

# SEARCH FOR DECAYS OF THE HIGGS BOSON INTO A Z BOSON AND A LIGHT HADRONICALLY DECAYING RESONANCE

Chonghao Wu

*Thesis submitted for the degree of  
Doctor of Philosophy*



Particle Physics Group,  
School of Physics and Astronomy,  
University of Birmingham.

*November 7, 2024*





---

## ABSTRACT

---

The work for this thesis was conducted within the ATLAS experiment. Initially, a noise study of the trigger system of ATLAS experiment was undertaken. This included the searching for noisy trigger towers and adjusting the noise thresholds for the new trigger component eFEX. Subsequently, a search for Higgs boson decays to a  $Z$  boson and a light resonance  $a$  is performed, using  $140 \text{ fb}^{-1} \sqrt{s} = 13 \text{ TeV}$   $pp$  collision data collected by the ATLAS experiment.  $a$  is considered to be a beyond standard model particle or a standard model charmonium state. The mass of  $a$  is less than 4 GeV and it is required to decay to quarks or gluons, while  $Z$  boson decays to two leptons. No significant excess above the expected background is observed. The observed 95% confidence-level upper limits are set on the branching ratio of Higgs boson decays to a  $Z$  boson and the  $a$  resonance, with values starting from 13.9% for  $a$  decays to gluons, and from 9.2% for  $a$  decays to a pair of quarks for different  $a$  masses. The search also considers the axion-like particle models. 95% confidence-level upper limits are set on the strength of the effective coupling  $C_{Zh}^{\text{eff}}$  with  $2.0 \text{ TeV}^{-1}$  for 0.5 GeV axion-like particle  $a$ , and  $0.9 \text{ TeV}^{-1}$  for 1.0 GeV  $a$ .

---

## DECLARATION OF AUTHORS CONTRIBUTION

---

This thesis shows the work which I have undertaken during my three-year post-graduate studies as a member of the ATLAS experiment at CERN.

Chapter 1 is a introduction to the thesis.

Chapter 2 recaps the theory background of Higgs boson as well as the experimental status of the measurement of Higgs properties and the beyond standard model scalar sectors, which helps to understand the content later in this thesis.

Chapter 3 describes the ATLAS experiment briefly, including different components of the ATLAS detector and the reconstruction of physics objects. This information is mainly from Technical Design Reports of the detectors and recent materials published by ATLAS.

Chapter 4 presents the work of my ATLAS authorship qualification task as part of ATLAS L1Calo group. The data samples of supercells used in this part is provided by the ATLAS LAr group, and the samples of trigger towers are generated from package developed by other people in L1Calo group. The noise analysis based on those samples are conducted by me. The results of this work make contributions to the data-taking of ATLAS in Run 3.

Chapter 5 discusses the mathematical structure of the neural network, which is used in the search for the Higgs boson decays into a  $Z$  boson and a light hadronically decaying resonance as shown in Chapter 6. I was involved almost every aspect of this analysis as part of a small analysis group from the University of Birmingham. I performed the neural network training for background reweighting, signal mass regression and signal-background classification. I also derived the signal models and different components of systematic uncertainties. The statistical treatment was

performed with a previously developed software which I modified for it to produce the results for this analysis with updated models and data. I am one of main editors of the internal note and the paper draft. The paper draft for this analysis is in circulation for comments within the ATLAS collaboration at the time of writing.

---

## ACKNOWLEDGEMENTS

---

After three years spent in Birmingham, I am now nearing the end of my PhD. My PhD study began in late 2021, when the world did not fully recover from the pandemic yet. It is not an easy story that crosses a half of the earth, live and study abroad, and then finally achieve a PhD. Therefore, I would like to thank everyone who has supported me in the last three years.

The first person I would like to thank is, of course, my supervisor, Prof. Kostas Nikolopoulos. His advice and supervision have helped me a lot, not only about the research to finish my PhD, but also lessons for my life.

I am also deeply grateful to Dr. Panagiotis Bellos, who gave me the most help in the last three years, both from the analysis and trigger sides. He engineered the analysis and guided me through the trigger analysis. Without his generous help, I could not finish my PhD study at this stage.

I would also like to acknowledge Dr. Eleni Skorda, who joined our analysis group after arriving in Birmingham in early 2023. Her significant contributions greatly accelerated the approval of our analysis by the EB. Additionally, I extend my thanks to Andy Chisholm who guided our analysis when Kostas was not in Birmingham, and to Dr. Tom Neep who gave us suggestions for our analysis.

My gratitude extends to Dr. Juraj Bracinik and Dr. Steve Hillier, who helped me to complete the ATLAS qualification task in the L1Calo group. I am also thankful to all other staffs in the Birmingham ATLAS group, they gave good advice for my research when I presented the student updates in the Local ATLAS weekly meeting.

During my time in Birmingham, I was fortunate to meet a group of wonderful and friendly PhD students in the ATLAS student office: Robert Ward, Antonio

Costa, Gareth Bird, Harry Cooke, Will George, Julia Silva, Adrien Auriol, Mihaela Marinescu, Josh Lomas, Eric Liu and Ilyas Benaoumeur. Many of them have already graduated and are now making valuable contributions to scientific and other fields. During the past three years, we often chatted about our research and daily life during the lunch time, which made my time in Birmingham more enjoyable.

I would like to extremely thank to my family. They gave me unwavering support for my PhD study as well as days before, The contact between us these years was not easy because of a 7 or 8-hour time zone difference, so they often woke up very early or stayed up late to call me at a convenient time. I could not have accomplished this without their love and encouragement.

Finally, I wish to thank my girlfriend, Dinghe Dai. Over the past three years, we spent much of our holidays traveling together, which make my PhD period much more wonderful and unforgettable. Thanks to her patience, understanding, love and support, I am completing my PhD in the UK, while she is doing her PhD in Germany.



我作为一名科技工作者，活着的目的就是为人民服务。如果人民最后对我的工作满意的话，那才是最高的奖赏。  
- 钱学森(1911-2009)

*As a scientific worker, the purpose of my life is to serve the people. If the people are satisfied with my work, that will be the highest award for me.*  
- *Qian, Xuesen (1911-2009)*



# Contents

1	Introduction	1
2	Particle Physics Theory	3
2.1	The Standard Model Particles	3
2.1.1	Leptons	4
2.1.2	Quarks	5
2.1.3	Gauge Bosons	5
2.1.4	Higgs boson	7
2.2	Gauge Symmetry	7
2.2.1	$U(1)$ Gauge Symmetry and QED	8
2.2.2	$SU(3)$ Gauge Symmetry and QCD	9
2.2.3	$SU(2) \times U(1)$ Gauge Symmetry and Electroweak Interaction	10
2.3	The Higgs Mechanism and the Higgs boson	11
2.3.1	Spontaneous Symmetry Breaking	12
2.3.2	Higgs Mechanism	13
2.3.3	The SM Higgs Boson	15
2.3.4	Higgs Boson Production at the LHC	17
2.3.5	Higgs Boson Decay	19
2.4	Extensions of the Standard Model	21
2.4.1	2HDM	21
2.4.2	2HDM+S	24
2.4.3	Experimental Searches for Extended Scalar Sectors	27
3	The ATLAS experiment at the LHC	29
3.1	The Large Hadron Collider	29
3.2	The ATLAS detector	32
3.2.1	Inner Detector	34
3.2.2	Calorimeters	37
3.2.3	Muon Spectrometer	41
3.2.4	Magnet System	44
3.2.5	Trigger and Data Acquisition System	44
3.3	Physics Objects Reconstruction	48
3.3.1	Tracks and Interaction Vertices	48
3.3.2	Electrons	50
3.3.3	Muons	52
3.3.4	Jets	55
3.3.5	Missing Transverse Energy	57

---

4	Noise Analysis of the Upgraded ATLAS L1Calo Trigger	59
4.1	L1Calo in LHC Run 3	60
4.1.1	Trigger Electronics	61
4.1.2	eFEX	62
4.2	Searching for Problematic Trigger Towers	65
4.3	Adjustment of Noise Thresholds of eFEX	67
4.3.1	$E_T$ calculation and the timing cut	67
4.3.2	$\eta$ -based Noise Thresholds	71
5	Elements of Machine Learning	76
5.1	Structure of Neural Networks	77
5.2	Training of NNs	80
5.2.1	Forward Propagation and Loss Functions	81
5.2.2	Backpropagation and Optimizers	83
5.2.3	Over-fitting treatment	91
5.3	Hyperparameter optimization	93
6	Searches for $H \rightarrow Z\alpha(\rightarrow \text{hadronic})$	95
6.1	Introduction	95
6.1.1	Light Scalar Bosons	98
6.1.2	Charmonium States	98
6.1.3	Axions	99
6.2	Experimental and Simulated Data Samples	100
6.2.1	Data Samples	100
6.2.2	Simulated Samples	100
6.2.3	Event Removal	103
6.3	Selection of Events	104
6.3.1	$ll + \text{jet}$ Pre-Selection	105
6.3.2	Track Selection	107
6.3.3	Track-based Neural Networks	114
6.3.4	Regression NN optimization	119
6.3.5	Alternative Classification NNs	121
6.4	Signal and Background Modelling	123
6.4.1	Signal Modelling	123
6.4.2	Background Modelling	126
6.4.3	Training	128
6.4.4	Reweighting results	129
6.4.5	Reweighting Validation	132
6.4.6	Standard Model Higgs contamination	133
6.5	Systematic Uncertainties	134
6.5.1	Background	134
6.5.2	Signal	145
6.6	Statistical Model	152
6.6.1	Statistical methods	152
6.6.2	Fit model	154
6.6.3	Asimov fits	154

6.6.4	Fit to data . . . . .	154
6.6.5	Nuisance parameter impact on the signal strength . . . . .	157
6.6.6	Alternative strategy . . . . .	160
6.7	Results and Interpretations . . . . .	161
6.7.1	Model Independent Interpretation . . . . .	161
6.7.2	$\eta_c$ and $J/\psi$ . . . . .	162
6.7.3	ALP Interpretation . . . . .	163
7	Conclusion	164

# CHAPTER 1

---

## Introduction

---

The Higgs boson,  $H$ , is the particle associated to the Higgs field which gives particles mass in the Standard Model (SM). The Higgs boson was predicted by three independent groups: François Englert and Robert Brout [1]; Peter Higgs [2]; Gerald Guralnik, Carl Hagen, and Tom Kibble [3] in 1964 to provide mechanism of mass generation of the  $W$  and  $Z$  bosons. On 4th of July 2012, the ATLAS and CMS experiments at the CERN Large Hadron Collider (LHC) announced the observation of a new particle consistent with the Higgs boson in the mass region around 125 GeV [4, 5]. Following this discovery, Peter Higgs and François Englert were awarded the 2013 Nobel Prize in physics [6].

Since the discovery, both the ATLAS and CMS experiments have performed several measurements of its properties, and have launched a range of searches for new physics in its production and decays. The results of all these searches are consistent with the predictions for the SM Higgs boson [7, 8]. However, given the current precision of the measurements and that some rare decays not being observed yet, there is still

great potential for the discovery of new physics beyond the Standard Model (BSM) in the Higgs sector.

The SM scalar sector is the simplest but not the only possible Higgs sector. There are various extensions of the SM featuring extended Higgs sectors, which motivate exotic Higgs boson decays involving new particles [9]. The two-Higgs-Doublet Model (2HDM) is one of the simplest extensions of the SM and there are many motivations for it, such as the baryon asymmetry problem [10], axion models [11] and supersymmetry [12]. Measuring the branching ratios of Higgs boson decays into BSM particles is the key method to constrain the 2HDM parameter space [13]. The ATLAS experiment has performed searches for the decays  $H \rightarrow Za$  and  $H \rightarrow aa$  in many different final state channels, where  $a$  is a light BSM pseudoscalar, which subsequently decays into photons, leptons, gluons or quarks [9]. This thesis discusses searches for the Higgs boson decays to a  $Z$  boson and a light hadronically decaying  $a$  at the ATLAS experiment.

This thesis is organized as follows. Chapter 2 recaps the theoretical background and experimental status of Higgs boson and the BSM scalar sectors. Chapter 3 describes the detector of the ATLAS experiment at the LHC, as well as the techniques used for the reconstruction of different physics objects. Chapter 4 presents the noise study of the trigger system of the ATLAS experiment, covering the search for the problematic components and the adjustment of the noise thresholds for the upgraded system. Chapter 5 discusses the machine learning technique, and the mathematical structure of the neural network (NN), a machine learning algorithm inspired by the structure of the human brain. Searches for decay of the Higgs boson to a  $Z$  boson and a light hadronically decaying resonance are presented in Chapter 6, where the neural networks are applied.

## CHAPTER 2

---

### Particle Physics Theory

---

#### 2.1 The Standard Model Particles

The Standard Model of Particle Physics is a theory that successfully describes three of the four fundamental interactions, i.e. the electromagnetic interaction, the weak interaction and the strong interaction. The SM is a quantum field theory (QFT), which combines quantum mechanics with special relativity. Each fundamental particle in the SM is associated with a quantum field.

The elementary particles of the SM can be seen in Figure 2.1, some of them, such as  $W$ ,  $Z$  bosons,  $\tau$  lepton and Higgs boson, were predicted before their discovery. In general, there are two types of elementary particles in the SM: fermions and bosons. The fermions are the building blocks of matter, including quarks and leptons. The bosons consist of gauge bosons, propagating the interactions between fermions, and the Higgs boson, which is associated with the Higgs mechanism that gives particles

masses.

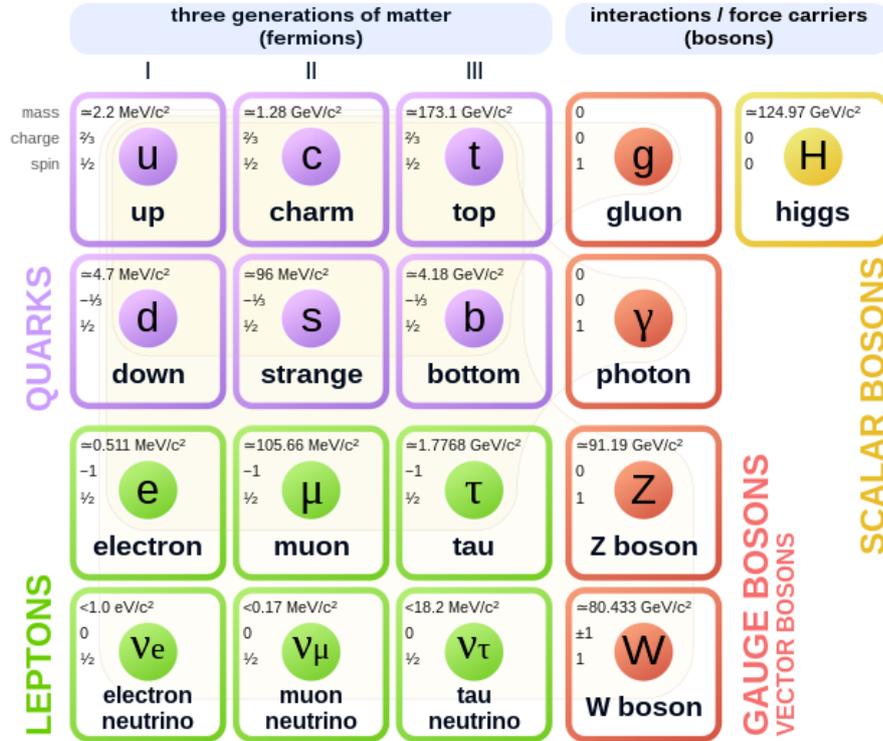


Figure 2.1: Elementary particles of the Standard Model [14]

### 2.1.1 Leptons

There are three generations of leptons, each generation has a charged lepton,  $l^-$ , and a neutrino,  $\nu$ . The three charged leptons are electron ( $e$ ), muon ( $\mu$ ), and tau ( $\tau$ ). They have the same electric charge  $-1$  and spin  $\frac{1}{2}$ . The mass of each charged lepton is greater than that of the charged lepton of the previous generation, i.e.  $m_e < m_\mu < m_\tau$ . Each charged lepton has a corresponding neutrino, named respectively electron neutrino ( $\nu_e$ ), muon neutrino ( $\nu_\mu$ ), and tau neutrino ( $\nu_\tau$ ). Neutrinos are electrically neutral. The masses of neutrinos are very small but not zero [15], which is one of the outstanding puzzles of the SM [16].

### 2.1.2 Quarks

There are in total six flavors of quarks ( $q$ ), including three ‘up-type’ quarks: up ( $u$ ), charm ( $c$ ), top ( $t$ ), and three ‘down-type’ quarks: down ( $d$ ), strange ( $s$ ), bottom ( $b$ ). Similar to leptons, quarks are also divided into three generations, each generation contains an ‘up-type’ quark and a ‘down-type’ quark. The ‘up-type’ quark is charged  $+\frac{2}{3}$ , and the ‘down-type’ quark has charge  $-\frac{1}{3}$ . Besides the electric charges, quarks have also color charge and interact through the strong force. There are three colors: red ( $R$ ), green ( $G$ ) and blue ( $B$ ). Quarks have colors and anti-quarks ( $\bar{q}$ ) have anti-colors. The equal mixture of a color and its corresponding anti-color, or the equal mixture of three colors (or three anti-colors) will generate a ‘colorless’ state.

The quark confinement principle postulates that quarks cannot exist in isolation, but quarks can form colorless compounds: hadrons. Hadrons are organised in two main categories: baryons and mesons. Baryons are composed of three quarks, each quark carrying a different color. For example protons and neutrons, which make up the atomic nuclei. Protons consist of two up quarks and one down quark ( $uud$ ), while neutrons consist of two down quarks and one up quark ( $udd$ ). Mesons are made of a quark and an anti-quark with opposite color charges. The lightest mesons are pions:  $\pi^+$ ,  $\pi^-$ ,  $\pi^0$ , made up of  $u\bar{d}$ ,  $\bar{u}d$ ,  $u\bar{u}$  or  $d\bar{d}$  respectively.

### 2.1.3 Gauge Bosons

In QFT, all forces of nature are a result of particle exchange. Gauge bosons are spin-1 vector particles, propagating the interactions. They can be derived from the local gauge symmetry of the fermion field, which is discussed in Section 2.2.

Photons ( $\gamma$ ) mediate the electromagnetic (EM) interactions of charged particles. Photons in the SM are massless. The theory describing the electromagnetic interactions between particles is quantum electrodynamics (QED).

Gluons ( $g$ ) are the propagators of the strong interaction. Similar to the EM inter-

action acting on electrically charged particles, the strong interaction acts on ‘color-charged’ particles, and gluons play the roles of mediating strong interactions between ‘color-charged’ particles. Gluons are massless, like photons. However, while photons are electric neutral, gluons themselves carry color charge. Therefore, gluons can directly interact with other gluons, i.e. the gluon self-coupling. Because of 3 types of color charges ( $R$ ,  $G$ ,  $B$ ) of quarks, there are 8 gluons with different color combinations:

$$R\bar{G}, R\bar{B}, G\bar{R}, G\bar{B}, B\bar{R}, B\bar{G}, \frac{1}{\sqrt{2}}(R\bar{R} - G\bar{G}), \frac{1}{\sqrt{6}}(R\bar{R} + G\bar{G} - 2B\bar{B}). \quad (2.1)$$

These 8 combinations belong a color octet. The remain combination is a color singlet:

$$\frac{1}{\sqrt{3}}(R\bar{R} + G\bar{G} + B\bar{B}), \quad (2.2)$$

which does not carry color and cannot mediate strong interactions. The strong interaction acts on color charges, so the theory describing it is called quantum chromodynamics (QCD).

The weak interaction is propagated via  $W$  and  $Z$  bosons, and acts on quarks and leptons. There are two types of weak interactions: ‘charged-current interaction’ mediated by  $W^\pm$  bosons, and ‘neutral-current interaction’ mediated by  $Z$  bosons. During the 1960s, Glashow, Weinberg, and Salam proposed the Electroweak theory that unifies the EM and the weak interaction [17, 18, 19]. The electroweak interaction conserves the weak isospin  $I_3$ . Particles with different handedness have different isospins. Usually the handedness is defined by the spin of particle: a particle is left-handed if the directions of spin and momentum are opposite, it is right-handed if the direction of spin is same as its momentum. The left-handed fermions carry isospin  $I_3 = \pm\frac{1}{2}$ , while the right-handed fermions have  $I_3 = 0$ . The electroweak theory also introduces a weak hypercharge ( $Y$ ) of a particle, which has the relation to the electric charge ( $Q$ ) and the isospin:

$$Q = I_3 + \frac{Y}{2} \quad (2.3)$$

Unlike other gauge bosons,  $W^\pm$  and  $Z$  bosons are not massless, which follows from the Higgs mechanism discussed in the section 2.3. The weak gauge bosons can also self-interact, such as  $WWZ$ ,  $WWZZ$  and  $4W$  interactions, as Feynman diagrams in Figure 2.2.

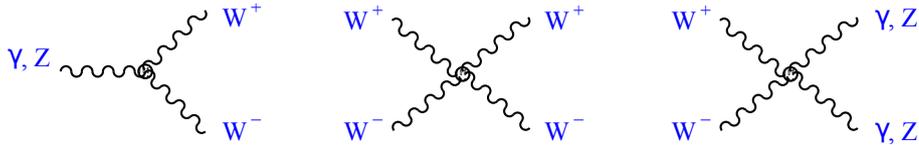


Figure 2.2: Feynman diagrams of gauge boson self-interactions. Taken from Ref. [20].

### 2.1.4 Higgs boson

In the SM, the Higgs boson is a scalar boson with spin 0. The Higgs field gives  $W^\pm$  and  $Z$  bosons masses through the Higgs mechanism. The interaction between Higgs boson and charged fermions is considered a Yukawa interaction, which gives the masses of fermions.

## 2.2 Gauge Symmetry

The gauge symmetry is a powerful principle that the SM obeys, indicates that the fields of the particles and the interactions are invariant under the local gauge transformation. The symmetry of the SM is denoted as  $SU(3)_C \times SU(2)_L \times U(1)_Y$ , which means that the Lagrangian ( $\mathcal{L}$ ) of the SM is invariant under  $SU(3)_C \times SU(2)_L \times U(1)_Y$  gauge transformation.  $SU(3)_C$  is the base of QCD and  $SU(2)_L \times U(1)_Y$  stands for the Electroweak theory, where  $L$  refers to the left-handedness since the weak interaction only interacts with left-handed fermions and  $Y$  refers to the weak hypercharge.

### 2.2.1 $U(1)$ Gauge Symmetry and QED

For a free fermion field  $\psi(x^\mu)$ , the Lagrangian is written as

$$\mathcal{L} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi. \quad (2.4)$$

where  $\gamma^\mu$  are gamma matrices and  $\bar{\psi} \equiv \psi^\dagger\gamma^0$ ,  $\psi^\dagger$  is the Hermitian conjugate of  $\psi$ . The Lagrangian is invariant under the global transformation:

$$\psi \rightarrow e^{i\theta}\psi. \quad (2.5)$$

The transformation is unitary and has single parameter, so this type of transformation forms a  $U(1)$  group. For a global  $U(1)$  transformation,  $\theta$  is a real constant so that the transformation is independent from space-time. If  $\theta$  is a function of  $x^\mu$ , that is, the transformation is not space-time independent, then it becomes a local  $U(1)$  transformation:

$$\psi \rightarrow e^{i\theta(x)}\psi. \quad (2.6)$$

The Lagrangian in (2.4) is no longer invariant under the local  $U(1)$  transformation. By demanding the local  $U(1)$  symmetry, the Lagrangian needs some modifications. The ‘covariant derivative’  $D_\mu$  is defined as

$$D_\mu = \partial_\mu - iQA_\mu, \quad (2.7)$$

where  $Q$  is the electric charge, and  $A_\mu$  is a vector field transforming as

$$A_\mu \rightarrow A_\mu + \frac{1}{Q}\partial_\mu\theta(x). \quad (2.8)$$

Under the local  $U(1)$  transformation,  $D_\mu$  satisfies:

$$D_\mu\psi \rightarrow e^{i\theta(x)}D_\mu\psi. \quad (2.9)$$

And the Lagrangian acquires the  $U(1)$  symmetry by replacing  $\partial^\mu$  by  $D_\mu$ :

$$\begin{aligned}\mathcal{L} &= i\bar{\psi}\gamma^\mu D_\mu\psi - m\bar{\psi}\psi \\ &= \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi + Q\bar{\psi}\gamma^\mu\psi A_\mu.\end{aligned}\tag{2.10}$$

Therefore, a vector field  $A_\mu$  is required to maintain the local  $U(1)$  gauge symmetry. The  $A_\mu$  is called the gauge field, and it couples to the charged fermions as shown in the last term in equation (2.10), just like the photon field. To regard  $A_\mu$  as the photon, additional terms of a vector boson are needed:

$$\mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \frac{1}{2}m_A^2 A^\mu A_\mu.\tag{2.11}$$

where the first term is the kinetic term and the second is the mass term.  $F_{\mu\nu}$  is called the field strength tensor, defined as

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu.\tag{2.12}$$

Since  $F_{\mu\nu}$  is local  $U(1)$  invariant but  $A^\mu A_\mu$  is not, the vector field  $A_\mu$  must be massless ( $m_A = 0$ ). The complete Lagrangian of QED can be written as:

$$\mathcal{L}_{QED} = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi + Q\bar{\psi}\gamma^\mu\psi A_\mu - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}\tag{2.13}$$

In summary, the theory of QED can be derived by requiring the local  $U(1)$  gauge symmetry of the free fermion field, and the gauge boson of electromagnetic interaction, photon, must be massless.

## 2.2.2 $SU(3)$ Gauge Symmetry and QCD

In QCD, the Lagrangian of quarks is similar to that in QED:

$$\mathcal{L} = i\bar{\psi}_i\gamma^\mu\partial_\mu\psi_i - m\bar{\psi}_i\psi_i,\tag{2.14}$$

where  $\psi$  stands for the spinor as for that in QED, and  $i = R, G, B$  indicates different colors. The Lagrangian (2.14) is invariant under the global  $SU(3)$  transformation:

$$\psi \rightarrow U\psi = e^{i\alpha_a T_a} \psi, a = 1, 2, \dots, 8 \quad (2.15)$$

where  $U$  is any  $3 \times 3$  unitary matrix with determinant 1 ( $U^\dagger U = 1$  and  $\det U = 1$ ).  $\alpha_a$  are group parameters, and  $T_a$  are 8 Gell-Mann matrices, the generators of  $SU(3)$  group, corresponding to 8 gluons, which transform as

$$G_\mu^a \rightarrow G_\mu^a + \frac{1}{g} \partial_\mu \alpha^a + f^{abc} G_\mu^b \alpha^c, \quad (2.16)$$

$f_{abc}$  are the structure constants of the  $SU(3)$  group. Comparing to the transformation of photon field in QED (2.8), (2.16) has an extra term because the  $SU(3)$  group is non-Abelian, meaning that  $3 \times 3$  matrices do not commute. By adding the kinetic terms of gluons, the full QCD Lagrangian can be written as

$$\mathcal{L}_{QCD} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi + g(\bar{\psi}\gamma^\mu T_a \psi)G_\mu^a - \frac{1}{4}G_a^{\mu\nu}G_{\mu\nu}^a, \quad (2.17)$$

where the field strength tensors are

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + gf_{abc}G_\mu^b G_\nu^c. \quad (2.18)$$

The higher order terms of  $G^a$  indicate the interaction between gluons. Just like the photon, the local gauge symmetry requires the gluons also to be massless.

### 2.2.3 $SU(2) \times U(1)$ Gauge Symmetry and Electroweak Interaction

The  $SU(2) \times U(1)$  transformation is

$$\psi \rightarrow e^{i\alpha^a \sigma^a / 2} e^{iY/2} \psi, a = 1, 2, 3 \quad (2.19)$$

where  $\sigma^a/2$  are the generators of  $SU(2)$  group, the Pauli matrices.  $Y/2$  is the generator of  $U(1)$  group.

Similar to the examples of QED and QCD, local  $SU(2) \times U(1)$  symmetry brings 4 gauge bosons:  $W_\mu^{1,2,3}$  corresponding to the three generators of  $SU(2)$  group, and  $B_\mu$  from  $U(1)$  generator. 4 gauge bosons of the electroweak interaction in the SM:  $W^\pm$ ,  $Z$ , and the photon  $\gamma$ , are the linear combinations of  $W^i$  and  $B$  bosons:

$$\begin{aligned}W_\mu^\pm &= \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2), \\Z_\mu &= \frac{1}{\sqrt{g^2 + g'^2}}(gW_\mu^3 - g'B_\mu), \\A_\mu &= \frac{1}{\sqrt{g^2 + g'^2}}(gB_\mu + g'W_\mu^3),\end{aligned}\tag{2.20}$$

where  $g$  and  $g'$  are coupling constants of  $SU(2)$  and  $U(1)$  groups. We know that the photon is massless, but  $W^\pm$  and  $Z$  bosons in the SM have masses. In principle, the mass of gauge bosons will break the local gauge symmetry. To solve this problem, three independent groups: Brout and Englert [1]; Higgs [2]; Guralnik, Hagen, and Kibble [3] introduced the Englert-Brout-Higgs-Guralnik-Hagen-Kibble mechanism almost simultaneously, which is also called the Higgs mechanism. The gauge bosons can acquire masses via the Higgs mechanism without violating the gauge symmetry.

## 2.3 The Higgs Mechanism and the Higgs boson

According to the SM, the Higgs mechanism gives masses to the  $W^\pm$  and  $Z$  bosons. David Griffiths once said that the Higgs mechanism is “the remarkable offspring of the marriage of local gauge invariance and spontaneous symmetry breaking” [21]. To understand how gauge bosons gain masses and why a Higgs boson is born, we need to start from spontaneous symmetry breaking.

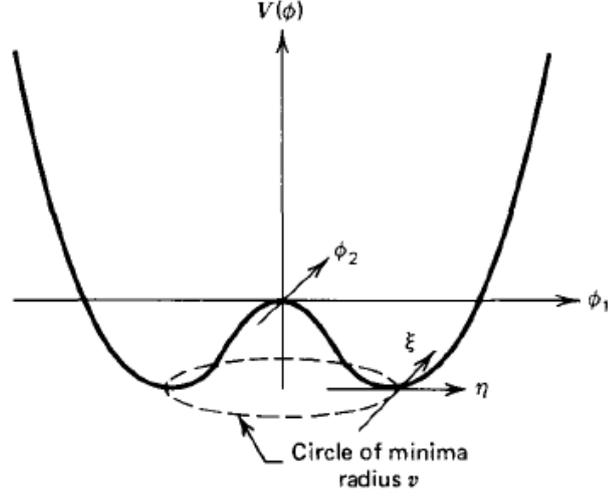


Figure 2.3: The potential  $V(\phi)$  for complex scalar field [22]

### 2.3.1 Spontaneous Symmetry Breaking

Spontaneous symmetry breaking (SSB) is a mechanism that generates the mass of a particle. For a complex scalar field  $\phi = (\phi_1 + i\phi_2)/\sqrt{2}$ , we can construct a Lagrangian

$$\mathcal{L} = T - V = (\partial_\mu \phi)^* (\partial^\mu \phi) - \mu^2 \phi^* \phi - \lambda (\phi^* \phi)^2, \quad (2.21)$$

which possesses a global  $U(1)$  symmetry. The potential  $V$  can be written as

$$V(\phi) = \frac{1}{2}\mu^2(\phi_1^2 + \phi_2^2) + \frac{1}{4}\lambda(\phi_1^2 + \phi_2^2)^2. \quad (2.22)$$

Consider the case of  $\lambda > 0$  and  $\mu^2 < 0$ , the minima of  $V(\phi)$  lie on a circle in the  $\phi_1, \phi_2$  plane:

$$\phi_1^2 + \phi_2^2 = v^2 = \frac{-\mu^2}{\lambda}, \quad (2.23)$$

as shown in Fig 2.3, the vacuum state is degenerate. Under the global  $U(1)$  transformation  $\phi \rightarrow e^{i\theta}\phi$ , the Lagrangian is invariant, but the point on the circle is moved to another point on the circle, hence the vacuum state is not invariant. This is the main idea of the SSB. We can pick a vacuum state

$$\phi_1 = v, \phi_2 = 0, \quad (2.24)$$

and expand the Lagrangian around the vacuum state by writing

$$\phi_1 = v + \eta(x), \phi_2 = \xi(x). \quad (2.25)$$

Then we have

$$\phi = \frac{1}{\sqrt{2}}(v + \eta(x) + i\xi(x)). \quad (2.26)$$

and

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \eta)^*(\partial^\mu \eta) + \mu^2 \eta^2 + \frac{1}{2}(\partial_\mu \xi)^*(\partial^\mu \xi) + \text{const.} + \text{higher order terms of } \eta \text{ and } \xi. \quad (2.27)$$

The Lagrangian with the form of

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \eta)^*(\partial^\mu \eta) - \frac{1}{2}m_\eta^2 \eta^2 \quad (2.28)$$

denotes a free scalar field with a mass  $m_\eta$ . Thus a massive scalar field  $\eta$  with the mass  $m_\eta = \sqrt{-2\mu^2}$  is generated from the spontaneously broken global symmetry.

Besides, we also obtain a massless scalar field  $\xi$ . It is called the ‘Goldstone boson’, which is the generic consequence of SSB for a continuous global symmetry, according to the Goldstone’s theorem [23]. However, the Goldstone boson is unwanted in the SM. That is reason to introduce the Higgs mechanism, the combination of SSB and the local gauge symmetry.

### 2.3.2 Higgs Mechanism

Again we consider a complex scalar field  $\phi$  with the Lagrangian (2.21). This Lagrangian is invariant under the global  $U(1)$  transformation, but not invariant under the local  $U(1)$  gauge transformation. With the knowledge in the section 2.2.1, by demanding the local  $U(1)$  gauge symmetry, a new gauge field  $A_\mu$  is defined. The

Lagrangian needs to be modified:

$$\mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + (D_\mu\phi)^*(D^\mu\phi) - \mu^2\phi^*\phi - \lambda(\phi^*\phi)^2. \quad (2.29)$$

where  $D_\mu$  has the definition

$$D_\mu = \partial_\mu - igA_\mu, \quad (2.30)$$

and  $A_\mu$  transformations as

$$A_\mu \rightarrow A_\mu + \frac{1}{g}\partial_\mu\theta(x). \quad (2.31)$$

Substituting  $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$  into (2.29), the full expression of Lagrangian is written as

$$\begin{aligned} \mathcal{L} = & -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + (\partial_\mu\phi)^*(\partial^\mu\phi) - \mu^2\phi^*\phi - \lambda(\phi^*\phi)^2 \\ & + igA_\mu\phi^*(\partial^\mu\phi) - ig(\partial_\mu\phi)^*A_\mu\phi + g^2A_\mu A^\mu\phi^*\phi. \end{aligned} \quad (2.32)$$

Same as the procedures in last section, the vacuum state is chosen to be  $\phi_1 = v$ ,  $\phi_2 = 0$  and the symmetry of Lagrangian is spontaneously broken. By expanding the field  $\phi$  about the chosen vacuum state:

$$\phi = \frac{1}{\sqrt{2}}(v + \eta(x) + i\xi(x)), \quad (2.33)$$

the Lagrangian can be written as

$$\begin{aligned} \mathcal{L} = & \frac{1}{2}(\partial_\mu\eta)^*(\partial^\mu\eta) + \mu^2\eta^2 + \frac{1}{2}(\partial_\mu\xi)^*(\partial^\mu\xi) - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \frac{1}{2}g^2v^2A_\mu A^\mu - gvA_\mu\partial^\mu\xi \\ & + \text{higher order terms,} \end{aligned} \quad (2.34)$$

from which we can see a massive scalar field  $\eta$ , a massless Goldstone boson  $\xi$  and a massive vector boson  $A_\mu$ . Notice that the gauge field  $A_\mu$  in section 2.2.1 was massless, but now it has mass. The term  $-QvA_\mu\partial^\mu\xi$  can be eliminated by implementing

the gauge transformation

$$A_\mu \rightarrow A'_\mu = A_\mu - \frac{1}{gv} \partial_\mu \xi(x). \quad (2.35)$$

The Lagrangian becomes

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \eta)^*(\partial^\mu \eta) + \mu^2 \eta^2 - \frac{1}{4} F^{\mu\nu} F_{\mu\nu} + \frac{1}{2} g^2 v^2 A'_\mu A'^\mu + \text{higher order terms}. \quad (2.36)$$

As choosing the gauge corresponding  $\theta(x) = -\xi(x)/v$  in (2.31), the Goldstone boson  $\xi$  disappears from the Lagrangian. The Lagrangian shows a massive scalar Higgs field  $\eta$  and a massive vector gauge boson  $A_\mu$ . The higher order terms indicates the Higgs self-coupling, and the interactions between Higgs boson and the gauge boson. The mass of gauge boson is given by

$$m_A = gv. \quad (2.37)$$

The Higgs boson has the mass:

$$m_H = \sqrt{-\mu^2} = \sqrt{2\lambda}v. \quad (2.38)$$

Notice that during the process of giving masses to the gauge bosons, the Goldstone boson is no longer appeared. We can say, the gauge field has ‘eaten’ the Goldstone boson and obtained the mass. This is the Higgs mechanism.

### 2.3.3 The SM Higgs Boson

In the SM, the spontaneously broken symmetry is local  $SU(2) \times U(1)$  symmetry.

Consider a complex scalar doublet  $\Phi = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}$  with hypercharge  $Y = 1$ . The

Lagrangian is given by

$$\mathcal{L} = -\frac{1}{4} F_a^{\mu\nu} F_{\mu\nu}^a - \frac{1}{4} B^{\mu\nu} B_{\mu\nu} + (D_\mu \Phi)^\dagger (D^\mu \Phi) - V(\Phi), \quad (2.39)$$

where the potential  $V(\Phi)$  has the form of

$$V(\Phi) = \lambda(\Phi^\dagger\Phi)^2 + \mu^2(\Phi^\dagger\Phi) \quad (2.40)$$

and the covariant derivative is defined as

$$D_\mu\Phi = (\partial_\mu - ig\frac{\sigma^a}{2}W_\mu^a - ig'\frac{Y}{2}B_\mu)\Phi \quad (2.41)$$

where  $\sigma^\alpha$  are Pauli matrices,  $Y$  is the hypercharge,  $g$  and  $g'$  are coupling constants,  $W_\mu$  and  $B_\mu$  are the gauge fields of gauge symmetry  $SU(2)_L$  and  $U(1)_Y$ ,  $F_{\mu\nu}^a$  and  $B_{\mu\nu}$  are the field strength. The Lagrangian is local  $SU(2) \times U(1)$  symmetric. For  $\mu^2 < 0$ , the minimum of  $V(\Phi)$  is located at

$$|\phi_1|^2 + |\phi_2|^2 = \frac{1}{2}v^2, \quad (2.42)$$

where  $v = -\frac{\mu}{\sqrt{\lambda}}$ . Therefore, we can choose a vacuum state  $\Phi = \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix}$ , which is not invariant under local  $SU(2)$  transformation.

As expanding the Lagrangian around the vacuum, we apply a small perturbation  $h$  on it:

$$\Phi = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}}(v+h) \end{pmatrix}. \quad (2.43)$$

Then the covariant derivative can be rewritten as

$$D_\mu\Phi = \begin{pmatrix} -\frac{ig}{2\sqrt{2}}(W_\mu^1 - iW_\mu^2)(v+h) \\ -\frac{i}{2\sqrt{2}}(g'B_\mu - gW_\mu^3)(v+h) + \frac{1}{\sqrt{2}}\partial_\mu h. \end{pmatrix}. \quad (2.44)$$

By substituting the gauge fields in (2.20) into the Lagrangian (2.39). The Lagrangian becomes

$$\mathcal{L} = \frac{1}{2}\partial_\mu h\partial^\mu h + \frac{1}{4}g^2v^2W_\mu^+W^{-\mu} + \frac{g^2 + g'^2}{8}v^2Z_\mu Z^\mu + \mu^2h^2 + \dots, \quad (2.45)$$

from which it is clear to see that there is a massless photon  $A_\mu$ , 2 massive complex

vector bosons  $W_\mu^\pm$  with the mass  $m_W = gv/2$ , a massive vector boson  $Z_\mu$  with the mass  $m_Z = \sqrt{g^2 + g'^2}v/2$  and a massive higgs field  $h$  with the mass  $m_h = \sqrt{2\lambda}v$ . By measuring the value of  $m_W$  and  $g$ , the vacuum expectation value (VEV)  $v$  is determined to be 246 GeV. The mass of the Higgs boson is one of the free parameters of the SM, it is measured at the LHC.

### 2.3.4 Higgs Boson Production at the LHC

After the discovery of the Higgs boson in 2012 [4, 5], the ATLAS and CMS experiments at LHC started to measure its properties. By combining Run 1 and Run 2 data of ATLAS, the result for the Higgs boson mass measurement is  $m_H = 125.11 \pm 0.11$  GeV [24]. Theoretically, with the knowledge of the mass of Higgs boson and the other parameters of the SM already measured, all the properties of SM Higgs boson can be predicted, including the cross sections of Higgs boson production and the decay rates.

In the proton-proton collisions at the LHC, the main Higgs boson production processes are gluon-gluon fusion (ggF), vector-boson fusion (VBF), and associated production with vector bosons ( $VH$ ), a pair of top quarks ( $t\bar{t}H$ ) or bottom quarks ( $b\bar{b}H$ ), or a single top quark ( $tH$ ). The Feynman diagrams of these production processes are shown in Figure 2.4 (a)-(e).

The ggF process (Figure 2.4 (a)) is the dominant Higgs boson production process at the LHC, about 87.2% of Higgs bosons are produced via ggF process. The Higgs boson production through ggF is mediated by a heavy quark loop, mainly the top quark due to its large Yukawa coupling to the Higgs boson [25].

The second dominant Higgs boson production process is the VBF process (Figure 2.4 (b)), which contributes about 6.8% of Higgs boson production. The Higgs boson couples to two  $W$  or  $Z$  bosons that link to two quarks. Two jets produced through the VBF process are most likely to be forward-backward directed, so the signature is used to suppress the background [26].

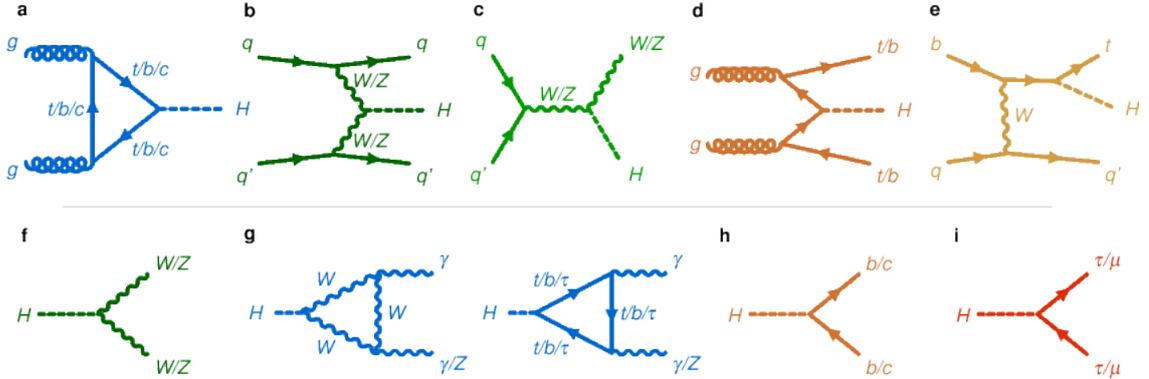


Figure 2.4: Feynman diagrams of Higgs boson production and decays. Top: Higgs production via (a) gluon-gluon fusion, (b) vector-boson fusion, and associate production with (c) weak vector bosons, (d)  $t$ - or  $b$ -quark pair, or (e) a single  $t$  quark. Bottom: The Higgs boson decays into (f) a pair of vector bosons ( $W^+W^-/ZZ$ ), (g) a pair of photons or a  $Z$  boson and a photon, (h) a pair of  $b$ - or  $c$ -quarks, and (i) a pair of charged leptons ( $\mu^-\mu^+/\tau^-\tau^+$ ) [7].

