UNIVERSIDADE D
COIMBRA

Susana Patrícia Amor dos Santos

STUDY OF HIGGS BOSON PRODUCTION IN ASSOCIATION WITH TOP QUARKS AT THE ATLAS EXPERIMENT AND SENSITIVITY OF NEW TOP QUARK OBSERVABLES

Tese no âmbito do Doutoramento em Física, ramo de Física Nuclear e das Partículas, orientada pelos Professor Doutor António Joaquim Onofre Abreu Ribeiro Gonçalves, Professor Doutor Filipe Manuel Almeida Veloso e Professor Doutor Juan António Aguilar Saavedra e apresentada ao Departamento de Física da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

Dezembro de 2022

UNIVERSIDADE D
COIMBRA

Susana Patrícia Amor dos Santos

Study of Higgs boson production in association with top quarks at the ATLAS experiment and sensitivity of new top quark observables

Tese no âmbito do Doutoramento em Física, ramo de Física Nuclear e das Partículas, orientada pelos Professor Doutor António Joaquim Onofre Abreu Ribeiro Gonçalves, Professor Doutor Filipe Manuel Almeida Veloso e Professor Doutor Juan Antonio Aguilar Saavedra e apresentada ao Departamento de Física da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

Dezembro 2022

Abstract

In this thesis, the search for the Standard Model (SM) Higgs boson produced in association with a top-quark pair, $t\bar{t}H$, was performed. A dedicated study of the dileptonic topology together with a Higgs boson decaying into a pair of b quarks is presented, using $20.3\,\text{fb}^{-1}$ of proton-proton collision data at a centre-of-mass energy of $\sqrt{s} = 8\,\text{TeV}$ collected with the ATLAS detector. Being mostly dominated by $t\bar{t}$ + jets production, the events were categorised according to their jet and b-tagged jet multiplicities, which suppresses the SM backgrounds without loss of events, while maintaining the highest possible signal sensitivity. In order to search for the signal, a combined statistical fit to signal-rich and signal-depleted regions was performed. An improved background prediction was achieved by combining information from differently populated regions. Signal-depleted regions were used either as background control regions, allowing to match the simulation prediction to that of data, or as sources of background information to constrain the statistical fit. A neural network was used to discriminate between signal and background events in the signal-rich regions. No significant excess of events above the background expectation was found and an observed (expected) 95% confidence level upper limit was set at 7.8 (4.5) times the Standard Model cross-section. The ratio of the measured signal strength to the Standard Model expectation was found to be $\mu = 3.7 \pm 2.2$ assuming a Standard Model Higgs boson with a mass of 125 GeV.

The sensitivity of several observables to the CP nature of the top quark couplings in neutral and charged currents was also investigated in a phenomenological approach. The top quark Yukawa coupling to the Higgs boson as well as the top quark electroweak coupling to the W boson were studied, considering different beyond the standard model (BSM) scenarios, while new observables were proposed to gauge their nature. The top quark-Higgs boson coupling is probed using dileptonic $t\bar{t}H/A$ events with either a scalar SM Higgs Boson, H, or a pure pseudoscalar CP-violating Higgs boson, A, considering also mixed states.

Angular distributions of the decay products, as well as CP angular asymmetries, are explored in order to separate the scalar from the pseudo-scalar components of the Higgs boson and reduce the contribution from the dominant irreducible background $t\bar{t}+b\bar{b}$ jets. Significant differences between angular distributions and between asymmetries of the $t\bar{t}H$, $t\bar{t}A$ and $t\bar{t}+b\bar{b}$ jets processes are observed. The nature of the top quark electroweak coupling to the W boson is also probed, using top quark (antiquark) polarisations in single top quark (antiquark) production processes, where possible deviations from the SM are considered in terms of anomalous Wtb couplings. Three orthogonal axes to investigate the top quark polarisation in the t-channel single top process were proposed and the expressions for the polarisation in these axes in terms of anomalous Wtb couplings were calculated. It was found that the polarisations in the two axes orthogonal to the spectator quark axis are very sensitive to an anomalous coupling involving a $\bar{b}_L\sigma^{\mu\nu}t_R$ dipole term. In particular, an asymmetry based on the polarisation normal to the production plane is more sensitive to the imaginary part of this coupling than previously studied observables.

In addition, the energy and noise response of the ATLAS hadronic calorimeter (TileCal) was investigated, including inter-correlations among cells of the calorimeter and with special attention to the effect of simultaneous multiple collisions (pile-up) during the first LHC run. A full survey of the TileCal cells was performed as a function of the pseudorapidity, bunch position in a train and number of interactions per bunch crossing. Several periods of data collected by the ATLAS experiment without any selective trigger at a center-of-mass energy of 7 TeV was analysed and compared to Monte Carlo simulation reweighted to the pile-up conditions of the data. The closely monitored response of the TileCal cells in different conditions contributed to a proper characterisation of the TileCal performance.

Resumo

Na presente tese é discutida a pesquisa da produção bosão de Higgs do Modelo Padrão (MP) em associação com pares de quarks top, $t\bar{t}H$. É apresentado um estudo dedicado à topologia dileptónica com decaimentos do bosão de Higgs para um par de quarks b, usando $20.3\,\text{fb}^{-1}$ de dados de colisões protão-protão com energia de centro-de-massa $\sqrt{s} = 8\,\text{TeV}$ recolhidos pelo detector ATLAS. Os eventos foram categorizados de acordo com as suas multiplicidades de jactos e jactos b, sendo o processo de fundo dominante a produção $t\bar{t}+$ jactos. A categorização em diferentes regiões de multiplicidade de jactos e jactos b suprime os fundos do MP sem perda de acontecimentos, enquanto mantém o máximo possível de sensibilidade ao sinal. De forma a procurar o sinal, foi efectuado um ajuste estatístico aos dados combinando regiões enriquecidas em sinal e fundo. Uma melhor previsão dos processos de fundo foi alcançada ao combinar a informação das diferentes regiões. Regiões ricas em fundo foram usadas como regiões de controlo dos processos de fundo, permitindo a normalização das simulações Monte Carlo dos fundos aos dados, ou alternativamente incluídas no ajuste estatístico para constranger as respectivas incertezas sistemáticas. Recorrendo a uma rede neuronal, foi possível discriminar os eventos de sinal dos de fundo nas regiões ricas em sinal. Não foi encontrado um excesso significativo de acontecimentos acima da expectativa de fundo previsto. Um limite superior observado (esperado) com um nível de confiança de 95% foi estabelecido em 7.8 (4.5) vezes a secção eficaz prevista pelo MP, para a produção de dileptónica de $t\bar{t}H$ em LHC a uma energia de centro de massa de 8 TeV. A razão entre as forças de sinal medida e a esperada pelo MP foi $\mu = 3.7 \pm 2.2$ assumindo um bosão de Higgs do MP com uma massa de 125 GeV.

A sensibilidade de vários observáveis à natureza CP dos acoplamentos do quark top em correntes neutras e carregadas foi investigada usando uma abordagem fenomenológica. Tanto o acoplamento do quark top ao bosão de Higgs, como o acoplamento electrofraco do quark top ao bosão W foram estudados

considerando diferentes cenários para além da física do MP, enquanto novas observáveis foram propostas para avaliar a sua natureza. O acoplamento do quark top ao bosão de Higgs foi explorado usando acontecimentos de $t\bar{t}H/A$ com um bosão de Higgs escalar, H, ou pseudoescalar, A, que viola a conservação de CP. Foram ainda considerados casos de mistura entre as componentes escalar e pseudoescalar. Distribuições angulares dos produtos de decaimento, bem como assimetrias angulares de CP, foram utilizadas para separar as componentes escalar e pseudoescalar do bosão de Higgs e reduzir a contribuição do fundo $t\bar{t}+b\bar{b}$ jactos cuja topologia do estado final é igual à do sinal. São observadas diferenças significativas entre distribuições angulares e assimetrias dos processos $t\bar{t}H$, $t\bar{t}A$ e $t\bar{t}+b\bar{b}$ jactos. A natureza do acoplamento electrofraco ao bosão W foi também investigada, recorrendo a polarizações do quark (antiquark) top em processos de produção simples do quark (antiquark) top, onde possíveis desvios ao MP foram considerados em termos de acoplamentos anómalos que possam existir no vértice Wtb. São propostos três eixos ortogonais para investigar a polarização do quark top no canal t do processo de produção simples. As expressões da polarização segundo esses eixos foram calculadas considerando acoplamentos anómalos. A polarização segundo os dois eixos ortogonais ao quark espectador é muito sensível a um acoplamento anómalo que envolva um termo $\bar{b}_L\sigma^{\mu\nu}t_R$. A assimetria baseada na polarização segundo a direcção perpendicular ao plano de produção é mais sensível à parte imaginária deste acoplamento do que outros observáveis já conhecidos.

Foi ainda investigada a resposta de energia e ruído do calorímetro hadrónico da experiência ATLAS (TileCal), incluíndo inter-correlações entre as células do calorímetro, com particular atenção ao efeito de colisões múltiplas simultâneas (*pile-up*) durante o primeiro run do LHC. Verificaram-se todas as células do TileCal em função da pseudorapidez, da posição do *bunch* num comboio de protões e do número de interações por *bunch crossing*. Foram analisados vários períodos de dados com energia de centro-de-massa 7 TeV recolhidos pela experiência ATLAS sem qualquer *trigger* de selecção, e comparados com simulação Monte Carlo de eventos pesados às condições de *pile-up* dos dados. A monitorização detalhada contribuiu para uma melhor compreensão da resposta do calorímetro hadrónico Tilecal.

Acknowledgements

I would like to thank my supervisors Professors António Onofre, Juan Antonio Aguilar-Saavedra, and Filipe Veloso for all the hard work, constant support and guidance along this journey. I would also like to thank Professor João Carvalho, who supervised my doctoral work in its first year. Working with each one of them was incredibly rewarding, and I grew a lot as a physicist and as a human being, through the passionate discussions shared over the years.

I acknowledge the support from Fundação para a Ciência e Tecnologia (FCT) through the Ph.D. scholarship SFRH/BD/73438/2010.

I would like to thank the University of Coimbra, in particular the Physics Department, and Laboratório de Instrumentação e Física Experimental de Partículas for providing the conditions to develop this doctoral work. The list of people from these institutions that have supported me in some way is almost as long as the document itself. Still, i would like to thank Professors Rui Marques, Amélia Maio, Patrícia Conde Muino, Helmut Wolters and Nuno Castro, as well as everyone in the LIP-ATLAS group. Thank you also to Rita, Alexandre, Paulo and Claudio with whom I had the pleasure of sharing an office in Coimbra for quite some time.

I would also like to thank Ana Henriques and Irene Vichou. When I first arrived at CERN, the TileCal group was my first home. Here, the guidance from Claudio Santoni was indispensable. To the experts with whom I learned so much, Sanya Solodkov and Luca Fiorini, thank you. A special thanks to Pavel Starovoitov, with whom i worked directly on the cell noise studies. I would also like to thank João Gentil for his insight whenever I was confused.

It really takes a village, and the analysis presented in this thesis would not have been possible without the ATLAS $t\bar{t}H$ group, with whom I had the chance to work with and learn from. In particular, Ricardo Gonçalo and Jared Adelman, whose shared enthusiasm for high-energy physics was always contagious.

I would like to thank Bruno and Juanpe for being my 'comrade in arms': your friendship made everything a lot more fun. A big thank you to everyone who made living at CERN a little less lonely: Joana, Lucia, Christian and Martin. I would also like to thank everyone I met in Departamento de Física Teórica y del Cosmos of the University of Granada for

their hospitality. In particular, Rafa, Miki, Fer, Patri and Maria del Mar. Thank you also to my friends in Braga, Emanuel and Henrique. A special thank you to Miguel, without whom I would probably have never gone to Coimbra in the first place. Thank you to Inês, Pedro, Diana, Nuno, Raquel, Magda, João Domingos, Melo, Joana, Fátima, Tiago, Márcio and Gafeira, who will always be Coimbra to me.

I would also like to mention my good friends Marta, Andreia, Sofia, Di, Sílvia, Sónia, Rute, Daniela, Lena, Filipa, and the rest of the gang who always made sure I was doing more than just this thesis. Thank you also to my book club cousins, Catarina and Inês, who made me fall in love with reading again. To the music enthusiasts, Stephanie and Selma, thank you for being close while so far away.

I would like to express my gratitude to my family: to my grandparents for always encouraging me to go after my dreams; to my mum for teaching me to be dedicated and hardworking; to my dad for passing along the curiosity of the unknown; and to my sister Ana whose love and support held me together through all of this.

This thesis is dedicated to Marco, who always asked me about the Higgs boson.

Finally, my appreciation to the Jury for reading this thesis.

This work was funded by Fundação para a Ciência e Tecnologia (FCT) through the grant SFRH/BD/73438/2010 and by the projects "Collaboration in the ATLAS Experiment at CERN" (CERN/FP/116346/2010 and CERN/FP/123595/2011), "Estudo do bosão de Higgs em produção associada de ttH no LHC" (EXPL/FIS-NUC/1705/2013), "Colaboração na experiência ATLAS no CERN" (CERN/FIS-NUC/0005/2015) and "Collaboration in the ATLAS experiment" (CERN/FIS-PAR/0008/2017).

"Adventure works in any strand — it calls to those who care more for living than for their lives."

Amal El-Mohtar and Max Gladstone,
This Is How You Lose The Time War

Contents

List of Figures

List of Tables

Glossary

2HDM two Higgs doublet model

ADC Analog-to-Digital Converter

ALICE A Large Ion Collider Experiment

ATLAS A Toroidal LHC ApparatuS

BCID Bunch Crossing Identifier

BCM Beam Condition Monitor

BDT Boosted Decision Trees

BDTG Gradient Boosted Decision Tree

BR Branching Ratio

BSM Beyond Standard Model

CERN European Organization for Nuclear Research

CKM Cabibbo-Kobayashi-Maskawa Matrix

CL Confidence Level

CMS Compact Muon Solenoid

CP Symmetry Charge Parity Symmetry

CR Control Region

CSC Cathode Strip Chambers

CTP Central Trigger Processor

DAQ Data AcQuisition System

DCS Detector Control System

DMU Data Management Unit

DQ Data Quality

EB Extended Barrel of the TileCal

ECal Electromagnetic Calorimeter

EF Event Filter

EMEC Electromagnetic Calorimeter End Cap

EW Electroweak

FB Forward-Backward

FCal Forward LAr Calorimeter

FCNC Flavour Changing Neutral Currents

FSR Final State Radiation

ggF Gluon-Fusion

GIM Mechanism Glashow, Iliopoulos and Maiani Mechanism

GRID Worldwide LHC Computing Grid

GRL Good Runs Lists

GUT Grand Unified Theory

HEC Hadronic End-Cap LAr Calorimeter

HF Heavy Flavour

HG High Gain

HLT High Level Trigger

ID Inner Detector

IP Interaction Point

JES Jet Energy Scale

JVF Jet Vertex Fraction

LAr Liquid Argon

LB Long Barrel of the TileCal

LCW Local Cluster Weighting

LEP Large Electron-Positron Collider

LG Low Gain

LH Likelihood

LHC Large Hadron Collider

LHCb Large Hadron Collider Beauty Experiment

LL Leading Logarithmic

LO Leading Order

LUCID LUminosity measurement using Cerenkov Integrating Detector

Lumiblock Luminosity Block

LVL1 Level 1 of Trigger System

LVL2 Level 2 of Trigger System

LVPS Low-Voltage Power Supply source of the TileCal front-end electronics

MBTS Minimum Bias Trigger Scintillators

MC Monte Carlo

MDT Monitored Drift Tubes

ME Matrix Element

ML Machine Learning

MPI Multiple Parton Interactions

MS Muon Spectrometer

MVA Multivariate Analysis

N^{3}LO Next-to-Next-to-Next-Leading-Order

NLO Next-to-Leading Order

NN Neural Networks

NNLL Next-to-Next-to-Leading Logarithmic

NNLO Next-to-Next-to-Leading Order

NP nuisance Parameter

OF Optimal Filter method

OS Opposite Sign

PDF Parton Distribution Functions

PMNS Pontecorvo-Maki-Nakagawa-Sakata Matrix

PMT Photomultiplier Tube

QCD Quantum Chromodynamics

QED Quantum Electrodynamics

RF Radiofrequency

RMS Root Mean Square

ROB Readout Buffers

ROC Receiver Operating Characteristic

ROD Readout Drivers

RoI Region of Interest

RPC Resistive Plate Chambers

SCT Semiconductor Tracker

SF Scale Factors

SFI Sub-Farm Input

SFO Sub-Farm Output

SM Standard Model

SPS Super Proton Synchrotron

SR Signal Region

SS Same Sign

SUSY Supersymmetry

TGC Thin Gap Chambers

TileCal Tile Calorimeter

TMVA Tool for Multivariate Analysis

TNF Tile Noise Filter

TRF Tag Rate Function

TRT Transition Radiation Tracker

TTC Trigger, Timing and Control system

UE Underlying Event

V-A Vector minus Axial-vector

VBF Vector-Boson Fusion

VEV Vacuum Expectation Value

WLS Wavelength Shifting

1

Introduction

The Standard Model of Particle Physics (SM) provides the most precise description, to date, of the fundamental forces and elementary particles composing known matter. The discovery of the Higgs boson at the Large Hadron Collider (LHC) in July 2012 by the ATLAS and CMS collaborations constituted an important validation of the SM theory. The top quark, being the heaviest known elementary particle in the SM, is expected to have a much stronger Yukawa coupling to the Higgs boson than any other fundamental particle. Couplings of the Higgs boson to heavy quarks are a strong focus of experimental searches at LHC, since new sources of CP violation are needed to explain the matter/antimatter asymmetry in the Universe. The SM Higgs boson production in association with a top quark pair ($t\bar{t}H$) with subsequent Higgs decay into bottom quarks ($H \to b\bar{b}$) addresses heavy quark couplings in both production and decay. Direct measurement of the top quark-Higgs boson coupling can be achieved through the measurement of the $t\bar{t}H$ production cross-section, which are complementary to indirect measurements of the coupling through loop effects, such as the production via gluon fusion ($gg \to H$) for instance. If new physics happens at the vertex, direct measurements are of utmost importance.

The search for the SM Higgs boson in the $t\bar{t}H$ production mode at a centre-of-mass energy of $\sqrt{s} = 8\,\text{TeV}$ using $20.3\,\text{fb}^{-1}$ of data collected with the ATLAS detector is studied in this thesis. A dedicated study of the dileptonic topology together with a Higgs boson decaying into a pair of b quarks is presented. The search for the signal is performed using a combined statistical fit to signal-rich and signal-depleted regions. The combined analysis for both semileptonic and dileptonic $t\bar{t}$ decays has been published in The European Physical Journal C [1].

In addition, the performance of the hadronic calorimeter of the ATLAS detector (TileCal) is studied. As the conditions of the LHC proton beams change towards the designed values, the energy and noise response of the TileCal are monitored under increasing multiplicity of simultaneous collisions. By contributing to a better characterisation of the TileCal response and performance, these studies improve object reconstruction, and consequently, provide more precise results from the ATLAS experiment [2].

Finally, the sensitivity of several top quark observables to the CP nature of the top quark couplings in neutral and charged currents is investigated using MC simulated events [3, 4, 5]. The study of the top quark Yukawa coupling to the Higgs boson, wether in line with the pure scalar SM Higgs boson, or not, is presented, considering angular observables in $t\bar{t}H/A$ signal events, where A is a pseudoscalar CP-violating Higgs-like boson. The nature of the top quark electroweak coupling to the W boson is investigated using top quark (antiquark) polarisations in single top quark (antiquark) production processes. Here, possible deviations from the SM are tested by considering anomalous top quark couplings. Both these exploratory studies make use of the spin information in the events to build asymmetries as possible sources of discrimination between the SM expectations and the physics beyond our present knowledge.

This dissertation is organised as follows. In Chapter 2 an introduction to the SM and the current theoretical understanding of top quark and Higgs Boson physics is presented. Chapter 3 describes the main features of the ATLAS detector and in Chapter 4 the monitoring of the TileCal energy response and associated noise is summarised. The analysis and data setup for the $t\bar{t}H$ search is shown in Chapter 5, including information on the reconstruction of physics objects and the event selection. In Chapter 6 the dileptonic $t\bar{t}H$ signal as well as the most relevant backgrounds of the analysis are investigated, and new data-driven methods are described. The analysis results, including the signal discrimination based on multivariate techniques, the treatment of systematic uncertainties on the statistical fit and the obtained limits on the cross-section of the $t\bar{t}H$ associate production, are discussed in Chapter 7. The sensitivity of top quark observables to BSM physics in top quark couplings in neutral and in charged currents is explored in Chapter 8, from the phenomenological point of view. Chapter 9 summarises the main conclusions of the studies developed in this dissertation.

2

Theoretical Framework

The Standard Model (SM) of particle physics is the most successful theory describing the elementary particles and interacting fields. Using quantum field theory, the SM incorporates three of the four fundamental forces known in the universe. It has been thoroughly tested and proven extremely accurate in describing experimental observations thus far.

In this chapter, the underlying constituents of matter and their interactions are presented. A brief introduction to the SM Lagrangian is also included, covering its symmetries, experimental results and shortcomings. The physics of the top quark is then discussed, with particular emphasis in its production in association with a Higgs boson. The experimental search for this production mode in dilepton final states is the general subject of this thesis and is particularly relevant as a thorough test of the theoretical model.

2.1 The Standard Model of Particle Physics

Developed since the 1960's, the SM is a renormalisable relativistic quantum field theory that provides an elegant description of the interaction among the constituents of matter (*fermions*) through the exchange of force mediators (*bosons*). Initially formulated to merge the quantum electrodynamics (QED) with the weak interaction by Glashow [6] and later Weinberg [7], the framework only completely described the electroweak interaction, including the different fermion and boson masses, after Salam [8] incorporated the Brout-Englert-Higgs mechanism [9]–[12]. The formulation of quantum chromodynamics (QCD), designed analogously to describe the strong interactions, is a result of independent efforts: Gell-Mann and Zweig [13], who suggested that hadrons were composite of different flavoured quarks; Han and Nambu [14] as well as Greenberg [15], who proposed the colour charge;

Interaction	Range [m]	Boson	Mass [GeV]	Electric Charge (Q)
Electromagnetic	∞	photon, γ	$< 1 \times 10^{-27}$	0
Weak	10^{-18}	Z^0	91.1876 ± 0.0021	0
		$W^{\pm}$	80.377 ± 0.012	± 1
Strong	10^{-15}	gluons, g	0	0
Gravitational	∞	graviton (hypothetical)	–	–

Table 2.1: Summary table of nature's fundamental forces, their ranges, mediator bosons and characteristics [19]. The electric charge, Q, is given by integer multiples of the elementary charge, $q_e = 1.602 \times 10^{-19} C$.

Politzer [16, 17], who disclosed the asymptotic freedom of strong interactions; and finally Gross and Wilczek [18] ensuring that perturbation theory techniques could be applied.

2.1.1 Fundamental Particles and Interactions

In the universe, the fundamental interactions can be described by the electromagnetic force, the weak force, the strong force and the gravitational force. In the SM, forces happen via the exchange of particles with integer spin, the gauge bosons. The electromagnetic force is mediated by massless photons γ, the weak force by massive neutral Z and charged $W^{\pm}$ bosons, whereas the strong force is propagated by eight massless colour charged gluons g. In Table 2.1, the fundamental interactions, their mediators and characteristics are listed. The gravitational interaction is the only force not described in the SM, since a renormalisable formulation as a quantum field theory is not yet known. It can, however, be neglected for energies lower than the Planck scale (1.22×10^{19} GeV) in the context of particle physics.[1]

Known matter is made by half-integer particles, the fermions, that follow the Pauli exclusion principle. These elementary particles can only interact via some force if they have the corresponding quantum number. For the electromagnetic force this is the electric charge. Similarly, for the weak force, the quantum number is the isospin, which all fermions carry. In the case of the strong force, only particles holding colour charge are able to interact.

Fermions can be subdivided into leptons and quarks, where the former do not have colour charge and therefore can only participate in EW interactions. There are twelve

[1] In this work, gravity is thus only discussed when appropriately needed.

Gen.	Fermion	Mass [GeV]	Electric Charge (Q)	Isospin (J)	Colour Charge (c)
		Leptons			
1st	Electron, e	0.5109×10^{-3} $\pm 0.15 \times 10^{-12}$	-1	$-1/2$	–
	Electron neutrino, ν_e	$< 1.1 \times 10^{-3}$	0	$1/2$	–
2nd	Muon, μ	$0.1057 \pm 0.23 \times 10^{-8}$	-1	$-1/2$	–
	Muon Neutrino, ν_μ	$< 0.19 \times 10^{-3}$	0	$1/2$	–
3rd	Tau, τ	$1.7769 \pm 0.12 \times 10^{-3}$	-1	$-1/2$	–
	Tau Neutrino, ν_τ	$< 0.182 \times 10^{-3}$	0	$1/2$	–
		Quarks			
1st	Up, u	$0.00216^{+0.00049}_{-0.00026}$	$+2/3$	$1/2$	✓
	Down, d	$0.00467^{+0.00048}_{-0.00017}$	$-1/3$	$-1/2$	✓
2nd	Charm, c	1.27 ± 0.02	$+2/3$	$1/2$	✓
	Strange, s	$0.0934^{+0.0086}_{-0.0034}$	$-1/3$	$-1/2$	✓
3rd	Top, t	172.69 ± 0.30	$+2/3$	$1/2$	✓
	Bottom, b	$4.18^{+0.03}_{-0.02}$	$-1/3$	$-1/2$	✓

Table 2.2: Summary table of quark and lepton generations in the SM, including the fermions's mass, electric charge, isospin and presence of colour charge [19]. The electric charge, Q, is given by multiples of the elementary charge, $q_e = 1.602 \times 10^{-19} C$.

different fermions, six leptons and six quarks, that can be grouped in three families or generations consisting of two particles each, as summarised in Table 2.2. To each fermion there is a corresponding anti-particle, with the same mass but opposite electric charge. The generations of quarks and leptons share the same quantum numbers except an increasingly higher mass value, the third generation being the heaviest one.

The first elementary particle discovered was the electron e [20, 21, 22]. It belongs to the first generation of leptons, together with the electron neutrino ν_e. The muon μ and muon neutrino ν_μ constitute the second generation of leptons and the tau τ and tau neutrino ν_τ the third generation. The neutrinos are neutrally charged particles, assumed massless in the SM, but experimentally proven to have a small non-zero mass [23, 24]. The remaining leptons have an electric charge proportional to the elementary charge and their masses vary by a range of four orders of magnitude.

Unlike the leptons, quarks can interact through both electroweak and strong forces, since

they have electric charge, isospin and colour charge. Quarks form bound states of colour-neutral particles, the hadrons, due to a property of the strong interaction called confinement. As such, quarks cannot be found isolated, forming instead hadronic states: baryons (three quarks) or mesons (quark-antiquark systems).[1] The third-generation's top quark, t, however, being the heaviest elementary particle known to date ($m_t = 172.69 \pm 0.30\,\text{GeV}$[19]) has such a short lifetime that, when produced, decays before hadronising. Baryons and mesons are short-lived, rapidly decaying through lighter particles. The proton is the only known stable baryon, made up of two up quarks and one down quark.

Finally, the last piece of the SM, discovered in 2012 by the ATLAS and CMS collaborations [25, 26], is the Higgs boson. The Brout-Englert-Higgs mechanism explains the spontaneous breaking of electroweak symmetry, accounting for the masses of the $W^{\pm}$ and Z bosons, as well as those of the fermions. The Higgs particle is a massive scalar boson ($m_H = 125.25 \pm 0.17\,\text{GeV}$ [19] and spin 0), neutrally charged and constituting its own anti-particle.

2.1.2 The SM Lagrangian

Incorporating the electromagnetic, the weak and the strong interactions, the SM is constructed to have total invariance under the gauge group:

$$G_{SM} = SU(3)_c \times SU(2)_L \times U(1)_Y \tag{2.1}$$

where $SU(3)_c$ and $SU(2)_L \times U(1)_Y$ are the symmetry groups of the strong and the electroweak interactions, respectively. Here, c refers to the colour charge, L to the left chiral nature of $SU(2)_L$ coupling and Y to the weak hypercharge. Following Noether's theorem, which states that each symmetry of a Lagrangian has a corresponding conserved quantity, the symmetry of the SM Lagrangian under these transformations guarantees that physics conservation laws, such as conservation of energy, momentum and angular momentum, are rendered. In order to assure locality, that is, invariance under local transformations, gauge invariance is required, which can be sustained by introducing gauge fields in the SM Lagrangian. The number of symmetry group generators defines the number of associated

[1] Exotic bond states of four or five quarks (tetraquarks and pentaquarks, respectively) are also theoreticaly possible, but only more recently experimentally discovered. It is not, however, completely clear whether these experimentally observed heavy states constitute a system of four (five) quarks tightly bound together in a true tetraquark (pentaquark) or instead a pair of mesons (a meson and a baryon) weakly bound in a molecule-like structure. Nevertheless, these new states are an interesting place to further study the strong interaction.

gauge boson fields, which in the SM results in eight massless gluons from the 8 generators of $SU(3)_c$ symmetry and four massless bosons from the 3 generators of the $SU(2)_L$ and 1 generator of the $U(1)_Y$ group (see Table 2.1). Furthermore, imposing local gauge symmetries allows the theory to be renormalisable, ensuring physics predictions at multiple energy scales [27].

The SM Lagrangian can be constructed considering the electroweak and strong interactions independently: $\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{EW}} + \mathcal{L}_{\text{QCD}}$.

The Electroweak Interaction

The electromagnetic and weak interactions can be described as different aspects of the same unified theory, the electroweak theory, described by the $SU(2)_L \times U(1)_Y$ group. Even though these forces seem very different at low energy scales, they are unified around the order of the electroweak energy scale, $v = 246\,\text{GeV}$, since the associated generators mix. To respect local invariance under both symmetry groups, the associated covariant derivative is given by:

$$D_\mu = \partial_\mu - ig\frac{\sigma_a}{2}W_\mu^a - ig'\frac{Y}{2}B_\mu\,, \tag{2.2}$$

where the Pauli matrices σ_a are the three $SU(2)_L$ generators[1] with three associated gauge fields W_μ^1, W_μ^2 and W_μ^3 and the weak hypercharge Y is the generator of $U(1)_Y$ with an associated gauge field B_μ. These generators satisfy the Lie Algebra: $[\sigma_a, \sigma_b] = i\epsilon_{abc}\sigma^c$ and $[\sigma_a, Y] = 0$. The coupling constants g and g' represent the interaction strengths associated to the weak ($SU(2)_L$) and quantum electrodynamic ($U(1)_Y$) forces, respectively.

In the symmetry group of quantum electrodynamics $U(1)_Y$, the hypercharge Y is the conserved quantum number:

$$Y = 2(Q - T_3)\,, \tag{2.3}$$

where Q is the already defined electric charge of the fermion and T_3 is the third component of the weak isospin.

For the symmetry group of the weak interaction, the associated quantum number is the isospin, T. The left- and right-handed chiral components[2] of the fermion fields transform

[1] The Pauli matrices are $\sigma_1 = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, $\sigma_2 = \left(\begin{smallmatrix} 0 & -i \\ i & 0 \end{smallmatrix}\right)$ and $\sigma_3 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$.

[2] For massive particles, chirality is not the same as helicity. Chirality is determined by whether a particle transforms in a right- or left-handed representation of the Poincaré group, whereas helicity is simply the projection of its spin into its direction of motion.

differently under $SU(2)_L$ transformations:

$$\begin{aligned} \psi_L &= \frac{1}{2}(1-\gamma^5)\psi \\ \psi_R &= \frac{1}{2}(1+\gamma^5)\psi\,. \end{aligned} \tag{2.4}$$

Left-handed fermions will transform as $SU(2)_L$ doublets, while right-handed fermions transform as a $SU(2)_L$ singlet, thus leading to the following grouping of quarks and leptons:

$$\begin{aligned} f_L^i &= \left(\begin{smallmatrix}\nu_L^i\\ l_L^i\end{smallmatrix}\right), \left(\begin{smallmatrix}u_L^i\\ d_L^i\end{smallmatrix}\right) \\ f_R^i &= l_R^i\,, u_R^i\,, d_R^i\,, \end{aligned} \tag{2.5}$$

where the index $i = 1, 2, 3$ is the fermion generation, ν represents the neutrinos, l the leptons and u and d the up and down-type quarks, respectively.[1] Left-chiral fermions carry isospin and can interact via the weak force, while right-chiral fermions have isospin zero and do not weakly interact. The left nature of $SU(2)_L$, where its gauge fields can only couple to left-chiral fermions, is an observed aspect of the weak force: the violation of parity symmetry [28, 29]. Parity, or space-inversion, transforms a left-chiral spinor transformation into a right-chiral spinor transformation and vice-versa.[2] Since the chiral components of the fermion fields transform differently, the SM Lagrangian is not invariant under parity transformations; hence parity can be violated by weak interactions.

Weak interactions produce only left-chiral particles, which means that the exclusively weakly interacting neutrinos would have only one possible chirality. However, for massive particles, chirality is not conserved during propagation. Experimentally observed neutrinos have a small non-zero mass and, therefore, a right-chiral neutrino component should exist in nature, even if it does not participate in any known interaction. Right-chiral components of all the other fermions are available since they can be produced via other interactions too. In general, mass eigenstates do not have a defined chiral state, ergo their representation in the Lagrangian by a Dirac spinor ψ with both chiral components.

The physically observable gauge bosons of the electroweak theory are not the W_μ^a and

[1] Up-type quarks, u^i, are the upper components of the $SU(2)_L$ quark doublet (u, c and t), whereas down-type quarks, d^i are the lower components of the doublets (d, s and b).

[2] For a massless fermion, parity reverses not only its chirality but also its helicity, since a parity transformation inverses the spatial coordinates of the fermion field.

B_μ, but instead a superposition of them:

$$\begin{aligned} A_\mu &= \sin\theta_W W_\mu^3 + \cos\theta_W B_\mu \\ Z_\mu^0 &= \cos\theta_W W_\mu^3 - \sin\theta_W B_\mu \\ W_\mu^\pm &= \sqrt{\frac{1}{2}(W_\mu^1 \mp i W_\mu^2)}\,. \end{aligned} \tag{2.6}$$

From these linear combinations of the electroweak gauge bosons, the photon field A_μ, the Z boson field Z_μ^0 and the W boson fields $W_\mu^\pm$ are obtained, where the weak mixing angle θ_W is defined as $\tan\theta_W = \frac{g'}{g}$.

An additional gauge invariant kinetic term must be added to the SM Lagrangian:

$$\mathcal{L}_\mathrm{K} = -\frac{1}{4} W_{\mu\nu}^i W_i^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}\,. \tag{2.7}$$

Here, the $SU(2)_L$ and $U(1)_Y$) field tensors, $W_{\mu\nu}^i$ and $B_{\mu\nu}$ respectively, are defined as:

$$\begin{aligned} W_{\mu\nu}^i &= \partial_\mu W_\nu^i - \partial_\nu W_\mu^i + g\epsilon^{ijk} W_\mu^j W_\nu^k \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \end{aligned} \tag{2.8}$$

where ϵ^{ijk} is the total antisymmetric Levi-Civita tensor. The non-abelian nature of the weak interaction can be seen precisely on this term $g\epsilon^{ijk} W_\mu^j W_\nu^k$, where the couplings between gauge fields is described.

The complete electroweak Lagrangian is then:

$$\mathcal{L}_\mathrm{EW} = \sum_{f=l,q} \bar{\psi}_f (i\gamma^\mu D_\mu - m)\psi_f + \mathcal{L}_\mathrm{K} \tag{2.9}$$

where the summation includes all existent fermions.

A crucial aspect of the electroweak interaction is the mass difference of the gauge bosons. Experimental observations have proven the photon to be massless, while the Z and W bosons are in fact massive. This directly reflects on the distinct ranges of both forces: the electromagnetic interaction is long-ranged due to its massless propagator, whereas the weak interaction is a short-ranged field. In order to account for the non-zero bosonic masses, additional mass terms should be added to the Lagrangian. However, the direct introduction of a mass term, such as $\frac{m^2}{2} W_\mu W_\mu$, would violate the local $SU(2)_L$ gauge invariance of the Lagrangian, due to the gauge transformation properties of the vector fields. Since breaking gauge invariance would spoil the renormalisability of the SM, a spontaneous symmetry

breaking mechanism was introduced, which generates non-zero masses while maintaining the theory renormalisability. This mechanism, explained in Section 2.1.3, will also account for the mixture of electroweak gauge boson fields needed to obtain the observed physical fields.

The Strong Interaction

Quantum Chromodynamics characterises the strong interactions in the SM by describing them within the underlying $SU(3)_c$ symmetry and introducing a new quantum number, the colour charge c. There are three different possible colour states and only colour charged particles can interact via the strong force. Quarks transform under the fundamental representation of the strong $SU(3)_c$ gauge group and are therefore represented in colour triplets. Leptons, however, are represented in colour singlets, since they have no colour charge and are thus invariant under $SU(3)_c$ transformations.

To guarantee that the SM Lagrangian is locally invariant under $SU(3)_c$ transformations of the fermion fields, the QCD covariant derivative is then given by:

$$D_\mu = \partial_\mu - ig_s \frac{\lambda_\alpha}{2} G_\mu^\alpha \,, \tag{2.10}$$

where the index $\alpha = 1, ..., 8$ runs over the eight gluon fields, G_μ^α, and the eight generators of the symmetry group, the Gell-Mann matrices λ_α. These generators satisfy the Lie Algebra:

$$[\lambda_\alpha \, \lambda_\beta] = i f^{\alpha\beta\gamma} \lambda_\gamma \,, \tag{2.11}$$

with $f^{\alpha\beta\gamma}$ the completely antisymmetric structure constants of the $SU(3)$ group. Under $SU(3)$, given the non-abelian nature of the symmetry group generators, the gluon gauge fields form a colour octet, thus being able to self-interact. The coupling constant g_s represents the strength of the QCD interactions,[1] varying with the energy of the strongly interacting particles. This is known as the QCD asymptotic freedom: g_s becomes weaker as the energy of the interacting particles increases. The colour confinement is a direct result of this property of the QCD coupling.

The complete QCD Lagrangian, including an additional kinematic term for the gluon fields, is thus given by:

$$\mathcal{L}_{\text{QCD}} = \bar{q}(i\gamma^\mu D_\mu)q - \frac{1}{4} G_{\mu\nu}^\alpha G_{\mu\nu}^\alpha \tag{2.12}$$

[1] The coupling constant is usually referred as $\alpha_S \equiv {g_s}^2/4\pi$.

where q is a vector of the three possible colour states of a specific quark type and $G^{\alpha}_{\mu\nu}$ is the field tensor given by:

$$G^{\alpha}_{\mu\nu} = \partial_{\mu}G^{\alpha}_{\nu} - \partial_{\nu}G^{\alpha}_{\mu} - g_s f_{\alpha\beta\gamma}G^{\beta}_{\mu}G^{\gamma}_{\nu}\,, \tag{2.13}$$

While the covariant derivative D_{μ} (Equation 2.10) describes the interactions between quarks and gluons, the kinetic term with the field tensor $G^{\alpha}_{\mu\nu}$ product describes the gluon self-interaction, including cubic and quartic gluon self-interactions.

2.1.3 The Brout-Englert-Higgs Mechanism

The SM Lagrangian, previously derived in Section 2.1.2 is an incomplete description of nature: it does not explain the different observed masses of the gauge fields and fermions; instead it is only comprised of massless objects. A spontaneous symmetry breaking process can answer the apparent inconsistency between having massive particles and the Lagrangian gauge invariance.

The relativistic mechanism that generates the masses of the gauge bosons via spontaneous symmetry breaking, also known as the Brout-Englert-Higgs Mechanism or Higgs Mechanism, is a renormalisable model which requires the introduction of a scalar $SU(2)$ doublet field:

$$\Phi = \begin{pmatrix} \phi^{+} \\ \phi^{0} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}\,, \tag{2.14}$$

with electrically charged (ϕ^{+}) and neutral (ϕ^{0}) complex scalar components.

In the Glashow-Salam-Weinberg model, the Higgs mechanism is set in the electroweak sector of the SM, breaking down the $SU(2)_L \times U(1)_Y$ local gauge symmetry to a $U(1)_{EM}$ symmetry. The SM Lagrangian corresponding to this scalar field can thus be written, using the covariant derivative in Equation 2.2, as:

$$\mathcal{L}_H = (D^{\nu}\Phi)^{\dagger}(D_{\nu}\Phi) - V(\Phi)\,, \tag{2.15}$$

where $V(\Phi)$ is called the Higgs potential, defined as:

$$V(\Phi) = \mu^2(\Phi^{\dagger}\Phi) + \lambda(\Phi^{\dagger}\Phi)^2\,. \tag{2.16}$$

The two parameters μ^2 and λ, regarding a mass term and the field's self-interaction strength respectively, will define the properties of the potential. For a physically meaningful solution λ cannot be negative, which would otherwise lead to having no stable minima. So, in the case of $\lambda > 0$, there are two different possibilities for the potential's shape depending on whether

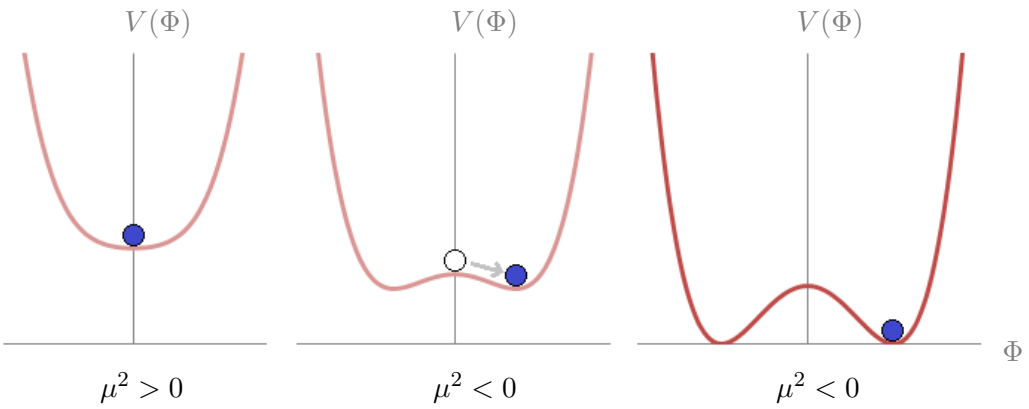


Figure 2.1: Illustrative 2D diagram of the spontaneous symmetry breaking mechanism [30].

μ^2 is positive or negative, as shown in the two-dimensional illustration in Figure 2.1. When $\mu^2 > 0$, the potential has a single minimum (or ground state), called the vacuum expectation value (VEV), corresponding to $|\Phi| = 0$. For $\mu^2 < 0$, as shown in the distributions on the right, the potential no longer has a single minimum, but instead an infinite set of possible minima (or ground states) satisfying:

$$\Phi^\dagger \Phi = -\frac{\mu^2}{2\lambda} = \frac{v^2}{2}\,, \tag{2.17}$$

which means that there is a non-zero vacuum expectation value, $\langle \Phi \rangle = (0, v/\sqrt{2})$. The minimum of the potential can then be represented as a circle of minima in the ϕ^+-ϕ^0 complex plane,

$$(\phi^+)^2 + (\phi^0)^2 = v^2, \tag{2.18}$$

as shown in Figure 2.2. The non-unique vacuum state is no longer invariant under $SU(2)_L \times U(1)_Y$, meaning that these two symmetries are broken. Although $V(\Phi)$ is symmetric under gauge transformations, the choice of the vacuum state breaks the symmetry of the Lagrangian, *i.e.* spontaneous symmetry breaking occurs. Whenever a continuous symmetry is broken, according to the Goldstone theorem, massless scalar bosons (Goldstone bosons) necessarily appear [32, 33]. These bosons can, however, be absorbed by a gauge field as a longitudinal polarisation component, resulting in a massive gauge boson. When choosing the vacuum state, the experimentally observed properties of the gauge bosons must, of course, be respected. Since the electrically neutral photon is the only massless electroweak gauge boson, only the neutral scalar field component of the Higgs potential, ϕ^0, should have a

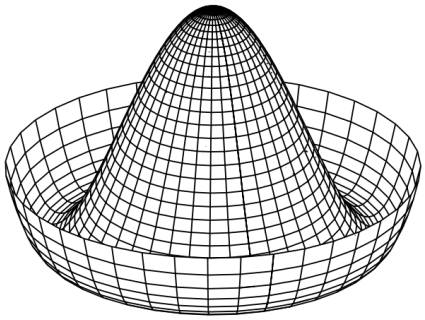


Figure 2.2: Typical shape of the vacuum potential for $\lambda > 0$ and $\mu^2 < 0$ in the ϕ^+-ϕ^0 complex plane [31].

non-zero vacuum expectation value:

$$\phi^0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \tag{2.19}$$

The fields can then be expanded around the true minimum of the theory:

$$\phi_v = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}, \tag{2.20}$$

where $H(x)$ represents the quantum fluctuations, or excitations, around the vacuum state. This is commonly known as the unitary gauge choice. As particles can be described as excitations of fields, $H(x)$ must correspond to an adicional electrically neutral spin-0 particle: the Higgs boson, H.

By choosing this unitary gauge, the Higgs doublet is now constricted to a single degree of freedom:

$$\phi_1 = \phi_2 = \phi_4 = 0, \qquad \phi_3^2 = v + H(x)\,. \tag{2.21}$$

The other three degrees of freedom were 'absorbed' during the unitary gauge choice, giving rise to the longitudinal degrees of freedom of the $W^{\pm}$ and Z bosons. This can be seen by explicitly writing the covariant-derivative terms of the Lagrangian (Equation 2.15) for the chosen gauge:

$$\begin{aligned} (D^\mu \Phi_\nu)^\dagger (D_\mu \Phi_\nu) &= \frac{1}{2}(\partial_\mu H)(\partial^\mu H) + \frac{1}{8}(v+H)^2 g^2 (W_\mu^1 + iW\mu^2)(W^{1\mu} - iW^{2\mu}) \\ &\quad + \frac{1}{8}(v+H)^2 (g' B_\mu - g\mu^3)(g' B^\mu - g W^{3\mu})\,. \end{aligned} \tag{2.22}$$

From the Higgs Lagrangian, including the explicit form of Higgs potential in Equation 2.16, the mass of the H boson can be identified as:

$$m_H = v\sqrt{2\lambda}\,. \tag{2.23}$$

It is important to notice that m_H does not have a SM predicted value, since the value of λ is unknown. It must then be determined experimentally.

The mass terms for the electroweak gauge bosons can also be derived from the following term of the Higgs Lagrangian (Equation 2.15) in the unitary gauge:

$$\begin{aligned}
&\Big|\Big(-ig\frac{\sigma_a}{2}W_\mu^a - i\frac{g'}{2}B_\mu\Big)\Phi\Big|^2 = \\
&= \frac{1}{8}\left|\begin{pmatrix} gW_\mu^3 + g'B_\mu & g(W_\mu^1 - iW_\mu^2) \\ g(W_\mu^1 + iW_\mu^2) & -gW_\mu^3 + g'B_\mu \end{pmatrix}\begin{pmatrix} 0 \\ v \end{pmatrix}\right|^2 \\
&= \frac{1}{8}v^2g^2((W_\mu^1)^2 + (W_\mu^2)^2) + \frac{1}{8}v^2(g'B_\mu - gW_\mu^3)(g'B^\mu - gW^{3\mu}) \\
&= \Big(\frac{1}{2}vg\Big)^2(W_\mu^+)(W^{-\mu}) + \frac{1}{8}v^2\begin{pmatrix} W_\mu^3 & B_\mu \end{pmatrix}\begin{pmatrix} g^2 & -gg' \\ -gg' & g'^2 \end{pmatrix}\begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix},
\end{aligned} \tag{2.24}$$

where $W_\mu^\pm = (W_\mu^1 \mp iW_\mu^2)/\sqrt{2}$ is as defined in Equation 2.6. The mixture of electroweak gauge boson fields needed to obtain the observed physical fields can be obtained from the diagonalisation of the mass matrix (last term of Equation 2.24). The mass eigenstates are thus obtained as function of the W_μ^3 and B_μ fields:

$$\begin{aligned}
&\frac{1}{8}v^2\Big(g^2(W_\mu^3)^2 - 2gg'W_\mu^3B_\mu + g'^2(B_\mu)^2\Big) = \\
&= \frac{1}{8}v^2\Big(gW_\mu^3 - g'B_\mu\Big)^2 + 0\Big(g'W_\mu^3 + gB_\mu\Big)^2 \\
&= \frac{1}{2}\Big(v\frac{g^2+g'^2}{2}\Big)^2 Z_\mu^2 + 0\cdot A_\mu^2\,,
\end{aligned} \tag{2.25}$$

where the fields Z_μ and A_μ result from orthogonal linear combinations of the fields W_μ^3 and B_μ and are associated to the Z boson and the photon respectively (cf. Equation 2.6).

The tree level predictions for the masses of the gauge bosons can then be inferred from the previous equations:

$$m_W = \frac{vg}{2}\,, \qquad m_Z = \frac{v\sqrt{g^2+g'^2}}{2} \quad \text{and} \quad m_\gamma = 0\,. \tag{2.26}$$

To summarise, after the electroweak spontaneous symmetry break the photon remains massless, while the W, Z and H boson masses are all directly proportional to v, the vacuum

expectation value, measured to be around 246 GeV [19]. The electroweak mixing angle, θ_W, can be measured directly from the ratio between the masses of the W and Z gauge bosons

$$\frac{m_W}{m_Z} = \cos\theta_W \,, \tag{2.27}$$

making its experimental observation an important argument in favour of the Higgs mechanism.

The Yukawa Terms

The spontaneous symmetry breaking of the $SU(2)_L \times U(1)_Y$ electroweak gauge group of the SM also generates the masses of the fermions, since there is nothing preventing the coupling of the Higgs doublet to the fermion fields. Even though the interaction between the scalar doublet and the fermion fields does not arise from the electroweak covariant-derivative previously outlined, the masses of fermions can be introduced in the form of the Yukawa-type Lagrangian:

$$\mathcal{L}_Y = -\sum_{f=l,q} y_f [\bar{f}_L \Phi f_R + \bar{f}_R \bar{\Phi} f_L] \,. \tag{2.28}$$

Here, the Yukawa coupling constant y_f is in fact a set of matrices describing the Yukawa couplings between the Higgs doublet and the fermions.

These extra Yukawa terms also give rise to fermion mass terms:

$$m_f = y_f \frac{v}{\sqrt{2}} \,, \tag{2.29}$$

where f stands for all the SM fermions. The Yukawa couplings can then be determined from the fermion observed masses and the vacuum expected value. For instance, the top quark with $m_t = 172.69\,\text{GeV}$ has a Yukawa coupling close to unity, which suggests a close interrelation of the top quark to the spontaneous electroweak symmetry breaking.

The Yukawa Lagrangian $\mathcal{L}_Y$ can then be added to the SM Lagrangian:

$$\mathcal{L}_{SM} = \mathcal{L}_{EW} + \mathcal{L}_{QCD} + \mathcal{L}_H + \mathcal{L}_Y \,. \tag{2.30}$$

2.1.4 The Cabibbo-Kobayashi-Maskawa Quark Mixing Matrix

Similarly to gauge bosons, the physically observable states of the propagating fermions (mass eigenstates) are different from the weak-interaction eigenstates that appear in the SM Lagrangian. Indeed, the weak-interaction and mass states can be connected by considering

a superposition of eigenstates. For simplicity, quarks and leptons can be treated separately, as can be seen from the Yukawa interactions with the Higgs field in Equation 2.28.

The quark sector of the Yukawa coupling y can thus be written in terms of the three generations of fermions:

$$\mathcal{L}_Y = -y^d_{ij}\overline{q_{Li}}\begin{pmatrix}\phi^+\\ \phi^0\end{pmatrix}d_{Rj} - y^u_{ij}\overline{q_{Li}}\epsilon\begin{pmatrix}\phi^+\\ \phi^0\end{pmatrix}^* u_{Rj} + \text{h.c.}\,, \tag{2.31}$$

where summations over the generation labels i and j are implied, q_L are the left-handed quark doublets, and d_R and u_R are the right-handed down and up-type quark singlets respectively, in the weak-eigenstate basis. It is clear that several gauge-invariant terms are allowed, including those with quark mixing among the different families.

The mass terms for quarks (Equation 2.29) are obtained after the spontaneous symmetry breaking, when Φ acquires a vacuum expectation value $\langle\Phi\rangle = (0, v/\sqrt{2})$:[1]

$$\begin{aligned}\mathcal{L}_Y &= -y^d_{ij}\overline{d_{Li}}\tfrac{v}{\sqrt{2}}d_{Rj} - y^u_{ij}\overline{u_{Li}}\tfrac{v}{\sqrt{2}}u_{Rj} + \text{h.c.}\\ &= -\overline{d_{Li}}M^d_{ij}d_{Rj} - \overline{u_{Li}}M^u_{ij}u_{Rj} + \text{h.c.}\,,\end{aligned} \tag{2.32}$$

where M^d_{ij} (M^u_{ij}) is the mass matrix of down-type (up-type) quarks. There is no *a priori* reason for these matrices to be diagonal, *i.e.* to have states with a defined mass. Nonetheless, the mass matrices can be diagonalised by making use of four unitary matrices, $V^{u,d}_{L,R}$:

$$M^{u,d}_{\text{diag}} = V^{u,d}_L M^{u,d}_{ij} V^{u,d}_R = V^{u,d}_L y^{u,d}_{ij} V^{u,d}_R \frac{v}{\sqrt{2}}\,. \tag{2.33}$$

In the mass eigenstates basis, the sector of the quarks coupling to the W boson in the electroweak Lagrangian (Equation 2.9) can be given by:

$$\mathcal{L}_{CC} = \frac{g}{\sqrt{2}}\overline{u'_{Li}}(V^u V^{d\dagger})_{ij}\gamma^\mu W^-_\mu d'_{Lj} + \frac{g}{\sqrt{2}}\overline{d'_{Li}}(V^d V^{u\dagger})_{ij}\gamma^\mu W^+_\mu u'_{Lj}\,, \tag{2.34}$$

where quark mixing terms have emerged as a result. $\mathcal{L}_{CC}$ is known as the charged current interaction of the left-handed quarks in the SM Lagrangian, reflecting the V-A (Vector minus Axial-vector) structure of the coupling to the W boson. Here, d' and u' represent the physical mass eigenstates of the quarks.

The Cabibbo-Kobayashi-Maskawa (CKM) matrix [34, 35], $V_{\text{CKM}} \equiv (V^u V^{d\dagger})_{ij}$ is a 3×3 unitary matrix which relates the weak and the mass eigenstates of quarks:

$$\begin{pmatrix} d\\ s\\ b\end{pmatrix}_L = \begin{pmatrix} V_{ud} & V_{us} & V_{ub}\\ V_{cd} & V_{cs} & V_{cb}\\ V_{td} & V_{ts} & V_{tb}\end{pmatrix}\begin{pmatrix} d'\\ s'\\ b'\end{pmatrix}_L . \tag{2.35}$$

[1] The charge-conjugated of the Higgs field is $\widetilde{\Phi} = \epsilon\Phi^*$.

The $|V_{ij}|^2$ can be interpreted as the probability of a quark i to transition to another quark j via the weak interaction. In fact, the CKM matrix introduces flavour-changing charged currents that describe the possible coupling of each up-type quark with every down-type quark.[1]

In a Standard Model with three generations, this unitary matrix has four independent parameters corresponding to three quark-mixing angles and one complex phase:

$$V_{\mathrm{CKM}} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta_{13}} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta_{13}} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta_{13}} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta_{13}} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta_{13}} & c_{23}c_{13} \end{pmatrix}. \tag{2.36}$$

This is known as the standard parametrisation [36] with mixing angles $c_{ij} = \cos\theta_{ij}$ and $s_{ij} = \sin\theta_{ij}$, where θ_{12} is also known as the Cabibbo angle. The phase δ_{13}, widely known as the Kobayashi-Maskawa phase, accounts for all Charge Parity (CP) violation in flavour-changing processes in the SM.[2] The unitarity of the CKM matrix imposes $\sum_i |V_{ij}|^2 = \sum_j |V_{ij}|^2 = 1$, for each i, j generation of quarks. Additionally, the orthogonality relations are given by $\sum_k V_{ik}V_{jk}^* = 0$ and $\sum_k V_{ki}V_{kj}^* = 0$ for any $i \neq j$.

The absolute magnitudes of the CKM-matrix elements can be determined from the combination of several dedicated experiments. For instance, $|Vtb|$ can be obtained from the single-top-quark production cross-section measurement. The current most precise magnitudes of the CKM elements are [19]:

$$V_{\mathrm{CKM}} = \begin{pmatrix} 0.97435 \pm 0.00016 & 0.22500 \pm 0.00067 & 0.00369 \pm 0.00011 \\ 0.22486 \pm 0.00067 & 0.97349 \pm 0.00016 & 0.04182^{+0.00085}_{-0.00074} \\ 0.00857^{+0.00020}_{-0.00018} & 0.04110^{+0.00083}_{-0.00072} & 0.999118^{+0.000031}_{-0.000036} \end{pmatrix}, \tag{2.37}$$

where a global fit to all available measurements was used, assuming three generation unitarity in the SM. It is particularly interesting to note that $|V_{ts}|$ and $|V_{td}|$ are much smaller than $|V_{tb}|$. Since the $t \to Wb$ vertex transition is proportional to $|V_{tb}|^2$ at leading order, the top quark will almost exclusively decay into a W boson and a bottom quark.

In the leptonic sector, the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix describes the lepton mixing analogously to the CKM matrix[37, 38, 39]. Unlike the CKM matrix, however, the PMNS matrix unitarity is not guaranteed, since it directly depends on the origin of neutrino masses [40].

[1] Interestingly, when considering neutral currents in the mass eigenstates, instead of the charged ones in Equation 2.34, the diagonalisation of the fermion fields leads to $V^{d\dagger}V^d = V^{u\dagger}V^u = 1$. Therefore, no flavour-changing neutral currents (FCNCs) are allowed in the SM, at least at tree-level.

[2] CP symmetry occurs when the laws of physics remain the same after converting a particle to its antiparticle while simultaneously mirroring the space coordinates. In nature, however, this symmetry is violated.

2.1.5 Successes and Shortcomings of the SM

The SM is a very successful theory, accurately explaining a wide range of phenomenological observations and meticulously tested in multiple experiments. Throughout the years, the SM has predicted the existence of new particles, all of which have been discovered and nicely accommodated into the model. It is precisely in the capability to adapt to new experimental discoveries that the beauty of the SM relies, having the development of the theory intertwined with the experimental findings. Some of the most remarkable successes of the Standard Model include:

The Charm Quark Discovery Postulated in the GIM Mechanism [41] by Glashow, Iliopoulos and Maiani as a fourth quark weak eigenstate that would couple to the strange weak eigenstate. Its existence in the SM would explain the suppression of flavour changing neutral current (FCNC) processes at one-loop level, whilst still favouring flavour changing processes with charged currents. The charm quark was directly observed for the first time in J/Ψ meson ($c\bar{c}$) decays in 1974, discovered independently by two teams led by Ting and Richter [42, 43] respectively at Brookhaven National Laboratory and SLAC.

The τ-lepton Discovery Discovered by Perl and his team through a series of experiments between 1974-77 at SLAC-LBL [44]. The τ-lepton was indirectly discovered by a significant excess of events of the type $e^+ + e^- \rightarrow e^\pm + \mu^\mp +$ at least 2 undetected particles, which lacked any conventional explanation at the time. No other muons, electrons, photons, or hadrons were detected and in order to have conservation of energy and momentum at least two extra undetected particles were necessary. Previously suggested by Zichichi and others at CERN [45, 46] in the 1960s and anticipated by Tsai [47] in 1971, the idea of a third sequential and heavier lepton was then considered to explain these peculiar events: $e^+ + e^- \rightarrow \tau^+ + \tau^- \rightarrow e^\pm + \mu^\mp + 4\nu$. The discovery of the τ-lepton spoiled the lepton and quark symmetry introduced by the GIM mechanism, which lead to the speculation of a fifth and sixth quark flavours. The first experimental evidence of the tau neutrino ν_τ was reported in 2000 by the DONUT collaboration at Fermilab [48].

The Bottom Quark Discovery First described in 1973 by Kobayashi and Maskawa to explain CP violation in the SM [35]. The quark mixing in the weak interaction was

extended from Cabibbo's initial proposal to account for the CP violation, directly predicting a third quark family comprised of heavier quarks. The subsequent discovery of the Υ meson ($b\bar{b}$) in 1977 by a team led by Lederman at Fermilab [49], unveiled the existence of the bottom quark and led to the speculation about its counterpart, the top quark.

The *W* and *Z* bosons Discoveries In the late 1960s, Glashow, Weinberg and Salam's unified electroweak theory [6, 7, 8] proposed the existence of two weak gauge bosons: the charged $W^{\pm}$ bosons, necessary to explain beta decay; and the neutral Z boson, previously unprecedented. The first experimental evidence of neutral weak interactions occurred a few years later, in 1973, in the Gargamelle bubble chamber [50] at CERN. The tracks of a few electrons were photographed moving' as neutrinos interacted with these leptons via the exchange of an unseen Z boson, altering the electron's momenta. During 1983, the first direct observations of the W [51] and Z [52] bosons were recorded at CERN in the UA1 and UA2 experiments led by Rubbia and Darriulat respectively. Accelerated in the Super Proton Synchrotron (SPS), proton and antiproton beams collided with $\sqrt{s} = 540\,\text{GeV}$ centre-of-mass energy in both UA1 and UA2, providing sufficient energy for the production of the weak bosons. This was only possible due to the work of Van der Meer in the development of stochastic cooling for proton accelerators [53]. The discovery of the $W^{\pm}$ and Z bosons with masses within the theoretical predictions, provided a stringent test of the electroweak unification in the SM.

The Top Quark Discovery Predicted in 1973 by Kobayashi and Maskawa to complement the bottom quark in the third weak isospin doublet, its discovery came in 1995 at the Fermilab by the Tevatron experiments, CDF and DØ [54, 55]. The inference of top quark mass was facilitated due to the theoretical and phenomenological studies developed by 't Hooft and Veltman in the renormalisation of non-abelian quantum field theories [27]. Their work provided a computing method for quantum corrections to several processes, allowing precision calculations of physical quantities and direct comparison with the experimental observations.

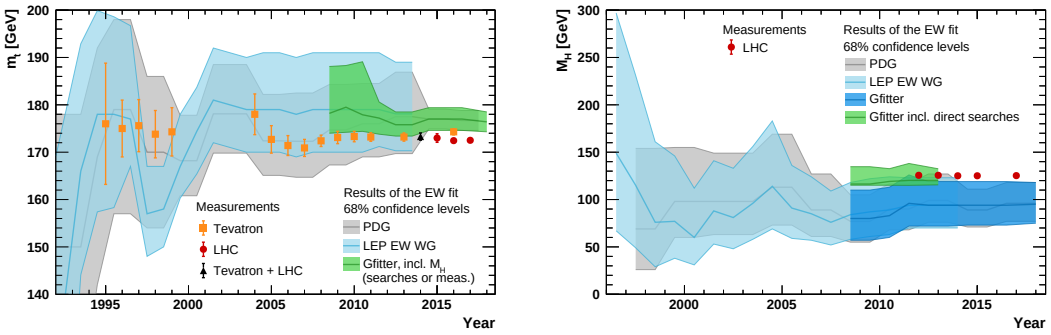
The Higgs Boson Discovery Arising from the Brout-Englert-Higgs mechanism, proposed independently by Engler and Brout [10], Higgs [11], and Guralnik, Hagen and Kibble [12] in 1964 to explain the different fermion and boson masses. The experimental

discovery of the Higgs boson was reported in 2012 by the LHC experiments ATLAS [25] and CMS [26] at CERN. Data subsequently collected at the LHC shows that the observed particle has spin zero, interacting and decaying in several of the ways predicted for the Higgs boson by the SM[56, 57]. The discovery of a Higgs boson, which is a direct manifestation of the existence of the Higgs field, constitutes yet another validation of the SM theory of particle physics. Whether the observed boson's nature and properties continue to behave in line with the SM predictions for the Higgs boson, or not, is now currently under further investigation.

Ultimately, the SM is an *ad hoc* model of theoretical ideas assembled together to reproduce the experimental findings, with free parameters[1] chosen to match the observations rather than originating from pure theoretical principles. The SM has 25 free parameters (or 26 if the neutrinos are not normal Dirac fermions): 14 are associated with the Higgs field (the fermion masses from the Yukawa couplings y_f, the Higgs mass and the vacuum expectation value); 8 with the flavour sector (the angles and phase of both quark and lepton mixing) and 3 with the gauge interactions (the coupling constants g, g' and g_s).

Throughout the years, the validity of the SM has been experimentally confirmed with precision measurements of several observables, as well as complex global fits of all its free parameters. The increasing high precision results from a variety of experiments (some of which within 0.1% precision) contributed to the constraining of searches and provided cross-check tests to the SM [58]. The discovery of the top quark, and later the Higgs boson, with mass values in agreement with the predictions from global fits impressively confirmed the SM at the quantum level. Figure 2.3 depicts the historical development of the global fit constraints and direct mass measurements as a function of time for the top quark (a) and the Higgs boson (b), respectively. The self-consistency of the SM can therefore be probed by relying on precise measurements of the SM parameters. One such test can be seen in Figure 2.4, where all relevant electroweak observable measurements are combined into a statistical model, while certain measurements are left free to float within their uncertainty to allow for a fit among multiple correlated measurements. Here, a least-square χ^2 fit [19] is shown for the Higgs boson mass M_H as a function of the top quark mass m_t including all relevant electroweak data with the exception of the ATLAS and CMS combined measurement of the Higgs boson mass. The SM global fit contour (represented in red) agrees well

[1] "Free", in this context means that the values of values of these parameters do not have theoretical projections or strict expected bounds.



(a) (b)

Figure 2.3: The 68% confidence level prediction bands of (a) the top quark mass and (b) the Higgs boson mass throughout the years, using electroweak precision data in various EW fits[58]. The direct mass measurements are shown by the data points.

with the direct M_H measurements (shown in the yellow dashed line), describing the data with a $\chi^2/\text{d.o.f.} = 40.8/41$. There seems to be, however, a strong constraint imposed by the direct precision data on the allowed top quark mass value, which is apparently favoured to be closer to the upper end of its 1σ uncertainty band. Other observables also present some difference between data and the SM fit,[1] but all still quite small, with only $g_\mu - 2$ showing a larger deviation of the order of 3.1σ. Despite the few discrepancies, it is remarkable how accurately the Standard Model is able to describe the electroweak data, continuously proving to be a self-consistent theory.

Regardless of the remarkable successes of the SM, there are still patterns amongst its different parameters that suggest the presence of some unknown symmetry principle. For instance, the fermions (excluding the neutrinos) seem to follow a yet unexplainable pattern of similar mass values within a single generation. The framework does not yet explain a number of theoretical and experimental evidences, which suggests that the SM is in fact an effective theory — perhaps a low energy version of a wider theory describing the full spectrum of observations. In fact, a Grand Unified Theory (GUT) of the forces appears plausible, since there is a similar order of magnitude between the coupling constants of the three gauge interactions, alluding that they might be the result of some broken symmetry, just different manifestations of the same force. Patterns like these, not yet accommodated

[1] For instance, using the Higgs boson mass in the global fit and not including the W boson mass and width, yields $m_W = 80.357 \pm 0.006\,\text{GeV}$, which is 1.7σ below the world average.

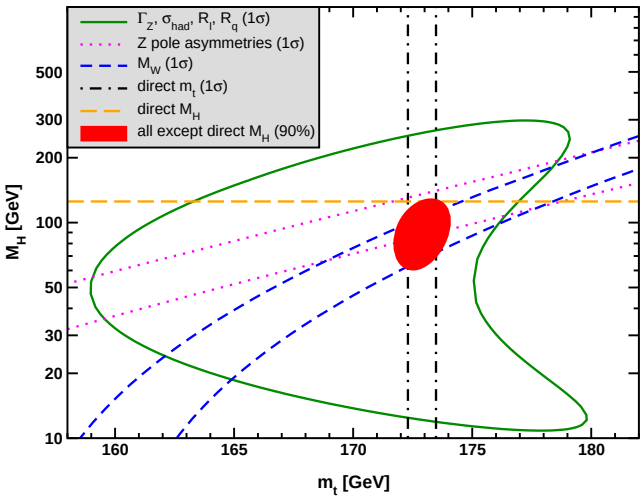


Figure 2.4: Global χ^2 fit result shown for the Higgs boson mass M_H as a function of the top quark mass m_t including all electroweak data except the direct Higgs boson mass measurement at the LHC, which is displayed in the yellow dashed line for completeness[19]. The red contour represents the 90% CL region ($\Delta\chi^2 = 4.605$) allowed by all data. The one-standard-deviation uncertainties of several inputs are also shown: the dashed lines delimit the 68% uncertainty bounds, whereas the filled lines the 39.35% uncertainty bounds.

into the framework, hint at physics beyond the Standard Model (BSM). Some of the yet unanswered questions of the Standard Model include:

Three Generations of Fermions There is no explanation for having only three fermion generations in the SM framework.

The Flavour Problem There are no predetermined theoretical values for the fermion masses, their mixing angles or CP violating phases.

The Matter/Antimatter Asymmetry Baryonic matter and antibaryonic matter should have been produced in equal amounts during the Big Bang. However, that is not what we observe in the universe. The CP-violating nature of the weak interaction in the SM can account for some asymmetry in the production rate of matter and antimatter, but it is not enough to account for the observed difference. The strong interaction does not seem to violate the CP symmetry, even with CP-violating terms allowed in the QCD Lagrangian.

Naturalness of the Higgs Boson Mass The mass of the Higgs boson is quite sensitive to top quark loop effects, since the Yukawa coupling between the H boson and the top quark is very large. These fermionic quantum corrections are quadratically divergent with the ultraviolet cut-off scale Λ_{UV} for new physics. Assuming the SM describes nature up to the Planck scale, then Λ_{UV} would be several orders of magnitude higher than the electroweak symmetry-breaking scale, which would require heavy fine-tuning of the Higgs mass. In order to cancel these fermionic corrections to the Higgs boson mass, new mechanisms or new particles at the TeV scale are considered in possible extensions of the SM, such as supersymmetry.

Neutrino Oscillations In the SM neutrinos are massless particles. The measurement of neutrino mass oscillations [23, 24], however, indicates different masses between the neutrino generations. Right-handed neutrinos or perhaps Majorana neutrinos could be an answer suitable within the SM.

Dark Matter and Dark Energy Only about 5% of the energy density in the universe is made of ordinary baryonic matter. The remaining energy density can only be accounted for considering the existence of dark matter and dark energy, which represent around 26.8% and 68.3% of the universe, respectively. The SM has no candidate particle, force or mechanism that can explain this large fraction of the energy density of the universe. Dark matter is extremely weakly interacting, but its gravitational effects on visible matter allow for an indirect probe. For instance, measurements of the rotation curves of galaxies and gravitational lensing can be understood considering the presence of dark matter. Dark energy is, however, a lot more aloof. Nonetheless, it is the most accepted justification for the acceleration of the expansion of the universe as of yet.

Gravity Einstein's theory of general relativity describes gravity as a distortion of the geometry of spacetime. Describing gravity as a quantum field theory, however, has proven to be quite a difficult task and has not yet been accomplished. While gravitational interactions between particles are assumed negligible for energies lower than the Planck scale, the high energy density of the early universe would certainly require a quantum-gravitational theory to accurately describe the fundamental particles and their interactions.

The SM is undoubtedly a successful theory, explaining and predicting a myriad of phenomena in high energy physics. Even with all the open questions, quantum field theory is until now the best way to describe particles and their interactions at the current accessible energies. Studying the SM, both where it accurately describes the experimental findings and where it fails to provide an answer, will deepen the understanding of the Universe at the most fundamental level.

2.2 The Top Quark and the Higgs Boson

By accelerating particles to very high energies and then colliding them, the SM and all of nature's fundamental elements can be thoroughly investigated. At particle colliders, the collision is just a way to force the particles to interact in a controlled environment and to register the properties of the obtained output. Typically, electrons or protons are accelerated, but heavier hadrons are also used in order to cover a wider physics spectrum. When colliding non-fundamental particles, such as the protons, at high energy the actual interacting objects are its constituent quarks and gluons. The hadron constituents, also known as partons, that undergo a significant momentum transfer constitute the hard-scattering system. In collisions involving hadrons, the full initial state of the hard-scattering system is unknown, since the experimental setup can only control the momentum of the initial beam of particles. The specific parton details are unknown *a priori*, and its measurement unfeasible without interrupting or ruining the interaction. Despite the unknown initial state of the hard-scattering system, a full consistent analysis can be done by using form-factor descriptions of the colliding hadrons that account for the fraction of momentum carried by the partons. These parton distribution functions (PDF) $f_i(x_i, {\mu_\mathrm{F}}^2)$ represent the probability density of a parton flavour i to have a longitudinal momentum fraction x_i within the accelerated hadron at the energy scale ${\mu_\mathrm{F}}^2$. Since the PDFs must reflect the underlying structure of the colliding hadron, the parton i can be one of the valence quarks/antiquarks or instead originate from the virtual hadronic sea, being in this case either a gluon or another quark/antiquark. The factorisation scale μ_F is the scale/threshold that separates the hard scattering regime in the PDF. The PDF dependence on an energy scale μ_F is described by the DGLAP equations, with calculations relying on perturbation theory and higher order corrections [59]–[62]. The functional form of the PDFs cannot, however, be fully computed from theory since they depend on low energy QCD processes. QCD has such a large coupling constant, therefore

not allowing perturbation theory to be applied, and so the dependence of the PDFs on x must be extracted from measurements of structure functions in other deep inelastic scattering experiments (see Section 5.2 for further details).

For collisions of two hadrons a and b like those performed at the LHC experiments, the inclusive hadron-hadron cross-section $\sigma_{ab\to X}$ can be obtained using the factorisation theorem,

$$\sigma_{ab\to X} = \sum_{i,j=q,\bar{q},g} \int dx_i dx_j f_i(x_i, {\mu_{\mathrm{F}}}^2) f_j(x_j, {\mu_{\mathrm{F}}}^2) \sigma_{ij\to X}(x_i, x_j, \alpha_s(\mu_{\mathrm{R}}^2)) \,, \tag{2.38}$$

where X is any possible final state. Since all possible parton-parton scattering processes must be considered, the inclusive cross-section is then the convolution of parton distribution functions and a partonic cross-section $\sigma_{ij\to X}(x_i, x_j, \alpha_s(\mu_{\mathrm{R}}^2))$ of the hard scattering process between two partons i and j. The hard scattering process $\sigma_{ij\to X}$ can be computed perturbatively from Feynman diagrams and depends on the strong coupling constant α_S and its renormalisation scale μ_{R}. The cross-section of a hard process can then be written from the scattering matrix between the initial and final state particles as a perturbative series,

$$\sigma_{ij\to X} = \sigma_{LO} \left(1 + \left(\frac{\alpha_S}{2\pi}\right) \sigma_1 + \left(\frac{\alpha_S}{2\pi}\right)^2 \sigma_2 + ...\right) \,, \tag{2.39}$$

where the leading order (LO) accuracy includes the minimal number of interacting vertices possible for the specific process. Next-to-leading order (NLO), or next-to-next-to-leading order (NNLO) accuracy in α_S can be obtained depending on where the power series expansion of the cross-section is truncated, that is the amount of vertices and loops included in the process. In theory, if the cross-section $\sigma_{ij\to X}$ is calculated to all orders, it is invariant under changes in these parameters. In reality, however, the truncated calculations/series deem it necessary to artificially choose the two scales in order to make the most accurate cross-section predictions. The numerical results will be different depending on the order of the corrections used — the more terms included in the perturbative expansion, the more exact the prediction — reflecting the uncertainty due to the unknown higher order corrections. μ_{F} and μ_{R} are usually set to the same value, typically the order of the energy scale of the hard scattering process, in order to avoid the reappearance of logarithmic divergences in the perturbation series [63].

Equation 2.38 only accounts for the most energetic interactions: the hard scattered system. The partons not taking part in the hard scattering process give rise to the underlying

event, that is the elastic or soft inelastic interactions between the remaining partons. These are also known as 'minimum bias' events. In this thesis, proton-proton collisions at the LHC are analysed with dedicated studies of both soft underlying processes as well as several hard processes, such as those involving top quarks or Higgs bosons.

2.2.1 Top Quark Physics

According to the SM, the top quark t, with electric charge $Q = +(\frac{2}{3})|e|$ and isospin $T_3 = +\frac{1}{2}$, is the up-type partner of the bottom quark in the left-chiral weak isospin doublet or the third generation right-chiral quark in the $SU(2)_L$ singlet (see Equation 2.5). The top quark is the heaviest known fermion (see Table 2.2) which makes it a rather unique quark. In fact, its large mass is the cornerstone of top quark phenomenology.

The top quark Yukawa interaction can be calculated from the Lagrangian in Equation 2.28, which, after spontaneous symmetry breaking, results in the top quark Yukawa coupling being directly obtained from the top quark mass as

$$y_t = \frac{\sqrt{2}m_t}{v} = 0.993 \pm 0.03\,, \tag{2.40}$$

resulting in the highest coupling within the SM (where other Yukawa couplings are of the order of 10^{-2}).

The CKM element V_{tb} has a much larger value than V_{ts} or V_{td}, being very close to unity as seen in the CKM matrix in Equation 2.37. Consequently, the top quark decays nearly exclusively via electroweak interactions into a W boson and a b quark, $t \to W^+ + b$ (or $\bar{t} \to W^- + \bar{b}$ for the top antiquark decay). Since the top quark is the only quark with a mass bigger than that of the W boson, the top quark is then the only quark able to decay into a real W boson, *i.e.* an on shell W boson. Also due to its large mass value, the top quark has an extremely brief lifetime, $1/\Gamma_t \approx 5 \times 10^{-25}$ s[19]. The formation of bound states involving top quarks is thus not possible, and, given the typical hadronisation timescale of $1/\Lambda_{QCD} \approx 10^{-23}$ s, the strong interaction can not depolarise the spin of the top quark before its decay. By keeping its spin coherent, the top quark properties can be transferred to its decay products through the V-A coupling structure of weak interactions. This then allows the study of top quark properties, such as the top quark polarisation, via angular distributions of the decay products.

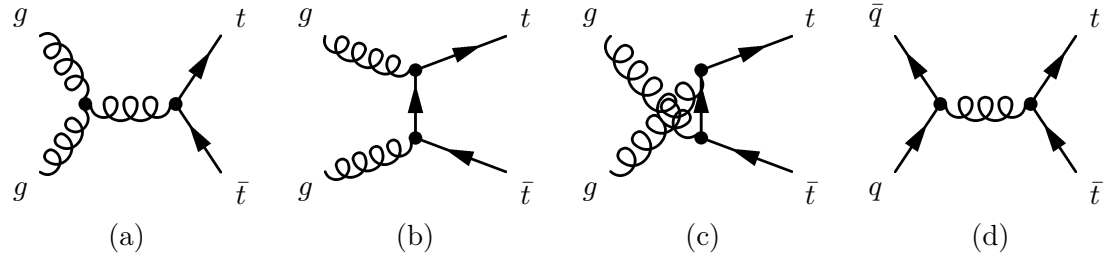


Figure 2.5: Leading-order Feynman diagrams of the dominant top quark pair production modes at the LHC: gluon fusion mechanism is shown in (a), (b) and (c) diagrams, whereas quark-antiquark annihilation can be seen in (d).

As a result of the top quark high mass, and all properties ensuing, the top quark is expected to play a special role in understanding the electroweak symmetry breaking mechanism and even possibly be a probe to new physics. The accurate knowledge of its properties is thus instrumental in providing information on fundamental interactions in the SM and beyond.

2.2.1.1 Top Quark Production

At hadron colliders, the top quark can be produced through either strong or electroweak mechanisms. The predominant processes of its production are the top quark pair production (also known as $t\bar{t}$) and the single top quark production (single top) via the strong and weak interactions, respectively. Table 2.3 summarises the theoretical cross-section predictions of both mechanisms at several centre-of-mass energies.

Pair Production

In proton collisions, the most probable top quark production mechanism is the $t\bar{t}$ pair production, occurring mainly via gluon ($gg \to t\bar{t}$) or quark fusion ($q\bar{q} \to t\bar{t}$) processes, at leading order in QCD. Due to the higher gluon density inside colliding protons, gluon-gluon fusion is the dominant mechanism at the LHC energy scale, constituting between 80% to 90% of the $t\bar{t}$ production cross-section in the LHC centre-of-mass energy $7 - 14$ TeV [19]. The different top quark pair production mechanisms are shown in the leading-order Feynman diagrams in Figure 2.5.

Several direct measurements of the $t\bar{t}$ production cross-section have been made since the discovery of the top quark and are summarised in Figure 2.6 as a function of the centre-of-

Collider	$\sqrt{s}$ [TeV]	Cross-Section $\sigma_{t\bar{t}}$ [pb]	Cross-Section $\sigma_{t+\bar{t}}$ [pb] t-channel	s-channel	Wt-channel
Tevatron	1.96	$7.164^{+0.110+0.169}_{-0.200-0.122}$	$2.02^{+0.13}_{-0.13}$	$1.03^{+0.05}_{-0.05}$	–
LHC	7	$177.31^{+4.56+9.02}_{-5.99-9.02}$	$63.89^{+1.92+2.19}_{-1.25-2.19}$	$4.29^{+0.12+0.14}_{-0.10-0.14}$	$15.74^{+0.40+1.10}_{-0.40-1.14}$
LHC	8	$252.89^{+6.39+11.67}_{-8.64-11.67}$	$84.69^{+2.56+2.76}_{-1.68-2.76}$	$5.24^{+0.15+0.16}_{-0.12-0.16}$	$22.37^{+0.60+1.40}_{-0.60-1.40}$
LHC	13	$831.76^{+19.77+35.06}_{-29.20-35.06}$	$216.99^{+6.62+6.16}_{-4.64-6.16}$	$10.32^{+0.29+0.27}_{-0.24-0.27}$	$71.7^{+1.80+3.40}_{-1.80-3.40}$
LHC	14	$984.50^{+23.21+41.31}_{-34.69-41.31}$	$248.09^{+7.58+6.98}_{-5.40-6.98}$	$11.39^{+0.32+0.29}_{-0.26-0.29}$	$84.4^{+2.00+3.00}_{-2.00-4.80}$

Table 2.3: Theoretical cross-section predictions for the $t\bar{t}$ pair production and single top quark processes for several centre-of-mass energies. Predictions for the Tevatron assumed a top quark mass of 173.3 GeV, whereas LHC predictions assumed 172.5 GeV. The uncertainties from scale dependence and from parton density functions are combined in quadrature or given separately (scale + PDF). The $t\bar{t}$ pair production cross-sections were calculated at NNLO in QCD including resummation of NNLL soft gluon terms with top++2.0 [64]–[70]. The scale uncertainty was determined from the independent variation of μ_{F} and μ_{R}, whereas the PDF and α_S uncertainties were calculated using the PDF4LHC prescription [71] with the MSTW2008 68% CL NNLO [72, 73], CT10 NNLO [74, 75] and NNPDF2.3 5f FFN [76] PDF sets. Single top quark production at Tevatron is shown with approximately NNLO order accuracy for both the t and s-channels [77, 78], while Wt events are negligible at this energy. Regarding the single top quark processes at the LHC, NLO accuracy was obtained with Hathor v2.1 [79, 80] for the cross-sections of the t and s-channels, whereas the Wt cross-section was calculated to NLO accuracy in QCD including NNLL soft-gluon corrections [81, 82]. The scale uncertainty was derived using the MSTW2008 NLO PDF set for the t- and s-channels, while the MSTW2008 NNLO PDF set was used for the Wt-channel. The PDF and α_S uncertainties were calculated using the PDF4LHC prescription with the MSTW2008 68% CL NLO, CT10 NLO and NNPDF2.3 PDF sets for the t- and s-channels. The PDF uncertainty of the Wt-production was calculated using the MSTW2008 NNLO PDF.

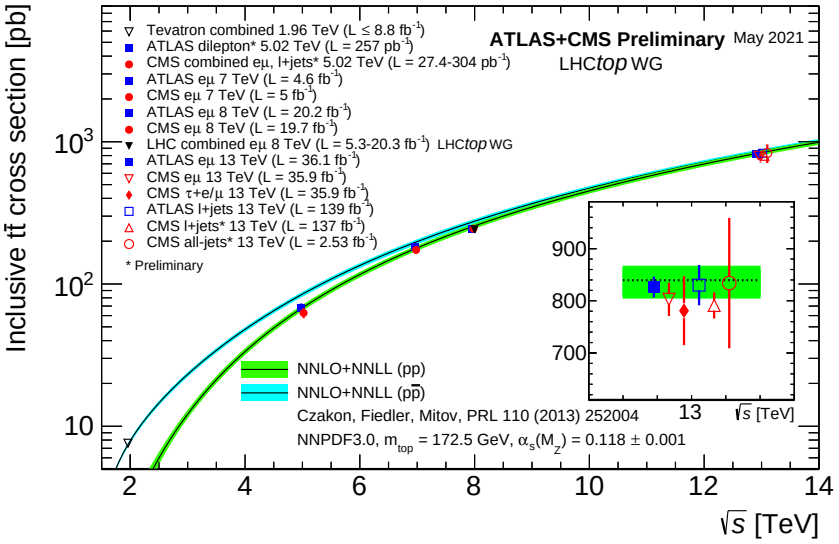


Figure 2.6: Top quark pair production cross-section measurements at the LHC and Tevatron as a function of the centre-of-mass energy $\sqrt{s}$ [83]. The measurements are compared to the NNLO QCD calculation with NNLL resummation (top++2.0 [70]), where the coloured bands represent theory uncertainties due to renormalisation and factorisation scale, parton density functions and the strong coupling. The mass of the top quark was set to $m_{top} = 172.5\,\text{GeV}$ in both experimental and theoretical results. Measurements performed at the same centre-of-mass energy are slightly offset for clarity.

mass energy $\sqrt{s}$. The SM theoretical predictions are in agreement with the experimental results from the Tevatron and LHC experiments.

Single Top Quark Production

Single top quark production occurs via the electroweak interaction in mechanisms mediated by virtual t-channel and s-channel W bosons ($qb \to q't$ and $q\bar{q}' \to t\bar{b}$ respectfully) or via the associated production of a W boson and a top quark ($bg \to W^- t$) commonly addressed as Wt-production. The leading order Feynman diagrams of the three processes are shown in Figure 2.7 for the top quark production, where the charge conjugate diagrams would represent the top antiquark production. The t-channel process can be described via a $2 \to 2$ or a $2 \to 3$ scheme, where in the latter the initial colliding b quark originates from gluon splitting.

The theoretical cross-sections for the sum of the singly produced top quarks and top antiquarks are shown in Table 2.3. At Tevatron, t-channel and s-channel single top quarks

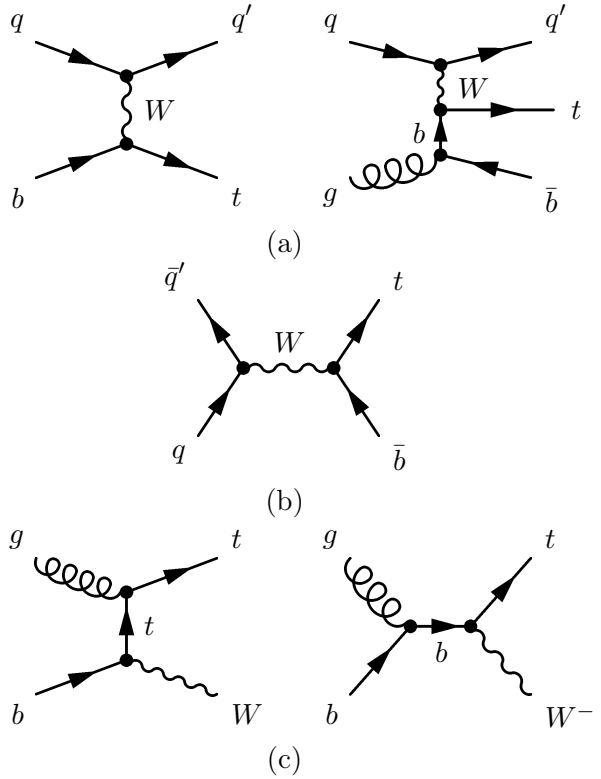


Figure 2.7: Leading-order Feynman diagrams of the dominant single top quark production mechanisms at the LHC: (a) t-channel (b) s-channel and (c) Wt-channel.

are produced at an identical rate as the top antiquarks, due to the initial parton densities in the $p\bar{p}$ collisions. In the pp collisions at the LHC, however, the charge asymmetry of the initial state partons leads to a production cross-section of single top quarks almost twice that of the top antiquarks in both t and s-channels. For Wt events, due to the density functions of the initial partons, an equal proportion of top quarks and top antiquarks is produced at the LHC. The weak coupling involved in these processes leads to smaller cross-sections than that of the top quark pair production. Among these three mechanisms, the t-channel production has the largest cross-section in proton collisions. The s-channel process cross-section is approximately the same order of magnitude as the t-channel one at the Tevatron, whereas at the LHC it is much smaller than the t-channel cross-section. Regarding the Wt production, while it has a negligible cross-section at the Tevatron energy scale, it becomes quite relevant at the LHC scale.

A summary plot of the single top quark production cross-section measurements is shown

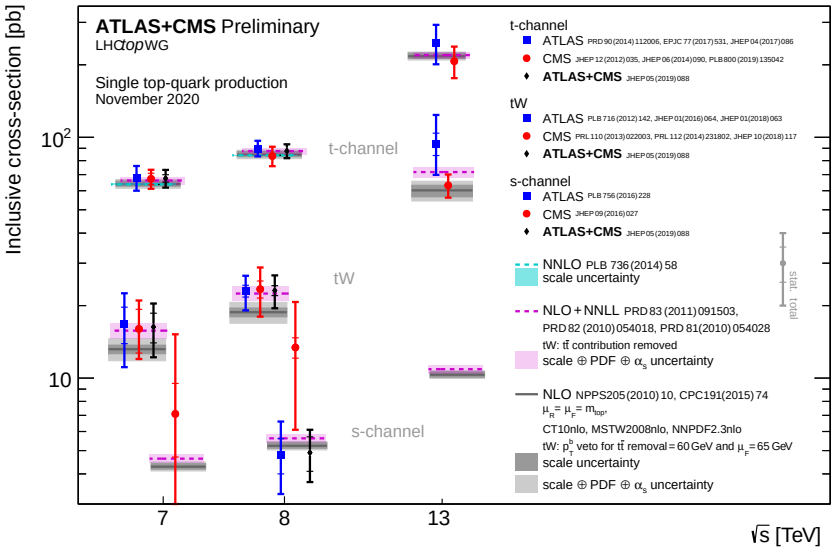


Figure 2.8: Single top production cross-section measurements at the LHC experiments in various channels as a function of the centre-of-mass energy $\sqrt{s}$ [83]. The theoretical predictions were calculated to NLO QCD, NLO QCD complemented with NNLL resummation and NNLO QCD (t-channel only).

in Figure 2.8 as a function of the centre-of-mass energy $\sqrt{s}$, where all data measurements are in agreement with the SM predictions.

The single top quark production cross-section measurement provides a direct determination of V_tb without assuming the number of quark generations. The single top quark production is particularly sensitive to new physics, such as anomalous couplings [84].

Top Quark Associated Production and Other Production Mechanisms

In addition to the previously introduced inclusive QCD and EW top quark production mechanisms, the top quark can be produced in other processes with typically much smaller cross-sections. For example, the production of a pair of top quark-antiquark can occur while simultaneously producing a SM boson. The associated production of top quark pairs with electroweak bosons, $t\bar{t}V$ where $V = \gamma, W$ or Z or with an Higgs boson, $t\bar{t}H$, are accessible processes at hadron colliders and were in fact already measured with LHC data [85]–[94]. The associated production of top quark pairs with gluons, commonly named $t\bar{t} + jets$, is usually included with the regular top quark pair production as higher order corrections. Analogously, single top quark production can happen associated with gauge bosons, such as

tZq, which was discovered in 13 TeV pp collisions at the LHC [95, 96, 97], and $tq\gamma$, which has been observed for the first time by the ATLAS experiment [98] Other more rare single top quark production processes, like $tW\gamma$, $tZ\gamma$, tHq, tHW, tHZ and tWZ, which have not yet been discovered, are interesting places to further establish the SM couplings.

Experimental evidence for processes producing four top quarks, i.e. two pairs of top quark-antiquark, was obtained by both ATLAS and CMS collaborations [99]–[103]. Such production mechanisms are quite interesting as they allow the study of the top quark-Higgs boson Yukawa coupling in a different process. This rare production process is particularly sensitive to many BSM models.

Mechanisms involving same-sign top quarks or other FCNC processes, including tZ, tH and $t\gamma$ productions, are not possible in the SM. However, the limits on their cross-sections provide invaluable information on the nature of the couplings involved and if found might be an indication of new physics.

As the recorded LHC data increases, searches for rare and very rare top quark production processes are becoming more viable, providing key information on the properties of the top quark and its couplings.

2.2.1.2 Top Quark Decay

As a SM test, direct measurements of the $t \to Wb$ branching ratio have been made by the CDF, DØ and CMS experiments, using the ratio $R = B(t \to Wb)/\sum_{q=d,s,b} B(t \to Wq)$ where the sum over q includes all possible down-type quarks [104, 105, 106]. The ratio R allows to probe the SM, since significant deviations from unity could imply a non-SM top quark decay. The most precise measurements are $R = 1.014 \pm 0.003(stat) \pm 0.032(syst)$ and $R > 0.955$ at 95% C.L. as determined by the CMS experiment in $\sqrt{s} = 8$ TeV pp collision data. This result, together with other beyond the SM searches, corroborates the SM prediction of a nearly exclusive decay of the top quark into a W boson and a b quark.

Due to the high top quark mass, the W boson from the decay of a top quark is produced nearly on-shell. Assuming $|V_{tb}| = 1$, the LO decay width of the top quark can then be calculated from the Feynman diagram of the $t \to Wb$ body channel in Figure 2.9 as:

$$\Gamma_t^{\mathrm{LO}} = \frac{G_F {m_t}^3}{8\pi\sqrt{2}} \left(1 - \frac{{m_W}^2}{{m_t}^2}\right)^2 \left(1 + 2\frac{{m_W}^2}{{m_t}^2}\right) \tag{2.41}$$

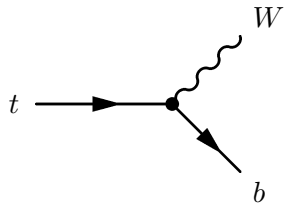


Figure 2.9: Leading-order Feynman diagram of the top quark decay.

where both the t quark and W boson are treated as on-shell particles[1] and the mass of the b quark is considered negligible [19]. For a top quark mass of 172.5 GeV, the predicted NNLO decay width is $\Gamma_t = 1.322\,\text{GeV}$, a value that increases with the mass [107].

Direct measurements of the top quark decay width have been performed by the ATLAS collaboration, yielding results in agreement with the SM prediction. For a top quark mass of 172.5 GeV, $\Gamma_t = 1.76 \pm 0.33(\text{stat.})^{+0.79}_{-0.69}(\text{syst.})\,\text{GeV}$ and $\Gamma_t = 1.28 \pm 0.30\,\text{GeV}$ have been measured using either the full $\sqrt{s} = 8\,\text{TeV}$ dataset [108] or a subset only exploiting events away from the $t\bar{t}$ resonance peak [109], respectively. Additionally, measurements at $\sqrt{s} = 13\,\text{TeV}$ have obtained a top width of $\Gamma_t = 1.9 \pm 0.5\,\text{GeV}$ [110].

The final states of top quark processes are then usually classified according to the W boson decay, which decays leptonically about one third of the time ($W^+ \to l^+ + \nu$ and the charge conjugate $W^- \to l^- + \bar{\nu}$) or hadronically about two thirds of the time ($W^+ \to u + \bar{d}$ and its charge conjugate $W^- \to d + \bar{u}$, where u and d represent up or down-type quarks). While the interactions occurring in particle accelerators are theoretically well described by parton-level gauge interactions yielding final state partons, partons are not visible in the detectors. The measurable quantities are the particle-level bound states of quarks and antiquarks, formed through hadronisation. Inside the detector, final state quarks evolve into jets of hadrons, whereas neutrinos pass through undetected. The final state of processes involving top quarks will then usually consist of a few jets, out of which one or two will be originating from b quarks, one or two charged leptons and missing energy carried away by the otherwise undetected neutrino(s). Despite the ambiguity due to the missing neutrino information, leptonic decays are typically easier to trigger since leptons leave a very clean signature in most detectors. The fully hadronic decay of the $W^\pm$ bosons is the hardest to trigger, particularly due to the severe QCD multijet background processes.

[1] Here m_t is the top quark pole mass.

2.2.2 Higgs Boson Physics

The SM Higgs boson is an electrically neutral CP-even scalar particle, and whose mass, given by $m_H = \nu\sqrt{2\lambda}$ (*c.f.* Equation 2.23), is a free parameter of the theory. There is no theoretical prediction of the Higgs boson mass since the quartic coupling λ is free in the model, depending only of the initial conditions of the universe. However, prior to its discovery at the LHC, the window for a SM Higgs boson was already very narrow. Since LEP direct searches found no signal, the Higgs mass was constrained by $m_H > 114\,\text{GeV}$ as a lower bound. In addition, indirect upper bounds obtained from the precision electroweak measurements at LEP and Tevatron, inferred that m_H was unlikely to be greater than 200 GeV. After its discovery in 2012 by the ATLAS and CMS experiments, the most recent experimental measures of the Higgs boson mass yield $m_H = 125.25 \pm 0.17\,\text{GeV}$, which leads to $\lambda \simeq 0.13$ [19].

The Higgs boson couples to all fundamental fermions and gauge bosons in the SM. The Higgs boson couplings to fermions are linearly proportional to the fermion masses, while the Higgs boson couplings to gauge bosons are proportional to the the square of the boson masses. This proportionality of the Higgs boson couplings to mass is fundamental in determining its dominant production and decay processes. In order to understand the Higgs boson nature, it is important to measure all properties of the Higgs boson couplings to SM particles.

2.2.2.1 Higgs Boson Production

At hadron colliders, Higgs boson production occurs through a variety of processes, since the Higgs boson couples to nearly every fundamental particle in the SM. The dominant mechanisms for its production involve the coupling of the H boson to heavy particles, such as the W and Z gauge bosons and the third generation of quarks and leptons. At the LHC, the Higgs boson can be produced via four main mechanisms, whose leading order diagrams are shown in Figure 2.10: gluon-fusion (ggF); vector-boson fusion (VBF); associated production with a heavy gauge boson (VH, where $V = W^{\pm}, Z$); and associated production with one or two top quarks (tHq or $t\bar{t}H$, respectively).

Despite the rather large variety of couplings possible, the Higgs boson production at the LHC is still a low rate process. In Figure 2.11 (a) the theoretical predictions for the different Higgs boson production modes are shown as a function of the centre-of-mass energy, for a

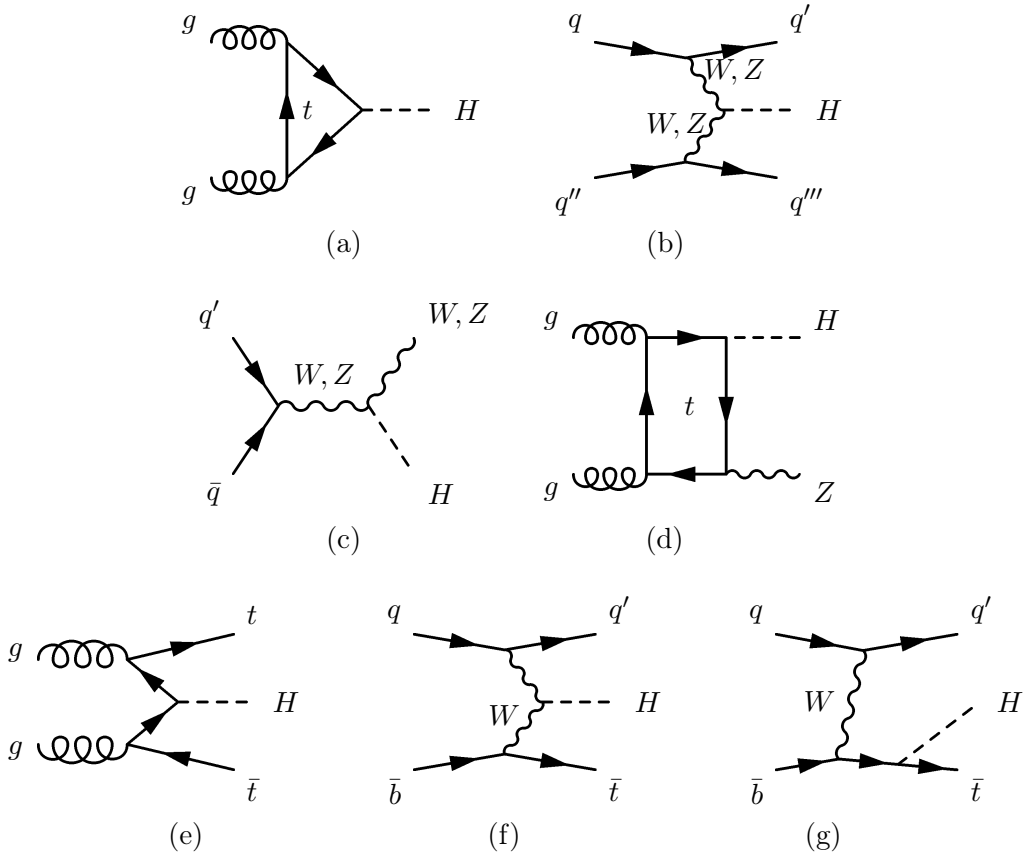


Figure 2.10: Leading-order Feynman diagrams of the main Higgs boson production processes at the LHC: (a) gluon-fusion (ggF); (b) vector-boson fusion (VBF); (c) Higgs-strahlung (VH); (d) associated production with a heavy gauge boson at loop level from a gluon-gluon interaction (ZH); (e) associated production with a pair of top quarks ($t\bar{t}H$); (f) and (g) associated production with a single top quark (tHq).

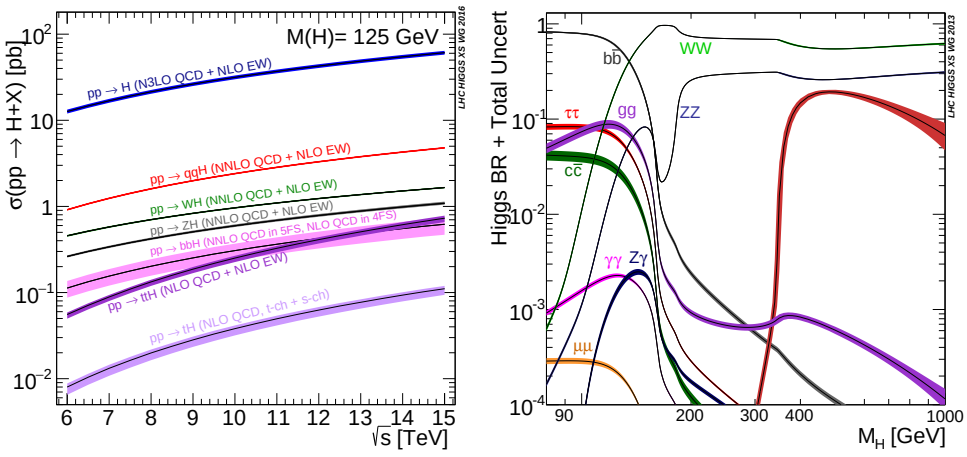
SM Higgs boson of 125 GeV. A summary of the expected theoretical cross-section values is also included in Table 2.4 for the most relevant centre-of-mass energies.

In proton collisions, the gluon-fusion process, $gg \to H + X$, is the production mechanism with the largest cross-section. The Higgs boson coupling to gluons is induced at leading order via a loop of heavy quarks, where a virtual pair of top quarks will couple to the Higgs boson (see Figure 2.10 (a)). Other virtual loops of lighter quarks could also contribute to the gluon-fusion process, but to a smaller extent than that of the massive top quark. Similarly, the Higgs boson coupling to photons must also occur through virtual loops. In this case, however, the dominant contribution will be from the coupling to a pair of massive W^+W^-

$\sqrt{s}$	Cross-Section [pb]			
[TeV]	$\sigma_{\mathbf{ggF}}$	$\sigma_{\mathbf{VBF}}$	σ_{WH}	σ_{ZH}
1.96	$0.95^{+17\%}_{-17\%}$	$0.065^{+8\%}_{-7\%}$	$0.13^{+8\%}_{-8\%}$	$0.079^{+8\%}_{-8\%}$
7	$16.82^{+4.4\%+3.3\%}_{-7.0-3.3\%}$	$1.240^{+0.2\%+2.2\%}_{-0.2\%-2.2\%}$	$0.5756^{+0.7\%+2.1\%}_{-0.8\%-2.1\%}$	$0.3384^{+2.6\%+1.7\%}_{-2.3\%-1.7\%}$
8	$21.39^{+4.4\%+3.2\%}_{-6.9\%-3.2\%}$	$1.600^{+0.3\%+2.2\%}_{-0.2\%-2.2\%}$	$0.7009^{+0.7+2.0\%}_{-0.8\%-2.0\%}$	$0.4199^{+2.8\%+1.7\%}_{-2.4\%-1.7\%}$
13	$48.61^{+4.27\%+1.85\%}_{-6.49\%-1.85\%}$	$3.766^{+0.43\%+2.1\%}_{-0.33\%-2.1\%}$	$1.358^{+0.51\%+1.35\%}_{-0.51\%-1.35\%}$	$0.880^{+3.50\%+1.65\%}_{-2.68\%-1.65\%}$
14	$54.72^{+4.28\%+1.85\%}_{-6.46\%-1.85\%}$	$4.260^{+0.45\%+2.1\%}_{-0.34\%-2.1\%}$	$1.498^{+0.51\%+1.35\%}_{-0.51\%-1.35\%}$	$0.981^{+3.61\%+1.90\%}_{-2.94\%-1.90\%}$

$\sqrt{s}$	Cross-Section [pb]
[TeV]	$\sigma_{t\bar{t}H}$
1.96	$0.004^{+10\%}_{-10\%}$
7	$0.08864^{+3.5\%+4.5\%}_{-9.2\%-5.0\%}$
8	$0.1326^{+4.0\%+4.3\%}_{-9.2\%-4.3\%}$
13	$0.5065^{+5.8\%+3.6\%}_{-9.2\%-3.6\%}$
14	$0.6128^{+6.0\%+3.5\%}_{-9.2\%-3.5\%}$

Table 2.4: Theoretical cross-section predictions for the main SM Higgs boson production processes at several centre-of-mass energies. For the Tevatron energy scale, $\sqrt{s} = 1.96\,\text{TeV}$, the predictions were obtained from Ref. [19] for $m_H = 125\,\text{GeV}$ and the combined total uncertainties are shown. As for the LHC energies, the predictions for the cross-section were taken from Refs [111, 112] for the Higgs mass $m_H = 125.09\,\text{GeV}$. Unless stated otherwise, the first uncertainty always represents the renormalisation and factorisation scale dependence for the LHC energies, whereas the second one is the combined PDF and α_S uncertainty. The cross-section predictions for the ggF channel have next-to-next-to-next-leading-order (N^3LO) precision in QCD and NLO EW accuracies [113], using the YR4 input parameters [111] and the PDF4LHC15_nnlo_100 set [71]. Here, the first uncertainty is the linear sum of various sources of theoretical uncertainties affecting the cross-section calculations, including the QCD scale. The second uncertainty corresponds to the combined PDF and α_S uncertainties computed following the PDF4LHC recommendation [114]. The VBF cross-section predictions were calculated at (approximate only for 7 and 8 TeV) NNLO QCD and NLO EW accuracies [115]–[119], using the same settings as in YR4 [111] with the PDF4LHC15_nnlo_100 [71] set for the QCD corrections. The EW and Photon corrections were computed using the NNPDF23_nlo_as_0118_qed [120] set at 7 and 8 TeV and the LUXqed_plus_PDF4LHC_nnlo_100 [121, 122] PDF set at 13 and 14 TeV. The VH cross-sections were calculated at NNLO QCD and NLO EW accuracies [119] [123]–[128], using PDF4LHC15_nnlo_mc (QCD) and NNPDF23_nlo_as_0118_qed (EW) PDF sets [120], and are shown separately for WH and ZH production. The theoretical predictions for the $t\bar{t}H$ associated production were calculated at NLO QCD [129]–[135] and NLO EW [134]–[136] accuracies, using MadGraph5_aMC@NLO [137, 138] and the PDF4LHC15_nlo_30_pdfas set.



(a) (b)

Figure 2.11: Standard Model Higgs boson (a) production cross-sections as a function of the centre-of-mass-energies [111] and (b) branching ratios in function of the Higgs boson mass [139]. The tH production cross-section includes only the t-channel and s-channel production.

bosons.

Other production modes always involve additional particles generated in association with the Higgs boson. At the LHC, vector boson fusion (VBF), $qq'' \to q'q'''H$, is the second most likely production mechanism, where two quarks from the initial protons (or two antiquarks from the virtual sea) radiate massive vector bosons which interact originating a Higgs boson. Due to the different fermionic couplings to the W and Z gauge bosons, and their different masses, the cross-section of the W fusion process is enhanced approximately three times more than that of the Z fusion. In this production mode, the resulting Higgs boson is accompanied by a pair of (anti)quarks, that give rise to two back-to-back hard jets in the most forward regions of the detectors. With no colour conversion occurring between the quarks in the weak gauge boson exchange, gluon radiation is mostly suppressed in the central regions of the detector. Therefore, experimentally VBF processes have a much cleaner signature than ggF ones, even allowing discrimination from the overwhelming QCD background.

Associated production with W and Z gauge bosons, VH, is the next most relevant Higgs boson production mode at the LHC. The VH associated production processes, $pp \to VH+X$ with $V = W^{\pm}, Z$ can occur either at tree level via quark-quark interaction, also called Higgs-strahlung (see Figure 2.10 (c)), or, in particular for the ZH mode, at loop level from a gluon-

gluon interaction (see Figure 2.10 (d)). Despite the additional gluon fusion contribution to the ZH production mechanism, the WH cross-section is twice that of the ZH, due to the different intensities of electroweak gauge couplings.

Lastly, the production of a Higgs boson in association with heavy quarks which have smaller cross-sections than the previous mechanisms, but are nonetheless very interesting for the study of the Yukawa couplings. The Higgs boson associated production with heavy quarks can happen involving a pair of quark-antiquark or a single (anti)quark (see Figures 2.10 (e) or (f) and (g) respectively). Given the large Yukawa coupling of the third generation of quarks, the top quarks have the highest associated production rates, followed by the bottom quarks. Out of these, the most favourable processes will then be those involving the production of a pair of top quarks, $gg \to t\bar{t}H$, usually referred as $t\bar{t}H$ production. The Higgs boson radiation off top quarks is particularly interesting since no virtual loop is involved in the production mechanism, which enables the direct probe of the top quark-Higgs boson Yukawa coupling, unlike the ggF channel for instance. For the $t\bar{t}H$ production, a detailed summary is presented in Section 2.2.3 below. The contribution of the Higgs boson production in association with a single top quark (tHq), though smaller, allows the determination of the sign of the top Yukawa coupling. Analogously to the single top quark production, the tHq can be divided into three production channels according to the virtuality of the W boson: t-channel production where a virtual space-like W is exchanged; s-channel production featuring a virtual time-like W; and W-associated production when an on-shell W boson is present in the final state.[1] In Figures 2.10 (f) and (g), two tHq t-channel examples are given at LO.

Other single Higgs boson production mechanisms are possible at the LHC. Besides having smaller contributions, they can still bring valuable information on the nature of the Higgs boson and its interactions with other SM particles. For instance, the Higgs boson production in association with a pair of bottom quarks ($b\bar{b}H$), involving the SM suppressed bottom-Higgs coupling, could be dominant at larger energies in beyond the SM theories, such as the two-Higgs-doublet model or a SUSY model. Likewise, the Higgs boson production in association with charm quarks ($c\bar{c}H$), whose expected production rate in the SM is quite small, may become relevant in specific models with enhanced charm Yukawa coupling.

[1] Although useful at event generation, this classification is only valid at LO, since when adding higher-order corrections some amplitudes can interfere.

