Search for the Higgs boson decay into a pair of taus in ATLAS

Faten Hariri

NNT: 2015SACLS065

THESE DE DOCTORAT
DE L'UNIVERSITE PARIS-SACLAY

École doctorale N°576: PHENIICS (Particules, hadrons, énergie et noyaux: Instrumentation, Imagerie, Cosmos et Simulation)

Spécialité physique des particules

Par

M^lle^ Faten Hariri

Recherche de la désintégration du boson de Higgs en deux leptons taus dans l'expérience ATLAS

Thèse présentée et soutenue au LAL, le 30 Octobre 2015
Composition du jury:

M. Maltoni Fabio	Prof., Chef du groupe Madgraph5 Université Catholique de Louvain, CP3	Rapporteur
M. Nakamura Koji	Prof. Assistant KEK, Tsukuba University	Rapporteur
M. Khoo Teng-Jian	College Research Fellow Cambridge University, Cavendish Lab	Examinateur
M. Beaudette Florian	Chercheur CNRS Laboratoire Leprince-Ringuet (LLR)	Examinateur
M. Jaffré Michel	Chercheur CNRS Laboratoire de l'Accélérateur Linéaire(LAL)	Directeur de thèse
M. Cavalier Fabien	Directeur Adjoint, LAL et Directeur-adjoint ED PHENIICS, Paris-Sud	Président du jury

Laboratoire de l'Accélérateur Linéaire n° LAL 15-346

université PARIS-SACLAY | ÉCOLE DOCTORALE PHENIICS

Search for the Higgs boson decay into a pair of taus in ATLAS

by

Faten Hariri

Submitted to the Université Paris-Saclay
on October 30, 2015, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy of Science (Physics)

Abstract

Keywords: ATLAS, Higgs boson, tau lepton, missing transverse energy, Higgs effective field theory (HEFT)

In the LHC project, one of the major goals was the search for the last missing piece of the standard model (SM), namely the Higgs boson (H). The quest was successful during the Run I data taking in 2012 with the discovery of a new scalar of mass ∼125 GeV, compatible with the SM Higgs boson, and decaying to two bosons (either two photons or two electroweak vector bosons ZZ or W^+W^-). To complete the picture, one needed to establish the couplings of the new particle to fermions. This motivated the search for the decay mode into two tau leptons predicted with a high branching ratio.

Inside the ATLAS collaboration, the analysis was divided into three channels according to the decay modes of the tau pair. The work reported in this Ph.D describes the 'lepton-hadron' analysis where one tau lepton decays leptonically into an electron or a muon and the other decays hadronically. Common features of all three analyses are the identification of the hadronic tau lepton and the presence of large missing transverse energy (MET) due to the escaping neutrinos from the tau decays.

An important contribution reported in this dissertation concerns the improvement brought by a new MET determination. By using charged tracks to estimate the contribution of the soft energy component produced in the proton-proton collision, the sensitivity to the overlaid events ('pile-up'), unavoidable in a high luminosity hadron collider, is very much reduced. The systematic uncertainties associated to this soft component were estimated, their dependence on physics modeling and pile-up conditions studied for various track MET definitions. It will contribute to an improved $H \to \tau^+\tau^-$ analysis with future data. In the lepton-hadron H analysis, the dominant background comes from events where a hadronic jet is misidentified as a hadronic tau ('fake-tau'). The work reports in detail how this fake-tau background has been estimated in the two most sensitive event configurations predicted for the H signal i.e. events where the H boson is highly boosted or where it is produced by fusion

of vector bosons (VBF); VBF events are characterized by two forward and backward jets in addition to the H decay products.
Finally, the thesis reports on a last contribution performed with the Higgs Effective Field Theory (HEFT) to study the H couplings and probe new physics beyond SM in a model independent way. The work consisted in testing and validating the 'TauDecay' model in association with the Higgs characterization framework in Madgraph5_aM@NLO. After fixing some encountered problems and implementing a tool to merge H production and decay in a single step (especially useful with NLO requirements), the validation was done in three different ways: direct matrix element generation, with the implemented merging tool and using MadSpin to decay taus. The combined package is ready for use in the LHC Run II context.

Résumé

Titre: Recherche de la désintégration du boson de Higgs en deux leptons taus dans l'expérience ATLAS

Mots clés: ATLAS, boson Higgs, lepton tau, énergie transverse manquante, théorie effective des champs (HEFT)

Au LHC, l'un des buts essentiels était de trouver la dernière pièce manquante du modèle standard (MS), i.e. le boson de Higgs (H). La recherche fut couronnée de succès avec les données prises en 2012 et la découverte d'une nouvelle particule scalaire de masse ∼125 GeV se désintégrant en deux bosons (deux photons ou deux bosons électrofaibles ZZ or W^+W^-). Pour vérifier la compatibilité de la nouvelle particule avec les prédictions du MS, son couplage aux fermions devait être établi, ce qui motiva la recherche du Higgs dans le mode de désintégration en deux leptons taus ayant un rapport d'embranchement important.

Dans ATLAS, cette analyse est divisée en trois canaux selon le mode de désintégration des leptons taus. Le travail présenté dans cette thèse concerne le canal 'lepton-hadron', où l'un des taus de l'état final se désintègre leptoniquement en un muon ou un électron, alors que l'autre se désintègre hadroniquement. Les canaux de l'analyse H→tau tau sont caractérisés par de larges valeurs de l'énergie transverse manquante (MET) dans l'état final et adoptent la même technique pour identifier le lepton tau.

Dans cette thèse, une contribution importante, mettant en relief l'amélioration obtenue avec une nouvelle estimation de MET, est montrée. En utilisant les traces chargées pour estimer la composante 'molle' de MET dans les événements issus de collisions p-p, la sensibilité à l'empilement (pile-up), inévitable dans les collisionneurs hadroniques à haute luminosité, est bien réduite. Les erreurs systématiques associées à la composante molle ont été évaluées et leur dépendance sur les conditions de pile-up et de modélisation de l'événement a été étudiée pour différentes définitions de MET. Ceci contribuera à améliorer les futures analyses H→tau tau.

Dans l'analyse 'lepton-hadron', le bruit de fond dominant provient des événements dont un jet de hadrons est mal identifié comme un tau se désintégrant hadroniquement ('fake tau'). Le travail discuté montre en détail l'estimation de ce bruit de fond pour les deux configurations les plus sensibles aux événements de signal H, i.e. les événements produits avec un Higgs bien boosté ou ceux produits par fusion de deux bosons vecteurs (mode VBF). L'état final de ces derniers est caractérisé par deux jets bien séparés en pseudorapidité, répartis sur les deux hémisphères, produits en association avec les produits de désintégration du H.

Enfin, cette thèse rapporte une dernière contribution utilisant la théorie effective des champs pour estimer la production du boson de Higgs et ses couplages aux particules du MS (HEFT), et explorer la nouvelle physique au delà du MS de façon indépendante du modèle théorique. Le travail consiste à tester et valider le modèle 'tauDecay' dans le cadre d'une caractérisation du Higgs utilisant HEFT au sein de Madgraph5_aMC@NLO. Après avoir écrit un outil permettant de fusionner les fichiers de production et de désintégration du Higgs (utile surtout en travaillant avec une précision au niveau NLO), la validation du modèle a été faite de 3 façons indépendantes: avec la génération d'événements au niveau d'éléments de matrice directement, avec l'outil créé et en désintégrant les taus avec MadSpin. Ce nouvel outil est prêt à être utilisé durant le Run-II du LHC.

Acknowledgements

First, I would like to express my sincere gratitude to the best supervisor ever, Michel Jaffré. Thanks for all the good times, the enjoyable discussions and for sharing your great knowledge with me. Thank you for believing in me, for encouraging and supporting me and for being there for me. You are a great friend. Without you, this journey would not have been possible.

My sincere thanks go to our group leader, David Rousseau, for his endless support on so many levels. You are a true friend and a leader like no other. Thank you for all the good moments we shared together, for all the excellent physics discussions, and for taking care of me like your own student. I could not have made it without you.

I am very grateful to Achille Stocchi and Fabien Cavalier. Thank you for being there for me, for believing in me and supporting me. Special thanks to you, Achille, for your faith in me from the moment I was your student. Without you, this thesis would not have been materialized.

A big thank you to Francois LeDiberder for his great support and encouragement. With you, great opportunities are not a dream but a reality.

Many thanks to convenor Elias Coniavitis. Special thanks to exceptional convenors Koji Nakamura and Teng-Jian Khoo for their help and support, and for all the great discussions.

A huge warm thank you to Fabio Maltoni for believing in me and giving me the opportunity to work with him and the great Madgraph5 team members. You are an excellent role model. Many thanks for the good discussions,for your guidance and for such a pleasurable working experience. Thanks to my good friend Pierre Artoisenet for his great help and support, for all the fruitful discussions and much more. Thank you Kentarou Mawatari for sharing your knowledge with me, your help and for all the good discussions. Thanks to my friends Federico and Ioannis for all the interesting discussions and help, and the memorable good times. Thanks to David, Eleni and Olivier. Thank you Ginette for being such a good, caring friend.

Many thanks to the MCnet community and to the network coordinator and project manager. Thank you for giving me this opportunity to be part of this great family.

Finally, last but most important, I would like to thank God for his great support and love, without which this work would have never been possible and this thesis would have never seen the light. I would also like to thank my dad for his uncondi-

tional love and support on so many countless levels. I am your student academically and in life. Thank you for all the things you've done for me. My love for physics grew through our discussions together and I became quite fond of particle physics thanks to you. Thank you for encouraging me to explore new fields and for your great passion for science that I admire. Without you, I would not be who I am today. And I would like to thank my mom and sisters for their great love that knows no limits and for their great support. Thanks for being there for me throughout this journey. Without you, I would not have made it. You are the wind beneath my wings. You're the reason I have come so far.

Thank you all for being such a blessing in my life.

Author's Contribution

I joined the ATLAS collaboration by the end of 2012 and started as an analyzer in the $H \to WW^* \to \ell\nu\ell\nu$ analysis group, working on top background estimation and signal extraction. Then, I was ready to work on the search for Higgs in the $H \to \tau_{\ell\text{ep}}\tau_{had}$ channel. In both channels, events are characterized by large missing transverse energy 'MET' in the final state, which also motivated the work on a promising pile-up robust MET definition based on track measurements. After exploring the experimental aspects of the search for the Higgs boson in Run-I, I tried probing the spin/CP measurement in the $H \to \tau^+\tau^-$ channel using the Higgs Effective Field Theory (HEFT) approach, which is now the default scheme for Run-II spin/CP studies. A summary of my contribution to ATLAS during my 2-year PhD work period is given below.

Track-based missing transverse energy

I was in charge of the evaluation of systematic uncertainties associated to soft terms of the track-based MET measurement 'track MET' (∼7 months) described in chapter 5. My work includes :

– Evaluating the soft term systematic uncertainties for various MET definitions for both 2011 and 2012 data sets, and studying their dependence on MC generators, pile-up, fast/slow simulation, event kinematics and potential correlations with the calorimeter based MET measurement. I wrote the systematics section of the track MET note :

F. Hariri, C. Lee, B. Liu, R. Mazini, and M. Testa, *Measurement of track-based missing transverse momentum in proton-proton collisions at* $\sqrt{s} = 8$ *TeV centre-of-mass energy with the ATLAS detector*, Tech. Rep. ATL-COM-PHYS-2013-1577, CERN, Geneva, Nov. 2013, https://cds.cern.ch/record/1645897/

– Implementing a user friendly tool for automatic systematics evaluation to be used by analysis groups.

Higgs Analyses: $H \to WW,\ \tau\tau$ channels

- **Work in $H \to \tau\tau$ channel:**

  1. Based on the τ-lepton decay modes (leptonic/hadronic), 3 orthogonal channels are defined: lep-lep, lep-had and had-had. I worked on the lep-had channel, where the major background contribution from $Z \to \tau\tau$ samples and from 'fakes', i.e. events with jets faking hadronic taus. I was in

charge of the evaluation of the fake tau background (fake factor method) and associated uncertainties for 7 TeV analysis, in addition to developing a functional analysis code and implementing the fake factor method in it. Results are explained in chapter 6. Additional details can be found in:

- *Measurement of the Higgs boson couplings in the $\tau\tau$ final state with the ATLAS detector - Supporting Note*, https://cds.cern.ch/record/1666539
- *Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector*, JHEP 1504 (2015) **117**

2. This work was followed by a study of track-based MET for global $H \to \tau\tau$ analysis to show the potential improvement in the lep-lep and lep-had channels as summarized in chapter 6 and explained in:

   F. Hariri, *Track MET for HSG4*, HSG4 workshop, Orsay, June 2014, https://indico.cern.ch/event/324233/contribution/4/material/slides/1.pdf

   This is useful for Run-II analyses in particular since track-based MET is less pile-up dependent than the calorimeter based definition, while giving similar results especially for the final analysis distributions.

- **Work in $H \to WW$ channel:** I worked on optimizing the cut-based analysis used for Higgs signal extraction for both 2011 and 2012 analyses, and succeeded to improve the significance of the standard analysis of the ATLAS HWW group. Results are summarized in Appendix A.

HEFT and Monte Carlo tools

During my 4-month MCnet internship @UCLouvain with the Madgraph5_aMC@NLO team, I worked on the validation of the TauDecay model with the Higgs Characterization (HC) framework, which uses a simple effective field Lagrangian (based on HEFT) below the EW symmetry breaking scale, yet is perfectly suitable to address questions on the strength of the Higgs coupling. The main advantage of HEFT is probing new physics in a model independent way while providing accurante spin/CP measurements. And the angular variables studied have a strong discriminating power regarding :

1. signal/background: $Z \to \tau^+\tau^-$ vs $H \to \tau^+\tau^-$ (important for background reduction in analysis)

2. $0^{\pm}$ spin hypotheses

which is promising for Run-II studies. In addition, a merging tool and an automation of a new MadSpin version were done as well as shown in chapter 7.

Contents

Abstract III

Acknowledgements VII

Author's Contribution IX

List of Figures XVII

List of Tables XXIX

Introduction 1

1 The Standard Model of Particle Physics 5
- 1.1 Introduction . . . 5
- 1.2 EWSB: The Brout-Englert-Higgs Mechanism . . . 8
  - 1.2.1 Fermion Masses in the SM before EWSB . . . 8
  - 1.2.2 The Brout-Englert-Higgs Mechanism . . . 9
  - 1.2.3 The Higgs Particle . . . 12
- 1.3 Higgs Production and Decay Modes at LHC . . . 14
  - 1.3.1 Higgs Production . . . 14
  - 1.3.2 Higgs Decay Modes . . . 14
- 1.4 Discovery of the Higgs Boson . . . 18
- 1.5 BSM, Run-II Prospectives and Conclusions . . . 18

2 LHC and the ATLAS Detector 23
- 2.1 The Large Hadron Collider . . . 23
  - 2.1.1 Purpose . . . 24
  - 2.1.2 Luminosity and Performance . . . 25
  - 2.1.3 LHC Experiments . . . 28
- 2.2 The ATLAS Experiment: Overview . . . 29
- 2.3 The ATLAS Coordinate System . . . 30
- 2.4 The ATLAS Detector . . . 31
  - 2.4.1 The Inner Detector (ID) . . . 31
  - 2.4.2 Calorimeters . . . 35
  - 2.4.3 The Muon Spectrometer . . . 45
  - 2.4.4 The Magnet System . . . 48

2.4.5 Trigger . . . 49
2.5 Conclusion and Prospectives . . . 52

3 Detector Simulation and Physics Modeling with MC Generators 59
3.1 Event Simulation : Introduction . . . 59
3.2 Detector Simulation with GEANT4 . . . 62
3.3 Full Sim Vs ATLFast . . . 63
3.4 Monte Carlo Generators . . . 63
3.5 Parton Distribution Functions and Tuning Parameters . . . 65
3.6 Monte Carlo Pile-up Modeling . . . 65
3.7 Conclusion . . . 66

4 Event and Object Reconstruction 69
4.1 Inner Detector Track Reconstruction . . . 69
4.1.1 Track Parametrization . . . 69
4.1.2 Track Reconstruction Steps . . . 70
4.1.3 Performance . . . 75
4.2 Vertex Reconstruction . . . 77
4.3 Calorimeter Clustering . . . 79
4.3.1 Sliding Window Algorithm . . . 79
4.3.2 Topological Algorithm . . . 82
4.4 Electrons . . . 86
4.4.1 Electron Reconstruction and Identification . . . 86
4.4.2 Gaussian Sum Filter . . . 90
4.4.3 Performance . . . 91
4.5 Muons . . . 92
4.5.1 Muon Reconstruction Efficiency . . . 94
4.5.2 Isolation . . . 95
4.5.3 Scale and Momentum Resolution . . . 96
4.6 Jets . . . 97
4.6.1 Jet Reconstruction and Calibration . . . 98
4.6.2 Pile-up Tools . . . 103
4.6.3 B-tagged Jets . . . 106
4.7 Taus . . . 107
4.7.1 Tau Trigger Operations . . . 108
4.7.2 Tau Reconstruction . . . 109
4.7.3 Discriminating Variables . . . 113
4.7.4 Discrimination Against Jets . . . 113
4.7.5 Discrimination Against Electrons . . . 115
4.7.6 Discrimination Against Muons . . . 116
4.7.7 Performance . . . 118
4.7.8 Energy Scale . . . 119
4.8 Missing Transverse Momentum . . . 120
4.8.1 Hard and Soft Terms Definitions . . . 121
4.8.2 $E_{\mathrm{T}}^{\mathrm{miss}}$ Reconstruction . . . 122

4.8.3 Basic $E_{\mathrm{T}}^{\mathrm{miss}}$ Definitions Used in Higgs Analyses . . . 123
4.8.4 Pile-up Suppression . . . 124
4.8.5 Performance . . . 126
4.9 Conclusion . . . 126

5 Track-based Missing Transverse Energy 133
5.1 Introduction . . . 133
5.1.1 Motivations . . . 134
5.1.2 Performance Limitations . . . 134
5.2 $p_{\mathrm{T}}^{\mathrm{miss}}$ Reconstruction . . . 135
5.2.1 Definitions . . . 135
5.2.2 Why Multiple Definitions ? . . . 137
5.2.3 Including Jets in Track MET-Cl-j: ΔR Vs Ghost Association . 137
5.3 MET Physics Object Selection . . . 138
5.3.1 Basic Track and Primary Vertex Selection . . . 138
5.3.2 Mis-reconstructed Tracks Removal . . . 139
5.3.3 Lepton Selection . . . 139
5.3.4 Jet selection . . . 140
5.4 Estimation of $p_{\mathrm{T}}^{\mathrm{miss}}$ Soft Systematic Uncertainties . . . 140
5.4.1 Soft Term Guidelines . . . 141
5.4.2 Method . . . 141
5.5 Monte Carlo Samples Used . . . 143
5.6 Results . . . 144
5.6.1 7+8 TeV Results with ΔR Method . . . 144
5.6.2 Track MET-Cl-j Systematic Uncertainties Using the Ghost Association Method . . . 157
5.6.3 Extrapolating Beyond $Z \to ll$ Events . . . 157
5.6.4 Influence of Hard Term on $p_{\mathrm{T,soft}}^{\mathrm{miss}}$ Systematics . . . 158
5.6.5 Generator and Simulation Dependence . . . 158
5.7 Application in $H \to WW^*$ Higgs Analysis . . . 169
5.8 Application in $H \to \tau^+\tau^-$ Higgs Analysis . . . 169
5.9 Application in $H \to b\bar{b}$ Higgs Analysis . . . 170
5.10 Conclusion and Prospectives . . . 171

6 Search for the Higgs Boson in the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ Final State 175
6.1 Motivation . . . 176
6.2 Event Experimental Signature and Processes Involved . . . 176
6.2.1 Signal Processes . . . 176
6.2.2 Background Processes . . . 177
6.3 Important Discriminating Variable Definitions . . . 178
6.4 Di-tau Mass Reconstruction in $H \to \tau_\ell\tau_{\mathrm{had}}$ . . . 180
6.4.1 Visible Di-tau Mass $m_{\tau\tau}^{\mathrm{vis}}$ . . . 180
6.4.2 Collinear Mass Approximation . . . 180
6.4.3 Missing Mass Calculator m_{MMC} . . . 182
6.5 Data and Monte Carlo Samples . . . 185

6.5.1 Signal Events . . . 186
6.5.2 Background Events . . . 188
6.6 Object Definitions . . . 188
6.7 Preselection . . . 189
6.7.1 Preliminary Step . . . 189
6.7.2 Preselection Requirements . . . 189
6.7.3 Background Normalization . . . 191
6.8 Event Categorization and Signal Extraction . . . 192
6.8.1 Analysis Categories and Signal Regions . . . 192
6.8.2 Control Regions . . . 192
6.8.3 Control Regions Use for the Final Analysis . . . 194
6.9 Background Estimation Methods . . . 194
6.9.1 OS-SS Method . . . 195
6.9.2 Fake Factor Method . . . 198
6.9.3 $Z \to \tau\tau$ Embedding . . . 210
6.10 Background Estimation . . . 214
6.10.1 $Z \to \tau\tau$ Background . . . 214
6.10.2 Fake Tau . . . 216
6.10.3 W+jets . . . 216
6.10.4 $Z \to \ell\ell$ +jets . . . 216
6.10.5 Top . . . 218
6.10.6 Diboson . . . 218
6.10.7 QCD . . . 218
6.11 Boosted Decision Trees . . . 218
6.12 Results and Systematic Uncertainties . . . 221
6.12.1 Systematic Uncertainties . . . 222
6.13 Track MET for $H \to \tau^+\tau^-$ Analysis . . . 229
6.14 Conclusion and Prospectives . . . 229

7 Higgs Effective Field Theory and Tau Model Validation in Madgraph5_aMC@NLO 239
7.1 Higgs Effective Field Theory (HEFT) . . . 239
7.1.1 Introduction . . . 239
7.1.2 Motivation . . . 240
7.1.3 Linear vs non-Linear Parametrization . . . 241
7.1.4 Building HEFT . . . 241
7.1.5 Effective Lagrangian: Basics . . . 242
7.1.6 HEFT Bases . . . 243
7.2 Importance of HEFT for LHC Run-II . . . 245
7.3 Higgs Properties (Spin/CP) and Monte Carlo Tools . . . 245
7.4 Higgs Characterization (HC) Framework . . . 246
7.4.1 HC Effective Lagrangian . . . 247
7.4.2 Comparison with JHU Results . . . 248
7.5 Application in Higgs Analyses . . . 252
7.6 Tau Model in Madgraph5_aMC@NLO . . . 252

7.7 Testing and Validation of HC+Tau Model in Madgraph5_aMC@NLO 253
7.8 Conclusions and Prospectives 255

Conclusions and Prospectives **263**

A Optimized $H \to W^+W^- \to \ell\nu\ell\nu$ analysis **265**
A.1 Optimized $H \to WW^*$ analysis with track MET 265

B Track MET for the $H \to \tau^+\tau^-$ analysis **279**

C Résumé Substantiel **283**
C.1 Track MET 284
C.2 Analyse $H \to \tau^+\tau^-$ 285
C.3 HEFT 286

List of Figures

1-1 Summary of the SM particle content. 7
1-2 Scalar potential shape for $\mu^2 > 0$ (left) and $\mu^2 < 0$ (right). In the latter, degenerate minima exist with different phases θ [9]. 10
1-3 Summary of Higgs couplings to bosons and fermions in addition to self-interaction terms [10]. 13
1-4 The 4 main Higgs production processes at LHC Run-I. 15
1-5 Cross section of the main Higgs production processes at LHC Run-I at $\sqrt{s} = 8$ TeV [14]. 15
1-6 Feynman diagrams illustrating the Higgs direct decay into a pair of bosons (left) and fermions (right) at leading order. The $H \to \gamma\gamma$ decay is not shown as it is loop induced (see Fig. 1-7). 16
1-7 Leading-order Feynman diagrams showing the Higgs decay into a pair of photons $\gamma\gamma$. 16
1-8 Higgs branching ratios and their uncertainties for the low mass range (left) and for the full mass range (right) [15]. The solid line in deep red for $m_H > 300$ GeVis for the $t\bar{t}$ mode. 18

2-1 Plot showing the LHC design (a) [2] and the various stages of particle acceleration at LHC (b) [5]. The energy values shown are for $\sqrt{s}$= 7 TeV in 2011. 24
2-2 Schematic representation showing the four major experiments and the two ring structure of the LHC. Interaction points (IPs) are labelled based on octant. For example, ATLAS is at IP1 [9]. 27
2-3 Average number of interactions per crossing at LHC in 2011 (blue) and 2012 (green) [11]. 28
2-4 A graphical representation of the various parts of ATLAS detector obtained by computer simulation [15]. 30
2-5 Layout of the ATLAS inner detector [14, 15]. The graphical representation in (b) does not include the additional Insertable B-Layer (IBL), which was added during the 2013-2014 shutdown after Run-I. 32
2-6 Figure showing where the IBL is inserted in the pixel detector [21]. . 33
2-7 Segmentation of the EM calorimeter barrel module in (η, ϕ) [14]. The various parts of the EM LAr calorimeter are shown in (b)[15]. 40
2-8 Jet energy resolution as determined in situ and extracted from Monte Carlo samples for Z-jet events (a) and γ-jet events (b) [28]. 41

2-9 Graphical representation of the tile calorimeter layers [15] (a) and schematic drawing of the tile calorimeter mechanical structure and readout (b) [14]. 42

2-10 Evolution of the percentage of non-reconstructed tile calorimeter faulty cells (masked cells) during Run-I 2011-2012 data taking period (left). The 2D-histogram (right) shows the (η, ϕ) positions of the masked cells at the end of 2012 [36]. The z-axis shows the number of tile cells corresponding to each (η, ϕ) position. 43

2-11 Graphical representation of a HEC module (a) and the corresponding readout electronics (b) [14]. 44

2-12 Schematic representation of the ATLAS muon system in plane containing the beam axis (bending plane). Blue dashed lines show trajectories of muons with infinite momentum [14]. 46

2-13 Figure showing the definition of the sagitta 's' as the maximum deviation of the curved trajectory from a straight line. The definition is shown for 3-point trajectories, which is the case of muon tracks in ATLAS. L is the distance between the end points 1 and 3. 48

2-14 Schematic representation of trigger performance in ATLAS during 2012. Design values are shown in addition to the actual peak values [40]. Items circled in red are the ones that need to be adjusted for Run-II requirements. 50

2-15 Figure showing trigger rates during 2012 data taking in Run-I [41]. . 50

2-16 Figure showing efficiency with respect to offline trigger for electrons (a), muons (b) and jets (c) as a function of the transverse momentum [41]. 51

2-17 Figure showing trigger towers for e/γ reconstruction [14]. 52

2-18 Graphical representation of the muon trigger [14]. High-p_{T} and low p_{T} refer to the L1 muon trigger thresholds. 53

3-1 ATLAS software simulation flow starting with event generators (top left) till reconstruction (top right) [1]. SDOs and RODs stand for Simulated Data Objects and Read Out Driver electronics during the dgitization step. 60

3-2 A graphical representation of the various event generation steps for a $t\bar{t}H$ event [9]. 62

4-1 Track perigee parameters illustrated in the transverse x-y plane (left) and r-z plane (right) as defined in ATLAS. 70

4-2 Sketch illustrating track seed parameters [8]. 71

4-3 Normalized distributions showing track parameter resolution for track seeds matched to the generated particles compared to those of the final tracks (final fit) in minimum bias MC simulation samples [8]. 75

4-4 Plot (a) shows the data and MC track candidates at different stages of the ambiguity solver for primary particles [8]. Plot (b) shows the primary track reconstruction efficiency as a function of the pseudorapidity η in minimum bias MC samples for two selections: default and robust requirements. Distributions are done for different pile-up conditions: no pile-up, average number of interactions $\langle\mu\rangle = 21, 41$ [19]. 76
4-5 2011 Data (closed markers) to MC samples (open markers) comparison of the average number of tracks as a function of the number of interactions per beam crossing for both default (black) and robust (red) requirements is shown in (a). The dashed lines in (a) show the linear fit of the track multiplicity for tracks meeting the robust requirements. Plot (b) shows the transverse impact parameter d_0 distributions for tracks meeting the robust requirements in data for various pile-up conditions [19]. 77
4-6 Plot showing different vertex definitions. 78
4-7 Distribution of the reconstructed number of vertices in various pile-up conditions (a) and the correlation with the average number of interactions per crossing [19, 20]. 78
4-8 (a) shows a qualitative fixed-size cluster representation for a single layer in the EM calorimeter. The sliding window algorithm combines such clusters layer by layer. An example of a topo-cluster is shown in (b). 80
4-9 Resolution of the invariant mass from $Z \to ee$ events after applying all energy corrections in the 2011 data set. The fit shown in red uses a Breit-Wigner convoluted with a Crystal Ball function [15]. 87
4-10 GSF impact on the electron track resolution and impact parameter in simulated MC events [16]. 91
4-11 Measured electron reconstruction efficiencies as a function of E_{T} integrated over the entire pseudorapidity range in (a) and as a function of η for E_{T} between 10 and 50 GeV for 2011 and 2012 data sets [30]. 92
4-12 Measured electron reconstruction efficiencies as a function of the number of primary vertices for 2012 data set is shown in (a). The combined reconstruction and identification efficiency as a function of E_{T} is plotted in (b) for cut-based loose, medium, tight and multilepton (They are mainly loose electrons with cuts on TRT variables used for Higgs→ZZ channel) selections. (b) shows the statistical and systematic uncertainties on the measured efficiencies along with the data to MC efficiency ratio, useful for SF calculation, drawn in the bottom plot [30]. 93
4-13 Chain 1 muon reconstruction efficiency as a function of η measured in $Z \to \mu\mu$ events for various muon reconstruction types. This is done for muons with $p_{\mathrm{T}} > 10$ GeV. CaloTag muons points are limited to the region in η where they are used in physics analyses. The CB + ST efficiency pile-up dependence is shown in (b) [33]. 94
4-14 ID muon reconstruction efficiency as a function η (a) and p_{T} (b) [33]. Pure statistical uncertainties are in green. Combined statistical and systematic uncertainties are shown in orange. 94

4-15 Muon isolation efficiency before (left) and after (right) pile-up corrections for 2011 data sets as done in $H \to WW^*$ analysis for $\Delta R = 0.3$ (top) and in $H \to ZZ$ analysis for $\Delta R = 0.2$ (bottom) with $\Sigma E_T/p_T^{\mu} \leq 0.14$ [41]. 96
4-16 Resolution correction terms for the MS (a) and ID (b) muons. The scale corrections for the MS and ID muons are shown in (c) and (d) respectively [26]. 98
4-17 Invariant mass distribution of Chain1 CB muons in $Z \to \mu\mu$ events before (a) and after (b) applying scale and momentum smearing corrections [26]. 99
4-18 Invariant mass distribution of $Z \to \mu\mu$ events with Chain 1 CB muons(a) [33]. The signal muons are plotted along with the background events. Filled histograms are for MC samples with muon momentum corrections applied, while the dashed histogram shows the MC distribution without these corrections. Plot (b) shows data to MC Chain 1 CB di-muon mass ratio for $Z \to \mu\mu$ (both muons with $p_{\mathrm{T}} > 25$ GeVand same η bin), $J/\Psi \to \mu\mu$ (both muons with $p_{\mathrm{T}} > 6$ GeVand same η bin) and $\Upsilon \to \mu\mu$ (both muons with $p_{\mathrm{T}} > 6.5$ GeV) events [26]. The η of the leading muon is shown. 100
4-19 Overview of the ATLAS jet reconstruction[37]. 101
4-20 Example of an event with random soft emissions at the parton-level clustered into jets using anti-k_T algorithm with $R = 1$[2]. 101
4-21 Data to MC response ratio for anti-k_T (R=0.4) LCW+JES jets in 2012 data as a function of the jet p_{T}. The error bands show the statistical uncertainties as well as the total (systematic and statistical uncertainties added in quadrature) uncertainties for three in-situ jet energy scale methods combined. The plot shows also the results for each method separately [23]. 103
4-22 JES fractional uncertainty at average pile-up conditions as a function of the jet transverse momentum and pseudorapidity [25]. 104
4-23 The hard scatter jet efficiency using JVT and JVF for anti-k_T LCW+JES jets (R = 0.4), with $20 < p_{\mathrm{T}} < 30$ GeV and $30 < p_{\mathrm{T}} < 40$ GeV, as a function of the number of primary vertices in Pythia dijet samples [35]. The inclusive sample efficiency is 90%. 105
4-24 Normalized distribution of the median p_T density ρ for several values of the number of primary vertices, N_{PV}, for $20 < \langle\mu\rangle < 21$ [25]. . . . 106
4-25 A graphical representation of the jet vertex fraction JVF. 107
4-26 Graphical representation of secondary vertex reconstruction: L_{xy} is the distance separating the secondary vertex from the primary vertex in the transverse plane, i.e. plane orthogonal to beam axis and d_0 is the track impact parameter in the transverse plane [40]. 108
4-27 MV1 algorithm light flavor rejection efficiency as a function of b-jet tagging efficiency(a) [29]. The b-jet efficiency p_{T} dependence is shown in (b) [28] for the MV1 algorithm at 70% working point. 109

4-28 (a): b-jet efficiency scale factors for MV1 b-tagging algorithm at 70% working efficiency in $t\bar{t}$ samples [28]. The green band shows total uncertainty (statistical+systematic). The scale factors to correct the efficiency with which the b-tagging algorithms identifies c-jets is shown in (b) [29]. 110
4-29 Diagram showing hadronically decaying tau and QCD jet cones. . . . 110
4-30 Tau trigger rates for 2012 data as a function of the instantaneous luminosity: Level-1 trigger L1(a) and Event Filter(EF) in (b) [32]. . . . 111
4-31 $\tau_{\text{had-vis}}$ track efficiency as a function of the average number of interactions per crossing for 1-prong taus in $Z \to \tau^+\tau^-$ events: the $\tau_{\text{had-vis}}$ candidates are required to have $p_T^\tau > 15$ GeV and match truth-tau within a distance $\Delta R < 0.2$. With TJVA, the efficiency is less pile-up sensitive with a small degradation at high pile-up values [21]. 112
4-32 Graphical representation of tau reconstruction steps summarized. . . 113
4-33 Normalized distributions of f_{cent} in (a), $N_{\text{track}}^{\text{iso}}$ in (b) and R_{track} in (c) in both signal samples ($Z, Z' \to \tau^+\tau^-$ and $W \to \tau\nu$) and background samples obtained using 2012 data. The inverse background efficiency as a function of the signal efficiency is plotted in (d) [32]. 115
4-34 Distributions of f_{HT} and f_{EM} in simulated $Z \to \tau^+\tau^-$ and $Z \to ee$ samples [32]. 116
4-35 Inverse background efficiency($Z \to ee$) as a function of the signal($Z \to \tau^+\tau^-$) efficiency for e-veto for various η regions [32]. 117
4-36 Pseudorapidity η (a) and f_{EM} (b) normalized distributions in reconstructed $\tau_{\text{had,vis}}$ matched to true muons and true taus. $\tau_{\text{had,vis}}$ candidates overlapping with reconstructed muons are discarded [22]. 118
4-37 $\tau_{\text{had-vis}}$ signal (top) and background (bottom) efficiency for 1-prong (left) and 3-prong(right) taus for the different BDT working points for 2012 data, as a function of the true visible tau p_{T} for signal and reconstructed p_{T} for background (QCD from data) [22]. 119
4-38 $\tau_{\text{had-vis}}$ signal efficiency as a function of the number of the vertices in the event for 1-prong (a) and 3-prong (b) taus [22]. 120
4-39 Tau SFs for all tau identification working points combined as a function of η for the muon and electron channels separately and together. The results are derived separately for 1-prong taus (a) and 3-prong taus (b) with $p_{\text{T}} > 20$ GeV. The error bars include combined statistical and systematic uncertainties [32]. 120
4-40 $E_{\text{T}}^{\text{miss}}$ distribution in $Z \to \mu\mu$ samples before (a) and after pile-up suppression with MET-STVF (b) and MET-JetFiltered(c) [24]. 125
4-41 MET resolution as a function of the scalar sum of the measured physics objects E_{T} in the event, ΣE_T, before and after pile-up suppression [24]. 126
4-42 $E_{\text{T}}^{\text{miss}}$ linearity in ggF(gluon-gluon fusion) and VBF(vector boson fusion) Higgs decaying in the $\tau\tau$ mode $H \to \tau^+\tau^-$ ($m_H = 125$ GeV) (b) as a function of the true $E_{\text{T}}^{\text{miss}}$ [24]. 127

5-1 Nominal Track MET distribution in Monte Carlo $Z \to ee$ events before and after the removal of events with Bremsstrahlung electrons using a cut on electron $E_T/p_T < 1.4$, compared to $Z \to \mu^+\mu^-$ results [1]. . . . 135

5-2 Resolutions of nominal calorimeter-based $E_{\mathrm{T}}^{\mathrm{miss}}$ and jet corrected $p_{\mathrm{T}}^{\mathrm{miss}}$ as a function of the number of reconstructed vertices in the event N_{PV} (a) and $\sum E_T$. The resolution is improved when switching to the track-based $p_{\mathrm{T}}^{\mathrm{miss}}$ estimation [1]. 137

5-3 Track p_T distribution in $Z \to ee$ events from 2012 data before and after mis-reconstructed track momentum rejection [1]. 140

5-4 Schematic representation of $p_{\mathrm{T}}^{\mathrm{hard}}$ for the jet corrected definition (right) and the remaining nominal and cluster definitions (left). For the latter, it is sometimes referred to as $p_{\mathrm{T}}^{\mathrm{Lep}}$ and the contribution from tracks associated to jets enters the soft term calculation. 142

5-5 Example showing the Gaussian fit of $p_{\mathrm{T,soft,L}}^{\mathrm{miss}}$ distribution, in a $Z \to \mu\mu$ data (at $\sqrt{s} = 8$ TeV) sample in the 1-jet bin, to extract the scale (mean) of $p_{\mathrm{T,soft}}^{\mathrm{miss}}$. The results are derived for Track MET-Cl-j using the ghost-association method in the first bin of $\mathbf{p}_{\mathrm{T}}^{\mathrm{hard}}$, for events with an average number of interactions per crossing $\langle\mu\rangle$ greater than 22. . 143

5-6 Projection of $p_{\mathrm{T,soft}}^{\mathrm{miss}}$ with respect to $\mathbf{p}_{\mathrm{T}}^{\mathrm{hard,trk}}$ direction to extract $p_{\mathrm{T,soft,L}}^{\mathrm{miss}}$ and $p_{\mathrm{T,soft,P}}^{\mathrm{miss}}$ 144

5-7 Distribution of the average number of interactions per crossing $\langle\mu\rangle$ in $Z \to \mu\mu$ events at $\sqrt{s} = 7$ TeV(left) and 8 TeV (right). The dashed lines show the limits defining the binning in $\langle\mu\rangle$ used for Track MET-Cl-j calculations. 147

5-8 Scale and resolution results for Track MET-Cl in the inclusive $Z \to \mu\mu$ samples. The scale uncertainty defined as the difference(shift) between data and MC $p_{\mathrm{T,soft,L}}^{\mathrm{miss}}$ is plotted in the lower panel of (a). The resolution uncertainty defined as the ratio of data to MC resolution values is plotted for $p_{\mathrm{T,soft,L}}^{\mathrm{miss}}$ and $p_{\mathrm{T,soft,P}}^{\mathrm{miss}}$ in the lower panels of (b) and (c) respectively. 149

5-9 Mean and resolution of $p_{\mathrm{T,soft,L}}^{\mathrm{miss}}$ components of Track MET-Cl-j in $Z \to \mu\mu$ 0-jet(left) and inclusive(right) samples respectively using the ΔR method 150

5-10 Mean and resolution of $p_{\mathrm{T,soft,L}}^{\mathrm{miss}}$ components of Track MET-Cl-j in $Z \to \mu\mu$ 1-jet sample (bottom) and $p_{\mathrm{T,soft,P}}^{\mathrm{miss}}$ component(top) using the ΔR method. 151

5-11 p_T^{hard} distribution in $Z \to \mu\mu$ inclusive sample for Track MET-Cl-j definition. The dotted lines show the binning limits as used for Track MET-Cl and the yellow dashed lines show the adjusted binning for the Track MET-Cl-j calculations. 152

5-12 Mean and resolution of $p_{\mathrm{T,soft,L}}^{\mathrm{miss}}$ components of Track MET-Cl-j in $Z \to \mu\mu$ 0-jet(left) and inclusive(right) samples respectively using the ΔR method at $\sqrt{s} = 7$ TeV. 153

5-13 Mean and resolution of $p_{\mathrm{T,soft,L}}^{\mathrm{miss}}$ components of Track MET-Cl-j in $Z \to \mu\mu$ 1-jet sample (bottom) and $p_{\mathrm{T,soft,P}}^{\mathrm{miss}}$ component(top) using the ΔR method at $\sqrt{s} = 7$ TeV. 154
5-14 Distributions of $p_{\mathrm{T,soft}}^{\mathrm{miss}}$ varied according to equation (5.7) are shown in (a) along with the corresponding total $p_{\mathrm{T}}^{\mathrm{miss}}$ distributions (b) for the $Z \to \mu\mu$ inclusive sample. The results are derived for Track MET-Cl-j derived using ΔR method at $\sqrt{s} = 8$ TeV. 155
5-15 Distributions of $p_{\mathrm{T,soft}}^{\mathrm{miss}}$ varied according to equation (5.7) are shown in (a) along with the corresponding total $p_{\mathrm{T}}^{\mathrm{miss}}$ distributions (b) for the $Z \to ee$ inclusive sample. The results are derived for Track MET-Cl-j derived using ΔR method at $\sqrt{s} = 8$ TeV. 155
5-16 $\sum E_T$ comparisons for different pile-up conditions at $\sqrt{s} = 8$ TeV in $Z \to \mu\mu$ data and AlpgenHerwig simulated samples. The pile-up modeling is in fact studied in terms of the average number of interactions per crossing $\langle\mu\rangle$ (a) and of the number of reconstructed primary vertices in the event N_{PV} (b). 156
5-17 Distributions of ($p_{\mathrm{T,soft,L}}^{\mathrm{miss}}$,$\sum E_T^{\text{soft term}}$) in the inclusive (left) and 0-jet bin (right) $Z \to \mu\mu$ events. 156
5-18 Mean and resolution of $p_{\mathrm{T,soft,L}}^{\mathrm{miss}}$ component of Track MET-Cl-j in $Z \to \mu\mu$ inclusive (top) and 0-jet (bottom) samples. The results are derived for Track MET-Cl-j using the ghost association method. 160
5-19 Resolution of $p_{\mathrm{T,soft,P}}^{\mathrm{miss}}$ component of Track MET-Cl-j in $Z \to \mu\mu$ events in the inclusive (left) and 0-jet bin(right). The mean and resolution of $p_{\mathrm{T,soft}}^{\mathrm{miss}}$ components of Track MET-Cl-j in $Z \to \mu\mu$ events in the 1-jet bin are shown in the bottom plots. The results are derived for Track MET-Cl-j using the ghost association method. 161
5-20 Resolution of $p_{\mathrm{T,soft,P}}^{\mathrm{miss}}$ in $Z \to \mu\mu$ 1-jet bin derived for Track MET-Cl-j using the ghost association method. 162
5-21 Distributions of $p_{\mathrm{T,soft}}^{\mathrm{miss}}$ varied according to equation (5.7) are shown in (a) along with the corresponding total $p_{\mathrm{T}}^{\mathrm{miss}}$ distribution in (b) for the $Z \to \mu\mu$ inclusive sample. The results are derived for Track MET-Cl-j derived using the ghost association method. 162
5-22 Plot showing the difference between Track MET-Cl-j and the true missing transverse energy originating from the neutrino in simulated AlpgenPythia $W \to \mu\nu$ samples at $\sqrt{s} = 8$ TeV. 163
5-23 Effects of $p_{\mathrm{T}}^{\mathrm{hard}}$ component systematics on the measured $p_{\mathrm{T,soft}}^{\mathrm{miss}}$ shift in scale (mean) and resolution (smearing) for Track MET-Cl-j in $Z \to \mu\mu$ samples at $\sqrt{s} = 8$ TeV [10]. 164
5-24 Scale and resolution results for calorimeter based MET before pile-up suppression in $Z \to \mu\mu$ data and AlpgenHerwig samples. 165
5-25 Scale results for calo MET before pileup suppression (top) and track MET (bottom) for $Z \to \mu\mu$ PowhegPythia8 (left) and Sherpa (right) samples. 166
5-26 Resolution plots for calo MET (top) and Track MET-Cl (bottom) for $Z \to \mu\mu$ events with PowhegPythia8 MC simulated samples. 167

5-27 Resolution plots for calo MET (top) and Track MET-Cl (bottom) for $Z \to \mu\mu$ events with Sherpa MC simulated samples. 168
5-28 MET-STVF definition (right) with an illustration of tracks coming from primary (PV) and secondary vertices in an event (left). 170
5-29 $p^{\text{miss}}_{\text{T,soft}}$ (left) and ΣE_T of soft term components (right) distributions for various MET definitions for $Z \to \tau^+\tau^-$ events in the $H \to \tau^+\tau^-$ analysis lep-lep channel. 170

6-1 Figure showing a schematic representation of the angular separation between the final state products and the MET vector, in signal and $Z \to \tau\tau$ lep-had events with additional jet(s) (left) and in W+jets lep-had samples (right). 179
6-2 Figure showing normalized distributions of the signal and $Z \to \tau\tau$ in the VBF signal region. 180
6-3 Figure showing an example of $Z/H \to \tau_\ell\tau_\ell$ decays with collinear mass approximation. The emitted tau decay products are collinear with the tau direction. The MET vector, assuming neutrinos are the only source of MET in the event, is illustrated as well. 181
6-4 Distribution of the ΔR separation between the neutrino(s) and the visible tau decay products in simulated $Z \to \tau\tau$ for hadronic 1-prong(a) and 3-prong(b), in addition to leptonic(c) tau decays for a chosen tau p_T [3] 183
6-5 Plots showing the probability distribution function of the 3-dimensional angular separation $\Delta\theta_{3D}$ between the neutrino(s) and the visible tau decay products in simulated $Z \to \tau\tau$ for hadronic 1-prong(a) and 3-prong(b), in addition to leptonic(c) tau decays [5]. The results are shown for taus with a generated momentum $45 < \text{p} \leq 50$ GeV. 184
6-6 Normalized m_{MMC} distribution in $Z \to \tau^+\tau^-$ and $H \to \tau^+\tau^-$ lep-had events for VBF and boosted analysis categories (defined in sec. 6.8) [2] .185
6-7 Figure showing normalized distributions of the transverse mass at the preselection level for the signal and W+jets samples. 193
6-8 7 TeV FF values for boosted (top) and VBF(bottom) categories as derived in each fake background type CR for 1-prong (left) and 3-prong(right) separately. 201
6-9 8 TeV FF values for VBF(bottom) and boosted (top) categories as derived in each fake background type CR for 1-prong (left) and 3-prong(right) separately [60]. 202
6-10 Graphic representation of the fake tau background composition for VBF $Z \to \tau\tau$ CR (left) and SR (right) at 7 TeV. 204
6-11 Graphic representation of the fake tau background composition for boosted $Z \to \tau\tau$ CR (left) and SR (right) at 7 TeV. 204
6-12 Effective FF value (FF_{mix}) in the boosted (top) and VBF (bottom) signal regions with systematic variations corresponding to the largest R_X variation for 1-prong (left) and 3-prong (right) taus at 7 TeV. . . 205

6-13 Effective FF value (FF_{mix}) in the boosted (top) and VBF (bottom) SR for 1-prong (left) and 3-prong (right) taus at 8 TeV [60]. 206
6-14 Effective FF value (FF_{mix}) in the boosted (up) and VBF (down) $Z \to \tau\tau$ CR for 1-prong (top) and 3-prong (right) taus at 7 TeV. 207
6-15 Fake background distribution for different FF statistical variations as a function of p_T^τ for VBF(left) and boosted(right) events at 7 TeV. . . 208
6-16 Fake background distribution for different FF statistical variations as a function of event BDT score for VBF (left) and boosted (right) events at 7 TeV. The distribution is used for fake tau statistical uncertainty estimations. 209
6-17 Fake background distributions for FF up/down systematic variations (see text) as a function of BDT score for VBF(left) and boosted(right) events at 7 TeV. 210
6-18 Closure plots showing fake tau distribution in W+jets CR in MC samples. 211
6-19 Closure plots showing fake tau distribution in $Z \to \ell\ell$ CR in MC samples. 212
6-20 Closure plots showing fake tau distribution in top CR in MC samples. 213
6-21 FF values at 7 TeV as derived directly from MC simulation samples except for QCD ones for VBF (left) and boosted (right). 214
6-22 Closure test in MC simulated samples boosted (left) and VBF(right) SR at 7 TeV. 215
6-23 Closure test in MC simulated samples boosted SR (right) with MC based MET correction weights (left) at 7 TeV. 215
6-24 Closure test in MC simulated samples boosted SR (right) with data-based MET correction weights (left). 216
6-25 7 TeV closure test of the FF method performed in the MC signal region for the boosted (left) and VBF (right) categories of the $\tau_\ell\tau_{had}$ channel. 217
6-26 Closure test in fake tau SS CR for boosted (left) and VBF(right) SR at 7 TeV. 218
6-27 FF in the SS CR for VBF (left) and boosted (right) events for both1-prong and 3-prong taus at 7 TeV. 219
6-28 7 TeV closure test of the FF method performed in the SS data control region for the boosted (left) and VBF (right) categories of the $\tau_\ell\tau_{had}$ channel. 220
6-29 Best-fit μ values per $H \to \tau^+\tau^-$ channel and per analysis category (VBF, boosted). The green band shows the $\pm 1\sigma$ uncertainty. The contributions of statistical (black), theoretical (red) and non-theoretical (blue) systematic uncertainties are plotted separately [2]. 223
6-30 BDT score distributions for the various $H \to \tau^+\tau^-$ channels, namely $\ell\ell$ (top),ℓh (middle) and hh (bottom) at $\sqrt{s} = 8$ TeV. Post-fit results are shown for VBF (left) and boosted (right) signal regions with statistical and systematic uncertainties. The background predictions as taken from the global fit ($\mu = 1.4$) are shown. And the data to model (background+Higgs signal prediction with strength μ) ratio is shown in the lower panel of each plot for $\mu = 0.0$ (solid black line), $\mu = 1.0$ (dashed red line) and $\mu = 1.4$ (solid red line) [2]. 226

6-31 Event yields as a function of $\log_{10}(S/B)$ for all channels combined ($\ell\ell$, ℓh, hh). The signal (S) and background (B) yields are estimated based on the BDT output bin of each event, with a signal strength $\mu = 1.4$ hypothesis. All categories are taken into account. Background events are displayed for the global fit (with $\mu = 1.4$). Signal yields are shown for both $\mu = 1$ and $\mu = 1.4$ (the best-fit value) at $m_H = 125$GeV. The dashed line corresponds to the background-only distribution obtained from the global fit with $\mu = 0$ [2]. 227

6-32 $p_{\mathrm{T,soft}}^{\mathrm{miss}}$ and $\Sigma E_T^{\mathrm{Soft\ Term}}$ comparison for different MET definitions in $Z \to \tau\tau$(top), lep-had VBF (middle) and lep-lep VBF (bottom) samples. The comparison is done between MET STVF (black), jet corrected track MET $\mathrm{TrkMET}_{c\ell}^{\mathrm{jet\ Corrected}}$(blue) and cluster Track MET (red). . . 232

6-33 Transverse mass distribution in VBF signal and $Z \to \tau^+\tau^-$ events in the lep-lep events with jet corrected track MET used as the MET definition entering the m_T calculation. 233

7-1 Figure showing a 4-fermion effective vertex (left) in a e^+p collision. The actual vertex with the W propagator is shown on the right. . . . 241

7-2 Figure showing the 5 angles fully characterizing the orientation of the decay chain in pp $\to H \to ZZ \to 4\ell^{\pm}$ events as used in the Madgraph5_aMC@NLO plots. The angles are defined in the corresponding particle rest frame. The illustrated production and decay of a H particle are for various spin hypotheses, e.g. spin 0, spin 1 [34]. 249

7-3 Distributions of the X $\to$ ZZ $\to 4\ell$ analysis using the effective Lagrangian as implemented in MG5 [34] for spin 0 (left), spin 1 (middle) and spin 2 (right) hypotheses. The Lagrangian parameter settings are shown on the plots. 250

7-4 Mass and angular distributions for the X $\to$ WW analysis using the effective Lagrangian as implemented in MG5 [34] for spin 0 (left), spin 1 (middle) and spin 2 (right) hypotheses. 251

7-5 Angular distributions $\Phi = \phi_1 + \phi_2$ and $\Delta\phi = \phi_1 - \phi_2$ for events produced directly at the matrix element (ME) level (top), obtained after merging X(J^P) production and decay MG5 output (middle) and decayed with MadSpin (bottom). Plots are obtained with 100k events generated in Madgraph5_aMC@NLO. 256

7-6 The coordinate system in the tau decay frame [51]. 257

7-7 Invariant mass distributions with 10 000 events (no cuts applied) for the a_1 and ρ tau decay modes in the 1-prong (a and b) and 3-prong (d) topologies. The pion pair $\pi^-\pi^0$ and $\pi^-\pi^+$ invariant mass distributions for the a_1 1-prong and 3-prong topologies are shown in (a) and (c) respectively. 257

7-8 Angular distributions $\Phi = \phi_1 + \phi_2$ and $\Delta\phi = \phi_1 - \phi_2$ obtained with new MadSpin to decay tau leptons. 258

A-1 Graph showing the soft hadronic recoil (in gray) against the dilepton system (in yellow) used in the calculation of f_{recoil} for the Z/Drell-Yan events. 267
A-2 Normalized distributions of $E_{\text{T,Rel,Cl}}^{\text{miss,track}}$ for the combined Higgs signal and total background events in the 1-jet bin, illustrating how the Higgs phase space is closed in $E_{\text{T,Rel,Cl}}^{\text{miss,track}}$ with finite limits. The results are shown for the 2012 $\ell\nu\ell\nu$ analysis 268
A-3 Plots showing significance for lower limit cut on $p_T^{\ell\ell}$ in the 0 jet bin before (left) and after (right) applying an additional upper limit cut in the 2012 $\ell\nu\ell\nu$ analysis. 269
A-4 Normalized distributions of METRel (calorimeter-based MET) at preselection for Higgs and background events in the SF (left) and OF (right) channels at the end of preselection in the 2012 $\ell\nu\ell\nu$ analysis. The pink lines show the limits taken into account during optimization for the definition of a METRel window cut. 271
A-5 Normalized Track METRel distributions for the Higgs and background events in the 0 jet bin (top) and 1 jet bin (bottom) with optimized cut limits for the 2012 $\ell\nu\ell\nu$ analysis. The pink lines show the limits taken into account during optimization for the definition of a track METRel window cut. 272
A-6 $p_T^{\ell\ell}$ distribution in the 1-jet bin for all channels (OF+SF) combined in the 2012 $H \to WW^*\ell\nu\ell\nu$ analysis. The pink lines show the proposed cut limits to improve the significance and enchance the obtained Higgs signal. 272
A-7 Distributions of $p_T^{\ell\ell}$ (left) and Track METRel (right) for events in the 0 jet bin kept in the ATLAS analysis but removed in the optimized analysis for the 2012 $\ell\nu\ell\nu$ analysis. 273

B-1 $p_{\text{T,soft}}^{\text{miss}}$ (left) and ΣE_T of soft term components (right) distributions for various MET definitions for $Z \to \tau^+\tau^-$ (top) and ggF(bottom) events in the lep-lep channel. 280
B-2 $p_{\text{T,soft}}^{\text{miss}}$ (left) and ΣE_T of soft term components (right) distributions for various MET definitions for lep-lep VBF(top) and lep-had ggF (bottom) events. 281
B-3 $p_{\text{T,soft}}^{\text{miss}}$ (left) and ΣE_T of soft term components (right) distributions for various MET definitions forVBF lep-had events 282
B-4 m_{MMC} plots for gFF(left) and VBF(right) lep-had events 282

List of Tables

1.1 Table giving a brief description of the SM symmetries. C and Y refer to the color quantum number and the hyper charge respectively. The suffix 'L' in $SU(2)_L$ stands for left-handed, where weak interactions are limited to left-handed fermions only. 6

1.2 Summary of the SM lepton generations and their basic properties [6]. I_3 refers to the weak isospin and the electric charge is given in units of [e]. 6

1.3 Summary of the SM quark generations and their basic properties [6]. I_3 refers to the weak isospin and the electric charge is given in units of [e]. 7

1.4 Table summarizing the basic properties of the SM gauge bosons and the associated interactions [6]. 7

1.5 Relative parametric (PU) and theoretical (THU) uncertainties on the SM Higgs partial widths for a selection of Higgs masses [15]. The PU are shown for each single parameter. 17

1.6 Table summarizing the signal strengths and statistical significance values in ATLAS for each Higgs search channel studied during Run-I at both $\sqrt{s} = 7$ and 8 TeV[19]. 19

2.1 Table showing 2012 beam parameters values justifying the choice of 50 ns bunch crossing over the design value of 25 ns [8]. 25

2.2 Table showing actual and design values of beam performance parameters in 2012 [8]. 26

2.3 Table summarizing the track-to-hit pull widths before and after scaling in addition to the corresponding C values for various parts of the ATLAS Inner detector [38]. 35

2.4 Table summarizing the measured values of the constant EM calorimeter energy resolution term for various η regions in ATLAS [30]. OW and IW stand for outer wheel and inner wheel in the end caps respectively. 38

2.5 The main parameters of the ATLAS muon system for both combined and stand-alone performances [37]. 45

4.1 Table summarizing the values of tracking cuts of the inside-out algorithm. (*) For full track reconstruction, selected tracks at the final stage are good tracks with at least 1 pixel hit [9, 12]. 73

4.2 Table showing qualitatively the effect of some track characteristics on the track score [39]. 74
4.3 EM Tower parametrization in the EM calorimeter barrel (EMB) and end-cap (EMC) [42]. 80
4.4 $\eta - \phi$ granularity of the three EM calorimeter layers [4, 42]. 80
4.5 Cluster size definition for each EM calorimeter layer as used in the sliding window algorithm. The layers are processed to build EM clusters in the order shown in the first column. The dimensions are expressed in tower bin units, i.e. in units of 0.025. $N_{\eta}^{\text{cluster}}$ and $N_{\phi}^{\text{cluster}}$ values depend on the particle type hypothesis and on the region of the EM calorimeter. Their definitions for electrons and photons in the barrel and endcap are given in Table 4.6 [42]. 82
4.6 Cluster size definition $N_{\eta}^{\text{cluster}} \times N_{\phi}^{\text{cluster}}$ for electrons and photons (converted and unconverted) in the barrel and end-caps of the ATLAS EM calorimeter for the sliding window algorithm [42]. 82
4.7 τ_{had} and QCD jet cone comparison. 108
4.8 Uncertainties on some basic tau performance measurements [32]. . . . 121

5.1 Comparison of the different contributions to soft and hard terms for the different track MET definitions when evaluating systematic uncertainties in $Z \to \mu\mu$ samples. 141

6.1 Table showing the m_{MMC} efficiency for $Z \to \tau\tau$ and Higgs signal events for various mass hypotheses, in the $\ell\tau_{\text{had}}$ channel [5]. 184
6.2 Table summarizing the luminosity information for the 7 and 8 TeV data samples [6]. 186
6.3 Table summarizing simulated MC signal and background samples information at $\sqrt{s} = 8$ TeV. The last column states the order of the applied QCD corrections, in addition to showing the product of cross section by the corresponding branching ratio ($\sigma \times B$) for each process, except for those marked with a $\circledast$. The latter are presented with the corresponding inclusive cross section value in the last column [2]. . . . 187
6.4 Table summarizing physics object requirements when entering the analysis and at the preselection level. 190
6.5 Table summarizing preselection cuts in the lep-had channel. 191
6.6 Table summarizing the VBF and boosted categories signal region (SR) and control regions (CRs) definitions. 194
6.7 Table showing the OS and SS normalization factors applied at the preselection level for MC simulated samples at 7 TeV. 198
6.8 Table summarizing the contamination of the electroweak processes in the QCD CR used for r_{QCD} calculations [59]. 198
6.9 Fraction R_X of the different fake background processes as derived in the anti-τ_{had} SR for 2011 and 2012 samples. 203

6.10 Table showing variation of the fake tau composition fraction R_X in the range $[R_X/2,\ R_X \times 2]$ for systematic uncertainties evaluation in VBF events at 7 TeV. 208
6.11 The most discriminating BDT variables for VBF and boosted categories listed in decreasing order of ranking [2]. 222
6.12 Post-fit event yields in the $H \to \tau_\ell \tau_{\rm had}$ channel for $m_H =$ 125 GeV at $\sqrt{s} = 8$ TeV. Background and signal normalizations are post-fit values. Full statistical and systematic uncertainties are shown for the events under 'Total background' and 'Total signal'. The individual background components uncertainties include the systematic uncertainties only [2]. 224
6.13 Table summarizing the expected and observed significances in various $H \to \tau^+\tau^-$ channel for both VBF and boosted categories[2]. 225
6.14 Most important sources of uncertainties affecting the signal strength measurements. Results are shown for the best-fit μ value [2]. 225
6.15 Summary of fake tau background various uncertainties values at 7 and 8 TeV. 228
6.16 Systematic uncertainties post-fit for the signal S and the background B for the 3 $H \to \tau^+\tau^-$ channels and for both VBF and boosted analysis categories at $\sqrt{s} = 8$ TeV. The correlation amongst channels is taken into account and the unaffected uncertainties are marked with a *. The ones with † have an important effect on the BDT distribution shape. UE= underlying event and PS=parton shower [2]. 231

7.1 HC model parameters for the effective Lagrangian [40]. 248

A.1 Table summarizing the cut-based event selection at preselection and in the 0- and 1-jet bin categories for the standard ATLAS Moriond 2013 for the 2012 $H \to WW^*\ell\nu\ell\nu$ analysis. 274
A.2 Table summarizing the optimized cut-based event selection in the 0-jet bin category for the 2012 $H \to WW^*\ell\nu\ell\nu$ analysis. The m_T cut is not shown since the final result of the cut-based analysis will be fitted for final signal extraction. 275
A.3 Table summarizing the optimized cut-based event selection in the 1-jet bin category for the 2012 $H \to WW^*\ell\nu\ell\nu$ analysis. The m_T cut is not shown since the final result of the cut-based analysis will be fitted for final signal extraction. 275
A.4 Table summarizing results of the optimized and standard ATLAS analysis (around Moriond 2013) for the 0 jet and 1 jet bin 2012 analysis. . 275
A.5 Table summarizing results of the optimized and standard ATLAS analysis (around Moriond 2013) for the 0 jet and 1 jet bin 2011 analysis. . 276

Introduction

"I think nature's imagination is so much greater than man's, she's never going to let us relax " - Richard Feynman

" Knowing a great deal is not the same as being smart; intelligence is not information alone but also judgement, the manner in which information is coordinated and used.[...] We make ourselves significant by the courage of our questions, and the depth of our answers." - Carl Sagan, Cosmos

---

The standard model (SM) of particle physics is a simple, yet extremely powerful theoretical model allowing to understand the fundamental physics of our universe. SM predictions have been successfully proven over the past years, from LEP era till now at LHC. During the past few years, a major emphasis in high energy physics was put on the search for the last missing piece of the SM, namely the Higgs boson. The quest was successful during the Run I data taking in 2012 with the discovery of a new scalar of mass $\sim$125 GeV, compatible with the SM Higgs boson, and decaying to two bosons (either two photons or two electroweak vector bosons ZZ or W^+W^-). To complete the picture, one needed to establish the couplings of the new particle to fermions. This motivated the search for the decay mode into two tau leptons predicted with a high branching ratio. Given the large values of missing transverse energy (MET) in the final state and the challenging pile-up conditions of Run-II, having a pile-up robust MET definition is crucial.

After confirming the discovery of a scalar particle compatible with the Higgs boson experimentally by both CMS and ATLAS, it is time to have precise measurements of its coupling strength to the SM particles (including itself). And this is indeed one of the major goals of the LHC Run-II. In order to have accurate measurements, and to probe new physics in a model independent way, using an effective field theory in a framework quantifying deviations from the SM becomes a must.

Thesis Organization

This thesis is divided in two main parts. The first four chapters introduce the theory behind the SM and give a detailed description of the experimental tools and techniques needed for analysis (ATLAS detector, simulation and physics modeling, event reconstruction). The remaining three chapters show the various studies done. Two central chapters cover the experimental/analysis side of the work presented in this thesis, discussing $H \to \tau^+\tau^-$ analysis and a pile-up robust MET definition useful for Run-II. And the last chapter, on the phenomenology/theory side, discusses the Higgs effective field theory (HEFT) and the validation of a Monte Carlo package for simulating tau decays keeping all relevant spin correlations.

Chapter 1 gives the basic theoretical background needed for Higgs physics, along with a brief description of the SM components. In addition to describing the ElectroWeak Symmetry Breaking (EWSB) BEH mechanism, this chapter gives a feeling of the feasible Higgs physics during the LHC Run-I (production and decay modes). The latest ATLAS results for the various search channels are summarized in Table 1.6.

To make physics studies concrete, one needs to understand well how the hadron collider, producing events of interest, works. Also, a knowledge of the detector and its limitations is needed to have correct simulation of background events, correct measurement of detected particles, suitable trigger levels, etc. And this is the main focus of chapter 2. The next steps to perform any analysis would be physics modeling with Monte Carlo (MC) simulation and event reconstruction. Chapter 3 describes the details of event simulation, from the hard scattering process till the detector response needed for event reconstruction. A brief overview of the various MC generators used in this thesis is given at the end of this chapter. On the other hand, the details of reconstruction of standard physics objects used in analyses (electrons, muons, jets...) from the signals read out of the ATLAS detector are given in chapter 4, including a description of the software algorithms that are used and run on both data and Monte Carlo (MC) simulation samples. Track and vertex reconstruction in addition to calorimeter clustering are discussed first. Subsequent sections describe the reconstruction efficiency of physics objects used in analyses such as electrons, muons, jets, taus, and missing transverse energy respectively.

Chapter 5 presents a track-based method to estimate the missing transverse energy, $E_{\mathrm{T}}^{\mathrm{miss}}$. The track-based estimate is indeed a complement to the existing Run-I calorimeter-based measurement of $E_{\mathrm{T}}^{\mathrm{miss}}$, and the equivalent of the default Run-II $E_{\mathrm{T}}^{\mathrm{miss}}$ definition, where the pile-up conditions will become even more challenging than the 2012 Run of LHC. The various definitions of track-based $E_{\mathrm{T}}^{\mathrm{miss}}$ are presented first, with the corresponding object selection requirements. Starting with a pure track-based definition, corrections to account for neutral jet components, accurate electron transverse momentum measurement and mis-reconstructed tracks removal are applied in the subsequent definitions. All definitions are presented since some Higgs analy-

ses use multiple track-based $E_{\mathrm{T}}^{\mathrm{miss}}$ definitions at the same time. The following sections discuss the soft term systematic uncertainties evaluation for various definitions, the event generator dependencies in addition to the correlation between track- and calorimeter-based definitions. This pile-up robust $E_{\mathrm{T}}^{\mathrm{miss}}$ definition is needed in studies with large missing transverse momentum in the final state, e.g. $H \to WW^*$ and $H \to \tau^+\tau^-$.

Chapter 6 presents one the interesting yet challenging channels to look for the Higgs boson i.e. the $H \to \tau^+\tau^-$ channel, where the Higgs boson decays leptonically into a pair of taus with a branching ratio BR $\approx 6.3\%$ for m_{H}=125 GeV. Based on the tau decay mode, i.e. hadronic or leptonic, 3 orthogonal channels are defined: lep-lep, lep-had and had-had. This chapter discusses the $H \to \tau^+\tau^-$ analysis mainly in the lep-had channel. The results are obtained using the full ATLAS data sets during Run-I, i.e. 2012 and 2011 periods corresponding to an integrated luminosity L= 20.3 fb^{-1} at center of mass energy $\sqrt{s} = 8$ TeV and L=4.5 fb^{-1} at $\sqrt{s} = 7$ TeV respectively. Starting with the motivations behind this analysis first in sec. 6.1, the experimental signatures of the $H \to \tau^+\tau^-$ events are presented next in section 6.2. In addition, the description of powerful discriminating variables used in the analysis is given in section 6.3, while the various mass calculation techniques are described separately in section 6.4. The data and Monte Carlo (MC) simulation samples are then described in section 6.5. This is followed by the analysis strategy: the objects definitions and event pre-selection are presented in sections 6.6 and 6.7. Then, event categorization and the analysis details for signal extraction are described in 6.8. In order to have a reliable result, a good background modeling is needed. The background estimation techniques are thus summarized in sec. 6.9 and applied as explained in sec. 6.10. Finally, the final analysis results using boosted decision trees background suppression (explained in sec. 6.11) and the associated systematic uncertainties are shown in sec. 6.12, followed by a study using track MET to improve the analysis results, in preparation for Run-II, as presented in section 6.13.

Finally, chapter 7 presents a brief discussion of the Higgs effective field theory (HEFT) and its importance for Run-II was presented. It can be applied in Monte Carlo simulations through the robust Higgs Characterization (HC) framework. The testing and validation of the tau model within the Monte Carlo generator Madgraph5_aMC@NLO is presented in this chapter. The final test results are done for the $H \to \tau^+\tau^-$ decay mode using HEFT as implemented in the HC framework. This work was done during a short MCnet internship with the Madgraph5_aMC@NLO team at the University of Louvain (UC Louvain).
HEFT is presented first in sec. 7.1, followed by a discussion of its importance for the LHC Run-II in sec. 7.2. Then, the Higgs spin/CP Monte Carlo tools are discussed in sec. 7.3. Afterwards, the HC framework is presented in sec. 7.4, with the results showing its application in various Higgs analyses summarized in sec. 7.5. The testing and validation of the tau model (presented in sec. 7.6) with the HC framework are discussed in sec. 7.7, with the conclusions for Run-II summarized at the end of this chapter (sec. 7.8).

During the Run-II at CERN, the emphasis is on Higgs effective field theory (HEFT) which will replace the κ-framework. The HC allows HEFT studies with NLO precision using Madgraph5_aMC@NLO. Since the Higgs$\rightarrow \tau\tau$ cross section is significantly increased, having a valid working tau model, which can be successfully combined with HC, becomes a must for precision measurements and spin/CP studies. This is now possible and in a user-friendly way.

Chapter 1

The Standard Model of Particle Physics

Most of the experimental observations in particle physics have been successfully described by the Standard Model (SM). In this chapter, a brief overview of the SM is presented since it is the essential basis to the work presented in this thesis, both for the experimental search for the Higgs boson and the physics beyond the SM (BSM) using the Higgs effective field theory.

The phenomenology of the SM is presented first in section 1.1, followed by a brief discussion of the SM electroweak symmetry breaking (EWSB) and the Brout-Englert-Higgs mechanism in section 1.2. The Higgs production and decay modes at the Large Hadron Collider (LHC) for the Run-I studies are then presented in section 1.3 in addition to the discovery of a Higgs like boson in section 1.4. Finally, prospectives for Run-II and conclusions are summarized at the end of this chapter.

1.1 Introduction

Particles physics focuses on the elementary particles constituting the universe and the fundamental interactions between them. The SM [1–4] provides a description of the electromagnetic, weak and strong forces seen in nature and has been so far very successful in describing the observed physics results. The SM in fact unifies these forces in a $SU(3)\otimes SU(2)_L\otimes U(1)_Y$ gauge theory described in Table 1.1, where the simple 'local gauge invariance' requirement explains the fundamental interactions. Combining these symmetries leads to conservation laws (Noether's theorem [5]); each symmetry gives rise to a force mediated by the corresponding spin-1 gauge bosons.

The 'matter' building blocks of the SM are spin 1/2 fermions, which are themselves classified as:

- **Leptons:** they are SU(3) singlets and interact only weakly or electromagnetically.

	$SU(3)_C$	$SU(2)_L$	$U(1)_Y$
Interaction	strong	Electroweak	
Abelian group	no	no	yes
coupling	g_S	g	g'
Conserved quantity	color (C)	weak isospin ($\overrightarrow{I}$)	hyper charge (Y)
Number of Gauge Fields (based on generators)	3^2-1=8 (8 $G^{\alpha}_{\mu\nu}$ in SU(3))	2^2-1= 3 (3 W^{b}_{μ}) b= 1, 2, 3	1 (B_μ)

Table 1.1: Table giving a brief description of the SM symmetries. C and Y refer to the color quantum number and the hyper charge respectively. The suffix 'L' in $SU(2)_L$ stands for left-handed, where weak interactions are limited to left-handed fermions only.

Generation	Leptons	Electric Charge [e]	I_3	Mass
1	ν_e	0	$+1/2$	< 2 eV
	e^-	-1	- $1/2$	511 keV
2	ν_μ	0	$+1/2$	< 2 eV
	μ^-	-1	- $1/2$	107.5 MeV
3	ν_τ	0	$+1/2$	< 2 eV
	τ^-	-1	- $1/2$	1.78 GeV

Table 1.2: Summary of the SM lepton generations and their basic properties [6]. I_3 refers to the weak isospin and the electric charge is given in units of [e].

- **Quarks:** they are SU(3) triplets and can undergo all 3 types of SM interactions, namely electromagnetic, strong and weak interactions.

The leptons and quarks are arranged each into three generations as shown in Table 1.2 and 1.3. In addition, each fermion is paired with its anti-fermion. Only left-handed fermions interact weakly and form $SU(2)_L$ doublets.
The fundamental forces in the SM are :

1. **Electromagnetic interaction:** mediated by the spin-1 massless photon (γ).
2. **Weak interaction:** carried by the $W^{\pm}$ bosons for charged weak currents and by Z^0 boson for the neutral weak current.
3. **Strong interaction:** mediated by gluons. This interaction type is limited to charged SU(3) particles i.e. quarks and gluons.

The SM particles and their basic properties are summarized in Fig. 1.1. An overview of the gauge bosons is given in Table 1.4, in addition to a summary of the SM

Generation	Quarks	Electric Charge [e]	I_3	Mass
1	u	$+2/3$	$+1/2$	2.3 MeV
	d	- 1/3	- 1/2	4.8 MeV
2	c	$+2/3$	$+1/2$	1.275 GeV
	s	- 1/3	- 1/2	95 MeV
3	t	$+2/3$	$+1/2$	173.34 GeV
	b	- 1/3	- 1/2	4.18 GeV

Table 1.3: Summary of the SM quark generations and their basic properties [6]. I_3 refers to the weak isospin and the electric charge is given in units of [e].

Bosons	Interaction	Electric Charge [e]	Color	Mass	Tree-level self-interaction
Photon γ	Electromagnetic	0	-	$< 1 \times 10^{-18}$ eV	no
$W^{\pm}$	weak	± 1	-	80.385 ± 0.015 GeV	yes
Z^0	weak	0	-	91.188 ± 0.0021 GeV	
Gluon g	strong	0	8 combinations	0	yes

Table 1.4: Table summarizing the basic properties of the SM gauge bosons and the associated interactions [6].

symmetries in Table 1.1. Since the work presented in this thesis focuses on Higgs physics, the next sections will describe the SM electroweak (EW) sector, with an emphasis on the Brout-Englert-Higgs mechanism and its search related issues at the LHC. For additional information on the SM, the reader can refer to [1, 3, 4] or any good introductory QFT (Quantum Field Theory) book e.g. [2].

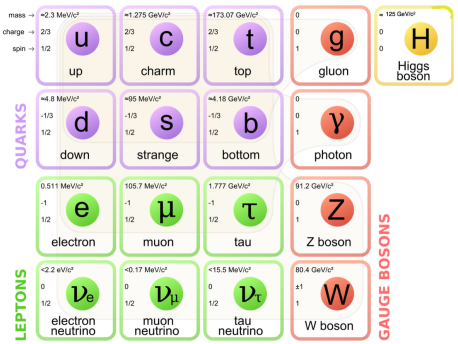


Figure 1-1: Summary of the SM particle content.

1.2 EWSB: The Brout-Englert-Higgs Mechanism

It is observed experimentally that the EW gauge bosons ($W^{\pm}$, Z) and fermions are massive. However, under SU(2) gauge symmetry, the gauge bosons are massless. In addition, a standard mass term of the form $m\bar{\psi}\psi$ for fermions in the Lagrangian is not invariant under SU(2) gauge-symmetry and the theory will not be renormalizable if it was included. As a result, the mass of gauge bosons is then introduced by electroweak symmetry breaking (EWSB) through the Brout-Englert-Higgs (BEH) mechanism [7, 8] and the fermions acquire mass through Yukawa couplings to the Higgs boson. A brief overview of the BEH mechanism is presented in sec. 1.2.2 after the discussion of the fermion masses in sec. 1.2.1.

1.2.1 Fermion Masses in the SM before EWSB

The covariant derivative describing the fermions, based on the SM electroweak sector before symmetry breaking, has the general form:

$$D_\mu = \partial_\mu - igA^a_\mu T^a - ig'YB_\mu \ , \tag{1.1}$$

where:

- T^a are the SU(2) generators: $T^a = \sigma^a/2$ with σ's being the Pauli spin matrices.
- Y refers to the weak hyper charge (U(1)).
- A^a_μ and B_μ are the vector fields for SU(2) and U(1) respectively.

In terms of mass eigenstates, D_μ is expressed as:

$$D_\mu = \partial_\mu - \frac{ig}{\sqrt{2}}(W^+_\mu T^+ + W^-_\mu T^-) - i\,\frac{g^2T^3 - g'^2Y}{\sqrt{g^2+g'^2}}Z_\mu - i\,\frac{gg'}{\sqrt{g^2+g'^2}}(T^3+Y)A_\mu \ , \tag{1.2}$$

where:

- The physical W bosons are obtained from linear combinations of W^1 and W^2

$$W^\pm_\mu = \frac{1}{\sqrt{2}}(W^1_\mu \mp iW^2_\mu). \tag{1.3}$$

- The A_μ and Z_μ fields, matched to the photon and Z^0 bosons, are defined in terms of the neutral fields W^3(SU(2)) and B_μ ($U(1)_Y$) as:

$$\begin{pmatrix} Z^0 \\ A \end{pmatrix} = \begin{pmatrix} \cos\theta_w & -\sin\theta_w \\ \sin\theta_w & \cos\theta_w \end{pmatrix} \begin{pmatrix} W^3 \\ B \end{pmatrix}.$$

- Q=T_3+Y is the electric charge operator (Q = -1 for electrons) and $e = \dfrac{gg'}{\sqrt{g^2+g'^2}}$.

  The derivation of the basic entities is given in sec. 1.2.2.

The left-handed leptons of each generation are arranged in doublets e.g.

$$E_L = \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L ,$$

with $Y_{\nu_e} = Y_{e_L}$= -1/2. On the other hand, for right-handed leptons (no right-handed neutrinos in the SM) form a singlet each e.g. e_R with $T_3 = 0$ and Y_{e_R} = -1. Similarly, left handed quarks of each generation form a doublet with Y_{Q_L} = -1/6 e.g.

$$Q_L = \begin{pmatrix} u \\ d \end{pmatrix}_L \qquad \text{(for the first generation)},$$

and the right-handed quarks form a singlet each with Y_{q_R} = quark charge ($Y_{u_R} = 2/3$ and Y_{d_R} = -1/3). The Lagrangian terms involving fermions based on SU(2)⊗U(1) gauge theory are then of the form

$$\begin{aligned} \mathcal{L} \;=\; & \overline{L}(i\partial\!\!\!/)L + \overline{\ell}_R(i\partial\!\!\!/)\ell_R + \overline{Q}_L(i\partial\!\!\!/)Q_L + \overline{q}_R(i\partial\!\!\!/)q_R \\ & + g(\mathrm{W}^+_\mu J^{\mu+}_W + \mathrm{W}^-_\mu J^{\mu-}_W + \mathrm{Z}^0_\mu J^\mu_Z) + eA_\mu J^\mu_{\mathrm{EM}} \ , \end{aligned} \tag{1.4}$$

with $L(\ell_R)$ referring to the left-handed (right-handed) lepton doublets (singlets), $J_{Z(W)}$ being the neutral (charged) weak currents and J^μ_{EM} denoting the EM current. The sum over all generations is implied. The fermion mass term cannot take a standard $m\overline{\psi}\psi$ form since this is not invariant under $\mathrm{SU}(2)_L$. The fermion mass terms are generated through Yukawa couplings to the Higgs doublet as described in sec. 1.2.2.

1.2.2 The Brout-Englert-Higgs Mechanism

In order to acquire mass, the non-abelian SM local symmetry needs to be broken such that [9–11]:

1. $\mathrm{W}^\pm$ and Z bosons masses are generated.

2. Photons γ remain massless and QED (Quantum ElectroDynamics) remains an exact symmetry.

3. Renormalization is not violated.

The simplest solution is a doublet of complex scalar fields [4]. For a scalar field ϕ, the Lagrangian and potential are defined as:

$$\mathcal{L} = \partial_\mu\phi^\dagger\partial^\mu\phi - V(\phi) \ , \tag{1.5}$$

$$V(\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2 \ . \tag{1.6}$$

The requirement $\lambda > 0$ is needed in order to have a ground state bounded from below. Depending on the sign of μ^2, two scenarios are possible as shown in Fig. 1-2. Only the case with $\mu^2 < 0$ is of interest since:

1. $\boldsymbol{\mu^2 > 0}$: This is the case of a scalar massive particle with mass μ and the minimum of the potential $V(\phi)$ has only one solution at $\phi = 0$.

2. $\boldsymbol{\mu^2 < 0}$: The potential reaches a minimum for

$$|\phi_0| = \sqrt{\frac{-\mu^2}{2\lambda}} \equiv \frac{v}{\sqrt{2}} > 0\,, \qquad\qquad V(\phi_0) = -\frac{\lambda}{4}v^4\,. \tag{1.7}$$

where v is the vacuum expectation value (vev).

$$\langle\,\Phi\,\rangle_0 \equiv \langle\,0\,|\,\Phi\,|\,0\,\rangle = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix} \quad \text{with} \quad v = \left(-\frac{\mu^2}{\lambda}\right)^{1/2}\,. \tag{1.8}$$

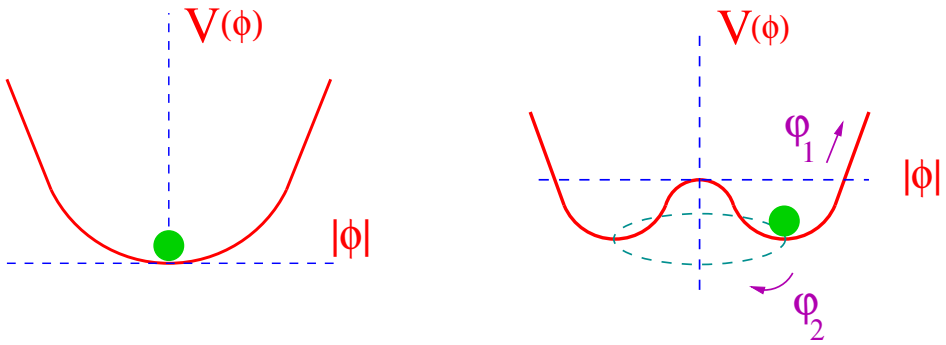


Figure 1-2: Scalar potential shape for $\mu^2 > 0$ (left) and $\mu^2 < 0$ (right). In the latter, degenerate minima exist with different phases θ [9].

The Lagrangian invariant terms corresponding to the Higgs doublet

$$\mathcal{L}_S = (D^\mu\Phi)^\dagger(D_\mu\Phi) - \mu^2\Phi^\dagger\Phi - \lambda(\Phi^\dagger\Phi)^2\,, \tag{1.9}$$

are then added to the SM Lagrangian in eq. (1.10) obtained from SU(3)⊗SU(2)$_L$⊗U(1):

$$\mathcal{L}_{\rm SM} = -\frac{1}{4}W^a_{\mu\nu}W^{\mu\nu}_a - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} + \overline{L}\,iD_\mu\gamma^\mu\,L + \overline{e}_R\,iD_\mu\gamma^\mu\,e_R + \text{strong interaction terms.} \tag{1.10}$$

Denoting the non-commutating 2×2 Pauli matrices by τ^a (a = 1, 2, 3), rewriting Φ in terms of four fields $\theta_{1,2,3}(x)$ and $H(x)$, and expanding around the minimum v at first order, one obtains:

$$\Phi(x) = \begin{pmatrix} \theta_2 + i\theta_1 \\ \frac{1}{\sqrt{2}}(v + H(x) - i\theta_3 \end{pmatrix} = e^{i\theta_a(x)\tau^a(x)/v} \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}}(v + H(x)\,) \end{pmatrix}\,. \tag{1.11}$$

Since the Lagrangian is invariant under SU(2)$_L$, the dependence on θ_i can be removed through rotation. Making a gauge transformation on the field above to have the

unitary gauge:

$$\Phi(x) \rightarrow e^{-i\theta_a(x)\tau^a(x)}\,\Phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \ . \tag{1.12}$$

The kinetic term of the Lagrangian $\mathcal{L}_S$ is then $|D_\mu\Phi|^2$:

$$\begin{aligned} |D_\mu\Phi|^2 &= \left|\left(\partial_\mu - ig_2\frac{\tau_a}{2}W_\mu^a - ig_1\frac{1}{2}B_\mu\right)\Phi\right|^2 \\ &= \frac{1}{2}\left|\begin{pmatrix} \partial_\mu - \frac{i}{2}(g_2W_\mu^3 + g_1B_\mu) & -\frac{ig_2}{2}(W_\mu^1 - iW_\mu^2) \\ -\frac{ig_2}{2}(W_\mu^1 + iW_\mu^2) & \partial_\mu + \frac{i}{2}(g_2W_\mu^3 - g_1B_\mu) \end{pmatrix}\begin{pmatrix} 0 \\ v + H \end{pmatrix}\right|^2 \\ &= \frac{1}{2}(\partial_\mu H)^2 + \frac{1}{8}g_2^2(v+H)^2|W_\mu^1 + iW_\mu^2|^2 + \frac{1}{8}(v+H)^2|g_2W_\mu^3 - g_1B_\mu|^2. \end{aligned} \tag{1.13}$$

Defining the physical fields $W_\mu^\pm$, Z_μ and A_μ as follows, with A_μ being orthogonal to Z_μ :

$$W^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2) \ , \ Z_\mu = \frac{g_2W_\mu^3 - g_1B_\mu}{\sqrt{g_2^2 + g_1^2}} \ , \ A_\mu = \frac{g_2W_\mu^3 + g_1B_\mu}{\sqrt{g_2^2 + g_1^2}} \ , \tag{1.14}$$

the masses of the gauge bosons arise from the bilinear terms in $\mathrm{W}^\pm$, Z and A of the form:

$$M_W^2W_\mu^+W^{-\mu} + \frac{1}{2}M_Z^2Z_\mu Z^\mu + \frac{1}{2}M_A^2A_\mu A^\mu. \tag{1.15}$$

The kinetic part of the scalar Lagrangian $\mathcal{L}_S$ gives thus rise to massive W and Z bosons, while the photon is still massless.

$$M_W = \frac{1}{2}vg_2 \ , \ M_Z = \frac{1}{2}v\sqrt{g_2^2 + g_1^2} \ , \ M_A = 0 \ . \tag{1.16}$$

The Z boson mass is related to the W boson mass though the "weak" or "Weinberg" angle according to:

$$M_Z\cos\theta_W \ = \ M_W \ = \ \frac{1}{2}vg \ . \tag{1.17}$$

The spontaneous symmetry breaking (SSB) of $\mathrm{SU(2)_L}\times\mathrm{U(1)_Y} \rightarrow \mathrm{U(1)_Q}$ lead to three Goldstone bosons, whose contribution can be removed from the Lagrangian using the local gauge invariance of the Lagrangian.
To generate the fermion masses, the same scalar field Φ, with hypercharge $Y = 1$, can be used to write the gauge-invariant fermion-scalar coupling (Yukawa-type Lagrangian), where $\tilde{\Phi} = i\tau_2\Phi^*$ ($Y =$ -1). For first generation fermions, the corresponding Lagrangian is :

$$\mathcal{L}_F = -\lambda_e\,\bar{E}_L\,\Phi\,e_R - \lambda_d\,\bar{Q}_{L,1}\,\Phi\,d_R - \lambda_u\,\bar{Q}_{L,1}\,\tilde{\Phi}\,u_R \ + \ h.c. \ , \tag{1.18}$$

$$\begin{aligned}\mathcal{L}_F &= -\frac{1}{\sqrt{2}}\lambda_e\,(\bar{\nu}_e, \bar{e}_L)\begin{pmatrix} 0 \\ v+H \end{pmatrix} e_R + \cdots \\ &= -\frac{1}{\sqrt{2}}\,(v\,+\,H)\,\left(\lambda_e\,\bar{e}e\,+\lambda_d\,\bar{d}d\,+\,\lambda_u\,\bar{u}u\right) \\ &= -(1+\frac{H}{v})\,\{m_e\,\bar{e}e\,+\,m_d\,\bar{d}d\,+\,m_u\,\bar{u}u\}\;. \end{aligned} \tag{1.19}$$

For each fermion type f, the constant in the mass term $-m_f\bar{f}f$ identifies with the fermion mass

$$m_f = \frac{\lambda_f\,v}{\sqrt{2}} \qquad (f\,=\,e,\,d,\,u)\;. \tag{1.20}$$

The same procedure is applied to the second and third fermion generations. Hence, using the same isodoublet Φ of scalar fields, the masses of weak vector bosons and fermions are generated.

The experimental measurements show that [12, 13]:

$$M_Z = 91.1875 \pm 0.0021 \;\; \text{GeV}\;,\;\; M_W = 80.398 \pm 0.025 \;\; \text{GeV}\;, \tag{1.21}$$

$$\sin^2\theta_W = 1 - \frac{M_W^2}{M_Z^2} \;=\; 0.223\;. \tag{1.22}$$

1.2.3 The Higgs Particle

Rewriting the scalar potential $V(\Phi) = \mu^2\Phi^\dagger\Phi + \lambda(\Phi^\dagger\Phi)^2$ in terms of the Higgs field

$$V \;=\; \frac{\mu^2}{2}(0, v+H)\begin{pmatrix} 0 \\ v+H \end{pmatrix} + \frac{\lambda}{4}\left|(0, v+H)\begin{pmatrix} 0 \\ v+H \end{pmatrix}\right|^2\;, \tag{1.23}$$

and using $v^2 = -\mu^2/\lambda$, the obtained potential expression is:

$$V \;=\; -\frac{1}{2}\lambda v^2\,(v+H)^2 + \frac{1}{4}\lambda(v+H)^4\;. \tag{1.24}$$

The Lagrangian part with Higgs field terms only is:

$$\mathcal{L}_H \;=\; \frac{1}{2}(\partial_\mu H)(\partial^\mu H) - V \;=\; \frac{1}{2}(\partial^\mu H)^2 - \lambda v^2\,H^2 - \lambda v\,H^3 - \frac{\lambda}{4}\,H^4\;.$$

The Higgs mass is extracted from the quadratic term. From this Lagrangian, one can see that the Higgs boson mass simply reads

$$M_H^2 = 2\lambda v^2 = -2\mu^2\;, \tag{1.25}$$

where the Higgs mass and interaction terms are encoded in the potential while the interaction term resides in the covariant derivative.
In addition, the Lagrangian terms involving fermions and gauge bosons are used to

extract the Higgs coupling to these particles:

$$\mathcal{L}_{M_V} \sim M_V^2 \left(1 + \frac{H}{v}\right)^2 \ , \quad \mathcal{L}_{m_f} \sim -m_f \left(1 + \frac{H}{v}\right) \ , \tag{1.26}$$

$$g_{Hff} = i\frac{m_f}{v} \ , \ g_{HVV} = -2i\frac{M_V^2}{v} \ , \ g_{HHVV} = -2i\frac{M_V^2}{v^2} \ . \tag{1.27}$$

Finally, the value of the vacuum expectation value v is dictated by the W boson mass M_W or the Fermi constant G_μ as estimated from the muon decay:

$$M_W = \frac{1}{2} g_2 v = \left(\frac{\sqrt{2} g^2}{8 G_\mu}\right)^{1/2} \Rightarrow v = \frac{1}{(\sqrt{2} G_\mu)^{1/2}} \simeq 246 \text{ GeV} \ . \tag{1.28}$$

The couplings of the Higgs boson to itself, fermions and gauge bosons are summarized in Fig. 1-3.

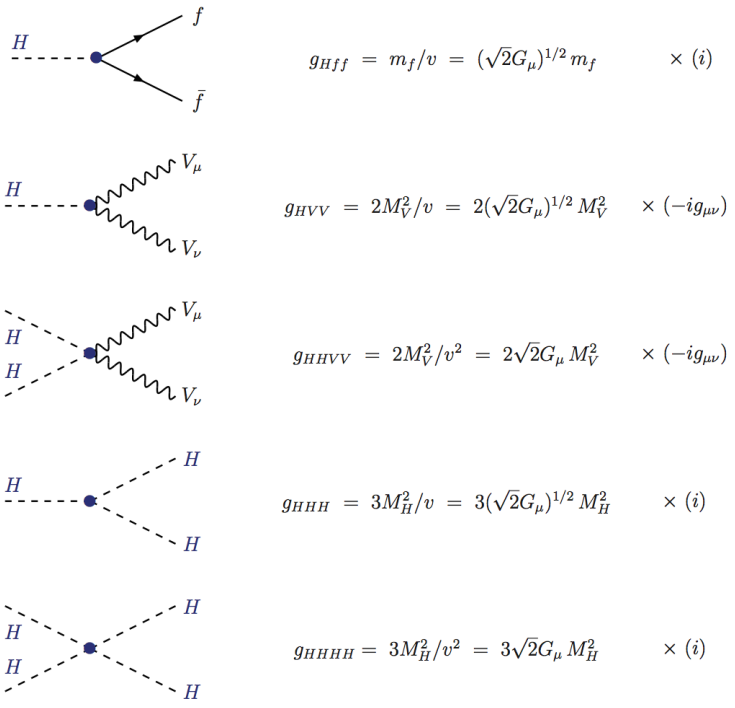


Figure 1-3: Summary of Higgs couplings to bosons and fermions in addition to self-interaction terms [10].

Finally, the Higgs mass is subject to various experimental (e.g. from electroweak precision data, from direct searches at LEP) and theoretical constraints (e.g. unitarity, triviality and stability bounds). The reader can refer to [10] for a thorough treatment.

1.3 Higgs Production and Decay Modes at LHC

1.3.1 Higgs Production

In proton-proton collisions at the Large Hadron Collider (LHC), many initial states can lead to the production of the Higgs boson. The vertices allowed by the SM predictions include Higgs couplings to fermions, W and Z bosons and to itself. The main Higgs production modes at the LHC Run-I energies ($\sqrt{s}$ = 7 TeV, 8 TeV) are:

1. **ggF:** The gluon-gluon fusion (ggF) is the dominant Higgs production mode $gg \to H$. At leading order, no jets are produced along with the Higgs boson, providing thus a clean experimental signature, with challenges limited to the decay mode.

2. **VBF:** Vector boson fusion (VBF) processes have the next highest Higgs cross section: $qq \to qqVV \to qq + H$. These events are characterized by two forward jets well separated in pseudorapidity and with high di-jet invariant mass.

3. **VH:** VH processes refer to associated production with W/Z bosons: $\mathrm{q\bar{q}} \to \mathrm{V{+}H}$

4. **$\mathbf{t\bar{t}H}$**: Associated production with heavy top quarks $gg, \mathrm{q\bar{q}} \to \mathrm{t\bar{t}} + \mathrm{H}$ have the smallest Higgs cross section.

The Feynman diagrams of the different Higgs production processes at leading order are shown in Fig. 1-4. The corresponding cross sections are illustrated in Fig. 1-5. The analyses presented in this thesis have Higgs signal contributions from ggF and VBF events mainly.

1.3.2 Higgs Decay Modes

The same set of allowed vertices in the SM for the Higgs production play a role in the Higgs decay modes. The Feynman diagrams for the Higgs decay into a pair of W or Z bosons, or fermions at leading order are illustrated in Fig. 1-6. The decay into a pair of photons is loop-induced as shown in Fig. 1-7.

The branching ratio (BR) depends on the Higgs mass and the particle to which the Higgs couples, with a preference of heavier particles. The Higgs branching ratios with the associated uncertainties are shown in Fig. 1-8 for the low mass and full mass ranges. The partial widths for the SM Higgs in addition to the corresponding parametric (PU) and theoretical (THU) uncertainties are summarized in Table 1.5, where:

- α_s is the strong coupling parameter.
- m_b, m_c and m_t are the bottom, charm and top quark masses respectively.
- For each parameter, the upper and lower variations are shown.

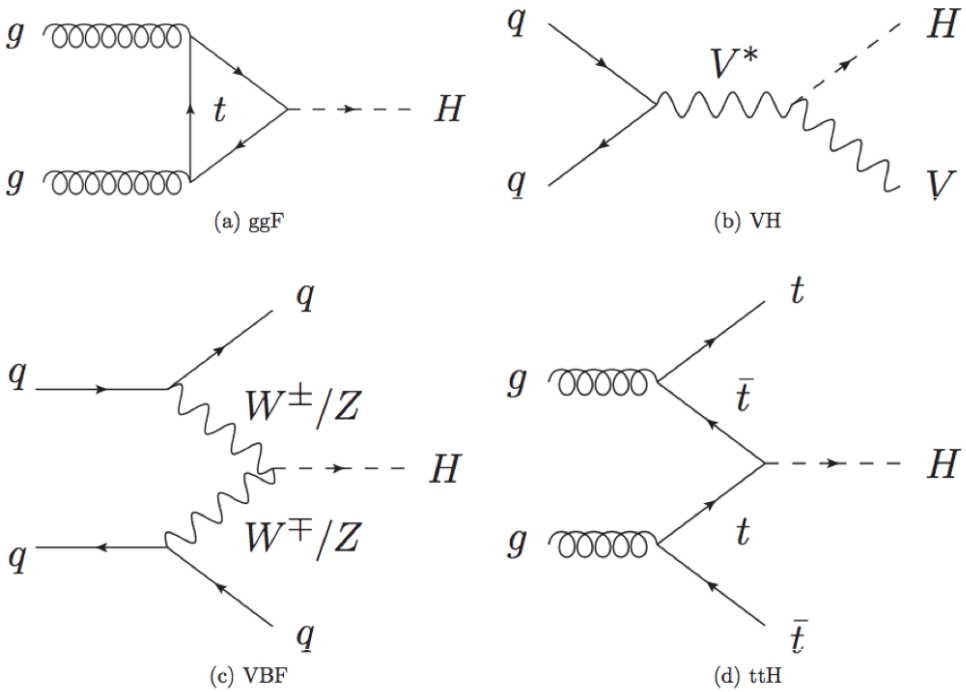


Figure 1-4: The 4 main Higgs production processes at LHC Run-I.

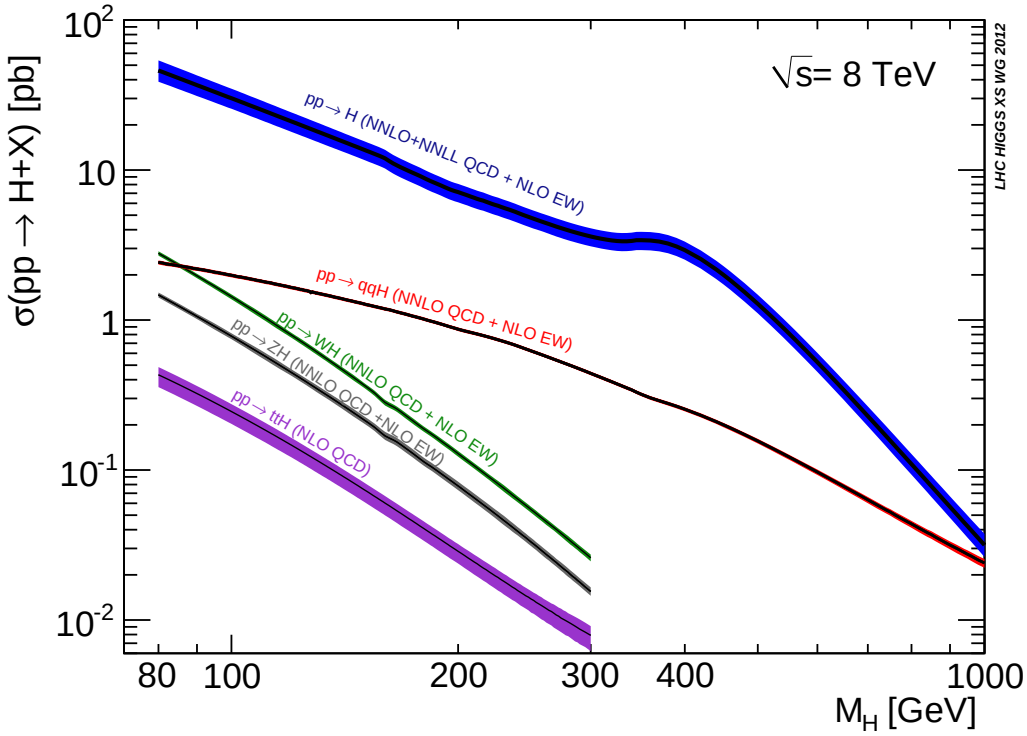


Figure 1-5: Cross section of the main Higgs production processes at LHC Run-I at $\sqrt{s} = 8$ TeV [14].

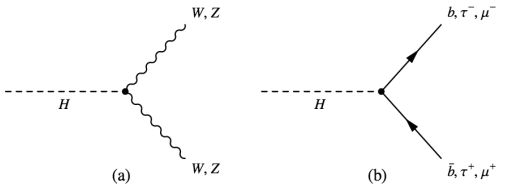


Figure 1-6: Feynman diagrams illustrating the Higgs direct decay into a pair of bosons (left) and fermions (right) at leading order. The $H \to \gamma\gamma$ decay is not shown as it is loop induced (see Fig. 1-7).

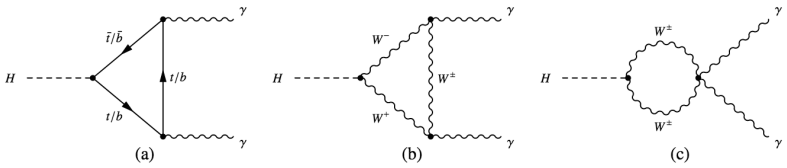


Figure 1-7: Leading-order Feynman diagrams showing the Higgs decay into a pair of photons $\gamma\gamma$.

PUs and THUs follow the definitions in [16, 17], with four input parameters to be varied for PUs and eight parameters for THUs. PUs are assumed to be Gaussian errors while THUs follow a flat probability distribution for a given range. All THUs are uncorrelated except for the ones corresponding to the 4-fermion final states [15]. Based on the chosen accuracy level, some uncertainties have been safely neglected. α_s and m_b variations affect BR predictions by the a few percent. m_c variations affect mainly the $H \to c\bar{c}$ channel. And the estimated THUs are 1% for the $H \to \gamma\gamma$ and 0.5% for $H \to WW/ZZ \to 4f$. For the remaining channels, THUs have a non-negligible impact if measurements from the channel in question are included or if errors of few per mille are important.

The partial widths at the leading order are given by [18]:

$$\Gamma(H \to ff) = \frac{G_F}{4\pi\sqrt{2}} m_f^2 m_H \left(1 - 4\frac{m_f^2}{m_H^2}\right)^{3/2}, \tag{1.29}$$

$$\Gamma(H \to WW) = \frac{G_F}{8\pi\sqrt{2}} m_H^3 \left(1 - 4\frac{m_W^2}{m_H^2}\right)^{1/2} \left(12\frac{m_W^4}{m_H^4} - \frac{m_W^2}{m_H^2} + 1\right), \tag{1.30}$$

$$\Gamma(H \to ZZ) = \frac{G_F}{16\pi\sqrt{2}} m_H^3 \left(1 - 4\frac{m_Z^2}{m_H^2}\right)^{1/2} \left(12\frac{m_Z^4}{m_H^4} - \frac{m_Z^2}{m_H^2} + 1\right). \tag{1.31}$$

Table 1.5: Relative parametric (PU) and theoretical (THU) uncertainties on the SM Higgs partial widths for a selection of Higgs masses [15]. The PU are shown for each single parameter.

