

USING VH ASSOCIATED PRODUCTION TO SEARCH
FOR THE $b\bar{b}$ DECAY OF THE HIGGS BOSON WITH
DATA FROM THE ATLAS DETECTOR AT $\sqrt{s} = 13\text{ TeV}$

Approved by:

Dr. Stephen Sekula
Associate Professor

Dr. Pavel Nadolsky
Associate Professor

Dr. Robert Kehoe
Professor

Dr. Giacinto Piacquadio
Assistant Professor



USING VH ASSOCIATED PRODUCTION TO SEARCH
FOR THE $b\bar{b}$ DECAY OF THE HIGGS BOSON WITH
DATA FROM THE ATLAS DETECTOR AT $\sqrt{s} = 13\text{ TeV}$

A Dissertation Presented to the Graduate Faculty of the
Dedman College

Southern Methodist University

in

Partial Fulfillment of the Requirements

for the degree of

Doctor of Philosophy

with a

Major in Physics

by

Jeffrey Hetherly

B.S., Physics, University of North Texas
Master of Science in Materials Science, University of North Texas
Master of Science in Physics, University of North Texas

May 20, 2017

Copyright (2017)

Jeffrey Hetherly

All Rights Reserved

ACKNOWLEDGMENTS

I would first like to thank my advisor, Stephen Sekula. His enthusiasm for particle physics and data analysis convinced me to pursue experimental rather than theoretical high-energy physics. I will forever look fondly on our time together and I couldn't have asked for a more knowledgeable, patient, and encouraging mentor.

Besides my advisor, I would like to thank the rest of my thesis committee: Pavel Nadolsky, Bob Kehoe, and Giacinto Piacquadio, for their encouragement and insights.

I want to also thank the entire SMU ATLAS group for their invaluable guidance and feedback throughout the years. They continually pushed me to be a better and better physicist. I am profoundly grateful for the numerous colleagues and friends I made during my time at CERN.

Finally, I could not have accomplished so much in life without the love and support of my grandparents, my parents, my brother, and my wife.

Hetherly, Jeffrey

B.S., Physics, University of North Texas
Master of Science in Materials Science, University of North Texas
Master of Science in Physics, University of North Texas

Using VH Associated Production to Search
for the $b\bar{b}$ Decay of the Higgs Boson with
Data from the ATLAS Detector at $\sqrt{s} = 13$ TeV

Advisor: Dr. Stephen Sekula

Doctor of Philosophy degree conferred May 20, 2017

Dissertation completed March 1, 2017

Two analyses, using the vector-boson-associated production of the Higgs boson in which the Higgs then decays to a pair of b -quarks using data collected with the ATLAS detector during 2015 and 2016 at a center-of-mass energy of $\sqrt{s} = 13$ TeV, are presented. The first analysis focuses on the Standard Model interpretation with the expressed goal of observing the $H \rightarrow b\bar{b}$ decay. The second analysis searches for an additional Higgs boson in the context of the Two-Higgs Doublet Model extension of the Standard Model that resolves some of the outstanding issues the Standard Model has such as baryogenesis. Both analyses aim to take advantage of the large predicted branching ratio of the Higgs boson to b -quarks. The primary backgrounds in both analyses are the production of a Standard Model vector boson with two additional jets and, also, t -quark pair-production.

The Standard Model analysis leverages a multivariate algorithm for better signal-to-background separation while the non-Standard Model analysis is primarily cut-based. Both use a global likelihood fit to assess the presence of signal within the constraints imposed by systematic uncertainties. A prominent source of uncertainty in both analyses is the uncertainties arising from the identification of b -quarks. A portion of my analytic and technical contribution is the estimation and ultimate reduction of these uncertainties.

TABLE OF CONTENTS

LIST OF FIGURES	xii
LIST OF TABLES	xvii
LIST OF SYMBOLS	xix
Chapter	Page
1. Introduction	1
2. Standard Model and 2HDM Higgs Theoretical Framework	3
2.1. Basic Principles of Quantum Field Theory	4
2.1.1. Scalar Fields	5
2.1.2. Spinor Fields	6
2.1.3. Vector Fields	6
2.1.4. Scattering	7
2.2. The Standard Model of Elementary Particle Physics	11
2.2.1. Electroweak Theory	13
2.2.2. Spontaneous Electroweak Symmetry Breaking and The Higgs Mechanism	15
2.2.3. Quantum Chromodynamics	17
2.2.4. Higgs Boson Production Mechanisms	21
2.2.5. Higgs Boson Decay Channels	23
2.2.6. Standard Model Executive Summary	25
2.3. The Two-Higgs Doublet Model Extension to the SM	26
2.4. Previous Higgs Measurements	30
2.5. Particle Colliders	35
3. The LHC and the ATLAS Detector	36
3.1. LHC	36

3.2.	ATLAS	39
3.2.1.	Magnets	42
3.2.1.1.	Barrel Solenoid	42
3.2.1.2.	Barrel & End-Cap Toroids	43
3.2.2.	Inner Detector	44
3.2.2.1.	Pixel	45
3.2.2.2.	SCT	45
3.2.2.3.	TRT	45
3.2.3.	Electromagnetic Calorimeter	47
3.2.3.1.	Barrel	47
3.2.3.2.	EMEC	48
3.2.4.	Hadronic Calorimeter	48
3.2.4.1.	Tile	49
3.2.4.2.	HEC	49
3.2.4.3.	FCal	49
3.2.5.	Muon Systems	50
3.2.5.1.	MDT	51
3.2.5.2.	CSC	52
3.2.5.3.	RPC	52
3.2.5.4.	TGC	53
3.2.6.	Trigger Systems	53
3.2.6.1.	L1	54
3.2.6.2.	HLT	55
4.	Simulation and Reconstruction	56
4.1.	Simulation Overview	57
4.1.1.	Generators	60

4.1.2.	Modeling Detector Responses and Digitization	62
4.2.	Object Reconstruction.....	63
4.2.1.	Tracks	63
4.2.2.	Calorimeter Clusters	67
4.2.2.1.	Sliding-Window Clustering	67
4.2.2.2.	Topological Clustering	68
4.2.3.	Electrons	69
4.2.4.	Muons	72
4.2.5.	Hadronic Jets	77
4.2.5.1.	Prelude	77
4.2.5.2.	Jet Objects	79
4.2.6.	Hadronic τ -leptons	83
4.2.7.	Flavor-Tagged Jets	84
4.2.7.1.	Truth Flavor Labels.....	85
4.2.7.2.	b -Tagging	85
4.2.8.	Overlap Removal.....	89
4.2.9.	Missing Transverse Energy and Momentum	90
4.2.10.	L1 Trigger	91
4.2.10.1.	Calorimeter.....	92
4.2.10.2.	Muon.....	92
4.2.11.	HLT Trigger	93
4.2.11.1.	Tracking.....	94
4.2.12.	Trigger Menu Items	94
4.3.	Summary	94
5.	Simulated & Detector Datasets	96
5.1.	MC Simulated Datasets	97

5.1.1.	$t\bar{t}$	97
5.1.2.	Single Top	98
5.1.3.	V +Jets	98
5.1.4.	Di-Boson	99
5.1.5.	QCD Processes	100
5.1.6.	VH	100
5.1.7.	ggA & $b\bar{b}A$	100
5.2.	ATLAS 2015 & 2016 Datasets	101
5.3.	Discussion	102
6.	Object and Event Selections	104
6.1.	Derivation-Level Cuts	106
6.2.	Object Selection	106
6.2.1.	Additional $A \rightarrow Zh$ Object Selection	111
6.3.	Event Selection	111
6.3.1.	SM 0-Lepton Selection	113
6.3.2.	SM 1-Lepton Selection	118
6.3.3.	SM 2-Lepton Selection	120
6.3.4.	$A \rightarrow Zh$ 0-Lepton Selection	120
6.3.5.	$A \rightarrow Zh$ 2-Lepton Selection	123
7.	Standard Model $VH(H \rightarrow b\bar{b})$	125
7.1.	b -Jet Energy Corrections	126
7.2.	2-Lepton Channel Kinematic Fit	127
7.3.	1-Lepton Multi-jet Background	129
7.4.	Truth-Tagging	131
7.5.	Boosted Decision Tree	133

7.6.	Fit Model & Statistical Analysis	136
7.6.1.	Systematic Uncertainties	138
7.6.1.1.	Experimental Uncertainties.....	139
7.6.1.2.	Simulated Background Modeling Uncertainties	140
7.6.1.3.	Multi-Jet Background Uncertainties	140
7.6.1.4.	Simulated Signal Modeling Uncertainties	142
7.6.2.	Statistical Analysis.....	144
7.7.	Results	149
7.7.1.	Higgs Boson Search	149
7.7.2.	Di-boson Validation	159
8.	2HDM $A \rightarrow Zh(h \rightarrow b\bar{b})$	160
8.1.	b -Jet Energy Corrections	161
8.2.	Analysis	161
8.2.1.	bbA	162
8.2.2.	CR Analysis	163
8.2.3.	Systematics	163
8.3.	Results	165
9.	Conclusions	176
9.1.	Summary	176
9.1.1.	SM VH	176
9.1.2.	2HDM $A \rightarrow Zh(h \rightarrow b\bar{b})$	177
9.2.	Discussion & Outlook	177
Appendix		Page
A.	b -Jet Trigger	180
A.1.	Software Architecture Overview	180

A.2. Overview of HLT and the b -Jet Trigger Slice	182
A.2.1. Run 1 and 2 HLT Overview	182
A.2.2. b -Jet Trigger Overview	183
A.3. Split-Chain Performance Studies	183
A.4. Migration of Offline Flavor-Tagging Algorithms	186
B. Flavor Tagging Calibrations	188
B.1. Local Polynomial Kernel Regression	189
B.1.1. Addition of Full Error Matrix	194
B.2. Reduction of Nuisance Parameters	194
B.2.1. Novel Summation	196
B.3. Generalization to PCA	198
C. Kernel Density Estimation with Systematic Variations	199
C.1. Motivation for Density Estimation	199
C.2. Description of Projection-KDE	202
C.3. Estimating Systematic Variations	209
C.4. Limit Curves After KDE Estimate	213
BIBLIOGRAPHY	215

LIST OF FIGURES

Figure	Page
2.3 Compton scattering Feynman diagrams	7
2.4 Particles of the Standard Model of Elementary Particle Physics	12
2.5 Higgs potential.....	16
2.6 Running of QCD coupling parameter	20
2.7 ggF and VBF Higgs production mechanisms	21
2.8 VH Higgs production mechanisms	22
2.9 ttH/bbH Higgs production mechanisms	22
2.10 Higgs fermionic and weak boson decay channels.....	23
2.11 Higgs photonic decay channels.....	24
2.12 Higgs gluon decay channel.....	24
2.13 Leading Higgs decay channels	25
2.14 2HDM A boson production mechanisms	29
2.15 $gg \rightarrow A \rightarrow Zh, h \rightarrow b\bar{b}$ amplitude ratio for Type II 2HDM ggF	30
2.16 LHC Run 1 Higgs couplings results	32
2.17 ATLAS 2HDM LHC Run 1 results	33
2.18 CMS 2HDM LHC Run 1 results	34
3.1 LHC overview.....	37
3.2 LHC schedule	38
3.3 Cutaway view of the ATLAS detector	40
3.4 Charged-particle multiplicities as a function of η	42
3.5 Solenoid magnetic field	43

3.6	Barrel toroid magnet	44
3.7	End-cap toroid	44
3.8	Cut-away view of the ATLAS inner detector	45
3.9	Cut-away view of the ATLAS Pixel detector	46
3.10	Cut-away view of the ATLAS calorimeter system	48
3.11	EM calorimeter accordion geometry	49
3.12	Tile calorimeter assembly	49
3.13	Schematic of the EM calorimeter crack region	50
3.14	positions of the FCal components	51
3.15	Cut-away view of the ATLAS muon system	52
3.16	Cross-section of an MDT tube	53
3.17	Cross-section of RPCs	53
3.18	Schematic of trigger terminology	54
4.1	GAUDI data flow	58
4.2	ATLAS simulation overview	59
4.3	Diagram of hadronization and showering in event generation	61
4.4	ATLAS Run 1 constrained and unconstrained track reconstruction	64
4.5	ATLAS Run 1 un/constrained track reconstruction distributions	65
4.6	ATLAS Run 1 track reconstruction TRT extension	66
4.7	ATLAS Run 2 track impact parameter performance	67
4.8	Comparison of sliding window versus topological clustering multiplicity	69
4.9	ATLAS Run 1 GSF electron corrections	70
4.10	ATLAS electron energy scale & resolution corrections	72
4.11	Electron identification efficiencies	73
4.12	Uncertainties regarding the muon scale factors	75
4.13	Muon identification efficiencies versus pseudorapidity	76

4.14	Performance of anti-kt algorithm	78
4.15	Performance of the JVT	80
4.16	Jet-energy calibration scales Run 1	82
4.17	Jet-energy calibration scales Run 2	82
4.18	JER smearing	84
4.19	b-jet decay topology	87
4.20	b-jet decay and transverse impact parameter	88
4.21	MV2c10 BDT output	88
4.22	MET calibration performance	91
4.23	L1Calo clustering algorithm	93
4.24	ATLAS Run 1 offline versus EF tracking	95
5.1	Pileup comparisons in data and MC	97
6.1	SM VH Feynman diagrams for the three lepton channels	105
6.2	BSM AZh Feynman diagrams for the two lepton channels	105
6.3	Signal regions for the SM VH analysis	112
6.4	Signal and control regions for the BSM $A \rightarrow Zh$ analysis	112
6.5	Two multi-jet variables before multi-jet suppression in the 0-lepton SM analysis	115
6.6	Two additional multi-jet variables before multi-jet suppression in the 0-lepton SM analysis	115
6.7	Two multi-jet variables after all multi-jet suppression in the 0-lepton SM analysis	116
6.8	Two additional multi-jet variables after all multi-jet suppression in the 0-lepton SM analysis	116
6.9	Missing energy trigger efficiencies as a function of offline missing energy and scalar sum of transverse momentum	117
6.10	Missing energy distributions in the 1-lepton SM analysis before any multi-jet suppression cuts	119
6.11	Run 1 SM 2-lepton multi-jet estimate	121

7.1	SM VH kinematic fit plots	128
7.2	SM VH 1-lepton multi-jet estimates	131
7.3	1-lepton mTop versus pTW for ttbar and WH signal	137
7.4	Toy example of MLE fit	137
7.5	Demonstration of CL_{S+B} and $1 - CL_B$	146
7.6	SM 0-lepton post-fit distributions	152
7.7	SM 1-lepton post-fit distributions	153
7.8	SM 2-lepton post-fit distributions	154
7.9	SM signal strength breakdown	156
7.10	SM event yield ratio	157
7.11	SM systematic ranking	158
8.1	Demonstration of the b-jet energy correction	162
8.2	Resolved V+jets control regions for both lepton channels	168
8.3	Resolved top- $e\mu$ control regions for the 2-lepton channel	169
8.4	Resolved post-fit invariant mass distributions	171
8.5	Merged post-fit invariant mass distributions	172
8.6	ggF expected 95% upper limit showing the “stitching” of the resolved and merged regimes	173
8.7	95% upper limit as a function of invariant mass	174
A.1	Run 1 trigger levels	182
A.2	Cartoon of Super-RoI	184
A.3	Performance of the online flavor-tagging algorithms	186
B.1	The b-jet scale factor distribution	191
B.2	Bin-uncorrelated c-jet systematics	192
B.3	Naive systematic reduction scheme	195
B.4	Robust systematic reduction scheme	197

C.1	$A \rightarrow Zh$ resolved Z +HF background before smoothing	200
C.2	Histogram of normally distributed data	201
C.3	Demonstration of histograms of double Gaussian distribution.....	201
C.4	Example of kernel density estimate	203
C.5	Example of weighted kernel density estimate	204
C.6	Performance comparison of RooFit KDE versus custom built KDE	206
C.7	Focus on smooth estimate of Z +HF problem area.....	206
C.8	Bin-to-bin correlation matrix for the Z +HF KDE.....	208
C.9	2-D systematic migration matrix	210
C.10	1-D projections of the 2-D migration matrix	210
C.11	1-D projections of systematic Δ s	211
C.12	ggF $A \rightarrow Zh$ limit curve after smoothing the Z +HF background components ..	214

LIST OF TABLES

Table		Page
2.1	Higgs production cross sections calculated at NNLO for a p - p collider at a center-of-mass energy of 13 GeV	22
3.1	Performance goals of the ATLAS detector	41
3.2	EM calorimeter parameters	47
3.3	Hadronic Calorimeter parameters	50
4.1	Muon identification efficiencies	75
4.2	JVT working points	79
4.3	MV2c20 b -tagging working points	86
4.4	MV2c10 b -tagging working points	86
5.1	Signal and background MC samples	103
6.1	Selection criteria for the HIGG5D1 DxAOs	107
6.2	Selection criteria for the HIGG5D2 DxAOs	108
6.3	Selection criteria for the HIGG2D4 DxAOs	109
6.4	Electron object selection requirements	109
6.5	Muon object selection requirements	109
6.6	AntiKt4EMTopoJets selection requirements	110
6.7	AntiKt10LCTopoTrimmedPtFrac5SmallR20Jets and AntiKt2PV0TrackJets selection requirements	111
6.8	Methodology for reconstructing Higgs boson from jets	113
6.9	SM 0-lepton event selection	114
6.10	SM 1-lepton event selection	119
6.11	SM 2-lepton event selection	121

6.12	BSM $A \rightarrow Zh$ 0-lepton event selection	122
6.13	BSM $A \rightarrow Zh$ 2-lepton event selection	124
7.1	BDT inputs for the SM VH analysis	135
7.2	SM VH background modeling uncertainties	141
7.3	SM VH signal modeling uncertainties	143
7.4	SM Post-fit floating background normalization	150
7.5	SM post-fit yields table	151
7.6	SM results	155
8.1	BSM AZh background and signal yields	170

LIST OF SYMBOLS

Symbol	Description	Units
HEP	H igh E nergy P hysics	
SM	S tandard M odel	
BSM	B eyond-the- S tandard M odel	
VH	V ector Boson-Associated Production of the H iggs Boson	
2HDM	Two H iggs D oublet M odel	
MC	M onte C arlo	
LHC	L arge H adron C ollider - circular proton-proton collider capable of center-of-mass energies of up to 14 TeV	
ATLAS	A T oroidal L H C A pparatu S - one of two general purpose dectectors located at the LHC	
CMS	C ompact M uon S olenoid - one of two general purpose dectectors located at the LHC	
SCT	S ilicon-Microstrip T racker	
TRT	T ransition R adiation T racker	
EMEC	E lectrom m agnetic E ndcap C alorimeter	
HEC	H adronic E ndcap C alorimeter	
FCal	F orward C alorimeter	
MDT	M onitored D rift T ubes	
CSC	C athode S trip C hambers	
RPC	R esistive P late C hambers	
TGC	T hin G ap C hambers	
CB	C ombined muons	
ST	S egment- t agged muons	
CT	C alorimeter- t agged muons	
ME	E xtrapolated m uons	

L1	First level of the ATLAS trigger (L evel 1)	
HLT	Second level of the ATLAS trigger (H igh- L evel T rigger)	
JVT	J et V ertex T agging	
ROOT	Software framework for statistical analysis and data processing	
GAUDI	Software architecture used in HEP experiments at CERN	
Athena	Software framework used in the ATLAS detector	
RDO	R aw D ata O bject	
SDO	S imulated D ata O bject	
AOD	A nalysis O bject D ata	
GSF	G aussian S um F ilter	
MVA	M ultivariate A nalysis	
BDT	B oosted D ecision T ree	
PCA	P rincipal C omponent A nalysis	
c	speed of light	1 unit
$\hbar$	Planck's constant divided by 2π	1 unit
p_T	transverse component of momentum	eV
η	pseudorapidity	
E_T^{miss}	missing transverse energy - measured using calorimeters	eV
$E_{T,\text{trk}}^{\text{miss}}$	missing transverse component of momentum - measured using tracking information	eV

To my beautiful and patient wife,
to my parents and grandparents for their boundless support,
and to my best friend for his invaluable camaraderie.
I am eternally in your debt.

Chapter 1

Introduction

Since the discovery of a Higgs-like boson in 2012, the particle physics community has been vigorously probing and measuring the various properties of this new particle. Several studies of the Higgs boson are underway using the two-general purpose experiments at the Large Hadron Collider, ATLAS and CMS. These detectors are currently collecting proton-proton collision data at a center-of-mass energy of 13 TeV and an instantaneous luminosity of about $\mathcal{L} = 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. Thus far, it appears to be consistent with the Standard Model Higgs boson first predicted in 1964¹. The Standard Model Higgs boson allows both the fermions and gauge bosons in the Standard Model to obtain mass without breaking local gauge invariance and was the last missing piece of the Standard Model of particle physics. The properties of the Higgs boson such as spin, CP, and its couplings to other particles are still being determined. The couplings to electroweak vector bosons (photons, W , Z) and the τ -lepton have been observed and there is evidence for its CP and spin aligning with predictions from the Standard Model. Because of the exact nature of the Higgs boson's couplings, observing all of its various properties could be a window into physics beyond the Standard Model. Specifically, the Two-Higgs Doublet Model extension, which remedies a few outstanding issues with the Standard Model, may alter the newly-discovered-boson's couplings from their Standard Model expectation.

One of the many properties of the Higgs boson yet to be directly observed is its Yukawa coupling to quark fermions. The $H \rightarrow b\bar{b}$ decay (the Higgs boson decaying to a pair of b -quarks) is predicted to be most favored by the Standard Model and is the focus of this thesis. Despite this favorability, observing such a decay proves quite difficult at a hadron collider. Detecting these b -quarks requires advanced particle identification techniques and

¹Electroweak symmetry breaking was independently predicted by three groups of theorists: Robert Brout and François Englert, Peter Higgs, and Gerald Guralnik, Carl Hagen, and Sir Thomas Kibble.

uses most of the detector components. My work in this area is highlighted in this thesis where applicable. Because of the difficulties of identifying these b -quarks, the analyses presented in this thesis look for a Higgs boson produced in association with a vector boson. This additional vector boson will be shown to help reduce the contamination from competing background processes.

For all the success of the Standard Model, it has many outstanding issues and known discrepancies with observations. One such issue is the apparent asymmetry of the amounts of baryons and anti-baryons that cannot be explained with our current understanding of the Standard Model. The Two-Higgs Doublet Model remedies this by introducing a second Higgs field into the Standard Model which (among other things) enhances for the inherent CP violation in the Standard Model and results in an excess of left-handed quarks over the corresponding anti-quarks. This extra Higgs field results in four additional Higgs bosons that should be observable at the Large Hadron Collider. Specifically, this thesis will focus on searches for a heavy, CP odd Higgs particle known as the A boson. The b -quark coupling to the Higgs is predicted to be especially sensitive to potential effects of this additional Higgs field.

In addition to my contributions to the aforementioned analyses, technical work I performed for the ATLAS b -jet trigger is detailed. Specifically, I describe how I contributed to improving the performance of identifying b -quark candidates during data collection.

This thesis is organized as follows: the theoretical foundations of the Standard Model and beyond-the-Standard Model Higgs bosons are given in Chapter 2. An overview of the Large Hadron Collider and ATLAS detector is given in Chapter 3, followed by detailed descriptions of simulated datasets and reconstruction algorithms relevant to this thesis in Chapter 4. A concise summary of the datasets used in the analyses is given in Chapter 5. A comprehensive overview of the object and event selection for both analyses is given in Chapter 6, which sets the stage for analysis overviews in Chapters 7 and 8. Concluding remarks are then offered in Chapter 9. Finally, appendices A, B, and C elaborate on some of my contributions.

Chapter 2

Standard Model and 2HDM Higgs Theoretical Framework

This chapter outlines the essential theoretical concepts necessary for motivating searches for the Standard Model (SM) Higgs boson decaying to a pair of b -quarks ($H \rightarrow b\bar{b}$) and using this decay signature to search for Higgs-like A boson in the context of a Two-Higgs Doublet Model (2HDM). Due to practical considerations detailed in Chapters 3, 7, and 8 emphasis will be placed on the vector boson-associated production of the Higgs through proton-proton (p - p) collisions. The following sections will give an overview of Quantum Field Theory (QFT), the SM, electroweak symmetry breaking via the Higgs mechanism, the relevant Higgs production and decay modes, Quantum Chromodynamics (QCD), the 2HDM extension to the SM, and why particle colliders are necessary tools for these searches.

Before starting this chapter I'll cover some essential terminology and concepts. This chapter assumes familiarity with special relativity, including concepts such as the Minkowski metric, 4-vectors, Lorentz invariance, and Einstein summation convention. For instance, the definition of the momentum 4-vector is $p = (E, \mathbf{p})$, where $E = \sqrt{\mathbf{p}^2 + m^2}$ such that $p^2 = m^2$ is the frame-independent, Lorentz-invariant mass squared. This chapter also relies on an understanding of Lagrangian and Hamiltonian mechanics. Recall that the Lagrangian (L) of a physical system is the function whose integral over time minimizes the action. This takes the generic form of the kinetic energy terms less the potential energy terms and is a function only of the generalized coordinates and their first derivatives (and possibly time). The primary result of the minimization of the action is the Euler-Lagrange equation ($\frac{\partial L}{\partial q_i} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} = 0$, where q_i is a generalized coordinate). The Hamiltonian (H) is related to the Lagrangian and (in virtually every case) is the sum of the kinetic and potential energy terms. Non-relativistic quantum mechanics (QM) makes extensive use of the Hamiltonian, which is a function only of the generalized coordinates and their conjugate momenta (and possibly time). In QM

the coordinates and momenta are promoted to operators that obey commutation relationships¹. These relationships have profound physical consequences that are too numerous to cover here. The quantum mechanical simple harmonic oscillator (SHO) has analogies in the next section’s discussion on QFT. In solving for the SHO energy eigenvalues the so-called “ladder operators” (denoted as a and $a^\dagger$) are introduced which raise and lower the energy levels of the SHO. The concept of intrinsic angular momentum (spin) also carries over to QFT. Finally, physically realizable states in QM exist in a (possibly infinite-dimensional) real, inner-product space known as a Hilbert space, which also has an analogue in QFT.

2.1. Basic Principles of Quantum Field Theory

QFT is a natural extension of quantum mechanics and classical, relativistic field theory, where the fields are promoted to quantum mechanical operators, and physical states now inhabit Fock space². QFT serves as the fundamental theoretical framework in particle physics, and the Lagrangian³ (or, more accurately, Lagrangian density $\mathcal{L}$) plays the leading role in determining how particles interact. These interactions are determined by the presence of low-order, polynomial terms of particular fields. A critically important result of QFT is that the number of particles is not fixed, as will be shown later. This allows the theory to naturally accommodate the empirical fact that particle number is not conserved in a number of physical processes, including high-energy collisions and radioactive decay. All the material regarding this section is taken from Refs. [1], [2], and [3].

The types of fields commonly used in particle physics are scalar (ϕ), spinor (ψ), and vector fields ($A_\mu, W_\mu^\pm, Z_\mu, G_\mu^C$) which will be shown to give rise to the Higgs boson, fermions (matter), and all non-Higgs bosons, respectively, in the next section. Each type of field is characterized by its transformation properties under parity inversion ($x_P = (x_0, -\mathbf{x})$), charge conjugation (particle $\leftrightarrow$ anti-particle exchange), and time reversal ($x_T = (-x_0, \mathbf{x})$). For the purposes of particle physics, all terms in the Lagrangian must be manifestly Lorentz-

¹The commutator between operators A and B is defined as $[A, B] = AB - BA$

²A Fock space is a set of Hilbert spaces for multi-particle states (it formally allows the number of particles to vary from state to state)

³The Lagrangian is used rather than the Hamiltonian as the Lagrangian is manifestly Lorentz-invariant.

invariant, locally gauge invariant, and be only functions of the fields and their first derivatives. Local gauge invariance is a property wherein an expression involving one or more fields is left unchanged after the addition of another spacetime-dependent field. For instance, the Lagrangian for a massless vector field A_μ is simply $\mathcal{L} = \frac{1}{4}F_{\mu\nu}F^{\mu\nu}$, where $F^{\mu\nu}$ is defined later. This happens to be invariant to a shift in A_μ by the gradient of a spacetime-dependent scalar field $Q\alpha(x)$, $A_\mu \rightarrow A_\mu - \frac{1}{Q}\partial_\mu\alpha(x)$, where Q is the electric charge of a (yet unspecified) particle. Thus, we would say that this Lagrangian is invariant under a local gauge transformation. This also allows for freedom to choose the “gauge” ($\alpha(x)$) in order to make computation easier. Local gauge invariance is a strong requirement that has a profound impact on the allowable physics.

2.1.1. Scalar Fields

The Lagrangian for a complex, non-self-interacting, free, and massive scalar field is $\mathcal{L} = \frac{1}{2}(\partial_\mu\phi)^2 - \frac{1}{2}m^2\phi^2$. The solution to the equation of motion of a free field for such a Lagrangian is

$$\phi = \int \frac{d^3\mathbf{p}}{(2\pi)^3 2E_p} \left[a(p)e^{-ip^\mu x_\mu} + b(p)^\dagger e^{ip^\mu x_\mu} \right]. \quad (2.1)$$

The $a(p)$ and $b(p)$ are creation and annihilation operators for this scalar particle. These have a direct analogy to the ladder operators for the SHO in QM. Instead of raising or lowering the energy level of the SHO, these QFT creation and annihilation operators act on a state in Fock space to either create or destroy a particular (anti-)particle. Also the complex scalar field carries electric charge (else, it wouldn't have an anti-particle).

Scalar fields transform like $\hat{P}\phi(x)\hat{P}^{-1} = \eta_P\phi(x_P)$ under the parity operator $\hat{P}$. The intrinsic parity η_P for the scalar field in the Standard Model (i.e. the Higgs field) is either 1 (scalar) or -1 (pseudo-scalar). Charge conjugation $\hat{C}$ takes the form of $\hat{C}\phi(x)\hat{C}^{-1} = \eta_C\phi^\dagger(x)$ and $\hat{C}\phi^\dagger(x)\hat{C}^{-1} = \eta_C^*\phi(x)$ with $\eta_C = \pm 1$ for hermitian ϕ . Time reversal $\hat{T}$ is similar in form to parity inversion (i.e. $\hat{T}\phi(x)\hat{T}^{-1} = \eta_T\phi(x_T)$ with $\eta_T = \pm 1$).

2.1.2. Spinor Fields

The Lagrangian⁴ for a free and massive spinor field is $\mathcal{L} = i\bar{\Psi}\not{\partial}\Psi - m\bar{\Psi}\Psi$. The equation of motion for a free spinor field $\bar{\Psi}$ is $(-i\not{\partial} + m)\Psi = 0$, which has solutions (after multiplying by $(i\not{\partial} + m)$)

$$\Psi(x) = \sum_{s=\pm} \int \frac{d^3\mathbf{p}}{(2\pi)^3 2E_p} [b_s(p)u_s(p)e^{ip^\mu x_\mu} + d_s^\dagger(p)v_s(p)e^{-ip^\mu x_\mu}], \quad (2.2)$$

where s is the helicity (component of spin along the direction of motion) state, and $b_s^{(\dagger)}(p)$ and $d_s^{(\dagger)}(p)$ are the “fermionic” creation and annihilation operators. The spinor properties of the field are manifest in the two $u_s(p)$ and $v_s(p)$ spinors. From the equation of motion these spinors each solve $(\not{p} + m)u(p) = 0$ and $(-\not{p} + m)v(p) = 0$. Spinor fields in the Standard Model Lagrangian are included in the form of so-called bi-linears, such as $\bar{\Psi}(x)\Psi(x)$, $\bar{\Psi}(x)\gamma_5\Psi(x)$, $\bar{\Psi}(x)\gamma^0\Psi(x)$, and $\bar{\Psi}(x)\boldsymbol{\gamma}\Psi(x)$.

Spinor fields transform like $\hat{P}\Psi(x)\hat{P}^{-1} = \eta_P\gamma^0\Psi(x_P)$ and $\hat{P}\bar{\Psi}(x)\hat{P}^{-1} = \eta_P^*\bar{\Psi}(x_P)\gamma^0$ under the parity operator. The intrinsic parity η_P for the spinor fields doesn’t depend on the spin of the field as the parity operator leaves spin unaltered. Charge conjugation is a bit more complicated for spinors than parity inversion. The general result is that the spinors used in the Standard Model are invariant under $\hat{C}$. However, there are additional choices to be made such as if a particle is described by a “Majorana” or “Dirac” spinor field⁵. Time reversal looks like $\hat{T}\Psi(x)\hat{T}^{-1} = \eta_T B^{-1}\Psi(x_T)$ and $\hat{T}\bar{\Psi}(x)\hat{T}^{-1} = \eta_T^* \bar{\Psi}(x_T)B$ where $B = \gamma_5 C$ in which C satisfies $C(\gamma^\mu)^t C^{-1} = -\gamma^\mu$ ($(\gamma^\mu)^t$ is the transpose of γ^μ).

2.1.3. Vector Fields

The Lagrangian for a non-self-interacting, free, and massive vector field V is $\mathcal{L} = -\frac{1}{4}F_{\mu\nu}(V)F^{\mu\nu}(V) + \frac{1}{2}m_V^2 V^\mu V_\mu$, where $F_{\mu\nu}(V) = \partial_\mu V_\nu - \partial_\nu V_\mu$ is the V -field tensor. The

⁴The “slash” notation ($\not{\partial}$) is defined as a contraction of the Dirac matrices γ_μ with another operator (i.e. $\not{\partial} = \gamma^\mu \partial_\mu$). Also, $\bar{\Psi} = \Psi^\dagger \gamma^0$.

⁵A Majorana spinor field describes particles that are their own anti-particle (i.e. they are invariant under $\hat{C}$). As such, Majorana particles cannot have electric charge. Whether or not neutrinos are described by Majorana spinor fields is an open question. All other fermions in the SM are Dirac fermions (i.e. they have distinct particle/anti-particle states).

free-field solution to the equation of motion for such a Lagrangian is

$$V_\mu = \int \frac{d^3\mathbf{k}}{(2\pi)^3 2E_k} \sum_{\text{spin}} \sum_{i=1}^{N_{\text{pol}}} \left(\epsilon_\mu^i(k) a_{k,i}^\dagger e^{ik^\nu x_\nu} + \epsilon_\mu^i(k) a_{k,i} e^{-ik^\nu x_\nu} \right). \quad (2.3)$$

If the mass parameter is zero (photon) then the free field will only have two polarization states ϵ_μ^i available; massive fields such as those for $W^\pm$ and Z have three. The spin states are summed over.

Vector fields transform like $\hat{P}V^\mu(x)\hat{P}^{-1} = \eta_P V_\mu(x_P)$ under the parity operator. The intrinsic parity η_P for the vector fields in the Standard Model is either 1 or -1 (pseudo-vector or axial-vector). Charge conjugation takes the form of $\hat{C}V_\mu(x)\hat{C}^{-1} = \eta_C V_\mu(x)^\dagger$ where η_C is determined by the details of the theory. Just as for scalar fields, time reversal for vector fields is similar in form to parity inversion (i.e. $\hat{T}V^\mu(x)\hat{T}^{-1} = \eta_T V_\mu(x_T)$).

2.1.4. Scattering

Of most relevance to particle physicists are the probabilities of producing certain particles (final state) given the momentum and quantum numbers of the incoming particles (initial state) during a collision. These probabilities are related to the computation of cross sections and decay widths. The cross section is defined as the probability of a process to occur while the decay width is probability per unit time a process will occur. These are commonly calculated by a perturbative series starting with the “free particle” solutions (Equations 2.1, 2.3) and sequentially adding more interaction terms. In general, the Lagrangian may be separated

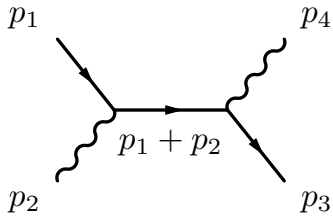


Figure 2.1

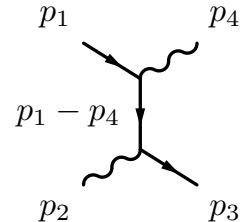


Figure 2.2

Figure 2.3: Leading-order Compton scattering Feynman diagrams.

into free particle terms ($\mathcal{L}_{\text{free}}$) and interaction terms ($\mathcal{L}_{\text{int}}$)

$$\mathcal{L} = \mathcal{L}_{\text{free}} + \mathcal{L}_{\text{int}}. \quad (2.4)$$

Suppose that at time t_0 a system starts in state $|t_0\rangle$ and evolves to a different state at time t . In the so-called “interaction picture” this looks like

$$|t\rangle = U(t, t_0) |t_0\rangle,$$

where the unitary⁶ time-evolution operator $U(t, t_0)$ contains the full description of how the initial state changes over time. In a scattering process (i.e. a collision) $t_0 \rightarrow -\infty$ and $t \rightarrow \infty$. In this limit the time-evolution operator is known as the S -matrix. The specific form of the S -matrix is

$$S = T \left[e^{i \int d^4x \mathcal{L}_{\text{int}}(x)} \right] = 1 + i \int d^4x \mathcal{L}_{\text{int}}(x) + \frac{i^2}{2!} \int d^4x_1 \int d^4x_2 T[\mathcal{L}_{\text{int}}(x_1) \mathcal{L}_{\text{int}}(x_2)] + \dots,$$

where $T[\dots]$ represents the time-ordered product⁷. The unitary S -matrix is used to compute probability amplitudes between initial and final states $\langle f|S|i\rangle$. The S -matrix can be written as $S = \mathbf{1} + \mathcal{T}$, where the transition matrix $\langle f|\mathcal{T}|i\rangle = i(2\pi)^4 \delta^4(p_f - p_i) \mathcal{M}_{i \rightarrow f}$. $\mathcal{M}$ is known as the scattering amplitude and (in the interaction picture) is essential to calculating the rate of quantum processes. Prescriptions for calculating $\mathcal{M}$ can be summarized by a theory’s “Feynman rules” along with corresponding diagrams. These “Feynman diagrams” are a compact representation of terms in the interaction picture’s perturbative expansion. For instance, the leading-order (LO) diagrams for inelastic scattering of a photon by an electron (Compton scattering) are given in Figure 2.3. LO refers to the fact that the fig-

⁶Unitarity is important for the preservation of probabilities.

⁷A time-ordered product for n spacetime-dependent operators $A_i(x)$ is defined as $T[A_1 A_2 \dots A_n] = \sum_p \theta(t_{p_1} > t_{p_2} > \dots > t_{p_n}) \varepsilon(p) A_{p_1}(t_{p_1}) A_{p_2}(t_{p_2}) \dots A_{p_n}(t_{p_n})$, where $\theta(\dots)$ is the Heaviside step function and $\varepsilon(p)$ is 1 or -1 depending on if the operator represents a boson or fermion, and the permutation p of the n operators.

ure only shows Feynman diagrams representing the first non-zero term⁸ in the S -matrix⁹. The convention in Figure 2.3 is that time increases from left to right (separately for each diagram). The straight-line segments represent fermions (e^+ or e^-) and the wavy lines are bosons (photons). Each line segment representing an initial or final state particle (“external” lines) is assigned momentum while the “internal” lines have momenta constrained by the external lines. Each vertex contributes a term to the scattering amplitude $\mathcal{M}$ whose form is prescribed by the theory’s Feynman rules. The terms for each diagram are combined to form the total amplitude $\mathcal{M}_T$. The initial and final spins states are then averaged as generally one is not concerned with these quantities for the incoming and outgoing electrons and photons. $|\mathcal{M}_T|^2$ is then used to compute the final differential cross section. For Compton scattering the differential cross section is normally given in terms of the angle of the “deflected” photon. In particle physics the general prescription of the development of a potential new theory is to postulate terms in the Lagrangian that satisfy some desired symmetry and what new particle(s) associated with these terms is(are) (or visa versa). Once the Lagrangian is known Feynman rules can be formulated which are used to compute scattering amplitudes for calculating (differential) cross sections. The formula for computing the differential cross section for two-body scattering (denoted i and j) where there are n decay products is

$$d\sigma = |\mathcal{M}|^2 \frac{S}{4\sqrt{((p^\mu)_i(p_\mu)_j)^2 - (m_i m_j)^2}} \left(\prod_{k=1}^n \frac{d^3 \mathbf{p}_k}{(2\pi)^3 2E_k} \right) \times (2\pi)^4 \delta^4 \left(p_i + p_j - \sum_{k=1}^n p_k \right), \quad (2.5)$$

where S is a product of statistical factors: $\frac{1}{m!}$ for each group of m identical particles in the final state. The formula for the differential decay width (inverse of proper lifetime) for a single particle that decays to n products is

$$d\Gamma = |\mathcal{M}|^2 \frac{S}{2m} \left(\prod_{k=1}^n \frac{d^3 \mathbf{p}_k}{(2\pi)^3 2E_k} \right) \times (2\pi)^4 \delta^4 \left(p - \sum_{k=1}^n p_k \right).$$

⁸Terms beyond these are commonly referred to as “higher-order” terms. These terms show up in Feynman diagrams as having additional external lines or internal “loops” that can significantly complicate computation. Calculations that include the next non-zero terms are referred to as NLO. Non-zero terms beyond this simply add another “N” to the “LO” base (i.e. NNLO, N3LO, etc.).

⁹The particular terms for Compton scattering are determined through quantum electrodynamics.

These two formulae along with Feynman rules and diagrams form the theoretical basis for testable calculations in particle physics. The explicit calculation of the scattering amplitude is quite involved, but a couple of items are worthy of note: the external lines in a Feynman diagram are represented by coefficients from the free-particle solutions given above, and internal lines are represented by “propagators” whose form is inversely proportional to the particle’s mass (or energy in the case of a photon), momentum, and decay width. In two-body interactions where the incoming particles have momenta p_1 and p_2 and the outgoing particles have momenta p_3 and p_4 , three useful variables are defined that typify the nature of the interaction:

- $s = (p_1 + p_2)^2 = (p_3 + p_4)^2 \geq 0$
- $t = (p_1 - p_3)^2 = (p_2 - p_4)^2 \leq 0$
- $u = (p_1 - p_4)^2 = (p_2 - p_3)^2 \leq 0$

Momentum transfers consistent with the s - and t -channels were shown in Figures 2.1 and 2.2, respectively. These are known as the Mandelstam variables and I’ll only discuss the one that is used ubiquitously in all of experimental, high-energy particle physics, s . The momentum transfer that s typifies is called the s -channel (for timelike) and is the channel responsible for the direct production and detection of particles. In this channel, two incoming particles create a single particle which then travels some distance before decaying to the two outgoing particles. Thus, at a particle collider the average energy at which the particles collide is presented in terms of $\sqrt{s}$ (since s has units of energy squared). The “pole mass” is the invariant mass of a particle that is independent of the kinematics of the particle. The name arises due to the fact that computations with propagators involve integrating in the complex plane¹⁰ and stable particles show up as poles. A particle with a mass at its pole mass is said to be “on-shell,” otherwise it’s said to be “off-shell.” Generally speaking, particles in an interaction “prefer” to occupy their pole mass (it’s most probable if it’s kinematically available after conservation of momentum, etc.).

¹⁰Propagators also mathematically resemble driven, damped harmonic oscillators from classical mechanics. Thus, the pole mass is also referred to as a particle’s resonant mass (as in resonant frequency).

2.2. The Standard Model of Elementary Particle Physics

Among the most successful and explanatory theories in all of physics is the Standard Model of Elementary Particle Physics (SM). This model describes with great accuracy the interactions between the most fundamental building blocks of nature. A summary of these building blocks, or particles, is given in Figure 2.4. This figure shows the so-called “generations” of fermions, the first of which is the first column with the second and third generations being the second and third columns, respectively. Fundamental fermions of the SM all have half-integer spin. The quarks all have electric charge in multiples of $\frac{1}{3}$ while the leptons all have integer charge. The leptons can be further classified by those that have electric charge (electron, muon, tau) and those that are absent charge (various neutrinos). Coincidentally, the leptons without charge are virtually massless with many global-scale experiments searching for their exact mass values. The fourth column contains the mediators particles for the electroweak and strong forces. The gluon mediates the strong force, which is necessary for binding of atomic nuclei and quarks within hadrons¹¹. The photon is the mediator of the electromagnetic (EM) force. Since it mediates this force, the photon is quantization of the EM field and is, thus, synonymous with what we commonly refer to as light as well as the entire EM spectrum. The two remaining W and Z bosons are the mediators of the weak force (weak bosons) which is responsible for radioactive decays. This column of force mediators is represented as vector fields in the SM. Each field may have “internal” symmetries¹² in addition to their Lorenz symmetry. These fields are also known as gauge fields (as in gauge invariant) for reasons that will be made clear in the next sections. The internal symmetry the electroweak sector of the SM obeys is $SU(2) \times U(1)$ while the strong sector obeys $SU(3)$. The last item in this table is the Higgs boson which is the mediator of the Higgs field. This field is responsible for the masses of the fermions and weak bosons. What follows is an overview of these fields along with the corresponding theory (terms in the Lagrangian and symmetry properties). All the material regarding this section is taken from Refs. [1, 3, 5, 6].

¹¹Hadrons (i.e. protons and neutrons) are composite objects made of quarks.

¹²Internal symmetries here means that a field may have additional indices that obey certain group theoretic constraints orthogonal to the field’s spacetime classification (i.e. scalar, vector, spinor).

mass →	$\approx 2.3 \text{ MeV}/c^2$	$\approx 1.275 \text{ GeV}/c^2$	$\approx 173.07 \text{ GeV}/c^2$	0	$\approx 126 \text{ GeV}/c^2$
charge →	$2/3$	$2/3$	$2/3$	0	0
spin →	$1/2$	$1/2$	$1/2$	1	0
	u up	c charm	t top	g gluon	H Higgs boson
QUARKS	$\approx 4.8 \text{ MeV}/c^2$ $-1/3$ $1/2$	$\approx 95 \text{ MeV}/c^2$ $-1/3$ $1/2$	$\approx 4.18 \text{ GeV}/c^2$ $-1/3$ $1/2$	0 0 1	
	d down	s strange	b bottom	γ photon	
	$0.511 \text{ MeV}/c^2$ -1 $1/2$	$105.7 \text{ MeV}/c^2$ -1 $1/2$	$1.777 \text{ GeV}/c^2$ -1 $1/2$	$91.2 \text{ GeV}/c^2$ 0 1	
	e electron	μ muon	τ tau	Z Z boson	
LEPTONS	$< 2.2 \text{ eV}/c^2$ 0 $1/2$	$< 0.17 \text{ MeV}/c^2$ 0 $1/2$	$< 15.5 \text{ MeV}/c^2$ 0 $1/2$	$80.4 \text{ GeV}/c^2$ ± 1 1	
	ν_e electron neutrino	ν_μ muon neutrino	ν_τ tau neutrino	W W boson	GAUGE BOSONS

Figure 2.4: The names, masses, charges, and spins of the particles contained in the Standard Model of particle physics [4]. The green and purple groups (left three columns) make up the fermions in the SM. The purple group (upper left 2×3 cells) are known as quarks, and the green group (lower left 2×3 cells) are known as leptons. The red group (fourth column) of particles are responsible for the mediation of forces (vector bosons) while the remaining particle in yellow (fifth column) is the Higgs boson which is responsible for the masses of the fermions and the W and Z bosons.

2.2.1. Electroweak Theory

Electroweak theory is the mathematical description of the leptons, quarks, vector bosons (except the gluon), and their interactions. Recall from quantum electrodynamics (QED) that the Lagrangian for a spinor field Ψ (possibly representing an electron) interacting with the EM field A_μ is $\mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + i\bar{\Psi}\not{D}\Psi - m\bar{\Psi}\Psi$, where the “covariant derivative” is defined as $D_\mu = \partial_\mu - iQA_\mu$ (Q is the electric charge of the field). This definition of the derivative is vital for recasting the Lagrangian in a manifestly gauge invariant form (under the gauge group $G = U(1)_Q$, where Q denotes electric charge¹³) as the derivatives will transform so as to cancel field transformations. While it doesn’t make sense to discuss a massive photon, in principle any boson described with a similar covariant derivative is forbidden from having a mass term in the Lagrangian. Also, recall the chirality of the spinor fields. A spinor Ψ may be written as a sum of “left-handed” and “right-handed” components $\Psi(x) = \Psi_L(x) + \Psi_R(x)$, where $\Psi_{R/L}(x) = \frac{1}{2}(1 \pm \gamma_5)\Psi(x)$ (‘+’ is associated with right-handed spinors and ‘-’ with left-handed). The Lagrangian of these decomposed spinors is $\mathcal{L} = \bar{\Psi}_R(x)i\gamma\partial\Psi_R(x) + \bar{\Psi}_L(x)i\gamma\partial\Psi_L(x) - m[\bar{\Psi}_R(x)\Psi_L(x) + \bar{\Psi}_L(x)\Psi_R(x)]$. When the Lagrangian is decomposed in this manner it’s clear that the mass term is responsible for the mixing of the left- and right-handed components. In other words, massless fermions can not change their “handedness.” Indeed, a simple calculation shows that massless (anti-)neutrinos are (right-)left-handed.

As in QED, the gauge group of the electroweak sector determines the properties of the gauge bosons and interactions. Whatever this group is, it should also incorporate the gauge group of QED, $U(1)$. Through several experimental sources of evidence [7,8] the gauge group for electroweak theory is $G = SU(2) \times U(1)_Y$ ¹⁴. Y is often referred to as weak-hypercharge and the $SU(2)$ is known as weak-isospin. The generators of $SU(2)$ are denoted $i\mathbf{T}$, where

¹³ $U(1)_Q$ is the group of all complex numbers that lie on the unit circle in the complex plane. To be explicit, in QED the spinor fields transform as $\Psi \rightarrow e^{iQ\alpha}\Psi$. The subscript Q simply denotes that the electric charge (unit-less in natural units) is to be an additional factor in the transformation.

¹⁴ $SU(n)$ is the Lie group of $n \times n$ unitary matrices with determinant of 1. This group is non-Abelian (elements of group are non-commuting). Thus, the gauge theories that obey these group symmetries are referred to as non-Abelian gauge theories.

$\mathbf{T} = (T_1, T_2, T_3)$ and $[T_a, T_b] = i\epsilon_{abc}T_c$. The degrees of freedom in this gauge group (three from $SU(2)$ + one from $U(1)$ makes four) implies four gauge bosons, $\mathbf{W}_\mu = (W_\mu^1, W_\mu^2, W_\mu^3)$ and B_μ . Note that the upper index specifying the three components of $\mathbf{W}_\mu$ have nothing to do with the spacetime index. This is an instance of an internal symmetry mentioned earlier. This gauge group implies all left-handed spinors are now doublets (2-D mathematical vectors) of spinors while the right-handed spinors are singlets (normal spinors) in the space of internal quantum numbers. The various doublets for the leptons (first, second, and third generations) are

$$L_e = \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}, \quad L_\mu = \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}, \quad L_\tau = \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix},$$

and for the quarks (first, second, and third generations) are

$$Q_u = \begin{pmatrix} u \\ d \end{pmatrix}, \quad Q_c = \begin{pmatrix} c \\ s \end{pmatrix}, \quad Q_t = \begin{pmatrix} t \\ b \end{pmatrix}.$$

Each element in these doublets represent a particle field.

The Lagrangian including only the first generation of fermions is (ignoring scalar field terms)

$$\mathcal{L}_{EW} = -\frac{1}{4}\mathbf{W}_{\mu\nu} \cdot \mathbf{W}^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} + \overline{Q}_u i \not{D} Q_u + \bar{u} i \not{D} u + \bar{d} i \not{D} d + \overline{L}_e i \not{D} L_e + \bar{e} i \not{D} e,$$

where $\mathbf{W}_{\mu\nu} = F_{\mu\nu}(\mathbf{W})$, $B_{\mu\nu} = F_{\mu\nu}(B)$, and the covariant derivative for this gauge group is

$$\mathbf{D}_\mu = \partial_\mu \mathbf{I}_2 - i\frac{g_1}{2}Y B_\mu \mathbf{I}_2 - i\frac{g_2}{2}\mathbf{W}_\mu \cdot \boldsymbol{\sigma},$$

where $\boldsymbol{\sigma}$ is the 3-D vector of Pauli matrices¹⁵. The group generators for $SU(2) \times U(1)_Y$

¹⁵The use of the Pauli matrices here isn't necessary. Any representation of the $SU(2)$ group will do. For dimensional consistency I've multiplied the first two terms by the 2×2 identity matrix $\mathbf{I}_2$, which usually isn't explicitly done in the literature.

looks like $g(\boldsymbol{\alpha}, \beta) = e^{-i\boldsymbol{\alpha}(x) \cdot \mathbf{T}} e^{-i\beta(x) Y \mathbf{I}_2}$. Notice that the exponential expression involving Y clearly represents the $U(1)_Y$ part of the gauge group. Now, consider adding a fermionic mass term to the Lagrangian of the form $-m\bar{\Psi}\Psi \rightarrow -m[\bar{\Psi}_R\Psi_L + \bar{\Psi}_L\Psi_R]$. Since the left- and right-handed components of the spinor couple differently to the gauge group (Ψ_L gets the full $SU(2) \times U(1)_Y$ $g(\boldsymbol{\alpha}, \beta)$ transformation while Ψ_R only sees the $U(1)_Y$ portion), terms like this are not invariant under this gauge group. The details of how the vector fields and derivatives transform in this gauge is omitted for brevity. Without showing the details, this Lagrangian is also incompatible with massive weak bosons (contrary to observation). The next section shows how to remedy this.

2.2.2. Spontaneous Electroweak Symmetry Breaking and The Higgs Mechanism

A complex scalar, $SU(2) \times U(1)_Y$ doublet with quadratic and quartic terms in its Lagrangian will now be shown to fix the shortcomings of the electroweak model mentioned in the previous section, namely that the fermions and gauge bosons cannot have mass terms. However, in order to give these particles mass the $SU(2) \times U(1)_Y$ symmetry must be spontaneously broken. Here, spontaneous symmetry breaking refers to the fact that the $SU(2) \times U(1)_Y$ symmetry is respected in all but the vacuum state in the SM. This is known as the ‘‘Higgs mechanism’’ [10–12] (or the Brout–Englert–Higgs mechanism, among others). From the previous definition of Y this ‘‘Higgs doublet’’ must have $Y = +\frac{1}{2}$ as we desire an uncharged vacuum state. The form of the Higgs doublet is

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

The Lagrangian terms for Higgs doublet are $\mathcal{L} = (\mathbf{D}^\mu \Phi)^\dagger \mathbf{D}_\mu \Phi - \mu^2(\Phi^\dagger \Phi) - \lambda(\Phi^\dagger \Phi)^2$, where $\mu^2 < 0$, $\lambda > 0$, and $\mathbf{D}_\mu$ is the familiar covariant derivative for $SU(2) \times U(1)_Y$. The vacuum state can be thought of as the state in which the potential terms in the Lagrangian for this doublet are minimized (w.r.t. the scalar fields). Recasting the Lagrangian in a unitary gauge

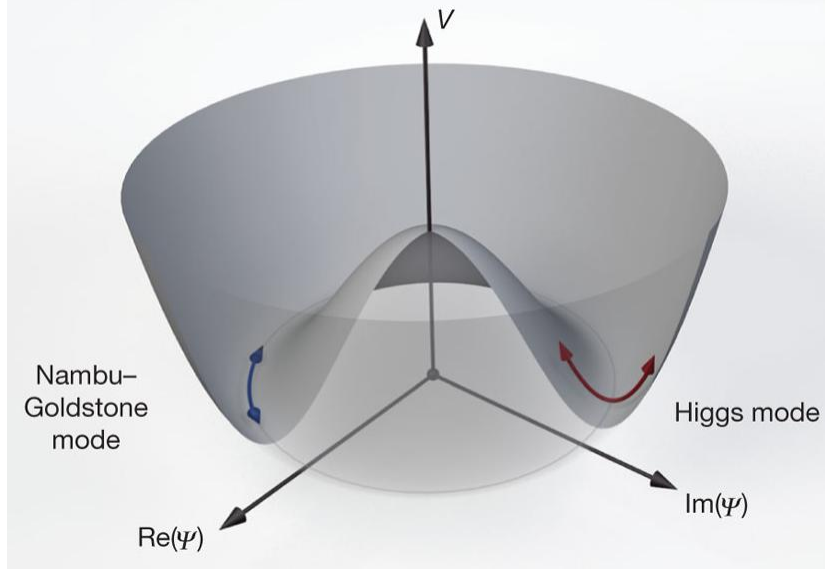


Figure 2.5: The potential for a complex scalar field (either ϕ^+ or ϕ^0 of the Higgs field doublet) with quadratic and quartic potential terms [9]. This illustrates the interpretation of the v as the ground state (vacuum) of the Higgs field (the circular minimum). The Goldstone bosons (labeled “Nambu-Goldstone”) are interpreted as quanta that move within this ground state.

allows the vector bosons present in $\mathbf{D}_\mu$ to absorb¹⁶ ϕ^+ and ϕ_4 , leaving only the real ϕ_3 field. The vacuum for this doublet can be shown to be (after recasting to a unitary gauge and finessing the equations)

$$\Phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix},$$

where $v = \langle \Phi_0 \rangle = \sqrt{\frac{-\mu^2}{\lambda}}$ and $h(x)$ is a real-valued scalar field. v represents the vacuum expectation value for the Higgs field $\langle \Phi_0 \rangle$. This form of the vacuum arises due to the quartic nature of the Higgs field. To help illustrate what this looks like in the space of scalar fields, a plot of the potential of a complex scalar field singlet is shown in Figure 2.5.

Multiplying this Φ_0 with the generators of the $SU(2) \times U(1)_Y$ gauge group demonstrates this broken symmetry (i.e. the results are non-zero). However, the $U(1)_Q$ symmetry is still preserved. After considerable manipulation the observable gauge bosons (photon, Z , and $W^\pm$) can be written in terms of the $\mathbf{W}_\mu$ and B_μ fields. These equations are $W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp$

¹⁶These are the so-called massless Goldstone bosons. Further detail can be found in Refs. [13] and [14].

iW_μ^2), $A_\mu = \sqrt{g_1^2 + g_2^2}^{-1}(g_1 W_\mu^3 + g_2 B_\mu)$, and $Z_\mu = \sqrt{g_1^2 + g_2^2}^{-1}(g_2 W_\mu^3 - g_1 B_\mu)$, where g_1 and g_2 are the coupling constants for the weak vector bosons. Rewriting the $(D^\mu \Phi_0)^\dagger (D_\mu \Phi_0)$ Lagrangian in terms of these observable bosons gives a massless photon, $M_{W^\pm} = \frac{1}{2}vg_2$, and $M_Z = \frac{1}{2}v\sqrt{g_1^2 + g_2^2}$. The SM gives no constraint on g_1 , g_2 , or v so the masses of the $W^\pm$ and Z bosons are free parameters to be determined from experiment¹⁷. The ratio $M_{W^\pm}/M_Z$ is expressed in terms of a mixing angle θ_W (weak mixing angle): $M_{W^\pm}/M_Z = \cos \theta_W$ which has a value of $\sin^2 \theta_W = 0.2397 \pm 0.0013$. The scalar field mass term gives a mass of the Higgs boson (referred to as the H -boson in the SM) as $m_H = \sqrt{2\lambda}v$. At the lowest order in electroweak perturbation theory, this can be measured as another parameter in the SM. v can be calculated independently [16] and has been found to be about 246 GeV (this is also known as the electroweak scale). Lepton masses depend on Yukawa couplings to the Higgs field. Just as in the case for a vector boson, the masses of the leptons are not predicted. However, in contrast to vector bosons there is no relationship amongst the various generations of leptons. The masses for quarks are somewhat more difficult to derive in electroweak theory. This will be touched on in the next section. The current values of all the SM masses can be found in Figure 2.4. As a final note of terminology the coupling between Dirac spinors and a scalar field is known as a Yukawa coupling. Thus, it is common to refer to the Yukawa couplings when searching i.e. the $b\bar{b}$ decay of the Higgs boson.

2.2.3. Quantum Chromodynamics

Quarks fermion fields are subject to an additional $SU(3)$ internal symmetry. This gauge group has eight degrees of freedom whose generators $i\mathbf{T}$ follow a similar commutation relationship as in $SU(2)$, namely $[\mathbf{T}_a, \mathbf{T}_b] = if_{abc}\mathbf{T}_c$. However, the $\mathbf{T}_i$ are now 3×3 matrices and f_{abc} are this group's structure constants. The usual representation of the $\mathbf{T}_i$ are the so-called Gell-Mann matrices $\frac{\lambda_i}{2}$. As with electroweak theory, the eight degrees of freedom lead to eight real, vector gauge bosons denoted $\mathbf{G}_\mu = G_\mu^C$, where $1 \leq C \leq 8$. These bosons are known as gluons and act as the mediators for the strong force. By a similar argument

¹⁷This is true in purely theoretical terms. However, Nature provides a clever mechanism for computing the masses of these bosons given θ_W and another constant known as G_F [15].

as in electroweak theory, the construction of a covariant derivative to act on quark spinors with G_μ^C is easy to imagine:

$$D_\mu = \partial_\mu \mathbf{I}_{\text{GaugeDim.}} - \langle \text{constant} \rangle \times \langle \text{generator} \rangle \times \langle \text{gauge field} \rangle \rightarrow D_\mu = \partial_\mu \mathbf{I}_3 - ig \mathbf{T} \cdot \mathbf{G}_\mu,$$

where g is the coupling constant. Note that the expression $i\mathbf{T} \cdot \mathbf{G}_\mu$ is the sum of eight 3×3 matrices, each with a spacetime index μ . It's unclear how this could operate on quark spinor fields if they remain $SU(3)$ singlets. This is remedied by promoting the quark spinor fields to $SU(3)$ triplets that have indices that match the $SU(3)$ matrix indices. Thus, the quark-gluon interaction term in the Lagrangian now looks like

$$\mathcal{L} = -g \bar{\Psi}_j G_C^\mu \frac{\lambda_{ij}^C}{2} \gamma_\mu \Psi_i, \quad (2.6)$$

where the $1 \leq i, j \leq 3$ are the $SU(3)$ matrix indices. These indices reflect the so-called color charge and, as indicated by the different indices in Equation 2.6 for the two spinors, quarks may change their color charge through interactions with the gluons. Every quark carries color charge, and gluons can be thought to carry two colors simultaneously despite being charge-less in the electromagnetic sense. All non-quark spinor fields in the SM are $SU(3)$ singlets (also called “color singlets” or color neutral) and, thus, do not interact with the gluon field. It is this concept of color charge that spawned the name quantum chromodynamics (QCD). There is, of course, a field-tensor term to add to the Lagrangian as well of the form $-\frac{1}{4} \mathbf{G}_{\mu\nu} \mathbf{G}^{\mu\nu}$, where $\mathbf{G}^{\mu\nu} = F^{\mu\nu}(\mathbf{G}) - g \frac{g}{2} [G_C^\mu \boldsymbol{\lambda}^C, G_C^\nu \boldsymbol{\lambda}^C]$. This is the origin of gluon self-interactions and the presence of the commutator allows for three- and four-gluon self-interactions.

An interesting consequence of enforcing that all terms in the Lagrangian involving quarks and gluons are color singlets is that this implies some sort of “color-confinement”: that nature prefers to configure quarks and gluons into systems that are color-neutral. This concept of color-confinement is bolstered by the empirical observation of quarks and gluons always bound in such color-neutral systems (i.e. no free quark or gluon has ever been observed

in nature or a laboratory). Color-neutral, bound systems of quarks and gluons are called hadrons. In a particle collider, quarks and gluons involved in the collision may disassociate from their original hadron. However, these quarks and gluons are always found to “hadronize” after the hard collision: they immediately form new color singlets (hadrons) such that no single quark or gluon is observed. A hadron with half-integer spin made from an odd number of quarks is called a baryon. A hadron with integer spin made from an even number of quarks is called a meson. The proton and neutron which make up the nuclei of atoms are baryons. These baryons make up most of the observable mass in our universe¹⁸ (electrons are tiny by comparison).

Of particular importance in particle physics is the concept of the running of the coupling parameter g with the renormalization scale μ_R ¹⁹ and asymptotic freedom of quarks. In the perturbative expansion necessary for calculating the scattering amplitude the coupling parameter isn’t a fixed constant in QCD and depends on how one absorbs infinities that naturally occur in the theory. The technique of absorbing these infinities is called renormalization and in QCD it depends critically on the scale at which the infinities reappear, denoted Λ . Additionally, there isn’t a single technique for this renormalization so somewhat arbitrary choices (influenced by the underlying physics) must be made. The renormalization scheme common in HEP is the modified minimal subtraction scheme ($\overline{\text{MS}}$) [18] where $\Lambda_{\overline{\text{MS}}} \approx 292 \text{ MeV}$ [17]²⁰. The convention for probing the values of $\alpha_s(\mu_R^2)$ is to enforce $\mu_R \approx Q$ where Q is the scale of the momentum transfer for a given process and is found through experiment [19, 20] to decrease with increasing Q . This is the origin of asymptotic freedom wherein the strong force becomes effectively weaker at high Q [21]. Asymptotic freedom

¹⁸One would be forgiven for thinking that the since atoms are made from nuclei, nuclei from quarks and that quarks get their masses from the Higgs, that the Higgs is responsible for all mass in the universe. One would be mistaken as the masses of the up and down quarks are wildly smaller than the mass of a nucleon (about 1 GeV). This discrepancy originates from that the fact that it’s actually the binding energy between the quarks that gives the nucleons inertia. Thus, it’s actually the strong force which gives objects the vast majority of their masses.

¹⁹ Technically all parameters in perturbative QCD including the quark masses depend on μ_R . However, I’ll ignore this important fact for brevity.

²⁰ This also depends of how many quarks masses are taken into account during the renormalization. The number quoted here assumes the masses of the up, down, strange, and charm quarks are zero (4-flavor scheme).

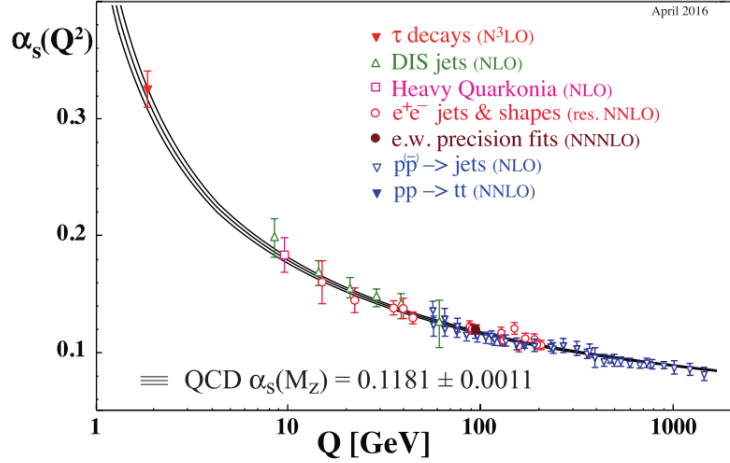


Figure 2.6: Running of the QCD coupling parameter (here denoted α_s) along with several experimental data points. As reference, the value of α_s at M_Z^2 is given as this is common benchmark value. Notice that α_s has an inverse dependence on $\log Q$. This allows for perturbative expansion in the higher Q regions. Figure taken from Ref. [17].

allows for perturbative expansion in the context of high-energy particle colliders. A plot of the coupling parameter versus Q is given in Figure 2.6.