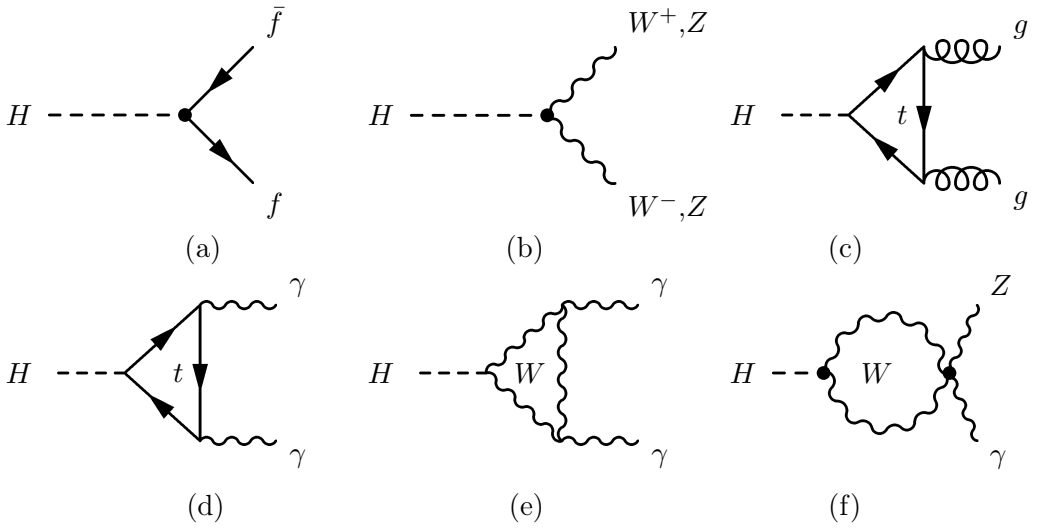


Figure 2.12: Lowest order Feynman diagrams of the SM Higgs boson decay: (a) $H \to f\bar{f}$; (b) $H \to W^+W^-$ or $H \to ZZ$; (c) $H \to gg$; (d) and (e) $H \to \gamma\gamma$ via a top quark or W boson virtual loop, respectively; (f) $H \to \gamma Z$ via a W boson virtual loop.

At hadron colliders, double Higgs boson production can also occur, although at much smaller rates than single Higgs boson production. In pp collisions, its dominant production mode happens via gluon fusion, $gg \to HH$, with an expected inclusive NNLO cross-section at $\sqrt{s} = 13\,\text{TeV}$ of $0.03105^{2.2\%}_{5.0\%}(\text{theory}) \pm 3\%(\text{PDF} + \alpha_\text{S}) \pm 2.6\%(\text{m}_\text{t})\,\text{pb}$ [140]. Double Higgs boson production can provide access to the Higgs trilinear self coupling, thus, being important to further understand the nature of the Higgs potential.

2.2.2.2 Higgs Boson Decay

The Higgs boson is not stable and can in principle decay to all Standard Model particles. The lowest order Feynman diagrams for the main decay modes are shown in Figure 2.12. For each decay, the matrix element is proportional to the mass of the particle coupling to the Higgs boson. Thus, the Higgs boson can decay into a pair of fermions, $H \to f\bar{f}$, as long as the channel is kinematically accessible, *i.e.* $m_H > 2m_f$. Similarly, the Higgs boson can also decay into a pair of weak gauge bosons, $H \to W^+W^-$ or $H \to ZZ$, where at least one of the bosons is produced off-mass-shell. It can even decay to massless particles like photons and gluons, although indirectly via boson and fermion loops.

The partial decay widths can be calculated from the Feynman rules for the interaction

Decay	Branching Ratio
$H \to b\bar{b}$	$58.09\%^{+0.65\%+0.72\%+0.77\%}_{-0.65\%-0.74\%-0.79\%}$
$H \to W^+W^-$	$21.52\%^{+0.99\%+0.98\%+0.64\%}_{-0.99\%-0.98\%-0.62\%}$
$H \to gg$	$8.180\%^{+3.40\%+1.12\%+3.70\%}_{-3.41\%-1.14\%-3.59\%}$
$H \to \tau^+\tau^-$	$6.256\%^{+1.17\%+0.98\%+0.62\%}_{-1.16\%-0.98\%-0.60\%}$
$H \to c\bar{c}$	$2.884\%^{+1.20\%+5.27\%+1.26\%}_{-1.20\%-0.94\%-1.25\%}$
$H \to ZZ$	$2.641\%^{+0.99\%+0.98\%+0.64\%}_{-0.99\%-0.98\%-0.62\%}$
$H \to \gamma\gamma$	$0.2270\%^{+1.73\%+0.97\%+0.66\%}_{-1.72\%-0.94\%-0.61\%}$
$H \to \gamma Z$	$0.1541\%^{+5.71\%+0.91\%+0.58\%}_{-5.71\%-1.00\%-0.64\%}$
$H \to \mu^+\mu^-$	$0.02171\%^{+1.23\%+0.97\%+0.60\%}_{-1.23\%-0.99\%-0.64\%}$

Table 2.5: Predicted branching ratios for 125.09 GeV SM Higgs boson [111]. The three estimated uncertainties correspond to the relative theoretical uncertainties resulting from missing higher order corrections, parametric uncertainties on the input quark masses, m_q where $q = c$, b and t, and on α_S, respectively.

vertex. For example, the partial decay width for a decay into a pair of b quarks is,

$$\Gamma(H \to b\bar{b}) = 3 \times \frac{{m_b}^2 m_H}{8\pi\nu 2}\,, \tag{2.42}$$

where the three possible colour states of the resulting $b\bar{b}$ pair are taken into account. Similarly, for other fermion pairs the partial decay rate is also proportional to the square of the fermion mass.[1] For the remaining Higgs boson decays, the corresponding partial width always depends on the mass of the particles involved in the coupling. When the output particles are massless, such as $H \to gg$ or $H \to \gamma\gamma$, the partial width of the decay depends instead on the mass of the particles inside the virtual loop. Since the particles in the virtual loops are quite massive, these decay rates are still comparable to those of fermions or off-mass-shell gauge bosons and provide indirect information on the Higgs couplings to W^+W^-, ZZ and $t\bar{t}$ in different combinations. Determining all relevant Higgs boson decay widths is fundamental in understanding and interpreting any Higgs boson analysis.

The branching ratios (BRs) for a SM Higgs boson with $m_H = 125.09\,\text{GeV}$ are shown in Table 2.5. The uncertainties in the BRs account for missing higher-order corrections in the

[1] Here, the quark masses run with q^2, just as the running of α_S. The energy scale is typically set to the Higgs boson mass, $q^2 = m_H^2$.

theoretical calculations and parametric uncertainties in the SM input parameters, such as the fermion masses, m_q, and the QCD gauge coupling, α_S [111]. For completeness, and since the BRs depend on the measured value of the Higgs mass, in Figure 2.11 (b) the different expected decay fractions are shown as a function of m_H. Given the different fermion and boson masses, the main decay processes are then $H \to b\bar{b}$ and $H \to W^+W^-$. Other relevant decay processes involve the strong coupling to gluons ($H \to gg$) or couplings to the third and second generation of fermions ($H \to \tau^+\tau^-$ and $H \to c\bar{c}$, respectively). Despite the somewhat low branching fraction, the Higgs boson decay into a pair of Z bosons, $H \to ZZ$, plays an important role at the LHC. The Higgs boson decays $H \to \gamma\gamma$, $H \to \gamma Z$ and $H \to \mu^+\mu^-$, although possible, have much smaller rates. The total predicted width, for a $m_H = 125.09\,\text{GeV}$ Higgs boson, is $\Gamma_H^{\text{Total}} = 4.100^{+0.73\%}_{-0.73\%}(\text{theory})^{+0.99\%}_{-0.97\%}(\text{m}_\text{q})^{+0.60\%}_{-0.61\%}(\alpha_\text{S})\,\text{MeV}$. Further information on the theoretical calculations of the Higgs boson decay and width can be found in [111].

Most of the particles produced in the Higgs boson decay will successively decay into lighter, more stable particles, even in the case hadronisation happens. Experimentally, the sensitivity to any channel will depend not only on the production cross-section and the decay BR, but also on the selection efficiency and the signal-to-background ratio in the final state. Therefore, while decay process such as $H \to W^+W^- \to l^+ + \nu_l + l'^-\bar{\nu}'_l$, $H \to b\bar{b}$ and $H \to \tau^+\tau^-$ have large branching fractions, their final states involve neutrinos or large background rates, thus leading to quite challenging analysis in order to isolate the Higgs boson signal. The most prominent decay channels at current LHC energies are thus $H \to \gamma\gamma$ and $H \to ZZ \to 4l$ where all final state particles can be precisely measured. With clear experimental signatures, these decays are easily identifiable from background processes, which makes them the most sensitive decay modes at the LHC, despite the relatively low branching ratios. In fact, due to the high resolution in the diphoton and $4l$-invariant masses, these decay modes were the two main discovery channels of the Higgs boson in 2012 [25, 26].

Since its discovery, Higgs boson properties have been under investigation at different energy runs of the LHC pp collisions. During the LHC Run 1 ($\sqrt{s} = 7$ and 8 TeV), measurements were only accessible via production and decay channels related to the couplings of the Higgs boson to the vector gauge bosons (γ, Z, W and g), whereas in LHC Run 2 ($\sqrt{s} = 13$ TeV) the couplings of the Higgs boson to the charged fermions of the third generation (t, b and τ) were established. So far, all observed Higgs Yukawa couplings are found in agreement,

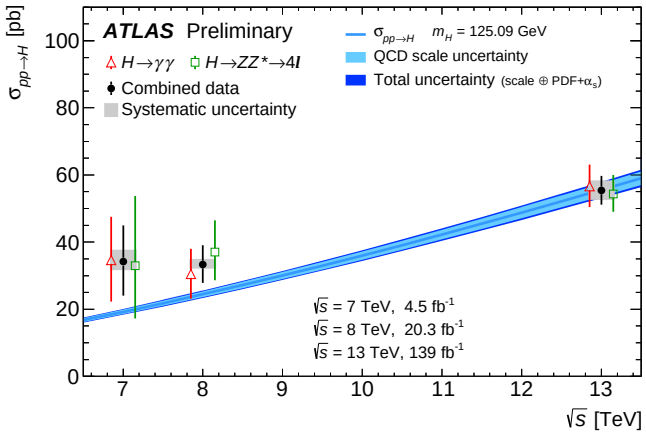


Figure 2.13: Total $pp \to H + X$ cross-sections measured at centre-of-mass energies 7, 8 and 13 TeV and compared to SM predictions. The $H \to \gamma\gamma$ (red triangles), $H \to ZZ^* \to 4l$ (green squares) and the combined (black dots) are shown, where the different energy results are offset along x-axis for better visualisation. The systematic uncertainties on the combined measurements are represented by the grey bands, whereas the error bars show the total uncertainty. The estimated uncertainty due to missing higher-order corrections is shown in the light blue band, and the total uncertainty in the dark blue band. The total theoretical uncertainty corresponds to the higher-order-correction uncertainty summed in quadrature with the sum of the PDF and α_S uncertainties, and is partially correlated across values of the centre-of-mass energy [141].

within experimental and theoretical uncertainties, with the SM expectations. As an example, in Figure 2.13, the measured total cross-sections for $H \to \gamma\gamma$ and $H \to ZZ \to 4l$ decay channels for $\sqrt{s} = 7$, 8 and 13 TeV pp collisions in the ATLAS experiment are compared to the SM predictions [141]. No significant deviations are found between the measured results and the state-of-the-art theoretical predictions.

The total decay width for a SM Higgs boson cannot be directly observable at the LHC, due to the nature of the experiments, where the identification of the Higgs boson decay products is necessary. Nevertheless, it can be indirectly inferred with additional assumptions. For instance, the combination of on-peak and off-peak contributions to the total width in Higgs decays to ZZ^* can provide such additional constraints.[1] The total width of the Higgs boson decay was measured to be $\Gamma_H^{\text{Total}} = 3.2^{+2.8}_{-2.2}$ MeV which is in good agreement with the SM prediction[19].

[1] Z^* represents a Z boson with an off-shell mass.

2.2.3 Top Quark Produced in Association with Higgs Bosons

The coupling of the Higgs boson to top quarks is quite special. Not only is it the largest Higgs boson coupling in the SM, but it may also provide insight into the scale of new physics [142]. The discovery of the Higgs boson is the first indirect evidence for this coupling, since its main production mechanism includes top quark virtual loops. Despite being one of the Higgs boson production mechanisms with lower cross-section, the associated production of a pair of top quark-antiquark and a Higgs boson, also known as $t\bar{t}H$, is the dominant process for a direct probe of the top quark-Higgs boson Yukawa coupling at current collider conditions. While challenging to measure with $\sqrt{s} = 8\,\text{TeV}$ data (having a cross-section of about 132.6 fb), the $\sqrt{s} = 13\,\text{TeV}$ cross-section at the LHC (506.5 fb) provides a substantial increment in $t\bar{t}H$ signal events, allowing more precise results. The measurement of the $t\bar{t}H$ final state is, however, not trivial, having a fairly complex signature and overwhelming background processes mimicking its imprint on the detectors. Ultimately, with the availability of ever increasing data and novel analysis techniques, the associated production of a pair of top quarks and a Higgs boson can provide a clean measurement of the top quark-Higgs boson Yukawa coupling, including its CP nature.

At the LHC, $t\bar{t}H$ events are mainly produced via gluon fusion or quark-antiquark annihilation, as shown in the Feynman diagrams in Figure 2.14. Since the top quark-antiquark pair as well as the Higgs boson will rapidly decay, there are several possible signatures of this mechanism, all of which include high multiplicity of quarks in their final state.

The associated production of a top quark-antiquark pair and a Higgs boson can be divided in terms of the decay modes of the two top quarks into one of three channels:

Fully Hadronic When both W bosons decay hadronically ($BR \approx 45.4\%$), originating a final signature with at least six jets, two of which originate from the b quarks coming from the Wtb vertex.

Semileptonic Occurs when one of the W bosons decays leptonically and the other hadronically ($BR \approx 44.1\%$). Its final state topology is characterised by: at least four jets, where at least two originate from b quarks of the t quark decay; one charged lepton; and a neutrino, whose presence is inferred via a large imbalance of the momentum in the transverse plane. Due to its experimental signature, this channel is also known has *lepton + jets*.

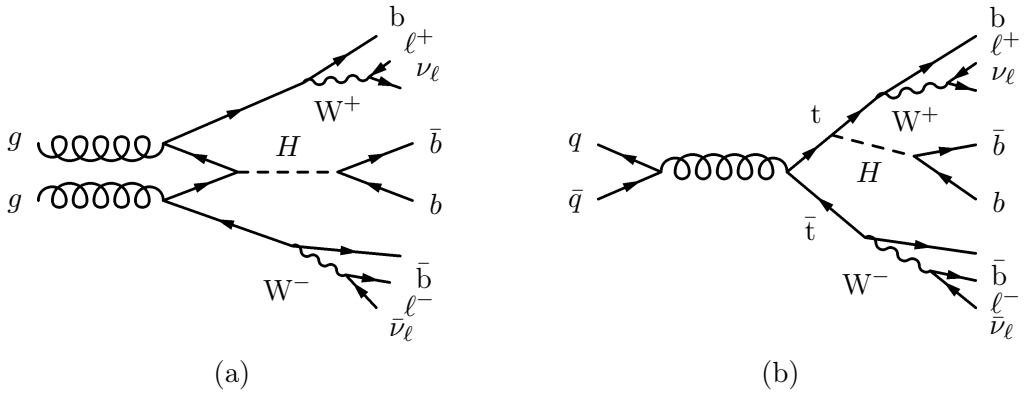


Figure 2.14: Tree-level Feynman diagrams of the dileptonic channel of the associated production of a top quark-antiquark pair and a SM Higgs boson with subsequent decay into a b quark-antiquark pair.

Dileptonic Where both W bosons decay leptonically ($BR \approx 10.6\%$). The final state topology includes: two jets coming from the two b quarks of the t decay; two opposite-sign charged leptons and two neutrinos, whose indirect detection occurs via the presence of missing energy in the transverse plane.

Hadronic decays have the highest branching fractions. However, a full hadronic signature is overwhelmed by QCD backgrounds, which have much larger cross-sections. Leptonic decays have typically more accessible detection because the leptons are fairly easy to identify, however, the neutrino presence makes it impossible to directly reconstruct the four-momentum of the top quark. In this thesis, the focus is on the associated production of Higgs boson with top quarks in the dilepton decay channel. Despite the comparatively small branching fraction, the dilepton channel has a very clear signature with two leptons, which contributes to a significant reduction of processes with similar final states. When produced, the τ-lepton will also further decay leptonically or hadronically, which may lead to the presence of extra electrons or muons in the final state. The resulting objects of the leptonic τ decays are included in the analysis performed in this thesis, as long as the selection criteria is fulfilled (see Chapter 5 for more details).

Regarding the decay of the Higgs boson, and considering the variety of possible decays (refer to Section 2.2.2.2), a classification can usually be done according to the final state topology:

Diphoton When the Higgs boson decays into a pair of photons, $H \to \gamma\gamma$ ($BR \approx 0.2\%$). Despite its very small branching fraction, the diphoton channel allows the reconstruction of the Higgs boson as a very narrow mass peak.

$\boldsymbol{H \to ZZ \to 4l}$ when the Higgs decays into a pair of Z bosons that decay exclusively into two pairs of same flavoured leptons ($BR \approx 0.01\%$ for electrons and muons). It has limited statistics due to the small branching fraction of the Z decays to leptons.

Multilepton Where the $H \to W^+W^-/ZZ$ and $H \to \tau^+\tau^-$ decays are considered inclusively with subsequent leptonic decays ($BR \approx 15\%$), originating a final state with the presence of multiple leptons.[1] The multilepton channel has a significant branching ratio, but the reconstruction of the Higgs boson is a bit more complex since it involves multiple leptons and/or neutrinos.

Hadronic $\boldsymbol{\tau}$-lepton where the Higgs boson originates a pair of τ-leptons, $H \to \tau^+\tau^-$, with a subsequent decay of both τ-leptons to hadrons ($BR \approx 4\%$).

$\boldsymbol{H \to b\bar{b}}$ When the Higgs boson decays into a b quark-antiquark pair ($BR \approx 58.1\%$). Typically this channel is analysed separately according to the top quark pair decays. The dilepton final state is characterised by four b-jets as can be seen in Figure 2.14.

Besides being the dominant one, the $H \to b\bar{b}$ decay in association with the top quark-antiquark pair significantly improves the the signal-to-background ratio. Due to the large backgrounds, however, the analysis of this channel is quite intricate. The $t\bar{t}$ production in association with extra jets, particularly $b\bar{b}$ jets, has the exact same final state topology, and is therefore the main irreducible background for the $t\bar{t}H$ with a $H \to b\bar{b}$ decay. The reconstruction of the Higgs boson is especially complex due to the combinatorial possibilities in resolving the $b\bar{b}$ system from the Higgs boson decay, in events with at least four b-tagged jets. In this thesis, the search for the SM Higgs boson in the $t\bar{t}H$ production mode is described, with predominant focus on the $H \to b\bar{b}$ decay sensitivity. Nevertheless, all other Higgs boson decay modes are treated as signal, in order to have a sensitive and somewhat competitive search.

The production of $t\bar{t}H$ events has been investigated using LHC data at 7, 8 and 13 TeV centre-of-mass energies by the ATLAS and CMS experiments. Searches of the $t\bar{t}H$ mechanism with the LHC Run 1 dataset ($\sqrt{s} = 7$ and 8 TeV) concentrate on having as many

[1] Events in the resonant $H \to 4l$ channel are not included, as to avoid overlap with other channels.

experimental signatures as possible, due to the limited level of statistics available. No significant excess of events above the background expectation is found by either experiment [143, 144, 145] at these center-of-mass energies. The ATLAS dedicated individual searches (using diphoton [146], multilepton [147], combined semilepton and dilepton [1] or full hadronic [144] events) show that while the diphoton channel has a limited sensitivity due to low statistics, all other channels are highly dominated by large systematic uncertainties, particularly on the background predictions. The ATLAS combination of all $t\bar{t}H$ searches yields an observed (expected) upper limit of 3.1 (1.4) times the SM expectation at 95% confidence level, with a signal strength $\mu = 1.7 \pm 0.8$ [144]. The equivalent CMS result has an observed $t\bar{t}H$ signal strength relative to the standard model cross-section $\mu = 2.8 \pm 1.0$ under the assumption that the Higgs boson decays as expected in the SM [143]. Combining ATLAS and CMS Run 1 results, a measured $t\bar{t}H$ signal significance of 4.4 is obtained where 2.0 is expected [148], which is not quite enough for a significant observation of this production mechanism. The work presented in this thesis actively contributed to the ATLAS search of the $t\bar{t}H$ with subsequent $H \to b\bar{b}$ decay, using 8 TeV semileptonic and dileptonic events [1]. This search reports an observed (expected) limit of 3.4 (2.2) times the Standard Model cross-section at 95% confidence level and a signal strength of $\mu = 1.5 \pm 1.1$ times the SM expectation, assuming a Higgs boson mass of 125 GeV. The sensitivity of this analysis is strongly influenced by systematic uncertainties on the irreducible background predictions, which hinders any strict constraints on the top Yukawa coupling.

Analysis of LHC Run 2 data ($\sqrt{s} = 13$ TeV) can provide better constraints on the top Yukawa coupling, seeing that the $t\bar{t}H$ cross-section has the largest relative increase comparatively to other background processes where top quarks are also produced. Both ATLAS [146]–[150] and CMS [143]–[155] experiments independently observe significant production of the $t\bar{t}H$ signal, including every sensitive decay channel and using datasets with center-of-mass energies $\sqrt{s} = 7$, 8, and 13 TeV combined [93, 94]. In order to maximise sensitivity, the different decay channel searches ($H \to \gamma\gamma$, WW, ZZ, $\tau\bar{\tau}$ and $b\bar{b}$) are investigated using orthogonal selection criteria in a statistically independent way and are later combined. CMS reports an observed (expected) excess of events with a significance of 5.2 (4.2) standard deviations over the expectation from the background-only hypothesis and a signal strength of $1.26^{+0.31}_{-0.26}$ times the SM prediction [93]. The equivalent $\sqrt{s} = 7$, 8, and 13 TeV combined result by the ATLAS experiment yields an observed (expected) significance of 6.3 (5.1) standard deviations, whereas using the Run 2 partial dataset alone 5.8 standard

deviations are observed compared to an expectation of 4.9 standard deviations [94]. With the full Run 2 data made available, results for most of these channels keep being updated by both experiments independently, using improved background rejection methods and more precise theory calculations [156]–[161].

3

The ATLAS Experiment

This chapter gives a brief description of the European Council for Nuclear Research (CERN) where the Large Hadron Collider (LHC) is located, focusing on a detailed discussion of the particle accelerator and of the ATLAS detector. The ATLAS sub-detectors, trigger and data acquisition systems are also described, as well as their performance.

3.1 CERN

CERN is a research center of particle physics, combining the efforts of thousands of scientists and engineers from all over the world. Located on the France-Swiss border, near Geneva, CERN was founded in 1954 and now comprises 23 member states. The organisation facilities cover six square kilometers scattered over twelve sites, providing infrastructures for high-energy physics research, such as the particle accelerators installed underground as well as huge caverns with particle detectors.

CERN has led to several discoveries, both in physics and technology-related areas. For instance, the discovery of neutral currents in the Gargamelle bubble chamber [50], the discovery of the $W^{\pm}$ and Z bosons in the UA1 and UA2 experiments [51, 52], the determination of the number of light neutrino families at the Large Electron-Positron Collider (LEP) [162] and the discovery of direct CP violation in the NA48 experiment [163]. Additionally, the first creation of anti-hydrogen atoms was performed at the PS210 experiment at CERN [164], as well as, isolating 38 anti-hydrogen atoms [165] and maintaining anti-hydrogen for over 15 minutes [166]. Furthemore, several technological discoveries were accomplished at CERN, such as the creation of the World Wide Web and the computer network infrastructure Worldwide LHC Computing Grid (GRID).

3.2 Large Hadron Collider

The LHC [167] is a high energy particle accelerator built at CERN with the intention to probe several theoretical predictions of particle physics, such as the SM Higgs boson. Even in the initial phase of operation, the LHC experiments were already able to measure important SM processes like the production of high transverse momentum jets, W and Z bosons, as well as b and t-quarks. The high luminosity and increased cross-sections at the LHC allow high precision tests of electroweak interactions, QCD and flavour physics. In addition, the available high collision energies enable the search for new particles in so far unexplored mass regions. Important constraints could be established on the existence of new particles predicted by BSM physics, such as heavy gauge bosons W' and Z'. Furthermore, the LHC data is quite sensitive to many models of new physics, such as searches for FCNCs and lepton flavour violation processes. The search for supersymmetric particles and their possible connection to the dark matter in the universe is also under investigation at the LHC.

The LHC is located in a 27 km underground circular tunnel, at a depth ranging from 50 to 175 m, originally built for LEP. On the LHC ring, two parallel beams of protons or heavy-ions are accelerated, traveling in opposite directions and in adjacent pipes that intersect at four interaction points (IP) where the particle collisions occur. Around each of the four collision points depicted in Figure 3.1, different detectors have been built to reconstruct and study final states of the collisions delivered by the LHC: ALICE (A Large Ion Collider Experiment) [168], ATLAS (A Toroidal LHC ApparatuS) [169], CMS (Compact Muon Solenoid) [170], and LHCb (Large Hadron Collider Beauty Experiment) [171]. In pp collisions, protons are extracted from the ionisation of Hydrogen gas and injected in a chain of linear and circular accelerators (PS, SPS), entering the LHC with an energy of 450 GeV. In order to avoid collisions of beam particles with gas molecules, the proton beampipes are kept at ultrahigh vacuum (10^{-13} atm). The protons are then further accelerated to the designated energy by passing through alternating electrical fields in radiofrequency (RF) cavities around the ring, which will group the protons in bunches and keep them at a constant energy.

There are eight RF cavities per beam/direction maintaining the protons tightly bunched and hence, guaranteeing high luminosity at the collision points. The LHC uses superconducting cavities with small energy losses and large stored energy. These metallic chambers

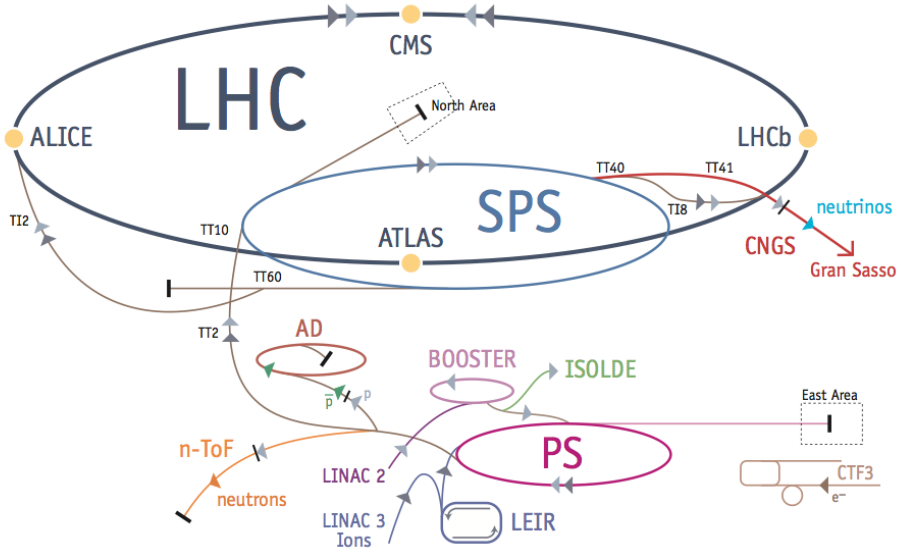


Figure 3.1: Scheme of CERN's accelerator complex.

were designed to make the oscillating electromagnetic field inside resonate. The oscillating field is tuned such that a proton with the desired energy suffers zero acceleration, while protons with slightly different energy values, arriving later or earlier, will be accelerated to match the desired particle energy in the beam. Each cavity delivers 2 MV (an accelerating field of 5 MV/m) at 400 MHz and operate at 4.5 K. These are arranged in groups of four in cryomodules, with two cryomodules per beam, and installed in a long straight section of the machine.

In order to bend and keep the beams in the circular trajectories, superconducting magnets are used. In the LHC 1232 dipole magnets, each 14.3 m long, are used to bend the beams around the ring and 392 quadrupole magnets keep the beams focused. These are copper-clad niobium-titanium (NbTi) type-II superconducting magnets which operate at an average temperature of 1.9 K, kept by using several tones of superfluid liquid helium He-II insulated from the exterior of the pipe by a vacuum vessel. Figure 3.2 shows a transverse view of the two pipes where the hadrons circulate and the superconducting coils. The superconducting dipole magnets produce two different magnetic fields in opposite directions, properly bending each beam. At the nominal beam energy of 7 TeV, the dipole magnets will produce a magnetic field of 8.3 T with electric currents of 11700 A.

The instantaneous luminosity of the collisions will depend on the beam properties for the

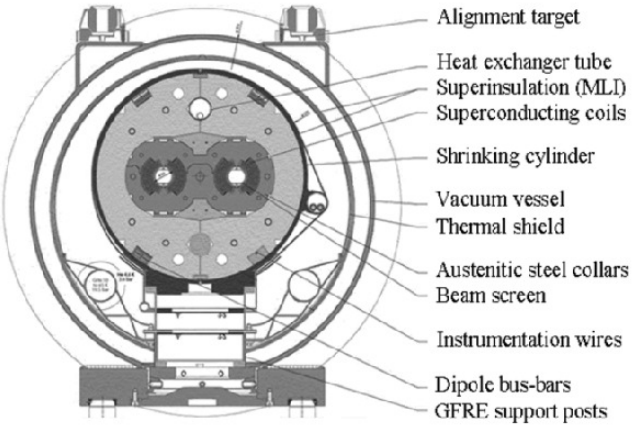


Figure 3.2: Diagram showing the cross-section of an LHC dipole magnet with cold mass and vacuum chamber.

different runs of the LHC. At designed conditions, the beams are expected to be composed by 2808 bunches of protons, 25 ns apart, leading to a collision rate of 40 MHz and a design instantaneous luminosity of 10^{34} cm^{-2}s^{-1}.

3.2.1 Luminosity and other Beam Parameters

The performance of a collider, as in the case of the LHC, can be assessed through the beam energy and luminosity of the accelerator. During a collision, the probability of a given process to occur will vary with the center of mass energy available. When studying interesting processes, which are often very rare (with small production cross-section), the accelerator must provide high beam energy and large number of useful interactions. The quantity that measures the ability of a particle accelerator to produce the required number of interactions is the instantaneous luminosity, $\mathcal{L}$:

$$\frac{dN}{dt} = \mathcal{L} \times \sigma \,, \tag{3.1}$$

which represents the proportionality factor between the number of events per unit of time dN/dt and the cross-section σ of the total collision.

When two particle beams collide, their characteristics, such as the beam density distribution, will define the collider's instantaneous luminosity [172]:

$$\mathcal{L} = \frac{n_b f_r n_1 n_2}{4\pi\sigma_x\sigma_y} \,, \tag{3.2}$$

where f_r represents the revolution frequency around the LHC ring, n_b is the number of bunch pairs crossing at the interaction point in each revolution, n_1 and n_2 are the number of protons in the two LHC colliding bunches, and σ_x and σ_y characterise the widths of the horizontal and vertical beam profiles. Since the currents of both LHC beams are measured, the luminosity can be determined simply by measuring σ_x and σ_y using van der Meer scans [173].[1]

Besides the van der Meer scan, the LHC relies on several different methods for luminosity determination, including standardised luminometers at each interaction point, luminosity monitors based on neutral particle flux counters, as well as the independent luminosity measurements made by each experiment.

In September 2008 the first proton beams circulated in the main ring of the LHC accelerator. In March 2010 the LHC started its operation at the highest centre-of-mass energy ever reached, with the first proton-proton (pp) collisions occurring at 3.5 TeV per beam. The collider and the associated experiments had a successful operation during Run 1, with the pp physics programme proceeding with center-of-mass energies of $\sqrt{s} = 7\,\text{TeV}$ and $\sqrt{s} = 8\,\text{TeV}$ during 2010–2011 and 2012, respectively. The world record on the instantaneous luminosity was reached with a luminosity of $4.7 \times 10^{32}\,\text{cm}^{-2}\text{sec}^{-1}$ in April 2011, and in August 2012 the LHC achieved a maximum instantaneous luminosity of $7.73 \times 10^{33}\,\text{cm}^{-2}\text{sec}^{-1}$. The LHC experiments also worked exceptionally well, being able to record the delivered luminosity with efficiencies of the order of 94%. During 2010 and 2011, with pp collisions at $\sqrt{s} = 7\,\text{TeV}$, an integrated luminosity of $5\,\text{fb}^{-1}$ was recorded by ATLAS, whereas during 2012, at a center-of-mass energy of 8 TeV, a total of approximately $21\,\text{fb}^{-1}$ were recorded. Table 3.1 includes a summary of the relevant parameters of pp beams distributed from the LHC accelerator to the experiments during Run 1, as these are relevant for the work developed in this thesis. In addition to pp collisions, the LHC has also produced lead ion ($PbPb$) collisions with a per-nucleon center-of-mass energy $\sqrt{s_{NN}} = 2.76\,\text{TeV}$ and proton-ion (pPb) collisions with $\sqrt{s_{NN}} = 5.02\,\text{TeV}$, during this first Run.

In the first long shutdown (LS1) the accelerator was upgraded in order to work at the design energy and luminosity expected to be obtained in Run 2. From 2015 to 2018, the LHC had pp collisions at $\sqrt{s} = 13\,\text{TeV}$, with ever-increasing luminosity delivered to its

[1] This operation consists on measuring the observed event rate of the two LHC beams crossing each other for different displacement steps of known distances, first in the horizontal and then in the vertical direction. These scans of the transverse profile of the beam will yield two bell-shaped curves, with maximum rate at zero displacement, from which one extracts the values of σ_x and σ_y (*i.e.* the RMS of the double gaussian distributions).

Parameter	Design Value	2010	2011	2012
Beam Energy (TeV)	7	3.5	3.5	4
Max. num. bunches / beam	2808	368	1380	1380
Max. num. protons / bunch	1.15×10^{11}	1.2×10^{11}	1.45×10^{11}	1.7×10^{11}
Bunch Spacing (ns)	25	150	75/50	50
Peak Luminosity ($\mathbf{cm^{-2}s^{-1}}$)	1×10^{34}	2.1×10^{32}	3.7×10^{33}	7.7×10^{33}
Max. $< \mu >$	19	4	17	37

Table 3.1: Overview of the parameters for the LHC performance comparing the design values with their time evolution during the first long run operation. $< \mu >$ is the mean number of inelastic pp interactions per bunch crossing.

experiments. Due to the technical improvements made, a total luminosity of $156\,\text{fb}^{-1}$ was delivered. Heavy ion collisions were also obtained: $PbPb$ collisions at $\sqrt{s_{NN}} = 5.0\,\text{TeV}$, as well as pPb and Pbp collisions at both $\sqrt{s_{NN}} = 5.0\,\text{TeV}$ and $\sqrt{s_{NN}} = 8.02\,\text{TeV}$. The accelerator was upgraded again during the second long shutdown (LS2) and provided $40.7\,\text{fb}^{-1}$ of pp collisions data at $\sqrt{s} = 13.6\,\text{TeV}$ during 2022.

3.3 ATLAS Detector

The ATLAS detector [169], shown in Figure 3.3, is a general-purpose detector designed to measure the myriad of particles produced by pp and heavy-ion collisions. It has been designed to optimize the search of the Higgs boson and to probe the widest range of physics possible.

At the LHC, a typical hadron collision consists on the hard scattering between two incoming partons, resulting on a set of particles with specific energy and momentum. However, these particles are produced amongst a wide number of other particles originated from the soft interactions of the remaining partons (*i.e.* the underlying event). In order to extract information regarding the fundamental physics mechanism behind the interaction, the produced particles, or their decay products, must be identified and their kinematic properties thoroughly measured.

The ATLAS detector was thus designed to cover a broad energy and angular range, ensuring the observation of several fundamental and composite particles. It is a large and complex system, approximately 44 m long and 25 m in diameter, composed by many concentric cylindrical layers of sub-detectors around the beam pipe and two perpendicular end-caps at both ends, also layer-shaped. This way, ATLAS covers the maximum possible solid angle

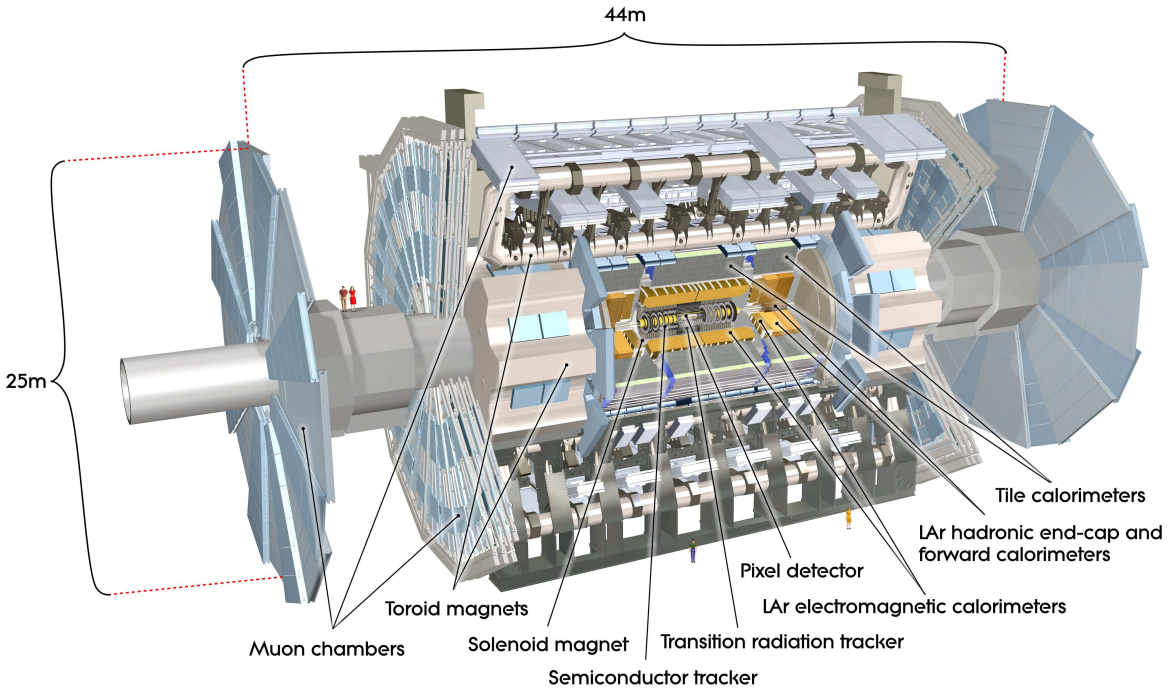


Figure 3.3: Schematic representation of the ATLAS detector.

around the interaction point and respects the forward-backward symmetry relative to the interaction point. The first hits of charged particles that interact with the detector material occur at the Inner Detector. The charged particle tracks are sampled under a magnetic field, allowing the measurement of their momentum and reconstruction of the collision vertices. Surrounding the Inner Detector are the hadronic and electromagnetic calorimeters, where most particles deposit all their energy. The Electromagnetic Calorimeter was designed to measure the energy of charged particles and photons through the electromagnetic interaction, while the hadronic calorimeter measures the energy deposited by hadrons. The last layer of the detector is the Muon Spectrometer and its associated toroidal magnet system, which allow the determination of the muons momenta from the deflection of their trajectories. In the following subsections, the sub-detector systems are described in more detail.

The coordinate system used at ATLAS is a cartesian right-handed reference system, with origin in the nominal interaction-point, as can be seen in Figure 3.4. It is defined in such a way that the positive x-axis points to the center of the LHC ring, the positive y-axis points upwards and the positive z-axis is tangent to the beam pipe. The z-axis defines the detector sides. The A-side (C-side) is defined as that of positive (negative) z. The azimuthal angle ϕ

Component	Resolution	η	*Trigger* (η)
ID	$\sigma_{p_T}/p_T = 0.05\% p_T \oplus 1\%$	± 2.5	
EM Cal.	$\sigma_E/E = 10\%/\sqrt{E} \oplus 0.7\%$	± 3.2	± 2.5
Hadronic Cal. barrel & end-cap	$\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$	± 3.2	± 3.2
Hadronic Cal. forward (FCal)	$\sigma_E/E = 100\%/\sqrt{E} \oplus 10\%$	$3.1 < \lvert\eta\rvert < 4.9$	$3.1 < \lvert\eta\rvert < 4.9$
Muon Det.	$\sigma_{p_T}/p_T = 10\%$ at p_T=1 TeV	± 2.7	± 2.4

Table 3.2: General detector performance, extracted from Ref. [169]. Notice that for high-p_T muons, the spectrometer performance is independent of the Inner detector system. Both E and p_T are in GeV.

can be obtained from the projections of the object momentum along the y-axis and z-axis, p_y and p_z respectively:

$$\phi = \arctan(p_y/p_x)\,. \tag{3.3}$$

Ranging from $-\pi$ to $+\pi$, ϕ is measured around the beam axis, so that $\phi = 0$ corresponds to the x-axis. The polar angle θ ranges from 0 to π and is defined as the angle between the positive z-axis and the object momentum.

In hadron colliders it is often used a convenient coordinate transformation,

$$\eta \equiv -\ln\big(\tan(\theta/2)\big), \tag{3.4}$$

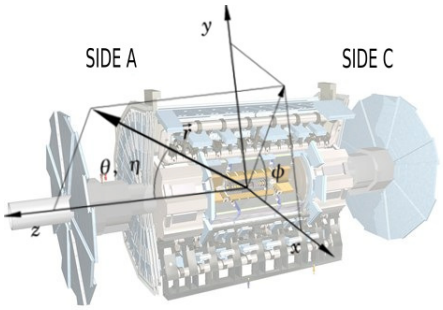


Figure 3.4: The ATLAS coordinate system.

where the pseudorapidity η replaces θ to describe the angle of the direction of movement of a particle with respect to the beam line. It ranges from zero, when the particle travels perpendicularly to the beam pipe, and infinity when it is parallel. In the massless particle limit, the variable η is equivalent to the relativistic rapidity y, which is defined in natural units as,

$$y \equiv \frac{1}{2} \ln(\frac{E + p_z}{E - p_z}) \,, \tag{3.5}$$

where E is the energy of the particle and p_z its momentum along the beamline. It should be noticed that rapidity differences Δy are boost-invariant along the z-axis. However, since it only requires knowing the particle polar position, the pseudorapidity is generally preferred. The distance ΔR in the pseudorapidity-azimuthal angle space is defined as

$$\Delta R^2 \equiv \Delta\eta^2 + \Delta\phi^2 \,. \tag{3.6}$$

Other common variables at ATLAS are the transverse momentum $p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2}$ and transverse energy $E_{\mathrm{T}} = E \sin\theta$ of a particle, which are measured in the xy plane, perpendicular to the beam pipe. Not only are both variables boost-invariant along the z-axis, but are also very convenient since the transverse component of the initial momentum of the colliding partons is zero.

Due to the nature of the LHC collisions, several experimental challenges to the identification of interesting physics signatures must be overcome. During the ATLAS design [169] a set of requirements were taken into account in order to enhance its performance. An essential characteristic of the detector is the already mentioned broad solid angle coverage, via the wide $\eta - \phi$ range obtained with the cylindrical shape structure. Given the high energy and collision rate at the LHC, the ATLAS detector has fast and radiation resistant electronics. The ATLAS detector also has high granularity in order to handle the expected high particle fluxes and multiplicities and to reduce the influence of multiple simultaneous collisions (pile-up). The general ATLAS detector performance goals are summarised in Table 3.2.

3.3.1 Magnet System

The ATLAS magnet system [174] features a unique hybrid system of four large superconducting magnets, which provide the magnetic field, orthogonal to the particle trajectory, over a volume[1] of 12000 m^3. It consists of one central solenoid, generating the magnetic

[1] Defined as the region in which the field exceeds 50 mT.

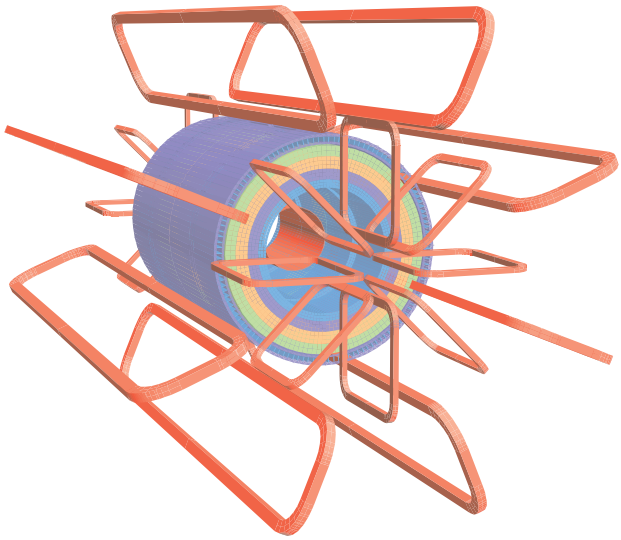


Figure 3.5: Layout of the ATLAS magnet system, with the toroidal coils and central solenoid represented in red. The Hadronic Tile Calorimeter steel is also represented in four layers of different magnetic properties [169]).

field for the Inner Detector, and three large air-core toroids (a barrel and two end-caps), producing the magnetic field for the muon spectrometer. Figure 3.5 shows a general layout of the superconducting magnet system with 22 m in diameter and 26 m long. The overall system has a stored energy of 1.6 GJ and operates at 4.5 K for optimal performance, therefore requiring a specific cooling systems, obtained with the resource to cryostats. The geometry of this huge system has influenced the ATLAS structure, size, and even name. It is fundamental for the measurement of any electrically charged particle momentum: particles produced in each event are deflected in the magnetic field, leading to the measurement of both their momenta and charge sign.

The inner solenoid [175] provides a 2 T magnetic field parallel to the beam pipe, in the tracking system region, with a peak magnetic field of 2.6 T at the superconductor itself. The solenoid is a cylindrical shell, 10 cm thick, extending over 5.3 m long and with a bore of 2.4 m. It is located between the barrels of the Inner Detector and the Electromagnetic Calorimeter, aligned with the beam axis. The minimisation of the radiative thickness in front of the electromagnetic calorimeter was mandatory, as to permit the desired calorimeter performance. Consequentially, both the electromagnetic calorimeter barrel and the solenoid are housed in the same vacuum vessel, thus eliminating two vacuum walls. With a stored

energy of 39 MJ, the central solenoid nearly uniform magnetic field can bend even very energetic particles and allows the precise measurement of their momentum by the tracking system. The detector system has, however, less sensitivity to low energy particles' momenta (of the order of hundreds of MeV).

The toroid system generates the magnetic field needed to deflect particles in the η direction at the Muon Spectrometer. It consists of a barrel toroid surrounding the calorimeters and two end-cap toroids inserted at both extremes of the detector and lining up with the central solenoid. Each of the three toroids is composed of eight super-conducting coils disposed radially and symmetrically around the beam pipe. On the barrel region, the 25.3 m toroid has an inner (outer) diameter of 9.4 m (20.1 m), while the end-cap toroids are 5 m long, with an inner (outer) diameter of 1.65 m (10.7 m). The end-caps are rotated by 22.5° in relation to the barrel toroid, leading to the overlap of the magnetic fields and optimal bending power in the interface regions of both coil systems. Contrary to the solenoidal magnetic field, the toroidal magnetic field is highly non-uniform. In the Muon Spectrometer, the magnetic field intensity ranges from 0.2 to 2.5 T in the barrel region ($|\eta| < 1.0$) and about 0.2 to 3.5 T in the end-caps ($1.4 < |\eta| < 2.7$). The open air toroid configuration was chosen to improve the muon reconstruction performance independently from the Inner Detector. Each coil is made of essentially the same Al-stabilised Nb/Ti/Cu conductor, winded into a pancake-shaped roll and kept in stainless-steel vacuum vessels. All the barrel coils are housed in individual cryostats taking up the forces between them, whereas the end-cap coils are kept together in a single cold mass with a rigid structure to withstand the Lorentz forces. The open toroid magnets efficiently provide the desired magnetic field over a large volume and do so by requiring a remarkably reduced amount of material. This is particularly important for the muon momentum resolution, since it is highly affected by muons multiple scattering.

3.3.2 Inner Detector

The Inner Detector (ID) [176, 177, 178] is the closest subdetector layer to the IP. It was designed as a tracking system used to identify and measure the properties of electrically charged particles arising from beam collisions. This detector provides information on the particle's position with enough granularity for a track to be identified and the interaction vertex to be reconstructed. Due to the high density of incoming particles and its closeness to the beam pipe, the ID is required not only high granularity, but also fast response electronics and good radiation resistance to successfully operate in the LHC environment.

Outside of the ID, the central superconducting solenoid provides a magnetic field parallel to the beam axis, curving the trajectories of electrically charged particles in the ϕ direction and allowing for momentum and charge sign measurement. As a result of the finite size of the solenoid, the magnetic field inside the ID is constant in the radial direction and decreases along the beam direction.

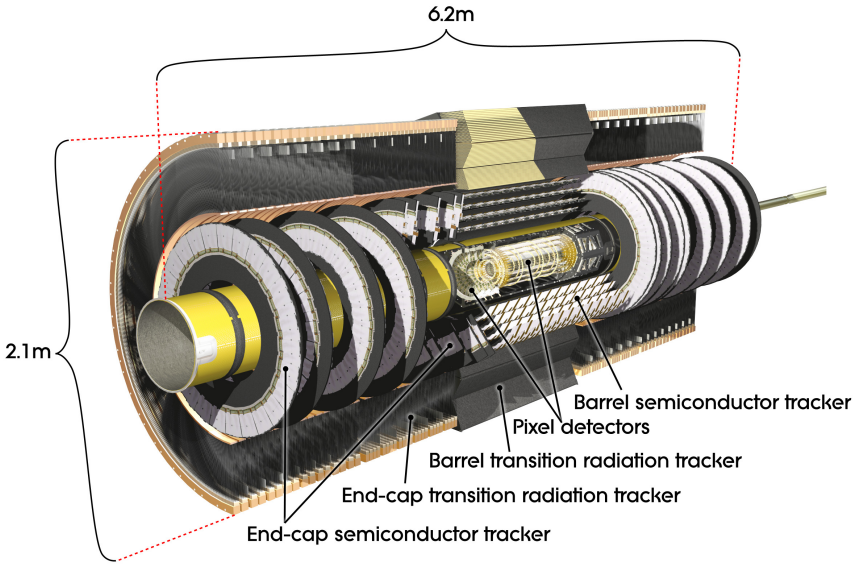
The Inner Detector is 6.2 m long and has a radius of about 1.1 m. Engineered to have the minimum amount of material before the calorimeter system, it allows the measurement of particle energy without degrading it. As illustrated in Figure 3.6, the ID is composed of three complementary sub-detectors: the Pixel Detector, the Semiconductor Tracker (SCT) and the Transition Radiation Tracker (TRT). In the barrel region, they are arranged in concentric cylinders around the beam axis while at the end-cap regions they are located on disks perpendicular to the beam axis.

Pixel Detector

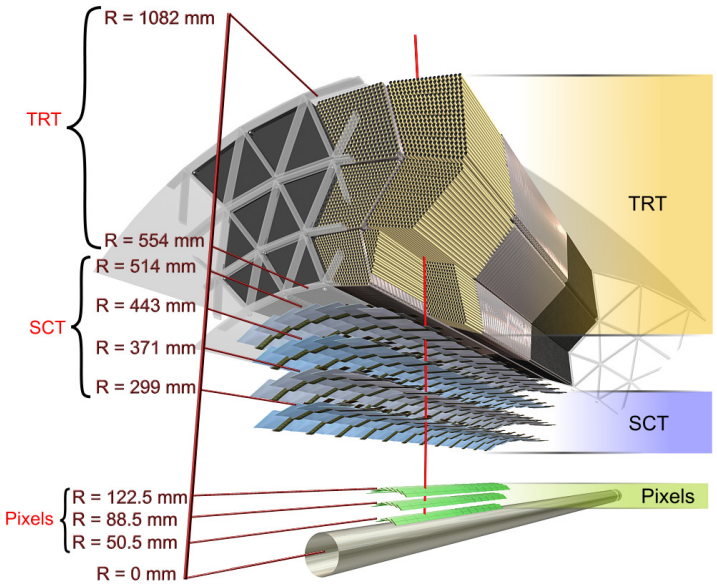
The innermost part of the tracking system is the Pixel Detector. It is composed by pixel sensors arranged in three cylindrical layers in the barrel region and five circular disks in both end-cap regions, which provide tracking coverage over $|\eta| < 2.5$. The fine granularity is achieved through the use of silicon wafers with very small radiation-hard pixel detectors ($50 \times 400\,\mu\text{m}^2$), resulting in ~ 80.4 million read-out channels. The intrinsic accuracies of the pixel detectors are $\sigma_{R-\phi} = 10\,\mu$m and $\sigma_z = 115\,\mu$m in the barrel, and $\sigma_{Z-\phi} = 10\,\mu$m and $\sigma_R = 115\,\mu$m in the end-caps. It is the most finely segmented of the ID components and typically provides three measurement points over the full acceptance, mostly determining the impact parameter resolution and the high precision vertex reconstruction. In particular, the very first layer (or B-layer), only ~ 4 cm away from the beam-interaction region, enables the precision measurement of secondary vertices associated with the production of long-lived particles such as B hadrons and τ leptons. In the busy environment of the LHC, this is specially useful to identify jets originating from the fragmentation of heavy flavoured b-quarks.

Semiconductor Tracker

The Semiconductor Tracker (SCT) is a silicon microstrip detector located in the middle section of the ID. The basic SCT unit is a pair of single-sided silicon microstrip sensors mounted back-to-back. The SCT is composed of a central barrel region, with four layers of



(a)



(b)

Figure 3.6: Schematic view of the ATLAS Inner Detector (a) and detail of the sensors and structural elements traversed by a charged track of 10 GeV p_T in the barrel region (b) [169].

strips, and two end-caps, each with nine disks arranged perpendicularly to the beam pipe. Each barrel layer consists of 768 microstrips rotated by a 40 mrad tilt angle with respect to the previous set, covering the full azimuthal angle. The strips have a 80 μm pitch and are 12 cm long. This layout creates overlapping between adjacent strips (c.f. Figure 3.6(b)), which is used to determine the space point position when both strips are hit by a traveling charged particle. In the barrel the silicon strips are arranged parallel to the beam line, while in the end-cap disks the strips are oriented radially. The radial range of each end-cap disk was adapted to obtain the optimal coverage of $|\eta| < 2.5$. The SCT has approximately 6.2 million readout channels and typically provides eight precision measurements per track (four space-points) in the intermediate radial range. The sensors intrinsic accuracies are 17 μm in $R-\phi$ and 580 μm in $z(R)$ for the barrel (end-cap) region. The SCT contributes to the measurement of momentum, impact parameter and vertex position. Due to their high granularity, both the SCT and Pixel Detector contribute to good pattern recognition and are referred to as the Precision Tracker.

Transition Radiation Tracker

The Transition Radiation Tracker (TRT) is the outermost subdetector layer of the ID. It consists of gaseous drift tubes (or straws) with a gold plated tungsten wire in the centre. The TRT operates at very high rates by virtue of the straw small diameter (4 mm) and the isolation of the sense wires within individual xenon-based gas volumes. The tubes are interleaved with polyethylene which acts as a radiator, producing transition radiation photons when relativistic charged particles cross the boundary between the different materials. The emission of photons will depend on the relativistic Lorentz boost γ (E/m) of the particle, thus allowing the discrimination between lighter and heavier particles. At the LHC energy range, the TRT was optimised for tracking at the expense of electron identification. In the barrel, the straw tubes are 144 cm long and oriented along the beam axis, while in the end-caps they are 37 cm long and disposed in a fan-layout. The total number of TRT read-out channels is about 351 thousand, each providing a drift-time measurement, with an azimuthal resolution of 170 μm per straw and two independent thresholds. The independent thresholds allow the discrimination between tracking hits (passing the lower threshold) and the transition radiation hits (passing the higher one). This technique is intrinsically radiation hard, and allows a large number of measurements per track (typically 36 points) continuously tracking the particle. The TRT covers the $|\eta| < 2.0$ range and the high density

of measurements in the outer part of the tracker contribute significantly to the momentum measurement.

The overall performance of the Inner Detector is very robust, with an acceptance in pseudorapidity of $|\eta| < 2.5$, and full coverage in ϕ. The combination of techniques used in each subdetector gives high precision in both ϕ and z coordinates and contribute to good pattern recognition. The transverse momentum resolution of the ID is included in Table 3.2. The relative precision of the three subdetectors is comparable, so that no single measurement dominates the momentum resolution. Besides the good resolution in the charged particles momentum, the Inner Detector also contributes to good jet reconstruction efficiency.

3.3.3 Calorimetry

The ATLAS calorimeter system [179] sits outside the ID and consists of two different type of detectors for energy measurements, the electromagnetic calorimeters and the hadronic ones. The system was designed to stop and contain the development of electromagnetic and hadronic particle showers (or jets) originating from the interaction of traveling particles with the medium. The energy measurement is performed by instrumented active material, which samples the energy loss in the dense absorber mediums.

The electromagnetic calorimeters cover the $|\eta| < 3.2$ range and serve the purpose of identifying and measuring the energy of electromagnetic-interacting particles, such as electrons and photons. On the other hand, the hadronic calorimeters were designed to identify and measure the energy of the hadrons that already crossed the electromagnetic calorimeters, covering the region of $|\eta| < 4.9$. Figure 3.7 is a cut-away view of the ATLAS calorimeter system.

In the acceptance region of the precision tracker, the electromagnetic calorimeter is finely segmented for accurate measurement of photons and electrons. Even though the hadronic and forward calorimeters have coarser segmentation, the precise measurement of jet kinematics is still available, and the pseudorapidity coverage is sufficient for the missing transverse energy calculation. The ATLAS calorimeters not only have to contain both electromagnetic and hadronic showers, but should also guarantee that the particles reaching the muon detectors are limited to muons and undetectable neutrinos.

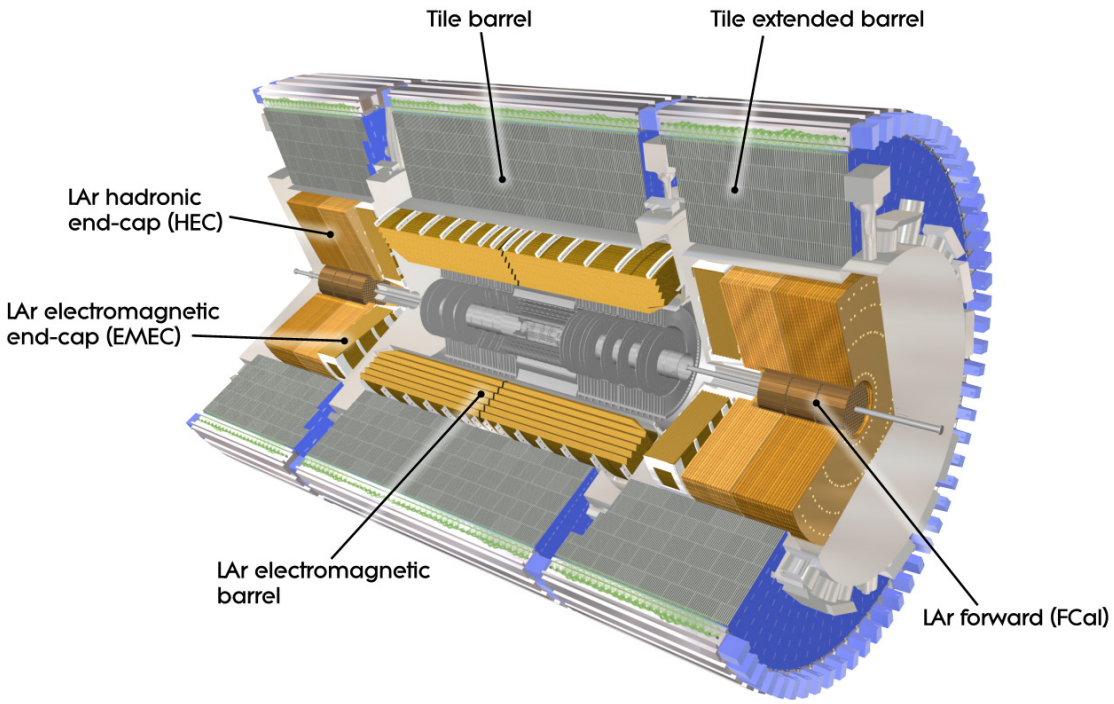


Figure 3.7: Cut-away view of the calorimeter system [169].

This system is formed by several sampling detectors, maintaining azimuthal symmetry and coverage around the beam axis. There are three cryostats that contain the calorimeters closest to the beam-pipe, one barrel and two end-caps, just like most sub-systems of the ATLAS detector. The barrel cryostat contains the electromagnetic barrel calorimeter, whereas the two end-cap cryostats each house an electromagnetic end-cap calorimeter (EMEC), followed by a hadronic end-cap calorimeter (HEC) and, covering the region closest to the beam, a forward calorimeter (FCal). The outermost subdetector is a cylindrical hadronic tile calorimeter (TileCal), also assembled as a central barrel and two end caps.

Electromagnetic Calorimeter

The Electromagnetic Calorimeter (ECal) [180, 181] is a sampling calorimeter composed of alternating layers of lead plates and liquid argon (LAr). The lead layers are disposed in an accordion shaped geometry, oriented in the radial direction and act as an absorber medium. The lead plates thickness was chosen to optimise the performance in energy resolution, and

range from 1.1 to 2.2 mm thick. The liquid argon solution distributed between the plates is the active material of the calorimeter. It was adopted for its intrinsic linear behaviour, high ionisation yield, stability and resistance to radiation. A cryogenic system maintains the LAr at a temperature of 88 K. The major cryostat components are housed between the ECal and the TileCal subdetectors. Pairs of ions and electrons, originated from interactions of crossing charged particles in the LAr, are accelerated and drift in opposite directions due to the presence of an electric field between the absorber plates. The ionisation pairs are then collected by kapton electrodes, leading to the signal amplification.

The accordion-like geometry provides complete azimuthal symmetry and coverage, without any dead paths between the different components. The ECal barrel covers the region with $|\eta| < 1.475$ and the two end-caps (EMEC) reach the $1.375 < |\eta| < 3.2$ region. Segmented in three longitudinal sections (c.f. Figure 3.8(a)), the ECal barrel has a cell structure in the $\eta \times \phi$ plane with higher resolution in the inner layers, as to precisely measure the particles η direction. The third and top layer has a coarser granularity and contributes to the estimation of the amount of energy lost beyond the ECal. The first two layers are of particular importance for photon and electron identification, and can be used to obtain information on the photon's production vertex. In particular, the second layer has a 0.025×0.025 granularity which provides the position measurement of the energy deposition cluster. In addition, an active LAr presampler layer is used to recover information on the energy lost by upstream particles, which started showering before reaching the calorimeter. Located inside the cryostat wall and covering the $|\eta| < 1.8$ region, the presampler has a thickness of 1.1 cm (0.5 cm) in the barrel (end-cap) region.

The calorimeter performance is obviously affected by the amount of material that the particles must cross before reaching it. In order to minimise the effect on the energy measurement, the ID radial depth has been kept to a minimum and a common vacuum vessel lodges both the central solenoid and the LAr calorimeter. The ECal has a required relative energy resolution of [181]

$$\frac{\sigma E}{E} = \frac{10\%}{\sqrt{E}} \oplus \frac{170\,\text{MeV}}{E} \oplus 0.7\%\,, \tag{3.7}$$

with E measured in GeV. The parametrisation of the noise term, 170 MeV/E, is the target value without pile-up. Specific regions of the calorimeter, however, are usually not used in precision physics involving photons, due to the particular quantities of material in the particle's path. For instance, for $1.37 < |\eta| < 1.52$, there is a significant amount of material,

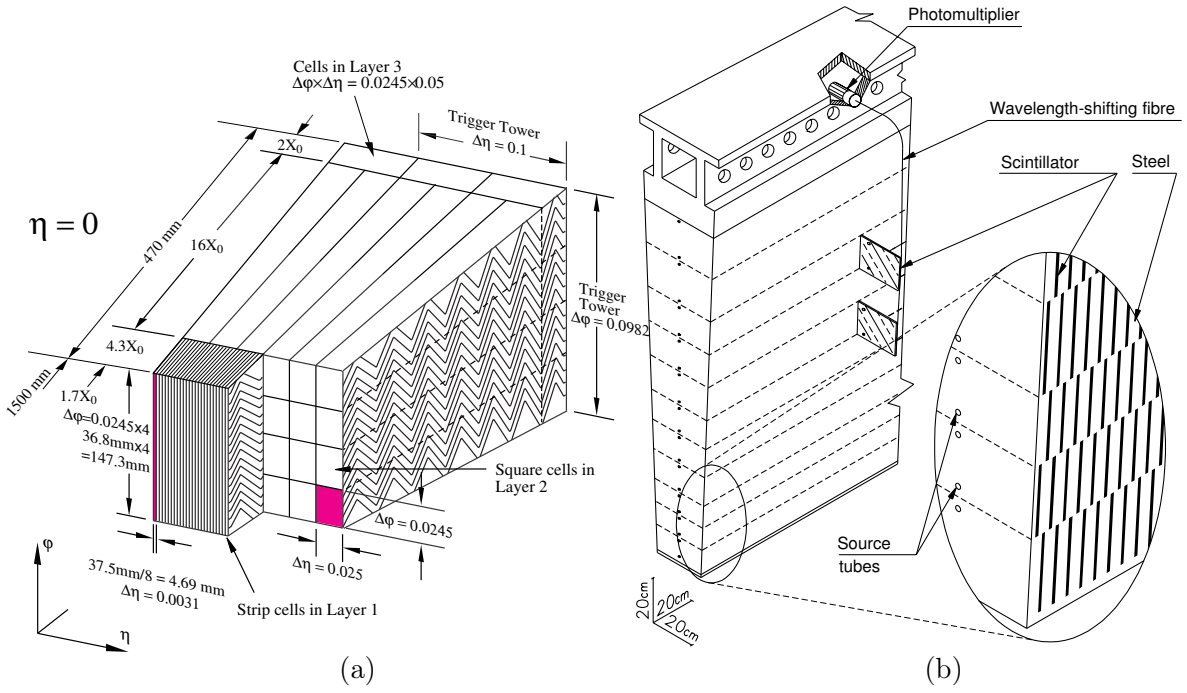


Figure 3.8: Detail sketches of (a) ECal and (b) TileCal barrel modules. The different structured layers are clearly visible, as well as the readout methods: the ECAL ganging of electrodes in ϕ and the TileCal fibres and photomultipliers in η [169].

corresponding to cables and service structures going to the ID. The total thickness of the ECal barrel is 1.2 m which guarantees a full containment of electrons and photons up to energies of a few TeV.

Hadronic Calorimeter

The ATLAS hadronic calorimeter is meant to detect the particle shower that results of the quark hadronisation and, after calibration, infer the energy of the original particle. Its composed of three different independent subdetectors: the Tile Calorimeter (TileCal), the Hadronic End-Cap LAr Calorimeter (HEC) and the Forward LAr Calorimeter (FCal). Each sampling calorimeter uses different techniques and materials best suited for the widely varying conditions in terms of radiation over the large pseudorapidity range ($\eta < 4.9$). Hadronic showers will typically penetrate further in the material than those dominated by the electromagnetic interaction and also result in broader shower shapes. The hadronic calorimeters are, therefore, located immediately after the electromagnetic ones.

In the most forward regions of the ATLAS calorimeter system, given the proximity to the beam, LAr calorimeters were selected for its radiation hardness. Housed in the two end-cap cryostats, HEC covers $1.5 < |\eta| < 3.2$ and FCal $3.1 < |\eta| < 4.9$ (c.f. Figure 3.7).

The Hadronic End-Cap calorimeter (HEC), is a copper sampling calorimeter with a flat-plate design. Each HEC is composed of two independent wheels (HEC1 and HEC2) with varying granularity and segmented longitudinally in layers. Closer to the IP in $1.5 < |\eta| < 2.5$, the HEC1 wheel has $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ in the first two longitudinal layers and 0.2×0.1 in the last one. This pseudorapidity region is used for precise measurements of energy and angles of jets and, at low luminosity, of single charged particles. For the HEC2 wheel, in the range $2.5 < |\eta| < 3.2$, the granularity is less refined: $\Delta\eta \times \Delta\phi = 0.2 \times 0.2$ in all the three samples. The HEC is particularly important due to its ability to detect muons and measure any radiative energy loss.

The Forward Calorimeter (FCal) covers the most forward region in pseudorapidity, located inside the end-cap cryostat, directly surrounding the beam pipe, with a front face at about 4.7 m from the interaction point. Given its vicinity to the IP it is bound to receive intense particle fluxes, and therefore, must be radiation resistant. Nonetheless, the integrated FCal allows uniform η coverage, as well as reduced background radiation levels in the muon spectrometer. Similarly to the other LAr calorimeters, it consists of layers of an absorber and the LAr in between as the sensitive medium. The FCal is divided into three modules according to the different passive materials used. The first module uses copper as passive medium, while the other two use tungsten. Each section of the calorimeter consists of a copper matrix with regularly spaced longitudinal channels filled with concentric metallic rods and tubes. The tubes and matrix are grounded, while the rods have positive high voltage. The LAr gaps are very small to avoid ion build-up problems and to provide the highest possible density. The tungsten rods were chosen in order to provide containment and minimise the lateral spread of hadronic showers. The FCal layout, as well as its coupling with the HEC wheels, minimizes the energy losses in cracks between the calorimeters, and is crucial for the good hermeticity of the ATLAS calorimeter system. Furthermore, a shielding plug made of a copper alloy, located behind the third FCal section, reduces the background

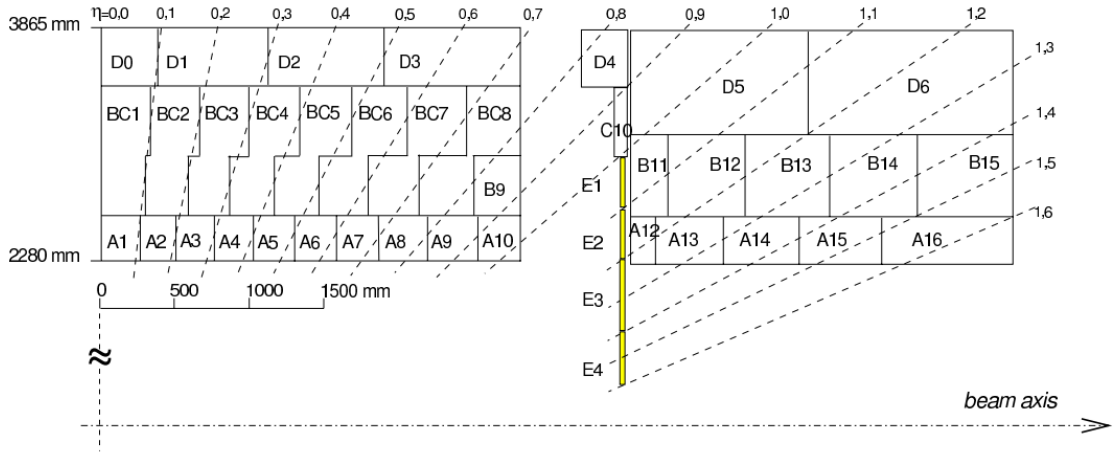


Figure 3.9: Cells and layers of the hadronic calorimeter TileCal. The regions between dashed lines cover $\Delta\eta = 0.1$ region. The gap-crack or special cells (E layer) are represented by the yellow shaded regions.

radiation in the end-cap muon system.

The Tile Calorimeter (TileCal) [182, 183] is a sampling calorimeter made of polystyrene scintillating plates (designated tiles) as active material, intercalated by low carbon steel (iron) plates as absorbing material. It covers the region $|\eta| < 1.7$ and is the first subdetector layer covering the ECal calorimeter, extending from an inner radius of 2.28 m to an outer radius of 4.25 m. The TileCal is divided into four cylindrical partitions: two long barrels (LB) and two extended barrels (EB). The different partitions are arranged along the beam axis, covering different pseudorapidity angles: the long barrels (LBA and LBC) cover the $|\eta| < 1.0$ region and the extended barrels (EBA and EBC) cover $0.8 < |\eta| < 1.7$ region. The partitions are named according to the pseudorapidity angle sign, and by convention, A is positive η and C negative η. Each of these partitions is further segmented into 64 azimuthal wedges, referred as modules, with $\Delta\phi$ granularity of $\Delta\phi = 2\pi/64 \sim 0.1$ radians. Figure 3.8(b) shows a detailed scheme of one TileCal barrel module. Radially, each module is segmented in three sampling layers of different depths (A, BC and D), with the alternating plates perpendicular to the beam direction. Designed to measure the energy deposition of strongly-interacting particles, the tiles will emit light when particles pass through them. The mechanical structure of TileCal includes the electronic and read-out systems, in order to collect the light produced.

When ionising particles cross the tiles, ultraviolet scintillation light production is induced in the base material. This light is collected and converted to visible light at the ϕ edges of each tile by wavelength shifting (WLS) optical fibers, which are bundled together to form readout cells. Each cell is defined with an η coverage of 0.1, having different sampling depths depending on the layer. The TileCal cells form a pseudo-projective geometry in η towards the interaction region, as shown in Figure 3.9. The bundles of fibres in each side of the scintillating tiles are read out by two separate channels to provide redundancy. Since each TileCal cell is read out by two independent photomultiplier tubes (PMT) each linked to one readout channel, every module will contain upto 48 readout channels, which are stored in metallic cases called drawers at the outer radius of the calorimeter. The transition region between the LB and the EB contain cells for specialised use, which enable the partial recovery of energy lost in the crack regions of the detector. These are single scintillator cells located in the plug extension and in the inter-cryostat regions, as can be seen represented in the yellow shaded cells in Figure 3.9. These cells are used for example in Cesium calibrations, and are labelled gap-crack or special cells, constituting the E layer.

With a total of 4672 cells, each read out by two PMTs, the TileCal comprises approximately 10000 PMT channels in the entire calorimeter. The TileCal was designed to have good time resolution ($\sim 1\,\text{ns}$) and a typical granularity of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ (0.2×0.1 for the last layer) in order to achieve good resolutions in jet energy. The performance of the TileCal cells with increasing pile-up conditions is discussed in Chapter 4, as well as a more detailed description of the calorimeter read-out system.

The ATLAS hadronic calorimeter system has a total radiation length[1] of about 11λ at $\eta = 0$, ensuring good containment of hadronic showers and very limited punch-through into the muon system. It has a required energy resolution of

$$\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E}} \oplus 3\%\,, \tag{3.8}$$

on the barrel region and

$$\frac{\sigma_E}{E} = \frac{100\%}{\sqrt{E}} \oplus 10\%\,, \tag{3.9}$$

in the end-caps. The hadronic calorimeter provides good resolution for high energy jets and, due to its large η coverage, guarantees accurate missing transverse energy measurement.

[1]Here, the characteristic absorption length λ is the nuclear interaction length describing the interaction of pions and kaons with the material.

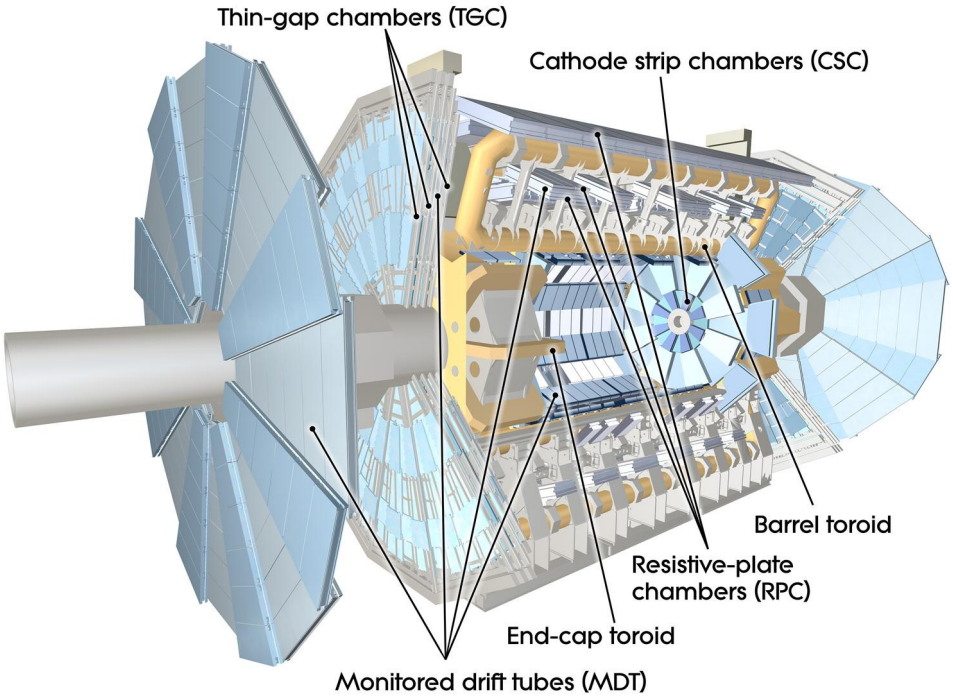


Figure 3.10: Overview of the muon chambers.

3.3.4 Muon Spectrometer

The Muon Spectrometer (MS) is the outermost system of ATLAS subdetectors [169, 184, 185] and was designed to identify and precisely measure high momentum muons. Based on the magnetic deflection of muon tracks, it consists on the combination of superconducting air-core toroid magnets and precision tracking chambers, instrumented with a separate trigger system. Before reaching the MS, particles originated in the collisions have to cross all the other ATLAS subdetectors, as well as the central solenoid, the cryogenic systems, cable structures and radiation shields. Thus, the large radiation length of the previous systems will assure that muons are the only charged particles expected to reach the spectrometer. The muon curvature, due to the magnetic field, is then measured by tracking detector chambers, using different technologies, chosen according to the expected particle flux and detector needs.

The ATLAS Muon Spectrometer and toroidal magnet system are shown in Figure 3.10. This detector extends from a 4.25 m radius around the calorimeters to the outer radius of the

detector ($\sim 12\,\mathrm{m}$) comprising several different structures immersed in the toroidal magnetic field. The precision-tracking chambers design and layout is such that particles originating from the interaction point traverse three layers of chambers, allowing the momentum determination from the sagitta of the trajectory. In the barrel region, these chambers are arranged in three cylindrical layers (or stations) around the beam axis with approximately 5, 7.5 and 10 m radius, located before, inside and after the eight coils of the toroid magnet respectively. In the transition and end-cap regions, however, the MS chambers are disposed vertically in front and behind the two end-cap toroid magnets, with the end-cap regions also segmented in three stations ($|z| \approx 7.4\,\mathrm{m}$, $10.8\,\mathrm{m}$, $14\,\mathrm{m}$, and $21.5\,\mathrm{m}$ from the interaction point). Figure 3.11 shows a detailed layout of the cross view of the Muon Spectrometer barrel. The structure of the muon chamber system, consisting of a 16-fold ϕ segmentation arranged in eight octants, is a reflection of the azimuthal symmetry of the toroids (c.f. section3.3.1). Each octant is arranged in a large and a small sector, leading to a region of overlap in ϕ. The large sectors cover the $\Delta\phi$ regions between the barrel toroid coils, whereas the small sectors cover the azimuthal range around or inside these open-aired coils. The azimuthal overlap of the chamber boundaries allows the relative alignment of adjacent sectors using tracks recorded by both large and small sectors, while minimising gaps in the spectrometer coverage.

The magnet configuration provides a magnetic field mostly orthogonal to the muon trajectories, deflecting them along the θ angle. In the barrel region ($|\eta| < 1.0$), the field intensity ranges from 0.2 to 2.5 T whereas in the end-caps ($1.4 < |\eta| < 2.7$) it fluctuates from 0.2 to 3.5 T. Over the transition region ($1.0 < |\eta| < 1.4$) the magnetic deflection is provided by a combination of barrel and end-cap fields. Due to the orthogonality of the magnetic field and the muon trajectories, the transverse momentum resolution is roughly independent of η over the whole detector acceptance. This magnetic layout also offers minimal degradation of the momentum resolution due to multiple scattering. The MS has coverage in the pseudorapidity $|\eta| < 2.7$, except around $|\eta| = 0$, where there is a 300 mm hole, allowing the passage of cables and services for the ID detector, inner solenoid and calorimeters, which will significantly degrade the muon reconstruction in that area.

Resorting to different technologies of multi-wire chambers, the MS can provide high-precision position and momentum measurements in η and ϕ as well as efficient triggering, all the while targeting the large particle flux variation from the central to the forward region of the spectrometer. The muon chambers consist of four different subdetector types: Monitored

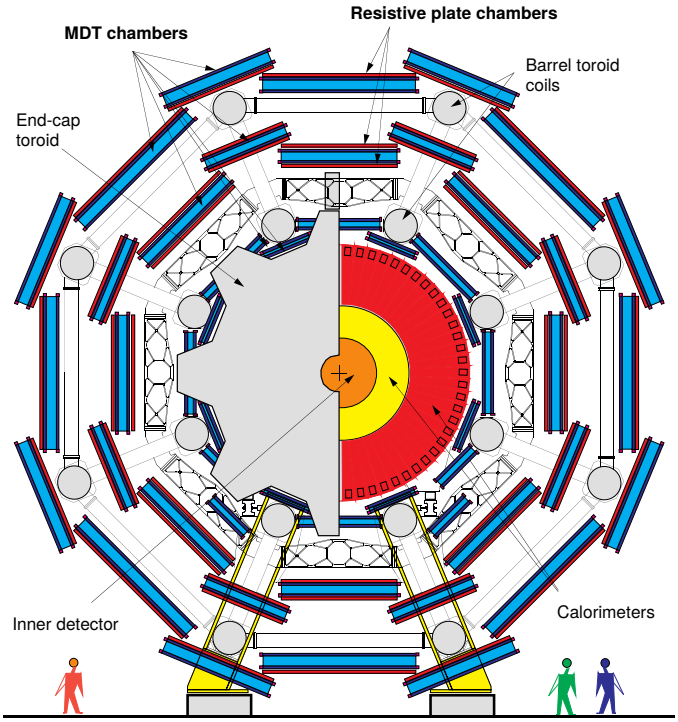


Figure 3.11: Detailed cross view of the barrel muon precision-tracking chambers [186]: the RPC's are coloured in red, the MDT chambers are represented in dark blue and in cyan there is the spacers in between them. The barrel of the ID and the Calorimeter systems is also depicted in the central part of the ATLAS detector.

Drift Tubes (MDT) and Cathode Strip Chambers (CSC) for tracking and Resistive Plate Chambers (RPC) and Thin Gap Chambers (TGC) for independent first event triggering (active in $|\eta| < 2.4$). The overall MS performance depends on the accurate alignment of the multi-wire chambers with respect to each other and with respect to the complete ATLAS detector. To this effect, the monitoring of the chamber positions rely on different strategies and sensor types through out the muon detector system.

The MDT cover most of the η-range, performing precision measurements of the track coordinates in the R-z projection and thus providing accurate muon momentum measurement. The passing muons will produce ionisation charges that are collected on the MDT wire and read out by an amplifier system at one end of the tube. The measurement of the minimum distance between the muon and the wire is obtained from the reconstructed drift time of the electrons resulting from the ionisation, allowing the reconstruction of the muon trajectory. The multiple tube layers improve the detector resolution and guarantee opti-

mal track reconstruction while reducing the fake track production from random background scattering.

At the end-caps in the most forward region, with pseudorapidities between 2.0 and 2.7, CSC with finer granularity are installed in the tracking layer closer to the beam pipe. The harsh background conditions and heavy particle fluxes of this region, particularly the high muon track densities, require the tracking chambers to have higher rate capability and time resolution. Designed to meet these requirements, the CSC consist in multi-wire proportional chambers that provide the simultaneous measurement of both η and ϕ coordinates from the interpolation between the induced charge distribution on neighbouring strips. Every ϕ sector contains four consecutive CSC rigid planes, resulting in four independent measurements along each track.

The trigger system covers the $|\eta| < 2.4$ region with Resistive Plate Chambers (RPC) used in the barrel and Thin Gap Chambers (TGC) in the end-caps (c.f. Figure 3.10. Both chambers provide first event triggering over the full azimuthal range, giving fast and coarse tracking information on muon traversing the detector, as well as evaluating their multiplicity and approximate energy range. The MS trigger chambers were designed to provide discrimination in well defined muon p_T thresholds, bunch crossing identification and complement the MDTs muon position measurements by recording the ϕ coordinate. The RPCs are located above and/or bellow the MDTs as shown in Figure 3.11. There are three RPC layers in the MS barrel. Assembled in the end-cap wheels, the TGC consist of multi-wire proportional chambers. Similarly to the RPCs, the TGCs complement the tracking chambers: in the innermost layer, covering the forward $1.92 <= |\eta| <= 2.4$ region, two TGC modules are installed in front of the CSC; and in the outer end-cap disk, covering $1.05 <= |\eta| <= 1.92$, one module in front of the MDTs and two modules behind them. The triggering of a signal is generated by a system of fast coincidences between the three outer layers along the path of the muon. In order to be used for trigger, a signal from a triggering-chamber must be compared with the two other chambers in the same sector and tower. From the sagitta of the curvature of the muon track, a coincidence pattern is defined as a criterion for the track to have passed a predefined momentum threshold. Coincidences between strips in the first two layers are used to create the low-p_{T} trigger, while a high-p_{T} trigger requires hits in all three layers.

The Muon Spectrometer provides good muon identification and momentum resolution over

a wide range of momenta. For a muon track with $p_{\mathrm{T}} = 1\,\mathrm{TeV}$, the MS momentum resolution is

$$\frac{\sigma_{p_{\mathrm{T}}}}{p_{\mathrm{T}}} = 10\%\,. \tag{3.10}$$

The trigger is very robust towards random hits due to neutron or photon backgrounds, originated from secondary interactions in the different material along the path of primary collision products.

3.3.5 Trigger and Data Acquisition System

In order to select potentially interesting events from the myriad of QCD background processes dominant at the LHC, the ATLAS physics program relies on the efficient performance of the trigger system. During the LHC runs of *pp* collisions, the increasing luminosity and pile-up conditions demanded the use of progressively higher energy thresholds and tighter selections to control the trigger rates.

The first run of *pp* data-taking at the LHC started by colliding bunches of protons at a center-of-mass energy of $\sqrt{s} = 7\,\mathrm{TeV}$, with 50 ns bunch spacing [188]. In the latter years, the center-of-mass energy of the collisions was increased to $\sqrt{s} = 8\,\mathrm{TeV}$ and the bunch spacing reduced to 25 ns, leading to approximately 23 interactions per beam crossing. This translates into a maximum input rate of events of 20 MHz, which is not only too large for permanent storage and analysis, but also much larger than the frequency of interesting hard-scattered interactions. The ATLAS trigger system [189] was designed to bring this rate to 100 Hz, in order to reliably select the events of interest and store their information permanently.

The trigger and data acquisition systems (DAQ) are based on signatures of particles with high transverse momentum and missing transverse energy, and are divided in three levels of event selection (c.f. Figure 3.12): Level 1 (LVL1), Level 2 (LVL2) and the Event Filter (EF). The most interesting events are identified in real time by the trigger system, each trigger level refining the decision made at the previous step and, if necessary, applying additional selection criteria.

The first level, LVL1, is a hardware-based trigger that selects large missing energy or high transverse momentum muons, electrons, photons, jets or τ-leptons decaying in hadrons. The decision is made under 2.5 μs, using coincidence or veto of certain defined thresholds for the different types of physics objects.

For the calorimeter triggers, the selection is made using reduced granularity signals covering $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ (trigger towers) from all ATLAS calorimeters to identify the

positions of Regions of Interest (RoI's) and compute the transverse energy E_{T} of electromagnetic clusters with a precision of 1 GeV. For each trigger tower, the cells of the electromagnetic or hadronic calorimeters are summed and a trigger is satisfied if the core-region, which is the central 2×2 trigger towers of a RoI, contains one pair of neighbouring towers with a combined energy that passes the threshold. In the case of the electron/photon and hadron/tau triggers, isolation can also be required.

For the muon identification, however, only the trigger chambers in the barrel and in the end-caps are used (c.f. Section 3.3.4). High transverse momentum muons are identified by the spatial and temporal coincidence of hits either in the RPCs or TGCs pointing to the beam interaction region. The degree of deviation from the hit pattern expected for a muon

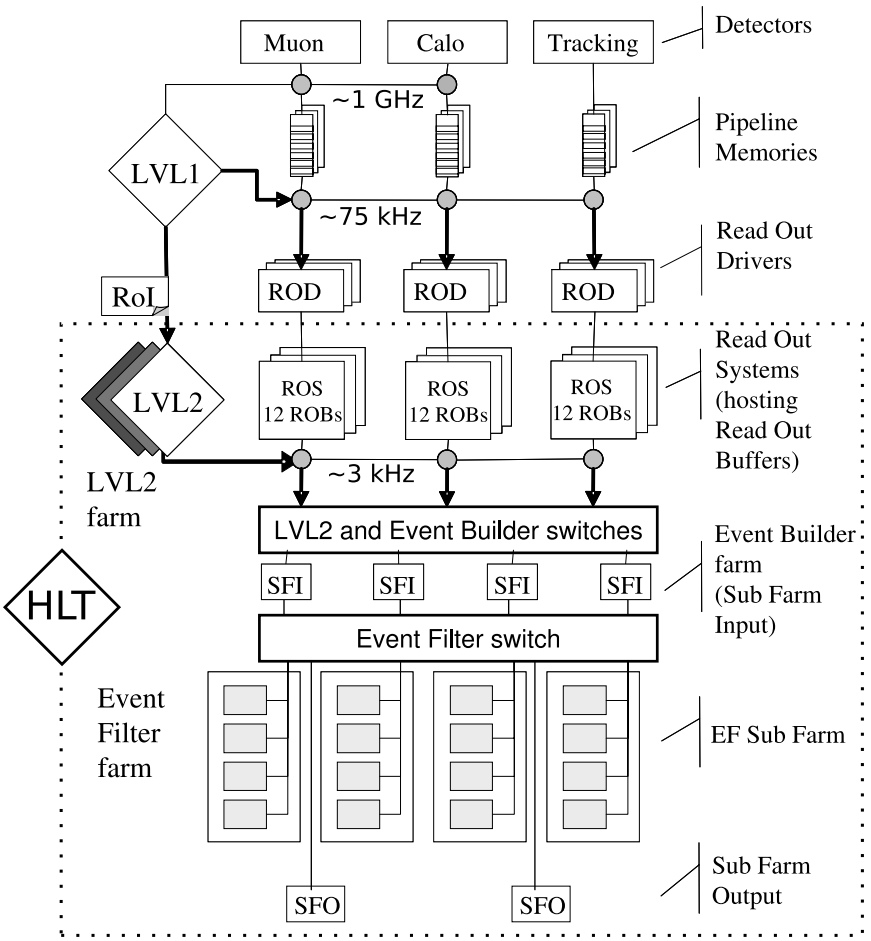


Figure 3.12: Block Diagram of the Trigger and DAQ system [187].

with infinite momentum is used to estimate the p_{T} of the muon, which must be above one of the possible thresholds, in order to be accepted.

If accepted by LVL1, the p_{T} thresholds and the corresponding detector RoIs are then sent to the LVL2 and EF for further consideration. During the information exchange period, *i.e.* the signal transmission between detector and trigger electronics, all of the data is kept in the pipeline memory. Data from events passing the LVL1 trigger selection are then transferred to the Readout Buffers (ROB's) via the Readout Drivers (ROD's). At this stage the event rate is about 75 kHz, which is still double the rate of the ATLAS physics programme.

Event data for the selected bunch crossing from all of the detectors are held in the ROBs either until the event is rejected by the LVL2 trigger (in which case the data are discarded) or, in case the event is accepted by LVL2, until the data have been successfully transferred by the DAQ system to storage associated with the EF. The process of moving data from the ROBs to the EF is called event building. This second and third levels of the trigger system are software based and constitute the High Level Trigger (HLT).

The LVL2 accesses nearly all the available sub-detector information in a RoI of specific $\eta \times \phi$ around the objects selected at LVL1. It will implement basic reconstruction algorithms in order to form a decision. The time taken to form and distribute the trigger decision at this level is 40 ms and it brings the rate of events down to approximately 1 kHz.

The data is then transferred to the EF, through the Sub-Farm Input (SFI). The EF uses refined reconstruction algorithms that are very close to those used offline. The available time for event processing is 4 s in average, and the rate of events is reduced to approximately to 100 Hz. Using the full granularity and precision of calorimeter and muon chamber data, as well as data from the inner detector, the HLT algorithms refine the trigger selections by having better information on energy deposition (which improves the threshold cuts) and by having inner track information (which enhances particle identification). The selected events will then go to the Sub-Farm Output (SFO) and will be stored to be analysed offline. The selected events are categorised into different datasets, or streams, according to specific interests of the analysis, e.g. electron streams, muon streams, jet streams, etc.

The trigger requirements are labeled with a code made out of a combination of letters and numbers. The letters indicate the type of object being triggered: whether it is an electromagnetic clusters (`EM`), electrons (`e`) or muons (`MU` or `mu`). Capital letters indicate that they are triggered in the LVL1 and the lower-case letters indicate the HLT.[1] The

[1]Further information on the trigger level can be specified, such as, EF for event filter triggers.

numbers right after these letters represent the transverse momentum of the object for which the trigger is 95% efficient. If the letter `i` or `I` is in the code, then an isolation criteria was applied. For example, `2e15i` means that there is a 95% efficiency for triggering an event with two isolated electrons with a p_{T} higher than 15 Gev. Further Information on the required trigger working point are expressed right after this sequence as `loose`, `medium` or `tight`.

During data-taking periods, the DAQ controls the movement of data down the trigger selection chain, as shown in Figure 3.12. However, it is also responsible for the configuration, control and monitoring of the ATLAS detector hardware and software by bi-directional communication with the Detector Control System (DCS), providing synchronisation of the detector with data-taking. The DCS sets the detector hardware operation conditions, coordinates all sub-detectors and LHC systems, continuously monitors and archives run-time parameters, and takes automatic corrective actions if necessary. The human interface for the full control of ATLAS and all sub-detectors is also guaranteed by the DCS.

3.3.6 ATLAS Recorded Luminosity and Pile-up

During normal LHC operations there is a non-zero probability of multiple proton-proton interactions per bunch crossing (pile-up). These simultaneous events, which originate from different pp collisions, are a side effect of the high frequency of collisions and the high density of bunches necessary for high luminosity. Pile-up events can be caused by additional interactions of protons in the same bunch (in-time pile-up) or by recording in the same event pp collisions from a different bunch crossing (out-of-time pile-up). During the $\sqrt{s} = 7\,\mathrm{TeV}$ run, the average number of interactions per crossing was approximately 9, increasing to 20 in the $\sqrt{s} = 8\,\mathrm{TeV}$ run.

Since protons are composite objects, the hard scatter interactions (where there is a large exchange of energy) can be described as an interaction between a pair of quarks or gluons. However, the underlying event can produce additional particles which will further disguise the signature of the main interaction. In order to detect and correct eventual bunch-by-bunch effects, the bunch-by-bunch luminosity measurement is required. The detectors, readout and acquisition systems must thus be capable of operating with a useful bandwidth of 40 MHz.

The average number of interactions per bunch crossing μ is a measure of the pile-up activity and from Equation 3.1 can be related to the luminosity as:

$$\mathcal{L} = \frac{\mu\, n_b f_r}{\sigma}\,. \tag{3.11}$$

In an experiment such as ATLAS, the observed or visible number of interactions is defined by $\mu_{\text{vis}} = \epsilon\mu$, where ϵ is the efficiency of a particular detector and algorithm used. Considering the detector (and associated algorithm), the visible inelastic cross-section σ_{vis} can be used for the luminosity measurement:

$$\mathcal{L} = \frac{\mu_{\text{vis}} n_b f_r}{\sigma_{\text{vis}}}\,, \tag{3.12}$$

where the value of σ_{vis} is initially unknown, but can be calibrated by making a measurement of the absolute luminosity at a certain point in time using Equation 3.2 and the beams transverse profiles from van der Meer scans. ATLAS monitors the delivered luminosity by measuring the observed interaction rate per crossing, μ_{vis} . In order to measure them, ATLAS uses several detectors, either bunch-sensitive or bunch-integrating.

Bunch-sensitive detectors will record the luminosity of each colliding bunch pair (BCID). These will primarily use counting algorithms. For example, a particular BCID will either pass or fail some trigger criteria (event counting algorithm) or instead, a number of detector channels with a signal above some value are counted per BCID (hit counting). For instance, the Beam Conditions Monitor (BCM) a station of four modules at either side of the ATLAS interaction point ($z = \pm 184\,\text{cm}$), designed to monitor beam particle loss [190], has fast response and readout time which enables bunch-by-bunch measurements. Another detector specifically designed for bunch-by-bunch measurements is LUCID. It measures the Cerenkov radiation produced by charged particles traveling above the local speed of light of the medium. It can perform more sophisticated measurements in addition to event counting, such as hit counting and particle counting. Both of these fast detectors are able to make statistically precise online luminosity measurements separately for each bunch crossing within the LHC active pattern without deadtime.

The ATLAS main data acquisition system can also contribute to luminosity measurements with bunch-by-bunch sensitivity. Such measurements rely on counting either the number of reconstructed tracks or vertices (interaction points). In the case of track counting, μ_{vis} is the average number of tracks per event. However, for a vertex-counting method, μ_{vis} is obtained by determining the average number of vertices per events satisfying some criteria and then applying corrections for known non-linear behaviour with increasing pile-up [191]. Whilst bunch-sensitive luminosity measurements can be made, these methods require extended time period to accumulate enough events, due to the limited read-out rate

of the triggers used. These algorithms are typically used as cross–checks of the primary online algorithms under special running conditions, where the trigger rates for these algorithms can be increased.

Bunch-integrating detectors, however, measure a time averaged quantity, such as a current in a calorimeter, and thus cannot distinguish between separate BCIDs. The ATLAS Tile Calorimeter system monitors the current drawn by each PMT, which is proportional to the total number of particles interacting in a given TileCal cell. This signal is proportional to the total luminosity summed over all the colliding bunches present at a given time. Similarly, the currents provided by the high voltage system of the Forward Calorimeter are directly proportional to the average flux of particles interacting in a given FCal sector. These are not exactly particle counting algorithms, since individual particles are not counted, but the measured currents should be proportional to the luminosity.

For most physics analysis, the relevant information is an integrated luminosity for a specific data sample. The Luminosity Block (lumiblock) is the basic time unit to store luminosity information for physics use. The boundaries of each luminosity block are defined by the ATLAS Central Trigger Processor (CTP), and can vary due to run conditions an other operational issues. Generally the duration of each lumiblock is one to two minutes of data taking. Each luminosity block contains data taken under uniform conditions (including luminosity), thus trigger configuration changes can only happen at luminosity block boundaries. The average luminosity for each detector and algorithm, along with a variety of general ATLAS data quality information, is stored for each lumiblock in a relational database. In order to define a physics data sample, lumiblocks are selected based on quality criteria. The integrated luminosity delivered in a selected lumiblock is obtained by multiplying the average luminosity in that lumiblock by the lumiblock duration. It should be noted that additional corrections can be made for trigger deadtime and trigger prescale factors, if needed.

The total integrated luminosity of an entire data sample can be obtained by integrating the instantaneous luminosity per lumiblock over the accelerator active time[1]:

$$N_{\text{tot}} = \sigma \int \mathcal{L} dt \,, \tag{3.13}$$

where N_{tot} is the total number of produced events. Therefore, the higher the integrated luminosity, the more data is available to analyse.

[1] Also refered to as a 'fill', period during which stable beams are kept colliding.

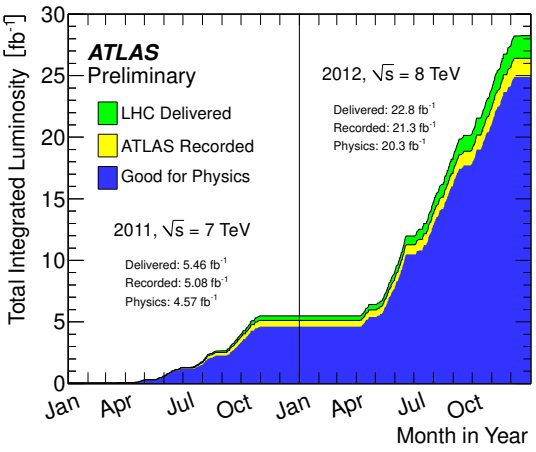


Figure 3.13: Cumulative luminosity versus time delivered to (green), recorded by ATLAS (yellow) and certified to be good quality data (blue) during stable beams for *pp* collisions are shown, at 7 and at 8 TeV center-of-mass energy for Run 1.

In Figure 3.13 the delivered luminosity by the LHC at the ATLAS experiment is shown. The differences between LHC delivered and ATLAS recorded luminosities are a result of inefficiencies in the data acquisition (DAQ). For recorded collision events to be used in data analysis, they must satisfy Data Quality (DQ) requirements assessed after reprocessing. Only data collected with the performance of all subdetectors meeting certain quality requirements are considered for the analysis. For each dataset, Good Runs Lists (GRL) are compiled, recording for each lumiblock which subdetectors satisfied the requirements. During Run 1, 95% of the ATLAS recorded data was considered as good, giving a total integrated luminosity of $20.3\,\text{fb}^{-1}$ of high quality data for physics analysis.

The distribution of the mean number of interactions per crossing is shown in Figure 3.14, weighted to the recorded luminosity and for both 7 and at 8 TeV center-of-mass energies. As can be seen, the mean number of interactions per BCID μ increases as the available center-of-mass energy of the collision increases. In order to obtain more data to analyse, higher energies and delivered luminosities are necessary, creating harder conditions in terms of pile-up. Here, μ corresponds the mean of the poisson distribution on the number of interactions per crossing calculated for each bunch. It was calculated from the instantaneous per bunch luminosity $\mathcal{L}$ in Equation 3.12, where σ_{vis} is the inelastic cross-section which was taken to be 71.5 mb for 7 TeV collisions and 73.0 mb for 8 TeV collisions [172].

For the $t\bar{t}H$ search performed in this thesis, the full ATLAS detector information was

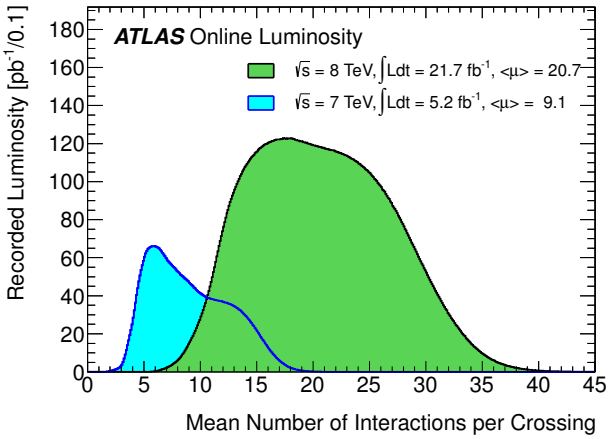


Figure 3.14: The luminosity-weighted distribution of the mean number of interactions per crossing for the 2011 and 2012 data.

used, requiring all sub-detector systems to meet the quality standards, resulting on a total integrated luminosity of $20.3\,\mathrm{fb}^{-1}$ of good data at $\sqrt{s} = 8\,\mathrm{TeV}$.

4

Description of Cell Noise in the Tile Calorimeter

For the reconstruction of the physics objects used in most data analysis, the ATLAS sub-detector systems must guarantee optimal performance under the LHC increasing energy and luminosity conditions. In this chapter, the monitoring of Tile hadronic calorimeter cells under different pile-up conditions is described. When multiple collisions are present, the energy deposits measured in each cell have overlapping contributions from the different interactions. A full survey of the energy response, associated noise and inter-correlations among cells is performed, contributing to a better knowledge of the TileCal response and performance in the increasing pile-up conditions of LHC's Run 1.

4.1 Tile Calorimeter Cells and Read Out System

Being the ATLAS barrel hadronic sampling calorimeter, the TileCal [182] provides accurate measurements of the energy deposition and direction of hadronic showers. Its structure and layout was previously introduced (section 3.3.3), however, additional details that are required for the cell response and noise evaluation are given here. Each longitudinal partition of the TileCal constitute an independent electronic readout block, with LBA and LBC covering the $|\eta| < 1.0$, while EBA and EBC cover $0.8 < |\eta| < 1.7$ region.[1] The light produced by incoming particles is read out from two sides by wavelength shifting (WLS) fibres, which are bundled together to form readout cells. As mentioned before (c.f. Figure 3.9), the cell structure in the calorimeter corresponds to pseudorapidity towers of 0.1, with different radial

[1] The partitions are named according to the pseudorapidity angle sign, and by convention, A is positive η and C negative η.

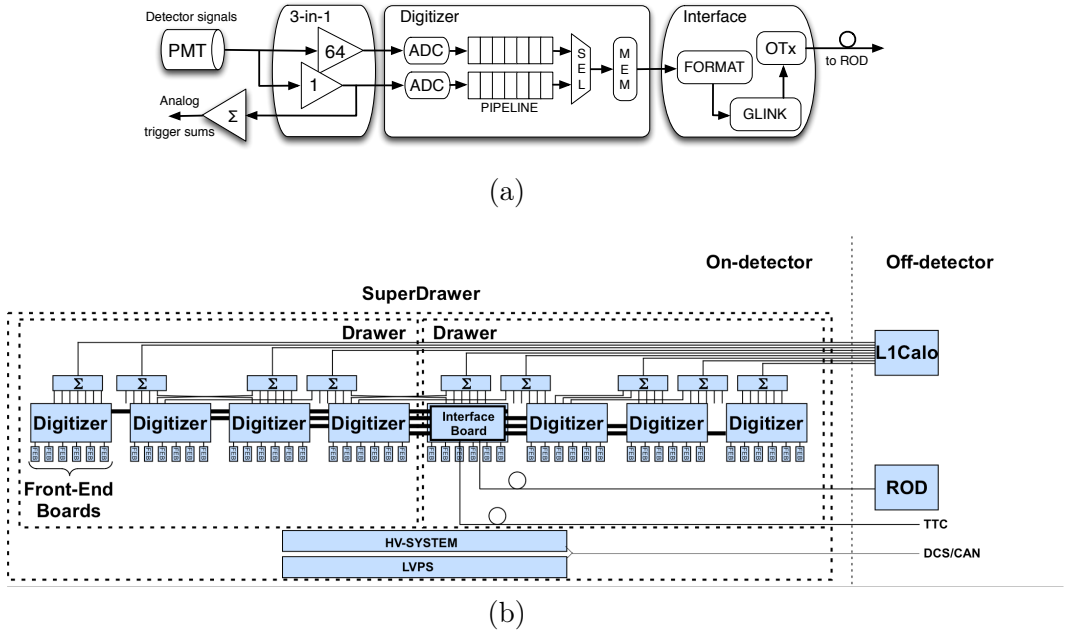


Figure 4.1: Diagrams of (a) a single readout channel [192] and (b) the front-end electronics [193] of TileCal, during run 1 of the LHC.

sampling depths named according to the relative position to the beam pipe: A cells for the innermost layer, BC and D cells to the middle and outer layers respectively. The bundles of fibres, in each side of the tiles, are guided into two separate channels and read out by photomultiplier tubes (PMT).

Stored in metallic cases (drawers) on the outermost part of each module, the front-end electronics of each LB (EB) has 45 (32) active channels. A channel consists of a magnetically shielded steel cylinder containing a light mixer, a photomultiplier tube, a voltage divider and a 3-in-1 card, as represented in Figure 4.1 (a). The light mixer is an optical plastic which ensures uniform illumination of the photo-cathode in the PMT by mixing the light from the readout fibres. With an amplitude proportional to the energy deposited in the cell by the incident particle, the signal from a PMT is then shaped to yield a fixed 50 ns pulsewidth, using a passive shaping circuit included in the 3-in-1 card. Providing all the analogue functions of the front-end electronics, the 3-in-1 board will also amplify the shaped signal in separate high (HG) and low (LG) gain branches (with a nominal gain ratio 64:1) [183].

The size of the TileCal signals can range from the typical muon energy deposition of a few hundreds of MeV to the highest-energy jet response, that can sometimes reach the TeV

level in a single cell. To cover this wide range, the digitiser system samples the incoming data from the 3-in-1 cards with a double readout, using an independent analog-to-digital converter (ADC) per gain branch. Housed in the TileCal data management unit chip (DMU), 10-bit ADCs sample the HG and LG signals with the LHC bunch-crossing frequency of 40 MHz (i.e. with timing steps of 25 ns). The digitisation timing of the ADCs is adjusted in multiples of ~ 0.1 ns guaranteeing that the full extension of the pulse is sampled and that the central sample is as close to the PMT pulse peak as possible. Seven consecutive analogue samples are kept: one close to the peak, four before the peak (with the two first samples providing a measurement of the signal pedestal) and two after the peak. To determine which of the gains is read out from the front-end electronics, a gain switch is used: the HG is always selected, unless any of the samples saturates the ADC; if the latter occurs then the LG ADC readout is chosen.

Each data management chip services six PMT channels at the same time, storing the digitised signals in fixed-length pipelines, while waiting for a LVL1 trigger decision. Figure 4.1 (b) shows the sketch of the front-end electronics on the calorimeter drawers, where the DMU are represented as digitiser boards. Organised as a chained system, the digitisers feed the data from the outermost readout channels to the Interface Board at the center of the superdrawer[1] via a low voltage differential signal. The Interface Board is responsible for the communication between the front-end and back-end electronics via optical fibres. The hardware-based trigger decision is made using the LG analogue signal from the digitisers, which is previously summed in adder boards within $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$, and then transmitted as trigger towers to the L1Calo trigger system, located outside the ATLAS detector. The front-end electronics is powered by a low-voltage power supply source (LVPS) per superdrawer. Each LVPS is located just outside of the superdrawer in a separate steel box. A system of four motherboards delivers the low-voltage power and control logic signals to the 3-in-1 cards, ADC integrators and adder boards. Each motherboard services up to twelve 3-in-1 cards and also contain specific circuitry for the TileCal calibration systems.

The back-end electronics is located approximately 100 m away from the ATLAS detector and contains three sub-systems: the LVL1 trigger receiver, the Trigger, Timing and Control system (TTC), and the readout Drivers (RODs). The TTC is responsible for the distribution of the the LHC clock, the trigger decisions and configuration commands to the Interface

[1] A superdrawer consists of two drawers: steel girders at the outer radius of each module that can be fully extracted while leaving the remaining module in place.

Boards, which then broadcasts them to the chain of digitiser boards. When an event is accepted by the LVL1 trigger, the TTC sends the trigger acceptance command to the front-end electronics. The corresponding digital signals from all the calorimeter channels, which were temporarily waiting in a pipeline memory, are transmitted by the TTC via optical fibers to the RODs. The TTC will also distribute the timing clock and trigger information to the RODs, which will process the data from the front-end electronics. At the RODs, the time and energy of the signals sampled in each channel are processed and reconstructed in order to provide the deposited energy and the signal timing in appropriate units. After being digitally processed by the back-end system, data collected by the front-end electronics is sent to the data acquisition system (DAQ) at a LVL1 trigger rate.

4.2 Signal Reconstruction and Energy Measurement

The electrical signal for each TileCal channel is reconstructed from the seven ADC samples sent to the RODs. The signal reconstruction is obtained with the Optimal Filtering method (OF) [183, 194], which calculates the signal properties, such as pulse amplitude, pedestal and time phase. This method compares the digitised samples with a reference pulse shape used for all channels[1], making use of weighted linear combinations. The OF algorithm computes the amplitude (A), time phase (t) and the pedestal (p) for a channel as:

$$A = \sum_{i=1}^{n=7} w_i S_i \qquad t = \frac{1}{A} \sum_{i=1}^{n=7} b_i S_i, \qquad p = \sum_{i=1}^{n=7} c_i S_i \tag{4.1}$$

where S_i is the ADC sample taken at time $t_i\,(i = 1, ..., n)$. The coefficients w_i, b_i and c_i are the OF weights derived for both high and low gain channels. The OF weights are evaluated while minimising the dispersion in both amplitude and time arising from electronics and pile-up noise, taking into account the pulse shape at the ADC input and noise autocorrelation matrix. During Run 1, the OF minimisation procedure only considered electronic noise for the weight computation.

Figure 4.2 shows the time shape of the signal, with the seven ADC samples in a 150 ns readout window, and whose magnitude was reconstructed using the Optimal Filtering weights. As can be seen, A is the distance between the peak and the baseline of the signal (pedestal). The phase is defined as the difference between the signal sampling clock and the maximum of the incoming pulses (peak amplitude of the signal).

[1]The reference pulse shape was taken as the average pulse shape from test beam data, with reference pulses for both high and low gain modes.

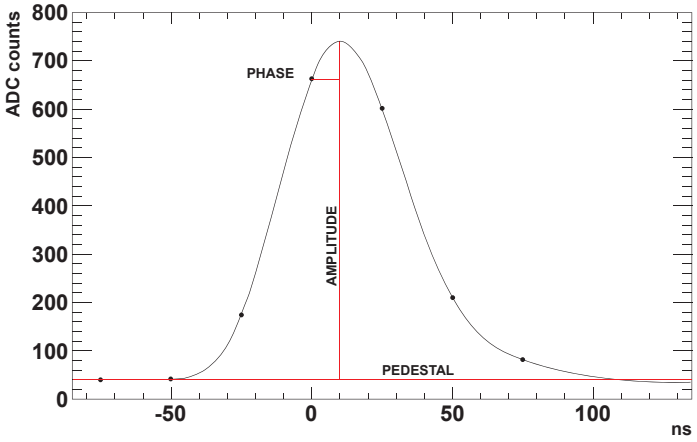


Figure 4.2: Reconstructed analog signal using the Optimal Filtering magnitudes for weights [195].