The production of a Higgs boson associated with a  $W/Z$  boson (Figure 2.4 (c)) is also called ‘Higgs-strahlung’, making up roughly 4% of LHC Higgs boson production. Before the Higgs boson was observed at 125 GeV, the Higgs-strahlung process was an important search mode for a low mass Higgs boson [27]. At the LHC, Higgs-strahlung is the most sensitive production mode to search for  $H \rightarrow b\bar{b}$  decay, because the leptonic decay of the vector boson significantly reduces the multijet background [28].

The contribution of  $t\bar{t}H$  or  $b\bar{b}H$  process (Figure 2.4 (d)) is approximately 1% each to the total Higgs production. As for other  $q\bar{q}H$  processes, the contributions are much smaller and not accessible for current experiments because of the small couplings between Higgs boson and lighter quarks.

The  $tH$  process accounts for only 0.05% of Higgs boson production. In the SM, the  $tH$  process is mediated by the weak interaction as shown in Figure 2.4 (e).

The SM prediction of the production cross sections of the processes above are shown in Figure 2.5. The measurement results of all production processes are in agreement with their SM predictions [7].

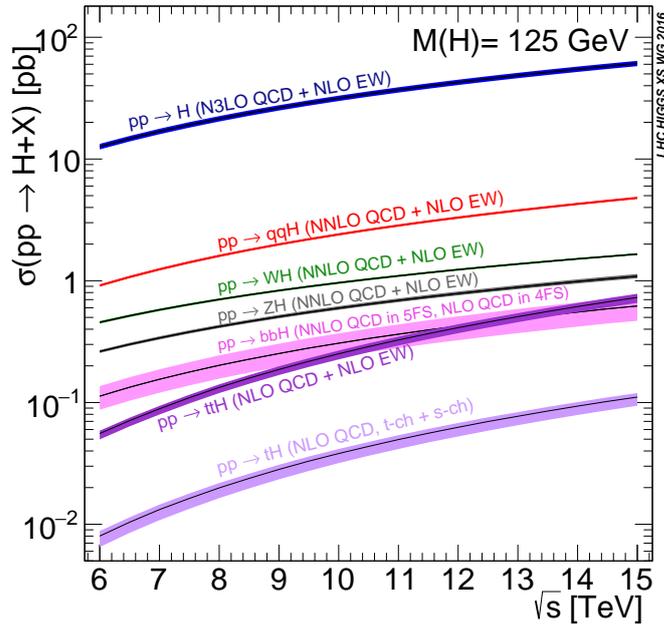


Figure 2.5: Predicted cross sections of different Higgs boson production processes as functions of LHC center-of-mass energy. Taken from Ref. [29].

### 2.3.5 Higgs Boson Decay

By the SM prediction, a 125 GeV Higgs boson is expected to have a natural width of about  $\Gamma_H = 4.1$  MeV [29]. Hence the expected lifetime of a Higgs boson is  $\tau_H = 1.6 \times 10^{-22}$  seconds. With such a short lifetime, the Higgs boson decays almost instantly after it is produced.

The Higgs boson can decay into two gauge bosons: two  $W$  bosons, two  $Z$  bosons, two photons, or a  $Z$  boson and a photon. Also the Higgs boson can decay into fermion pairs:  $b$ -quarks,  $c$ -quarks,  $\tau$ -leptons, or muons. These eight decay modes contribute more than 90% of total Higgs boson decays, their Feynman diagrams can be seen in Figure 2.4 (f)-(i), and the Figure 2.6 shows their branching ratios, also called branching fractions, describing the probability of the Higgs boson undergoing the decay mode.

The branching ratios of different decay modes with different Higgs boson production processes are also measured, the ratios of measured values to the SM predictions are shown in Figure 2.7. The measurement results are compatible with the SM pre-

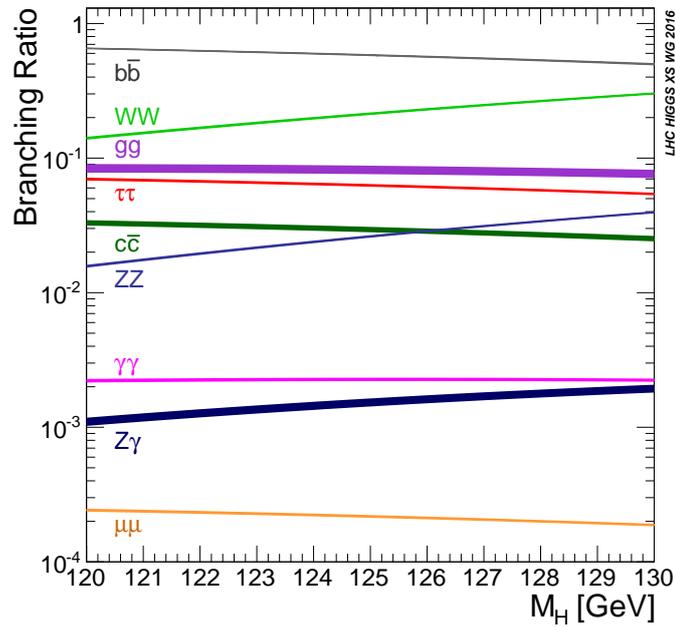


Figure 2.6: Branching ratios of different Higgs boson decay modes as functions of Higgs boson mass. Taken from Ref. [29].

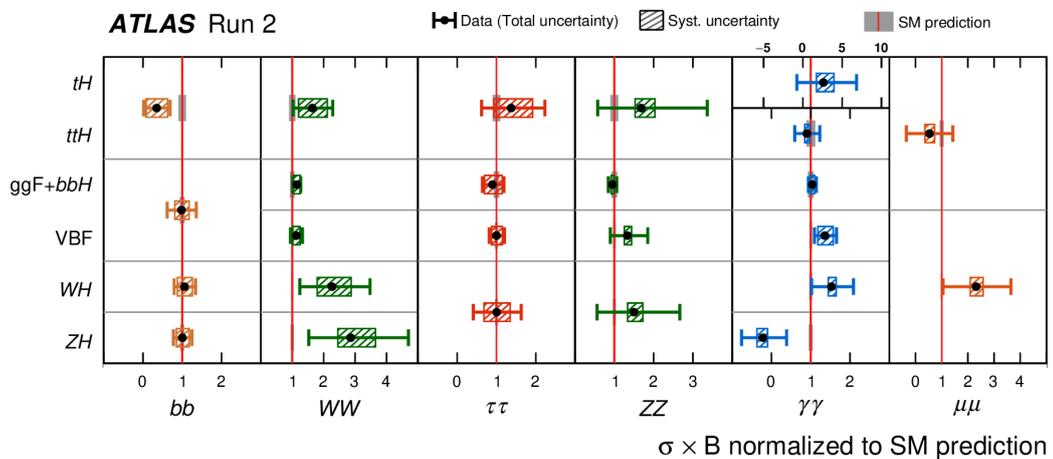


Figure 2.7: The ratio of measured event rate to the SM prediction for different combination of Higgs boson production and decay processes [7].

dictions. The Higgs boson may also decay into invisible particles or BSM particles, the ATLAS and CMS experiments are also searching for such decays [30, 31].

## 2.4 Extensions of the Standard Model

The SM has achieved remarkable success and has been confirmed by experiments. However, the SM is not a complete theory, there are still many unsolved questions. There is no dark matter (DM) [32] candidate in the SM. By the SM prediction neutrinos are massless, but the observation of neutrino oscillations indicates that neutrinos have mass [15]. The SM does not describe the gravity, one of the fundamental interactions. To solve these problems, a variety of theories that extend the SM is proposed.

As discussed in Section 2.3, the SM Higgs sector consists of only one complex Higgs doublet. As a result, there is only one neutral Higgs scalar in the SM. The SM has the ‘minimal’ Higgs sector but there is still room for non-minimal Higgs structures. The two-Higgs-Doublet Model (2HDM) [13] and the two-Higgs-Doublet Model with an additional singlet (2HDM+S) [9, 33] are two models with simplest possible extensions to the SM Higgs sector. These models introduce new phenomena, for example the charged Higgs bosons. Such extended Higgs sectors are required in supersymmetry (SUSY) [34]. There are many more complicated Higgs sectors [35, 36], but they will not be discussed here.

### 2.4.1 2HDM

The 2HDM introduces two complex doublet scalar fields  $\Phi_1$  and  $\Phi_2$  with hypercharge  $Y = 1$ , where  $\Phi_i = (\phi_i^+, \phi_i^0)^T$ . The most general potential is then given by [13]

$$\begin{aligned}
 V = & m_{11}^2 \Phi_1^\dagger \Phi_1 + m_{22}^2 \Phi_2^\dagger \Phi_2 - m_{12}^2 (\Phi_1^\dagger \Phi_2 + \Phi_2^\dagger \Phi_1) + \frac{\lambda_1}{2} (\Phi_1^\dagger \Phi_1)^2 + \frac{\lambda_2}{2} (\Phi_2^\dagger \Phi_2)^2 \\
 & + \lambda_3 \Phi_1^\dagger \Phi_1 \Phi_2^\dagger \Phi_2 + \lambda_4 \Phi_1^\dagger \Phi_2 \Phi_2^\dagger \Phi_1 + \frac{\lambda_5}{2} [(\Phi_1^\dagger \Phi_2)^2 + (\Phi_2^\dagger \Phi_1)^2],
 \end{aligned} \tag{2.46}$$

where all the parameters  $m_{ij}$ ,  $\lambda_i$  are real. The vacuum state, i.e. the minimum of the potential, gives

$$\Phi_1 = \begin{pmatrix} 0 \\ v_1/\sqrt{2} \end{pmatrix}, \Phi_2 = \begin{pmatrix} 0 \\ v_2/\sqrt{2} \end{pmatrix}, \quad (2.47)$$

where  $\sqrt{v_1^2 + v_2^2} = v = 246$  GeV is the SM vacuum expectation value. After the SSB, we can expand two complex scalar doublets around the vacuum state:

$$\Phi_a = \begin{pmatrix} \phi_a^+ \\ \frac{1}{\sqrt{2}}(v_a + \rho_a + i\eta_a) \end{pmatrix}, \quad a = 1, 2. \quad (2.48)$$

There are 8 degrees of freedom in total. Three of those corresponding to the Goldstone bosons: two charged scalars ( $G^\pm$ ) and a neutral pseudoscalar ( $G^0$ ). Three Goldstone bosons are eaten after SSB to give mass to  $W^\pm$  and  $Z$  bosons. The five remaining become physical ‘Higgs’ fields: two charged scalars ( $H^\pm$ ), two neutral scalars ( $h^0$  and  $H^0$ ) with  $m_{h^0} < m_{H^0}$ , and a pseudoscalar ( $a$ ).

The most important parameter of the model is the ratio of VEVs, noted as

$$\tan \beta = v_2/v_1. \quad (2.49)$$

The angle  $\beta$  describes the mixing between eigenstates in the charged sectors and the pseudoscalars. The charged Goldstone boson can be written as

$$G^\pm = \phi_1^\pm \cos \beta + \phi_2^\pm \sin \beta, \quad (2.50)$$

and the physical charged Higgs state is orthogonal to  $G^\pm$

$$H^\pm = -\phi_1^\pm \sin \beta + \phi_2^\pm \cos \beta. \quad (2.51)$$

The neutral Goldstone boson is

$$G^0 = \eta_1 \cos \beta + \eta_2 \sin \beta, \quad (2.52)$$

and the physical pseudoscalar  $a$  is orthogonal to  $G^0$

$$a = \eta_1 \sin \beta - \eta_2 \cos \beta. \quad (2.53)$$

The mixing between the eigenstates of the neutral Higgs fields is governed by the angle  $\alpha$ , which is defined by

$$\tan 2\alpha = \frac{-2m_{12}^2 + 2(\lambda_3 + \lambda_4 + \lambda_5)v_1v_2}{m_{12}^2(\frac{v_2}{v_1} - \frac{v_1}{v_2}) + \lambda_1v_1^2 - \lambda_2v_2^2}. \quad (2.54)$$

Two physical neutral Higgs are

$$h^0 = \rho_1 \sin \alpha - \rho_2 \cos \alpha, \quad (2.55)$$

$$H^0 = -\rho_1 \cos \alpha - \rho_2 \sin \alpha. \quad (2.56)$$

The SM Higgs can be identified as

$$H_{SM} = \rho_1 \cos \beta + \rho_2 \sin \beta = h^0 \sin(\alpha - \beta) - H^0 \cos(\alpha - \beta). \quad (2.57)$$

The masses of 5 physical Higgs are given by [37]:

$$m_{H^\pm}^2 = \left[ \frac{m_{12}^2}{v_1v_2} - \frac{1}{2}(\lambda_4 + \lambda_5) \right] (v_1^2 + v_2^2), \quad (2.58)$$

$$m_a^2 = \left[ \frac{m_{12}^2}{v_1v_2} - \lambda_5 \right] (v_1^2 + v_2^2), \quad (2.59)$$

$$m_{h^0, H^0}^2 = \frac{1}{2} (\text{tr} \mathcal{M}^2 \pm \sqrt{[\text{tr} \mathcal{M}^2]^2 - 4 \det \mathcal{M}^2}), \quad (2.60)$$

where  $\mathcal{M}$  is the mass matrix of neutral Higgs bosons:

$$\mathcal{M} = \begin{pmatrix} m_{12}^2 \frac{v_2}{v_1} + \lambda_1 v_1^2 & -m_{12}^2 + (\lambda_3 + \lambda_4 + \lambda_5) v_1 v_2 \\ -m_{12}^2 + (\lambda_3 + \lambda_4 + \lambda_5) v_1 v_2 & m_{12}^2 \frac{v_1}{v_2} + \lambda_2 v_2^2 \end{pmatrix}. \quad (2.61)$$

According to the different couplings of the Higgs fields to the SM particles, 2HDMs can be categorized into four types [13]: type-I, type-II, type-III (lepton-specific),

and type-IV (flipped), listed in Table 2.1. Figure 2.8 shows the current constraints on the parameters  $\cos(\beta - \alpha)$  and  $\tan\beta$  for different types of 2HDM. The data is consistent with the ‘alignment limit’ [13]:  $\cos(\beta - \alpha) = 0$ , which means the observed Higgs boson is identified with the light scalar  $h$  predicted by 2HDM.

2HDM type	Up-type quarks couple to	Down-type quarks couple to	Charged leptons couple to
Type-I	$\Phi_2$	$\Phi_2$	$\Phi_2$
Type-II	$\Phi_2$	$\Phi_1$	$\Phi_1$
Type-III (Lepton-specific)	$\Phi_2$	$\Phi_2$	$\Phi_1$
Type-IV (Flipped)	$\Phi_2$	$\Phi_1$	$\Phi_2$

Table 2.1: The couplings of quarks and charged leptons to the scalar doublets in different types of 2HDM.  $\Phi_1$  and  $\Phi_2$  correspond to vacuum expectation values  $v_1$  and  $v_2$  respectively [13].

### 2.4.2 2HDM+S

The 2HDM+S model is an extension of 2HDM with a complex singlet. The 2HDM+S is well motivated because its Higgs sector constitutes the scalar sector of the Next-to-Minimal Supersymmetric Standard Model (NMSSM) [39]. The Higgs sector of 2HDM+S contains 5 physical neutral Higgs bosons and two charged Higgs bosons ( $H^\pm$ ) [33]. The 5 neutral Higgs bosons are 3 scalars:  $(h_{125}, h^0, H^0)$ , and 2 pseudoscalars:  $(a, A)$ . The state  $h_{125}$  can represent the 125 GeV Higgs boson observed at LHC. The remaining states have the orders of masses:  $m_{h^0} < m_{H^0}$ , and  $m_a < m_A$ .

The pseudoscalar  $a$  can decay to SM fermions or gauge bosons. Figure 2.9 shows the branching ratios of  $a$  in different types of 2HDM+S model. The branching ratios of  $a$  in type-I 2HDM+S is  $\tan\beta$  independent, while in type-II, III and IV the branching ratios depend on  $\tan\beta$  [9].

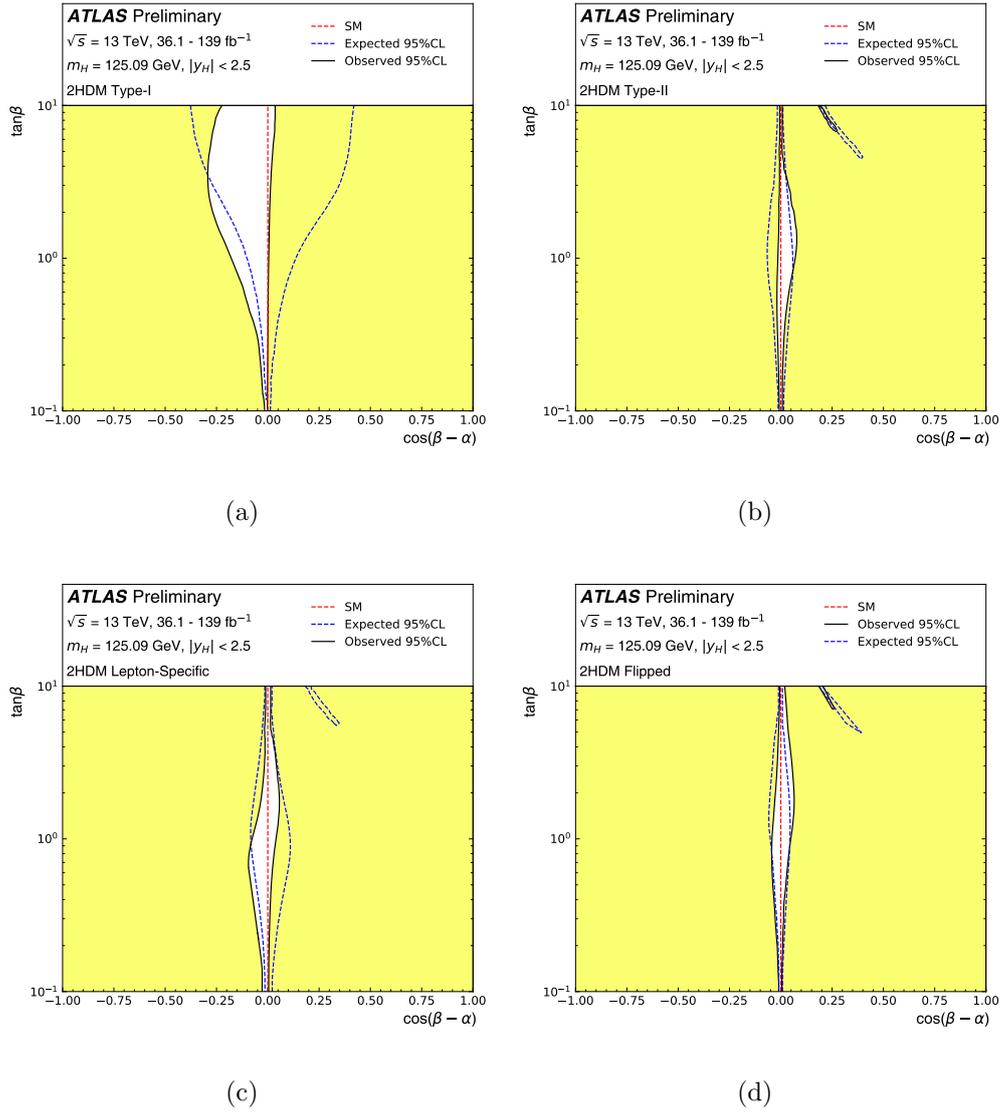


Figure 2.8: Regions of  $(\cos(\beta - \alpha), \tan \beta)$  plane excluded at 95% confidence level for (a) type-I, (b) type-II, (c) type-III and (d) type-IV 2HDM [38].

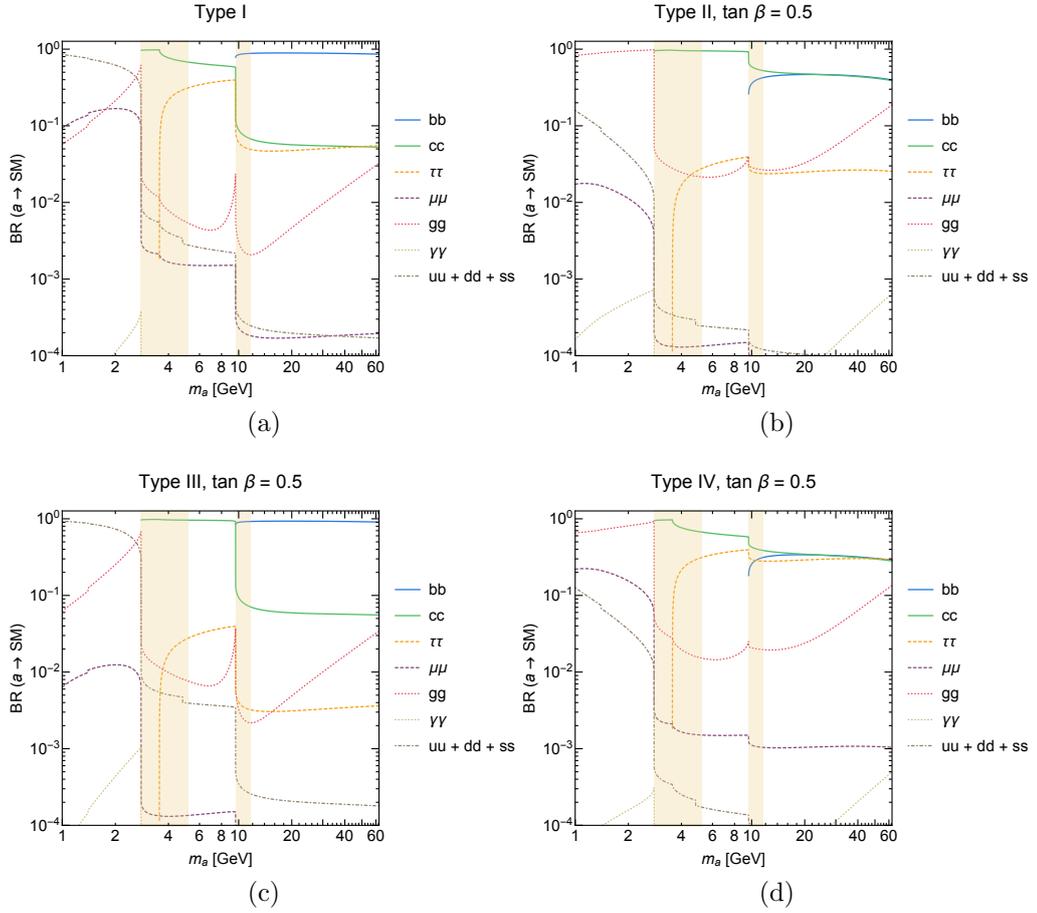


Figure 2.9: Branching ratios of  $a$  in the (a) type-I, (b) type-II, (c) type-III and (d) type-IV 2HDM+S. For type-II, III and IV, the plots show the branching ratios with  $\tan \beta = 0.5$  [9].

### 2.4.3 Experimental Searches for Extended Scalar Sectors

The extended scalar sectors of both 2HDM and 2HDM+S contain several BSM Higgs bosons, which, by the prediction, can interact with SM particles. Such interactions open the window of searches for new particles, and provide the possibility to explore new physics experimentally.

At the LHC, the ATLAS and CMS experiments have performed searches for the pseudoscalar  $a$  in various channels. To search for such  $a$ , we can assume  $a$  is lighter than the SM Higgs. Then there will be two possible exotic Higgs decays:  $H \rightarrow Za$  if the mass of  $a$  is less than the mass difference between SM Higgs and  $Z$  boson,  $H \rightarrow aa$  if the mass of  $a$  is below a half of SM Higgs mass. Most of searches were focusing on the  $H \rightarrow aa$  decays, such as  $H \rightarrow aa \rightarrow bb\mu\mu$  [40],  $H \rightarrow aa \rightarrow 4b$  [41, 42],  $H \rightarrow aa \rightarrow b\bar{b}\mu^+\mu^-$  [43],  $H \rightarrow aa \rightarrow \mu^+\mu^-\tau^+\tau^-$  [44] from ATLAS, and  $H \rightarrow aa \rightarrow \tau^+\tau^-\tau^+\tau^-/\mu^+\mu^-\tau^+\tau^-$  [45] and  $H \rightarrow aa \rightarrow b\bar{b}\tau^+\tau^-, b\bar{b}\mu^+\mu^-$  [46] from CMS. Since any significant excess over SM backgrounds were not found, the upper limits at 95% confidence level for the decay  $H \rightarrow aa$  were set by these searches. The results are show in Figure 2.10.

By the theory, the branching ratios of  $H \rightarrow Za$  and  $H \rightarrow aa$  can be adjusted independently [9]. This motivates searches for the decay  $H \rightarrow Za$ , even though many  $H \rightarrow aa$  searches have been done. However, the decay mode of  $H \rightarrow Za$  remains less studied. Few searches have been performed on  $H \rightarrow Za \rightarrow l^+l^- + \gamma\gamma$  [48, 49] and  $H \rightarrow Za \rightarrow l^+l^- + \text{jet}$  [50]. So far, no BSM particle is observed.

Chapter 6 of this thesis presents a search for  $H \rightarrow Za \rightarrow l^+l^- + \text{jet}$ . The previous search in this channel was published in 2020 [50], which did not find any excess. The observed upper limits at 95% confidence level were set on the Higgs boson production cross section times the branching ratio of  $H \rightarrow Za$  with values in the range of 17-340 pb for  $a$  with the mass between 0.5 and 4 GeV [50]. The corresponding upper limits of branching ratio of  $H \rightarrow Za$  is above 30%, which is quite trivial. This is one of the motivations to do the search for  $l^+l^- + \text{jet}$  final states again.

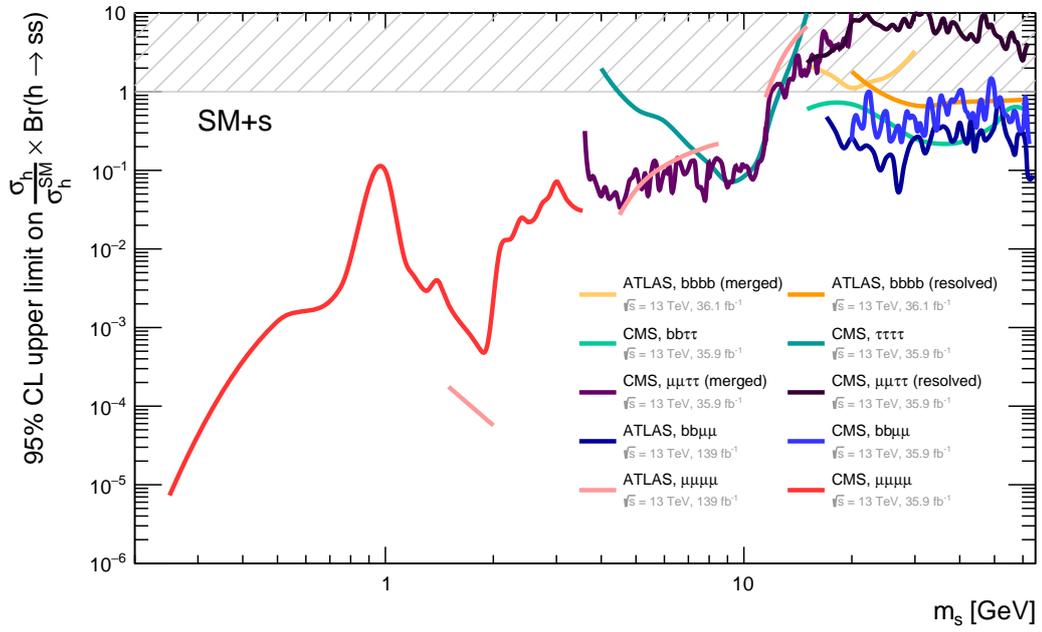


Figure 2.10: Observed 95% Confidence Level upper limits on the branching ratio of Higgs boson decays into a pair of additional (pseudo)scalars. Plot taken from [47]. The branching fractions of the new scalar to SM particles are taken from [29].

---

### The ATLAS experiment at the LHC

---

#### 3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [51] is currently the highest centre-of-mass energy particle accelerator and collider in the world, it is located at approximately 100 meters underground on the border between Switzerland and France at CERN. The LHC is installed in a circular tunnel with an approximately 27km circumference.

The LHC most often collides protons, i.e.  $pp$  collisions. The protons are supplied by an accelerator complex, shown in Figure 3.1. The proton beams are produced and accelerated in the following steps: the Linear accelerator 4 (Linac4) boosts negative hydrogen ions ( $\text{H}^-$ , consists of one proton and two electrons) to 160 MeV. Two electrons are stripped from the ions by passing through a thin carbon foil during the injection from Linac4 into the Proton Synchrotron Booster (PSB), where protons are accelerated to 2 GeV and then injected into the Proton Synchrotron (PS). The

PS accelerates the protons up to 26 GeV and delivers them to the Super Proton Synchrotron (SPS), which operates at up to 450 GeV. Finally, protons are injected into the LHC in two directions, one clockwise and another anti-clockwise, and are accelerated further. The LHC is designed to accelerate two proton beams up to 7 TeV, corresponding to 99.9999991% of the speed of light, and bring them to collision with a center-of-mass energy ( $\sqrt{s}$ ) of up to 14 TeV at four locations, where the four main experiments lie: ATLAS (A Toroidal LHC Apparatus) [52], CMS (Compact Muon Solenoid) [53], LHCb (Large Hadron Collider beauty) [54] and ALICE (A Large Ion Collider Experiment) [55]. The LHC also accelerates and collides other types of particle species, such as Pb-Pb collision.

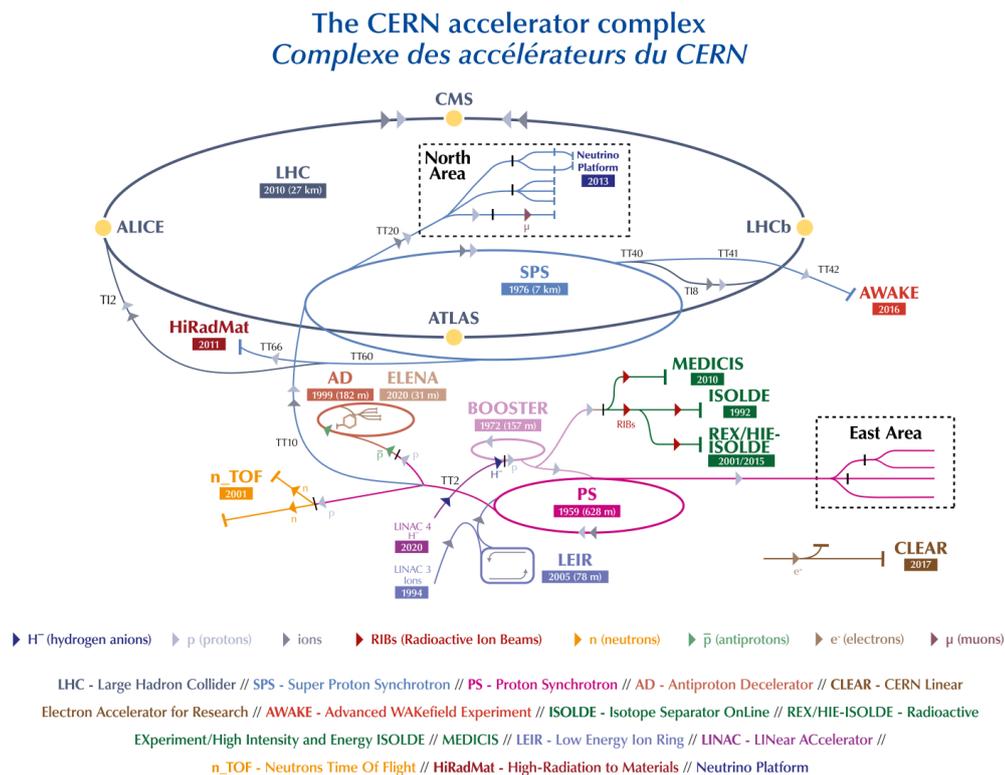


Figure 3.1: The CERN accelerator complex in 2022 [56].

By design, the peak instantaneous luminosity of the LHC is  $10^{34} \text{ cm}^{-2}\text{s}^{-1}$  for  $pp$  collisions at  $\sqrt{s} = 14 \text{ TeV}$ . Each proton beam in the LHC contains 2808 bunches, and each bunch has about  $10^{11}$  protons. At the full luminosity, these bunches are spaced at 25 ns time intervals, which means that bunches collide 40 million times

and generate about 1 billion particle collisions per second. The rate of collision is  $R = L \times \sigma$ , where  $L$  is the instantaneous luminosity,  $\sigma$  is the total inelastic proton-proton cross section. This effect of multiple  $pp$  collisions in the same bunch crossing is called the pile-up. As an example, Figure 3.2 shows the mean number of collisions per bunch crossing,  $\langle\mu\rangle$ , recorded by ATLAS each year during the LHC Run 2 (year of 2015-2018) and Run 3 (year of 2022-recent).

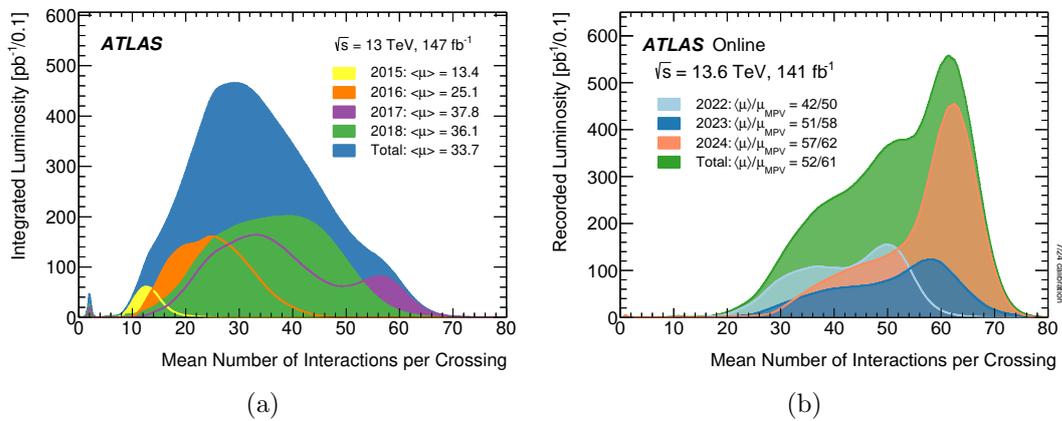


Figure 3.2: The distribution of the mean number of collisions per bunch crossing recorded by ATLAS for each year of (a) Run 2  $pp$  collisions [57] and (b) Run 3  $pp$  collisions (Until 19 August 2024) [58].

The Run 1 of the LHC started in 2009. The centre-of-mass energy of the stable beams for  $pp$  collision reached 7 TeV in 2010 and then 8 TeV in 2012. The ATLAS recorded intergrated luminosity of 5 fb<sup>-1</sup> and 21 fb<sup>-1</sup> of data at  $\sqrt{s} = 7$  and 8 TeV\*, respectively. The peak instantaneous luminosity was 3.6 (7.7)  $\times 10^{33}$  cm<sup>-2</sup>s<sup>-1</sup> at  $\sqrt{s} = 7$  (8) TeV. The Run 1 ended in 2013 and the first long shutdown (LS1) began. Run 2 operated between 2015 and 2018. During the Run 2, the collision energy reached 13 TeV and the peak instantaneous luminosity reached  $2.1 \times 10^{34}$  cm<sup>-2</sup>s<sup>-1</sup>. ATLAS recorded 147 fb<sup>-1</sup> of data, 140 fb<sup>-1</sup> of which have good quality for physics analysis. The second long shutdown (LS2) lasted between 2018 and 2022, while the LHC and detectors got Phase-I upgrade during this period. The first beam of Run 3 came in July 2022 and the collision energy reached a new record: 13.6 TeV. Until the end of year 2023, ATLAS has already recorded 66 fb<sup>-1</sup> of data in Run 3.

\*1 b = 10<sup>-24</sup> cm<sup>2</sup>, 1 fb = 10<sup>-15</sup> b.

According to the plan, Run 3 will end at 2025 and then will be the long shutdown 3 (LS3). In the 3-year LS3 period, the LHC will undergo a major upgrade to become High-Luminosity LHC (HL-LHC) in 2029, which will deliver proton beams with the instantaneous luminosity reaching  $5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  [59].

## 3.2 The ATLAS detector

ATLAS [52] is the largest volume detector at the LHC. It weighs around 7000 tonnes, and has an approximately cylindrical geometry with 46 m length and a 25 m diameter. The ATLAS detector is designed to cover a wide range of physics, including measurements of Higgs boson properties and searches for the BSM physics. The overview of the ATLAS detector can be seen in Figure 3.3.

The detector is divided into three main subsystems: the inner tracking detector system, the calorimeter system, and the muon spectrometer. With the combination of these subsystems, the ATLAS detector is able to detect different physics objects, and covers a near  $4\pi$  solid angle around the collision point. There are also few smaller detector systems covering the ATLAS forward region. A two-level trigger system is used to select interesting events for recording and reduce the event rate. These subsystems are discussed in the following.

The ATLAS experiment uses a right-handed coordinate system, which has the origin at the nominal interaction point (IP) in the center of the detector. The  $x$ -axis points from the IP to the center of the LHC ring, the  $y$ -axis points upwards, and the  $z$ -axis is along the beam pipe so that a right-handed reference frame is formed. For a particle with momentum  $p$  and the angle  $\theta$  with respect to the  $z$ -axis, the transverse momentum is  $p_T = p \sin \theta$ , the azimuthal angle in the transverse plane is  $\phi$ , and the rapidity  $y$  is defined as

$$y = \ln \left( \frac{E + p_L}{E - p_L} \right), \quad (3.1)$$

where  $E$  is the particle energy, and  $p_L$  is the longitudinal momentum  $p_L = p \cos \theta$ .

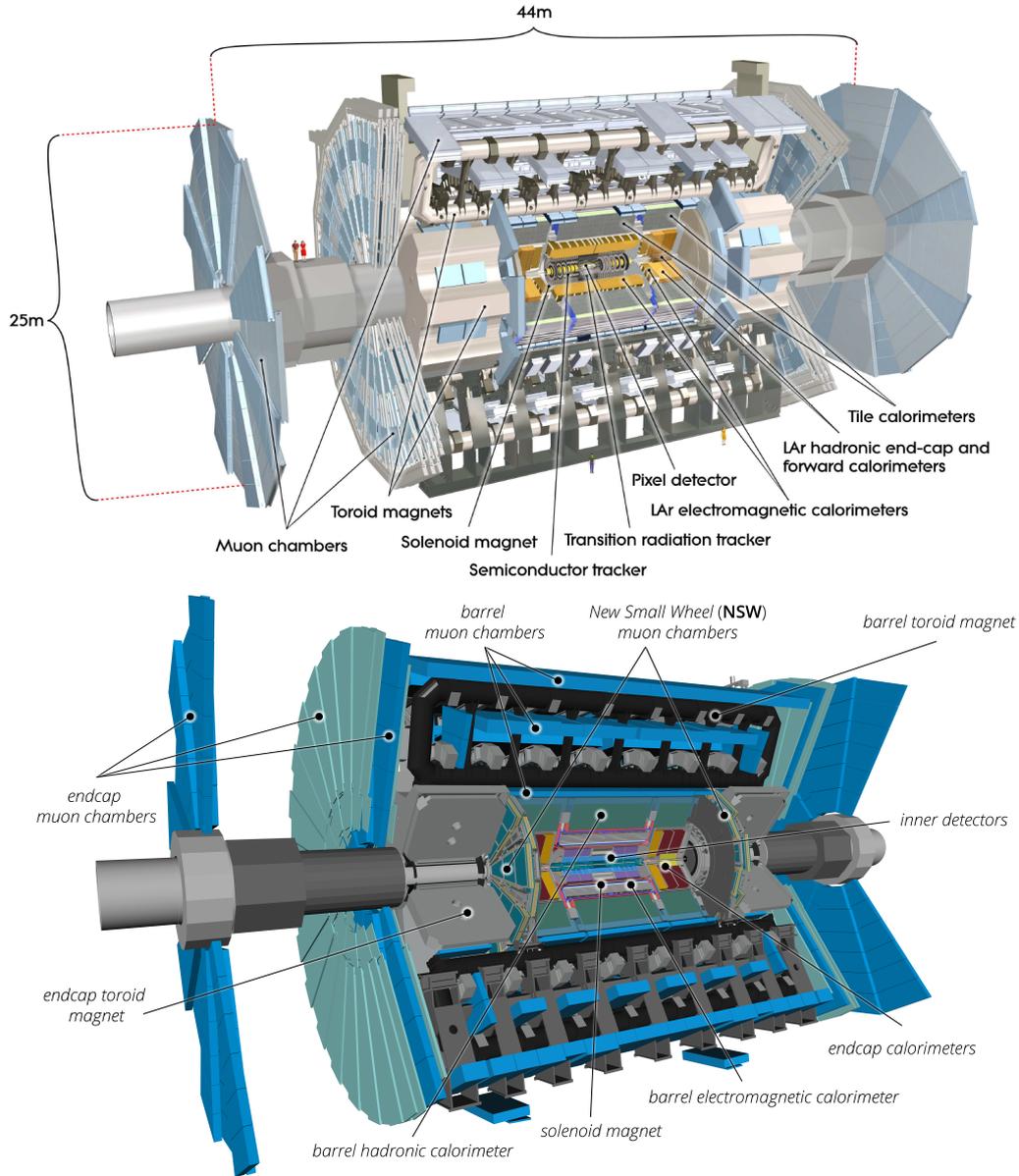


Figure 3.3: An overview of the ATLAS detector (top) in Run 2, and (bottom) in Run 3 [52, 59]. The main difference is the addition of New Small Wheels in Run 3.

The rapidity of particles emitted along the direction of  $z$ -axis is infinity, while particles orthogonal to the  $z$ -axis have zero rapidity. Notice that the difference in rapidity,  $\Delta y$ , is invariant under the Lorentz boosts along the  $z$ -axis. For a particle near the speed of light, its rapidity can be approximated by the pseudorapidity  $\eta$ , given by

$$\eta = -\ln \left[ \tan \left( \frac{\theta}{2} \right) \right]. \quad (3.2)$$

Thus the kinematics of particles can be expressed on the  $(\eta, \phi)$  plane. The angular distance between particles is therefore  $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ .

### 3.2.1 Inner Detector

The Inner Detector (ID) system [60, 61] is designed to measure the tracks of charged particles in ATLAS with high precision. It has a cylindrical envelope 7024 mm long and 1150 mm in radius, immersed in a 2 T solenoidal magnet field. It is able to cover the charged trajectories within the pseudorapidity range  $|\eta| < 2.5$ .