Channel	M_H [GeV]	Γ [MeV]	$\Delta\alpha_s$	Δm_b	Δm_c	Δm_t	THU
$H \to b\bar{b}$	122	2.30	-2.3% $+2.3\%$	$+3.2\%$ -3.2%	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+2.0\%$ -2.0%
	126	2.36	-2.3% $+2.3\%$	$+3.3\%$ -3.2%	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+2.0\%$ -2.0%
	130	2.42	-2.4% $+2.3\%$	$+3.2\%$ -3.2%	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+2.0\%$ -2.0%
$H \to \tau^+\tau^-$	122	$2.51 \cdot 10^{-1}$	$+0.0\%$ $+0.0\%$	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+0.0\%$ -0.1%	$+2.0\%$ -2.0%
	126	$2.59 \cdot 10^{-1}$	$+0.0\%$ $+0.0\%$	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+0.1\%$ -0.1%	$+2.0\%$ -2.0%
	130	$2.67 \cdot 10^{-1}$	$+0.0\%$ $+0.0\%$	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+0.1\%$ -0.1%	$+2.0\%$ -2.0%
$H \to \mu^+\mu^-$	122	$8.71 \cdot 10^{-4}$	$+0.0\%$ $+0.0\%$	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+0.1\%$ -0.1%	$+2.0\%$ -2.0%
	126	$8.99 \cdot 10^{-4}$	$+0.0\%$ $+0.0\%$	$+0.0\%$ -0.0%	-0.1% -0.0%	$+0.0\%$ -0.1%	$+2.0\%$ -2.0%
	130	$9.27 \cdot 10^{-4}$	$+0.1\%$ $+0.0\%$	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+0.1\%$ -0.0%	$+2.0\%$ -2.0%
$H \to c\bar{c}$	122	$1.16 \cdot 10^{-1}$	-7.1% $+7.0\%$	-0.1% $+0.1\%$	$+6.2\%$ -6.0%	$+0.0\%$ -0.1%	$+2.0\%$ -2.0%
	126	$1.19 \cdot 10^{-1}$	-7.1% $+7.0\%$	-0.1% $+0.1\%$	$+6.2\%$ -6.1%	$+0.0\%$ -0.1%	$+2.0\%$ -2.0%
	130	$1.22 \cdot 10^{-1}$	-7.1% $+7.0\%$	-0.1% $+0.1\%$	$+6.3\%$ -6.0%	$+0.1\%$ -0.1%	$+2.0\%$ -2.0%
$H \to gg$	122	$3.25 \cdot 10^{-1}$	$+4.2\%$ -4.1%	-0.1% $+0.1\%$	$+0.0\%$ -0.0%	-0.2% $+0.2\%$	$+3.0\%$ -3.0%
	126	$3.57 \cdot 10^{-1}$	$+4.2\%$ -4.1%	-0.1% $+0.1\%$	$+0.0\%$ -0.0%	-0.2% $+0.2\%$	$+3.0\%$ -3.0%
	130	$3.91 \cdot 10^{-1}$	$+4.2\%$ -4.1%	-0.1% $+0.2\%$	$+0.0\%$ -0.0%	-0.2% $+0.2\%$	$+3.0\%$ -3.0%
$H \to \gamma\gamma$	122	$8.37 \cdot 10^{-3}$	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+1.0\%$ -1.0%
	126	$9.59 \cdot 10^{-3}$	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+1.0\%$ -1.0%
	130	$1.10 \cdot 10^{-2}$	$+0.1\%$ -0.0%	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+1.0\%$ -1.0%
$H \to Z\gamma$	122	$4.74 \cdot 10^{-3}$	$+0.0\%$ -0.1%	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+0.0\%$ -0.1%	$+5.0\%$ -5.0%
	126	$6.84 \cdot 10^{-3}$	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+0.0\%$ -0.1%	$+0.0\%$ -0.1%	$+5.0\%$ -5.0%
	130	$9.55 \cdot 10^{-3}$	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+5.0\%$ -5.0%
$H \to WW$	122	$6.25 \cdot 10^{-1}$	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+0.5\%$ -0.5%
	126	$9.73 \cdot 10^{-1}$	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+0.5\%$ -0.5%
	130	1.49	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+0.5\%$ -0.5%
$H \to ZZ$	122	$7.30 \cdot 10^{-2}$	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+0.5\%$ -0.5%
	126	$1.22 \cdot 10^{-1}$	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+0.5\%$ -0.5%
	130	$1.95 \cdot 10^{-1}$	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+0.0\%$ -0.0%	$+0.5\%$ -0.5%

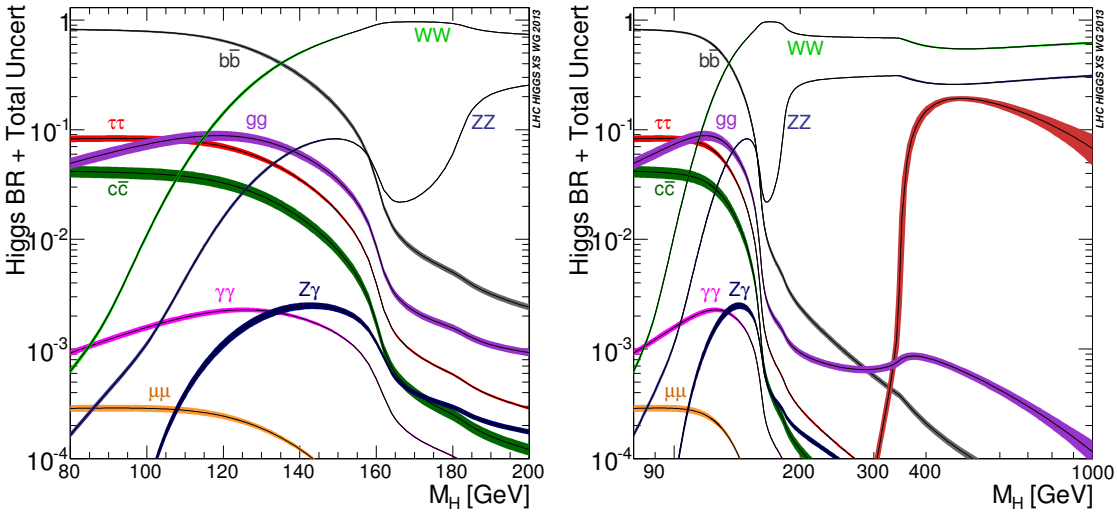


Figure 1-8: Higgs branching ratios and their uncertainties for the low mass range (left) and for the full mass range (right) [15]. The solid line in deep red for $m_H > 300$ GeVis for the $t\bar{t}$ mode.

1.4 Discovery of the Higgs Boson

In June 2012, ATLAS and CMS announced the discovery of a new particle with early results in the bosonic decay modes, showing a significance of more than 5 σ (probability of a statistical background fluctuation $p_0 \sim 1.7 \times 10^{-9}$). Results in the fermionic modes were missing, which motivated the search in channels such as $H \to \tau^+\tau^-$ (see chapter 6) to check the compatibility with a SM like Higgs boson. The latest results from ATLAS and CMS for Higgs searches in various decay modes show significance of at least $\sim 5\ \sigma$ in each bosonic mode. Full results are summarized in [19, 20] and the complete ATLAS results are illustrated in Table 1.6.

1.5 BSM, Run-II Prospectives and Conclusions

The SM phenomenology was presented with an emphasis on the Higgs mechanism, since the work presented in this thesis focuses on the experimental search for the Higgs boson and the physics beyond the SM (BSM) using the Higgs effective field theory. Higgs search studies, e.g. $H \to \tau^+\tau^-$ presented in chapter 6, require a good knowledge of the various Higgs production and decay modes at LHC Run-I as discussed briefly earlier. The latest ATLAS and CMS combination results for the various Higgs production and decay modes were also summarized.

In addition, the properties of the newly discovered particle need to be determined to check if it is indeed a SM Higgs boson or not. In order to explore all possibilities in a model independent way, with options including SM and beyond SM theories,

Table 1.6: Table summarizing the signal strengths and statistical significance values in ATLAS for each Higgs search channel studied during Run-I at both $\sqrt{s} = 7$ and 8 TeV[19].

Analysis	Signal		$\int \mathcal{L}dt$ [fb^{-1}]	
Categorisation or final states	Strength μ	Significance [s.d.]	7 TeV	8 TeV
$H \to \gamma\gamma$	1.17 ± 0.27	5.2 (4.6)	4.5	20.3
ttH: leptonic, hadronic			✓	✓
VH: one-lepton, dilepton, $E_{\mathrm{T}}^{\mathrm{miss}}$, hadronic			✓	✓
VBF: tight, loose			✓	✓
ggF: 4 $p_{\mathrm{T_t}}$ categories			✓	✓
$H \to ZZ^* \to 4\ell$	$1.44^{+0.40}_{-0.33}$	8.1 (6.2)	4.5	20.3
VBF			✓	✓
VH: hadronic, leptonic			✓	✓
ggF			✓	✓
$H \to WW^*$	$1.16^{+0.24}_{-0.21}$	6.5 (5.9)	4.5	20.3
ggF: (0-jet, 1-jet) $\otimes$ $(ee + \mu\mu,\ e\mu)$			✓	✓
ggF: $\geq$ 2-jet and $e\mu$				✓
VBF: $\geq$ 2-jet $\otimes$ $(ee + \mu\mu,\ e\mu)$			✓	✓
VH: opposite-charge dilepton, three-lepton, four-lepton			✓	✓
VH: same-charge dilepton				✓
$H \to \tau^+\tau^-$	$1.43^{+0.43}_{-0.37}$	4.5 (3.4)	4.5	20.3
Boosted: $\tau_{\mathrm{lep}}\ \tau_{\mathrm{lep}}$,$\tau_{\mathrm{lep}}\ \tau_{\mathrm{had}}$,$\tau_{\mathrm{had}}\ \tau_{\mathrm{had}}$			✓	✓
VBF: $\tau_{\mathrm{lep}}\ \tau_{\mathrm{lep}}$,$\tau_{\mathrm{lep}}\ \tau_{\mathrm{had}}$,$\tau_{\mathrm{had}}\ \tau_{\mathrm{had}}$			✓	✓
$VH \to Vb\bar{b}$	0.52 ± 0.40	1.4 (2.6)	4.7	20.3
0ℓ ($ZH \to \nu\nu b\bar{b}$): $N_{\mathrm{jet}} = 2, 3$, N_{btag} =1,2, $p_{\mathrm{T}}^V \in$ 100-120 and >120 GeV			✓	✓
1ℓ ($WH \to \ell\nu b\bar{b}$): $N_{\mathrm{jet}} = 2, 3$, N_{btag} =1,2, $p_{\mathrm{T}}^V <$ and >120 GeV			✓	✓
2ℓ ($ZH \to \ell\ell b\bar{b}$): $N_{\mathrm{jet}} = 2, 3$, N_{btag} =1,2, $p_{\mathrm{T}}^V <$ and >120 GeV			✓	✓
		95% CL limit		
$H \to Z\gamma$		$\mu < 11$ (9)	4.5	20.3
10 categories based on $\Delta\eta_{Z\gamma}$ and $P_{\mathrm{T_t}}$			✓	✓
$H \to \mu\mu$		$\mu < 7.0$ (7.2)	4.5	20.3
VBF and 6 other categories based on η_μ and $p_{\mathrm{T}}^{\mu\mu}$			✓	✓
ttH production			4.5	20.3
$H \to b\bar{b}$: single-lepton, dilepton		$\mu < 3.4$ (2.2)		✓
$ttH \to$multileptons: categories on lepton multiplicity		$\mu < 4.7$ (2.4)		✓
$H \to \gamma\gamma$: leptonic, hadronic		$\mu < 6.7$ (4.9)	✓	✓
Off-shell H^* production		$\mu < 5.1$ – 8.6 (6.7 – 11.0)		20.3
$H^* \to ZZ \to 4\ell$				✓
$H^* \to ZZ \to 2\ell 2\nu$				✓
$H^* \to WW \to e\nu\mu\nu$				✓

Higgs effective field theory (HEFT) is the most suitable option. BSM features are included in HEFT only when new states appear at scales not directly accessible at the LHC. Effects of heavy particles are encoded into the contact interaction of low energy particles, keeping only the relevant degrees of freedom at low energies.
HEFT, which is the option adopted by CERN for Run-II studies, is already well implemented in Madgraph5_aMC@NLO and a first look at HEFT in the $H \rightarrow \tau^+\tau^-$ channel is shown in chapter 7.

This effective parametrization of the SM 'HEFT' allows a systematic study of EW precision tests and triple-gauge couplings, in addition to Higgs couplings in a unified and efficient framework. These measurements, along with flavor physics studies, are the major focus of Run-II studies and will contribute to constraining BSM models [21].

Bibliography

[1] S. L. Glashow, *Partial-symmetries of weak interactions*, Nucl.Phys. **22** (1961) 579-588

[2] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory*, Westview Press, 1995

[3] A. Salam *Elementary particle theory: Relativistic groups and analyticity*, Nobel Symposium No. 8 (1968)

[4] S. Weinberg, *A model of leptons*, Phys. Rev. Lett. **19** (1967) 1264-1266

[5] E. Noether, *Invariante Variationsprobleme*, Nachr. d. Koenig. Gesellsch. d. Wiss. zu Goettingen, Math-phys. Klasse (1918)

[6] Particle Data Group Collaboration(K. A. Olive et al.), *The Review of Particle Physics*, Chin. Phys. C **38** (2014) 090001

[7] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321?323

[8] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys.Lett. **12** (1964) 132-133

[9] A. Pich, *The Standard model of electroweak interactions*, arXiv:0705.4264 [hep-ph]

[10] A. Djouadi, *The Anatomy of electro-weak symmetry breaking. I: The Higgs boson in the standard model*, Phys.Rept. 457 (2008) 1-216, arXiv:hep-ph/0503172

[11] J. Iliopoulos, *Introduction to the Standard Model of the Electro-Weak Interactions*, arXiv:1305.6779

[12] ALEPH and DELPHI and L3 and OPAL and LEP Electroweak Working Group Collaborations (Alcaraz, J. et al.), *A Combination of preliminary electroweak measurements and constraints on the standard model* , (2006)arXiv:hep-ex/0612034

[13] The ALEPH, DELPHI, L3, OPAL and SLD Collaborations, the LEP Electroweak Working Group and the SLD Electroweak and Heavy Flavour Groups

Collaborations (Schael, S. et al.), *Precision electroweak measurements on the Z resonance*, Phys. Rept. **427** (2006) 257

[14] LHC Higgs Cross section working group, *CERN Yellow Page Report At 8 TeV*, https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt8TeV

[15] LHC Higgs Cross Section Working Group, S. Heinemeier, C. Mariotti, G. Passarino, R. Tanaka (Eds.), *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, CERN-2013-004 (CERN, Geneva, 2013), DOI: 10.5170/CERN-2013-004

[16] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.), *Handbook of LHC Higgs Cross Sections: 2. Differential Distributions*, CERN-2012-002 (CERN, Geneva, 2012) , arXiv:1201.3084 [hep-ph].

[17] A. Denner, S. Heinemeyer, I. Puljak, D. Rebuzzi, and M. Spira, *Standard Model Higgs-boson branching ratios with uncertainties*, Eur. Phys. J. C**71** (2011) 1753, arXiv:1107.5909 [hep-ph].

[18] E. A. Paschos, *Electroweak theory*, Cambridge University Press, 2007

[19] ATLAS Collaboration, *Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s}$=7 and 8 TeV in the ATLAS experiment* , (2015) arXiv:1507.04548 [hep-ex]

[20] *Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s}$ = 7 and 8 TeV*, ATLAS-COM-CONF-2015-053

[21] J. Ellis, *The Beautiful Physics of LHC Run 2*, PoS Beauty2014 (2015) **056**, arXiv:1412.2666 [hep-ph], KCL-PH-TH-2014-50, LCTS-2014-51, CERN-PH-TH-2014-248

Chapter 2

LHC and the ATLAS Detector

The measurements presented in this thesis are based on the data taken with the ATLAS detector, which is one the LHC detectors. In this chapter, a brief LHC introduction given first is followed by a description of the ATLAS detector, explaining the different parts of the detector, the magnetic system and the trigger system.

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC), as illustrated in Fig 2-1(a), is currently the largest particle physics accelerator. Developed by CERN[1], this circular accelerator is installed in a 27 Km circumference underground tunnel with superconducting magnets and accelerating structures, 175 meters beneath the franco-swiss border near Geneva [1, 2]. In fact, one of the major challenges when building the LHC was the magnetic system required to work well in a high luminosity, high radiation environment. The LHC has indeed 1232 dipole magnets bending the beams, 392 quadrupole magnets to focus the beams and increase the interaction probability when crossing at each of the four intersection points, in addition to sextuples, octupoles, etc. for a total of 9593 magnets. These NbTi superconductors are cooled using superfluid helium at -271.3°C (below 2K) [3].

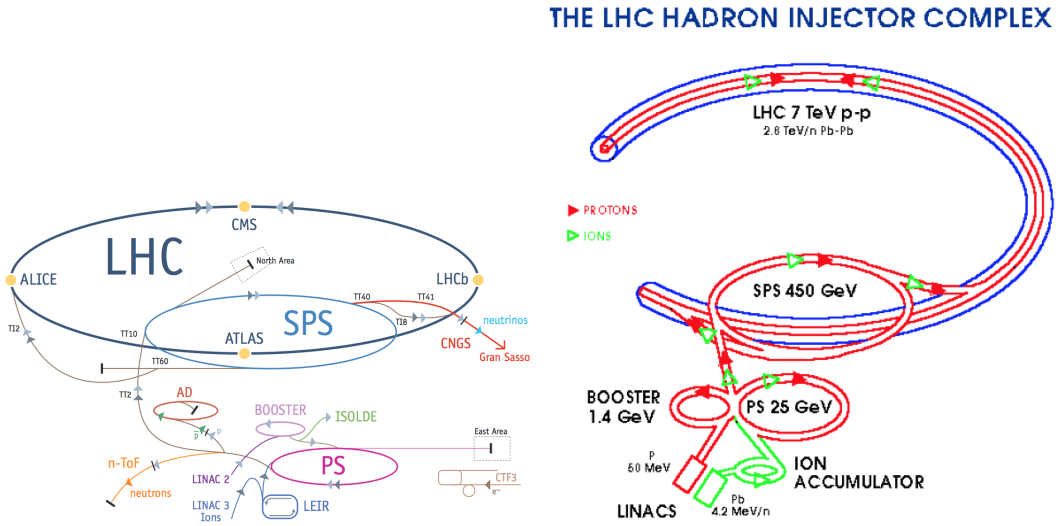
In the 3.8 m wide tunnel, which was formerly used as the LEP (Large Electron-Positron) [4] collider ring, two proton (p-p) beams circulate in opposite directions in the ring at nearly the speed of light (Lorentz factor of 7500 at the 14 TeV design center of mass (CM) energy), intersecting at 4 points. The beams had 3.5 TeV(4 TeV) energy each, corresponding to CM energy $\sqrt{s} = 7\,(8)$ TeV in 2011 (2012)[2].

The proton beams, obtained by stripping electrons from hydrogen atoms, leave Linac2 (Linear accelerator) at 50 MeV energy into the proton synchrotron booster (PSB). In the booster, they are accelerated up to 1.4 GeV in energy and to 25 GeV in the

[1] European Center for Nuclear Research.

[2] LHC Run-I lasted from 2010-2013, with data being taken in 2011 and 2012. The Run-II started in the spring of 2015 with 13 TeV CM energy.

proton synchrotron (PS), before reaching 450 GeV in the super proton synchrotron (SPS) and being sent to the LHC, where they are accelerated to their nominal energy e.g. 7 TeV in 2011. In addition to p-p beams, heavy ions (e.g. lead nuclei) collisions are performed at the end of the year. The various stages of particle acceleration at the LHC are illustrated in Fig 2-1 [5]. Additional details on the LHC design and performance during the period 2010-2013 can be found in [6].



(a) (b)

Figure 2-1: Plot showing the LHC design (a) [2] and the various stages of particle acceleration at LHC (b) [5]. The energy values shown are for $\sqrt{s}$= 7 TeV in 2011.

2.1.1 Purpose

The LHC was built to test the Standard Model (SM) in order to prove/deny the existence of the lightest scalar boson, namely the Higgs boson, in addition to testing BSM (Beyond Standard Model) theories e.g. supersymmetry,and looking for dark matter, dark energy. One of the major goals of the LHC is to answer some fundamental open questions in physics concerning the interactions of elementary particles and the laws governing them. A major emphasis was thus put on testing and validating either the SM predictions or Higgsless models, while allowing many theoretical extensions. Some of the issues/questions explored include [2]:

- Does the lightest scalar boson predicted by SM, i.e. Higgs boson exist with spin-parity 0^+?
- Is supersymmetry realized in nature?
- What is the nature of dark energy?
- The flavor hierarchy

- Gauge coupling unifications
- Baryon asymmetry and strong CP violation
- Properties of quark-gluon plasma.

2.1.2 Luminosity and Performance

The number of interactions per second, or interaction rate, is proportional to the event cross section, with 'luminosity' L being the constant of proportionality. For a given event cross section σ_{event}, the number of events produced per second is defined as:

$$\mathbf{R} = L \times \sigma_{\text{event}} \ , \tag{2.1}$$

and the observed number of events is $N_{\text{obs}} = \int L \, \mathrm{d}t \, \sigma \, \epsilon$, where ϵ depends on the detection efficiency.
As a result, the design value of the luminosity is dictated by the cross section of the events of interest. For example, with an actual instantaneous LHC luminosity $\sim 10^{33}\text{cm}^{-2}\text{s}^{-1}$ ($\sim 10^{34}\text{cm}^{-2}\text{s}^{-1}$ design value), Higgs events produced through vector boson fusion (VBF) with a cross section of about 1.2 pb (1 pb = 10^{-36} cm^2) at 7 TeV [7], of interest for e.g. $H \to \tau\tau$ analysis, will be recorded at the rate of 1 event per day (all decay modes included). So for interesting events with small cross section values, a sufficiently large number of events needs to be recorded within a reasonable time.

The LHC during Run-I operated with 50 ns bunch spacing for both 2011 and 2012, leading to high instantaneous luminosity with high pile-up in particular during 2012. This value was favored over the design value 25 ns since it allows for a larger number of protons per bunch while keeping a smaller geometrical emittance as shown in Table 2.1. The average number of interactions per crossing $\langle\mu\rangle$, as defined in sec. 2.1.2.1, was 12 (30) in 2011 (2012), with the maximum number of interactions per crossing around 30 (40) events/crossing in 2011(2012) [8].

Bunch spacing[ns]	Protons per bunch	Emittance [μm]
50	1.7×10^{11}	1.8
25	1.2×10^{11}	2.7

Table 2.1: Table showing 2012 beam parameters values justifying the choice of 50 ns bunch crossing over the design value of 25 ns [8].

The luminosity at LHC is derived from eq. (2.2) assuming round beams with circular symmetry as shown in eq. 2.3 where [3, 8]:

$$L = \frac{N^2 k_b f}{4\pi\sigma_x^*\sigma_y^*} F \ . \tag{2.2}$$

$$L = \frac{N^2 k_b f \gamma}{4\pi \epsilon_n \beta^*} F \ , \tag{2.3}$$

- N: number of particles per bunch.
- k_b: number of bunches.
- f: revolution frequency.
- γ: relativistic γ function.
- $\sigma^*_{x(y)}$: beam width at the interaction point (IP) in the horizontal (vertical) direction.
- ϵ_n: normalized beam emittance.
- β^*: beta function value at the interaction point.
- F: geometrical reduction factor due to the crossing angle at the IP defined as:

$$F = \left(1 + \left(\frac{\theta_c \sigma_z}{2\sigma^*}\right)^2\right)^{-1/2} \ , \tag{2.4}$$

  where θ_c is the crossing angle at the IP, σ_z is the RMS bunch length and σ^* is transverse beam width (assuming round beams).

A comparison between the actual and design values of the basic beam parameters for 2012 is shown in Table 2.2, with the interaction points associated with each octant of the LHC ring illustrated in Fig 2-2.

Parameter	Value in 2012	Design value
Beam energy[TeV]	4	7
β^* in IP 1,2,5,8[m]	0.6, 3.0, 0.6, 3.0	0.55
Bunch spacing [ns]	50	25
Number of bunches	1374	2808
Average bunch intensity [protons per bunch]	1.6-1.7$\times 10^{11}$	1.5$\times 10^{11}$
Peak luminosity [cm$^{-2}s^{-1}$]	7.7×10^{33}	1.0×10^{34}
Max. mean number of events per bunch crossing	≈ 40	19
Stored beam energy[MJ]	$\approx$140	362

Table 2.2: Table showing actual and design values of beam performance parameters in 2012 [8].

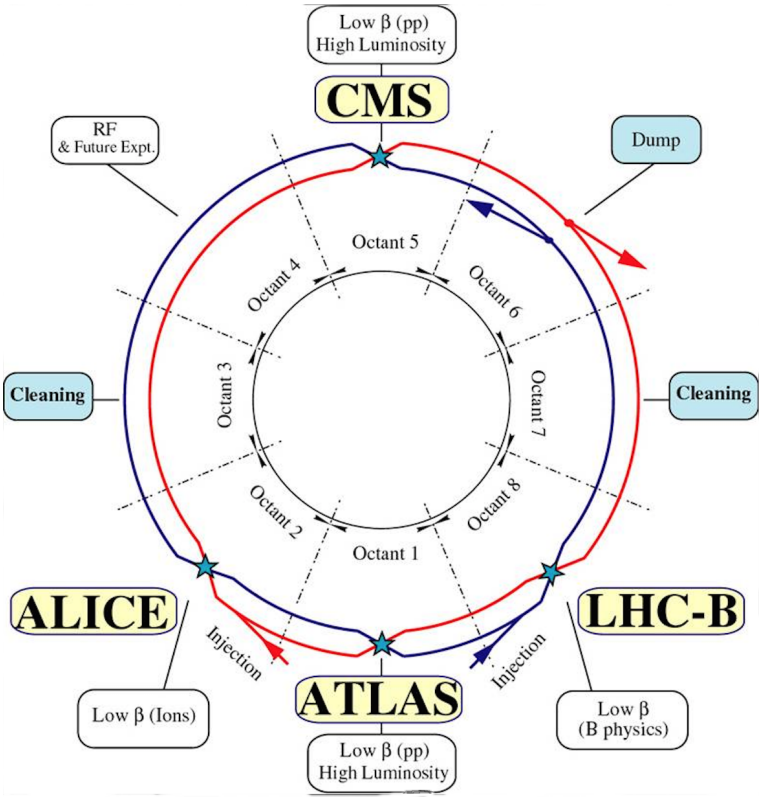


Figure 2-2: Schematic representation showing the four major experiments and the two ring structure of the LHC. Interaction points (IPs) are labelled based on octant. For example, ATLAS is at IP1 [9].

2.1.2.1 Pile-up

With high luminosity colliders, several events per bunch crossing are produced including the one of interest and the 'pile-up' events. Two different types of pile-up exist in ATLAS [34]:

- **In-time pile-up:** occurs when the collision of interest is surrounded by additional proton-proton collisions in the same bunch crossing. The number n_{PV} of reconstructed primary vertices (PV) in the inner detector (ID) is the estimator of its magnitude.

- **Out-of-time pile-up:** is due to the long charge collection time and signal shaping techniques in calorimeters. One correlates its magnitude to the average number of interactions per crossing $\langle\mu\rangle$, shown in Fig. 2-3, defined as:

$$\langle\mu\rangle = \frac{L \times \sigma_{\mathrm{inel}}}{k_b \times f} \ , \tag{2.5}$$

  where L is the instantaneous luminosity and σ_{inel} is the proton-proton inelastic collision cross section. This effect is seen mainly in the liquid argon calorimeter, while being small in the tile calorimeter.

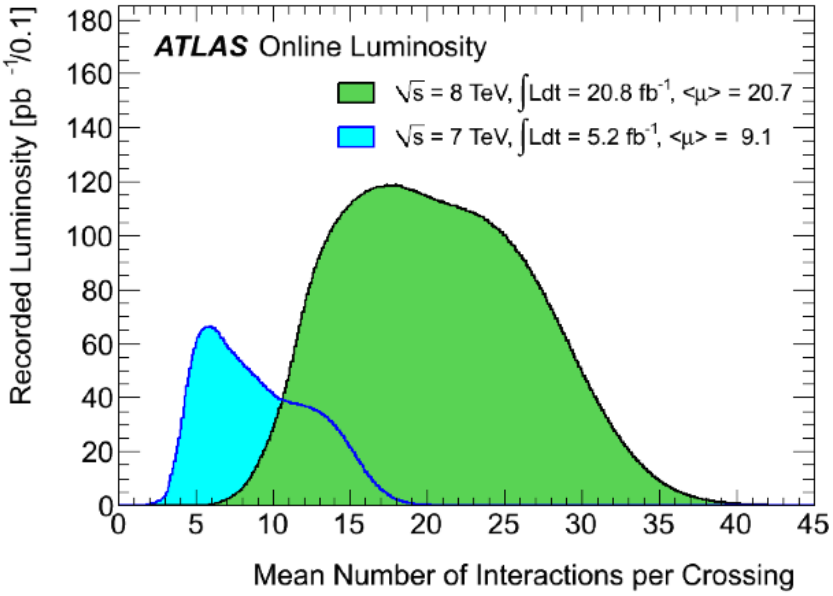


Figure 2-3: Average number of interactions per crossing at LHC in 2011 (blue) and 2012 (green) [11].

2.1.3 LHC Experiments

To observe the particles resulting from collisions, particle detectors are needed. The different traces that a particle leaves in a detector (hits, energy deposits, etc.) are used to reconstruct particles as described in chapter 4. The obtained measurements allow to measure the particle properties e.g. masses, lifetime, charge, spin, etc. At LHC, there are seven experimental detectors to analyze the collision products:

1. ATLAS (**A** **T**oroidal **L**HC **A**pparatu**S**)
2. CMS (**C**ompact **M**uon **S**olenoid)
3. LHCb (**LHC-b**eauty)
4. ALICE (**A** **L**arge **I**on **C**ollider **E**xperiment)
5. TOTEM (**TOT**al **E**lastic and diffractive cross section **M**easurement)
6. LHCf (**LHC-f**orward)
7. MoEDAL (**Mo**nopole and **E**xotics **D**etector **A**t the **L**HC)

ATLAS and CMS are the largest experiments, exploring a various range of particle physics, looking for signs of new physics. Their detectors were designed independently and allow for cross-checks and cross-confirmation of any new discovery. The LHCb, on the other hand, is built for flavor physics studies and ALICE is designed to study

massive nuclei collisions, e.g. lead nuclei, focusing on the quark-gluon plasma created by strongly interacting particles at very high energy densities. Finally, the smallest experiments are TOTEM and LHCf, which focus on "forward particles"[3], whereas MoEDAL searches for the hypothetic particle "magnetic monopole" [12].
At LHC, the accelerated colliding beams consist of:

- protons at 4 TeV each (8 TeV collision energy) in 2012 and at 6.5 TeV (13 TeV collision energy) in 2015.
- lead nuclei (574 TeV per nucleus, 2.76 TeV per nucleon). The lead nuclei collisions are done for about a month at the end of the year after all proton-proton (p-p) collisions have been completed.
- proton and lead nuclei i.e. p-Pb collisions.

2.2 The ATLAS Experiment: Overview

ATLAS has a general-purpose detector designed to scan a large range of possible particle physics signatures. The goal is to explore the full LHC potential with an emphasis on SM Higgs boson and BSM studies amongst others (e.g. CP violation studies and top quark properties measurements).
The major parts of the ATLAS detector (described in subsequent sections) are the inner detector (ID), the calorimeters, the muon spectrometer and the magnet system. The main features of this detector are [13, 14]:

1. A cylindrical symmetry mainly in the central barrel region, where the beam pipe is surrounded by various cylindrical detector layers.
2. Good electromagnetic calorimeters for detection and identification of electrons and photons, surrounded by hadronic calorimeters for accurate measurements of the jets and missing transverse energy in particular.
3. Precise muon measurements with a well-performing muon spectrometer (MS) in the high pile-up conditions guaranteed to give accurate measurements even if used alone.
4. Large acceptance in pseudorapidity and azimuthal angle covering the range ($0 \leq |\eta| \leq 4.9$). The detector is hermetic, with nearly 4π coverage. The energy of hadrons needs to be well measured in order to have reliable missing transverse energy estimates originating from the p_{T} imbalance in the transverse plane.

In the following sections, a summary of the design and structure of the ATLAS detector (shown in Fig. 2-4) is given.

[3] protons or heavy ions not undergoing head-on collisions but merely passing by each other.

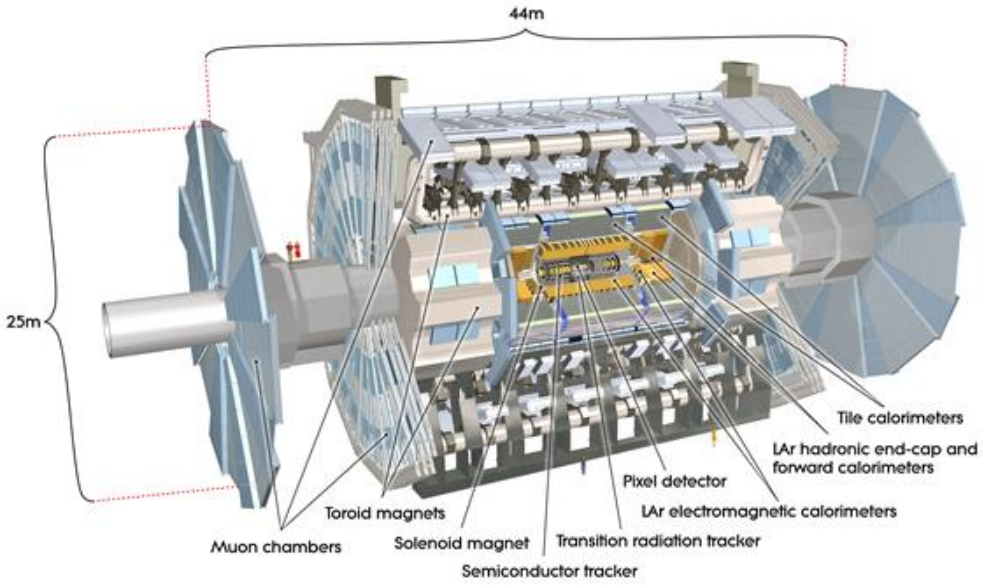


Figure 2-4: A graphical representation of the various parts of ATLAS detector obtained by computer simulation [15].

2.3 The ATLAS Coordinate System

To define the basic parameters of a detected particle needed for energy and position reconstruction, ATLAS uses a coordinate system such that :

1. **The origin** is the detector center, which corresponds to the nominal interaction point (IP) as well.

2. **The z-axis** is parallel to the beam axis in a counterclockwise direction. The positive z-axis corresponds to 'side A' of the detector, whereas the other half of the detector given by the negative z-axis defines 'side C'.

3. **The x-axis** points from the IP to the center of the LHC ring.

4. **The y-axis** points upwards to the surface. (z, x, y) form a right-handed basis.

The radial distance is then defined as $r = \sqrt{(x^2 + y^2)}$. The xy-plane is thus referred to as the transverse plane. Given the cylindrical symmetry of the sub-detector parts, in particular in the barrel[4] region which is the largest region, cylindrical coordinates are more suitable. The angular coordinates are defined as follows:

- $\phi \equiv$ the **azimuthal angle** with the positive x-axis in the xy-plane. $\phi \in [-\pi, \pi]$. It increases in the clockwise direction when looking down the positive z-axis.

[4]The detector main regions are the barrel region with cylindrical symmetry, and the end caps region where the detector parts are organized in disks perpendicular to the z-axis. The barrel region occupies most of the detector volume while the end caps are at each of the cylinder edges.

- $\theta \equiv$ **polar angle** with respect to the z-axis. $\theta \in [0, \pi]$. The θ measurement is actually replaced by the pseudorapidity measurement η as defined in eq. (2.6), since the particle flux is almost constant as a function of η.

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

Objects are assigned (η, ϕ) measurements, and the opening angle between two objects is defined as:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \ . \tag{2.7}$$

2.4 The ATLAS Detector

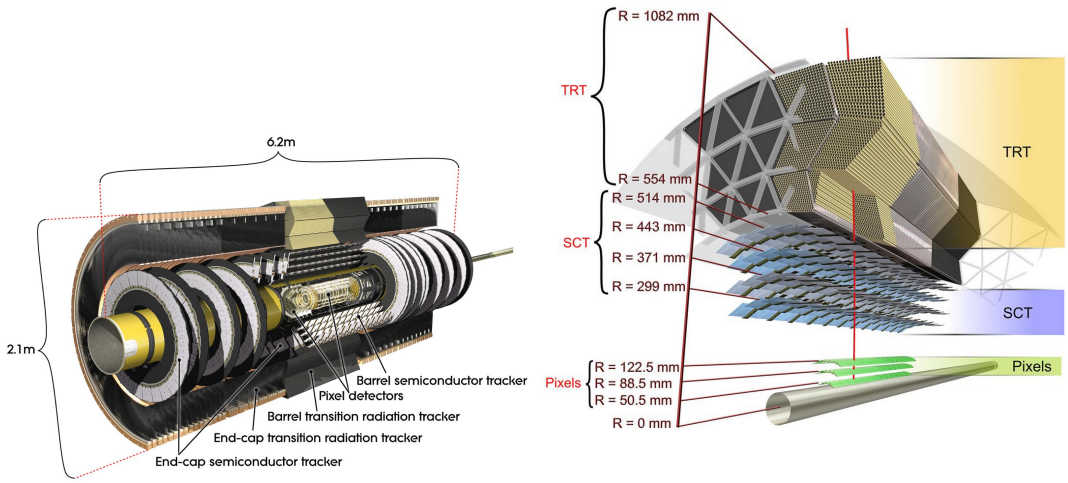
The major parts of the ATLAS detector shown in Fig. 2-4 are:

- a central tracking detector, also called inner detector (ID), composed of 3 sub-detectors layered in a concentric way: a pixel detector surrounded by silicon microstrip trackers (SCT), followed by a transition radiation tracker (TRT).
- high granularity calorimeters (liquid argon LAr and tile calorimeters).
- large high-precision muon spectrometer (MS) with micro drift tubes (MDT), resistive plate chambers (RPC), cathode strip chambers (CSC) and thin-gap chambers (TGC).
- A central solenoid providing the tracker with a 2 T magnetic field, in addition to toroidal magnets covering the muon spectrometer and the area outside the calorimeters.

Strict requirements are imposed on the precision, efficiency and acceptance of the detector, affecting its performance, by the wide range of particle energy to be detected, the various particle types produced from p-p collisions and the high instantaneous luminosity needed. This results in constraints on the choice of the absorber/detector material in calorimeters and on the read-out electronics for example.

2.4.1 The Inner Detector (ID)

The inner detector is designed to reconstruct charged particle tracks and vertices in the event with high precision, allowing good charged particle/physics object reconstruction along with calorimeter and muon system measurements, and provide information about the short-lived particles decay vertices. In the barrel region, the 3 subdetectors (pixel, SCT, TRT), immersed in a 2 T magnetic field, are designed as concentric cylindrical layers, while in the end-cap regions they are in the form of disks perpendicular to the beam axis [16]. The inner detector (ID), as illustrated in Fig. 2-5 with an acceptance $|\eta| < 2.5$ (< 2 in TRT) and full ϕ coverage, was designed to have a transverse momentum resolution $\frac{\sigma\,(p_T)}{p_T} = 0.05\%\ p_T\ [\text{GeV}] + 1\%$, with transverse impact parameter resolution of about 10 μm for high-p_T particles [14].



(a) (b)

Figure 2-5: Layout of the ATLAS inner detector [14, 15]. The graphical representation in (b) does not include the additional Insertable B-Layer (IBL), which was added during the 2013-2014 shutdown after Run-I.

2.4.1.1 ID subdetectors

1. **Pixel Detector:** It is designed to give precise measurements of particles produced near the interaction point.

   - **Run-I:** During Run-I, the barrel region had 3 layers at radii $r = 5$ cm (B-Layer or layer-0), 9 cm (Layer-1), 12 cm (Layer-2), while the two end caps had 3 disks each with 48 modules per disk. The pixel detector, providing typically three precise measurements, consists of 1744 modules (1456 barrel modules + 288 forward modules), tilted at 20° in the r-ϕ plane to compensate for the Lorentz angle [17, 18]. Each of these modules has more than 46 000 silicon pixels with dimensions 50 μm $(r\phi)\times$ 400 μm (η) and 250 μm thickness each, leading to a total of 80.4 million channels. In each module, the pixels are arranged in 328 rows $(x_{\text{local}}) \times 144$ columns (y_{local}), providing measurements with a resolution of 10 μm in $r\phi$ and 115 μm in η [17–19]. The detector parts are cooled on average to $-13°C$.
   - **Run-II:** An additional Insertable B-Layer (IBL), as shown in Fig. 2-6, was later added to the existing detector, during the upgrade in the 2013-2014 shutdown, at 3.3 cm from the beam pipe[5] axis with a coverage in $|\eta|$ up to 2.9. IBL has 14 staves 64 cm long each, tilted at 14° to compensate for the Lorentz angle, with 26880 pixels in total arranged in 336 rows$\times$ 80 columns. The pixels in IBL have a smaller size with 50 μm $(r\phi)\times$250 μm(η) [20]. The main goal of IBL is to improve measurements when increasing the luminosity, where the added smaller pixels allow better tracking per-

[5]The beam pipe is replaced for Run-II by a smaller beryllium beam pipe of 25 mm radius.

formance and the higher granularity leads to better tagging of short lived hadrons e.g. b-tagging[6].

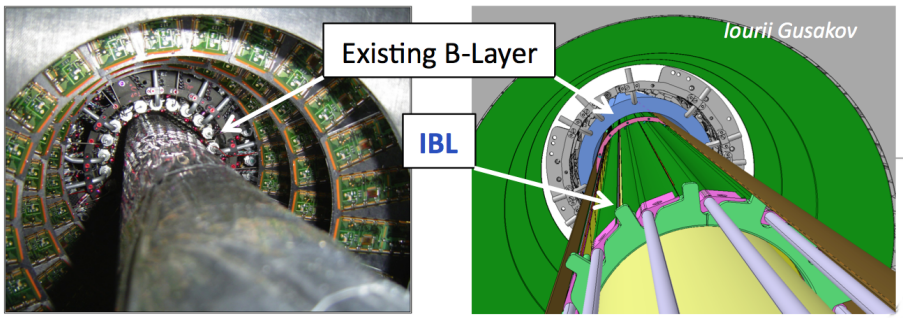


Figure 2-6: Figure showing where the IBL is inserted in the pixel detector [21].

2. **The SemiConductor Tracker (SCT):** Its sensitive elements cover distances between 299 mm and 560 mm radially, with 4088 silicon-strip modules distributed over the barrel carbon-fibre cylinders at radii 30.0, 37.7, 44.7 and 52.0 cm respectively [22], and 2 end caps with 9 disks each. The direction and strip pitch value varies depending on the detector region: in the barrel area, the strips approximately parallel to z axis have a constant pitch of 80 μm, whereas, in the end-caps, the strips are arranged along the radial direction with variable pitch values [23–26].
   In the barrel region, each module is double layered with two sets of single sided detectors mounted back to back : the strips are parallel to the beam pipe in one layer (axial) and are rotated by 40 mrad on the other one (back side) to give a stereo view, allowing to obtain the z-measurement. Similarly, in the end-cap, one set is aligned radially while the other has a 40 mrad stereo angle α.
   The SCT provides thus 8 strip measurements corresponding to 4 space points, with a total of ∼6.3 million readout channels. The corresponding intrinsic spatial resolution is 17 μm in r-ϕ direction and 580 μm in $z(r)$ obtained using the small 40 mrad stereo angle between the back to back sensors in the barrel (end-cap) region [16].

3. **The Transition Radiation Tracker (TRT):** It is a gaseous detector consisting of drift tubes (straws) of 4 mm diameter each with a central wire, with good single point intrinsic resolution of 120 μm. The long lever arm allows to explore wide ranges in transverse momentum p_T and continuous tracking with typically at least 30 hits/track. The TRT provides measurements for electron identification as well since the use of Xe gas allows the detection of transition radiation photons.

   TRT Barrel: It consists of about 50 000 longitudinal (parallel to beam pipe) straws of 1.44 m maximum length, arranged in 3 layers of 32 modules each. The

[6]see subsection 4.6.3 for b-tagging details in ATLAS.

sense wires are split in two and read out at both ends. The sensors cover radial distances between $r = 0.563$ m to $r = 1.066$ m with $|\eta| < 1$. There are 105088 readout channels in total.

TRT Endcaps (sides A and C): They have radial ($\sim$0.4 m) straws arranged radially in 20 wheels (8 inner wheels and 12 outer wheels), covering the region $1 < |\eta| < 2.0$. The wires are read out at the outer end, which corresponds to 245 760 readout channels in total.

Gas Filling: A gaseous mixture of (Xe,CO_2, O_2)[7] at concentrations 70%, 27%, 3% respectively is chosen for stability and transition radiation performance.

Structure: The straws are Kapton tubes reinforced with carbon fibers and have a 31 μm gold plated tungsten wire in the center. The straw walls are at -1.5 kV while the central wire is grounded.

Position resolution: Charged particles with $p_T > 0.5$ GeV and $|\eta| < 2.0$ pass through 30 straws at least, yielding measurements with 118 (132) μm residual resolution in the barrel (end-caps) region, surpassing the design resolution.

TRT thresholds: To increase the probability for a charged particle to emit a photon (with typical energy 5-30 keV), multiple transitions are needed. Many thin radiator fibres (barrel)/foils (end-caps) inserted between straws are thus used to increase the emission probability and the xenon[8], the major component of the gaseous mixture, is a good X-ray absorber. The widely varying indices of refraction of the material between the straws causes even highly relativistic particles ($\gamma_{\text{Lorentz}} \geq 10^3$) to undergo transition radiation and leave a strong signal in straws. As a result, high threshold (HT) TRT hits are used to distinguish electrons from hadrons and MIPs. The TRT has thus a low threshold (300 eV) for MIPs and a high threshold (6 keV) for transition radiation hits.

2.4.1.2 Error Scaling

The measurements after the initial alignment show for high momentum tracks a 22.1$\pm$0.9 μm transverse impact parameter resolution, and a relative momentum resolution $\frac{\sigma_p}{p} = (4.83 \pm 0.16) \times 10^{-4}$ GeV$^{-1} \times p_T$ [17].

Ideally, the intrinsic resolution reflects the uncertainty on the detector measurements and the corresponding track-to-hit pull distributions are Gaussian centered at zero, with a width equal to one. However, data track-to-hit pull distributions have a width larger than one due to misalignments and other effects. In order to correct for such

[7] X-ray absorber+quenchers
[8] gas with high Z required for photon absorption

effects, an uncertainty is applied to the width based on the pull Gaussian fit width in a recursive way until the fitted width converges to unity. The final uncertainty is then expressed in terms of the intrinsic uncertainty σ_0 and the error scaling term C, accounting for misalignment and other effects as

$$\sigma \;=\; \sigma_0 \oplus C \;. \tag{2.8}$$

The pull width values before and after the error scaling in addition to the associated C values for 2012 data are summarized in Table 2.3 for the various ID parts. Additional details can be found in [38].

Detector	Pull σ before	Pull σ after	C (μm)
Pixel Barrel X	1.18	1.00	4.5
Pixel Barrel Y	1.13	1.00	33.5
Pixel End-cap X	1.06	1.00	3.6
Pixel End-cap Y	1.00	0.99	0
SCT Barrel	1.08	1.00	8.1
SCT End-cap	1.07	1.00	8.7
TRT Barrel	0.96	0.96	0.0
TRT End-cap	0.96	0.96	0.0

Table 2.3: Table summarizing the track-to-hit pull widths before and after scaling in addition to the corresponding C values for various parts of the ATLAS Inner detector [38].

2.4.2 Calorimeters

Calorimeters are present in high energy physics experiments and have been introduced to measure particle energies. The main advantages of calorimeters are [27]:

- In addition to estimating particle energies from energy deposits, calorimeter measurements can also be used to determine the position and direction (angle) of the particle shower, distinguish different particle types (e.g. electrons and photons from pions) and timing for charged and neutral particles.
- The calorimeters produce fast, easy to interpret signals ($<$ 100 ns feasible)[9] which can be used for triggering.
- Unlike spectrometers, calorimeters allow the detection of both charged and neutral particles. Also, for calorimeters, the higher the energy, the better the resolution, in contrast with magnetic spectrometers whose energy resolution is degraded as the momentum increases.
- The shower depth dependence on the particle energy is logarithmic. Thus, they are space and cost effective.

[9]This is the case of the $PbWO_4$ homogeneous CMS calorimeter. The sampling Pb-LAr EM calorimeter in ATLAS is not as fast (450 ns).

2.4.2.1 Basic Definitions

Calorimeters can be categorized as '*electromagnetic*' calorimeters mainly used for photon and electron measurements based on electromagnetic (EM) interactions (bremsstrahlung, pair production, etc.), or as '*hadronic*' calorimeters to detect hadrons.

If the calorimeter is built using a single material type for energy degradation and generation of the detectable signal, it is called '*homogeneous*'. In a '*sampling*' calorimeter, on the other hand, layers of 'absorber' medium needed for energy degradation and 'active medium' (generating signals from the showers of lower energy obtained at 'absorber' level) are alternated.

2.4.2.2 Performance Requirements of Calorimeters in ATLAS

Various technical and physics performance requirements are imposed on calorimeter operation. A good calorimeter should have/allow [27]:

- Fast response with high performance read out and good pile-up handling (enhanced with fine granularity).
- Radiation hardness because the large particle flux at LHC is accompanied by high radiation levels in particular in the forward region near $|\eta|$=5.
- Good angular coverage for optimal particle identification and total transverse energy measurement.
- Excellent electromagnetic energy resolution.
- Angular measurements with good resolution (needed for mass calculations, missing transverse energy, etc).
- Large particle energy range coverage, varying between a few GeV and a few TeV.
- Jet energy resolution and linearity: For the LHC physics goals, a jet resolution of $\sim \frac{50\%}{\sqrt{E\ [GeV]}} \oplus 3\%$ is needed with a linearity better than 2% on the jet energy (up to ∼4 TeV). Run-I measurements show a linearity of 1-2% for 35 GeV $\leq p_{\mathrm{T}}^{\mathrm{jet}} <$ 210 GeV (derived using Z+jets events), and 1% for 120 GeV $\leq p_{\mathrm{T}}^{\mathrm{jet}} <$ 700 GeV (derived using γ+ jets) [28]. Details on the hadronic energy resolution and the corresponding LHC Run-I measurements are presented in sec. 2.4.2.5.
- Efficient particle identification and rejection of fake (mis-identified) ones.

2.4.2.3 ATLAS Calorimeters : Overview

ATLAS has in fact electromagnetic and hadronic sampling calorimeters with ϕ-symmetry and full coverage: liquid argon calorimeters (LAr) are closest to the beam axis and are followed by tile calorimeter(s) (central barrel and two extended barrels) as shown in Fig. 2-4. The latter are hadronic calorimeters in the barrel region, while LAr calorimeters are arranged in one barrel and two end-cap cryostats such that the barrel has a LAr electromagnetic calorimeter (EMB), whereas each end-cap contains an electromagnetic end-cap calorimeter (EMEC), a hadronic end-cap calorimeter (HEC) and a forward calorimeter (FCAL) arranged as shown in Fig. 2-7(b). The choice of LAr as the active detector medium was motivated by the intrinsic linear and uniform behavior of liquid argon, its stable response over time and its radiation-hardness [14]. The expected electron and photon calorimeter energy measurements at LHC span the range [5-5000] GeV. A corresponding energy linearity better than 0.5% is required to meet the physics requirements at LHC.

2.4.2.4 Electromagnetic Calorimeters

EM calorimeter measurements are provided by the LAr sampling calorimeters with a coverage in pseudorapidity up to $|\eta| < 4.9$. The precision EM calorimeters are lead-liquid argon detectors with accordion geometry (described hereafter), where liquid argon is the active medium and lead is the absorber. These calorimeters are segmented in depth with several active longitudinal layers: the precision-measurement region ($0 < |\eta| < 2.5$) has 3 layers, while only two layers are associated to the higher-η region ($2.5 < |\eta| < 3.2$) and the EMB-EMEC overlap region. As shown in Fig. 2-7, the first of the 3 EMB layers is finely segmented in η and is 4.3 X_0 deep, allowing π^0 rejection and separation of 2 photons. The second layer is 16 X_0 deep and collects most of the EM shower energy (∼80%). And the third layer, with 2 X_0 depth, has a coarser segmentation in η since it is mainly to contain late starting showers or shower tails. The coverage in η is extended through FCAL to cover regions with $3.1 < |\eta| < 4.9$. The energy lost in front of the EM calorimeters outside the ID is recovered through a liquid argon layer only, referred to as PreSampler (PS), covering the region $0 < |\eta| < 1.8$. A summary of the basic details of EMB, EMEC and FCAL is given below.

The precision measurement region of the EM calorimeter is restricted to the area defined by $|\eta| < 2.5$ [14, 27, 29].

- **EM Calorimeter Energy Resolution:**
  The energy resolution of an EM calorimeter is parametrized as follows:

$$\frac{\sigma(E)}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c \qquad \oplus \text{ means quadratic sum ,} \tag{2.9}$$

  where:

  – *a:* stochastic term describing fluctuations related to physical development of the shower. It is about (1/3)% in scintillating crystals, 0.5% in

Cherenkov radiators and 10% in ATLAS Pb-LAr sampling calorimeters.

- *b:* is the noise term coefficient, leading to lowered resolution at lower energies. Contributions to this term come from electronic noise or pile-up in the shower area.
- *c:* is the constant term and is the dominant term at high energies. It is related to various nuisance sources e.g. calibration stability, radiation effects, energy leakage, non-uniformity of material, energy losses in dead material, etc.

In ATLAS, mass measurements from $Z \to ee$ Monte Carlo (MC) simulation and data samples are used to check the modeling of the stochastic and noise terms. The constant term, c_{data} is then evaluated in data using the width of the mass distributions [30]:

$$c_{\text{data}} = \sqrt{2 \times \left(\left(\frac{\sigma}{m_Z} \right)^2_{\text{data}} - \left(\frac{\sigma}{m_Z} \right)^2_{\text{MC}} \right) + c^2_{\text{MC}}} \ , \tag{2.10}$$

where c_{MC} is the constant term in MC samples, m_Z is the Z boson mass. The distributions are fitted using a convolution of Breit-Wigner and Crystal Ball functions, and σ is the Gaussian component of the associated resolution. The values of c_{data} for all the η ranges covered by the calorimeter are summarized in Table 2.4.

Subsystem	η**-range**	**Effective constant term,** c_{data}
EMB	$\lvert\eta\rvert < 1.37$	$1.2\% \pm 0.1\%(\text{stat.})^{+0.5\%}_{-0.6\%}(\text{syst})$
EMEC-OW	$1.52 < \lvert\eta\rvert < 2.47$	$1.8\% \pm 0.4\%(\text{stat.}) \pm 0.4\%(\text{syst})$
EMEC-IW	$2.5 < \lvert\eta\rvert < 3.2$	$3.3\% \pm 0.2\%(\text{stat.}) \pm 1.1\%(\text{syst})$
FCal	$3.2 < \lvert\eta\rvert < 4.9$	$2.5\% \pm 0.4\%(\text{stat.})^{+1.0\%}_{-1.5\%}(\text{syst})$

Table 2.4: Table summarizing the measured values of the constant EM calorimeter energy resolution term for various η regions in ATLAS [30]. OW and IW stand for outer wheel and inner wheel in the end caps respectively.

The design resolution values are $\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E[\text{GeV}]}} \oplus 0.7\%$ (EMEC+EMB) and $\frac{\sigma_E}{E} = \frac{100\%}{\sqrt{E[\text{GeV}]}} \oplus 10\%$ for the FCal. Combined test beam measurements show a stochastic term of $(10.1 \pm 0.4)\% \sqrt{\text{GeV}}$ for both EMEC and EMB, and $(28.5 \pm 1.0)\% \sqrt{\text{GeV}}$ for the FCal (electron test beams).

- **EM Calorimeters (EMEC+EMB):**
  - Pb absorber

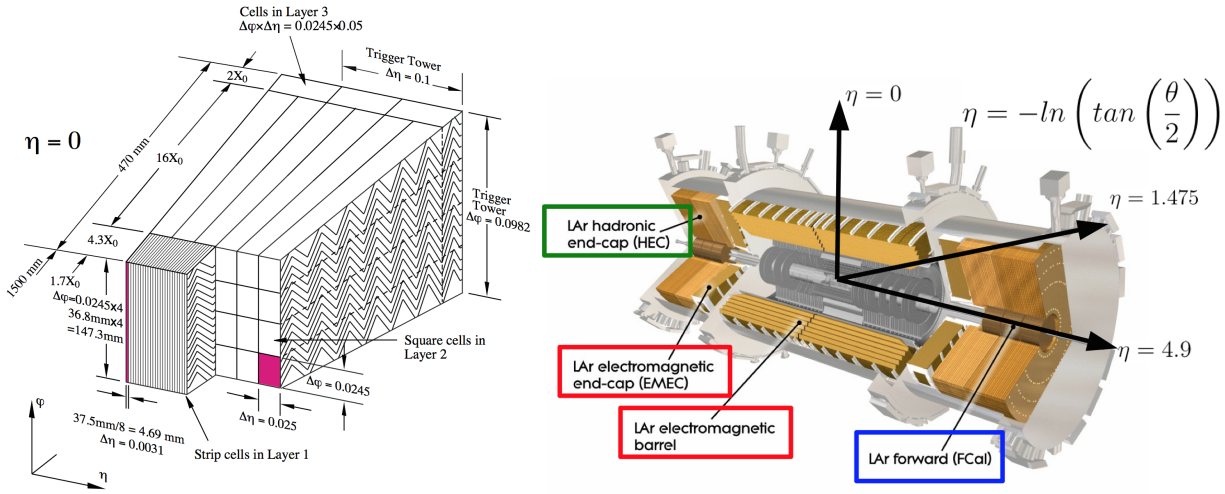
- Pb thickness varying in regions of η to limit the degradation of the sampling fraction. The Pb plates thickness is 1.53 mm for $|\eta| < 0.8$ and 1.13 mm otherwise in the barrel region, while it is 1.7 mm for $|\eta| < 2.5$ and 2.2 mm otherwise in the end-caps.
- Coverage: $|\eta| < 3.2$
- 3(2) longitudinal layers for $|\eta| < 2.5$ ($2.5 < |\eta| < 3.2$) and a presampler for $0 < |\eta| < 1.8$
- ∼170k channels (99.9% operational)
- Moliere radius $R_M = 7.3$ cm
- Temperature sensitivity to deposited energy $5\%/^\circ K$

- **FCAL (EM Section):**
  - Cu/W absorber: Cu absorber in FCAL1 used as EM calorimeter (1 layer).
  - Coverage: $3.2 < |\eta| < 4.9$.
  - 3524 channels, 99.8% of which are operational (FCAL total).

- **The Accordion Structure:**
  The accordion structure was chosen for the precision EM calorimeters (EMEC+EMB) to ensure coverage without any cracks in ϕ, and a fast signal extraction at both ends of the electrodes. The accordion waves are oriented axially and run in ϕ in the barrel region, while those at end-caps are along the radial direction, running axially. In both cases, since the liquid argon gap varies with the radius, the waves folding angle is varied accordingly. As a result, the accordion geometry ensures linearity and uniform resolution in ϕ.

- **Detector Status and Performance:**
  76 of the 173312 (0.04%) LAr EM calorimeter channels were not functional during Run-I due to technical defects either inside the cryostat or in the read-out electronics. For the FCAL, the number is 8 out of 3524 (0.23% in total). In addition, trigger and signal control were lost for 6 FEBs (front-end boards) and a control board in 2011, and the problem was completely solved after six months during the winter shutdown. The amount of collected data suitable for physics went from 97% in 2011 to more than 99% in 2012 due to the reduction of the high-voltage (HV) trips and noise bursts, leading to 0.46% and 0.2% of data losses respectively (in 2012).

  During the 2013-2014 shutdown, the LAr calorimeter was upgraded to be able to handle high trigger levels in Run-II. Additional upgrade plans are foreseen for the High Luminosity LHC expected to run with an integrated luminosity of ∼3000 fb^{-1}, which is beyond the radiation levels and luminosity for which the calorimeter was designed to work with [31, 32].



(a) (b)

Figure 2-7: Segmentation of the EM calorimeter barrel module in (η, ϕ) [14]. The various parts of the EM LAr calorimeter are shown in (b)[15].

2.4.2.5 Hadronic Calorimeter

In this section, a brief overview of the hadronic calorimeter energy resolution is presented first. Then, the various ATLAS hadronic calorimeters are described: the tile calorimeter, HEC and FCAL.

- **Hadronic Calorimeter Energy Resolution:**

  Hadronic showers have invisible energy due to nuclear effects (e.g. soft neutrons, binding energy, etc.), which is not detectable with practical instruments or detectable with reduced efficiency. The energy resolution in hadronic calorimeters is subject to strong intrinsic fluctuations of nuclear interactions producing EM secondaries (e.g. $\pi^0 \rightarrow \gamma\gamma, \eta \rightarrow \gamma\gamma$), excitation and break-up energy rarely detected, energy needed to extract nucleons, high energy muons and neutrinos escaping the calorimeter. The resulting invisible energy accounts for 40% of the non-EM energy loss. The main components of hadronic showers are [33]:

  - visible EM ($e^{\pm}, \gamma, \pi^0 \rightarrow \gamma\gamma$)$\sim$O(50%)
  - visible non-EM energy (e.g. ionization of $\pi^{\pm}$)$\sim$ O(25%)
  - invisible (nuclear break-up, nuclear excitation) and escaped energy$\sim$ O(25%)

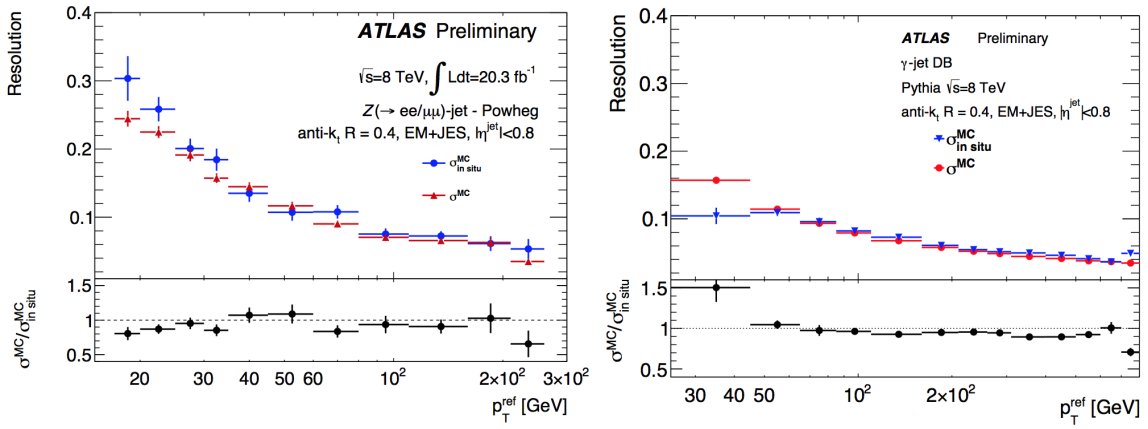
  In the detector, only the visible energy is sampled. The response is in fact non-linear and requires compensation since the efficiency of the hadronic and EM components of the hadronic shower energy measurements are different, and the fraction of energy deposited hadronically (e.g. visible non-EM energy deposited fraction) is energy dependent. The hadronic calibration as described in chapter 4 provides software compensation and accounts for the invisible and

escaped energies.

The fractional hadronic energy resolution for the tile calorimeter, HEC and FCAL is parametrized as follows:

$$\frac{\sigma(E)}{E} = \frac{a}{\sqrt{E\,[\mathrm{GeV}]}} \oplus b \tag{2.11}$$

where a is the stochastic term and b is the constant term related to calorimeter non-uniformities. The overall jet energy resolution is illustrated in Fig. 2-8. The associated uncertainty is about 20% if 20 GeV $\leq p_{\mathrm{T}}^{\mathrm{jet}} <$ 80 GeV (using Z-jet events), and 10% for 80 GeV $\leq p_{\mathrm{T}}^{\mathrm{jet}} <$ 700 GeV (using γ-jet events)[28].



(a) Z–jet anti-k_t $R = 0.4$ EM+JES

(b) γ–jet anti-k_t $R = 0.4$ EM+JES

Figure 2-8: Jet energy resolution as determined in situ and extracted from Monte Carlo samples for Z-jet events (a) and γ-jet events (b) [28].

- **Tile Calorimeter:**

  Basic description: [14, 34]

  - Central hadronic calorimeter in ATLAS.
  - Pseudorapidity coverage: $|\eta| < 1.7$.
  - 3 cylindrical layers (radial depth is $\sim 7.4\ \lambda$ (interaction length)).
  - ∼10 000 channels or $\sim$ 5 000 cells.
  - 256 modules arranged in 4 partitions (EBA, LBA, LBC, EBC) segmented in 64 wedge-shaped modules each (EB/LB refers to extended barrel/long barrel regions).
  - 12 m long (5.8 m for central barrel and two extended barrels 2.6 m long each) with 2.28 (4.25) m as inner (outer) radius.

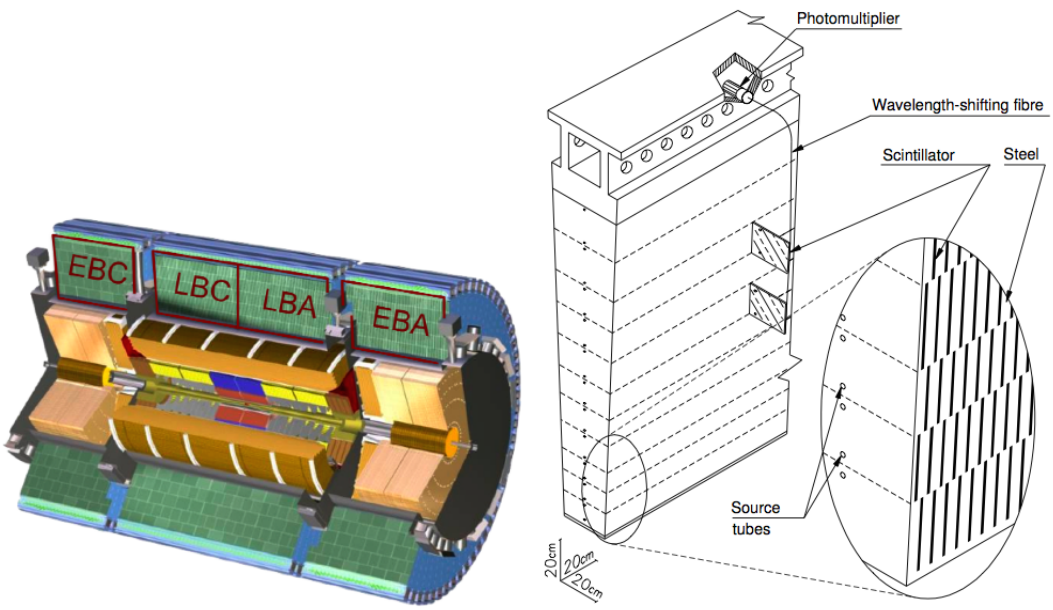
– Sampling calorimeter: the absorber is made of steel and the active medium is a scintillator (plastic scintillating tiles).

– The hadronic energy resolution measured using test beams is $\frac{\sigma(E)}{E} = \frac{52.9\%}{\sqrt{E}} \oplus 5.7\%$ (see eq. (2.11)). It is consistent with the design expectations $\frac{\sigma(E[GeV])}{E[GeV]} \sim \frac{50\%}{\sqrt{E[GeV]}} \oplus 3\%$ [35].

– e/π response = 1.33.

Structure:

Each tile cylinder has 64 independent modules as shown in Fig. 2-9 with an almost periodic steel-scintillator structure. The scintillator tiles are inserted in the steel structure. The produced light, proportional to the deposited energy in the plastic scintillating tiles, is then collected at the tile edges by wavelength shifting fibers (WSF) and sent to photomultipliers (PMT). Fibres are grouped to define 3-dimensional cells, which form three resulting layers of radial depths 1.5, 4.1 and 1.8 λ at $\eta = 0$. The corresponding readout granularity is $\Delta\eta \times \Delta\phi = 0.1 \times 0.1 (0.2 \times 0.1$ in the outermost layer) with most cells being read by 2 PMTs.



(a) (b)

Figure 2-9: Graphical representation of the tile calorimeter layers [15] (a) and schematic drawing of the tile calorimeter mechanical structure and readout (b) [14].

Detector Status:

In 2011, 99.2% of the collected measurements are good data for physics analysis. Of the 5182 tile calorimeter cells, ∼5% were non-usable. This number was reduced to 3% in 2012 (1% up to August 2012), with a 99.6% tile data quality efficiency for p-p collisions [36]. Figure 2-10 shows the evolution of the percentage of non-reconstructed tile calorimeter faulty cells as a function of time during 2011-2012 data taking period. The major reasons of the observed inefficiency

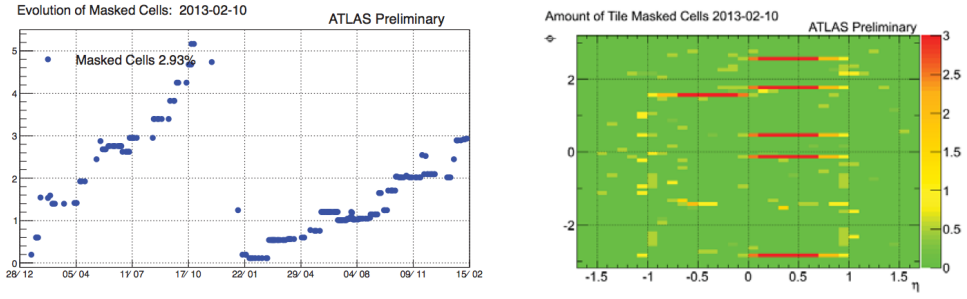


Figure 2-10: Evolution of the percentage of non-reconstructed tile calorimeter faulty cells (masked cells) during Run-I 2011-2012 data taking period (left). The 2D-histogram (right) shows the (η, ϕ) positions of the masked cells at the end of 2012 [36]. The z-axis shows the number of tile cells corresponding to each (η, ϕ) position.

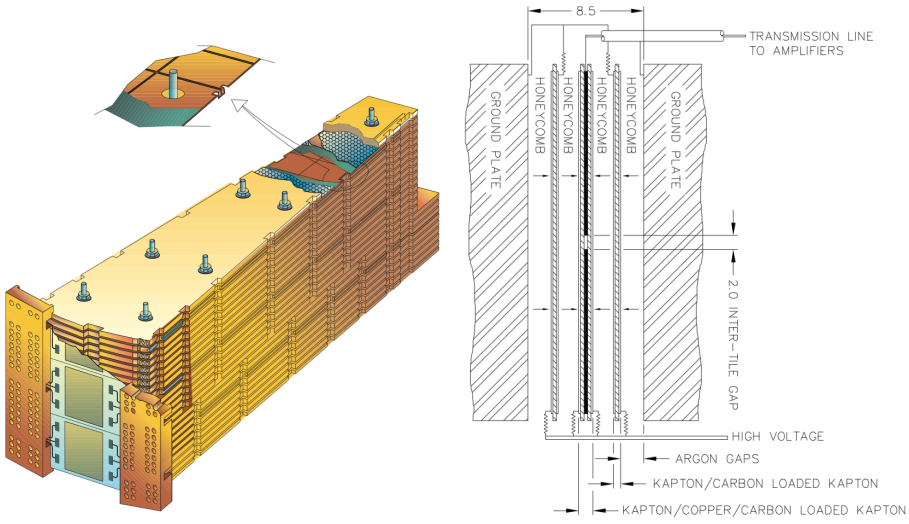
are:

- high voltage (HV) PMT trips or technical problems with the back-end electronics, which leads to improper read-out of at least four consecutive modules.
- channel timing problems with 25 ns reconstruction delay.

- **HEC:**
  - LAr active medium and Cu absorber (flat plate)
  - Pseudorapidity coverage: $1.5 < |\eta| < 3.2$
  - 2 wheels in each end-cap with 32 ϕ modules and 2 readout depths each
  - 4 longitudinal layers
  - 5632 channels (99.6% operational for the same reasons mentioned earlier in sec. 2.4.2.4 under "Detector status and performance")
  - Test beam measurements fitted according to eq. (2.11) show a hadronic resolution with $a = (70.6 \pm 1.5)\% \sqrt{\text{GeV}}$ and $b = (5.8 \pm 0.2)\%$. In the transition region to the FCal, near $|\eta| = 3.2$, the resolution degrades with a = 81-85% $\sqrt{\text{GeV}}$ (close to the 100% FCal design value). Results are compatible with the design hadronic fractional resolution: $\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E}} \oplus 3\%$ (E in GeV).

– Granularity: $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ if $|\eta| < 2.5$ and (0.2×0.2) otherwise

A graphical representation of a HEC module and the corresponding readout electronics is given in Fig. 2-11.



(a) (b)

Figure 2-11: Graphical representation of a HEC module (a) and the corresponding readout electronics (b) [14].

- **FCAL (Hadronic Section):**
  - 2 layers: FCAL2 and FCAL3, W absorber
  - Pseudorapidity coverage: $3.2 < |\eta| < 4.9$
  - Exposed to high radiation levels since it is a forward calorimeter, covering high η regions with large particle flux. Consequently, small LAr gaps are needed ($<$ 2 mm gap in the EMB).
  - Tungsten rods used in FCAL2 and FCAL3 for optimized absorption length [14]
  - Test beam measurements fitted according to eq. (2.11) show a hadronic resolution with $a = (94.2 \pm 1.6)\% \sqrt{\text{GeV}}$ and $b = (7.5 \pm 0.4)\%$ using a single weight (flat) per module method. Applying a more sophisticated method, with radial weights and taking into account the FCal segmentation, improves the stochastic term value from 94% to 70% and the constant term goes from 7.5% to 3.0%. These values are to be compared with design hadronic resolution: $\frac{\sigma_E}{E} = \frac{100\%}{\sqrt{E}} \oplus 5\%$ for $3 < |\eta| < 5$ (E in GeV).

2.4.3 The Muon Spectrometer

The muon spectrometer (MS) consists of:

- three large air-core toroids (described in sec. 2.4.4) delivering a magnetic field varying with (η, ϕ).
- precision tracking chambers i.e. monitored drift tubes (MDTs) and cathode strip chambers (CSC)
- fast response chambers used for trigger, namely resistive plate chambers (RPC) in the barrel and thin-gap chambers (TGC) in the end-cap

Muons are detected within $|\eta| < 2.7$ (in MDTs and CSCs) and triggered on within $|\eta| < 2.4$ (in RPCs and TGCs) [14]. The MS measurements cover muons momenta falling in the range [3 GeV- 3 TeV]. The main parameters of the ATLAS muon system are summarized in Table 2.5 for both combined and stand-alone performances, while the main MS parts are illustrated in Fig. 2-12

Parameter	**ATLAS**
Pseudorapidity coverage	
-Muon measurement	$\|\eta\|$<2.7
-Triggering	$\|\eta\|$<2.4
Dimensions (m)	
-Innermost (outermost) radius	5.0 (10.0)
-Innermost (outermost) disk (z-point)	7.0 (21-23)
Segments/superpoints per track for barrel (end caps)	3 (4)
Magnetic Field B (T)	
-Bending power (BL, in T.m) at $\|\eta\| \approx 0$)	3
-Bending power (BL, in T.m) at $\|\eta\| \approx 2.5$)	8
Combined (stand-alone) momentum resolution at	
-p = 10 GeV and $\eta \approx 0$	1.4%(3.9%)
-p = 10 GeV and $\eta \approx 2$	2.4%(6.4%)
-p = 100 GeV and $\eta \approx 0$	2.6%(3.1%)
-p = 100 GeV and $\eta \approx 2$	2.1%(3.1%)
-p = 1000 GeV and $\eta \approx 0$	10.4%(10.5%)
-p = 1000 GeV and $\eta \approx 2$	4.4%(4.6%)

Table 2.5: The main parameters of the ATLAS muon system for both combined and stand-alone performances [37].

The precision measurements are provided by MDTs covering the range $0 < |\eta| < 2.7$ everywhere in the MS except for the forward region $2 < |\eta| < 2.7$ in the end-cap innermost layer, where they are replaced by CSC chambers with higher granularity

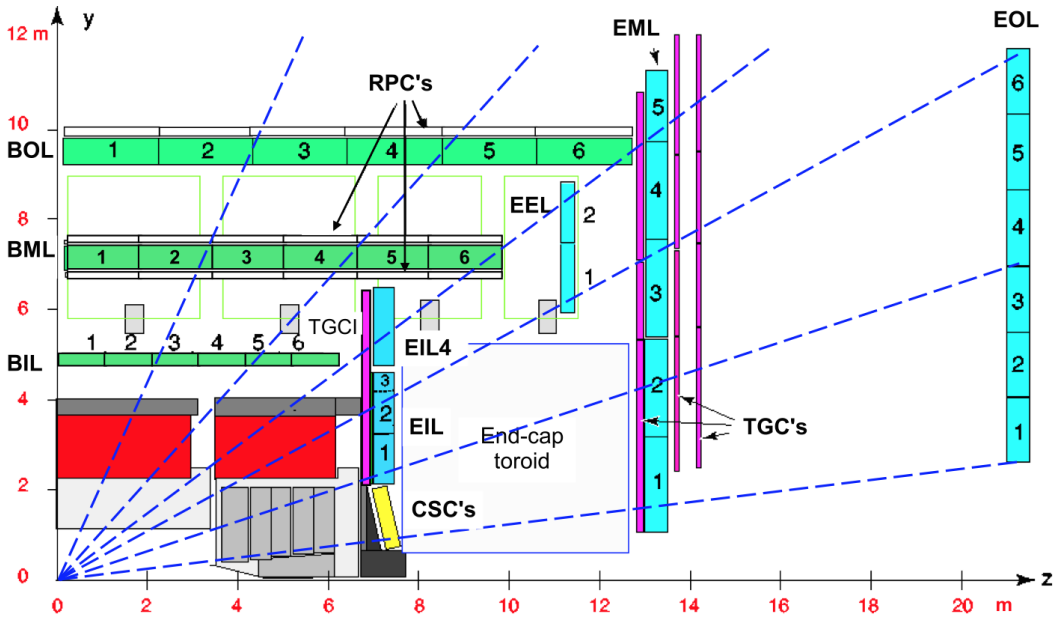


Figure 2-12: Schematic representation of the ATLAS muon system in plane containing the beam axis (bending plane). Blue dashed lines show trajectories of muons with infinite momentum [14].

because of the higher counting rates. The MS is mainly designed for stand-alone muon reconstruction, with [14]:

- 30-50 μm precision in the drift tubes
- fast trigger rates in the barrel (RPC with fast rate $\sim$ 5 ns) and the endcap (TGC)
- p_T resolution $\leq$ 4% up to 100 GeV and $\leq$ 10% up to 1 TeV
- $\sim$ 1.1 million channels

The details of muon reconstruction are given in chapter 4 and the muon trigger is described in section 2.4.5.

- **MDT:**
  - for muon tracking
  - Two multi-layered 2.997 cm diameter drift tubes along the ϕ direction, within a gaseous mixture of (Ar,CO_2), at concentrations 93% and 7% respectively, operated at an absolute pressure of 3 bars (gas gain$\sim 2 \times 10^4$)
  - Gold-plated W/Re (97/3) 50 μm diameter wires (3080 V)
  - 1150 chambers, 354 k channels
  - Average resolution of 80 μm/tube or 35 μm/chamber

- η coverage: range $|\eta| < 2.7$ except for the forward region $2 < |\eta| < 2.7$ in the innermost end-cap layer

- **CSC:**
  - for muon tracking
  - proportional wire chambers with cathodic strip readout in 4 planes
  - high granularity, providing high space, time and track resolution, in addition to high capability to handle high particle rates and low sensitivity to neutrons
  - 32 chambers, 30.7 k channels
  - covers the forward region $2 < |\eta| < 2.7$ in the innermost end-cap layers
  - resolution of 40 μm (r direction) in the bending plane and 5 mm in the transverse plane[10]

- **RPC:**
  - Level-1 muon trigger chambers made of gaseous parallel-plate detectors (in saturated avalanche mode)
  - 606 chambers, 373 k channels
  - coverage up to $|\eta| < 1.05$
  - 10 mm spatial resolution in both z and ϕ direction

- **TGC:**
  - multi-wire proportional chambers in saturated mode with high rate capability used for trigger
  - gas mixture of carbon dioxide and n-pentane (55% - 45%)
  - 3588 chambers, 318 k channels
  - covering $1.05 < |\eta| < 2.4$ regions
  - spatial resolution of 2-6 mm (r) and 3-7 mm ϕ

The sagitta (defined in Fig. 2-13) resolution in the precision tracking chambers is achieved for MDT wires and CSC strips locations known to better than 30 μm. The alignment is done using a very precise optical alignment system in addition to track-based algorithms. Additional details can be found in [14, 39].

[10] this difference is due to different readout pitches.

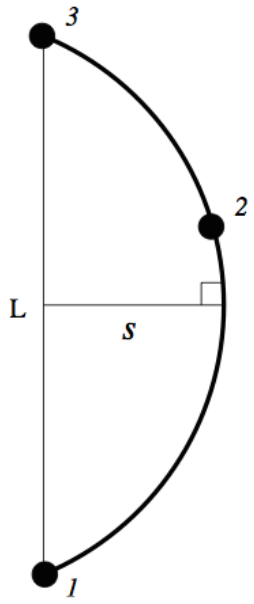


Figure 2-13: Figure showing the definition of the sagitta 's' as the maximum deviation of the curved trajectory from a straight line. The definition is shown for 3-point trajectories, which is the case of muon tracks in ATLAS. L is the distance between the end points 1 and 3.

2.4.4 The Magnet System

ATLAS has 4 large superconducting magnets, namely one central solenoid and three toroids (one in the barrel and two at the end caps), providing the magnetic field in a 12 000 m^3 volume for various detector parts. Charged particles are bent by the magnetic field and the trajectory curvature allows good momentum measurements.

- **Inner solenoid:** provides the inner detector with a 2 T axial magnetic field aligned along the beam axis. With such a high intensity field, even the trajectory of very energetic particles is bent enough for momentum calculations. The material thickness in front of the EM calorimeter is reduced to best minimum value allowing optimal calorimeter performances. The stored energy is 38 MJ.

- **Outer toroidal magnetic field:** within the muon system and outside the calorimeter area is provided by the barrel toroid (8 air-core superconducting barrel loops) and the two end-cap air-toroidal magnets needed to optimize the bending power in the muon spectrometer in the end-cap regions. The toroidal magnetic field is not uniform, with bending power varying between 2 and 8 Tm[11], with 1.6 GJ stored energy of which 1.10 GJ are for the barrel toroid. The barrel toroid has indeed 8 air-core superconducting coils evenly positioned around the beam axis, with a 20 m outer diameter. Each of these coils is inside a vacuum vessel of dimensions 26 m × 5 m × 1 m and weighs 85

[11] 4 T peak field in the barrel.

tons. Its structure allows for other ATLAS services e.g. cooling pipes, cables, accessibility for maintenance, etc.

The magnetic systems requires a critical current of 20.5 kA at 4 T for the solenoid conductor and 65 kA for the barrel and end-cap toroids. The superconductors are cooled at 4.8 K [14].

2.4.5 Trigger

The simplest definition of a trigger would be : *a system allowing to decide, in real time, whether to keep or discard the measurements coming from each observed interaction.* It is needed to filter the event stream based upon a 'quick look' at the data, testing a large number of hypotheses (*trigger chains*) independently. The main goal is to keep only interactions of interest for later analysis, getting rid of uninteresting collisions at high efficiency. Higgs and BSM physics, one of the main reasons for which the ATLAS detector was built, put thus important constraints on the trigger system. Yet, the interactions acceptance rate must be low enough to allow storage and reconstruction. This filtering and information reduction is needed for 2 main reasons:

1. Today's technology does not allow the recording of the entire amount of raw data at the current instantaneous luminosity $\approx$ 30 TB/s on average in 2012 (peak value of 32 TB/s during the 2012 data taking) [40].

2. If all events are recorded, we end up saving a lot of uninteresting "background" events. New interesting physics is buried under a lot of "uninteresting" events. The solution to this is a fast processing, flexible, affordable trigger with a high rejection factor and a high efficiency for interesting physics.

2.4.5.1 Online Vs Offline

Online trigger happens during data taking, while the offline selection is applied on stored data sets for event reconstruction and particle identification. Various limitations prevent the same online/offline performance in event reconstruction. Using the saved trigger decision as part of the raw data, analyses can match the objects reconstructed offline and the triggered objects.

2.4.5.2 ATLAS Trigger

The ATLAS trigger system consists of a level 1 (L1) hardware trigger and of a high level trigger (HLT), where the latter includes level 2 (L2) and the Event Filter (EF). In 2012, starting with 20 MHz crossing rate, the rate at the EF is $\simeq$ 700 Hz on average ($\simeq$ 1kHz peak value) with a factor of rejection of about 30 000 and 1 GB/s (1.6 GB/s peak value) average bandwidth. The "online" reduction is by about 99.99% [40]. The trigger and DAQ[12] were successful at and beyond the design values during Run-I. Fig. 2-14 shows the design and actual peak values for 2012 various trigger and DAQ

[12]Data Acquisition

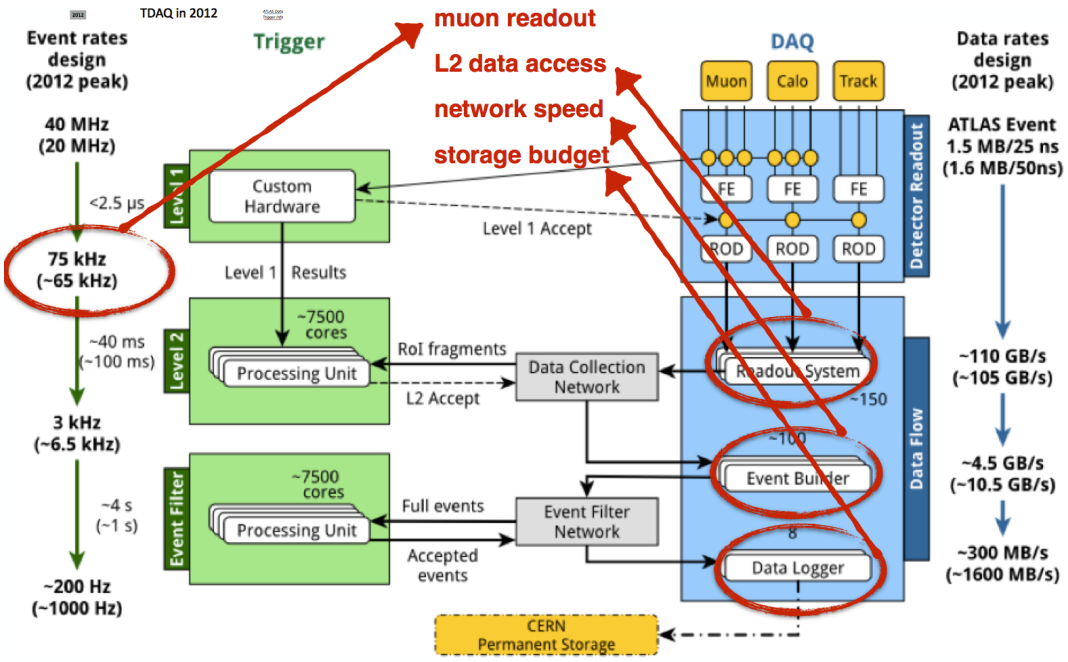


Figure 2-14: Schematic representation of trigger performance in ATLAS during 2012. Design values are shown in addition to the actual peak values [40]. Items circled in red are the ones that need to be adjusted for Run-II requirements.

stages, where the items circled in red do not scale well with Run-II specifications. The corresponding rate evolution of various triggers during 2012 is shown in Fig 2-15 in addition to the efficiency with respect to offline trigger for electrons, muons and jets in Fig 2-16. At the EF level, efficiency measurements as a function of transverse momentum show a plateau at 95% for electrons and at 86(70)% for muons in the end-cap(barrel) regions. The plot in Fig 2-16(c) is for EM and EM+JES[13] jets. The tau triggers and their efficiency are discussed, in the context of object reconstruction in ATLAS, in sec. 4.7.

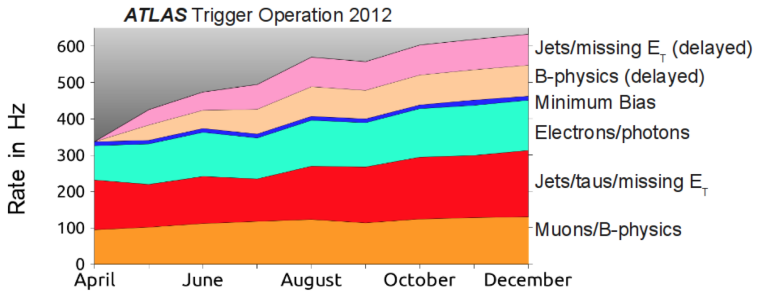
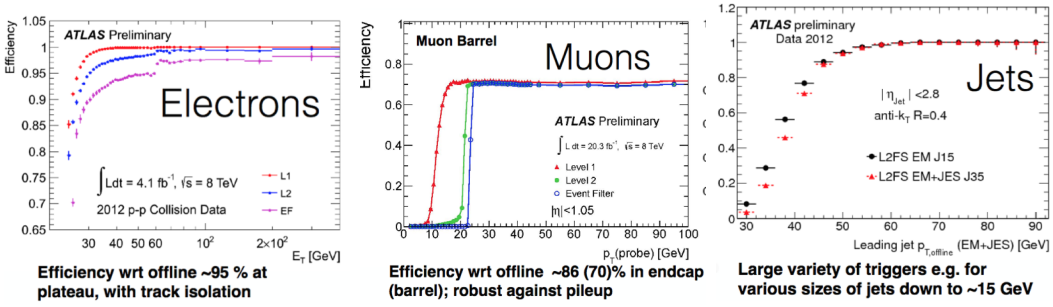


Figure 2-15: Figure showing trigger rates during 2012 data taking in Run-I [41].

[13]EM and EM+JES (jet energy scale) jets are defined in section 4.6.



(a) (b) (c)

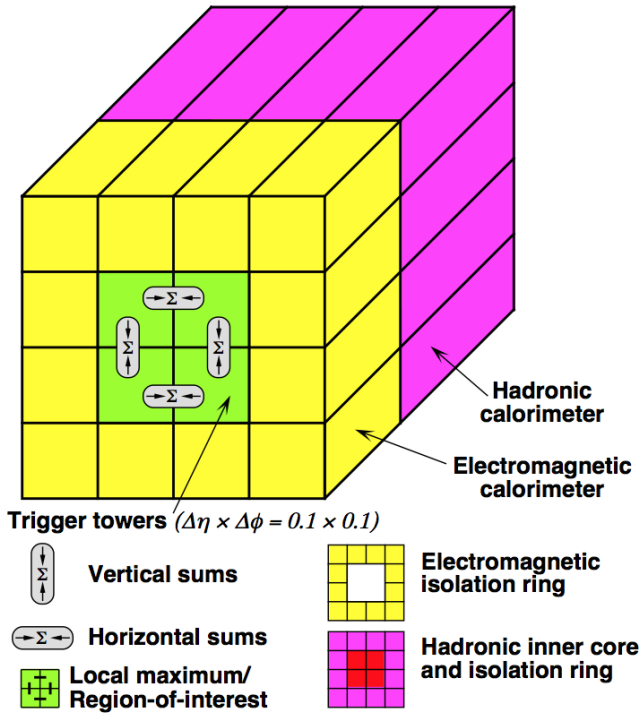
Figure 2-16: Figure showing efficiency with respect to offline trigger for electrons (a), muons (b) and jets (c) as a function of the transverse momentum [41].

1. **Level-1 (L1) Trigger**
   The L1 hardware trigger uses coarse-granularity information from the calorimeter and the muon spectrometer trigger hardware, i.e. L1Calo and L1Muon respectively, in addition to information from forward detectors, to make the decision and identify the regions of interest 'ROIs' (μ, jet, e/γ, τ, E_T^{miss}) [14, 42]. The trigger uses calorimetry trigger towers. An example of such towers is shown in Fig. 2-17.
   L1 trigger has 3 main parts: the PreProcessor, the Cluster Processor and the Jet/Energy-sum Processor. The calorimeter trigger runs in parallel the jet/energy sum processor and the cluster processor used to identify electrons, photons and taus. The decision is made by the central trigger processor (CTP) and accepted data is sent to the read-out system (ROS) and will be accessed by HLT afterwards [43].

   - **L1 muon trigger** uses information from RPCs (Resistive Plate Chambers) in the barrel and TGCs (Thin Gap Chambers) in the end caps. Preliminary rough reconstruction of (p_{T}, η, ϕ) muon parameters is made using hits in these detectors. A graphical representation is given in Fig. 2-18. The third layer of RPC is used for high-p_T muons.

2. **Level-2 (L2) Trigger**

   - Software based trigger using the full detector granularity (sub detector info can be combined).
   - Uses L1 ROIs as seeds of reconstruction. Each chain has various steps and rejection can occur at any step. At each of these steps, feature extraction (FEX) and hypothesis testing (Hypo) algorithms are applied. For an event to be accepted, at least one chain needs to survive the final step.



(a)

Figure 2-17: Figure showing trigger towers for e/γ reconstruction [14].

- Only data within ROI are used. The data information of accepted events is collected from ROS to form event data structure, that will then be communicated to the EF. The L2 algorithms are very specialized and fast.
- Only data requested by the algorithms are read out [44].

3. **EF**

- Uses complete event information and is seeded by L2.
- Uses offline algorithms.
- Pure software trigger (high flexibility).
- Accepted events are written to mass storage.
- Event Rate in 2012 of $\simeq$ 700 Hz on average ($\simeq$ 1kHz peak value) is better than the design value.

Additional information on the trigger performance during Run-I and the upgrade preparation for Run-II can be found in [45–47].

2.5 Conclusion and Prospectives

In this chapter, a brief overview of the LHC operation and performance was presented in addition to presenting the details of the various parts of the ATLAS detector : the

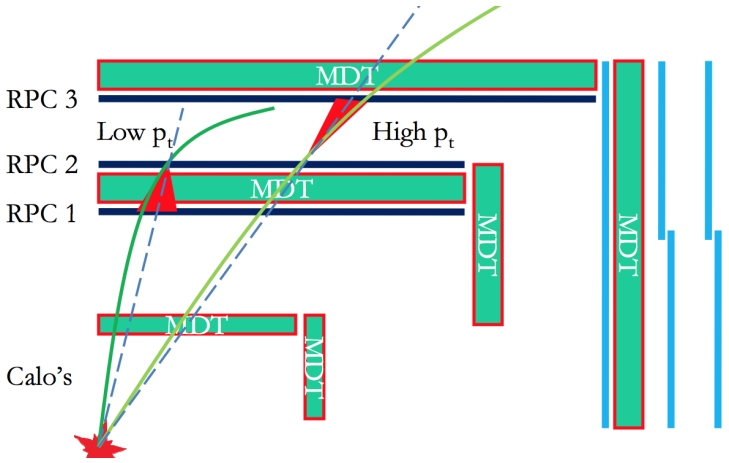


Figure 2-18: Graphical representation of the muon trigger [14]. High-p_{T} and low p_{T} refer to the L1 muon trigger thresholds.

inner detector (tracking system), the calorimeters (electromagnetic and hadronic), the muon spectrometer, in addition to the magnet system and the trigger. Various upgrades were performed during the long shutdown I (LS-I during 2013-2014) to make the necessary adjustments before the LHC Run-II. With the increasing number of collisions, the inserted (innermost fourth) layer in the pixel detector will lead to better vertex identification, which is crucial to identify interesting collisions. The improvements during the LS-I include also the following areas [48]:

- General infrastructure (electrical power, cryogenic and cooling systems).
- Simulation, reconstruction and analysis software.
- Management tools of data generation on the GRID (new ones were implemented).

Another round of improvements is foreseen during LS-II.

Bibliography

[1] http://home.web.cern.ch/topics/large-hadron-collider

[2] CERN, *The LHC guide*, http://cds.cern.ch/record/1165534/files/CERN-Brochure-2009-003-Eng.pdf

[3] L. Evans and Philip Bryant, *LHC Machine*, JINST **3** (2008), S08001

[4] http://home.web.cern.ch/about/accelerators/large-electron-positron-collider

[5] CERN proton synchrotron division, *Project to prepare the PS Complex to be a Pre-injector for the LHC*, http://ps-div.web.cern.ch/ps-div/LHC-PS/LHC-PS.html

[6] S.Myers, *The large hadron collider 2008-2013*, International Journal of Modern Physics A28 (2013), 25, 1330035

[7] LHC Higgs Cross Section Working Group, S. Heinemeier, C. Mariotti, G. Passarino, R. Tanaka (Eds.),*Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, CERN-2013-004 (CERN, Geneva, 2013), DOI: 10.5170/CERN-2013-004

[8] M. Lamont, *Status of the LHC*, Journal of Physics: Conference Series 455 (2013) 012001

[9] http://lhc-machine-outreach.web.cern.ch/lhc-machine-outreach/lhc_in_pictures.htm

[10] The ATLAS Collaboration, *Pile-up Correction in the Missing Transverse Momentum Reconstruction in the ATLAS experiment in Proton-Proton Collisions at* $\sqrt{s}$ *= 8 TeV*, ATLAS-CONF-2014-019 (2014)

[11] ATLAS Collaboration, *ATLAS Twiki: Luminosity Public Results*, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults

[12] http://home.web.cern.ch/about/experiments

[13] ATLAS Collaboration, Detector and physics performance technical design report (1999) Vol. 1

[14] G. Aad et al., *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008), S08003

[15] ATLAS Collaboration, http://www.atlas.ch/photos/full-detector-cgi.html

[16] ATLAS Inner Detector Community, Technical Design Report (1997) Vol.1, CERN/LHCC 97-16

[17] ATLAS Collaboration, *The ATLAS Inner Detector commissioning and calibration*, Eur. Phys. J. **C 70** (2010),787-821 arXiv:1004.5293 [physics.ins-det]

[18] A. Andreazza, *Status of the ATLAS Pixel Detector at the LHC and its performance after three years of operation*, Proceedings of the 6$^{\text{th}}$ International Workshop on Semiconductor Pixel Detectors for Particles and Imaging, PIXEL (2012)

[19] ATLAS Collaboration, *ATLAS pixel detector electronics and sensors*, JINST**3** (2008) P07007

[20] A. Miucci, *The ATLAS Insertable B-Layer project*, Proceedings of the 13; th Topical seminar on innovative particle and radiation detectors, JINST **9** (2014), C02018

[21] ATLAS IBL Collaboration (A. La Rosa for the collaboration), *ATLAS-IBL Pixel Upgrade*, Proceedings of the 12$^{\text{th}}$ Topical Seminar on Innovative Particle and Radiation Detectors, IPRD (2010)

[22] ATLAS Collaboration, *Detector Description*, http://www.atlas.ch/sct.html

[23] A. Abdesselam et al., *The barrel modules of the ATLAS semiconductor tracker*, Nucl. Inst. Meth. **A 568** (2006) 642-671

[24] A. Abdesselam et al., *The ATLAS semiconductor tracker end-cap module*, Nucl. Inst. Meth. **A 575** (2007) 353-389

[25] A. Ahmad et al., *The Silicon microstrip sensors of the ATLAS semiconductor tracker*, Nucl. Inst. Meth. **A 578** (2007) 98-118

[26] F. Campabadal et al., *Design and performance of the ABCD3TA ASIC for readout of silicon strip detectors in the ATLAS semiconductor tracker*, Nucl. Inst. Meth. **A 552** (2005) 292-328

[27] C.W. Fabjan and F. Gianotti, *Calorimetry for Particle Physics*, Rev. Mod. Phys. **75** (2003) 1243-1286, CERN-EP-2003-075

[28] ATLAS collaboration, *Determination of the jet energy scale and resolution at ATLAS using Z/γ-jet events in data at $\sqrt{s}$=8 TeV*, ATLAS-CONF-2015-057 (2015)

[29] ATLAS Collaboration (H. AbouZeid for the collaboration), *Status of the ATLAS Liquid Argon Calorimeter: Performance after 2 years of LHC Operation*, Proceedings of the XVth International Conference on Calorimetry in High Energy Physics, J. Phys. Conf. Ser. **404** (2012) 012004

[30] ATLAS Collaboration (N. Nikiforou for the collaboration), *Performance of the ATLAS Liquid Argon Calorimeter after three years of LHC operation and plans for a future upgrade*, arXiv:1306.6756 [physics.ins-det]

[31] ATLAS collaboration (P. Strizenec for the collaboration), *Performance of the ATLAS Liquid Argon Calorimeter after three years of LHC operation and plans for a future upgrade*, Proceedings of the International Conference on Instrumentation for Colliding Beam Physics, JINST **9** (2014) C09007

[32] ATLAS collaboration (N. Ilic for the collaboration), *Performance of the ATLAS Liquid Argon Calorimeter after three years of LHC operation and plans for a future upgrade*, Proceedings of the 13$^{\text{th}}$ Topical Seminar on Innovative Particle and Radiation Detectors, JINST **9** (2014) C03049

[33] R. Arcidiacono, *Calorimetry in particle physics experiments*, http://personalpages.to.infn.it/∼arcidiac/calo_2.pdf

[34] ATLAS Tile Calorimeter system (Y. Smirnov for the collaboration), *The ATLAS Tile Calorimeter and its upgrades for the high luminosity LHC*, Proceedings of the XXIII International Workshop on Deep-Inelastic Scattering and Related Subjects, DIS (2015)

[35] P. Adragna et al., *Testbeam studies of production modules of the ATLAS tile calorimeter*, Nucl. Instrum. Meth. A**606** (2009) 362-394, ATL-TILECAL-PUB-2009-002, ATL-COM-TILECAL-2009-004

[36] ATLAS Collaboration (L.Heelan for the collaboration), *Performance of the ATLAS Tile Calorimeter*, Proceedings of the International Workshop on Discovery Physics at the LHC, J. Phys. Conf. Ser. **623** (2015) 012014

[37] U. Husemann, K. Reygers and U. Uwer, *Probing the High Energy Frontier at the LHC*, http://www.kip.uni-heidelberg.de/veranstaltungen/vorlesungen/lhc2009/slides/lhc_2009_02.pdf

[38] ATLAS Collaboration, *Alignment of the ATLAS Inner Detector and its Performance in 2012*, ATLAS-CONF-2014-047 (2014)

[39] N. Benekos, *Muon Physics at Run-I and its upgrade plan*, EPJ Web of Conferences **95** (2015) 05001

[40] ATLAS Collaboration (F. Pastore for the collaboration), *The ATLAS Trigger System: Past, Present and Future*, Proceedings of the 37$^{\text{th}}$ International Conference on High Energy Physics, Nucl. Phys. B Proceedings Supplement **00** (2014) 1-7

[41] ATLAS Collaboration (F. Pastore for the collaboration), *The ATLAS Trigger System: Past, Present and Future*, ATL-DAQ-SLIDE-2014-373

[42] ATLAS Collaboration (W. Buttinger for the collaboration), *The ATLAS Level-1 Trigger System*, Proceedings of the International Conference on Computing in High Energy and Nuclear Physics 2012, J. Phys. Conference Series **396** (2012) 012010

[43] C. Gabaldon, *Performance of the ATLAS Trigger System*, Proceedings of the Topical Workshop on Electronics for Particle Physics, JINST **7** (2012) C01092

[44] ATLAS Collaboration (T. Bold for the collaboration), *The core trigger software framework of the ATLAS experiment*, https://cds.cern.ch/record/1609597/files/ATL-DAQ- SLIDE-2013-837.pdf

[45] L. Heinrich, *ATLAS Trigger and Data Acquisition in preparation for Run 2*, ATL-COM-DAQ-2014-013

[46] ATLAS Collaboration (W. Panduro for the collaboration), *ATLAS Data Acquisition: from Run I to Run II*, Proceedings of the International Conference on High Energy Physics (2014)

[47] The ATLAS Trigger System, *The ATLAS Trigger System*, ATLAS Software and Computing Workshop, https://indico.cern.ch/event/352570/session/5/contribution/16/material/ slides/ 0.pdf

[48] CERN Press, *LHC Season 2: Major work at the experiments for Run 2*, http://press.web.cern.ch/backgrounders/lhc-season-2-major-work-experiments-run-2

Chapter 3

Detector Simulation and Physics Modeling with MC Generators

In any analysis study, the measurements from data need to be compared to theoretical predictions for testing purposes and to be able to extract significant conclusions. In searches for new particles e.g. Higgs boson, data events need to be compared to the background ones to look for possible deviations from the expectations of the model tested. Such predictions are provided in particle physics, whenever possible, by Monte Carlo (MC) simulation. For Higgs physics, good MC simulation is crucial for both the discovery of the Higgs boson itself and the precision measurements.

In this chapter, a brief overview of the various steps of event simulation is given first in sec. 3.1. A discussion of the detector simulation with GEANT4 is given next in sec. 3.2, with a comparison of the 'full' and 'fast' detector simulation techniques in sec. 3.3. Then, a very brief description of the various generators used for the MC samples needed by the analyses presented in this thesis is given in sec. 3.4. Finally, parton distribution functions and MC pile-up modeling are briefly presented in sec. 3.5 and 3.6 respectively.