During the LHC operation, the ADC digitisation should be synchronous with the TTC clock, allowing the phase of the pulses to be constant within very small fluctuations and predictable using the TileCal calibration system. For a precise signal reconstruction of the energy deposited in the TileCal, the phase is evaluated and minimised, so that all the channels are expected to have their mean time $< t_{\text{channel}} >= 0\,\text{ns}$. In Run 1 of the LHC operations, two modes of OF reconstruction were applied in the RODs: an iterative and a non-iterative implementation. The iterative OF method was used to reconstruct data from the 2010 proton collisions (without out-of-time pile-up). In this iterative method, when the difference between the maximum and minimum sample is above a certain noise threshold, the pulse shape is recursively fit: the initial time phase is taken as the time of the maximum sample; and in the following steps the previous time phase is used as the initial input for the fit. For signals below the threshold, only a single iteration is performed, assuming a pulse with the peak in the central maximum sample. The non-iterative OF method was applied in 2011-2012 proton collision data, since it is more robust against pile-up. Here the time phase is fixed for each individual channel and only a single fit to the samples is performed.

The digital signal processor in the ROD performs the online signal reconstruction using the OF weights, as well as the conversion from the signal amplitude A in ADC counts to energy units of GeV by applying channel-dependent calibration constants:

$$E_{\text{channel}} = A \cdot C_{\text{ADC}\to\text{pC}} \cdot C_{\text{pC}\to\text{GeV}} \cdot C_{\text{Cs}} \cdot C_{\text{Laser}} \ . \tag{4.2}$$

The calibration constant $C_{\mathrm{ADC}\to\mathrm{pC}}$ is the conversion factor of ADC counts to charge in pC. It is determined for each channel using a well defined injected charge with the CIS (Charge Injection System) calibration System. The calibration constant $C_{\mathrm{pC}\to\mathrm{GeV}}$ is the conversion factor of charge to energy in GeV and is obtained from the cell response to an electron test-beam of known momentum. Additional corrections are applied via the calibration constant C_{Cs}, which accounts for residual non-uniformities after the gain equalisation of all channels. It is obtained by circulating a Cs radioactive source in a specific calibration system of the TileCal. Finally, the calibration constant C_{Laser} corrects non-linearities of the PMT response and is measured by the Laser calibration system.

Both the channel energy (in GeV) and the phase time (in ns) are then transmitted to the HLT. If selected by the HLT, the raw signal is reconstructed offline by using the same iterative or non-iterative OF method as online. The 2010 HLT-selected data was fully reconstructed offline, only using the RODs amplitude and time estimates for the HLT decision. However, with the increasing luminosity of later data taking periods, the ROD output become saturated. In order to avoid this saturation, from 2011 onwards, a signal amplitude threshold of 5 ADC counts ($\approx 60\,\mathrm{MeV}$) was required for offline reconstruction. If the difference between maximum and minimum S_i samples is below that threshold, the ROD signal reconstruction results are used in the offline data processing.

The energy deposited in a single cell is obtained by summing the energy measurements of the two readout channels:

$$E_{\mathrm{cell}} = E_{\mathrm{channel\,1}} + E_{\mathrm{channel\,2}}\,. \tag{4.3}$$

The cell time, however, should be the same in both channels. By averaging the time measurement of the corresponding readout channels of each cell a more precise cell time measurement is obtained:

$$t_{\mathrm{cell}} = \frac{1}{2}(t_{\mathrm{channel\,1}} + t_{\mathrm{channel\,2}})\,. \tag{4.4}$$

For single readout cells, the cell energy measurement is twice the available channel energy. If one of the readout channels is masked out, the same algorithm is applied. This ensures the robustness of the energy measurement of the TileCal cells to failures on one of the readout channels.

For reconstruction of physics objects, such as jets, the cell energy is the fundamental building block. The energy deposited in the TileCal cells is combined into clusters with the topological clustering algorithm [196]. The significance of the absolute reconstructed cell

energy value relative to the respective noise ($S = |E|/\sigma$) is used to identify meaningful energy depositions from noise fluctuations in the cells. Using an iterative process, the clustering algorithm searches for cells with energy deposits above 4σ of the energy distribution of randomly triggered events. It then surveys neighbouring cells for energy deposits above a 2 σ threshold. The energy cluster is formed by grouping together the energy deposits of the cells verifying the 4 σ and 2 σ thresholds, as well as including their immediate adjacent cells. The topological clusters are then used as inputs to jet reconstruction algorithms and transverse energy algorithms. A good description of both the electronic noise and the pile-up noise contribution is thus crucial for the creation of topological clusters of calorimeter cells.

4.3 Tile Calorimeter Cell Noise

From the PMT analog signal extraction and processing by fast electronics on the detector, to the transmission via optical fibers to back-end electronics for final signal processing and storage, the data signal will have been through a complex chain of fast electronic circuits and computer farms. During this process undesirable effects may happen perturbing the transition or recording of the input signal. Current or voltage fluctuations due to thermal motion or charge granularity may occur, as a result of the readout hardware architecture. Such unwanted fluctuations are usually defined as noise and represent the uncertainty to which the signal can be measured. The characteristics of the complex readout chain will therefore determine the precision limits of the measurement, as will be discussed in the following sections.

Designed for fast signal processing, the TileCal front-end electronics of each readout channel was chosen as to minimise the inherent electronic noise associated with the hardware. In fact, the output of each channel is not simply signal from the energy deposition, it also includes pedestal and a random noise component. The total electronic noise of a single cell comprises contributions from the intrinsic electronic noise of each single photomultiplier, the digitisation (digital noise), and the cell architecture in the calorimeter (where a single PMT is shared by a group of tiles). In the front-end electronics common signals, such as power, timing, trigger and control, as well as electromagnetic field emissions, can also produce coherence effects like cross-talk between neighbouring channels [197]. Since these perturbations are mainly hardware related, and as so, largely independent of external LHC beam conditions, their immediate effect can be assessed by the variation of the baseline signal

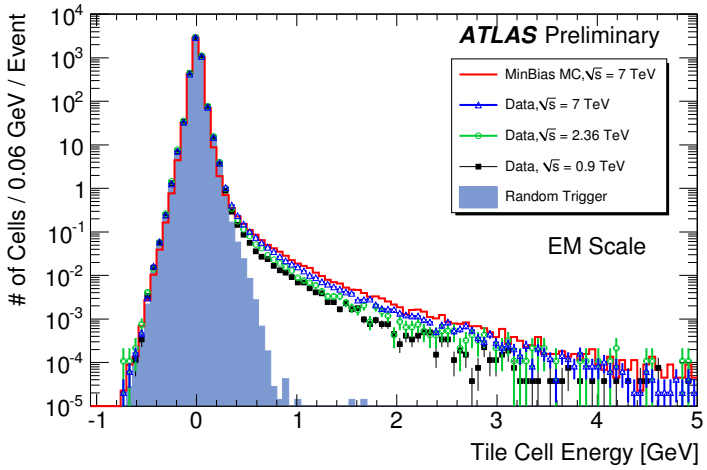


Figure 4.3: Energy distribution of the TileCal cells for collision data at center-of-mass energies of 0.9, 2.36 and 7 TeV. Each distribution is normalised to the respective number of events and compared with Pythia minimum bias Monte Carlo simulation (in red) and random triggered events (in the blue shaded histogram).

(pedestal) of individual PMT channels, as well as their sum. In the TileCal, the electronic only pedestal signal can be measured in empty event runs, where the detector is active, but there are no beams colliding. In these conditions, the fluctuations of the pedestal represent the level of electronic noise in each channel and are typically calculated as the standard deviation or root mean square (RMS) of the energy distribution.

During normal collision LHC runs, the physics signals arising from shower fluctuations create a wider dispersion of the energy measurements, much larger than the noise from any source. Figure 4.3 shows the energy distributions of TileCal cells for collision events of different center-of-mass energies, normalised to the total number of events per run. The measurement is performed using 2010 data where the pile-up contribution is negligible. Here, the events were triggered using the Minimum Bias Trigger Scintillators (MBTS) [182], which select data with signal amplitudes just above the electronic pedestal noise. Data selected with a minimum bias trigger is usually used to characterise the properties of inelastic hadron collisions with the smallest bias possible. The minimum bias trigger criteria avoids dependence on complex object reconstruction, as opposed to the nominal ATLAS triggers. At ATLAS, events collected in minimum bias dedicated runs account for the majority of interactions resulting from beam collisions, without significant hard-scattering collisions occurring in each event. In Figure 4.3, the cell response spectrum to minimum bias triggered

events can be compared with the distribution resulting from random triggered events, which is represented by the shaded histogram. As can be seen, the energy distribution of events passing a random trigger has a much narrower standard deviation than those passing a minimum bias trigger, suggesting that there is an even rarer presence of hard-scattered signal events on the randomly triggered data than on the minimum bias triggered. Particles from hard-scattered collisions produce quite large energy depositions. Ordinarily, the calorimeter measurement of interest is the energy deposition of these particles. However, when trying to assess the noise associated with the measurement, it is important to avoid the bias that the actual physics signal entails. For cell noise estimation purposes, the calorimeter response to random triggered collision events is more suitable, since these are rich in low energy particles produced in soft collisions. The standard deviation of the energy distribution obtained with randomly selected events during collision runs provides a good estimate of the cell noise in data.

Proton-proton collisions at the LHC, in addition to the collision of interest, also yield several simultaneous soft-scattered interactions. The mean number of inelastic pp interactions per bunch crossing, $< \mu >$, is proportional to the instantaneous luminosity, $\mathcal{L}$:

$$< \mu >= \frac{\mathcal{L}\, \sigma_{inel}}{N_{bunch}\, f_{LHC}} \tag{4.5}$$

where σ_{inel} is the total inelastic pp cross-section[1] and $N_{bunch}\, f_{LHC}$ is the average frequency of bunch crossings in the LHC [172]. Since the lumiblock is the basic time unit to store luminosity information,[2] all possible measures of the number of collisions will also be an average over a given lumiblock.

The simultaneous pile-up interactions are a consequence of the beam characteristics, particularly the high number of protons per bunch and the short time interval between bunch crossings. The increasing luminosity per bunch crossing, and hence increasing $< \mu >$, leads to higher probability of multiple soft proton-proton interactions, also called pile-up events. Consequently, in-time energy deposits originated from multiple collisions of the same bunch crossing (but different pp pair) can occur in the same TileCal cell. This in-time pile-up can be studied by determining its average effect on the measured calorimeter energies. Pile-up also arises when the time interval between consecutive BCIDs is lower

[1] For $\sqrt{s} = 7\,\text{TeV}$, ATLAS uses the PYTHIA value $\sigma_{inel} = 71.5\,\text{mb}$

[2] Recall that one lumiblock contains roughly 2 minutes of data taking, but its duration can vary due to run conditions.

than the detector integration time and so *pp* collisions from different bunch crossings may be integrated in the same event. For example, during Run 1, as the luminosity and pile-up conditions change towards higher values, with consecutive proton bunches crossing every 50 ns (25 ns by the end of the run), the ADCs large readout window of 150 ns (c.f. Figure4.2) lead to a significant fraction of the TileCal cells receiving energy from more than one bunch crossing within the same readout window [188]. This is called the out-of-time pile-up and can considerably degrade the measurement of the energy deposited in a cell. Both in-time and out-of-time pile-up depend on the luminosity, and can be inclusively estimated by the μ value of collision data, as previously mentioned in Chapter3.3.6. Throughout Run 1, for every bunch crossing, the number of *pp* collisions was progressively increased from 5 up to 35 simultaneous events.

In the ATLAS detector, every physics object is affected by pile-up to some extent. It is essential to carefully assess the detector response to pile-up events, by studying its characteristics in data and model Monte Carlo simulations. Particularly in the Tile Calorimeter, the effect of the pile-up on the cell output is evident by changes in the associated noise value. Since the topological clustering algorithm, employed to reconstruct the energy deposits, uses the cell noise to estimate the compatibility of the energy measurement with a noise fluctuation, it is very important to, not only minimise, but also precisely determine the noise level per cell, as to reduce the impact of noise on the jet energy measurement.

During collision runs, the total noise per cell is a result of electronic noise in the readout chain and also of contributions from pile-up interactions (pile-up noise). Taking into account these two components, the total noise is usually estimated by adding in quadrature the two contributions:

$$\sigma_{el.noise+pile-up} = \sqrt{\sigma^2_{el.noise} + \sigma^2_{pile-up}} \tag{4.6}$$

where pile-up and electronic noise are assumed to be uncorrelated.

Electronic Noise and Coherence Effects

In the TileCal, the electronic noise associated with the energy depositions is measured by fluctuations of the pedestal, without the presence of any high energy particles originated on hard-scattered interactions. Special standalone bi-gain runs with no physics signal are used to study the electronic noise of the readout chain. These calibration runs taken without colliding beams (pedestal runs) have two different electronic configurations - the low gain and the high gain. Since this electronic component of noise is independent of the LHC beam

conditions, noise estimated from pedestal data mirrors that of the data-taking period, using the OF technique.

For correct estimation of the significance of the energy deposits, during run 1, the template of the cell energy shape was described to a good approximation by a double Gaussian distribution [183]. A normalised function is used, with a narrow single Gaussian core and a second wider Gaussian function to describe the tails:

$$f_{\text{p.d.f.}} = \frac{1}{1+R}\left(\frac{1}{\sqrt{2\pi}\sigma_1}e^{\frac{x^2}{2\sigma_1^2}} + \frac{R}{\sqrt{2\pi}\sigma_2}e^{\frac{x^2}{2\sigma_2^2}}\right) \quad . \tag{4.7}$$

Here the relative normalisation of the two Gaussians, R, and the standard deviations of the first and second Gaussians, σ_1 and σ_2, are the three independent parameters used to fit the function. In pedestal data, the means of the two Gaussian functions is zero, which is a good approximation for the cell noise. For the topological clustering algorithm, an equivalent noise estimate, $\sigma_{\text{eq}}(E)$, is used to measure the significance of the signal, where the significance for the double Gaussian is the same as in the 1σ region of a standard single Gaussian function. During Run 1, the standard deviation of the energy distributions in pedestal events fluctuated by an average of 1.2% for high gain and 1.8% for low gain across all channels, indicating stable electronic noise constants [198].

In Monte Carlo simulations, the electronic noise is implemented in the digitisation of the signals. From the cell noise measured in data, the ADC noise can be extrapolated and added to the individual MC digitised samples. The signal shape at the sample level is modelled to follow the double Gaussian distribution, as seen in data. The parameters used in the simulations are based on the cell noise coefficients measured in data. More details about the electronic noise description in both data and MC simulations in TileCal can be found in [199, 200].

When using the sum of many signals in one measurement, such as in the topological cluster algorithms, the understanding of the correlations between signals from different cells is of major importance. The readout chain encompasses several groupings of electric signals: the six PMT channels read out by each DMU; the twelve channels fed by each motherboard; the single interface board per superdrawer; and finally the LVPS powering the motherboards. Since each signal grouping may lead to coherence effects, it is important to assess its impact not only on individual channels but also on their sum. The coherent effects due to the architecture of the electronic readout chain can be studied by looking at the correlations between the energy distributions of neighbouring cells [197]. During the initial TileCal test

phase, a considerable level of correlation was found among channels belonging to the same motherboard. However, the gaussian model of the energy shape does not take into account these correlation effects.

Pile-up Noise

During nominal conditions, the major source of noise in the TileCal should be from pile-up events. Under high number of simultaneous collisions, the electronic noise does not play a significant role in the performance of the calorimeter. The presence of multiple collisions degrades the detector response by possibly integrating energies from more than one collision in a single measurement. Calorimeter cells can receive energies either from additional in-time collisions within the same bunch crossing or from collisions in bunches crossing just before and after the collision of interest (out-of-time pile-up). Although out-of-time collisions are more susceptible to affect the measurement of the interesting signal depositions, both pile-up sources may lead to overlapping energy deposits. At ATLAS, the pile-up effects are reduced by using vertex information and individually reconstructing each collision. However, distinguishing energy depositions in the same readout window is almost impossible. Due to the harsher operation conditions, pile-up is a serious challenge to the hadronic calorimeter. The impact of increasing pile-up conditions on the TileCal response can then be directly assessed by monitoring the noise changes in pile-up data.

For collision data, just as for pedestal data, the signals sampled in each TileCal channel have a Gaussian shape with an amplitude proportional to the energy deposition. The signal width is, however, larger. In the presence of hard-scattered interactions, the high energy influence on the signal width maybe too big to retrieve unbiased information on the noise of each cell. Thus, for pile-up noise estimation unbiased data is required. As already mentioned, random or zero bias triggered events are the best sources of simultaneous soft-scattered collisions, without compromising the integrity of the measurement. Recording events in coincidence with the *pp* collisions, the zero bias trigger is a randomly seeded trigger with a rate proportional to the instantaneous bunch luminosity. This triggering provides a data sample, which is unbiased by a significant energy deposition in calorimeter cells.

Even though the zero-bias sample will contain a small fraction of the hard scattering events with quite large energy deposits, the contribution of these events is small on average, such that cells will be mostly exposed to low energy particles produced from soft pile-up collisions. Estimating the noise of each TileCal channel as the significance of the signal

deposition is still a valid assumption. Here, particles from pile-up events will produce additional energy depositions in the tile cells widening the energy distribution and, possibly, pulling the mean energy from zero. The width of the signal measurement of each cell can be determined as the RMS of the energy distribution:

$$RMS = \sqrt{< E^2 > - < E >^2}, \tag{4.8}$$

where $< E^2 >$ is the mean value of the square of the energy and $< E >^2$ is the square of the mean energy deposit. The RMS describes the total noise per cell, including both pile-up and electronic noise. In zero bias data, the estimated noise per channel is a combination of electronic noise and pile-up effects, mirroring that of normal data-taking periods.

The simulation of pile-up events in Monte Carlo is done by mixing the hard scattered events with an appropriate number of simulated minimum bias events. The pile-up conditions in the simulations are defined by the beam energy, bunch spacing and the expected average number of the pile-up collisions per bunch crossing, just as in data. The number of pile-up collisions per bunch crossing follows the Poisson distribution with mean value $< \mu >$. The minimum bias interactions are produced with the Pythia generator [201] and are merged with the simulated high energy signal events during the seven sample digitisation process, matching the ADC's sampling every 25 ns of signal in collision data. The pedestal and electronic noise are implemented in the merged samples, which will then enter the signal reconstruction in the OF method. In the MC simulations the pile-up noise is estimated as the RMS of the cell energy measurement. For different pile-up conditions, an approximation assuming that the pile-up term increases with $\sqrt{< \mu >}$ for fixed bunch spacing was used [202].

In spite of ATLAS having several methods to deal with the presence of pile-up events, the TileCal's cell noise associated with the energy measurement needs to be addressed. The noise dependency with beam energy, luminosity and pile-up is investigated in Section 4.5, contributing to a better characterisation of the TileCal response and performance under increasing pile-up conditions.

4.4 Analysis of the TileCal Correlated Noise

The coherent component of the electronic noise is an inherent consequence of the architecture of the TileCal front-end readout electronics and is responsible for significant pedestal

degradation [197]. Since the possible sources of correlation in the readout chain are diverse, the description of the coherent effects starts by considering the correlations between signals at the digitisation level. The correlations among digitised signals can be described according to a data driven covariance matrix and its effect propagated to all subsequent steps of the energy reconstruction, thus avoiding the direct simulation of possible coherence causes.

Considering the electronic structure of a single TileCal readout channel (c.f. Section 4.1), the two different gain signals should not be correlated. Furthermore, the ADC digitisation will not contribute with sample-to-sample correlations within the same channel. Different channel samples with the same trigger time can, however, be correlated. Since the readout electronics is contained inside individual metallic drawers, correlations between channels from different partitions or modules are not considered. The correlation between channels i and j can be evaluated with a covariance matrix $\mathbf{cov}(x_i, x_j)$, where x describes the noise contribution of each channel involved:

$$\mathbf{cov}(x_i, x_j) = E[(x_i - \bar{x}_i)(x_j - \bar{x}_j)] = E[x_i x_j] - \bar{x}_i \bar{x}_j \ . \tag{4.9}$$

The channel signals, x_i and x_j, are the seven sampled pulses reconstructed with the OF method, $\bar{x}_i$ and $\bar{x}_j$ are the mean values of each channel, and the operator E is the expectation value.

The correlation matrix, $\rho(x_i, x_j)$, can be defined according to:

$$\rho(x_i, x_j) = \frac{\mathbf{cov}(x_i, x_j)}{\sqrt{E[(x_i - \bar{x}_i)^2]}\sqrt{E[(x_j - \bar{x}_j)^2]}} = \frac{\mathbf{cov}(x_i, x_j)}{\sigma_i . \sigma_j}, \tag{4.10}$$

where σ_i and σ_j are the standard deviations of channel i and j respectively.

In order to assess the general behaviour of the coherent noise component in the TileCal, correlation matrixes for the full set of modules were used as a diagnosis tool. Initial testing using pedestal only data already yielded significant correlations, particularly between channels sharing the same motherboard or DMU.

In Figure 4.4, pedestal only correlation matrices calculated according to Equation 4.10 are shown, for the four TileCal partitions. The same azimutal module is depicted for comparison purposes, in this case module 32. Since the correlations were only visible between high gain channels, only these are considered here. For each partition, all 48 channels of module 32 were mapped and the covariance and correlation between channels was calculated. The observed pattern shows regions of high and low correlation in all partitions,

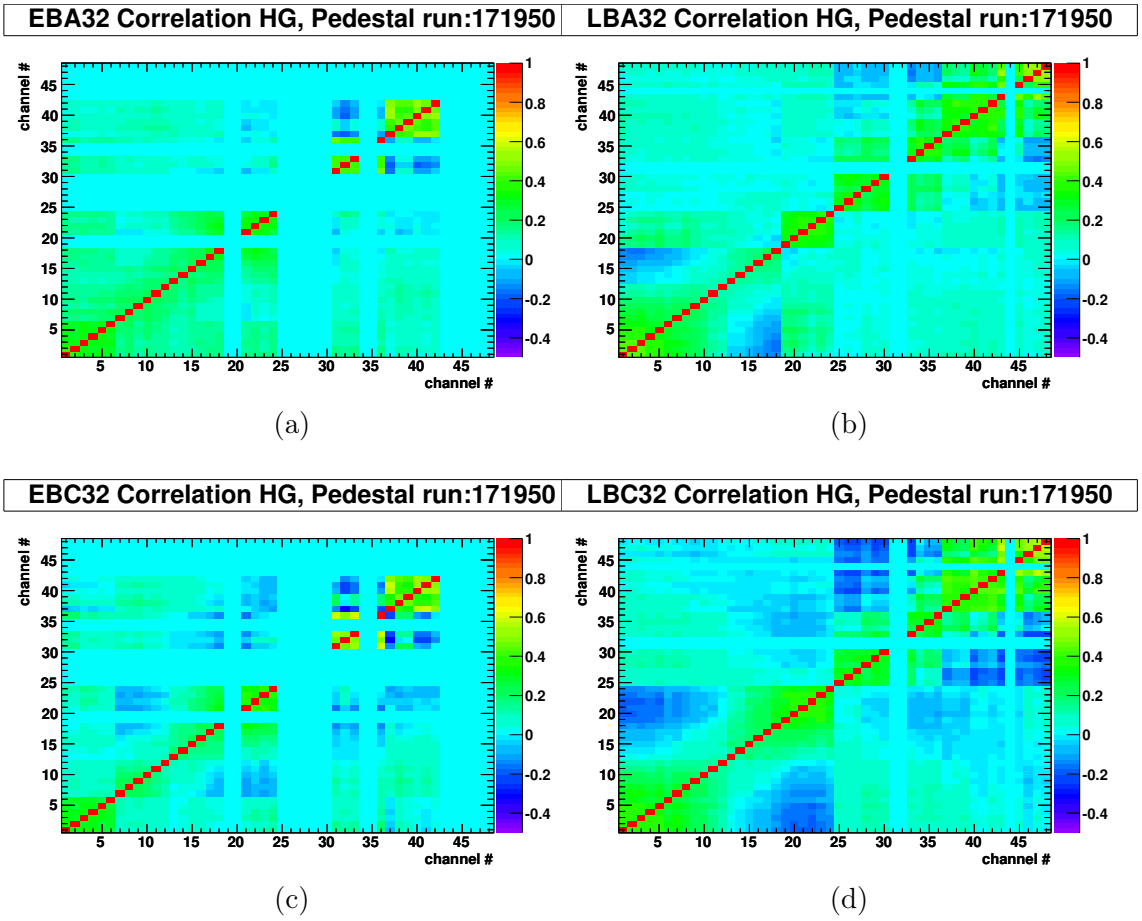


Figure 4.4: Maps of the correlation between the pedestal noise signals of high-gain channels, for all four TileCal partitions: (a) EBA, (b) LBA, (c) EBC and (d) LBC. The same arbitrary azimuthal module is shown.

consistent with the TileCal front-end architecture. The presence of *open channels* in the TileCal modules is also visible.[1] The correlation is stronger between groups of 6 or 12 neighbour channels, where possible interference occurs between channels read by the same digitiser card or motherboard, respectively. Channels closer to the LVPS (channels > 40) show higher correlations, as do channels near the sub-assemblies (channels 20 to 24). These problematic regions of the front-end super-drawers are particularly sensitive to interference between electromagnetic signals of the LVPS and of the readout chain.

To understand the presence of these correlations, and their effect on the signal dis-

[1]Channels 31, 32 and 44 of long barrel modules are *open channels*: no tile cell exists at those locations due to the passage of cables for the ID, and as so, no PMTs are installed. For extended barrel modules, the open channels are 19, 20, 25, 26, 27, 28, 29, 30, 34, 35, 42, 43, 44, 45, 46, 47 and 48.

turbance of individual cells, software studies were developed side-by-side with hardware improvements. Two methods were developed to mitigate the coherent noise component, based on different algorithms: the χ^2 method [203] and the Tile Noise Filter (TNF)[204]. Both approaches effectively remove the observed coherent noise component in pedestal runs, without modelling the specific source of the problem.

χ^2 method

The χ^2 method assumes that the noise of a particular TileCal channel can be described as a linear combination of an intrinsic noise component (β_i) and a contribution that depends on the response of all readout channels in the same module:

$$x_i \quad = \quad \beta_i + \sum_{j \neq i}^{N} \alpha_{i,j} x_j, \tag{4.11}$$

where x is the OF reconstructed signal for channels i and j of the same module. The unknown $\alpha_{i,j}$ ensure the weighted sum of the signals of all the other N readout channels in the module. In pedestal data, the signal distribution of a single channel is expected to be consistent with zero energy deposited. Given that β_i is the pedestal subtracted signal, its value is also expected to be zero. For each channel, a χ^2 function, fitting the model with the measured noise values, can be defined as:

$$\chi^2_i = \sum_{\text{Events}} \frac{\left[x_i - (\beta_i + \sum_{k \neq i}^{N} \alpha_{i,k} x_k)\right]^2}{\sigma_i^2}. \tag{4.12}$$

Using the χ^2 minimisation method, with respect to each model parameter ($\alpha_{i,j}$ and β_i) for all individual channels, the α matrix of a specific module can be extracted. This algorithm assumes that the noise correlations are independent of the signal amplitude, thus only the first ADC sample of each signal deposition is used. The signal can then be re-evaluated by removing the offset β_i and applying the α matrix to the measured signals of all the other channels:

$$s_i^{rec} \quad = \quad s_i - (\alpha_{i,1} x_1 + \alpha_{i,2} x_2 + ... + \beta_i + ... + \alpha_{i,N} x_N) \tag{4.13}$$

Tested through a systematic survey of the TileCal modules in pedestal runs, this method removes the coherent noise component from the cell energy depositions, leading to narrower signal distributions, with RMS values closer to the intrinsic noise ones.

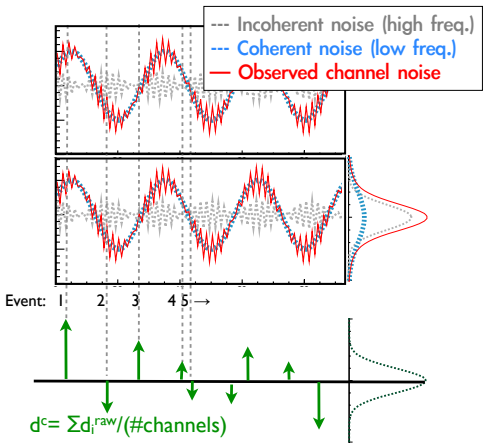


Figure 4.5: Schematic interpretation of the coherence effect on the observed noise of a TileCal channel. The TNF common mode determination is also explained [204].

Tile Noise Filter

The TNF approach is based on the strong correlation observed among signals from channels connected to the same motherboard (12 channels each). The noise characteristics observed, in a single channel, indicate a possible grounding problem: besides the intrinsic channel noise associated with an event (be it an actual physics object or a random triggered event), a lower frequency oscillation is present, and it is synchronized with the motherboard power supply. Figure 4.5 is a diagram representing the oscillations in the signal of two neighbouring channels. Typically, when an event activates a specific channel, the noise associated with the energy deposition shows an incoherent behaviour, which is seen by the oscilations in the channel signal (dashed grey line). However, the observed channel noise (represented by the solid red line) seems to also have a coherent behaviour, consistent with the LVPS lower frequency (represented by the bold dashed blue line). The cell signal for the sum of all events will reflect this combination of incoherent and coherent noise sources: the signal distribution encompasses a narrower Gaussian due to the inherent channel noise and a widder Gaussian due to the coherence effects. For each motherboard, the common mode, d_c, can be defined by estimating the average pedestal, on a event by event basis, as:

$$d_c = \frac{\sum_i^{N_{ch}} d_i}{N_{ch}} \tag{4.14}$$

where d_i is the raw pedestal of the i-th channel (in ADC counts) and $N_{ch} = 12$ is the number of channels per motherboard. Using dedicated runs, the common mode distribution can be

measured and a significant mean value, $< d_c >$, can be estimated. For each event, the common-mode fluctuations constitute a good estimate of the coherent noise component:

$$d_{coherent} = d_c - < d_c >= d_c - \left(\frac{\sum^{N_{ev}} d_c}{N_{ev}} \right) \tag{4.15}$$

with N_{ev} the number of events. The common mode shift can then be subtracted from the raw pedestal value per channel,

$$s_{rec} = d_i - d_{coherent}\,, \tag{4.16}$$

which gives the corrected channel signal, s_{rec}. The signal correction is applied on the digitised signal samples, in ADC counts, and then propagated through the normal energy reconstruction algorithms. An energy threshold is used in order to avoid biases due to problematic cells or high energy signals from hard-scattered events. The TNF method does not affect the TileCal calibration constants, nor the cell timing. It is a powerful tool that effectively removes the correlations between the calorimeter cells, resulting in the suppression of the second gaussian component of the digitised sample, up to $\sim 5\%$ on the barrel cells.

Both unfolding methods were validated using pedestal data, as well as minimum bias triggered data collected in the beginning of Run 1, *i.e.* with no significant pile-up presence. The TileCal output reconstruction software was optimised to remove the correlation noise using the TNF algorithm. Despite being mostly a result of the hardware features, the coherence pattern can be influenced by the presence of multiple energetic signals in the calorimeter cells, which can impact the performance of the unfolding method used to remove the correlations. The study of the coherent noise component in the presence of simultaneous collisions was investigated and the impact of the correlation removal method was tested, resulting in several internal presentations to ATLAS community, as well as a poster summary at the European School of High-Energy Physics [205].

4.4.1 Pile-up effect on the TileCal Correlated Noise

As already mentioned, in order to achieve the luminosity necessary for the physics analysis, the ATLAS detector is subjected to several simultaneous collisions, more so the higher the luminosity. In the following analysis, the TileCal coherence noise characteristics is investigated in the presence of increasing number of pile-up collisions, during Run 1 of the LHC.

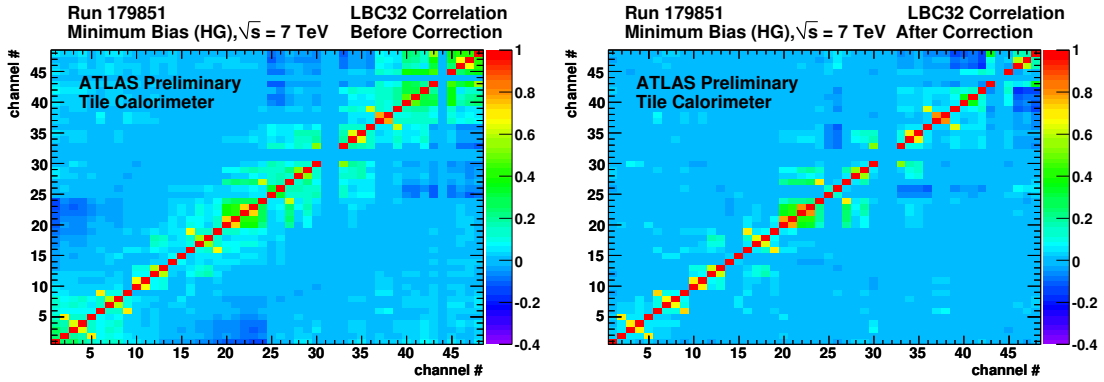
Period	Run Number	Peak Luminosity ($cm^{-2}s^{-1}$)	Peak μ
Apr 13 2011	179581	1.95×10^{32}	5.83
Period D5	180225	4.3×10^{32}	7.11
Period F2	182454	10.72×10^{32}	8.05

Table 4.1: Reference table of the $\sqrt{s} = 7\,\text{TeV}$ data runs analysed.

For pile-up sensitivity purposes, minimum bias triggered data at a center-of-mass energy of $\sqrt{s} = 7\,\text{TeV}$ and 50 ns bunch crossing separation was analysed. Different luminosity runs were studied, as the mean number of multiple collisions $< \mu >$ is directly related with $\mathcal{L}$ of the collision (c.f. Equation 4.5). Table 4.1 summarises the data information, including the maximum number of simultaneous collisions registered per run. A full survey of the TileCal was performed, by constructing covariance and correlation matrices between two readout channels in the same module. Since the correlations were only visible between high gain channels, only these are considered here.

Figure 4.6 shows two dimensional correlation matrices of the high gain signal outputs, where the x and y axis run over all 48 channels of the LBA 32 module. The correlation pattern before (a) and after (b) the TNF filtering of the samples can be compared. Before the signal filtering, in Figure 4.6 (a), regions of high and low correlation values are visible, which reflect the configuration of the TileCal hardware. As already seen in pedestal data, stronger correlations appear between groups of 6 or 12 neighbour channels and are negligible for other pairs of channels. The regions of the front-end drawers which are closer to the LVPS (channels > 40) or in the sub-assembly area (channels 20 to 24), remain the most affected. After the TNF filtering of the samples (Figure 4.6 (b)), the coherence effect on the observed channel noise is reduced, specially for the channels closer to the LVPS source. The need for a double Gaussian distribution, which typically describes the noise behaviour of the TileCal cannot, however, be fully ruled out after removing the correlated noise component within a module.

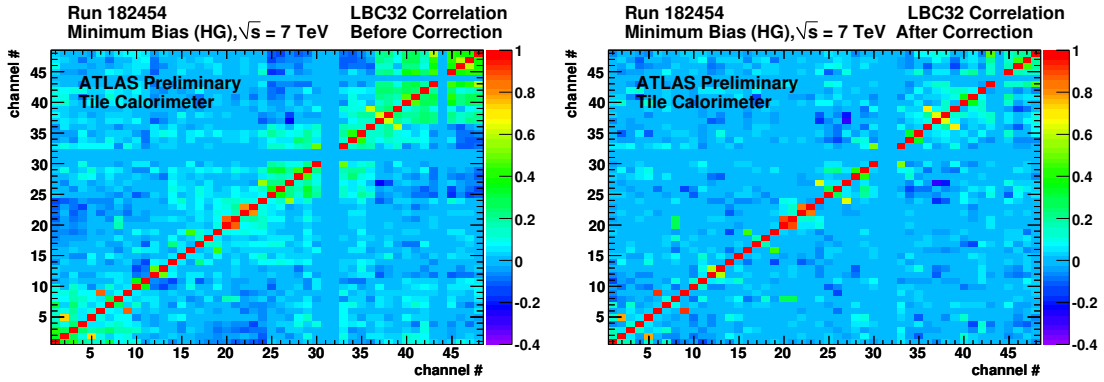
For higher luminosity data, as shown in Figure 4.7, stronger correlations appear. However, these do not follow a coherent behaviour throughout the TileCal modules. The origin of these strong correlations can probably be attributed to the irregular frequency of incoming particles. The soft-scattered particles, selected in this minimum bias run, randomly illuminate clusters of adjacent channels. The pile-up presence deteriorates the channel resolution



(a) (b)

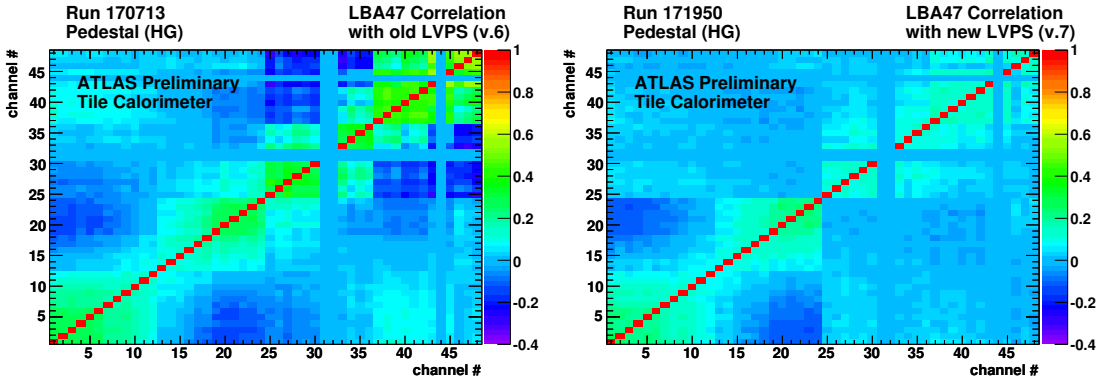
Figure 4.6: Correlation plot of the noise value for all high-gain channels in module LBA32 before (left) and after (right) the unfolding of the noise correlation effect with TNF. The correlations where calculated using minimum bias data with a maximum of 5.8 simultaneous collisions.

and leads to interference between signals of neighbouring channels. Comparing the correlation pattern before (a) and after (b) signal filtering with the TNF method, one can conclude that the presence of pile-up does not affect the performance of the unfolding method, which reduces the correlation.



(a) (b)

Figure 4.7: Correlation plot of the noise value for all high-gain channels in module LBA32 before (left) and after (right) the unfolding of the noise correlation effect with TNF. The correlations where calculated using minimum bias data with a maximum of 8 pile-up collisions.



(a) (b)

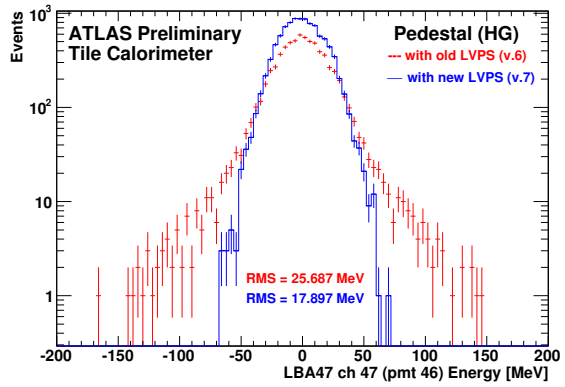
Figure 4.8: Distribution of the LBA47 correlations before (a) and after (b) the replacement of the LVPS (from v.6 to v.7). High gain pedestal data is used, without any correlation removal method applied.

4.4.2 TileCal Correlated Noise: Hardware Improvements

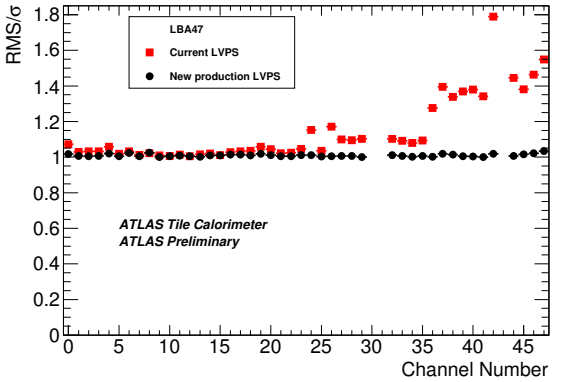
The software studies, developed to deal with the correlations found on the TileCal modules, lead to a deeper understanding of the coherence characteristics. A double approach was then taken: while studies were made to filter the noise effects, an hardware upgrade to effectively solve the source of the problem was also considered. From the correlation noise survey, the LVPS influence on the correlation pattern was evident. This was further corroborated by the positive effect of the algorithms developed to mitigate the correlations. Since the LVPSs seemed the most probable source of interference, the more problematic ones were replaced in the ATLAS cavern . The impact of new generation of LVPS in five modules of the TileCal was tested on pedestal data, without any correction method applied.

Module 47 of the long barrel is a particularly noisy module, with a considerable level of correlation between channels (see Figure 4.8 (a)). After the replacement of the power supply, Figure 4.8 (b), the correlation pattern for channels closer to the LVPS (channels > 40) was mostly removed. The correlation between channels below channel 25, however, remains unchanged. Considering the whole module, the replacement of the LVPS yields $\sim 30\%$ reduction on the correlations between pairs of channels.

Channel 47 is an example of a highly correlated channel with the old LVPS. In Figure 4.9 (a), the observed signal distribution is shown for pedestal data taken before the LVPS upgrade (in red) and after (in blue). The signal follows a double gaussian shape, as mentioned before. The mean energy deposited is zero, as expected from a pedestal run. The equivalent



(a)



(b)

Figure 4.9: Impact of the LVPS upgrade on the electronic noise of module LBA47. (a) Signal distribution of channel 47 of LBA 47 for high gain pedestal data taken before the LVPS upgrade (in red) and after (in blue). Here, no correction method is applied to account for the coherence effect. With the new LVPS, the tails of the distribution are significantly reduced. (b) Ratio between the signal RMS and the width σ of a single Gaussian fit to the signal shape, before (in red) and after (in black) the LVPS change.

standard deviation of the distribution, calculated as the RMS to a single Gaussian, gives an estimate of the electronic noise present on the channel, including the coherent component. The tails of the distribution are significantly reduced with the new LVPS, as can be seen by the RMS values: 25.687 MeV with the original LVPS (v.6) and 17.897 MeV with the new LVPS (v.7). With the new LVPS generation, the smaller fraction of events on the tails of the double Gaussian distribution results in a more single-Gaussian-like behaviour of the electronic noise per channel.

In Figure 4.9 (b), the Gaussian-like shape of the electronic noise, before (in red) and after

(in black) the LVPS change is evaluated. For all channels of LBA 47, the ratio between the measured RMS of the distribution and the width σ of a single Gaussian fit is shown. The ratio of several channels powered by the original LVPS differers significantly from 1, which should be the expected value for gaussian-like noise [206]. Indeed, with the new LVPS generation, the ratio RMS/σ is uniformly closer to unity, confirming the single-Gaussian characteristics of the pedestal signals collected in the TileCal. Considering that the electronic noise in test beam data followed a single Gaussian distribution[1] and the observed improvements of using the new LVPS, the double Gaussian behaviour of the signal is attributed to the presence of the previous LVPS. The survey of the other upgraded modules is consistent with the results shown here for LBA47. The upgrade of the LVPS reduces $\sim 20\%$ the average RMS value of the TileCal modules, which is 20.6 MeV with the new power suppliers. Subsequently, during the 2011-2012 LHC winter shutdown, 40 other LVPS were replaced.

4.4.3 Conclusions

The coherence pattern present on the noise of the TileCal cells was a problematic feature that needed to be addressed, either by correcting these effects or by directly changing the problematic hardware sources. Scans of covariance and correlation matrices of all modules and partitions of the calorimeter were made, as to completely survey the coherence effect on the TileCal. The TNF method, used to correct the observed channel correlation, was validated in pile-up collision data. Although the hardware-related correlations are effectively reduced, the presence of soft scattered multiple collisions leads to irregular clusters of neighbour channels, which determine the behaviour in magnitude and shape of the signal responses. However, correcting or removing these uninteresting pile-up generated signals is not the purpose of such method. In fact, ATLAS has several other methods to identify and deal with these soft signals, usually using vertex information.

The impact of a new generation of LVPS is clearly visible on the Gaussian shape of the channel noise and on the magnitude of the correlations in TileCal modules. Data obtained with the new LVPS shows that the intrinsic white noise distribution of each channel has a smaller RMS and is closer to a single Gaussian shape.

[1] For test beam data, the power source configuration used temporary power supplies located far from the detector.

Period	Run Number	Peak Luminosity ($cm^{-2}s^{-1}$)	Peak μ	$< \mu >$
Period F2	182424	8.5×10^{32}	6.62	4.8
Period G4	183081	11.0×10^{32}	6.9	5.2
Period I2	185747	10.0×10^{32}	6.24	6.1
Period L2	189049	30.0×10^{32}	14.4	11.3
Period L4	189598	26.0×10^{32}	12.4	11.4
Period M2	190617	24.0×10^{32}	11.5	11.3
Period M5	191190	35.0×10^{32}	16.7	10.9

Table 4.2: Summary table of analysed samples for the pile-up impact on the TileCal channel noise.

4.5 Analysis of the TileCal Pile-up Noise

The LHC was designed to collide proton bunches every 25 ns, resulting in approximately 40 extra interactions per beam crossing. The presence of extra interactions, and its ever growing number, has a significant impact on the TileCal performance. For instance, the calorimeter energy resolution can be degraded due to fluctuations on the standard deviation of the energy deposition.Understanding the sensibility of the calorimeter to increasing contributions from pile-up effects is of major importance. In this section, the noise behaviour of TileCal cells under crescent number of pile-up collisions is studied, using data from the first LHC run. The noise dependence with the cell pseudorapidity and with bunch position in a train of proton bunches was investigated, as well as, the noise dependence with the average number of interactions per bunch crossing.

In the following, a full survey of the noise in the TileCal was performed, using data selected with a zero bias trigger, for different luminosities and number of pile-up collisions. Several runs recorded by the zero-bias stream over the course of 2011 data-taking were analysed. In these periods, the LHC was colliding bunches of up to 10^{11} protons 40 million times per second, at a centre-of-mass energy of 7 TeV. Since both beam and detector setups were changing with time, the performance of the detector in varying conditions was surveyed. Table 4.2 is a summary of the data samples used. With $\sqrt{s} = 7\,\text{TeV}$ and 50 ns bunch spacing throughout all samples, each data period corresponds to different beam properties, leading to different luminosity and pile-up conditions. For each run, the maximum instantaneous luminosity and number of pile-up collisions is shown, as well as, the average number of pile-up collisions throughout the run.

As already mentioned, zero bias streamed data may include a few extreme energy values, but will be mostly constituted of soft pile-up events. In order to reduce the dependence of the results on these high energy outliers, the 0.1% fraction of the events in both positive and negative tails of the energy distribution were excluded for all cells. By removing extreme values in the tails of the distribution, the symmetry of the signal shape was improved. Therefore, an accurate estimation of the total noise per channel can still be made using the RMS of the Gaussian-like signal.

In order to simulate the data samples and predict the impact of future pile-up conditions, a dedicated set of MC samples was generated with a single neutrino or a muon as a hard-scattering particle overlaid with a number of soft pile-up events generated by Pythia 6 [201]. In the case of MC sample with a single muon, only the cells with $\Delta R(\text{cell}, \text{muon}) > 0.2$ were considered.[1] The simulation of pile-up in MC events does not completely account for the full pile-up conditions of LHC data, in particular the distribution of the number of pile-up collisions per bunch. A special reweighting technique was applied to the simulated events to accurately describe this effect, using a dedicated tool [207]. The MC simulations were then reweighted to the pile-up conditions of each run to allow a data/MC comparison on run-by-run basis.

As an example, the energy deposition on a TileCal cell is shown in Figure 4.10. Two data periods with different pile-up conditions are represented, in blue for $< \mu >= 6$ and in red for $< \mu >= 12$. Both distributions have a Gaussian shaped core and are centered at ~ 0 MeV, as expected. The sampled signals have long non-Gaussian tails, which are larger the higher the number of pile-up collisions. Since the energy distribution gets wider (with longer negative and positive tails) in the presence of more pile-up events, it is expected that, for data with higher pile-up conditions, calorimeter cells will also have higher values of noise. The data/MC ratio is represented in the lower plot, each colour representing the different pile-up conditions of the collected data. The MC simulations show very good agreement with data in the core of the distribution and in the positive tail. However, in the negative tail larger discrepancies are observed. This reflects the presence of out-of-time energy deposits, which are not well modelled in simulations and are a major source of degradation in the sampling of the signal.

At ATLAS, the pile-up conditions are typically assessed using two common observables: the average number of interactions per crossing and the actual number of interactions per

[1] $\Delta R^2 = \Delta\phi^2 + \Delta\eta^2$

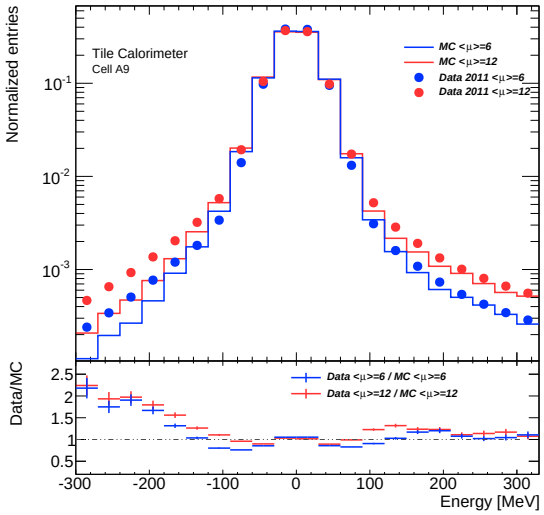


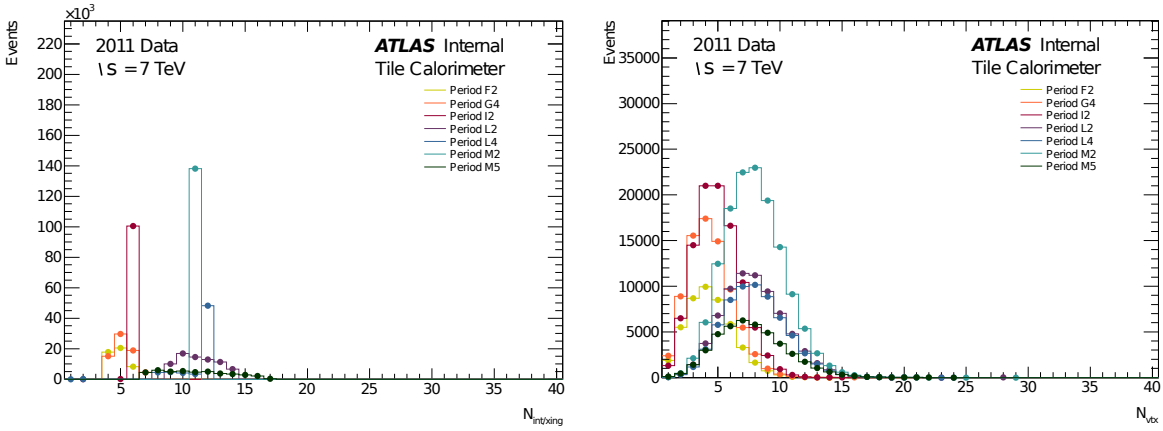
Figure 4.10: Energy distribution in a cell (cell A9) for 7 TeV data with 50 ns bunch spacing collected in 2011. Two different pile-up conditions are depicted: in blue $\mu = 6$ and in red $\mu = 12$. The markers represent data and the histograms the MC sample. The data/MC ratio is represented in the lower plot, each colour representing the different pile-up conditions of the collected data [2].

crossing. The first one consists in calculating the number of interactions per crossing, averaged across all bunch crossing identifiers (BCIDs) in the lumiblock, $< \mu >_{\mathrm{LB,BCID}}$.[1] Since each bunch can have a different emittance and a different number of protons, μ values can vary for different BCIDs. The second estimate of the pile-up level, is the number of interactions in a given BCID, averaged over the lumiblock, $< \mu >_{\mathrm{LB}}$ (BCID). Since the Monte Carlo simulation does not include the bunch structure, both quantities are the same in simulated events.[2]

The average number of interactions per crossing, $< \mu >_{\mathrm{LB,BCID}}$, is represented in Figure 4.11 (a), for all the run periods included in Table 4.2. For comparison purposes, the data distributions are normalised to unity. As the instantaneous luminosity changes from run to run, so does the number of pile-up collisions. For example, period I2, with maximum luminosity of $10 \times 10^{32}\mathrm{cm}^{-2}\mathrm{s}^{-1}$, has an average number of interactions per bunch crossing $< \mu >_{\mathrm{LB,BCID}} \approx 6$, while period M2, with maximum luminosity of $24 \times 10^{32}\mathrm{cm}^{-2}\mathrm{s}^{-1}$, has

[1] Here, LB stands for the luminosity block and is not to be confused with the long barrel of the TileCal.

[2] Note that the pile-up reweighting of the MC to data is calculated using the average number of interactions per crossing ($< \mu >_{\mathrm{LB,BCID}}$) [207].



(a) (b)

Figure 4.11: Distributions of (a) the average interactions per bunch crossing and (b) the number of primary vertices for several periods of 7 TeV data with 50 ns bunch spacing. The markers represent the data sample, one run from each period F2, G4, I2, L2, L4, M2 and M5 according to the colours in the legend, and the histograms the MC sample [2].

$< \mu >_{\mathrm{LB,BCID}} \approx 11$. Run periods F2 and G4 have the fewer interactions per bunch crossing and have also registered the smallest peak luminosity throughout their duration. The highest peak luminosity analysed in this study was recorded in periods L2 and M5, which also have the highest maximum number of interactions per BCID, 14.4 and 16.7 respectively. However, during these runs the $< \mu >_{\mathrm{LB,BCID}}$ values are not constant and span over a wider distribution, leading to $< \mu >$ values slightly lower (~ 11).

In Figure 4.11 (b) the distribution of the number of primary vertices for each period is shown. As expected, for higher luminosity data, the mean value of vertices per event increases, which is consistent with the higher number of pile-up interactions. Runs from periods F2, G4 and I2 show similar behaviours in the distributions of the average number of pile-up interactions and number of primary vertices. By comparison with the other four runs (from periods L2, L4, M2, M5) one can say F2, G4 and I2 are the low pile-up samples in this study. The four runs with higher pile-up have an average number of primary vertices of the order of 10, which is consistent with the extra 8 collisions expected from pile-up for these luminosity conditions.

4.5.1 Energy and Noise Pseudorapidity Dependence

The TileCal cells have a diverse geometry, specifically designed according to their location on different radial layers and at different pseudorapidity angles (c.f. Figure 3.9). As seen in the previous study, the electronic noise component varies with the pseudorapidity of the cell, due to the readout system architecture. In the presence of collisions, the cells which are closer to the beam pipe are exposed to a larger amount of low-energy particles, which results in differences in both electronic and pile-up noise components for the various cells. Having two long barrels within the $|\eta| < 1.0$ region and two extended barrels covering $0.8 < |\eta| < 1.7$ region, the TileCal is symmetric in η. In order to study the calorimeter dependence on the pseudorapidity, the η symmetry was verified both for energy and noise measurements. Moreover, due to the calorimeter cylindrical shape, there is homogeneity in the ϕ modules. In this study, the energy and noise dependence on ϕ was proven to be approximately constant for the same η tower.

In Figure 4.12 the mean energy deposited in a specific layer l as a function of pseudorapidity η is represented. The mean value of the OF reconstructed signals, after pedestal subtraction and corrected for the calorimeter constants were used. It was averaged over all 64 modules of both negative and positive η values. The different periods are depicted in different colours and can be compared with the lowest luminosity sample from period F2, in yellow. Each marker represents a TileCal cell layer in both extended and long barrel of the calorimeter: A-cells being the closest to the electromagnetic calorimeter and D-cells the outermost layer. For these zero bias triggered runs, the mean energies are of the order of a few MeV. In the central region of the calorimeter, where the long barrel (LB) covers $|\eta| < 1.0$, the mean energy depositions are always less than 1 MeV for all the samplings. In the extended barrel partitions, depositions are more energetic and the different layers of cells show distinct behaviours. The cells in the gap-crack region exhibit higher levels of energy deposits which is expected due to their special nature. When comparing different run periods, it is noticeable that the higher the luminosity, the higher the average energy depositions in the calorimeter, particularly in the extended barrel cells. The presence of higher energy collisions tends to be in the forward regions of the ATLAS detector, which is in agreement with the results.

The Monte Carlo to data agreement was studied in Figure 4.13. The differences range from $-2\,\text{MeV}$ to $2\,\text{MeV}$ with a discrepancy of the order of $1\,\text{MeV}$ for the long barrel region.

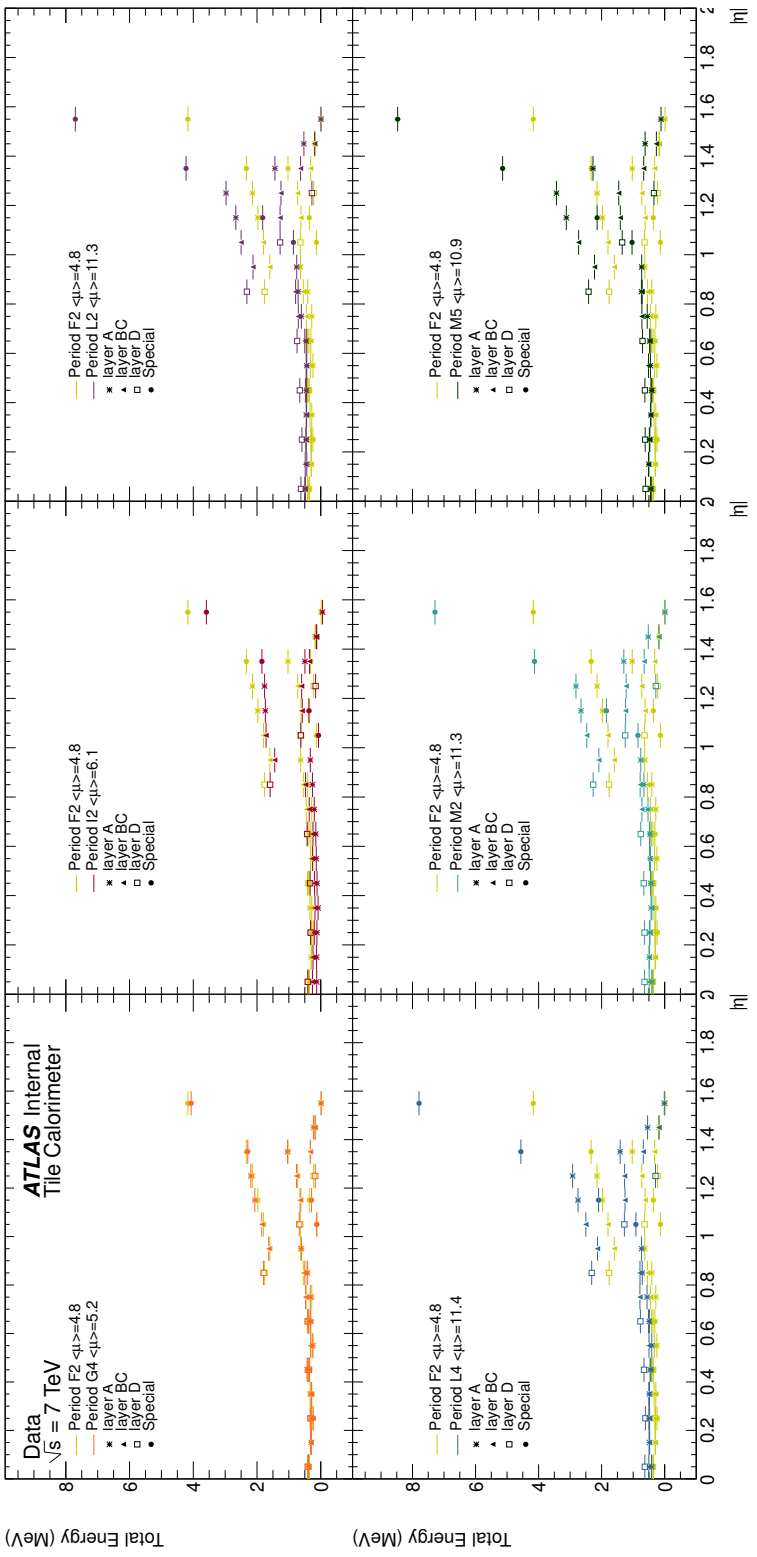


Figure 4.12: The mean energy deposited in each cell, averaged over all modules, is represented as a function of $|\eta|$ for several periods of 7 TeV data, according to the colours in the legend. The reference value is the F2 period run in yellow [2].

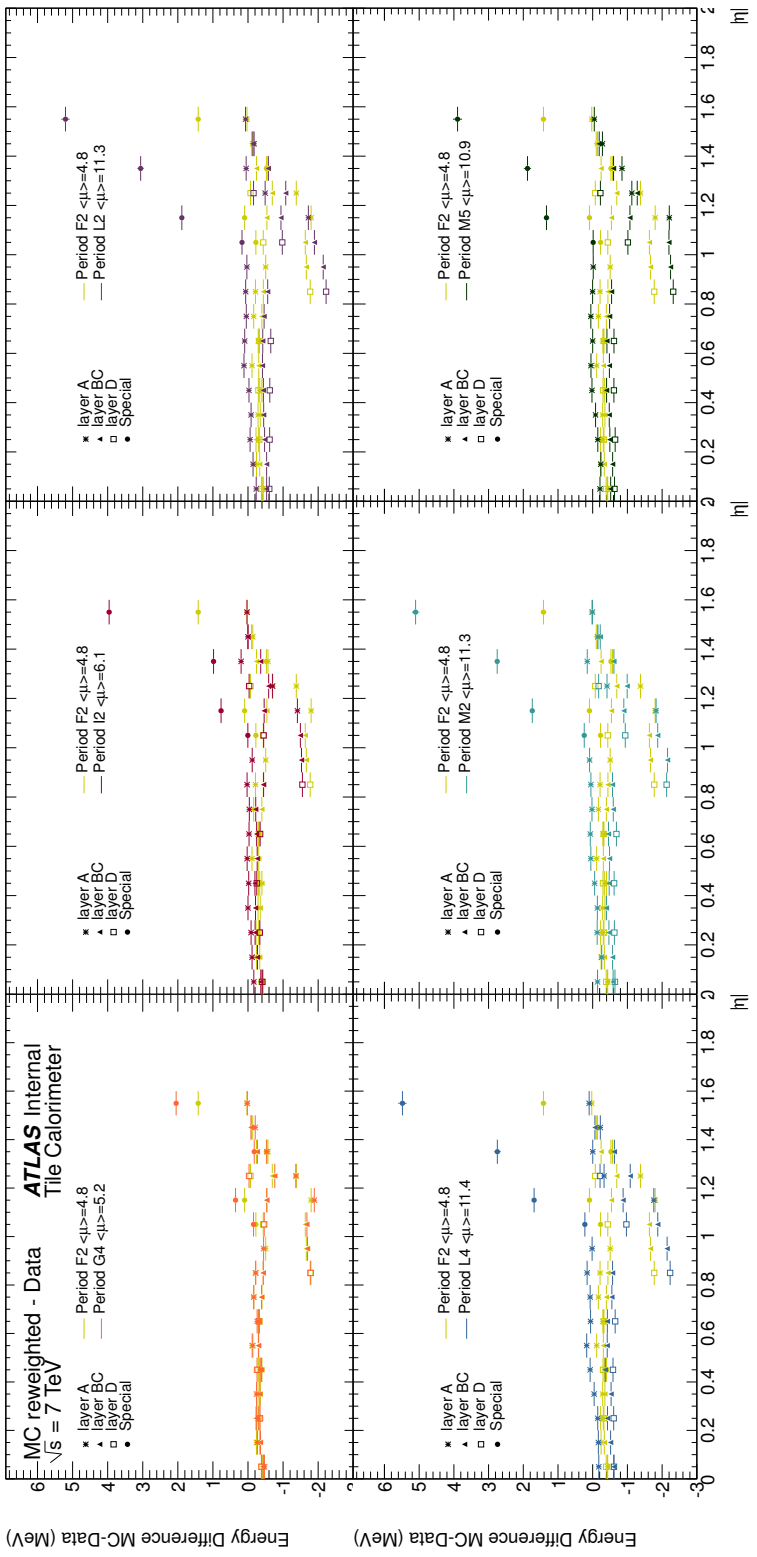


Figure 4.13: The MC-Data difference of the mean energy deposited in each tower is represented for several runs of 2011 data. In each histogram the distribution of one run is compared with the F2 run represented in yellow (reference value) [2].

For the gap-crack cells, the differences between simulation and data are still consistent with 2 MeV in the low pile-up runs, but vary up to 5 MeV in the high pile-up samples. Differences of this magnitude between data and MC are much smaller than the precision of the energy measurement.

The noise of a cell, including both the electronics and the pile-up contributions, was estimated as the RMS of the energy deposited within that cell, according to Equation 4.5. The uncertainty on the noise is obtained from the error of the RMS of a gaussian distribution. Taking advantage of the TileCal structure, i.e. ϕ homogeneity and η symmetry, the mean noise was obtained inclusively over all modules and $|\eta|$ as the RMS of the reconstructed energy per cell. In Figure 4.14 the total noise of each TileCal cell is represented as a function of the cell pseudorapidity, $|\eta|$, for several periods of data-taking. Data periods are shown in different colours and are compared to the lowest luminosity sample, period F2, in yellow. In the long barrel partitions, the noise values range from 20 to 50 MeV for the low pile-up periods (F2, G4 and I2) and the calorimeter response is constant within the same type of cell layer. The noise increases with proximity to the beam pipe, thus A-cells have higher noise than D-cells. In the extended barrel partitions, noise values are higher (up to 180 MeV) and the behaviour of different layers of cells is more erratic. For the higher luminosity runs (periods L2, L4, M2 and M5), the cell noise increases proportionally to the extra pile-up events throughout the whole calorimeter: going up to 60 MeV in the LB regions and up to 50 MeV in the EB. Figure 4.15 shows the difference between the mean value of noise as a function of $|\eta|$ measured in MC and in Data. Differences up to 10 MeV in the long barrel partitions and up to 40 MeV in the extended barrel are observed.

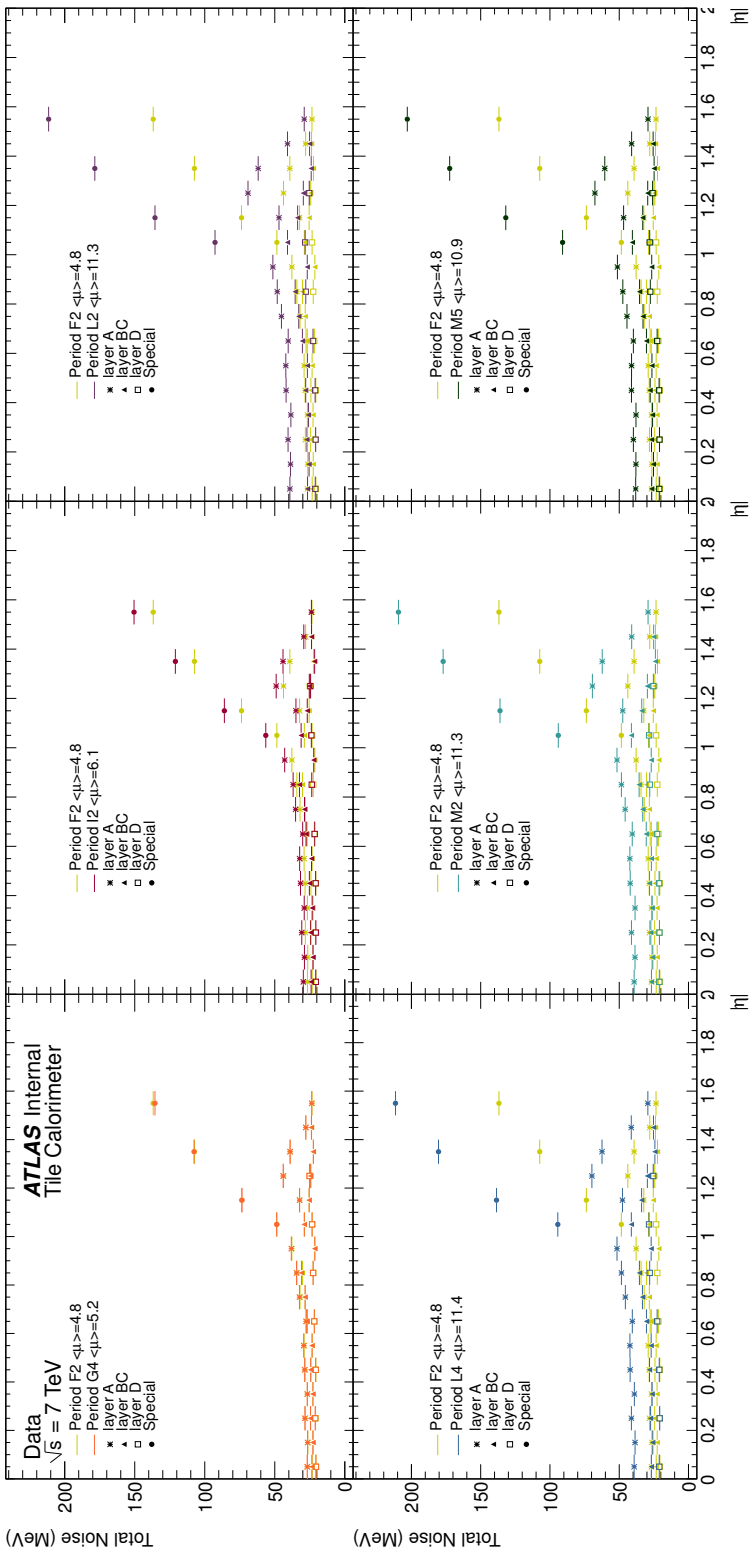


Figure 4.14: The total noise in a cell computed inclusively over all modules is represented as a function of $|\eta|$ for several periods of 7 TeV data. The reference value is the F2 run in yellow. The markers represent the cell layer and different colours are used for each period F2, G4, I2, L2, L4, M2 and M5 according to the legend [2].

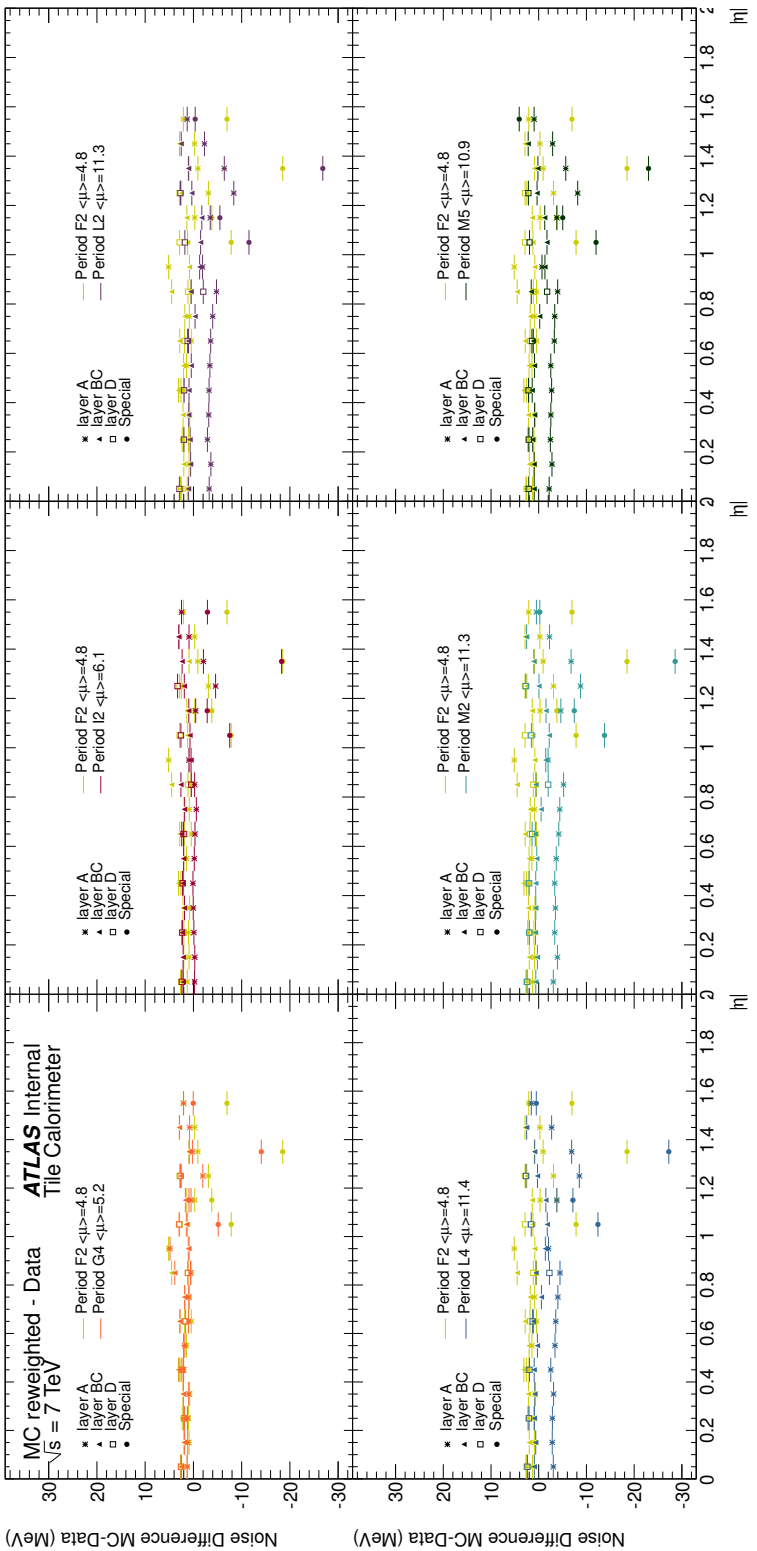


Figure 4.15: The MC-Data difference of the total noise in a cell, computed inclusively over all modules, is represented for several runs from different periods of 7 TeV data. In each histogram the distribution of one run is compared with the F2 run represented in yellow (reference value) [2].

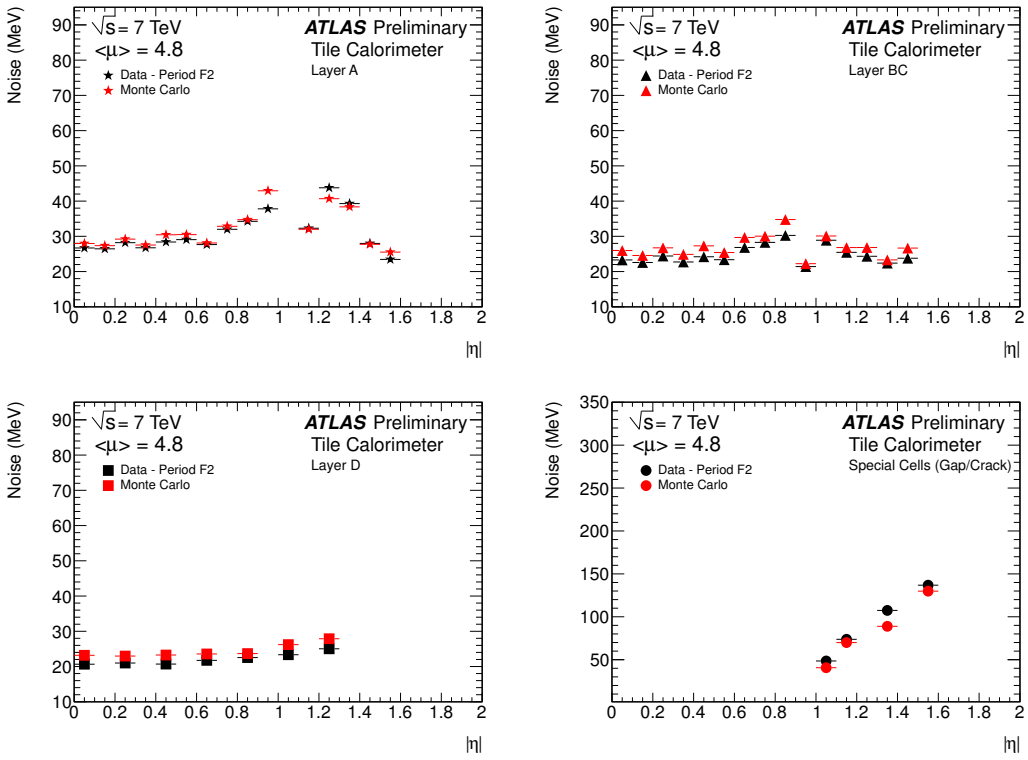


Figure 4.16: The noise distribution in different TileCal cells is represented as a function of $|\eta|$ of the cells for zero bias run 182424 of period F2 of 2011 data with an average number of interactions $< \mu >= 4.8$ per bunch crossing [2].

Figures 4.16 and 4.17 respectively for a low pile-up run (period F2 with $< \mu >= 4.8$) and a high pile-up run (period M2 with $< \mu >= 11.3$) are shown for better visual discrimination of the noise behaviour from different cell layers when subjected to pile-up collisions. Each plot corresponds to a different cell layer: A-cells on the top left, BC-cells on the top right, D-cells on the bottom left and special cells on the bottom right. As already mentioned, given the TileCal azimuthal symmetry, the information around the 64 modules was integrated in ϕ, and $|\eta|$ was used. In the case of the reference sample F2, in Figure 4.16, noise values range from 20 MeV to up to 45 MeV for A, BC and D layers. Here, the extended-barrel ($0.8 < |\eta| < 1.7$) is slightly noisier than the long-barrel. Nevertheless, the noisier cells are the special cells, where noise values range from $\sim$ 40 MeV to 150 MeV. Due to their location in the gap-crack region, these cells are particularly exposed to forward high energetic radiation coming from the IP. The MC describes data up to a 2 MeV (5 MeV) difference for

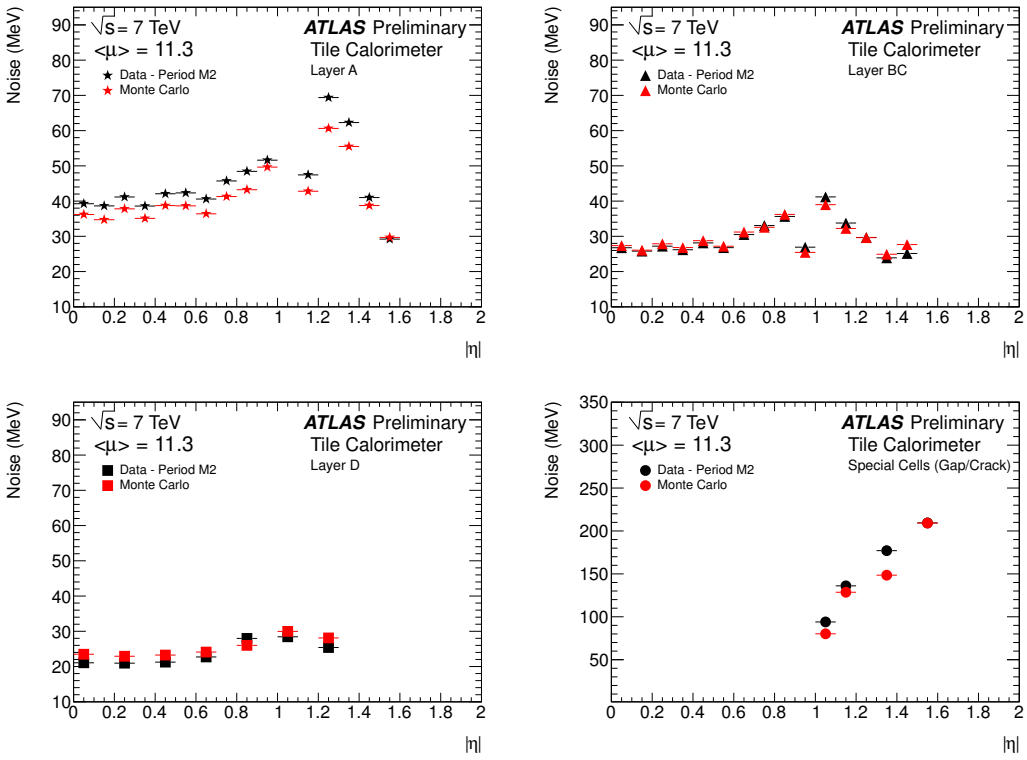


Figure 4.17: The noise distribution in different TileCal cells is represented as a function of $|\eta|$ of the cells for zero bias run 190617 of period M2 of 2011 data with an average number of interactions $< \mu >= 11.3$ per bunch crossing [2].

the A, BC and D layers of the LB (EB). The special cells show differences between the mean value of noise measured in MC and in Data up to 20 MeV.

For high pile-up conditions, the noise values for A-cells range from 40 MeV to up to 70 MeV, as can be seen in Figure 4.17. Noise values are around 20 MeV for D cells and varies from 30 MeV up to 40 MeV for BC cells. This is consistent with the TileCal cell structure, since cell layers closer to beam pipe receive more energy deposition from particles originated in the pile-up events than the ones in the outer layers of the calorimeter. Just as in the low pile-up run, the extended-barrel is slightly noisier than the long-barrel, mostly due to the more forward coverage it entails. In the presence of more simultaneous collisions, the special cells present a much noisier behaviour, with noise values ranging from $\sim$ 70 MeV to 200 MeV. As already mentioned, this is understandable due to the nature of their location, immediately after the IP and before the LB partitions. The data to MC agreement shows

differences of the order of 3 MeV for the A layer in the LB and goes up to 8 MeV in the EB. For the BC and D layers the data to MC differences throughout the whole calorimeter are of the order of 2 MeV and 3 MeV respectively. The differences between simulation and data in the gap-crack cells vary up to 30 MeV. The imperfect modelling of the out-of-time energy depositions in the simulations accounts for most of the observed differences between data and MC.

In the $1.0 \leq |\eta| \leq 1.2$ region, there is a change in the noise behaviour for both low and high pile-up setups. This is simply a reflection of the TileCal geometry for here is where the gap with inactive materials (such as cables) between the TileCal Long-Barrel and Extended-Barrel partitions is located (c.f. Figure 3.9). In these pseudorapidity towers, the cells must be smaller in order to accommodate the service structures going to the ID, and thus have less scintillating material. This, together with the fact that some of the energy deposition in the inactive materials is lost, explains the slight reduction of the noise value for $1.1 \leq |\eta| \leq 1.2$.

4.5.2 Energy and Noise Dependence on the Bunch Position

During the analysed data periods (referred on Table 4.2) bunch trains with approximately 60 filled bunches, spaced by 50 ns intervals, were accelerated by the LHC. Each bunch train was followed by a large time interval with empty bunches. Since the readout window of the signal sampling (150 ns) was much larger than the time interval between consequent bunches crossing (50 ns), the energy reconstruction of a given event could be biased by energy deposits in the same cell during the previous (or following) event. As a result the shape of the energy distribution could depend on the position of the bunch crossing in the train.

In Figure 4.18 the number of events as a function of the bunch crossing position in a train for all the studied periods is shown. Events throughout the different runs show a similar distribution, the majority of them having train lengths of up to 60 bunches and a small fraction of events having trains up to 70 bunches.

TileCal cells from the same layer and partition exhibit similar energy and noise performances, even though they are located in different pseudorapidity towers, as proven in the previous section. In order to simplify the analysis and simultaneously increase the statistical power, the BCID dependence of both the mean energy deposits and its respective associated

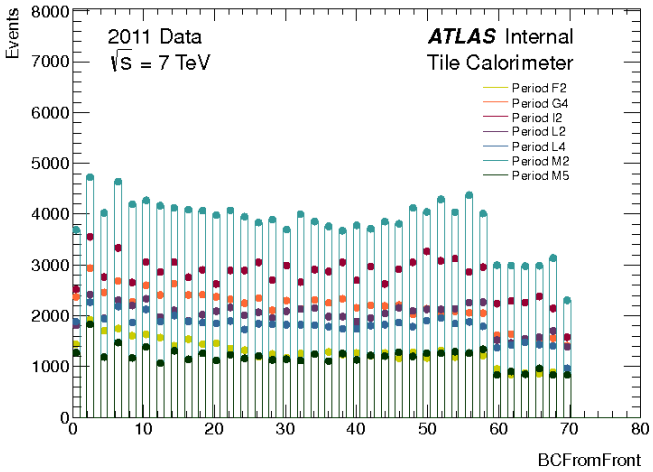


Figure 4.18: Distribution of the bunch crossing IDs for several periods of $\sqrt{s} = 7\,\text{TeV}$ data. The dots represent the data, one run from each period F2, G4, I2, L2, L4, M2 and M5 according to the colours in the legend [2].

noise values was done cumulating over ϕ modules and η towers, within the same layer and partition.

In Figure 4.19 the mean energy deposits as function of the BCID, for cell layer A of the long barrel is represented. A full survey of the TileCal cells can be seen in Appendix A.1, for both barrels and all layers of the calorimeter. The mean energy deposition fluctuates with the BCID, having a slight tendency to be higher for the first and last bunches in the train. A similar behaviour is seen for all layers of the TileCal. The data events in the D layer, however, show slightly lower mean energy values for the last bunches in a train. As seen in the pseudorapidity distributions, the energy value increases with proximity to the beam pipe: for A-cells the deposited energy fluctuates around higher mean levels than in the D-cells. The extended barrels show slight higher values of energy depositions than the long barrel ones, as expected.

The cell total noise was also surveyed, as the RMS of the energy distribution per cell. For long barrel A-cells, the total noise as a function of the BCID is represented in Figure 4.20. In Appendix A.2 the full survey of the TileCal layers is included. Just as the mean energy measurements, the noise is constant within a few MeV ($> 5\,\text{MeV}$ for the A-cells). All layers of TileCal have slightly noisier signal measurements for the first and last bunches in each train. Periods with higher multiplicity of pile-up collisions have noisier cell responses. In the special cells, for instance, the mean noise fluctuates around 150 MeV for the low pile-up

periods and around 220 MeV for the higher pile-up runs. The further away from the beam pipe in the radial direction (from A cells to D cells), the noise tends to decrease from layer to layer, as seen in the η studies.

4.5.3 Noise Dependence on the Number of Interactions per BCID

The increasing number of pile-up collisions results in the widening of energy distributions measured at the TileCal cells. Therefore, higher values of noise in each cell are expected for increasing multiplicity of simultaneous collisions. The proportionality of the noise with the average number of interactions per bunch crossing was investigated for all the different cell types. Here only LBA and EBA cell measurements are presented, due to the already confirmed similar response of the negative pseudorapidity partitions.

In Figure 4.21, the inclusive noise of cells from different layers (A, BC, D and the special cells E) as a function of $< \mu >$ is shown. For both long and extended barrel, a representative cell from each layer was selected and only signals measured by that readout cell are included. Located at $\eta = 0.45$ or $\eta = 1.15$, the selected cells display typical and consistent behaviour of LB or EB cells respectively. Similarly to the previous sections, the noise is estimated as the RMS of signal measurement in each cell, thus including components from electronic and pile-up sources. In all four cell types, the noise smoothly grows as the number of interactions per bunch crossing increases.

In order to quantify the noise increment with $< \mu >$ linear fits were performed to the typical cell noise response. In Figure 4.21 the fit functions are represented by the dashed lines and their values are included in the legend. The noise dependence is more steep for the A cells than for the BC cells, and for these ones more than for the D cells. This is a consequence of the calorimeter layer structure, where the layers closer to the beam pipe are more susceptible to pile-up and display noisier readouts. As already seen in previous sections, the cells in the gap-crack region show higher values of noise. These special cells are also more sensible to changes in pile-up conditions, showing a more prominent rate in the noise as function of the number of interactions per bunch crossing. When the linear fits are propagated to $< \mu >= 0$, the noise values are consistent with the electronic noise of these cells, as expected.

In the outer layers of Tile calorimeter (BC and D), the noise is directly proportional to $< \mu >$, increasing linearly with the level of pile-up. However, for cells located closer to the beam pipe, such as A-cells or gap-crack cells (which are under more intense pile-up

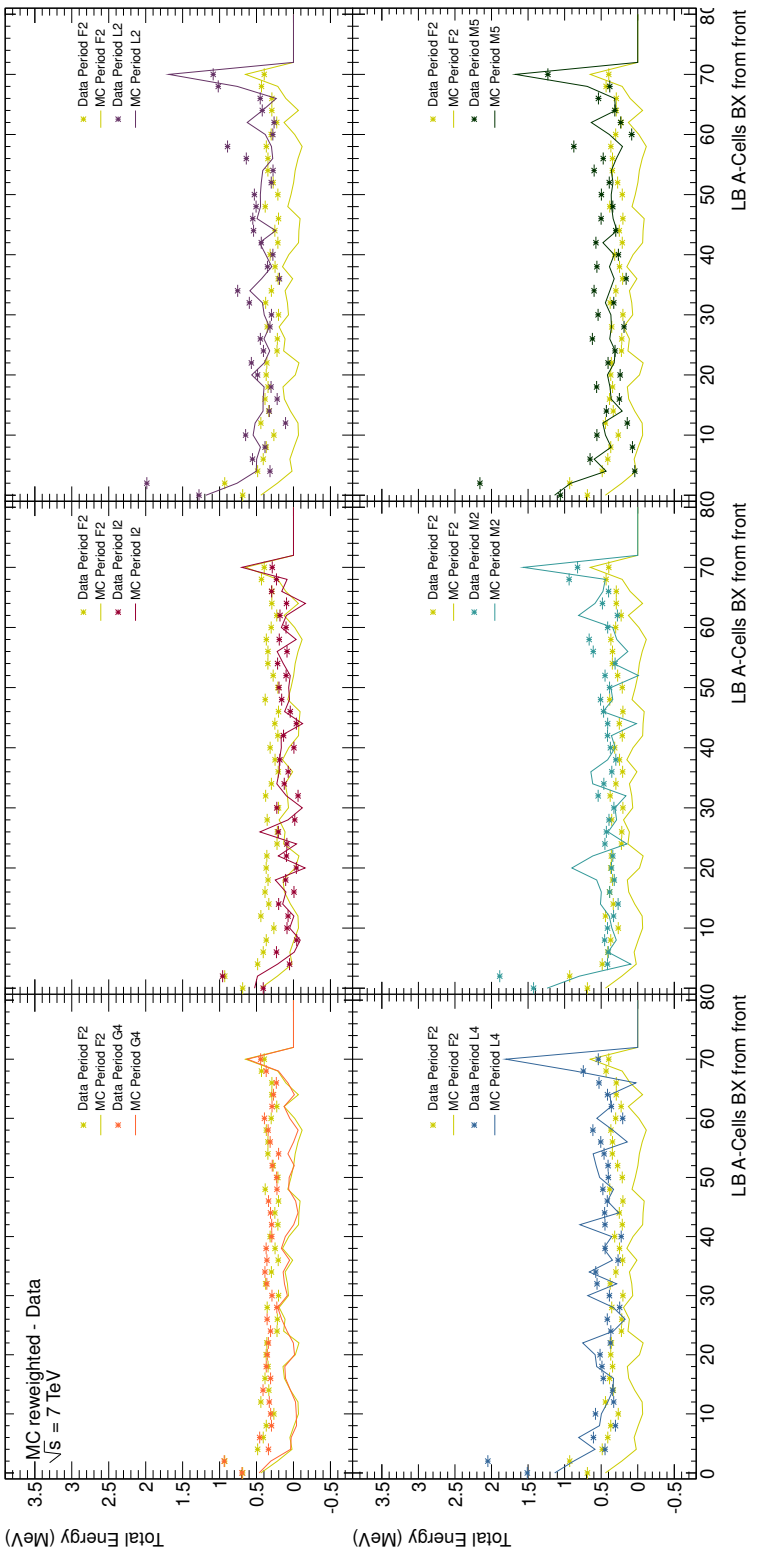


Figure 4.19: The mean energy measurements as a function of the BCID for sample A of the long barrel (LB). Represented for several runs of $\sqrt{s} = 7\,\text{TeV}$ data, with the lowest luminosity period (F2) chosen as a reference. The marker points represent the data sample and the lines are the Monte Carlo distribution.

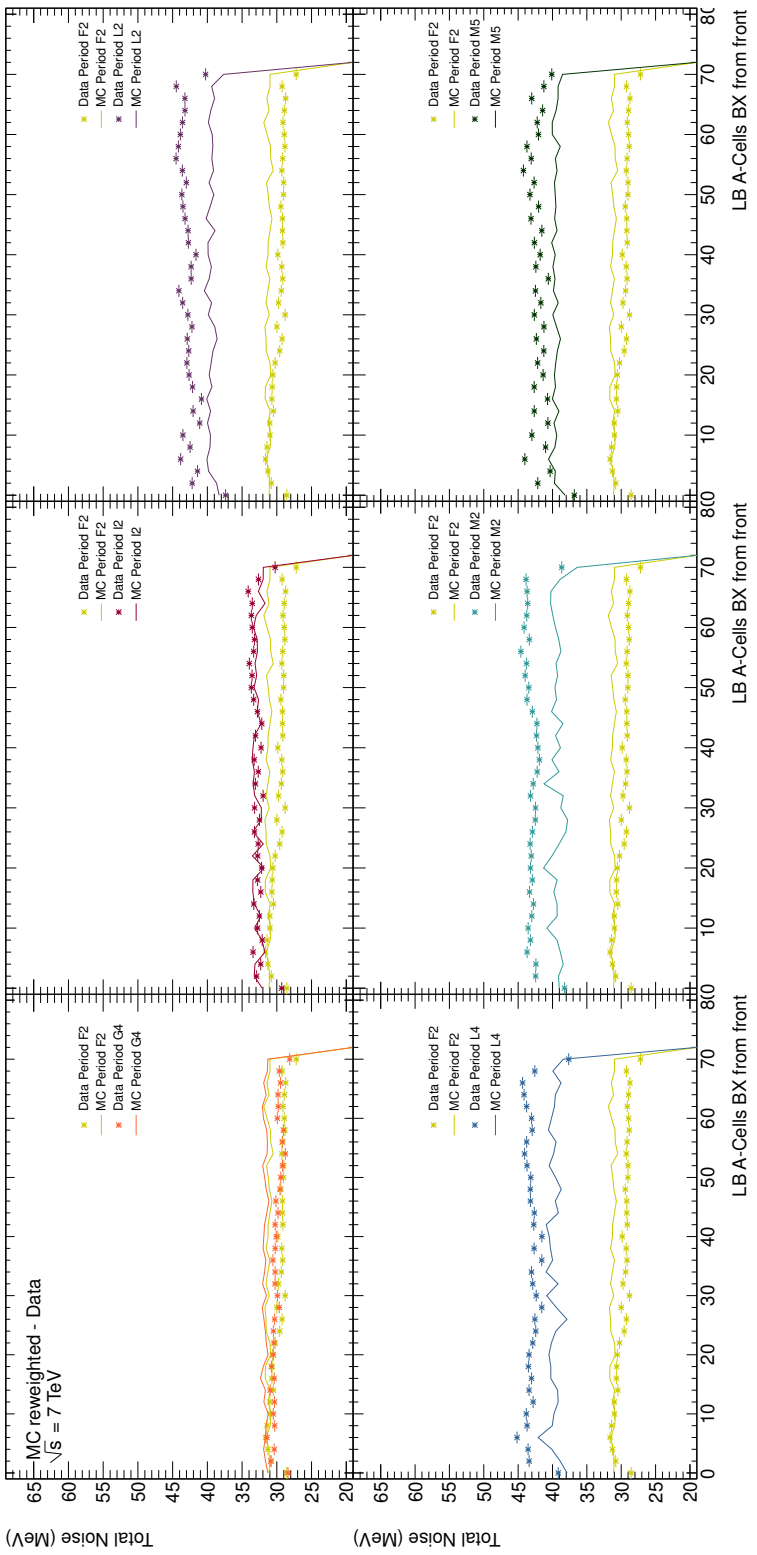


Figure 4.20: The total noise as a function of the BCID for sample A of the long barrel (LB), for several runs of $\sqrt{s} = 7\,\text{TeV}$ data. The lowest luminosity period (F2) was chosen as a reference. The marker points represent the data sample (one run for each period) and the lines are the Monte Carlo distributions (mc11c reweighted to each data period).

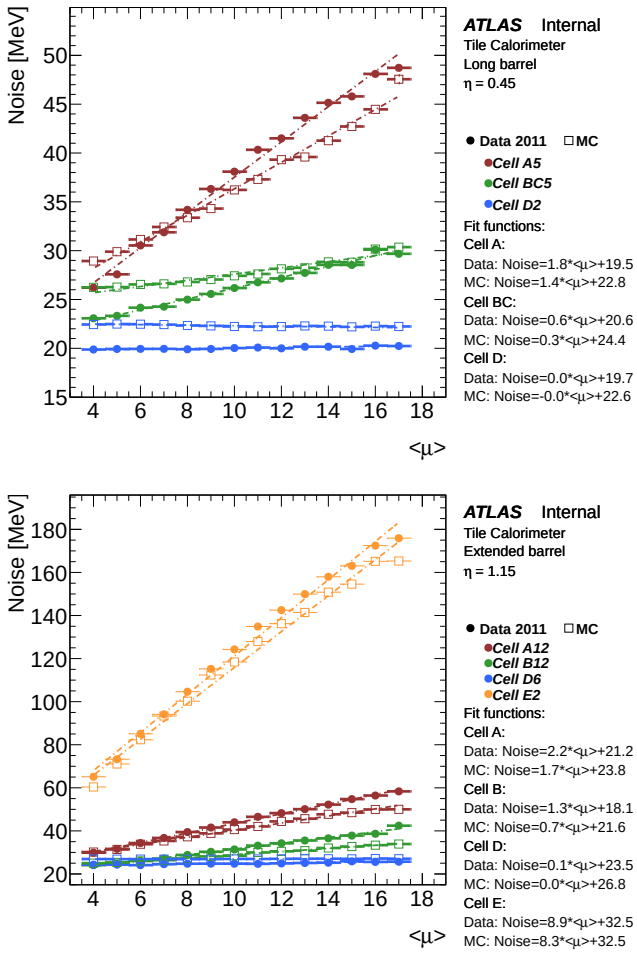


Figure 4.21: Noise distribution as a function of the average number of interactions per crossing for different cell layers of the LBA (top) and EBA (bottom) TileCal partitions. Typical cells from each layer and partition are represented by different colours: cell A in bordeaux, BC in green, D in blue and E in orange [2]. The points represent $\sqrt{s} = 7\,\text{TeV}$ data and the MC simulation is represented by squares. The dashed lines represent linear fits to the data distributions.

radiation), the noise does not depend linearly on the number of pile-up events, showing linear fits with high χ^2. Here, the noise distribution is better described by a square root growth with the increasing number of simultaneous *pp* collisions. In fact, this is exactly what is simulated in the MC samples, where the pile-up noise component is estimated as increasing with $\sqrt{<\mu>}$.

4.5.4 Conclusion

An innovative study of the TileCal readout cells response with the increasing pile-up level of the LHC collisions was performed. The dependence of both the measured signal deposits and associated noise on the multiplicity of pile-up events was investigated. The noise of a cell was estimated as the RMS of the energy deposit, including contributions from electronic noise and pile-up sources. In zero bias data the electronic noise was small when compared to the pile-up effect on the noise of each cell. A complete survey of the calorimeter was carried out, including all partitions, modules and layers. The dependencies with η, $BCID$ and number of interactions per bunch crossing were investigated.

Throughout the data periods analysed, the luminosity of the runs was consistently increased, and as such, a wide variety of average number of interactions per event were examined. As the multiplicity of pile-up interactions increases, the mean energy deposited in a cell also increased. The behaviour of the cells was constant from run to run and reflects the geometry of the TileCal: cells closer to the IP (A-cells) or in more forward regions (gap-crack cells) receive higher energy deposits. Cells subjected to higher pile-up radiation were also more prone to noisier outputs. Furthermore, cell measurements of data with higher number of pile-up interactions showed higher values of noise. These values, however, were not significant to the energy measurement, since they were inferior to 250 MeV. Given the special structure of the gap-crack cells, these tend to be more susceptible to the pile-up noise than the normal cells of the calorimeter. However, measurements form these cells are not used in normal data-taking runs, as per their design.

Overall, the cell noise increased with the number of interactions per bunch crossing. The rate to which the noise depends on the number of interactions is specific of the cell location in the calorimeter. As expected, cells which undergo higher levels of pile-up radiation, such as A-cells or special cells, present typically noisier measurements, with values growing with $\sqrt{<\mu>}$. The noise of the outermost cells, located in the BC or D layers, showed linear proportionality to the average number of interactions in a bunch crossing.

The Monte Carlo simulations, reweighted to the pile-up conditions of each period, described the $\sqrt{7}$ TeV data. The simulations can be used to remove the noise from the TileCal measurements, contributing to higher resolution in the jet energy reconstruction. A correct model of the noise dependence with pile-up contributes to a better description of noise in simulation, allowing better and more precise results of the ATLAS experiment. This

study was summarised in 'Description of the Tile Calorimeter noise with increasing Pile-up for $\sqrt{s} = 7\,\text{TeV}$ data collected during 2011' [2]. It was the first complete survey of the TileCal cell noise dependence with the multiplicity of pile-up events, and gave way for its continuous monitoring throughout Run 1 of the LHC, with different centre-of-mass energies, luminosities and bunch spacings.

5

General Setup for the $t\bar{t}H$ Search

This chapter describes the analysis and data setup for the search for the SM Higgs boson, in the dileptonic channel of $t\bar{t}H$ production, at a centre-of-mass energy of $\sqrt{s} = 8$ TeV. An overview of the dataset and simulated MC samples is given in Section 5.2, including a brief description of the relevant MC generators, as well as, the ATLAS detector simulation. Section 5.3 presents the trigger criteria used for the selection of the $\sqrt{s} = 8$ TeV dataset. The reconstruction of the main physics objects and the event preselection criteria are detailed in sections 5.4 and 5.5, respectively.

5.1 Introduction

The search for the SM Higgs boson in the $t\bar{t}H$ production mode with the ATLAS detector is the main data analysis of this thesis. Here, the dedicated study of the dileptonic topology together with a $H \to b\bar{b}$ decay is presented. The combined analysis of semileptonic and dileptonic $t\bar{t}H$ decays has been published in the European Physical Journal C [1].

The final state of the $t\bar{t}H$ dileptonic production has quite a clean signature, as already mentioned in Chapter 5.1. The top quark decays mostly to a W boson and a b quark, and in the case of a leptonic decaying W boson, a charged lepton and a neutrino will also be present. Since the neutrinos are weakly interacting particles, they traverse the ATLAS detector undetected. Despite escaping direct detection, the neutrinos presence can be inferred using conservation of momentum in the transverse plane. This analysis was designed to be mostly sensitive to the Higgs boson decay into a pair of b quarks, since this is the expected predominant decay. Nonetheless, all Higgs decays are treated as signal for a more inclusive analysis. The final topological signature will thus involve two oppositely

charged leptons, arising from the leptonic decay of both W bosons, as well as at least two b-tagged jets from the top quark decays. By selecting collision events with this topology, the presence of the $t\bar{t}H$ process can be investigated.

Following the signal topology is important for the selection of $t\bar{t}H$ events, but it does not guarantee that only signal events are accepted, especially considering the small predicted signal cross-section. The main sources of event contamination will then come from processes emulating the signal signature. Processes involving top quarks, particularly top quark pairs, are therefore the major source of background. Additionally, high cross-section processes including leptons will also mimic the targeted topology. Vector boson production in the presence of extra jets (Z and W+jets), as well as diboson production (WW, WZ and ZZ), can originate real leptons and contribute to the background. Another important source of background to the dileptonic $t\bar{t}H$ process are events where mis-identification of leptons occur. Either real leptons with mis-measured properties, such as electrical charge sign, or mis-identified jets can lead to the selection of events with fake leptons.

In order to search for the SM Higgs boson in the dileptonic $t\bar{t}H$ production, the capability to reproduce the observed physics processes of interest is evaluated. This is commonly done by comparing data with the simulation of both signal and background, from the evolution of the hard process to the final objects observed with the detector. Monte Carlo (MC) simulation is used to model both signal and background processes. In this thesis, dedicated background studies were developed to match the MC predictions to the data observations, by resorting to control regions orthogonal to those included in the analysis.

The analysis focuses on identifying and reconstructing the particles present in the final signal topology. For this purpose, a clear and unequivocal physics object definition is implemented, according to detector specifications, data quality criteria and targeted analysis efficiency.

Events are then preselected, requiring several criteria to enhance the signal significance and mitigate the background presence. Detector inefficiencies are also considered when selecting the best quality data for a reliable analysis. The general setup for the $t\bar{t}H$ search is introduced in the following sections.

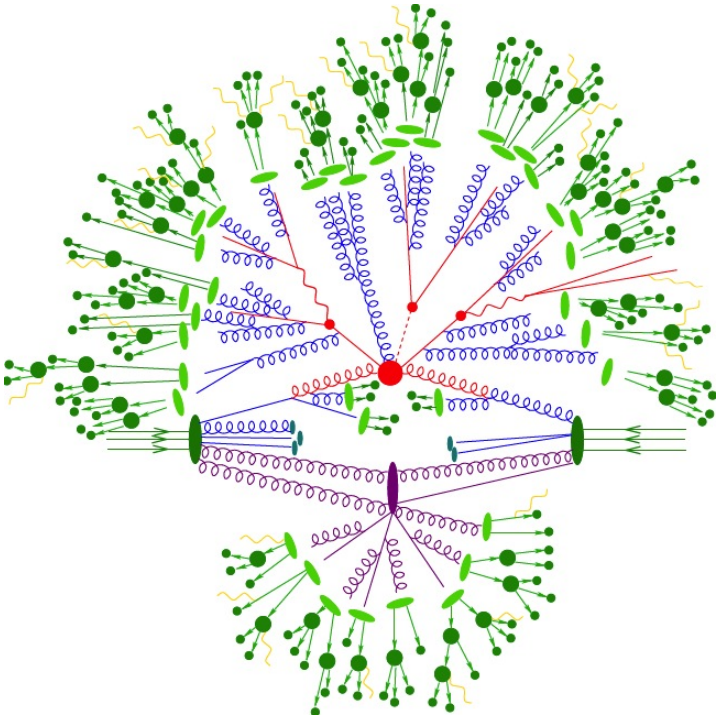


Figure 5.1: Scheme of a *pp* collision simulation with the different factorised steps [208].

5.2 Data and Monte Carlo Samples

The analysis presented in this chapter uses the full dataset of *pp* collisions at a centre-of-mass energy of $\sqrt{s} = 8$ TeV recorded by ATLAS during Run I. Taking into consideration the high quality requirements for physics analysis, an integrated luminosity of $20.3\,\text{fb}^{-1}$ of data was analysed, with pile-up reaching up to about 40 simultaneous interactions per bunch crossing (c.f. Figures 3.13 and 3.14 in Chapter 3.3.6).

In order to evaluate the consistency with the SM, the observed data is compared with detailed theoretical predictions. The simulated samples are created using Monte Carlo (MC) event generators, which reproduce the high energy events from the deep-inelastic scattering between the initial partons, as well as the low energy interactions of the resulting final state objects. Since a wide range of energy scales is involved, the full simulation of a *pp* collision is factorised into different steps: the hard interaction is computed up to a fixed order in perturbation theory, whereas the softer scales are calculated using phenomenological models.

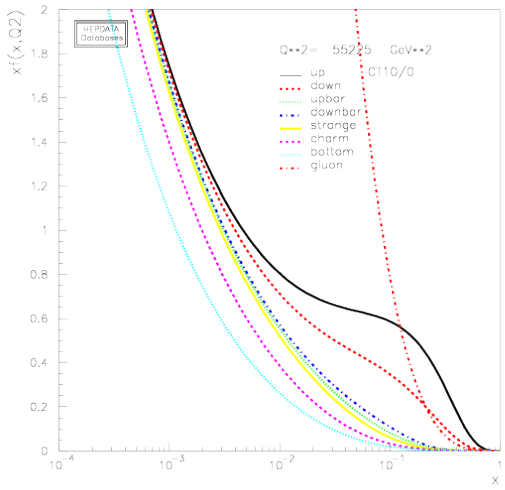


Figure 5.2: Dependency of the CT10 PDF set on the momentum fraction x for $Q^2 = {\mu_R}^2 = {\mu_\mathrm{F}}^2 = (m_t + m_H/2)^2 = 55225\,\mathrm{GeV}^2$ [209].

Figure 5.1 is a representative scheme of the factorised steps necessary for a full simulation of hadronic collisions. When there is sufficient momentum transfer, the partons can be considered free, which allows the treatment of hadron-hadron scattering as a single parton-parton interaction. Inside the colliding protons, the initial state of the partons (represented by the dark green arrows pointing to the shaded oval protons) can thus be modelled independently from the actual interaction by resorting to universal parton density functions (PDFs), as introduced in equation 2.38. The PDFs, $f_i(x_i, Q^2)$, encode the probability of finding a certain type of parton i within the proton carrying a fraction x_i of the proton's momentum, $\vec{p}_{parton} = x\,\vec{p}_{hadron}$, where x is referred to as the Björken scaling variable and Q^2 is the factorisation scale. As an example, in Figure 5.2 the PDFs used in this analysis, the CT10 set, are shown. These partonic distributions were obtained using lepton-proton collision data from the HERA collider and proton-antiproton collisions from the Tevatron [74, 75]. Here, the energy scale was set to $Q^2 = {\mu_\mathrm{F}}^2$ with $\mu_\mathrm{F} = \mu_R = m_t + m_H/2$ as used in the $t\bar{t}H$ analysis. For values of the Björken scale close to unity, the parton densities of the proton's valence quarks, the up and down quarks, dominate the momentum fraction. However, at smaller values of x, the gluon density is the largest one. The bottom, the charm, the anti-up, and the anti-down quarks, all originating from the virtual hadronic sea, show lower densities in the proton.

When a collision happens, two of these partons undergo an interaction with a significant momentum transfer, also known as a hard-scatter. This process can be explicitly computed at fixed order in perturbation theory, due to the high energy scale involved. The hard scatter, represented in red in Figure 5.1, is obtained through the calculation of the scattering matrix between the initial and final state particles of the process, *i.e.* the Matrix Element (ME) calculation.

QCD Bremsstrahlung radiation (in blue) will arise as a result of the color charge of the partons. Partons will successively emit gluons, that either split into quark-antiquark pairs or radiate more gluons, leading to the creation of patron showers. The MC simulation of such showers, both for initial and final state partons, gives a more complete description of the collision, since it accounts for the remainder of higher order contributions. It is, however, an approximation, as the parton emissions are considered independent. The continuous production of this radiation ends when partons reach the $Q \approx 1\,\text{GeV}$ energy scale. At this scale, the QCD partons (represented in green) recombine into colorless hadrons. The hadronisation process is simulated using phenomenological models, as is the simulation of the hadron decay into the final state particles observed by the detector.

The underlying event (UE) (represented in purple) constitutes the secondary soft interactions of spectator partons from the colliding protons. Since these secondary interactions are low energy scale processes, with large cross-section values, they are simulated with multiple interaction models. The multiple parton interactions (MPI) are determined using phenomenological models with parameters tuned using experimental data. Similarly to the UE, pile-up events are also modelled using phenomenological models and tuned using observed data.

The different physics processes are then simulated by either a full simulation chain (multi-purpose generators) or by an interface between ME generators with additional parton shower simulation. The generator's output includes information on the final state of stable particles, after the decay and hadronisation. The samples at the particle level are then passed through the full ATLAS detector simulation [210], based on GEANT4 [211], which models the interaction of the particles with the detector and subsequent reconstruction, yielding output samples at the reconstruction level, just like observed data. During the detector simulation, the energy deposits originated from the interaction are converted into electronic signals according to the ATLAS geometry, materials and readout system. For faster computation

there is also a less exhaustive simulation, AtlasFast-II [212], which applies a parameterised description of the particle showers in the calorimeters.

The MC samples, including the ATLAS simulation, will allow to estimate the experimental sensitivity for this luminosity and the associated systematic uncertainties. Simulations of the relevant physics processes, with accurate description of the hard interactions, the underlying physics and detector performance are thus chosen according to the signal topology. In order to reproduce the best theoretical cross-section estimates (NLO or NNLO), small corrections are then applied to re-scale object identification efficiencies, energy scales, and the pile-up based on control regions from data.

In this analysis, the masses of the top quark and the Higgs boson are set to 172.5 GeV and 125 GeV respectively, in all MC samples. PHOTOS 2.15 [213] is used to simulate photon radiation and TAUOLA 1.20 [214] to simulate the τ decays. The pile-up interactions are simulated as minimum bias events and superimposed on the simulated MC events, matching the luminosity profile of recorded data. The Pythia 8.1 generator [215] is used to simulate the pile-up events with the MSTW2008 LO PDF set and the AUET2 [216] tune. Contributions from interactions within the same bunch crossing and between neighbouring bunch crossings are considered. All MC samples are processed through the same reconstruction software as data, and corrections are applied to match the object identification efficiencies, energy scales and energy resolutions determined in data control samples.

