

**Search for Dark Matter in Proton-Proton
Collisions at a Center-of-Mass Energy of 13 TeV in
the Higgs Boson associated b-anti-b quark channel**

Jue Chen

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2020

©2020

Jue Chen

All Rights Reserved

ABSTRACT

Search for Dark Matter in Proton-Proton Collisions at a Center-of-Mass Energy of 13 TeV in the Higgs Boson associated b-anti-b quark channel

Jue Chen

This work presents the search for Dark Matter particles associated with the Higgs Boson decaying into a $b\bar{b}$ quark pair. The dark matter search result is based on proton-proton collision data collected at a center-of-mass energy of 13 TeV by the ATLAS detector during Run II with an integrated luminosity of 160 fb^{-1} . Two simplified models, as dark matter theory candidate in particle physics, are the target of searching with collision data. The new powerful Higgs tagging techniques, which exploits the jet substructure and heavy flavor information on large extent, are developed to improve the search sensitivity of the search. The target physics signals are signatured with optimized search region, and interpreted with background estimation result in the statistical manner.

Table of Contents

I	Introduction	1
1	Introduction	2
II	The standard model and Dark Matter	4
2	The Standard Model	5
2.1	Introduction	5
2.2	Particles	6
2.3	Interactions	8
2.4	Gauge theory	8
2.5	The formation of the Lagrangian of the Standard Model	9
2.6	Spontaneous Symmetry Breaking and the Higgs mechanism	10
2.7	Challenges	11
3	Dark matter	13
3.1	Dark matter: From astrophysics to particle physics	13
3.2	Weakly Interacting Massive Particles	15
3.3	Searches for dark matter in collider experiments	16
3.3.1	Effective field theory	16
3.3.2	The simplified model	18
3.3.3	Z' and Two-Higgs-doublet models	18

III	The LHC and ATLAS experiment	22
4	The LHC	23
4.1	The LHC performance	24
4.1.1	Beam energy	24
4.1.2	Luminosity	24
4.2	The LHC operation	26
5	The ATLAS experiment	28
5.1	Inner detector	29
5.1.1	Pixel detector	32
5.1.2	Semiconductor Tracker (SCT)	33
5.1.3	Transition Radiation Tracker (TRT)	33
5.2	Calorimeter	33
5.2.1	Liquid Argon Calorimeter	35
5.2.2	Tile Calorimeter	36
5.3	Muon Spectrometer (MS)	38
5.4	Forward Detectors	39
5.5	Trigger and Data Acquisition	40
IV	ATLAS Physics Objects and Event Reconstruction	43
6	Physics Objects	44
6.1	Inner Detector Tracks and Primary Vertex	44
6.2	Electrons	45
6.3	Muons	46
6.4	Taus	47
6.5	Jets	48
6.5.1	Calorimeter jets	49
6.5.2	Track jets	51
6.5.3	Ghost Association	52

6.5.4	Flavor tagging	53
6.6	Missing Transverse Momentum (E_T^{miss}) and Missing Transverse Momentum Significance	56
6.7	Higgs tagging	57
6.7.1	Higgs tagging with advanced subjets	57
7	Double-b tagger	60
7.1	Introduction	60
7.2	Monte Carlo Simulation Samples	61
7.3	Double-b tagger algorithm	62
7.3.1	Input features	62
7.4	Neural network architecture	64
7.5	Signal efficiency and Background mistag rate measurement	65
V	Dark Matter search in the Higgs Boson associated $b\bar{b}$ decay	67
8	Data and MC samples	68
8.1	Data samples	68
8.2	Trigger	68
8.3	MC samples	72
8.3.1	2HDM+a and Z'-2HDM signal	72
8.3.2	V+jets	74
8.3.3	$t\bar{t}$	74
8.3.4	Single top	75
8.3.5	Diboson	75
8.3.6	SM $Vh(b\bar{b})$	75
8.3.7	$t\bar{t} + Z/H$	76
9	Signal selection	77
9.1	Physics objects definition	77
9.1.1	Leptons	77

9.1.2	Jets	78
9.1.3	Missing transverse momentum	79
9.2	Signal regions	80
9.2.1	Common event selection	80
9.2.2	Resolved region	83
9.2.3	Merged region	84
9.3	Control regions	84
9.3.1	1-lepton channel	84
9.3.2	2-lepton channel	86
10	Systematic uncertainty	88
10.1	Experimental systematics	88
10.2	Theoretical systematics	88
10.3	Data-MC comparison	91
11	Result	101
11.1	Sample section	101
11.1.1	Sample subsection	101
11.1.2	Sample subsubsection	101
11.2	Sample section	102
11.2.1	Sample subsection	102
VI	Conclusions	103
12	Conclusions	104
VII	Appendices	105
A	The ATLAS detector service work	106
A.1	Sample section	106
A.1.1	Sample subsection	106
A.2	Sample section	106

A.2.1	Sample subsection	107
B	CombinedXbbScore Tagger implementation	108
B.1	Introduction	108
B.2	Performance	108
B.3	Signal selection in Merged region using CombinedXbbScore	109
B.3.1	Backgrounds breakdown of signal region and control regions	110
B.3.2	Preliminary significance study	114
VIII	Bibliography	116
	Bibliography	117

List of Figures

2.1	Particles of the Standard Model of particle physics	6
3.1	The rotation curve for spiral galaxy NGC 3198. The blue diamonds are the observations and reveal the almost flat-like nature of the curve in the outer regions of the galaxy. The dashed green line is the curve for Newtonian gravity. It shows that the rotational velocity should decrease with distance from the galaxy center. The solid red line through the data points is the curve obtained by assuming a simple Gaussian energy scale variation and a simple Gaussian density distribution for the galaxy.	14
3.2	Left: an s-channel Feynman diagram of the interaction between quarks and dark matter candidates via a scalar mediator. Right: the corresponding effective field theory contact interaction.	17
3.3	Feynman diagram of the interaction between Standard Model quarks and a Z' boson.	20
3.4	The Feynman diagram of the interaction between the Standard Model quarks and Z' boson.	20
3.5	The Feynman diagram of $b\bar{b}$ decayed Higgs channel in $Z' + 2HDM$	21
4.1	The LHC full injection chain.	24
4.2	Cumulative luminosity versus time delivered to ATLAS (green), recorded by ATLAS (yellow) during stable beams for pp collisions at 13 TeV centre-of-mass energy in 2015-2018.	25
5.1	Cut-away view of the ATLAS detector	29

5.2	An image of the ATLAS detector, showing the different layers and the interactions of different particle types through the layers.	30
5.3	Cut-away view of the ATLAS Inner Detector.	31
5.4	A schematic view of the active region of the Pixel detector consisting of barrel and end-cap layers (??) and the IBL detector before the insertion (5.4b) . .	32
5.5	Cut-away view of ATLAS calorimeter system	34
5.6	Accordion structure of the barrel. The left figure is a view of a small sector of the barrel calorimeter in a plane transverse to the LHC beams. The right figure is a diagram of a sketch of barrel module showing both accordion structure and granularity in $\eta - \phi$ of the cells on each of layer	36
5.7	Illustration of a Tile calorimeter module.	37
5.8	Geometric layout of muon sub-detectors in barrel (5.8a) and end-cap (5.8b) region.	38
5.9	Cut-away view of the ATLAS Inner Detector.	39
5.10	b	40
5.11	Cross sections of physics processes produced by hadron colliders 5.10	40
5.12	The ATLAS TDAQ system in Run 2 with the relevant components for triggering.	41
6.1	Calibration stages for reconstructed jets.	49
6.2	The Expected performance of the b -tagging algorithm (MV2c10) for b -jet triggers in 2018 data-taking (blue solid line) is compared to the same b -tagging algorithm trained on the Hybrid training sample (red solid line). . .	55
6.3	A cartoon depicting using VR track jets instead of $R = 0.2$ track jets.	57
6.4	Higgs efficiency using two VR track jets with different R parameter (Left) and ρ parameter (Right)	58
6.5	Higgs efficiency using four different subjet techniques: $R = 0.2$ track jets, VR track jets, ExKt calorimeter jets, and CoM calorimeter jets.	58
6.6	ROC curves using four different subjet techniques: $R = 0.2$ track jets, VR track jets, ExKt calorimeter jets, and CoM calorimeter jets in two different jet p_T ranges	59

7.1	ROC plot: Comparison between Hbb taggers (Left) and Top taggers (Right).	65
7.2	Higgs efficiency (left) and Background rejection (right) as a function of large-radius jet p_T	66
8.1	Measured trigger efficiencies as a function of offline $E_{T,\mu}^{\text{miss}}_{\text{invis.}}$ in data and MC for the E_T^{miss} triggers used in 2015-2018. The plots are shown for 0,1 and 2 tags together. The lower panels provide the ratio of data and MC events (the scale factor).	70
8.2	Measured scale factors as a function of offline $E_{T,\mu}^{\text{miss}}_{\text{invis.}}$ for the E_T^{miss} triggers used in 2015-2010. The scale factors were derived for 0,1 and 2 tags together. The hatched band shows the 1σ fit uncertainty.	72
8.3	Validation plots showing E_T^{miss} trigger efficiencies and scale factors as function of offline $E_{T,\mu}^{\text{miss}}_{\text{invis.}}$ after applying scale factor corrections for $E_{T,\mu}^{\text{miss}}_{\text{invis.}} > 100 \text{ GeV}$ to the MC in the full single lepton region. Good agreement is observed between data and MC.	73
9.1	Decay categories for $t\bar{t}$ events in the signal region. The classification of the decay processes was performed by examining the truth particle information within the reconstructed signal region events. The three dominant decay modes are the semileptonic decays with either a light lepton, a leptonically or a hadronically decaying τ -lepton.	86
9.2	Normalised distributions of E_T^{miss} , the Higgs boson candidate mass and the number of small-radius jets for $t\bar{t}$ events. The distributions are shown separately for the three semileptonic decay modes, which are the dominant decays modes in the signal region. The shapes of the distributions agree well within the statistical uncertainties.	87
10.1	Higgs candidate mass spectra in the different E_T^{miss} regions with 2 b -tagged jets in the 0-lepton channel.	92
10.2	Higgs candidate mass spectra in the different E_T^{miss} regions with at least 3 b -tagged jets in the 0-lepton channel.	96

10.3 Muon charge distribution in the different E_T^{miss} regions with 2 b -tagged jets in the 1-lepton channel.	97
10.4 Muon charge distribution in the different E_T^{miss} regions with least 3 b -tagged jets in the 1-lepton channel.	98
10.5 Total yields in the 2-lepton control region for different E_T^{miss} regions with 2 b -tagged jets.	99
10.6 Total yields in the 2-lepton control region for different E_T^{miss} regions with more than 2 b -tagged jets.	100
B.1 ROC plot: Comparison between different working points of CombinedXbbScore	109
B.2 Leading Fatjet mass in Merged region (left) and 2b Merged region defined by 2-b tagged VR (right)	110
B.3 Leading Fatjet mass in 2b Merged region defined by CombinedXbbScore . .	110
B.4 Large-R jet mass distribution in the 2-btagged singal and control regions. .	111
B.5 Fraction of large-R jets with bb labeling (left) and without bb labeling (right) in with W+jets samples.	112
B.6 Fraction of large-R jets with bb labeling (left) and without bb labeling (right) in with Z+jets samples.	112
B.7 Flavor breakdown of backgrounds with truth labeling in singal and control regions.	114
B.8 Ratio of singal significance of 2b merged region defined by CombinedXbb-Score compared to the orignal method with or without the Higgs window. .	115

List of Tables

2.1	Three fundamental interactions in Standard Model.	8
4.1	LHC beam modes.	27
6.1	Selection criteria for tracks used in the reconstruction of to reconstruct vertices.	45
6.2	Definitions for the different categories of electron.	46
6.3	Definitions for the different categories of muons.	47
6.4	Selection criteria for tracks to cluster track jets.	52
7.1	Input features used for training double-b tagger.	63
8.1	E_T^{miss} and single-lepton triggers used in the analysis.	71
9.1	Electron selection criteria.	78
9.2	Muon selection criteria.	78
9.3	Tau selection criteria.	79
9.4	Small- R jet reconstruction criteria.	80
9.5	Large- R jet reconstruction criteria.	81
9.6	b-jets selection criteria.	82
9.7	E_T^{miss} reconstruction criteria.	82
9.8	Summary of the resolved and merged event selection applied in the 0-lepton channel. \mathcal{S} denotes the object based E_T^{miss} significance.	85
10.1	A summary of the experimental systematic uncertainties.	93
10.2	A summary of the experimental systematic uncertainties.	94
10.3	A summary of the experimental systematic uncertainties.	95

B.1	Preliminary working points of CombinedXbbScore	109
B.2	Flavor breakdown of backgrounds with truth labeling in singal region.	113
B.3	Flavor breakdown of backgrounds with truth labeling in 1 lepton control region.	113
B.4	Flavor breakdown of backgrounds with truth labeling in 2 lepton control region.	114

Acknowledgments

The acknowledgments go here. The acknowledgments go here.

Dedication text

Part I

Introduction

Chapter 1

Introduction

Over the past few decades, the Standard Model has been established to describe the fundamental building blocks of the universe and their interactions. Not only is the Standard Model elegant and self-consistent, but also scientists observed all force carriers, predicted by the Standard Model with the evolution of particle colliders to the corresponding energy scale. The Higgs boson, which plays a key role in the Standard Model through spontaneous symmetry breaking was discovered in 2012 by both ATLAS and CMS collaborations, and completes the Standard Model[21] [60].

Despite being remarkably successful, the Standard Model has limitations and leaves us with some questions. New hypothetical models have been developed, like supersymmetry (SUSY) and extra dimensions, to break down the huge iceberg in front of us. The increase of collision energy and luminosity at the large hadron collider will benefit the search for new physics.

In order to probe new physics at an energy scale higher than electroweak scale, we measure the particles produced in proton-proton collisions using A Toroidal LHC ApparatuS (ATLAS) detector system, which is one of the two general purpose detectors at the CERN LHC. This work presents a search for Dark Matter particles produce in association with the Higgs Boson, using 160 fb^{-1} of proton-proton collision data collected at a center-of-mass energy of 13 TeV by the ATLAS detector at the Large Hadron Collider during Run 2. Two theoretical dark matter models, Z'+2HDM and 2HDM+a, are used to interpret the thesis.

In particular, we examine the channel where the Higgs boson decays to a pair of bottom

quarks, taking advantage of the large branching ratio of the $b\bar{b}$ mode. Two Higgs boson reconstruction methods are used, the resolved channel where the Higgs boson is reconstructed as two separate b-quark jets, and the boosted channel where the Higgs boson is reconstructed as a single large-radius jet.

This thesis is structured as follows. In Chapter 2, the Standard Model of particle physics and its open questions are briefly described. Dark matter and two Higgs doublet models, as a candidate theory to resolve the dark matter challenges, are exposed in Chapter 3. Then, the CERN Large Hadron Collider and ATLAS detector are explained in detail as an experimental solution to examine new particle physics theories in Chapter 4 and Chapter 5. The physics analysis workflow and statistical interpretation are demonstrated from Chapter 8 to 11. Finally, the conclusions are drawn in Chapter 12.

Part II

The standard model and Dark Matter

Chapter 2

The Standard Model

2.1 Introduction

What are the most fundamental particles that constitute our world? How do the particles interact with each other? These are two of the essential questions.

By the early twentieth century, scientists believed that atoms were the building blocks of nature. People trusted Newtonian laws of motion which solved most problems of physics. Fundamental precepts of physics at that time were challenged by the establishment of Einstein's theory of relativity and quantum mechanics. Just as Max Born predicted, "Physics as we know it will be over in six months".

Guided by both theoretical models and experimental discoveries, quantum field theory became very successful in describing particles as well as electromagnetic and weak interactions between them. The first evidence for a subatomic particle, proton, was found in 1919 and followed by the emergence of the strong force in 1921 as a new interaction type which holds the subatomic particles.

Starting from 1964 when Murray Gell-Mann and George Zweig put forth the idea of quarks, the Standard Model was gradually completed and finally summarized by John Iliopoulos in 1974.

In Sections 2.2 and 2.3, an overview of the Standard Model of particle physics and its interactions are given, while a brief description of gauge theory follows in Section 2.4. We introduce the Standard Model Lagrangian of in Section 2.5 to explain how particles gain

their masses via the Higgs mechanism in Section 2.6.

2.2 Particles

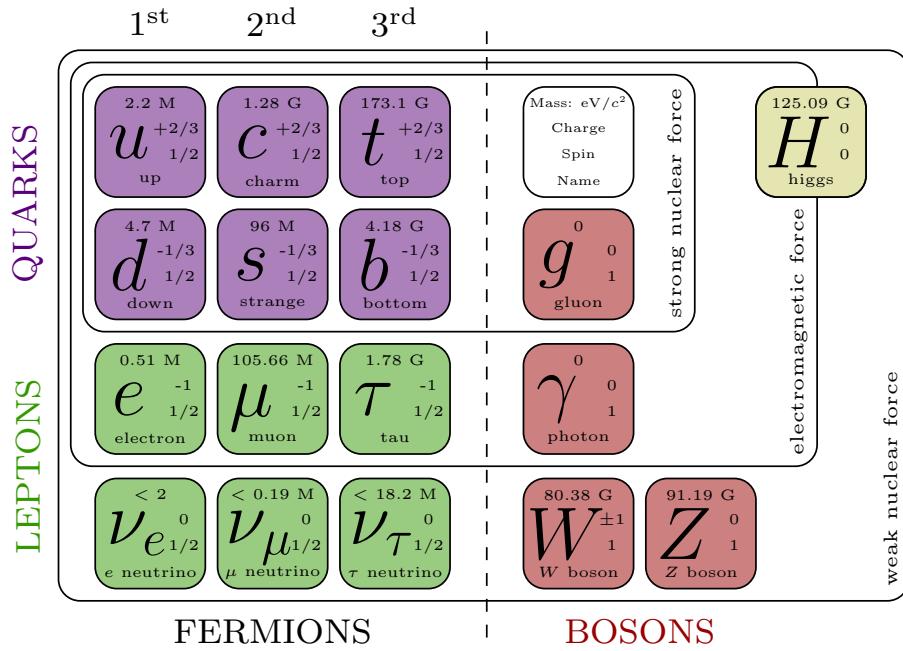


Figure 2.1: Particles of the Standard Model of particle physics

Shown in Fig 2.1, the fundamental building blocks of matter are fermions with $\frac{1}{2}$ spin in the Standard Model, while the mediators of forces of their interactions are gauge bosons with one spin. The spin-0 scalar Higgs boson is expected to be at the origin of the mass of other particles.

- **Fermions** can be divided into two categories: quarks, and leptons. There are three generations of both kinds: the first generation made up common matter, and the higher generations can be accessed at higher energies.
 - **Quarks** Each generation of quarks has two types: up-type with a charge of $\frac{2}{3}$ and down-type with a charge of $-\frac{1}{3}$ and both can interact through the electromagnetic, strong and weak interactions. The color charge carried by quarks makes them able to interact via the strong interactions. Asymptotic freedom explains why

quarks need to bind together at distance largert than , resulting in the color-neutral particles called *hadrons*. There are two types of hadrons: *baryons* and *mesons*. Protons and neutrons are two examples of baryons while pions and kaons are examples of mesons. Weak isospin makes quarks able to couple via weak interaction like fermions.

- **Leptons** Each generation has a charged particle and a electrically neutral neutrino. Leptons don’t have color charge so they can’t interact via strong interactions. While the charged particles can interact through the electromagnetic and weak forces as they have both charge and weak isospin, neutrinos only interact via weak interactions, making them extremely hard to detect in experiments.
- **Bosons** as mediators for the three types of force in the Standard Model are listed below.
 - **Photons:** discovered very early, it is the mediator of the electromagnetic force. It’s a massless, electromagnetic-charge-neutral, spin-1 particle.
 - **Gluons:** first discovered at DESY in the late 1970s, it is the mediator of the strong force. Gluons are massless, spin-1 particles. The gluon caries color charge itself and interacts with quarks. There are eight varieties of gluons as there are nine different combinations of the color charge, but the singlet state $\frac{r\bar{r}+b\bar{b}+g\bar{g}}{\sqrt{3}}$ does not exist.
 - The W^\pm and Z bosons: discovered in the late 1980s, serve as mediators of the weak force. They are spin-1 particles with masses around 80 and 91 GeV. The W^\pm bosons carry the weak charged current and have electromagnetic charge of ± 1 , while the Z boson is the mediator of the weak neutral current and is electromagnetic-charge-neutral.
 - **Higgs** boson, discovered by the CMS and ATLAS collaborations in 2012[21][60], is a spin-0 scalar neutral particle with mass 125GeV. Fermions and bosons acquire their mass through Yukawa coupling with the Higgs field.

2.3 Interactions

Force type	Mediator	Affected particles	Acts on	Coupling Constant
Electromagnetic	γ	Electrically charged fermions	Electric charge	α
Strong	g	Quarks	Color charge	α_s
Weak	W^\pm, Z	Left-handed fermions	Weak hypercharge	α_W

Table 2.1: Three fundamental interactions in Standard Model.

Table 2.1 summarizes the properties of the three different interactions, including the values of coupling constants at low energy. At low energies, the weak force is much weaker than the electromagnetic force, even though the weak coupling constant is relatively larger. This is because the strength of weak interactions is suppressed by the large masses of the W and Z bosons.

2.4 Gauge theory

Gauge Theories are field theories in which the Lagrangian is invariant under certain Lie groups of local transformations.

As a gauge theory, the Standard Model Lagrangian is invariant under transformations of the group $SU(3)_c \times SU(2)_L \times U(1)_Y$. The fermions of the Standard Model are described by representations of the symmetry group, while the local gauge symmetry is represented by a force mediated by gauge bosons.

There are twelve gauge bosons in total: eight gluons which correspond to the generators of $SU(3)_c$, W^\pm bosons which correspond to generators of $SU(2)_L$, and the Z boson and γ which correspond to linear combinations of generators for $SU(2)_L \times U(1)_Y$.

For fermions, two different representations are used based on chirality: left-handed fermions are doublets under $SU(2)$ and interact with the weak bosons, while right-handed fermions are singlets.

$$\psi_L^j = \begin{pmatrix} \psi_{L+}^j \\ \psi_{L-}^j \end{pmatrix}, \psi_{R\sigma}^j \quad (2.1)$$

Here $j = 1, 2, 3$ is the generation index. For quarks, $\sigma = +$ represents up-type quark and $\sigma = -$ represents down-type quarks, while for leptons, $\sigma = +$ represents neutrinos and

$\sigma = -$ represents charged leptons. As neutrinos are considered to be massless, there are no right-handed neutrinos:

2.5 The formation of the Lagrangian of the Standard Model

To understand why the Higgs field is responsible for the masses of other particles, the Lagrangian of the Standard Model is examined.

$$\mathcal{L}_{\text{Standard Model}} = \mathcal{L}_{\text{Gauge}} + \mathcal{L}_{\text{Fermion}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}}, \quad (2.2)$$

where $\mathcal{L}_{\text{Gauge}}$ describes the kinematics of the gauge fields, which are written as

$$\mathcal{L}_{\text{Gauge}} = -\frac{1}{4}G_{a\mu\nu}G_a^{\mu\nu} - \frac{1}{4}W_{a\mu\nu}W_a^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}. \quad (2.3)$$

Here $W_{\mu\nu}^a$ and $B_{\mu\nu}$ are the field tensors corresponding to non-Abelian SU(2) and Abelian U(1) respectively. $\mathcal{L}_{\text{Fermion}}$ describes the fermion kinematics and interactions with gauge bosons and is written as:

$$\mathcal{L}_{\text{Fermion}} = \sum_j \overline{\psi_L^j} i\gamma^\mu D_\mu^L \psi_L^j + \sum_{j,\sigma} \overline{\psi_{R\sigma}^j} i\gamma^\mu D_\mu^R \psi_{R\sigma}^j, \quad (2.4)$$

where γ^μ are the Dirac matrices and D_μ is the covariant derivative operator which is defined as

$$D_\mu = \partial_\mu - \frac{ig_1 Y}{2} B_\mu - \frac{ig_2 \tau^i}{2} \mathbf{W}_\mu^i - i \frac{ig_3 \lambda^a}{2} \mathbf{G}_\mu^a. \quad (2.5)$$

Here Y , τ^i , and λ are the generators for the U(1), SU(2) and SU(3) gauge symmetry groups, and g_1 , g_2 and g_3 are coupling constants between fermion and gauge fields. B_μ is the spin-one field needed to maintain the U(1) gauge invariance, and \mathbf{W}_μ and \mathbf{G}_μ are composed of 2×3 and 3×3 traceless Hermitian matrices. They are associated with the field tensors above via:

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu,$$

$$\mathbf{W}_{\mu\nu} = \partial_\mu \mathbf{W}_\nu - \partial_\nu \mathbf{W}_\mu + ig_2 \frac{(\mathbf{W}_\mu \mathbf{W}_\nu - \mathbf{W}_\nu \mathbf{W}_\mu)}{2}, \text{ and}$$

$$\mathbf{G}_{\mu\nu} = \partial_\mu \mathbf{G}_\nu - \partial_\nu \mathbf{G}_\mu + (\mathbf{G}_\mu \mathbf{G}_\nu - \mathbf{G}_\nu \mathbf{G}_\mu).$$

The gauge bosons, the charged W^\pm , photon γ (A_μ) and Z_μ bosons arise as follows:

$$A_\mu = W_{11\mu} \sin(\theta_w) + B_\mu \cos(\theta_w),$$

$$Z_\mu = W_{11\mu} \cos(\theta_w) - B_\mu \sin(\theta_w),$$

$$W_\mu^+ = W_\mu^{-*} = \frac{W_{12\mu}}{\sqrt{2}}.$$

Higgs term describes the kinematic and potential energies of the Higgs field ϕ :

$$\mathcal{L}_{Higgs} = T - V = \overline{(D_\mu \phi)} D^\mu \phi - \mu^2 \bar{\phi} \phi - \lambda (\bar{\phi} \phi)^2 \quad (2.6)$$

Finally, the Yukawa term describes the interactions between matter particles and the Higgs field is given by

$$\mathcal{L}_{Yukawa} = -g_l \overline{L}_l \phi l_R - g_d \overline{Q}_L \phi d_R - g_u \overline{Q}_L \phi u_R + (h.c.) \quad (2.7)$$

Fermion mass terms $m(\overline{\psi}_R \psi_L + \overline{\psi}_L \psi_R)$ and gauge boson mass term $s \frac{1}{2} m^2 B^\mu B_\mu$ would break the SU(2) invariance of the Lagrangian. Thus the Higgs mechanism is introduced to explain mass generation through the electroweak symmetry breaking mechanism.

2.6 Spontaneous Symmetry Breaking and the Higgs mechanism

The Higgs field ϕ is introduced to break the electroweak symmetry in vacuum (spontaneous symmetry breaking). As a doublet in SU(2), we denote ϕ as

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \quad (2.8)$$

We need to look at \mathcal{L}_{Higgs} to understand how the gauge bosons gain masses via the Higgs mechanism Model.

For $\mu^2 < 0$, the minimum of potential V in Equation 2.6 is at $\bar{\phi}\phi = -\frac{\mu^2}{2\lambda} = \frac{\nu^2}{2}$

Since the potential depends only on the combination $\bar{\phi}\phi$, we can arbitrarily choose the vacuum:

$$\phi_0 = \begin{pmatrix} 0 \\ v \end{pmatrix}. \quad (2.9)$$

We can expand Higgs field around vacuum as:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v+h \end{pmatrix}. \quad (2.10)$$

The kinematic term $T = \overline{(D_\mu \phi)} D^\mu \phi$ gives us the mass terms of bosons: $\frac{1}{2}(\frac{1}{2}\nu g_2)^2 W_\nu^+ W^{-\nu}$ and $\frac{1}{2}(\frac{1}{2}\nu \sqrt{g_1^2 + g_2^2})^2 Z_\nu Z^\nu$.

By plugging the simplified Higgs doublet into the Yukawa lagrangian term in Equation 2.7, we obtain the reduced form:

$$\mathcal{L}_{Yukawa} = - \sum_{fermions} m_f \overline{\psi}_f \psi_f - \sum_{fermions} \frac{m_f}{v} \overline{\psi}_f \psi_f h, \quad (2.11)$$

where $m_f = \frac{1}{\sqrt{2}} g_f v$ yields the mass of the fermions and the second term represents the interaction between fermions and the Higgs boson with the interaction Yukawa coupling proportional to the fermion mass.

2.7 Challenges

Despite being remarkably successful, the Standard Model has its limitations and leaves us with some significant questions.

First of all, while the Higgs boson gives mass to other particles via couplings, the Standard Model gives no prediction for the Higgs boson mass. The current measurement of the Higgs mass indicates that the electroweak scale is $\mathcal{O}(100\text{GeV})$, while the Planck scale is at $\mathcal{O}(10^{19}\text{GeV})$. This so-called hierarchy problem indicates that either unnatural fine-tuning exists, or there is some new physics which cancels out the divergent terms in the Higgs boson mass. !!!(more explanation needed.)

The source of the huge imbalance between matter and antimatter (baryon asymmetry) in the universe is unexplained. It's reasonable to assume that equal amounts of matter

and antimatter are created during the Big Bang. However, the universe is now dominated by matter while antimatter has essentially vanished. This requires Charge Parity(CP) violation. CP violation in the Standard Model is insufficient to account for the level of matter-antimatter asymmetry observed currently.

Finally, one striking evidence of physics beyond the Standard Model is dark matter. Given astrophysics theory and observations, 22.7% of the total mass-energy of the universe is dark matter, which is about five times more than visible matter. Although both direct and indirect evidence prove the existence of dark matter, dark matter still remains a mystery. So far, the properties of dark matter have only been probed via its gravitational interaction. The leading hypothesis suggests that most of the dark matter is in the form of stable, electrically neutral, massive particles, i.e. weakly Interacting Massive Particle(WIMB).

To answer these questions, there are several well-motivated models that predict dark matter interacting with Standard Model particles weakly, perhaps via a new mediator. If this is the case, then there is a good reason to search for dark matter production in high energy collisions, such as those provided by the Large Hadron Collider.

Chapter 3

Dark matter

As mentioned in Chapter 2, about one quarter of the total mass-energy in the universe is dark matter. No known particle that can account for the dark matter in the universe, and single-particle model is disfavored. Therefore, current models consider dark matter as a whole sector of new particles, instead of the existence of a single new type of particle.

This chapter is organized as follows: the astrophysical measurements establishing of dark matter and its particle physics explanation are discussed in Section 3.1. In Section 3.2, dark matter as a microscopic particle, Weakly Interacting Massive Particle, is described. Finally, Section 3.3 outlines how dark matter particles might be produced by particle colliders and what signatures they might generate.

3.1 Dark matter: From astrophysics to particle physics

The concept of dark matter originates from physical observations. Multiple astrophysical observations at different distance scales prove the existence of invisible matter in the universe. The concept of dark matter was formally brought up by Jacobus Kapteyn in the studies of the velocity distribution of stars in nearby galaxies[93]. These studies showed that the amount of visible matter from stars and interstellar gas near the solar system was not sufficient to explain the motions of the stars perpendicular to the Milky Way disk. The visible matter doesn't provide enough gravitational force attraction which suggests the existence of the unobserved matter. Another proof on galactic scales is the motion of the

Coma galaxy cluster observations by Fritz Zwicky in 1933[123], which suggested a large contribution to the gravitational forces that is not visible.