The structure of the ID system can be seen in Figure 3.4. The ID is composed of three sub-detectors that surround the beampipe: the silicon pixel detector (Pixel), the semiconductor tracker (SCT), and the transition radiation tracker (TRT). During the LS1 period, a new innermost pixel layer, Insertable B-Layer (IBL), was installed. The positions and radii of these sub-detectors are shown in Figure 3.5.

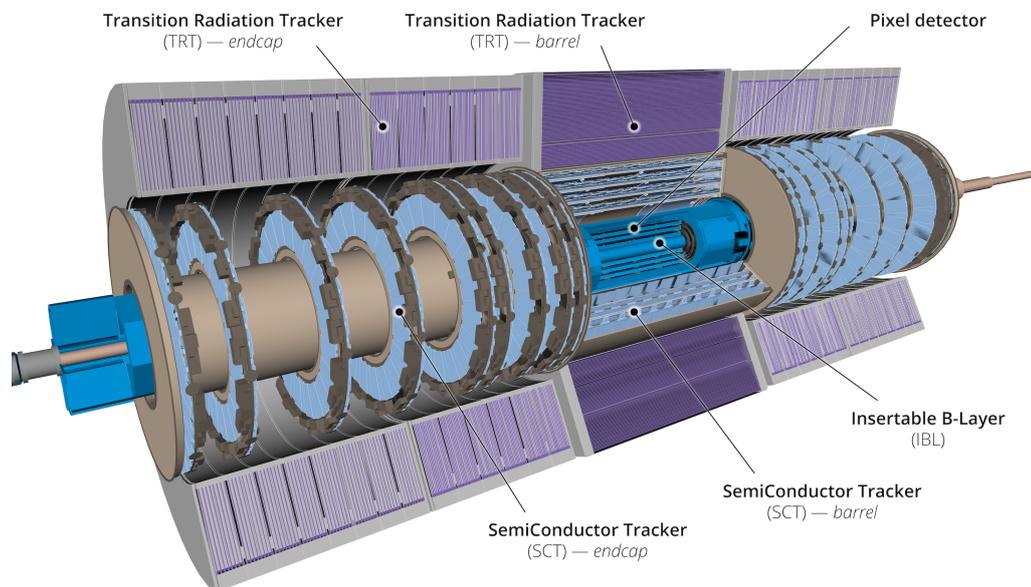


Figure 3.4: An overview of the ATLAS Inner Detector system [59].

The Pixel detector [62] is the innermost part of the ID. The active region of the Pixel detector contains three barrel layers: B-Layer, Layer 1 and Layer 2, with the radius of 50.5 mm, 88.5 mm and 122.5 mm respectively, and three disk layers at each

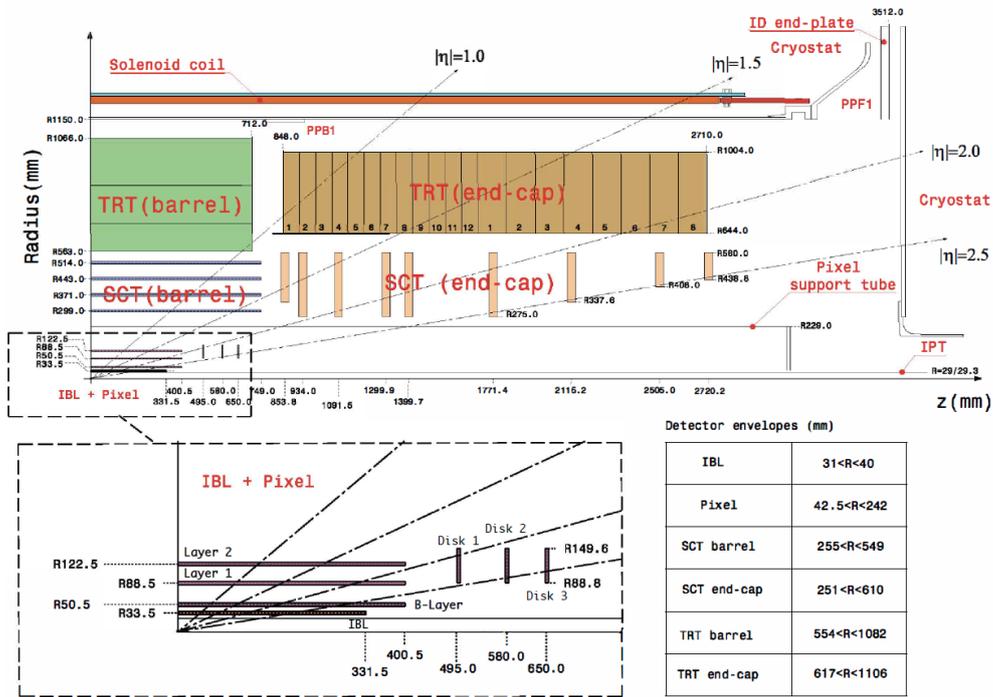


Figure 3.5: The layout of one quadrant of the ATLAS ID, with the IBL [59]. All dimensions are in mm.

end of the barrel. Many particles such as  $B$ -hadrons decay inside the beampipe, the performance of the ATLAS experiment strongly rely on the innermost pixel layer. Therefore, a new fourth layer, the IBL [63], was added between the existing B-Layer and the new narrower beampipe. The radius of the IBL is 33.5 mm, while the new beampipe is between 23.5 and 30 mm. There are 92 million pixel channels in total, the size of silicon pixel sensors in three outer layers is  $50 \times 400 \mu\text{m}^2$ , and it is  $50 \times 250 \mu\text{m}^2$  in the IBL. The resolutions of the Pixel detector in the barrel are  $10 \mu\text{m}$  in azimuth ( $R-\phi$ ) and  $115 \mu\text{m}$  in axial ( $z$ ), whereas in the end-caps, the resolutions are  $10 \mu\text{m}$  in azimuth ( $R-\phi$ ) and  $115 \mu\text{m}$  in radius ( $R$ ). By design, the charged track originating from the collision region is identified with at least three points provided by the Pixel detector. Figure 3.6 shows the improvement on the resolution of track impact parameters. The track impact parameters are explained in Section 3.3.1.

The SCT [64] covers radial distance from 299 mm to 560 mm, surrounding the Pixel detector. The SCT has four cylinders in the barrel region and nine disks each side of

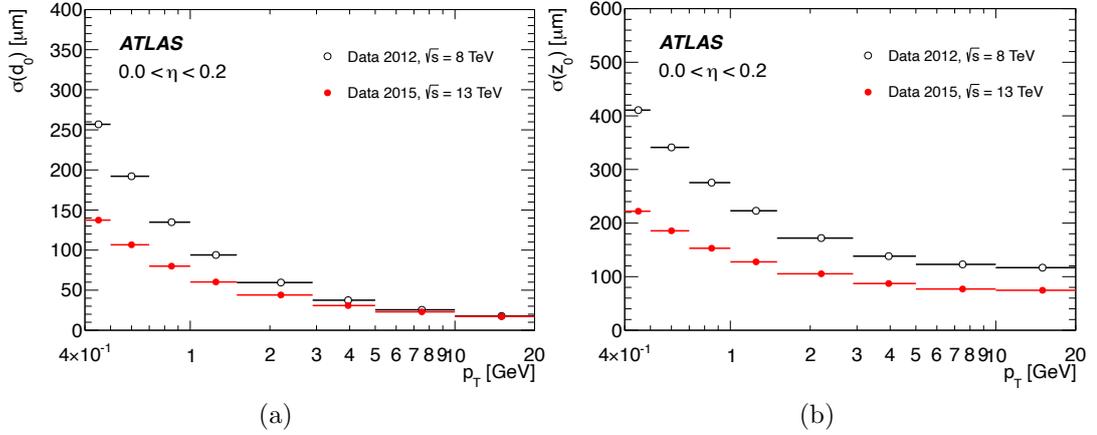


Figure 3.6: Comparison between 2012 and 2015 (before and after IBL installed) for resolutions of (a) the transverse impact parameter, and (b) the longitudinal impact parameter as functions of track  $p_T$  [59].

the end-cap region, with each layer has sensors on both sides so that able to read out a position in two dimensions. Usually, the SCT provides eight strip measurement, corresponding to four space-points, for a charged track, with the resolutions of  $17 \mu\text{m}$  in  $R\text{-}\phi$  and  $580 \mu\text{m}$  in  $z$  ( $R$ ) for barrels (end-caps). There are 4088 modules and approximately 6.3 million strips in the SCT, together with the Pixel detector, they provide high-precision tracking of the charged particles within  $|\eta| < 2.5$  for the ATLAS experiment.

The outermost part of the ID is the TRT [65]. The basic elements of the TRT are 4 mm diameter straw drift tubes. 52544 straws in the barrel region are 144 cm long and are parallel to the beam axis, while in each end-cap, 122880 straws are 37 cm long and arranged radially. These straws are embedded in transition radiation material, and filled with gas mixture of 70% Xe, 27%  $\text{CO}_2$  and 3%  $\text{O}_2$  during the normal operation. However, by the end of Run 1, leaks were observed in several TRT modules. Leaking modules were partially repaired during LS1, but some of those are located in inaccessible areas and therefore impossible to repair. To reduce the cost on Xenon, during Run 2 leaking modules are filled with the Argon-based gas mixture: 70% Ar, 27%  $\text{CO}_2$  and 3%  $\text{O}_2$  [66]. When charged particles pass through the transition radiation material, it will emit transition radiation photons. For instance, electrons emit X-rays that will be absorbed by the gas mixture. Other

particles with low Lorentz factors ( $< 1000$ ) do not produce transition radiation photons. Hence the TRT has the ability to identify the electrons and reject hadrons such as pions. The TRT straw has a anode at the center with a diameter of  $31 \mu\text{m}$ , made up of gold-plated tungsten wire. The anodes of straws operate at ground potential, while the cathodes have a high voltage about  $1530 \text{ V}$ . When a charged particle goes across a straw, it ionises gas and creates clusters of electrons and ions. These ionised electrons then drift to the central anode wire in a strong electric field and produce a detectable signal. By measuring the time between the ionisation and electron clusters hit the anode, the straw can measure the distance from the charged track to the central anode. On average, a charged particle with  $|\eta| < 2$  and  $p_T > 0.5 \text{ GeV}$  can cross 36 straws, which make the TRT able to provide continuous tracking between the radii of  $563 \text{ mm}$  and  $1066 \text{ mm}$  with a resolution of  $130 \mu\text{m}$  in  $R\text{-}\phi$  per straw.

In summary, The ATLAS ID measures several high-precision space-points and more than 30 straw hits on average for a charged track with  $|\eta| < 2.5$ . And the ID can implement the particle identification based on the transition radiation. The precise tracking measurements in ID is the fundamental of most analysis in ATLAS.

### 3.2.2 Calorimeters

The ATLAS Calorimeter system [67, 68] is used to measure the energies and positions of most of the particles coming from collisions, except muons and neutrinos. The ATLAS calorimeters are sampling calorimeters, in which most of the incoming particles are stopped and their energies absorbed by layers of passive absorber alternating with active media. The structure of the ATLAS Calorimeter system can be seen in Figure 3.7. It has two major parts: the electromagnetic (EM) calorimeter and the hadronic calorimeter. These calorimeters cover the full  $\phi$  range and the region of  $|\eta| < 4.9$ . The Calorimeter system uses two different technologies: Liquid Argon (LAr) for all of the electromagnetic calorimeters, and the hadronic calorimeters in the endcap and forward regions, while scintillating Tiles are used for the

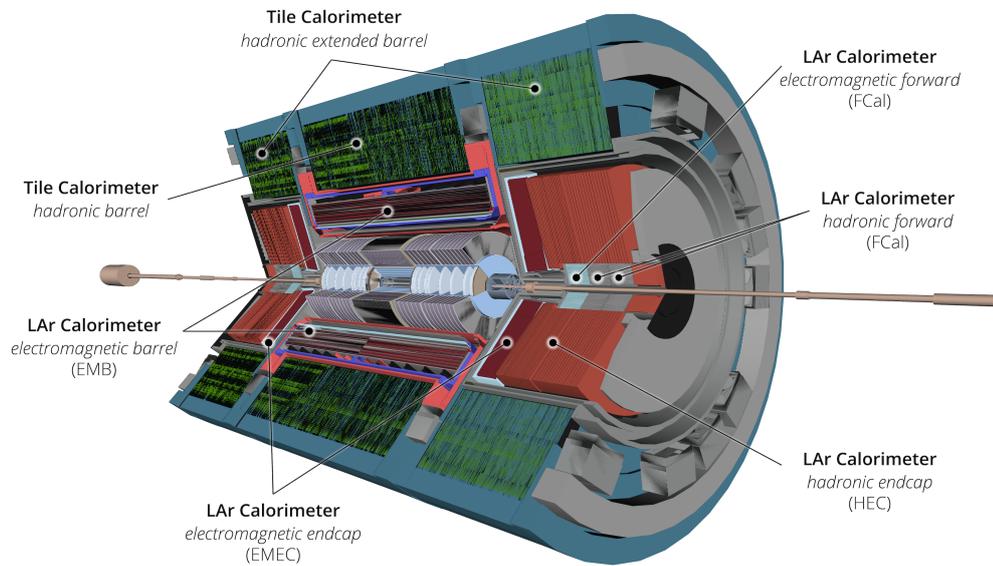


Figure 3.7: An overview of the ATLAS Calorimeter system [59].

hadronic calorimeters at larger radius region.

### LAr EM calorimeter

The LAr EM calorimeter system includes the LAr electromagnetic barrel calorimeter (EMB) covering  $|\eta| < 1.475$ , and the LAr electromagnetic endcap calorimeter (EMEC) covering  $1.375 < |\eta| < 3.2$ . The LAr sampling calorimeters use liquid argon as active media and lead (Pb) as the absorber, the lead plates are arranged in the shape of accordion. Once an incoming electromagnetic particle (electron, photon,...) hits the absorber, it will produce an electromagnetic shower in the liquid argon, which consist of a bunch of electrons and photons. The development of the shower is governed by the radiation length ( $X_0$ ) of the detectors. The radiation length of a material is the mean distance that the energy of a electron is reduced by a factor of  $1/e$ . With the design of sandwich of Pb absorber and LAr active media, the total thickness of EM calorimeters is  $> 22X_0$  in the barrel and  $> 24X_0$  in the endcaps. Thus the EM shower deposits almost all its energy in the EM calorimeters.

The module of the EM calorimeter has three layers in depth, called front, middle

and back layer. A layer of pre-sampler detector is added in front of the calorimeters in the region  $|\eta| < 1.8$ . It is used to estimate the energy loss because of the material upstream of the calorimeter. These layers have fine granularities up to  $\Delta\eta \times \Delta\phi = 0.025/8 \times 0.1$ , so the detector is providing precision measurements of electrons and photons for the experiment. Figure 3.8 shows the performance of different layers of LAr calorimeters in Run 2.

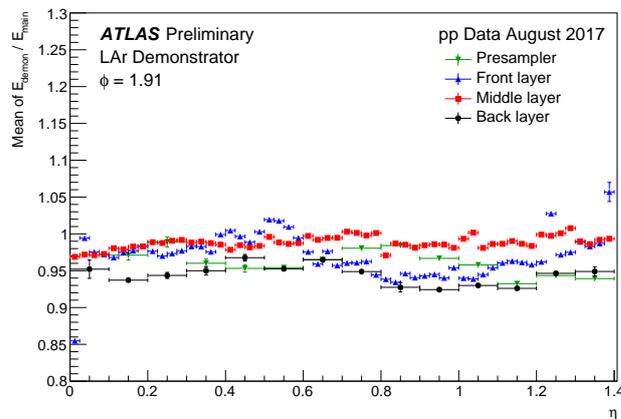


Figure 3.8: Ratio of the LAr calorimeter Phase-I readout to the summed cell energies as a function of  $\eta$  [69].

## Hadronic calorimeter

The hadronic calorimeter covering  $|\eta| < 4.9$  uses different techniques in different regions. It is divided into three main parts: the LAr hadronic endcap calorimeter (HEC), the LAr forward calorimeter (FCAL) and the scintillating tile calorimeter.

The scintillating tile calorimeter is a cylindrical structure surrounding the LAr calorimeters. It is divided into a central barrel covering  $|\eta| < 1.0$  and two extended barrels covering  $0.8 < |\eta| < 1.7$ , as shown in Figure 3.7. The tile calorimeter uses layers of steel as the absorber and scintillator tiles as the active material. A hadronic shower is generated when a hadron hits the absorber, and the particles of the shower stimulate the scintillators to produce photons. The effective nuclear interaction length of the tile calorimeter is  $\lambda = 20.7$  cm, representing the mean-free-path for inelastic nuclear interaction. The tile calorimeter has three layers and the total thickness

is about  $7.4 \lambda$ , so most of hadronic shower energy is absorbed. These three layers have different granularities. The first and second layers have the granularity  $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ , while the last layer has  $\Delta\eta \times \Delta\phi = 0.2 \times 0.1$ . The granularity of the hadronic calorimeter is coarser than that of the EM calorimeter, but it is sufficient to the jet reconstruction and missing transverse momentum ( $E_T^{\text{miss}}$ ) measurements. Figure 3.9 shows the good agreement between experimental data and simulated multijet events.

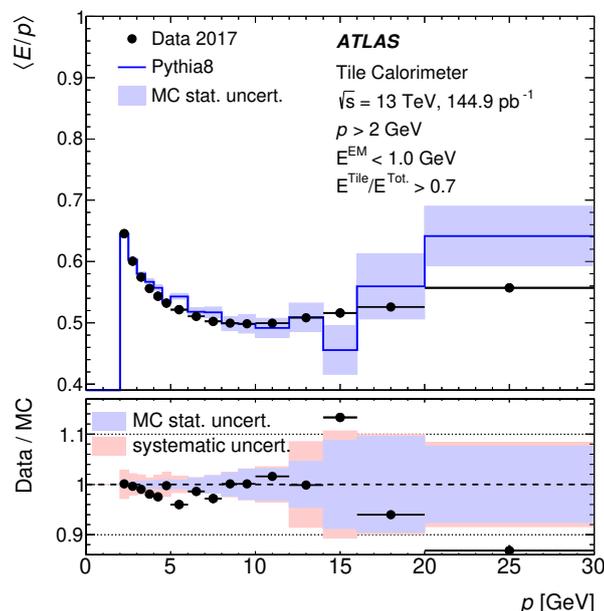


Figure 3.9: The average ratio of energy to momentum ( $\langle E/p \rangle$ ) measured by the tile calorimeter as a function of hadron momentum [70].

The hadronic endcap calorimeter uses flat copper (Cu) plates and LAr as the absorber and active material respectively. It consists of two independent wheels in each endcap: front wheel and rear wheel, both of which cover the range  $1.5 < |\eta| < 3.2$ . The granularity of HEC is coarser than EM calorimeters. In the region  $1.5 < |\eta| < 2.5$ , the granularity of HEC is  $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ , and it is  $0.2 \times 0.2$  in the region  $2.5 < |\eta| < 3.2$ .

The LAr forward calorimeters modules are located at high- $\eta$  region:  $3.1 < |\eta| < 4.9$  and it is about 4.7 m from the interaction point. The FCAL is split into three sections with 45 cm depth each: the first section (FCAL1) uses copper as the absorber

to optimise the EM measurements, the other two (FCAL2, FCAL3) are made of tungsten (W) for the hadronic measurements.

With the combination of different calorimeters, the ATLAS Calorimeter system absorbs EM and hadronic showers effectively, and significantly reduces the particles going into the muon system.

### 3.2.3 Muon Spectrometer

The muon spectrometer (MS) [71] is the outer part of the ATLAS detector. It is designed to detect charged particles escaping from the barrel and endcap calorimeters within  $|\eta| < 2.7$ , and measure their momentum. The overall layout of the muon spectrometer is shown in Figure 3.10. It consists of several chambers with different functions: Cathode Strip Chambers (CSC) and Monitored drift tubes (MDT) for precision tracking, Resistive plate chambers (RPC) and Thin gap chambers (TGC) for triggering. For the Run 3, the original TGCs in the inner wheels of the endcap are not sufficient for the triggering and tracking, therefore the endcap inner wheels have been completely removed and replaced by the New Small Wheels (NSW). Furthermore, the muon system is immersed in a toroidal magnetic field that is mostly orthogonal to the muon trajectories, thus the muon trajectory is bent in  $\eta$  plane but  $\phi$  keeps unchanged. Based on the magnetic deflection of tracks, the detector is able to measure the momentum of muons.

Providing continuous tracking is impossible for the muon spectrometer due to its large size. The track is reconstructed based on three detected stations: the first station is close to where the track exits the calorimeter, the second station is inside the magnetic field, the third station is outside the magnetic field. The detectors at each station are multilayered, so in total at least six space-points are provided. These points are connected with straight lines, then fitted by a curved path and match the track from the ID.

In the barrel region, these stations are called barrel inner, barrel middle and barrel

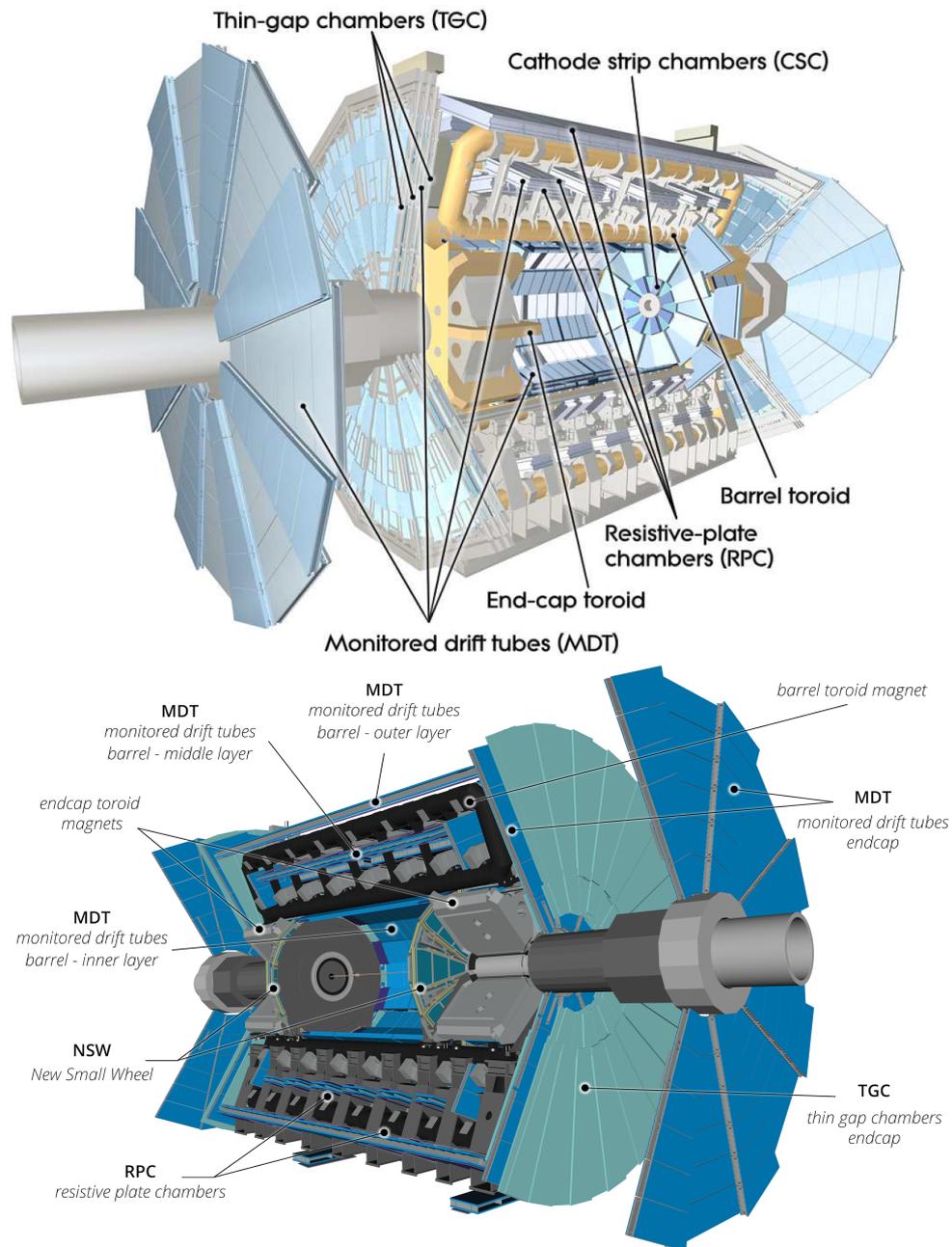


Figure 3.10: An overview of the ATLAS Muon system (top) in Run 2, and (bottom) in Run 3 [52, 59].

out, each station uses layers of MDT for the precision momentum measurement. The drift tube in MDT chamber has a diameter of 30 mm, and is filled with gas mixture. The electrons generated from the ionization by passing particles drift and are collected by the anode wire at the center. With the layers of drift tubes, the MDT chamber can achieve  $35 \mu\text{m}$  on the resolution of track position measurement.

The RPC is responsible for the triggering and the  $\eta$ - $\phi$  measurement in the barrel region at  $|\eta| < 1.05$ . It delivers the signal of the passing particle in a few tens of nanoseconds with a 2 ns overall resolution.

Each endcap of MS consists of three wheels: inner, middle and outer, and an extended endcap ring. The endcaps use MDTs for precision tracking, which covers the range up to  $|\eta| = 2.7$ . The fast triggering and the measurement of non-bending coordinate are performed by TGC, covering  $1.05 < |\eta| < 2.4$ . The outer wheels only have MDTs, while the middle wheels contain both MDTs and TGCs. The extended endcap rings of MDT are located between the inner and middle wheels, providing the tracking measurement in the region  $1.05 < |\eta| < 1.3$ , which are not covered by the outer wheels. During the Run 1 and 2, the inner wheels used MDTs and CSCs for precision tracking. While in Run 3, NSWs [72] have capabilities of both precision tracking and triggering.

The muon reconstruction performance of MS in Run 2 can be seen in Figure 3.11.

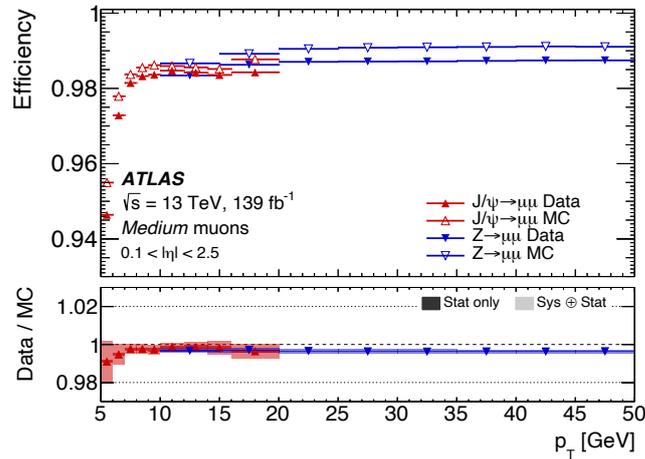


Figure 3.11: Muon reconstruction efficiency in  $J/\Psi \rightarrow \mu\mu$  and  $Z \rightarrow \mu\mu$  events as a function of muon  $p_T$  [73].

### 3.2.4 Magnet System

As shown in previous sections, the ID and Muon Spectrometer are immersed in the magnetic fields to measure the momenta of the charged particles. These magnet fields are generated by superconducting magnets of the ATLAS magnet system. The layout of the ATLAS magnet system can be seen in Figure 3.12. It is divided into four parts: a central solenoid, a barrel toroid and two endcap toroids.

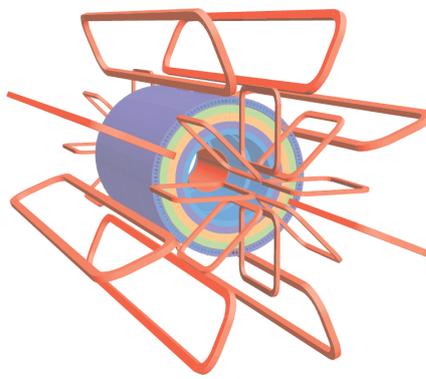


Figure 3.12: The geometry of the ATLAS magnet system [52].

The central solenoid is placed parallel to the beam pipe and surrounds the ID. It is designed to produce a 2 T axial magnetic field for the ID. The solenoid is installed in front of the calorimeters, so its thickness is minimised to reduce the impact on the measurement of particle kinematics. The barrel and endcap toroids generate toroidal magnetic fields that are mostly perpendicular to the tracks of muons. Each toroid magnet is made up of eight coils as in Figure 3.12. The locations of toroid magnets in the muon system can be seen in Figure 3.10. The barrel toroid and endcap toroids provide 0.5 T and 1 T toroid magnetic fields for the muon spectrometer respectively.

### 3.2.5 Trigger and Data Acquisition System

The ATLAS Trigger and Data Acquisition System (TDAQ) [74] is designed to select events that interesting for physics analyses, and read out the data for offline processing. The trigger system of ATLAS experiment in Run 2 and 3 has two levels, the

hardware-based level-1 trigger (L1) and the software-based high-level trigger (HLT). Both levels of triggers make decisions and send them to the Data Acquisition (DAQ) system, where data is buffered and then transferred to offline storage. A diagram of the TDAQ system is shown in Figure 3.13. In comparison with the Run 2, the main parts of TDAQ system upgraded in Run 3 are the Level-1 Calorimeter trigger electronics, including the new eFEX, jFEX and gFEX processors, the L1Muon trigger logic for the endcap, the Central Trigger Processor (CTP) and the muon-to-CTP interface (MUCTPI) [74].

The L1 trigger consists of the Level-1 Calorimeter trigger (L1Calo), the Level-1 Muon trigger (L1Muon), the Level-1 Topological Processor (L1Topo) and the Central Trigger. The L1Calo trigger receives reduced-granularity data from calorimeters and searches for high- $p_T$  electrons, photons, taus, jets and also calculates missing transverse energy. The details of L1Calo and its electronics are discussed in section 4.1. The L1Muon trigger identifies high- $p_T$  muons based on the inputs from RPCs, TGCs and NSWs of the muon system, and then sends data to CTP through MUCTPI. The L1Topo receives trigger objects (TOBs) from L1Calo and L1Muon, including  $E_T$  and coordinates information. L1Topo uses these data to select events based on event topology (e.g. rapidity gaps or invariant masses). The CTP makes the final L1 decision based on the trigger signals from L1Calo, L1Muon and L1Topo. The time from a bunch crossing to the final L1 trigger decision, also called the L1 latency, is less than  $2.5 \mu\text{s}$ . The rate of LHC collision is 40 MHz, and the L1 trigger selects up to 100k events per second.

The events accepted by L1 trigger will be sent to the HLT, where the selection is based on more detailed information than L1, for example the full-granularity data from the calorimeters. The HLT is software-based, and operates on a large farm of CPUs. It executes the algorithms to analyse the properties of events and makes decisions. On average, the HLT reduces the event rate from 100 kHz to 3 kHz.

The Fast TracKer (FTK) project [75] was a hardware-based fast tracking system, designed for providing full track information to HLT after L1 trigger makes decisions.

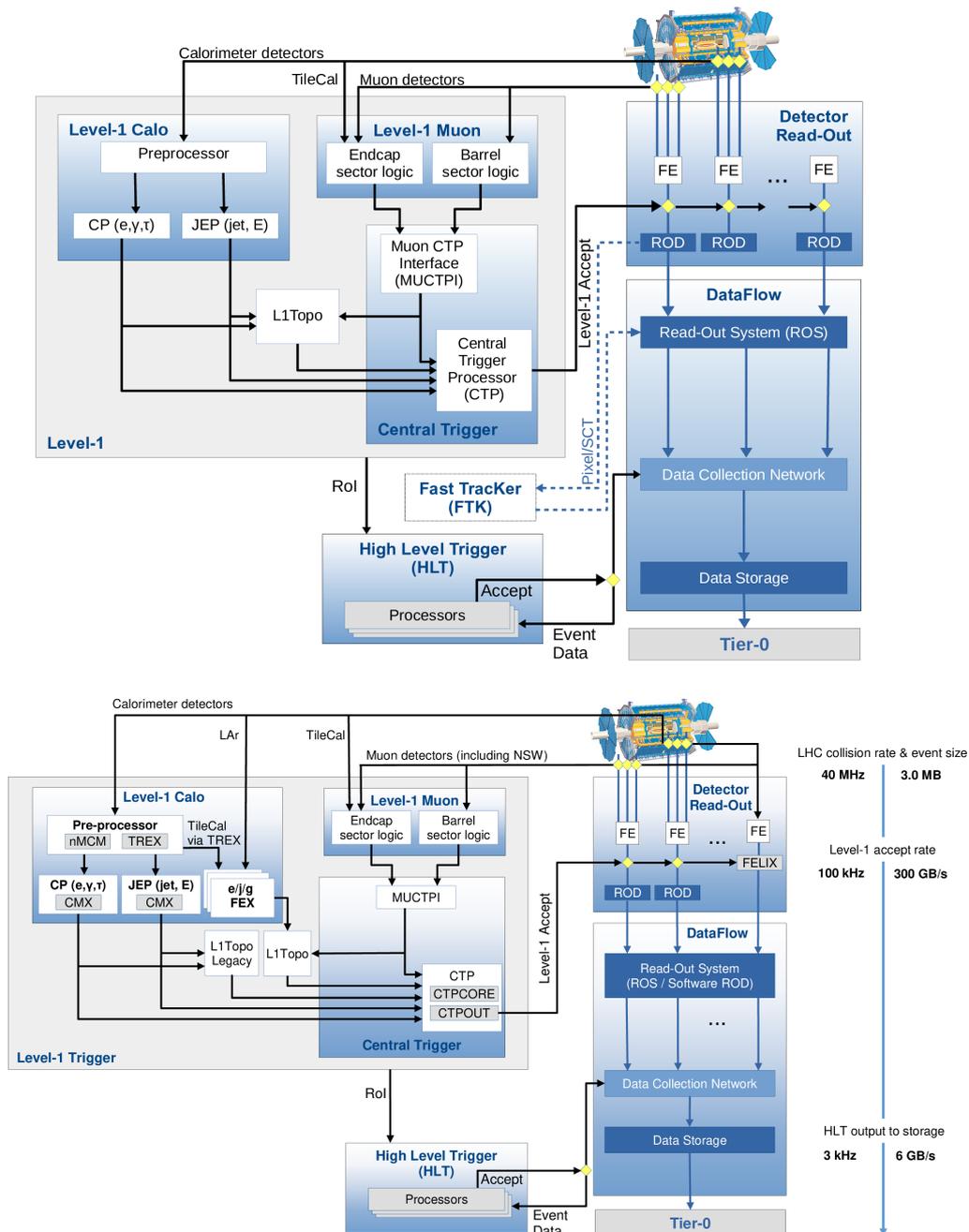


Figure 3.13: An overview of the trigger and data acquisition system of the ATLAS experiment (top) in Run 2 and (bottom) at the beginning of Run 3. The Fast Tracker (FTK) system [75] was used for commissioning at the end of Run 2 and during LS2, but it was stopped in 2019. Figure taken from [76, 59]

The FTK boards were installed during the Run 2, but the project was stopped in 2019.

The DAQ system is responsible for readout of the data from the detectors and recording them to the storage. The new Front-End LInk eXchange (FELIX) system receives readout data from detector electronics. According to the L1 trigger decisions sent to FELIX, the relevant data is passed to the readout system and buffered there. Once the HLT makes the decisions, the data of accepted events is packed, compressed, and finally transferred to offline storage.

The efficiencies of single-electron trigger and single-muon trigger in Run 2 are shown in Figure 3.14 and 3.15.

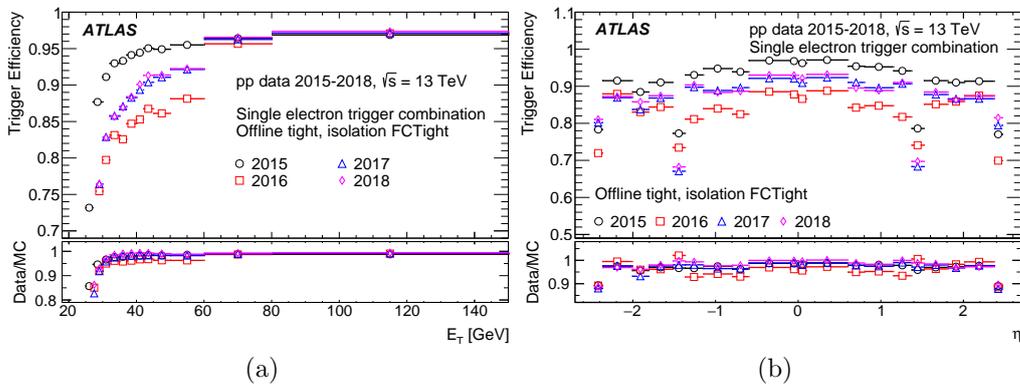


Figure 3.14: The efficiency of single-electron trigger combination as a function of the offline electron (a)  $E_T$  and (b)  $\eta$  during Run 2 [77].

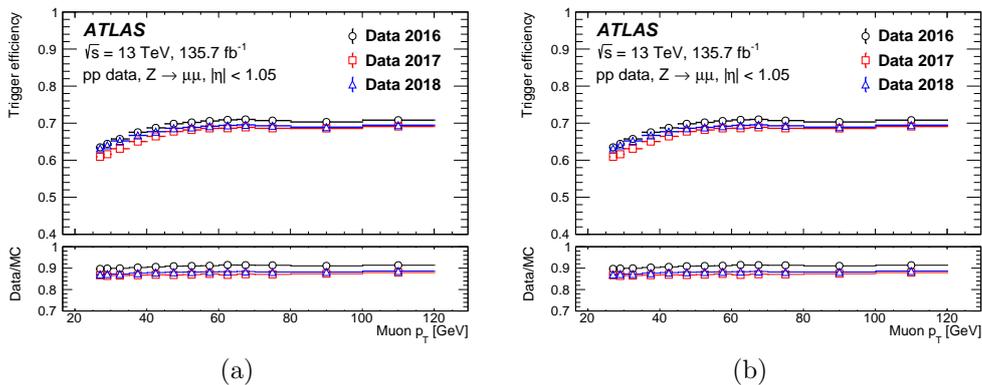


Figure 3.15: Efficiency of single-muon trigger in the (a) barrel and (b) endcaps as a function of the muon  $p_T$  during 2016-2018 [78].

### 3.3 Physics Objects Reconstruction

The ATLAS detector records vast amounts of collision data. In order to perform a physics analysis, it is important to reconstruct the physics objects from these complex data. In this section, algorithms for the reconstruction of various physics objects are discussed.

#### 3.3.1 Tracks and Interaction Vertices

The track of a particle is described by five track parameters [79]: the transverse impact parameter  $d_0$ , defined as the distance in the transverse plane between the point of the closest approach of the track to the interaction vertex and the beam line; the longitudinal impact parameter  $z_0$ , defined as the longitudinal distance along the beam line between the point where  $d_0$  is measured and the average position of the  $pp$  interactions; the azimuthal angle  $\phi$  and the polar angle  $\theta$ ; and  $q/p$ , the charge of the track divided by its momentum. These parameters can be seen in Figure 3.16. To reconstruct tracks is to find the hits created by the tracks in the detectors and to estimate these parameters.

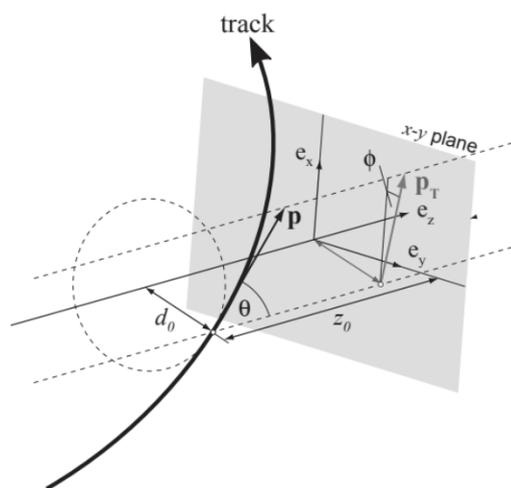


Figure 3.16: A brief illustration of track parameters [80].

The tracks of charged particles are mainly reconstructed in the ID with a series of algorithms [81]. In general, the track reconstruction has two sequences, the primary

tracking “inside-out” and the back-tracking “outside-in”. The inside-out reconstruction is optimized for particles produced from the primary  $pp$  interactions. The track candidate is formed in Pixel and SCT and then extended to TRT. The outside-in reconstruction is a complementary strategy to improve the efficiency for secondary tracks, it is used for particles produced at a larger distance from the beamline.

The quality of reconstructed tracks depends on the number of hits and holes of tracks, where the hole means the measurement on the detector is expected by the track prediction but absent. Two sets of quality criteria are defined: Loose and Tight Primary [82]. The Loose track selection is the default requirement of inside-out reconstruction, while the Tight Primary is optimized for selecting primary tracks. Figure 3.17 shows the efficiencies of track reconstruction with early Run 2 condition for two selections.

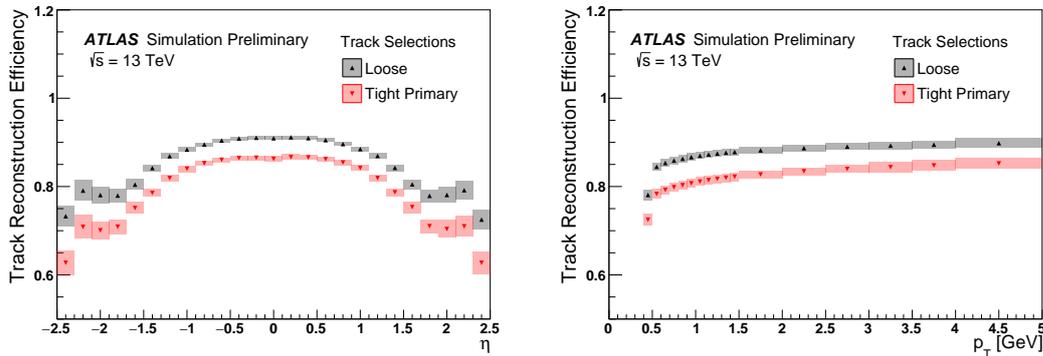


Figure 3.17: Efficiencies of track reconstruction with both Loose and Tight Primary selections as a function of (left)  $\eta$  and (right)  $p_T$ , using minimum bias events [82].

The reconstruction of interaction vertices is split into two stages [83]: vertex-finding and vertex-fitting. At the vertex-finding stage, the algorithms need reconstructed tracks as the inputs. In Run 2, the input tracks need to pass following cuts [84]:  $p_T > 400$  MeV and  $|\eta| < 2.5$ ; number of silicon hits (Pixel and SCT)  $\geq 9$  if  $|\eta| < 1.65$ ,  $\geq 11$  if  $|\eta| > 1.65$ ; IBL hits + B-Layer hits  $\geq 1$ ; shared module  $\leq 1$  (1 shared Pixel hit or 2 shared SCT hits)\*; no Pixel hole; SCT holes  $\leq 1$ .