5.2.1 $t\bar{t}H$ Signal Simulation

In order to access the acceptance and efficiency of event selection, as well as the overall performance of the analysis, a detailed prediction of the $t\bar{t}H$ signal is necessary. Furthermore, the compatibility test of a possible data excess is only feasible if using accurate MC simulations.

The $t\bar{t}H$ process is modelled with NLO matrix elements from HELAC-Oneloop [217] and interfaced with POWHEG-BOX for parton shower simulation [218, 219, 220]. These MC samples are commonly referred to as POWHEL samples. The PDF set used in this production is the CT10nlo and the factorisation (μ_F) and renormalisation (μ_R) scales are set to $\mu_F = \mu_R = m_t + m_H/2$. Showering was performed with PYTHIA 8.1 using the CTEQ6L1 PDF set [221] and the AU2 UE tune [222].

In the generation of the $t\bar{t}H$ samples, all Higgs boson decays are included. The signal sample was normalised using the NLO cross-section [129, 130, 133] and the Higgs decay

branching ratios [223, 224, 225, 226] from reference [227]. In this analysis, all combinations of top quark-antiquark pair decays are considered as signal: all hadronic ($t\bar{t}H \rightarrow$ allHad+H), semi-leptonic ($t\bar{t}H \rightarrow l$jets + H) and dileptonic($t\bar{t}H \rightarrow ll + H$). The signal MC samples are summarised in Table5.1.

Process	Generator	σ (fb)	ATLAS Simulation
$t\bar{t}H \rightarrow$ allHad + H	PowHel+Pythia8	59.09	Full Simulation
$t\bar{t}H \rightarrow l$jets + H	PowHel+Pythia8	56.63	Full Simulation
$t\bar{t}H \rightarrow ll + H$	PowHel+Pythia8	13.58	Full Simulation

Table 5.1: Monte Carlo samples used for signal description.

5.2.2 Background Simulation

Table 5.2 summarises the sample information for all MC backgrounds used in this analysis. The multijet background is not included, since it was estimated purely using data-driven techniques.

Process	Generator	ATLAS Simulation
$t\bar{t}$ + jets	Powheg+Pythia6.425	AtlasFast
Z + jets	Alpgen+Pythia6.425	Full Simulation
W + jets	Alpgen+Pythia6.425	Full Simulation
WW, WZ, ZZ	Alpgen+Herwig+Jimmy	Full Simulation
Single t	Powheg+Pythia6.425	Full Simulation
Single tZ	Madgraph5+Pythia8	Full Simulation
$t\bar{t}Z$ + jets	Madgraph5+Pythia6.425	Full Simulation
$t\bar{t}W$ + jets	Madgraph5+Pythia6.425	Full Simulation
$t\bar{t}WW$	Madgraph5+Pythia8	AtlasFast

Table 5.2: Monte Carlo samples used for background description. The multijet background is not included, since it was purely estimated using data-driven techniques.

5.2.2.1 $t\bar{t}$ + jets background

The dileptonic $t\bar{t}$ + jets process is generated using the Powheg-Box 3.0 NLO generator with the CT10 PDF set [74]. It is interfaced to Pythia 6.425 [201] with the CTEQ6L1 PDF set [221] and the Perugia2011C underlying-event tune [228]. The top++2.0 [70] theoretical calculation ($\sigma_{t\bar{t}+jets} = 26.55\,\text{pb}$) is used to normalise the sample to the next-to-

next-to-leading order (NNLO) in QCD, including resummation of next-to-next-to-leading logarithmic (NNLL) soft gluon terms [65, 66, 67, 68, 69].

Due to its final state topology, the $t\bar{t}+$jets production is the expected dominant background contribution in this analysis. Furthermore, the cross-section of $t\bar{t}$ processes is several orders of magnitude larger than the predicted SM $t\bar{t}H$ associated production cross-section. Accurate description of the different $t\bar{t}$ topologies, especially the emission of additional jets and the heavy-flavour component, is therefore of major importance. Particular attention to both the normalisation and modelling of final state objects' properties is considered in this study. Here, the $t\bar{t}+$jets samples are generated inclusively, but events are categorised in orthogonal sets depending on the flavour of the additional jets: $t\bar{t}+$ light jets, $t\bar{t}+c\bar{c}$ jets and $t\bar{t}+b\bar{b}$ jets. The classification is based on an algorithm matching hadrons to particle jets built from stable particles. The partons, which do not originate from the decay of the $t\bar{t}$ pair, are matched to particle jets, using a $\Delta R < 0.4$ requirement. Particle jets are reconstructed by clustering stable particles, excluding muons and neutrinos, using the anti-k_t algorithm with a radius parameter $R = 0.4$. The matching algorithm considers all b and c hadrons with $p_{\mathrm{T}} > 5\,\mathrm{GeV}$ which do not originate from the $t\bar{t}$ decay. Particle jets with $p_{\mathrm{T}} > 15\,\mathrm{GeV}$ and $|\eta| < 2.5$ are matched to the set of heavy flavour hadrons, and the particle jet is labeled as a b or c-jet, as long as the $\Delta R < 0.4$ requirement is satisfied. The 15 GeV p_{T} threshold for particle jets is chosen to allow for resolution effects in the reconstructed jets. Events are labeled $t\bar{t}+b\bar{b}$ jets, if they contain at least one b-jet not originating from the top pair decay. $t\bar{t}+c\bar{c}$ jets events are those failing the previous criteria and containing at least one c-jet not from a W decay. In this thesis, the set of $t\bar{t}+c\bar{c}$ jets and $t\bar{t}+b\bar{b}$ jets events is referred as heavy flavor (HF) events ($t\bar{t}+\mathrm{HF}$). $t\bar{t}+$light jets events are the remainder events which fail all above requirements, including those without additional jets.

The large number of $t\bar{t}$ events produced at the LHC has allowed for detailed studies of the top quark pair production. Differential cross-section measurements are available at $\sqrt{s} = 7$ TeV for several kinematic variables. In the phase space of this analysis, both at 7 TeV and 8 TeV, the MC prediction of most generators over-predicts the energy spectrum of data, leading to visible differences, not accounted by the statistical and systematic uncertainties of the measurements. Several studies were developed to correct for this effect. The analysis described in this thesis differs from that of Ref.[1] in the way the modelling of the $t\bar{t}$ background processes are derived. This is described in detail in Chapter 6.2.2.

5.2.2.2 Z + jets background

The background contribution from the Z + jets production is generated using the ALPGEN 2.14 [229] leading order generator with the CTEQ6L1 PDFs [221]. PYTHIA 6.425 is used to model both the parton showers and the fragmentation scheme. The Z + jets background is generated with up to five additional partons, separated according to the flavour of the partons: Z + light jets, $Z + c\bar{c}$ jets and $Z + b\bar{b}$ jets. The Z + jets samples are normalised to the inclusive NNLO theoretical cross-section [230]. In the case of the Z + light jets samples, the overlap between $Z + Q\bar{Q}$ events (where $Q = b, c$) generated from the matrix element calculation and those from parton-shower evolution is removed by an algorithm based on the angular separation between the extra heavy quarks: if $\Delta R(Q, \bar{Q}) > 0.4$, the matrix element prediction is used, otherwise the parton shower prediction is used.

In previous Z + jets studies [231], the comparison of data and ALPGEN simulation has shown harder jet and lepton transverse momentum predictions in MC than in real collision data. Unfortunately, these differences can not be accounted for with the statistical and systematic uncertainties alone, and lead to visible disagreement between data and MC in the higher energy/momentum regions of most physics objects, be it a simple lepton or a more complex Z boson. Making use of orthogonal control regions, a data driven method is implemented. The MC is reweighed to not only correct for this transverse momentum effect, but also to give a more precise estimate of the Z + jets normalisation. This method is discussed in Section 6.2.1.

5.2.2.3 W + jets background

The W +jets background is generated using the CTEQ6L1 PDF set in ALPGEN 2.14, just as the Z + jets production. Similarly, PYTHIA 6.425 is used to model the parton showers and the fragmentation. The W+jets samples are also generated with up to five extra partons, and separately according to the additional parton flavour (W +light jets, $W + c\bar{c}$ jets, $W + b\bar{b}$ jets and W+c jet). The inclusive NNLO theoretical cross-section is used to normalise the W+jets samples, and the same angular separation algorithm ($\Delta R(Q, \bar{Q}) > 0.4$)employed for Z +jets is implemented to remove the overlap of ME generated events and those from PS evolution.

5.2.2.4 Diboson background

The diboson production (WW, WZ, ZZ) in association with up to three jets is generated using ALPGEN 2.14. HERWIG 6.520 is used to model the parton showers and the fragmenta-

tion of these samples [232]. The diboson samples are also interfaced with JIMMY 4.31 [233] to simulate the underlying event. The NLO theoretical cross-sections are used to normalise the MC samples to the highest order calculations available [234].

5.2.2.5 Single t background

The single top quark production is simulated independently considering the s-channel, t-channel and Wt production mechanisms. The samples are generated with POWHEG-BOX 3.0 using the CT10 PDF set and interfaced to PYTHIA 6.425 with the CTEQ6L1 set of PDFs and Perugia2011C underlying-event tune. In order to remove the overlap of Wt production with the $t\bar{t}$ final state starting at NLO, the diagram removal scheme [235] is used. The single top quark samples are normalised to the approximate NNLO theoretical cross sections [77, 78, 81] using the MSTW2008 NNLO PDF set [72, 73].

The single top production in association with a Z boson is also considered for completeness. Simulated with MADGRAPH5 [236] and interfaced with PYTHIA8.1 [215] for showering, this small background was normalised to its LO theoretical prediction.

5.2.2.6 $t\bar{t}V$ background

Contributions from the associated production of a vector boson V (where $V = Z$, W) and a $t\bar{t}$ pair are generated using MADGRAPH5 and the CTEQ6L1 PDF set. The parton shower is simulated using PYTHIA 6.425 with the AUET2B tune [237]. The $t\bar{t}V$ samples are normalised to the NLO cross-section predictions [238, 239]. Additionally, the production of $t\bar{t}WW$ is also included. It is generated with MADGRAPH5 interfaced with PYTHIA8.1 for showering.

5.3 Trigger Requirements

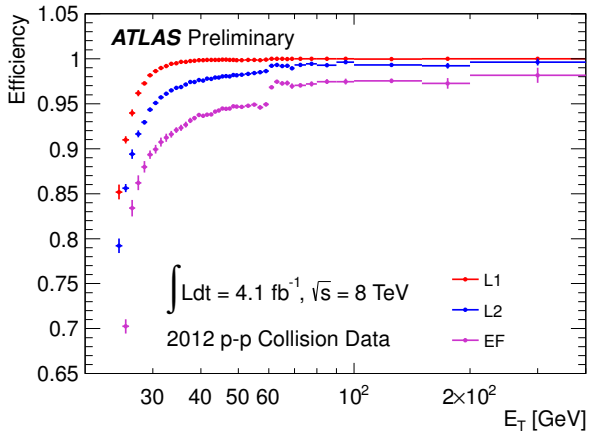
In this analysis, only pp collisions under stable beam conditions and for which all detector subsystems were operational are considered. Events are accepted using unsuppressed single-electron or single-muon triggers, with the lowest transverse momentum thresholds possible at the EF trigger stage. In order to maximise the overall efficiency, the different p_{T} threshold triggers are combined in a logical OR. For the single electron trigger either `EF_e24vhi_medium1` or `EF_e60_medium1` were used, with p_{T} thresholds of 24 and 60 GeV

respectively [240].[1] As for the single muon trigger p_{T} thresholds of 24 or 36 GeV were used on tightly isolated muons, corresponding to `EF_mu24i_tight` or `EF_mu36_tight` respectively [241]. Figure 5.3 shows the single lepton efficiencies measured for the logical OR.

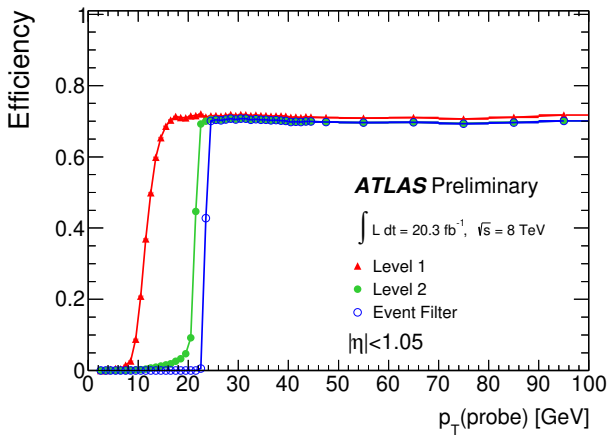
The triggers with the lower p_{T} threshold include isolation requirements on the lepton candidate. Although this result in inefficiency at high p_{T}, the efficiency is recovered by the triggers with the higher p_{T} threshold, where this condition is relaxed. For instance, regarding the electron candidates, the low p_{T} trigger only selects events with an electron passing the medium identification requirement and loose tracking isolation, whereas the trigger above 60 GeV dropped the isolation requirement and slightly loosened the identification. Figure 5.3 (a) shows the clear improvement in the electron trigger efficiency for higher transverse energy. Regarding the muon trigger, good inner detector track and matching hits in the muon spectrometer are required. Loose tracking isolation is required for the low p_{T} trigger, while for the 36 GeV trigger it is dropped. Furthermore, offline isolation requirements are tighter than the trigger included, which prevents the analysis from being affected by the isolation requirement applied at the trigger level. As can be seen in Figures 5.3 (b) and (c) the single muon triggers show an efficiency plateau ($p_{\mathrm{T}} > 25$ GeV) of approximately 70% in the barrel region ($\eta < 1.05$) and of approximately 86% in the endcap region ($\eta > 1.05$).

As discussed in Chapter 3.3.6, with the elevated number of proton collisions per bunch crossing, several interaction vertices can be reconstructed in a single the event. Each vertex can be determined by combining the reconstructed trajectories, or tracks, of the charged particles in the ID with a vertex fitting algorithm [242]. Since the profiles of the two proton beams overlap around the interaction point, the primary vertex algorithm can further constrain the selected tracks to lie within the estimated position of the beam spot. Trigger-selected events were thus required to have at least one well-defined primary vertex, with at least five associated tracks (with $p_{\mathrm{T}} > 500$ MeV) consistent with the beam collision region in the $x - y$ plane. In-time pile-up can then be assessed by the number of reconstructed primary vertices per event. The main vertex of the event, corresponding to the hardest pp collision, is assumed to be the one with highest squared sum of p_{T} of tracks, while the remaining vertices are considered pile-up interactions. Displaced vertices, *i.e.* those which are not contained within the beam collision region, are classified as secondary vertices.

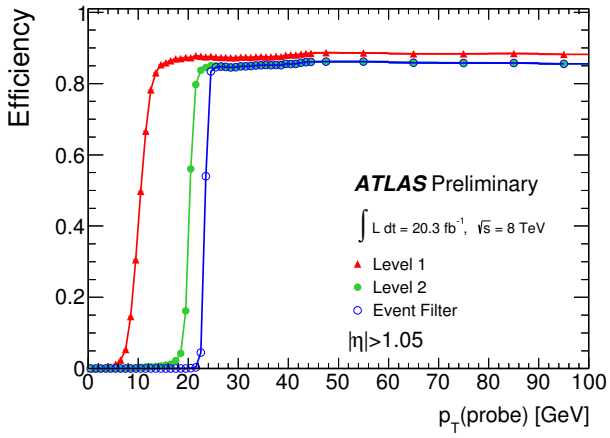
[1] The nomenclature used on trigger labels was previously explained in 3.3.5. Here, `e24vhi` means a 95% efficiency of EF triggering an event with an isolated electron with a $p_{\mathrm{T}} > 24$ GeV.



(a)



(b)



(c)

Figure 5.3: Trigger efficiencies at L1, L2 and EF for (a) the single electron triggers (`EF_e24vhi_medium1` OR `EF_e60_medium1`), (b) the barrel and (c) the endcap single muon triggers (`EF_mu24i_tight` or `EF_mu36_tight`) in the 2012 dataset [240, 241].

5.4 Object Reconstruction

In this search, the relevant physics objects to consider are electrons, muons, jets, b-jets and the magnitude and direction of the missing transverse momentum. The reconstruction and identification of such objects was performed as follows.

5.4.1 Leptons

The only relevant charged leptons for this analysis are the electrons and muons, since tau-leptons are not explicitly used. Whether a τ decays leptonically or hadronically and even if no τ reconstruction is applied, the decay products still contribute to the object reconstruction: either as isolated electrons or muons, or as low multiplicity jets.

5.4.1.1 Electrons

The identification and reconstruction of electrons is performed the ID and the electromagnetic calorimeter As a charged particle, the electron will cross the ID trackers emitting transition radiation. Since the lighter electrons have a larger characteristic Lorentz factor than charged hadrons, a higher emission of photons occurs. This in turn leads to more ionization charge being produced in the TRT (higher TRT hits above a certain threshold), which can be used as a discrimination variable for electron identification. The electron leaving the ID will then enter the electromagnetic calorimeter, where it will deposit most of its energy. The high granularity of this calorimeter, both in η and in ϕ, enables the precise determination of the electron's impact point, which can be associated to a ID track.

Electron candidates are therefore selected by searching for narrow and well defined energy deposits, or *clusters*, in the electromagnetic calorimeter, which are matched to a reconstructed track in the inner detector [243]. A cluster is defined as a group of cells within a fixed-size window around a local maxima of transverse energy (cluster seed). Only energy deposits with a total transverse energy above a threshold of 2.5 GeV are considered for cluster reconstruction. For electron cluster identification, a sliding-window algorithm [244], that searches for local maxima of energy within a window of dimensions 3×5 in units of 0.025×0.025 in $\Delta\eta \times \Delta\phi$ space, is used.

The matching of tracks and clusters is performed by extrapolating the track from the ID to the middle layer of the ECal and matched to the cluster seed. For the matching to occur, the $|\Delta\eta|$ between the cluster and the track must be inferior to 0.05 and the condition

$0.05 < q \cdot \Delta\phi < 0.10$ must be satisfied. Note that this $\Delta\phi$ condition is sign (q) dependent and thus takes into account the bending direction of the electrons in the inner magnetic field. Matched clusters are then reconstructed with slightly larger window: 3×7 in the barrel or 5×5 in the end-caps. If no matched track is found, the cluster is classified as an unconverted photon candidate, which are not relevant for this analysis.

The electron four-momentum is then computed with the final cluster energy and the (η, ϕ) spatial coordinate information from the associated track at the interaction vertex. The final cluster energy is obtained by correcting for the energy losses in the material in front of the calorimeter, and for energy leakage due to the fixed cluster size. Such corrections are obtained from dedicated studies using MC simulation, test beam data and $Z \to e^+e^-$ events [245].

The electron reconstruction algorithm can identify the isolated leptons as desired. However, not all resulting output objects are prompt electrons (*signal electrons*). Contributions from mis-identified hadrons, non-isolated electrons from semileptonic decays of heavy-flavour particles, or electrons from photon conversion can also be included (*background electrons*). Therefore, electrons must be identified with increasing degrees of purity, in order to suppress mis-identification while maintaining the high efficiency of signal electron detection. Several methods of *background electron* rejection are available. For instance, by requiring tracks to have hits within the B-layer, electrons from photon conversions can be rejected. Further track quality criteria can be useful to reduce mis-matching of tracks with photons, or electrically neutral charged mesons, such as π^0 or ρ. Hadronic jets can be discarded by requiring different cluster shapes and leakage conditions, since electrons produce narrower showers than hadrons and the hadronic leakage is smaller.

For the present search, a multivariate analysis (MVA) technique was used to improve the electron selection. A discriminant variable, $d_{\mathcal{L}}$, combining the likelihoods of a given object to be identified as signal or background, $\mathcal{L}_{\mathrm{s}}$ and $\mathcal{L}_{\mathrm{b}}$ respectfully, was defined as:

$$d_{\mathcal{L}} = \frac{\mathcal{L}_{\mathrm{s}}}{\mathcal{L}_{\mathrm{s}} + \mathcal{L}_{\mathrm{b}}} \,. \tag{5.1}$$

Here, the likelihood $\mathcal{L}_{\mathrm{s}}$ ($\mathcal{L}_{\mathrm{b}}$) is obtained from the signal (background) probability density function $P_{\mathrm{s},i}(x_i)$ ($P_{\mathrm{b},i}(x_i)$):

$$\mathcal{L}_{\mathrm{s}}(\overrightarrow{x}) = \prod_{i=1}^{n} P_{\mathrm{s},i}(x_i) \,, \tag{5.2}$$

where $\overrightarrow{x}$ is a vector of variable values describing different properties of the electron candidate. These probability density functions are obtained from data, using a set of discriminating distributions: variables describing the properties of the tracks in the inner detector, the longitudinal and transverse shapes of the electromagnetic showers in the calorimeters, and the matching between tracks and energy clusters [246]. Finally, there are three electron likelihood (LH) selections available, with similar electron efficiency as the equivalent cut-based selection, but with increasing background rejection: loose, medium and tight. Each LH selection is performed by cutting on a LH discriminant obtained from a different set of variables. In this analysis, the tight LH selection was used, which provided a very low rate of misidentified leptons. The longitudinal impact parameter z_0 of the electron track with respect to the primary vertex is also required to be inferior to 2 mm.

An additional isolation requirement is applied to further reduce the selection of non-prompt electrons, Based only on tracker information, this requirement suppresses the selection of electrons originated from hadron decays produced in jets. The isolation is guaranteed by demanding that within $\Delta R = 0.3$ of the electron track, the sum of the transverse momentum of the tracks with $p_{\mathrm{T}} > 0.4\,\mathrm{GeV}$ is not larger than 12% of the electron transverse momentum: $p_{\mathrm{T}}^{cone30}/p_{\mathrm{T}} \leq 0.12$. Only good quality tracks (in terms of number of silicon hits) originating from the primary vertex associated to the electron track are considered in the sum inside the cone. After this LH optimised selection, the electron efficiency is improved by roughly 7% per electron.

As final requirements, the electrons must be contained within the barrel or end-cap calorimeters ($|\eta_{\mathrm{cluster}}| < 2.47$) and outside the transition region between them ($1.37 < |\eta_{\mathrm{cluster}}| < 1.52$).

Reconstruction, Identification and Isolation Efficiencies

The efficiency of the reconstruction, identification and isolation selection is measured using the *tag-and-probe* method in unbiased samples of electrons, such as $Z \to e^+e^-$ and $J/\Psi \to e^+e^-$ events [246]. This method selects a clean sample of electrons from Z and J/Ψ decays, using strict selection criteria and full identification on one of the leptons from the decay, denoted *tag*. Taking into account the kinematic properties of the event, specifically the di-electron invariant mass near the Z or J/Ψ pole mass, there is reasonable confidence that the second object is an electron. The second electron candidate (called *probe*) is used for

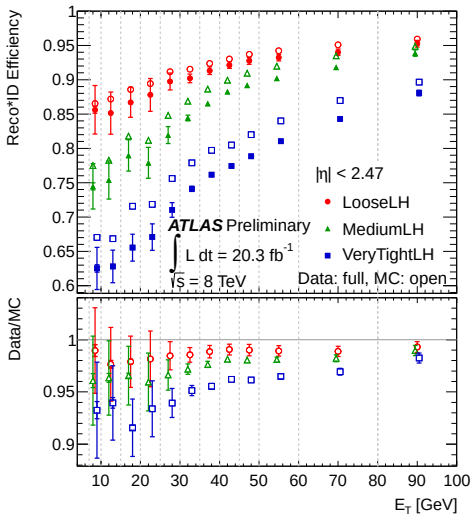


Figure 5.4: Measured combined reconstruction and identification efficiency as a function of E_{T} for the three LH selections available, compared to MC expectation for electrons from $Z \to e^+e^-$ decay [246]. The lower distribution shows the data-to-MC efficiency ratios. The data efficiency is derived from the measured data-to-MC efficiency ratios and the MC prediction for electrons from $Z \to ee$ decays. The uncertainties are statistical (inner error bars) and statistical+systematic (outer error bars).

the efficiency measurement. The electron selection is then applied to the unidentified and unbiased *probes*, allowing the efficiency measurement:

$$\epsilon_{\mathrm{total}} = \epsilon_{\mathrm{reco}} \times \epsilon_{\mathrm{ID}} \times \epsilon_{\mathrm{trigger}} \times \epsilon_{\mathrm{additional}} \,. \tag{5.3}$$

Here ϵ_{reco} is the reconstruction efficiency of clusters in the ECal, which is measured with the high-statistics electrons in the *probe* sample originated from Z boson decays. The identification efficiency, ϵ_{ID}, is determined with respect to reconstructed electrons and is measured in high-purity events: $Z \to e^+e^-$ events with $E_{\mathrm{T}} > 10\,\mathrm{GeV}$ and $J/\Psi \to e^+e^-$ events with $7\,\mathrm{GeV} < \mathrm{E_T} < 20\,\mathrm{GeV}$. The trigger efficiency, $\epsilon_{\mathrm{trigger}}$, is measured in Z boson decay events with respect to the medium++ electron identification (equivalent working point to the one used at trigger level, `medium1`). Finally, the $\epsilon_{\mathrm{additional}}$ is related to additional selection criteria as the lepton isolation.

Unfortunately, contamination from fake electrons make the *tag-and-probe* method rather difficult. Background template shapes or combined functional fits of background and signal are used to model data and estimate the fake lepton contamination. This allows the removal of fake electron contamination. The final efficiencies are calculated with the independently

estimated number of electrons at probe level and those passing the tested criteria. Figure 5.4 shows the combined reconstruction and identification efficiency as a function of E_{T} for the three LH selections available.

The electron efficiencies have slightly different behaviour for simulation (ϵ_{MC}) and data (ϵ_{data}), which needs to be accounted for reliable physics results. Calibration scale factors are thus derived from data and applied on the MC samples to obtain data-like efficiencies:

$$SF = \frac{\epsilon_{\mathrm{data}}}{\epsilon_{\mathrm{MC}}} \,. \tag{5.4}$$

These scale factors are usually close to unity, as can be seen in the lower distributions of Figure 5.4. The deviations mostly occur when mis-modelling of either track properties or shower shapes are present. The electron efficiencies are dependent on the transverse energy and pseudorapidity, which is reflected on the SF corrections as well. The combined uncertainties on the reconstruction, identification and isolation requirement SF are $\approx 2\%$. An additional uncertainty of 2% is added in quadrature to the systematic uncertainties, due to the extrapolation of electron isolation efficiencies from the Z boson environment to the t quark related one [247].

Electron Energy Scale and Resolution

The absolute energy scale of electrons was also measured using $Z \to e^+e^-$ and $J/\Psi \to e^+e^-$ events and has a total uncertainty at the sub-percent level [245]. By fitting the di-electron invariant mass distribution of both resonances, correction factors were calculated as a function of the electron pseudorapidity. In the central η region, the total uncertainty on the electron calibration is less than 1%, increasing up to a few percent in the most forward regions of the calorimeter. These are dominated by uncertainties from the detector material and the presampler energy scale, but also include the event selection, pile-up, and hardware modelling .

For the electron energy resolution, the main probe is obtained via the Z resonance width in both data and simulation. In the calorimeter the electron energy resolution can be parametrised as,

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c \,, \tag{5.5}$$

with a, b and c dependent on pseudorapidity. The first term, a, is the sampling term which is well described in simulation within 10%. This term dominates the energy resolution for low energy electrons and is computed from the di-electron invariant mass in the J/Ψ resonance.

The second term, b, is the noise term. Finally, c is the resolution constant term, typically 1% larger in data than in simulation, with the exception of the regions $1.37 < |\eta| < 1.82$ where 3% is observed. Since the resolution in data is slightly worse than that obtained in the simulation, appropriate corrections are derived and applied to the simulation reconstructed electrons to match the energy smearing in data.

5.4.1.2 Muons

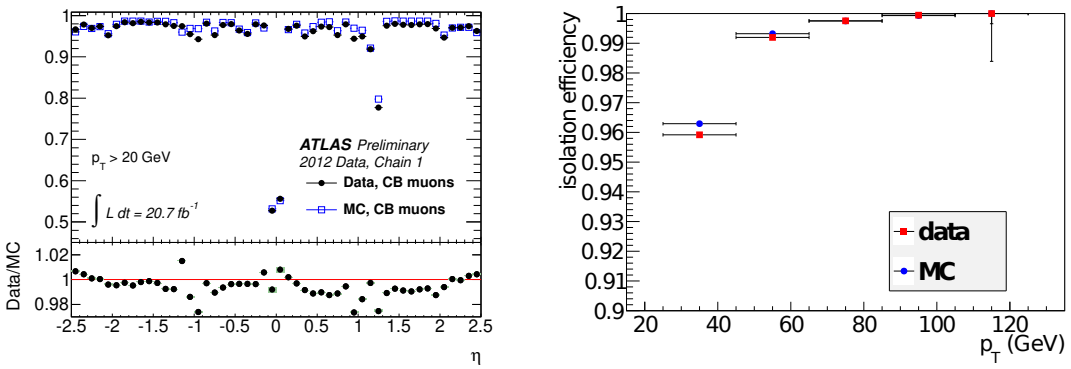
The identification and reconstruction of muons relies on combining information from several ATLAS sub-detectors, such as the inner detector and the muon spectrometer. As minimum ionizing particles, muons will usually traverse the ID and the calorimeters without significant energy loss. There are several muon reconstruction methods available at ATLAS.

In this analysis, a muon is defined by the *MuId* algorithm [248] for *combined muons* which uses the independent reconstruction of tracks in the ID and in the MS. Both the muon transverse momentum and charge sign can be measured by matching the track segments in the MS layers with the reconstructed tracks found in the ID, covering the total $|\eta| < 2.7$ region. Taking into account the muon energy loss in the calorimeter, a combined track can be formed after re-fitting the hits of both tracks. This algorithm provides the highest muon purity by requiring good quality combined fits of ID hits and MS segments.

In order to reduce the mis-identification rate, further selection criteria are applied to the muon candidates. For instance, the final muon candidates are required to be within $|\eta| < 2.5$, where the ID coverage is available. The muon tracks are required to pass a minimal number of hits in the Pixel, SCT and TRT sub-detectors, as well as a hit in the innermost pixel layer when the track crosses an active module. Muons must also originate from the primary vertex, with a longitudinal impact parameter relative to the primary vertex, z_0, smaller than 2 mm.

Similarly to the electrons, muons are also required to be isolated from additional tracking or calorimeter energy. The overlap cone $\Delta R_{(\mu,j)}$ between a muon and any jet with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|JVF| > 0.5$ must be larger than 0.4 to guarantee muon isolation.[1] Additionally, an isolation requirement is applied to distinguish prompt muons arising from the hard interaction and those originating from decay chains of b/c-hadrons or kaons. The mini-isolation variable, I^{μ}_{mini}, is defined as the sum of the transverse momentum of all the tracks verifying $\Delta R_{(\mu,track)} < 10\,\mathrm{GeV}/\mathrm{p}^{\mu}_{\mathrm{T}}$, with p^{μ}_{T} the transverse momentum of the muon. The

[1] JVF is the jet vertex fraction and will be defined in the following section.



(a) (b)

Figure 5.5: Data and MC comparison of muon efficiencies: (a) reconstruction efficiency for combined muons as function of η [249]; (b) isolation efficiency as a function of p_{T} [247].

mini-isolation requirement selects events with I^{μ}_{mini} smaller than 5% of the muon p^{μ}_{T}. As the muon transverse momentum increases, the mini-isolation cut is relaxed and the size of the considered cone is smaller. This leads to an isolation cut less susceptible to pile-up effects and more efficient for boosted configurations, when a real muon is in the vicinity of a jet.

Muon Reconstruction, Identification and Isolation Efficiencies

Just as for the electrons, the reconstruction, identification and isolation efficiencies have been measured with the *tag-and-probe* method. High-statistics and high-purity $Z \to \mu^+\mu^-$ and $J/\Psi \to \mu^+\mu^-$ events are used to determine the efficiencies in data.

The muon reconstruction efficiency for the complete $\sqrt{s} = 8$ TeV dataset can be seen in Figure 5.5 (a). Slight discrepancies are observed between data and MC, which are accounted for by calculating correction scale factors to improve the MC predictions. These data/MC scale factors are also included in the figure and have uncertainties on the percent level. The largest data-to-MC discrepancy is in the $1.5 < \eta < 2$ region, where two faulty pixel B-layer modules caused a lower reconstruction efficiency than that predicted by the simulation.

The efficiency of the isolation requirement is measured by determining if the unbiased *probe* muon also passes the isolation requirement. Figure 5.5 (b) shows the measured efficiencies in data and simulation. Here, the MC samples include $Z \to \mu^+\mu^-$ and $Z \to \tau^+\tau^-$ events, as well as $b\bar{b}$ and $c\bar{c}$ processes, where at least one of the heavy quarks is forced to decay into a muon. Background samples ($\sim 0.1\%$) are added to the simulation, and not

subtracted from data as was done for electrons. Even so, the net effect on the scale factors should be similar. The level of agreement between data and simulation is within 0.5%, and stable throughout several kinematic quantities.

Muon Momentum Scale and Resolution

The muon momentum scale and resolution are also determined from pure muon samples, such as those originated in $Z \to \mu^+\mu^-$ and $J/\Psi \to \mu^+\mu^-$ events. Due to the large number of events collected, both measurements can be obtained by performing a fit to the di-muon invariant mass distributions of the Z and J/Ψ resonances [248].

Correction factors for the muon momentum scales are evaluated separately for the ID and MS regions, allowing a direct understanding of the sources of discrepancies between data and MC simulation. These corrections are then propagated to the combined momentum reconstruction using a weighted average. The weight is derived individually for each muon from the combined transverse momentum, which is assumed as a linear combination of the $p_{\mathrm{T}}^{\mathrm{ID}}$ and $p_{\mathrm{T}}^{\mathrm{MS}}$. For combined muons in the range $5 \geq p_{\mathrm{T}} \geq 100\,\mathrm{GeV}$, the momentum scale is extremely well measured, with an uncertainty not larger than 0.2%. The correction factors and their relative uncertainties are used to accurately describe the data observations by scaling the muon p_{T}.

Additionally, the smearing of the momentum of muons in the MC is performed, to match the momentum resolution seen in data. The di-muon mass resolution is $2 - 3\%$, presenting slightly lower values for low-p_{T} central muons and slightly higher values at larger pseudorapidity or high transverse momentum. This smearing correction factor reflects the differences in the momentum resolution measurement between the two separate parametrisations used for the ID and MS. The resolution observed in data is reproduced by the corrected simulation within relative uncertainties of 3% to 10% depending on η and p_{T}.

5.4.2 Jets

Quarks and gluons originated from the hard interactions can not be found isolated due to the color confinement. As a result of fragmentation and hadronisation of the hard partons, collimated sprays of energetic hadrons, called jets, will traverse the experimental apparatus. In order to reconstruct a physics object with the same characteristics of the initial hard parton, a jet is defined as the ensemble of particles produced in the hadronisation.

Several jet reconstruction algorithms are available based on different types of detector-level input: from track only information to combined track and calorimeter records. Jets reconstructed from charged particle tracks in the ID, with origin in the primary vertex, are called *track jets*, which are insensitive to pile-up effects. *Calorimeter-tower jets* are built from the energy deposits in the hadronic and electromagnetic calorimeter cells. These jets present a resolution that reflects the static granularity of the calorimeter cells and towers. In ATLAS analyses, the most commonly used reconstruction algorithm (anti-k_t) uses topological clustering, whose output objects are called *topo-cluster jets* [196, 244]. Topo-cluster jets are also built from energy deposits in the calorimeter, but are reconstructed to form three-dimensional topological clusters following the shower development of a particle interacting with the calorimeter. Non-collision backgrounds can, however, affect the quality and purity of calorimeter signals, as described in the Chapter 4. Fortunately, such contaminations are already considered in the topo-cluster jet reconstruction by using the cell noise as threshold for identification of significant energy depositions in the cells. Remnant contaminations can then be removed through further jet quality criteria. In MC simulations *particle* (or *truth*) *jets* can also be reconstructed from truth stable particles.

In this study, the jets are reconstructed from calibrated topo-clusters built in the calorimeters. As was briefly described in Section 4.2, cells with significant signal-to-noise ratio ($s \geq 4\ \sigma$) in randomly triggered events are considered as seed cells. Here, the noise, σ, is estimated from the standard deviation of the cell energy gaussian-like shape, including the electronic noise of the current gain and the pile-up noise contribution. If the signal-to-noise ratio of neighbouring cells verify $s \geq 2\ \sigma$ then they are iteratively added to the developing cluster. Perimeter cells, with no noise threshold, are also added to the topological cluster to guarantee a complete shower reconstruction.

The topo-clusters calibration is performed using the local cluster weighting (LCW) method. The LCW classifies topo-clusters as electromagnetic or hadronic, according to the longitudinal shower depth and measured energy density. Energy corrections are then derived according to this classification using single charged and neutral pion MC simulations. This method improves the cluster resolution and reduces fluctuations due to the non-compensating nature of the ATLAS calorimeters [247, 250].

Jets can then be reconstructed from the calibrated topological clusters using a sequential recombination algorithm, the anti-k_t [251, 252]. This algorithm is a powerful jet-finding algorithm used to decide which inputs are aggregated into each individual jet. Considering

two particles, i and j, the grouping of these into a jet using the anti-k_t algorithm depends on the distance d_{ij} between the constituents:

$$d_{ij} = \frac{\Delta R_{ij}^2}{R^2} \min\Big(\frac{1}{k_{\mathrm{T\,i}}^2}, \frac{1}{k_{\mathrm{T\,j}}^2}\Big), \tag{5.6}$$

where k_{T} are the transverse momenta and $\Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$ is the square of the angular distance between particles i and j. R is a constant parameter of the jet algorithm, specifically optimised for the experiment and analysis, since it somewhat controls the expected size of the jets. In order to distinguish constituents coming from the hard-scatter interaction, the distance between a constituent and the beam, d_{iB}, is also defined:

$$d_{iB} = \frac{1}{k_{\mathrm{T\,i}}^2} \tag{5.7}$$

The sequential jet clustering algorithm identifies the smallest of the distances, thus grouping the most energetic particles first. If the minimum found is d_{ij}, then i and j are merged into a single particle, summing their four-momenta. Otherwise, if the minimum is a distance to the beam, d_{iB}, particle i is declared as a final jet and removed from the particle list. The procedure is then repeated with the remaining particles (including the recombined ones), until none are left.

Since the anti-k_t algorithm starts by combining the higher p_{T} constituents, as the algorithm progresses the lower p_{T} emissions with wider-angles are combined with the core jet, resulting is a circular jet in the (η, ϕ) plane. When two hard particles are separated by $\Delta R_{ij} < 2R$, i.e. when overlap of jets occur, the shape of the reconstructed jets loosely deviate from circular.

For the purpose of jet definition in this analysis the anti-k_t algorithm with $R = 0.4$ was used. The jet four momentum is thus defined as the four momentum sum of all its constituents.

5.4.2.1 Jet Calibration

Each reconstructed jet energy must be calibrated in order to correspond to the energy of the truth particle jet. As already mentioned, the topo-clusters are calibrated [253, 254] prior to the jet finding with the anti-k_t algorithm. These topo-cluster corrections account for the effects of dead material, non-compensation and out-of-cluster leakage.

After reconstruction, the jets are calibrated to the mean energy of stable particles inside the jets, using energy and pseudorapidity dependent calibrations factors derived from simulations [255]. The calibration scheme for LCW jets consists of four sequential steps:

Pile-up Correction The presence of simultaneous extra collisions can modify the measured jet energy. Derived from MC simulations, this correction accounts for the energy offset introduced by pile-up interactions. A jet-by-jet estimation and subsequent removal of the energy added by pile-up events is performed, depending on the number of primary vertices and on $< \mu >$ as a measure of in-time and out-of-time pile-up respectively [256].

Origin Correction Taking into consideration the location of the collision primary vertex, a correction is applied to the calorimeter jet direction, forcing it to point to the primary event vertex and not to the centre of the ATLAS detector. This correction improves the angular resolution, whilst the energy of the jet remains unchanged, resulting in a small improvement in the jet p_{T} response.

Energy Calibration Restores the reconstructed jet energy and η to the MC particle-level values of truth jets. It accounts for detector effects, such as the mis-measurement of the deposited energy, lost energy in inactive regions of the detector or unclustered energy deposits. Derived from MC, the isolated calorimeter jets are matched to isolated particle-level jets within $\Delta R = 0.3$.

In-situ Calibration Residual correction applied to jets reconstructed in data. Data-to-MC differences are evaluated using in-situ techniques based on the transverse momentum balance between a jet and well-measured photons, Z bosons or jets.

In the $t\bar{t}H$ process, jets are mostly originated from the decay of massive resonances. In order to reduce contamination from backgrounds with softer radiation, in this search, jets are required to be central ($|\eta| < 2.5$) and have high momentum ($p_{\mathrm{T}} > 25\,\mathrm{GeV}$).

During the jet reconstruction process, no discrimination is made on the source of the energy deposits. In order to distinguish between identified electrons and jet energy deposits, an isolation requirement is applied. If any jets are within $\Delta R < 0.2$ of a reconstructed electron, the single closest jet is discarded, thus avoiding double counting of electrons as

jets. An additional isolation cut is then applied, which removes electrons within $\Delta R < 0.4$ from a reconstructed jet. This further suppresses background from non-isolated electrons.

5.4.2.2 Jet Energy Scale

The jet energy scale (JES) is determined from data events using a combination of in-situ techniques, where the p_{T} of a jet is compared to a well-measured reference photon or a Z boson [255]:

$$\mathcal{R}(p_{\mathrm{T}}^{\mathrm{jet}}, \eta) = \frac{(p_{\mathrm{T}}^{\mathrm{jet}}/p_{\mathrm{T}}^{\mathrm{ref}})_{\mathrm{data}}}{(p_{\mathrm{T}}^{\mathrm{jet}}/p_{\mathrm{T}}^{\mathrm{ref}})_{\mathrm{mc}}} \tag{5.8}$$

The JES uncertainty was evaluated in $\sqrt{s} = 8$ TeV data and includes multiple sources of systematic uncertainties:

Statistics, Detector, Modelling and Mixed — Using in-situ techniques, uncertainties originated from numerous sources (statistics/method, detector, physics modelling and mixed detector and modelling) can be assessed for central η jets, while preserving the correlation among them across different p_{T} and η bins. The most relevant sources of uncertainty are identified and diagonalised into eigenvectors so that they can be combined in an uncorrelated way [255].

Pile-up — To correct for residual mis-modelling of simultaneous interactions in the MC, uncertainties are assigned to the pile-up correction of the jet calibration, which rapidly decrease for higher p_{T}.

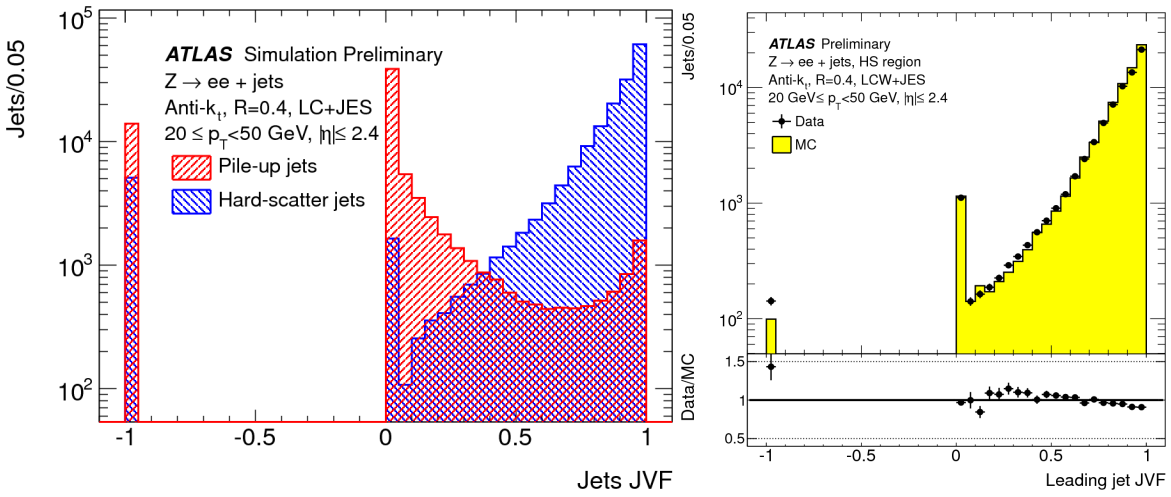
η-intercalibration — These uncertainties account for the calibration of jets in the forward η region relative to jets in the central region. Determined separately in a model component and a statistical one in di-jet events, the p_{T} of a well calibrated central jet and a jet in forward η is balanced and the detector response is made uniform throughout all η regions. These are the dominant source of JES in the most forward regions ($|\eta| > 3$).

High p_{T}	In-situ techniques for very high p_{T} jets ($p_{\mathrm{T}} > 2$ TeV) have limited statistics, thus the detector response is extrapolated from both MC studies and single-hadron test-beam results [257]. In order to perform this extrapolation, jets are considered as a superposition of single particles' energy deposits. The calorimeter response in the combined test-beam are then extrapolated to high p_{T} jets and the systematic uncertainties are evaluated.
Flavour Response	These uncertainties take into consideration that in-situ techniques are mainly sensitive to jets arising from quark sources rather than gluon ones. Due to differences in the fragmentation and showering properties of the jet origin, the calorimeter response also varies according to the jet source. MC estimates of the response difference between quark and gluon sources are evaluated and the baseline uncertainty increased [258].
***b* quark Response**	Jets arising from b quarks have a specific detector response that needs to be considered. Using an inclusive jet sample and a sample enriched in jets arising from b quarks, the jet calibration is compared to an estimate of jet p_{T} obtained with track jets, therefore determining the uncertainty to which the jet energy scale can vary [259].

5.4.2.3 Jet Energy Resolution

The jet energy resolution was measured using the bisector in-situ technique [260] in $\sqrt{s} =$ 8 TeV dijet events. The sum of the transverse momenta of the two leading jets $\overrightarrow{P}_{\mathrm{T},1}$ and $\overrightarrow{P}_{\mathrm{T},2}$ defines the transverse balance vector, $\overrightarrow{P}_{\mathrm{T}}$. By projecting $\overrightarrow{P}_{\mathrm{T}}$ on two orthogonal directions (Ψ, η) in the transverse plane (where η was chosen in the direction that bisects $\Delta\phi_{12}$ the azimuthal angle between $\overrightarrow{P}_{\mathrm{T},1}$ and $\overrightarrow{P}_{\mathrm{T},2}$), the balance of the dijet event can be evaluated.

A perfectly balanced dijet event would yield $\overrightarrow{P}_{\mathrm{T}} = 0$, however, several sources can originate significant fluctuations around this value and thus give rise to non-zero variance of both η and Ψ components. The jet energy resolution is then determined by comparing data and MC simulation and taking advantage of the different sensitivities of each projection to the underlying physics of the dijet system. The MC prediction for the energy resolution



(a) (b)

Figure 5.6: Distributions of the jet vertex fraction [256]: (a) using simulated Z+jets events, with the hard-scatter jets in blue and the pile-up jets in red[1]; (b) using data and MC samples enriched in hard-scatter jets, *i.e.* where a jet is balanced against a Z boson.

reasonably describes the observed values in data, apart from particular regions of the phase space, such as high p_{T} or high η, where the resolution in data has been found to be larger than expected. Instead of directly correcting the nominal measurement, this effect was considered as a source of systematic uncertainty, has will be described later on.

5.4.2.4 Jet Reconstruction Efficiency

The jet reconstruction efficiency of calorimeter jets was derived using a tag-and-probe technique, where the reconstruction efficiency is compared to that of track-jets [261, 262]. Defined as the fraction of probe track-jets matched to a calorimeter jet, the reconstruction efficiency was found to be approximately 0.2% lower in MC than in data, for jets in the $p_{\mathrm{T}} < 30\,\mathrm{GeV}$. Similarly to the energy resolution, this is considered as a source of systematic uncertainty.

5.4.2.5 Jet Selection: Cleaning and Jet Vertex Fraction

In order to identify and reject *bad jets*, such as those arising from transient hardware problems in the calorimeter, LHC beam-gas interactions, or cosmic-ray induced showers, a set of jet quality criteria were required. Jets with in-time real energy deposits are discriminated from fake or background jets using information on the shape of the electrical signal in every

calorimeter cell (jet quality factor), the fraction of jet energy that is deposited in each sub-detector of the calorimeter system, and even the fraction of jet energy deposited per layer of the calorimeter. In this analysis, the criteria used to reduce contamination from bad jets was the looser jet quality selection [247]. This working point has a jet selection efficiency above 99.8% with MC simulation closely describing data.

Another important source of background jets is pile-up. Simultaneous collisions can originate jets which should not be included in the hard-scatter event. To reduce the effect of in-time pile-up, track information associated to each jet is used. The vertex to which a jet originates from is identified using the jet vertex fraction (JVF) [256]. This quantity is defined as the ratio of the scalar sum of the p_{T} of tracks matched to the jet, which originates from the primary vertex, to the scalar sum of the p_{T} of all tracks matched to the jet, independently of their origin. For each vertex, a JVF value can be defined per jet in the event. Figure 5.6 (a) shows the expected discrimination power of the JVF, which peaks at one for jets arising from the hard-scatter and at zero for jets arising from other vertices (pile-up jets). Figure 5.6 (b) is the JVF distribution in a nearly pure sample of hard-scatter jets. The sample is obtained by selecting $Z \rightarrow ee$ events with jets produced back to back with the Z boson. As can be seen, the JVF of hard scatter jets strongly peaks at $JVF = 1$ and the MC accurately describes the data behaviour.

The optimal working point for top quark analyses is the rejection of jets with $|JVF| < 0.50$ which gives a 95% selection efficiency for jets from primary interaction and rejects 75% of the pile-up jets. At high p_{T} the pile-up contribution becomes negligible, thus the JVF requirement is only applied to jets with $p_{\mathrm{T}} < 50\,\mathrm{GeV}$. Since tracking information is required, the JVF is only valid for $|\eta| < 2.4$ jets.

5.4.2.6 *b*-tagged Jets

Hadronic showers containing b or c quarks can be distinguished from showers made of lighter quarks due to their longer lifetimes ($\tau \sim 1.5\,\mathrm{ps}$) and leptonic decay signatures. For instance, B-hadrons produced with significant transverse momentum ($> 20\,\mathrm{GeV}$) will typically fly a few millimetres within their lifetime before decaying. This results in a displaced secondary vertex of a B-hadron decay with respect to the primary one, from which several displaced tracks originate. Figure 5.7 is a representation of a B-hadron decay inside a jet shower, including the important variables used to discriminate heavy flavour jets from light flavour

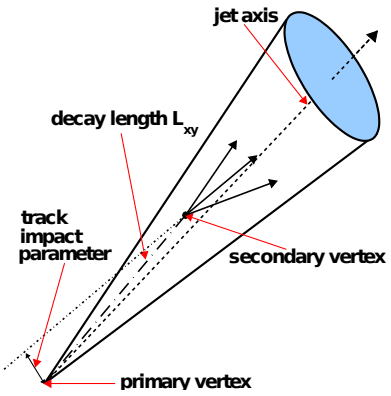


Figure 5.7: Sketch of B-hadron decay inside a jet shower [263].

ones: the track impact parameter,[1] the reconstructed secondary vertex and the length of the decay. [2] Atlas b-tagging algorithms rely on the use of these variables (standalone or combined) in order to identify true b-jets, while simultaneously reducing the contamination from true light jets that are mis-identified as b-jets.

Identifying or tagging jets originating from b quarks is particularly important in this analysis, due to the elevated number of b-jets in the signal topology. There are several b-tagging algorithms available. The following are the most relevant for the present analysis:

IP3D Compares input variables to MC templates for both the b-jet and light-jet hypotheses, using the longitudinal and transverse impact parameter of the tracks to built a two dimensional likelihood ratio discriminant [264].

[1]The transverse track impact parameter, d_0, is the distance of closest approach of the track to the PV in the $x-y$ plane. The longitudinal impact parameter, z_0, is the distance of closest approach of the track to the PV in the z direction.

[2]The length of the decay is defined as the distance of the secondary vertex from the PV in the plane orthogonal to the proton beam direction

SV0 and SV1 Both algorithms rely on the secondary vertex reconstruction and combine several variables using a likelihood ratio technique [264]. SV0 is based on the decay length significance in order to reduce the effect of poorly-measured vertices. SV1 is an extension of SV0, including additional variables: the invariant mass of all tracks associated with the vertex, the ratio of the sum of the energies of the tracks in the vertex to the sum of the energies of all tracks in the jet, and the number of two-track vertices.

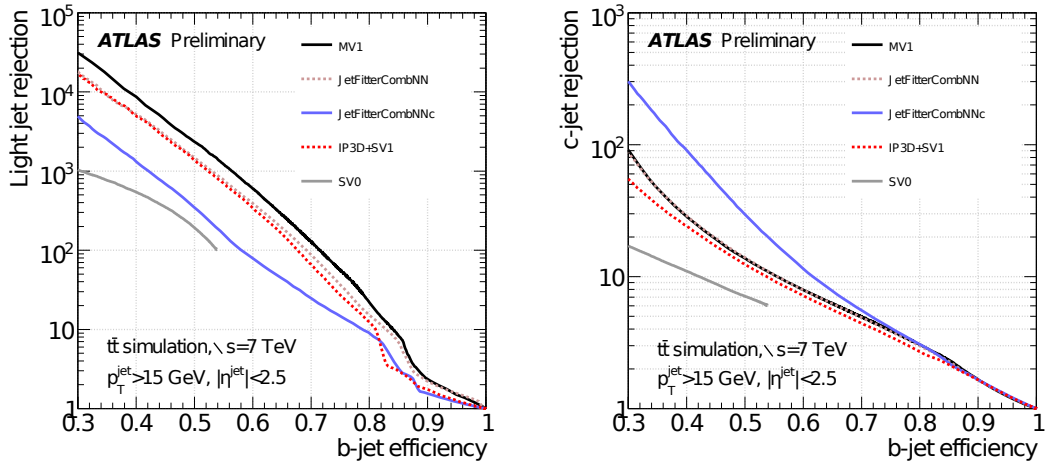
JetFitter Based on the topological structure of B and C-hadron decays inside jet showers, this algorithm tries to reconstruct the internal decay chain of a jet. Secondary and tertiary vertices are identified using a Kalman-fitter approach, assuming that they lie on the flight direction of the B-hadron[265].

JetFitterComNN Combines the output of the IP3D and JetFitter algorithms in a neural network, in order to improve the discrimination power. There are two versions of the algorithm, JetFitterComNN and JetFitterComNNc, using explicitly trained NN to identify b-jets and c-jets respectively [264].

MV1 This algorithm uses a neural network which combines the output of the IP3D, SV1 and JetFitterCombNN algorithms. A dedicated version, MV1c, which achieves better separation between jets originating from b quarks and jets originating from c quarks, was also developed [266].

The performance of each b-tagging algorithm can be evaluated through its efficiency to tag jets that do originate from b quarks, while simultaneously having high rejection rates for c and light-jets. These are usually determined as the inverse of the c-tagging efficiency (rate at which the b-tagging algorithm identifies jets that originate from c quarks) and mis-tagging rate (rate at which the b-tagging algorithm identifies jets that actually originate from light quarks), respectively. In Figure 5.8, the b-tagging efficiency is compared to the light-jet and c-jet rejection rates. The b-tagging efficiencies measured in Monte-Carlo simulation as a function of the jet transverse momentum are calibrated to the values observed in data, using $t\bar{t}$ events.

In this analysis, the MV1 algorithm was used, as it provides the best light-jet rejection. Here, the MV1 was implemented by selecting a 70% efficiency working point, which translates



(a) (b)

Figure 5.8: Jet rejection of several b-tagging algorithms as a function of the b-jet efficiency for (a) light-jets and (b) c-jets [266].

into c and light mis-tagging rates of 20% and 1%, respectively. These rates were determined for b-tagged jets with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $|\eta| < 2.5$ in simulated $t\bar{t}$ events.

b-tagging calibration

The results, which correct the efficiency in simulation to that measured in data, are presented in the form of jet transverse momentum dependent (and in the case of the light-flavour jets, pseudorapidity dependent) scale factors (SF).

The tagging efficiencies of each operating point in simulation are then calibrated to that measured in data, using samples enriched in b-jets, c-jets and light jets, respectively [267]. The mis-modelling of the input variables used in the tagging algorithms can thus be corrected using the resulting scale factors, $\mathrm{SF} = \epsilon_{\mathrm{data}}/\epsilon_{\mathrm{MC}}$.

The b-tagging calibration was derived using a combinatorial likelihood approach, which takes into account the correlation between the multiple jets in the event. This calibration is done relying on a pure sample of b-jets, obtained from dileptonic $t\bar{t}$ events. For jets within 30 and 200 GeV, the calibration was done with a precision of a few percent. Since the main analysis of this thesis also requires dileptonic events, the b-tagging calibration used was derived from dileptonic $t\bar{t}$ events with exactly two jets, thus avoiding overlap of data.

The c-tagging calibration was derived using a sample of jets containing D mesons and comparing the yield of D mesons before and after the MV1 tagging cut [268]. This calibration required the reconstruction of D mesons within a jet, according to $D^{*+} \to D^0\pi^+ \to K^-\pi^+\pi^+$.

The mis-tag rate was determined using the 'negative tag' method in an inclusive jet sample [268]. Light-jets (originated from either gluons or u, d or s quarks) are expected to have symmetric distributions of track impact parameters or vertex decay length significance. To evaluate the mis-tag performance, tracks (vertices) with negative impact parameter (decay length significance) are used and their sign is reversed within the algorithm. Since the mis-tag rate depends on the jet kinematics, the measurement is performed in bins of jet p_{T} and η.

The scale factors were then applied to MC samples as event weight corrections: for jets tagged by the b-tagging algorithm, a weight equal to the SF of the corresponding jet flavour is applied; whereas for jets failing the b-tagging criteria, a weight corresponding to $(1-\mathrm{SF}\cdot\epsilon_{\mathrm{MC}})/(1-\epsilon_{\mathrm{MC}})$ is assumed. The event-level weight is then obtained by multiplying the individual jet weights for all the selected jets.

5.4.3 Missing Transverse Energy

Weakly interacting neutral particles, such as neutrinos, escape the ATLAS volume undetected. Before any collision occurs, the momentum of the incoming partons is mostly limited to the z direction. After the collision, if such non-interacting particles are created, then an apparent imbalance of the measured momentum in the transverse plane can arise. The transverse energy of the escaping particles can be inferred, using the conservation of momentum, as the vectorial sum of all the calorimeter energy depositions. In each pp collision, the negative vector sum of the p_{T} of all particles detected is usually defined as the missing transverse momentum, $\overrightarrow{p}_{\mathrm{T}}^{\mathrm{miss}}$, whose magnitude is the missing transverse energy, $E_{\mathrm{T}}^{\mathrm{miss}}$. In this analysis, where the signal is dileptonic $t\bar{t}H$ events, neutrinos are part of the expected topological final state particles and thus $E_{\mathrm{T}}^{\mathrm{miss}}$ is foreseen.

For the $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction [269] it is necessary to have maximum detector coverage, including all reconstructed objects in the event (after calibration) and any remaining unmatched calorimeter deposits and tracks. Thus, the $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction comprises all the energy deposits in the calorimeter, as well as, reconstructed information from the muon spectrometer. In order to recover the contribution from low transverse momentum particles,

otherwise lost in the calorimeter, track information is also considered. Using reconstructed and calibrated objects, the calorimeter energy deposits are considered in the following order: electrons, jets and muons[1]. By adding the energy measurement associated to physics objects in a specific order, double counting of deposits is avoided. In this analysis, the $E_{\mathrm{T}}^{\mathrm{miss}}$ was calculated from both E_x^{miss} and E_y^{miss} components, which result of the sum of the following ordered terms:

$$E_x^{\mathrm{miss}} = E_x^{\mathrm{miss},\,e} + E_x^{\mathrm{miss,\,jet}} + E_x^{\mathrm{miss},\,\mu} + E_x^{\mathrm{miss,\,soft-jet}} + E_x^{\mathrm{miss,\,cell-out}}\,. \tag{5.9}$$

The term $E_x^{\mathrm{miss,\,soft-jet}}$ includes the contribution from LCW jets with $10 < p_{\mathrm{T}} < 20\,\mathrm{GeV}$ without jet area correction. For energy deposits not associated to any physics object, a *Cell-Out* term, $E_x^{\mathrm{miss,\,cell-out}}$ is calculated.

Other sources of missing transverse energy can lead to a mis-measurement of the transverse missing energy. The most common fake missing transverse energy sources are the inactive regions of the detector, noise sources and overall detector resolution, as well as cosmic-ray and beam-halo muons interacting with the detector material [270]. Such effects are considered in the systematic uncertainties associated to this variable.

5.5 Preselection of Events

For the $t\bar{t}H$ dileptonic search, an event preselection tailored to the targeted topology is implemented based on a set of selection cuts. Further requirements relative to the quality of the event reconstruction and the detector status, commonly referred to as *event cleaning*, are also applied. From all the data collected with $\sqrt{s} = 8$ TeV, only that which verifies good quality criteria is considered (as described in Section 3.3.6), giving a total integrated luminosity of $20.3\,\mathrm{fb}^{-1}$.

Event Cleaning

Non-collision sources or detector noise are reduced by rejecting events where no well reconstructed primary vertex[2] is found within the beam collision region. Additionally, particles produced by cosmic interactions are rejected by discarding events where two opposite sign

[1] Note that hadronically-decaying taus and photons are not considered in this analysis.

[2] In this analysis, a good reconstructed primary vertex is defined as that with at least five associated tracks with $p_{\mathrm{T}} > 500MeV$ (see Section 5.3)

muons, with $|d_0| > 0.5\,\text{mm}$ and opposite direction ($\Delta\phi > 3.10$), are present. An electron-muon overlap consistency check is also applied, by rejecting events where a selected electron and muon share an inner detector track ($\Delta\theta < 0.005$ and $\Delta\phi < 0.005$). However, contamination from mis-identified leptons (either from jets or by mis-measurement of charge sign) can still occur. The estimation of this fake lepton background was derived from data and is discussed in the following Chapter 6. To avoid double counting in fake lepton estimation, the reconstructed leptons are required to match the true leptons from the MC simulation. The jet quality is verified by rejecting events where a bad jet with $p_{\text{T}} > 20\,\text{GeV}$ and $|\eta| < 4.5$ is found. This quality requirement avoids contamination of the $E_{\text{T}}^{\text{miss}}$ calculation from mis-measured jets.

Preselection Cuts

Events are required to have exactly two isolated leptons of opposite charge[1], defined according to the quality criteria mentioned in Section 5.4.1. The leading and sub-leading lepton must have $p_{\text{T}} > 25\,\text{GeV}$ and $p_{\text{T}} > 15\,\text{GeV}$, respectively. Events can then be categorised according to lepton flavour into ee, $e\mu$ and $\mu\mu$ selections. As already mentioned, selected events are triggered with a single-lepton trigger. Therefore, at least one of the leptons must be matched within $\Delta R = 0.15$ to the one reconstructed at the trigger level. In order to be above the trigger threshold, the reconstructed lepton matching the trigger one must have $p_{\text{T}} > 25\,\text{GeV}$. Furthermore, events with additional leptons passing this selection are removed to avoid statistical overlap between other ATLAS analysis.

Jets must satisfy the criteria defined in Section 5.4.2 with $p_{\text{T}} > 25\,\text{GeV}$ and $|\eta| < 2.5$. In the $t\bar{t}H$ dileptonic signal, at least four b quarks are expected. Despite the quite high number of b quarks involved, events are only required to have at least two jets tagged as b-jets using the MV1 algorithm (see 5.4.2.6). By establishing this compromise selection, non $t\bar{t}$ backgrounds are highly reduced, while still achieving high signal efficiency.

A set of conditions was also required according to the specific dilepton combination, in order to suppress non-$t\bar{t}$ background contributions. In the $e\mu$ selection, the scalar sum of the transverse energy of all selected leptons and jets, H_{T}, is required to be above 130 GeV. This H_{T} cut significantly reduces $Z/\gamma^* + \text{jets}$ production, where the Z boson decayed into leptonically decaying τ leptons. In the ee and $\mu\mu$ selections, a cut of low values of the invariant mass of the two leptons, $m_{\ell\ell}$, is applied, requiring $m_{\ell\ell} \geq 15\,\text{GeV}$ in events with

[1]Throughout this thesis the lepton pair is typically referred to as opposite sign (OS) leptons.

Preselection Cuts	
Exactly 2 OS leptons with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $p_{\mathrm{T}} > 15\,\mathrm{GeV}$	
Lepton matching trigger within $\Delta R = 0.15$ with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$	
≥ 2 b-tagged jets with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ & $\lvert\eta\rvert < 2.5$	
***ee* and *μμ* selections:**	***eμ* selection:**
$m_{\ell\ell} \geq 15\,\mathrm{GeV}$ and $\lvert m_{\ell\ell} - m_Z\rvert > 8\,\mathrm{GeV}$	$H_{\mathrm{T}} > 130\,\mathrm{GeV}$
$m_{\ell\ell} \geq 60\,\mathrm{GeV}$ if exactly 2 b-tagged jets	

Table 5.3: Preselection requirements for dileptonic $t\bar{t}H$.

more than two b-jets. This lepton cut suppresses contributions from the decay of hadronic resonances, such as the J/Ψ and Υ. For same-flavour lepton events with exactly two b-jets an extra requirement is made on the $m_{\ell\ell}$, requesting it to be larger than 60 GeV, due to mis-agreement between data and MC prediction. Events in ee and $\mu\mu$ selections are particularly sensitive to Z + jets production. A further $m_{\ell\ell}$ cut, requiring $|m_{\ell\ell} - m_Z| > 8\,\mathrm{GeV}$, is thus applied to reject events close to the Z boson mass resonance.

Table 5.3 includes a summary of the preselection requirements.

Considering the high jet and b-jet multiplicity of the signal final states, pre-selected events are classified into exclusive categories according to the number of reconstructed jets and b-tagged jets. These categories are referred here as $(m\mathrm{j}, n\mathrm{b})$ regions for having m jets of which n are b-tagged. This classification, without loss of events, is very effective in suppressing SM backgrounds, while maintaining the highest possible signal sensitivity. In order to search for the signal, a combined fit to signal-rich and signal-depleted regions will then be performed. By combining information from these differently populated regions, an improved background prediction can also be achieved. Signal depleted regions are used either as orthogonal control regions, allowing to match the simulation prediction to that of data, or as sources of background information to constrain the statistical fit.

For the purpose of the background and signal studies present in this thesis, the dilepton flavour selections, ee, $e\mu$ and $\mu\mu$, were kept separated, unless stated otherwise. Nevertheless, these dilepton selections were ultimately combined and treated coherently for the statistical and discriminant analysis as described in Chapter 7.

6

Signal and Background Modelling

High signal sensitivity and precise background modelling are crucial for the $t\bar{t}H$ search. Therefore, dedicated studies of the signal and background characteristics were developed, as mentioned in the previous chapter.

An overview of the dileptonic $t\bar{t}H$ signal is presented in Section 6.1, as well as the challenges associated with such a search.

The most relevant backgrounds for the $t\bar{t}H$ study are investigated, using dedicated event selections and control regions. In Section 6.2, data-driven corrections were determined for the Z+jets, the $t\bar{t}$+jets and multijet processes, which are subsequently applied to the MC simulated samples in order to match the observed data.

6.1 Dileptonic $t\bar{t}H$ Signal

The dileptonic $t\bar{t}H$ associated production has a very busy final state. Depending on the Higgs Boson decay mode, the signal's signature may vary, but it will unavoidably originate high jet multiplicity. In order to identify the signal, high jet and b-tag jet multiplicities are required in this analysis. Events verifying the preselection cuts are categorized in $(m\text{j}, n\text{b})$ multiplicity regions, which are treated individually. In order to have statistical significance, the $t\bar{t}H$ dilepton analysis considered a total of six independent regions: (2j, 2b), (3j, 2b), (3j, 3b), ($\geq$ 4j, 2b), ($\geq$ 4j, 3b) and ($\geq$ 4j, $\geq$ 4b). Signal significance will be different in each of them, as well as background composition. The separate treatment of these regions guarantees the maximum selection of signal events possible, while allowing for dedicated background studies in regions with minimum to no signal presence.

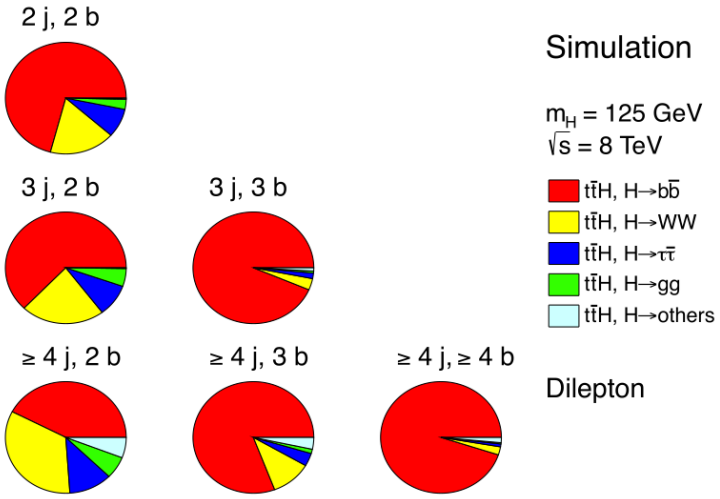


Figure 6.1: Fractional contributions of the various Higgs Boson decays to the total signal prediction, after preselection, in all $(m\text{j}, n\text{b})$ regions of the analysis. Each row shows the distribution for a specific jet multiplicity (2, 3, ≥ 4), and the columns show the b-jet multiplicity (2,3, ≥ 4).

In Figure 6.1 the contributions of different Higgs boson decay modes, in each of the analysis regions, are shown. The $H \to b\bar{b}$ decay is the dominant contribution in the expected signal-rich regions, ($\geq$ 4j, 3b) and ($\geq$ 4j, $\geq$ 4b). Other decays such as $H \to WW$, $H \to \tau\bar{\tau}$ and $H \to gg$ are also significant, though to a lesser extent in the higher multiplicity bins. In this study, all Higgs boson decays are treated as signal. Nonetheless, the analysis was optimised for the most dominant decay by taking into account its kinematic properties.

The $t\bar{t}H$ dilepton final state has a very clean signature, that allowed for simple preselection cuts, optimised for the signal sensitivity and background reduction. However, its low cross-section is still a challenge. Even after event preselection, all regions of the analysis are contaminated by several background processes at a much larger magnitude than that of the signal. The $t\bar{t}$ production, with a cross-section $\sim$ 500 times larger than that of the $t\bar{t}H$ signal and an almost identical final state signature, is the dominant background in all the regions of the analysis. Other major background contributions originate from processes containing at least two leptons, such as the Z boson production. Processes involving at least a top quark, such as the single top quark production or the $t\bar{t}V$ processes, have very small contributions to the background due to their lower cross-sections. While these overwhelming background rates may hinder the dileptonic $t\bar{t}H$ search, the detailed knowledge of the

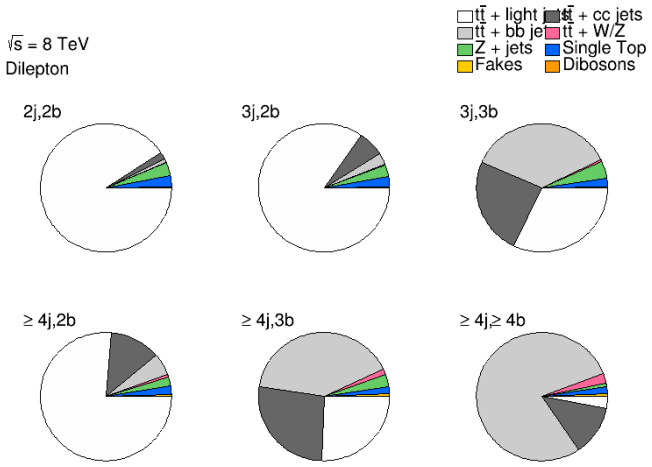


Figure 6.2: Fractional contributions of the various backgrounds to the total background prediction after preselection in all $(m\text{j}, n\text{b})$ regions of the analysis.

different backgrounds will considerably improve the performance of the analysis.

In order to assess the expected analysis performance and test the compatibility of a possible excess in data with a given hypothesis, it is necessary to increase the signal sensitivity as best as possible while also improving the background modelling. As already mentioned, the classification in $(m\text{j}, n\text{b})$ regions allows the highest possible signal sensitivity without loss of events, while the event preselection reduces the contamination from several SM backgrounds, in particular processes without the presence of top quarks. The surviving backgrounds, however, have substantially larger cross-sections than those of the $t\bar{t}H$ associated production. It is because the dilepton $t\bar{t}H$ search has such significant backgrounds, that their accurate knowledge is of major importance. Dedicated methods, tailored to each of the main background production mechanisms, were developed in order to have the best possible accuracy, *i.e.* that of data. These data driven methods to improve background modelling will be discussed in the following section.

6.2 Background Estimation

The background composition across different jet and b-tagged jet multiplicity regions is shown in Figure 6.2.

In this analysis, the dominant background is the $t\bar{t}$+jets production, accounting for more than 90% of the background in all regions. The HF contribution, specially the $t\bar{t} + b\bar{b}$ jets production, is particularly relevant since it is the main irreducible background in the signal regions, having an almost identical signature as the signal itself. While precise measurements of the $t\bar{t} + b\bar{b}$ jets cross-section were underway, the thorough modelling of the heavy flavour component of the associate top quark pair production was necessary for this analysis. Based on tagging efficiencies and using control regions with low or no b-tagged jets, MC corrections were estimated to match data precision. Being the dominant background for the $t\bar{t}H$ analysis, it is important that its modelling be determined in the most accurate way possible. As such, contamination from other background sources needs to be precisely assessed in both control and signal regions. Given the nature of the dilepton $t\bar{t}$ production, other processes involving two leptons, such as the Z+jets production, must then be previously modelled to data. This way, the discrepancies seen in $t\bar{t}$+jets control regions will constitute purer sources of information on the $t\bar{t}$ process behaviour.

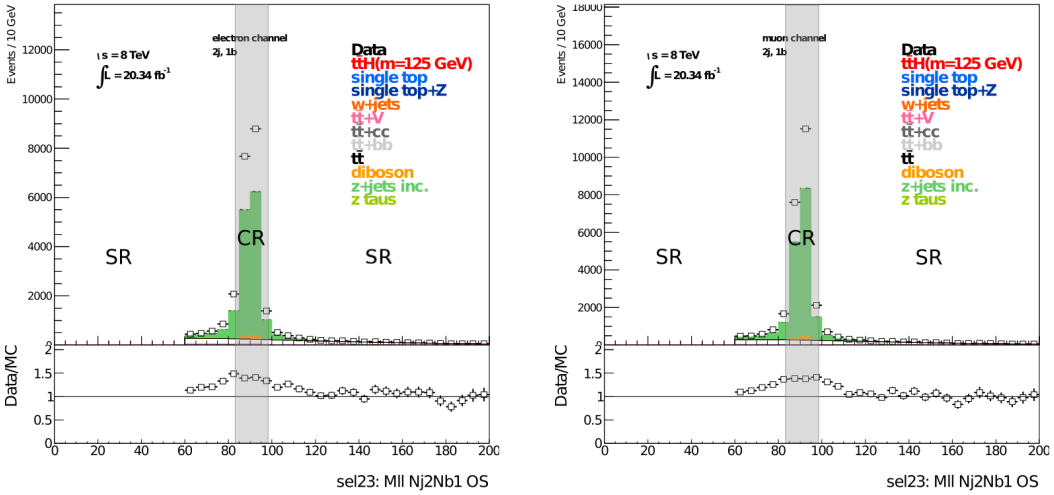
Another important background contribution originates from the production of Z bosons in association with jets. This contribution, although not dominant, has a significant impact on the data and MC agreement, with the Alpgen simulation predicting harder transverse momentum of the Z boson than observed. Derived from otherwise discarded dilepton events within the Z boson mass peak, the data driven corrections for this background were the first to be evaluated. The selected control region is a clean sample of the Z boson events and completely orthogonal to the analysis.

Finally, contributions from multijet events are also considered. Non-prompt leptons or misidentification of jets or photons as electrons can lead to the selection of events with fake leptons. The expected yield of misidentified lepton background is derived from data using dilepton events where the lepton pair has the same electric charge.

In the following the data-driven methods developed for the precise estimation of each background are presented.

6.2.1 Z+jets Data Driven Estimation

The production of a Z boson in association with jets constitutes a significant source of background in same flavor dilepton channels. Although Drell-Yan processes involving photons have significantly higher cross-sections, for this study, where high energy final state objects are required, Z boson decays are a more relevant background. In this analysis, the Z+jets



(a) (b)

Figure 6.3: Dilepton mass distribution in the 2 jet, 1b−tagged jet region for the dielectron (a) and dimuon (b) channels, where the control regions (CR) are shaded.

production accounts for approximately 3% of the total background processes. In order to reach good sensitivity for $t\bar{t}H$ dilepton signals, in addition to the basic rejection cuts used to suppress background events with a Z boson decay (*c.f.* preselection criteria in Section 5.5), it is crucial to control the yield of events, as well as the shape of the kinematic distributions. In the MC simulations, these processes are predicted at NLO accuracy, but suffer from large theoretical uncertainties originating for example from the choice of the PDFs and the renormalisation and factorisation scales. The MC predictions have reasonable agreement with data and allow the use of data driven information to constrain its systematic uncertainties [231]. In this section a fully data-driven method has been developed to estimate the Z+jets contribution to the signal region (SR), minimizing the dependence on simulation by making use of the data information in orthogonal control regions (CR).

The Z+jets background normalisation was estimated in all jet and b-tagged jet multiplicity bins of the analysis. The treatment of the dielectron and dimuon channels was done independently. The contributions of the light and heavy flavour components of the Z+jets background vary with the jet and b-tagged jet multiplicity, both in normalisation and transverse momentum spectrum of the final state objects. Given the lack of MC statistics associated with heavy flavoured quarks, in this study the Z+jets events are treated

inclusively in terms of the flavour of the extra partons (light, c and b quarks). Although this discrimination is not performed, it is still possible to correct the Z+jets background in each $(m\mathrm{j}, n\mathrm{b})$ multiplicity bin individually. This treatment of the Z+jets population keeps the systematic sources of the data-driven correction to a minimum by avoiding further uncertainties associated with tagging algorithms.

The control regions were chosen in such a way as to be a pure source of Z+jets events, with the smallest contamination possible from both signal events and remaining background processes. Thus, only events with a Z boson decay were considered in the CRs, through a cut based selection of events with exactly two opposite sign same flavour leptons, whose corresponding dilepton invariant mass value, $m_{\ell\ell}$, was included in a window of 8 GeV around the Z mass peak, 91 GeV [19]. This means that all the selection cuts of the SR are applied, but the CR only includes events with $|m_{\ell\ell} - 91\,\mathrm{GeV}| < 8\,\mathrm{GeV}$, as opposed to the SRs where these events are excluded. In Figure 6.3 an example of the signal and control regions of a given jet and b-tagged jet combination is shown, for the dielectron and dimuon channels separately.

The Z+jets contamination in the CR from other background physics processes is subtracted from data, and a scale factor, $w_{j\,b}$, is derived using Drell-Yan simulations to convert observations in the CR into background estimates for the SR:

$$w_{j\,b} = \frac{Z\text{+jets}_{\,\mathrm{data}}}{Z\text{+jets}_{\,\mathrm{MC}}}\bigg|_{j\,b} = \frac{N_{\mathrm{data}} - N_{\mathrm{other\,MC}}}{N_{Z\text{+jets\,MC}}}\bigg|_{j\,b} \tag{6.1}$$

Here Z+jets$_{\,\mathrm{data}}$ and Z+jets$_{\,\mathrm{MC}}$ are the number of events consistent with a dilleptonic decaying Z boson within the CR for data and MC simulation respectively. The Z+jets$_{\,\mathrm{data}}$ is obtained by removing the contamination from all of other MC predictions in the CR ($N_{\mathrm{other\,MC}}$, which include the di-boson and other top quark production processes) from the total number of events observed in data (N_{data}).

The scale factor $w_{j\,b}$ is then used to renormalise the Z+jets background sample in the signal region, providing a better estimation of this background for each jet and $b-$tagged jet combination:

$$Z\text{+jets}_{\,\mathrm{j\,b}} = \mathrm{MC}^{(\mathrm{SR})}_{\mathrm{Z+jets}}\big|_{\,\mathrm{j\,b}} \times \mathrm{w}^{(\mathrm{CR})}_{\,\mathrm{j\,b}}. \tag{6.2}$$

The normalisation scale factors for all the jet and b-tagged jet multiplicity bins are summarised in Tables 6.1 and 6.2. The uncertainties were estimated by taking into account the statistical fluctuation of the data in the CR. The mismodelling of the Z+jets heavy

Number of Jets	ee-channel: Number of b-tagged Jets 0	1	2	3	≥4
2	0.9808 ± 0.0029	1.437 ± 0.012	1.573 ± 0.046		
3	1.036 ± 0.0065	1.372 ± 0.021	1.475 ± 0.062	1.519 ± 0.36	
≥4	0.9896 ± 0.012	1.223 ± 0.031	1.192 ± 0.077	1.696 ± 0.33	0.3556 ± 0.82

Table 6.1: The Z+jets renormalisation scale factor, $w_{j\,b}$, for all $(m\text{j}, n\text{b})$ multiplicity bins of interest in the analysis for the dielectron channel.

Number of Jets	$\mu\mu$-channel: Number of b-tagged Jets 0	1	2	3	≥4
2	0.9720 ± 0.0030	1.417 ± 0.011	1.541 ± 0.042		
3	1.014 ± 0.0059	1.341 ± 0.019	1.446 ± 0.057	1.326 ± 0.33	
≥4	1.001 ± 0.011	1.285 ± 0.030	1.341 ± 0.076	0.7308 ± 0.25	1.341 ± 1.08

Table 6.2: The Z+jets renormalisation scale factor, $w_{j\,b}$, for all $(m\text{j}, n\text{b})$ multiplicity bins of interest in the analysis for the dimuon channel.

flavour component in the MC predictions leads to data-driven scale factors typically above unity for bins with higher b-tagged jet multiplicity. Although the scale factors are different for the two lepton channels, in both cases they start a little bellow unit for regions without b-tagged jets and increase with the number of b-tagged jets up to ~ 1.5. For events with exactly 2 and 3 jets, the ee channel requires higher scale factors when compared with the $\mu\mu$ channel results. In regions with at least 4 jets, however, the corrections for both lepton channels are within statistical uncertainty of each other, except in the ≥ 4 jets, 3 b-tagged jets, where the electron channel scale factor is much higher.

In both dielectron and dimuon channels, the agreement between data and the signal-plus-background model after the cross-section renormalisation is improved. For illustration purposes, the dielectron invariant mass distribution before and after the correction of the Z+jets cross-section are shown in Figures 6.4 and 6.5 for exactly two, three and at least four jets and all the pertinent b-tagged jet multiplicity bins. Here, the Z+jets process is represented by the green histogram and the error on the data points includes only the statistical uncertainties on the measurement.

The kinematic properties of the leptons, as well as other significant objects, such as the jets, missing energy and H_T were investigated after the Z+jets data driven cross-section cor-

rection. In Figures 6.6, 6.7 and 6.8 the leading and subleading electron kinematic properties are summarised for multiplicity regions with exactly two jets. The angular distributions show a good MC to data agreement, however in the transverse momentum distributions it is evident a small mis-prediction of the behaviour seen in data. For other jet multiplicity bins, identical conclusions are drawn. The muon distributions show analogous properties and as such are omitted here.

The distributions of the transverse momentum of the leading and subleading jets, which can be seen in Figure 6.9, presents similar properties as the lepton ones, with the MC slightly mis-predicting the data behaviour. Just as for the leptons, in the jet angular distributions the data is well described by the MC.

For completeness, the missing transverse energy, $P_{\mathrm{T}}^{\mathrm{miss}}$, and the scalar sum of the transverse energy of all selected leptons and jets, H_{T}, were also investigated. Figure 6.10 shows these distributions for the two jet region of the dielectron channel as an example. Due to their nature, these distributions reflect the disagreement seen in the transverse momenta of individual objects, as expected. Similar results were observed for the 3 jet and ≥ 4 regions, as well as for the dimuon channel.

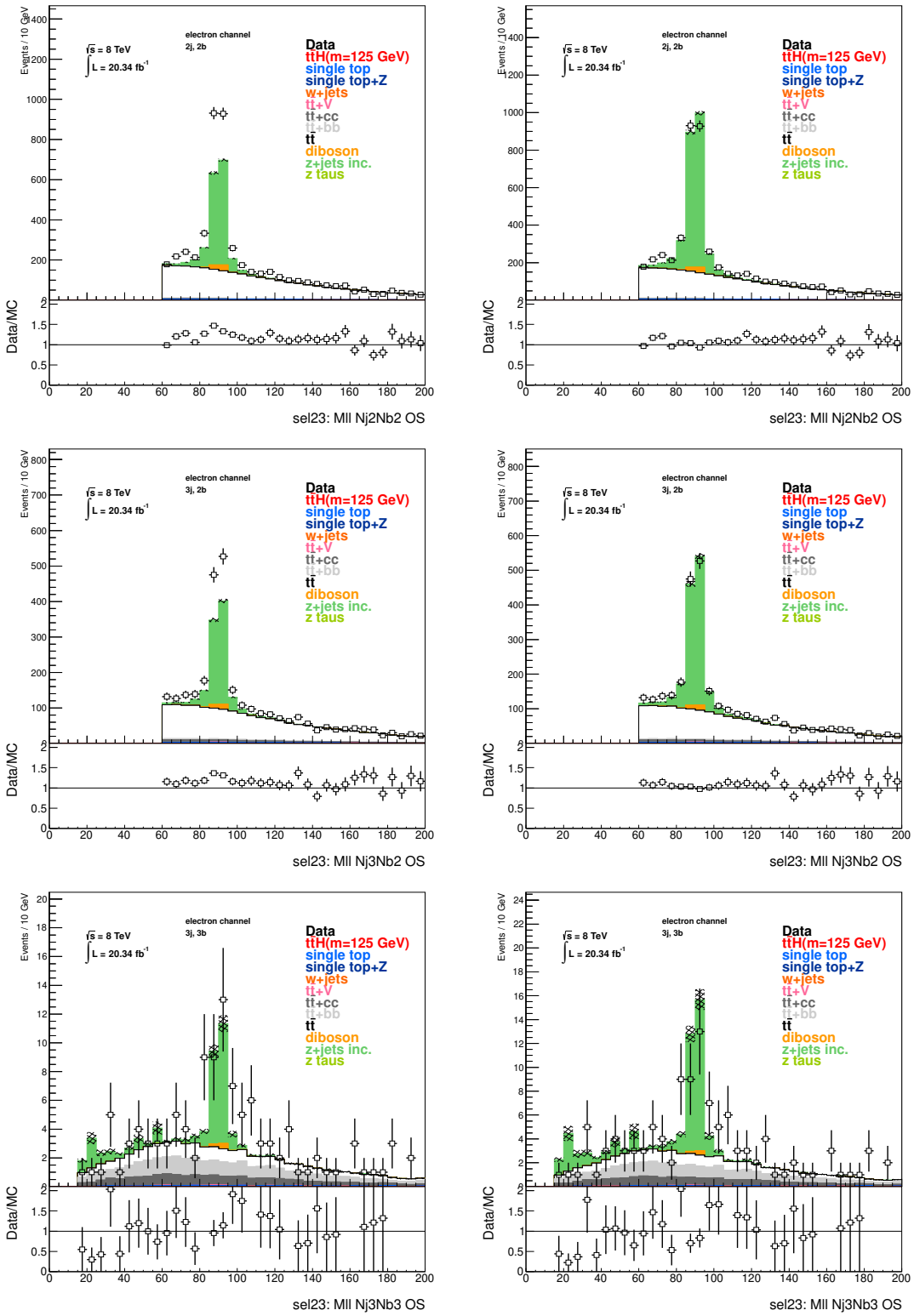


Figure 6.4: Dilepton mass distribution before (on the left) and after (right) the renormalisation of the Z+jets population, for exactly two and exactly three jets and all the relevant b-tagged jet multiplicity bins of the dielectron channel. The error bars associated to the data points include only the statistical uncertainties on the measurement.

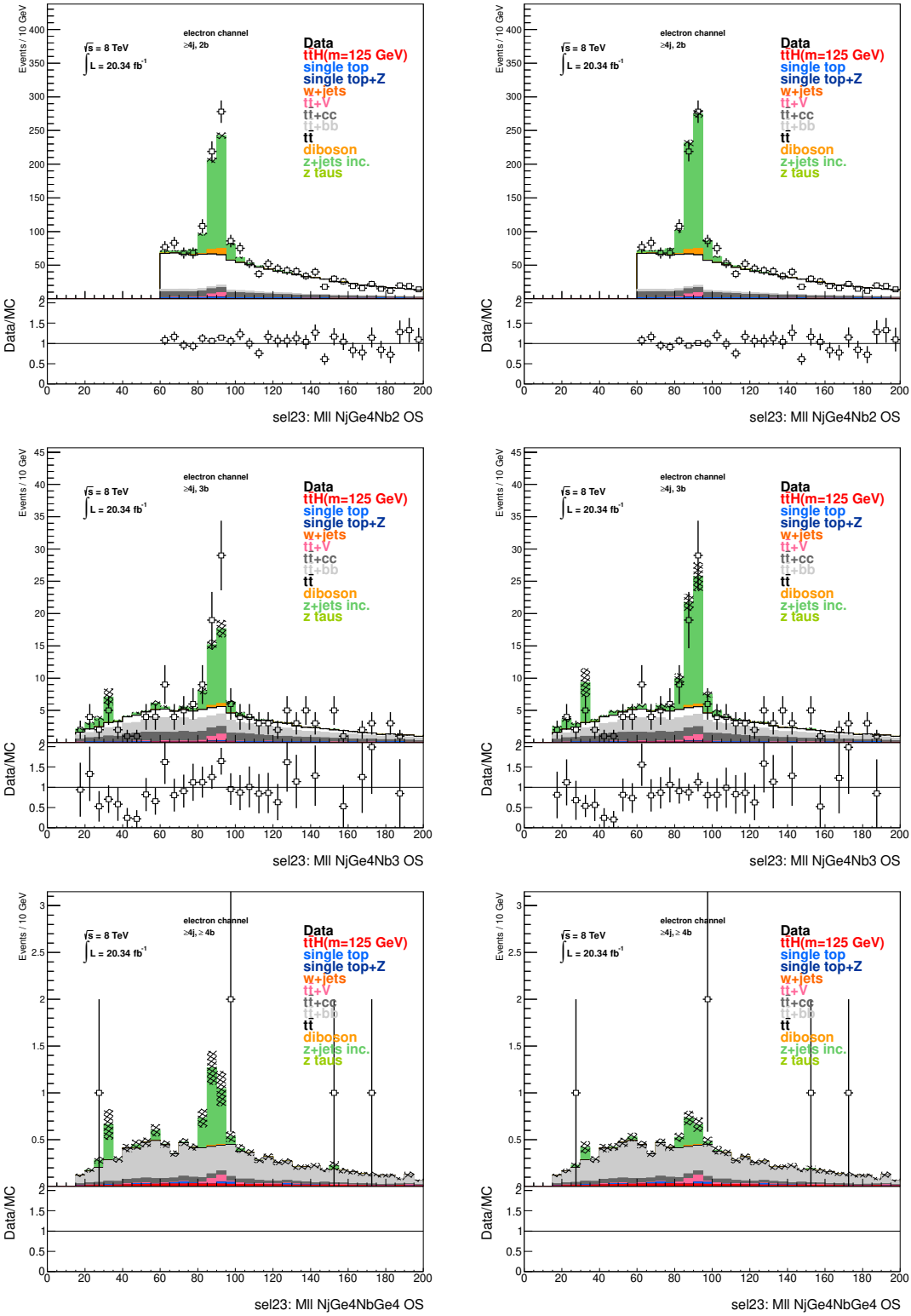
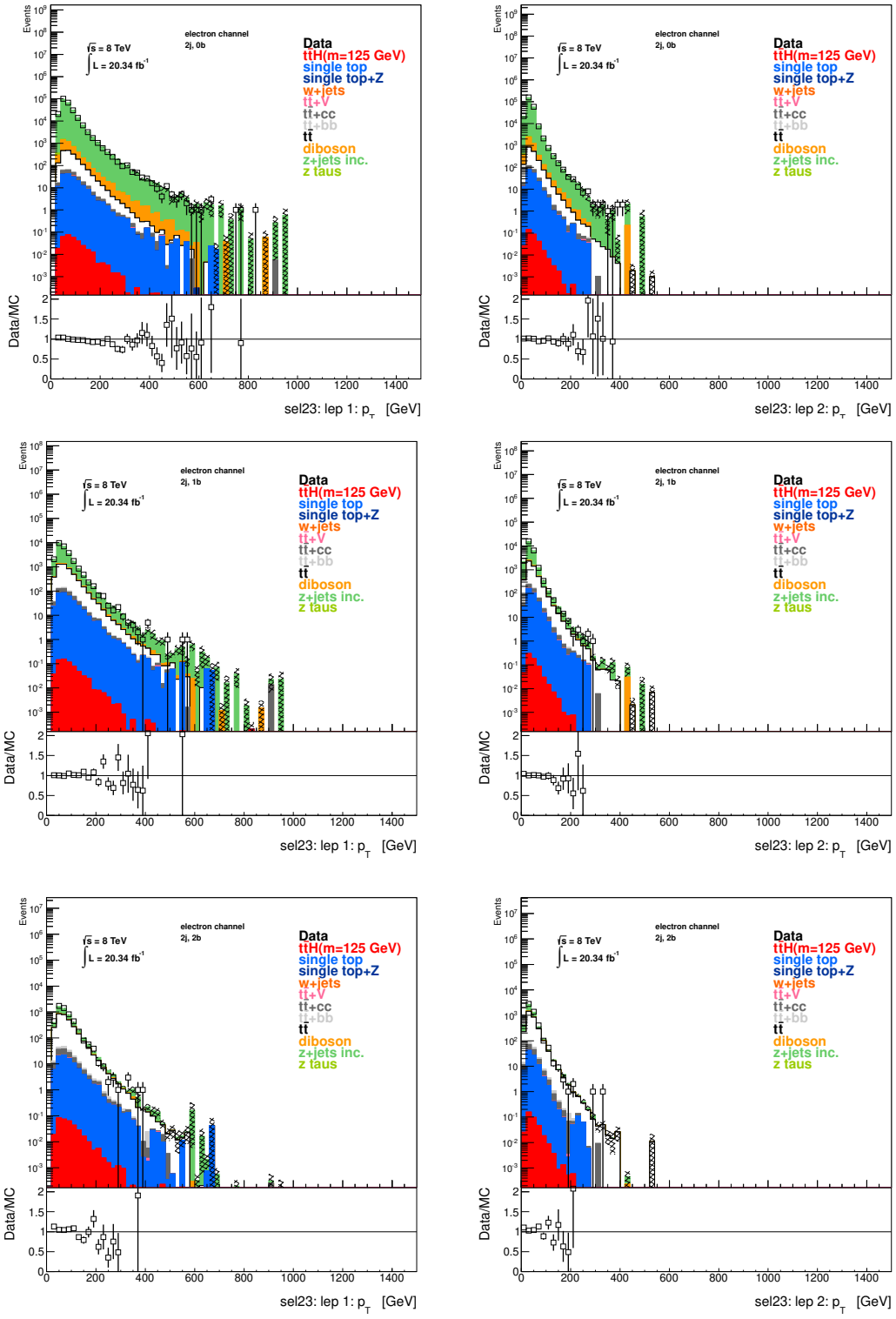
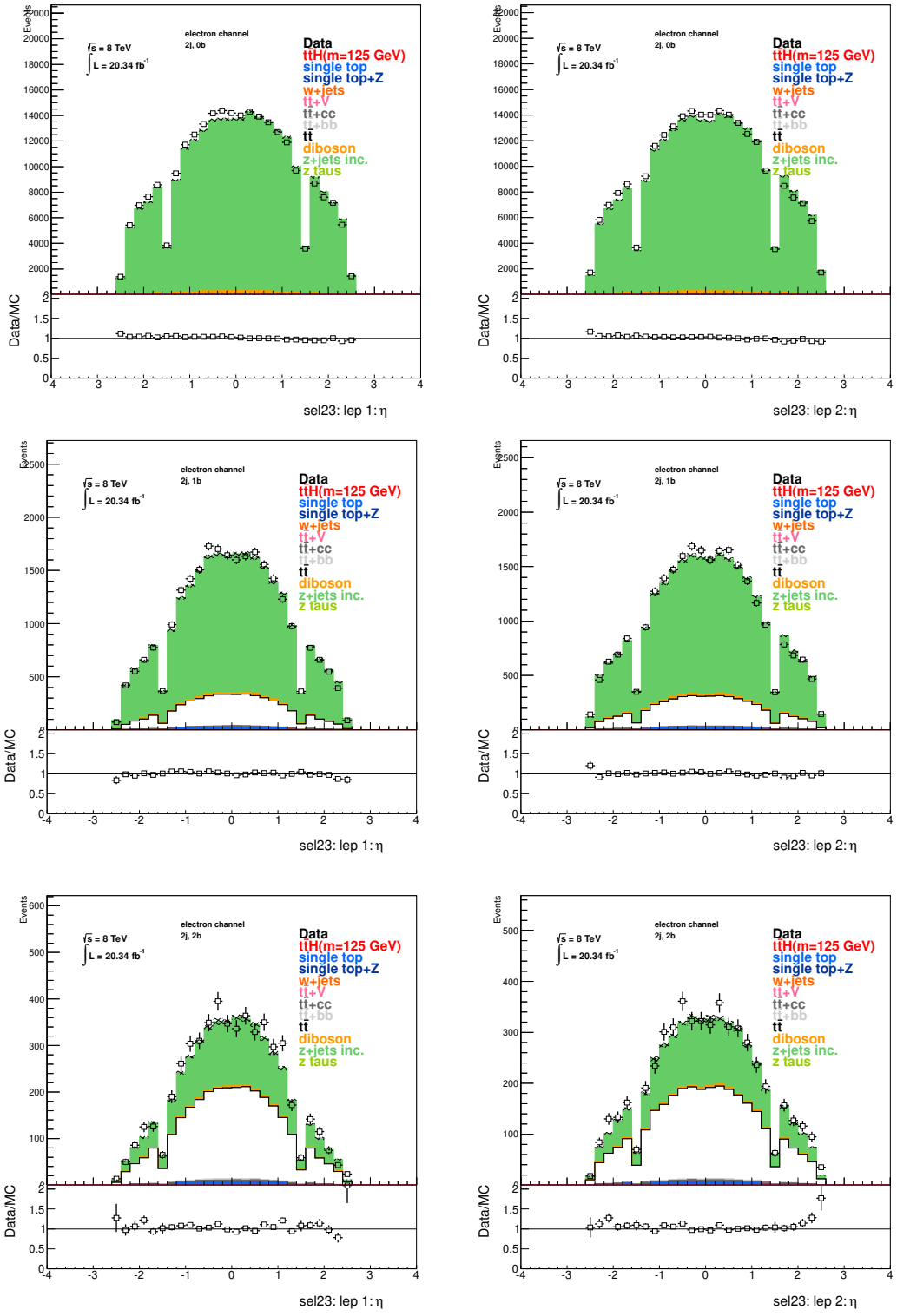


Figure 6.5: Dilepton mass distribution before (on the left) and after (right) the renormalisation of the Z+jets population, for at least four jets and the relevant b-tagged jet multiplicity bins of the dielectron channel. The error bars associated to the data points include only the statistical uncertainties on the measurement.



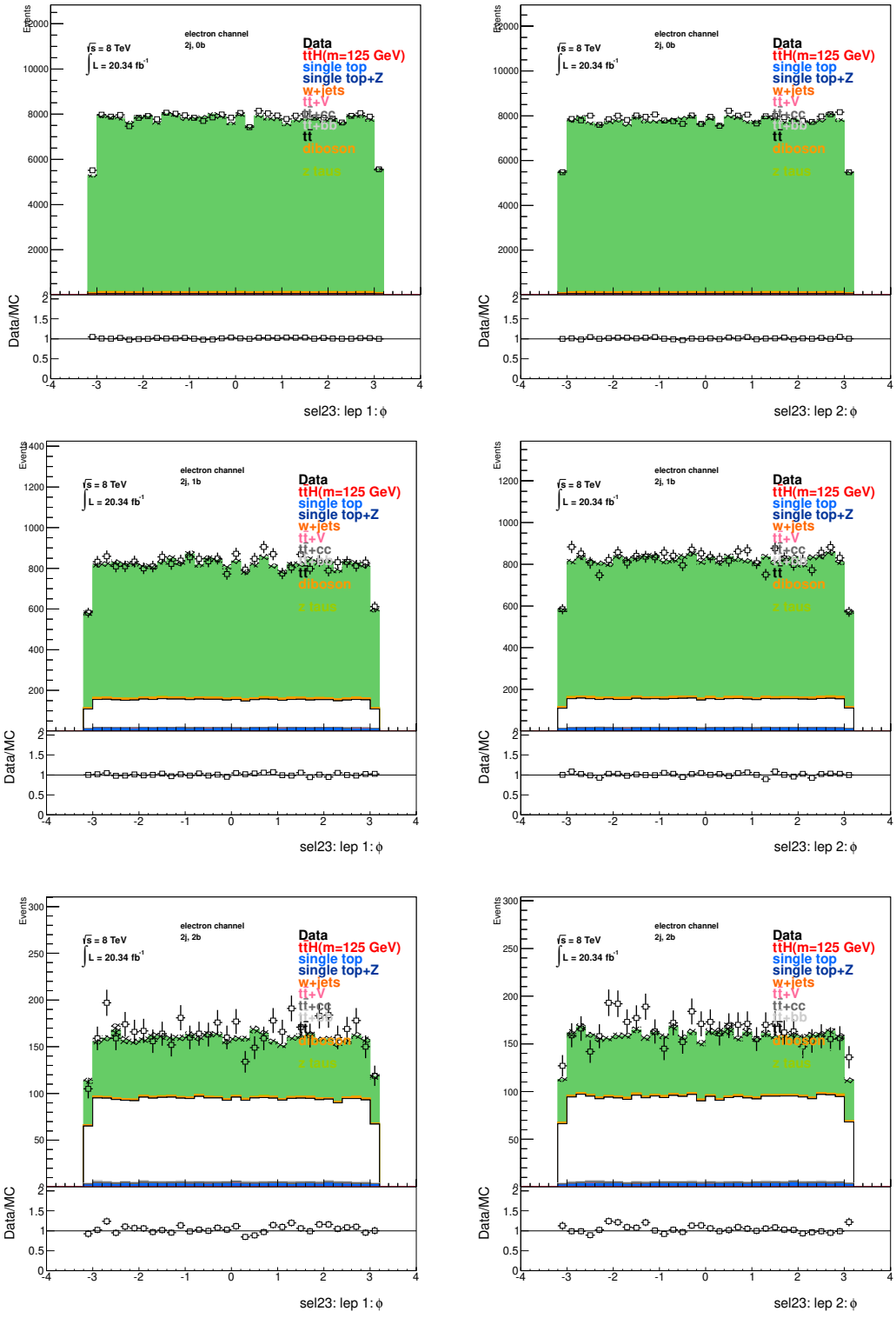
(a) (b)

Figure 6.6: Transverse momentum of the leading (a) and subleading (b) lepton in the two jet region of the dielectron channel after the Z+jets data driven cross-section correction. The error bars associated to the data points include only the statistical uncertainties on the measurement.



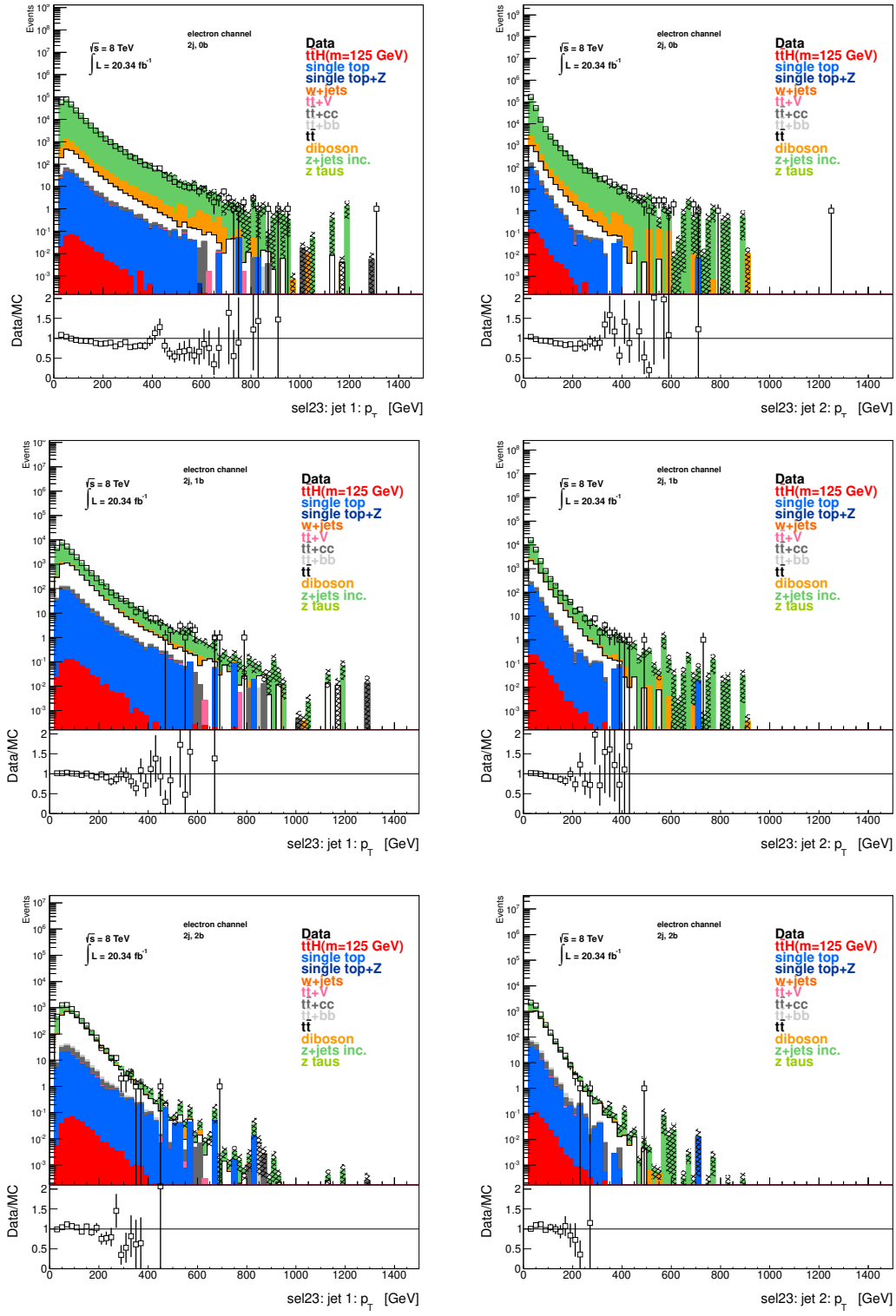
(a) (b)

Figure 6.7: Pseudorapidity, η, of the leading (a) and subleading (b) lepton in the two jet region of the dielectron channel after the Z+jets data driven cross-section correction. The error bars associated to the data points include only the statistical uncertainties on the measurement.



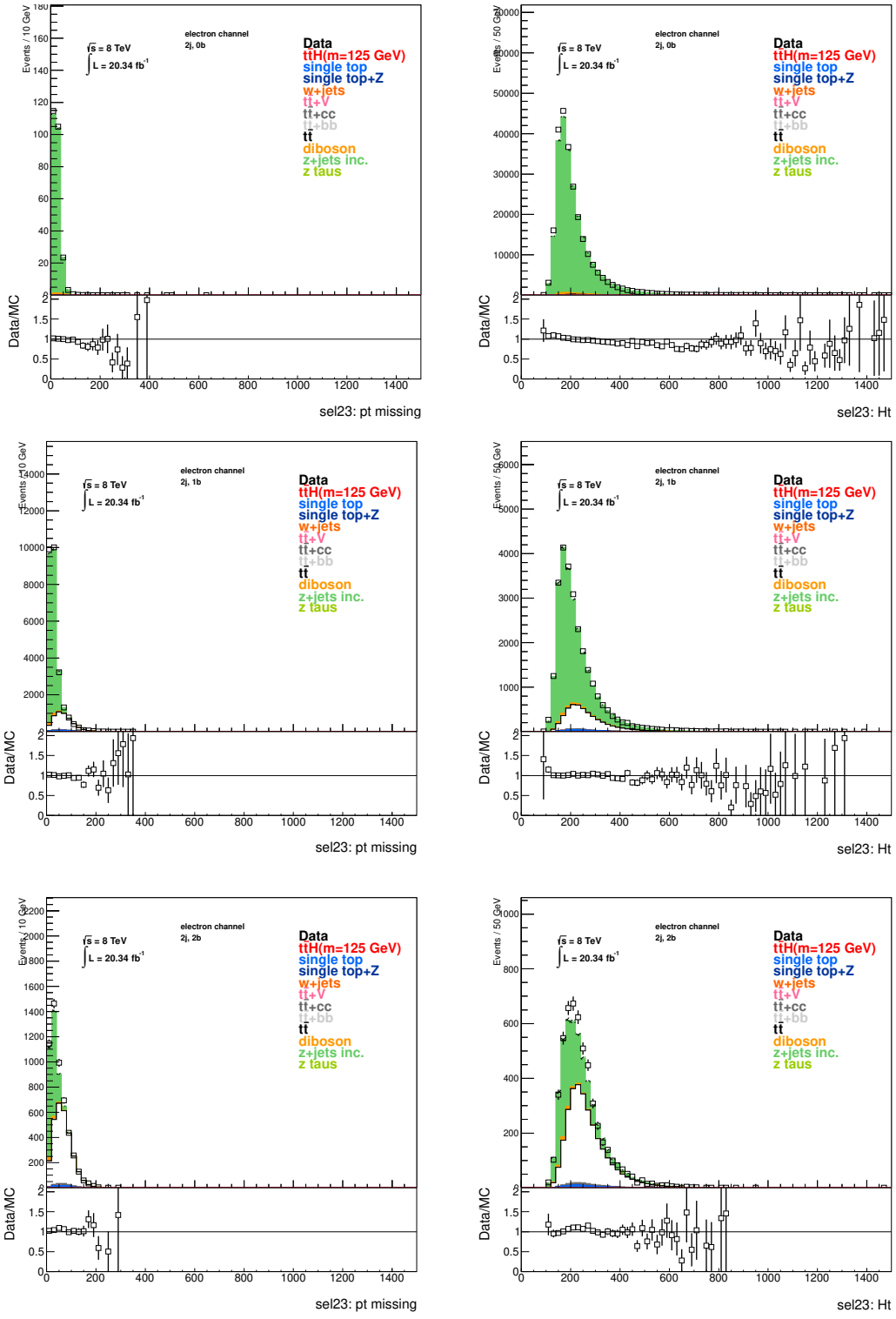
(a) (b)

Figure 6.8: Azimuthal angle, ϕ, of the leading (a) and subleading (b) lepton in the two jet region of the dielectron channel after the Z+jets data driven cross-section correction. The error bars associated to the data points include only the statistical uncertainties on the measurement.



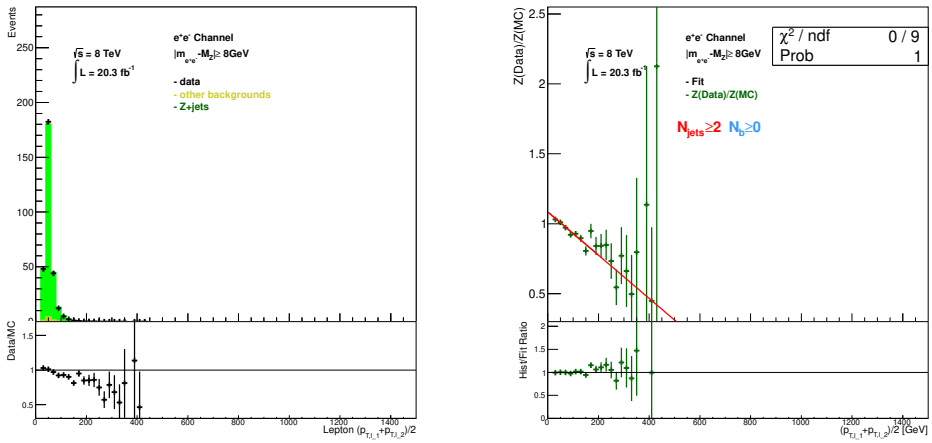
(a) (b)

Figure 6.9: Transverse momentum of the leading (a) and subleanding (b) jet in the two jet region of the dielectron channel after the Z+jets data driven cross-section correction. The error bars associated to the data points include only the statistical uncertainties on the measurement.



(a) (b)

Figure 6.10: Missing transverse momentum (a) and H_T (b) in the two jet region of the di-electron channel after the Z+jets data driven cross-section correction. The error bars associated to the data points include only the statistical uncertainties on the measurement.



