographs CERN-2020-006, CERN, 2020,
URL: https://cds.cern.ch/record/2736208 (Cited on pages 33, 37).

[125] CERN LHC Experiments Committee, *ALICE: A Large Ion Collider Experiment at the CERN LHC*, LHC Technical Proposal CERN-LHCC-95-71, LHCC-P-3, CERN, 1995,
URL: https://cds.cern.ch/record/293391 (Cited on page 35).

[126] E. Quercigh and J. Rafelski, *A strange quark plasma*, Physics World **13** (2000) 37,
URL: https://cds.cern.ch/record/715770 (Cited on page 35).

[127] CERN LHC Experiments Committee, *LHCb: A Large Hadron Collider Beauty Experiment for Precision Measurements of CP Violation and Rare Decays*, LHC Technical Proposal CERN-LHCC-98-004, LHCC-P-4, CERN, 1998,
URL: https://cds.cern.ch/record/622031 (Cited on page 35).

[128] ATLAS Collaboration, *Luminosity determination in* pp *collisions at* $\sqrt{s} = 13\,\mathrm{TeV}$ *using the ATLAS detector at the LHC*, The European Physical Journal C **83** (2023) 982,
URL: https://cds.cern.ch/record/2844887 (Cited on pages 36, 219, 226).

[129] R. P. Feynman, "The behavior of hadron collisions at extreme energies", *Proceedings of the 3rd International Conference on High Energy Collisions of Hadrons*, Stony Brook: Brookhaven National Laboratory, 1969, pp. 237–258 (Cited on page 38).

[130] Yu. L. Dokshitzer, *Calculation of the Structure Functions for Deep Inelastic Scattering and e^+e^- Annihilation by Perturbation Theory in Quantum Chromodynamics*, Soviet Physics JETP **46** (1977) 641 (Cited on page 39).

[131] V. N. Gribov and L. N. Lipatov, *Deep inelastic ep scattering in perturbation theory*, Soviet Journal of Nuclear Physics **15** (1972) 438 (Cited on page 39).

[132] G. Altarelli and G. Parisi, *Asymptotic freedom in parton language*, Nuclear Physics B **126** (1977) 298 (Cited on page 39).

[133] A. Buckley et al., *LHAPDF6: Parton density access in the LHC precision era*, The European Physical Journal C **75** (2015) 132 (Cited on pages 39, 93).

[134] NNPDF Collaboration, *Parton distributions for the LHC Run-2*, Journal of High Energy Physics **2015** (2015) 040 (Cited on pages 39, 93).

[135] G. Sterman and S. Weinberg, *Jets from Quantum Chromodynamics*, Physical Review Letters **39** (1977) 1436 (Cited on pages 40, 154).

[136] G. P. Salam, *Towards jetography*, The European Physical Journal C **67** (2010) 637 (Cited on pages 40, 154).

[137] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, Journal of Instrumentation **3** (2008) S08003,
URL: https://cds.cern.ch/record/1129811 (Cited on pages 42, 62, 64, 65).

[138] ATLAS Collaboration, *The ATLAS experiment at the CERN Large Hadron Collider: a description of the detector configuration for Run-3*, Journal of Instrumentation **19** (2024) P05063,
URL: https://cds.cern.ch/record/2859916 (Cited on pages 42, 45, 49, 58, 69).

[139] ATLAS Collaboration, *The ATLAS Upgrade for the HL-LHC*, Note ATL-UPGRADE-PUB-2025-001, CERN, 2025,
URL: https://cds.cern.ch/record/2928798 (Cited on page 42).

[140] ATLAS Collaboration, *ATLAS Inner Detector, Vol. I*, Technical Design Report CERN-LHCC-97-016, ATLAS-TDR-4, CERN, 1997,
URL: https://cds.cern.ch/record/331063 (Cited on page 44).

[141] ATLAS Collaboration, *ATLAS Inner Detector, Vol. II*, Technical Design Report CERN-LHCC-97-017, ATLAS-TDR-5, CERN, 1997,
URL: https://cds.cern.ch/record/331064 (Cited on page 44).

[142] ATLAS Collaboration, *ATLAS Pixel Detector*, Technical Design Report CERN-LHCC-98-013, ATLAS-TDR-11, CERN, 1998,
URL: https://cds.cern.ch/record/381263 (Cited on page 44).

[143] ATLAS Collaboration, *ATLAS Insertable B-Layer*, Technical Design Report CERN-LHCC-2010-013, ATLAS-TDR-19, CERN, 2010,
URL: https://cds.cern.ch/record/1291633 (Cited on pages 46, 66).

[144] ATLAS Collaboration, *The ATLAS TRT Barrel Detector*, Journal of Instrumentation **3** (2008) P02014,
URL: https://cds.cern.ch/record/1094548 (Cited on page 47).

[145] ATLAS Collaboration, *The ATLAS TRT End-Cap Detectors*, Journal of Instrumentation **3** (2008) P10003,
URL: https://cds.cern.ch/record/1151338 (Cited on page 47).

[146] ATLAS Collaboration, *Performance of the ATLAS Transition Radiation Tracker in Run-1 of the LHC: tracker properties*, Journal of Instrumentation **12** (2017) P05002,
URL: https://cds.cern.ch/record/2253059 (Cited on page 47).

[147] C. W. Fabjan and F. Gianotti, *Calorimetry for Particle Physics*, Reviews of Modern Physics **75** (2003) 1243,
URL: https://cds.cern.ch/record/692252 (Cited on pages 47, 53).

[148] CMS Collaboration, *The CMS Electromagnetic Calorimeter project*, Technical Design Report CERN-LHCC-97-033, CMS-TDR-4, CERN, 1997,
URL: https://cds.cern.ch/record/349375 (Cited on page 48).

[149] ATLAS Collaboration, *ATLAS Calorimeter Performance*, Technical Design Report CERN-LHCC-96-040, ATLAS-TDR-1, CERN, 1996,
URL: https://cds.cern.ch/record/331059 (Cited on page 48).

[150] Y. S. Tsai, *Pair production and bremsstrahlung of charged leptons*, Reviews of Modern Physics **46** (1974) 815 (Cited on page 48).

[151] B. Rossi, *High-energy particles*, Prentice-Hall physics series, New York: Prentice-Hall, 1952,
URL: https://cds.cern.ch/record/99081 (Cited on page 50).

[152] W. R. Nelson, T. M. Jenkins, R. C. McCall and J. K. Cobb, *Electron-Induced Cascade Showers in Copper and Lead at* $1\,\mathrm{GeV}$, Physical Review **149** (1966) 201 (Cited on page 50).

[153] ATLAS Collaboration, *ATLAS Liquid-Argon Calorimeter*, Technical Design Report CERN-LHCC-96-041, ATLAS-TDR-2, CERN, 1996,
URL: https://cds.cern.ch/record/331061 (Cited on page 50).

[154] ATLAS Collaboration, *ATLAS Liquid Argon Calorimeter Phase-I Upgrade*, Technical Design Report CERN-LHCC-2013-017, ATLAS-TDR-022, CERN, 2013,
URL: https://cds.cern.ch/record/1602230 (Cited on pages 50, 51, 67).

[155] ATLAS Collaboration, *ATLAS TDAQ System Phase-I Upgrade*, Technical Design Report CERN-LHCC-2013-018, ATLAS-TDR-023, CERN, 2013,
URL: https://cds.cern.ch/record/1602235 (Cited on pages 51, 64, 65).

[156] T. A. Gabriel, D. E. Groom, P. K. Job, N. V. Mokhov and G. R. Stevenson, *Energy dependence of hadronic activity*, Nuclear Instruments and Methods in Physics Research Section A **338** (1994) 336,
URL: https://cds.cern.ch/record/254124 (Cited on page 53).

[157] The IDEA Study Group, *The IDEA detector concept for FCC-ee*, Report FERMILAB-PUB-25-0189-PPD, 2025,
URL: https://cds.cern.ch/record/2926782 (Cited on page 54).

[158] C. Leroy and P. G. Rancoita, *Physics of cascading shower generation and propagation in matter: principles of high-energy, ultrahigh-energy and compensating calorimetry*, Reports on Progress in Physics **63** (2000) 505 (Cited on page 54).

[159] ATLAS Collaboration, *ATLAS Tile Calorimeter*, Technical Design Report CERN-LHCC-96-042, ATLAS-TDR-3, CERN, 1996,
URL: https://cds.cern.ch/record/331062 (Cited on pages 54, 55).

[160] ATLAS Collaboration, *The ATLAS Forward Calorimeter*, Journal of Instrumentation **3** (2008) P02010,
URL: https://cds.cern.ch/record/1094547 (Cited on page 56).

[161] D. E. Groom, N. V. Mokhov and S. I. Striganov, *Muon stopping power and range tables* $10\,\mathrm{MeV}$ *to* $100\,\mathrm{TeV}$, Atomic Data and Nuclear Data Tables **78** (2001) 183 (Cited on page 57).

[162] ATLAS Collaboration, *ATLAS Muon Spectrometer*, Technical Design Report CERN-LHCC-97-022, ATLAS-TDR-10, CERN, 1997,
URL: https://cds.cern.ch/record/331068 (Cited on page 57).

[163] ATLAS Collaboration, *ATLAS Muon Spectrometer Phase-II Upgrade*, Technical Design Report CERN-LHCC-2017-017, ATLAS-TDR-026, CERN, 2017,
URL: https://cds.cern.ch/record/2285580 (Cited on pages 59, 67).

[164] ATLAS Collaboration, *ATLAS New Small Wheel*, Technical Design Report CERN-LHCC-2013-006, ATLAS-TDR-020, CERN, 2013,
URL: https://cds.cern.ch/record/1552862 (Cited on page 61).

[165] H. Raether, *Die Entwicklung der Elektronenlawine in den Funkenkanal: Nach Beobachtungen in der Nebelkammer*, Zeitschrift für Physik **112** (1939) 464 (Cited on page 61).

[166] ATLAS Collaboration, *ATLAS Magnet System*, Technical Design Report CERN-LHCC-97-018, ATLAS-TDR-6, CERN, 1997,
URL: https://cds.cern.ch/record/338080 (Cited on page 62).

[167] ATLAS Collaboration, *ATLAS Central Solenoid*, Technical Design Report CERN-LHCC-97-021, ATLAS-TDR-9, CERN, 1997,
URL: https://cds.cern.ch/record/331067 (Cited on page 62).

[168] ATLAS Collaboration, *ATLAS Barrel Toroid*, Technical Design Report CERN-LHCC-97-019, ATLAS-TDR-7, CERN, 1997,
URL: https://cds.cern.ch/record/331065 (Cited on page 62).

[169] ATLAS Collaboration, *ATLAS End-Cap Toroids*, Technical Design Report CERN-LHCC-97-020, ATLAS-TDR-8, CERN, 1997,
URL: https://cds.cern.ch/record/331066 (Cited on page 62).

[170] ATLAS Collaboration, *Expected performance of the ATLAS experiment: detector, trigger and physics*, Report CERN-OPEN-2008-020, CERN, 2009,
URL: https://cds.cern.ch/record/1125884 (Cited on page 63).

[171] ATLAS Collaboration, *ATLAS Level-1 Trigger*, Technical Design Report CERN-LHCC-98-014, ATLAS-TDR-12, CERN, 1998,
URL: https://cds.cern.ch/record/381429 (Cited on page 64).

[172] ATLAS Collaboration, *ATLAS Computing*, Technical Design Report CERN-LHCC-2005-022, ATLAS-TDR-17, CERN, 2005,
URL: https://cds.cern.ch/record/837738 (Cited on page 64).

[173] ATLAS Collaboration, *Software and computing for Run-3 of the ATLAS experiment at the LHC*, The European Physical Journal C **85** (2025) 234,
URL: https://cds.cern.ch/record/2895022 (Cited on page 64).

[174] ATLAS Collaboration, *ATLAS High-Level Trigger, Data-acquisition and Controls*, Technical Design Report CERN-LHCC-2003-022, ATLAS-TDR-16, CERN, 2003,
URL: https://cds.cern.ch/record/616089 (Cited on page 65).

[175] CERN, *LHC Computing Grid*, Technical Design Report CERN-LHCC-2005-024, LCG-TDR-001, CERN, 2005,
URL: https://cds.cern.ch/record/840543 (Cited on page 65).

[176] S. van der Meer, *Calibration of the effective beam height in the ISR*, Technical Design Report CERN-ISR-PO-68-31, ISR-PO-68-31, CERN, 1968,
URL: https://cds.cern.ch/record/296752 (Cited on page 65).

[177] ATLAS Collaboration, *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, Journal of Instrumentation **13** (2018) P07017,
URL: https://cds.cern.ch/record/2633501 (Cited on page 65).

[178] ATLAS Collaboration, *The LUCID 3 detector for the ATLAS Phase-II Upgrade*, Technical Design Report CERN-LHCC-2021-016, LHCC-P-018, CERN, 2021,
URL: https://cds.cern.ch/record/2780604 (Cited on pages 66, 67).

[179] ATLAS Collaboration, *The ATLAS Beam Conditions Monitor*, Journal of Instrumentation **3** (2008) P02004,
URL: https://cds.cern.ch/record/1094819 (Cited on page 66).

[180] ATLAS Collaboration, *The ATLAS Diamond Beam Monitor*, Journal of Instrumentation **9** (2014) C02026,
URL: https://cds.cern.ch/record/1630832 (Cited on page 66).

[181] ATLAS Collaboration, *ATLAS Liquid Argon Calorimeter Phase-II Upgrade*, Technical Design Report CERN-LHCC-2017-018, ATLAS-TDR-027, CERN, 2017,
URL: https://cds.cern.ch/record/2285582 (Cited on page 67).

[182] ATLAS Collaboration, *ATLAS Tile Calorimeter Phase-II Upgrade*, Technical Design Report CERN-LHCC-2017-019, ATLAS-TDR-028, CERN, 2017,
URL: https://cds.cern.ch/record/2285583 (Cited on page 67).

[183] ATLAS Collaboration, *ATLAS Trigger and Data Acquisition System Phase-II Upgrade*, Technical Design Report CERN-LHCC-2017-020, ATLAS-TDR-029, CERN, 2017,
URL: https://cds.cern.ch/record/2285584 (Cited on page 67).

[184] ATLAS Collaboration, *ATLAS TDAQ System Phase-II Upgrade – EF Tracking Amendment*, Technical Design Report CERN-LHCC-2022-004, ATLAS-TDR-029-ADD-1, CERN, 2022,
URL: https://cds.cern.ch/record/2802799 (Cited on pages 67, 68).