3.1 Event Simulation : Introduction

The basic steps of event simulation in ATLAS are the following [1–3]:

1. **Event Generation:**
   At this step, a simulation of a proton-proton (pp) collision is done and the transition from the hard scattering process to the final physics objects that can be measured in the detector is done through 4 separate steps:

   (a) *Hard Scattering:* This is the event at the partonic level. Calculations at this stage are done based on matrix element (ME) computations for a fixed order in perturbation theory to estimate the probability distribution of a particular 'hard scatter', having the highest momentum transfer in the event. Probability Distribution Functions (PDFs) are used to define the

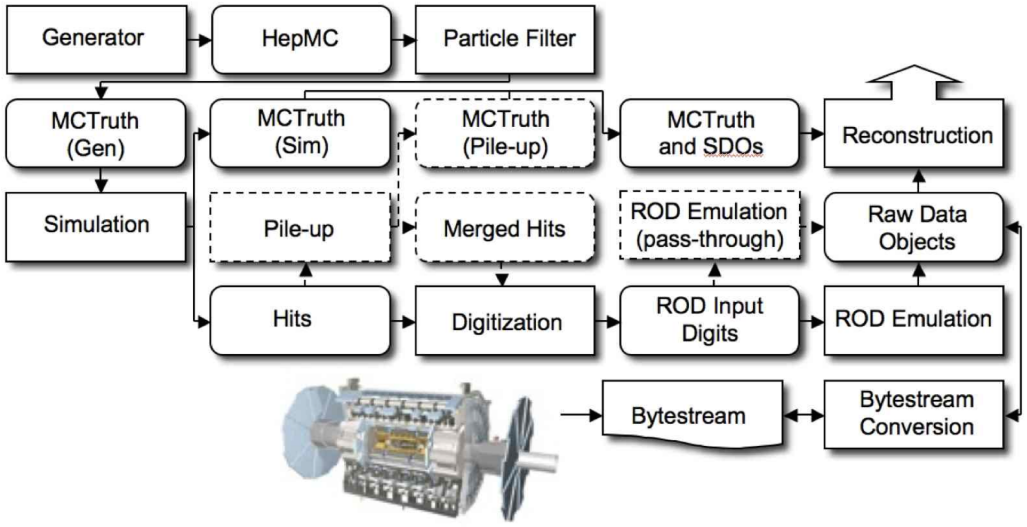


Figure 3-1: ATLAS software simulation flow starting with event generators (top left) till reconstruction (top right) [1]. SDOs and RODs stand for Simulated Data Objects and Read Out Driver electronics during the dgitization step.

incoming partons.
If short-lived 'resonances', e.g. top, $W^{\pm}$ or Z^0, are produced in the hard process, the corresponding decay is considered as part of this process itself, taking into account spin correlations. This step is process-dependent.

(b) *Parton Shower (PS):* This step is needed to go from the hard scattering process with colored partons in the final state to the hadronization scale (low-energy) with colorless hadrons, where the perturbation theory breaks down. After simulating the initial hard scattering, gluon radiations are implemented. The produced gluons can either radiate more gluons or decay into a $q\overline{q}$ (quark,anti-quark) pair. This is 'parton showering' and it can be performed with PYTHIA [4], HERWIG [5] and SHERPA [6] in addition to other generators. Initial-State Radiation (ISR) emitted by the incoming colliding partons is modeled by space-like parton showers, while Final-State Radiation (FSR) associated with the outgoing partons is modeled usinh time-like parton showers.

(c) *Hadronization:* At this stage, colorless hadrons are obtained at the QCD hadronization scale Λ_{QCD} , using hadronization models (e.g. string/cluster models) to take into account the confinement of a partonic system, modeling its non-perturbative dynamics. This is process independent and can be implemented in different ways, e.g. the methods used in PYTHIA, HERWIG. In the detector, the decay products of these hadrons can be observed.

(d) *Underlying event:* Hard or semi-hard interactions can arise from the remnants of the colliding protons and this is referred to as 'underlying event'. In p-p collisions, a colored parton is taken out of each initial, colorless proton during the hard scattering process. Since protons are made of many

partons, additional multiparton interactions (MPI)[1] can occur. Indeed, the probability of secondary interactions arising from the proton remnants is high. Each of these interactions can be associated with ISRs and FSRs.

Many of the produced hadrons are heavy, unstable resonances leading to secondary particle decays. QCD radiations can occur at any stage and the ME to parton shower transition is ensured by parton-jet matching schemes.

The process matrix element and parton distribution functions (PDFs) are needed as input for the event generation. The differences in the cross section resulting from the use of different PDFs is taken into account when calculating the uncertainties on MC samples. The overall event generation is done differently when comparing event generators: SHERPA, HERWIG, PYTHIA can perform the ME+PS for $2 \to 1$ and $2 \to 2$ processes only, while $2 \to n$ events need multi-leg generators such as MADGRAPH5 [7] and Alpgen [8], which can be interfaced with Pythia or Herwig for example for parton showering. Sherpa also has a matrix element generator called COMIX inside. An illustration of the various event generation steps is given in Fig. 3-2 for a $t\bar{t}H$ event.

In ATLAS, the event generation is run within the Athena [12] framework, where ATLAS packages are interfaced with the MC generator packages, which are usually developed and maintained by external authors. Each generated event is associated with an interaction emanating from the vertex defined at the geometric origin. Beam properties are taken into account at this stage, before passing the event to the detector simulation step. The MC generator output, usually in HEP MC format, can be filtered at the generation phase to keep only events of interest, e.g. with particular cuts on the missing transverse energy, etc.

Events are produced using general-purpose MC generators either in a standalone manner or in association with specialized generators for a better description of the final states. A brief overview of MC generators is given in 3.4.

2. **Detector Simulation:**
   The interactions of the generated particles in the detector need to be implemented at this step. In ATLAS, this is done using GEANT4 [10, 11] as described in sec. 3.2.

3. **Digitization:**
   This is basically the translation of the output from the previous step of 'core simulation' to detector responses 'digits'. A digit is produced when the voltage of a readout channel exceeds a pre-configured threshold during a defined time lapse. Each subdetector digitization software takes into account the particular features of the subdetector charge collection such as the dependence of the detector response on readout channels, electronic noise and cross-talk.

[1] MPIs should not be confused with pile-up events corresponding to overlaid events in a bunch crossing.

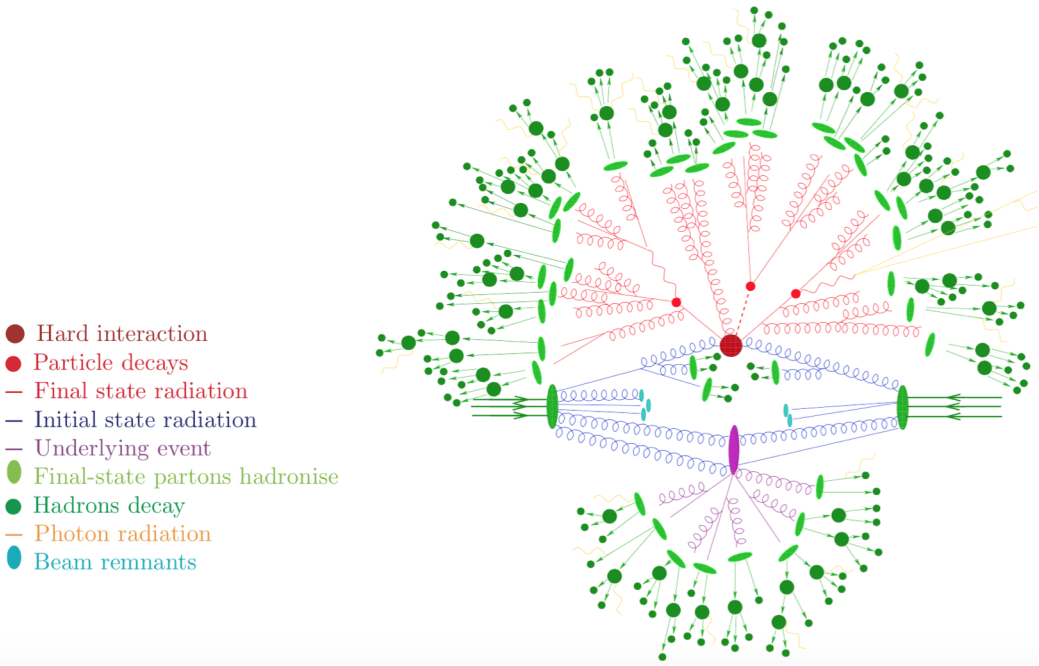


Figure 3-2: A graphical representation of the various event generation steps for a $t\bar{t}H$ event [9].

Digits are written as raw data objects (RDOs). Simulated Data Objects (SDOs) can also be produced, storing information about all the event particles, the detector noise information contributing to the produced signal in a given sensor, etc.

4. **Event reconstruction:**
   After simulating a collision with the appropriate detector response, the reconstruction of physics objects is performed in a similar way to that applied for real data. A detailed description of event and object reconstruction in ATLAS is given in chapter 4.

3.2 Detector Simulation with GEANT4

GEANT4 is a general purpose Monte Carlo tool used to simulate elementary particles passing through and interacting with matter. Simulation is needed in high energy physics to determine the experimental setup (optimize detector parameters, study radiation level, etc.), for calibration purposes in addition to performing comparisons with real data in analysis studies. Inputs that must be provided by the user include amongst many others:

1. Material and geometry of the detector

2. Particle types and their interactions

3. Event generation information (the standard HEPMC format given by event generators is fed to GEANT4)

4. Detector configuration (e.g. misalignments and distortions)

While GEANT4 tools are used to described the infrastructure of particle transportation through the detector, ATLAS specific parts are provided as user codes. The detector geometry is defined in the Geant4 format, and all particle scoring ('sensitive detector' classes) is performed using Athena [12]. The scoring is optimized so that only relevant necessary information to reproduce the detector performance. Additional details can be found in [10, 11].

3.3 Full Sim Vs ATLFast

The detector response in ATLAS can be simulated in two ways [1, 13]:

– Full simulation or 'FullSim': The simulation of the detector response is done using GEANT4 for all detector parts. This method is usually time-consuming.

– Fast simulation or 'ATLFAST-II': Fast simulation is needed to provide the various physics studies with the required simulated statistics, while overcoming the time limitations of full simulation imposed by the complicated detector geometry. 80% of FullSim time is spent on the particle interactions in calorimeters. Simulating electromagnetic particles takes 75% of FullSim time.

  The goal is to speed up the simulation process while allowing to run the standard ATLAS reconstruction. The two components of ATLFAST-II are : the Fast ATLAS Tracking Simulation (Fatras) for the inner detector and muon system simulation, and the Fast calorimeter Simulation (FastCaloSim) package for the calorimeter simulation (fast simulation of electromagnetic and hadronic showers). Pre-simulated showers (stored in memory) are used instead of low energy electromagnetic particles in the calorimeter. Geant4 can be used for the simulation of any subdetector in order to achieve higher accuracy. If the simulation of the inner detector (ID) and muon system response is done with GEANT4, while the calorimeter part is performed using the FastCaloSim package, the simulation time is improved by a factor of 10. With Fatras and FastCaloSim, this factor goes up to 100.

3.4 Monte Carlo Generators

Monte Carlo simulation techniques aim at generating pseudorandom weighted/unweighted numbers using 'hit-and-miss' or 'veto algorithm' methods. The details of event generation were presented in sec. 3.1. A brief basic description is given hereafter of the various generators used for signal and background processes in the different studies presented in this thesis:

- ALPGEN: This is a ME Monte Carlo generator doing calculations with a fixed leg number (number of partons) in the final state. ME calculations give accurate estimations for events with multiple high-momentum jets in the final state. It is most commonly used for W and Z bosons decays generations, and is usually interfaced with Herwig or Pythia for parton showering.

- HERWIG (Hadron Emission Reactions With Interfering Gluons): Its main feature is the angular-ordered parton showering and cluster hadronization model. This general purpose MC generator allows the simulation of lepton-lepton, lepton-hadron, hadron-hadron scattering and soft hadron-hadron collisions. Herwig++ is an improved version of Herwig, allowing BSM (Beyond Standard Model) physics for various models and including spin-correlations and off-shell effects. Additional details can be found in [14].

- PYTHIA: This self-contained general purpose generator is commonly used for parton showering (using string fragmentation), and can be interfaced with subprocesses defined externally, PDF libraries, tau decay libraries, etc. It has a large number of adjustable parameters allowing to cover various physics processes with different configurations. The basic steps in Pythia are 'initialization', 'event generation loop' and 'printing and histogramming'. Details on the latest PYTHIA 8.2 can be found in [15, 16].

- SHERPA (Simulation of High-Energy Reactions of PArticles): It originated in the studies of hard-emission ME matching to PS. The latest Sherpa version includes NLO calculations, incorporates the UFO (Universal FeynRules Output) model format, allows scale uncertainties computations for a single event sample on-the-fly, and hosts a new parton shower [17].

- POWHEG (Positive Weight Hardest Emission Generator): This MC generator performs calculations at NLO and is used for calculations involving charm, bottom and top hadronic production [18]. POWHEG needs to be interfaced with PYTHIA or HERWIG for PS and hadronization. Spin correlations amongst decay products are taken into account. Details on the latest implementations with POWHEG Box can be found in [19].

- MC@NLO: The MC@NLO generator allows matching QCD NLO calculation with the parton shower from MC simulation. The basic idea of MC@NLO is well described in [20] and full documentation can be accessed at [21].

- MADGRAPH5_AMC@NLO: The MadGraph5_aMC@NLO framework provides all the elements necessary for SM and BSM phenomenology, including cross section calculations, hard event generation and multiple tools allowing to perform analyses and tailor events to the user needs. The matrix elements can be obtained at both tree-level (leading order) and at the one loop level (next-to-leading order). This version combines LO and NLO calculations from Madgraph5 and aMC@NLO respectively. It thus supersedes all the 1.5.x MadGraph5 versions and the beta versions of aMC@NLO. It needs to be interfaced

with PYTHIA or HERWIG for PS and hadronization. The reader can refer to [22] for additional information on the general SM and BSM features, LO and NLO calculations, spin correlations with MadSpin, etc.

The choice of MC generator made for each signal/background sample used is given in the same chapter where the analysis (e.g. $H \to \tau^+\tau^-$, track MET) is described.

3.5 Parton Distribution Functions and Tuning Parameters

The substructure of the colliding protons is described using parton distribution functions (PDFs), given as input to MC generators. In ATLAS, CTEQ PDFs are used by default in addition to the wide PDF selection available through Les Houches Accord PDF Interface (LHAPDF) library. Since PDFs and ISR tuning parameters in the event generator are correlated, whenever the PDF set is changed, the tuning parameters are adjusted accordingly to avoid getting inconsistent results.

Tuning parameters are needed during event generation if the default parameters set by generator authors are not suitable for the LHC conditions. The tunings are estimated based on comparisons with data and can be done during run time. Additional details can be found in [1].

3.6 Monte Carlo Pile-up Modeling

At high luminosity values, overlaid events ('pile-up') are unavoidable in the proton-proton collisions. The pile-up seen in data is simulated in MC generated events using the hits read from various event types configured at run time, including minimum bias, cavern background, beam gas, and beam halo events. The number of events to be overlaid depends on the luminosity and follows a Poisson distribution with a long tail and a mean value $\langle\mu\rangle$. The readout is triggered by the hard scattering event, but the subdetectors are sensitive to hits in bunch crossings before and after the one containing this event. Detector and electronic effects are taken into account before pile-up merging.

Monte Carlo pile-up simulations are done using the suitable number of simulated minimum bias events along with the high energy signal event. The separation between bunch crossings Δt and the average number of interactions per crossing $\langle\mu\rangle$ are used to define the pile-up conditions. Pile-up interactions are simulated using PYTHIA. High energy signal events and minimum bias ones have separate treatments. Additional details can be found in [1].

3.7 Conclusion

The basic steps for event simulation in ATLAS , namely event generation, detector simulation, digitization and event reconstruction, were presented. In addition, a brief description of the various MC generators used in the different studies presented in this thesis, parton distribution functions and MC pile-up modeling was given as well.

Bibliography

[1] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C**70** (2010) 823-874, arXiv:1005.4568

[2] M. Seymour and M. Marx, *Monte Carlo Event Generators*, (2013) arXiv:1304.6677 [hep-ph], MCNET-13-05

[3] T. Sjöstrand, *Monte Carlo Tools*, (2009) arXiv:0911.5286 [hep-ph], LU-TP-09-31, MCNET-09-17

[4] T. Sjöstrand, S. Mrenna, and P. Skands, *Pythia 6.4 Physics and Manual*, JHEP 05 (2006) **026**

[5] G. Corcella et al., *HERWIG 6: an event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)*, JHEP 01 (2001) **010**

[6] T. Gleisberg et al., *Event generation with SHERPA 1.1,*, JHEP 0902 (2009) **007**

[7] J. Alwall et al., *MadGraph/MadEvent v4: the new web generation,*, JHEP 09 (2007) **028**

[8] M. Mangano et al., *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, JHEP 0307 (2003) **1**

[9] T. Gleisberg et al., Event generation with SHERPA 1.1, JHEP 0902 (2009) **007**, arXiv:0811.4622 [hep-ph]

[10] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823

[11] S. Agostinelli et al., *Geant4: a simulation toolkit*, Nucl. Instr. Meth. A **506** no. 3, (2003) 250

[12] ATLAS Collaboration, *ATLAS Computing Technical Design Report*, ATLAS-TDR-017, CERN- LHCC-2005-022 (2005), See also http://atlas-computing.web.cern.ch/atlas-computing/packages/athenaCore/athenaCore.php

[13] T. Yamanaka and the ATLAS Collaboration, *The ATLAS calorimeter simulation FastCaloSim,* , Journal of Physics: Conference Series 331 (2011) 032053

[14] M. Bahr et al., *Herwig++ Physics and Manual*, Eur.Phys.J. C**58** (2008) 639-707, arXiv:0803.0883 [hep-ph]

[15] T. Sjöstrand et al. *An Introduction to PYTHIA 8.2*, Comput.Phys.Commun. **191** (2015) 159-177, arXiv:1410.3012 [hep-ph], LU-TP-14-36, MCNET-14-22

[16] T. Sjöstrand, http://home.thep.lu.se/ torbjorn/Pythia.html

[17] HepForge project, *Sherpa wiki*, https://sherpa.hepforge.org/trac/wiki/SherpaDownloads/Sherpa-2.2.0

[18] S. Frixione et al., *A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, JHEP 0709 (2007) **126**, arXiv:0707.3088 [hep-ph]

[19] Powheg Box Project, http://powhegbox.mib.infn.it/

[20] S. Frixione et al., *Matching NLO QCD computations and parton shower simulations*, JHEP 0206 (2002) **029**, hep-ph/0204244

[21] B. Webber, http://www.hep.phy.cam.ac.uk/theory/webber/MCatNLO/

[22] 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*, JHEP 1407 (2014) **079**, arXiv:1405.0301 [hep-ph]

Chapter 4

Event and Object Reconstruction

To obtain standard physics objects used in analyses (electrons, muons, jets...) from the signals read out from the ATLAS detector, software algorithms are used and run on both data and Monte Carlo (MC) simulation samples. This process is called 'reconstruction'. Differences between data and MC samples are taken into account and necessary corrections are applied to MC samples. The reconstruction and identification of objects used in the analysis presented in this thesis are described in this chapter. Track and vertex reconstruction in addition to calorimeter clustering are discussed first. Subsequent sections describe the reconstruction efficiency of physics objects used in analyses such as electrons, muons, jets, taus, and missing transverse energy respectively.

4.1 Inner Detector Track Reconstruction

Tracks are reconstructed within the full inner detector (ID) acceptance($|\eta| < 2.5$) using a sequence of algorithms. Once the information from the pixel, SCT (Semi Conductor Tracker) and TRT (Transition Radiation Tracker) detectors is treated and interpreted to generate space-points from hits and clusters, these are used iteratively to build tracks using mainly two algorithms: the 'inside-out' and the 'outside-in' tracking.
The track parametrization is presented first. Then, the various track reconstruction steps are discussed in detail. The performance of the track reconstruction algorithms in ATLAS with data/MC comparisons is presented at the end of this section.

4.1.1 Track Parametrization

A reconstructed track in ATLAS is parametrized at the point of closest approach to the global z-axis using the perigee representation (d_0, z_0, ϕ_0, θ, q/p) where:

- q/p: charge to momentum ratio of the particle.

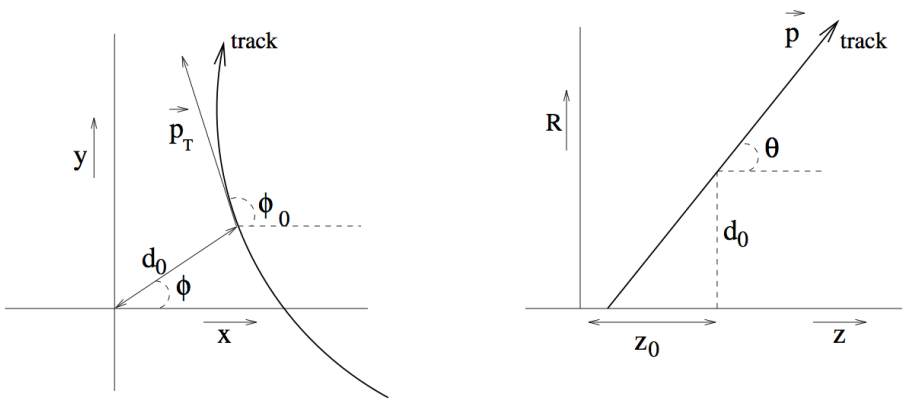


Figure 4-1: Track perigee parameters illustrated in the transverse x-y plane (left) and r-z plane (right) as defined in ATLAS.

- ϕ_0 : the azimuthal angle with the x-axis at the perigee point in the transverse plane i.e. $x - y$ plane.
- θ_0 : the angle with the z-axis in the $r - z$ plane.
- d_0 : transverse impact parameter defined as the signed distance to the z-axis. Its sign is opposite to the sign of the angular momentum of the track about the z axis, i.e. d_0 is positive if $\phi - \phi_0 = \frac{\pi}{2}[2\pi]$ where ϕ is the angle to the perigee point in the $x - y$ plane (Fig. 4-1).
- z_0 : longitudinal impact parameter at the perigee. When defined with respect to a reconstructed vertex, this is the longitudinal distance of closest approach and cuts on $|z_0|$ are cuts on $|z_{\text{at perigee}} - z_{\text{vertex}}|$.

Reconstructed track impact parameters, d_0 and z_0, are also defined with respect to the reconstructed vertex, usually the primary vertex in the event, to be used in analyses later on [4, 9].

4.1.2 Track Reconstruction Steps

First, the 'inside-out' algorithm is applied. Then, the 'outside-in' algorithm is applied, followed by a second inside-out round with less strict pattern recognition requirements as explained hereafter.

A- The 'Inside-out' Algorithm

The 'inside-out' algorithm starts with the silicon space points close to the interaction point in order to build track segments and extrapolating them to the TRT. The major steps in this standard pattern recognition strategy can be summarized as follows:

1. **Track Seed Finding in the Pixel Detector and SCT :** First, silicon (pixel detector and SCT) clusters are deduced from raw hits. The goal is to create

three-dimensional representations, called space-points, of the ID measurements. Pixel detector clusters translate directly into space-points since they give a local two-dimensional representation on a fixed surface. On the other hand, a SCT strip gives a precise measurement in one direction, which requires combining the information from pairs of SCT clusters to form a space-point. Track seeds, i.e. initial trajectories, are then formed from sets of three space-points coming from a different silicon layer each. The requirement on the number of space points is chosen such that the number of possible combinations is maximum while allowing for a rough first estimation of the transverse momentum. The 3 space-points can have four possible configurations: all points originating from the pixel detector or the SCT, two points in the pixel detector and one in the SCT, and one point from the pixel detector with 2 points coming from the SCT. The inside-out algorithm uses the first three categories and requires a minimum distance between the space-points to suppress candidates with multiple hits from the same layer. A preselection step is applied by setting minimum transverse momentum or maximum impact parameter requirements [8].

2. **Estimation of Track Seed Parameters:** At this stage, an early estimation of the track perigee parameters is made assuming a perfect helical model. The circular trajectory given by the track projection in the transverse plane is described by three parameters: the transverse momentum p_{T}, the transverse impact parameter d_0 (shown in Fig. 4-2) and the azimuthal angle ϕ_0. Using the radius, ρ, of the trajectory and assuming a homogeneous magnetic field B parallel to the z-axis, the transverse momentum p_{T} is given by [8]:

$$\rho[\mathrm{mm}] = \frac{p_T[\mathrm{GeV}]}{3.10^{-4} \times q[e] \times B[\mathrm{T}]} \ . \tag{4.1}$$

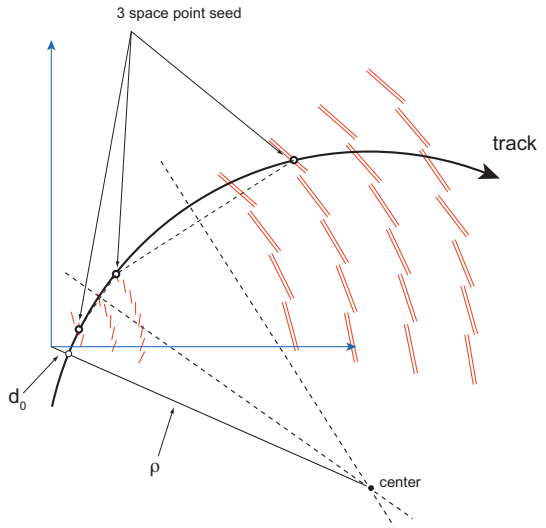


Figure 4-2: Sketch illustrating track seed parameters [8].

d_0 is the distance of closest approach of the track in the transverse plane with respect to a reference point. Using the (x, y) coordinates of the circle center, (c_x, c_y), d_0 is calculated as :

$$d_0 = \sqrt{c_x^2 + c_Y^2} - \rho \ . \tag{4.2}$$

The reference point can be the origin or a fast reconstructed vertex. The details of vertex reconstruction are given in sec. 4.2.

3. **Select Good Track Candidates, Full Track Fit :** A combinatorial Kalman Filter (KF) is applied to search for further hits in the silicon detectors. The trajectory is extrapolated from one layer to the next one, accounting for multiple scattering and energy loss. On the new layer, new trajectories are constructed with updated parameters (and errors). The number of trajectories to grow is limited according to their χ^2 and the number of missing hits. If a measurement point degrades significantly the χ^2 of the fit, it is removed [39]. Each seed can be associated with a single track candidate at most. This is referred to as 'seed survival' and it occurs only if the seed satisfies quality requirements on track seed parameters and passes cuts on the number of holes, hits and additional parameters. Here are some basic definitions :

   - A hit is a measurement point assigned to a track.
   - A hole, on the other hand, is associated with an expected non-existing measurement point for a given track trajectory. It can be due to material interaction, silicon inefficiency or problems with the pattern recognition. Inactive modules are excluded from the hole definition.
   - An outlier is a measurement point that degrades the fit quality.
   - A shared hit is a hit associated with 2 tracks.
   - A hole detected on both sides of the SCT module is called a double hole.

   Good tracks are selected by applying a set of good quality cuts on the number of outliers, good hits and holes. Table 4.1 gives the track selection used at the start of the LHC running with low pile-up environment. Later, the conditions needed to be more robust to minimize the number of fake tracks, requiring 9 hits or more in the silicon detectors (SCT+pixel) and no hole in the pixel detector. Tracks reconstructed by the 'inside-out' algorithm are required to have a transverse momentum $p_{\mathrm{T}} > 400$ MeV. A Kalman fitter-smoother approach, which is intrinsic to the KF, is used to fit the trajectories and add successive hits for the track fit candidate. This fitting approach is the default one used along with the global χ^2 minimization. Modified KF versions can be applied in specific cases e.g. the Gaussian Sum Filter (GSF) used to fit electron tracks [39].

4. **Resolve Ambiguities :** Some tracks candidates, at this level, share hits or are incomplete. In addition, some track candidates can be 'fake tracks' i.e. tracks

Cut	Value
Minimum p_{T}	500 MeV
Maximum z	250 mm
Maximum η	2.7
Maximum d_0	10 mm
$N_{\text{SCT hits}}$	≥ 6
$N^{(*)}_{\text{pixel hits}}$	No cut
$N_{\text{Shared Hits}}$	< 4
$N_{\text{Pixel Holes}}$	< 3
$N_{\text{SCT Holes}}$	< 3
$N_{\text{Double Holes}}$	< 2

Table 4.1: Table summarizing the values of tracking cuts of the inside-out algorithm. (*) For full track reconstruction, selected tracks at the final stage are good tracks with at least 1 pixel hit [9, 12].

where most of the associated measurements do not come from the same particle. The likelihood of a track to describe the trajectory of a particle from the underlying physics event is an essential criteria used to rank tracks and discriminate real tracks from 'fake' ones [39]. Therefore, before extrapolating into the TRT, ambiguities resulting from hits associated to more than one track are treated. A track refit using the refined reconstruction geometry, giving the parameter χ^2/(number of degrees of freedom) is not sufficient. Consequently, a *track score* is associated to each track. The *track score* combines the fit quality with additional morphologic track parameters : a beneficial or penalty score is associated with each track characteristic and the combined result forms the track score. A qualitative example of such characteristics is shown in Table 4.2. Generally, associating hits to tracks improves the track score value: the main goal is to obtain fully reconstructed tracks instead of small segments. Based on the *track score*, tracks with holes are disfavored. Tracks with hits overlapping on different detectors or on successive layers of the same detector type are favored. In the case of hits belonging to multiple tracks, the hits are assigned to the track with the highest track score. A track refit is applied whenever hits are removed [39]. In addition, the track score contains a term dependent on the logarithm of the transverse momentum. Hence, in the case of two track candidates having similar hit composition, the one with the highest transverse momentum is favored.

Rejection Stages:

(a) **Initial Stage:** Before evaluating the track score, tracks can be rejected by the ambiguity solver if they fail to pass certain predefined requirements such as a minimum transverse momentum cut or a cut on the number

of hits. At this level, rejected tracks, such as tracks with silicon (pixel detector +SCT) holes, are attributed a score of zero.

(b) **Scoring Stage:** A track fails the ambiguity solving stage if the track score is relatively too low compared to other tracks entering the ambiguity solver, or if it was combined with other candidates to favor complete track segments over incomplete ones or random hits combinations.

5. **Extend Resolved Tracks into TRT :** Tracks failing the set of good quality cuts applied in the previous steps are discarded and not considered any further when adding hits from the TRT. If the fit based on silicon measurements only is better than the one with TRT hits, contributions from the TRT are considered outliers and neglected for that track. The track score is reduced.

Track characteristics	Detector	Effect on the track score
B layer hole	pixel	strong penalty
Layer hole	pixel	penalty
Overlap hit	pixel, SCT	strong benefit
Sensor hole	SCT	weak penalty
Layer hole (module)	SCT	strong penalty

Table 4.2: Table showing qualitatively the effect of some track characteristics on the track score [39].

6. **Refit of extensions and replace original if better.**

The 'inside-out' algorithm is used to reconstruct the primary particle tracks, which are typically used in physics analyses. Primary particles are ones with lifetime greater than 3×10^{-11} s originating directly from proton-proton (pp) interactions, or from subsequent decays or interactions of shorter-lived particles (with lifetime $< 3\times10^{-11}$ s [19].

B- The 'Outside-in' Algorithm

The next stage includes applying the 'back-tracking', i.e. the 'outside-in' tracking algorithm. This is used to reconstruct particles resulting from the interactions or decays of primaries i.e. to reconstruct secondaries. Secondaries may also arise from decays of other secondaries. As an important complement to the 'inside-out' algorithm, the 'back-tracking' algorithm allows track reconstruction of long-lived particles which may not leave any hits in the Pixel Detector and of positrons that leave holes in the silicon detectors if converted to photons. In such cases, there is no seed for the 'inside-out' algorithm even though an actual particle was present. Additionally, in case of bremsstrahlung and electrons radiating energetic photons, the TRT hits and the silicon detectors hits cannot be added together for 'inside-out' track reconstruction. To handle such cases, seeds are created from the TRT hits. A KF uses the seeds

to search for tracks working backwards towards the silicon detector, adding hits to the track.

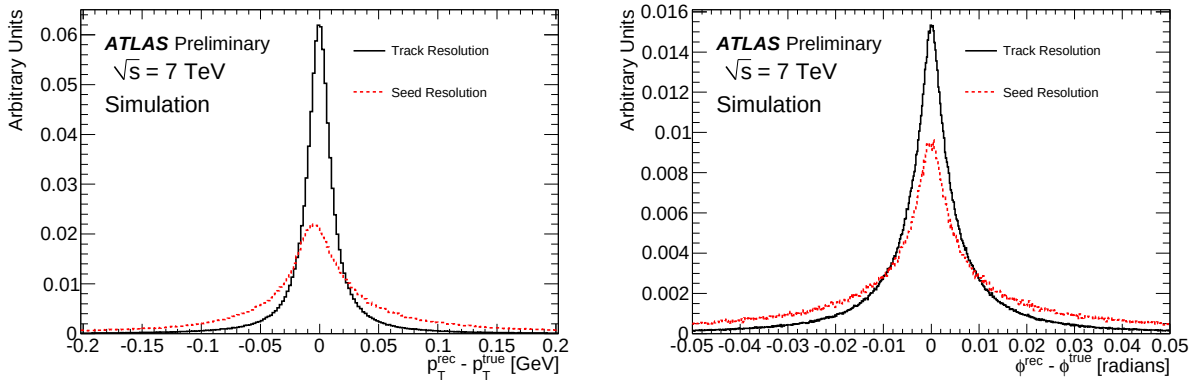
C- Second 'Inside-out' Round

The next step is a second inside-out round with less strict pattern recognition requirements. This allows low-p_{T} tracking: tracks with p_{T} as low as 150 MeV can be recovered. Tracks issued from a TRT segment without extensions to the silicon detectors are called TRT-standalone tracks.

D- Robust Requirements and Fake Tracks

Finally, a set of robust requirements is applied to reduce the amount of fake tracks [19]. To select samples with mostly primary particles, $|d_0|_{\mathrm{PV}} < 1.5$ mm and $|z_0 \sin(\theta)|_{\mathrm{PV}} < 1.5$ mm cuts on the track transverse and longitudinal impact parameters respectively, defined with respect to the reconstructed vertex (usually the primary vertex), are applied in analyses [9, 12].

4.1.3 Performance

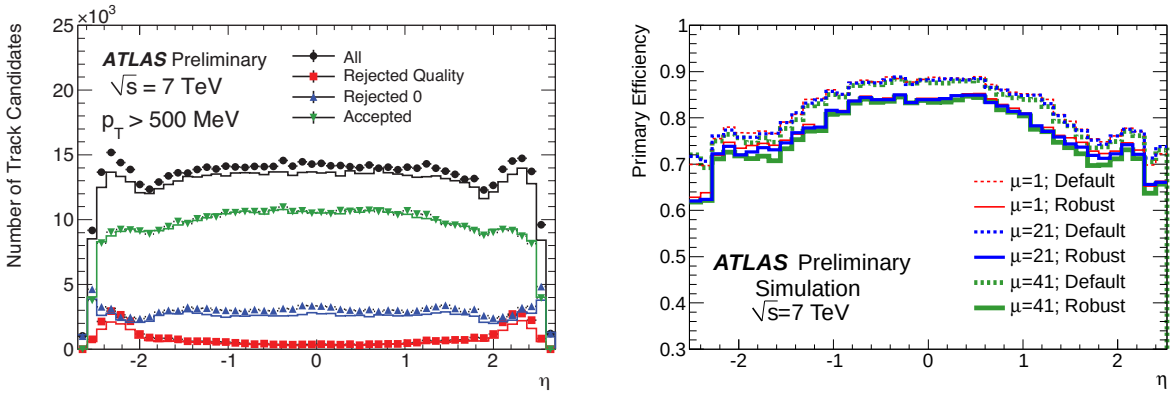


(a) Transverse momentum resolution

(b) Azimuthal resolution

Figure 4-3: Normalized distributions showing track parameter resolution for track seeds matched to the generated particles compared to those of the final tracks (final fit) in minimum bias MC simulation samples [8].

Performance plots showing track parameter resolution and track reconstruction efficiency are shown in Figs. 4-3 and 4-4. Compared to the final tracks, track seeds momentum resolution is worse by a factor of three and the average momentum value is a bit shifted since the track seeds do not include energy loss corrections. However, the



(a) (b)

Figure 4-4: Plot (a) shows the data and MC track candidates at different stages of the ambiguity solver for primary particles [8]. Plot (b) shows the primary track reconstruction efficiency as a function of the pseudorapidity η in minimum bias MC samples for two selections: default and robust requirements. Distributions are done for different pile-up conditions: no pile-up, average number of interactions $\langle\mu\rangle = 21$, 41 [19].

azimuthal resolution of tracks seeds is very close to that of the final tracks. The data to MC comparison of track candidates at different stages of the ambiguity solver is shown in Fig. 4-4-(a) as a function of the pseudorapidity η: the histograms showing the number of all track candidates entering the ambiguity solver before any rejection are drawn in black, the blue histograms show the ones rejected because they were assigned a zero track score, the red histograms correspond to the ones rejected based on quality cuts. And the resolved tracks are shown in green under the label 'accepted'. A fair data/MC agreement is observed for these different stages. Figure 4-4-(b) shows the track reconstruction efficiency for primary particles, i.e. fraction of primary particles with $p_{\mathrm{T}} > 400$ MeV and $|\eta| < 2.5$ matched to a reconstructed track, for different pile-up conditions in minimum bias MC simulation samples: samples with no pile-up($\langle\mu\rangle = 1$) and samples with significant pile-up for both the default requirements (at least 7 silicon (pixel+SCT) hits and at most 2 holes in the pixel detector for 2011 samples) and robust requirements (at least 9 silicon hits and exactly zero holes in the pixel detector for 2011 samples) cases. As pile-up increases, the efficiency decreases by about 1% for the default scenario and by 5% for the robust requirements [8, 19]. The number of tracks reconstructed per interaction using the default requirements increases with pile-up as shown in Fig. 4-5(a) and the fraction of fake tracks increases. The robust requirements give a stable number of reconstructed tracks per interaction regardless of pile-up [19] as shown by the linear dependence in Fig. 4-5(a). The distributions of the track transverse impact parameter d_0 in Fig. 4-5(b) for tracks passing the robust requirements in data samples are in good agreement for different pile-up conditions.

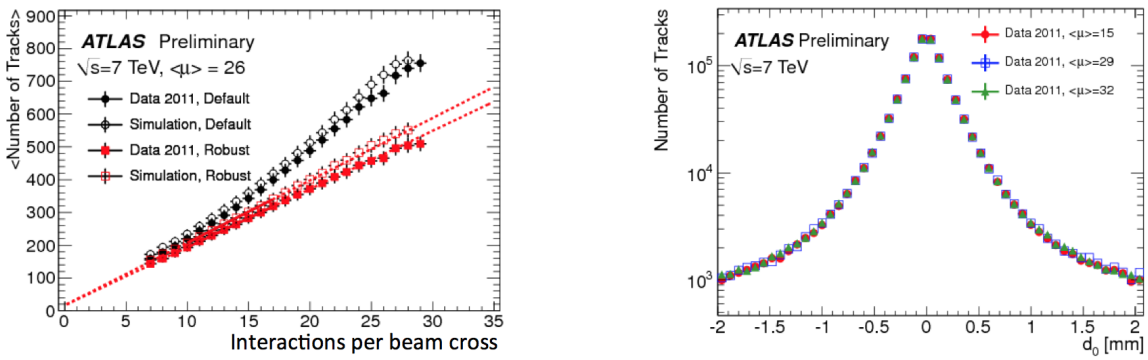


Figure 4-5: 2011 Data (closed markers) to MC samples (open markers) comparison of the average number of tracks as a function of the number of interactions per beam crossing for both default (black) and robust (red) requirements is shown in (a). The dashed lines in (a) show the linear fit of the track multiplicity for tracks meeting the robust requirements. Plot (b) shows the transverse impact parameter d_0 distributions for tracks meeting the robust requirements in data for various pile-up conditions [19].

4.2 Vertex Reconstruction

Primary vertices, as illustrated in Fig. 4-6, refer to the p-p interaction points while secondary vertices are displaced away from primary ones. Secondary vertices are associated with displaced tracks, with a non-negligible impact parameter significance with respect to the primary vertex, as seen for e.g. b-hadrons, K_s^0 mesons, Λ baryon and photon conversions. The primary vertices need to be reconstructed before secondary ones.

In order to reconstruct vertices in ATLAS, an iterative procedure, namely ***adaptive vertex finding*** [47], is used first to associate reconstructed tracks with vertex candidates. Then, performing a vertex fit allows to extract the vertex parameters. Finally, the tracks associated with the chosen vertex are refit constraining them to come from the reconstructed interaction point i.e. the primary vertex (PV). The vertex finding algorithm steps can be summarized as follows [6]:

1. Preselect reconstructed tracks with $p_{\mathrm{T}} > 400$ MeV that are consistent with the interaction region based on d_0 and z_0 such that $d_0 < 4$ mm, $\sigma(d_0) < 5$ mm, $\sigma(z_0) < 10$ mm and the track has at least 4 associated hits in the SCT detector and 6 silicon (pixel detector +SCT) hits.

2. The global maximum of the z_0 distribution for all tracks is selected as the primary vertex seed and $x = y = 0$.

3. The 'adaptive vertex finding algorithm' is applied to identify the vertex position.

4. If the track agreement with the vertex position is within 7 σ, it is kept. Otherwise, it is used to seed a new vertex. A χ^2 with 2 degrees of freedom is used

to describe the track-vertex compatibility, which will be reflected by a weight associated to the track [19]. A loose cut on χ^2 ($\chi^2 > 49$) is applied to reduce the number of single vertices splitting into two due to outlying track measurements.

This procedure is repeated iteratively until no new vertices are found and the beam-spot is used as a constraint. The vertex with the highest Σp_T^2, where the sum runs over the tracks associated to the vertex, is chosen as the primary vertex (PV). In analyses, events are required to have a PV with at least 3 tracks. The secondary vertices and converted photon vertices are reconstructed by dedicated algorithms.

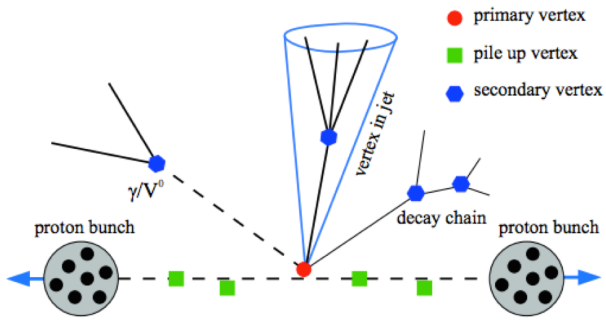
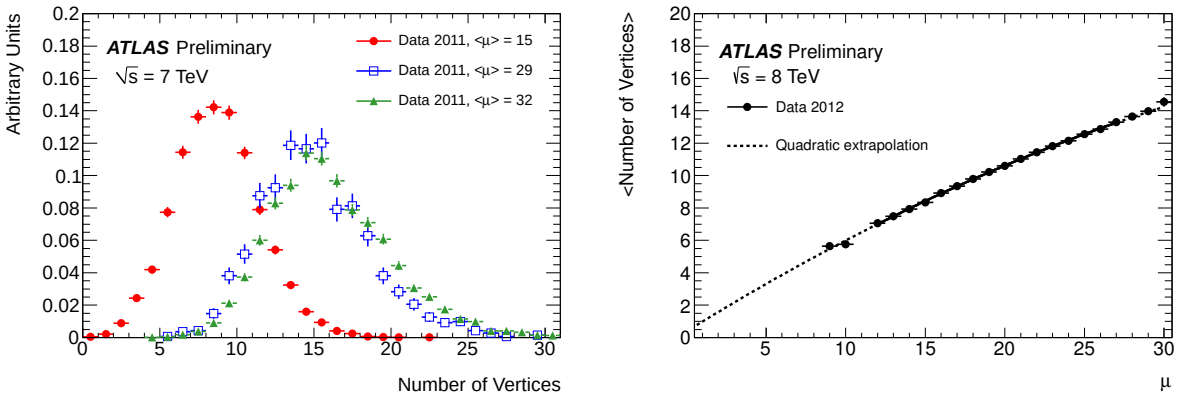


Figure 4-6: Plot showing different vertex definitions.



(a) (b)

Figure 4-7: Distribution of the reconstructed number of vertices in various pile-up conditions (a) and the correlation with the average number of interactions per crossing [19, 20].

Figure 4-7 shows the dependence of the vertex reconstruction on pile-up. As pile-up increases, the fraction of fake tracks increases and the impact of the 'shadowing effect', i.e. when neighboring vertices are merged because the interactions are not far enough to be resolved, is greater. This affects the precision on the vertex position and

the vertex reconstruction efficiency is reduced. Hence, the number of reconstructed vertices depends indeed on the pile-up conditions [19, 20]. The distribution of the number of vertices in data for different pile-up conditions is shown in Fig. 4-7-(a). And the quadratic correlation between the average number of vertices and the average number of interactions per bunch crossing is illustrated in Fig. 4-7-(b) for 2012 data at $\sqrt{s} = 8$ TeV [20]. For 2011 events, the dependence was linear up to $\langle\mu\rangle = 15$ [19]. For 2012 data at $\sqrt{s} = 8$ TeV, a disagreement is seen between data and default MC simulation in the minimum bias vertex multiplicity due to differences in the visible cross sections [14]. This can be corrected for by applying a μ-rescaling of 1.11 ± 0.03 to the MC samples. In addition, the track selection requirements also have an impact on the vertex reconstruction efficiency. In fact, using the track default requirements (shown in section 4.1), the fraction of fake vertices, i.e. vertices whose weight mainly comes from fake tracks, goes up to 7% at $\langle\mu\rangle = 41$. However, applying the robust track requirements leads to an enhanced reconstruction efficiency at high pile-up and a decrease by 5% at low pile-up conditions. The vertex reconstruction efficiency is about 80% for single interactions including minimum bias events with low particle multiplicity and low-p_{T} particles. It goes up to 90% once the requirement of at least two charged primary particles is imposed [19].

4.3 Calorimeter Clustering

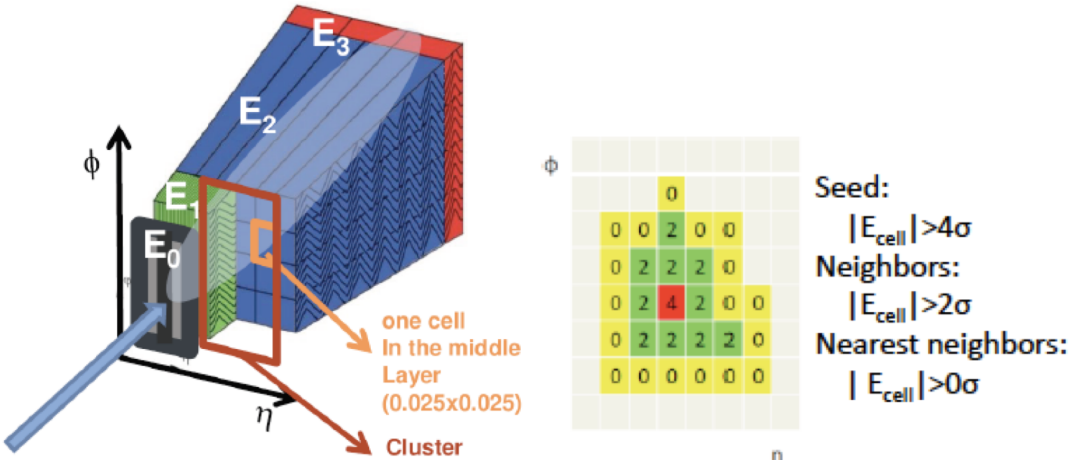
An important information for particle identification comes from calorimeter clustering algorithms. In fact, a cluster is the main reconstruction object for calorimetry and is defined as a group of cells (or even fraction of cells) formed around a seed cell with either a fixed size in pseudorapidity-azimuthal angle $\Delta\eta \times \Delta\phi$ (sliding window), or variable borders based on the significance of the cells (topological clusters). ATLAS uses both the "sliding window" (used for e, γ identification) and the "topological" (used for jets and missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction) algorithms.

These clustering algorithms group calorimeter cells and estimate the energy deposited in the clusters (see Fig. 4-8), which needs to be calibrated afterwards [42].

4.3.1 Sliding Window Algorithm

The basic steps of the sliding window algorithm are tower building, seed (precluster) finding and cluster filling described hereafter.

1. **Tower Building:**
   The calorimeter cells are projected in the pseudorapidity azimuthal angle $(\eta-\phi)$ space into a fixed grid of $N_\phi \times N_\eta$ elements of size $\Delta\eta \times \Delta\phi$ each. Table 4.3 shows the parametrization used for tower building in the EM calorimeter (barrel (EMB) and end cap (EMC)). The goal is to obtain $N_\phi \times N_\eta$ towers, called 'calo towers', used to find cluster seeds where the tower energy is the sum over all the cells in all the calorimeter layers, falling within $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$. The tower bin size $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$ is different from the granularity of



(a) (b)

Figure 4-8: (a) shows a qualitative fixed-size cluster representation for a single layer in the EM calorimeter. The sliding window algorithm combines such clusters layer by layer. An example of a topo-cluster is shown in (b).

Tower Type	EM
Calorimeters	EMB,EMC
$\lvert\eta_{\max}\rvert$	2.5
$N_\eta(\Delta\eta)$	200(0.025)
$N_\phi(\Delta\phi)$	256(0.025)

Table 4.3: EM Tower parametrization in the EM calorimeter barrel (EMB) and end-cap (EMC) [42].

Layer	$\Delta\eta \times \Delta\phi$
1 or 'Strips'	0.003×0.1
2 or 'Middle'	0.025×0.025
3 or 'Back'	0.050×0.025

Table 4.4: $\eta - \phi$ granularity of the three EM calorimeter layers [4, 42].

all EM calorimeter layers except the middle layer as Table 4.4 shows. Hence, if a calorimeter cell is shared among several towers, its contribution to each tower energy is weighted by the fraction of the cell area overlapping with the tower in question [4, 42].

2. **Seed (Precluster) Finding:**
The window size in all the above precluster finding steps is optimized for maximum efficiency with minimum noise.

   - A candidate fixed size rectangular window of dimension $N_\eta^{\text{window}} \times N_\phi^{\text{window}} = 5 \times 5$ (in units of tower bin size $\Delta\eta \times \Delta\phi$) is used to scan the grid towers, looking for a 'local' maximum. The window is shifted by one cell either up/down($\Delta\phi$ step), or left/right ($\Delta\eta$ step).
   - The window position is chosen such that the contained transverse energy, i.e. the transverse energy of the towers within the window, is a local maximum. A precluster, or seed, is obtained if the window transverse energy is greater than the 3 GeV threshold.

- A smaller window size, $N_\eta^{\text{pos}} \times N_\phi^{\text{pos}} = 3 \times 3$ (in tower bin units), is used for precluster position calculation. Duplicate preclusters are defined as ones with positions contained within a window of size $= 3 \times 3$ in tower bin units. In this case, the cluster with the largest transverse energy is selected and the others are removed.

3. **Cluster Building:**

   - To build EM clusters, a window of size $N_\eta^{\text{cluster}} \times N_\phi^{\text{cluster}}$, as defined in Table 4.5, with the seed position as its center, is used to select the corresponding cells.
   - This process starts with the seed position in the middle layer, where 80% of the energy of the EM shower is deposited. After that, the "strips" (layer 1), the presampler (PS) and the "back" (layer 3) are processed respectively.
   - The cluster size, $N_\eta^{\text{cluster}} \times N_\phi^{\text{cluster}}$, depends on the particle type generating the shower and on both the calorimeter part (barrel/end-cap) and layer treated. Table 4.5 shows the cluster size definition used for each calorimeter layer. The definitions for electrons and photons in the EM barrel and end-cap are given in Table 4.6. N_ϕ^{cluster} for electrons is larger than the one for photons because the electron shower in the barrel is wider than the photon shower. Unlike photons, electrons are subject to deviations in the ϕ direction in the presence of the magnetic field. Moreover, converted photons give rise to electron-positron pairs, which explains the cluster size difference between converted and unconverted photons in the barrel. N_η^{cluster} is bigger in the end-cap than in the barrel since, for a given $\Delta\eta$, the corresponding physical size is smaller in the endcap than in the barrel.
   - Energy deposits outside the cluster and in dead material are taken into account when calibrating the cluster energy. This calibration is also dependent on the particle type. In fact, the cluster filling is done as part of the $e-\gamma$ (electron-photon) identification algorithm. Knowing the particle type hypothesis before calibration allows for a single calibration step i.e. there is no need to calibrate an EM cluster as both an electron and a photon candidate [42].

4.3.1.1 Duplicate Clusters and Energy Sharing

The cluster efficiency evaluation is done using single particle simulation samples. In this case, duplicate clusters are encountered when more than one cluster is associated to the single particle event. They are mainly produced by real physics processes such as bremsstrahlung and photon conversions, where the secondary emitted particle leads to a second cluster. The two clusters are treated as duplicates if the distance separating the original particle cluster and the secondary particle one $\Delta R = \sqrt{\Delta\eta^2 + \Delta\Phi^2} < 0.3$. In addition, clusters obtained at the end of the sliding-window algorithm may overlap and share cells if they are very close. This is handled

Order	Layer	$\Delta\eta_{\rm cl}$(units of 0.025)	$\Delta\phi_{\rm cl}$(units of 0.025)
1	Middle	$N_\eta^{\rm cluster}$	$N_\phi^{\rm cluster}$
2	Strips	$N_\eta^{\rm cluster}$	6 or 8
3	PS	$N_\eta^{\rm cluster}$	6 or 8
4	Back	$N_\eta^{\rm cluster}+1$	$N_\phi^{\rm cluster}$

Table 4.5: Cluster size definition for each EM calorimeter layer as used in the sliding window algorithm. The layers are processed to build EM clusters in the order shown in the first column. The dimensions are expressed in tower bin units, i.e. in units of 0.025. $N_\eta^{\rm cluster}$ and $N_\phi^{\rm cluster}$ values depend on the particle type hypothesis and on the region of the EM calorimeter. Their definitions for electrons and photons in the barrel and endcap are given in Table 4.6 [42].

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

Table 4.6: Cluster size definition $N_\eta^{\rm cluster}\times N_\phi^{\rm cluster}$ for electrons and photons (converted and unconverted) in the barrel and end-caps of the ATLAS EM calorimeter for the sliding window algorithm [42].

by the "energy sharing algorithm" in order to avoid cell double counting and to have correct cluster energy estimation : the energy of a cell 'i' shared among N clusters is distributed according to the energy fraction of each cluster $w_{i,j}=\frac{E_j}{\Sigma_{k=1}^{N}E_k}$ [42].

4.3.2 Topological Algorithm

The main steps of the topological clustering are :

1. *Cluster Making :*

   - In order to find the seeds, the calorimeter cells are scanned to identify the ones having a high energy significance, i.e. a signal to noise ratio $S/N \geq 4$: $|E_{\rm seed}| > 4$ (σ(electronics noise) $\oplus$ σ(pile-up noise)). The seed signal is defined as the absolute value of the cell energy, while the noise includes the electronics noise RMS and the pile-up noise RMS added in quadrature. Each obtained seed forms a 'proto-cluster'.
   - Next, the seed cells are processed in decreasing order of energy significance: Starting with the seed with the highest signal to noise ratio, the algorithm adds iteratively neighboring cells, if S/N $\geq$ 2, to the adjacent proto-clusters. If two proto-clusters share the same neighboring cell, they are merged. Perimeter cells defined as cells with $|E_{\rm cell}| > 0$ σ on the

perimeter of the group of seed and neighbors are then added (whatever their energy value is).

- Once all seeds on the 'seed list' are processed, the neighbor cell list becomes the new seed list and the procedure is repeated until all candidates on the seed list are processed.
- All proto-clusters with $|E_T| > 0$ GeV are converted into clusters.
- The cluster expansion is driven by neighbors in 3 dimensions : the set of neighbors definition includes the 8 neighboring cells within the same calorimeter layer in addition to cells partially overlapping in η and ϕ with the central cell in the previous and next layers. If the calorimeter layers granularity was uniform, then this would correspond to 10 neighbors per cell. But, in ATLAS, the corresponding number is greater than 10 since the calorimeter granularity depends on the layer treated and the calorimeter region.
- Electronics and pile-up noise is significantly reduced by the high seeds and neighbors thresholds.
- The absence of threshold for the perimeter cells allows to take into account shower tails.
- Using $|E_T|$ ensures a symmetric noise contribution.

2. *Cluster Splitting :*
A cluster splitting procedure is needed in ATLAS at this stage since most of the clusters obtained at the end of the previous step may cover a large area of the calorimeter merging showers of several particles. However, individual particles may still be found around local maxima of energy deposits and are thus separable. The basic steps to achieve that are the following :

   - **Finding Local Maxima :** Cluster cells with $E_{\text{seed}} > 500$ MeV are chosen as local maxima if they have at least 4 neighbors and their energy is the highest compared to the neighboring cells. The search for primary local maxima is done in the EM samplings (hadronic samplings are secondary). But the presampler and strip layer cells are excluded from the search by default since no large maxima are expected there, and removing them lowers the probability of building noise clusters. On the other hand, local maxima from the strips and hadronic calorimeter not overlapping in η nor in ϕ with primary local maxima are taken into account. Finally, if a parent cluster has no local maximum cell, it will not be considered furthermore for splitting.
   - **Finding Neighbors :** A re-clustering algorithm around the local maximum with the same neighbor is performed under the following conditions:
     - Only originally clustered cells are included
     - No energy thresholds
     - No merging

– If shared cells were added to the neighbor list or to the proto-cluster, they will be discarded. Shared cells are handled separately in the next step.

- **Shared Cells and Finalizing :** Every cell shared between two proto-clusters will be added to them with energy and distance dependent weights (w_1, w_2) reflecting the probability of the cell belonging to the proto-cluster:

$$w_1 = \frac{E_1}{E_1 + rE_2}, \quad w_2 = 1 - w_1, \quad r = exp(d_1 - d_2) \ , \tag{4.3}$$

where E_1, E_2 are the proto-clusters energies and d_i (i = 1, 2) is the distance separating the cell from the center of proto-cluster 'i' in units of EM shower scale. The weight values are usually close to zero or one. At this stage, each local maximum lead to a proto-cluster. Parent clusters with no local maxima found in the first step are added back to the list of final proto-clusters, which will be converted to clusters. The uncertainties on electronics and pile-up noise have a direct impact on the clustering algorithm and its efficiency, leading sometimes to an increase of the number of fake clusters, lowering consequently the efficiency of the clustering [42].

4.3.2.1 Energy Sharing

If some cells are shared by the borders of 2 clusters, the weights given by the cluster splitting algorithm to cells belonging to multiple clusters add up to one. Hence, there is no double counting of energy.

4.3.2.2 Local Hadronic Calibration (LC)

Since the calorimeter in ATLAS is a sampling non-compensating one, its response to electromagnetic (EM) and non-EM components of a shower is different. Consequently, a software calibration of the calorimeter energy measurements is required to correct the signal non-linearity and energy resolution degradation due to the under-compensation. Two methods can be used to achieve this goal :

- **Global Hadronic Calibration :** The hadronic final state objects such as jets and missing transverse energy are reconstructed using the calorimeter measurements fixed at the EM scale, then calibrated using in-situ measurements and normalized using true MC measurements.

- **Local Hadronic Calibration :** Unlike the global hadronic calibration, this method starts with the calorimeter clusters as input and calibrates them, applying local normalization and corrections, before physics objects (e.g. jets) reconstruction.

The basic steps of local hadronic calibration (LC) in ATLAS are [46] :

1. **Cluster Classification :** Clusters obtained at the end of the topological clustering algorithm are classified into "em-like" and "hadron-like" clusters, based on shower variables and phase space population predictions for charged/neutral pions from single pion MC measurements. Good discriminating variables include :

   - Shower depth, width, length, lateral/longitudinal shape variables
   - Energy fractions (in EM calorimeters, in most energetic cells, etc.)
   - Average energy density

   The shower axis is needed for many of these calculations. In fact, the default method uses the depth of the shower center λ_{center} and the first moment in energy density $\langle\rho\rangle$ (= E/V) of the cell constituent variables, in addition to a cut on the EM fraction of a cluster determined from single pion MC measurements: Hadronic showers are deeper and have a smaller average energy density than EM showers. On the other hand, the alternative classification method uses second moments related to the shower width and length $\langle r^2\rangle$ and $\langle\lambda^2\rangle$. If the shower axis is defined along $\overrightarrow{s}$ with a shower center $\overrightarrow{c}$, then :

$$r_i = |(\overrightarrow{x_i} - \overrightarrow{c}) \times \overrightarrow{s}| \,, \tag{4.4}$$

$$\lambda_i = (\overrightarrow{x_i} - \overrightarrow{c}).\overrightarrow{s} \quad . \tag{4.5}$$

2. **Calibration/Weighting :** The "hadron-like" clusters are weighted to correct for the e/π different calorimeter response. The weights are derived using the true and reconstructed cell energies in GEANT4 simulation samples. The true energy is indeed estimated using the calibration hits in GEANT4 available for both active and inactive material: $E_{\text{true}} = E_{\text{EM}} + E_{\text{non-EM}} + E_{\text{inv}} + E_{\text{escaped}}$ with:

   - $E_{\text{EM}} + E_{\text{non-EM}}$ = EM and non-EM visible energy.
   - E_{inv} = invisible energy from non-ionizing processes like the energy lost in breaking nuclei for example.
   - E_{escaped} = energy of non-interacting particles, e.g. neutrinos, escaping the detector.

   The weight associated to a cell is thus defined as $w_i = \langle E_{\text{cell}}^{calib}/E_{\text{cell}}^{reco}\rangle$ in bins of the energy of the cluster to which the cell belongs and cell energy density $\rho_{\text{cell}} = E_{\text{cell}}^{reco}/V_{\text{cell}}$ for various η regions. The index i refers to the bin number.

   - $i = \text{bin}\,(E_{\text{cluster}}, \rho_{\text{cell}})$.
   - E_{cell}^{calib}= energy from all calibration hits , $E_{\text{cell}}^{calib} > \sigma_{\text{noise}}$.
   - E_{cell}^{reco}= reconstructed energy from visible energy deposits, i.e. $E_{\text{EM}} + E_{\text{non-EM}}$, $E_{\text{cell}}^{reco} > 2\ \sigma_{\text{noise}}$.
   - V_{cell} = cell volume, $V_{\text{cell}} > 0$.

– σ_{noise} = electronics noise.

3. **Out of Cluster Correction:** This correction is applied to correct for energy deposits inside the calorimeter cells but not in topo-clusters. Single pions MC simulation samples are used to estimate the total fraction of energy deposited outside clusters based on the calibration hits falling within the calorimeter but outside of clusters. The calculation is done with respect to the total energy inside the clusters before any dead material corrections, and the resulting measurements are expressed in bins of pion energy, $|\eta|$ and calorimeter depth λ. Then, the out-of-cluster energy fraction is multiplied by the cluster degree of isolation. It follows that the energy resolution and the linearity are improved.

4. **Dead Material Correction:** This correction is important to recover energy deposits in dead material. The calibration hits in single pion simulation samples are used to calculate the correction coefficients. The cell weights in clusters are thus changed accordingly to have the correct cluster energy.

4.4 Electrons

4.4.1 Electron Reconstruction and Identification

Electron reconstruction covers cluster reconstruction and electron identification. Information from energy deposits in the EM calorimeter and ID tracks is used for this purpose. Additional information from the hadronic calorimeter can be used to discriminate electrons from hadrons faking electrons.

A-Reconstruction Algorithms

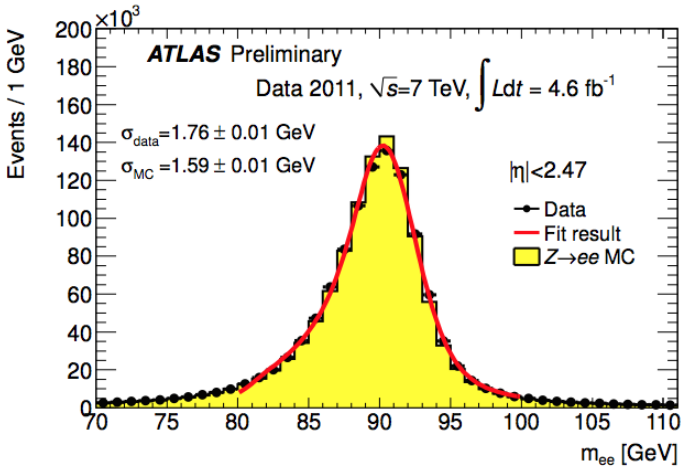
Electron reconstruction is done using two algorithms :

1. EM seeded outside-in algorithm :

   - The reconstruction is outside-in starting with EM clusters and ending with ID tracks matching them.
   - It is the standard algorithm.

2. Track seeded inside-out algorithm :

   - Non-standard algorithm.
   - It starts with reconstructed ID tracks and extends to EM clusters matching them.
   - Useful to reconstruct low p_{T} electrons and to find electrons within or close to jets.



(a)

Figure 4-9: Resolution of the invariant mass from $Z \to ee$ events after applying all energy corrections in the 2011 data set. The fit shown in red uses a Breit-Wigner convoluted with a Crystal Ball function [15].

B- Reconstruction Steps

Once the electron candidate object is obtained after matching EM clusters with ID reconstructed tracks, electron identification is the next step. Quality criteria are applied to select electrons with the best possible efficiency, separating electrons from background contamination coming from jets and photons. A baseline energy calibration based on test-beam data followed by a residual energy calibration based on $Z \to ee$ events is applied. Figure 4-9 shows the $Z \to ee$ invariant mass distribution in both data and MC samples after applying the electron energy calibration for 2011 data [31].

1. ***Cluster Reconstruction*:** The standard algorithm uses EM clusters with total transverse energy E_{T} of at least 2.5 GeV. The clusters are obtained using the sliding-window algorithm discussed in section 4.3. Cluster reconstruction efficiency is high and increases as the electron transverse energy E_{T} increases: it is $\simeq 95\%$ for electrons with $E_{\mathrm{T}} = 7$ GeV, goes up to 99% at $E_{\mathrm{T}} = 15$ GeV and reaches 99.9% at $E_{\mathrm{T}} = 45$ GeV in MC simulations [30, 31].

2. ***Track Association with Cluster*:** ID tracks with $p_{\mathrm{T}} > 0.5$ GeV are extrapolated to the middle EM calorimeter layer at the level of shower maximum (80% of the EM shower is deposited in this layer). A track is said to match the cluster if its extrapolation is within a window $\Delta\eta \times \Delta\phi = 0.05 \times 0.1$ from the cluster position. The 0.1 window in $\Delta\phi$ takes into account the extrapolated track curvature due to the solenoid magnetic field and energy loss by bremsstrahlung. In case of no track-cluster match, the cluster is defined as an unconverted photon candidate. The tracks used for the electron reconstruction are fitted by

the Gaussian Sum Filter (GSF) algorithm for 2012 data as described in section 4.4.2. Default KF technique using the pion particle hypothesis was used for 2011 data [16, 30].

3. ***Electron candidate reconstruction*:** A successful track-cluster match is followed by an optimization of the cluster size to take into account the overall energy distribution in different parts of the calorimeter. The final total energy of the electron therefore combines the estimated energy deposited in the material *in front of* the EM calorimeter, *outside* the cluster (lateral leakage) and *beyond* the EM calorimeter (longitudinal leakage) with the energy measured *within* the cluster window. Indeed, all these four contributions are used to calculate the final cluster's energy [31].

The resulting electron has a four-momentum calculated using the information from the final cluster and its best matching track. The electron energy is associated to the cluster energy while the track parameters are used for the (η, ϕ) coordinate measurements.

C- Electron Identification and Isolation

At this point, electrons are classified into 3 different categories with different efficiency and fake rejection levels. In 2011, a cut-based method was used whereas for 2012, multivariate analysis (MVA) techniques were developed and have been used in some Higgs analyses, with the cut-based technique being the default one for 2012. The categorization is based on several variables related to hadronic leakage, shower shape, calorimetry and tracking classifying electrons into 'loose', 'medium' and 'tight' (for cut-based). These categories, described hereafter, are ordered with respect to decreasing signal efficiency and increasing background rejection: with increasing tightness, the cuts on the variables already used are tightened and new variables are introduced to cut on.

- **Loose**: uses the EM calorimeter middle layer information on shower shape and hadronic leakage variables. This has the best efficiency but the lowest background rejection.

- **Medium**: adds to the 'loose' requirements cuts on variables from the EM calorimeter strip layer related track quality and track-cluster matching. Compared to the 'loose' selection, the background rejection is increased by a factor of 3 or 4 and the reconstruction efficiency is reduced by $\simeq 10\%$.

- **Tight**: adds to the 'medium' selection criteria requirements on E/p (E = cluster energy and p = associated track momentum), on the number of TRT hits and identification with TRT along with the discrimination against photon conversions.

Electron Isolation

In addition to these cuts, an isolation cut is applied to further reduce background from non-prompt electrons (i.e. from hadron decays within jets) and jets faking electrons. Two options are available :

- ***Track based isolation*** : $\mathrm{I}\,(p_{\mathrm{T}},\,\Delta R) = \dfrac{{p_{\mathrm{T}}}^{\mathrm{isolation}}}{{p_{\mathrm{T}}}^{e\ell}}$, where :
  -${p_{\mathrm{T}}}^{\mathrm{isolation}}$ is the isolation energy calculated using the sum of the transverse momentum of the tracks in a cone of size ΔR around the electron track.
  -${p_{\mathrm{T}}}^{e\ell}$ is the electron transverse momentum.
  Typical ΔR values used are 0.2, 0.3 or 0.4 depending on the analysis. Tracks must be good quality tracks with $p_{\mathrm{T}} > 400$ MeV and coming from the same primary vertex as the electron.

- ***Calorimeter based isolation*** : $\mathrm{I}\,(E_{\mathrm{T}}\,,\,\Delta R) = \dfrac{{E_{\mathrm{T}}}^{\mathrm{isolation}}}{{E_{\mathrm{T}}}^{e\ell}}$, where :
  -${E_{\mathrm{T}}}^{\mathrm{isolation}}$ is the isolation energy calculated as the sum of the transverse energy of the calorimeter topological clusters found within a cone of ΔR around the cluster seed.
  -${E_{\mathrm{T}}}^{e\ell}$ is the electron transverse energy.
  The cluster contribution within $\Delta\eta \times \Delta\phi = 0.125 \times 0.175$ around the electron cluster barycenter is not included [31]. Typical ΔR values used are 0.2, 0.3 or 0.4 depending on the analysis.

Each of these methods has its advantages. While the track-based isolation calculation is limited to charged particles, the calorimeter-based one includes contributions from both neutral and charged components, which makes it more sensitive to the activity in the electron neighborhood. On the other hand, the energy estimated from the calorimeter cluster is sensitive to both in-time and out-of-time pile-up, whereas the track-based measurement is less pile-up dependent : $\mathrm{I}\,(p_{\mathrm{T}},\,\Delta R)$ values are sensitive to in-time pile-up only, but the effect is reduced since only tracks associated to the same primary vertex as the electron are considered.

4.4.1.1 Electron Pseudorapidity

Electrons, defined this way, have a cluster pseudorapidity $|\eta_{\mathrm{cluster}}| < 2.47$; however, the ones falling in the barrel transition region $1.37 < |\eta_{\mathrm{cluster}}| < 1.52$ are excluded. In addition to this standard reconstruction, a different method is used to reconstruct forward electrons ($|\eta_{\mathrm{cluster}}| > 2.47$) without associated tracks. The relative details will not be discussed since the latter are not used in the analyses presented in this thesis.

4.4.1.2 Data/MC Differences

Data/MC efficiency differences can be seen at all the stages discussed earlier : reconstruction, identification and isolation. These efficiency measurements are in fact

estimated using a tag-and-probe method on $Z \to ee$ and $J/\Psi \to ee$, where one electron passing the standard cuts is chosen as 'the tag'. Then, a search for the second electron with looser cuts, i.e. 'the probe', is performed and the probability to reconstruct it in terms of η and E_{T} gives the efficiency estimate. A cut on the di-electron invariant mass (around Z boson or J/Ψ mass) is usually applied to increase the purity of the samples and reduce the background contribution. Additional efficiency measurements are also performed on $W \to e\nu$ samples with high statistical power, tagging on the missing transverse momentum [30, 31]. At the analysis level, a scale factor (SF) is applied to the MC samples to take care of the small differences seen with respect to data. A scale factor is just a correction factor defined as the ratio of data to MC efficiencies.

$$\mathrm{SF} = \frac{\epsilon_{\mathrm{data}}}{\epsilon_{\mathrm{MC}}} \tag{4.6}$$

$\epsilon_{\mathrm{total}} = \epsilon_{\mathrm{reco}} \times \epsilon_{\mathrm{ID}} \times \epsilon_{\mathrm{trigger}} \times \epsilon_{\mathrm{additional}}$ with:

- $\epsilon_{\mathrm{total}}$: total efficiency to detect an electron.
- ϵ_{reco}: reconstruction efficiency as measured in $Z \to ee$ events.
- ϵ_{ID}: identification efficiency as calculated using $Z \to ee$ events for electrons with $E_{\mathrm{T}} > 10$ GeV and $J/\Psi \to ee$ for electrons with lower p_{T}.
- $\epsilon_{\mathrm{trigger}}$: trigger efficiency as measured in $Z \to ee$ events.
- $\epsilon_{\mathrm{additional}}$: efficiency corresponding to additional selection cuts such as isolation cuts for example.

SFs are close to unity and are dependent on (E_{T} , η) since the calibrated electron energy depends on them. SF values have deviations of a few percent away from unity in regions with low E_{T} or high η. In fact, for high-E_{T} measurements, a high statistical precision is achieved for various binnings in η (coarse, middle, fine) with an uncertainty of the order of few per mille at 35 GeV . For tight electrons with $E_T > 20$ GeV, the uncertainty is within 1-2%. However, at low-E_{T} , the $J/\Psi \to ee$ are the main contributing events. The efficiency measurements are statistically limited and are done with a single η binning ('*coarse*'). The associated uncertainties vary from 3% in the barrel region to 7% in the end caps. On the other hand, the data/MC efficiency discrepancy at high η values is due to calorimeter shower shape mismodelling in the MC samples and to the tighter requirement imposed on shower shapes for $|\eta| > 2$, where no TRT information is accessible, to keep a satisfactory rejection rate [31].

4.4.2 Gaussian Sum Filter

Track fitting is done with the pion particle hypothesis by default. For 2011, the track fitting procedure was the same for all charged particles. For 2012, in order to

take bremsstrahlung into account , the electron track refitting was done using the Gaussian Sum Filter (GSF) algorithm. If there is no Bremsstrahlung (e.g. for the pion or muon tracks), the ratio of the true p_{T} to the reconstructed one p_{reco} has a Gaussian distribution. The GSF is a generalization of the Kalman filter where it estimates energy loss by a sum of Gaussian functions instead of just one Gaussian. The track parameters are improved after refitting and used again for the track-cluster matching and the electron four-momentum calculation. This algorithm leads to better reconstruction and identification efficiencies. The impact of the GSF algorithm on electron track resolution and impact parameters is shown in Fig. 4-10. The overall reconstruction efficiency is increased by 5% [16, 31].

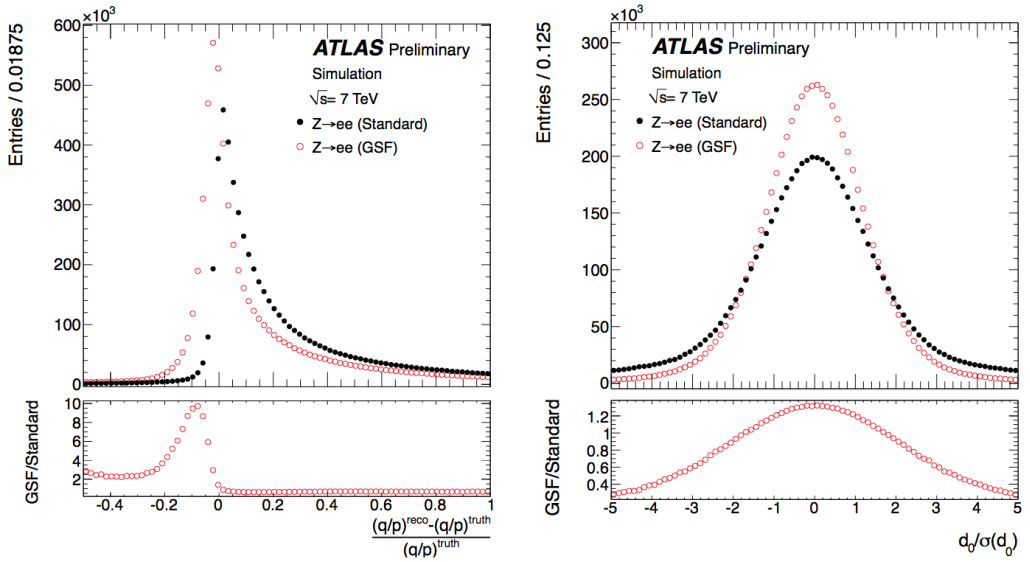
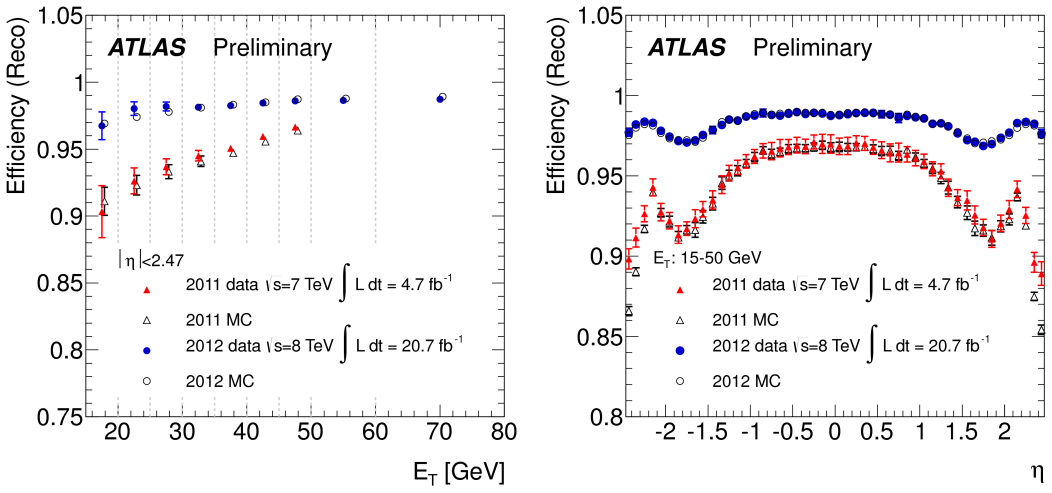


Figure 4-10: GSF impact on the electron track resolution and impact parameter in simulated MC events [16].

4.4.3 Performance

Integrating over $|\eta|$, the electron reconstruction efficiency goes from 97% at $E_{\mathrm{T}} = 15$ GeV to 99% at $E_{\mathrm{T}} = 50$ GeV. Below 15 GeV, the efficiency is not measured given the large background contamination. Above this threshold, the efficiency decreases with $|\eta|$: it varies between 99% at low η and 95% at high $|\eta|$. The pile-up dependence is less than 4% for events with less than 30 reconstructed primary collision vertices regardless of the selection criteria. Finally, the electron identification total uncertainties are small and decrease as E_{T} increases: they are near 5-6% for electrons with $E_{\mathrm{T}} < 25$ GeV and 1-2% otherwise [31]. A summary of the reconstruction efficiencies in 2011 and 2012 data sets is given in Figs. 4-11 and 4-12. Unlike 2012 events where GSF is applied, 2011 reconstructed electrons were not corrected for bremsstrahlung and have a lower, less uniform reconstruction efficiency in terms of E_{T} and η. The dip around $|\eta| = 1.5$ seen in (b) corresponds to the barrel-end cap transition region ($1.37 < \eta < 1.52$).



(a) (b)

Figure 4-11: Measured electron reconstruction efficiencies as a function of E_{T} integrated over the entire pseudorapidity range in (a) and as a function of η for E_{T} between 10 and 50 GeV for 2011 and 2012 data sets [30].

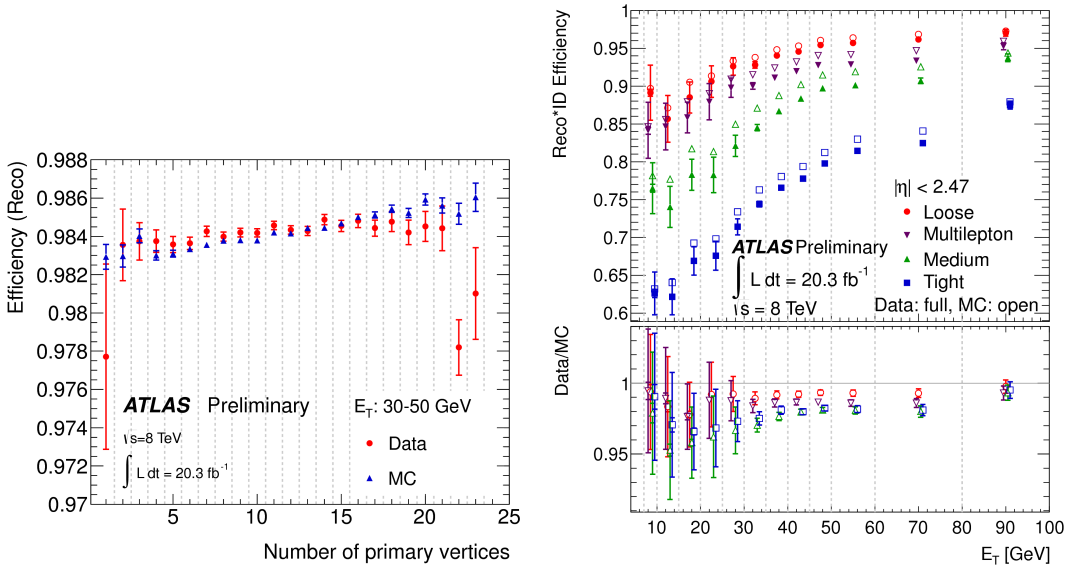
4.5 Muons

There are 4 types of muons reconstructed in ATLAS using different criteria based on the information from one or more of the following sub detectors: ID, the muon spectrometer (MS) and the calorimeter. Muons are classified as :

- **Stand-alone(SA) muons:** The muon trajectory is reconstructed in the MS only. To obtain the muon parameters at the interaction point e.g. track parameters (angular parameters and impact parameter for example) and momentum, the MS track is extrapolated close to the beam line and the momentum value is adjusted to take into account energy losses of the muon.

- **Combined (CB) muons:** The muon trajectory is reconstructed using the ID and the MS only. SA tracks are combined with the reconstructed ID tracks based on the χ^2 difference in the corresponding parameters and refit to obtain the 'combined' muon. This is used by the STACO (*STAtistical COmbination*) algorithm. Given the limited acceptance of the ID, CB muons $|\eta|$ values can not exceed 2.5. The momentum of the reconstructed muon is a weighted sum of the two momenta values used in the combination. This is the standard muon type in ATLAS.

- **Segment-tagged (ST) muons:** If an ID track matches a segment in the muon precision chambers (Monitored Drift Tubes (MDT) or Cathod Strip Chambers(CSC)) once extrapolated to the MS, it is identified with a ST muon. This is mainly used for low-p_{T} muons or for muon reconstruction in MS regions with reduced acceptance.



(a) (b)

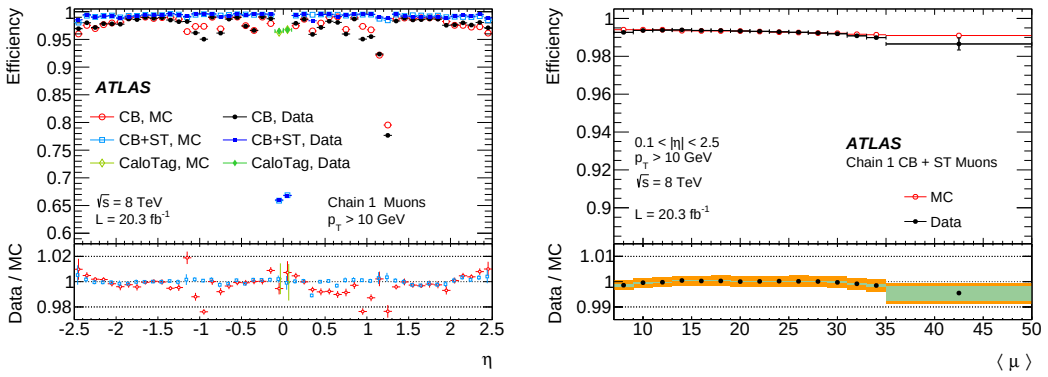
Figure 4-12: Measured electron reconstruction efficiencies as a function of the number of primary vertices for 2012 data set is shown in (a). The combined reconstruction and identification efficiency as a function of E_{T} is plotted in (b) for cut-based loose, medium, tight and multilepton (They are mainly loose electrons with cuts on TRT variables used for Higgs→ZZ channel) selections. (b) shows the statistical and systematic uncertainties on the measured efficiencies along with the data to MC efficiency ratio, useful for SF calculation, drawn in the bottom plot [30].

- **Calorimeter-tagged (CaloTag) muons:** When an ID track can be matched to a minimum ionizing particle (MIP) energy deposit in the calorimeter, this corresponds to a CaloTag muon. This type has the lowest purity but allows to recover acceptance in regions with no MS coverage. Optimal performance for this type is achieved for regions with $|\eta| < 0.1$ for a transverse momentum range $25 \leq p_{\mathrm{T}} \leq 100$ GeV.

Muon reconstruction in ATLAS during Run I is performed using two methods: STACO, or Chain 1, combines the ID track parameters with the MS information in a statistical way and MUID, or Chain 2, which refits entirely the tracks using hits from the ID and the MS sub detectors [26].
STACO combined muons are commonly used especially that CB candidates have the highest purity. All analyses presented in this thesis use STACO combined muons, which is the default option for ATLAS analyses. However, reconstructing muons this way is subject to the following limitations:

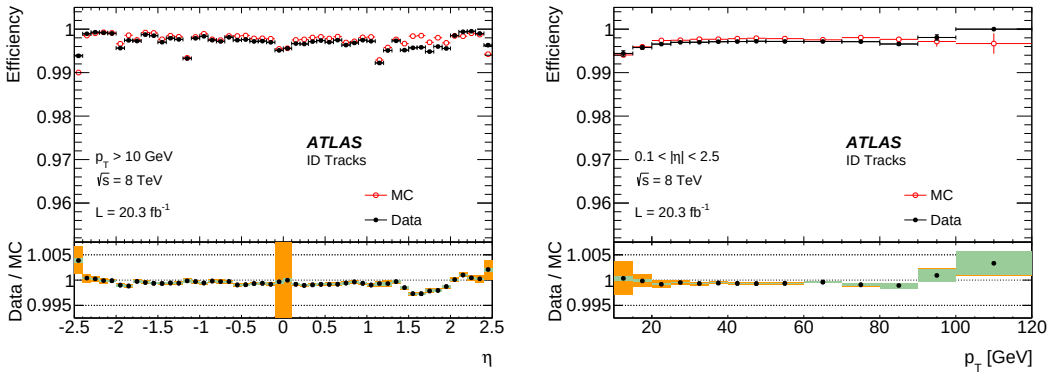
- The MS is not well equipped with muon chambers at $\eta = 0$ in order to leave space for ID and calorimeter services. In this region, muon reconstruction relies on ID and calorimeter information (CaloTag muons).

- In the region ($1.1 < \eta < 1.3$), only one layer of the MS is crossed by muons. This is true for Run I since the installation of all necessary muon chambers in this region was completed during the 2013-2014 shutdown. Hence, for Run I analyses, the momentum for SA muons is not available there.



(a) (b)

Figure 4-13: Chain 1 muon reconstruction efficiency as a function of η measured in $Z \to \mu\mu$ events for various muon reconstruction types. This is done for muons with $p_{\mathrm{T}} > 10$ GeV. CaloTag muons points are limited to the region in η where they are used in physics analyses. The CB + ST efficiency pile-up dependence is shown in (b) [33].



(a) (b)

Figure 4-14: ID muon reconstruction efficiency as a function η (a) and p_{T} (b) [33]. Pure statistical uncertainties are in green. Combined statistical and systematic uncertainties are shown in orange.

4.5.1 Muon Reconstruction Efficiency

The muon reconstruction efficiency is defined as $\epsilon(\text{type}) = \epsilon(\text{type} \mid \text{ID}) \times \epsilon(\text{ID})$ with :

- ϵ(type | ID) = matching efficiency
- ϵ(ID) : efficiency in the ID
- type = CB, ST

The efficiency is not the same for all (η, ϕ) regions. In fact, the efficiency decreases as expected for regions near $|\eta| \simeq 0$ where the ID and the calorimeter services are provided, and around $|\eta| \simeq 1.2$ where some MS muon chambers were missing in Run I.

To estimate the muon reconstruction efficiency, a tag-and-probe method is applied on $Z \to \mu\mu$ and $J/\Psi \to \mu\mu$ events in a similar way to the electron case [33]. A cut on the di-muon invariant mass compatible with a Z boson or J/Ψ is applied to reduce the background from non-prompt or fake muons. With the tag-and-probe method, the requested quality of the muon labeled as a tag is usually high. The probability to reconstruct the other muon, i.e. *the probe*, correctly is then studied as a function of η and ϕ. The difference in reconstruction efficiencies between data and MC simulation samples is corrected for using a scale factor defined as: SF = $\epsilon_{\text{data}}/\epsilon_{\text{MC}}$. This correction is applied to MC muons. The reconstruction efficiency for 2012 is shown in Fig. 4-13-(a) : the efficiency is $\simeq$ 99% and calorimeter measurements are used to recover the gap at $|\eta| = 0$ where there is not enough muon chambers. The reconstruction efficiency of ID muons with $p_{\text{T}} > 10$ GeV in $Z \to \mu\mu$ events as a function of p_{T} and η is shown in Fig. 4-14. The data/MC discrepancy seen in the region $1.5 < \eta < 2$ is due to incorrect treatment of non-operating B-layer pixel modules in data reconstruction, causing an efficiency in data lower than that expected in the MC. A hit in this layer is requested though to ensure precise impact parameter measurements. This is corrected for by the SFs shown in the ratio plot at the bottom. Away from this region, the data/MC agreement is excellent with an efficiency greater than 99%. In addition, the reconstruction efficiency was studied as a function of pile-up. Figure 4-13(b) shows the CB + ST muon reconstruction efficiency as a function of the average number of interactions per crossing $\langle\mu\rangle$. High efficiency values are obtained (on average about 0.97) with a good stability against pile-up. The efficiency drop of about 2% for $\langle\mu\rangle$ near 35 is due to the reduced ID reconstruction efficiency in data as explained earlier.

4.5.2 Isolation

As for electrons, the muons are required to be well isolated. This is done using track and calorimeter isolation estimators. The track-based isolation method defines the isolation energy as the sum of the transverse momenta of the tracks in a cone ΔR (= 0.2, 0.3 or 0.4 depending on the analysis) around the muon track. The calorimeter-based method defines the isolation energy as that being deposited in calorimeter clusters in a cone ΔR (= 0.2, 0.3 or 0.4 depending on the analysis) around the muon's trajectory. If there are any data/MC isolation efficiency differences, this is taken care of by the corresponding scale factors. Figure 4-15 shows the calorimeter isolation efficiency for 2011 before and after pile-up correction [41]. For 2012 data sets, the

data/MC disagreement is within 0.5% so the isolation efficiency scale factor used is SF = 1.00±0.5%.

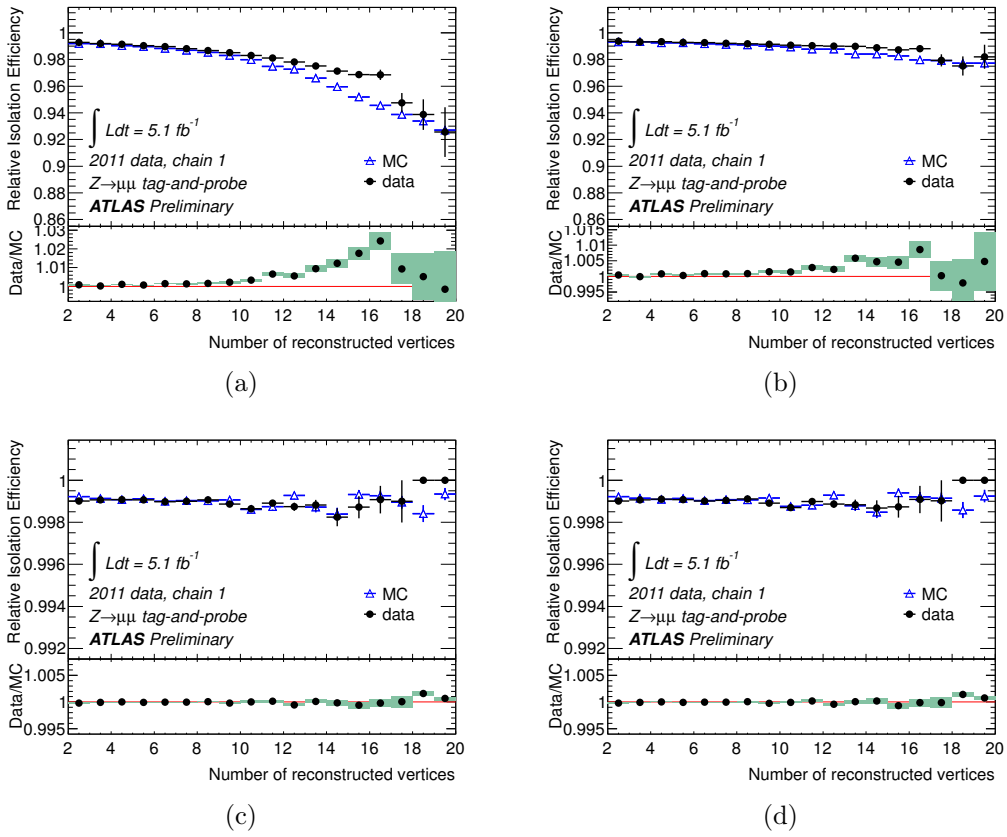


Figure 4-15: Muon isolation efficiency before (left) and after (right) pile-up corrections for 2011 data sets as done in $H \to WW^*$ analysis for $\Delta R = 0.3$ (top) and in $H \to ZZ$ analysis for $\Delta R = 0.2$ (bottom) with $\Sigma E_T/p_T^{\mu} \leq 0.14$ [41].

4.5.3 Scale and Momentum Resolution

Di-muon Z, J/Ψ and Υ resonance samples, with a background of the order of 0.1%, are used to extract the momentum scale and resolution, in order to have a complete description of the muon reconstruction performance. A good approximation of the resolution on the muon momentum for p_T >20 GeVis given by [10, 26] :

$$\frac{\sigma(p_T)}{p_T} = a \oplus b \ . \ p_T \ , \tag{4.7}$$

where :

- $\frac{\sigma(p_T)}{p_T}$ is the fractional momentum resolution.

- 'a' is constant in p_{T} describing multiple scattering effect.
- The term 'b . p_T' describes detector related contribution, i.e. intrinsic resolution given by the detector spatial resolution.

However, when comparing data and MC samples, one notices a small shift in momentum scale. Furthermore, the momentum resolution of MC muons is slightly better. To account for these differences, one corrects the MC muon momentum by a scale correction factor and allows for additional smearing of the resolution. The corrected MC momentum is then defined as:

$$p_{\mathrm{T}}^{\mathrm{Cor,det}} = p_{\mathrm{T}}^{\mathrm{MC,det}} . s^{\mathrm{det}}(\eta)(1 + \Delta a^{\mathrm{det}}(\eta)G(0,1) + \Delta b^{\mathrm{det}}G(0,1)p_{\mathrm{T}}^{\mathrm{MC,det}}) \ , \tag{4.8}$$

where

- det = MS, ID.
- G(0,1) is a random variable having a normal distribution with mean at 0 and width equal to 1.
- $s^{\mathrm{det}}(\eta)$: momentum scale correction factor.
- $\Delta a^{\mathrm{det}}(\eta), \Delta b^{\mathrm{det}}(\eta)$: resolution correction terms derived in 16 different η regions of the detector.

The correction factors are obtained using a MC template fitting technique. The scale and resolution terms needed for the muon momentum correction for 2012 are shown in Fig. 4-16, while Fig. 4-17 shows the di-muon mass distribution in $Z \rightarrow \mu\mu$ events before (a) and after (b) the scale correction. The final distribution with muon momentum scale correction and smearing is shown in Fig. 4-18(a). Figure 4-18(b) shows the data/MC mass ratio with the scale uncertainty for Chain 1 CB muons for the di-muon decays of the $Z, J/\Psi$ and Υ resonances as a function of the pseudorapidity η.

4.6 Jets

Quarks and gluons cannot be seen as free particles due to confinement. Instead, they are observed as collimated bundles of hadrons, created by the fragmentation and hadronization of the partons. The purpose of the jet algorithms is to collect those hadrons into a physical object called a jet. Also, depending on the input used, we can define different types of jets as shown in Fig. 4-19 :

1. Particle (or truth) jets: reconstructed from stable[1] truth particles in MC simulation
2. Track-jets: built using tracks of charged particles coming from PV.
3. Topo-cluster jets: obtained using topological calorimeter clusters (defined in section 4.3)

[1]Stable particles are defined to be particles with lifetime longer than 10 ps.

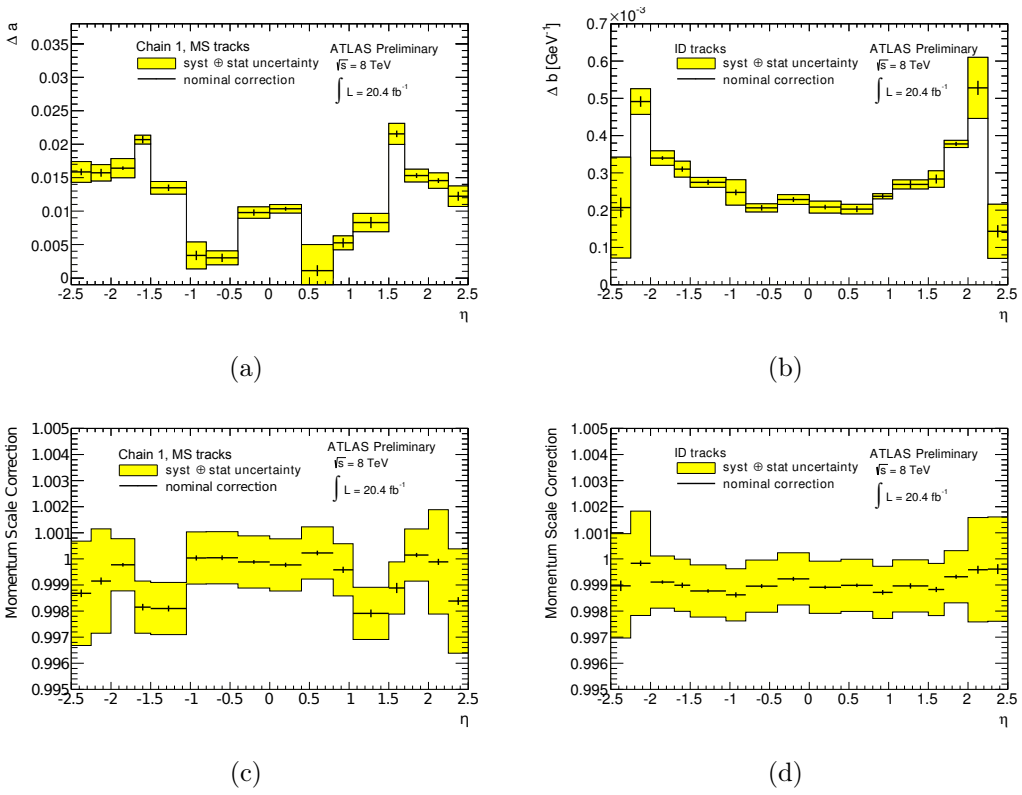


Figure 4-16: Resolution correction terms for the MS (a) and ID (b) muons. The scale corrections for the MS and ID muons are shown in (c) and (d) respectively [26].

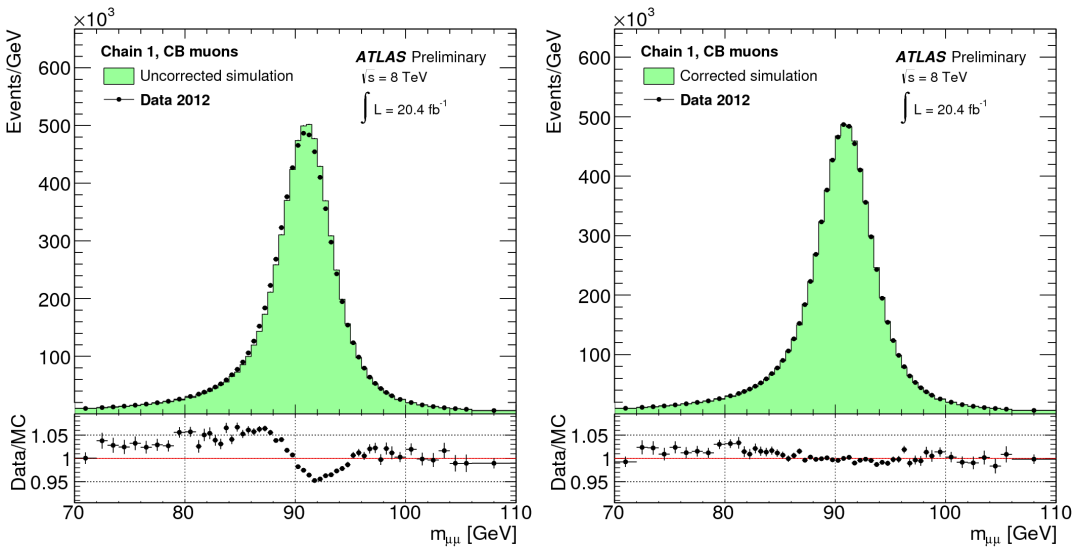
4.6.1 Jet Reconstruction and Calibration

In ATLAS, the analyses presented in this thesis use topo-cluster jets. The jet reconstruction and calibration steps are as follows:

1. *Choice of topo-clusters :*
   Topo-clusters are available at two energy scales: the EM scale and the LCW scale as described in section 4.3. Based on the used cluster type, the jets will be called either jet_{EM} or jet_{LCW}.

2. *Jet algorithm :*
   Jets are built using the anti-k_T algorithm [2, 43] as illustrated in Fig. 4-20. The algorithm is a general successive clustering algorithm doing iterative pair-wise clustering. It consists in finding the minimum distance d_{min} of all calculated distances d_{ij} and d_{iB} defined as:

   - Distance between 2 objects i and j : $d_{ij} = min(p_{T_i}^{-2}, p_{T_j}^{-2}))\frac{\Delta R_{ij}^2}{R^2}$.
   - Distance to the beam : $d_{iB} = \frac{1}{p_{T_i}^2}$.
   - $\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ where y and ϕ are the rapidity and the azimuthal angle of object i.



(a) (b)

Figure 4-17: Invariant mass distribution of Chain1 CB muons in $Z \to \mu\mu$ events before (a) and after (b) applying scale and momentum smearing corrections [26].

with p_{T_i} denoting the transverse momentum of object i. The parameter R defines the jet size. In ATLAS, the R values are 0.4 or 0.6. Starting with the identification of the minimum distance, if the minimum is a d_{ij}, then entities i and j are recombined and their momenta are summed over. Otherwise, if $\min(d_{ij}) > d_{iB}$, then i is a jet and is removed from the list of particles. This is repeated iteratively, recalculating the distances, until no more possible combinations can be found. The algorithm clusters all soft particles around hard particles first. Hence the jets are perfectly round cones with radius R around the hardest particle. Consequently, jet shapes are not modified by soft radiation. The anti-k_T algorithm satisfies 2 important conditions that are required for any good jet finding algorithm: it is infrared (soft radiation/gluon emission) and collinear safe (collinear splitting). The jets used in the analyses presented here are reconstructed with $R = 0.4$ within the *FastJet* software[2] [3]. At this stage, the EM (LCW) jets need to be calibrated applying corrections for pile-up, the jet direction and the jet energy scale [37]. The calibration consists of 4 steps:

(a) Pile-up corrections

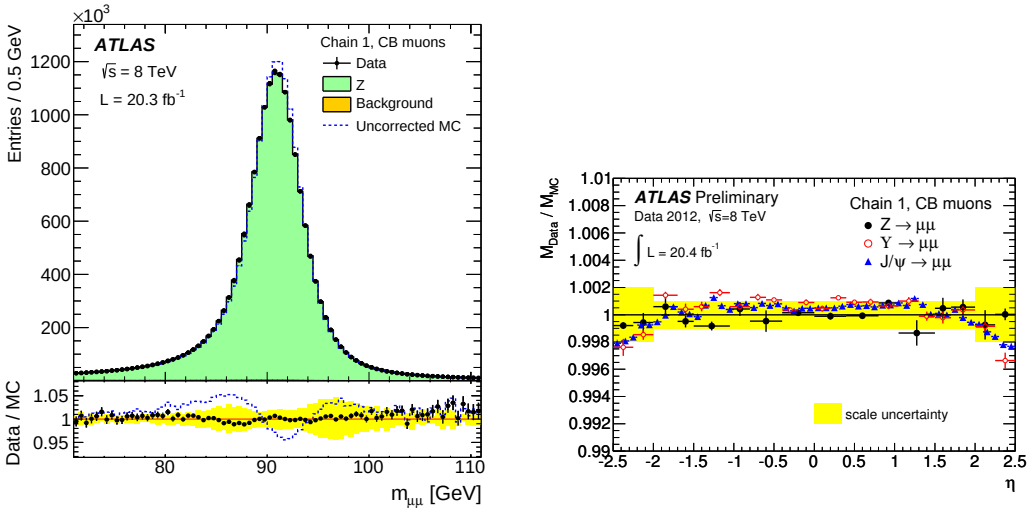
(b) Origin correction

(c) Calibration of the jet pseudorapidity and energy

(d) Residual calibration using in-situ measurements

3. *Pile-up corrections:*
Some of the main pile-up challenges affecting jets caused by additional energy

[2]FastJet is a package giving a geometrically-based implementation of various jet finding algorithms such as k_T, anti-k_T and Cambridge-Aachen algorithms.