(a) (b)

Figure 6.11: Average transverse momentum of the electron pair (a) and linear fit on $f_{\text{av.lep}}$ (b) for events with at least 2 jets within the CR ($|m_{\ell\ell} - 91\,\text{GeV}| \leq 8\,\text{GeV}$).

From the above study, it is clear that the ALPGEN generator predicts harder spectrums for the transverse momentum of all involved objects than those seen in data. This suggests a bad parameterisation of the radiation in the MC generator. Since this data to MC misprediction remains present even after adjusting the Z+jets rates to data, a second correction was developed. Linear fits were performed on the transverse momentum of leptons and jets separately, maintaining the Z+jets normalisation achieved in the previous correction.

As a first step, the lepton transverse momentum was corrected. The average transverse momentum of the pair of leptons,

$$< p_{\text{T}}^{l_1+l_2} >= \frac{p_{\text{T}}^{l_1} + p_{\text{T}}^{l_2}}{2}\,, \tag{6.3}$$

was chosen as it results in a softer correction than that obtained only from the leading lepton momentum. In Figure 6.11 (a) the average momentum of the electron pair is shown for events with at least two jets within the CR $|m_{\ell\ell} - 91\,\text{GeV}| \leq 8\,\text{GeV}$. The harder spectrum of the MC is clearly reflected on the slope present in the Data/MC distribution (bottom distribution).

Using the same control regions as before, linear fits are performed on

$$f_{\text{av.lep}} = \left.\frac{< p_{\text{T}}^{l_1+l_2} >_{\text{data}}}{< p_{\text{T}}^{l_1+l_2} >_{\text{MC}}}\right|_{j\,b}\,. \tag{6.4}$$

The linear function obtained per $(m\text{j}, n\text{b})$ multiplicity bin is then used as a correction of the Z+jets MC samples in the SR. Since the already corrected normalisation must be maintained, the linear fits are constrained to a single free parameter, the slope. Fits were performed in all $(m\text{j}, n\text{b})$ multiplicity regions individually. Due to the larger statistical uncertainties on events with higher average lepton transverse momentum, the linear fits were determined using only CR events below 200 GeV. Interestingly, similar nominal values were obtained for the slopes across all regions. A global fit was then performed including all multiplicity regions for each lepton channel and the obtained slope was used in the correction function. A conservative approach was implemented and regions where the obtained fit had a high χ^2 where not corrected. As an example, in Figure 6.11 (b), the linear fit of the electron pair is shown.

A second step was then implemented by analogously correcting the average transverse momentum of the most energetic pair of jets,

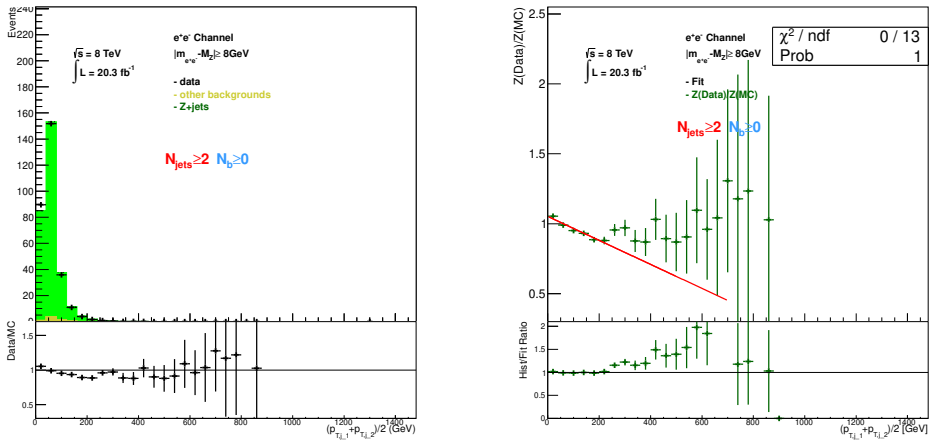
$$< p_{\text{T}}^{j_1+j_2} >= \frac{p_{\text{T}}^{j_1} + p_{\text{T}}^{j_2}}{2} \,. \tag{6.5}$$

The average momentum of the leading jet pair is shown in Figure 6.12 (a), where once again the harder spectrum of the MC is visible. Linear fits were performed on

$$f_{\text{av.jet}} = \left. \frac{< p_{\text{T}}^{j_1+j_2} >_{\text{data}}}{< p_{\text{T}}^{j_1+j_2} >_{\text{MC}}} \right|_{j\,b} \,, \tag{6.6}$$

and the obtained function was used to correct the Z+jets MC samples. These linear fits have a single degree of freedom in order to maintain the previously obtained Z+jets sample normalisation, just like those of the average lepton momentum.

All $(m\text{j}, n\text{b})$ multiplicity regions were individually fitted and similar nominal slope values were obtained once again. This just reiterates the fact that the source of disagreement is independent of final state multiplicities and affects all physics objects in the MC sample. In order to reduce the uncertainty on the slope estimate, a global fit was performed including all multiplicity regions for each leading jet pair. The linear fit of the average momentum of leading jets is shown in Figure 6.12 (b) for events with at least two jets within the CR. Once again, taking into account the larger statistical uncertainties of events with high average transverse momentum of the jet pair, the linear fits were calculated using CR events with $< p_{\text{T}}^{j_1+j_2} >$ below 500 GeV. The determined slope is then used in the correction function applied on the SR events for regions where a high χ^2 was not obtained.

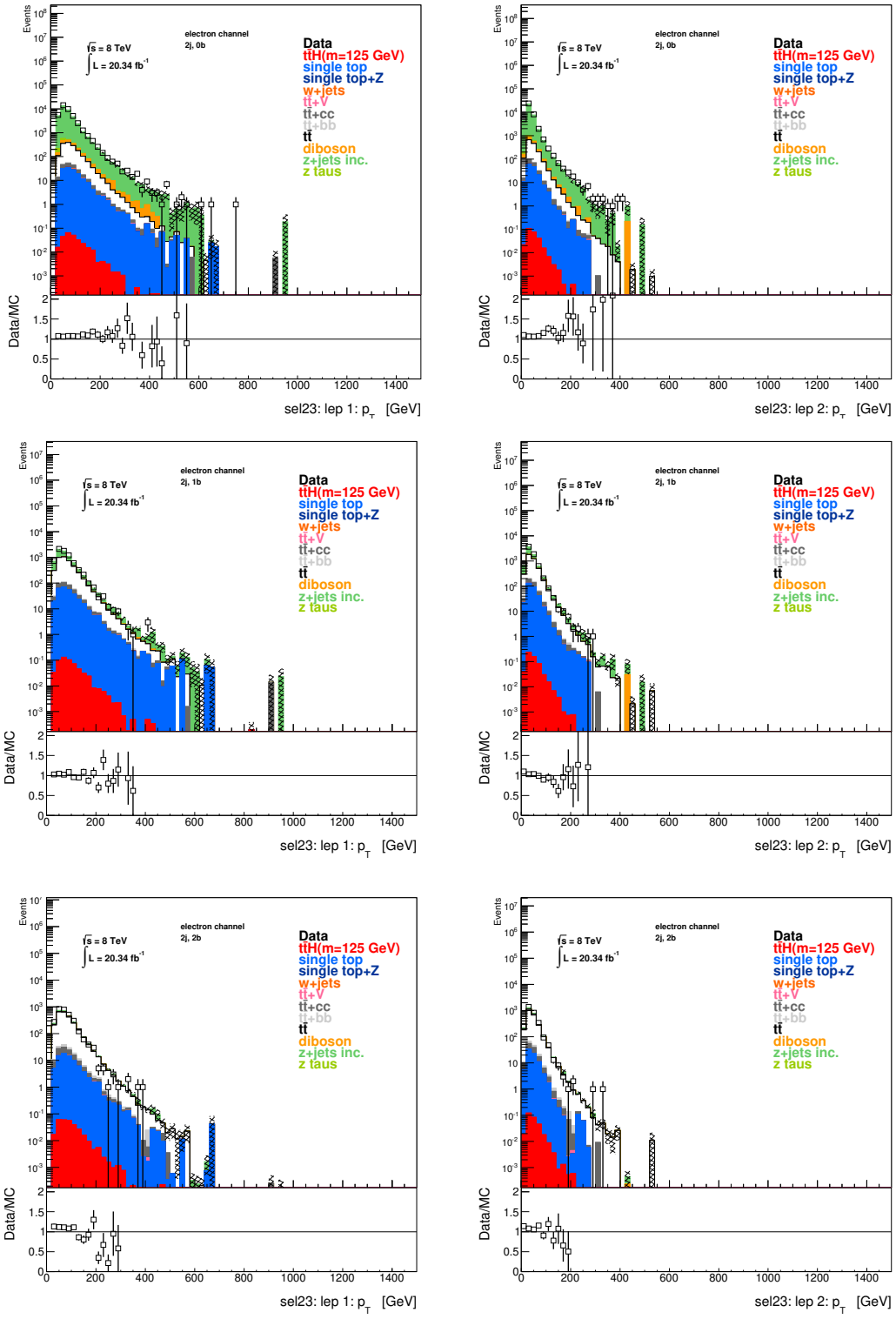


(a) (b)

Figure 6.12: Average transverse momentum of the leading jet pair (a) and linear fit on $f_{\text{av.jet}}$ (b) for events with at least 2 jets within the CR ($|m_{\ell\ell} - 91\,\text{GeV}| \leq 8\,\text{GeV}$).

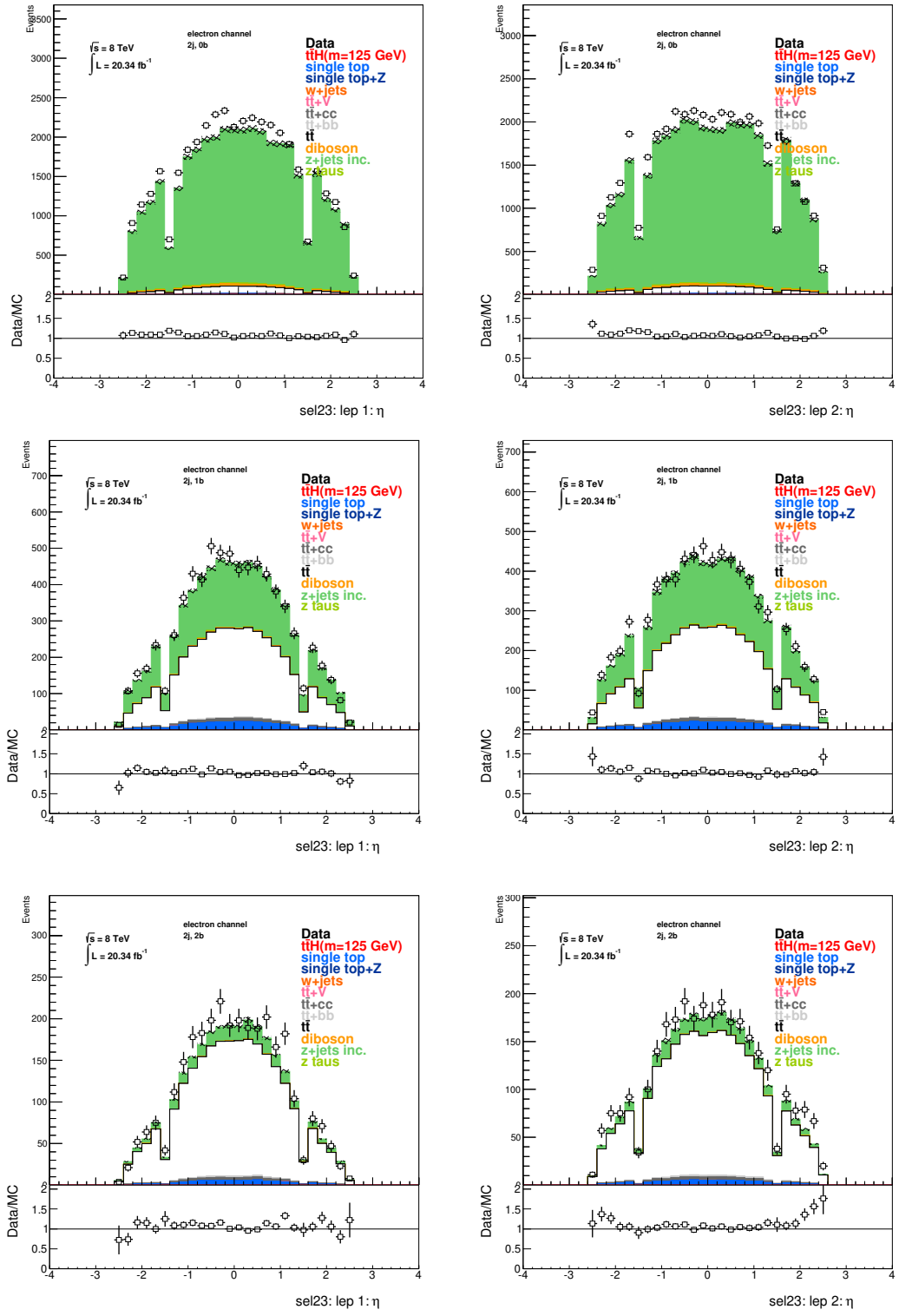
The impact of the complete Z+jets data driven correction can be seen in Figures 6.13–6.17 for electrons and Figures 6.18–6.22 for muons for several kinematic distributions of the two jet regions. The Z+jets MC description of the data was improved in all regions of interest, having directly corrected the bad parametrisation of leptons and jets alike.

In conclusion, the Z+jets data driven correction implemented in this thesis was a three step method developed to minimize the impact of systematics and lack of statistics in regions with higher multiplicity of b-tag jets. The rates of Z events were corrected from data in control regions completely orthogonal to the signal regions of the analysis. In each $(m\text{j}, n\text{b})$ multiplicity region, the cross-section normalisation compensates the total Z+jets scales, independently of the true type of jet simulated in the MC production. The normalisation correction accounts for differences seen in dilepton and dimuon events, by computing different corrections for each of the leptonic channels. Additional transverse momentum corrections were performed both for leptons and jets. The average lepton pair transverse momentum and the average leading jet pair transverse momentum were corrected to account for the issues in the MC radiation parametrisation. A single parameter fit was implemented to determine inclusively over all multiplicity regions the slope of the correction function in both the lepton and jet corrections. Having both the lepton and jet momentum corrected brings a much better description of all the kinematic distributions.



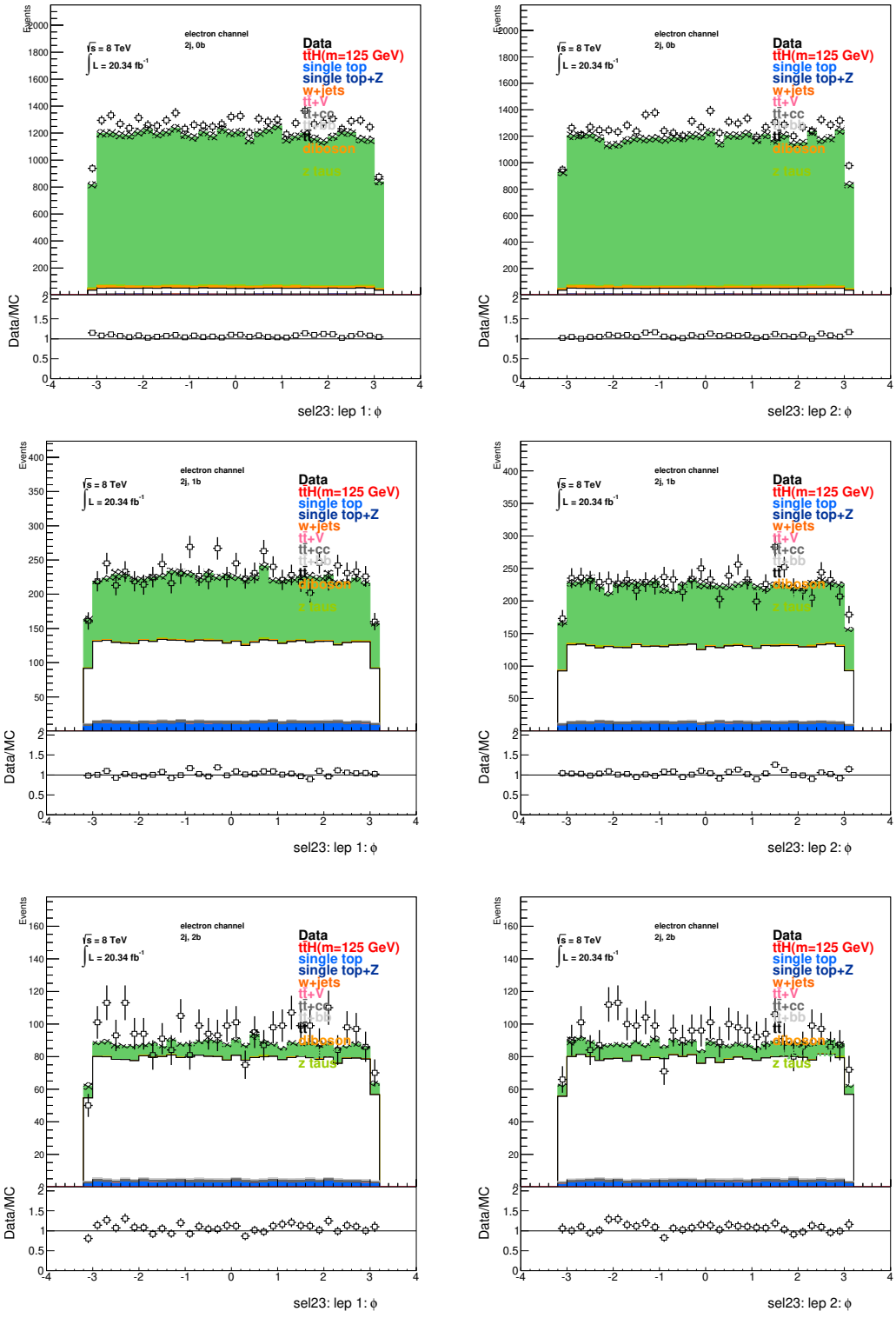
(a) (b)

Figure 6.13: Transverse momentum of the leading (a) and subleading (b) lepton in the two jet region of the dielectron channel after the full Z+jets data driven correction. The error bars associated to the data points include only the statistical uncertainties on the measurement.



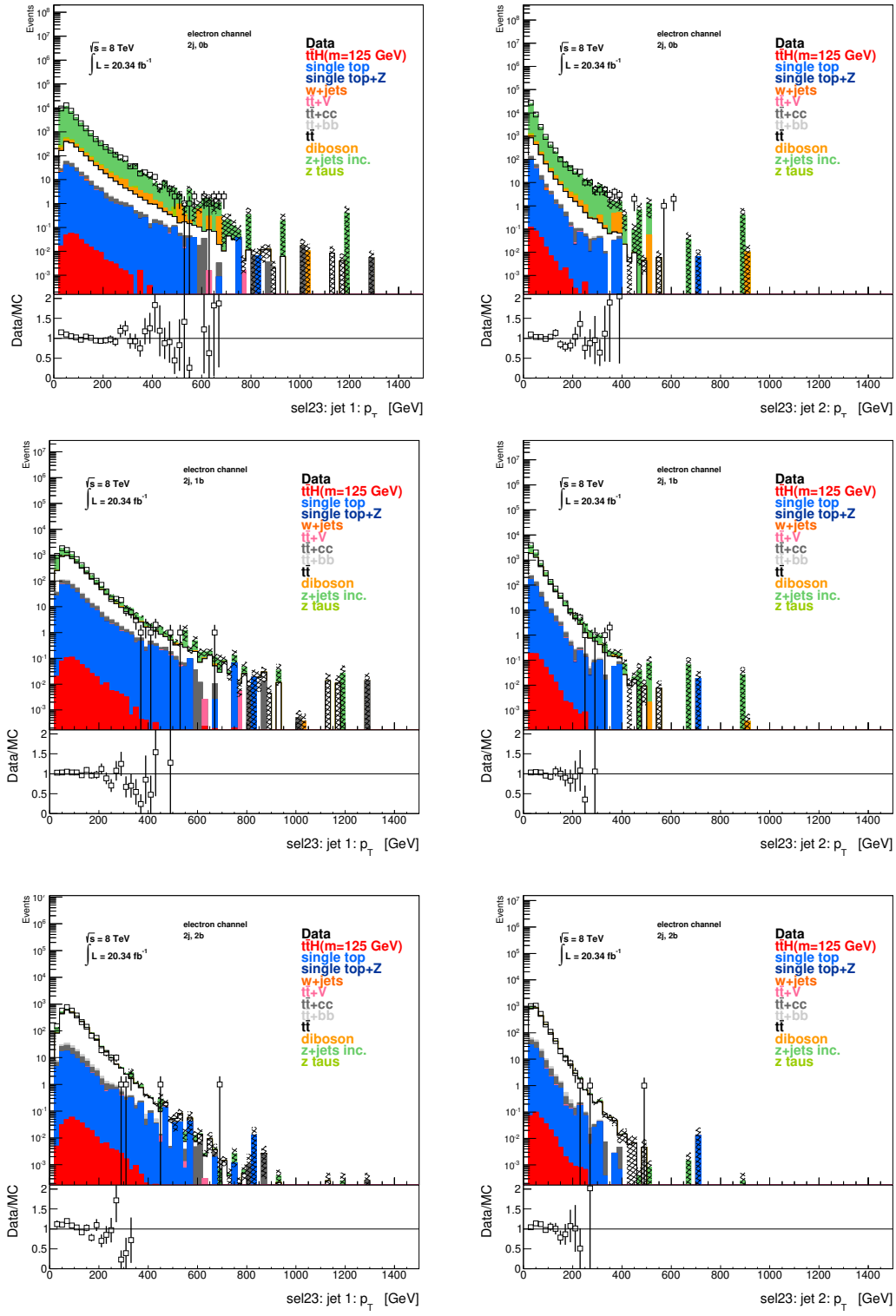
(a) (b)

Figure 6.14: Pseudorapidity, η, of the leading (a) and subleading (b) lepton in the two jet region of the dielectron channel after the full Z+jets data driven correction. The error bars associated to the data points include only the statistical uncertainties on the measurement.



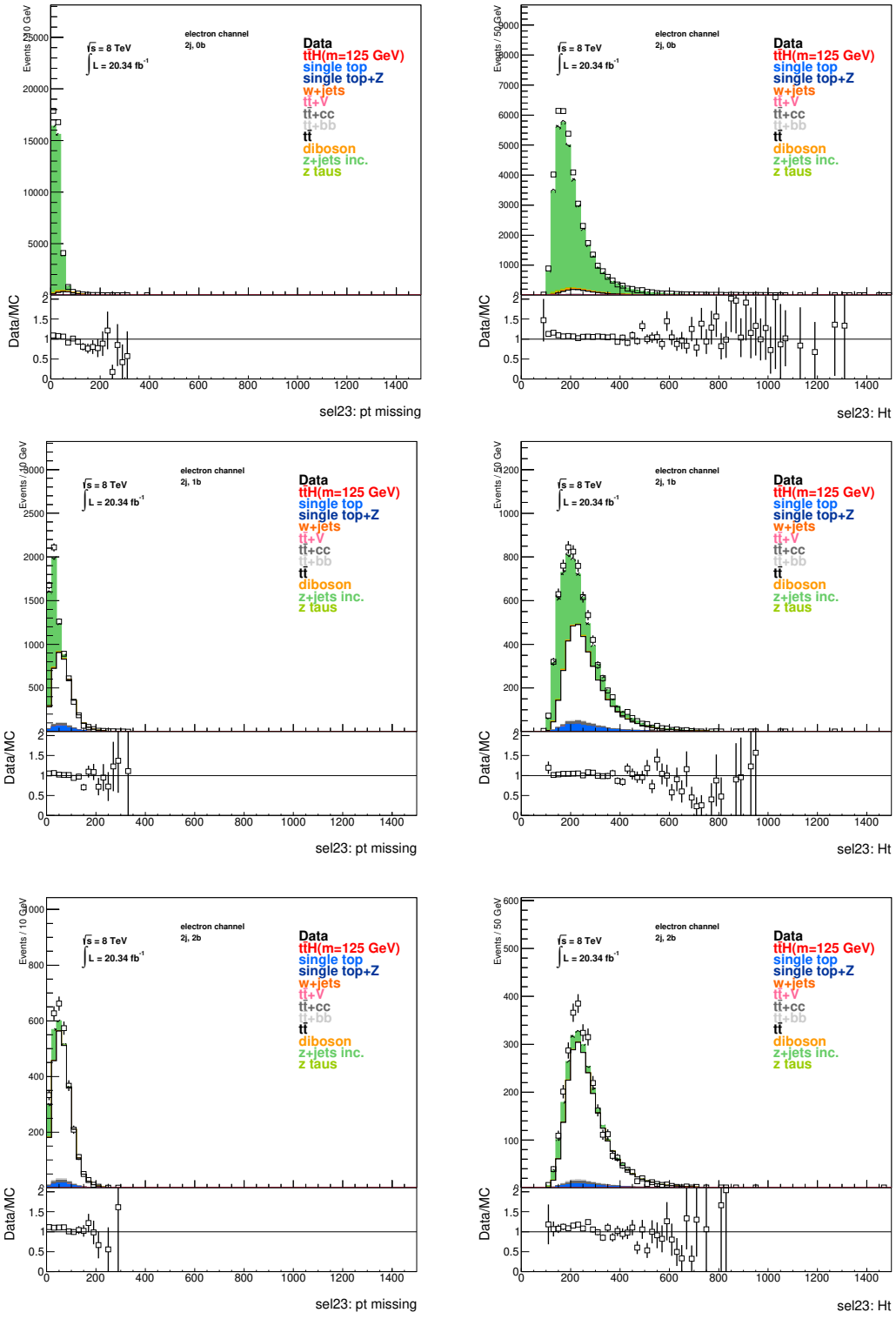
(a) (b)

Figure 6.15: Azimuthal angle, ϕ, of the leading (a) and subleading (b) lepton in the two jet region of the dielectron channel after the full Z+jets data driven correction. The error bars associated to the data points include only the statistical uncertainties on the measurement.



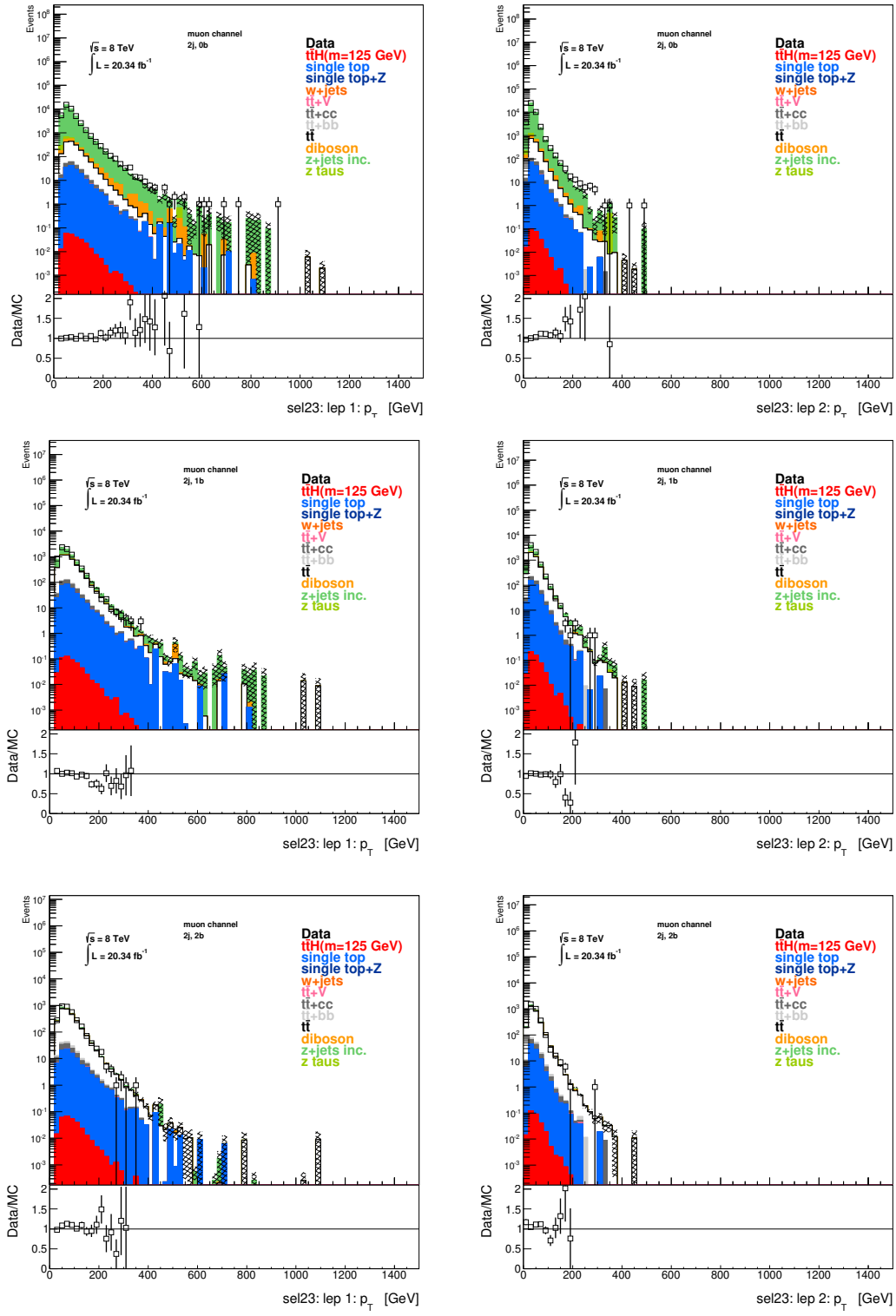
(a) (b)

Figure 6.16: Transverse momentum of the leading (a) and subleanding (b) jet in the two jet region of the dielectron channel after the full Z+jets data driven correction. The error bars associated to the data points include only the statistical uncertainties on the measurement.



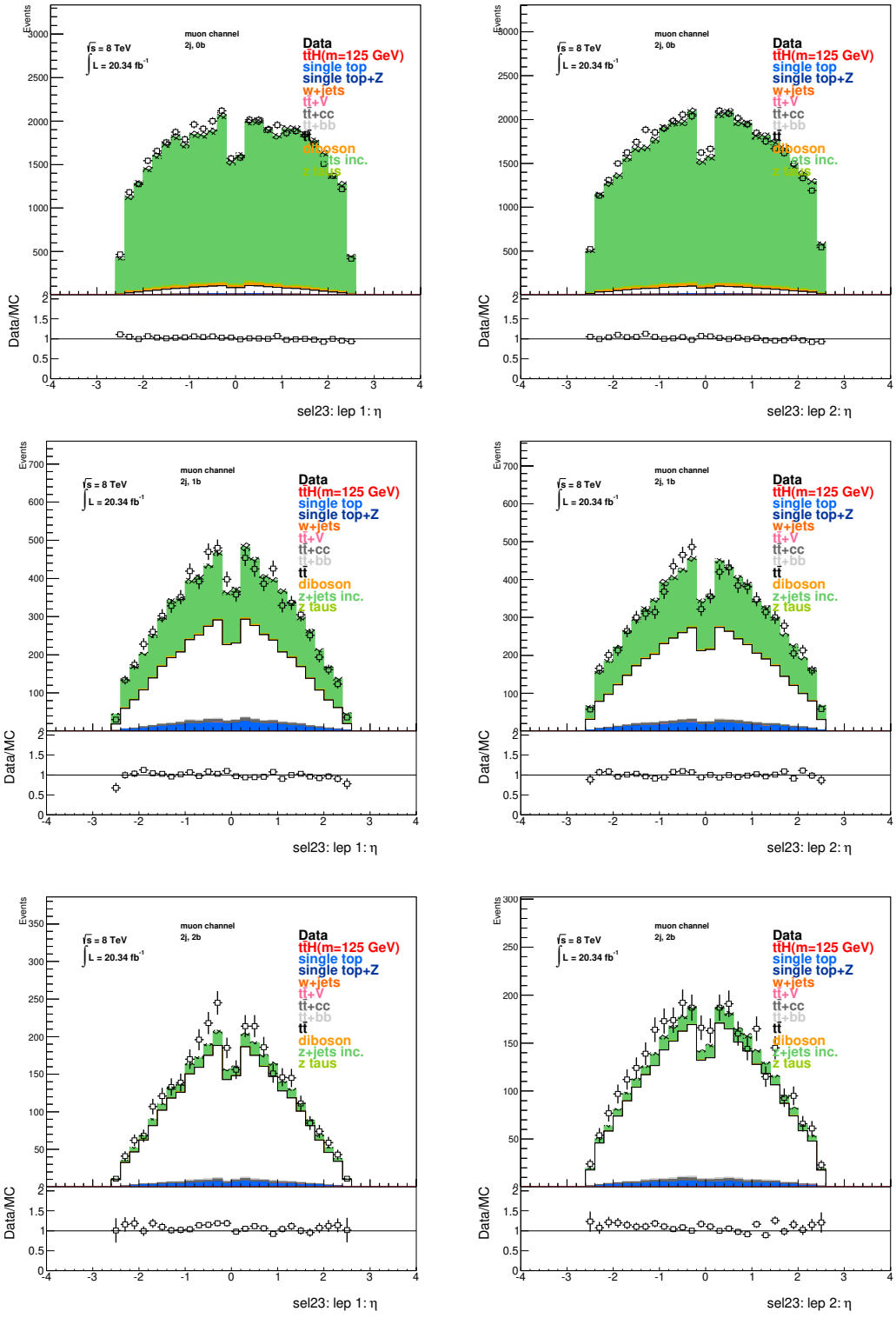
(a) (b)

Figure 6.17: Missing transverse momentum (a) and H_T (b) in the two jet region of the di-electron channel after the full Z+jets data driven correction. The error bars associated to the data points include only the statistical uncertainties on the measurement.



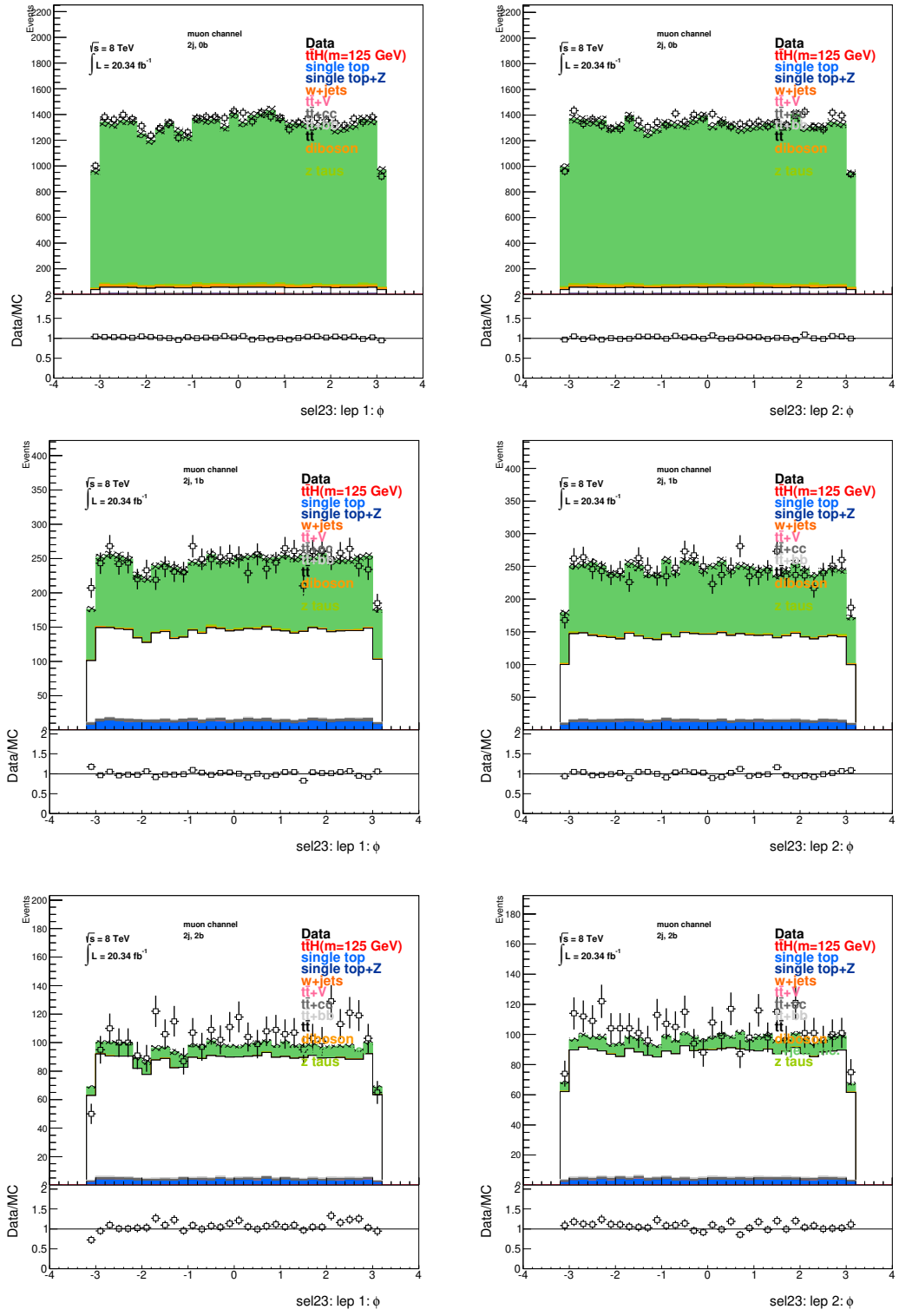
(a) (b)

Figure 6.18: Transverse momentum of the leading (a) and subleading (b) lepton in the two jet region of the dimuon channel after the full Z+jets data driven correction. The error bars associated to the data points include only the statistical uncertainties on the measurement.



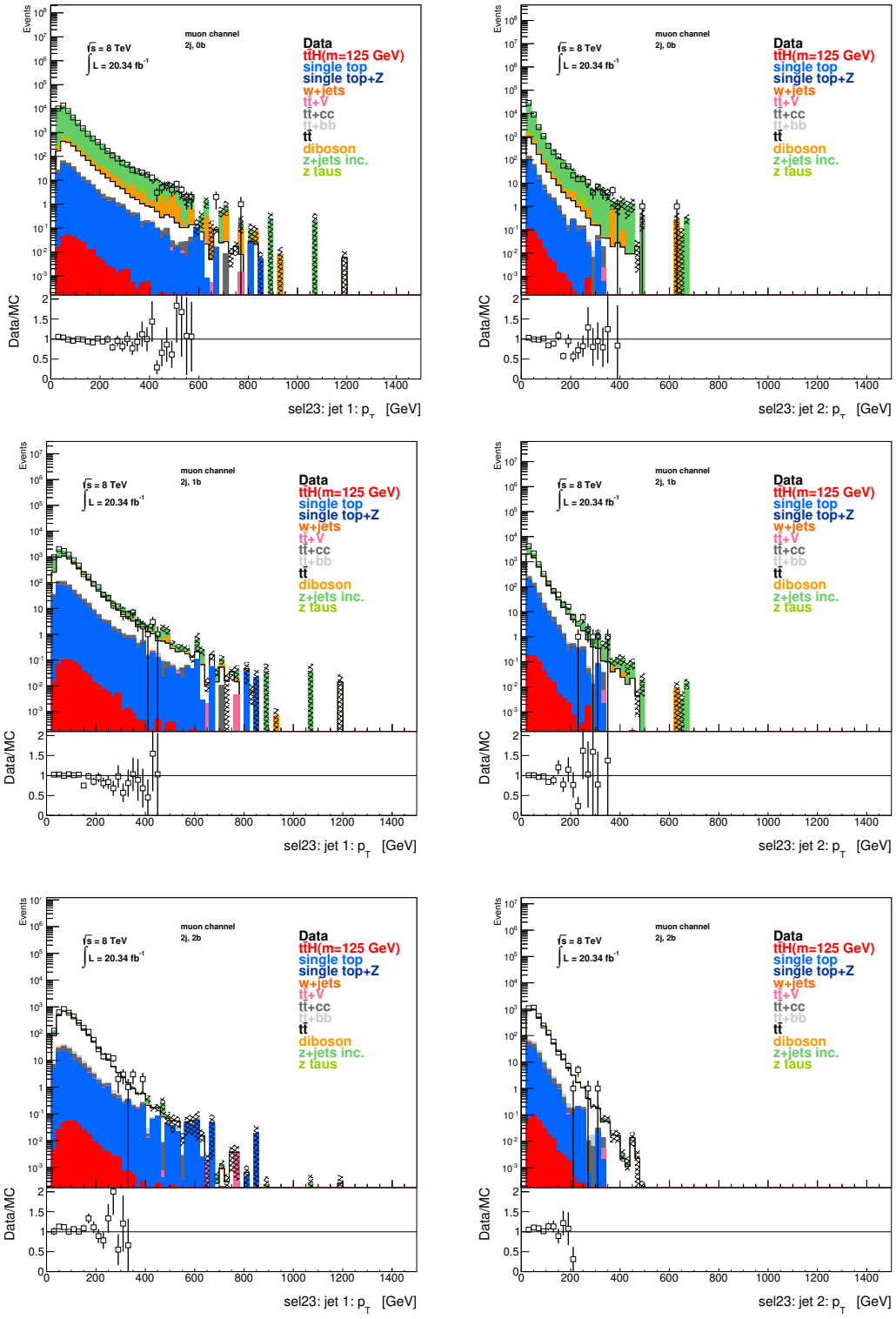
(a) (b)

Figure 6.19: Pseudorapidity, η, of the leading (a) and subleading (b) lepton in the two jet region of the dimuon channel after the full Z+jets data driven correction. The error bars associated to the data points include only the statistical uncertainties on the measurement.



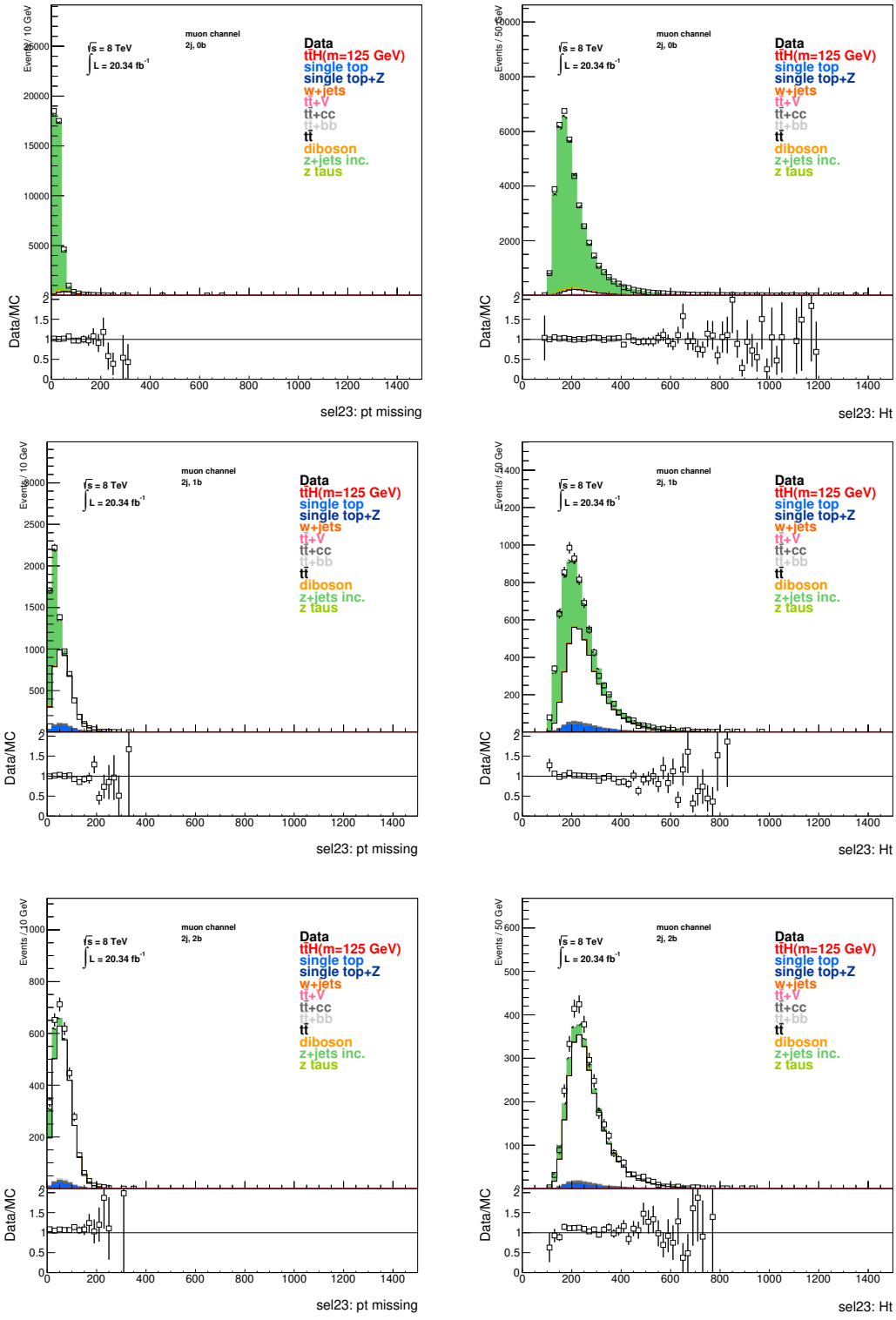
(a) (b)

Figure 6.20: Azimuthal angle, ϕ, of the leading (a) and subleading (b) lepton in the two jet region of the dimuon channel after the full Z+jets data driven correction. The error bars associated to the data points include only the statistical uncertainties on the measurement.



(a) (b)

Figure 6.21: Transverse momentum of the leading (a) and subleanding (b) jet in the two jet region of the dimuon channel after the full Z+jets data driven correction. The error bars associated to the data points include only the statistical uncertainties on the measurement.



(a) (b)

Figure 6.22: Missing transverse momentum (a) and H_T (b) in the two jet region of the dimuon channel after the full Z+jets data driven correction. The error bars associated to the data points include only the statistical uncertainties on the measurement.

6.2.2 $t\bar{t}$ + jets Data Driven Estimation

A key aspect of the $t\bar{t}H$ analysis rests in achieving the best possible modelling of the $t\bar{t}$+jets background. After event preselection it is not only the most dominant background, but it is also an irreducible one, due to the similar final state objects as the $t\bar{t}H$ signal. The $t\bar{t}$+jets description in the simulated samples must be the most accurate possible, particularly in the emission of additional jets and their heavy flavour fraction.

The MC simulation available at 8 TeV was comprehensively compared with the $t\bar{t}$ events produced at the LHC in all ATLAS analysis involving the top quark. The large phase space covered by all the analyses led to the conclusion that the MC prediction for most generators, and in particular for POWHEG+PYTHIA, over predicts the data at high transverse momentum.

In order to correct this feature, several methods were developed. For the purpose of this thesis, a simple approach based on jet efficiencies and basic event rates was implemented. A reweighting scale factor was calculated from events in dedicated control regions and used to correct inclusively the $t\bar{t}+light$, $t\bar{t}+c\bar{c}$ and $t\bar{t}+b\bar{b}$ samples in the signal regions.

The $t\bar{t}$ + jets correction method was estimated considering only $e\mu$ events, purest in $t\bar{t}$ + jets, and maintaining the jet multiplicity bins as independent regions, just as in the nominal analysis. For each particular jet combination the number of events that survive the $e\mu$ channel selection, $N^j_{e\mu}$, can be determined by:

$$N^j_{e\mu} = L\,\sigma_{t\bar{t}}\epsilon^j_{e\mu} \tag{6.7}$$

where L is the integrated luminosity of the sample, $\sigma_{t\bar{t}}$ is the inclusive $t\bar{t}$ + jets production cross-section and $\epsilon^j_{e\mu}$ is the efficiency for a $t\bar{t}$ event to pass the opposite-sign $e\mu$ selection. Further separation according to b-tag jet multiplicity is achieved in both data and MC events by considering the b-tag efficiencies ϵ_b^{D} and ϵ_b^{MC} respectively. For example, for events with exactly 2 jets, the possible b-tag combinations result in:

$$\begin{aligned} N_{2j\,0b} &= N^{2j}_{e\mu}\,(1-\epsilon_b)^2 \ , \\ N_{2j\,1b} &= N^{2j}_{e\mu}\,2\epsilon_b(1-\epsilon_b) \ , \\ N_{2j\,2b} &= N^{2j}_{e\mu}\,\epsilon_b^2 \ , \end{aligned} \tag{6.8}$$

where ϵ_b is the b-tagged jet efficiency of either data or MC depending on the event sample. The number of selected events in each region can thus be expressed in terms of the possible b-tagged jet combination and depends only on the b-tag jet efficiency.

Scale factors, k_{jb}, to estimate the $t\bar{t}+$jets events from data can be determined similarly to what was previously done for the Z+jets samples:

$$\begin{aligned} \text{Estimate}_{jb}^{t\bar{t}+jets\ (\text{SR})} &= k_{jb}^{(\text{CR})} \times \text{MC}_{jb}^{t\bar{t}+jets\ (\text{SR})} \\ &= \left.\frac{N_{j\,b}^{D}}{N_{j\,b}^{MC}}\right|^{(\text{CR})} \times \text{MC}_{jb}^{t\bar{t}+jets\ (\text{SR})} , \end{aligned} \tag{6.9}$$

where $N_{j\,b}^{D}$ ($N_{j\,b}^{MC}$) is the number of $t\bar{t}+$jets events in data (MC) samples on the bin with j jets and b b-tagged jets. For the determination of data $t\bar{t}+$jets events, just as in the $Z+$jets data driven study, the subtraction of the MC prediction for all other background processes is performed. By counting data events in dedicated CR, the correction scale factors can be calculated and posteriorly used to normalise the MC events in the SR. For each jet and b-tagged jet combination, a k_{jb} scale factor can be estimated from the ratio of data and MC events in that region. So for example in regions with exactly 2 jets we have:

$$\begin{aligned} k_{2j\,0b} &= \frac{N_{2j\,0b}^{D}}{N_{2j\,0b}^{MC}} = \frac{N_{e\mu}^{2j\,D}\,(1-\epsilon_b^{D})^2}{N_{e\mu}^{2j\,MC}\,(1-\epsilon_b^{MC})^2} \\ k_{2j\,1b} &= \frac{N_{2j\,1b}^{D}}{N_{2j\,1b}^{MC}} = \frac{N_{e\mu}^{2j\,D}\,2\epsilon_b^{D}(1-\epsilon_b^{D})}{N_{e\mu}^{2j\,MC}\,2\epsilon_b^{MC}(1-\epsilon_b^{MC})} \\ k_{2j\,2b} &= \frac{N_{2j\,2b}^{D}}{N_{2j\,2b}^{MC}} = \frac{N_{e\mu}^{2j\,D}\,\epsilon_b^{D\,2}}{N_{e\mu}^{2j\,MC}\,\epsilon_b^{MC\,2}} \ . \end{aligned} \tag{6.10}$$

In theory one could write 12 k_{jb} factors, one per jet and b-tagged jet combination of the analysis. These, however, should not be directly calculated from events in the SR where the correction is of the most interest. Therefore an indirect way of determining these scale factors was created.

From the previous set of equations it can be demonstrated that the ratio of $k_{j\,b}$ for consecutive b-tag multiplicity regions is constant:

$$R_{j\,b+1;j\,b} = \frac{k_{j\,b+1}}{k_{j\,b}} = \frac{\epsilon_b^{\text{D}}(1-\epsilon_b^{\text{MC}})}{\epsilon_b^{\text{MC}}(1-\epsilon_b^{\text{D}})} \ . \tag{6.11}$$

	2 jets	**3 jets**	**≥4 jets**
0 *b*-tagged jets	CR	CR	CR
1 *b*-tagged jets	CR	CR	CR
2 *b*-tagged jets	CR	SR	SR
3 *b*-tagged jets		SR	SR
≥4 *b*-tagged jets			SR

Table 6.3: Control (CR) and signal regions (SR) of the $t\bar{t}$ + jets data driven calculation.

Furthermore, this ratio $R_{j\,b+1;j\,b}$ depends only on the b-tag efficiencies of data and MC. Indirect calculation of the necessary $k_{j\,b}$ to correct the SR can thus be obtained by using regions with low b-tagged jet multiplicity, which aren't included in the analysis:

$$k_{j\,b+1} = R_{j\,b+1;j\,b} \times k_{j\,b} \,. \tag{6.12}$$

Scale factors can then be determined using only information from signal depleted bins (CR). Considering the $t\bar{t}H$ signal topology, reflected in the implemented selection of events, only regions with at least 2 b-tagged jets are considered regions of interest. Table 6.3 shows the control and signal regions considered for this correction. As compromise in the $t\bar{t}$ + jets purity and statistical significance, the ratio $R_{2\,2;2\,1} = 1.056 \pm 0.015$, between scale factors $k_{2\,2}$ and $k_{2\,1}$, was used to calculate all other scale factors. For each specific jet multiplicity, all scale factors were recursively computed starting from the directly measured $k_{j\,0}$ of regions without b-tagged jets. That is:

$$\begin{aligned} k_{j\,1} &= R_{2\,2+1;2\,1} \times k_{j\,0} \\ k_{j\,2} &= R_{2\,2+1;2\,1} \times k_{j\,1} \\ &\ldots \\ k_{j\,b+1} &= R_{2\,2+1;2\,1} \times k_{j\,b} \end{aligned}$$

In Table 6.4 the obtained scale factors are shown for each jet and b-tagged jet region. The renormalisation scale factors seem to increase with the b-tagged jet number of the region, which is concurrent with what is expected. These regions rich in b-tagged jets are populated by a significant heavy flavoured component of the $t\bar{t}$, which were only predicted to LL accuracy by POWHEG+PYTHIA generator [1] [1]. It is therefore expected a slightly higher

[1] POWHEG includes only $gb \to t\bar{t}b$ diagrams, thus the production of $b\bar{b}$ pairs is done by PYTHIA at the parton shower.

Region (j,b)	$k_{j\,b}$
(2j,1b)	1.03 ± 0.03
(2j,2b)	1.09 ± 0.03
(3j,1b)	1.04 ± 0.03
(3j,2b)	1.09 ± 0.04
(3j,3b)	1.15 ± 0.04
(4j,1b)	0.96 ± 0.04
(4j,2b)	1.02 ± 0.04
(4j,3b)	1.08 ± 0.05
(4j,≥4b)	1.14 ± 0.05

Table 6.4: The $t\bar{t}+$jets renormalisation scale factors, $k_{j\,b}$, for all $(m\text{j}, n\text{b})$ multiplicity bins of interest in the analysis.

mis-modelling of the $t\bar{t}+$HF component, resulting in an overall increased mis-prediction of the inclusive $t\bar{t}+$jets sample in these bins of the analysis.

Determined from $e\mu$ events, the $t\bar{t}+$jets scale factors summarised in Table 6.4 were used to compensate the mis-prediction of the MC events in SR for all lepton channels. In Figure 6.23 some kinematic distributions are shown where the impact of the data driven correction can be seen for $e\mu$ events in the 3 jet, 2 b-tagged jet region, which are dominated by the $t\bar{t}$+HF. As can be seen, the inclusive $t\bar{t}$+jets cross-section normalisation compensates the total scales of the $t\bar{t}$ events. A slight over prediction at high transverse momentum is still present. Further correction was considered using a similar method to that applied in the $Z+$jets events. That is, performing linear fits to the data/MC ratio of the average jet transverse momentum of events without b-tagged jets and using the determined slope to correct SR events. However, this was found to bring no significant impact on the analysis and only add to harsher systematic errors. As such and for the purpose of this study, only the data driven renormalisation was applied.

In summary, the $t\bar{t}+$jets cross-section was corrected to data inclusively for all extra jet flavour. In each $(m\text{j}, n\text{b})$ multiplicity bin, the total scale of events was compensated using correction factors determined recursively from CR orthogonal to the analysis. Determined from $e\mu$ events, which are the purest in $t\bar{t}$ events, and using CR with no b-tagged jet, the scale factors were applied in the SR of the analysis for all lepton channels. This simple correction method effectively deals with the lack of statistics in regions with higher multiplicity of b-tagged jets, while simultaneously minimizing the sources of systematic uncertainties. The discrimination in both jet number and flavour was of major importance since the corrections

differs significantly through out the bins of the analysis. Additional transverse momentum correction was tested and found to be too harsh. After applying the data driven correction to $t\bar{t}$ + jets simulated events, the effect of the re-normalisation on the agreement between data and MC simulation improves considerably in all kinematic distributions of the various objects.

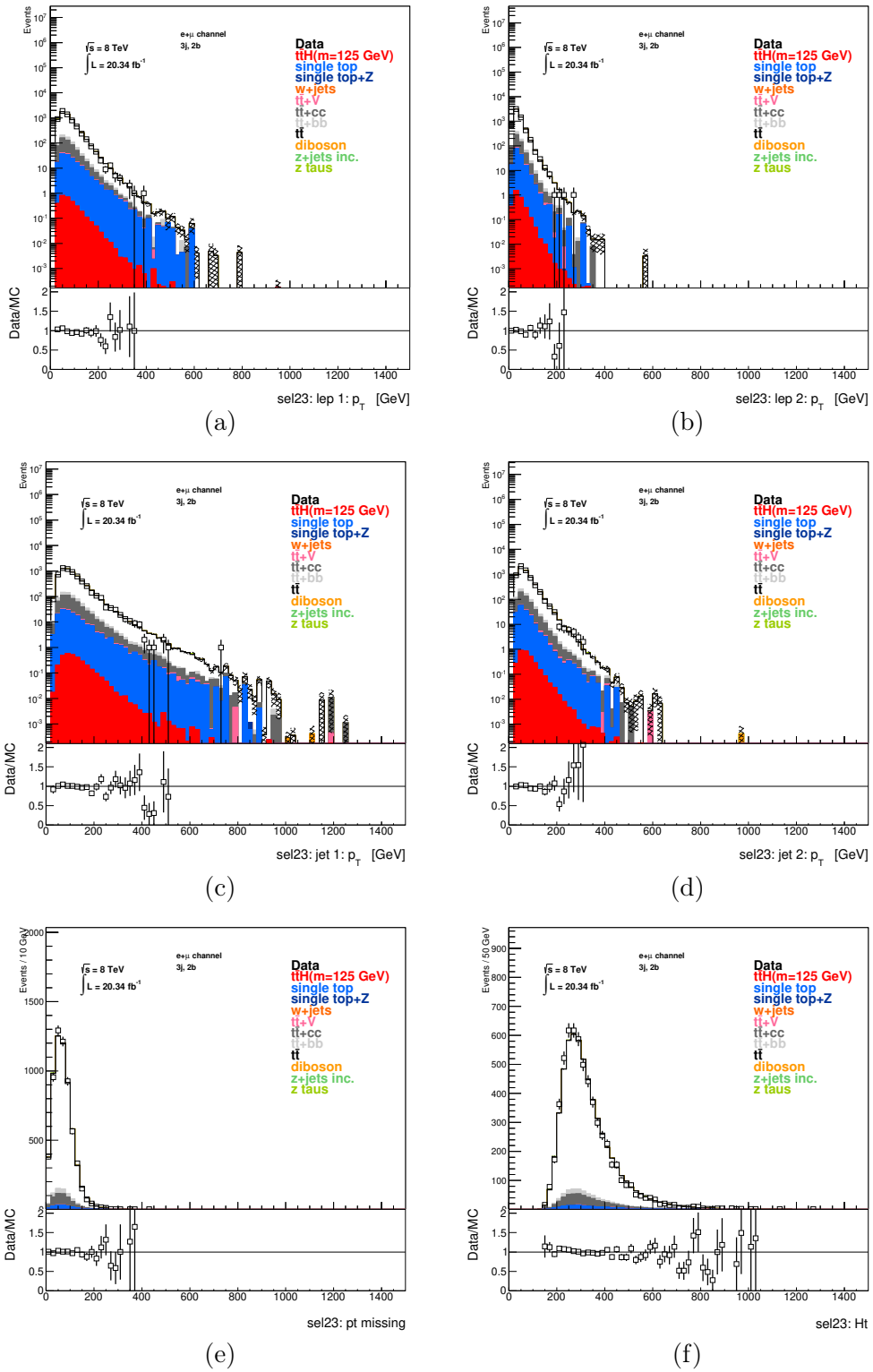


Figure 6.23: Kinematic distributions after the $t\bar{t}$ + jets data driven correction for the $e\mu$ channel: p_{T} of the leading (a) and sub-leading (b) lepton; p_{T} of the leading (c) and sub-leading (d) jet; missing transverse momentum (e) and H_T (f).

6.2.3 Misidentified Lepton Data Driven Estimation

The dileptonic $t\bar{t}H$ analysis relies heavily on the identification of exactly two isolated leptons with opposite electric charge. However, non-prompt leptons or non-leptonic particles may satisfy the selection criteria, giving rise to non-prompt or fake lepton backgrounds. Fake electrons can occur from the misidentification of semileptonic decays of b- and c-quarks, photon conversions and jets with large electromagnetic energy. In the muon case, fakes can also arise from the semileptonic decays of b- and c-quarks, but also from charged hadron decays either in the tracking volume or in hadronic showers, or from punch-through particles emerging from high-energy hadronic showers.

In this analysis, most backgrounds include the presence of prompt leptons as final state objects. Processes containing at least two prompt leptons include the $t\bar{t}$ + jets, the Z + jets, the diboson, the $Wt-$ channel of the single top quark production and the $t\bar{t}+V$. Backgrounds that can contain non-prompt leptons passing the lepton isolation requirements or jets misidentified as leptons are the W+jets, the $t\bar{t}$+jets with one final state prompt lepton and the single top quark production in both $t-$ and $s-$channels. Using MC simulation, the previous processes are estimated including their fake lepton contribution. Background processes with no prompt lepton involved can still enter the selected data sample via the lepton misidentification. In this analysis, the major source of such backgrounds are the multijet events, originating from QCD processes with huge production cross-sections. Being quite difficult to accurately model with MC simulation, several data-driven methods have been developed for its estimation in the final selection sample. A simple approach based on the selection of dileptonic events with the same sign electric charge was developed for the purpose of this thesis.

In order to estimate the contribution of fake leptons in the analysis, the same final kinematic selection criteria as in the nominal analysis is applied. Instead of the opposite sign requirement on the charge of the leptons, here exactly two same electric charge leptons are requested, creating two orthogonal regions: the opposite sign (OS) and the same sign (SS). Assuming equal probability of non-isolated leptons or misidentification of non-leptons to occur in both the same sign and the opposite sign dileptonic regions, a fake lepton estimation can be obtained:

$$Fakes = N^{\mathrm{SS}}_{\mathrm{Data}} - N^{\mathrm{SS}}_{\mathrm{MC}} \ . \tag{6.13}$$

The removal of all MC contributions with same sign lepton pairs, $N_{\mathrm{MC}}^{\mathrm{SS}}$, assures no double counting of misidentified lepton events occurs.

Unfortunately, the same sign dilepton selection has very limited statistics, particularly in high b-tag and jet multiplicity bins. Therefore, all b-tag regions were considered inclusively, compromising between increasing the available statistics and maintaining jet discrimination. The contribution of fake leptons was estimated in all inclusive b-tag regions and propagated to each specific b-tag bin using the MC simulation behaviour as transfer functions:

$$Fakes_{j\,b} = \left[N_{\mathrm{data}}^{\mathrm{SS}} - N_{\mathrm{MC}}^{\mathrm{SS}}\right]_{j\,\mathrm{All\,btag\,Inc.}} \times \frac{N_{\mathrm{MC}\,j\,b}^{\mathrm{SS}}}{N_{\mathrm{MC}\,j\,\mathrm{All\,btag\,Inc.}}^{\mathrm{SS}}} \ . \tag{6.14}$$

Here, $N_{\mathrm{data}}^{\mathrm{SS}}$ is the selected same sign data and $N_{\mathrm{MC}}^{\mathrm{SS}}$ the contribution of all MC backgrounds in the same sign region. The transfer functions $N_{\mathrm{MC}\,j\,b}^{\mathrm{SS}}/N_{\mathrm{MC}\,j\,\mathrm{All\,btag\,Inc.}}^{\mathrm{SS}}$ are also determined from same sign events and depend both on jet and b-tagged jet multiplicity. These are exclusively determined from MC events and provide a modulation rate for the final fake lepton yield. The developed method to determine the fake lepton contamination is particularly useful, since both the shape and normalisation are directly obtained from SS data, whose information would be otherwise discarded in this analysis. In order to improve the statistical uncertainties on the fake lepton background, all lepton combinations were merged in a single channel of analysis. From here onwards, all lepton channels will be considered inclusively.

Figure 6.24 shows several kinematic distributions, in the two jets and two b-tagged jets region, including the estimate of the fake lepton background in yellow. The statistical uncertainty on the misidentified lepton estimation was obtained via error propagation. The fake lepton background amounts to 0.2% in the low multiplicity regions and varies up to 0.7% in high b-tag and jet multiplicity bins. After the fake estimation, the data and MC agreement is quite good.

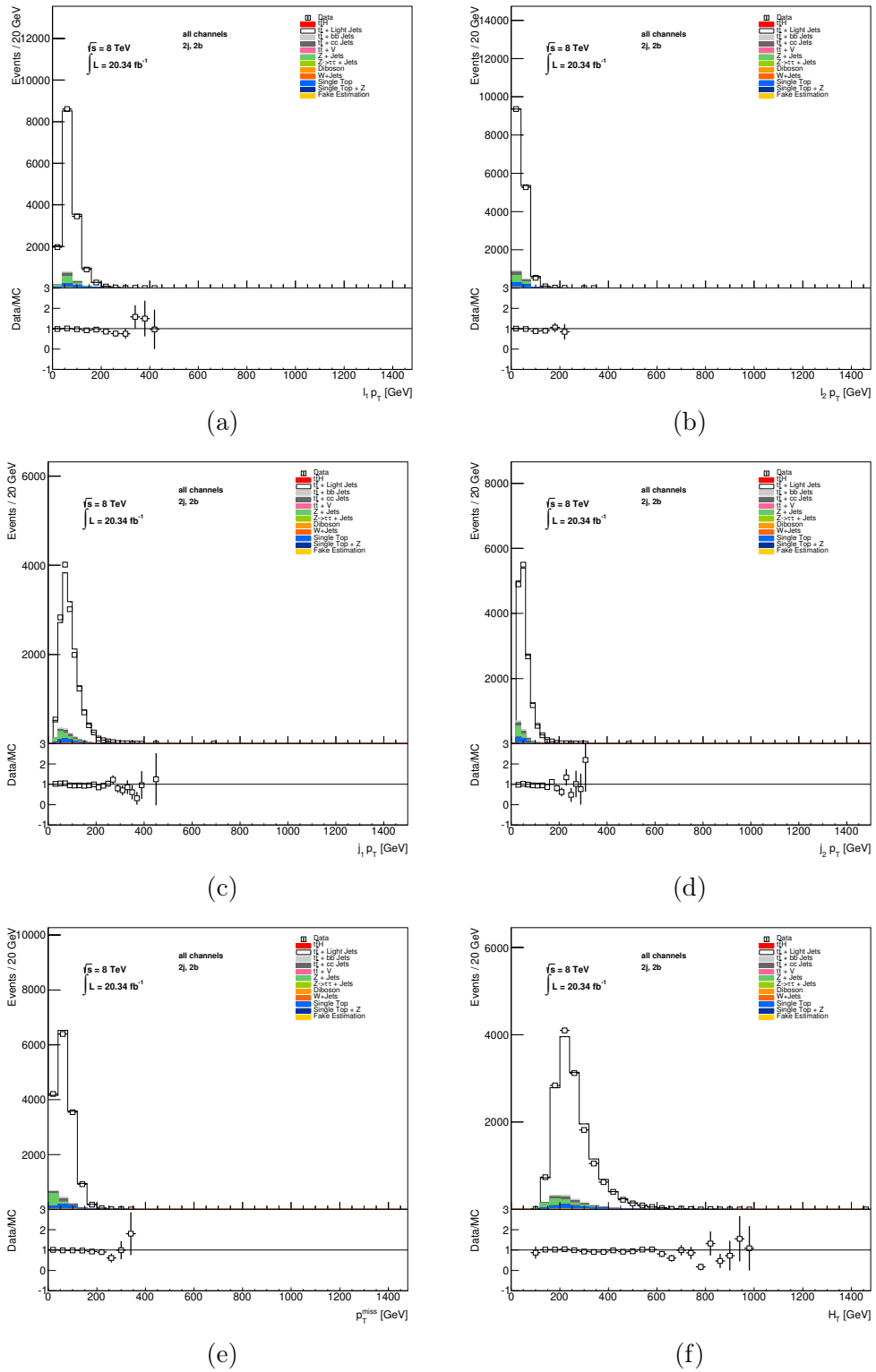


Figure 6.24: Kinematic distributions including the fake estimation in yellow, in the 2 jet, 2 b-taggged jet bin: p_{T} of the leading (a) and sub-leading (b) lepton; p_{T} of the leading (c) and sub-leading (d) jet; missing transverse momentum (e) and H_T (f).

7

Statistical Analysis and Results

In this chapter, the search of the SM Higgs boson in the dileptonic channel of $t\bar{t}H$ production is presented. The implemented methodology is described in detail in Section 7.1, including a brief account of the statistical tools used and the sources of systematic uncertainty considered in the analysis. Section 7.2 details the discriminant analysis performed. The results are discussed in Sections 7.3 and 7.4, where a combined likelihood fit performed to five regions of the analysis is presented and a limit on the $\sigma_{t\bar{t}H} \times BR(dileptonic)$ is set at 95% CL, respectively. The work described in this chapter was also included in [1], where a combined limit on the $t\bar{t}H$ cross-section is set together with the analysis results of the semi-leptonic decay channel.

7.1 Analysis Method

After applying the preselection requirements (described in Section 5.5) and background corrections (described in Section 6.2), the selected events, which are classified in six independent $(m\text{j}, n\text{b})$ regions, are used to construct discriminant variables that provide the highest separation between signal and background. The analysis of the dileptonic $t\bar{t}H$ search is optimised considering the selection criteria in terms of signal-to-background ratio. The dileptonic ee, $e\mu$ and $\mu\mu$ channels are, from here onwards, coherently combined and treated as a single dileptonic selection.

Given the $t\bar{t}H$ topology, signal events will mostly populate high jet and b-tag multiplicity regions. In Figure 7.1, the signal-to-background ratio, S/B, and the signal significance, $S/\sqrt{B}$, are shown for all considered regions, where S and B are the expected $t\bar{t}H$ signal for a SM Higgs boson and background, respectively. Regions with $S/B > 1\%$ and $S/\sqrt{B} > 0.3$

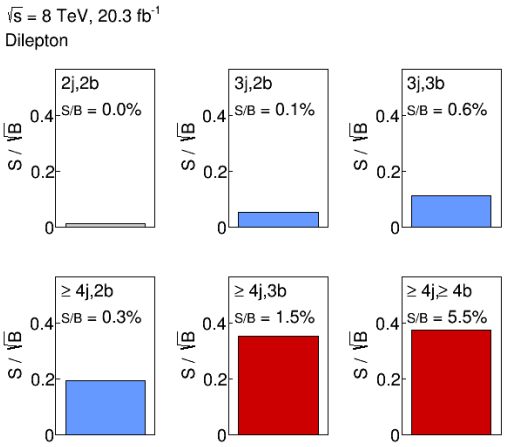


Figure 7.1: Pre-fit $S/\sqrt{B}$ ratio for each of the $(m\text{j}, n\text{b})$ regions, assuming SM cross-sections, branching fractions and $m_H = 125$ GeV. Signal-rich regions are represented in dark red, while the signal-depleted ones are shown in light blue. Each region also includes the respective S/B ratio.

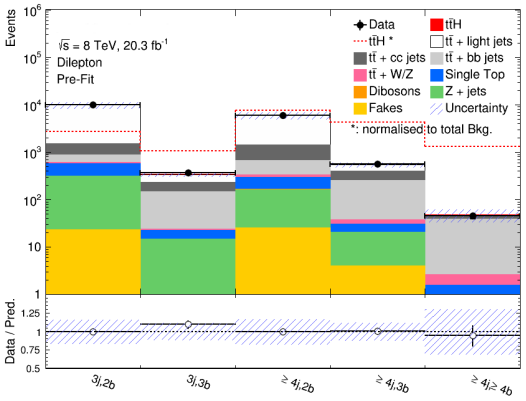


Figure 7.2: Summary of the pre-fit MC prediction to data comparison in all analysis regions. The signal is shown in red, both normalised to the SM prediction (filled red area stacked on the background) and normalised to the total number of background events (dashed red line). The total uncertainty on the yields is represented by the hashed area.

provide the highest sensitivity to the signal and are appropriately referred to as signal-rich regions. In this analysis, the most signal sensitive regions are ($\geq$ 4j, 3b) and ($\geq$ 4j, $\geq$ 4b). The overall signal contribution is expected to be approximately 90% $H \to b\bar{b}$ decays in the most sensitive region ($\geq$ 4j, $\geq$ 4b), as depicted in Figure 6.1. The remainder multiplicity regions, (2j, 2b), (3j, 2b), (3j, 3b) and ($\geq$ 4j, 2b), are considered signal-depleted regions, being mostly populated by background events.

To inquiry the presence or absence of the $t\bar{t}H$ process, a combined profile likelihood fit to both signal-rich and signal-depleted regions will be performed, based on powerful discriminant observables that are suited to deal with signal extraction in such complex final states. The signal-depleted regions are quite useful in constraining systematic uncertainties and background normalisations, which will thus allow better background prediction in signal-rich regions. The (2j, 2b) region, which has been used as an orthogonal control region for the $t\bar{t}+$ jets background in this analysis (*c.f.* Section 6.2.2), is, however, not included in the combined fit, being present here for validation reasons only. The detailed description of the combined fit and statistical analysis is carried out in Section 7.3.

A comparison between the MC predicted yields and data is shown in Figure 7.2 for the analysis regions included in the statistical fit. The $t\bar{t}H$ signal is included normalised to the SM prediction and stacked on the background (filled red area), as well as, normalised to the total number of background events (dashed red line). The black hashed area represents the total uncertainty on the yields, before performing the combined fit (labelled pre-fit). The SM expectation accurately describes the data within 10–30% uncertainty. Table 7.1 shows the detailed event yields for each region, including the validation region (2j, 2b). The quoted uncertainties include the MC statistical and systematic uncertainties. The sources of systematic uncertainty are explained in Section 7.1.2 below.

7.1.1 Tag Rate Function Method

In an analysis that requires such a high number of jets, in particular b-tagged jets, the available MC statistics gets drastically reduced. In the signal regions, where the jet and b-tagged jet multiplicities are the highest, having scarcity of events can result in large fluctuations in the distributions for certain samples. The sensitivity of the analysis, which relies on the shape information in each of the event categories, can be jeopardised by large statistical uncertainties on the templates or unreliable systematic uncertainties given the fluctuations in the shape of the distribution. To attenuate this effect, the tag rate function (TRF) method is used in this analysis.

The TRF method is an alternative way to b-tag jets in MC simulated events. Rather than tagging jets by applying the MV1 b-tagging algorithm, this method assigns a probability (or weight) of being b-tagged to each event based on parameterised efficiencies. By computing the probability of each event to have a specific number of b-jets, the TRF does not cut events according to the required b-tag multiplicity, instead all the events present in a sample

	2j, 2b	**3j, 2b**	**3j, 3b**
$t\bar{t}H$	1.51 ± 0.246	5.32 ± 0.419	2.08 ± 0.236
$t\bar{t}$ + light jets	14 000 ± 1950	8500 ± 1390	109 ± 29.1
$t\bar{t} + c\bar{c}$ jets	278 ± 186	647 ± 244	83.0 ± 63.2
$t\bar{t} + b\bar{b}$ jets	147 ± 116	284 ± 155	123 ± 65.6
$t\bar{t} + W/Z$	8.40 ± 2.60	20.6 ± 6.35	1.90 ± 0.604
Single Top	430 ± 44.7	266 ± 28.5	7.56 ± 1.11
Dibosons	4.59 ± 3.46	2.62 ± 2.90	0.0346 ± 0.0477
Z + jets	538 ± 51.1	295 ± 33.7	14.2 ± 4.41
Fakes	25.7 ± 13.1	23.8 ± 12.2	0.885 ± 0.485
Total	15 500 ± 2220	10 000 ± 1660	342 ± 38.6
Data	15296	9996	374

	$\geq$ 4j, 2b	**$\geq$ 4j, 3b**	**$\geq$ 4j, $\geq$ 4b**
$t\bar{t}H$	15.1 ± 0.808	8.27 ± 0.570	2.55 ± 0.306
$t\bar{t}$ + light jets	4570 ± 881	140 ± 30.9	1.33 ± 0.931
$t\bar{t} + c\bar{c}$ jets	756 ± 207	147 ± 113	5.86 ± 5.98
$t\bar{t} + b\bar{b}$ jets	326 ± 166	222 ± 114	36.2 ± 19.9
$t\bar{t} + W/Z$	44.6 ± 15.7	7.66 ± 2.97	1.07 ± 0.357
Single Top	127 ± 27.9	9.69 ± 3.21	0.765 ± 0.232
Dibosons	2.41 ± 2.27	0.134 ± 0.157	0.0179 ± 0.0381
Z + jets	140 ± 28.9	16.8 ± 9.98	0.480 ± 0.546
Fakes	25.3 ± 13.1	4.07 ± 2.18	0.329 ± 0.210
Total	6010 ± 1040	556 ± 67.6	48.6 ± 15.1
Data	6002	561	46

Table 7.1: Event yields for $t\bar{t}H$ signal, relevant backgrounds and observed data in all multiplicity bins considered in the analysis. The uncertainties on the expected event yields include both statistical and systematic sources.

before any b-tagging is applied can be used to predict the normalisation and shape after the b-tagging [271].

For each event, all possible permutations of labelling n_b jets as "tagged" are considered. In every permutation assessed, each jet of the event is attributed a weight according to their probability of being tagged. The probability of each jet to be b-tagged, ε_i, is parameterised as a function of the true jet flavour, p_{T} and η. Therefore, "tagged" jets are given a weight equal to the tagging probability, while "un-tagged" jets are given a weight of one minus the tagging efficiency. Assuming that the probability of tagging a jet is independent of the rest of the jets in the event,[1] the total probability for each event to contain a selected number of tagged jets can be obtained by multiplying all jet weights in the event. For a given event with N jets, the probability of containing exactly one b-tagged jet ($n_b = 1$) is calculated by the TRF method as:

$$P_{n_b=1} = \sum_{i=1}^{N} \left(\varepsilon_i \prod_{i \neq j} (1 - \varepsilon_j) \right) , \tag{7.1}$$

where the tagging probability of the ith jet is given by $\varepsilon_i = \varepsilon_i(\text{flavour}, |\eta|, p_{\mathrm{T}})$. For inclusive b-tagging selections, for instance $n_b \geq 1$, the probability is given as:

$$P_{n_b \geq 1} = 1 - P_{n_b=0} \, , \tag{7.2}$$

with $P_{n_b=0}$ the probability of the event containing exactly zero b-tagged jets. This way all events in the pre b-tagged selection are used to predict the normalisation and shape after the b-tagging requirements of the analysis. For a specific b-tagging requirement, the shape of the distribution is built by selecting one of the possible permutations based on their relative probability.

The TRF method assumes that the parameterisation of the tagging efficiencies, ε_i, as a function of the jet flavour, p_{T} and η is sufficient to describe the b-tagging dependencies. The MC true jet flavour is determined for partons with $p_{\mathrm{T}} > 5\,\mathrm{GeV}$ by checking if a b quark is found within a $\Delta R < 0.3$ cone around the jet direction, in which case the jet is labelled with b origin. When no b quark is found, c quark presence is gauged instead. If no heavy flavoured quark is found, the jet is labelled as a light jet. The efficiencies ε_i are derived from an inclusive $t\bar{t}$+jets simulated sample. The production mechanism of the jet is taken into account, since b-jets coming from top quark decay products have a slightly higher b-tagging probability than those of a b-jet with the same p_{T} and η but originating from other sources.

[1] This is also the case in the calibration of b-tagging algorithms.

Jet Flavour	Source of Origin	Efficiency [%]
***b*-jet**	top quark decay	72
	additional jets	66
	MPI	49
***c*-jet**	W boson decay	22
	additional jets	18
light-jet	W boson decay or additional jets	0.7

Table 7.2: Summary of the average b-tagging efficiency for jets with different origin sources.

Thus, for the different types of jets, the b-tagging efficiency, averaged over p_{T} and η, are derived separately. In Table 7.2 a summary of the b-tagging efficiency averaged over p_{T} and η for the different types of jets is shown.

In order to validate the performance of the parameterisation, tests on simulated events were performed. The TRF method provides good agreement with the yields and shapes obtained by applying the MV1 b-tagging algorithm directly, within the statistical uncertainty. Given the reasonable agreement within the phase space of the analysis, no additional systematic uncertainty of the tagging efficiency is needed.

In this thesis, the TRF method was applied to all signal and background samples.

7.1.2 Systematic Uncertainties

In this analysis, aside from the statistical uncertainty associated with the finite number of events, several sources of systematic uncertainty are taken into consideration. The normalisation of both signal and background processes, as well as the shape of the kinematic distributions, can be affected by systematic uncertainties, which in turn impacts the discriminant distributions and the sought-after measurements. Here, individual sources of systematic uncertainty are considered in order to be treated as uncorrelated amongst each other. The correlations of a given systematic effect are kept across processes and analysis regions. The sources of systematic uncertainties considered in this study can be categorised into detector-related, background modelling and signal modelling systematics. Detector-related systematic uncertainties are estimated from auxiliary measurements, resulting in the nominal and the $\pm 1\sigma$ variations. For theory-related systematics, however, the $+1\sigma$ variation is determined as the difference between the alternative and the nominal model, which is then

Systematic uncertainty	Type	Components
Detector Modelling		
Luminosity	N	1
Electron	SN	5
Muon	SN	6
Jet energy scale	SN	22
Jet vertex fraction	SN	1
Jet energy resolution	SN	1
Jet reconstruction	SN	1
b-tagging efficiency	SN	6
c-tagging efficiency	SN	4
Light-jet tagging efficiency	SN	12
High-p_{T} tagging efficiency	SN	1
Background Modelling		
Z+jets normalisation	N	1
Z+jets modelling: $< p_{\mathrm{T}}^{l_1+l_2} >$ and $< p_{\mathrm{T}}^{j_1+j_2} >$ reweighting	S	2
$t\bar{t}$ normalisation	N	1
$t\bar{t}$ modelling	SN	3
$t\bar{t}$ + heavy flavour: normalisation	N	2
$t\bar{t}$+$b\bar{b}$	SN	8
$t\bar{t}$+$c\bar{c}$	SN	5
Lepton misID normalisation	N	1
Single top cross-section	N	1
Diboson+jets normalisation	N	3
$t\bar{t} + V$ cross-section	N	1
Signal Modelling		
$t\bar{t}H$ scale	SN	2
$t\bar{t}H$ generator	SN	1
$t\bar{t}H$ hadronisation	SN	1

Table 7.3: Summary list of systematic uncertainties considered in the analysis. The systematic uncertainties are taken to be normalisation-only (N), shape-only (S) or both shape and normalisation (NS) across all processes and regions. Some of the systematic uncertainties are split into several components for a more accurate treatment.

symmetrised to obtain the -1σ variation. These are summarised in Table 7.3, while also indicating whether they are taken to be normalisation-only (N), shape-only (S), or to affect both shape and normalisation (NS).

Luminosity

The integrated luminosity for the data set used has an uncertainty of 2.8%, determined using van der Meer scans [191]. This systematic uncertainty affects all processes whose contributions are determined with simulated samples.

Uncertainties on Physic Objects

While defining physics objects, such as leptons, jets, $E_{\mathrm{T}}^{\mathrm{miss}}$ and even jet flavour-tagging, the reconstruction and calibration processes will lead to associated uncertainties, which were already discussed in detail in Chapter 5.4. In this analysis, the largest physics objects related uncertainties are the the b-tagging uncertainties and the jet energy scale.

Leptons

There are in total five (six) systematic uncertainties associated with electrons (muons), which pertain to the trigger, reconstruction, identification and isolation efficiencies, as well as lepton energy-momentum scale and resolution.

Jets

In the case of jet selection, the uncertainties are associated with the jet energy scale (JES), jet vertex fraction (JVF) requirement, jet energy resolution (JER) and jet reconstruction efficiency. In this analysis, the JES has one of the largest impacts on the fitted signal strength μ of the binned likelihood fit. It is split into 22 components arising from uncorrelated sources, which can have different jet p_{T} and η dependencies. The modelling dependence of the JES calibration has the largest post-fit impact on μ.

Heavy and Light Flavour Tagging

The heavy flavour tagging efficiencies were already briefly explained in Section 5.4.2.6. The uncertainties on the b-tagging efficiency of jets originating from b, c and light quarks or gluons, can be split into several uncorrelated components using the eigenvector method, which consists of a diagonalisation procedure that preserves the correlation among the jet p_{T} bins of the measured SFs [267]. Therefore, there are a total of six (four) independent sources of uncertainty affecting the $b(c)$-tagging efficiency according to the six (four) jet p_{T} bins used in the efficiency measurement. An additional uncertainty is assigned due to the extrapolation of the b-tagging efficiency measurement to the high transverse momentum region. Twelve components are considered for the mistag rate or light-jet tagging, where the efficiency is parametrised in six bins in jet p_{T}, and in two η regions.

As already mentioned, no additional uncertainty is included due to the use of the TRF method, since the difference on the systematic uncertainties between using the TRF approach or applying the MV1 b-tagging algorithm directly is negligible compared to the other sources.

The b-tagging uncertainty on the fifth p_{T} bin is the third leading uncertainty in terms of the impact on the fitted signal strength.

Uncertainties on Background Modelling

Z+jets Modelling

The Z+jets data driven modelling, discussed in Section 6.2.1, was obtained from data behaviour in control regions not included in the nominal analysis. To improve the modelling of the Z boson p_{T} spectrum, the method included a cross-section normalisation, an average p_{T} reweighting of the two leptons and an average p_{T} reweighting of the two leading jets, determined for different multiplicity bins of jets and b-tagged jets. The variation of the scale factors within their statistical errors was taken as systematic uncertainty of the Z+jets modelling, resulting in three systematic components symmetrised with respect to the nominal value.

$t\bar{t}$+jets Modelling

As already discussed in Section 6.2.2, the inclusive $t\bar{t}+$ jets normalisation was determined using data information in control regions selected for different multiplicity bins of jets and b-tagged jets. The effect of varying the correction scale factors within their statistical errors is then taken as a systematic uncertainty of the $t\bar{t}+$ jets normalisation.

The $t\bar{t}$+jets modelling can be affected by sources of systematic uncertainties, such as the parton shower and hadronisation model. These uncertainties are derived by comparing events produced by Powheg interfaced with Pythia or Herwig. By changing the parton shower, different predictions of parton-level kinematics are observed both in the number of jets and on the heavy flavour content. Since this fragmentation uncertainty affects the relative population of flavour jets it is represented by three uncorrelated parameters, one acting on the $t\bar{t}+$ light jets contribution and two others on the $t\bar{t}+b\bar{b}$ jets and $t\bar{t}+c\bar{c}$ jets contributions. Considering that in this analysis no specific shape correction is applied to the $t\bar{t}+$ jets background processes, and a slight mis-prediction in the higher b-tagged bins of the analysis is present, a conservative systematic on the $t\bar{t}+$ jets modelling, is included. The Powheg+Herwig production is then reweighted to relative differential cross-section measurements of $t\bar{t}+$ light jets, $t\bar{t}+b\bar{b}$ jets and $t\bar{t}+c\bar{c}$ jets contributions according to the difference between $\sqrt{s}=7\,\mathrm{TeV}$ data and simulation in transverse momentum distributions of the top quark and of the $t\bar{t}$ system [272]. This reweight technique is used in the combined

semi-leptonic and dileptonic ATLAS analyses [1] to correct the nominal $t\bar{t}$ background instead of the alternative data driven method developed in this thesis (Section 6.2.2), which makes use of the statistics available at $\sqrt{s} = 8\,\text{TeV}$ to correct the inclusive $t\bar{t}$ normalisation. Since the data driven method used in this thesis does not attempt to correct the shape of the distributions, or the relative population of $t\bar{t}$ + light jets, $t\bar{t} + b\bar{b}$ jets and $t\bar{t} + c\bar{c}$ jets contributions, this reweight correction is then used as a possible systematic variation of the $t\bar{t}$+jets modelling. The $t\bar{t}$+jets modelling is the leading systematic uncertainty in the results of this thesis, as will be seen in Section 7.3.

$t\bar{t}$ + Heavy Flavour Modelling

The MC generator that best models the whole $t\bar{t}$ + jets production is Powheg+Pythia, which is used in this analysis as nominal reference. However, the Powheg generator only includes $gb \to t\bar{t}b$ diagrams, which can only provide leading logarithmic (LL) accuracy to the modelling of the $t\bar{t} + b\bar{b}$ jets. Being the main irreducible background in the signal regions, a precise $t\bar{t} + b\bar{b}$ jets modelling is essential for this analysis. To account for the limited knowledge on the HF components, supplementary uncertainties are added on the $t\bar{t} + b\bar{b}$ jets and $t\bar{t} + c\bar{c}$ jets simulations by comparing different MC generators and parton-showers to the nominal Powheg+Pythia. Specifically, an exclusive $t\bar{t} + b\bar{b}$ jets NLO prediction with massive b quarks within the Sherpa framework [273] interfaced with the OpenLoops library [274] sample is considered, as it allows a complete cover of the full $t\bar{t} + b\bar{b}$ jets phase space. The Sherpa+OpenLoops NLO sample was generated using the Sherpa 2.0 pre-release with the CT10 PDF set following the four-flavour scheme, where the b quark can only be generated as a massive final state and does not contribute to the proton PDF. An inclusive $t\bar{t}$ + jets sample generated with the Madgraph5 1.5.11 generator [236], again with the CT10 PDF set, and with parton shower and hadronisation performed by Pythia 6.427 is also investigated. This Madgraph+Pythia production was generated including tree-level diagrams with up to three extra partons following a five-flavour scheme, with b and c quarks treated as massless partons in the ME calculation as well as the possibility to have been originated inside the proton.

The cross-section contributions of the $t\bar{t}$ including different b particle-jet topologies are shown in Figure 7.3, for Powheg+Pythia, Sherpa+OpenLoops and Madgraph+Pythia samples. The categorisation on the x axis was done using the number of quarks and particle jets present per event, including the matching details: $t\bar{t} + b$ when a single particle jet is

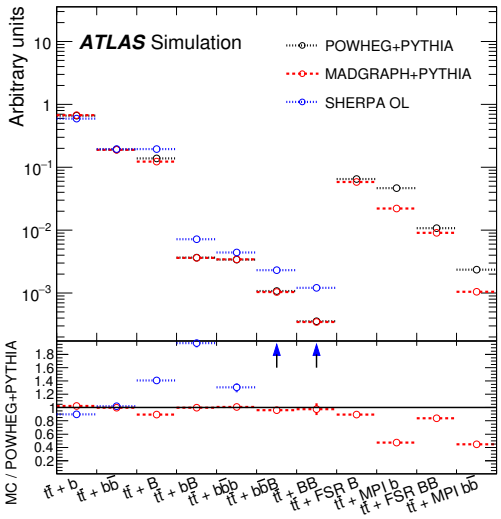


Figure 7.3: POWHEG+PYTHIA, SHERPA+OPENLOOPS and MADGRAPH+PYTHIA comparisons of the $t\bar{t} + b\bar{b}$ jets production.

matched to a single b quark; $t\bar{t} + b\bar{b}$ when two particle jets are matched to two different b quarks; and $t\bar{t} + B$ when a single particle jet is matched to a $b\bar{b}$ pair. The remaining categories result of the possible combinations of the previous three. The overall $t\bar{t}+b$ particle-jet cross-section agrees within 20% between all MC generators, however, the relative distribution across categories shows that SHERPA+OPENLOOPS predicts higher contributions of production processes where a second pair of $b\bar{b}$ is required, which is also were the biggest disagreement between inclusive $t\bar{t}$ + jets and data is seen. A similar classification is also done for the $t\bar{t} + c\bar{c}$ sample.

It is important to note that the SHERPA+OPENLOOPS NLO sample does not include certain diagrams of $t\bar{t} + b\bar{b}$ jets-like processes. For instance, events such as multiple parton interactions (MPI) that could originate $b\bar{b}$ pairs overlaying $t\bar{t}$ + jets events, or gluons radiated of top decay products that could also decay into a $b\bar{b}$ pair, known as final-state radiation (FSR), are not incorporated. The effects of these contributions, MPI and FSR, are shown in Figure 7.3 for the other two generators. In the following comparisons to the SHERPA+OPENLOOPS MC simulation, these contributions are identified and excluded to avoid double counting their systematic effect.

For most of the relevant kinematic distributions in this analysis, POWHEG+PYTHIA and SHERPA+OPENLOOPS show reasonable agreement in the predicted shapes. Slight differ-

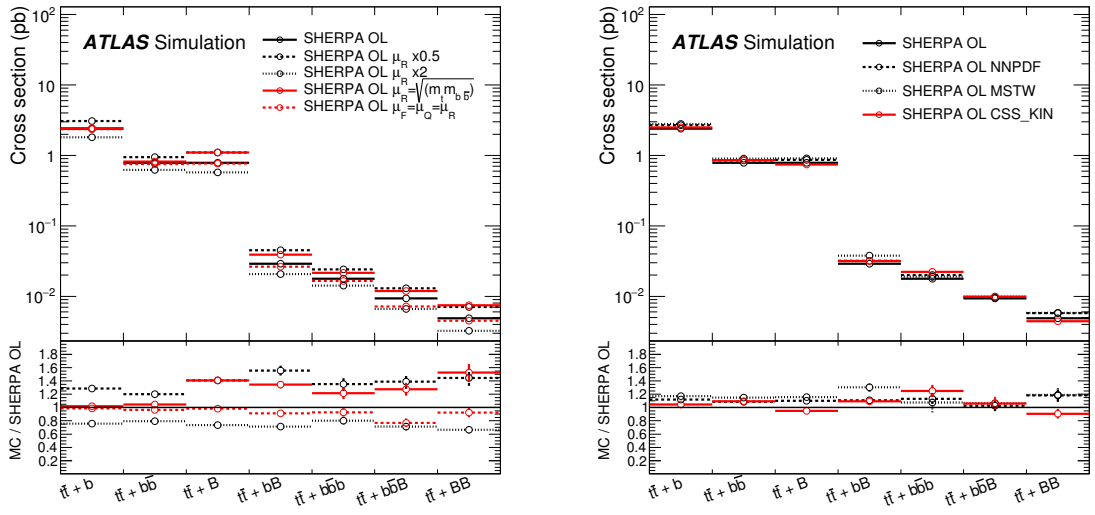
ences are present, however, in distributions of the transverse momentum of the top pair, as well as the ΔR of the $b\bar{b}$ pair. When comparing distributions of Powheg+Pythia and Madgraph+Pythia, good agreement is found. This could be a result of using the same parton shower program in both samples, since $b\bar{b}$ pairs are only produced in the showering process for Powheg+Pythia simulation.

From these detailed comparisons of the $t\bar{t}$+$b\bar{b}$ jets between the nominal Powheg+Pythia and the NLO Sherpa+OpenLoops prediction, a cross-section agreement within 50% of each other can be seen. As a result, a 50% systematic uncertainty on the normalisation of the $t\bar{t}+b\bar{b}$ component of the nominal $t\bar{t}+$ jets simulation is included. The same 50% systematic uncertainty is applied to the normalisation of the $t\bar{t}+c\bar{c}$ jets component of the $t\bar{t}+$ jets background since no NLO prediction of the $t\bar{t}+c\bar{c}$ jets component was available. These $t\bar{t}+$heavy flavour normalisation uncertainties on $t\bar{t}+b\bar{b}$ jets and $t\bar{t}+c\bar{c}$ jets are treated as uncorrelated. While having the second highest impact on the signal strength of the statistical fit performed in this thesis results, the available data sample allows for a significant constraining of the $t\bar{t}+$ heavy flavour normalisation uncertainties, as will be shown later.

Shape modelling uncertainties can also be estimated by taking the difference between parton level differential distributions of Powheg+Pythia and Sherpa+OpenLoops simulation. Although the full $t\bar{t}+$ jets cross-section is similar for both generators due to the DD method applied, the relative cross-sections in each of the heavy-hadrons and particle jet categories are distinct for each MC sample. Two independent kinematic variables are then considered according to the event topology: for events without additional HF jets, the distributions of the p_{T} of the top and $t\bar{t}$ pair are used; for events with only one additional HF jet, the distributions of the p_{T} and η of the heavy-flavour jet are used; and for events with two or more HF jets, the systematic is based on the ΔR and p_{T} of the dijet system.[1]

Other specific $t\bar{t}+b\bar{b}$ jets modelling uncertainties are derived through variations of the NLO Sherpa+OpenLoops sample. In this analysis, uncertainties are obtained by changing the functional form of the renormalisation scale to $\mu_{\mathrm{R}}=(m_t m_{b\bar{b}})^{1/2}$, changing the functional form of the factorisation μ_{F} and resummation μ_Q scales to $\mu_{\mathrm{F}}=\mu_{\mathrm{Q}}=\prod_{i=t,\bar{t},b,\bar{b}} E_{\mathrm{T,i}}^{1/4}$ and varying the renormalisation scale, μ_{R}, up and down by a factor of two. The shower recoil model uncertainty and two additional uncertainties due to the PDF choices (obtained by taking the full difference between the nominal CT10 PDF and two alternative PDF sets,

[1] These variables were chosen, since correcting them on the nominal sample to the NLO behaviour would improve the overall modelling of the remaining kinematic variables [1].



(a) (b)

Figure 7.4: $t\bar{t} + b\bar{b}$ jets systematic uncertainties based on (a) scale variations and (b) PDF choice and shower recoil model of the SHERPA+OPENLOOPS simulation, across different $t\bar{t}+b\bar{b}$ parton level categories [1].

MSTW and NNPDF) are also evaluated for completeness. In Figure 7.4, the different SHERPA+OPENLOOPS samples are shown across the b particle-jet topologies considered. The renormalisation scale choice has a large effect on the shape modelling of $t\bar{t} + b\bar{b}$ jets, resulting in the ninth leading uncertainty in the statistical fit. The cross-section prediction of these systematic uncertainties is corrected with the same data driven method as the nominal $t\bar{t} +$ jets, since the normalisation uncertainty is already considered.

Systematic uncertainties related to the MPI and FSR sources of $b\bar{b}$ pairs are also considered. For the first one, a dedicated sample with increased MPI activity is used. The difference between this enhanced sample and the nominal one is taken as systematic variation, which is around 25% per b-jet arising from MPI. For the latter, and given the nature of FSR, which originates from gluon splitting, events in the $t\bar{t} + B$ category are used to estimate a systematic variation, since these are dominantly produced by collinear $g \to b\bar{b}$. The difference between the normalised predictions of POWHEG (which include FSR) and SHERPA+OPENLOOPS (which do not include FSR) in the $t\bar{t} + B$ category, is then taken as a systematic variation, resulting in an uncertainty of about 40% per extra b-jet arising from FSR.

As stated previously, for the $t\bar{t}+c\bar{c}$ jets no NLO calculations were available at the time of

this analysis. Given the good agreement between the POWHEG+PYTHIA modelling and that of the multi-leg LO prediction MADGRAPH+PYTHIA, the full difference between the two is taken as a systematic uncertainty, which is particularly interesting as the latter includes the $t\bar{t} + c\bar{c}$ jets process in the matrix element calculation, unlike the POWHEG+PYTHIA where it is absent. In the MPI categories, however, a large difference is observed between the two generators. It is possible that these are due to the differences in the Pythia settings used when producing both samples, since the MPI prediction is modelled by the parton shower. Here, where POWHEG used PYTHIA 6.426, CTEQ61L pdf and Sandhoff colour-reconnection scheme [275], MADGRAPH was instead interfaced with PYTHIA 6.427 and using CT10 PDF and Wicke colour-reconnection scheme [276]. Nonetheless, the observed MPI difference is already covered by the preceding systematic. Similarly to the $t\bar{t}+b\bar{b}$ jets systematic variations, the cross-section prediction of the samples used for systematic uncertainty calculations are corrected with the DD method, since the normalisation uncertainty is already covered. Four additional systematic uncertainties in the $t\bar{t} + c\bar{c}$ jets background estimate are derived from the variation of factorisation and renormalisation scales simultaneously, matching threshold and c quark mass (within the range $1.50 \pm 0.8\,\text{GeV}$) in the MADGRAPH+PYTHIA $t\bar{t}+$ jets simulation. The difference between the $t\bar{t} + c\bar{c}$ jets events in MADGRAPH+PYTHIA and POWHEG+PYTHIA is then taken as a systematic uncertainty.

Misidentified Lepton Background Modelling

Systematic uncertainties on the misidentified lepton background are estimated as a 50% normalisation uncertainty, since the implemented data driven method depended in same-sign dilepton events in both data and MC simulation. The uncertainty is considered correlated across different jet and b-tag multiplicity bins.

Other Background Modelling

Sources of uncertainty related to the remainder background processes were also considered.

The uncertainty associated with the theoretical cross-sections of the single top production is $\pm 6.8\%$, corresponding to the theoretical uncertainty on the Wt-channel production [235], which is the only non-null single top process contributing to the dileptonic final state.

In the case of the diboson background, in order to account for the extrapolation to high jet multiplicity, an inclusive theoretical NLO cross-section uncertainty of $\pm 5\%$ is applied [234].

For the $t\bar{t}+V$ processes, an uncertainty of $\pm 30\%$ is assumed for the theoretical cross-sections [239, 277].

Uncertainties on Signal Modelling

For the $t\bar{t}H$ signal kinematics, in order to account for the uncertainty on the MC generator modelling, several systematic uncertainties are considered. The impact of the choice of factorisation and renormalisation scales as $\mu_0 = \mu_\mathrm{F} = \mu_\mathrm{R} = (m_t + m_H)/2$ on the $t\bar{t}H$ signal kinematics is assessed with dedicated NLO POWHEL samples, where the default scale is varied by a factor of two up and down. The effect of changing the functional form of the scale from the chosen static scale to a dynamical one is also considered systematic, which is taken as the difference between the two scales where the latter dynamical scale is defined as a geometrical average of the transverse mass of the generated partons, $\mu_\mathrm{dyn} = \mu_\mathrm{F} = \mu_\mathrm{R} = (m_\mathrm{T}^t . m_\mathrm{T}^{\bar{t}} . m_\mathrm{T}^H)^{1/3}$ [139]. Systematic uncertainties on the choices of generator as well as parton shower and fragmentation model are estimated as the full difference between the nominal POWHEL+PYTHIA8 prediction and MADGRAPH5_AMC@NLO [137] interfaced with HERWIG++ [278, 279] or POWHEL+HERWIG, respectively. Due to the low statistic of the alternative samples, and in order to avoid further statistical fluctuations, the effect of the systematic variations on $t\bar{t}H$ distributions were studied at particle level and the nominal POWHEL $t\bar{t}H$ sample are then reweighted to reproduce these variations. The PDF envelope uncertainty is evaluated following the PDF4LHC recommendation, taking as uncertainty the maximum difference between the predictions with the MSTW2008 68% CL NNLO, CT10 NNLO and NNPDF2.3 5f FFN PDF sets, which was found to be negligible.

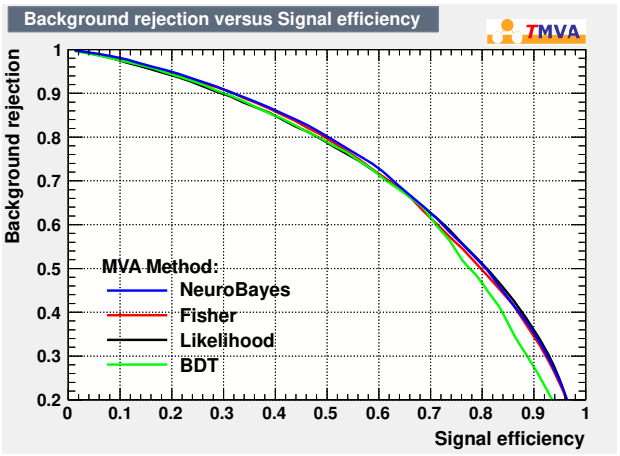
7.2 Multivariate Discriminant Analysis

After the common preselection of events, background modelling, event categorisation and systematic uncertainty description, in order to probe the dileptonic $t\bar{t}H$ signal, several variables are investigated based on their discriminant power to distinguish the signal processes from the background. While the event categorisation in jet multiplicity bins allows for high signal sensitivity by creating signal-rich regions without any loss of events, after the categorisation it is rather difficult to further suppress the irreducible $t\bar{t}+b\bar{b}$ background with simple selection cuts. The distinct kinematic properties of the $t\bar{t}H$ signal and of the $t\bar{t}+b\bar{b}$ background could allow differentiating the signal from the background, in particular those

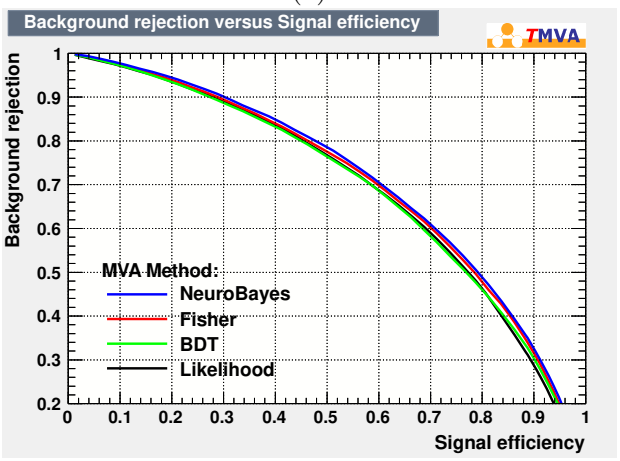
produced by the $H \to b\bar{b}$ decay or those of the angles between the momenta of the produced particles. The correct identification of the b-tagged jets (whether originated from the t quark, the $\bar{t}$ quark or the H boson) is not trivial, and neither is the reconstruction of the two neutrinos involved in the dileptonic decay, which makes it significantly harder to use as single sources of discrimination. Kinematic features such as these are, however, taken into consideration when constructing discriminant distributions for each region of the analysis, in particular for the signal-rich regions, where the information from diverse sources is combined in multivariate discriminant variables. In this thesis, these properties are additionally explored at a phenomenological level, including a full kinematic reconstruction algorithm for the dileptonic $t\bar{t}H$ process, described in Chapter 8.

Robust signal discrimination is obtained by performing a multivariate analysis (MVA) based on machine learning (ML) algorithms, where a non-trivial discriminating variable is created using the shape information of several distributions with significant signal-background contrast as input. In this analysis, three multivariate discriminants are built for regions with the highest signal significance: (3j, 3b), ($\geq$ 4j, 3b) and ($\geq$ 4j, $\geq$ 4b). It should be noted that (3j, 3b) is not considered a signal-rich region due to the small expected $S/\sqrt{B}$, but the use of a NN discriminant in this region adds sensitivity to the signal. The discrimination between $t\bar{t}H$ events and background events with similar experimental signatures is tested with different multivariate classification methods based on machine learning techniques using TMVA [280], which is an open source software included in ROOT [281]. The correlations among variables are exploited by training on events with a known classification (either signal or background) and determining a mapping function that describes the classification or an approximation of the underlying behaviour defining the target value (regression).

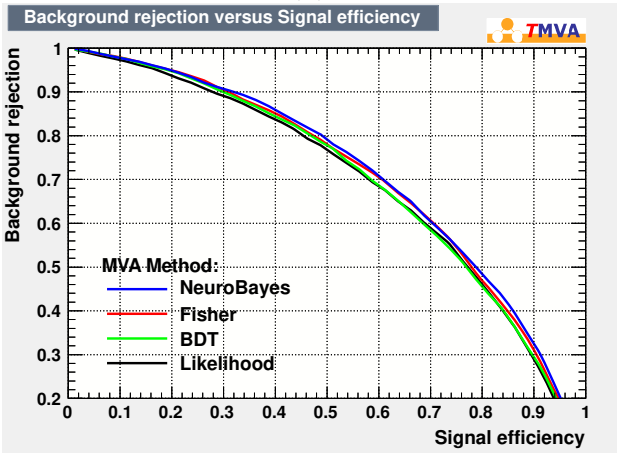
In this thesis, the methods of maximum likelihood, boosted decision trees (BDT), Fisher's discriminants and neural networks (NN) are tested for the classification of events. The discriminating power of each method is assessed by analysing the receiver operating characteristic (ROC) curve. The ROC curve represents the background rejection over the signal efficiency, and therefore the best classifier is the one maximising the area under it. Figure 7.5 presents the performance of the different multivariate algorithms tested for the three regions (3j, 3b), ($\geq$ 4j, 3b) and ($\geq$ 4j, $\geq$ 4b). Several input variables were considered, from simple kinematic distributions to more complex ones, and tested in different combinations for each of the regions. The results in Figure 7.5 were calculated using the same input distributions as in the final discriminant calculation shown below. Although, a similar performance can



(a)



(b)



(c)

Figure 7.5: ROC curves for different multivariate methods: neural networks (blue), Fisher's discriminant (red), maximum likelihood (black) and boosted decision trees (green). (a) for (3j, 3b); (b) for ($\geq$ 4j, 3b) and (c) for ($\geq$ 4j, $\geq$ 4b). The NeuroBayes method is the best one in all three regions.

be expected from the MVA methods examined, given that the highest performance was obtained with the neural network method, this was the preferred technique to build the discriminants in each of the considered regions of this analysis.

For the remaining regions of the analysis, (3j, 2b) and ($\geq$ 4j, 2b), since the signal sensitivity is lower, no MVA technique is applied. Instead, the scalar sum of the jet and lepton transverse momenta, H_{T}, is used in the likelihood fit that will search for the $t\bar{t}H$ presence. The signal-depleted bins, rich in data and background events, will allow better control of the systematic uncertainties, in particular those affecting the $t\bar{t}+$ light jets production. Being quite sensitive to the background modelling, as well as systematic uncertainties with transverse momentum dependence, the H_{T} is a suitable variable for the statistical fit.

7.2.1 Neural Network Training

The NN technique consists in calculating a final output from several input variables by recurring to one or more intermediate hidden layers of computation. Every layer n can have multiple nodes, with $n = 0$ being the first layer where each input variable is a node. An output is returned by each node j in layer n, which is calculated from a weighted sum of the outputs of the previous layer $n-1$, via a non-linear transfer function. The information will thus transfer from input to output consecutively, since each node can only be connected to the layer immediately after. In order to determine the best set of weights, the error/loss function is minimised while training the NN using events whose classification is known.

The analysis uses the NeuroBayes [282] package to calculate the NN discriminant independently for each signal region. This software includes an additional Bayesian regularisation method during the training stage, that intensifies the weights supported by data while pruning the node connections that are insignificant. This procedure changes the architecture of the NN reducing the number of free parameters. In this study, the NN was trained with a single hidden layer with $J+2$ nodes, where J input variables were fed to the NN algorithm, and a final layer with one output node.[1] All other hyperparameters of the NeuroBayes were set to the default ones. Both signal and background processes were divided in two sets, using one set for training and the other one for evaluation.

Considering the rich event topology and the diverse signal-background population in each jet multiplicity bin, the choice of the input variables is made independently in each

[1] The use of a single hidden layer with a sufficiently large number of nodes can approximate the Bayes optimal decision boundary [283].

Variable	Definition	NN rank		
		$\geq$ 4j, $\geq$ 4b	$\geq$ 4j, 3b	3j, 3b
$\Delta\eta_{jj}^{\max\Delta\eta}$	Maximum $\Delta\eta$ between any two jets in the event	1	1	1
$m_{bb}^{\min\Delta R}$	Mass of the combination of the two b-tagged jets with the smallest ΔR	2	8	-
$m_{b\bar{b}}$	Mass of the two b-tagged jets from the Higgs candidate system	3	-	-
$\Delta R_{hl}^{\min\Delta R}$	ΔR between the Higgs candidate and the closest lepton	4	5	-
N_{30}^{Higgs}	Number of Higgs candidates within 30 GeV of the Higgs mass of 125 GeV	5	2	5
$\Delta R_{bb}^{\max p_T}$	ΔR between the two b-tagged jets with the largest vector sum p_T	6	4	8
$\text{Aplan}_{\text{jet}}$	$1.5\lambda_2$, where λ_2 is the second eigenvalue of the momentum tensor built with all jets	7	7	-
$m_{jj}^{\min m}$	Minimum dijet mass between any two jets	8	3	2
$\Delta R_{hl}^{\max\Delta R}$	ΔR between the Higgs candidate and the furthest lepton	9	-	-
m_{jj}^{closest}	Dijet mass between any two jets closest to the Higgs mass of 125 GeV	10	-	10
H_T	Scalar sum of jet p_T and lepton p_T values	-	6	3
$\Delta R_{bb}^{\max m}$	ΔR between the two b-tagged jets with the largest invariant mass	-	9	-
$\Delta R_{lj}^{\min\Delta R}$	Minimum ΔR between any lepton and jet	-	10	-
Centrality	Sum of the p_T divided by sum of the E for all jets and both leptons	-	-	7
$m_{jj}^{\max p_T}$	Mass of the combination of any two jets with the largest vector sum p_T	-	-	9
$H4$	Fifth Fox–Wolfram moment computed using all jets and both leptons	-	-	4
$p_T^{\text{jet 3}}$	p_T of the third leading jet	-	-	6

Table 7.4: The definitions and rankings of the variables considered in each of the regions where a NN is used in the analysis.

region, with several variables surveyed for their discriminating power. The ranking procedure included in the NeuroBayes package is used to select ten input observables[1] in each signal-rich region according to their statistical separation power and the correlation amongst variables. Table 7.4 lists the highest-ranked variables selected by the NN and their respective ranking. The considered input observables varied from straightforward object kinematics, such as object (or pair) p_T and η properties, to slightly more complex observables, such as event shape distributions or global event variables that required information from all final state objects. In the regions with ≥ 4 jets, a maximum of five jets are considered to construct the

[1]The selection of ten input variables results of balancing the robustness of the NN limiting the complexity of the analysis, and obtaining the best separation between signal and background possible.

kinematic variables, first considering the b-tagged jets and then incorporating the untagged jets with the highest p_{T}. By resorting to this approach, one ensures that the discrimination power is not arising from the presence of soft jets, whose correct modelling is quite difficult to achieve in signal simulation with only NLO accuracy.