Once the tracks are selected, the primary vertex seed is derived from the global

\*a shared hit is one that could belong to either of 2 tracks.

maximum in the distribution of the reconstructed tracks on the  $z$ -axis. The seed and nearby tracks are then used in an iterative  $\chi^2$  fit, where in each iteration the contribution of outlying tracks is down-weighted. Tracks that deviate from the vertex by more than  $7\sigma$  are removed from the vertex candidate and used to form a new vertex seed. This procedure is repeated until all tracks are associated with vertices or no additional vertex can be found. Vertices that have at least two associated tracks are valid primary vertex candidates. The efficiency of vertex reconstruction with low-pileup data is shown in Figure 3.18. The efficiency is the ratio between events with a reconstructed vertex and events with at least two reconstructed tracks.

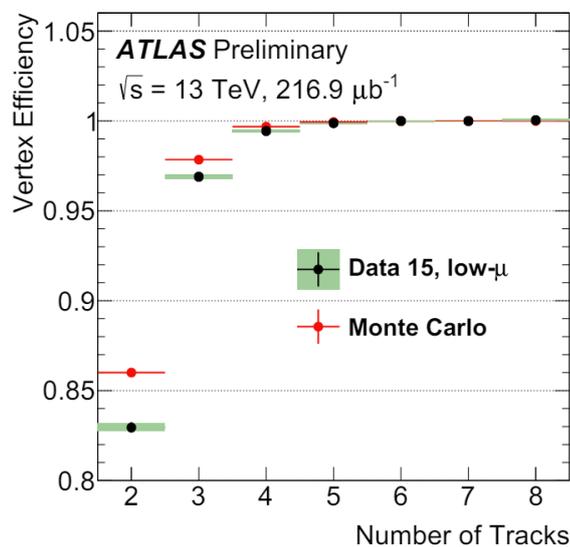


Figure 3.18: Efficiency of vertex reconstruction as a function of the number of tracks with early Run 2 data [85].

### 3.3.2 Electrons

The electron reconstruction is normally implemented by matching an energy cluster in the EM calorimeter to a reconstructed track in the ID. The cluster is reconstructed using sliding-window algorithm [86], which is based on the energy deposited in a fixed-size rectangular window in  $\eta \times \phi$ . The seeds of clusters are searched with a window size of  $3 \times 5$  middle-layer calorimeter cells. The window moves in steps of  $\Delta\eta$  and  $\Delta\phi$  and scans all components of the calorimeter. The seed is formed where the sum of transverse energy reaches a local maximum and is above a threshold.

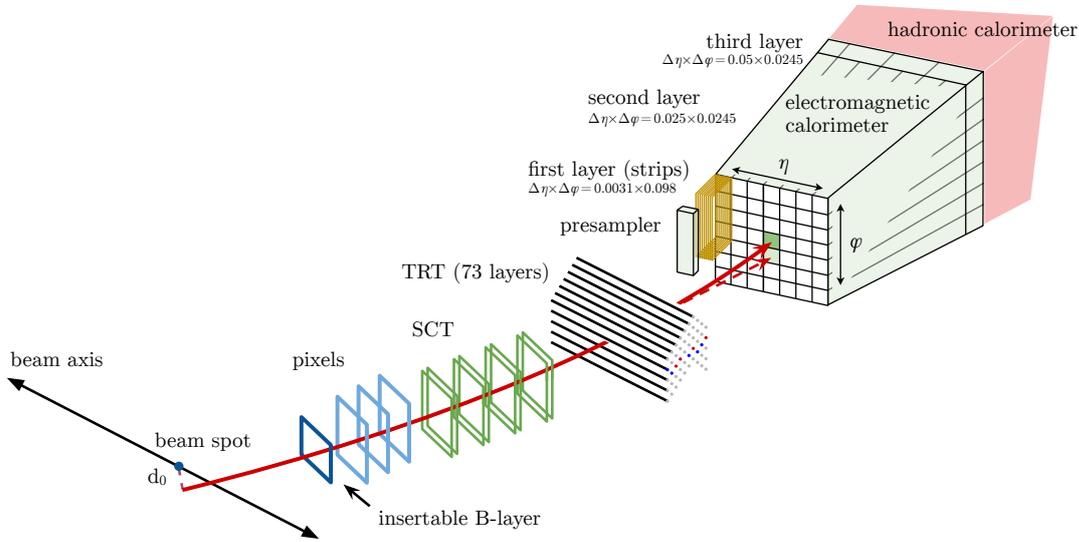


Figure 3.19: An illustration of an electron passing through the ID and entering the EM calorimeter. The red line is the trajectory of an electron, while the dashed red line is the path of a photon produced by the interaction of the electron and the material of the ID [87].

However, the electron can interact with the material and emit a photon because of bremsstrahlung when it is traveling through the ID or even the beam pipe. It is possible that the radiated photon converts into an electron-positron pair, which can also interact with the detector material. As shown in Figure 3.19, these electrons, positrons and photons can generate clusters near the primary electron and produce multiple tracks in the ID. To account for the challenge of the bremsstrahlung effect, ID tracks matched clusters are re-fit using a Gaussian sum filter (GSF) algorithm [88] for correcting the information of the electron track.

To recover energy from bremsstrahlung, an improved dynamic superclustering algorithm is used [89]. The seed of cluster is selected based on the significance (i.e. the signal to noise ratio) of the cell. A cluster is built around the seed by adding cells that pass a lower significance cut. A supercluster consists of a seed cluster and several satellite clusters near the seed. The electron is then defined as an object containing a supercluster and a matched track (or tracks) [89].

To improve the purity of selected electrons, the identification selections are applied to all reconstructed electrons. The selections are based on a likelihood identification,

with input variables depending on the information from the ID and the calorimeters [87]. Three levels of identification, Loose, Medium and Tight, are defined with different requirements on the likelihood discriminant, which are also called operating points or working points (WPs). The efficiencies of different identification levels in Run 2 data are shown in Figure 3.20. Electrons are required to be isolated. There are calorimeter-based and track-based isolation variables describing the activity in a cone of radius  $\Delta R$  around the electron candidate [87]. Different levels of isolation are defined by different cuts on isolation variables, resulting different efficiencies.

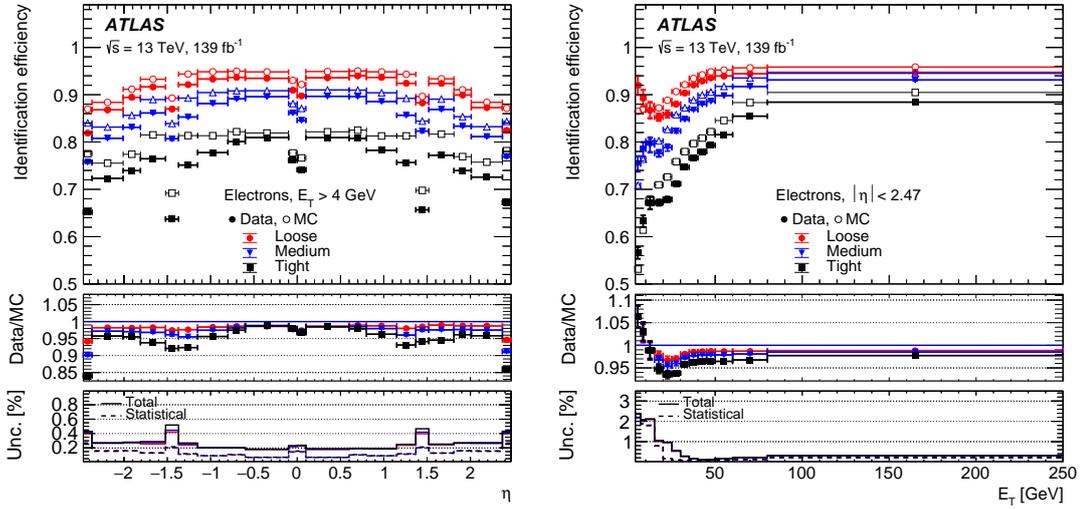


Figure 3.20: Identification efficiencies of the electrons from  $Z \rightarrow ee$  decays in both full Run 2 data and simulation as a function of  $\eta$  (left), and as a function of  $E_T$  (right) for the Loose, Medium and Tight identification levels. The top panels show the efficiencies. The middle panels show the ratio of data to MC simulation. The bottom panels show the total and statistical uncertainties [90].

### 3.3.3 Muons

The reconstruction of muons is primarily performed in the ID and MS independently, it also uses information from calorimeters for the combined reconstruction. The reconstruction of muon track in the ID is same as that of other charged particles, described in section 3.3.1. In the MS, the track of muon is reconstructed based on the hits on the chambers [91]. For the combined reconstruction, the strategies are

different for different types of muons. There are five main muon types according to the sub-detectors used in the reconstruction algorithms [73]:

- Combined (CB) muon: CB muons are identified by matching tracks in the MS to tracks in the ID. The combined track is built with a global fit using hits from both the MS and ID, as well as the energy loss in the calorimeters. Most muons are reconstructed in this type.
- Inside-out combined (IO) muon: IO muons are reconstructed in the ID first and then extrapolated to aligned hits in the MS. The combined track is fit based on the track in the ID, the energy loss in the calorimeters and the hits in the MS. Notice that this algorithm does not require independently tracking in the MS, so it has advantage for the reconstruction in the regions the MS does not fully cover and for low- $p_T$  muons which do not reach the middle station of the MS.
- Muon-spectrometer extrapolated (ME) muon: the reconstruction of ME muons only depends on the MS track, it does not associated with the ID tracks. Hence this algorithm is mainly used in the region of  $2.5 < |\eta| < 2.7$ , which is not covered by the ID but covered by the MS.
- Segment-tagged (ST) muon: a ST muon requires a reconstructed ID track and at least one matched segment in the MS. The properties of ST muon are taken from the ID track measurements. This muon type is usually used for muons hit only one layer of MS chambers due to low- $p_T$ .
- Calorimeter-tagged (CT) muon: a CT muon is identified if the ID track can be matched to an energy deposit in the calorimeters which is compatible with a minimum-ionizing particle. This type of muon has the lowest purity because of the large background contamination, but it allows the muon reconstruction with only the ID and calorimeters.

Similar to the electron objects, reconstructed muons are also categorised by several identification criteria [73]. The Loose, Medium and Tight selections are based on

the quality of muons, including the difference between measurements in the ID and MS,  $\chi^2$  of the combined track fit, and the number of precision stations (at least three hits in sub-detectors), etc. The difference between measurements in the ID and MS is defined by  $q/p$  compatibility:

$$q/p \text{ compatibility} = \frac{|q/p_{\text{ID}} - q/p_{\text{MS}}|}{\sqrt{\sigma^2(q/p_{\text{ID}}) + \sigma^2(q/p_{\text{MS}})}}, \quad (3.3)$$

where  $q/p_{\text{ID}}$  and  $q/p_{\text{MS}}$  are ID and MS measured ratios of the charge to the momentum of the muon, while the  $\sigma(q/p_{\text{ID}})$  and  $\sigma(q/p_{\text{MS}})$  are corresponding uncertainties. The Medium and Tight identification WPs only accept CB and IO muons, while the Loose selection WP includes CB, IO, CT and ST muons [73]. From Loose to Medium to Tight identification WP, the purity of muons is getting higher, but some efficiency is sacrificed. Figure 3.21 shows the efficiencies of muon identification in Run 2 with Loose, Medium and Tight selections. The muon identification also has a fourth criterion: High- $p_T$  selection, which is optimized for tracks with transverse momentum  $p_T > 100$  GeV [92, 93].

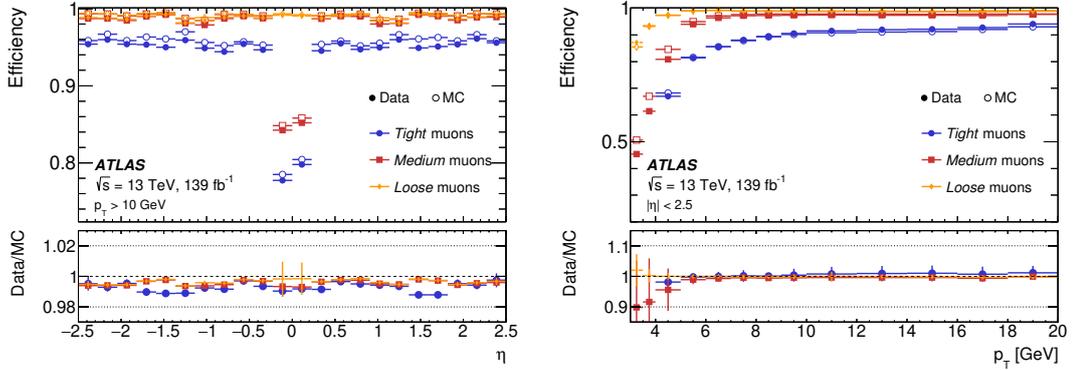


Figure 3.21: Efficiencies of muon identification for Loose, Medium and Tight selections as a function of (left)  $\eta$  and (right)  $p_T$ . Both the prediction from MC and the measurement from Run 2 data are shown [73].

### 3.3.4 Jets

Quarks and gluons cannot exist in isolation because of the color-confinement. After quarks and gluons are produced from the collisions, they are hadronized to form sprays of color-less hadrons and are detected as jets in the experiment. In the ATLAS experiment, jet candidates are reconstructed using the anti- $k_t$  jet clustering algorithm [94] with a radius parameter of  $R = 0.4$ . The inputs to the anti- $k_t$  algorithm are various. The inputs can be topological clusters [86], which contains the information of calorimeter energy deposits. The topological clusters are weighted by EM calorimeter calibration, so reconstructed jets are referred to as EMTopo jets [95]. An alternative approach is using particle-flow objects [96], which combine the information from topological clusters and the tracks of charged particles. The reconstructed jets is called PFlow jets. These jets are then calibrated a jet energy scale (JES) correction factor [95] to restore the jet energy to the particle level. The jet energy resolution (JER) describes the width of the energy distribution of reconstructed jets. The total JER uncertainty in PFlow and EMTopo jets are compared in Figure 3.22. Because the momentum resolution of the tracker is significantly better than the energy resolution of the calorimeter [52, 96], as seen in Table 3.1, PFlow jets have lower JER uncertainty in the low- $p_T$  and low- $\eta$  region.

Detector component	Designed resolution
Inner detector	$\sigma_{p_T}/p_T = 0.05\% p_T \oplus 1\% \dagger$
EM calorimeter	$\sigma_E/E = 10\%/\sqrt{E} \oplus 0.7\%$
Hadronic calorimeter	$\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$ (barrel and end-cap) $\sigma_E/E = 100\%/\sqrt{E} \oplus 10\%$ (forward)
Muon spectrometer	$\sigma_{p_T}/p_T = 10\%$ at $p_T = 1$ TeV

Table 3.1: Transverse momentum and energy resolutions of the ATLAS detector by design [52]. The units for  $E$  and  $p_T$  are in GeV.

Jets can also be reconstructed from particles not produced in the hard-scatter interaction because of the pile-up [96], where the hard-scatter is defined as the vertex with the largest sum of  $p_T^2$  of tracks associated with it in an event [97]. The jet-

$\dagger \oplus$  represents the quadrature sum.  $a \oplus b = \sqrt{a^2 + b^2}$

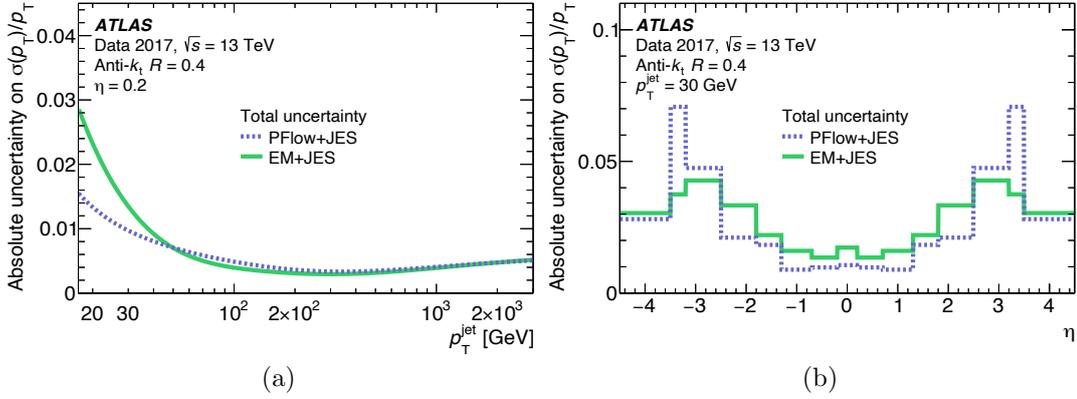


Figure 3.22: The JER uncertainty for PFlow and EMTopo jets (a) as a function of jet  $p_T$ , and (b) as a function of  $\eta$  [95].

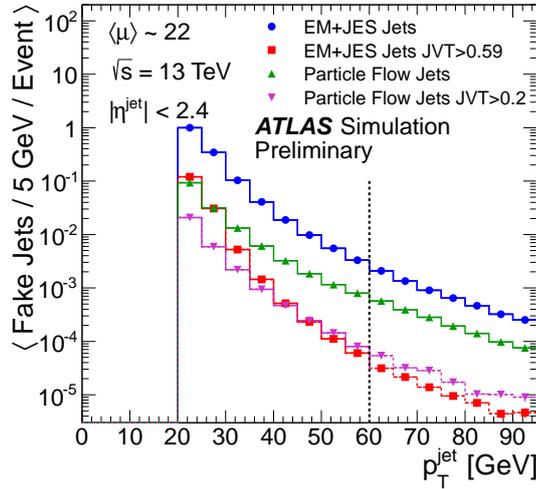


Figure 3.23: The number of reconstructed fake jets as a function of  $p_T^{\text{jet}}$  before and after JVT requirements. Both EMTopo and PFlow jets are shown [99].

vertex-tagger (JVT) discriminant is used to tag and suppress these fake jets created by the pileup interaction [98]. Figure 3.23 shows the number of reconstructed fake jets before and after JVT requirements.

To study the substructure of a jet, the ghost-association technique [100] is used, which provides a robust matching of the tracks to the calorimeter subjet. In this method, the  $p_T$  of each track in the event is set to be infinitesimal, while only its direction is retained. After rerun the jet finding algorithm, tracks that are clustered in the jet are considered as ghost-associated. The ghost-associated track can be

identified by the subjet it is clustered with [101]. Therefore the substructure of a jet can be learned through those tracks ghost-associated to the jet.

### 3.3.5 Missing Transverse Energy

Missing transverse energy (MET,  $E_T^{\text{miss}}$ ) is also referred to missing transverse momentum ( $p_T^{\text{miss}}$ ). It is the transverse momentum not recorded by the ATLAS detector. Its non-zero value may indicate the production of undetected particles, such as neutrinos or new BSM particles escaping detection [97]. The reconstruction of  $E_T^{\text{miss}}$  in ATLAS is contributed from two aspects [97]. The first aspect is hard objects, including fully reconstructed and calibrated electrons, photons,  $\tau$ -leptons, muons and jets. The second is soft signals, currently consisting of reconstructed charged tracks that are associated with the hard-scatter vertex but not with a hard object.

As the performance of MET reconstruction is hugely affected by the measurements of jets, MET selection working points are based on the requirements on jets [102]. There are four WPs supported for PFlow jets: Loose, Tight, Tighter and Tenacious, with different  $p_T^{\text{jet}}$  thresholds and JVT requirements [98]. Figure 3.24 show the modelling of  $p_T^{\text{miss}}$  distribution in MC simulation and data with PFlow jets for Loose and Tighter selection WPs.

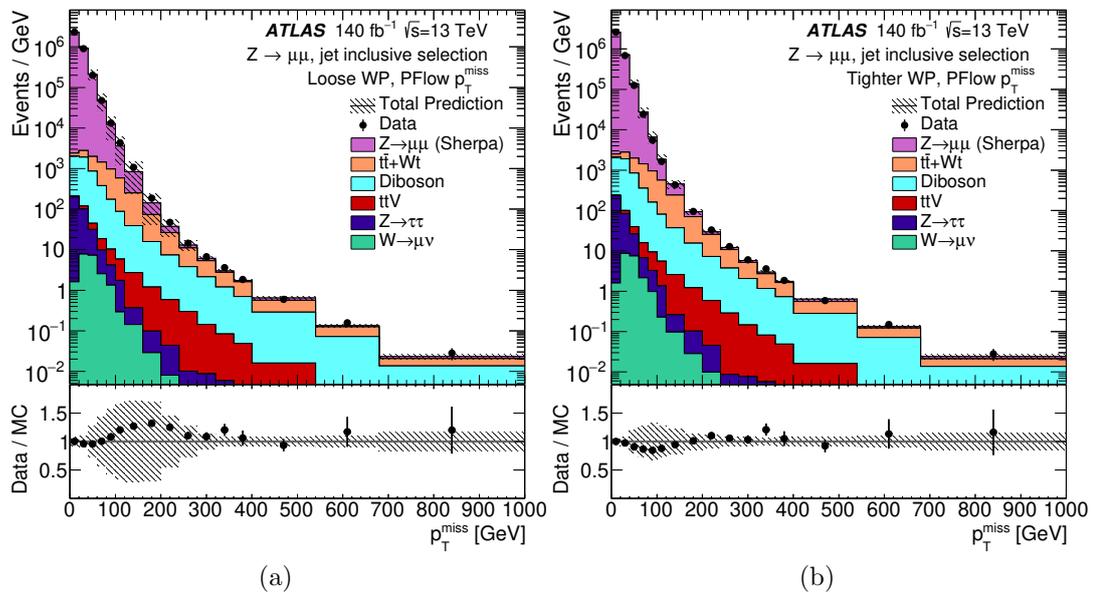


Figure 3.24: Distributions of  $p_T^{\text{miss}}$  in MC simulation and full Run 2 data with PFlow jets for (a) Loose and (b) Tighter working points [102].

## CHAPTER 4

---

### Noise Analysis of the Upgraded ATLAS L1Calo Trigger

---

As the LHC was being upgraded during the LS2, the Phase-I Upgrade of ATLAS TDAQ system has been undertaken to manage increasing LHC luminosity and pile-up rates in Run 3. The upgraded trigger system is designed to maintain the trigger efficiency as the LHC luminosity is increased beyond that for which ATLAS was originally designed. During the Phase-I Upgrade, several new components of the trigger system were added. To make sure the new system working properly during the Run 3 data-taking, it is necessary to do the noise analysis.

In this chapter, the ATLAS Level-1 Calorimeter Trigger (L1Calo) and its inputs are presented in Section 4.1 and 4.1.1. Section 4.2 shows the noise analysis of trigger towers, while Section 4.3 discusses the noise-level of supercells. Based on the noise analysis of supercells, the noise thresholds of eFEX, one of new components of L1Calo in Run 3, are adjusted.

## 4.1 L1Calo in LHC Run 3

The L1Calo processes signals from the electromagnetic and hadronic calorimeters in real-time and produces trigger signals to the Central Trigger Processor (CTP). The system identifies electrons, photons, taus, jets and calculates missing transverse energy [74].

The components of L1Calo in different LHC operating periods are shown in Figure 4.1. During Run 2, the L1Calo system sampled and digitalized signals in the Pre-Processor Modules (PPMs), and then the digital data were sent to the Cluster Processor (CP) and Jet Energy Processor (JEP). The L1Calo in Run 3 has three new feature extractors: the electron feature extractor (eFEX), the jet feature extractor (jFEX) and the global feature extractor (gFEX). These processors are designed to identify different physics objects. The Fibre Optics Exchange (FOX) and Topo-FOX ensure the high-speed data transmission between calorimeter inputs and the trigger processors [103]. During the beginning of Run 3, the components that were operating in the Run 1 and Run 2, which is also called the ‘legacy’ system, are running in parallel with the Phase-I upgrade system.

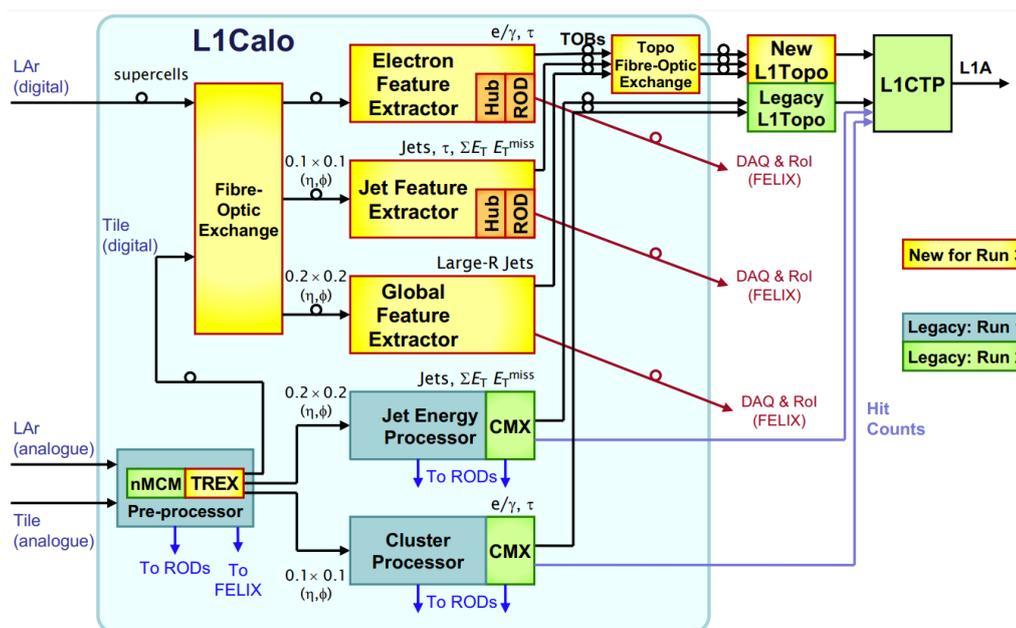


Figure 4.1: An overview of L1Calo in Run 3 [104].

### 4.1.1 Trigger Electronics

The trigger system receives in real time reduced granularity information from ATLAS calorimeters. An upgrade for electronics is provided during the Phase-I Upgrade of ATLAS [105]. The legacy system and new systems in Run 3 can be seen in Figure 4.2. The inputs to the legacy system are called Trigger Towers, the inputs to the Phase 1 system are called Supercells. Trigger Towers signals are analogue, while supercells signals are digitized.

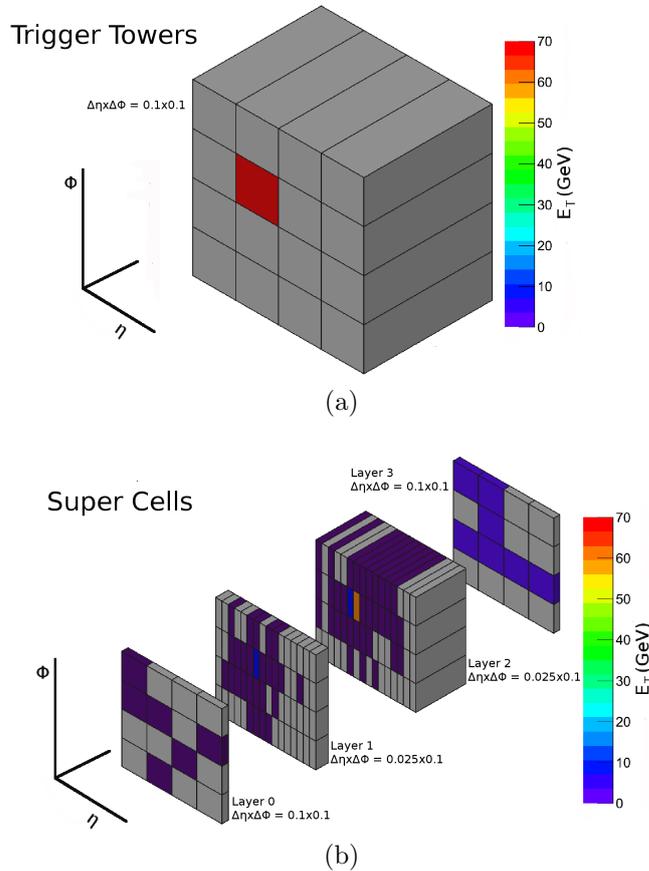


Figure 4.2: (a) The trigger electronics “trigger towers” of the legacy system and (b) the trigger electronics “supercells” of the upgraded system [105].

The Trigger Tower carries information about the sum of the deposited energy across the layers of the calorimeters, while keeping EM and HAD energies separate [105]. This summation is happening in calorimeter front end electronics, mounted on the detector. There are 7168 trigger towers in total, split between electromagnetic and hadronic layers of the calorimeter. The granularity of each tower is  $\Delta\eta \times \Delta\phi =$

Layer	Elementary Cell $\Delta\eta \times \Delta\phi$	Trigger Tower $\Delta\eta \times \Delta\phi$	Super Cell $\Delta\eta \times \Delta\phi$
0. Pre-sampler	$0.025 \times 0.1$	$0.1 \times 0.1$	$0.1 \times 0.1$
1. Front layer	$0.003125 \times 0.1$		$0.025 \times 0.1$
2. Middle layer	$0.025 \times 0.025$		$0.025 \times 0.1$
3. Back layer	$0.05 \times 0.025$		$0.1 \times 0.1$

Table 4.1: Comparison of the Trigger Tower, the Super Cell granularity in the EM barrel calorimeter in terms of the granularity [105].

$0.1 \times 0.1$  in the central region, where  $\phi$  is the azimuthal angle in the transverse plane, and  $\eta$  is the pseudorapidity. In the forward region, the granularity is coarser, which is up to  $\Delta\eta \times \Delta\phi = 0.4 \times 0.4$ .

The supercells have 4 layers: layer 0-3, corresponding the pre-sampler, front, middle and back layer of the EM calorimeter. Compared to the trigger towers, the supercells have much finer granularity. Table 4.1 shows the difference between these two trigger inputs in terms of granularity and layers. The 2<sup>nd</sup> and 3<sup>rd</sup> layer of supercells have four times finer granularity than trigger towers. Therefore, in the Run 3, there are 10 supercells in the area of one barrel trigger tower ( $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ ). In total, there are 34048 supercells in the new readout system.

#### 4.1.2 eFEX

The electron Feature Extractor (eFEX) is one of the new feature identification systems added to L1Calo during the Phase-I upgrade. A total of 24 eFEX modules cover the area of  $-2.5 < \eta < 2.5$  and  $0 < \phi < 2\pi$ , and read higher granularity data from supercells. The eFEX is used to identify electrons, photons and taus with improved algorithms processed in a bank of 4 processor FPGAs, which can be seen in the Figure 4.3.



Figure 4.3: A photo of the eFEX module. There are four processor FPGAs under the copper heatsinks. Photo taken from Ref. [106].

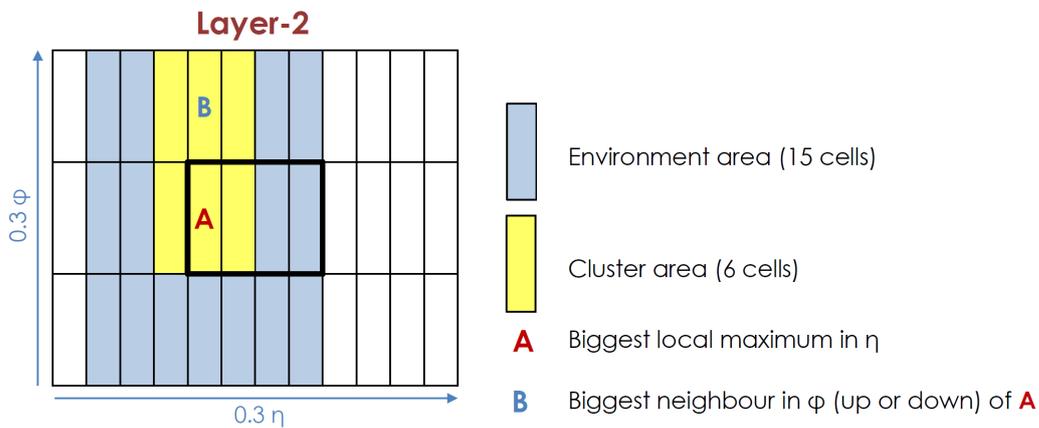


Figure 4.4: The seed, the cluster and the environment area built by the electron algorithms of the eFEX. Each cell stands for a supercell in layer 2. The supercell *A* is the seed located by the algorithms. The yellow area is the cluster area and the blue is the environment area. The square in bold means the area of a tower ( $0.1 \times 0.1$ ). Plot taken from Ref. [107].

### Electron algorithms

The  $e/\gamma$  trigger has two key functions, the first is to form a cluster containing the energy of an electromagnetic shower, the second is to reject jets [74].

The electron algorithms of the eFEX run in a  $\Delta\eta \times \Delta\phi = 0.3 \times 0.3$  window with steps of  $0.1 \times 0.1$ . The algorithms is designed to find the ‘seed’, which is the supercell with the highest energy in the second layer of the window. The cluster area of  $3 \times 2$  supercells is built around the seed. The cluster contains both neighbours of the seed in the  $\eta$  direction, and of the largest neighbour in the  $\phi$  direction. The environment

area with 15 cells is defined around the cluster area, as shown in the Figure 4.4. To discriminate electrons from jets, the eFEX calculates three following isolation variables:

- Isolation condition  $R_\eta$ , defined as

$$R_\eta = \frac{E_{clu}}{E_{clu} + E_{env}}, \quad (4.1)$$

where  $E_{clu}$  is the transverse energy measured in the cluster area of layer 2,  $E_{env}$  is the transverse energy measured in the environment area of layer 2.  $R_\eta$  has much narrower distribution close to 1 for electrons than jets, a good electron should have  $R_\eta > \text{threshold}$ .

- Cluster-width condition  $W_S$ .  $W_S$  describes the spread of shower, defined by

$$W_S^2 = \frac{\sum E_i (i - i_{max})^2}{\sum E_i}, \quad (4.2)$$

where  $i$  loops over supercells of the cluster area in layer 1. The electromagnetic shower is narrower than jet shower, so good electrons should have  $W_S < \text{threshold}$ .

- Hadronic condition  $R_{had}$ , defined by

$$R_{had} = \frac{E_{had}}{E_{had} + E_{EM}}, \quad (4.3)$$

where  $E_{had}$  stands for the energy deposited in the hadronic towers (tiles for instance), and  $E_{EM}$  is the total energy deposited in the electromagnetic layers. Good electrons should have  $R_{had} < \text{threshold}$ , because jets deposit more energy in hadronic layers than electrons.

Each variables has three thresholds defining the loose, medium and tight isolation. For each cluster, the eFEX calculates these three isolation variables and compares them to the thresholds. Figure 4.5 compares the electron trigger efficiencies between

the legacy system and the Phase-I system, where the electron identification in Phase-I system is made by the eFEX.

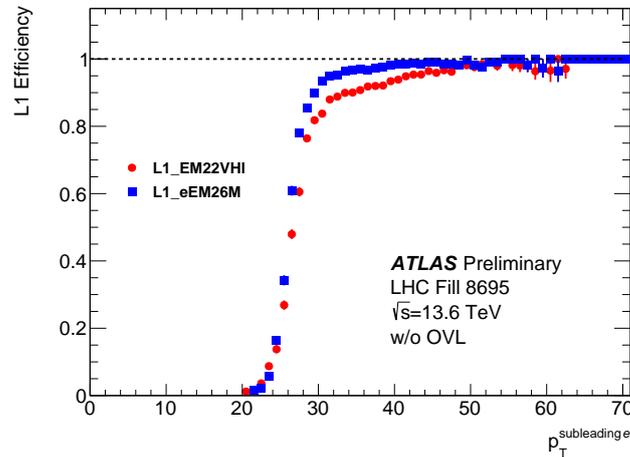


Figure 4.5: The efficiencies of L1Calo single electron trigger for (red) the legacy system and (blue) the Phase-I system in Run 3 as functions of electron  $p_T$ . The Phase-I electron identification is provided by the eFEX. Plot taken from Ref. [108].

## 4.2 Searching for Problematic Trigger Towers

As mentioned in Section 4.1.1, the electromagnetic L1 triggering was performed by the legacy system using trigger tower energies as an input. The legacy system was running in the beginning of Run 3 and it was used as the main triggering system for first data taking. Part of the upgraded system will also route information from Tile Calorimeter to the Phase-I system. Therefore, it is necessary to check if the trigger towers work as expected.

The analogue data from trigger towers was sent to the L1Calo legacy system. These signals are sampled in the pre-processor and digitized by the analogue-to-digital conversion (ADC) [109]. The signal-to-noise ratio is then improved using a digital filter. ADC pulse is converted into transverse energy based on Look-Up Tables (LUT), which perform pedestal subtraction, noise cut and find calibration. There are two LUTs working in parallel, one for Cluster Processor (CP) and one for Jet Energy Processor (JEP). For this test, there is a linear filter working on ADC values,

in which the relation between the ADC pulse and LUT values can be described as:

$$LUT_{CP} = \frac{ADC_{peak} - 32}{2}, \quad (4.4)$$

$$LUT_{JEP} = \frac{ADC_{peak} - 32}{4}, \quad (4.5)$$

where  $ADC_{peak}$  indicates the largest ADC count in a pulse. The number 32 in the equations is an average pedestal, but in reality pedestals for individual trigger towers vary and these tower-dependent pedestals are used. The size of least significant bit (LSB) is 250 MeV, which means the resolutions of CP and JEP LUT values are 0.5 GeV and 1 GeV respectively. The LUT values are stored in 8 bits, so the CP LUT values cover 0-127.5 GeV and JEP LUT values cover 0-255 GeV.

To search for problematic trigger towers from offline, it is good to look at cosmic data, which have lots of events with zero energy and only few with energy. The background noise is everywhere so it is normal that trigger towers output some low none-zero energy. But a problematic trigger tower may output higher energy multiple times. So the strategy is searching for trigger towers that produce many high amplitude ADC pulses. First, a cut  $ADC_{peak} > 52$  is applied to get rid of the low energy noise. Then we count the number of high energy pulses from each trigger tower. Those trigger towers producing high energy pulses more than 10 times will be marked as ‘problematic’. Table 4.2 shows three problematic trigger towers and their location collected by this method. Those problematic towers had been already added to the list of problematic trigger towers of LAr. Some were fixed before the Run 3 and some are masked.

Problematic Trigger Towers				
ID in Decimal	ID in Hexadecimal	$\eta$	$\phi$	Location
51842816	03170f00	-1.9500	2.9943	EMEC
85722624	051c0600	-3.1500	-1.8653	HEC
119475202	071f0c02	-0.23	-0.34	Tile

Table 4.2: Problematic trigger towers and their location.

### 4.3 Adjustment of Noise Thresholds of eFEX

The eFEX receives digitized data from supercells. Before the data are fed into the algorithms of eFEX, noise cuts are applied. At the beginning of Run 3, the eFEX has one noise threshold per layer: 800 MeV for pre-sampler, 600 MeV for other layers. However, the noise levels in different regions and layers are not the same. Therefore, to have better performance in Run 3, eFEX needs noise thresholds based on  $\eta$  and layers.

To adjust the noise thresholds properly, it is necessary to understand the input of eFEX. Therefore, a study of supercell data is taken. This is the first time that looking at supercell data from the L1Calo side.

#### 4.3.1 $E_T$ calculation and the timing cut

Before being sent to the eFEX, the supercell data is processed in the the LAr Trigger Processing Mezzanines (LATOME) boards, which is a part of LAr Digital Processing System (LDPS), the new back-end component of LAr readout system as shown in Figure 4.6. The LDPS receives ADC samples from LAr Trigger Digitizer Boards (LTDB), where the high-granularity information is digitized (bottom-left of Figure 4.6). The function of LATOME is to reconstruct the supercell energy and to identify bunch crossing.

The transverse energy of supercell and the pulse timing are reconstructed from ADC samples using the Optimal Filtering (OF) technique [110] with following equations:

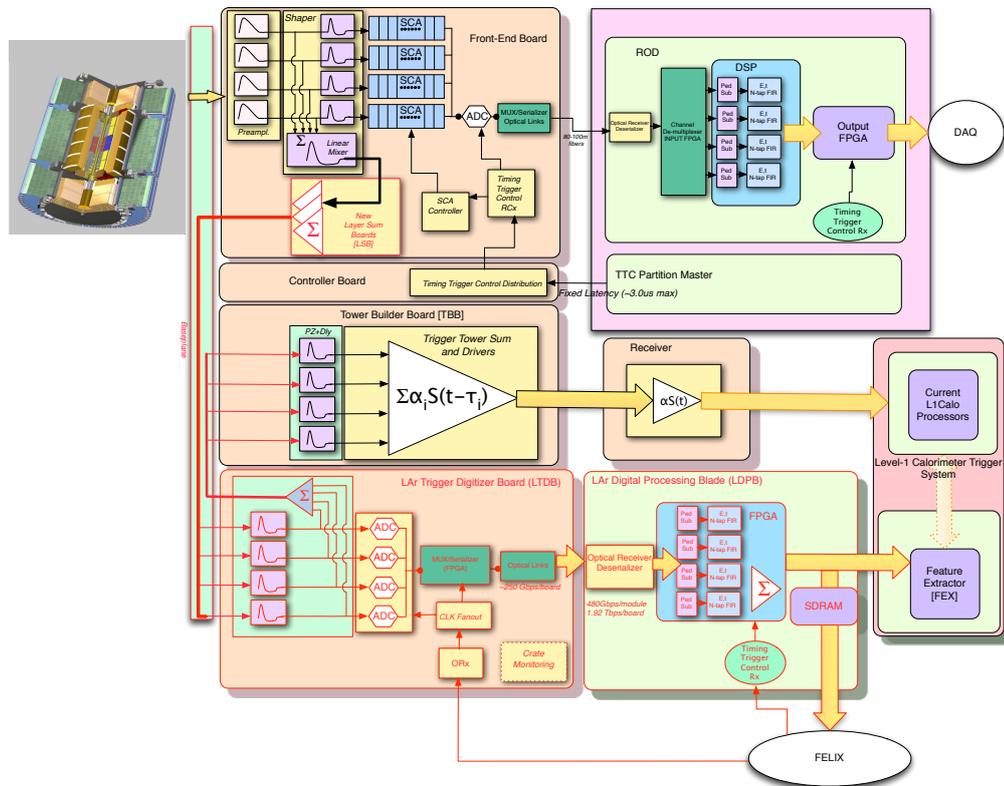


Figure 4.6: Overview of the LAr trigger readout system of ATLAS experiment after Phase-I upgrade. The components in red are the new subsystems installed during Phase-I upgrade. The new back-end, the LAr Digital Processing System (LDPS), is at the bottom-right of the figure. The LAr Trigger Digitizer Boards (LTDB) is at the bottom-left. Plot taken from Ref. [105].