(a) (b)

Figure 4-18: Invariant mass distribution of $Z \to \mu\mu$ events with Chain 1 CB muons(a) [33]. The signal muons are plotted along with the background events. Filled histograms are for MC samples with muon momentum corrections applied, while the dashed histogram shows the MC distribution without these corrections. Plot (b) shows data to MC Chain 1 CB di-muon mass ratio for $Z \to \mu\mu$ (both muons with $p_{\mathrm{T}} > 25$ GeVand same η bin), $J/\Psi \to \mu\mu$ (both muons with $p_{\mathrm{T}} > 6$ GeVand same η bin) and $\Upsilon \to \mu\mu$ (both muons with $p_{\mathrm{T}} > 6.5$ GeV) events [26]. The η of the leading muon is shown.

deposits are offsets in the jet energy, degradation of the energy resolution and creation of additional fake jets. For 2011, the pile-up correction consisted of estimating the offset in the jet transverse momentum due to pile-up and subtracting it [37]. The offset is calculated as a function of the number of primary vertices N_{PV} and average number of interactions per crossing $\langle\mu\rangle$ for each jet definition separately. The term proportional to (N_{PV} -1) describes the in-time pile-up while the term proportional to $\langle\mu\rangle$ is sensitive to the out-of-time pile-up. The correct jet p_{T} is given by:

$$p_T^{\mathrm{corr}} = p_T - \alpha \times (N_{PV} - 1) - \beta \times \langle\mu\rangle \ . \tag{4.9}$$

α and β are coefficients obtained from MC dijet samples and depend on the pseudorapidity η. The correction as defined in eq. (4.9) assumes correct identification of tracks and vertices. With the increasing pile-up conditions for 2012 data, the method was improved and the implemented correction consists of a jet area-based correction followed by residual offset corrections [1, 25, 35, 38]. The corrected jet p_{T} is thus defined for 2012 data by:

$$p_T^{\mathrm{corr}} = p_T - A \times \rho \ , \tag{4.10}$$

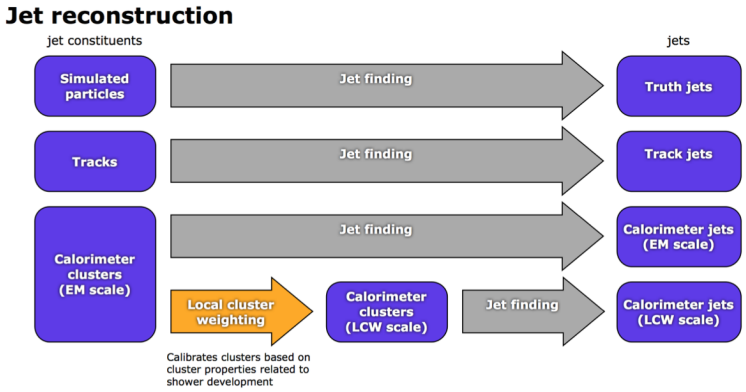


Figure 4-19: Overview of the ATLAS jet reconstruction[37].

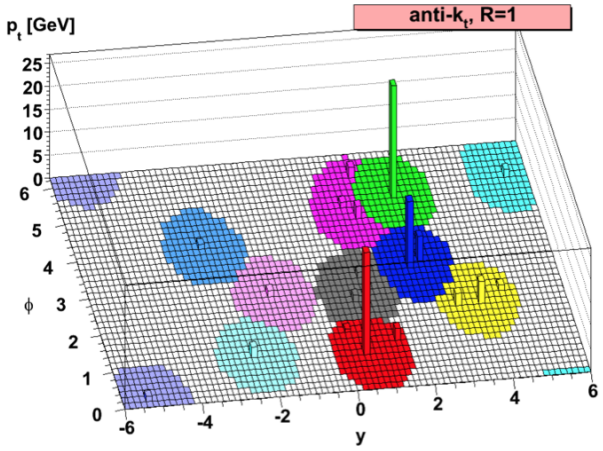


Figure 4-20: Example of an event with random soft emissions at the parton-level clustered into jets using anti-k_T algorithm with $R = 1$[2].

where A is the jet area and ρ is the median p_{T} density as defined later in sections 4.6.2.1 and 4.6.2.2 respectively. The residual offset correction is similar to that applied for 2011 data. The main advantages of adding a jet area-based correction are improvements seen in the resolution and the pile-up jets rejection factor as pile-up fluctuations are taken into account event by event through the term in ρ, in addition to the fact that the correction is largely less dependent on the jet definition and pile-up effect on vertex reconstruction.

4. *Origin corrections :*
The jet direction is corrected so that it points to the event primary hard-scattering vertex (PV). This is achieved by adjusting the kinematic terms of each constituent topo-cluster in a way that it points to the PV. The raw measurement of the jet four-momentum, defined as the sum of the four momenta of the constituent topo-clusters, is thus recalculated. And the jet raw pseudo-rapidity measurement is replaced by η_{origin} [37].

5. *Calibration of the jet pseudorapidity and energy scale (absolute JES):*
The pseudorapidity η and energy correction factors are derived using reconstructed and truth jets in inclusive jet MC samples. The energy calibration corresponds to a multiplication by a correction factor defined as the inverse of average energy response given by [36, 37]:

$$R^{\mathrm{EM(LCW)}} = \langle E_{\mathrm{jet}}^{\mathrm{EM(LCW)}} / E_{\mathrm{jet}}^{\mathrm{truth}} \rangle \ . \tag{4.11}$$

Jets originally calibrated at the EM (LCW) scale are referred to, after this correction, as EM (LCW)+JES calibrated jets. The responses 'R' are defined in bins of the jet pseudorapidity and transverse momentum. The origin and energy corrections are then followed by a jet pseudorapidity η calibration consisting in the addition of a MC-derived correction factor $\Delta\eta = \eta_{\mathrm{truth}} - \eta_{\mathrm{origin}}$, defined in terms of the jet energy and the uncorrected pseudorapidity. The values of $\Delta\eta$ are very small, going up to 0.05 near the transition regions [37].

6. *Residual calibration using in-situ measurements:*
This is applied to data only, where the jet p_{T} correction factor is defined as the MC to data response ratio:

$$\frac{\mathrm{Response}_{\mathrm{MC}}}{\mathrm{Response}_{\mathrm{Data}}} = \frac{\langle p_T^{\mathrm{jet}} / p_T^{\mathrm{ref}} \rangle_{MC}}{\langle p_T^{\mathrm{jet}} / p_T^{\mathrm{ref}} \rangle_{Data}} \ . \tag{4.12}$$

The ratio $\langle p_T^{jet} / p_T^{ref} \rangle$ is obtained from the transverse momentum balance between the jet and a reference object. Several methods are used:

   (a) γ+jet balance

   (b) Z+jet balance

   (c) multi-jet balance

   (d) di-jet η-intercalibration: this is the η-intercalibration (calibration from central to forward region of the detector)

Methods (a)-(c) apply to central jets where photons and Z bosons are used as reference objects for jets with $p_{\mathrm{T}} < 800$ GeV, and low-p_{T} jets recoiling against a high-p_{T} jet are taken as a reference otherwise. The η-intercalibration using QCD di-jet events is applied to equalize the jet response in η, where well calibrated central jets are used to probe forward ones to have a uniform response. These methods are described in detail in [17, 37]. Fig. 4-21 shows that the data to MC jet response ratio for 2012 data samples, for LCW+JES jets with $|\eta| < 0.8$, as a function of the jet p_{T}, combining three different in-situ jet residual calibration methods. The total in-situ uncertainties are of the order of 2%.

Once calibrated, good jet candidates are defined as jets with $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.5$. Higgs analyses use jets with p_{T} greater than 25 (e.g. $H \to WW^*$) or 30 GeV (e.g. $H \to \tau^+\tau^-$). Main sources of systematic uncertainties include:

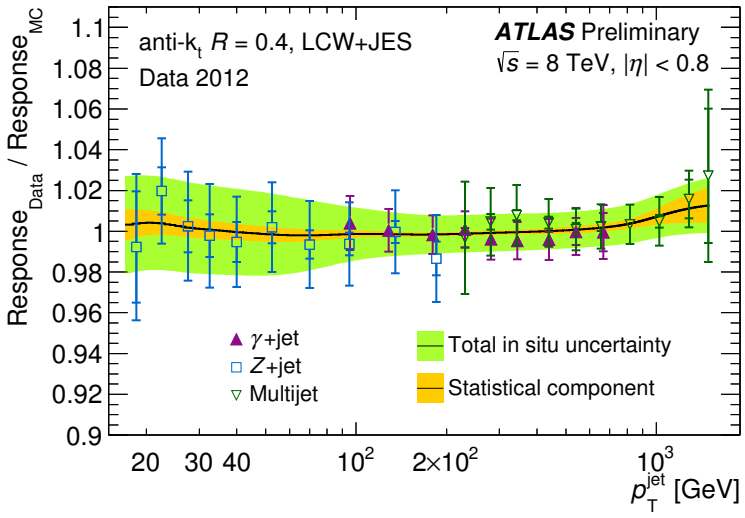


Figure 4-21: Data to MC response ratio for anti-k_T (R=0.4) LCW+JES jets in 2012 data as a function of the jet p_{T}. The error bands show the statistical uncertainties as well as the total (systematic and statistical uncertainties added in quadrature) uncertainties for three in-situ jet energy scale methods combined. The plot shows also the results for each method separately [23].

1. Baseline ones e.g. baseline jet calibration uncertainties related to in-situ calibration, forward JES uncertainty related to the η-intercalibration, in-time event pile-up uncertainty depending on the number of primary vertices and out-of-number pile-up uncertainty based on the average number of interactions $\langle\mu\rangle$

2. Jet flavor and topology related uncertainties: close-by jet uncertainties reflecting the response to non-isolated jets, flavor composition and response uncertainties (quark/gluon jets), b-jet energy scale

Figure 4-22 shows the JES uncertainties as a function of the pseudorapidity and transverse momentum. The maximum fractional uncertainty for central jets is less than 4% for $p_T > 20$ GeV and less than 2% for p_T within the range [100, 1000] GeV. As a function of the pseudorapidity η for jets with $p_T > 40$ GeV, the fractional JES uncertainty is about 4% for central jets, while for forward jets the maximum is 7%.

4.6.2 Pile-up Tools

As the jet size and $\langle\mu\rangle$ increase, the contamination from pile-up increases. To reduce the number of additional jets from pile-up, several possible corrections can be applied [25, 35, 37] :

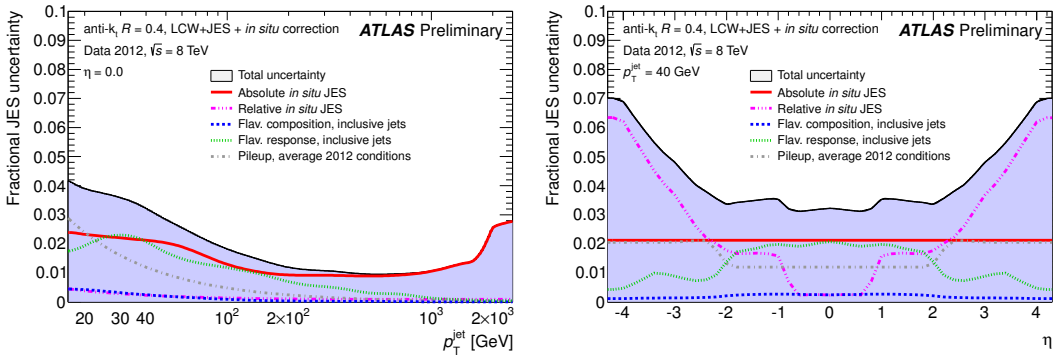
– a cut on the jet vertex fraction (JVF) (see the definition in sec. 4.6.2.3).

– a pile-up correction based on jet area A and momentum density ρ in 2011 events. This correction, as defined in eq. (4.10), is already done for 2012 events.

– Jet Vertex Tagger (JVT), which is a multivariate combination of two track based variables namely corrJVF and R_{pT} [35], where corrJVF is a JVF corrected for the number of vertices, and

$$R_{\mathrm{pT}} = \frac{(\text{scalar } p_{\mathrm{T}} \text{ sum of jet tracks originating from hard scattering vertex})}{p_T^{\mathrm{jet}}}.$$

The hard scatter jet efficiency improvement using JVT compared to JVF is shown in Fig 4-23 for anti-k_T LCW+JES jets ($R = 0.4$) as a function of the the number of primary vertices. JVT was not applied during Run-I. It is used for 2015 data.



(a) (b)

Figure 4-22: JES fractional uncertainty at average pile-up conditions as a function of the jet transverse momentum and pseudorapidity [25].

4.6.2.1 Jet Area

The jet area [1, 25] is measured using "ghost" particles, i.e. particles with extremely soft p_T not affecting the jet clustering. The jet area four-momentum is defined as:

$$A_j = \frac{1}{\nu_g \langle g_t \rangle} \sum_{g_i \in j} g_i \ , \tag{4.13}$$

where $\nu_g \langle g_t \rangle$ is the ghost transverse momentum density and g_i is the four-momentum of ghost particle 'i'. The jet area is the transverse component of A_j, which translates into the scalar area in the limit of small jets. In practice, a constant ghost transverse momentum is chosen and the scalar area is simply the number of ghosts associated to the jet j divided by the ghost number density ν_g. Isolated jets and high p_{T} jets obtained with the anti-k$_T$ algorithm have a circular shape and a surface close to πR^2. The surface of non-isolated jets is smaller as some of their surface is "eaten" by higher p_{T} jets. More details can be found in [1, 25].

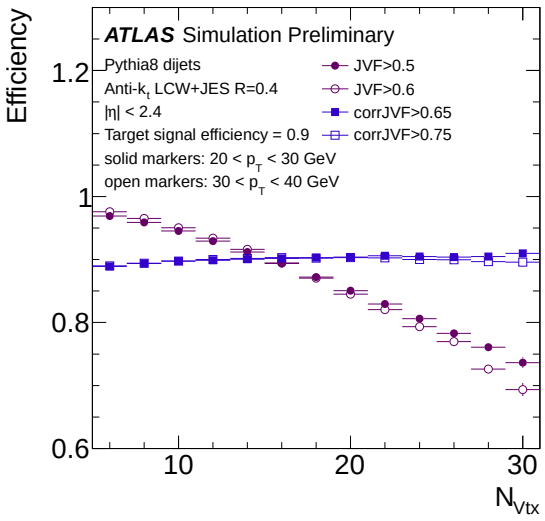


Figure 4-23: The hard scatter jet efficiency using JVT and JVF for anti-k_T LCW+JES jets (R = 0.4), with $20 < p_{\mathrm{T}} < 30$ GeV and $30 < p_{\mathrm{T}} < 40$ GeV, as a function of the number of primary vertices in Pythia dijet samples [35]. The inclusive sample efficiency is 90%.

4.6.2.2 Median p_T Density ρ

The original paper describing the method is [1]. The determination of the amount of transverse momentum per unit area, ρ, added to a hard collision event by pile-up is performed on an event by event basis. It has to be blind to the presence of hard jets. When no lower cut on the jet p_{T} is applied, the k$_T$ jet algorithm (as well as the Cambridge-Aachen algorithm), provides a large number soft pile-up jets with similar properties in p_T/A. This algorithm has clustered a uniform background into jets. The idea is then to estimate ρ as the median $\left\{p_{T,i}^{\mathrm{jet}}/A_i^{\mathrm{jet}}\right\}$, where:

- A_i^{jet} is the jet area.
- $p_{T,i}^{\mathrm{jet}}$ is the jet transverse momentum.
- The index i is associated with all jets in the event reconstructed with k_T algorithm [44, 45], with a distance parameter $R = 0.4$.

This density is independent of the jet algorithm. The jets entering the calculation are built from positive energy topo-clusters with $|\eta| < 2.0$ at both EM and LCW energy scales. Figure 4-24 shows some ρ distributions for several values of the number of primary vertices, N_{PV}, at a fixed level of the instantaneous luminosity ($20 < \langle\mu\rangle < 21$). The results show large fluctuations, which increase with N_{PV}. For more details, see Ref. [25].

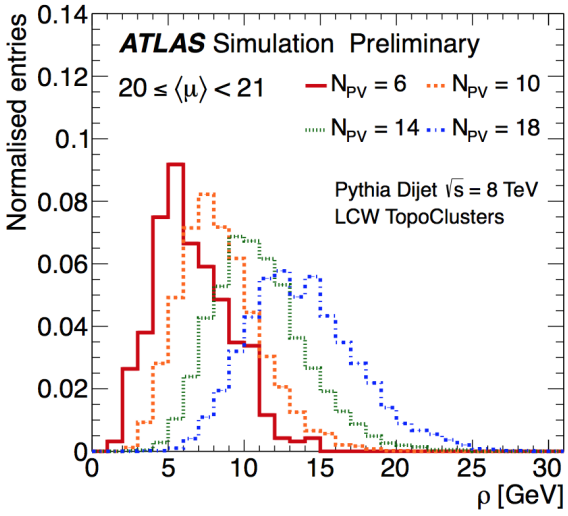


Figure 4-24: Normalized distribution of the median p_T density ρ for several values of the number of primary vertices, N_{PV}, for $20 < \langle\mu\rangle < 21$ [25].

4.6.2.3 Jet Vertex Fraction (JVF)

JVF, as illustrated in Fig. 4-25, is defined as the ratio of the sum of p_{T} of tracks associated to a jet from a PV candidate to the sum of p_{T} of all tracks from any PV in the event:

$$\mathrm{JVF}(jet_a, PV_j) = \frac{\Sigma_k\ p_T(track_k^{\ jet_a}, PV_j)}{\Sigma_n \Sigma_l\ p_T(track_l^{\ jet_a}, PV_n)}\ , \tag{4.14}$$

where k runs over all jet_a tracks originating from PV_j and passing required selection cuts, n runs over all PV in the event and l for the tracks originating from PV_n and matched to jet_a. JVF possible values are : ' 1' for jets with little to no contamination from pile-up, i.e. the hard scatter signal, '< 1' (and positive) for jets with a mixture of hard scatter tracks and tracks originating from pile-up interactions, and '0' for pure pile-up jets where all charged tracks come from pile-up interactions. If a jet has no tracks matched to it e.g. a forward jet, then JVF = -1 [25, 34].

4.6.3 B-tagged Jets

The identification of jets from b-quark decays are important for many LHC studies especially top quark physics and Standard Model (SM) backgrounds in many Higgs analyses. Since B-hadrons have relatively long life-time ($\simeq$ 1.5 ps), which leads to a displaced secondary vertex (as shown in Fig. 4-26) with large impact parameter with respect to the primary vertex, the resulting b-jets produced by the b-quark fragmentation and hadronization can be distinguished from light flavor jets. In order to identify b-jets, several algorithms have been developed based on [11, 13, 27]:

- displaced track properties using transverse and longitudinal track impact parameters [5].

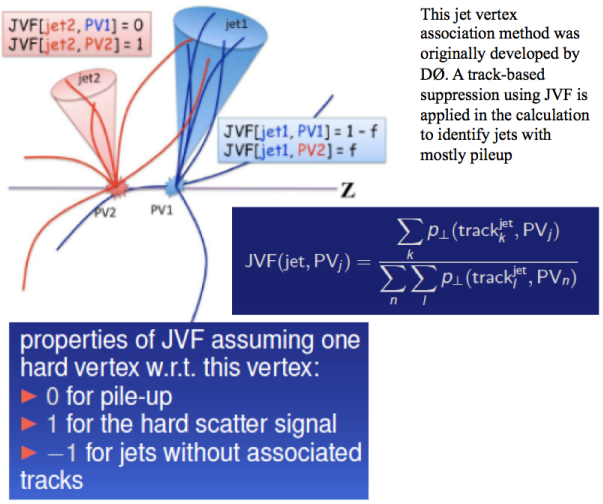


Figure 4-25: A graphical representation of the jet vertex fraction JVF.

- secondary vertex information [7, 11, 13].
- a combination of both track information and secondary vertex properties using multivariate techniques to take into account the correlation amongst variables. One such algorithm, called 'MV1', combines in a neural network the weights provided by some of these algorithms [29].

For Run I analyses in ATLAS, a standard MV1 algorithm [18, 29] was used with various working points corresponding to different inclusive b-tag efficiency in simulated $t\bar{t}$ events: 60%, 70%, 75% and 85%. The higher the b-tag (signal) efficiency, the lower the light flavor (background) rejections as shown in Fig. 4-27(a). These working points are derived using $t\bar{t}$ samples with leptonically decaying top quarks using a combinatorial likelihood method, taking into account correlations among jets in the events, leading to reduced scale factors (SF) [28]. Background events with real isolated leptons include Wt events, Z+jets with $Z \to \tau\tau$ where taus decay leptonically and diboson events where bosons decay leptonically. The b-jet efficiency for MV1 as a function of the jet p_{T} is illustrated in Fig. 4-27(b). Statistical and systematic uncertainties are shown where the latter come mainly from $t\bar{t}$ modeling, hadronization modeling of $t\bar{t}$ events, top p_{T} reweighting, jet experimental uncertainties, and mis-tag uncertainties. The calibration SFs for b- and c- jets MV1 tagging efficiencies at 70% working point are shown in Fig. 4-28 [28, 29].

4.7 Taus

Tau leptons can decay either leptonically ($\tau_\ell \to \ell\overline{\nu_\ell}\nu_\tau$, $\ell = e, \mu$) or hadronically ($\tau_{\mathrm{had}} \to$ hadrons ν_τ). Taus are heavy leptons with mass $m_\tau = 1.77$ GeV and proper decay length of 87 μm: they are the only leptons that decay both leptonically and hadronically. With such a short lifetime, they usually decay inside the beam pipe before reaching the detector active region. Thus, taus are identified based on their

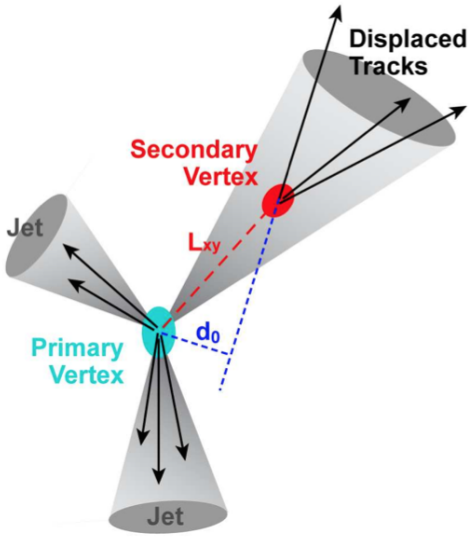


Figure 4-26: Graphical representation of secondary vertex reconstruction: L_{xy} is the distance separating the secondary vertex from the primary vertex in the transverse plane, i.e. plane orthogonal to beam axis and d_0 is the track impact parameter in the transverse plane [40].

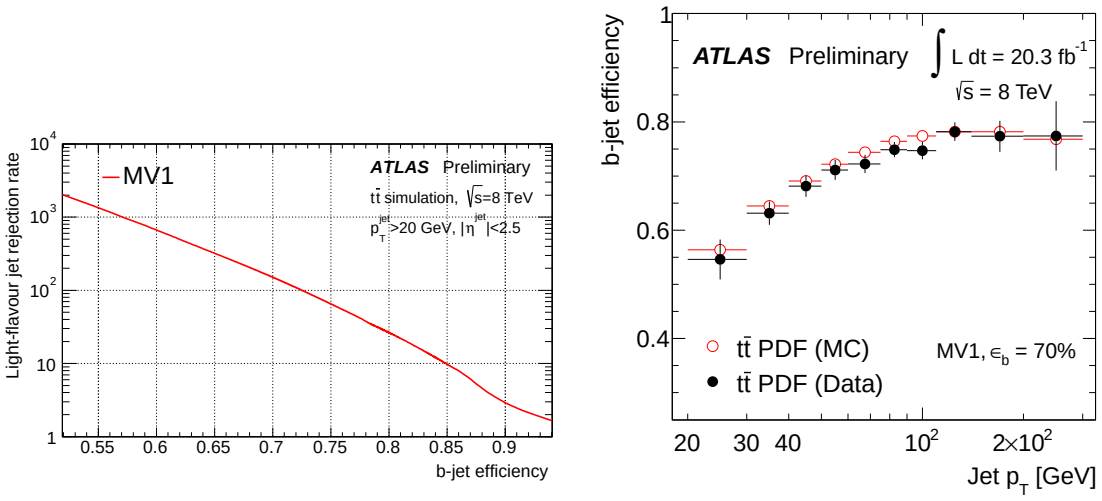
decay products. Visible leptonic tau decay products, i.e. electrons and muons, are treated like primary electrons and muons. Only hadronic tau decays with branching ratio BR = 65% are considered when discussing tau reconstruction and identification in this section. In the hadronic mode, tau decay products include one or three charged tracks in 72% and 22% of the cases respectively. Therefore, taus are classified based on the number of associated tracks, or 'prongs'. The challenge is to identify hadronic tau decays amidst a much larger production of hadron jets. The major background to hadronic tau decays comes from the jets of energetic hadrons. Figure 4-29 shows the composition of a typical hadronically decaying tau cone and a QCD jet cone. A quick comparison between hadronic tau τ_{had} decays and QCD jet cones is given in table 4.7.

Typical τ_{had} decay	Typical QCD jet
Collimated decay products	Typically wider
Mostly pions: $\pi^{+/-}$ and π^0	Contains all sorts of hadrons
Few $\pi^{+/-}$ (1 or 3) and π^0(0 or 2)	Mutliplicities increasing with jet energy

Table 4.7: τ_{had} and QCD jet cone comparison.

4.7.1 Tau Trigger Operations

The visible hadronic tau decay products i.e. neutral and charged hadrons are referred to as $\tau_{\text{had-vis}}$. Various tau triggers have been implemented to accommodate different



(a) (b)

Figure 4-27: MV1 algorithm light flavor rejection efficiency as a function of b-jet tagging efficiency(a) [29]. The b-jet efficiency p_{T} dependence is shown in (b) [28] for the MV1 algorithm at 70% working point.

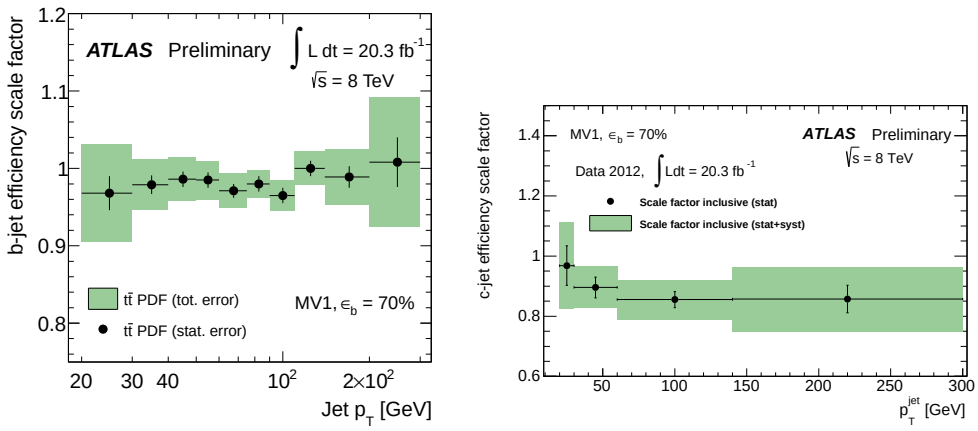
physics processes and the associated kinematic requirements on reconstructed objects [32]. Figure 4-30 shows the tau trigger rate at L1 and EF[3] levels in terms of instantaneous luminosity at 8 TeV. The dependence is linear everywhere except for the $\tau_{\text{had-vis}} + E_T^{miss}$ trigger at high luminosity values due to high pile up effects on the missing transverse energy reconstruction in the event. The rates for double hadronic triggers are higher than those of the single hadronic trigger, since the threshold for the former (29 and 20 GeV) is much lower than that of the shown single tau trigger $p_T(\tau) > 115$ GeV. Applying the isolation and full identification requirements for the $\tau_{\text{had-vis}}$ in the double hadronic triggers at the EF level varies depending on the physics process considered (e.g. Standard Model Higgs, H_{SUSY}). Since the $\tau_{\text{had-vis}}$ background rejection rate is smaller than that of the electron/muon background rejection, tau trigger p_{T} requirement is higher than the electron/muon ones. Analyses requesting low-p_{T} $\tau_{\text{had-vis}}$ in 2012 have to rely on combined tau triggers (combining tau trigger requirements with those of other objects) based on lower tau p_{T} threshold. In the $H \to \tau^+\tau^-$ analysis studied in the leptonic-hadronic mode, only single lepton trigger (SLT) events were used. Taus coming from the combined tau triggers are not considered since the estimation of the fake taus for the analysis in that case is not feasible given the poor statistics.

4.7.2 Tau Reconstruction

The basic τ lepton reconstruction steps in ATLAS can be summarized as follows:

- $\tau_{\text{had-vis}}$ reconstruction algorithm is seeded with jets (see 4.6) satisfying the fol-

[3]Trigger levels L1 and EF(event filter) are described in chapter 2.



(a) (b)

Figure 4-28: (a): b-jet efficiency scale factors for MV1 b-tagging algorithm at 70% working efficiency in $t\bar{t}$ samples [28]. The green band shows total uncertainty (statistical+systematic). The scale factors to correct the efficiency with which the b-tagging algorithms identifies c-jets is shown in (b) [29].

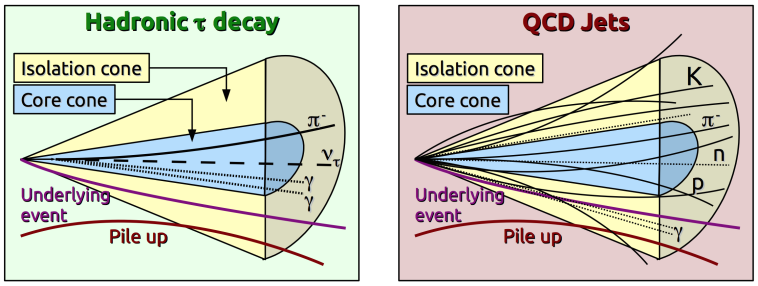


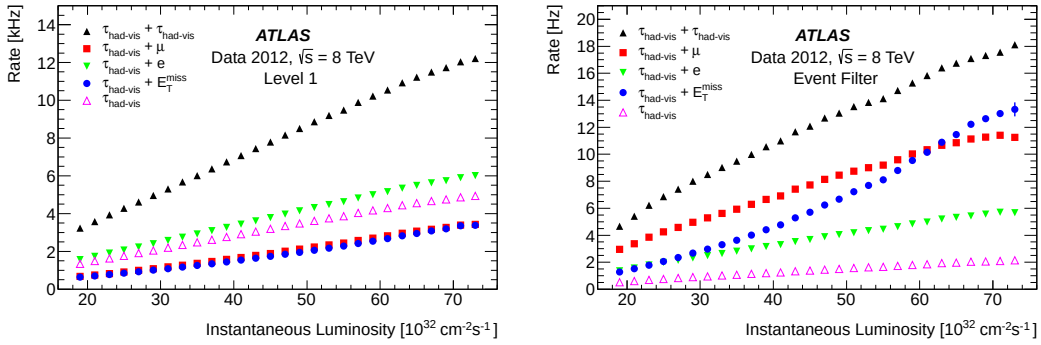
Figure 4-29: Diagram showing hadronically decaying tau and QCD jet cones.

lowing conditions:

1. reconstructed from the LC calibrated topo-clusters using the anti-k_T algorithm with distance parameter R = 0.4.
2. having $p_T > 10$ GeV (after calibration) and $|\eta| \leq 2.5$.

Events are required to have a primary vertex (PV) with at least 3 associated tracks. Given the high pile-up conditions, the vertex corresponding to the tau lepton may not necessarily coincide with the event PV chosen as the vertex with the highest Σp_T^2 amongst all PV candidates. In fact, one observes that in $Z \to \tau^+\tau^-$ events, under the same conditions of pile-up, the tau vertex and the event PV coincide with each other about 90% of the time. With a large number of interactions per crossing, a pile-up vertex is sometimes chosen as the event primary vertex. This wrong choice leads to tracks failing to pass the z_0 requirement and a degraded tau track selection efficiency [21, 22, 32].

Therefore, the tau vertex (TV) needs to be identified in order to reduce pile-up



(a) (b)

Figure 4-30: Tau trigger rates for 2012 data as a function of the instantaneous luminosity: Level-1 trigger L1(a) and Event Filter(EF) in (b) [32].

and noise effects, and increase the reconstruction efficiency. This is done using the tau jet vertex association (TJVA).
TJVA uses the '*Tau Jet Vertex Fraction*' (f_{TJVF}), which is JVF (defined in section 4.6.2.3) with modified track selection cuts. Tracks are required to pass the same cuts as tracks associated to taus (detailed hereafter) with the conditions on the impact parameters being dropped. The primary vertex candidate having the largest f_{TJVF} is chosen as the TV [21, 22, 32]. This vertex is needed to define the tau direction and identify the tracks associated to the τ candidate. The track selection efficiency as a function of the average number of interactions per crossing for 1-prong taus (taus with 1 associated track) in $Z \to \tau^+\tau^-$ MC events is shown in Fig. 4-31, where the efficiency is defined as the probability to match a real pion from a hadronic tau decay with a reconstructed $\tau_{\text{had-vis}}$ candidate.

- The barycenter of the jet seed topo-cluster calibrated at the local hadronic calibration (LC) scale is used to get the $\tau_{\text{had-vis}}$ 3-momentum components [22].
- All clusters within a cone of $\Delta R = 0.2$ of the barycenter are collected.
- The 4-momenta of the cluster and cluster position with respect to the tau vertex are recalculated to obtain the momentum p^{LC} and $\tau_{\text{had-vis}}$ direction. We have now a tau candidate with its clusters.
- Tracks are assigned to $\tau_{\text{had-vis}}$ if:
  1. they are within the core region $\Delta R < 0.2$.
  2. they have $p_{\text{T}} > 1$ GeV.
  3. they have ≥ 2 hits in the pixel detector and ≥ 7 hits in both pixel and SCT layers.
  4. their impact parameters defined with respect to TV satisfy the conditions: $|d_0| < 1.0$ mm and $|z_0\ sin(\theta)| < 1.5$ mm, where d_0 and z_0 denote

the distance of closest approach of the track to the TV in the transverse and longitudinal planes respectively.

- Tracks falling in the isolation annulus ($0.2 < \Delta R < 0.4$) and satisfying the criteria mentioned above are used to evaluate the isolation of the candidate (Fig. 4-32).

Based on the number of associated tracks, a tau candidate is classified as a 1-prong or 3-prong tau. Candidates with other track multiplicities are not considered.

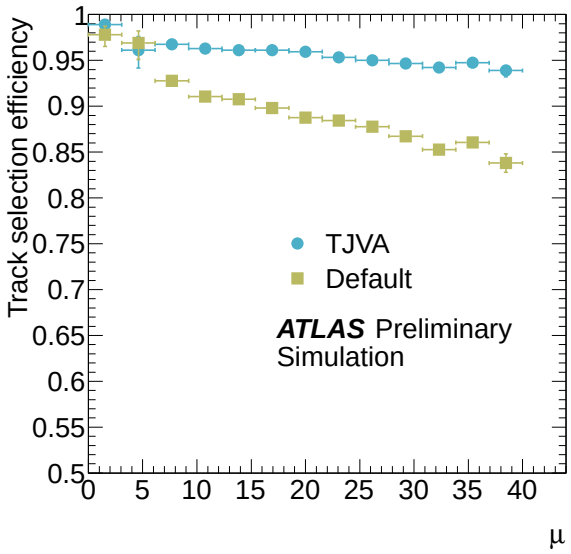


Figure 4-31: $\tau_{\text{had-vis}}$ track efficiency as a function of the average number of interactions per crossing for 1-prong taus in $Z \to \tau^+\tau^-$ events: the $\tau_{\text{had-vis}}$ candidates are required to have $p_T^\tau > 15$ GeV and match truth-tau within a distance $\Delta R < 0.2$. With TJVA, the efficiency is less pile-up sensitive with a small degradation at high pile-up values [21].

A π^0 algorithm is also applied to find the number of neutral pions in the tau core region using the calorimeter and tracking information. A π^0 is matched with up to two clusters in the core cone. The matched clusters are assigned a π^0 likeliness score based on the cluster properties, after removing the contribution of pile-up, noise and the underlying event as estimated in the isolation region. The clusters with the highest likeliness scores are used, and provide along with the tracks in the core region the input for the tau identification variables.
After reconstruction, tau identification is done using Boosted Decision Trees (BDTs) and likelihood methods, which provide 3 different working points: loose, medium and tight for τ selection. Identification variables formed during reconstruction based on tracking and calorimeter variables, are combined into multivariate discriminants against contamination from jets and leptons (e, μ). Then, preselection criteria can be applied at the analysis level depending on the needed level of signal efficiency and background rejection [32].

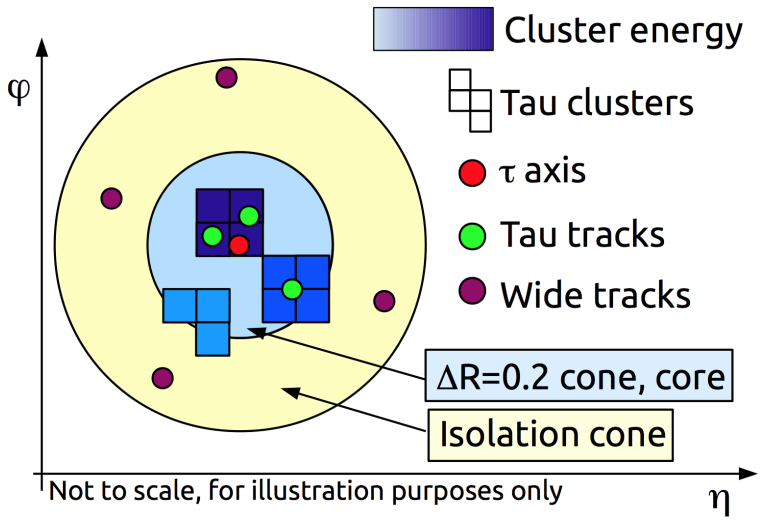


Figure 4-32: Graphical representation of tau reconstruction steps summarized.

4.7.3 Discriminating Variables

The rejection efficiency of jets and leptons faking taus at the reconstruction level is poor. Consequently, the *tau* signal-background discrimination is done through tau identification (tau ID) algorithms. Background rejection variables entering these algorithms include the invariant mass in topological clusters, the tracks invariant mass, the leading track momentum, the fraction of energy deposited in EM calorimeter f_{EM}, the energy fraction in the core cone and the number of tracks in the isolation annulus ($0.2 < \Delta R < 0.4$) amongst others. Variables with high discriminating power against background from jets are described in the next section. A detailed summary of all the variables used in tau ID algorithm is given in [22, 32].

4.7.4 Discrimination Against Jets

τ-jet discrimination is done using boosted decision trees (BDTs) and log-likelihood requirements, defined each for 1-prong and 3-prong taus. Some important powerful discriminating variables are [22, 32] :

- **Shower shape variables in the calorimeter :** QCD jets are wider than typical τ_{had}. The calorimeter shower shape is given by the core energy fraction f_{core} defined by :

$$f_{\mathrm{core}} = \frac{\Sigma_{i \in all}^{\Delta R_i < 0.1} E_{T,i}^{EM}}{\Sigma_{j \in all}^{\Delta R_j < 0.2} E_{T,j}^{EM}} \ , \tag{4.15}$$

  where $E_{T,i(j)}^{EM}$ is the transverse energy calibrated at the EM energy scale in cell i(j). The index *i(j)* runs over all calorimeter cells matched to the $\tau_{\text{had-vis}}$ in the central region ($\Delta R < 0.1$) (core region ($\Delta R < 0.2$)).

  This is the fraction of transverse energy in the central region ($\Delta R < 0.1$) with respect to the energy deposited in the core cone $\Delta R < 0.2$. The pile-up corrected

version removing pile-up bias based on the number of good vertices in the event is given by $f_{\text{core}}^{\text{corr}}$, also referred to as central energy fraction f_{cent}. The pile-up corrected fraction $f_{\text{core}}^{\text{corr}}$ enters tau ID calculations.

$$\begin{aligned} f_{\text{cent}} = f_{\text{core}}^{\text{corr}} &= f_{\text{core}} + 0.003 * N_{vtx} \quad \text{for } p_{\text{T}} < 80 \text{ GeV}. \\ &= f_{\text{core}} \qquad\qquad\qquad\qquad \text{otherwise}. \end{aligned}$$

where N_{vtx} is the number of good vertices (pile-up vertices with at least 2 tracks and primary vertex with at least 4 tracks) in the event, and p_T is the $\tau_{\text{had,vis}}$ momentum calibrated at the tau energy scale (TES).

- **Number of tracks in the isolation annulus** ($N_{\textbf{track}}^{\textbf{iso}}$)**:** This is the number of tracks within $0.2 < \Delta R < 0.4$ associated to the $\tau_{\text{had,vis}}$ candidate.

- **Average p_{T}-weighted track distance from the tau axis** $R_{\textbf{track}}$: All tracks within the core and isolation region, i.e. within $\Delta R < 0.4$, are included. For 1-prong taus, this is just the distance ΔR separating the track from the tau axis.

$$R_{\text{track}} = \frac{\Sigma_i^{\Delta R_i \leq 0.4} p_{T,i} \Delta R_i}{\Sigma_i^{\Delta R_i \leq 0.4} p_{T,i}} \ , \tag{4.16}$$

  where i runs over all tracks within $\Delta R < 0.4$.

- **The maximal distance** ΔR **(**$\Delta R_{\textbf{Max}}$**):** between a track associated with a $\tau_{\text{had,vis}}$ candidate and the tau axis. It is used for multi-prong taus and only core tracks are taken into account.

- **Transverse flight path significance** S_T^{flight}: This is the significance of the decay length of the secondary vertex in the transverse plane, i.e. the ratio of decay length of the secondary vertex with respect to the TV to its uncertainty. The secondary vertex is defined as the reconstructed vertex obtained using core tracks. It is applied for multi-prong $\tau_{\text{had,vis}}$ candidates.

$$S_T^{\text{flight}} = \frac{L_T^{\text{flight}}}{\delta L_T^{\text{flight}}} \ . \tag{4.17}$$

Figure 4-33 shows the normalized distribution of some jet discriminating variables (f_{cent} in (a), $N_{\text{track}}^{\text{iso}}$ in (b) and R_{track} in (c)) in both signal samples ($Z, Z' \to \tau^+\tau^-$ and $W \to \tau\nu$) and background samples obtained using 2012 data. The signal and background peaks are well separated in the f_{cent} and R_{track} distributions. The normalized $N_{\text{track}}^{\text{iso}}$ distribution in Fig. 4-33-(b) for 1-prong taus shows that about 90% of the signal events have no tracks in the isolation cone as expected, while background events have a varied track multiplicity in the isolation cone. Finally, to summarize, Fig. 4-33-(d) shows the inverse background efficiency as a function of the signal efficiency for high-p_{T} taus ($p_T > 40$ GeV). The signal efficiency is defined as the fraction of $\tau_{\text{had,vis}}$ candidates matched to true hadronic taus passing the identification requirements, whereas the background efficiency is defined as the number of background

$\tau_{\text{had,vis}}$ candidates passing tau ID requirements divided by the total number of background $\tau_{\text{had,vis}}$ candidates. Background rejection factors of 10-30 are reached at 70% signal efficiency, and going up to 500 at 35% signal efficiency [32].

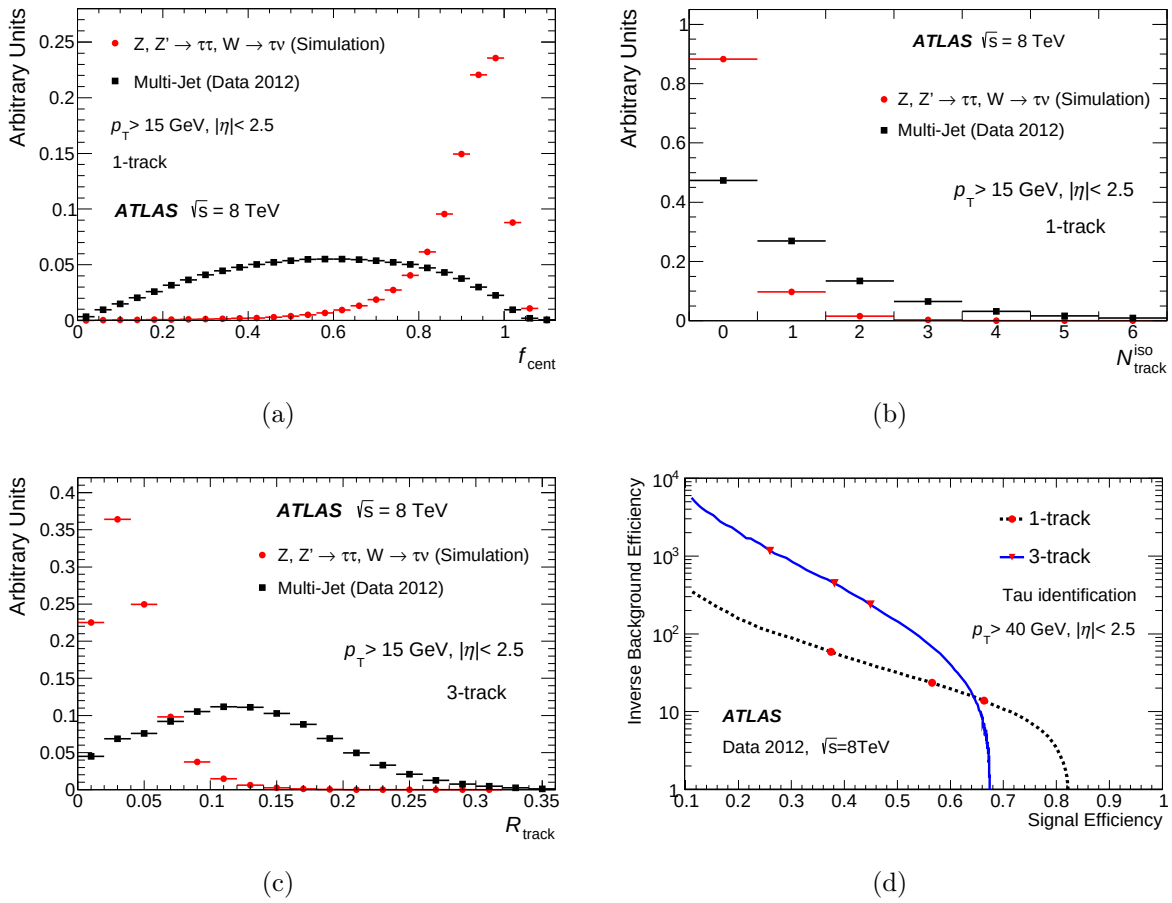
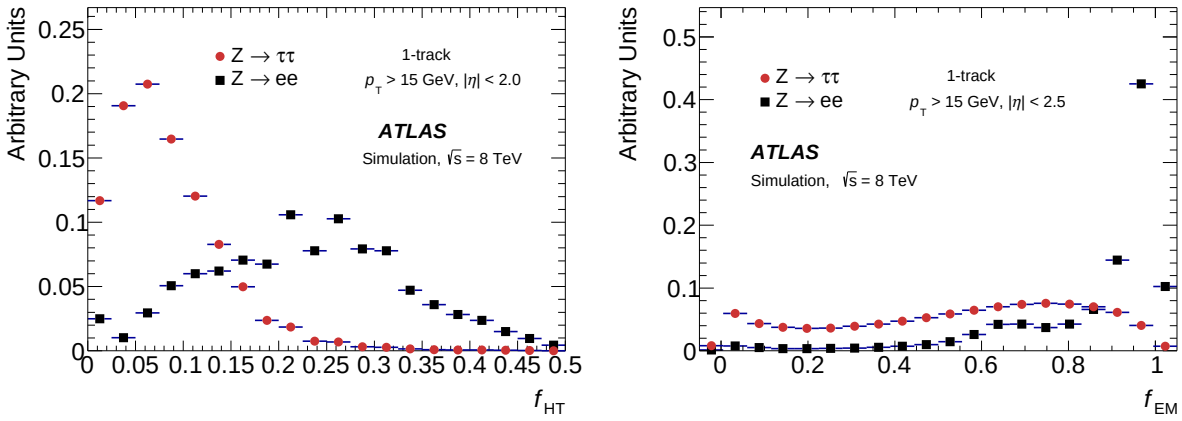


Figure 4-33: Normalized distributions of f_{cent} in (a), $N^{\text{iso}}_{\text{track}}$ in (b) and R_{track} in (c) in both signal samples ($Z, Z' \to \tau^+\tau^-$ and $W \to \tau\nu$) and background samples obtained using 2012 data. The inverse background efficiency as a function of the signal efficiency is plotted in (d) [32].

4.7.5 Discrimination Against Electrons

For 1-prong taus, even after the rejection of background events coming from jets, a significant background contribution coming from electrons still needs to be suppressed. This is achieved using the 'electron veto' algorithm. Powerful discriminating variables used to separate electrons faking τ_h from $\tau_{\text{had,vis}}$ candidates are based on the following [32]:

- transition radiation: electrons are lighter than pions and have therefore higher Lorentz factors. They are more likely thus to emit transition radiation, leading



(a) (b)

Figure 4-34: Distributions of f_{HT} and f_{EM} in simulated $Z \to \tau^+\tau^-$ and $Z \to ee$ samples [32].

to high threshold TRT hits. Consequently, electrons have a higher TRT HT fraction f_{HT}, which is the ratio of high-threshold to low-threshold hits (outliers (as defined in section 4.1) included).

- R_{track} (see section 4.7.4).
- fraction of energy in EM calorimeter f_{EM} with respect to the total calorimetric (EM+hadronic) energy deposits.
- amount of energy deposited in the hadronic calorimeter.
- f_{cent} (defined in section 4.7.4).
- shower depth (electrons have shorter and narrower showers than the decay products of τ_{had}).

A BDT based on the discriminating variables is used for the electron-$\tau_{\mathrm{had,vis}}$ separation. The complete list of variables entering the e-veto algorithm is given in [22]. Figure 4-34 shows the distributions of f_{HT} and f_{EM} in simulated $Z \to \tau^+\tau^-$ and $Z \to ee$ samples, where $\tau_{\mathrm{had,vis}}$ are required to match the true $\tau_{\mathrm{had,vis}}$ at the generator level. $Z \to ee$ background events are chosen such that $\tau_{\mathrm{had,vis}}$ decay products are matched to the true electrons produced at the generator level. To summarize, the e-veto efficiency is shown in Fig. 4-35. Three working points are defined, labeled tight, medium and loose corresponding respectively to signal efficiencies of 75%, 85% and 95%.

4.7.6 Discrimination Against Muons

The energy deposits of muons, minimum ionizing particles, in the calorimeter are usually smaller than those required for $\tau_{\mathrm{had,vis}}$ reconstruction. However, if a muon is

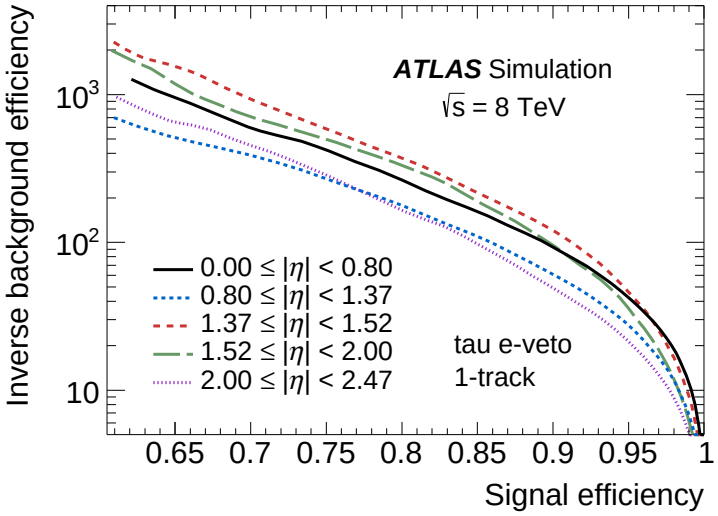


Figure 4-35: Inverse background efficiency($Z \to ee$) as a function of the signal($Z \to \tau^+\tau^-$) efficiency for e-veto for various η regions [32].

associated with a calorimeter cluster with enough energy deposits, it may incorrectly be reconstructed as a $\tau_{\text{had,vis}}$ candidate. The possible scenarios are:

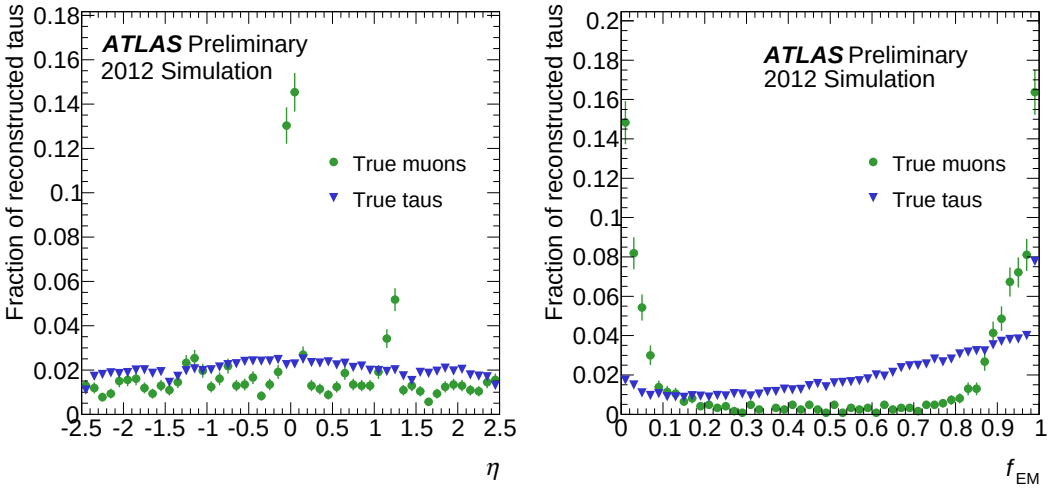
1. Overlap of a reconstructed muon and $\tau_{\text{had,vis}}$ candidate: in this case, the $\tau_{\text{had,vis}}$ is discarded.

2. The muon passed through regions in the muon spectrometer (MS) with poor (near $\eta \sim 1.2$) or no sensitivity (in the central region of the detector near $\eta = 0$), and fails consequently the MS reconstruction (see section 4.5 for details). The fake rate in these regions is the highest.

3. A high-p_{T} muon has large energy deposits in the calorimeter (muon bremsstrahlung), mainly the hadronic calorimeter. Therefore, compared to usual $\tau_{\text{had,vis}}$ matched to true τ_{had}, it has a very small f_{EM} and a large track-p_{T} to E_T ratio. Such muons have a strongly bent trajectory and fail the MS reconstruction.

4. A muon, with p_{T} low enough to be stopped in the calorimeter and not reach the spectrometer, overlaps with a calorimetric cluster energetic enough to meet $\tau_{\text{had,vis}}$ reconstruction requirements. In this case, the $\tau_{\text{had,vis}}$ candidate has a small f_{EM} value with low p_{T} to E_{T} ratio.

Fake muon rejection is done using a simple cut-based selection, using the two variables f_{EM} and the track-p_{T} to E_T ratio. This 'muon veto' algorithm has an efficiency better than 96%, reducing the muon fakes by 40% [22, 32]. Figure 4-36 shows the pseudorapidity η and f_{EM} normalized distributions in reconstructed $\tau_{\text{had,vis}}$ matched to true muons and true taus.



(a) (b)

Figure 4-36: Pseudorapidity η (a) and f_{EM} (b) normalized distributions in reconstructed $\tau_{\mathrm{had,vis}}$ matched to true muons and true taus. $\tau_{\mathrm{had,vis}}$ candidates overlapping with reconstructed muons are discarded [22].

4.7.7 Performance

The signal identification efficiency for 1-prong and 3-prong taus as a function of the true visible tau p_{T} for various BDT working points is shown in Fig. 4-37, where the signal efficiency is defined as the fraction of true $\tau_{\mathrm{had\text{-}vis}}$ decays reconstructed with n decay products using n tracks that pass the tau identification criteria. The BDT requirements are selected such that there is no efficiency dependence on the $\tau_{\mathrm{had\text{-}vis}}$ p_{T}. The dip seen at 80 GeV is due to variable calculation changes, in $f_{\mathrm{core}}^{\mathrm{corr}}$ mainly (see section 4.7.4). In addition, the signal efficiency is stable as a function of the number of vertices as shown in Fig. 4-38. Background efficiencies are summarized as well in Fig. 4-37, where the background efficiency is defined as the number of background $\tau_{\mathrm{had,vis}}$ candidates passing tau ID requirements divided by the total number of background $\tau_{\mathrm{had,vis}}$ candidates.

To correct for data/MC differences in $\tau_{\mathrm{had,vis}}$ reconstruction efficiencies, scale factors are applied to simulated MC samples, where a scale factor is defined as the ratio of data to MC efficiencies: $\mathrm{SF} = \epsilon_{\mathrm{data}}/\epsilon_{\mathrm{MC}}$. In order to extract SFs, tau reconstruction efficiency was studied directly in data and in MC simulated samples using tag and probe method [22, 32]. The MC measurements are done using $Z \to \tau^+\tau^-$, $W \to \tau\nu$, and $t\bar{t} \to \tau$ + jets samples independently. The SFs obtained from all these channels are compatible with each other within the associated uncertainties. The $Z \to \tau^+\tau^-$ SFs have the smallest uncertainties and provide the main measurements, while the $W \to \tau\nu$ results are used as cross checks. The values of these SFs, shown in Fig. 4-39 for signal samples, are close to unity except those of *tight* 1-prong taus in the endcap region, where the SF is about 0.9. The SFs thus obtained have an uncertainty

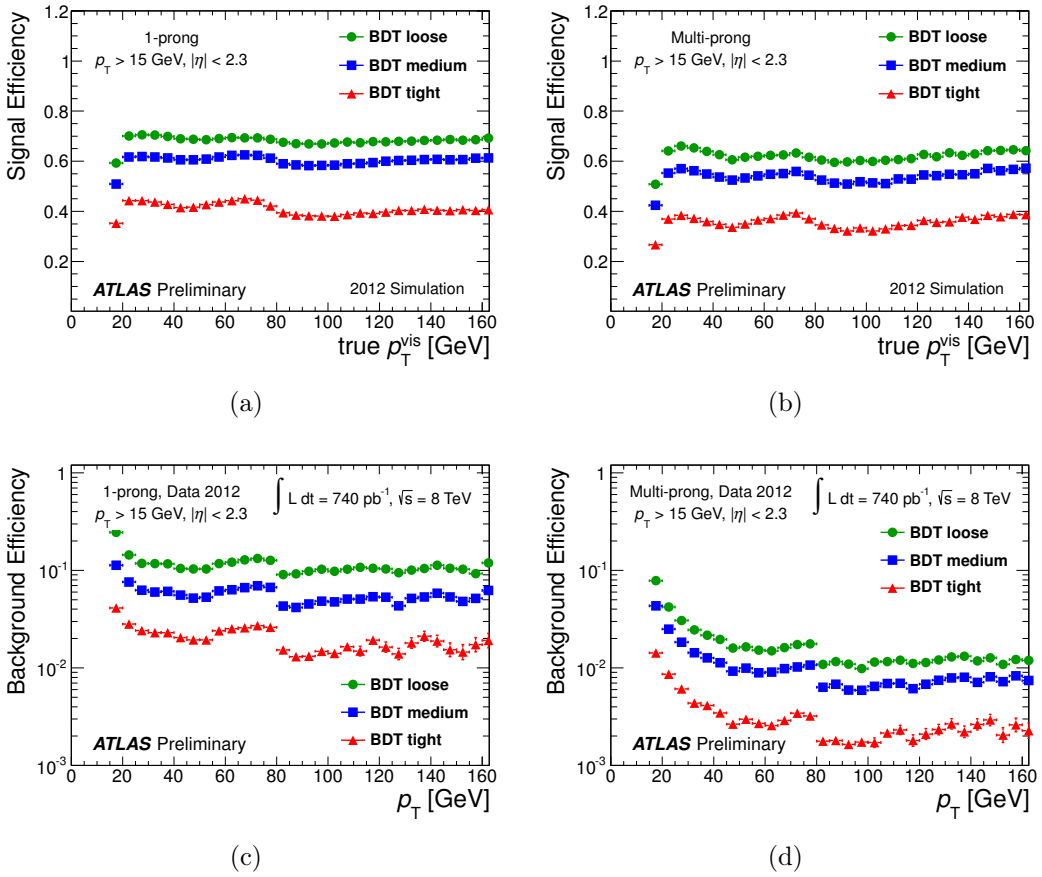
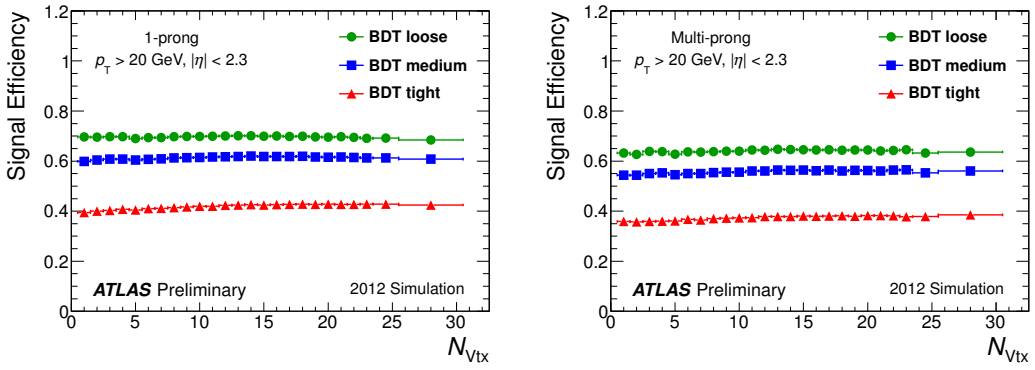


Figure 4-37: $\tau_{\text{had-vis}}$ signal (top) and background (bottom) efficiency for 1-prong (left) and 3-prong(right) taus for the different BDT working points for 2012 data, as a function of the true visible tau p_{T} for signal and reconstructed p_{T} for background (QCD from data) [22].

of 2-3% in inclusive samples in the most sensitive channel ('1-track'). The SFs for background events are however analysis-dependent (kinematics, final state, etc). The SF calculation is also done for various combinations of the electron and muon vetoes working points. Finally, uncertainties on the tau identification efficiency, tau trigger identification efficiency and tau energy scale are summarized in Table 4.8.

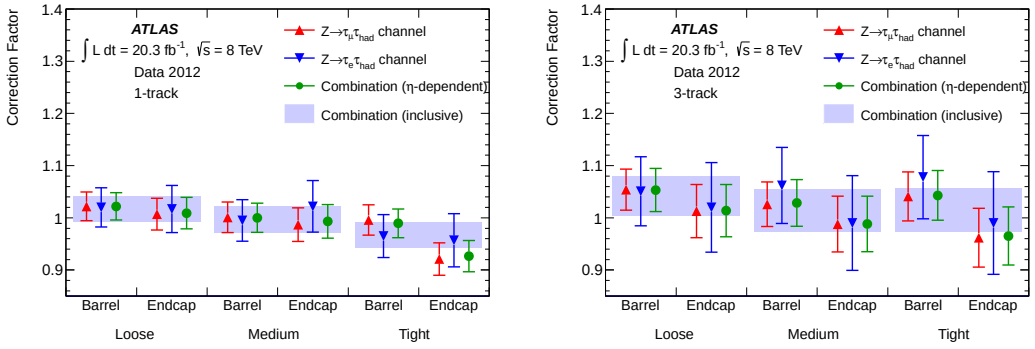
4.7.8 Energy Scale

The jet energy scale (JES) cannot be used to correct for the τ energy since τ-leptons have their specific particle content. The tau energy scale (TES) is applied to correct for the momentum scale: the pile-up contribution is subtracted first from the reconstructed τ momentum then it is scaled to the true momentum. The uncertainty on this is below 4%. This is studied under various conditions: studying pile-up behavior, E/p measurements, behavior in bins of p_T^{τ}, varying underlying event model, etc. in a similar way to the jet energy scale (JES) calculations discussed in section 4.6.



(a) (b)

Figure 4-38: $\tau_{\text{had-vis}}$ signal efficiency as a function of the number of the vertices in the event for 1-prong (a) and 3-prong (b) taus [22].



(a) (b)

Figure 4-39: Tau SFs for all tau identification working points combined as a function of η for the muon and electron channels separately and together. The results are derived separately for 1-prong taus (a) and 3-prong taus (b) with $p_{\text{T}} > 20$ GeV. The error bars include combined statistical and systematic uncertainties [32].

4.8 Missing Transverse Momentum

The exact initial energy carried by the partons in the main hard scattering event in the proton-proton collision cannot be known. Hence, we cannot benefit from energy-momentum conservation for calculations in the longitudinal direction along the beam axis. Given that initial transverse momentum of the colliding partons is negligible, the vector sum of the momenta of the particles in the final state in the transverse plane is expected to be zero and momentum conservation can be applied $\Sigma\overrightarrow{p_T} = 0$. Some particles may escape the detector without being measured, leaving no tracks nor any energy deposit like, for example, the Standard Model neutrinos and the light-

	Uncertainties
Tau ID efficiency	2-3% for 1-prong τ 4-5% for 3-prong τ
Tau trigger identification efficiency	2-8%: 2% for $p_{\mathrm{T}} \in$[30-50]GeV 8% for p_{T}=100GeV
Tau energy scale(TES) measurement	2-4%
medium BDT tau measurement (used in most analyses)	Total(statistical+systematic): 2.5%(1-prong) 4% (3-prong)

Table 4.8: Uncertainties on some basic tau performance measurements [32].

est supersymmetric particle. This leads to true missing transverse momentum that can be estimated based on the transverse momentum of the well reconstructed particles. The missing transverse momentum is also referred to as missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ (or MET) defined as: $E_{\mathrm{T}}^{\mathrm{miss}} = p_{\mathrm{T}}^{\mathrm{miss}} = |\mathbf{p}_{\mathrm{T}}^{\mathrm{miss}}|$, where $\mathbf{p}_{\mathrm{T}}^{\mathrm{miss}}$ is the negative vector sum of the transverse momentum of all the visible measured particles $\mathbf{p}_{\mathrm{T}}^{\mathrm{miss}} = -\Sigma\overrightarrow{p_T}^{reco}$. Missing transverse momentum can also arise due to limitations on detector efficiencies, poor resolutions in some cases, mis-reconstructed objects, dead regions in the detector, cosmic rays and noise sources, leading to an imbalance in the transverse momentum in the event. The latter is a fake $E_{\mathrm{T}}^{\mathrm{miss}}$.

In ATLAS, two generic $E_{\mathrm{T}}^{\mathrm{miss}}$ types were used in Run-I, where the values of the reconstructed objects transverse momenta values are taken from either track-based or calorimeter-based measurements. The calorimeter and track based MET basic definitions are orthogonal and give complementary measurements. From the latter, various MET versions (sometimes called flavors) encountered in analyses are derived, combining both tracking and calorimeter measurements for a better performance. In particular, the default track-based MET and its improved versions are discussed in detail in chapter 5, and this section will focus mainly on calorimeter-based measurements. The definitions of all the MET flavors used in the analyses presented are given at the end of this section.

4.8.1 Hard and Soft Terms Definitions

The $E_{\mathrm{T}}^{\mathrm{miss}}$ algorithm is performed after the reconstruction of the particles (electrons, photons, muons, jets and taus), with energy calibration and pile-up suppression applied to each particle. One distinguishes two classes of contributions in any $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation [34], namely hard terms and soft terms :

- **Hard Terms:** This refers to the contribution from "hard" reconstructed physics objects, i.e. particles and/or jets in the event, coming from the hard scatter interaction. Particles and jets passing kinematic thresholds (e.g. p_{T} threshold)

and quality cut requirements are called "hard" physics objects, and they are added to $E_{\mathrm{T}}^{\mathrm{miss}}$ calculations based on the associated reconstruction efficiency and precision: physics objects with the highest precision are added first and those with the least precision are added last. That's why the electron terms are added before the jet terms for example.

- **Soft Terms:** The soft terms receive all contributions (either tracks p_{T} and/or energy deposits in calorimeter cells) from:
  - particles in the underlying event
  - diffusely scattered particles from pile-up interactions
  - reconstructed physics objects failing the "hard" term quality cuts or kinematic thresholds (e.g. soft jets) [24, 34]
  - low-p_{T} tracks to recover particles with low momentum not reaching the calorimeter

To summarize, with ΣE_T being the scalar sum of transverse momenta:

$$E_{\mathrm{x(y)}}^{\mathrm{miss}} = E_{\mathrm{x(y)}}^{\mathrm{miss,HardTerm}} + E_{\mathrm{x(y)}}^{\mathrm{miss,SoftTerm}} = -\sum_{\text{hard objects}} p_{\mathrm{x(y)}} - \sum_{\text{soft signals}} p_{\mathrm{x(y)}} \ . \tag{4.18}$$

$$\Sigma E_T = \Sigma E_T^{\mathrm{HardTerm}} + \Sigma E_T^{\mathrm{SoftTerm}} = \sum_{\text{hard objects}} p_T + \sum_{\text{soft signals}} p_T \ . \tag{4.19}$$

4.8.2 $E_{\mathrm{T}}^{\mathrm{miss}}$ Reconstruction

In this section, the basic calorimeter based MET definition, i.e. MET_RefFinal[4] from which many MET flavors can be defined, will be discussed.
MET_RefFinal reconstruction uses energy deposits in the calorimeter along with reconstructed muons in the spectrometer. ST muons are also included to take into account low-p_{T} muons. To recover the missed low-p_{T} contribution from particles not having enough energy to reach the calorimeters, low-p_{T} tracks can be added. When a low-p_{T} track is matched to a calorimeter cluster, the momentum information is taken from the track, since tracks have better momentum resolution than calorimeter clusters at low-p_{T}. The calorimeter energy deposits are associated with a "hard" physics object, as defined in sec. 4.8.1, in the following order: electrons, photons, hadronically decaying tau leptons, jets and muons. If there is no match between a reconstructed physics object and the calorimeter energy deposits, the latter are still included under the *soft term* contribution to MET. In short:

$$E_{\mathrm{x(y)}}^{\mathrm{miss}} = E_{\mathrm{x(y)}}^{\mathrm{miss,e}} + E_{\mathrm{x(y)}}^{\mathrm{miss},\gamma} + E_{\mathrm{x(y)}}^{\mathrm{miss},\tau_{\mathrm{h}}} + E_{\mathrm{x(y)}}^{\mathrm{miss,jets}} + E_{\mathrm{x(y)}}^{\mathrm{miss,SoftTerm}} + E_{\mathrm{x(y)}}^{\mathrm{miss},\mu} \ , \tag{4.20}$$

$$E_{\mathrm{x(y)}}^{\mathrm{miss,Physics\ object}} = -\Sigma p_{\mathrm{x(y)}}^{\mathrm{Physics\ object}} \ , \tag{4.21}$$

[4] RefFinal refers to refined calibration of the calorimeter cell measurements based on the physics objects to which they are associated.

$$\boldsymbol{E_T^{\mathrm{miss}}} = (E_{\mathrm{x}}^{\mathrm{miss}}, E_{\mathrm{y}}^{\mathrm{miss}}) \ , \tag{4.22}$$

$$E_T^{\mathrm{miss}} = |\mathbf{E}_T^{\mathrm{miss}}| = \sqrt{(E_{\mathrm{x}}^{\mathrm{miss}})^2 + (E_{\mathrm{y}}^{\mathrm{miss}})^2} \ , \tag{4.23}$$

$$\sum E_T = E_{\mathrm{T}}^{\mathrm{e}} + E_{\mathrm{T}}^{\gamma} + E_{\mathrm{T}}^{\tau} + E_{\mathrm{T}}^{\mathrm{jets}} + E_{\mathrm{T}}^{\mathrm{SoftTerm}} \ . \tag{4.24}$$

where:

- 'physics object' = e, γ, τ_h, jet and μ.
- $\sum E_T$ is the scalar sum of the transverse energy E_T of each term contributing to $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation.

The MET calculation, as defined in eq.(4.24), requires standard ATLAS calibrated electrons, i.e. 'medium' (see sec. 4.4) electrons with $p_{\mathrm{T}} > 10$ GeV, and photons calibrated at the EM scale with $p_{\mathrm{T}} > 10$ GeV. Hadronic taus should have $p_{\mathrm{T}} > 20$ GeV and are calibrated using local cluster weighting (LCW) with a subtraction of an offset to correct for pile-up. TES is also applied to correct for the energy scale. Jets are calibrated using LCW+JES scheme (see sec. 4.6) and should pass $p_{\mathrm{T}} > 20$ GeV cut. Finally, muons are required to have $p_{\mathrm{T}} > 6$ GeV. The measured momentum is already corrected to include energy losses in the calorimeter, whose contribution is omitted from $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation to avoid energy double counting.

4.8.3 Basic $E_{\mathrm{T}}^{\mathrm{miss}}$ Definitions Used in Higgs Analyses

Various definitions for $E_{\mathrm{T}}^{\mathrm{miss}}$ exist in ATLAS, sometimes called MET flavors such as:

- MET_RefFinal: This is the default MET definition that was used for Run I analyses. The soft term estimation is cluster-based with contributions from low-p_{T} tracks. No pile-up suppression is applied for the soft term at this level. MET_CST (cluster-based soft terms) is the same as MET_RefFinal with soft terms estimated from cluster energy deposits. The nomenclature was developed to separate it from MET_TST for Run-II studies.
- MET_TST (Track Soft Terms): This is the similar to MET_RefFinal but with track-based soft terms, referred to as MET TST. This definition was developed at the end of Run I and will be the default one for Run II analyses.
- MET_Track: track based missing transverse energy, giving an independent measurement from the calorimeter based one if using the default (nominal) MET_Track version. Three versions have been implemented: nominal, cluster corrected and jet corrected. The last definition ('jet corrected'), however, is similar to MET TST. This is discussed in detail in chapter 5.
- MET_STVF: This is derived from MET_RefFinal with a track-based correction factor for the soft terms as explained in section 4.8.4.

4.8.4 Pile-up Suppression

The measured energy performance is affected by pile-up which can also create additional jets. The main terms of $E_{\mathrm{T}}^{\mathrm{miss}}$ affected by pile-up are the jet and soft terms mainly because of their dependence on hadronic energy. Various methods are used to suppress pile-up and correct for its effects:

- **Selection Affecting Jets :** Making use of the JVF, the jet vertex fraction as defined in section 4.6.2.3, some jets (with $p_{\mathrm{T}} < 50$ GeV and $|\eta| < 2.4$) are discarded if none of the tracks associated to them originate from the event primary vertex (i.e. $|\mathrm{JVF}| = 0$). This mainly removes jets in the central region that have no associated tracks coming from the leading primary vertex (PV) [25, 34].

- **Soft Term Correction with STVF (SoftTerm Vertex Fraction) :** Similarly to the JVF variable defined for a jet, one can define STVF as the fraction of the soft term tracks momenta matched to the PV among all the soft term tracks momenta [25, 34] :

$$\mathrm{STVF} = \frac{\Sigma_{k,PV}\ p_T^{\mathrm{track,SoftTerm}}}{\Sigma_k\ p_T^{\mathrm{track,SoftTerm}}}\ , \tag{4.25}$$

  where k runs over all tracks not associated to high-p_{T} physics object. Once STVF is calculated, the soft term is scaled by its value. The resulting $E_{\mathrm{T}}^{\mathrm{miss}}$ is called MET STVF.

$$E_{\mathrm{x(y),corr}}^{\mathrm{miss,Soft\ Term}} = \mathrm{STVF} \times E_{\mathrm{x(y)}}^{\mathrm{miss,Soft\ Term}}\ . \tag{4.26}$$

  STVF corrected soft term is the CST.

- **Jet Area method and Track Filter for $E_{\mathrm{T}}^{\mathrm{miss}}$ Soft Term Pile-up Suppression:** This pile-up suppression method is based on the jet area correction (described in eq. (4.10)) used for jet pile-up suppresion. This mainly requires :

  1. Reconstructing jets with k_T algorithm down to $p_{\mathrm{T}} = 0$ GeV and energy reclustering afterwards.
  2. Subtracting pile-up using the jet area method for each k_T jet,

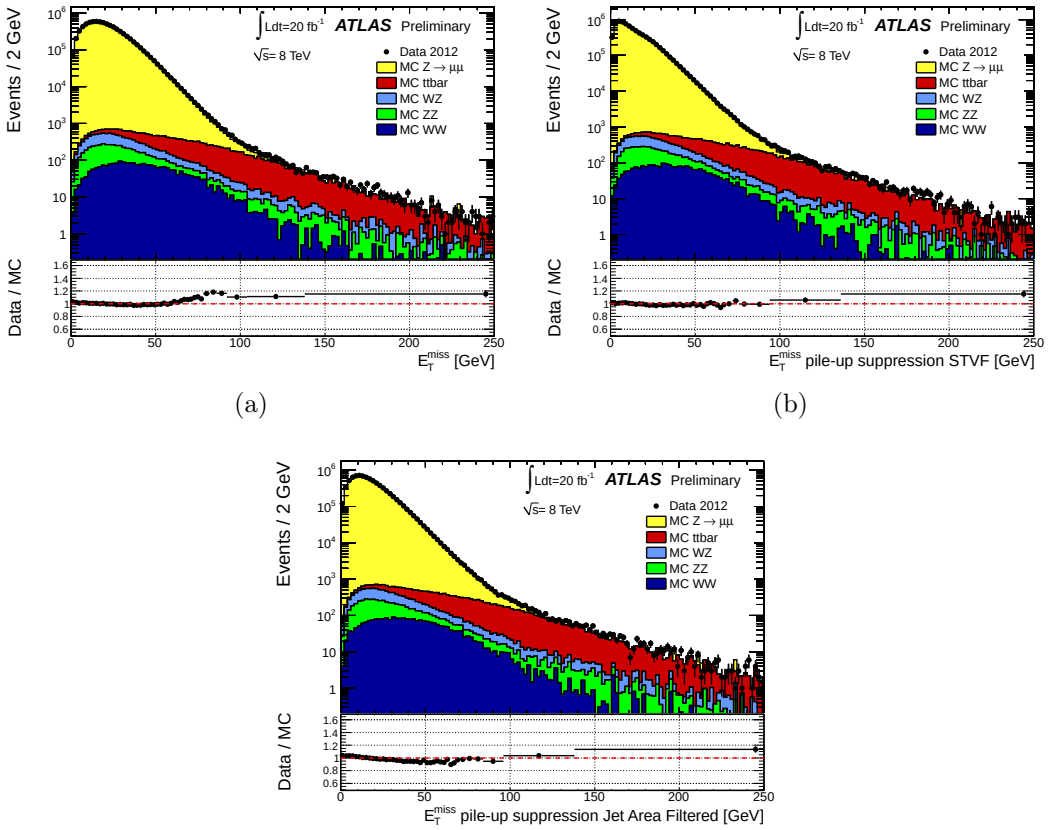
$$p_T^{corr} = p_T^{jet} - \rho \times A^{jet}\ ,$$

     where ρ is the event transverse momentum density.
  3. Applying a track based filter using only k_T jets with $|\mathrm{JVF}| > 0.25$.

  The resulting MET is called MET Jet Area Filtered [34].

- **Track Based Soft Term:** The soft term in MET can be replaced by a track based estimation of the soft term. All tracks originating from the primary vertex and not associated to any selected hard object are included. This definition,

however, suffers from the limited acceptance of the ID and misses the contribution of the soft neutral particles. Once this is applied, we get MET TST, as defined earlier.



(a) (b)

(c)

Figure 4-40: $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution in $Z \to \mu\mu$ samples before (a) and after pile-up suppression with MET-STVF (b) and MET-JetFiltered(c) [24].

4.8.5 Performance

The distribution of $E_{\mathrm{T}}^{\mathrm{miss}}$ in $Z \to \mu\mu$ samples before and after pile-up suppression with two different methods, namely STVF and jet area filtered ones, is shown in Fig. 4-40.

Soft term scale and resolution systematic uncertainties are estimated in $Z \to \mu\mu$ events, where there is no real missing transverse energy in the final state and the events have a very clean experimental signature, which allows for a very pure sample in Z events. Systematics of the hard terms inherit their uncertainties from the component objects. The resolution on the $x - y$ components is illustrated in Fig. 4-41. In addition, the linearity of MET defined as $(E_{\mathrm{T}}^{\mathrm{miss}} - E_{\mathrm{T}}^{\mathrm{miss,\,true}})/E_{\mathrm{T}}^{\mathrm{miss,\,true}}$ is shown in Fig. 4-42 as a function of $E_{\mathrm{T}}^{\mathrm{miss,\,true}}$. The difference plotted is expected to be zero for $E_{\mathrm{T}}^{\mathrm{miss}}$ estimation at the correct scale and with correct resolution [24].

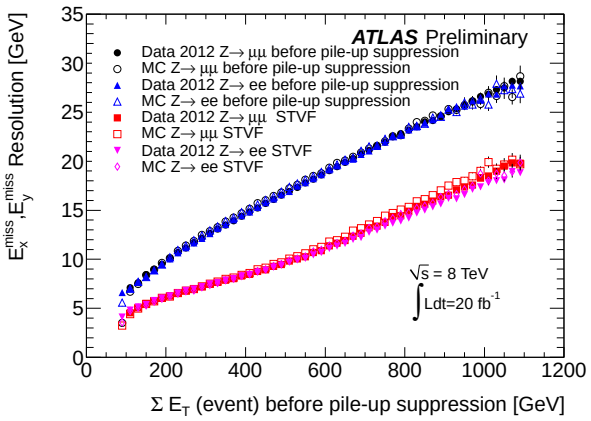
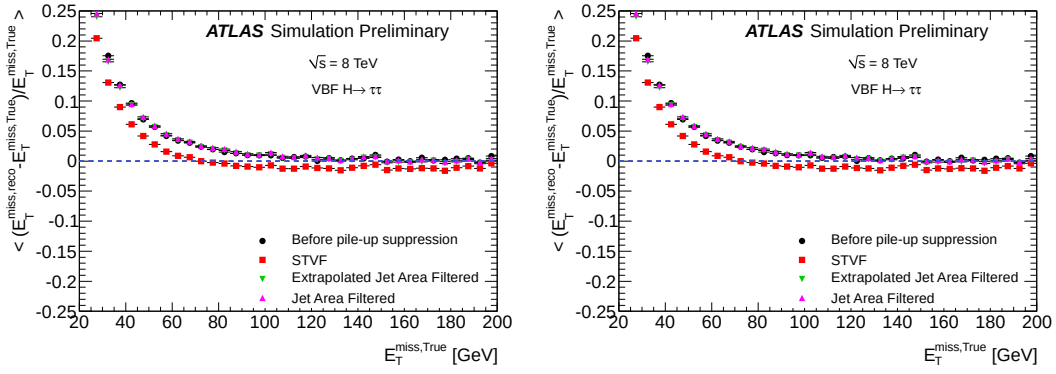


Figure 4-41: MET resolution as a function of the scalar sum of the measured physics objects E_{T} in the event, ΣE_T, before and after pile-up suppression [24].

In fact, the positive bias observed at low $E_{\mathrm{T}}^{\mathrm{miss,\,true}}$ values is due to the finite resolution on $E_{\mathrm{T}}^{\mathrm{miss}}$. At higher $E_{\mathrm{T}}^{\mathrm{miss,\,true}}$ values, the bias is within 5%. Different methods are used to evaluate the systematics. The highest values of systematic uncertainties are about 20% on scale and 5% on resolution. The soft term method used for both calorimeter-based and track-based MET definitions is described in detail in chapter 5. Another soft term systematics estimation method used only for calorimeter based MET is described in [24].

4.9 Conclusion

The reconstruction details of the various physics objects (tracks, electrons, muons, jets, taus, etc.) needed for the analysis discussed in this thesis were presented. The different calorimeter clustering methods were discussed as well, in addition to the various pile-up suppression and calibration techniques for each physics object. Finally,



(a) (b)

Figure 4-42: $E_{\mathrm{T}}^{\mathrm{miss}}$ linearity in ggF(gluon-gluon fusion) and VBF(vector boson fusion) Higgs decaying in the $\tau\tau$ mode $H \to \tau^{+}\tau^{-}$ ($m_H = 125$ GeV) (b) as a function of the true $E_{\mathrm{T}}^{\mathrm{miss}}$ [24].

the missing transverse energy, estimated using measurements of well reconstructed physics objects, was described.

Bibliography

[1] M. Cacciari and G. P. Salam, *pile-up subtraction using jet areas*, Phys.Lett. **B659** (2008) 119-126, arXiv: 0707.1378 [hep-ph]

[2] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-kt jet clustering algorithm*, JHEP **0804** (2008) 063.

[3] M. Cacciari, G. P. Salam, and G. Soyez., *Fastjet user manual*, Eur. Phys. J. **C72** (2012) 1896.

[4] The ATLAS Collaboration, *Expected Performance of the ATLAS Experiment-Detector, Trigger and Physics*, arXiv:0901.0512 [hep-ex] SLAC-R-980, CERN-OPEN-2008-020.

[5] The ATLAS Collaboration, *Performance of Impact Parameter-Based b-tagging Algorithms with the ATLAS Detector using Proton-Proton Collisions at* $\sqrt{s}$*= 7 TeV*, ATLAS-CONF-2010-091

[6] The ATLAS Collaboration, *Performance of primary vertex reconstruction in proton-proton collisions at* $\sqrt{s}$ *=7 TeV in the ATLAS experiment*, ATLAS-CONF-2010-069.

[7] The ATLAS Collaboration, *Performance of the ATLAS Secondary Vertex b-tagging Algorithm in 7 TeV Collision Data*, ATLAS-CONF-2010-042).

[8] The ATLAS Collaboration, *Performance of the ATLAS Silicon Pattern Recognition Algorithm in Data and Simulation at* $\sqrt{s}$ *= 7 TeV*, ATLAS-CONF-2010-072.

[9] The ATLAS Collaboration, *Tracking Results and Comparison to Monte Carlo simulation at* $\sqrt{s}$ *= 900 GeV*, ATLAS-CONF-2010-011.

[10] The ATLAS Collaboration, *ATLAS Muon Momentum Resolution in the First Pass Reconstruction of the 2010 p-p Collision Data at*$\sqrt{s}$*= 7 TeV*, ATLAS-CONF-2011-046.

[11] The ATLAS Collaboration, *Calibrating the b-Tag Efficiency and Mistag Rate in 35* pb^{-1} *of Data with the ATLAS Detector*, ATLAS-CONF-2011-089.

[12] The ATLAS Collaboration, *Charged-particle multiplicities in pp interactions measured with the ATLAS detector at the LHC*, New J. Phys. **13** (2011) 053033.

[13] The ATLAS Collaboration, *Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data*, ATLAS-CONF-2011-102.

[14] The ATLAS Collaboration, *Measurement of the inelastic proton-proton cross-section at $\sqrt{s}$ = 7 TeV with the ATLAS detector*, Nature Commun. **2** (2011) 463, arXiv:1104.0326 [hep-ex], CERN-PH-EP-2011-047.

[15] The ATLAS Collaboration, *Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data*, Eur.Phys.J. **C72** (2012) 1909.

[16] The ATLAS Collaboration. *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung.* ATLAS-CONF-2012-047

[17] The ATLAS Collaboration, *In situ jet pseudorapidity intercalibration of the ATLAS detector using dijet events in $\sqrt{s}$= 7 TeV proton-proton 2011 data*, ATLAS-CONF-2012-124

[18] The ATLAS Collaboration, *Measurement of the b-tag Efficiency in a Sample of Jets Containing Muons with 5 fb^{-1} of data from the ATLAS Detector*, ATLAS-CONF-2012-043

[19] The ATLAS Collaboration, *Performance of the ATLAS Inner Detector Track and Vertex Reconstruction in the High Pile-Up LHC Environment*, ATLAS-CONF-2012-042

[20] The ATLAS Collaboration, *Performance of the ATLAS Inner Detector Track and Vertex Reconstruction in the High Pile-Up LHC Environment*, ATLAS-COM-PHYS-2012-474

[21] The ATLAS Collaboration, *Performance of the Reconstruction and Identification of Hadronic tau Decays in ATLAS with 2011 Data*, ATLAS-CONF-2012-142

[22] The ATLAS Collaboration, *Identification of Hadronic Decays of Tau Leptons in 2012 Data with the ATLAS Detector*, ATLAS-CONF-2013-064

[23] The ATLAS Collaboration, *Jet global sequential corrections with the ATLAS detector in proton-proton collisions at $\sqrt{s}$= 8 TeV*, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/%20JetEtmissApproved2013JESUncertainty, 2013.

[24] The ATLAS Collaboration, *Performance of Missing Transverse Momentum Reconstruction in ATLAS studied in Proton-Proton Collisions recorded in 2012 at $\sqrt{s}$ = 8 TeV*, ATLAS-CONF-2013-082

[25] The ATLAS Collaboration, *Pile-up subtraction and suppression for jets in ATLAS*, ATLAS-CONF-2013-083

[26] The ATLAS Collaboration, *Preliminary results on the muon reconstruction efficiency, momentum resolution, and momentum scale in ATLAS 2012 pp collision data*, ATLAS-CONF-2013-088.

[27] The ATLAS Collaboration, *b-tagging in dense environments*, ATLAS-CONF-2014-014.

[28] The ATLAS Collaboration, *Calibration of b-tagging using dileptonic top pair events in a combinatorial likelihood approach with the ATLAS experiment*, ATLAS-CONF-2014-004.

[29] The ATLAS Collaboration, *Calibration of the performance of b-tagging for c and light-flavour jets in the 2012 ATLAS data*, ATLAS-CONF-2014-046.

[30] The ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton-proton collision data*, ATLAS-CONF-2014-032.

[31] The ATLAS Collaboration, *Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton-proton collision data*, Eur.Phys.J. **C74** (2014) 7, 2941 arXiv:1404.2240 [hep-ex] CERN-PH-EP-2014-040.

[32] The ATLAS Collaboration, *Identification and energy calibration of hadronically decaying tau leptons with the ATLAS experiment in pp collisions at* $\sqrt{s}$*=8 TeV*, Eur. Phys. J. **C75** (2015) 7, 303 arXiv:1412.7086 [hep-ex] CERN-PH-EP-2014-227.

[33] The ATLAS Collaboration, *Measurement of the muon reconstruction performance of the atlas detector using 2011 and 2012 lhc proton-proton collision data*, Eur. Phys. J. **C74** (2014) 3130.

[34] The ATLAS Collaboration, *Pile-up Suppression in Missing Transverse Momentum Reconstruction in the ATLAS Experiment in Proton-Proton Collisions at* $\sqrt{s}$*= 8 TeV*, ATLAS-CONF-2014-019.

[35] The ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*, ATLAS-CONF-2014-018.

[36] The ATLAS Collaboration, *Data-driven determination of the energy scale and resolution of jets reconstructed in the ATLAS calorimeters using dijet and multijet events at* $\sqrt{s}$*= 8 TeV*, ATLAS-CONF-2015-017.

[37] The ATLAS Collaboration, *Jet energy measurement and its systematic uncertainty in proton-proton collisions at* $\sqrt{s}$*=7TeV with the ATLAS detector*, Eur.Phys.J. **C75** (2015) 1, 17 arXiv:1406.0076 [hep-ex] CERN-PH-EP-2013-222.

[38] The ATLAS Collaboration, *Jet global sequential corrections with the ATLAS detector in proton-proton collisions at* $\sqrt{s}$*= 8 TeV*, ATLAS-CONF-2015-002.

[39] T. Cornelissen et al. *Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)*, Soft-pub-2007-007.

[40] A. B. Galtieri, F. Margaroli, and I. Volobouev. *Precision measurements of the top quark mass from the Tevatron in the pre-LHC era*, Rept.Prog.Phys. **75** (2012) 056201 arXiv:1109.2163 [hep-ex] FERMILAB-PUB-11-458.

[41] The ATLAS Muon Combined Performance Group, *Pile-up Dependence of the ATLAS Muon Performance*, ATL-COM-PHYS-2011-1640.

[42] W. Lampl et al, *Calorimeter Clustering Algorithms: Description and Performance*, ATL-LARG-PUB-2008-002.

[43] Gavin P. Salam. *Towards Jetography.* Eur. Phys. J. **C67** (2010) 637-686 arXiv:0906.1833 [hep-ph].

[44] S.Catani, Y.L.Dokshitzer, M.Seymour, and B.Webber, *Longitudinally invariant $k_\perp$ clustering algorithms for hadron collisions*, Nucl. Phys. **B406** (1993) 187-224.

[45] S.D.Ellis and D.E. Soper, *Successive combination jet algorithm for hadron collisions*, Phys.Rev. **D48** (1993) 3160-3166.

[46] T.Barillari et al., *Local Hadronic Calibration*, ATL-LARG-PUB-2009-001.

[47] W. Waltenberger, R. Fruhwirth, and P. Vanlaer, *Adaptive vertex fitting*, J.Phys. **G34** (2007) N343, CERN-CMS-NOTE-2007-008, CMS-NOTE-2007-008.

Chapter 5

Track-based Missing Transverse Energy

5.1 Introduction

This chapter presents a track-based method to estimate the missing transverse energy, $E_{\mathrm{T}}^{\mathrm{miss}}$, which is calculated based on the measurements of transverse momentum of visible measured particles in the detector:

$$E_{\mathrm{T}}^{\mathrm{miss}} = p_{\mathrm{T}}^{\mathrm{miss}} = |\mathbf{p}_{\mathrm{T}}^{\mathrm{miss}}| \ . \tag{5.1}$$

where $\mathbf{p}_{\mathrm{T}}^{\mathrm{miss}}$ is the opposite of vector sum of the transverse momentum of all the visible measured particle. The track-based estimate is indeed:
- a complement to the existing Run-I calorimeter-based (defined in chapter 4) measurement of $E_{\mathrm{T}}^{\mathrm{miss}}$ (mainly estimated from calorimetric energy deposits with the inclusion of low-pt tracks in the soft term calculation) in a less pile-up dependent way.
- the equivalent of the default Run-II $E_{\mathrm{T}}^{\mathrm{miss}}$ definition, where the pile-up conditions will become even more challenging than the 2012 Run of LHC.

The various definitions of track-based $E_{\mathrm{T}}^{\mathrm{miss}}$ are presented first, with the corresponding object selection requirements. Starting with a pure[1] track-based definition, corrections to account for neutral jet components, accurate electron transverse momentum measurement and mis-reconstructed tracks removal are applied in the subsequent definitions[2]. The following sections discuss the systematic uncertainties evaluation for various definitions, the event generator dependencies in addition to the correlation between track- and calorimeter-based definitions. My personal contribution is the estimation of the soft term systematic uncertainties for various definitions of track-based $E_{\mathrm{T}}^{\mathrm{miss}}$, in addition to personal studies in the $H \to WW^*$ and $H \to \tau^+\tau^-$ Higgs channels. My work covers most of this chapter (sec. 5.4 till sec. 5.8).

[1] Only tracks enter the calculation (no corrections).

[2] All definitions are presented since some Higgs analyses use multiple track-based $E_{\mathrm{T}}^{\mathrm{miss}}$ definitions at the same time.

5.1.1 Motivations

Even with an existing calorimeter-based measurement of missing transverse energy, the need for a track-based one was driven by the following:

- Measuring $E_{\mathrm{T}}^{\mathrm{miss}}$, the missing transverse energy, is essential for many physics studies at the LHC especially for Higgs analysis and SUSY studies. A precise measurement of $E_{\mathrm{T}}^{\mathrm{miss}}$ is crucial for better background rejection and signal extraction in Higgs analyses such as $H \to WW^*$ and $H \to \tau^+\tau^-$ with real $E_{\mathrm{T}}^{\mathrm{miss}}$ from neutrinos in the final state (The dilepton channel of the $H \to \tau^+\tau^-$ has at least 4 neutrinos in the final state).

- Several factors such as mis-reconstructed objects, detector inefficiencies, dead or noisy read-out channels, etc. can lead to large fake $E_{\mathrm{T}}^{\mathrm{miss}}$ or degrade its resolution, which increases the contamination from background events with no real intrinsic transverse missing energy.

- Keeping in mind the large number of pile-up interactions for the 2012 LHC run at 8 TeV, and the fact that the calorimeter based $E_{\mathrm{T}}^{\mathrm{miss}}$ measurement is based on the calorimeter deposits originating from all interaction vertices, the need to develop a $E_{\mathrm{T}}^{\mathrm{miss}}$ that has little pile-up dependence is obvious. Tracks can be easily associated to each collision vertex using the tracking and vertexing algorithms in ATLAS. Hence, a calculation of $E_{\mathrm{T}}^{\mathrm{miss}}$ using tracks from the primary vertex leads to lower pile-up dependence.

- Finding the best possible $E_{\mathrm{T}}^{\mathrm{miss}}$ measurement allows to improve the mass reconstruction directly affected by $E_{\mathrm{T}}^{\mathrm{miss}}$ in Higgs analyses, such as the transverse mass m_T in $H \to WW^*$ events and $\mathrm{m_{MMC}}$[3] in $H \to \tau^+\tau^-$ studies, with real $E_{\mathrm{T}}^{\mathrm{miss}}$ from neutrinos in the final state. As a result, better signal/background discrimination is achieved.

- Since the cosmic-ray muons and the beam halo background tracks have usually different time and geometry signatures than the collision tracks, a track-based measurement of $E_{\mathrm{T}}^{\mathrm{miss}}$ allows lower sensitivity than the calorimeter based one to such unwanted events [1].

5.1.2 Performance Limitations

The precision of the track-based estimation of $p_{\mathrm{T}}^{\mathrm{miss}}$ is limited by the following:

1. Its basic definition includes terms from charged particles only. A correction for neutral particles can be applied ('jet-corrected' definition).

2. Limited geometrical coverage since the ATLAS tracking detector covers only the region with $|\eta| < 2.5$.

[3]MMC stands for missing mass calculator, which is a likelihood algorithm used in $H \to \tau^+\tau^-$ analyses to calculate di-tau mass (details in chapter 6). The CMS equivalent is the secondary vertex (SV) fit [2].

3. The low resolution on high-p_{T} tracks, which bend less in a magnetic field than the low-p_{T} ones, affects the resolution on the track momentum and hence the resolution on the measurement of $p_{\mathrm{T}}^{\mathrm{miss}}$.

4. The resolution on $p_{\mathrm{T}}^{\mathrm{miss}}$ is degraded when highly boosted jets are involved due to a degradation of the tracking efficiency in the jet core [1].

5.2 $p_{\mathrm{T}}^{\mathrm{miss}}$ Reconstruction

5.2.1 Definitions

There are 3 definitions of track based missing transverse momentum, $p_{\mathrm{T}}^{\mathrm{miss}}$, two of which are derived from the nominal definition. The selection criteria of physics objects (tracks, leptons and jets) used for $p_{\mathrm{T}}^{\mathrm{miss}}$ calculation are defined in sec. 5.3, and apply to all 3 definitions presented hereafter.

5.2.1.1 The Nominal Track MET Definition

The nominal $p_{\mathrm{T}}^{\mathrm{miss}}$ components as well as the $\sum p_T$ of the event are calculated as follows:

$$p_{x,y}^{\mathrm{miss}} = -\sum_{tracks} p_{x,y}^{track} \quad , \quad \sum p_T = \sum_{tracks} p_T^{track} \ . \tag{5.2}$$

Tracks are taken into account if they pass the quality cuts. However, tracks associated with leptons are taken into account in the $p_{\mathrm{T}}^{\mathrm{miss}}$ calculation, even if they fail some of the track high quality requirements.

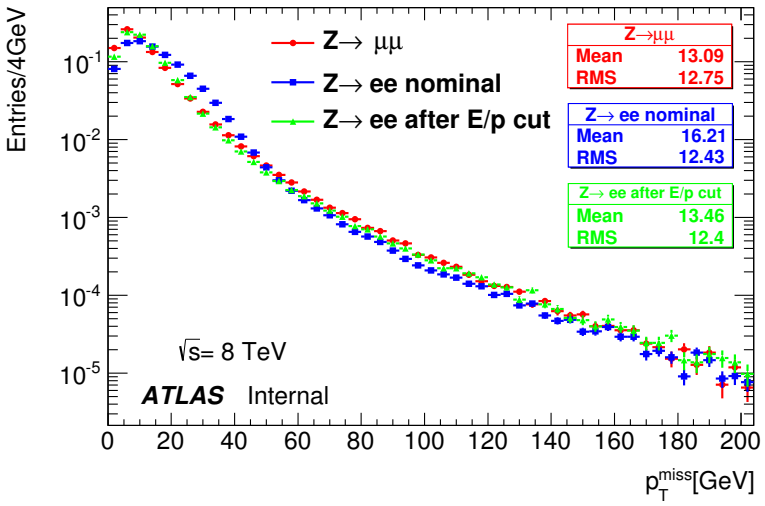


Figure 5-1: Nominal Track MET distribution in Monte Carlo $Z \to ee$ events before and after the removal of events with Bremsstrahlung electrons using a cut on electron $E_T/p_T < 1.4$, compared to $Z \to \mu^+\mu^-$ results [1].

5.2.1.2 The Cluster Track MET-Cl Definition

The $p_{\mathrm{T}}^{\mathrm{miss}}$ resolution is worse in $Z \to ee$ events than in $Z \to \mu\mu$ events. This is due to the track resolution being worse in the case of electrons because of bremsstrahlung losses. This is demonstrated in Fig. 5-1, where a $E_{\mathrm{T}}/p_{\mathrm{T}} < 1.4$ cut is applied to electrons to reduce bremsstrahlung losses, resulting in a comparable resolution for both $Z \to ee$ and $Z \to \mu\mu$ samples. E_T is the electron calorimeter cluster transverse energy and p_T is the electron track transverse momentum.

While the direction of the electron is determined by (η, ϕ) as calculated at the track level, the electron track p_{T} is replaced by the cluster p_{T} in the Track MET-Cl definition.

5.2.1.3 The Jet Corrected Track MET-Cl-j Definition

In the Track MET-Cl-j definition, the jet associated tracks[4] contribution is replaced by the pileup-corrected jet p_{T}, in addition to applying the cluster p_{T} correction for the electrons. This correction allows to account for the neutral components of jets.