On galactic scales, the most convincing evidence is the direct observation of the rotation curves of galaxies, which shows the orbital velocities of stars and gas inside a galaxy as a function of their distance from the galactic center. According to the Newtonian theory of gravity, the rotation velocity, $v(r) \propto \frac{1}{\sqrt{r}}$, which is in contrast with observed curve in Fig 3.1. It is the discrepancy between the observed and expected velocities that has led to the belief that some form of dark matter must exist. And this suggests the existence of a dark (invisible) halo with $M(r) \propto r$ and $\rho \propto \frac{1}{r^2}$.

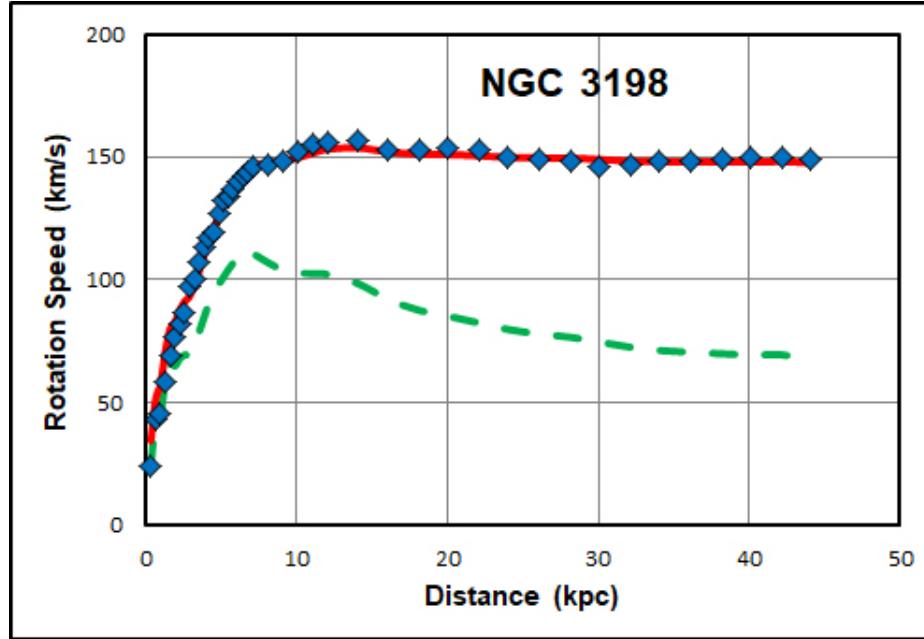


Figure 3.1: The rotation curve for spiral galaxy NGC 3198. The blue diamonds are the observations and reveal the almost flat-like nature of the curve in the outer regions of the galaxy. The dashed green line is the curve for Newtonian gravity. It shows that the rotational velocity should decrease with distance from the galaxy center. The solid red line through the data points is the curve obtained by assuming a simple Gaussian energy scale variation and a simple Gaussian density distribution for the galaxy.

The Cosmic Microwave Background (CMB), is electromagnetic radiation remnant from an early stage of the universe, also known as “relic radiation”. On the cosmological scale,

not only does CMB show evidence of dark matter, in the context of specific models it also quantifies the amount of dark matter in the Universe. Experiments measure the power spectrum of the CMB. Now the Lambda-CDM model best fits the observation data, giving values for the density of baryonic matter, dark matter, and Dark Energy.

Hot dark matter (HDM) theory[122] was established in 1980 by Zeldovich's team. HDM assumes that the light neutrinos make up the majority of the dark matter. It is natural to assume that the dark matter is the weakly coupled particle that already exists in the Standard Model. However, the conventional neutrino-dominated picture is ruled out by early universe simulation studies in 1983[119]. Afterwards, scientists realized that only with a slow-paced particle, so-called “cold” dark matter, the diffusion of small scale fluctuation can be prevented. The early universe structure can be formed on all scales is consistent with this astrophysical observations. Therefore, the cold dark matter theory (CDM theory)[104] was established in 1983 to explain the cosmic microwave background observation result. Nowadays, as an important part of the standard cosmological model (Lambda-CDM theory), the concept of cold dark matter is widely accepted.

3.2 Weakly Interacting Massive Particles

Although the existence of cold dark matter around galaxies and clusters is supported by cosmological observation, scientists still have poor knowledge regarding to the nature of cold dark matter is. Under the assumption it is a particle, the observation of cold dark matter indicates the existence of weakly interacting particles. The density fluctuations and the small scale fluctuations indicate the mass of cold dark matter candidate can not be small. Therefore, WIMPs - weakly interacting massive particles, become one of the best candidate particles to characterize the feature of cold dark matter.

WIMP masses range from a few GeV to $\mathcal{O}(\text{TeV})$, to match the observed relic density from the CMB analysis. Both popular beyond SM models Supersymmetry (SUSY) and Extra-dimensions, there are WIMP candidates of DM particles. The electrically neutral lightest supersymmetric particle (LSP) predicted in an R-parity conserving scenario of supersymmetric extension of the Standard Model is an ideal candidates for dark matter.

Among all possible choices, the most promising one is the lightest neutralino, which is the lightest state of the mixtures of neutral electroweak gauginos and the neutral higgsinos[72]. Similarly to the phenomenology of SUSY, in universal extra dimensions (UED) models, each SM particle is accompanied by a partner particle at the first Kaluza-Klein (KK) mode level. Also, UED models preserve a geometric parity (KK parity). The lightest partner state (which in UED is the partner of the $U(1)_Y$ gauge boson) is a good dark matter candidate which can yield the observed dark matter relic density[113].

3.3 Searches for dark matter in collider experiments

Dark matter is observed through its gravitational interactions in the universe. To explore its potential particle properties, several complementary detection methods are used[98]:

- **Direct detection:** Direct detection focuses on recoils of nuclei with few keV induced by interactions with dark matter particles propagating through the solar system.
- **Indirect detection:** Indirect detection searches for astrophysical dark matter signals from annihilation products, usually using cosmic rays.
- **Production at colliders:** Production at colliders, as its name suggests, looks for WIMPs in collision data. WIMPs can be detected indirectly as missing energy and momentum that escapes the detectors.

In the rest of this thesis, only the “Production at collider” approach will be considered. Several beyond the Standard Model models that have dark matter candidates will be described in the rest of this section. These models are the search target and will be used to interpret the collision data in statistical manner.

3.3.1 Effective field theory

Before we dive into specific particle physics models, some generic principles need to be illustrated. The first concept is effective field theory. An effective field theory is an approximation for an underlying physical theory, for example quantum field theory. Assuming

that all the current physical observations can be explained under one single physics theory. However, since this theory of everything is supposed to explain the experiment at all energy scales, the physical observables are always resulted in infinite under the quantum field theory framework. Furthermore, the important features of the known physics, namely the Standard Model, are not guaranteed to be extendable with higher energy scale. Therefore, the ultraviolet energy cut-off is needed for quantum field theory based particle physics model to guarantee the model itself renormalizable. The effective field theory[106] is designed to resolve this issue.

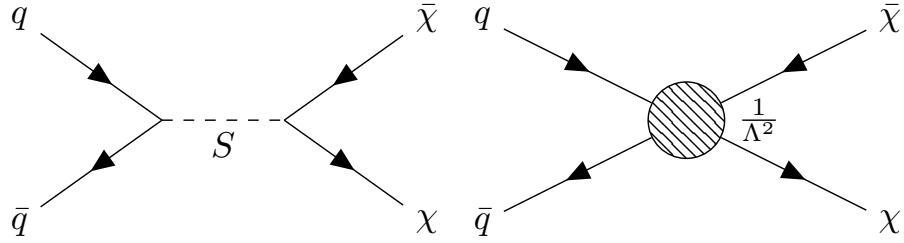


Figure 3.2: Left: an s-channel Feynman diagram of the interaction between quarks and dark matter candidates via a scalar mediator. Right: the corresponding effective field theory contact interaction.

In the context of particle physics, effective field theories are simplifications of the complete quantum field theories where the mass of the interaction mediator is large enough with respect to the momentum exchange of a physical process. Therefore, an s-channel process can be replaced by a contact operator as demonstrated in Fig 3.2. In Fig 3.2, a new scalar mediator, labeled S , is introduced to couple with both the Standard Model quarks and a dark matter candidate. The Lagrangian of this simple model is 3.1:

$$L = \frac{1}{2}M^2S^2 - g_q q\bar{q}S - g_\chi \chi\bar{\chi}S, \quad (3.1)$$

where M is the mass of newly introduced scalar, g_χ is the coupling strength between the scalar mediator and dark matter candidates, and g_q is the coupling constant between the scalar mediator and the Standard Model quarks. The knowledge of the quantum field theory tells us that the s-channel cross section is proportional to the factor $\frac{1}{Q_{tr}^2 - M^2}$, where

Q_{tr} is the momentum transfer of the process. Using Taylor expansion, this factor can be expanded in powers of $\frac{Q_{tr}}{M}$:

$$\frac{1}{Q_{tr}^2 - M^2} = -\frac{1}{M^2}(1 + \frac{Q_{tr}^2}{M^2} + \mathcal{O}(\frac{Q_{tr}^4}{M^4})) \quad (3.2)$$

If $Q_{tr} \ll M$, the propagator term can be simplified as $-\frac{1}{M^2}$. Therefore, the interaction vertices of the s-channel collapse into a contact point: $O_s = \frac{g_q g_\chi}{M^2} \chi \bar{\chi} q \bar{q}$, where $\Lambda = \frac{g_q g_\chi}{M^2}$ can be viewed as the mass scale of the effective field theory. Given the condition that $Q_{tr} \ll M$, the constraint of the new energy scale can be derived as 3.3:

$$\Lambda \gg \frac{Q_{tr}}{\sqrt{g_q g_\chi}} \quad (3.3)$$

Therefore, the calculation can be simplified for any process that satisfies the condition in Equation 3.3.

3.3.2 The simplified model

Although the energy scale can be limited with the effective field theory, the simulation can still be complex due to the model structure itself. For example, in the simplest supersymmetry scenario, the Minimal Supersymmetric Standard Model, there are more than 100 free parameters. It is almost impossible to explore them all within a search for one signature. Therefore, simplified models[35] are used.

A simplified model describes the particles relevant to the search process only. With the help of simplified models, the complex interaction Feynman diagram group can be trimmed into one or several tree-level diagrams that describe the signal of interest. As a result, the model calculation is simplified while its characteristics are still preserved.

With the simplification of effective field theory and simplified model, several models are introduced in the rest of this section. The physical events from these models are used as search signal for the rest of this thesis.

3.3.3 Z' and Two-Higgs-doublet models

A Z' model[86] is one of the simplest Standard Model extensions that can have a dark matter candidate. As mentioned in Section 3.2, the dark matter candidate as fundamental

particle needs to be both weakly interacting and massive. Therefore, a natural way to introduce a weakly couples particle into the Standard Model is to add a new vector boson that coupled with WIMPs. This newly added particle is a massive, neutral boson. The simplest gauge structure for this beyond the Standard Model is a U(1) extension. However, this newly added U(1) can be a derivative from more a complex symmetry, like SU(2), or E_6 . This simplest extension of the Standard Model is called a Z' model.

The Z' boson can couple with both the Standard Model quarks and dark matter candidates. A typical Feynman diagram is shown in Fig 3.3. However, the mass of the dark matter candidate is highly constrained by the Higgs boson mass, therefore, it is less possible to find a massive dark matter candidate within the current experiment constraints in Z' model. But mixing the Z' model with other beyond Standard Model models we can gain more room in parameter space. One of the examples is a $Z' - 2HDM$ model, a mixture of a Z' model and a Two-Higgs-doublet model.

Before describing the mixed model, let us summarize the features of Two-Higgs-doublet models[48], abbreviated as 2HDM. 2HDM are multiple Higgs doublet extensions of Standard Model. 2HDM models are inspired by supersymmetric models[99] and Axion model[105]. Therefore, 2HDMs can not only provide a dark matter candidate, but also bring enough CP violation to support baryon asymmetry universe.

As mentioned, Z' and Two-Higgs-doublet models can be combined into mixed models to provide much richer phenomenological predictions. The Higgs-like field in the 2HDM is introduced into the Z' model as a dark matter mediator. As a result, the mixture $Z' + 2HDM$ model[46] can provide sufficiently massive dark matter candidate while respecting experimental data constraints with its larger parameter space.

With the discovery of the Higgs boson, the study of Higgs properties becomes the frontier of particle physics. The Higgs boson is widely used as a physics object, like the top quark after 1995, in new physics searches. Given our limited knowledge of the Higgs boson, it is quite possible that is couples dark matter sector in some way. Among all the Higgs decay channels, Higgs $\rightarrow b\bar{b}$ is the one with the largest branching ratio. Therefore, it is natural for experimentalists to search dark matter candidates produced in association with the Higgs boson in the channel with large missing transverse momentum and Higgs $\rightarrow b\bar{b}$ decay as

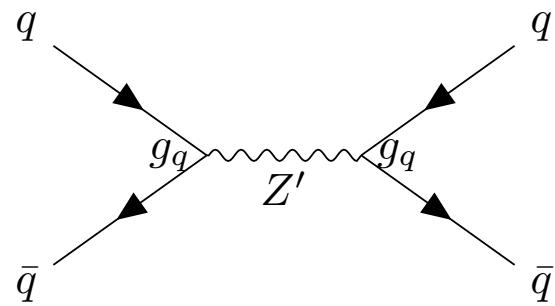


Figure 3.3: Feynman diagram of the interaction between Standard Model quarks and a Z' boson.

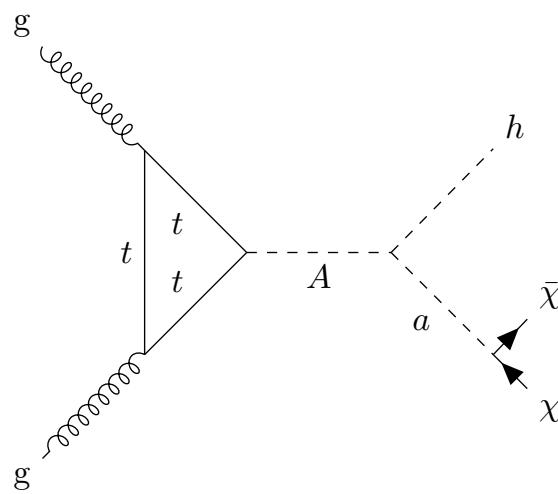


Figure 3.4: The Feynman diagram of the interaction between the Standard Model quarks and Z' boson.

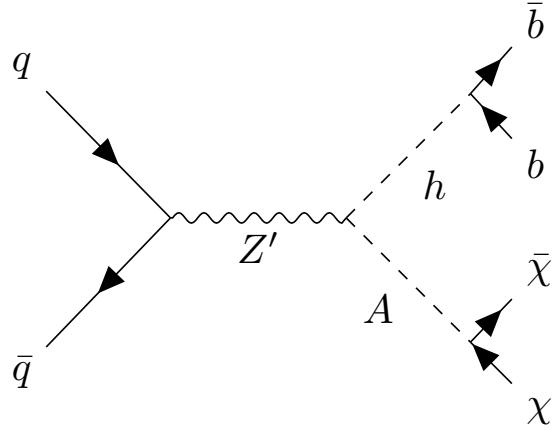


Figure 3.5: The Feynman diagram of $b\bar{b}$ decayed Higgs channel in $Z' + 2HDM$.

its signature. The Feynman diagram of $H \rightarrow b\bar{b}$ channel is shown in Fig 3.4 for a Two-Higgs-doublet model, and in Fig 3.5 for the $Z' + 2HDM$. Both are studied as the signal in the rest of this thesis.

Part III

The LHC and ATLAS experiment

Chapter 4

The LHC

The Large Hadron Collider, abbreviated as LHC, is the most powerful proton-proton collider, located at CERN, Geneva, Switzerland. The LHC is installed in a 26.7 km tunnel that was built for last generation lepton collider between 1984 and 1989. The LHC tunnel, which is located from 45m to 170m below the surface, is composed of 8 straight sections and 8 arcs. Therefore, the LHC can be viewed as 8 octants. Each octant has an access point, which includes an elevator from surface to underground. Half of the LHC points are hosting the detector systems currently: ATLAS[20] at Point 1, ALICE[28] at Point 2, CMS[59] at Point 5 and LHCb[34] at Point 8. The other 4 points are designed for LHC operation purposes.

Rome wasn't built in one day, neither was the LHC. The LHC was built on the infrastructure of the previous generation of colliders located at CERN. The LHC is the current frontier of the evolution chain shown in Fig 4.1: from Proton Synchrotron (1954)[79], Super Proton Synchrotron (1976)[69], Large Electron-Positron Collider (1984)[1][2], to Large Hadron Collider (2008)[49][50]. Both infrastructure and technology are utilized in an economic manner to support the most powerful collider in the world, the LHC. The proton beam can be accelerated at 7 TeV by the current LHC setup. Although the LHC has reached the frontier of the high energy of human experiment, it is not the end of the CERN collider evolution chain. Both the LHC upgrade plan[37] and next generation collider design[29] have been proposed to sustain the prosperity of the CERN collider family.

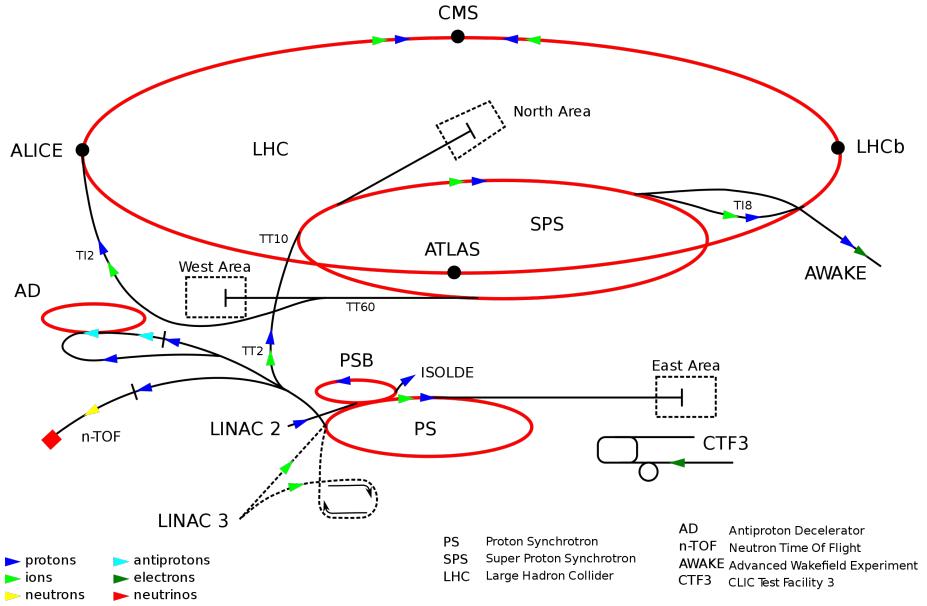


Figure 4.1: The LHC full injection chain.

4.1 The LHC performance

The aim of the LHC is to provide a stable, large amount of proton-proton collision events at the high energy frontier. Therefore, the LHC performance can be viewed as two parameters: the center-of-mass energy and integrated luminosity.

4.1.1 Beam energy

The center-of-mass energy is defined as the proton pair kinetic energy in the center-of-mass frame. Currently, the LHC is running at 13 TeV center-of-mass energy, with 6.5 TeV for each beam. Protons are accelerated and bended by electric field strips and 1232 superconducting dipole magnets around the LHC ring. Higher beam energy requires a stronger magnetic field, which requires a higher electric current flowing in the magnets' superconducting coils.

4.1.2 Luminosity

The integrated luminosity, defined as the number of collision events produced by the LHC inside the particle detector, reflects the number of collisions delivered by the LHC. The

integrated luminosity is determined by the LHC instantaneous luminosity, the beam cross-section and total LHC collision time.

The LHC luminosity can be calculated by Eq 4.1, where N_b is the number of protons per bunch, n_b is number of filled bunches per beam, f_r is the frequency of the beam circling the ring, γ_r is the relativistic gamma factor of the protons, ϵ_n is the normalized transverse beam emittance, β^* is the measure of beam width in the longitudinal direction, F is a geometric factor which accounts for the non-zero crossing angle between two beams:

$$L = \frac{N_b^2 n_b f_r \gamma_r 4\pi \epsilon_n \beta^*}{F}, \quad (4.1)$$

The LHC can be running on different luminosity modes, while the detector system needs to be tuned to adapt the various LHC luminosity. The last parameter, total LHC collision time, is an operational parameter that depends on the collider and detector operation teams. The LHC delivered around 156 fb^{-1} data during the Run 2 period as shown in Fig ??.

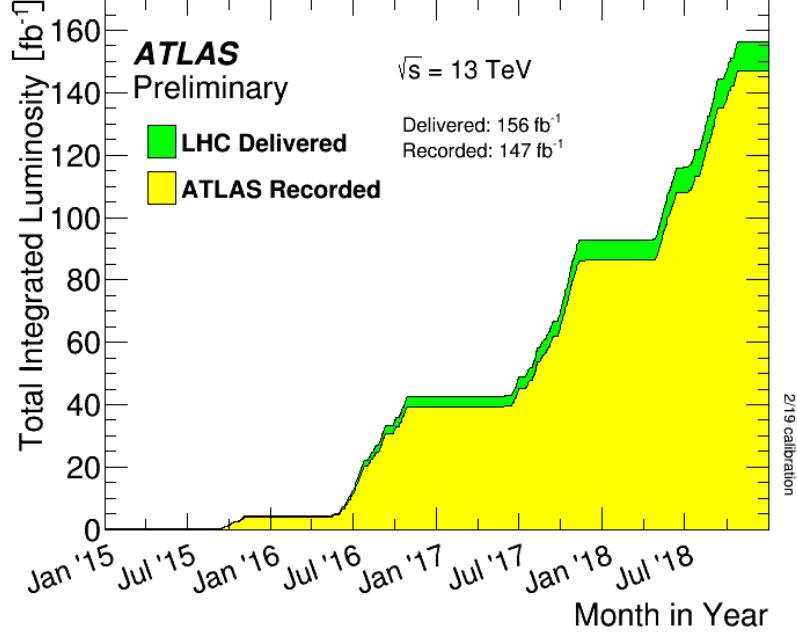


Figure 4.2: Cumulative luminosity versus time delivered to ATLAS (green), recorded by ATLAS (yellow) during stable beams for pp collisions at 13 TeV centre-of-mass energy in 2015-2018.

4.2 The LHC operation

The LHC beam modes are explained in Table 4.1. The beam modes describe the status of the beam related activities in the LHC. A successful beam injection starts with BEAM SETUP mode, which means beam circulating inside the Super Proton Synchrotron and waiting to be injected into the LHC. Then, a probe beam will be injected into the LHC ring as a trial, which is INJECTION PROBE BEAM mode. After that, INJECTION SETUP BEAM comes to measure the beam properties. After all the previous preparation, the physics injection finally comes into the LHC, which is called INJECTION PHYSICS BEAM. Normally, the detector operators need to get the detector system prepared when an injection physics beam happens. Then, the proton beam will be accelerated in the LHC with the PRERAMP and RAMP mode. The LHC system operators work on final machine adjustment at the FLAT TOP mode, and the beam impact parameter is reduced during the SQUEEZE mode. Finally, the beam is aligned in the ADJUST mode and STABLE BEAM mode will happen. Only the data taken during the STABLE BEAM mode will be used in physics analysis.

Table 4.1: LHC beam modes.

Mode Name	Description
SETUP	Beam in transfer line, but not in the ring
ABORT	Recovery mode following beam drop
INJECTION PROBE BEAM	Ring is injected with test beam for safe circulating
INJECTION SETUP BEAM	Beam measurement going on after probe beam but before injection physics beam
INJECTION PHYSICS BEAM	Beam for physics is injected in the ring
PRERAMP	Injection done, prepare for ramp
RAMP	Ramp up the beam energy
FLAT TOP	Ramp done, pre-squeeze checks
SQUEEZE	Squeezing the beam size
ADJUST	Preparing for collision or after collision
STABLE BEAMS	Stable collision, detector should taking data
UNSTABLE BEAMS	Unstable beam because of sudden beam degradation
BEAM DUMP WARNING	Beam dump warning in case of emergency beam dump
BEAM DUMP	End of physics collision
RAMP DOWN	Ramp down beam energy after programmed dump
CYCLING	Pre-cycle before injection following access, recovery, etc
NO BEAM	No beams exist

Chapter 5

The ATLAS experiment

ATLAS (A Toroidal LHC ApparatuS) is one of the major experiments, located at Point 1 of the Large Hadron Collider (LHC) at CERN [20]. It is a general-purpose particle physics experiment, which is designed to exploit the huge range of physics opportunities that the LHC provides. Located at 92 m below ground, the ATLAS detector has a cylindrical shape with a length of 46 m, a diameter of 25 m and a weight of over 7000 tons. The detector was built and is operated by a collaboration involving roughly 3,000 physicists from over 175 institutions in 38 countries [63].

A right-handed Cartesian coordinate system is used in this thesis: the coordinate origin is at the geometric center of the ATLAS detector, with a z-axis as the direction of beam pipe.

The x-y plane is perpendicular to the z-axis, with x pointing from origin point to the center of the LHC ring and y pointing upward. Therefore, polar angle θ is measured with respect to z-axis and the azimuthal angle ϕ is measured around the beam axis.

The pseudorapidity is defined as $\eta = -\ln \tan(\frac{\theta}{2})$. The distance ΔR in the pseudorapidity-azimuthal plane is defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$. Transverse momentum p_T and missing transverse momentum E_T^{miss} are defined in the x-y plane. A cut-away view of ATLAS detector is shown in Fig 5.1.

The four major components of the ATLAS detector are the Inner Detector, the Calorimeter, the Muon Spectrometer, and the Magnet System. Integrated with the detector components are: the Trigger and Data Acquisition System, a specialized multi-level computing

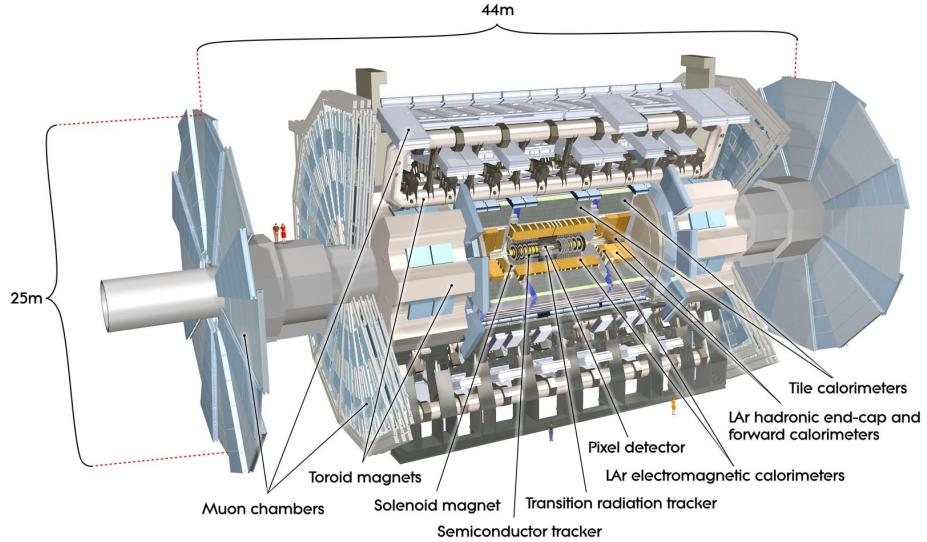


Figure 5.1: Cut-away view of the ATLAS detector

system, which selects physics events with distinguishing characteristics; and the Computing System, which develops and improves computing software used to store, process and analyze vast amounts of collision data at 130 computing centers worldwide[62].

Different particles deposit energy in different layers as shown in Fig 5.2, where inner Detectors measure $\frac{q}{p_T}$ (charge over transverse momentum), the measurement of charged particles, Calorimeters measure energy of electromagnetically and hadronically interacting particles, and the Muon Spectrometer serves as an tracker for muons.

This chapter is intended as a brief introduction to the ATLAS sub-systems, including Inner Detectors in Section 5.1, calorimeters in Section 6.5.1, muon system in Section 5.3, Forward detectors in Section 5.4, and trigger and data acquisition system in Section 5.5.

5.1 Inner detector

The Inner Detector tracker is important for track reconstruction as well as both primary and secondary vertex measurements for charged tracks in the pseudorapidity range of $\eta < 2.5$. The Inner Detector tracker is contained within a cylindrical envelope with a length of ± 3512 millimeter and a radius of 1150 millimeter.

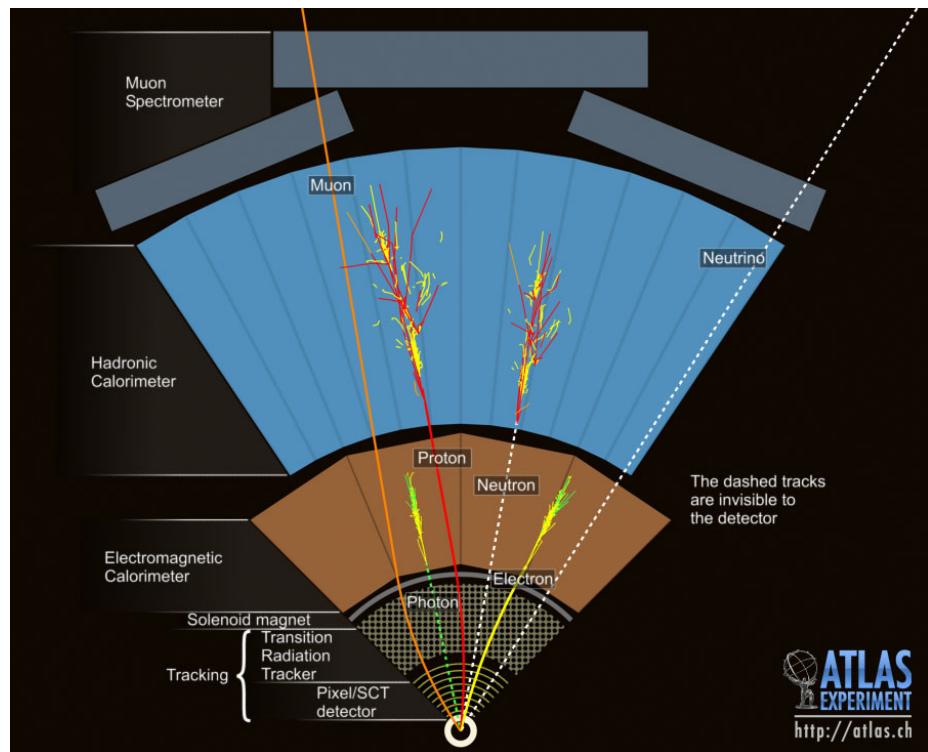


Figure 5.2: An image of the ATLAS detector, showing the different layers and the interactions of different particle types through the layers.