Computing cross sections involving hadrons requires a somewhat more involved approach than a naive application of perturbative QFT would suggest. This is due to the inherently non-perturbative nature of QCD at low energies. A common technique to computing total cross sections in p - p collisions is to construct an expression for the total cross section such as in Eq. 2.5. A so-called inclusive collinear factorization theorem is then applied to this expression. This factorization allows for the decoupling of the universal (i.e. process independent) parton²¹ distribution functions and (perturbatively calculable) partonic cross sections. [22] In this factorization it's necessary to know the parton distribution functions (denoted PDF) to be able to test the perturbative calculations. A PDF ($f_i(x, Q^2)$) is defined as a probability distribution function (denoted pdf, f) of finding a specific parton, i , at a given longitudinal momentum (x) at a given energy scale (Q^2). QCD gives no prediction for the values of the PDFs and they must be determined by fits to experimental data [23]. A potential complication of this model (among many others) is that constituent partons in

²¹A parton is a generic term for the constituents of a hadron (i.e. quarks and gluons).

a hadron are only defined w.r.t. a certain energy scale (this is why PDFs are functions of Q^2). This implies that a single parton at one energy scale can really be multiple partons at a different scale. For instance, the proton is well-modeled by three “valence” partons at low energies, but at higher energies there exists a multitude of “non-valence” partons (parton sea) within the proton that must be taken into account by the PDF. These details on the computation of α_s and of PDFs will prove important in later discussions on Monte Carlo simulation.

2.2.4. Higgs Boson Production Mechanisms

The Higgs can be produced in a virtually infinite number of ways. However, most relevant for this analysis is the production of the Higgs boson through proton-proton collisions (p - p). Eschewing a pedagogical treatment of computing cross sections from p - p interactions (a rich subject far beyond the scope of this thesis) I will simply assert the leading production mechanisms at a p - p collider with a center-of-mass energy of 13 TeV. The phrase “center-of-mass energy” refers to the energy of the colliding bodies as measured in the center of mass reference frame (i.e. the frame in which the 3-momentum of the system sums to zero). The four major production modes in this type of collider are (in decreasing order of cross section): gluon-gluon fusion (ggF), vector-boson fusion (VBF), vector boson-associated production (VH), and $t\bar{t}/b\bar{b}$ -associated production (ttH/bbH) [24]. The leading-order Feynman diagrams for these processes are shown in Figures 2.7a, 2.7b, 2.8, and 2.9 and the values of the total cross sections for a $m_H = 125$ GeV Higgs boson are given in Table 2.1. At a collider of



Figure 2.7: Sketches of two LO Feynman diagrams for Higgs production (more information is needed for these to be useful for computation). (a) Gluon-gluon fusion via a quark loop. (b) Vector boson fusion.

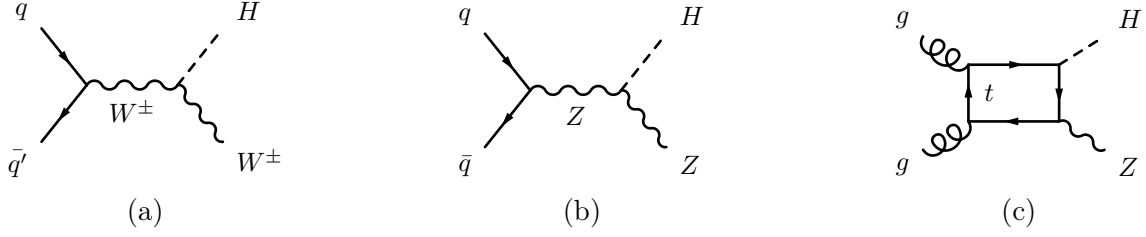


Figure 2.8: Sketches of the LO Feynman diagrams for vector boson-associated Higgs production. (a) $W^\pm$ boson-associated production of the Higgs. (b) Z boson-associated production of the Higgs. (c) Z boson-associated production of the Higgs through a t -quark loop.



Figure 2.9: Sketches of two LO Feynman diagrams for $t\bar{t}/b\bar{b}$ -associated Higgs production. (a) Production through a t/b -quark radiating a Higgs boson. (b) Production from gluons through exchange of a t/b -quark.

	ggF	VBF	VH	ttH	bbH
Total Cross Section (pb)	49	3.8	2.3	0.51	0.49

Table 2.1: Higgs ($m_H = 125$ GeV) production cross sections for a p - p collider at a center-of-mass energy of 13 TeV [24]

this type gluon-gluon interactions are the most probable as evident in the fact that ggF dominates the production cross-sections. The analyses in this thesis search for the Higgs produced via VH. Despite the lower cross section w.r.t. ggF and VBF, VH gives these analyses an additional vector boson to identify which then aids in locating the Higgs boson while suppressing the copious background processes.

2.2.5. Higgs Boson Decay Channels

The Higgs boson couples to all but the gluon and photon fields in the SM and, thus, can decay to any pair of these particles at leading order. Of course, some of these decay channels can be highly suppressed and the Higgs-to-gluons and -photons channels are possible at higher order as will be shown. Neglecting the Higgs self-coupling, as this is a higher-order effect and suppressed w.r.t. LO processes, the Higgs will decay to either a pair of fermions (Figure 2.10a), an on- and off-shell pair of weak vector bosons (Figure 2.10b), a pair of photons (Figure 2.11), or gluons (Figure 2.12). Decays to massless gauge bosons are facilitated through massive particles as shown in the figures. The following give the general relationship between the decay widths and the SM parameters. The decay width for $H \rightarrow f\bar{f}$ is related to the fermion and Higgs masses like

$$\Gamma(H \rightarrow f\bar{f}) \propto \frac{N_C}{v^2} m_f^2 m_H,$$



Figure 2.10: Higgs fermionic and weak boson decay channels. (a) Higgs decaying to a fermion and an anti-fermion. (b) Higgs decaying to on- and off-shell weak vector bosons. The reason both vector bosons are not on-shell is due to the constraint of the known mass of the Higgs not being large enough to decay to two on-shell weak vector bosons.

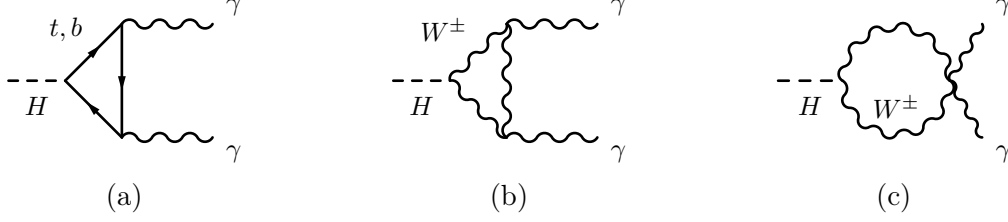


Figure 2.11: Higgs photonic decay channels. (a) Higgs decaying via a quark loop. (b) & (c) Higgs decaying via $W^\pm$ loops.

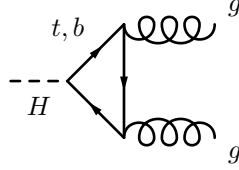


Figure 2.12: Higgs decaying to gluons via a quark loop.

where $N_C = 3$ is a “color” factor only used for quarks ($N_C = 1$ for leptons) to account for the $SU(3)$ internal symmetry of QCD. The decay width for the on- and off-shell weak vector boson is related by

$$\Gamma(H \rightarrow VV^*) \propto \frac{M_V^4}{v^4} m_H b_V,$$

where $b_{W^\pm} = 1$ and $b_Z \propto (\sin \theta_W)^4$. Both the decays to photons and gluons are related by

$$\Gamma(H \rightarrow \gamma\gamma/gg) \propto \frac{m_H^3}{v^2}.$$

Figure 2.13 shows a complete summary for all leading decay channels of the Higgs boson. The figure shows so-called “branching ratios” rather than decay rates. Branching ratios (BR) are easily computed by taking the decay width for a desired channel (i.e. $H \rightarrow b\bar{b}$) and dividing by the total decay width (inverse of the total proper lifetime of the particle). The analyses in this thesis use the $b\bar{b}$ decay of a SM Higgs boson at 125 GeV. The reason for choosing this decay channel is twofold: this channel has the largest predicted BR at $57.7^{+3.2}_{-3.3}\%$ ²² with $m_H = 125$ GeV and this channel has yet to be directly observed. Searching

²²This BR is computed using NLO electroweak corrections. The uncertainties arise due to missing, higher-order terms and experimental error on the SM model parameters [24].

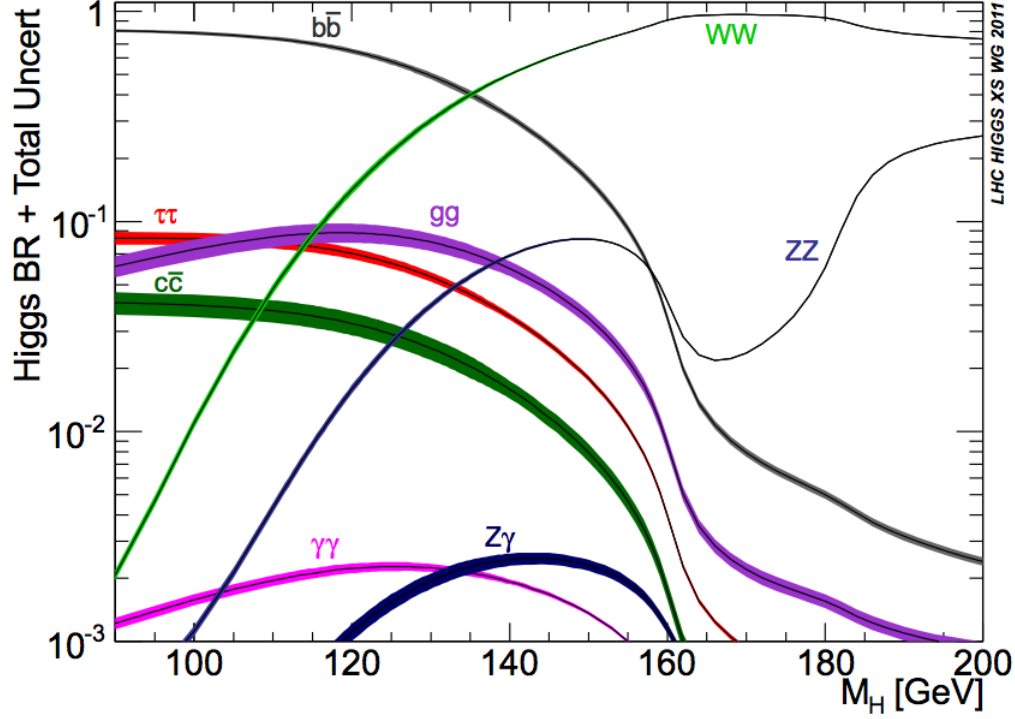


Figure 2.13: Leading Higgs decay channels [24].

for this decay channel is an opportunity to either make a new SM observation, probe physics beyond the SM (described later), or both.

2.2.6. Standard Model Executive Summary

The SM gives particle physicists an extremely expressive and accurate theory by which they can explore the fundamental nature of reality. It's most compactly stated as a Lagrangian density of fields wherein its Lorentz and non-Abelian $SU(3) \times SU(2) \times U(1)$ symmetries are manifest. Its electroweak vacuum state is degenerate and breaks the $SU(2) \times U(1)_Y$ symmetry in order to give mass to the fermions and gauge bosons through the Higgs field. Generally speaking, bosons are mediators of forces and have integer spin ($s = 1$ for all but the Higgs boson, which has $s = 0$), while the fermions are matter particles that have half-integer spin and have three different generations. Leptons are fermions that don't couple to the gluon field and have weak hypercharge since they couple to the $SU(2)$ portion of the SM. Quarks are fermions with electrical charges in multiples of $\frac{1}{3}e$. Since quarks couple to the

gluon field they are virtually always found in bound states known as hadrons²³. Mesons are hadrons with integer spin while baryons are hadrons with half-integer spin. The common proton and neutron are half-integer spin bound states of up-up-down and up-down-down quarks, respectively. While the SM can also be used to calculate quantum properties of hadrons, these details are omitted in this thesis for brevity. In addition to the symmetries already discussed in this thesis, there exist additional symmetries that separately conserve electron-, muon-, and tau-lepton number, along with a so-called “baryon number²⁴.” The computation of cross sections and decay widths can be done with Feynman diagrams and scattering amplitudes. The scattering amplitude is computed using a perturbative expansion that includes the first few non-zero terms in the electroweak and QCD portions of the SM. However, not every aspect of the SM is computed to the same order in the perturbative series (i.e. a particular process may be compute to NLO in the $SU(2) \times U(1)_Y$ electroweak part of the model while NNLO $SU(3)$ QCD “corrections” may be added for additional accuracy). In total, there are 19 free parameters assuming massless neutrinos.

2.3. The Two-Higgs Doublet Model Extension to the SM

In addition to the single, complex, scalar $SU(2) \times U(1)_Y$ doublet that is present in the SM, another doublet of this type may be introduced into the model. This two-Higgs doublet model (2HDM) is virtually identical to the SM save for this extra doublet. The motivation for adding this extra doublet comes from supersymmetric theory (needed to give masses to the quarks and anti-quarks as a single Higgs doublet cannot couple to its own complex conjugate in the Lagrangian), axion theory (low-energy approximation to the additional axion field at high energies), and baryogenesis (origin of the asymmetry of the number of baryons to anti-baryons observed in nature) [25]. A generic 2HDM model has numerous parameters and is beyond the scope of this thesis. In the specific type of 2HDM models explored here, the Higgs doublets are CP invariant, CP isn’t broken in the vacuum state, and discrete symmetries

²³A notable exception is the top quark as this particle is expected to have too short a lifetime (owing to its large mass) to form a bound state before decaying. This is backed by the lack of experimental evidence for top-quark bound states.

²⁴Quarks are assigned a baryon number of $\frac{1}{3}$ and anti-quarks have a baryon number of $-\frac{1}{3}$.

discussed later prevent flavor-changing neutral currents²⁵ (an observational fact). These restrictions give a potential for two complex, scalar doublets Φ_1, Φ_2 of the form

$$V = m_{11}^2 \Phi_1^\dagger \Phi_1 + m_{22}^2 \Phi_2^\dagger \Phi_2 - m_{12}^2 (\Phi_1^\dagger \Phi_2 + \Phi_2^\dagger \Phi_1) + \frac{\lambda_1}{2} (\Phi_1^\dagger \Phi_1)^2 + \frac{\lambda_2}{2} (\Phi_2^\dagger \Phi_2)^2 + \lambda_3 \Phi_1^\dagger \Phi_1 \Phi_2^\dagger \Phi_2 + \lambda_4 \Phi_1^\dagger \Phi_2 \Phi_2^\dagger \Phi_1 + \frac{\lambda_5}{2} \left[(\Phi_1^\dagger \Phi_2)^2 + (\Phi_2^\dagger \Phi_1)^2 \right],$$

where all parameters are real numbers. Similar to the SM Higgs doublet, these fields have vacuum states of

$$[\Phi_a]_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v_a \end{pmatrix}, a = 1, 2.$$

Notice that there are now two vacuum expectation values. This is a central feature of the theory with implications for related experimental searches. A common representation of these values is their ratio $\tan \beta = \frac{v_2}{v_1}$. The two doublets have eight fields and a generic form of

$$\Phi_a = \begin{pmatrix} \phi_a^+ \\ (\nu_a + \rho_a + i\eta_a)/\sqrt{2} \end{pmatrix}, a = 1, 2.$$

As in the SM, three of these fields are absorbed by the gauge freedom of the weak bosons in the covariant derivative. Thus, five physically realizable fields remain. There are two charged scalars with the same mass, two CP-even, neutral scalars of different mass, and a CP-odd pseudo-scalar. The mass term in the Lagrangian for the pseudo-scalar (denoted A) is

$$\frac{m_A^2}{\nu_1^2 + \nu_2^2} \begin{pmatrix} \eta_1 & \eta_2 \end{pmatrix} \begin{pmatrix} v_2^2 & -v_1 v_2 \\ -v_1 v_2 & v_1^2 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix}.$$

²⁵Flavor changing neutral currents refer to interactions where the flavor of a fermion changes with no change to its electric charge.

Relevant for later discussion is the mass term for the two scalar fields

$$-\begin{pmatrix} \rho_1 & \rho_2 \end{pmatrix} \begin{pmatrix} m_{12}^2 \frac{v_2}{v_1} + \lambda_1 v_1^2 & -m_{12}^2 + \lambda_{345} v_1 v_2 \\ -m_{12}^2 + \lambda_{345} v_1 v_2 & m_{12}^2 \frac{v_1}{v_2} + \lambda_2 v_2^2 \end{pmatrix} \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix},$$

where $\lambda_{345} = \lambda_3 + \lambda_4 + \lambda_5$. The above matrix can be diagonalized and the angle of rotation that accomplishes this is known as α . The angle β defined earlier diagonalizes the matrices for the pseudo-scalar and charged scalars. Together, these two angles determine several aspects of the theory including how the scalar fields couple to the weak bosons and fermions. For instance, the relationship of the masses of the light uncharged scalar h , heavy uncharged scalar H , and pseudo-scalar A to that of the neutral weak boson Z is [26] (at LO)

$$\cos^2(\beta - \alpha) = \frac{M_h^2(M_Z^2 - M_h^2)}{M_A^2(M_H^2 - M_h^2)}.$$

Using the fact that flavor-changing neutral currents have never been observed along with assumptions about the nature of the vacuum mentioned earlier, there exist two, popular 2HDM theories known as type I and type II. In the type I 2HDM theory quarks only couple to one of the $SU(2)$ Higgs doublets (Φ_2). This is achieved through the discrete symmetry requirement that $\Phi_1 \rightarrow -\Phi_1$. In the type II model all up-type quarks (u, c, t) couple to one of the doublets (Φ_2) while the down-type quarks (d, s, b) couple to the other doublet (Φ_1). This is enforced by requiring $\Phi_1 \rightarrow -\Phi_1$ and that the down-type quarks behave like $d_R^i = -d_R^i$, where d_R^i is the right-handed spinor field of the down-type of quark ($SU(2)$ singlet) with color index i . The fact that these Higgs doublets only act on a subset of quarks prevents flavor-changing neutral currents. In both theories the physical pseudo-scalar field is written as $A = \eta_1 \sin \beta - \eta_2 \cos \beta$ while the lighter scalar field is $h = \rho_1 \sin \alpha - \rho_2 \cos \alpha$. The SM Higgs can be similarly written $H^{\text{SM}} = h \sin(\alpha - \beta) - H \cos(\alpha - \beta)$. With this in mind the so-called alignment limit is achieved by letting $\cos(\alpha - \beta) \rightarrow 0$ so that the lighter uncharged scalar is identical to the SM Higgs. Due to requirements on the top coupling to Higgs an empirical restriction of $\tan \beta > 0.3$ is placed on the explorable phase-space for the

analyses in this thesis [25]. In total there are seven parameters in the type I & II theories: m_h , m_H , m_A , $m_{H^\pm}$, β , α , and m_{12} .

Among the most prominent decay channels for the A boson is $A \rightarrow Zh$ [26, 27]. This decay of the A boson will have very similar final states to that of the SM ZH associated production with the $Z - H$ system now forming the resonant A -boson mass. In fact, there exists so much similarity between the SM and 2HDM at LO (in electroweak theory) that the couplings of the 2HDM A - and h -boson are usually given in term of the equivalent couplings of the SM Higgs boson (a fact that will be abbreviated in the following discussion as simply “times the SM Higgs boson coupling”). The production modes of the A boson that are explored in this thesis are the gluon-gluon fusion and $b\bar{b}$ -associated production. Diagrams for these processes are given in Figure 2.14. Using these two production mechanisms allows for analyses searching for an A boson to explore the $\tan\beta$ parameter space. This is due to the fact that in the type I model the couplings of the A boson to the t - and b -quarks are both proportional to $\cot\beta$ (the t -quark coupling plays a role in the ggF production channel through the quark loop). However, in the type-II model the A boson couples to the t -quark proportionally to $\cot\beta$ while it couples as $\tan\beta$ to the b -quark. Furthermore, the mixing-angle dependence of the coupling of the h boson to down-type quarks in type-II 2HDM is $\sqrt{1 - (\cos^2(\beta - \alpha))} - \tan(\beta) \cos(\beta - \alpha)$ while the AZh vertex is proportional to $\cos(\beta - \alpha)$ regardless of the 2HDM type (this makes sense since the models only differ in the fermion couplings) [28]. To illustrate this point, a 2-D plot in the $\tan(\beta)$ - $\cos(\beta - \alpha)$ plane is given in



Figure 2.14: 2HDM A boson production mechanisms. (a) ggF production of the A boson through a t -quark loop. (b) $b\bar{b}$ -associated production of the A boson through the exchange of a b -quark between gluons.

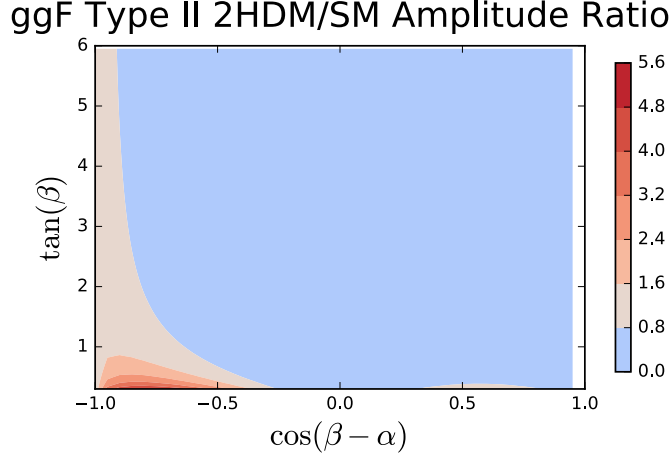


Figure 2.15: Ratio of scattering amplitudes (at LO) between a type II 2HDM vs. SM for the $gg \rightarrow A \rightarrow Zh, h \rightarrow b\bar{b}$ process (discussion in the text).

Figure 2.15 that shows the multiplicative factor needed to modify the SM in order to compute the scattering amplitude for the $gg \rightarrow A \rightarrow Zh, h \rightarrow b\bar{b}$ process (assuming the top-quark loop dominates ggF production). This shows that for the ggF production of the A boson an enhancement over the SM prediction is greatest at the low and high $\cos(\beta - \alpha)$ when $\tan(\beta)$ is low while there exists a deficit virtually everywhere else. Thus, one may naively think that observation would be more difficult in the bluer regions of this plot (remember, this is LO only).

2.4. Previous Higgs Measurements

After decades of searching in several experimental thrusts [29, 30], the discovery of the Higgs boson was announced on July 4th, 2012 at CERN using data from two independent detectors [31, 32]. The three decay channels for the initial discovery are the $H \rightarrow ZZ^* \rightarrow 4l$ [33, 34], $H \rightarrow \gamma\gamma$ [35, 36], and $H \rightarrow WW^* \rightarrow \nu e \nu \mu$ [37, 38] (here, V^* =off-shell vector boson, l =lepton, and γ =photon). Since then the Higgs mass has been more precisely measured to be $m_H = 125.09 \pm 0.21(\text{stat.}) \pm 0.11(\text{syst.})\text{GeV}$ and the $H \rightarrow \tau\bar{\tau}$ channel observed [39] (the sources of the error on the mass measurement are found in this reference). The $H \rightarrow \tau\bar{\tau}$ decay is markedly different than the initial three decay channels in that it explicitly involves

the Yukawa coupling of the Higgs to a type of fermion (τ -lepton). The analyses in this thesis focus on the yet-to-be-directly-observed $H \rightarrow b\bar{b}$ decay. This observation is important for two primary reasons. First, a direct observation of the Higgs Yukawa coupling to quark fermions consistent with the SM has yet to be made. Secondly, the Higgs Yukawa coupling to quark fermions is sensitive to beyond-the-SM (BSM) phenomena including the 2HDM discussed in this thesis.

As a point of reference, the best measurements for $H \rightarrow b\bar{b}$ and exclusion limits on the 2HDM A boson prior to the work in this thesis are given. The most comprehensive summary of the state of SM measurement of the $H \rightarrow b\bar{b}$ decay channel is given in Ref. [39]. A summary of all the production modes and decay channels analyzed during the Run 1 period of the operation of the LHC collider (discussed in the next chapter) is given in Figure 2.16. From this figure it is clear we, as a community, haven't reached the level of discovery with the vector-boson associated production of the Higgs that then decays to $b\bar{b}$. The precise number for the $WH, H \rightarrow b\bar{b}$ mode is $1.0^{+0.5}_{-0.5}$, and for the ZH mode it is $0.4^{+0.4}_{-0.4}$, where the central value is unity in the case of perfect agreement with the SM and the nature of the errors are described in the aforementioned reference. This measure of agreement with the SM described in the previous sentence is known as the signal strength (μ). It is defined as the ratio of the observed or expected cross section times branching ratio versus the SM expectation,

$$\mu = \frac{(\text{branching ratio} \times \sigma)_{\text{exp/obs}}}{(\text{branching ratio} \times \sigma)_{\text{SM}}}, \quad (2.7)$$

where the numerator is the expected or observed cross section multiplied by the branching ratio and the denominator is the SM expectation of this quantity. The previous results for the 2HDM A -boson search are given in Figures 2.17 and 2.18. These results are interpreted in terms of exclusion regions in the two-dimensional, $\tan(\beta)$ - $\cos(\beta - \alpha)$ plane that characterizes the 2HDM theory as discussed earlier. Construction (and, thus, interpretation) of these plots is as follows:

- Exclusion limits (detailed in Chapter 8) on the cross section times branching ratio for

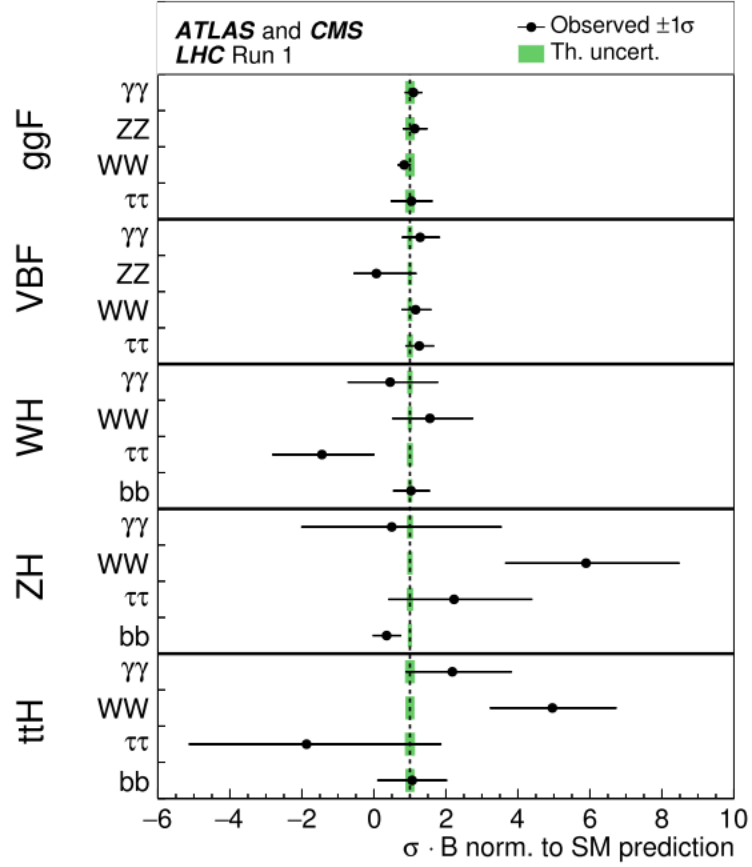
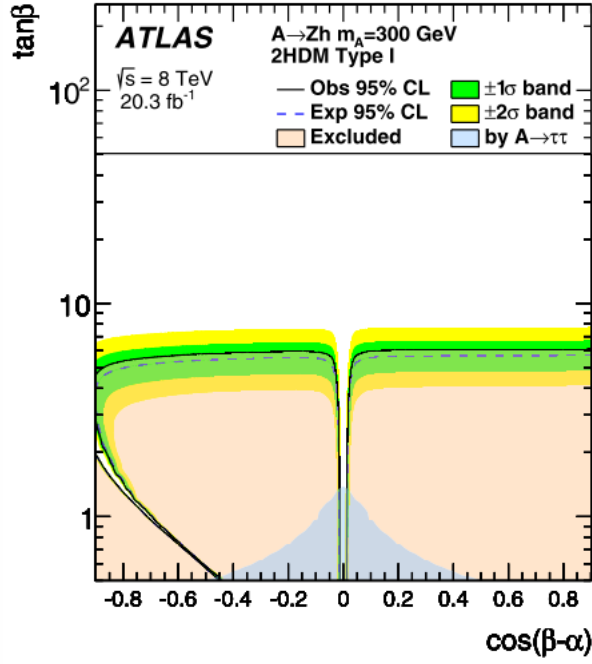
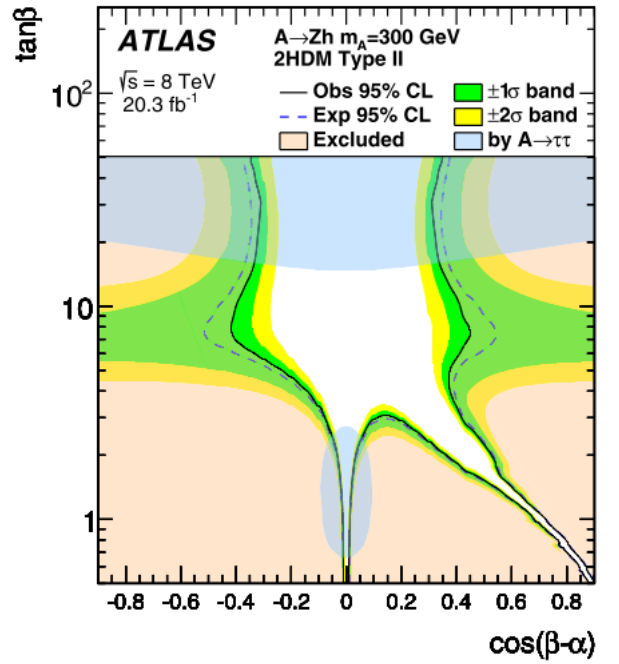


Figure 2.16: LHC Run 1 Higgs couplings results. The production modes of the Higgs are given on the far left, while the decay channels are given in the boxes on the left. Here, the measured cross sections times branching ratios are given relative to their SM expectation such that a value of unity represents perfect agreement with the SM. Further description can be found in Ref. [39].



(a)



(b)

Figure 2.17: ATLAS 2HDM LHC Run 1 results for an A boson mass of 300 GeV. Both $Z \rightarrow \nu\bar{\nu}$ and $Z \rightarrow l\bar{l}$ decays of the Z are accounted for and the Higgs decays to a pair of b -quarks. For reference, similar results where the Higgs decays to τ -leptons is shown in light blue. (a) Type I (b) Type II

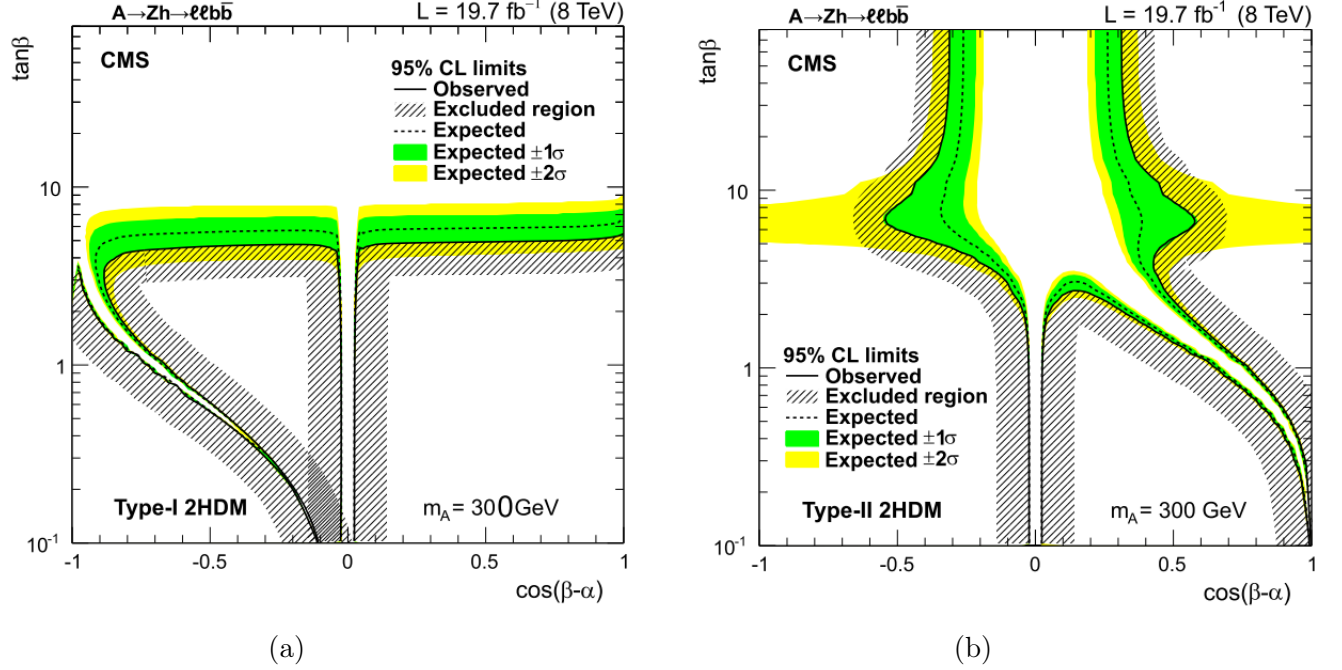


Figure 2.18: CMS 2HDM LHC Run 1 results for an A boson mass of 300 GeV. Only the $Z \rightarrow \ell\bar{\ell}$ decay of the Z is accounted for and the Higgs decays to a pair of b -quarks. (a) Type I (b) Type II

various A -boson mass hypotheses are determined for both the ggF and $b\bar{b}$ -associated production from experiment.

- Exclusion limits are also derived after the empirical mass distributions of the A -boson signal are convolved with Breit-Wigner distributions at varying widths (thus, altering the signal cross section and exploring regions of $\cos(\beta - \alpha)$). The relative fractions of the $b\bar{b}$ -associated production signal to ggF are also varied in order to explore regions of $\tan(\beta)$.
- With handles for the exclusion limits with regards to both the $\tan(\beta)$ and $\cos(\beta - \alpha)$ 2HDM parameters, the observed and expected exclusion limits (per mass hypothesis) are translated onto the 2-D $\tan(\beta)$ - $\cos(\beta - \alpha)$ plane after correcting for unequal signal acceptances in the $b\bar{b}$ -associated and ggF production mechanisms.
- Finally, a region is excluded if the exclusion limit on the cross section times branching ratio is below the rate predicted by theory.

2.5. Particle Colliders

Observations of the Higgs boson and searches for the A boson require the use of high-energy particle colliders. These machines accelerate subatomic particles (usually ions, electrons, neutrons, or protons) to speeds close to that of light and steer them in such a manner as to cause collisions with either a fixed target or other high-velocity particles. The reason this is necessary is related to the fundamental physics of object resolution and quantum fields. Recall the de Broglie wavelength of a particle $\lambda = \frac{h}{|p|}$ that relates the wavelength of a particle to its momentum (magnitude) and effectively says that the more energetic (faster) a particle, the smaller its wavelength. In order to resolve an object at everyday length scales (about a $100\text{ }\mu\text{m}$) one must collide particles of smaller wavelength (photons) and then have some way of detecting these particles (eyes). However, in particle colliders we wish to collide particles with wavelengths on the order of 1 fm (i.e. the scale of the constituent quarks and gluons of a proton [40]). Additionally, we desire that the particles we collide have enough energy to create more exotic particles than those we started with (i.e. gluons interacting to produce a Higgs boson). Thus, we need a collider to accelerate particles to high velocities to first be able to “see” each other in order to interact and secondly have enough energy during the interaction to make these exotic particles kinematically accessible (as determined through the scattering amplitude). The current states-of-the-art for particle accelerators and detectors is described in the next chapter.

Chapter 3

The LHC and the ATLAS Detector

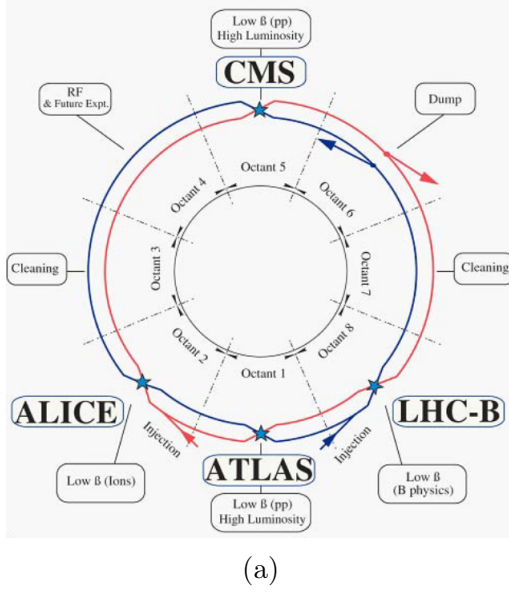
This chapter gives a brief overview of the LHC machine along with descriptions of the ATLAS detector’s components. The LHC demands particle detectors capable of operating in unprecedented center-of-mass collision energies and luminosities. The ATLAS detector is specially designed for this environment with sub-systems custom built for tracking, calorimetry, and identification. My contributions to the ATLAS detector’s trigger system (described in Section 3.2.6) are detailed in Appendix A. Unless otherwise stated, all material in this chapter including figures and tables is taken from Ref. [41].

3.1. LHC

The LHC [42] is a circular particle accelerator designed to deliver luminosities¹ of up to $\mathcal{L} = 10^{34} \text{ cm}^{-2} \text{ s}^{-1} = 0.036 \text{ fb}^{-1} \text{ h}^{-1}$ for proton-proton collisions (denoted $p\text{-}p$)² and $\mathcal{L} = 10^{27} \text{ cm}^{-2} \text{ s}^{-1}$ for lead ion collisions with $p\text{-}p$ center-of-mass energies reaching $\sqrt{s} = 14 \text{ TeV}$. Each proton beam can nominally hold up to 2808 packets of protons (called “bunches,” each of which contain about 1.1×10^{11} protons) with a temporal spacing of about 25 ns between consecutive bunches. The layout of the LHC is given in Figure 3.1a which shows the two circulating beams along with the two general purpose experiments at Points 1 & 5 (ATLAS and CMS respectively). Each octant contains both straight and curved portions of accelerator with detector experiments occupying four of the available eight straight segments. Figure 3.1b is an overview of the accelerators that are used to boost the protons in several high intensity bunches for use in the LHC. Protons are supplied through the following chain: Linac2 $\rightarrow$ ProtonSynchrotron(PS) $\rightarrow$ SuperProtonSynchrotron(SPS).

¹Denoting instantaneous luminosity as $\mathcal{L}$ represents an unfortunate reuse of symbols as this is identical to the Lagrangian density from Chapter 2. Hopefully the context makes it clear which meaning is intended.

²Luminosities and cross sections are frequently expressed in terms of units of barns, where $1 \text{ b} = 10^{-24} \text{ cm}^2$.



The LHC injection complex

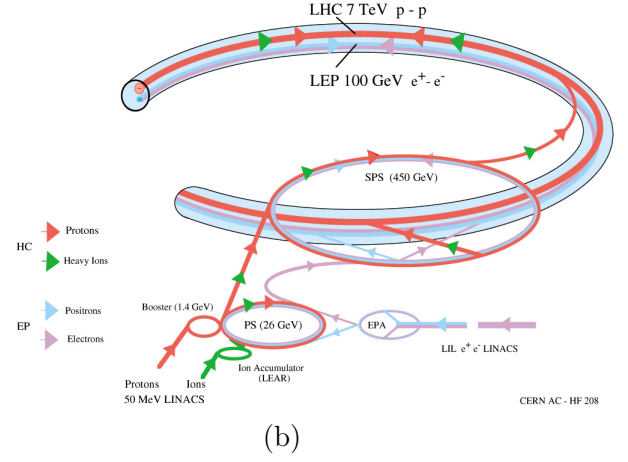


Figure 3.1: LHC overview. (a) Schematic layout of the LHC (beam 1 travels clockwise, beam 2 travels anti-clockwise). (b) Accelerators that feed the LHC.

The LHC occupies the 26.7 km of tunnels previously used in the LEP electron-positron collider experiment. This has practical implications in that the curvature of portions of the tunnel is optimized for synchrotron radiative loss for electrons which is somewhat conservative for hadron acceleration. Also, the design choice for a twin-bore magnet system (i.e. two proton beams sharing one mechanical housing and cryostat) is a consequence of the limited tunnel diameter. Together, these and other comprises place limitations on peak luminosities and collision energies.

The operation schedule for the LHC involves several “Run” phases with interspersed “Long Shutdown” (LS) phases. The Run phases are designated for p - p and lead ion collision data taking while the LS phases are absent beam collisions and primarily used for installing upgrades and performing repairs. A tentative schedule of the LHC up through the year 2035 is given in Figure 3.2. The analyses presented in this thesis are performed on data taken during Run 2 (p - p collisions at $\sqrt{s} = 13$ TeV). Some technical work in this thesis was performed during the LS1 period between Runs 1 & 2.

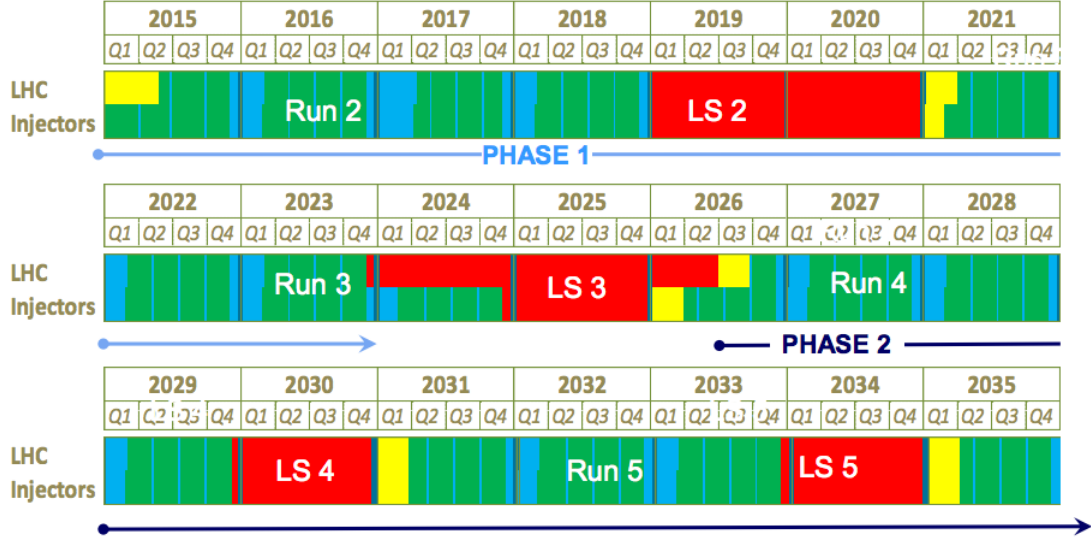


Figure 3.2: Tentative long-term LHC schedule. Phase 2 will bring a drastic increase in luminosity w.r.t. Phase 1. Section colors: green=physics Run, blue=technical stop, red=Long Shutdown, yellow=beam commissioning

The physics goals of the LHC are expressly to search for physics beyond the Standard Model (referred to as BSM physics). During the conception of the LHC the Standard Model was nearing experimental verification, the top quark and Higgs boson being the two missing pieces. It was clear that even with their discovery that our understanding of particle physics would be unsatisfactory and that accelerators capable of unprecedented center-of-mass energies and luminosities would be needed to probe the rare particle decay signatures indicative of BSM physics. Since, the predicted cross-sections for Higgs and BSM physics are less than 1 pb and the cross-section for p - p inelastic collisions at $\sqrt{s} = 13$ TeV is measured to be 79.3 ± 2.9 mb [43] the signal-to-background ratio w.r.t. these inelastic collisions is quite small. This necessitates general-purpose detector experiments sensitive to a wide range of particle decays and interactions.

3.2. ATLAS

ATLAS is one of two general-purpose detectors at the LHC. It is designed to operate in both $\mathcal{L} = 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ p - p and $\mathcal{L} = 10^{27} \text{ cm}^{-2} \text{ s}^{-1}$ lead-lead collision environments³. The ATLAS experiment is comprised of several sub-detector components centered around an interaction region each of which is specialized for various aspects of particle detection such as calorimetry, muon detection, tracking, and triggering. This layout is comparable to an onion, where each “layer” represents a different sub-detector as seen in Figure 3.3. The required resolution and coverage of the sub-detector components are given in Table 3.1.

The use of p - p collisions at the LHC drives many of the technical specifications of the ATLAS detector. Our ignorance of the longitudinal component of the momentum of the colliding partons within the protons dictates the detector have cylindrical symmetry rather than the spherical symmetry preferred at, for instance, lepton-lepton colliders. Closely related to this is the fact that charged-particle multiplicities are approximately uniform in pseudorapidity (η) within the central regions of hadron colliders (Figure 3.4). This fact has profound design implications for the physical layout of the detector components.

Our event topologies take full advantage of the many sub-detectors of the ATLAS experiment. Final states with electrons necessitate tracking and calorimetric systems capable of avoiding the misidentification of high energy photons and hadronic jets as electrons. The presence of muons requires dedicated tracking and magnet systems that accurately reconstruct the momentum of muons for both event reconstruction and trigger decisions. Hadronic jets need additional calorimetry beyond what is needed for electrons and photons. Finally, since neutrinos interact weakly with matter, their presence is only noticed through an imbalance of both transverse momentum (p_T) and energy deposits in the calorimeters. Thus, accurate reconstruction of neutrinos requires the use of all detector sub-systems.

The following paragraph defines common terms and concepts useful in understanding the layout of the detector and particle physics in general. The standard ATLAS right-handed

³This thesis only covers analyses using data from p - p collisions. Thus, ATLAS’ role in lead-lead collisions is ignored in the rest of this document.

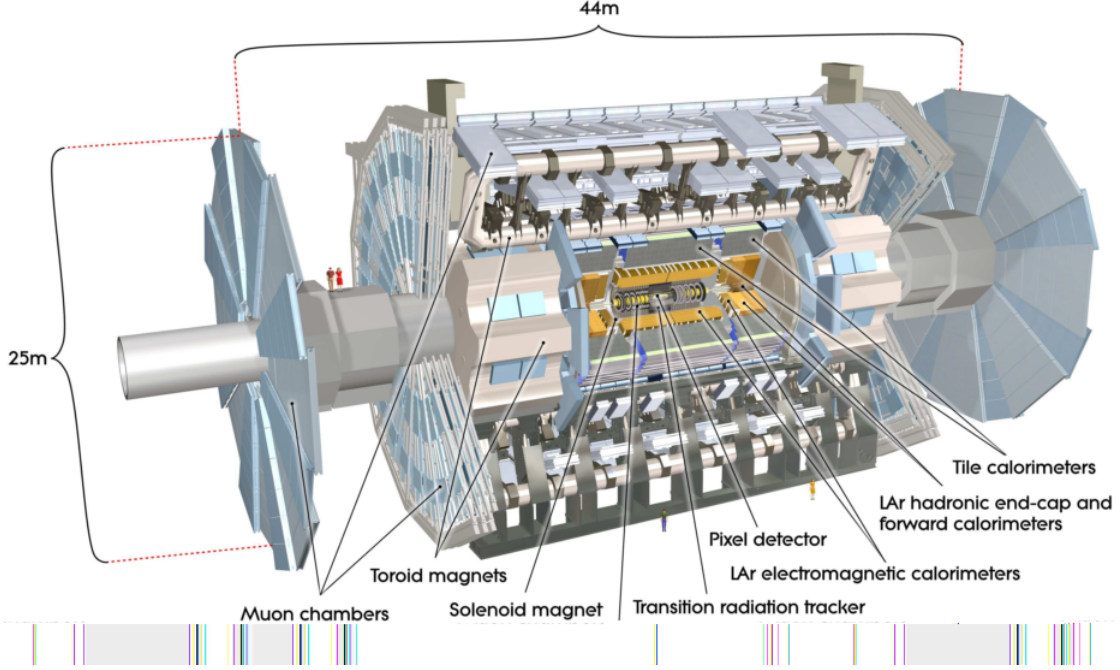


Figure 3.3: Cutaway view of the ATLAS detector along with several sub-detectors.

coordinates are defined by the positive z -axis directed along the beam pipe pointing eastward (the “A” side of the detector cavern), the positive x -axis pointing toward the center of the LHC ring (northward), and the positive y -axis pointing skyward. The adjective “longitudinal” (“transverse”) refers to a quantity that exists solely along (in the plane orthogonal to) the direction of motion (z -axis (x - y plane) w.r.t. the ATLAS detector). The azimuthal angle ϕ is defined in the usual way, and the z -axis is the reference for the polar angle θ with $\theta = 0$ for vectors $(0, 0, \pm 1)$. Rapidity is defined as $y = \frac{1}{2} \ln \frac{E+p_z}{E-p_z}$, where E and p_z are the energy and z -component of momentum of a particle, respectively. When the mass of a particle is negligible compared to its momentum, pseudorapidity defined as $\eta = -\ln(\tan \theta/2)$ can be used in place of rapidity. ΔR is defined as the distance in η - ϕ space, $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$. The volume in 3-space where the two proton beams collide is called the interaction region. It’s nominally described by a 3-D Gaussian interaction probability profile with a standard deviation of 5.6 cm along the beam axis, 15 μm in the transverse plane, and an origin located at the geometric center of the detector [45]. During a crossing of the proton beam bunches

Detector component	Required resolution	η coverage	
		Measurement	Trigger
Tracking	$\sigma_{p_T}/p_T = 0.5\% \ p_T \oplus 1\%$	± 2.5	
EM calorimetry	$\sigma_E/E = 10\%/\sqrt{E} \oplus 0.7\%$	± 3.2	± 2.5
Hadronic calorimetry (jets)			
barrel and end-cap	$\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$	± 3.2	± 3.2
forward	$\sigma_E/E = 100\%/\sqrt{E} \oplus 10\%$	$3.1 < \eta < 4.9$	$3.1 < \eta < 4.9$
Muon spectrometer	$\sigma_{p_T}/p_T = 10\%$ at $p_T = 10$ TeV	± 2.7	± 2.4

Table 3.1: General performance goals of the ATLAS detector. Note that, for high- p_T muons, the muon-spectrometer performance is independent of the inner-detector system. The units for E and p_T are in GeV. The $\oplus$ symbol means the sum in quadrature of the two components.

(called a bunch crossing), several protons collide to create cascades of particle interactions and decays within the detector. Vertices are reconstructed for each collision within the interaction region. Interesting physical processes originating at these vertices are discernible by criteria described later in Chapter 4. A bunch crossing is also called an “event.” A single vertex is selected per bunch crossing as the most likely to produce interesting physics. Interactions between the detector and particles originating from vertices other than the selected vertex within the same event are known as “in-time pile-up” effects. Residual effects from previous bunch crossings such as long-lived particle decays, etc. that can interfere with the current event are known as “out-of-time pile-up” effects. The physical units commonly used in the high energy physics are called Natural units. In these units $c = 1$ and $\hbar = 1$ so that length and time have inverse units of energy, momentum and mass have units of energy, and angular momentum, speed, and electric charge are unit-less. These units simplify the form of equations and allow for easier computation during analysis.

This section presents a brief overview of the detector components along with additional context as to how they relate to my signal processes.

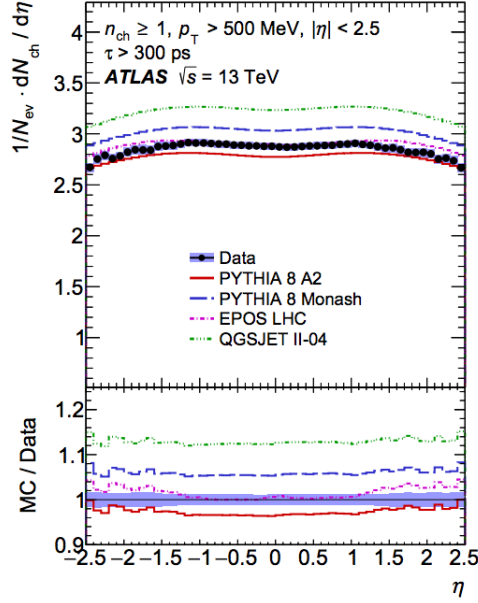


Figure 3.4: Primary-charged-particle multiplicities as a function of η [44]. Primary charged particles are defined as charged particles with a mean lifetime $\tau > 300\text{ps}$, either directly produced in p - p interactions or from subsequent decays of directly produced particles with $\tau < 30\text{ps}$. Here n_{ch} is the number of primary charged particles in an event, N_{ev} is the number of events with $n_{\text{ch}} \geq 1$, and N_{ch} is the total number of primary charged particles in the data sample. Note the uniformity (flatness) of the data points (black dots).

3.2.1. Magnets

The primary purposes of the magnet system are to aid in momentum resolution through the deflection of charged particle trajectories and the determination of the sign of electric charge for particles. It is comprised of three main components: a barrel solenoid, barrel toroid, and end-cap toroids.

3.2.1.1. Barrel Solenoid

The barrel solenoid provides a roughly 2 T magnetic field for the inner detector as shown in Figure 3.5. The solenoid is located just outside the inner detector (Figure 3.3) and shares a cryostat with the electromagnetic calorimeter to minimize radiative loss of particles traveling through the solenoid. It is made from a single winding (1154 turns) of Al-stabilized NbTi conductor with an Al support structure. The solenoid's role is to deflect all charged particles

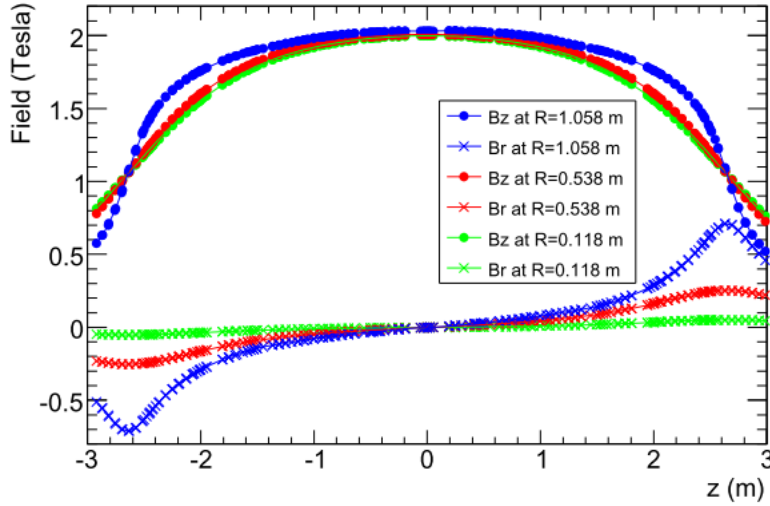


Figure 3.5: Modeling of the radial (B_r) and axial (B_z) components of the magnetic field within the barrel solenoid as a function of axial location at three selected radii ($R = 1.058$ m, $R = 0.538$ m, and $R = 0.118$ m).

incoming to the detector and plays an important part in tracking.

3.2.1.2. Barrel & End-Cap Toroids

The barrel (Figure 3.6) and end-cap (Figure 3.7) toroids are used together to deflect muons within the full range ($|\eta| \leq 2.7$) of the muon spectrometers. The barrel toroid is an air-core toroid located radially beyond the hadronic tile calorimeters (Figure 3.3) and is intermixed with parts of the muon system. The barrel toroid is made from eight coils (120 turns each) of pure Al-stabilized Nb/Ti/Cu conductor with each coil housed in a “race-track” shaped vacuum chamber. Its primary purpose is to deflect muons within the barrel region ($|\eta| \leq 1.4$). The end-cap toroids are located axially beyond the cathode strip chambers of the muon system and are used to deflect muons within the end-cap region ($1.6 \leq |\eta| \leq 2.7$). Each end-cap toroid is made from eight square-coil units (116 turns per coil) along with eight keystone wedges. The conductor material used in them is the same as the barrel toroid. The region in $|\eta|$ ($1.4 < |\eta| < 1.6$) that isn’t explicitly covered by either toroid system has reduced bending power and is known as the transition region. It is properly accounted for in the muon system.

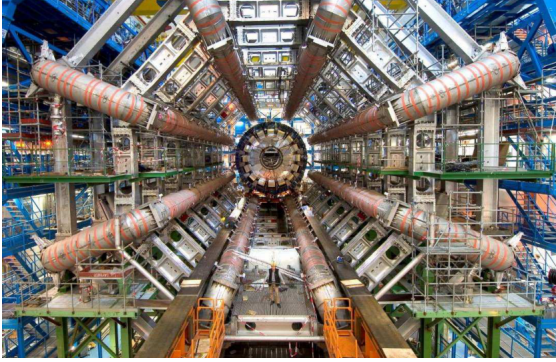


Figure 3.6: View of the barrel toroid before installation of the detector. The eight, metal tube-like structures are the vacuum housings for the coils.

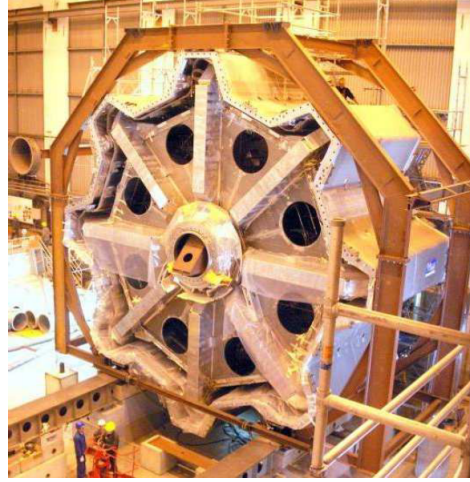


Figure 3.7: One of the two end-cap toroids before installation. The eight flat, square coil units and eight keystone wedges (with the circular holes) are visible.

3.2.2. Inner Detector

ATLAS' Inner Detector (ID) focuses on momentum resolution for charged particles and track and vertex reconstruction through robust pattern recognition. It has three sub-detectors each based on a different technology as shown in Figure 3.8. The ID is fully submersed in the 2 T magnetic field of the barrel solenoid. The innermost layer is known as the Pixel layer and made from densely configured silicon-based pixel sensors. The next layer is the silicon-microstrip tracker (SCT) arranged similarly to the Pixel layer. Together, the Pixel and SCT layers comprise the precision tracking layer (also called the silicon layers). All barrel layers in the precision tracker have sensors that are slightly rotated (pitch) along the z -axis with azimuthal overlap between azimuthal neighbors. Similarly, the end-cap sensors are rotated along the radial axis with azimuthal overlap between azimuthal neighbors. These rotations ensures complete azimuthal coverage with no gaps. Finally, the Transition Radiation Tracker (TRT) is the last of the ID layers and is made from straw tubes filled with xenon gas. The overall η coverage of the ID is $|\eta| < 2.5$ with performance goals of $\sigma_{d_0} = 140 [\mu\text{mGeV}]/p_{\text{T}}[\text{GeV}] \oplus 10 \mu\text{m}$ for the transverse and $\sigma_{z_0 \sin \theta} = 209 [\mu\text{mGeV}]/p_{\text{T}}[\text{GeV}] \oplus 91 \mu\text{m}$ for the longitudinal track impact parameters.

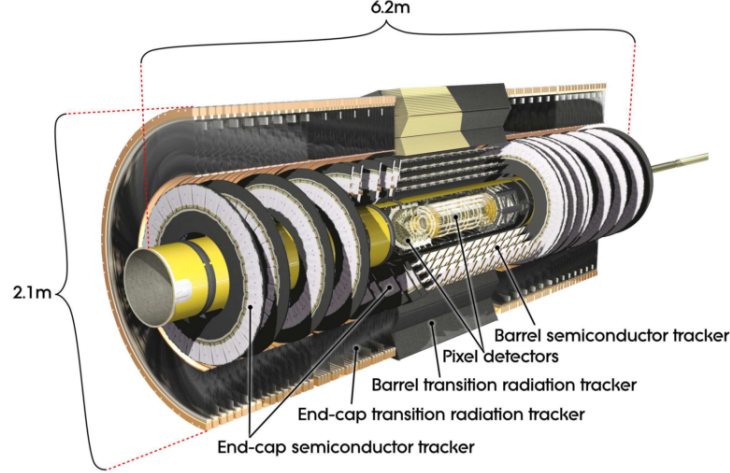


Figure 3.8: Cut-away view of the ATLAS inner detector. All three technologies used in the inner detector have a barrel and end-cap configuration.

3.2.2.1. *Pixel*

The Pixel detector (Figure 3.9) is made from four barrel and six end-cap disks. The innermost barrel layer is called the Insertable B-Layer [46] (IBL) and is a new addition to ATLAS for Run 2. Installation of the IBL required replacing the central portion of the beryllium beampipe in ATLAS for one that has a reduced radius as the existing inner layer was already just outside the beampipe. The radial extent of the beam pipe is 29 mm with the IBL starting at 31 mm. The remaining barrel layers are dubbed the B-layer, Layer 1, and Layer 2 with corresponding end-cap layers making the total number of pixel in the Pixel detector roughly 80.4 million. The intrinsic accuracy of each layer is approximately $10\text{ }\mu\text{m}$ in $R\text{-}\phi$ and $115\text{ }\mu\text{m}$ in z (R) for barrel (end-cap).

3.2.2.2. *SCT*

The SCT is made up of four barrel and 18 end-cap disks of silicon-microstrips. The η range of the SCT is $|\eta| < 2.5$ with the end-cap disks spaced optimally for track hits given the η distribution of charged particles and the configuration of the TRT. The intrinsic accuracy of each layer is approximately $17\text{ }\mu\text{m}$ in $R\text{-}\phi$ and $580\text{ }\mu\text{m}$ in z (R) for barrel (end-cap).

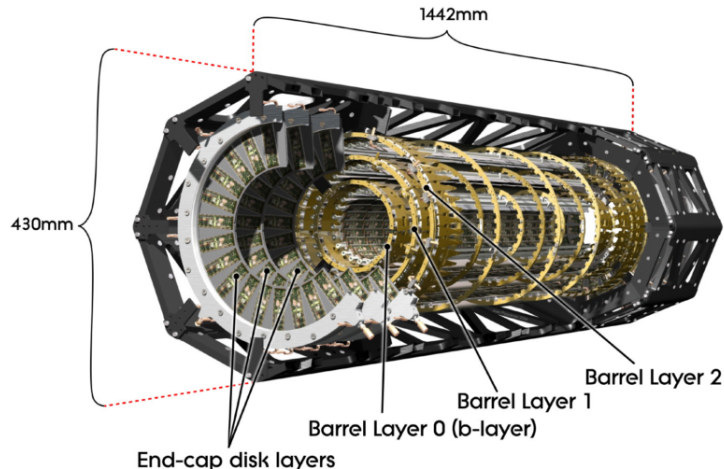


Figure 3.9: Cut-away view of the ATLAS Pixel detector. This figure is absent the new IBL discussed in the text.

3.2.2.3. TRT

The TRT [47] is a tracking system, electron (vs. pion) identification detector, and is the outermost ID component. It is made from proportional drift tubes (straws) for the tracking of all charged particles and transition radiative material for electron identification. The arrangement of straws is such that the straws in the barrel TRT are aligned with the beam pipe while the end-cap straws are perpendicular to the beam pipe. They are non-projective, thus have no resolution along the length of a straw. There are 73 straw planes in the barrel and 160 straw planes in the end-cap regions. The walls of the straws act as the cathodes while the gold-plated tungsten wire along the centers act as the anode. The gas mixture was chosen for good X-ray absorption, increased electron drift velocity, and photon-quenching. The material used to generate the transition radiation is polypropylene either in fiber (in the barrel) or foil (at the end-caps) form. The intrinsic accuracy for tracking is $130\text{ }\mu\text{m}$ in the direction perpendicular to the straws. It is important to note that the TRT only extends to $|\eta|=2.0$ — less than the SCT.

3.2.3. Electromagnetic Calorimeter

The purpose of the electromagnetic (EM) calorimeter is to measure the energies and directions of incident electrons and photons within $|\eta| < 3.2$. The layout of the entire calorimeter system (both EM and hadronic) is shown in Figure 3.10. The EM calorimeter’s core technology is liquid argon (LAr) with lead absorber plates in an accordion geometry (Figure 3.11). The accordion geometry is designed to give full azimuthal coverage and fast readout of the signal from either the front or rear of the electrodes. As with the previous detector systems, the EM calorimeter has both barrel and end-cap components (EMEC & FCal1 are discussed later). The region $1.37 < |\eta| < 1.52$ represents a transition region between the barrel and end-cap cryostats as shown in Figure 3.13. This region along with the fact the EM barrel calorimeter is joined at $\eta = 0$ play important roles during object reconstruction. In each of these components sections evenly spaced in η - ϕ are grouped together to form trigger towers that are used in the L1 Trigger. This allows for faster readout at the L1 level, but with reduced granularity.

Calorimeter Section	Minimum Distance (X_0)	Range
Barrel	22	$ \eta < 1.475$
EMEC	24	$1.375 < \eta < 3.2$

Table 3.2: Table of selected EM calorimeter parameters.

3.2.3.1. Barrel

The barrel EM calorimeter is made from two halves centered at $z = 0$ with a 4 mm gap. It shares a common cryostat chamber with the barrel solenoid magnet as to reduce the material in front of the calorimeters. To reduce punch-through of electrons and photons, the barrel EM calorimeter is designed to have a minimum distance for a particle to travel through measured in radiation lengths (X_0) given in Table 3.2. The geometric configuration of the accordion waves in the barrel “are axial and run in ϕ , and the folding angles of the

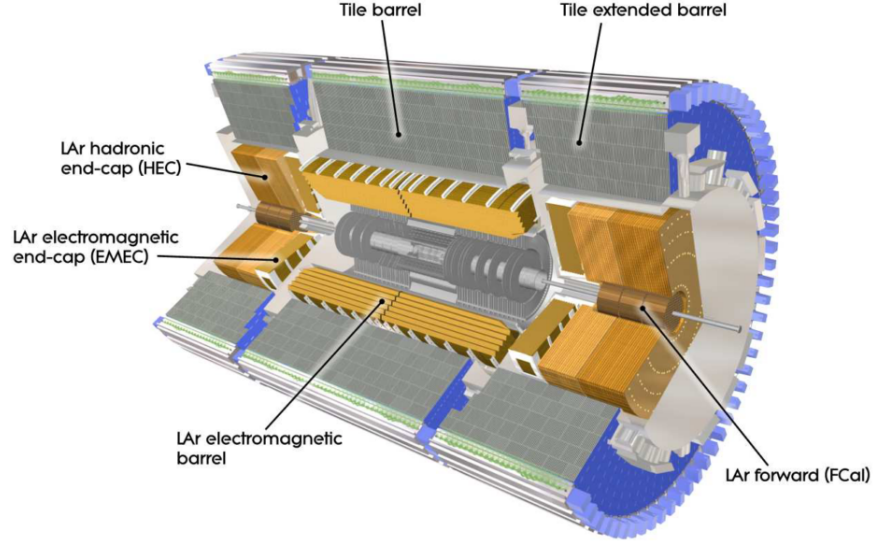


Figure 3.10: Cut-away view of the ATLAS calorimeter system (both EM and hadronic calorimeter systems are shown).

waves vary with radius to keep the liquid-argon gap constant. [41]” The lead plates decrease in thickness past $|\eta| > 0.8$ for a more consistent readout.