In addition, this definition is very close to the calorimeter-based missing transverse energy (MET) with track-based soft terms[5] 'MET TST' (defined in chapter 4), which will be the default MET definition for Run-II. If we add the photon term to the track MET-Cl-j calculation, we get the equivalent of MET TST. The calorimeter-based MET measurement is defined as:

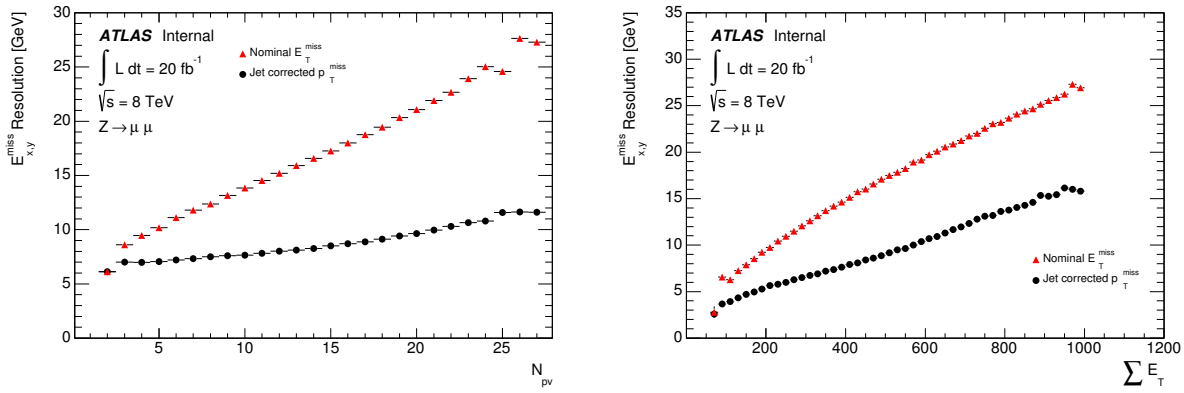
$$E_{\mathrm{x(y)}}^{\mathrm{miss}} = (E_{\mathrm{x(y)}}^{\mathrm{miss,el}} + E_{\mathrm{x(y)}}^{\mathrm{miss},\gamma} + E_{\mathrm{x(y)}}^{\mathrm{miss},\tau} + E_{\mathrm{x(y)}}^{\mathrm{miss,jets}} + E_{\mathrm{x(y)}}^{\mathrm{miss,softTerms}} + E_{\mathrm{x(y)}}^{\mathrm{miss},\mu}) \ . \quad (5.3)$$

Compared to the nominal track MET, Track MET-Cl-j leads to better scale and resolution with a good stability against pileup, and is closer to the true MET in the event. Fig. 5-2 shows the resolution of different MET definitions as a function of the number of vertices N_{pv} in (a) and ΣE_{T} (b). The dependence on N_{pv} is reduced significantly when switching to Track MET-Cl-j, with a stable resolution on most of the spectrum. The observed increase in the resolution with pile-up is attributed to a loose jet selection. The resolution in terms of ΣE_{T} is also improved with the jet corrected track MET.

Similarly, the contribution to Track MET-Cl-j coming from tracks associated to a τ lepton is replaced by the calorimeter p_{T}^{τ} measurement.

[4] Tracks can be associated to jets using either ΔR or ghost association methods as described in [3].

[5] 'Hard' components of MET are terms coming from well measured physics objects entering the calculation of the MET definition. Everything else is referred to as 'soft' terms of MET.



(a) (b)

Figure 5-2: Resolutions of nominal calorimeter-based $E_{\mathrm{T}}^{\mathrm{miss}}$ and jet corrected $p_{\mathrm{T}}^{\mathrm{miss}}$ as a function of the number of reconstructed vertices in the event N_{PV} (a) and $\sum E_T$. The resolution is improved when switching to the track-based $p_{\mathrm{T}}^{\mathrm{miss}}$ estimation [1].

5.2.2 Why Multiple Definitions ?

The baseline default(nominal) track MET definition was meant to be an independent measurement of MET in the event, orthogonal to the calorimeter based MET. This is possible since the former is based on tracks only while the calorimeter based MET contributions come mainly from energy deposits in the calorimeter. The cluster corrected version is needed to recover losses due to Bremsstrahlung while the jet corrected version takes into account charged/neutral contributions. In all definitions, there are no tracks for pile-up interactions basically: tracks entering the definition should pass the quality cuts. The exception is made for tracks matched to leptons in the event. Thus, track MET is still interesting for the soft term evaluation. In addition, some Higgs analyses, e.g. $H \to W^*W$, use multiple definitions of track MET at different stages of the analysis, especially when the better resolution obtained with track MET when compared to MET STVF[6] leads to better signal/background discrimination.

5.2.3 Including Jets in Track MET-Cl-j: ΔR Vs Ghost Association

The information from calorimetric energy deposits alone is not enough to have a complete understanding of jets. As a result, information from the inner detector needs to be taken into account for better performance. The need to use tracks arises for example in cases of b-tagging[7] and pile-up suppression techniques (e.g. jet vertex fraction).

Track MET-Cl-j in the following sections uses ghost association for track-jet association. However, for $H \to WW^*$ results, an older alternative method, namely ΔR, is applied. A brief description of the two methods is given below, while the detailed

[6]Calorimeter based MET definition with pile-up correction. Exact definition is given in chapter 4.
[7]Identification ("tagging") of jets originating from bottom quarks as described in chapter 4.

cuts for the $H \rightarrow WW^*$ study are presented in sec. 5.6.1. During Run-II, ghost association will be the default method.

- ΔR **method:** Denoting the pseudorapidity by η and the azimuthal angle by ϕ, ΔR is defined as:
$$\Delta R = \sqrt{(\Delta\eta^2) + (\Delta\phi^2)} \ . \tag{5.4}$$
At low pile-up conditions, it is safe to assume that the jet is cone-shaped and that anti-k_T jets are circular. Hence, a simple association of tracks falling within a cone of ΔR (0.4 or 0.6 depending on the analysis) around a jet works well [3].

- **Ghost association:** Starting 2011, given the increasing pile-up conditions, jets are starting to overlap and even anti-k_T jets are not always circular. In addition, some physics analyses use "fat" jets i.e. anti-k_T with $\Delta R = 1.0$ or Cambridge/Aachen (C/A) with R = 1.2 jets with k_T[8] subjets. The original hypothesis on the shape on the jet fails, especially for k_T and C/A jets, and the ΔR method breaks down, degrading the performance. Consequently, the ghost association method was developed. The method is based on the 'active area' concept as defined in [5]. The active area is calculated assuming a uniform distribution of 'ghost' particles with tiny energy[9] (~ 0). The sea of evenly distributed 'ghosts' is appended at the topo-cluster level before jet clustering is performed. Similarly, 'ghost' tracks are created with very low p_{T} using the same (η, ϕ) track measurements. After running the jet clustering algorithm, ghost tracks associated to the jet are identified. This is followed by the jet-track association based on the ghost-real track matching[5, 7, 8].

The tracks associated to jets are removed from Track MET-Cl-j and replaced by the jet p_{T}.

5.3 MET Physics Object Selection

The selection criteria for tracks, electrons, muons and jets entering the track MET definitions are defined in this section. Additional information on the reconstruction of physics objects and the identification criteria for different working points is given in chapter 4.

5.3.1 Basic Track and Primary Vertex Selection

The primary vertex (PV), associated with the hard interaction process, is assigned to the vertex with the highest $\sum(p_T^{track})^2$. Events are rejected if they fail to have at least three reconstructed tracks with $p_T > 500$ MeV associated with the PV.
Tracks entering the $p_{\mathrm{T}}^{\mathrm{miss}}$ calculation should pass certain requirements to have a good momentum measurement, an efficient rejection of mis-reconstructed tracks and a

[8] Jet algorithms are described in detail in [4].
[9] This method is infrared safe.

very good track-to-vertex association. Tracks are considered for $p_{\mathrm{T}}^{\mathrm{miss}}$ calculation if the following criteria are satisfied:

- $p_T > 500$ MeV
- $|\eta| < 2.5$
- at least 1 pixel detector hit
- at least 6 SCT hits
- Transverse impact parameter with respect to the PV $|d_0| < 1.5$ mm
- Longitudinal impact parameter with respect to the PV $|z_0 \times \sin(\theta)| < 1.5$ mm.

5.3.2 Mis-reconstructed Tracks Removal

The momentum of the selected tracks can be mis-reconstructed in some cases when they interact early in the ID material or when there is an early bremsstrahlung. Such tracks are mainly observed in the regions $|\eta| < 1.5$ for tracks with $p_T > 200$ GeV, and $|\eta| > 1.5$ for track $p_T > 120$ GeV leading to large tails in the reconstructed momentum distribution. This affects 0.1% fo the reconstructed tracks and the effect is more important in $Z \to ee$ events than in $Z \to \mu\mu$ ones.
To eliminate mis-reconstructed tracks in the regions selected above, the following selections are applied:

- the relative uncertainty on the charge to track-momentum ratio (q/p) should be:

$$\frac{\sigma(q/p)}{(q/p)} \; < \; 0.4 \; .$$

- the energy in calorimeter clusters in a cone of 0.1 radius, in the (η, ϕ) plane, around the track should satisfy $E_T^{0.1}/p_T^{track} > 0.65$ in order to reflect the reconstructed track momentum.

The improvement after the mis-reconstructed track removal in the $Z \to ee$ samples is illustrated in Fig. 5-3 showing the distribution of the track transverse momentum p_T in $Z \to ee$ events in 2012 data.

5.3.3 Lepton Selection

– Selected electrons have $p_T^{\mathrm{cluster}} > 10$ GeV and $|\eta| < 2.47$. They should pass the "medium"[10] selection criteria for $Z \to ll$ events and "tight" selection cuts for $W \to l\nu$ events. Electron candidates falling in the barrel transition region $(1.37 < |\eta| < 1.52)$ are discarded.

– Staco combined[11] muons with $p_T > 6$ GeV and $|\eta| < 2.5$ are chosen. A maximum

[10] Electron identification criteria detailed in sec. 4.4.
[11] Muon identification criteria detailed in sec. 4.5.

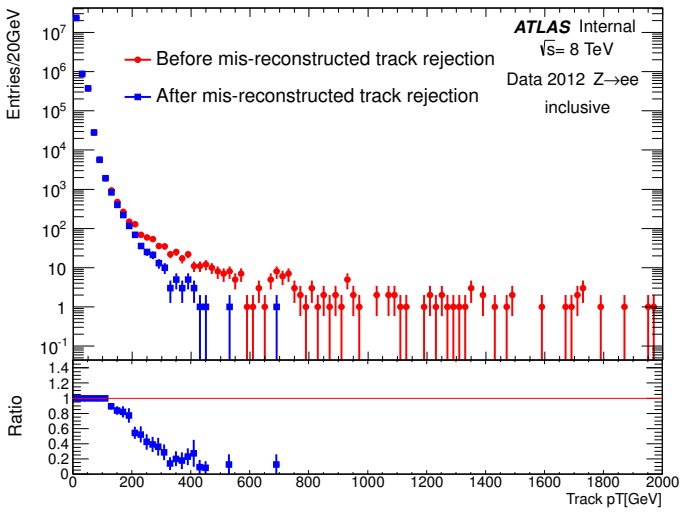


Figure 5-3: Track p_T distribution in $Z \to ee$ events from 2012 data before and after mis-reconstructed track momentum rejection [1].

z displacement of 10 mm from the main primary vertex is required. And the "STACO" algorithm requires the reconstructed muon in the spectrometer to have a matched track in the inner detector.

5.3.4 Jet selection

Anti-k$_{\mathrm{T}}$ algorithm with $\Delta R = 0.4$ is used for jet reconstruction. Pileup corrections using the LCW+JES scheme are applied at the topocluster level and subsequent calibration levels. Jets entering the $p_{\mathrm{T}}^{\mathrm{miss}}$ calculation satisfy the following cuts:

- calibrated $p_T > 20$ GeV
- $|\eta| < 4.4$
- no overlap with any of the selected leptons
- Jet Vertex Fraction cut: JVF > 0.25 if calibrated $p_T < 50$ GeV and $|\eta| < 2.4$ to reduce the contamination from pile-up jets in the central region. For Run-II, the results can be further improved using the Jet Vertex Tagger and JVFCorr (JVF with additional corrections from primary vertex) [9] (see details in sec. 4.6).

5.4 Estimation of $p_{\mathrm{T}}^{\mathrm{miss}}$ Soft Systematic Uncertainties

An important consideration in the study of $p_{\mathrm{T}}^{\mathrm{miss}}$ is the systematic uncertainty arising from soft components in events with hard objects (well measured high-p_{T} leptons and jets), where soft denotes the $p_{\mathrm{T}}^{\mathrm{miss}}$ components which remains after removing the hard objects terms (jets are considered as hard objects in the Track MET-Cl-j definition,

whereas their associated tracks are included in the soft term[12] in the other definitions). Fig. 5-4 shows the TrackMET composition where $p_{\mathrm{T}}^{\mathrm{hard}}$ is referred to as p_T^{lep}. A summary of the different soft/hard terms related components is given in Table 5.1.

Term	Track MET and Track MET-Cl	Track MET-Cl-j
$p_{\mathrm{T}}^{\mathrm{hard}}$	contribution from p_T of e/μ/ν referred to also as p_T^{lep}	contribution from p_T of e/μ/ν/jets $\equiv p_T^{\mathrm{lep}} + p_T^{\mathrm{jets}}$
$p_{\mathrm{T,soft}}^{\mathrm{miss}}$	-objects failing thresholds and quality cuts (e.g. soft jets) -objects not entering track MET definition e.g. jets -tracks not matched to any reconstructed physics object	-objects failing thresholds and quality cuts (e.g. soft jets) - tracks not matched to any reconstructed physics object
interest	what recoils against the Z boson	what recoils against the Z(+jet(s)) system

Table 5.1: Comparison of the different contributions to soft and hard terms for the different track MET definitions when evaluating systematic uncertainties in $Z \rightarrow \mu\mu$ samples.

The developed systematics evaluation method exploits the balance between the total transverse momentum of the high-p_{T} leptons, jets (and neutrinos in the event e.g. in the $W \rightarrow \ell\nu$ events), $p_{\mathrm{T}}^{\mathrm{hard}}$, and $p_{\mathrm{T,soft}}^{\mathrm{miss}}$. The systematic uncertainties are calculated by evaluating data-MC differences in the mean and resolution of $\vec{p}_{\mathrm{T,soft}}^{\mathrm{miss}}$ in $Z \rightarrow \mu\mu$ events. Such events have a clean event signature, relatively large cross section and are characterized by the absence of any real $E_{\mathrm{T}}^{\mathrm{miss}}$ source.

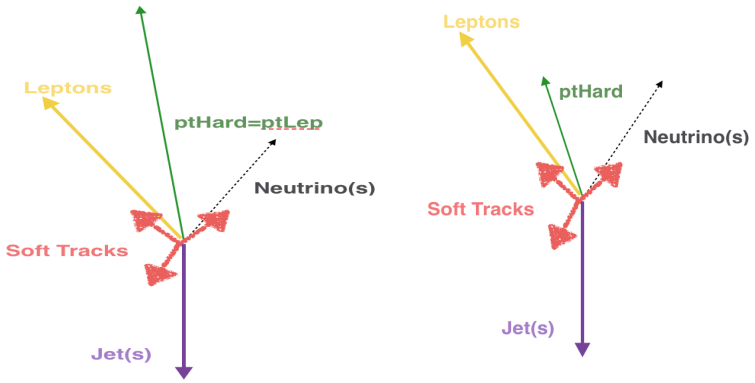
5.4.1 Soft Term Guidelines

The soft term in systematics evaluation (as described in Table 5.1) was first defined in $Z \rightarrow \mu\mu$ events as what recoils against the Z boson, using the Z boson p_{T} direction to define scale and resolution. The same principle is used for calorimeter based MET systematics using this method. However, with jet corrections, the soft term was adjusted to accommodate all hard terms entering the definition of track MET, resulting thus in a definition that works well regardless of the jet multiplicity providing more stable results.

5.4.2 Method

To evaluate the systematic uncertainties related to the mean and resolution of $p_{\mathrm{T,soft}}^{\mathrm{miss}}$, $\mathbf{p}_{\mathrm{T,soft}}^{\mathrm{miss}}$ is decomposed into a longitudinal component $p_{\mathrm{T,soft,L}}^{\mathrm{miss}}$ along the $\mathbf{p}_{\mathrm{T}}^{\mathrm{hard,trk}}$ direc-

[12]The qualification 'soft term' is not very appropriate for definitions other than Track MET-Cl-j, where jets enter the soft term calculation. This nomenclature was inherited from the calorimeter-based MET soft term systematic method terminology.



(a) (b)

Figure 5-4: Schematic representation of $p_{\mathrm{T}}^{\mathrm{hard}}$ for the jet corrected definition (right) and the remaining nominal and cluster definitions (left). For the latter, it is sometimes referred to as $p_{\mathrm{T}}^{\mathrm{Lep}}$ and the contribution from tracks associated to jets enters the soft term calculation.

tion and into a perpendicular component $p_{\mathrm{T,soft,P}}^{\mathrm{miss}}$ lying along the direction orthogonal to $\mathbf{p}_{\mathrm{T}}^{\mathrm{hard,trk}}$ in the transverse plane (see Fig. 5-6). Exact expressions of $\mathbf{p}_{\mathrm{T,soft}}^{\mathrm{miss}}$ and $\mathbf{p}_{\mathrm{T}}^{\mathrm{hard,trk}}$ are given in equations (5.5) and (5.6). The neutrino term in equation (5.5) has no contribution to $\mathbf{p}_{\mathrm{T}}^{\mathrm{hard,trk}}$ in the case of $Z \to \mu\mu$ events from which the systematic uncertainties are estimated. For non-Z events, where there is a true $E_{\mathrm{T}}^{\mathrm{miss}}$, the neutrino term helps recover the $\mathbf{p}_{\mathrm{T}}^{\mathrm{hard,trk}}$ axis direction as defined for Z events, so that the systematic uncertainties from Z events can be applied to such samples.

The mean and resolution of $p_{\mathrm{T,soft}}^{\mathrm{miss}}$ components are extracted from the mean and width parameters of the Gaussian fits to the corresponding distributions for both data and MC samples, as illustrated in Fig. 5-5 for a data sample in the 1-jet bin with $\langle\mu\rangle \geq 22$. The mean of the longitudinal component of $p_{\mathrm{T,soft}}^{\mathrm{miss}}$ reflects the $p_{\mathrm{T,soft}}^{\mathrm{miss}}$ scale since the longitudinal direction is sensitive to the balance between high-p_{T} leptons and jets, and $\mathbf{p}_{\mathrm{T,soft}}^{\mathrm{miss}}$.

$$p_{\mathrm{x(y)}}^{\mathrm{hard,trk}} = \Sigma_{\mu} p_{x(y)}^{\mu} + \Sigma_{ele} p_{x(y)}^{ele} + \Sigma_{\nu} p_{x(y)}^{\nu} + \Sigma_{jets} p_{x(y)}^{jet} \ . \tag{5.5}$$

$$p_{\mathrm{x(y),soft}}^{\mathrm{miss}} = p_{\mathrm{x(y)}}^{\mathrm{miss}} + \Sigma_{\mu} p_{x(y)}^{\mu} + \Sigma_{ele} p_{x(y)}^{ele} + \Sigma_{jets} p_{x(y)}^{jet} \ . \tag{5.6}$$

The contribution from jet terms in eq. (5.5) and (5.6) is only valid for Track MET-Cl-j and omitted for other definitions.

The systematic uncertainties are evaluated in bins of $p_{\mathrm{T}}^{\mathrm{hard}}$, of jet multiplicity, and of the average number of interactions per crossing $\langle\mu\rangle$. The first two take into account event topology dependence, while the latter reflects the pile-up dependence. To evaluate the systematic uncertainties of $p_{\mathrm{T,soft}}^{\mathrm{miss}}$ in $Z \to \mu\mu$ events, six systematic variations are considered: shifting $\langle p_{\mathrm{T,soft,L}}^{\mathrm{miss}} \rangle$ up and down, as well as scaling the

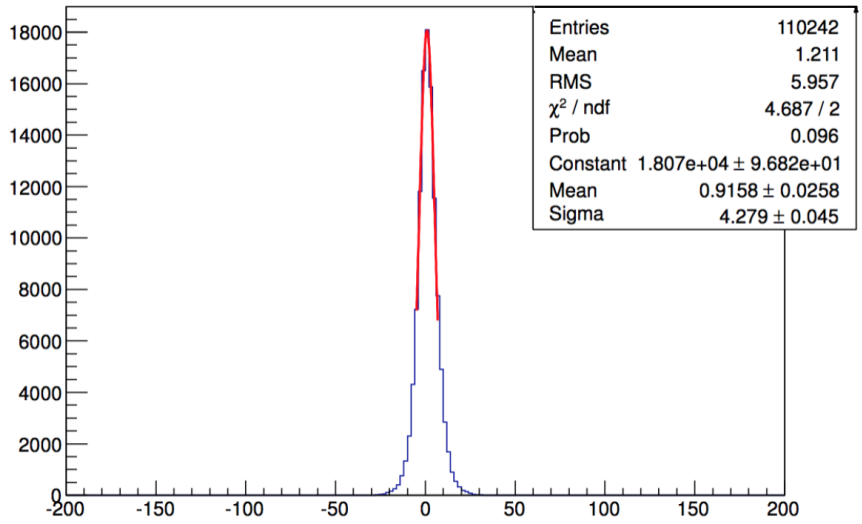


Figure 5-5: Example showing the Gaussian fit of $p^{\text{miss}}_{\text{T,soft,L}}$ distribution, in a $Z \to \mu\mu$ data (at $\sqrt{s} = 8$ TeV) sample in the 1-jet bin, to extract the scale (mean) of $p^{\text{miss}}_{\text{T,soft}}$. The results are derived for Track MET-Cl-j using the ghost-association method in the first bin of $\mathbf{p}^{\text{hard}}_{\text{T}}$, for events with an average number of interactions per crossing $\langle\mu\rangle$ greater than 22.

$p^{\text{miss}}_{\text{T,soft,L}}$ and $p^{\text{miss}}_{\text{T,soft,P}}$ resolutions up and down either together in a correlated manner or in a fully anti-correlated manner. These systematic studies provide 6 independent variations that are governed by the following equation :

$$p^{\text{miss}}_{\text{T,soft,L(P)}} = \left(1 \pm \Delta R_{L(P)}\right) \times \left(p^{\text{miss}}_{\text{T,soft,L(P)}} - \langle p^{\text{miss}}_{\text{T,soft,L(P)}}\rangle\right) + \langle p^{\text{miss}}_{\text{T,soft,L(P)}}\rangle \pm \Delta^{trk} \ , \quad (5.7)$$

where :

- All terms are estimated in the associated ($p^{\text{hard}}_{\text{T}}$, $\langle\mu\rangle$, jet multiplicity) bin.
- $R_{L(P)}$: data/MC resolution ratio for the longitudinal (perpendicular) component of $p^{\text{miss}}_{\text{T,soft}}$.
- $\Delta R_{L(P)} = |1 - R_{L(P)}|$.
- $\langle p^{\text{miss}}_{\text{T,soft,L(P)}}\rangle$: mean of the longitudinal (perpendicular component) of $p^{\text{miss}}_{\text{T,soft}}$.
- Δ^{trk} : data-MC difference in mean longitudinal component estimation. This shift applies only to the longitudinal component calculation.

For $Z \to \mu\mu$ events, $p^{\nu}_{x(y)} \equiv 0$ in equations (5.5) and (5.6).

5.5 Monte Carlo Samples Used

For the performance studies, the inclusive $Z \to ll$ and $W \to l\nu$ (where $l = e, \mu, \tau$) Monte Carlo (MC) samples were generated with the next-to-leading order (NLO)

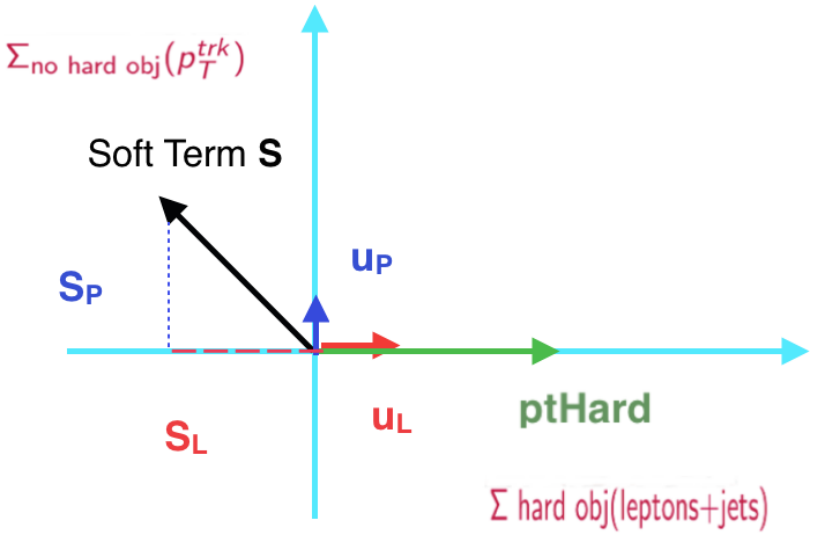


Figure 5-6: Projection of $p^{\text{miss}}_{\text{T,soft}}$ with respect to $\mathbf{p}^{\text{hard,trk}}_{\text{T}}$ direction to extract $p^{\text{miss}}_{\text{T,soft,L}}$ and $p^{\text{miss}}_{\text{T,soft,P}}$.

POWHEG and showered by PYTHIA using the CT10 NLO parton distribution function (PDF) set and the ATLAS AU2[13] tune. The generation of multi-leg processes such as $pp \rightarrow W/Z + n$ jets, with up to 6 partons in the final state, was done with ALPGEN+HERWIG. The $t\bar{t}$ production was done with the MC@NLO event generator. The description of the samples for the systematics evaluation and the application in the Higgs H $\rightarrow$ WW, $\tau\tau$ channels is given in the corresponding sections.

5.6 Results

The Track MET-Cl-j soft term systematic uncertainties results are split into two categories based on the track-jet association technique, namely results with ΔR and ghost association methods. During the LHC Run-I, the standard method feasible in Higgs analyses is the ΔR one and the corresponding results are presented first. The ghost association method was developed in preparation for Run-II. The results with this method are more stable and slightly more compatible with data than the ones obtained using ΔR. During Run-II, results are derived with the ghost association method by default[14].

5.6.1 7+8 TeV Results with ΔR Method

5.6.1.1 Lepton Selection

Leptons are called 'analysis' leptons if they passing the analysis specific selection criteria (e.g. p_{T} threshold), and are referred to as 'non-analysis' ones otherwise.

[13] A description of Monte Carlo generators, parton showers and tunes is given in chapter 3.
[14] The samples format required for ghost-association was available at the end of Run-I.

The lepton selection requirements are similar to the ones presented in sec. 5.3.3. In addition, a $|z_0 \sin\theta| < 1$ mm cut[15] is applied to non-analysis leptons in order to reduce lepton candidates from pile-up interactions. In the $H \to WW^* \to \ell\nu\ell\nu$, the p_{T} threshold is 25 GeV for the leading lepton and 22 GeV for the subleading one. These cuts are analysis specific.

5.6.1.2 Jet Selection

The jet selection requirements are similar to the ones presented in sec. 5.3.4, but with a higher jet p_{T} threshold (25 GeV if $|\eta| < 2.4$ and 30 GeV otherwise) compared the 20 GeV threshold used in other analyses [10]. The anti-k$_{\mathrm{T}}$ algorithm with $\Delta R = 0.4$ is used for jet reconstruction. Pileup corrections using the LCW+JES scheme are applied at the topo-cluster level and subsequent calibration levels.

5.6.1.3 Mis-reconstructed Track Removal

The procedure for mis-reconstructed track removal in $H \to WW^*$ analysis is different from what has been presented in sec. 5.3.2. Mis-reconstructed tracks do not enter the calculation of track MET and are removed based on the following steps:

- Tracks matched to jets passing the 10 GeV threshold ($p_T^{\mathrm{jet}} > 10$ GeV) within ΔR (track, jet) < 0.4 such that $\frac{p_T^{track}}{p_T^{jet}} > 1.4$ are removed.
- Tracks with $p_{\mathrm{T}} > 100$ GeV are removed if they do not match any reconstructed jet, muon or electron (tracks are matched to leptons within $\Delta R < 0.05$) or if they match the conditions above.

Removing mis-reconstructed tracks affects the nominal Track MET more than the object corrected Track MET.

5.6.1.4 Track MET-Cl-j Calculation

The ΔR method was implemented by the $H \to WW^*$ group and gives similar results to the track-jet matching with ghost association method that is applied in Track MET-Cl-j . The method corrects for the neutrals in analysis jets by replacing the contribution of tracks falling within a cone of ΔR=0.4 around the jet by the pileup corrected calorimeter jet p_{T}. The resulting Track MET-Cl-j is defined as follows:

1. In addition to the selection requirements defined in the sections above, mis-reconstructed tracks are removed as explained in sec. 5.6.1.3 and do not enter the Track MET-Cl-j calculations.
2. Match the tracks to the leptons based on ΔR and relative p_{T} matching. Then, replace electron tracks with cluster E_T and the muon ID track with the combined track:

[15] specific to $H \to WW^*$ analysis.

- Tracks matched to leptons should satisfy the $\Delta p_T/p_T = |(p_T^\ell - p_T^{\text{track}}|/p_T^\ell < 10^{-5}$ requirement, with ΔR(track, lepton track in ID) < 0.05.
- Tracks matched to combined (CB) muons with $\Delta p_T/p_T = |(p_T^\ell - p_T^{\text{track}}|/p_T^\ell < 10^{-5}$ and ΔR(track, CB muon) < 0.05 are removed and their contribution is replaced by the p_{T} as measured from CB muons algorithm.
- **Electron track-cluster correction:** Tracks matched to electron clusters within ΔR(track, electron cluster) < 0.05 are removed. The electron track p_{T} is replaced by the cluster p_{T} (Track MET-Cl definition).

3. Apply object overlap removal. This step is mainly needed to treat electrons overlap, electron-jet overlap and to remove all tracks associated with leptons (electrons and/or muons), keeping tracks associated to jets in addition to those not matched to any well measured physics object passing the analysis requirements. For example, a good well reconstructed jet with $20 < p_{\text{T}} < 25$ GeV does not pass the analysis p_{T} threshold, and its associated tracks enter thus the soft term. The overlap removal steps are summarized as follows:

   - In the case of multiple electrons overlapping with each other within $\Delta R < 0.10$, the one with the highest p_{T} is kept.
   - If an analysis electron and a jet overlap within $\Delta R < 0.3$, the electron is kept and the jet is discarded.
   - If a jet overlaps with a non-analysis electron within $\Delta R < 0.3$, the jet is kept since "medium"[16] non-isolated electrons are more likely to be jets mis-identified as electrons.

   Muons and jets, as well as muons and electrons There is no overlap removal done for muon-jet or muon-electron overlap for track MET calculations.

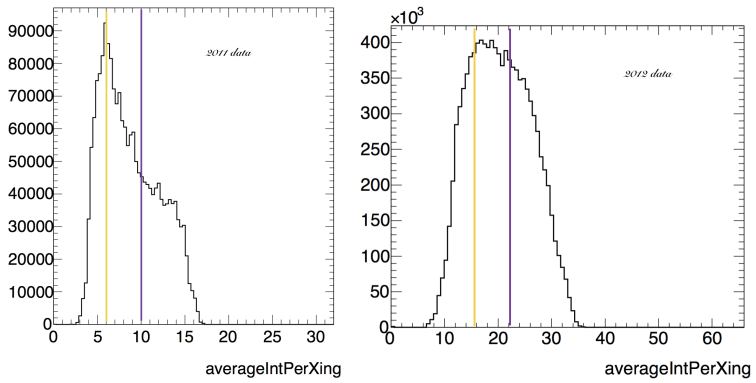
4. Find tracks associated to selected jets using the ΔR method as described in sec. 5.2.3, and replace their contribution to track MET by the pileup corrected (using jet area method defined in sec. 4.6.2.1) calorimeter jet p_{T} . A jet should pass the p_{T} threshold of 25 GeV for $|\eta| < 2.4$ and 30 GeV otherwise, in addition to the JVF[17] cut for central jets ($|\eta| < 2.4$): JVF > 0.5 if $p_T^{\text{jet}} < 50$ GeV.

5.6.1.5 Systematic Uncertainties

The corresponding systematic uncertainties are evaluated according to the method in sec. 5.4, and the results for both Track MET-Cl and Track MET-Cl-j used in the $H \to WW^*$ analysis are shown. MC simulation samples used are AlpgenHerwig $Z \to \mu\mu$ events. And the $\langle\mu\rangle$ binning choice used for the uncertainties evaluations in 2011 and 2012 is illustrated in Fig. 5-7, with the limits adjusted to accommodate the

[16] The electron identification criteria are described in detail in sec. 4.4.

[17] JVF stands for Jet Vertex Fraction. This is the fraction estimated from tracks by which the jet p_{T} is corrected for pile-up effects as explained in sec. 4.6.2.3.



(a) (b)

Figure 5-7: Distribution of the average number of interactions per crossing $\langle\mu\rangle$ in $Z \to \mu\mu$ events at $\sqrt{s} = 7$ TeV(left) and 8 TeV (right). The dashed lines show the limits defining the binning in $\langle\mu\rangle$ used for Track MET-Cl-j calculations.

differences in pile-up conditions between 2011 and 2012. Based on the distribution in Fig. 5-7, the 2011 $\langle\mu\rangle$ bins are chosen in a way to ensure statistically significant samples in each of them: [0-6, 6-10, $\geq$ 10]. Results are also split based on jet multiplicity (0-jet, 1-jet and ≥ 2 jets bins) and in bins of $p_{\mathrm{T}}^{\mathrm{hard}}$.

The results with Track MET-Cl are illustrated in Fig. 5-8. The total systematic uncertainty is 30% on the track MET values. The pile-up dependence is very small. As expected, scale increases as a function of $p_T^{\mathrm{hard}}(\equiv p_T^{\mathrm{lep}}$ for this track MET definition). The results with Track MET-Cl-j are illustrated in Fig. 5-9 and 5-10. The change in the p_T^{hard} binning seen when comparing Fig. 5-8 and 5-9 or 5-10 is due to the change in the p_T^{hard} distribution between the Track MET-Cl and Track MET-Cl-j definitions.

As shown in Fig. 5-11, the Track MET-Cl p_T^{hard} binning [0-30, 30-60, 60-200] GeV is not suitable for Track MET-Cl-j. Consequently, in order to have more meaningful results from a statistically significant sample, this binning used for Track MET-Cl inclusive and 1-jet bin samples is thus replaced by [0-10, 10-25, 25-80] GeV for Track MET-Cl-j . This binning is already used for 0-jet $Z \to \mu\mu$ events, resulting in a uniform systematic uncertainties derivation amongst all jet bins.

The 2011 systematic results at 7 TeV are also compatible with the 2012 results as shown in Fig. 5-12 and 5-13. the resolution of $p_{\mathrm{T,soft,P}}^{\mathrm{miss}}$ is assigned for the 1 jet bin with data/MC ratio = 1.11. Given the different pile-up conditions in 2011, the binning in $\langle\mu\rangle$, the average number of interactions per crossing, is adjusted accordingly.

5.6.1.6 Closure Test

The 6 different systematic variations on the scale (up/down) and resolution (parallel: up/down and perpendicular: up/down) are applied to $Z \to \mu\mu$ samples to check if the data/MC discrepancies are well covered by these variations. Then, the same procedure is applied to $Z \to ee$ samples using the systematic uncertainties derived in $Z \to \mu\mu$ events as a cross-check. The distributions of $p_{\mathrm{T,soft}}^{\mathrm{miss}}$ varied according to (5.7)

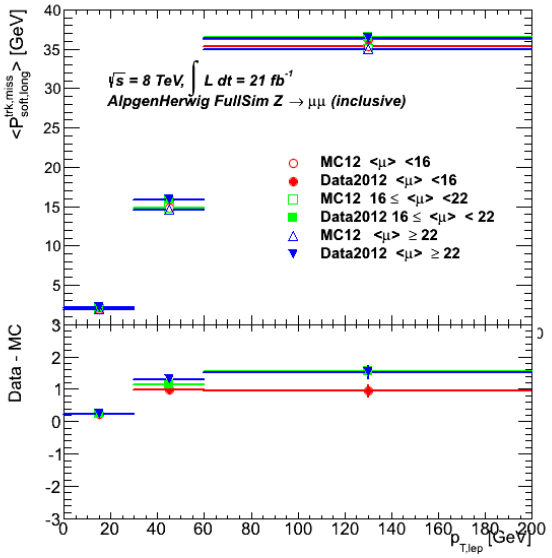
are plotted in Fig. 5-14(a). The corresponding $p_{\mathrm{T}}^{\mathrm{miss}}$ distributions, i.e. after adding the $p_{\mathrm{T}}^{\mathrm{miss}}$ contributions of the leptons and the jet(s), are shown in Fig. 5-21(b). The results for $Z \to \mu\mu$ events in Fig. 5-14 show good data-MC agreement for $p_{\mathrm{T,soft}}^{\mathrm{miss}}$ distributions. The data/MC ratio at small $p_{\mathrm{T,soft}}^{\mathrm{miss}}$ values, corresponding to small p_{T} of (Z+jet(s)), is expected due to the $Z \to \ell\ell$ mis-modeling in Alpgen, where MC cross sections need to be reweighted [11, 12]. When the leptons and jet(s) are added to $p_{\mathrm{T,soft}}^{\mathrm{miss}}$, the data agrees well with the MC predictions at low $p_{\mathrm{T}}^{\mathrm{miss}}$ values. The data/MC discrepancies are well covered by the systematic variations up to $p_T^{\mathrm{hard}} = 70$ GeV. The disagreement is resolved when adding the associated systematic uncertainties of the hard objects in the event. Applying $Z \to \mu\mu$ systematics to $Z \to ee$ events as shown in Fig. 5-15, data-MC comparisons show a behavior close to that of $Z \to \mu\mu$ samples: there is a good data-MC agreement without any real dependence on pileup.

5.6.1.7 Soft Term Comparisons

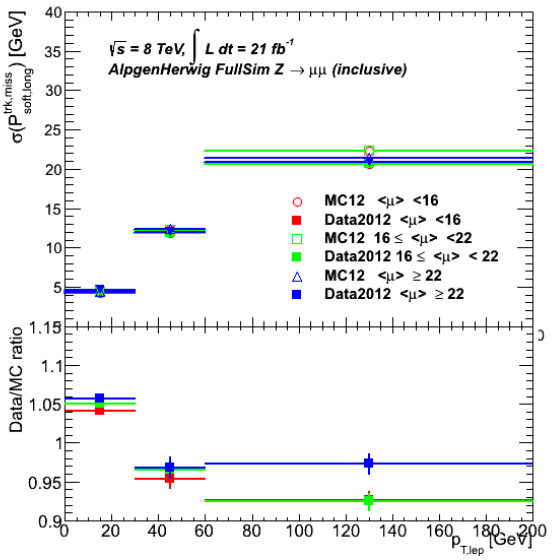
The almost absence of any pile-up dependence effect is not only observed at the systematic uncertainty level but also when calculating the soft term. In fact, $\sum E_T$ distributions in $Z \to \mu\mu$ samples, shown in Fig. 5-16 for different pile-up conditions, show no significant pile-up dependence. The data/MC discrepancies are explained by mis-modeling of the Z+jets events cross sections by MC generators. Events generated by Alpgen or Powheg for example require some reweighting as explained in [11, 12].

The results show no real pile-up dependence. And the data/MC differences are well explained by the Z+jets cross section modeling problem by MC generators, which requires some reweighting.

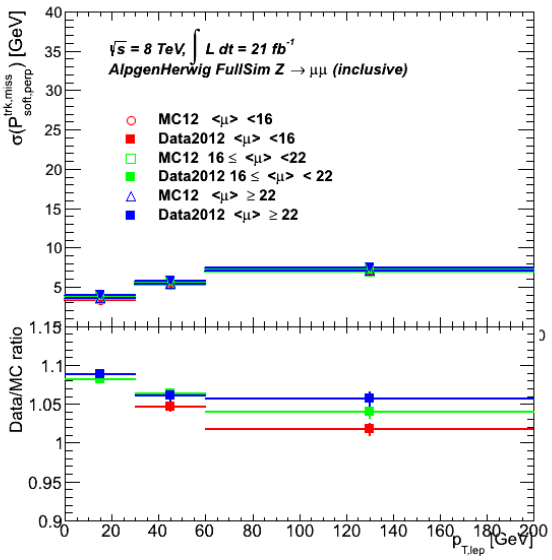
Moreover, the comparison of the soft term 'longitudinal' and 'perpendicular' components between Track MET-Cl-j and the calorimeter based MET (calo MET) definition shows similar effects. The 2D distributions of ($p_{\mathrm{T,soft,L}}^{\mathrm{miss}}$, $\sum E_T^{\text{soft term}}$) in $Z \to \mu\mu$ inclusive samples in Fig. 5-17 show that most events have small $p_{\mathrm{T,soft,L}}^{\mathrm{miss}}$ values.



(a)

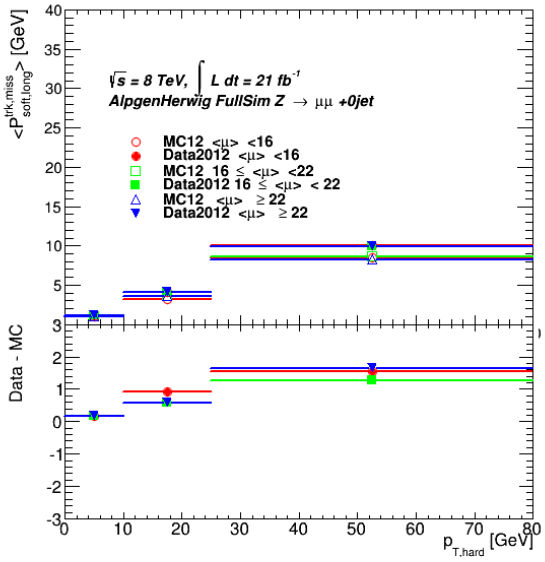


(b)

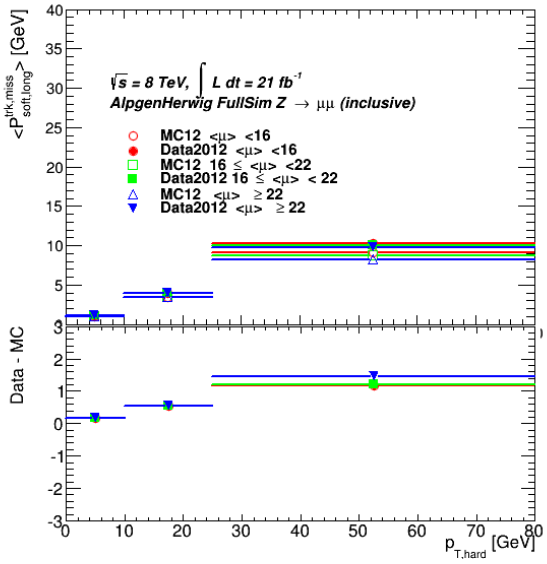


(c)

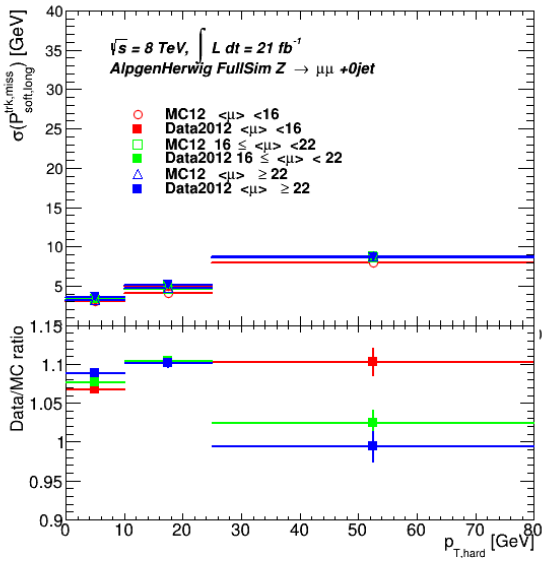
Figure 5-8: Scale and resolution results for Track MET-Cl in the inclusive $Z \rightarrow \mu\mu$ samples. The scale uncertainty defined as the difference(shift) between data and MC $p^{\text{miss}}_{\text{T,soft,L}}$ is plotted in the lower panel of (a). The resolution uncertainty defined as the ratio of data to MC resolution values is plotted for $p^{\text{miss}}_{\text{T,soft,L}}$ and $p^{\text{miss}}_{\text{T,soft,P}}$ in the lower panels of (b) and (c) respectively.



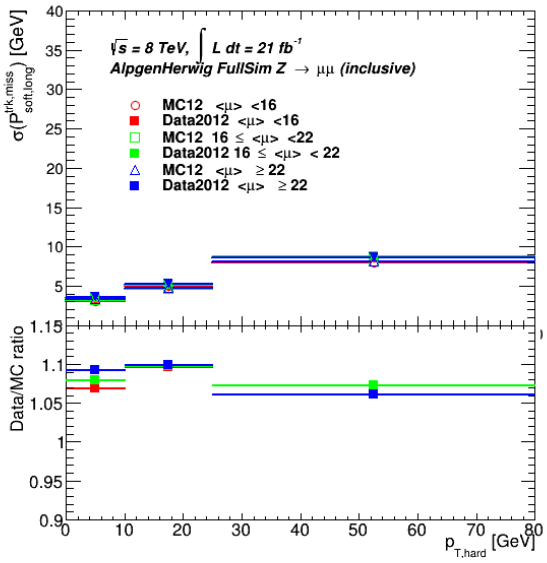
(a) Mean of $p^{\text{miss}}_{\text{T,soft,L}}$ in the 0-jet bin



(b) Mean of $p^{\text{miss}}_{\text{T,soft,L}}$ in the inclusive-jet bin

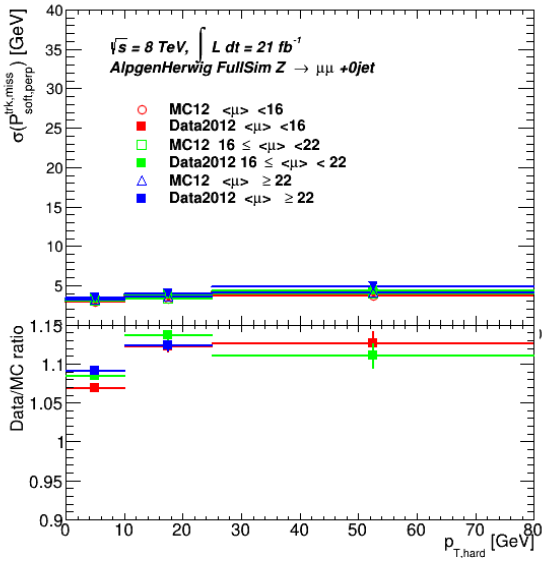


(c) Resolution of $p^{\text{miss}}_{\text{T,soft,L}}$ in the 0-jet bin

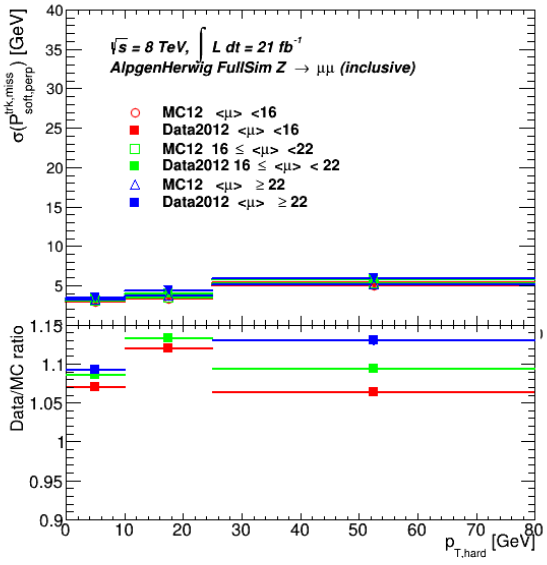


(d) Resolution of $p^{\text{miss}}_{\text{T,soft,L}}$ in the inclusive-jet bin

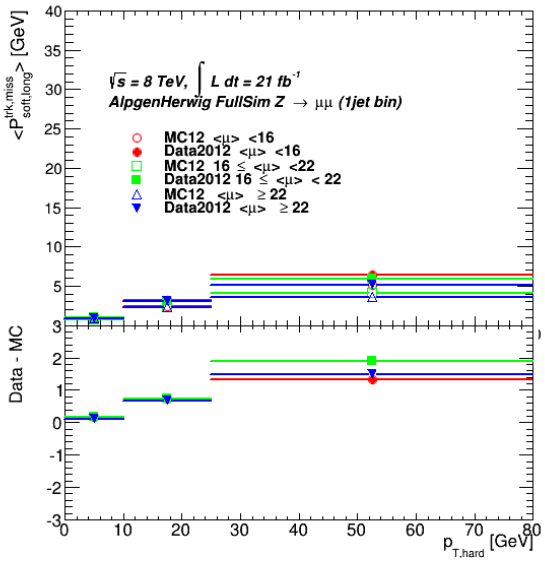
Figure 5-9: Mean and resolution of $p^{\text{miss}}_{\text{T,soft,L}}$ components of Track MET-Cl-j in $Z \to \mu\mu$ 0-jet(left) and inclusive(right) samples respectively using the ΔR method



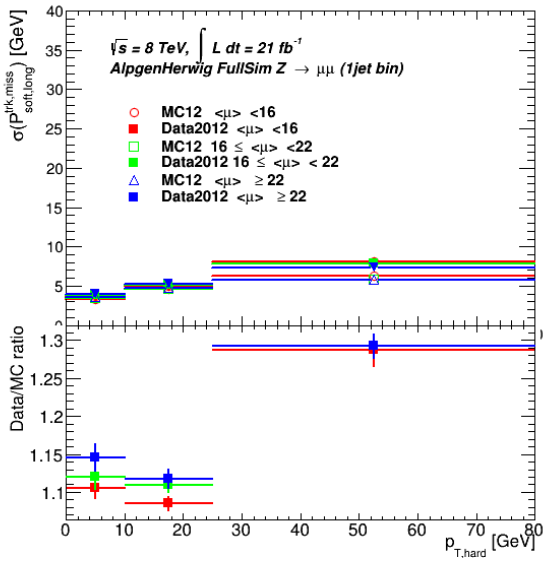
(a) Resolution of $p^{\text{miss}}_{\text{T,soft,P}}$ in the 0-jet bin



(b) Resolution of $p^{\text{miss}}_{\text{T,soft,P}}$ in the inclusive-jet bin



(c) Mean of $p^{\text{miss}}_{\text{T,soft,L}}$ in the 1-jet bin



(d) Resolution of $p^{\text{miss}}_{\text{T,soft,L}}$ in the 1-jet bin

Figure 5-10: Mean and resolution of $p^{\text{miss}}_{\text{T,soft,L}}$ components of Track MET-Cl-j in $Z \to \mu\mu$ 1-jet sample (bottom) and $p^{\text{miss}}_{\text{T,soft,P}}$ component(top) using the ΔR method.

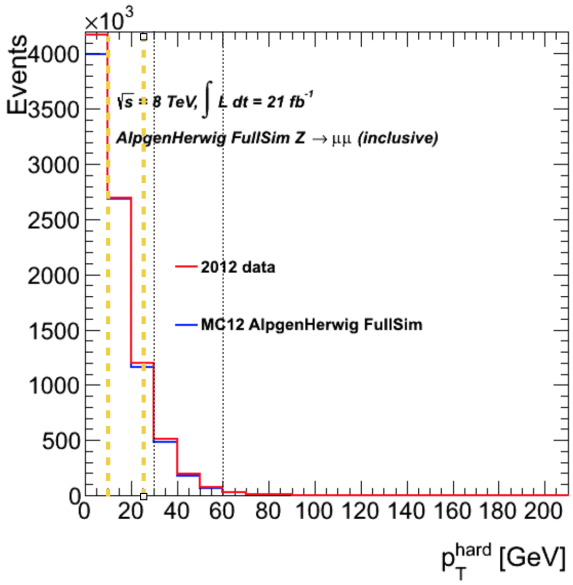
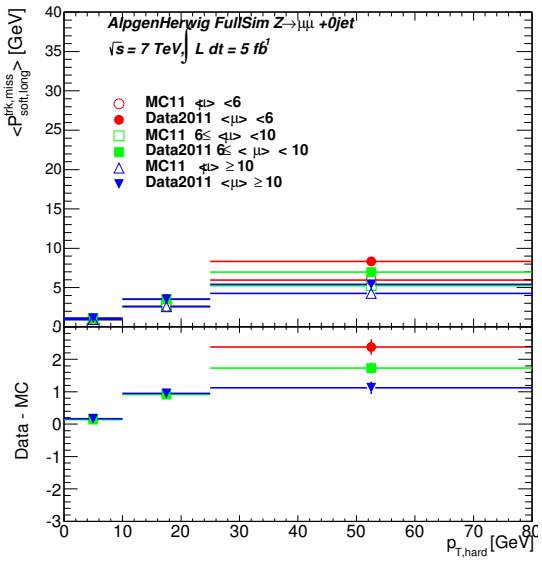
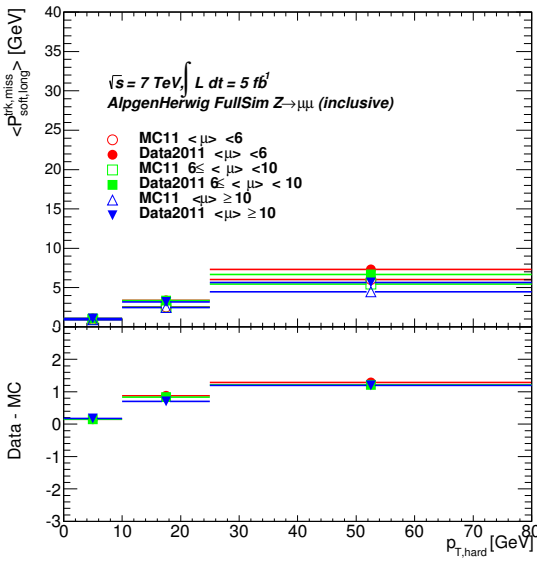


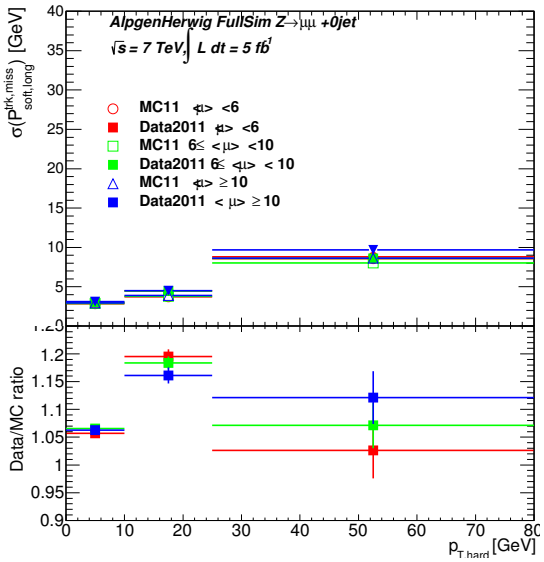
Figure 5-11: p_T^{hard} distribution in $Z \to \mu\mu$ inclusive sample for Track MET-Cl-j definition. The dotted lines show the binning limits as used for Track MET-Cl and the yellow dashed lines show the adjusted binning for the Track MET-Cl-j calculations.



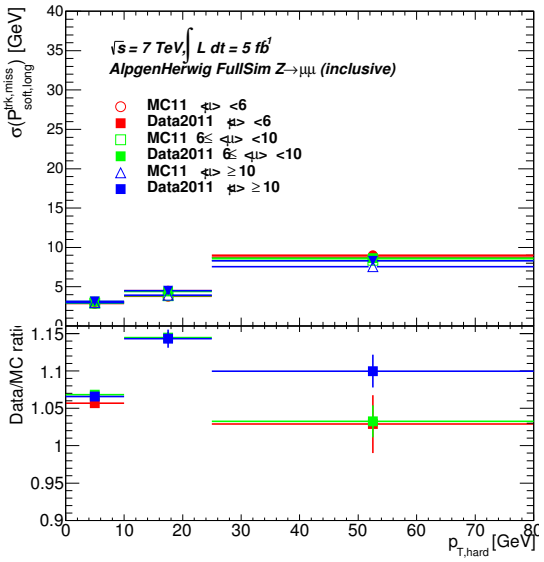
(a) Mean of $p_{\mathrm{T,soft,L}}^{\mathrm{miss}}$ in the 0-jet bin



(b) Mean of $p_{\mathrm{T,soft,L}}^{\mathrm{miss}}$ in the inclusive-jet bin

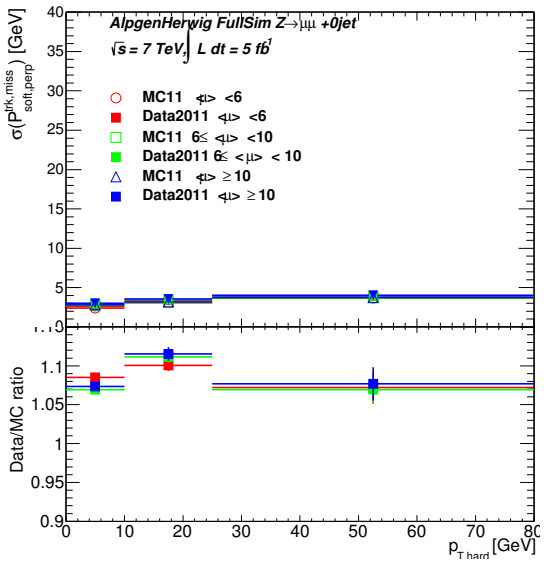


(c) Resolution of $p_{\mathrm{T,soft,L}}^{\mathrm{miss}}$ in the 0-jet bin

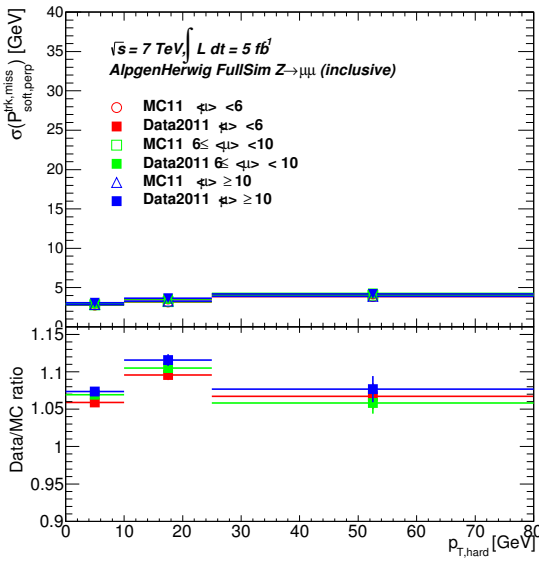


(d) Resolution of $p_{\mathrm{T,soft,L}}^{\mathrm{miss}}$ in the inclusive-jet bin

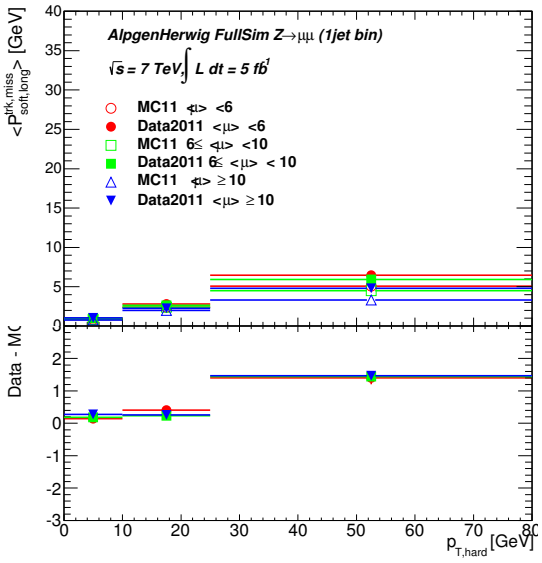
Figure 5-12: Mean and resolution of $p_{\mathrm{T,soft,L}}^{\mathrm{miss}}$ components of Track MET-Cl-j in $Z \to \mu\mu$ 0-jet(left) and inclusive(right) samples respectively using the ΔR method at $\sqrt{s} = 7$ TeV.



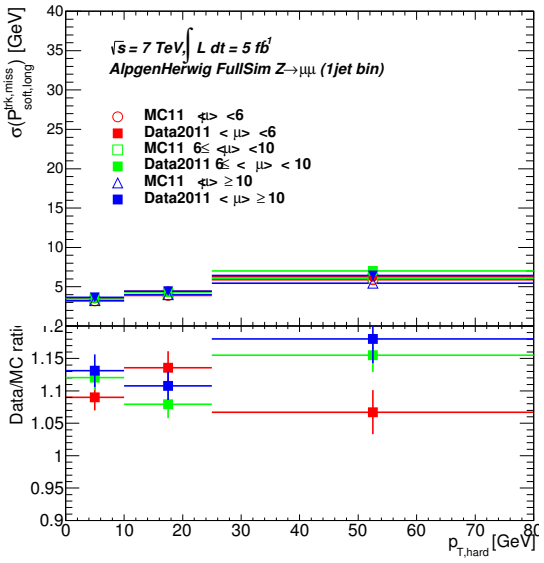
(a) Resolution of $p_{\mathrm{T,soft,P}}^{\mathrm{miss}}$ in the 0-jet bin



(b) Resolution of $p_{\mathrm{T,soft,P}}^{\mathrm{miss}}$ in the inclusive-jet bin

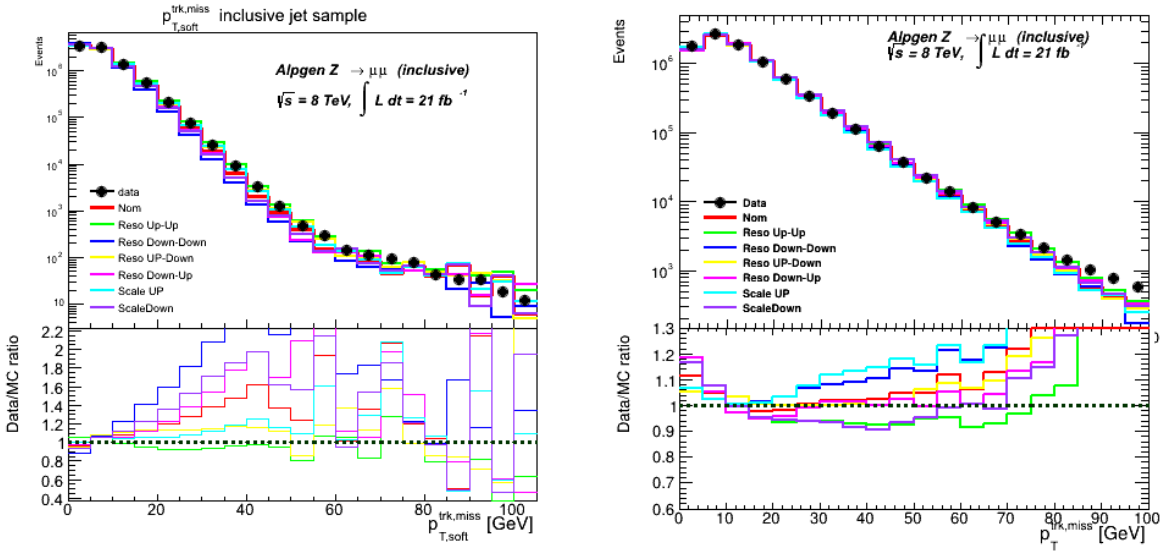


(c) Mean of $p_{\mathrm{T,soft,L}}^{\mathrm{miss}}$ in the 1-jet bin



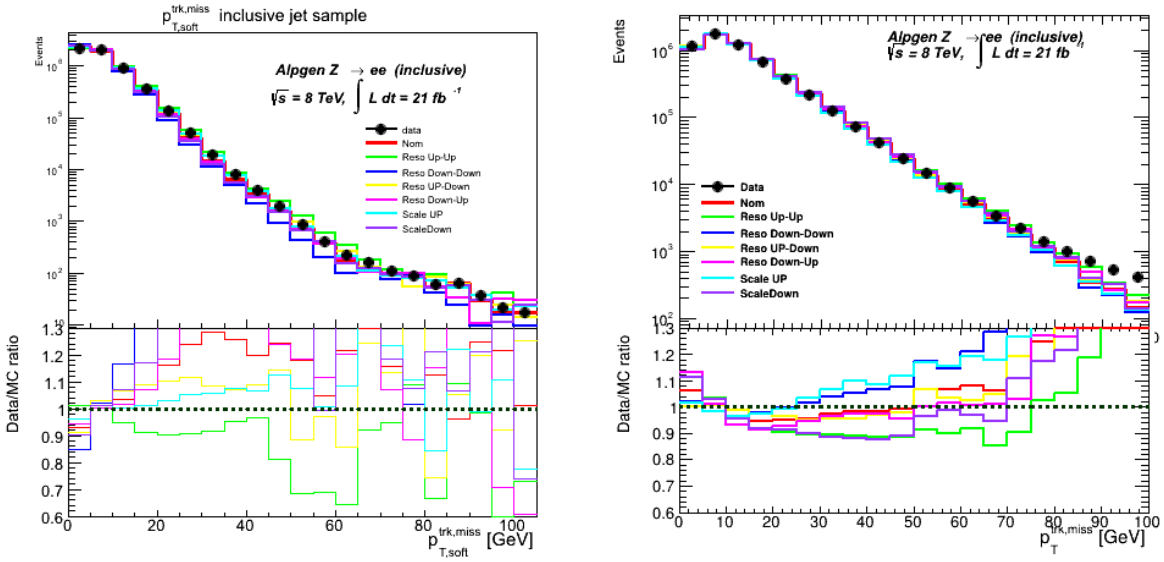
(d) Resolution of $p_{\mathrm{T,soft,L}}^{\mathrm{miss}}$ in the 1-jet bin

Figure 5-13: Mean and resolution of $p_{\mathrm{T,soft,L}}^{\mathrm{miss}}$ components of Track MET-Cl-j in $Z \to \mu\mu$ 1-jet sample (bottom) and $p_{\mathrm{T,soft,P}}^{\mathrm{miss}}$ component(top) using the ΔR method at $\sqrt{s} = 7$ TeV.



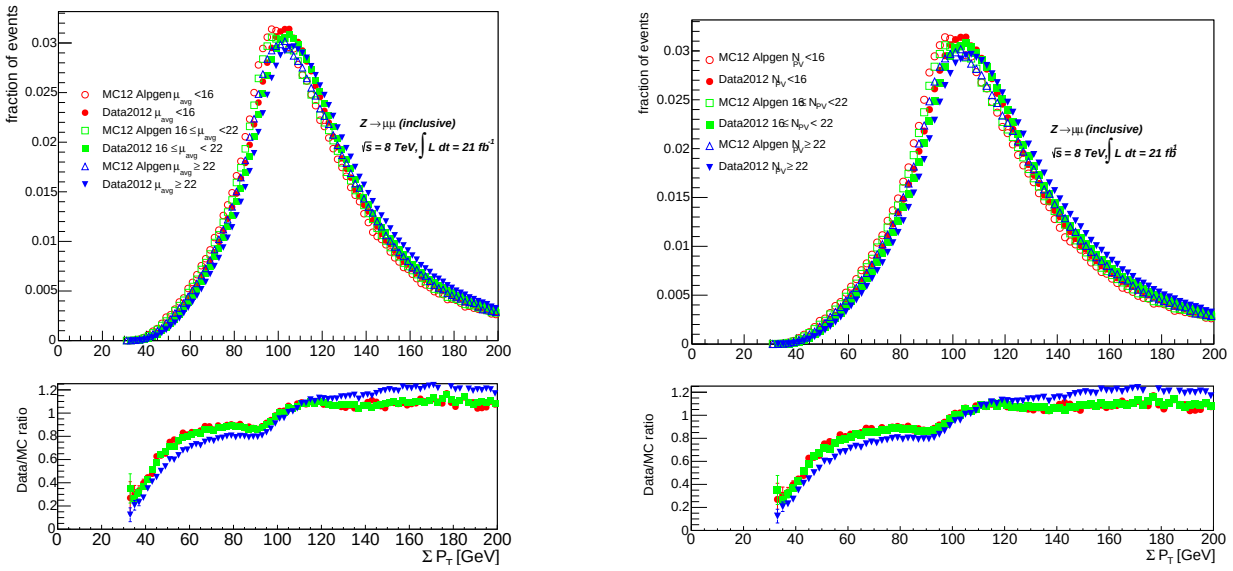
(a) (b)

Figure 5-14: Distributions of $p_{\mathrm{T,soft}}^{\mathrm{miss}}$ varied according to equation (5.7) are shown in (a) along with the corresponding total $p_{\mathrm{T}}^{\mathrm{miss}}$ distributions (b) for the $Z \to \mu\mu$ inclusive sample. The results are derived for Track MET-Cl-j derived using ΔR method at $\sqrt{s} = 8$ TeV.



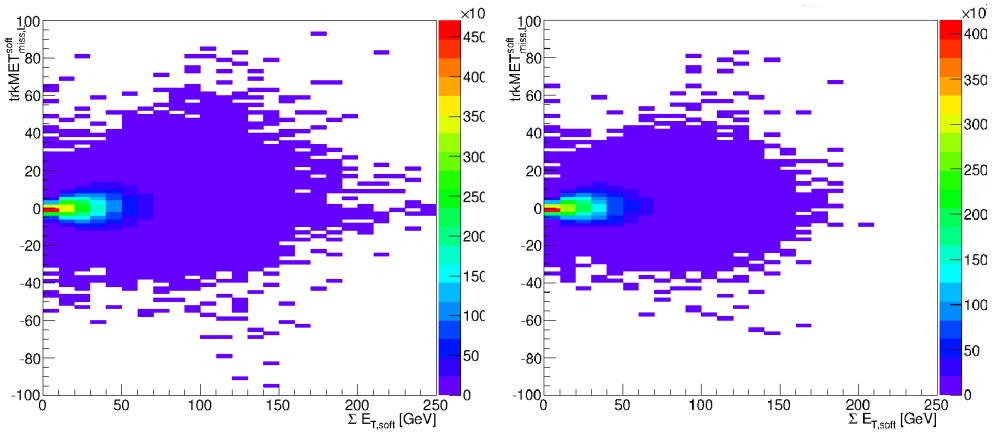
(a) (b)

Figure 5-15: Distributions of $p_{\mathrm{T,soft}}^{\mathrm{miss}}$ varied according to equation (5.7) are shown in (a) along with the corresponding total $p_{\mathrm{T}}^{\mathrm{miss}}$ distributions (b) for the $Z \to ee$ inclusive sample. The results are derived for Track MET-Cl-j derived using ΔR method at $\sqrt{s} = 8$ TeV.



(a) (b)

Figure 5-16: $\sum E_T$ comparisons for different pile-up conditions at $\sqrt{s} = 8$ TeV in $Z \to \mu\mu$ data and AlpgenHerwig simulated samples. The pile-up modeling is in fact studied in terms of the average number of interactions per crossing $\langle\mu\rangle$ (a) and of the number of reconstructed primary vertices in the event N_{PV} (b).



(a) (b)

Figure 5-17: Distributions of ($p^{\mathrm{miss}}_{\mathrm{T,soft,L}}$,$\sum E_T^{\text{soft term}}$) in the inclusive (left) and 0-jet bin (right) $Z \to \mu\mu$ events.

5.6.2 Track MET-Cl-j Systematic Uncertainties Using the Ghost Association Method

The systematic uncertainties of $p_{\mathrm{T,soft}}^{\mathrm{miss}}$ using $Z \to \mu\mu$ events are estimated for events in the 0-jet bin, 1-jet bin and the inclusive sample where the jet p_{T} cut is 25 GeV. The inclusive case is used to have a general systematic uncertainties evaluation of the $Z \to \mu\mu$ sample taking into account the small contribution of Z events with at least 2 jets. The dominant contribution to the inclusive sample calculations, whether from the 0 jet bin or the 1-jet bin, depends on the region of $p_{\mathrm{T}}^{\mathrm{hard}}$ in question. All MC results were derived using AlpgenHerwig $Z \to \mu\mu$ samples with the jet corrections being applied to $p_{\mathrm{T}}^{\mathrm{miss}}$ .The mean of the perpendicular component $\langle p_{\mathrm{T,soft,P}}^{\mathrm{miss}} \rangle$ is consistent with zero in all cases. Figures 5-18 and 5-19 show good data-MC agreement for the mean and resolution of $p_{\mathrm{T,soft}}^{\mathrm{miss}}$ components for different jet multiplicities separately, and no significant dependence on the average number of interactions per crossing $\langle \mu \rangle$ for all $p_{\mathrm{T}}^{\mathrm{hard}}$ bins regardless of the jet multiplicity. In fact, these results for Track MET-Cl-j systematic uncertainties are the same for different jet multiplicities as expected, given that the physics describing whatever recoils against the Z+jet(s) system is similar. The results for the soft term, as defined for Track MET-Cl-j , are thus valid for any jet bin and are pile-up independent. In addition, the mean and resolution of the longitudinal component increase as a function of $p_{\mathrm{T}}^{\mathrm{hard}}$. As the p_{T} of the Z+jet(s) increases, the p_{T} of the system recoiling against them gets higher as required by the p_{T} balance. As a result, the projection of the p_{T} of the system recoiling against the Z+jet(s), i.e. the mean or 'scale' of the soft term, has automatically higher values. For the 1-jet bin, the resolution of $p_{\mathrm{T,soft,P}}^{\mathrm{miss}}$ is almost independent of $p_{\mathrm{T}}^{\mathrm{hard}}$ and $\langle \mu \rangle$ values as Fig. 5-20 shows. This effect is seen in different jet η (central i.e. $|\eta| < 2.4$ and forward) regions, so a single value for the resolution of $p_{\mathrm{T,soft,P}}^{\mathrm{miss}}$ is assigned for the 1 jet bin with a data/MC ratio = 1.118 ± 0.005. These results are similar to the ones obtained using the ΔR method (Fig. 5-9 to 5-10) in sec. 5.6.1.5.

The distributions of $p_{\mathrm{T,soft}}^{\mathrm{miss}}$ varied according to (5.7) are plotted in Fig. 5-21(a). The corresponding $p_{\mathrm{T}}^{\mathrm{miss}}$ distributions, i.e. after adding the $p_{\mathrm{T}}^{\mathrm{miss}}$ contributions of the leptons and the jet(s), are shown in Fig. 5-21(b). The results for $Z \to \mu\mu$ events in Fig. 5-21 show good data-MC agreement when $p_{\mathrm{T}}^{\mathrm{miss}}$ contributions of the leptons and jet(s) are added to $p_{\mathrm{T,soft}}^{\mathrm{miss}}$.

5.6.3 Extrapolating Beyond $Z \to ll$ Events

Since the mean and resolution parametrization in terms of bins in $p_{\mathrm{T}}^{\mathrm{hard}}$ and $\langle \mu \rangle$ takes into account the response of the $p_{\mathrm{T,soft}}^{\mathrm{miss}}$ in different event topologies, then the parametrization determined from $Z \to \mu\mu$ events can be used to evaluate the systematic uncertainties on $p_{\mathrm{T,soft}}^{\mathrm{miss}}$ in other MC samples too. To propagate systematic uncertainties from $Z \to \mu\mu$ samples to other samples, $p_{\mathrm{T,soft}}^{\mathrm{miss}}$ and $p_{\mathrm{T}}^{\mathrm{hard}}$ equations as defined in (5.5) and (5.6) are applied, where the neutrinos are added to $p_{\mathrm{T}}^{\mathrm{hard}}$ if high-p_{T} neutrinos are expected in the event of the MC sample e.g. in the case of W events.

Figure 5-22 shows the distribution of the difference between Track MET-Cl-j and the true missing transverse energy originating from the neutrino in simulated Alpgen-Pythia $W \rightarrow \mu\nu$ samples at $\sqrt{s} = 8$ TeV. The results are plotted for events with an average number of interactions per crossing $\langle\mu\rangle$ greater than 22. The difference is to be compared to the resolution uncertainties on $p^{\text{miss}}_{\text{T,soft,L}}$ in the 3$^{\text{rd}}$ bin in $\langle\mu\rangle$ in Fig. 5-18 and 5-19. The obtained mean value is zero, showing a good estimation of the missing transverse energy in the event. The RMS of 9.7 GeV is compatible with the resolution values of ∼10 GeV obtained from $Z \rightarrow \mu\mu$ data and MC events used for systematics derivation.

5.6.4 Influence of Hard Term on $p^{\text{miss}}_{\text{T,soft}}$ Systematics

In the previous sections, the systematic uncertainties on $p^{\text{miss}}_{\text{T,soft}}$ were evaluated assuming a perfect knowledge of $p^{\text{hard}}_{\text{T}}$. The latter, however, plays an important role when defining $p^{\text{miss}}_{\text{T,soft}}$. Therefore, for a complete estimation of $p^{\text{miss}}_{\text{T,soft}}$ systematic uncertainties, effects from the systematic uncertainties on the hard physics objects, i.e. jets and leptons in $p^{\text{hard}}_{\text{T}}$, should be included. The dominant contribution stems from JES/JER (jet energy scale/jet energy resolution) uncertainties, whereas the lepton uncertainties are smaller.
Fig. 5-23 shows the results for the $p^{\text{miss}}_{\text{T,soft,L}}$ scale uncertainty, denoted by p^{soft}_{T}long.shift, and resolution uncertainty denoted by p^{soft}_{T}long.smear, taking into account the effects of MC generators and parton shower model (discussed in the next section), pile-up, JES/JER variations in addition to the forward/central jet treatment difference. The obtained envelop is used to assign the final error on the systematic uncertainties values. The largest variation comes from the parton shower model. The effects on the $p^{\text{miss}}_{\text{T,soft,P}}$ resolution are shown as well.

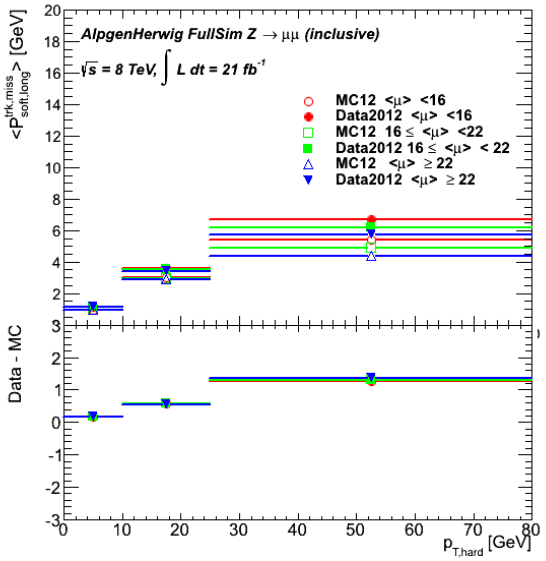
5.6.5 Generator and Simulation Dependence

To study the scale and resolution generator dependence, $Z \rightarrow ll$ samples of Algen-Herwig, PowhegPythia8 and Sherpa have been used for the comparison. Figures 5-8 and 5-24 to 5-27 show the scale and resolution uncertainties for Track MET-Cl and calorimeter MET (calo MET for short) definitions. Both calo and Track MET-Cl results have ∼no dependence on $\langle\mu\rangle$. Results can be summarized as follows:

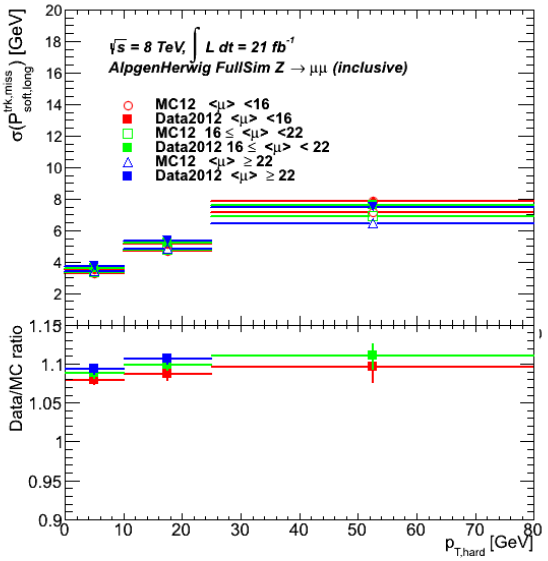
- AlpgenHerwig shows positive scale corrections.
- However, unlike Alpgen, the data-MC differences (scale corrections) are negative for PowhegPythia8 and consistent with zero for Sherpa.
- Highest resolution uncertainties are obtained for Alpgen and the smallest are achieved with Sherpa. Results with Sherpa show that the resolution on the parallel and perpendicular components is almost the same for calo MET before pileup suppression. For Track MET-Cl , the resolution on the perpendicular component is very small compared to the longitudinal one and has a smaller p^{lep}_{T} dependence.

- The results for each generator are almost the same whether we use ATLAS Fast (AFII) or full simulation (FullSim) samples. Comparing FullSim and AFII Sherpa samples, the systematic uncertainties remained unchanged thus giving no dependence.

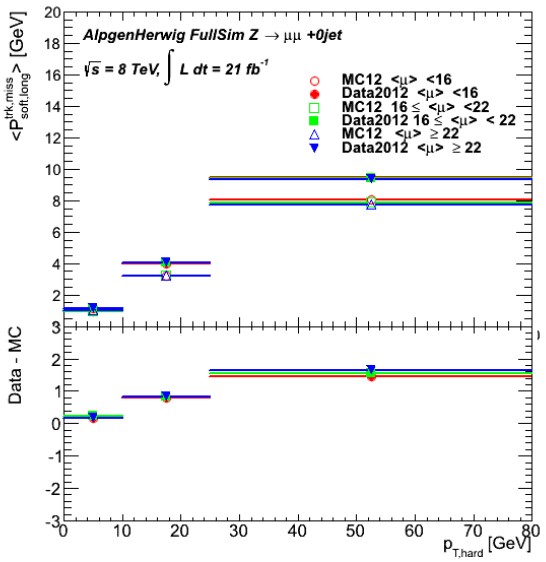
In order to take into account the different systematic uncertainties of various MC generators, a conservative approach was adopted where the envelop covering the largest variations is used to assign the final uncertainties.



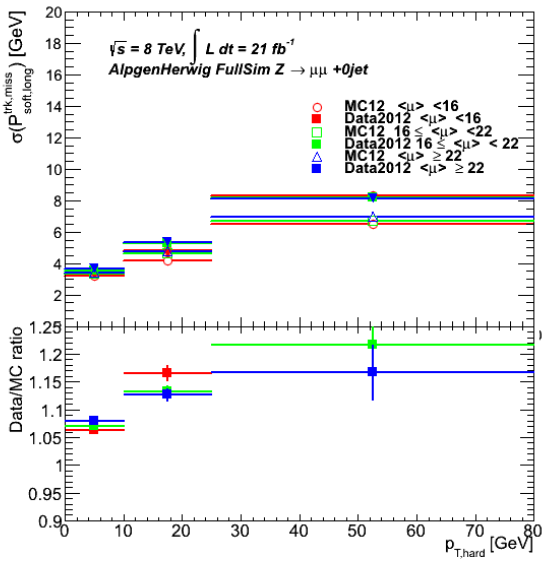
(a) Mean of $p^{\text{miss}}_{\text{T,soft,L}}$ in the inclusive sample



(b) Resolution of $p^{\text{miss}}_{\text{T,soft,L}}$ in the inclusive sample

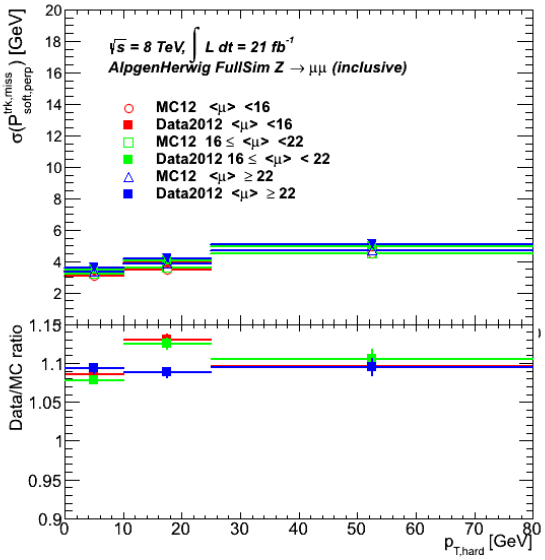


(c) Mean of $p^{\text{miss}}_{\text{T,soft,L}}$ in the 0-jet bin

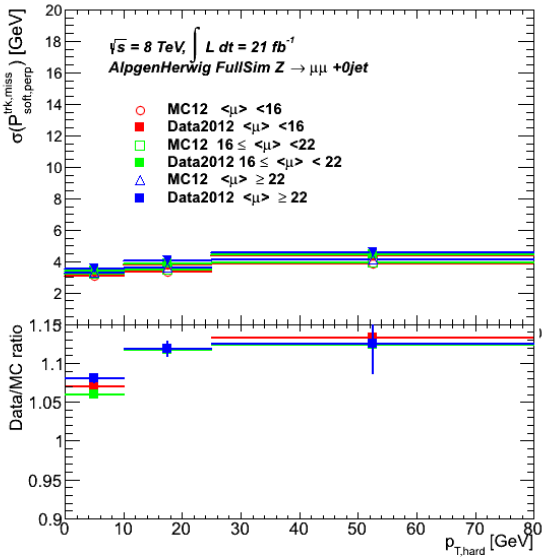


(d) Resolution of $p^{\text{miss}}_{\text{T,soft,L}}$ in the 0-jet bin

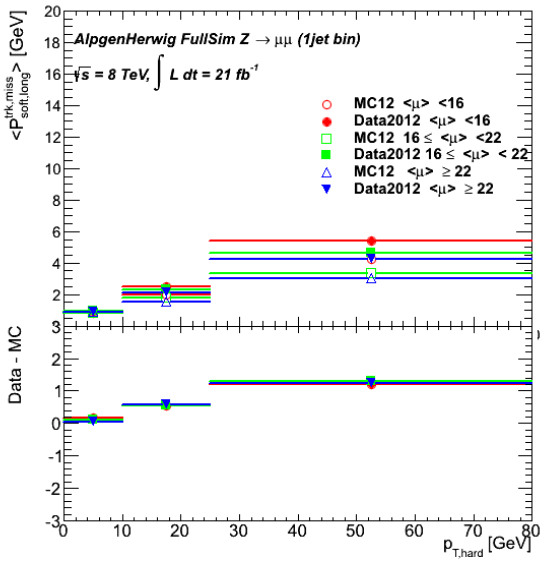
Figure 5-18: Mean and resolution of $p^{\text{miss}}_{\text{T,soft,L}}$ component of Track MET-Cl-j in $Z \to \mu\mu$ inclusive (top) and 0-jet (bottom) samples. The results are derived for Track MET-Cl-j using the ghost association method.



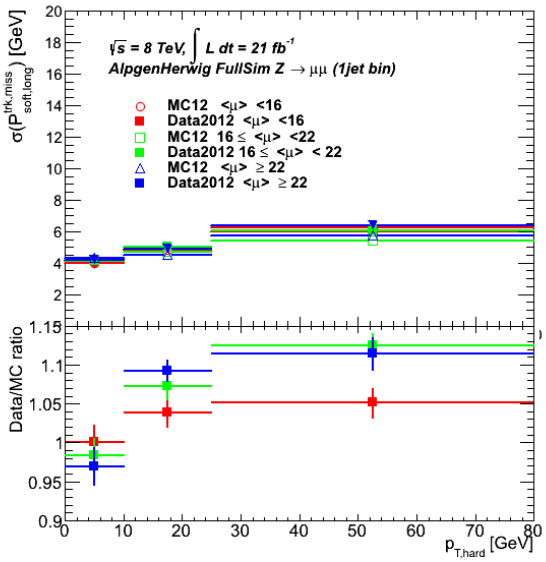
(a) Resolution of $p^{\text{miss}}_{\text{T,soft,P}}$ in the inclusive sample



(b) Resolution of $p^{\text{miss}}_{\text{T,soft,P}}$ in the 0-jet bin



(c) Mean of $p^{\text{miss}}_{\text{T,soft,L}}$ in the 1-jet bin



(d) Resolution of $p^{\text{miss}}_{\text{T,soft,L}}$ in the 1-jet bin

Figure 5-19: Resolution of $p^{\text{miss}}_{\text{T,soft,P}}$ component of Track MET-Cl-j in $Z \to \mu\mu$ events in the inclusive (left) and 0-jet bin(right). The mean and resolution of $p^{\text{miss}}_{\text{T,soft}}$ components of Track MET-Cl-j in $Z \to \mu\mu$ events in the 1-jet bin are shown in the bottom plots. The results are derived for Track MET-Cl-j using the ghost association method.

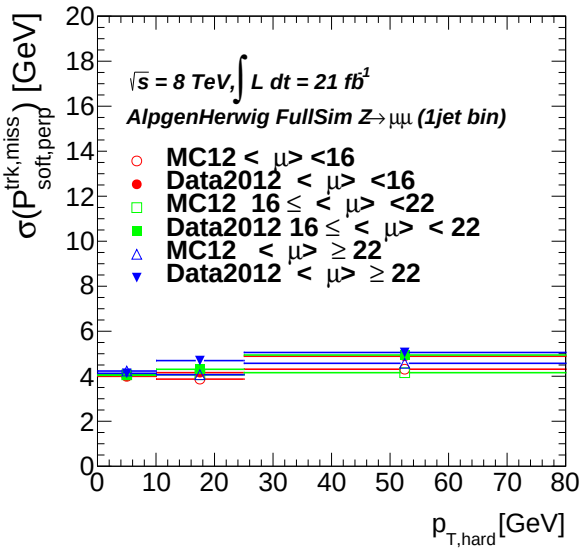
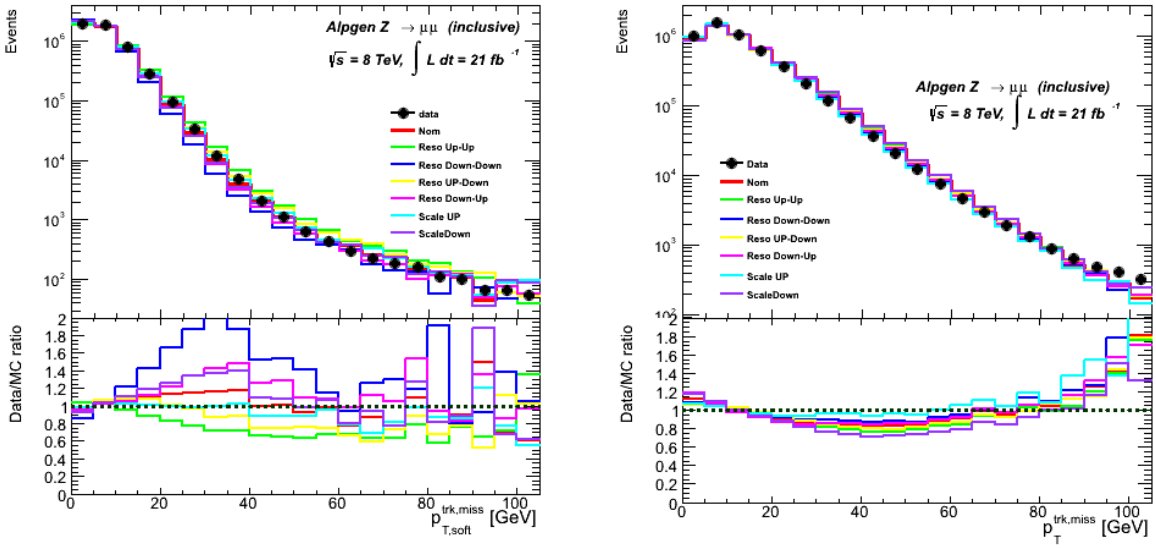


Figure 5-20: Resolution of $p^{\text{miss}}_{\text{T,soft,P}}$ in $Z \to \mu\mu$ 1-jet bin derived for Track MET-Cl-j using the ghost association method.



(a) (b)

Figure 5-21: Distributions of $p^{\text{miss}}_{\text{T,soft}}$ varied according to equation (5.7) are shown in (a) along with the corresponding total $p^{\text{miss}}_{\text{T}}$ distribution in (b) for the $Z \to \mu\mu$ inclusive sample. The results are derived for Track MET-Cl-j derived using the ghost association method.

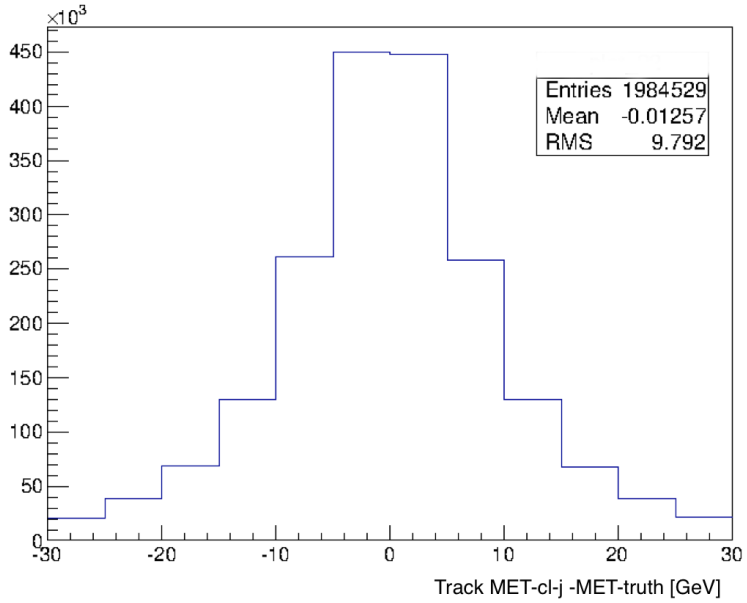


Figure 5-22: Plot showing the difference between Track MET-Cl-j and the true missing transverse energy originating from the neutrino in simulated AlpgenPythia $W \to \mu\nu$ samples at $\sqrt{s} = 8$ TeV.

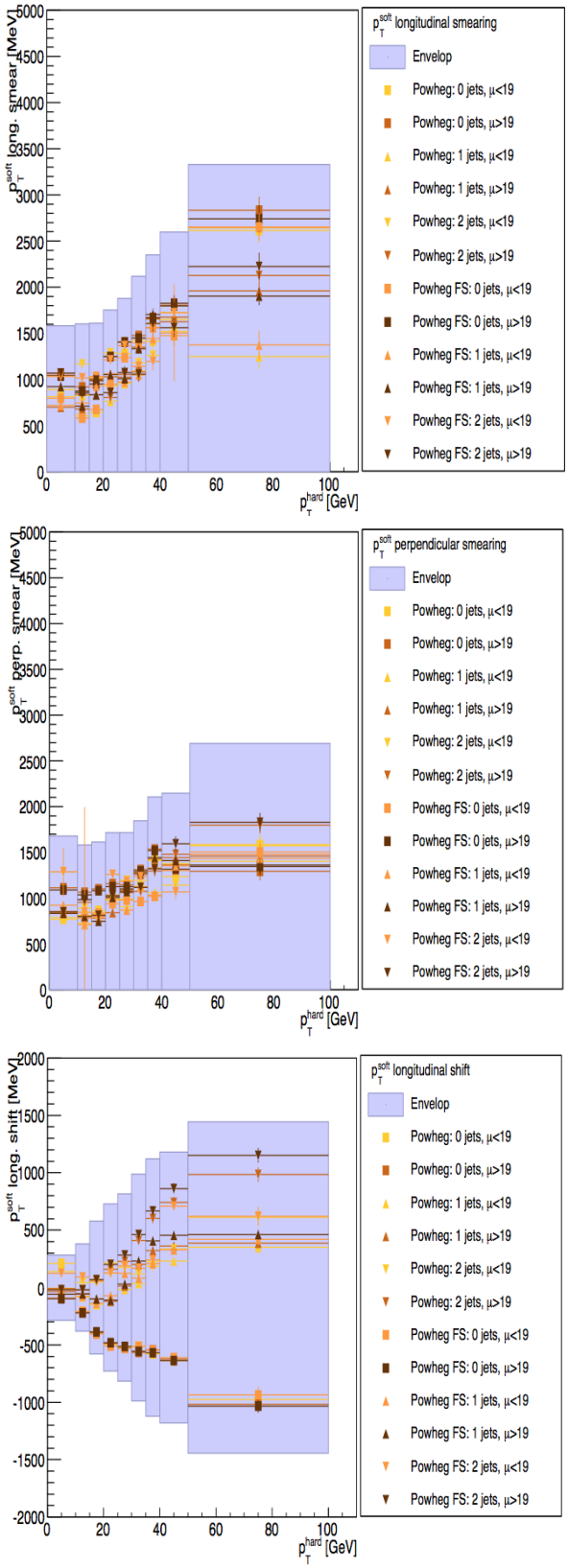
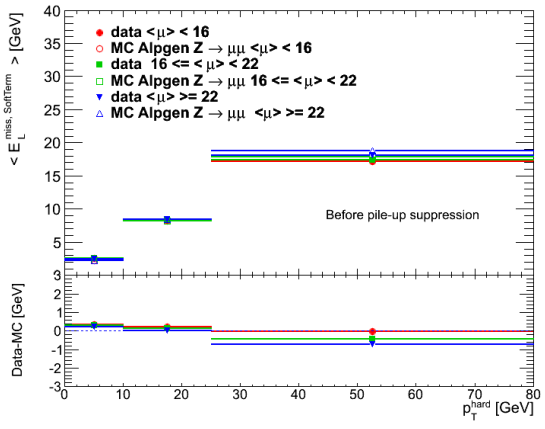
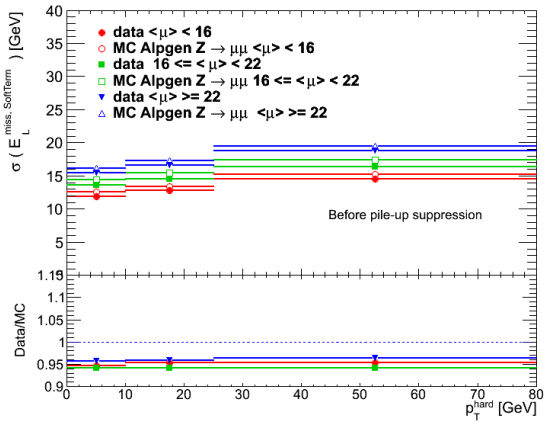


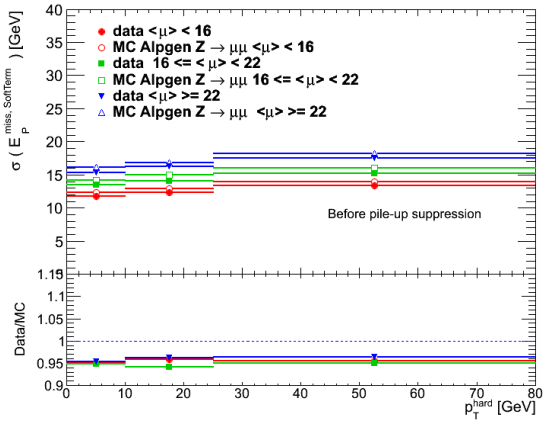
Figure 5-23: Effects of $p_{\mathrm{T}}^{\mathrm{hard}}$ component systematics on the measured $p_{\mathrm{T,soft}}^{\mathrm{miss}}$ shift in scale (mean) and resolution (smearing) for Track MET-Cl-j in $Z \to \mu\mu$ samples at $\sqrt{s} = 8$ TeV [10].



(a)



(b)



(c)

Figure 5-24: Scale and resolution results for calorimeter based MET before pile-up suppression in $Z \to \mu\mu$ data and AlpgenHerwig samples.

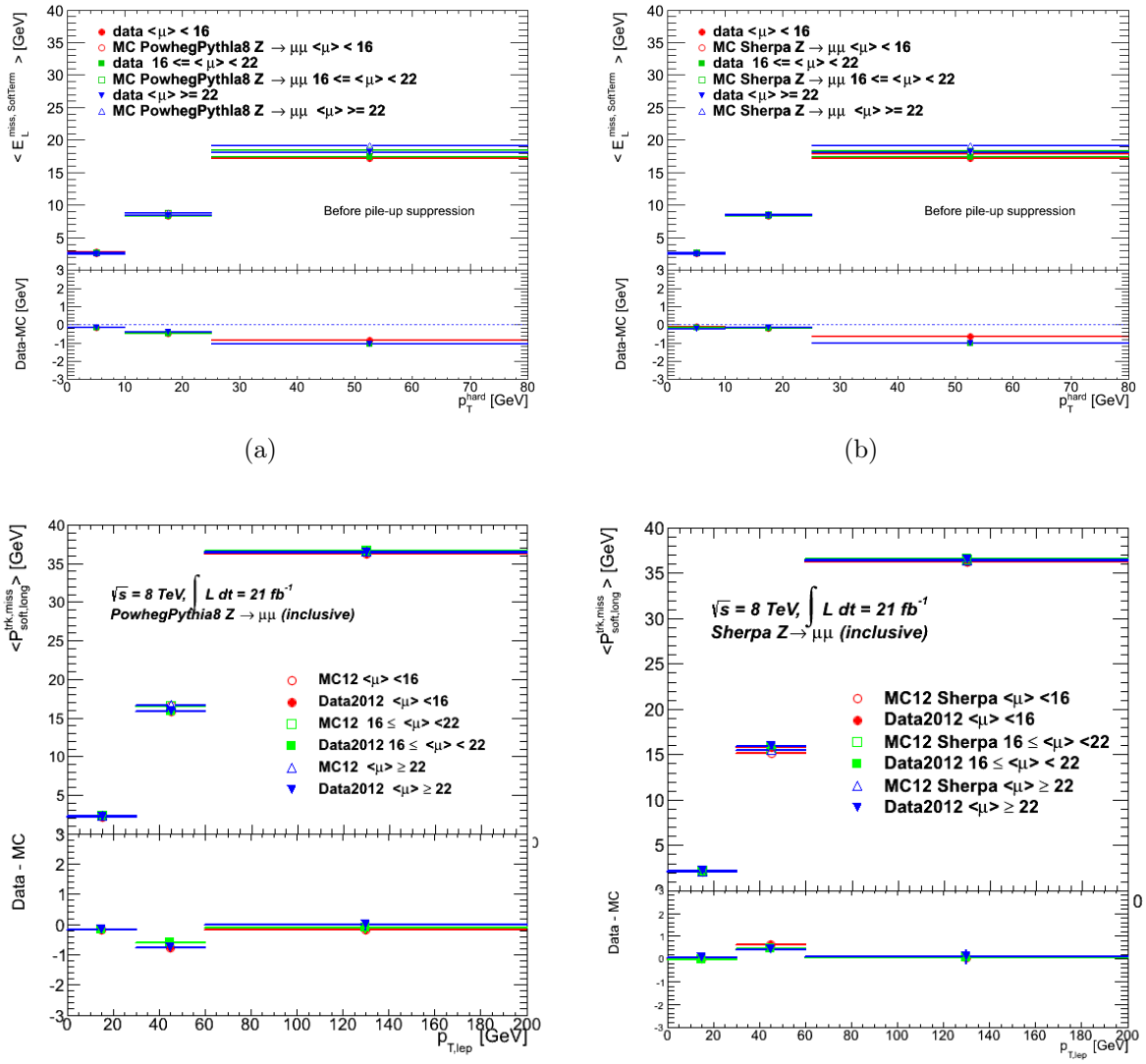


Figure 5-25: Scale results for calo MET before pileup suppression (top) and track MET (bottom) for $Z \to \mu\mu$ PowhegPythia8 (left) and Sherpa (right) samples.

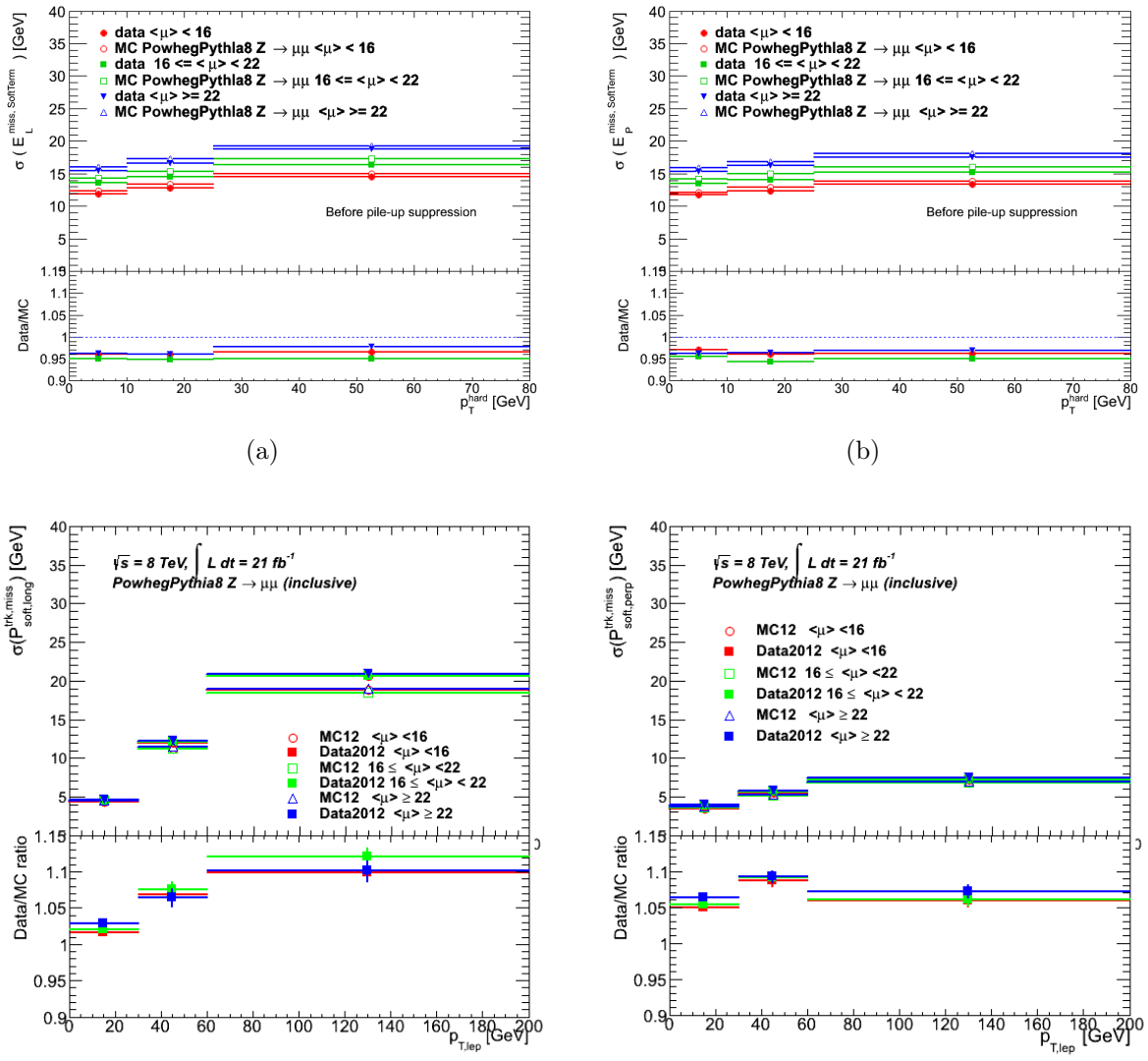


Figure 5-26: Resolution plots for calo MET (top) and Track MET-Cl (bottom) for $Z \to \mu\mu$ events with PowhegPythia8 MC simulated samples.