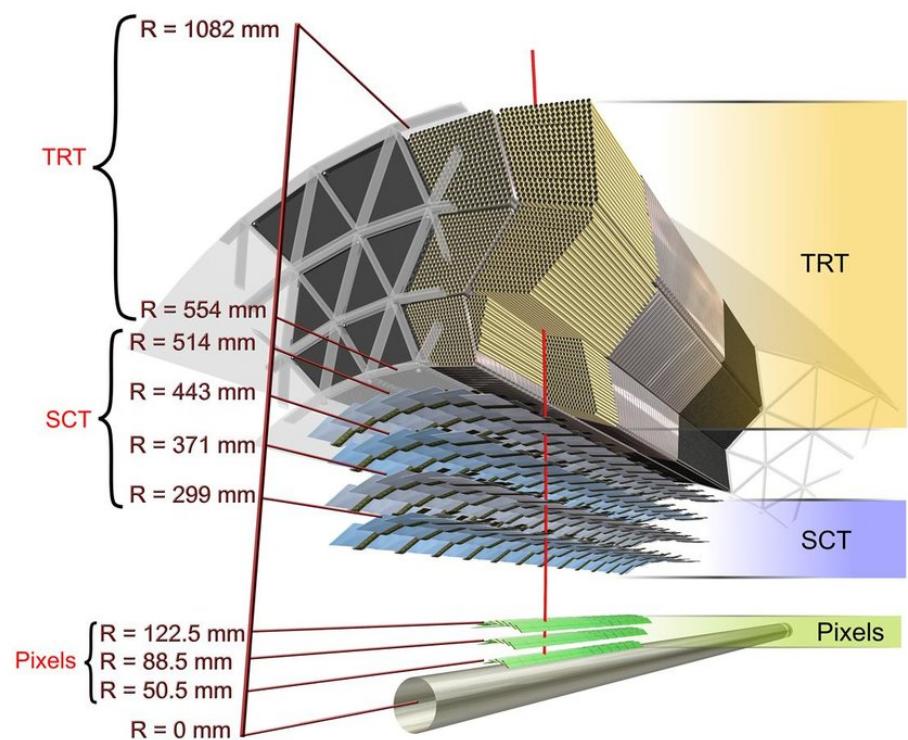
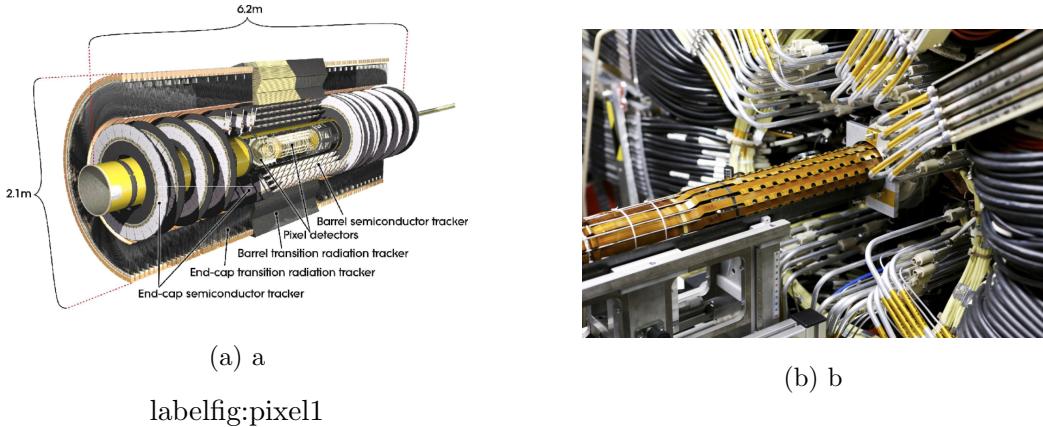


Figure 5.3: Cut-away view of the ATLAS Inner Detector.

As shown in Fig 5.3, the Inner Detector tracker comprises three detector types dedicated to tracking. Moving from inside out we find the Silicon Pixel Detector, the SemiConductor Tracker (SCT), and the Transition Radiation Tracker (TRT). An extra pixel detector layer (IBL)[56] was inserted before the Run 2 and improves the identification of b-jets [13]. The Pixel system provides a coverage of $|\eta| < 2.5$. The SCT system consists of four barrel double layers and 18 end-cap layers (9 on each end) [24], and provides a coverage of $|\eta| < 2.5$. The TRT consists of 70 barrel layers and 280 end-cap layers (140 on each end), and provides a coverage of $|\eta| < 2.0$ [24].

5.1.1 Pixel detector



labelfig:pixel1

Figure 5.4: A schematic view of the active region of the Pixel detector consisting of barrel and end-cap layers (??) and the IBL detector before the insertion (5.4b)

The Pixel detector is the innermost element of the Inner Detector [88]. With the fine granularity of the pixel sensors, the pixel detector is designed to provide the identification and reconstruction of secondary vertices from the long-lived particles. It provides high resolution for primary vertices reconstruction to suppress pile-ups due to the increase of luminosity for LHC. A schematic view of the active region of the pixel detector consisting of a barrel and end-cap layers can be found in Fig ??.

The IBL makes it possible for the pixel detector to have further improved resolution and a picture of the IBL being inserted into the Pixel detector is shown in Fig 5.4b. The pixel sensor pitch of the IBL has a minimum size in $R - \phi \times z$ of $50 \times 250 \mu\text{m}^2$ compared to other

pixel detector layers with a size of $50 \times 400 \mu\text{m}^2$. The IBL provides an intrinsic spatial resolution for hits of $14 \mu\text{m}$ in the $R - \phi$ plane, and $72 \mu\text{m}$ in the z-direction, compared to an intrinsic spatial resolution for hits of $14 \mu\text{m}$ in the $R - \phi$ plane, and $115 \mu\text{m}$ in the z-direction of the three outer pixel barrel layers.

The Pixel detector's ability to associate tracks correctly to secondary vertices is essential to tagging algorithms of the b -hadrons decays which are important for this thesis, as well as other non-prompt decays.

5.1.2 Semiconductor Tracker (SCT)

The SCT [6] is designed to provide a good measurement of momentum, impact parameter, and vertexes. It includes 4 cylindrical barrel layers and 18 planar end-cap disks, covered by 61 m^2 of silicon detectors and 6.2 million readout channels. The spatial resolution is about $16 \mu\text{m}$ in the $R - \phi$ plane and $580 \mu\text{m}$ in the z-direction.

5.1.3 Transition Radiation Tracker (TRT)

The TRT[30] is the outermost part of the Inner Detector. The TRT is composed of thin proportional chambers either in the form of straws embedded in fibers or with foils. The 4-mm diameter Kapton straw drift tubes, which can be operated at 6 to 18 MHz at the barrel while 7 to 19 MHz at the end-caps, are filled with a mixture gas. The mixture gas is composed with 70% of Xe, 20% of CF_4 and 10% of CO_2 .

The TRT is designed as a dual-purpose detector: quasi-continuous tracking and particle identification. The quasi-continuous track reconstruction is achieved by the large amount of hits provided by 370000 straws in TRT. The particle is identified by its transition radiation product when it is traveling across the material boundaries (CO_2 gas and polypropene/polyethylene fibers).

5.2 Calorimeter

The ATLAS calorimeters [3] are designed to measure the energy of the outgoing particles. Particles passing the calorimeters will initiate particle showers, either electromagnetic (EM)

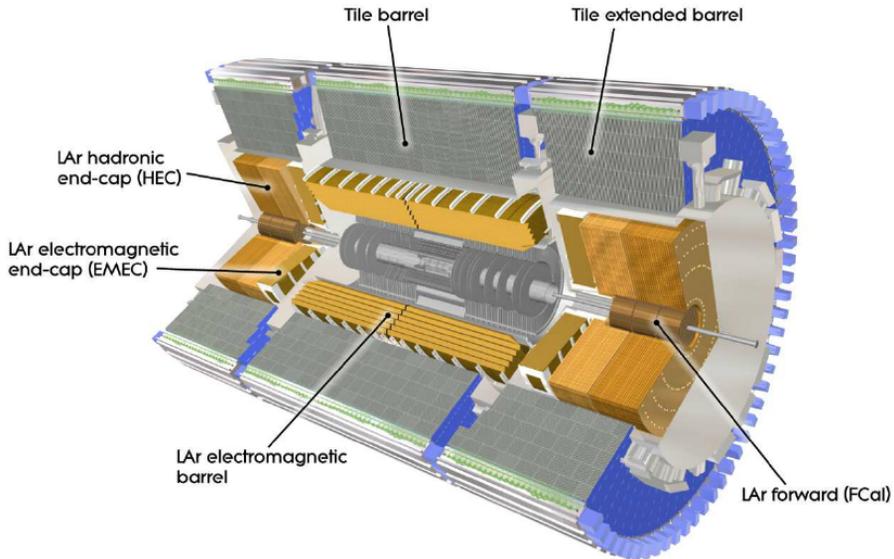


Figure 5.5: Cut-away view of ATLAS calorimeter system

showers (consisting of electrons and photons) or hadronic showers (consisting of pions , protons, ...) in layers of passive material made of dense materials. Such showers grow as the outgoing particles continue to interact with the absorber until their energies are lower than the critical energy where the ionization becomes dominant over radiation or interactions. The ATLAS calorimeters are sampling calorimeters with layers where the final state particles can induce ionization or scintillation in the active layers to record the energy of the showering particles.

There are two types of calorimeters in ATLAS as shown in Fig 5.5: the EM calorimeter, a lead/liquid-argon sampling calorimeter, covering the pseudorapidity range of $|\eta| < 1.475$ in the barrel region and $1.375 < |\eta| < 3.2$ in the end-cap region; the hadronic calorimeter, consisting of the tile calorimeter covering $|\eta| < 1.7$, the liquid-argon hadronic end-cap calorimeter (HEC) covering $1.5 < |\eta| < 3.2$; and the liquid-argon forward calorimeter (FCal) covering $3.1 < |\eta| < 4.9$.

To contain all particles showers in the calorimeters and thus avoiding particles (except muons and neutrinos) penetrating into the muon spectrometer, the materials and thickness of each layers are optimized [64]. The thickness of the EM calorimeter is greater than $22 X_0$

in the barrel and 24 X_0 in the end-cap region, where the radiation length of the material X_0 represents the thickness of material that reduces the mean energy of an electron by a factor e . The thickness of the hadronic calorimeter is approximately 9.7λ in the barrel and 10λ in the end-cap regions, where λ is the nuclear interaction length.

Ideally, if all showers are counted, the measured energy would be proportional to the signal strength and thus the error is proportional to \sqrt{E} :

$$\frac{\sigma(E)}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c , \quad (5.1)$$

where a is stochastic term, which is a combination of the intrinsic statistical shower fluctuation, signal quantum fluctuation and sampling fluctuation, b is a noise term from readout electronics noise, radioactivity and pile-up fluctuations, and c is a constant term representing any inhomogeneities and imperfections in calorimeter construction, and non-linearity of readout electronics [45]. From Equation 5.1, we could the energy resolution improves with increasing energy in general.

5.2.1 Liquid Argon Calorimeter

The LAr EM Calorimeter uses lead as its passive material, and the active material is liquid argon (LAr). It consists of the barrel EM calorimeter (EMB) and one end-cap (EMEC) on each side, with the inner wheels (IW) covering $1.375 < |\eta| < 2.5$ and outer wheels (OW) covering $2.5 < |\eta| < 3.2$.

As shown in Fig 5.6a, the LAr calorimeter uses a novel accordion geometry to avoid gaps at boundaries. It comprises accordion-shaped copper-kapton electrodes positioned between lead absorber plates and kept in position by honeycomb spacers while the system is immersed in LAr. The accordion geometry provides complete ϕ coverage without azimuthal cracks.

For most of the EM calorimeter, EMB and EMEC-OW, each module has three layers in depth with different granularities, as can be seen in Fig 5.6b, while each EMEC-IW module has only two layers. Table 1.3 of Ref [64] shows the granularity of the EM calorimeter for different ranges. The fine segmentation of the first layers helps distinguishing photons from π_0 mesons decaying to two photons, as well as providing flight direction of neutral particles. The second layers has a relatively coarser granularity but it is quite thick so the largest

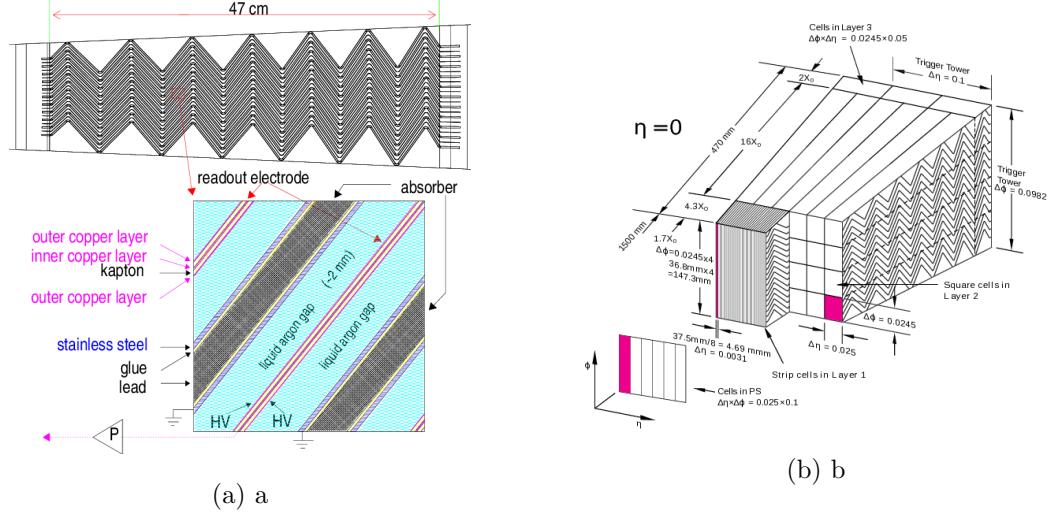


Figure 5.6: Accordion structure of the barrel. The left figure is a view of a small sector of the barrel calorimeter in a plane transverse to the LHC beams. The right figure is a diagram of a sketch of barrel module showing both accordion structure and granularity in $\eta - \phi$ of the cells on each of layer

fraction of the energy is deposited there, and only a small tail of the EM shower penetrates in the last layer which measures the remaining energy of the most energetic particles.

The principle of operation of the HEC and FCal is similar to LAr, but they use copper and tungsten as passive material and the design details are different and vary with position.

5.2.2 Tile Calorimeter

Located outside of the LAr EM calorimeter, the tile calorimeter [4] is a sampling calorimeter using steel as the absorber and scintillator as the active medium. Its barrel covers the region $|\eta| < 1.0$, and its two extended barrels covers the range $0.8 < |\eta| < 1.7$.

The barrel and extended barrels are divided azimuthally into 64 modules with a span $\Delta\phi = 0.1$. It is segmented in depth in three layers, approximately 1.5λ , 4.1λ and 1.8λ thick for the barrel and 1.5λ , 2.6λ , and 3.3λ for the extended barrel.

As shown in Fig 5.7, scintillator tiles are oriented radially, with wavelength-shifting readout fiber connected at the tile edge. Readout fibers are then grouped together and are connected to the readout photomultiplier.

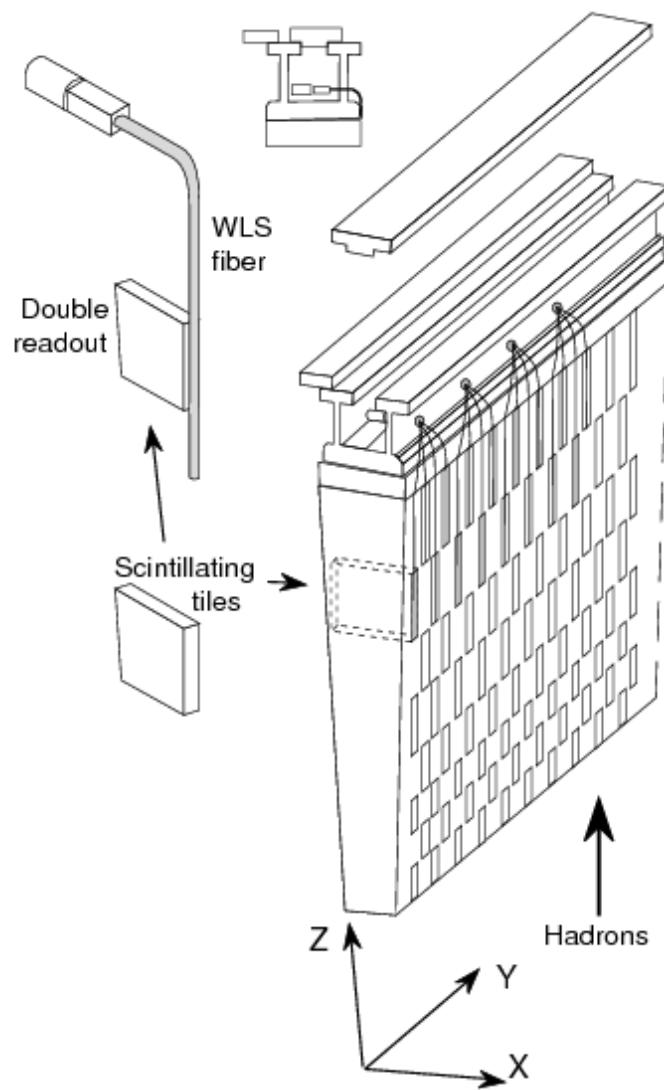


Figure 5.7: Illustration of a Tile calorimeter module.

5.3 Muon Spectrometer (MS)

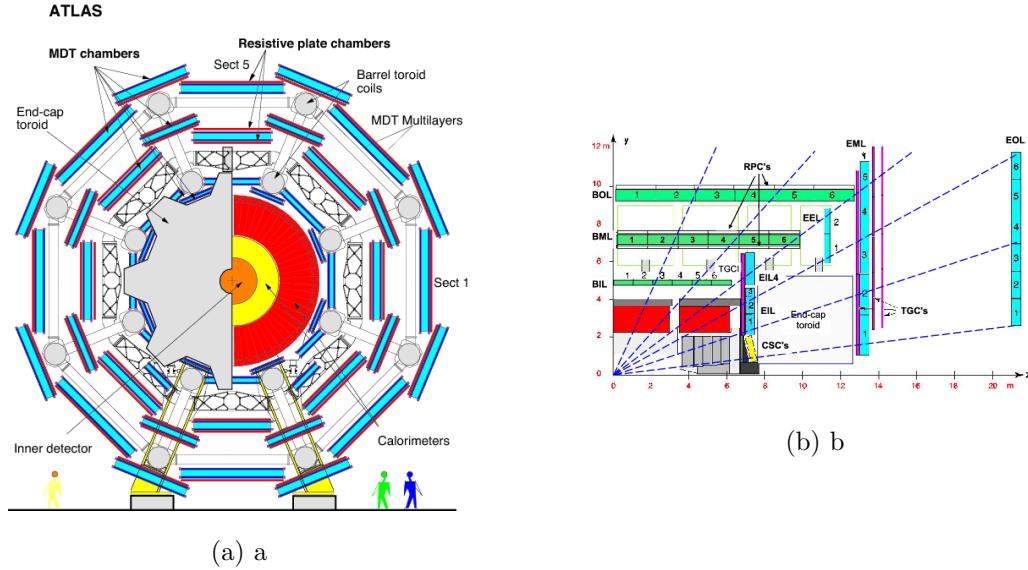


Figure 5.8: Geometric layout of muon sub-detectors in barrel (5.8a) and end-cap (5.8b) region.

The MS [5] is designed to measure the trajectory of transversing muons, as well as to provide muon trigger signals using separate sets of detection chambers. The MS is immersed in a toroidal magnetic field of about 0.5 T and 1 T in the barrel and end-cap regions, respectively. The MS is designed to measure muons standalone in a wide range of 3 GeV up to about 3 TeV. Being located farthest from interaction point, cells are relatively large due to low occupancy and the radiation level is typically smaller in the muon system. Also, it should be able to perform standalone measurement of high-momentum muon [92]. The targeted p_T resolution is 10% for 1 TeV muon tracks, which is a sagitta along the z-axis of about 500 μm with a resolution of about 50 μm .

As illustrated in Fig 5.8, the MS is composed of the precision tracking detectors, Monitored Drift-tube Chambers (MDT) and Cathode Strip Chambers (CSC), as well as the triggering detectors, Resistive Plate Chambers (RPC) in the barrel region and Thin Gap Chambers (TGC) in the end-cap region.

MDTs have a spatial resolution of about 80 μm and cover $|\eta| < 2.7$, except in the innermost end-cap layer, where it is $|\eta| < 2.0$. The CSCs with higher rate capability are

installed on the forward region of $2.0 < |\eta| < 2.7$ in the innermost end-cap layer. They have a resolution of about $60\mu m$, but since the cathode segmentation is coarser, the resolution is 5 mm in the non-bending direction. RPCs in the barrel region covering $|\eta| < 1.05$ and TGCs in the end-cap region covering $1.05 < |\eta| < 2.4$ provide a very fast response to muon hits.

Muon reconstruction and identification algorithms rely on information from both the Inner Detector and the MS.

5.4 Forward Detectors

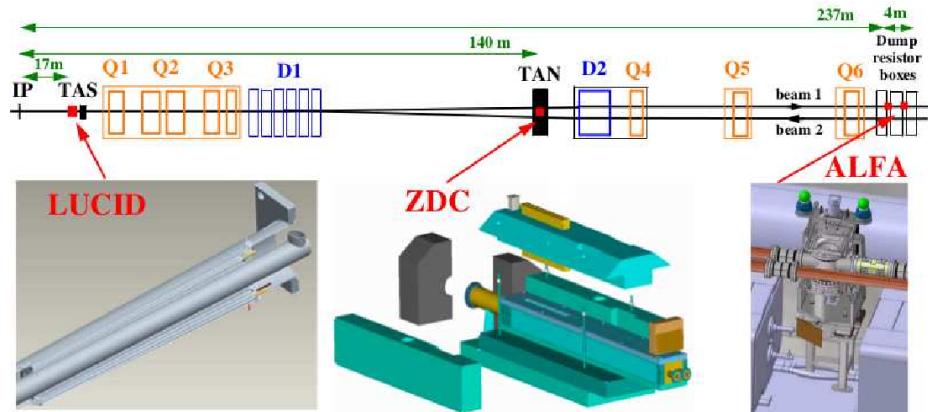


Figure 5.9: Cut-away view of the ATLAS Inner Detector.

Shown in Fig 5.9, LUCID (Luminosity measurement using Cerenkov Integrating Detector), ALFA (Absolute Luminosity For ATLAS) and ZDC (Zero-Degree Calorimeter) are together called Forward Detectors.

The LUCID and ALFA are dedicated to luminosity measurement at ATLAS experiment. The LUCID, situated at 17 meter from the interaction point on both side, is a luminosity monitor that able to monitor online bunch-by-bunch luminosity and provide absolute luminosity after calibration. The ALFA, housed in Roman pots at 240m, is designed to measure the total cross section. It will provide simultaneously total cross section and luminosity with an uncertainty of 2-3% in the LHC working range. The ZDC, located at 140 meter from the interaction point, although it is designed to study heavy ions and proton-proton

physics, it can also be used as luminosity monitor to tune the LHC parameters during the first day machine adjustment (for example, Van der Meer scan).

5.5 Trigger and Data Acquisition

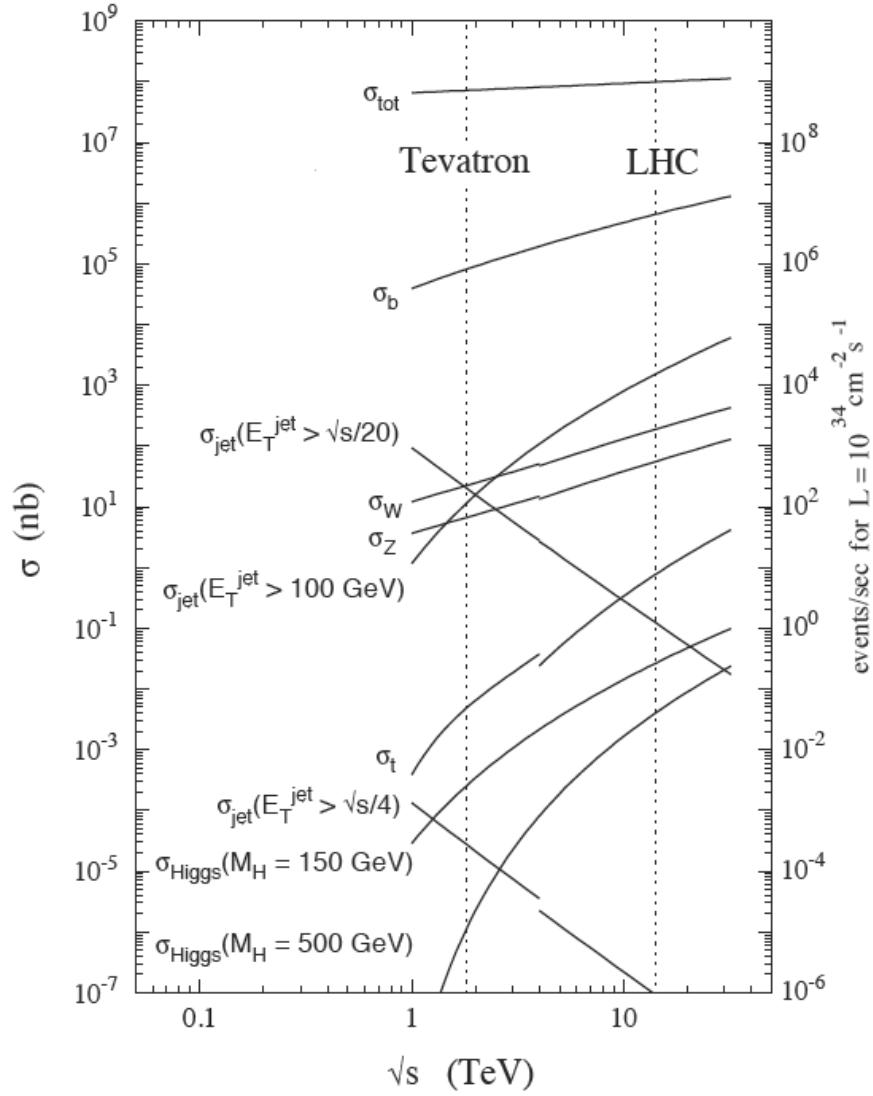


Figure 5.10: b

Figure 5.11: Cross sections of physics processes produced by hadron colliders 5.10

As illustrated in Fig 4.2, the cumulative luminosity delivered by LHC during Run 2 is

about 156 fb^{-1} . However, as shown in Fig 5.10, the rate of events containing electroweak bosons, top quarks or high p_T jets physics phenomena is a tiny fraction of total events. So the trigger system is a key component of hadron collider experiments. In order to select events for a limited data storage capacity, the trigger system needs to provide fast online selection of events from a variety of physics processes to record for later analysis.

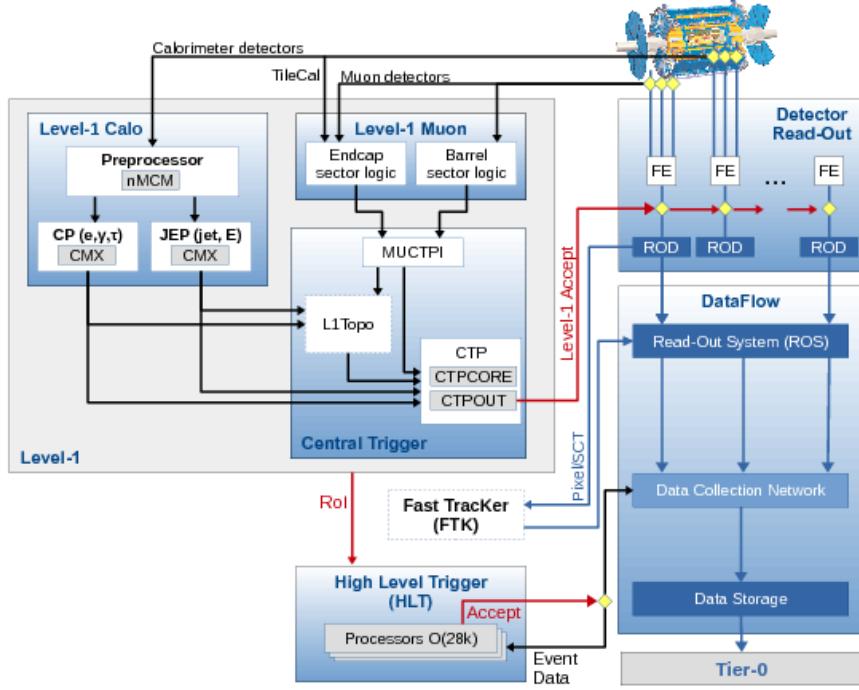


Figure 5.12: The ATLAS TDAQ system in Run 2 with the relevant components for triggering.

The ATLAS Trigger and Data Acquisition (TDAQ) system [109] is illustrated in Fig 5.12. In Run 2, the trigger system consists of two levels of event selections: the Level 1 trigger (L1) is a hardware-based trigger using reduced-granularity information from subdetectors, the ATLAS calorimeter (L1Calo) and Muon Spectrometer (L1 muon trigger). These send trigger information to the L1 Central Trigger Processor (CTP) which generates a pre-scale and final L1 decision (L1A) signal. The front-end of each sub-detector receives the L1A and transmits the data for selected events off-detector. The rate of L1As is about 100 kHz, reducing the total 40 MHz event rate by a factor of 400. This is followed by a software-based

High Level Trigger (HLT) that reduces the rate to 1 kHz on average. Accepted events are reconstructed and sent to long-term storage. As a result, the L1 and HLT triggers together reduce the accepted event rate by a factor of 40000.

Part IV

ATLAS Physics Objects and Event Reconstruction

Chapter 6

Physics Objects

This chapter introduces the “physics objects” used in this thesis in the reconstruction of events. The reconstruction of primary vertices is described in Section 6.1. The reconstruction of electrons, muons and taus is described in Section 6.2, 6.3 and 6.4. Different types of jets that are used by this analysis in different kinematic regions are described in Section 6.5. The reconstruction of missing transverse momentum (E_T^{miss}) and missing transverse momentum significance, is discussed in Section 6.6. Finally Higgs tagging is described in detail in Section 6.7.

6.1 Inner Detector Tracks and Primary Vertex

The tracks in the inner detector are based on fitting a trajectory model to a set of measurements using a sequence of algorithms[65].

The inside-out algorithm is the baseline algorithm and is designed for efficient reconstruction of primary particles. It starts with three-point seeds in the silicon detectors and adds hits moving away from the interaction point using a combinatorial Kalman filter and tracks are extended into the TRT.

Then reconstruction of secondary particles produced by the interactions of the primary particles is achieved by back-tracking. Back-tracking means the track search starts from segments reconstructed in the TRT and extends inwards by adding silicon hits. TRT-standalone tracks refer to tracks from TRT segment without extension into the silicon

detectors.

The transverse and longitudinal impact parameters of a track are referred to as d_0 and z_0 and their resolutions as σ_{d_0} and σ_{z_0} .

Each beam crossing generates multiple proton-proton interactions, leading to multiple track vertices reconstructed from the available inner detector tracks.