3.2.3.2. EMEC

The EMEC are made of two coaxial wheels. The minimum distance for a particle in the EMEC is given Table 3.2. The geometric configuration of the accordion waves at the end-caps “are parallel to the radial direction and run axially. [41]” The lead plates increase in thickness past $|\eta| > 2.5$ for a more consistent readout.

3.2.4. Hadronic Calorimeter

The Hadronic Calorimeter measures the energy and direction of hadrons and their decay products. It’s made of three components (Figure 3.10): the barrel and extended-barrel Tile Calorimeters, the LAr-based Hadronic End-cap Calorimeter (HEC), and the LAr-based Forward Calorimeter (FCal2 & FCal3). Similarly to the EM calorimeter, the Hadronic Calorimeter is sectioned into trigger towers for use in the L1 Trigger. The full range of the Hadronic Calorimeter extends up to $|\eta| < 4.9$.

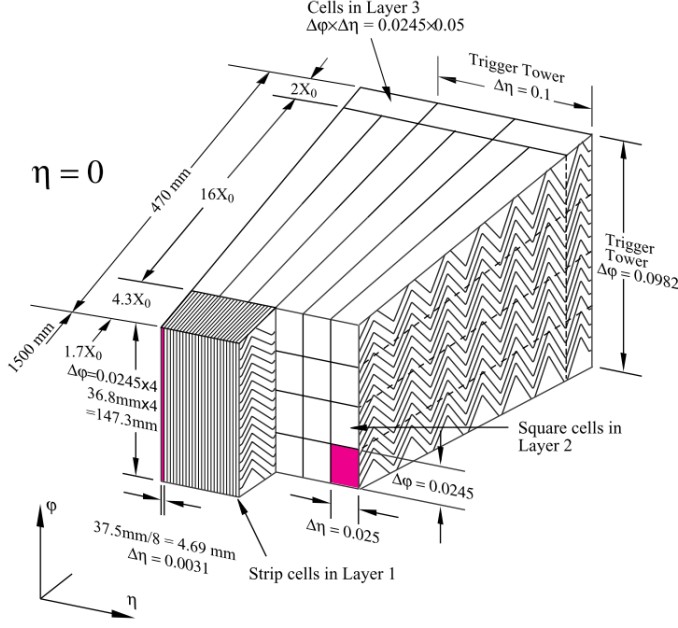


Figure 3.11: Sketch of a barrel module for the EM calorimeter demonstrating the accordion geometry.

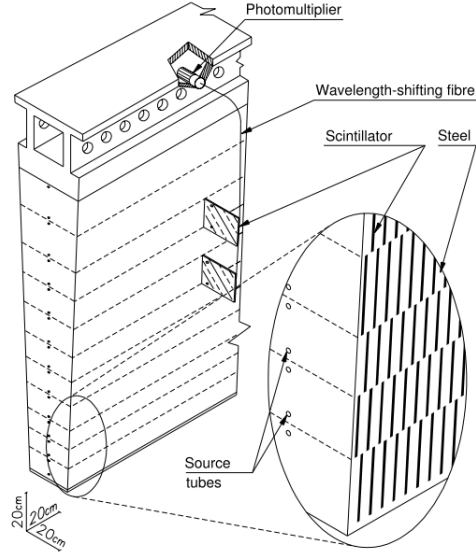


Figure 3.12: Schematic showing how the mechanical assembly and the optical read-out of the tile calorimeter are integrated together.

3.2.4.1. Tile

The Tile Calorimeter has two sections: the barrel which covers $|\eta| < 1.0$ and the extended barrel which covers $0.8 < |\eta| < 1.7$. It is placed just outside the barrel EM calorimeter and is made from alternating steel absorbers and scintillating tiles (Figure 3.12) with photomultiplier tubes used to measure the light from the scintillators. The thickness of the Tile Calorimeter in units of interaction length is given in Table 3.3.

3.2.4.2. HEC

The HEC has two segments of copper/LAr sampling calorimeter wheels configured in such a way that the wheel closest to the interaction point is called HEC1 and the furthest is called HEC2. Here, the copper plates needn't conform to the accordion geometry of the EM calorimeter for full azimuthal coverage. The thickness of the copper plates increases and the granularity decreases in HEC2 versus HEC1.

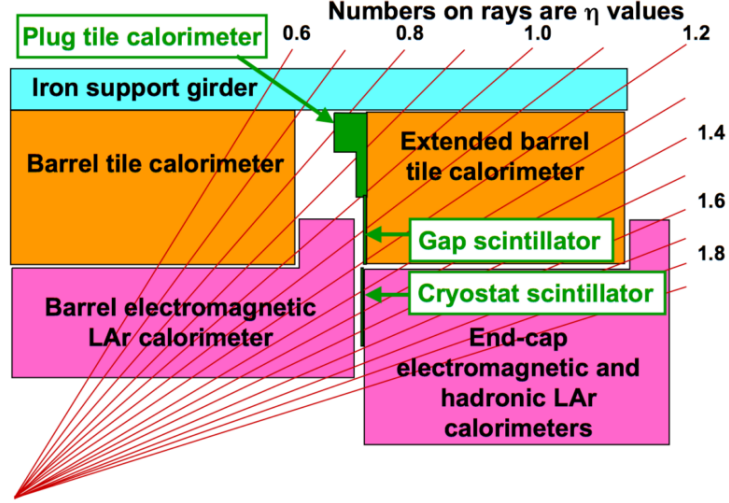


Figure 3.13: Schematic of the transition region between the barrel and end-cap cryostats, where additional scintillator elements are installed to provide corrections for energy lost in inactive material (not shown).

Calorimeter Section	Minimum Distance (λ)	Range
Tile	9.7	$ \eta < 1.7$
HEC	10	$1.5 < \eta < 3.2$
FCAL	10	$3.1 < \eta < 4.9$

Table 3.3: Table of selected Hadronic Calorimeter parameters. Here, λ represents interaction lengths.

3.2.4.3. FCal

The FCal has three LAr-based components as shown in Figure 3.14. The component closest to the interaction region is devoted to electromagnetic calorimetry (FCal1) and uses copper as the absorber material. The next two segments (FCal2 & FCal3) are devoted to hadronic calorimetry and use tungsten as the absorber to maximize absorption length. In all FCal components LAr is the sensitive medium and fills the space between copper tubes and copper or tungsten rods. These rods are aligned with the beam axis and form a closely-packed structure.

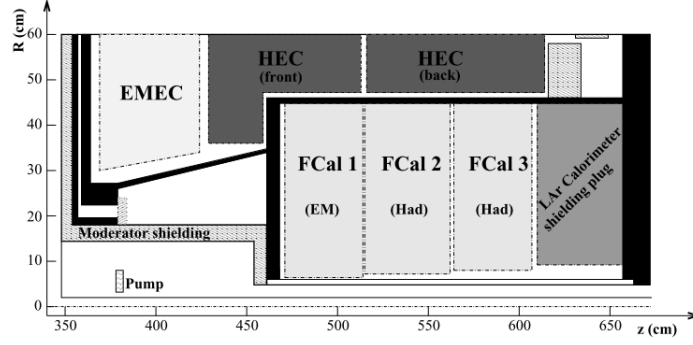


Figure 3.14: Diagram of the relative positions of the FCal components in the r - η plane.

3.2.5. Muon Systems

The muon spectrometers (Figure 3.15) are designed to identify, track, and trigger on muon leptons originating from the interaction region. It relies heavily on the toroidal magnet systems to sufficiently deflect muons for momentum and charge resolution. The Monitored Drift Tubes (MDT) cover the barrel and end-cap regions with the Cathode Strip Chambers (CSC) lending additional coverage in the end-cap region. Together, the MDT and CSC provide precision tracking for muons. Special muon systems are devoted to triggering on muons and also provide secondary tracking information. These are the Resistive Plate Chambers (RPC) in the barrel and Thin Gap Chambers (TGC) at the end-caps. The precision tracking is available within $|\eta| < 2.7$ while triggering is only available within $|\eta| < 2.4$. Muon momentum measurements are unavailable for $|\eta| < 0.1$ and $|\eta| = 1.3$ due to gaps in coverage by the spectrometers.

3.2.5.1. MDT

The MDT is based on arrays of proportional drift tubes (made of aluminum) filled with an argon gas mixture and tungsten-rhenium anode wire (Figure 3.16). Arrays of these drift tubes are packaged in rectangular (barrel) or wedge (end-cap) shaped sections which are themselves placed into larger configurations shown in Figure 3.15. The three barrel layers of the MDT are located in between the tile calorimeter and the barrel toroid (two layers) and amongst the coils of the barrel toroid (one layer) while the end-cap sections are located

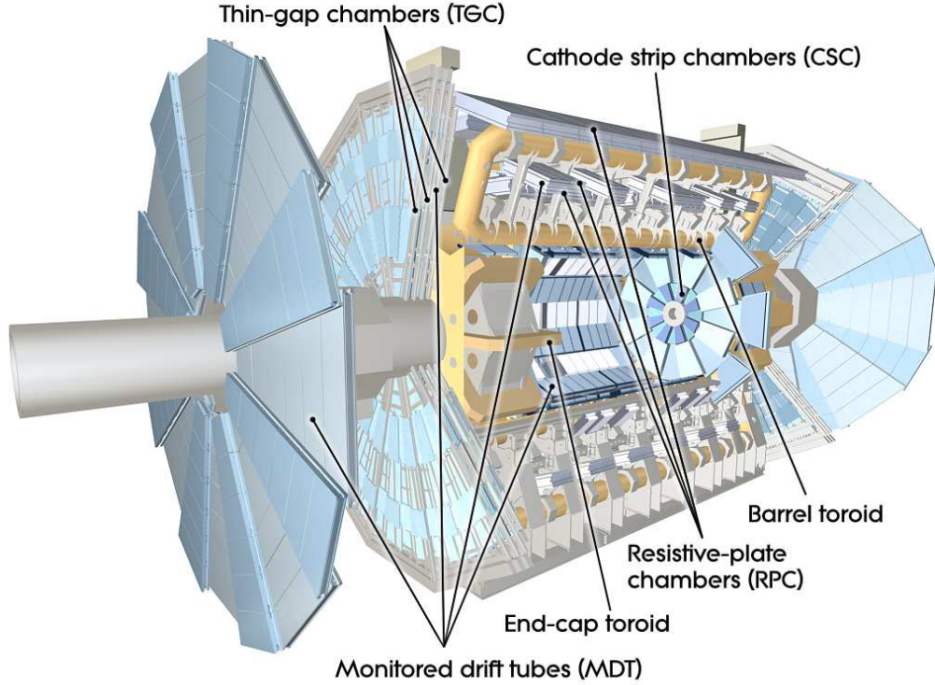


Figure 3.15: Cut-away view of the ATLAS muon system.

within and beyond the TGC. The total coverage of the MDT is $|\eta| < 2.7$.

3.2.5.2. *CSC*

The CSC provides complementary end-cap coverage ($2.0 < |\eta| < 2.7$) to the MDT. The CSC are made from multi-wire proportional chambers and offer higher granularity and rate tolerance than the MDT. The CSC is a single wheel located in between the end-cap calorimeters and the end-cap toroid.

3.2.5.3. *RPC*

The RPC provides muon triggering capabilities in the barrel region. It is located alongside the MDT within the barrel toroid (Figure 3.17) with a coverage of $|\eta| < 1.05$. The RPC is a gaseous parallel electrode-plate detector which allows for the fast operational requirements of the trigger.

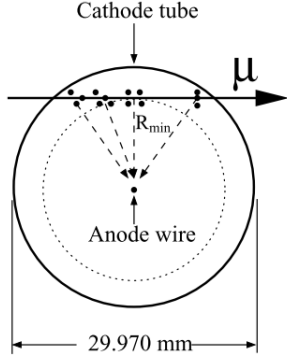


Figure 3.16: Cross-section of an MDT tube with an incident muon shown to illustrate the build up of ions around the muon's path.

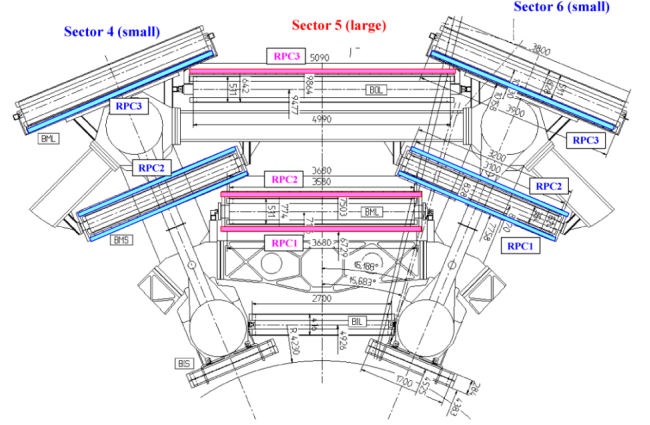


Figure 3.17: Cross-section through the upper part of the barrel with the RPCs marked in color. All dimensions are in mm.

3.2.5.4. TGC

The TGC are multi-wire proportional chambers designed to provide muon triggering capabilities at the end-cap regions and offer another coordinate for the end-cap MDT for precision tracking. The coverage of the TGC is $1.05 < |\eta| < 2.4$ for the trigger and $1.05 < |\eta| < 2.7$ for the precision muon tracking. It is located just beyond the end-cap toroid and has a total of seven layers.

3.2.6. Trigger Systems

The high rate of $p-p$ collisions (40 MHz during Run 2) at the LHC requires that the detectors have a mechanism for quickly determining which events are interesting enough (defined by ATLAS' physics goals) to store for further analysis. The ATLAS trigger system has two components: the Level 1 (L1) trigger and the High Level Trigger (HLT) [48]. The L1 trigger is a hardware system for quickly analyzing detector readout from the calorimeters and muon RPC & TGC. Once an event passes the L1 trigger it's transferred to the software-based HLT which uses the full detector readout to make a final decision to record the event for future processing. Each component of the trigger has its own trigger menu based made from a variety of signatures from each trigger slice. A signature corresponds to the specific

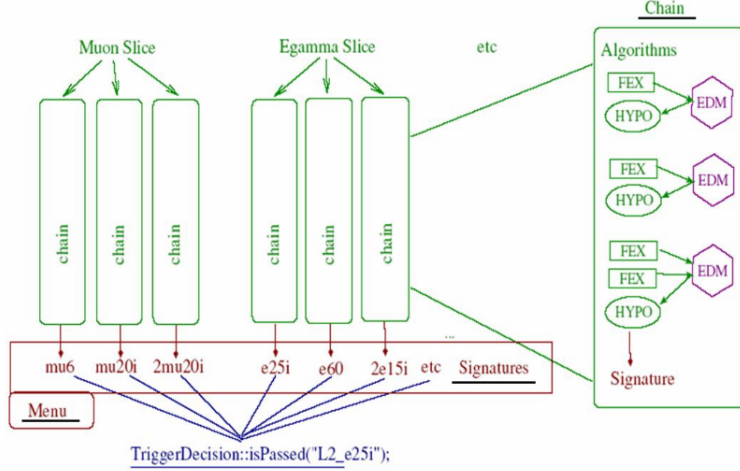


Figure 3.18: Conceptual schematic of trigger terminology. Signatures (in dark red) are grouped into a trigger menu. Each signature is made of a trigger chain (in green). In the HLT each trigger chain is made of a combination of FEX and HYPO algorithms which interact with the trigger’s event data model (EDM). Analyzers can query the trigger decision through a uniform interface (in blue).

detector readout of interest to the trigger. This could correspond to minimum bias events, cosmic muons, hadronic jets above a given p_T threshold, etc. A trigger slice is a conceptual grouping of signatures which corresponds to roughly the same physics object. For instance, there are muon, e-gamma, jet, and b-jet slices. A schematic of this terminology is given in Figure 3.18. The trigger also has the ability to pre-scale menu items (accept one out every n events of a given signature) to adapt to changes in luminosity or detector conditions. I will delay giving an overview of the trigger algorithms until Chapter 4 since understanding them requires more general knowledge of event reconstruction.

3.2.6.1. L1

The L1 trigger is a hardware-based trigger system. It’s designed to look for high- p_T muons, electrons, photons, jets, hadronically decaying τ -leptons, and missing and total transverse energy. It has to reduce the event rate from about 40 MHz to 100 kHz due to the operational requirements of the HLT. Information from the ID is not used at L1 and only a subset of the remaining detector is used. The L1 identifies regions in η - ϕ space that

satisfy criteria for certain signatures called regions of interest (RoIs). RoIs are more coarsely grained than the full detector resolution, thus L1 is used to make rough estimates of what physics objects could be in an event. The output of the L1 trigger is added to a readout stream via dedicated read-out-drivers (RODs).

3.2.6.2. HLT

The HLT uses the full granularity of the detector and offline reconstruction algorithms to make its decision. The HLT needs to reduce the event rate to roughly 1 kHz in order to safely record those events for later processing as higher rates than this can lead to a loss of data. Uninteresting events need to be rejected as soon as possible. This leads to a software architecture optimally segmented into algorithms to extract relevant features (FEX) and test object hypotheses (HYPO). This is diagrammed in Figure 3.18 where the sequential nature of the hypothesis tests are clearly shown for a given chain. More details of the HLT software are given in Appendix A.

Chapter 4

Simulation and Reconstruction

In the most generic sense, simulation ultimately serves as the “theoretical” model to which particle physicists compare actual observations in the detector. This model must incorporate a multitude of factors including state-of-the-art calculations in particle physics, phenomenological effects such as hadronization and material interactions, and the exact manner in which particles manifest in the detector as electronic signals. These steps are commonly known as event (or 4-vector) generation, detector and physics simulation, and digitization, respectively. These are the building blocks of simulations in HEP and are the basis by which theoretical predictions are judged against reality. Because these simulated events are constructed from first principles, simulation is often (somewhat paradoxically) referred to as “truth” when comparing to actual detector data¹.

Reconstruction is the process of applying pattern recognition to detector information to infer the existence of particles. This involves such tasks as searching for tracks² left by charged particles from “hits” in the ID, differentiating a photon from an electron using those tracks, and determining a proper method for calibrating the resulting photons or electrons. All of these tasks require specialized algorithms that are the result of decades of expert physics and detector knowledge. Finally, careful calibrations of the reconstructed objects are required for good agreement to detector data. A popular method for calibration is known as the tag-and-probe method where an object to be calibrated is “tagged” (selected using stringent requirements) and separately a “probe” (selected using very loose requirements) is used to calculate the efficiency of locating this type of object using high-purity simulated and data samples. Here, the word “object” refers to anything whose components are either raw

¹One could argue that detector data is “truth” while simulation is a “best guess.” But, I digress...

²Tracks are a trajectory that indicates the flight path of a charged particle as it deposited energy in the detector.

detector informations (ID hits, calorimeter deposits) or made from other objects (photons made from calorimeter deposits). The phrase “physics object” means detector information collections that represent fundamental physical particles or assemblies of these particles (e.g. jets).

The common ATLAS software infrastructure for both simulation and reconstruction is named Athena [50]. This framework is built on the GAUDI library³ [49] and serves as a control architecture for the various software components. The bulk of Athena is written in C++ with Python acting as an accompanying scripting language. The general Athena workflow consists of three modular steps centered around event processing. The first is initialization and configuration where all requested tools (self-contained software for specific tasks), services (software for tasks such as logging, etc.), and algorithms (software designed to be called once per event and in a well-defined sequence) are loaded into memory and appropriately configured (usually through Python). The next step is the event loop where a defined sequence of algorithms is repeatedly executed until the finalization begins. The same tools may be used in several different algorithms and the same services are generally shared between all algorithms (i.e. no need for different loggers within the same execution). Data flow within the event loop may appear to flow from the input source, down through the sequence of algorithms, and finally through to output files or the console. However, data actually flows through a service called the transient event data store as shown in Figure 4.1. This service facilitates data sharing between algorithms without having to directly interface these algorithms at the software level. The last step is finalization (after all event have been processed) where all algorithms can summarize their performance and clean up any memory they’ve used. More details about how this is used in the context of the b -jet trigger is given in Appendix A.

An overview of the simulation and reconstruction procedures relevant for the ATLAS detector and the analyses in this thesis is given in this chapter.

³Itself built on ROOT [51].

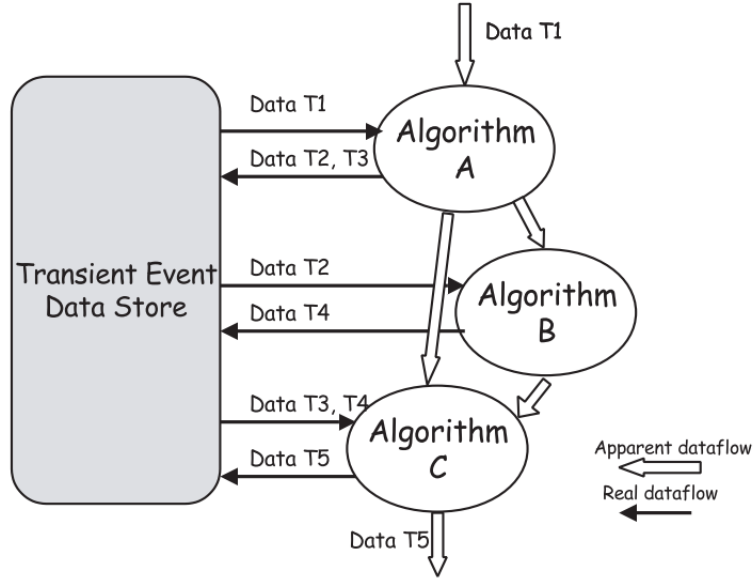


Figure 4.1: The actual versus perceived flow of data in a GAUDI program [49]. The transient event data store actually holds all data on a per event basis.

4.1. Simulation Overview

An overview of the ATLAS simulation workflow is given in Figure 4.2 (unless otherwise stated the information in this section is taken from Ref. [52]). The process starts with event generation which is stored in the common “HepMC” format. The abbreviation “MC” stands for Monte Carlo⁴ and is synonymous with the simulation of particle physics phenomena. This decoupling between generation and rest of the simulation allows the same events to be used in many different detector scenarios without having to rerun event generation software. These events are then passed through a particle filter that selects subsets of events based on the presence of particular particles (i.e. leptons, large p_T neutrino, etc.) and are stored as MCTruth (Gen). These “truth particles” are then used as input to a simulator of the interactions between the event particles and the detector material while respecting the detector geometry. This is by far the most computationally taxing step. The results of the simulation are stored as both MCTruth (Sim) and as a “hit file.” The MCTruth (Sim) is a

⁴This is due to Monte Carlo, Monaco being home to famous casinos. The connection is that the nature of both particle physics simulations (along with several other quantitative fields) and casinos rely quite a bit on random numbers and chance.

offline reconstruction algorithms operate solely on these RDO files. Note that since many of the steps are decoupled, this allows for disjoint MC simulation productions and asynchronous development of the various components.

4.1.1. Generators

Simulated event generation for particle physics is accomplished through a variety of software packages independent of Athena. This software’s primary focus is the generation of the type, quantity, and kinematic description of fundamental particles and possibly the bound states of such particles. The event generators and simulation programs relevant for this thesis are PYTHIA [54, 55], HERWIG [56], SHERPA [57], Jimmy [58], POWHEG [59], MADGRAPH [60], AMC@NLO [61], and EVTGEN [62]. All of these generators have several parameters that govern how they simulate particle physics. A specific set of parameters for a generator is called a “tune” and will be elaborated on in Chapter 5. Athena has wrappers to interface all generators to allow reconfiguration without the need for recompilation of the generator codes. In order to be passed along for further simulation a generated particle must satisfy the condition $c\tau > 10\text{ mm}$ (speed of light times proper lifetime) to be considered a “stable particle”. Since the Athena framework must accommodate a plethora of generators, the MCTruth (Gen) record of particles is necessarily approximate. For instance, the record holds (as best it can) the decay chains of particles in what resembles a tree structure. However, this is a somewhat classical picture of a purely quantum phenomenon and can be misleading if taken too literally (i.e. a tree cannot capture the true relationship of particles arising from quantum interference). The generators listed above can be categorized as being either general-purpose parton showering programs (PYTHIA, SHERPA, HERWIG), fixed-order QCD computations (MADGRAPH, AMC@NLO) and specialized programs (EVTGEN) or program interfaces (Jimmy, POWHEG). General purpose generators are capable of starting from descriptions of the incoming proton beams and producing a complete final state ready for further detector simulation. Additionally, MADGRAPH is an NLO parton-level calculator, PYTHIA, SHERPA, and HERWIG are LO parton showering

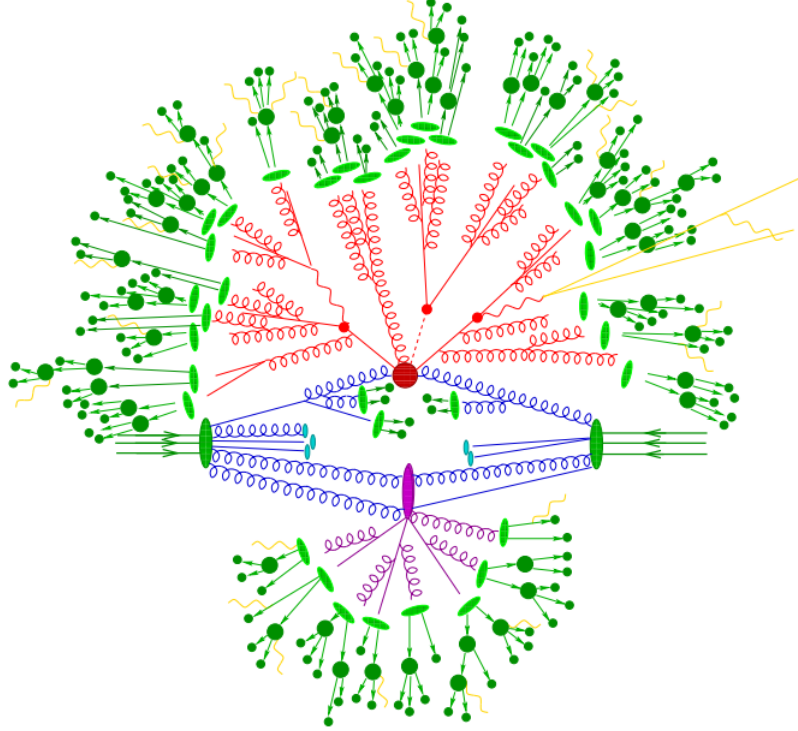


Figure 4.3: Diagram of hadronization and showering in event generation [53]. From the reference: “[t]he red blob in the center represents the hard collision, surrounded by a tree-like structure representing bremsstrahlung as simulated by parton showers. The purple blob indicates a secondary hard scattering event. Parton-to-hadron transitions are represented by light green blobs, dark green blobs indicate hadron decays, while yellow lines signal soft photon radiation.”

programs, and Jimmy, AMC@NLO, and POWHEG are interfaces between NLO fixed-order and LO showering programs. PYTHIA is ATLAS’s default general-purpose event generator. It starts from calculating the fundamental scattering process of interest (referred to as the hard-scatter process) to lowest order in QCD and adds additional radiation calculated with a so-called showering algorithm. These showering algorithms add additional partons (quarks and gluons) and photons. PYTHIA can also simulate additional scattering processes within the same event to approximate what is known as “underlying activity” (i.e. in-time pile-up effects). A similar approach using processes from previous events is used to simulate out-of-time pile-up. It also approximates the hadronization of the generated gluons and quarks. A diagram illustrating the complexity of the showering and hadronization is

given in Figure 4.3. There are two versions of PYTHIA and although PYTHIA8 has more features, improved modeling, and more easily interfaces with Athena, both PYTHIA6 and PYTHIA8 are used in ATLAS. SHERPA is a general-purpose generator that uses PYTHIA for hadronization but is expected to give a better description for events with many isolated jets (see Section 4.2.5.2) than PYTHIA. HERWIG is another general purpose generator that uses Jimmy to simulate underlying events⁶. Specialized generators are used to increase the accuracy of or enable certain final states. For instance, EVTGEN is used exclusively to better model b - and c -hadron decays; MADGRAPH, AMC@NLO, and POWHEG are used to compute the hard-scatter event to higher orders in QCD. In the case of MADGRAPH, it also allows for the inclusion of non-SM processes including the 2HDM A -boson searched for in this thesis.

4.1.2. Modeling Detector Responses and Digitization

The interactions of particles with the ATLAS detector material are modeled by the Geant4 [63] simulation software. This software takes into account the geometry and material composition of the entire detector including infrastructure. The inputs to this stage of simulation are either the MCTruth (Gen) records or particles created by Geant4. The outputs of this stage of simulation are a record of the particles as they interacted with the detector, MCTruth (Sim), and collections of hits in each sub-detector. As there are a very large number of particles produced during simulation, the MCTruth (Sim) does not store all of these particles and only saves those according to predefined strategies. The hits are a record of the energy deposition at a given position and time. Each sub-detector then implements their own modeling of the responses to these hits.

To simulate pile-up events, additional hits are overlaid before digitization. The type of events utilized for this step is usually underlying events, cavern background (constant electron and proton background due to neutrons), or various beam effects (beam interaction with the collimators and imperfect vacuum conditions in the beam pipe). Only minimum

⁶Underlying events are essentially the remnants of hadrons (protons in the case of the LHC) that hadronize and may show up in the detector.

bias events require special treatment as all other types of pile-up can be simulated and, thus, can be digitized the same way as normal events. Minimum bias events are overlaid after the digitization step as they are stored in the RDO format (collected as they were in data).

Digitization reproduces how the detector sensors and electronics react to particle energy depositions. This usually consists of requiring a hit to be above a certain threshold in a given amount of time and possibly have complex dependence on the shape of the signal as well. This stage of modeling also includes sub-detector cross-talk, electronic noise, and dead channels along with many other details of their responses. The final outputs of digitization are the RDO and SDO formats. RDOs are then fed to reconstruction algorithms discussed in the next section. SDOs are optional and mainly used for tracking efficiency and fake rate studies.

4.2. Object Reconstruction

Object reconstruction in ATLAS happens at both the online trigger and offline event reconstruction level. Both types of reconstruction are similar, but the detector information used in the trigger is not as complete or well-calibrated as in the offline case. Outlines of the reconstruction algorithms and calibration techniques used for the physics objects relevant for the analyses in this thesis are given in this section.

4.2.1. Tracks

Tracking in ATLAS uses both the ID and muon spectrometer (MS) detector components. Broadly speaking, there are two conceptual ways to reconstruct tracks: *inside-out* and *outside-in* [64]. Inside-out refers to building up track candidates from the inner components of the ID (Pixel and SCT) first and using the outer component (TRT) to refine the trajectories. Likewise, outside-in refers to building initial tracks hypotheses from the TRT and then back-tracking into the SCT and Pixel layers. The primary techniques employed in track reconstruction are Kalman filters and χ^2 fits. Comprehensive overviews of how these techniques are applied in the context of tracking and vertex finding are given in Refs. [65],

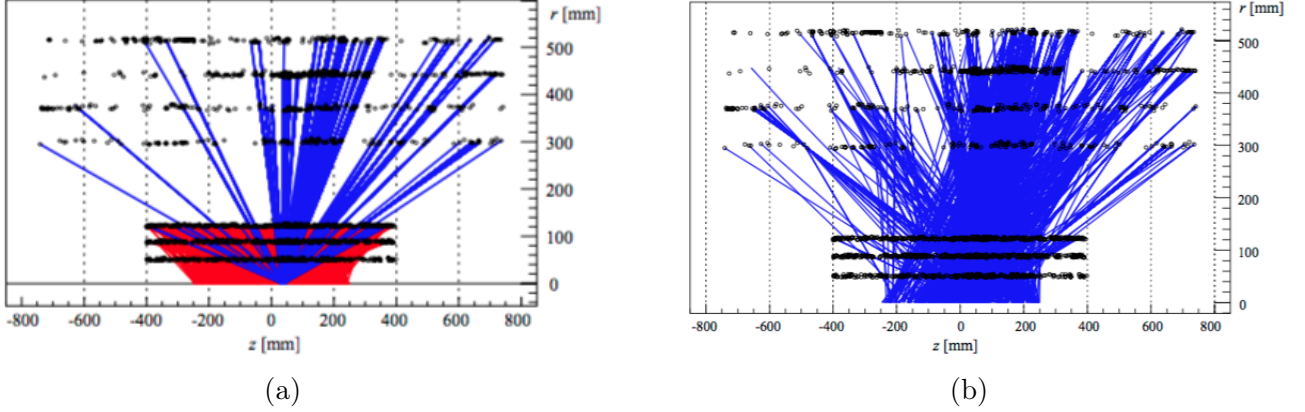


Figure 4.4: Barrel-only view of (a) constrained and (b) unconstrained, three-spacepoint tracks in the Pixel and SCT layers of the ID for Run 1 (no IBL) [64].

[66], [67], and [68]. The process of inside-out track reconstruction is as follows:

- “Spacepoints” (3D representations of the silicon detector measurements) are constructed from Pixel hits and SCT readout. The strips in an SCT module are sandwiched together and rotated w.r.t. one another. This relative orientation along with the location of the beam allow for full 3D determination of a hit.
- Spacepoints from the Pixel detector are then used to seed tracks in constrained and unconstrained algorithms (Figure 4.5).
 - The constrained track seeds are made by taking all tracks made from two spacepoints, extrapolating these tracks back to the beamline (z -axis), and constructing a primary vertex (PV) from the best track-pair candidates (i.e. the largest sum- p_T of tracks). This PV is then used to constrain the number of track seeds (Figure 4.4a).
 - The unconstrained track seeds are simply made from all two-spacepoint pairs from Pixel hits. Obviously, the sheer combinatorics are quite large in this case, but it’s necessary for signatures such as $H \rightarrow \gamma\gamma$ where few ID tracks are expected to point directly to a PV (Figure 4.4b).
- Before extending the tracks to the TRT an ambiguity solving algorithm is run that produces much higher quality tracks than initially constructed. This step revisits all

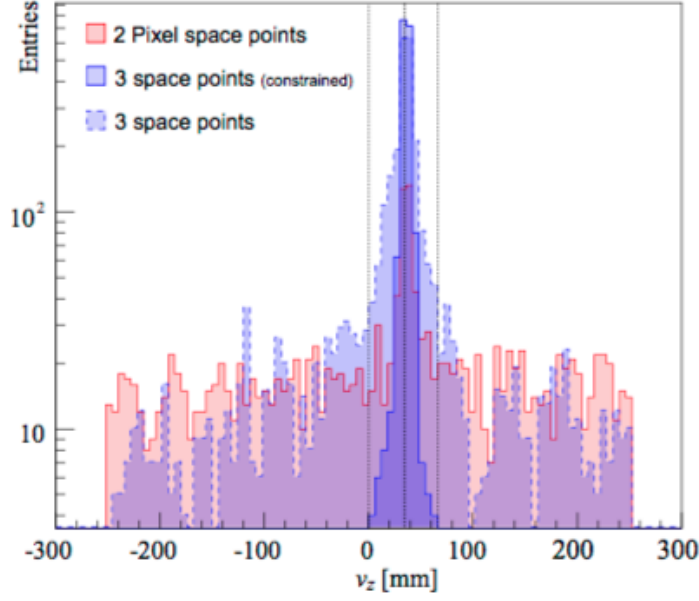


Figure 4.5: Distributions in projected z -coordinates of the constrained and unconstrained two- and three-spacepoint track for Run 1 (no IBL) [64]. The solid histograms are the constrained tracks while the dashed histograms are the unconstrained. The range for selecting a primary vertex is shown by the vertical lines at 0 and ~ 65 mm.

hit-to-track assignments using much more detailed information about the material in the detector and, in addition to χ^2/dof values, determines the tracks' quality scores. These scores are based on the presence of “holes” (absence of hits along trajectory) and hits. Additionally, in Run 2 an artificial neural network [69] (NN) is trained to further differentiate shared track hits [70, 71].

- Finally the tracks are then extended to the TRT region and matched as best as possible to trajectories measured in the TRT (Figure 4.6).

The inside-out approach is susceptible to many different systemic defects including non-existent track seeds and incorrect extrapolation into the TRT through energy loss in the outer layers of the SCT. Thus, this method is complemented by the outside-in reconstruction of tracks. Here, a Kalman filter is employed in a two step process of global pattern searches combined with subsequent local pattern matching. Because of the projective nature of the TRT, spacepoints cannot be reconstructed at this stage. Track candidates are then backtracked into SCT and Pixel layers. More details of the outside-in method are given in

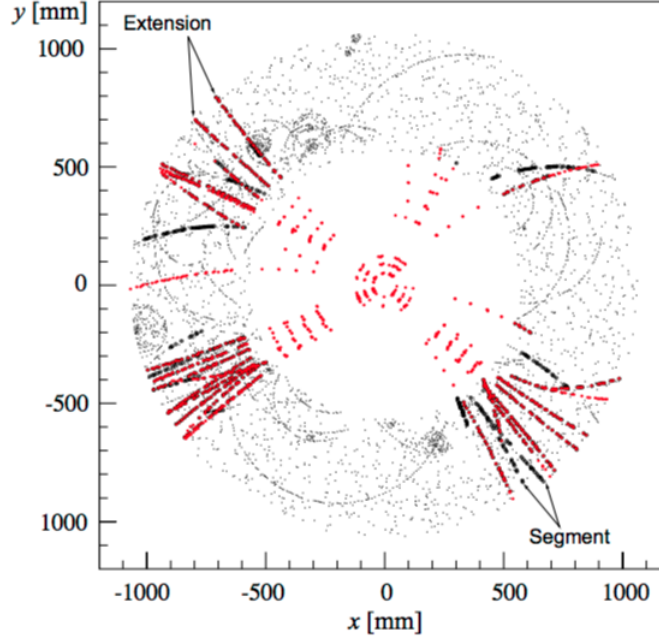


Figure 4.6: Same event as shown in Figure 4.4a showing those tracks extended out to the TRT looking down the barrel of the ID along the beam axis [64].

Ref. [64]. After both the inside-out and outside-in methods find quality tracks, vertexing algorithms find PVs followed by secondary vertexes (SVs) and photon conversions.

A summary of tracking performance is given in Ref. [72] which shows good agreement between simulation and data for most measures of track fitness (i.e. ID hits for both individual tracks and within calorimeter-based jets). However, the resolution of the transverse (d_0) and longitudinal (z_0) impact parameters is not as well-described by simulation as shown in Figure 4.7. This disagreement between data and MC simulation is primarily due to inaccurate material descriptions at low p_T , residual misalignment in the ID at high p_T , and mis-modeling of the energy deposition in the silicon detectors. Additionally, since the impact parameters are measured w.r.t. the PV the resolution of the PV is folded into the impact parameters and, thus, must be taken into account when comparing impact parameters in data and MC simulation.

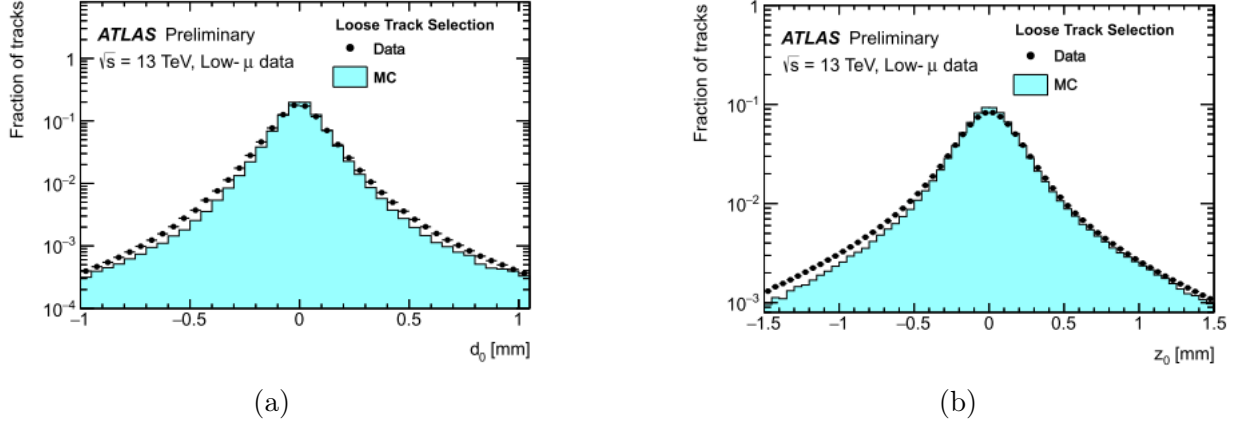


Figure 4.7: (a) Transverse and (b) longitudinal impact parameter distributions taken from special, low pile-up data in 2015. Details of the data-MC disagreement are given in the text [72].

4.2.2. Calorimeter Clusters

Object reconstruction requires that the relatively fine-grained, raw output of the calorimeters be grouped together in so-called “clusters.” Calorimeter clusters can be constructed from either EM calorimeter (ECAL) or a combination of the ECAL and hadronic calorimeter (HCAL) energy deposits. There are two types of clustering algorithms used in ATLAS, sliding-window and topological clustering. The following are brief reviews of these technique along with their performance (all material in this section is taken from Ref. [73]).

4.2.2.1. Sliding-Window Clustering

This type of clustering has three conceptual steps (although in practice they may be merged for efficiency): tower building, seed finding, and cluster filling. In tower building, a calorimeter (either just the ECAL or both the ECAL and HCAL) is divided into $N_\eta \times N_\phi$ tower (denoted T_i) each of area $\Delta\eta \times \Delta\phi$ in η - ϕ space. These T_i also include all layers of the calorimeter along the radial direction within their area of $\Delta\eta \times \Delta\phi$. For each T_i the calorimeter cells are summed to form the total energy of the tower. Overlapping cells share their energy proportionally to how much of their η - ϕ area overlaps with a tower. Clusters that are built from towers reference only the individual cells while jets keep reference to

towers and ignore the individual cells. This has implications in data storage and processing. Seed finding uses a sliding window of area⁷ $N_\eta^w \Delta\eta \times N_\phi^w \Delta\phi$ to find groups of towers with transverse energy larger than a given threshold (E_T^{thresh}) and locally maximum. The size of the window (N_η^w and N_ϕ^w) is optimized for balancing efficiency of seed finding against being robust against noise. The E_T -weighted center (in η - ϕ space) of the seed is computed from the cells inside the sliding window. This center position is expressed as (usually) smaller multiples ($N_{\eta/\phi}^{\text{pos}}$) of the T_i dimensions. Overlapping seeds are removed by requiring that if two or more seeds are within $\Delta\eta_{\text{dup}} \times \Delta\phi_{\text{dup}}$ of each other, the seed with the largest E_T is kept while the rest are discarded. Cluster filling is different for pure ECAL clusters or combined ECAL and HCAL clusters. In the combined scenario there exists an option to “pre-fill” the clusters by grouping all towers centered at a seed and within the sliding window’s dimensions and designating this as a cluster. For ECAL-only clusters cluster filling is done by taking the varying windows of towers depending on location and layer of the calorimeter.

For the range of electron transverse energies relevant for this thesis the sliding window algorithm is $\sim 99\%$ efficient at finding at least one cluster. Energy sharing of cells that overlap between towers is important as not doing so could lead to double counting energy deposits and artificially increase the energy of the reconstructed objects.

4.2.2.2. Topological Clustering

Topological clustering involves two main steps: cluster making and cluster splitting. It first finds all cells above a high signal-to-noise ratio (known as a “seed-threshold”). Neighboring cells above a different signal-to-noise ratio (“neighbor-threshold”) are then recursively added to the “proto-cluster” until there are no further neighboring cells above the neighbor-threshold. This stage also ensures that overlapping proto-clusters are merged. The final set of proto-clusters is made from those clusters that have an E_T greater than a given threshold. The parameters that govern these process are optimized against fake rate in the ECAL only scenario while maximizing efficiency in the combined scenario. Due to the nature of

⁷ $N_{\eta/\phi}^w$ are integer multiples in the η/ϕ directions.

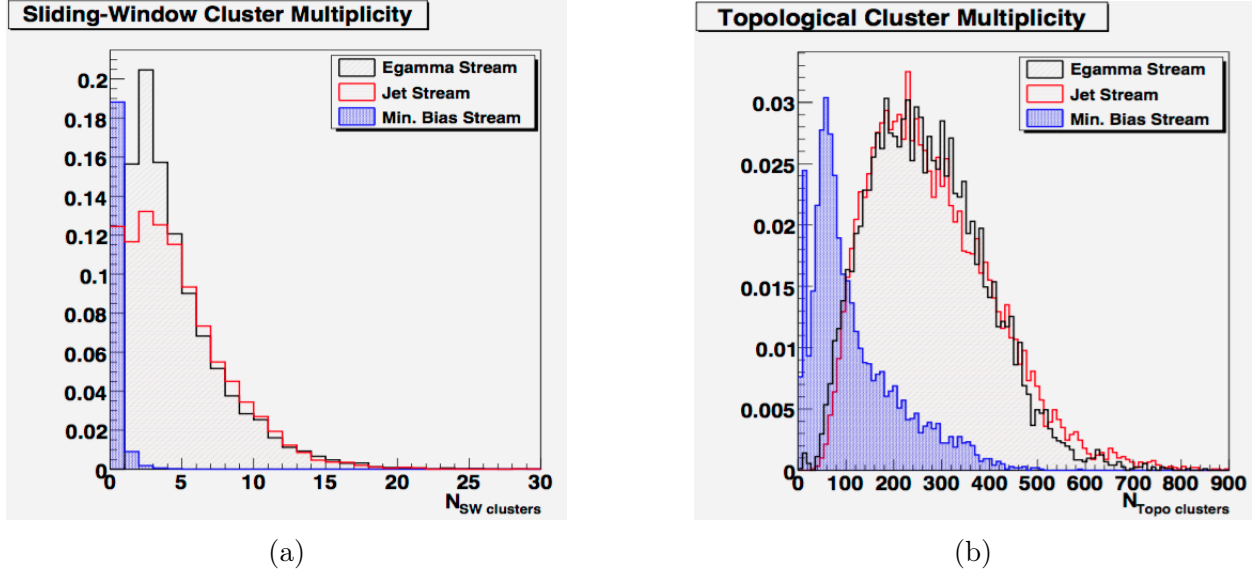


Figure 4.8: Comparison of sliding window versus topological clustering multiplicity for electron/photon, jet, and minimum bias simulated events [73].

the clustering algorithm, energy deposits from distinct particles can be merged into a single topological cluster. To cope with this a splitting step is run after the clusters are made (the details of which can be found in the aforementioned reference). After splitting, the clusters are now three dimensional blobs that may have shared cells at their edges where the energy is split appropriately.

An important difference between the sliding window and topological clustering is the fact that the resulting number of clusters is much larger in topological clustering as shown in Figure 4.8. This fact is relevant in regards to jet reconstruction where topological clusters are used as the basic building blocks. Another important difference is that since the clusters in topological clustering are of irregular shape and size, they are more difficult to calibrate than their sliding-window-based counterparts.

4.2.3. Electrons

Electron reconstruction [75] starts with raw signals from the ECAL cells after reshaping and sampling⁸. These cells are converted to energy depositions using hardware calibrations

⁸The physics signal from an ECAL cell is a sawtooth wave. However, the period of this signal is too

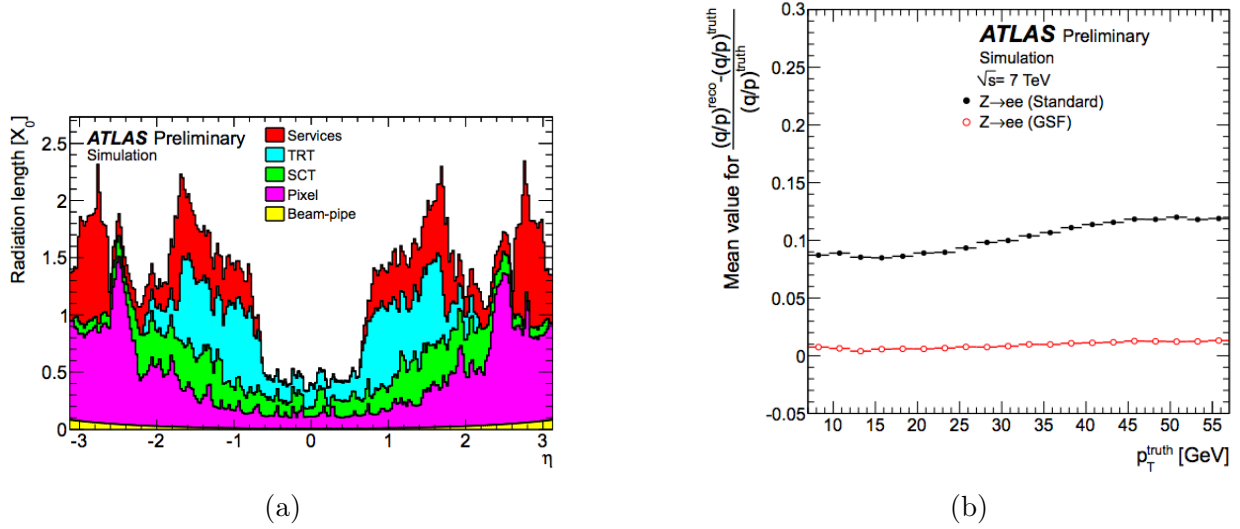


Figure 4.9: (a) Simulated radiation-length profile of the beampipe and ID for Run 1 (thicker beampipe and no IBL versus Run 2). Note the dependence of thickness on η . (b) The mean relative bias of the track inverse momentum multiplied by the charge q/p for GSF and standard (pre-GSF) truth-matched electrons for Z -boson decays [74].

that are periodically updated. The barrel portion of the ECAL ($|\eta| < 2.47$) clusters of size $\Delta\eta \times \Delta\phi = 0.075 \times 0.175$ are used to reconstruct EM particles (photons and electrons) using sliding-window clustering while clusters of size $\Delta\eta \times \Delta\phi = 0.125 \times 0.125$ are used at the endcaps. Tracks from the ID are then matched to these calorimeter clusters to determine whether or not the deposits are due to a photon or electron. The general rules for photon and electron identification are:

- Unconverted photons are those clusters with no matching ID tracks
- Converted photons are those clusters matched to a pair of oppositely charged tracks that are collinear from the PV to the cluster and satisfy signatures in the TRT consistent with electrons⁹
- Electrons are all other cases of matched tracks to ECAL clusters

Radiative losses (mostly bremsstrahlung) in the ID from electrons is corrected for using long given the LHC bunch spacing. Thus, the signal is reshaped to have a narrower peak and only digitally sampled (in the crest) four or five times.

⁹Since the detection of the track pairs is inefficient at large conversion radii, a single track may also count as a converted photon if there are no hits in the innermost portions of the ID.

a Gaussian Sum Filter (GSF) technique [74]. This technique performs a bremsstrahlung refitting prior to full electron reconstruction and bremsstrahlung recovery at the initial step of electron trajectory calculation. Ultimately, these corrections help match tracks from the electrons in the ID to ECAL deposits and, thus, greatly aid in electron identification. The performance of this technique is shown in Figure 4.9.

The simulation-based calibration of electrons is performed by a multivariate (MVA) regression which corrects for energy deposited in front of the calorimeter and outside the clusters. This calibration for a 100 GeV electron is less than 20% and ~5% for the energy lost in front of the calorimeter and outside clusters, respectively. In-situ calibrations are performed with $Z \rightarrow e^+e^-$ events using a parameterization (α) of the energy mis-calibration in bins of η (i):

$$E_i^{\text{data}} = E_i^{\text{MC}}(1 - \alpha_i)$$

The relative energy resolution (Equation 4.1) is parameterized by three terms, all of which are determined by the MVA regression.

$$\frac{\sigma(E)}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c \quad (4.1)$$

To first order the difference in energy resolution doesn't depend on energy, thus

$$\left(\frac{\sigma(E)}{E}\right)_i^{\text{data}} = \left(\frac{\sigma(E)}{E}\right)_i^{\text{MC}} \oplus c'_i.$$

The measured energy scale factors and constant term in the energy resolution are given in Figure 4.10. These calibrations have numerous, associated systematic uncertainties including those from event selection during calibration, the calibration procedure itself, material mis-modeling, and detector conditions and limited statistics during in-situ calibrations. This results in over over 70 different sources of uncertainty. However, there is an option to sum these sources in quadrature¹⁰ in analyses that are insensitive to the electron scale such as

¹⁰This is somewhat conservative, but saves time during statistical fitting.

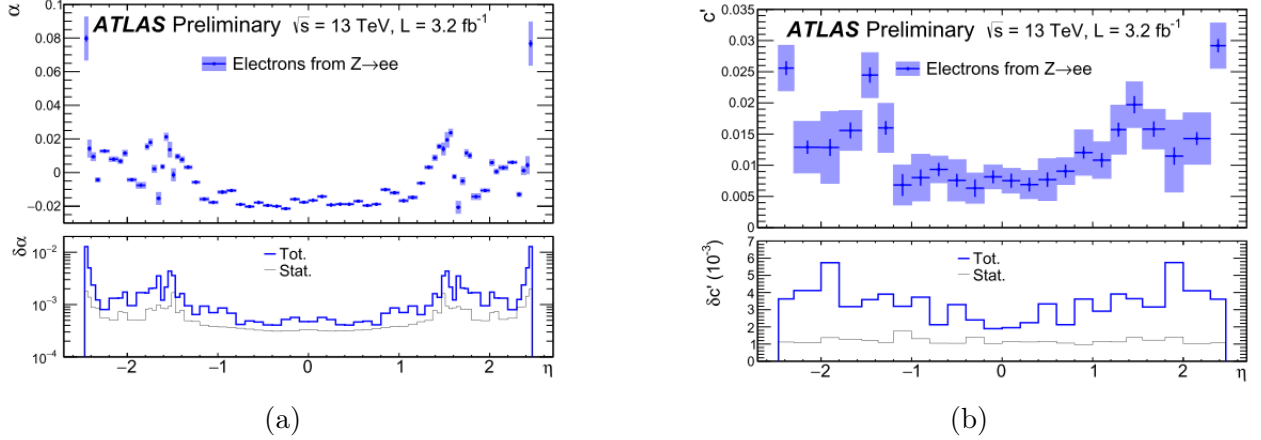


Figure 4.10: Electron (a) energy scale and (b) resolution corrections as functions of η . The total systematic and statistical errors are given in the bottom plots [75].

the analyses in this thesis. For detailed overview of the calibration procedure and derivation of the corresponding systematics, see Ref. [76].

Varying levels of electron purity are available to analyses through a likelihood-based identification discriminant that is robust against pileup. The available levels are denoted *loose*, *medium*, and *tight* (corresponding to roughly 95, 90, and 80% efficiency at $E_T = 40$ GeV). The inputs to this discriminant are detailed in Ref. [77] and are related to calorimeter and tracking quality. The efficiencies of the likelihood classifications are validated with a tag-and-probe technique in $Z \rightarrow e^+e^-$ events (Figure 4.11). Additional purity criteria orthogonal to those already mentioned are available in the so-called isolation variables. These are measures of how well-isolated the reconstructed electrons are from surrounding detector signals and include a variable, p_T -dependent track isolation cone as well as a calorimeter-based isolation cone. Most relevant for the analyses in this thesis for electrons are the *LooseTrackOnly* and *FixedCutTight* requirements (the details of which can be found in section 7.2 of Ref. [78] in the context of muon reconstruction).

4.2.4. Muons

Reconstruction of muons takes place independently from tracks computed in the ID and MS [78]. Muon tracks are reconstructed in the ID as any other charged particle. The MS

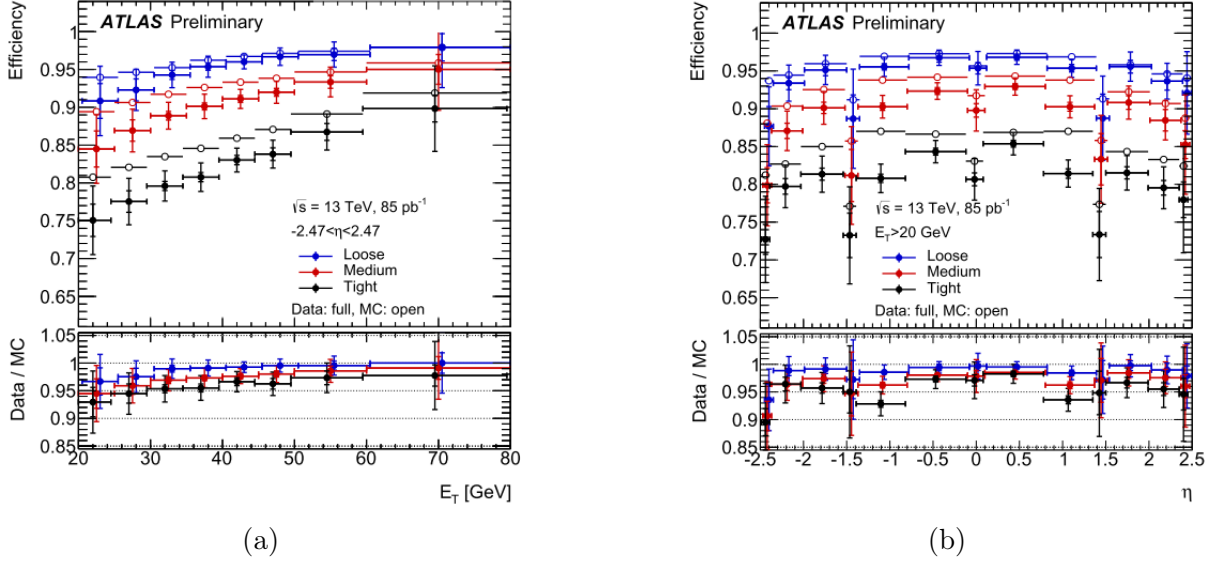


Figure 4.11: Electron identification efficiencies as functions of (a) p_T and (b) η . Note that the dependence on η maps with detector realities such as the dips at $\eta = 0$ and $|\eta| = 1.5$ representing the joining of the barrel ECAL and the transition region between the barrel and endcaps, respectively [77].

reconstruction algorithms search for hit patterns from each muon chamber (the MDT, RPC, TGC, and CSC) to create segments unique to those chambers. A combinatorial search starting with the middle layers of the detector is then performed to create muon trajectories. Much like track reconstruction in the ID, track quality criteria can be imposed to ensure accurate reconstruction in the MS. At least two matching segments are required for all tracks except those at the barrel-endcap transition region where the magnetic field weakens. There, only one, high-quality segment is necessary. Overlapping tracks with shared hits are also resolved. Final χ^2 values for the tracks constructed in the MS are computed and they are now ready to be compared with tracks from the ID. Several different types of muons are available for Run 2:

- Combined (CB) muons are made from combined, global refits of ID and MS tracks using both an outside-in (MS-to-ID) and inside-out (ID-to-MS) scheme.
- Segment-tagged (ST) muons are computed using an inside-out (ID-to-MS) scheme where only one segment in the MS is required (good for low- p_T muons).

- Calorimeter-tagged (CT) muons are made from ID tracks and an ECAL deposit consistent with a minimum ionizing particle (lowest purity but good acceptance).
- Extrapolated (ME) muons are made from MS tracks with loose or no ID track requirements (most useful in $2.5 < |\eta| < 2.7$ where there is no ID coverage).

Overlap between the categories of muons follows the priority CB over ST and ST over CT. Conflict with ME needs track quality information to resolve.

Calibrations for muons are derived with $J/\psi \rightarrow \mu\mu$ (for low p_T) and $Z \rightarrow \mu\mu$ events (for higher p_T) using a combination of MC simulation and data. The calibrations are ultimately expressed in terms of efficiency scale factors (SFs),

$$\text{SF} = \frac{\epsilon^{\text{Data}}(\dots)}{\epsilon^{\text{MC}}(\dots)}, \quad (4.2)$$

where ϵ is the efficiency in either data or MC simulation as a function any number of (here unspecified) parameters (usually the p_T and $|\eta|$ of an object as these quantities correlate to varying detector geometry and hardware responses). The systematic uncertainties related to these SF as a function of p_T are shown in Figure 4.12.

Varying levels of muon purity based on χ^2 and other measures of quality are available to analyses that help to suppress the mostly pion and kaon backgrounds. These levels of purity are denoted *Loose*, *Medium*, *Tight*, and *High- p_T* :

- *Loose* muons maximize efficiency and are constructed from all reconstruction categories.
- *Medium* muons are the default purity level meant to minimize the number of muon-related systematic uncertainties.
- *Tight* muons maximize purity at the expense of some efficiency.
- *High- p_T* muons maximize the momentum resolution of muons with $p_T > 100 \text{ GeV}$

Efficiencies for these purity levels are obtained from $t\bar{t}$ MC simulation (Table 4.1) and validated in data using a tag-and-probe method in $J/\psi \rightarrow \mu\mu$ and $Z \rightarrow \mu\mu$ events [78]. The effi-

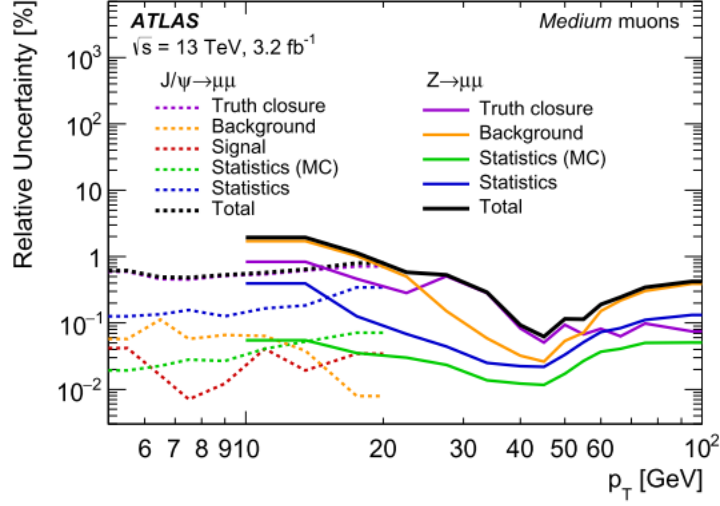


Figure 4.12: Total systematic and statistical uncertainty associated with *Medium* muon calibrations [78]. “Truth closure” accounts for a possible bias in the tag-and-probe method by comparing the efficiency measured with the tag-and-probe method with the “true” efficiency given by the fraction of generator-level muons that are successfully reconstructed. “Signal” and “Background” refer to mis-modelings of the signal and background processes. “Statistics” and its MC variant are uncertainties related to the finite numbers of events in either data or MC simulation.

	$4 < p_T < 20 \text{ GeV}$		$20 < p_T < 100 \text{ GeV}$	
Selection	$\epsilon_{\mu}^{\text{MC}} [\%]$	$\epsilon_{\text{Hadron}}^{\text{MC}} [\%]$	$\epsilon_{\mu}^{\text{MC}} [\%]$	$\epsilon_{\text{Hadron}}^{\text{MC}} [\%]$
<i>Loose</i>	96.7	0.53	98.1	0.76
<i>Medium</i>	95.5	0.38	96.1	0.17
<i>Tight</i>	89.9	0.19	91.8	0.11
<i>High-p_T</i>	78.1	0.26	80.4	0.13

Table 4.1: Muon identification efficiencies derived in $t\bar{t}$ MC simulation in different p_T regimes [78]. The efficiencies for hadronic contamination are also given.

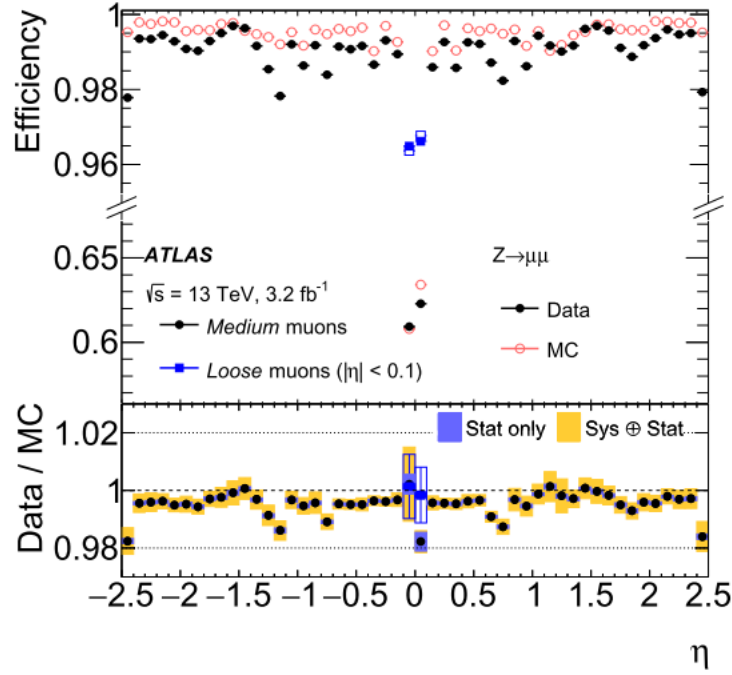


Figure 4.13: Muon identification efficiencies for *Medium* purity versus η [78]. The drastic dip at $|\eta|=0$ is due to the known gap in acceptance in the MS barrel components due to service infrastructure. The slight dips at $|\eta|=1.3$ are due to the small gap in coverage between the barrel and endcap MS components.

ciencies versus η for *Medium* are shown in Figure 4.13. Isolation criteria based on calorimeter- and track-based cones may also be applied to muons to further enhance purity. Of the many categories of isolation, the two that are relevant for the analyses in this thesis for muons are *LooseTrackOnly* and *FixedCutTightTrackOnly* (both only use the track-related isolation cone).

4.2.5. Hadronic Jets

Quarks, gluons, and some τ -leptons manifest in the detector as a directional spray of tracks and calorimeter deposits known as hadronic jets (henceforth referred to as simply jets¹¹). This is true of the b -quark initiated jets (b -jets) produced by the Higgs boson in the signal processes relevant for this thesis as well as the jets initiated from quarks and gluons present in all of the background processes as well. Thus, jets play an important role in identifying event topologies and kinematics consistent with the relevant signal processes for this thesis.

4.2.5.1. Prelude

A “hadron-level jet” is a directional spray of hadrons that results from a color-charged object being scattered or produced during a collision. Gluons and quarks are scattered (asymptotic freedom) or produced in a collision and begins to radiate gluons and quarks in order to achieve color-neutral states (confinement). These scattered and radiated particles hadronize into more stable mesons and baryons that in turn form a collimated spray of particles. A “reconstructed jet” is a grouping of charged tracks and calorimeter deposits into a single object that reliably captures most of the hadron-level jet’s characteristics such as direction, flavor, charge, and p_T . The objective of jet reconstruction is to try to map each hadron-level jet with a reconstructed jet. In simulation it is (at least in principle) possible to trace though the decay chain and fully reconstruct a hadron-level jet. However, this is beyond what an experimentalist can do starting only with charged tracks and calorimeter clusters.

¹¹Note that in general any directional spray of particles may be called a jet in literature. This includes the purely electromagnetic showers in the ECAL for instance.