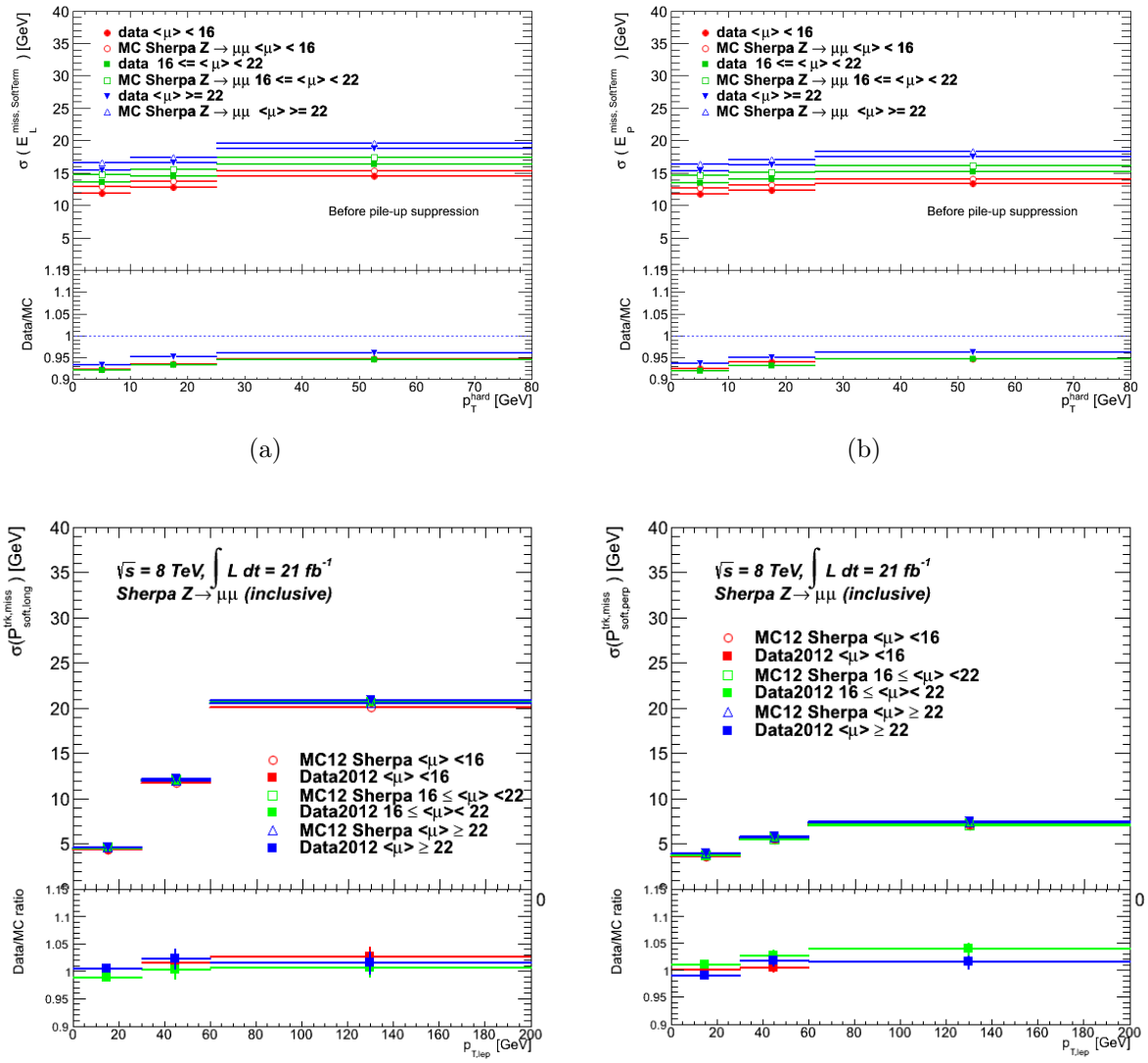


Figure 5-27: Resolution plots for calo MET (top) and Track MET-Cl (bottom) for $Z \to \mu\mu$ events with Sherpa MC simulated samples.

5.7 Application in $H \to WW^*$ Higgs Analysis

For Moriond 2013 results, Track MET-Cl was used in the $H \to \mathrm{WW}^* \to \ell\nu\ell\nu$ analysis after preselection in the 0- and 1-jet bins in order to reduce the Drell-Yan background in the same flavor channels i.e. in the ee and $\mu\mu$ channels with about 30% systematics. The results shown earlier for Track MET-Cl systematics were applied. An optimized version of that $H \to WW^*$ analysis (as described in Appendix A), using track MET to define an optimized phase space volume, increases the expected significance furthermore.
The final re-analysis results use Track MET-Cl and Track MET-Cl-j given that TrackMET has the best resolution among various MET definitions. TrackMET is used in 2 flavors:

1. Track MET-Cl and Track MET-Cl-j

2. Relative MET projection:

$$\begin{aligned} \mathrm{MET}^{rel} &= \mathrm{MET} \times \sin(\Delta\Phi) && \text{if } \Delta\Phi < \pi/2 \ , \\ \mathrm{MET}^{rel} &= \mathrm{MET} && \text{otherwise} \ , \\ \Delta\Phi &= min[\Delta\Phi(\mathrm{MET}, \mathrm{lepton}),\ \Delta\Phi(\mathrm{MET},\ \mathrm{jet})] \ . \end{aligned} \tag{5.8}$$

Cuts on relative Track MET-Cl , namely $E_{\mathrm{T,Rel}}^{\mathrm{miss,track}}$, are applied in same flavor (SF) channels (ee,$\mu\mu$) and on Track MET-Cl-j are applied in opposite flavor (OF) channels($e\mu,\ \mu e$) [10]:

- relative TrackMET > 40(35) GeV in the SF 0-jet bin (1-jet bin).
- Track MET-Cl-j >20 GeVin OF 0 jet bin, 10 GeVin OF 1-jet bin and 50 GeVfor VBF($n_{jets} \geq$2).

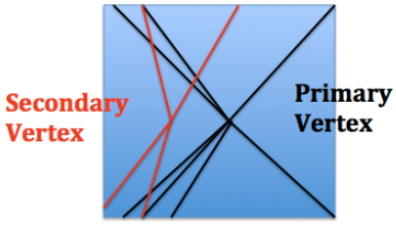
5.8 Application in $H \to \tau^+\tau^-$ Higgs Analysis

TrackMET is studied in $H \to \tau^+\tau^-$ channels where at least one tau decays leptonically (i.e. the lep-lep and lep-had channels). Figure 5-28 shows the MET definition used in Run-I for this analysis, namely the calorimeter MET-STVF. Compared to the calorimeter-based MET-STVF, Track MET-Cl-j :

– gives equivalent results in terms of resolution.

– has better performance against pile-up.

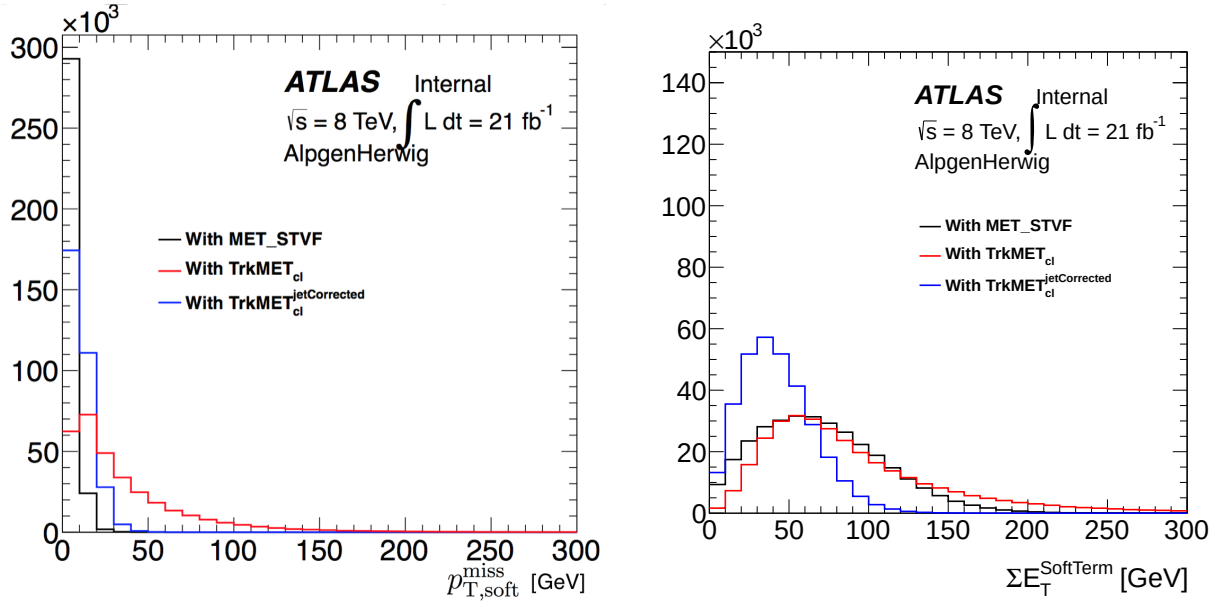
– has smaller $p_{\mathrm{T,soft}}^{\mathrm{miss}}$ and $\Sigma E_{\mathrm{T}}^{\mathrm{Soft\ Term}}$ tails, and slightly smaller tails in the overall MET value.



$$STVF = \frac{\sum_{\text{Soft term tracks, PV}} p_T}{\sum_{\text{All soft term tracks}} p_T}$$

Figure 5-28: MET-STVF definition (right) with an illustration of tracks coming from primary (PV) and secondary vertices in an event (left).

An example comparing MET-STVF and track MET in the $H \to \tau^+\tau^-$ dilepton channel is given in 5-29, where $|p^{\text{miss}}_{\text{T,soft}}|$ ('softTerm') and ΣE_T of soft term components are shown for AlpgenHerwig $Z \to \tau^+\tau^-$ samples. A complete list of similar figures for the lep-lep and lep-had $Z \to \tau^+\tau^-$ and Higgs samples in both ggF and VBF production modes is given in Appendix B, in addition to some m_{MMC} plots.



(a) (b)

Figure 5-29: $p^{\text{miss}}_{\text{T,soft}}$ (left) and ΣE_T of soft term components (right) distributions for various MET definitions for $Z \to \tau^+\tau^-$ events in the $H \to \tau^+\tau^-$ analysis lep-lep channel.

5.9 Application in $H \to b\bar{b}$ Higgs Analysis

Track MET is also used in the $H \to b\bar{b}$ analysis in addition to the calorimeter-based MET definition. In the 0-lepton channel, the cuts related to the missing transverse momentum are the following [13]:

- $E_{\mathrm{T}}^{\mathrm{miss}} > 120$ GeV
- $p_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV
- $\Delta\phi(E_T^{\mathrm{miss}}, p_T^{\mathrm{miss}}) < \pi/2$
- $\min[\Delta\phi(E_T^{\mathrm{miss}}, \mathrm{jet})] > 1.5$
- $\Delta\phi(E_T^{\mathrm{miss}}, b\bar{b}) > 2.8$

where $p_{\mathrm{T}}^{\mathrm{miss}}$ refers to track MET and $E_{\mathrm{T}}^{\mathrm{miss}}$ refers to the missing transverse momentum estimated based on calorimeter energy deposits.

5.10 Conclusion and Prospectives

In this chapter, a track-based, pile-up robust method to estimate the missing transverse energy, namely track MET, was presented. Besides being a complement to the existing Run-I calorimeter-based measurement of $E_{\mathrm{T}}^{\mathrm{miss}}$, it has little pile-up dependence, which makes it very important Run-II studies where the pile-up conditions will become even more challenging than the 2012 Run of LHC.

Soft term systematics were explored in detail and the corresponding results for various track MET definitions were presented for both 2011 and 2012 data sets. Finally, a brief overview of studies done in $H \to WW^*$ and $H \to \tau^+\tau^-$ channels, where the final state is characterized by a large MET value, was given. Improving the MET value and resolution will lead to better mass measurement and resolution in Higgs studies.

Bibliography

[1] F. Hariri, C. Lee, B. Liu, R. Mazini, and M. Testa, *Measurement of track-based missing transverse momentum in proton-proton collisions at $\sqrt{s}$ = 8 TeV centre-of-mass energy with the ATLAS detector*, Tech. Rep. ATL-COM-PHYS-2013-1577, CERN, Geneva, Nov. 2013.

[2] L. Bianchini et al., *Reconstruction of the Higgs mass in $H \to \tau^+\tau^-$ Events by Dynamical Likelihood techniques*, J. Phys. Conf. Ser. 513 (2014) 022035

[3] The ATLAS Collaboration, *Pile-up subtraction and suppression for jets in ATLAS*, ATLAS-CONF-2013-083

[4] Gavin P. Salam. *Towards Jetography.* Eur. Phys. J. **C67** (2010) 637-686 arXiv:0906.1833 [hep-ph].

[5] M. Cacciari, and G. P. Salam, *The Catchment Area of Jets*, JHEP **0804** (2008) 005 arXiv:0802.1188 [hep-ph] LPTHE-07-02.

[6] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas*, Phys.Lett. **B659** (2008) 119-126, arXiv: 0707.1378 [hep-ph]

[7] A. Schwartzman, and M. Swiatlowski, *New Track-jet Association and Heavy Flavor Labeling Tools*, ATLAS b-tagging workshop.

[8] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas*, Phys.Lett. **B659** (2008) 119-126, arXiv: 0707.1378 [hep-ph]

[9] ATLAS Collaboration *Tagging and suppression of pileup jets with the ATLAS detector*, ATLAS-CONF-2014-018

[10] *Object selections in the $H \to WW \to \ell\nu\ell\nu$ analysis with 20.3 fb^{-1} of data collected with the ATLAS detector at $\sqrt{s}$ = 8 TeV* Tech. Rep. ATL-COM-PHYS-2013-1504, CERN, Geneva, July 2014

[11] ATLAS Collaboration, *Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector*, *JHEP* **09** (2014) 145, [arXiv:1406.3660].

[12] ATLAS Collaboration, *Measurement of the electroweak production of dijets in association with a Z-boson and distributions sensitive to vector boson fusion in*

proton-proton collisions at $\sqrt{s} = 8$ *TeV using the ATLAS detector*, *JHEP* **04** (2014) 031, [arXiv:1401.7610].

[13] ATLAS Collaboration, *Search for the* $b\bar{b}$ *decay of the Standard Model Higgs boson in associated (W/Z)H production with the ATLAS detector*, ATLAS-CONF-2013-079

Chapter 6

Search for the Higgs Boson in the $\tau_{\text{lep}}\tau_{\text{had}}$ Final State

One of the main goals of the LHC Run-I was to check experimentally the existence of the lightest scalar boson predicted by the Standard Model (SM), i.e. the Higgs boson. And one of the interesting yet challenging channels to look for the Higgs boson is the $H \to \tau^+\tau^-$ channel, where the Higgs boson decays leptonically into a pair of taus with a branching ratio BR$\approx$ 6.3% [1] for m_{H}=125 GeV. Based on the tau decay mode, i.e. hadronic or leptonic, 3 orthogonal channels are defined: lep-lep, lep-had and had-had. This chapter discusses the $H \to \tau^+\tau^-$ analysis mainly in the lep-had channel. The results are obtained using the full ATLAS data sets during Run-I, i.e. 2012 and 2011 periods corresponding to integrated luminosity L= 20.3 fb^{-1} at center of mass energy $\sqrt{s}$ = 8 TeV and L=4.5 fb^{-1} at $\sqrt{s}$ = 7 TeV respectively. Starting with the motivations behind this analysis first in sec. 6.1, the experimental signatures of the $H \to \tau^+\tau^-$ events are presented next in section 6.2. In addition, the description of powerful discriminating variables used in the analysis is given in section 6.3, while the various mass calculation techniques are described separately in section 6.4. The data and Monte Carlo (MC) simulation samples are then described in section 6.5. This is followed by the analysis strategy: the objects definitions and event pre-selection are presented in sections 6.6 and 6.7. Then, event categorization and the analysis details for signal extraction are described in sec. 6.8. In order to have a reliable result, a good background modeling is needed. The background estimation techniques are thus summarized in sec. 6.9 and applied as explained in sec. 6.10. Finally, the final analysis results using boosted decision trees background suppression (explained in sec. 6.11) and the associated systematic uncertainties are shown in sec. 6.12, followed by a study using track MET[1] to improve the analysis results, in preparation for Run-II, is presented in section 6.13. My personal work in the $H \to \tau^+\tau^-$ analysis is presented in sec. 6.9.2 ('fake tau' estimation using the fake factor method) and sec. 6.13.

[1]Track based definition of missing transverse energy. Details can be found in chapter 5.

6.1 Motivation

Even with discovery of the Higgs in the bosonic modes $\gamma\gamma^*, ZZ^*$ in June 2012, the need for the search in the $H \to \tau^+\tau^-$ channel is driven by the following:

- The Higgs boson coupling to fermions is essential to check the compatibility with the SM prediction, and eventually excluding a fermiophobic Higgs boson hypothesis.
- The $H \to \tau^+\tau^-$ channel allows direct measurements of the Yukawa couplings of Higgs to leptons, which makes it thus a very interesting channel for spin studies and electroweak measurements.
- In the vector boson fusion mode for Higgs production[2], the $H \to \tau^+\tau^-$ decay mode is one of the dominant modes for low mass Higgs ($80\ \text{GeV} < m_H < 180\ \text{GeV}$), making it the most sensitive VBF channel (see chapter 1). As a consequence, this allows for Higgs couplings measurements in different production modes and is interesting for precise electroweak measurements.

6.2 Event Experimental Signature and Processes Involved

The experimental signature of $H \to \tau_\ell\tau_{\text{had}}$ events is determined by both the tau decay mode and the Higgs production process.

- Based on the tau decay mode, in the lep-had channels, events are expected to have in the final state a lepton (e or μ), a narrow jet from the hadronically decaying tau and at least 3 neutrinos (2 ν from τ_ℓ and one from τ_{had}) leading to large missing transverse energy MET. Thus, the analysis requires, at the reconstruction level, a single lepton, a tau and large MET.
- Based on the Higgs production mechanism, the number of reconstructed jets and their kinematic properties vary. VBF processes require at least 2 forward jets well separated in pseudorapidity η, hence no central jets.

6.2.1 Signal Processes

Of the four Higgs production mechanisms at LHC, ggF and VBF are the dominant modes and contribute the most to detected signals[3]. Events with Higgs bosons produced in association with a vector boson (VH) have a significantly smaller cross section and little contribution to the $H \to \tau_\ell\tau_{\text{had}}$ analysis after applying selection criteria. At LHC Run-I, the top-pair associated production mechanism has no significant contribution and such events will not be considered in the lep-had analysis.

[2]Higgs is mainly produced through gluon-gluon fusion (ggF) and vector boson fusion (VBF) processes at LHC during Run-I.

[3]see chapter 1.

- **ggF:**
  - The largest cross section and dominant contribution
  - At the tree level, there are no additional partons in the final state, resulting thus in no additional jets. However, higher order QCD corrections for the gluon radiation and top loop are important and lead to additional partons in the final state. Thus, the fraction of multi-jet ggF lep-had events is non-negligible.
- **VBF:**
  - The second largest cross section
  - At leading order, VBF lep-had processes have 2 additional jets. The production mechanism is based on electroweak interactions in this case and the QCD corrections are smaller than in the ggF case.
  - VBF events have at least 2 forward jets, well separated in η, and a high di-jet mass with no significant hadronic activity in the central region. In such electroweak processes, there is color coherence between the 2 forward jets. The latter can be used for signal/background discrimination i.e. by applying a central jet veto.

6.2.2 Background Processes

1-Irreducible background process:

It consists of Z boson events '$Z \to \tau\tau$(+jets)' decaying into a pair of tau leptons in the final state. Signal-background discrimination is done through cuts on di-tau mass and on kinematic properties of jets (τ_{had} excluded) produced in the final state.

2-Reducible background processes:

- W+**jets :** The W+jets events can produce an experimental signature similar to the signal one, where the W boson decays leptonically and a jet fakes a hadronic tau decay. The emitted neutrino will lead to real missing transverse energy in the event and such events are mainly reduced with appropriate cuts on the transverse mass m_T.
- **QCD multi-jet events :** The signal experimental signature can be also produced by QCD multi-jet events, with one jet faking τ_{had} and another a lepton (e,μ). Such events are mainly reduced by requiring well identified taus and tight isolated leptons.
- $Z/\gamma^* \to \ell\ell$+**jets :** A signal signature can also be obtained for the $Z/\gamma^* \to \ell\ell$+jets production when :

- one lepton is mis-identified as a τ_{had} , which is not very likely since electron and muon reconstruction efficiencies are very high. This is related to the electron isolation : it can happen for non-isolated electrons or in case of failure of the electron-jet overlap removal.
- one lepton does not pass the selection requirements or is out of the detector acceptance, and a jet fakes τ_{had} . Such events have the major contribution to $Z/\gamma^* \to \ell\ell$+jets processes, and can be reduced by requiring events to have exactly one lepton in the final state while increasing the acceptance of low-p_{T} leptons.

- **top ($t\bar{t}$ and single top) :** Another source of background is the top quark production. The jets produced in the event may fake τ_{had} while leptons originate from W decays. Such events are significantly reduced when vetoing b-jets using adequate b-tagging (see section 4.6.3) algorithms.

- **diboson (W^+W^-, ZZ and $W^\pm Z$) :** Various scenarios are possible in this case but they will not be discussed since this background type has an extremely negligible contribution to the event yield in the lep-had analysis.

6.3 Important Discriminating Variable Definitions

The major powerful signal-background discriminating variables in the $H \to \tau_\ell \tau_{\text{had}}$ analysis are defined in the following:

1. MET: missing transverse energy. When added to the visible decay products information, MET allows to reconstruct the di-tau mass, which is necessary to distinguish the irreducible background (e.g. Z mass peak) from the signal processes (peaked at the Higgs mass). MET is used for resonant/non-resonant di-tau processes discrimination(e.g. QCD from signal events). Thus, a good estimation of MET in the event with high resolution is crucial. The MET definition used in this analysis is MET STVF (defined in detail in chapter 4).

2. m_{vis} (or $m_{\tau\tau}^{\text{vis}}$): mass reconstructed using the visible tau decay products and used in the definition of the VBF signal region. Details are given in section 6.4.

3. $m_{\tau\tau}$: di-tau invariant mass. Since there are at least 3 neutrinos in the final state, the complete information about the final state is not available, and assumptions about the neutrino momenta are done. $m_{\tau\tau}$ is used for signal/background discrimination in resonant di-tau processes ($Z \to \tau_{\text{lep}}\tau_{\text{had}}$ vs Higgs).
   The estimation of $m_{\tau\tau}$ can be done using the collinear mass approximation or the MMC approach. In both cases, MET is added to the information from the visible decay products.

4. m_{MMC}(or $m_{\tau\tau}^{\text{MMC}}$): missing mass calculator [3], which is a reconstruction technique for resonances decaying to $\tau\tau$. It uses a likelihood method taking into

account available kinematic variables and their resolution, in addition to the τ-mass constraint. Details are given in section 6.4.

5. m_T: transverse mass used to define signal phase space regions and reduce background contamination. A cut on m_T is very effective in W+jets reduction for example, especially that the angular term in m_T calculation, as shown in eq.(6.1), is different in W+jets events when compared to di-tau resonances (Higgs and Z boson decays) with additional jet(s) in the final state.

$$m_T = \sqrt{2p_T^{\ell} E_T^{\text{miss}} \left(1 - \cos\left(\Delta\phi_{\ell, E_T^{\text{miss}}}\right)\right)} \,. \tag{6.1}$$

In fact, in this case, the Higgs boson recoils against the jet(s) and is boosted. Stronger boost leads to smaller angular separation (opening angle) between the decay products as shown in Fig 6-1, since the taus resulting from the boosted Higgs decay tend to be collinear. The MET direction falls thus between the τ_{had} candidate and the lepton. On the other hand, in the case of W+jets events, the angular separation between the jet faking τ_{had} and the lepton is larger and consequently the m_T values are large.

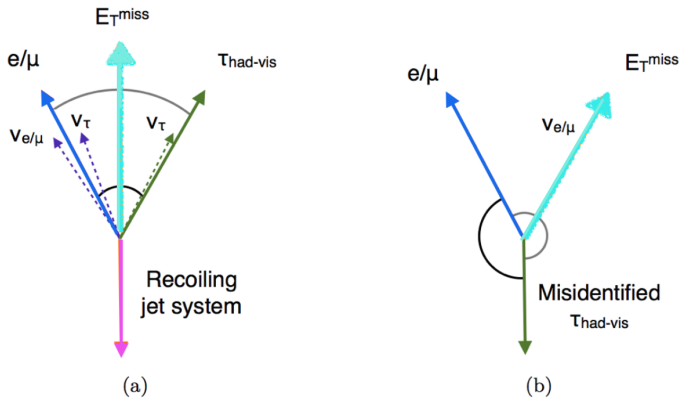


Figure 6-1: Figure showing a schematic representation of the angular separation between the final state products and the MET vector, in signal and $Z \to \tau\tau$ lep-had events with additional jet(s) (left) and in W+jets lep-had samples (right).

6. m_{jj}: di-jet invariant mass (needed for VBF).

7. $\Delta\eta_{\text{jj}}$: di-jet pseudorapidity difference (needed for VBF).

8. p_T^H: is the Higgs transverse momentum needed to defined the "boosted" category (defined in section 6.8). p_T^H is reconstructed by doing a vector sum of the event MET and the visible tau decay products transverse momentum [2].

6.4 Di-tau Mass Reconstruction in $H \to \tau_\ell \tau_{\mathbf{had}}$

Since the final state of lep-had events has neutrinos resulting in real MET values, the invariant mass of the tau pair can not be reconstructed and approximations are made to estimate the di-tau mass. Di-tau mass resolution thus depends strongly on how well MET is reconstructed. The various mass reconstruction techniques used in $H \to \tau_\ell \tau_{\text{had}}$ analysis are presented hereafter.

6.4.1 Visible Di-tau Mass $m_{\tau\tau}^{\text{vis}}$

m_{vis} (or $m_{\tau\tau}^{\text{vis}}$) is the mass estimate obtained using tau visible decay products only. Neutrinos are thus neglected [2].

$$m_{\text{vis}} = \sqrt{(E_\ell + E_{\tau_{\text{had-vis}}})^2 - (\overrightarrow{p}_\ell + \overrightarrow{p}_{\tau_{\text{had-vis}}})^2} \ . \tag{6.2}$$

The Z boson (in $Z \to \tau\tau$ events) and Higgs resulting mass distributions peaks are not well separated as shown in Fig. 6-2. In addition, since the neutrinos are omitted from the calculation, the entire mass distribution is shifted to lower values as seen clearly for the Z mass peak. This variable is mainly used at the final stages of VBF signal region (signal phase space) definition and for $Z \to \tau\tau$ background model validation.

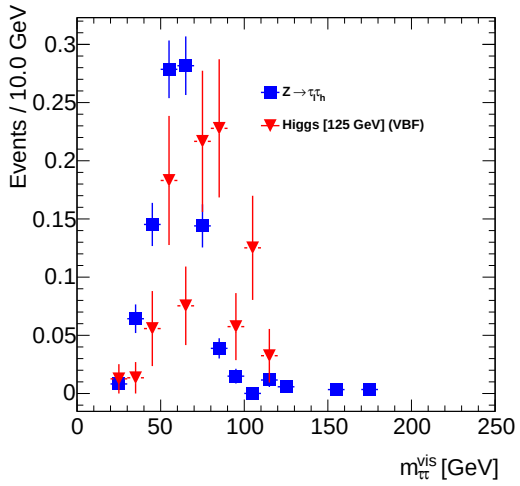


Figure 6-2: Figure showing normalized distributions of the signal and $Z \to \tau\tau$ in the VBF signal region.

6.4.2 Collinear Mass Approximation

This mass variable is not currently used in $H \to \tau_\ell \tau_{\text{had}}$ analysis; however, it was used at early stages of the analysis. It is presented here for completeness and for better understanding of improvements when switching to m_{MMC}.
The Collinear Approximation (CA) assumes the following :

– The tau decay products are collinear with the τ direction in the lab frame. This is justifiable since in Z/H events (Z boson and Higgs), the di-tau resonance mass is significantly larger than $2\times m_\tau$ and the produced taus are very boosted [3].

– MET in the event originates from neutrinos coming from tau decays only.

– The lepton (e/μ) masses are negligible along with the neutrino mass ($m_\nu < 2$ eV).

With the assumptions above, the collinear mass is expressed in terms of the visible mass as [4]:

$$m_{\mathrm{coll}} = (m_{\mathrm{vis}})/(\sqrt{x_\ell x_{\mathrm{had}}}) \ , \tag{6.3}$$

where x_ℓ and x_{had} are the fractions of the tau transverse momentum carried by the visible decay products in leptonic and hadronic decays respectively.

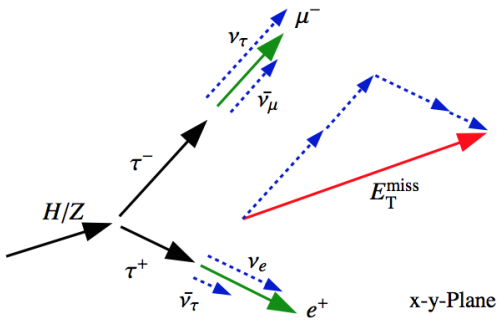


Figure 6-3: Figure showing an example of $Z/H \to \tau_\ell\tau_\ell$ decays with collinear mass approximation. The emitted tau decay products are collinear with the tau direction. The MET vector, assuming neutrinos are the only source of MET in the event, is illustrated as well.

Considering the $H \to \tau_\ell\tau_\ell$ as shown in Fig. 6-3 for illustrative purposes, the tau transverse momenta can be written as (neglecting the opening angle between the tau daughters and their masses):

$$p_{\tau^\pm} = (\alpha_\pm + 1)p_{\ell_\pm} \ , \tag{6.4}$$

where $\alpha_\pm p_{\ell_\pm}$ is the neutrinos momenta expressed in terms of the associated lepton transverse momentum. With CA assuming that all MET comes from neutrinos of tau decays:

$$\overrightarrow{\not{E}}_T = \alpha_+ \overrightarrow{p}_{T,\ell^+} + \alpha_- \overrightarrow{p}_{T,\ell^-} \ , \tag{6.5}$$

$$\alpha_\pm = \pm \frac{(\overrightarrow{\not{E}}_T \times \overrightarrow{p}_{T,\ell^\mp}).\hat{z}}{(\overrightarrow{p}_{T,\ell^+} \times \overrightarrow{p}_{T,\ell^-}).\hat{z}} \ . \tag{6.6}$$

Limitations:

As a result, as suggested by eq.(6.6), CA fails when the angular separation between the two taus is close to zero or π. The former is rarely realized experimentally. CA can not thus be applied in ggF with no additional jets in the final state, where the taus are rather back-to-back. In addition, the mass measurement are very sensitive to the measured MET value. This method requires boosted $\tau\tau$ system and large MET, since low MET values mean that the neutrinos are back-to-back or close enough to such configuration. CA fails also if the MET vector direction does not fall within the opening angle of the two tau leptons. If there is no physical solution (i.e. negative neutrino energies), then the event is rejected and not considered any further. For these reasons, a more powerful method with higher efficiencies was developed, namely the missing mass calculator (MMC) discussed in sec. 6.4.3.

6.4.3 Missing Mass Calculator m_{MMC}

The di-tau invariant mass m_{MMC}, calculated taking into account all the event kinematics, is a powerful discriminating variable used to separate signal from background processes in the lep-had analysis. The lep-had analysis uses m_{MMC} for $m_{\tau\tau}$ in final results. The mass resolution is good enough that the distribution peaks for $Z \to \tau_\ell \tau_{\mathrm{had}}$ and Higgs samples are well separated. An important fraction of irreducible background events can thus be eliminated.
The MMC technique is a likelihood based method, allowing improved mass reconstruction in $\tau\tau$ final states, without the limitations of the collinear mass approximation described in the previous section. The method is valid for various event topologies, delivering di-tau invariant mass estimates with good resolution [3].

6.4.3.1 The method

Assuming perfect detector resolution[4] and that all neutrinos in the event originate from tau decays, the event kinematics can be described by a system of 6 to 8 unknowns, depending on tau decay mode [3]. For lep-had events, 7 unknowns are needed to describe the neutrino systems, namely the $\nu_{\tau_{\mathrm{had}}}$ and $\nu_{\tau_\ell}\nu_\ell$ systems: two 3-vector components for each neutrino(s) system in addition to the angular separation between ν_ℓ and ν_{τ_ℓ} to describe the invariant mass of the neutrinos from the leptonic decay. The obtained system is underconstrained since there are only 4 equations (in x, y components of MET and invariant mass of each tau) to describe the 7 unknowns. To solve this, various grid points in $(\Delta\phi_1, \Delta\phi_2)$ are considered, where $\Delta\phi_i$ is the azimuthal angular separation between the visible and invisible decay products of $\tau_{i\,(i=1,2)}$. Not all solutions are equally probable and the likelihood of the best di-tau mass estimate, amongst all scanned grid points, is determined using variables describing the tau kinematics. Candidates for such variables are :

– the distance separating the neutrino(s) and the visible tau decay products

[4]The detector resolution effects are presented in the next section.

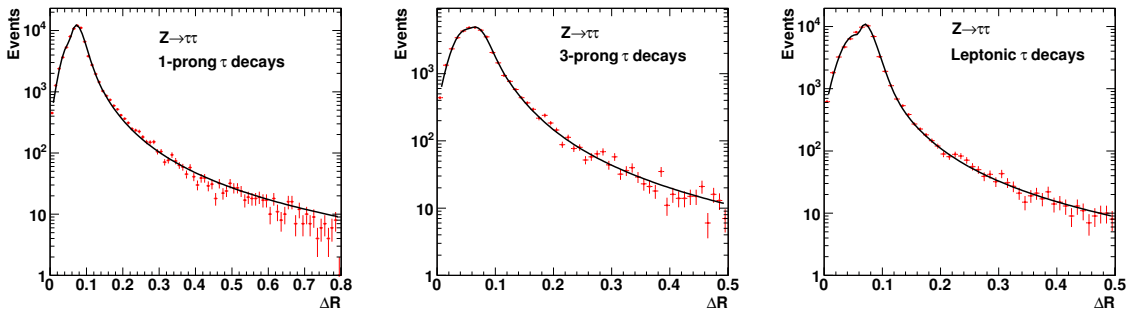
$$\Delta R = \sqrt{(\eta_{\text{vis}} - \eta_{\text{mis}})^2 + (\phi_{\text{vis}} - \phi_{\text{mis}})^2} \ ,$$

where η is the pseudorapidity and ϕ is the azimuthal angle in the transverse plane (x-y plane). Neglecting the tau polarization, the ΔR distribution depends only on the decay type and the tau momentum but not on the tau lepton origin. The distribution of ΔR in $Z \to \tau\tau$ for leptonic and hadronic tau decays is shown in Fig. 6-4.

– the 3-dimensional angular separation $\Delta\theta_{3\text{D}}$ between the invisible neutrino(s) momentum vector and the visible tau decay products momentum [5]. The corresponding distributions, similar to ΔR ones, are given in Fig. 6-5.

These kinematic distributions are parametrized for each p_{T}^{τ} range. They define probability density functions (PDFs) that enter the likelihood algorithm, by which the mass candidate $m_{\tau\tau}$ will be weighted. It was found that these angular distributions from $Z \to \tau\tau$ events are similar to those obtained with Higgs signal events, within the associated uncertainties. It should be noted that in these calculations, the tau lepton polarization is neglected and ΔR dependencies on the tau momentum and decay mode are taken into account only.

In the lep-had channel, the 3-dimensional angular separation $\Delta\theta_{3\text{D}}$ is used to obtain the PDFs entering the m_{MMC} likelihood algorithm.

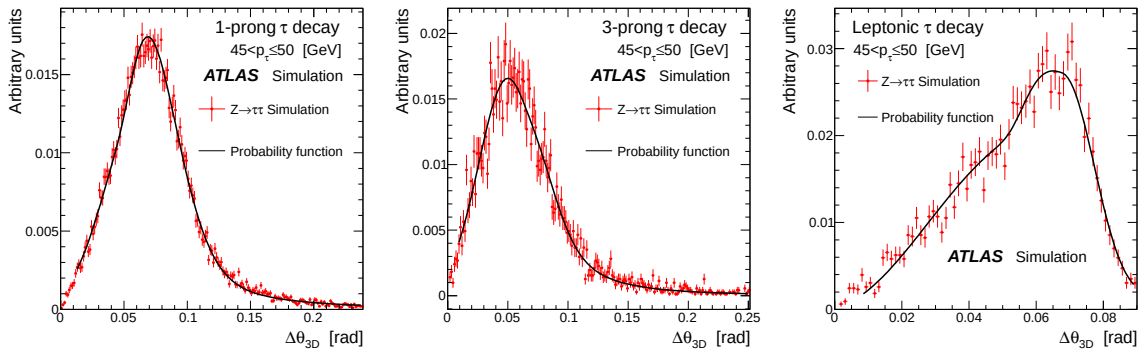


(a) (b) (c)

Figure 6-4: Distribution of the ΔR separation between the neutrino(s) and the visible tau decay products in simulated $Z \to \tau\tau$ for hadronic 1-prong(a) and 3-prong(b), in addition to leptonic(c) tau decays for a chosen tau p_T [3]

6.4.3.2 Performance and efficiency

Smearing the resolution on MET $\overrightarrow{\not{E}}_T$ and the visible tau decay momenta to study the detector effects results in:



(a) (b) (c)

Figure 6-5: Plots showing the probability distribution function of the 3-dimensional angular separation $\Delta\theta_{3D}$ between the neutrino(s) and the visible tau decay products in simulated $Z \to \tau\tau$ for hadronic 1-prong(a) and 3-prong(b), in addition to leptonic(c) tau decays [5]. The results are shown for taus with a generated momentum $45 < \mathrm{p} \leq 50$ GeV.

– a slight degradation of the mass reconstruction efficiency $\sim$ 3 - 7% due to mismeasured visible tau decay products momenta. However, the mass peak and resolution are not affected.

– a significant degradation of the mass reconstruction efficiency $\sim$ 30 - 40% for 5 GeV $\overrightarrow{E}_T$ resolution ($\sigma_x = \sigma_y = \sigma = 5$ GeV), if it wasn't taken into account. In addition, the resolution on the reconstructed mass is degraded by $\sim$ 8% - 18%, resulting in longer tails. This is mainly due to assuming that neutrinos originating from tau decays are the only source of MET in the event.

As a result, to have credible experimental measurements, m_{MMC} take into account the resolution on MET [3]. The m_{MMC} reconstruction efficiency in $Z \to \tau_\ell\tau_{\mathrm{had}}$ events and various Higgs mass hypotheses lep-had samples is summarized in Table 6.1. The MMC efficiency is defined as the fraction of the number of input events for which the MMC algorithm finds a solution. The efficiency is more than 96% for Higgs mass hypotheses less than 250 GeV and for Z events.

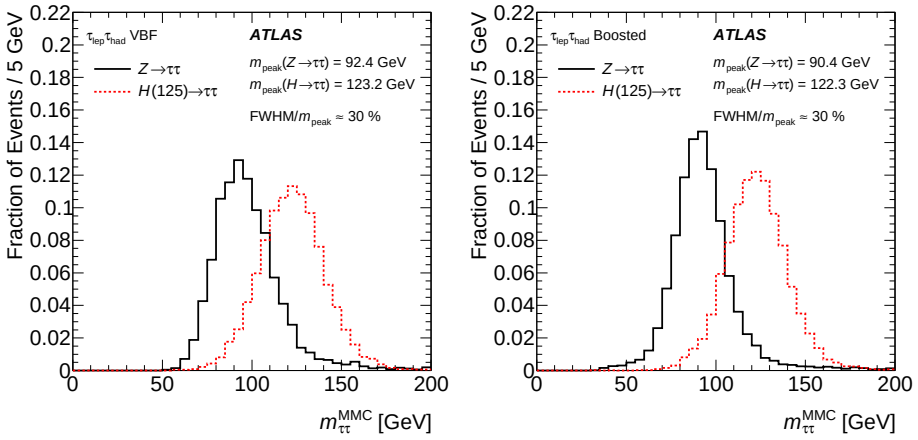
	Z	H(120)	H(150)	H(200)	H(250)	H(300)	H(400)
Efficiency(%)	98	99	98	98	96	91	88

Table 6.1: Table showing the m_{MMC} efficiency for $Z \to \tau\tau$ and Higgs signal events for various mass hypotheses, in the $\ell\tau_{\mathrm{had}}$ channel [5].

The efficiency is not exactly 100% for Z/H events, and is less than 96% for non-Z background events (e.g. 87% for W events and 93% for QCD processes) for the following reasons :

- having a false hypothesis where the topology of $Z \to \tau^+\tau^-$ events is assumed for other background events. The topology of the latter may not be consistent with the $Z \to \tau\tau$ decays as shown in Fig. 6-1 for example.
- having a mis-reconstructed MET, where the wrong MET value causes the m_{MMC} algorithm to fail (no convergence).
- assuming a perfect Gaussian MET resolution, which is not always the case.

Figure 6-6 shows the m_{MMC} distribution in $Z \to \tau\tau$ and Higgs signal samples in the lep-had channel. The Z/H mass peaks are well separated with a good mass resolution. Defining the resolution as the ratio of the full width at half maximum (FWHF) to the peak value gives $\sigma(m_{\mathrm{MMC}}) \approx 30\%$.



(a) (b)

Figure 6-6: Normalized m_{MMC} distribution in $Z \to \tau^+\tau^-$ and $H \to \tau^+\tau^-$ lep-had events for VBF and boosted analysis categories (defined in sec. 6.8) [2] .

6.5 Data and Monte Carlo Samples

The analysis presented here shows $H \to \tau_\ell\tau_{\mathrm{had}}$ results for the full Run-I data sets in both 2011 and 2012. The center of mass (CM) energy and corresponding luminosity for each year are summarized in Table 6.2. Events are required to pass the data quality requirements depending on the data taking period. During 2012, the data quality efficiency was ∼96%. Simulated MC signal and background samples information is summarized in Table 6.3, showing the related MC generator used at the matrix element level for the hard-scattering process, the perturbative QCD correction level applied, the parton shower and hadronization models in addition to the cross sections values used to normalize the simulated samples [2]. In all samples, the ATLAS detector simulation was done using GEANT4 [41] and the same reconstruction algorithms applied to data are used. Minimum-bias events are generated using Pythia8 with the

Year	CM energy $\sqrt{s}$ in TeV	Integrated luminosity in fb^{-1}	Peak Instantaneous luminosity in cm$^{-2}s^{-1}$
2011	7	4.5	3.7×10^{33}
2012	8	20.3	7.7×10^{33}

Table 6.2: Table summarizing the luminosity information for the 7 and 8 TeV data samples [6].

AU2 [42] tune and their pile-up contribution is taken into account in the simulated samples. MC generators, parton showers and PDFs are described in chapter 3.

6.5.1 Signal Events

Higgs events produced by ggF, VBF and VH are considered while those coming from $t\bar{t}H$ are neglected as in section 6.2.1.

1. **ggF:** Events are generated using POWHEG [7–10] interfaced with PYTHIA6 in 2011 and PYTHIA8 [11] in 2012 for parton shower and hadronization, where PYTHIA8 has the additional advantage of including the bottom and top quark masses in the quark loop compared to Pythia6. The parton distribution functions (PDFs) are taken from the CT10 [12] set. The cross sections are calculated with NNLO (next-to-next-to-leading order) QCD corrections [13–18], with soft gluon resummation up to NNLL (next-to-next-to-leading logarithm) [19] in addition to NLO electroweak (EWK) corrections [20, 21]. The resummation is in fact needed since the convergence of fixed order calculations at small p_T values is lost due to large logs. The NNLO QCD calculation with NNLL corrections given by HRES2.1 [29] is also used to correct the shape of the Higgs p_T distribution to include effects of finite top and bottom quark masses [29, 30], which have a large impact on the distribution, with renormalization μ_R and factorization μ_F scales: $\mu_R = \mu_F = \sqrt{m_H^2 + p_T^2}$ (dynamic scale) [2]. The reweighting for events with exactly one jet and those with at least 2 jets is done separately.

2. **VBF:** POWHEG with CT10 PDFs is used to generate VBF events and is interfaced with Pythia8 for parton shower and hadronization. The overall normalization includes approximate NNLO QCD corrections and full NLO QCD and EWK ones [22, 23, 26, 27]. The EWK corrections derived from the difference between POWHEG+PYTHIA and HAWK [22, 23] calculations vary with the Higgs boson p_T and are of the order of a few percent at low mass, going up to 20% for p_T=300 GeV [25].

3. **VH:** Events are generated using PYTHIA8 and CTEQ6L1 PDFs. The cross section normalization includes QCD NNLO [28] and NLO EW corrections [24].

All signal samples are generated for different Higgs mass hypothesis covering the range [100-150] GeV, in steps of 5 GeV.

Signal ($m_H = 125$ GeV)	MC generator	$\sigma \times \mathcal{B}$ [pb] $\sqrt{s} = 8$ TeV		
ggF, $H \to \tau\tau$	POWHEG [7–10] + PYTHIA8 [11]	1.22	NNLO+NNLL	[13–18, 25]
VBF, $H \to \tau\tau$	POWHEG + PYTHIA8	0.100	(N)NLO	[22, 23, 25, 26]
WH, $H \to \tau\tau$	PYTHIA8	0.0445	NNLO	[25, 28]
ZH, $H \to \tau\tau$	PYTHIA8	0.0262	NNLO	[25, 28]
Background	**MC generator**	$\sigma \times \mathcal{B}$ [pb] $\sqrt{s} = 8$ TeV		
$W(\to \ell\nu)$, $(\ell = e, \mu, \tau)$	ALPGEN [37]+PYTHIA8	36800	NNLO	[43, 44]
$Z/\gamma^*(\to \ell\ell)$, 60 GeV$< m_{\ell\ell} <$ 2 TeV	ALPGEN+PYTHIA8	3910	NNLO	[43, 44]
$Z/\gamma^*(\to \ell\ell)$, 10 GeV$< m_{\ell\ell} <$ 60 GeV	ALPGEN+HERWIG [45]	13000	NNLO	[43, 44]
VBF $Z/\gamma^*(\to \ell\ell)$	SHERPA [46]	1.1	LO	[46]
$t\bar{t}$	POWHEG + PYTHIA8	253$^{\circledast}$	NNLO+NNLL	[47–52]
Single top : Wt	POWHEG + PYTHIA8	22$^{\circledast}$	NNLO	[53]
Single top : s-channel	POWHEG + PYTHIA8	5.6$^{\circledast}$	NNLO	[54]
Single top : t-channel	AcerMC [40]+PYTHIA6 [32]	87.8$^{\circledast}$	NNLO	[55]
$q\bar{q} \to WW$	ALPGEN+HERWIG	54$^{\circledast}$	NLO	[56]
$gg \to WW$	GG2WW [39]+HERWIG	1.4$^{\circledast}$	NLO	[39]
WZ, ZZ	HERWIG	30$^{\circledast}$	NLO	[56]
$H \to WW$	same as for $H \to \tau\tau$ signal	4.7$^{\circledast}$		

Table 6.3: Table summarizing simulated MC signal and background samples information at $\sqrt{s} = 8$ TeV. The last column states the order of the applied QCD corrections, in addition to showing the product of cross section by the corresponding branching ratio ($\sigma \times B$) for each process, except for those marked with a $\circledast$. The latter are presented with the corresponding inclusive cross section value in the last column [2].

6.5.2 Background Events

An important contribution to the lep-had background processes comes from $Z/\gamma^* \to \tau_\ell\tau_{\text{had}}$ events, which are estimated using $Z/\gamma^* \to \mu\mu$ data samples in which muons are replaced by simulated tau decay products kinematics. The modeling of these samples, also called 'embedded' $Z \to \tau\tau$ events, is described in detail in section 6.10. The dependence of such events on simulation is limited to tau decay products and the corresponding detector response.

For the remaining processes, various MC generators and parton shower models are used. The samples generated through HERWIG [33, 34] have taus simulated using TAUOLA [35]. In all samples, PHOTOS [36] is used to model the lepton charge radiation. For W/Z+jets samples, the MLM [38] matching and merging scheme is used. Finally, CTEQ6L1 PDFs were used in most samples except for the WW production through GG2WW [39], where CT10 PDFs were used. The cross section normalizations of background processes are done using values shown in Table 6.3 or based on control region calculations as explained in section 6.10.

6.6 Object Definitions

The requirements on reconstructed objects to be considered by the lep-had analysis before applying any event selections are given below. The definition of the selection criteria and reconstruction terminology used below is given in chapter 4.

1. **Leptons:** Electrons are required to pass the 'medium' identification criteria, with p_T >15 GeV, and $|\eta| < 2.47$. Electrons falling within the transition region $1.37 < |\eta| < 1.52$ are discarded. On the other hand, muons must be of type 'combined' or 'ST', passing identification cuts, with $p_T > 10$ GeV and falling within the central region $|\eta| < 2.5$.

2. **Hadronic taus:** Taus should pass the 'medium' BDT identification selection, falling within $|\eta| < 2.47$, and have 1 or 3 prongs[5]. With charge $= \pm 1$, the p_T threshold is set at 20 GeV.

3. **Jets**: Jets are built with the anti-k_T algorithm with distance parameter R = 0.4, using locally calibrated (LC) topoclusters. They should pass the threshold p_T >30 GeV and fall within $|\eta| < 4.5$. JVF requirements are applied to suppress pile-up, where the cut |JVF| > 0.5 (0.75) is applied to central jets within $|\eta| < 2.4$ in 2012 (2011). For the 8 TeV dataset, this is only applied to jets if they have $p_T < 50$ GeV.
   Moreover, b-jets are identified using the MV1 b-tagging algorithm with a 70% efficiency.

[5]Term referring to the number of charged tracks associated to a hadronic tau.

4. **Missing Transverse Energy (MET):** The MET STVF definition is used to evaluate the missing transverse energy in the event. This is mainly a calorimeter based MET definition, with a track corrected soft term (see sec. 4.8).

Overlap removal (OLR) is applied to all reconstructed objects if they fall within $\Delta R < 0.2$, and selected objects are chosen in the following decreasing order of priority: muons, electrons, τ_{had} and jets. The order is dictated by the object reconstruction efficiency, being the largest in case of muons. For OLR calculations, leptons are required to pass looser identification selections than the ones mentioned earlier to ensure maximum reduction of leptons mis-identified as τ_{had}. Moreover, the muon p_T threshold in these calculations is lowered to 4 GeV.

In addition, before any event selection, such loose leptons are also used to identify di-lepton $Z \to \ell\ell$ background processes, defining thus a di-lepton veto to be used later in analysis.

6.7 Preselection

The '*preselection*' stage here refers to the basic event selection requirements done in analysis before proceeding to categorization (e.g. defining VBF candidates) and performing cut-based or multi-variate analyses (MVA).

6.7.1 Preliminary Step

Before starting any analysis, a preliminary check is done per event to make sure that it is the result of a proton-proton interaction, to reduce contamination from cosmic rays, etc. In order to enter the analysis, an event must have thus a primary vertex with at least 4 associated tracks with $p_T > 400$ MeV. The position of the vertex must be compatible with the beam spot as well. Events failing such requirements are discarded at this level.

6.7.2 Preselection Requirements

Once the events with good primary vertices and objects satisfying pre-analysis selections described in sections 6.6 and 6.7.1 are identified, the analysis is performed next, starting with the preselection stage.

1. **Single lepton events:** At this stage, only one lepton is required per event. Leptons (e,μ) should be well isolated, satisfying tighter identification criteria, with higher p_T thresholds e.g. electrons with $p_T > 25\,(26)$ GeV and muons with $p_T > 22\,(26)$ GeV in 2011 (2012) respectively, as summarized in Table 6.4. Electrons should pass the 'tight' identification criteria for better rejection of electrons faking τ_{had} . Both track-based I $(p_T, \Delta R)$ and calorimeter-based I $(E_T, \Delta R)$ isolation[6] requirements are applied such that I $(p_T, 0.4) < 0.06$ and I $(E_T, 0.2) < 0.06$

[6]Described in detail for electrons and muons in sections 4.4 and 4.5 .

Object	Before Preselection requirements	Preselection additional requirements
Muons	Combined or ST $p_T > 10$ GeV (4 GeV for ORL) $\|\eta\| < 2.5$ ID selection	Combined $p_T > 26\,(22)$ GeV for 2012 (2011) analysis $\mathrm{I}\,(p_T, 0.4) < 0.06$ $\mathrm{I}\,(E_T, 0.2) < 0.06$
Electrons	Medium $p_T >$15 GeV $\|\eta\| < 2.47$ discarded if $1.37 < \|\eta\| < 1.52$	Tight $p_T > 26\,(25)$ GeV for 2012 (2011) analysis $\mathrm{I}\,(p_T, 0.4) < 0.06$ $\mathrm{I}\,(E_T, 0.2) < 0.06$
Taus	BDT medium (with e veto) $p_T >$20 GeV $\|\eta\| < 2.47$ 1 or 3-prongs	
Jets	LC topo anti-k_T $p_T > 30$ GeV $\|\eta\| < 4.5$ JVF requirement (see text)	

Table 6.4: Table summarizing physics object requirements when entering the analysis and at the preselection level.

for both muons and electrons in 2012 and 2011 as well. The p_T thresholds in analysis are greater than the online trigger ones by 2 GeV to ensure maximum efficiency. So combined STACO muons are chosen, with $p_T > 26\,(22)$ GeV in 2011 (2012) respectively.
The orthogonal set of events with loose, non-isolated leptons is used to estimate QCD contribution based on data events and calculate the amount of events with jets faking τ_{had} .

2. **Charge requirement:** The single lepton and the τ_{had} candidate (exactly one lepton and one τ_{had} candidate per event) should have charges of opposite sign (OS). The SM Higgs boson, to be explored in the analysis, is a neutral scalar decaying into the pair $\tau^+\tau^-$.

3. **Mass cut:** A cut on the transverse mass, $m_T < 70$ GeV, is applied to reduce significantly the W+jets contribution. m_T is defined in eq. (6.1).

4. **b-jet veto:** A b-veto is applied to reject events with b-jet(s) with $p_T > 30$ GeV. This is needed to reduce the background contamination from top events.

Finally, even though the m_{MMC} efficiency is $\sim$98-99%, events where m_{MMC} fails to converge are discarded. A summary of the preselection cuts is given in Table 6.5.

Preselection cuts
Exactly one lepton and one medium BDT τ_{had} candidate with OS charges
Event passes the dilepton veto and b-jet veto with jet p_{T} threshold = 30 GeV
$m_T < 70$ GeV

Table 6.5: Table summarizing preselection cuts in the lep-had channel.

6.7.3 Background Normalization

At the preselection level, each background process type is handled separately and the MC sample is normalized to data using dedicated control regions. All background types undergo this normalization process, except QCD processes which are estimated from data. The background processes in question are: W+jets, top, embedded[7] $Z \to \tau\tau$, and $Z/\gamma^* \to \ell\ell$+jets. The corresponding control regions[8](CRs) are defined, in general, by the same preselection cuts except for some which are reverted (or modified).

- Z CR ($Z/\gamma^* \to \ell\ell$ + jets events): This sample is defined by cuts similar to the preselection stage but with the dilepton veto reverted. In fact, such events need 2 opposite sign (OS) same flavor leptons, with invariant mass $m_{\ell\ell}$ in the neighborhood of the Z mass peak: $61 < m_{\ell\ell} < 121$ GeV. In addition, at least one of the leptons in the event must be isolated.

- W + jets: the m_T cut is reversed, i.e. $m_T > 70$ GeV.

- Top CR: the b-veto is reversed and $m_T > 50$ GeV.

- Embedded $Z \to \tau\tau$ CR: $m_T < 40$ GeV and $m_{\tau\tau}^{\text{MMC}} < 110$ GeV cuts are applied in addition to the preselection ones.

In each CR for background type X, if N denotes the number of events of a given process, then appropriate normalization factor k is derived such that:

$$k_{\text{bkg X}} = \frac{(N_{data} - N_{\text{other bkg}})^{\text{X CR}}}{N_{bkgX}^{\text{MC, X CR}}} \ . \tag{6.7}$$

The k factors are derived at this stage to be applied in the "embedded $Z \to \tau\tau$" CR and normalize these samples correctly. An example of the k factors applied at preselection is given in Table 6.7. The details of the derivation are explained in sec. 6.9.1.

[7] Explained in detail in sec. 6.9.3.

[8] A control region is a phase space region enriched with the corresponding background type.

6.8 Event Categorization and Signal Extraction

At this level, either MVA or cut-based analysis can be performed. Since MVA is the main analysis approach, while keeping the cut-based one for cross-check, the following sections will present the details used in MVA.
After preselection, based on the event topology, two main mutually exclusive categories are defined, namely **VBF** and **boosted** categories. The phase space is thus divided in a way allowing maximum sensitivity to different signal types (VBF vs ggF+VH).

6.8.1 Analysis Categories and Signal Regions

- **VBF:** This category is designed to search for the Higgs signal produced by the vector boson fusion mechanism. Events selected in this category satisfy the following requirements :

  - ≥ 2 jets, with $p_T^{\mathrm{j1}} > 50$ GeV and $p_T^{\mathrm{j2}} > 30$ GeV
  - $\Delta\eta_{\mathrm{jj}} > 3.0$
  - $m_{\tau\tau}^{\mathrm{vis}} > 40$ GeV to remove low-mass Z/γ^* events

  Simulation studies show that 64% of signal contributions to this category come from VBF events, while 35% originate from ggF events.

- **Boosted:** aimed at studying Higgs boson in the boosted mode produced through ggF. The events in this category though have a small contribution from VBF and VH processes. Events should fail the VBF categorization cuts and have a reconstructed Higgs boson candidate transverse momentum $p_T^H > 100$ GeV. If an event passes the VBF dijet p_{T} and $\Delta\eta_{\mathrm{jj}}$ requirements while failing the $m_{\tau\tau}^{\mathrm{vis}}$, it is dropped from the analysis and is not considered for the boosted category calculations.

The categories are thus matched to two signal regions (SRs), namely VBF and boosted. The event categorization is mainly done to increase the sensitivity and have an optimized signal extraction. The signal/background discrimination itself is done effectively using boosted decision trees, as described in section 6.11.

6.8.2 Control Regions

To each category defined in sec. 6.8.1, one associates a single signal region and multiple control regions. Control regions are indeed phase space regions enriched with the corresponding background type processes. They are used in analysis to validate the background model and/or get the values of parameters related to that particular background type. For example, a W+jets CR is a phase space region dominated by W+jets background events, with little contribution from other processes.
The definitions of the CRs used in the analysis are :

1. W+**jets CR :** same as SR but with m_T cut inverted i.e. $m_T > 70$ GeV. As Fig. 6-7 shows, the W+jets transverse mass distribution peaks at 70 GeV, while most of the signal is contained within the range [0-50] GeV. At 70 GeV, the signal (ggF+VBF) and W+jets distributions plotted at the preselection level are well separated, which justifies the definition for the W+jets CR base cut $m_T > 70$ GeV in both boosted and VBF signal regions.

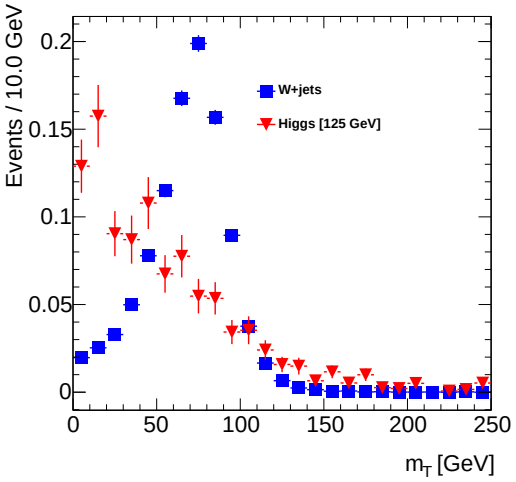


Figure 6-7: Figure showing normalized distributions of the transverse mass at the preselection level for the signal and W+jets samples.

2. **Top CR :** same as SR with inverted b-veto, i.e. with ≥ 1 b-jet(s). Two sub-control regions are defined, both with the inverted b-veto :

   – A region with $m_T < 70$ GeV, similar to the signal region m_T cut, is used to estimate the number of events with a jets mis-identified as a τ_{had} candidate, referred to as fake taus. The 'fake tau' estimate is a data-based calculation and is explained in detail in sec. 6.9.2.

   – A second orthogonal region with $m_T > 70$ GeV, used to validate and extract the correct normalization for top events with real taus in the final state (no jet faking taus events). Such top events are simulated by MC tools.

3. **Z**/$\gamma^* \to \tau\tau$ **CR :** same as SR with $m_T < 40$ GeV and $m_{\tau\tau}^{\text{MMC}} < 110$ GeV .

4. **Fake tau CR:** same as SR but using same-sign (SS) events. This region is used to validate data-based fake tau estimates, as described in sec. 6.9.2, and is needed for closure tests.

5. Z **CR** ($Z/\gamma^* \to \ell\ell$+**jets events) :** The basic definition for Z CR, on top of which SR selections or pre-selection cuts can be applied, is given in section 6.7.3. This is used in the fake tau estimations as described in section 6.9.2.

A summary of the signal and control regions cuts for the different categories is given in Table 6.6. No control region is defined for the diboson background events as they are derived from MC samples and normalized to NLO cross sections shown in Table 6.3, in addition to the fact that their contribution to the background events is negligible.

Region	VBF category	Boosted category
Signal region (SR)	≥ 2 jets $p_T^{j1(j2)} > 50\,(30)$ GeV	Fails VBF category $p_T^H > 100$ GeV
	$m_{\text{vis}} > 40$ GeV	
W+jets CR (Control region)	same as SR with inverted m_T cut : $m_T > 70$ GeV	same as SR with inverted m_T cut : $m_T > 70$ GeV
Top CR	same as SR but ≥ 1 b-tagged jet $m_T > 70$ GeV	same as SR but ≥ 1 b-tagged jet $m_T > 70$ GeV
$Z \to \tau\tau$ CR (a subset of the SR)	same as SR with: m_T <40 GeV $m_{\tau\tau}^{\text{MMC}} < 110$ GeV	same as SR with: $m_T < 40$ GeV $m_{\tau\tau}^{\text{MMC}} < 110$ GeV
Fake tau CR	same as SR but same sign	same as SR but same sign

Table 6.6: Table summarizing the VBF and boosted categories signal region (SR) and control regions (CRs) definitions.

6.8.3 Control Regions Use for the Final Analysis

It is worth noting that the lep-had final analysis results will involve events:

- with jets faking taus coming from W+jets, QCD, top (mainly $t\bar{t}$) and $Z \to \ell\ell$ + jets. The main estimation is data-based and the control regions definitions mentioned in the previous sections are used in the intermediate calculations (see section 6.9.2). These events, referred to as "fakes", have the major background contribution and sub-processes, treated equally, are not distinguishable.
- with real taus such as embedded $Z/\gamma^* \to \tau\tau$ and top events.
- with leptons faking taus e.g. $Z/\gamma^* \to \ell\ell$ + jets.

6.9 Background Estimation Methods

Background processes are estimated using data-driven methods or based on MC simulation and normalized to data in corresponding control regions. In the following

sections, the (OS-SS) method used for various background types at preselection is described first, in addition to the calculation of 'fakes' i.e. jets mis-identified as τ_{had} using the fake factor method. Then, the embedding technique for $Z \to \tau\tau$ background estimation is discussed. A summary illustrating the application of the different techniques to a given background process type, depending on the analysis stage, is given in sec. 6.10.

Background estimation with fake tau requirements

In fact, the simulation of background events with jets mis-identified as τ_{had} ($j \to \tau_{\text{had}}$) presents various challenges such as:

- simulating jets with tau-like experimental signature in terms of shower shape, tau identification criteria, particle content and track multiplicity. MC simulation is tuned to reproduce well the standard QCD jet production with event-by-event fluctuations, but it fails to do a good job in case of the unusual tau-like jets.

- achieving correct quark/gluon contribution simulation as the properties of jets originating from gluons are different from those of quark-based ones (e.g. shower width).

Data-driven techniques are thus suitable for estimating such background events and sections 6.9.1 and 6.9.2 will present two different methods used for such calculations.

6.9.1 OS-SS Method

The **'OS-SS'** background estimation method is applied at the preselection level in the $H \to \tau_\ell\tau_{\text{had}}$ analysis for all background processes [57–59]. It exploits the charge asymmetry between the opposite sign (OS) and same sign (SS) events for a given background type. The various background types in question are:

- events with a true τ_{had} and a single lepton: Contributions to this type come from $Z/\gamma^* \to \tau\tau$ mainly, in addition to diboson (W^-W^+, ZZ, $W^\pm Z$) and top (single top and $t\bar{t}$) events. The charge correlation between the single lepton and the τ_{had} is very strong[9] and N_{OS} is much larger than N_{SS}, where N_{OS} is the number of OS events ($q_\ell q_\tau < 0$) and N_{SS} is that of SS events.

- events with jets faking τ_{had} : this type of events have major contributions from QCD multi-jet and W+jets events, in addition to top events and $Z/\gamma^* \to \ell\ell$+jets. All these background processes except the latter have some charge asymmetry ($N_{\text{OS}} > N_{\text{SS}}$).

- events with a lepton mis-identified (faking) a τ_{had} candidate: these events come mainly from $Z/\gamma^* \to \ell\ell$ ($\ell = e, \mu$) and exhibit strong charge asymmetry.

[9]The charge correlation is less strong in diboson and top events than in $Z/\gamma^* \to \tau\tau$ events.

6.9.1.1 Assumptions made

1. Denoting the ratio of OS to SS QCD multi-jet events by r_{QCD}, the OS event yield can be estimated based on SS events using r_{QCD} assuming no OS/SS shape difference in observable distributions.

$$r_{\text{QCD}} = \frac{N_{\text{OS}}^{\text{QCD}}}{N_{\text{SS}}^{\text{QCD}}} \ . \tag{6.8}$$

   This assumption was found to be valid and the distributions of the discriminating variables $m_{\tau\tau}^{\text{vis}}$, MET and m_T are compatible between OS and SS samples. The SS sample is defined in the same way as the OS signal region at preselection but with the charge requirement inverted. In these estimations, the contamination from electroweak processes, between $\sim$1% and 27% depending on charge and lepton in question, was subtracted using the MC simulation samples. The highest electroweak contamination is seen in OS $e + \tau_{\text{had}}$ events. The r_{QCD} value thus measured is $r_{\text{QCD}} = 1.00 \pm 0.05\,(\text{stat.}) \pm 0.12\,(\text{syst.})$ for the $e + \tau_{\text{had}}$ subchannel and $r_{\text{QCD}} = 1.10 \pm 0.06\,(\text{stat.}) \pm 0.13\,(\text{syst.})$[10] for $\mu + \tau_{\text{had}}$ events. The systematic uncertainty seen is obtained through variations of the lepton isolation requirement in the calorimeter and of some definition requirements of the control region e.g. using 'BDT medium' taus instead of 'loose' BDT taus.

2. The MC background to data normalization factor k, derived in dedicated control regions, is the same in the signal and control region for a background process. However, the expected charge asymmetry implies different k values in OS and SS regions, namely k_{OS} and k_{SS} which are thus derived separately.

6.9.1.2 Background estimation

Since the QCD multi-jet events, where one jet simulates a lepton while another fakes a τ_{had} candidate, are not well modeled by MC generators, they are estimated based on data. A large fraction of data events in the SS control region corresponds to QCD events and they are thus used to model QCD background processes. The SS control region used is defined as the phase space region obtained using similar cuts to the OS signal region[11] while inverting the charge requirement, i.e. the single lepton and the τ_{had} candidate should have same charges ($q_\ell q_\tau > 0$). Based on the SS predictions, the total background estimation in the OS signal region at preselection is obtained using [59]:

$$N_{\text{OS}}^{bkg} \ = \ r_{\text{QCD}} \times N_{\text{SS}}^{\text{data}} + N_{\text{add-on}}^{Z\to\tau\tau} + N_{\text{add-on}}^{Z\to\ell\ell} + N_{\text{add-on}}^{W+\text{jets}} + N_{\text{add-on}}^{\text{top}} + N_{\text{add-on}}^{\text{diboson}}, \tag{6.9}$$

where:

[10] 'stat.' stands for statistical and 'syst.' for systematic uncertainty.

[11] at preselection level here.

– $Z \to \ell\ell$ denotes $Z \to \ell\ell(+\text{jets})$ where either a lepton or a jet fakes a τ_{had} :

$$N_{\text{add-on}}^{Z\to\ell\ell} = N_{\text{add-on}}^{Z\to\ell\ell(\ell\to\tau_{\text{had}})} + N_{\text{add-on}}^{Z\to\ell\ell(j\to\tau_{\text{had}})} \ . \tag{6.10}$$

– $N_{\text{SS}}^{\text{data}}$: data event yield in the SS CR.

– $N_{\text{add-on}}^{\text{bkg type}}$ terms are to account for the OS/SS differences for various background processes, and are thus needed to correct for the SS data based estimation scaled by r_{QCD} for all OS background processes, instead of being limited to QCD processes. For a background process type X (different from QCD):

$$N_{\text{add-on}}^{X} = k_{\text{OS}}^{X} \times N_{\text{OS}}^{X} - r_{\text{QCD}} \times k_{\text{SS}}^{X} \times N_{\text{SS}}^{X} \ . \tag{6.11}$$

- $k_{\text{OS (SS)}}^{X}$ is the background-to-data normalization factor, applied to the OS (SS) yield of background type X derived in the corresponding OS (SS) control region. The shape of variable distributions is still taken from simulated MC samples.

- The calculation of k follows the definition in eq. (6.7). The definition of the various CRs is given earlier in sec. 6.7.3, while the values of the normalization factors used for the 7 TeV analysis are summarized in Table 6.7. The diboson events have no dedicated CR and one assumes a unity normalization factor. The main normalization factors extracted at this level and used later on in the analysis are the $Z \to \tau\tau$ embedded k factors: $k_{Z\to\tau\tau}^{\text{el}} = 1.02$ and $k_{Z\to\tau\tau}^{\mu} = 0.68$ (at 7 TeV). When comparing the values in the e/μ channels in 2011, one should keep in mind the difference in the lepton p_{T} thresholds and the missing 3-prong $e + \tau$ veto (see sec. 4.7.5). The discrepancy seen between the e/μ normalizations is less important in 2012 samples ($k_{Z\to\tau\tau}^{\mu} \sim 0.56$, $k_{Z\to\tau\tau}^{\text{el}} \sim 0.6$), where the same lepton p_{T} threshold of 26 GeV (24 GeV trigger level) is used with a 'tight' muon selection requirement. On the other hand, in 2011, 'medium' muons are chosen with $p_{\text{T}} > 18$ GeV at the trigger level, the emulated trigger uncertainties[12] are much larger than in 2012 and the embedding systematic uncertainties are larger ($\sim 30\%$). In addition, the 3-prong $e + \tau$ veto is applied on top of the tau identification for 2012 samples. For the final fit, no normalization constraints are applied to the embedded $Z \to \tau\tau$ samples.

The 'OS-SS' method does not take into account the p_T - dependence of tau misidentification since this is not incorporated in the derivation of the normalization factors k^X described above. In addition, this method faces statistical limitations, in particular in SS data samples after applying selection cuts. As a result, a stronger, more statistically powerful method is needed such as the '*fake factor*' method described in the next section. And the 'OS-SS' estimations are limited to the preselection calculations.

[12] This is needed since the trigger simulation is not available for embedded samples.

Background process	e+τ_{had}	μ+τ_{had}	Combined
$Z \to \ell\ell$ (OS)	0.598	0.598	0.598
$Z \to \ell\ell$ (SS)	0.63	0.63	0.63
W+jets (OS)	0.588	0.55	0.565
W+jets (SS)	0.848	0.748	0.787
Top (OS)	0.982	0.933	0.954
Top(SS)	1.25	1.134	1.186

Table 6.7: Table showing the OS and SS normalization factors applied at the preselection level for MC simulated samples at 7 TeV.

6.9.1.3 r_{QCD} Calculation

Events in the QCD enriched control region used for r_{QCD} (defined in eq.(6.8)) calculations should pass $E_T^{\text{miss}} < 15$ GeV and $m_T < 30$ GeV cuts. Taus should meet the BDT 'Loose' selection criteria while the lepton should satisfy the following isolation requirements: $\text{I}(p_T, 0.4) < 0.06$ for the track isolation as in the signal region and $\text{I}(E_T, 0.2) > 0.06$ for the calorimeter isolation. The background from electroweak processes in this CR as predicted by MC simulation samples (see Table 6.8) is removed from the calculation, with the W+jets event yield normalized to data for OS and SS cases separately.
r_{QCD} for events with leptons passing the default isolation requirements was calculated as well based on a linear fit as a function of the calorimeter isolation criteria[13]. The systematic uncertainties were evaluated varying the tau BDT selection criteria (BDT medium instead of BDT loose), the isolation requirements and the fit variable. The final results are $r_{\text{QCD}} = 1.00 \pm 0.05$ (stat.) $\pm$ 0.12 (syst.) for the $e + \tau_{\text{had}}$ channel and 1.10 ± 0.06 (stat.) $\pm$ 0.13 (syst.) for $\mu + \tau_{\text{had}}$ events. The combined measurement $r_{\text{QCD}} = 1.05 \pm 0.16$ is used in the final analysis [59].

	$e + \tau_{\text{had}}$ (OS)	$e + \tau_{\text{had}}$ (SS)	$\mu + \tau_{\text{had}}$ (OS)	$\mu + \tau_{\text{had}}$ (SS)
EWK fraction	27.4%	12.6%	6.4%	1.7%

Table 6.8: Table summarizing the contamination of the electroweak processes in the QCD CR used for r_{QCD} calculations [59].

6.9.2 Fake Factor Method

6.9.2.1 Motivation

Events with jets mis-identified as ("faking") τ_{had} candidates constitute the dominant background type in the VBF category as shown in Fig. 6-30(c), and the second dom-

[13]The track isolation criteria is not considered here since it is the same for both the SR and this QCD CR but is varied in the systematic uncertainties calculations.

inant type for boosted events. They need then to be well modeled and estimated as accurately as possible. A direct MC simulation based estimation is not used for all the reasons mentioned at the beginning of sec. 6.9. Hence, a data-based technique is needed to estimate such $j \to \tau_{\text{had}}$ processes. Unlike the 'OS-SS' method, the 'fake factor' method, based on anti-τ_{had} calculation as described hereafter, is more powerful statistically and is thus chosen for the final results. This method is applied at the categorization level.

6.9.2.2 Basic Terminology

- **tau ID :** refers to τ_{had} candidates passing the tau identification criteria e.g. tau BDT medium, along with other selection cuts

- **anti-τ_{had} :** refers to τ_{had} candidates passing all tau selection criteria except the tau identification cuts, i.e. passing 'loose' but not 'medium' tau BDT requirements. In the analysis presented here, the tau candidates with a very low BDT score are not included in order to avoid significant differences between the τ_{had} and anti-τ_{had} samples, especially that the quark/gluon fraction changes considerably at low tau BDT score : For both 1-prong and 3-prong τ_{had} candidates, quark jets are dominant at high BDT score, whereas gluon and pile-up jets are dominant at very low BDT score. Therefore, a lower limit cut at 0.7×BDT 'loose' working point was chosen for anti-τ_{had}. This cut is not at a fixed BDT value since the BDT working points are chosen such that the τ efficiency is rather independent of the tau p_{T} range.

  However, for the 7 TeV analysis, the lower limit cut on anti-τ_{had} is omitted to have statistically significant samples and be able to apply the '*fake factor*' method.

- **anti-τ_{had} CR/SR :** phase space region defined in the same way as the corresponding tau ID phase space region, with the τ_{had} requirement being inverted. It could be the analog of a tau ID signal region (SR) or a background enriched control region (CR).

- **fake tau :** jet mis-identified as a τ_{had} candidate.

6.9.2.3 Method Overview

The '*fake factor*' method is based on:

1. Estimating the jet faking tau ($j \to \tau_{\text{had}}$) events coming from QCD multi-jet events, W+jets, top and $Z \to \ell\ell$+jet(s) processes in the corresponding anti-τ_{had} control region (CR), where the signal contribution is negligible. These phase space regions have high statistical power and allow good modeling of such events. An anti-τ_{had} CR can have up to 10-20 times more events than

the analog tau ID CR. This is particularly true for W+jets and QCD multi-jet events calculations for example.

2. Extrapolating $j \to \tau_{\text{had}}$ events from the anti-τ_{had} SR back to the corresponding tau ID signal region, using transfer factors called '*fake factors*'(FFs). The fake factor of a given region (CR for direct derivation or SR for validation) is defined as:

$$\text{FF}^{\text{CR(SR)}} = \frac{N_{\text{tau ID}}^{j\to\tau_{\text{had}},\text{CR(SR)}}}{N_{\text{anti-}\tau_h}^{j\to\tau_{\text{had}},\text{CR(SR)}}} \ , \tag{6.12}$$

$$N_{\text{anti-}\tau_h,CR(SR)}^{j\to\tau_{\text{had}}} = N_{\text{anti-}\tau_h,CR(SR)}^{\text{data}} - N_{\text{anti-}\tau_h,CR(SR)}^{\text{other}} \ , \tag{6.13}$$

$$N_{\text{tau ID},CR(SR)}^{j\to\tau_{\text{had}}} = N_{\text{tau ID},CR(SR)}^{\text{data}} - N_{\text{tau ID},CR(SR)}^{\text{other}} \ . \tag{6.14}$$

where "other" denotes mainly non $j \to \tau_{\text{had}}$ events, i.e. events with real taus or lepton mis-identified as τ_{had} in the final state, including $Z \to \tau_\ell\tau_h$, top (single top and $t\bar{t}$), $Z \to \ell\ell(\ell \to \tau)$ and diboson events[14] . These events labelled "other" are estimated from MC simulated samples, except for embedded $Z \to \tau\tau$ events. As a result, assuming similar background composition in the anti-τ_{had} and tau ID signal regions, the total number of fake tau events in the tau ID SR for a given category (VBF or Boosted), is given by:

$$N_{\text{fake bkg}}^{\text{SR}} = (N_{\text{anti-}\tau_h}^{\text{data, SR}} - N_{\text{anti-}\tau_h}^{\text{other, SR}}) \times \text{FF}^{\text{derived for SR}} \ . \tag{6.15}$$

6.9.2.4 FF CR Definitions

For a given background type X entering the fake tau calculation, the corresponding CR is defined as follows :

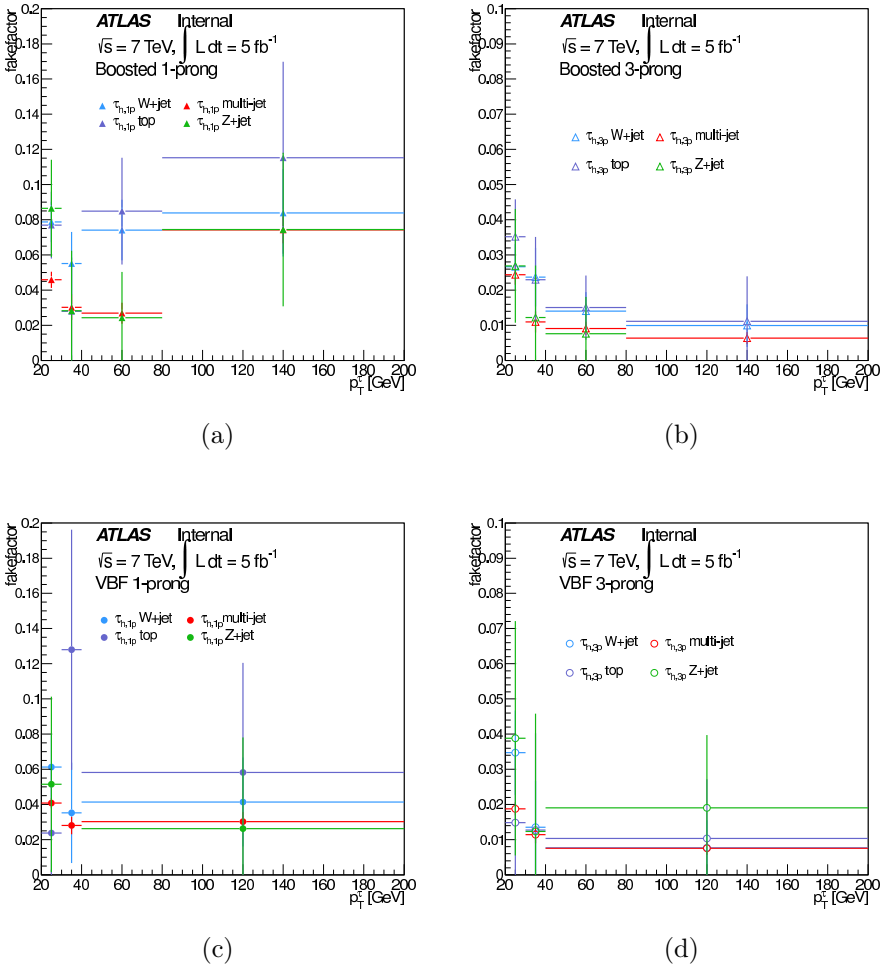
$$\text{FF CR of X} = \text{X CR at preselection with categorization cuts and appropriate } m_{\text{T}} \text{ cut applied}$$

while inverting the tau ID requirement to get anti-τ_{had} candidates. The preselection and categorization cuts follow the definitions presented in sec. 6.7.2 and 6.8.1. So the FF CRs definitions follow mainly the analysis CR definitions in sec. 6.8.2 except for the m_T requirement in some cases.

- QCD FF CR is defined with loose, non-isolated leptons (inverted track isolation and relaxed calorimeter isolation), with anti-τ_{had} candidates passing preselection and categorization cuts, in addition to m_T <70 GeV.
- Top FF CR: same as top CR used for analysis except with m_T < 70 GeV.
- W+jets FF CR: same as the analysis W+jets CR.

[14]Even though diboson events ($VV \to \ell\tau_h$, $VV \to \ell j$) may have $j \to \tau_{\text{had}}$ events, they are not included in the FF calculation and are thus listed under "other".

- $Z \to \ell\ell$+jets FF CR: same as $Z \to \ell\ell$ in analysis but without any m_T cut at any level for the 7 TeV calculations. For 7 TeV calculations, the OS requirement is dropped, i.e. (OS+SS) events are included to have a statistically significant control region.



(a) (b)

(c) (d)

Figure 6-8: 7 TeV FF values for boosted (top) and VBF(bottom) categories as derived in each fake background type CR for 1-prong (left) and 3-prong(right) separately.

6.9.2.5 Procedure

The FF value depends on the tau kinematics and properties. It is therefore defined in bins of τ-p_{T} for each of the four $j \to \tau_{\mathrm{had}}$ included process types, for both 1-prong and 3-prong taus separately. Since the FF values are to be applied to the sample of $N^{j\to\tau_{\mathrm{had}}}_{\text{anti-}\tau_h}$ data events, no direct distinction can be made amongst the various processes included, which are treated equally. Consequently, an effective FF needs to be defined for the SR calculations. Therefore, to have a correct estimate, the fraction

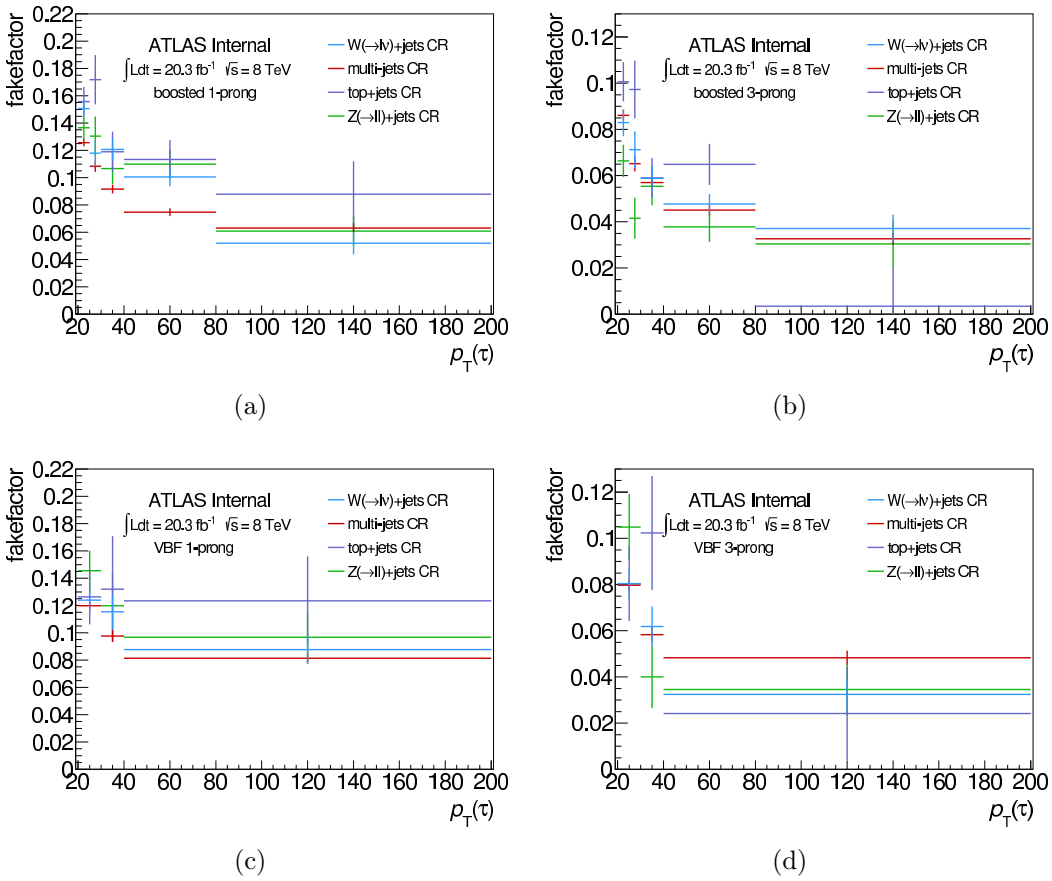


Figure 6-9: 8 TeV FF values for VBF(bottom) and boosted (top) categories as derived in each fake background type CR for 1-prong (left) and 3-prong(right) separately [60].

corresponding to each "fake tau" background process type is calculated in the anti-τ_{had} SR. This fraction of events is then multiplied by the corresponding FF_{CR}. The basic steps of FF calculation leading to "fake tau" background estimation in the tau ID signal region for a given category can be summarized as follows:

1. For each of the four background types considered for fake tau calculations (W+jets, QCD multi jet, top, $Z \to \ell\ell$+jets), calculate the corresponding FF in the dedicated control region for 1-prong and 3-prong taus separately, for both VBF and boosted (different quark/gluon jet composition). The exact definitions of all CRs used for FF calculations are given separately below. For a given background type X, $\text{FF}_X = N_{\text{tau ID}}/N_{\text{anti-}\tau_h}$ in X CR, with N as defined in eq. (6.13) and (6.14). The p_{T} binning used is [20-30,30-40,40-200] GeV and [20-30, 30-40,40-80, 80-200] GeV for VBF and boosted processes respectively. The FF values derived in the various CRs for VBF and boosted event are summarized in Fig. 6-8 and 6-9 for 7 TeV and 8 TeV analysis respectively. At 7 TeV, the FF central values are similar within statistical uncertainties. In some p_{T} bins, for 7 TeV values, the FF_{top} or $\text{FF}_{Z\ell\ell}$ central value is significantly differ-

ent from other FFs and this is mainly due to the fact that the corresponding CR is statistically poor, which is the case of $\mathrm{FF_{top}}$ in the second bin in tau p_{T} for VBF 1-prong processes for example.
On the other hand, the statistical uncertainty is not dominant at 8 TeV and the individual FF values from different CRs in a same p_{T}^{τ} bin are distinguishable in some cases, especially for boosted events and in particular at low p_{T}^{τ} where most of the boosted events are.

2. Estimate the fraction R_X of anti-τ_{had} SR data events corresponding to each background type X:

$$R_X = \frac{N_X^{\text{anti-}\tau_h, MC}}{N_{\mathrm{data}}^{\text{anti-}\tau_h}} \ . \tag{6.16}$$

The definition in eq. (6.16) applies to all included non-QCD background types except for W+jets events, where it is normalized by a correction factor, namely the background to data normalization factor derived in anti-τ_{had} W+jets CR, defined in a similar way to the one used at preselection in sec. 6.7.3 :

$$R_W = \frac{N_{\mathrm{W,\ SR}}^{\text{anti-}\tau_h, MC} \times k_W^{\text{anti-}\tau_h \mathrm{W\ CR}}}{N_{\mathrm{data,SR}}^{\text{anti-}\tau_h}} \ . \tag{6.17}$$

R_{QCD} is estimated last and is defined as :

$$R_{\mathrm{QCD}} = 1.0 - (R_W + R_{\mathrm{top}} + R_{Z\to\ell\ell}) \ . \tag{6.18}$$

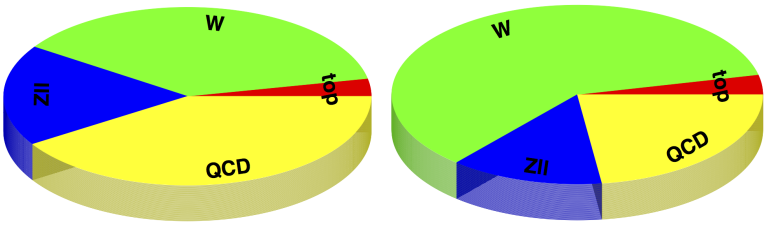
	Boost (2011)	VBF (2011)	Boost (2012)	VBF (2012)
R_W	0.75	0.60	0.62	0.46
R_{QCD}	0.13	0.24	0.26	0.40
R_Z	0.06	0.13	0.05	0.11
R_{Top}	0.06	0.03	0.07	0.03

Table 6.9: Fraction R_X of the different fake background processes as derived in the anti-τ_{had} SR for 2011 and 2012 samples.

The R_X values for the different fake tau background processes, for each of the 7 and 8 TeV analyses, are summarized in Table 6.9. The fake tau background is dominated by W+jets events (60% in VBF and 75% in boosted events at 7 TeV) and QCD multi-jet events.

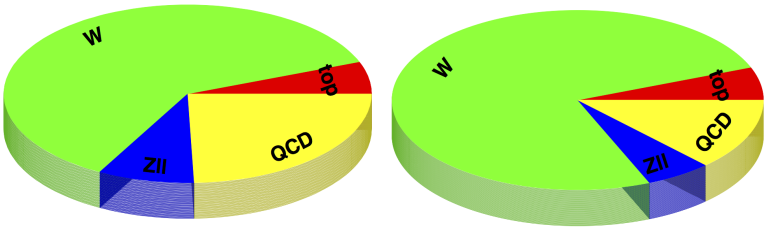
3. Calculate the effective FF defined as:

$$\begin{aligned} \mathrm{FF_{mix}}\ (p_T, n_{\mathrm{prongs}}, \mathrm{category}) &= \sum_i \mathrm{R}_i \mathrm{FF}_i\ (p_T, n_{\mathrm{prongs}}, \mathrm{category}) \\ &= \mathrm{R}_W \times \mathrm{FF}_W + \mathrm{R_{top}} \times \mathrm{FF_{top}} + \mathrm{R}_{Z\ell\ell} \times \mathrm{FF}_{Z\ell\ell} + \mathrm{R_{QCD}} \times \mathrm{FF_{QCD}} \ , \end{aligned} \tag{6.19}$$



(a) (b)

Figure 6-10: Graphic representation of the fake tau background composition for VBF $Z \to \tau\tau$ CR (left) and SR (right) at 7 TeV.



(a) (b)

Figure 6-11: Graphic representation of the fake tau background composition for boosted $Z \to \tau\tau$ CR (left) and SR (right) at 7 TeV.

where i runs over all fake tau background processes, n_{prongs} and p_{T} denote the track multiplicity and transverse momentum of the tau candidate respectively. The calculation is done independently for each of the event categories, i.e. 'VBF' and 'boosted' here. It is worth noting that the R_i values in eq.(6.19) are calculated in the anti-τ_{had} SR for the fake tau estimation in the signal region. However, this calculation is not limited to SR calculations and can be applied to any desired phase space region, as long as the corresponding fractions R_i are calculated in the anti-τ_{had} analog of this region. For example, FF_{mix} and R_i values have been derived for the $Z \to \tau_\ell\tau_h$ CR for VBF and boosted categories, for calculations needed for the signal extraction fitting procedure at later stages of the analysis. Compared to its value in the SR at 7 TeV, R_W is reduced by 20% in boosted events and by 41% in VBF events for example, while R_{QCD} is almost doubled for both categories. The difference in composition fractions R_i between SRs and $Z \to \tau_\ell\tau_h$ CR (which is a subset of the SR) of the same category at 7 TeV is illustrated in Fig. 6-10 and 6-11, while the FF_{mix} values for VBF and boosted categories are summarized in Fig. 6-12 and in Fig. 6-13 for the analyses at 7 and 8 TeV respectively. The plots show the systematic and statistical uncertainties, whose evaluation is described afterwards. The corresponding 7 TeV plots for $Z \to \tau_\ell\tau_h$ are shown in Fig. 6-14.

4. With the obtained FF_{mix} value, the total fake tau background is estimated according to eq. (6.15), replacing $\text{FF}^{\text{derived for SR}}$ by $\text{FF}_{\text{mix}}(p_{\text{T}}, n_{\text{prongs}}, \text{category})$

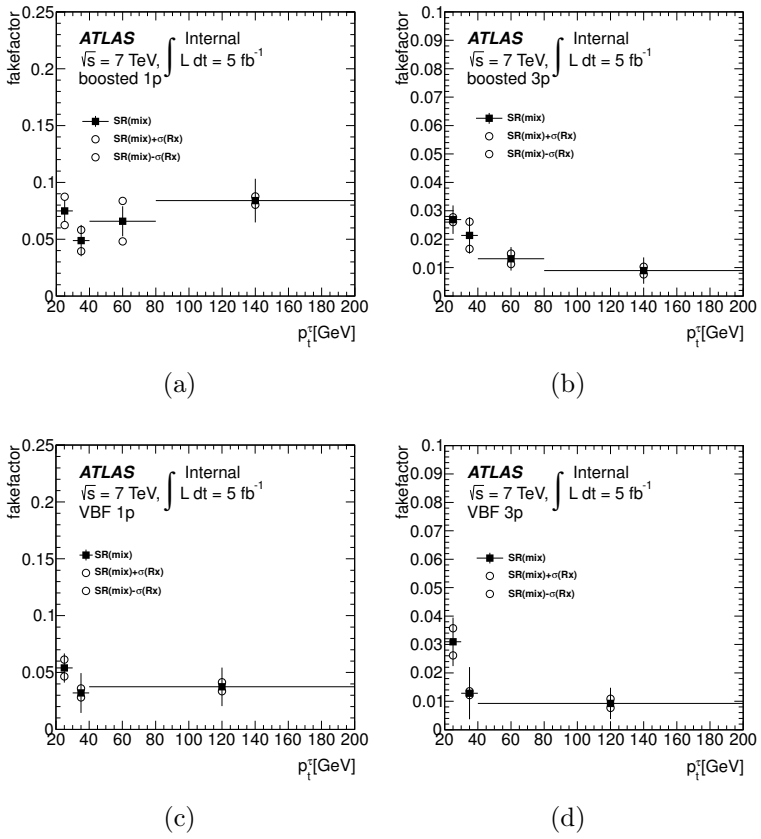


Figure 6-12: Effective FF value (FF_{mix}) in the boosted (top) and VBF (bottom) signal regions with systematic variations corresponding to the largest R_X variation for 1-prong (left) and 3-prong (right) taus at 7 TeV.

and summing in each category over all bins in p_T and n_{prongs}.

6.9.2.6 Statistical Uncertainties

In each tau p_T bin, $FF = FF_{mix} \pm \sigma(FF_{mix})$, where $\sigma(FF_{mix})$ denotes here the statistical uncertainty associated to the central FF value. So in order to estimate the statistical uncertainties, the FF estimate in each tau p_T bin is varied up/down by the statistical associated error $\sigma(FF_{mix})$. Since most of the events occur in the first one or two bins in p_T as shown in Fig. 6-15 for 7 TeV VBF events, the uncertainty from these bins has the dominant contribution. Varying FF_{mix} up/down, the fake tau background distribution as a function of the event BDT score[15] is almost flat for both VBF and boosted events as shown in Fig. 6-16. The largest uncertainty value is chosen as the statistical uncertainty. This corresponds to ± 22% (± 11%) for VBF (boosted) events at 7 TeV and to ± 4% (± 2.3%) in the 8 TeV VBF (boosted) category.

[15]The BDT score in FF calculations refers to the event BDT score. BDTs are described in sec. 6.11.

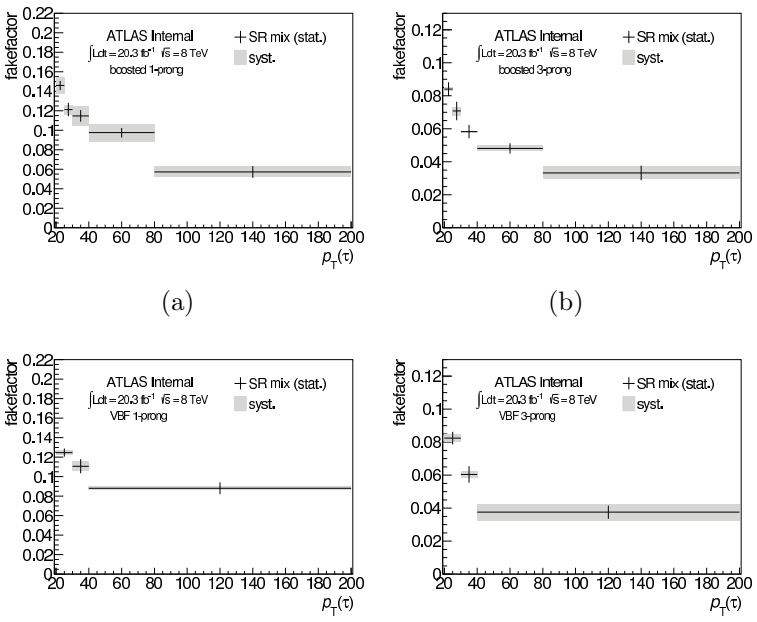


Figure 6-13: Effective FF value ($\mathrm{FF_{mix}}$) in the boosted (top) and VBF (bottom) SR for 1-prong (left) and 3-prong (right) taus at 8 TeV [60].

6.9.2.7 Systematic Uncertainties

The major source of systematic uncertainties on the effective FF, $\mathrm{FF_{mix}}$, stems from the fake background composition estimation i.e. R_X calculation. Assuming 100% uncertainty on the R_X values and varying each in the range $[R_X/2,\ R_X \times 2]$, 7 variations (including nominal) are obtained. In cases where the upper value is larger than 1.0 for a fake background process X, e.g. R_W for VBF and boosted categories at 7 TeV, R_X is set to 1 and all the remaining composition fractions $\mathrm{R}_i (i \neq X)$ vanish by definition $(\sum_j \mathrm{R}_j = 1)$ as shown in Table 6.10.
In each tau p_{T} bin, the variation resulting in the largest fluctuation of FF central value, enveloping all others, is chosen for the systematic uncertainties evaluation. Therefore, there are 2 values in each tau p_{T} bin: one for maximum upward variation and another for maximum downward variation. These FF up/down variations do not necessarily correspond to the same R_X. Indeed, the R_X change responsible for FF up/down varies from one p_T^τ bin to another. The fake background distributions based on the built up/down FF variations are then studied as a function of the event BDT score. The ratios of these distributions to the nominal one are indeed flat everywhere in the BDT score range for both VBF and boosted events as shown in Fig. 6-17. The largest of the up/down uncertainties is chosen as the final systematic uncertainty value, symmetrizing up/down variations. A $\pm$ 3% ($\pm$ 6%) systematic uncertainty is associated to VBF (boosted) events at 8 TeV, whereas the 7 TeV FF uncertainty is $\pm$ 10% ($\pm$ 15%) for VBF and boosted events. The 7 TeV analysis is thus dominated by statistical uncertainty.

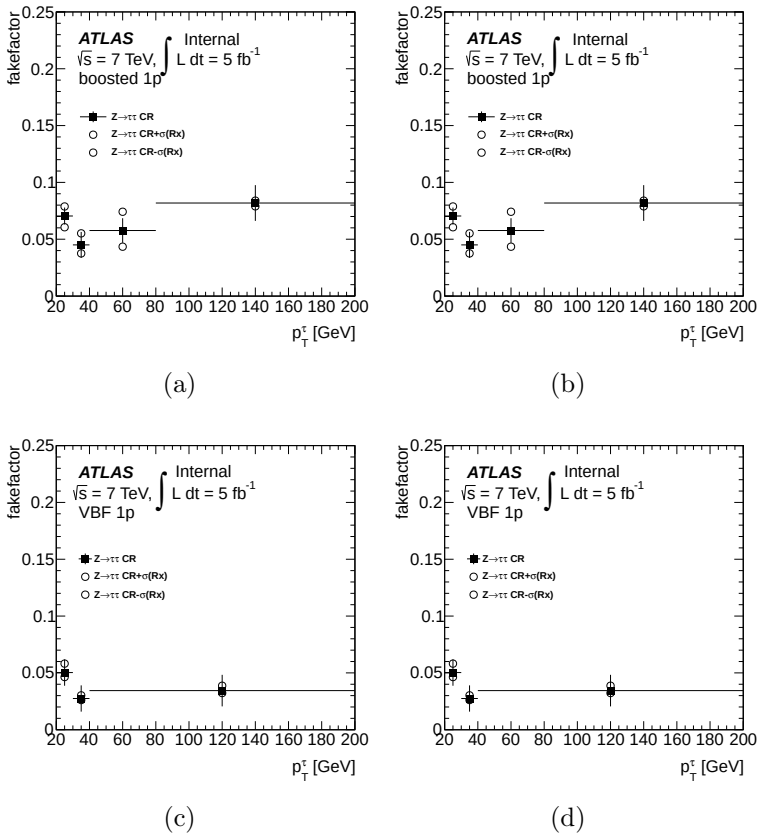
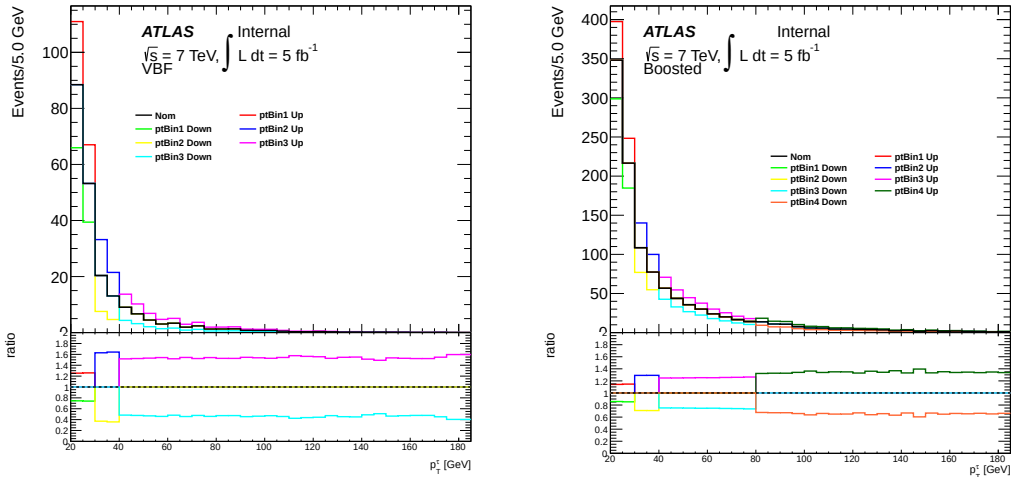


Figure 6-14: Effective FF value ($\mathrm{FF_{mix}}$) in the boosted (up) and VBF (down) $Z \to \tau\tau$ CR for 1-prong (top) and 3-prong (right) taus at 7 TeV.