All vertices with at least two associated tracks are retained as valid primary vertex candidates. The output of the vertex reconstruction algorithm is a set of three dimensional vertex positions and their covariance matrices, i.e. the vertex corresponding to the hard scatter that generated the physics objects of interest, is selected as the one with the largest $\sum p_T^2$, where the sum is over all associated tracks. The basic track selection criteria are summarized in Table 6.1.

Aim	Selection
Reject soft fake tracks	$p_T > 0.4 \text{ GeV}$
In ID fiducial volume	$ \eta < 2.5$
Enough hits for track reconstruction	More than 9 hits between the Pixel&SCT detectors for $ \eta \leq 1.65$
Enough hits for track reconstruction	More than 11 hits between the Pixel&SCT detectors for $ \eta \geq 1.65$
Good hit quality	Less than 2 hits in a SCT detector layer shared by multiple tracks
Good hit quality	Less than 1 hits in a Pixel detector layer shared by multiple tracks
Good hit quality	0 missing hit in the Pixel detector when a hit is expected
Good hit quality	Less than 1 missing hits in the SCT detector when hits are expected

Table 6.1: Selection criteria for tracks used in the reconstruction of to reconstruct vertices.

6.2 Electrons

Electron candidates are clusters of energy associated with ID tracks, where the final track-cluster matching is performed after the tracks have been fitted with a Gaussian-sum filter.

A few variables are checked to identify electrons while suppressing background objects such as hadronic jets or converted photons [11]. They are the hits in the silicon detectors, including a hit in the IBL, and a likelihood discriminator, which combines the shower shape information provided by the highly segmented calorimeter, high-threshold hits in the TRT, compatibility of the tracking and calorimeter information, track quality information, as well

as the impact parameter in the transverse plane ($|d_0|$) and its significance ($\frac{|d_0|}{\sigma_{d_0}}$).

Electron isolation measures the detector activity around an electron candidate, and can be used to further reject backgrounds such as electrons originating from converted photons produced in hadron decays, electrons from heavy flavor hadron decays, and light hadrons misidentified as electrons.

There are several working points of the likelihood variable corresponding to the strictness of requirements imposed on electrons. This analysis uses the LooseLLHBLayer working point. In addition, two types of electrons are selected in this analysis: VHLoose and ZHSignal. The definitions of VHLoose and ZHSignal electrons are summarized in Table 6.2.

Electron Type	$ p_T $	$ \eta $	$\frac{ d_0 }{\sigma_{d_0}}$	$z_0 \sin\theta$	Likelihood	Isolation
VHLoose	> 7	< 2.47	< 5	< 0.5	LooseLLHBLayer	LooseTrackOnly
ZHSignal	> 27	< 2.47	< 5	< 0.5	LooseLLHBLayer	LooseTrackOnly

Table 6.2: Definitions for the different categories of electron.

The VHLoose definition is used in the Signal Region whereas the ZHSignal definition is used in some control regions.

6.3 Muons

Muon reconstruction is performed based on information from the inner detector, muon spectrometer and calorimeters. Explained in [25], there are five types depending on different reconstruction methods.

- Combined muons are reconstructed by combining the hits of the ID track and MS track and the energy loss in the calorimeter.
- Segment-tagged muons are formed from a track in the ID if it is associated with at least one track segment in the MDT or CSC chambers. It captures muons passing only one layer of MS chambers, due to their low p_T or reduced MS acceptance in the region.

- Extrapolated muons are reconstructed based only on the MS track and a loose requirement on compatibility with originating from the IP. They are used to extend the acceptance for muon reconstruction into the region $2.5 < \eta < 2.7$, where three layers are required.
- Calorimeter-tagged muons. In the region of $|\eta| < 0.1$, ID tracks with $15 \text{ GeV} < p_T < 100 \text{ GeV}$ are identified as muons if their energy deposits in the calorimeter match that expected from minimum ionizing particles. They recover muon acceptance in the region where the MS is only partially instrumented.

Similar to Electron reconstruction, there are different muon identification working points. This analysis chooses “Loose”, defined as muons reconstructed using any reconstruction methods, and “Medium”, reconstructed using either the Combined muon or Extrapolated muon methods. The VHLoose definition is used in the signal region whereas the ZHSignal definition is used in some control regions.

As shown in Table 6.3, this analysis applies additional criteria to muons in different regions. A tighter selection is required in control regions for a high muon purity and a looser selection when muons are not desired and are vetoed in the signal region.

Electron Type	$ p_T $	$ \eta $	$\frac{ d_0 }{\sigma_{d_0}}$	$z_0 \sin\theta$	Likelihood	Isolation
VHLoose	> 7	< 2.7	< 3	< 0.5	Loose	LooseTrackOnly
WHSignal	> 25	< 2.5	< 3	< 0.5	Medium	FixedTrackTTTight
ZHSignal	> 25	< 2.5	< 3	< 0.5	Loose	LooseTrackOnly

Table 6.3: Definitions for the different categories of muons.

6.4 Taus

The Tau lepton, unlike the lighter leptons, electron and muon, is the only lepton that can decay hadronically. The branching ratio to hadrons is 64.79%. Therefore, it is important to reject events with hadronically decaying tau lepton.

Hadronically decaying tau leptons can be reconstructed as small-radius jets and identified with a tree-based machine learning algorithm. A typical tau jet contains either one or three charged hadrons, which can be exploited in the tree-based classifiers to differentiate from QCD jets. With the trade-off of selection efficiencies and fake rate, three classifiers, loose, medium, tight, can be determined. More details can be found in Ref[14]. In order to suppress the tau background to a maximum extent, the loose working point is chosen to build the tau veto condition, together with the baseline section of $p_T > 25\text{GeV}$ and $|\eta| < 2.5$.

6.5 Jets

In proton-proton collisions, almost immediately after being produced, a quark or gluon fragments and hadronizes, leading to a collimated spray of energetic hadrons – a jet [110]. There exist many different jet clustering algorithms to measure the jet properties.

Among all jet finding algorithms, the anti- k_T algorithm [55] is now most commonly used. This algorithm favors clusterings that involve hard particles, and the jets then grow outwards around hard “seeds”. The radius R represents the clustering distance metric. It determines the radial size of the jet in the $\eta - \phi$ plane.

Calorimeter jets are reconstructed by combining topological clusters of energy deposits in the calorimeter [22], while the inputs to track jet clustering are inner detector (ID) tracks. Small-radius ($R=0.4$) calorimeter jets are designed to capture the four-momentum of a single hard quark or gluon produced in the pp collision. For jets that are from collimated decay products of heavy particles or low p_T jets, a large cone size is used to contain all of the decay products. Large-radius jets are defined as calorimeter jets reconstructed with the anti- k_T algorithm with $R = 1.0$. The large-radius calorimeter jets are used for Higgs boson reconstruction in the boosted topology in this analysis.

In this iteration of the analysis, variable radius track jets used for b-quark identification inside the large-radius jets.

6.5.1 Calorimeter jets

Topological clusters (topo-clusters) that are cells are clustered together to reconstruct particle showers, exploiting the fine granularity of the calorimeters. The topo-cluster reconstruction is based on cell signal significance S/N , defined as the ratio of cell energy at electromagnetic energy scale over average expected noise. The reconstruction starts from a seed cell with signal significance above the threshold of $S/N = 4$. Then the neighbouring cells with signal significance over $S/N = 2$ are included iteratively. Finally, all calorimeter cells neighbouring the formed topo-cluster are added.

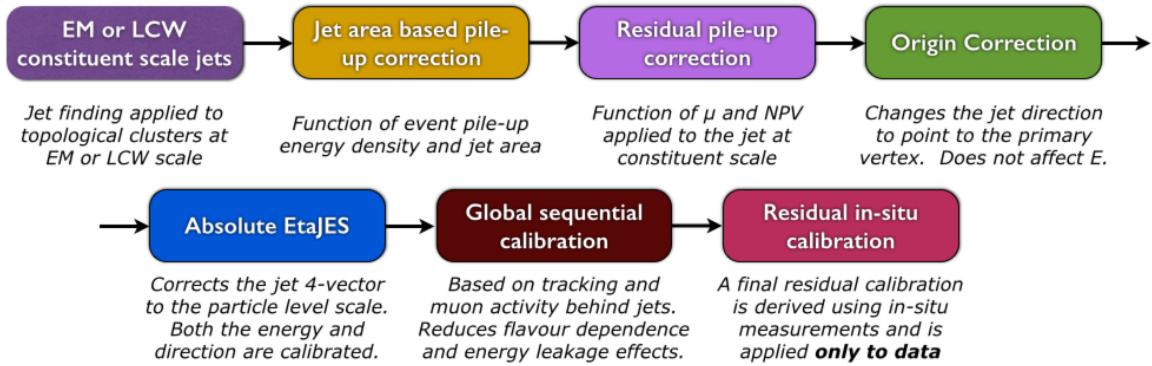


Figure 6.1: Calibration stages for reconstructed jets.

After jets are reconstructed, a calibration is applied to the jet four momentum to recover the particle-level energy scale. A series of calibrations are applied after jet clustering as shown in Fig 6.1.

- **Jet area-based pile-up correction** This step are designed to remove the excess energy from pile-up interactions. The area-based p_T density subtraction is applied event-by-event [54]. The p_T density is estimated using the medium of p_T density of all jets in the event calculated by p_T/A , where A is the jet area.
- **Residual pile-up correction** A residual p_T dependence on in-time pile-up evaluated through the number of reconstructed primary vertices and out-of-time pile-up μ is roughly linear. So a linear correction is applied with coefficients derived from MC simulation.

- **Origin correction** The kinematic observables of each topo-cluster are recalculated using the vector from the primary hard-scattering vertex to the topo-cluster centroid as its direction. The resolution of η can be improved in this step. The jet energy is unaffected.
- **Absolute MC-based calibration** An absolute jet energy and η correction is derived from MC simulation. The average energy response is defined as the mean of E_{reco}/E_{truth} binned in E_{truth} and η_{jet} . A similar correction is done for η .
- **Global sequential calibration** The global sequential calibration (GSC) method [22] is applied to improve resolution of the Jet Energy Scale (JES) resolution. In this step, JES depends on five features which account for different aspects and the calibration factor is derived from MC. The procedure is similar to the absolute JES calibration.
- **Residual in situ calibration** This steps aims at correcting the difference between data and MC. For jets up to 950 GeV and with $|\eta| < 0.8$, calibration factors are derived from Z+jet and η +jet balance, while, for high- p_T jets up to 2 TeV, the multijet balance is used.

The larger the jet size, the more chances there are that particles from the underlying events, pile-up interactions and soft radiation contaminate the jet. Thus for large R jets, three techniques of jet grooming are developed: split-filtering [52], pruning [70] and trimming [96].

- **Split-filtering** Large-radius jets are rebuilt using the Cambridge/Aachen (C/A) algorithm. C/A jets are then de-clustered by splitting the jet into two pieces into pieces, starting from the last C/A merge. De-clustering continues with the highest mass piece until the requirements on mass-drop $\mu_{12} < \mu_{max}$ and momentum balance $\sqrt{y_{12}} > \sqrt{y_{min}}$ are met. The momentum balance and mass drop are defined as $y_{12} = \frac{\min(p_{T1}, p_{T2})}{m_{12}} \Delta R_{12}$, $\mu_{12} = \max(m_1, m_2)m_{12}$, where m_{12} is the invariant mass of two pieces. If the requirements are not satisfied, the jet is discarded. Filtering aims at removing soft radiations that is irrelevant to the hard scattering process. In

the filtering stage, constituents of the surviving jet are reclustered with subjet size $R_{sub} = \min(0.3, \Delta R_{12})$, where ΔR_{12} is taken from the splitting stage.

- **Trimming** Large-R jets are reclustered using the k_T algorithm, which favors clustering low p_T constituents first. This creates a set of subjets with radius parameter R_{sub} . Subjets with an energy below some threshold fraction of the energy of the large-radius jet are removed from the large R jet.
- **Pruning** Pruning is similar to trimming as it removes constituents with relative small p_T , while it has an additional wide-angle radiation veto. Constituents of ungroomed large-radius jet are then re-clustered with the C/A algorithm. In each pairwise clustering step, secondary constituents with wide-angle $\Delta R_{12} > R_{cut} \times \frac{2M}{p_T}$ or soft property are discarded. The definition of being soft is that $f_2 < Z_{cut}$ where f_2 is the fraction of the softer constituent p_T with respect to the pair.

Trimmed large-radius jets are used to reconstruct Higgs candidates when the b-quark decay products of the Higgs are too collimated to resolve by two small-radius jets.

Small-radius jets used in the analysis can be divided into two categories: central jets and forward jets. Forward jets are small-radius jets with $2.5 < |\eta| < 4.5$ and $p_T > 30GeV$, and can be used to reduce backgrounds. Small-radius jets with $|\eta| < 2.5$ and $p_T > 20GeV$ are called central jets and they are used to reconstruct Higgs candidates with low or moderate p_T . For central jets with $|\eta| < 2.4$ and $20GeV < p_T < 60GeV$, an additional requirement on the jet vertex tagger value $JVT > 0.59$ is imposed. Details on the jet vertex tagger (JVT) can be found in Ref [26].

6.5.2 Track jets

Track jets are jets built entirely from tracks reconstructed in the inner detector. In Run 2, track jet b-tagging became the standard approach to resolve the heavy flavor components from boosted decay of heavy resonances [12; 15]. Studies of track jet calibration can be found in Ref [116].

In this thesis, two types of track jets are discussed. $R = 0.2$ track jets will be referred to as FR track jets, for fixed radius track jets, in contrast to the variable radius (VR) track

Aim	Selection
Reject soft fake tracks	$p_T > 0.4\text{GeV}$
In ID fiducial volume	$ \eta < 2.5$
enough hits for track reconstruction	More than 7 hits between the Pixel and SCT detectors
good hit quality	Less than 1 hit in a Pixel shared by multiple tracks
good hit quality	Less than 1 missing hit in the Pixel detector when a hit is expected
good hit quality	Less than 2 missing hits in the SCT detector when hits are expected
reject tracks from pile-up	$z_0 \sin(\theta) < 3\text{mm}$

Table 6.4: Selection criteria for tracks to cluster track jets.

jets.

A track selection of input ID tracks is applied in order to suppress fake tracks and tracks from pile-up vertices. The selection criteria are summarized in Table 6.4. With a relatively looser requirement on hits in the ID detectors compared to the track selection criteria for primary vertex reconstruction, a small longitudinal impact parameter of the tracks with respect to the primary vertex is required to reject pile-up vertices.

FR track jets are then built by applying the anti- k_T algorithm with $R = 0.4, 0.3, 0.2$ on selected tracks. Track jets in the fiducial region with $p_T > 7\text{GeV}$ and $|\eta| < 2.5$ are accepted as track jets with $p_T < 7\text{GeV}$ are dominated by light jets [10].

VR track jets are clustered using the anti- k_t algorithm from tracks with the same selection criteria used for $R = 0.2$ track jets. The main feature of the VR track jet reconstruction [95], compared to the FR jets is the p_T dependence of the jet radius:

$$R \rightarrow R_{eff}(p_T) \approx \frac{\rho}{p_T} \quad (6.1)$$

where the parameter ρ shows how the effective jet radius scales with the p_T of the pseudo-jet during the jet finding procedure, R_{min} and R_{max} are a lower and an upper cut-off on the jet radius. To optimize the efficiency of double b-tagging over a wide mass range, $\rho = 30\text{GeV}$, $R_{min} = 0.02$ and $R_{max} = 0.4$ [18].

6.5.3 Ghost Association

In practice, track jets are seldom used alone for object reconstruction and instead used along with the large-radius calorimeter jets. Calorimeter jets are in charge of providing jet

reconstruction while track jets are in charge of providing flavor tagging information. To match track jets to calorimeter jets is the first step for flavor tagging of calorimeter jets.

Ghost association [54; 55] is a method to associate the “ghosts” (particles, jets or tracks) to jets by giving them negligible momentum and clustering them within the jets. This is to make sure that the hard-particle content of the jet is unaltered by the addition of the soft ghost particles during jet reclustering. The jet substructure after reclustering is unchanged compared to the previous jet, but with the addition of “ghosts” as constituents.

An object (track, jet, truth particle) is ghost-associated to a jet if its “ghost” is clustered as a constituent of the jet.

Compared to ΔR association which matches objects based on angular distance, ghost association is more robust when dealing with overlapping jets or jets that are not cone-shaped.

6.5.4 Flavor tagging

The identification of b-jets is referred to as b-tagging or flavor tagging. After the fragmentation of b-quarks, about 70% of the b-quark energy goes into the weakly decaying b-hadrons (5 GeV). With an intrinsic life-time of 1.5×10^{12} s, the average decay lengths for b-hadrons of 30 GeV is about 3 mm, which can be measured with the high-precision tracking system in ATLAS. The c-hadrons from b-hadrons decay have similar average lifetimes, and thus lead to an additional decay vertex. Several b-tagging algorithms are in use to exploit the decay patterns.

ID tracks are used for these b-tagging algorithms, and tracks need to pass the basic quality cuts.

These three b-tagging algorithms below are used to provide complementary information and combined using a boosted decision tree (BDT) to provide a score to distinguish between different flavors (c, b, light).

- **Impact Parameter Based Algorithms (IP2D and IP3D)** As shown in Fig.2 from Ref [13], tracks from a displaced vertex have larger impact parameter than those coming from the primary vertex. The probability density functions (PDFs) for the signed impact parameter significance of these tracks $\frac{d_0}{\sigma_{d_0}}$ and $\frac{z_0}{\sigma_{z_0}}$ are used to

define probability ratios of the jet hypotheses of different flavors, and these are then combined in a single log likelihood ratio discriminant (LLR). While IP2D only uses the transverse impact parameter, IP3D uses the 2D template with correlation between the transverse and longitudinal direction.

Both IP2D and IP3D assume that all tracks are independent and use naive Bayesian models which ignore correlations between tracks. Besides, while it is easy to build 2D or 3D template histograms, it is technically hard to encode too many variables at the same time. To improve the results, new algorithms which make use of recurrent neural networks with the same input tracks as IP2D/IP3D are developed in ATLAS b-tagging [16].

- **Inclusive secondary vertex reconstruction algorithm (SV)** The secondary vertex based algorithm explicitly reconstructs an inclusive displaced secondary vertex within the jet. Tracks passing quality cuts are first paired to form two-track vertex. These two-track vertices are then be filtered to reject those coming from decays of long-lived particles such as K_s , Λ , photon conversions or hadronic interactions with detector material. An inclusive secondary vertex is fit using surviving tracks using a Kalman filter in an iterative way with certain quality cuts applied.
- **Decay chain multi-vertex reconstruction algorithm (JetFitter)** The decay chain multi-vertex reconstruction algorithm is called JetFitter and exploits the cascade decay structure of b- and c-hadrons to reconstruct the decay chain $Primary\ Vertex \rightarrow b \rightarrow c$ using a Kalman filter.

Unlike the SV algorithm, the vertex in JetFitter refers to the intersection between tracks and estimated decay chain direction and thus a vertex can have only one track associated to it. Thus JetFitter typically has a much higher reconstruction efficiency, as well as a higher fake rate. To reduce the fake rate, a set of topological variables are built in JetFitter for discrimination purposes.

After running each b-tagging algorithm independently, the output discriminative variables of each algorithm, as well as the kinematic properties of jet p_T and η are combined using a BDT. The p_T and η distributions for signal and backgrounds are re-weighted to match

each other before training so that these kinematic variables are not treated as discriminative variables.

The training is performed on high purity $t\bar{t}$ and Z' MC samples with at least millions of events. While b-jets are used as signal jets, the background is a mixture of c-jets and light-jets. Three versions of the mixture are used: MV2c00, MV2c10 and MV2c20. The number after the character “MV2c” shows the percentage of c-jets in the background mixture. Depending on the physics processes, users can choose any of these three version. The higher the number, the better the discrimination power against c-jets at the cost of reduced light-jet rejection. The performance of MV2c10 tagger [117] for discrimination against light-jet and c-jets can be found in Fig 6.2.

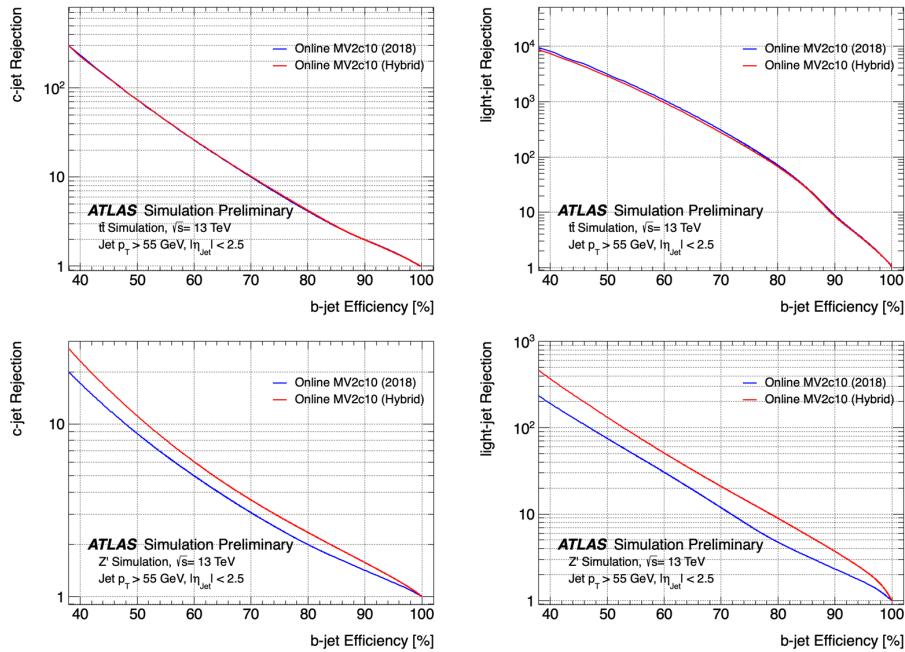


Figure 6.2: The Expected performance of the b -tagging algorithm (MV2c10) for b -jet triggers in 2018 data-taking (blue solid line) is compared to the same b -tagging algorithm trained on the Hybrid training sample (red solid line).

6.6 Missing Transverse Momentum (E_T^{miss}) and Missing Transverse Momentum Significance

According to the conservation of momentum, the sum of the transverse momenta of all particles produced in collisions is zero. Considering the existence of non-detected objects, the transverse momenta from detected objects need not be balanced. The unbalanced part of the transverse momentum is called “missing transverse momentum” (E_T^{miss}), or MET for short.

As mentioned in the last paragraph, the E_T^{miss} is derived from all detected objects. The E_T^{miss} calculation is the most difficult object not only because it needs accurate measurement of all detected objects from various detectors, but also of the small fraction of energy deposits that are not clustered by any algorithm. As a result, the E_T^{miss} scale and E_T^{miss} resolution are affected by many factors, such as missing muons, mismeasured jets, beam pile-up, etc. In the ATLAS experiment, the E_T^{miss} reconstruction uses energy deposits in the calorimeters and muon tracks reconstructed in the muon detectors. The Trackers’ information is also used to recover the low p_T fraction missed in the calorimeters [7]. Therefore, both hard objects (high p_T), like electrons, jets, muons, etc, and soft objects, like track soft terms, are considered in the E_T^{miss} reconstruction.

The uncertainty of E_T^{miss} calculation result can be large due to the complexity of the reconstruction algorithm. Therefore, a significance variable can be introduced to describe the reliability of the derived E_T^{miss} . The E_T^{miss} significance is defined as

$$S = \frac{E_T^{\text{miss}}}{\sqrt{\sum_i E_{Ti}}},$$

where the numerator is the amplitude of the derived E_T^{miss} , and the denominator is the scalar sum of the detected objects that are used in reconstructing the E_T^{miss} . A high value of E_T^{miss} significance suggests the event is more likely to contain an invisible object than be due to resolution smearing.

6.7 Higgs tagging

For a Higgs particle with sufficiently low p_T , two outgoing b-quarks can be reconstructed individually as small-radius equals to 0.4 calorimeter jets. However, for a Higgs particle with high p_T , the two outgoing b-quarks are too collimated to be reconstructed using small-R jets. “Higgs tagging” refers to the techniques used to identify boosted Higgs decays to b-quarks. The Higgs candidate is reconstructed as a trimmed large-R jet with two ghost-associated b-tagged subjets, reconstructing the b-hadrons.

6.7.1 Higgs tagging with advanced subjets

Three techniques are used to tag boosted Higgs bosons: variable-radius track jets (Fig 6.3), Exclusive-kT (ExKt) and Center-of-Mass (CoM) tagging. These allow to probe the very boosted region [18], where b-hadron decay products are too collimated to resolve even with $R = 0.2$ track jets.

For variable-radius track jets, the parameters $\rho = 30\text{GeV}$, $R_{min} = 0.02$ and $R_{max} = 0.4$ are chosen after scanning different ρ and radius as values in Fig. 6.4. ExKt refers to exclusive regions of interest using exclusive-kT declustering of the large-radius jets. CoM jets are constructed via exclusive-kT declustering in the jet’s center of mass frame instead of in the lab frame.

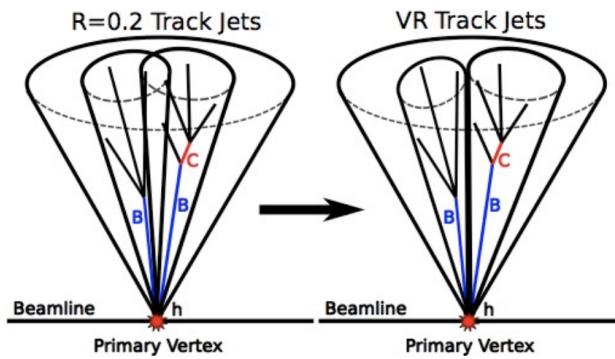


Figure 6.3: A cartoon depicting using VR track jets instead of $R = 0.2$ track jets.

The double b-tag performance for the FR track jet, VR track jet, ExKt and CoM techniques can be found in Fig. 6.5. The plot shows a dramatic decrease for the $R = 0.2$

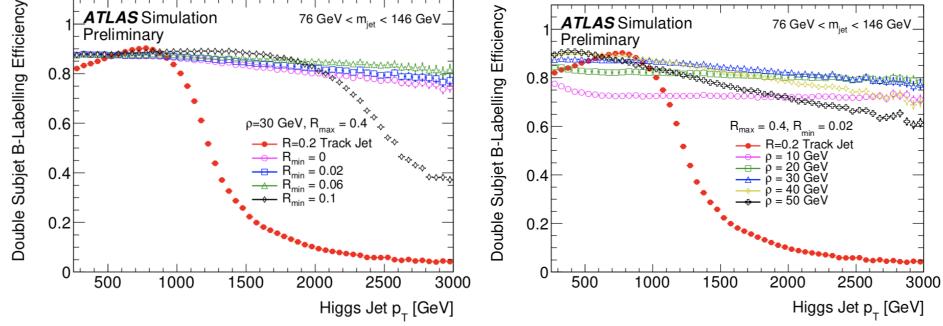


Figure 6.4: Higgs efficiency using two VR track jets with different R parameter (Left) and ρ parameter (Right)

track jet technique as the Higgs jet p_T becomes larger than 1.2 TeV where the $R = 0.2$ track jets are expected to merge. These new techniques, however, can reconstruct Higgs jets with a p_T of 3 TeV. Receiver operating characteristic (ROC) curves showing Higgs jet tagging performance versus QCD jet rejection are shown in Fig. 6.6.

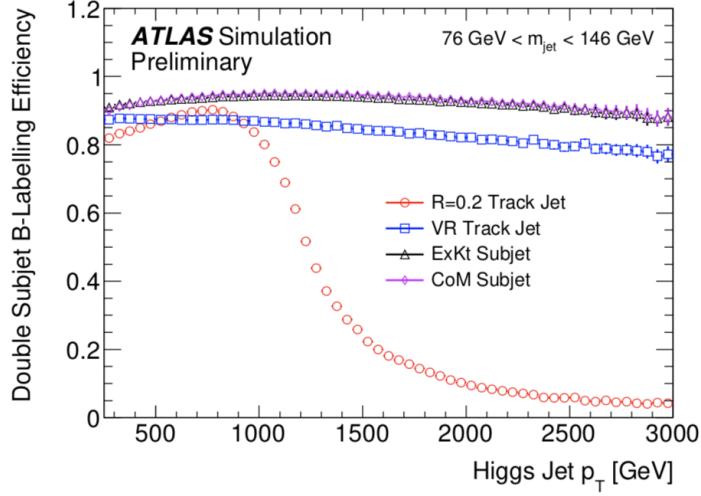


Figure 6.5: Higgs efficiency using four different subjet techniques: $R = 0.2$ track jets, VR track jets, ExKt calorimeter jets, and CoM calorimeter jets.

The variable-radius track jet technique was chosen to use as subjets in the mono-Higgs analysis as an existing framework to calibrate track jet b-tagging efficiency was readily available in ATLAS.

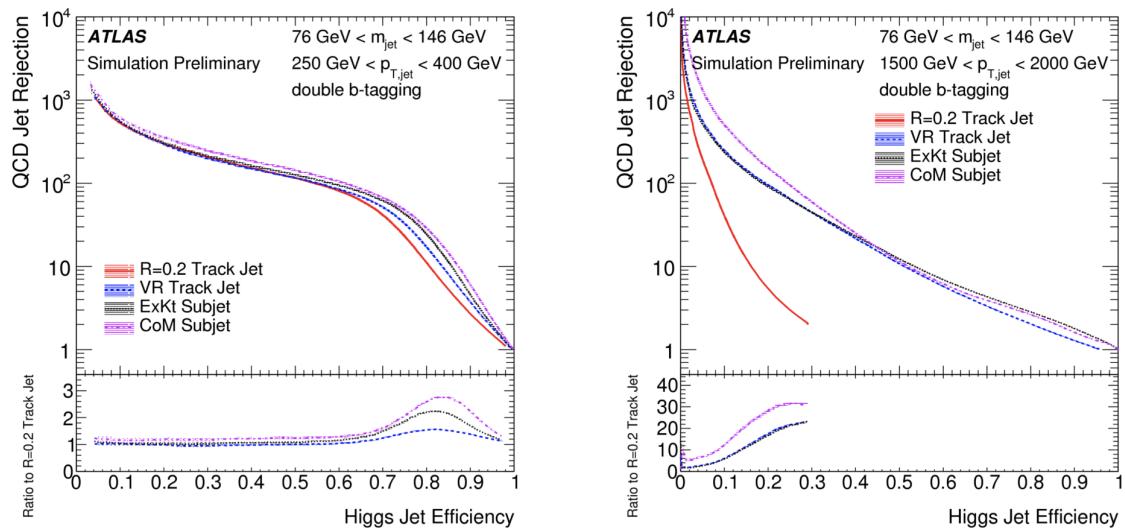


Figure 6.6: ROC curves using four different subjet techniques: R = 0.2 track jets, VR track jets, ExKt calorimeter jets, and CoM calorimeter jets in two different jet p_T ranges

Chapter 7

Double-b tagger

This chapter introduces a new tagger to reconstruct boosted Higgs bosons. Section 7.1 gives a brief overview of this “Xbb” tagger. Monte Carlo simulation samples used for training the tagger are summarized in Section 7.2. The algorithm is introduced in Section 7.3. The neural network architecture is described in Section 7.4. Section 7.5 shows the performance of the tagger, which includes the signal efficiency and background mistag rate measurement.