[185] ATLAS Collaboration, *A High-Granularity Timing Detector for the ATLAS Phase-II Upgrade*, Technical Design Report CERN-LHCC-2020-007, ATLAS-TDR-031, CERN, 2020,
URL: https://cds.cern.ch/record/2719855 (Cited on pages 67, 70).

[186] A. L. Samuel, *Some Studies in Machine Learning Using the Game of Checkers*, IBM Journal of Research and Development **3** (1959) 210 (Cited on page 72).

[187] A. M. Turing, *Computing Machinery and Intelligence*, Mind **LIX** (1950) 433 (Cited on page 72).

[188] W. S. McCulloch and W. Pitts, *A logical calculus of the ideas immanent in nervous activity*, The Bulletin of Mathematical Biophysics **5** (1943) 115 (Cited on page 72).

[189] D. O. Hebb, *The organization of behavior*, New York: Wiley and Sons, 1949 (Cited on page 72).

[190] F. Rosenblatt, *The perceptron: A perceiving and recognizing automaton*, Report 85-460-1, Cornell Aeronautical Laboratory, 1957 (Cited on page 72).

[191] D. E. Rumelhart, G. E. Hinton and R. J. Williams, *Learning representations by back-propagating errors*, Nature **323** (1986) 533 (Cited on pages 72, 77).

[192] K. Hornik, *Approximation capabilities of multilayer feedforward networks*, Neural Networks **4** (1991) 251 (Cited on page 74).

[193] M. Leshno, V. Ya. Lin, A. Pinkus and S. Schocken, *Multilayer feedforward networks with a nonpolynomial activation function can approximate any function*, Neural Networks **6** (1993) 861 (Cited on page 74).

[194] H. Robbins and S. Monro, *A Stochastic Approximation Method*, The Annals of Mathematical Statistics **22** (1951) 400 (Cited on page 75).

[195] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, 2014, arXiv: 1412.6980 [cs.LG],
URL: https://arxiv.org/abs/1412.6980 (Cited on page 76).

[196] B. T. Polyak, *Some methods of speeding up the convergence of iteration methods*, USSR Computational Mathematics and Mathematical Physics **4** (1964) 1 (Cited on page 76).

[197] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks", *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*, Chia: PMLR, 2010, pp. 249–256 (Cited on page 78).

[198] K. Fukushima, *Visual Feature Extraction by a Multilayered Network of Analog Threshold Elements*, IEEE Transactions on Systems Science and Cybernetics **5** (1969) 322 (Cited on page 79).

[199] C. M. Bishop, "Regularization and Complexity Control in Feed-forward Networks", *Proceedings of the International Conference on Artificial Neural Networks*, vol. 1, Paris: EC2 & Cie, 1995, pp. 141–148 (Cited on page 80).

[200] S. J. Hanson and L. Y. Pratt, "Comparing Biases for Minimal Network Construction with Back-Propagation", *Proceedings of the 2nd International Conference on Neural Information Processing Systems*, Denver: NeurIPS, 1988, pp. 177–185 (Cited on page 83).

[201] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever and R. R. Salakhutdinov, *Improving neural networks by preventing co-adaptation of feature detectors*, 2012, arXiv: 1207.0580 [cs.NE],
URL: https://arxiv.org/abs/1207.0580 (Cited on page 83).

[202] S. Ioffe and C. Szegedy, *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*, 2015, arXiv: 1502.03167 [cs.LG],
URL: https://arxiv.org/abs/1502.03167 (Cited on page 83).

[203] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision", *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas: IEEE, 2016, pp. 2818–2826 (Cited on page 83).

[204] K. Fukushima, *Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position*, Biological Cybernetics **36** (1980) 193 (Cited on page 83).

[205] J. L. Elman, *Finding Structure in Time*, Cognitive Science **14** (1990) 179 (Cited on page 83).

[206] A. Vaswani et al., *Attention Is All You Need*, 2017, arXiv: 1706.03762 [cs.CL],
URL: https://arxiv.org/abs/1706.03762 (Cited on page 83).

[207] D. P. Kingma and M. Welling, *An Introduction to Variational Autoencoders*, Foundations and Trends in Machine Learning **12** (2019) 307 (Cited on page 83).

[208] I. J. Goodfellow et al., *Generative Adversarial Networks*, 2014, arXiv: 1406.2661 [stat.ML],
URL: https://arxiv.org/abs/1406.2661 (Cited on page 83).

[209] M. Gori, G. Monfardini and F. Scarselli, "A new model for learning in graph domains", *Proceedings of the 2005 IEEE International Joint Conference on Neural Networks*, vol. 2, Montreal: IEEE, 2005, pp. 729–734 (Cited on page 83).

[210] S. J. D. Prince, *Understanding Deep Learning*, The MIT Press, 2023,
URL: http://udlbook.com (Cited on page 84).

[211] B. Denby, *Neural networks and cellular automata in experimental high energy physics*, Computer Physics Communications **49** (1988) 429 (Cited on page 84).

[212] C. Peterson, *Track finding with neural networks*, Nuclear Instruments and Methods in Physics Research Section A **279** (1989) 537 (Cited on page 84).

[213] DELPHI Collaboration, *Classification of the hadronic decays of the Z^0 into b and c quark pairs using a neural network*, Physics Letters B **295** (1992) 383 (Cited on page 84).

[214] J. K. Köhne et al., *Realization of a second level neural network trigger for the H1 experiment at HERA*, Nuclear Instruments and Methods in Physics Research Section A **389** (1997) 128 (Cited on page 84).

[215] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman and A. Schwartzman, *Jet-images – deep learning edition*, Journal of High Energy Physics **2016** (2016) 069 (Cited on page 84).

[216] S. Farrell et al., *Novel deep learning methods for track reconstruction*, 2018, arXiv: 1810.06111 [hep-ex],
URL: https://arxiv.org/abs/1810.06111 (Cited on page 84).

[217] Keras Developers, *Keras*, 2015,
URL: https://keras.io (Cited on page 84).

[218] TensorFlow Developers, *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, 2015,
URL: https://www.tensorflow.org/ (Cited on page 84).

[219] NumPy Developers, *Array programming with NumPy*, Nature **585** (2020) 357 (Cited on page 84).

[220] Matplotlib Developers, *Matplotlib: A 2D graphics environment*, Computing in Science & Engineering **9** (2007) 90 (Cited on page 84).

[221] L. Breiman, J. H. Friedman, R. A. Olshen and C. J. Stone, *Classification and Regression Trees*, Monterey: Wadsworth and Brooks/Cole Advanced Books & Software, 1984 (Cited on page 84).

[222] CDF Collaboration, *Observation of Electroweak Single Top-Quark Production*, Physical Review Letters **103** (2009) 092002 (Cited on page 86).

[223] DØ Collaboration, *Observation of Single Top-Quark Production*, Physical Review Letters **103** (2009) 092001 (Cited on page 86).

[224] A. Hoecker et al., *TMVA – Toolkit for Multivariate Data Analysis with ROOT*, Technical Report CERN-OPEN-2007-007, CERN, 2007,
URL: https://cds.cern.ch/record/1019880 (Cited on pages 86, 112).

[225] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, Nuclear Instruments and Methods in Physics Research Section A **389** (1997) 81 (Cited on page 86).

[226] E. W. Forgy, *Cluster analysis of multivariate data: efficiency versus interpretability of classifications*, Biometrics **21** (1965) 768 (Cited on page 86).

[227] S. Lloyd, *Least Squares Quantization in PCM*, IEEE Transactions on Information Theory **28** (1982) 129 (Cited on page 86).

[228] Scikit-learn Developers, *Scikit-learn: Machine Learning in Python*, Journal of Machine Learning Research **12** (2011) 2825 (Cited on pages 87, 91).

[229] G. Voronoi, *Nouvelles applications des paramètres continus à la théorie des formes quadratiques, Deuxième mémoire: Recherches sur les parallélloèdres primitifs*, Journal für die reine und angewandte Mathematik (Crelles Journal) **134** (1908) 198 (Cited on page 87).

[230] G. Voronoi, *Nouvelles applications des paramètres continus à la théorie des formes quadratiques, Premier mémoire: Sur quelques propriétés des formes quadratiques positives parfaites*, Journal für die reine und angewandte Mathematik (Crelles Journal) **133** (1908) 97 (Cited on page 87).

[231] SciPy Developers, *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python*, Nature Methods **17** (2020) 261 (Cited on page 87).

[232] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions", *Advances in Neural Information Processing Systems 30*, Long Beach: NeurIPS, 2017, pp. 4765–4774 (Cited on page 88).

[233] L. Breiman, *Random Forests*, Machine Learning **45** (2001) 5 (Cited on page 88).

[234] K. Simonyan, A. Vedaldi and A. Zisserman, "Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps", *Proceedings of the 2nd International Conference on Learning Representations, Workshop Track*, Banff: ICLR, 2014 (Cited on page 89).

[235] J. Bergstra and Y. Bengio, *Random Search for Hyper-Parameter Optimization*, Journal of Machine Learning Research **13** (2012) 281 (Cited on page 89).

[236] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams and N. de Freitas, *Taking the Human Out of the Loop: A Review of Bayesian Optimization*, Proceedings of the IEEE **104** (2016) 148 (Cited on page 89).

[237] J. Močkus, "On Bayesian Methods for Seeking the Extremum", *Optimization Techniques IFIP Technical Conference*, Novosibirsk: Springer, 1975, pp. 400–404 (Cited on page 89).

[238] J. Močkus, V. Tiešis and A. Žilinskas, "The Application of Bayesian Methods for Seeking the Extremum", *Towards Global Optimization 2*, Amsterdam: North-Holland, 1978, pp. 117–129 (Cited on page 89).

[239] D. R. Jones, M. Schonlau and W. J. Welch, *Efficient Global Optimization of Expensive Black-Box Functions*, Journal of Global Optimization **13** (1998) 455 (Cited on page 89).

[240] H. J. Kushner, *A New Method of Locating the Maximum Point of an Arbitrary Multipeak Curve in the Presence of Noise*, Journal of Basic Engineering **86** (1964) 97 (Cited on page 90).

[241] N. Srinivas, A. Krause, S. M. Kakade and M. Seeger, "Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design", *Proceedings of the 27th International Conference on Machine Learning (ICML)*, Haifa: Omnipress, 2010, pp. 1015–1022 (Cited on page 90).

[242] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, Cambridge: MIT Press, 2006 (Cited on page 90).

[243] J. C. Collins, D. E. Soper and G. Sterman, "Factorization of Hard Processes in QCD", *Perturbative Quantum Chromodynamics*, ed. by A. H. Mueller, vol. 5, Advanced Series on Directions in High Energy Physics, World Scientific, 1989, pp. 1–91 (Cited on page 93).

[244] G. T. Bodwin, *Factorization of the Drell–Yan cross section in perturbation theory*, Physical Review D **31** (1985) 2616 (Cited on page 93).

[245] NNPDF Collaboration, *The path to proton structure at $1\,\%$ accuracy*, The European Physical Journal C **82** (2022) 428 (Cited on page 93).

[246] The PDF4LHC Working Group, *The PDF4LHC21 combination of global PDF fits for the LHC Run III*, Journal of Physics G **49** (2022) 080501 (Cited on page 93).

[247] R. D. Ball et al., *Parton distributions with LHC data*, Nuclear Physics B **867** (2013) 244 (Cited on page 93).

[248] S. Catani and M. H. Seymour, *A general algorithm for calculating jet cross sections in NLO QCD*, Nuclear Physics B **485** (1997) 291 (Cited on page 94).

[249] S. Frixione, Z. Kunszt and A. Signer, *Three-jet cross sections to next-to-leading order*, Nuclear Physics B **467** (1996) 399 (Cited on page 95).

[250] R. Frederix, S. Frixione, F. Maltoni and T. Stelzer, *Automation of next-to-leading order computations in QCD: the FKS subtraction*, Journal of High Energy Physics **2009** (2009) 003 (Cited on pages 95, 96).

[251] S. Frixione and B. R. Webber, *Matching NLO QCD computations and parton shower simulations*, Journal of High Energy Physics **2002** (2002) 029 (Cited on pages 95, 96).

[252] S. Höche, F. Krauss, M. Schönherr and F. Siegert, *A critical appraisal of NLO+PS matching methods*, Journal of High Energy Physics **2012** (2012) 049 (Cited on pages 95, 293).

[253] S. Höche, F. Krauss, M. Schönherr and F. Siegert, *QCD matrix elements + parton showers. The NLO case*, Journal of High Energy Physics **2013** (2013) 027 (Cited on pages 95, 96, 293).

[254] S. Catani, F. Krauss, R. Kuhn and B. R. Webber, *QCD Matrix Elements + Parton Showers*, Journal of High Energy Physics **2001** (2001) 063 (Cited on pages 95, 96, 183, 293, 294).

[255] S. Höche, F. Krauss, S. Schumann and F. Siegert, *QCD matrix elements and truncated showers*, Journal of High Energy Physics **2009** (2009) 053 (Cited on pages 95, 96, 183, 293, 294).

[256] B. Andersson, G. Gustafson, G. Ingelman and T. Sjöstrand, *Parton fragmentation and string dynamics*, Physics Reports **97** (1983) 31 (Cited on page 95).

[257] B. R. Webber, *A QCD model for jet fragmentation including soft gluon interference*, Nuclear Physics B **238** (1984) 492 (Cited on page 95).

[258] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, Journal of High Energy Physics **2014** (2014) 079 (Cited on page 96).

[259] V. Hirschi and O. Mattelaer, *Automated event generation for loop-induced processes*, Journal of High Energy Physics **2015** (2015) 146 (Cited on page 96).

[260] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, Journal of High Energy Physics **2010** (2010) 043 (Cited on page 96).

[261] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, Journal of High Energy Physics **2007** (2007) 070 (Cited on page 96).

[262] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, Computer Physics Communications **191** (2015) 159 (Cited on page 96).

[263] C. Bierlich et al., *A comprehensive guide to the physics and usage of PYTHIA 8.3*, SciPost Physics Codebases **8** (2022) 8 (Cited on page 96).

[264] P. Artoisenet, R. Frederix, O. Mattelaer and R. Rietkerk, *Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations*, Journal of High Energy Physics **2013** (2013) 015 (Cited on page 96).

[265] P. Golonka and Z. Wąs, *PHOTOS Monte Carlo: a precision tool for QED corrections in* Z *and* W *decays*, The European Physical Journal C **45** (2006) 97 (Cited on page 96).