6.9.2.8 Closure Tests

1. **Closure in MC Simulated Samples:**
   To check the validity of the *'fake factor'* method and assess its performance, it is tested directly on MC simulated samples. With FFs derived from p_{T}^{τ} distributions in MC fake tau background samples, a perfect agreement is expected in p_{T}^{τ} distributions by construction. It is thus necessary to check the agreement in distributions of other variables. This is denoted by 'closure test' hereafter. It was studied in $j \to \tau_h$ MC W+jets, top and $Z \to \ell\ell$ samples. QCD multi jet events can not be used for this test. 8 TeV results show no particular discrepancy within the statistical uncertainties.

   However, at 7 TeV, unlike the VBF events, boosted ones show a consistent discrepancy as shown in Fig. 6-18 to 6-20, where the fake tau background distribution, as estimated from both MC samples directly and using MC-based FF, is shown as a function of the BDT score. Using the FFs illustrated in Fig. 6-21 to estimate the fake background in the signal region as shown in Fig. 6-22, a 3% discrepancy is seen for VBF events but well covered by statistical uncertainties.



(a) (b)

Figure 6-15: Fake background distribution for different FF statistical variations as a function of p_{T}^{τ} for VBF(left) and boosted(right) events at 7 TeV.

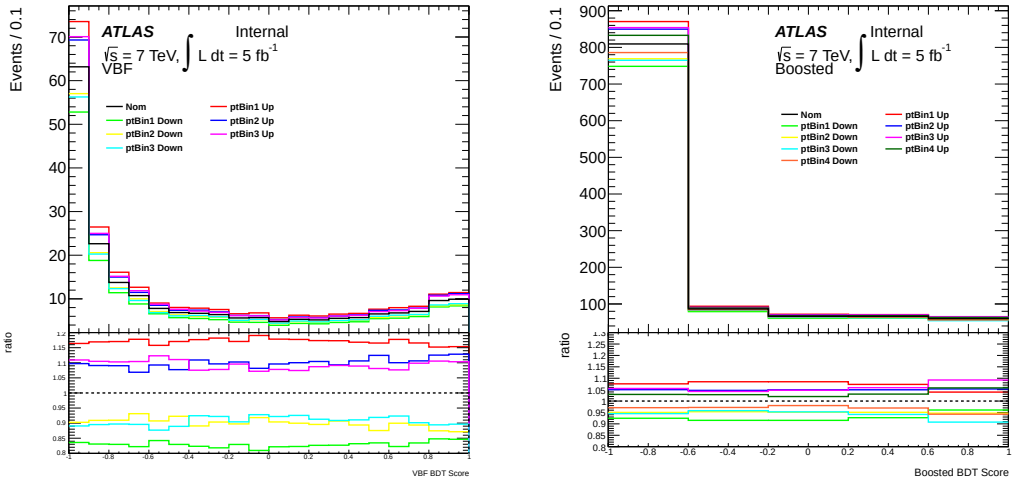
Variation Process	Nominal	up	Down
$W \to \ell\nu$	0.602	1.0	0.301
$Z \to \ell\ell$	0.134	0.0	0.134
Top	0.034	0.0	0.034
QCD	0.23	0.0	0.531

Table 6.10: Table showing variation of the fake tau composition fraction R_X in the range $[R_X/2,\ R_X \times 2]$ for systematic uncertainties evaluation in VBF events at 7 TeV.

However, the boosted events show a particular trend, with a discrepancy that is not due to the MET mismodeling in anti-τ_{had} events. The discrepancy persists even after applying a MET reweighting (see sec. 6.9.2.9) with data-based weights as in Fig. 6-24 and MC-based weights as in 6-23. As a result, a 10% systematic uncertainty is assigned to cover this discrepancy. The distributions of some basic kinematic variables are shown in Fig. 6-25.

2. **Final Closure in SS Data CR:**

A final uncertainty is to be added if there is any mis-modeling in data that is not already covered by the uncertainties derived above in the following order: statistical, systematic then MC closure uncertainties. This data-based closure test is performed in the SS CR, which is a control region dominated by fake tau events with an extremely negligible contribution from other background processes. The 8 TeV calculations show good agreement between data and fake



(a) (b)

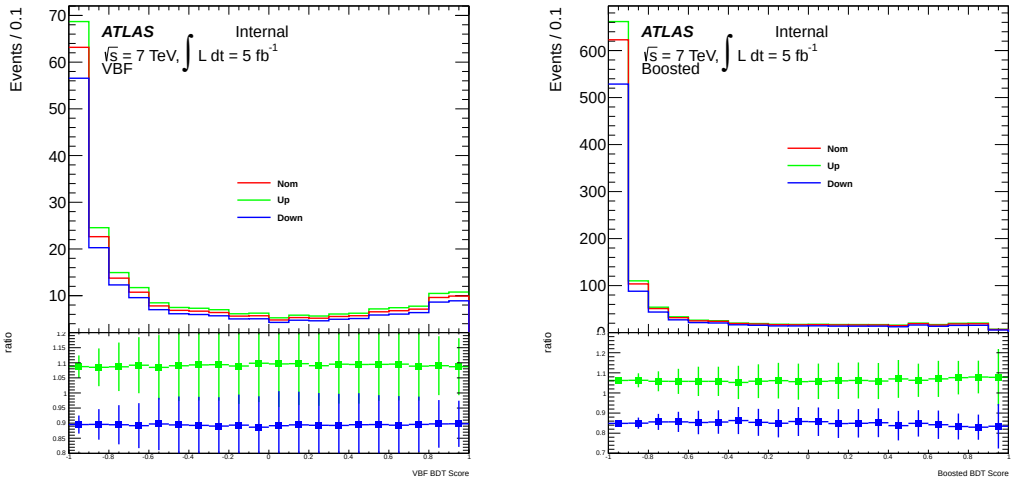
Figure 6-16: Fake background distribution for different FF statistical variations as a function of event BDT score for VBF (left) and boosted (right) events at 7 TeV. The distribution is used for fake tau statistical uncertainty estimations.

tau modeling. However, at 7 TeV, a consistent discrepancy is seen for boosted events as shown in Fig. 6-26, where the fake estimates are compared to data yields. The FFs and the SS closure plots of some kinematic variable distributions at 7 TeV are shown in Fig. 6-27 and 6-28.

6.9.2.9 MET for Anti-τ_{had}

Since the anti-τ_{had} fails the tau identification requirement, it is treated as a jet in the MET calculation. This results in mis-modelling of MET related variables, especially with the jet energy scale(JES) applied to the anti-τ_{had} while the tau energy scale (TES) would be more appropriate[16]. The MET evaluation for the 8 TeV analysis was adjusted to treat anti-τ_{had} like τ_{had} candidates, whereas a simple reweighting based on MET and tau p_{T} was done in the 7 TeV analysis. The weights are derived in the W+jets CR using the distribution of the variable $\Xi = \frac{E_T^{\text{miss}}}{p_T^{\tau_{\text{had}}}} \cos[\Delta\phi(\tau_{\text{had}}, E_T^{\text{miss}})]$. They are defined as the ratio of the number of $j \to \tau$ events in MC samples to the number of estimated 'fake tau' background, where the 'fake tau' events can be estimated either from data (Fig. 6-24) or from the MC simulation (Fig. 6-23). This results in two sets of weights, namely data-based and MC-based weights, derived separately for VBF and boosted categories. Ξ is a projection of the missing transverse energy onto the tau transverse momentum. And the choice of the W+jets CR is motivated by the purity of this CR in fake tau background events. These weights derived for fake tau background processes based on W+jets only are assigned a conservative 100% systematic uncertainty [60].

[16] JES and TES are discussed in chapter 4.



(a) (b)

Figure 6-17: Fake background distributions for FF up/down systematic variations (see text) as a function of BDT score for VBF(left) and boosted(right) events at 7 TeV.

6.9.2.10 Impact on the Final Analysis

At 8 TeV, the "fake tau" background accounts for 60% of the total background yield in the VBF category and for 44% of the boosted background events. The corresponding numbers for the 7 TeV analysis are 55% and 38% for VBF and boosted categories respectively. The fake tau background is thus the dominant background type in the VBF category, which has the highest signal sensitivity (see sec. 6.12). On the other hand, boosted events are dominated by the irreducible background $Z \to \tau\tau$.

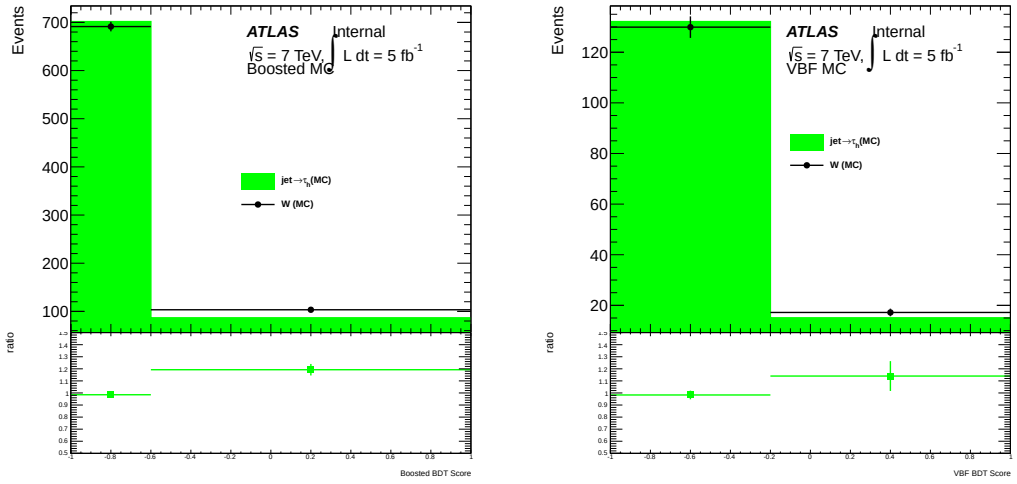
Even in the last two bins of the event BDT score distribution, where the excess corresponding to signal events is observed, 'fake tau' events account for :

- 45% of total background events in the second to last bin for the VBF(most sensitive) category. This number goes up to 48% in the last bin.
- 35.7% of total background events in the second to last bin for the boosted category. In the last bin, the corresponding number is 27.2%.

6.9.3 $Z \to \tau\tau$ Embedding

6.9.3.1 Motivation

$Z/\gamma^* \to \tau\tau$ events, which form the irreducible background, have a major contribution to background events in the $H \to \tau_\ell\tau_{\text{had}}$ analysis. A good reliable modeling of these events is thus crucial to the analysis and to have credible signal extraction results.



(a) (b)

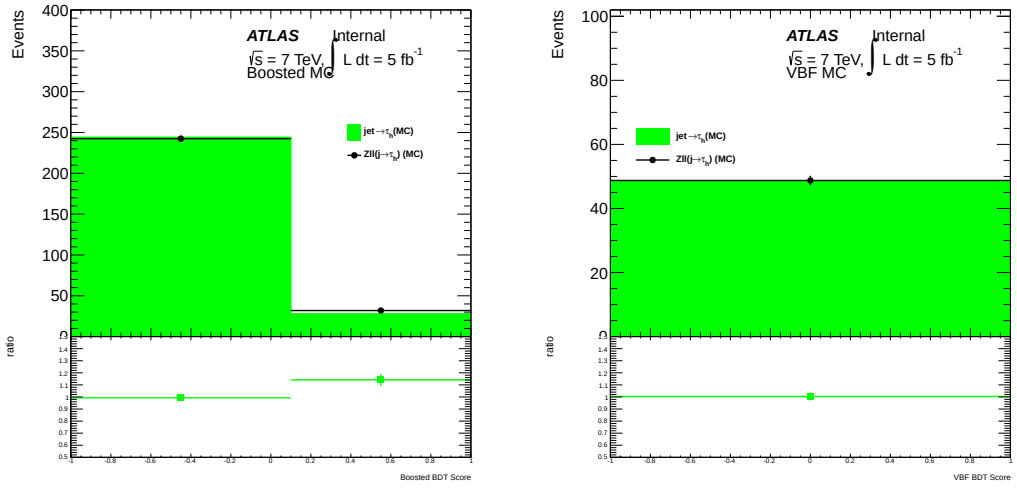
Figure 6-18: Closure plots showing fake tau distribution in W+jets CR in MC samples.

However, the $Z/\gamma^* \to \tau\tau$ estimation can not be done based on simulated MC samples only nor directly extracted from data solely for the following reasons, which explains the need of the 'embedding' procedure [2, 61]:

- The complexity of the final state and event kinematics require more than just using MC simulated samples.
- The Z + jets samples as simulated in ATLAS need to be reweighed, at the reconstruction and particle level, to model the corresponding data correctly as shown in other ATLAS analyses [62–64]. This is mainly due to some MC generator mismodelling. The reweighing of Zjj (j stands for jets) samples, where the Z boson is produced along with 2 jets, leads to better dijet invariant mass, compatible with data measurements, in all fiducial cross sections. The reweighting depends on various kinematic and topological variables e.g. jet multiplicity, pseudorapidity, jet p_{T},etc [64].
- The estimation can not be done based on data events only since it is very difficult to get a pure, signal free $Z \to \tau\tau$ sample from data, without including Higgs boson decays into di-tau pair.

6.9.3.2 Why $Z \to \mu\mu$ Events?

It is possible to get a sufficiently pure $Z \to \mu\mu$ sample from data, with negligible contamination from signal events. In fact, given the small muon mass and thus the small Higgs coupling to muons, in addition to the high efficiency of muon reconstruction, the assumption that $Z \to \mu\mu$ events with high-energy, well isolated muons are pure



(a) (b)

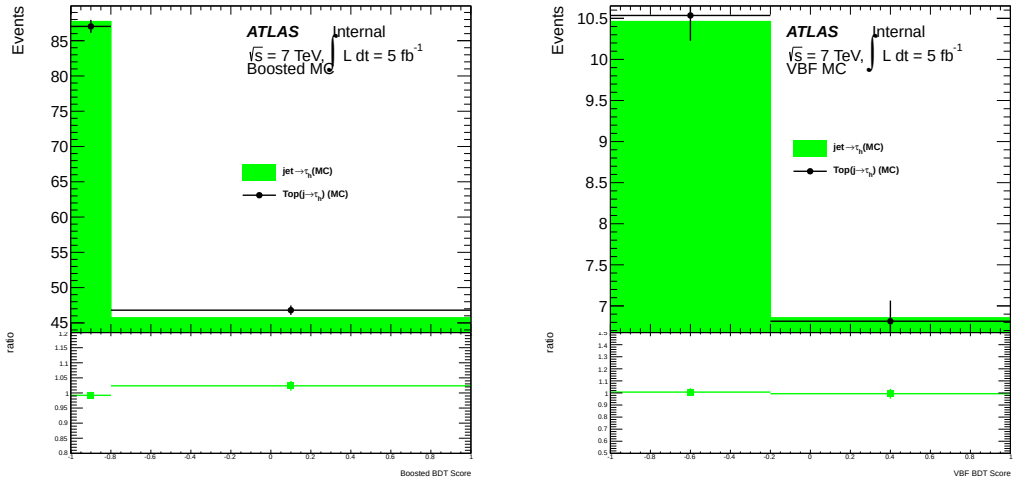
Figure 6-19: Closure plots showing fake tau distribution in $Z \to \ell\ell$ CR in MC samples.

and almost signal free can be made safely. Under the assumption of lepton universality, and apart from mass related effects, taus and muons are expected to have similar event kinematics. In addition, the 'Z+jets' event kinematics are not flavor dependent. As a result, a data-driven estimation method based on $Z \to \mu\mu$ can be used to model $Z/\gamma^* \to \tau\tau$ events, with τ related corrections derived from MC samples whenever needed.
This technique was also adapted to single-τ processes for charged Higgs studies and $W \to \tau\nu_\tau$ analyses. A similar method is used as well in the corresponding analyses in CMS [65, 66].

6.9.3.3 Procedure

Selecting $Z \to \mu\mu$ data events from collision as input, the tracks and calorimetric energy deposits left by muons in the detector are replaced by the corresponding ones coming from tau decay products. The τ kinematics are taken from the original muons though after applying necessary adjustments related to $\mu - \tau$ mass differences and spin correlation between the tau decay products. The simulation role is thus limited to generating $Z \to \tau\tau$ and τ decays, and covering the detector response to τ decay products. On the other hand, event features such as the kinematics of the Z boson, the hadronic activity (produced jets and underlying event) and pile-up, are taken from data events [61].
Then, based on reconstructed tracks and measured energy deposits, the information from both data and simulation is merged to form 'hybrid' samples called "embedded" $Z \to \tau\tau$ events. This is followed by a re-reconstruction step of the obtained events. The main embedding steps can be summarized as follows:

1. Selecting the input $Z \to \mu\mu$ data events. In case of multiple muons, the invariant



(a) (b)

Figure 6-20: Closure plots showing fake tau distribution in top CR in MC samples.

mass of muon pairs is calculated and the di-muon system having the most compatible mass with a Z boson is selected as $Z \to \mu\mu$ input candidate.

2. Generating $Z \to \tau\tau$ decays:

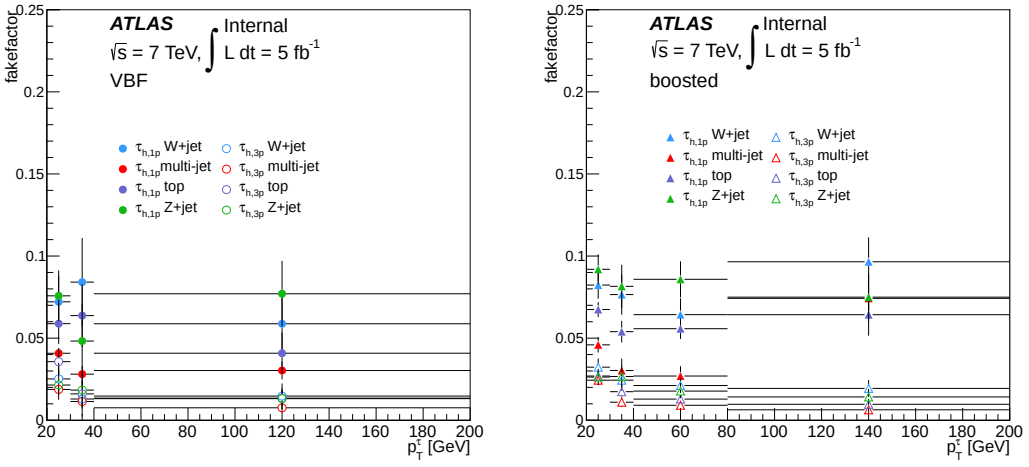
   – The τ kinematics are derived based on the data muon kinematics, while keeping the reconstructed vertex as the event primary vertex. The muon energy is kept and the resulting τ momentum is :

   $$p_\tau = \sqrt{E_\mu^2 - m_\tau^2} \ . \tag{6.20}$$

   The event generation uses PHOTOS for radiation emitted by charged leptons, while TAUOLA is used to take care of spin and polarizations of taus. However, in TAUOLA, the actual non-zero Z polarization is not obtained, since a $\pm$ 1 helicity is selected randomly for each Z boson. To correct for this, weights derived from TauSpinner [31] are applied based on calculating the most probable initial state parton configuration.

   – A kinematic filter is applied to enhance the number of $Z \to \tau\tau$ decays. This is needed because the number of embedded samples generated would be limited if it is based solely on the number of $Z \to \mu\mu$ data events passing the selection requirements before being used as input. Without the kinematic filter, very few embedded events would pass the kinematic selection of physics analyses such as $H \to \tau^+\tau^-$.

3. $Z \to \tau\tau$ detector simulation without underlying event contribution and no calorimetric noise to avoid double counting later.



(a) (b)

Figure 6-21: FF values at 7 TeV as derived directly from MC simulation samples except for QCD ones for VBF (left) and boosted (right).

4. Data-MC information merging: the detector response to muons is replaced by the τ corresponding one from simulated $Z \to \tau\tau$ obtained at the previous step.

5. Embedded events reconstruction: The ATLAS reconstruction algorithms, except the track related ones, are run over the obtained hybrid events to have a complete final state with full reconstructed physics objects.

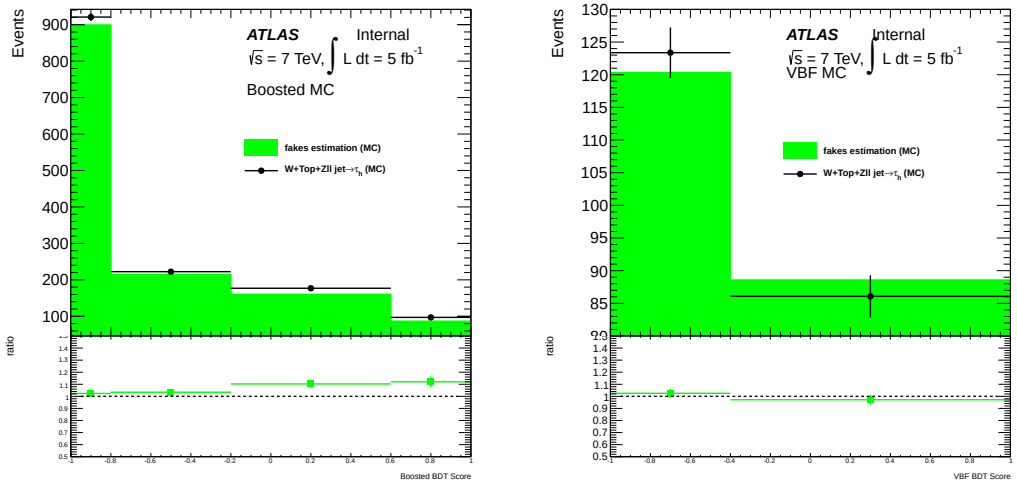
6.9.3.4 Validation

The embedded $Z \to \tau\tau$ samples have been validated against $Z \to \mu\mu$ data samples, showing that the implementation technique is reliable and produces successfully the expected results replacing muons by taus without affecting other event features. Another validation test against simulated $Z \to \tau\tau$ events, testing the tau modeling, shows agreement for various physics variables, taking into account the associated statistical and systematic uncertainties.

6.10 Background Estimation

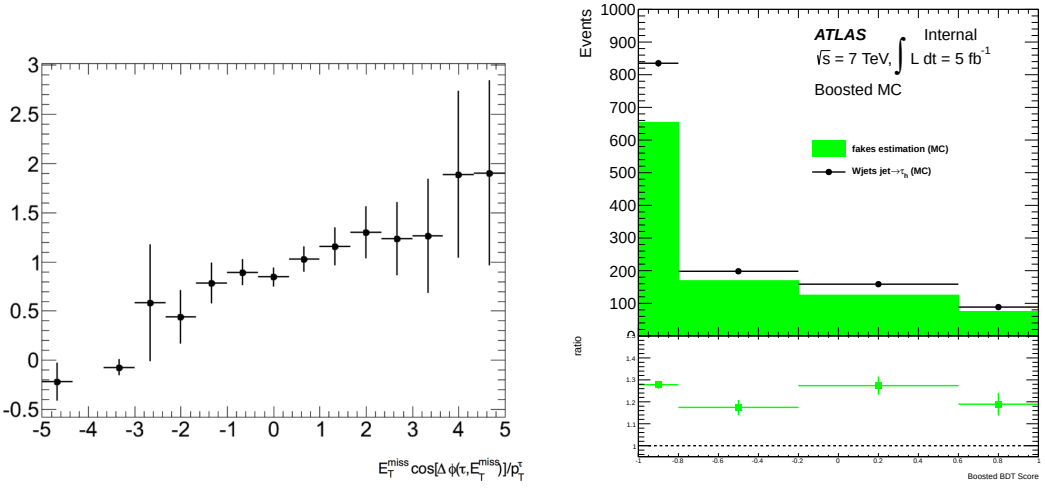
6.10.1 $Z \to \tau\tau$ Background

The $Z \to \tau\tau$ background is estimated using the embedding technique described in sec. 6.9.3. At the preselection level, (OS-SS) method (see sec. 6.9.1) is used to estimate the corresponding event yields. At the categorization level, OS yields are used. At all analysis stages, the corresponding normalization factors, derived in the $Z \to \tau\tau$ CR at the preselection level, are applied.



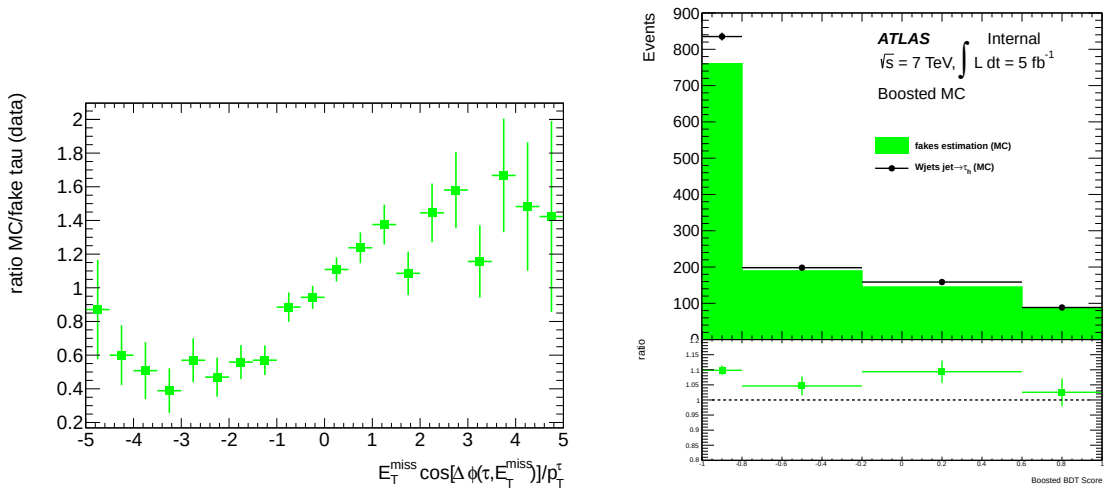
(a) (b)

Figure 6-22: Closure test in MC simulated samples boosted (left) and VBF(right) SR at 7 TeV.



(a) (b)

Figure 6-23: Closure test in MC simulated samples boosted SR (right) with MC based MET correction weights (left) at 7 TeV.



(a) (b)

Figure 6-24: Closure test in MC simulated samples boosted SR (right) with data-based MET correction weights (left).

6.10.2 Fake Tau

This background type summarizes contributions to jet faking τ_{had} events from W+jets, top (mainly $t\bar{t}$), QCD and Z+jets as described in sec. 6.9.2. The data-driven estimation of such events is applied at the categorization level.

6.10.3 W+jets

At the preselection level, W+jets events are estimated using the (OS-SS) method (see sec. 6.9.1). At the categorization level,W+jets events have jets faking τ_{had} and are estimated using the fake factor method described in sec. 6.9.2.

6.10.4 $Z \to \ell\ell$ +jets

For $Z \to \ell\ell$ (+jets) events, where $\ell = e, \mu$, the background estimation and correction depends on whether a lepton or a jet is mis-identified as τ_{had} .

- The (OS-SS) method is applied at the preselection level, regardless of the source of mis-identified τ_{had} .
- In the case of a jet mis-identified as τ_{had} candidate, the fake factor method (sec. 6.9.2) is applied at the categorization level.
- If a lepton is mis-identified as a τ_{had} candidate, then a data-based correction from corresponding tag-and-probe studies is applied.

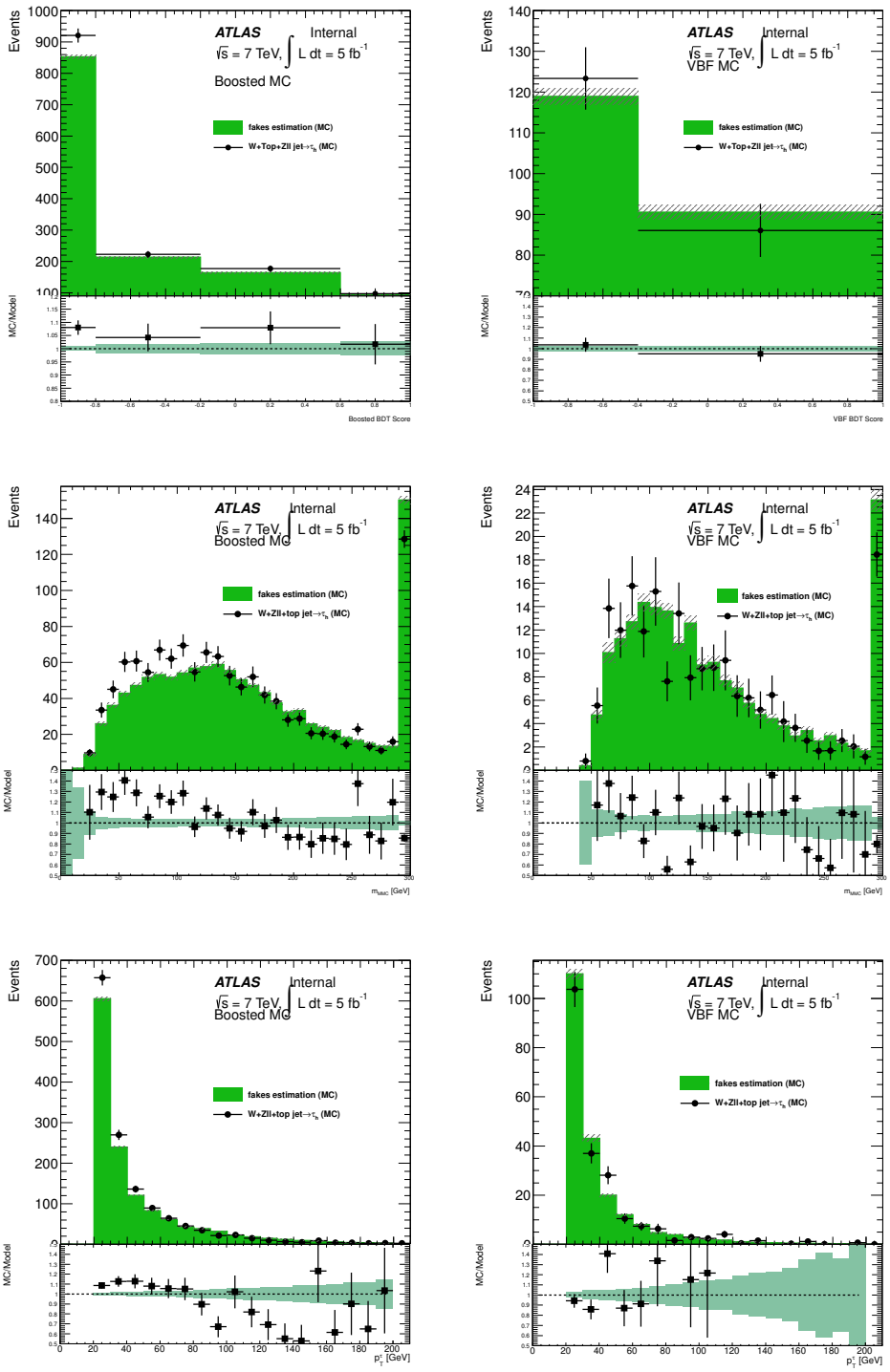
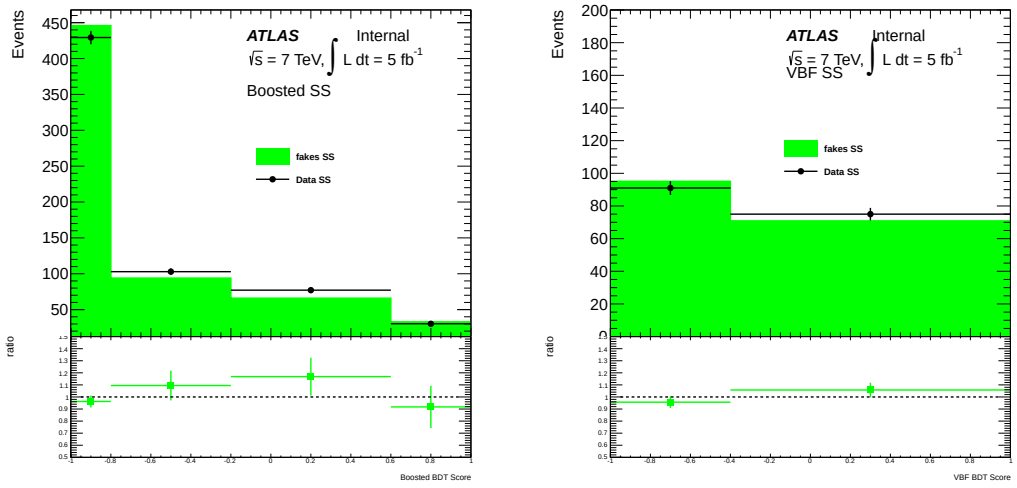


Figure 6-25: 7 TeV closure test of the FF method performed in the MC signal region for the boosted (left) and VBF (right) categories of the $\tau_\ell\tau_{\text{had}}$ channel.



(a) (b)

Figure 6-26: Closure test in fake tau SS CR for boosted (left) and VBF(right) SR at 7 TeV.

6.10.5 Top

At the preselection level, top events are estimated using the (OS-SS) method (see sec. 6.9.1). At the categorization level, OS yields for events with real taus are used after being normalized in the corresponding top CR. However, events with jets faking τ_{had} are estimated using the fake factor method described in sec. 6.9.2.

6.10.6 Diboson

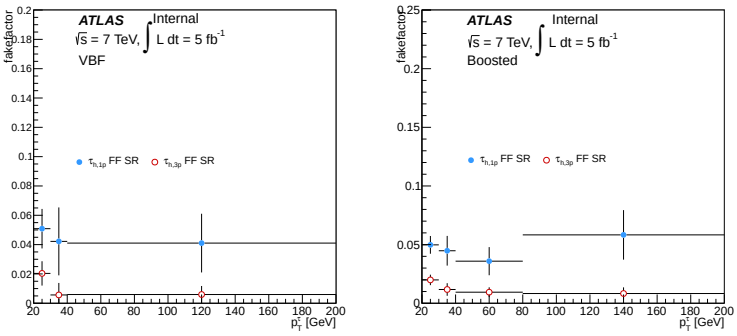
Diboson (W^+W^-, ZZ and $W^\pm Z$) processes have negligible contribution to the event yields in the lep-had analysis. At the preselection level, they are estimated using the (OS-SS) method (see sec. 6.9.1). At the categorization level, OS yields are used. At all analysis stages, simulation MC samples are used after being normalized to cross sections as shown in Table 6.3.

6.10.7 QCD

QCD multi-jet events are estimated based on data. At the preselection level, same sign (SS) data is used. At the categorization level, events with jets faking τ_{had} are estimated from data using the fake factor method described in sec. 6.9.2, where the fraction reflecting the QCD contribution is calculated in the corresponding QCD CR.

6.11 Boosted Decision Trees

The categorization cuts as defined in sec. 6.8 are mainly meant to define VBF and boosted events phase space regions. The VBF category definition does not include a



(a) (b)

Figure 6-27: FF in the SS CR for VBF (left) and boosted (right) events for both1-prong and 3-prong taus at 7 TeV.

cut on the two forward jet invariant mass m_{jj}, which is a typical VBF powerful cut for example. An additional powerful background suppression is required to achieve a good sensitivity to any potential signal. This can be done either using a cut-based approach or multi-variate analysis (MVA) techniques. The standard ATLAS $H \to \tau_\ell \tau_{\mathrm{had}}$ analysis is a MVA, with a cut-based analysis used for results cross-check purposes.
To extract the signal from overwhelming background processes, boosted decision trees (BDTs) [67–69] are used for VBF and boosted categories. BDT divide the phase space recursively in signal/background enriched regions. The final output is matched to a score varying between -1 and 1, where scores close to -1 characterize "most background like" samples while those near '+1' describe samples with highest signal purity. The BDTs were trained for signal and background processes for each analysis category for a Higgs mass m_H=125 GeV, and the most discriminating variables are kept: 9 kinematic variables are needed for VBF category and 6 are required for boosted events as shown in Table 6.11, where the variables are listed in decreasing order of ranking. The definitions of the BDT variables not described in sec. 6.3 are given below [2]:

- $\Delta \mathrm{R}\,(\tau_1, \tau_2)$: separation $\Delta \mathrm{R}\,(\ell, \tau)$ between the single lepton and τ_{had} .

- p_T^{total}: magnitude of the vector sum of the transverse momenta of forward jets, visible tau decay products(lepton, τ_{had}) and MET in the VBF final state

$$p_T^{\mathrm{total}} = |\mathbf{p}_T^{\ell} + \mathbf{p}_T^{\tau_h} + \mathbf{p}_T^{j1} + \mathbf{p}_T^{j2} + E_T^{\mathrm{miss}}| \ .$$

  Without any additional radiation in the final state, and with correct E_T^{miss} calculation, p_T^{total} vanishes. It is thus a measure of additional hadronic activity in the final state.

- Σp_T : scalar sum of the measured transverse momenta in the event (visible tau decay products (ℓ+τ_{had}) and jets).

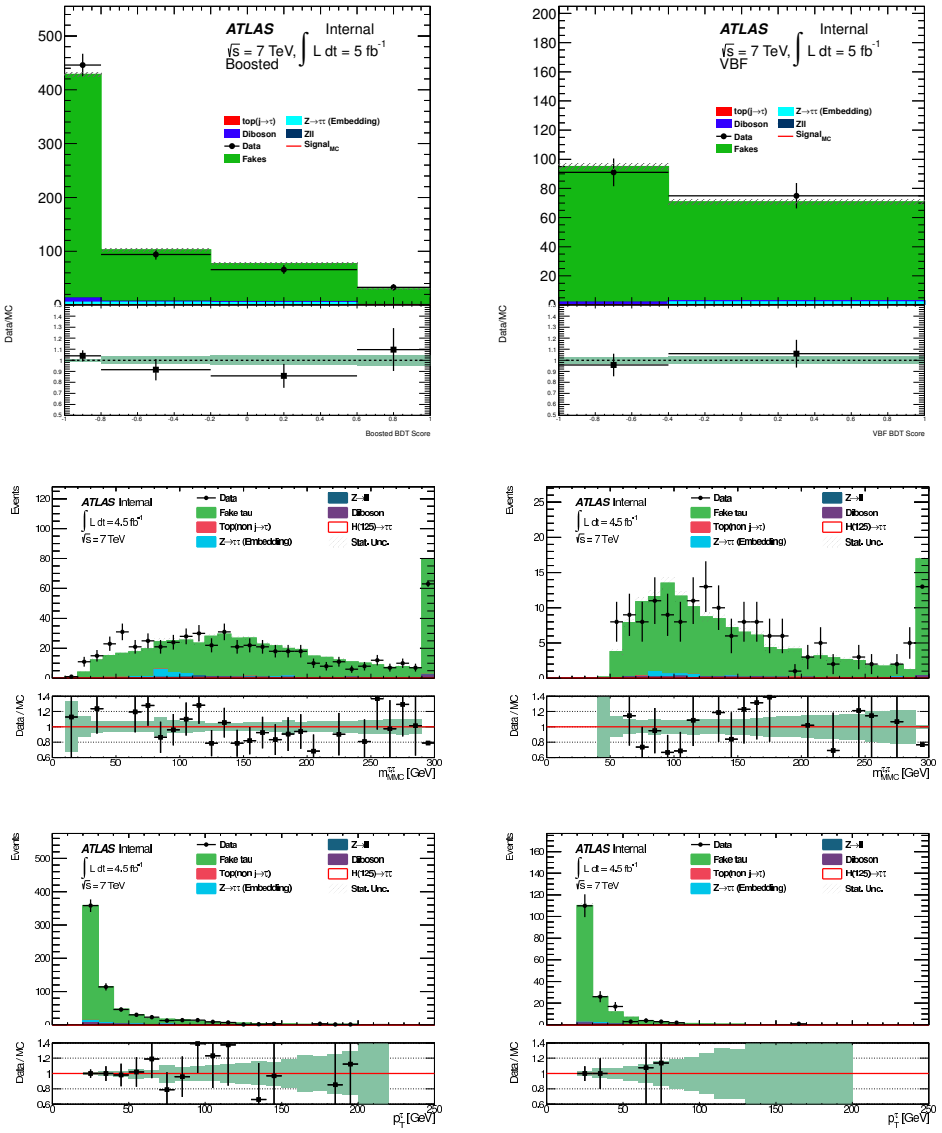


Figure 6-28: 7 TeV closure test of the FF method performed in the SS data control region for the boosted (left) and VBF (right) categories of the $\tau_\ell\tau_{\text{had}}$ channel.

- $E_T^{\text{miss}}\phi$ centrality : This variable describes the E_T^{miss} orientation with respect to the lepton and τ_{had} in the final state. The calculation is performed in the transverse xy-plane using a basis defined by the directions of the visible tau decay products. Projecting E_T^{miss} onto this basis, the E_T^{miss}centrality is defined as follows:

$$E_T^{\text{miss}}\text{centrality} = \text{MET centrality} = \frac{A+B}{\sqrt{A^2+B^2}} , \tag{6.21}$$

$$A = \frac{\sin(\phi_{MET} - \phi_{\tau_h})}{\sin(\phi_\ell - \phi_{\tau_h})}, B = \frac{\sin(\phi_\ell - \phi_{MET})}{\sin(\phi_\ell - \phi_{\tau_h})} . \tag{6.22}$$

  The variable has a value of $\sqrt{2}$ if $\vec{E}_{\text{T}}^{\text{miss}}$ falls exactly centrally between the τ_{had} and lepton ℓ, (-$\sqrt{2}$) if $\vec{E}_{\text{T}}^{\text{miss}}$ falls in the opposite of this direction, 1 if $\vec{E}_{\text{T}}^{\text{miss}}$ is aligned with one of visible tau decay products and < 1 if $\vec{E}_{\text{T}}^{\text{miss}}$ falls outside the angle spanned by the lepton and τ_{had} candidate.

- Lepton η_ℓ centrality $C_{\eta_1,\eta_2}(\eta_\ell)$: relevant for VBF events, where the indices 1 and 2 refer to the two forward jets. This variable describes the pseudorapidity position of the single lepton with respect to the two leading jets.

$$C_{\eta_1,\eta_2}(\eta_\ell) \;=\; \exp\left[\frac{-4}{(\eta_1-\eta_2)^2}\left(\eta_\ell - \frac{\eta_1+\eta_2}{2}\right)^2\right] . \tag{6.23}$$

  The variable has a value of:

  - 1 when the lepton falls centrally halfway between the two jets ($\eta_\ell = \frac{\eta_1+\eta_2}{2}$).
  - 1/e when η_ℓ= η_1 or η_2.
  - $<$ 1/e when the lepton falls outside the range delimited by the two jets.

  In VBF events, the lepton usually falls within the angle defined by the two jets. This is not the case of $Z \to \tau\tau$ events.

- p_T^τ/p_T^ℓ : this is taken as the ratio of τ_{had} p_{T} to the lepton p_T^ℓ.

- $\eta_1 \times \eta_2$: product of the pseudorapidity of two jets. This variable provides information about the di-jet separation and is used for the two VBF leading jets. Signal VBF processes usually have two well separated jets in the opposite hemispheres, leading to large negative values of $\eta_1 \times \eta_2$.

6.12 Results and Systematic Uncertainties

Based on the maximum likelihood fit of the BDT output, the signal strength is determined for each channel separately ($\ell\ell$, ℓh, hh) in addition to a combined result. The fit uses a binned likelihood function $\mathcal{L}(\mu, \vec{\theta})$, where $\vec{\theta}$ refers to nuisance parameters, with Poisson probability distributions, reflecting the effect of systematic uncertainties

VBF BDT variables	Boosted BDT variables
$\lvert\eta_1 - \eta_2\rvert$	$m_{\tau\tau}^{\text{MMC}}$
$m_{\tau\tau}^{\text{MMC}}$	$\Delta R_{\tau\ell}$
m_T	p_T^{τ}/p_T^{ℓ}
$E_T^{\text{miss}}\phi$ centrality	$E_T^{\text{miss}}\phi$ centrality
p_T^{total}	m_T
m_{jj}	Σp_T
$\Delta R_{\tau\ell}$	
$C_{\eta_1,\eta_2}(\eta_\ell)$	
$\eta_1 \times \eta_2$	

Table 6.11: The most discriminating BDT variables for VBF and boosted categories listed in decreasing order of ranking [2].

on the signal and background expectations. The signal strength μ is defined as the ratio of the signal yield to the SM expectation: $\mu = 0$ if it is compatible with the background-only hypothesis, $\mu = 1$ for the SM Higgs.

For the combined $H \to \tau^+\tau^-$ ($7 + 8$ TeV analyses),

$$\mu = 1.43^{+0.27}_{-0.26}\,(\text{stat.})\,^{+0.32}_{-0.25}\,(\text{syst.}) \pm 0.09\ (\text{theory syst.})\ \text{for}\ m_H = 125.36\ \text{GeV}.$$

The m_H value is the Higgs mass as obtained from combination studies in ATLAS. The corresponding probability of a statistical background fluctuation, p_0, is 2.7×10^{-6}, which is matched to 4.5 σ deviation from the *'background-only'* hypothesis. Figure 6-29 shows μ measurements per channel and per analysis category (VBF, boosted), in addition to the combined results for 7 and 8 TeV separately [2]. Table 6.12 shows the final post-fit event yield for the lep-had channel. Similar tables for the $\ell\ell$ and hh can be found in [2].

The expected and observed significance value for each category and per channel are summarized in Table 6.13, and the important sources of uncertainties affecting the signal strength measurements are summarized in Table 6.14. The corresponding BDT output for each channel is shown in Fig. 6-30 per analysis category and in Fig. 6-31. The excess of data events with respect to the background predictions is seen in last bins of BDT score.

6.12.1 Systematic Uncertainties

The systematic uncertainties affect the BDT output and signal/background discrimination. Their effect is found to be similar in 7 and 8 TeV analyses. The various sources of systematic uncertainties are described below and the values associated to the main sources of systematic uncertainties are summarized in Table 6.16.

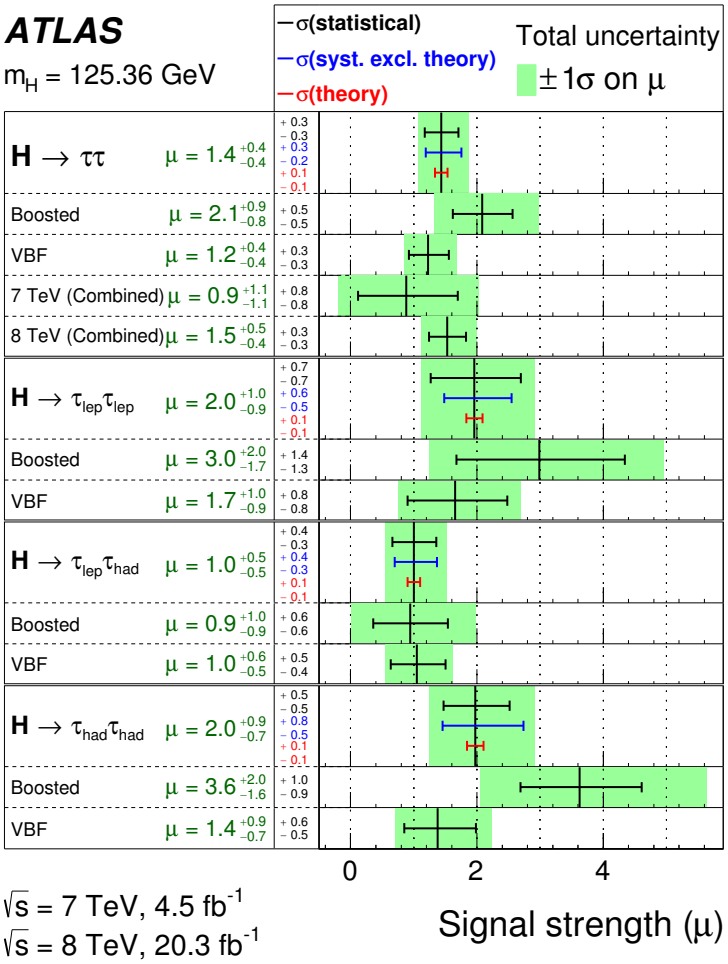


Figure 6-29: Best-fit μ values per $H \to \tau^+\tau^-$ channel and per analysis category (VBF, boosted). The green band shows the $\pm 1\sigma$ uncertainty. The contributions of statistical (black), theoretical (red) and non-theoretical (blue) systematic uncertainties are plotted separately [2].

Process/Category	VBF			Boosted		
BDT output bin	All bins	Second to last bin	Last bin	All bins	Second to last bin	Last bin
Fake background	1680 ± 50	8.2 ± 0.9	5.2 ± 0.7	5640 ± 160	51.0 ± 2.5	22.3 ± 1.8
$Z \to \tau\tau$	877 ± 29	7.6 ± 0.9	4.2 ± 0.7	6210 ± 170	57.5 ± 2.8	41.1 ± 3.2
Top	82 ± 15	0.3 ± 0.4	0.5 ± 0.4	380 ± 50	12 ± 4	4.8 ± 1.5
$Z \to \ell\ell(\ell \to \tau_{\text{had}})$	54 ± 26	1.0 ± 0.7	0.30 ± 0.28	200 ± 50	13 ± 4	8.6 ± 3.5
Diboson	63 ± 11	1.0 ± 0.4	0.48 ± 0.20	430 ± 40	9.7 ± 2.2	4.7 ± 1.6
ggF: $H \to \tau\tau$ ($m_H = 125$ GeV)	16 ± 6	1.0 ± 0.4	1.2 ± 0.6	60 ± 20	9.2 ± 3.2	10.1 ± 3.4
VBF: $H \to \tau\tau$	31 ± 8	4.5 ± 1.1	9.1 ± 2.2	16 ± 4	2.5 ± 0.6	2.9 ± 0.7
$WH: \ H \to \tau\tau$	0.6 ± 0.4	< 0.1	< 0.1	9.1 ± 2.3	1.3 ± 0.4	1.9 ± 0.5
$ZH: \ H \to \tau\tau$	0.16 ± 0.07	< 0.1	< 0.1	4.6 ± 1.2	0.77 ± 0.20	0.93 ± 0.24
Total background	2760 ± 40	18.1 ± 2.3	10.7 ± 2.7	12860 ± 110	143 ± 6	82 ± 6
Total signal	48 ± 12	5.5 ± 1.3	10.3 ± 2.5	89 ± 26	14 ± 4	16 ± 4
Data	2830	22	21	12952	170	92

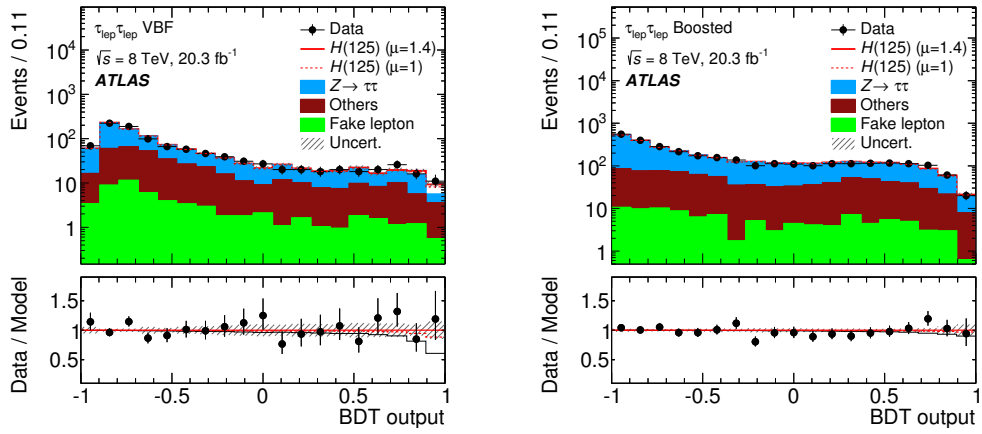
Table 6.12: Post-fit event yields in the $H \to \tau_\ell\tau_{\text{had}}$ channel for $m_H = 125$ GeV at $\sqrt{s} = 8$ TeV. Background and signal normalizations are post-fit values. Full statistical and systematic uncertainties are shown for the events under 'Total background' and 'Total signal'. The individual background components uncertainties include the systematic uncertainties only [2].

Channel and Category	Expected Significance (σ)	Observed Significance (σ)
$\tau_{\text{lep}}\tau_{\text{lep}}$ VBF	1.15	1.88
$\tau_{\text{lep}}\tau_{\text{lep}}$ Boosted	0.57	1.72
$\tau_{\text{lep}}\tau_{\text{lep}}$ Total	1.25	2.40
$\tau_{\text{lep}}\tau_{\text{had}}$ VBF	2.11	2.23
$\tau_{\text{lep}}\tau_{\text{had}}$ Boosted	1.11	1.01
$\tau_{\text{lep}}\tau_{\text{had}}$ Total	2.33	2.33
$\tau_{\text{had}}\tau_{\text{had}}$ VBF	1.70	2.23
$\tau_{\text{had}}\tau_{\text{had}}$ Boosted	0.82	2.56
$\tau_{\text{had}}\tau_{\text{had}}$ Total	1.99	3.25
Combined	3.43	4.54

Table 6.13: Table summarizing the expected and observed significances in various $H \to \tau^+\tau^-$ channel for both VBF and boosted categories[2].

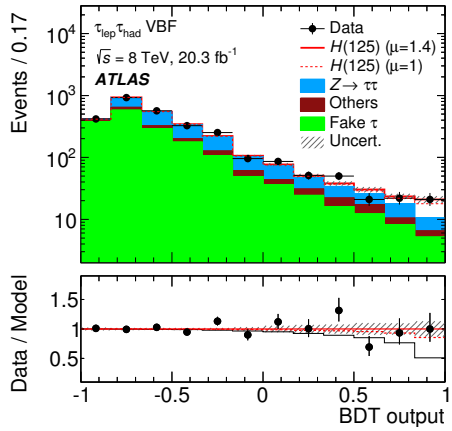
Source of Uncertainty	Uncertainty on μ
Signal region statistics (data)	$^{+0.27}_{-0.26}$
Jet energy scale	$\pm$ 0.13
Tau energy scale	$\pm$ 0.07
Tau identification	$\pm$ 0.06
Background normalisation	$\pm$ 0.12
Background estimate stat.	$\pm$ 0.10
BR ($H \to \tau\tau$)	$\pm$ 0.08
Parton shower/Underlying event	$\pm$ 0.04
PDF	$\pm$ 0.03
Total sys.	$^{+0.33}_{-0.26}$
Total	$^{+0.43}_{-0.37}$

Table 6.14: Most important sources of uncertainties affecting the signal strength measurements. Results are shown for the best-fit μ value [2].

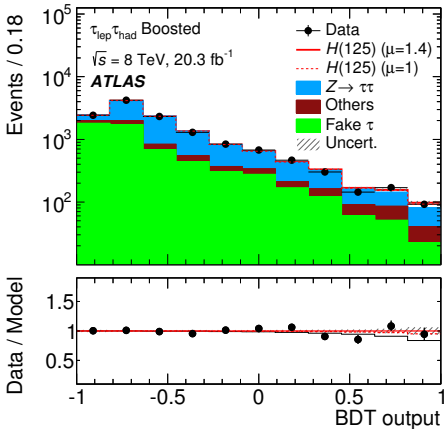


(a)

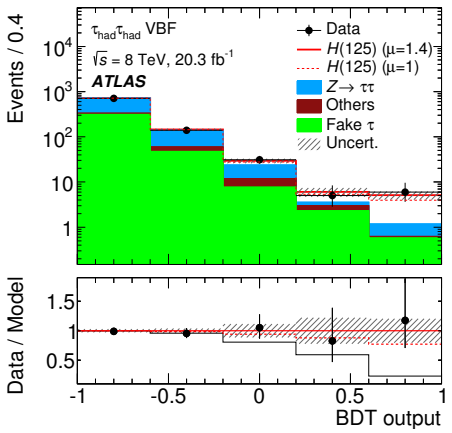
(b)



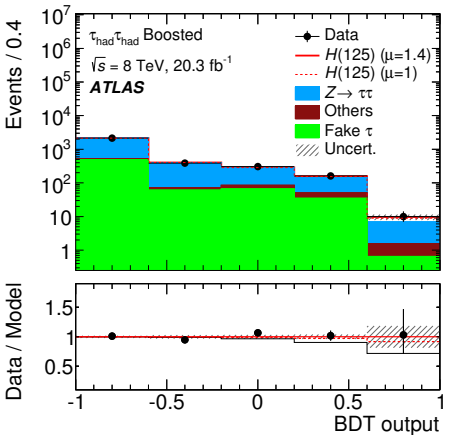
(c)



(d)



(e)



(f)

Figure 6-30: BDT score distributions for the various $H \rightarrow \tau^+\tau^-$ channels, namely $\ell\ell$ (top),ℓh (middle) and hh (bottom) at $\sqrt{s} = 8$ TeV. Post-fit results are shown for VBF (left) and boosted (right) signal regions with statistical and systematic uncertainties. The background predictions as taken from the global fit ($\mu = 1.4$) are shown. And the data to model (background+Higgs signal prediction with strength μ) ratio is shown in the lower panel of each plot for $\mu = 0.0$ (solid black line), $\mu = 1.0$ (dashed red line) and $\mu = 1.4$ (solid red line) [2].

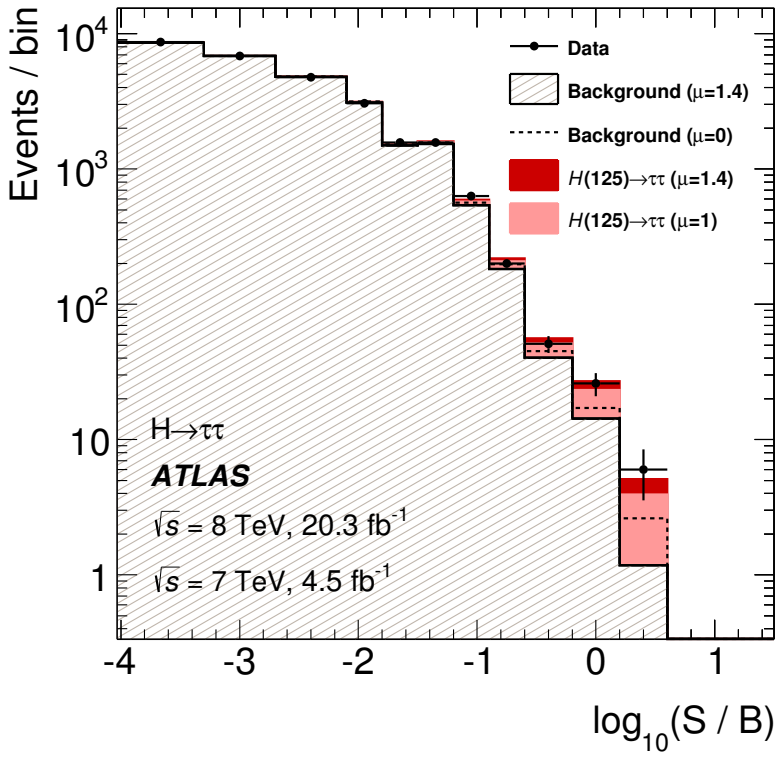


Figure 6-31: Event yields as a function of $\log_{10}(S/B)$ for all channels combined ($\ell\ell$, ℓh, hh). The signal (S) and background (B) yields are estimated based on the BDT output bin of each event, with a signal strength $\mu = 1.4$ hypothesis. All categories are taken into account. Background events are displayed for the global fit (with $\mu = 1.4$). Signal yields are shown for both $\mu = 1$ and $\mu = 1.4$ (the best-fit value) at $m_H = 125$ GeV. The dashed line corresponds to the background-only distribution obtained from the global fit with $\mu = 0$ [2].

6.12.1.1 Experimental Systematic Uncertainties

The sources of experimental systematic uncertainties include:

- Luminosity: the integrated luminosity is assigned ± 2.8% uncertainty at 8 TeV and 1.8% at 7 TeV. It estimated from beam calibration scans as explained in [70].

- Efficiencies: mainly reconstruction, identification and triggering efficiencies of detected physics objects. The τ_{had} identification efficiency is ± (2-3)% for 1-prong taus and ± 3-5% for 3-prong ones. The efficiencies for each physics object type (electron, muon, b-jet, etc.) is described in chapter 4.

- Energy scales: include jet energy scale (JES) and tau energy scale (TES). The effect of JES is limited since an important fraction of background events is estimated from data (fake taus calculations are data-based and the embedding $Z \to \tau\tau$ calculations start with data events). The energy scale variation affects the MET calculation as well.

- Energy resolutions: include systematic energy resolutions on electrons, muons, taus, jets and MET. The impact of the tau energy resolution systematics on the total uncertainty is $< 1\%$. Energy resolution uncertainty estimations are described in chapter 4.

6.12.1.2 Background Modeling Systematic Uncertainties

Uncertainty	VBF category	Boosted category
statistical	± 4.3% at 8 TeV ± 22% at 7 TeV	± 2.3% at 8 TeV ± 11% at 7 TeV
systematic	± 3% at 8 TeV ± 10% at 7 TeV	± 6% at 8 TeV ± 15% at 7 TeV
closure	-	± 10% at 7 TeV

Table 6.15: Summary of fake tau background various uncertainties values at 7 and 8 TeV.

In the lep-had channel, a major contribution to background modeling uncertainties come from fake tau calculations. The statistical and systematic uncertainties, as explained in sec. 6.9.2, are summarized in Table 6.15. An additional contribution comes from the normalization of $t\bar{t}$, whereas the largest uncertainties obtained from the global fit are associated with JES and b-tagging efficiency. The uncertainties on $Z \to \tau\tau$ embedding are $\sim \pm 1.5\%(1.2\%)$ for VBF and boosted categories respectively.

6.12.1.3 Theoretical Systematic Uncertainties

Theoretical systematic uncertainties sources include higher-order QCD corrections, simulation of the underlying event, parton shower model, generator modeling, electroweak(EWK) corrections, PDF uncertainties and branching ratio (BR) estimation uncertainty. To estimate higher order QCD corrections uncertainties in VBF and VH events, the renormalization and factorization scales are varied by a factor of 2 around the W mass. This results in ± 2-4% uncertainties. The EWK corrections lead to an additional 2% uncertainty. Similarly, the higher order QCD corrections uncertainties for ggF cross sections are estimated by varying the renormalization μ_{R} and the factorization μ_{F} scales around the central values $\mu_{\mathrm{R}} = \mu_{\mathrm{F}} = \sqrt{m_H^2 + p_T^2}$, resulting in ±24% uncertainty in boosted category and ±20% for VBF category.
The underlying event simulation uncertainties are estimated through Powheg+Pythia to Powheg+Herwig comparisons of ggF and VBF events. Additional details can be found in [2].

6.13 Track MET for $H \to \tau^+\tau^-$ Analysis

Since the standard ATLAS analysis uses MET STVF in $H \to \tau^+\tau^-$ analysis channels, the possible improvement with jet corrected track MET (see chapter 5) was studied in preparation for Run-II. MET STVF is indeed a calorimeter-based definition of missing transverse energy, with a track-based soft term correction. The soft term estimations are thus strongly affected by pile-up and the effect becomes more important at higher luminosity for the LHC Run-II, with beyond "design" average number of interactions per bunch crossing. Track MET, a track-based missing transverse energy definition, offers less pile-up dependence with no additional cost in terms of performance and efficiencies. On the contrary, the resolution with track MET is improved with signal/background discrimination. The plots in Fig. 6-32 show significant tail reduction in ΣE_T distributions upon switching from MET STVF (black curve) to jet corrected track MET (blue curve). The transverse mass distribution in the lep-lep case comparing VBF signal events with $Z \to \tau^+\tau^-$ events is shown in Fig. 6-33. Additional details on track MET can be found in chapter 5.

For Run-II, the official MET definition used by default is MET TST (Track Soft Term), which has the soft term as estimated in track MET and the hard term including the contributions from all well measured detectable physics objects calculated from calorimeter energy deposits. This definition combines the advantages of both calorimeter and track based definitions used in Run-I.

6.14 Conclusion and Prospectives

The ATLAS $H \to \tau^+\tau^-$ lep-had multi-variate analysis details are presented, showing results for both 7 and 8 TeV LHC Run-I data sets corresponding to an integrated

luminosity of 4.5 fb^{-1} and 20.3 fb^{-1} respectively. The various background estimation methods are explained, with an emphasis on the particularities of those used for the dominant $Z \to \tau\tau$ and the 'fake tau' background processes in the lep-had (ℓh) channel. At 8 TeV, the 'fake tau' ℓh background accounts for 60% of the total background yield in the VBF category (most sensitive one) and for 44% of the boosted background events. The corresponding numbers for the 7 TeV analysis are 55% and 38% for VBF and boosted categories respectively. Even in the last two bins of the event BDT score distribution, they are not negligible.

Combining measurements from all $H \to \tau^+\tau^-$ channels, for both the vector boson fusion and boosted event categories, shows an evidence of the Higgs boson with an excess of 4.5 (3.4) observed (expected) standard deviations over the expected background. The corresponding signal strength is found to be compatible with the Yukawa coupling strength of the Standard Model (SM) Higgs boson.
$\mu = 1.43^{+0.27}_{-0.26}(stat.)^{+0.32}_{-0.25}(syst.) \pm 0.09$ (theory syst.) for m_H=125.36 GeV .
Possible improvements with track-based missing transverse energy in view of Run-II were also presented.

As the center-of-mass energy at the LHC Run-II increases from 8 to 13 TeV, and the delivered luminosity is multiplied by a factor 5, $H \to \tau^+\tau^-$ events can be observed with at least 5 σ (to confirm discovery in the fermionic modes). In addition, Higgs couplings to fermions needs to be well studied to probe new physics beyond the SM. An effective parametrization of the SM 'HEFT', allowing Higgs coupling measurements in a unified and efficient framework, is presented in the next chapter.

Table 6.16: Systematic uncertainties post-fit for the signal S and the background B for the 3 $H \to \tau^+\tau^-$ channels and for both VBF and boosted analysis categories at $\sqrt{s} = 8$ TeV. The correlation amongst channels is taken into account and the unaffected uncertainties are marked with a *. The ones with † have an important effect on the BDT distribution shape. UE= underlying event and PS=parton shower [2].

Source	Relative signal and background variations [%]											
	$\tau_{\text{lep}}\tau_{\text{lep}}$ VBF		$\tau_{\text{lep}}\tau_{\text{lep}}$ Boosted		$\tau_{\text{lep}}\tau_{\text{had}}$ VBF		$\tau_{\text{lep}}\tau_{\text{had}}$ Boosted		$\tau_{\text{had}}\tau_{\text{had}}$ VBF		$\tau_{\text{had}}\tau_{\text{had}}$ Boosted	
	S	B	S	B	S	B	S	B	S	B	S	B
Experimental												
Luminosity	±2.8	±0.1	±2.8	±0.1	±2.8	±0.1	±2.8	±0.1	±2.8	±0.1	±2.8	±0.1
Tau trigger*	–	–	–	–	–	–	–	–	$^{+7.7}_{-8.8}$	< 0.1	$^{+7.8}_{-8.9}$	< 0.1
Tau identification	–	–	–	–	±3.3	±1.2	±3.3	±1.8	±6.6	±3.8	±6.6	±5.1
Lepton ident. and trigger*	$^{+1.4}_{-2.1}$	$^{+1.3}_{-1.7}$	$^{+1.4}_{-2.1}$	$^{+1.1}_{-1.5}$	±1.8	±0.5	±1.8	± 0.8	–	–	–	–
b-tagging	±1.3	±1.6	±1.6	±1.6	< 0.1	±0.2	±0.4	±0.2	–	–	–	–
τ energy scale†	–	–	–	–	±2.4	±1.3	±2.4	±0.9	±2.9	±2.5	±2.9	±2.5
Jet energy scale and resolution†	$^{+8.5}_{-9.1}$	±9.2	$^{+4.7}_{-4.9}$	$^{+3.7}_{-3.0}$	$^{+9.5}_{-8.7}$	±1.0	±3.9	±0.4	$^{+10.1}_{-8.0}$	±0.3	$^{+5.1}_{-6.2}$	±0.2
$E_{\text{T}}^{\text{miss}}$ soft scale & resolution	$^{+0.0}_{-0.2}$	$^{+0.0}_{-1.2}$	$^{+0.0}_{-0.1}$	$^{+0.0}_{-1.2}$	$^{+0.8}_{-0.3}$	±0.2	±0.4	< 0.1	±0.5	±0.2	±0.1	< 0.1
Background Model												
Modelling of fake backgrounds*†	–	± 1.2	–	±1.2	–	±2.6	–	±2.6	–	±5.2	–	±0.6
Embedding†	–	$^{+3.8}_{-4.3}$	–	$^{+6.0}_{-6.5}$	–	±1.5	–	±1.2	–	±2.2	–	±3.3
$Z \to \ell\ell$ normalisation*	–	± 2.1	–	±0.7	–	–	–	–	–	–	–	–
Theoretical												
Higher-order QCD corrections†	$^{+11.3}_{-9.1}$	±0.2	$^{+19.8}_{-15.3}$	±0.2	$^{+9.7}_{-7.6}$	±0.2	$^{+19.3}_{-14.7}$	±0.2	$^{+10.7}_{-8.2}$	< 0.1	$^{+20.3}_{-15.4}$	< 0.1
UE/PS	± 1.8	< 0.1	± 5.9	< 0.1	±3.8	< 0.1	±2.9	< 0.1	±4.6	< 0.1	±3.8	< 0.1
Generator modelling	±2.3	< 0.1	±1.2	< 0.1	±2.7	< 0.1	±1.3	< 0.1	±2.4	< 0.1	±1.2	< 0.1
EW corrections	±1.1	< 0.1	±0.4	< 0.1	±1.3	< 0.1	±0.4	< 0.1	±1.1	< 0.1	±0.4	< 0.1
PDF†	$^{+4.5}_{-5.8}$	± 0.3	$^{+6.2}_{-8.0}$	± 0.2	$^{+3.9}_{-3.6}$	± 0.2	$^{+6.6}_{-6.1}$	± 0.2	$^{+4.3}_{-4.0}$	± 0.2	$^{+6.3}_{-5.8}$	± 0.1
BR ($H \to \tau\tau$)	± 5.7	–	±5.7	–	±5.7	–	±5.7	–	± 5.7	–	±5.7	–

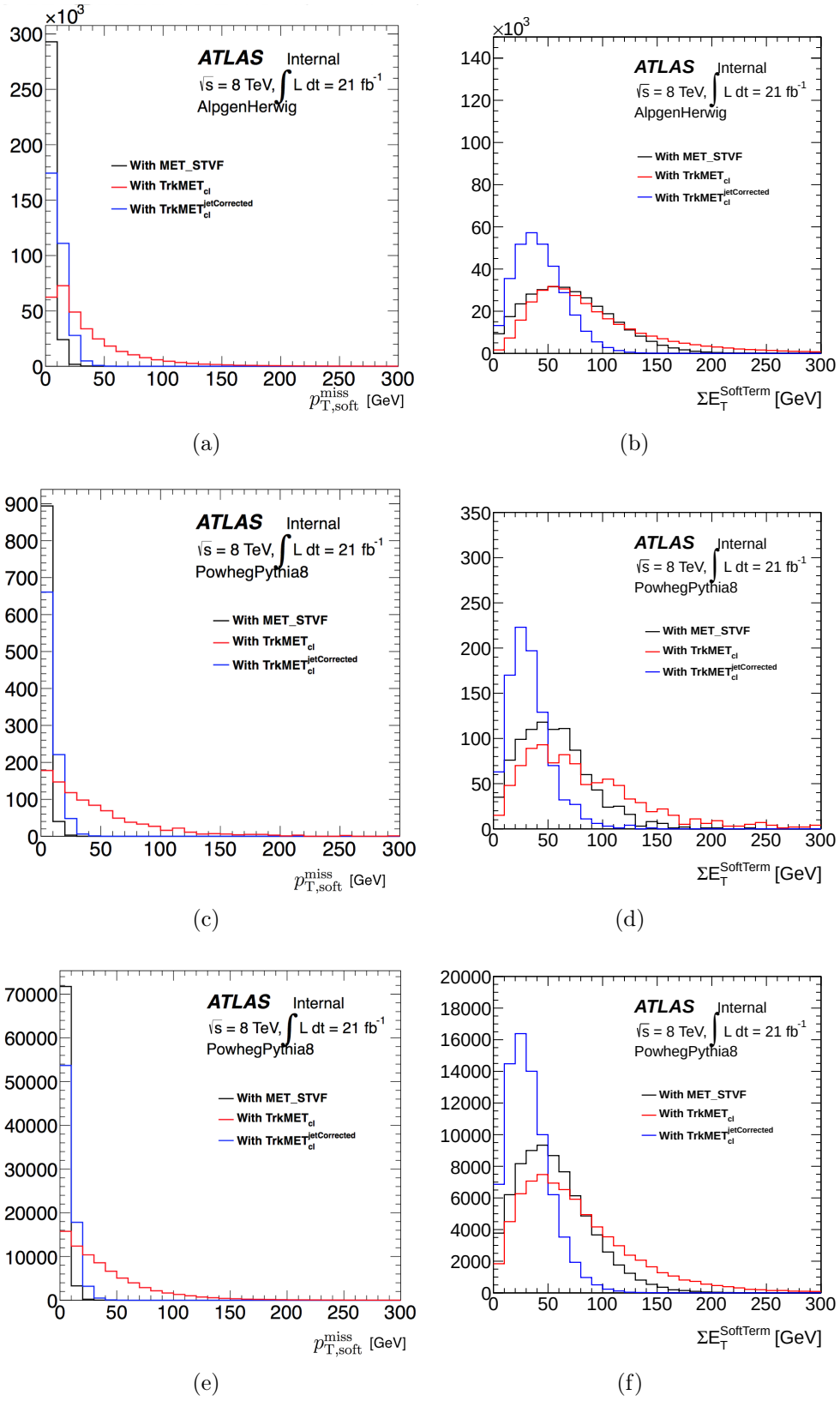


Figure 6-32: $p^{\mathrm{miss}}_{\mathrm{T,soft}}$ and $\Sigma E_T^{\mathrm{Soft\ Term}}$ comparison for different MET definitions in $Z \to \tau\tau$(top), lep-had VBF (middle) and lep-lep VBF (bottom) samples. The comparison is done between MET STVF (black), jet corrected track MET $\mathrm{TrkMET}_{c\ell}^{\mathrm{jet\ Corrected}}$(blue) and cluster Track MET (red).

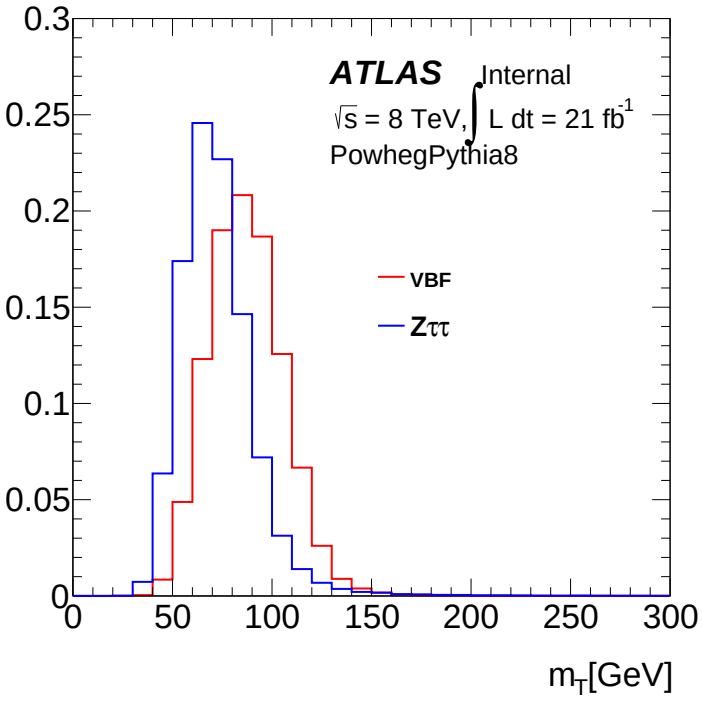


Figure 6-33: Transverse mass distribution in VBF signal and $Z \to \tau^+\tau^-$ events in the lep-lep events with jet corrected track MET used as the MET definition entering the m_T calculation.

Bibliography

[1] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, R. Tanaka (Eds.), *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*, CERN-2011-002 (CERN, Geneva, 2011)

[2] ATLAS Collaboration, *Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector*, JHEP **1504** (2015) 117

[3] A. Elagin et al., *A New Mass Reconstruction Technique for Resonances Decaying to $\tau\tau$*, Nucl. Instrum. Meth. **A654** (2011) 481-489, [arxiv:1012.4686]

[4] ATLAS Collaboration, *Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics*, arXiv:0901.0512

[5] ATLAS Collaboration,*Search for the neutral MSSM Higgs bosons decaying to $\tau^+\tau^-$ pairs in proton-proton collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector*, ATLAS-CONF-2011-132 (2011)

[6] ATLAS Collaboration, *ATLAS Twiki: Luminosity Public Results*, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults

[7] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, *JHEP* **11** (2004) 040, [hep-ph/0409146].

[8] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the* POWHEG *method*, *JHEP* **11** (2007) 070, [arXiv:0709.2092].

[9] S. Alioli et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the* POWHEG *BOX*, *JHEP* **06** (2010) 043, [arXiv:1002.2581].

[10] E. Bagnaschi, G. Degrassi, P. Slavich, and A. Vicini, *Higgs production via gluon fusion in the POWHEG approach in the SM and in the MSSM*, *JHEP* **02** (2012) 088, [arXiv:1111.2854].

[11] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A brief introduction to* PYTHIA *8.1*, *Comput. Phys. Commun.* **178** (2008) 852, [arXiv:0710.3820].

[12] H.-L. Lai et al., *New generation of parton distributions with uncertainties from global QCD analysis*, *Phys. Rev.* **D 82** (2010) 074024, [arXiv:1007.2241].

[13] A. Djouadi, M. Spira, and P. Zerwas, *Production of Higgs bosons in proton colliders: QCD corrections*, *Phys. Lett.* **B 264** (1991) 440.

[14] S. Dawson, *Radiative corrections to Higgs boson production*, *Nucl. Phys.* **B 359** (1991) 283.

[15] M. Spira et al., *Higgs boson production at the LHC*, *Nucl. Phys.* **B 453** (1995) 17, [hep-ph/9504378].

[16] R. V. Harlander and W. B. Kilgore, *Next-to-next-to-leading order Higgs production at hadron colliders*, *Phys. Rev. Lett.* **88** (2002) 201801, [hep-ph/0201206].

[17] C. Anastasiou and K. Melnikov, *Higgs boson production at hadron colliders in NNLO QCD*, *Nucl. Phys.* **B 646** (2002) 220, [hep-ph/0207004].

[18] V. Ravindran, J. Smith, and W. L. van Neerven, *NNLO corrections to the total cross-section for Higgs boson production in hadron hadron collisions*, *Nucl. Phys.* **B 665** (2003) 325, [hep-ph/0302135].

[19] S. Catani et al., *Soft gluon resummation for Higgs boson production at hadron colliders*, *JHEP* **07** (2003) 028, [hep-ph/0306211].

[20] U. Aglietti et al., *Two loop light fermion contribution to Higgs production and decays*, *Phys. Lett.* **B 595** (2004) 432, [hep-ph/0404071].

[21] S. Actis et al., *NLO electroweak corrections to Higgs boson production at hadron colliders*, *Phys. Lett.* **B 670** (2008) 12, [arXiv:0809.1301].

[22] M. Ciccolini, A. Denner, and S. Dittmaier, *Strong and electroweak corrections to the production of Higgs + 2-jets via weak interactions at the LHC*, *Phys. Rev. Lett.* **99** (2007) 161803, [arXiv:0707.0381].

[23] M. Ciccolini, A. Denner, and S. Dittmaier, *Electroweak and QCD corrections to Higgs production via vector-boson fusion at the LHC*, *Phys. Rev.* **D 77** (2008) 013002, [arXiv:0710.4749].

[24] M. Ciccolini, S. Dittmaier, and M. Krämer, *Electroweak radiative corrections to associated WH and ZH production at hadron colliders*, *Phys. Rev.* **D 68** (2003) 073003, [hep-ph/0306234].

[25] S. Heinemeyer et al. (LHC Higgs Cross Section Working Group), *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, *CERN-2013-004* (2013) , [arXiv:1307.1347].

[26] K. Arnold et al., *VBFNLO: A parton level Monte Carlo for processes with electroweak bosons*, *Comput. Phys. Commun.* **180** (2009) 1661, [arXiv:0811.4559].

[27] P. Bolzoni et al., *Higgs production via vector-boson fusion at NNLO in QCD*, *Phys. Rev. Lett.* **105** (2010) 011801, [arXiv:1003.4451].

[28] O. Brein, A. Djouadi, and R. Harlander, *NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders*, *Phys. Lett.* **B 579** (2004) 149, [hep-ph/0307206].

[29] M. Grazzini and H. Sargsyan, *Heavy-quark mass effects in Higgs boson production at the LHC*, *JHEP* **09** (2013) 129, [arXiv:1306.4581].

[30] D. de Florian et al., *Higgs boson production at the LHC: transverse momentum resummation effects in the $H \to 2\gamma, H \to WW \to \ell\nu\ell\nu$ and $H \to ZZ \to 4\ell$ decay modes*, *JHEP* **06** (2012) 132, [arXiv:1203.6321].

[31] Z. Czyczula, T. Przedzinski, and Z. Was, *TauSpinner Program for Studies on Spin Effect in tau Production at the LHC*, *Eur. Phys. J.* **C 72** (2012) 1988, [arXiv:1201.0117].

[32] T. Sjostrand, S. Mrenna, and P. Skands, PYTHIA *6.4 physics and manual*, *JHEP* **05** (2006) 026, [hep-ph/0603175].

[33] G. Marchesini, B.R. Webber, G.Abbiendi, I.G.Knowles, M.H.Seymour, L.Stanco *Computer Physics Communications* **67** (1992) 465.

[34] G.Corcella, I.G.Knowles, G.Marchesini, S.Moretti, K.Odagiri, P.Richardson, M.H.Seymour, B.R.Webber *JHEP* **0101** (2001) 010, [hep-ph/0011363].

[35] S. Jadach et al., *The tau decay library Tauola: Version 2.4*, *Comput. Phys. Commun.* **76** (1993) 361.

[36] N. Davidson, T. Przedzinski, and Z. Was, *Photos Interface in C++: Technical and Physics Documentation*, (2010), [arXiv:1011.0937].

[37] M. Mangano et al., ALPGEN*, a generator for hard multiparton processes in hadronic collisions*, *JHEP* **07** (2003) 001, [hep-ph/0206293].

[38] M. Mangano et al., *Multijet matrix elements and shower evolution in hadronic collisions: $Wb\bar{b}$ + n jets as a case study*, *Nucl. Phys.* **B 632** (2002) 343, [hep-ph/0108069].

[39] T. Binoth et al., *Gluon-induced W-boson pair production at the LHC*, *JHEP* **12** (2006) 046, [hep-ph/0611170].

[40] B. P. Kersevan and E. Richter-Was, *The Monte Carlo event generator AcerMC versions 2.0 to 3.8 with interfaces to PYTHIA 6.4, HERWIG 6.5 and ARIADNE 4.1*, *Comput. Phys. Commun.* **184** (2013) 919, [hep-ph/0405247].

[41] GEANT4 Collaboration, S. Agostinelli et al., *Geant4– A Simulation toolkit*, *Nucl. Instrum. Meth.* **A 506** (2003) 250.

[42] ATLAS Collaboration, *Summary of ATLAS* PYTHIA *8 tunes*, ATL-PHYS-PUB-2012-003 (2012). http://cds.cern.ch/record/1474107.

[43] S. Catani et al., *Vector boson production at hadron colliders: A fully exclusive QCD calculation at next-to-next-to-leading order*, *Phys. Rev. Lett.* **103** (2009) 082001, [arXiv:0903.2120].

[44] S. Catani and M. Grazzini, *Next-to-next-to-leading-order subtraction formalism in hadron collisions and its application to Higgs-boson production at the Large Hadron Collider*, *Phys. Rev. Lett.* **98** (2007) 222002.

[45] G. Corcella et al., HERWIG *6.5 release note*, (2002), [hep-ph/0210213].

[46] T. Gleisberg et al., *Event generation with SHERPA 1.1*, *JHEP* **02** (2009) 007, [arXiv:0811.4622].

[47] M. Cacciari et al., *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*, *Phys. Lett.* **B 710** (2012) 612, [arXiv:1111.5869].

[48] P. Bärnreuther, M. Czakon and A. Mitov, *Percent Level Precision Physics at the Tevatron: First Genuine NNLO QCD Corrections to* $q\bar{q} \to t\bar{t} + X$, *Phys. Rev. Lett.* **109** (2012) 132001, [arXiv:1204.5201].

[49] M. Czakon and A. Mitov, *NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels*, *JHEP* **12** (2012) 054, [arXiv:1207.0236].

[50] M. Czakon and A. Mitov, *NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction*, *JHEP* **01** (2013) 080, [arXiv:1210.6832].

[51] M. Czakon, P. Fiedler, A. Mitov, *The total top quark pair production cross-section at hadron colliders through* $\mathcal{O}(\alpha_s^4)$, *Phys. Rev. Lett.* **110** (2013) 252004, [arXiv:1303.6254].

[52] M. Czakon and A. Mitov, *Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders*, *Comput. Phys. Commun.* **185** (2014) 2930, [arXiv:1112.5675].

[53] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated production with a* W^- *or* H^-, *Phys. Rev.* **D 82** (2010) 054018, [arXiv:1005.4451].

[54] N. Kidonakis, *Next-to-next-to-leading logarithm resummation for s-channel single top quark production*, *Phys. Rev.* **D 81** (2010) 054028, [arXiv:1001.5034].

[55] N. Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production*, *Phys. Rev.* **D 83** (2011) 091503, [arXiv:1103.2792].

[56] J. M. Campbell, K. R. Ellis, and C. Williams, *Vector boson pair production at the LHC*, *JHEP* **07** (2011) 018, [arXiv:1105.0020].

[57] ATLAS Collaboration, *Search for the Standard Model Higgs boson in $H \to \tau^+\tau^-$ decays in proton-proton collisions with the ATLAS detector*, ATLAS-CONF-2012-160.

[58] ATLAS Collaboration, *Search for the Standard Model Higgs boson in the $H \to \tau^+\tau^-$ decay mode in $\sqrt{s} = 7$ TeV pp collisions with ATLAS*, JHEP **1209** (2012) 070, [arXiv:1206.5971].

[59] ATLAS Collaboration, *Evidence for Higgs Boson Decays to the $\tau^+\tau^-$ Final State with the ATLAS detector*, ATLAS-CONF-2012-160.

[60] ATLAS Collaboration, *Measurement of the Higgs boson couplings in the $\tau\tau$ final state with the ATLAS detector*, (2014), https://cds.cern.ch/record/1666539/

[61] ATLAS Collaboration, *Modelling $Z \to \tau\tau$ processes in ATLAS with τ-embedded $Z \to \mu\mu$ data*, arXiv:1506.05623 [hep-ex], CERN-PH-EP-2015-130

[62] ATLAS Collaboration, *Measurement of the production cross section of jets in association with a Z boson in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *JHEP* **07** (2013) 032, [arXiv:1304.7098].

[63] ATLAS Collaboration, *Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *JHEP* **09** (2014) 145, [arXiv:1406.3660].

[64] ATLAS Collaboration, *Measurement of the electroweak production of dijets in association with a Z-boson and distributions sensitive to vector boson fusion in proton-proton collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, *JHEP* **04** (2014) 031, [arXiv:1401.7610].

[65] CMS Collaboration, *Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons*, *JHEP* **1405** (2014) 104, [arXiv:1401.5041].

[66] CMS Collaboration, *Search for neutral Higgs bosons decaying to tau pairs in pp collisions at $\sqrt{s} = 7$ TeV*, Phys. Lett. B**713** (2012) 68-90, [arXiv:1202.4083].

[67] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*. Chapman & Hall, New York, 1984.

[68] J. Friedman, *Stochastic gradient boosting*, *Comput. Stat. Data Anal.* **38** (2002) 367.

[69] Y. Freund and R. E. Schapire, *A decision-theoretic generalization of on-line learning and an application to boosting*, *J. Comput. Syst. Sci.* **55** (1997) 119.

[70] ATLAS Collaboration, *Improved luminosity determination in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector at the LHC*, *Eur. Phys. J.* **C 73** (2013) 2518, [arXiv:1302.4393].

Chapter 7

Higgs Effective Field Theory and Tau Model Validation in Madgraph5_aMC@NLO

The testing and validation of the tau model within the Monte Carlo generator Madgraph5_aMC@NLO is presented in this chapter. The final test results are done for the $H \to \tau^+\tau^-$ decay mode using the Higgs effective field theory (HEFT) as implemented in the Higgs characterization (HC) framework. This work was done during a short MCnet internship with the Madgraph5_aMC@NLO team at the University of Louvain (UC Louvain).
HEFT is presented first in sec. 7.1, followed by a discussion of its importance for the LHC Run-II in sec. 7.2. Then, the Higgs spin/CP Monte Carlo tools are discussed in sec. 7.3. Afterwards, the HC framework is presented in sec. 7.4, with the results showing its application in various Higgs analyses summarized in sec. 7.5. My personal contribution consisting in testing and validating the tau model (presented in sec. 7.6) with the HC framework is discussed in sec. 7.7. Conclusions and prospectives for LHC Run-II are summarized at the end of this chapter (sec. 7.8).

7.1 Higgs Effective Field Theory (HEFT)

7.1.1 Introduction

The LHC Run-I experimental results by the ATLAS and CMS collaborations [1, 2] confirm the discovery of a 125 GeV scalar resonance compatible with the Standard Model (SM) Higgs boson. This suggests that any new physics (NP) scale is much larger than the electroweak (EW) one, unless NP couples very weakly to SM particles so that the properties of the SM-like Higgs boson are not affected. As a result, an effective field theory (EFT) can be used to study the Higgs physics, parametrizing NP at low energies using a series of local operators. The higher order operators are built using SM fields and are organized in inverse powers of the new physics scale Λ [3, 4].

During the LHC Run-I, studies put a major emphasis on the on-shell single-production of the Higgs boson and its decay modes, probing thus the EW symmetry breaking (EWSB) physics at the scale Q = m_h ≡ Higgs mass. As a result, the corresponding Higgs coupling corrections are of the order of $\delta c/c \sim (g_*^2/g_{\rm SM}^2)m_h^2/m_*^2$, where m_* denotes the mass of new states, g_* refers to their coupling strength and $g_{\rm SM}$ is the SM coupling. However, the higher center of mass energy and luminosity during Run-II will allow to probe the EWSB dynamics directly at a higher scale Q ∼ E » m_h, where E is the 2 →2 processes scattering energy. The expected NP effects are then approximately $(g_*^2/g_{\rm SM}^2)E^2/m_*^2 = E^2/m_h^2 \times$ values from on-shell Higgs processes [4].

7.1.2 Motivation

The main advantages of an EFT are [5, 6]:

- Probing NP in a model independent way: neutrino masses, dark matter, inflation and baryon asymmetry suggest that physics beyond the Standard Model (BSM) exists and there are several theoretical hints pushing in that direction e.g. the strong CP problem, flavor hierarchies, gauge coupling unifications, etc. There are many models describing BSM physics with different assumptions but there is no strongly preferred model. With the EFT approach, however, BSM NP effects are described in a model independent way. One can then extrapolate from the observed processes, considered as effective limits of more fundamental physics, to a more complete theory of high energy physics.

- Having a reduced number of degrees of freedom at low energies, where only the relevant ones are kept in the EFT framework.

- Studying physics when the high and low energy physics matching is not well known: With EFT, heavy particle effects are incorporated in the contact interactions of low energy particles and the ones that cannot be produced at low energies are integrated out.

- Automatic inclusion of constraints of gauge invariance, analyticity and unitarity while having a finite number of constants for a given order in $(1/\Lambda)$.

A thorough treatment of EFTs is given in [5, 7–9].

7.1.2.1 EFT Examples

There are many examples in particle physics where effective theories are useful such as:

– Fermi theory of weak interactions where a muon decay, for example, can be described using an effective vertex with 4-fermion interactions (electron, muon and two neutrinos). An example of a 4-fermion effective vertex is illustrated in Fig. 7-1.

– Chiral perturbation theory describing low energy interactions of pions.

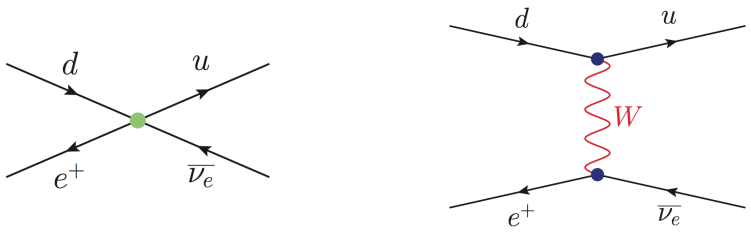


Figure 7-1: Figure showing a 4-fermion effective vertex (left) in a e^+p collision. The actual vertex with the W propagator is shown on the right.

- Heavy quark effective theory used to describe mesons with one heavy quark (c or b).
- Non-relativistic QED (Quantum electro-dynamics).

7.1.3 Linear vs non-Linear Parametrization

With the HEFT approach, two formulations are possible:

1. **'Linear Lagrangian' assumption:** In this case, the Higgs belongs to a weak doublet with $SU(2)_L \times U(1)_Y$ being linearly realized at high energies.
2. **'Non-linear Lagrangian' assumption:** The $SU(2)_L \times U(1)_Y$ is realized in a non-linear way in this parametrization, and the Higgs boson is assumed to be a singlet not belonging to a weak doublet.

Both parametrizations have been studied in [10]. The LHC Run-I results provide strong evidence that the newly discovered particle is compatible with the SM 0^+ Higgs boson, with couplings and branching ratios consistent with the SM predictions (with large errors). The collected data is consistent with the predictions of a linearly realized $SU(3) \times SU(2) \times U(1)$ local symmetry, which motivates the use of the 'linear Lagrangian' assumption. Under the latter, the deviations from the SM are small and new states are expected at energy scales far from the weak scale. In the following, the 'linear Lagrangian' assumption is applied.

7.1.4 Building HEFT

The basic ideas of a HEFT can be summarized as follows [6, 16]:

- Assuming $SU(3) \times SU(2) \times U(1)$ local symmetry is linearly realized, the SM prediction is the lowest order approximation.
- New physics is described using higher dimensional gauge-invariant operators built using SM fields (fermion fields, the scalar doublet H and $SU(3) \times SU(2) \times U(1)$ gauge fields). The higher order operators are organized in inverse powers of the new physics scale Λ. In the limit $\Lambda \to \infty$, the SM is recovered.

- New particles can be included as additional heavy fields. Yet, the EWSB mechanism is the Higgs mechanism.
- Even though HEFT comes with many free parameters, various constraints on new physics are imposed using results from Higgs searches, electroweak precision observables, gauge boson pair production, dijet production, etc [11–15].
- The resulting effective Lagrangian is then of the form:

$$\mathscr{L}_{\text{EFT}} = \mathscr{L}_{\text{SM}} + \sum_{n>4} \sum_{i=1}^{N_n} \frac{a_i^n}{\Lambda^{n-4}} \mathscr{O}_i^{(\text{d=n})} , \tag{7.1}$$

  where $\mathscr{L}_{\text{SM}}$ is the SM 4-dimensional Lagrangian, a_i are arbitrary dimensionless Wilson coefficients and Λ is the NP scale. For Higgs physics, the dominant NP contribution comes from the dimension-6 (dim-6) operators[1] at scale $\Lambda \sim$1 TeV beyond which NP effects are too small to be detected [16]. Many frameworks, including the latest HEFT effort at CERN, restrict studies to dimension-6 operators. A brief discussion of the various effective Lagrangian terms is given in section 7.1.5.
  Some major HEFT key points to consider, before discussing operators are:
  - From dimensional analysis and power counting, the size of the coefficients can be estimated.
  - The operators at a given dimension form a vector space and require thus a basis (see sec. 7.1.6).
  - Higher dimension operators have small contributions to low-energy processes at tree-level. When considering loop effects, these operators are not suppressed. However, their effect is limited to correcting lower dimensional operators [17].
- Poincaré invariance (Lorentz+translations) is assumed in addition to conservation of baryon and lepton numbers with dimension-6 operators.
- CP (charge-parity) violating operators are allowed. In fact:

$$\mathscr{L}_{\text{EFT}} = \mathscr{L}_{\text{CPC}} + \mathscr{L}_{\text{CPV}} , \tag{7.2}$$

  where CPC refers to CP-conserving terms and CPV denotes CP-violating ones.

7.1.5 Effective Lagrangian: Basics

The effective Lagrangian can be written as :

$$\mathscr{L}_{\text{eff}} = \mathscr{L}_{\text{SM}} + \frac{1}{\Lambda}\mathscr{L}^{\text{D=5}} + \frac{1}{\Lambda^2}\mathscr{L}^{\text{D=6}} + ... \tag{7.3}$$

[1] Dimension-6 operators are the least suppressed

where:

- $\mathscr{L}_{\text{SM}}$ is the SM Lagrangian with operators up to dimension 4 (renormalizable) and 19 free parameters (all measured). The predictions for the lowest order include:
  - W and Z boson mass ratio in terms of the Weinberg angle θ_W
  - The Higgs coupling to fermions is linear in their mass and its coupling to gauge bosons is proportional to their mass squared.
  - Triple and quartic gauge boson couplings are proportional to gauge couplings.

- The dimension-5 terms correspond to the Weinberg operators, which give rise to Majorana mass terms for left-handed neutrinos. The scale Λ_5 is very high and does not affect Higgs physics [16]. In addition, the only SU(2) invariant dimension-5 term violates lepton number conservation, which is one the basic HEFT assumptions considered here. It will not be included in the following calculations.

- At dimension 6, a complete set of non-redundant operators was identified. The first attempt to construct dim-6 operators was in [19] with 80 independent parameters (for one generation). Imposing lepton and baryon number conservation leads to 2499 non-redundant RG (Renormalization Group) parameters [18]. This number is further reduced if flavor symmetries are assumed [22] and CP conservation is imposed [20] (see section 7.1.6).

- Operators with dimension D $>$ 6 are neglected as current experimental precision does not allow to constrain them.

For additional details, the reader can refer to [3, 6, 10, 11, 22] for a thorough treatment of HEFT using dimension-6 operators with different approximations.

7.1.6 HEFT Bases

Various bases have been developed for HEFT studies, with the most popular ones being the following:

- **Warsaw basis:** This is a descendant of the first attempt to construct a basis ('Buchmuller-Wyler') described in [19]. Starting with 80 dimension-6 operators, this basis removed the redundant [23, 24] ones and the number of operators is reduced to 59 [22]. The redundancy seen in dimension-6 operators consisted in having operators equivalent to linear combinations of other dimension-6 operators. This is not seen for the SM operators but is a feature of dimension-6 ones.

- **HISZ ('Hagiwara') basis[2]:** This basis [25] focuses on operators related to Higgs and electroweak gauge bosons only, not including fermion operators. Even though it is not a complete basis, it is widely used for Higgs studies and can be extended to include fermion operators.

- **SILH basis:** The SILH (**S**trongly **I**nteracting **L**ight **H**iggs) basis was first introduced in [21] and the complete basis is described in [10, 11, 26, 27]. One of the main motivations behind the development of this basis is the Higgs boson discovery.

- **Higgs basis:** This basis was recently developed during 2015 in preparation for Run-II by the CERN HEFT team, in an effort to have a unified basis for the LHC Run-II studies. The basic ideas and assumptions defining this basis are the following:

  - The basis is built using a complete, non-redundant set of dimension-6 operators.
  - The effective Lagrangian is limited to dimension-6 operators. The Higgs boson is assumed to belong to a weak doublet with linear[3] SU(3)×SU(2)×U(1) local symmetry.
  - The complete basis is constructed with SU(2)-invariant terms after identifying *'independent coupling'* terms as described in [28].
  - The Z and W partial decay widths impose constraints on HEFT parameters and the independent couplings are chosen such that the inclusion of such constraints is easy.
  - The Higgs basis operators are mapped to the SU(3)×SU(2)×U(1) invariant operators before EWSB of the Warsaw basis.
  - Flavor universality is not imposed, CP violating operators are allowed and the basis assumes the conservation of lepton and baryon numbers.

  The translation from the *'Higgs basis'* to each of the 'Warsaw', 'SILH' and 'HISZ' bases is explained in Appendix A in [15] (and being added to [29]).

Other bases exist as well e.g. the basis in [20] falls between 'Warsaw' and 'SILH' bases [31]. A comparison of the 'Warsaw' and 'SILH' bases is summarized in [30].
In Madgraph5_aMC@NLO, which was used for the derivation of the results presented here, the implementation of 'Warsaw', 'SILH' and 'Higgs' bases is available in the ROSETTA framework (being finalized) [32].

[2]HISZ ≡ **H**agiwara **I**shihara **S**zalapski **Z**eppenfeld

[3] A discussion of the non-linear effective field theory for the CERN Yellow Report 4 is expected soon in September 2015

7.2 Importance of HEFT for LHC Run-II

A major emphasis during the LHC Run-II is put on precision measurements and testing/checking for all deviations from the SM Higgs couplings. That is where HEFT becomes very useful, probing NP in a model independent way. In fact, the HEFT approach will be used as the default option for Run-II, replacing thus the κ-framework [34] which was used during Run-I for spin/CP studies. The switch to HEFT for Run-II is also motivated by the fact that one can overcome the limitations of the κ-framework (e.g. proper treatment of NLO corrections, EW corrections, etc.), while having a consistent treatment of observables (even beyond the Higgs sector) within a single BSM framework [33].

7.3 Higgs Properties (Spin/CP) and Monte Carlo Tools

After the evidence showing that the scalar resonance discovered at LHC is compatible with the SM Higgs boson, it was time to study the spin/CP properties of the new particle to confirm/establish whether it is indeed a 0^+ SM Higgs boson or not. The spin-1 hypothesis was ruled out by the Landau-Yang theorem [35, 36] as the new scalar was observed the $H \to \gamma\gamma$ decay mode. The spin-2 hypothesis was eliminated as well by the measurements in the diboson decay modes, namely $H \to ZZ, WW$. In addition, the spin-2 hypothesis is extremely challenging from a theoretical point of view [34]. On one hand, the Velo-Zwanziger problem arises in the case of a naive coupling of a massive spin-2 field with a U(1) gauge field, leading to superluminal modes and other extreme features. On the other hand, using an effective description where the spin 2 particle is treated as a Kaluza Klein (KK) graviton cannot be adopted, since no analogous KK excitations of the SM gauge bosons have ever been observed. Also, the couplings of a graviton-like massive spin 2 boson to WW and ZZ are significantly smaller than those to $\gamma\gamma$, which is not compatible with the measurements extracted from the observed data.
The last step was thus to look for $0^{+/-}$ discriminating signatures to check if it is CP-even (0^+) or CP-odd (0^-) in both bosonic and fermonic (to test fermio-phobic Higgs hypotheses) modes. Assuming a mixture of CP-even and CP-odd in the data samples, one can base the spin analysis on specific kinematic variables only or try exploiting the full event information using the *matrix element method*. The latter can be used to discriminate signal/background samples or two signal samples based on a likelihood constructed at the tree level amplitude. The *matrix element method* can be applied using either of the following approaches:

- **The effective Lagrangian method:** Using HEFT as described in the previous sections, the most general Lagrangian with Lorentz and gauge invariance is chosen. This method is adopted by the MADGRAPH5 team[4] and can be ex-

[4]The 'MadWeight' tool in MADGRAPH5 can be used for the calculations of the 'anomalous couplings approach'.

tended to NLO calculations.
Results with this method include (to name a few) the latest global constraints on anomalous triple gauge couplings in effective field theory using the CERN 'Higgs basis' as shown in [15], diagrammatic computation of multi-Higgs processes at very high energies [37], NLO QCD corrections effects for various spin 0 hypotheses [38],etc. The implementation of the effective Lagrangian is described in [39–41].

- **Anomalous couplings approach:** This method chooses the most general amplitude with Lorentz and gauge invariance and is limited to LO (Leading Order) calculations. This method is implemented with 'MELA' (**M**atrix **E**lement **L**ikelihood **A**pproach) in the JHU[5] generator (see [42, 43]), and is used for Higgs off-shell measurements and constraints for example [44–46].

The results presented hereafter were derived using the effective Lagrangian approach as implemented in the Higgs characterization framework within Madgraph5_aMC@NLO. The details of the method are explained in section 7.4.

7.4 Higgs Characterization (HC) Framework

The Higgs characterization (HC) framework uses a simple effective field Lagrangian below the EW symmetry breaking scale, yet is perfectly suitable to address questions on the strength of the Higgs coupling [40]. It provides accurate simulations in the context of multi-parton tree-level (Madgraph5) and next-to-leading order (aMC@NLO) computations. It also allows to study the Higgs quantum numbers, coupling strengths and structure in a model independent way, in addition to the possiblity of going beyond the 'one-process, one-decay, one-observable' approach to the 'any-process, any-decay, any-observable' one.