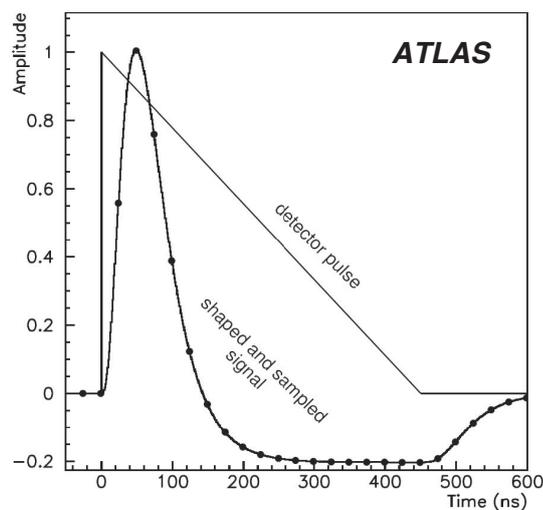


Figure 4.7: A pulse in the detector and digitized samples. The dots are ADC samples. The original amplitude and peak time can be restored by the equations (4.6) and (4.7). Plot taken from Ref. [105].

$$E_T = \sum_{i=1}^4 a_i(S_i - P), \quad (4.6)$$

$$E_T \cdot \tau = \sum_{i=1}^4 b_i(S_i - P), \quad (4.7)$$

where  $a_i$ ,  $b_i$  are Optimal Filter coefficients (OFC),  $P$  is pedestal,  $S_i$  is ADC sample amplitude,  $E_T$  is the transverse energy deposited in the supercells and  $\tau$  is relative phase, describing the time between pulse peak and current bunch crossing, as shown in Figure 4.7. By this method, the original amplitude and the real-time are estimated, and the contribution of electronics and pile-up noise is minimized [111].

However, the calculation above is not directly used in LATOME. Due to the limitation on the resource, LATOME uses integer coefficients instead of double or float precision. The pedestals and filter coefficients are bit shifted and saved as integers. Therefore, in reality, the supercell energy is reconstructed as following:

$$E_T = \left[ \sum_{i=1}^4 a_i(S_i \ll 3 - P) \right] \gg 8, \quad (4.8)$$

where " $\ll n$ " and " $\gg n$ " indicate left and right shift by  $n$  bits. ADC samples are first left shifted by three bits to get same scale as pedestals, the overall sum is right shifted by 8 bits. As the result, the transverse energy reconstructed by LATOME is stored in steps of 12.5 MeV. The  $E_T \cdot \tau$  calculation in practice is similar. Instead of right shifting the overall sum by 8 bits,  $E_T \cdot \tau$  needs a 6-bit right shift:

$$E_T \cdot \tau = \left[ \sum_{i=1}^4 b_i(S_i \ll 3 - P) \right] \gg 6, \quad (4.9)$$

Large value of the pulse phase indicates peak may come from a previous bunch crossing. In order to make sure the signal comes from current bunch crossing, a timing cut is applied, which actually is also called the  $\tau$  selection: for the supercell  $E_T > 10$  GeV, the requirement is  $-8 \text{ ns} < \tau < 16 \text{ ns}$ ; for the supercell  $E_T < 10$  GeV, the requirement is  $-8 \text{ ns} < \tau < 8 \text{ ns}$ . During Run3, if the supercell  $E_T$  pass the  $\tau$

selection, the supercell  $E_T$  will be sent to FEXs, otherwise zero energies will be sent.

To check the reliability of the supercells, it is good to check if the supercells have similar output as the legacy system for the same event. There are three energies that can be used: a) Supercell energies, these are energies read out from LATOME for each event. The copy of energies can be read out from the eFEX input, it is called pre-scaled readout; b) Trigger Tower energies, these are energies used by the legacy system and can be read out from pre-processor system; c) Calorimeter cell energies, these are energies that are used for analysis and are the most precise information with best granularity and are read out from LAr calorimeter front end electronics. The comparison of these three energies is shown in Figures 4.8.

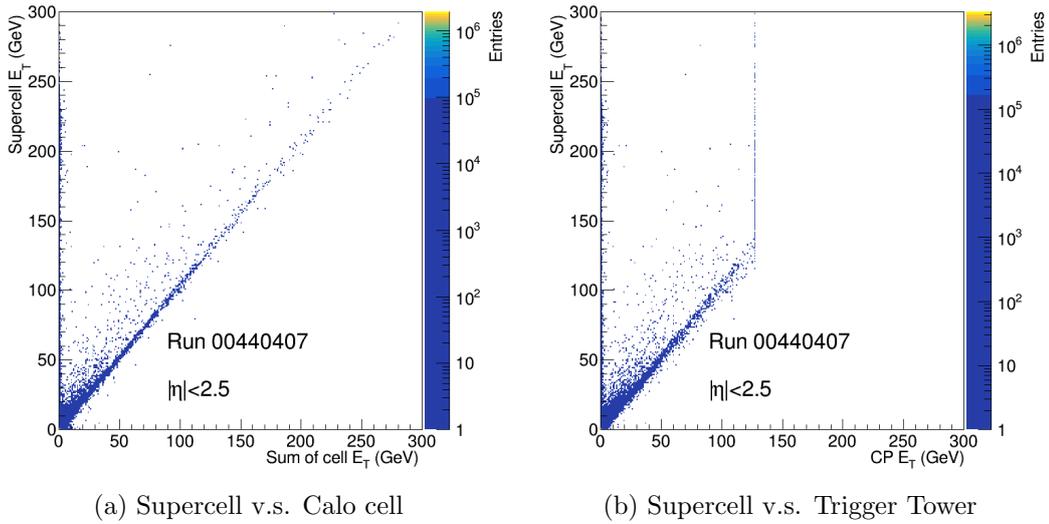


Figure 4.8: The Comparison between reconstructed energy of supercells and the legacy system in the region  $|\eta| < 2.5$ , with the data from ATLAS Run 440407. (a) The comparison between supercells and cells. The supercell  $E_T$  is reconstructed by LATOME, the Calo cell energy is re-computed offline by summing the calorimeter cells' energies in an area of  $0.1 \times 0.1$ . (b) The comparison between supercell  $E_T$  and the  $E_T$  used by the legacy system. The Trigger Tower  $E_T$  has a maximum at 127.5 GeV because it is stored in 8 bits with a step of 500 MeV.

Most supercell output are consistent with the legacy system and the deposited energy in the calorimeter cells. There are some bad supercell output, shown at the y-axis of Figures 4.8, have output of tens of or hundreds of GeV while there is only few

GeV energy deposited in the calorimeter. Most of bad supercells are masked before sending energy to the FEXes.

### 4.3.2 $\eta$ -based Noise Thresholds

Besides of reconstructed transverse energy of supercells, we also need to know the locations of supercells to derive the  $\eta$ -based noise thresholds of the eFEX for the data taking. According to the location information stored in the supercell data, the supercells are separated into four regions according to the LAr detectors: Electromagnetic Barrel (EMB), Electromagnetic End-Cap (EMEC), Hadronic End-Cap (HEC) and Forward Calorimeters (FCAL). The Figures 4.9 show the location and layers of supercells in these different regions. For the EMB and EMEC, the layer 0-3 indicate pre-sampler, front, middle and back layer respectively.

The thresholds require no more than a certain percentage of supercells in events pass the thresholds, and that percentage is called the occupancy. To derive the  $\eta$ -based noise thresholds, the following method is used:

1. Obtain the  $E_T$  distribution of supercells in each  $|\Delta\eta| = 0.1$  bin and each layer. Figure 4.10 shows an example of supercell  $E_T$  distribution in the  $0 < \eta < 0.1$  region of EMB pre-sampler. To minimize effect of faulty supercells and of energy deposits from high- $p_T$  collision, only  $-10 \text{ GeV} < E_T < 10 \text{ GeV}$  events are considered. The peak at zero in the Figure 4.10 indicates that most of supercells output zero energy or do not pass the  $\tau$  selection, and there is a few GeV tail near the zero peak caused by the noise.
2. Calculate the cut on  $E_T$  for each bin so that certain percentage of events in each bin pass the cut. First, obtain the total number of supercell entries sending energy between  $-10$  and  $10 \text{ GeV}$  to FEXes. Second, starting with cut on  $10 \text{ GeV}$ , calculate the ratio of number of entries having  $E_T$  greater than the cut to the total number. Then lower the cut in steps of  $12.5 \text{ MeV}$  and calculate the ratio, until the ratio hits the expected occupancy.

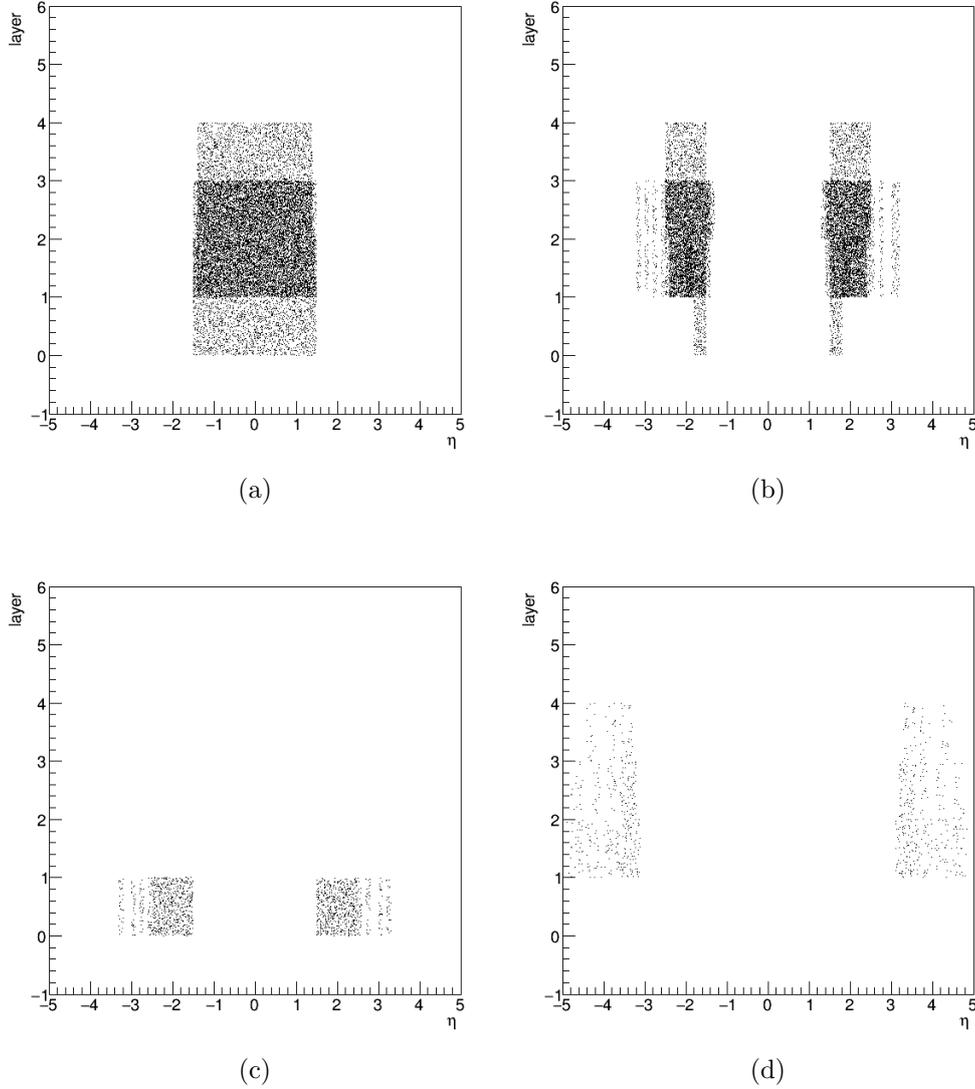


Figure 4.9: The  $\eta$  location and layers of supercells in the different part of the calorimeters: (a) Electromagnetic Barrel (EMB), (b) Electromagnetic End-Cap (EMEC), (c) Hadronic End-Cap (HEC), (d) Forward Calorimeters (FCAL). Each dot in the plots above represents a supercell. More dense dots mean that supercells have finer granularity in that region.

With this method, we can obtain a set of  $E_T$  cuts based on  $\eta$  and meeting the requirement of occupancy. The  $E_T$  cuts with 0.5% and 1% occupancy are can be seen in Figure 4.11. These cuts are derived by using Run 440407 Stable beam data, which is an ATLAS run while LHC was delivering 6.8 TeV beams in November 2022. There is few asymmetry between A side ( $\eta > 0$ ) and C side ( $\eta < 0$ ), believed to be

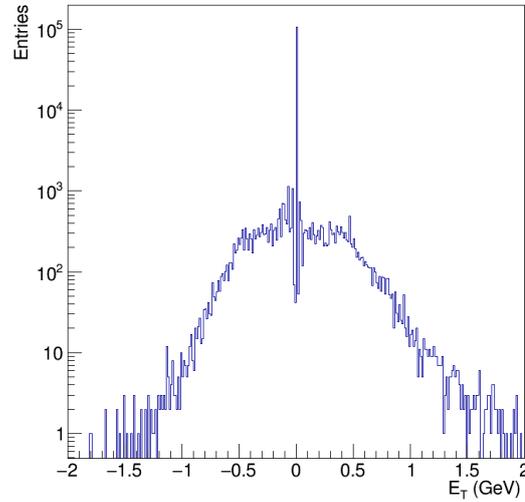
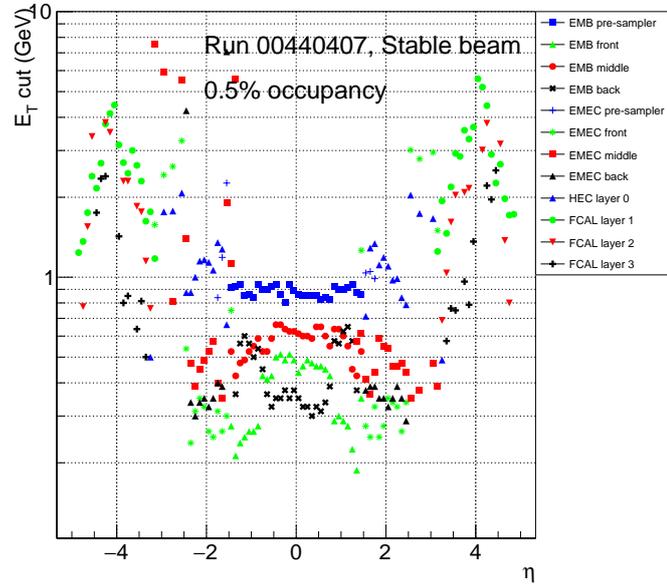


Figure 4.10: An example of supercell  $E_T$  distribution, each bin is 12.5 MeV wide. The  $E_T$  is output from the supercells in the bin  $0 < \eta < 0.1$  of EMB pre-sampler. The data is collected from a stable beam lumi block of ATLAS Run 440407. The mean number of interactions per bunch crossing of this lumi block  $\langle \mu \rangle = 53.7$

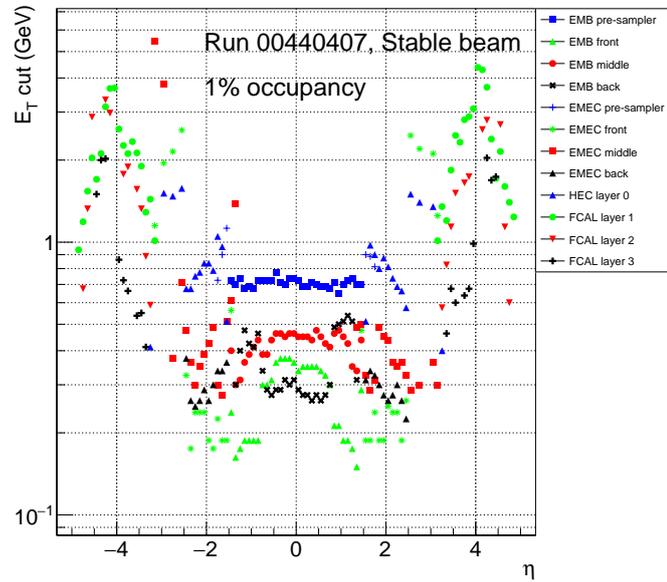
caused by the influence of noisy cells.

The purpose is to derive noise thresholds for eFEX, so the study focuses the region of  $|\eta| < 2.5$ . There are 50 bins in the region  $|\eta| < 2.5$ . To simplify the analysis and to avoid the influence of bad supercells, these 50 bins are symmetrized. For each  $\eta$  bin, we pick the lower noise cut between  $\eta$  and  $-\eta$  bin as the threshold. The thresholds for different bins and layers and the comparison with old  $\eta$ -constant thresholds can be seen in the Figure 4.12.

This set of noise thresholds are the first  $\eta$ -based noise threshold of eFEX. These thresholds were loaded into the database and used during subsequent data taking.



(a)



(b)

Figure 4.11: The noise cuts of supercells of each  $|\Delta\eta| = 0.1$  bin and layer in different parts of LAr calorimeter with (a) 0.5% and (b) 1% occupancy

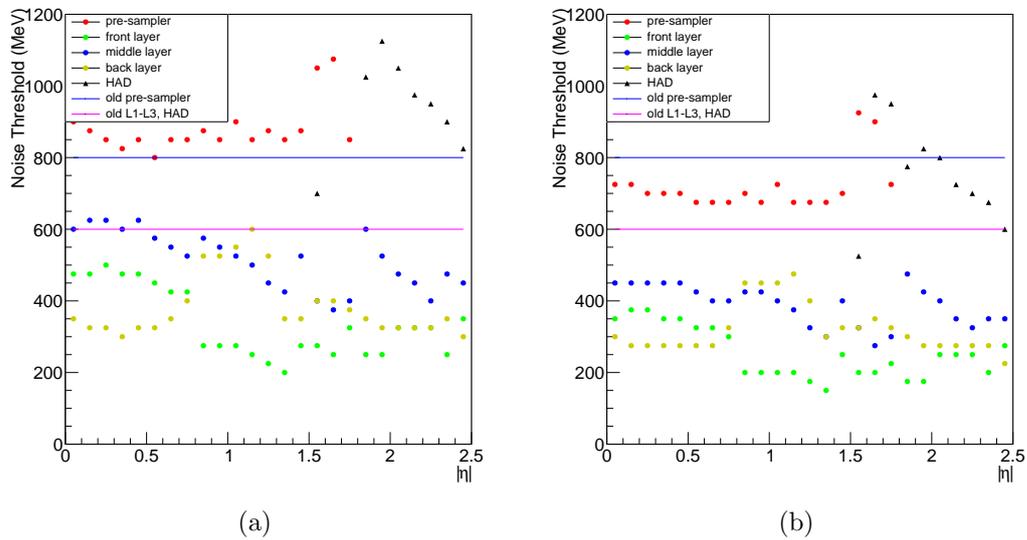


Figure 4.12: The comparison of  $\eta$ -based thresholds and  $\eta$ -constant thresholds of eFEX with (a) 0.5% and (b) 1% occupancy. The circles are the  $\eta$ -based noise thresholds of different layers in the EM region. The black triangles are the  $\eta$ -based noise thresholds of hadronic layer. The blue straight line at 800 MeV is  $\eta$ -constant threshold of pre-sampler used for eFEX at the beginning of Run 3. The purple line at 600 MeV is the threshold of other layers at the beginning of Run 3.

---

### Elements of Machine Learning

---

“Can machines think?” Alan Turing proposed this question in 1950 [112], which has inspired exploration into artificial intelligence (AI) and machine learning (ML) for decades. ML is a subfield of AI that learns from complex data and produces desired results or predicts unseen data. The ML algorithms do not need explicit programming, and only need small human intervention [113]. A typical ML algorithm consists of three components: a decision process making predictions or classifications based on input data, an error function used to measure the discrepancy of the model from the known samples, and an optimization process adjusting the model autonomously to reduce the discrepancy. The processes are repeated until the accuracy of the model reaches a threshold.

ML is now widely used in particle physics experiments, for example the detector simulation, event triggering, and particle identification, etc [114]. Most analyses in high-energy physics (HEP), including the discovery of Higgs boson [4, 5], benefit greatly from using ML algorithms. Currently the most frequently used ML algo-

gorithms in HEP are boosted decision trees (BDT) and artificial neural networks (ANN or NN) [115]. This chapter focuses on NNs that are used for the analysis presented in Chapter 6.

## 5.1 Structure of Neural Networks

The NN is a class of algorithms inspired from the structure of the human brain. In the brain, the basic elements are neurons, which are highly connected and exchange chemical signals with each other. In the NN, the neuron is modelled as a perceptron [116]. Figure 5.1 illustrates a perceptron with  $n$  inputs  $x_1, x_2, \dots, x_n$  carrying weights  $w_1, w_2, \dots, w_n$ , a input bias  $b$ , and a output  $g(z)$ , where  $z$  is the weighted sum of the inputs and  $g$  is the activation function.

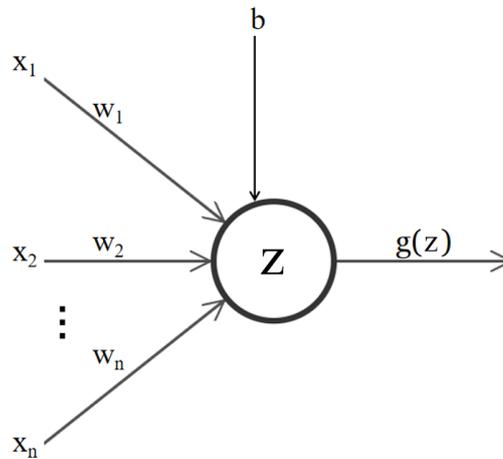


Figure 5.1: A illustration of a perceptron in the NN.

The inputs ( $x_i$ ) of a perceptron come from data or outputs of other perceptrons, each input is assigned a weight ( $w_i$ ) reflecting its importance. The bias ( $b$ ) is an additional term used to offset the result. The perceptron firstly calculates the weighted sum ( $z$ ) of these inputs:

$$z = \sum_{i=1}^n w_i x_i + b, \quad (5.1)$$

then the sum is passed through an activation function  $g$ . The activation function is used to add non-linearity to the NN, otherwise the NN would be a composition

---

of linear functions and will not be able to learn any complex task or make complex decisions. The function  $g$  transforms  $z$  to a value within a pre-defined range, and identifies if the neuron is activated. There are many activation functions for different uses, here some commonly used examples are shown:

- Binary step:

$$g(z) = \begin{cases} 0 & \text{for } z < 0 \\ 1 & \text{for } z \geq 0 \end{cases}. \quad (5.2)$$

The binary step function is a threshold-based function, can be used in the output of a NN for a binary classification problem.

- Identity/Linear:

$$g(z) = z. \quad (5.3)$$

The linear activation function does not provide any non-linearity and has no range restriction. It is mainly used in the output of a regression NN or some simple tasks.

- Sigmoid:

$$g(z) = \frac{1}{1 + e^{-z}} \quad (5.4)$$

The sigmoid function has output values between 0 and 1, and it is an S-curve shown as the red line in Figure 5.2. The output gets closer to 0 or 1, as the input value decreases or increases.

- Hyperbolic Tangent (TanH):

$$g(z) = \tanh z = \frac{e^z - e^{-z}}{e^z + e^{-z}}. \quad (5.5)$$

The TanH function is like the sigmoid function, it also has an S-shape shown as the purple line in Figure 5.2. But the output range of TanH function is from  $-1$  to  $1$ .

- Rectified Linear Unit (ReLU):

$$g(z) = \max(0, z) = \begin{cases} 0 & \text{for } z < 0 \\ z & \text{for } z \geq 0 \end{cases}. \quad (5.6)$$

The ReLU function behaves as a linear function for positive input values, and returns 0 if the input is negative, so it does not activate all the neurons at the same time.

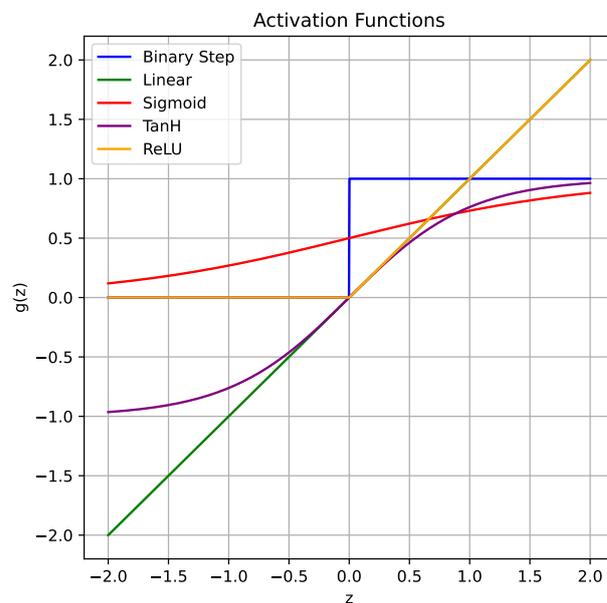


Figure 5.2: Five different activation functions: (blue) Binary step, (green) Linear, (red) Sigmoid, (purple) TanH, and (orange) ReLU.

After the calculations above, the output of a perceptron  $g(z)$  is used as the input to other perceptrons. With many perceptrons arranged in layers and connected, a NN is formed.

The structure of a basic NN can be seen in Figure 5.3. The basic form of a NN contains three layers: an input layer, an output layer, and a hidden layer in the middle, having no connection with the external world. The hidden layer is the critical part of a NN, which controls the learning process. If a NN has more than one hidden layer, then it is called a deep neural network (DNN) [118]. After years of rapid

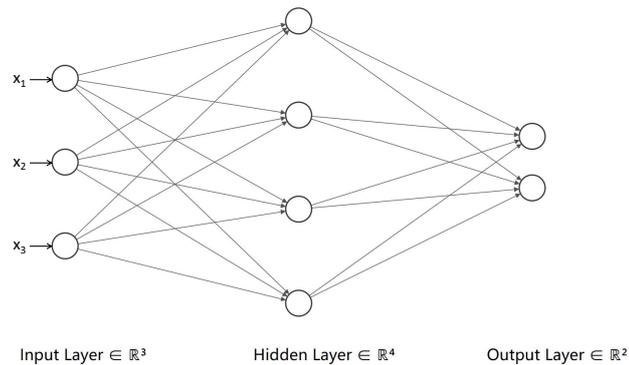


Figure 5.3: A basic form of a NN with an input layer, a hidden layer and an output layer. Plot made using the tool NN-SVG [117].

development, NNs are now incredibly complex and a great number of types of NNs have been proposed to address different problems. For example, the convolutional neural network (CNN) originally used for computer vision [119], the recurrent neural network (RNN) used for voice recognition [120], and the transformer which gave birth to ChatGPT [121, 122].

In this chapter, a simple type of NN, the feed-forward neural network (FFNN) is discussed. In a FFNN, each perceptron in one layer is connected to every perceptron in the next layer, and no connection among perceptrons in the same layer, just like the one shown in Figure 5.3.

## 5.2 Training of NNs

To train a NN simply means to obtain the set of weights and biases that maximize the accuracy of the NN or, equivalently, minimize the loss function. The two main steps of the training process are forward propagation and backpropagation.

### 5.2.1 Forward Propagation and Loss Functions

Forward propagation describes the calculation through the network that produces outputs based on inputs. Consider a NN with  $n$  input neurons, a hidden layer with  $m$  neurons, and only one output neuron. To make it simpler, bias are set to 0, and the activation functions  $g$  of the hidden and output layer neurons are taken to be the same. The neurons in the input layer receives the raw input data and passes the data to the hidden layer. For the first neuron of the hidden layer, its output can be written as

$$y_1^{[1]} = g(w_{11}^{[1]}x_1 + w_{21}^{[1]}x_2 + \dots + w_{n1}^{[1]}x_n) = g\left(\sum_{i=1}^n w_{i1}^{[1]}x_i\right), \quad (5.7)$$

the superscripts of  $w$  and  $y$  indicate that they are input weights and outputs of the neurons in the first hidden layer, this will be helpful if generalising to the NN with multiple hidden layers.  $w_{ij}$  denotes the weight of  $i$ -th input to the  $j$ -th neuron. Therefore, it is easy to write all outputs of  $j$  hidden neurons:

$$y_j^{[1]} = g\left(\sum_{i=1}^n w_{ij}^{[1]}x_i\right), \quad j = 1, 2, \dots, m. \quad (5.8)$$

These outputs from the hidden layer are then sent to the single output layer perceptron with weights  $w_{j1}^{[2]}$ , the output of this NN is written as

$$y_{NN} = g\left(\sum_{j=1}^m w_{j1}^{[2]}y_j^{[1]}\right) = g\left(\sum_{j=1}^m w_{j1}^{[2]} \cdot g\left(\sum_{i=1}^n w_{ij}^{[1]}x_i\right)\right) \quad (5.9)$$

Note that the activation function  $g$  in equation (5.9) is non-linear, so the NN is a massive non-linear function of inputs  $x_1, x_2, \dots, x_n$ . The weights  $(w_{ij}^{[1]}, w_{ij}^{[2]})$  are the parameters of this function. The number of weights depends on the number of neurons in each layer, for example, in equation (5.9) there are  $n \times m$   $w^{[1]}$ s and  $m \times 1$   $w^{[2]}$ s. Therefore, the number of hidden layers and the number of neurons in each layer are called ‘‘Hyperparameters’’. Besides of these two, the activation function is

also seen as a hyperparameter.

During the training, the NN is supplied with  $N$  data events  $\mathbf{x}_a = (x_1, x_2, \dots, x_n)_a$  corresponding to expected outputs  $\hat{y}_a$ ,  $a = 1, 2, \dots, N$ . For each data event the NN gives an output  $y_a$  through forward propagation. The disagreement between NN outputs and the desired predictions is measured by the loss function ( $L$ ). There are different types of loss functions applicable for different tasks. For instance, a regression NN is expected to predict continuous values, so it can use Mean Absolute Error (MAE), Mean Square Error (MSE), Mean Squared Logarithmic Error (MSLE), Mean Absolute Percentage Error (MAPE) or Huber Loss [123] as the loss function. For a binary classification task or other tasks expecting discrete output, Binary Cross-Entropy Loss and Hinge Loss would do better [124]. The Cross-Entropy Loss is often used for multi-class classification. The mathematical equations of these loss functions are shown below:

- Mean Absolute Error (MAE) /  $L_1$  Loss:

$$L = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|. \quad (5.10)$$

- Mean Square Error (MSE) /  $L_2$  Loss:

$$L = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2. \quad (5.11)$$

- Mean Squared Logarithmic Error (MSLE):

$$L = \frac{1}{N} \sum_{i=1}^N (\log(1 + y_i) - \log(1 + \hat{y}_i))^2. \quad (5.12)$$

- Mean Absolute Percentage Error (MAPE):

$$L = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100. \quad (5.13)$$

- Huber Loss [123]:

$$L_\delta = \begin{cases} \frac{1}{2}(y - \hat{y})^2 & \text{for } |y - \hat{y}| \leq \delta, \\ \delta(|y - \hat{y}| - \frac{1}{2}\delta) & \text{for } |y - \hat{y}| > \delta. \end{cases} \quad (5.14)$$

- Binary Cross-Entropy Loss / Log Loss:

$$L = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]. \quad (5.15)$$

- Hinge Loss:

$$L = \frac{1}{N} \sum_{i=1}^N \max(0, 1 - y_i \hat{y}_i). \quad (5.16)$$

- Cross-Entropy Loss:

$$L = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{ic} \log(\hat{y}_{ic}), \quad (5.17)$$

where  $C$  denotes the number of classes.

The loss function is also seen as one of hyperparameters. As shown in Figure 5.4, different loss functions have different levels of sensitivity to outlying predictions. Therefore, it is important to select a proper loss function for a NN.

Since the loss function quantifies the discrepancy between the NN prediction and the actual data, the purpose of a NN is to minimize the loss function. Notice that the loss function is not a function of  $x_1, x_2, \dots, x_n$ , but a function of weights ( $w_{ij}^{[1]}$ ,  $w_{ij}^{[2]}$ ). Hence to train a NN means to find an optimal set of weights corresponding to the global minimum of the loss function.

## 5.2.2 Backpropagation and Optimizers

In a NN, the weights are adjusted repeatedly based on their contribution to the overall loss. Backpropagation [125] is an algorithm to fast compute the gradient of

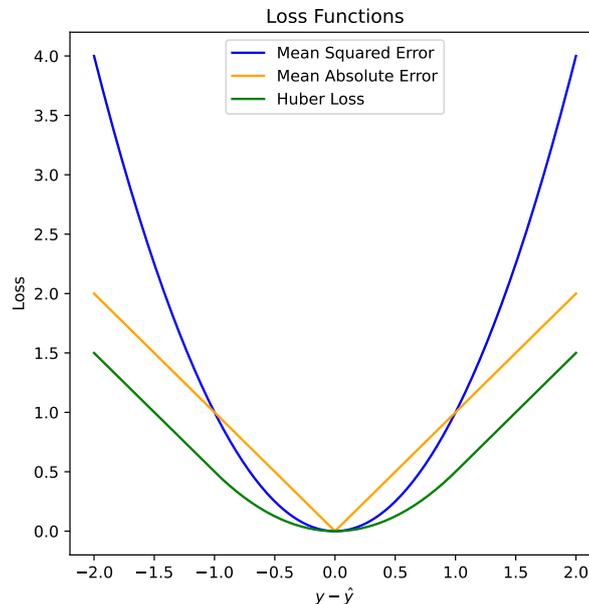


Figure 5.4: The loss: (blue) Mean Square Error , (orange) Mean Absolute Error and (green) Huber Loss, as functions of the difference between the NN prediction and the expected value with  $N = 1$ .

the loss function. The rule of weight-adjustment is provided by the optimization algorithm, also called the optimizer. A basic optimizer is the gradient descent algorithm, where the gradients of the loss function ( $L$ ) with respect to the weights are needed. Since the activation function is designed to be differentiable, the partial derivative of  $L$  with respect to each weight can be derived from the chain rule. For the weights of the output layer:

$$\frac{\partial L}{\partial w_{j1}^{[2]}} = \frac{1}{N} \sum_{a=1}^N \frac{\partial L_a}{\partial w_{j1}^{[2]}} = \frac{1}{N} \left( \sum_{a=1}^N \frac{\partial L_a}{\partial y} \frac{\partial y}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial w_{j1}^{[2]}} \right), \quad (5.18)$$

where  $z^{[2]}$  is the wighted sum in the output neuron and  $L_a$  indicates the loss calculated from the  $a$ -th data  $\mathbf{x}_a$ . By omitting the summation and the subscript  $a$  for simplification, the equation above can be written as

$$\frac{\partial L}{\partial w_{j1}^{[2]}} = \frac{\partial L}{\partial y} \frac{\partial y}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial w_{j1}^{[2]}} = \frac{\partial L}{\partial y} \cdot g'(z^{[2]}) \cdot y_j^{[1]}. \quad (5.19)$$

For the hidden layer:

$$\begin{aligned}\frac{\partial L}{\partial w_{ij}^{[1]}} &= \frac{\partial L}{\partial y} \frac{\partial y}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial y_j^{[1]}} \frac{\partial y_j^{[1]}}{\partial z_j^{[1]}} \frac{\partial z_j^{[1]}}{\partial w_{ij}^{[1]}} \\ &= \frac{\partial L}{\partial y} \cdot g'(z^{[2]}) g'(z_j^{[1]}) \cdot w_{j1}^{[2]} x_i\end{aligned}\quad (5.20)$$

Equation (5.19) and (5.20) are derived from a very simple NN with only one hidden layer and one output neuron. This method can be generalised to a NN with multiple hidden layers and multiple output neurons. For the output layer of a NN with  $(S-1)$  hidden layers, it can be written that

$$\frac{\partial L}{\partial z_j^{[s]}} = \frac{\partial L}{\partial y} \frac{\partial y}{\partial z_j^{[s]}} = \frac{\partial L}{\partial y} \cdot g'(z_j^{[s]}), \quad (5.21)$$

where  $z_j^{[s]}$  denotes the weighted sum in the  $j$ -th output neuron. In the  $s$ -th hidden layer ( $s = 1, 2, \dots, S-1$ ), the change of a neuron affects all neurons in the next layer, therefore

$$\frac{\partial L}{\partial z_j^{[s]}} = \sum_{k=1}^K \left( \frac{\partial L}{\partial z_k^{[s+1]}} \frac{\partial z_k^{[s+1]}}{\partial y_j^{[s]}} \frac{\partial y_j^{[s]}}{\partial z_j^{[s]}} \right) = \sum_{k=1}^K \left( \frac{\partial L}{\partial z_k^{[s+1]}} \cdot w_{jk}^{[s+1]} \right) \cdot g'(z_j^{[s]}), \quad (5.22)$$

where  $K$  is the number of neurons in the next layer. Because the weighted sum of a neuron is

$$z_j^{[s]} = \sum_i w_{ij}^{[s]} y_i^{[s-1]}, \quad (5.23)$$

the derivative of  $L$  with respect to the weights can be written as

$$\frac{\partial L}{\partial w_{ij}^{[s]}} = \frac{\partial L}{\partial z_j^{[s]}} \frac{\partial z_j^{[s]}}{\partial w_{ij}^{[s]}} = \frac{\partial L}{\partial z_j^{[s]}} \cdot y_i^{[s-1]}, \quad s = 1, 2, \dots, S, \quad (5.24)$$

the term  $\frac{\partial L}{\partial z_j^{[s]}}$  can be derived by combining equations (5.21) and (5.22). Note that the calculation in these equations starts from the output layer, then proceeds layer by layer backward. That is the reason of ‘backpropagation’ algorithm getting its name. To minimize the loss function, weights are initialized randomly, then adjusted

by the rule of the gradient descent:

$$\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} - \eta \cdot \frac{\partial L}{\partial \mathbf{w}^{(t)}}, \quad (5.25)$$

where  $\mathbf{w}^{(t)}$  indicates the set of weights in the  $t$ -th iteration step.  $\eta$  is the learning rate, a hyperparameter presenting the step size to move against gradients in each iteration. The updating of weights continues until the loss reaching a minimum, as shown in Figure 5.5. The learning rate affects the speed and the quality of the training, so it should be chosen appropriately. A low learning rate ensures precise learning, but increases the time cost of the training (Figure 5.6(a)), while a high learning rate has the risk of ‘jumping over’ the global minimum (Figure 5.6(b)) or even make the training diverges (Figure 5.6(c)). In some training processes, the learning rates are designed to slowly reduce over iterations, such a technique is called the decay of the learning rate. The most common decay is as following:

$$\eta = \eta_0 \cdot \frac{1}{1 + \text{Decay rate} \times \text{Iteration number}}, \quad (5.26)$$

where  $\eta_0$  is the initial learning rate,  $\eta$  is the learning rate of current iteration. The decay rate is a hyperparameters that need to be tuned.

The NN is usually designed for a large dataset, and its loss function is very high dimensional. The minimum reached by basic gradient descent in such a high-dimension space is not likely to be the global minimum. For the loss function to converge faster and get better performance, several types of optimizers are designed for NNs. For example, stochastic gradient descent (SGD) [126], as well as some more advanced optimizers like SGD with momentum [127], AdaGrad [128], RMSprop [129] and Adam [130].

- The SGD (or mini-batch SGD) algorithm is a variant of gradient descent. Instead of using the entire training data to calculate the loss function, SGD only uses a randomly selected subset of data. Usually, the optimizer using a single data event for the loss is called SGD. If the number of events used

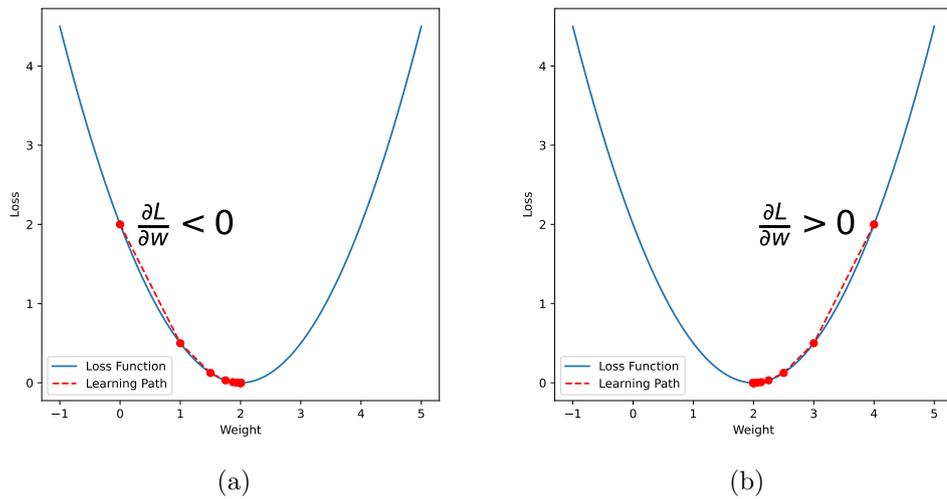


Figure 5.5: A simple example of the gradient descent, where the loss is a function of single weight. (a) The gradient of the Loss function is negative. (b) The gradient of the Loss function is positive.

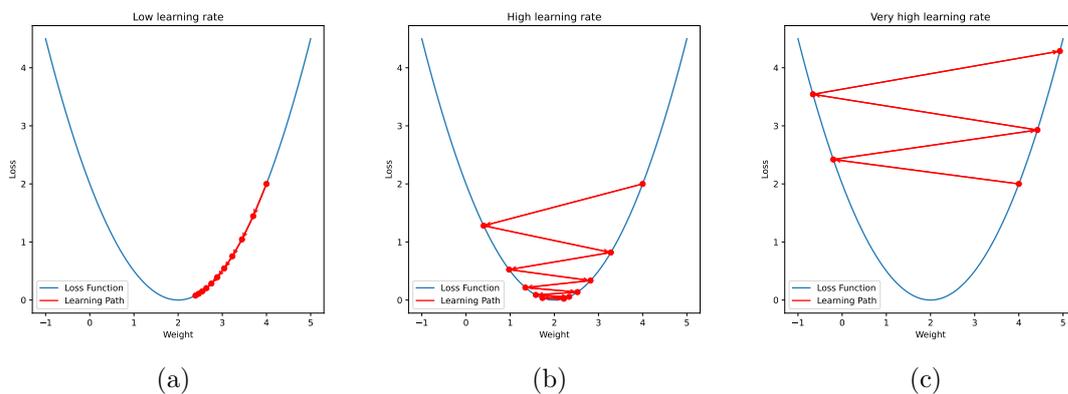


Figure 5.6: The training with different learning rates. (a) The training converges very slowly with a low learning rate. (b) The training passes the minimum with a high learning rate. (c) The training diverges if the learning rate is very high,

for the loss,  $k$ , is in the range  $1 < k < N$ , the optimizer is called mini-batch SGD, and  $k$  is the hyperparameter: batch size. The updating of weights can be described as

$$\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} - \eta \cdot \frac{\partial L(w^{(t)}; x_i, y_i)}{\partial \mathbf{w}^{(t)}}, \quad (5.27)$$

where  $(x_i, y_i)$  is a mini-batch. Since the calculation of the loss function in SGD does not include all the training data, the NN training process is faster and more efficient than that in the original gradient descent. However, SGD may obtain noisy gradients due to reduced information with respect to the full dataset. With a larger batch size, the loss function becomes smoother. So the batch size is a hyperparameter that balances between training speed and accuracy.

- SGD with momentum introduces a “momentum” term to SGD, which includes the direction of previous updates during the current update. The rule of SGD with momentum is as following:

$$w^{(t+1)} = w^{(t)} - v_t, \quad (5.28)$$

where  $v_t$  is the ‘moving’ of the weight in the  $t$ -th iteration, which also contains the ‘moving’ information from previous iterations:

$$v_t = \gamma v_{t-1} + \eta \frac{\partial L}{\partial w^{(t)}}, \quad (5.29)$$

where  $\gamma$  is the hyperparameter: momentum, and  $\eta$  is the learning rate. If the update has the same direction as previous updates, the SGD with momentum tends to take a larger step to converge faster and smoother than SGD. The momentum term can let the training escape a local minimum.