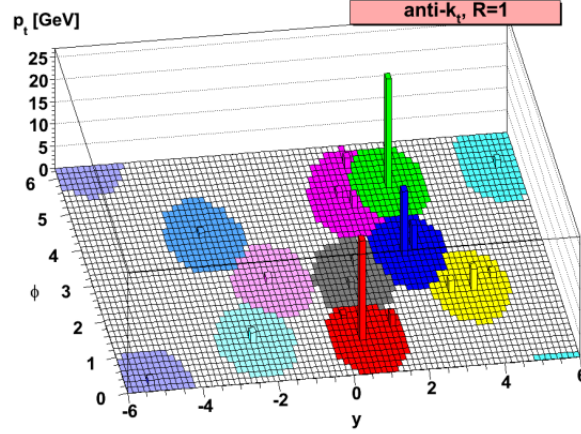


Figure 4.14: A toy simulation demonstrating the clustering performance of the anti- k_t jet clustering algorithm [79]. The particles here are submersed in a randomized assortment of soft gluons that cause other algorithms (not shown) to create non-uniform boundaries. This property is highly desirable for many theoretical and experimental reasons (i.e. passive and active jet areas and related quantities).

Thus, a common algorithm for clustering together objects in both MC simulations and real data must be invented in order to make meaningful comparisons of jets. Two important criteria for such an algorithm are known as infrared and collinear safety [80]. Infrared safety refers to the fact that soft, low-energy tracks can confuse naive algorithms into grouping distinct jets together. Collinear safety refers to a problem some algorithms have with high-energy tracks that are too close to one another wherein these algorithms may needlessly split a single jet into multiple jets. Two different types of jet finding algorithms are popular in HEP [81]: cone and sequential recombination algorithms. Cone algorithms are top-down approaches in that they identify broad regions (cones) of energy or track flow. Sequential recombinations are bottom-up approaches wherein smaller elements are repeatedly combined in accordance with a governing distance measure. The sequential recombination algorithm relevant to jet reconstruction in ATLAS is the anti- k_t algorithm¹² [79] (Figure 4.14). This algorithm (like all sequential recombination algorithms) works by iteratively performing the following steps over some collection of objects:

¹²Here, $k_t = p_T$

1. Compute all pairwise distances between objects i and j , d_{ij}
2. Compute all distances between objects i and the beamline B , d_{iB}
3. Find the smallest distance
 - If it's a pairwise distance between objects, merge them
 - If it's a distance of an object to the beamline, remove that object and label it a jet
4. Repeat from step 1 until there are no more objects

The distances are defined by

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta_{ij}^2}{R^2}$$

$$d_{iB} = k_{ti}^{2p},$$

where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, y is the rapidity (or η for use in ATLAS), and p and R are free parameters. What distinguishes the anti- k_t from other, similar algorithms is that it sets $p = -1$, thus the label “anti” ($p = 1$ for the k_t -algorithm). The advantage for this choice of p is illustrated by Figure 4.14.

4.2.5.2. Jet Objects

The primary type of jet used in the analyses relevant for this thesis is reconstructed from topological calorimeter-cell clusters calibrated using the ECAL scale¹³ [84] (denoted

¹³This is discussed later.

Working Point	JVT Cut	Average efficiency [%]
Loose	0.11	97
Default	0.59	92
Tight	0.91	85

Table 4.2: JVT working points and their average efficiencies [82].

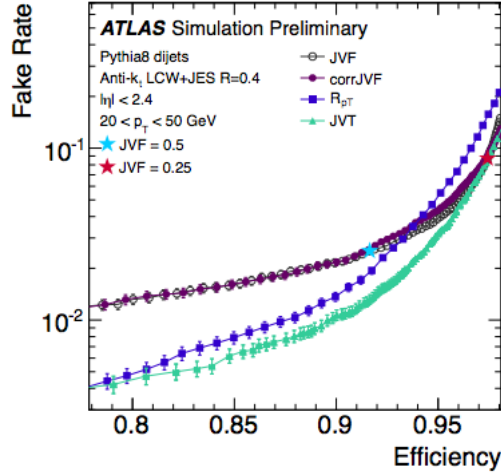


Figure 4.15: Performance of the JVT versus other discriminants against pileup (details found in Ref. [83]). The two inputs to JVT in construction of its likelihood are the R_{p_T} and corrJVF that are also plotted for comparison (described in the reference). The starred values are working points for a Run 1 equivalent.

EMTopo). These clusters are then used as input to the anti- k_t algorithm with a radius parameter of $R=0.4$ to reconstruct jet objects (referred to as small- R calo-jets). To suppress jets arising from pileup a likelihood-based discriminant is used called the Jet Vertex Tagger (JVT) [83]. JVT uses information about the primary vertex, jet p_T , and track p_T to construct a 2-D likelihood that is robust against potential jet flavor and training sample biases. It is trained on jets within a kinematic range that is expected to be sensitive to pileup ($20 \text{ GeV} < p_T < 60 \text{ GeV}$ and $|\eta| < 2.4$) and calibrated with $Z \rightarrow \mu\mu + \text{jets}$ samples. Thus, JVT is only ever used on jets within this kinematic range. JVT is calculated on uncalibrated jets during reconstruction and, therefore, must be recalculated during analysis. Efficiency scale factors at various efficiency working points are available to correct for data-to-MC discrepancies. The list of available working points are given in Table 4.2 and a plot comparing the performance versus other pileup discriminants is given in Figure 4.15. For $A \rightarrow Zh$ signal processes with a large resonant mass the b -jets produced by the Higgs may be too close together to be resolved by the small- R calo-jets. This effect is expected to be noticeable when $p_T^H > 600 \text{ GeV}$ (using the rule of thumb $\Delta R = 2m/p_T$ [85], where $m = m_H$ and $\Delta R = 0.4$). Our approach to selecting the signal b -jets in this “merged” regime is to use a calorimeter-based jet with a large

anti- k_t parameter (referred to as a large- R calo-jet) with large invariant mass and search for smaller- R track-based jets (track-jets) ghost-associated¹⁴ to this large- R calo-jet that pass relevant b -tagging requirements (Section 6.2). Specifically, the analyses in this thesis define a large- R jet with $R=1.0$ and smaller track-jet with $R=0.2$. In order to minimize the effects from pileup, $R=0.2$ calo-jets with less than 5% of the large- R jet p_T are removed from consideration. b -tagging is then applied to the ghost-associated, $R=0.2$ track-jets.

Jet calibrations must take into account several compounding factors to accurately reflect the energy and direction of a jet [84]. Accounting for the energy offset from the effects of in-time and out-of-time pileup is the first correction to be applied. The jet is then made to point to the event's PV rather than the center of the detector (this doesn't alter the energy). An η -dependent jet-energy scale (JES) derived from $t\bar{t}$ MC simulation is then applied to the jet. The JES ECAL scale correctly measures the energy deposited by electromagnetic showers. However, since hadronic jets have a strong interaction¹⁶ with the HCAL, calibrating to the ECAL scale results in an underestimation of the jet energy. Another method for jet-energy calibration is known as a local-cluster (LC) calibration and can take the HCAL into account depending on shower depth. However, this type of calibration is not yet complete for Run 2 at the time of the data release used in this thesis. A comparison of the two calibration schemes is given in Figure 4.16. Additionally, the inverse of the response in each bin of this figure is the average JES correction. The Run 2 JES values and bias used in the analyses in this thesis are shown in Figure 4.17. An additional series of multiplicative corrections known as the Global Sequential Calibrations (GSCs) are also applied to the jet. To quote Ref. [87]:

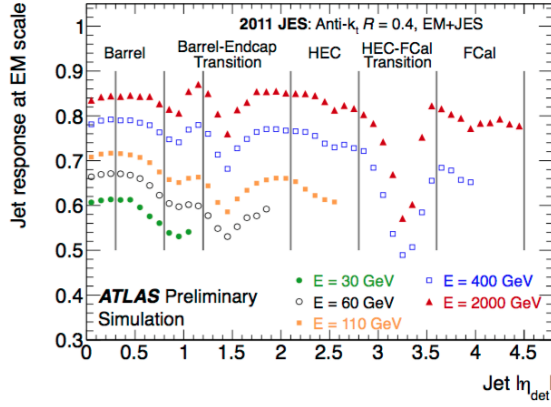
[GSCs] are based on global jet observables such as the longitudinal structure of the energy depositions within the calorimeters, tracking information associated to the jet, and information related to the activity in the muon chambers behind a jet.

Tracking information is used to reduce the flavor dependence of the calorimeter

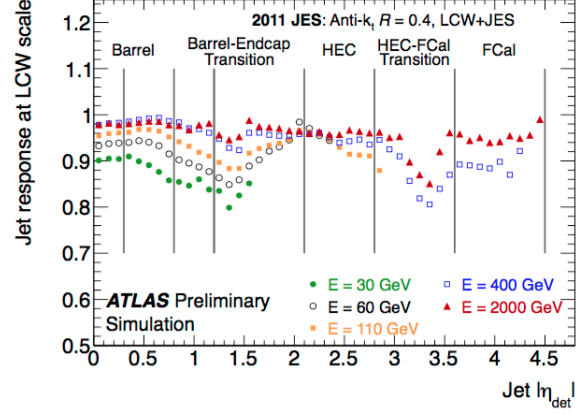
¹⁴Ghost-matching is a technique used to associate an object with a jet. The object's p_T is set to zero (or a negligible value) but retains its direction. The jet clustering algorithm is then rerun and if the object is clustered into the jet it's said to be ghost-associated.

¹⁶See what I did there...¹⁷

¹⁷Attempted humor circa 2016

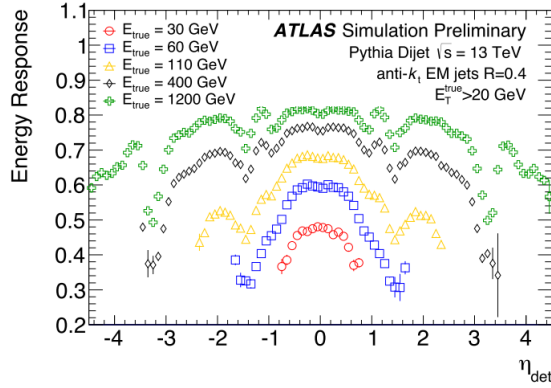


(a)

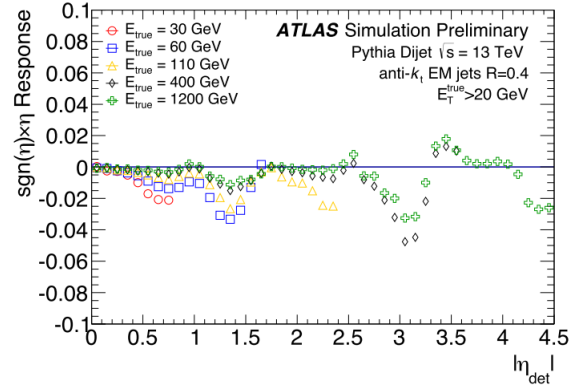


(b)

Figure 4.16: Run 1 average jet response for jets calibrated at the (a) EM scale and (b) LCW scale w.r.t. truth jet¹⁵ energy ($E_{\text{jet}}^{\text{EM/LC}}/E_{\text{jet}}^{\text{truth}}$) [84]. The quantity η_{det} is the jet direction as measured from the center of the detector (i.e. before correcting the direction of the jet).



(a)



(b)

Figure 4.17: Run 2 EMTopo jets (a) average jet response and (b) its resulting bias as a function of η_{det} [86].

response to jets, i.e. the difference in the observed calorimeter response to jets initiated by different parton flavors. Calorimeter energy deposit and, in the case of high energy jets penetrating the full depth of the calorimeter, muon spectrometer information are used to improve the jet energy resolution.

A final *in-situ*, data-driven correction based on the jet p_T and derived with γ/Z +jets and multi-jets events is then applied. Uncertainties from these calibrations are numerous (~ 70) and include effects from punch-through (jets not being fully absorbed by the HCAL), modeling uncertainties from the *in-situ* technique, pileup conditions, and many more. ATLAS analysis groups are given the option to reduce the number of systematic uncertainties by instead using reduced sets wherein systematics dominant in certain kinematic regions are merged. This can compress the number of systematics down to a more reasonable 18 or even 3 in the most extreme cases [86].

Jets reconstructed in MC simulation must also have their jet-energy resolution (JER) “smeared” to more accurately reflect jet reconstruction in real data. These smearing factors account for several effects detailed in Ref. [86] and are shown in Figure 4.18. This smearing also introduces additional systematics that must be taken into account by analyses.

4.2.6. Hadronic τ -leptons

Hadronic decay of τ -leptons (τ_{had} , sometimes referred to as τ -jets) are characterized by having one or three charged pions (“prongs”) along with a ν_τ and possibly neutral pions [17]. The reconstruction of visible τ decays uses the anti- k_t jet reconstruction algorithm with $R = 0.4$ [88]. The $\tau_{\text{had-vis}}$ p_T is set to the total energy of topological clusters in the calorimeter within $\Delta R < 0.2$. Tracks associated to the $\tau_{\text{had-vis}}$ are required to be within $\Delta R < 0.2$ around the τ direction. BDT-based¹⁸ τ identification is used to reject the background from jets. Derivations of calibrations are detailed in Ref. [88] and will be omitted from this section as τ -leptons are only used to exclude events and are not part of any relevant final state.

¹⁸The boosted decision tree algorithm is discussed a greater detail in Chapter 7.

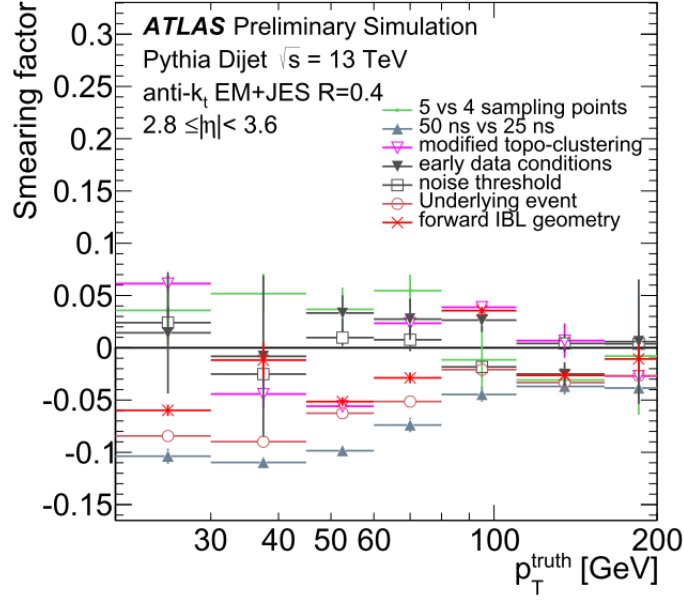


Figure 4.18: Different sources that contribute to the smearing of JER in MC simulation [86]. Details of these sources are given in the aforementioned reference.

4.2.7. Flavor-Tagged Jets

The flavor tagging of jets generally refers to assigning a “flavor” to a jet that represents whether it came from the production of a specific type of quark or gluon. Of the many types of flavor tagging the two that are most relevant are so-called truth tagging and b -tagging. Truth tagging is procedure for assigning a flavor label to a truth jet as described in the following section. b -tagging is a more generic form of flavor tagging wherein a jet is assigned a flavor label (b tag) from an analysis of the properties of the jet available to both MC simulation and data. To this end, some notable properties of a b -initiated jet and of B -mesons (a meson containing a b -quark) are [17]:

- mean lifetime (τ) of 1.5 ps
- mass of about 5 GeV
- $c\tau \approx 0.5$ mm
- 20% decay to lepton + neutrino + anything
- 90% decay to a charm meson

This decay length ($c\tau$) is appreciable enough to be detected as a secondary vertex (SV) since the vertex resolution is $23\,\mu\text{m}$ (r - ϕ plane) by $40\,\mu\text{m}$ (z axis) [89, 90]. These properties give b -jets a distinct signature that are utilized by algorithms.

4.2.7.1. Truth Flavor Labels

Truth-level flavor labelling of a jet in MC simulation is accomplished through the following procedure (after selecting an R_j cone-matching parameter):

1. If a simulated weakly decaying b -hadron is found within $\Delta R < R_j$ of the reconstructed jet direction, the jet is labeled a b -jet.
2. If a b -hadron isn't found, then if a weakly decaying c -hadron is found within $\Delta R < R_j$ of the jet direction, the jet is labeled a c -jet.
3. Otherwise, if a τ -lepton is found within $\Delta R < R_j$ of the jet direction, the jet is labeled a τ -jet.
4. If any hadron/ τ -lepton matches more than one jet, the closest jet is chosen.
5. All unlabeled jets after steps 1 through 4 are labeled light-jets.

For the analyses in this thesis $R_j = 0.3$.

4.2.7.2. b -Tagging

The jets from the Higgs boson in the signal processes relevant for this thesis should originate from b -quarks. Therefore, it's desirable to be able to distinguish b -jets from τ -, c -, and light-jets. A diagram of a prototypical b -jet decay is given in Figure 4.19. Since the signed track impact parameters for b -jets tend to be large and positive (as shown in Figures 4.20a and 4.20b), these parameters are useful in discriminating b - and c -jets from light-jets. The standard algorithms for b -jet tagging are the MV2c20 and MV2c10 BDT-based discriminants (for Athena releases before and after 20.7, respectively). These algorithms (referred to

Tagging Scheme	MV2c20 weight cut	b -tagging efficiency [%]	c RR	light RR
Anti- k_t $R=0.4$ EMTopo Jets				
FixedCutBEff_60	0.4496	60.27	21.32	1704.54
FixedCutBEff_70	-0.0436	70.26	8.03	405.37
FixedCutBEff_77	-0.4434	77.25	4.51	125.93
FixedCutBEff_85	-0.7887	85.22	2.06	27.16
Anti- k_t $R=0.2$ Track Jets				
FixedCutBEff_70	-0.3098	70.03	5.65	173.35
FixedCutBEff_77	-0.6134	77.05	9.86	71.77

Table 4.3: b -tagging working points and average performance (over p_T and η) available for MV2c20 (pre-Athena release 20.7). RR is the rejection rate (the inverse of efficiency) [94].

Tagging Scheme	MV2c10 weight cut	b -tagging efficiency [%]	c RR	light RR
Anti- k_t $R=0.4$ EMTopo Jets				
FixedCutBEff_60	0.9349	60.03	34.54	1538.78
FixedCutBEff_70	0.8244	69.97	12.17	381.32
FixedCutBEff_77	0.6459	76.97	6.21	134.34
FixedCutBEff_85	0.1758	84.95	3.10	33.53
Anti- k_t $R=0.2$ Track Jets				
FixedCutBEff_70	0.6455	70.00	7.09	119.69
FixedCutBEff_77	0.3706	77.00	8.09	57.90

Table 4.4: b -tagging working points and average performance (over p_T and η) available for MV2c10 (Athena release 20.7). RR is the rejection rate (the inverse of efficiency) [94].

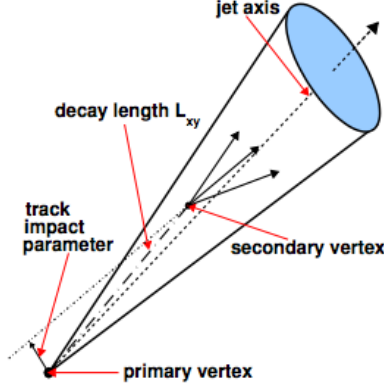


Figure 4.19: Prototypical b -jet decay [91]. Since the backtrace of the middle track within the SV falls behind the PV along the jet axis, this track would have a negative transverse impact parameter w.r.t. b -tagging.

collectively as MV2cXX) use log-likelihood ratios from IP2D and IP3D¹⁹, secondary vertex information from SV1²⁰, information about the relationship between the primary and secondary vertices from JetFitter²¹, and the p_T and η of the jet [92, 95, 96]. Both algorithms are trained on 5 million $t\bar{t}$ events with a mixture of (100-XX)% light- and XX% c -jets as background. This additional c -jet contamination gives better c -jet rejection with minimal degradation of light-jet rejection. The BDT output of MV2c10 for light-, c -, and b -jets is given in Figure 4.21. Practical usage of MV2cXX involves either selecting jets based on a fixed working point (i.e. a specific BDT output value) or requiring a constant efficiency or purity over the entire kinematic range of the jet (efficiencies are parameterized in jet p_T and η). Calibrations are only available for the fixed working point and constant efficiency usage schemes. However, in practice the performance of the fixed MV2cXX working point is close to a constant light- and c -jet rejection scheme. The available b -tagging efficiencies for the two algorithms are shown in Tables 4.3 and 4.4.

¹⁹IP2D uses transverse impact parameter information, while IP3D additionally uses the longitudinal impact parameter.

²⁰SV1 is a high-performance likelihood ratio technique that uses the decay length significance, the invariant mass of all tracks associated to 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 to find an SV.

²¹JetFitter is a likelihood technique that exploits the topology of weak b - and c - hadron decays is used to find a common line on which the PV and the b - and c -vertices lie using a Kalman filter.

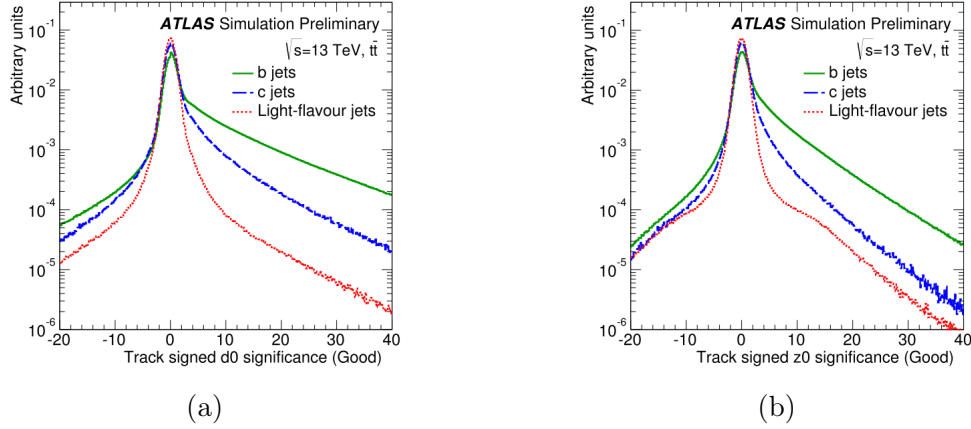


Figure 4.20: (a) Distribution of the signed transverse impact parameter significance $d_0/\sigma(d_0)$ [92] in simulated $t\bar{t}$ events. (b) Distribution of the signed longitudinal impact parameter significance $z_0/\sigma(z_0)$ [92]. Notice the bias of the b -jet distributions toward positive values in (a) and (b). The label “Good” refers to track quality requirements found in the noted reference.

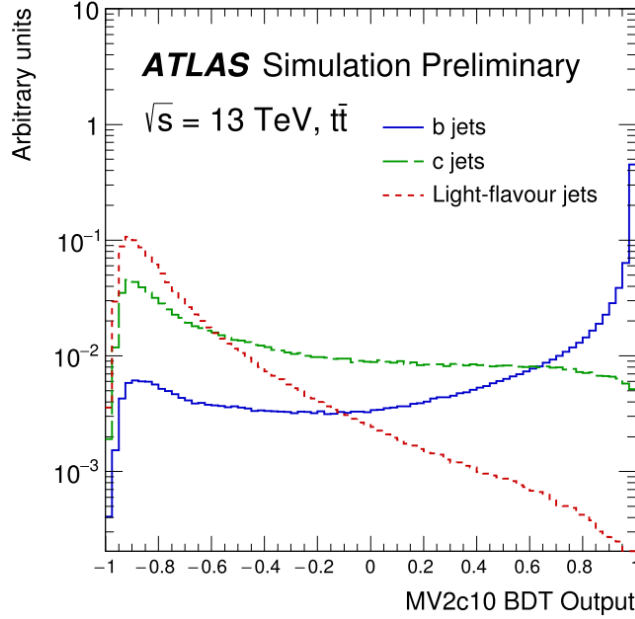


Figure 4.21: BDT output for the MV2c10 algorithm [93] in simulated $t\bar{t}$ events. Note that the b -jets are well separated from the light- and c -jets.

Calibrations for jet flavor-tagging are available in the form of efficiency scale factors (see Equation 4.2). The calibrations are derived on leptonic $t\bar{t}$ events (i.e. where the tops decay to a W +quark and the W then decay to an e or μ and a neutrino) as the purity of such events is quite high and the presence of b -initiated jets is almost certain. An eigenvalue-decomposition takes advantage of the correlations in p_T and is an area to which I made significant contributions. The methods I developed for the smoothing of the scale factors and reduction of systematics is detailed in Appendix B.

4.2.8. Overlap Removal

Before the reconstruction of missing transverse momentum, a procedure for removing objects that geometrically overlap must be applied as the algorithms used to reconstruct electrons, muons, and jets are “greedy” (i.e. they generally don’t take other species of object into account). Thus, to avoid double counting objects created from the same tracks and calorimeter deposits within an event, well-established rules for object priority are applied for removing objects in close proximity to one another. This is known as overlap removal. In practice analyses use a standard interface to perform overlap removal and only select from a predefined list of removal priority options (i.e. which objects to remove or keep). The procedure is as follows (order matters) [97]:

- remove any calorimeter-tagged muons sharing an ID track with an electron
- remove any electrons sharing an ID track with remaining muons
- reject jets within $\Delta R < 0.2$ of any remaining electron
- reject electrons within $\Delta R < 0.4$ of remaining non-pileup jets (i.e. ($p_T < 60$ GeV, $|\eta| < 2.4$ and the Default JVT requirement) or ($p_T \geq 60$ GeV and $|\eta| < 2.5$))
- reject jets satisfying both:
 - within $\Delta R < 0.2$ or ghost-matched to a muon
 - less than three associated tracks or $(p_T^{\text{muon}}/p_T^{\text{jet}} > 0.5$ and $p_T^{\text{muon}}/\Sigma_{p_T}^{\text{tracks in jet}} > 0.7)$

- reject muons within $dR < 0.4$ of remaining non-pileup jets

One important note is that overlap removal is not applied to a jet if it's b -tagged. This is done to prevent semi-leptonic b -decays from being unduly removed.

4.2.9. Missing Transverse Energy and Momentum

The missing transverse energy (MET) and missing transverse momentum (MPT) play key roles in two of the signal channels relevant for this thesis: those consistent with events containing $Z \rightarrow \nu\nu$ and $W \rightarrow l\nu$, respectively. In either case, the neutrino is not directly detectable and, thus, appears only as an imbalance in the total transverse energy and momentum of the vector sum of all the objects in an event. Transverse momentum is defined as the components of momentum in the plane perpendicular to the beam axis while transverse energy is nominally made from the perpendicular component of the vector sum of all calorimeter deposits. Analyses have the option to select between using the calorimeter-based jets (denoted E_T^{miss} and most resembles MET with TST) or track-based jets (denoted $E_{T,\text{trk}}^{\text{miss}}$ and most resembles MPT). MET is reconstructed as of the negative vector sum of p_T of physics objects (leptons, jets, etc.) and a so-called soft term. This track-based soft term (TST) [98] incorporates good-quality tracks from additional sources other than the hard terms arising from photons, electrons, muons, τ -leptons, and jets. Since these quality tracks are associated with the primary vertex, E_T^{miss} constructed with an additional TST term is more robust against pile-up. However, $E_{T,\text{trk}}^{\text{miss}}$ more naturally provides a robust estimate of the MET while being less sensitive to effects from pileup than E_T^{miss} . However, since it's mostly based on tracks left by charged particles it cannot account for the momenta of neutral particles.

MET is calibrated using $Z \rightarrow \mu\mu$ events in data and MC simulation. These events should contain very little, if any, MET and so provide an excellent test of the robustness of MET reconstruction. A measure of the MET scale is the mean projection of the MET along the axis of the reconstructed Z boson, $\langle E_T^{\text{miss}}, \mathbf{A}_Z \rangle$ ($\mathbf{A}$ is the unit vector along the transverse component of the Z boson momentum). This quantity is sensitive to the balance between

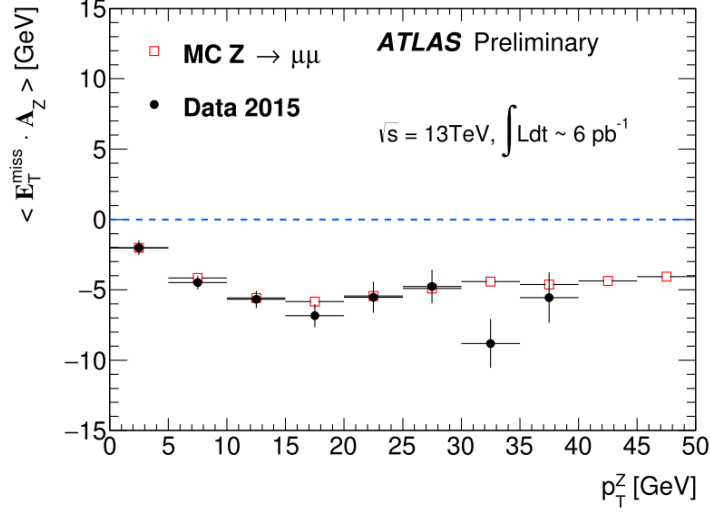


Figure 4.22: MET scale as a function of $Z p_T$ [98]. The negative bias at 5 GeV is due to an underestimation of the TST (the limit acceptance of the ID and its exclusion of neutral, soft particles).

the hard objects and potential recoil of the initial-state hadron due to radiation (known as soft hadronic recoil²²) [98].

An additional variable called METHT, defined as,

$$\text{METHT} = \frac{E_T^{\text{miss}}}{\sqrt{\sum p_T^e + \sum p_T^\mu + \sum p_T^{\text{jet}}}}, \quad (4.3)$$

can be used in event selection. It is the MET divided by the (scalar) p_T sum-in-quadrature of the hard objects in the event. Since the denominator in Equation (4.3) is proportional to the MET resolution, cutting on METHT is equivalent to making harder cuts on better-resolved MET and looser cuts on worse-resolved MET.

4.2.10. L1 Trigger

The primary role of the L1 trigger is to unambiguously and expediently find interesting bunch crossings²³. Practicalities such as the times of flight for muons (owning to the

²²Soft hadronic recoil refers to initial state QCD radiation that can cause a systematic bias in i.e. the Z -boson mass measurement in $Z \rightarrow \mu\mu$ if you naively assume that the Z boson lacks any initial transverse momentum.

²³All information in this section is taken from Ref. [41]

magnitude of the MS) and the width of the calorimeter signals (hinted at in Section 4.2.3) all contribute to the fact that the event data stream must be held temporarily in pipeline memories that are made of custom electronics that must be radiation hard and reliable. The approximate latency of the L1 trigger is roughly $2.5\text{ }\mu\text{s}$ in which $1\text{ }\mu\text{s}$ is needed for the signals to traverse cables and another $0.5\text{ }\mu\text{s}$ is needed as a buffer. The L1 trigger is designed to look for high- p_T muons, “egamma” objects²⁴, jets, and hadronic τ ’s as well as the presence of large amounts of MET and total transverse energy. As alluded to in Chapter 3, only the RPCs and TGCs of the MS system are involved in L1 muon trigger decisions while the entire calorimeter (both EM and hadronic) is employed for all other L1 triggers. The final stage of the L1 trigger is devoted to constructing so-called Regions of Interest (RoIs) that are used as indicators to the HLT of which regions of the detector warrant further investigation.

4.2.10.1. Calorimeter

All L1 trigger objects other than muons are reconstructed in the calorimeters in special hardware components collectively called the L1Calo. The L1Calo has roughly 7000 channels from using towers of size 0.1×0.1 in $\Delta\eta \times \Delta\phi$. These towers are used in a sliding-window-like clustering algorithm that quickly locates isolated regions of calorimeter activity above a given threshold (a brief overview of which is given in Figure 4.23). Multiple E_T thresholds for each object are available in the L1Calo depending on the specific bunch crossing. The L1Calo can also enforce isolation requirements on objects that are made of cuts along the ϕ “axis.” Electrons and jets at the L1 trigger level are made solely from ECAL energy deposits. These deposits are made from coarser clusters than during reconstruction due to time constraints.

4.2.10.2. Muon

The RPCs and TGCs in the MS provide L1 triggering capabilities in the the barrel and endcap regions, respectively. These two detector subsystems look for patterns consistent with high- p_T muons. The L1 muon trigger can, like the L1Calo, adapt its thresholds to

²⁴Since the L1 trigger lacks ID access, electrons and photons are indistinguishable. Thus, calorimeter deposits are referred to as egamma.

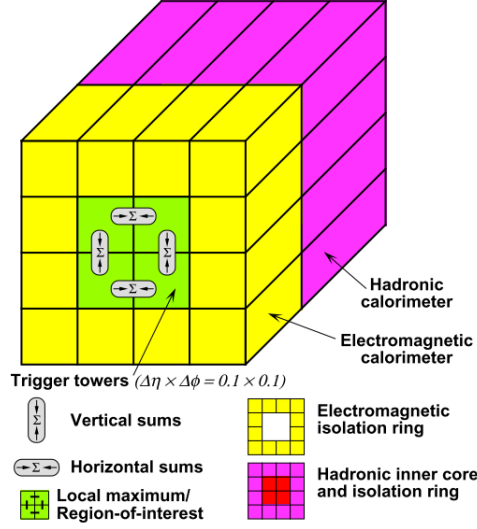


Figure 4.23: Diagram of the L1Calo clustering algorithm. This algorithm identifies 2×2 tower clusters in which at least one of four possible two-tower sums of nearest-neighbor electromagnetic towers exceeds a pre-defined threshold. Isolation-veto thresholds are set for the 12-towers surrounding ring in the electromagnetic calorimeter, as well as for the 2×2 hadronic-tower core sum behind the cluster and the 12-tower hadronic ring around it. (Description taken from Ref. [41].)

the beam crossing conditions. Importantly, muons are not double counted across different triggering thresholds. L1 muon reconstruction is a drastically scaled down version of its offline counterpart.

4.2.11. HLT Trigger

The HLT trigger uses the entire readout of the detector. Its latency of about 1 s allows for the use of sophisticated, offline-like reconstruction algorithms. For a detailed discussion on how this pertains to the the b -jet trigger, see Appendix A. Since most of the reconstruction is quite similar to what has already been discussed earlier in this chapter, only tracking will be outlined in this section as this is a major component missing from the L1 trigger that is present in the HLT.

4.2.11.1. Tracking

Tracking in the trigger is only done at the HLT level as ID information isn't available at L1. The HLT track reconstruction algorithms mimic their offline counterparts. However, these algorithms are modified to run efficiently within the timeframe of the HLT (~ 1 s per event) and are initially run only based on the presence of MS activity or calorimeter RoIs. A visual comparison of offline versus HLT track reconstruction is given in Figure 4.24.

4.2.12. Trigger Menu Items

The final trigger menu item that an analysis sees usually (but not always) takes the form “HLT_<underscore-separated HLT objects>_L1<underscore-separated L1 objects>.” Any L1 objects that are listed are dependencies of the HLT items. The relevant L1 trigger object descriptions are “EM” (egammas), “MU” (muons), “J” (jets), and “XE” (E_T^{miss}). Since the L1 trigger operates almost exclusively with transverse energy, most L1 object descriptions are followed by a minimum E_T threshold. “DPHI” represents a $\Delta\phi$ requirement on the two objects that follow this specifier. For instance, “L1J40_DPHI-J20XE50” means an L1 jet of minimum transverse energy 40 GeV with a minimum separation in $\Delta\phi$ between an L1 jet of minimum transverse energy 20 GeV and E_T^{miss} of 50 GeV. The HLT object descriptions are similar to those in the L1 trigger except the HLT uses lowercase rather than all uppercase. Because the HLT reconstructions tracks, the egamma objects from L1 are able to be resolved to either being electrons (“e”) or photons (“g”). An example of a full menu item that is used in the analyses in this thesis is “HLT_j100_xe80_L1J40_DPHI-J20XE50.” This is read as the same L1 item as described earlier as a prerequisite for the presence of both a 100 GeV HLT jet and 80 GeV of E_T^{miss} .

4.3. Summary

This chapter has outlined the major components of simulation and reconstruction which serve as the foundation of the next two chapters. Object selection depends critically on how accurately the objects are reconstructed. The simulation process also determines many

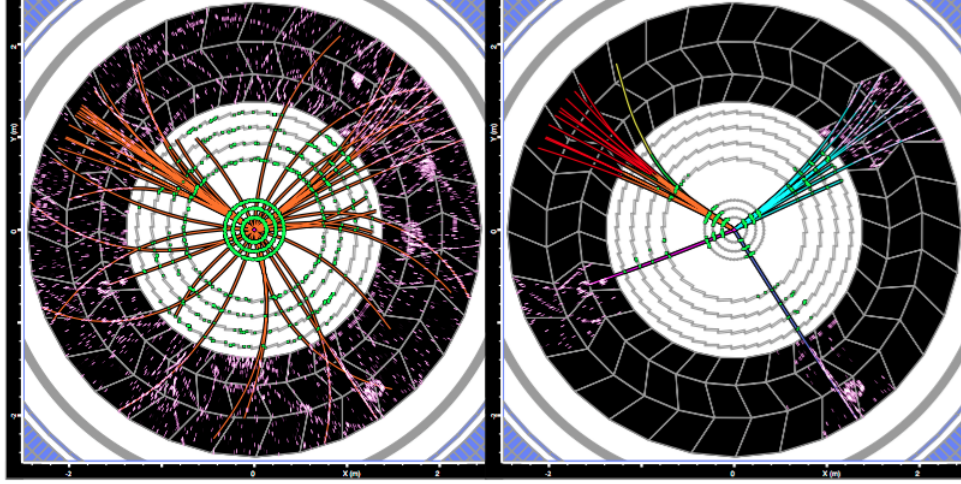


Figure 4.24: Comparison of offline tracking (left) to HLT tracking (right) during Run 1 [64]. Note that the tracking is initiated by the presence of the RoIs in the ECAL (not shown).

aspects of both object and event selection of different simulation schemes can lead to substantially different event selection outcomes.

Chapter 5

Simulated & Detector Datasets

MC simulated events are used to compare expected yields of background and signal processes to that in data. The various processes along with their respective generators are given in Table 5.1. To determine competing background processes to a signal process, one must look at processes that produce a similar final state to the signal and in what quantity those processes occur (i.e. cross sections). To recap, the signal processes for the two analyses in this thesis have similar final states. The two analyses both have $\nu\nu b\bar{b}$ and $llb\bar{b}$ final states. The SM analysis has an additional $l\nu b\bar{b}$ final state. Thus, the prominent background processes are t -quark pair-production where each t -quark decays to a W boson and b -quark ($t\bar{t} \rightarrow WbW\bar{b}$, as the W -bosons can decay to leptons and E_T^{miss}), single top-quark production (for similar reasons as $t\bar{t}$), single vector bosons produced in association with hadronic jets (the jets will occasionally “look” like (fake) a Higgs boson decaying to two b -quarks), pairs of vector bosons (one of them may fake a Higgs boson), and multiple-jet final states (also known as QCD multi-jet). The t -quark pair-production process is commonly referred to as “ $t\bar{t}$ bar” or $t\bar{t}$ and the vector-boson-in-association-with-jets productions is commonly abbreviated as simply V +jets. An overview of the simulated and real detector datasets used for the two analyses in this thesis is given in this chapter. This includes discussion of the generators used, cross sections computed, and tuning that the simulated datasets use. All information in this chapter is taken from Refs. [100, 101] unless otherwise noted.

Before moving on to the specific descriptions of the datasets, it’s important to note that each MC event must be reweighted in order to make accurate comparisons to data. This involves rescaling the total MC yield to the measured luminosity, weighing each MC event by a factor that accounts for the differences in simulated and actual detector conditions (PU weight [103, 104]), and MC generator weights that are often found in physics processes

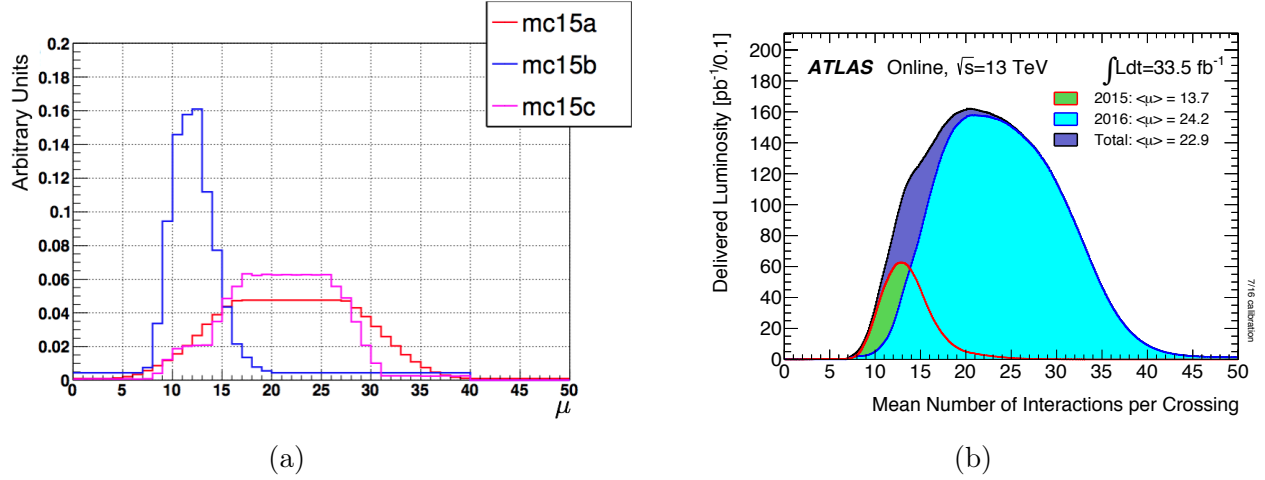


Figure 5.1: Comparison of pileup (μ) distributions (a) between Monte Carlo simulation sets MC15a, MC15b and MC16c [99] (normalized to unit area) and (b) the corrected luminosity for the complete 2015 & 2016 ATLAS datasets [102].

where different amplitudes contribute to a given process and account for possible interference. Samples with interference may have events with negative weights, which are essential in correctly representing physical distributions affected by the quantum interference. The rescaling of the total MC yield is straightforward once the luminosity is obtained from an official ATLAS Luminosity Calculator [105]. The PU reweighting is accomplished through an official ATLAS tool that provides corrections of MC events that account for differences in pileup distributions between MC and data shown in Figure 5.1.

5.1. MC Simulated Datasets

5.1.1. $t\bar{t}$

For the generation of $t\bar{t}$, the POWHEG v2 generator with the CT10 [106] PDF set is used in the matrix-element calculations along with EVTGEN. The parton shower, fragmentation, and the underlying event are simulated using PYTHIA 6.428 with the CTEQ6L1 PDF set and the corresponding Perugia 2012 tune (P2012) [107]. The top-quark mass is set to 172.5 GeV. Events are filtered such that at least one W boson in the event decays leptoni-

cally. The samples are normalised to their next-to-NLO(NNLO) cross-section including the resummation of soft gluon emission at next-to-leading-logarithmic accuracy using Top++ 2.0 [108]. Samples produced with alternative generators are used to assign uncertainties on the $t\bar{t}$ background. Alternative samples are generated with varying schemes: different parton showering generator (POWHEG+HERWIG++), alternative matrix-element generator (MADGRAPH5_AMC@NLO+HERWIG++) and varying the choice of generator parameters to increase or decrease the amount of radiation. The $A \rightarrow Zh$ analysis uses $t\bar{t}$ samples generated under entirely SM assumptions (i.e. that no other 2HDM particles play a role in the production of the top quark or its decay).

5.1.2. Single Top

Wt -associated and s -channel single top-quark production uses the POWHEG v2 generator with the CT10 PDF set for in the matrix-element calculations. Electroweak t -channel single top-quark events are generated using the POWHEG v1 generator. This generator uses the 4-flavor scheme for the next-to-leading order (NLO) matrix-element calculations together with the fixed 4-flavor PDF set CT10f4. The parton shower, fragmentation, and the underlying event are simulated using the same generator, PDFs, and tune as the $t\bar{t}$ samples. The top-quark mass is set to 172.5 GeV and events are filtered such that at least one W boson in the event decays leptonically as in the $t\bar{t}$ samples. The s -, t -, and Wt -channel samples are normalized to cross-sections using NLO calculations in QCD using Hathor v2.1 [109]. The same alternative generators that are used in producing the alternative $t\bar{t}$ samples are used in single-top production.

5.1.3. V +Jets

Events containing W or Z bosons with jets (V +Jets) are simulated using the SHERPA 2.2.0 generator. Matrix elements are calculated for up to two partons at NLO and four partons at leading order (LO) using the COMIX [110] and OPENLOOPS [111] matrix-element generators and merged with the SHERPA parton shower using the ME+PS@NLO prescrip-

tion [112]. The CT10 PDF set is used in conjunction with a dedicated parton shower tuning developed by the SHERPA authors. The V +Jets samples are produced with a simplified scale-setting prescription in the multi-parton matrix elements, to improve the event generation speed. A theory-based reweighing of the jet multiplicity distribution is applied at event level, which is derived from an event generation with the strict scale prescription. The V +Jets events are normalized to the NNLO QCD theoretical cross sections. Alternative versions of the SHERPA samples are generated to evaluate modeling uncertainties, where the factorization, renormalisation, CKKW merging¹ and re-summation scales have been varied. These samples are generated using SHERPA 2.1.1 but it has been verified that the relative variations due to the scale variations are consistent with those found for SHERPA 2.2.0. Alternative samples are also generated using MADGRAPH5_AMC@NLO v2.2.2 at LO interfaced to the PYTHIA 8.186 parton shower model. The A14 [115] tune is used together with the NNPDF23LO PDF set.

Simulated V +Jets samples are labelled according to the flavor label of the two jets chosen for the reconstructed Higgs boson candidate (cover in the next chapter): $V+bb$, $V+bc$, $V+cc$, $V+bl$, $V+cl$, $V+l$ (short for $V+ll$). An inclusive category V +HF is defined as containing the first four jet labels: $V+bb$, $V+bc$, $V+cc$, $V+bl$.

5.1.4. Di-Boson

Di-boson processes (WW , WZ , and ZZ) are generated using SHERPA 2.1.1 which calculates up to one (ZZ) or zero (WW , WZ) additional partons at NLO and up to three additional partons at LO. These samples are normalized to cross sections calculated at NLO. Alternative SHERPA samples are generated with the factorization, normalization and resummation scales varied. Additional di-boson samples are produced using the POWHEG v2 generator, interfaced to either the PYTHIA 8.2 or the HERWIG++ parton shower generators.

¹CKKW merging refers to the CKKW matrix element-parton shower merging algorithm wherein tree-level matrix elements are combined with parton showering algorithms, for a given process, for all parton multiplicities below some maximum N [113, 114]. The basic approach is to divide the phase space for parton emissions into two regions by a merging scale y_{MS} , defined by either the jet resolution or jet p_{T} . Above the merging scale, emissions are described by exact matrix elements while below it emissions are produced by the parton shower.

The CT10NLO PDF set is used for the hard process, while the CTEQ6L1 PDF set is used for the parton shower.

5.1.5. QCD Processes

Soft QCD used in minimum bias events are simulated with PYTHIA 8.186 with an A2 tune [116] and using the MSTW2008LO PDF [117].

5.1.6. VH

Simulated events for the SM $qq \rightarrow VH$ production are generated at leading order for the hard-scatter and leading-log for the parton shower using PYTHIA 8.186, using the A14 tune and NNPDF23LO PDF set. The samples for $gg \rightarrow ZH$ (gluon-induced) production are simulated using the POWHEG v2 matrix-element generator interfaced with PYTHIA 8.2, using the AZNLO [118] tune with CT10NLO PDF sets. The mass of the Higgs boson is fixed to 125 GeV, its width treated as narrow compared to the resolution of the detector. The inclusive cross sections are calculated at NNLO (QCD) and NLO (EW). For the gluon-induced ZH production the cross section is calculated at NLO+NLL (QCD). This is then subtracted from the inclusive cross section to estimate the quark-induced cross section. An additional NLO (EW) scale factor is applied as a function of the vector boson transverse momentum to take into account effects not already modeled in the simulation. This makes use of the VH differential cross section computed with HAWK [119]. The simulated samples include all final states where the Higgs boson decays to $b\bar{b}$ and the vector boson to a leptonic final state, including those with a τ -lepton.

5.1.7. ggA & $b\bar{b}A$

Samples for the $A \rightarrow Zh$ signal process are generated for gluon fusion and b -quark associated production using MADGRAPH5_AMC@NLO 2.2.2 and NNPDF2.3LO PDFs. Parton showering and hadronization are performed with PYTHIA 8.186 using the NNPDF2.3LO PDFs. The SM value of the $h \rightarrow b\bar{b}$ branching fraction is $58 \pm 2\%$, but in assessing con-

straints on the 2HDM this is controlled by the parameters of the model. The mass of the A boson, m_A , is varied between 220 GeV and 2000 GeV and the narrow-width approximation is used in the simulation. When considering the constraints on 2HDMs, the A boson width is allowed to deviate from the narrow-width approximation, as described in Chapter 8. The production cross section is arbitrarily set to 500 fb, but after fitting to data the normalization is restricted to its 95% confidence level (CL) upper limit (described later) in the plots shown in Chapter 8. This is done purely for the purpose of allowing the reader to see what a signal at this level would look like in the ATLAS data.

5.2. ATLAS 2015 & 2016 Datasets

The datasets used in this analysis include only p - p collision data recorded in stable beam conditions and with all relevant sub-detectors providing high-quality data. Both analyses use data recorded at a center-of-mass energy of $\sqrt{s} = 13$ TeV. The integrated luminosity of the 2015 dataset for both $A \rightarrow Zh$ and SM VH analyses is $3.2 \pm 0.1 \text{ fb}^{-1}$. The SM VH analysis uses an additional $10.0 \pm 0.4 \text{ fb}^{-1}$ of data collected in 2016. High data quality is maintained by pruning events through a Good Runs List (GRL [120]) XML file provided by the ATLAS Data Quality Group. This list uses data-quality status flags from specific areas of the detector to determine if an event within a given luminosity block² is suitable for physics analysis. In addition to using a GRL, events corrupted from bad tile calorimeter and LAr noise bursts are removed as well. The GRL is necessary to determine the correct amount of luminosity processed. A Luminosity Calculator is also provided that computes the total integrated luminosity for a given GRL XML file [121]. Data event weights may also be adjusted by an official pileup reweighing tool [122] to account for trigger pre-scales (although, the analyses in this thesis don't use triggers that require pre-scales). A more accurate estimate of the mean number of interactions per crossing $\langle\mu\rangle$ can also be obtained from this tool.

²The term luminosity block refers to the fact data is collected and recorded according to the average luminosity during a given temporal period.

5.3. Discussion

The datasets described in this chapter serve as the foundation for both the SM VH and BSM $A \rightarrow Zh$ analyses in this thesis. As shown in the previous chapters, both theoretical and practical issues prevent a complete, *ab initio* model of what we should expect to see in the detector. To this end, the MC simulated datasets act as the theoretical model when comparing to detector data. However, even though these simulated datasets are carefully constructed to produce the desired final state of particles, they require a virtually identical analysis procedure to that of actual detector data as simulated datasets are produced in such a way that tries to mimic observed data. This fact is reflected in the following chapters that detail the analysis procedures.

Process	Generator
SM Signal	
$q\bar{q} \rightarrow ZH \rightarrow \nu\bar{\nu}b\bar{b}/l\bar{l}b\bar{b}$	PYTHIA 8 + EVTGEN
$gg \rightarrow ZH \rightarrow \nu\bar{\nu}b\bar{b}/l\bar{l}b\bar{b}$	POWHEG + PYTHIA 8 + EVTGEN
$q\bar{q} \rightarrow WH \rightarrow l\bar{\nu}b\bar{b}$	PYTHIA 8 + EVTGEN
ggA $\rightarrow ZH$ Signal	
$l\bar{l}b\bar{b}, \nu\bar{\nu}b\bar{b}$, all masses	MADGRAPH + aMC@NLO + PYTHIA 8 + EVTGEN
$b\bar{b}A \rightarrow ZH$ Signal	
$l\bar{l}b\bar{b}, \nu\bar{\nu}b\bar{b}$, all masses	MADGRAPH + aMC@NLO + PYTHIA 8 + EVTGEN
QCD	
multi-jet	PYTHIA 8 + EVTGEN
Vector boson + jets	
$(W \rightarrow l\bar{\nu}) + \text{jets}$	SHERPA
$(Z \rightarrow l\bar{l}) + \text{jets}$	SHERPA
$(Z \rightarrow \nu\bar{\nu}) + \text{jets}$	SHERPA
Top-quark	
$t\bar{t}$	POWHEG + PYTHIA 6 + EVTGEN
single top t -channel	POWHEG + PYTHIA 6 + EVTGEN
single top s -channel	POWHEG + PYTHIA 6 + EVTGEN
single top Wt -channel	POWHEG + PYTHIA 6 + EVTGEN
Diboson	
WW	SHERPA
WZ	SHERPA
ZZ	SHERPA

Table 5.1: List of simulated signal and background samples for the SM and $A \rightarrow Zh$ analyses along with their respective generators [99].

Chapter 6

Object and Event Selections

In the context of HEP, object selection refers to the process of searching for physical constructs (i.e. muons, electrons, jets, etc.) reconstructed from detector information that satisfy basic detector quality and CP¹ requirements. The overarching motivation for object selection is the event selection. Event selection is the process of searching for events (defined in Chapter 3) that match as closely as possible the “signal” topology using the physical objects selected during object selection. Object and event selections are two distinct but related steps in every analysis. They serve as the foundation of more sophisticated aspects of a measurement or search. They are largely driven by the topology of signal processes and detector realities. Hence, object and event selections for the SM VH and BSM $A \rightarrow Zh$ analyses are separately optimized depending on the number of leptons in each event as this reflects the presence of a W or Z boson. While the SM analysis contains an extra lepton channel, both selections are quite similar. Feynman diagrams for the different SM VH lepton channels are given in Figure 6.1. The analogous diagrams for the BSM $A \rightarrow Zh$ lepton channels are given in Figure 6.2. These event topologies are the driving force behind the object and event selections given in this chapter. In the following chapter details of the selections are given for both analyses and differences between the analyses are explicitly stated whenever necessary.

¹ Combined performance (CP) groups are a collection of experts in specific detector components or physics objects. Through careful analysis they make recommendations for use of these objects in analysis. For instance, there are electron, muon, and jet/ E_T^{miss} CP groups to name a few.

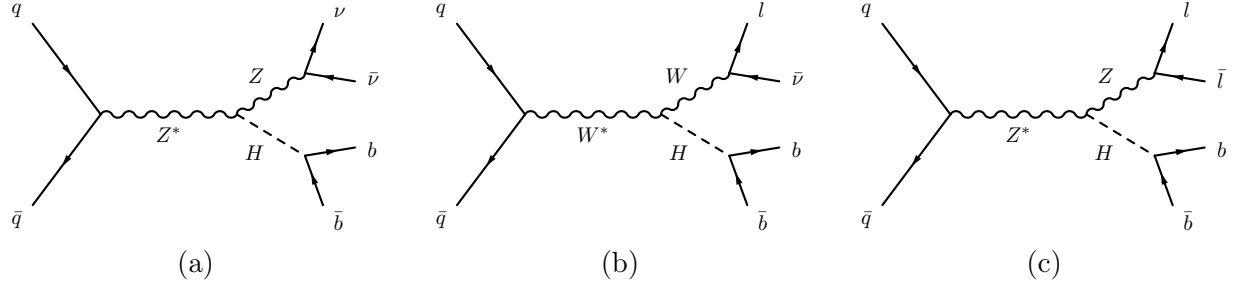


Figure 6.1: SM VH signal event topologies for the three lepton channels. In all channels the Higgs boson and vector boson are back-to-back in the transverse plane of the detector's frame of reference and the Higgs boson always decays to two b -quarks. Events are classified by the number of reconstructed charged leptons. The 0-, 1-, and 2-lepton channels are illustrated in (a), (b), and (c), respectively. Not shown are the $gg \rightarrow Z$ production diagrams.

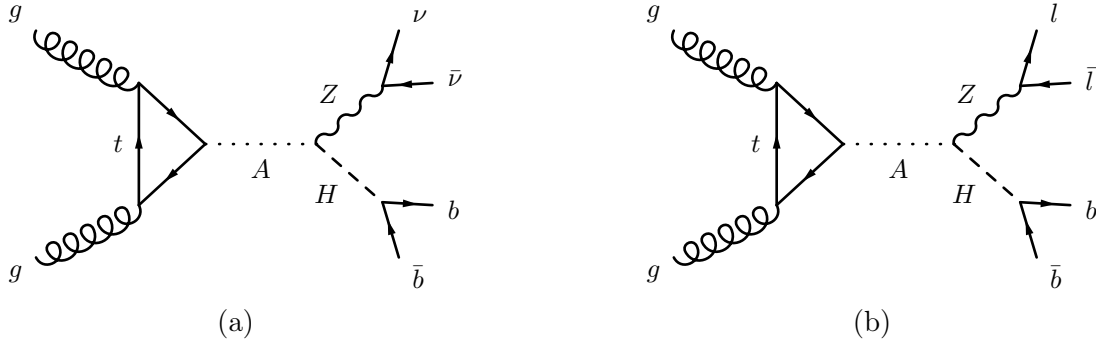


Figure 6.2: BSM $A \rightarrow Zh$ signal event topologies for the two lepton channels. In all channels the Higgs boson and Z boson are back-to-back in the transverse plane of the detector's frame of reference and the Higgs boson always decays to two b -quarks. The Zh system forms the resonant, A boson particle. Events are classified by the number of reconstructed charged leptons. The 0- and 2-lepton channels are illustrated in (a) and (b), respectively. Not shown are the $b\bar{b}A$ associated production diagrams.

6.1. Derivation-Level Cuts

Data and MC samples are originally stored in a large, comprehensive format known as “xAOD”² [123]. These samples are reduced in size on a per-analysis basis. These “derived” samples, or DxAODs, are produced centrally with the expressed goals of reducing sample sizes, adding analysis-specific variables, applying basic object and event selection, and using the latest CP recommendations. The derivations relevant to this analysis are “HIGG5D1” for the 0-lepton channel, “HIGG5D2” for the 1-lepton channel, and “HIGG2D4” for the 2-lepton channel. The list of selection criteria for these derivations are given in Tables 6.1, 6.2, and 6.3. All derivations have jet requirements consistent with either small- or large- R calorimeter jets. More notable features of these derivations are [99]:

- HIGG5D1 (0-lepton) has trigger requirements largely related to missing transverse momentum;
- HIGG5D2 (1-lepton) has trigger requirements related to missing transverse momentum or single leptons at varying purity and a lepton requirement for the presence of at least one lepton of low purity;
- HIGG2D4 (2-lepton) has trigger requirements for single leptons at all purities, as well as requirements for the presence of at least two leptons of low purity with minimum requirements on the di-lepton invariant mass.

6.2. Object Selection

Object selection defines the basic physics objects such as electrons, muons, and jets to be used in event-level selections³. Since much of the physics is different between the lepton channels, the definitions of leptons are vital. Leptons are separated into “Loose,” “Medium,” and “Tight” categories which vary in overall purity as summarized in Tables 6.4

²xAOD stands for **extensible Analysis Object Data**. This format contains (software) objects such as reconstructed muons as well as the reconstructed constituents of the physics objects such as tracks and calorimeter deposits.

³All information in the section is taken from Ref. [99]

Selection	Value
HIGG5D1	
Trigger	HLT_xe35 or HLT_xe50 or HLT_xe60 or HLT_xe70 or HLT_xe80 or HLT_xe80_L1XE50 or HLT_xe80_L1XE70 or HLT_xe100 or HLT_j80_xe80_dphi1_L1J40_DPFI-J20XE50 or HLT_j80_xe80_dphi1_L1J40_DPFI-J20s2XE50 or HLT_j100_xe80_L1J40_DPFI-J20XE50 or HLT_j100_xe80_L1J40_DPFI-J20s2XE50 or HLT_g120_loose
Jets	count((jet1 $p_T > 15$ GeV) and ($ \text{jet1 } \eta < 2.6$)) > 1 or count((jet1 $p_T > 100$ GeV) and ($ \text{jet1 } \eta < 2.6$)) > 0 or count((jet2 $p_T > 100$ GeV) and ($ \text{jet2 } \eta < 2.6$)) > 0 or count((jet3 $p_T > 100$ GeV) and ($ \text{jet3 } \eta < 2.6$)) > 0 or count((fatjet1 $p_T > 100$ GeV) and ($ \text{fatjet1 } \eta < 2.6$)) > 0 or count((fatjet2 $p_T > 100$ GeV) and ($ \text{fatjet2 } \eta < 2.6$)) > 0

Table 6.1: List of selection criteria for the HIGG5D1 DxAODs used in the analyses. The “count” function is simply shorthand for “number of objects satisfying the following criteria.” A list of the xAOD jet containers used is as follows: jet1=AntiKt4EMTopoJets, jet2=CamKt10LCTopoJets, jet3=CamKt12LCTopoJets, fatjet1=AntiKt10LCTopoTrimmedPtFrac5SmallR20Jets, and fatjet2=CamKt12LCTopoBDRSFilteredMU100Y15Jets.

Selection	Value
HIGG5D2	
Trigger	HLT_e17_lhloose_L1EM15 or HLT_e24_medium_iloose_L1EM18VH or HLT_e24_medium_iloose_L1EM20VH or HLT_e24_lhmedium_iloose_L1EM20VH or HLT_e24_tight_iloose or HLT_e26_tight_iloose or HLT_e26_tight1_iloose or HLT_e26_lhtight_iloose or HLT_e60_medium or HLT_e60_medium1 or HLT_e60_lhmedium or HLT_mu14 or HLT_mu20_iloose_L1MU15 or HLT_mu24_iloose_L1MU15 or HLT_mu24_imedium or HLT_mu26_imedium or HLT_mu50 or HLT_xe35 or HLT_xe50 or HLT_xe60 or HLT_xe70 or HLT_xe80 or HLT_xe80_L1XE50 or HLT_xe80_L1XE70 or HLT_xe100 or HLT_j80_xe80_dphi1_L1J40_DPHI-J20XE50 or HLT_j80_xe80_dphi1_L1J40_DPHI-J20s2XE50 or HLT_j100_xe80_L1J40_DPHI-J20XE50 or HLT_j100_xe80_L1J40_DPHI-J20s2XE50
Jets	count((jet1 $p_T > 15$ GeV) and ($ \text{jet1 } \eta < 2.6$)) > 1 or count((jet1 $p_T > 100$ GeV) and ($ \text{jet1 } \eta < 2.6$)) > 0 or count((jet2 $p_T > 100$ GeV) and ($ \text{jet2 } \eta < 2.6$)) > 0 or count((jet3 $p_T > 100$ GeV) and ($ \text{jet3 } \eta < 2.6$)) > 0 or count((fatjet1 $p_T > 100$ GeV) and ($ \text{fatjet1 } \eta < 2.6$)) > 0 or count((fatjet2 $p_T > 100$ GeV) and ($ \text{fatjet2 } \eta < 2.6$)) > 0
Leptons	count((Muon $p_T > 20$ GeV) and ($ \text{Muon } \eta < 2.6$) and (DFCommonGoodMuon)) > 0 or count((Electron $p_T > 20$ GeV) and ($ \text{Electron } \eta < 2.6$) and (DFCommonElectronsLHLoose)) > 0

Table 6.2: List of selection criteria for the HIGG5D2 DxAODs used in the analyses. The “count” function is simply shorthand for “number of objects satisfying the following criteria.” A list of the xAOD jet containers used is as follows: `jet1=AntiKt4EMTopoJets`, `jet2=CamKt10LCTopoJets`, `jet3=CamKt12LCTopoJets`, `fatjet1=AntiKt10LCTopoTrimmedPtFrac5SmallR20Jets`, and `fatjet2=CamKt12LCTopoBDRSFilteredMU100Y15Jets`. “DFCommonMuons” and “DFCommonElectrons” (+ i.d. and quality requirements) are Derivation Framework leptons.

Selection	Value
HIGG2D4	
Trigger	L1_EM.* or L1_MU.*
Jets	count((jet $p_T > 15$ GeV) and ($ \text{jet } \eta < 2.6$)) > 1 or count((jet $p_T > 100$ GeV) and ($ \text{jet } \eta < 2.6$)) > 0 or count((fatjet1 $p_T > 100$ GeV) and ($ \text{fatjet1 } \eta < 2.6$)) > 0 or count((fatjet2 $p_T > 100$ GeV) and ($ \text{fatjet2 } \eta < 2.6$)) > 0
Leptons	(count((Muon $p_T > 6$ GeV) and (DFCommonMuonsLoose)) + count((Electron $p_T > 6$ GeV) and (DFCommonElectronsLHVeryLoose)) > 1) and $\Delta R(\text{electron, jet}) > 0.05$ and $m_{ll} > 5$ GeV

Table 6.3: List of selection criteria for the HIGG2D4 Dx AODs used in the analyses. The “count” function is simply shorthand for “number of objects satisfying the following criteria.” A list of the xAOD jet containers used is as follows: jet=AntiKt4EMTopoJets, fatjet1=AntiKt10LCTopoTrimmedPtFrac5SmallR20Jets, and fatjet2=CamKt12LCTopoBDRSFilteredMU100Y15Jets. “DFCommonMuons” and “DFCommonElectrons” (+ i.d. and quality requirements) are Derivation Framework leptons.

Electron Selection	p_T	η	ID	d_0^{sig}	$ \Delta z_0 \sin \theta $	Isolation
Loose	> 7 GeV	$ \eta < 2.47$	LH Loose	< 5	< 0.5 mm	LooseTrackOnly
Medium	> 25 GeV			Same as Loose		
Tight	Same as Medium		LH Tight	Same as Loose		FixedCutTight

Table 6.4: Electron object selection requirements.

Muon Selection	p_T	η	ID	d_0^{sig}	$ \Delta z_0 \sin \theta $	Isolation
Loose	> 7 GeV	$ \eta < 2.7$	Loose quality	< 3	< 0.5 mm	LooseTrackOnly
Medium	> 25 GeV	$ \eta < 2.5$		Same as Loose		
Tight	Same as Medium		Medium quality	Same as Loose		FixedCutTightTrackOnly

Table 6.5: Muon object selection requirements.

Jet Category		Selection Requirements	
Forward Jets		jet cleaning	
		$p_T > 30 \text{ GeV}$	
		$2.5 \leq \eta < 4.5$	
Signal Jets		jet cleaning	
		$p_T \geq 60 \text{ GeV}$	$20 \text{ GeV} \leq p_T < 60 \text{ GeV}$
		$ \eta < 2.5$	or $ \eta < 2.4$
		JVT > 0.59 for SM VH , 0.64 for $A \rightarrow Zh$	

Table 6.6: **AntiKt4EMTopoJets** selection requirements. Jet cleaning refers to the quality criteria interfaced through the **JetCleaningTool**. This tool removes jets in regions with hot, noisy calorimeter cells.