7.1 Introduction

For a Higgs boson with p_T above about 250 GeV, the two b quark jets merge into a single jet for a jet cone size of $R = 1$ ($\Delta R = \frac{2m}{p_T}$) and Higgs boson reconstruction can exploit this topology. The decaying object is reconstructed within a large-radius jet. The previous chapter showed studies that explored $H \rightarrow b\bar{b}$ tagging algorithms using ghost-associated subjets, but the performance can be further optimized using both the substructure information of the large-radius jet and the track and vertex information related to the b hadron lifetime.

The approach presented here utilizes both the jet substructure and the low-level b-tagging information from the bb pair within the same large-radius jet.

A universal boosted Xbb tagger is built by avoiding a strong performance dependence on the large-radius jet p_T and mass. As a result, the tagger can be used in many different topologies and kinematical regimes, such as searches for the Higgs boson in ttH, VH and

VBF production modes, resonant HH and VH production, t' and b' in the tH and bH final states. The boosted Higgs tagger is also essential for this thesis - a search for a boosted Higgs boson produced in association with dark matter. In principle, the algorithm should also be able to identify $Z \rightarrow b\bar{b}$, as well as any hypothetical particle decaying into a $b\bar{b}$ pair with a mass close to the W/Z/H boson mass.

7.2 Monte Carlo Simulation Samples

A sample of background jets initiated by light quarks and gluons is derived from a multijet process simulated using PYTHIA8 [114] with the NNPDF2.3 leading order (LO) parton distribution function (PDF) set [43] and the A14 [8] tune for underlying event parameterisations. Jets from this sample are referred to as QCD jets. Fully hadronic top quark pair events are used for jets originating from hadronic top quark decays (top jets) and are generated using POWHEG [77] [103] interfaced to PYTHIA6 [?] with the PERUGIA 2012 [?] underlying event tune parameter set and the four flavor scheme of the CT10 PDF set [?].

For the boosted Higgs jets, a sample of high p_T Higgs bosons is obtained from the BSM physics simulation of a Randall-Sundrum graviton decaying to a pair of Higgs bosons with both Higgs bosons subsequently decaying to $b\bar{b}$ pairs ($g \rightarrow HH \rightarrow b\bar{b}b\bar{b}$) [?]. This process is generated using MadGraph5 [36] interfaced with PYTHIA8 and with the ATLAS A14 tune, and the NNPDF2.3 LO PDF set. The mass splitting between the massive graviton and the Higgs boson provides a boost proportional to the mass of the graviton. Therefore, signal samples have been generated with graviton masses between 300 GeV and 6000 GeV to fully populate the kinematic region of interest. The signal samples of various masses are merged and both top and signal samples are reweighted on a jet-by-jet basis such that the large-radius calorimeter jet kinematics in signal matches that of the QCD jet sample in p_T and η . This reweighting is intended to mitigate the effects of any difference in kinematics present between the unweighted samples. EvtGen [?] is used to model the decays of b- and c-flavoured hadrons.

The Monte Carlo samples are processed through the full ATLAS detector simulation [?] with the tool Geant4 [?], and reconstructed using the standard ATLAS reconstruction

software. Simulated minimum-bias events are added to the events to match the pile-up distribution in data.

7.3 Double-b tagger algorithm

As shown in the Chapter 6, the previous approach requires matching the large-radius jet which are used to reconstruct Higgs boson with two track jets.

The standard single-b tagging algorithm is applied to the track jets to identify subjets from single b quarks. With its focus on individual subjets, the subjet single-b tagging does not make use of the global properties of the large-radius jets which contain two b hadrons. Even though the algorithm performs well in the high purity regime, relying heavily on the reconstruction of secondary vertices associated with the subjets, at high p_T , the subjets start to overlap and cause the standard single b tagging techniques to break down due to double-counting of tracks and secondary vertices when computing the subjet single b-tagging discriminants.

To discriminate $b\bar{b}$ pairs originated from Higgs bosons from QCD jets initiated by single partons, a boosted Higgs tagger with large signal-to-background ratio is built using a deep neural network (DNN) which fully exploits the presence of two b quarks inside a large radius jet and their topology in relation to the jet substructure.

7.3.1 Input features

The inputs fed into the neural networks are both features of the subjets declustered from the large-radius jet and features that describes the correlation between the subjets.

For each subjet,

- To reconstruct b hadron decay vertices, we apply the Inclusive secondary vertex (SV1) algorithm which identifies secondary vertices independently of the jet clustering.
- We reconstruct the decay chains of the two b hadrons by associating reconstructed secondary vertices to the subjet axis via JetFitter.
- We use the output scores of two algorithm, IP3D/IP2D (using likelihood-ratio method)

and RNNIP (using recurrent neural networks) which process charged particle tracks associated to jets without reliance on secondary vertex finding to augment secondary-vertex based taggers.

Similarly to the single-b tagger (MV2c10), the scores from these low-level taggers are all fed into the neural network to complement each other's performance.

All input features constructed within each subjet that are used for the training are listed in Table 7.1.

Low-level Algorithm	Algorithm Outputs (Training inputs)
IP2D(IP3D)	IP2D_pu, IP2D_pb, IP2_pc, IP3D_pu, IP3D_pb, IP3D_pc
RNNIP	RNNIP_pu, RNNIP_pb, RNNIP_pc, RNNIP_tau
SV1	sv1_ntkv, sv1_n2t, sv1_mass, sv1_efrc, sv1_dR, SV1_dstToMatLay, SV1_Lxy, SV1_L3d, sv1_sig3
JetFitter	jf_n2tv, jf_ntrkv, jf_ntvx, jf_ntvx1t, jf_mass, jf_efrc, jf_dR, jf_sig3, JetFitter_deltaphi, JetFitter_deltaeta

Table 7.1: Input features used for training double-b tagger.

In addition, the following features are used to describe the correlations:

- Subjet kinematic variables: p_T and η .
- Subjet angular variables: dR , $d\eta$, $d\phi$ with respect to the parent large-radius jet.
- Event-based variables: p_T imbalance between subjets which indicates the ratio of each subjet momentum to the total momentum.

7.4 Neural network architecture

A feed-forward, fully connected neural network [?] with 6 hidden layers, 164 input features and 2 or 3 outputs is designed. While a batch normalization layer is applied at the input layer, each hidden layer of the neural network is built from the following components:

- Dense layer: defined as a linear combination of all outputs from the previous layer.
- Batch normalization layer: to transform the inputs to zero-mean and unit-variance. [?]
- Dropout layer: an operation that drops a fixed fraction of randomly chosen nodes, used as a regularization handle. The dropout rate is one of the optimized hyperparameters of the neural network. [?]
- Activation unit: we use the Rectified Linear Unit (ReLU) [?]:

$$f(x) = \begin{cases} 0 & \text{for } x < 0, \\ x & \text{for } x \geq 0, \end{cases} \quad (7.1)$$

To optimize the performance of the neural network, three hyperparameters are considered: the depth of the network architecture, the dropout rate and the learning rate. The following values are considered:

- Dropout rate: 0.1, 0.2, 0.3, 0.4, 0.5;
- Number of hidden layers: 3, 4, 5, 6, 7;
- Learning rate: 10^1 , 10^2 , 10^3 , 10^4 , 10^5 .

Out of 125 grid points in the hyperparameter space, 50 combinations are randomly sampled. From all the configurations with comparable performance, the network with the smallest number of trainable parameters was chosen.

We start with learning rate = 0.1 and uses an adaptive learning rate which means decreasing the learning rate by a factor of three once the performance of the model plateaus. Also, early stopping [?] which allows 20 training epochs and stops training once the model performance stops improving on a hold out validation dataset are applied as regularization.

7.5 Signal efficiency and Background mistag rate measurement

Both signal efficiency and background mistag rate of the double-b tagger are evaluated with the testing samples which are independent of the training samples.

Performances of the following taggers are evaluated: DL-tag refers to a bi-class discriminator that distinguishes QCD from Higgs with input features from each subjet, and subjet correlation. XbbScore is a multiclass discriminator (QCD, Higgs, Top) which also includes other large-radius jet substructure features in the input features. The baseline boosted Hbb tagger, which identifies a large-radius jet with two ghost-associated track jets (2-tag) or one b-tagged track jet (1-tag leading jet), is used for comparison.

The performance can be expressed in terms of its receiver operator characteristic (ROC) plot, which shows the change of background rejection as a function of the Higgs boson tagging efficiency. Fig 7.1 (left) compares double-b taggers with the baseline boosted Higgs tagger while the right plot compares XbbScore which trains Hbb against QCD and top jet with a top tagger built on a neural network and the jet substructure variable τ_{32} .

In general, the double-b tagger outperforms the baseline boosted Hbb and top tagger. At the same signal efficiency, the mistag rate of the new method is uniformly lowered by about a factor of 2 to 5.

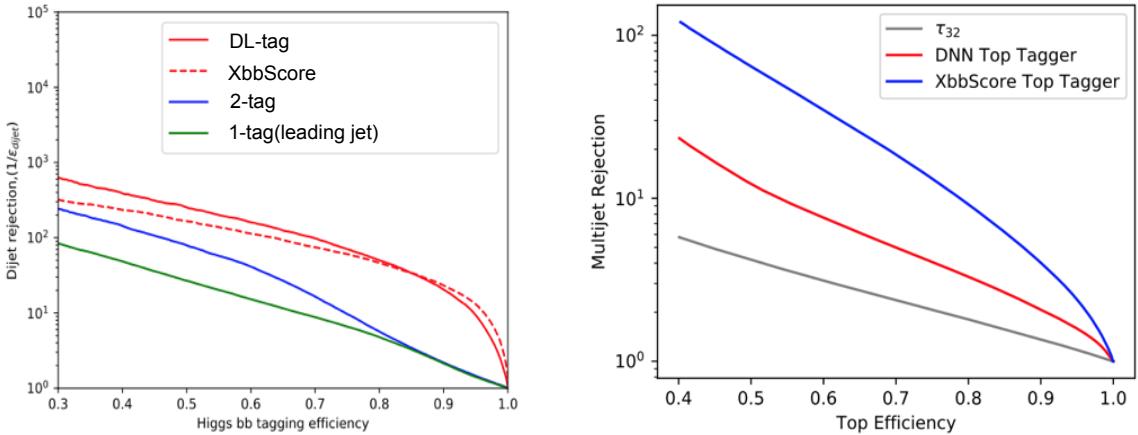


Figure 7.1: ROC plot: Comparison between Hbb taggers (Left) and Top taggers (Right).

In Fig. 7.2, the signal efficiencies and mistag rates for the double-b tagger as a function of large-radius jet p_T are reported with the baseline boosted Hbb tagger for comparison.

The left plot shows the Higgs efficiency for a fixed 70% working point (WP) cut as a function of large-radius jet p_T . To isolate the effect of a changing Higgs efficiency from that of the changing dijets rejection, the right plot shows a flat-efficiency 70% WP, which means all taggers have a 70% efficiency across p_T and only the rejection is varying. The mistag rate is relatively flat across the p_T range, while the signal efficiency decreases with increasing p_T , as expected from the degradation of the tracking performance inside high p_T jets. Again, double-b tagger outperforms the baseline boosted Hbb tagger.

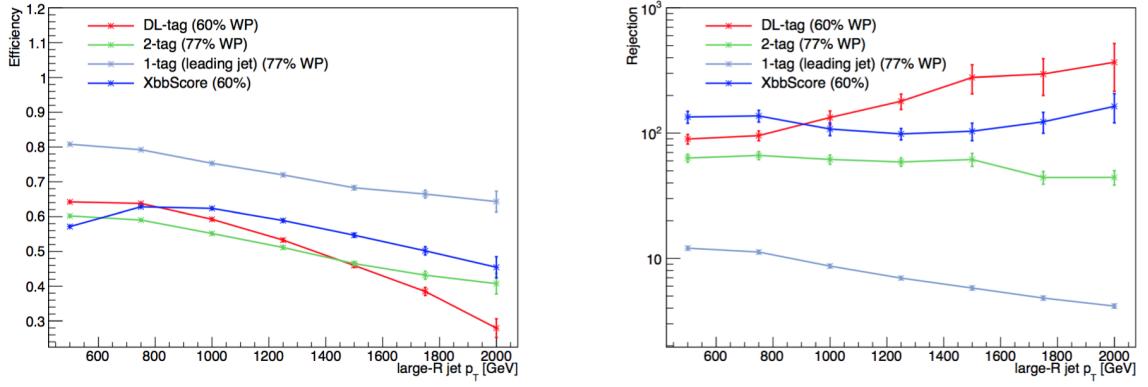


Figure 7.2: Higgs efficiency (left) and Background rejection (right) as a function of large-radius jet p_T

Part V

Dark Matter search in the Higgs Boson associated $b\bar{b}$ decay

Chapter 8

Data and MC samples

8.1 Data samples

The pp collisions at a centre-of-mass energy of 13 TeV recorded between 2015 and 2017 are used in this search, while the Good-Run-Lists are showed below:

- `data15_13TeV.periodAllYear_DetStatus-v89-pro21-02_Unknown_PHYS_StandardGRL_All_Good_25ns.xml`
- `data16_13TeV.periodAllYear_DetStatus-v89-pro21-01_DQDefects-00-02-04_PHYS_StandardGRL_All_Good_25ns.xml`
- `data17_13TeV.periodAllYear_DetStatus-v99-pro22-01_Unknown_PHYS_StandardGRL_All_Good_25ns_Triggerno17e33prim.xml`
- `data18_13TeV.periodAllYear_DetStatus-v102-pro22-04_Unknown_PHYS_StandardGRL_All_Good_25ns_Triggerno17e33prim.xml`

The resulting dataset corresponds to an integrated luminosity of 3.2 fb^{-1} , 33.0 fb^{-1} , 44.3 fb^{-1} and 58.5 fb^{-1} for each Good-Run-Lists, respectively. The total integrated luminosity is 139.0 fb^{-1} . The proton bunch gap equals to 25 ns.

8.2 Trigger

E_T^{miss} triggers is used in 0 and 1 lepton regions and unprescaled single-muon triggers in the 2 lepton region. Their thresholds are determined by requiring lowest unprescaled single-muon triggers [40].

An offline cut $E_T^{\text{miss}} > 150 \text{ GeV}$ is applied as the triggers are not fully efficient in low E_T^{miss} region. Since the trigger turn-on curve is not well modeled in MC, trigger efficiencies

are measured in both data and MC from a single-muon measurement region, and scale factors are calculated to correct turn-ons in MC to match data in the signal region and the single-muon control region.

$E_{T,\mu \text{ invis.}}^{\text{miss}}$ compensates the online E_T^{miss} which is reconstructed using only calorimeter information without the contribution of the muon. Events collected with the muon trigger can be used to measure the trigger turn-on curve of the E_T^{miss} trigger.

The efficiencies of E_T^{miss} triggers are derived in a single-muon measurement region. The event selection is same as the selection in the resolved regime in Chapter 9, except for the cut on E_T^{miss} . The efficiencies are calculated inclusively in the number b -jets to allows for a larger statistics in the measurement region.

The E_T^{miss} trigger efficiency is defined by:

$$\text{efficiency} = \frac{\#\text{Events passed selection AND } E_T^{\text{miss}} \text{ trigger requirement}}{\#\text{Events passed selection}} \quad (8.1)$$

The efficiencies are calculated for each E_T^{miss} trigger separately within the data-taking periods in which it was used, and separately for data and MC. The trigger efficiency curves for each E_T^{miss} trigger are shown in Fig. 8.1 as a function of $E_{T,\mu \text{ invis.}}^{\text{miss}}$ to mimic the E_T^{miss} topology on trigger level.

Scale factors (SF) are defined as the ratio of E_T^{miss} trigger efficiencies for data and MC:

$$\text{SF} = \frac{\text{Efficiency}_{\mu}^{\text{data}}}{\text{Efficiency}_{\mu}^{\text{MC}}} \quad (8.2)$$

To calculate the data-driven corrections for the MC trigger turn-on curves, the scale factors are fitted for each E_T^{miss} trigger starting in the range $100 \text{ GeV} < E_{T,\mu \text{ invis.}}^{\text{miss}} < 300 \text{ GeV}$ in E_T^{miss} bins of 10 GeV using the following fit function:

$$f(x) = p_0 \cdot \left[1 + \text{erf} \left(\frac{x - p_1}{\sqrt{2}p_2} \right) \right] + p_3 \quad (8.3)$$

where $x = E_{T,\mu \text{ invis.}}^{\text{miss}}$.

The scale factor applied to the MC in the signal region and one-lepton control region is given by evaluating $f(E_T^{\text{miss}})$ or $f(E_{T,\mu \text{ invis.}}^{\text{miss}})$, respectively. The scale factors measured for the different E_T^{miss} are shown in Fig. 8.2 together with the fitted SF curves. Good agreement

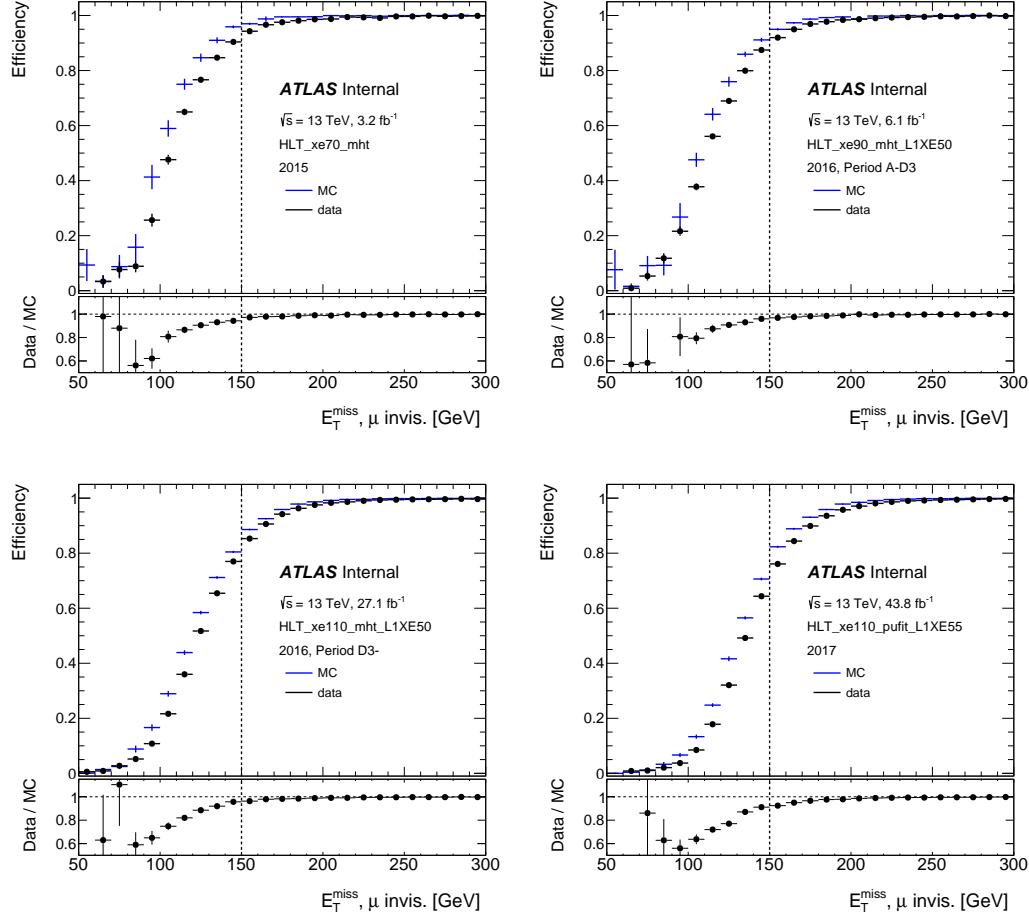


Figure 8.1: Measured trigger efficiencies as a function of offline $E_{T,\mu}^{\text{miss}}$ in data and MC for the E_T^{miss} triggers used in 2015-2018. The plots are shown for 0,1 and 2 tags together. The lower panels provide the ratio of data and MC events (the scale factor).

is observed in data and MC efficiencies comparison in Figure 8.3, after the scale factors are applied to the simulation.

The triggers are listed in Table 8.1.

Period	0 lepton	1 lepton	2 lepton + E_T^{miss} trigger SF measurement
2015	HLT_XE70_MHT	HLT_XE70_MHT	HLT_E24_LHMEDIUM_L1EM20VH OR HLT_E120_LHLOOSE OR HLT_MU20_ILOOSE_L1MU15 OR HLT_MU50
2016	HLT_XE90_MHT_L1XE50 (A)	HLT_XE90_MHT_L1XE50	HLT_E60_LHMEDIUM_NOD0 OR HLT_E140_LHLOOSE_NOD0 OR HLT_MU40 OR HLT_MU50
2016	HLT_XE90_MHT_L1XE50 (B-D3)	HLT_XE90_MHT_L1XE50	HLT_E60_LHMEDIUM_NOD0 OR HLT_E140_LHLOOSE_NOD0 OR HLT_MU24_IVARMEDIUM OR HLT_MU50
2016	HLT_XE110_MHT_L1XE50 (D4-E3)	HLT_XE110_MHT_L1XE50	HLT_E26_LHTIGHT_NOD0_IVARLOOSE OR HLT_E60_LHMEDIUM_NOD0 OR HLT_E140_LHLOOSE_NOD0 OR HLT_MU24_IVARMEDIUM OR HLT_MU26_IVARMEDIUM OR HLT_MU50
2016	HLT_XE110_MHT_L1XE50 (F1)	HLT_XE110_MHT_L1XE50	HLT_E26_LHTIGHT_NOD0_IVARLOOSE OR HLT_E60_LHMEDIUM_NOD0 OR HLT_E140_LHLOOSE_NOD0 OR HLT_MU26_IVARMEDIUM OR HLT_MU50
2016	HLT_XE110_MHT_L1XE50 (F2-)	HLT_XE110_MHT_L1XE50	HLT_E26_LHTIGHT_NOD0_IVARLOOSE OR HLT_E60_LHMEDIUM_NOD0 OR HLT_E140_LHLOOSE_NOD0 OR HLT_MU26_IVARMEDIUM OR HLT_MU50
2017	HLT_XE110_PUFIT_L1XE55	HLT_XE110_PUFIT_L1XE55	HLT_E60_LHMEDIUM_NOD0 OR HLT_E140_LHLOOSE_NOD0 OR HLT_MU26_IVARMEDIUM OR HLT_MU50
2018	HLT_XE110_PUFIT_70_L1XE55 (B-C5)	HLT_XE110_PUFIT_70_L1XE55	HLT_E60_LHMEDIUM_NOD0 OR HLT_E140_LHLOOSE_NOD0 OR HLT_MU26_IVARMEDIUM OR HLT_MU50
2018	HLT_XE110_PUFIT_65_L1XE55 (C5-)	HLT_XE110_PUFIT_65_L1XE55	HLT_E60_LHMEDIUM_NOD0 OR HLT_E140_LHLOOSE_NOD0 OR HLT_MU26_IVARMEDIUM OR HLT_MU50

Table 8.1: E_T^{miss} and single-lepton triggers used in the analysis.

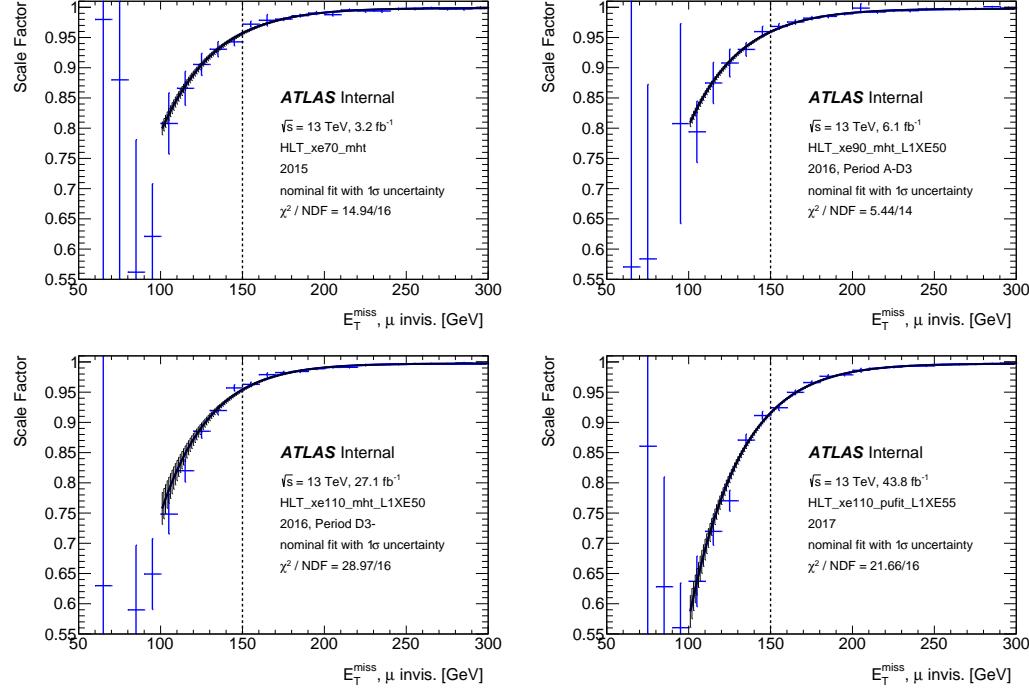


Figure 8.2: Measured scale factors as a function of offline $E_{T,\mu}^{\text{miss}}$ for the E_T^{miss} triggers used in 2015–2010. The scale factors were derived for 0,1 and 2 tags together. The hatched band shows the 1σ fit uncertainty.

8.3 MC samples

8.3.1 2HDM+a and Z'-2HDM signal

Simulated events corresponding to the $pp \rightarrow h\chi\bar{\chi}$ process were generated with MADGRAPH5_AMC@NLO (MG5_AMC) 2.6.1 [36] based on the UFO model developed in [31] at leading-order (LO) accuracy using the NNPDF 3.0 next-to-leading order (NLO) PDF set with $\alpha_s = 0.118$ [44]. The signal samples are generated separately for the loop-induced gluon–gluon fusion (ggF) process and the b -initiated production.

Simulated events corresponding to the $pp \rightarrow Z' \rightarrow hA(\chi\bar{\chi})$ process were generated with MADGRAPH5_AMC@NLO (MG5_AMC) 2.2.3 [36] at leading-order (LO) accuracy using the NNPDF 2.3 NLO PDF set with $\alpha_s = 0.119$ [43]. The 4-flavour scheme was used for the calculation of the matrix elements. Several samples were generated for different values of $m(Z')$ and $m(A)$. The masses of the additional Higgs bosons were fixed to $m(H) =$

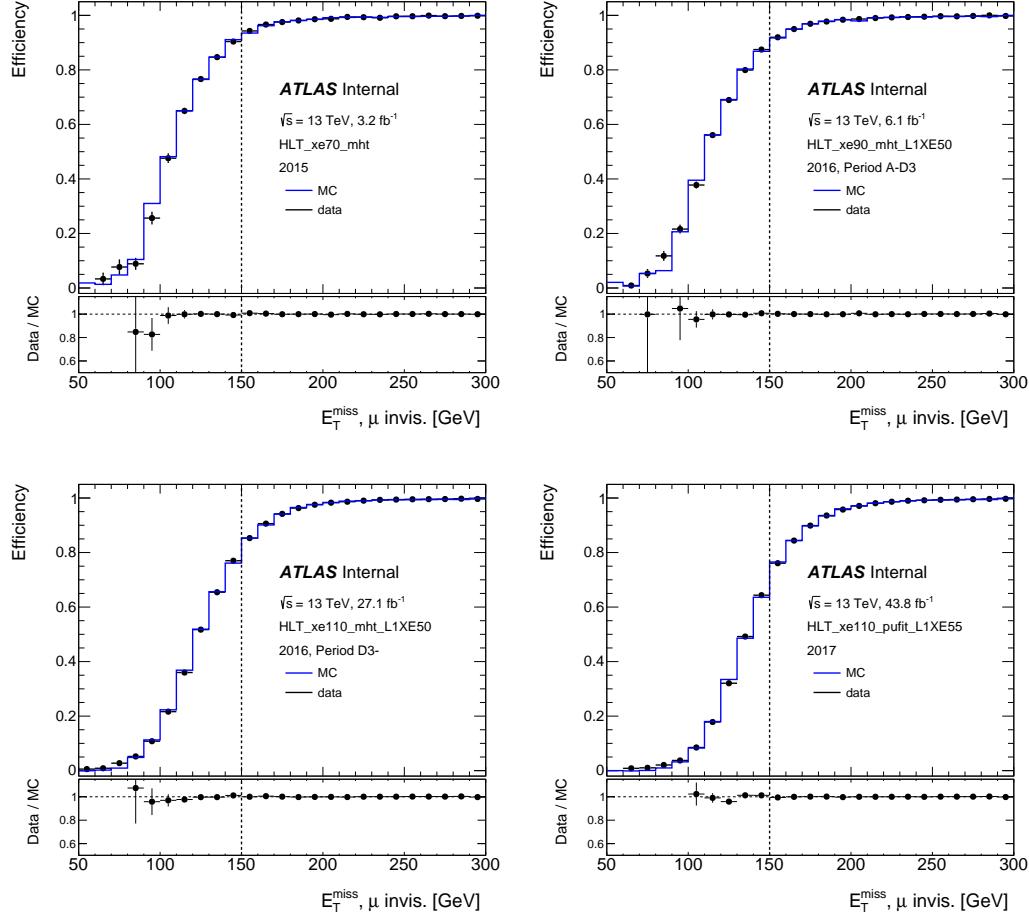


Figure 8.3: Validation plots showing E_T^{miss} trigger efficiencies and scale factors as function of offline $E_{T,\mu}^{\text{miss}}$ after applying scale factor corrections for $E_{T,\mu}^{\text{miss}} > 100 \text{ GeV}$ to the MC in the full single lepton region. Good agreement is observed between data and MC.

$m(H^\pm) = 300 \text{ GeV}$ and the mass of the DM candidates was fixed to $m(\chi) = 100 \text{ GeV}$.

The parton shower and hadronisation were simulated with PYTHIA 8.230 [115] and PYTHIA 8.186 [114], respectively for 2HDM+a and Z' 2HDM signal, using the A14 set [8] of tuned parameters together with the NNPDF 2.3 LO PDF set [42]. Higgs boson decays into $b\bar{b}$ pairs were also simulated with according PYTHIA versions with a branching fraction fixed to the SM prediction.

8.3.2 $V+jets$

$V+jets$ production is simulated with the SHERPA v2.2 [47] generator. In this setup, NLO-accurate matrix elements for up to two jets, and LO-accurate matrix elements for up to four jets are calculated with the Comix [82] and OpenLoops [57; 68] libraries. They are matched with the SHERPA parton shower [112] using the MEPS@NLO prescription [89; 90; 58; 91]. Samples are generated using the NNPDF3.0nnlo set [44].