[266] E. Bothmann et al., *Event generation with Sherpa 2.2*, SciPost Physics **7** (2019) 034 (Cited on page 96).

[267] E. Bothmann et al., *Event generation with Sherpa 3*, Journal of High Energy Physics **2024** (2024) 156 (Cited on page 96).

[268] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, Journal of High Energy Physics **2008** (2008) 039 (Cited on page 96).

[269] S. Schumann and F. Krauss, *A parton shower algorithm based on Catani–Seymour dipole factorisation*, Journal of High Energy Physics **2008** (2008) 038 (Cited on page 96).

[270] F. Cascioli, P. Maierhöfer and S. Pozzorini, *Scattering Amplitudes with Open Loops*, Physical Review Letters **108** (2012) 111601 (Cited on page 96).

[271] S. Actis et al., *RECOLA – REcursive Computation of One-Loop Amplitudes*, Computer Physics Communications **214** (2017) 140 (Cited on pages 96, 186).

[272] S. Agostinelli et al., *Geant4 – a simulation toolkit*, Nuclear Instruments and Methods in Physics Research Section A **506** (2003) 250 (Cited on page 96).

[273] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, The European Physical Journal C **70** (2010) 823,
URL: https://cds.cern.ch/record/1267853 (Cited on page 96).

[274] D. J. Lange, *The EvtGen particle decay simulation package*, Nuclear Instruments and Methods in Physics Research Section A **462** (2001) 152 (Cited on pages 96, 292).

[275] ATLAS Collaboration, *Electron and photon reconstruction and performance in ATLAS using a dynamical, topological cell clustering-based approach*, Note ATL-PHYS-PUB-2017-022, CERN, 2017,
URL: https://cds.cern.ch/record/2298955 (Cited on page 98).

[276] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at* $\sqrt{s} = 13\,\mathrm{TeV}$, The European Physical Journal C **79** (2019) 639,
URL: https://cds.cern.ch/record/2657964 (Cited on page 99).

[277] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data*, Journal of Instrumentation **14** (2019) P12006,
URL: https://cds.cern.ch/record/2684552 (Cited on page 99).

[278] ATLAS Collaboration, *Electron and photon efficiencies in LHC Run-2 with the ATLAS experiment*, Journal of High Energy Physics **2024** (2024) 162,
URL: https://cds.cern.ch/record/2868520 (Cited on pages 99, 131, 218).

[279] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run-1*, The European Physical Journal C **77** (2017) 490,
URL: https://cds.cern.ch/record/2138166 (Cited on page 99).

[280] ATLAS Collaboration, *Calorimeter Clustering Algorithms: Description and Performance*, Note ATL-LARG-PUB-2008-002, CERN, 2008, URL: https://cds.cern.ch/record/1099735 (Cited on page 100).

[281] ATLAS Collaboration, *The New ATLAS track reconstruction (NEWT)*, Journal of Physics: Conference Series **119** (2008) 032014, URL: https://cds.cern.ch/record/1176900 (Cited on page 101).

[282] ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run-2*, The European Physical Journal C **77** (2017) 673, URL: https://cds.cern.ch/record/2261156 (Cited on page 101).

[283] C. Runge, *Über die numerische Auflösung von Differentialgleichungen*, Mathematische Annalen **46** (1895) 167 (Cited on page 101).

[284] W. Kutta, *Beitrag zur näherungsweisen Integration totaler Differentialgleichungen*, Zeitschrift für Mathematik und Physik **46** (1901) 435 (Cited on page 101).

[285] P. V. C. Hough, "Method and Means for Recognizing Complex Patterns", US 3,069,654, 1962 (Cited on page 101).

[286] R. E. Kalman, *A New Approach to Linear Filtering and Prediction Problems*, Journal of Basic Engineering **82** (1960) 35 (Cited on page 101).

[287] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*, Nuclear Instruments and Methods in Physics Research Section A **262** (1987) 444 (Cited on page 101).

[288] ATLAS Collaboration, *The Global χ^2 track fitter in ATLAS*, Journal of Physics: Conference Series **119** (2008) 032013, URL: https://cds.cern.ch/record/1176901 (Cited on page 102).

[289] H. Bethe and W. Heitler, *On the stopping of fast particles and on the creation of positive electrons*, Proceedings of the Royal Society of London. Series A **146** (1934) 83 (Cited on page 102).

[290] ATLAS Collaboration, *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung*, Conference Note ATLAS-CONF-2012-047, CERN, 2012, URL: https://cds.cern.ch/record/1449796 (Cited on page 102).

[291] ATLAS Collaboration, *Reconstruction of primary vertices at the ATLAS experiment in Run-1 proton–proton collisions at the LHC*, The European Physical Journal C **77** (2017) 332, URL: https://cds.cern.ch/record/2235651 (Cited on page 103).

[292] W. Waltenberger, R. Frühwirth and P. Vanlaer, *Adaptive vertex fitting*, Journal of Physics G **34** (2007) 343 (Cited on page 103).

[293] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using LHC Run-2 data*, Journal of Instrumentation **19** (2024) P02009, URL: https://cds.cern.ch/record/2870086 (Cited on pages 107, 218).

[294] W. E. Cleland and E. G. Stern, *Signal processing considerations for liquid ionization calorimeters in a high rate environment*, Nuclear Instruments and Methods in Physics Research Section A **338** (1994) 467 (Cited on page 107).

[295] ATLAS Collaboration, *Identification of electrons using a deep neural network in the ATLAS experiment*, Note ATL-PHYS-PUB-2022-022, CERN, 2022, URL: https://cds.cern.ch/record/2809283 (Cited on page 117).

[296] M. J. Oreglia, *A study of the reactions $\psi' \to \gamma\gamma\psi$*, PhD thesis: Stanford University, 1980 (Cited on page 118).

[297] S. Catani, Yu. L. Dokshitzer, M. H. Seymour and B. R. Webber, *Longitudinally-invariant $k_\perp$-clustering algorithms for hadron–hadron collisions*, Nuclear Physics B **406** (1993) 187 (Cited on pages 118, 157).

[298] M. Cacciari, G. P. Salam and G. Soyez, *FastJet user manual*, The European Physical Journal C **72** (2012) 1896 (Cited on pages 118, 157).

[299] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas*, Physics Letters B **659** (2008) 119 (Cited on pages 118, 160).

[300] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, The European Physical Journal C **77** (2017) 466, URL: https://cds.cern.ch/record/2257597 (Cited on pages 119, 152, 155, 157).

[301] ATLAS Collaboration, *Evidence for the associated production of the Higgs boson and a top quark pair with the ATLAS detector*, Physical Review D **97** (2018) 072003, URL: https://cds.cern.ch/record/2299050 (Cited on pages 120, 152).

[302] ATLAS Collaboration, *Search for flavour-changing neutral current interactions of the top quark and the Higgs boson in events with a pair of τ-leptons in pp collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ with the ATLAS detector*, Journal of High Energy Physics **2023** (2023) 155, URL: https://cds.cern.ch/record/2825300 (Cited on pages 120, 152).

[303] ATLAS Collaboration, *Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run-1 data*, The European Physical Journal C **76** (2016) 666, URL: https://cds.cern.ch/record/2158117 (Cited on page 120).

[304] ATLAS Collaboration, *Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run-2 data collected in 2015 and 2016*, The European Physical Journal C **79** (2019) 205, URL: https://cds.cern.ch/record/2643391 (Cited on page 120).

[305] ATLAS Collaboration, *Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton–proton collision data*, The European Physical Journal C **74** (2014) 2941, URL: https://cds.cern.ch/record/1694142 (Cited on page 121).

[306] G. Breit and E. P. Wigner, *Capture of slow neutrons*, Physical Review **49** (1936) 519 (Cited on page 124).

[307] L. D. Landau, *On the Energy Loss of Fast Particles by Ionization*, Journal of Physics (USSR) **8** (1944) 201 (Cited on page 124).

[308] T. Dozat, "Incorporating Nesterov momentum into Adam", *Proceedings of the 4th International Conference on Learning Representations, Workshop Track*, San Juan: ICLR, 2016 (Cited on page 128).

[309] Y. E. Nesterov, *A method for solving the convex programming problem with convergence rate $\mathcal{O}(1/k^2)$*, Soviet Mathematics Doklady **27** (1983) 372 (Cited on page 128).

[310] ATLAS Collaboration, *Athena*, 2019, DOI: 10.5281/zenodo.2641997 (Cited on page 138).

[311] R. Franceschini, G. Panico, A. Pomarol, F. Riva and A. Wulzer, *Electroweak precision tests in high-energy diboson processes*, Journal of High Energy Physics **2018** (2018) 111 (Cited on page 144).

[312] OPAL Collaboration, *Measurement of the mass of the W boson in $\mathrm{e^+e^-}$ collisions at $\sqrt{s} = 161\,\mathrm{GeV}$*, Physics Letters B **389** (1996) 416, URL: https://cds.cern.ch/record/313688 (Cited on pages 144, 146).

[313] ALEPH Collaboration, *Measurement of the $\mathrm{e^+e^-} \to \mathrm{ZZ}$ production cross section at centre-of-mass energies of $183\,\mathrm{GeV}$ and $189\,\mathrm{GeV}$*, Physics Letters B **469** (1999) 287, URL: https://cds.cern.ch/record/404700 (Cited on pages 144, 146).

[314] CDF Collaboration, *Observation of WZ production*, Physical Review Letters **98** (2007) 161801 (Cited on pages 144, 146).

[315] DØ Collaboration, *Measurement of the $\mathrm{p\bar{p}} \to \mathrm{WZ} + X$ cross section at $\sqrt{s} = 1.96\,\mathrm{TeV}$ and limits on WWZ trilinear gauge couplings*, Physical Review D **76** (2007) 111104 (Cited on pages 144, 146).

[316] DØ Collaboration, *Measurement of the WW production cross section in $\mathrm{p\bar{p}}$ collisions at $\sqrt{s} = 1.96\,\mathrm{TeV}$*, Physical Review Letters **94** (2005) 151801 (Cited on pages 144, 146).

[317] CDF Collaboration, *Measurement of the $\mathrm{W^+W^-}$ production cross section in $\mathrm{p\bar{p}}$ collisions at $\sqrt{s} = 1.96\,\mathrm{TeV}$ using dilepton events*, Physical Review Letters **94** (2005) 211801 (Cited on pages 144, 146).

[318] CDF Collaboration, *Strong evidence for ZZ production in $\mathrm{p\bar{p}}$ collisions at $\sqrt{s} = 1.96\,\mathrm{TeV}$*, Physical Review Letters **100** (2008) 201801 (Cited on pages 144, 146).

[319] DØ Collaboration, *Observation of ZZ production in $\mathrm{p\bar{p}}$ collisions at $\sqrt{s} = 1.96\,\mathrm{TeV}$*, Physical Review Letters **101** (2008) 171803 (Cited on pages 144, 146).

[320] CMS Collaboration, *Measurement of $\mathrm{W^+W^-}$ production and search for the Higgs boson in pp collisions at $\sqrt{s} = 7\,\mathrm{TeV}$*, Physics Letters B **699** (2011) 25, URL: https://cds.cern.ch/record/1332217 (Cited on pages 144, 146).

[321] ATLAS Collaboration, *Measurement of the* WW *cross section in* $\sqrt{s} = 7\,\mathrm{TeV}$ pp *collisions with the ATLAS detector and limits on anomalous gauge couplings*, Physics Letters B **712** (2012) 289, URL: https://cds.cern.ch/record/1435199 (Cited on pages 144, 146).

[322] CMS Collaboration, *Measurement of the* $\mathrm{W^+W^-}$ *cross section in* pp *collisions at* $\sqrt{s} = 8\,\mathrm{TeV}$ *and limits on anomalous gauge couplings*, The European Physical Journal C **76** (2016) 401, URL: https://cds.cern.ch/record/2033478 (Cited on pages 144, 146).

[323] ATLAS Collaboration, *Measurement of total and differential* $\mathrm{W^+W^-}$ *production cross sections in proton–proton collisions at* $\sqrt{s} = 8\,\mathrm{TeV}$ *with the ATLAS detector and limits on anomalous triple-gauge-boson couplings*, Journal of High Energy Physics **2016** (2016) 029, URL: https://cds.cern.ch/record/2136855 (Cited on pages 144, 146).

[324] ATLAS Collaboration, *Measurement of the* $\mathrm{W^\pm Z}$ *production cross section and limits on anomalous triple gauge couplings in proton–proton collisions at* $\sqrt{s} = 7\,\mathrm{TeV}$ *with the ATLAS detector*, Physics Letters B **709** (2012) 341, URL: https://cds.cern.ch/record/1401254 (Cited on pages 144, 146).

[325] ATLAS Collaboration, *Measurements of* $\mathrm{W^\pm Z}$ *production cross sections in* pp *collisions at* $\sqrt{s} = 8\,\mathrm{TeV}$ *with the ATLAS detector and limits on anomalous gauge boson self-couplings*, Physical Review D **93** (2016) 092004, URL: https://cds.cern.ch/record/2137188 (Cited on pages 144, 146, 188, 189, 284).

[326] CMS Collaboration, *Measurement of the* WZ *production cross section in* pp *collisions at* $\sqrt{s} = 7$ *and* $8\,\mathrm{TeV}$ *and search for anomalous triple gauge couplings at* $\sqrt{s} = 8\,\mathrm{TeV}$, The European Physical Journal C **77** (2017) 236, URL: https://cds.cern.ch/record/2216557 (Cited on pages 144, 146).

[327] ATLAS Collaboration, *Measurement of* ZZ *production in* pp *collisions at* $\sqrt{s} = 7\,\mathrm{TeV}$ *and limits on anomalous* ZZZ *and* $\mathrm{ZZ}\gamma$ *couplings with the ATLAS detector*, Journal of High Energy Physics **2013** (2013) 128, URL: https://cds.cern.ch/record/1496097 (Cited on pages 144, 146).

[328] CMS Collaboration, *Measurement of the* ZZ *production cross section and search for anomalous couplings in* $2\ell 2\ell'$ *final states in* pp *collisions at* $\sqrt{s} = 7\,\mathrm{TeV}$, Journal of High Energy Physics **2013** (2013) 063, URL: https://cds.cern.ch/record/1495152 (Cited on pages 144, 146).