Higgs spin/CP and couplings to SM particles studies using the EFT approach can be done in all channels in a complete, systematic way with good precision using the 'Higgs characterization' framework. HC is indeed a framework where the various spin hypotheses ($J^P = 0^+, 0^-, 1^+, 1^-, 2^+, 2^+$) were implemented to test the spin/CP properties of the newly discovered scalar boson at LHC. The Higgs effective Lagrangian can be easily incorporated in any matrix element (ME) generator following FeynRules, using the Universal FeynRules Output (UFO) [47] interface. And it is already implemented in Madgraph5_aMC@NLO[6]. The main advantages offered by such a framework are:

1. All Higgs production (gluon-gluon fusion 'ggF', vector-boson fusion 'VBF', boson associated 'VH' and $t\bar{t}$ associated modes) and decay modes are available in a single model, and the generation is fast.

2. The model can be extended to allow additional four-point interactions or higher-order dimension interactions.

[5] John Hopkins University
[6] See the generator description in chapter 3.

3. NLO corrections are automatically implemented with 'aMC@NLO'. The NLO accuracy on the Higgs production is provided in a straightforward manner to the user [34, 39, 40].

In the following sections, the effective Lagrangian is described with an emphasis on spin 0 hypothesis implementation. Spin-1 and spin-2 hypotheses are already ruled out by experimental results, and their implementation in the framework will not be presented here but can be found in [40]. In addition, a brief comparison with the JHU results is shown.

7.4.1 HC Effective Lagrangian

The framework implementation follows the guidelines highlighted in [48]. 0^+ and 0^- spin states are allowed to mix. The separation is done through the parameter $\cos(\alpha)$, whose value is one if the resonance is a SM scalar 0^+ and zero in case of a pseudoscalar (CP-odd). The effective Lagrangian with dimension-6 operators assumes SU(2)×U(1) linear realization and is currently limited to three-point Higgs interactions for practical reasons. The resulting Lagrangian is equivalent to:

- the one in [21, 49] (SILH basis) and [10] (without 4-point interactions terms) for CP-even states

- the one in [10] (Warsaw basis) for CP-odd states

During the Run-I Higgs spin studies, a major emphasis was on the Higgs coupling to SM particles and this is described by the following terms in the effective Lagrangian:

1. **Coupling to fermions:** This is described by the term $\mathcal{L}_0^f$ where X_0 refers to the new boson field (scalar or pseudoscalar), and $g_{Hff} = m_f/v$, g_{Aff} refer to the scalar 0^+ and pseudoscalar 0^- couplings to fermions respectively:

$$\mathcal{L}_0^f = -\sum_{f=t,b,\tau} \bar{\psi}_f \big(c_\alpha \kappa_{Hff} g_{Hff} + i s_\alpha \kappa_{Aff} g_{Aff}\, \gamma_5\big) \psi_f X_0\,, \tag{7.4}$$

$$c_\alpha \equiv \cos\alpha\,, \quad s_\alpha \equiv \sin\alpha\,, \tag{7.5}$$

$$g_{Hff} = m_f/v\,, (g_{Aff} = m_f/v)\,. \tag{7.6}$$

   The coupling of scalar Higgs to fermions is assumed for the third generation only for simplicity. The SM Higgs is recovered for $c_\alpha = 1$ and $\kappa_{Hff} = 1$. On the other hand, setting $s_\alpha = 1$ and $\kappa_{Aff} = \cot\beta$ ($\tan\beta$) gives access to the up (down) parts of a type-II CP-conserving 2HDM or SUSY.

2. **Coupling to vector bosons:** This is described by the term $\mathcal{L}_0^V$ component

parameter	reference value	description
Λ [GeV]	10^3	cutoff scale
$c_\alpha (\equiv \cos\alpha)$	1	mixing between 0^+ and 0^-
κ_i	0 , 1	dimensionless coupling parameter

Table 7.1: HC model parameters for the effective Lagrangian [40].

defined as [40]:

$$\begin{aligned}\mathcal{L}_0^V = \Big\{ & c_\alpha \kappa_{\rm SM} \big[\frac{1}{2} g_{HZZ}\, Z_\mu Z^\mu + g_{HWW}\, W_\mu^+ W^{-\mu}\big] \\ & - \frac{1}{4}\big[c_\alpha \kappa_{H\gamma\gamma} g_{H\gamma\gamma}\, A_{\mu\nu}A^{\mu\nu} + s_\alpha \kappa_{A\gamma\gamma} g_{A\gamma\gamma}\, A_{\mu\nu}\widetilde{A}^{\mu\nu}\big] \\ & - \frac{1}{2}\big[c_\alpha \kappa_{HZ\gamma} g_{HZ\gamma}\, Z_{\mu\nu}A^{\mu\nu} + s_\alpha \kappa_{AZ\gamma} g_{AZ\gamma}\, Z_{\mu\nu}\widetilde{A}^{\mu\nu}\big] \\ & - \frac{1}{4}\big[c_\alpha \kappa_{Hgg} g_{Hgg}\, G^a_{\mu\nu}G^{a,\mu\nu} + s_\alpha \kappa_{Agg} g_{Agg}\, G^a_{\mu\nu}\widetilde{G}^{a,\mu\nu}\big] \\ & - \frac{1}{4}\frac{1}{\Lambda}\big[c_\alpha \kappa_{HZZ}\, Z_{\mu\nu}Z^{\mu\nu} + s_\alpha \kappa_{AZZ}\, Z_{\mu\nu}\widetilde{Z}^{\mu\nu}\big] \\ & - \frac{1}{2}\frac{1}{\Lambda}\big[c_\alpha \kappa_{HWW}\, W^+_{\mu\nu}W^{-\mu\nu} + s_\alpha \kappa_{AWW}\, W^+_{\mu\nu}\widetilde{W}^{-\mu\nu}\big] \\ & - \frac{1}{\Lambda} c_\alpha \big[\kappa_{H\partial\gamma}\, Z_\nu \partial_\mu A^{\mu\nu} + \kappa_{H\partial Z}\, Z_\nu \partial_\mu Z^{\mu\nu} + \big(\kappa_{H\partial W}\, W^+_\nu \partial_\mu W^{-\mu\nu} + h.c.\big)\big]\Big\} X_0\,, \end{aligned} \tag{7.7}$$

The definitions of the reduced field tensors and dual tensors entering $\mathcal{L}_0^V$ are given below and the model parameters are summarized in Table 7.1.

$$V_{\mu\nu} = \partial_\mu V_\nu - \partial_\nu V_\mu \quad (V = A, Z, W^\pm)\,, \tag{7.8}$$

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s f^{abc} G^b_\mu G^c_\nu\,, \tag{7.9}$$

$$\widetilde{V}_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma} V^{\rho\sigma}\,. \tag{7.10}$$

7.4.2 Comparison with JHU Results

A very brief comparison with some JHU results is presented in this section for illustrative purposes. Plots are shown for pp $\to$ X $\to VV^* \to 4f$ events ($V = W$ or Z, f = fermion). The comparison between the JHU generator and the Madgraph5_aMC@NLO results is done for various spin hypotheses of the new boson $X(J^P)$ at $m_X = 125$ GeV for a LHC center of mass (CM) energy $\sqrt{s} = 8$ TeV. Unlike experimental settings, no kinematical cuts are applied to leptons. The generation details for the JHU plots are summarized in [50] and the Madgraph5_aMC@NLO (or MG5 for simplicity) settings are summarized in [40]. The angle definitions as

used in the MG5 plots are illustrated in Fig. 7-2 for pp $\to H \to ZZ \to 4\ell^{\pm}$ events, and those used for the JHU results are shown in Fig. 1 of [50] for gg or $q\bar{q} \to X \to V_1(q_1)V_2(q_2)$, $V_1 \to f(q_{11})\bar{f}(q_{12})$, $V_2 \to f(q_{21})\bar{f}(q_{22})$ events (f and $\bar{f}$ referring to fermions and anti-fermions respectively). Finally, the masses of the bosons V_i ($i = 1, 2$) in $pp \to X \to V_1V_2$ are denoted by m_1 and m_2 for $V = W, Z$ in the Madgraph5_aMC@NLO plots.

When comparing the results, one needs to do the following translation from the def-

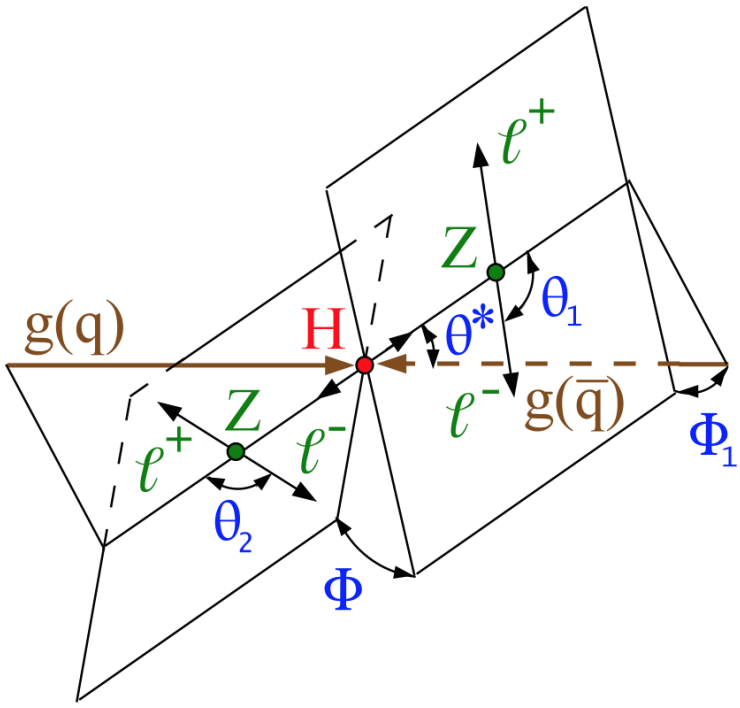


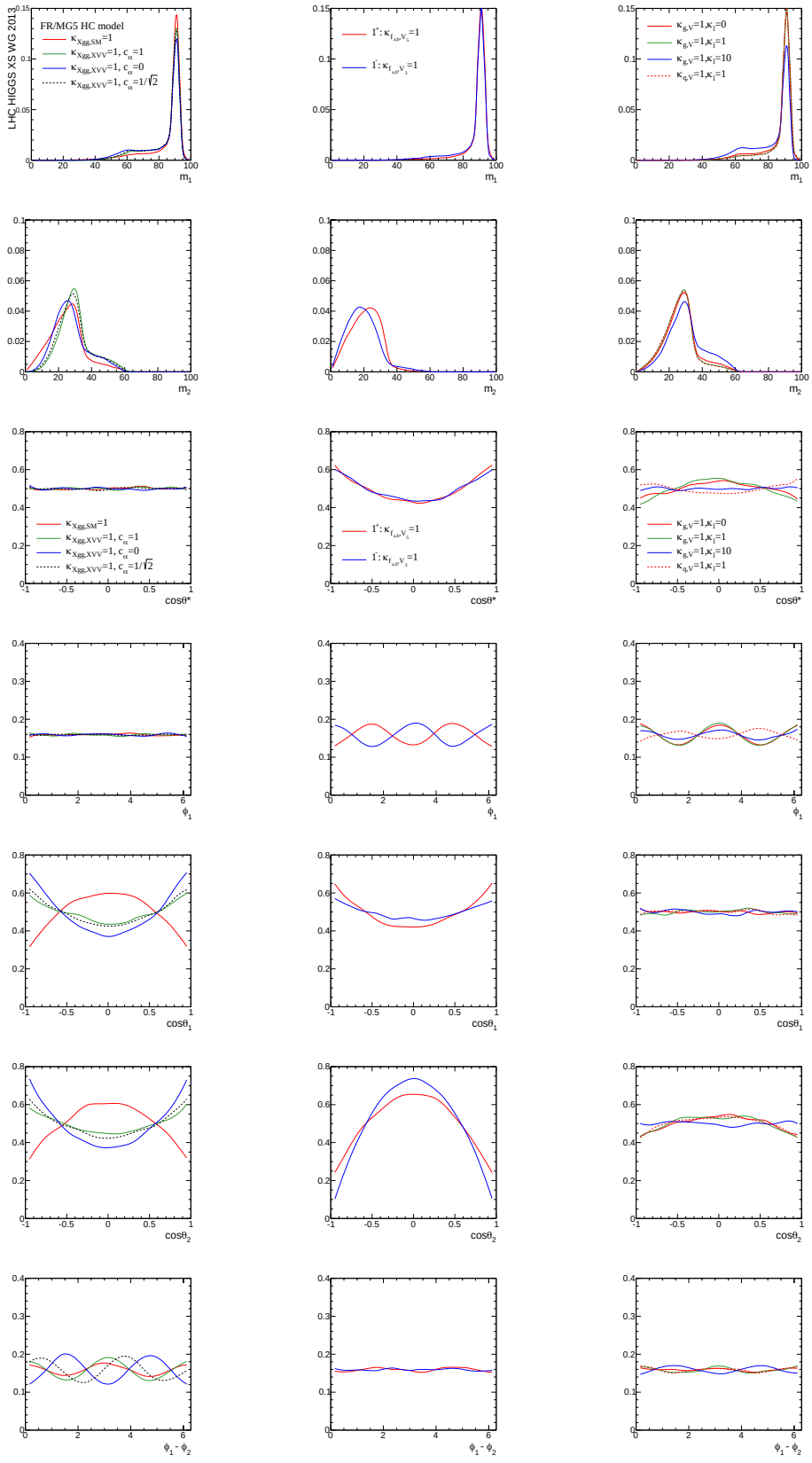
Figure 7-2: Figure showing the 5 angles fully characterizing the orientation of the decay chain in pp $\to H \to ZZ \to 4\ell^{\pm}$ events as used in the Madgraph5_aMC@NLO plots. The angles are defined in the corresponding particle rest frame. The illustrated production and decay of a H particle are for various spin hypotheses, e.g. spin 0, spin 1 [34].

initions in Madgraph5_aMC@NLO plots to the notations used for the kinematical variables in JHU results.

$$\phi_1 \to \Phi_1, \quad \phi_1 - \phi_2 \to \Phi \ , \tag{7.11}$$

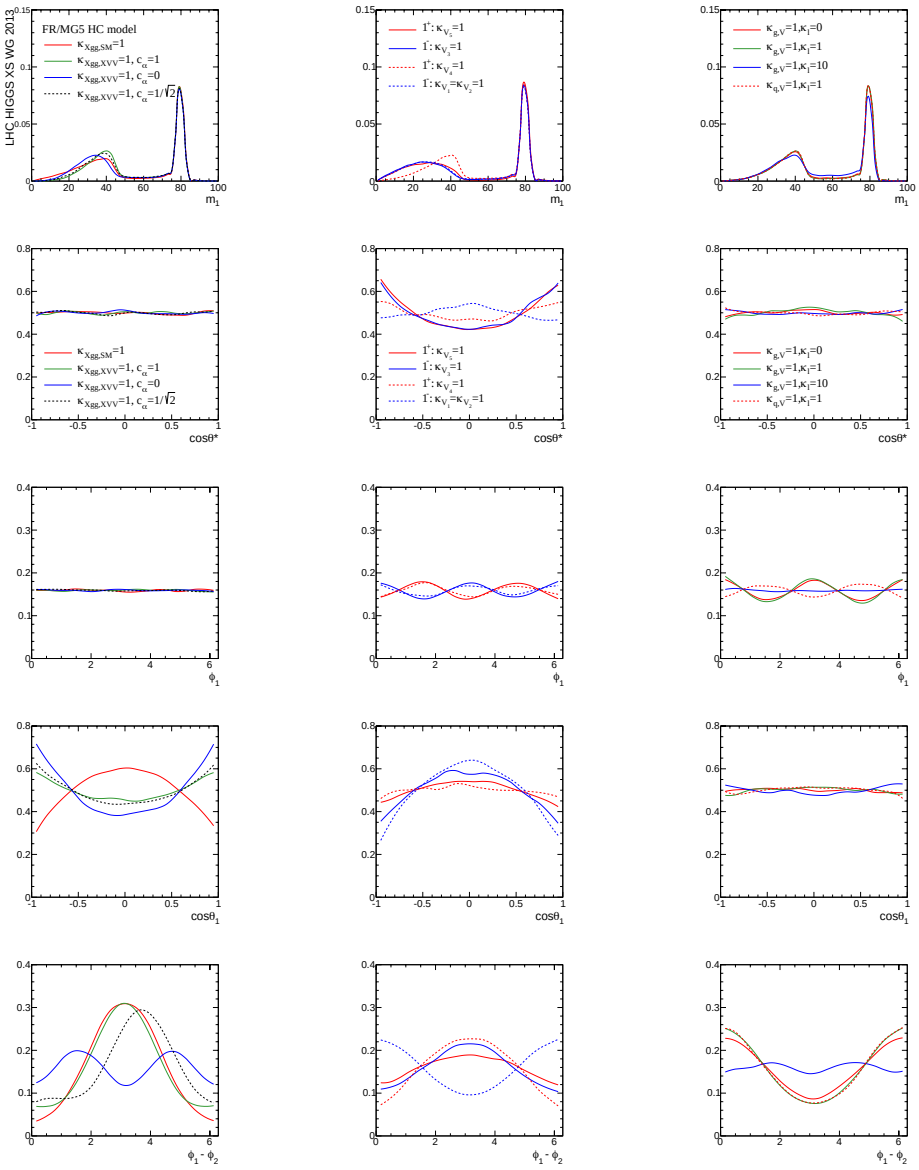
$$\text{Azimuthal angles: } [0, 2\pi] \to [-\pi, \pi] \ . \tag{7.12}$$

According to [34], the MG5 results in the ZZ and WW decay modes, illustrated in Fig. 7-3 and 7-4, agree with the JHU ones.



spin-0 spin-1 spin-2

Figure 7-3: Distributions of the X → ZZ → 4ℓ analysis using the effective Lagrangian as implemented in MG5 [34] for spin 0 (left), spin 1 (middle) and spin 2 (right) hypotheses. The Lagrangian parameter settings are shown on the plots.



spin-0 spin-1 spin-2

Figure 7-4: Mass and angular distributions for the X → WW analysis using the effective Lagrangian as implemented in MG5 [34] for spin 0 (left), spin 1 (middle) and spin 2 (right) hypotheses.

7.5 Application in Higgs Analyses

The HC framework can be used for Higgs spin/CP studies for various Higgs production and decay modes in addition to BSM search studies. The distributions of standard kinematic and mass variables in addition to some results for spin/CP expectations in the bosonic modes ($ZZ, WW, \gamma\gamma$) have been presented in [40] and validated as explained in the previous section. Some Run-I BSM studies in CMS for the $H \to \tau^+\tau^-$ decay mode have used the HC framework as well. In addition, the framework has been recently used to estimate the latest global constraints on anomalous triple gauge couplings in HEFT using the CERN 'Higgs basis' as shown in [15].
For complete testing of the SM, the coupling of the discovered Higgs particle to fermions needs to be checked as well. The dominant Higgs fermionic decay mode at the LHC Run-I is the $H \to \tau^+\tau^-$, on which the work presented in the next sections focuses. The tau model in Madgraph5_aMC@NLO is used for its advantages over the TAUOLA version used for LHC Run-I studies.

7.6 Tau Model in Madgraph5_aMC@NLO

The tau model 'TAUDECAY' [51] is a library allowing the simulation of polarized taus, taking into account the spin correlation amongst the tau decay products. It is constructed using FEYNRULES packages within Madgraph5_aMC@NLO and has the effective vertices of hadronic tau decays built-in. Compared to TAUDECAY, TAUOLA [52–54], which is the standard interface used by HEP[7] experiments to generate taus, suffers from the following limitations (for LHC-Run I):

1. The produced tau leptons need to be on-shell.
2. Only standard processes such as $Z/\gamma^* \to \tau^+\tau^-$ and H/A[8] $\to \tau^+\tau^-$ are feasible.
3. Transverse spin effects require particular treatment.

To overcome these problems automatically avoided with 'TAUDECAY', the TauSpinner package [55, 56] was used with the TAUOLA outcome during LHC Run-I studies. TauSpinner is of particular interest for data driven analysis using $Z/\gamma^* \to \tau^+\tau^-$ (especially embedded $Z \to \tau\tau$ samples used in the $H \to \tau^+\tau^-$ search analysis), and $W^\pm \to \tau^\pm\nu$ samples. In addition, with the tau spinner, various helicity states can be obtained and new model studies become feasible if tau leptons are generated with TAUOLA.

An improved version, namely TAUOLA++, is developed but was not used during LHC Run-I. The reader can refer to [57] for a review of the performance of TAUOLA with a discussion of the τ-lepton generation status.
Additional details on the 'TAUDECAY' library and its performance in comparison with TAUOLA are given in [51].

[7] HEP≡ high energy physics
[8] H/A refer to the scalar 0^+ (H) and pseudoscalar 0^- (A) spin hypotheses.

7.7 Testing and Validation of HC+Tau Model in Madgraph5_aMC@NLO

After working on the search for the Higgs boson experimentally during the LHC Run-I, it was time to prepare for spin/CP studies especially for the LHC Run-II. As a result, a short-term internship was done with the Madgraph5_aMC@NLO team at the University of Louvain (UC Louvain), with a major emphasis on testing and validating the tau model with the HC framework within MG5. The major steps consisted in the following:

- After reproducing plots in [40], an example of which is given in Fig. 7-3 and 7-4, and getting familiar with the HC framework and Madgraph5_aMC@NLO generator, I developed a merging tool that allows full Higgs production and decay. This is especially useful for NLO Higgs production and/or decays with an off-shell particle (e.g. WW^* and ZZ^*). The tool provides automated LHE[9] files with necessary corrections within an optimized CPU time, and can handle multiple sub-decay chains automatically. The tool was validated in different Higgs decay modes (e.g. ZZ and WW channels). The MG5 LHE parser was updated as well to handle multiple sub-decay chains and/or events with off-shell particles.

- While testing and validating HC and tau models within MG5, I reported several encountered problems: the tau model 'FRBlock' parameters needed to be redefined to allow proper model merging, form factors for the a1 and ρ hadronic τ-decay modes were missing and coupling parameters for the 0^- Higgs when decayed in the $\tau\tau$ mode were not implemented correctly (e.g. κ_{SM}) leading to zero cross section. This was taken care of by K.Mawatari[10] and O.Mattlaer[11].

- Since MadSpin uses the decay chain approximation [58], and given the tau width Γ_τ, the efficiency of the generation of unweighted events becomes a serious issue. That's why an effective width is needed. Otherwise, narrow width approximation needs to be implemented at the matrix element level. So an effective tau width of 1 MeV, needed to decay taus in MadSpin using decay chain approximation, was correctly applied. Details on the narrow width approximation and the treatment of resonances beyond the Breit-Wigner approximation can be found in [59].

  As a result, the HC+tau model was validated in 3 different ways: with full ME production, using implemented merging tool and with MadSpin decays as shown in Fig. 7-5 for hadronic tau decays for A/H/Z$\rightarrow \tau\tau$ events (H refers to the scalar

[9] LHE stands for Les Houches Eventfile. This is the standard file format to store the information about the generated events (initial, intermediate and final states particles, etc.) used by many Monte Carlo generators.

[10] One of the main tau model authors in MG5.

[11] MG5 expert.

0^+ spin hypothesis and A for the pseudoscalar 0^- spin hypothesis). In these figures, two discriminating angular variables $\Phi = \phi_1 + \phi_2$ and $\Delta\phi = \phi_1 - \phi_2$ are plotted. Φ and $\Delta\phi$ are defined as the sum and difference respectively of azimuthal angles of produced pions in the tau decay planes. The $\pi - \nu$ tau decay mode is considered here for simplicity. The exact variable definition is illustrated in Fig. 7-6, where :

(a) With the process of interest being pp $\to$ X $\to \tau^+\tau^-$, the 4-momenta p_1, p_2, k_1 and k_2 are defined in the X rest frame as follows:

$$\begin{aligned} p_1 &= \tfrac{\sqrt{\hat{s}}}{2}(1, -\sin\Theta, 0, \cos\Theta), \\ p_2 &= \tfrac{\sqrt{\hat{s}}}{2}(1, \sin\Theta, 0, -\cos\Theta), \\ q_1 &= \tfrac{\sqrt{\hat{s}}}{2}\big(1 + \tfrac{q_1^2 - q_2^2}{\hat{s}}, 0, 0, \beta\big), \\ q_2 &= \tfrac{\sqrt{\hat{s}}}{2}\big(1 + \tfrac{q_2^2 - q_1^2}{\hat{s}}, 0, 0, -\beta\big) \end{aligned} \tag{7.13}$$

Θ is the scattering angle and $\sqrt{\hat{s}}$ is the CM energy.

(b) The z-axis is defined along the direction of the τ^- 3-momentum vector.

(c) The y-axis is defined along the direction of $\mathbf{p_1} \times \mathbf{q_1}$.

(d) The same (x,y,z) coordinate system is used in all 3 frames: X rest frame, τ^- and τ^+ frames.

(e) $Z \to \tau\tau$ processes are background events for H/A$\to \tau\tau$ ones.

The plots in Fig. 7-5 show that Φ has a strong signal/background discriminating power. In addition, the azimuthal angular difference $\Delta\phi$ can be used for $0^\pm$ hypothesis testing. The plots successfully match the analytic functions and the spin/CP powerful discriminating variable distributions in [51]. The $Z \to \tau\tau$ plots are not shown for the merging method for the following reason : since Z is a spin-1 vector boson, the Z production and decay processes are not factorizable. The full matrix element is spin-dependent. More sophisticated techniques than the merging method, such as the ones in MadSpin, are required in this case. However, for spin 0 bosons (A and H), the full matrix element is factorizable as

$$\sigma(\mathrm{A/H})_{\mathrm{full}} = \sigma_{\mathrm{prod}} \times \mathrm{BR}_{\mathrm{A/H\ decay\ mode}} \quad , \tag{7.14}$$

where BR refers to the corresponding branching ratio.

In addition to the results in the $\pi\nu$ hadronic decay mode, tau results in the a_1 and ρ mode were also explored. The definitions of the major hadronic tau

decay modes are given in eq. (7.15).

$$\pi \text{ mode :} \qquad \tau^- \to \nu_\tau \pi^-, \tag{7.15a}$$
$$\rho \text{ mode :} \qquad \tau^- \to \nu_\tau \rho^- \to \nu_\tau \pi^- \pi^0, \tag{7.15b}$$
$$a_1 \text{ mode :} \qquad \tau^- \to \nu_\tau a_1^- \to \nu_\tau \pi^0 \rho^- \to \nu_\tau \pi^0 \pi^0 \pi^-, \tag{7.15c}$$
$$\tau^- \to \nu_\tau a_1^- \to \nu_\tau \pi^- \rho^0 \to \nu_\tau \pi^- \pi^- \pi^+, \tag{7.15d}$$

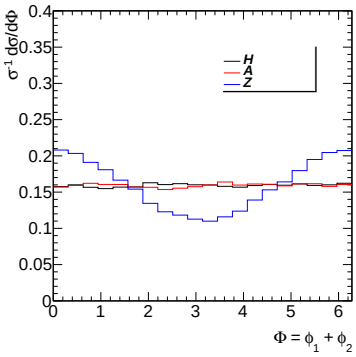
For illustrative purposes, the invariant mass distributions in the a_1 and ρ modes are shown in Fig. 7-7.

– My last contribution was participating in the implementation (automatizing) of a new version of MadSpin to allow $1 \to 3$ body decay for different spin hypotheses. The tau spin/CP plots were also successfully produced with the new MadSpin version (work in progress) as well as shown in Fig. 7-8. The curves are less smooth than the ones in Fig. 7-5 because they are done with a significantly smaller number of events (only 10 000 compared to 100 000 in Fig. 7-5).

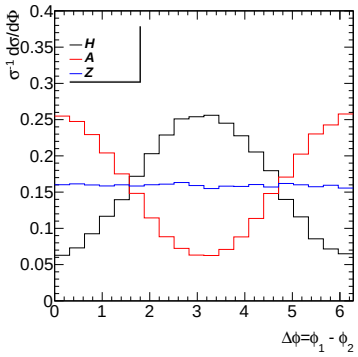
7.8 Conclusions and Prospectives

In this chapter, a brief discussion of the Higgs effective field theory (HEFT) and its importance for Run-II was presented. Since no BSM features were discovered during the LHC Run-I, new physics can be probed through Higgs coupling measurements instead of using a particular BSM theory. HEFT then is the most suitable option to explore all possibilities in a model independent way, with options including SM and beyond SM theories. In this case, the effects of heavy particles are encoded into the contact interaction of low energy particles, keeping only the relevant degrees of freedom. This effective parametrization of the SM 'HEFT' allows a systematic study of EW precision tests and triple-gauge couplings, in addition to Higgs couplings in a single, unified and efficient framework. HEFT can be applied in Monte Carlo simulations with NLO precision through the robust Higgs characterization (HC) framework implemented in Madgraph5_aMC@NLO.

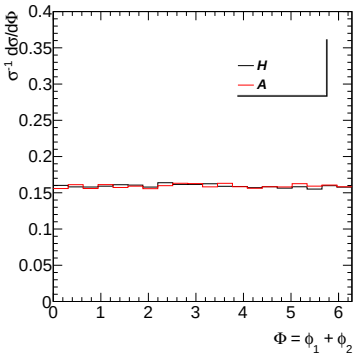
During the LHC Run-II at CERN, the emphasis is on HEFT which will replace the κ-framework in spin/CP studies. Since the Higgs$\to \tau\tau$ cross section is significantly increased, having a valid working tau model, which can be successfully combined with HC, becomes a must for precision measurements and spin/CP studies. This is now possible and in a user-friendly way after the work done during the internship, most of which focused on testing and validating HC+tau models with MG5 and correct model merging. And the tau model presented is free from all the limitations of the TAUOLA version used for the LHC Run-I.



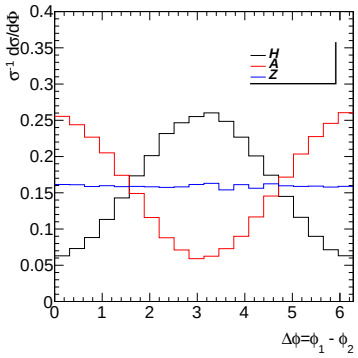
(a) Φ(ME)



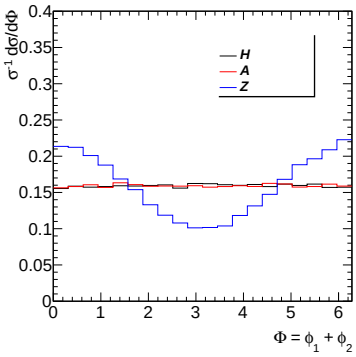
(b) $\Delta\phi$(ME)



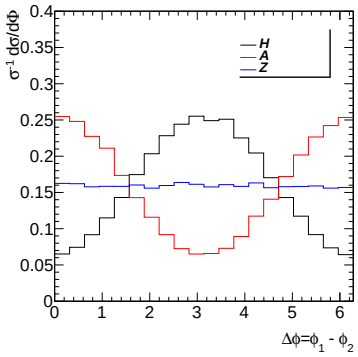
(c) Φ(Merged)



(d) $\Delta\phi$(Merged)



(e) Φ(Madspin)



(f) $\Delta\phi$(Madspin)

Figure 7-5: Angular distributions $\Phi = \phi_1 + \phi_2$ and $\Delta\phi = \phi_1 - \phi_2$ for events produced directly at the matrix element (ME) level (top), obtained after merging X(J^P) production and decay MG5 output (middle) and decayed with MadSpin (bottom). Plots are obtained with 100k events generated in Madgraph5_aMC@NLO.

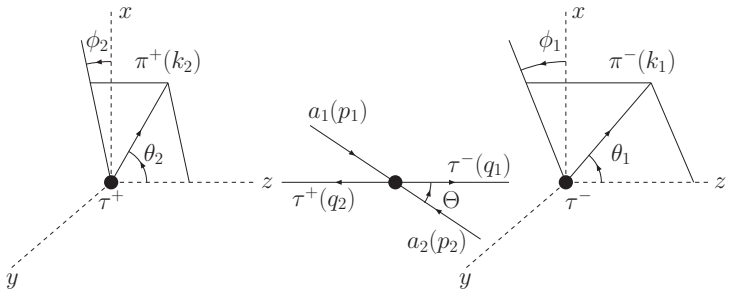


Figure 7-6: The coordinate system in the tau decay frame [51].

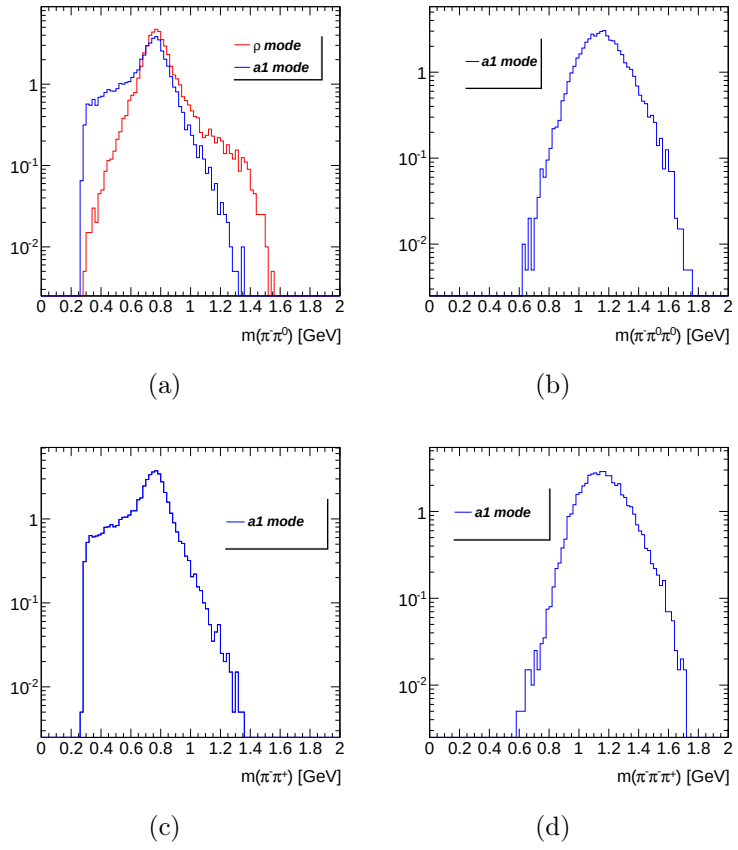


Figure 7-7: Invariant mass distributions with 10 000 events (no cuts applied) for the a_1 and ρ tau decay modes in the 1-prong (a and b) and 3-prong (d) topologies. The pion pair $\pi^-\pi^0$ and $\pi^-\pi^+$ invariant mass distributions for the a_1 1-prong and 3-prong topologies are shown in (a) and (c) respectively.

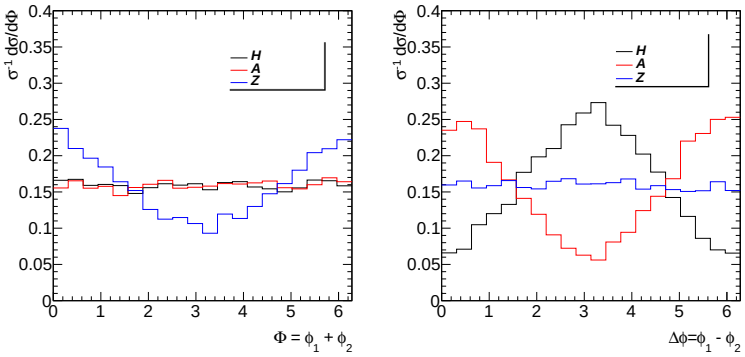


Figure 7-8: Angular distributions $\Phi = \phi_1 + \phi_2$ and $\Delta\phi = \phi_1 - \phi_2$ obtained with new MadSpin to decay tau leptons.

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*, Phys.Lett. **B716** (2012) 1-29, arXiv:1207.7214 [hep-ex]

[2] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys.Lett. **B716** (2012) 30-61, arXiv:1207.7235 [hep-ex]

[3] C. Chiang et al., *Standard Model Effective Field Theory: Integrating out a Generic Scalar*, arXiv:1505.06334 [hep-ph] (2015) IPMU15-0073

[4] A. Azatov et al., *Effective field theory analysis of double Higgs boson production via gluon fusion*, Phys.Rev. **D92** (2015) 3, 035001 arXiv:1502.00539 [hep-ph]

[5] D. Kaplan, *Five lectures on effective field theory*, nucl-th/0510023 (2005)

[6] A. Falkowski, *Effective field theory approach to LHC Higgs data*, arXiv:1505.00046 (2015)

[7] J. Polchinski, *Effective field theory and the Fermi surface*, hep-th/9210046 (1999)

[8] A. V. Manohar, *Effective field theories*, hep-ph/9606222 (1996)

[9] I. Z. Rothstein, *TASI lectures on effective field theories*, hep-ph/0308266 (2004)

[10] R. Contino et al., *Effective Lagrangian for a light Higgs-like scalar*, JHEP 1307 (2013) **035** arXiv:1303.3876 [hep-ph]

[11] A. Pomarol and F. Riva, *Towards the Ultimate SM Fit to Close in on Higgs Physics*, JHEP 1401 (2014) **151**, arXiv:1308.2803 [hep-ph]

[12] J. Ellis et al., *Complete Higgs Sector Constraints on Dimension-6 Operators*, JHEP 1407 (2014) **036**, arXiv:1404.3667

[13] M. Trott, *On the consistent use of Constructed Observables*, JHEP 1502 (2015) **046**, arXiv:1409.7605 [hep-ph]

[14] A. Falkowski and F. Riva, *Model-independent precision constraints on dimension-6 operators*, JHEP 1502 (2015) **039**, arXiv:1411.0669 [hep-ph]

[15] A. Falkowski et al., *Global constraints on anomalous triple gauge couplings in effective field theory approach*, (2015) arXiv:1508.00581 [hep-ph]

[16] A. Manohar, Higgs Effective Field Theories, HEFT 2013 Workshop (December 2013)

[17] B. Gripaios, *Lectures on Physics Beyond the Standard Model*, (2015) arXiv:1503.02636

[18] R. Alonso et al., *Renormalization Group Evolution of the Standard Model Dimension Six Operators III: Gauge Coupling Dependence and Phenomenology*, JHEP 1404 (2014) **159**, arXiv:1312.2014 [hep-ph]

[19] W. Buchmuller and D. Wyler, *Effective Lagrangian analysis of new interactions and flavour conservation*, Nucl. Phys. **B268** (1986) 621-653

[20] J. Elias-Miró et al., *Scaling and tuning of EW and Higgs observables*, JHEP 1405 (2014) **019**, arXiv:1312.2928 [hep-ph], SISSA-57-2013-FISI

[21] G. F. Giudice et al., *The Strongly-Interacting Light Higgs*, JHEP 0706 (2007) **045** hep-ph/0703164

[22] B. Grzadkowski et al., *Dimension-Six Terms in the Standard Model Lagrangian*, JHEP 1010 (2010) **085**, arXiv:1008.4884

[23] B. Grzadkowski et al., *Optimal-observable analysis of possible new physics using the b quark in $\gamma\gamma \to t$ anti-$t \to bX$*, Phys.Lett. **B593** (2004) 189-197, hep-ph/0403174

[24] J. A. Aguilar-Saavedra, *A Minimal set of top anomalous couplings -, [hep-ph]*, J.A. Nucl.Phys. **B812** (2009) 181-204, arXiv:0811.3842

[25] K. Hagiwara et al., *Low-energy effects of new interactions in the electroweak boson sector*, Phys.Rev. **D48** (1993) 2182-2203

[26] E. Massó et al., *Limits on anomalous couplings of the Higgs boson to electroweak gauge bosons from LEP and the LHC*, Phys.Rev. **D87** (2013) 3, 033001 arXiv:1211.1320 [hep-ph]

[27] E. Miro et al., *Renormalization of dimension-six operators relevant for the Higgs decays $h \to \gamma\gamma, \gamma Z$*, JHEP 1308 (2013) **033**, arXiv:1302.5661 [hep-ph]

[28] R. Gupta et al., *BSM Primary Effects*, Phys.Rev. **D91** (2015) 3, 035001 arXiv:1405.0181 BSM Primary Effects - Gupta, Rick S. et al. Phys.Rev. D91 (2015) 3, 035001 arXiv:1405.0181 [hep-ph]

[29] LHC Higgs Cross Section Working Group 2 Collaboration, *Higgs Basis: Proposal for an EFT basis choice for LHC HXSWG*, LHCHXSWG-INT-2015-001(2015) cds.cern.ch/record/2001958

[30] R. Contino, *HEFT Basis Overview*, Higgs (N)NLO MC and Tools Workshop for LHC RUN-2, December 17-19, 2014

[31] S. Willenbrock et al., *Effective Field Theory Beyond the Standard Model*, Ann. Rev. Nucl. Part. Sci. **64** (2014) 83-100, arXiv:1401.0470 [hep-ph]

[32] B. Fuks and S. Sekmen, *Summary of the activities of the Tools & MC working group*, 9$^{\text{th}}$ Les Houches Workshop on Physics at TeV Colliders (June 2015)

[33] M. Duehrssen, *Higgs property measurements for Run 2: kappa's and beyond*, LHC Higgs XS WG workshop, June 2014

[34] LHC Higgs Cross Section Working Group, S. Heinemeier, C. Mariotti, G. Passarino, R. Tanaka (Eds.), *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, CERN-2013-004 (CERN, Geneva, 2013), DOI: 10.5170/CERN-2013-004

[35] L. Landau, *On the angular momentum of a two-photon system*, Dokl. Akad. Nauk Ser. Fiz. 60 (1948) 207-209.

[36] C. N. Yang, *Selection Rules for the Dematerialization of a Particle Into Two Photons*, Phys. Rev. 77 (1950) 242-245.

[37] V. Khoze, *Diagrammatic computation of multi-Higgs processes at very high energies: Scaling $F_{holy\ grail}$ with MadGraph*, Phys.Rev. D**92** (2015) 1, 014021 arXiv:1504.05023 [hep-ph]

[38] F. Maltoni et al., *Higgs characterisation via vector-boson fusion and associated production: NLO and parton-shower effects*, Eur.Phys.J. C**74** (2014) 1, 2710 arXiv:1311.1829 [hep-ph]

[39] A. Alloul, *Phenomenology of the Higgs Effective Lagrangian via FEYNRULES*, JHEP 1404 (2014) 110 arXiv:1310.5150 [hep-ph]

[40] P. Artoisenet, et al., *A framework for Higgs characterisation*, JHEP 1311 (2013) 043 arXiv:1306.6464 [hep-ph]

[41] P. de Aquino et al., *Characterising a Higgs-like resonance at the LHC* , arXiv:1307.5607 [hep-ph]

[42] Y. Gao et al., *Spin determination of single-produced resonances at hadron colliders*, Phys. Rev. D**81** (2010) 075022, arXiv:1001.3396 [hep-ph]

[43] P. Avery et al., *Precision Studies of the Higgs Golden Channel $H \to ZZ \to 4\ell$ with* MEKD, Phys.Rev. D**87** (2013) 5, 055006 arXiv:1210.0896 [hep-ph].

[44] A. Gristan, *How Wide is the Higgs Boson: off-shell constraints from CMS*, LBNL Research Progress Meeting, May 2014

[45] M. Schulze, *Higgs off-shell measurements with the JHUGen/MELA +MCFM framework*, Higgs off-shell coupling measurements and related constraints on the Higgs width CERN meeting, June 2015

[46] CMS Collaboration (Khachatryan, Vardan et al.) , *Limits on the Higgs boson lifetime and width from its decay to four charged leptons*, arXiv:1507.06656 [hep-ex]

[47] C. Degrande et al., *UFO-The Universal FeynRules Output*, Comput.Phys.Commun. 1**83** (2012) 1201-1214, [arXiv:1108.2040]

[48] N. D. Christensen et al., *A Comprehensive approach to new physics simulations*, Eur.Phys.J. C71 (2011) 1541, [arXiv:0906.2474]

[49] J.R. et al. , *First Glimpses at Higgs' face* JHEP 1212 (2012) **045**, arXiv:1207.1717 [hep-ph]

[50] S. Bolognesi et al., *On the spin and parity of a single-produced resonance at the LHC*, Phys. Rev. D **86** (2012) 095031, arXiv:1208.4018 [hep-ph].

[51] K. Hagiwara et al., *TauDecay: a library to simulate polarized tau decays via FeynRules and MadGraph5*, Eur.Phys.J. C**73** (2013) 2489, arXiv:1212.6247 [hep-ph]

[52] S. Jadach et al., *A library of monte carlo programs to simulate decays of polarized tau leptons*, Comput. Phys. Commun. 64 (1990) 275

[53] M. Jezabek et al., *The tau decay library tauola, update with exact o(alpha) qed corrections in tau→mu. (e) neutrino anti-neutrino decay modes*, Comput. Phys. Commun. 70 (1992) 69

[54] S. Jadach et al., *The tau decay library tauola: Version 2.4*, Comput. Phys. Commun. 76 (1993) 361

[55] Z. Czyczula et al., *TauSpinner Program for Studies on Spin Effect in tau Production at the LHC* , Eur.Phys.J. C**72** (2012) 1988, arXiv:1201.0117 [hep-ph]

[56] A. Kaczmarska et al., *Application of TauSpinner for Studies on τ-Lepton Polarization and Spin Correlations in Z, W and H Decays at the LHC*, Acta Phys.Polon. B45 (2014) 10, 1921-1946, arXiv:1402.2068 [hep-ph], IFJPAN-IV-2013-19

[57] Z. Was, *The τ leptons theory and experimental data: Monte Carlo, fits, software and systematic errors*, Nucl.Part.Phys.Proc. **260** (2015) 47-51, arXiv:1412.2937 [hep-ph], IFJ-PAN-IV-2014-18

[58] P. Artoisenet et al., *Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations*, JHEP 1303 (2013) **015**, arXiv:1212.3460 [hep-ph]

[59] G. Cacciapaglia et al., *Nearby resonances beyond the Breit-Wigner approximation*, Phys.Lett. B**682** (2009) 43-49, arXiv:0906.3417 [hep-ph]

Conclusions and Prospectives

The work presented in this thesis focuses on the experimental search for the Higgs boson and the physics beyond the SM (BSM) using the Higgs effective field theory. The first four chapters introduce the theory behind the Standard model (SM) and give a detailed description of the experimental tools and techniques needed for analysis (ATLAS detector, Monte Carlo simulation and physics modeling, event and physics object reconstruction). In the last three chapters, the author's work (done within 2 years) on Higgs physics both experimentally (in $H \to WW^*$ and $H \to \tau^+\tau^-$ channels) and phenomenologically/theoretically (Higgs effective field theory) is discussed, in addition to studies of a track-based definition of missing transverse energy (MET).

In chapter 5, a track-based, pile-up robust method to estimate the missing transverse energy, namely track MET, was presented. Besides being a complement to the existing Run-I calorimeter-based measurement of $E_{\mathrm{T}}^{\mathrm{miss}}$, it has little pile-up dependence, which makes it very important Run-II studies where the pile-up conditions will become even more challenging than the 2012 Run of LHC. Soft term systematics were explored in detail and the corresponding results for various track MET definitions were presented for both 2011 and 2012 data sets. Finally, a brief overview of studies done in $H \to WW^*$ and $H \to \tau^+\tau^-$ channels, where the final state is characterized by a large MET value, was given. Improving the MET value and resolution could lead to better mass measurement and resolution in Higgs studies.

In chapter 6, the ATLAS $H \to \tau^+\tau^-$ multi-variate analysis details are presented, showing results for both 7 and 8 TeV LHC Run-I data sets corresponding to an integrated luminosity of 4.5 fb^{-1} and 20.3 fb^{-1} respectively. The various background estimation methods are explained, with an emphasis on the particularities of those used for the dominant $Z \to \tau\tau$ and the 'fake tau' background processes in the lep-had (ℓh) channel. At 8 TeV, the 'fake tau' ℓh background accounts for 60% of the total background yield in the VBF category (most sensitive one) and for 44% of the boosted background events. The corresponding numbers for the 7 TeV analysis are 55% and 38% for VBF and boosted categories respectively. Even in the last two bins of the event BDT score distribution, where the excess corresponding to signal events is observed, 'fake tau' events are not negligible, accounting for :

– 45% of total background events in the second to last bin for the VBF(most sensitive) category. This number goes up to 48% in the last bin.

– 35.7% of total background events in the second to last bin for the boosted category. In the last bin, the corresponding number is 27.2%.

Potential improvements with track-based missing transverse energy in view of Run-II were also presented.

As the center-of-mass energy at the LHC Run-II increases from 8 to 13 TeV, and the delivered luminosity is multiplied by a factor 5, $H \to \tau^+\tau^-$ events can be observed with at least 5 σ (to confirm discovery in the fermionic modes). In addition, Higgs couplings to fermions needs to be well studied to probe new physics beyond the SM. An effective parametrization of the SM 'HEFT', allowing Higgs coupling measurements in a unified and efficient framework, is presented in the next chapter.

Since no BSM features were discovered during the LHC Run-I, new physics can be probed through Higgs coupling measurements instead of using a particular BSM theory. In order to explore all possibilities in a model independent way, with options including SM and beyond SM theories, Higgs effective field theory (HEFT) is the most suitable option as explained in chapter 7. BSM features are included in HEFT only when new states appear at scales not directly accessible at the LHC. Effects of heavy particles are encoded into the contact interaction of low energy particles, keeping only the relevant degrees of freedom. This effective parametrization of the SM 'HEFT' allows a systematic study of electroweak precision tests and triple-gauge couplings, in addition to Higgs couplings in a single framework. HEFT is now the option adopted by CERN groups for Run-II studies, and is already well implemented in Madgraph5_aMC@NLO.

Therefore, I did a 4-month MCnet ITN internship with the Madgraph5_aMC@NLO team at the Louvain University. My personal contribution consisting in testing and validating the tau model (presented in sec. 7.6) with the Higgs characterization (HC) framework is discussed in sec. 7.7.
During the LHC Run-II at CERN, the emphasis is on electroweak precision measurements and HEFT which will replace the κ-framework in spin/CP studies. Since the Higgs$\to \tau\tau$ cross section is significantly increased, having a valid working tau model, which can be successfully combined with HC, becomes a must for precision measurements and spin/CP studies. This is now possible and in a user-friendly way after the work done during the internship. And the tau model presented is free from all the limitations of the TAUOLA version used for the LHC Run-I.

Appendix A

Optimized $H \to W^+W^- \to \ell\nu\ell\nu$ analysis

A.1 Optimized $H \to WW^*$ analysis with track MET

Another important improvement obtained with track MET is seen when it is applied to the Higgs $H \to W^+W^- \to \ell\nu\ell\nu$ analysis. The optimized analysis improves the performance of the standard ATLAS analysis at the time, enhancing the extracted signal yield in 2012 analysis, while finding also a small signal in 2011 analysis compared to no signal in the official ATLAS analysis. For both 2011 and 2012 analyses, the statistics is significantly increased as well, with many Higgs candidates at the end of the selection cuts.
Both the official ATLAS analysis and the optimized one are cut-based blinded analyses. The main idea behind the improvement is having a well defined phase space volume, enhancing the Higgs signal and suppressing the contribution from various background events.

A.1.0.1 Event experimental signature and processes involved

The event experimental signature is dictated by both the W boson decay and the Higgs production mode. In the full leptonic $H \to W^+W^- \to \ell\nu\ell\nu$ mode, two well reconstructed leptons (electrons and/or muons) are required in addition to large MET[1] value. On the other hand, the Higgs production mechanism plays an important role in the number of jets in the final state, their η separation and in the event kinematics.

- **Signal processes:** have a major contribution from gluon-gluon fusion (ggF) events in the 0-jet and 1-jet bins, while the dominant contribution in the category with at least 2 jets comes from Higgs events produced through vector boson fusion (VBF). VBF events are characterized by 2 forward jets well separated in η, belonging to different hemispheres $\eta_{j1} \times \eta_{j2}$ = -1 and with high di-jet invariant mass ($m_{\mathrm{jj}} > 500$ GeV).

[1] missing transverse energy

- **Background processes:** have a major contribution from the W^+W^- SM production. Other background processes include diboson WZ/ZZ/Wγ , top (single top and $t\bar{t}$), Z+jets and W+jets events. All background types are estimated from MC simulation samples except the W+jets, which are calculated using a data-driven method.

Additional details about signal and background processes can be found in [1].

A.1.0.2 Event categorization and analysis strategy

Based on the W boson decay mode (leptonic $W \to \ell\nu$ or hadronic $W \to q_1\overline{q_2}$), a fully leptonic channel ($\ell\nu\ell\nu$) and a semi-leptonic channel ($\ell\nu qq$) are defined. The third possible option with at least 4 jets in the final state and no leptons is not studied. Moreover, the $\ell\nu\ell\nu$ analysis has 3 categories based on the jet multiplicity: 0 jet bin, 1 jet bin and $\geq$ 2 jet bin. The signal in first two categories is dominated by ggF (gluon gluon fusion) events, whereas the latter has major signal contribution from VBF (vector boson fusion). The study presented here focuses on the $\ell\nu\ell\nu$ analysis for both 2011 and 2012 data sets, in particular on 0- and 1-jet bin categories, given that the VBF analysis is completely orthogonal with very limited statistics: it has only 4 events after all cuts from which Higgs candidates should be found if any.

Before classifying events into the corresponding category (0 jet, 1 jet or VBF like), a preselection stage to select relevant candidates for the $\ell\nu\ell\nu$ analysis is needed. The ATLAS cuts as they were used for the $\ell\nu\ell\nu$ Moriond 2013 analysis are summarized in Table A.1. The $\ell\nu\ell\nu$ analysis events are labelled according to the final state leptons as either SF (same flavor), e.g. ee or $\mu\mu$ events, or OF(opposite flavor) events in the case of $e\mu\nu\nu$ final states.

A.1.0.3 Variables description

The basic definitions of powerful discriminating variables used in the optimized analysis are presented in this section. Additional details, including variables not used in this analysis, blinding criteria, and brief background description can be found in [1].

- $p_T^{\ell\ell}$: dilepton transverse momentum.

- $m_{\ell\ell}$: dilepton invariant mass. Requiring $m_{\ell\ell}$>10(12) GeV in OF(SF) events removes low mass resonances, while $m_{\ell\ell} < 50$ GeV enhances the Higgs signal phase space.

- Track MET: this analysis was performed using the cluster corrected version of $p_{\mathrm{T}}^{\mathrm{miss}}$ as defined earlier in chapter 5. Track MET-Cl-j was not implemented at the time when this analysis was performed.

- $\Delta\phi_{\ell\ell}$: dilepton azimuthal angular separation.

- MET^{rel}:

$$\begin{aligned} \text{MET}^{rel} &= \text{MET} \times \sin(\Delta\Phi) && \text{if } \Delta\Phi < \pi/2 \ , \\ \text{MET}^{rel} &= \text{MET} && \text{otherwise} \ , \\ \Delta\Phi &= min[\Delta\Phi(\text{MET}, \text{lepton}),\ \Delta\Phi(\text{MET},\ \text{jet})] \ . \end{aligned} \tag{A.1}$$

  Two missing transverse energy (MET) definitions are used: MET STVF and Track MET-Cl (also called $E_{\text{T,Rel,Cl}}^{\text{miss,track}}$).

- m_T: transverse mass estimated as:

$$m_T = \sqrt{(p_T^{\ell\ell} + E_T^{\text{miss}})^2 - |\mathbf{p}_T^{\ell\ell} + \mathbf{E}_T^{\text{miss}}|^2} \ . \tag{A.2}$$

- $\Delta\phi_{\ell\ell,\text{MET}}$: azimuthal angular separation between the dilepton system and the missing transverse energy (MET) vector

- $\eta_{\ell\ell} = |\eta_{\ell_1} - \eta_{\ell_2}|$ is the dilepton separation in pseudoradity η

- f_{recoil}: soft hadronic recoil fraction defined as

$$f_{\text{recoil}} = \frac{|\sum |\text{JVF}| \times \vec{p_T}|}{p_T^{\ell\ell}} \ , \tag{A.3}$$

  where JVF is the jet vertex fraction, and the sum in $\vec{p_T}$ runs over all soft jets in the quadrant opposite to the dilepton ($\ell\ell$) system as illustrated in Fig. A-1. The soft jets entering the calculation of f_{recoil} do not overlap with electrons, have $p_{\text{T}} > 10$ GeV. The JVF is used for weighting without any JVF cut being applied. This variable is not used in the optimized analysis.

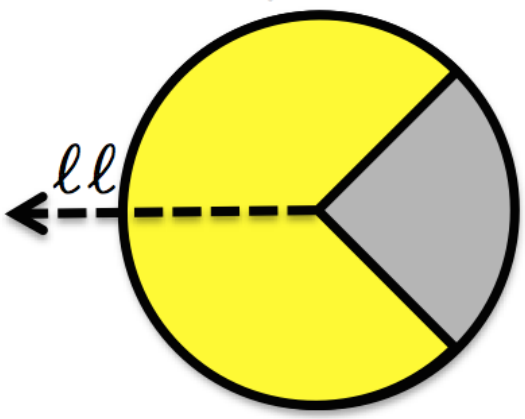


Figure A-1: Graph showing the soft hadronic recoil (in gray) against the dilepton system (in yellow) used in the calculation of f_{recoil} for the Z/Drell-Yan events.

A.1.0.4 Features of new analysis

The optimized analysis replaces single cuts by window cuts, i.e. cuts with both upper and lower limits, defining thus a closed phase space volume and maximizing the Higgs cross section. An example showing the normalized distributions of $E_{\mathrm{T,Rel,Cl}}^{\mathrm{miss,track}}$ in the 1 jet-bin signal and background events is illustrated in Fig. A-2. The background

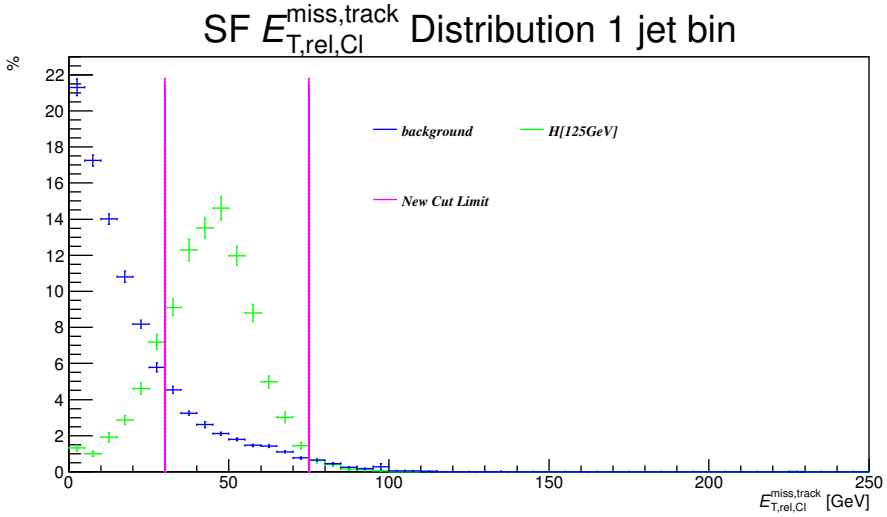


Figure A-2: Normalized distributions of $E_{\mathrm{T,Rel,Cl}}^{\mathrm{miss,track}}$ for the combined Higgs signal and total background events in the 1-jet bin, illustrating how the Higgs phase space is closed in $E_{\mathrm{T,Rel,Cl}}^{\mathrm{miss,track}}$ with finite limits. The results are shown for the 2012 $\ell\nu\ell\nu$ analysis

plot in blue corresponds to the total background events combined and Higgs signal events include contributions from events mainly. The Higgs events are confined in a limited phase space, falling within [30,75] GeV mainly. The lower limit cut at 30 GeV suppresses background significantly without significant losses in signal events. The upper cut contributes also to better signal-background separation. This procedure is repeated for other dynamic variables, defining thus the Higgs signal in a closed phase space.
The main features of the optimized analysis, compared to the official ATLAS analysis, are the following:

- Type of cuts: Only topological cuts on kinematic variables are used. Since a cut-based approach is adopted, there is correlation amongst cuts on kinematic variables. In addition, some cuts are flavor dependent since the kinematics of the event vary from opposite to same flavor channels.

- Use of window cuts to define an optimized closed phase space volume, enhancing Higgs signal events and suppressing background. The advantages of using window cuts are summarized in the next section.

- Higher expected significance and higher signal

- Higher statistics after all cuts
- Stability of the obtained signal regardless of the MC simulation changes, unlike the ATLAS analysis results with important fluctuations depending on the MC version used.
- Stability of the result regardless of the ordering of the cuts.
- Better results than the MVA (Multi-Variate Analysis)

A.1.0.5 Advantages of window cuts

The main advantage of using 'window' cuts, i.e. cuts with both upper and lower limit cuts, are:

- Well defined closed phase space maximizing the number of observed Higgs candidates and increasing the significance. To illustrate this idea, the significance based on different $p_T^{\ell\ell}$ values in the 0-jet bin is plotted for in Fig. A-3, showing that applying an upper limit on $p_T^{\ell\ell}$ leads to an increased significance beyond 2 when applying the lower limit cut. If the $p_T^{\ell\ell}$ cut is limited to a single cut with a lower limit as done in the ATLAS analysis, the significance is lower.

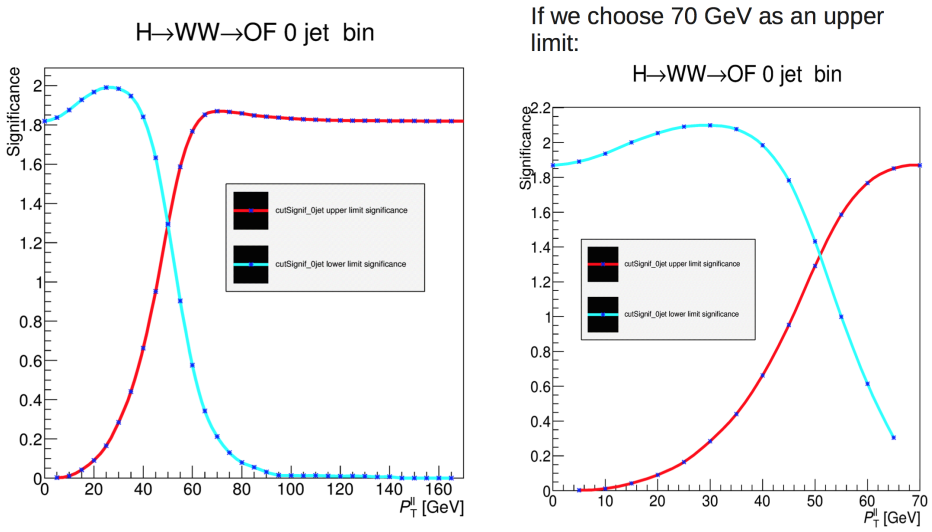


Figure A-3: Plots showing significance for lower limit cut on $p_T^{\ell\ell}$ in the 0 jet bin before (left) and after (right) applying an additional upper limit cut in the 2012 $\ell\nu\ell\nu$ analysis.

- Systematics are limited to finite phase space volume rather than an open one
- Better control and suppression of background events compared to single cuts
- Removing badly measured events e.g. events with very high p_{T} seen in the tails of track MET distributions up to 1TeV.

A.1.0.6 Cuts and event selection

For each variable describing topology and/or kinematics of the event, the significance is calculated for upper and lower cut limits. The cut values are chosen in a way to maximize the significance, which is defined as follows with 'S' being the signal, 'B' the total background and 'δ_B' the error on the background yield:

$$\text{significance} = \frac{S}{\sqrt{B + \delta_B \times \delta_B}} \ . \tag{A.4}$$

The significance was also studied using the definition:

$$\begin{aligned} \text{significance} &= \sqrt{2 \times (A - S)} \ , \\ A &= (S + B) \times \log(1 + S/B) \ . \end{aligned} \tag{A.5}$$

In the optimized analysis, the limits of the best phase space volume enhancing Higgs signal are defined based on the normalized signal and background (blinded) distributions, and the upper and lower cut limits are defined in parallel.

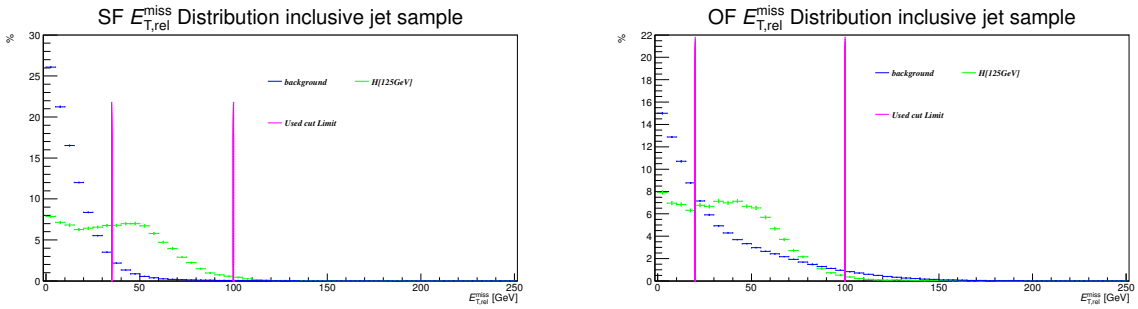
Starting with the cuts as defined in Table A.1, they are optimized at the various analysis stages:

- **Preselection:** The p_{T} of the leading lepton is not modified. A separate study shows that lowering the leading lepton p_{T} to 22 GeV increases the significance, but this was not adopted as it would require re-estimating the fakes from W+jets, which is beyond the scope of this analysis. All cuts are unchanged in the inclusive sample at preselection, except for METRel where:

  1. 20< METRel < 100 GeV for OF
  2. 35< METRel < 100 GeV for SF

  From the OF and SF METRel normalized distributions shown in Fig. A-4, it may seem that the upper limit cut is unnecessary, since beyond 100 GeV no Higgs events are there and few background events survive. In fact, the 100 GeV upper limit removes ∼8700 background events, a few hundreds of which will survive after all selection cuts are applied in the ATLAS analysis for the 0-jet and 1-jet bin events.

- **0-jet bin:** Since only cuts on kinematic variables related to the event topology and kinematics are kept, the f_{recoil} cut applied in the ATLAS analysis is dropped. It does not improve the significance. The same is seen with the $\Delta\phi_{\ell\ell,\mathrm{MET}}$, which is why it is dropped. And the track METRel application is extended from SF channel only to SF and OF channels. Track MET distributions with the new cut limits are shown in Fig A-5. The same procedure is indeed repeated for the remaining cuts in the 0-jet bin, adjusting existing cut limits and adding an upper or lower limit. A cut on $\eta_{\ell\ell}$ is added before the $\Delta\phi_{\ell\ell}$ cut for additional significance enhancement. The exact cut values are summarized in Table A.2.



(a) (b)

Figure A-4: Normalized distributions of METRel (calorimeter-based MET) at preselection for Higgs and background events in the SF (left) and OF (right) channels at the end of preselection in the 2012 $\ell\nu\ell\nu$ analysis. The pink lines show the limits taken into account during optimization for the definition of a METRel window cut.

- **1-jet bin:** Similarly to the 0 jet bin, the track METRel cut is extended to OF+SF channels, with upper limit cuts added as shown in Fig A-5. The same procedure is indeed repeated for the remaining cuts in the 1-jet bin. Moreover, two cuts on $\Delta\phi_{\ell\ell,\mathrm{MET}}$ and $p_T^{\ell\ell}$ are added before and after the $m_{\ell\ell}$ respectively in addition to a cut on $\eta_{\ell\ell} = |\eta_{\ell_1} - \eta_{\ell_2}|$. The $p_T^{\ell\ell}$ distributions in Fig. A-6 show that at least a simple cut with a lower limit is needed to enhance the Higgs signal events and suppress the background. On the other hand, the cut on $\Delta\phi_{\ell\ell,\mathrm{MET}}$ is a soft for additional enhancement while the cut on $\eta_{\ell\ell}$ is for completeness. The f_{recoil} is dropped in this category as well, especially that it adds nothing to the already obtained significance. The exact cut values are summarized in Table A.3.

A.1.0.7 Results

The results of the optimized MC based blinded analysis for 2011 and 2012 are summarized in Table A.4 and A.5, where the nominal ATLAS analysis results are shown as well.

A.1.0.8 What was missing in the ATLAS analysis

The official ATLAS winter 2013 selection has some unnecessary background events and badly measured ones, that are removed with the optimized 'window' cuts. The distributions of the events surviving the ATLAS analysis selection for various variables, which are removed by the optimized analysis, are shown in Fig. A-7 for the 0 jet bin for example, with about 140 unnecessary background events.

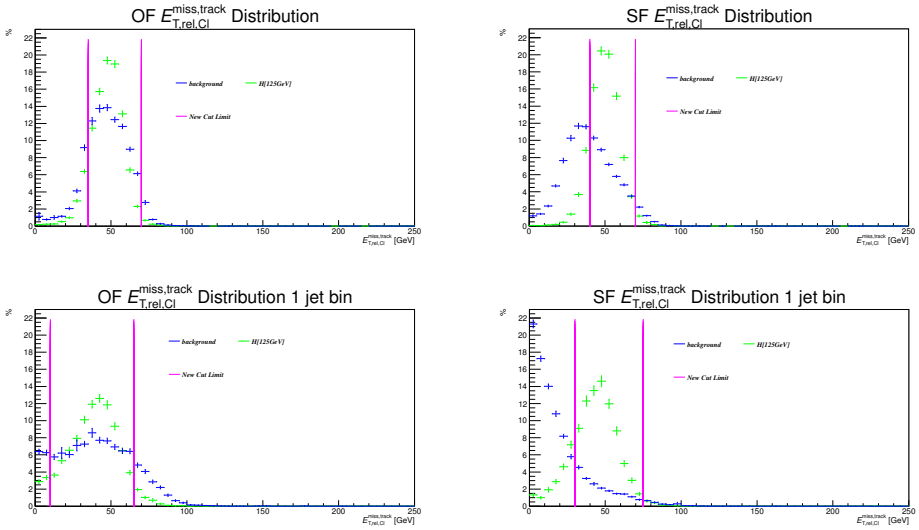


Figure A-5: Normalized Track METRel distributions for the Higgs and background events in the 0 jet bin (top) and 1 jet bin (bottom) with optimized cut limits for the 2012 $\ell\nu\ell\nu$ analysis. The pink lines show the limits taken into account during optimization for the definition of a track METRel window cut.

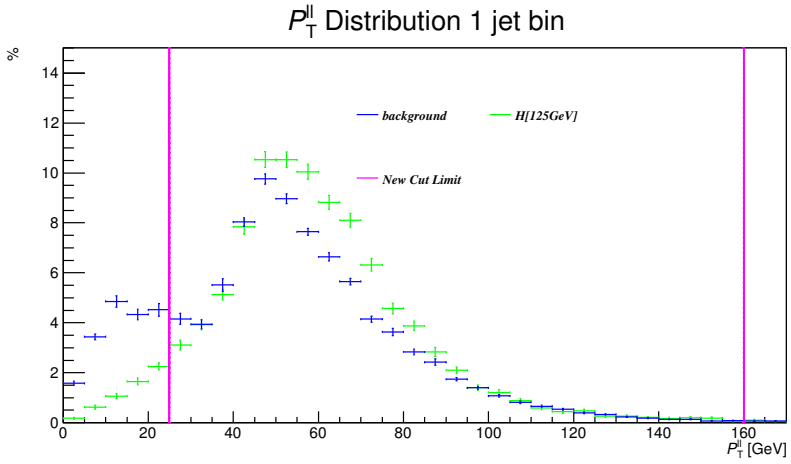


Figure A-6: $p_T^{\ell\ell}$ distribution in the 1-jet bin for all channels (OF+SF) combined in the 2012 $H \rightarrow WW^*\ell\nu\ell\nu$ analysis. The pink lines show the proposed cut limits to improve the significance and enchance the obtained Higgs signal.

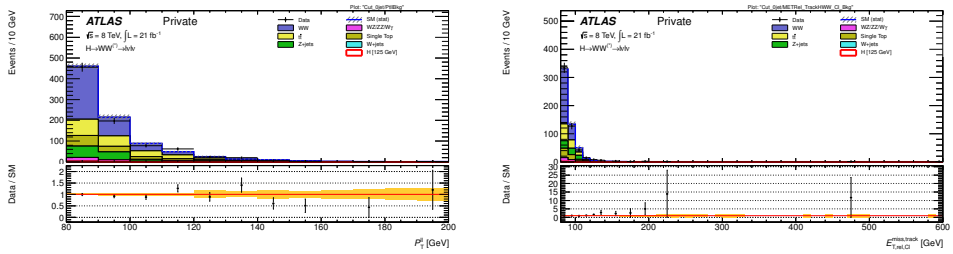


Figure A-7: Distributions of $p_T^{\ell\ell}$ (left) and Track METRel (right) for events in the 0 jet bin kept in the ATLAS analysis but removed in the optimized analysis for the 2012 $\ell\nu\ell\nu$ analysis.

A.1.0.9 Outcome used by ATLAS and Conclusion

An optimized blinded $H \to WW^* \to \ell\nu\ell\nu$ analysis was developed, with improved significance and statistics (multiplied by a factor of [4,4.5] depending on the jet bin). Where no signal was found by ATLAS in the 2011 analysis, the optimized analysis found the beginning of a signal, which is also confirmed by 2012 analysis results. This analysis is classified as 'high-p_{T}' analysis given the p_{T} of the leptons involved. Adding low-p_{T} leptons enhances the signal found even more. The 'window' cut approach defines well the best phase space, especially with track METRel cuts , enhancing the Higgs signal events yield.

The main ideas applied in the optimized analysis that were used later on in the subsequent versions of the ATLAS analysis are:

- Applying track MET cut for both OF and SF cuts.
- Removing tails of track MET by applying 100 GeV upper limit cut on tracks in the track MET calculation. This contribution was removed by the upper limit on the track MET cut added in the optimized analysis.
- The 2011 data sets were re-analyzed and a small signal was found eventually (see [2]).

Analysis Stage	Variable	Cut value	Channel
Preselection	leading lepton p_{T}	$>$ 25 GeV	all
	product of lepton charges	-1 (opposite sign)	all
	$m_{\ell\ell}$	$>$ 10 GeV	OF
		$>$ 12 GeV	SF
	$\mathrm{MET}^{\mathrm{rel}}$ (calorimeter)	$>$ 25 GeV	OF
		$>$ 45 GeV	SF
	Z veto	$\lvert m_{\ell\ell}$- 91.1876$\rvert >$ 15 GeV	SF
		-	OF
0 jet bin			
	n_{jet}	$= 0$	all
	$\Delta\phi_{\ell\ell,\mathrm{MET}}$	> 1.57	all
	$p_T^{\ell\ell}$	$>$ 30 GeV	all
	$m_{\ell\ell}$	$<$ 50 GeV	all
	Track METRel	$>$ 45 GeV	SF
		-	OF
	$\Delta\phi_{\ell\ell}$	< 1.8	all
	f_{recoil}	< 0.05	SF
		-	OF
	m_T	$0.75\times m_H < m_T < m_H$	all
		m_H=Higgs mass hypothesis	
1-jet bin			
	n_{jet}	$= 1$	all
	b-jet veto	b-tag counter= 0	all
	$Z \to \tau\tau$ veto	not (x1 $>$ 0 and x2 $>$ 0 and $\lvert m_{\tau\tau}$- 91.1876$\rvert \leq$ 25 GeV)	OF
		-	SF
	$m_{\ell\ell}$	$<$ 50 GeV	all
	Track METRel	$>$45 GeV	SF
		-	OF
	$\Delta\phi_{\ell\ell}$	< 1.8	all
	f_{recoil}	< 0.2	SF
		-	OF
	m_T	$0.75\times m_H < m_T < m_H$	all

Table A.1: Table summarizing the cut-based event selection at preselection and in the 0- and 1-jet bin categories for the standard ATLAS Moriond 2013 for the 2012 $H \to WW^*\ell\nu\ell\nu$ analysis.

Analysis Stage	Variable	Cut value	Channel
0 jet bin			
	n_{jet}	$= 0$	all
	$p_T^{\ell\ell}$	$>$ 30 GeV and $<$ 70 GeV	OF
		$>$ 35 GeV and $<$ 80 GeV	SF
	$m_{\ell\ell}$	$<$ 60 GeV	all
	Track METRel	$>$ 40 GeV and $<$ 70 GeV	SF
		$>$ 35 GeV and $<$ 70 GeV	SF
	$\eta_{\ell\ell}$	$<$ 3.0	all
	$\Delta\phi_{\ell\ell}$	$<$ 2.5	all

Table A.2: Table summarizing the optimized cut-based event selection in the 0-jet bin category for the 2012 $H \to WW^*\ell\nu\ell\nu$ analysis. The m_T cut is not shown since the final result of the cut-based analysis will be fitted for final signal extraction.

Analysis Stage	Variable	Cut value	Channel
1-jet bin			
	n_{jet}	$= 1$	all
	b-jet veto	b-tag counter$= 0$	all
	$Z \to \tau\tau$ veto	not (x1 $>$ 0 and x2 $>$0 and	OF
		$\lvert m_{\tau\tau}$- 91.1876$\rvert \leq$ 25 GeV)	
		-	SF
	$\Delta\phi_{\ell\ell,\text{MET}}$	$>$ 0.6	all
	$m_{\ell\ell}$	$<$ 60 GeV	all
	$p_T^{\ell\ell}$	$>$ 25 GeV and $<$ 160 GeV	all
	Track METRel	$>$30 GeV and $<$ 75 GeV	SF
		$>$10 GeV and $<$ 65 GeV	OF
	$\eta_{\ell\ell}$	$<$ 3.0	all
	$\Delta\phi_{\ell\ell}$	$<$ 2.5	all

Table A.3: Table summarizing the optimized cut-based event selection in the 1-jet bin category for the 2012 $H \to WW^*\ell\nu\ell\nu$ analysis. The m_T cut is not shown since the final result of the cut-based analysis will be fitted for final signal extraction.

	0 jet bin optimized	0 jet bin ATLAS	1 jet bin optimized	1 jet bin ATLAS	0+1 jet bins Combined
Number of 2012 data events selected	3162	801	1686	309	4848
Expected significance	4.22	3.57	2.85	2.46	5.09

Table A.4: Table summarizing results of the optimized and standard ATLAS analysis (around Moriond 2013) for the 0 jet and 1 jet bin 2012 analysis.

	0 jet bin optimized	0 jet bin ATLAS	1 jet bin optimized	1 jet bin ATLAS	0+1 jet bins Combined
Number of 2011 data events selected	640	154	297	62	937
Expected significance	1.95	1.99	0.87	1.09	2.14
Expected Higgs candidates	47.4	24.86	14	7.34	61.4

Table A.5: Table summarizing results of the optimized and standard ATLAS analysis (around Moriond 2013) for the 0 jet and 1 jet bin 2011 analysis.

Bibliography

[1] ATLAS Collaboration , *Measurements of the properties of the Higgs-like boson in the $WW^{(*)} \to \ell\nu\ell\nu$ decay channel with the ATLAS detector using 25 fb^{-1} of proton-proton collision data*, ATLAS-CONF-2013-030

[2] ATLAS Collaboration, *Observation and measurement of Higgs boson decays to WW^* with the ATLAS detector*, Phys.Rev. D**92** (2015) 1, 012006 arXiv:1412.2641 [hep-ex]

Appendix B

Track MET for the $H \to \tau^+\tau^-$ analysis

Track MET-Cl-j gives a similar performance to MET-STVF with less pile-up dependence showing small tails in the $p_{\mathrm{T,soft}}^{\mathrm{miss}}$ distribution and significantly smaller tail in the ΣE_T of soft term components distributions (Fig. [B-1- B-3]). m_{MMC} plots with Track MET-Cl-j are shown in Fig. B-4.

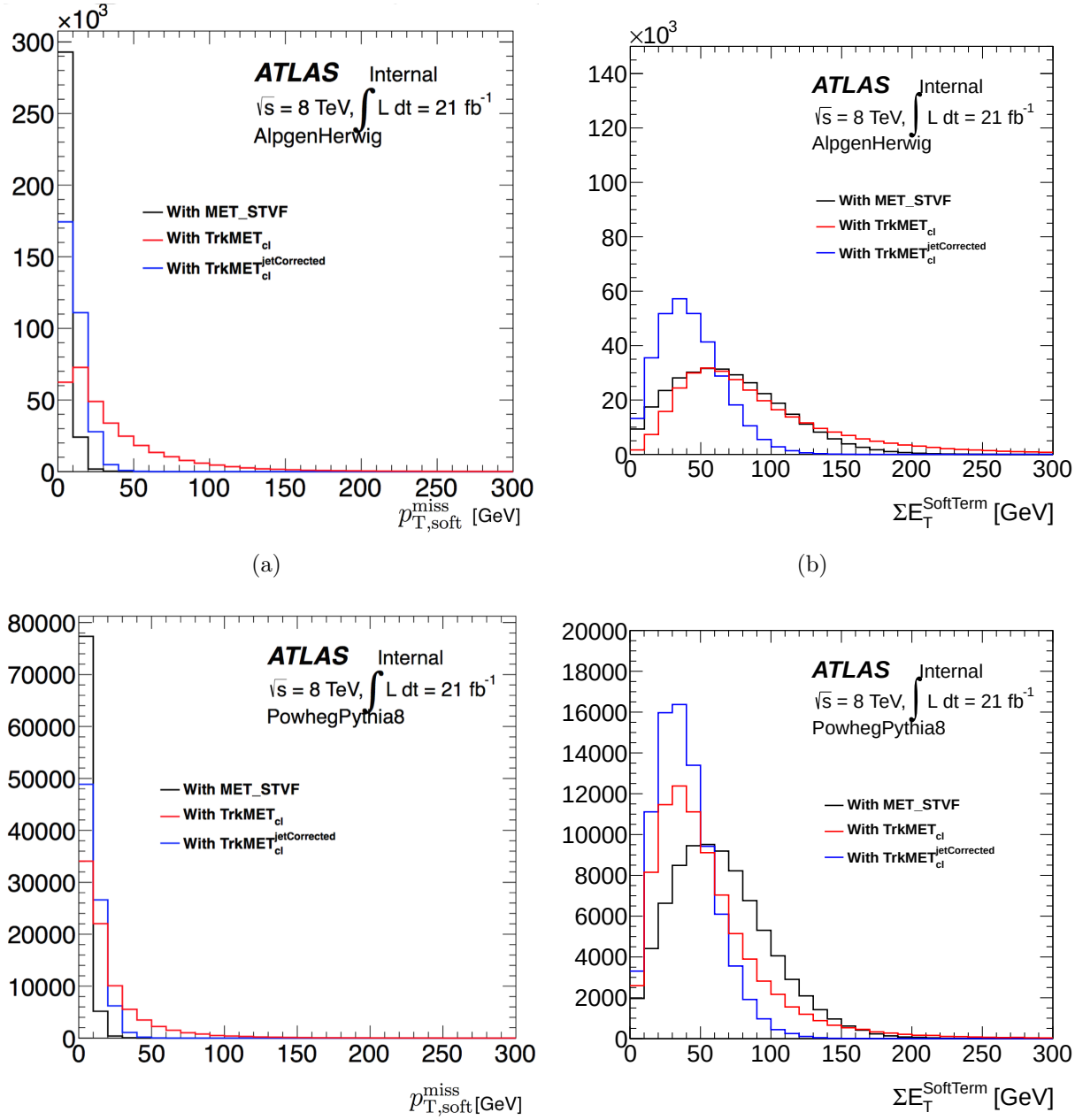
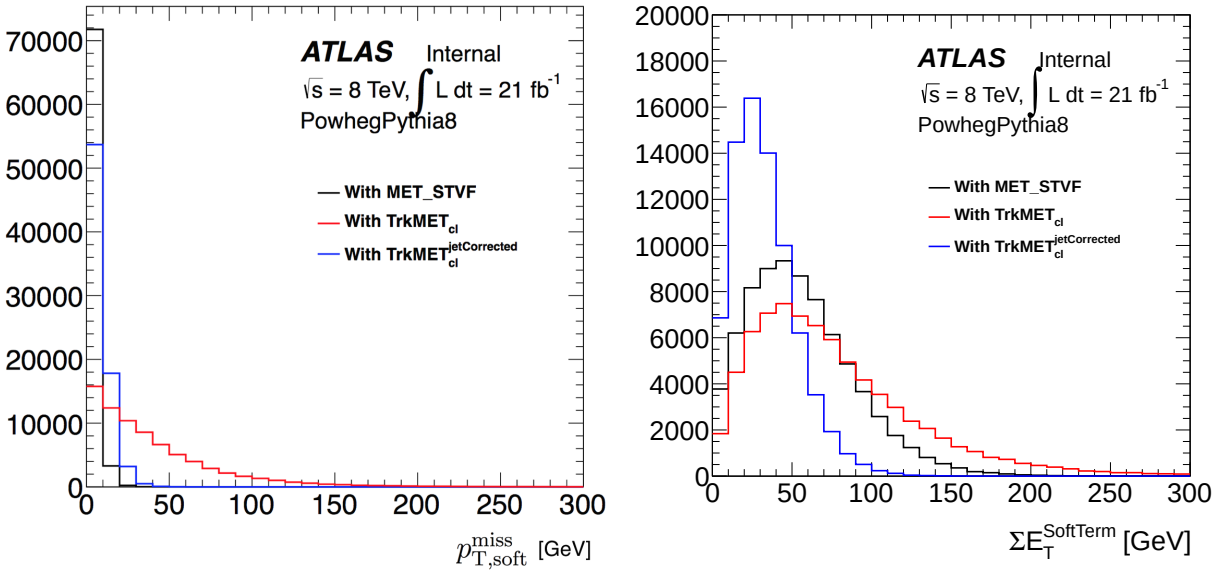
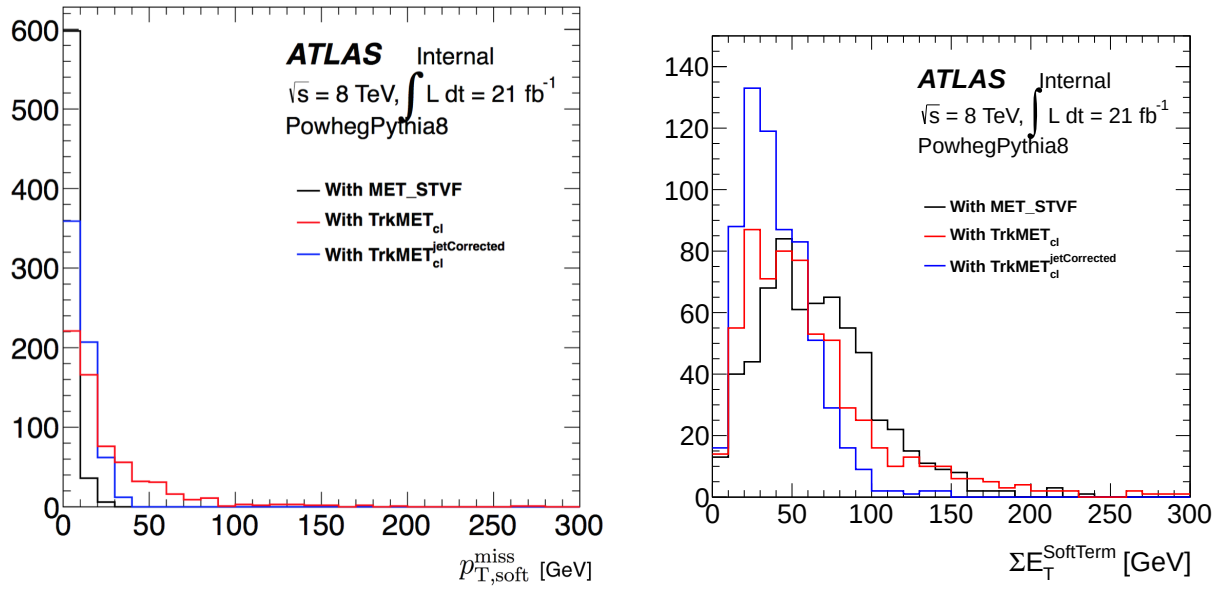


Figure B-1: $p_{\mathrm{T,soft}}^{\mathrm{miss}}$ (left) and ΣE_T of soft term components (right) distributions for various MET definitions for $Z \to \tau^+\tau^-$ (top) and ggF(bottom) events in the lep-lep channel.

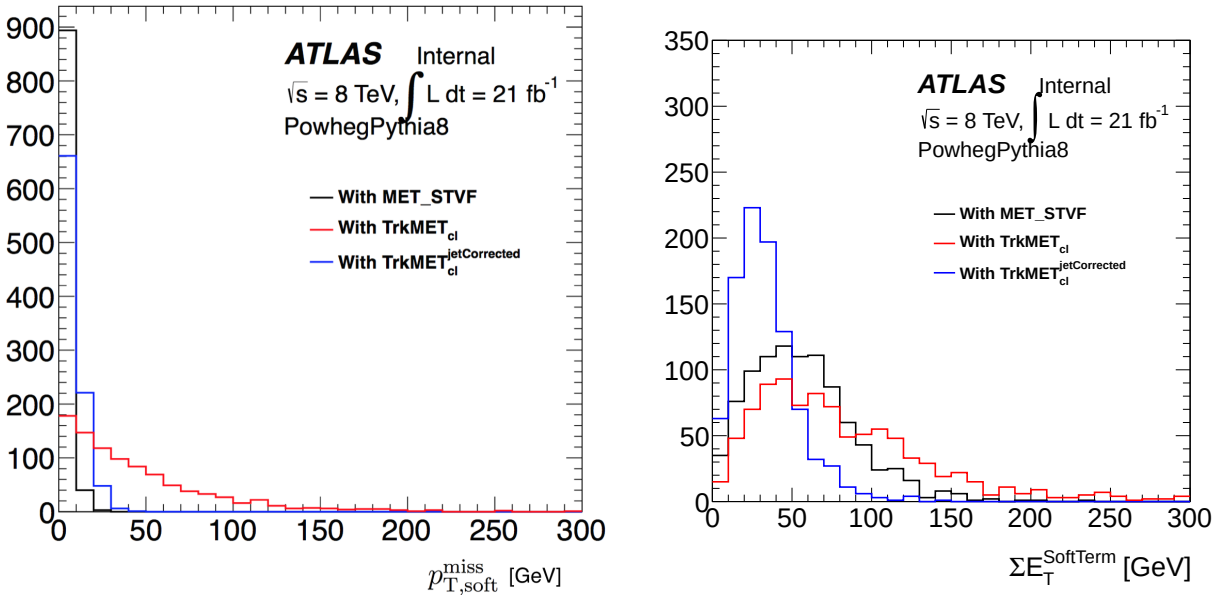


(a) (b)



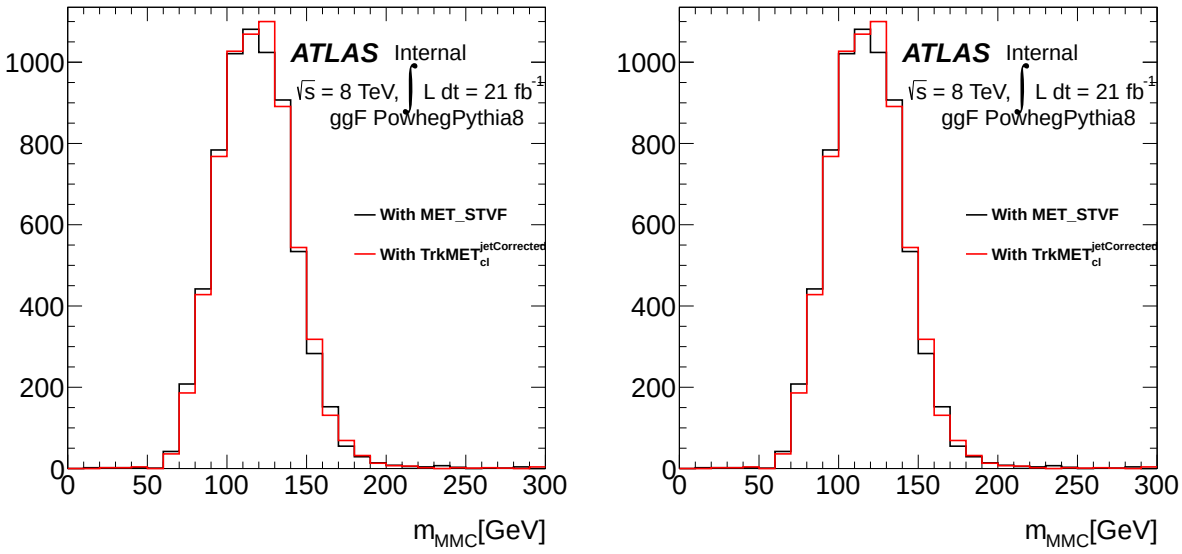
(c) (d)

Figure B-2: $p_{\mathrm{T,soft}}^{\mathrm{miss}}$ (left) and ΣE_T of soft term components (right) distributions for various MET definitions for lep-lep VBF(top) and lep-had ggF (bottom) events.



(a) (b)

Figure B-3: $p_{\mathrm{T,soft}}^{\mathrm{miss}}$ (left) and ΣE_T of soft term components (right) distributions for various MET definitions forVBF lep-had events .



(a) (b)

Figure B-4: m_{MMC} plots for gFF(left) and VBF(right) lep-had events .

Appendix C

Résumé Substantiel

Au LHC, l'un des buts essentiels était de trouver la dernière pièce manquante du modèle standard (MS), i.e. le boson de Higgs (H). La recherche fut couronnée de succès avec les données prises en 2012 et la découverte d'une nouvelle particule scalaire de masse ∼125 GeV, se désintégrant en deux bosons (deux photons ou deux bosons électrofaibles ZZ or W^+W^-). Pour vérifier la compatibilité de la nouvelle particule avec les prédictions du MS, son couplage aux fermions doit être établi, ce qui motiva la recherche du Higgs dans le mode de désintégration en deux leptons taus prédit dans le MS avec un rapport d'embranchement important. Cette thèse présente le travail effectué au sein du groupe d'analyse H→tau tau dans ATLAS, ainsi que l'étude d'une nouvelle méthode de reconstruction de l'énergie transverse manquante (MET) peu sensible aux effets d'empilement d'événements ('pile-up'), ce qui est prometteur pour les analyses futures (Run-II). En plus du travail lié à la recherche du Higgs expérimentalement, le travail présenté dans cette thèse, orienté vers le futur, montre aussi la possibilité d'explorer la nouvelle physique au delà du MS au LHC, de façon indépendante du modèle théorique. Cette option, récemment adoptée au CERN pour le Run-II du LHC, consiste à utiliser la théorie effective des champs pour estimer la production du boson de Higgs et ses couplages aux particules du MS (HEFT). L'application de cette méthode effective en utilisant le cadre de caractérisation de Higgs (HC) de Madgraph5_aMC@NLO a été étudiée, surtout pour les événements H→ $\tau\tau$.

Le premier chapitre présente le MS, montrant les détails de la brisure de symmétrie électrofaible nécessaire pour la génération des masses des particules élémentaires et les bosons de jauge médiateurs des forces fondamentales du MS. En plus, les modes de production et de désintégration du boson de Higgs analysés durant le Run-I sont présentés. La découverte du boson de Higgs est confirmée avec plus de 5 σ dans la plupart des modes étudiés, comme le montre tableau 1.6 pour chaque canal étudié dans ATLAS. Une description détaillée du LHC et du détecteur ATLAS est donnée dans le chapitre 2. Les générateurs Monte Carlo (MC) nécessaires pour la simulation d'événements et les détails de reconstruction des objets physiques (électrons, muons, jets, leptons taus, énergie transverse manquante) dans ATLAS sont présentés dans les chapitres 3 et 4 respectivement.

C.1 Track MET

Dans cette thèse, une contribution importante, mettant en relief l'amélioration obtenue avec une nouvelle estimation de l'énergie transverse manquante i.e. 'track-MET' (voir eq. (C.1)), est montrée.

$$\text{track-MET}_{x,y} \equiv E_{T,x,y}^{\text{miss,trace}} = -\sum_{traces} p_{x,y}^{trace} \quad , \quad \Sigma p_T = \sum_{\text{traces}} p_T^{\text{trace}} \tag{C.1}$$

Track-MET est calculée à partir de la somme des impulsions transverses p_T^{trace} des traces des particules chargées issues du vertex d'interaction de l'événement, tout en appliquant les corrections nécessaires pour:

- tenir compte des composantes neutres de jets en remplaçant la contribution provenant du p_{T} des traces associées au jet par le p_{T} du jet calibré et corrigé pour les effets de pile-up. L'association trace-jet peut être faite en utilisant 2 méthodes: ΔR ou 'ghost-association'. Les résultats montrés considèrent les 2 options.

- avoir une estimation précise de $p_T^{\text{électron}}$, en tenant compte de l'information acquise au niveau du cluster dans le calorimètre.

- éliminer les traces mal mesurées.

Track-MET est le complément de la définition de MET calculée à partir des dépôts d'énergie dans le calorimètre (calo-MET), tout en ayant l'avantage d'être plus robuste contre le pile-up. De plus, track-MET est équivalente à la définition de MET qui est utilisée par défaut pour le Run-II. En fait, en utilisant les traces chargées pour estimer la composante 'molle' (soft) de MET (définie dans eq. (C.2)) dans les événements issus de collisions p-p, la sensibilité au pile-up, inévitable dans les collisionneurs hadroniques à haute luminosité, est bien réduite. Ceci contribuera à améliorer les futures analyses $H \to \tau^+\tau^-$.

$$E_T^{\text{miss, track}} = E_{T,\text{ soft}}^{\text{miss, track}} + E_{T,\text{ hard}}^{\text{miss, track}} \tag{C.2}$$

où la composante 'molle' refère aux termes de track-MET ne provenant pas d'objets physiques bien reconstruits passant les critères de sélection requis (électrons, photons, jets, taus, muons), entrant dans le calcul de la composante 'dure' (hard).

Les erreurs systématiques associées à la composante molle ont été évaluées et leur dépendance sur les conditions de pile-up et de modélisation de l'événement (générateurs Monte Carlo, méthode de simulation, effet provenant de la composante 'dure') a été étudiée pour différentes définitions de MET. Les tests de clôture montre un bon accord data-MC, comme illustré dans Fig. 5-21(b). De plus, les corrélations entre

track-MET et calo-MET ont été étudiées.

C.2 Analyse $H \to \tau^+\tau^-$

Dans ATLAS, l'analyse $H \to \tau^+\tau^-$ est divisée en trois canaux selon le mode de désintégration (leptonique ou hadronique) des leptons taus: lepton-lepton, lepton-hadron et hadron-hadron. Les trois canaux de l'analyse sont caractérisés par de larges valeurs de l'énergie transverse manquante (MET) dans l'état final et adoptent la même technique pour identifier le lepton tau. Le travail présenté dans cette thèse concerne le canal 'lepton-hadron', où l'un des taus de l'état final se désintègre leptoniquement en un muon ou un électron, alors que l'autre se désintègre hadroniquement.

Dans l'analyse 'lepton-hadron', le bruit de fond dominant provient des événements dont un jet de hadrons est mal identifié comme un tau se désintégrant hadroniquement ('fake tau'). Le travail discuté montre en détail l'estimation de ce bruit de fond pour les deux configurations les plus sensibles aux événements de signal H, i.e. les événements produits avec un Higgs bien boosté ou ceux produits par fusion de deux bosons vecteurs (mode VBF). L'état final de ces derniers est caractérisé par deux jets bien séparés en pseudorapidité, répartis sur les deux hémisphères, produits en association avec les produits de désintégration du Higgs. La figure 6-30(c) met en relief la contribution importante des jets identifiés comme taus dans la distribution des événements VBF après toutes les sélections.

Les événements 'fake tau' ($j \to \tau$) dans la région de signal (SR) sont estimés en utilisant la méthode 'fake factor'. Le facteur de transfert FF est calculé comme le rapport des candidats tau passant les critères d'identification du lepton tau à ceux échouant ces critères (appelés aussi anti-τ_{had}), dans les régions de contrôle (CR) des bruits de fond QCD, W+jets, Z+jets et top.

$$\text{FF}^{\text{CR}} = \frac{N^{j\to\tau_{\text{had}},\text{CR}}_{\text{tau ID}}}{N^{j\to\tau_{\text{had}},\text{CR}}_{\text{anti-}\tau_h}} \tag{C.3}$$

La fraction de chacun de ces bruits de fond R_X (voir Tableau 6.9) dans la SR est estimée à partir de la région anti-τ_{had} équivalente. Les erreurs ont été évaluées séparément pour les données prises en 2011 (7 TeV) et en 2012 (8 TeV). D'un côté, les erreurs statistiques associés sont $\pm 22\%(\pm 11\%)$ pour les événements VBF (boosted) à 7 TeV et $\pm 4\%(\pm 2.3\%)$ pour les catégories VBF(boosted) à 8 TeV. De l'autre côté, les erreurs systématiques, évaluées en supposant une incertitude de 100% sur les valeurs de R_X, sont $\pm 3\%(\pm 6\%)$ pour les événements VBF(boosted) events à 8 TeV, alors qu'à 7 TeV elles sont $\pm 10\%(\pm 15\%)$ pour les catégories VBF(boosted). A 7 TeV, les erreurs statistiques sont dominantes.

Enfin, des tests de clôtures sur les événements de bruit de fond avec un 'fake tau' sont

conduits dans les CR MC et les CR data pour des candidats tau de même charge, montrant une différence consistante de 10% pour les événements 'boosted' à 7 TeV comparée à une différence négligeable à 8 TeV en tenant compte des erreurs statistiques.

A 8 TeV, le bruit de fond "fake tau" compte pour 60% du bruit de fond total des événements VBF et 43% pour les événements 'boosted'. Les chiffres correspondants à 7 TeV sont 55% et 38% pour les catégories VBF et boosted respectivement. Le résultat final de l'analyse montre une évidence de l'existence du boson de Higgs dans $H \to \tau^+\tau^-$ avec 4.54 σ, avec une contribution de 2.33 σ par le canal lepton-hadron. Enfin, l'amélioration de l'analyse $H \to \tau^+\tau^-$ dans le futur (Run-II) en utilisant une définition de MET a été étudiée aussi, surtout pour les canaux lepton-lepton et lepton-hadron.

C.3 HEFT

Enfin, cette thèse rapporte une contribution utilisant la théorie effective des champs pour estimer la production du boson de Higgs et ses couplages (HEFT), et explorer la nouvelle physique au delà du MS (BSM[1]) de façon indépendante du modèle théorique. La théorie de Fermi pour les intéractions électrofaibles est un bon exemple illustrant le pouvoir et l'utilité des théories effectives. Plusieurs bases de HEFT existent e.g. Warsaw, HISZ, SILH et Higgs. La dernière est développée par le groupe du CERN. Le travail présenté utilise le cadre de caractérisation de Higgs (HC), avec un Lagrangian effectif linéaire de dimension-6, comme implémenté dans Madgraph5_aMC@NLO. HC est compatible avec la plupart de ces bases.

HEFT est d'importance majeure pour le Run-II du LHC, comme les mesures de précision électrofaible et la recherche de toute déviation des couplages du MS du bosons de Higgs sont les sujets d'intérêt et de haute priorité pour ce run. En fait, HEFT sera utilisée par défaut durant le Run-II, remplaçant le κ-framework utilisé pour le Run-I, surtout que cette théorie permet de surmonter les limitations de ce dernier (e.g. traitement correct des corrections NLO[2], corrections électrofaibles, etc.) tout en offrant un traitement consistent des observables MS et BSM.

Le travail effectué à l'Université de Louvain, avec le groupe Madgraph5_aMC@NLO durant un stage au sein du groupe européen MCnet, consiste à tester et valider le modèle de désintégration des leptons taus ('TauDecay') dans le cadre d'une caractérisation du Higgs utilisant HEFT au sein de Madgraph5_aMC@NLO. Après avoir écrit un outil permettant de fusionner les fichiers de production et de désintégration du boson de Higgs (utile surtout en travaillant avec une précision au niveau NLO et/ou en cas de particules produites hors de leur couche de masse), la validation du modèle a été faite de 3 façons indépendantes: avec la génération d'événements au niveau

[1] Beyond the Standard Model
[2] Next-to-Leading-Order

d'éléments de matrice directement, avec l'outil créé et en désintégrant les taus avec MadSpin. La figure 7-5 montre les résultats de cette validation pour les distributions de la somme et la différence des angles azimuthaux des pions pour les événements $A/H \to \tau^+\tau^- \to \pi^+\nu\pi^-\overline{\nu}$ (hypothèse de spin $0^{-/+}$) et $Z \to \tau^+\tau^-$.

Cette validation a été possible après avoir résolu les problèmes liés aux paramètres nécessaires pour fusionner le modèle TauDecay et HC et la largeur effective du lepton tau. En plus, la validation de ces outils, a été faite aussi avec la nouvelle version de MadSpin. Ce nouvel outil est prêt à être utilisé durant le Run-II du LHC. Ma dernière contribution a été l'automatisation de cette version MadSpin.