The samples are split according to whether they contain a B hadron or no B with $p_T > 5$ GeV and $|\eta| < 2.9$, a C hadron with $p_T > 4$ GeV and $|\eta| < 3$ (with the filtered samples called Bfilter, CFilterBVeto, CVetoBVeto respectively). They are further split by either:

- using the transverse momentum of the V boson produced by Sherpa ($p_T(V)$), or
- using the max $p_T(V), H_T$, where H_T is the scalar sum of the p_T of the vector boson and the jets.

8.3.3 $t\bar{t}$

The production of $t\bar{t}$ events is modelled using the POWHEGBOX [78; 103; 77; 33] v2 generator which provides matrix elements at NLO with the NNPDF3.0NLO [44] parton distribution function (PDF) and the h_{damp} parameter¹ set to $1.5 m_{\text{top}}$ [41]. The functional form of the renormalisation and factorisation scale is set to the default scale $\sqrt{m_{\text{top}}^2 + p_T^2}$. The events are interfaced with PYTHIA8.230 [115] for the parton shower and hadronisation, using the A14 set of tuned parameters [8] and the NNPDF23LO PDF set. The decays of bottom and charm hadrons are simulated using the EvtGEN v1.6.0 program [67].

The events are filtered to select dilepton or semi-leptonic $t\bar{t}$ decays. Two sets of samples are used in the analysis

- an inclusive $t\bar{t}$ sample
- a set of $E_{\text{T}}^{\text{miss}}$ -filtered $t\bar{t}$ samples, which was not used in the previous iteration of the

¹The h_{damp} parameter controls the transverse momentum p_T of the first additional emission beyond the leading-order Feynman diagram in the parton shower and therefore regulates the high- p_T emission against which the $t\bar{t}$ system recoils.

analysis [19], and was specifically developed to reduce the MC statistical uncertainties in the high- E_T^{miss} region.

8.3.4 Single top

Single-top tW associated production, single-top t-channel and s-channel production are all modelled using the PowhegBox [74; 103; 77; 33] v2 generator which provides matrix elements at NLO with the NNPDF3.0NLOnf4 [44] PDF set. The diagram removal scheme [76] was employed to handle the interference with $t\bar{t}$ production [41]. The events are interfaced with PYTHIA8 .230 [115] using the A14 tune [8] and the NNPDF23LO PDF set. The decays of bottom and charm hadrons are simulated using the EVTGEN v1.6.0 program [67]. For the t-channel, the functional form of the renormalisation and factorisation scale is set to $\sqrt{m_b^2 + p_{T,b}^2}$ following the recommendation of [74] while for the other two, the functional form of the renormalisation and factorisation scale is set to the default scale, which is equal to the top-quark mass.

8.3.5 Diboson

Semileptonically decaying diboson and loop-induced diboson samples are simulated with the SHERPA v2.2 [47] generator. In this setup multiple matrix elements are matched and merged with the SHERPA parton shower based on Catani-Seymour dipole [82; 112] using the MEPS@NLO prescription [89; 90; 58; 91]. The virtual QCD correction for matrix elements at NLO accuracy are provided by the OPENLOOPS library [57; 68]. Samples are generated using the NNPDF3.0nnlo set [44].

8.3.6 SM $Vh(b\bar{b})$

The production of $q\bar{q} \rightarrow Wh \rightarrow \ell\nu b\bar{b}$ and $q\bar{q} \rightarrow Zh \rightarrow \ell\ell b\bar{b}$ events is modelled using the PowhegBox v2 generator [33] using the Powheg MiNLO procedure [84; 97] with the NNPDF3.0NLO [44] PDF set. The events are interfaced with PYTHIA8 .212 [115] using the AZNLO tune [23] and the CTEQ6L1 [108] PDF set.

The loop-induced $gg \rightarrow Zh \rightarrow \ell b\bar{b}, \nu\bar{\nu} b\bar{b}$ process is modelled using the PowhegBox v2 generator [33] with the NNPDF3.0NLO [44] PDF set. Parton showering and hadronisation

are provided by PYTHIA8 .212 [115] with the same settings as the one used for the $q\bar{q}$ process.

8.3.7 $t\bar{t} + Z/H$

$t\bar{t}H$ events are modelled using the POWHEGBOX [78; 103; 77; 33; 85] generator at NLO with the NNPDF3.0NLO [44] PDF set. The events are interfaced with PYTHIA8 .230 [115] using the A14 tune [8] and the NNPDF2.3LO [44] PDF set.

$t\bar{t}V$ events are modelled using the MADGRAPH5_AMC@NLO v2.3.3 [36] generator which provides matrix elements at next-to-leading order (NLO) with the NNPDF3.0NLO [44] parton distribution function (PDF). The functional form of the renormalization and factorization scale is set to the default scale $0.5 \times \sum_i \sqrt{m_i^2 + p_{T,i}^2}$. Top quarks are decayed at LO using MADSPIN [75; 38] to preserve all spin correlations. The events are interfaced with PYTHIA8 .210 [115] for the parton shower and hadronisation, using the A14 set of tuned parameters [8] and the NNPDF23LO [44] PDF set. The decays of bottom and charm hadrons are simulated using the EVTGEN v1.2.0 program [67].

Chapter 9

Signal selection

Signal selection is a process to select target events with multiple filtering criteria based on the event signature. In this article, in order to obtain a high signal over background efficiencies, the physics object level selection is applied first, then, signal regions are designed. Moreover, control regions are also defined to constrain the background contribution.

9.1 Physics objects definition

As mentioned previously, leptons, jets and missing transverse momentum are the variables that are helpful to increase signal selection efficiencies.

9.1.1 Leptons

- **Electrons:** As mentioned in Section 6.2, electrons can be identified using various likelihood-base criteria, for example, shower profile selections, track quality, and high threshold TRT hits. The electron identification criteria that used in this analysis are listed In Table 9.1.
- **Muons:** Muons are reconstructed with a combined information from inner tracker and muon spectrometer. Moreover, two different working points are applied in this analysis for muon identification. A loose criteria is applied as baseline muon selection to obtain larger acceptance, while medium working point is used in signal selection for higher purity. More details are listed in Table 9.2.

- **τ -Leptons:** As described in Section 6.4, τ -lepton is reconstructed using inner tracker and calorimeter. Since the τ -lepton is vetoed in both signal and control regions, a loose τ -lepton working point is enough for this analysis. More information can be found in Table 9.3.

Table 9.1: Electron selection criteria.

Feature	Criterion
Pseudorapidity range	$ \eta < 2.47$
Transverse momentum	$p_T > 7\text{GeV}$
Track to vertex association	$ d_0^{\text{BL}}(\sigma) < 5$ $ \Delta z_0^{\text{BL}} \sin \theta < 0.5\text{mm}$
Identification	<code>FCLoose</code>
Isolation	<code>LooseTrackOnly / FCHighPtCaloOnly</code>

Table 9.2: Muon selection criteria.

Feature	Baseline criterion	Signal criterion
Selection working point	<code>Loose</code>	<code>Medium</code>
Isolation working point	<code>FCLoose</code>	<code>FCTightTrackOnly</code>
Momentum calibration	Sagitta correction	Sagitta correction
p_T cut	7GeV	7GeV
$ \eta $ cut	< 2.7	< 2.5
d_0 significance cut	3	3
z_0 cut	0.5mm	0.5mm

9.1.2 Jets

- **Small-radius jets:** As mentioned in Section 6.5, the small-radius jets can be divided into two categories: central jets and forward jets. In this analysis, only the central jets are used to reconstruct Higgs boson, while both central and forward jets are involved

Table 9.3: Tau selection criteria.

Feature	Criterion
Pseudorapidity range	$ \eta < 2.5$
Track selection	1 or 3 tracks
Charge	$ Q = 1$
Tau energy scale	MVA TES
Transverse momentum	$p_T > 20\text{GeV}$
Jet rejection	BDT-based (Loose)
Electron rejection	BDT-based
Muon rejection	Via overlap removal in $\Delta R < 0.2$ and $p_T > 2\text{GeV}$. Muons must not be Calo-tagged

in E_T^{miss} calculation. More information about small-radius jets identification can be found in Table 9.4.

- **Large-radius jets:** The large-radius jets are reconstructed with large cone size ($R=1.0$). These large cone size jets are used to reconstruct the boosted Higgs boson. More details are listed in Table 9.5.
- **Variable-radius track jets:** The variable-radius track jets are reconstructed from inner tracker using anti- k_t algorithm. The jet radius is dependent on the jet p_T : $R = \frac{30\text{GeV}}{p_T}$. The variable-radius track jets provide a better acceptance when reconstructing boosted Higgs boson in this analysis.
- **b-jets:** The b-tagging is a technology to identify the b-jets, which is applied on central jets to reconstruct the resolved Higgs boson. More detailed information can be found in Table 9.6.

9.1.3 Missing transverse momentum

As described in Section 6.6, the missing transverse momentum (E_T^{miss}) is defined as the magnitude of missing transverse momentum vector, which is calculated by all well-reconstructed

Table 9.4: Small- R jet reconstruction criteria.

Feature	Criterion
Algorithm	Anti- k_t
R -parameter	0.4
Input constituent	PFlow
CalibArea tag	00-04-82
Calibration configuration	JES_MC16Recommendation_Consolidated_EMTopo_Apr2019_Rel21.config
Calibration sequence (Data)	JetArea_Residual_EtaJES_GSC_Insitu
Calibration sequence (MC)	JetArea_Residual_EtaJES_GSC
Jet cleaning	TightBad
p_T	$> 20\text{GeV}$ (central) / $> 30\text{GeV}$ (forward)
$ \eta $	< 2.5 (central) / $2.5 < \eta < 4.5$ (forward)
JVT	Medium working point, applied only to central jets with $p_T < 120\text{GeV}$

physics object, like electron, muon and jets. More information about E_T^{miss} reconstruction is listed in Table 9.7.

9.2 Signal regions

Signal regions are a set of phase spaces that is defined by a combination of selection conditions for signal selection. In this analysis, the most important signature is Higgs boson, which have different signatures under different energy scales. Therefore, Resolved region and Merged region are designed as sub signal regions to probe signals effectively under different Higgs boson momentum signatures.

9.2.1 Common event selection

Several common event selection criteria are applied for both resolved and merged regions as a baseline selection.

Table 9.5: Large- R jet reconstruction criteria.

Feature	Criterion
Algorithm	Anti- k_t
R-parameter	1.0
Input constituent	LCTopo
Grooming algorithm	Trimming
Subjet p_T fraction for trimming	0.05
R_{trim}	0.2
CalibArea tag	00-04-82
Calibration configuration (Data)	JES_MC16recommendation_FatJet _Trimmed_JMS_comb_3April2019.config
Calibration configuration (MC)	JES_MC16recommendation_FatJet_ Trimmed_JMS_comb_17Oct2018.config
Calibration sequence (Data)	EtaJES_JMS_In situ
Calibration sequence (MC)	EtaJES_JMS
p_T	$> 200 GeV$
$ \eta $	< 2

- **Event cleaning:** Event cleaning is a process to apply a set of filters to veto the corrupted events. In this analysis, LAr noise burst, tile corruption, SCT recovery procedure and incomplete events are applied as the event cleaning filters. More information can be found in [61].
- **Loose jet Veto:** In analysis, the reconstructed jets can be a fake jet from noise, collision background, etc.. TightBadJet criteria, provided by Jet/ E_T^{miss} group, is applied in this analysis to veto events that contain the fake jets.
- **E_T^{miss} selection:** In signal regions, a constrain $E_T^{\text{miss}} > 150 GeV$ is applied. In control region, the lepton momentum is also treated as part of missing transverse momentum and same E_T^{miss} cut is applied.

Table 9.6: b-jets selection criteria.

Feature	Criterion
Jet collection	AntiKt4EMPFlow / AntiKtVR30Rmax4Rmin02
Algorithm	DL1
Operating point	Eff = 77
CDI	2017-21-13TeV-MC16-CDI-2019-07-30_v1

Table 9.7: E_T^{miss} reconstruction criteria.

Parameter	Value
Algorithm	Calo-based
Soft term	Track-based (TST)
E_T^{miss} operating point	Tight

- **Light lepton and τ -lepton veto:** Events that contain a electron, muon or τ -lepton, as defined in Section 9.1, are rejected.
- **Extended τ -lepton veto:** An extended τ veto is applied in case of τ -lepton failed to be identified. A hadronically decayed τ can be viewed as a jet. Therefore, a jet is considered as a τ candidate if it fullfills two conditions: (1) The track multiplicity in the jet cone is between 1 to 4. (2) The angular separation between the jet and E_T^{miss} is $\Delta\phi \leq 22.5^\circ$. The cut on the track multiplicity in the jet cone makes sure that the hadronic decay products in the jet are compatible with the τ -lepton decay products, namely with charged pions. The tracks considered are associated with the primary vertex and have a p_T greater than 1 GeV. The cut on the angular separation between jet and E_T^{miss} makes sure that the τ -candidate comes from a W boson.
- $\min_{j \in \{1,2,3\}} \Delta\phi(E_T^{\text{miss}}, j)$: QCD multijet events can be generated by jet energy mis-measurement. In case of fake E_T^{miss} , the E_T^{miss} will generally point to the direction of leading jets. Therefore, an angular cut between E_T^{miss} and leading jets is applied to veto these events.

After these baseline selection, signal region needed to be analyzed on different sub regions based on Higgs boson signature. In the target model, the sum of E_T^{miss} vector and Higgs boson momentum vector should be zero based on momentum conservation. Therefore, one can apply a single cut on E_T^{miss} to separate events with resolved or merged Higgs boson. In this analysis, the cut $E_T^{\text{miss}} < 500\text{GeV}$ is defined as resolved region while $E_T^{\text{miss}} > 500\text{GeV}$ is merged region. The Higgs boson identification, which will be demonstrate in the rest of this section, is different in these two signal regions.

Moreover, there is one additional condition to select events with at lease one b-jets. The b-tagging is applied on all small-radius central jets in resolved region, while only on two leading track jets that associated to large-radius jet in merged region.

9.2.2 Resolved region

In resolved region, the Higgs boson momentum is relative low. Therefore, the Higgs boson decay products, two b-jets, can be resolved into two separate jets. To reconstruct the Higgs boson, a set of jet candidates needed to be selected first. There are two categories in the Higgs boson jets candidate set: (1) b-tagged jets, (2) central non b-tagged jets. The jets are ordered in decreasing transverse momentum within each category. The first two out of this set of jets (referred as j_1 and j_2 below) are used to reconstruct Higgs boson candidate, which referred as H_{reco} .

There are several criteria to select Higgs boson in resolved channel:

- **p_T of H_{reco} :** p_T of H_{reco} is expected to increase with E_T^{miss} . Therefore, a cut $p_T(H_{\text{reco}}) > 100\text{GeV}$ is applied when $E_T^{\text{miss}} < 350\text{GeV}$, while $p_T(H_{\text{reco}}) > 300\text{GeV}$ when $350\text{GeV} < E_T^{\text{miss}} < 500\text{GeV}$.
- **m_{jj} :** A window cut on the mass of reconstructed Higgs boson is helpful to identify the Higgs boson. The mass of reconstructed Higgs boson can be calculated from the invariant mass of jets pair (m_{jj}) that form H_{reco} . A window cut $50\text{GeV} < m_{jj} < 280\text{GeV}$ is applied in this analysis.
- **m_T^b :** In resolved region, the major background is top quark pair production. m_T^b is introduced to reject these background events. The m_T^b is defined as: $m_T^b =$

$\sqrt{2p_T^b E_T^{\text{miss}}(1 - \cos\Delta\phi(\vec{p}_T, \vec{E}_T^{\text{miss}}))}$. The m_T^b is calculated on the b-jets that closest to and furtherest from E_T^{miss} . The closest b-jet m_T^b is required to be greater than 170 GeV, while the furtherest cut is greater than 200 GeV.

In addition, it is possible that the measured E_T^{miss} of an event is due to a resolution fluctuation. Therefore, a E_T^{miss} significance cut $S > 12$ is introduced to reject fake E_T^{miss} events.

9.2.3 Merged region

In merged region, the momentum of Higgs boson is relative high, and therefore, the products of this high energy Higgs boson are merged into one fat cone, which is reconstructed as a large-radius jet in detector.

Several criteria are applied on the large-radius to identify the Higgs boson in merged region:

- m_J : The invariant mass of the large-radius jet is an expression of Higgs boson mass. A window cut $50\text{GeV} < m_J < 270\text{GeV}$ is applied.
- **Variable-radius track jets association**: At least two variable-radius track jets needed to be associated to the large-radius jet.

As a summary of Section 9.2, both resolved and merged signal regions are listed in Table 9.8.

9.3 Control regions

Control regions are a set of phase spaces that are orthogonal to signal regions. Control regions are designed to estimate the contribution of backgrounds in signal regions.

9.3.1 1-lepton channel

In the signal region the W+jets related background is mostly composed of events in which the W -boson decays leptonically. The $t\bar{t}$ background in the signal region originates almost only from semileptonic $t\bar{t}$ decays. Events with leptons can enter the signal region if the

Resolved	Merged
lowest unprescaled E_T^{miss} trigger	
veto on baseline light leptons and τ -leptons	
extended τ -veto	
$E_T^{\text{miss}} > 150\text{GeV}$	
$\min_{j \in \{1,2,3\}} \Delta\phi(E_T^{\text{miss}}, j) > 20^\circ$	
$E_T^{\text{miss}} < 500\text{GeV}$	$E_T^{\text{miss}} > 500\text{GeV}$
$N(\text{central small-R jets}) \geq 2$	$N(\text{central large-R jets}) \geq 1$
$N(b\text{-tagged small-R jets}) \geq 1$	$N(b\text{-tagged associated track jets}) \geq 1$
$S > 12$	—
$p_T(jj) > 100\text{GeV}$ if $E_T^{\text{miss}} < 350\text{GeV}$	—
$p_T(jj) > 300\text{GeV}$ if $E_T^{\text{miss}} \geq 350\text{GeV}$	—
$m_T^{b,\text{min}} > 170\text{GeV}$	—
$m_T^{b,\text{max}} > 200\text{GeV}$	—
$50\text{GeV} < m_{jj} < 280\text{GeV}$	$50\text{GeV} < m_J < 270\text{GeV}$
b-tagging on small-radius jets	b-tagging on variable-radius track jets

Table 9.8: Summary of the resolved and merged event selection applied in the 0-lepton channel. S denotes the object based E_T^{miss} significance.

leptons fail to be identified or if they are outside the detector acceptance. The largest contribution comes from decays involving τ -leptons, as shown in Figure 9.1. Figure 9.2 shows the distribution of different kinematic variables for the dominant decay modes. From these it can be seen that shapes of the event variables agree well within the statistical uncertainties. Therefore the $W + \text{jets}$ and $t\bar{t}$ backgrounds can be estimated using a 1-lepton control region, in which the lepton could have any flavor in principle.

The 1-muon control region is chosen to constrain $W + \text{jets}$ related background. Events in this control region is required to have exactly muon and no other baseline leptons.

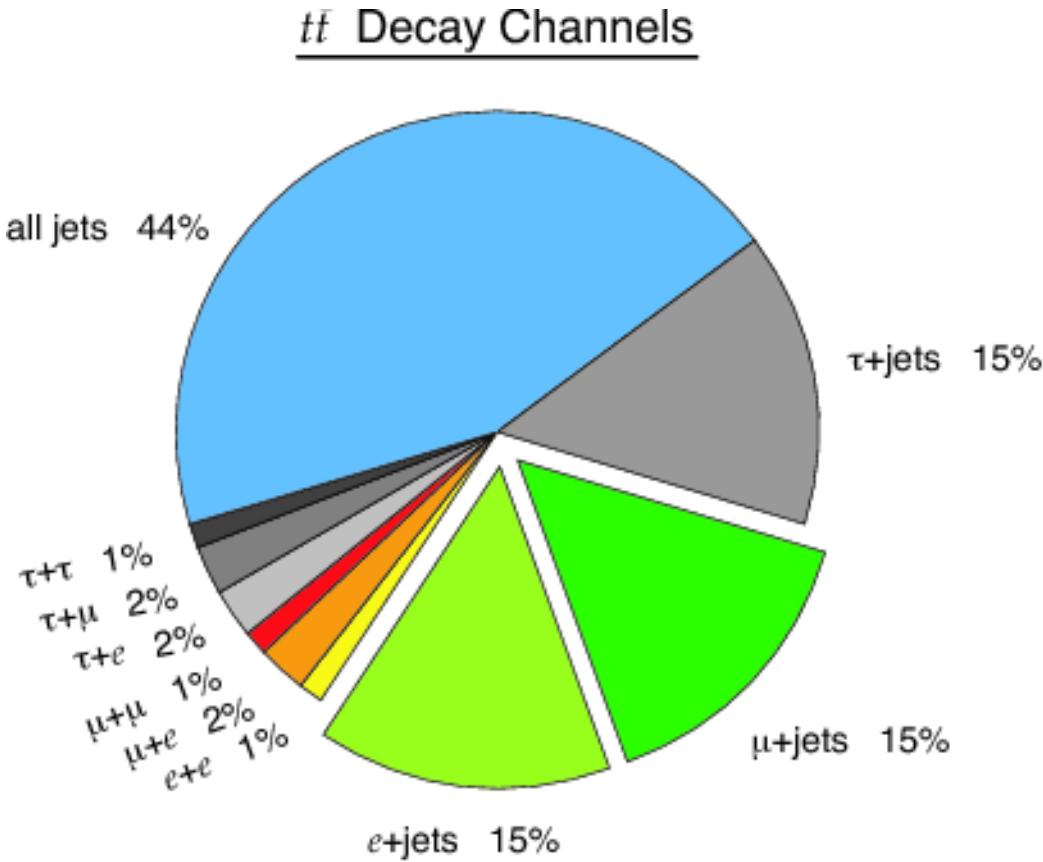


Figure 9.1: Decay categories for $t\bar{t}$ events in the signal region. The classification of the decay processes was performed by examining the truth particle information within the reconstructed signal region events. The three dominant decay modes are the semileptonic decays with either a light lepton, a leptonically or a hadronically decaying τ -lepton.

9.3.2 2-lepton channel

A 2-lepton control region is used to estimate the $Z + \text{jets}$ background. In the signal region the $Z(\rightarrow \nu\nu) + \text{jets}$ production leads to a significant amount of background events, which have the same decay topology as $Z(\rightarrow \ell\ell) + \text{jets}$, because the momentum of the Z boson does not depend on its decay products. Hence the normalisation of $Z(\rightarrow \nu\nu) + \text{jets}$ events can be estimated with the help of a $Z(\rightarrow \ell\ell) + \text{jets}$ control region.

The 2-lepton control region is defined by selecting exactly two muons or two electrons. At least one of the lepton need to satisfy the signal lepton criteria with enough p_T to pass

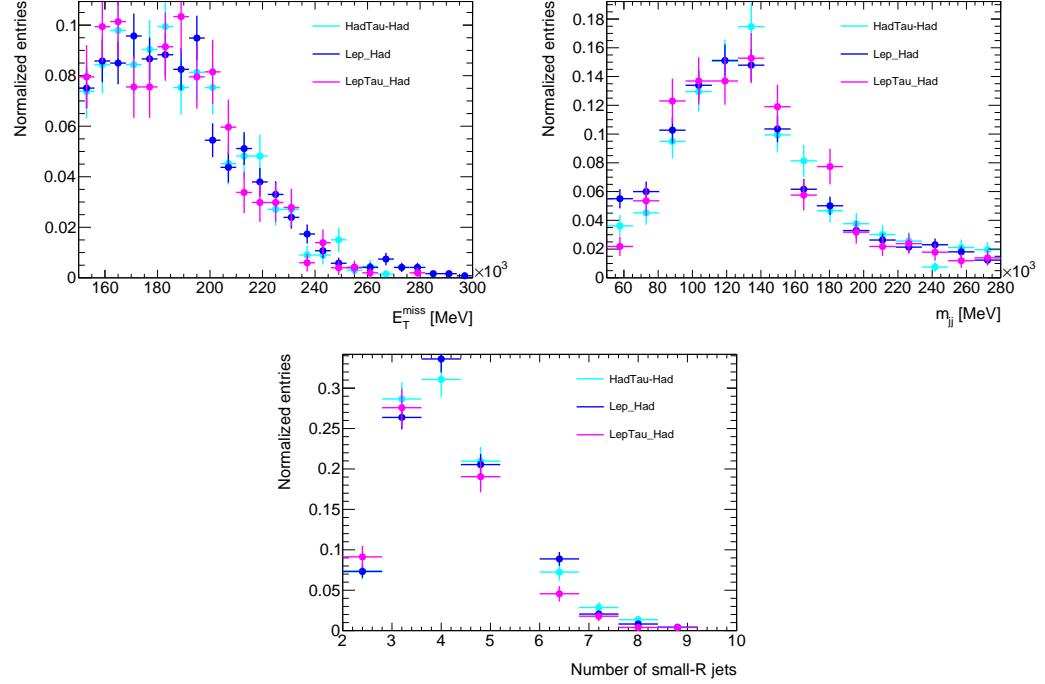


Figure 9.2: Normalised distributions of E_T^{miss} , the Higgs boson candidate mass and the number of small-radius jets for $t\bar{t}$ events. The distributions are shown separately for the three semileptonic decay modes, which are the dominant decay modes in the signal region. The shapes of the distributions agree well within the statistical uncertainties.

the trigger. The signal muon is required to have $p_T > 25\text{GeV}$, while signal electron is $p_T > 27\text{GeV}$. The other lepton is not required to pass signal lepton selection to increase the acceptance. More over, since the two lepton is supposed to reconstruct Z boson, a mass window requirement is implemented to identify the Z mass peak: $|m_Z - m_{ll}| < 10\text{GeV}$.

Chapter 10

Systematic uncertainty

Once signal regions are determined, one can use the data observations and its systematic uncertainties in signal regions to do statistical analysis. Data observation is easy to obtain with signal region determined, while its systematic uncertainties are not trivial to calculate. In this analysis, both experimental and theoretical uncertainties are considered.

10.1 Experimental systematics

Experimental uncertainty is the systematic uncertainty that comes from reconstruction of physics objects. It is a reflection of uncertainty related to detectors or reconstruction algorithms. Multiple physical objects, like muon, electron, jets, E_T^{miss} , etc., are used in this analysis. Therefore, the systematic uncertainties of the reconstruction of these variables needed to be considered carefully when interpreting the analysis result. A summary of these experimental uncertainties can be viewed in Table 10.1, Table 10.2 and Table 10.3.

10.2 Theoretical systematics

Theoretical uncertainty is the systematic uncertainty that comes from Monte-Carlo simulation of background and signal processes. In this analysis, both the signal and background yields come from monte carlo simulation samples rather than data driven estimation from control regions. Therefore, a careful estimation on theoretical uncertainties is needed on all

related simulation processes.

The sources of theoretical uncertainty are listed below:

- **Missing higher orders in the calculation of the inclusive matrix elements:**

For all processes the calculation of the cross-section relies on a perturbative expansion of the scattering matrix, which is truncated at a certain order. The effect of the missing higher orders is estimated by varying the renormalisation and factorisation scales (μ_R and μ_F) independently by a factor of 2, excluding the $(\mu_R, \mu_F) = (\frac{1}{2}, 2), (2, \frac{1}{2}) \times \mu_{\text{central}}$ variations, which may lead to large logarithms.

- **Uncertainties from the choice of PDFs and α_s :** This kind of uncertainty arises from uncertainties in the experimental measurements that are used to determine the PDF sets used in each calculation, uncertainties from the choice of the functional form used in the PDF fits, and uncertainties associated to the experimental determination of α_s . These are estimated using the PDF4LHC prescription[51].

- **Merging scale uncertainties:** For samples generated by merging matrix elements (ME) corresponding to different multiplicities, e.g. $V + jets$, an uncertainty related to the choice of the merging scale, i.e. the scale that separates soft from hard jets, is evaluated by varying the merging scale by a factor of 2 up and down.

- **Resummation scale uncertainties:** For SHERPA[81] samples an additional uncertainty related to the energy cut-off for the integration of MC counterterms in the parton shower (PS).

- **Matching uncertainties:** For samples generated using a NLO matrix element and matched to a parton shower, a comparison between a POWHEG and a MG5AMC sample can probe uncertainties related to the ME/PS matching procedure.

- **Parton shower/Hadronisation uncertainties:** Uncertainties related to algorithmic or parametric differences in the modelling of the PS and hadronisation can be assessed by comparing samples generated with different showering/hadronisation (SHG) generators, typically PYTHIA 8 with HERWIG 7.

- **Eigentune uncertainties:** This kind of uncertainties are related to uncertainties in the choice of the free parameters that are used in the SHG programmes, derived so as to encompass the data used in the ATLAS tuning program[8].
- **Other implementation-specific uncertainties:** The variation of the h_{damp} scale in the Powheg samples, etc..

The sources mentioned above are considered and grouped into Monte-carlo template, which is used as an input of the likelihood fit. There are four types of uncertainties considered in the statistical interpretation:

1. **Inclusive cross-section uncertainties:** Inclusive cross-section uncertainties are implemented in the fit using Gaussian priors that affect the normalisation of a given samples in all regions in a correlated manner (`OverallSys` in the HISTFACTORY terminology [66]). These uncertainties are applied only on the samples whose normalisation is not freely floating in the fit.
2. **Shape uncertainties:** Shape uncertainties are uncertainties on the shape of the fitted discriminant ($m(b\bar{b})$). These uncertainties are estimated by comparing the shape of the fitted variables for the nominal MC samples and the alternative samples that probe the uncertainties that mentioned as sources of theoretical uncertainties. A comparison of the $m(b\bar{b})$ distribution between the nominal and the alternative MCs provide additional templates (histograms) that define the $\pm 1\sigma$ variations (`HistoSys` in the HISTFACTORY terminology)
3. **Relative acceptance uncertainties:** The theory uncertainties can also alter the shape of the observables used to separate the different fit regions. These shape differences induce normalisation/acceptance differences between the regions that are used in the fit. For example, differences in the E_T^{miss} shape can induce relative acceptance differences between the adjacent E_T^{miss} bins and differences in the lepton p_T spectrum can induce relative acceptance differences between the 0 and 1-lepton channels. These are included in the fit as Gaussian priors (`OverallSys`) whose magnitude is estimated

by

$$\sigma_{\text{accept}} = \sqrt{\sum_i \left(1 - \frac{N_A^{\text{alt},i}}{N_B^{\text{alt},i}} \middle/ \frac{N_A^{\text{nom}}}{N_B^{\text{nom}}} \right)^2}, \quad (10.1)$$

where i runs over all alternative MC generators considered for a given process and $(A, B) = (\text{SR,CR1}), (\text{CR1,CR2}), (E_T^{\text{miss}} \text{ bin1}, E_T^{\text{miss}} \text{ bin2}), (E_T^{\text{miss}} \text{ bin2}, E_T^{\text{miss}} \text{ bin3})$, (resolved, merged). Since the uncertainty is relative between region A and region B it is only applied on region B in the fit.

4. **Flavour composition uncertainties:** An additional uncertainty on the flavour composition is assigned on the W and $Z+\text{heavy}$ flavour components (Zhf, Whf), which consist of bb, cc, bc, bl , in order to allow the cc, bc, bl components to vary individually from the total Zhf, Whf normalisation which is freely floating in the fit.