[329] ATLAS Collaboration, *Measurement of the* ZZ *production cross section in proton–proton collisions at* $\sqrt{s} = 8\,\mathrm{TeV}$ *using the* $\mathrm{ZZ} \to \ell^-\ell^+\ell'^-\ell'^+$ *and* $\mathrm{ZZ} \to \ell^-\ell^+\nu\bar{\nu}$ *decay channels with the ATLAS detector*, Journal of High Energy Physics **2017** (2017) 099, URL: https://cds.cern.ch/record/2227023 (Cited on pages 144, 146).

[330] CMS Collaboration, *Measurement of the* $\mathrm{pp} \to \mathrm{ZZ}$ *production cross section and constraints on anomalous triple gauge couplings in four-lepton final states at* $\sqrt{s} = 8\,\mathrm{TeV}$, Physics Letters B **740** (2015) 250, URL: https://cds.cern.ch/record/1706050 (Cited on pages 144, 146).

[331] ATLAS Collaboration, *Measurement of the* $\mathrm{W^\pm Z}$ *boson pair-production cross section in* pp *collisions at* $\sqrt{s} = 13\,\mathrm{TeV}$ *with the ATLAS detector*, Physics Letters B **762** (2016) 1, URL: https://cds.cern.ch/record/2160353 (Cited on pages 144, 146, 188).

[332] CMS Collaboration, *Measurement of the* WZ *production cross section in* pp *collisions at* $\sqrt{s} = 13\,\mathrm{TeV}$, Physics Letters B **766** (2017) 268, URL: https://cds.cern.ch/record/2201221 (Cited on pages 144, 146).

[333] ATLAS Collaboration, *Measurement of the* ZZ *production cross section in* pp *collisions at* $\sqrt{s} = 13\,\mathrm{TeV}$ *with the ATLAS detector*, Physical Review Letters **116** (2016) 101801, URL: https://cds.cern.ch/record/2115496 (Cited on pages 144, 146).

[334] CMS Collaboration, *Measurements of the* $\mathrm{pp} \to \mathrm{ZZ}$ *production cross section and the* $\mathrm{Z} \to 4\ell$ *branching fraction, and constraints on anomalous triple gauge couplings at* $\sqrt{s} = 13\,\mathrm{TeV}$, The European Physical Journal C **78** (2018) 165, URL: https://cds.cern.ch/record/2285803 (Cited on pages 144, 146).

[335] CMS Collaboration, $\mathrm{W^+W^-}$ *boson pair production in proton–proton collisions at* $\sqrt{s} = 13\,\mathrm{TeV}$, Physical Review D **102** (2020) 092001, URL: https://cds.cern.ch/record/2729339 (Cited on pages 144, 146).

[336] ATLAS Collaboration, *Measurements of* $\mathrm{W^+W^-}$ *production cross-sections in* pp *collisions at* $\sqrt{s} = 13\,\mathrm{TeV}$ *with the ATLAS detector*, Journal of High Energy Physics **2025** (2025) 142, URL: https://cds.cern.ch/record/2932300 (Cited on pages 144, 146).

[337] CMS Collaboration, *Measurements of the* $\mathrm{pp} \to \mathrm{WZ}$ *inclusive and differential production cross section and constraints on charged anomalous triple gauge couplings at* $\sqrt{s} = 13\,\mathrm{TeV}$, Journal of High Energy Physics **2019** (2019) 122,
URL: https://cds.cern.ch/record/2653379 (Cited on pages 144, 146).

[338] ATLAS Collaboration, *Measurements and interpretations of* $\mathrm{W^{\pm}Z}$ *production cross-sections in* pp *collisions at* $\sqrt{s} = 13\,\mathrm{TeV}$ *with the ATLAS detector*, Journal of High Energy Physics **2025** (2025) 006,
URL: https://cds.cern.ch/record/2937475 (Cited on pages 144, 146, 188, 189, 284).

[339] ATLAS Collaboration, *Measurements of differential cross-sections in four-lepton events in* $13\,\mathrm{TeV}$ *proton–proton collisions with the ATLAS detector*, Journal of High Energy Physics **2021** (2021) 005,
URL: https://cds.cern.ch/record/2753518 (Cited on pages 144, 146, 286, 290).

[340] CMS Collaboration, *Measurements of* $\mathrm{pp} \to \mathrm{ZZ}$ *production cross sections and constraints on anomalous triple gauge couplings at* $\sqrt{s} = 13\,\mathrm{TeV}$, The European Physical Journal C **81** (2021) 200,
URL: https://cds.cern.ch/record/2729607 (Cited on pages 144, 146).

[341] CMS Collaboration, *Measurement of inclusive and differential cross sections for* $\mathrm{W^{+}W^{-}}$ *production in proton–proton collisions at* $\sqrt{s} = 13.6\,\mathrm{TeV}$, Physics Letters B **861** (2025) 139231,
URL: https://cds.cern.ch/record/2900127 (Cited on pages 144, 146).

[342] CMS Collaboration, *Measurement of the inclusive* WZ *production cross section in* pp *collisions at* $\sqrt{s} = 13.6\,\mathrm{TeV}$, Journal of High Energy Physics **2025** (2025) 115,
URL: https://cds.cern.ch/record/2918719 (Cited on pages 144, 146).

[343] ATLAS Collaboration, *Measurement of the* ZZ *production cross section in* pp *collisions at* $\sqrt{s} = 13.6\,\mathrm{TeV}$ *with the ATLAS detector*, Physics Letters B **855** (2024) 138764,
URL: https://cds.cern.ch/record/2880115 (Cited on pages 144, 146).

[344] CMS Collaboration, *Measurement of the inclusive and differential* WZ *production cross sections, polarization angles, and triple gauge couplings in* pp *collisions at* $\sqrt{s} = 13\,\mathrm{TeV}$, Journal of High Energy Physics **2022** (2022) 032,
URL: https://cds.cern.ch/record/2786853 (Cited on pages 145, 146).

[345] L. Selem, *Measurement of gauge boson joint-polarisation states in* WZ *production at the LHC with the ATLAS detector*, PhD thesis: Université Savoie Mont Blanc, 2022, DOI: 10.70675/4e747725ze59ez4e92zb144z12707afbbc8e (Cited on pages 145, 166).

[346] ATLAS Collaboration, *Evidence of pair production of longitudinally polarised vector bosons and study of CP properties in* $\mathrm{ZZ} \to 4\ell$ *events with the ATLAS detector at* $\sqrt{s} = 13\,\mathrm{TeV}$, Journal of High Energy Physics **2023** (2023) 107,
URL: https://cds.cern.ch/record/2874981 (Cited on pages 145, 146).

[347] ATLAS Collaboration, *Studies of the energy dependence of diboson polarization fractions and the radiation-amplitude-zero effect in* WZ *production with the ATLAS detector*, Physical Review Letters **133** (2024) 101802,
URL: https://cds.cern.ch/record/2890051 (Cited on pages 145, 146, 167).

[348] ATLAS Collaboration, *Evidence for electroweak production of* $\mathrm{W^{\pm}W^{\pm}jj}$ *in* pp *collisions at* $\sqrt{s} = 8\,\mathrm{TeV}$ *with the ATLAS detector*, Physical Review Letters **113** (2014) 141803,
URL: https://cds.cern.ch/record/1703745 (Cited on pages 145, 147).

[349] CMS Collaboration, *Study of vector boson scattering and search for new physics in events with two same-sign leptons and two jets*, Physical Review Letters **114** (2015) 051801,
URL: https://cds.cern.ch/record/1957133 (Cited on pages 145, 147).

[350] CMS Collaboration, *Observation of electroweak production of same-sign* W *boson pairs in the two jet and two same-sign lepton final state in proton–proton collisions at* $\sqrt{s} = 13\,\mathrm{TeV}$, Physical Review Letters **120** (2018) 081801,
URL: https://cds.cern.ch/record/2284582 (Cited on pages 145, 147).

[351] ATLAS Collaboration, *Observation of electroweak production of a same-sign* W *boson pair in association with two jets in* pp *collisions at* $\sqrt{s} = 13\,\mathrm{TeV}$ *with the ATLAS detector*, Physical Review Letters **123** (2019) 161801,
URL: https://cds.cern.ch/record/2677868 (Cited on pages 145, 147).

[352] CMS Collaboration, *Measurements of production cross sections of polarized same-sign* W *boson pairs in association with two jets in proton–proton collisions at* $\sqrt{s} = 13\,\mathrm{TeV}$, Physics Letters B **812** (2021) 136018,
URL: https://cds.cern.ch/record/2730856 (Cited on pages 145, 147, 252).

[353] ATLAS Collaboration, *Measurement and interpretation of same-sign* W *boson pair production in association with two jets in* pp *collisions at* $\sqrt{s} = 13\,\mathrm{TeV}$ *with the ATLAS detector*, Journal of High Energy Physics **2024** (2024) 026,
URL: https://cds.cern.ch/record/2882929 (Cited on pages 145, 147, 236).

[354] ATLAS Collaboration, *Evidence for longitudinally polarized* W *bosons in the electroweak production of same-sign* W *boson pairs in association with two jets in* pp *collisions at* $\sqrt{s} = 13\,\mathrm{TeV}$ *with the ATLAS detector*, Physical Review Letters **135** (2025) 111802,
URL: https://cds.cern.ch/record/2926878 (Cited on pages 145, 147, 252).

[355] CMS Collaboration, *Observation of electroweak* $\mathrm{W^+W^-}$ *pair production in association with two jets in proton–proton collisions at* $\sqrt{s} = 13\,\mathrm{TeV}$, Physics Letters B **841** (2023) 137495,
URL: https://cds.cern.ch/record/2809311 (Cited on pages 145, 147).

[356] ATLAS Collaboration, *Observation of electroweak production of* $\mathrm{W^+W^-}$ *in association with jets in proton–proton collisions at* $\sqrt{s} = 13\,\mathrm{TeV}$ *with the ATLAS detector*, Journal of High Energy Physics **2024** (2024) 254,
URL: https://cds.cern.ch/record/2891254 (Cited on pages 145, 147).

[357] ATLAS Collaboration, *Observation of electroweak* $\mathrm{W^\pm Z}$ *boson pair production in association with two jets in* pp *collisions at* $\sqrt{s} = 13\,\mathrm{TeV}$ *with the ATLAS detector*, Physics Letters B **793** (2019) 469,
URL: https://cds.cern.ch/record/2652670 (Cited on pages 145, 147).

[358] CMS Collaboration, *Measurement of electroweak* WZ *boson production and search for new physics in* WZ *+ two jets events in* pp *collisions at* $\sqrt{s} = 13\,\mathrm{TeV}$, Physics Letters B **795** (2019) 281,
URL: https://cds.cern.ch/record/2653586 (Cited on pages 145, 147).

[359] CMS Collaboration, *Measurements of production cross sections of* WZ *and same-sign* WW *boson pairs in association with two jets in proton–proton collisions at* $\sqrt{s} = 13\,\mathrm{TeV}$, Physics Letters B **809** (2020) 135710,
URL: https://cds.cern.ch/record/2716981 (Cited on pages 145, 147).

[360] ATLAS Collaboration, *Measurements of electroweak* $\mathrm{W^\pm Z}$ *boson pair production in association with two jets in* pp *collisions at* $\sqrt{s} = 13\,\mathrm{TeV}$ *with the ATLAS detector*, Journal of High Energy Physics **2024** (2024) 192,
URL: https://cds.cern.ch/record/2892680 (Cited on pages 145, 147, 148, 188, 201, 208, 214, 236, 245, 247, 255, 288, 306).

[361] CMS Collaboration, *First measurements of vector boson scattering in* $\mathrm{W^\pm W^\pm}$ *and* WZ *production in all-leptonic final states at* $\sqrt{s} = 13.6\,\mathrm{TeV}$, Preprint CMS-SMP-25-013, CERN, 2026, arXiv: 2605.15396 [hep-ex],
URL: https://cds.cern.ch/record/2962072 (Cited on pages 145, 147).

[362] CMS Collaboration, *Evidence for electroweak production of four charged leptons and two jets in proton–proton collisions at* $\sqrt{s} = 13\,\mathrm{TeV}$, Physics Letters B **812** (2021) 135992,
URL: https://cds.cern.ch/record/2728134 (Cited on pages 145, 147).

[363] ATLAS Collaboration, *Observation of electroweak production of two jets and a* Z*-boson pair*, Nature Physics **19** (2023) 237,
URL: https://cds.cern.ch/record/2715804 (Cited on pages 145, 147).

[364] CMS Collaboration, *Observation of electroweak production of pairs of* Z *bosons at* $13\,\mathrm{TeV}$ *with the CMS detector*, Physics Analysis Summary CMS-PAS-SMP-23-001, CERN, 2026,
URL: https://cds.cern.ch/record/2957284 (Cited on pages 145, 147).

[365] ATLAS Collaboration, *Search for the electroweak diboson production in association with a high-mass dijet system in semileptonic final states in* pp *collisions at* $\sqrt{s} = 13\,\mathrm{TeV}$ *with the ATLAS detector*, Physical Review D **100** (2019) 032007,
URL: https://cds.cern.ch/record/2675009 (Cited on pages 145, 147).

[366] CMS Collaboration, *Evidence for* $\mathrm{WW/WZ}$ *vector boson scattering in the decay channel* $\ell\nu qq$ *produced in association with two jets in proton–proton collisions at* $\sqrt{s} = 13\,\mathrm{TeV}$, Physics Letters B **834** (2022) 137438,
URL: https://cds.cern.ch/record/2792547 (Cited on pages 145, 147).

[367] CMS Collaboration, *Vector boson scattering and anomalous quartic couplings in final states with* $\ell\nu qq$ *or* $\ell\ell qq$ *plus jets using proton–proton collisions at* $\sqrt{s} = 13\,\mathrm{TeV}$, Journal of High Energy Physics **2026** (2026) 022,
URL: https://cds.cern.ch/record/2944855 (Cited on pages 145, 147).

[368] ATLAS Collaboration, *Electroweak diboson production in association with a high-mass dijet system in semileptonic final states from* pp *collisions at* $\sqrt{s} = 13\,\text{TeV}$ *with the ATLAS detector*, The European Physical Journal C **85** (2025) 1275, URL: https://cds.cern.ch/record/2927622 (Cited on pages 145, 147).

[369] E. Accomando, A. Ballestrero, A. Belhouari and E. Maina, *Isolating vector boson scattering at the CERN LHC: Gauge cancellations and the equivalent vector boson approximation versus complete calculations*, Physical Review D **74** (2006) 073010 (Cited on page 149).