& 6.5. In these tables “ID” refers to identification purity. For electrons this is based on a likelihood method described in Section 4.2.3. For muons this requirement is based on four muon reconstruction algorithms (Subsection 4.2.4) using ID tracks and/or MS tracks and deposited energy in the calorimeter. $d_0^{\text{sig}} = |d_0|/\sigma_{d_0}$ is the impact parameter significance and $|\Delta z_0 \sin \theta|$ is the absolute value of the longitudinal impact parameter (as measured from the primary vertex) times the sine of the polar angle. Lastly, “Isolation” refers to isolation requirements detailed in Subsections 4.2.3 & 4.2.4. Small- R calo-jets are classified as either “Signal” or “Forward.” Signal jets are eligible for b -tagging and used in reconstructing the Higgs boson. As such, the requirements placed on signal jets are consistent with the flavor tagging CP group’s recommendation for b -jet tagging. The full set of selection requirements is given in Table 6.6. Both analyses use the 70% WP for b -tagging both small- R calo- and track-jets. This WP was chosen based on studies conducted on the BSM $A \rightarrow Zh$ analysis at various A -boson mass points in both the 0- and 2-lepton channels. In both lepton channels the 70% WP for MV2c20 was found to give the highest significance (as computed by a binned log-likelihood ratio [124]) [125]. This WP was kept after the move to MV2c10 as the rejection rate and efficiency are nearly identical to MV2c20. The SM analysis also adopts this WP as the low-mass A -boson samples are kinematically very similar to the SM signal.

6.2.1. Additional $A \rightarrow Zh$ Object Selection

In the 0-lepton channel contamination from $t\bar{t}$ and V +jets that include a τ -lepton that decays to hadrons (pions) is non-negligible. A hadronic- τ veto helps to reduce this source of background contamination. τ -lepton candidates have the following requirements:

- $p_T > 20 \text{ GeV}$
- $|\eta| < 2.5$ excluding LAr crack region ($1.37 < |\eta| < 1.52$)
- 1 or 3 matched tracks

The full set of selection requirements for the large- R calo-jets and track-jets are given in Table 6.7. These objects are used exclusively for events where the b -jets from the Higgs boson are too collimated to be resolved by the small- R calo-jets.

6.3. Event Selection

Although event selections are similar between the SM VH and (resolved) BSM $A \rightarrow Zh$ analyses, there are enough differences to warrant separate presentations. The $A \rightarrow Zh$ analysis makes extensive use of control regions and has both a “resolved” and “merged” regime where the b -jets originating from the Higgs boson are either able or unable to be resolved by the small- R calo-jets, respectively. However, the SM VH analysis is absent any control regions and doesn’t consider the scenario of a merged regime. A summary of various

Jet Category	Selection Requirements
Large- R Jets	$p_T > 250 \text{ GeV}$
	$ \eta < 2.0$
Track-Jets	$p_T > 10 \text{ GeV}$
	$ \eta < 2.5$
	has at least 2 track constituents

Table 6.7: `AntiKt10LCTopoTrimmedPtFrac5SmallR20Jets` (large- R calo-jets) and `AntiKt2PV0TrackJets` (track-jets) selection requirements.

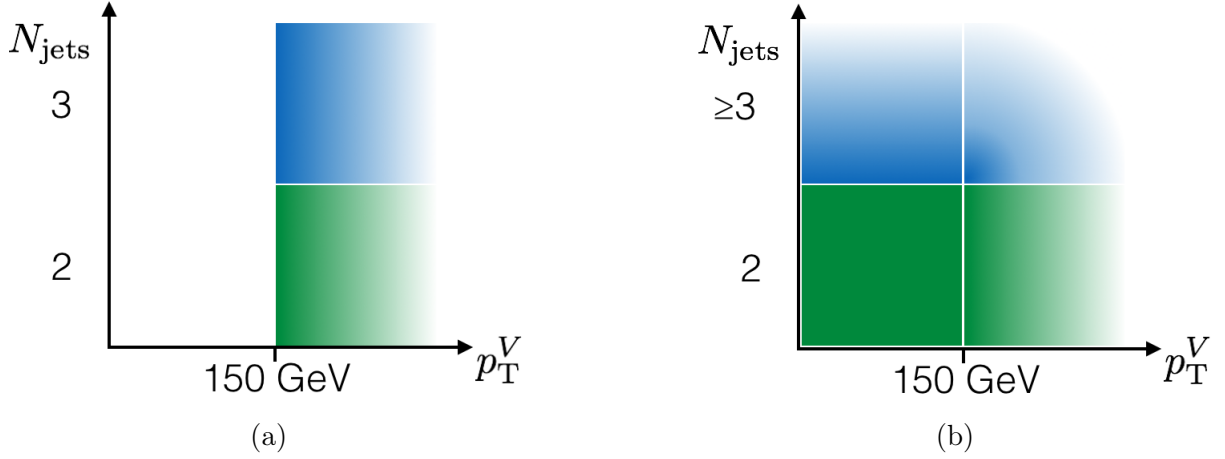


Figure 6.3: Signal regions for the SM VH analysis. The 0- & 1-lepton regions are shown in (a). Here, the only p_T^V region extends from 150 GeV through ∞ and the two N_{jets} regions consist of events with of two (green regions) or three (blue regions) jets. The 2-lepton regions are shown in (b). Here, there is an additional p_T^V region that extends from 0 GeV through 150 GeV and the three-jet region now extends to three-or-more.

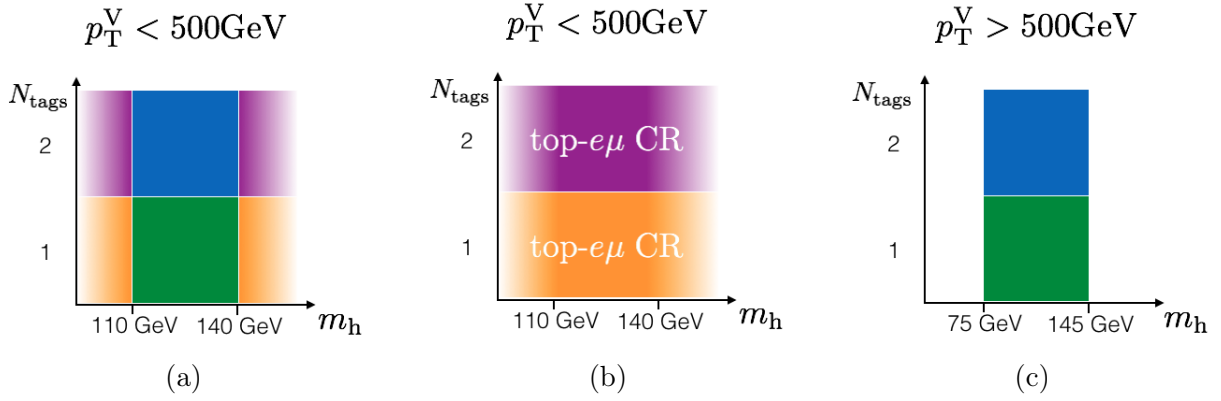


Figure 6.4: Signal and control regions for the BSM $A \rightarrow Zh$ analysis. The resolved signal and control regions are shown in (a). Here, there are one (green regions) and two (blue regions) b -tag signal and m_h sideband regions. The signal regions have either one or two N_{tags} and an m_h window about the nominal Higgs mass (125 GeV). The m_h sideband regions (orange and purple regions for 1- and 2- N_{tags} in (a)) act as control regions that effectively isolate $V + \text{jets}$ backgrounds. Note that in the 0-lepton channel there is an implicit $p_T^V > 150$ GeV requirement in the resolved regime due to a minimum E_T^{miss} cut. The resolved control regions that isolates the $t\bar{t}$ background (orange and purple regions for 1- and 2- N_{tags}) by an opposite lepton flavor ($e - \mu$) requirement are shown in (b). They have a looser m_{ll} requirement of $m_{ll} > 40$ GeV. The merged signal regions are shown in (c). Here, the m_h window is loosened and there are no control regions (colors are identical as in (a)).

Number of b -Tagged Jets	Jets Used in Reconstructing the Higgs
0	Two highest p_T signal jets
1	Highest p_T b -tagged jet, then highest p_T signal jet
2	Two highest p_T b -tagged jets

Table 6.8: Methodology for reconstructing Higgs boson from jets in the different b -tag regions. Note that in the 1-tag region the signal jets from which to select the highest p_T jets is absent the b -tagged jet already selected.

regions in these analyses is given in Figures 6.3 & 6.4. In these figures, m_h refers to the reconstructed Higgs boson mass from either small- R or large- R calo-jets. For events in the merged regime, this mass is simply that of the leading, large- R calo-jet while the resolved regime has a slightly more involved procedure for calculating the reconstructed Higgs mass. The Higgs mass in the resolved regime is always reconstructed from two signal jets, the rules for which are given in Table 6.8. This section gives a brief overview of the event selections for the two analyses in each lepton channel. In both analyses and all lepton channels the dominant background processes are $t\bar{t}$ and $V + \text{jets}$. Multi-jet background is negligible in all but the SM 1-lepton channel. The p_T cut on the hardest jet is a remnant of the Run 1 analysis and is present in all lepton channels and in both the SM and BSM analyses. This cut was found to aid in rejection of the multi-jet background and was correlated with a Run 1-analogous quantity to the JVT [126]. Sub-dominant backgrounds include single-top and di-boson (WW , WZ , and ZZ) events. The choice of physical observables and cuts is driven by the goal of excluding these backgrounds.

6.3.1. SM 0-Lepton Selection

Events are selected using E_T^{miss} triggers with thresholds of 70 and 90 GeV for 2015 and 2016 data, respectively. The 90 GeV trigger is used for higher pile-up conditions that necessitate an increase in the trigger threshold. Events are required to have no loose leptons present and a missing transverse energy of $E_T^{\text{miss}} > 150$ GeV. This E_T^{miss} cut achieves a high trigger efficiency of about 85(100)% for $E_T^{\text{miss}} = 150(200)$ GeV [101]. A selection based on the scalar

Selection	0-lepton
Trigger	$E_T^{\text{miss}} > 70 \text{ GeV}$ for data15 (90 GeV for data16)
Number of Leptons	0 “Loose”
Number of Jets	2 or 3
Number of b -jets	exactly 2
Leading Jet p_T [GeV]	> 45
E_T^{miss} [GeV]	> 150
S_T [GeV]	$> 120(150)$ for 2(3) jets
Multi-jet Suppression	
$\Delta\Phi(E_T^{\text{miss}}, E_{T,\text{trk}}^{\text{miss}})$	$< 90^\circ$
$\Delta\Phi(\text{jet}_1, \text{jet}_2)$	$< 140^\circ$
$\Delta\Phi(E_T^{\text{miss}}, \text{Higgs})$	$> 120^\circ$
$\min[\Delta\Phi(E_T^{\text{miss}}, \text{jets})]$	$> 20^\circ$

Table 6.9: SM 0-lepton event selection. S_T is the scalar sum of the transverse momentum of available jets (2 or 3). The number of jets is two signal jets and possibly an additional signal or forward jet. $\min[\Delta\Phi(E_T^{\text{miss}}, \text{jets})]$ is the minimum azimuthal angular difference between the E_T^{miss} and the signal jets.

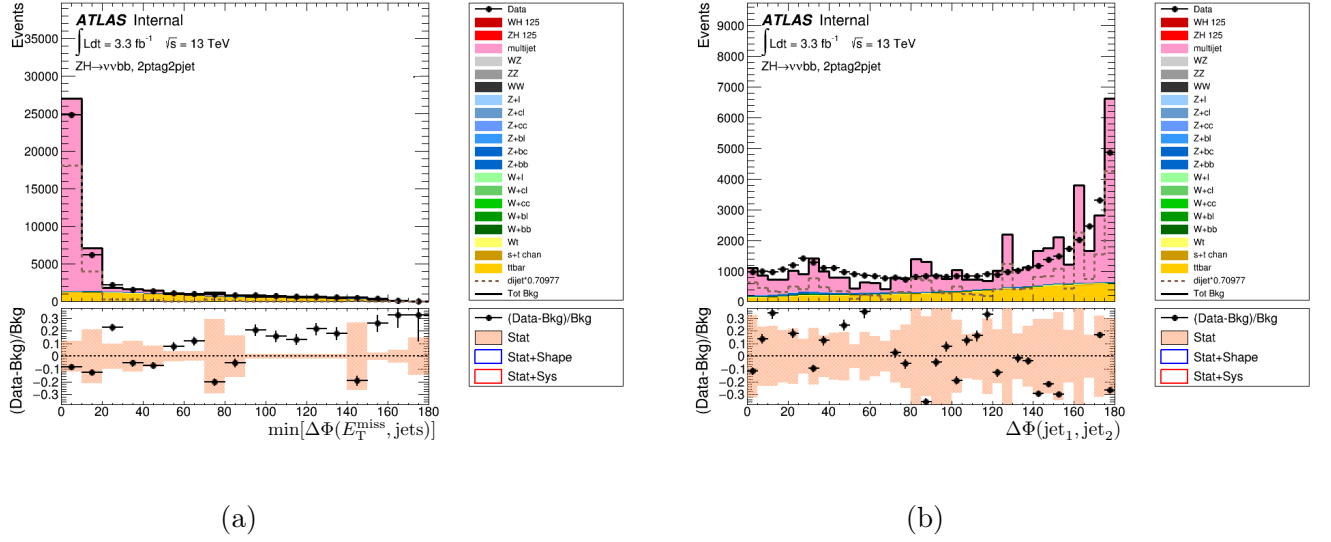


Figure 6.5: (a) $\min[\Delta\Phi(E_T^{\text{miss}}, \text{jets})]$ and (b) $\Delta\Phi(\text{jet}_1, \text{jet}_2)$ before any multi-jet suppression cuts in the 2-jet region of the 0-lepton SM analysis. Compare these distributions with the cut values in Table 6.9 [99].

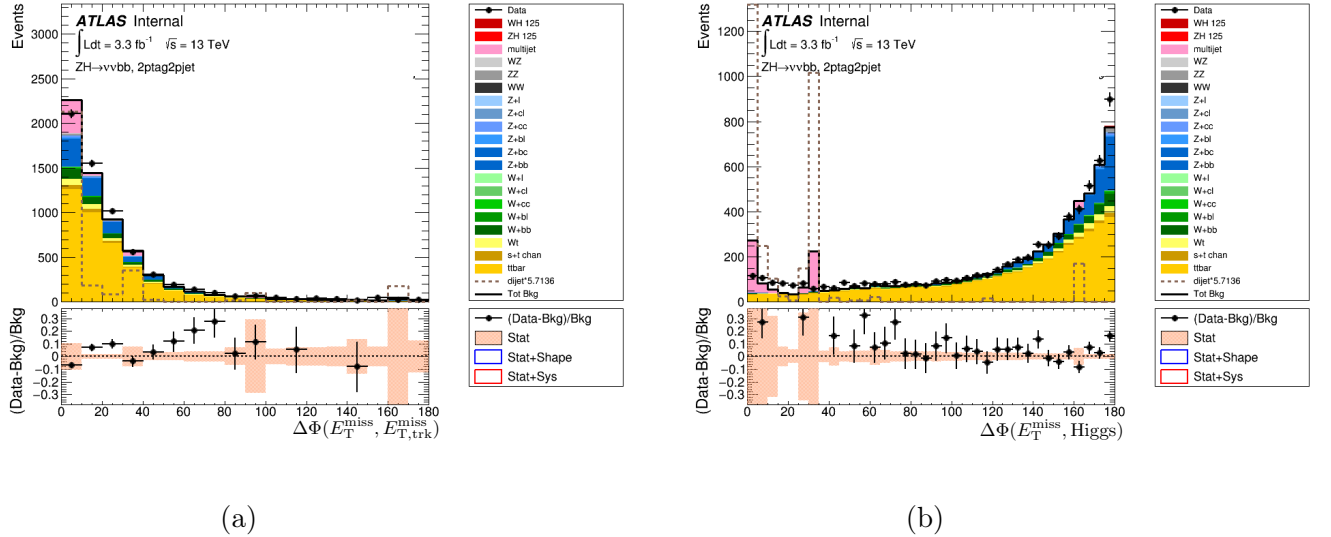


Figure 6.6: (a) $\Delta\Phi(E_T^{\text{miss}}, E_T^{\text{miss}}, \text{trk})$ and (b) $\Delta\Phi(E_T^{\text{miss}}, \text{Higgs})$ after multi-jet suppression cuts with $\min[\Delta\Phi(E_T^{\text{miss}}, \text{jets})] > 20^\circ$ and $\Delta\Phi(\text{jet}_1, \text{jet}_2) < 140^\circ$ in the 2-jet region of the 0-lepton SM analysis. Compare these distributions with the cut values in Table 6.9 [99].

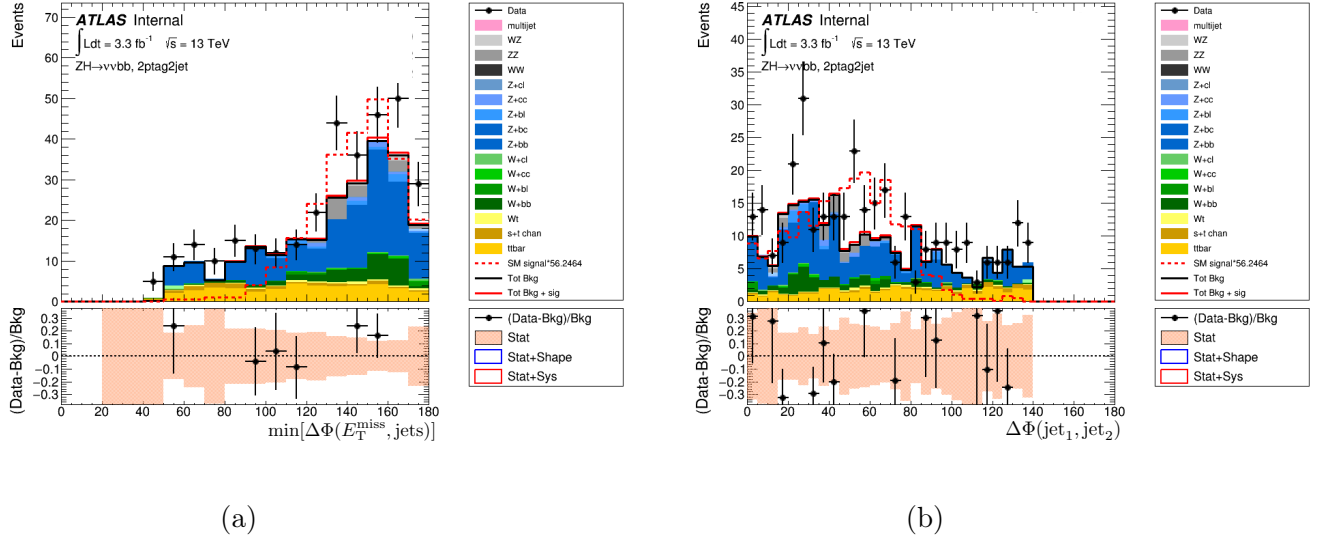


Figure 6.7: (a) $\min[\Delta\Phi(E_T^{\text{miss}}, \text{jets})]$ and (b) $\Delta\Phi(\text{jet}_1, \text{jet}_2)$ after all multi-jet suppression cuts in the 2-jet region of the 0-lepton SM analysis [99].

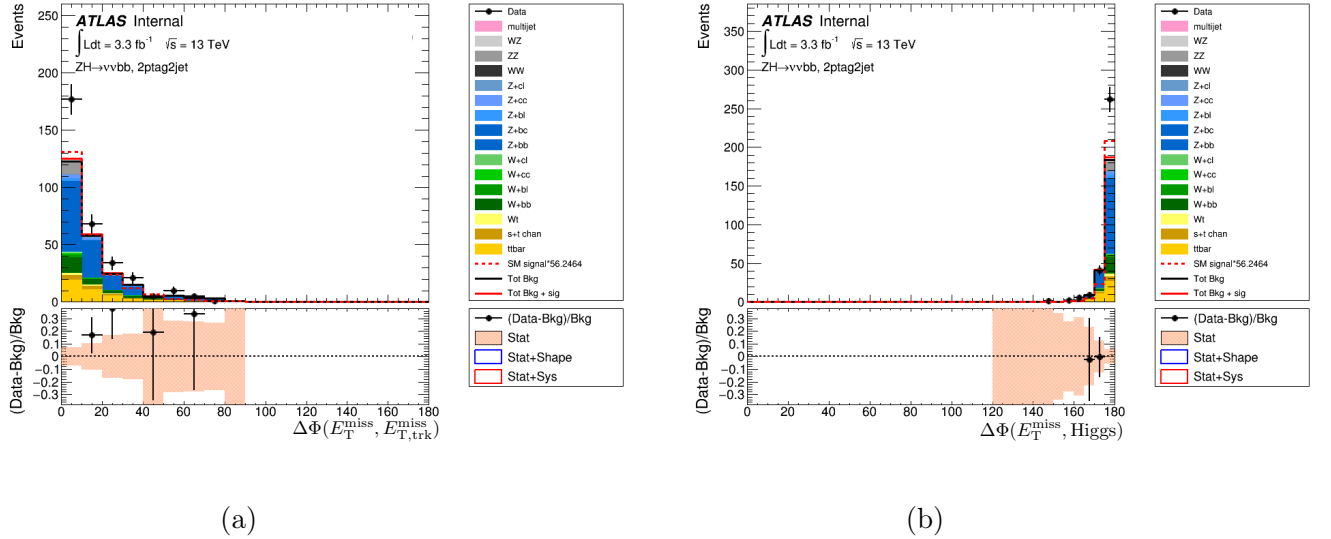


Figure 6.8: (a) $\Delta\Phi(E_T^{\text{miss}}, E_T^{\text{miss}}, \text{trk})$ and (b) $\Delta\Phi(E_T^{\text{miss}}, \text{Higgs})$ after all multi-jet suppression cuts in the 2-jet region of the 0-lepton SM analysis [99].

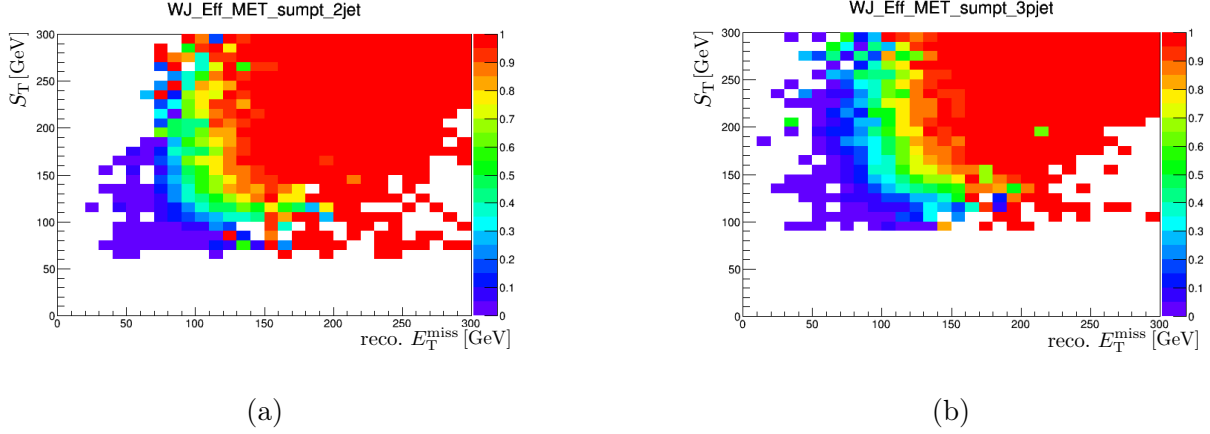


Figure 6.9: Missing energy trigger efficiencies as a function of offline missing energy and scalar sum of transverse momentum [99]. Both 2-jet (a) and 3-jet (b) regions shown use a special set of cuts aimed at isolating the $W(\mu, \nu) + \text{jets}$ simulated background events (selection is detailed in the aforementioned reference). Efficiency is defined here as the number of events that pass the event selection and trigger divided by those that pass event selection. Here “reco. E_T^{miss} ” is the offline value of E_T^{miss} .

sum of the transverse momentum of the two or three signal and forward jets, S_T , is used to remove poorly modeled regions of the phase-space arising from a non-trivial dependence of the trigger on the jet multiplicity [101]. Plots showing the dependence of the E_T^{miss} trigger on offline E_T^{miss} and S_T are given in Figure 6.9. Much of the multi-jet background originates as mis-identified E_T^{miss} [127] from a poor measurement of jet energies. Thus, comparing the direction of E_T^{miss} to $E_{T,\text{trk}}^{\text{miss}}$ aids in reducing this source of multi-jet contamination. If E_T^{miss} is correlated with a mis-measurement of a certain jet’s energy in the multi-jet background then the E_T^{miss} should be somewhat aligned with this jet. Thus, cutting on the minimum angle between E_T^{miss} and the set of jets in an event should help reduce this background. The Z and Higgs bosons should be relatively back-to-back in the transverse plane in the signal process, but no such relationship is expected in the multi-jet background. So, a comparison of the azimuthal angle between the E_T^{miss} (Z) and di-jet system (Higgs) should help further reduce the multi-jet background. In the signal process the jets that make up the Higgs boson are more collimated than in the multi-jet background, so a cut on the angle between these jets is also applied. In total, four angular selections between E_T^{miss} and $E_{T,\text{trk}}^{\text{miss}}$

($\Delta\Phi(E_T^{\text{miss}}, E_{T,\text{trk}}^{\text{miss}})$), the jets in the di-jet system ($\Delta\Phi(\text{jet}_1, \text{jet}_2)$), E_T^{miss} and the di-jet system ($\Delta\Phi(E_T^{\text{miss}}, \text{Higgs})$), and E_T^{miss} and a signal jet ($\min[\Delta\Phi(E_T^{\text{miss}}, \text{jets})]$) are applied. The distributions for $\min[\Delta\Phi(E_T^{\text{miss}}, \text{jets})]$ and $\Delta\Phi(\text{jet}_1, \text{jet}_2)$ before any of the multi-jet suppression cuts are applied are shown in Figure 6.5. The distributions for $\Delta\Phi(E_T^{\text{miss}}, E_{T,\text{trk}}^{\text{miss}})$ and $\Delta\Phi(E_T^{\text{miss}}, \text{Higgs})$ after $\min[\Delta\Phi(E_T^{\text{miss}}, \text{jets})] > 20^\circ$ and $\Delta\Phi(\text{jet}_1, \text{jet}_2) < 140^\circ$ cuts are given in Figure 6.6. The plots of the multi-jet suppression variables after applying their cut values as given by Table 6.9 are shown in Figures 6.7 and 6.8, after which no multi-jet background is expected to remain. It should be noted that a τ -lepton veto (present in the 0-lepton channel of the $A \rightarrow Zh$ analysis) causes significant signal loss and is not applied [99]. A summary of the SM 0-lepton event selection cuts is given in Table 6.9.

6.3.2. SM 1-Lepton Selection

This event category has two sub-channels: single-electron and single-muon. For the electron sub-channel, events are selected using single-electron triggers with thresholds of 24, 60 and 120 GeV for the 2015 data and with an increased threshold of 140 GeV for the highest threshold trigger in 2016. The lowest threshold trigger includes an isolation requirement. This is removed for the second-highest-threshold trigger and the identification requirements are relaxed for the highest-threshold trigger.

The muon sub-channel uses the same missing transverse energy triggers as the 0-lepton channel. As muons are not included in the E_T^{miss} calculation at trigger level, in signal events where a muon is present, this trigger is effectively a selection based on the p_T^V . This trigger has a higher efficiency than the single muon trigger, which has limited muon trigger chamber coverage in some regions of the detector [99]. Events are required to contain exactly one “Tight” lepton and no additional “Loose” leptons. In the electron sub-channel where it is easier for a jet to fake the lepton an additional selection of $E_T^{\text{miss}} > 30 \text{ GeV}$ is applied to reduce multi-jet backgrounds. Figure 6.10 shows a plot demonstrating the estimated amount of multi-jet removed by such an E_T^{miss} cut. However, this channel still contains a non-negligible amount of multi-jet contamination. Since it is not feasible to estimate this

Selection	1-lepton
Trigger	μ -sub-channel: same E_T^{miss} trigger as 0-lepton e -sub-channel: lowest unrescaled single e -trigger
Number of Leptons	exactly 1 “Tight” lepton
Number of Jets	2 or 3
Number of b -jets	exactly 2
Leading Jet p_T [GeV]	> 45
E_T^{miss} [GeV]	> 30 (e -sub-channel only)

Table 6.10: SM 1-lepton event selection. The number of jets is two signal jets and possibly an additional signal or forward jet.

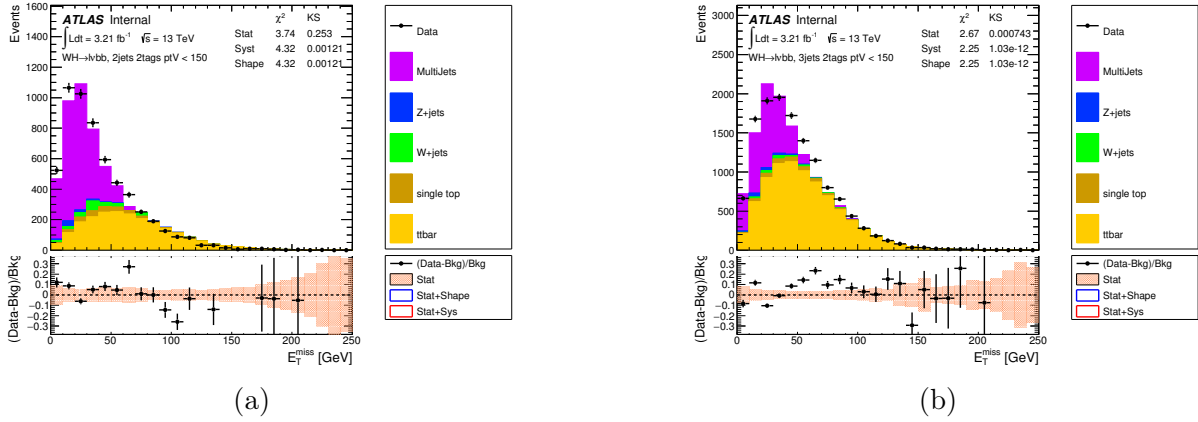


Figure 6.10: Missing energy distributions in the 1-lepton SM analysis before the E_T^{miss} multi-jet suppression cut for the (a) 2- and (b) 3-jet regions. An estimate for the multi-jet contribution by a template method [127] is given in purple. From this it’s clear that a 30 GeV cut removes a large amount of multi-jet contamination.

background contribution using only MC simulations, a data-driven estimate known as the fake-factor method is used that is described in Chapter 7. A summary of the SM 1-lepton event selection cuts is given in Table 6.10.

6.3.3. SM 2-Lepton Selection

This event category also has two sub-channels: electron and muon (denoting the leptons using to reconstruct the Z boson). Events are selected in the electron sub-channel using the same electron triggers as for the 1-lepton channel. For the muon sub-channel triggers with thresholds of 20 and 40 GeV are used for 2015 data and 24 and 50 GeV for 2016 data. As with the electron triggers the lowest threshold triggers include an isolation requirement that is removed for the higher threshold trigger. Exactly two “Loose” leptons of the same flavor are required to be reconstructed, one of which must also pass the “Medium” requirements. In the muon sub-channel an opposite sign requirement is also applied. This is not used in the electron sub-channel where the charge misidentification rate is higher [125]. The invariant mass of the di-lepton system, m_{ll} , must be consistent with the Z boson mass. This window is relatively large as this observable is an input to the final multivariate discriminant and is asymmetric to take into account the corresponding asymmetry in the backgrounds (determined in Run 1 and shown in Figure 6.11). A summary of the SM 2-lepton event selection cuts is given in Table 6.11.

6.3.4. $A \rightarrow Zh$ 0-Lepton Selection

Most of the event selection criteria in the resolved 0-lepton channel are identical to the SM VH 0-lepton selection with the addition of a τ -lepton veto and an $E_{T,\text{trk}}^{\text{miss}}$ cut. The transition between the resolved and merged regions is determined by a simple cut on p_T^V . The determination of the specific value of this variable is made through an optimization detailed in Ref. [125] where the p_T^V cut value is chosen through a binned log-likelihood ratio [124]. An m_h mass window is separately optimized for the resolved and merged regimes with efficiencies of 70%–80% and 90% in the resolved and merged regimes, respectively [125]. The

Selection	2-lepton
Trigger	μ -sub-channel: lowest unscaled single μ trigger e -sub-channel: lowest unscaled single e trigger
Number of Leptons	exactly 1 “Loose” lepton and 1 “Medium” lepton of the same flavor
Lepton Charge	opposite charge in μ -sub-channel
m_{ll} [GeV]	$71 < m_{ll} < 121$
Number of Jets	2 or ≥ 3
Number of b -jets	exactly 2
Leading Jet p_T [GeV]	> 45

Table 6.11: SM 2-lepton event selection. The number of jets is two signal jets and possibly one or more additional signal or forward jets.

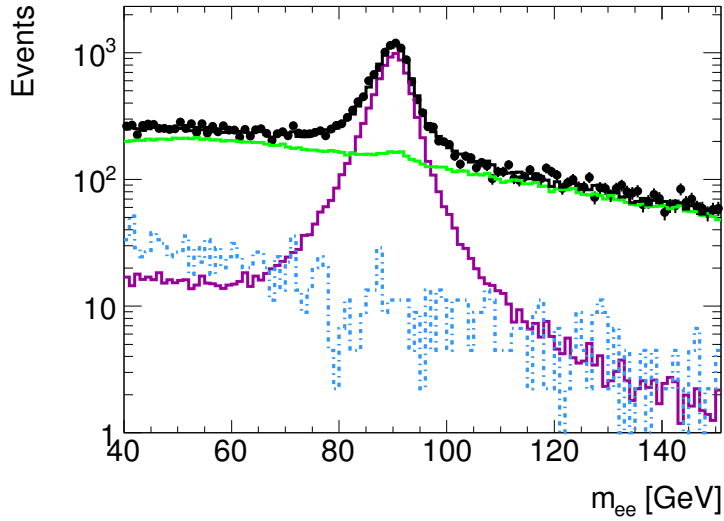


Figure 6.11: Run 1 SM 2-lepton multi-jet estimate [126]. Here, the black points with error bars are data, the purple histogram is the $Z + \text{jets}$, the green histogram is all non- $Z + \text{jets}$, and the dashed, blue histogram is the multi-jet background. Notice the asymmetry in the background composition. This leads to an asymmetric di-lepton mass window.

Selection	0-lepton	
	Resolved	Merged
Trigger	$E_{\text{T}}^{\text{miss}} > 70 \text{ GeV}$	
Number of Leptons	0 “Loose”	
Number of $R = 0.2$ Track-Jets	-	$\geq 2^*$
Number of $R = 0.4$ Calo-Jets	≥ 2	-
Number of $R = 1.0$ Calo-Jets	-	≥ 1
$E_{\text{T}}^{\text{miss}}$ [GeV]	> 150	-
Number of b -jets	1 or 2	
Leading Jet p_{T} [GeV]	> 45	-
p_{T}^Z [GeV]	< 500	≥ 500
m_h [GeV]	$110 \leq m_{bb} \leq 140$	$75 \leq m_{\text{jet}} \leq 145$
Number of hadronic τ -leptons	0	
S_{T} [GeV]	$> 120(150)$ for 2(3) jets	-
b -tagged track-jets not associated to the leading large- R jet	-	0
$E_{\text{T, trk}}^{\text{miss}}$ [GeV]	> 30	
Multi-jet Suppression		
$\Delta\Phi(\text{jet}_1, \text{jet}_2)$	$< 140^\circ$	-
$\Delta\Phi(E_{\text{T}}^{\text{miss}}, E_{\text{T, trk}}^{\text{miss}})$	$< 90^\circ$	$< 90^\circ$
$\Delta\Phi(E_{\text{T}}^{\text{miss}}, \text{Higgs})$	$> 120^\circ$	$> 120^\circ$
$\min[\Delta\Phi(E_{\text{T}}^{\text{miss}}, \text{jets})]$	$> 20^\circ$	$> 20^\circ$

Table 6.12: BSM $A \rightarrow Zh$ 0-lepton event selection. S_T is the scalar sum of the transverse momentum of available jets (2 or 3). $\min[\Delta\Phi(E_T^{\text{miss}}, \text{jets})]$ is the minimum azimuthal angular difference between the E_T^{miss} and the signal jets. *The two leading (in p_T) b -tagged track-jets must be ghost-associated to the leading large- R jet.

$E_{\text{T}}^{\text{miss}}$ cut is removed in the merged regime as this requirement is naturally satisfied when $p_{\text{T}}^Z \geq 500 \text{ GeV}$ ($E_{\text{T}}^{\text{miss}} = p_{\text{T}}^Z$ in this lepton channel). The multi-jet and non-collision backgrounds in the 0-lepton channel are suppressed by imposing an $E_{\text{T, trk}}^{\text{miss}}$ threshold (inherited from Run 1) and similar angular requirements as in the SM analysis in both the resolved and merged regimes. The $t\bar{t}$ and $W + \text{jets}$ backgrounds are further suppressed by vetoing events with at least one hadronically decaying τ -lepton candidate in both the resolved and merged regimes [125]. Specifically, events having any τ objects of medium quality [128, 129] are rejected. In addition, the $t\bar{t}$ background is suppressed by vetoing events with at least one b -tagged track jet not associated with the $h \rightarrow b\bar{b}$ candidate in the resolved regime. For the low- p_{T}^Z category, the four-momentum of the di-jet system is scaled by $m_h/m_{\text{di-jet}}$, where $m_h = 125 \text{ GeV}$. No rescaling is applied for the large- R jet in the high- p_{T}^Z category. A summary of the $A \rightarrow Zh$ 0-lepton event selection cuts is given in Table 6.12.

6.3.5. $A \rightarrow Zh$ 2-Lepton Selection

The 2-lepton channel is defined using events containing exactly two muons of opposite electric charge or two electrons, where at least one of the leptons has $p_{\text{T}} > 25 \text{ GeV}$ and $|\eta| < 2.5$. In the merged regime, both leptons must have $p_{\text{T}} > 25 \text{ GeV}$ and $|\eta| < 2.5$. The invariant mass of the two leptons must be in the range 70–110 GeV; this window is widened to 55–125 GeV for muons in the merged regime due to a decline in muon momentum resolution at high p_{T} . To suppress the $t\bar{t}$ background in the resolved regime, a requirement is made on $E_{\text{T}}^{\text{miss}}/H_{\text{T}}$, where H_{T} is the scalar sum of the p_{T} of the leptons and the small- R jets in the event. Since the invariant mass of the Zh system is used as the final discriminant, improvements in the resolution of the $A \rightarrow Zh$ reconstructed invariant mass are made through rescaling the four-momentum of the di-muon system by $m_Z/m_{\mu\mu}$ ($m_Z = 91.2 \text{ GeV}$) in both resolved and merged regimes and the di-jet system by $m_h/m_{\text{di-jet}}$ ($m_h = 125 \text{ GeV}$) in the resolved regime. The resulting resolution of the invariant mass m_{Zh} is in the range 2.5%–3.5% [125]. The events passing this selection define the signal regions of this channel.

A $t\bar{t}$ control region in the resolved regime is made to isolate this channel's $t\bar{t}$ contribution

Selection	2-lepton	
	Resolved	Merged
Trigger	μ -sub-channel: lowest unscaled single μ trigger e -sub-channel: lowest unscaled single e trigger	
Number of Leptons	exactly 1 “Loose” lepton and 1 “Signal” lepton of the same flavor	
Lepton Charge	opposite charge in μ -sub-channel	
m_{ee} [GeV]	$70 < m_{ll} < 110$	
$m_{\mu\mu}$ [GeV]	$70 < m_{ll} < 110$	$55 < m_{ll} < 125$
Number of $R = 0.2$ Track-Jets	-	$\geq 2^*$
Number of $R = 0.4$ Calo-Jets	≥ 2	-
Number of $R = 1.0$ Calo-Jets	-	≥ 1
E_T^{miss}/H_T [$\sqrt{\text{GeV}}$]	< 3.5	-
Number of b -jets	1 or 2	
Leading Jet p_T [GeV]	> 45	-
p_T^Z [GeV]	< 500	≥ 500
m_h [GeV]	$110 \leq m_{bb} \leq 140$	$75 \leq m_{\text{jet}} \leq 145$

Table 6.13: BSM $A \rightarrow Zh$ 2-lepton event selection. *The two leading (in p_T) b -tagged track-jets must be ghost-associated to the leading large- R jet.

by making a different flavor lepton requirement and changing the di-lepton mass window to a simple lower bound $m_{ll} \geq 40 \text{ GeV}$. A summary of the $A \rightarrow Zh$ 2-lepton event selection cuts is given in Table 6.13.

Chapter 7

Standard Model $VH(H \rightarrow b\bar{b})$

The SM VH analysis in Run 2 continues the ATLAS Run 1 search for the SM Higgs boson decaying to a pair of b -quarks. The Higgs boson discovered in 2012 (Section 2.4) has yet to be directly observed decaying to b -quark fermions despite this decay channel having the largest predicted branching ratio. As explained earlier, this is largely due to substantial background contamination from $t\bar{t}$ and V +jets production. The analysis presented in this chapter is a continuation of the Run 1 effort in ATLAS to observe this decay mode of the Higgs in the context of its vector-boson associated production. The leptonic decays of the associated vector boson allows for the possibility of triggering on either the presence of charged leptons or missing momentum while also suppressing multi-jet background contributions as detailed in Chapter 6. In addition to the analysis information given in previous chapters, five topics related to event reconstruction (Section 7.1, 7.2), background estimation (Section 7.3), modeling improvement (Section 7.4), and signal extraction using a BDT (Section 7.5) are covered in this chapter before presenting the analysis of the data through a global likelihood fit (Section 7.6) and results (Section 7.7). Although the primary focus of this analysis is the measurement of the $VH, H \rightarrow b\bar{b}$ process, a parallel analysis where the Higgs boson is replaced by the Z boson is performed as a cross check. This distinction is noted in the chapter wherever relevant. Unless otherwise noted, the information in this chapter has appeared in the ATLAS public note on the subject (Ref. [101]).

This analysis is the product of a team to which I contributed. My contributions to this analysis are technical implementation and upkeep of the flavor-tagging recommendations, creation of an analysis tool that allows users the ability to create database tables quickly and easily, preliminary flavor-tagging studies on which working point and systematics set to use, and discussions on the technical implementation of the truth tagging. I also served

as co-editor of an internal document on the selection of physics objects. In this chapter, I describe this analysis. Conclusions are presented in later chapters.

7.1. b -Jet Energy Corrections

The nominal jet energy calibrations (see Section 4.2.5.2) are derived on collections of jets that are primarily made of jets originating from light quarks or gluons. These calibrations are subsequently applied uniformly to all jets, irrespective of how they originated. Since jets of different flavors have different behavior in terms of the GSC input variables, the same GSC correction will affect a given jet differently depending on whether it originated from a b -, c -, or light-quark (or gluon). Since our signal contains two b -jets originating from a resonance and the di- b -jet invariant mass ($m_{b\bar{b}}$) is a good signal-to-background discriminant, further specific b -jet energy corrections are needed to account for b -jet and resonance-specific effects.

Two additional energy corrections are applied to b -tagged jets. From purely theoretical considerations one can expect the presence of a muon and an associated muon-neutrino in about 20% of the b -jet decays (from weak decays of the b - or c -quark) [99]. The “muon-in-jet” correction is used when one or more muons are matched to a jet within $\Delta R < 0.4$. When more than one muon is found, the closest to the jet axis is used. The muon 4-vector is added to that of the jet and the energy deposited in the calorimeter by the muon is removed. Systematics associated with the muon-in-jet correction were found to be negligible as they are closely associated with muon systematics which have very little impact in this analysis [99]. For the 0- and 1-lepton channels the second correction, denoted “PtReco,” is a scaling of the jet 4-vector as a function of the jet p_T after the muon-in-jet correction, derived in simulation by comparing the calibrated jet energy to the energy of matching truth b -quark partons (binned in p_T of the reconstructed jet). The correction is derived from simulated $ZH \rightarrow llb\bar{b}$ signal events and corrects for biases in the response arising from effects due to resolution, out-of-cone energy and unreconstructed neutrinos: taken from signal MC simulation, it automatically folds in the expected signal kinematics. This results

in a correction of around 12% or 24% at low p_T and plateauing at high p_T at around 1% or 6% for hadronic or semi-leptonic b -jets, respectively. In the 2-lepton channel, where the full $ZH \rightarrow llb\bar{b}$ event kinematics can be reconstructed, an improvement in the Higgs boson mass reconstruction is achieved through the use of a likelihood-based kinematic fit detailed in the next section, instead of the PtReco correction. It's important to note that these corrections to the b -jets' energies do not affect event selection as they are applied after all event-selection related cuts are made (see Tables 6.9, 6.10, and 6.11).

7.2. 2-Lepton Channel Kinematic Fit

The event topology of the 2-lepton channel can be exploited to better constrain the final state products through a Kinematic Fit (KF) procedure in order to improve the reconstructed Higgs boson mass resolution. The KF is a maximum likelihood estimate built from the expected signal topology without explicitly considering the Higgs mass. It is performed after the muon-in-jet correction in place of the PtReco correction. Recall that the final state in the 2-lepton channel should contain negligible E_T^{miss} and that the angular relationships amongst the leptons and b -jets should be well defined. At the level of MC simulation any E_T^{miss} should appear only as a consequence of the inherent transverse momentum of the colliding partons and from a possible semi-leptonic b -hadron decay. The inherent transverse momentum in the events determines the fundamental resolution of the total E_T^{miss} of the KF and must be determined through simulation. The 14 inputs to the KF are:

- energies of the two electrons or $1/p_T$ of the two muons
- p_T of the two b -jets
- η and ϕ of the leptons and jets
- p_T of the $llb\bar{b}$ system (or $llb\bar{b}j$ in case of three jets)
- reconstructed di-lepton mass m_{ll}

The constraints placed on the inputs are:

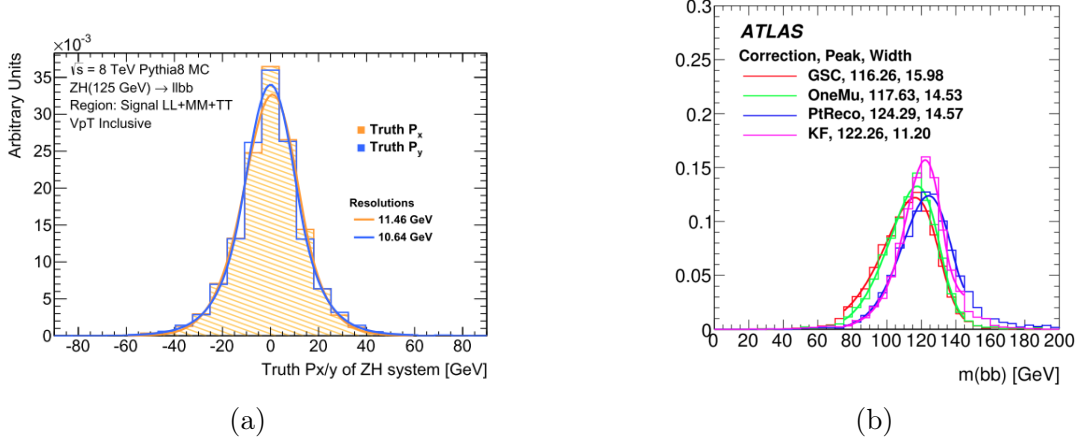


Figure 7.1: Different aspects of the KF used in the 2-lepton channel. The inherent resolution of the $llb\bar{b}$ system from a Run 1 study is shown in (a) [126]. The Run 2 improvement in resolution of the KF over just the muon-in-jet correction (denote OneMu), muon-in-jet+PtReco (denoted PtReco), and muon-in-jet+KF (denoted KF) is given in (b) [127]. This plot shows the 2-jet, low p_T^V region in a ZH simulated sample at $\sqrt{s} = 13$ TeV. Note that even though the PtReco gives a better match to the true value of the Higgs mass, the KF has a better resolution which translates to a higher significance when comparing to background processes.

- distribution of m_{ll} is assumed to behave as a Breit-Wigner centered about the pole mass of the Z boson with a width the same as the natural width of the Z boson
- the p_x and p_y of the $llb\bar{b}$ system are each centered at zero with a width of 9 GeV (found from MC simulation studies, see Figure 7.1a)
- all energy and azimuthal angle parameters follow a Gaussian distribution aside from the jets' p_T which have p_T - and η -dependent asymmetric transfer functions that are derived in ZH signal samples

The jet transfer functions are derived in bins of reconstructed jet p_T (after the muon-in-jet correction) and determined by fitting a Bukin function [130] to distributions of the ratio of jet p_T to matching-truth jet p_T . The KF is applied to both the 2- and 3-jet regions. In the 3-jet region the additional jet is accounted for in the total vector-sum of the system. The resolutions for the leptons and jets (modeled as Gaussian widths in the fit) are provided by the respective ATLAS CP groups. Ultimately, the KF improves the resolution of the $m_{b\bar{b}}$ mass by about 23%(13%) over that of just the muon-in-jet correction in the 2(3)-jet

region without “sculpting” the Z +jets and $t\bar{t}$ backgrounds. A demonstration of the improved resolution is given in Figure 7.1b.

7.3. 1-Lepton Multi-jet Background

As discussed in Chapter 6, the 0- and 2-lepton channels have negligible QCD multi-jet contamination after the event selection. For the 0-lepton channel the potential multi-jet contamination was studied in detail using samples of simulated and real data events prior to the application of the anti-multi-jet criteria in a variety of observables for events with 0, 1 and 2 b-tagged jets. The multi-jet contribution is modeled for these studies using PYTHIA 8 MC samples with the A14 tune and NNPDF2.3LO PDFs and used to estimate both the overall multi-jet contribution and the contribution in the region of $m_{b\bar{b}}$ close to the Higgs boson mass. Simulation and data generally agree well but a conservative approach of inflating the uncertainty on the background prediction where there are signs of possible mis-modeling is taken. The overall contribution is $<1\%$ of the total background and $<10\%$ of the estimated Higgs boson signal yield in the region $60 \leq m_{b\bar{b}} \leq 160$ GeV. This is a small enough contribution that this background is neglected in the global likelihood fit. In the 2-lepton channel, requiring two isolated leptons with a di-lepton invariant mass compatible with that of a Z boson strongly suppresses the contributions from multi-jet events. The residual contribution is estimated using a fit to the di-lepton distribution using a sample of events where the two lepton candidates have the same sign charge. The fit model includes expected contributions from electroweak (EW) backgrounds¹ from simulation and an exponential model for the multi-jet background. An estimate is then made of the fraction of the background in a mass window around the Z boson peak in the signal region that could be attributed to multi-jet events based on the assumption that the opposite sign and same sign events are symmetric for the multi-jet background. Inside a mass window $71 < m_{ll} < 121$ GeV the fraction of the background coming from multi-jet events is estimated to be 0.3% and 1.4% for the muon and electron sub-channels, respectively. This is small

¹EW backgrounds are any background processes other than multi-jet.

enough to have a negligible impact on the signal extraction and so is not included in the statistical analysis.

However, the 1-lepton channel must account for multi-jet contamination as both the electron and muon channels have contributions from multi-jet events faking the isolated lepton signature. The multi-jet background in the muon channel is predominately due to semi-leptonic decays of heavy-flavor hadrons to a muon. The size of the multi-jet background in the electron channel, which is mostly due to jets or photon conversions faking leptons or semi-leptonic heavy-flavor hadron decays, results in a larger background. Given the different sources of the multi-jet background events in the two channels the backgrounds are estimated separately, but using a similar procedure. Control regions are constructed, separately for each of the 2- and 3-jet categories, enriched in fake leptons that are kinematically close to but not overlapping with the corresponding signal regions. Correction factors are then applied to map each control region into the appropriate signal region to estimate the yield and shape of the multi-jet background. The control regions are selected using the nominal signal selection but with the requirement that the lepton should fail the tight identification requirements but pass a similar selection with looser identification and isolation requirements. For determining the background normalization, control regions with two b -tagged signal jets are used, whereas to estimate the shape of the background a larger sample with a relaxed requirement of one b -tagged jet is used. To correctly extrapolate the background measured in control regions to the signal region, the EW backgrounds with real leptons are subtracted based on MC simulation estimates and two correction factors are applied. The first correction factor is derived from measurements of the b -tagging efficiencies and takes into account the fact that only one b -tag is required for the shape control region, but two are required for the signal region. The second factor corrects for the differing lepton selection requirements. It is measured from an additional sample of data, highly enriched in di-jet events, selected by requiring exactly one signal jet and exactly one lepton meeting the loose requirements. The scale factor is derived from the ratio of the number of tight leptons to the number of more loosely ID'ed leptons after subtracting the contribution from the simulated EW

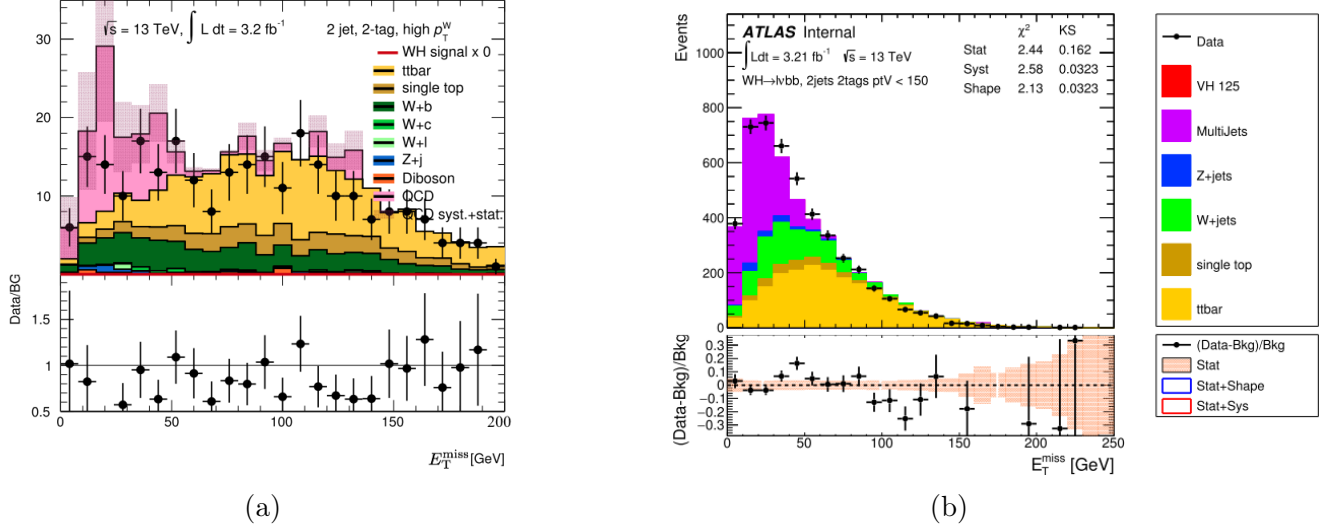


Figure 7.2: 1-lepton multi-jet estimates in the (a) electron and (b) muon sub-channels using the method described in the text [127]. These are both in the 2-jet region. The muon sub-channel is plotted in the low- p_T^V region which is not used in the analysis, but the general idea is the same.

backgrounds with real leptons and is measured in bins of p_T , $|\eta|$, and E_T^{miss} (the latter for the electron sub-channel only). The determination is quite sensitive to the modeling of the EW backgrounds and so a data-driven constraint on the normalization of these backgrounds is obtained by performing a fit to the data in the region $150 < E_T^{\text{miss}} < 250$ GeV. The final background template is constructed in the relevant final discriminating variable by taking the shape and normalization from the control regions and applying the correction factors discussed above. Plots of the multi-jet estimates in the E_T^{miss} distributions for the electron and muon sub-channels are given in Figure 7.2.

7.4. Truth-Tagging

Removing events which do not pass a given b -tagging requirement (known as direct tagging) with too few events surviving this criterion can create instabilities during the training and testing of the BDT and the global likelihood fit. One method to overcome these instabilities is not to cut away events due to b -tagging, but apply a weight to each MC simulation event based on the desired number of b -jets, tagging efficiencies, and scale factors. This

is broadly referred to as “truth tagging” because of the use of the truth flavor label (Section 4.2.7.1) of the jet in determining the event weight. Specifically, we select a pair of jets for truth tagging based on their probability to be tagged. This method is used for the $V+cl$, $V+l$, and WW samples in order to reduce the statistical uncertainties on the prediction from MC simulation. Direct tagging is used for all the other samples.

The three primary components to compute for truth tagging are the event weight (w_{TT}), which jets to “tag,” and what tag weight to assign to the truth-tagged jets. Given a set of n jets, J , the set of all m -tagged-jets combinations of jets is denoted $\binom{J}{m}$, the i^{th} combination $\binom{J}{m}_i$, and $\overline{\binom{J}{m}_i} = J - \binom{J}{m}_i$. The efficiency and inefficiency products are

$$\mathcal{E}(J, x) = \prod_{j \in J} \epsilon_x(j)$$

and

$$\mathcal{E}_{\text{in}}(J, x) = \prod_{j \in J} (1 - \epsilon_x(j))$$

where $\epsilon_x(j)$ is the tagging efficiency for jet j at or above efficiency working point x . Thus, the event weight for m truth-tagged jets can be expressed as

$$w_{\text{TT}}(x) = \sum_i^{|\binom{J}{m}|} P_i(x)$$

where

$$P_i(x) = \mathcal{E} \left(\binom{J}{m}_i, x \right) \mathcal{E}_{\text{in}} \left(\overline{\binom{J}{m}_i}, x \right)$$

The i^{th} combination is given a probability of being truth-tagged equal to $P_i(x)/w_{\text{TT}}(x)$. A combination is randomly selected to be tagged based on these probabilities. Once tagged, the jets are assigned artificial tagweight values through the following steps. Given the set of efficiency working points (WPs) X and a selected WP $\mathbf{x}$ the set of WPs equal to or less $\mathbf{x}$ in terms of efficiency is denoted $X_{\leq \mathbf{x}}$. For truth-tagged jet j the probability to associate this

jet with a specific WP x ($x \leq \mathbf{x}$) is

$$P_{\mathbf{x}}(j, x) = \frac{\epsilon_x(j)}{\sum_{i \in X_{\leq \mathbf{x}}} \epsilon_i(j)}$$

A specific WP is randomly selected for jet j from $X_{\leq \mathbf{x}}$ using the probabilities computed from by $P_{\mathbf{x}}(j, x)$. Finally, a tagweight tw_{TT} is computed for a randomly selected WP with tagweight tw_{sel} by

$$tw_{\text{TT}} = \frac{1}{2} (tw_{\text{sel}} + tw_{\text{sel}-1}),$$

where $tw_{\text{sel}-1}$ is the next-lowest tagweight from the selected one. For jets not selected to be tagged, the procedure for computing the tagweight is the same, except the next-highest WP is chosen rather than the next-lowest.

7.5. Boosted Decision Tree

A multivariate technique known as a BDT [131] is used to more optimally distinguish VH signal from background processes [127]. At the core of the BDT is the decision tree (DT). A DT is a hierarchical series of rectangular cuts² that benefits from easy interpretation. However, a traditional DT suffers from instability with respect to statistical fluctuations in the training sample (i.e. a fluctuation may unduly influence the decision to cut on a certain feature, thus propagating to all subsequent tree nodes). This is remedied by instead training an assortment (forest) of trees that are relatively shallow (only four nodes deep at most in this analysis) where the many trees “vote” to make a final classification decision. Another very effective technique for improving the accuracy of the DTs is to use a so-called “boosting” training procedure. Boosting is a generic technique applicable to many different learning algorithms. Among the most popular type of boosting and the kind used in this analysis is known as adaptive boost (AdaBoost [132]). AdaBoost in the context of a BDT takes events that were misclassified during the training of a decision tree and gives them

²Rectangular owing to the fact that in the hyperspace of features, the DT makes simple less-than or greater-than comparisons amongst that features that results in a hyper-rectangle of acceptance.

higher event weights in the training of the following tree. Taken together, a forest of shallow decision trees combined with AdaBoost make a robust classifier that is simple to train and use.

As with most machine learning algorithms, a BDT will only perform as well as the space in which the data are embedded. To this end, separate BDTs are trained for each region with specific feature sets (physical observables) designed for those regions. Multivariate discriminants making use of BDTs are constructed, trained and evaluated in each of the 0-, 1-, and 2-lepton channels separately for events with two and three jets. In the 0- and 1-lepton channels, only events with $p_T > 150$ GeV are used. For the 2-lepton channel two BDTs are used in each of the jet multiplicity categories, one for $p_T < 150$ GeV and one for $p_T > 150$ GeV. Two versions of the BDTs are trained: one to separate the $(VH, H \rightarrow b\bar{b})$ signal from the sum of the expected background processes, referred to as BDT_{VH} , and another to separate the $(VZ, Z \rightarrow b\bar{b})$ di-boson from the sum of the expected background processes, referred to as BDT_{VZ} . The input variables used to construct the BDTs are chosen in order to maximize the separation, while avoiding the use of variables not improving the performance significantly (performance here is measured with a binned likelihood metric [124]). Starting from the most powerful variable in terms of signal versus background discrimination, the di-jet mass ($m_{b\bar{b}}$), additional variables are tried one at a time and the one yielding the best separation gain is kept. This procedure is repeated until adding more variables does not result in a significant performance gain (or even a performance loss)³. The final sets of variables for the different channels are listed in Table 7.1. The b -tagged jets belonging to the di-jet system are labelled in decreasing p_T as b_1 and b_2 , and their separation in pseudorapidity is $|\Delta\eta(b_1, b_2)|$. In 3-jet events, the third jet is labelled as jet_3 and the mass of the 3-jet system is denoted $m_{b\bar{b}j}$. The angular separation, in the transverse plane, of the vector boson and the di-jet system of b -tagged jets and their pseudorapidity separation are denoted $\Delta\phi(V, b_1b_2)$ and $|\Delta\eta(V, b_1b_2)|$, respectively. In the 0-lepton channel, H_T is defined as the scalar sum of

³The specific threshold in performance is somewhat subjective. The general rule here is that if the performance doesn't improve by, say, 5% relative to the performance of the last variable, then stop. Also, adding additional variables can even degrade performance when holding the BDT parameters fixed (for obvious reasons).

Variable	0-lepton	1-lepton	2-lepton
p_T^V		X	X
E_T^{miss}	X	X	X
$p_T^{b_1}$	X	X	X
$p_T^{b_2}$	X	X	X
$m_{b\bar{b}}$	X	X	X
$\Delta R(b_1, b_2)$	X	X	X
$ \Delta\eta(b_1, b_2) $	X		X
$\Delta\phi(V, b_1 b_2)$	X	X	X
$ \Delta\eta(V, b_1 b_2) $			X
H_T	X		
$\min[\Delta\phi(l, b)]$		X	
m_T^W		X	
m_{ll}			X
m_{Top}		X	
$ \Delta Y(V, H) $		X	
	Only in 3-jet events		
$p_T^{\text{jet}_3}$	X	X	X
$m_{b\bar{b}j}$	X	X	X

Table 7.1: Variables used as inputs to the BDT for the 0-, 1-, and 2-lepton channels.

the transverse momenta of all jets and E_T^{miss} . In the 1-lepton channel, the angle between the lepton and the closest b -tagged jet in the transverse plane is denoted $\min[\Delta\phi(l, b)]$. The W boson transverse mass is defined as $m_T^W = \sqrt{2p_T^l E_T^{\text{miss}}(1 - \cos(\Delta\phi(l, E_T^{\text{miss}}))}$ where p_T^l is the lepton transverse momentum.

In the 1-lepton channel two variables are used to improve the rejection of the $t\bar{t}$ background: the rapidity difference between the W and Higgs bosons, $|\Delta Y(V, H)|$ and, under the hypothesis that the event is $t\bar{t}$ the reconstructed top-quark mass, m_{Top} . To construct each of these variables an estimate of the 4-vector of the neutrino in the W boson decay is required. The vector E_T^{miss} is assumed to give the transverse components and then p_z' can be determined up to a possible two-fold ambiguity by constraining the mass of the lepton + neutrino to be consistent with the known W boson mass. The top quark is then reconstructed by considering the reconstructed W boson and one of the two b -tagged jets. The choice of the two possible p_z' and b -tagged jet is made such that the value of m_{Top} is minimized. The inclusion of $|\Delta Y(V, H)|$ gives up to a 10% improvement in the ROC curve⁴ for the 1-lepton BDTs [134]. Two-dimensional histograms illustrating the discriminating power of the m_{Top} variable are given in Figure 7.3.

A parallel analysis, intended as a cross-check, that performs a likelihood fit to the $m_{b\bar{b}}$ (rather than the BDT output) distribution (as was done for the 2011 Run 1 ATLAS dataset) performed 30–40% lower in expected significance when compared to the BDT-based analysis [127].

7.6. Fit Model & Statistical Analysis

Arguably the most crucial step in an analysis is to compare theoretical predictions to observed data. This can take the form of analyzing non-biasing regions of data⁵ in an attempt

⁴A receiver operating characteristic (ROC) curve is a common performance metric for a multivariate algorithm in which the true-positive rate (signal efficiency) is plotted against the false-positive rate (purity). An algorithm should strive for the area under the ROC curve to be unity (perfect separation in the two-class case).

⁵What is meant by non-biasing in this context is that while you are developing analysis strategies, you try as best you can to exclude regions of data that are predicted to contain large portions of signal. In this way, you attempt to limit the possibility of biasing your analysis by forcing the data to conform to the signal

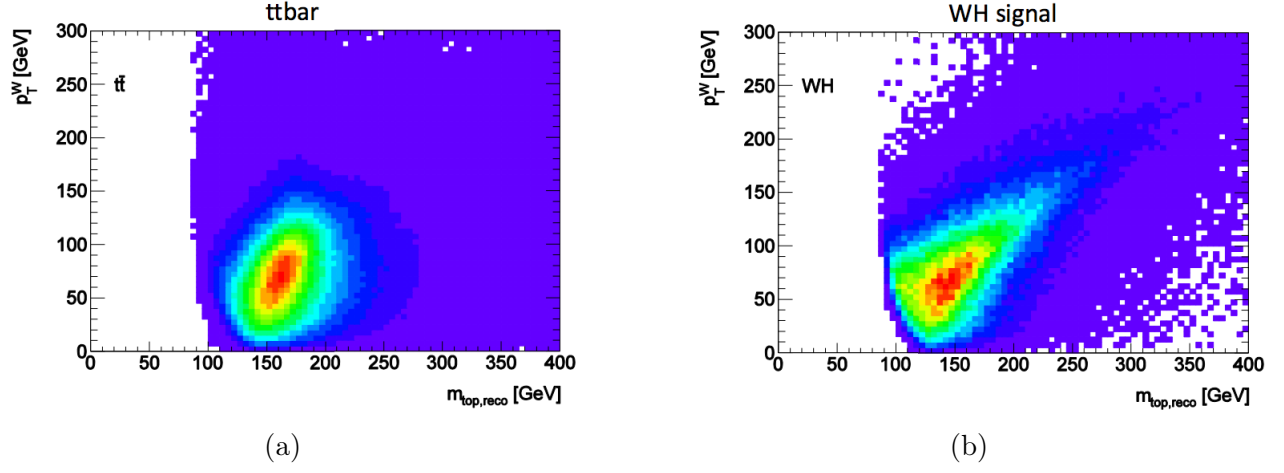


Figure 7.3: m_{Top} versus p_T^W in the 1-lepton, 2+3 jet regions for (a) $t\bar{t}$ and (b) WH [133]. Notice there exists a noticeable shape difference between the background $t\bar{t}$ and signal WH that the BDTs will likely take advantage of.