Due to the lack of statistical discrimination between signal and background, a full kinematic reconstruction of the events is not performed in this analysis. Instead, object pair properties are used to collect information, even if partially, on the Higgs boson and top quark pair. The jet assignment is done so that the two b-tagged jets closest in ΔR to the leptons are considered as having originated from top quarks; from the remainder b-tagged jets, the ones with highest transverse momenta are assigned to the Higgs candidate. Here, the use of a NN is particularly advantageous, allowing the combination of the information of partial event reconstruction variables, such as p_{T}, invariant mass, $\Delta\eta$ and ΔR of jet pairs or lepton-jet pairs, without the complexities of a full kinematic reconstruction.

Global and event shape variables are less sensitive to the loss of jets through acceptance effects and can be accessed in every jet and b-tagged jet multiplicity region. The event shape distributions explored in this study are the sphericity, aplanarity, centrality and the Fox-Wolfram moments. They can be constructed either using all leptons and jets in the event or using only the jets in the event (referenced with suffix "jet"). The sphericity and aplanarity are combinations of the eigenvalues of the linear momentum tensor with his elements defined as $M^{\alpha\beta} = \sum_i p_i^\alpha p_i^\beta / \sum_i |p_i|^2$ where i runs over all particles considered, and the indices α β run over the x, y, z component of the vector momentum of the particle [284]. The centrality is defined as $C = \sum p_{\mathrm{T}} / \sum E$, where the summation runs over all jets and both leptons in the event. The Fox-Wolfram moments describe the geometrical correlation among objects in the event in terms of spherical harmonics [285]. The fifth Fox-Wolfram moment is defined as $H_4 = \sum_{i,j}((|\vec{p}_i||\vec{p}_j|)/E^2)P_4 \cos\theta_{ij}$ where i, j run over all final state objects in the event, $\vec{p}_i$ is the momentum vector of i, E is the sum of the energy of all considered objects, P_4 is the fourth-degree Legendre polynomial and θ_{ij} is the angle between the objects i and j. Out of all global event variables considered in this analysis, the H_{T}, the number of Higgs boson candidates above a p_{T} threshold of 30 GeV and the minimum ΔR between the Higgs candidate and the closest lepton are among the top ten discriminating observables ranked by the NeuroBayes package.

In Figure 7.6 the normalised distributions used as input for the NN training are shown for the ($\geq$ 4j, $\geq$ 4b) region. For completeness, the correlation between the variables is shown

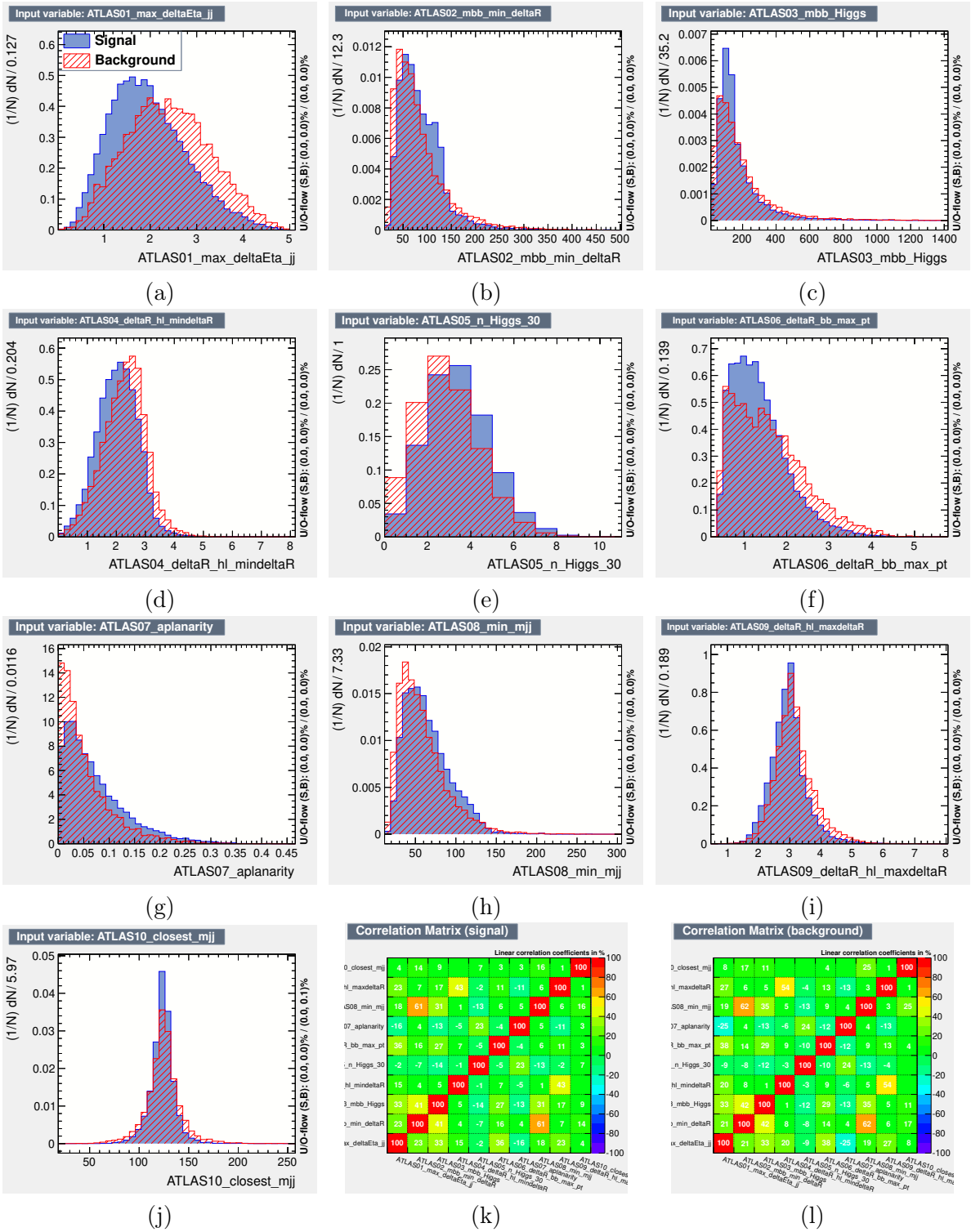
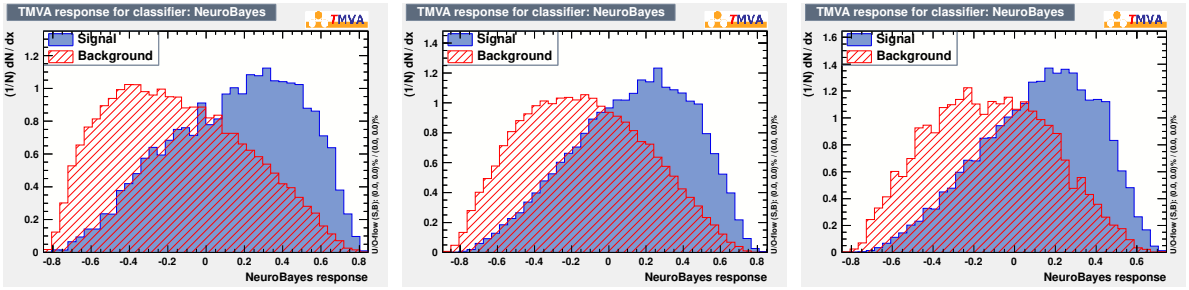
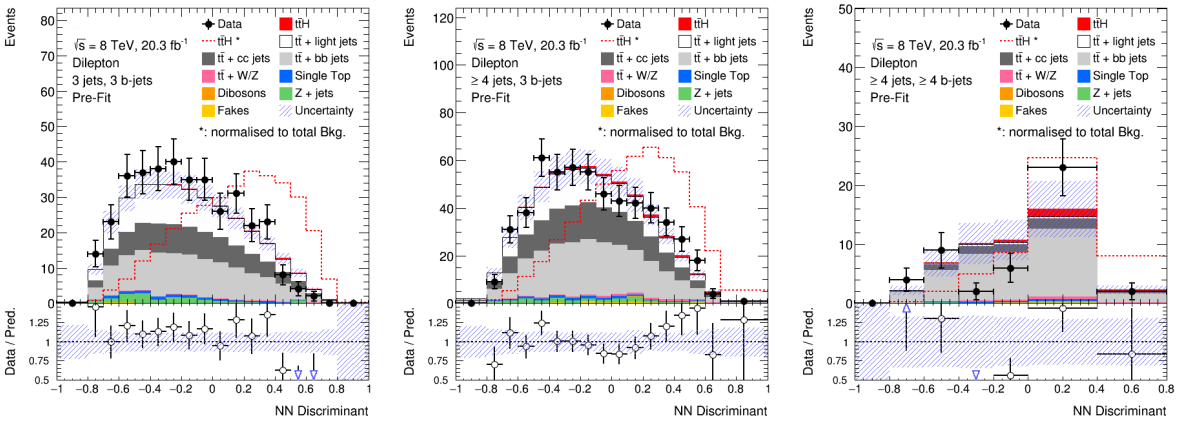


Figure 7.6: Normalised distributions of the NN highest-ranked input variables in the ($\geq$ 4j, $\geq$ 4b) region: (a) $\Delta\eta_{jj}^{\max\Delta\eta}$; (b) $m_{bb}^{\min\Delta\mathrm{R}}$; (c) $m_{b\bar{b}}$; (d) $\Delta R_{hl}^{\min\Delta\mathrm{R}}$; (e) N_{30}^{Higgs} ; (f) $\Delta R_{bb}^{\max p_{\mathrm{T}}}$; (g) $\mathrm{Aplan}_{\mathrm{jet}}$; (h) $m_{jj}^{\min m}$; (i) $\Delta R_{hl}^{\max\Delta\mathrm{R}}$; and (j) $m_{jj}^{\mathrm{closest}}$. The correlation among the variables are also shown, (k) for the signal and (l) for the background events.



(a) (b) (c)

Figure 7.7: NN output for the different regions: (a) for (3j, 3b); (b) for ($\geq$ 4j, 3b) and (c) for ($\geq$ 4j, $\geq$ 4b). The distributions are normalised to unit area. Blue filled histogram represents the $t\bar{t}H$ signal, whereas the dashed red histogram is background.



(a) (b) (c)

Figure 7.8: NN output discriminants in data and MC simulation for the different regions: (a) for (3j, 3b); (b) for ($\geq$ 4j, 3b) and (c) for ($\geq$ 4j, $\geq$ 4b). The uncertainty bands include both MC statistical uncertainty and systematic uncertainties before the fit procedure. The red dashed line shows the $t\bar{t}H$ signal normalised to the background yield.

in Figures 7.6 (k) for signal and (l) for background events. The correlation is mostly small (under 40%), although it reaches up to approximately 60% between $m_{bb}^{\min\Delta\mathrm{R}}$ and $m_{jj}^{\min\mathrm{m}}$. The input discriminant variables for the other two regions can be seen in Appendix B. All observables and their pairwise correlations are well described in simulation in all control regions.

As can be seen in Table 7.4, the NeuroBayes selection of the input variables according to their discriminating power and correlation is quite similar for the regions with $\geq$ 4 jets,

mostly due to the identical kinematics of events between the two regions. However, the ranking among the selected variables is different. This is mostly due to the signal topology, where ≥ 4 b-tagged jets are expected, as well as the different populations of heavy flavoured jets of the dominant background processes in each region. In the ($\geq$ 4j, 3b) region, global variables or variables which do not directly dependent on the b-tagging of jets become more important for the NN algorithm. This is even more noticeable in the (3j, 3b) region where the signal sensitivity is smaller and the discriminating power of event shape variables becomes more prominent. Furthermore, the maximum $\Delta\eta$ between any two jets in the event is the highest ranked discriminant observable in all three signal regions.

The NN output is a set of weights, and a response function, that can be applied to any signal or background process to determine if the process is background-like or signal-like. In order to separate signal from background, the algorithm is applied to the data in the aforementioned regions. Figure 7.7 illustrates the distribution of the NN discriminant for the $t\bar{t}H$ signal and background. For all jet and b-tagged jet multiplicity regions considered in the fit, the $t\bar{t}H$ signal includes all Higgs decay modes, which were also included in the NN training. The NN output distributions in data and MC events are shown in Figure 7.8 for the three regions with highest signal significance. A good agreement between data and simulation is seen in the NN discriminants, as a result of the already good description of the selected input variable distributions.

7.3 Binned Likelihood Fit to Data

The distributions of the discriminants from each of the five regions considered are combined to test for the presence of a signal, assuming a Higgs boson mass of $m_H = 125\,\text{GeV}$. The dileptonic $t\bar{t}H$ search performed in the framework of this thesis is based on the maximisation of a likelihood function. In this statistical test, the likelihood function gives the probability that a given hypothesis is compatible with the observed data. In order to validate the simulated $t\bar{t}H$ production process two hypothesis are tested: one describing the already known physics processes (H_0, *null* or *background-only hypothesis*), and one which also includes the $t\bar{t}H$ production (H_1, *test* or *signal plus background hypothesis*). A signal strength modifier, μ, defined as:

$$\mu = \frac{\sigma_{t\bar{t}H}}{\sigma^{\text{SM}}_{t\bar{t}H}}\,, \tag{7.3}$$

whose values will depend on the physics hypothesis under study — 0 in the null hypothesis, 1 in the test hypothesis — is the parameter of interest in the fit. The modified frequentist approach [286, 287] for confidence levels, is used to construct a likelihood function $\mathcal{L}(\text{data}|\mu,\theta)$ in terms of μ, the nuisance parameters θ and the observed data n. Assuming data follows a Poisson distribution and considering the minimum division in which the observed data is classified, i.e. one single histogram bin, i, in one region, j, the binned likelihood function describing the probability of the observed n events to have been produced by the model is:

$$\mathcal{L}(\text{data}|\mu,\theta) = \prod_{j}^{\text{all reg.}} \text{Poisson}\left(\text{data}|\,\mu s(\theta)+b(\theta)\right). \prod_{k}^{\text{all NP}} \rho(\theta) \tag{7.4}$$

$$= \prod_{j}^{\text{all reg.}} \prod_{i}^{\text{all bins }(j)} \frac{\left(\mu s_{ij}(\theta)+b_{ij}(\theta)\right)^{n_{ij}}}{n_{ij}!} e^{-(\mu s_{ij}(\theta)+b_{ij}(\theta))}. \prod_{k}^{\text{all NP}} \rho(\theta)\,, \tag{7.5}$$

where the event yields of signal and background processes are denoted as s and b, and the nuisance parameters (NP) θ account for the systematic and statistical uncertainties on these predictions. The probability density function $\rho(\theta)$ is usually referred as a prior on θ and assumes different functional forms depending on the nature of the NP: Gaussian or log-normal, for shape or normalisation systematic uncertainties respectively; a Gamma pdf can be additionally used for statistical uncertainties. In the limit of small uncertainties, both the log-normal and Gamma pdfs can be approximated to a Gaussian distribution. In this study, the $\rho(\theta)$ functions are broken down in such a way as to be treated as either 100% correlated or uncorrelated, allowing the inclusion of all constraints in the likelihood in a clean factorised form, as mentioned in Section 7.1.2. For each uncertainty break, $\theta = 0$ corresponds to the template of the nominal distribution, whereas $\theta = \pm 1$ corresponds to the $\pm 1\sigma$ variation of the systematic uncertainty.

The likelihood function can then be maximised in two distinct ways, resulting in: an unconditional maximum likelihood, $\mathcal{L}(\hat{\mu},\hat{\theta})$, where both the signal strength and the nuisance parameters are fitted; or a conditional maximum likelihood, $\mathcal{L}(\mu,\hat{\hat{\theta}}(\mu))$, when maximised for a fixed value of the signal strength μ. Conventionally, the profile likelihood ratio defined as $\lambda(\mu) = \mathcal{L}(\mu,\hat{\hat{\theta}}(\mu))/\mathcal{L}(\hat{\mu},\hat{\theta})$ is used as the test statistic for most LHC searches. Ranging from $0 < \lambda < 1$, the profile likelihood implies that the hypothesised value of μ has a good description of data when λ is close to unity. In this analysis, the test statistic used is

$$q_\mu = -2\ln(\lambda(\mu)) = -2\ln(\mathcal{L}(\mu,\hat{\hat{\theta}}_\mu)/\mathcal{L}(\hat{\mu},\hat{\theta}))\,, \tag{7.6}$$

where the best fit values $\hat{\mu}$ and $\hat{\theta}$ are the values of the parameters that maximise the likelihood function (with the constraints $0 \leq \hat{\mu} \leq \mu$), and $\hat{\hat{\theta}}_\mu$ are the values of the nuisance parameters that maximise the likelihood function for a given value of μ [288].

The probability that the observed data originates from the considered hypothesis can be computed from the test statistic as a p-value:

$$p_\mu = \int_{q_{\mu\,\text{obs.}}}^{\infty} f(q_\mu|\mu) dq_\mu \tag{7.7}$$

with $q_{\mu\,\text{obs.}}$ the observed value of the test statistic in data and $f(q_\mu|\mu)$ the pdf of q_μ assuming the hypothesis μ. For example, p_0 is the p-value in the background-only hypothesis where $\mu = 0$, which refers to the probability of the observed data to be caused by a fluctuation in the background-only hypothesis. Additionally, the test statistic can provide statistical inferences about μ, such as upper limits using the confidence level CL_s method [286]–[288]. A more general p-value, which assesses the compatibility of an hypothesis μ when the data is originated from a model with μ', can be determined using the pdf $f(q_\mu|\mu')$ with $\mu \neq \mu'$. For instance, to characterise the expected performance of a search-like analysis, $f(q_0|1)$ is used to compute the median significance for a discovery, while $f(q_1|0)$ returns the expected 95% CL limit in the absence of a signal.

An alternative way to access the probability that the observed data originates from the background-only hypothesis ($\mu = 0$) is the significance, Z_0, which can be determined from to the p_0 value as:

$$Z_0 = \Phi^{-1}(1 - p_0)\,, \tag{7.8}$$

where Φ^{-1} is the inverse of the cumulative distribution of the standard Gaussian. This means that a Gaussian distributed variable which is Z_0 standard deviations above the background-only hypothesis, has an upper-tail probability equal to p_0. Conventionally, an appropriate significance value to reject the background-only hypothesis is $Z_0 = 5$ standard deviations, which in the particle physics community constitutes a discovery of a new signal.

Since the expected number of events depends on variations of the nuisance parameter value, during the maximisation of the likelihood, adjustments in θ will occur in order to improve the agreement between the model expectation and the observed data. A shift, or pull, of a nuisance parameter's central value can occur to improve the data and simulation agreement; and/or a reduction (or constraint) of a nuisance parameter's uncertainty with respect to the nominal value, if the initial large uncertainty effect is not compatible with the

range allowed by data. The fitted values of the nuisance parameters will then correspond to the amount that best fits the data. In particular, when several NP create a similar effect and the total variation is larger than the precision supported by data, a correlation (or anti-correlation) is established, so that the combined effect is at the level of the data statistics. This procedure of including the nuisance parameters in the fit is referred to as 'profiling', which results in additional information on the systematic uncertainties and therefore an improvement in the analysis sensitivity. By taking advantage of the highly populated background-dominated control regions included in the likelihood fit, the impact of systematic uncertainties on the search sensitivity can be reduced.[1] Such a powerful fit procedure requires a good understanding of the effect of each systematic uncertainty on the shapes of the discriminant distributions per fit region and process. In order to identify the leading systematic uncertainties affecting the sensitivity of the search, a ranking procedure of nuisance parameters is also used, which allows for more detailed study of the systematic sources.

For the dileptonic $t\bar{t}H$ search, a simultaneous fit to the data on the discriminant distributions of the five analysis regions ((3j, 2b), (3j, 3b), ($\geq$ 4j, 2b), ($\geq$ 4j, 3b) and ($\geq$ 4j, 4b)) is performed under the signal-plus-background hypothesis. The signal-strength parameter μ, which is the parameter of interest in the fit, is required to be the same in all fit regions. Each analysis region is differently populated and has diverse contributions of systematic uncertainties. The normalisation of each background is determined from the fit simultaneously with μ, while the profiling of the NPs allows the fit to constrain the systematic uncertainties. The statistical uncertainties in each bin of the discriminant distributions are also considered in the fit with dedicated parameters. Figures 7.9 and 7.10 show the distributions before and after the fit for the signal-depleted and signal-rich regions respectively. For the (3j, 2b) and ($\geq$ 4j, 2b) regions, the H_{T} distribution is used in the fit, while for the remaining regions the NN output is used. The performance of the fit is tested using simulated events by injecting $t\bar{t}H$ signal with free signal strength, which is then compared to the fitted value. Good agreement between the injected and measured signal strength is observed. Compared to the pre-fit distributions, the background uncertainty is significantly reduced after the fit. Contributions from $t\bar{t}$, W/Z+jets production, single top quark production, diboson and

[1] It should be noted that a too simplistic systematic treatment might induce overconstraints, which should be avoided.

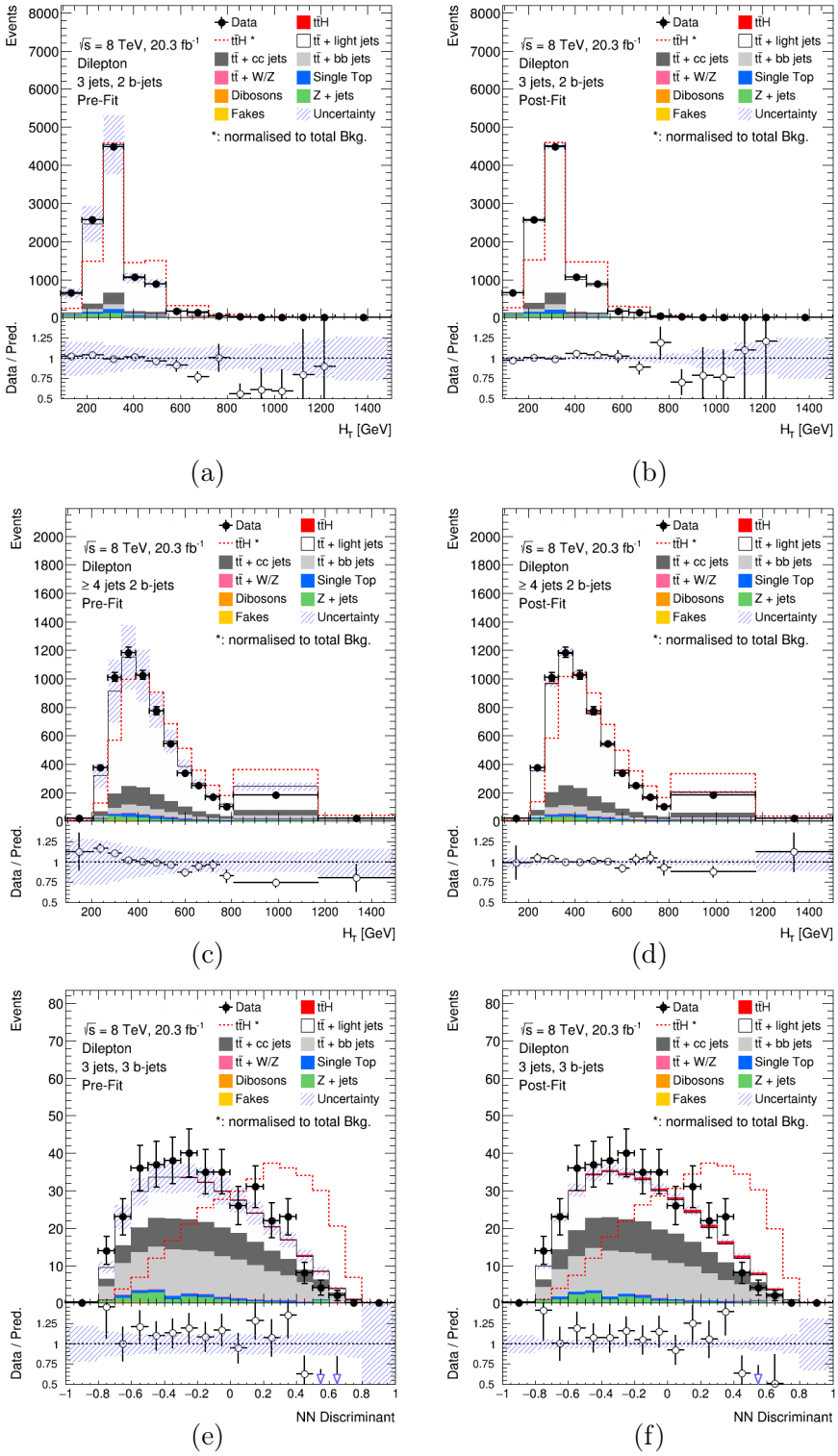


Figure 7.9: Comparison between data and prediction for the discriminant distributions on the background dominated regions before (left) and after (right) the fit: (a) and (b) for (3j, 2b); (c) and (d) for ($\geq$ 4j, 2b); and (e) and (f) for (3j, 3b). The $t\bar{t}H$ signal yield is shown normalised to the fitted μ (solid red histogram) and normalised to the total background prediction (hashed red line). The hashed area represents the total uncertainty on the background and the first (last) bin in all figures contains the underflow (overflow).

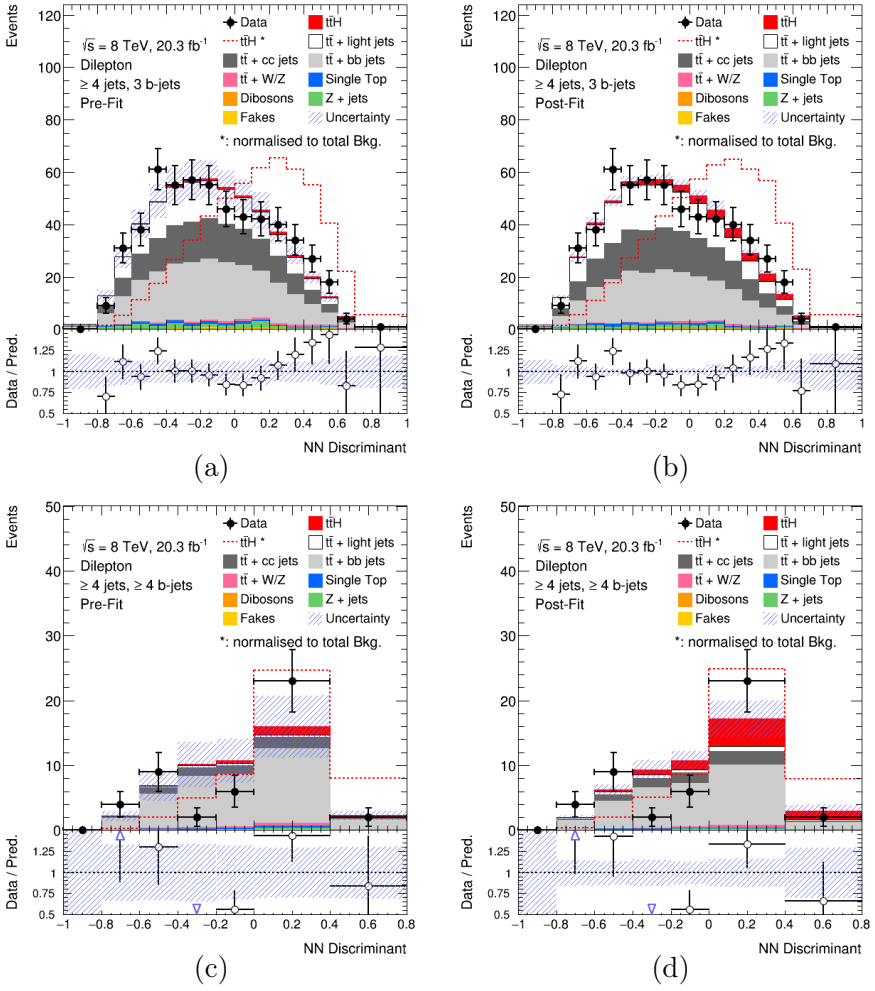


Figure 7.10: Comparison between data and prediction for the NN discriminant distributions on the signal-rich regions before (left) and after (right) the combined fit: (a) and (b) for ($\geq$ 4j, 3b); and (c) and (d) for ($\geq$ 4j, 4b). The $t\bar{t}H$ signal yield is shown normalised to the fitted μ (solid red histogram) and normalised to the total background prediction (hashed red line). The hashed area represents the uncertainty on the background and the first (last) bin in all figures contains the underflow (overflow).

$t\bar{t}V$ backgrounds are constrained by the uncertainties of the respective theoretical calculations, of the luminosity and of the data themselves. In addition, the anti-correlations among sources of systematic uncertainty, that result of the fit to the data, further restrain the total background uncertainty. For example, the highly populated (3j, 2b) provides a powerful constraint on the overall normalisation of the $t\bar{t}$ background, while both (3j, 2b) and ($\geq$ 4j, 2b) regions of the analysis, which are almost pure in $t\bar{t}$+light-jets background, provide an

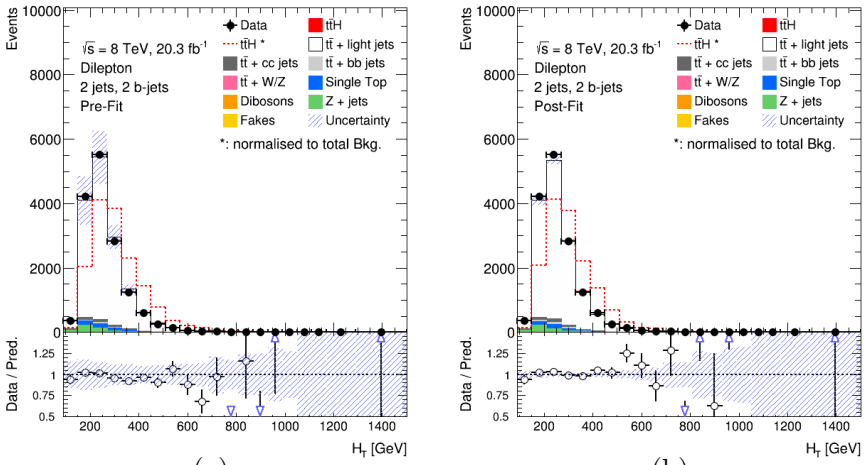


Figure 7.11: Comparison between data and prediction for the validation region (2j, 2b) before (a) and after (b) the statistical fit. The $t\bar{t}H$ signal yield is shown normalised to the fitted μ (solid red histogram) and normalised to the total background prediction (hashed red line). The hashed area represents the uncertainty on the background and the first (last) bin in all figures contains the underflow (overflow).

important constraint on $t\bar{t}$ modelling uncertainties both in terms of normalisation and shape. The fact that regions with exactly 3 and ≥ 4 b-jets have different fractions of $t\bar{t}$+$b\bar{b}$ jets and $t\bar{t}$+$c\bar{c}$ jets backgrounds, allow the possibility to constrain uncertainties on the normalisations of these processes. Ultimately, the combined fit improves the overall signal sensitivity of the analysis. Further confidence in the fit procedure is attained, since the agreement between MC simulation and data in the validation region (2j, 2b), which was not used in the fit, has also improved after the statistical fit, as can be seen in Figure 7.11.

A summary of the yields after the fit in the five fitted regions is included in Figure 7.12. Table 7.5 shows the event yields after the fit in all analysis regions, including the validation region (2j, 2b). The uncertainties are calculated, considering the correlations among nuisance parameters and among processes, as the sum in quadrature of statistical and systematic uncertainties on the yields. As previously mentioned, the uncertainties decrease significantly in all regions due to constraints provided by data and correlations between different sources of uncertainty introduced by the fit to the data.

Figure 7.13 summarises the fitted NPs under the signal-plus-background hypothesis. For each NP, the deviation of the fitted values $\hat{\theta}$ from its nominal value θ_0 is shown (black points) in units of the pre-fit standard deviation $\Delta\theta$. The post-fit uncertainty of each

	2j, 2b	**3j, 2b**	**3j, 3b**
$t\bar{t}H$	5.37 ± 3.39	19.2 ± 11.9	6.98 ± 4.63
$t\bar{t}$ + light jets	13 600 ± 331	8410 ± 149	115 ± 15.0
$t\bar{t} + c\bar{c}$ jets	340 ± 46.1	693 ± 69.6	88.6 ± 28.8
$t\bar{t} + b\bar{b}$ jets	177 ± 42.5	294 ± 71.6	116 ± 30.4
$t\bar{t} + W/Z$	7.67 ± 2.49	18.7 ± 6.1	1.68 ± 0.574
Single Top	394 ± 33.3	242 ± 21.5	6.24 ± 0.892
Dibosons	5.69 ± 3.19	3.44 ± 2.64	0.0571 ± 0.0538
Z + jets	502 ± 36.5	275 ± 23.9	12.7 ± 3.72
Fakes	21.8 ± 12.1	20.2 ± 11.3	0.754 ± 0.453
Total	15 000 ± 352	9980 ± 175	348 ± 14.3
Data	15296	9996	374

	$\geq$ 4j, 2b	**$\geq$ 4j, 3b**	**$\geq$ 4j, $\geq$ 4b**
$t\bar{t}H$	54.5 ± 33.6	29.2 ± 18.4	8.45 ± 5.66
$t\bar{t}$ + light jets	4550 ± 120	156 ± 19.3	2.32 ± 0.542
$t\bar{t} + c\bar{c}$ jets	764 ± 72.6	157 ± 51.6	6.82 ± 3.08
$t\bar{t} + b\bar{b}$ jets	299 ± 80.6	188 ± 52.2	29.0 ± 8.75
$t\bar{t} + W/Z$	39.4 ± 14.8	6.45 ± 2.78	0.868 ± 0.326
Single Top	121 ± 21.4	8.93 ± 2.78	0.611 ± 0.166
Dibosons	3.12 ± 2.18	0.179 ± 0.209	0.006 29 ± 0.0419
Z + jets	127 ± 20.4	13.3 ± 8.50	0.386 ± 0.492
Fakes	21.6 ± 12.1	3.47 ± 2.03	0.280 ± 0.200
Total	5980 ± 138	562 ± 22.0	48.8 ± 6.64
Data	6002	561	46

Table 7.5: Post-fit event yields for $t\bar{t}H$ signal, relevant backgrounds and observed data in all multiplicity bins considered in the analysis. The uncertainties on the expected event yields include both statistical and systematic sources.

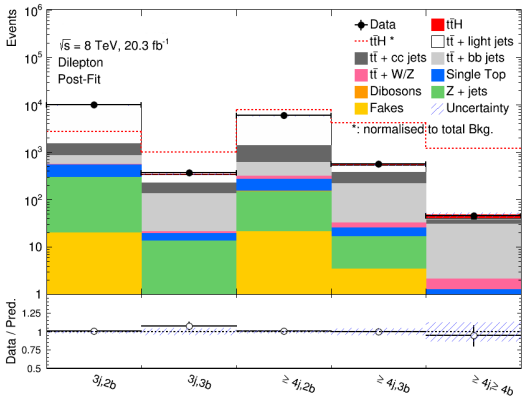


Figure 7.12: Post-fit summary of the MC prediction to data comparison after in all analysis regions. The signal is shown in red, both normalised to the SM prediction (filled red area stacked on the background) and normalised to the total number of background events (dashed red line). The total uncertainty on the yields is represented by the hashed area.

NP, σ_θ, is represented by the associated error bar in units of the prior uncertainty. The correlation values amongst the fitted NP are mostly below 5%. When data does not provide any further constraint on an uncertainty, σ_θ is close to unity. However, for values of σ_θ much smaller than one, a significant reduction or constraining with respect to the original uncertainty occurred. The most relevant pulls and constraints of the NP by the available data in the regions considered in the fit are mainly associated with the large uncertainties on $t\bar{t}$ modelling, and in particular the irreducible $t\bar{t} + b\bar{b}$ jets background.

For completeness, the effect of the various systematic uncertainties on the fitted value of μ, including the constraints provided by the data is shown in Figure 7.14. Only the systematic uncertainties with the largest impact on the measured signal strength are shown. Similarly to Figure 7.13, the deviation of each fitted parameter from its nominal value is represented (black points) in units of the pre-fit standard deviation $\Delta\theta$ and are measured by the scale in the bottom axis. The NP are sorted according to the post-fit effect each has on μ (filled blue and cyan areas) conforming to the scale of the top axis, with those with the largest impact at the top. The post-fit effect on μ is calculated by fixing the corresponding nuisance parameter at $\hat{\theta} \pm \sigma_\theta$ and performing the fit again. The effect a particular systematic uncertainty has on μ is then represented as the difference between the default and the modified signal strength, $\Delta\mu$.

The leading systematic uncertainty in the results of this thesis is the variation of the $t\bar{t}$ modelling (named 'tt_herwig' in the figures), which is not supported by data and is

significantly constrained to approximately 30% of its pre-fit value. The reduction of the systematic uncertainty was already somewhat expected, since it was a rather conservative approach based on 7 TeV data, which has less statistics than the dataset used for this thesis. The post-fit normalisations of each of these background processes change slightly, for example increasing the $t\bar{t} + c\bar{c}$ jets population in all the regions, always being within the uncertainty limits of the respective pre-fit values.

The $t\bar{t}$ + heavy flavour normalisation uncertainties are the second leading effect on the fitted signal strength μ. These uncertainties pertaining to the $t\bar{t}+c\bar{c}$ jets and $t\bar{t}+b\bar{b}$ jets backgrounds, are reduced by more than one half from the initial 50%, while the pull is marginally negative. The post-fit yields for $t\bar{t} + c\bar{c}$ jets and $t\bar{t} + b\bar{b}$ jets backgrounds change slightly, but remain in agreement with their pre-fit values within the total systematic and statistical uncertainties. This effect is caused by the interplay between the $t\bar{t}$ + heavy flavour normalisation uncertainties and several other systematic uncertainties affecting these background yields. Nevertheless, the available data allows the reducing of the uncertainty, improving the sensitivity of the search.

The $t\bar{t} + b\bar{b}$ jets modelling uncertainties affecting the shape of this background, and with no correlation with other $t\bar{t}$ backgrounds, are in agreement with the nominal prediction. The generator uncertainty ('ttbb_SHERPAOL') and the renormalisation scale uncertainty ('ttbb_R_Mbb' in the figures) have negative pulls and are constrained by the available data to a fraction of the respective pre-fit value. The uncertainty on the renormalisation scale is the ninth ranked nuisance parameter of the fit and the effect on μ is already very small.

Uncertainties arising from detector modelling, such as those related to jets, have also shown considerable pulls and constraints within the fit. The uncertainty on the jet flavour composition ('flavour_comp') affects analyses that are performed using a different flavour fraction than the sample used to derive the jet energy scale calibration. This is due to the jet energy response being different for quark-initiated jets than for gluon-initiated jets [289]. The NP has a negative pull, with constrained uncertainty, that agrees with the nominal prediction. The jet vertex fraction uncertainty ('JVF') is assessed by changing the JVF cut in such a way as to cover differences in the data to MC simulation agreement in a sample with one single jet. Here, the data in the fit does not support the total uncertainty of the simultaneous variation of the JVF cut for all the jets in the event, and as such constrains it to a fraction of its pre-fit effect. The jet energy resolution uncertainty (´JER') was conservatively estimated from the uncertainty for low transverse momenta jets as explained

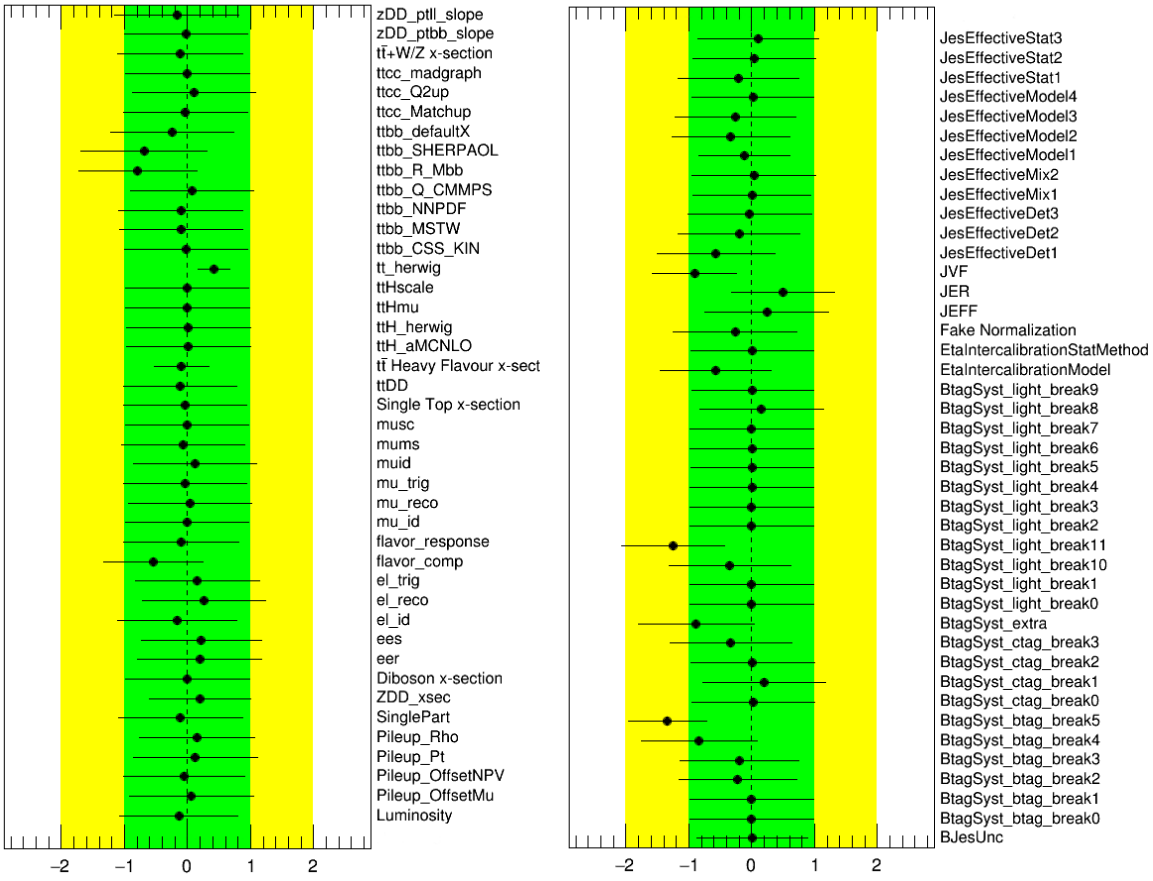


Figure 7.13: Fitted nuisance parameters under the signal-plus-background hypothesis. The deviation of each of the fitted parameters, $\hat{\theta}$, from its nominal value θ_0, is represented (black points) in units of the pre-fit standard deviation $\Delta\theta$. The associated error bar represents the post-fit uncertainty, σ_θ, in units of the prior uncertainty. The 1σ (2σ) error bands are represented in green (yellow).

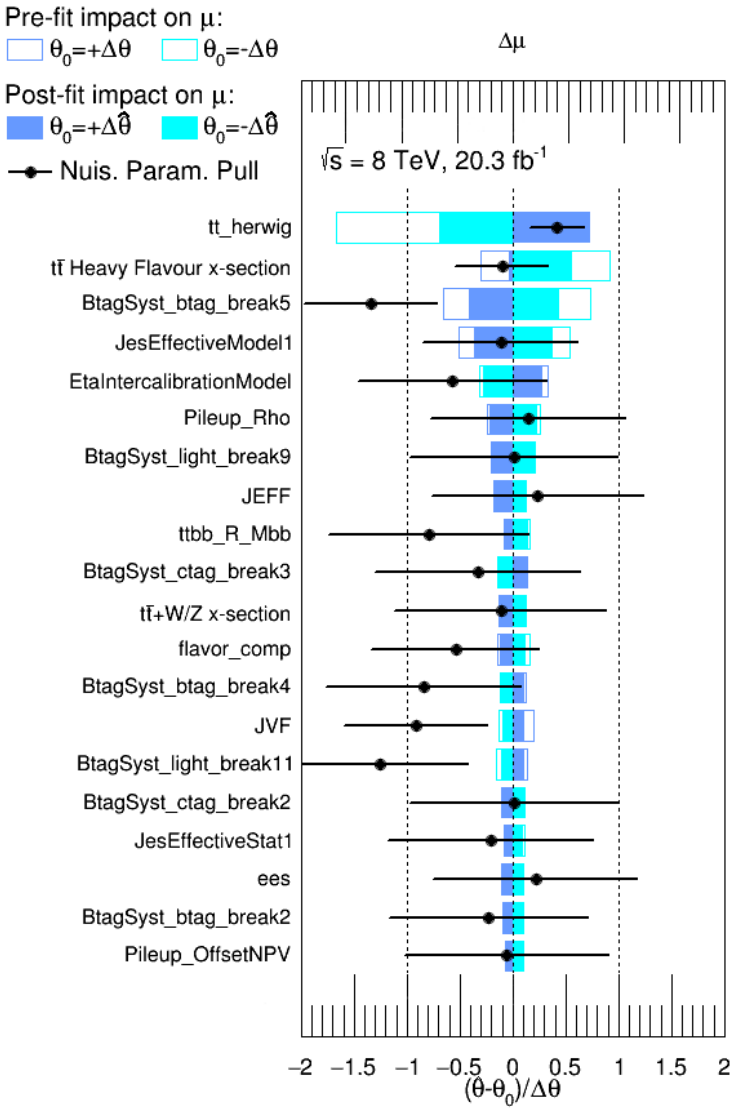


Figure 7.14: Fitted nuisance parameters under the signal-plus-background hypothesis, sorted according to their impact on the measured signal strength μ, with those with the largest impact at the top. The points, which are drawn conforming to the scale of the bottom axis, represent the deviation of each of the fitted nuisance parameters, $\hat{\theta}$, from its nominal value θ_0, in units of the pre-fit standard deviation $\Delta\theta$. The error bars show the post-fit uncertainties σ_θ. The post-fit effect of each NP on μ (filled blue and cyan areas) is conforming to the scale of the top axis.

in Section 5.4.2.3 and is constrained almost 50% of its initial values by the statistical fit, which uses the high multiplicity of jets available in the analysis. The big variations of most of the JES related uncertainties (whose nomenclature in the figures includes 'JES'

and 'EtaIntercalibration') are not allowed by the high data statistics of the analysis regions, which therefore constrains their values. All of these jet related uncertainties affect primarily low p_{T} jets, being constrained in the signal-depleted regions and do not have a significant effect on the fitted value of μ.

The noticeable pulls of some uncertainties related to the flavour tagging of jets in the analysis can be understood by the relatively small amount of data in the regions with the highest b-jets multiplicity. The light-tag systematic uncertainty 'BtagSyst_light_break11' is the component with the largest eigenvector after diagonalisation of the light-tagging uncertainties. Since the $\geq 4b$ region has limited data, the statistical fit cannot significantly constrain the uncertainty. Its impact on the fitted value of μ is quite small. The b-tagging eigenvector 5 ('BtagSyst_btag_break5'), corresponds to the largest eigenvector after diagonalisation of the b-tagging uncertainties, which produces a large variation that is not supported by data and is therefore slightly constrained by the data in the fit. Similarly to the previous systematic, being the largest eigenvector component, it is sensitive to the reduced availability of events in the $\geq 4b$ region, where possible mistagging can occur. Although this constitutes the third leading nuisance parameter of the fit, the impact on the signal strength μ is already sufficiently small.

The normalisation of the $t\bar{t} + W/Z$ background, with $Z \to b\bar{b}$, is among the highest ranked NP impacting the signal strength μ of the fit. Being an irreducible background of the $t\bar{t}H$ signal, with a very small contribution in most regions of the analysis which are mainly dominated by $t\bar{t}$+jets events, the uncertainty on the $t\bar{t} + W/Z$ normalisation cannot be significantly constrained by the fit. Nevertheless, its impact on μ is meagre.

The remainder systematic uncertainties do not affect the sensitivity of the analysis substantially, and have less significant pulls and constraints.

7.4 Limit on $t\bar{t}H$ Production

From the statistical fit performed over the five regions of the dileptonic analysis, on the signal-plus-background hypothesis, the obtained fitted value for the signal strength for $m_H = 125\,\mathrm{GeV}$ is:

$$\mu[m_H = 125\,(\mathrm{GeV})] = 3.7 \pm 2.2. \tag{7.9}$$

The expected uncertainty for the signal strength (assuming $\mu = 1$) is ± 2.1.

95% CL upper limit	Observed	-2σ	-1σ	Median	-1σ	$+2\sigma$	Median ($\mu = 1$)
Dilepton	7.8	2.4	3.3	4.5	6.4	8.8	4.6

Table 7.6: Observed and expected (median, for the background-only hypothesis) 95% CL upper limits on $\sigma(t\bar{t}H)$ relative to the SM prediction, assuming $m_H = 125\,\text{GeV}$. The 68% and 95% confidence intervals around the expected limits under the background-only hypothesis are also provided, denoted by $\pm 1\sigma$ and $\pm 2\sigma$, respectively. The expected (median) 95% CL upper limits assuming the SM prediction for $\sigma(t\bar{t}H)$ are shown in the last column.

In order to test the compatibility of data with the background-only hypothesis, the observed (expected) significance of the signal is 1.8 (0.5) standard deviations.

Since no significant excess over the background-only hypothesis is found, a 95% CL upper limit can be set on the signal strength modifier, as implemented in the RooFit package [290, 291]. The observed limits, those expected with and without assuming a SM Higgs boson with $m_H = 125\,\text{GeV}$, are shown in Table 7.6. A signal 7.7 times larger than predicted by the SM is excluded at 95% CL using the CLs method. The expected upper limit at a 95% CL is $\mu < 4.5$ under the background-only hypothesis. A signal 4.6 times larger than the SM prediction is expected to be excluded in the case a SM Higgs boson.

The combination of single lepton and dilepton analyses to search for the $t\bar{t}H\,(H \to b\bar{b})$ has been performed in ATLAS [1]. Using orthogonal datasets at $\sqrt{s} = 8\,\text{TeV}$ and equivalent analysis procedures in both channels, a combined fit is performed to nine regions of the single lepton search and six regions from the dilepton search. For this combined result, alternative methods to estimate the $t\bar{t}$+jets background to those described in Section 6.2.2 are applied as previously mentioned. These methods rely in shape modelling variables based on the top quark and top quark pair momenta, determined from 7 TeV data. Instead, in this thesis, no direct shape modelling correction of the $t\bar{t}$+jets is attempted and the yield estimation is based on 8 TeV data from the (2j, 2b) region, which is then used as validation of the dilepton fit. As a complementary approach, the method implemented in the combined analysis is used in this thesis as a systematic uncertainty of the $t\bar{t}$+jets modelling. The result of the combined fit is shown in Figure 7.15, including the individual single lepton and dilepton fits. Table 7.7 summarises the 95% CL exclusion limits with their corresponding error bands. A good agreement in the fitted values is observed between the individual

95% CL upper limit	Observed	-2σ	-1σ	Median	-1σ	$+2\sigma$	Median ($\mu = 1$)
Semi Lepton	3.6	1.4	1.9	2.6	3.7	4.9	3.6
Dilepton	6.7	2.2	3.0	4.1	5.8	7.7	4.7
Combination	3.4	1.2	1.6	2.2	3.0	4.1	3.1

Table 7.7: Summary of the combination of single lepton and dilepton observed and expected (median, for the background-only hypothesis) 95% CL upper limits on $\sigma(t\bar{t}H)$ relative to the SM prediction, assuming $m_H = 125\,\text{GeV}$ [1]. The 68% and 95% confidence intervals around the expected limits under the background-only hypothesis are also provided, denoted by $\pm1\sigma$ and $\pm2\sigma$, respectively. The expected (median) 95% CL upper limits assuming the SM prediction for $\sigma(t\bar{t}H)$ are shown in the last column.

and the combined analyses, as well as between the dileptonic analysis in this thesis. The combination of the two analysis improves the expected sensitivity, resulting in the exclusion at 95% CL of a signal 3.4 times larger than predicted by the SM. A signal 2.2 times larger than the SM prediction is expected to be excluded in the absence of the $t\bar{t}H$process, and 3.1 times larger than the SM prediction if the $t\bar{t}H$ process is present with SM strength. The combined analysis has a fitted signal strength of $\mu = 1.5 \pm 1.1$. The observed (expected) significance of the signal is 1.4 (1.1) standard deviations.

For completeness, the results of the CMS Collaboration for the equivalent 8 TeV analysis are included here[292]. The CMS individual dileptonic observed (background-only expected) exclusion limit at 95% CL is $\mu < 7.7$ (6.9), and the best-fit value of μ is $\mu = 1.0^{+3.3}_{-3.0}$. The combined single lepton and dilepton observed (background-only expected) limit at 95% CL is $\mu < 4.2$ (3.3), corresponding to a best-fit value $\mu = 1.2^{+1.6}_{-1.5}$.

Although, no evidence of a $t\bar{t}H$ signal is found within the 8 TeV statistics analysed, the dileptonic analysis documented in this thesis provides competitive results compared to those obtained at the same centre-of-mass energy and for the same final states.

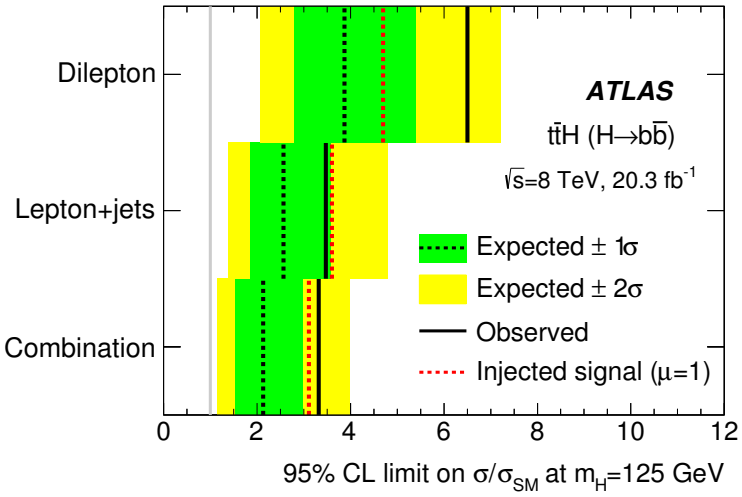


Figure 7.15: 95% CL upper limits on $\sigma(t\bar{t}H)$ relative to the SM prediction, σ/σ_{SM}, for the individual channels as well as their combination [1]. The observed limits (solid lines) are compared to the expected (median) limits under the background-only hypothesis and under the signal-plus-background hypothesis assuming the SM prediction for $\sigma(t\bar{t}H)$ and pre-fit prediction for the background. The surrounding shaded bands correspond to the 68 and 95% confidence intervals around the expected limits under the background-only hypothesis, denoted by $\pm 1\sigma$ and $\pm 2\sigma$, respectively.

8

Top Quark Couplings at the LHC

In this chapter, top quark couplings are explored in both neutral and charged currents. New observables are introduced that allow probing the nature of the top quark couplings to the Higgs boson (in Section 8.2) and to the W bosons (in Section 8.3). Although studied here at LHC energies, these observables can also be considered in future high luminosity datasets, with possible different discrimination powers.

8.1 Introduction

As already mentioned in Chapter 2, the Standard Model does not fully explain the matter/antimatter asymmetry of the Universe, for which new sources of CP violation are needed. The magnitude of the observed asymmetry could potentially be accounted for with BSM physics, for which the top quark, with a mass close to the electroweak scale, could be a sensitive probe. In particular, the top quark couplings, whether via neutral or charged processes, can provide a rigorous test of the SM framework and are a prolific playground for new physics. In this thesis, the top quark Yukawa coupling to the Higgs boson as well as the top quark electroweak coupling to the W boson are investigated considering different BSM scenarios, while angular based observables are proposed to gauge their nature.

The chapter is organised as follows. First, in Section 8.2, a study of the top quark Yukawa coupling to the Higgs boson is presented, which resulted in the publication of "Angular distributions in $t\bar{t}H(H \to b\bar{b})$ reconstructed events at the LHC" [3] and "Probing the CP nature of the Higgs coupling in $t\bar{t}h$ events at the LHC" [4]. Then, in Section 8.3, the nature of the top quark couplings in charged currents is probed using the polarisation of the top

quark, whose results were published in "New directions for top quark polarisation in the t-channel process" [5].

8.2 Top quark and Higgs boson Yukawa Coupling

The top quark-Higgs boson Yukawa coupling can be investigated using dileptonic $t\bar{t}h$ events, as the nature of this production process allows for a direct probe of the coupling. The possibility of CP violation can be guaranteed by considering BSM models where the Higgs boson, h, has no definite CP quantum number, such as supersymmetry (SUSY) or two Higgs doublet model (2HDM). The most general Yukawa coupling for the top quark is defined as,

$$\mathcal{L} = \kappa\, y_t\, \bar{t}\,(\cos(\alpha) + i\gamma_5 \sin(\alpha))\, t\, h, \tag{8.1}$$

where y_t is the SM Higgs Yukawa coupling and α represents a CP phase. The resulting Yukawa coupling has then two components — a CP even associated to a SM-like spin 0, parity even h boson, here named H boson for simplicity; and a CP odd with a pseudoscalar spin 0, parity odd h boson, here called A boson (see for example [293]). The SM Higgs boson can be restored with $\cos(\alpha) = \pm 1$, whilst the pure pseudoscalar is attained by setting $\cos(\alpha) = 0$. Departures from the SM nature of the Higgs boson can be accessed by comparing the kinematical properties of $t\bar{t}h$ samples with a SM $J^{CP} = 0^+$ Higgs boson (labelled $t\bar{t}H$) to samples of $t\bar{t}h$ signal with pure pseudoscalar $J^{CP} = 0^-$ Higgs boson (labelled $t\bar{t}A$). Given the state of the art results, the pure pseudoscalar scenario, $t\bar{t}A$, has been excluded at a 99% CL, with most results being consistent with a SM-like spin 0, parity even boson. A CP admixture of the two components is, however, still possible. Recently, ATLAS and CMS studied the CP-nature of the 125 GeV Higgs boson in $t\bar{t}H$ ($H \to \gamma\gamma$) events, resulting in the exclusion of the pure CP-odd hypothesis at 3.9σ and setting a 95% CL observed (expected) exclusion upper limit for the CP mixing angle of 43° (63°) [160]. The decay of the tau lepton can also be used to probe the ratio between the CP-even and the CP-odd components on the Yukawa coupling. The tau lepton CP mixing angle was measured by CMS, using $\sqrt{s} = 13$ TeV events, as $4° \pm 17°$, with an observed (expected) exclusion upper limit of 36° (55°) [158].

In this exploratory study, several angular observables are tested to gauge the mixing between the CP-even and the CP-odd components of the top quark Yukawa coupling to a 125 GeV Higgs boson. Most of the observables rely on kinematic and spatial information of

the particles produced in the hard scatter, i.e. t, $\bar{t}$ and h, which are only accessible experimentally through a reconstruction algorithm. In Section 8.2.1 a simple $t\bar{t}h$ reconstruction method, built specifically for this purpose, is explained. A set of new observables are proposed and thoroughly investigated in section 8.2.2, while also testing several others already presented in literature. The $t\bar{t}H$ signal discrimination against the irreducible $t\bar{t}b\bar{b}$ is investigated using these observables. The CP-even and CP-odd components in the $t\bar{t}h$ process are probed in Section 8.2.3, using angular distributions of the decay products that present good discrimination between the scalar and the pseudoscalar components of the Higgs.

8.2.1 $t\bar{t}H$ Event Reconstruction

The $t\bar{t}H$ dileptonic events pose quite a challenging kinematic reconstruction. Despite having a very clean signature with two oppositely charged leptons, the final state events are substantially busy, with two neutrinos escaping direct detection and high jet activity, particularly in the $H \to b\bar{b}$ decay channel where the high multiplicity of b quarks leads to an overwhelming presence of jets in the detector. In this study, the kinematic reconstruction of $t\bar{t}H$ ($H \to b\bar{b}$) events in the dilepton channel of the top quark pair decay is done by fully reconstructing the undetected neutrinos and matching all final state particles to their possible mother particles, *i.e.* top quark, top antiquark and Higgs boson. Although challenging, the full kinematic reconstruction of the $t\bar{t}H$ dileptonic events allows the access to spatial information, such as the angular distribution of the particles resulting from the hard scatter which would otherwise be unreachable. The chosen reconstruction method consists on a kinematic fit to the events using mass constraints and energy-momentum conservation. The kinematic fit was developed for the $t\bar{t}H$ analysis framework within this thesis and used here to provide access to the spin information of the t, $\bar{t}$ and H produced in the hard scattered interaction.

The $t\bar{t}H(H \to b\bar{b})$ dileptonic decay has a final state signature with four b-jets, two charged leptons, and two neutrinos. In order to reconstruct these events, final states with exactly two charged leptons and at least four reconstructed jets are required. Only leptons and jets with $p_T \geq 20\,\text{GeV}$ and $\eta \leq 2.5$ are considered. In the reconstruction method, no b-tagging is used to identify the flavoured jets.[1] Despite being used in the reconstruction to account for the sum of the momenta of the two neutrinos in the $x \times y$ plane, no additional

[1] However, in the following studies b-tag requirements are occasionally needed. Unless stated otherwise, no b-tagging prerequisite is made.

selection cuts are applied to the missing transverse energy of the events, E_T^{miss}. Considering only the measured information − the masses of the detectable particles, the momenta of charged leptons and jets, and the missing energy in the transverse plane − is not enough to directly reconstruct the event kinematics, since the presence of two neutrinos creates an under-constrained kinematic system. The reconstruction fit takes into consideration additional constraints, allowing the full reconstruction of the $t\bar{t}H$ dilepton system.

The performance of the kinematic reconstruction algorithm was done at simulation level using $t\bar{t}H$ events generated at LO using MADGRAPH5_aMC@NLO [137] with the NNPDF2.3 PDF sets [76] at $\sqrt{s} = 13\,\text{TeV}$. The event generation was done with the default dynamic factorisation and renormalisation scales. Also set to the generator default values were the masses of the top quark and the SM Higgs boson. The Higgs boson is assumed to be a scalar particle ($CP = 1$) according to the SM. The decay of the heavy particles is processed by MADSPIN [294] which preserves the full spin correlations of the $t \to bW^+ \to b\ell^+\nu_\ell$, $\bar{t} \to \bar{b}W^- \to \bar{b}\ell^-\bar{\nu}_\ell$ and $h \to b\bar{b}$ decays, with $\ell^\pm \in \{e^\pm, \mu^\pm\}$. The shower and hadronisation processes were simulated with PYTHIA 6 [201] and fast detector-simulation was performed using DELPHES [295] with the ATLAS detector as default.

Neutrino Reconstruction

Assuming the neutrinos are the only source of missing transverse energy, a kinematic solution for the x and y components of the neutrinos four momenta can be found:

$$p_x^\nu + p_x^{\bar{\nu}} = E_x^{\text{miss}}\,, \tag{8.2}$$

$$p_y^\nu + p_y^{\bar{\nu}} = E_y^{\text{miss}}\,, \tag{8.3}$$

where p^ν ($p^{\bar{\nu}}$) represents the neutrino (antineutrino) momentum and $E_{x,y}^{\text{miss}}$ the components of the transverse missing energy. The masses of the $W^\pm$ bosons and the top quark/antiquark (m_W and m_t respectively) can be used to constrain the dileptonic decay of the top quark pair, such that the full momenta of the two neutrinos can be obtained from:

$$(p_{\ell+} + p_\nu)^2 = m_W^2\,, \tag{8.4}$$

$$(p_{\ell-} + p_{\bar{\nu}})^2 = m_W^2\,, \tag{8.5}$$

$$(p_{W^+} + p_b)^2 = m_t^2\,, \tag{8.6}$$

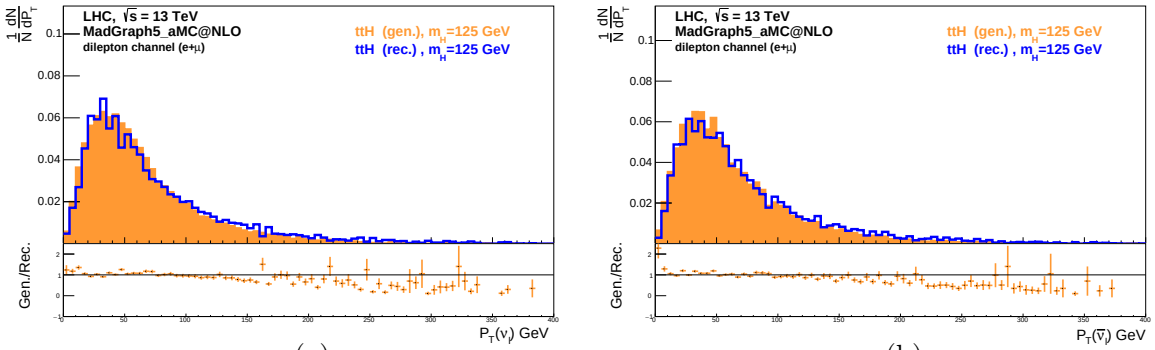
$$(p_{W^-} + p_{\bar{b}})^2 = m_t^2\,, \tag{8.7}$$

where $p_{\ell+}$ and p_b ($p_{\ell-}$ and $p_{\bar{b}}$) correspond to the lepton and b-jet (anti-lepton and b-jet) four momenta, respectively from the t ($\bar{t}$) decay. The masses of the intermediate particles W^+ and W^-, as well as those of the t and $\bar{t}$, are randomly generated using probability density functions obtained from the corresponding generator-level mass distributions. The correlations between the masses of these particles are preserved in the reconstruction by generating random mass values for t and $\bar{t}$ from a two-dimensional PDF for $(m_t, m_{\bar{t}})$, while generating the mass values of W^+ and W^- from two-dimensional PDFs of (m_t, m_{W^+}) and $(m_{\bar{t}}, m_{W^-})$, respectively.

The neutrino momentum is then obtained by solving equations 8.2–8.7 for the randomly generated masses, resulting in a quadratic equation with up solutions solutions per neutrino/anti-neutrino. When no solution is found for the particular combination of top quark and W boson masses, the generation of mass values is repeated until a solution is obtained. The event is only discarded as not compatible with the topology under study after 500 mass values are attempted without obtaining at least one valid solution. Finally, a probability is given to each pair of neutrino and anti-neutrino solutions based on parton level density functions of the neutrino/anti-neutrino transverse momenta, $P(p_{\mathrm{T}\nu})$ and $P(p_{\mathrm{T}\bar{\nu}})$, respectively. However, having assumed that the neutrinos are the only source of missing transverse energy, is saying that the momenta of the neutrino and anti-neutrino also accommodate any energy loss in the event, be it by QCD radiation or detector effects. This may result in larger estimated neutrino and anti-neutrino p_{T} reconstruction, relatively to their transverse momenta at parton level (or real momenta in collision data). The higher probability is thus given to the solution that minimises the product of the reconstructed neutrino and anti-neutrino transverse momenta, *i.e.* the solution with higher $\big(P(p_{\mathrm{T}\nu})P(p_{\mathrm{T}\bar{\nu}})\big)/\big(p_{\mathrm{T}\nu}p_{\mathrm{T}\bar{\nu}}\big)$.

In order to evaluate the performance of the reconstruction of the two neutrinos, the algorithm is first tested on MC simulated events by using truth-matched objects, i.e., jets and leptons which are matched respectively to their parton-level generated quarks and charged leptons using a ΔR criterion[1] to ensure the matching. The efficiency using truth-matched objects is 62%. In Figure 8.1 the truth-matched reconstructed transverse momentum distributions for the neutrino (a) and antineutrino (b) are shown. The performance of the

[1]The minimum distance in the pseudorapidity-azimuthal angle plane, ΔR, between the reconstructed jet or lepton and the parton-level quark or charged lepton: $\Delta R < 0.4$ for jets and $\Delta R < 0.1$ for leptons, respectively.



(a) (b)

Figure 8.1: Distributions of the transverse momentum of the neutrino (a) and antineutrino (b), where the kinematical fit reconstruction with truth match (full line) is compared with the generated distribution (shaded region) [3].

truth-matched reconstruction (solid lines) can be gauged by comparing with the respective MC generated distribution (filled histograms), as well as the ratio between the two. The reconstructed p_{T} distributions have a very good agreement with the parton-level neutrino distributions, thus corroborating the kinematical reconstruction of both neutrinos in $t\bar{t}H$ events. Due to radiation effects, a slight slope is visible in the distribution ratios for high p_{T}, which was not explicitly corrected here.

Jet-Lepton Pairing

In order to reconstruct the top quark (antiquark), the b quark ($\bar{b}$) must be paired with the charged lepton from the same decay, *i.e.* two $b\ell$ pairings from all possible combinations of reconstructed jets and charged leptons must be chosen. Since the lepton with positive (negative) charge is associated with the top quark (antiquark), the t and the $\bar{t}$ will be fully reconstructed once the $b\ell$ pairings are done. Using equations 8.2–8.7, the probability $P_{t\bar{t}}$ of each jet-lepton combination per event to be compatible with a dileptonic $t\bar{t}$ decay is assessed. Again, assuming the p_T of the neutrino and anti-neutrino, the mass of the $W^{\pm}$ bosons and the mass of top quark and antiquark to be distributed at parton level as in the SM, probability density functions are used to estimate $P_{t\bar{t}}$ for each possible $b\ell$ pairing and neutrino solution per event:

$$P_{t\bar{t}} \sim \frac{P(p_{\mathrm{T}\nu})P(p_{\mathrm{T}\bar{\nu}})}{p_{\mathrm{T}\nu}p_{\mathrm{T}\bar{\nu}}} P(m_t)P(m_{\bar{t}})P(m_{W^-})P(m_{W^+})\,. \tag{8.8}$$

Here, $P(m_t)$, $P(m_{\bar{t}})$, $P(m_{W^-})$ and $P(m_{W^+})$ are the probabilities given to each specific combination from the respective PDF distributions. The set of two jet-lepton pairs that maximise $P_{t\bar{t}}$ are then chosen as the correct $b\ell$ pairings originating from the top quark and antiquark.

$H \to b\bar{b}$ Reconstruction

The Higgs boson is reconstructed considering the $H \to b\bar{b}$ decay by selecting the two-jet combination, among the ones not used in the $t\bar{t}$ reconstruction, that best matches jets from a Higgs boson decay. There are several possible methods to do this selection and, on a first attempt, the kinematic characteristics of the $t\bar{t}H$ decay were explored, specifically the spatial distribution of particles in the transverse plane where in a first order approximation $p_{\mathrm{T}H} = -p_{\mathrm{T}t\bar{t}}$ for energy-momentum conservation. Given, however, the dependence on the neutrino reconstruction, a simpler relation was considered to estimate the likelihood, P_H, of a specific jet pairing to be originated from the Higgs boson decay,

$$P_H = \frac{1}{\left|\sqrt{(p_i + p_j)^2} - m_H\right|}, \tag{8.9}$$

related to how close the expected SM Higgs boson mass ($m_H = 125$ GeV) is to the invariant mass of each particular jet-pair combination. This criteria can obviously lead to an increase selection of incorrect jet assignment combinations within the same mass window of the Higgs boson, but the kinematics of this combinatorial background is in most cases distinct from the right combinations.

The Dilepton $t\bar{t}H$ ($H \to b\bar{b}$) Likelihood

For every combination of jet and lepton pair, neutrino solution and two-jet pair assignment to the Higgs boson decay, the likelihood $L_{t\bar{t}H}$ of it being originated from a $t\bar{t}H$ dileptonic decay is evaluated based on the kinematic information of the reconstructed objects:

$$\begin{aligned} L_{t\bar{t}H} &= P_{t\bar{t}} \times P_H \\ &\sim \frac{P(p_{\mathrm{T}\nu})P(p_{\mathrm{T}\bar{\nu}})}{p_{\mathrm{T}\nu}p_{\mathrm{T}\bar{\nu}}} P(m_t)P(m_{\bar{t}})P(m_{W^-})P(m_{W^+}) \times \frac{1}{|\sqrt{(p_i + p_j)^2} - m_H|}\,. \end{aligned} \tag{8.10}$$

The solution which has higher $L_{t\bar{t}H}$ compared to remaining ones is then chosen for the full kinematical reconstruction of the events, which fixes the assignment of jets and charged leptons to their parent particles.

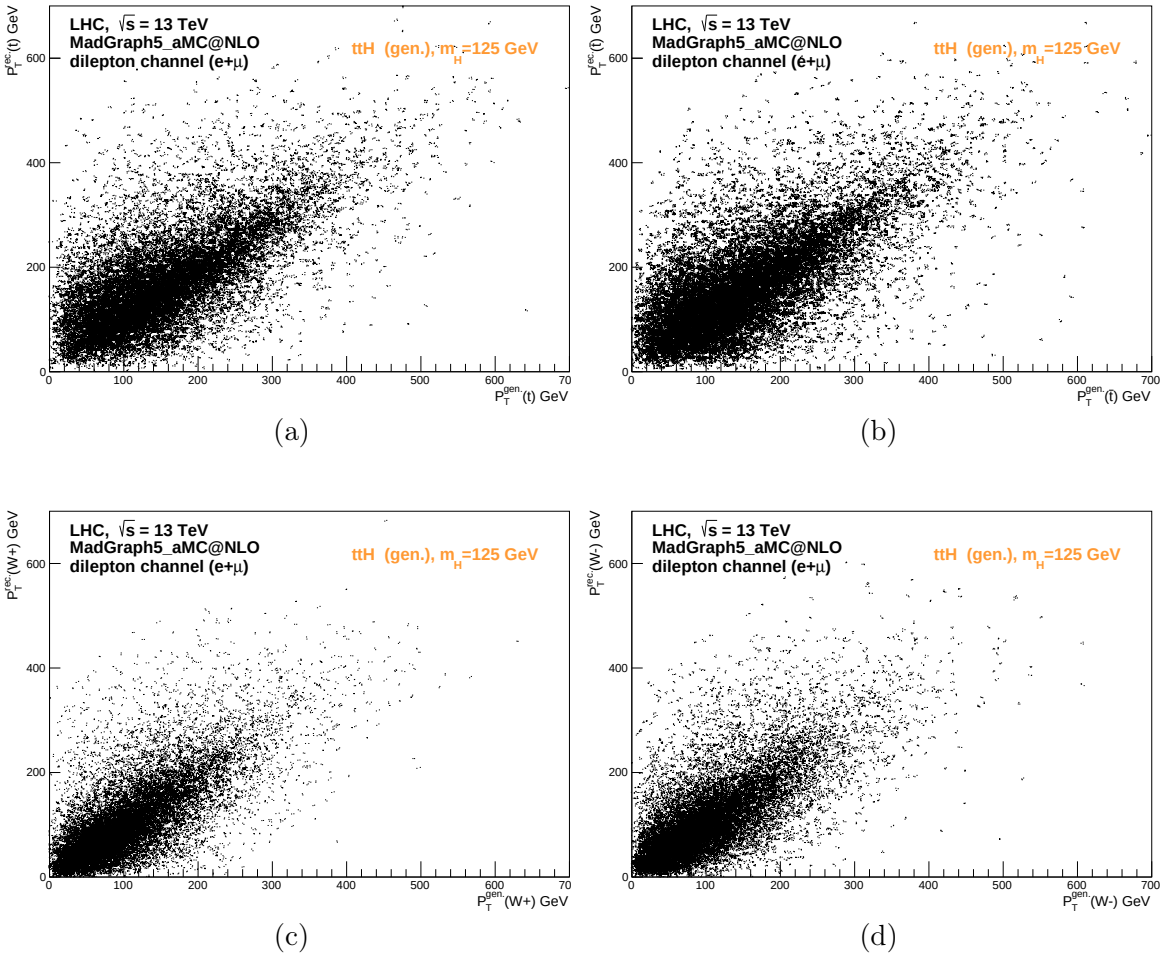


Figure 8.2: Distributions of the reconstructed transverse momentum, $p_{\mathrm{T}}^{\mathrm{rec}}$, of the top quark (a), the top antiquark (b), the W^+ boson (c) and the W^- boson (d) using the kinematical fit (without truth match) as a function of the $p_{\mathrm{T}}^{\mathrm{gen}}$ at parton level [3].

To evaluate the performance of the full kinematic reconstruction, an efficiency rate is determined by comparing the number of reconstructed events with the number of parton-level dileptonic $t\bar{t}H$ events in the MC sample. Here, the kinematic fit is applied to MC simulated events after detector simulation, just as it would be to collider data. The kinematic fit reconstructs 88% of all events, choosing the best $b\ell$ combination for each neutrino solution and assigning them to either t or $\bar{t}$, while also selecting the best jet combination for a 125 GeV Higgs boson. The transverse momentum distributions of the reconstructed top quarks and

W bosons are shown in Figure 8.2. The reconstructed transverse momentum, $p_{\mathrm{T}}^{\mathrm{rec}}$, without any jet or lepton truth matching, is compared to the true parton-level distribution. Visible correlations between the kinematically reconstructed distributions and the parton-level ones ensure the reasonable efficiency of the full kinematic reconstruction.

8.2.2 Angular Distributions in $t\bar{t}H$

Angular variables can be used to further understand the dynamics of the $t\bar{t}H$ process, providing not only insight on the signal itself, but perhaps alternative ways to discriminate the $t\bar{t}H$ ($H \to b\bar{b}$) signal from the main irreducible background process $t\bar{t}b\bar{b}$. Having fully reconstructed the $t\bar{t}H$ final states in the dileptonic topology, grants access to new observables which rely on spin information (such as the presence or lack of spin correlations)[296]–[294], where the top quark and antiquark are natural spin analysers. The sensitivity to new angular distributions and asymmetries is investigated in the present study.

In this analysis, angular distributions involving three-dimensional angles between the decay products of the $t\bar{t}H$ dileptonic final states are considered. After the full reconstruction of events, two frames of reference are defined:

Frame 1 — the $t\bar{t}H$ centre-of-mass system (determined by using the laboratory four-momenta);

Frame 2 — the $\bar{t}H$ centre-of-mass system recoiling against the t quark in the $t\bar{t}H$ system (i.e. Frame 1).

Considering the complete $t\bar{t}H$ decay chain[1] various angles can be investigated. θ_Y^X is then defined as the angle between the direction of the Y system in the rest frame of X and the direction of the X system in the rest frame of its parent system. For example, $\theta_H^{\bar{t}H}$ is defined as the angle between the direction of the momentum of the Higgs boson (in the $\bar{t}H$ centre-of-mass, *i.e.* Frame 2) and the $\bar{t}H$ direction (in the $t\bar{t}H$ reference system, *i.e.* Frame 1). The momentum of Y in the centre-of-mass of X is determined using the laboratory four-momenta in a direct rotation-free boost. The direction of the momentum of the $t\bar{t}H$ system is measured with respect to the laboratory frame.

In order to evaluate how the spin information can be used to improve the $t\bar{t}H$ analysis, comparisons between MC predictions of the $t\bar{t}H$ signal and the main background $t\bar{t}b\bar{b}$ are

[1]Not only the final state particles, such as ℓ^+,ℓ^-, b and $\bar{b}$ jets, but also intermediate objects like the W^+ and W^- bosons, as well as two body systems such as $\bar{t}h$ are considered.

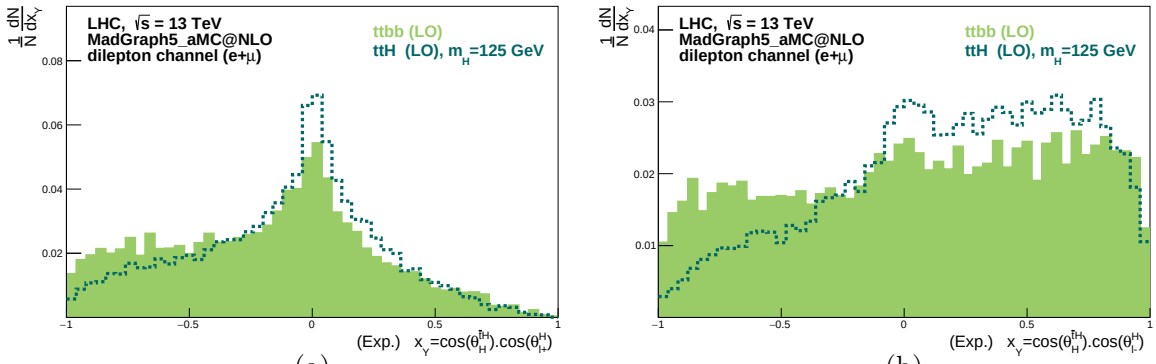


Figure 8.3: Double angular product distributions after the event selection and full kinematic reconstruction. The distributions represent the product of the cosine of $\theta_H^{\bar{t}H}$ and the cosine of $\theta_{\ell^+}^H$ (a) or $\theta_{\ell^-}^H$ (b).

investigated using simulated pp collisions at $\sqrt{s} = 13\,\text{TeV}$. Here, only the $t\bar{t}H$ signal and the main background $t\bar{t}b\bar{b}$ are considered for simplicity. Both processes are generated for the LHC at LO with the same setup as the one used to test the kinematic reconstruction performance. A similar event selection is also required, followed by the full kinematic reconstruction described above. It is important to note that the $t\bar{t}H$ dileptonic event selection and kinematic reconstruction preserve the information of the particles produced in the hard-scatter, in particular the spin properties present in the matrix element. To demonstrate this, a comparison of several angular distributions is shown in Appendix C.

Following an extensive survey of angular distributions, double angular products such as

$$x_Y = \cos\left(\theta_H^{\bar{t}H}\right) \times \cos\left(\theta_Y^H\right), \tag{8.11}$$

where Y can be any electrically charged decay product (ℓ^+,ℓ^-, b, $\bar{b}$, W^+ or W^-), were found to provide discrimination between the $t\bar{t}H$ signal and its main background $t\bar{t}b\bar{b}$. As an example, Figure 8.3 displays x_{ℓ^+} on the left and x_{ℓ^-} on the right. Here, as previously stated, $\theta_H^{\bar{t}H}$ is the angle between the direction of the momentum of the H boson (in the $\bar{t}H$ centre-of-mass frame) with respect to the direction of the momentum of the $\bar{t}H$ system (in the $t\bar{t}H$ centre-of-mass frame), whereas $\theta_{\ell^\pm}^H$ is the angle between the direction of the momentum of the ℓ^+(left) or ℓ^-(right) in the H centre-of-mass system, and the direction of the H boson momentum in the $\bar{t}H$ centre-of-mass frame. The four-momenta obtained after applying the full kinematic fit reconstruction is used to construct the angular variables, without any

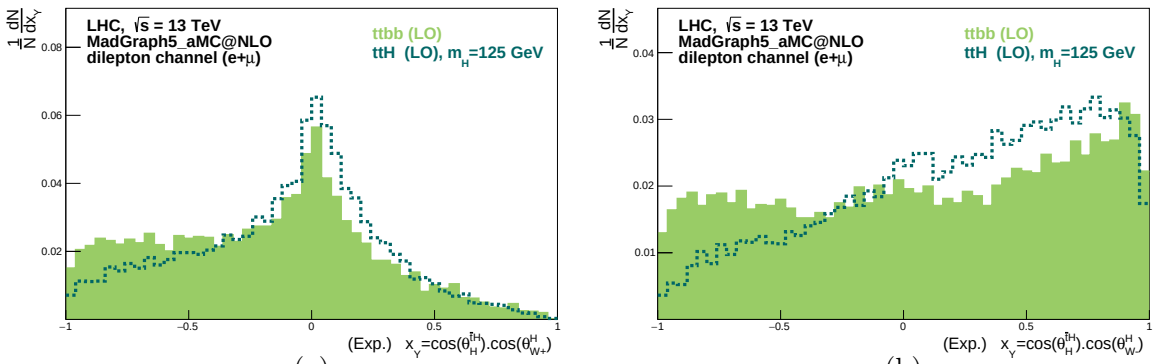


Figure 8.4: Angular distribution $\cos(\theta_H^{\bar{t}H}) \times \cos(\theta_{W^\pm}^H)$ after event selection and full kinematic reconstruction. $\theta_H^{\bar{t}H}$ is the angle between the direction of the momentum of the H boson (in the $\bar{t}H$ centre-of-mass frame) with respect to the direction of the momentum of the $\bar{t}H$ system (in the $t\bar{t}H$ centre-of-mass frame), while $\theta_{W^\pm}^H$ is the angle between the direction of the momentum of the W^+ (a) or W^- (b) (in the H centre-of-mass system) and the direction of the H boson momentum (in the $\bar{t}H$ centre-of-mass frame).

truth matched object, just as it would be done in collision data. In these distributions, as well as in the following ones, signal and background are normalised to unity. The distinct behaviour of the two processes is quite visible. Indeed, having a kinematic reconstruction fit allows the direct access to the original spin information of each event, which opens the possibility to explore angular differences among processes whose final signatures are identical in every other way. Figures 8.3 (a) and (b) show that different angular observables can have various sensitivity to the different spin information in the $t\bar{t}H$ signal and the $t\bar{t}b\bar{b}$ background events: where the $\cos(\theta_H^{\bar{t}H}) \times \cos(\theta_{\ell^+}^H)$ distribution (e) shows similar signal and background distributions, the $\cos(\theta_H^{\bar{t}H}) \times \cos(\theta_{\ell^-}^H)$ distribution (f) seems to have a much more contrasting behaviour between the $t\bar{t}H$ signal and the $t\bar{t}b\bar{b}$ background. Similarly, in Figures 8.4 and 8.5 $\cos(\theta_H^{\bar{t}H}) \times \cos(\theta_Y^H)$ distributions are shown for $Y = W^\pm$ (the direction of the momentum of the $W^\pm$ boson) and $Y = b_H, \bar{b}_H$ (the direction of the momentum of the b jets originated from the Higgs decay), respectively. Again it is noticeable that some angular distributions provide higher discrimination between signal and background than others.

In order to further explore the contrasting behaviour of signal and background, several forward-backward asymmetries, A_{FB}^Y, are defined using the double angular product (Ex-

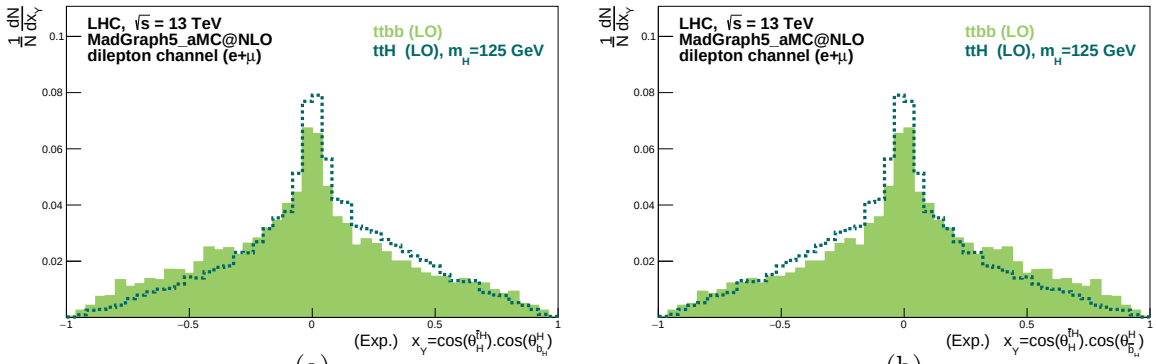


Figure 8.5: Angular distributions (a) $\cos(\theta_H^{\bar{t}H}) \times \cos(\theta_{b_H}^H)$ and (b) $\cos(\theta_H^{\bar{t}H}) \times \cos(\theta_{\bar{b}_H}^H)$ after event selection and full kinematic reconstruction. $\theta_H^{\bar{t}H}$ is the angle between the direction of the momentum of the H boson (in the $\bar{t}H$ centre-of-mass frame) with respect to the direction of the momentum of the $\bar{t}H$ system (in the $t\bar{t}H$ centre-of-mass frame), while $\theta_{b_H}^H$ ($\theta_{\bar{b}_H}^H$) is the angle between the direction of the momentum of the b ($\bar{b}$) jet coming from the Higgs boson (in the H centre-of-mass system) and the direction of the H boson momentum (in the $\bar{t}H$ centre-of-mass frame).

pression 8.11):

$$A_{FB}^{Y} = \frac{N(x_Y > 0) - N(x_Y < 0)}{N(x_Y > 0) + N(x_Y < 0)}, \tag{8.12}$$

where $N(x_Y > 0)$ ($N(x_Y < 0)$) are the total number of events in the corresponding angular distribution with x_Y above (below) zero. These asymmetries can have quite different values for $t\bar{t}H$ signal and $t\bar{t}b\bar{b}$ background immediately at generation level.

In Table 8.1, the values of LO asymmetries for different final state particles (Y) boosted to the centre of mass of the Higgs boson are shown, both at parton level (without any selection cuts applied to the events) and after the kinematic fit reconstruction (without any truth-matched particle). Among these results, some asymmetry values are evidently different between signal and background, such as $A_{FB}^{\ell^-}$, $A_{FB}^{W^-}$, $A_{FB}^{\bar{b}}$ with $\bar{b}$ from $\bar{t}$, even after the kinematical reconstruction fit. Please note that the asymmetries using the b and $\bar{b}$ from the Higgs boson decay, which are zero at parton level for both signal and background, due to a non-perfect reconstruction of the events, yield misleading non-zero values at the reconstructed level.

On the whole, a $t\bar{t}H$ analysis focused on final states in the dileptonic topology is viable and, even with a reconstruction which is not optimised, can still be sensitive to the angular

(Asymmetries @ LO)	Parton level		Reconstruction	
	$t\bar{t}H$	$t\bar{t}b\bar{b}$	$t\bar{t}H$	$t\bar{t}b\bar{b}$
$A_{FB}^{Y=\ell+}$	−0.157	−0.137	−0.141	−0.268
$A_{FB}^{Y=\ell-}$	+0.291	+0.056	+0.331	+0.118
$A_{FB}^{Y=W+}$	−0.154	−0.119	−0.119	−0.275
$A_{FB}^{Y=W-}$	+0.317	+0.067	+0.348	+0.127
$A_{FB}^{Y=b}$ (b from t)	−0.155	−0.141	−0.179	−0.306
$A_{FB}^{Y=\bar{b}}$ ($\bar{b}$ from $\bar{t}$)	+0.293	+0.053	+0.334	+0.117
$A_{FB}^{Y=b}$ (b from H)	+0.000	+0.001	+0.086	−0.048
$A_{FB}^{Y=\bar{b}}$ ($\bar{b}$ from H)	+0.000	−0.001	−0.086	+0.048

Table 8.1: Asymmetry values for $t\bar{t}H$ and $t\bar{t}b\bar{b}$ events at the LHC. The second and third column show the observed asymmetries at the parton level (without any cuts), while the fourth and last column show same asymmetries after applying the selection cuts and the kinematical reconstruction (without truth match). These asymmetries have a statistical uncertainty smaller than 1%.

distributions and asymmetries. These observables retain their kinematic properties and show significant differences between the $t\bar{t}H$ signal and the $t\bar{t}$ + jets background, regardless of the small distortions that naturally occur in the distributions throughout a full analysis. As a result, angular variables, such as these, can contribute to a discriminant analysis of $t\bar{t}H$ events.

8.2.3 Sensitivity to the CP Nature of $t\bar{t}h$

The sensitivity provided by the angular distributions to the CP nature of the top Yukawa coupling to the Higgs boson can be explored by comparing $t\bar{t}h$ signal samples with a SM-like Higgs boson to samples with a pure pseudoscalar Higgs boson. In order to cover a wide range of CP states — from purely CP-even to purely CP-odd — several $t\bar{t}h$ signal samples were produced, varying $\cos(\alpha)$ (introduced in Expression 8.1) incrementally from -1 to 1, in steps of 0.1. For this study, NLO precision $t\bar{t}h$ events are used, which were generated using MadGraph5_aMC@NLO [137] and the NNPDF2.3 PDF sets [76]. The default `sm` model in MadGraph_aMC@NLO was used for the SM $t\bar{t}H$ signal generation, whereas for samples with a non-zero CP-odd Higgs component ($t\bar{t}A$) the generation was done using the `HC_NLO_X0` model [300]. Adjustments of effective couplings between the Higgs boson and any vector boson were set to zero, except $H_{\gamma\gamma}$, $A_{\gamma\gamma}$, $H_{Z\gamma}$ and $A_{Z\gamma}$.

All the relevant SM processes that can originate similar final states to the signal ones are considered. The dominant background process, $t\bar{t}b\bar{b}$, was generated at NLO with the

same generator and PDF set as the $t\bar{t}h$ signal. The remaining background processes were simulated at LO with MadGraph5_aMC@NLO [137]: $t\bar{t} + jets$ including up to three additional c-jets or light-flavoured jets; t-channel, s-channel, and Wt channel (including up to one extra jet) of single top quark production; $t\bar{t}V + jets$, where $V = Z, W^{\pm}$ (including up to one extra jet); diboson ($WW, WZ, ZZ + jets$ with up to three additional jets); $W + jets$ and $Z + jets$ (with up to four additional jets); and $Wb\bar{b} + jets$ and $Zb\bar{b} + jets$ (including up to two extra jets). The $t\bar{t} + jets$ was normalised to the QCD NNLO cross-section with NNLL resummation of soft gluons [70, 71, 73, 75, 76]. For the single top quark production cross-section a normalisation to the approximate NNLO theoretical predictions [77, 78] was done following the prescription defined in [69], assuming the NNPDF2.3 PDF sets.

MC simulated events were generated for LHC pp collisions at $\sqrt{s} = 13\,\text{TeV}$, similarly to what was done in the previous section, with MadSpin [294] preserving the spin information. The renormalisation and factorisation scales were set to default, as were the masses of the W boson (m_W), the top quark (m_t), and Higgs bosons (both scalar, m_H, and pseudoscalar, m_A). The parton shower and hadronisation processes were also done using Pythia6 [201], with the matching between generator and parton shower performed by the MLM [301] scheme for LO events and the MC@NLO [302] for NLO events. Fast simulation of a general-purpose collider experiment was implemented with the Delphes [295] package, using the default ATLAS parameter card.

A selection of dileptonic $t\bar{t}h$ events is done analogously to the previous section. Following this selection, a 16% (17%) $t\bar{t}H$ ($t\bar{t}A$) event acceptance rate is achieved. Events are then submitted through a full $t\bar{t}h\,(h \to b\bar{b})$ kinematic reconstruction, based on the reconstruction algorithm previously described and including some additional improvements that are explained in Appendix D. In particular, by no longer calculating a solution per each possible jet combination, but instead doing the jet-lepton pairing to the top quark/antiquark and $b\bar{b}$ jet pair assignment to the Higgs boson with a mutivariate method, the computation runtime reduced significantly and better signal acceptance is achieved. In this section, whenever the full kinematic reconstruction is said to be used, this is the method applied.

Angular Distributions in $t\bar{t}H$, $t\bar{t}A$ and $t\bar{t}b\bar{b}$

Similarly to what was done in Section 8.2.2, angular distributions, such as those introduced in Expression 8.11, are explored for their promising discrimination power — not only between the SM $t\bar{t}H$ and its irreducible $t\bar{t}b\bar{b}$ background, but also considering the pure CP odd process

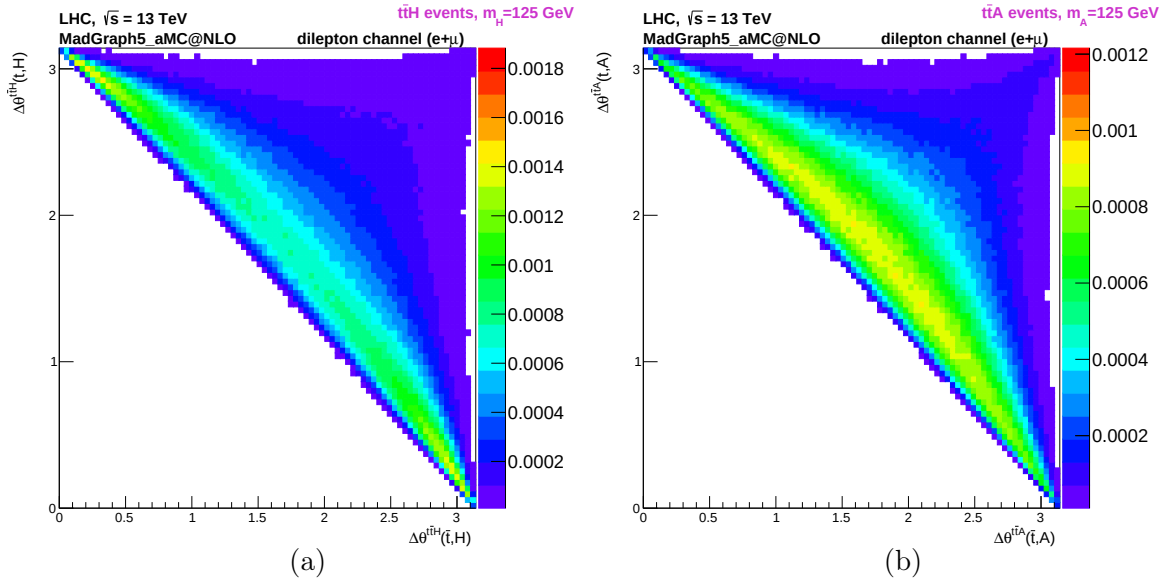


Figure 8.6: Two-dimensional distributions of $\theta_t^{t\bar{t}h}$ versus $\theta_{\bar{t}}^{t\bar{t}h}$ at parton level with NLO+shower effects accuracy: (a) is the distribution for the SM Higgs boson, $t\bar{t}H$, and (b) for a pure pseudoscalar Higgs boson, $t\bar{t}A$ [4]. $\theta_t^{t\bar{t}h}$ is the angle between the t quark and Higgs boson in the $t\bar{t}h$ centre-of-mass system, while $\theta_{\bar{t}}^{t\bar{t}h}$ is the angle between the $\bar{t}$ quark and Higgs boson also in the $t\bar{t}h$ centre-of-mass system.

$t\bar{t}A$. Angular distributions based on θ_Y^X, the angle between the direction of the Y system in the rest frame of X and the direction of the X system in the rest frame of its parent system, are then investigated. A survey of all the possible angles is attempted by following the $t\bar{t}h$ system dileptonic decay chain, here labeled as (123), and going through a parametrisation of successive two-body systems: $(123) \to 1 + (23)$, $(23) \to 2 + (3)$ and $(3) \to 4 + 5$. Particles 1, 2 and 3 are either the t quark, the $\bar{t}$ quark, or the h boson, Particles 4 and 5, on the other hand, can be any of the decay products of t, $\bar{t}$ or h, including the intermediate W bosons: W^+, W^-, ℓ^+,ℓ^-, b_t, $\bar{b}_{\bar{t}}$, b_h, $\bar{b}_h$.

In order to explore the nature of top quark Yukawa coupling, it is important to investigate the kinematic differences between purely scalar $t\bar{t}H$ events and the pseudoscalar $t\bar{t}A$ ones. The discriminating power of the different possible angles to differentiate between the two $t\bar{t}h$ processes can be easily gauged in Figures 8.6, 8.7 and 8.8. First, in Figure 8.6, the distinct behaviour of $t\bar{t}H$ events (a) and $t\bar{t}A$ ones (b) in two-dimensional distributions of the angle between the t quark and Higgs boson versus the angle between the $\bar{t}$ quark and Higgs boson, in the $t\bar{t}h$ centre-of-mass system, are shown with NLO+Shower effects accuracy at parton

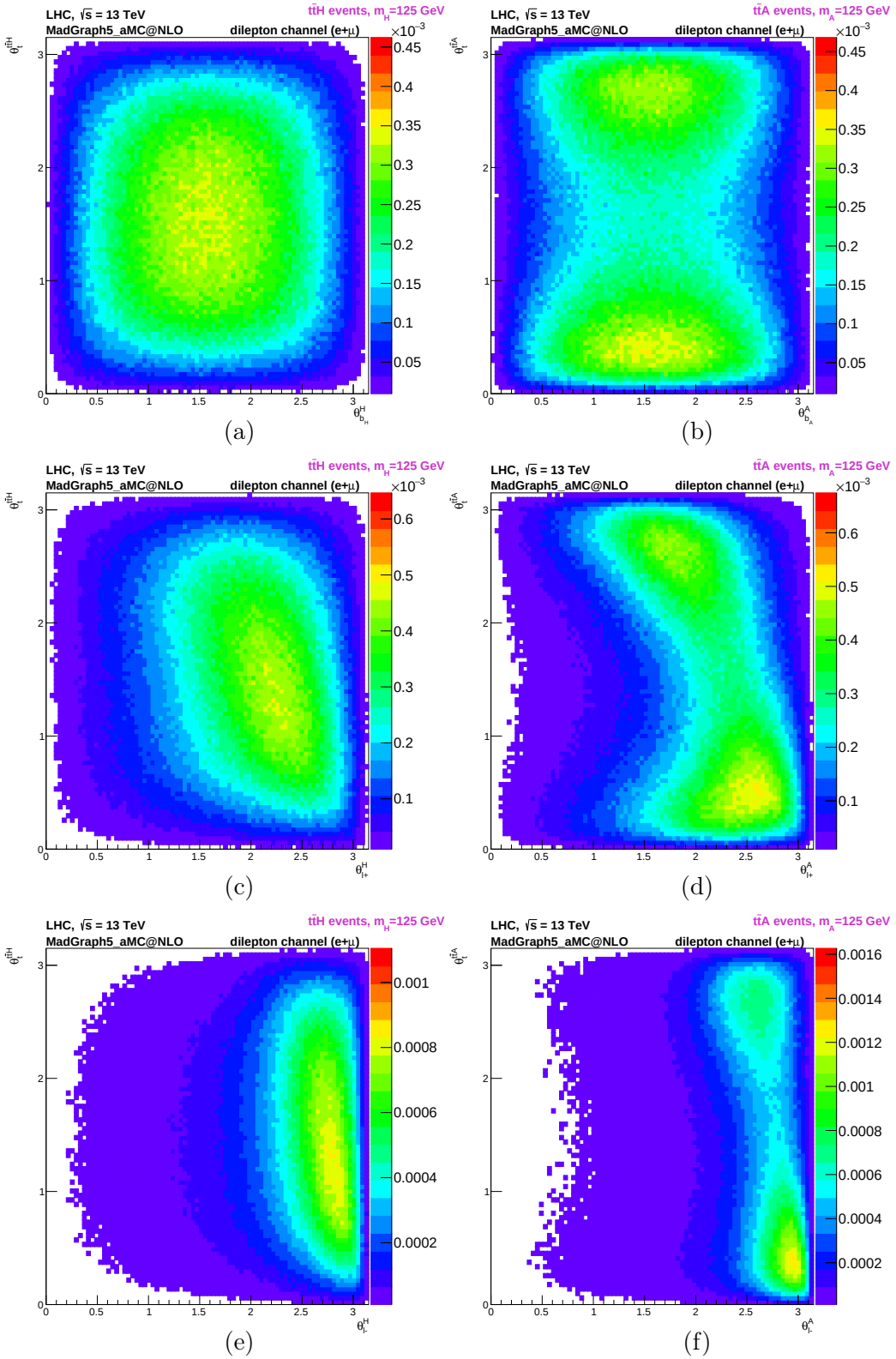


Figure 8.7: Two dimensional distribution at NLO+Shower of the angle between the top quark, in the $t\bar{t}h$ centre-of-mass frame, and the $t\bar{t}h$ direction in the lab frame, $\theta_t^{t\bar{t}h}$, versus the angle between the Higgs direction, in the $\bar{t}h$ rest frame, and the direction of several decay products (all boosted to the Higgs centre-of-mass): (a) and (d) b quark from h, (b) and (e) ℓ^+ from top quark and (c) and (f) ℓ^- from $\bar{t}$ [4]. The left (right) row distributions correspond to $t\bar{t}H$ ($t\bar{t}A$).

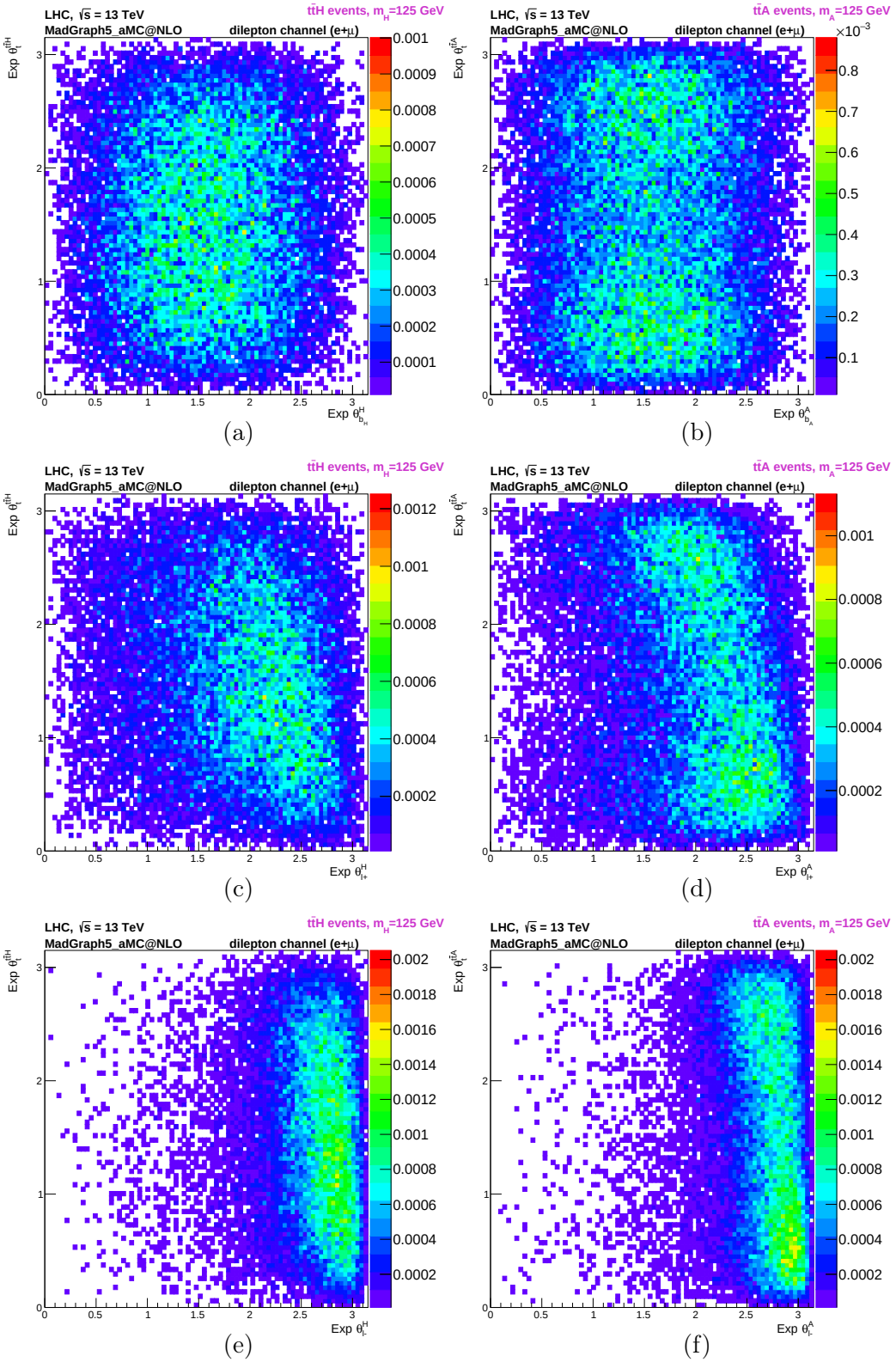


Figure 8.8: In the same disposition of Figure 8.7, but now after all selection cuts and full kinematic reconstruction, two dimensional distribution of $\theta_t^{t\bar{t}h}$ versus θ_Y^h for (a) and (d) $Y = b_h$, (b) and (e) $Y = \ell+_t$, (c) and (f) $Y = \ell-_{\bar{t}}$, where h represents the SM Higgs boson H in left distributions, and the pseudoscalar A in the right ones [4].

level. It is clear that there are kinematic differences between the pure CP even process of the SM and the CP odd process that could involve BSM physics. Second, the sensitivity of different θ_Y^X angles is assessed in Figure 8.7 at parton level and Figure 8.8 after event selection and full kinematic reconstruction. Again, two-dimensional distributions are shown here: the angle between the t quark direction, in the $t\bar{t}h$ centre-of-mass frame, and the $t\bar{t}h$ momentum in the lab frame, $\theta_t^{t\bar{t}h}$, is plotted against:

$\theta_{b_h}^h$ — the angle between the direction of the b quark from h boson decay, in the h boson centre-of-mass, and the h boson direction, in the $\bar{t}h$ rest frame — (a) for $t\bar{t}H$ sample and (d) for $t\bar{t}A$;

$\theta_{\ell+}^h$ — the angle between the ℓ^+ originating from t quark decay, in the h boson centre-of-mass, and the h boson in the $\bar{t}h$ rest frame — (b) for $t\bar{t}H$ sample and (e) for $t\bar{t}A$;

$\theta_{\ell-}^h$ — the angle between the ℓ^- originating from $\bar{t}$ quark decay, in the h boson centre-of-mass, and the h boson in the $\bar{t}h$ rest frame — (c) for $t\bar{t}H$ sample and (f) for $t\bar{t}A$.

The differences between the scalar and pseudoscalar signals are quite visible at parton level. Due to the selection cuts that events underwent, as well as the kinematic fit necessary to reach the hard scattered particles, the distributions in Figure 8.8 are distorted, with a significant reduction on the total number of events when compared with the distributions at parton level (Figure 8.7). Although there is a smaller density of points, the selection and reconstruction of events did not change the observed patterns: the $t\bar{t}H$ and the $t\bar{t}A$ signals have distinct angular behaviours.

In this study, we intend to explore these different kinematic properties to discriminate the two signals and gauge the CP nature of the top quark Yukawa coupling to the Higgs boson. Double angular products, like those introduced in Expression 8.11 and whose nomenclature x_y is kept for consistency, are considered, constituting three major families of observables:

$$\begin{aligned} &f(\theta_1^{123})g(\theta_4^3) \\ &f(\theta_1^{123})g(\theta_3^{23}) \\ &f(\theta_3^{23})g(\theta_4^3)\,, \end{aligned} \tag{8.13}$$

with $f = \{\sin, \cos\}$, $g = \{\sin, \cos\}$ and the momentum direction of the (123) system measured in the laboratory frame. Here, to calculate θ_Y^X it is necessary to boost particle Y to the centre-of-mass of particle X, which can be performed in two different ways:

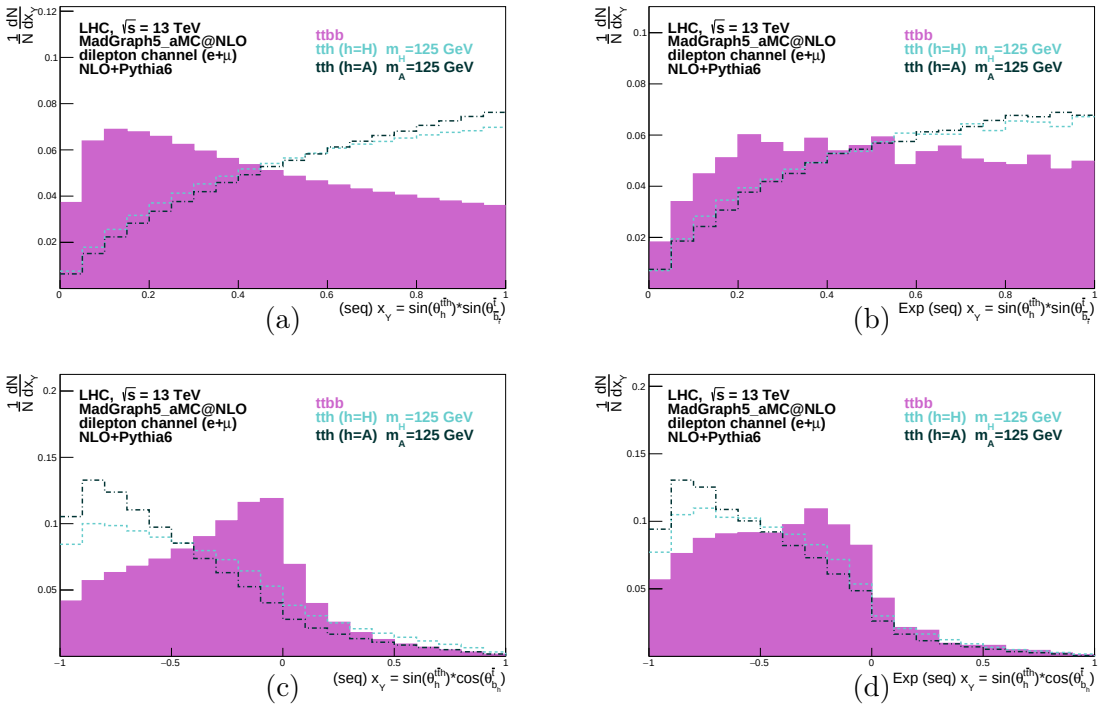


Figure 8.9: Distributions of $x_Y = \sin(\theta_h^{t\bar{t}h})\sin(\theta_{b_{\bar{t}}}^{\bar{t}})$ (top) and $x_Y = \sin(\theta_h^{t\bar{t}h})\cos(\theta_{b_h}^{\bar{t}})$ (bottom). The distributions at parton level ((a) and (c)) and after selection cuts, full kinematic reconstruction and extra b-tag multiplicity and $m_{\ell\ell}$ requirements ((b) and (d)), are shown. The dashed line represents the $t\bar{t}h$ SM model signal (where h represents H, the SM Higgs boson with $CP = +1$) and the dashed-dotted line corresponds to the pure pseudoscalar distribution $t\bar{t}h$ (where h represents A, the pseudoscalar boson with $CP = -1$). Here, the laboratory four-momentum of b quarks is boosted sequentially to the Higgs centre of mass system [4]

Direct Boost — using the laboratory four-momentum of both particles X and Y;

Sequential Boost — boosting Y and X sequentially through all intermediate centre-of-mass systems, until particle Y is evaluated in the centre-of-mass of X.