The rule in equations (5.28) and (5.29) can also be expressed in the following form:

$$v_t = \beta v_{t-1} + (1 - \beta) \frac{\partial L}{\partial w^{(t)}}, \quad (5.30)$$

$$w^{(t+1)} = w^{(t)} - \eta v_t, \quad (5.31)$$

where  $\beta$  is a parameter to be tuned in the range  $0 \leq \beta < 1$ ,  $\eta$  is the learning rate scaled by a factor of  $1/(1-\beta)$ . The term  $v_t$  in equation (5.30) is also called the exponentially weighted moving average (EWMA) or exponential moving average (EMA) of previous gradients.

- AdaGrad stands for the adaptive gradient descent optimizer. ‘Adaptive’ means that the learning rate of the algorithm is adaptively scaled for different neurons and different layers based on iterations. The weight-updating rule of AdaGrad is given by:

$$w^{(t+1)} = w^{(t)} - \frac{\eta}{\sqrt{\alpha_t + \epsilon}} \frac{\partial L}{\partial w^{(t)}}, \quad (5.32)$$

where  $\eta$  is the initial learning rate,  $\epsilon$  is a small quantity to avoid zero denominator,  $\alpha_t$  is the sum of squares of all previous gradients of the weight:

$$\alpha_t = \sum_{i=1}^t \left( \frac{\partial L}{\partial w^{(i)}} \right)^2. \quad (5.33)$$

By using the AdaGrad optimizer, weights with large gradients or frequent updates get low learning rates so that the training does not jump over the minimum. For the weights with small gradients or infrequent updates, the learning rates are scaled up for faster converging. Since  $\alpha_t$  is monotonically increasing over iterations, the learning rate may become extremely small if the training takes too many iterations. This is the main weakness of the AdaGrad optimizer in the deep NN training.

- RMSprop (Root Mean Square Propagation) is an optimizer similar to AdaGrad. However, instead of using the sum of all previous squares of gradients, RMSprop is using the EWMA of previous squares of gradients to scale the learning rates. The rule is given by:

$$w^{(t+1)} = w^{(t)} - \frac{\eta}{\sqrt{E[(\partial L/\partial w)^2]_t + \epsilon}} \frac{\partial L}{\partial w^{(t)}}, \quad (5.34)$$

where  $E[(\partial L/\partial w)^2]_t$  is the EWMA of previous squares of gradients:

$$E \left[ \left( \frac{\partial L}{\partial w} \right)^2 \right]_t = \beta \cdot E \left[ \left( \frac{\partial L}{\partial w} \right)^2 \right]_{t-1} + (1 - \beta) \left( \frac{\partial L}{\partial w^{(t)}} \right)^2, \quad 0 \leq \beta < 1. \quad (5.35)$$

Other parameters are same as AdaGrad,  $\eta$  is the initial learning rate, and  $\epsilon$  is a small quantity. The hyperparameter  $\beta$  is less than 1, hence older gradients contribute less to  $s_t$  so that the algorithm prevents the learning rate being too small.

- The Adam (Adaptive Moment Estimation) [130] optimizer combines the ideas of momentum and RMSprop. The scaling of learning rate in Adam uses both the EWMA of previous gradients  $v_t$  and the EWMA of previous squares of gradients  $s_t$ . The full form of its rule is given by:

$$v_t = \beta_1 v_{t-1} + (1 - \beta_1) \frac{\partial L}{\partial w^{(t)}}, \quad 0 \leq \beta_1 < 1, \quad (5.36)$$

$$s_t = \beta_2 s_{t-1} + (1 - \beta_2) \left( \frac{\partial L}{\partial w^{(t)}} \right)^2, \quad 0 \leq \beta_2 < 1, \quad (5.37)$$

$$w^{(t+1)} = w^{(t)} - \eta \frac{\hat{v}_t}{\sqrt{\hat{s}_t + \epsilon}}, \quad (5.38)$$

where  $\hat{v}_t$  and  $\hat{s}_t$  are  $v_t$  and  $s_t$  after the bias correction:

$$\hat{v}_t = \frac{v_t}{1 - \beta_1^t}, \quad \hat{s}_t = \frac{s_t}{1 - \beta_2^t}. \quad (5.39)$$

The bias correction is used to correct the bias from the initialization, which are quite large in the first few iterations. The Adam optimizer is easy to implement and has little memory requirements [130]. It often has faster convergence than other optimizers, especially in a high-dimensional complex model [130].

The weights of a NN are updated each iteration following the rule of the optimizer. Once the model processes all data in the entire training dataset, it can be said that the algorithm finishes one epoch. If the optimizer is implemented with mini-batches, the update takes place in every batch. Therefore for a dataset of size  $N$ , the weight

update takes place  $N/k$  times in each epoch, where  $k$  is the batch size.

### 5.2.3 Over-fitting treatment

Sometimes the NN learns statistical fluctuations within the training dataset so that the NN model has a too good performance on the training data, but does not generalise well and cannot predict unseen data with high accuracy. Such a NN is called over-trained, and the model is over-fit. Figure 5.7 shows example results from a well-fit model and an over-fit model of a classification task.

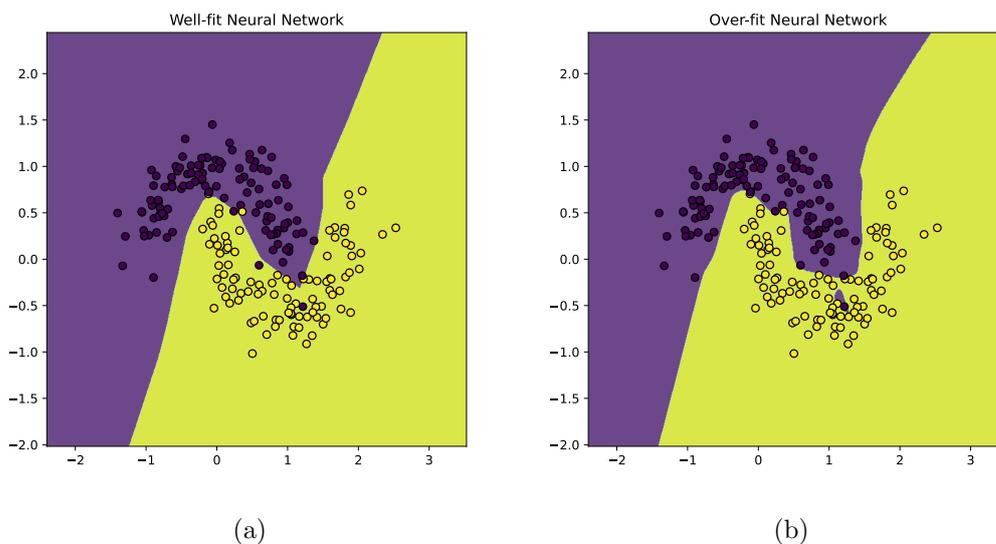


Figure 5.7: Simple examples of (a) a well-fit classification NN and (b) an over-fit classification NN. The dots in purple and yellow are training data points with two different classes. The purple and yellow areas are classifications from the NN prediction. The over-fit NN has high accuracy on classifying the training data, but the model includes the noise in the dataset and is unable to generalize well to unseen data.

Over-fitting is not rare in deep NN training. To detect over-fitting, a popular method is to split the dataset into two subsets with equal sizes. The NN is trained on one of the subsets, while the other is used for validation. The NN becomes over-fit when the loss of test dataset starts to increase while the loss of training dataset is decreasing. The training process is then stopped to prevent the NN from further

over-training.

Other than early stopping, there are several methods to avoid over-fitting. For a complex NN, the regularization method can be helpful, which applies a penalty to the NN parameters to limit their variance. Three common used regularization schemes are introduced in this section: L1 regularization, L2 regularization and dropout.

- Both L1 and L2 regularization are based on similar ideas: adding an extra penalty term on the loss function to penalize large weights. An over-fit model may have large weights that provide big changes in the output. By penalizing the large weights, the NN can be more stable. The loss of L1 regularization is given by [131]:

$$L_{L1} = L_0 + \lambda_1 \sum_w |w|, \quad (5.40)$$

and the loss of L2 regularization is [132]:

$$L_{L2} = L_0 + \lambda_2 \sum_w w^2, \quad (5.41)$$

where  $\lambda_1$  and  $\lambda_2$  are called regularization factors, they are hyperparameters controlling the effect of the penalty. The penalty shrinks weights in the NN. The larger the penalty is, the smaller the NN weights. The L1 regularization is often used in linear regression and simple tasks, while L2 regularization is better for a complex model.

- Dropout reduces over-fitting by temporarily removing random neurons and their connections with other neurons [133], as shown in Figure 5.8. In each iteration of training, different sets of neurons are deactivated and the model gets different structures. In such way, the training of a NN combines several different models efficiently, and the combination always improves the performance of the NN [133]. Neurons are usually dropped with a fixed probability  $p$ , which is defined as a hyperparameter: dropout rate.

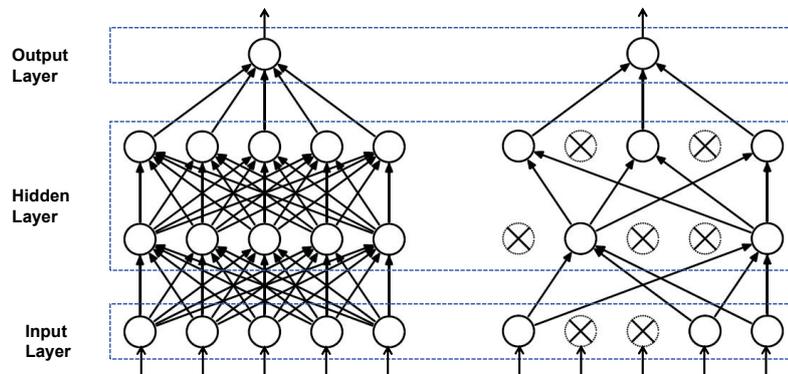


Figure 5.8: The structures of (left) a standard NN and (right) a NN applying dropout [133].

### 5.3 Hyperparameter optimization

Several NN hyperparameters were introduced in previous sections, summarized in Table 5.1. Those hyperparameters are critical in the training process and highly affect the performance of the NN. So it is important to optimize hyperparameters of a NN.

Hyperparameters like activation function, loss function and optimizer are sensitive to different types of the tasks. For example, the regression NN often uses the linear activation function at the output layer, while the classification NN usually uses the sigmoid function. For the hidden layers of a FFNN, ReLU is the most commonly used activation function. These hyperparameters can be simply chosen following recommendations, or by selecting ones with the best performance after test trials for different settings.

Other hyperparameters, such as the number of hidden layers, the number of neurons in each layer, learning rate, momentum, etc, have numeric values that should be tuned. There are two naive methods to tune these hyperparameters, called grid search and randomized search [134]. Just as their names, the grid search means that

Hyperparameter	Description
Number of layers	Depth of the neural network, the number of hidden layers.
Neurons per layer	Number of neurons in each hidden layer.
Activation function	Non-linear transformation applied to neuron outputs.
Loss function	Function measuring the disagreement between expected predictions and NN outputs.
Optimizer	Algorithm used to minimize the loss function.
Learning rate	Controls the step size during gradient descent.
Decay rate	Factor to reduce the learning rate after a certain number of epochs.
Momentum	Factor that accelerates gradient descent in the relevant direction.
Regularization	Adding extra penalty terms to the loss to penalize large weights.
Dropout rate	Probability of neurons being dropped during the training to prevent overfitting.
Batch size	Number of training data samples used in one iteration.
Number of epochs	Number of times the entire training dataset passes through the network.

Table 5.1: Commonly used NN hyperparameters.

hyperparameters are tried in different values with fixed intervals, and the randomized search selects values randomly. Both of these methods are commonly used, but they have some disadvantages that the chosen values may not improve the performance of the NN, and the optimization process can be slow for a complex NN.

There are some more advanced algorithms for hyperparameter optimization, such as Bayesian optimization (BO) [135] and Tree-structured Parzen Estimator (TPE) [136]. They can provide results faster and more robustly compared to grid search and randomized search. Both of BO and TPE algorithms use probabilistic methods, TPE is a variant of BO methods. Their aims are to find a good hyperparameter combination that has higher probability to achieve better NN performance. These algorithms iterate between calculating the probability distribution over different combinations based on observations and collecting new observations based on this probability. The iteration keeps going until the probability distribution converges or the iteration number reaches a set number. Such type of algorithms are called the Sequential Model-Based Optimization (SMBO) [137].

---

## Searches for $H \rightarrow Za(\rightarrow \text{hadronic})$

---

### 6.1 Introduction

This chapter presents a search for decays of the Higgs boson to a  $Z$  boson and a light hadronically decaying resonance. The  $Z$  boson is required to decay to a pair of same-flavour leptons ( $\ell$ ), i.e. electrons, muons or  $\tau$ -leptons if both  $\tau$  decay leptonically. The target light resonance is required to decay hadronically. It is reconstructed as a single jet as its mass is lower than 4 GeV and it is highly boosted. 140 fb<sup>-1</sup> of ATLAS Run 2 data is used in this search. This analysis has three interpretations: first, a Standard Model charmonium resonance ( $\eta_c$  or  $J/\psi$ ); second, the light pseudo-scalar resonance ( $a$ ) from 2HDM or 2HDM+S; at last an axion interpretation.

The Higgs decays to non-SM particles are categorised into decays to invisible particles and other undetected decays because of large backgrounds [7], the 95% upper limits on the corresponding branching fractions are about 13% and 12% respec-

Decay channel	Experiment	Mass range (GeV)
$H \rightarrow aa \rightarrow b\bar{b}b\bar{b}$ [41]	ATLAS	15-30
$H \rightarrow aa \rightarrow b\bar{b}b\bar{b}$ [42]	ATLAS	20-60
$H \rightarrow aa \rightarrow b\bar{b}b\bar{b}$ [140]	CMS	20-60
$H \rightarrow aa \rightarrow b\bar{b}\tau^+\tau^-$ [141]	CMS	15-60
$H \rightarrow aa \rightarrow b\bar{b}\tau^+\tau^-, b\bar{b}\mu^+\mu^-$ [46]	CMS	20-62.5, 20-65
$H \rightarrow aa \rightarrow b\bar{b}\mu^+\mu^-$ [40]	ATLAS	16-62
$H \rightarrow aa \rightarrow b\bar{b}\mu^+\mu^-$ [43]	ATLAS	20-60
$H \rightarrow aa \rightarrow b\bar{b}\mu^+\mu^-$ [142]	CMS	20-62.5
$H \rightarrow aa \rightarrow b\bar{b}\mu^+\mu^-$ [143]	CMS	20-65
$H \rightarrow aa \rightarrow \tau^+\tau^-\tau^+\tau^-$ [144, 145]	CMS	4-8, 4-15
$H \rightarrow aa \rightarrow \tau^+\tau^-\tau^+\tau^-, \mu^+\mu^-\tau^+\tau^-$ [45]	CMS	4-15
$H \rightarrow aa \rightarrow \mu^+\mu^-\tau^+\tau^-$ [146]	ATLAS	4-50
$H \rightarrow aa \rightarrow \mu^+\mu^-\tau^+\tau^-$ [147]	CMS	15-62.5
$H \rightarrow aa \rightarrow \mu^+\mu^-\tau^+\tau^-$ [148]	CMS	20-65
$H \rightarrow aa \rightarrow \mu^+\mu^-\mu^+\mu^-$ [149]	CMS	0.2-2
$H \rightarrow aa \rightarrow \gamma\gamma jj$ [150]	ATLAS	20-60
$H \rightarrow aa \rightarrow \gamma\gamma\gamma\gamma$ [151]	CMS	0.1-1.2
$H \rightarrow aa \rightarrow \gamma\gamma\gamma\gamma$ [152]	ATLAS	10-60
$H \rightarrow aa \rightarrow \gamma\gamma\gamma\gamma$ [153]	CMS	15-62
$H \rightarrow (Z/a)a \rightarrow \ell_1^+\ell_1^-\ell_2^+\ell_2^-$ [154]	ATLAS	1-60
$H \rightarrow Za \rightarrow l^+l^-b\bar{b}$ [155]	CMS	$<10^3$
$H \rightarrow Za \rightarrow l^+l^-\gamma\gamma$ [48]	ATLAS	0.1-33
$H \rightarrow Za \rightarrow l^+l^-\gamma\gamma$ [49]	CMS	1-30
$H \rightarrow Za \rightarrow l^+l^-\text{jet}$ [50]	ATLAS	$<4$

Table 6.1: Searches for Higgs boson decays to light resonances by ATLAS and CMS.

tively [7].

Previous analyses have set limits on the branching ratios of the Higgs boson decays to light resonances ( $a$ ). The DØ experiment at the Tevatron set limits in the  $H \rightarrow aa \rightarrow \mu^+\mu^-\mu^+\mu^-$  and  $H \rightarrow aa \rightarrow \mu^+\mu^-\tau^+\tau^-$  decay modes [138] for a mass range of 4-18 GeV. The ATLAS and CMS experiments have also performed searches at the LHC using various channels that led to upper limits in various mass ranges. A summary of the searches is shown in Table 6.1. These searches mostly focused on  $H \rightarrow aa$  decays, and the final states of  $a$  include leptons, photons or bottom quark pairs. The pseudoscalar  $a$  can decay also into coloured particles, the relevant modes are  $a \rightarrow gg$  and  $a \rightarrow q\bar{q}$ . These decays are allowed if  $m_a > m_\pi$  [139].

The branching ratios of  $H \rightarrow aa$  and  $H \rightarrow Za$  can be constrained independently. The first search for hadronic decays of the  $a$  over the low mass range ( $\leq 4$  GeV), and one of the few searches for  $H \rightarrow Za$  was published by the ATLAS experiment in 2020 [50]. It set 95% confidence level upper limits on the cross section of Higgs boson production times the branching ratio of  $H \rightarrow Za$ , the results are shown in Figure 6.1. That analysis also performed searches for decays of Higgs boson to a  $Z$  boson and a charmonium resonance ( $\eta_c$  or  $J/\psi$ ). No significant excess is found. The main factor limiting the analysis sensitivity was the uncertainty of the background model, including the pure MC statistical uncertainty and the shape uncertainty. This analysis was documented in the PhD thesis of Dr Elliot Reynolds [156]. In this iteration of the analysis we aspire to go beyond the limitations observed in the first implementation.

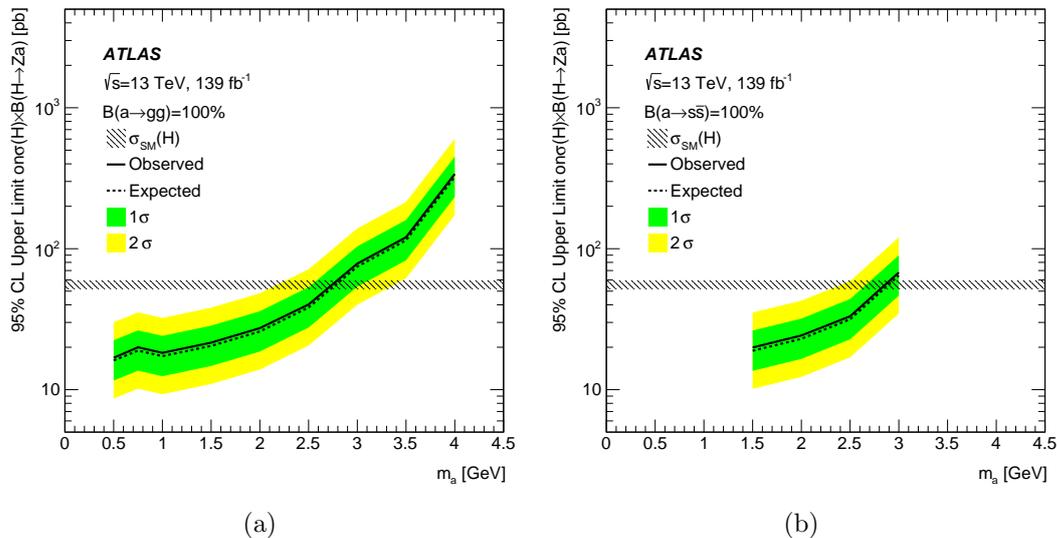


Figure 6.1: Observed and expected 95% confidence level upper limits on  $\sigma(H)BR(H \rightarrow Za)$  for the assumption (a)  $BR(a \rightarrow gg)=100\%$  and (b)  $BR(a \rightarrow s\bar{s})=100\%$ . Limits set by Ref. [50].

By changing to a higher statistics POWHEG background MC sample from SHERPA samples, and introducing a shape fit instead of a cut and count approach, the sensitivity is expected to improve significantly. Also a dedicated neural network for reweighting is used to account for mismatches between data and MC. For the jet

reconstruction, PFlow jets are used instead of EMTopo jets since PFlow jets are better in terms of reconstruction efficiency and energy resolution, as discussed in Section 3.3.4. So this second iteration is strongly motivated by the great improvement on the exclusion limits, as well as by a new axion interpretation.

### 6.1.1 Light Scalar Bosons

As discussed in Section 2.4, both 2HDM and 2HDM+S feature extended Higgs sectors, allowing for a light pseudoscalar ( $a$ ) or a lighter scalar Higgs boson, and the  $a$  can have a large coupling to the observed Higgs boson. These two models are necessary to generate the masses of quarks in the Minimal Supersymmetric Model (MSSM), and the Next-to-MSSM (NMSSM), respectively [157]. Due to the narrow width of a SM Higgs boson with a mass of 125 GeV, even a small coupling to a non-SM Higgs boson could result in new decay modes with large branching ratios.

In the low mass range ( $m_a < 4$  GeV), the pseudoscalar  $a$  mostly decays to hadronic final states [9], as shown in Figure 2.9.  $a$  decays to leptons are also suppressed in the low (high)  $\tan\beta$  phase space of the Type-II and Type-III (Type-I and Type-IV) 2HDM and 2HDM+S [9]. These parts of the 2HDM(+S) phase space are largely unconstrained by previous searches, but can be probed with this search.

### 6.1.2 Charmonium States

The coupling of the Higgs boson to quarks is one of the windows to search for new physics [158]. The couplings of the Higgs boson to third generation quarks have been established experimentally [159, 160]. However, the coupling of the Higgs boson to first and second generation quarks have not yet been measured. Loose constraint has been made on the couplings of the Higgs boson to quarks via decays of the Higgs boson to a  $Z$  boson and light SM resonances, but there is still a large phase space not yet constrained [161, 162]. This thesis is also searching for decays of the

Higgs boson to a  $Z$  boson and a SM charmonium resonance ( $J/\psi$  or an  $\eta_c$ ), which is sensitive to the new physics [158], while the SM branching ratios of  $H \rightarrow Z\eta_c$  and  $H \rightarrow ZJ/\psi$  are  $1.4 \times 10^{-5}$  and  $2.2 \times 10^{-6}$  respectively [163].

### 6.1.3 Axions

Axions and Axion-Like Particles (ALP) [139] are also models predicting light pseudoscalar resonances. Axions were initially introduced to solve the strong CP problem [164]. ALPs have similar properties as axions but with much wider interests. Decay modes for two ALP models are shown in Figure 6.2. Figure 6.3 shows Feynman diagrams contributing to the decay  $H \rightarrow Za$  for an ALP  $a$ .

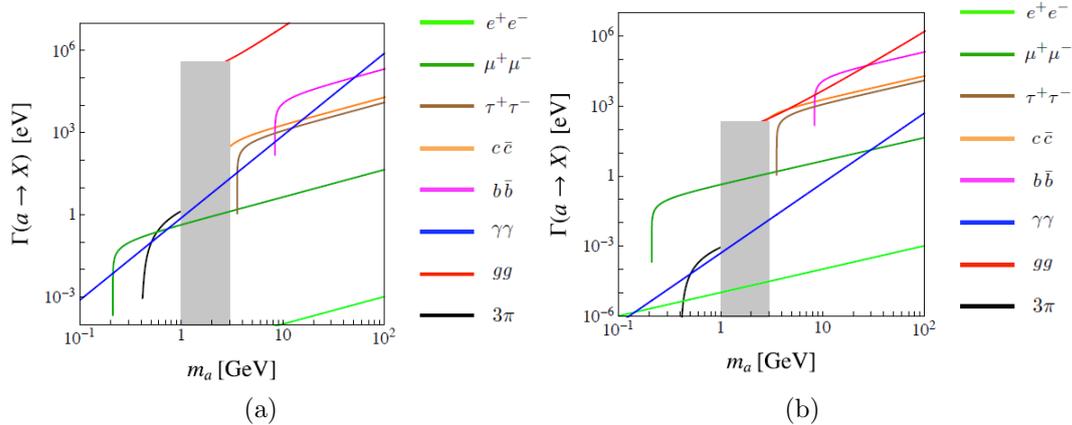


Figure 6.2: ALP decay rates into pairs of SM particles obtained (a) by setting the relevant effective Wilson coefficients to 1, or (b) by setting the ALP fermion couplings coefficients to 1 and the ALP boson couplings to  $1/(4\pi^2)$ . The gray area between 1 and 3 GeV shows the region in which various exclusive hadronic decay channels dominate, which is difficult to calculate [139].

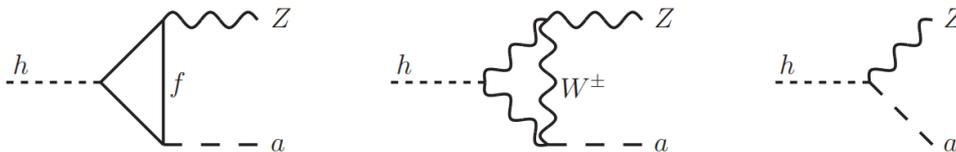


Figure 6.3: Feynman diagrams contributing to the decay  $H \rightarrow Za$  [139].

## 6.2 Experimental and Simulated Data Samples

### 6.2.1 Data Samples

The analysis uses the full Run 2 dataset collected by the ATLAS detector at  $\sqrt{s} = 13$  TeV, corresponding to a total integrated luminosity of  $140 \text{ fb}^{-1}$ . For each year between 2015-2018, the integrated luminosity of data is  $3.2 \text{ fb}^{-1}$ ,  $33.4 \text{ fb}^{-1}$ ,  $44.6 \text{ fb}^{-1}$  and  $58.8 \text{ fb}^{-1}$  respectively. In order to reduce the data size, the dataset is processed through the dedicated ‘‘HDBS3 derivation’’, which has some loose selection on the events and includes the jet substructure variables used to separate the light resonance from multijet production.

### 6.2.2 Simulated Samples

Monte-Carlo (MC) simulation samples used in this search are summarised in Table 6.2. All background and signal MC samples have been processed using the HDBS3 derivation, and are used for the background and signal modelling.

The signal MC samples are generated via gluon-gluon Fusion (ggF) process. The SM Higgs boson is produced in POWHEG [165, 166, 167], using the AZNLO tune parameters [168]. Signal events are modelled using PYTHIA8 [169] (v8.212) and EVTGEN [170], interfaced to the CT10 [171] and CTEQ6L1 PDF sets. Because the signal is a BSM process, the SM Higgs boson is replaced by the heavy neutral scalar Higgs ( $H^0$ ) from the 2HDM, and subsequently decays to a  $Z$  boson and a neutral pseudoscalar  $a$ . Decays of the BSM Higgs bosons are generated with the default PYTHIA8 2HDM settings, and  $a$  can decay to any channel. Branching ratios of  $a$  decays to quarks and gluons are shown in Table 6.3. Alternative signal samples are generated using HERWIG7 event generator [172, 173]. The  $a$  is only allowed to decay to gluons or quarks,  $a \rightarrow qq$  and  $a \rightarrow gg$  events are modelled separately. These HERWIG samples are used to determine the signal modelling uncertainties.

Signal	$gg \rightarrow H \rightarrow Z(\ell\ell) + \eta_c$	POWHEG+PYTHIA8	PYTHIA8+EVTGEN	
	$gg \rightarrow H \rightarrow Z(\ell\ell) + J/\psi$	POWHEG+PYTHIA8	PYTHIA8+EVTGEN	
	$gg \rightarrow H \rightarrow Z(\ell\ell) + a$	POWHEG+PYTHIA8	PYTHIA8+EVTGEN	$m_a \in (0.5, 0.75, 1, 1.5, 2, 2.5, 3, 3.5, 4, 8)$ GeV
	$gg \rightarrow H \rightarrow Z(\ell\ell) + a(gg)$	HERWIG+EVTGEN		$m_a \in (2, 2.5, 3, 3.5, 4)$ GeV
	$gg \rightarrow H \rightarrow Z(\ell\ell) + a(q\bar{q})$	HERWIG+EVTGEN		$m_a \in (2, 2.5, 3, 3.5, 4)$ GeV
Background	$pp \rightarrow Z(\ell\ell) + \text{jets}$	POWHEG+PYTHIA8	PYTHIA8+EVTGEN	$Z \rightarrow ee$ $Z \rightarrow \mu\mu$
	$pp \rightarrow Z(\ell\ell) + \text{jets}$	SHERPA2.2.1		Max( $H_T, p_T^V$ ) slices and flavor filtered
	$gg \rightarrow Z(\ell\ell) + Z(q\bar{q})$	SHERPA2.2.2		
	$pp \rightarrow Z(\ell\ell) + Z(q\bar{q})$	SHERPA2.2.1		
	$pp \rightarrow Z(\ell\ell) + W(q\bar{q})$	SHERPA2.2.1		
	$pp \rightarrow t\bar{t}$	POWHEG	PYTHIA8+EVTGEN	$\geq 2\ell$
	VBF Higgs	POWHEG	PYTHIA8+EVTGEN	

Table 6.2: Simulated samples for signal and background.

$a$ Mass (GeV)	Branching Ratio
0.5	$gg$ (97%)
0.75	$gg$ (91%)
1	$gg$ (90%)
1.5	$gg$ (80%), $s\bar{s}$ (14%)
2	$gg$ (75%), $s\bar{s}$ (11%)
2.5	$gg$ (90%), $s\bar{s}$ (7%)
3	$gg$ (88%), $s\bar{s}$ (8%)
3.5	$gg$ (11%), $s\bar{s}$ (1%)
4	$gg$ (6%), $c\bar{c}$ (56%)

Table 6.3: Branching ratios of the gluons and quarks decay modes, for various  $a$  mass points. Values are derived from PYTHIA8 signal samples.

---

The background is comprised mostly by  $Z + \text{jets}$  events. POWHEG+PYTHIA and PYTHIA8+EVTGEN are used for event generation and showering, respectively. The PYTHIA8 background process is  $V + 0j@NLO$  without any lepton filter, also using the AZNLO tune [168]. The choice is motivated by the higher MC statistics than SHERPA2.2.1 samples. Only samples with  $Z \rightarrow ee$  and  $Z \rightarrow \mu\mu$  decays are used. An additional contribution of about 5% is expected from  $Z \rightarrow \tau\tau$  decays, with very similar event kinematics. It can be corrected by the background reweighting procedure. For the  $Z + \text{jets}$  background a cross-check sample is produced, it is modelled using an alternative SHERPA2.2.1 [174] interfaced to the NNPDF [175] PDF sets. The samples are produced in parts based on  $H_T$ , which is the maximum of the scalar  $p_T$  for all jets and leptons in the event, and the generator-level  $Z$  boson  $p_T$ . They are also categorised by the number of generator-level heavy flavour partons. The corresponding inclusive production cross sections are calculated to NNLO in QCD [176]. It is noted that weights of anomalous high-weight (magnitude above 100) events in the SHERPA samples are set to be 1 according to ATLAS recommendations.

The diboson ( $ZZ$  and  $ZW$ ) processes contribute  $< 1\%$  of total backgrounds, they are modelled by SHERPA2.2.1 with the NNPDF 3.0 PDF set. The  $t\bar{t}$  production contributes less than 1% of the background. It is modelled using POWHEG interfaced to PYTHIA8 and EVTGEN.

All these background processes are merged and called Combined background.

The above mentioned samples are processed through a complete GEANT4 [177] simulation of ATLAS [178].

### 6.2.3 Event Removal

The decays of the light scalar  $a$  to hadrons is modelled in PYTHIA and HERWIG through the hadronisation of the parton-level decays  $a \rightarrow qq$  and  $a \rightarrow gg$ , rather than through the definition of  $a$  dedicated decay table for the  $a$  describing the possible exclusive decays (for which no such predictions exist). Since the hadronisation is an event level process (i.e. energy, momentum and colour must be conserved at the event level), one can expect that in some cases the kinematics of the  $a$  hadronic decay products will not be accurate for the decay process. In particular, cases can occur in which the invariant mass of all stable particles associated with the  $a$  decay are inconsistent with the true mass of  $a$ , which is generated with a negligible natural width. As shown in Table 6.4, with around 40% being rejected for the 0.5 GeV PYTHIA sample, reducing to  $< 5\%$  for the 1.5 GeV and above. Any event with final state invariant mass inconsistency larger than 10% is removed. Those removed events lie mostly on low  $p_T$  or high- $\eta$  regions as shown in Figures 6.4, where most events will fail the event selection. The event removal process can affect the normalization by about 1%, which is far less than the applied normalization uncertainty coming from the difference between Pythia-Herwig.

Signal mass (GeV)		0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
Pythia (%)	$a \rightarrow gg$	54.3	83.1	92.9	96.2	94.7	93.3	92.0	91.1
	$a \rightarrow qq$	-	-	95.2	99.8	99.8	99.4	98.9	91.5
Herwig (%)	$a \rightarrow gg$	-	-	-	74.8	69.0	63.7	58.8	53.8
	$a \rightarrow qq$	-	-	-	81.7	77.5	73.2	68.7	75.9

Table 6.4: The fraction of remaining events after event removal.



### 6.3.1 $\ell\ell + \text{jet}$ Pre-Selection

The pre-selection is aiming at events with the final state: a Same-Flavor Opposite-Sign (SF-OS) lepton pair + jet, since the signal requires a leptonically decaying  $Z$  boson and a hadronically decaying  $a$ . For triggering, the lowest  $p_T$  single electron/muon item without a prescale is used, see in Table 6.5. The trigger objects are required to match the corresponding offline object, i.e.  $\Delta R_{(\text{trigger object, offline lepton})} < 0.1$ .

Period	$e$ Triggers
2015	Medium identification with $p_T > 24$ GeV at HLT, $p_T > 20$ GeV at L1 Medium identification with $p_T > 60$ GeV at HLT Loose identification with $p_T > 120$ GeV at HLT
2016-2018	Tight identification with $p_T > 26$ GeV at HLT, and Loose isolation Medium identification with $p_T > 60$ GeV at HLT Loose identification with $p_T > 140$ GeV at HLT

Period	$\mu$ Triggers
2015	Loose isolation with $p_T > 20$ GeV at HLT, $p_T > 15$ GeV at L1 $p_T > 40$ GeV at HLT
2016 A	Loose isolation with $p_T > 20$ GeV at HLT, $p_T > 15$ GeV at L1 Medium isolation with $p_T > 26$ GeV at HLT $p_T > 40$ GeV at HLT
2016 B-L, 2017, 2018	Loose isolation with $p_T > 20$ GeV at HLT, $p_T > 15$ GeV at L1 Medium isolation with $p_T > 26$ GeV at HLT $p_T > 50$ GeV at HLT

Table 6.5: Single electron and muon triggers for the Run 2 dataset in different periods.

The method of electron candidates reconstruction is described in Section 3.3.2. The Medium likelihood-based identification and the Loose isolation are required [90]. The selection requirements of electrons are summarized in Table 6.6. Note that  $|z_0 \sin \theta|$  is the track longitudinal impact parameter for the selected electron, and  $|d_0/\sigma(d_0)|$  is the significance of the track transverse impact parameter, where  $\sigma(d_0)$  is its estimated uncertainty and both are measured relative to the primary vertex.

Muons are reconstructed as discussed in Section 3.3.3. Selected muons have to be reconstructed within the region of  $|\eta| < 2.7$  covered by the muon spectrometer, and

Sub-leading electron	Leading electron
Identification WP: Medium	
Isolation WP: Loose_varRad	
$ \eta  < 1.37$ and $1.52 <  \eta  < 2.47$	
$p_T > 18 \text{ GeV}$	$p_T > 27 \text{ GeV}$
$ z_0 \sin \theta  < 0.5 \text{ mm},  d_0/\sigma(d_0)  < 5$	

Table 6.6: Electron kinematic, identification, and isolation requirements.

Sub-leading muon	Leading muon
Identification WP: Medium	
Isolation WP: Loose_varRad	
$ \eta  < 2.7$	
$p_T > 18 \text{ GeV}$	$p_T > 27 \text{ GeV}$
(If ID track presents) $ z_0 \sin \theta  < 0.5 \text{ mm},  d_0/\sigma(d_0)  < 3$	

Table 6.7: Summary of muon selection requirements.

pass the Medium quality and the Loose isolation working points [73]. The Medium quality WP accepts only CB and IO muons in the range  $|\eta| < 2.5$ , requires at least two precision stations. For ME muons in the range  $2.5 < |\eta| < 2.7$ , at least three precision stations are required. The  $q/p$  compatibility (equation (3.3)) of CB and IO muon is required to be less than 7. The selection requirements of muons are summarized in Table 6.7.

Since the target resonance has low mass ( $< 4 \text{ GeV}$ ), it is highly boosted. The products of  $a$  decay spread in a cone with the width  $\Delta R \approx 2m_a/p_{T_a}$ , giving the cone size of  $\Delta R < 0.4$  for the lowest jet  $p_T$  of 20 GeV. Therefore, the resonance  $a$  is reconstructed as a single anti- $k_T$  jet with a radius parameter of 0.4 [94], built from particle flow objects, as described in Section 3.3.4. Jets are required to pass the ‘‘cleansing’’ selection to reject jets from secondary collisions [179]. The selection requirements of jets are summarized in Table 6.8.

An Overlap Removal (OLR) procedure [180] is applied on reconstructed electrons, muons and jets to avoid counting the same energy deposits or same tracks multiple times. The OLR procedure sequentially compares two types of objects, removing one or another based on a pre-defined priority. For example, the electron sharing a track with a muon is removed, the jet within  $\Delta R = 0.2$  of an electron is removed,

Jets
AntiKt4EMPflowJets
$ \eta  < 2.5$
$p_T^{\text{leading jet}} > 20 \text{ GeV}$
JVT $> 0.59$ for jets with $p_T < 60 \text{ GeV}$ and $ \eta  < 2.4$

Table 6.8: Summary of selection requirements for small- $R$  jets.

etc. At least two SF-OS leptons are required to pass the OLR. The requirement for the invariant mass of this lepton pair is  $|m_{ll} - m_Z| < 10 \text{ GeV}$ . If multiple SF-OS pairs meet the requirement, the selection chooses the pair closest to the  $Z$  boson mass. The invariant mass of the three body system (two leptons and a jet) is required to pass  $50 \text{ GeV} < m_{llj} < 180 \text{ GeV}$ . If multiple jets pass the selection in an event, the largest  $p_T$  jet is chosen.

The  $p_T$  distributions of the reconstructed  $Z$  boson (2 leptons) and the selected jet are shown in Figure 6.5. The  $p_T$  and the invariant mass of the reconstructed Higgs boson (2 leptons + jet) are shown in Figure 6.6. The signal peaks of  $m_H$  distribution (Figure 6.6 (b)) are not at 125 GeV, because the PFlow jets calibration is designed for QCD jets [96], rather than jets produced from BSM light resonances.

### 6.3.2 Track Selection

In order to distinguish signal resonances from background jets, a track-based discriminant is used since the inner tracker provides better angular and  $p_T$  resolution than the calorimeter. The tracks are selected by the ghost-association [100], as discussed in Section 3.3.4. In order to suppress the large contamination coming from pile-up fake tracks [50], the Loose track quality and Loose track-to-vertex association (TTVA) working points are applied to these ghost-associated tracks. The selection requirements of ghost-associated tracks are summarized in Table 6.9.

As discussed above, full event selection is summarised in Table 6.10, and cutflows for a 1.5 GeV signal and different backgrounds are shown in Table 6.11.

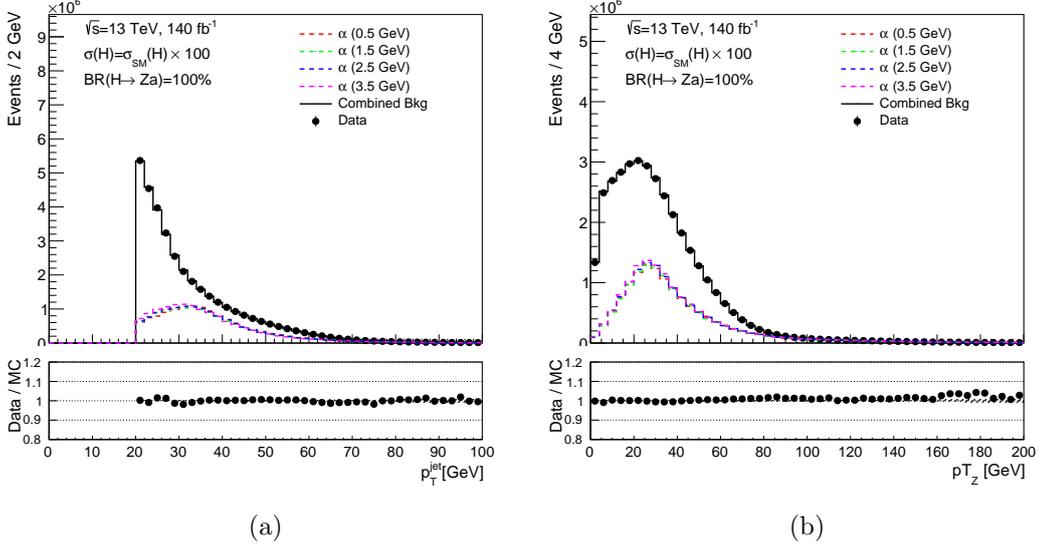


Figure 6.5: (a)  $p_T$  of the leading jet and (b)  $p_T$  of the reconstructed  $Z$  boson after the full pre-selection for (dots) data, (dash lines) signal and (black line) background. The lower panel shows the ratio of data to background. Only statistical uncertainties are considered. The background has been reweighted as per Section 6.4.2.

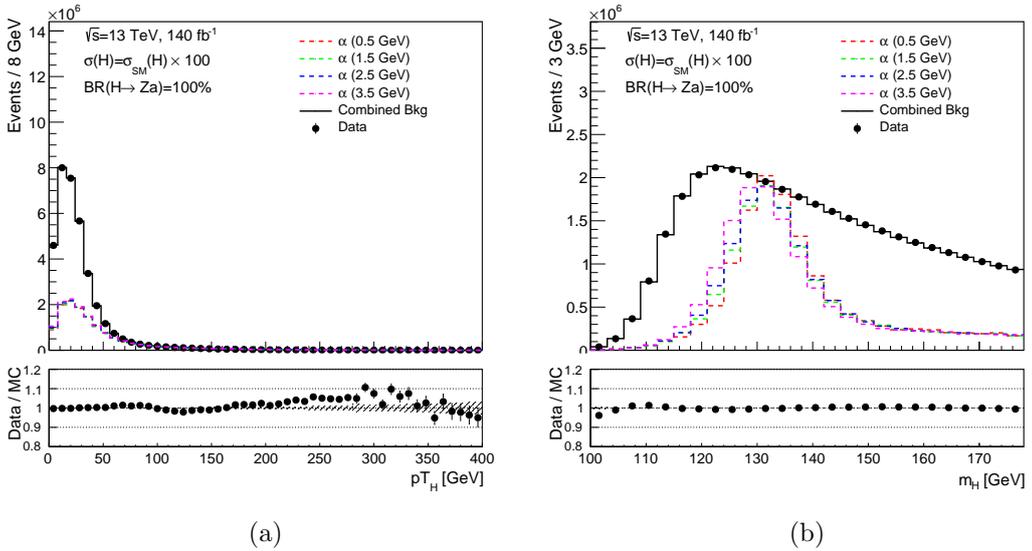


Figure 6.6: (a)  $p_T$  and (b) invariant mass distributions of the reconstructed Higgs boson after the full pre-selection for (dots) data, (dash lines) signal and (black line) background. The lower panel shows the ratio of data to background. Only statistical uncertainties are considered. The background has been reweighted as per Section 6.4.2.