[370] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at* $\sqrt{s} = 13\,\text{TeV}$, The European Physical Journal C **76** (2016) 292, URL: https://cds.cern.ch/record/2139897 (Cited on pages 150, 219).

[371] ATLAS Collaboration, *Muon reconstruction and identification efficiency in ATLAS using the full Run-2* pp *collision data set at* $\sqrt{s} = 13\,\text{TeV}$, The European Physical Journal C **81** (2021) 578, URL: https://cds.cern.ch/record/2746302 (Cited on pages 151, 219).

[372] ATLAS Collaboration, *Studies of the muon momentum calibration and performance of the ATLAS detector with* pp *collisions at* $\sqrt{s} = 13\,\text{TeV}$, The European Physical Journal C **83** (2023) 686, URL: https://cds.cern.ch/record/2844624 (Cited on pages 152, 219).

[373] ATLAS Collaboration, *Jet energy scale and resolution measured in proton–proton collisions at* $\sqrt{s} = 13\,\text{TeV}$ *with the ATLAS detector*, The European Physical Journal C **81** (2021) 689, URL: https://cds.cern.ch/record/2722869 (Cited on pages 157, 160, 218, 225).

[374] M. Cacciari, G. P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, Journal of High Energy Physics **2008** (2008) 063 (Cited on page 157).

[375] S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions*, Physical Review D **48** (1993) 3160 (Cited on page 157).

[376] M. Cacciari, G. P. Salam and G. Soyez, *The catchment area of jets*, Journal of High Energy Physics **2008** (2008) 005 (Cited on pages 158, 160).

[377] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in* pp *collisions at* $\sqrt{s} = 8\,\text{TeV}$ *using the ATLAS detector*, The European Physical Journal C **76** (2016) 581, URL: https://cds.cern.ch/record/2058295 (Cited on pages 161, 219, 225).

[378] ATLAS Collaboration, *Identification and rejection of pile-up jets at high pseudorapidity with the ATLAS detector*, The European Physical Journal C **77** (2017) 580, URL: https://cds.cern.ch/record/2262419 (Cited on pages 161, 225).

[379] ATLAS Collaboration, *ATLAS flavour-tagging algorithms for the LHC Run-2* pp *collision dataset*, The European Physical Journal C **83** (2023) 681, URL: https://cds.cern.ch/record/2842028 (Cited on pages 162, 218).

[380] ATLAS Collaboration, *Transforming jet flavour tagging at ATLAS*, Nature Communications **17** (2026) 541, URL: https://cds.cern.ch/record/2933089 (Cited on pages 162, 200, 218).

[381] ATLAS Collaboration, *Transformer networks for constituent-based b-jet calibration with the ATLAS detector*, Note ATL-PHYS-PUB-2025-012, CERN, 2025, URL: https://cds.cern.ch/record/2905688 (Cited on page 163).

[382] ATLAS Collaboration, *Calibration of the jet energy scale and resolution of small-radius jets using semileptonic* $\text{t}\bar{\text{t}}$ *events with the ATLAS detector*, Preprint CERN-EP-2025-262, CERN, 2025, arXiv: 2512.17482 [hep-ex], URL: https://cds.cern.ch/record/2951723 (Cited on pages 163, 218).

[383] ATLAS Collaboration, *The performance of missing transverse momentum reconstruction and its significance with the ATLAS detector using* $140\,\text{fb}^{-1}$ *of* $\sqrt{s} = 13\,\text{TeV}$ pp *collisions*, The European Physical Journal C **85** (2025) 606, URL: https://cds.cern.ch/record/2888751 (Cited on pages 163, 219).

[384] ATLAS Collaboration, *Reconstruction of hadronic decay products of tau leptons with the ATLAS experiment*, The European Physical Journal C **76** (2016) 295, URL: https://cds.cern.ch/record/2116436 (Cited on page 164).

[385] ATLAS Collaboration, *Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment*, Note ATL-PHYS-PUB-2019-033, CERN, 2019, URL: https://cds.cern.ch/record/2688062 (Cited on page 165).

[386] U. Baur, T. Han and J. Ohnemus, *Amplitude zeros in $\mathrm{W}^{\pm}\mathrm{Z}$ production*, Physical Review Letters **72** (1994) 3941 (Cited on page 167).

[387] C. Cohen-Tannoudji, B. Diu and F. Laloë, *Mécanique Quantique – Tome 1: Nouvelle édition*, EDP Sciences, 2020 (Cited on page 171).

[388] C. Bourrely, E. Leader and J. Soffer, *Polarization phenomena in hadronic reactions*, Physics Reports **59** (1980) 95 (Cited on pages 171, 266).

[389] J. A. Aguilar-Saavedra, J. Bernabéu, V. A. Mitsou and A. Segarra, *The Z boson spin observables as messengers of new physics*, The European Physical Journal C **77** (2017) 234 (Cited on pages 171, 266).

[390] Z. Bern et al., *Left-handed W bosons at the LHC*, Physical Review D **84** (2011) 034008 (Cited on page 171).

[391] E. P. Wigner, *Gruppentheorie und ihre Anwendung auf die Quantenmechanik der Atomspektren*, Braunschweig: Vieweg+Teubner Verlag, 1931 (Cited on page 173).

[392] M. Jacob and G. C. Wick, *On the General Theory of Collisions for Particles with Spin*, Annals of Physics **281** (2000) 774 (Cited on pages 173, 266, 271).

[393] A. Aeppli, G. J. van Oldenborgh and D. Wyler, *Unstable particles in one loop calculations*, Nuclear Physics B **428** (1994) 126 (Cited on page 177).

[394] C. F. Uhlemann and N. Kauer, *Narrow-width approximation accuracy*, Nuclear Physics B **814** (2009) 195 (Cited on page 177).

[395] A. Ballestrero, E. Maina and G. Pelliccioli, *Polarized vector boson scattering in the fully leptonic WZ and ZZ channels at the LHC*, Journal of High Energy Physics **2019** (2019) 087 (Cited on pages 177, 178, 266).

[396] A. Denner, S. Dittmaier, M. Roth and D. Wackeroth, *Predictions for all processes $\mathrm{e}^{+}\mathrm{e}^{-} \to$ fermions $+\gamma$*, Nuclear Physics B **560** (1999) 33 (Cited on page 180).

[397] J. C. Ward, *An Identity in Quantum Electrodynamics*, Physical Review **78** (1950) 182 (Cited on page 181).

[398] A. A. Slavnov, *Ward identities in gauge theories*, Theoretical and Mathematical Physics **10** (1972) 99 (Cited on page 181).

[399] J. C. Taylor, *Ward identities and charge renormalization of the Yang–Mills field*, Nuclear Physics B **33** (1971) 436 (Cited on page 181).

[400] D. B. Franzosi, O. Mattelaer, R. Ruiz and S. Shil, *Automated predictions from polarized matrix elements*, Journal of High Energy Physics **2020** (2020) 082 (Cited on page 182).

[401] M. Hoppe, M. Schönherr and F. Siegert, *Polarised cross sections for vector boson production with Sherpa*, Journal of High Energy Physics **2024** (2024) 001 (Cited on pages 183, 294).

[402] G. Pelliccioli and G. Zanderighi, *Polarised-boson pairs at the LHC with NLOPS accuracy*, The European Physical Journal C **84** (2024) 16 (Cited on pages 183, 294).

[403] P. Richardson, *Spin correlations in Monte Carlo simulations*, Journal of High Energy Physics **2001** (2001) 029 (Cited on page 183).

[404] G. Bozzi, B. Jäger, C. Oleari and D. Zeppenfeld, *Next-to-Leading-Order QCD corrections to $\mathrm{W}^{+}\mathrm{Z}$ and $\mathrm{W}^{-}\mathrm{Z}$ production via vector-boson fusion*, Physical Review D **75** (2007) 073004 (Cited on page 184).

[405] B. Jäger, A. Karlberg and J. Scheller, *Parton-shower effects in electroweak WZjj production at the next-to-leading order of QCD*, The European Physical Journal C **79** (2019) 226 (Cited on pages 184, 188).

[406] B. Jäger, A. Karlberg and S. Reinhardt, *QCD effects in electroweak WZjj production at current and future hadron colliders*, The European Physical Journal C **84** (2024) 587 (Cited on page 184).

[407] A. Denner, S. Dittmaier, P. Maierhöfer, M. Pellen and C. Schwan, *QCD and electroweak corrections to WZ scattering at the LHC*, Journal of High Energy Physics **2019** (2019) 067 (Cited on pages 184, 186, 188).

[408] A. Denner, C. Haitz and G. Pelliccioli, *NLO EW and QCD corrections to polarised same-sign WW scattering at the LHC*, Journal of High Energy Physics **2024** (2024) 115 (Cited on pages 184, 186–188).

[409] A. Ballestrero et al., *Precise predictions for same-sign W-boson scattering at the LHC*, The European Physical Journal C **78** (2018) 671 (Cited on page 184).

[410] A. Denner and S. Pozzorini, *One-loop leading logarithms in electroweak radiative corrections, I. Results*, The European Physical Journal C **18** (2001) 461 (Cited on pages 184, 185, 280, 282, 283).

[411] A. Denner and S. Pozzorini, *One-loop leading logarithms in electroweak radiative corrections, II. Factorization of collinear singularities*, The European Physical Journal C **21** (2001) 63 (Cited on pages 184, 185, 280, 282, 283).

[412] A. Denner and S. Dittmaier, *Electroweak radiative corrections for collider physics*, Physics Reports **864** (2020) 1 (Cited on pages 184, 185, 279, 280).

[413] A. Denner, R. Franken, C. Haitz, D. Lombardi and G. Pelliccioli, *Electroweak corrections to doubly polarised* WZ *scattering at the LHC*, Journal of High Energy Physics **2026** (2026) 120 (Cited on pages 185–188, 193, 194, 208, 282).

[414] A. Denner, S. Dittmaier, M. Roth and L. H. Wieders, *Electroweak corrections to charged-current* $\mathrm{e^+e^-} \to 4$ *fermion processes: Technical details and further results*, Nuclear Physics B **724** (2005) 247 (Cited on page 185).

[415] A. Denner, D. Lombardi, S. Lopez Portillo Chavez, M. Pellen and G. Pelliccioli, *MoCaNLO: a Monte Carlo integrator for NLO calculations*, 2026, arXiv: 2602.19842 [hep-ph], URL: https://arxiv.org/abs/2602.19842 (Cited on pages 186, 229).

[416] A. Denner, S. Dittmaier and L. Hofer, *COLLIER: a fortran-based Complex One-Loop LIbrary in Extended Regularizations*, Computer Physics Communications **212** (2017) 220 (Cited on page 186).

[417] B. Biedermann, A. Denner and M. Pellen, *Large Electroweak Corrections to Vector-Boson Scattering at the Large Hadron Collider*, Physical Review Letters **118** (2017) 261801 (Cited on pages 187, 188).

[418] A. Ballestrero, A. Belhouari, G. Bevilacqua, V. Kashkan and E. Maina, *PHANTOM: A Monte Carlo event generator for six parton final states at high energy colliders*, Computer Physics Communications **180** (2009) 401 (Cited on page 188).

[419] ATLAS Collaboration, *Proposal for particle-level object and observable definitions for use in physics measurements at the LHC*, Note ATL-PHYS-PUB-2015-013, CERN, 2015, URL: https://cds.cern.ch/record/2022743 (Cited on page 188).

[420] ATLAS Collaboration, `TRExFitter`, 2026, DOI: 10.5281/zenodo.14845712 (Cited on pages 214, 358).

[421] K. Cranmer, G. Lewis, L. Moneta, A. Shibata and W. Verkerke, `HistFactory`*: A tool for creating statistical models for use with* `RooFit` *and* `RooStats`, Report CERN-OPEN-2012-016, CERN, 2012, URL: https://cds.cern.ch/record/1456844 (Cited on pages 214, 358).

[422] ATLAS Collaboration, *ATLAS* b*-jet identification performance and efficiency measurement with* $\mathrm{t\bar{t}}$ *events in* pp *collisions at* $\sqrt{s} = 13\,\mathrm{TeV}$, The European Physical Journal C **79** (2019) 970, URL: https://cds.cern.ch/record/2682119 (Cited on page 218).

[423] ATLAS Collaboration, *Measurement of the* c*-jet mistagging efficiency in* $\mathrm{t\bar{t}}$ *events using* pp *collision data at* $\sqrt{s} = 13\,\mathrm{TeV}$ *collected with the ATLAS detector*, The European Physical Journal C **82** (2022) 95, URL: https://cds.cern.ch/record/2781747 (Cited on page 218).

[424] ATLAS Collaboration, *Calibration of the light-flavour jet mistagging efficiency of the* b*-tagging algorithms with* Z*+jets events using* $139\,\mathrm{fb}^{-1}$ *of ATLAS proton–proton collision data at* $\sqrt{s} = 13\,\mathrm{TeV}$, The European Physical Journal C **83** (2023) 728, URL: https://cds.cern.ch/record/2846342 (Cited on page 218).

[425] ATLAS Collaboration, *Eigenvector recomposition: a new method to correlate flavour-tagging systematic uncertainties across analyses*, Note ATL-PHYS-PUB-2022-024, CERN, 2022, URL: https://cds.cern.ch/record/2809651 (Cited on page 218).

[426] ATLAS Collaboration, *A method for the construction of strongly reduced representations of ATLAS experimental uncertainties and the application thereof to the jet energy scale*, Note ATL-PHYS-PUB-2015-014, CERN, 2015, URL: https://cds.cern.ch/record/2037436 (Cited on page 218).

[427] ATLAS Collaboration, *Measurement of the Inelastic Proton–Proton Cross Section at* $\sqrt{s} = 13\,\mathrm{TeV}$ *with the ATLAS Detector at the LHC*, Physical Review Letters **117** (2016) 182002, URL: https://cds.cern.ch/record/2159578 (Cited on page 219).

[428] ATLAS Collaboration, *Preliminary luminosity calibration of the ATLAS $13.6\,\mathrm{TeV}$ data recorded in 2024 and combination with the 2022 and 2023 measurements*, Note ATL-DAPR-PUB-2025-001, CERN, 2025,
URL: https://cds.cern.ch/record/2948582 (Cited on pages 219, 226).

[429] ATLAS Collaboration, *Multi-Boson Simulation for $13\,\mathrm{TeV}$ ATLAS Analyses*, Note ATL-PHYS-PUB-2016-002, CERN, 2016,
URL: https://cds.cern.ch/record/2119986 (Cited on page 227).