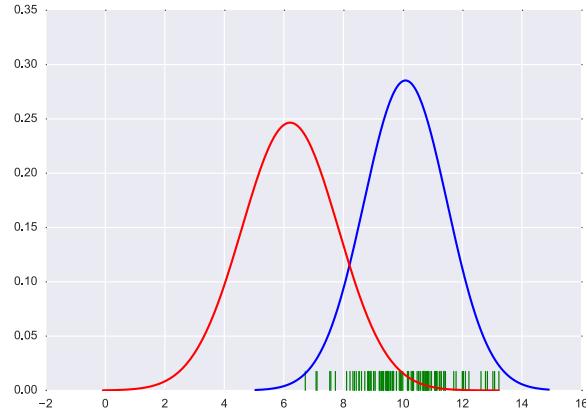


Figure 7.4: Toy example of an MLE fit to normally distributed data (green rug plot). The red Gaussian of the left has parameters less consistent with the data than the blue Gaussian on the right. Thus, the blue Gaussian will have a more favorable MLE score and will be selected as the more likely model.

to fit to data one or more expertly chosen (or theoretically motivated) analytic functions to model the predicted background and signal processes. The process of fitting a functional form to the data is typically accomplished through maximum likelihood estimation (MLE). MLE relies on the properties of pdfs to find a “best fit” configuration of pdf parameters that most likely produced the given data. A toy demonstration of such a fit is given in Figure 7.4. Other methods of fitting to data include so-called non-parametric techniques that eschew the simplicity of an analytic form in favor of additional complexity with the hopes of better adapting to data. However, one must be extremely careful with non-parametric estimates as, despite their name, they have hyper-parameters⁶ that must be tuned so as to not over- or under-fit to the data (the exact interpretation depends on the context). Two examples of these methods in the context of regression and density estimation are given in Appendices B and C. As explained in Appendix C, a histogram is an example of a non-parametric estimate for the pdf (and not a particularly good one for reasons explained in the appendix) and is used ubiquitously in high energy physics. Although it’s desirable to devise a parametric form for the distributions one is fitting, in many real-world scenarios, including this analysis, there exists no known analytic shape that reasonably describes the predicted or observed data. Since this is a frequent occurrence in high energy physics, a framework has been created that parameterizes histograms for use in an MLE fit⁷ that is detailed in Section 7.6.2. Before moving to a discussion of the formal data analysis, an overview of the systematic uncertainties that impact this analysis is presented. This section then outlines the statistical analysis performed in comparing the simulated, theoretic model (MC simulation) versus detector data.

7.6.1. Systematic Uncertainties

There are several sources of systematic uncertainty that will impact the measurement.

hypothesis or by trying to design cuts that reduce or enhance an unexpected data yield.

⁶As stated in the appendices, hyper-parameters are quite different than usual parameters in that there usually exists a set of hyper-parameters that fit the data perfectly. This is almost always undesirable and represents a case of over-fitting.

⁷This is a somewhat perverse use of a histogram by parametrically fitting a non-parametric estimator.

These can broadly be grouped into four categories: those of an experimental nature, those relating to the multi-jet background estimation, those relating to the modeling of the simulated background and Higgs boson signal samples.

7.6.1.1. *Experimental Uncertainties*

The dominant sources of experimental uncertainty arise from the flavor tagging simulation-to-data efficiency correction factors and the jet energy scale and resolution corrections. The flavor tagging simulation-to-data efficiency correction factors are measured separately for b -, c -, and light-flavor jets as detailed in Chapter 4 and Appendix B (this is an area where I made significant contributions). The correction factors for c - and light-flavor jets are based on measurements carried out with the Run 1 data and thus have Run 1-to-Run 2 extrapolation uncertainties. A dedicated Run 2 correction factor has been derived for b -jets using $t\bar{t}$ events. All three correction factors have many sources of uncertainty and are decomposed into uncorrelated components, which are then treated independently resulting in four uncertainties for both b - and c -jets and six for light-flavor jets. Uncertainties on the jet energy scale (JES) and resolution (JER) are estimated from 13 TeV data as detailed in Chapter 4. The many sources of uncertainties on the JES correction are decomposed into nineteen uncorrelated components which are treated as independent sources. Uncertainties on the reconstruction, identification, isolation and trigger efficiencies of muons and electrons, along with the uncertainty on their energy scale and resolution, are also evaluated based on 13 TeV data and are found to have only a small impact on the result. The uncertainty on the energy scale and resolution of the jets and leptons are propagated to the calculation of the E_T^{miss} , which also has additional dedicated uncertainties on the scale, resolution and efficiency of the tracks associated with the TST, along with the modeling of the underlying event. An uncertainty is applied on the simulation-to-data E_T^{miss} trigger scale factor, relating to the statistical uncertainty on the measured scale factor and differences between the scale factor determined on W +Jets, Z +Jets and $t\bar{t}$ events. The uncertainty on the luminosity is 2.1% for the 2015 and 3.7% for the 2016 data, giving an uncertainty of 2.9% on the combined

dataset. It is derived, following a methodology similar to that detailed in Ref. [135], from a preliminary calibration of the luminosity scale using x - y beam-separation scans performed in August 2015 and May 2016.

7.6.1.2. *Simulated Background Modeling Uncertainties*

Modeling uncertainties are derived on the simulated backgrounds and broadly cover three areas: normalization, acceptance⁸ differences that affect the relative normalization between analysis regions with a common background normalization, and the shapes of the most important kinematic variables. These uncertainties are derived either from truth-particle comparisons based on the nominal and alternative samples (Section 5.1) using the RIVET framework [136] or from comparisons to data in control regions. When normalization and acceptance uncertainties are evaluated all the nominal and alternative samples are normalized to the same production cross-section. Such uncertainties are generally evaluated by adding the difference between the nominal and appropriate alternative samples in quadrature. Shape uncertainties are considered in each of the analysis regions separately, with the samples scaled to have the same normalization in each region. The uncertainty is taken from the alternative generator which has the largest shape difference compared to the nominal sample. Shape uncertainties are only derived on the $m_{b\bar{b}}$ and p_T^V variables, as it was found that it is sufficient to only consider systematic variations of these variables to cover the full variation when evaluating the BDT_{VH} discriminant for an alternative generator. A summary of the systematic uncertainties affecting the modeling of the background samples is given in Table 7.2.

7.6.1.3. *Multi-Jet Background Uncertainties*

Three sources of uncertainty are considered on the estimate of the multi-jet background in the electron channel: the relative amount of fake leptons from jets versus those from semi-leptonic heavy flavor hadron decays, the amount of EW background in the reversed

⁸Acceptance in this context refers the fraction of events that survive selection criteria.

<i>Z</i> +jets	
<i>Zl</i> normalization	18%
<i>Zcl</i> normalization	23%
<i>Zbb</i> normalization	Floating
<i>Zbc</i> -to- <i>Zbb</i> ratio	14–27%
<i>Zcc</i> -to- <i>Zbb</i> ratio	7–31%
<i>Zbl</i> -to- <i>Zbb</i> ratio	15–38%
0-to-2 lepton ratio	26%
2-to-3 jet ratio	28% (0-lepton) and 25% (2-lepton)
$p_T^V, m_{b\bar{b}}$	S
<i>W</i> +jets	
<i>Wl</i> normalization	32%
<i>Wcl</i> normalization	37%
<i>Wbb</i> normalization	Floating
<i>Wbl</i> -to- <i>Wbb</i> ratio	17% (0-lepton) and 31% (1-lepton)
<i>Wbc</i> -to- <i>Wbb</i> ratio	42% (0-lepton) and 21% (1-lepton)
<i>Wcc</i> -to- <i>Wbb</i> ratio	17% (0-lepton) and 31% (1-lepton)
2-to-3 jet ratio	23%
0-to-1 lepton ratio	17%
$p_T^V, m_{b\bar{b}}$	S
<i>t$\bar{t}$</i> (all are de-correlated between the 0+1 and 2-lepton channels)	
<i>t$\bar{t}$</i> normalization	Floating
2-to-3-jet ratio	9% (0+1-lepton) and 24% (2-lepton)
$p_T^V, m_{b\bar{b}}$	S
Single top	
Cross section	4.4% (<i>s</i> -channel), 4.6% (<i>t</i> -channel), 6% (<i>Wt</i>)
Acceptance 2-jet	16% (<i>t</i> -channel), 25% (<i>Wt</i>)
Acceptance 3-jet	19% (<i>t</i> -channel), 32% (<i>Wt</i>)
$m_{b\bar{b}}, p_T^V$	S (p_T^V uncorrelated between 2 and 3-jet channels <i>Wt</i>)
<i>ZZ</i>	
Single top	
Cross section	4.4% (<i>s</i> -channel), 4.6% (<i>t</i> -channel), 6% (<i>Wt</i>)
Acceptance 2-jet	16% (<i>t</i> -channel), 25% (<i>Wt</i>)
Acceptance 3-jet	19% (<i>t</i> -channel), 32% (<i>Wt</i>)
$m_{b\bar{b}}, p_T^V$	S (p_T^V uncorrelated between 2 and 3-jet channels <i>Wt</i>)
<i>ZZ</i>	
Normalization	20%
0-to-2 lepton ratio	30%
2-to-3 jet ratio	19 %
$m_{b\bar{b}}, p_T^V$	S (correlated with <i>WZ</i> uncertainties)
<i>WZ</i>	
Normalization	26%
2-to-3 jet ratio	14% (0-lepton) and 11% (1-lepton)
0-to-1 lepton ratio	12%
$m_{b\bar{b}}, p_T^V$	S (correlated with <i>ZZ</i> uncertainties)
<i>WW</i>	
Normalization	25%
Multi-jet (1-lepton)	
Normalization	14–81% (electron), 5–50%(muon)
Template variations	S

Table 7.2: Summary of the systematic uncertainties on the background modeling. An “S” symbol is used when only a shape uncertainty is assessed.

isolation region and the impact of the energy scale applied to the “fake-electrons” on the E_T^{miss} calculation. In the muon channel the sources of uncertainty are due to the size of the EW background in the reversed isolation region and the bias caused by the different E_T^{miss} that isolated and non-isolated muon events may have.

7.6.1.4. Simulated Signal Modeling Uncertainties

The signal samples are normalized to the inclusive cross sections as described in Chapter 5. Uncertainties on the calculations of the normalization are assigned following the recommendations of the LHC Higgs Cross Section working group⁹ [137]. An uncertainty is estimated by varying the factorization and renormalization scale factors independently between a third and three times their nominal values. For WH production this gives an uncertainty of 0.7%. For ZH production it is somewhat larger at around 3% due to the contributions from the gluon fusion process. To assess uncertainties independently for the quark-quark and gluon-gluon initiated processes it is assumed that the uncertainty on $qq \rightarrow ZH$ is identical to WH at 0.7%. This assumption then leads to an estimate of the uncertainty on the $gg \rightarrow ZH$ cross section of 27%. The PDF and α_S contributions to the uncertainty are 1.9% for WH , 1.6% for $qq \rightarrow ZH$, and 5% for $gg \rightarrow ZH$. A further uncertainty is assessed to account for missing higher order electroweak effects as the maximum of either: 1%, the relative NLO electroweak corrections squared, or the uncertainty on the photon induced cross section relative to the total VH cross section. The uncertainty on the $H \rightarrow b\bar{b}$ branching ratio is 1.7%, which takes into account missing higher order QCD and electroweak effects as well as uncertainties on the Higgs boson decay width.

Acceptance uncertainties are estimated using event samples generated using MADGRAPH5_AMC@NLO v2.2.2 at LO interfaced to the PYTHIA 8.186 parton shower model. The A14 tune is used together with the PDF4LHC2015 NLO PDF set. The uncertainty is estimated by varying the renormalization and factorization scales between a half and twice their nominal values independently and taking the envelope of the variations. The

⁹This is a group of expert theorists and phenomenologists that compile Higgs related cross section information for use across all LHC experiments.

Signal	
Cross section (scale)	0.7% ($q\bar{q}$), 27% (gg)
Cross section (PDF)	1.9% ($q\bar{q} \rightarrow WH$), 1.6% ($q\bar{q} \rightarrow ZH$), 5% (gg)
Branching ratio	1.7 %
Acceptance (scale)	1.4%–5%
3-jet acceptance (scale)	1.4%–4.7%
p_T^V shape (scale)	S
Acceptance (PDF)	0.3%–0.7%
p_T^V shape (NLO EW correction)	S
Acceptance (parton shower)	4%–7.5%

Table 7.3: Summary of the systematic uncertainties on the signal modeling. An “S” symbol is used when only a shape uncertainty is assessed.

uncertainties are decomposed into three components that correctly take into account the use of exclusive jet bins. The uncertainties for $qq \rightarrow VH$ in the 2-jet category are 4%, 5%, and 1.4% for the 0-, 1-, and 2-lepton channels respectively and 2.7%, 4.7%, and 1.4% for the 3-jet category. The uncertainties for $gg \rightarrow ZH$ are assumed to be the same as for $qq \rightarrow ZH$.

The PDF and α_S contributions to the acceptance uncertainty are estimated for $qq \rightarrow VH$ to be: 0.6%, 0.7%, and 0.3% for the 0-, 1-, and 2-lepton selections respectively. An uncertainty of 3% is used for the $gg \rightarrow ZH$ contribution. Contributions from the modeling of the underlying event, parton showering and hadronization (denoted UE/PS/HAD) are estimated using the eigen-tunes of the A14 tune. This procedure leads to uncertainties correlated between the 2- and 3-jet categories of 7.5%, 7.5%, and 6.5% for the 0-, 1-, and 2-lepton selections, with additional uncertainties for the 3-jet category of 5%, 6% and 4% respectively.

Where significant variations are observed in the modeling of the p_T^V and $m_{b\bar{b}}$ distributions under the QCD scale variations, PDF, and α_S variations; variations accounting for missing higher order electroweak effects; or variations associated with UE/PS/HAD modeling, shape

uncertainties are assessed by taking the envelope of the observed variations and propagating it through to the relevant final discriminant distributions. Variations from the same source affecting the same distribution are correlated in the global likelihood fit. For the scale, electroweak, and UE/PS/HAD variations a shape systematic is assessed for both p_T^V and $m_{b\bar{b}}$ for all analysis regions. For PDF and α_S contributions to the shape variation no significant impact is seen for $m_{b\bar{b}}$ and so shape variations are assessed only for their impact on p_T^V .

A summary of the systematic uncertainties affecting the modeling of the signal samples is given in Table 7.3.

7.6.2. Statistical Analysis

A statistical fitting procedure based on the RooStats framework [138, 139] is used to extract the signal strength (μ , discussed in Chapter 2) from the data. A binned likelihood function is constructed as the product of Poisson-probability terms over the bins of the input distributions involving the numbers of data events and the expected signal and background yields, taking into account the effects of the floating background normalizations and the systematic uncertainties. In the context of fitting to data, μ is known as the parameter of interest (PoI) while all other parameters that must be fit yet are not the explicit focus of the analysis are known as nuisance parameters (NPs). The specific form of the likelihood function is given by the HistFactory template [140]:

$$\mathcal{P}(n_{cb}, a_p | \phi_p, \alpha_p, \gamma_b) = \prod_{c \in \text{channels}} \prod_{b \in \text{bins}} \text{Pois}(n_{cb} | \nu_{cb}) \cdot G(L_0 | \lambda, \Delta_L) \cdot \prod_{p \in \mathbf{S} + \mathbf{\Gamma}} f_p(a_p | \alpha_p).$$

Here, a channel represents a non-overlapping region of an analysis space (i.e. the 2-lepton, high p_T^V , 2-jet region is a single channel in this context). n_{cb} is the observed bin content for bin b and channel c . The expected (mean) bin content, ν_{cb} , is a function of a bin-by-bin scale factor γ_b , a luminosity parameter λ , a product of unconstrained normalization factors (including the PoI μ) ϕ_p , and the parameterized normalization and histogram (i.e. shape) uncertainties. The luminosity has a Gaussian constraint $G(L_0 | \lambda, \Delta_L)$ and the $f_p(a_p | \alpha_p)$

represent constraints on systematics $\mathbb{S}$ and bin-by-bin parameters¹⁰ $\mathbf{\Gamma}$. The parameterized normalization and histogram uncertainties are constructed from template histograms representing the nominal and $\pm 1\sigma$ variations of the respective systematic. From these templates interpolation and extrapolation schemes are chosen wherein partial variations and variations beyond 1σ can be approximated. Much more detail can be found in the aforementioned reference on HistFactory.

In the rest of the thesis all of the various parameters in the fit that are not the PoI (i.e. the NPs) will be referred to collectively by θ and the dependence of the likelihood on the PoI and NPs will be abbreviated as $\mathcal{P}(n_{cb}, a_p | \phi_p, \alpha_p, \gamma_b) \rightarrow \mathcal{L}(\mu | \theta)$. The impact of systematic uncertainties on the signal and background expectations is described by NPs which are constrained by Gaussian or log-normal probability density functions, the latter being used for normalization uncertainties to prevent normalization factors from becoming negative in the fit. The parameterization of each NP is chosen such that the predicted signal and background yields in each bin are log-normally distributed for a normally distributed θ . For each NP, the prior is added as a penalty term to the likelihood, $\mathcal{L}(\mu | \theta)$, which decreases its numerical value as soon as θ is shifted away from its nominal value. The statistical uncertainties associated with the background predictions from simulation are included through bin-by-bin nuisance parameters ($\mathbf{\Gamma}$ parameters).

The test statistic¹¹ q_μ is then constructed from the profile likelihood ratio

$$q_\mu = -2 \ln \Lambda(\mu, \hat{\hat{\theta}}, \hat{\mu}, \hat{\theta}) \quad \text{with} \quad \Lambda(\mu, \hat{\hat{\theta}}, \hat{\mu}, \hat{\theta}) = \frac{\mathcal{L}(\mu | \hat{\hat{\theta}})}{\mathcal{L}(\hat{\mu} | \hat{\theta})},$$

where the $\hat{\mu}$ and $\hat{\hat{\theta}}$ are the parameters that maximize the likelihood with the constraint $0 \leq \hat{\mu} \leq \mu$, and $\hat{\theta}$ are the nuisance parameter values that maximize the likelihood for a given μ . Distributions of the test statistic are generated through pseudo-data by repeatedly

¹⁰These parameters allow for varying each bin independently from the other bins. In the context of this analysis, these constraints correspond to the total MC statistical uncertainty in each bin.

¹¹A test statistic is any function of the data (MC simulated data, pseudo-data generated from MC distributions, or actual data) which characterizes a full dataset in a single number [141]. It is critical to determining confidence levels as these quantities are based on distributions of the test statistic.

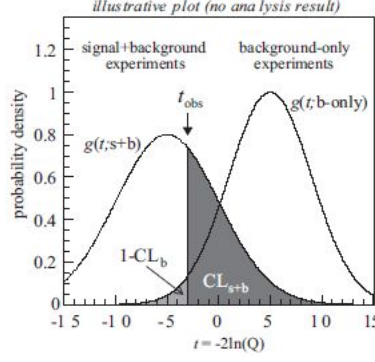


Figure 7.5: Demonstration of the meaning of CL_{S+B} and $1 - CL_B$ on a toy data set. In practice, the test statistic, t , distributions for the background-only and signal+background hypotheses are built from pseudo-data. Observed limits are derived by evaluating the test statistic on data, t_{obs} , and plugging this into the limits mentioned in the text [141].

varying the parameters of the likelihood within their uncertainties and recomputing the values of the test statistic. The integrals of the test-statistic distribution play an important role in evaluating the statistical power of an analysis. These are $g(q_\mu | b - \text{only or } s + b)$ ¹², w.r.t. a reference value, q_{ref} ,

$$I_{b-\text{only or } s+b}^+[q_{\text{ref}}] = \int_{q_{\text{ref}}}^{\infty} g(q' | b - \text{only or } s + b) dq',$$

and

$$I_{b-\text{only or } s+b}^-[q_{\text{ref}}] = \int_{-\infty}^{q_{\text{ref}}} g(q' | b - \text{only or } s + b) dq'.$$

The q_{ref} in these equations can either be the observed (in actual data) or expected (from the median of the signal-plus-background distribution) test statistic value. Which integral is used depends on whether the signal+background hypothesis tends to more positive or negative values of the test statistic. I will denote the integral in the direction of more signal+background-like as I^{sig} and in the opposite direction as I^{back} . A confidence level (CL) is synonymous with I^{back} while a p -value is $1 - CL = I^{\text{sig}}$. The test statistic is used for exclusion intervals derived with the CL_S method [142, 143] (the largest value of μ with a corresponding p -value less than a set $1 - CL$ value is quoted). This method for exclusion

¹² $b - \text{only}$ denotes background only and $s + b$ denotes signal+background

limits is chosen to safeguard against prematurely excluding a signal due to a downward fluctuation in data (or if the analysis is too insensitive to the presence of signal, but this isn't the case here) [141]. The CL_S is a ratio of confidence levels (CL_{S+B}/CL_B) and must be interpreted with caution. CL_{S+B} and CL_B are both derived from pseudo-datasets that produce distributions of the test statistic (denoted $g(q|b - \text{only or } s + b)$, Figure 7.5). The CL_B is $I_{b-\text{only}}^{\text{back}}$ and CL_{S+B} is I_{s+b}^{back} . To measure the compatibility of the background-only hypothesis with the observed data, the test statistic used is $q_0 = -2 \ln \Lambda(0, \hat{\theta}_0, \hat{\mu}, \hat{\theta})$. The results are presented in terms of: the 95% CL_S upper limit on the signal strength; the probability p_0 of the background-only hypothesis ($1 - CL_B$); and the best-fit signal-strength value $\hat{\mu}$ with its associated uncertainty σ_μ . These represent the potential for exclusion (i.e. if $CL_S < 5\%$), the potential for discovery (i.e. if $p_0 < 5.73 \times 10^{-7}$ (5σ)), and the best-fit signal strength compatible with the data given the associated uncertainties, respectively. The effectiveness of any experimental search can be summarized by its expected exclusion or discovery power, collectively known as the expected sensitivity. The expected exclusion limit is given by the CL_S using the $b - \text{only}$ hypothesis evaluated at the median of the $b - \text{only}$ test statistic distribution. The expected discovery power is given by the p -value using the $b - \text{only}$ hypothesis evaluated at the median of the $s + b$ test statistic distribution. The fitted $\hat{\mu}$ value is obtained by maximizing the likelihood function with respect to all parameters. The uncertainty σ_μ is obtained from the variation of $2 \ln \Lambda(\mu, \hat{\theta}, \hat{\mu}, \hat{\theta})$ by one unit, where $\Lambda(\mu, \hat{\theta}, \hat{\mu}, \hat{\theta})$ in this context is now defined without the constraint $0 \leq \hat{\mu} \leq \mu$. Expected results are obtained in the same way as the observed results by replacing the data in each input bin by the expectation from simulation with all NPs set to their best-fit values, as obtained from the fit to the data¹³.

The 8 categories used in the global likelihood fit are outlined in Chapter 6. The data have sufficient statistical power to constrain the largest background-normalization NPs, which are left free to float in the fit. This applies to the $t\bar{t}$, W +HF, and Z +HF processes. The $t\bar{t}$ background is normalized separately in the 0+1 and 2 lepton channels due to the different

¹³This type of pseudo-data sample is referred to as an Asimov dataset [143].

regions of phase space probed in the two cases. In order to assess the robustness of the fitting procedure and isolate the fit’s performance in different regions, additional fits are performed: separately for each lepton channel, combined for all channels but with three independent signal strength parameters one for each lepton channel, and combined for all channels with two independent signal strength parameters for the ZH and WH processes separately.

A “ranking” for the NPs is established by performing the fit again for each NP, with the corresponding NP fixed to its previously fitted value, $\hat{\theta}$, shifted up or down by its previously fitted uncertainty with all the other parameters allowed to vary to take properly into account the correlations between systematic uncertainties. The magnitude of the shift in the fitted signal strength $\hat{\mu}$ is a measure of the observed impact of the considered NP. The same procedure is repeated, using the nominal values of the NP and of its associated uncertainty to provide its expected impact.

The fit uses templates constructed from the predicted yields from MC simulation for the signal and the various backgrounds in the bins of the input distribution in each region. As explained earlier, the systematic uncertainties are encoded in templates of variations relative to the nominal template for each up and down ($\pm 1\sigma$) variation. To reduce the impact of statistical fluctuations, which can be introduced when a systematic variation alters the event selection, a smoothing procedure is applied where bins are merged based on the constraints that the statistical uncertainty in each bin should be less than 5%.

7.7. Results

The results of the Higgs boson search and the di-boson analysis are given below. In the rest of the chapter the fitted signal-strength parameters are denoted μ and μ_{VZ} rather than $\hat{\mu}$ and $\hat{\mu}_{VZ}$.

7.7.1. Higgs Boson Search

The factors applied to the nominal background normalizations resulting from the global likelihood fit are shown in Table 7.4 and shows that the overall background normalizations before the fit ($\theta_0 = 1$) are roughly within the 1σ range of their post-fit values suggesting that the rates of the background processes are well-modeled. The post-global likelihood fit signal and background yields are shown in Table 7.5 for all the analysis regions. As expected, the $V+HF$ and $t\bar{t}$ backgrounds dominate all analysis regions. The yields in data are all within the 1σ error band of the background rates suggesting that extracting evidence of a signal process will be quite difficult (barring any additional shape information not in this table). A selection of characteristic post-fit plots including the BDT output distributions (for the most sensitive 2-jet, high p_T^V categories) for each of the lepton channels are shown in Figures 7.6, 7.7, and 7.8. In all of these plots the background contributions after the global likelihood fit are shown as filled histograms. The Higgs boson signal ($m_H = 125$ GeV) is shown as a filled histogram on top of the fitted backgrounds as expected from the SM (indicated as $\mu = 1.0$). The dashed histogram shows the total background as expected from the pre-fit MC simulation. The size of the combined statistical and systematic uncertainty on the sum of the signal and fitted background is indicated by the hatched band. The ratio of the data to the sum of the signal and fitted background is shown in the lower panel. All of the plots show marked improvement in the agreement between data and MC simulation post fit indicating that any disagreement pre-fit is covered by the systematic variations and floating normalizations.

Sample	Scale factor
$t\bar{t}$ 0+1-lepton	0.86 ± 0.13
$t\bar{t}$ 2-lepton	0.94 ± 0.09
W +HF	1.59 ± 0.39
Z +HF	1.04 ± 0.11

Table 7.4: Factors applied to the nominal normalizations ($\theta_0 = 1$) of the $t\bar{t}$, W +HF, and Z +HF backgrounds, as obtained from the global likelihood fit. The $t\bar{t}$ background is normalized independently in the 0+1 and 2 lepton channels. The errors include the statistical and systematic uncertainties.

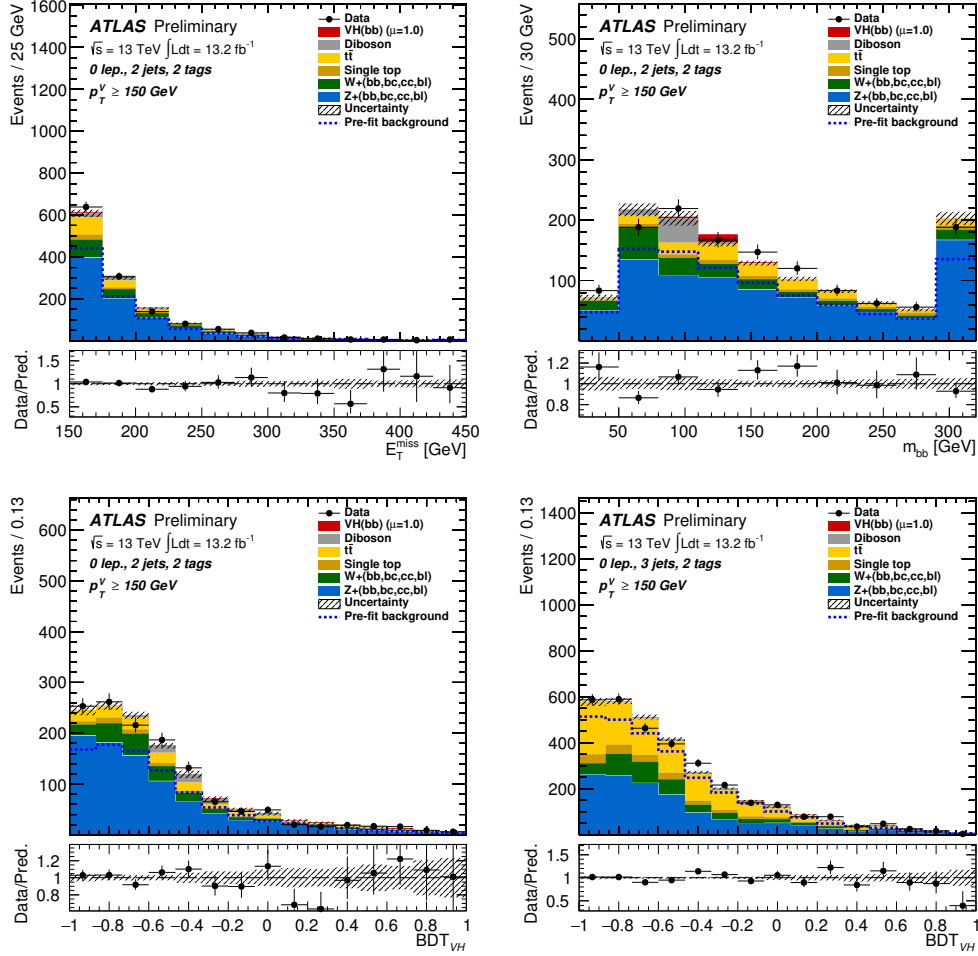


Figure 7.6: The E_T^{miss} (top left) and $m_{b\bar{b}}$ (top right) post-fit distributions in the 0-lepton channel for 2-jet, 2 b -tag events in the high p_T^V region. The BDT_{VH} post-fit distributions (actual fitted distribution) in the 0-lepton channel for 2-tag events, in the 2-jet (bottom left) and 3-jets (bottom right) categories in the high p_T^V region. Description in the text.

For all lepton channels combined, the observed 95% confidence level¹⁴ upper limit on μ is 1.2, to be compared to an expected 95% CL upper limit of $1.0^{+0.4}_{-0.3}$. The probability p_0 of obtaining from background alone a result at least as signal-like as the observation is 34% for a tested Higgs boson mass of 125 GeV. In the presence of a Higgs boson with that mass and the SM signal strength, the expected p_0 value is 3%. This corresponds to an observed excess with a significance¹⁵ of 0.42σ , to be compared to an expectation of 1.94σ . Table 7.6

¹⁴A 95% CL implies that 95% of the confidence intervals would include the true value parameter.

¹⁵Statistical significance is often given in terms of σ 's. The relationship between a p value and $\sqrt{Z}\sigma$ is

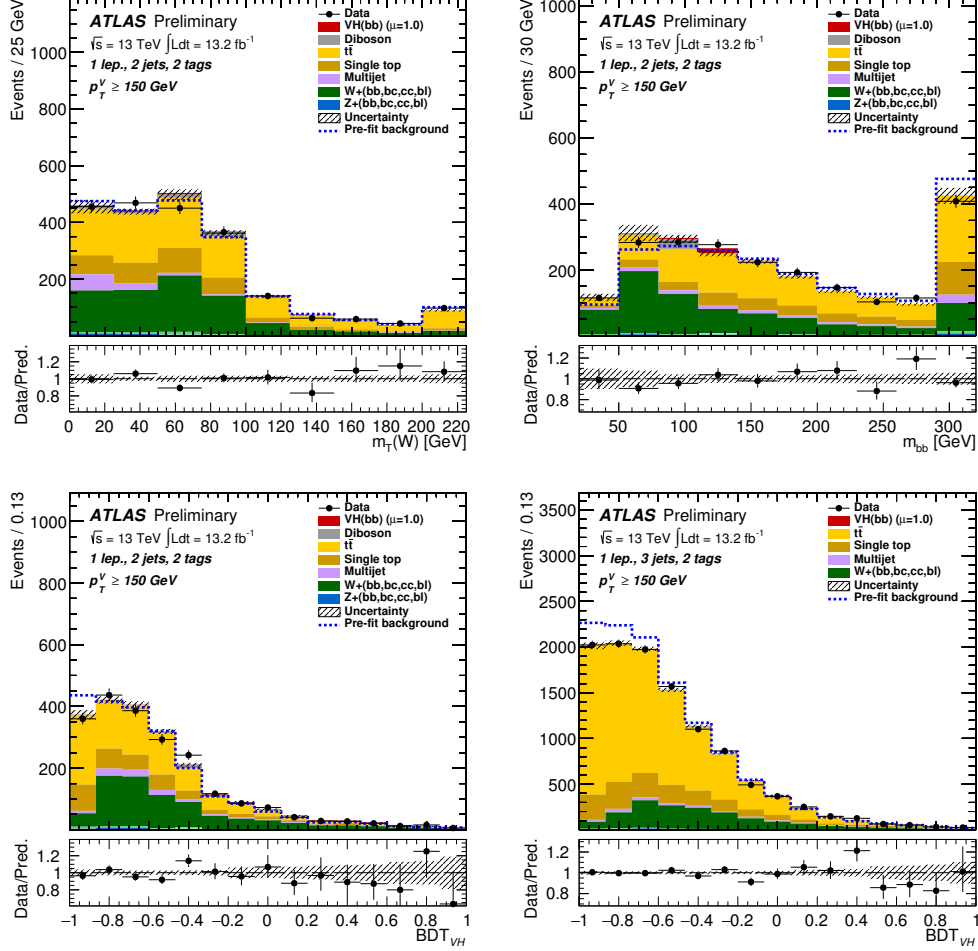


Figure 7.7: The m_T^W (top left) and $m_{b\bar{b}}$ (top right) post-fit distributions in the 1-lepton channel for 2-jet, 2 b -tag events in the high p_T^V region. The BDT_{VH} post-fit distributions (actual fitted distribution) in the 1-lepton channel for 2-tag events, in the 2-jet (bottom left) and 3-jets (bottom right) categories in the high p_T^V region. Description in the text.

shows the expected and observed 95% CL limits, p_0 , and significance values for the separate lepton channel fits and for the lepton channels combined in the global fit. For all channels combined the fitted value of the signal-strength parameter is $\mu = 0.21^{+0.36}_{-0.35}(\text{stat.}) \pm 0.36(\text{syst.})$. Fits are also performed in the case of the three lepton channels combined, where the signal strengths are allowed to float independently for (i) the WH and ZH production processes, or (ii) the three lepton channels, but leaving all other NPs with the same correlations as the nominal result. The results of these fits are shown in Figure 7.9. Figure 7.10 shows

$$\int_Z^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt = p$$

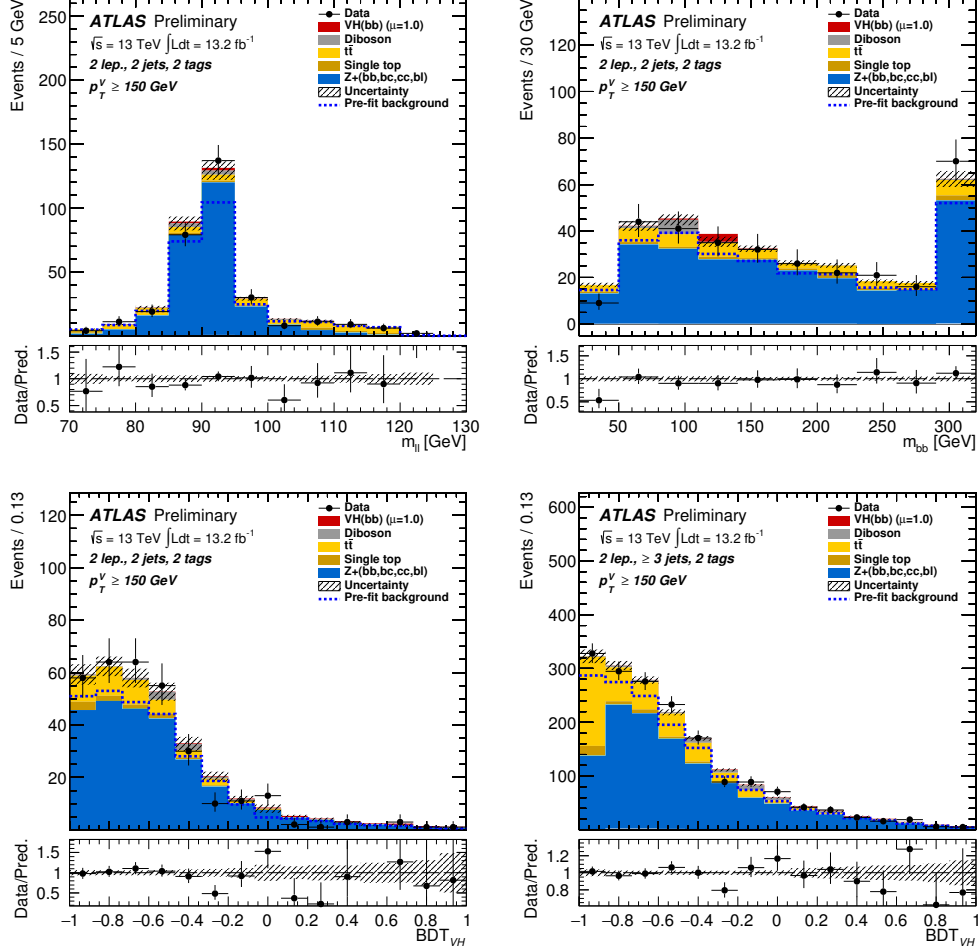


Figure 7.8: The m_{ll} (top left) and m_{bb} (top right) post-fit distributions in the 2-lepton channel for 2-jet, 2 b -tag events in the high p_T^V region. The BDT_{VH} post-fit distributions (actual fitted distribution) in the 2-lepton channel for 2-tag events, in the 2-jet (bottom left) and ≥ 3 -jets (bottom right) categories in the high p_T^V region. Description in the text.

the data, background and signal yields, where final-discriminant bins in all signal regions are combined into bins of $\log S/B$. Here, S is the expected signal yield and B is the fitted background yield. As shown in the plot, the bins with the most signal seem to have a downward statistical fluctuation in the data. It remains to be seen if this will be resolved after collecting more data.

The ranking of the systematic uncertainties in the global likelihood fit are shown in Figure 7.11. The NPs are ordered by decreasing post-fit impact on μ . The five systematic uncertainties with the largest impact are due to the b -jet efficiency uncertainties, the floating

Dataset	Upper Limit		p_0		Significance	
	Exp.	Obs.	Exp.	Obs.	Exp.	Obs.
0-lepton	$1.4^{+0.6}_{-0.4}$	2.0	0.07	0.15	1.45	1.02
1-lepton	$2.0^{+0.8}_{-0.6}$	2.1	0.15	0.46	1.04	0.10
2-lepton	$1.8^{+0.7}_{-0.5}$	1.7	0.13	0.57	1.14	-0.17
Combined	$1.0^{+0.4}_{-0.3}$	1.2	0.03	0.34	1.94	0.42

Table 7.6: The expected and observed 95% CL limits on μ and p_0 with their respective significance values for the individual lepton channels and their combination. The expected limits are evaluated assuming the absence of signal and the expected p_0 and significance assuming a Higgs boson of 125 GeV mass with the SM signal strength.

background normalizations, and the $\text{jet}/E_{\text{T}}^{\text{miss}}$ systematics. Overall, the largest sources of uncertainties are due to the limited number of real data events (denoted Data Stat.) and too few MC simulated events passing selection (denote MC Stat.).

Since the expected and observed signal strengths and significances are quite different (though within 2σ), a di-boson signal cross check is made to ensure a sound analysis.

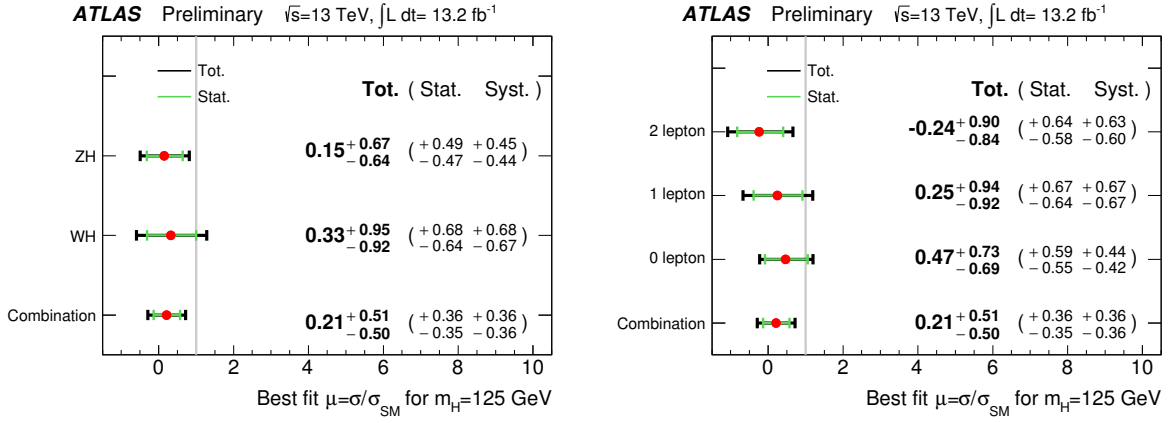


Figure 7.9: The fitted values of the Higgs boson signal-strength parameter, μ , for $m_H = 125$ GeV for the WH and ZH processes and their combination (left) and for the 0-, 1-, and 2-lepton channels and their combination (right). The individual μ values in either the case of the $(W/Z)H$ processes or individual lepton channels are obtained from a simultaneous fit with the signal strength for each of the processes or lepton channels floated independently.

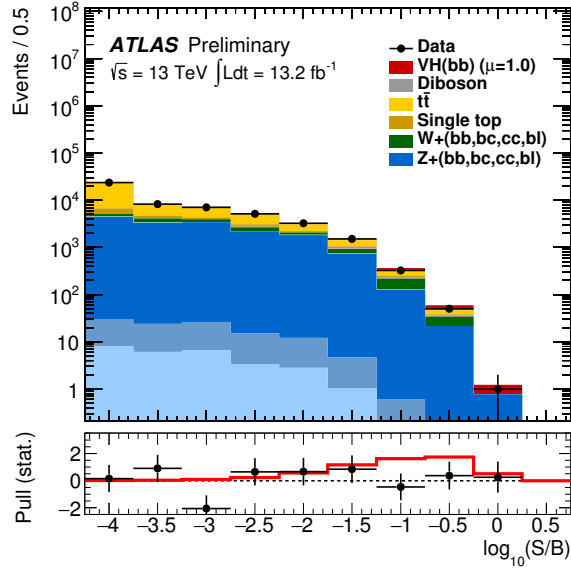


Figure 7.10: Event yields as a function of $\log(S/B)$ for data, background and Higgs boson signal with $m_H = 125 \text{ GeV}$. Final-discriminant bins in all signal regions are combined into bins of $\log(S/B)$. The signal S and background B yields are the expected and fitted values, respectively. The Higgs boson signal contribution is shown as expected for the SM cross section (indicated as $\mu = 1.0$). The difference between the data with respect to the background-only prediction (pull) is shown without systematic uncertainties. The solid red line indicates the pull of the prediction for signal ($\mu = 1.0$) and background with respect to the background-only prediction.

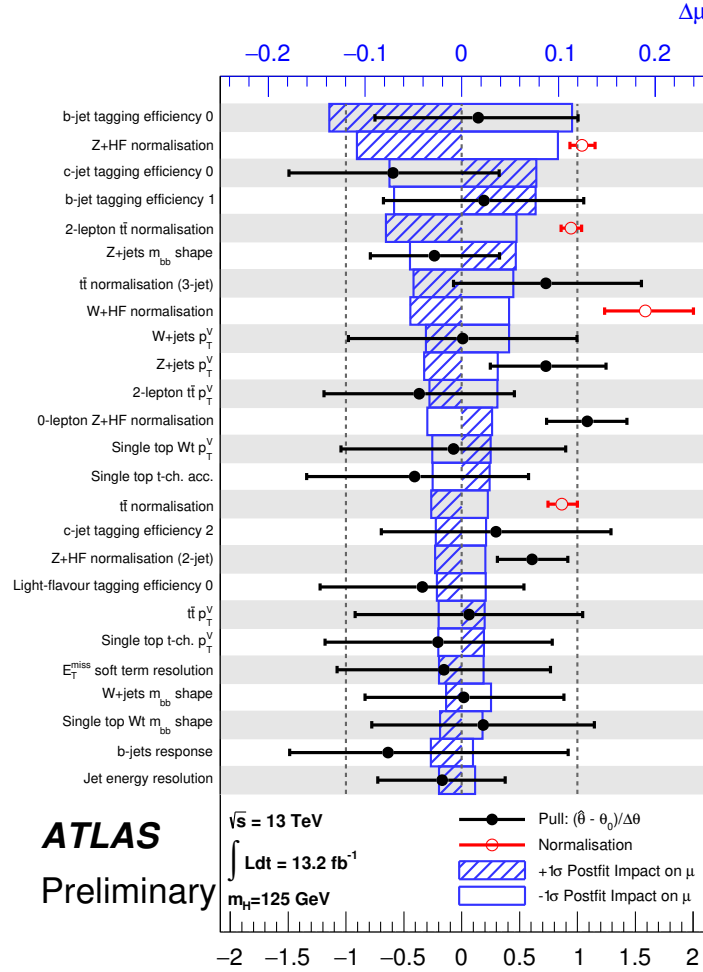


Figure 7.11: Impact of systematic uncertainties on the fitted signal-strength parameter μ . The systematic uncertainties are listed in decreasing order of their impact on μ on the y -axis. The boxes show the variations of μ , referring to the top x -axis, when fixing the corresponding individual nuisance parameter θ to its post-fit value $\hat{\theta}$ modified upwards or downwards by its post-fit uncertainty, and repeating the fit as explained in the text. The hatched and open areas correspond to the upwards and downwards variations, respectively. The filled circles, referring to the bottom x -axis, show the deviations of the fitted nuisance parameters $\hat{\theta}$ from their nominal values θ_0 , expressed in terms of standard deviations with respect to their nominal uncertainties $\Delta\theta$. The associated error bars show the post-fit uncertainties of the nuisance parameters, relative to their nominal uncertainties. The open circles with their error bars, also referring to the bottom x -axis, show the fitted values and uncertainties of the normalization parameters that are freely floating in the fit. The normalization parameters have a pre-fit value of one. As explained in Section 7.6.1, the jet energy scale and b -tagging uncertainties are decomposed into uncorrelated components; the numerical labels refer to such components.

7.7.2. Di-boson Validation

The validation results are obtained using the maximum likelihood fits described in Section 7.6.2 in an identical manner as for the signal fit, but using the BDT_{VZ} output as the final discriminant and with the VZ signal-strength parameter, μ_{VZ} , freely floating. The di-boson and Higgs boson BDTs provide sufficient separation between the VZ and VH processes such that they only have a weak correlation in their results. The Higgs boson signal-strength parameter, μ , is set to the SM prediction with a 50% uncertainty assigned on the normalization. A value of $\mu_{VZ} = 0.91 \pm 0.17(\text{stat.})^{+0.32}_{-0.27}(\text{syst.})$ is obtained. The VZ signal is observed with a significance of 3.0σ , to be compared to an expected significance of 3.2σ . Although the VZ signal has slight kinematic differences from the VH signal such as lower p_{T}^V , the consistency between the signal strength and significance for this cross check implies that the Higgs boson analysis is sound and needs more data in order to resolve its issue with low signal strength (this was also hinted at in the yields table among other things).

Chapter 8

2HDM $A \rightarrow Zh(h \rightarrow b\bar{b})$

The BSM $A \rightarrow Zh(h \rightarrow b\bar{b})$ analysis in Run 2 continues the ATLAS Run 1 search for the 2HDM heavy pseudo-scalar particle (A boson) that then decays to a Z boson and a 2HDM light scalar h (taken to be very similar to the SM Higgs boson). The decays of the Z and h bosons mirror the decay channels of the SM VH analysis with the h standing in for the SM H . The A -boson masses (m_{Zh}) used in this search are 220–500 GeV in 20 GeV steps, 500–1000 GeV in 50 GeV steps, and 1000–2000 GeV in 100 GeV steps for a total of 35 mass hypotheses. As hinted in the chapter on object and event selection (Chapter 6), several aspects of this analysis are different from the SM VH in addition to the unnecessary 1-lepton channel. Most important is the fact that there are several additional analysis regions specifically designed to isolate the various background components. These additional regions are denoted control regions (CRs) and the details of how they are defined is given in Chapter 6. The discriminating variable used in this analysis is the reconstructed invariant mass of the A boson rather than the output of a BDT. However, in the 0-lepton channel it is not possible to fully reconstruct the invariant mass of the Zh system due to the neutrinos present in the final state. Therefore, the transverse mass is used as the final discriminant

$$m_{\text{T}}^{Zh} = \sqrt{(E_{\text{T}}^h + E_{\text{T}}^{\text{miss}})^2 - (\vec{p}_{\text{T}}^h + \vec{E}_{\text{T}}^{\text{miss}})^2},$$

where $\vec{p}_{\text{T}}^h$ (E_{T}^h) is the transverse momentum (energy) of the $h \rightarrow b\bar{b}$ candidate. Additionally, there is no use of a BDT for signal extraction, no need for a multi-jet estimate, no truth tagging, and a kinematic fit for the 2-lepton channel is not possible as the likelihood would need to be retrained for each A -boson mass hypothesis. The material presented in this section is documented in the dedicated ATLAS paper on this search (Ref. [100]).

This analysis is the product of a team to which I contributed. My contributions to this analysis are technical implementation and upkeep of the flavor-tagging recommendations, preliminary flavor-tagging studies on which working point, systematics set to use, and large-R jet tag categorization, technical development of the statistical framework, and several studies related to the global likelihood fit including how to incorporate the flavor-tagging systematics.

8.1. *b*-Jet Energy Corrections

The resolved regime of the $A \rightarrow Zh$ analysis uses an identical set of *b*-jet energy corrections as in the SM VH (Figure 8.1). The muon-in-jet correction is applied to the merged regime (with a radius of association increased to $\Delta R < 1.0$), but not the PtReco correction as this was found to have a negligible impact on the reconstructed mass of the Higgs boson [99].

8.2. Analysis

The $A \rightarrow Zh$ uses a fundamentally similar approach to the SM VH analysis in constructing a theoretical model (MC simulation) and data analysis. However, since the A boson's mass is unknown several mass hypotheses need to be explored. Thus, instead of a single signal strength PoI, there are as many PoI as mass hypotheses. The upper limits at the 95% CL for the multiple PoIs at sequential mass hypothesis are usually presented in terms of a so-called Brazil¹-band plot where all mass hypotheses are shown simultaneously. My contributions in this area consist of several studies regarding the fit performance (i.e. altering the way or which systematics are included in the fit model, which regions are correlated in the fit model, etc.), exploring techniques for smoothing the tails of background distributions, and updating the statistical framework codebase.

¹Brazil, because the colors chosen resemble the Brazilian flag.

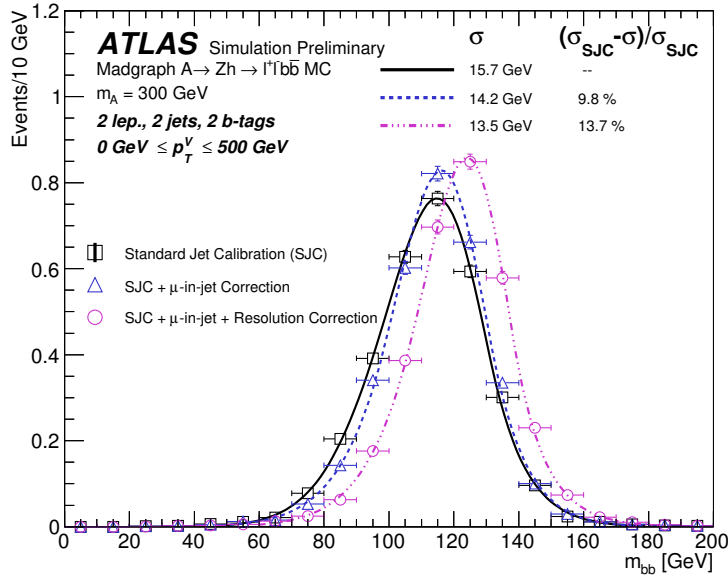


Figure 8.1: Di-jet invariant mass distribution for the decay products of a Higgs boson with $m_h = 125$ GeV in the 2-lepton low- p_T^Z 2-tag category in simulated $gg \rightarrow A \rightarrow Zh \rightarrow l^+l^-b\bar{b}$ events. The distributions are shown using jets after the standard jet calibration, after the muon-in-jet correction, and finally after correcting for bias effects specific to the kinematics of the decay of a Higgs boson with $m_h = 125$ GeV (resolution correction). The distributions are fit using the Bukin function and the parameter representing the width of the core of the distribution, σ_{SJC} , is shown above, as well as the relative improvement in the resolution with respect to jets after the standard jet calibration.

8.2.1. bbA

The efficiencies of selecting events in the 2-tag signal region for an $A \rightarrow Zh$ resonance with $m_A = 600$ GeV generated via gluon-gluon fusion are 12.3% and 15.3% for $A \rightarrow Zh \rightarrow \nu\nu b\bar{b}$ and the average of $A \rightarrow Zh \rightarrow (e^+e^-, \mu^+\mu^-)b\bar{b}$, respectively. In the 1-tag signal region, the efficiencies are 7.0% and 7.4%, respectively. While the analysis has been optimized for the generation via gluon-gluon fusion, it has significant efficiency for the b -quark associated production (58%–64% of that found for the gluon-gluon fusion process for the same A boson mass). This is important to take into account when exploring the full two-dimensional 2HDM parameter space discussed later.

8.2.2. CR Analysis

As detailed in Chapter 6, the dominant backgrounds are studied using CRs that invert selection criteria to create independent regions of the data and MC simulation that allow specific backgrounds to be enhanced. The CRs are represented graphically in Figure 6.4.

The Z +jets background is studied using a set of “ W/Z +jets” CRs defined by inverting the selection criteria on $m_{\text{di-jet}}$ (m_{jet}) in the low- p_{T}^Z (high- p_{T}^Z) category. Two sub-regions of this CR are defined by requiring either that two jets are b -tagged, or that only one jet is b -tagged. The reconstructed m_{T}^{Zh} and m_{Zh} mass distributions for events passing the selections for these CRs are shown in Figure 8.2.

In the 2-lepton low- p_{T}^Z channel, two additional $t\bar{t}$ -enriched CRs (one each for 1-tag and 2-tag) are defined by requiring that the two charged leptons have different lepton flavors and that $m_{e\mu} > 40 \text{ GeV}$. Events must satisfy all other selection criteria for the SR except the requirement on $E_{\text{T}}^{\text{miss}}/\sqrt{H_{\text{T}}}$ and the $m_{\text{di-jet}}$ mass range. The reconstructed m_{Zh} mass distributions for events passing the selections for these $t\bar{t}$ -enriched CRs are shown in Figure 8.3.

The level of agreement obtained between data and background estimates in the CRs implies good control of the background modeling.

8.2.3. Systematics

The systematics used in this analysis are quite similar to the ones in the SM VH with several additional systematics added for the A -boson signal MC simulation. Of course, there are no multi-jet background uncertainties as this background is negligible in the 0- and 2-lepton channels. Due to this similarity with Chapter 7, I will give a shorter discussion here.

All experimental uncertainties are virtually identical to the SM VH analysis save that the b -jet flavor-tagging uncertainties were still based on a Run 1-to-Run 2 extrapolation as data-driven estimates were not yet available². Additionally, the uncertainty on the scale and

²This analysis was completed chronologically before the SM VH analysis.

resolution of large- R jet energy and mass are evaluated by comparing the ratio of calorimeter-based to track-based measurements in di-jet events in data and simulation [85].

Uncertainties are also taken into account for possible differences between data and the simulation model that is used for each process. For the most significant background sources, $Z + (bb, bc, cc)$ production, $Z + (bl, cl)$, and $t\bar{t}$ production, the normalizations are fully determined by the global likelihood fit. For the other sources (with normalizations taken from theoretical predictions), in addition to the 5% uncertainty on the integrated luminosity, the following normalization uncertainties are assigned: 26% for $Z + l$, 30% for $W + (bb, bc, cc, bl)$, 30% for $W + cl$, and 10% for $W + l$. These additional contributions to the normalization uncertainty take into account the uncertainty on the inclusive W/Z +jets cross section [144] and the uncertainty on the experimental acceptance from comparisons of the nominal SHERPA 2.1.1 W/Z +jets samples with alternative samples where the factorization, renormalization, and resummation scales have been either doubled or halved, or where the matching scale has been varied in the range 15 GeV–30 GeV. In addition, the following normalization uncertainties are also assigned to the luminosity uncertainty on background processes: 4% for single-top in the s - and t -channels, 7% for single-top in the Wt -channel [109, 145], and 50% for the associated $W/Z + h$ production [146]. An additional 13% (35%) uncertainty is assigned to the relative normalization of $Z + cl$ to $Z + bl$ ($W + bl$ to $W + bb$) from comparisons of the SHERPA 2.1.1 samples.

Uncertainties of about 30% on the relative normalization of the high- p_T^Z and low- p_T^Z regions and in the range 9% to 17% on the relative normalization of the 0- and 2-leptons channels have been estimated for each individual background source by comparing SHERPA 2.1.1 and MADGRAPH5_AMC@NLO 2.2.2 at leading order interfaced to PYTHIA 8.186 for the W/Z +jets backgrounds. The comparison of the predictions from SHERPA 2.1.1 and MADGRAPH5_AMC@NLO 2.2.2 are used to estimate the uncertainties on the modeling of the m_{Zh} and m_T^{Zh} distributions for the W/Z +jets backgrounds.

For the $t\bar{t}$ background, an uncertainty of 30% on the relative normalization of the high- and low- p_T^Z categories, and uncertainties in the range of 2.5%–10% on the relative normaliza-

tion of the signal regions and the CRs, have been estimated by comparing samples produced with POWHEG and with MADGRAPH5_AMC@NLO 2.2.2, and samples produced with the factorization and renormalization scale either doubled or halved. This comparison is also used to estimate the uncertainties on the shape m_{Zh} and m_T^{Zh} distributions in $t\bar{t}$ events.

The sources of uncertainty considered on the cross sections for the di-boson production (WW , WZ and ZZ) are the renormalization and factorization scales (in the range 10% - 17% for $p_T^Z > 200$ GeV, with a dependence on p_T^Z and on the number of jets), the choice of PDFs (in the range 2% to 4%) and the choice of parton-shower and hadronization model (of about 20% for $m_{\text{di-jet}}$ around 125 GeV).

Uncertainties on the signal acceptance from the choice of PDFs (in the range 1% to 4%), from the choice of factorization and renormalisation scales (of about 1%), and from the choice of parton-shower and underlying event tune (in the range 1% to 4%) have been taken into account in the analysis. In addition, uncertainties arising from the limited number of simulated events are taken into account.

The uncertainty on the cross section of an $A \rightarrow Zh$ signal process for $m_A = 600$ GeV after the maximum likelihood fit to data is mainly due to the limited number of events in data (which is responsible for almost 80% of the total uncertainty). The limited number of MC simulated events, the modeling systematics, and the experimental uncertainties give a smaller contribution to the uncertainty on the cross section.

8.3. Results

The data are fitted using models for the expected signal and background contributions with a binned maximum likelihood approach (using an identical HistFactory template as in the SM VH analysis) in which the low- and high- p_T^Z categories are handled simultaneously. The fit is performed on the m_{Zh} distribution in the 2-lepton channel, while in the 0-lepton channel the fit is performed over m_T^{Zh} . Both lepton channels are handled simultaneously in the fit. The CRs are included in the fit to constrain backgrounds in the SRs. In addition, the data are fitted in categories according to the number of b -tagged jets in the events

(1-tag and 2-tag) to provide additional information about the flavor components of jets in the W/Z +jets background events. Systematic uncertainties are incorporated into the fit as nuisance parameters with Gaussian constraints, with correlations across regions and processes taken into account.

The expected and observed numbers of events in each signal region are shown in Table 8.1. The numbers predicted for an $A \rightarrow Zh$ signal process assume $\sigma(gg \rightarrow A) \cdot \text{BR}(A \rightarrow Zh) \cdot \text{BR}(h \rightarrow b\bar{b}) = \sigma(pp \rightarrow A b\bar{b}) \cdot \text{BR}(A \rightarrow Zh) \cdot \text{BR}(h \rightarrow b\bar{b}) = 113 \text{ fb}$ for $m_{Zh} = 600 \text{ GeV}$. This cross section corresponds to the expected 95% CL upper limit on the $\sigma(gg \rightarrow A) \cdot \text{BR}(A \rightarrow Zh) \cdot \text{BR}(h \rightarrow b\bar{b})$ process at this mass. The reconstructed m_{T}^{Zh} and m_{Zh} distributions for events passing the selections are shown in Figures 8.4 and 8.5 for the low- and the high- p_{T}^Z categories. The background expectation is shown after the binned maximum likelihood fit to the data. No significant excess beyond the expectation from the background prediction is observed in the data. Two upward deviations from the background-only hypothesis occur at masses $m_A = 260 \text{ GeV}$ and 440 GeV , each corresponding to a local significance of about 2σ .

Exclusion limits are determined using the modified frequentist CL_s method and the same profile-likelihood test statistic as the SM VH analysis. The 95% confidence level (CL) upper limits on the product of production cross section ($pp \rightarrow A$) and branching fractions ($A \rightarrow Zh$ and $h \rightarrow b\bar{b}$) is shown as a function of the resonance mass in Figure 8.7 for different production mechanisms: pure gluon-gluon fusion and pure b -quark associated production. The expected upper limits for $gg \rightarrow A \rightarrow Zh \rightarrow l^+ l^- b\bar{b}$ for the 2-lepton and 0-lepton categories are shown separately and combined in Figure 8.7(a). The 2-lepton category contributes most strongly in the combined limit for $m_A < 800 \text{ GeV}$ due to the higher resolution on m_{Zh} , while the 0-lepton category contributes most strongly for $m_A > 800 \text{ GeV}$ due to the large value of $\text{BR}(Z \rightarrow \nu\bar{\nu})$ compared to $\text{BR}(Z \rightarrow l\bar{l})$. The high- p_{T}^Z category contributes most strongly for $m_A > 1100 \text{ GeV}$. The limits presented in Figure 8.7 improve upon the results presented in Ref. [147] for $m_A \gtrsim 800 \text{ GeV}$.

The data are further interpreted in the context of Type-I and Type-II 2HDMs³ as shown in Figure 8.8. The 95% CL exclusion regions are shown for $m_{Zh} = 600 \text{ GeV}$ in the plane of $\tan \beta$ and $\cos(\beta - \alpha)$. The width of the A boson is assumed in simulation to be narrow, but is corrected to the width (Γ_A) predicted by the 2HDM for each point in the parameter space using a relativistic Breit-Wigner function. Only points in parameter space where $\Gamma_A/m_{Zh} < 5\%$ are considered.

³see Section 2.4 for a detailed description in how interpret these plots

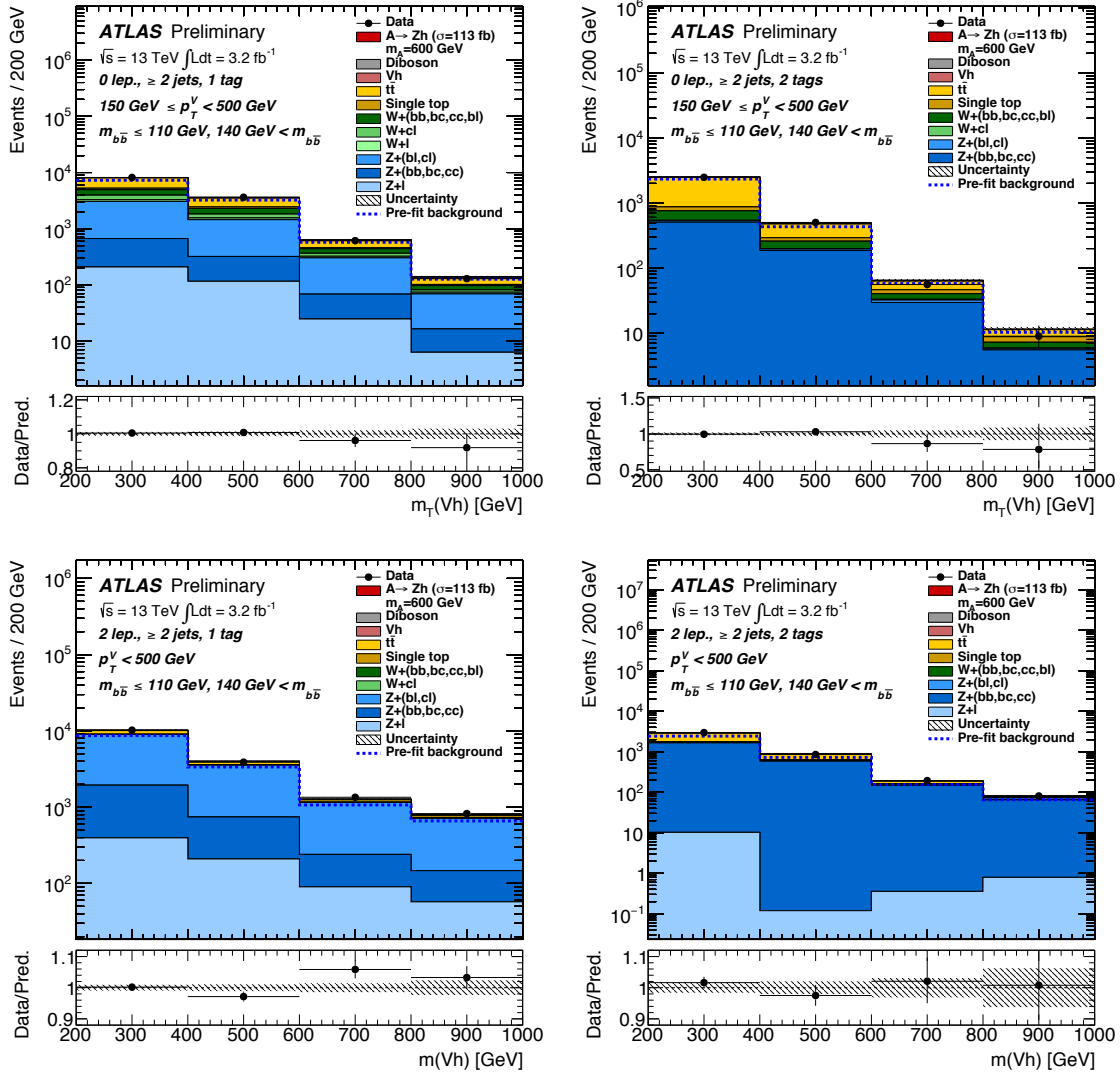
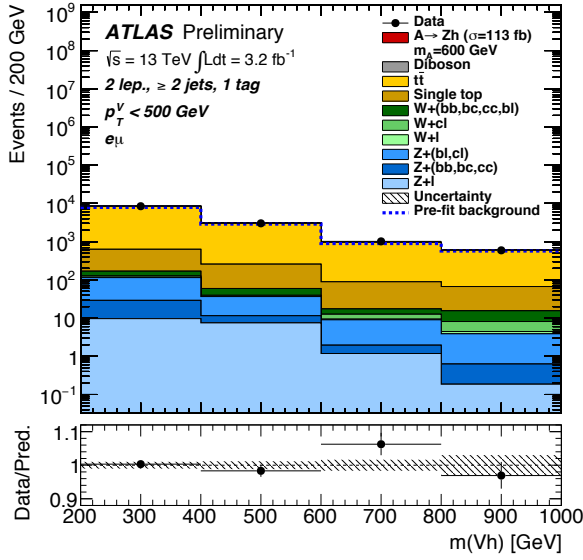
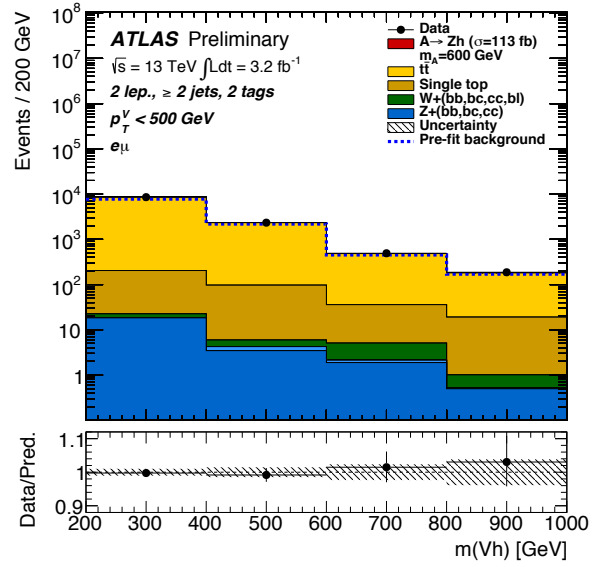


Figure 8.2: Distributions of m_T^{Zh} and m_{Zh} in the low- p_T^Z categories for the 0-lepton (top), and 2-lepton (bottom) channels in the “ W/Z +jets” control region. The left (right) column corresponds to the 1-tag (2-tag) control regions. The background expectation is shown after the maximum likelihood fit to the data; the total background expectation before the fit is denoted by the dotted blue line. The ratio of the data to background prediction is shown below each plot, and contains the total uncertainty band (black diagonal hashing) on the background prediction.