Ghost-associated tracks
Track Quality WP: Loose
TTVA WP: Loose
$ \eta  < 2.5$
$p_T > 500 \text{ MeV}$
silicon hits $\geq 7$ , shared Pixel hit $\leq 1$ or shared SCT hits $\leq 2$
Pixel hole $\leq 1$ , silicon hole $\leq 2$
$ d_0  < 2 \text{ mm}$ and $ z_0 \sin \theta  < 3 \text{ mm}$

Table 6.9: Summary of selection requirements for ghost-associated tracks to jet.

Cut	Description
HDBS3	Single lepton triggers, $> 2$ leptons $p_T > 10 \text{ GeV}$ , $m_{ll} : [50-106 \text{ GeV}]$
Triggers	Leading lepton $p_T > 27 \text{ GeV}$
Leptons	$\geq 2$ leptons
Z bosons	SF-OS lepton pair, $ m_{ll} - m_Z  < 10 \text{ GeV}$ & $p_T^{\text{lead}} > 27 \text{ GeV}$
Trigger Matching	At least one of the Z leptons triggers the event
Jet selection	Leading jet with $p_T > 20 \text{ GeV}$
Invariant mass	$50 \text{ GeV} < m_{llj} < 180 \text{ GeV}$
$p_T$ requirement	Leptons $p_T > 18 \text{ GeV}$
Ghost tracks	$> 2$ ghost associated tracks

Table 6.10: Full pre-selection of events.

Cut	1.5 GeV $a$	$Z \rightarrow \mu\mu$	$Z \rightarrow ee$	$t\bar{t}$	ZZ	WZ
HDBS3	100%	47%	40%	43%	50%	44%
Triggers	54%	47%	40%	42%	50%	44%
Leptons	35%	40%	27%	12%	37%	30%
Z bosons	30%	35%	24%	2%	32%	26%
Trigger Matching	30%	35%	24%	2.0%	32%	26%
Jet selection	30%	35%	24%	2.0%	32%	25%
Invariant mass	24%	11%	8.7%	1.9%	27%	21%
$p_T$ requirement	22%	10%	8.2%	1.9%	26%	21%
Ghost tracks	21%	10%	8.0%	1.7%	25%	19%

Table 6.11: Cutflow for 1.5 GeV signal and backgrounds. No selection is applied to the signal samples at derivation level.

Quantity	Description
$n_{\text{tracks}}$	Number of Ghost-Associated tracks
$p_{\text{T}}^{\text{lead track}}/p_{\text{T}}^{\text{tracks}}$	Fraction of the leading track transverse momentum to the total
$\Delta R^{\text{lead track, calo jet}}$	$\Delta R$ between the leading track and the reconstructed jet
$\tau_2$	NSubJettiness 2 [181]
$U_1(0.7)$	Modified energy correlation function ${}_1e_2^{(0.7)}$ [182]
$M_2(0.3)$	Modified energy correlation functions ${}_1e_3^{(0.3)}/{}_1e_2^{(0.3)}$ [182]
angularity(2)	Angularity 2 [183]

Table 6.12: Quantities used for signal to background separation. A ghost association of the tracks to the calorimeter jets is applied.

7 dimensionless variables are selected to distinguish the signal from the background, including the number of ghost-associated tracks and 6 jet-substructure variables. These variables are summarised in Table 6.12 and their correlation with the final states invariant mass  $m_H$  had been studied in [50], showing all of them have correlation less than 20%.

$n_{\text{tracks}}$  is the number of Ghost-Associated tracks.  $p_{\text{T,leading track}}/p_{\text{T,tracks}}$  is the fraction of the highest track transverse momentum to the total transverse momentum.  $\Delta R^{\text{lead track, calo jet}}$  is the distance between the leading track and the jet in  $\eta - \phi$  plane. The NSubJettiness 2 variable  $\tau_2$  is designed to identify boosted hadronically-decaying objects [181]. It describes the degree of the jet can be composed of 2 subjets. It is defined by

$$\tau_2 = \sum_k p_{T,k} \min \{ \Delta R_{1,k}, \Delta R_{2,k} \} / \sum_k p_{T,k} R_0, \quad (6.1)$$

where the sum runs over ghost-associated tracks, and  $R_0$  is the jet radius parameter 0.4.  $U_1(0.7)$  and  $M_2(0.3)$  are generalized energy correlation functions [182], defined as

$$U_1(0.7) = {}_1e_2^{(0.7)} \quad (6.2)$$

and

$$M_2(0.3) = \frac{{}_1e_3^{(0.3)}}{{}_1e_2^{(0.3)}} \quad (6.3)$$

respectively, where  ${}_\nu e_n^{(\beta)}$  are the  $n$ -point energy correlation functions, describing the

angular separation ( $\Delta R$ ) between particles of the jet [182]:

$${}_1e_2^{(\beta)} = \sum_{1 \leq i < j \leq n_J} z_i z_j \theta_{ij}^\beta, \quad (6.4)$$

and

$${}_1e_3^{(\beta)} = \sum_{1 \leq i < j < k \leq n_J} z_i z_j z_k \min \left\{ \theta_{ij}^\beta, \theta_{ik}^\beta, \theta_{jk}^\beta \right\}, \quad (6.5)$$

where  $z_i$  is the generic energy fraction of the  $i$ -th particle of the jet, calculated by

$$z_i = \frac{p_{Ti}}{\sum_{j \in \text{jet}} p_{Tj}}, \quad (6.6)$$

and  $\theta_{ij}$  is the generic angular variable between the  $i$ -th and  $j$ -th particle of the jet, given by

$$\theta_{ij}^2 = R_{ij}^2 = (\phi_i - \phi_j)^2 + (y_i - y_j)^2, \quad (6.7)$$

where  $\phi$  and  $y$  are azimuthal angle and rapidity of particles. The last one is an angularity variable, which gives different weights to particles near the center and those at the edge of jet cone [183]:

$$\text{angularity}(2) = \sum_k p_{T,k} \sin^2\left(\frac{\pi\theta_k}{2R}\right) \left(1 - \cos\left(\frac{\pi\theta_k}{2R}\right)\right)^{-1}, \quad (6.8)$$

where  $\theta_k$  is the angle of the particles relative to the jet momentum axis. The distributions of these discriminate variables for data, combined background and signals are shown in Figures 6.7 to 6.10

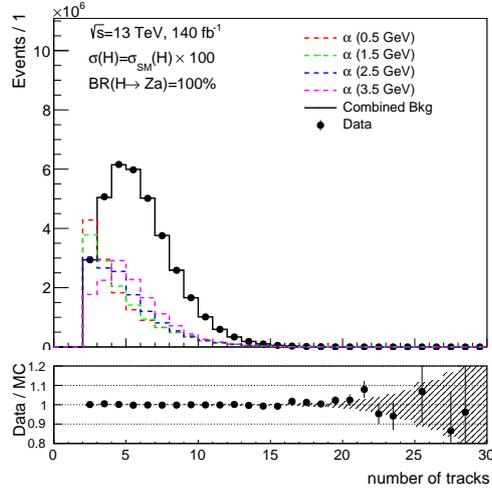


Figure 6.7: NN input distributions based on the number of Ghost-Associated tracks after the full pre-selection for (dots) data, (dash lines) signal and (black line) background. The lower panel shows the ratio of data to background. Only statistical uncertainties are considered. The background has been reweighted as per Section 6.4.2.

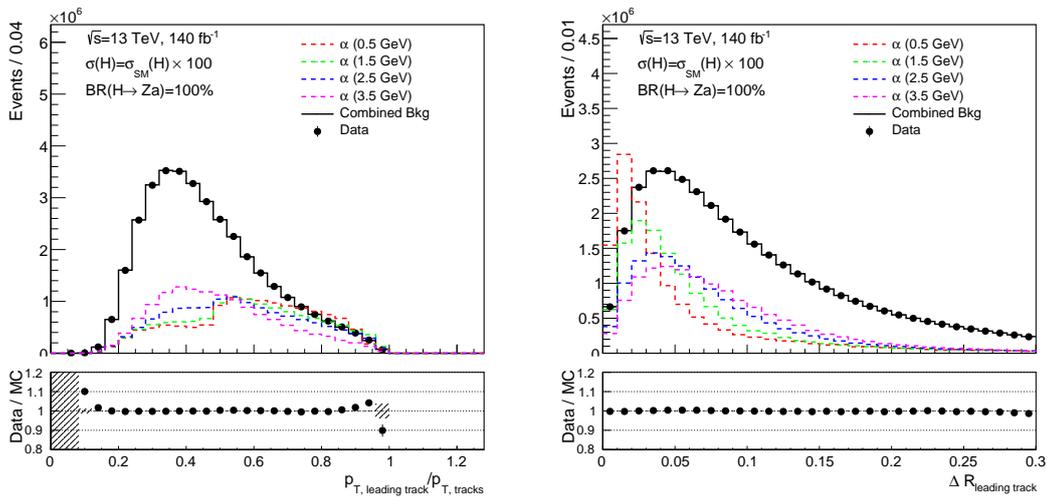


Figure 6.8: NN input distributions based on the leading track Ghost-Associated to the jet after the full pre-selection for (dots) data, (dash lines) signal and (black line) background. The lower panel shows the ratio of data to background. Only statistical uncertainties are considered. The background has been reweighted as per Section 6.4.2.

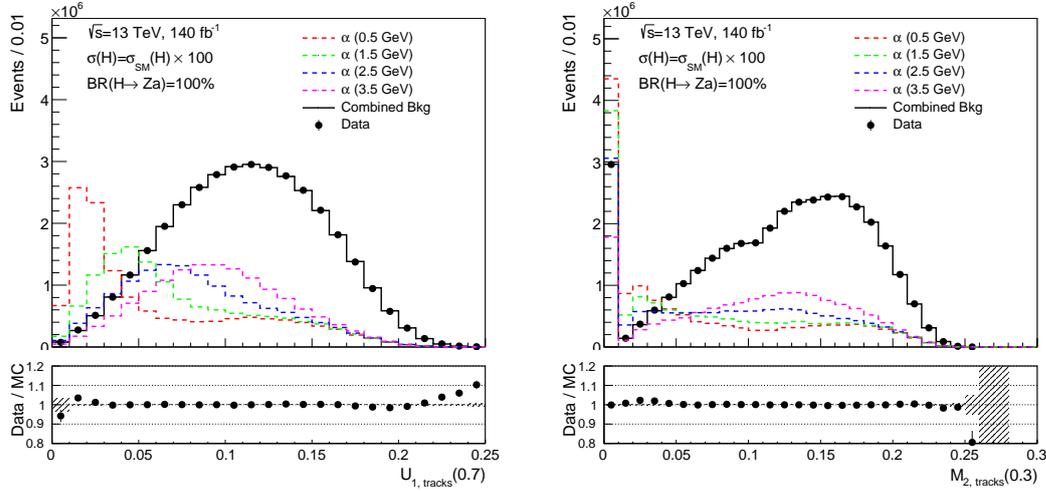


Figure 6.9: NN input distributions based on the modified correlation functions after the full pre-selection for (dots) data, (dash lines) signal and (black line) background. The lower panel shows the ratio of data to background. Only statistical uncertainties are considered. The background has been reweighted as per Section 6.4.2.

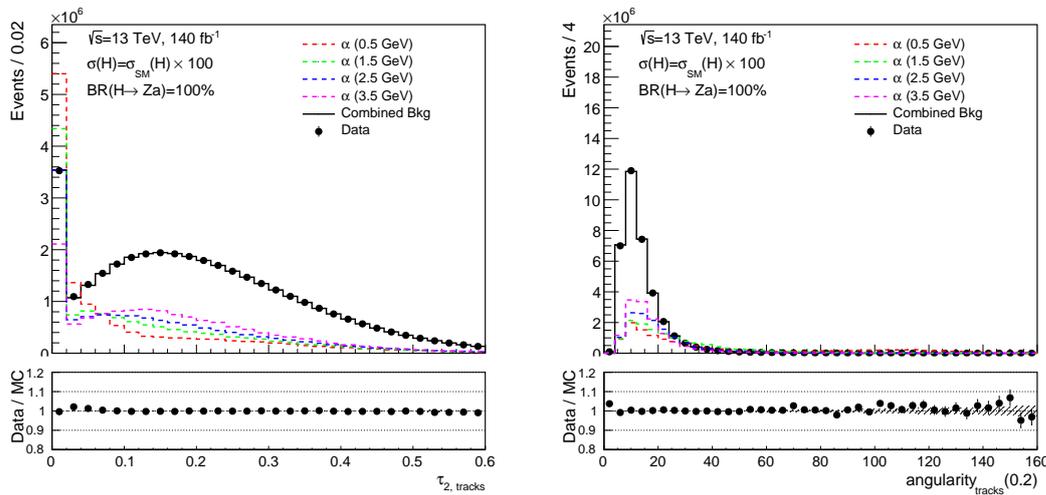


Figure 6.10: NN input distributions based on the NSubJettiness and angularity variables after the full pre-selection for (dots) data, (dash lines) signal and (black line) background. The lower panel shows the ratio of data to background. Only statistical uncertainties are considered. The background has been reweighted as per Section 6.4.2.  $\tau_2$  is defined to be 0 for events less than three tracks.

### 6.3.3 Track-based Neural Networks

A Feed-Forward classification NN is used to distinguish jets produced by hadronic  $a$  decays from the background, using the input variables listed in Table 6.12. Typically the signal classification is done with Parameterized NNs (PNN) applying the  $a$ 's truth mass as an additional NN parameter. But this will lead to multiple outputs based on different signal mass hypotheses, which makes the analysis more complicated. One strategy is using the mass of reconstructed jet as a NN training variable. However, as distributions of reconstructed jet mass for different signal masses show in Figure 6.11, the calorimeter resolution is not good enough in such low mass region. The reconstructed jet mass cannot provide effective information to the classification NN. Therefore, a regression NN is trained to predict the  $a$  mass and feed this information to the classification NN.

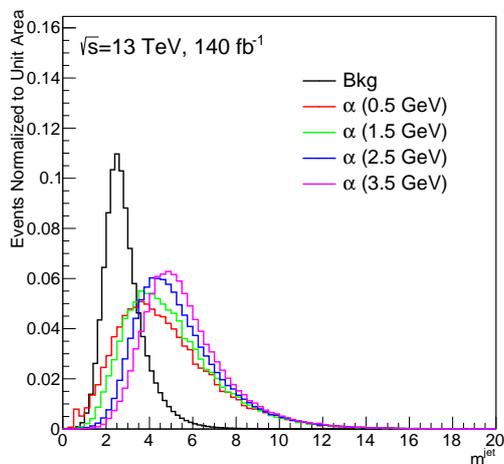


Figure 6.11: Distributions of reconstructed jet mass for background and different signal samples after the pre-selection. The area of each distribution is normalized to 1.

#### Regression NN

The regression NN is trained on 0.5, 0.75, 1, 1.5, 2, 2.5, 3, 3.5 and 4 GeV  $a$  signal samples. Its aim is to estimate  $m_a$  based on jet substructure variables and number of tracks (Table 6.12). Only events with  $120 < m_{lj} < 145$  GeV are used for the

training. The hyperparameters of the regression NN are optimized based on the following strategies:

- Recommendations for the the kind of NN,
- Self optimizing algorithms,
- Bayesian optimization (Tree-structured Parzen Estimator (TPE) [136], discussed in Section 5.3),
- Choice based on analysis criteria.

The target of the Bayesian optimization for the this NN is to find the minimum of the loss function. In addition, the background-data agreement in the region of  $m_{llj} < 120$  GeV and  $m_{llj} > 145$  GeV, for the regression NN output is monitored by the  $\chi^2$ -test for every trial, and trials with excessive disagreement are rejected. The optimal hyperparameters are shown in Table 6.13. The output of the regression NN is shown in Figure 6.12.

NN Hyperparameter	Optimization Strategy	Value
Layers	Bayesian Optimization	4
Neurons per Layer	Bayesian Optimization	35
Hidden Layer Activation Function	Recommendation for Feed-Forward NNs	ReLU
Output Layer Activation Function	Recommendation for regression tasks	Linear
Loss Function	Choice based on analysis criteria	Huber
Optimizer	Recommendation for NNs	Adam
R2 Regularization	Bayesian Optimization	$10^{-11}$
Batch Size	Bayesian Optimization	160
Epochs	Best Validation Loss	50

Table 6.13: Optimal hyperparameters and their optimization strategies for the Regression NN

The regression output variable is then sent together with other input variables in Table 6.12, to a Classification NN. Therefore, rather than using PNNs for different mass points, the classification NN obtains information from regression mass to understand that it is searching for a spectrum of signals.

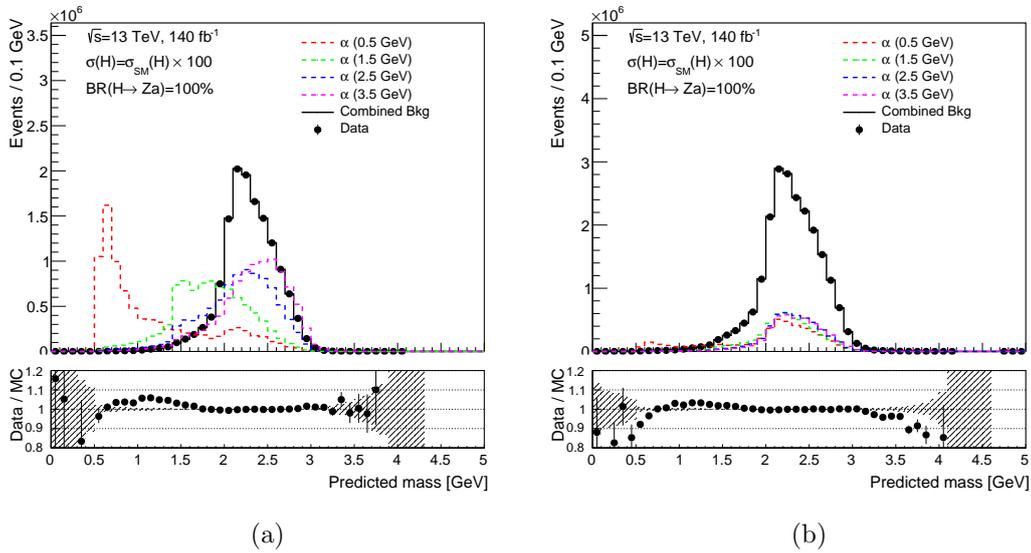


Figure 6.12: Output of the regression NN, for data, background and four signal hypotheses in (a) the region  $120 < m_{llj} < 140$  GeV, and (b) the region  $m_{llj} < 120$  GeV and  $m_{llj} > 140$  GeV. Events are required to pass the full pre-selection. The background normalisation is set equal to that of the data, and the signal normalisations assume the SM Higgs production cross section and  $\text{BR}(H \rightarrow Za) = 100\%$  with the signal normalisation is scaled up by a factor of 100. Only statistical uncertainties are considered. The background has been reweighted as per Section 6.4.2

## Classification NN

The Classification NN is designed to discriminate signal events from background. It is trained on 0.5, 0.75, 1, 1.5, 2, 2.5, 3, 3.5, 4 GeV  $a$  signal samples, and combined background samples. Only events with  $120 < m_{llj} < 135$  GeV are used for the training. The input variables include 7 variables described in Table 6.12 and the output of the regression NN. The choice of the best classification result is made based on the background-data agreement for the final state invariant mass, as well as the maximum significance ( $S/\sqrt{B}$ ), where  $S$  and  $B$  are the event numbers of the signal and the combined background respectively. The hyperparameters of the best classification NN are shown in Table 6.14. The output of the classification NN is shown in Figure 6.13, and the efficiency and significance curves in Figure 6.14.

NN Hyperparameter	Optimization Strategy	Value
Layers	Bayesian Optimization	3
Neurons per Layer	Bayesian Optimization	10
Hidden Layer Activation Function	Recommendation for Feed-Forward NNs	ReLU
Output Layer Activation Function	Recommendation for classification tasks	sigmoid
Loss Function	Recommendation for classification tasks	Binary Crossentropy
Optimizer	Recommendation for NNs	Adam
R2 Regularization	Bayesian Optimization	$10^{-10}$
Batch Size	Bayesian Optimization	100
Epochs	Best Validation Loss	29

Table 6.14: Optimal hyperparameters and their optimization strategies for the Classification NN

The classification NN cut value is decided to be 0.93 yielding a 99.3% background rejection, while the signal efficiency is more than 40% for the 0.5 GeV mass. This cut gives a significance quite close to the maximum for the low masses without degrading too much for the high masses.

The full event selection is summarized in Table 6.15. The final state invariant mass  $m_{llj}$  after the full selection is shown in Figure 6.15. The signal is normalized assuming  $\text{BR}(H \rightarrow Za) = 100\%$ .

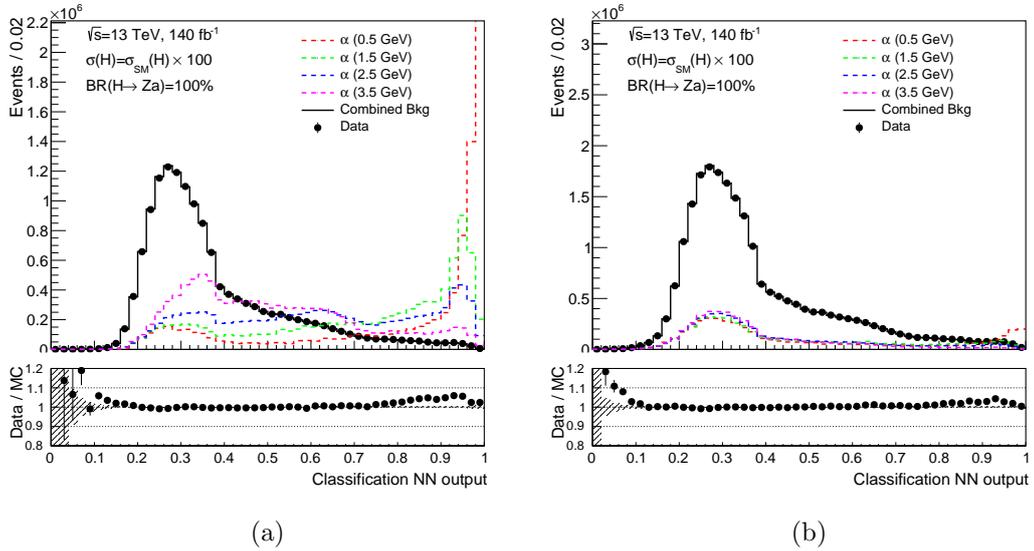


Figure 6.13: Output of the classification NN. Events are required to pass the full event selection. The left plot includes the  $120 < m_{llj} < 140$  GeV requirement. The plot on the right includes the events in the regions  $m_{llj} < 120$  GeV and  $m_{llj} > 140$  GeV. The background has been reweighted and normalized to match the data. The signal normalisations assume  $\text{BR}(H \rightarrow Za) = 100\%$  and scale up the SM Higgs boson production cross section by a factor of 100 and. Only statistical uncertainties are considered.

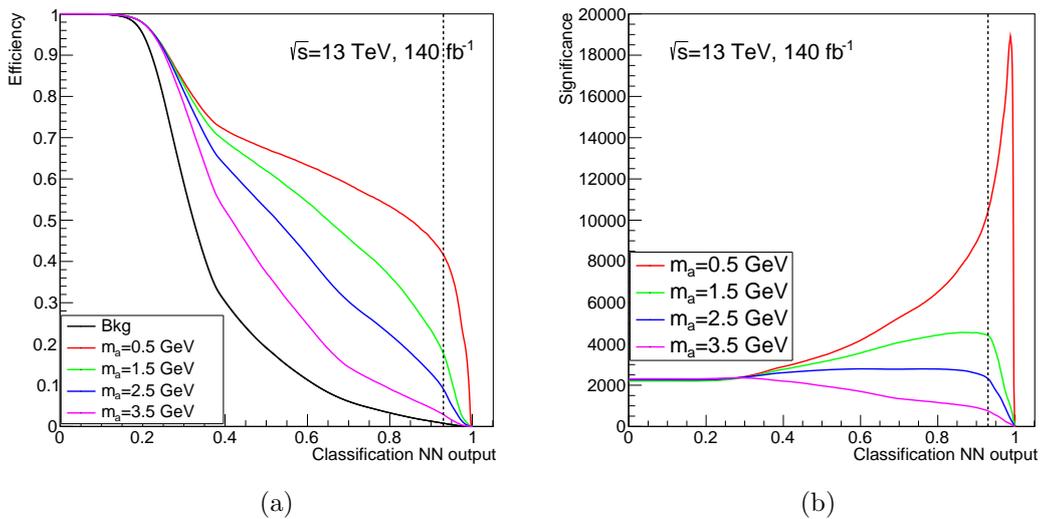
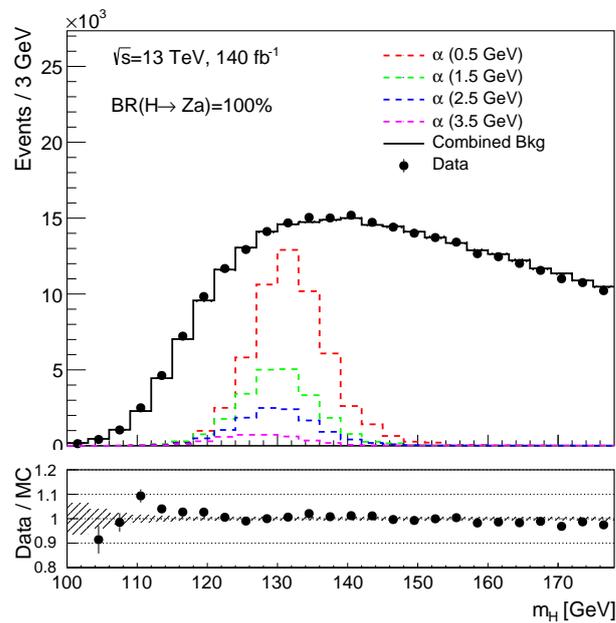


Figure 6.14: (a) Background and signal NN efficiencies and (b) signal significance ( $S/\sqrt{B}$ ) as functions of the Classification NN output cut. The dashed line is the NN output cut on 0.93

Cut	Description
Triggers	Leading lepton $p_T > 27$ GeV
Leptons	$\geq 2$ leptons with $p_T > 18$ GeV
Z boson	SF-OS lepton pair, $ m_{ll} - m_Z  < 10$ GeV & $p_T^{\text{lead}} > 27$ GeV
Select $a$ -candidate as jet ( $p_T^{\text{leading jet}} > 20$ GeV), for which $m_{lj} < 180$ GeV	
Tracks	$> 2$ ghost associated tracks
Classification NN	NN output $> 0.93$

Table 6.15: Summary of full event selection.

Figure 6.15: The final state invariant mass  $m_H$  after the full event selection.

### 6.3.4 Regression NN optimization

Besides of using Huber loss function, MSE, MSLE and MAPE loss functions are also tried for the regression NN training. The strategies of hyperparameter optimization are consistent with the nominal. In order to provide effective information to the classification NN, the regression output is expected to be able to distinguish different signal masses. The background-data agreement in the mass range of  $m_{lj} < 120$  GeV and  $m_{lj} > 140$  GeV is also monitored for each trial.

For each type of loss function, multiple trials have been trained. Figure 6.16, 6.17 and 6.18 show the regression NN outputs after minimizing MSE, MSLE and MAPE

loss functions, respectively. By comparing these with the output of the regression NN using Huber Loss (Figure 6.12), it can be seen that the Huber loss gives the best regression performance in this case.

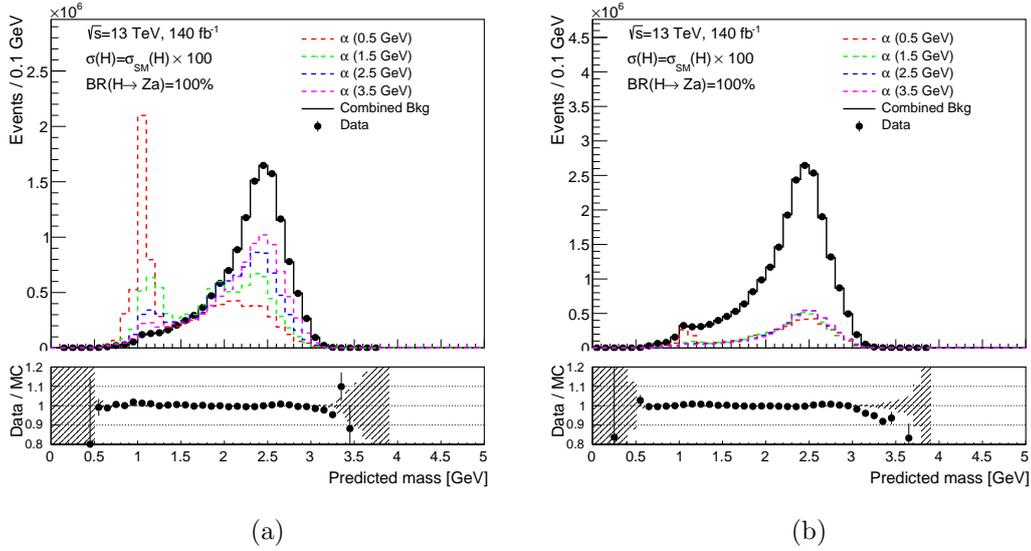


Figure 6.16: Output of the regression NN with the Mean Square Error (MSE) in (a) the range of  $120 < m_{Uj} < 140$  GeV, and (b) the range of  $m_{Uj} < 120$  GeV and  $m_{Uj} > 140$  GeV. The full pre-selection is required.

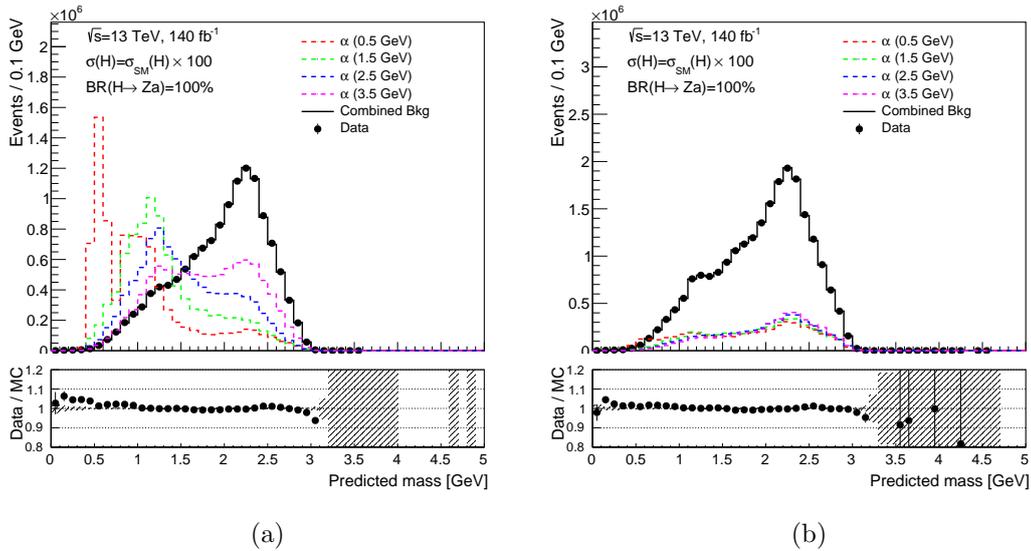


Figure 6.17: Output of the regression NN with the Mean Squared Logarithmic Error (MSLE) (a) the range of  $120 < m_{Uj} < 140$  GeV, and (b) the range of  $m_{Uj} < 120$  GeV and  $m_{Uj} > 140$  GeV. The full pre-selection is required.

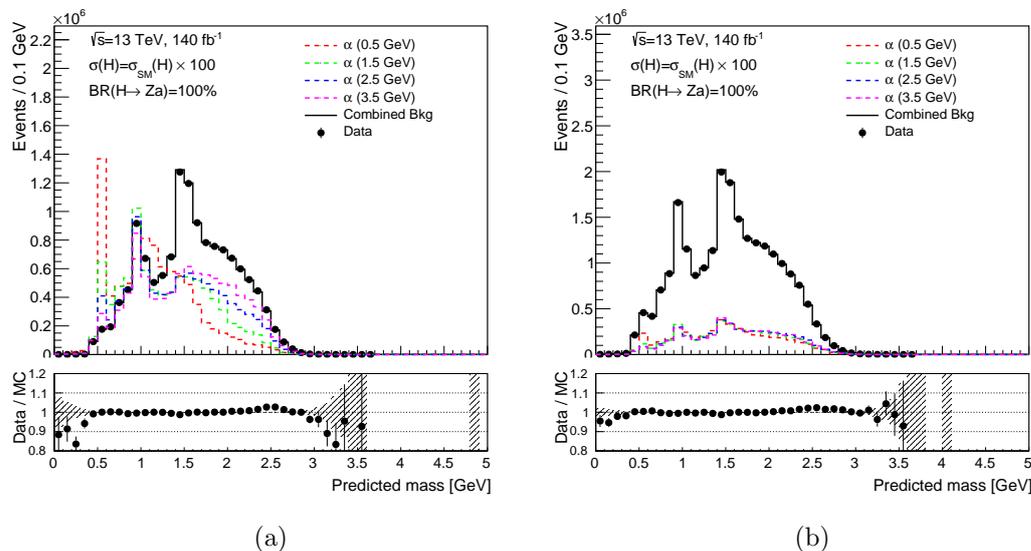


Figure 6.18: Output of the regression NN with the Mean Absolute Percentage Error (MAPE)(a) the range of  $120 < m_{llj} < 140$  GeV, and (b) the range of  $m_{llj} < 120$  GeV and  $m_{llj} > 140$  GeV. The full pre-selection is required

### 6.3.5 Alternative Classification NNs

In order to check the robustness of the classification approach in Section 6.3.3, alternative approaches of classification NN training also had been tried. Instead of using the Regression NN output as one of the input variables, the first alternative approach is training the Classification NN without Regression NN result. To improve the sensitivity and accuracy of Classification NN for heavier signals, the second alternative approach is giving heavier input signal samples larger weights. The weights for different signal samples can be seen in Table 6.16. The third approach is training the Classification NN with 4 GeV signal samples only. The output of these NNs are shown in Figure 6.19, and their significance curves in Figure 6.20. By comparing the signal significance between the nominal and alternative NNs (Figure 6.13 and 6.14), no improvement is observed.

Signal mass (GeV)	0.5	0.75	1.0	1.5	2.0	2.5	3.0	3.5	4.0
Weight	1	1	2	2	2	2	4	4	4

Table 6.16: The weights given to different signal samples

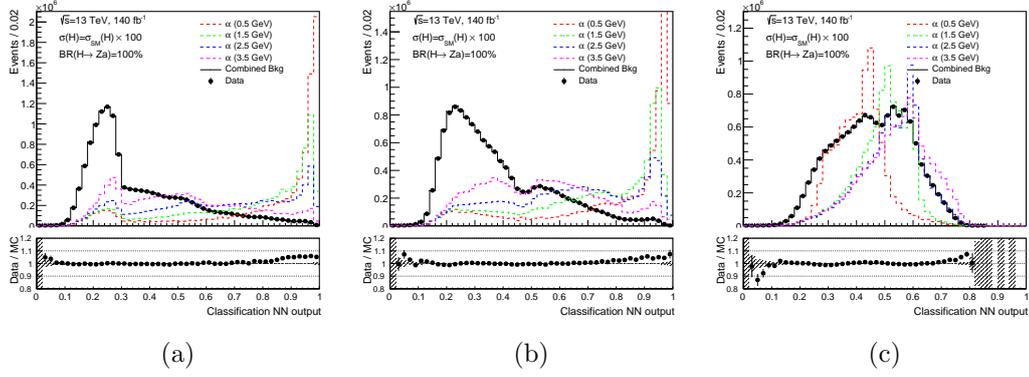


Figure 6.19: Output of the classification NN in different approaches. (a) Training without Regression mass. (b) Training with weighted signal samples. (c) Training with 4 GeV signal samples only. Events are required to pass the full event selection. Only events with  $120 < m_{Uj} < 140$  GeV are shown. The background has been reweighted and normalized to match the data. The signal normalisations assume  $\text{BR}(H \rightarrow Za) = 100\%$  and scale up the SM Higgs boson production cross section by a factor of 100 and. Only statistical uncertainties are considered.

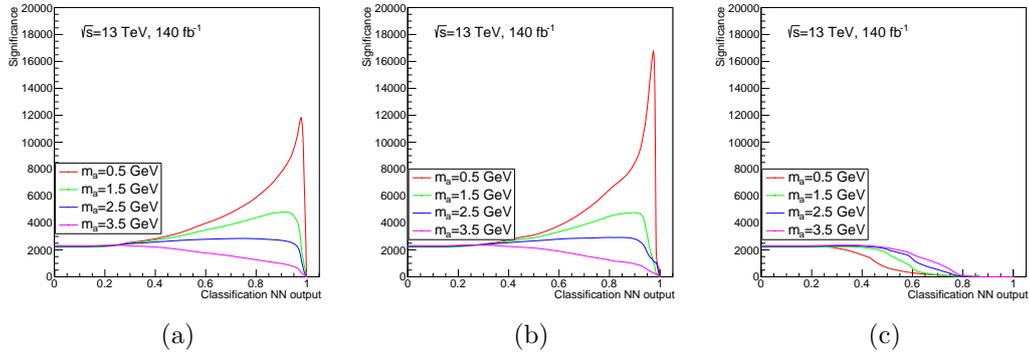


Figure 6.20: Various signals significance as a function of the Classification NN output cut in different approaches. (a) Training without Regression mass. (b) Training with weighted signal samples. (c) Training with 4 GeV signal samples only.

## 6.4 Signal and Background Modelling

### 6.4.1 Signal Modelling

The selection efficiency for the signal is derived from MC samples. It is used to scale the expected Higgs production yield. The MC samples are generated using ggF production mode, but the cross section used in the normalization is for the full inclusive Higgs production, including ggF, VBF,  $ZH$ ,  $WH$ ,  $b\bar{b}H$ ,  $t\bar{t}H$  and  $tH$  associated production. A systematic uncertainty is accounted for the difference. The expected signal yield is given by

$$N = L \times \sigma \times \text{BR}(H \rightarrow Za) \times \text{BR}(Z \rightarrow l^+l^-) \times \text{BR}(a \rightarrow gg/qq) \times \text{selection efficiency}, \quad (6.9)$$

where  $L = 140 \text{ fb}^{-1}$  is the total luminosity, and  $\sigma = 55.7 \text{ pb}$  is the total SM Higgs boson production cross section.  $\text{BR}(Z \rightarrow l^+l^-)$  is taken as 10.1%.  $a \rightarrow gg/qq$  BRs are given in Table 6.3. The selection efficiencies and the expected signal yields are shown in Table 6.17, assuming  $\text{BR}(H \rightarrow Za) = 10\%$ .

The final state invariant mass  $m_{lj}$  after applying the Classification NN cut for each signal mass hypothesis is fitted with a gaussian function. The MC histograms and the fits for the different mass points and decay modes are shown in Figure 6.21 and 6.22. The gaussian fit parameters, mean ( $\mu$ ) and sigma ( $\sigma$ ), for each mass point are shown in Figure 6.23.

After the classification NN cut, only 8-16% of the events lie out of the 120-140 GeV region. For the quark decays of the 4 GeV signal this number is higher (38%) because of the low mean value and the long tail.

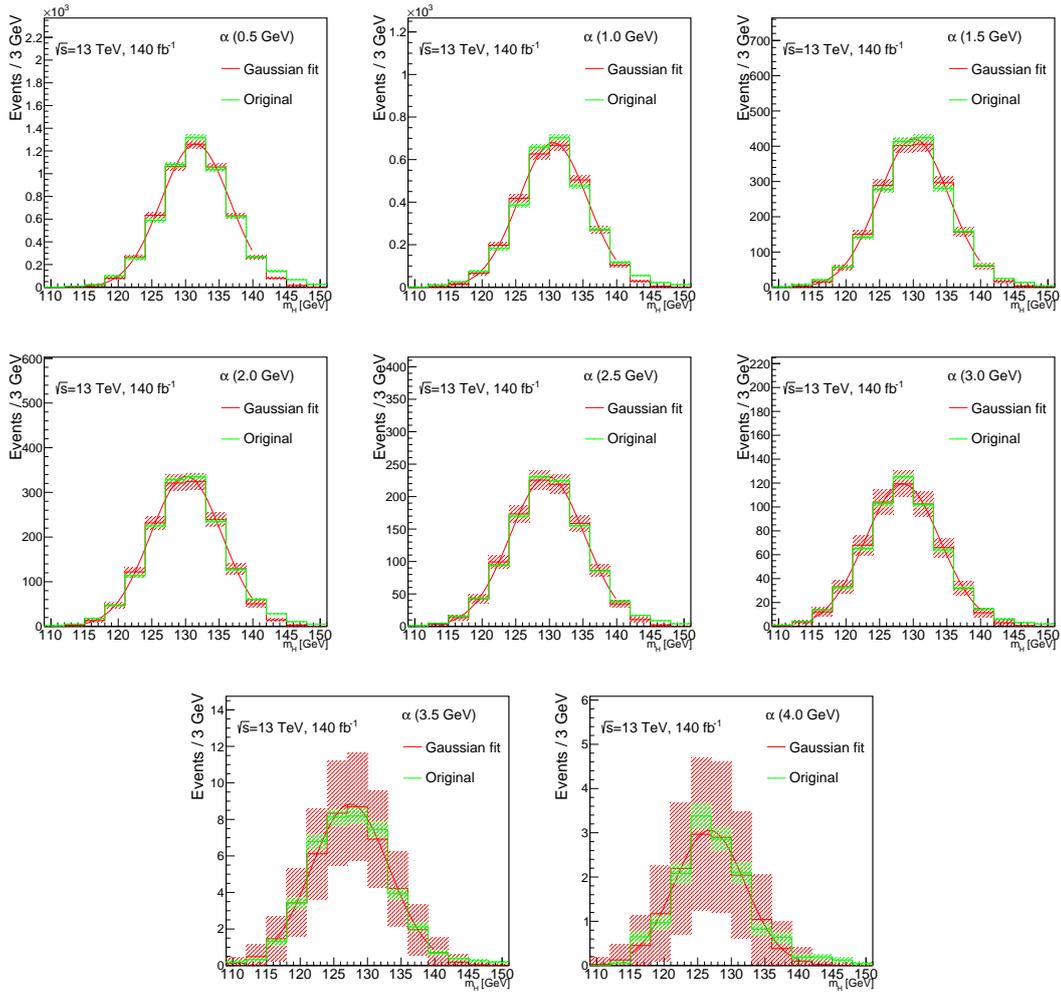


Figure 6.21: The  $m_{lj}$  distributions after the full event selection (both pre-selection and classification NN cut) for 0.5 - 4.0 GeV signals with the  $a \rightarrow gg$  decay mode. Green histograms are derived from MC samples, and the red histograms are generated from the gaussian fits. Shadows show the statistical uncertainties.