[430] J. Butterworth et al., *PDF4LHC recommendations for LHC Run-2*, Journal of Physics G **43** (2016) 023001 (Cited on page 228).

[431] A. Pinto et al., *Uncertainty components in profile likelihood fits*, The European Physical Journal C **84** (2024) 593 (Cited on pages 242, 362).

[432] ATLAS Collaboration, *Evidence for longitudinally polarised Z bosons in electroweak $\mathrm{W^{\pm}Z}$ production in association with two jets from pp collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ and $13.6\,\mathrm{TeV}$ with the ATLAS detector*, 2026, arXiv: 2607.20246 [hep-ex],
URL: https://cds.cern.ch/record/2967028 (Cited on pages 245, 254).

[433] K. Cranmer, *Practical Statistics for the LHC*, 2015, arXiv: 1503.07622 [physics.data-an],
URL: https://arxiv.org/abs/1503.07622 (Cited on page 251).

[434] ATLAS Collaboration, *Observation of the associated production of a top quark and a Z boson in pp collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ with the ATLAS detector*, Journal of High Energy Physics **2020** (2020) 124,
URL: https://cds.cern.ch/record/2710130 (Cited on pages 286, 290).

[435] ATLAS Collaboration, *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model*, Note ATL-PHYS-PUB-2016-017, CERN, 2016,
URL: https://cds.cern.ch/record/2206965 (Cited on page 292).

[436] ATLAS Collaboration, *ATLAS Pythia 8 tunes to $7\,\mathrm{TeV}$ data*, Note ATL-PHYS-PUB-2014-021, CERN, 2014,
URL: https://cds.cern.ch/record/1966419 (Cited on pages 295, 297).

[437] R. Frederix and S. Frixione, *Merging meets matching in MC@NLO*, Journal of High Energy Physics **2012** (2012) 061 (Cited on page 296).

[438] K. Pearson, *Note on regression and inheritance in the case of two parents*, Proceedings of the Royal Society of London **58** (1895) 240 (Cited on page 320).

[439] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, The European Physical Journal C **71** (2011) 1554 (Cited on pages 358, 363).

[440] S. S. Wilks, *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*, The Annals of Mathematical Statistics **9** (1938) 60 (Cited on page 361).

[441] A. Wald, *Tests of statistical hypotheses concerning several parameters when the number of observations is large*, Transactions of the American Mathematical Society **54** (1943) 426 (Cited on page 363).

[442] J. Neyman, *Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability*, Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences **236** (1937) 333 (Cited on page 363).

[443] G. J. Feldman and R. D. Cousins, *Unified approach to the classical statistical analysis of small signals*, Physical Review D **57** (1998) 3873 (Cited on page 363).

[444] T. Junk, *Confidence level computation for combining searches with small statistics*, Nuclear Instruments and Methods in Physics Research Section A **434** (1999) 435 (Cited on page 364).

[445] S. Baker and R. D. Cousins, *Clarification of the use of chi-square and likelihood functions in fits to histograms*, Nuclear Instruments and Methods in Physics Research **221** (1984) 437 (Cited on page 365).

[446] U. de Noyers, *Phenomenological aspects of scotogenic models*, PhD thesis: Université Savoie Mont Blanc, 2025,
URL: https://theses.hal.science/tel-05493910 (Cited on page 406).

[447] U. de Noyers, M. Sarazin and B. Herrmann, *Phenomenology of a singlet–doublet–triplet scotogenic framework*, The European Physical Journal C **85** (2025) 922 (Cited on pages 406, 408, 410).

[448] E. Ma, *Verifiable radiative seesaw mechanism of neutrino mass and dark matter*, Physical Review D **73** (2006) 077301 (Cited on page 406).

[449] D. Restrepo, O. Zapata and C. E. Yaguna, *Models with radiative neutrino masses and viable dark matter candidates*, Journal of High Energy Physics **2013** (2013) 011 (Cited on page 406).

[450] M. Sarazin, J. Bernigaud and B. Herrmann, *Dark matter and lepton flavour phenomenology in a singlet–doublet scotogenic model*, Journal of High Energy Physics **2021** (2021) 116 (Cited on page 406).

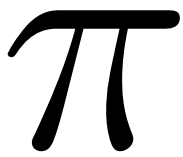
[451] J. A. Casas and A. Ibarra, *Oscillating neutrinos and* $\mu \to e, \gamma$, Nuclear Physics B **618** (2001) 171 (Cited on page 407).

[452] I. Esteban et al., *NuFit-6.0: updated global analysis of three-flavor neutrino oscillations*, Journal of High Energy Physics **2024** (2024) 216 (Cited on page 407).

[453] F. Staub, *SARAH*, 2008, arXiv: 0806.0538 [hep-ph], URL: https://arxiv.org/abs/0806.0538 (Cited on page 408).

[454] W. Porod, *SPheno, a program for calculating supersymmetric spectra, SUSY particle decays and SUSY particle production at* e^+e^- *colliders*, Computer Physics Communications **153** (2003) 275 (Cited on page 408).

[455] G. Bélanger, F. Boudjema, A. Pukhov and A. Semenov, *micrOMEGAs: A program for calculating the relic density in the MSSM*, Computer Physics Communications **149** (2002) 103 (Cited on page 408).



Acronyms

μ-PIC micro-pixel chamber. 68

μ-RWELL micro-resistive WELL. 68

pp proton–proton. 1–3, 35–37, 39, 42–44, 48, 64, 65, 95–98, 102, 103, 214, 219, 254

1D one-dimensional. 46, 176, 198, 214, 242, 287, 363, 365, 390

2D two-dimensional. i, 46, 103, 121, 176, 198, 199, 201, 202, 209, 214, 238, 242, 243, 272, 298, 348, 350, 363, 365, 410

3D three-dimensional. 33, 45, 46, 49, 58, 69, 70, 118, 122, 172, 198, 257, 298, 299

4D four-dimensional. 12, 94, 271

4FNS 4-flavour number scheme. 93, 293–295

5FNS 5-flavour number scheme. 93, 293–295, 297

ADAM ADAptive Moment estimation. 76, 77, 128, 200, 409

ADC analogue-to-digital converter. 107, 109

AE AutoEncoder. 83

AF3 AtlFast3. 219, 225

AI artificial intelligence. 72

AL active learning. 4, 93, 257, 406, 408, 410, 411

Al aluminium. 57, 62

ALEPH Apparatus for LEP PHysics. 32, 144

ALICE A Large Ion Collider Experiment. 35

AMVF adaptive multi-vertex finder. 103

Ar argon. 47, 57, 60, 61

ASIC application-specific integrated circuit. 64

ATLAS A Toroidal LHC ApparatuS. i, ii, 1–3, 5–7, 23, 35, 37, 42–44, 48, 50, 51, 54, 57, 62–67, 70, 84, 86, 96, 98, 99, 101, 103, 107, 108, 116, 117, 119–122, 124, 138, 141, 144, 145, 150, 155–157, 162, 163, 188, 194, 197, 214, 217, 242, 245, 252, 254, 255, 386, 445

Au gold. 10, 35, 47, 57, 61

AUC area under curve. 82, 88, 89, 200–203, 300

BAGGING Bootstrap AGGregatING. 86

BCE binary cross entropy. 75, 132, 200, 202, 409

BCM Beam Conditions Monitor. 66

BDT boosted decision tree. 3, 84–86, 88, 109, 116, 120, 152, 208, 214, 220, 222, 232, 234

BEE Barrel End-cap Extra. 57

BEH Brout–Englert–Higgs. 20, 21, 25

BI Barrel Inner. 67

BIS Barrel Inner Small. 57, 59, 67, 154

BLL binomial log-likelihood. 86

BML Barrel Middle Large. 60

BMS Barrel Middle Small. 60

BNL Brookhaven National Laboratory. 17, 18

BO Bayesian optimisation. 2, 3, 89, 91, 200, 257, 300, 358

BOL Barrel Outer Large. 60

BOS Barrel Outer Small. 60

BR branching fraction. 96, 179, 180

BSM beyond the Standard Model. 1, 3, 4, 11, 28, 40, 93, 145, 257, 406

CB combined. 151, 219, 226

CC charged current. 185, 277, 281

CCE categorical cross entropy. 75, 200, 201

CDF Collider Detector at Fermilab. 19, 86, 144

CE cross entropy. 85

CERN European Organization for Nuclear Research. 1, 5, 32–35, 42, 65, 87, 88, 446

CKF combinatorial Kalman filter. 93, 101–103

CKM Cabibbo–Kobayashi–Maskawa. 18, 24, 27

CL confidence level. i, 227, 228, 245, 247, 251, 254, 363–365

CLIC Compact Linear Collider. 40

CM centre-of-mass. i, 1, 21, 22, 31, 35, 37–39, 124, 144, 145, 148, 169, 204, 245, 252, 260–264, 288, 293–295

CMS Compact Muon Solenoid. i, ii, 1, 5, 23, 35, 37, 48, 84, 144, 145, 252

CNN convolutional neural network. 83, 84, 154

CO_2 carbon dioxide. 47, 57, 60, 61, 68

CP charge-parity. 18, 19, 24, 28, 35

CPU central processing unit. 65, 116

CR control region. 192, 193, 196–198, 205, 214–217, 220, 222, 227, 228, 230–232, 234, 236–238, 245, 252, 286–290, 294, 361, 362, 365, 378–384, 392, 405

CS Catani–Seymour. 94, 184, 187, 293–295

CSC Cathode-Strip Chamber. 59, 60, 154

CTP Central Trigger Processor. 64

Cu copper. 50, 51, 55, 56, 59, 61, 62

DAQ Data AcQuisition. 3, 64, 67

DBM Diamond Beam Monitor. 66

DELPHI DEtector with Lepton, Photon and Hadron Identification. 32, 84

DESY Deutsches Elektronen-Synchrotron. 16, 84

DGLAP Dokshitzer–Gribov–Lipatov–Altarelli–Parisi. 39

DM dark matter. 1, 3, 27, 257, 406–408, 410

DNN deep neural network. i, 2, 4, 72, 83, 116, 117, 120, 125, 127–129, 132, 134–143, 152, 162, 166, 186, 190, 198–211, 214, 215, 218, 228–230, 254, 255, 257, 287, 298–304, 306–318, 320–346, 406, 408–410

DPA double pole approximation. 180–183, 185–188, 193, 194, 293, 294

DT decision tree. 72, 84–86

EEL Extra End-cap Large. 57

EES Extra End-cap Small. 57

EI expected improvement. 89

EIL End-cap Inner Large. 59, 61, 67

EIS End-cap Inner Small. 59, 61

EM electromagnetic. 15, 20, 21, 25, 47, 48, 50–56, 62–64, 98–102, 104–111, 113, 114, 116, 118, 120, 121, 125, 126, 138, 152, 154, 155, 158, 160, 164, 185, 189, 190, 280, 281

EMB ElectroMagnetic Barrel. 50, 51

EMEC ElectroMagnetic End-cap. 50, 51, 55, 65, 121, 123, 124, 127

EW electroweak. i, 1–3, 20, 21, 25, 28, 30, 31, 40, 108, 117, 145, 148–150, 178, 179, 183–188, 193, 194, 196–198, 200, 208, 214, 229, 230, 236, 238, 242, 245, 247, 252, 254, 255, 257, 276, 279–283, 288, 292–296, 306, 366, 378, 390, 404, 408

EYETS Extended Year-End Technical Stop. 37

FCal Forward Calorimeter. 48, 55, 56, 64, 65, 121, 123–125, 127, 129, 134, 138, 141, 219, 225

FCC Future Circular Collider. 40

Fe iron. 35, 50, 55, 57, 62

FELIX Front-End Link eXchange. 67

Fermilab Fermi National Accelerator Laboratory. 19, 86

fJVT forward jet vertex tagger. 161, 162, 190, 191, 219, 225

FKS Frixione–Kunszt–Signer. 95, 96, 184

FN false negative. 82

FP false positive. 82

FPGA field-programmable gate array. 64

FS full simulation. 219, 225

FSR final-state radiation. 39, 95, 96, 124, 188, 189, 296, 297

FWHM full width at half maximum. 112

GAN generative adversarial network. 83, 410

GBET Goldstone boson equivalence theorem. 1, 30, 31, 181, 258–264, 276, 280

GN graph network. 162, 190, 191, 200, 218

GNN graph neural network. 83, 84, 162

GP Gaussian process. 89–91, 300

GPU graphics processing unit. 67, 72, 84

GSC global sequential calibration. 160, 163

GSF Gaussian sum filter. 102, 104, 116, 124

GUT grand unified theory. 28

GWS Glashow–Weinberg–Salam. 20, 21, 23, 25, 258

H hydrogen. 33, 53, 55, 60

HAD hadronic. 18, 40, 47, 48, 51, 53–56, 63, 64, 93, 98, 100, 101, 111, 113, 114, 117, 119–122, 125, 126, 154, 155, 158, 160, 164, 165, 196, 219, 279

HEC Hadronic End-cap Calorimeter. 54–56, 64, 121, 125

HERA Hadron-Elektron-Ringanlage. 84

HF heavy-flavour. 102, 111, 112, 114, 116, 117, 120, 152, 162, 165, 284, 287, 288, 294

HGTD High-Granularity Timing Detector. 67, 70, 71, 124–126, 138, 140, 141, 163

HL high-luminosity. 6, 7, 66, 120

HL-LHC High-Luminosity LHC. i, 2, 3, 37, 38, 40, 51, 68, 120, 124, 129, 141, 154, 252, 255

HLT High-Level Trigger. 64, 65, 116

HV high voltage. 60, 61, 109

IBL Insertable b-Layer. 46, 66, 70

ID Inner Detector. 2, 3, 42, 44–47, 51, 55, 57, 62, 63, 65, 67, 68, 70, 98, 99, 101, 102, 104, 106, 113, 116–118, 121, 124, 141, 150–152, 154, 155, 190, 192

ILC International Linear Collider. 40

IP interaction point. 35–38, 42–44, 46, 56, 59, 64, 66, 68, 70, 87, 98, 107, 122, 151, 154, 155