10.3 Data-MC comparison

After obtaining the data and monte-carlo simulation yields and evaluating the systematic uncertainties, data versus monte-carlo simulation comparison, as an important validation of this analysis, can be demonstrated. The data-MC comparison is shown along several physics variables, which are used in this analysis, in both signal regions and control regions. One can view the signal regions comparison along m_{jj} or m_J in Figure 10.1 and Figure 10.2, 1-lepton control regions comparison along muon charge in Figure 10.3 and Figure 10.4, and 2-lepton control regions comparison along number of b-jets in Figure 10.5 and Figure 10.6.

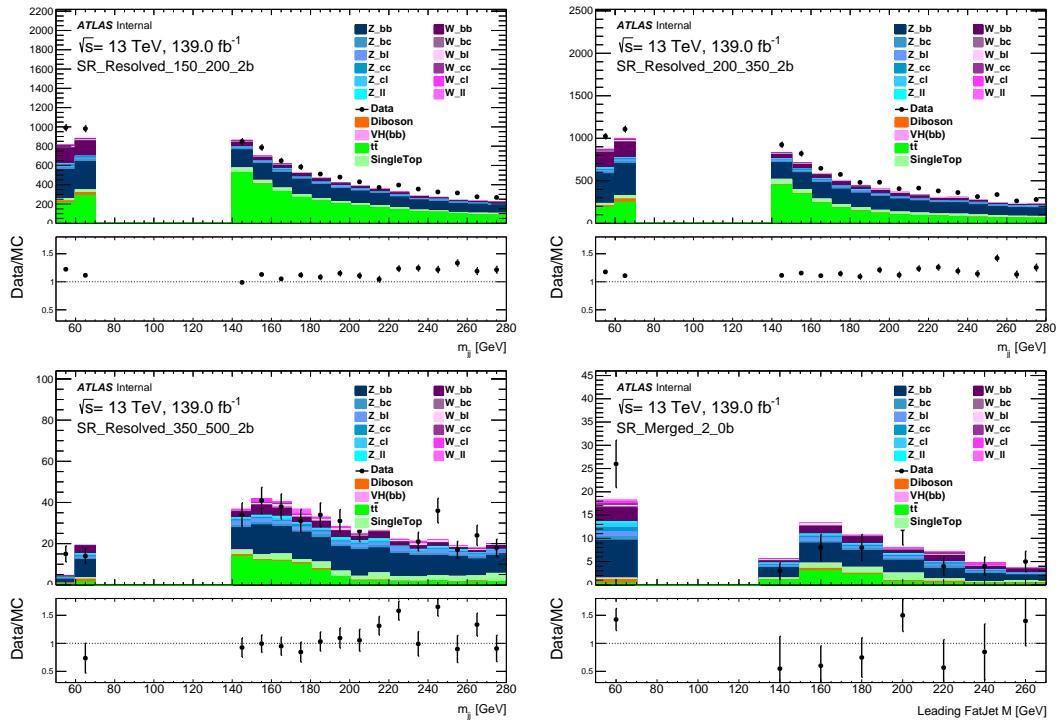


Figure 10.1: Higgs candidate mass spectra in the different E_T^{miss} regions with 2 b -tagged jets in the 0-lepton channel.

Systematic uncertainty	Short description
Event	
Luminosity	Uncertainty on the total integrated luminosity
Electrons	
EL_EFF_Trigger_TOTAL_1NPCOR_PLUSUNCOR	Trigger efficiency uncertainty
EL_EFF_Reco_TOTAL_1NPCOR_PLUSUNCOR	Reconstruction efficiency uncertainty
EL_EFF_ID_TOTAL_1NPCOR_PLUSUNCOR	ID efficiency uncertainty
EL_EFF_Iso_TOTAL_1NPCOR_PLUSUNCOR	Isolation efficiency uncertainty
EG_SCALE_ALL	Energy scale uncertainty
EG_RESOLUTION_ALL	Energy resolution uncertainty
Muons	
mu20_iloose_L1MU15_OR_HLT_mu40_MUON_EFF_Trig	Trigger efficiency uncertainties
mu24_ivarmed_OR_HLT_mu40_MU_EFF_TrigStat	None
mu24_ivarmed_OR_HLT_mu50_MU_EFF_TrigStat	None
mu26_ivarmed_OR_HLT_mu50_MU_EFF_TrigStat	None
MUON_EFF_RECO_STAT	Reconstruction uncertainty for $p_T > 15\text{GeV}$
MUON_EFF_RECO_SYS	None
MUON_EFF_RECO_STAT_LOWPT	Reconstruction and ID efficiency uncertainty for $p_T < 15\text{GeV}$
MUON_EFF_RECO_SYS_LOWPT	None
MUON_ISO_STAT	Isolation efficiency uncertainty
MUON_ISO_SYS	None
MUON_TTVA_STAT	Track-to-vertex association efficiency uncertainty
MUON_TTVA_SYS	None
MUONS SCALE	Energy scale uncertainty
MUONS_SAGITTA_RHO	Variations in the scale of the momentum
MUONS_SAGITTA_RESBIAS	Variations in the scale of the momentum
MUONS_ID	Energy resolution uncertainty from inner detector
MUONS_MS	Energy resolution uncertainty from muon system
E_T^{miss} related	
METTrigStat	Trigger efficiency uncertainty
METTrigSyst	None
MET_SoftTrk_ResoPerp	Track-based soft term related to transversal resolution uncertainty
MET_SoftTrk_ResoPara	Track-based soft term related to longitudinal resolution uncertainty
MET_SoftTrk_Scale	Track-based soft term related to longitudinal scale uncertainty
MET_JetTrk_Scale	Track E_T^{miss} scale uncertainty due to tracks in jets
PRW_DATASF	Uncertainty on data scale factor used for the computation of pileup reweighting

Table 10.1: A summary of the experimental systematic uncertainties.

Systematic uncertainty	Short description
Small-radius jets	
JET_EtaIntercalibration_Modelling	η -intercalibration: MC generator modelling uncertainty
JET_EtaIntercalibration_TotalStat	η -intercalibration: statistical uncertainty
JET_EtaIntercalibration_NonClosure_highE	η -intercalibration: non-closure uncertainty of jet response, high energy component
JET_EtaIntercalibration_NonClosure_negEta	η -intercalibration: non-closure uncertainty of jet response, negative η component
JET_EtaIntercalibration_NonClosure_posEta	η -intercalibration: non-closure uncertainty of jet response, positive η component
JET_Pileup_OffsetMu	Pileup: Offset, term for number of interactions per crossing μ
JET_Pileup_OffsetNPV	Pileup: Offset, term for number of primary vertices
JET_Pileup_PtTerm	Pileup: Offset, p_T term
JET_Pileup_RhoTopology	Pileup: Offset, ρ topology uncertainty on jet areas
JET_Flavor_Composition	Flavor composition uncertainty
JET_Flavor_Response	Flavor response uncertainty (dominated by gluon response)
JET_PunchThrough_MC16	Punch-through correction uncertainty
JET_EffectiveNP_Statistical	Statistical components of effective jet energy scale uncertainties, split into 6 components
JET_EffectiveNP_Modelling	Modelling components of effective jet energy scale uncertainties, split into 4 components
JET_EffectiveNP_Detector	Detector components of effective jet energy scale uncertainties, split into 2 components
JET_EffectiveNP_Mixed	Effective jet energy scale uncertainties coming from various sources, split into 3 components
JET_SingleParticle_HighPt	Uncertainty related to high p_T jets
JET_RelativeNonClosure_MC16	Closure of the calibration, relative to MC12a
JET_BJES_Response	Jet energy scale uncertainty for b -jets
JET_JER_DataVsMC_MC16	Nuisance parameter covering when jet energy resolution in data smaller than resolution in MC
JET_JER_EffectiveNP	Effective jet energy resolution uncertainty; split into 6 components
FT_EFF_EIGEN_B	b -tagging efficiency uncertainties ("BTAG_MEDIUM")
FT_EFF_EIGEN_C	None
FT_EFF_EIGEN_L	None
FT_EFF_EIGEN_extrapolation	b -tagging efficiency uncertainty on the extrapolation on high p_T -jets
FT_EFF_EIGEN_extrapolation_from_charm	b -tagging efficiency uncertainty on τ -jets

Table 10.2: A summary of the experimental systematic uncertainties.

Systematic uncertainty	Short description
Large-radius jets	
JET_EtaIntercalibration_Modelling	η -intercalibration: MC generator modelling and method uncertainty
JET_EtaIntercalibration_R10_TotalStat	η -intercalibration: statistical uncertainty
JET_Flavor_Composition	Flavor composition uncertainty
JET_Flavor_Response	Flavor response uncertainty (dominated by gluon response)
JET_EffectiveNP_R10_Statistical	Statistical components of effective jet energy scale uncertainties, split into 6 components
JET_EffectiveNP_R10_Modelling	Modelling components of effective jet energy scale uncertainties, split into 4 components
JET_EffectiveNP_R10_Detector	Detector components of effective jet energy scale uncertainties, split into 2 components
JET_EffectiveNP_R10_Mixed	Effective jet energy scale uncertainties coming from various sources, split into 3 components
JET_SingleParticle_HighPt	Uncertainty related to high p_T jets (for R=0.4)
JET_CombMass_Baseline	Baseline uncertainty of the jet mass scale accounting for data-MC differences
JET_CombMass_Modelline	Modelling uncertainty of the jet mass scale accounting for different MC generators
JET_CombMass_Tracking	Uncertainty of the jet mass scale accounting for tracking variations; 3 variations in total
Variable-radius track jets	
FT_EFF_EIGEN_B	b -tagging efficiency uncertainties
FT_EFF_EIGEN_C	None
FT_EFF_EIGEN_L	None
FT_EFF_EIGEN_extrapolation	b -tagging efficiency uncertainty on the extrapolation on high p_T -jets
FT_EFF_EIGEN_extrapolation_from_charm	b -tagging efficiency uncertainty on τ -jets

Table 10.3: A summary of the experimental systematic uncertainties.

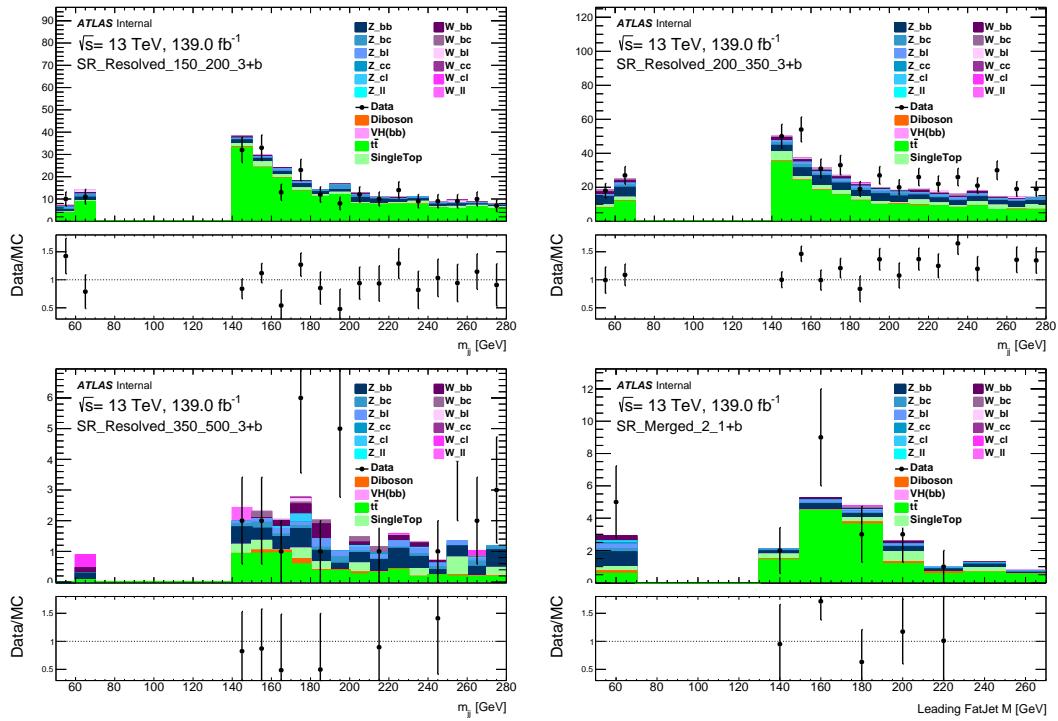


Figure 10.2: Higgs candidate mass spectra in the different E_T^{miss} regions with at least 3 b -tagged jets in the 0-lepton channel.

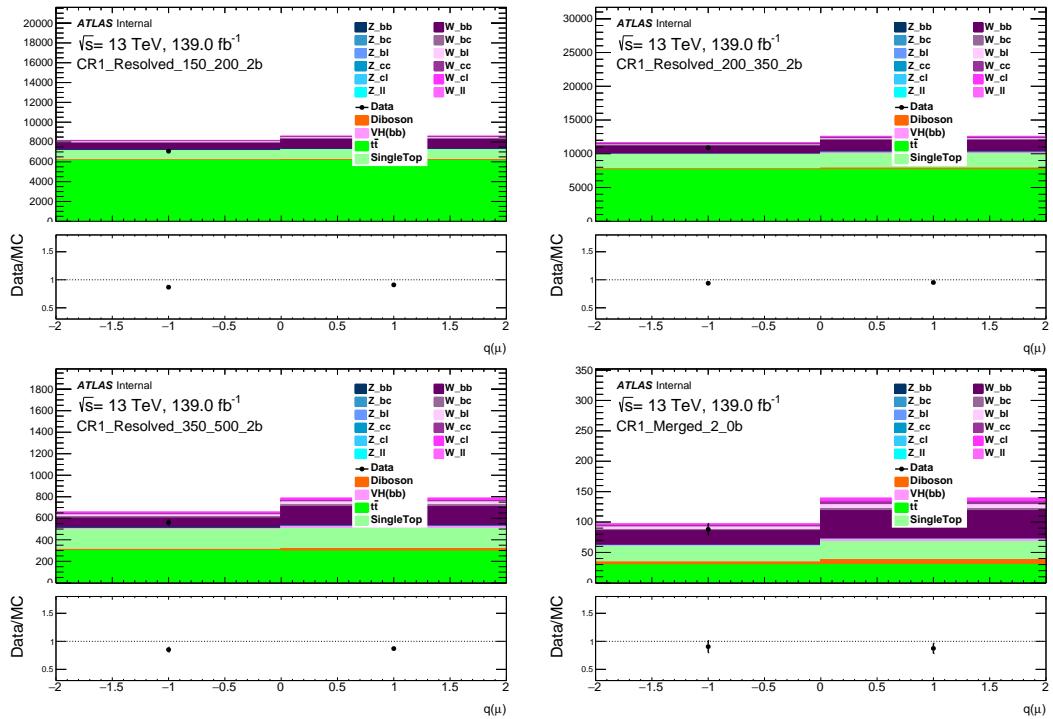


Figure 10.3: Muon charge distribution in the different E_T^{miss} regions with 2 b -tagged jets in the 1-lepton channel.

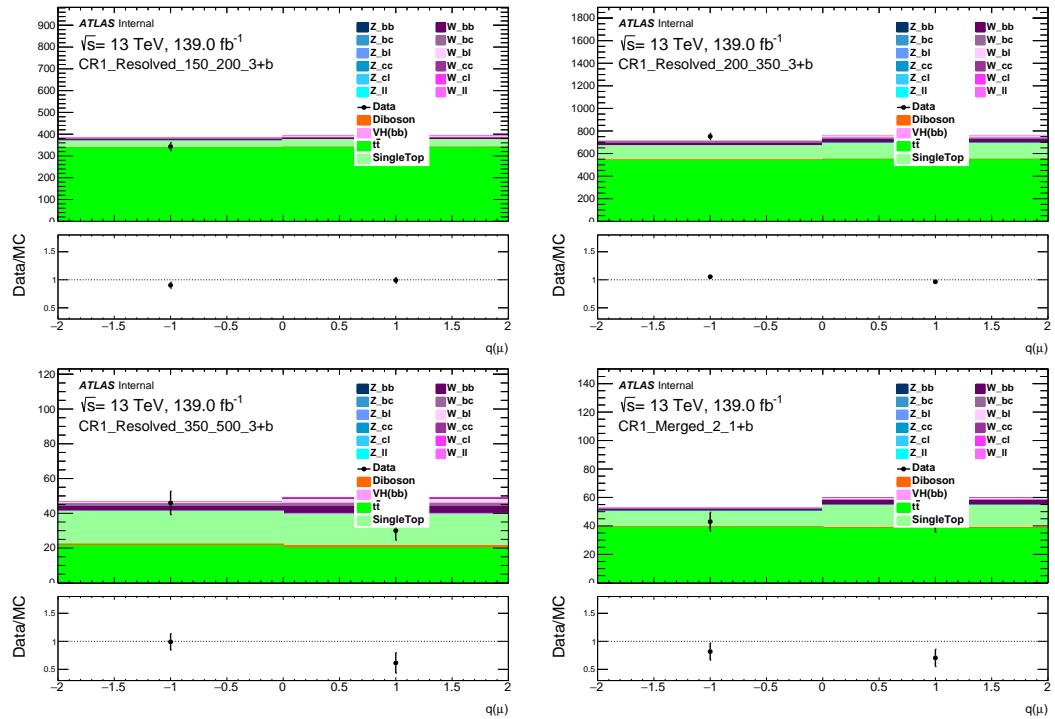


Figure 10.4: Muon charge distribution in the different E_T^{miss} regions with least 3 b -tagged jets in the 1-lepton channel.

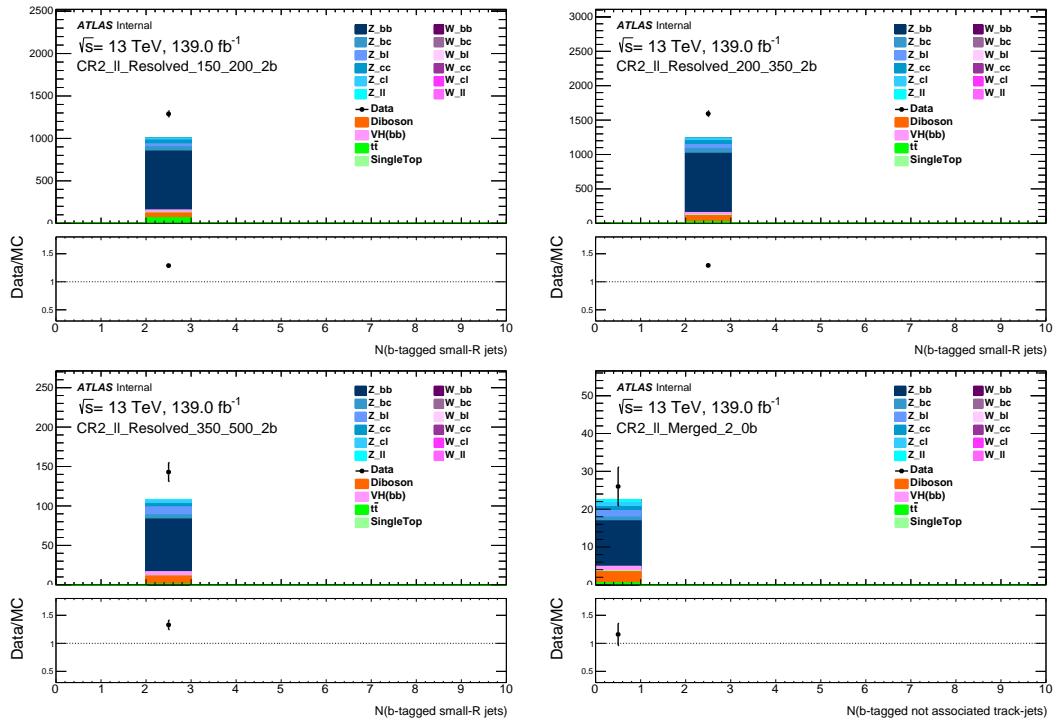


Figure 10.5: Total yields in the 2-lepton control region for different E_T^{miss} regions with 2 b -tagged jets.

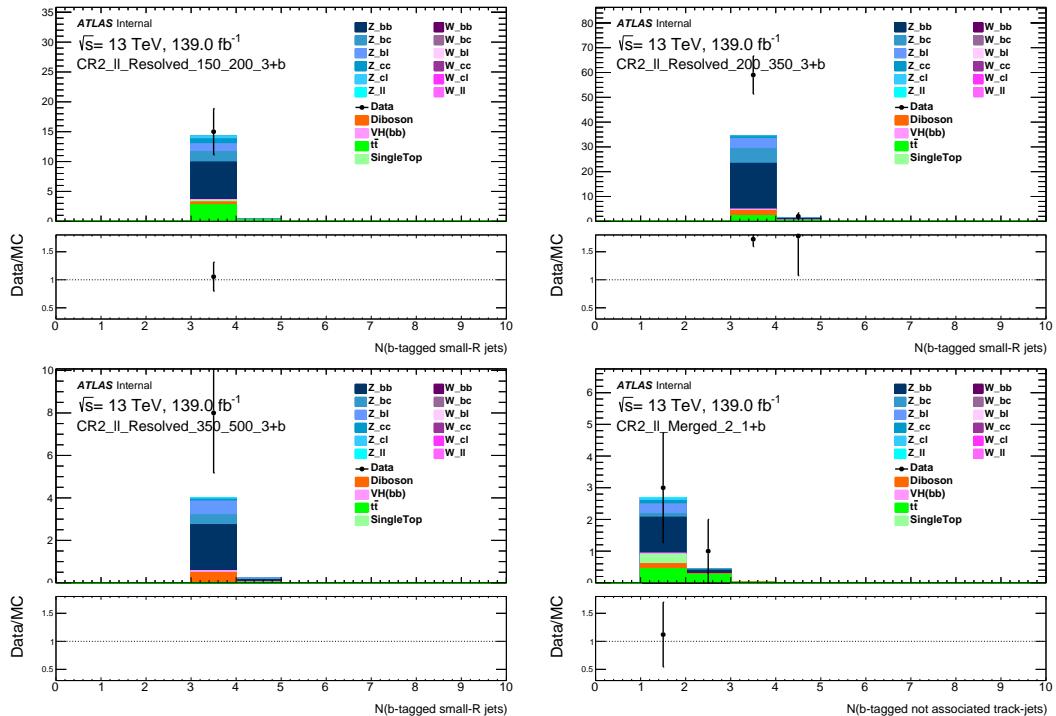


Figure 10.6: Total yields in the 2-lepton control region for different E_T^{miss} regions with more than 2 b -tagged jets.

Chapter 11

Result

TODO, background predictions in signal region, stack chart and table.

Sample text sample text sample text. Sample text sample text sample text.

11.1 Sample section

Sample text sample text sample text. Sample text sample text sample text.

11.1.1 Sample subsection

Sample text sample text sample text. Sample text sample text sample text.

11.1.2 Sample subsubsection

Sample text sample text sample text. Sample text sample text sample text.

sample text. Sample text sample text sample text.

11.2 Sample section

Sample text sample text sample text. Sample text sample text sample text. Sample text sample text sample text sample text. Sample text sample text sample text.

11.2.1 Sample subsection

Sample text sample text sample text. Sample text sample text sample text. Sample text sample text sample text sample text. Sample text sample text sample text.

Part VI

Conclusions

Chapter 12

Conclusions

The general conclusions go here.

Part VII

Appendices

Appendix A

The ATLAS detector service work

Sample text sample text sample text. Sample text sample text sample text. Sample text sample text sample text sample text. Sample text sample text sample text.

A.1 Sample section

Sample text sample text sample text. Sample text sample text sample text. Sample text sample text sample text sample text. Sample text sample text sample text.

A.1.1 Sample subsection

Sample text sample text sample text. Sample text sample text sample text. Sample text sample text sample text sample text. Sample text sample text sample text.

A.2 Sample section

Sample text sample text sample text. Sample text sample text sample text. Sample text sample text sample text sample text. Sample text sample text sample text.

A.2.1 Sample subsection

Sample text sample text sample text. Sample text sample text sample text. Sample text sample text sample text sample text. Sample text sample text sample text.

Appendix B

CombinedXbbScore Tagger implementation

B.1 Introduction

B.2 Performance

The Boosted Xbb Tagger is based on a neural network (roughly 10 layers) which uses b-tagging and jet substructure information. The classification has three outputs which represents how likely the jet is Higgs, Top and QCD jet.

With different fraction of the the three classes, CombinedXbbScore Tagger is defined as

$$D = \ln\left(\frac{p_H}{(1 - f) \times p_{QCD} + f \times p_{top}}\right), \quad (\text{B.1})$$

where f is the mixing fraction of top jet in the background, and XbbScore Top/Higgs/QCD are the three output of the Xbb Tagger.

The performance can be expressed in terms of its receiver operator characteristic (ROC) plot, which shows the change of background rejection as a function of the Higgs boson tagging efficiency. Fig.B.1 compares CombinedXbbScore aggers with different mixing fraction.

The preliminary working points of the CombinedXbbScore are listed in Table.B.1.

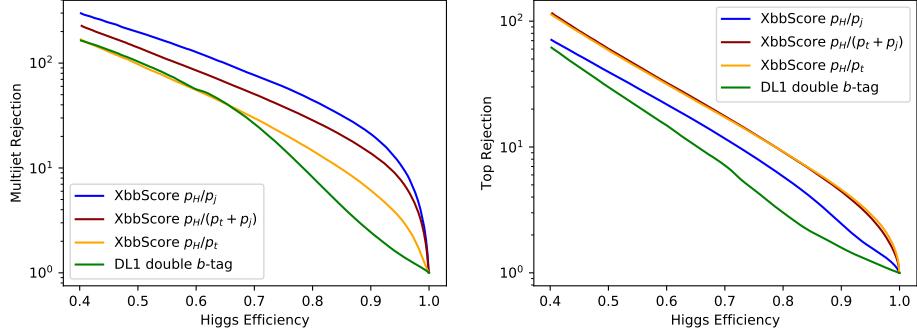


Figure B.1: ROC plot: Comparison between different working points of CombinedXbbScore

Higgs efficiency [%]	mixing fraction (f)	cut value
50	0	5.1
60	0	4.8
70	0	3.9
50	0.2	4.5
60	0.2	3.9
70	0.2	3.0
50	1	3.6
60	1	3.0
70	1	2.1

Table B.1: Preliminary working points of CombinedXbbScore

B.3 Signal selection in Merged region using CombinedXbb-Score

For the merged region, instead of requiring two b-tagged VR trackjets with 77% working point as described in Section 9.1. The background large-R jets ($R = 1.0$) are suppressed by applying a cut on CombinedXbbScore. To have a fair comparison within these two methods, 77% working point is chosen for the VR trackjets b-tagging and 60% working points are chosen for the CombinedXbbScore tagger with different mixing fractions for all plots showed in this chapter.

The large-R jet mass distribution of the Z+2HDM signal with $m_Z = 2800 \text{ GeV}$ and

$m_A = 300 \text{ GeV}$ is scaled by a factor of 1000 and compared to backgrounds in both Fig.B.2 and Fig.??.

The left plot in Fig.B.2 shows the large-R jet mass distribution in the merged region while the right plot shows the large-R jet mass distribution after requiring two b-tagged VR trackjets. Fig.B.2 shows the large-R jet mass distribution after applying cutting on CombinedXbbScore with mixing fraction $f = 1$ (left) and $f = 0$ (right).

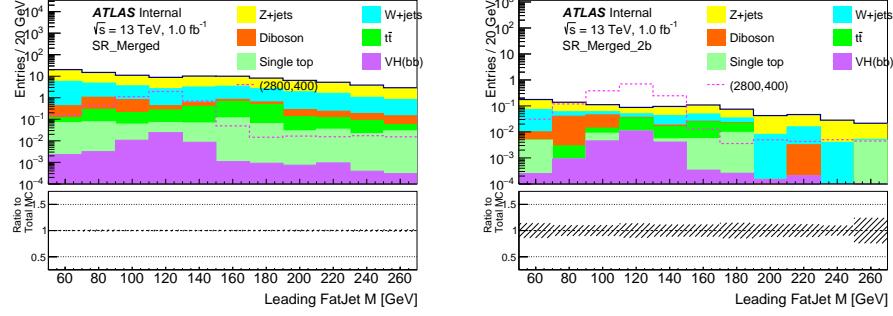


Figure B.2: Leading Fatjet mass in Merged region (left) and 2b Merged region defined by 2-b tagged VR (right)

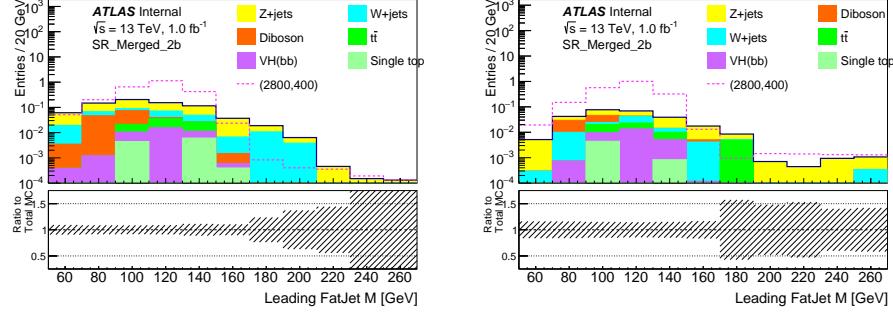


Figure B.3: Leading Fatjet mass in 2b Merged region defined by CombinedXbbScore

While the CombinedXbbScore tagger did a much better job at suppressing backgrounds, it also shaped the background distributions.

B.3.1 Backgrounds breakdown of signal region and control regions

To study the modelling and systematics, truth labeling is implemented in ntuples based on the truth particles ghost-associated to the large-R jets. And the backgrounds composition

on the 1L/2L control regions are examined.

Fig.B.4 shows the large-R jet mass distribution in the 2-btagged singal and control regions defined by the CombinedXbbScore.

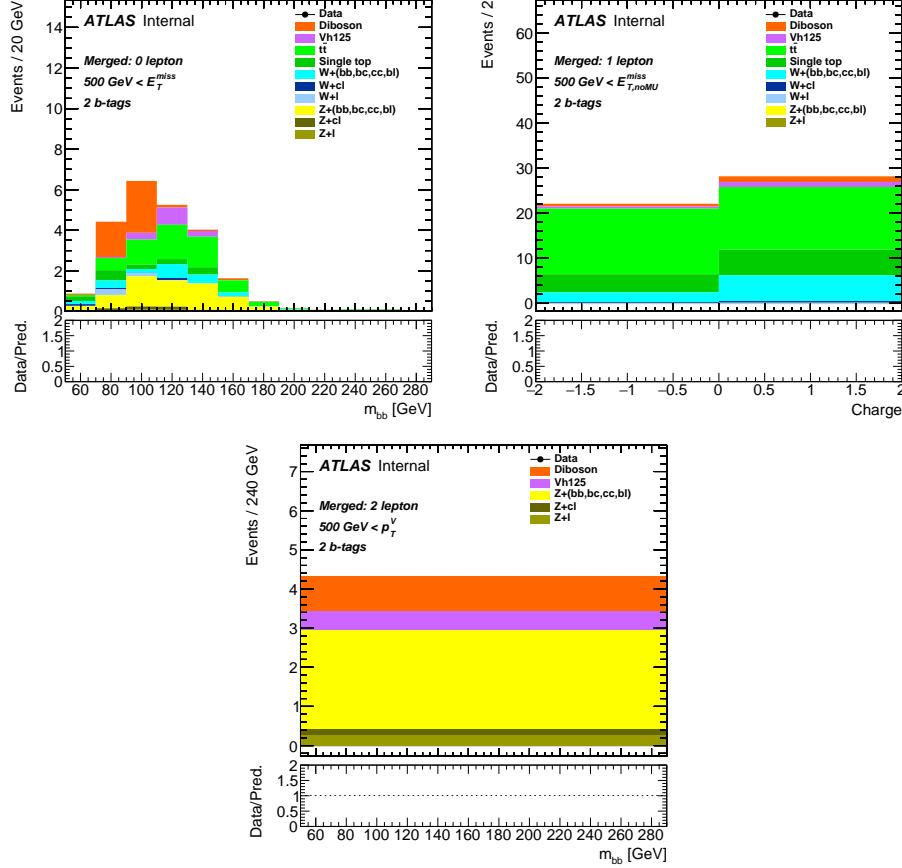


Figure B.4: Large-R jet mass distribution in the 2-btagged singal and control regions.