(a)

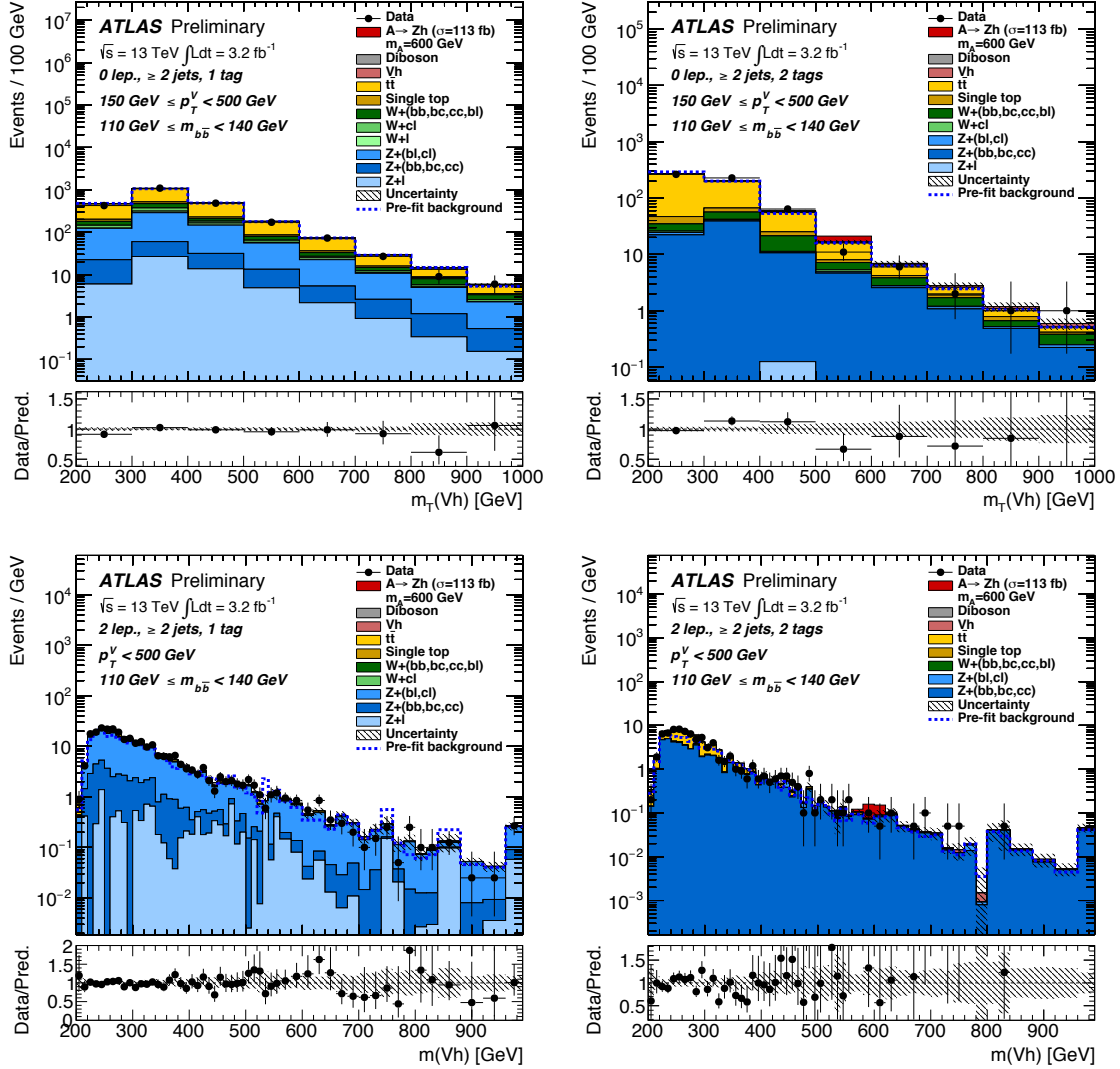


(b)

Figure 8.3: Distributions of m_{Zh} in the $t\bar{t}$ -enriched top- $e\mu$ control region. The left (right) plot corresponds to the 1-tag (2-tag) region. The background expectation is shown after the maximum likelihood fit to the data; the total background expectation before the fit is denoted by the dotted blue line. The ratio of the data to background prediction is shown below each plot, and contains the total uncertainty band (black diagonal hashing) on the background prediction.

One b -tag				
	0-leptons low- p_T^Z	0-leptons high- p_T^Z	2-leptons low- p_T^Z	2-leptons high- p_T^Z
$Z + l$	55 ± 31	2.0 ± 1.0	118 ± 38	0.57 ± 0.28
$Z + (cl, bl)$	518 ± 54	8.2 ± 1.8	1943 ± 65	3.18 ± 0.65
$Z + (bb, bc, cc)$	82 ± 13	1.82 ± 0.35	391 ± 23	0.74 ± 0.13
$W + l$	48 ± 22	–	–	–
$W + cl$	111 ± 42	1.17 ± 0.45	0.67 ± 0.28	–
$W + (bb, bc, cc, bl)$	185 ± 71	1.80 ± 0.63	10.1 ± 4.2	–
$t\bar{t}$	1202 ± 77	2.05 ± 0.72	276 ± 22	–
Single Top	99 ± 10	0.49 ± 0.11	15.3 ± 1.6	–
Di-boson	27.2 ± 4.6	5.3 ± 1.0	26.3 ± 6.1	0.67 ± 0.32
$(W/Z)h$	3.3 ± 1.6	0.16 ± 0.08	3.7 ± 1.8	0.04 ± 0.02
Total Background	2332 ± 45	23.0 ± 2.6	2784 ± 47	5.2 ± 0.9
Expected $A \rightarrow Zh$ (ggF)	5.00 ± 0.41	–	1.73 ± 0.12	–
Expected $A \rightarrow Zh$ ($b\bar{b}A$)	3.05 ± 0.26	–	1.01 ± 0.08	–
Data	2295	25	2769	8
Two b -tags				
	0-leptons low- p_T^Z	0-leptons high- p_T^Z	2-leptons low- p_T^Z	2-leptons high- p_T^Z
$Z + l$	0.13 ± 0.44	0.01 ± 0.01	–	–
$Z + (cl, bl)$	4.7 ± 1.8	0.06 ± 0.03	9.6 ± 4.0	0.02 ± 0.01
$Z + (bb, bc, cc)$	81 ± 13	0.75 ± 0.15	490 ± 22	0.36 ± 0.05
$W + l$	–	0.02 ± 0.01	–	–
$W + cl$	3.6 ± 2.1	0.04 ± 0.02	–	–
$W + (bb, bc, cc, bl)$	37 ± 14	0.28 ± 0.09	0.42 ± 0.20	–
$t\bar{t}$	392 ± 24	0.22 ± 0.11	284 ± 22	–
Single Top	27.8 ± 2.9	–	6.39 ± 0.60	–
Di-boson	6.1 ± 3.1	0.05 ± 0.03	0.57 ± 0.23	–
$(W/Z)h$	5.4 ± 2.6	0.10 ± 0.05	7.6 ± 3.7	0.03 ± 0.01
Total Background	557 ± 18	1.53 ± 0.25	799 ± 23	0.42 ± 0.05
Expected $A \rightarrow Zh$ (ggF)	9.21 ± 0.76	–	3.65 ± 0.29	–
Expected $A \rightarrow Zh$ ($b\bar{b}A$)	5.85 ± 0.47	–	2.21 ± 0.17	–
Data	577	0	788	0

Table 8.1: The numbers of expected and observed events for all the signal regions. The expectation is shown after the maximum likelihood fit to the data. The quoted uncertainties are the combined systematic and statistical uncertainties. The numbers predicted for a $A \rightarrow Zh$ signal process in gluon-gluon fusion and b -quark-associated production assume that $\sigma(gg \rightarrow A) \cdot \text{BR}(A \rightarrow Zh) \cdot \text{BR}(h \rightarrow b\bar{b}) = \sigma(pp \rightarrow A b\bar{b}) \cdot \text{BR}(A \rightarrow Zh) \cdot \text{BR}(h \rightarrow b\bar{b}) = 113 \text{ fb}$ for $m_{Zh} = 600 \text{ GeV}$. This cross section corresponds to the expected 95% CL upper limit on these effective cross sections at this mass. A dash indicates a negligible yield of events.



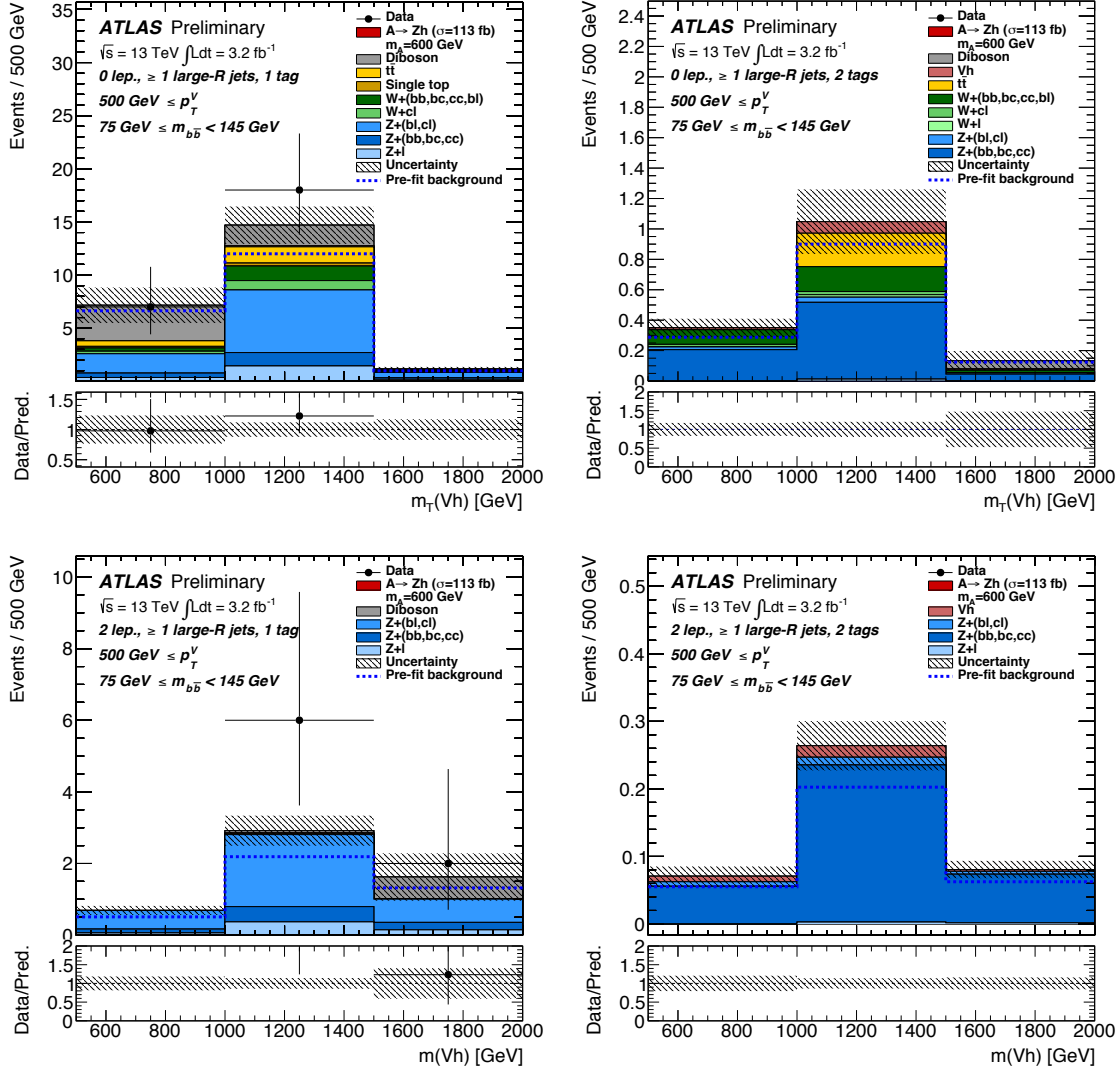


Figure 8.5: Distributions of m_T^{Zh} and m_{Zh} in the high- p_T^Z categories for the 0-lepton (2-lepton) channels are in the top (bottom) row. The left (right) column corresponds to the 1-tag (2-tag) signal regions. The background expectation is shown after the maximum likelihood fit to the data; the total background expectation before the fit is denoted by the dotted blue line. The signal for $gg \rightarrow A \rightarrow Zh$ with $m_A = 600$ GeV is shown stacked on top of the background and normalized to 113 fb, corresponding to the expected 95% CL upper limit on $\sigma(gg \rightarrow A) \cdot \text{BR}(A \rightarrow Zh) \cdot \text{BR}(h \rightarrow b\bar{b})$. The ratio of the data and background prediction is shown below each plot, and contains the total uncertainty band (black diagonal hashing) on the background prediction. The right-most bin in each plot contains all events in and above that bin.

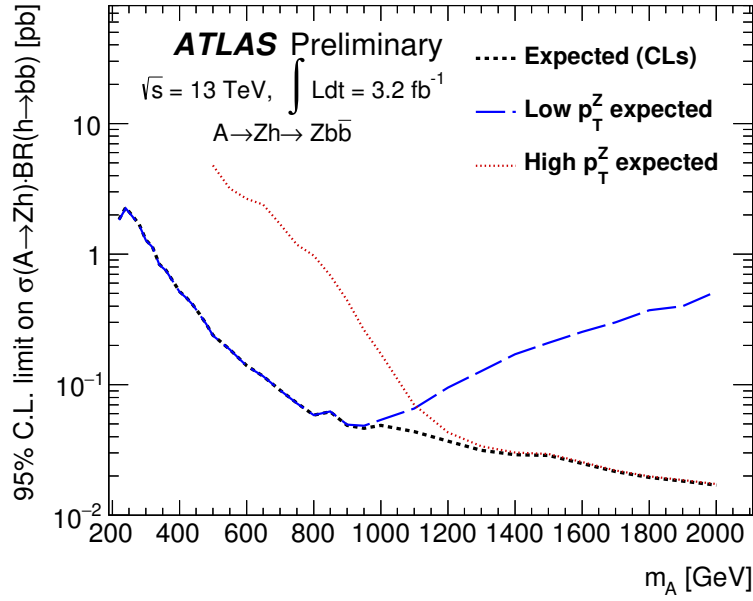
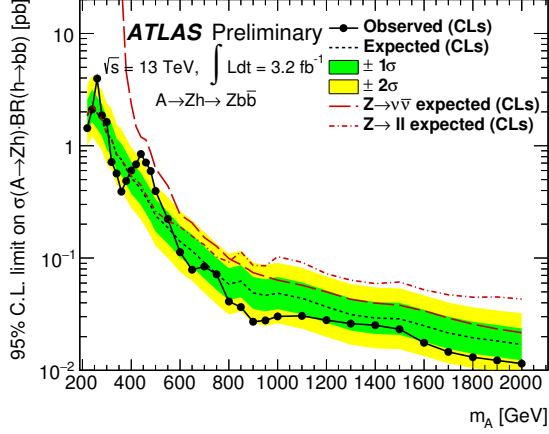
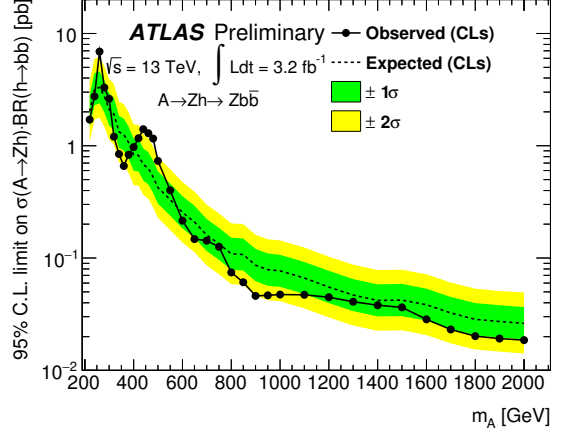


Figure 8.6: Expected upper limits at the 95% CL on $\sigma(gg \rightarrow A) \cdot \text{BR}(A \rightarrow Zh) \cdot \text{BR}(h \rightarrow b\bar{b})$ evaluated by combining the 0-lepton and 2-lepton channels. The long-dashed line is the upper limit expected from the low- p_T^Z analysis, while the fine-dotted line is that from the high- p_T^Z analysis. The remaining dotted line is the combination of the two analysis channels.

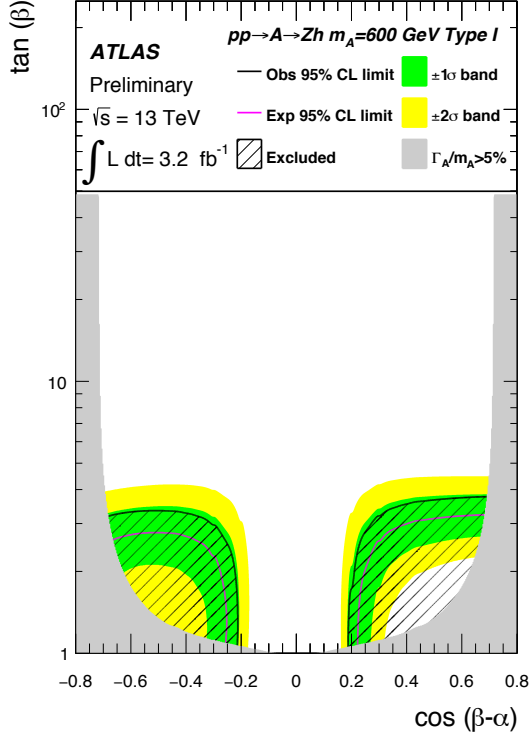


(a)

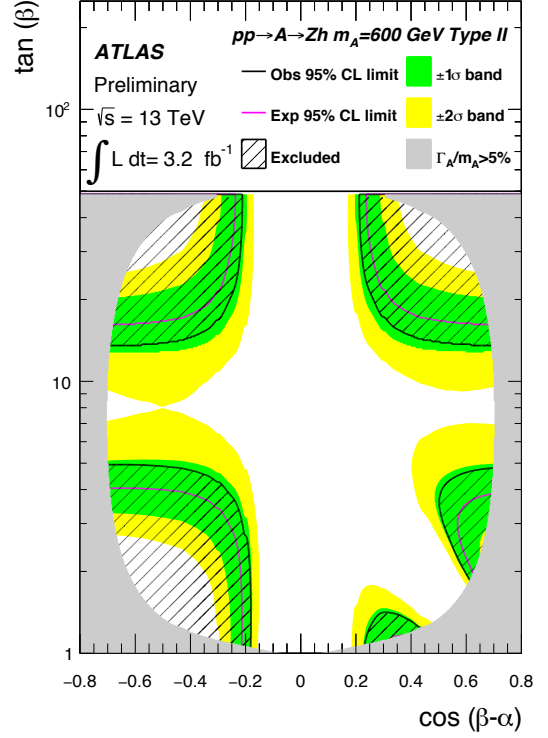


(b)

Figure 8.7: Upper limits at the 95% CL on the product of the production cross section for $pp \rightarrow A$ and the branching ratios for $A \rightarrow Zh$ and $h \rightarrow b\bar{b}$ evaluated by combining the 0-lepton and 2-lepton channels. The possible signal components of the data are interpreted using the following production mechanism hypotheses: Assuming (a) pure gluon-gluon fusion production, and (b) pure b -quark-associated production. The green (yellow) band indicates the 1σ (2σ) statistical and systematic uncertainty on the expected exclusion limit (dotted black line). The solid black line is the observed exclusion limit. The long-dashed (fine-dotted) red line indicates the limit for the 0-lepton (2-lepton) analysis alone. The small increase in the expected limit for $A \rightarrow Zh \rightarrow l^+l^-b\bar{b}$ with m_A around 850 GeV is due to statistical uncertainties on the background estimation (more detail in Appendix C), while for m_A around 1 TeV the effect is due to the transition from the low- to high- p_T^Z category (see Figure 8.6).



(a)



(b)

Figure 8.8: The interpretation of the cross section limits in the context of the 2HDM (a) Type-I and (b) Type-II models as a function of the parameters $\tan\beta$ and $\cos(\beta - \alpha)$ for $m_A = 600$ GeV. Variations of the natural width up to $\Gamma_A/m_A = 5\%$ and different mixtures of gluon-gluon fusion and b -quark-associated production are taken into account. Only points in parameter space where $\Gamma_A/m_A < 5\%$ are considered.

Chapter 9

Conclusions

Two analyses that take advantage of the large predicted branching ratio of the Higgs boson to b -quarks have been presented and my contributions highlighted. A discussion of the results follows the summary of both analyses given below.

9.1. Summary

9.1.1. SM VH

A search for the decay of a Standard Model Higgs boson into a $b\bar{b}$ pair when produced in association with a W or Z boson has been performed with the ATLAS experiment using data collected in proton-proton collisions from Run 2 of the LHC. This dataset corresponds to an integrated luminosity of 13.2 fb^{-1} collected at a center-of-mass energy of $\sqrt{s} = 13 \text{ TeV}$. For $m_H = 125 \text{ GeV}$ the measured signal strength with respect to the SM expectation is found to be $\mu = 0.21^{+0.36}_{-0.35}(\text{stat.}) \pm 0.36(\text{syst.})$. This corresponds to a significance of 0.42 standard deviations compared with an expectation of 1.94 and is consistent with previous results in this channel. In absence of observation (5σ) or evidence (3σ), the CL_S exclusion limits are calculated to determine if the SM Higgs boson can be excluded in this decay channel. The observed 95% CL upper limit is set on the ratio of the cross section times branching ratio with respect to the SM expectation for $pp \rightarrow (W/Z)(H \rightarrow b\bar{b})$ of 1.2, to be compared to the expected 95% CL upper limit in the absence of a signal of $1.0^{+0.4}_{-0.3}$. This expected upper limit does not exclude the SM expectation observed in data as the 1σ error bands are more than enough to explain the difference. The analysis procedure has been validated by measuring the yield of $(W/Z)Z$ with $Z \rightarrow b\bar{b}$, where the ratio of the observed yield to that expected in the SM is found to be $0.91 \pm 0.17(\text{stat.})^{+0.32}_{-0.27}(\text{syst.})$, corresponding to a significance of 3.0σ .

compared with an expectation of 3.2σ .

9.1.2. 2HDM $A \rightarrow Zh(h \rightarrow b\bar{b})$

A search for a heavy, CP-odd Higgs boson, A , decaying into a Z boson and a 125 GeV Higgs boson, h , is presented. The Z boson is reconstructed in the decay channels $Z \rightarrow e^+e^-$, $Z \rightarrow \mu^+\mu^-$, or $Z \rightarrow \nu\bar{\nu}$, and the h boson is reconstructed in the decay channel $h \rightarrow b\bar{b}$. The search has been performed using proton–proton collision data at a center-of-mass energy of 13 TeV corresponding to an integrated luminosity of 13.2fb^{-1} , collected by the ATLAS detector at the LHC. No significant excess beyond the SM expectation from background is observed in the data for all A -boson mass hypotheses. The data are used to determine 95% confidence level upper limits on the product of production cross sections and branching fractions, $\sigma(pp \rightarrow A) \cdot \text{BR}(A \rightarrow Zh) \cdot \text{BR}(h \rightarrow b\bar{b})$, in the range of $[4.0, 0.017]\text{pb}$ ($[6.9, 0.026]\text{pb}$) for $m_A = [220, 2000]\text{GeV}$ assuming gluon-gluon fusion (b -quark-associated) exclusive production. Constraints have also been presented in the context of the Type-I and Type-II 2HDMs using the $\tan\beta\text{-cos}(\beta - \alpha)$ parameterization.

9.2. Discussion & Outlook

The di-boson cross check performed in the SM VH analysis demonstrated that the analysis is robust and needs more data before any conclusions can be drawn. However, this check does indicate this analysis is capable of making a measurement. Future improvements include the potential to perform a 1-lepton kinematic fit that parallels the 2-lepton fit. Ultimately, however, more data is needed in order to make an observation of the $H \rightarrow b\bar{b}$ decay. There is a total of 100fb^{-1} of data expected by the end of Run 2 that will likely result in a 5σ expected significance. However, due to the downward fluctuations in a fraction of the data, this may not result in a 5σ observation in data. By comparison, CMS have so far produced no Run 2 results for this decay channel and the particle physics community is looking forward to their results.

The exclusion region plots in the BSM analysis (Figure 8.8) are not yet competitive with the Run 1 ATLAS results (Figure 2.17). This is mostly driven by the small amount of data as this is the most prominent uncertainty especially for large A -boson mass hypotheses. Additionally, by extending the selection to events with one or more additional b -tagged jets, the efficiency for the b -quark associated production would become almost 80% of those found for the gluon-gluon fusion process allowing also for a better separation of the two production modes. Future developments for this analysis should likely focus on improving the b -quark associated production signal acceptance as this will have the most impact on the 2-D 2HDM exclusion region plots. Lastly, two approximately 2σ excesses are observed in the ggF production mode at 260 GeV and 440 GeV as seen in Figure 8.7a. The inclusion of more data should resolve if these excesses result in the discovery of new boson or were just statistical fluctuations.

Both analyses will benefit greatly from more data as the limited size of the data samples is the single largest contribution to the uncertainties. More data will likely improve several experimental sources of uncertainty as well owing to the fact many (flavor-tagging, jets) of these uncertainties are data-driven. Areas of improvement for both analyses include reducing the statistical error on the MC simulation. This can be accomplished through either producing more MC simulation or devising more sophisticated techniques to preserve MC simulation events such as truth tagging. In terms of signal extraction, one could explore more advanced machine learning techniques that actually discover discriminating features themselves rather than relying on hand-crafted feature sets. This has shown some promise in an admittedly restricted context [148]. The possibility of using pseudo-continuous b -tagging as was done in Run 1 would likely help in isolating regions of maximal sensitivity for both analysis. Related to my expertise in non-parametric density estimation is the possibility of better approximating the underlying distributions with a simplified vine-copula model [149]. These sorts of robust density estimates can be used to estimate both a Bayesian discriminant and event selection efficiencies. Both analyses highlight the difficulties in suppressing the large contributions from the dominant V +jets and $t\bar{t}$ backgrounds.

Appendix A

b-Jet Trigger

During the LS1 upgrade period between Runs 1 and 2 several aspects of the ATLAS detector were improved in preparation for the increases in both center-of-mass energy (from 8 TeV to 13 TeV) and luminosity (from 50 ns to 25 ns bunch spacing). As noted on several occasions in this thesis, the ATLAS trigger decision must be made within CPU-time and memory constraints. Thus, increased performance of the HLT *b*-jet trigger was among these expected improvements during LS1. I was a key contributor to two projects that resulted in improved execution time of the *b*-jet trigger and increased identification rate of *b*-jets. Both of these projects as well as an overview of the HLT and *b*-jet trigger are outlined in this Appendix. The information from this Appendix is taken from Ref. [150] unless otherwise noted.

A.1. Software Architecture Overview

Before going to the details of what I contributed to the *b*-jet trigger group I will give more details about the software architecture of both Athena and the ATLAS trigger. This section builds on the information given in Chapter 4 regarding the GAUDI and Athena frameworks. Recall that GAUDI serves as the scaffolding upon which Athena constructs the ATLAS simulation and reconstruction frameworks. GAUDI realizes several software architectural patterns (Pipes and Filters, Blackboard), design patterns (Plug-in Components/Strategy, Finite-State Machine (FSM), Handle/Proxy), and domain patterns (Event Loop, Transient/Persistent Separation, Collections and Streams). For instance, the Pipes and Filters, Blackboard, Event Loop, and Transient/Persistent Separation aspects of GAUDI are diagramed in Figure 4.1. Here, algorithms runs in pre-defined sequences (Pipes) that can be terminated by a Filter. During the Event Loop, algorithms consume and produce

transient data that is shared via an event store (Blackboard). Athena implements several Plug-in Components such as `AthAlgorithm` (Event Loop building block, called once per event, adds data to the event), `AthAlgTool` (helps an Algorithm perform a function, typically used to implement a Strategy pattern), and `AthService` (provides a common service to multiple Athena components). A few examples of `AthService` components are the `StoreGateSvc` (a transient data store), the `IMessageSvc` (logging), and the `IAtRndmGenSvc` (random number management). The `StoreGateSvc` is of particular relevance to the trigger. This service manages memory for data objects (i.e. it owns them) and interacts with the Event Store to write/read data objects from disk/tape including persist-able references (`Data/ElementLinks`). Every `AthAlgorithm` interacts with the Event Loop through methods representing the various states in the GAUDI FSM. The constructor of an algorithm declares configurable (in Python) properties in addition to initializing data members needed while executing. The `initialize` method is called once per job and before the Event Loop. The `execute` method is called once per event. The `finalize` method is called once per job after exiting the Event Loop. An `AthAlgTool` has all the same FSM methods w.r.t. the `AthAlgorithm` except the `execute` method is replaced by any number of custom methods needed for the tool to accomplish its stated goal.

The HLT software framework is built atop the Athena framework [151]. It has a more elaborate FSM and restricted data flow as compared to the normal Athena framework. Algorithms in the HLT do not have access to the `StoreGateSvc` directly as output needs to be carefully curated since excessive IO can lead to unacceptable increases memory and CPU consumption. Instead, HLT algorithms operate on units of data storage called Trigger Elements (TE) that are passed in and out of algorithms in a standardized way. This throttling of the flow of data allows for the efficient operation of the HLT. There are four types of algorithms allowed in the HLT framework: feature extraction (FEX), hypothesis testing (HYPO), combo, and full scan. The most relevant for the b -jet trigger are the FEX and HYPO algorithms. FEX algorithms receive all necessary data through an input TE and attach newly created features (data) to an output TE that is passed to either other FEX

algorithms or to HYPO algorithms. HYPO algorithms makes a pass or fail decision based on data from an input TE.

A.2. Overview of HLT and the b -Jet Trigger Slice

An understanding of the Run 1 HLT (comprised of two parts, L2 and the event filter (EF)) is needed in order to understand the improvements to the HLT for Run 2. This along with the changes to the HLT for Run 2 are given below in the context of the b -jet trigger.

A.2.1. Run 1 and 2 HLT Overview

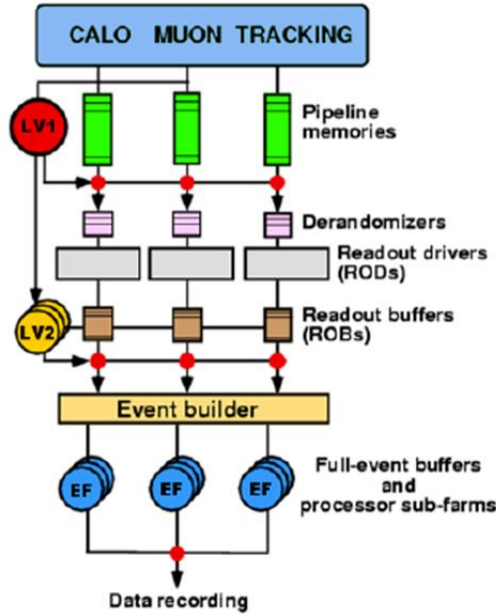


Figure A.1: Sketch of ATLAS Run 1 trigger levels [48].

The ATLAS trigger is an arrangement of hardware and software layers that act to reduce the rate of events written to disk. This is accomplished with the three logical layers of the trigger known as Level 1 (L1), Level 2 (L2), and Event Filter (EF) shown in Figure A.1. The Run 1 L1 is conceptually identical in Run 2 while L2 and the EF are merged into one layer simply known as the HLT. The output of this layer is a set of RoIs that contain the η , ϕ , and E_T information of specific regions within the detector from the calorimeter or muon

system and are used to seed the next levels of the trigger. Level two (L2) is a software-based trigger that used the RoIs generated from L1, but has access to the full detector readout within the RoIs. However, B -physics triggers had access to the entire detector at L2. The last level of the Run 1 ATLAS trigger is the EF. It is also a software-based system that applies offline-like reconstruction algorithms to the L2 output much like the HLT in Run 2 does when seeded with the L1 RoIs. The output from the EF is called “partial reconstruction” because incomplete and incorrectly calibrated information is used as input to the full reconstruction algorithms. The trigger works in successive steps with each level being seeded by the previous level. Each particle candidate is reconstructed in a sequence of algorithms (HYPO and FEX, described in Section 3.2.6) and that sequence is terminated as soon as it fails a HYPO algorithm [48].

A.2.2. b -Jet Trigger Overview

Much like offline flavor tagging, the general aim of the b -jet trigger is to maximize the b -jet tagging efficiency while maximizing the light- and c -jet rejection. Run 1 saw extensive use of b -jet triggers from a variety of physics groups and Run 2 brought a number of different upgrades related to the b -jet trigger in both hardware and software, such as the Insertable B-Layer (IBL) [152], L1 topological triggers (L1Topo), the Fast TracKer (FTK) [153], Super-RoIs (described later), and adoption of offline-based taggers (the last two I was personally involved with). Because the b -jet trigger relies so much on HLT jets and tracks, the shape of the RoIs and p_T threshold of the tracks play an important role as to how well b -jets are identified at the trigger level.

A.3. Split-Chain Performance Studies

A new approach for Run 2 to processing RoIs and reconstructing tracks and the PV is to merge the individual RoIs into a single RoI into what is known as a “Super-RoI.” The procedure is as follows:

1. Merge the L1 RoIs at the HLT level into a single Super-RoI

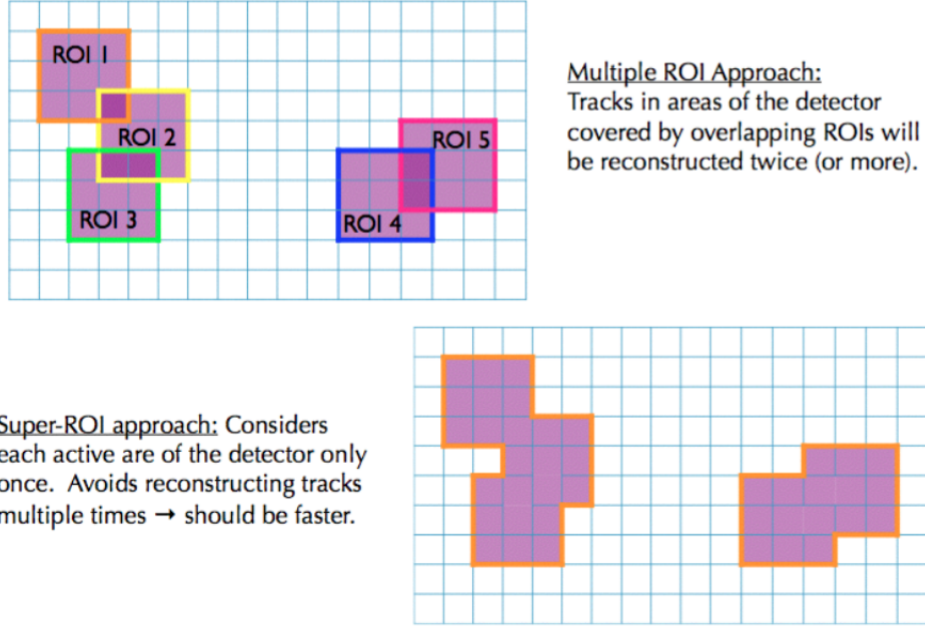


Figure A.2: Cartoon of how separate RoIs are merged to form a single Super-RoI. The grids in this figure represent η - ϕ space. The Super-RoI may not be topologically connected but still counts as just a single object.

2. Perform fast tracking and find a PV for the Super-RoI
3. Perform precision tracking on the individual RoIs, but constrain their tracks by the PV found in the previous step.

In high-pileup scenarios this will reduce the number of redundant track calculations and track duplication. A cartoon of what a Super-RoI looks like is given in Figure A.2.

This section evaluates the performance of the Event Filter Inner Detector (EFID) tracking and Event Filter b -jet feature extraction (EFBjetFex) algorithms¹. These algorithms are assessed in the context of a new Super-RoI approach where the individual RoIs are merged to reduce track overlap. The previous technique of using only those tracks found in an individual RoI will be referred to as the separate-RoI approach. The performance of the algorithms in the new Super-RoI versus the previous separate-RoI is studied by varying the p_T threshold of the tracks and RoI widths for the input RoIs (these are the RoIs that are

¹Despite the merging of the L2 and EF into the HLT in Run 2, these algorithms still retain the legacy names that refer to the EF.

merged into one Super-RoI) and recording the average time per event. The event times are obtained through standard ATLAS Athena monitoring services that create histograms of time per RoI for a variety of algorithms. The specific algorithms from which I evaluated the execution times in $t\bar{t}$ MC simulation versus pileup, track p_T threshold, and RoI width are:

- EFBjetFex (HLT b -jet feature extraction),
- TrigAmbiguitySolver (HLT precision tracking ambiguity resolution),
- TrigFastTrackFinder (fast HLT track finding that runs before precision tracking).

The average pileup values, track p_T thresholds, and RoI widths explored are:

- average pileup values: 0, 20, and 50 average interactions per bunch crossing,
- track p_T thresholds: 1-8 GeV in steps of 1 GeV,
- RoI widths: 0.4×0.4 , 0.6×0.6 , and 0.8×0.8 in $\Delta\eta \times \Delta\phi$.

The qualitative results are summarized as (specific numbers are not reliable since the computing farm used for testing is subject to random, non-isolation effects):

- EFBjetFex: no dependence on track p_T threshold or RoI width, separate RoI approach has increasing execution time with pileup while Super-RoI approach executes much faster with virtually no dependence on pileup
- TrigAmbiguitySolver: separate RoI approach execution time has an inverse relationship with track p_T threshold, proportional relationship to RoI width and pileup while Super-RoI approach has no relationship to any of these and has a negligible execution time
- TrigFastTrackFinder: separate RoI and Super-RoI approaches execution times have inverse relationships with track p_T threshold, proportional relationships to RoI width and pileup, but the Super-RoI approach has drastically smaller execution times.

Thus, the Super-RoI approach was adopted for use in Run 2 as several key algorithms improved their execution time.

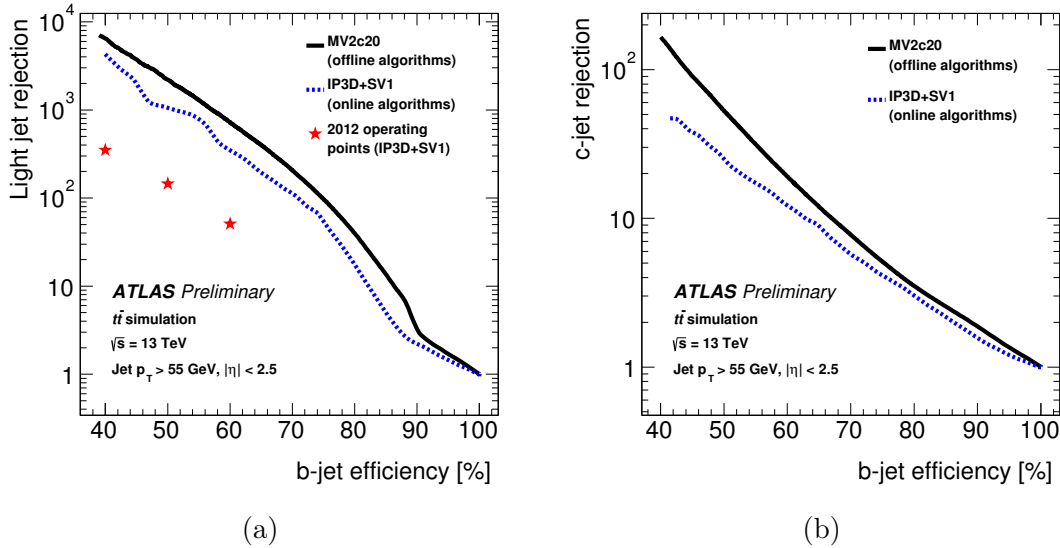


Figure A.3: The expected online performance in terms of (a) light-jet rejection and (b) c -jet rejection of the MV2c20 tagger (solid black line) is shown together with the expected performance of the IP3D+SV1 tagger in Run 2 (dashed blue line) and the actual performance of the IP3D+SV1 tagger that was achieved during Run 1 (red star) [154].

A.4. Migration of Offline Flavor-Tagging Algorithms

Another improvement in software comes from the adoption of b -tagging algorithms used in offline reconstruction. These algorithms are more robust and their use in the trigger allows for better correlations between the online and offline working points. The use of the same algorithms and codebase for online and offline flavor tagging also reduces redundancy which also reduces the chance for software bugs. However, because these algorithms were originally (and still primarily) designed for use in reconstruction, implementing them in the HLT environment was challenging. One of the central issues in the migration of the offline flavor tagging software is the fact that all input objects to an HLT algorithm are `const` in C++ parlance. The offline algorithms expect writable jet objects, which are not natively available in the HLT. Thus, our eventual solution was to create temporary jet objects and run the flavor tagging algorithms on these, copy the pieces of information necessary for the b -jet HYPO to make a decision, and properly delete the remaining information². This solution

²This seems obvious and simple, but implementing this in a codebase that never intended for the fracturing of its data structures led to weeks of crashes and debugging.

created another problem of its own: training of the offline algorithms requires that all data be perfectly preserved which isn't the case in the HLT. Thus, a short-term solution to rerun the flavor-tagging algorithms with the online configuration on HLT jets before training. A long-term solution is to build a special run mode for the b -jet slice so to preserve all the necessary data. The primary algorithm used for online b -tagging is the boosted-decision-tree-based discriminant called MV2c20 detailed in Chapter 4. As a reminder, MV2c20 refers to the second generation of multivariate discriminate (MV2 versus MV1 from Run 1) with 20% c -jet contamination to obtain better c -jet rejection while retaining excellent light-jet rejection. The improvements in rejection power achieved by adopting the offline algorithms are shown in Figure A.3. In summary, the adoption of new b -tagging algorithms results in a large improvement in tagging efficiency for similar levels of rejection versus Run 1.

Appendix B

Flavor Tagging Calibrations

Flavor-tagging scale factors (SFs) are ratios of efficiencies (ϵ) that are applied to simulated events in order to correct the tagging rate in simulation to what is measured in data. They are provided in bins of jet p_T and $|\eta|$ and have the form [92, 93, 155]

$$SF_{\text{Flavor}}(p_T, \eta) = \frac{\epsilon_{\text{Flavor}}^{\text{Data}}(p_T, \eta)}{\epsilon_{\text{Flavor}}^{\text{MC}}(p_T, \eta)}.$$

The scale factor associated with the jet depends on whether or not the jet is tagged by the MV2cXX algorithm. If the jet is tagged, the per-jet weight is simply defined as

$$w_{\text{jet}} = SF_{\text{Flavor}}(p_T, \eta),$$

while if it is not the SF in simulation is a necessary ingredient

$$w_{\text{jet}} = \frac{1 - \epsilon_{\text{Flavor}}^{\text{Data}}(p_T, \eta)}{1 - \epsilon_{\text{Flavor}}^{\text{MC}}(p_T, \eta)} = \frac{1 - SF_{\text{Flavor}}(p_T, \eta)\epsilon_{\text{Flavor}}^{\text{MC}}(p_T, \eta)}{1 - \epsilon_{\text{Flavor}}^{\text{MC}}(p_T, \eta)}.$$

The bins are chosen according to the available statistics in data, typically constrained by the large pre-scale factors applied at the trigger level or the lack of $t\bar{t}$ candidate events at high p_T . The calibrations derived with Run-I data used 6 bins in jet p_T for the b -jet scale factor, calibrated in the range between 20 and 300 GeV, 4 bins for the c -jet scale factor, calibrated in the same jet p_T range, and 6 bins for the light-jet scale factor, calibrated up to 750 GeV. The light-jet calibration is also provided as a function of two $|\eta|$ bins.

To avoid boundary effects and prevent distortions in the distributions of interest for the analyses relying on flavor tagging, the scale factors are smoothed. The method I developed for this smoothing is the local polynomial kernel estimation modified to account for the full

covariance matrix which is detailed in this appendix.

B.1. Local Polynomial Kernel Regression

Among the most common and crucial tasks in an analysis is fitting a curve with some functional form to a set of data. These curves are often used to interpolate or extrapolate into critical regions of phase space. Thus, the functional form of the fitted curve is of great importance. This form is usually chosen with a combination of domain specific knowledge (physical insight, etc.) and selecting from those forms with previous empirical success. However, sometimes the covariates and responses follow no obvious or previously encountered relationship. Additionally, one might not wish to impose a functional form on the data. In either case a “non-parametric” approach to constructing a smooth curve from data is attractive.

Broadly speaking, non-parametric curves have no innate functional form and can in principle adapt to any set of data. Splines, kernel estimation, and support vector regression are all forms of non-parametric regression. Each method will model the data differently and have its own unique “parameters” to determine the amount of “smoothness” of the curve. These are not parameters in the usual sense of fitting because one can normally find a value of the parameter(s) that can interpolate through all of the data. This is undesirable in this context as we are not seeking the “best” fit, but rather wish to follow the general trend of the data (otherwise, we could just use a histogram or empirical distribution function).

The non-parametric method used in the calibrations file and described here is known as local polynomial kernel estimation (LPKE). One can think of this method as fitting a polynomial by weighted least squares where the weights are functions of the covariates. These “weight” functions are referred to as kernels. Kernels are usually chosen to be even and integrate to unity, although for regression the requirement of integrability can be relaxed. A common kernel in physics and the one used here is a normalized gaussian kernel. A general, multivariate local polynomial kernel estimator of polynomial order p , covariate dimension d , with kernel $K(\mathbf{x})$, $d \times d$ bandwidth matrix $\mathbf{H}$, and constructed from n data points with

covariates $\mathbf{X}_i$ and responses Y_i has the form:

$$\hat{m}(\mathbf{x}; p, \mathbf{H}) = \hat{\mathbf{e}}_1^T \cdot (\mathbf{X}_x^T \mathbf{W}_x \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_x \mathbf{Y}$$

where,

$$\mathbf{W}_x = \begin{bmatrix} K_{\mathbf{H}}(\mathbf{X}_1 - \mathbf{x}) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & K_{\mathbf{H}}(\mathbf{X}_n - \mathbf{x}) \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \quad \mathbf{X}_x = \begin{bmatrix} 1 & (\vec{\Delta}_1^1)^T & \dots & (\vec{\Delta}_1^p)^T \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (\vec{\Delta}_n^1)^T & \dots & (\vec{\Delta}_n^p)^T \end{bmatrix}$$

$$K_{\mathbf{H}}(\mathbf{x}) = |\mathbf{H}|^{-1/2} K(\mathbf{H}^{-1/2} \mathbf{x})$$

$$\vec{\Delta}_i^m = \sum_{j=1}^d ((\mathbf{X}_i - \mathbf{x}) \cdot \hat{\mathbf{e}}_j)^m \hat{\mathbf{e}}_j$$

where the $\mathbf{e}$ are unit vectors. The derivation of this for the univariate case is straightforward in the context of least squares minimization [156]. In areas of phase space sufficiently far away from any data points, the weight matrix $\mathbf{W}_x$ may approach machine zero due to small kernel values. The convention followed here is that in the case of a zero weight matrix the kernel evaluations are just replaced by 1. Replacing the kernels this way reduces this method to the least-squares fit of a multidimensional “polynomial”. In this way, bandwidth values that are “very large” or “very small” yield similar results.

I also modified the weights matrix to allow for unequally weighted data by multiplying the kernel by $w_i = 1/\sigma_i^2$ since σ_i ’s are the variance (for uncorrelated errors) in the responses. This was motivated by wanting to approximate the best linear unbiased estimator (BLUE) where the weight matrix should be the inverse of the variance-covariance matrix [157]. Thus,

the weight matrix looks like:

$$\mathbf{W}_{\mathbf{x}} = \begin{bmatrix} w_1 K_{\mathbf{H}}(\mathbf{X}_1 - \mathbf{x}) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & w_n K_{\mathbf{H}}(\mathbf{X}_n - \mathbf{x}) \end{bmatrix}.$$

Lastly, for the initial implementation of the LPKE I used a simplified bandwidth matrix:

$$\mathbf{H} = \begin{bmatrix} h_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & h_d \end{bmatrix}.$$

Thus, the kernel here is a function of the “scaled” covariates:

$$K(\mathbf{H}^{-1/2}\mathbf{x}) \rightarrow K\left(\frac{x_1}{h_1}, \dots, \frac{x_d}{h_d}\right).$$

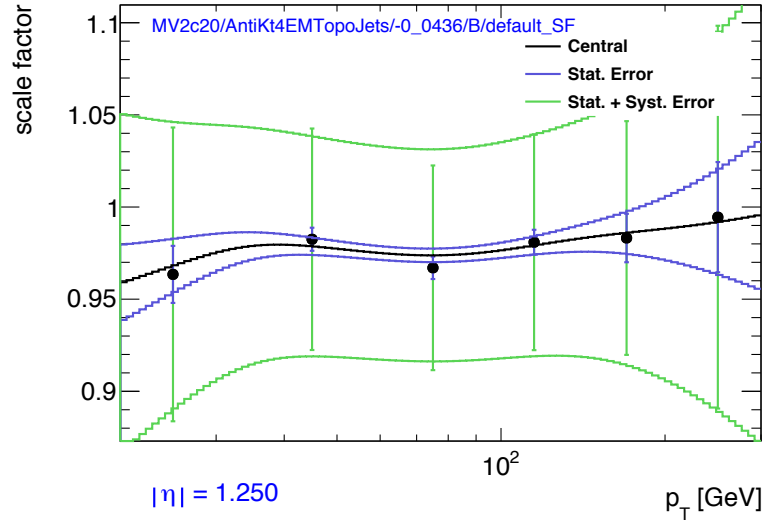


Figure B.1: The b -jet scale factor for the 70% working point smoothed from the calibrated 6 bins to 100 bins with the Gaussian kernel method.

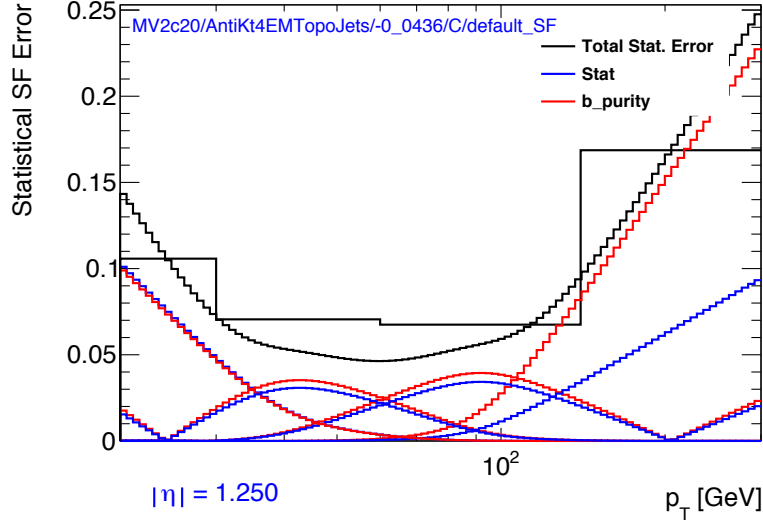


Figure B.2: Result of the bin-uncorrelated variational technique on the c -jet scale factor distribution. The blue curves are the results from the statistical bin uncertainties while the red are from the “ b -purity” systematic.

The $p = 1$ case of the estimator is called a local linear kernel estimator and has less bias at the ends of the data than a local constant kernel estimator ($p = 0$). Note that all odd orders have less bias at the end points than even powers [156]. The $p = 1$ local kernel estimator was chosen for this reason in addition to better accommodating the spacing in the SF distributions.

The bandwidth plays the role of the smoothing parameter in an LPKE. There are a number of methods for constraining the smoothing parameters, some of which are [156]: cross validation, minimization of squared residuals, or aesthetics. I implemented a cross-validation scheme that gave us a bandwidth of 0.4 (on a logarithmic scale in p_T). This is demonstrated in Fig. B.1.

The procedure for smoothing the scale factor distributions is as follows:

- Scan through entire CDI file to find the light-, tau-, c -, and b -jet CalibrationDataHistogramContainers objects. Specifically, select the ones labeled default_SF.
- For each flavor, add in quadrature all of the bin-uncorrelated systematics along with

the statistical bin uncertainties for the central SF distribution. These will serve as the weights given to the smoother when smoothing the central SF, denoted $\hat{m}_W$. For comparison with other smoothed curve later in the procedure, we also smooth the central distribution where all bins are weighted equally, denoted $\hat{m}_{NW}$.

- Bin-uncorrelated systematics and the statistical bin uncertainties are smoothed via a variational technique. In this approach, each bin i of the central distribution is varied by the content of the systematic or statistical uncertainty. This varied distribution is then smoothed with equally weighted bins and $\hat{m}_{NW}$ is subtracted away, leaving only the smoothed variation, $\hat{m}_{\delta_i}$. This is repeated for each bin of the systematic until there are as many smoothed, varied systematics as there are bins in the original SF distribution (Fig. B.2).
- Bin-correlated systematics are smoothed by varying the entire central SF distribution in all bins at once, smoothing, then subtracting away $\hat{m}_{NW}$.
- The up-variations are stored as the bin contents and the down-variations are stored as the bin errors for each uncertainty
- Extrapolation uncertainties are not smoothed. However, for technical reasons the “empty” region that overlaps with the other uncertainties has a bin structure identical to the smoothed bins.
- For each η bin, smooth only along the p_T axis scaled logarithmically using a $p = 1$ LPKE with a bandwidth of 0.4.

Lastly, note that the form of the estimator has the following property. Let $\hat{m}_Y$ be associated with responses Y . Now shift the Y by $\vec{\delta}$ so that $\hat{m}_{Y+\vec{\delta}}$ is the estimator associated with the “shifted” responses. Notice that $\hat{m}_{Y+\vec{\delta}} - \hat{m}_Y = \hat{m}_{\vec{\delta}}$. That is to say, the estimator of the “shifted” responses less the estimator for the original responses is simply the estimator associated with the shifts themselves, $\hat{m}_{\vec{\delta}}$. This is useful in the variational technique where the nominal distribution is shifted and normally one would need to perform two LPKEs and

a subtraction. This can introduce undesirable numerical effects. Fortunately, one can skip that procedure altogether and simply perform an LPKE on the $\vec{\delta}$'s themselves.

B.1.1. Addition of Full Error Matrix

As mentioned earlier, the BLUE can inform how to incorporate the kernel matrix with information about the covariance matrix ($\Sigma_{\mathbf{X}}$). Using the fact that the inverse covariance matrix ($\Sigma_{\mathbf{X}}^{-1}$) is used in place of the kernel matrix in the context of the BLUE, the kernel and covariance matrix can be simultaneously taken into account through:

$$\mathbf{W}_{\mathbf{x}} = \begin{bmatrix} (\Sigma_{\mathbf{X}}^{-1})_{11} K_{\mathbf{H}}(\mathbf{X}_1 - \mathbf{x}) & \dots & (\Sigma_{\mathbf{X}}^{-1})_{1n} \times \sqrt{K_{\mathbf{H}}(\mathbf{X}_1 - \mathbf{x}) K_{\mathbf{H}}(\mathbf{X}_n - \mathbf{x})} \\ \vdots & \ddots & \vdots \\ (\Sigma_{\mathbf{X}}^{-1})_{n1} \times \sqrt{K_{\mathbf{H}}(\mathbf{X}_n - \mathbf{x}) K_{\mathbf{H}}(\mathbf{X}_1 - \mathbf{x})} & \dots & (\Sigma_{\mathbf{X}}^{-1})_{nn} K_{\mathbf{H}}(\mathbf{X}_n - \mathbf{x}) \end{bmatrix}.$$

Thus, it is achieved that the full error matrix is taken into account when smoothing. This is the form of the kernel matrix used in the smoothing the SF distributions.

B.2. Reduction of Nuisance Parameters

Each scale factor distribution in a CDI data set has many (~ 40) systematic variations as well as statistical variations that need to be taken into account in order to properly access the uncertainty of the distribution. For a physics analysis this can lead to a large number of nuisance parameter when one needs to assess the effects of various systematics. Therefore, it is highly desirable to somehow reduce the number of systematics under consideration.

One method of reducing the number of systematics that preserves the bin-to-bin correlations and total error is to perform an eigenvalue decomposition on the covariance matrix of systematic and statistical variations, also known as a principal component analysis (PCA). It is clear that the resulting number of eigen-variations (EV) will simply be the number of bins in the scale factor distribution. This is already a sizable reduction of systematic variations to consider if your distribution has 10 bins. However, after smoothing the number of bins can

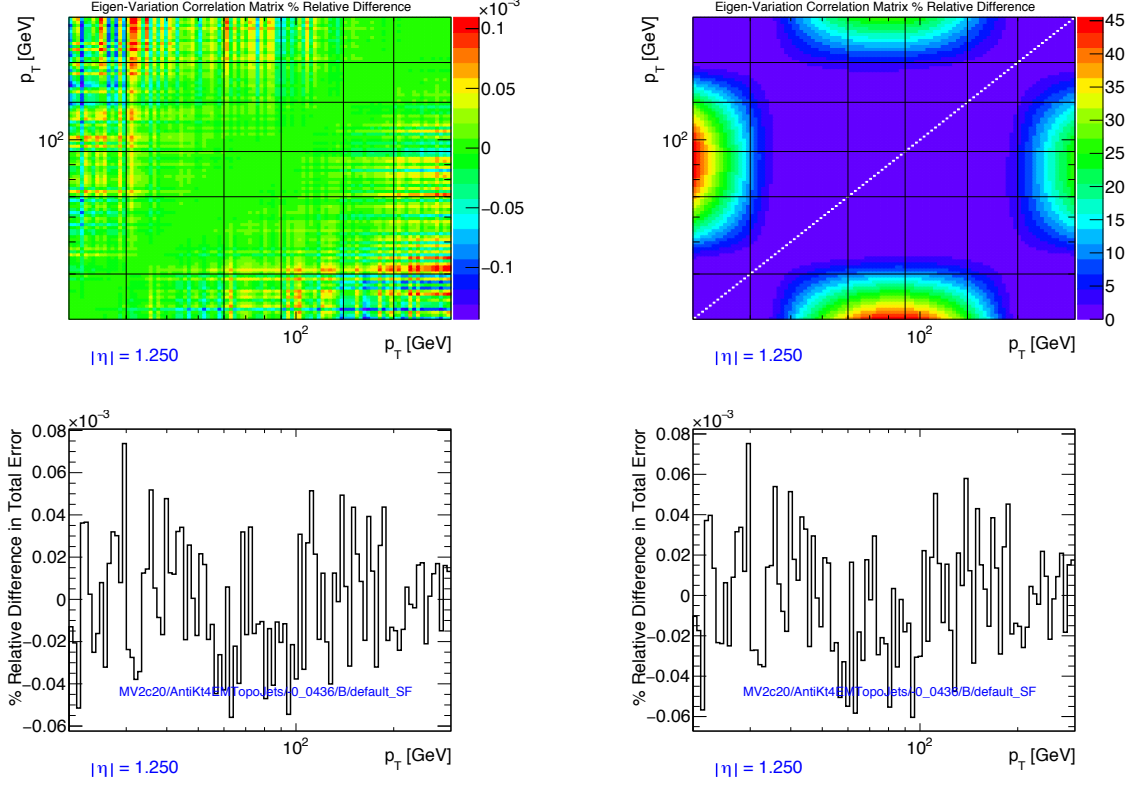


Figure B.3: % relative-difference in b-jet scale factor uncertainty correlations (top) and total errors (bottom) between the full set of statistical and systematic uncertainties (full set) and a reduced set of EV. **Left column:** Comparison of full set with complete set of EV. **Right column:** Comparison of full set with the S_Q , $M = 3$ reduced set of EV.

far exceed the original number of systematics. If the smoothing of the systematics is done in a careful way, most of the EV are very small in magnitude (numerically zero) and can be thrown out without fear of losing correlations or total error. In principle one should recover the original number of “significant” EV after this pruning (i.e. number of “pre-smoothed” bins). The remaining EV can be further reduced by summing up the M least significant EV (ranked by eigenvalue). Whatever summing technique is used, the overall goal is to try to satisfy the following constraint on the newly summed histogram, H' :

$$H'(x_i)H'(x_j) = \sum_{k=M}^{N_{EV}} H_k(x_i)H_k(x_j)$$

This condition may only be satisfied if one of two criteria are met:

1. The $H'(x_i)$ are vectors with elements $H_k(x_i)$
2. The $H'(x_i)$ are scalars but the $H_k(x_i)$ conspire in such a way that makes this equality hold

Unfortunately, neither of these criteria are applicable for our case. Thus, our essential problem is that of how one can store the information of a vector in a single scalar. Using the (false) equality above one can set $i = j$ and take the positive square root to yield:

$$H'(x_i) = \sqrt{\sum_{k=M}^{N_{EV}} H_k^2(x_i)}$$

Using this equation to sum the M least significant EV (S_Q) preserves the total errors in the resulting covariance matrix. However, because all the information about the sign of the original systematics is lost this method of merging severely distorts the correlations. This is demonstrated in Figure B.3 where the percent relative-difference in correlation matrices between the full set of systematics and the S_Q , $M = 3$ case of reduced EV set shows severe distortions of up to 45%.

B.2.1. Novel Summation

The loss in correlations when performing the quadrature sum is undesirable as analyses could be sensitive to this loss. Thus, I explored several methods of combining the eigenvariations in an effort to better preserve that correlations. Another manipulation of the (false) equality is to first sum over j , sum over i , then finally back substitute to solve for $H'(x_i)$:

$$H'(x_i)S_{H'} = \sum_j^{N_{\text{bins}}} \sum_{k=M}^{N_{EV}} H_k(x_i)H_k(x_j)$$

$$S_{H'}^2 = \sum_i^{N_{\text{bins}}} \sum_j^{N_{\text{bins}}} \sum_{k=M}^{N_{EV}} H_k(x_i)H_k(x_j).$$

After combining these two results I arrive at an alternative method for summing the least

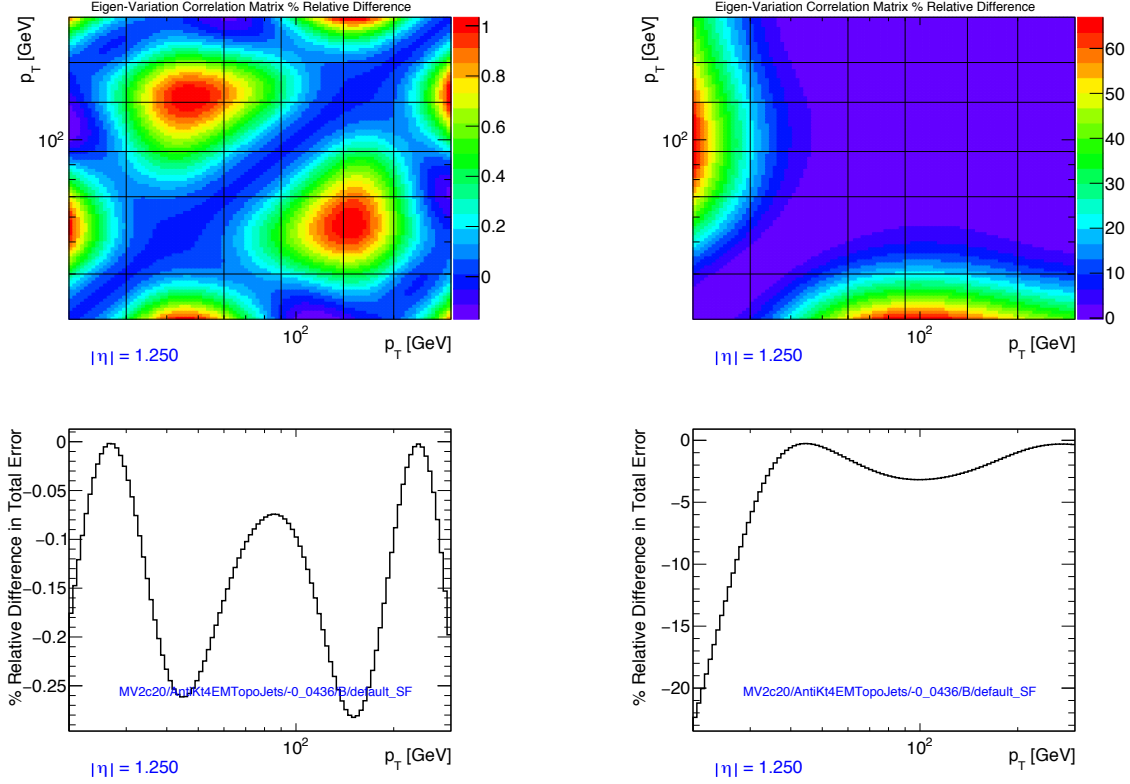


Figure B.4: % relative-difference in b-jet scale factor uncertainty correlations (top) and total errors (bottom) between the full set of statistical and systematic uncertainties (full set) and a reduced set of EV. **Left column:** Comparison of full set with the S_A , $M = 3$ reduced set of EV. Notice the better agreement in correlations but slightly worse agreement in total error versus the S_Q method. **Right column:** Comparison of full set with the S_A , $M = 4$ reduced set of EV.

eigen-variations

$$H'(x_i) = \frac{\sum_j^{N_{\text{bins}}} \sum_{k=M}^{N_{\text{EV}}} H_k(x_i) H_k(x_j)}{\sqrt{\sum_n^{N_{\text{bins}}} \sum_m^{N_{\text{bins}}} \sum_{k=M}^{N_{\text{EV}}} H_k(x_n) H_k(x_m)}}.$$

In principle this should be equal to the first method of merging, S_Q . But, because the original equation is false they give different results. This alternative method of summing (S_A) the M lowest EVs is found empirically to better preserve the correlations over the S_Q method (Figure B.4). However, S_A doesn't preserve the total error as S_Q does. Thus, a trade off must be made as to how much total error can be sacrificed for extra gains in correlations.