Due to Wigner rotations, different directions are obtained for the momentum of particle Y by using either of these boosting procedures. In this study, the use of either the sequential or direct prescriptions is tested.

By extensively surveying the angular observables in Expression 8.13, it is possible to identify distributions where the $t\bar{t}H$ and the $t\bar{t}A$ signals have a very similar behaviour, while other distributions show an impressive degree of separation between the two signals. The

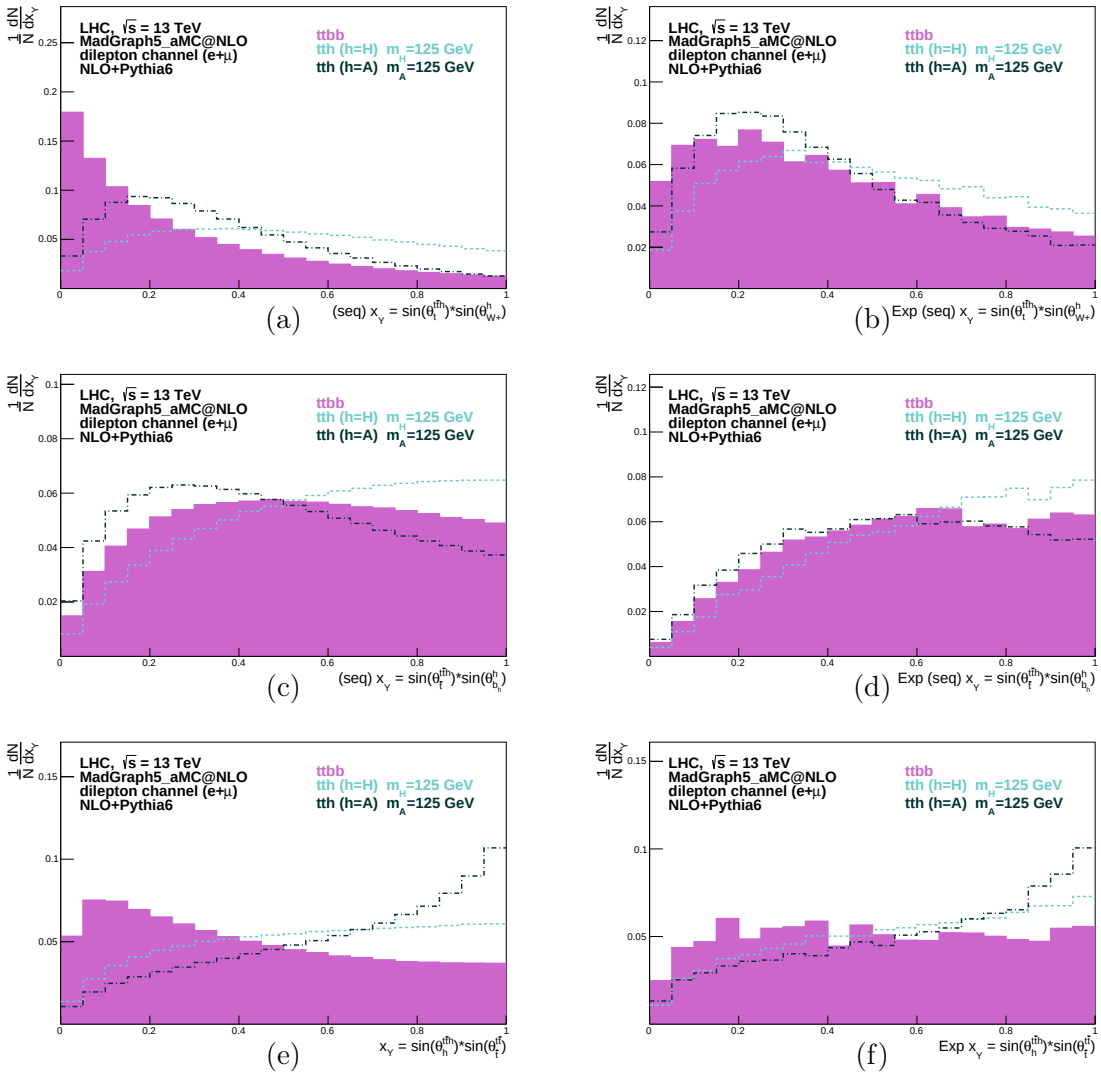


Figure 8.10: Angular distributions at parton level before selection cuts (left) and after event selection, full kinematic reconstruction and extra b-tag multiplicity and $m_{\ell\ell}$ requirements (right) of: (top) x_Y=sin $(\theta_t^{t\bar{t}H})$ sin (θ_{W+}^H), (middle) x_Y=sin $(\theta_{\bar{t}}^{t\bar{t}H})$ sin $(\theta_{b_H}^H)$ and (bottom) x_Y=sin $(\theta_H^{t\bar{t}H})$ sin $(\theta_{\bar{t}}^{t\bar{t}})$. The $t\bar{t}H$ SM model signal with $CP = +1$ (dashed line), the pure pseudoscalar $t\bar{t}A$ signal with $CP = -1$ (dashed-dotted line) and the dominant background $t\bar{t}b\bar{b}$ (shadowed region) are shown [4].

former observables can be used to search for, or set limits to, the total $t\bar{t}h$ production cross-section, since they are insensitive to the CP nature of the top Yukawa coupling to the Higgs boson. Some of these observables can even be used to further discriminate the $t\bar{t}h$ signal

from the irreducible $t\bar{t}b\bar{b}$ background events. The latter ones, however, present incompatible distributions between $t\bar{t}H$ and $t\bar{t}A$ samples, which are ideal for experimentally measuring, or setting limits to, a pseudoscalar component of the top quark-Higgs boson coupling.

In Figure 8.9, angular observables with higher discrimination between the $t\bar{t}b\bar{b}$ background (colour shaded) and the signal samples, $t\bar{t}H$ (dashed line) and $t\bar{t}A$ (dashed-dotted line), are shown. The distributions of $\sin(\theta_h^{t\bar{t}h})\sin(\theta_{\bar{b}_{\bar{t}}}^{\bar{t}})$ (top) and $\sin(\theta_h^{t\bar{t}h})\cos(\theta_{b_h}^{\bar{t}})$ (bottom) are represented without any selection cuts at parton level with NLO+Shower accuracy (Figures (a) and (c)) and after the event selection and full kinematic reconstruction (Figures (b) and (d)). The last-mentioned distributions include additional requirements of at least 3 b-tagged jets and $|m_{\ell\ell} - m_Z| > 10$ GeV, which are needed in an experimental setting to reduce background events and increase the signal significance.

In contrast, distributions of angular observables for which the $t\bar{t}H$ and $t\bar{t}A$ samples are the least compatible are shown in Figure 8.10: (top) x_Y=$\sin(\theta_t^{t\bar{t}H})\sin(\theta_{W^+}^{H})$, (middle) x_Y=$\sin(\theta_{\bar{t}}^{t\bar{t}H})\sin(\theta_{b_H}^{H})$ and (bottom) x_Y=$\sin(\theta_H^{t\bar{t}H})\sin(\theta_{\bar{t}}^{t\bar{t}})$. Even if the discrimination between the signal samples is slightly degraded, the kinematic properties seen at parton level (left) are still visible after the dileptonic event selection and the full reconstruction (right). For completeness the $t\bar{t}b\bar{b}$ process is also included. Again, the supplementary cuts on b-tag multiplicity and $m_{\ell\ell}$ are applied in the distributions on the right.

Previous studies, intent on distinguishing between the two CP components of the $t\bar{t}h$ signal, had already proposed several observables [294, 299, 303, 304, 305, 306], which are investigated in this study, with fully reconstructed $t\bar{t}h$ signal and $t\bar{t}b\bar{b}$ background events. The discriminating performance of most of these observables is identical, particularly after the full dileptonic $t\bar{t}h$ reconstruction. In Figure 8.11, the two most prominent variables, out of those proposed in literature, are shown:

$$\beta_{b\bar{b}}\Delta\theta^{\ell h}(\ell^+,\ell^-)\,, \tag{8.14}$$

proposed in [299], where $\theta^{\ell h}(\ell^+,\ell^-)$ is the angle between the ℓ^+ and ℓ^- directions projected onto the plane perpendicular to the h direction in the lab frame, β is the sign of $(\vec{p_b} - \vec{p_{\bar{b}}}) \cdot (\vec{p_{\ell^-}} \times \vec{p_{\ell^+}})$ and b and $\bar{b}$ are the resulting quarks of t and $\bar{t}$ decays, respectively; and

$$b_4 \;=\; \frac{p_t^z \cdot p_{\bar{t}}^z}{|\vec{p}_t| \cdot |\vec{p}_{\bar{t}}|}\,, \tag{8.15}$$

proposed in[303], which requires the full t and $\bar{t}$ four-momenta reconstruction. The $t\bar{t}H$, $t\bar{t}A$ and $t\bar{t}b\bar{b}$ samples are represented at parton level without any selection cuts (Figures 8.11

(a) and (b)), at parton level after event selection (Figures 8.11 (c) and (d)) and after the full reconstruction kinematic fit including the additional b-tag multiplicity and $m_{\ell\ell}$ requirement (Figures 8.11 (e) and (f)). The distributions in the top and middle row, obtained before and after the event selection respectively, the events are shown at parton level. The distributions on the bottom row are obtained using the reconstructed four momenta of all objects, without any truth-matched information, just as it would be done on collision data. As expected, a deterioration of the discriminating power of these observables is seen from parton level (at NLO with shower effects) to the final stage of the analysis (which includes event selection and kinematic reconstruction). The biggest distortion on the distributions comes from the event selection applied, as can be seen from the Figures 8.11 (c) and (d), and not from the kinematic reconstruction itself. Although the kinematic algorithm applied for the reconstruction cannot reconstruct every event, it does not change the behaviour of the distributions.

Similarly to what was done in Section 8.2.2, for each angular observable, x_y, a forward-backward asymmetry can be defined,

$$A_{FB}^{Y} = \frac{\sigma(x_Y > 0) - \sigma(x_Y < 0)}{\sigma(x_Y > 0) + \sigma(x_Y < 0)}, \tag{8.16}$$

where $\sigma(x_Y > 0)$ and $\sigma(x_Y < 0)$ correspond to the total cross-section with x_Y above and below zero, respectively. In Table 8.2, a selection of significantly discriminant asymmetries are presented. These are calculated both at parton level without any cuts and after the event selection and kinematic fit. This allows to understand whether or not the selection cuts and reconstruction method degrade the original parton level asymmetry or instead create artificial asymmetries in observables that did not have one to begin with. The most interesting variable choices are:

$$\cos(\theta_h^{\bar{t}h})\cos(\theta_{\ell^-}^{h}) \quad \text{for} \quad A_{FB}^{\ell-(h)}\,, \tag{8.17}$$

$$b_4 = (p_t^z.p_{\bar{t}}^z)/(|\vec{p}_t|.|\vec{p}_{\bar{t}}|) \quad \text{for} \quad A_{FB}^{b_4}\,, \tag{8.18}$$

$$\sin(\theta_h^{t\bar{t}h})\sin(\theta_{b_{\bar{t}}}^{\bar{t}}) \quad \text{for} \quad A_{FB}^{\bar{b}_{\bar{t}}(\bar{t})}(\text{seq. boost})\,, \tag{8.19}$$

$$\sin(\theta_h^{t\bar{t}h})\cos(\theta_{b_h}^{\bar{t}}) \quad \text{for} \quad A_{FB}^{b_h(\bar{t})}(\text{seq. boost})\,, \tag{8.20}$$

$$\sin(\theta_t^{t\bar{t}h})\sin(\theta_{W+}^{h}) \quad \text{for} \quad A_{FB}^{W+(h)}(\text{seq. boost})\,, \tag{8.21}$$

$$\sin(\theta_{\bar{t}}^{t\bar{t}h})\sin(\theta_{b_h}^{h}) \quad \text{for} \quad A_{FB}^{b_h(h)}(\text{seq. boost})\,, \tag{8.22}$$

$$\sin(\theta_h^{t\bar{t}h})\sin(\theta_{\bar{t}}^{t\bar{t}}) \quad \text{for} \quad A_{FB}^{\bar{t}(t\bar{t})}\,, \tag{8.23}$$

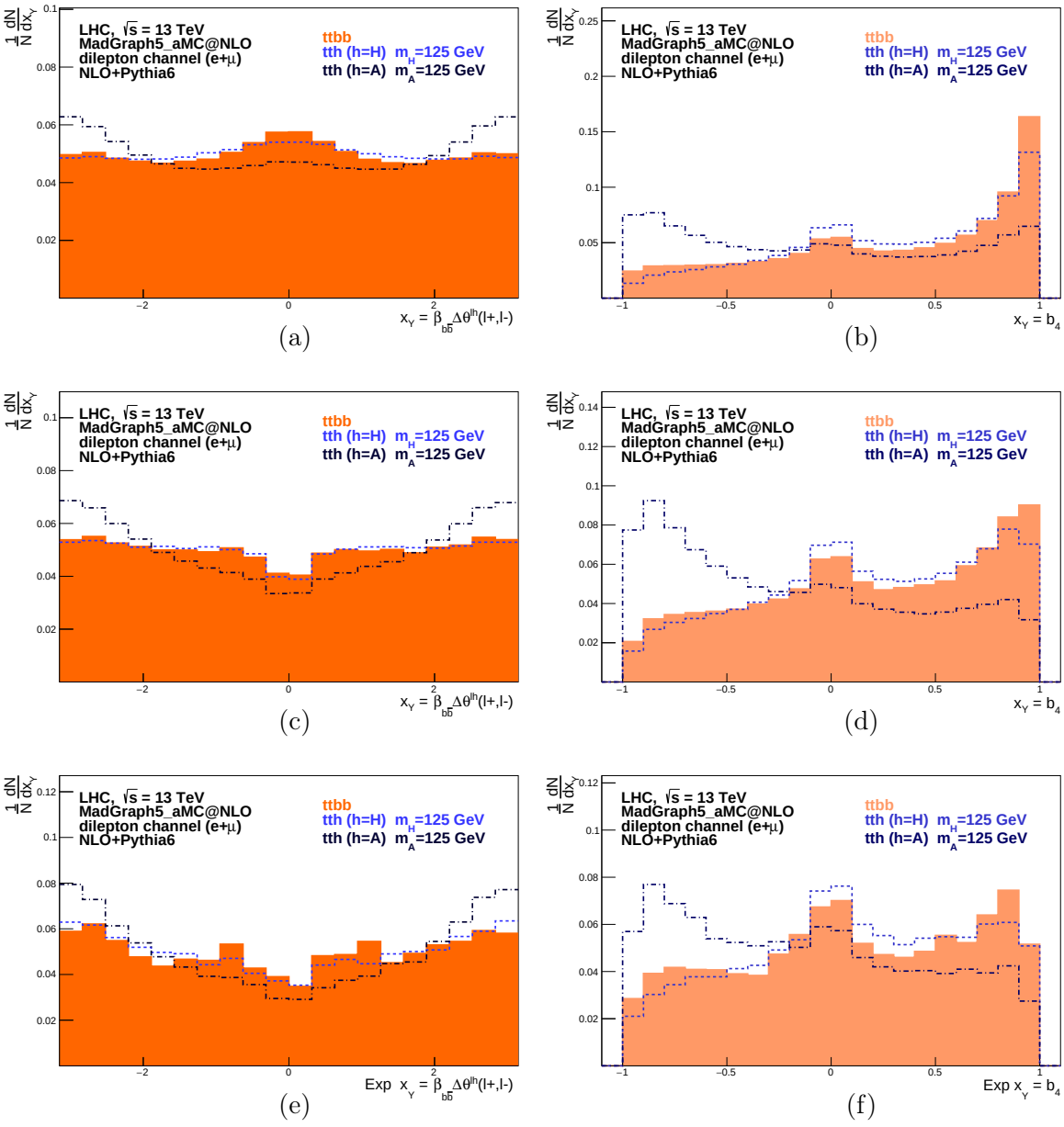


Figure 8.11: Distributions of $\beta\Delta\theta^{\ell h}(\ell+,\ell-)$ (left) and b_4 (right): (a) and (b) at parton level without cuts, (c) and (d) at parton level with event selection, and (e) and (f) after the full kinematic reconstruction, respectively. The $t\bar{t}H$ SM model signal with $CP = +1$ (dashed line), the pure pseudoscalar $t\bar{t}A$ signal with $CP = -1$ (dashed-dotted line) and the dominant background $t\bar{t}b\bar{b}$ (shadowed region) are shown [4].

(Asymmetries)	Parton level		Reconstruction	
	$t\bar{t}H/t\bar{t}A$	$t\bar{t}b\bar{b}$	$t\bar{t}H/t\bar{t}A$	$t\bar{t}b\bar{b}$
$A_{FB}^{\ell-(h)}$	+0.37/+0.41	+0.17	+0.42/+0.39	+0.24
$A_{FB}^{b_4}$	+0.35/−0.10	+0.33	+0.16/−0.17	+0.12
$A_{FB}^{\bar{b}_{\bar{t}}(\bar{t})}$(*seq.* **boost**)	+0.28/+0.33	−0.17	+0.25/+0.28	+0.03
$A_{FB}^{b_h(\bar{t})}$(*seq.* **boost**)	−0.65/−0.77	−0.62	−0.78/−0.83	−0.76
$A_{FB}^{W+(h)}$(*seq.* **boost**)	−0.03/−0.46	−0.60	+0.17/−0.06	−0.04
$A_{FB}^{b_h(h)}$(*seq.* **boost**)	+0.25/−0.08	+0.07	+0.37/+0.16	+0.23
$A_{FB}^{\bar{t}(t\bar{t})}$	+0.16/+0.37	−0.21	+0.23/+0.31	+0.01

Table 8.2: Asymmetry values for $t\bar{t}H$, $t\bar{t}A$ and $t\bar{t}b\bar{b}$ at NLO+Shower (without any cuts) and after applying the selection criteria and kinematic reconstruction. These asymmetries have a statistical uncertainty better than 1%.

which were already shown in Figures 8.9–8.11. In general, the discriminating performance of the newly proposed angular observables is comparable to that of the ones proposed in the literature. The reconstruction of the t, $\bar{t}$ quarks and h boson is almost always necessary in any of these observables, yet by using fully reconstructed $t\bar{t}h$ events, the spin information of signal and background processes can be reached and the nature of the quark-Higgs boson Yukawa coupling, and in particular of the $t\bar{t}h$ production, can be investigated.

Expected Limits

In order to estimate the experimental sensitivity of a dileptonic $t\bar{t}h$ analysis, all relevant SM backgrounds are considered. After the dileptonic $t\bar{t}H$ event selection and full kinematic reconstruction, further selection criteria is applied, as a means to deplete other backgrounds. The Z+jets background is significantly reduced by vetoing events with a dilepton invariant mass $|m_{\ell^+\ell^-} - m_Z| > 10$ GeV. The presence of background events involving t quark production, notably $t\bar{t}$+jets events, is also decreased by selecting events with at least 3 b-tagged jets. These additional event cuts, where already included in the previous Figures 8.9,8.10 and 8.11 and asymmetry results. It should be noted that both the additional event cuts and the improvements made to the dileptonic reconstruction are necessary to reduce the event rates of the SM backgrounds included, as well as to increase signal significance.

The expected effective cross-sections are shown in Table 8.3, at different levels of the event selection, for dileptonic (ee, $\mu\mu$ and $e\mu$) signal and SM backgrounds at a centre-of-mass energy of 13 TeV at the LHC. For comparison, numbers and distributions are also

	$N_{jets} \geq 4$ $N_{lep} \geq 2$	Kinematic Fit	m_Z cut	$N_b \geq 3$	$N_b \geq 4$
$t\bar{t}$+$c\bar{c}$, $t\bar{t}$+lf	2160	1300	1110	4.78	0.06
$t\bar{t}$+$b\bar{b}$	87.1	51.9	44.5	2.91	0.27
$t\bar{t}$+V(V=Z, W)	7.9	4.5	3.9	0.09	0.01
Single t	54	26	23	0.12	0.00
V+jets (V=W, Z)	2700	1200	200	0.00	0.00
V+$b\bar{b}$(V=W, Z)	570	280	20	0.00	0.00
Diboson	130	53	14	0.00	0.00
Total background	5700	2900	1410	7.90	0.34
$t\bar{t}H$	4.04	2.49	2.15	0.26	0.033
$t\bar{t}A$	4.43	2.69	2.36	0.31	0.041

Table 8.3: Expected cross-sections (expressed in fb) as a function of selection cuts, at 13 TeV, for dileptonic signal and background events at the LHC.

shown for events with at least four b-jets. Here, the $t\bar{t}A$ pseudoscalar signal was scaled to the $t\bar{t}H$ scalar cross-section for illustration purposes. In Figure 8.12, the angular distribution x_Y=$\sin(\theta_H^{t\bar{t}H})\cos(\theta_{b_H}^{\bar{t}})$ (top) and the invariant mass of the Higgs boson $m_{b\bar{b}}$ (bottom) are shown, for the expected number of signal and background events at a luminosity of 100 fb^{-1} at the LHC, on the left plots for a selection of at least three b-jets and on the right plots for at least four b-jets. Here, fake data points were included as a visual guide of the total number of expected events and related statistical uncertainties. The fake data are randomly calculated with a pseudo-experiment from the expected SM $t\bar{t}H$ signal and background distributions.

Similarly to what was done in the ATLAS 8 TeV analysis shown in Chapter 7, in this study, various multivariate methods are tested to provide better signal significance. The multivariate methods are fed information on the kinematic properties of the events through the input of several distributions, including the new angular distributions discussed previously. A boosted decision tree with gradient boost (BDTG) has returned the best performance and its output is used to test the sensitivity of the analysis to discriminate the scalar versus pseudoscalar component of the top quark-Higgs boson couplings, as a function of $\cos(\alpha)$. From the exhaustive set of observables tested, the multivariate method ranked the top 15 ones to discriminate between the $t\bar{t}H$ and $t\bar{t}A$ signals:

— the b_4 variable from [303];

— the invariant mass of the reconstructed Higgs boson ($m_{b\bar{b}}$);

— angular distributions with direct boost, $\cos(\theta_h^{\bar{t}h})\cos(\theta_{\ell^-}^h)$, $\sin(\theta_h^{t\bar{t}h})\sin(\theta_{\bar{t}}^{t\bar{t}})$;

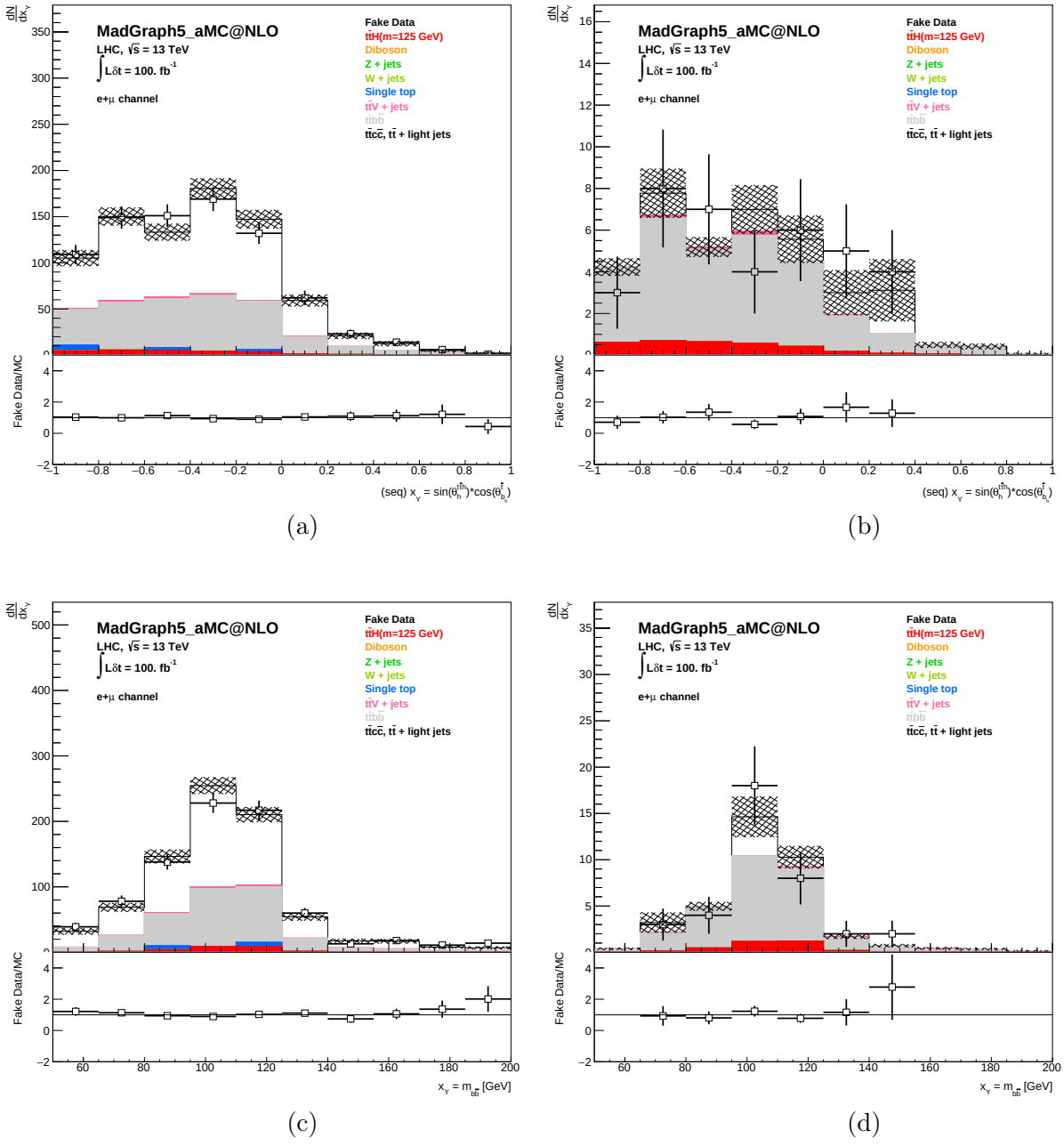
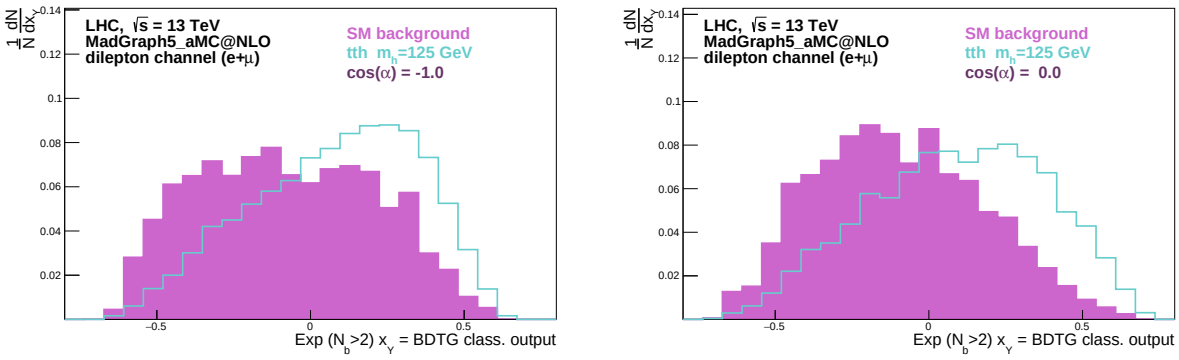


Figure 8.12: Distributions of $x_Y = \sin(\theta_H^{t\bar{t}H}) \cos(\theta_{b_H}^{\bar{t}})$ (top) and $m_{b\bar{b}}$ (bottom) after the final selection at 13 TeV for 100 fb^{-1}. The distributions on the left (right) correspond to events with at least three (four) jets from the hadronisation of b-quarks [4].



(a) (b)

Figure 8.13: Normalised distributions of the BDTG output discriminant variable for the pure scalar (a) and pure pseudoscalar (b) Higgs bosons. The line represents the $t\bar{t}h$ signal, while the shaded region corresponds to the total SM background [4].

— variables with sequential boost, $\sin(\theta^{t\bar{t}h}_{\bar{t}})\sin(\theta^{h}_{b_h})$, $\sin(\theta^{t\bar{t}h}_{h})\cos(\theta^{\bar{t}}_{b_h})$, $\sin(\theta^{t\bar{t}h}_{h})\sin(\theta^{\bar{t}}_{b_{\bar{t}}})$, $\sin(\theta^{t\bar{t}h}_{t})\sin(\theta^{h}_{W^+})$;

— the $\Delta\eta$ distribution between the jets with maximum $\Delta\eta$ ($\Delta\eta_{jj}^{\max\Delta\eta}$)

— the invariant mass of the two b-tagged jets with lowest ΔR ($m_{bb}^{\min\Delta R}$);

— the ΔR between the Higgs candidate and the closest ($\Delta R_{hl}^{\min\Delta R}$) and farthest ($\Delta R_{hl}^{\max\Delta R}$) leptons;

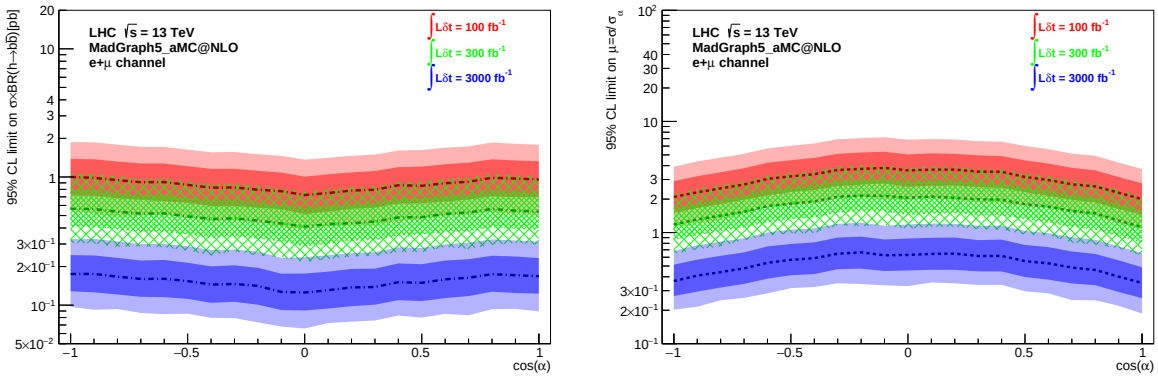
— the ΔR between the b-tagged jets with highest p_T ($\Delta R_{bb}^{\max p_T}$)

— the invariant mass of the two jets with closest value to the Higgs mass ($m_{jj}^{\text{closest to 125 GeV}}$);

— the jets aplanarity.

The BDTG output classifier is shown in Figure 8.13 for the pure scalar (a) and pseudoscalar (b) Higgs bosons. For each value of $\cos(\alpha)$, the BDTG used for the limit extraction is trained on a signal sample generated with that exact same $\cos(\alpha)$ value. This accounts for the different shape of the total background (shaded region) seen in these distributions.

Finally, using the BDTG output distribution, expected limits at 95% CL for $\sigma \times BR(h \to b\bar{b})$ and for signal strength μ, can be extracted in the background-only scenario. The possibility of a Higgs boson with a mixed-CP state is investigated using the samples of $t\bar{t}h$ with $\cos(\alpha)$ ranging from -1 to 1, discussed previously. The expected limits for integrated luminosity values of 100, 300 and 3000 fb^{-1} are presented in Figure 8.14. Using the dileptonic decay channel alone, sensitivity to the SM $t\bar{t}H$ production (μ=1) should be achieved shortly



(a) (b)

Figure 8.14: Expected limits at 95% CL in the background-only scenario, as a function of $\cos(\alpha)$. Limits on $\sigma \times BR(h \to b\bar{b})$ (a) and μ (b) obtained with the BDTG output discriminant for integrated luminosity values of 100, 300 and 3000 fb^{-1}. The lines correspond to the median, while the narrower (wider) bands correspond to the $1\sigma(2\sigma)$ intervals [4].

after the 300 fb^{-1} milestone. Individual limits on $\sigma \times BR(h \to b\bar{b})$ at 300 fb^{-1}, obtained from fits to $sin(\theta_h^{t\bar{t}h})sin(\theta_{\bar{t}}^{t\bar{t}})$ (a), $\beta_{b\bar{b}}\Delta\theta^{\ell h}(\ell^+,\ell^-)$ (b), $m_{b\bar{b}}^{\min\Delta R}$ (c) and b_4 (d) are shown in Figure 8.15. The same overall dependence on $\cos(\alpha)$ is seen, yet each of these distributions has a different sensitivity when used as input to the BDTG. For $\cos(\alpha) \sim 0$, *i.e.* closer to the pure pseudoscalar region, the 95% CL limit on $\sigma \times BR(h \to b\bar{b})$ is the most stringent for all variables.

Figure 8.16 displays a comparison between the expected limits on $\sigma \times BR(h \to b\bar{b})$, at 300 fb^{-1}, determined from each individual distribution used as input in the BDTG multivariate discriminant. For completeness, the limits obtained from the BDTG itself are also included, as well as, those from the $\beta_{b\bar{b}}\Delta\theta^{\ell h}(\ell^+,\ell^-)$ distribution, which was not used as input in the BDTG. The limits computed from the angular observables, $\beta_{b\bar{b}}\Delta\theta^{\ell h}(\ell^+,\ell^-)$, b_4 and $m_{b\bar{b}}$ are shown on the left, whereas the limits obtained from the remaining individual observables, that were used as input for the BDTG method, are shown on the right. The bottom plots represent the distributions of the ratio of each individually obtained limit over the BDTG one. As can be seen, most individual angular variables result in limits 15 to 20% worse than the limit obtained with the BDTG discriminant, in particular for the pure pseudoscalar case ($\cos(\alpha) = 0$) where it is slight more imposing. From the remaining variables investigated, including kinematical distributions commonly discussed in the literature,

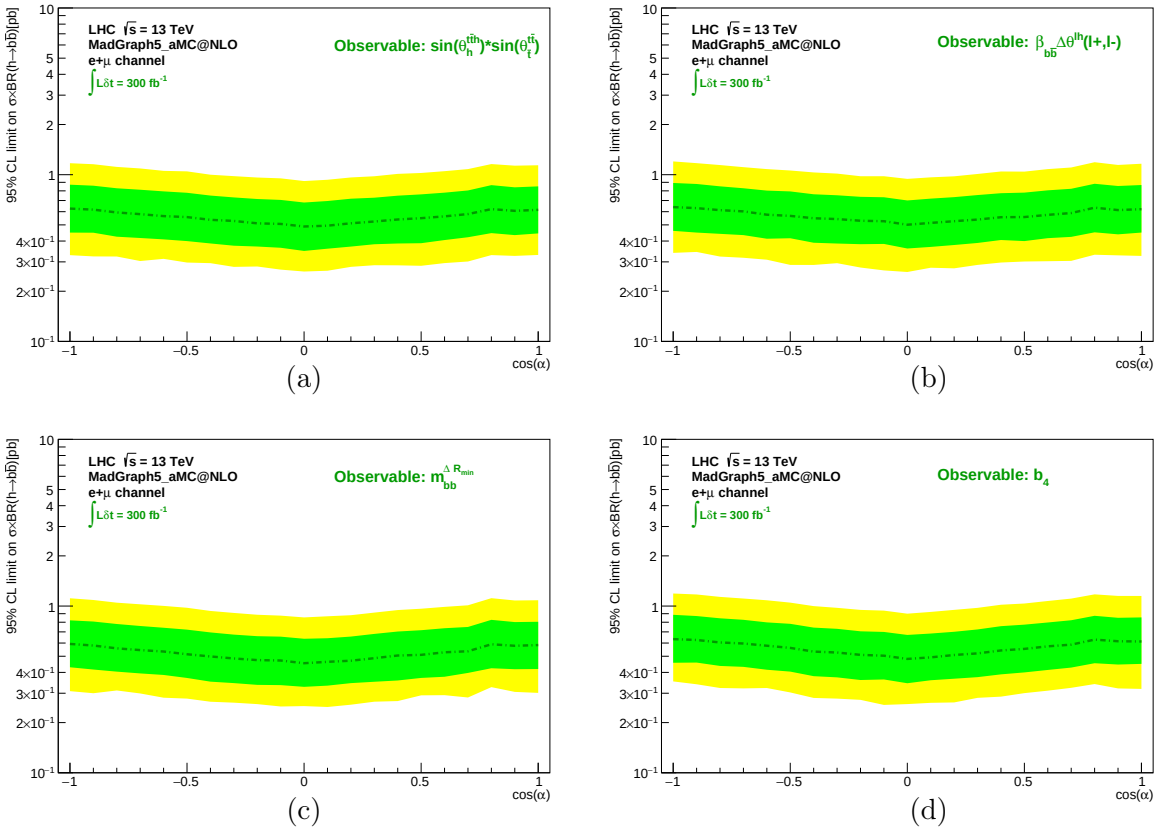


Figure 8.15: Expected 95% CL limits in the background-only scenario on $\sigma \times BR(h \to b\bar{b})$ at 300 fb^{-1} as a function of cos(α), using individual observables: $sin(\theta_h^{t\bar{t}h})sin(\theta_{\bar{t}}^{t\bar{t}})$ (a) and $\beta_{b\bar{b}}\Delta\theta^{\ell h}(\ell^+,\ell^-)$ (b); $m_{bb}^{\min\,\Delta R}$ (c) and b_4 (d). The lines correspond to the median, while the narrower (wider) bands correspond to the $1\sigma(2\sigma)$ intervals [4].

$m_{bb}^{\min\,\Delta R}$ seems to have a higher discriminating power, with expected limits slightly closer to those of the BDTG discriminant. By using a multivariate method to build a discriminant that includes all the information available, the computed expected limits are unsurprisingly the most stringent possible. Here, the BDTG method improves the individual limits up to 25%. The sensitivity to the nature of the top quark Yukawa coupling to Higgs boson, when evaluated individually, is overall similar between the newly proposed angular observables and the more common variables referenced in literature. The expected limits do not have a particularly strong dependence on the CP-phase (α). In this study, only the dileptonic final states of $t\bar{t}h$ process are considered. By combining these with other decay channels, using

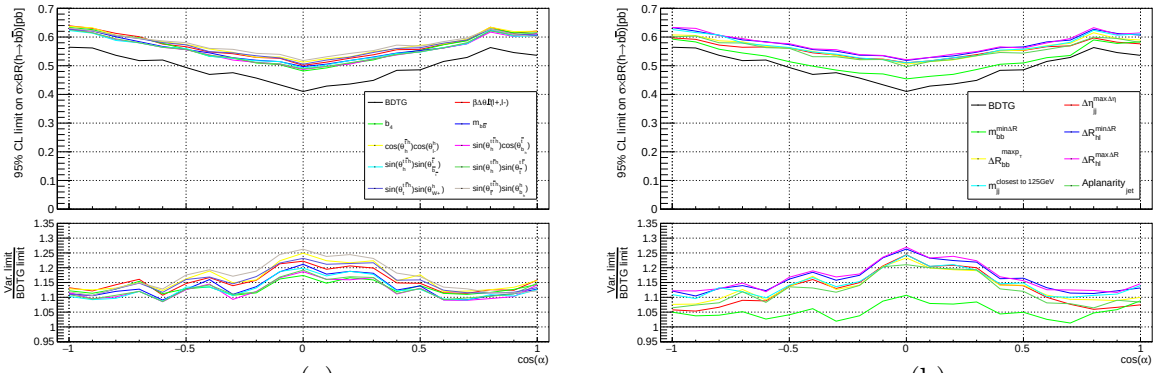


Figure 8.16: Comparison between limits on $\sigma \times BR(h \to b\bar{b})$, at 300 fb^{-1}, obtained from each one of the individual distributions used in the BDTG: $\beta_{b\bar{b}}\Delta\theta^{\ell h}(\ell^+,\ell^-)$, b_4, $m_{b\bar{b}}$ and angular distributions (left), and remaining distributions used as input for the BDTG (right). The ratios with respect to the limit obtained from the BDTG distribution are also represented [4].

fully reconstructed states and similar angular and kinematic observables, these results are expected to improve significantly.

8.3 Top Quark and W boson Coupling

In order to study the top quark electroweak coupling to the W boson, single top quark events are explored. In the presence of a BSM coupling, the top quark polarisation along some axis (defined as twice the expectation value of its spin operator in that axis, $P_i = 2\langle S_i \rangle$) may depart from its SM value. This can provide a direct test to the presence of an anomalous Wtb coupling [5]. The t-channel process, being the predominant electroweak production channel of top quarks, is used to probe the top quark polarisation.

The polarisations are studied in the framework of an effective electroweak theory [307] expanding the SM Lagrangian for a general Wtb vertex with anomalous effective left- and right-handed vector and tensor couplings, $V_{\mathrm{L,R}}$ and $g_{\mathrm{L,R}}$ respectively:

$$\begin{aligned}\mathcal{L}_{Wtb} &= -\frac{g}{\sqrt{2}}\bar{b}\,\gamma^{\mu}\left(V_{\mathrm{L}}P_{\mathrm{L}} + V_{\mathrm{R}}P_{\mathrm{R}}\right)t\,W^{-}_{\mu} \\ &\quad -\frac{g}{\sqrt{2}}\bar{b}\,\frac{i\sigma^{\mu\nu}q_{\nu}}{M_W}\left(g_{\mathrm{L}}P_{\mathrm{L}} + g_{\mathrm{R}}P_{\mathrm{R}}\right)t\,W^{-}_{\mu} + \mathrm{h.c.}\,, \end{aligned} \tag{8.24}$$

allowing the sensitivity to these non-standard contributions to be estimated [308]. Here, $P_{\mathrm{L,R}}$

are the left- and right-handed chirality projection operators, M_W is the W boson mass, and g the weak coupling constant. Alternatively, the effects of BSM physics at a high energy scale Λ could be described by an effective Lagrangian [309, 310, 311]:

$$-\mathcal{L}_{eff} = \mathcal{L}_{SM} + \sum_x \frac{C_x}{\Lambda^2}\mathcal{O}_x + \mathcal{O}(1/\Lambda^3) + ... \,, \tag{8.25}$$

where $\mathcal{O}_x$ are dimension-six gauge-invariant operators and C_x represent complex dimensionless constants, commonly known as Wilson coefficients. The relevant operators affecting the general effective Wtb vertex are then:

$$\begin{aligned} \mathcal{O}_{\phi\phi} &= i(\tilde{\phi}^\dagger D_\mu \phi)(\bar{t}_{\rm R}\gamma^\mu b_{\rm R}) \\ \mathcal{O}_{tW} &= (\bar{q}_{\rm L}\sigma^{\mu\nu}\tau^I t_{\rm R})\tilde{\phi}W^I_{\mu\nu} \\ \mathcal{O}_{bW} &= (\bar{q}_{\rm L}\sigma^{\mu\nu}\tau^I b_{\rm R})\phi W^I_{\mu\nu} \,, \end{aligned} \tag{8.26}$$

where ϕ represents the weak doublet of the Higgs field and d (u) is the down (up) type quarks, $t_{\rm R}$ and $d_{\rm R}$ are the weak singlets, $q_{\rm L} = (t, b)_{\rm L}$ is the $SU(2)_{\rm L}$ weak doublet of the third family of quarks and τ^I is the Pauli matrix. The anomalous couplings definition in terms of the dimension-six effective operators parametrising the new physics is:

$$\begin{aligned} V_{\rm L} &= V_{tb} + C_{\phi q}\frac{v^2}{\Lambda^2} \,, & V_{\rm R} &= \frac{1}{2}C^*_{\phi\phi}\frac{v^2}{\Lambda^2} \,, \\ g_{\rm L} &= \sqrt{2}C^*_{dW}\frac{v^2}{\Lambda^2} \,, & g_{\rm R} &= \sqrt{2}C_{uW}\frac{v^2}{\Lambda^2} \,. \end{aligned} \tag{8.27}$$

Within the SM, $V_{\rm L}$ matches the CKM element V_{tb}, whereas $V_{\rm R}$, $g_{\rm L}$ and $g_{\rm R}$ are absent at tree level. A direct study of the Wtb vertex properties can provide constraints on the former and consequently on the effective operator coefficients. In order to distinguish between left and right effective couplings, observables involving the measurement of the top quark spin are necessary, to which its polarisation is, by definition, conveniently suitable.

Several direct and indirect limits on the anomalous right-handed vector (V_R), as well as on the left- and right-handed tensor couplings ($g_{\rm L}$ and $g_{\rm R}$) have been set over the years. A combined 95% CL limit has been defined by the ATLAS and the CMS collaborations from the measurement of the W-boson helicity fractions on data collected at 8 TeV, as $\mathrm{Re}\, V_R \in [-0.11, 0.16]$, $\mathrm{Re}\, g_L \in [-0.08, 0.05]$ and $\mathrm{Re}\, g_R \in [-0.04, 0.02]$, or in terms of the Wilson coefficients $C^*_{\phi\phi} \in [-3.48, 5.16]$, $C^*_{bW} \in [-0.96, 0.67]$ and $C_{tW} \in [-0.48, 0.29]$, where the couplings were assumed real without any additional CP violation introduced, and only

one was allowed to vary while fixing all others to their SM values [312]. Additionally, two-dimensional 95% CL bounds on g_R/V_L were set by ATLAS using single top quark events at $\sqrt{s} = 7TeV$, yielding $\mathrm{Re}\, g_R/V_L \in [-0.36, 0.10]$ and $\mathrm{Im}\, g_R/V_L \in [-0.17, 0.23]$ with a correlation of 0.11 [313]. $\mathrm{Im}\, C_{tW}$ can then be inferred to be within $[-2.3, 3.0]$ at 95% CL. These measurements were included in a combination of electric dipole moment analyses, collider observables, precision electroweak tests and flavour physics in [314], where very tight individual limits on $\mathrm{Im}\, C_{tW}$ are obtained by allowing only $\mathrm{Im}\, C_{tW}$ to be a non-zero coefficient. However, when multiple coefficients are allowed to vary simultaneously these bounds become significantly less stringent. Direct limits on the imaginary part of g_R were also measured to be within $[-0.18, 0.06]$ at 95% CL by ATLAS [315], using A_{FB}^{N} and A_{FB}^{ℓ} asymmetries and while assuming $V_L = 1$ and all other anomalous couplings to be null ($V_R = g_L = 0$ and $\mathrm{Re}\, g_R = 0$). A global fit including real and imaginary parts of all Wtb operators [316], determined by combining Tevatron and LHC precise measurements of W boson helicity fractions, single top quark production cross-sections and forward-backward asymmetries, gives $\mathrm{Re}\, g_R \in [-0.07, 0.08]$ and $\mathrm{Im}\, g_R \in [-0.19, 0.13]$ at 95% CL (or $\mathrm{Re}\, C_{tW} \in [-0.8, 0.7]$ and $\mathrm{Im}\, C_{tW} \in [-2.3, 1.6]$), where the bound on imaginary part is dominated by the previous ATLAS result [315]. A later ATLAS result [317] improved the 95% CL interval of the real and imaginary parts of g_R/V_L to be within $[-0.12, 0.17]$ and $[-0.07, 0.06]$ respectively, while the magnitudes of the ratios $|V_R/V_L|$ and $|g_L/V_L|$ are constrained at 95% CL to be less than 0.37 and 0.29, respectfully. The bounds on $\mathrm{Im}\, C_{tW}$ are then $[-0.8, 0.7]$ at 95% CL.

Considering a reference system with axes (x, y, z) in the top quark rest frame, the state of an ensemble of polarised top quarks can then be described as:

$$\rho = \frac{1}{2} \begin{pmatrix} 1 + P_z & P_x - iP_y \\ P_x + iP_y & 1 - P_z \end{pmatrix} . \tag{8.28}$$

The three dimensional polarisation vector, $\vec{P}$, satisfies $|\vec{P}| \leq 1$, and can only be $|\vec{P}| = 1$ if the top quarks are produced in a pure spin state. For the t-channel of single top production (whose corresponding LO Feynman diagrams were shown in Figure 2.7), three orthogonal axes were defined relative to the direction of the momentum of the light flavoured quark produced with the top (anti)quark, also known as the spectator quark. Hereafter, we will treat the light quark q' as the spectator quark j. The three orthogonal axes are then chosen as:

longitudinal direction, $\hat{z} = \frac{\vec{p}_j}{|\vec{p}_j|}$ — the direction of the spectator quark momentum $\vec{p}_j$;

normal direction, $\hat{\boldsymbol{y}} = \frac{\vec{\boldsymbol{p}}_j \times \vec{\boldsymbol{p}}_q}{|\vec{\boldsymbol{p}}_j \times \vec{\boldsymbol{p}}_q|}$ — orthogonal to $\vec{p}_j$ and the initial quark momentum $\vec{p}_q$;

transverse direction, $\hat{\boldsymbol{x}} = \hat{\boldsymbol{y}} \times \hat{\boldsymbol{z}}$ — determined by requiring the coordinate system to be right-handed;

with $\vec{p}_j$ and $\vec{p}_q$ in the top quark rest frame [5].[1] The momentum direction of the initial quark in the t-channel process cannot be determined unambiguously in hadronic collisions. Therefore a criterion to select the direction of one of the two incoming proton beams will be implemented and explained later on.

The top quark polarisation can be determined from an analysis of angular distributions of its decay products in the top quark rest frame. In this study, a complete description of the top quark (antiquark) polarisation in the t-channel of single production is done by considering all three orthogonal axes and providing, for the first time, the polarisation expressions along these axes for a general Wtb effective vertex including anomalous couplings. Given P_x, P_y and P_z, the top quark (antiquark) spin density matrix can be determined and therefore allow the calculation of the polarisation in any other direction. Using these observables, the sensitivity to new physics in the Wtb vertex structure is estimated. The measurements are then used to set limits on anomalous contributions of the Wtb couplings.

The t-channel single top quark process in proton-proton collisions at 8 TeV was simulated with the LO generator Protos [84], using CTEQ6L1 [221] parton distribution functions. For the SM Lagrangian, the obtained top quark and antiquark polarisations are:

$$\begin{aligned} \vec{P} &\simeq (0, 0, 0.90) && (t)\,, \\ \vec{P} &\simeq (-0.14, 0, -0.86) && (\bar{t})\,. \end{aligned} \tag{8.29}$$

calculated for the $2 \to 3$ process $qg \to q't\bar{b}$. As can be seen, both top quarks and antiquarks are produced mostly polarised along the P_z direction. Indeed, due to the $V-A$ nature of the SM Wtb vertex, their spins are aligned along the direction of the down-type quarks. While the newly defined polarisations P_x and P_y vanish in the SM for the single top quarks, for top antiquarks, however, P_x is of the order of $\mathcal{O}(0.1)$ with a still vanishing P_y component. In this case, the transverse direction P_x is not null, since for the leading top antiquark production

[1] The nomenclature was chosen such that in the $2 \to 2$ approximation to the t-channel production, $qb \to q't$, the vectors $\vec{p}_j$ and $\vec{p}_q$ determine the production plane, therefore containing $\hat{x}$ in that plane, whereas $\hat{y}$ is orthogonal to it.

process $dg \to \bar{u}t\bar{b}$, the spectator quark is not a down-type quark but instead an up-type quark [318, 319]. The longitudinal polarisations obtained with the PROTOS generator are compatible with the NLO calculation [320], $P_z = 0.91$ (t) and $P_z = -0.86$ ($\bar{t}$). Although calculated at LO, the PROTOS generator samples show similar kinematical distributions to NLO generated ones [321, 322]. These SM predictions change substantially when in the presence of new physics in the Wtb vertex. As a result, their measurement constitute a strong BSM physics probe.

In Figure 8.17, the polarisation, P_i, is shown for all three axes as a function of each anomalous coupling variation. The SM coupling V_L is fixed to unity, as well as hereinafter. The anomalous couplings, V_R, g_L and g_R, are varied within a $[-0.2, 0.2]$ range, since previous direct and indirect results [323]–[328] excluded values outside that range. A fit can now be used to extract the behaviour of the top quark polarisation when in the presence of anomalous couplings. The cross-sections of the single top quark (t) and single top antiquark ($\bar{t}$) production processes can be calculated for anomalous Wtb couplings by resorting to correction factors, f_t and $f_{\bar{t}}$ respectively:

$$\begin{aligned} f_t &= 1 + 0.90\,|V_R|^2 + 1.47\,|g_L|^2 + 2.31\,|g_R|^2 - 0.11\,\mathrm{Re}\,V_L^* V_R - 0.53\,\mathrm{Re}\,V_L^* g_R \\ f_{\bar{t}} &= 1 + 1.09\,|V_R|^2 + 2.36\,|g_L|^2 + 1.58\,|g_R|^2 - 0.12\,\mathrm{Re}\,V_L^* V_R - 0.56\,\mathrm{Re}\,V_R^* g_L\,, \end{aligned} \quad (8.30)$$

where the combinations of couplings with numerical pre-factors smaller than 0.1 are considered negligible. The resulting top quark polarisations in the three orthogonal axes are then

$$\begin{aligned} P_x &= \left(-0.13|V_R|^2 + 0.25\,|g_L|^2 - 0.90\,|g_R|^2 + 2.14\,\mathrm{Re}\,V_L^* g_R - 1.53\,\mathrm{Re}\,V_R^* g_L\right)/f_t\,, \\ P_y &= \left(-2.12\,\mathrm{Im}\,V_L^* g_R - 1.54\,\mathrm{Im}\,V_R^* g_L\right)/f_t \\ P_z &= \left(0.90 - 0.76\,|V_R|^2 + 1.15\,|g_L|^2 - 1.50\,|g_R|^2 - 0.60\,\mathrm{Re}\,V_L^* g_R + 0.36\,\mathrm{Re}\,V_R^* g_L\right)/f_t\,, \end{aligned} \quad (8.31)$$

and for the top antiquarks

$$\begin{aligned} P_x &= \left(-0.14 - 0.96\,|g_L|^2 + 0.34\,|g_R|^2 - 1.71\,\mathrm{Re}\,V_L^* g_R + 2.31\,\mathrm{Re}\,V_R^* g_L\right)/f_{\bar{t}}\,, \\ P_y &= \left(1.72\,\mathrm{Im}\,V_L^* g_R + 2.30\,\mathrm{Im}\,V_R^* g_L\right)/f_{\bar{t}} \\ P_z &= \left(-0.86 + 0.99\,|V_R|^2 - 1.56\,|g_L|^2 + 1.20\,|g_R|^2 + 0.42\,\mathrm{Re}\,V_L^* g_R - 0.67\,\mathrm{Re}\,V_R^* g_L\right)/f_{\bar{t}}\,, \end{aligned} \quad (8.32)$$

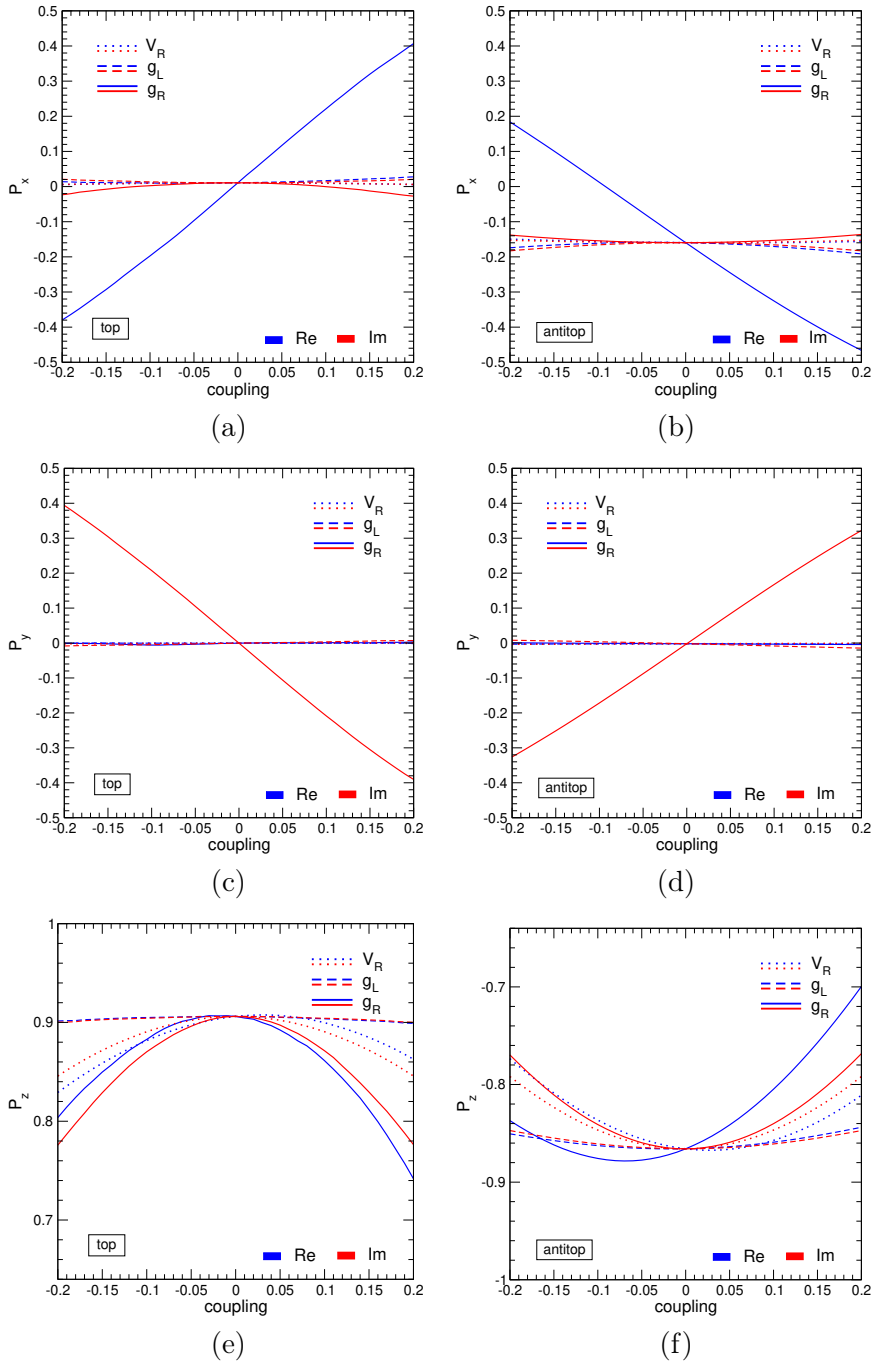


Figure 8.17: Polarisation of single top quark (antiquark) in the three previously defined axes, while in the presence of either purely real (blue) or imaginary (red) anomalous Wtb couplings [5]. In the left column, the polarisation of the top quark, whereas in the right column, the polarisation for the top antiquark is shown. The initial quark direction $\vec{q}$ is assumed known for both P_x and P_y.

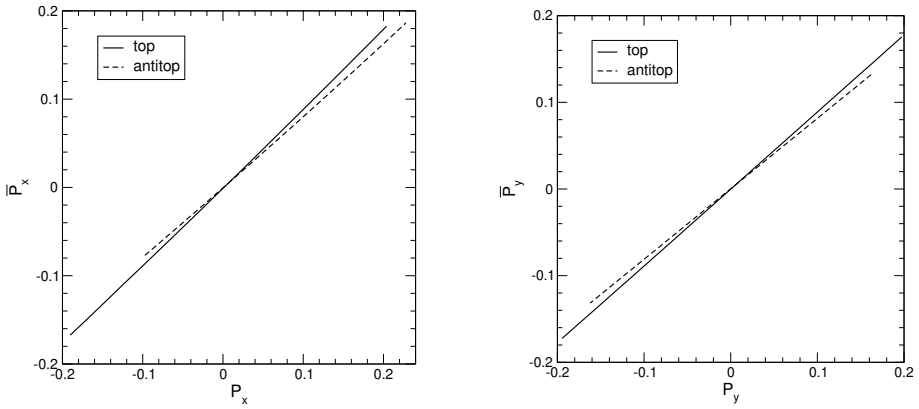


Figure 8.18: Relation between the polarisations $P_{x,y}$, in which the initial quark direction is known, and the polarisations $\bar{P}_{x,y}$, in which it is estimated from the spectator quark. [5].

where the dependence with the Wtb anomalous couplings can be seen. The polarisations along the normal direction, P_y, have a dependence on imaginary parts of coupling products, while the polarisation components along the transverse and longitudinal directions, P_x and the P_z respectively, do not. From the $2 \to 2$ approximation to the t-channel process, where the differential cross-section is proportional to the squared matrix element, with no absorptive parts, these imaginary terms can only arise from traces of Dirac matrices $\mathrm{tr}\,\gamma^5\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma = -4i\epsilon^{\mu\nu\rho\sigma}$ contracted with four different tetravectors.[1] The four tetravectors involved in the the $2 \to 2$ process are the three independent four-momenta, associated to three of the involved particles, plus the spin vector s_t. In the top quark rest frame, the Lorentz-invariant contraction $\epsilon_{\mu\nu\rho\sigma}p_q^\mu p_j^\nu p_t^\rho s_t^\sigma$ is proportional to the triple product $(\vec{p}_q \times \vec{p}_j)\cdot\vec{s}_t$. When $\vec{s}_t$ is in either $\hat{x}$ or $\hat{z}$ directions, the triple product vanishes, whereas it does not in the $\hat{y}$ direction. Considering instead the $2 \to 3$ process, since the extra b quark is mostly collinear to the beam direction, its effect should be negligible to this argument.

In hadronic collisions, the direction of the initial quark in the $qg \to q't\bar{b}$ single top quark production process cannot be unequivocally known. As the base of the construction of our (x, y, z) reference system, a selection of the direction of the initial quark q coincident with either one of the two incoming proton beam directions must be made. In the laboratory frame, the momentum of the spectator quark $j = q'$ mostly follows that of the initial quark: 97% of the times for $ug \to dt\bar{b}$ and 98% of the time for $dg \to u\bar{t}b$. Since these two processes

[1] $\epsilon^{\mu\nu\rho\sigma}$ is the totally antisymmetric tensor verifying $\epsilon^{0123} = 1$.

constitute the main channels of single top quark and antiquark production, in this study we choose the direction of the initial quark according to that of the spectator quark. For the remainder production channels, however, the rate of having correctly estimated the initial quark direction is lower. For this reason, the 'observed' polarisations $\bar{P}_{x,y}$ are slightly smaller than those that would be measured had the initial quark direction been known with 100% certainty (the 'true' polarisation $P_{x,y}$). In Figure 8.18, the relation between the polarisations $P_{x,y}$, in which the initial quark direction is known, and the polarisations $\bar{P}_{x,y}$, in which it is estimated from the spectator quark, is shown. $\bar{P}_{x,y}$ is linearly dependent on $P_{x,y}$, and independent of the anomalous couplings in the considered intervals. The ranges of the polarisations displayed are those that correspond to varying $\mathrm{Re}\, g_R$ and $\mathrm{Im}\, g_R$ between -0.2 and 0.2. The obtained numerical relation between $\bar{P}_{x,y}$ and $P_{x,y}$ is

$$\begin{aligned} \bar{P}_{x,y} &= 0.89 P_{x,y} && (t)\,, \\ \bar{P}_{x,y} &= 0.81 P_{x,y} && (\bar{t})\,, \end{aligned} \tag{8.33}$$

where the same numerical factor was obtained for both the transverse and normal directions. This occurs due to the fact that choosing the other proton direction as the initial quark orientation, will only result in a negative sign in the $\hat{x}$ or $\hat{y}$ directions.

For LHC collisions at a center-of-mass energy of 14 TeV, the correction factors for the total cross-sections with anomalous couplings are

$$\begin{aligned} f_t &= 1 + 0.92\,|V_R|^2 + 1.82\,|g_L|^2 + 2.60\,|g_R|^2 - 0.11\,\mathrm{Re}\, V_L^* V_R - 0.47\,\mathrm{Re}\, V_L^* g_R \\ f_{\bar{t}} &= 1 + 1.07\,|V_R|^2 + 2.61\,|g_L|^2 + 1.92\,|g_R|^2 - 0.12\,\mathrm{Re}\, V_L^* V_R - 0.12\,\mathrm{Re}\, V_L^* g_R \\ &\quad -0.49\,\mathrm{Re}\, V_R^* g_L\,. \end{aligned} \tag{8.34}$$

The polarisations obtained for the three orthogonal axes are then:

$$\begin{aligned} P_x &= \left(-0.10|V_R|^2 - 0.84\,|g_R|^2 + 0.31|g_L|^2 + 2.19\,\mathrm{Re}\, V_L^* g_R - 1.68\,\mathrm{Re}\, V_R^* g_L\right)/f_t\,, \\ P_y &= \left(-2.17\,\mathrm{Im}\, V_L^* g_R - 1.69\,\mathrm{Im}\, V_R^* g_L\right)/f_t\,, \\ P_z &= \left(0.88 - 0.77\,|V_R|^2 + 1.38\,|g_L|^2 - 1.71\,|g_R|^2 - 0.53\,\mathrm{Re}\, V_L^* g_R + 0.35\,\mathrm{Re}\, V_R^* g_L\right)/f_t\,, \\ \bar{P}_{x,y} &= 0.88 P_{x,y} \end{aligned} \tag{8.35}$$

for top quarks and

$$\begin{aligned} P_x &= \left(-0.11 - 0.88|g_L|^2 + 0.37\,|g_R|^2 - 1.84\,\mathrm{Re}\,V_L^* g_R + 2.31\,\mathrm{Re}\,V_R^* g_L\right)/f_{\bar{t}}\,, \\ P_y &= \left(1.85\,\mathrm{Im}\,V_L^* g_R + 2.31\,\mathrm{Im}\,V_R^* g_L\right)/f_{\bar{t}}\,, \\ P_z &= \left(-0.85 + 0.95\,|V_R|^2 - 1.76\,|g_L|^2 + 1.43\,|g_R|^2 + 0.41\,\mathrm{Re}\,V_L^* g_R - 0.601, \mathrm{Re}\,V_R^* g_L\right)/f_{\bar{t}}\,, \\ \bar{P}_{x,y} &= 0.79 P_{x,y} \end{aligned} \tag{8.36}$$

for top antiquarks.

In order to experimentally probe the top polarisation, angular distributions of the top quark decay products can be used. Taking into account their spin analising properties, this study starts by considering the charged lepton, $\ell^{\pm}$, with three-momentum $\vec{p}_\ell$, and whose direction in the (x, y, z) reference system is parameterised by the angles (θ_ℓ, ϕ_ℓ). θ_ℓ is the angle between the three-momentum of the lepton (in the t rest frame) and the direction in which we want to measure the polarisation of the top quark. The angular distributions of the lepton in the top quark rest frame are given by

$$\frac{1}{\Gamma}\frac{d\Gamma}{d\cos\theta_\ell} = \frac{1}{2}(1 + P_z \alpha_\ell \cos\theta_\ell)\,, \tag{8.37}$$

where α_ℓ is the 'spin analyzing power' of the lepton, a constant which can range between -1 and 1 [329]. In the pure SM case, at LO, α_ℓ corresponds to unity. In the more general case, however, its value will depend on possible anomalous Wtb couplings.[1] Since leptons have the highest spin analyzing power, a forward-backward (FB) asymmetry is defined using the charged leptons as spin analysers for the top quark and antiquark decays:

$$A_{\mathrm{FB}}^z = \frac{\sigma(\cos\theta_\ell > 0) - \sigma(\cos\theta_\ell < 0)}{\sigma(\cos\theta_\ell > 0) + \sigma(\cos\theta_\ell < 0)} = \frac{1}{2}\alpha_\ell P_z\,, \tag{8.38}$$

which is sensitive to P_z. Considering the angles between $\vec{p}_\ell$ and the $\hat{x}$, $\hat{y}$ directions, two additional FB asymmetries can be established in the transverse and normal directions:

$$A_{\mathrm{FB}}^x = \frac{1}{2}\alpha_\ell \bar{P}_x\,, \quad A_{\mathrm{FB}}^y = \frac{1}{2}\alpha_\ell \bar{P}_y\,, \tag{8.39}$$

which are sensitive to $\bar{P}_x$ and $\bar{P}_y$ respectively.[2] This way, the polarisation of top quarks, and similarly for top antiquarks, can be measured.

[1] NLO corrections to α_ℓ are small [330].

[2] Alternatively, the transverse and normal polarisations could also be extracted from the azimuthal angle dependence of the double angular distribution $d\Gamma/d\cos\theta_\ell\, d\phi_\ell$. For instance, in [331] the azimuthal distributions were investigated for tW production, although no corresponding polarisations were extracted.

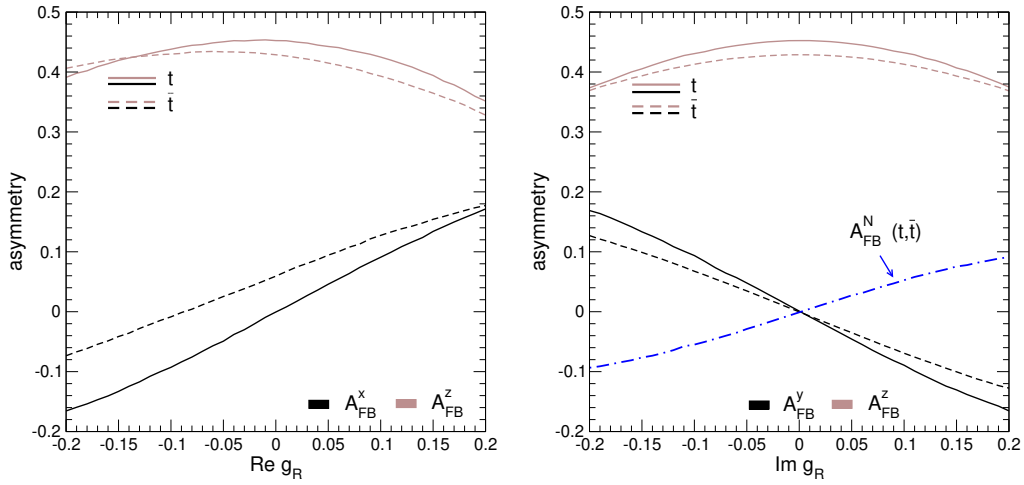


Figure 8.19: Forward-backward asymmetries sensitive to $\mathrm{Re}\, g_R$ (a) and $\mathrm{Im}\, g_R$ (b). For comparison, the normal W polarisation asymmetry A_{FB}^{N} is included on the right panel (b).

It is important to notice, however, that the top quark polarisation obtained via these three asymmetries is model-dependent, as new physics in the decay vertex can enter the distributions through the α_ℓ factor. For a model-independent measurement, double angular distributions involving the W boson rest-frame momenta would be necessary [332]. For the purpose of this study, however, probing the presence of new physics via anomalous Wtb couplings is enough. Studies involving other BSM sources, such as four-fermion ones, have been explored elsewhere [333, 334, 335].

In this study, it is from the measurement of A_{FB}^{x}, A_{FB}^{y} and A_{FB}^{z} that limits on the anomalous couplings are derived. Figure 8.19 depicts the asymmetries dependence on the real and imaginary parts of g_R, given via the polarisations P_i and α_ℓ. For completeness, in the right panel, the dependence of the normal W polarisation asymmetry A_{FB}^{N} [336] on the imaginary part of g_R is also included. The first direct limit on $\mathrm{Im}\, g_R$ was imposed using this asymmetry, which considers the charged lepton distribution in the W boson rest frame, as measured by the ATLAS Collaboration using 7 TeV: $A_{\mathrm{FB}}^{N} = 0.031 \pm 0.065\ (\mathrm{stat})^{+0.029}_{-0.031}\ (\mathrm{syst})$ [326]. Considering the proposed asymmetries, A_{FB}^{x} and A_{FB}^{y}, in Figure 8.19, a stronger dependence on $\mathrm{Re}\, g_R$ and $\mathrm{Im}\, g_R$ than for the A_{FB}^{z} asymmetry is seen. Experimentally, the measurement of either of these asymmetries ($A_{\mathrm{FB}}^{x,y,z}$ and A_{FB}^{N} alike) requires the full kinematical recon-

struction of single top quark events. Since both $A_{\rm FB}^{x,y}$ and $A_{\rm FB}^{N}$ vanish in the SM, [1] it is reasonable to expect experimental uncertainties with similar magnitudes. Additionally, it may be that the systematic uncertainties on $A_{\rm FB}^{x,y}$ are smaller, as these FB asymmetries have null values in the pure SM scenario, with the exception of $A_{\rm FB}^{x}$ in the single $\bar{t}$ production.

In the last couple of years, in order to measure the top quark polarisation in single top quark events, several techniques are applied by the LHC experiments. In general, and as previously suggested in this study too, the angular distribution of the charged lepton $\ell = e, \mu$ from the decay $t \to Wb \to \ell\nu b$ is used, since it is the most sensitive distribution to the top quark polarisation. The first of such measurements was done by the CMS experiment [337], where the top polarisation in the spectator quark direction was measured to be $P_z = 0.82 \pm 0.12\,(\text{stat}) \pm 0.32\,(\text{sys})$, at a centre-of-mass energy of 8 TeV, assuming the maximum spin analysing power of the lepton ($\alpha_\ell = 1$). Also at $\sqrt{s} = 8\,\text{TeV}$, the ATLAS experiment set a limit for the single top quark polarisation $|P_z| > 0.72$ (at 95% CL) at parton level [317]. Regarding the lepton spin asymmetry A_{FB}^{z}, it was measured by both experiments at $\sqrt{s} = 8\,\text{TeV}$ as $A_{FB}^{z} = 0.49 \pm 0.06$ by ATLAS [338] and $A_{FB}^{z} = 0.26 \pm 0.11$ by CMS [337]. At $\sqrt{s} = 13\,\text{TeV}$, the CMS collaboration measured $A_{FB}^{z} = 0.440 \pm 0.070$ [339]. All of these results were done at parton level and an average over top quarks and antiquarks was taken.

In this study, by introducing two new directions, transverse and normal, orthogonal to the already used spectator quark direction, a complete description of the top (anti)quark polarisation in single top production can be achieved. These new polarisations are particularly sensitive to an anomalous coupling involving a $\bar{b}_L \sigma^{\mu\nu} t_R$ dipole term. Specifically, the $A_{\rm FB}^{y}$ asymmetry, based on the polarisation normal to the production plane, is more sensitive to the imaginary part of this coupling than previously studied observables. The three FB asymmetries proposed in this study were recently measured by the ATLAS collaboration at a centre-of-mass energy of 13 TeV resulting in $P_x = 0.01 \pm 0.18$, $P_y = -0.029 \pm 0.027$ and $P_z = 0.91 \pm 0.10$ for top quarks, and $P_x = -0.02 \pm 0.20$, $P_y = -0.007 \pm 0.051$ and $P_z = -0.79 \pm 0.16$ for top antiquarks [340]. These measurements are consistent with the predictions made for the SM case, with the top quarks (antiquarks) polarised mostly in the longitudinal direction (opposite direction) and P_y null for both the top quarks and antiquarks. By combining the results from these new asymmetries with previously known observables, the overall sensitivity to anomalous right-handed vector(V_R), and left- and right-handed tensor (g_L,g_R) couplings and on the corresponding Wilson coefficients can be significantly improved. The

[1] Except $A_{\rm FB}^{x}$ for $\bar{t}$ as previously seen.

ATLAS collaboration fitted the measured differential cross-sections with an effective field theory prediction in order to set exclusion limits on the real and imaginary parts of the complex Wilson coefficient, C_{tW}. A non-zero value for the imaginary part of C_{tW} could be a possible indicator of beyond the SM CP violation in the tWb vertex. The real and the imaginary parts of C_{tW} were found to be within $[-0.9, 1.4]$ and $[-0.8, 0.2]$ respectively, at 95% CL, which is compatible with the SM predictions [340].

9

Conclusions

This dissertation covers a substantial range of aspects of a high energy physics programme: from detector monitoring, to an experimental physics search and the proposal of new phenomenological observables to probe the most successful theoretical framework as of yet, the SM of particle physics. The main focus of this doctoral work was the study of top quark physics, with particular emphasis in the production of top quark pairs in association with a Higgs boson, which is especially relevant, since the top quark-Higgs boson Yukawa coupling constitutes a test of the SM. As the doctoral work was carried out during the LHC's ramp up to its designed luminosity, interesting challenges arose, not only from the particle detection point of view, with detectors being put to the test in increasingly harsher conditions, but also from an analysis perspective, where the search of an already rare physics process is overwhelmed by a much larger background, with the constraint of having reduced data statistics in the regions of interest.

In order to reach precise measurements, an impressive and scrutinous surveillance of the ATLAS detector is required. In the first part of this doctoral work, the monitoring of the TileCal was performed under different pile-up conditions during the LHC's Run 1. As the conditions of the proton-proton collisions changed, the rising number of multiple collision lead to overlapping contributions in the energy deposits measured in each TileCal cell. The energy response of the TileCal cells was thus surveyed, and the associated noise was addressed, including inter-correlations among cells of the calorimeter. It was of particular importance to understand the sources of correlated pedestal noise in the energy measurement to correct them if possible. When hardware changes were not yet possible, software to reduce the correlated pedestal noise was tested in pile-up conditions ahead of the physics pp runs. Scans of covariance and correlation matrices between cell responses of all modules and

partitions of the calorimeter were made, as to completely survey the coherence noise effect on the TileCal. The TNF method, used to correct the observed channel correlation, was validated in pile-up collision data. Since the coherent noise component is mostly related to the hardware of the readout, by upgrading to a new generation of LVPS, the Gaussian shape of the channel noise and the magnitude of the correlations in TileCal modules improved: the intrinsic white noise distribution of each channel showed smaller RMS and is closer to a single Gaussian shape.

Additionally, the dependence on the multiplicity of pile-up events of the signal deposits measured on the TileCal cells and the associated noise was investigated. The dependencies with η, $BCID$ and number of interactions per bunch crossing were surveyed for the whole calorimeter, including all partitions, modules and layers. As the luminosity of data increased, the multiplicity of pile-up interactions was also higher, and it was observed that the mean energy deposited in a cell also increased. Periods with higher multiplicity of pile-up collisions also had noisier cell responses. The cell response was constant from run to run and reflects the geometry of the TileCal: cells closer to the IP (A-cells) or in more forward regions (gap crack cells), which are subjected to higher pile-up radiation, receive higher energy deposits and also show noisier outputs. It was also observed that data runs with higher number of pile-up interactions lead to cell measurements with higher values of noise, which follows the same pattern as the mean energy, being more significant for cells closer to the IP and for higher η values. The BCID dependence of the mean energy deposits averaged over ϕ modules and η towers, within the same layer and partition was also surveyed. The mean energy is mostly constant throughout the bunches position in the train. However, a slight tendency for higher values in the first and last bunches of the train was seen. The associated noise dependence with BCID showed a similar response, with fluctuations within a few MeV along the train and slightly noisier signal measurements for the first and last bunches in each train. In addition, it was observed that as the number of interactions per bunch crossing increases, the cell noise also increased. The rate to which the noise depends on the number of interactions is specific of the cell location in the calorimeter. As expected, cells which undergo higher levels of pile-up radiation, such as A-cells or special cells, present typically noisier measurements, with values growing with $\sqrt{<\mu>}$. The noise of the outermost cells, located in the BC or D layers, showed linear proportionality to the average number of interactions in a bunch crossing. The closely monitored response of the TileCal cells in different conditions contributed to a proper characterisation of the TileCal response and

performance, allowing better object reconstruction and thus more precise results of the ATLAS experiment. This study encompasses complete survey of the TileCal cell noise dependence with the multiplicity of pile-up events, summarised in 'Description of the Tile Calorimeter noise with increasing Pile-up for $\sqrt{s} = 7\,\text{TeV}$ data collected during 2011'[2], and gave way for the continuous monitoring of the TileCal cell response throughout Run 1 of the LHC.

The search for $t\bar{t}H$ production in the dileptonic channel of top quark decays, focusing on a SM Higgs boson $H \to b\bar{b}$ decay, was performed. The experimental search for this production mode was done using ATLAS collected data at a centre-of-mass energy of $\sqrt{s} = 8\,\text{TeV}$ with a luminosity of $20.3\,\text{fb}^{-1}$. The results presented in this dissertation led to the publication of the ATLAS Collaboration paper 'Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s} = 8\,\text{TeV}$ with the ATLAS detector' [1]. After a dedicated dileptonic event selection, the events were classified into jet and b-tagged jet multiplicity regions, improving the signal sensitivity without loss of events. In fact, the improvement of the signal sensitivity was the cornerstone of the analysis, for which various measures were implemented. As the precise modelling of the different background processes was one of the main challenges, a significant effort was devoted to estimate the processes to the best precision possible. Data-driven methods were developed, using dedicated control-regions and event selection criteria for the Z + jets, the $t\bar{t}$ + jets and multijet processes. The backgrounds normalisations (and shape modelling in the case of the Z + jets) were then corrected to the best precision available at the time, *i.e.* that of the 8 TeV data. It should be noted that these data-driven methods were performed for the first time as exploratory alternatives to those implemented in the ATLAS published result, giving similar sensitivities to the top quark-Higgs boson vertex. In particular, for the $t\bar{t}$ + jets, where no shape modelling was estimated in this thesis, a systematic variation was included based on the top quark and top quark pairs transverse momenta correction used in the nominal semi-leptonic and dileptonic ATLAS analyses [1]. Additionally, for the $t\bar{t}$ production with extra heavy-flavour jets, for which no differential measurements had been performed yet, systematic uncertainties on the modelling of the individual $t\bar{t} + b\bar{b}$ jets and $t\bar{t} + c\bar{c}$ jets processes were estimated. The signal to background discrimination was achieved by using neural networks in signal-rich regions. The neural networks were trained using kinematic variables of the final state objects such as the p_{T} and η distributions, as well as event shape distributions and global event variables requiring final state information from

all objects in the event. A combined likelihood fit to data in all the jet multiplicity regions was employed, including systematic uncertainties as nuisance parameters in the fit. The systematic uncertainties, both theoretical and experimental, were significantly constrained by the available data. By exploiting signal-depleted regions with higher statistics to constrain in-situ the leading uncertainties, the background modelling also improved. The previously mentioned systematic variations had the highest impact on the likelihood fit. No significant excess of $t\bar{t}H$ events above the background expectation was found for a Standard Model Higgs boson with a mass of 125 GeV. An observed (expected) 95% confidence-level upper limit was set at 7.8 (4.5) times the Standard Model cross section. The ratio of the measured signal strength to the Standard Model expectation was found to be $\mu = 3.7 \pm 2.2$ under the signal-plus-background hypothesis. This result is in agreement with those of both ATLAS and CMS for the same centre-of-mass energy and equivalent luminosities.

As the LHC continued data-taking, providing pp collisions at 13 TeV, the production of SM Higgs boson in association with top quark pairs was finally observed [93, 94]. Benefiting from the increase in energy, as well as further developing the experimental strategies used in Run 1, a more powerful search was achieved. The combination of different Higgs boson decay channel searches provided statistical significance making the discovery possible. Although, no deviations from the SM predictions have been observed so far, this opens the door to precision measurements of the top quark couplings to the Higgs boson.

In this dissertation, new top quark observables were also proposed to investigate the nature of the top quark couplings, both in neutral and charged currents. These studies were published in 'Angular distributions in $t\bar{t}H(H \to b\bar{b})$ reconstructed events at the LHC' [3], 'Probing the CP nature of the Higgs coupling in $t\bar{t}h$ events at the LHC' [4] and 'New directions for top quark polarisation in the t-channel process' [5].

The top quark-Higgs boson yukawa coupling was studied in $t\bar{t}h$ production, for scalar and pseudo-scalar Higgs bosons, at the phenomenological level. Using the expertise developed for the experimental search, a dileptonic selection and full kinematic reconstruction of t, $\bar{t}$ and and $h \to b\bar{b}$ decays was performed using $\sqrt{s} = 13$ TeV simulated pp collisions at the LHC. Although, there is a reduction on the total number of events when compared with the initial parton level, the kinematical reconstruction preserves the different spin information of the signal and background processes. Having access to relevant spin information allowed the exploratory study of several observables, including newly proposed angular distributions and asymmetries that provided better discrimination not only between signals of different

nature (scalar or pseudoscalar) but also between $t\bar{t}h$ events and the main $t\bar{t}$ background at the LHC. For completeness, variables already proposed in literature particularly suited to the discrimination between the CP components of the $t\bar{t}h$ signal, such as the b_4, were also investigated. Expected limits at 95% CL were extracted on the $\sigma \times BR(h \to b\bar{b})$ and signal strength μ using a boosted decision tree. A comparison between the sensitivities of the individual variables as a function of $\cos(\alpha)$ was also performed, showing that a MVA method combining all variables can improve the individual limits up to 25%. The obtained expected limits do not have a strong dependence on the particular choice of the CP-phase (α), however, the analysis of the SM Higgs case (CP-even) is a good starting point to probe mixtures with CP-odd contributions. These results are expected to improve when other decay channels are combined, using fully reconstructed final states. This exploratory work led to continuous phenomenological studies, including studies with semileptonic top quark pair decays [341]–[345]. On the experimental front of the search for $t\bar{t}H/A$ production, the ATLAS Collaboration is currently preparing a new publication 'Probing the CP nature of the top-Higgs Yukawa coupling in $t\bar{t}H$ and tH events with $H \to b\bar{b}$ using the ATLAS detector at the LHC' based on the conference note [346]. This soon to be published result makes use of CP sensitive observables first tested on reconstructed events in this dissertation. It is worth mentioning that the search for four top quarks production is also sensitive to the top quark Yukawa coupling [347].

The top quark coupling to the charged W bosons was also explored in this thesis, making use of single top quark simulated events. A complete description of the top quark and top antiquark polarisations was achieved by introducing two directions, transverse and normal, which are orthogonal to the already commonly used spectator quark direction. Given the polarisations P_x, P_y and P_z in the three proposed axes, the top (anti)quark spin density matrix can be determined and the polarisation in any other direction can be computed. It was observed that the polarisations P_x and P_y vanish in the SM for the single top quark production, whereas for top antiquark production P_x is of the order of $\mathcal{O}(0.1)$ with null P_y. The dependence of these polarisations on possible anomalous Wtb couplings was investigated. Two asymmetries, involving the transverse (P_x) and normal (P_y) polarisations, were shown to be sensitive to an anomalous coupling g_R involving a dipole term of the form $\bar{b}_L \sigma^{\mu\nu} t_R$. The three FB asymmetries proposed in this study were recently measured by the ATLAS collaboration at a centre-of-mass energy of 13 TeV [340], and are consistent with the predictions made for the SM case, with the top quarks (antiquarks) polarised mostly

in the longitudinal direction (opposite direction) and P_y null for both the top quarks and antiquarks. By combining the results from these new asymmetries with previously known observables, the overall sensitivity to anomalous right-handed vector(V_R), and left- and right-handed tensor (g_L,g_R) couplings and on the corresponding Wilson coefficients can be significantly improved.

The experimental and phenomenological analysis developed in this thesis, have been fruitful in laying the ground work for continuous investigation in top quark and Higgs boson physics, by contributing to more precise detector measurements, testing new experimental strategies and proposing novel CP sensitive observables.

Appendix A

Energy and Noise Dependence on the Bunch Crossing ID in TileCal

A systematic survey of the TileCal cell signal dependence on the BCID is presented. The mean energy deposits and the respective significance (RMS) were estimated in zero bias data from pp colisions at $\sqrt{s} = 7\,\text{TeV}$. All sampling layers of the calorimeter were individually monitored, maintaining the LB and EB separate, but inclusively averaging all the azimuthal modules. The points represent the data samples, one run for each period in Table 4.2. The lines are the MC simulation reweighted to the pile-up conditions of each period. The lowest luminosity period (F2), represented by the yellow line, was chosen as a reference to guide the eye. Overall, the noise of the TileCal cells increases for data with higher multiplicity of pile-up.

A.1 Energy dependence with the BCID

Figures A.1–A.6 show the mean energy deposits as a function of the BCID for the sampling layers of the LB and EB. These are analogous to Figure 4.19 for sample A of the LB shown in Chapter 4. The mean energy deposits flutuate a few MeV as a function of the BCID. On average, the first and last bunches of the trains lead to higher energy deposits in the cells. For both barrels, sampling layers closer to the beam pipe have more energetic deposits than those on the outermost parts of the calorimeter. Due to their forward location, the EB cells record higher energy deposits than the LB ones.

A.2 Noise dependence with the BCID

The total noise, estimated as the RMS of the energy signal, is shown in Figures A.7–A.12 as a function of the BCID for the sampling layers of the LB and EB partitions of the TileCal. These are analogous to Figure 4.20 for sample A of the LB shown in Chapter 4. The noise is constant within a few MeV, with the first and last bunches of the trains showing, on average, slightly noisier signal measurements. Just as expected, cells near the IP have noisier measurements, because of the intense exposure to pile-up events. The LB cells have higher values of noise, which reflect the forward pseudorapidity coverage of these modules. The gap-crack cells provide noisier outputs, mainly because of their location and special layout.

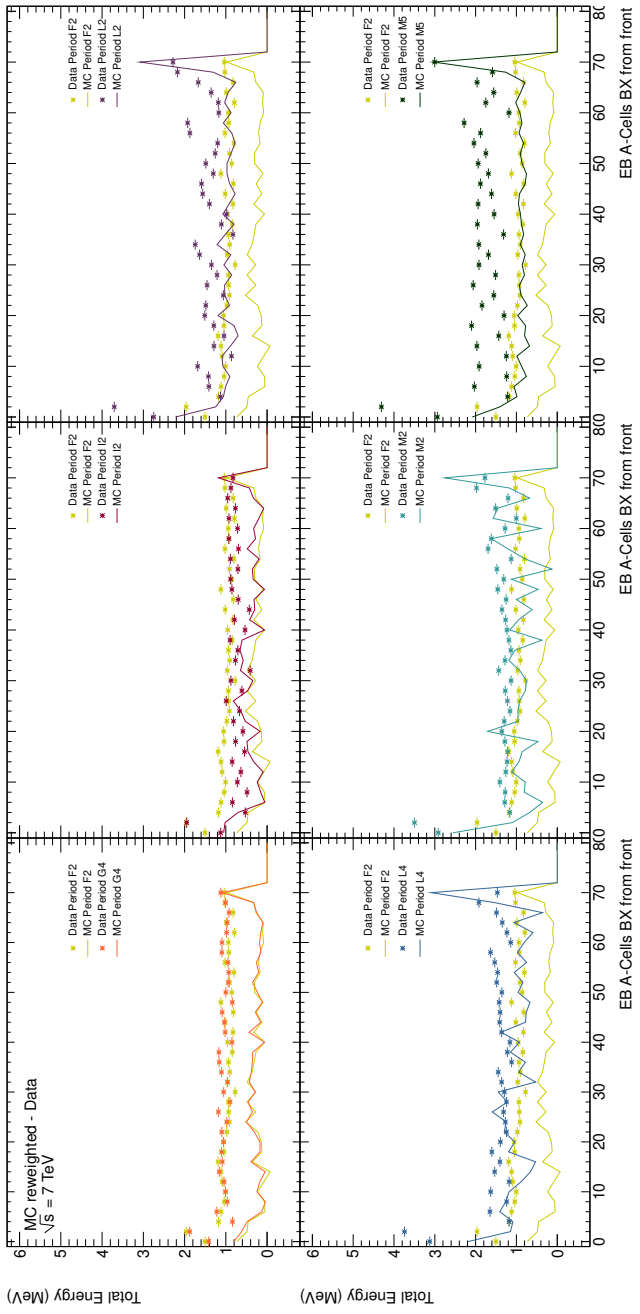


Figure A.1: The mean energy depositions as a function of the BCID for sample A of the extended barrel is represented, for several runs of 2011 data. The lowest luminosity period (F2) was chosen as a reference. The marker points represent the data sample (one run for each period) and the lines are the Monte Carlo distribution (mc11c reweighted to each data period). The MC was reweighted to the data, according to the weight given by the PileUpReweighting tool (based on the average number of interactions per bunch crossing).

A. ENERGY AND NOISE DEPENDENCE ON THE BUNCH CROSSING ID IN TILECAL

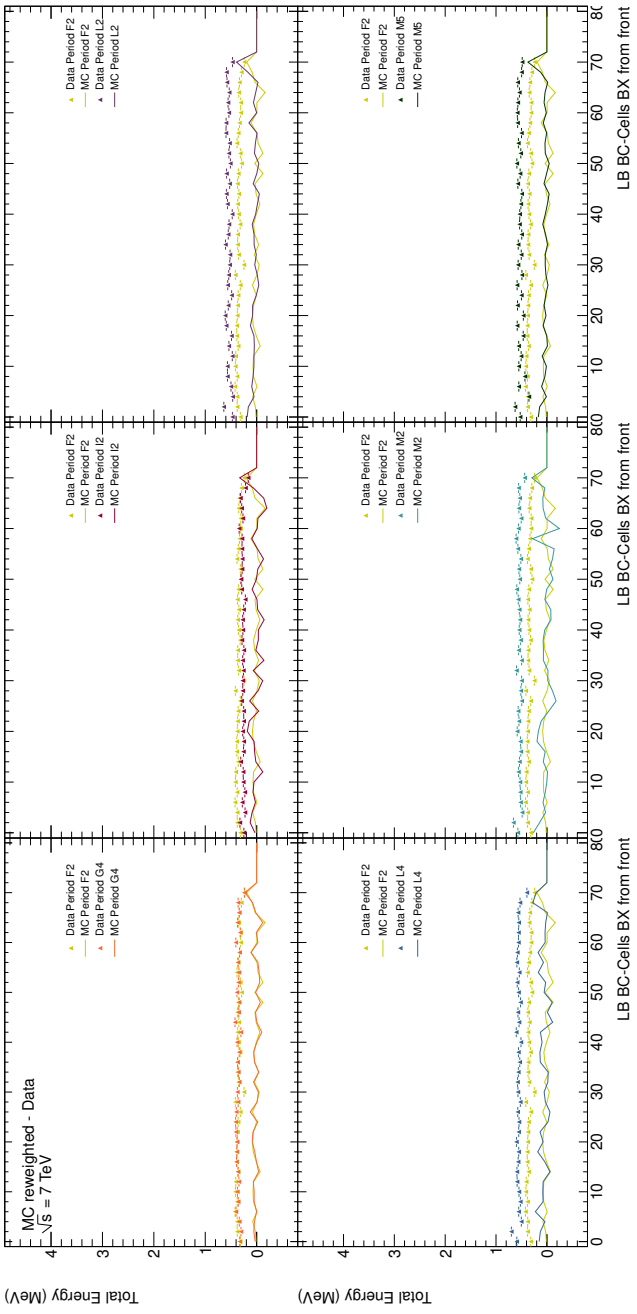


Figure A.2: The mean energy depositions as a function of the BCID for sample BC of the long barrel is represented, for several runs of 2011 data. The lowest luminosity period (F2) was chosen as a reference. The marker points represent the data sample (one run for each period) and the lines are the Monte Carlo distribution (mc11c reweighted to each data period). The MC was reweighted to the data, according to the weight given by the PileUpReweighting tool (based on the average number of interactions per bunch crossing).

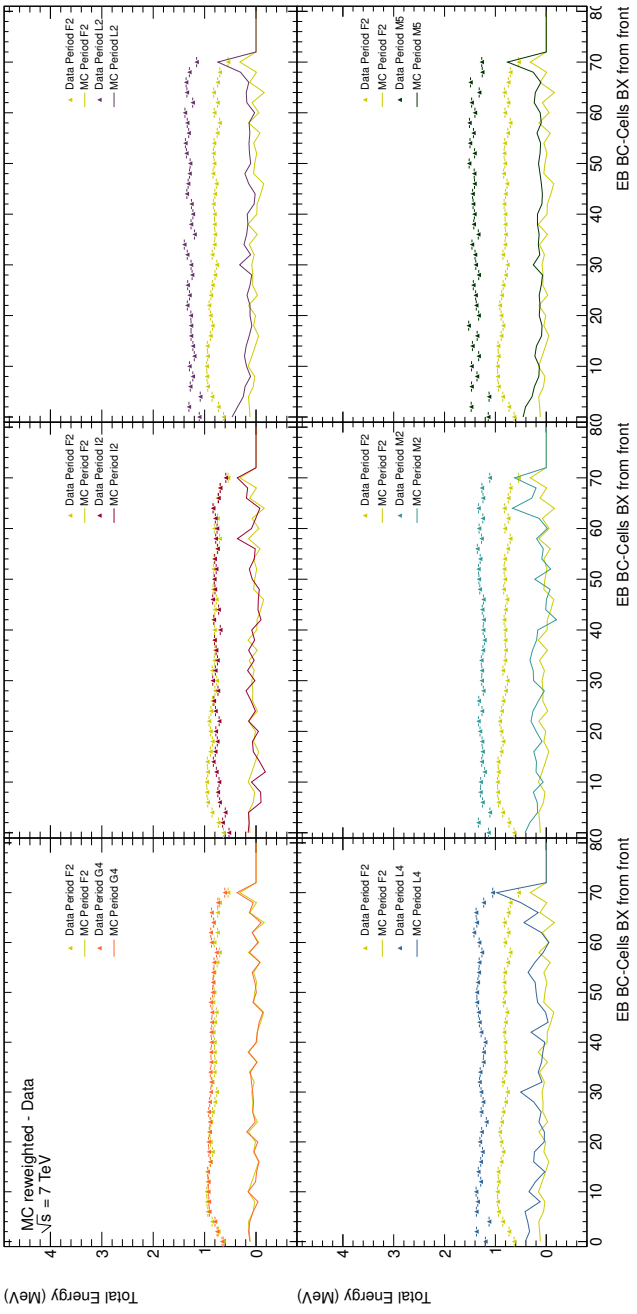


Figure A.3: The mean energy depositions as a function of the BCID for sample BC of the extended barrel is represented, for several runs of 2011 data. The lowest luminosity period (F2) was chosen as a reference. The marker points represent the data sample (one run for each period) and the lines are the Monte Carlo distribution (mc11c reweighted to each data period). The MC was reweighted to the data, according to the weight given by the PileUpReweighting tool (based on the average number of interactions per bunch crossing).

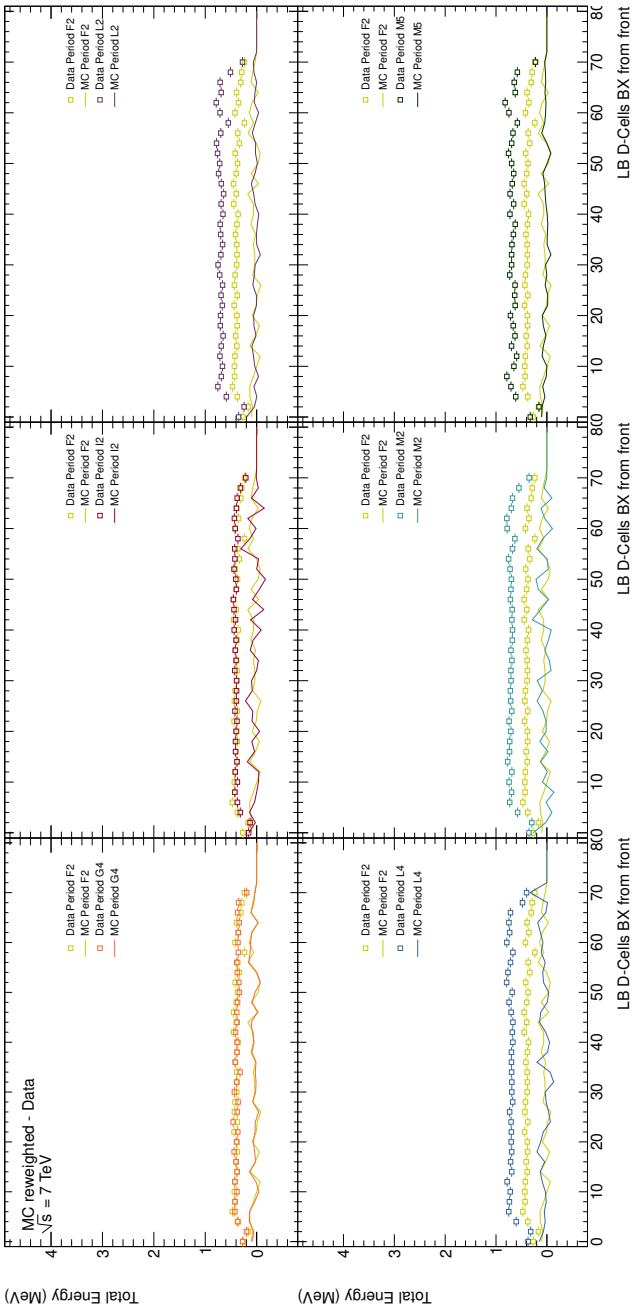


Figure A.4: The mean energy depositions as a function of the BCID for sample D of the long barrel is represented, for several runs of 2011 data. The lowest luminosity period (F2) was chosen as a reference. The marker points represent the data sample (one run for each period) and the lines are the Monte Carlo distribution (mc11c reweighted to each data period). The MC was reweighted to the data, according to the weight given by the PileUpReweighting tool (based on the average number of interactions per bunch crossing).

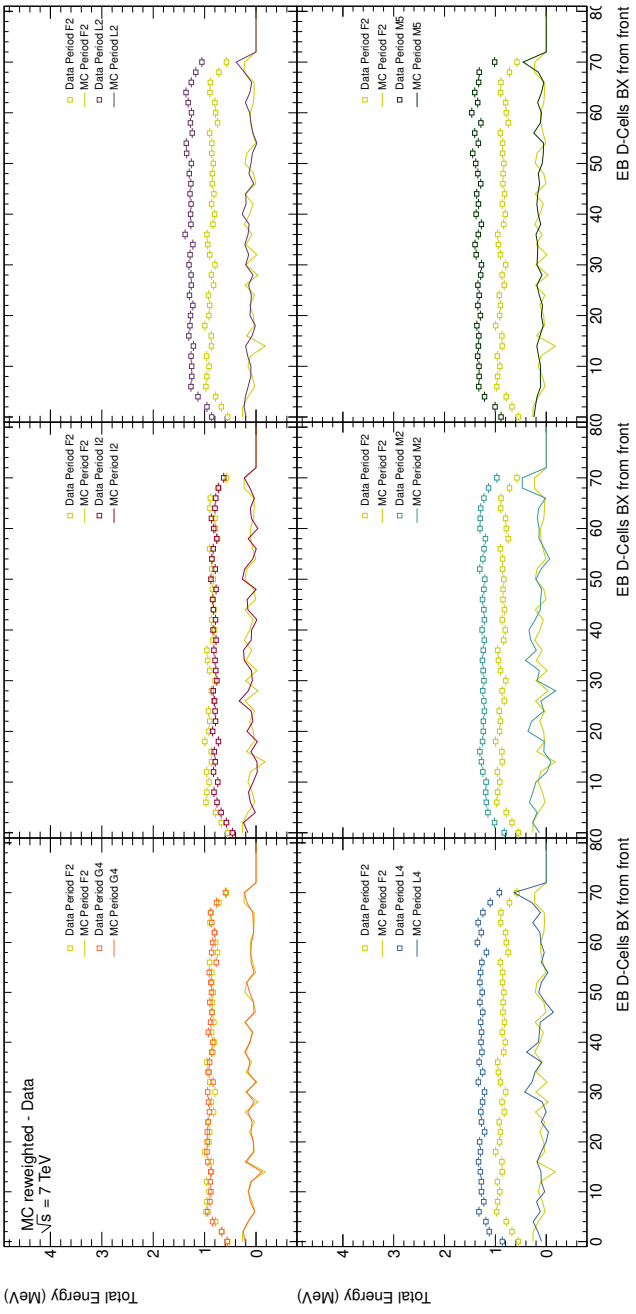


Figure A.5: The mean energy depositions as a function of the BCID for sample D of the extended barrel is represented, for several runs of 2011 data. The lowest luminosity period (F2) was chosen as a reference. The marker points represent the data sample (one run for each period) and the lines are the Monte Carlo distribution (mc11c reweighted to each data period). The MC was reweighted to the data, according to the weight given by the PileUpReweighting tool (based on the average number of interactions per bunch crossing).

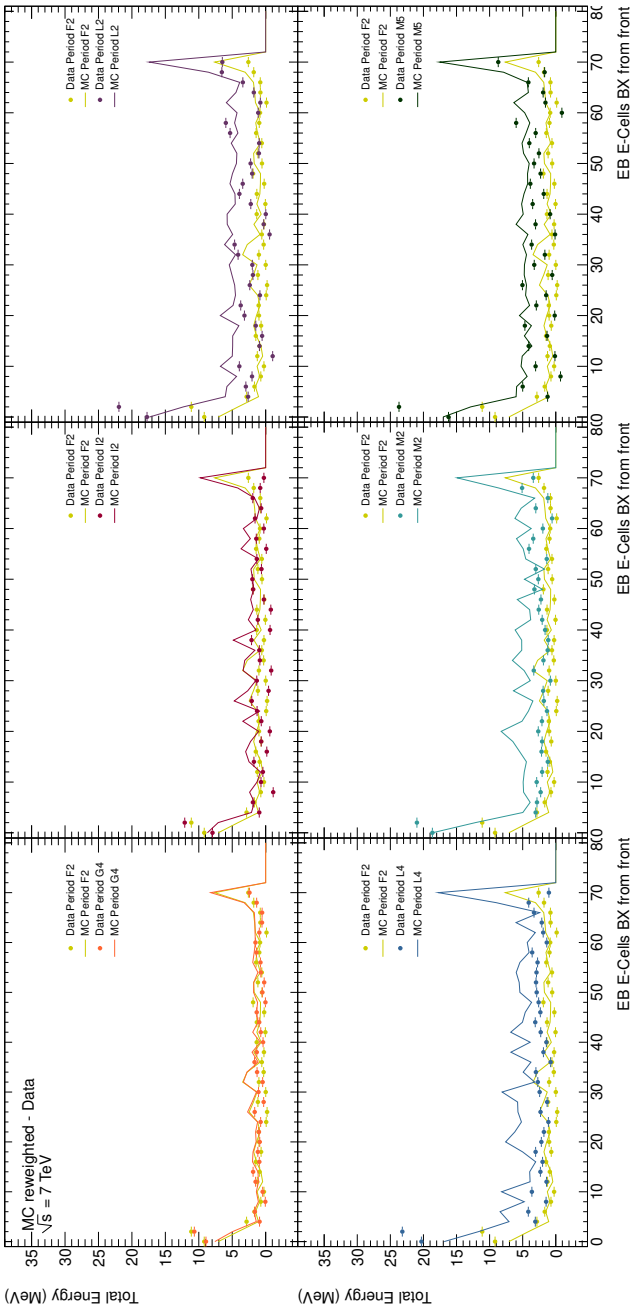


Figure A.6: The mean energy depositions as a function of the BCID for the special cells of the extended barrel is represented, for several runs of 2011 data. The lowest luminosity period (F2) was chosen as a reference. The marker points represent the data sample (one run for each period) and the lines are the Monte Carlo distribution (mc11c reweighted to each data period). The MC was reweighted to the data, according to the weight given by the PileUpReweighting tool (based on the average number of interactions per bunch crossing).

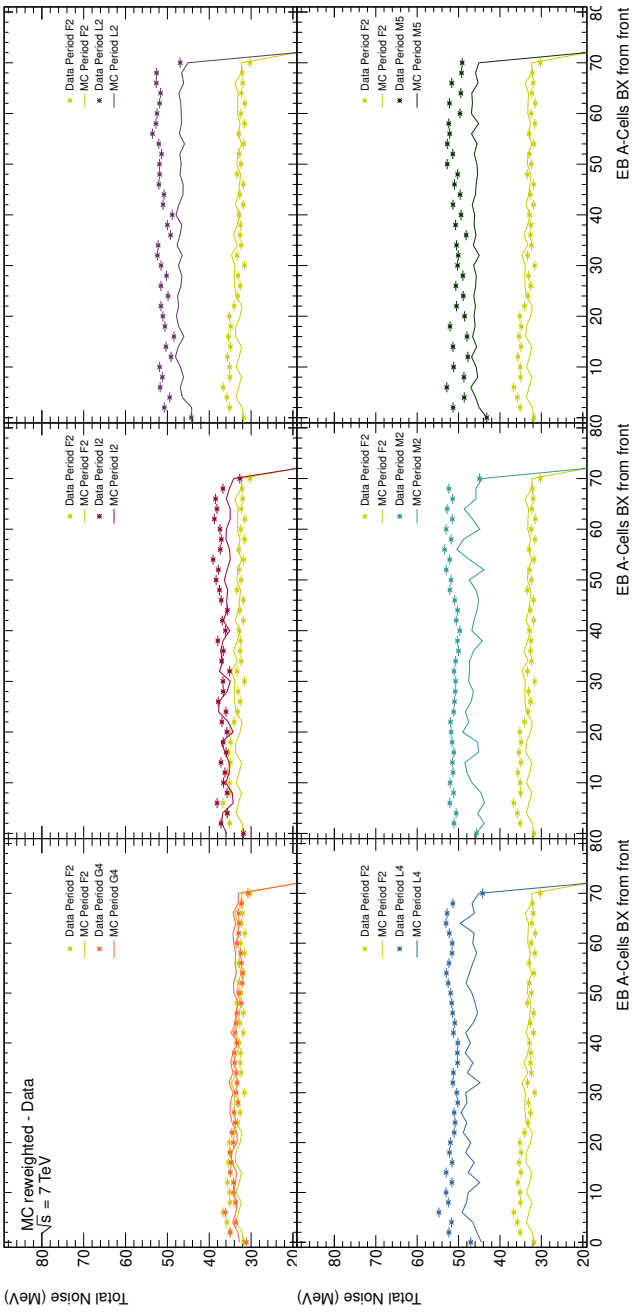


Figure A.7: The total noise as a function of the BCID for sample A of the extended barrel is represented, for several runs of 2011 data. The lowest luminosity period (F2) was chosen as a reference. The noise was estimated as the standard deviation of the energy value per cell per bcid. The marker points represent the data sample (one run for each period) and the lines are the Monte Carlo distributions (mc11c reweighted to each data period). The MC was reweighted to the data, according to the weight given by the PileUpReweighting tool (based on the average number of interactions per bunch crossing).

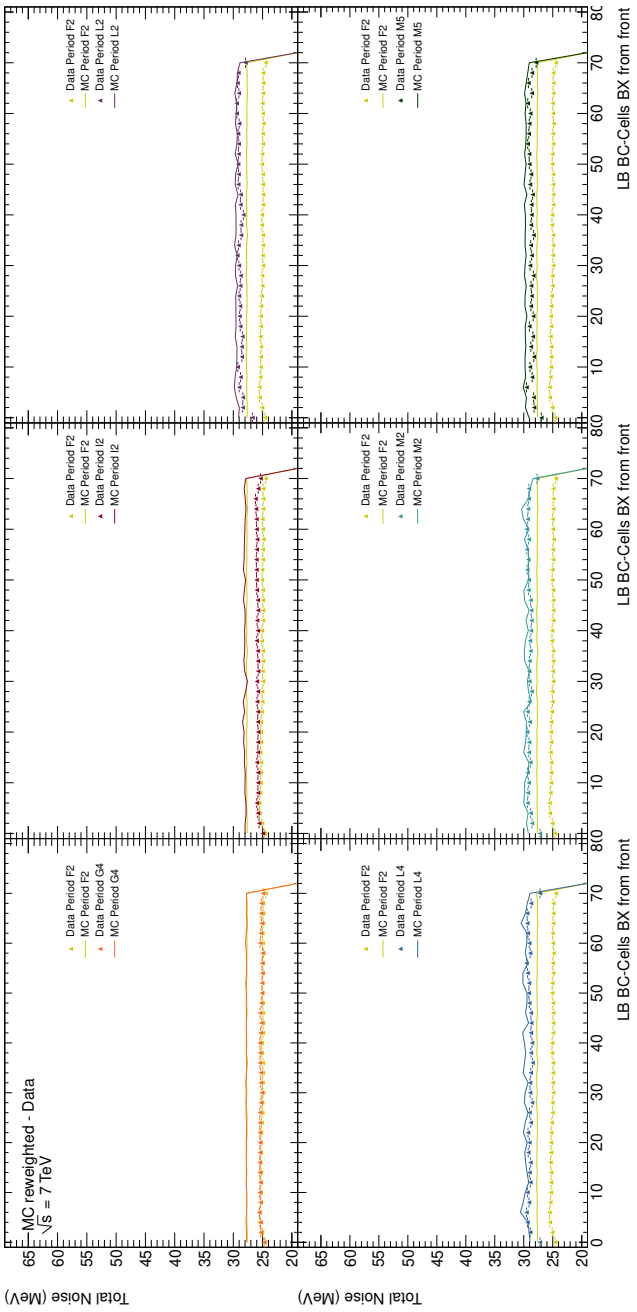


Figure A.8: The total noise as a function of the BCID for sample BC of the long barrel is represented, for several runs of 2011 data. The lowest luminosity period (F2) was chosen as a reference. The noise was estimated as the standard deviation of the energy value per cell per bcid. The marker points represent the data sample (one run for each period) and the lines are the Monte Carlo distributions (mc11c reweighted to each data period). The MC was reweighted to the data, according to the weight given by the PileUpReweighting tool (based on the average number of interactions per bunch crossing).