Ideally, large-R jets selected by cutting on CombinedXbbScore are likely to have two b-hadrons ghost-associated to them and have a truch labeling of bb.

To have a clear look at the fraction of large-R jets labling, the Xbb tag fraction which refer to the fraction of large-R jets with labeling bb are examined for 1 lepton region with W+jets samples in Fig.B.5 and for 2 region with Z+jets samples in Fig.B.6.

Xbb tag fraction of signal region vs 1b control regions in Fig.B.5 matches within uncertainty, as well as signal region vs 2b control regions in Fig.B.6. And the Xbb tag fraction peaks in around Higgs mass.

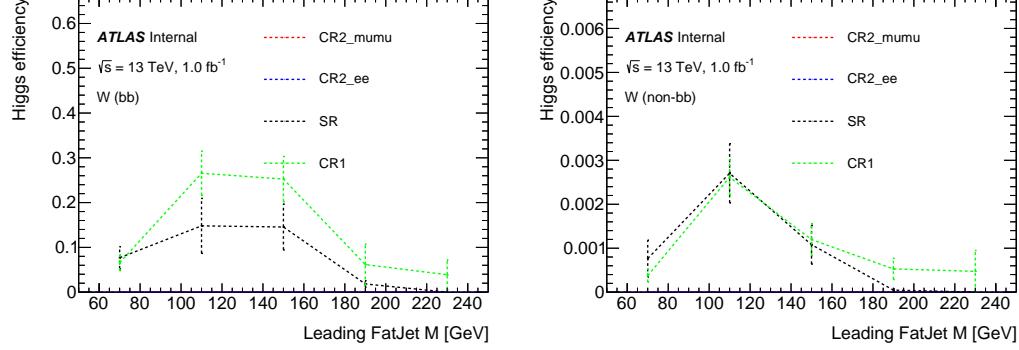


Figure B.5: Fraction of large-R jets with bb labeling (left) and without bb labeling (right) in with $W + \text{jets}$ samples.

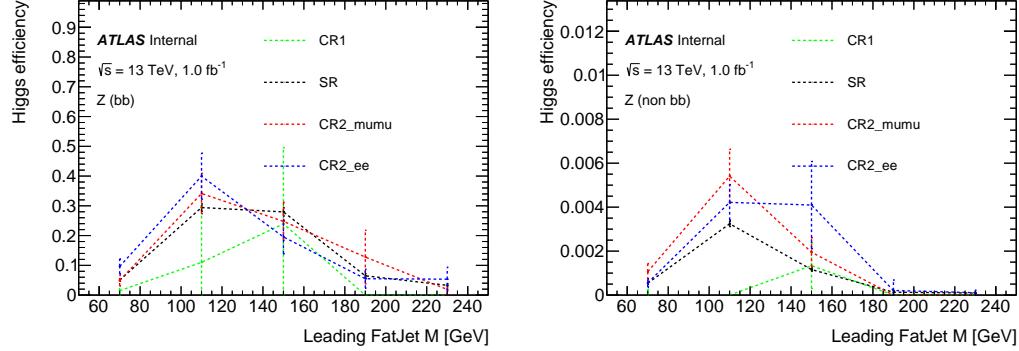


Figure B.6: Fraction of large-R jets with bb labeling (left) and without bb labeling (right) in with $Z + \text{jets}$ samples.

The flavor breakdown of backgrounds with truth labeling in singal and control regions is showed in Fig.B.7

To further quantify the flavor breakdown in signal and control regions are showed in Table.B.2, Table.B.3 and Table.B.4.

According to the tables above, the Wbb fraction in $W + \text{jets}$ in signal region is $46.62\% \pm 10.76\%$ compared to $56.60\% \pm 7.68\%$ in 1 lepton control reigon. And the Zbb fraction in $Z + \text{jets}$ is $56.53\% \pm 1.64\%$ compared to $54.42\% \pm 6.21\%$ in 2 lepton control reigon.

Zero lepton 2 tag merged, $E_T^{miss} > 500$ GeV	
signal mzp1400_mA600	0.0458 ± 0.0005
WZ	0.6106 ± 0.2170
ZZ	3.9200 ± 0.3486
Wl	0.4992 ± 0.2104
Wcl	0.2709 ± 0.0918
Wbb	1.3348 ± 0.2478
Wbl	0.7481 ± 0.1999
Wbc	0.0100 ± 0.0070
Zl	0.3124 ± 0.0293
VHbb	1.5404 ± 0.0162
WW	0.1818 ± 0.1285
Zcc	0.3775 ± 0.0293
Zbl	1.5497 ± 0.0580
Zbb	3.8587 ± 0.0879
Zcl	0.4465 ± 0.0431
Zbc	0.2807 ± 0.0233
ttbar	6.0874 ± 0.2225
stopWt	1.5491 ± 0.4528
stops	0.0200 ± 0.0141
Total Bkgd	23.5977 ± 3.3084

Table B.2: Flavor breakdown of backgrounds with truth labeling in singal region.

One lepton 2 tag merged, $E_T^{miss} > 500$ GeV	
WZ	1.7570 ± 0.2641
ZZ	0.0775 ± 0.0264
Wl	0.2212 ± 0.0978
Wcl	0.5467 ± 0.2127
Wbb	4.8836 ± 0.5282
Wcc	0.5283 ± 0.1443
Wbl	1.5783 ± 0.2929
Wbc	0.8691 ± 0.2415
Zl	0.0060 ± 0.0043
VHbb	1.5967 ± 0.0162
Zbl	0.0093 ± 0.0093
Zbb	0.0333 ± 0.0276
ttbar	28.4487 ± 0.8372
stopWt	9.6438 ± 1.1237
stops	0.0704 ± 0.0321
Total Bkgd	50.2699 ± 3.1672

Table B.3: Flavor breakdown of backgrounds with truth labeling in 1 lepton control region.

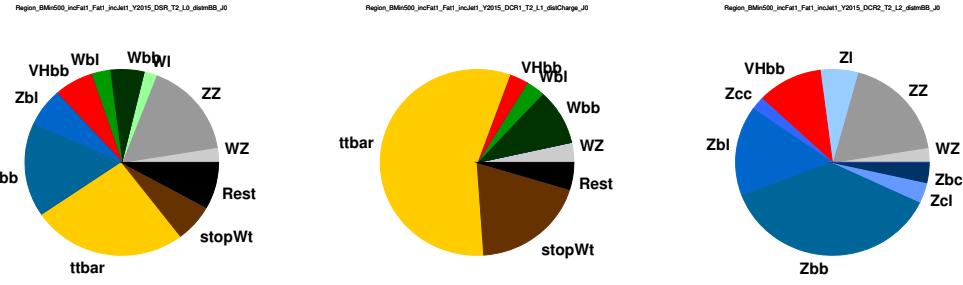


Figure B.7: Flavor breakdown of backgrounds with truth labeling in singal and control regions.

Two lepton 2 tag merged, $E_T^{miss} > 500$ GeV	
WZ	0.1108 ± 0.0427
ZZ	0.7880 ± 0.1048
Zl	0.2722 ± 0.1364
VHbb	0.4736 ± 0.0053
Zcc	0.0991 ± 0.0360
Zbl	0.6740 ± 0.0969
Zbb	1.6105 ± 0.1363
Zcl	0.1514 ± 0.0428
Zbc	0.1524 ± 0.0404
Total Bkgd	4.3320 ± 5.8464

Table B.4: Flavor breakdown of backgrounds with truth labeling in 2 lepton control region.

B.3.2 Preliminary significance study

To quantify the improvement brought by the CombinedXbbScore, the signal significances of these two methods are compared. Singal Significance in each bin is defined as

$$S_i = \sqrt{2(s + b) \ln(1 + \frac{s}{b})s}, \quad (\text{B.2})$$

where s, b is the count of signal and background in the i-th bin.

And the Bin-by-bin signal significance is difined as

$$S_{bin-by-bin} = \sqrt{\sum_i S_i^2}, \quad (\text{B.3})$$

Fig.?? (left) shows the ratio of signal significance of the 2b region defined by CombinedXbbScore vs 2b b-tagged VR trackjets with the Higgs window [70, 140] GeV while the

right plot is without the Higgs window.

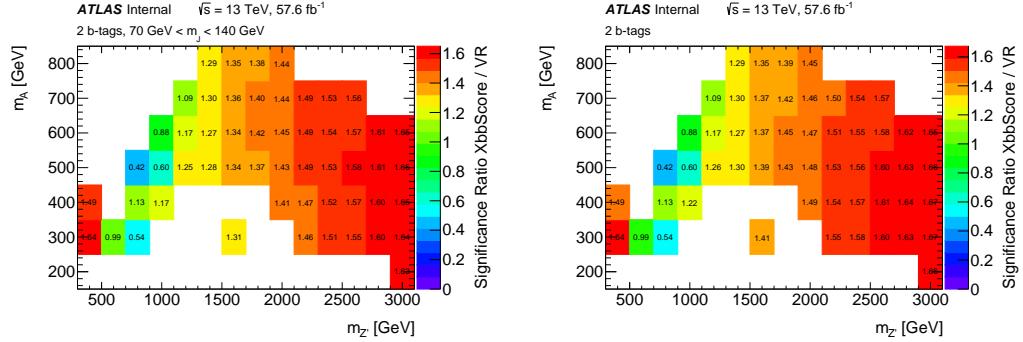


Figure B.8: Ratio of singal significance of 2b merged region defined by CombinedXbbScore compared to the orignal method with or without the Higgs window.

Part VIII

Bibliography

Bibliography

- [1] LEP Design Report Vol.1. 1983.
- [2] LEP Design Report: Vol.2. The LEP Main Ring. 1984.
- [3] *ATLAS liquid-argon calorimeter: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1996.
- [4] *ATLAS tile calorimeter: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1996.
- [5] *ATLAS muon spectrometer: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1997.
- [6] The silicon microstrip sensors of the atlas semiconductor tracker. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 578(1):98 – 118, 2007.
- [7] Performance of Missing Transverse Momentum Reconstruction in ATLAS studied in Proton-Proton Collisions recorded in 2012 at 8 TeV. Technical Report ATLAS-CONF-2013-082, CERN, Geneva, Aug 2013.
- [8] ATLAS Run 1 Pythia8 tunes. Technical Report ATL-PHYS-PUB-2014-021, CERN, Geneva, Nov 2014.
- [9] Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton-proton collision data. Technical Report ATLAS-CONF-2014-032, CERN, Geneva, Jun 2014.

- [10] Flavor Tagging with Track Jets in Boosted Topologies with the ATLAS Detector. Technical Report ATL-PHYS-PUB-2014-013, CERN, Geneva, Aug 2014.
- [11] 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.
- [12] Expected Performance of Boosted Higgs ($\rightarrow b\bar{b}$) Boson Identification with the ATLAS Detector at $\sqrt{s} = 13$ TeV. Technical Report ATL-PHYS-PUB-2015-035, CERN, Geneva, Aug 2015.
- [13] Expected performance of the ATLAS b -tagging algorithms in Run-2. Technical Report ATL-PHYS-PUB-2015-022, CERN, Geneva, Jul 2015.
- [14] 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.
- [15] Boosted Higgs ($\rightarrow b\bar{b}$) Boson Identification with the ATLAS Detector at $\sqrt{s} = 13$ TeV. Technical Report ATLAS-CONF-2016-039, CERN, Geneva, Aug 2016.
- [16] Identification of Jets Containing b -Hadrons with Recurrent Neural Networks at the ATLAS Experiment. Technical Report ATL-PHYS-PUB-2017-003, CERN, Geneva, Mar 2017.
- [17] Optimisation and performance studies of the ATLAS b -tagging algorithms for the 2017-18 LHC run. Technical Report ATL-PHYS-PUB-2017-013, CERN, Geneva, Jul 2017.
- [18] Variable Radius, Exclusive- k_T , and Center-of-Mass Subjet Reconstruction for Higgs($\rightarrow b\bar{b}$) Tagging in ATLAS. Technical Report ATL-PHYS-PUB-2017-010, CERN, Geneva, Jun 2017.
- [19] Search for Dark Matter Produced in Association with a Higgs Boson decaying to $b\bar{b}$ at $\sqrt{s} = 13$ TeV with the ATLAS Detector using 79.8 fb^{-1} of proton-proton collision data. Technical Report ATLAS-CONF-2018-039, CERN, Geneva, Jul 2018.

- [20] G. Aad et al. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.
- [21] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett.*, B716:1–29, 2012.
- [22] Georges Aad et al. Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 7$ TeV. *Eur. Phys. J.*, C73(3):2304, 2013.
- [23] Georges Aad et al. Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *JHEP*, 09:145, 2014.
- [24] Georges Aad et al. Operation and performance of the ATLAS semiconductor tracker. *JINST*, 9:P08009, 2014.
- [25] 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.
- [26] Georges Aad et al. Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector. *Eur. Phys. J.*, C76(11):581, 2016.
- [27] Georges Aad et al. Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1. *Eur. Phys. J.*, C77:490, 2017.
- [28] K. Aamodt et al. The ALICE experiment at the CERN LHC. *JINST*, 3:S08002, 2008.
- [29] A. Abada et al. FCC-hh: The Hadron Collider. *Eur. Phys. J. ST*, 228(4):755–1107, 2019.
- [30] E. Abat et al. The ATLAS Transition Radiation Tracker (TRT) proportional drift tube: Design and performance. *JINST*, 3:P02013, 2008.
- [31] Tomohiro Abe et al. LHC Dark Matter Working Group: Next-generation spin-0 dark matter models. *Phys. Dark Univ.*, page 100351, 2018.
- [32] R. Adam et al. Planck 2015 results. I. Overview of products and scientific results. *Astron. Astrophys.*, 594:A1, 2016.

- [33] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010.
- [34] A. Augusto Alves, Jr. et al. The LHCb Detector at the LHC. *JINST*, 3:S08005, 2008.
- [35] Daniele Alves, Nima Arkani-Hamed, Sanjay Arora, Yang Bai, Matthew Baumgart, Joshua Berger, Matthew Buckley, Bart Butler, Spencer Chang, Hsin-Chia Cheng, Clifford Cheung, R Sekhar Chivukula, Won Sang Cho, Randy Cotta, Mariarosaria D’Alfonso, Sonia El Hedri, Rouven Essig, Jared A Evans, Liam Fitzpatrick, Patrick Fox, Roberto Franceschini, Ayres Freitas, James S Gainer, Yuri Gershtein, Richard Gray, Thomas Gregoire, Ben Gripaios, Jack Gunion, Tao Han, Andy Haas, Per Hansson, JoAnne Hewett, Dmitry Hits, Jay Hubisz, Eder Izaguirre, Jared Kaplan, Emanuel Katz, Can Kilic, Hyung-Do Kim, Ryuichiro Kitano, Sue Ann Koay, Pyungwon Ko, David Krohn, Eric Kuflik, Ian Lewis, Mariangela Lisanti, Tao Liu, Zhen Liu, Ran Lu, Markus Luty, Patrick Meade, David Morrissey, Stephen Mrenna, Mihoko Nojiri, Takemichi Okui, Sanjay Padhi, Michele Papucci, Michael Park, Myeonghun Park, Maxim Perelstein, Michael Peskin, Daniel Phalen, Keith Rehermann, Vikram Rentala, Tuhin Roy, Joshua T Ruderman, Veronica Sanz, Martin Schmaltz, Stephen Schnetzer, Philip Schuster, Pedro Schwaller, Matthew D Schwartz, Ariel Schwartzman, Jing Shao, Jessie Shelton, David Shih, Jing Shu, Daniel Silverstein, Elizabeth Simmons, Sunil Somalwar, Michael Spannowsky, Christian Spethmann, Matthew Strassler, Shufang Su, Tim Tait, Brooks Thomas, Scott Thomas, Natalia Toro, Tomer Volansky, Jay Wacker, Wolfgang Waltenberger, Itay Yavin, Felix Yu, Yue Zhao, and Kathryn Zurek and. Simplified models for LHC new physics searches. *Journal of Physics G: Nuclear and Particle Physics*, 39(10):105005, sep 2012.
- [36] 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.

- [37] G. Apollinari, I. Bjar Alonso, O. Brning, P. Fessia, M. Lamont, L. Rossi, and L. Tavian. High-Luminosity Large Hadron Collider (HL-LHC). *CERN Yellow Rep. Monogr.*, 4:1–516, 2017.
- [38] Pierre Artoisenet, Rikkert Frederix, Olivier Mattelaer, and Robbert Rietkerk. Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations. *JHEP*, 03:015, 2013.
- [39] R W Assmann. Preliminary Beam-based specifications for the LHC collimators. Technical Report LHC-PROJECT-NOTE-277, CERN, Geneva, Jan 2002.
- [40] ATLAS Collaboration. TWiki.
- [41] ATLAS Collaboration. Studies on top-quark Monte Carlo modelling for Top2016. ATL-PHYS-PUB-2016-020, 2016.
- [42] Richard D. Ball, Valerio Bertone, Francesco Cerutti, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Jose I. Latorre, Juan Rojo, and Maria Ubiali. Impact of Heavy Quark Masses on Parton Distributions and LHC Phenomenology. *Nucl. Phys.*, B849:296–363, 2011.
- [43] Richard D. Ball et al. Parton distributions with LHC data. *Nucl. Phys.*, B867:244–289, 2013.
- [44] Richard D. Ball et al. Parton distributions for the LHC Run II. *JHEP*, 04:040, 2015.
- [45] David Barney. Calorimetry by examples. https://indico.cern.ch/event/526766/attachments/1269795/1881900/CalorimetryByExamples_2016.pdf.
- [46] Asher Berlin, Tongyan Lin, and Lian-Tao Wang. Mono-Higgs Detection of Dark Matter at the LHC. *JHEP*, 06:078, 2014.
- [47] Enrico Bothmann et al. Event Generation with Sherpa 2.2. 2019.
- [48] 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.

- [49] Oliver S. Bruning, P. Collier, P. Lebrun, S. Myers, R. Ostojaic, J. Poole, and P. Proudlock. LHC Design Report Vol.1: The LHC Main Ring. 2004.
- [50] O. Buning, P. Collier, P. Lebrun, S. Myers, R. Ostojaic, J. Poole, and P. Proudlock. LHC Design Report. 2. The LHC infrastructure and general services. 2004.
- [51] Jon Butterworth, Stefano Carrazza, Amanda Cooper-Sarkar, Albert De Roeck, Jol Feltesse, Stefano Forte, Jun Gao, Sasha Glazov, Joey Huston, Zahari Kassabov, and et al. Pdf4lhc recommendations for lhc run ii. *Journal of Physics G: Nuclear and Particle Physics*, 43(2):023001, Jan 2016.
- [52] Jonathan M. Butterworth, Adam R. Davison, Mathieu Rubin, and Gavin P. Salam. Jet substructure as a new Higgs search channel at the LHC. *Phys. Rev. Lett.*, 100:242001, 2008.
- [53] Nicola Cabibbo. Unitary symmetry and leptonic decays. *Phys. Rev. Lett.*, 10:531–533, Jun 1963.
- [54] Matteo Cacciari and Gavin P. Salam. Pileup subtraction using jet areas. *Phys. Lett.*, B659:119–126, 2008.
- [55] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The Catchment Area of Jets. *JHEP*, 04:005, 2008.
- [56] M Capeans, G Darbo, K Einsweiller, 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.
- [57] Fabio Cascioli, Philipp Maierhofer, and Stefano Pozzorini. Scattering Amplitudes with Open Loops. *Phys. Rev. Lett.*, 108:111601, 2012.
- [58] S. Catani, F. Krauss, R. Kuhn, and B. R. Webber. QCD Matrix Elements + Parton Showers. *JHEP*, 11:063, 2001.
- [59] S. Chatrchyan et al. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

- [60] Serguei Chatrchyan et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett.*, B716:30–61, 2012.
- [61] ATLAS Collaboration. Atlas data preparation check list for physics analysis. <https://twiki.cern.ch/twiki/bin/view/Atlas/DataPreparationCheckListForPhysicsAnalysis>.
- [62] ATLAS Collaboration. Atlas detector & technology, cern official page. <https://atlas.cern/discover/detector>.
- [63] ATLAS Collaboration. Atlas fact sheets. <https://atlasexperiment.org/fact-sheets-1-view.html>.
- [64] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003. 437 p, 2008. Also published by CERN Geneva in 2010.
- [65] 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.
- [66] 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.
- [67] D. J. Lange. The EvtGen particle decay simulation package. *Nucl. Instrum. Meth. A*, 462:152, 2001.
- [68] Ansgar Denner, Stefan Dittmaier, and Lars Hofer. Collier: A fortran-based complex one-loop library in extended regularizations. *Comput. Phys. Commun.*, 212:220–238, 2017.
- [69] Niels Doble, Lau Gatignon, Kurt Hbner, and Edmund Wilson. The Super Proton Synchrotron (SPS): A Tale of Two Lives. *Adv. Ser. Direct. High Energy Phys.*, 27:135–177, 2017.

- [70] Stephen D. Ellis, Christopher K. Vermilion, and Jonathan R. Walsh. Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches. *Phys. Rev.*, D81:094023, 2010.
- [71] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964.
- [72] Jonathan L. Feng. Dark Matter Candidates from Particle Physics and Methods of Detection. *Ann. Rev. Astron. Astrophys.*, 48:495–545, 2010.
- [73] Rikkert Frederix and Stefano Frixione. Merging meets matching in MC@NLO. *JHEP*, 12:061, 2012.
- [74] Rikkert Frederix, Emanuele Re, and Paolo Torrielli. Single-top t-channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO. *JHEP*, 09:130, 2012.
- [75] Stefano Frixione, Eric Laenen, Patrick Motylinski, and Bryan R. Webber. Angular correlations of lepton pairs from vector boson and top quark decays in Monte Carlo simulations. *JHEP*, 04:081, 2007.
- [76] Stefano Frixione, Eric Laenen, Patrick Motylinski, Bryan R. Webber, and Chris D. White. Single-top hadroproduction in association with a W boson. *JHEP*, 07:029, 2008.
- [77] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. *JHEP*, 11:070, 2007.
- [78] Stefano Frixione, Paolo Nason, and Giovanni Ridolfi. A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction. *JHEP*, 09:126, 2007.
- [79] Simone Gilardoni and Django Manglunki, editors. *Fifty years of the CERN Proton Synchrotron*. 2011.
- [80] T. Gleisberg, Stefan Höche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, and J. Winter. Event generation with SHERPA 1.1. *JHEP*, 02:007, 2009.

- [81] T Gleisberg, S Hche, F Krauss, M Schnherr, S Schumann, F Siegert, and J Winter. Event generation with sherpa 1.1. *Journal of High Energy Physics*, 2009(02):007007, Feb 2009.
- [82] Tanju Gleisberg and Stefan Höche. Comix, a new matrix element generator. *JHEP*, 12:039, 2008.
- [83] David J. Gross and Frank Wilczek. Ultraviolet behavior of non-abelian gauge theories. *Phys. Rev. Lett.*, 30:1343–1346, Jun 1973.
- [84] Keith Hamilton, Paolo Nason, and Giulia Zanderighi. MINLO: Multi-Scale Improved NLO. *JHEP*, 10:155, 2012.
- [85] Heribertus B. Hartanto, Barbara Jäger, Laura Reina, and Doreen Wackerlo. Higgs boson production in association with top quarks in the POWHEG BOX. *Phys. Rev. D*, 91(9):094003, 2015.
- [86] Xiao-Gang He, Girish C. Joshi, H. Lew, and R. R. Volkas. Simplest Z-prime model. *Phys. Rev.*, D44:2118–2132, 1991.
- [87] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964.
- [88] Toko Hirono, Marlon Barbero, Pierre Barrillon, Siddharth Bhat, Patrick Breugnon, Ivan Caicedo, Zongde Chen, Michael Daas, Yavuz Degerli, Stephanie Godiot, Fabrice Guilloux, Tomasz Hemperek, Fabian Hugging, Hans Kruger, Patrick Pangaud, Piotr Rymaszewski, Philippe Schwemling, Maxence Vandenbroucke, Tianyang Wang, and Norbert Wermes. Depleted Fully Monolithic Active CMOS Pixel Sensors (DMAPS) in High Resistivity 150 nm Technology for LHC. *Nucl. Instrum. Methods Phys. Res., A*, 924(arXiv:1803.09260):87–91. 5 p, Mar 2018.
- [89] Stefan Höche, Frank Krauss, Marek Schönherr, and Frank Siegert. A critical appraisal of NLO+PS matching methods. *JHEP*, 09:049, 2012.
- [90] Stefan Höche, Frank Krauss, Marek Schönherr, and Frank Siegert. QCD matrix elements + parton showers. The NLO case. *JHEP*, 04:027, 2013.

- [91] Stefan Höche, Frank Krauss, Steffen Schumann, and Frank Siegert. QCD matrix elements and truncated showers. *JHEP*, 05:053, 2009.
- [92] Kerstin Hoepfner. Muon detection at the lhc. https://indico.cern.ch/event/115060/attachments/49316/70878/CERN_May2011_Hoepfner_Muonsystems.pdf.
- [93] J. C. Kapteyn. First Attempt at a Theory of the Arrangement and Motion of the Sidereal System. *Astrophys. J.*, 55:302–328, 1922.
- [94] Makoto Kobayashi and Toshihide Maskawa. CP Violation in the Renormalizable Theory of Weak Interaction. *Prog. Theor. Phys.*, 49:652–657, 1973.
- [95] David Krohn, Jesse Thaler, and Lian-Tao Wang. Jets with Variable R. *JHEP*, 06:059, 2009.
- [96] David Krohn, Jesse Thaler, and Lian-Tao Wang. Jet Trimming. *JHEP*, 02:084, 2010.
- [97] Gionata Luisoni, Paolo Nason, Carlo Oleari, and Francesco Tramontano. $HW^\pm/HZ + 0$ and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO. *JHEP*, 10:083, 2013.
- [98] Teresa Marrodn Undagoitia and Ludwig Rauch. Dark matter direct-detection experiments. *J. Phys.*, G43(1):013001, 2016.
- [99] Stephen P. Martin. A Supersymmetry primer. pages 1–98, 1997. [Adv. Ser. Direct. High Energy Phys.18,1(1998)].
- [100] Y. Nambu and G. Jona-Lasinio. Dynamical model of elementary particles based on an analogy with superconductivity. i. *Phys. Rev.*, 122:345–358, Apr 1961.
- [101] Y. Nambu and G. Jona-Lasinio. Dynamical model of elementary particles based on an analogy with superconductivity. ii. *Phys. Rev.*, 124:246–254, Oct 1961.
- [102] Yoichiro Nambu. Axial vector current conservation in weak interactions. *Phys. Rev. Lett.*, 4:380–382, Apr 1960.
- [103] Paolo Nason. A new method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11:040, 2004.

- [104] Heinz Pagels and Joel R. Primack. Supersymmetry, cosmology, and new physics at teraelectronvolt energies. *Phys. Rev. Lett.*, 48:223–226, Jan 1982.
- [105] R. D. Peccei. The Strong CP problem and axions. *Lect. Notes Phys.*, 741:3–17, 2008.
[,3(2006)].
- [106] Antonio Pich. Effective field theory: Course. In *Probing the standard model of particle interactions. Proceedings, Summer School in Theoretical Physics, NATO Advanced Study Institute, 68th session, Les Houches, France, July 28-September 5, 1997. Pt. 1, 2*, pages 949–1049, 1998.
- [107] H. David Politzer. Reliable perturbative results for strong interactions? *Phys. Rev. Lett.*, 30:1346–1349, Jun 1973.
- [108] J. Pumplin et al. New Generation of Parton Distributions with Uncertainties From Global QCD Analysis. *JHEP*, 07:012, 2002.
- [109] Aranzazu Ruiz-Martinez and ATLAS Collaboration. The Run-2 ATLAS Trigger System. Technical Report ATL-DAQ-PROC-2016-003, CERN, Geneva, Feb 2016.
- [110] Gavin P. Salam. Towards Jetography. *Eur. Phys. J.*, C67:637–686, 2010.
- [111] Martin Schmaltz and David Tucker-Smith. Little Higgs review. *Ann. Rev. Nucl. Part. Sci.*, 55:229–270, 2005.
- [112] Steffen Schumann and Frank Krauss. A Parton shower algorithm based on Catani-Seymour dipole factorisation. *JHEP*, 03:038, 2008.
- [113] Geraldine Servant and Timothy M. P. Tait. Is the lightest Kaluza-Klein particle a viable dark matter candidate? *Nucl. Phys.*, B650:391–419, 2003.
- [114] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008.
- [115] Torbjørn Sjstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z.

- Skands. An Introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159–177, 2015.
- [116] Wouter Van Den Wollenberg, Michael Kagan, Emily Thompson, and Lei Zhou. Measurement of the Track-Jet b -Tagging Efficiency of the MV1 algorithm in pp Collisions at $\sqrt{s} = 8$ TeV using $e\mu$ Dilepton $t\bar{t}$ events. Technical Report ATL-COM-PHYS-2015-009, CERN, Geneva, Jan 2015.
- [117] Carlo Varni and Stephen Jacob Sekula. b -tagging hybrid performances for b -jet triggers. Technical Report ATL-COM-DAQ-2019-018, CERN, Geneva, Feb 2019.
- [118] Steven Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967.
- [119] Simon D. M. White, C. S. Frenk, and M. Davis. Clustering in a Neutrino Dominated Universe. *Astrophys. J.*, 274:L1–L5, 1983. [,80(1984)].
- [120] Jan-Christopher Winter, Frank Krauss, and Gerhard Soff. A modified cluster-hadronization model. *Eur. Phys. J. C*, 36:381–395, 2004.
- [121] C. N. Yang and R. L. Mills. Conservation of isotopic spin and isotopic gauge invariance. *Phys. Rev.*, 96:191–195, Oct 1954.
- [122] Ya. B. Zeldovich, J. Einasto, and S. F. Shandarin. Giant Voids in the Universe. *Nature*, 300:407–413, 1982.
- [123] F. Zwicky. Die Rotverschiebung von extragalaktischen Nebeln. *Helv. Phys. Acta*, 6:110–127, 1933. [Gen. Rel. Grav.41,207(2009)].