As a final note on EV reduction, another way of attempting to satisfy the (false) equality is to perform a multidimensional minimization against some metric (i.e. minimizing correlation

differences) treating each bin of the merged histogram (H') as the variables to minimize. An implementation of this using the Amoeba routine [158] was explored, but yielded results similar to using S_A . Thus, it is not currently being used. Also, the jet systematics are reduced using a slightly different method. These systematics are reduced by isolating kinematic regions where certain systematics dominate and merging (by quadrature sum) the least dominant systematics in those regions [159]. This method suffers from highly distorted correlations, but only in isolated regions.

B.3. Generalization to PCA

For a D -dimensional dataset with a covariance matrix Σ that has a set of eigenvalues $\{\lambda_1 \dots \lambda_D\}$ in decreasing order with their corresponding D -dimensional eigenvectors as $\{v_1 \dots v_D\}$, let $\{\mathbf{\Lambda}_1 \dots \mathbf{\Lambda}_D\}$ be the set of eigenvectors multiplied by their corresponding eigenvalues. These are referred to as the principle components of the dataset. The generalization of my result presented in the last section to combine the last M principle components into a single component $\mathbf{\Lambda}_{D-M+1}^A$ would be

$$\Lambda_{D-M+1,i}^A = \frac{\sum_{j=1}^D \sum_{k=D-M+1}^D \Lambda_{k,i} \Lambda_{k,j}}{\sqrt{\sum_{n=1}^D \sum_{m=1}^D \sum_{k=D-M+1}^D \Lambda_{k,n} \Lambda_{k,m}}},$$

where $\Lambda_{i,j}$ represents the j^{th} component of the i^{th} principle component vector $\mathbf{\Lambda}_i$.

Appendix C

Kernel Density Estimation with Systematic Variations

During the course of deriving the limit curves in the $A \rightarrow Zh$ analysis, an artificial-looking “bump” was seen in the two-lepton limit curve around $m_A = 850$ GeV as shown in Figure 8.7. This bump was found to originate from the Z +heavy flavor background component in the sparsely populated, high-invariant-mass tail (Figure C.1). This is suspected to arise from a statistical fluctuation and not an underlying physics effect¹. In order to avoid such a situation in the future, a method to “smooth”² this background component was sought. Additionally, every background component has several systematic variations that must be accurately estimated as well. The material shown in this chapter is my work on how to more accurately estimate the densities of the background distributions and systematic variations thereof.

C.1. Motivation for Density Estimation

A common task during all analyses is the estimation of frequency distributions from MC simulations (i.e. for a given process what is the frequency that an object of interest (jet, lepton, etc.) has a particular value of a quantity of interest (p_T , η , etc.)). A widely used, computationally efficient, and non-parametric way of creating such a distribution is to store the value of the quantities of the objects of interest in histograms. A histogram can naturally be expressed in terms of empirical cumulative distribution [160] ($\hat{F}(x)$)³ and is ubiquitous in expressing frequencies of processes. An example of a histogram of 1000

¹Indeed, this bump went away when more MC simulation samples were added after publication.

²I personally feel that the word “smoothing” is used in a somewhat pejorative sense here. Really, what we needed was a more accurate estimate of the distribution. The key difference here is that we need a better estimate of the true density and smoothness is just a happy byproduct, not the goal.

³The empirical cumulative distribution is $\hat{F}(x) = \frac{\#\{x_i \leq x\}}{n}$. Histogram estimates can be expressed in terms of $\hat{F}(x)$ through $\hat{f}(x) = \frac{\hat{F}(b_{j+1}) - \hat{F}(b_j)}{h}$, where h is the bin width of the histogram and $b_j(b_{j+1})$ is the low(high) edge of the bin containing x .

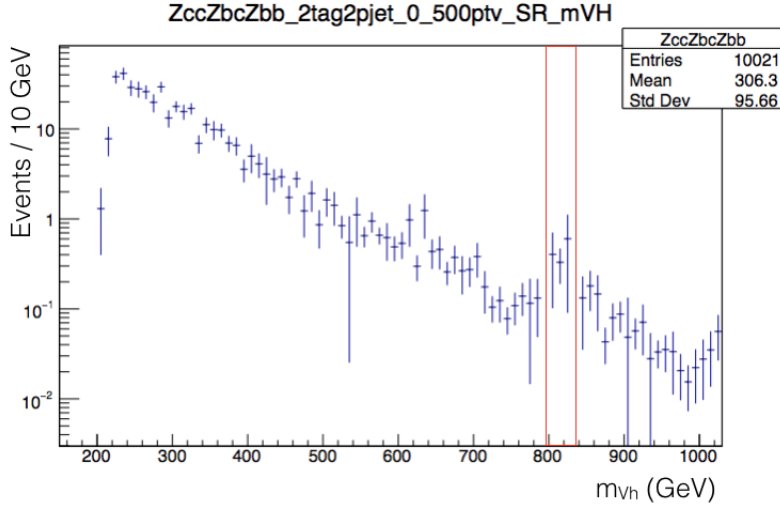


Figure C.1: Z +HF background component in the resolved, 2-tag region. The anomalous bump around 850 GeV is highlighted in red.

samples taken from a normal distribution is shown in Figure C.2. However, a histogram has several issues. The shape of a histogram can change drastically with a different choices of bin width (Figure C.3) and anchor position (the low edge of the first bin). It's often desirable to convert a histogram to a probability density function. Using a histogram in this way can be problematic if any sort of smoothness is required (i.e. differentiability) as histograms contain none of this information (either zero or infinite derivatives). Although this can be somewhat alleviated with frequency polygons [160], these also suffer from distorted shapes due to different choices of bin width and anchor position and have undefined derivatives at certain points. Additionally, statistical fluctuations, small numbers of entries in tails of distributions, and poor choices in bin width can lead to histograms with unrealistic unsmooth behavior (depending on the domain). A different approach to histogramming data that addresses some of these issues is to use data to construct a kernel density estimate (KDE).

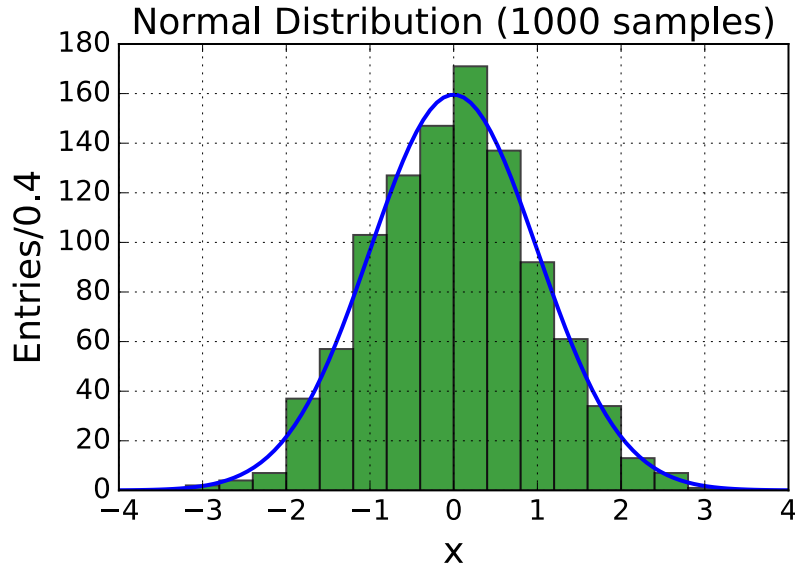


Figure C.2: 20-bin histogram of 1000 samples of normally distributed data. A normal distribution scaled to the same area as the histogram is given for comparison.

Double Unit Gaussian Distribution (1000 total samples)

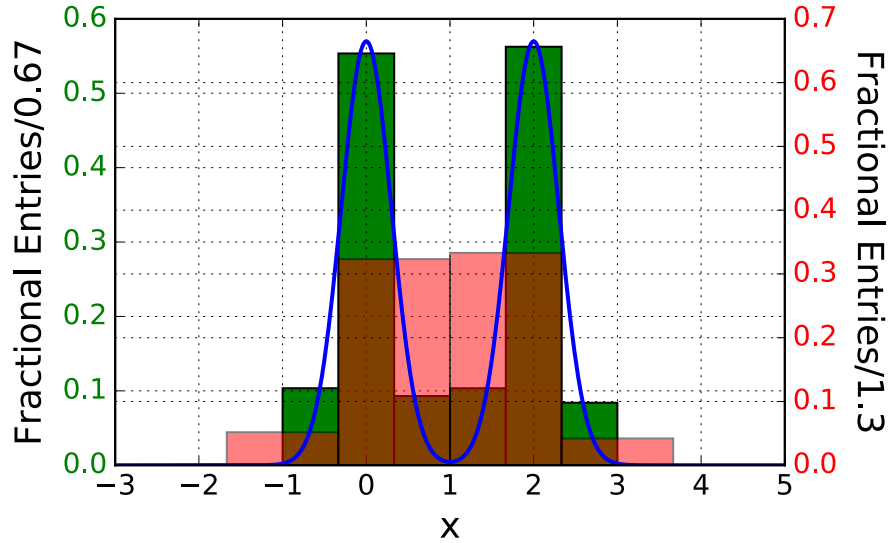


Figure C.3: Demonstration of histograms of different bin widths for data sampled from a distribution made from two unit Gaussian. The red histogram uses bins of length 1.3 while the green uses bins of length 0.67 (i.e. half the length of the red histogram). Note that the shapes for the two histograms are substantially different. The red histogram misses the bimodal nature of the underlying distribution whereas a histogram with half the bin width accurately captures this feature.

C.2. Description of Projection-KDE

The kernel-based formulation of density estimation is also expressible in terms of $\hat{F}(x)$ [160]:

$$\hat{f}(x) = \frac{\hat{F}(x+h) - \hat{F}(x-h)}{2nh} \rightarrow \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right)$$

where,

$$K(u) = \begin{cases} \frac{1}{2} & \text{if } -1 < u \leq 1 \\ 0 & \text{otherwise} \end{cases}.$$

In general, the kernel ($K(u)$) can be any function that is non-negative, integrates to one, and has a mean of zero (i.e. unit Gaussian). Here, h is known as the bandwidth rather than the bin width as in the context of histograms. The bandwidth plays extremely important role in the resulting shape of the estimate. There are several methods for choosing a particular bandwidth value. However, one must be careful to not blindly use a method like maximum likelihood as, of course, this estimator will achieve the best score with such a metric at bandwidths small enough to resolve every individual data point (i.e. a very un-smooth estimate that likely doesn't capture the true nature of the underlying distribution). An example of a KDE with a bandwidth optimized by cross validation is given in Figure C.4. Also, it is convenient to express the kernel argument in normalized coordinates $u_i = (x - x_i)/h$. It is self evident that if $K(u)$ obeys the constraints as listed, then $\hat{f}(x)$ is a proper density function normalized on the same interval (accounting for displacements of the kernel centers) as $K(u)$.

As histograms are typically what one encounters in practice, a way to translate a histogram to a KDE is desirable. A simple recasting of a histogram to weighted data where the bin centers of the histogram act as the data points and the bin contents act as the weights will be used as the basis for conceptually translating between histograms to data points.

Gaussian Kernel Density Estimation (1000 total samples, bw=0.1)

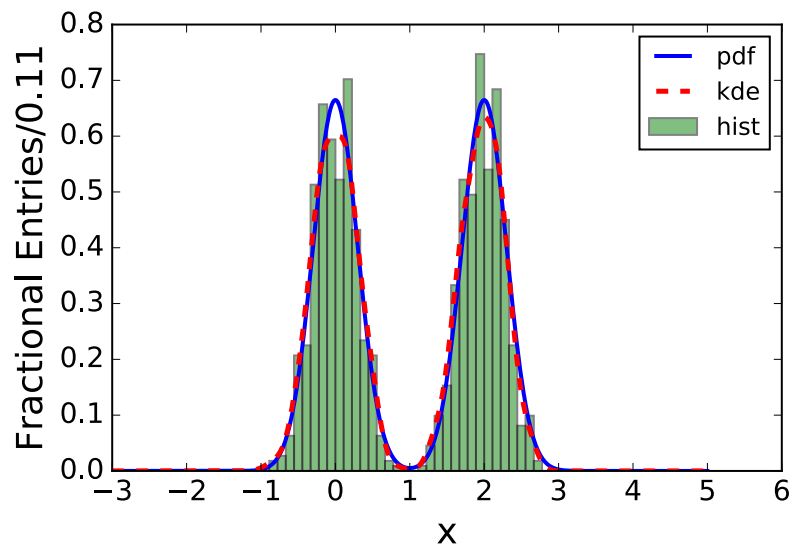


Figure C.4: Demonstration of KDE with a bandwidth (indicated in the title) optimized through the holdout, cross-validation method (i.e. half the sample is reserved for training the KDE while the other half is used for testing its performance w.r.t. a maximum likelihood metric). A histogram whose bin width (indicated in the y -axis label) is optimized in a similar manner is given for reference. Although both the bin width and bandwidth are similar in this case, this is a coincidence and not always the case.

Weighted Gaussian Kernel Density Estimation (1000 total samples, bw=0.15)

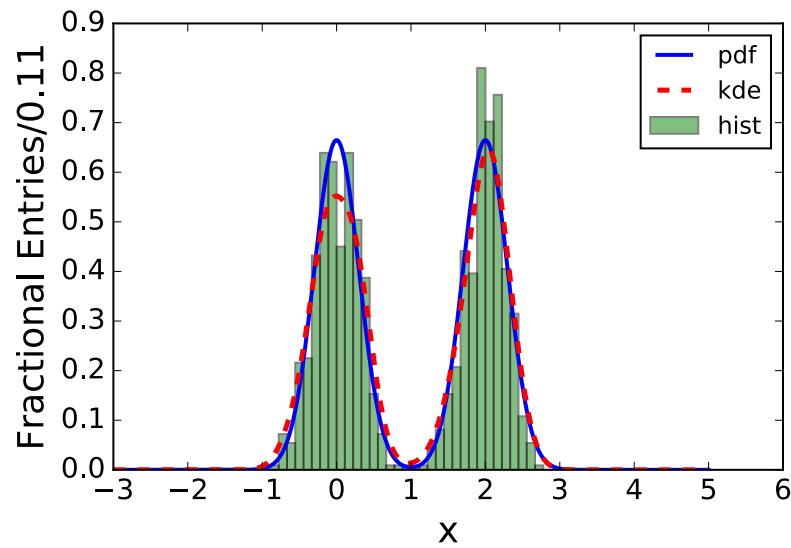


Figure C.5: Demonstration of weighted KDE derived from the histogram (green) with a bandwidth optimized in similar fashion as in Figure C.4. However, optimizing the weighted version of the KDE necessarily introduces biases due to only ever comparing performance relative to the source histogram (i.e. anomalous “spikes” in the histogram pull the weighted KDE in that direction). Additionally, since the number of data points is drastically reduced by the using the histogram’s bin centers, the holdout method uses 2/3 of the available data for training rather than half as in Figure C.4.

The weighted KDE now looks like,

$$\hat{f}(x) = \frac{1}{h} \sum_{i=1}^{N_{\text{bins}}} w_i K(u_i).$$

In this form, $\hat{f}(x)$ is now normalized to the sum of weights, $W = \sum_i w_i$. This is shown in Figure C.5.

The KDE can be further modified such that the bandwidth can take into account the local variance of the data (i.e. that the tails are more sparsely populated than the bulk). This is known as an adaptive bandwidth KDE and comes in two different forms. The first form is known as a balloon estimator where the bandwidth is now a function of the covariate(s),

$$\hat{f}(x) = \frac{1}{h(x)} \sum_{i=1}^{N_{\text{bins}}} w_i K(u_i).$$

This allows for precise control over the behavior of the values of the bandwidth, but is difficult to generalize. Most importantly is that a balloon estimator is no longer a density as the kernels now belie extra hidden dependence on the covariates that can cause them to no longer have a unit integral. Another adaptive bandwidth approach and one that preserves the unit integrals of the kernel is known as a point-wise estimator. A point-wise estimator allows the bandwidth to vary depending on the source data point, u_i ,

$$\hat{f}(x) = \sum_{i=1}^{N_{\text{bins}}} \frac{w_i}{h_i} K(u_i).$$

This allows for the bandwidth to spatially vary without being an explicit function of the covariates and is the method I adopted for estimating the distributions of background and systematic variations. The specific form of the point-wise bandwidth I developed is heavily inspired by that used in the RooFit `RooNDKeysPDF` class [138, 161]. However, I added factors to aid in smoothing the tails of the distributions as my initial attempt to use `RooNDKeysPDF`

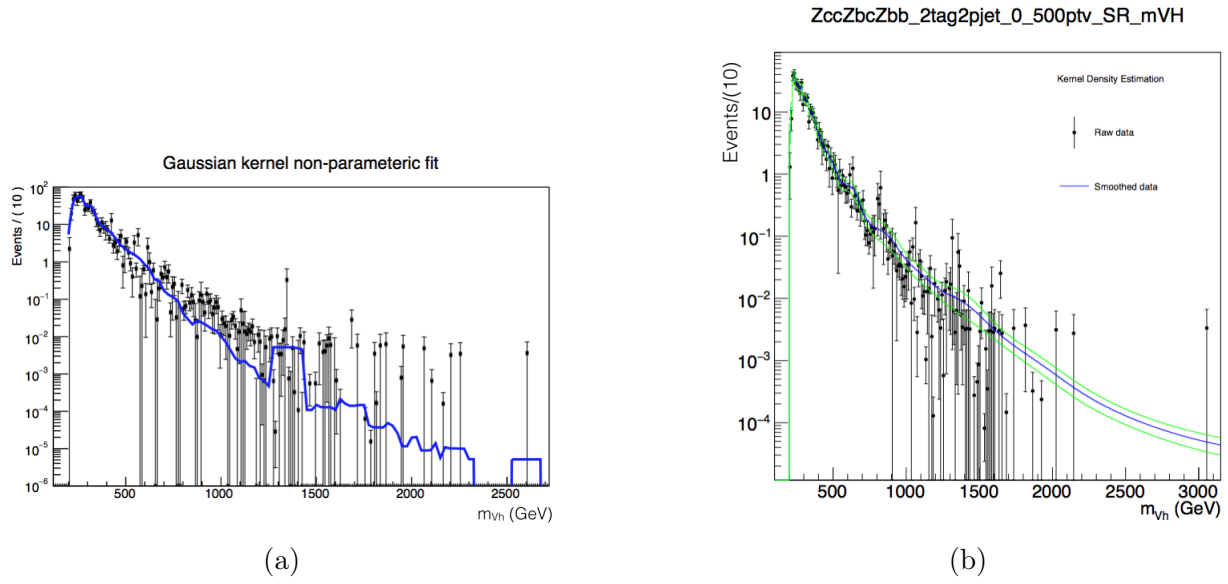


Figure C.6: Performance comparison of the (a) RooFit KDE versus (b) custom built KDE for the distribution shown in Figure C.1. Notice the drastically different behavior in the tails ($m_{Vh} \geq 750$ GeV). The green error bands in (b) are discussed in the text.

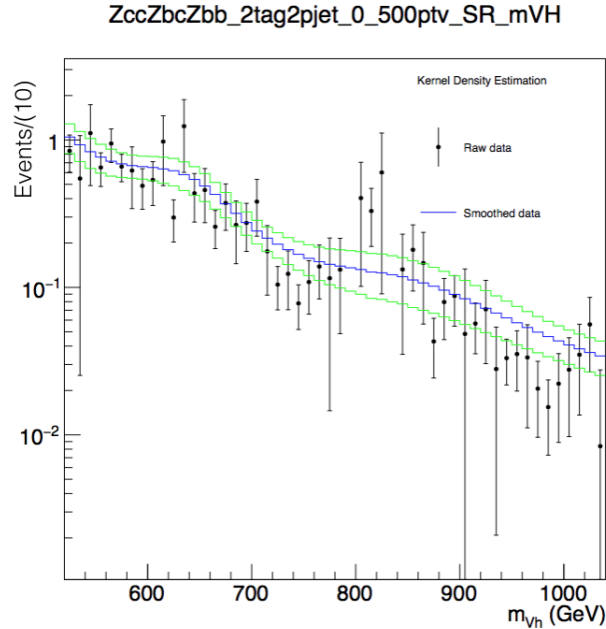


Figure C.7: Plot showing the smooth estimate and un-smooth data near the 850 GeV region in the Z +HF background component shown in Figure C.6b. Notice the blue KDE eliminates the “bump” around 850 GeV.

failed in the tail regions as shown in Figure C.6a. Thus, I arrived at

$$h_i = \text{KNN}_i \times \left(\frac{4}{3 \times \text{EF} \times N} \right)^{(1/5)} \sqrt{\frac{\sigma_i}{N}} \Delta,$$

where Δ is a characteristic length, KNN_i is a data-adaptive, nearest-neighbors factor, EF is an overall scale factor applied uniformly to all bandwidths, σ_i is either the local or total sample (square root of the) variance⁴, and N is the sum of weights. Prior to computing the KDE, the data are linearly binned as described in Ref. [156] in order to increase computational performance while introducing less bias than traditional binning (i.e. histogramming). This binning procedure uses a regularly-spaced grid in which the spacing between grid points is Δ . Thus, the factor of Δ in the bandwidth formula acts as a way to encode all of the information of the scale of the coordinates and leaves the other factors (including σ_i) scale-free. In order to smooth the low-density tails more aggressively while preserving the information in the high-density bulk regions, a unique “KNN” factor modulates each bandwidth’s magnitude. This factor is calculated for each data point as the radial distance (in units of Δ) from point i that encompasses a user-defined fraction (denoted f_{KNN}) of the event weights. In low density regions this radial distance tends to be large and increases the bandwidth while in high-density regions this radial distance is usually close to unity and thereby leaves the bandwidth unchanged. The EF factor scales the sum of weights and acts as an overall scale factor for all bandwidths (albeit to the inverse fifth root). The results of this estimation algorithm are given in Figures C.6b and C.7.

Often, each bin in a histogram will have an associated error. This is translated to the KDE through first order error propagation w.r.t. the weights. Since a KDE is a linear transform of the weights, error propagation is straightforward (errors on the weights are assumed uncorrelated)⁵. If $\mathbf{w}$ is the vector of weights and $\mathbf{K}(x)$ is the vector of kernel evaluations (both of length N_{bins}), then $\hat{f}(x) = \mathbf{w} \cdot \mathbf{K}(x)$ (i.e. the scalar product between $\mathbf{w}$

⁴Whether the local or total sample variance is used depends on criteria given in Ref. [161].

⁵Taking into account errors on the covariates is a much harder problem in density estimation and will be ignored here.

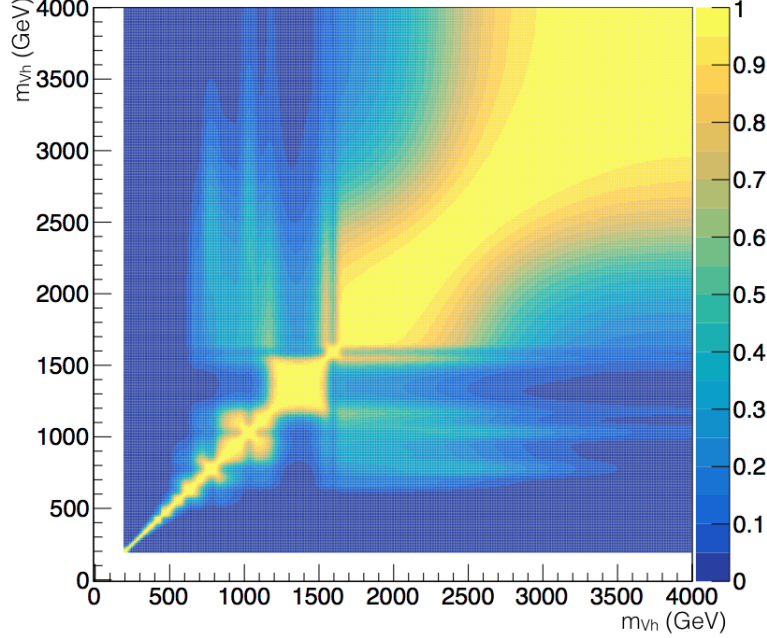


Figure C.8: Bin-to-bin correlation matrix for the Z +HF KDE in the Z +HF background component shown in Figure C.6b. The strange structures are due to negative bin weights.

and $\mathbf{K}(x)$). Thus, first order error propagation is exact and takes the form

$$\sigma_f^2 = \mathbf{K}^\top \Sigma^{\mathbf{w}} \mathbf{K} \quad (\text{C.1})$$

where $\Sigma^{\mathbf{w}}$ is the $N_{\text{bins}} \times N_{\text{bins}}$ covariance matrix of bin errors. Since the bin errors are assumed uncorrelated, this reduces to

$$\sigma_f^2 = (\delta \mathbf{w}_2) \cdot \mathbf{K}_2,$$

where $\delta \mathbf{w}$ is the vector of bin errors and the subscript represents the power to which each element of the vector is raised. Although the original bin errors were independent from each other, after the estimation procedure the propagated errors are now correlated. An example correlation matrix of the KDE bin errors is given in Figure C.8.

Everything discussed thus far naturally extends to multiple dimensions. For the adaptive KDE just described, my implementation uses the 1-D projections (empirical marginal distributions) along each axis to compute the σ_i and KNN_i quantities. The motivation for using

the projections is given in the next section. Lastly, there are several practical issues that need to be addressed such as memory and CPU consumption as well as sparsity issues as the number of dimensions grows (curse of dimensionality). Due to these practical issues, only dimensions of one or two are considered in the following discussion except where explicitly stated.

C.3. Estimating Systematic Variations

A straightforward solution (and the one I eventually adopted) to the estimation of the systematic variations of the backgrounds is to simply apply the same 1-D KDE technique to the variations as to the nominal. This approach is computationally fast and efficient. However, any correlations between the nominal distribution and systematic variation are ignored when the estimation is performed independently in this way. A proposal suggested by my colleagues is to first construct 2-D “migration” matrices that have the nominal value along one axis and a systematic variation along another axis (Figure C.9a). For each migration matrix (one for each systematic) one should construct a 2-D KDE (Figure C.9b) and project along each axis to get a smooth estimate for the nominal distribution and systematic variation (Figure C.10). This method for estimation accounts for the correlations among the nominal and each individual systematic. Additionally, it provides more smooth behavior when looking at the so-called “deltas” (systematic variation less the nominal) than estimating the 1-D nominal distribution and systematic variation separately as shown in Figure C.11.

However, there are two conceptual and one practical issues with this with this approach. The practical issue is that the space (memory) and time (CPU consumption) requirements for storing several 2-D histograms and smoothing each of them individually with the method presented in this chapter is currently impractical⁶. The two conceptual issues are even more troubling. The first is that this approach only accounts for correlations among a systematic and nominal distribution and not amongst the systematics themselves. This likely isn’t an issue between unrelated systematics like, say, the jet and muon systematics. However, it

⁶Even the relatively small number of 1000 bins along each axis in 2-D requires the storage of 1000000 floating point numbers. At, say, four bytes per float, this is minimally 4 MB of storage per systematic.

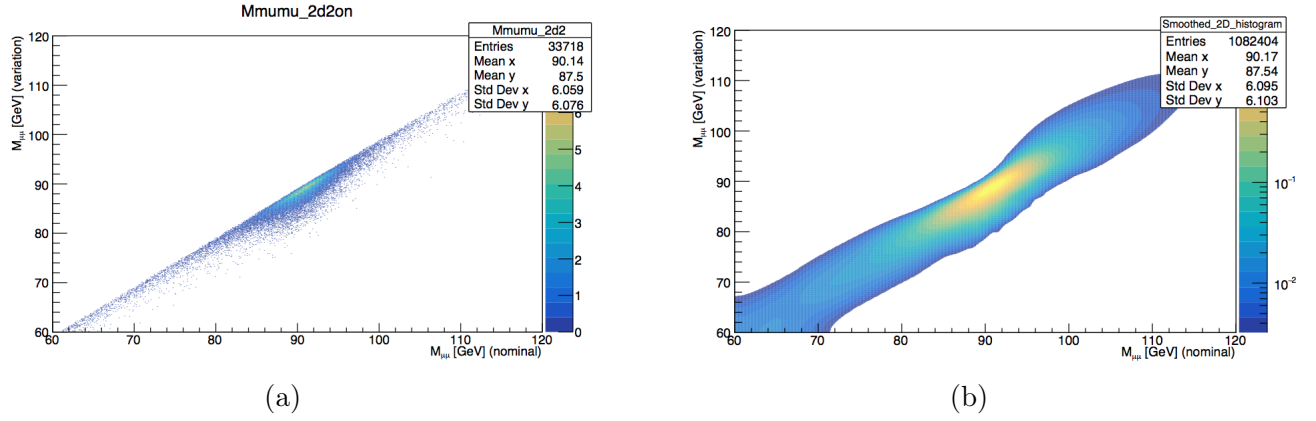


Figure C.9: (a) The 2-D migration matrix for a toy systematic variation on the di-muon invariant mass where the nominal is along the horizontal axis and the systematic variation is along the vertical axis. The systematic effectively biases the peak away from the Z -boson mass and slightly smears the resolution. (b) The smooth KDE of the same migration matrix. Notice that the means along each axis stay virtually the same while the resolution is slightly worse (as is expected when using a Gaussian kernel).

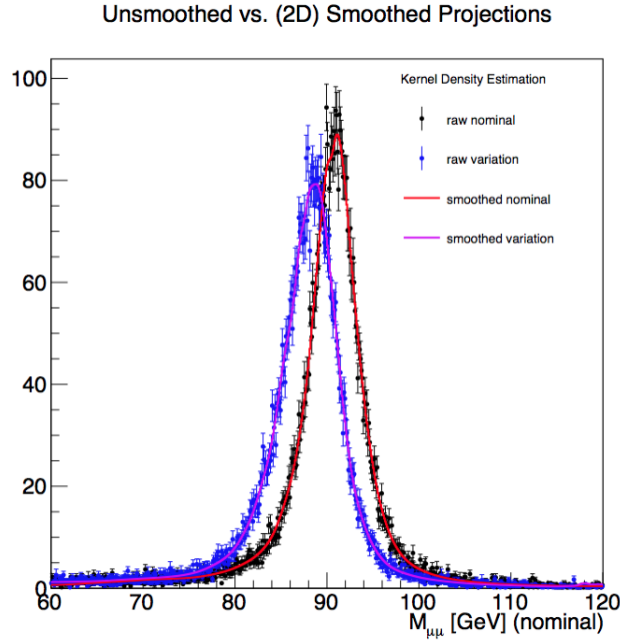


Figure C.10: 1-D projections of the nominal distribution and systematic variation of the 2-D migration matrix before (data points with error bars) and after the KDE smoothing (solid line (error envelope omitted)).

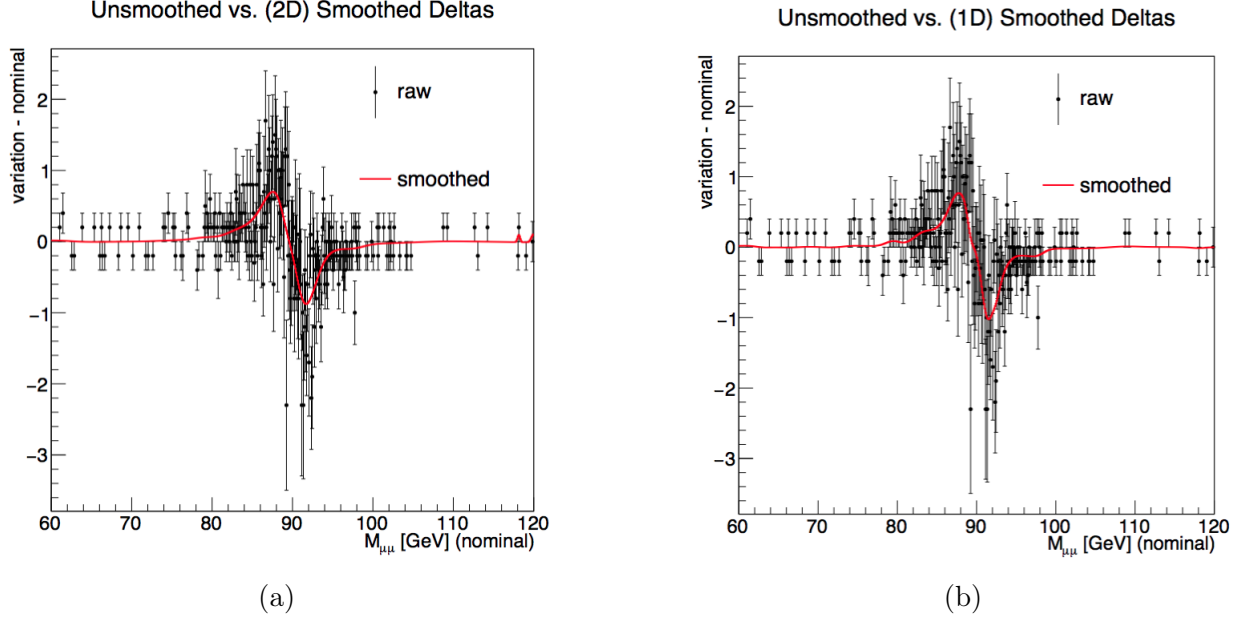


Figure C.11: 1-D projections of systematic Δ s for the (a) 2-D estimation of the migration matrix versus the (b) estimation of the 1-D projections. Notice that in (a) the Δ s have smoother behavior than in (b). This is due in part to the 2-D smoothing taking into account the correlations between the nominal and systematic variation.

likely affects the jet systematic amongst themselves. Secondly and most of all, there are now several different “nominal” distributions, one for each systematic. How to select the “correct” nominal is somewhat arbitrary in this situation.

A proposal for future work presented here would solve all of these issues. The conceptual basis for this approach is to construct an $(N_{\text{sys}} + 1)$ -dimensional histogram of the nominal and all systematics, smooth this taking into all correlations amongst the entire set of systematics and nominal, then projecting along each axis to get the all individual systematic variations and a single, nominal distribution. As always, this is far easier said than done. However, this is a solvable problem as I’ll outline here. The practical issues with this are:

1. There is currently no widely available n -D histogram
2. The space requirements for such a data structure are be unfeasible
3. Smoothing in n -D would take just as much space as the n -D histogram and the CPU time to compute this would be impractical

4. Sparsity issues in n -D are a major concern

Issue 1 is actually not true. ROOT has an efficient class for sparse, n -D histograms called `THnSparse` that can also perform projections along each axis. But, are “sparse” data structures appropriate here? Absolutely, as most of the bins that are filled will be along the “hyper-diagonal” of the histogram, a sparse data structure is perfect. Thus, issues 1 and 2 are solved. Smoothing this sparse histogram with the method described in this chapter is not practically possible. However, a smearing approach similar to Gaussian blurring in image processing is possible. The steps for this are outlined here:

- Compute the sample covariance, $\mathbf{Q}$, from the data (this is an $O(n^2 + nN_{\text{data}})$ operation)⁷
- Scale $\mathbf{Q}$ by some desired amount, s (this is tunable parameter)
- Construct a multidimensional Gaussian, G_n , using the scale sample covariance, $s\mathbf{Q}$, and its inverse⁸
- Create a bounding box, B , of whose dimensions are the standard deviations of $s\mathbf{Q}$ (i.e. the square root of its diagonal elements) scaled by some factor f_v (this will determine the “range” of the Gaussian, another tunable parameter)
- Ready an empty sparse histogram, S , to store the smoothed values
- Loop over each bin, i , with non-zero content in the sparse histogram and all bins with non-zero content within the bounding box B centered on bin i (accounting for reflecting boundaries if necessary) are filled in S with the value of the G_n centered at the bin center of bin i . These values are then multiplied by the bin original content of bin i and an additional factor to scale up the new bin contents so they add to bin i ’s original content (possibly taking into account different bin sizes if necessary). This is an $O(N_{\text{data}}^2)$ operation at worst (depends on B).

⁷See https://en.wikipedia.org/wiki/Sample_mean_and_covariance

⁸The multidimensional Gaussian in this case is $G(\mathbf{x}|\boldsymbol{\mu}, s\mathbf{Q}) = \sqrt{|2\pi s\mathbf{Q}|}^{-1} \exp -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'(s\mathbf{Q})^{-1}(\mathbf{x} - \boldsymbol{\mu})$

- If the points are too sparse as to have no overlap within B , points in the immediate vicinity of bin i can be used to fill with values in the same manner as stated in the previous step
- The errors at each bin are the quadrature sum of the Gaussian-smeared contributions to that bin’s content (this need to be computed at the same time as the bin contents)

This smoothing procedure can be repeated any number of times. There is, of course, an inherent un-smooth quality to this method as only bins that originally had content (and maybe a few more) are filled. This will certainly leave the tails of a distribution sparsely populated after projecting to 1-D. However, this can be overcome by simply interpolating between the smoothed points in the tail region. Every event will always have an associated weight. A weight for a given event could potentially vary across the different systematics, so the question is now what event weight should be assigned to a given event that is correlated across all the systematics? A solution here is to assign only bin counts to the histogram’s bin content and separately keep track of the total event weights for each axis. After projecting along each axis one would simply scale the projections by the total weight of that axis. One important detailed left to address is that some systematics may not have entries while the others will for a given event. This is (conceptually) easy to accommodate by allocating a special bin along each axis that acts as a special “null bin” that is excluded from smoothing. Thus, one would have to smooth these points using a similar approach as detailed above, but in fewer dimensions.

C.4. Limit Curves After KDE Estimate

Now that the procedure for estimating the distributions and systematic variations is thoroughly established, the effects on the 95% CL curve can be evaluated. The procedure for calculating the limit curve here is identical to the methodology described in Chapter 8 except that the Z +HF background components are merged and smoothed together. This has no impact on the number of NPs in the fit as these background components are already treated together. The resulting curve is given in Figure C.12. As expected, the limit curve

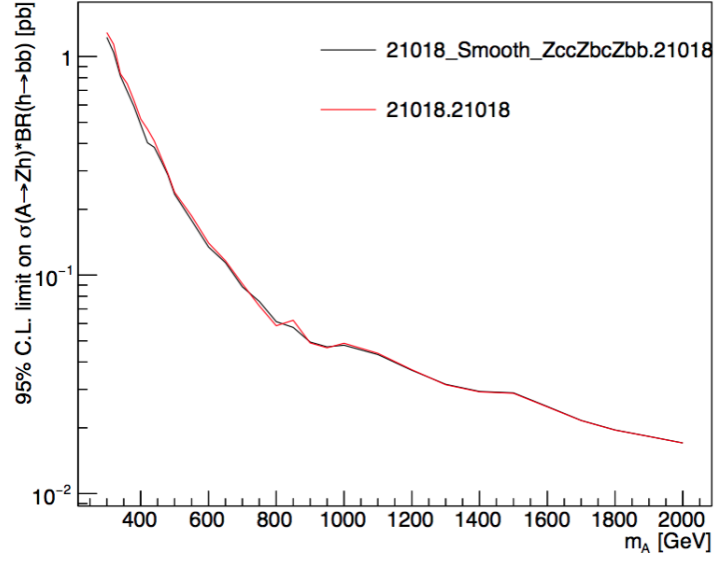


Figure C.12: The ggF $A \rightarrow Zh$ limit curve before (red) and after (black) smoothing the Z +HF background components.

after the kernel estimation is absent the bump at 850 GeV.

BIBLIOGRAPHY

- [1] Simon Dalley. Lecture notes in the Standard Model, 2002.
- [2] Henriette Elvang and Yu-tin Huang. Scattering Amplitudes. 2013.
- [3] I.J.R. Aitchison and A.J.G. Hey. *Gauge Theories in Particle Physics*. Taylor & Francis; 3rd edition, 2004.
- [4] Wikipedia. Standard Model particle table. https://en.wikipedia.org/wiki/Standard_Model. Accessed: 2016-11-11.
- [5] Ivo van Vulpen. Lecture notes in the Standard Model Higgs Boson, October 2013.
- [6] Fabio Maltoni, Giovanni Ridolfi, and Maria Ubiali. b-initiated processes at the LHC: a reappraisal. *JHEP*, 07:022, 2012. [Erratum: *JHEP*04,095(2013)].
- [7] Carl H. Albright. Neutrino cross-sections in a quark parton model with the Weinberg neutral current. *Phys. Rev. D*, D8:3162–3169, 1973.
- [8] F.J. et al. Hasert. Search for elastic muon-neutrino electron scattering. *Physics Letters B*, 46(1):121 – 124, 1973.
- [9] F. R. Villatoro. Higgs potential with Goldstone mode and higgs mode. <http://francis.naukas.com/files/2013/07/dibujo20130704-higgs-potential-goldstone-mode-higgs-mode.jpg>. Accessed: 2016-11-11.
- [10] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964.
- [11] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964.
- [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.
- [13] Yoichiro Nambu. Quasi-particles and gauge invariance in the theory of superconductivity. *Phys. Rev.*, 117:648–663, Feb 1960.
- [14] J. Goldstone. Field theories with « superconductor » solutions. *Il Nuovo Cimento (1955-1965)*, 19(1):154–164, 1961.
- [15] G. et al. Arnison. Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s}=540$ gev. *Physics Letters B*, 122(1):103 – 116, 1983.
- [16] Tilman Plehn and Michael Rauch. Quartic higgs coupling at hadron colliders. *Phys. Rev. D*, 72:053008, Sep 2005.
- [17] K. A. Olive et al. Review of Particle Physics. *Chin. Phys.*, C38:090001, 2014.

- [18] G. 't Hooft. Dimensional regularization and the renormalization group. *Nuclear Physics B*, 61:455 – 468, 1973.
- [19] The CMS Collaboration. Measurement of the ratio of the inclusive 3-jet cross section to the inclusive 2-jet cross section in pp collisions at $\sqrt{s} = 7$ tev and first determination of the strong coupling constant in the tev range. *The European Physical Journal C*, 73(10):2604, 2013.
- [20] Kyle J. Knoepfel. Standard Model Higgs boson searches at the Tevatron. *Mod. Phys. Lett.*, A29(10):1430009, 2014.
- [21] David J. Gross and Frank Wilczek. Ultraviolet behavior of non-abelian gauge theories. *Phys. Rev. Lett.*, 30:1343–1346, Jun 1973.
- [22] Jeffrey R. Forshaw, Michael H. Seymour, and Andrzej Siodmok. On the Breaking of Collinear Factorization in QCD. *JHEP*, 11:066, 2012.
- [23] Ringaile Placakyte. Parton Distribution Functions. In *Proceedings, 31st International Conference on Physics in collisions (PIC 2011): Vancouver, Canada, August 28-September 1, 2011*, 2011.
- [24] LHC Higgs Cross Section Working Group, S. Heinemeyer, C. Mariotti, G. Passarino, and R. Tanaka (Eds.). Handbook of LHC Higgs Cross Sections: 3. Higgs Properties. *CERN-2013-004*, CERN, Geneva, 2013.
- [25] G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, Marc Sher, and Joao P. Silva. Theory and phenomenology of two-Higgs-doublet models. *Phys. Rept.*, 516:1–102, 2012.
- [26] Abdelhak Djouadi. The Anatomy of electro-weak symmetry breaking. II. The Higgs bosons in the minimal supersymmetric model. *Phys. Rept.*, 459:1–241, 2008.
- [27] Rick S. Gupta and James D. Wells. Next Generation Higgs Bosons: Theory, Constraints and Discovery Prospects at the Large Hadron Collider. *Phys. Rev.*, D81:055012, 2010.
- [28] Nathaniel Craig, Jamison Galloway, and Scott Thomas. Searching for Signs of the Second Higgs Doublet. 2013.
- [29] CDF Collaboration and D0 Collaboration. Higgs boson studies at the tevatron. *Phys. Rev. D*, 88:052014, Sep 2013.
- [30] ALEPH Collaboration DELPHI Collaboration L3 Collaboration OPAL Collaboration The LEP Working Group for Higgs Boson Searches. Search for the standard model higgs boson at {LEP}. *Physics Letters B*, 565:61 – 75, 2003.
- [31] The ATLAS Collaboration. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Physics Letters B*, 716(1):1 – 29, 2012.
- [32] The CMS Collaboration. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Physics Letters B*, 716(1):30 – 61, 2012.
- [33] The ATLAS Collaboration. Measurements of higgs boson production and couplings in the four-lepton channel in pp collisions at center-of-mass energies of 7 and 8 tev with the atlas detector. *Phys. Rev. D*, 91:012006, Jan 2015.
- [34] The CMS Collaboration. Measurement of the properties of a higgs boson in the four-lepton final state. *Phys. Rev. D*, 89:092007, May 2014.

- [35] The ATLAS Collaboration. Measurement of higgs boson production in the diphoton decay channel in pp collisions at center-of-mass energies of 7 and 8 tev with the atlas detector. *Phys. Rev. D*, 90:112015, Dec 2014.
- [36] The CMS Collaboration. Observation of the diphoton decay of the higgs boson and measurement of its properties. *The European Physical Journal C*, 74(10):3076, 2014.
- [37] The ATLAS Collaboration. Observation and measurement of higgs boson decays to WW^* with the atlas detector. *Phys. Rev. D*, 92:012006, Jul 2015.
- [38] The CMS Collaboration. Measurement of higgs boson production and properties in the ww decay channel with leptonic final states. *Journal of High Energy Physics*, 2014(1):96, 2014.
- [39] The ATLAS Collaboration and CMS Collaboration. 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):45, 2016.
- [40] Stanley J. Brodsky and Robert Shrock. Maximum wavelength of confined quarks and gluons and properties of quantum chromodynamics. *Physics Letters B*, 666(1):95 – 99, 2008.
- [41] The ATLAS Collaboration. The atlas experiment at the cern large hadron collider. *Journal of Instrumentation*, 3(08):S08003, 2008.
- [42] Lyndon Evans and Philip Bryant. Lhc machine. *Journal of Instrumentation*, 3(08):S08001, 2008.
- [43] The ATLAS Collaboration. Measurement of the Inelastic Proton-Proton Cross Section at $\sqrt{s} = 13$ TeV with the ATLAS Detector at the LHC. 2016.
- [44] The ATLAS Collaboration. Charged-particle distributions in $\sqrt{s}=13$ TeV pp interactions measured with the ATLAS detector at the LHC. *Phys. Lett.*, B758:67–88, 2016.
- [45] G Piacquadio, K Prokofiev, and A Wildauer. Primary vertex reconstruction in the atlas experiment at lhc . *Journal of Physics: Conference Series*, 119(3):032033, 2008.
- [46] M Capeans, G Darbo, K Einsweiler, M Elsing, T Flick, M Garcia-Sciveres, C Gemme, H Pernegger, O Rohne, and R Vuillermet. ATLAS Insertable B-Layer Technical Design Report. Technical Report CERN-LHCC-2010-013. ATLAS-TDR-19, Sep 2010.
- [47] Vasiliki A. Mitsou. The ATLAS transition radiation tracker. In *Astroparticle, particle and space physics, detectors and medical physics applications. Proceedings, 8th Conference, ICATPP 2003, Como, Italy, October 6-10, 2003*, pages 497–501, 2003.
- [48] The ATLAS Collaboration. Trigger for Dummies, 2015.
- [49] M Cattaneo. GAUDI - The Software Architecture and Framework for building LHCb data processing applications . Feb 2000.
- [50] *ATLAS Computing: technical design report*. Technical Design Report ATLAS. CERN, Geneva, 2005.
- [51] R. Brun and F. Rademakers. ROOT: An object oriented data analysis framework. *Nucl. Instrum. Meth.*, A389:81–86, 1997.
- [52] The ATLAS Collaboration. The atlas simulation infrastructure. *The European Physical Journal C*, 70(3):823–874, 2010.

- [53] Stefan Höche. Introduction to parton-shower event generators. In *Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders (TASI 2014) Boulder, Colorado, June 2-27, 2014*, 2014.
- [54] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006.
- [55] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008.
- [56] G. Corcella, I.G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, M.H. Seymour, and B.R. Webber. HERWIG 6.5. *JHEP*, 0101, 2001.
- [57] Tanju Gleisberg, Stefan Hoeche, Frank Krauss, Andreas Schalicke, Steffen Schumann, and Jan-Christopher Winter. SHERPA 1. alpha: A Proof of concept version. *JHEP*, 02:056, 2004.
- [58] J. M. Butterworth, Jeffrey R. Forshaw, and M. H. Seymour. Multiparton interactions in photoproduction at HERA. *Z. Phys.*, C72:637–646, 1996.
- [59] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. *JHEP*, 11:070, 2007.
- [60] 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. *JHEP*, 07:079, 2014.
- [61] Stefano Frixione, Paolo Nason, and Bryan R. Webber. Matching NLO QCD and parton showers in heavy flavor production. *JHEP*, 08:007, 2003.
- [62] D. J. Lange. The EvtGen particle decay simulation package. *Nucl. Instrum. Meth.*, A462:152–155, 2001.
- [63] J. et al. Allison. Geant4 developments and applications. *IEEE Transactions on Nuclear Science*, 53(1):270–278, Feb 2006.
- [64] T Cornelissen, M Elsing, S Fleischmann, W Liebig, E Moyse, and A Salzburger. Concepts, Design and Implementation of the ATLAS New Tracking (NEWT). Technical Report ATL-SOFT-PUB-2007-007. ATL-COM-SOFT-2007-002, CERN, Geneva, Mar 2007.
- [65] R. Frühwirth. Application of Kalman filtering to track and vertex fitting. *Nucl. Instrum. Meth.*, A262:444–450, 1987.
- [66] R. Frühwirth, A. Strandlie, T. Todorov, and M. Winkler. Recent results on adaptive track and multitrack fitting. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 502(2–3):702 – 704, 2003. Proceedings of the {VIII} International Workshop on Advanced Computing and Analysis Techniques in Physics Research.
- [67] A. Strandlie and J. Zerubia. Particle tracking with iterated kalman filters and smoothers: the pmht algorithm. *Computer Physics Communications*, 123(1):77 – 86, 1999.
- [68] Lars Bugge and Jan Myrheim. Tracking and track fitting. *Nuclear Instruments and Methods*, 179(2):365 – 381, 1981.

- [69] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, pages 65–386, 1958.
- [70] The ATLAS Collaboration. A neural network clustering algorithm for the atlas silicon pixel detector. *Journal of Instrumentation*, 9(09):P09009, 2014.
- [71] Robustness of the Artificial Neural Networks Used for Clustering in the ATLAS Pixel Detector. Technical Report ATL-PHYS-PUB-2015-052, CERN, Geneva, Dec 2015.
- [72] Track Reconstruction Performance of the ATLAS Inner Detector at $\sqrt{s} = 13$ TeV. Technical Report ATL-PHYS-PUB-2015-018, CERN, Geneva, Jul 2015.
- [73] 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.
- [74] Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung. Technical Report ATLAS-CONF-2012-047, CERN, Geneva, May 2012.
- [75] Electron and photon energy calibration with the ATLAS detector using data collected in 2015 at $\sqrt{s} = 13$ TeV. Technical Report ATL-PHYS-PUB-2016-015, CERN, Geneva, Aug 2016.
- [76] The ATLAS collaboration. Electron and photon energy calibration with the atlas detector using lhcb run 1 data. *The European Physical Journal C*, 74(10):3071, 2014.
- [77] Electron identification measurements in ATLAS using $\sqrt{s} = 13$ TeV data with 50 ns bunch spacing. Technical Report ATL-PHYS-PUB-2015-041, CERN, Geneva, Sep 2015.
- [78] Georges Aad et al. Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV. *Eur. Phys. J.*, C76(5):292, 2016.
- [79] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti- k_t jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063, 2008.
- [80] Gavin P. Salam and Gregory Soyez. A Practical Seedless Infrared-Safe Cone jet algorithm. *JHEP*, 05:086, 2007.
- [81] Gregory Soyez. The SIScone and anti- $k(t)$ jet algorithms. In *Proceedings, 16th International Workshop on Deep Inelastic Scattering and Related Subjects (DIS 2008): London, UK, April 7-11, 2008*, page 178, 2008.
- [82] The ATLAS Collaboration. JetVertexTagger for run-2 in reco and analysis. <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/JetVertexTaggerTool>. Accessed: 2016-12-08.
- [83] The ATLAS Collaboration. Tagging and suppression of pileup jets with the ATLAS detector. Technical Report ATLAS-CONF-2014-018, CERN, Geneva, May 2014.
- [84] Jet energy scale and its systematic uncertainty in proton-proton collisions at $\sqrt{s}=7$ TeV with ATLAS 2011 data. Technical Report ATLAS-CONF-2013-004, CERN, Geneva, Jan 2013.
- [85] Identification of boosted, hadronically-decaying W and Z bosons in $\sqrt{s} = 13$ TeV Monte Carlo Simulations for ATLAS. Technical Report ATL-PHYS-PUB-2015-033, CERN, Geneva, Aug 2015.
- [86] Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at $\sqrt{s} = 13$ TeV. Technical Report ATL-PHYS-PUB-2015-015, CERN, Geneva, Jul 2015.

- [87] Jet global sequential corrections with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV. *ATLAS-CONF-2015-002*, 2015.
- [88] The ATLAS Collaboration. Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC. Technical Report ATL-PHYS-PUB-2015-045, CERN, Geneva, Nov 2015.
- [89] Ina Chalupkova. Inner tracking performance in ATLAS: vertex and mass resolution. *14th International Conference on B-Physics at Hadron Machines*, 2013.
- [90] J. Fleckner. Performance of Track and Vertex Reconstruction and b -Tagging Studies with ATLAS in pp Collisions at $\sqrt{s} = 7$ TeV. *35th International Conference of High Energy Physics*, 2010.
- [91] b -tagging in dense environments. Technical Report ATL-PHYS-PUB-2014-014, CERN, Geneva, Aug 2014.
- [92] The ATLAS Collaboration. Expected performance of the ATLAS b -tagging algorithms in Run-2. Technical Report ATL-PHYS-PUB-2015-022, CERN, Geneva, Jul 2015.
- [93] Optimisation of the ATLAS b -tagging performance for the 2016 LHC Run. Technical Report ATL-PHYS-PUB-2016-012, CERN, Geneva, Jun 2016.
- [94] ATLAS Flavor Tagging CP Group. Btaggingbenchmarks. <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/BTaggingBenchmarks>. Accessed: 2016-12-08.
- [95] 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.
- [96] 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.
- [97] Steve Farrell. Overlap Removal Tools. https://indico.cern.ch/event/539619/contributions/2191033/attachments/1284239/1909241/Farrell_ort_asg.pdf. Accessed: 2016-12-08.
- [98] The ATLAS Collaboration. Performance of missing transverse momentum reconstruction for the ATLAS detector in the first proton-proton collisions at $\sqrt{s} = 13$ TeV. Technical Report ATL-PHYS-PUB-2015-027, CERN, Geneva, Jul 2015.
- [99] Tatsuya Masubuchi, David Olivier Jamin, Jeffrey Wayne Hetherly, and Spyridon Argyropoulos. Object selections for analyses involving a Higgs boson produced in association with a vector boson in which $h \rightarrow b\bar{b}$ and V decays leptonically with Run-2 data collected with the ATLAS detector at $\sqrt{s} = 13$ TeV. Technical Report ATL-COM-PHYS-2016-480, CERN, Geneva, May 2016.
- [100] Search for a CP-odd Higgs boson decaying to Zh in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. Technical Report ATLAS-CONF-2016-015, CERN, Geneva, Mar 2016.
- [101] Search for the Standard Model Higgs boson produced in association with a vector boson and decaying to a $b\bar{b}$ pair in pp collisions at 13 TeV using the ATLAS detector. Technical Report ATLAS-CONF-2016-091, CERN, Geneva, Aug 2016.
- [102] The ATLAS Collaboration. LuminosityPublicResultsRun2, 2017. <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>.

- [103] William Buttinger. Using Event Weights to account for differences in Instantaneous Luminosity and Trigger Prescale in Monte Carlo and Data. Technical Report ATL-COM-SOFT-2015-119, CERN, Geneva, 2015.
- [104] Jonas Strandberg. Data Preparation Report, 2016.
<https://indico.cern.ch/event/544429/contributions/2210913/attachments/1295579/1932223/DataPreparation.pdf>.
- [105] ATLAS Luminosity Working Group. LuminosityForPhysics, 2017.
<https://twiki.cern.ch/twiki/bin/view/Atlas/LuminosityForPhysics>.
- [106] Hung-Liang Lai, Marco Guzzi, Joey Huston, Zhao Li, Pavel M. Nadolsky, Jon Pumplin, and C. P. Yuan. New parton distributions for collider physics. *Phys. Rev.*, D82:074024, 2010.
- [107] Peter Zeiler Skands. Tuning Monte Carlo Generators: The Perugia Tunes. *Phys. Rev.*, D82:074018, 2010.
- [108] Michał Czakon and Alexander Mitov. Top++: A program for the calculation of the top-pair cross-section at hadron colliders. *Computer Physics Communications*, 185(11):2930 – 2938, 2014.
- [109] M. Aliev, H. Lacker, U. Langenfeld, S. Moch, P. Uwer, and M. Wiedermann. {HATHOR} – {HADronic} top and heavy quarks cross section calculator. *Computer Physics Communications*, 182(4):1034 – 1046, 2011.
- [110] Tanju Gleisberg and Stefan Höche. Comix, a new matrix element generator. *Journal of High Energy Physics*, 2008(12):039, 2008.
- [111] F. Cascioli, P. Maierhöfer, and S. Pozzorini. Scattering amplitudes with open loops. *Phys. Rev. Lett.*, 108:111601, Mar 2012.
- [112] Stefan Höche, Frank Krauss, Marek Schönherr, and Frank Siegert. Qcd matrix elements + parton showers. the nlo case. *Journal of High Energy Physics*, 2013(4):27, 2013.
- [113] S. Catani, F. Krauss, R. Kuhn, and B. R. Webber. QCD matrix elements + parton showers. *JHEP*, 11:063, 2001.
- [114] Keith Hamilton, Peter Richardson, and Jon Tully. A Modified CKKW matrix element merging approach to angular-ordered parton showers. *JHEP*, 11:038, 2009.
- [115] ATLAS Run 1 Pythia8 tunes. Technical Report ATL-PHYS-PUB-2014-021, CERN, Geneva, Nov 2014.
- [116] The ATLAS Collaboration. Summary of ATLAS Pythia 8 tunes. 2012.
- [117] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt. Parton distributions for the lhc. *The European Physical Journal C*, 63(2):189–285, 2009.
- [118] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt. Measurement of the z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ tev with the atlas detector. *Journal of High Energy Physics*, 2014(9):145, 2014.
- [119] Ansgar Denner, Stefan Dittmaier, Stefan Kallweit, and Alexander 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, 2015.

- [120] M et al. Baak. Data Quality Status Flags and Good Run Lists for Physics Analysis in ATLAS. Technical Report ATL-COM-GEN-2009-015, CERN, Geneva, 2009.
- [121] ATLAS Luminosity Working Group. ATLAS Luminosity Calculator, 2017.
<https://atlas-lumicalc.cern.ch/>.
- [122] The ATLAS Collaboration. ExtendedPileupRewighting, 2015.
<https://twiki.cern.ch/twiki/bin/view/AtlasProtected/ExtendedPileupRewighting>.
- [123] A Buckley, T Eifert, M Elsing, D Gillberg, K Koenek, A Krasznahorkay, and E Moyse. Report of the xAOD design group. Technical Report ATL-COM-SOFT-2013-022, CERN, Geneva, Sep 2013.
- [124] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J.*, C71:1554, 2011. [Erratum: *Eur. Phys. J.*C73,2501(2013)].
- [125] Song-Ming et al. Wang. Supporting Note for the 2HDM Higgs $A \rightarrow Zh$, $h \rightarrow b\bar{b}$ searches. Technical Report ATL-COM-PHYS-2016-030, CERN, Geneva, Jan 2016.
- [126] F et al. Ahmadov. Supporting Document for the Search for the $b\bar{b}$ decay of the Standard Model Higgs boson in associated (W/Z)H production with the ATLAS detector. Technical Report ATL-COM-PHYS-2014-051, CERN, Geneva, Jan 2014.
- [127] T. et al. Masubuchi. Search for a Standard Model Higgs boson produced in association with a vector boson and decaying to a pair of b-quarks. Technical Report ATL-COM-PHYS-2016-429, CERN, Geneva, May 2016.
- [128] The ATLAS Collaboration. 2015 Tau Recommendations , 2015.
<https://twiki.cern.ch/twiki/bin/view/AtlasProtected/TauPreRecommendations2015>.
- [129] P et al. Bechtle. Reconstruction of hadronic decay products of tau leptons with the ATLAS experiment. Technical Report ATL-COM-PHYS-2015-214, CERN, Geneva, Mar 2015.
- [130] A.D. Bukin. Fitting function for asymmetric peaks. 2007.
- [131] Andreas Hoecker, Peter Speckmayer, Joerg Stelzer, Jan Therhaag, Eckhard von Toerne, and Helge Voss. TMVA: Toolkit for Multivariate Data Analysis. *PoS, ACAT:040*, 2007.
- [132] Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119 – 139, 1997.
- [133] Yohei Yamaguchi. $t\bar{t}b\bar{b}$ rejection for 1 lepton, 2016.
https://indico.cern.ch/event/484076/contributions/2002318/attachments/1249283/1841196/20160331_Mtop.pdf.
- [134] Pedro, Rute and Conde Muino, P. and Maneira, J. WH BDT input variables, 2015.
https://indico.cern.ch/event/384507/contributions/915174/attachments/1141852/1635851/2015_08_19_MVAinputVars.pdf.
- [135] The ATLAS Collaboration. 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.
- [136] Andy Buckley, Jonathan Butterworth, Leif Lonnblad, David Grellscheid, Hendrik Hoeth, James Monk, Holger Schulz, and Frank Siegert. Rivet user manual. *Comput. Phys. Commun.*, 184:2803–2819, 2013.

- [137] LHC Higgs Cross Section Working Group. CERN Yellow Report 4.
- [138] Wouter Verkerke and David P. Kirkby. The RooFit toolkit for data modeling. *eConf*, C0303241:MOLT007, 2003. [,186(2003)].
- [139] L. Moneta and et al. The RooStats Project. *PoS*, ACAT:057, 2010.
- [140] Kyle Cranmer, George Lewis, Lorenzo Moneta, Akira Shibata, and Wouter Verkerke. HistFactory: A tool for creating statistical models for use with RooFit and RooStats. Technical Report CERN-OPEN-2012-016, New York U., New York, Jan 2012.
- [141] Olaf Behnke, Kevin Kröninger, Grégory Schott, and Thomas Schörner-Sadenius. *Data analysis in high energy physics: a practical guide to statistical methods*. Wiley-VCH, Weinheim, 2013.
- [142] A L Read. Presentation of search results: the cl_s technique. *Journal of Physics G: Nuclear and Particle Physics*, 28(10):2693, 2002.
- [143] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *The European Physical Journal C*, 71(2):1554, 2011.
- [144] Kirill Melnikov and Frank Petriello. Electroweak gauge boson production at hadron colliders through $O(\alpha_s^2)$. *Phys. Rev.*, D74:114017, 2006.
- [145] P. Kant, O.M. Kind, T. Kintscher, T. Lohse, T. Martini, S. Molbitz, P. Rieck, and P. Uwer. 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, 2015.
- [146] The ATLAS Collaboration. Search for the $b\bar{b}$ decay of the Standard Model Higgs boson in associated $(W/Z)H$ production with the ATLAS detector. *JHEP*, 01:069, 2015.
- [147] The ATLAS Collaboration. Search for a CP-odd Higgs boson decaying to Zh in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Phys. Lett.*, B744:163–183, 2015.
- [148] Daniel Whiteson. Deep Learning in Analysis, 2015.
- [149] T. Nagler and C. Czado. Evading the curse of dimensionality in nonparametric density estimation with simplified vine copulas. *ArXiv e-prints*, March 2015.
- [150] Hetherly, Jeffrey Wayne and ATLAS Collaboration. Real-time Flavour Tagging Selection in ATLAS. Technical Report ATL-DAQ-PROC-2015-035, CERN, Geneva, Oct 2015.
- [151] The ATLAS Collaboration. HLT Steering, guide for algorithm developers, 2016.
- [152] M. et al. Capeans. *ATLAS Insertable B-Layer Technical Design Report*. (CERN-LHCC-2010-013. ATLAS-TDR-19), Sep 2010.
- [153] M. et al. Shochet. *Fast Tracker (FTK) Technical Design Report*. (CERN-LHCC-2013-007. ATLAS-TDR-021), Jun 2013.
- [154] The ATLAS Collaboration. Public b-Jet Trigger Plots for Collision Data: Expected performance of the ATLAS b-jet trigger in Run 2, 2015.
<https://twiki.cern.ch/twiki/bin/view/AtlasPublic/BJetTriggerPublicResults>.
- [155] Commissioning of the ATLAS b -tagging algorithms using $t\bar{t}$ events in early Run-2 data. Technical Report ATL-PHYS-PUB-2015-039, CERN, Geneva, Aug 2015.

- [156] M.P. Wand and M.C. Jones. *Kernel Smoothing*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1994.
- [157] A. C. Aitken. On Least Squares and Linear Combinations of Observations. *Proceedings of the Royal Society of Edinburgh* **vol. 55**, **pp. 42–48** (1935).
- [158] John A. Nelder and Mead. R. A simplex method for function minimization. *Computer Journal* **7**, **pp. 308–313** (1965).
- [159] Steven Schramm, Sophio Pataraiia, Daniela Boerner, Bogdan Malaescu, and Christopher John Young. A method for the construction of strongly reduced representations of ATLAS experimental uncertainties and the application thereof to the jet energy scale. Technical Report ATL-COM-PHYS-2015-374, CERN, Geneva, May 2015.
- [160] Jeffrey S. Simonoff. *Smoothing Methods in Statistics*. Springer, 1998.
- [161] Kyle S. Cranmer. Kernel estimation in high-energy physics. *Comput. Phys. Commun.*, 136:198–207, 2001.
- [162] The ATLAS Collaboration. Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at $\sqrt{s} = 13$ TeV. Technical Report ATL-PHYS-PUB-2015-015, CERN, Geneva, Jul 2015.
- [163] Gordon Watts, Frank Filthaut, and Giacinto Piacquadio. Extrapolating Errors for b-tagging. Technical Report ATL-COM-PHYS-2015-711, CERN, Geneva, Jul 2015. Internal collaboration document.
- [164] The ATLAS Collaboration. Usage of Missing ET in analyses: rebuilding and systematics, 2015. <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/METUtilities>.
- [165] Michael E. Peskin and Daniel V. Schroeder. *An Introduction To Quantum Field Theory*. Addison-Wesley, 1995.
- [166] Data-driven determination of the energy scale and resolution of jets reconstructed in the ATLAS calorimeters using dijet and multijet events at $\sqrt{s}=8$ TeV . *ATLAS-CONF-2015-017*, 2015.