ISR initial-state radiation. 39, 93, 95, 169, 188

ITC Intermediate Tile Calorimeter. 54

ITk Inner Tracker. i, ii, 2, 3, 6, 7, 67–70, 98, 124, 129, 138, 140, 141, 255

JER jet energy resolution. 157, 161, 163, 217–219, 225, 242, 246

JES jet energy scale. 160, 161, 163, 217–219, 224, 242, 246

JTE junction termination extension. 70

JVT jet vertex tagger. 161, 162, 190, 191, 217, 219, 225

KDE kernel density estimation. 111, 112

L0 Level 0. 67

L1 Level 1. 64, 65, 84, 107, 154

L1 Layer 1. 51, 53, 100, 108

L1Calo Level 1 Calorimeter. 64, 65

L1Muon Level 1 Muon. 64, 65

L1Topo Level 1 Topological. 64

L2 Level 2. 65

L2 Layer 2. 51, 53, 108, 109, 155

L3 Layer 3. 51, 53, 108

LAPP Laboratoire d'Annecy de Physique des Particules. 445

LAr liquid argon. 50, 51, 55, 56, 65, 67, 106–109, 193

LBL Lawrence Berkeley Laboratory. 24

LEIR Low Energy Ion Ring. 35

LEP Large Electron–Positron Collider. 32, 33, 144

LF light-flavour. 111, 112, 114, 117, 162, 218, 224, 287, 288

LFV lepton flavour violation. 407, 408

LGAD low-gain avalanche detector. 70

LH left-handed. 14, 17, 18, 22, 29, 169, 173–175, 177, 270, 272, 276, 277, 282, 407

LH likelihood. i, 91, 111–114, 116, 117, 120, 123, 134, 141, 161, 162, 190, 198, 214, 218, 230, 242, 243, 252, 358–365, 378, 386, 390, 391, 402, 403, 408, 410

LHC Large Hadron Collider. i, ii, 1–3, 5–7, 23, 32–39, 42, 43, 47, 59, 65, 66, 86, 108, 144, 154, 186, 188, 214, 254, 255, 446

LHCb Large Hadron Collider beauty experiment. 35

LINAC LINear ACcelerator. 33, 35, 37, 87

LIPS Lorentz-invariant phase-space. 13, 94, 182, 186, 263, 279

LL log-likelihood. 242, 251, 358, 359, 362

LLM large language model. 83

LO leading order. 12, 31, 94, 96, 149, 180, 182–184, 186–188, 194, 202, 215, 228, 229, 292–297

LR linear regression. 72, 129, 131, 141

LRD local-recoil dipole. 293, 295

LS Long Shutdown. 37, 46, 47, 51, 59, 61, 65, 67, 68, 70, 124, 154

LUCID LUminosity Cherenkov Integrating Detector. 65–67, 219

MBS momentum-balance significance. 151

MC Monte Carlo. 3, 54, 92–97, 109–111, 120, 124, 129, 149, 160, 180, 182, 186, 189, 193–197, 199–203, 205, 206, 209, 211, 214–219, 225, 227–229, 246, 248–251, 255, 284, 286–290, 292, 306–310, 350, 352, 362, 392, 393, 396

MCMC Markov chain Monte Carlo. 4, 93, 406, 408, 410

MDT Monitored Drift Tube. 57, 59–61, 67, 154

ME matrix element. 30, 93–96, 173, 180, 182, 183, 186–188, 228, 259–263, 277–279, 292–297, 396

MEPS matrix element plus parton shower. 96, 183, 188, 293–296

MET missing transverse momentum. 2, 47, 54, 64, 165, 189, 204, 205, 226, 246, 248–250

MIP minimum ionising particle. 57, 150, 151

ML machine learning. 2, 3, 67, 72, 84–88, 90, 93, 149, 162, 209, 254, 255, 406

MLE maximum likelihood estimate. 358, 359

MM MicroMegas. 59, 61, 62, 68, 154

MPGD micro-pattern gas detector. 68

MPI multi-parton interactions. 39, 95

MS Muon Spectrometer. 42, 47, 57, 58, 60, 63, 67, 98, 150–152, 160

MSE mean square error. 74, 127, 128, 166

NADAM Nesterov-accelerated ADAptive Moment estimation. 127, 128, 132, 409

NC neutral current. 32, 277

NF normalisation factor. 196, 205, 215, 216, 227, 236, 238, 242, 245, 252, 361, 362, 365, 386, 392, 398, 402, 404

NLO next-to-leading order. i, 12, 17, 94–96, 180, 183–188, 193, 194, 200, 201, 229, 230, 242, 245, 246, 252, 254, 255, 280, 292–297, 366, 404

NN neural network. 3, 67, 72–75, 77–86, 88, 89, 97, 162, 165, 190, 191, 200, 219, 225, 300, 408

NNLO next-to-next-to-leading order. 17, 95, 297

NP nuisance parameter. 111, 160, 216–219, 224, 225, 227, 230, 236, 238, 242, 244, 251, 255, 360–363, 378–384, 386–389, 392, 393, 396, 401, 404

NSW New Small Wheel. 61, 65, 68, 154

NWA narrow-width approximation. 177, 179–181, 293

O_2 dioxygen. 47

OFC optimal filter coefficient. 107

OPAL Omni-Purpose Apparatus at LEP. 32, 144

Pb lead. 35, 48, 50, 51, 54, 109

PC photon conversion. 99, 102, 104, 109, 112–114, 116, 117, 125, 165, 284, 286–289

PCB printed circuit board. 61

PDF parton distribution function. 38, 39, 93, 94, 96, 182, 217, 219, 227, 228, 245, 246, 248–250, 292, 295, 297, 366

PDF probability density function. 38, 89, 92, 111–113, 117, 359, 360

PDG Particle Data Group. 11, 96, 189, 358

PETRA Positron-Elektron-Tandem-Ring-Anlage. 16

PFlow particle flow. 119, 152, 154, 155, 157, 158, 161

PFO particle-flow object. 155, 157, 190, 191

PI probability of improvement. 90

PLIV prompt-lepton improved veto. 120, 152

PLV prompt-lepton veto. 120, 152

PMNS Pontecorvo–Maki–Nakagawa–Sakata. 24, 27, 407

PMT PhotoMultiplier Tube. 54, 65, 66

POI parameter of interest. 205, 215, 216, 230, 238–242, 244–246, 298, 360–364, 390, 392, 396, 398, 401, 402

PreRec pre-recommendation. 219, 225

PS parton shower. 39, 95, 96, 183, 188, 189, 217, 227, 228, 292–297, 396

PS pre-sampler. 51, 53, 100, 101, 106, 108, 109

PS Proton Synchrotron. 32, 33, 87

PSB Proton Synchrotron Booster. 33, 37

QCD quantum chromodynamics. 16, 17, 27, 35, 40, 93, 120, 129, 148–150, 161, 167, 168, 183, 184, 187, 188, 196, 197, 203, 214, 227–229, 236, 238, 242, 245, 246, 251, 279, 280, 292, 294–297, 306, 338, 366, 386, 396

QED quantum electrodynamics. 11, 15–18, 21, 94, 96, 187–189, 281, 296, 297

QFT quantum field theory. 12

QGP quark–gluon plasma. 17, 35

RAZ radiation amplitude zero. 145, 146, 167

ReLU rectified linear unit. 79, 80, 127, 128, 132, 200, 409

RF radio frequency. 33, 35, 37

RH right-handed. 14, 17, 18, 22, 23, 25, 29, 169, 173–175, 177, 270, 272, 277, 282, 407

RMS root mean square. 36

RNN recurrent neural network. 83, 84, 165

ROC receiver operating characteristic. 82, 88, 134, 136, 141, 143, 201–204, 206, 324, 335–337

ROD ReadOut Driver. 64

RoI region of interest. 65, 102, 104

ROOT ROOT data-analysis framework. 86

ROS ReadOut System. 65

RPC Resistive Plate Chamber. 60, 61, 64, 67, 154

SCT Semi-Conductor Tracker. 44, 46, 63, 66, 68, 101, 102, 104, 113, 114, 150, 193

SDM spin density matrix. 171–173, 176–178, 182, 183, 198, 266, 268, 270–272, 294

SF scale factor. 108–110, 116, 120, 121, 124, 131, 153, 154, 162, 189, 217–219, 224–226

SFOC same-flavour opposite-charge. 192, 193, 284, 286

SGD stochastic gradient descent. 75, 76, 83

SHAP SHapley Additive exPlanations. 87, 88, 134, 138, 139, 324–334

Si silicon. 44, 46, 66, 68, 70, 102, 104–106, 113, 118, 124, 141, 150, 151, 155

SLAC Stanford Linear Accelerator Center. 18, 24

SLHA SUSY Les Houches Accord. 408, 410

SM Standard Model. i, 1–3, 11, 25–28, 31, 35, 98, 144, 163, 179, 183, 188, 198, 215, 216, 220–223, 231–236, 242, 245, 251, 252, 254, 255, 276, 363–365, 390, 402, 406, 407

sMDT small-diameter Monitored Drift Tube. 59, 67

SNO Sudbury Neutrino Observatory. 24

SPEAR Stanford Positron–Electron Asymmetric Ring. 24

SPS Super Proton Synchrotron. 32, 33

SR signal region. 97, 168, 169, 179, 187, 192, 193, 196–198, 205, 214–216, 221, 223, 227, 228, 230, 231, 233, 235–237, 245, 252, 285, 287–290, 293, 361, 378–384, 396, 400, 405

SSB spontaneous symmetry breaking. 1–3, 19–21, 23, 25, 28, 30, 144, 198, 254, 255, 407

sTGC small-strip Thin Gap Chamber. 61, 62, 64, 154

TASSO Two Arm Spectrometer SOlenoid. 16

TDAQ Trigger and Data AcQuisition. 64, 65, 67, 68

TGC Thin Gap Chamber. 60, 61, 64, 67, 154

TMVA Toolkit for MultiVariate data Analysis. 86, 112

TN true negative. 82

TP true positive. 82

TR transition radiation. 47, 114

TRT Transition Radiation Tracker. 44, 47, 101, 102, 104–106, 113, 114

TTVA track-to-vertex association. 154, 219, 226

UCB upper confidence bound. 90, 91, 300

UV ultraviolet. 47, 54, 93, 184, 281

VBS vector boson scattering. 2, 3, 30, 31, 144, 145, 147–149, 162, 167, 169, 178, 179, 183, 184, 187–189, 192, 193, 196, 198, 203–205, 229, 252, 254, 255, 257–259, 264, 265, 276, 286, 288, 290, 293, 306

vdM van der Meer. 65, 66, 219

VEV vacuum expectation value. 1, 19, 20, 22, 23, 259, 264, 265, 406

VR validation region. 192, 193, 196

W tungsten. 47, 50, 56, 57, 61

WIMP weakly interacting massive particle. 406

WP working point. 3, 112, 113, 116, 117, 119–122, 124, 129, 131, 132, 134, 136–138, 140–143, 151, 152, 154, 161, 162, 164, 165, 189–192, 218, 255

Xe xenon. 47

Λ Acknowledgements

Ce manuscrit conclut trois ans de travail et de recherche en physique des particules. Il s'inscrit cependant dans la continuité d'une démarche bien plus longue, qui remonte au lycée, quand j'étais en seconde. En effet, une année auparavant, en troisième, j'ai souhaité comprendre ce qui faisait réellement la différence entre du bois et du fer. Cela m'a conduit à descendre de plus en plus bas dans l'échelle des tailles, jusqu'aux molécules, puis aux atomes et enfin aux quarks. La découverte de ce monde infiniment petit était également infiniment intéressante, et je me suis pris de passion pour ce nouveau pan de la nature qui s'ouvrait à moi, me renseignant autant que possible à son sujet. J'ai su à ce moment-là que je voulais devenir un expert, ou du moins tout savoir à ce sujet. C'est donc en seconde, au moment de choisir ma voie d'études, que je me suis renseigné sur la carrière à envisager. Le métier de chercheur en physique (des particules, plus précisément, vous l'aurez compris) est alors apparu comme une évidence, et cela signifiait que je devais obtenir le plus haut diplôme de l'enseignement supérieur : le doctorat. Si vous lisez ceci, j'ai donc mené à bien cette très longue quête de plus de dix ans.

La première personne que je voudrais remercier, c'est donc mon jeune moi-même, qui a eu soif de savoir et qui a persévéré pour trouver les réponses à ces questions sur l'infiniment petit. Comme l'a dit Socrate : *« Ce que je sais, c'est que je ne sais rien »*. En emmagasinant toujours plus de savoir sur la physique des particules, je me rends compte que je suis loin de tout savoir à ce sujet et que je ne le saurai probablement jamais, mais c'est ce qui me donne envie de continuer. Je pense cependant être en mesure de répondre à toutes les questions que le jeune Mathis pouvait se poser et donc, en un sens, j'ai accompli mon objectif principal : comprendre ce qui nous compose.

Cette longue épopée, je ne l'ai pas accomplie seul, et j'ai une liste presque aussi longue de personnes à remercier.

Ma famille

En premier lieu, je voudrais remercier toute ma grande famille. Mes parents, évidemment, sans qui je ne serais physiquement pas là, mais également pour tout leur soutien depuis que je suis tout petit et pour m'avoir permis de faire les études que je souhaitais sans que j'aie à me soucier de problèmes financiers, même lorsque je suis parti étudier à Lyon. Je remercie mon père Jean-Christophe pour m'avoir transmis sa passion pour l'informatique et la logique de la programmation ainsi que de bonnes jambes pour le vélo, et ma mère Nathalie pour m'avoir transmis sa passion pour la lecture, la rigueur et la persévérance. Je remercie aussi mes deux petits frères, Maël et Maevan : étant leur aîné, j'ai toujours voulu montrer l'exemple et ne pas les décevoir, et cela m'a permis d'accomplir énormément de choses. Je remercie ensuite mes grands-parents, Mamie Marina, Papi Olivier et Mamie Marie-France, Papi Christian et Mamie Laurence. Ils m'ont toujours soutenu et encouragé dans mes projets, et je chérirai toujours les vacances scolaires que j'ai passées chez eux (même si cela signifiait faire des cahiers de vacances ou des dictées). Je remercie également mon Tonton Fred, Tatie Christiane et Tonton Jean-Marc, Tatie Marie-Jo et Tonton Claude, ainsi que leurs enfants, mes cousins et cousines, Rudi, Lise, Teddy, Carole et Nicolas (Φ), qui ont eux aussi été de précieux soutiens, toujours présents pour les bons moments.

Je voudrais aussi remercier mon Papi Michel, qui n'a malheureusement pas pu voir l'aboutissement de ce travail, mais à qui j'ai dédié la citation de ce manuscrit. C'est une phrase qu'il me répétait très souvent lorsque je lui parlais de mon souhait de devenir chercheur, et qui résonne encore aujourd'hui en moi.

Mes amis

Il y a la famille que l'on a et la famille que l'on choisit. Ces longues années d'études ont été accompagnées d'innombrables rencontres sur les bancs de l'école, et certaines d'entre elles se sont transformées en véritables amitiés.