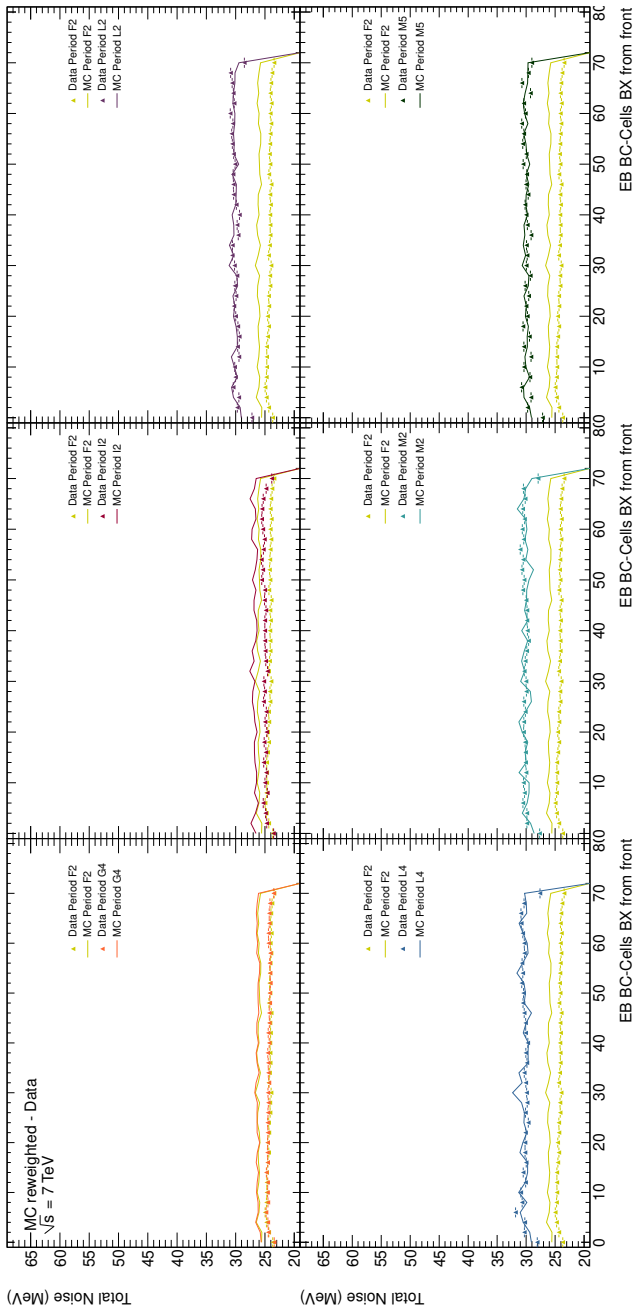


Figure A.9: The total noise as a function of the BCID for sample BC of the extended barrel is represented, for several runs of 2011 data. The lowest luminosity period (F2) was chosen as a reference. The noise was estimated as the standard deviation of the energy value per cell per bcid. The marker points represent the data sample (one run for each period) and the lines are the Monte Carlo distributions (mc11c reweighted to each data period). The MC was reweighted to the data, according to the weight given by the PileUpReweighting tool (based on the average number of interactions per bunch crossing).

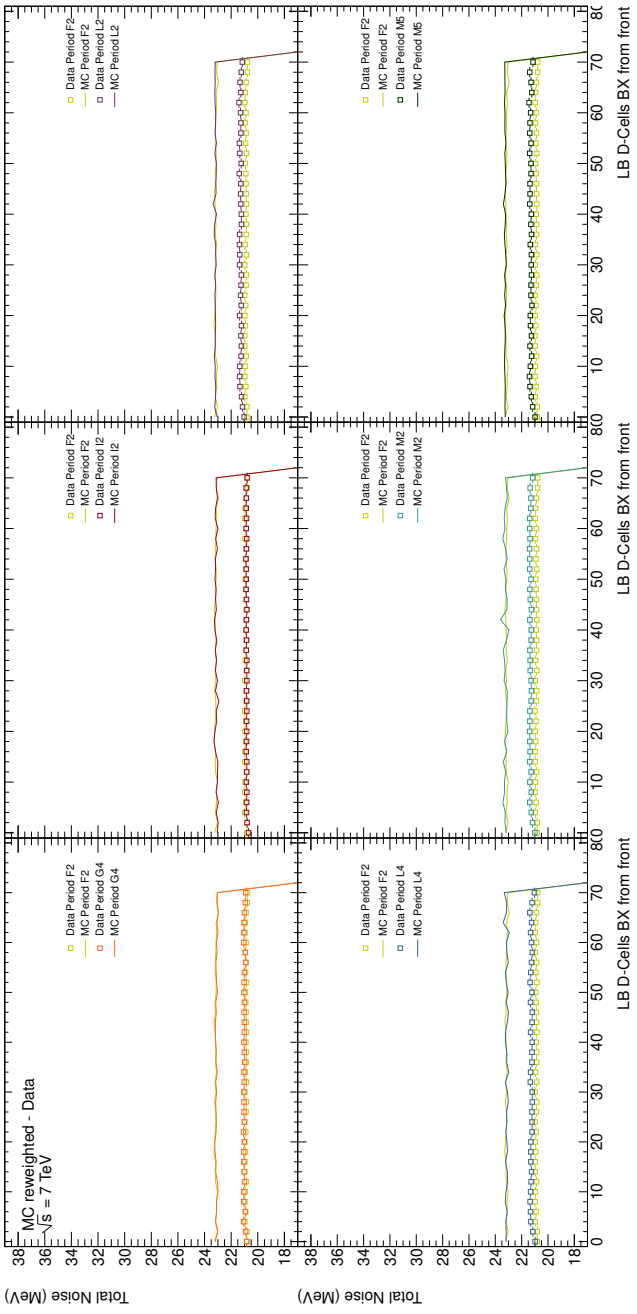


Figure A.10: The total noise as a function of the BCID for sample D of the long barrel is represented, for several runs of 2011 data. The noise was estimated as the standard deviation of the energy value per cell per bcid. The marker points represent the data sample (one run for each period) and the lines are the Monte Carlo distributions (mc11c reweighted to each data period). The MC was reweighted to the data, according to the weight given by the PileUpReweighting tool (based on the average number of interactions per bunch crossing).

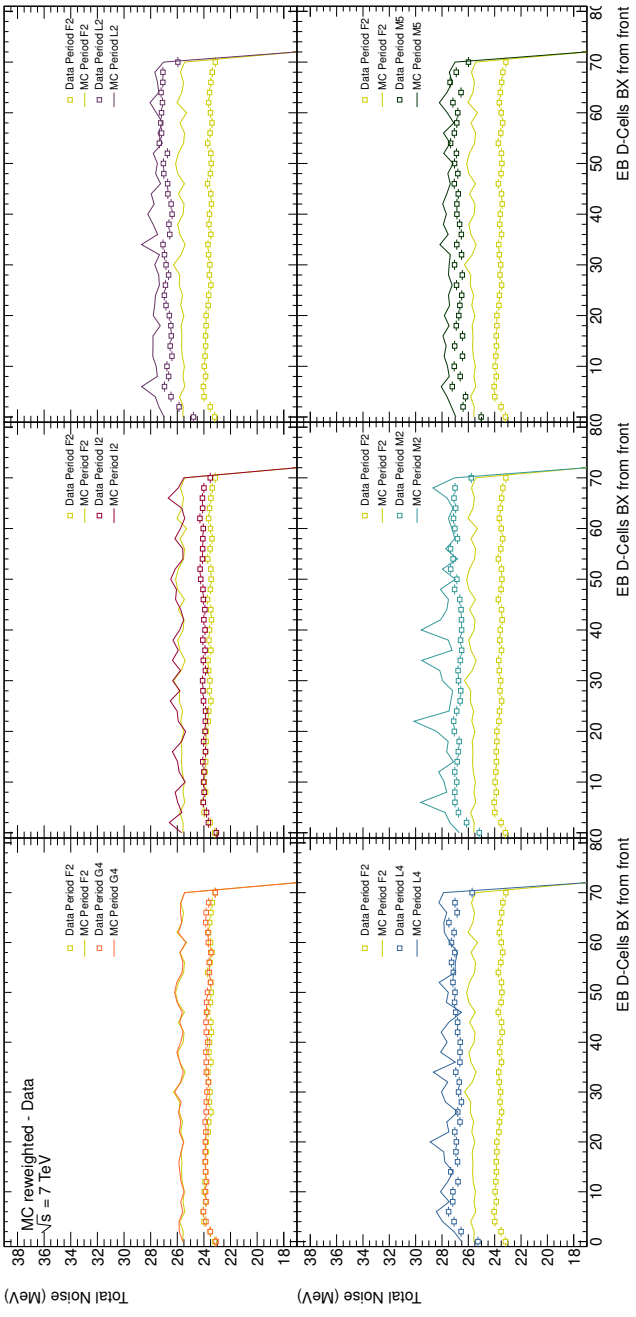


Figure A.11: The total noise as a function of the BCID for sample D of the extended barrel is represented, for several runs of 2011 The noise was estimated as the standard deviation of the energy value per cell per bcid. The marker points represent the data sample (one run for each period) and the lines are the Monte Carlo distributions (mc11c reweighted to each data period). The MC was reweighted to the data, according to the weight given by the PileUpReweighting tool (based on the average number of interactions per bunch crossing).

A. ENERGY AND NOISE DEPENDENCE ON THE BUNCH CROSSING ID IN TILECAL

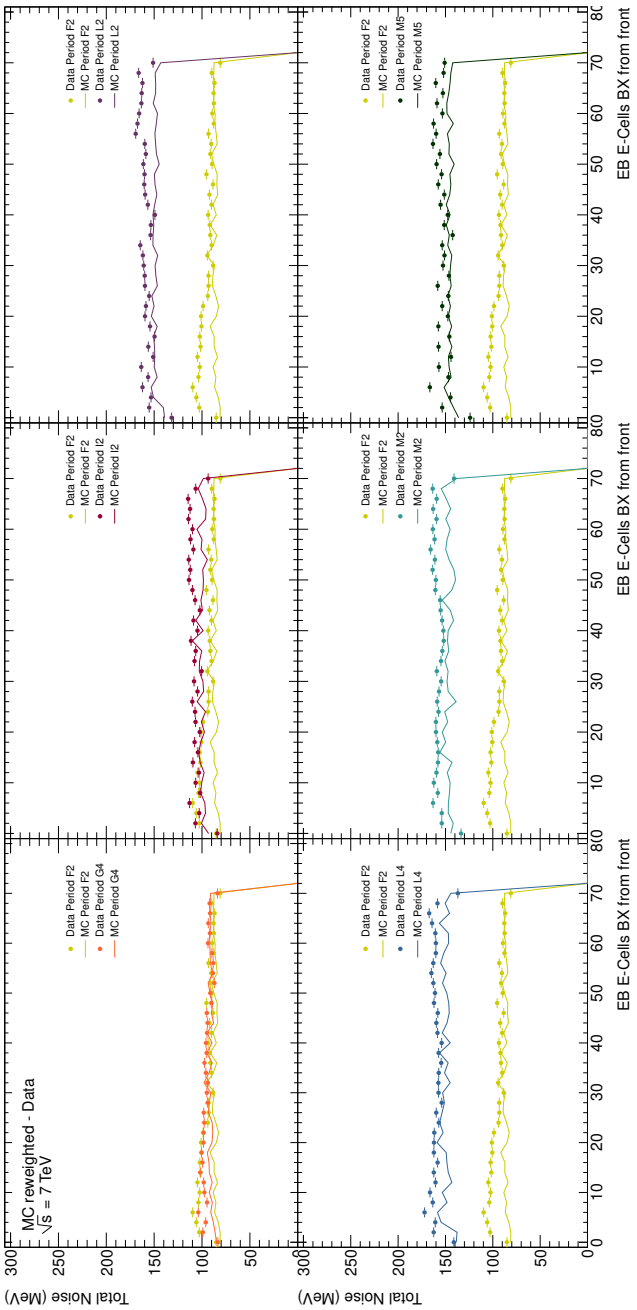


Figure A.12: The total noise as a function of the BCID for the special cells of the extended barrel is represented, for several runs of 2011 data. The lowest luminosity period (F2) was chosen as a reference. The noise was estimated as the standard deviation of the energy value per cell per bcid. The marker points represent the data sample (one run for each period) and the lines are the Monte Carlo distributions (mc11c reweighted to each data period). The MC was reweighted to the data, according to the weight given by the PileUpReweighting tool (based on the average number of interactions per bunch crossing).

Appendix B

Neural Network Input Variables

The discrimination between signal and background is shown in Figures B.1 and B.2 for the ten highest ranked observables in the ($\geq$ 4j, 3b) and (3j, 3b) regions where NN is used. These are analogous to Figure 7.6 for the ($\geq$ 4j, $\geq$ 4b) region with the highest signal sensitivity. The distributions are normalised to unity. The correlations between observables is also included for both signal and background.

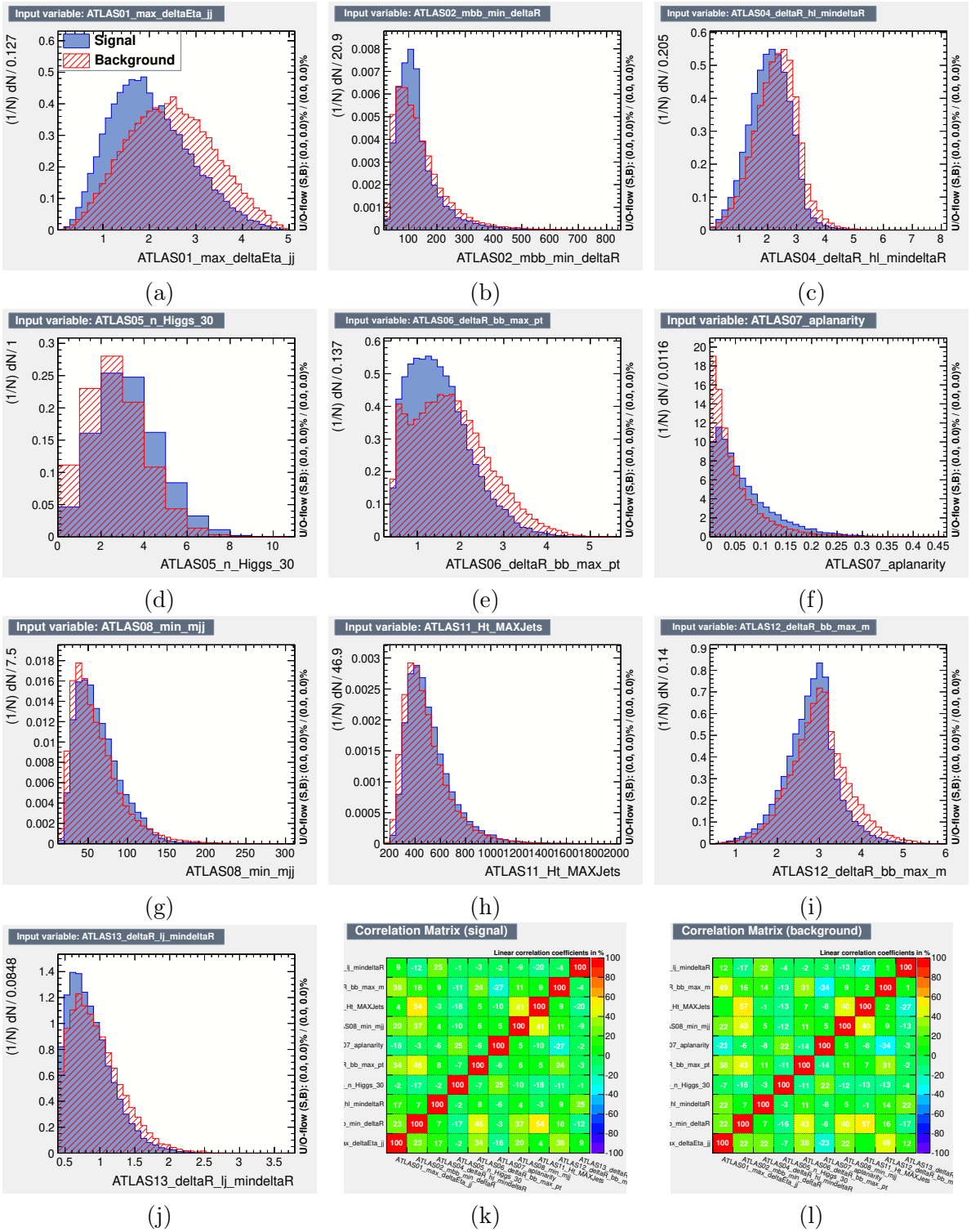


Figure B.1: Normalised distributions of the NN highest-ranked input variables in the ($\geq$ 4j, 3b) region: (a) $\Delta\eta_{jj}^{\max\Delta\eta}$; (b) N_{30}^{Higgs}; (c) m_{jj}^{minm}; (d) $\Delta R_{bb}^{\max \mathrm{p_T}}$; (e) $\Delta R_{hl}^{\min\Delta\mathrm{R}}$; (f) H_{T}; (g) $\mathrm{Aplan}_{\mathrm{jet}}$; (h) $m_{bb}^{\min\Delta\mathrm{R}}$; (i) $\Delta R_{bb}^{\max \mathrm{m}}$; (j) $\Delta R_{lj}^{\min\Delta\mathrm{R}}$; The correlation among the variables are also shown, (k) for the signal and (l) for the background events.

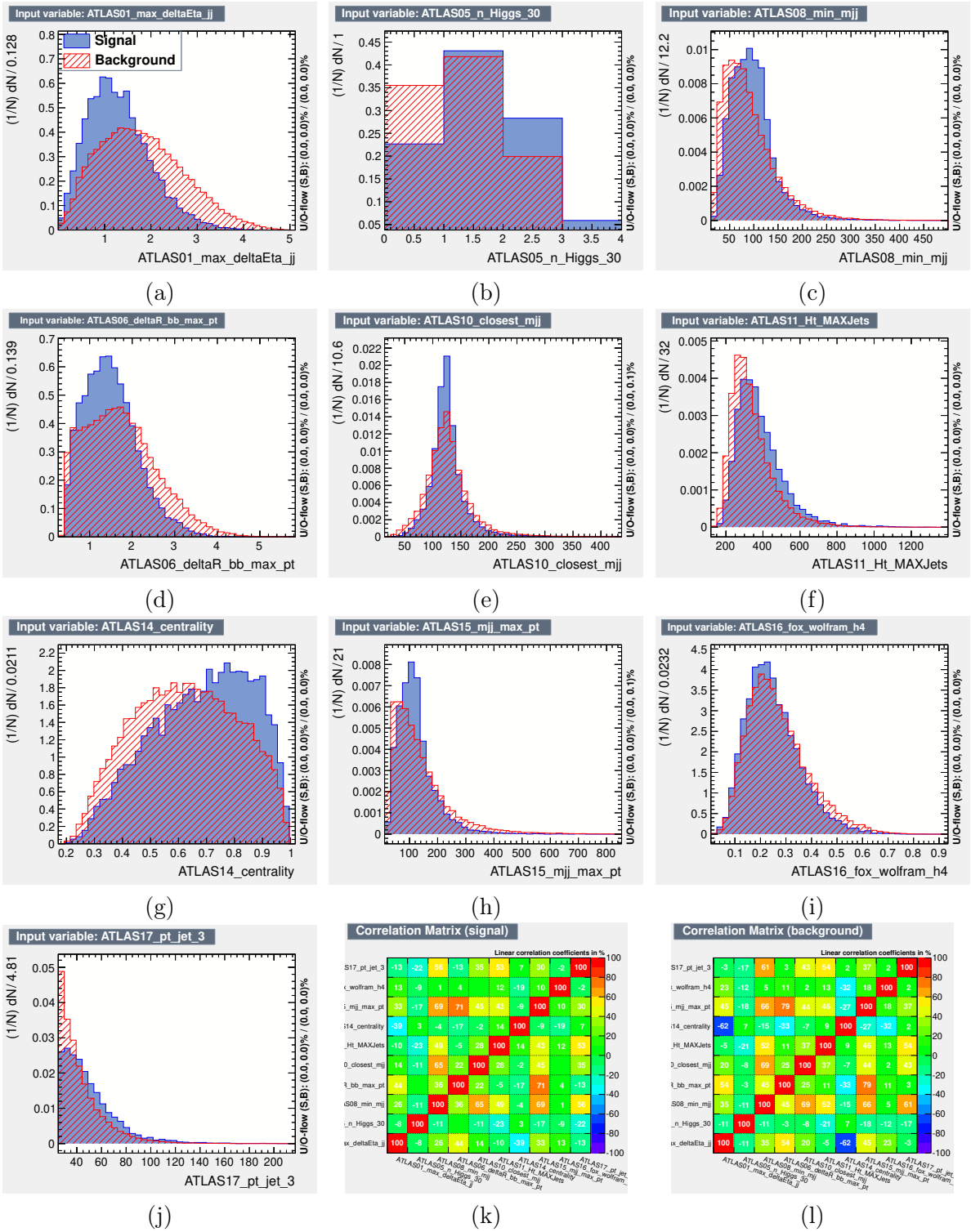


Figure B.2: Normalised distributions of the NN highest-ranked input variables in the (3j, 3b) region: (a) $\Delta\eta_{jj}^{\max \Delta\eta}$; (b) $m_{jj}^{\min \mathrm{m}}$; (c) H_{T}; (d) $H4$; (e) N_{30}^{Higgs}; (f) $p_{\mathrm{T}}^{\mathrm{jet}\,3}$; (g) Centrality; (h) $\Delta R_{bb}^{\max \mathrm{p_T}}$; (i) $m_{jj}^{\max \mathrm{p_T}}$; (j) $m_{jj}^{\mathrm{closest}}$. The correlation among the variables are also shown, (k) for the signal and (l) for the background events.

Appendix C

$t\bar{t}h$ Event Reconstruction Performance

Comparisons of three distinct steps of a dileptonic analysis using the kinematic reconstruction method in Section 8.2.1:

1. after event selection — using the parton-level four-momenta of all relevant objects;
2. after event selection and a truth-matched kinematic reconstruction — using the reconstructed four-momenta of all relevant objects, while using truth-matched information to select the jet-lepton pairing and the jets originating from the Higgs boson;
3. after event selection and the full kinematic fit reconstruction — using the fully reconstructed four-momenta, without any truth-matched objects.

Angular distributions of x_{ℓ^+} on the left and x_{ℓ^-} on the right are shown in Figures C.1, for both the $t\bar{t}h$ signal and the main background $t\bar{t}b\bar{b}$. The two processes are normalised for better visualisation. Figures C.1 (a) and (b) show the parton level angular distributions after the event selection. In Figures C.1 (c) and (d), the angular distributions calculated at truth level, after selection cuts and kinematical fit with truth-matched objects; Finally, in Figures C.1 (e) and (f), the angular distributions obtained using the four-momenta after the full event analysis: event selection cuts and kinematic fit reconstruction (without any truth-matched object, just as it would be applied on collision data). The shape of the distributions is naturally altered by the event selection, where only events fulfilling certain requirements survive. When the kinematic reconstruction is applied, small distortions became visible in comparison to the parton level forms after event selection, but the overall behaviour of the distributions does not change. Consequently, the information of the particles produced in

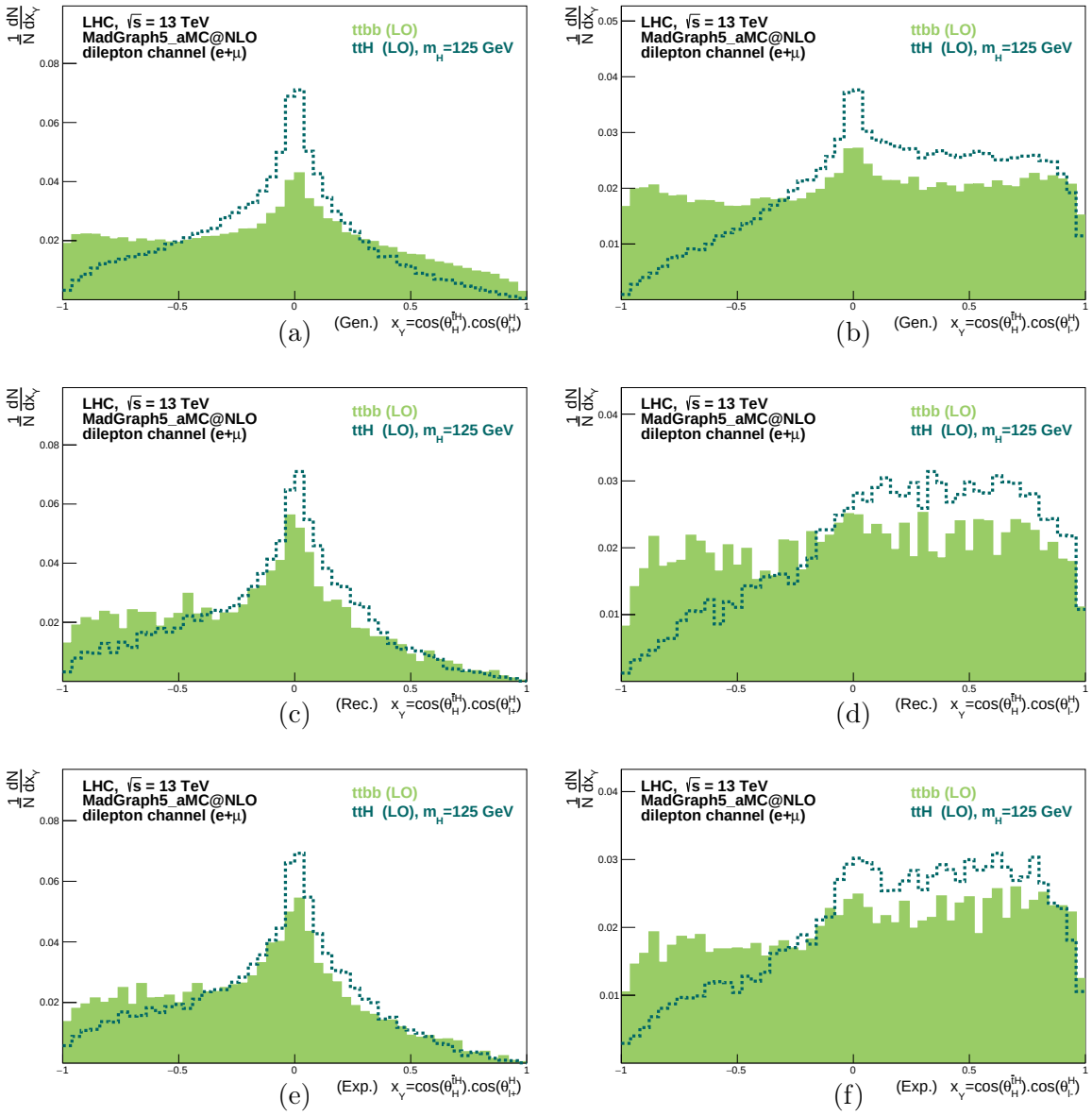


Figure C.1: Double angular product distributions: (a) and (b) at truth-level, after the selection cuts; (c) and (d) at truth level, after selection cuts and kinematical fit; (e) and (f) after the full analysis (selection cuts and kinematic reconstruction without any truth-match object). The distributions represent the product of the cosine of $\theta^{t\bar{t}H}_H$ and the cosine of $\theta^H_{\ell^+}$ (left) or $\theta^H_{\ell^-}$ (right) [3].

the hard-scatter, in particular the spin properties present in the matrix element, is preserved by the analysis throughout $t\bar{t}H$ dileptonic event selection and kinematic reconstruction.

Appendix D

$t\bar{t}h$ Event Reconstruction: jet assignment with TMVA

The full reconstruction of $t\bar{t}h$ events is achieved by attributing a total likelihood for being a dileptonic $t\bar{t}h\,h \to b\bar{b}$ event based on the kinematic information of the event, similarly to what was done in Section 8.2.1:

$$L_{t\bar{t}h} \quad \sim \quad P_{t\bar{t}}\,P(m_h)\,, \tag{D.1}$$

which is computed for every possible combination of neutrino solutions, jet-lepton pair and two-jet pair assignment. $P_{t\bar{t}}$ is the likelihood of a specific combination of objects in the event to have originated from a $t\bar{t}$ pair, whereas $P(m_h)$ is the probability of a certain pair of jets to have been the result of a Higgs boson decay. The top quark (antiquark) reconstruction is determined by correctly pairing the b ($\bar{b}$) quark to the lepton ℓ^+ (ℓ^-) and the neutrino ν_ℓ ($\bar{\nu}_\ell$). Following the previous neutrino reconstruction method, the kinematic information of the two escaping neutrinos is obtained by choosing the solution with higher $\big(P(p_{\mathrm{T}\nu})P(p_{\mathrm{T}\bar{\nu}})\big)/\big(p_{\mathrm{T}\nu}p_{\mathrm{T}\bar{\nu}}\big)$. In order to achieve better signal acceptance and faster runtime, the jet assignment is refined in the algorithm by including additional PDF information as well as using a MVA method to select the $b\ell$ and $b\bar{b}$ pairs most probable to have originated from the top quark/antiquark and the Higgs boson, respectively.

Instead of using PDF distributions of the mass of the $W^\pm$ bosons and the mass of top quark and antiquark, here, PDFs of the transverse momentum of the top quark, $P(p_{\mathrm{T}t})$, of the top antiquark, $P(p_{\mathrm{T}\bar{t}})$ and of the $t\bar{t}$ system, $P(p_{\mathrm{T}t\bar{t}})$, are used to determine a probability for each solution. Additionally, a probability is also given to each solution based on the two-dimensional PDF of the top quark masses, $P(m_t, m_{\bar{t}})$. The probability $P_{t\bar{t}}$ of each jet-lepton

combination per event and per neutrino solution to be compatible with a dileptonic $t\bar{t}$ decay is calculated as

$$P_{t\bar{t}} \sim \frac{P(p_{\mathrm{T}\nu})P(p_{\mathrm{T}\bar{\nu}})}{p_{\mathrm{T}\nu}p_{\mathrm{T}\bar{\nu}}} P(p_{\mathrm{T}t})P(p_{\mathrm{T}\bar{t}})P(p_{\mathrm{T}t\bar{t}})P(m_t, m_{\bar{t}}) \,, \tag{D.2}$$

assuming the p_{T} distributions of the neutrino, anti-neutrino, top quark, top antiquark, and $t\bar{t}$ system to be distributed at parton level as in the SM.

The Higgs boson reconstruction is simplified to selecting two jets amongst the remaining available ones. The probability of the $b\bar{b}$ pair to have been originated by the Higgs boson is given by the PDF of the SM Higgs boson mass, $P(m_h)$.

One of the main challenges of this reconstruction is, however, the number of possible jet combinations available to fully reconstruct the $t\bar{t}h$ system. Choosing a wrong combination of jets to kinematically reconstruct the signal events will give rise to, what is known as, combinatorial background. In an effort to reduce the number of possible combinations, only the six highest p_{T} jets are considered for the reconstruction, which include 95% of the time the jets produced from the hadronisation of b quarks from either the top quark pair or the Higgs boson decays, for both $t\bar{t}H$ and $t\bar{t}A$ signal events. Additionally, jet combinations are required to verify the following criteria:

$$m_{b_{t,\bar{t}}\ell^{\pm}} < 150\,\mathrm{GeV} \;\text{ and }\; 50\,\mathrm{GeV} \leq \mathrm{m}_{\mathrm{b_h}\bar{\mathrm{b}}_{\mathrm{h}}} \leq 200\,\mathrm{GeV} \,, \tag{D.3}$$

where $m_{b_t\ell^+}$ ($m_{b_{\bar{t}}\ell^-}$) is the invariant mass of the $b_t\ell^+$ ($b_{\bar{t}}\ell^-$) pair assigned in the reconstruction to the t ($\bar{t}$) quark decay and $m_{b_h\bar{b}_h}$ is the mass of the jet pair assigned to the Higgs boson. Finally, in order to increase the choice of correct combinations among those fulfilling the previous requirements, machine learning methods were tested with TMVA [348] to provide the correct jet assignment to the jet-lepton pairs and the Higgs boson.

For the training of the multivariate methods, nine parton level distributions were used as input: ΔR, lab-frame angles $\Delta\theta$ and $\Delta\Phi$ between the particle pairs (b_t, ℓ^+), $(\bar{b}_{\bar{t}}, \ell^-)$ and $(b_H, \bar{b}_H)$. To account for detector resolution effects, the invariant masses of these three particle pairs computed at reconstruction level with truth-matched events, were additionally enclosed in the TMVA inputs. The TMVA training and testing were done on an independent sample of $t\bar{t}H$ events (with a SM CP-even Higgs boson). The signal sample was created with the correct parton level jet combination. The correct jet combination includes the two jet-lepton pairs from the top quarks and the $b\bar{b}$ pair from the Higgs boson. The combinatorial background, however, was selected by computing the variables three times per event, each

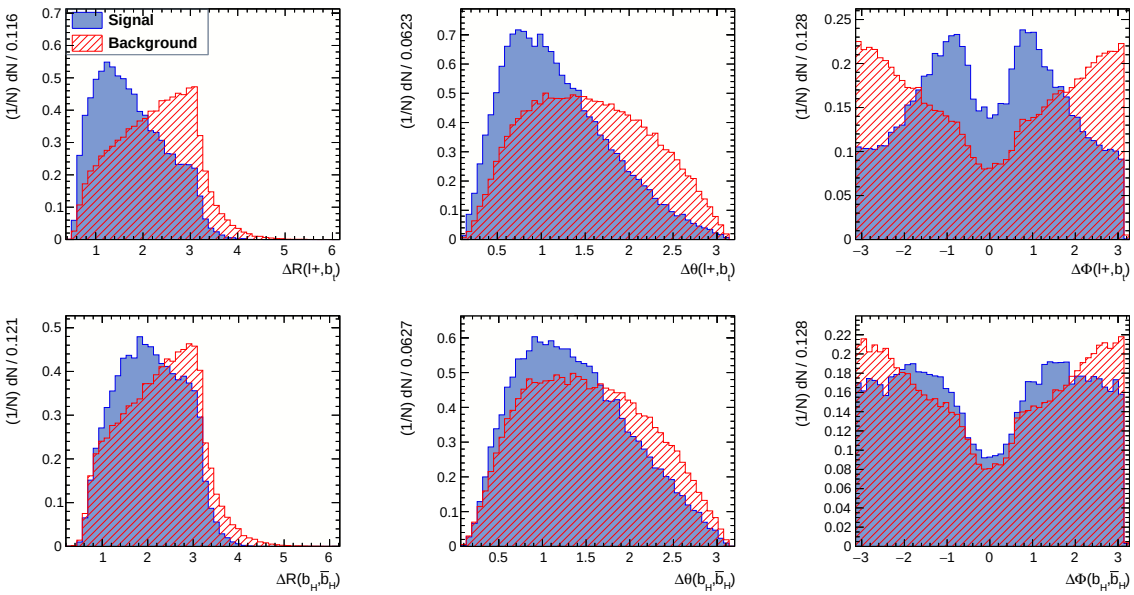


Figure D.1: Distributions of TMVA input variables for correct (filled blue, labelled 'Signal') and wrong combinations (red shaded, labelled 'Background') of jets and leptons from the same parent decaying particle: $\Delta R(\ell^+, b_t)$ (top left) and $\Delta R(b_H, \bar{b}_H)$ (bottom left); $\Delta\theta(\ell^+, b_t)$ (top middle) and $\Delta\theta(b_H, \bar{b}_H)$ (bottom middle); $\Delta\Phi(\ell^+, b_t)$ (top right) and $\Delta\Phi(b_H, \bar{b}_H)$ (bottom right).

one corresponding to a wrong permutation of the 4 b and $\bar{b}$ partons, and guaranteeing that all the variables computed in a permutation are different from the ones in any other, including the correct one. In Figures D.1 and D.2 (left), distributions of the input variables used on the TMVA trainning are shown for the signal and combinatorial background samples. The correlations between variables are shown in Figure D.2 (right), for the correct (top) and wrong combinations (bottom) samples.

A gradient boost decision tree (BDTG), which provided the best separation between signal (correct jet pairing) and combinatorial background (wrong jet pairing), is shown in Figure D.3. The distributions of the BDTG discriminant for the correct combinations (blue filled histogram) and for the wrong combinations (red dashed histogram) are shown, for both the training and the test samples. The jet combination with highest BDTG discrimination is then selected in the full kinematic reconstruction of events, improving the correct jet assignment rate. By using the TMVA to select the $b\ell$ pairs and assign the b and $\bar{b}$ originating from Higgs boson, the kinematic fit reconstruction has a faster runtime since it no longer needs to compute a total probability per each possible jet combination in the event. Instead,

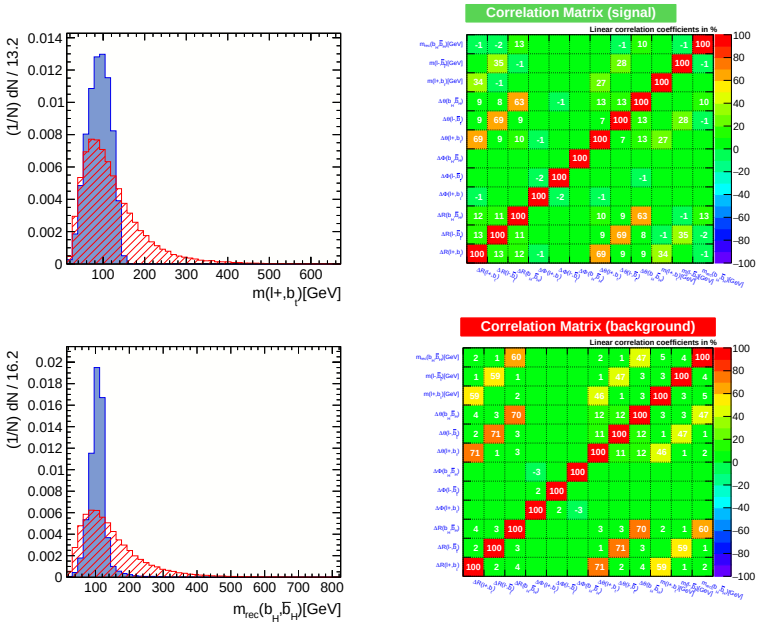


Figure D.2: Mass distributions (left) for right (filled blue, 'signal') and wrong (red shaded, 'background') combinations of jets and leptons from the same parent decaying particle: (upper-left) the $m(\ell^+, b_t)$ and (lower-left) $m(b_H, \bar{b}_H)$; (right-top) TMVA input variables correlations for 'signal' and (right-bottom) background'.

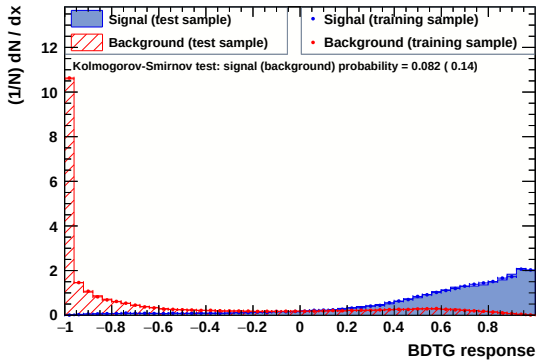


Figure D.3: BDTG response for signal and combinatorial background.

the kinematic fit will use the TMVA as decisive input in determining the jet assignment and proceed with the neutrino solution calculation, t, $\bar{t}$ and H reconstruction using the total likelihood for being a dileptonic $t\bar{t}h\,(h \to b\bar{b})$ event, $L_{t\bar{t}h}$, in Equation D.1. For $t\bar{t}H$ ($t\bar{t}A$) signal events, a 62% (61%) reconstruction rate is obtained, where in 31% (34%) of the signal

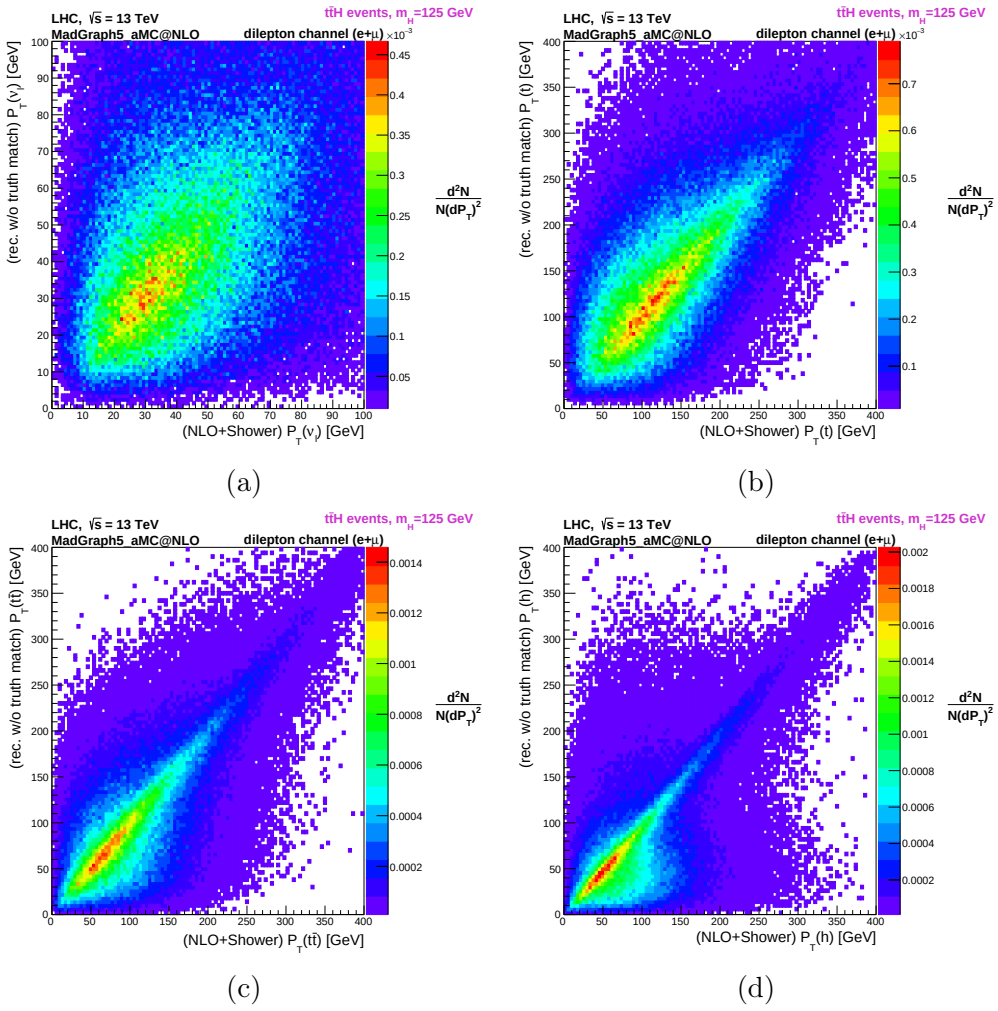


Figure D.4: Two-dimensional distributions of transverse momenta in $t\bar{t}H$ events: (a) for the neutrino ($p_{\mathrm{T}(\nu_{\mathrm{t}})}$); (b) for the top quark ($p_{\mathrm{T}(\mathrm{t})}$); (c) for the $t\bar{t}$ pair ($p_{\mathrm{T}(\mathrm{t}\bar{\mathrm{t}})}$); (d) for the Higgs boson ($p_{\mathrm{T}(\mathrm{h})}$). The horizontal axes represent truth level variables recorded at NLO+Shower, and the vertical axes represent the corresponding variables recorded after the reconstruction (without truth-match). Similar distributions were found for the antiparticles, $\bar{\nu}$ and $\bar{t}$ [4].

events the full kinematic reconstruction chooses the same jet combination as the parton level one. In section8.2.3, whenever the full kinematic reconstruction is said to be used, this is the method applied.

The correlation between the reconstructed objects, after the full $t\bar{t}h$ kinematic fit, and their parton level distributions can be seen in Figure D.4, where two dimensional transverse momentum distributions for (a) the neutrino, (b) the top quark, (c) the $t\bar{t}$ pair and (d)

the Higgs boson are presented. All distributions show a clear correlation between the reconstructed objects and their generated parton level. The neutrino p_{T} has a slightly larger spread of values, which is a result of the nature of the dileptonic events where two neutrinos are present, thus, requiring two neutrinos to be reconstructed per event.

Although the reconstruction could still be improved by resorting to alternative or more elaborate methods, it was deemed sufficient for the purposes of this study.

References

[1] G. Aad *et al.* **Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector**. *Eur. Phys. J.*, **C75**(7):349, 2015. 1, 46, 127, 134, 191, 199, 208, 210, 211, 234, 235, 236, 281

[2] S. Amor Santos *et al.* **Description of the Tile Calorimeter noise with increasing Pile-up for $\sqrt{s} = 7$ TeV data collected during 2011**. Technical report, CERN, Geneva, 2016. 2, 108, 109, 111, 112, 114, 115, 116, 117, 119, 123, 125, 281

[3] S. P. Amor dos Santos, *et al.* **Angular distributions in $t\bar{t}H(H \to b\bar{b})$ reconstructed events at the LHC**. *Phys. Rev. D*, **92**(3):034021, 2015. 2, 237, 242, 244, 282, 304

[4] S. Amor Dos Santos *et al.* **Probing the CP nature of the Higgs coupling in $t\bar{t}h$ events at the LHC**. *Phys. Rev. D*, **96**(1):013004, 2017. 2, 237, 251, 252, 253, 255, 256, 259, 262, 263, 264, 265, 266, 282, 309

[5] J. A. Aguilar-Saavedra and S. Amor dos Santos. **New directions for top quark polarization in the t-channel process**. *Physical Review D*, **89**(11), Jun 2014. 2, 238, 266, 269, 271, 272, 282

[6] S. L. Glashow. **Partial-symmetries of weak interactions**. *Nuclear Physics*, **22**(4):579 – 588, 1961. 3, 19

[7] S. Weinberg. **A Model of Leptons**. *Phys. Rev. Lett.*, **19**:1264–1266, Nov 1967. 3, 19

[8] N. Svartholm. *Elementary particle theory: relativistic groups and analyticity.* Nobel symposium. John Wiley & Sons, 1968. 3, 19

[9] P. W. Anderson. **Plasmons, Gauge Invariance, and Mass**. *Phys. Rev.*, **130**:439–442, Apr 1963. 3

[10] F. Englert and R. Brout. **Broken Symmetry and the Mass of Gauge Vector Mesons**. *Phys. Rev. Lett.*, **13**:321–323, Aug 1964. 19

[11] P. W. Higgs. **Broken Symmetries and the Masses of Gauge Bosons**. *Phys. Rev. Lett.*, **13**:508–509, Oct 1964. 19

[12] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. **Global Conservation Laws and Massless Particles**. *Phys. Rev. Lett.*, **13**:585–587, Nov 1964. 3, 19

[13] G. Zweig. **An SU(3) model for strong interaction symmetry and its breaking**. Technical Report CERN-TH-401, CERN, Geneva, Jan 1964. 3

[14] M. Y. Han and Y. Nambu. **Three-Triplet Model with Double SU(3) Symmetry**. *Phys. Rev.*, **139**:B1006–B1010, Aug 1965. 3

[15] O. W. Greenberg. **Spin and Unitary Spin Independence in a Paraquark Model of Baryons and Mesons**. *Phys. Rev. Lett.*, **13**:598–602, Oct 1964. 3

[16] H. David Politzer. **Reliable Perturbative Results for Strong Interactions?** *Phys. Rev. Lett.*, **30**:1346–1349, Jun 1973. 4

[17] H. David Politzer. **Asymptotic Freedom: An Approach to Strong Interactions**. *Phys.Rept.*, **14**:129–180, Jul 1974. 4

[18] D. J. Gross and F. Wilczek. **Ultraviolet Behavior of Non-Abelian Gauge Theories**. *Phys. Rev. Lett.*, **30**:1343–1346, Jun 1973. 4

[19] R. L. Workman *et al.* **Review of Particle Physics**. *PTEP*, **2022**:083C01, 2022. 4, 5, 6, 15, 17, 20, 22, 26, 27, 33, 34, 36, 42, 166

[20] J.J. Thomson. **Cathode rays**. *The Electrician*, **37**:104–11, 1897. 5

[21] J.J. Thomson. **Cathode rays**. *Phil. Mag. Ser. 5*, **44**:293–316, 1897. 5

[22] J.J. Thomson. **Les rayons**. *Roentgen Revue Sci.*, **6**:289–95, 1896. 5

[23] Y. Fukuda *et al.* **Evidence for Oscillation of Atmospheric Neutrinos**. *Physical Review Letters*, **81**(8):1562–1567, Aug 1998. 5, 23

REFERENCES

[24] Y. Abe *et al.* **Indication of Reactor $\overline{\nu}_e$ Disappearance in the Double Chooz Experiment**. *Phys. Rev. Lett.*, **108**:131801, Mar 2012. 5, 23

[25] G. Aad *et al.* **Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC**. *Physics Letters B*, **716**(1):1–29, 2012. 6, 20, 41

[26] S. Chatrchyan *et al.* **Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC**. *Physics Letters B*, **716**(1). 6, 20, 41

[27] G. 't Hooft and M.J.G. Veltman. **Regularization and Renormalization of Gauge Fields**. *Nucl. Phys. B*, **44**:189–213, 1972. 7, 19

[28] T. D. Lee and C. N. Yang. **Question of Parity Conservation in Weak Interactions**. *Phys. Rev.*, **104**:254–258, Oct 1956. 8

[29] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson. **Experimental Test of Parity Conservation in Beta Decay**. *Phys. Rev.*, **105**:1413–1415, Feb 1957. 8

[30] **Spontaneous Symmetry Breaking (Explanatory Diagram)**, June 2020. <http://commons.wikimedia.org/wiki/File:Spontaneous_symmetry_breaking_(explanatory _diagram).png>. 12

[31] **Mexican Hat Potential Polar**, June 2020. <http://commons.wikimedia.org/wiki/File:Mexican_hat _potential_polar.svg>. 13

[32] Y. Nambu. **Quasi-Particles and Gauge Invariance in the Theory of Superconductivity**. *Phys. Rev.*, **117**:648–663, Feb 1960. 12

[33] J. Goldstone. **Field theories with "Superconductor" solutions**. *Il Nuovo Cimento (1955-1965)*, **19**:154–164, 1961. 10.1007/BF02812722. 12

[34] N. Cabibbo. **Unitary Symmetry and Leptonic Decays**. *Phys. Rev. Lett.*, **10**:531–533, Jun 1963. 16

[35] M. Kobayashi and T. Maskawa. ***CP*-Violation in the Renormalizable Theory of Weak Interaction**. *Progress of Theoretical Physics*, **49**(2):652–657, 1973. 16, 18

[36] L.L. Chau and W.Y. Keung. **Comments on the Parametrization of the Kobayashi-Maskawa Matrix**. *Phys. Rev. Lett.*, **53**:1802–1805, Nov 1984. 17

[37] B. Pontecorvo. **Mesonium and anti-mesonium**. *Sov. Phys. JETP*, **6**:429, 1957. 17

[38] Z. Maki, M. Nakagawa, and S. Sakata. **Remarks on the Unified Model of Elementary Particles**. *Progress of Theoretical Physics*, **28**(5):870–880, 11 1962. 17

[39] B. Pontecorvo. **Neutrino Experiments and the Problem of Conservation of Leptonic Charge**. *Sov. Phys. JETP*, **26**:984–988, 1968. 17

[40] Z. Xing and D. Zhang. **Radiative decays of charged leptons as constraints of unitarity polygons for active-sterile neutrino mixing and CP violation**. *The European Physical Journal C*, **80**(12), Dec 2020. 17

[41] S. L. Glashow, J. Iliopoulos, and L. Maiani. **Weak Interactions with Lepton-Hadron Symmetry**. *Phys. Rev. D*, **2**:1285–1292, Oct 1970. 18

[42] J. J. Aubert *et al.* **Experimental Observation of a Heavy Particle *J***. *Phys. Rev. Lett.*, **33**:1404–1406, Dec 1974. 18

[43] J. E. *et al.* Augustin. **Discovery of a Narrow Resonance in e^+e^- Annihilation**. *Phys. Rev. Lett.*, **33**:1406–1408, 1974. 18

[44] M. L. Perl *et al.* **Evidence for anomalous lepton production in e^+e^- annihilation**. *Phys. Rev. Lett.*, **35**(22):1489–1492, Dec 1975. 18

[45] G. Charpak *et al.* **Measurement of the anomalous magnetic moment of the muon**. *Phys. Rev. Lett.*, **6**:128–132, 1961. 18

[46] M. Bernardini *et al.* **A proposal to search for leptonic quarks and heavy leptons produced by ADONE**. pages 202–226, 3 1967. 18

[47] Y.S. Tsai. **Decay Correlations of Heavy Leptons in $e^+ + e^- \to l^+ + l^-$**. *Phys. Rev. D*, **4**:2821–2837, Nov 1971. 18

[48] K. Kodama *et al.* **Observation of tau neutrino interactions**. *Physics Letters B*, **504**(3):218 – 224, 2001. 18

[49] S. W. Herb *et al.* **Observation of a Dimuon Resonance at 9.5-GeV in 400-GeV Proton-Nucleus Collisions**. *Phys. Rev. Lett.*, **39**:252–255, 1977. 19

[50] F.J. Hasert *et al.* **Observation of neutrino-like interactions without muon or electron in the gargamelle neutrino experiment**. *Phys. Lett. B*, **46**:138–140, 1973. 19, 49

[51] UA1 Collaboration. **Further evidence for charged intermediate vector bosons at the SPS collider**. *Phys. Lett. B*, **129**:273–282, Sep 1983. 19, 49

[52] UA1 Collaboration. **Experimental observation of lepton pairs of invariant mass around 95 GeV/c^2 at the CERN SPS collider**. *Phys. Lett. B*, **126**:398–410, Jul 1983. 19, 49

[53] S. van der Meer. **Stochastic damping of betatron oscillations in the ISR**. pages 261–269, 8 1972. 19

[54] CDF Collaboration. **Observation of top quark production in $\bar{p}p$ collisions**. *Phys. Rev. Lett.*, **74**:2626–2631, 1995. 19

[55] D0 Collaboration. **Search for high mass top quark production in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV**. *Phys. Rev. Lett.*, **74**:2422–2426, 1995. 19

[56] S. Chatrchyan *et al.* **Study of the Mass and Spin-Parity of the Higgs Boson Candidate via Its Decays to Z Boson Pairs**. *Physical Review Letters*, **110**(8), Feb 2013. 20

[57] G. Aad *et al.* **Evidence for the spin-0 nature of the Higgs boson using ATLAS data**. *Physics Letters B*, **726**(1-3):120–144, Oct 2013. 20

[58] J. Haller *et al.* **Update of the global electroweak fit and constraints on two-Higgs-doublet models**. *The European Physical Journal C*, **78**(8), Aug 2018. 20, 21

[59] L. N. Lipatov. **The parton model and perturbation theory**. *Yad. Fiz.*, **20**:181–198, 1974. 24

[60] V. N. Gribov and L. N. Lipatov. **Deep inelastic e p scattering in perturbation theory**. *Sov. J. Nucl. Phys.*, **15**:438–450, 1972.

[61] G. Altarelli and G. Parisi. **Asymptotic Freedom in Parton Language**. *Nucl. Phys. B*, **126**:298–318, 1977.

[62] Y. L. Dokshitzer. **Calculation of the Structure Functions for Deep Inelastic Scattering and e^+ e^- Annihilation by Perturbation Theory in Quantum Chromodynamics.** *Sov. Phys. JETP*, **46**:641–653, 1977. 24

[63] J. M. Campbell, J. W. Huston, and W. J. Stirling. **Hard Interactions of Quarks and Gluons: A Primer for LHC Physics**. *Rept. Prog. Phys.*, **70**:89, 2007. 25

[64] M. Beneke, P. Falgari, S. Klein, and C. Schwinn. **Hadronic top-quark pair production with NNLL threshold resummation**. *Nuclear Physics B*, **855**(3):695–741, Feb 2012. 28

[65] M. Cacciari, M. Czakon, M. Mangano, A. Mitov, and P. Nason. **Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation**. *Physics Letters B*, **710**(4):612–622, 2012. 134

[66] P. Bärnreuther and 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**:132001, 2012. 134

[67] M. Czakon and A. Mitov. **NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels**. *Journal of High Energy Physics*, **2012**(12):54, Dec 2012. 134

[68] M. Czakon and A. Mitov. **NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction**. *JHEP*, **01**:080, 2013. 134

[69] M. Czakon, P. Fiedler, and A. Mitov. **Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_S^4)$**. *Physical Review Letters*, **110**(25), Jun 2013. 134, 250

[70] M. Czakon and A. Mitov. **Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders**. *Comput. Phys. Commun.*, **185**:2930, 2014. 28, 29, 133, 250

[71] M. Botje *et al.* **The PDF4LHC Working Group Interim Recommendations**, 2011. 28, 36, 250

[72] A. D. Martin *et al.* **Parton distributions for the LHC**. *The European Physical Journal C*, **63**(2):189–285, Sep 2009. 28, 136

[73] A. D. Martin *et al.* **Uncertainties on $\alpha(S)$ in global PDF analyses and implications for predicted hadronic cross sections**. *Eur. Phys. J.*, **C64**:653–680, 2009. 28, 136, 250

[74] H.-L. Lai *et al.* **New parton distributions for collider physics**. *Phys. Rev. D*, **82**:074024, Oct 2010. 28, 130, 133

[75] J. Gao *et al.* **CT10 next-to-next-to-leading order global analysis of QCD**. *Phys. Rev. D*, **89**:033009, Feb 2014. 28, 130, 250

[76] R. D. Ball *et al.* **Parton distributions with LHC data**. *Nuclear Physics B*, **867**(2):244–289, 2013. 28, 240, 249, 250

[77] N. Kidonakis. **Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production**. *Phys. Rev. D*, **83**:091503, May 2011. 28, 136, 250

[78] N. Kidonakis. **Next-to-next-to-leading logarithm resummation for s-channel single top quark production**. *Phys. Rev. D*, **81**:054028, Mar 2010. 28, 136, 250

REFERENCES

[79] M. Aliev *et al.* **HATHOR – HAdronic Top and Heavy quarks crOss section calculatoR**. *Computer Physics Communications*, **182**(4):1034–1046, Apr 2011. 28

[80] P. Kant *et al.* **HATHOR for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions**. *Computer Physics Communications*, **191**:74–89, Jun 2015. 28

[81] N. Kidonakis. **Two-loop soft anomalous dimensions for single top quark associated production with a W^- or H^-**. *Phys. Rev. D*, **82**:054018, Sep 2010. 28, 136

[82] N. Kidonakis. **Top Quark Production**, 2013. 28

[83] **LHCTop WORKING GROUP: LHCTopWG Summary plots**, 2021. <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCTopWGSummaryPlots>. 29, 31

[84] J.A. Aguilar-Saavedra. **Single top quark production at LHC with anomalous Wtb couplings**. *Nuclear Physics B*, **804**(1-2):160–192, Nov 2008. 31, 269

[85] G. Aad *et al.* **Measurements of inclusive and differential cross-sections of combined $t\bar{t}\gamma$ and $tW\gamma$ production in the eμ channel at 13 TeV with the ATLAS detector**. *JHEP*, **09**:049, 2020. 31

[86] A. Tumasyan *et al.* **Measurement of the inclusive and differential $t\bar{t}\gamma$ cross sections in the single-lepton channel and EFT interpretation at $\sqrt{s}$ = 13 TeV**. (CMS-TOP-18-010, CERN-EP-2021-117), 7 2021.

[87] A. Tumasyan *et al.* **Measurement of the inclusive and differential $t\bar{t}\gamma$ cross sections in the dilepton channel and effective field theory interpretation in proton-proton collisions at $\sqrt{s}$ =13 TeV**. *JHEP*, **05**:091, 2022.

[88] A. M. Sirunyan *et al.* **Measurement of the cross section for top quark pair production in association with a W or Z boson in proton-proton collisions at $\sqrt{s}$ = 13 TeV**. *JHEP*, **08**:011, 2018.

[89] M. Aaboud *et al.* **Measurement of the $t\bar{t}Z$ and $t\bar{t}W$ cross sections in proton-proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector**. *Phys. Rev. D*, **99**(7):072009, 2019.

[90] A. M. Sirunyan *et al.* **Measurement of top quark pair production in association with a Z boson in proton-proton collisions at $\sqrt{s}$ = 13 TeV**. *JHEP*, **03**:056, 2020.

[91] G. Aad *et al.* **Measurements of the inclusive and differential production cross sections of a top-quark-antiquark pair in association with a Z boson at $\sqrt{s}$ = 13 TeV with the ATLAS detector**. *Eur. Phys. J. C*, **81**:737, 2021.

[92] **Measurement of the cross section of top quark-antiquark pair production in association with a W boson in proton-proton collisions at $\sqrt{s}$ = 13 TeV**. 8 2022.

[93] A.M. Sirunyan *et al.* **Observation of $t\bar{t}H$ Production**. *Physical Review Letters*, **120**(23), Jun 2018. 46, 282

[94] M. Aaboud *et al.* **Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector**. *Physics Letters B*, **784**:173–191, Sep 2018. 31, 46, 47, 282

[95] A. M. Sirunyan *et al.* **Observation of Single Top Quark Production in Association with a Z Boson in Proton-Proton Collisions at $\sqrt{s}$ =13 TeV**. *Phys. Rev. Lett.*, **122**(13):132003, 2019. 32

[96] G. Aad *et al.* **Observation of the associated production of a top quark and a Z boson in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector**. *Journal of High Energy Physics*, **2020**(7), Jul 2020. 32

[97] A. Tumasyan *et al.* **Inclusive and differential cross section measurements of single top quark production in association with a Z boson in proton-proton collisions at $\sqrt{s}$ = 13 TeV**. *JHEP*, **02**:107, 2022. 32

[98] **Observation of single-top-quark production in association with a photon at the ATLAS detector**. Technical report, CERN, Geneva, 2022. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2022-013. 32

[99] A. Sirunyan *et al.* **Search for the production of four top quarks in the single-lepton and opposite-sign dilepton final states in proton-proton collisions at $\sqrt{s}$ = 13 TeV**. *JHEP*, **11**:082, 2019. 32

[100] A. Sirunyan *et al.* **Search for production of four top quarks in final states with same-sign or multiple leptons in proton-proton collisions at $\sqrt{s}$ = 13 TeV**. *Eur. Phys. J. C*, **80**(2):75, 2020.

[101] G. Aad *et al.* **Evidence for $t\bar{t}t\bar{t}$ production in the multilepton final state in proton-proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector**. *Eur. Phys. J. C*, **80**(11):1085, 2020.

[102] **Evidence for the simultaneous production of four top quarks in proton-proton collisions at $\sqrt{s}$ = 13 TeV**. Technical report, CERN, Geneva, 2022.

[103] G. Aad *et al.* **Measurement of the $t\bar{t}t\bar{t}$ production cross section in pp collisions at $\sqrt{s}$=13 TeV with the ATLAS detector**. (CERN-EP-2021-075), 6 2021. 32

[104] D. Acosta, J. Adelman, T. Affolder, T. Akimoto, M. G. Albrow, D. Ambrose, S. Amerio, D. Amidei, A. Anastassov, and K. Anikeev *et al.* **Measurement of $B(t \to Wb)/B(t \to Wq)$ at the Collider Detector at Fermilab**. *Physical Review Letters*, **95**(10), Aug 2005. 32

[105] V. M. Abazov *et al.* **Precision Measurement of the Ratio $B(t \to Wb)/B(t \to Wq)$ and Extraction of V_{tb}**. *Physical Review Letters*, **107**(12). 32

[106] V. Khachatryan *et al.* **Measurement of the ratio $B(t \to Wb)/B(t \to Wq)$ in pp collisions at $\sqrt{s} = 8$ TeV**. *Physics Letters B*, **736**. 32

[107] J. Gao, C. S. Li, and H. X. Zhu. **Top-Quark Decay at Next-to-Next-to-Leading Order in QCD**. *Physical Review Letters*, **110**(4), Jan 2013. 33

[108] M. Aaboud *et al.* **Direct top-quark decay width measurement in the $t\bar{t}$ lepton+jets channel at $\sqrt{s} = 8$ TeV with the ATLAS experiment**. *The European Physical Journal C*, **78**(2). 33

[109] C. Herwig, T. Ježo, and B. Nachman. **Extracting the Top-Quark Width from Nonresonant Production**. *Physical Review Letters*, **122**(23), Jun 2019. 33

[110] T. Dado, T. Peiffer, A. Quadt, M. Racko, E. Shabalina, and S. Tokar. **Measurement of the top-quark decay width in top-quark pair events in the dilepton channel at $\sqrt{s} = 13$ TeV with the ATLAS detector**. Technical report, CERN, Geneva, Aug 2019. 33

[111] D. de Florian, C. Grojean, F. Maltoni, C. Mariotti, A. Nikitenko, M. Pieri, P. Savard, M. Schumacher, and R Tanaka (Eds.) *et al. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*. CERN Yellow Reports: Monographs. CERN, Geneva, Oct 2016. 36, 37, 40, 41

[112] M. Cepeda, S. Gori, P. Ilten, M. Kado, and F. Riva *et al.* **Higgs Physics at the HL-LHC and HE-LHC**, 2019. 36

[113] F. Dulat, A. Lazopoulos, and B. Mistlberger. **iHixs 2 - Inclusive Higgs cross sections**. *Computer Physics Communications*, **233**:243–260, Dec 2018. 36

[114] C. Anastasiou, C. Duhr, F. Dulat, E. Furlan, T. Gehrmann, F. Herzog, A. Lazopoulos, and B. Mistlberger. **High precision determination of the gluon fusion Higgs boson cross-section at the LHC**. *JHEP*, **05**:058, 2016. 36

[115] M. Cacciari, F. A. Dreyer, A. Karlberg, G. P. Salam, and G. Zanderighi. **Fully Differential Vector-Boson-Fusion Higgs Production at Next-to-Next-to-Leading Order**. *Physical Review Letters*, **115**(8), Aug 2015. 36

[116] F. A. Dreyer and A. Karlberg. **Vector-Boson Fusion Higgs Production at Three Loops in QCD**. *Physical Review Letters*, **117**(7), Aug 2016.

[117] M. Ciccolini, A. Denner, and S. Dittmaier. **Strong and Electroweak Corrections to the Production of Higgs +2jets via Weak Interactions at the Large Hadron Collider**. *Physical Review Letters*, **99**(16), Oct 2007.

[118] M. Ciccolini, A. Denner, and S. Dittmaier. **Electroweak and QCD corrections to Higgs production via vector-boson fusion at the CERN LHC**. *Physical Review D*, **77**(1), Jan 2008.

[119] A. Denner, S. Dittmaier, S. Kallweit, and A. Mück. **HAWK 2.0: A Monte Carlo program for Higgs production in vector-boson fusion and Higgs strahlung at hadron colliders**. *Computer Physics Communications*, **195**:161–171, Oct 2015. 36

[120] R. D. Ball, V. Bertone, S. Carrazza, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, and J. Rojo. **Parton distributions with QED corrections**. *Nuclear Physics B*, **877**(2):290–320, Dec 2013. 36

[121] A. Manohar, P. Nason, G. P. Salam, and G. Zanderighi. **How Bright is the Proton? A Precise Determination of the Photon Parton Distribution Function**. *Physical Review Letters*, **117**(24), Dec 2016. 36

[122] A. V. Manohar, P. Nason, G. P. Salam, and G. Zanderighi. **The photon content of the proton**. *Journal of High Energy Physics*, **2017**(12), Dec 2017. 36

[123] R. V. Harlander, J. Klappert, S. Liebler, and L. Simon. **vhnnlo-v2: new physics in Higgs Strahlung**. *Journal of High Energy Physics*, **2018**(5), May 2018. 36

[124] R. V. Harlander, A. Kulesza, V. Theeuwes, and T. Zirke. **Soft gluon resummation for gluon-induced Higgs Strahlung**. *Journal of High Energy Physics*, **2014**(11), Nov 2014.

[125] L. Altenkamp, S. Dittmaier, R. V. Harlander, H. Rzehak, and T. J. E. Zirke. **Gluon-induced Higgs-strahlung at next-to-leading order QCD**. *Journal of High Energy Physics*, **2013**(2), Feb 2013.

[126] A. Denner, S. Dittmaier, S. Kallweit, and A. Mück. **Electroweak corrections to Higgs-strahlung off W/Z bosons at the Tevatron and the LHC with Hawk**. *Journal of High Energy Physics*, **2012**(3), Mar 2012.

[127] O. Brein, A. Djouadi, and R. Harlander. **NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders**. *Physics Letters B*, **579**(1-2):149–156, Jan 2004.

[128] M. L. Ciccolini, S. Dittmaier, and M. Krämer. **Electroweak radiative corrections to associated WH and ZH production at hadron colliders**. *Physical Review D*, **68**(7), Oct 2003. 36

[129] W. Beenakker, S. Dittmaier, M. Krämer, B. Plümper, M. Spira, and P. M. Zerwas. **Higgs Radiation Off Top Quarks at the Tevatron and the LHC**. *Phys. Rev. Lett.*, **87**:201805, Oct 2001. 36, 132

REFERENCES

[130] W. Beenakker, S. Dittmaier, M. Krämer, B. Plümper, M. Spira, and P.M. Zerwas. **NLO QCD corrections to $t\bar{t}H$ production in hadron collisions**. *Nuclear Physics B*, **653**(1):151 – 203, 2003. 132

[131] L. Reina and S. Dawson. **Next-to-Leading Order Results for $t\bar{t}H$ Production at the Tevatron**. *Physical Review Letters*, **87**(20), Oct 2001.

[132] S. Dawson, L. H. Orr, L. Reina, and D. Wackeroth. **Associated Top Quark-Higgs Boson Production at the LHC**. *Physical Review D*, **67**(7), Apr 2003.

[133] S. Dawson, C. Jackson, L. H. Orr, L. Reina, and D. Wackeroth. **Associated Higgs boson production with top quarks at the CERN Large Hadron Collider: NLO QCD corrections**. *Phys. Rev. D*, **68**:034022, Aug 2003. 132

[134] Y. Zhang, W.-G. Ma, R.-Y. Zhang, C. Chen, and L. Guo. **QCD NLO and EW NLO corrections to $t\bar{t}H$ production with top quark decays at hadron collider**. *Physics Letters B*, **738**:1–5, Nov 2014. 36

[135] S. Frixione, V. Hirschi, D. Pagani, H. S. Shao, and M. Zaro. **Electroweak and QCD corrections to top-pair hadroproduction in association with heavy bosons**, 2015. 36

[136] S. Frixione, V. Hirschi, D. Pagani, H.-S. Shao, and M. Zaro. **Weak corrections to Higgs hadroproduction in association with a top-quark pair**. *Journal of High Energy Physics*, **2014**(9), Sep 2014. 36

[137] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. **The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations**. *Journal of High Energy Physics*, **2014**(7), Jul 2014. 36, 213, 240, 249, 250

[138] R. Frederix, S. Frixione, V. Hirschi, D. Pagani, H.-S. Shao, and M. Zaro. **The automation of next-to-leading order electroweak calculations**. *Journal of High Energy Physics*, **2018**(7), Jul 2018. 36

[139] S Heinemeyer, C Mariotti, G Passarino, and R Tanaka (Eds.) *et al. Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group*. CERN Yellow Reports: Monographs. CERN, Geneva, Jul 2013. 37, 213

[140] B. Di Micco, M. Gouzevitch, J. Mazzitelli, and C. Vernieri. **Higgs boson potential at colliders: Status and perspectives**. *Reviews in Physics*, **5**:100045, Nov 2020. 39

[141] ATLAS Collaboration. **Combined measurement of the total and differential cross sections in the $H \to \gamma\gamma$ and the $H \to ZZ^* \to 4\ell$ decay channels at $\sqrt{s}$ =13 TeV with the ATLAS detector**. Technical report, CERN, Geneva, ATLAS-CONF-2019-032, Jul 2019. 42

[142] F. Bezrukov and M. Shaposhnikov. **Why should we care about the top quark Yukawa coupling?** *Journal of Experimental and Theoretical Physics*, **120**(3):335–343, Mar 2015. 43

[143] V. Khachatryan *et al.* **Search for the associated production of the Higgs boson with a top-quark pair**. *Journal of High Energy Physics*, **2014**(9), Sep 2014. 46

[144] G. Aad *et al.* **Search for the Standard Model Higgs boson decaying into $b\bar{b}$ produced in association with top quarks decaying hadronically in pp collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector**. *Journal of High Energy Physics*, **2016**(5), May 2016. 46

[145] V. Khachatryan *et al.* **Search for a Standard Model Higgs Boson Produced in Association with a Top-Quark Pair and Decaying to Bottom Quarks Using a Matrix Element Method**. *Eur. Phys. J. C*, **75**(6):251, 2015. 46

[146] G. Aad *et al.* **Search for $H \to \gamma\gamma$ produced in association with top quarks and constraints on the Yukawa coupling between the top quark and the Higgs boson using data taken at 7 TeV and 8 TeV with the ATLAS detector**. *Physics Letters B*, **740**:222–242, 2015. 46

[147] G. Aad *et al.* **Search for the associated production of the Higgs boson with a top quark pair in multilepton final states with the ATLAS detector**. *Physics Letters B*, **749**:519–541, 2015. 46

[148] G. Aad *et al.* **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**. *Journal of High Energy Physics*, **2016**(8), Aug 2016. 46

[149] M. Aaboud *et al.* **Search for the standard model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s}$ = 13TeV with the ATLAS detector**. *Phys. Rev. D*, **97**:072016, Apr 2018.

[150] M. Aaboud *et al.* **Evidence for the associated production of the Higgs boson and a top quark pair with the ATLAS detector**. *Phys. Rev. D*, **97**:072003, Apr 2018. 46

[151] A. M. Sirunyan *et al.* **Measurements of properties of the Higgs boson decaying into the four-lepton final state in pp collisions at $\sqrt{s}$ = 13 TeV**. *Journal of High Energy Physics*, **2017**(11), Nov 2017.

[152] A.M. Sirunyan *et al.* **Search for $t\bar{t}$H production in the all-jet final state in proton-proton collisions at $\sqrt{s}$ = 13 TeV**. *Journal of High Energy Physics*, **2018**(6), Jun 2018.

[153] A.M. Sirunyan *et al.* **Evidence for associated production of a Higgs boson with a top quark pair in final states with electrons, muons, and hadronically decaying τ leptons at $\sqrt{s}$ = 13 TeV**. *Journal of High Energy Physics*, **2018**(8), Aug 2018.

[154] A.M. Sirunyan *et al.* **Measurements of Higgs boson properties in the diphoton decay channel in proton-proton collisions at $\sqrt{s} = 13$ TeV**. *Journal of High Energy Physics*, **2018**(11), Nov 2018.

[155] A.M. Sirunyan *et al.* **Search for $t\bar{t}H$ production in the $H \to b\bar{b}$ decay channel with leptonic $t\bar{t}$ decays in proton-proton collisions at $\sqrt{s} = 13$ TeV**. *Journal of High Energy Physics*, **2019**(3), Mar 2019. 46

[156] ATLAS Collaboration. **Analysis of $t\bar{t}H$ and $t\bar{t}W$ production in multilepton final states with the ATLAS detector**. ATLAS-CONF-2019-045, 10 2019. 47

[157] ATLAS Collaboration. **Measurement of the Higgs boson decaying to b-quarks produced in association with a top-quark pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector**. ATLAS-CONF-2020-058, 11 2020.

[158] G. Aad *et al.* **CP Properties of Higgs Boson Interactions with Top Quarks in the $t\bar{t}H$ and tH Processes Using $H \to \gamma\gamma$ with the ATLAS Detector**. *Phys. Rev. Lett.*, **125**(6):061802, 2020. 238

[159] A. M. Sirunyan *et al.* **Combined measurements of Higgs boson couplings in proton–proton collisions at $\sqrt{s} = 13$ TeV**. *Eur. Phys. J. C*, **79**(5):421, 2019.

[160] A. M. Sirunyan *et al.* **Measurements of $t\bar{t}H$ Production and the CP Structure of the Yukawa Interaction between the Higgs Boson and Top Quark in the Diphoton Decay Channel**. *Phys. Rev. Lett.*, **125**:061801, Aug 2020. 238

[161] A. M. Sirunyan *et al.* **Measurement of the Higgs boson production rate in association with top quarks in final states with electrons, muons, and hadronically decaying tau leptons at $\sqrt{s} = 13$ TeV**. *Eur. Phys. J. C*, **81**(4):378, 2021. 47

[162] ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, SLD Collaboration, LEP Electroweak Working Group, SLD Electroweak, and Heavy Flavour Groups. **Precision electroweak measurements on the Z resonance**. *Physics Reports*, **427**(5-6):257 – 454, 2006. 49

[163] V. Fanti *et al.* **A New measurement of direct CP violation in two pion decays of the neutral kaon**. *Phys. Lett. B*, **465**:335–348, 1999. 49

[164] G. Baur *et al.* **Production of antihydrogen**. *Phys. Lett. B*, **368**:251–258, 1996. 49

[165] G.B. Andresen *et al.* **Trapped antihydrogen**. *Nature*, **468**:673–676, Dec 2010. 49

[166] G.B. Andresen *et al.* **Confinement of antihydrogen for 1,000 seconds**. *Nature Physics*, **7**:558–564, Jul 2011. 49

[167] Lyndon Evans and Philip Bryant. **LHC Machine**. *Journal of Instrumentation*, **3**:S08001, 2008. 50

[168] K. Aamodt *et al.* **The ALICE experiment at the CERN LHC**. *JINST*, **3**:S08002, 2008. 50

[169] G. Aad *et al.* **The ATLAS Experiment at the CERN Large Hadron Collider**. *JINST*, **3**:S08003, 2008. 50, 54, 56, 57, 58, 61, 64, 66, 70

[170] S. Chatrchyan *et al.* **The CMS experiment at the CERN LHC**. *JINST*, **3**:S08004, 2008. 50

[171] A. Augusto Alves Jr. *et al.* **The LHCb Detector at the LHC**. *JINST*, **3**:S08005, 2008. 50

[172] G. Aad *et al.* **Luminosity Determination in pp Collisions at $\sqrt{s} = 7$ TeV using the ATLAS Detector at the LHC**. *Eur. Phys. J.*, **C71**:1630, 2011. 52, 80, 91

[173] S. van der Meer. **Calibration of the Effective Beam Height in the ISR**. 1968. 53

[174] *ATLAS magnet system: Technical Design Report, 1*. Technical Design Report ATLAS. CERN, Geneva, 1997. 57

[175] A. Yamamoto *et al.* **The ATLAS central solenoid**. *Nucl. Instrum. Meth.*, **A584**:53–74, 2008. 58

[176] ATLAS Collaboration. **ATLAS inner detector: Technical design report. Vol. 1**. Technical Report CERN-LHCC-97-16, ATLAS-TDR-4, CERN, Geneva, 1997. 59

[177] ATLAS Collaboration. **ATLAS inner detector: Technical design report. Vol. 2**. Technical Report CERN-LHCC-97-17, CERN, Geneva, 1997. 59

[178] ATLAS Collaboration. **The ATLAS Inner Detector commissioning and calibration**. *Eur. Phys. J. C*, **70**:787–821, 2010. 59

[179] ATLAS Collaboration. **ATLAS calorimeter performance Technical Design Report**. Technical Report CERN-LHCC-96-40, CERN, Geneva, 1996. 63

[180] ATLAS Collaboration. **ATLAS liquid argon calorimeter: Technical design report**. Technical Report CERN-LHCC-96-41, CERN, Geneva, 1996. 64

[181] G. *et al.* Aad. **Readiness of the ATLAS Liquid Argon Calorimeter for LHC Collisions**. *Eur. Phys. J.*, **C70**:723–753, 2010. 64, 65

[182] ATLAS Collaboration. **ATLAS tile calorimeter: Technical design report**. Technical Report CERN-LHCC-96-42, CERN, Geneva, 1996. 68, 83, 90

REFERENCES

[183] ATLAS Collaboration. **Readiness of the ATLAS Tile Calorimeter for LHC collisions**. *Eur. Phys. J. C*, **70**:1193–1236, Dec 2010. 68, 84, 86, 93

[184] ATLAS Collaboration. **ATLAS muon spectrometer: Technical design report**. Technical Report CERN-LHCC-97-22, ATLAS-TDR-10, CERN, Geneva, 1997. 70

[185] ATLAS Collaboration. **Commissioning of the ATLAS Muon Spectrometer with Cosmic Rays**. *Eur. Phys. J. C*, **70**:875–916, 2010. 70

[186] D. G. Drakoulakos, E. Gschwendtner, J. M. Maugain, F. Rohrbach, and Y. Sedykh. **The high-precision x-ray tomograph for quality control of the ATLAS MDT muon spectrometer**. (CERN-OPEN-97-023), 7 1997. 72

[187] F. Veloso. *Study of ATLAS sensitivity to FCNC top quark decays*. PhD thesis, LIP, Coimbra, 2008. 75

[188] C. Clement and P. Klimek. **Identification of Pile-up Using the Quality Factor of Pulse Shapes in the ATLAS Tile Calorimeter**. Technical Report ATL-TILECAL-PROC-2011-014, CERN, Geneva, Nov 2011. 74, 92

[189] ATLAS Collaboration. **ATLAS high-level trigger, data acquisition and controls: Technical design report**. CERN-LHCC-2003-022. 74

[190] V. Cindro, D. Dobos, I. Dolenc, H. Frais-Kölbl, A. GorišieFullanak, E. Griesmayer, H. Kagan, G. Kramberger, B. Macek, I. Mandić, M. Mikuž, M. Niegl, H. Pernegger, D. Tardif, W. Trischuk, P. Weilhammer, and M. Zavrtanik. **The ATLAS Beam Conditions Monitor**. *Journal of Instrumentation*, **3**(02):P02004, 2008. 78

[191] G. Aad *et al.* **Improved luminosity determination in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector at the LHC**. *Eur. Phys. J.*, **C73**(8):2518, 2013. 78, 205

[192] C.A. Solans and A. Valero. **Tile front-end diagram**. Technical Report ATL-COM-TILECAL-2012-048, CERN, Geneva, Oct 2012. 84

[193] A. Valero, C. Solans, C. Bohm, and M. Oreglia. **Drawings for TileCal electronics Upgrade**. Dec 2012. 84

[194] E. Fullana *et al.* **Digital signal reconstruction in the ATLAS hadronic tile calorimeter**. *IEEE Trans. Nucl. Sci.*, **53**:2139–2143, 2006. 86

[195] A. Valero. **The ATLAS TileCal Read-Out Drivers Signal Reconstruction**. Technical Report ATL-TILECAL-PROC-2009-004, CERN, Geneva, Nov 2009. 87

[196] ATLAS Collaboration. **Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics**. (SLAC-R-980, CERN-OPEN-2008-020), 2009. e-Print: 0901.0512 [hep-ex]. 88, 147

[197] F. Spano. **Simulation of the full noise pattern in tile calorimeter front end electronics: a phenomenological approach to coherent effects**. 10 2008. 89, 93, 96

[198] M. Aaboud et al. **Operation and performance of the ATLAS Tile Calorimeter in Run 1**. *Eur. Phys. J. C*, **78**(12):987, 2018. 93

[199] A. Artamonov *et al.* **Description of the Tile Calorimeter Electronic Noise**. Technical report, CERN, Geneva, 2011. 93

[200] O. Lundberg. **Studies of the TileCal electronic noise in 2011 and early 2012 data**. Technical report, CERN, Geneva, 2013. 93

[201] T. Sjöstrand, S. Mrenna, and P. Skands. **PYTHIA 6.4 physics and manual**. *Journal of High Energy Physics*, **2006**(05):026, 2006. 95, 107, 133, 240, 250

[202] J. Novakova. **Monte Carlo Simulations of the Tile Calorimeter and Measurement of the Z to tau tau Cross Section with the ATLAS Detector**, 2012. Presented 05 Sep 2012. 95

[203] M C N Fiolhais. **Correlated noise unfolding on a Hadronic Calorimeter**. Technical report, CERN, Geneva, 2011. 98

[204] S. Yamamoto. **Coherent Noise Unfolding and Effect on Clusters and MET**. https://indico.cern.ch/event/85005/contributions/2101146/attachments/1076555/1535509/TileWeek_ShimpeiYamamoto_20100218.pdf accessed in August 2017. 98, 99

[205] S Santos. **Study of the Correlated Noise in TileCal**. Technical report, CERN, Geneva, 2011. 100

[206] L. Fiorini and B.T. Martin. **Validation of TileCal double Gaussian noise description**. https://indico.cern.ch/event/66869/contributions/2063944/attachments/1019634/1451274/jet-et-31-08-2009.pdf accessed in August 2017. 105

[207] K. Koenneke. **Pileup Reweighting**. https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/PileupReweighting accessed in August 2017. 107, 108

[208] **Sherpa and Open Science Grid: Predicting the emergence of jets**, June 2018. «https://sciencenode.org/feature/sherpa-and-open-science-grid-predicting-emergence-jets.php». 129

[209] **The Durham HepData Project**, May 2021. <http://hepdata.cedar.ac.uk/pdf/pdf3.html>. 130

[210] G. Aad *et al.* **The ATLAS Simulation Infrastructure**. *Eur. Phys. J.*, **C70**:823–874, 2010. 131

[211] S. Agostinelli *et al.* **Geant4 - a simulation toolkit**. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **506**(3):250 – 303, 2003. 131

[212] M. Beckingham *et al.* **The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim**. Technical Report ATL-PHYS-PUB-2010-013, CERN, Geneva, Oct 2010. 132

[213] P. Golonka and Z. Was. **PHOTOS Monte Carlo: a precision tool for QED correctionsin Z and W decays**. *The European Physical Journal C - Particles and Fields*, **45**(1):97–107, Jan 2006. 132

[214] S. Jadach, J. H. Kühn, and Z. Wąs. **TAUOLA - a library of Monte Carlo programs to simulate decays of polarized τ leptons**. *Computer Physics Communications*, **64**(2):275 – 299, 1991. 132

[215] T. Sj$^{'}$ostrand, S. Mrenna, and P.Z. Skands. **A Brief Introduction to PYTHIA 8.1**. *Comput. Phys. Commun.*, **178**:852–867, 2008. 132, 136

[216] ATLAS Collaboration. **ATLAS tunes of PYTHIA 6 and Pythia 8 for MC11**. Technical Report ATL-PHYS-PUB-2011-009, CERN, Geneva, Jul 2011. 132

[217] G. Bevilacqua, M. Czakon, M. V. Garzelli, A. van Hameren, A. Kardos, C. G. Papadopoulos, R. Pittau, and M. Worek. **HELAC-NLO**. *Comput. Phys. Commun.*, **184**:986–997, 2013. 132

[218] P. Nason. **A New method for combining NLO QCD with shower Monte Carlo algorithms**. *JHEP*, **11**:040, 2004. 132

[219] S. Frixione, P. Nason, and C. Oleari. **Matching NLO QCD computations with Parton Shower simulations: the POWHEG method**. *JHEP*, **11**:070, 2007. 132

[220] S. Alioli, P. Nason, C. Oleari, and E. Re. **A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX**. *JHEP*, **06**:043, 2010. 132

[221] Pavel M. Nadolsky, Hung-Liang Lai, Qing-Hong Cao, Joey Huston, Jon Pumplin, Daniel Stump, Wu-Ki Tung, and C.-P. Yuan. **Implications of CTEQ global analysis for collider observables**. *Phys. Rev. D*, **78**:013004, Jul 2008. 132, 133, 135, 269

[222] ATLAS Collaboration. **Summary of ATLAS Pythia 8 tunes**. Technical Report ATL-PHYS-PUB-2012-003, CERN, Geneva, Aug 2012. 132

[223] A. Djouadi, J. Kalinowski, and M. Spira. **HDECAY: a program for Higgs boson decays in the Standard Model and its supersymmetric extension**. *Computer Physics Communications*, **108**(1):56 – 74, 1998. 133

[224] A. Bredenstein, A. Denner, S. Dittmaier, and M. M. Weber. **Precise predictions for the Higgs-boson decay $H \to WW/ZZ \to 4$ leptons**. *Phys. Rev. D*, **74**:013004, Jul 2006. 133

[225] LHC Higgs Cross Section Working Group, A. Denner, S. Heinemeyer, I. Puljak, D. Rebuzzi, and M. Spira. **Standard model Higgs-boson branching ratios with uncertainties**. *The European Physical Journal C*, **71**(9):1753, Sep 2011. 133

[226] Stefano Actis, Giampiero Passarino, Christian Sturm, and Sandro Uccirati. **NNLO computational techniques: The cases $H \to \gamma\gamma$ and $H \to gg$**. *Nuclear Physics B*, **811**(1):182 – 273, 2009. 133

[227] S. Dittmaier *et al.* **Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables**. 2011. 133

[228] P.Z. Skands. **Tuning Monte Carlo generators: The Perugia tunes**. *Phys. Rev. D*, **82**:074018, Oct 2010. 133

[229] M. L. Mangano, F. Piccinini, A. D. Polosa, M. Moretti, and R. Pittau. **ALPGEN, a generator for hard multiparton processes in hadronic collisions**. *Journal of High Energy Physics*, **2003**(07):001, 2003. 135

[230] K. Melnikov and F. Petriello. **Electroweak gauge boson production at hadron colliders through $\mathcal{O}(\alpha_s^2)$**. *Phys. Rev.*, **D74**:114017, 2006. 135

[231] 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**. *Journal of High Energy Physics*, **2013**(7):32, Jul 2013. 135, 165

[232] G. Corcella, I. G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, M. H. Seymour, and B. R. Webber. **HERWIG 6: an event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)**. *Journal of High Energy Physics*, **2001**(01):010, 2001. 136

[233] J. M. Butterworth, J. R. Forshaw, and M. H. Seymour. **Multiparton interactions in photoproduction at HERA**. *Zeitschrift für Physik C: Particles and Fields*, **72**(4):637, Dec 1996. 136

[234] J. M. Campbell and R. K. Ellis. **Update on vector boson pair production at hadron colliders**. *Phys. Rev. D*, **60**:113006, Nov 1999. 136, 212

[235] S. Frixione, E. Laenen, P. Motylinski, C. White, and B. R. Webber. **Single-top hadroproduction in association with a W boson**. *Journal of High Energy Physics*, **2008**(07):029, 2008. 136, 212

REFERENCES

[236] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer. **MadGraph 5 : Going Beyond**. *JHEP*, **06**:128, 2011. 136, 208

[237] ATLAS Collaboration. **New ATLAS event generator tunes to 2010 data**. Technical Report ATL-PHYS-PUB-2011-008, CERN, Geneva, Apr 2011. 136

[238] J.M. Campbell and R.K. Ellis. **$t\bar{t}W^{\pm}$ production and decay at NLO**. *JHEP*, **07**:052, 2012. 136

[239] M. V. Garzelli, A. Kardos, C. G. Papadopoulos, and Z. Trocsanyi. **$t\bar{t}W^{\pm}$ and $t\bar{t}Z$ Hadroproduction at NLO accuracy in QCD with Parton Shower and Hadronization effects**. *JHEP*, **11**:056, 2012. 136, 213

[240] ATLAS Collaboration. **Egamma Trigger Public Results**, January 2018. <https:/twiki.cern.ch/twiki/bin/view/AtlasPublic/EgammaTriggerPublicResults>. 137, 138

[241] ATLAS Collaboration. **Muon Trigger Public Results**, January 2018. <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/MuonTriggerPublicResults>. 137, 138

[242] ATLAS Collaboration. **Performance of primary vertex reconstruction in proton-proton collisions at $\sqrt{s}$ =7 TeV in the ATLAS experiment**. Technical Report ATLAS-CONF-2010-069, CERN, Geneva, Jul 2010. 137

[243] ATLAS Collaboration. **Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton–proton collision data**. *The European Physical Journal C*, **74**(7):2941, Jul 2014. 139

[244] W. Lampl, S. Laplace, D. Lelas, P. Loch, H. Ma, S. Menke, S. Rajagopalan, D. Rousseau, S. Snyder, and G. Unal. **Calorimeter Clustering Algorithms: Description and Performance**. Technical Report ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, CERN, Geneva, Apr 2008. 139, 147

[245] G. Aad *et al.* **Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data**. *Eur. Phys. J.*, **C74**(10):3071, 2014. 140, 143

[246] ATLAS Collaboration. **Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton-proton collision data**. Technical Report ATLAS-CONF-2014-032, CERN, Geneva, Jun 2014. 141, 142

[247] B. Acharya *et al.* **Object selection and calibration, background estimations and MC samples for top quark analyses using the full 2012 data set**. Technical Report ATL-COM-PHYS-2013-1016, CERN, Geneva, Jul 2013. 143, 145, 147, 153

[248] G. Aad *et al.* **Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton-proton collision data**. *Eur. Phys. J.*, **C74**(11):3130, 2014. 144, 146

[249] M. Goblirsch, M. Vanadia, and O. Kortner. **Muon reconstruction efficiency in 2012 pp collisions: Reconstruction efficiency measured with the tag and probe method using $Z \to \mu\mu$ decays in 2012 8 TeV data**. Technical Report ATL-COM-MUON-2013-015, CERN, Geneva, Jul 2013. 145

[250] ATLAS Collaboration. **Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s}$ = 7 TeV**. *The European Physical Journal C*, **73**(3):2304, Mar 2013. 147

[251] M. Cacciari, G. P. Salam, and G. Soyez. **The anti-k_t jet clustering algorithm**. *JHEP*, **04**:063, 2008. 147

[252] G. P. Salam. **Towards Jetography**. *Eur. Phys. J.*, **C67**:637–686, 2010. 147

[253] C. Cojocaru *et al.* **Hadronic calibration of the ATLAS liquid argon end-cap calorimeter in the pseudorapidity region $1.6 < |\eta| < 1.8$ in beam tests**. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **531**(3):481 – 514, 2004. 148

[254] T. Barillari *et al.* **Local Hadronic Calibration**. Technical Report ATL-LARG-PUB-2009-001-2. ATL-COM-LARG-2008-006. ATL-LARG-PUB-2009-001, CERN, Geneva, Jun 2008. 148

[255] Atlas Collaboration. **Jet energy measurement and its systematic uncertainty in proton–proton collisions at $\sqrt{s}$ = 7TeV with the ATLAS detector**. *The European Physical Journal C*, **75**(1):17, Jan 2015. 149, 150

[256] ATLAS Collaboration. **Pile-up subtraction and suppression for jets in ATLAS**. Technical Report ATLAS-CONF-2013-083, CERN, Geneva, Aug 2013. 149, 152, 153

[257] ATLAS Collaboration. **Single hadron response measurement and calorimeter jet energy scale uncertainty with the ATLAS detector at the LHC**. *The European Physical Journal C*, **73**(3):2305, Mar 2013. 151

[258] G. Aad *et al.* **Light-quark and gluon jet discrimination in pp collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector**. *Eur. Phys. J.*, **C74**(8):3023, 2014. 151

[259] ATLAS Collaboration. **Jet energy measurement and systematic uncertainties using tracks for jets and for b-quark jets produced in proton-proton collisions at $\sqrt{s}$ = 7 TeV in the ATLAS detector**. Technical Report ATLAS-CONF-2013-002, CERN, Geneva, Jan 2013. 151

[260] The ATLAS Collaboration. **Jet energy resolution in proton-proton collisions at $\sqrt{s}$ = 7 TeV recorded in 2010 with the ATLAS detector**. *The European Physical Journal C*, **73**(3):2306, Mar 2013. 151

[261] B. Butler and A. Schwartzman. **Track-Jet Reconstruction and Performance**. Technical Report ATL-PHYS-INT-2010-040, CERN, Geneva, Apr 2010. 152

[262] ATLAS Collaboration. **Jet energy resolution and selection efficiency relative to track jets from in-situ techniques with the ATLAS Detector Using Proton-Proton Collisions at a Center of Mass Energy $sqrts = 7$ TeV**. Technical Report ATLAS-CONF-2010-054, CERN, Geneva, Jul 2010. 152

[263] ATLAS Collaboration. **b-tagging in dense environments**. Technical Report ATL-PHYS-PUB-2014-014, CERN, Geneva, Aug 2014. 154

[264] ATLAS Collaboration. **Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data**. Technical Report ATLAS-CONF-2011-102, CERN, Geneva, Jul 2011. 154, 155

[265] G. Piacquadio and C. Weiser. **A new inclusive secondary vertex algorithm for b-jet tagging in ATLAS**. *Journal of Physics: Conference Series*, **119**(3):032032, 2008. 155

[266] 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**. 2012. 155, 156

[267] ATLAS Collaboration. **Calibration of b-tagging using dileptonic top pair events in a combinatorial likelihood approach with the ATLAS experiment**. Technical Report ATLAS-CONF-2014-004, CERN, Geneva, Feb 2014. 156, 206

[268] ATLAS Collaboration. **Calibration of the performance of b-tagging for c and light-flavour jets in the 2012 ATLAS data**. Technical Report ATLAS-CONF-2014-046, CERN, Geneva, Jul 2014. 157

[269] G. Aad *et al.* **Performance of Missing Transverse Momentum Reconstruction in Proton-Proton Collisions at 7 TeV with ATLAS**. *Eur. Phys. J.*, **C72**:1844, 2012. 157

[270] ATLAS collaboration. **Performance of Missing Transverse Momentum Reconstruction in ATLAS studied in Proton-Proton Collisions recorded in 2012 at 8 TeV**. (ATLAS-CONF-2013-082), 2013. 158

[271] V. M. Abazov *et al.* **Measurement of the $t\bar{t}$ production cross section in $p\overline{p}$ collisions at $\sqrt{s} = 1.96$ TeV using secondary vertex b tagging**. *Phys. Rev. D*, **74**:112004, Dec 2006. 203

[272] G. Aad and *et al.* **Measurements of normalized differential cross sections for $t\bar{t}$ production in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector**. *Phys. Rev. D*, **90**(7):072004, 2014. 207

[273] T. Gleisberg and *et al.* **Event generation with SHERPA 1.1**. *JHEP*, **02**:007, 2009. 208

[274] F. Cascioli *et al.* **Scattering Amplitudes with Open Loops**. *Phys. Rev. Lett.*, **108**:111601, 2012. 208

[275] C. Buttar *et al.* **Les houches physics at TeV colliders 2005, standard model and Higgs working group: Summary report**. In *4th Les Houches Workshop on Physics at TeV Colliders*, 4 2006. 212

[276] Peter Z. Skands and Daniel Wicke. **Non-perturbative QCD effects and the top mass at the Tevatron**. *Eur. Phys. J. C*, **52**:133–140, 2007. 212

[277] John M. Campbell and R. Keith Ellis. **$t\bar{t}W^{+-}$ production and decay at NLO**. *JHEP*, **07**:052, 2012. 213

[278] M. Bahr and *et al.* **Herwig++ Physics and Manual**. *Eur. Phys. J. C*, **58**:639–707, 2008. 213

[279] J. Bellm and *et al.* **Herwig++ 2.7 Release Note**. 10 2013. 213

[280] A. Hoecker *et al.* **TMVA - Toolkit for Multivariate Data Analysis**, 2007. 214

[281] Rene Brun and Fons Rademakers. **ROOT — An object oriented data analysis framework**. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **389**(1):81–86, 1997. New Computing Techniques in Physics Research V. 214

[282] M. Feindt and U. Kerzel. **The NeuroBayes neural network package**. *Nucl. Instrum. Meth. A*, **559**:190–194, 2006. 216

[283] M. Leshno, V. Ya. Lin, A. Pinkus, and S. Schocken. **Multilayer feedforward networks with a nonpolynomial activation function can approximate any function**. *Neural Networks*, **6**(6):861–867, 1993. 216

[284] Vernon D. Barger, J. Ohnemus, and R. J. N. Phillips. **Event shape criteria for single lepton top signals**. *Phys. Rev. D*, **48**:R3953–R3956, 1993. 218

[285] C. Bernaciak *et al.* **Fox-Wolfram Moments in Higgs Physics**. *Phys. Rev. D*, **87**:073014, 2013. 218

[286] Thomas Junk. **Confidence level computation for combining searches with small statistics**. *Nucl. Instrum. Meth. A*, **434**:435–443, 1999. 222, 223

[287] Alexander L. Read. **Presentation of search results: The CL(s) technique**. *J. Phys. G*, **28**:2693–2704, 2002. 222

[288] G. Cowan *et al.* **Asymptotic formulae for likelihood-based tests of new physics**. *Eur. Phys. J. C*, **71**:1554, 2011. [Erratum: Eur.Phys.J.C 73, 2501 (2013)]. 223

REFERENCES

[289] G. Aad *et al.* **Light-quark and gluon jet discrimination in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector**. *Eur. Phys. J. C*, **74**(8):3023, 2014. 230

[290] Wouter Verkerke and David Kirkby. **The RooFit toolkit for data modeling**, 2003. 234

[291] W. Verkerke and D. Kirkby. **RooFit Users Manual**. 234

[292] V. Khachatryan *et al.* **Search for a Standard Model Higgs Boson Produced in Association with a Top-Quark Pair and Decaying to Bottom Quarks Using a Matrix Element Method**. *Eur. Phys. J. C*, **75**(6):251, 2015. 235

[293] Duarte Fontes, Jorge C. Romão, Rui Santos, and João P. Silva. **Large pseudoscalar Yukawa couplings in the complex 2HDM**. *JHEP*, **06**:060, 2015. 238

[294] Pierre Artoisenet, Rikkert Frederix, Olivier Mattelaer, and Robbert Rietkerk. **Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations**. *JHEP*, **03**:015, 2013. 240, 245, 250, 257

[295] J. de Favereau *et al.* **DELPHES 3: a modular framework for fast simulation of a generic collider experiment**. *Journal of High Energy Physics*, **2014**(2), Feb 2014. 240, 250

[296] J. Ellis *et al.* **Disentangling Higgs-top couplings in associated production**. *Journal of High Energy Physics*, **2014**(4), Apr 2014. 245

[297] S. Biswas *et al.* **Enhancing the $t\bar{t}H$ signal through top-quark spin polarization effects at the LHC**. *Journal of High Energy Physics*, **2014**(7), Jul 2014.

[298] F. Demartin *et al.* **Higgs characterisation at NLO in QCD: CP properties of the top-quark Yukawa interaction**. *The European Physical Journal C*, **74**(9), Sep 2014.

[299] Fawzi Boudjema, Rohini M. Godbole, Diego Guadagnoli, and Kirtimaan A. Mohan. **Lab-frame observables for probing the top-Higgs interaction**. *Phys. Rev. D*, **92**(1):015019, 2015. 257

[300] P. Artoisenet et al. **A framework for Higgs characterisation**. *JHEP*, **11**:043, 2013. 249

[301] Johan Alwall et al. **Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions**. *Eur. Phys. J. C*, **53**:473–500, 2008. 250

[302] Stefano Frixione and Bryan R. Webber. **Matching NLO QCD computations and parton shower simulations**. *JHEP*, **06**:029, 2002. 250

[303] John F. Gunion and Xiao-Gang He. **Determining the CP nature of a neutral Higgs boson at the LHC**. *Phys. Rev. Lett.*, **76**:4468–4471, 1996. 257, 261

[304] John Ellis, Dae Sung Hwang, Kazuki Sakurai, and Michihisa Takeuchi. **Disentangling Higgs-Top Couplings in Associated Production**. *JHEP*, **04**:004, 2014. 257

[305] S. Biswas, R. Frederix, E. Gabrielli, and B. Mele. **Enhancing the $t\bar{t}H$ signal through top-quark spin polarization effects at the LHC**. *JHEP*, **07**:020, 2014. 257

[306] Federico Demartin, Fabio Maltoni, Kentarou Mawatari, Ben Page, and Marco Zaro. **Higgs characterisation at NLO in QCD: CP properties of the top-quark Yukawa interaction**. *Eur. Phys. J. C*, **74**(9):3065, 2014. 257

[307] D. Espriu and J. Manzano. **Measuring effective electroweak couplings in single top quark production at the CERN LHC**. *Physical Review D*, **65**(7), Mar 2002. 266

[308] J.A. Aguilar-Saavedra. **A minimal set of top anomalous couplings**. *Nuclear Physics B*, **812**(1-2):181–204, may 2009. 266

[309] W. Buchmuller and D. Wyler. **Effective Lagrangian Analysis of New Interactions and Flavor Conservation**. *Nucl. Phys. B*, **268**:621–653, 1986. 267

[310] D. Barducci et al. **Interpreting top-quark LHC measurements in the standard-model effective field theory**. 2 2018. 267

[311] Q.H. Cao, B. Yan, J.H. Yu, and C. Zhang. **A general analysis of Wtb anomalous couplings**. *Chinese Physics C*, **41**(6):063101, jun 2017. 267

[312] Georges Aad et al. **Combination of the W boson polarization measurements in top quark decays using ATLAS and CMS data at $\sqrt{s} = 8$ TeV**. *JHEP*, **08**(08):051, 2020. 268

[313] A. Georges *et al.* **Search for anomalous couplings in the Wtb vertex from the measurement of double differential angular decay rates of single top quarks produced in the t-channel with the ATLAS detector**. *JHEP*, **04**:023, 2016. 268

[314] V. Cirigliano, W. Dekens, J. de Vries, and E. Mereghetti. **Constraining the top-Higgs sector of the Standard Model Effective Field Theory**. *Phys. Rev. D*, **94**(3):034031, 2016. 268

[315] Morad Aaboud et al. **Probing the W tb vertex structure in t-channel single-top-quark production and decay in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector**. *JHEP*, **04**:124, 2017. 268

[316] Frédéric Déliot, Ricardo Faria, Miguel C. N. Fiolhais, Pedro Lagarelhos, António Onofre, Christopher M. Pease, and Ana Vasconcelos. **Global constraints on top quark anomalous couplings**. *Phys. Rev. D*, **97**(1):013007, 2018. 268

[317] ATLAS Collaboration. **Analysis of the Wtb vertex from the measurement of triple-differential angular decay rates of single top quarks produced in the t-channel at $\sqrt{s} = 8$ TeV with the ATLAS detector**. *JHEP*, **12**:017, 2017. 268, 276

[318] G. Mahlon and S. Parke. **Improved spin basis for angular correlation studies in single top quark production at the Fermilab Tevatron**. *Physical Review D*, **55**(11):7249–7254, jun 1997. 270

[319] G. Mahlon and S. Parke. **Single top quark production at the LHC: Understanding spin**. *Physics Letters B*, **476**(3-4):323–330, mar 2000. 270

[320] R. Schwienhorst, C.P. Yuan, C. Mueller, and Q.H. Cao. **Single top quark production and decay in the t-channel at next-to-leading order at the LHC**. *Physical Review D*, **83**(3), feb 2011. 270

[321] J. M. Campbell, R. Frederix, F. Maltoni, and F. Tramontano. **Next-to-Leading-Order Predictions for t-Channel Single-Top Production at Hadron Colliders**. *Physical Review Letters*, **102**(18), may 2009. 270

[322] J. M. Campbell, R. Frederix, F. Maltoni, and F. Tramontano. **NLO predictions for t-channel production of single top and fourth generation quarks at hadron colliders**. *Journal of High Energy Physics*, **2009**(10):042–042, oct 2009. 270

[323] Georges Aad et al. **Measurement of the W boson polarization in top quark decays with the ATLAS detector**. *JHEP*, **06**:088, 2012. 270

[324] G. L. Kane, G. A. Ladinsky, and C. P. Yuan. **Using the top quark for testing standard-model polarization and CP predictions**. *Phys. Rev. D*, **45**:124–141, Jan 1992.

[325] **Measurement of the W-boson helicity in top decays from ttbar production in lepton+jets events at the LHC at sqrt(s)=8 TeV**. 2013.

[326] **Search for $\mathcal{CP}$ violation in single top quark events in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector**. 3 2013. 275

[327] Bohdan Grzadkowski and Mikolaj Misiak. **Anomalous Wtb coupling effects in the weak radiative B-meson decay**. *Phys. Rev. D*, **78**:077501, 2008. [Erratum: Phys.Rev.D 84, 059903 (2011)].

[328] Jure Drobnak, Svjetlana Fajfer, and Jernej F. Kamenik. **Interplay of $t \to bW$ Decay and B_q Meson Mixing in Minimal Flavor Violating Models**. *Phys. Lett. B*, **701**:234–239, 2011. 270

[329] N. Castro A. Onofre J.A. Aguilar-Saavedra, J. Carvalho and F. Veloso. **Probing anomalous Wtb couplings in top pair decays**. *The European Physical Journal C*, **50**(3):519–533, apr 2007. 274

[330] P. González W. Bernreuther and C. Mellein. **Decays of polarized top quarks to lepton, neutrino, and jets at NLO QCD**. *The European Physical Journal C*, **74**(3), mar 2014. 274

[331] Saurabh D. Rindani and Pankaj Sharma. **Probing anomalous tbW couplings in single-top production using top polarization at the Large Hadron Collider**. *JHEP*, **11**:082, 2011. 274

[332] J. A. Aguilar-Saavedra *et al.* **The fully differential top decay distribution**. *Eur. Phys. J. C*, **77**(3):200, 2017. 275

[333] Qing-Hong Cao, Jose Wudka, and C.-P. Yuan. **Search for new physics via single-top production at the LHC**. *Physics Letters B*, **658**(1-3):50–56, dec 2007. 275

[334] Cen Zhang and Scott Willenbrock. **Effective-field-theory approach to top-quark production and decay**. *Physical Review D*, **83**(3), feb 2011. 275

[335] J.A. Aguilar-Saavedra. **Effective four-fermion operators in top physics: A roadmap**. *Nuclear Physics B*, **843**(3):638–672, 2011. 275

[336] J.A. Aguilar-Saavedra and J. Bernabéu. **W polarisation beyond helicity fractions in top quark decays**. *Nuclear Physics B*, **840**(1-2):349–378, nov 2010. 275

[337] **Measurement of top quark polarization in t-channel single-top production**. Technical report, CERN, Geneva, 2013. 276

[338] ATLAS Collaboration. **Probing the Wtb vertex structure in t-channel single-top-quark production and decay in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector**. *JHEP*, **04**:124, 2017. 276

[339] CMS Collaboration. **Measurement of differential cross sections and charge ratios for t-channel single top quark production in proton–proton collisions at $\sqrt{s} = 13$ TeV**. *Eur. Phys. J. C*, **80**:370, 2020. 276

[340] Georges Aad et al. **Measurement of the polarisation of single top quarks and antiquarks produced in the t-channel at $\sqrt{s} = 13$ TeV and bounds on the tWb dipole operator from the ATLAS experiment**. 2 2022. 276, 277, 283

[341] D. Azevedo *et al.* **CP tests of Higgs couplings in $t\bar{t}h$ semileptonic events at the LHC**. *Phys. Rev. D*, **98**:033004, Aug 2018. 283

[342] A. Ferroglia *et al.* **Role of the $t\bar{t}h$ rest frame in direct top-quark Yukawa coupling measurements**. *Phys. Rev. D*, **100**:075034, Oct 2019.

[343] D. Azevedo *et al.* **Scalar mass dependence of angular variables in $t\bar{t}\phi$ production**. *JHEP*, **06**:155, 2020.

[344] D. Azevedo *et al.* **Light Higgs searches in $t\bar{t}\phi$ production at the LHC**. *JHEP*, **04**:077, 2021.

[345] D. Azevedo *et al.* **CP-violation, asymmetries and interferences in $t\bar{t}\phi$**. *JHEP*, **09**:246, 2022. 283

[346] **Probing the CP nature of the top-Higgs Yukawa coupling in $t\bar{t}H$ and tH events with $H \to b\bar{b}$ using the ATLAS detector at the LHC**. Technical report, CERN, Geneva, 2022. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2022-016. 283

[347] **Search for $t\bar{t}H/A \to t\bar{t}t\bar{t}$ production in the multilepton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector**. Technical report, CERN, Geneva, 2022. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2022-008. 283

[348] Andreas Hocker et al. **TMVA - Toolkit for Multivariate Data Analysis**. 3 2007. 306