Le premier est le plus ancien de tous : Maxence, rencontré au CP et, selon moi, véritable expert en mathématiques. Je le remercie pour l'inspiration qu'il m'a donnée, et me donne encore, dans cette matière. Zacharie, ou plus communément Zac, est mon partenaire de toujours quand il s'agit de faire quelque chose d'insensé. Parmi la longue liste, je retiendrai les innombrables nuits blanches à jouer à la console, un marathon (le plus long du monde) et plusieurs triathlons. Je suppose que la liste va continuer à s'allonger.

Je dois à présent remercier la grande troupe rencontrée à l'université de Bordeaux. Melvin, le mordu de lasers, compagnon sans faille de longues soirées à réfléchir, à travailler sur divers projets ou tout simplement à jouer. Thaïs, ma colocataire à Lyon, grande amatrice d'art, de philosophie et de politique, qui m'a ouvert à ces nombreux domaines et m'y a fait grandir. Grégoire, grand aventurier et fin cuisinier, avec qui je partage autant de points communs que de différences : je ne compte plus les aventures improbables qui nous sont arrivées. Même si nous nous voyons de façon irrégulière, chacune de nos retrouvailles semble prolonger la précédente, comme si aucun temps ne s'était écoulé entre les deux. Je remercie également Lucie, Mathis, Chris, Fabio et Romain, qui ont rendu les longues heures sur les bancs de la fac plus heureuses et plus drôles, et pour toutes nos soirées en dehors.

Enfin, je remercie toutes celles et tous ceux que j'ai rencontrés en chemin.

Le LAPP

Cette thèse a été effectuée au Laboratoire d'Annecy de Physique des Particules (LAPP), dans la magnifique ville d'Annecy. Au sein de ce laboratoire, j'ai également fait un grand nombre de rencontres. En premier lieu, je remercie les services administratifs, qui m'ont toujours aidé, que ce soit pour les démarches ou pour les nombreux problèmes liés aux missions.

I want to thank the many other PhD students and postdocs I met and with whom I shared many adventures. The generation before me: Cervane, Cyann, Daniel, Jérémy, Maël and Tom. My own generation, bound by the same writing timeline: Léo, Margaux, Nathan and Yann. And the generation after me: Hind, Maël, Maxime, Timothy and Ward.

Je voudrais également remercier mon premier collègue de bureau, Léo, avec qui j'ai formé le « bureau de la chocolatine » et partagé de nombreuses bières devant les matchs. I would also like to thank my second office mate, Panagiotis, with whom it was a pleasure to share the last year.

Pour finir avec les doctorants, je tiens à remercier Ugo qui, en plus d'être devenu un ami, est également devenu un coauteur. Outre les nombreuses séances de musculation, je le remercie d'avoir apprécié mon idée de scan et de l'avoir appliquée à ses propres modèles théoriques, ce qui nous a conduits à coécrire un article. Je remercie également pour cela Björn, son directeur de thèse, de nous avoir fait confiance.

Je remercie Jean, ingénieux et toujours au fait des dernières nouveautés en matière d'informatique et d'apprentissage automatique, pour ses conseils et son aide.

I would also like to thank the whole ATLAS group at LAPP for their warm welcome and the many meals they treated us to.

Je souhaite remercier en particulier le groupe diboson, avec qui j'ai évolué depuis mon premier jour au laboratoire. Ses membres m'ont guidé et aiguillé dans ma recherche et ont toujours été de très bon conseil. Je tiens à remercier tout particulièrement Emmanuel, véritable pilier de l'analyse, dont le *framework* robuste et polyvalent m'a permis de travailler efficacement en me concentrant sur la partie apprentissage automatique. J'ai beaucoup appris de lui en rigueur de travail, et les résultats obtenus seraient vraisemblablement de moindre qualité sans lui.

Gli italiani

Vorrei ringraziare moltissimi italiani. Non sapevo che avrei incrociato la strada dell'Italia e dei suoi abitanti. Alle scuole medie avevo persino scelto lo spagnolo come seconda lingua straniera (una scelta non proprio azzeccata, tutto sommato). È successo però che, durante questa tesi, l'italiano, che sto cercando di imparare, abbia occupato una parte importante della mia vita. Innanzitutto vorrei ringraziare Marco, il capogruppo, che è il motivo stesso per cui ho presentato la mia candidatura al LAPP la prima volta. Vorrei inoltre ringraziare Lucia, per aver condiviso con me le sue vaste conoscenze in materia di fisica, Alberto, per aver coordinato l'analisi e per il suo aiuto nella risoluzione dei problemi, e Ludovica, per la sua competenza e il suo aiuto nel gruppo e/γ. Infine, la persona più importante, Laura: inizialmente mia collega al workshop e/γ di Valencia, in Spagna, è diventata molto più di quanto avrei mai potuto

immaginare, un pilastro della mia vita negli ultimi due anni (e oltre), in particolare durante il lungo periodo dedicato alla stesura del manoscritto e agli intensi allenamenti di triathlon. Ringrazio inoltre tutte le persone straordinarie che ho avuto modo di incontrare grazie a lei, ovviamente tutte italiane.

Mon jury

Je remercie les membres de mon jury, Nicolas, Nikola, Corinne, Rodanthi et Gilles, pour leur lecture attentive de mon bien long manuscrit durant l'été. Je les remercie pour leurs précieux commentaires et suggestions d'amélioration, ainsi que pour les échanges que nous avons eus à l'issue de la soutenance. Je remercie aussi Gilles de m'avoir initié à l'enseignement lors de la première année de ma thèse, et pour ses conseils dans cette activité, qui a été très enrichissante.

Mon directeur de thèse

J'ai gardé le meilleur pour la fin, mon directeur de thèse : Olivier. Une thèse est le travail du doctorant, mais elle ne serait rien sans un bon directeur. À la manière de l'électron et du photon de mon chapitre 5, ce sont les deux faces d'une seule et même pièce. Je remercie donc Olivier d'être venu me proposer ce stage de Master 2 lors de mon entretien avec Marco. Je le remercie de m'avoir fait confiance et de m'avoir choisi, financement à la clé. Il m'a toujours soutenu et conseillé sur tous les aspects de la thèse. J'ai pu jouir d'une grande liberté et d'une grande autonomie, mais il a toujours répondu présent quand j'avais besoin de lui. J'ai toujours reçu des suggestions d'amélioration, quelle que soit l'idée farfelue que j'explorais, et je lui en suis infiniment reconnaissant. Je pense que les résultats de cette thèse sont à la hauteur de notre collaboration. Merci !

CERN

I would like to finish by thanking CERN and the people who make it what it is. In my quest to understand particle physics, one place made me dream: CERN. I therefore slowly moved closer to it, from my birthplace in Bordeaux to Lyon for my Master's degree, and finally to Annecy, home of the closest French laboratory to CERN. I had high expectations for this place and it lived up to them. I remember my very first day at this magnificent laboratory. I was overjoyed when I received my badge with my name on it next to the CERN logo: it was a huge accomplishment. I had the unique chance to work there. Many brilliant scientists, including Nobel laureates, have walked the same hallways and eaten in the same restaurants, and I even attended conferences in the very auditorium where the discovery of the Higgs boson was announced to the world. I am very grateful to have been part of this journey, which started long before me and will probably continue long after me, and I am happy to have added my own small piece to the study of the Standard Model, as so many have done here. I enjoyed exploring the impressive wealth of particle physics history that this place has to offer, from its first accelerator, the Synchrocyclotron, to the four large LHC experiments and the Antimatter Factory. I am grateful that such a place exists, where people from all over the world live and work together, across language barriers and politics, towards the common goal of understanding the laws of physics.

Pasta alla carbonara

As a final addition to these long (but, to me, important) acknowledgements, I would like to share with you, who have read this far, my state-of-the-art recipe for *pasta alla carbonara*. This comfort recipe has been like my journey in physics, a constant search for improvement, until I reached the perfect result the night before my defence. It has accompanied me since I left home in Quinsac, and over the five years since, I have refined it through many trials until it reached its final form. This dish has always brought a moment of joy in difficult times, and I wish everyone who tries it the same happiness it brought me.

Let N be the number of people, and take $N \times [200, 250]\,\mathrm{g}$ of pasta. I recommend a cylindrical pasta maximising the perimeter $2\pi r$ of its cross-section, which holds the sauce. For this I advise *pici*, but they are difficult to find outside Italy, so spaghetti is the common choice. Cook them for the required time. In parallel, cut the guanciale into parallelepipeds $1\,\mathrm{cm}$ to $2\,\mathrm{cm}$ long and cook them until the fat turns transparent. Take them out of the pan and set them aside. In a bowl, put i eggs, with $i \in \{2N-1, 2N\}$, of which a whole eggs and b yolks only, defined as:

$$a(i) = \left\lceil \frac{i}{2} \right\rceil ,$$
$$b(i) = i - a(i) .$$

Add grated pecorino to the bowl until the sauce is viscous enough to flow like honey rather than like water. You can add pepper while mixing. The guanciale should already bring enough salt. Once the pasta is cooked, and this is a crucial step, add $i + 1$ tablespoons of the pasta cooking water to the egg bowl. This prevents the eggs from turning into an omelette when the pasta is added. Stir the eggs and pecorino sauce while adding this pasta water, then add the cooked pasta and the cooked guanciale. *Et voilà ! Bon appétit !*

Le lecteur / The reader

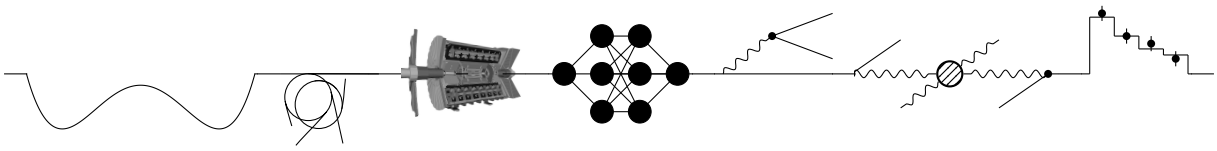
Merci de m'avoir lu jusqu'ici. Si vous vous apprêtez à lire ma thèse, tout d'abord merci, et surtout bon courage. J'espère qu'elle vous apportera les réponses que vous cherchez. Si vous l'avez déjà lue, je vous remercie d'avoir pris le temps de la lire et j'espère qu'elle vous a plu.

Thank you for reading this far. If you are about to read my thesis, first of all thank you, and above all, good luck. I hope it provides the answers you are looking for. If you have already read it, thank you for taking the time to read my work, and I hope you enjoyed it.

Multivariate study of polarisation states in electroweak vector boson scattering WZ events using ATLAS data at the LHC

Abstract. The spontaneous breaking of the electroweak symmetry gives the W and Z bosons a longitudinal polarisation state, in which they behave at high energy like the Goldstone bosons of the unbroken theory. Their scattering is therefore a unique probe of the Standard Model. This thesis presents the first measurement of the longitudinal polarisation states of one or both bosons in WZjj-EW processes, using $140\,\mathrm{fb}^{-1}$ of ATLAS Run-2 data at $13\,\mathrm{TeV}$ combined with $164\,\mathrm{fb}^{-1}$ of Run-3 at $13.6\,\mathrm{TeV}$. Deep neural networks separate the signal from the background and then disentangle the polarisation states. Their outputs define a two-dimensional discriminant plane, in which an optimised Voronoi binning maximises the sensitivity of a binned maximum-likelihood fit. The fraction of events in which the Z boson is longitudinally polarised is measured to be $f_{\mathrm{X0}} = 0.324^{+0.092}_{-0.085}$, an observed (expected) significance of 4.0 (2.9) standard deviations and the first evidence for such states in WZjj-EW. Upper limits at $95\,\%$ confidence level are set on the fractions in which both bosons are longitudinal ($f_{00} < 0.12$) and in which the W boson is ($f_{0\mathrm{X}} < 0.19$). The signal strength is $\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}} = 1.06 \pm 0.18$ and the fiducial cross-section is reported at $13.6\,\mathrm{TeV}$ for the first time. A first calibration and identification of forward electrons with the new Inner Tracker (ITk) is also presented, in preparation for Run-4.

Keywords: WZjj-EW, electroweak, vector boson scattering, ATLAS, machine learning, polarisation, LHC.



Étude multivariée des états de polarisation dans les événements WZ de diffusion de bosons vecteurs électrofaibles dans les données d'ATLAS au LHC

Résumé. La brisure spontanée de la symétrie électrofaible confère aux bosons W et Z un état de polarisation longitudinale, dans lequel ils se comportent à haute énergie comme les bosons de Goldstone de la théorie non brisée. Leur diffusion constitue ainsi une sonde unique du Modèle Standard. Cette thèse présente la première mesure des états de polarisation longitudinale de l'un ou des deux bosons dans les processus WZjj-EW, à partir de $140\,\mathrm{fb}^{-1}$ de données ATLAS du Run-2 à $13\,\mathrm{TeV}$, combinées à $164\,\mathrm{fb}^{-1}$ du Run-3 à $13{,}6\,\mathrm{TeV}$. Des réseaux de neurones profonds séparent le signal du bruit de fond puis dissocient les états de polarisation. Leurs sorties définissent un plan discriminant à deux dimensions, dans lequel une segmentation de Voronoï optimisée maximise la sensibilité d'un ajustement par maximum de vraisemblance. La fraction d'événements où le boson Z est polarisé longitudinalement vaut $f_{\mathrm{X0}} = 0{,}324^{+0{,}092}_{-0{,}085}$, soit une significance observée (attendue) de $4{,}0$ $(2{,}9)$ déviations standards : la première mise en évidence de tels états en WZjj-EW. Des limites supérieures à $95\,\%$ de niveau de confiance sont posées sur les fractions où les deux bosons sont longitudinaux ($f_{00} < 0{,}12$) et où le boson W l'est ($f_{0\mathrm{X}} < 0{,}19$). La force du signal vaut $\mu^{\mathrm{fid.}}_{\mathrm{WZjj\text{-}EW}} = 1{,}06 \pm 0{,}18$ et la section efficace fiducielle est rapportée pour la première fois à $13{,}6\,\mathrm{TeV}$. Une première calibration et identification des électrons vers l'avant avec le nouvel Inner Tracker (ITk) est également présentée, en préparation du Run-4.

Mots-clés : WZjj-EW, électrofaible, diffusion de bosons vecteurs, ATLAS, apprentissage automatique, polarisation, LHC.