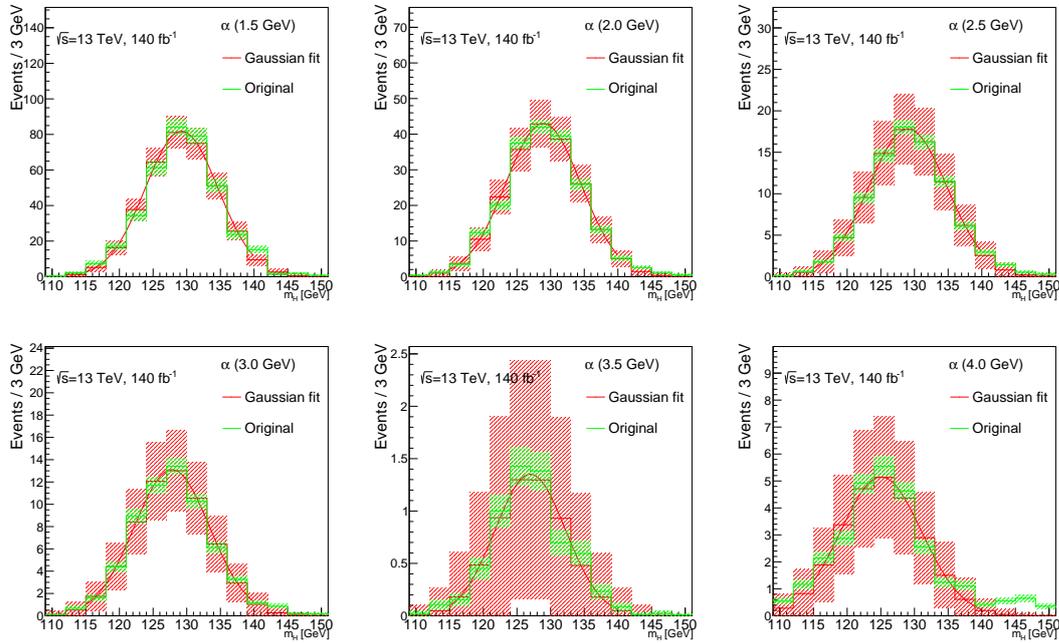


Figure 6.22: The  $m_{llj}$  distributions after the full event selection (both pre-selection and classification NN cut) for 1.5 - 4.0 GeV signals with the  $a \rightarrow qq$  decay mode. Green histograms are derived from MC samples, and the red histograms are generated from the gaussian fits. Shadows show the statistical uncertainties.

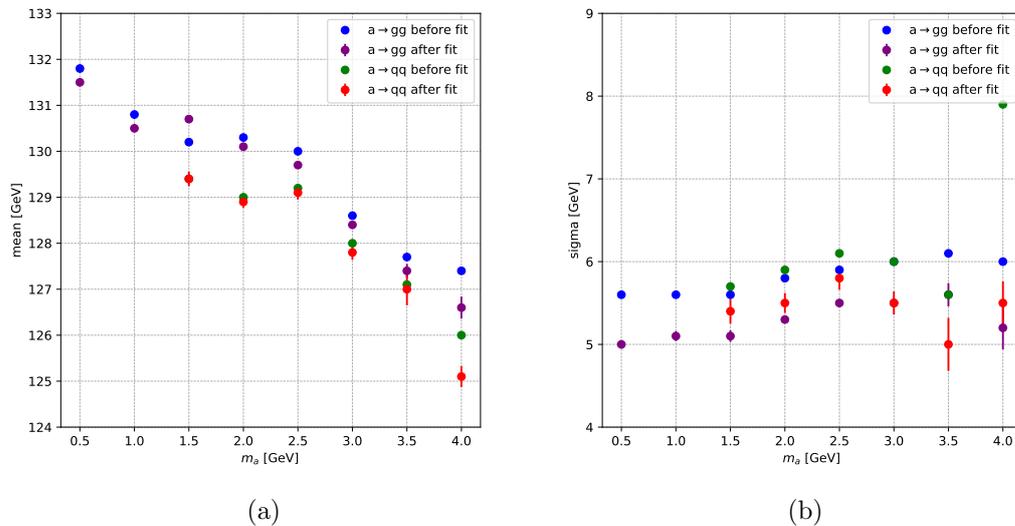


Figure 6.23: Gaussian fit parameters: (a) mean and (b) sigma of different signals, for both  $a \rightarrow gg$  and  $a \rightarrow q\bar{q}$  decays.

$m_a$ (GeV)	$a$ decay channel	Generated events (k)	SR events (k)	SR yield (k)	Selection efficiency (%)
0.5	$gg$	175	13.9	5.6	7.5
1.0	$gg$	249	11	3.0	4.3
1.5	$gg$	246	8	1.9	3.1
	$qq$	43	1.6	0.38	3.6
2.0	$gg$	832	20	1.6	2.7
	$qq$	106	2.8	0.2	2.5
2.5	$gg$	1731	29	1.1	1.6
	$qq$	141	2.4	0.09	1.7
3.0	$gg$	1675	15	0.57	0.84
	$qq$	162	1.7	0.06	1.03
3.5	$gg$	385	2	0.045	0.51
	$qq$	40	0.30	0.006	0.73
4.0	$gg$	190	0.65	0.015	0.32
	$qq$	1839	1.5	0.034	0.08

Table 6.17: Selection efficiencies and expected yields for each signal sample in the  $m_{uj}$  range 120-140 GeV. Assuming  $\text{BR}(H \rightarrow Za)=10\%$ . The decay mode  $a \rightarrow qq$  stands for  $a \rightarrow s\bar{s}$  for 0.5 - 3.5 GeV signals, and  $a \rightarrow c\bar{c}$  for 4.0 GeV signal.

## 6.4.2 Background Modelling

Before being sent to the classification NN, the simulated background events are reweighted to improve the NN input variable modelling (listed in Table 6.12). More kinematic variables, such as  $p_{T_H}$ ,  $p_{T_Z}$  and  $p_T^{\text{jet}}$ , are also added in the reweighting procedure to improve the its robustness. To reduce the bias caused by the signals in data, the events within the range  $120 < m_{uj} < 140$  GeV are excluded from the reweighting.

The reweighting is done by training a NN to estimate the multidimensional density ratio of the background to the data probability density function (PDF). The background and data PDF are denoted as  $f_0(X)$  and  $f_1(X)$  respectively, where  $X$  is the multidimensional vector containing NN input variables. Suppose  $r(X) = \frac{f_1(X)}{f_0(X)}$ , the aim of the reweighting NN is to estimate  $r(X)$  by minimizing the cost function [184]:

$$\mathcal{J}(u) = E_0[\phi(u(X)) + r(X)\psi(u(X))] = E_0[\phi(u(X))] + E_1[\psi(u(X))] \quad (6.10)$$

where  $E_0[\cdot]$  and  $E_1[\cdot]$  are the expectations with respect to  $f_0(X)$  and  $f_1(X)$  respectively, and  $u(X)$  is the NN output.  $\phi(z)$  and  $\psi(z)$  are scalar functions designed to satisfy that the global minimizer of  $\mathcal{J}(u)$  is equal to  $u(X) = \omega(r(X))$ , where  $\omega(r)$  is called the transformation function. By the Ref. [184], in the case of Log-Likelihood Ratio estimation  $\omega(r) = \log r$ , these two scalar functions can be chosen as

$$\phi(z) = e^{0.5z}, \psi(z) = e^{-0.5z}. \quad (6.11)$$

In practice, the cost function  $\mathcal{J}(u)$  is rewritten as

$$\mathcal{J}(u) \approx \hat{\mathcal{J}}(\theta) = \frac{1}{n_0} \sum_{i=0}^{n_0} \phi(u(X_i^0, \theta)) + \frac{1}{n_1} \sum_{i=0}^{n_1} \psi(u(X_i^1, \theta)), \quad (6.12)$$

where  $\theta$  stands for a set of NN parameters,  $X_i^0$  and  $X_i^1$  are random samples from background and data, and  $u(X, \theta)$  means the NN output with the parameters summarized by  $\theta$ . Comparing to the equation (6.10), the statistical expectations are replaced with averages over background and data samples. This approximation is supported by the law of large numbers, which indicates that the average result of repeating an experiment a large number of times independently should be close to its expected value. Since  $\phi(z)$  and  $\psi(z)$  are designed based on the transformation function  $\omega(r)$ , the cost function in equation 6.12 only depends on  $\omega(r)$  and two datasets, requires no knowledge of  $f_0(X)$ ,  $f_1(X)$ . Therefore, the problem of minimizing the cost function has been converted from finding the global minimizer  $u(X)$  to finding the optimal NN parameters ( $\theta$ ), which can be implemented by training the NN.

By substituting  $\phi$  and  $\psi$  in equation 6.11, the cost function for Log-Likelihood Ratio estimation can be written as

$$\hat{\mathcal{J}}(\theta) = \frac{1}{n_0} \sum_{i=0}^{n_0} e^{u(X_i^0, \theta)/2} + \frac{1}{n_1} \sum_{i=0}^{n_1} e^{-u(X_i^1, \theta)/2}. \quad (6.13)$$

The NN output  $u(X)$  estimates the log-likelihood ratio, while the likelihood ratio  $r(X)$  is estimated by  $e^{u(X)}$ .

### 6.4.3 Training

For the reweighting NN training, 11 input variables are used, including the number of tracks and 6 jet substructure variables which are also used in the classification and regression NNs, event kinematic variables, as well as the final state invariant mass. Those training variables are shown in Table 6.18. Only events with  $120 < m_{llj} < 140$  GeV and  $140 < m_{llj} < 180$  GeV are used in the training.

Variable	Description
$m_{llj}$	Final state invariant mass
$p_{T_H}$	Transverse momentum of reconstructed Higgs boson
$p_{T_Z}$	Transverse momentum of reconstructed Z boson
$p_T^{\text{jet}}$	Transverse momentum of reconstructed jet
$n_{\text{tracks}}$	Number of Ghost-Associated tracks
$p_T^{\text{lead track}}/p_T^{\text{tracks}}$	Fraction of the leading track transverse momentum to the total
$\Delta R^{\text{lead track, calo jet}}$	$\Delta R$ between the leading track and the reconstructed jet
$\tau_2$	NSubJettiness 2 [181]
$U_1(0.7)$	Modified energy correlation function ${}_1e_2^{(0.7)}$ [182]
$M_2(0.3)$	Modified energy correlation functions ${}_1e_3^{(0.3)}/{}_1e_2^{(0.3)}$ [182]
angularity(2)	Angularity 2 [183]

Table 6.18: Training variables for the Reweighting NN.

The optimization of the hyperparameters follows the same strategy as the regression NN, combining typical suggestions and probabilistic optimization. A  $\chi^2$  test is performed on the invariant mass  $m_{llj}$  for each set of hyperparameters to monitor the disagreement between data and reweighted background.

The final decision of reweighting result is made in two stages. In the first stage, hyperparameters are altered based on different strategy and the  $\chi^2$  test on  $m_{llj}$  is performed for each trial. After tens of trials, the set of hyperparameters getting the lowest  $\chi^2$  value, i.e. the smallest data-background disagreement, is selected. In the second stage, the hyperparameters are fixed as the selected and more trials are trained with randomly initialised weights. The trial with the lowest  $\chi^2$  value is chosen for the background modelling. Optimal hyperparameters are shown in Table 6.19.

<i>a</i> Hyperparameter	Optimization Strategy	Value
Layers	Bayesian Optimization	3
Neurons per Layer	Bayesian Optimization	100
Hidden Layer Activation Function	Recommendation for Feed-Forward NNs	ReLU
Output Layer Activation Function	Smoothly limited linear	$N \cdot \tanh(x/N)$
Loss Function	Exponential Loss for Log-Likelihood Ratio Estimation [184]	Equation (6.13)
Optimizer	Choice between ADAM and SGD	SGD
Learning rate	Bayesian Optimization	0.11
Momentum	Bayesian Optimization	0.975
Decay rate	Bayesian Optimization	0.001
Batch Size	Bayesian Optimization	200
Epochs	Best Validation Loss	50

Table 6.19: Optimal hyperparameters and their optimization strategy for the reweighting NN.

#### 6.4.4 Reweighting results

The distributions of 11 input variables before and after the reweighting are shown in Figures 6.24. After the reweighting, the background modellings for all 11 variables are improved significantly.

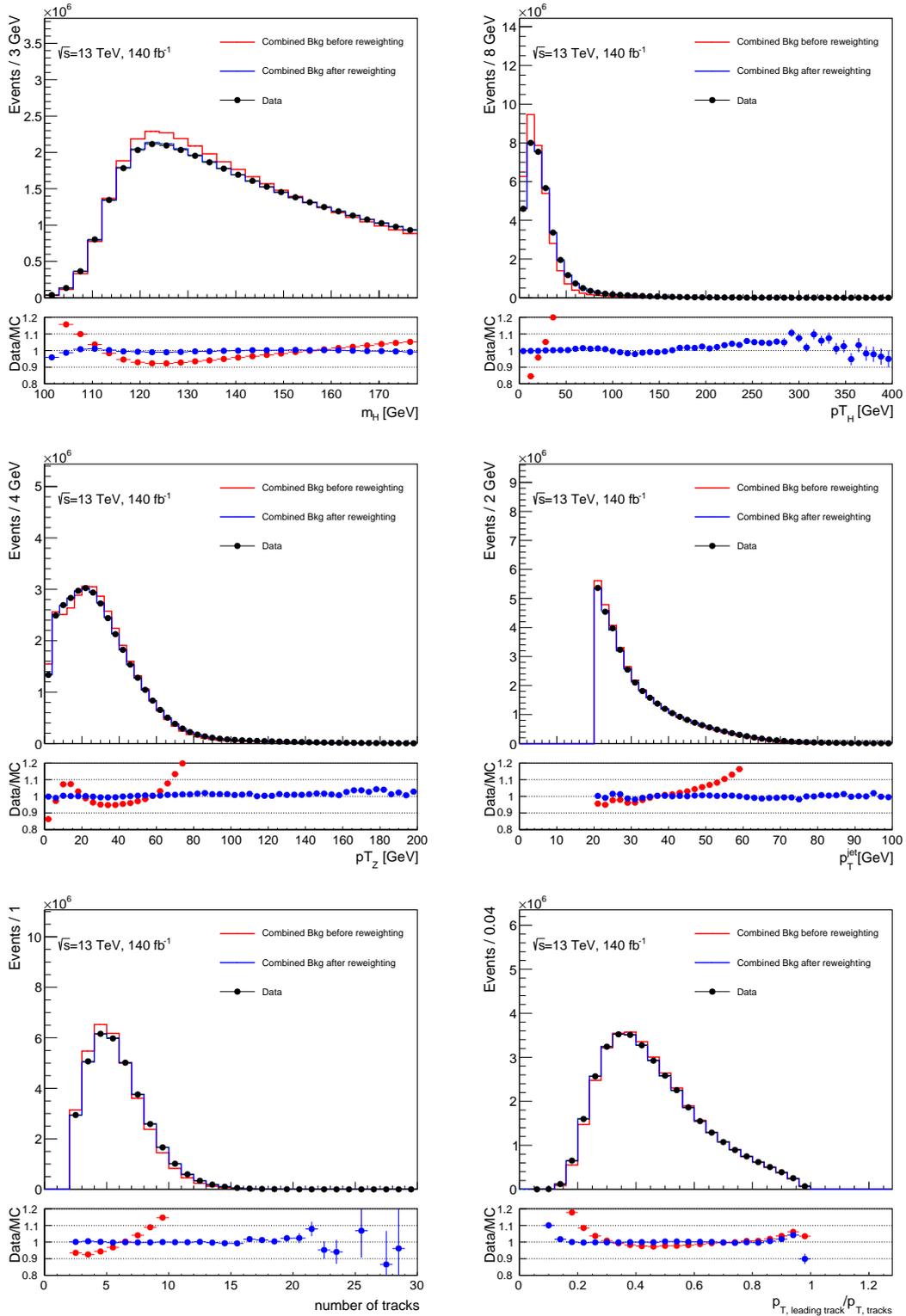


Figure 6.24: (Part 1) Distributions of 11 variables used to reweight the background, after the full pre-selection. The red and blue histograms in the upper panels show the background distribution before and after the reweighting, respectively. In the lower panels, the red and blue dots represent the ratio of data to non-reweighted background and data to reweighted background, respectively.

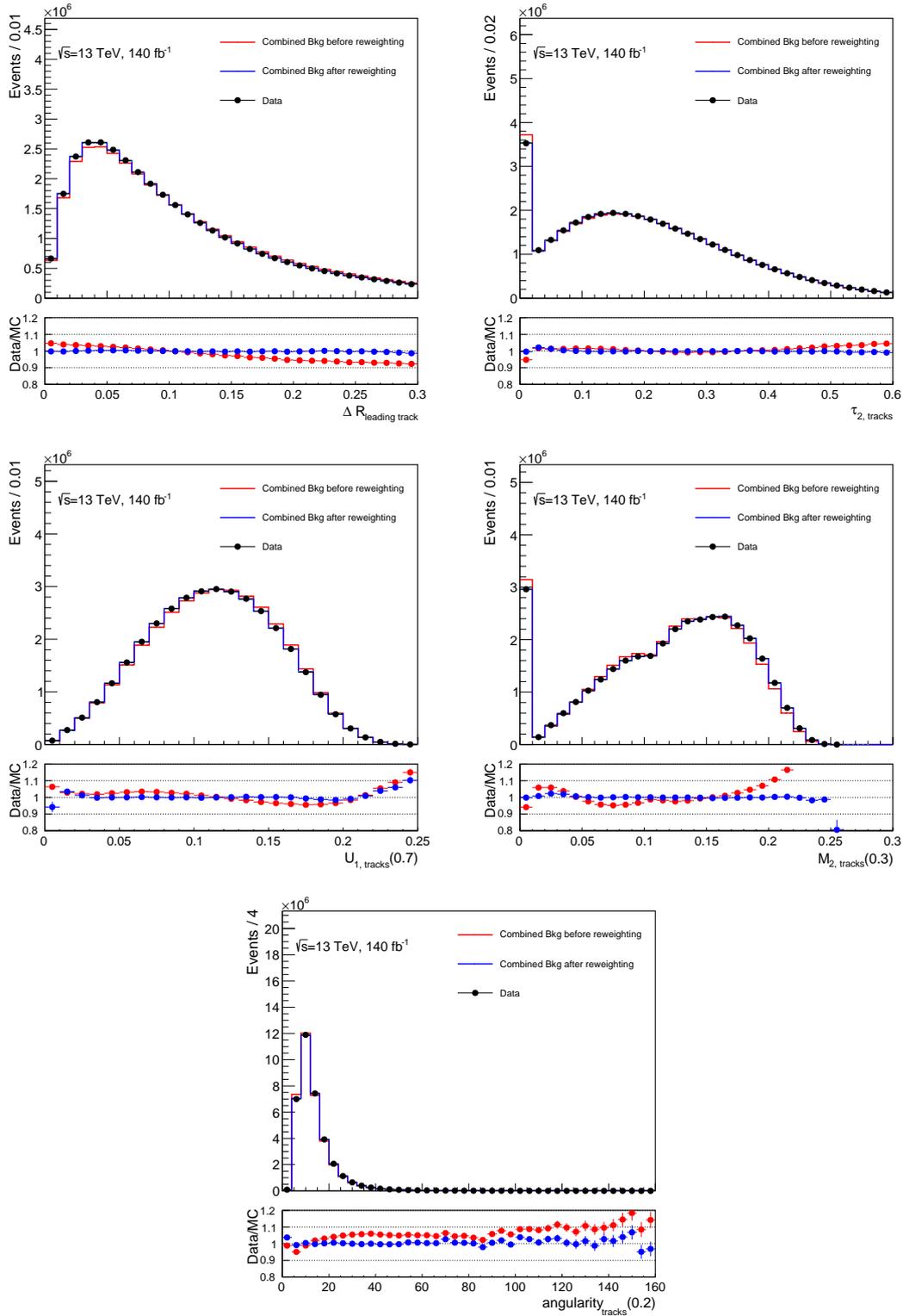


Figure 6.24: (Part 2) Distributions of 11 variables used to reweight the background, after the full pre-selection. The red and blue histograms in the upper panels show the background distribution before and after the reweighting, respectively. In the lower panels, the red and blue dots represent the ratio of data to non-reweighted background and data to reweighted background, respectively.

### 6.4.5 Reweighting Validation

As shown in the signal distributions in Figure 6.6, there are few signal events lie out of the region  $120 \text{ GeV} < m_{llj} < 140 \text{ GeV}$ . To check the validity of the nominal background reweighting in section 6.4.4, an alternative reweighting NN is trained. Instead of excluding events in the range  $120 \text{ GeV} < m_{llj} < 140 \text{ GeV}$ , the alternative NN blinds a wider region:  $115 \text{ GeV} < m_{llj} < 145 \text{ GeV}$ . By comparing the  $m_{llj}$  distributions after two reweightings (Figure 6.25) and other variables in the sideband region ( $115 \text{ GeV} < m_H < 120 \text{ GeV}$  &  $140 \text{ GeV} < m_H < 145 \text{ GeV}$ ) (Figure 6.26), there is no significant difference between the results of these two NNs. Therefore, a wider blind region does not affect the performance of the reweighting NN, the signal contamination on the nominal reweighting NN is negligible.

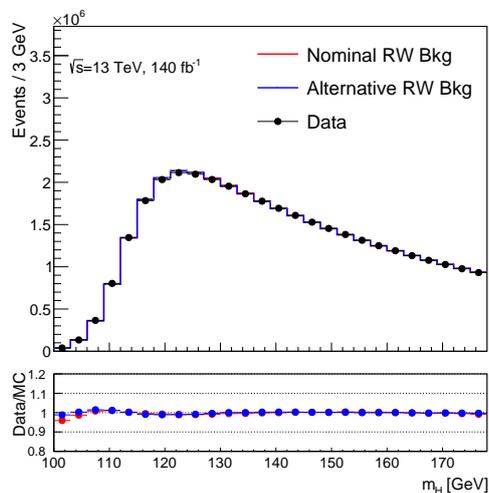


Figure 6.25: Comparison of the  $m_{llj}$  after reweighting via two reweighting NNs with different blind regions. The nominal NN blinds the region 120-140 GeV, while the alternative NN blinds the region 115-145 GeV. Red and blue dots in the lower panel show the ratios of data to background reweighted by the nominal NN and the alternative NN respectively.

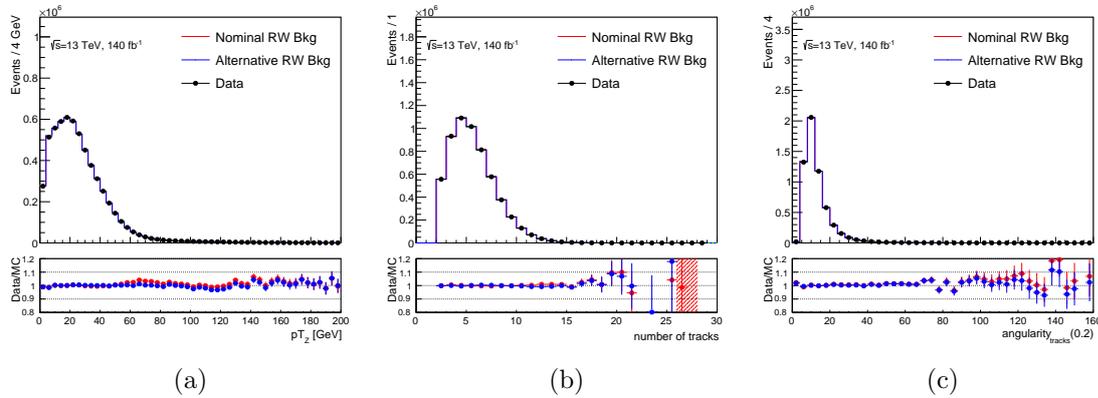


Figure 6.26: Comparison of the variables (a)  $pT_Z$ , (b) number of tracks, and (c) angularity after reweighting via two neural networks with different blind regions. The nominal NN blinds the region 120-140 GeV, while the alternative NN blinds the region 115-145 GeV. Red and blue dots in the lower panel show the ratios of data to background reweighted by the nominal NN and the alternative NN respectively. Only events in the  $m_{llj}$  region of 115-120 GeV and 140-145 GeV are shown.

#### 6.4.6 Standard Model Higgs contamination

In order to estimate the contamination from SM Higgs decays, VBF Higgs,  $H \rightarrow ZZ \rightarrow 4l$  and  $H \rightarrow Z\gamma \rightarrow ll + \gamma$  MC samples are used. The VBF Higgs inclusive sample is scaled to the total Higgs production cross section. The greatest contamination is expected from the  $H \rightarrow ZZ \rightarrow 2l + \text{hadrons}$  decays. Its contamination is estimated by assuming the pre-selection and NN selection efficiency is same as the diboson background ( $\sim 30\% \times 2\% = 0.6\%$ ), and the branching ratio is  $\text{BR}(H \rightarrow ZZ) \times \text{BR}(Z \rightarrow ll) \times \text{BR}(Z \rightarrow \text{hadrons}) \sim 3\% \times 7.5\% \times 70\% = 0.16\%$ .  $ZH \rightarrow ll + \text{hadrons}$  events are also estimated. The Expected events numbers in the SR caused by SM Higgs contamination are listed in Table 6.20.

These few contamination events are added as a gaussian distribution ( $N=50$ ,  $\mu = 130$  GeV,  $\sigma = 7$  GeV) on the top of the expected background which counts  $89900 \pm 150$  events in the  $m_{llj}$  120-140 GeV region. Therefore the effect is negligible.

Process	expected SR yield	Notes
$H \rightarrow ZZ \rightarrow 4l$	2	
$H \rightarrow Z\gamma \rightarrow ll + \gamma$	11	
VBF Higgs	60	Scaled to 55.7 pb, half of events lie in the $m_{ll}$ 120-140 GeV region
$H \rightarrow ZZ \rightarrow 2l + \text{hadrons}$	70	Only a part of $Z$ bosons decays will be boosted enough
$ZH \rightarrow ll + \text{hadrons}$	< 10	The process is non-resonant

Table 6.20: Expected events numbers in the signal region of each SM Higgs process that may cause contamination.

## 6.5 Systematic Uncertainties

There are two types of systematic uncertainties: theoretical and experimental (detector and reconstruction). They are used in the statistical model as nuisance parameters (NP).

### 6.5.1 Background

In the subsection, the background statistical uncertainty and four background modelling uncertainties are discussed.

#### Background Statistical Uncertainty

The background statistical uncertainty can be derived from the  $m_{ll}$  distribution histograms in Figure 6.27. In the  $120 \text{ GeV} < m_{ll} < 140 \text{ GeV}$  region, the pure background MC yield and uncertainty is  $89900 \pm 150$ . After the reweighting it becomes  $88400 \pm 180$  events. Comparing to  $82400 \pm 2900$  events in the previous analysis [50], the MC statistical uncertainty has been reduced from 3.5% to 0.22% in total. This reduction comes mainly from the higher statistics of the Powheg  $Z$ +jets sample compared to the Sherpa samples. In the final fit, the statistical uncertainty is calculated and implemented bin-by-bin of the  $m_{ll}$  distribution histogram.

## Background Modelling Uncertainties

The background modelling uncertainty consists of four main uncertainties:

The first one is data-driven shape uncertainty. It is estimated from a Control Region (CR) including same amount of background events as the Signal Region (SR). The CR is defined near the SR, by changing the classification NN requirement from  $\text{NN output} > 0.93$  to  $0.883 < \text{NN output} < 0.93$ . The  $m_{Uj}$  difference between data and background in CR is shown in Figure 6.28 (left). The data-driven uncertainty is derived by multiplying the SR background by the data/background ratio in the CR. Then it is symmetrized to generate the up (or down) variation, as shown in Figure 6.28 (right). The signal contamination is also shown as assuming  $\text{BR}(H \rightarrow Za) = 10\%$  which is close to the upper limit for undetected non-SM Higgs decays [7].

To check the stability of deriving the data-driven uncertainty from the CR, two more validation regions (VR) are defined: VR1:  $0.907 < \text{NN output} < 0.93$  and VR2:  $0.843 < \text{NN output} < 0.93$ . These two VRs are also near the signal region, and have half of and double of background events as the signal region respectively. The  $m_{Uj}$  distributions in these VRs and the uncertainties derived from these VRs can be seen in the Figure 6.29. As shown in Figure 6.30, the difference between uncertainties derived from VRs and the CR after smoothing is negligible.

A signal injection test is also implemented to check the reliability of the background shape uncertainty. 2.5 GeV and 3.5 GeV signal events are injected to the control region ( $0.883 < \text{NN output} < 0.93$ ) with  $\text{BR}(H \rightarrow Za) = 10\%$ . The Figures 6.31 shows the results of the signal injection tests. As shown in Figure 6.32 The effect of signal injection on background shape uncertainty is negligible.

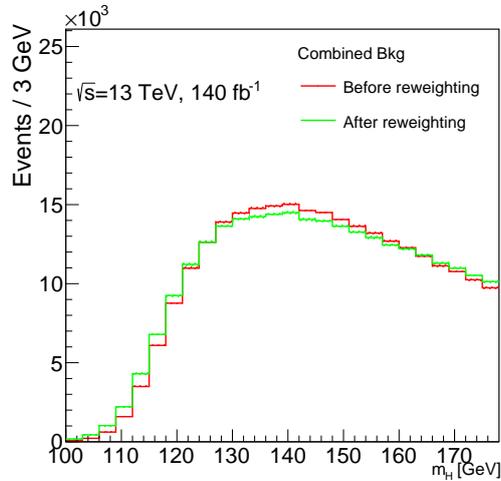


Figure 6.27: Background  $m_{llj}$  distributions after both pre-selection and the classification NN cut, (red) before and (green) after reweighting.

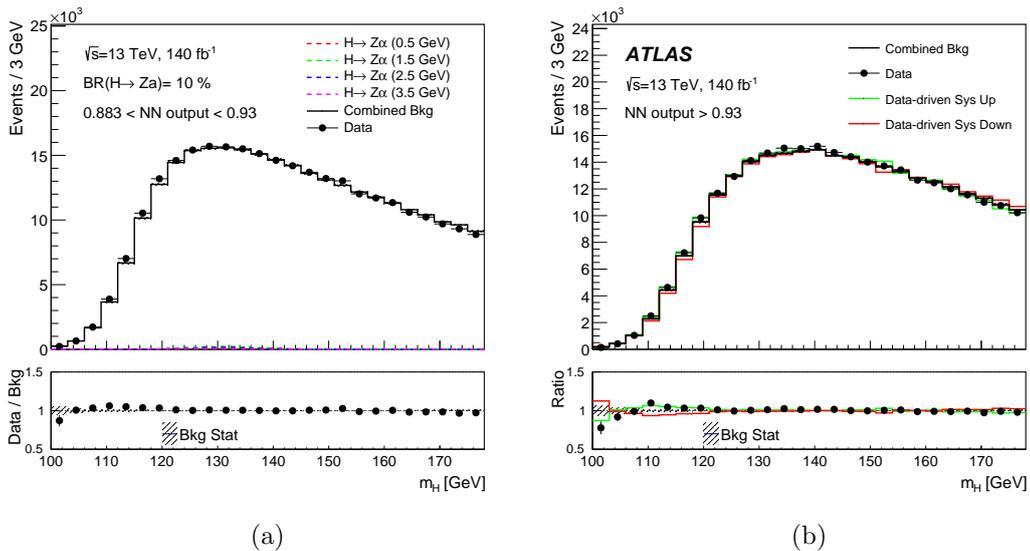


Figure 6.28: (a) Data and MC  $m_{llj}$  distributions in the control region  $0.883 < \text{NN output} < 0.93$ . The data-MC disagreement is used to apply a Background modelling systematic uncertainty (b) in the Signal Region  $\text{NN output} > 0.93$ . The red and green lines represent the down and up variations respectively. The lower panel shows the ratios of data and systematic variations to the background.

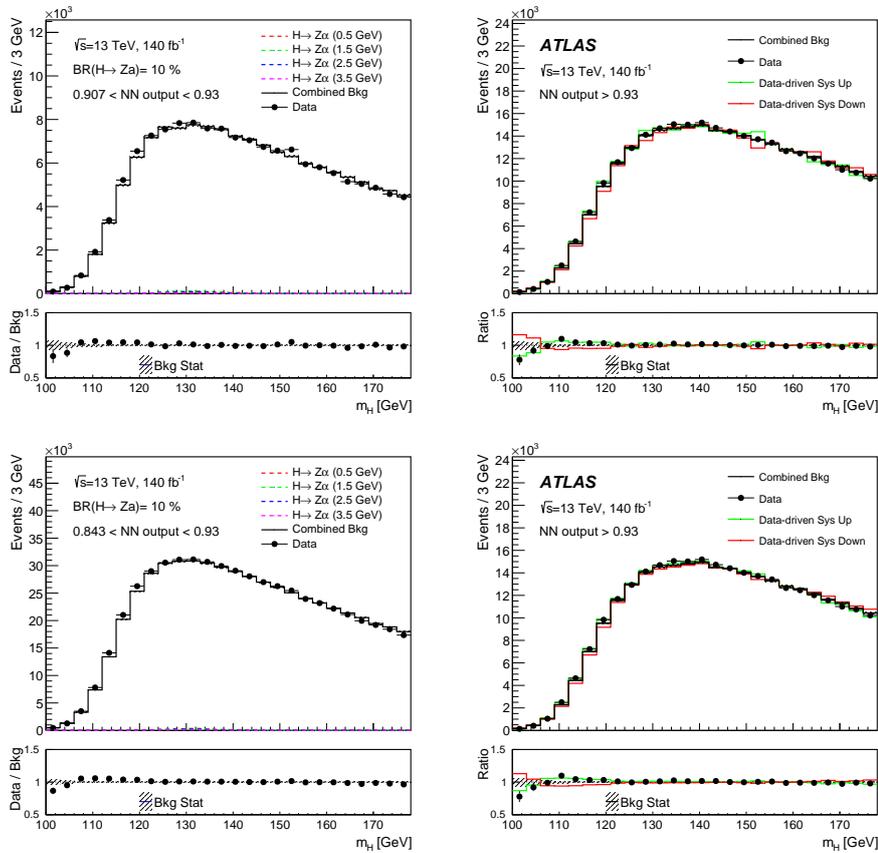


Figure 6.29: Data-driven uncertainty validation checks. Data and MC  $m_{llj}$  distributions in the validation regions (top-left) VR1:  $0.907 < \text{NN output} < 0.93$  and (bottom-left) VR2:  $0.843 < \text{NN output} < 0.93$ . Two plots on the right are their corresponding data-driven systematic uncertainties in the Signal Region  $\text{NN output} > 0.93$ . The red and green lines represent the down and up variations respectively.

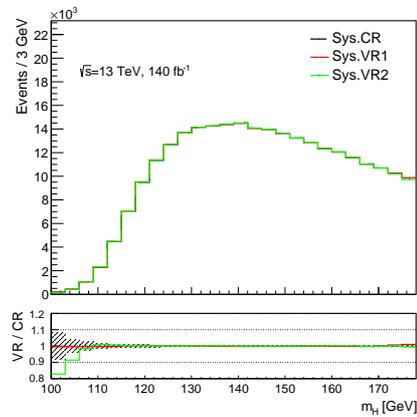


Figure 6.30: Ratio of the data-driven systematics derived from 2 VRs to the CR. The histograms are smoothed.

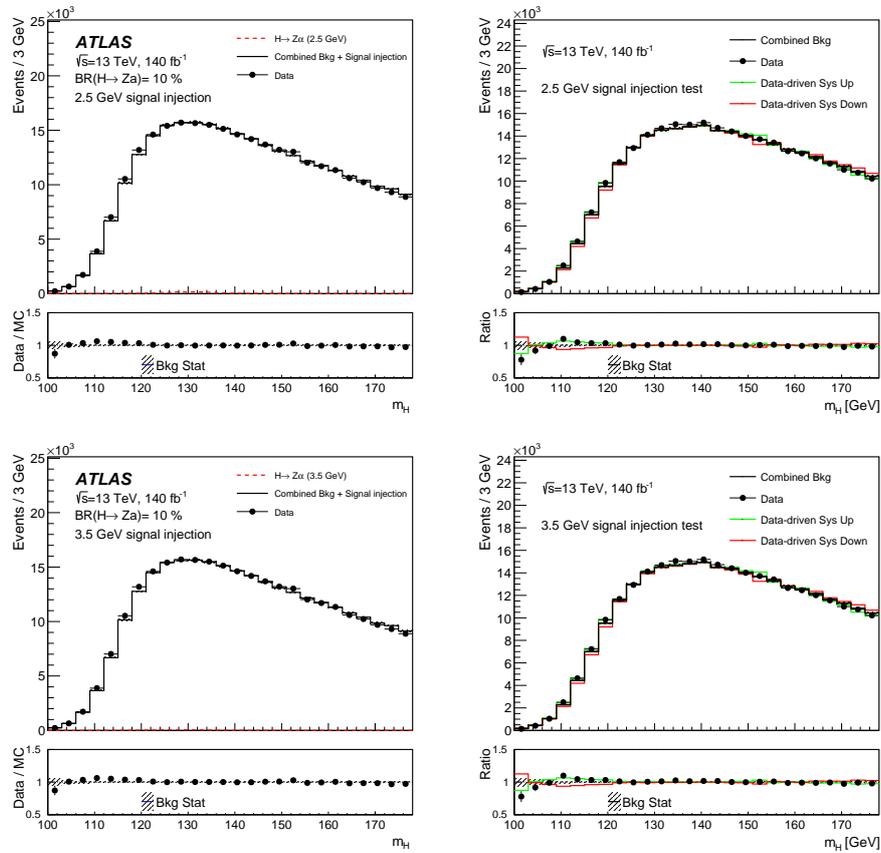


Figure 6.31: The results of (top) 2.5 GeV and (bottom) 3.5 GeV signal injection tests. The left plots show the data, signal injected background and signal distributions in the control region. The signal is injected to the control region with  $\text{BR}(H \rightarrow Za) = 10\%$ . The right plots show the corresponding data-driven uncertainties in the signal region  $\text{NN} > 0.93$ . The red and green lines represent the down and up variations respectively.

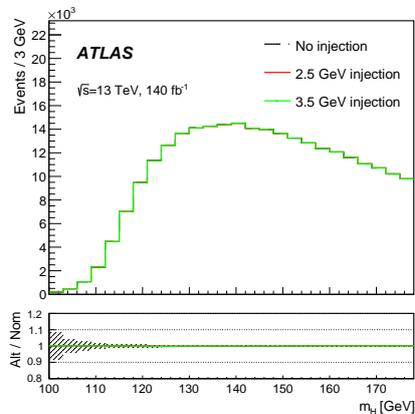


Figure 6.32: Ratio of the data-driven systematics derived from signal injections to the nominal. The histograms are smoothed.

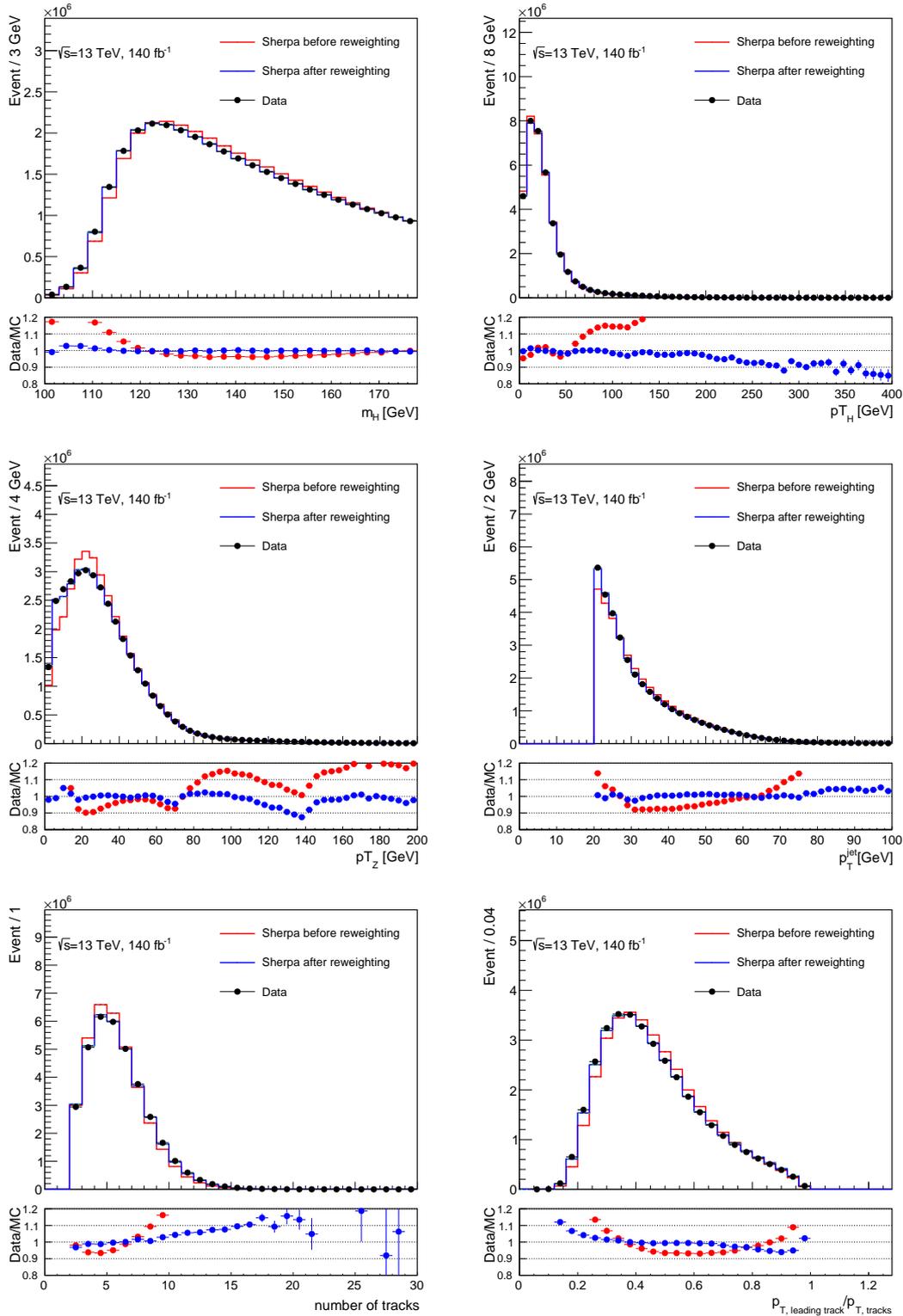


Figure 6.33: (Part 1) Distributions of 11 variables used to reweight the SHERPA background, after the full pre-selection. The red and blue histograms in the upper panels show the background distribution before and after the reweighting, respectively. In the lower panels, the red and blue dots represent the ratio of data to non-reweighted background and data to reweighted background, respectively.

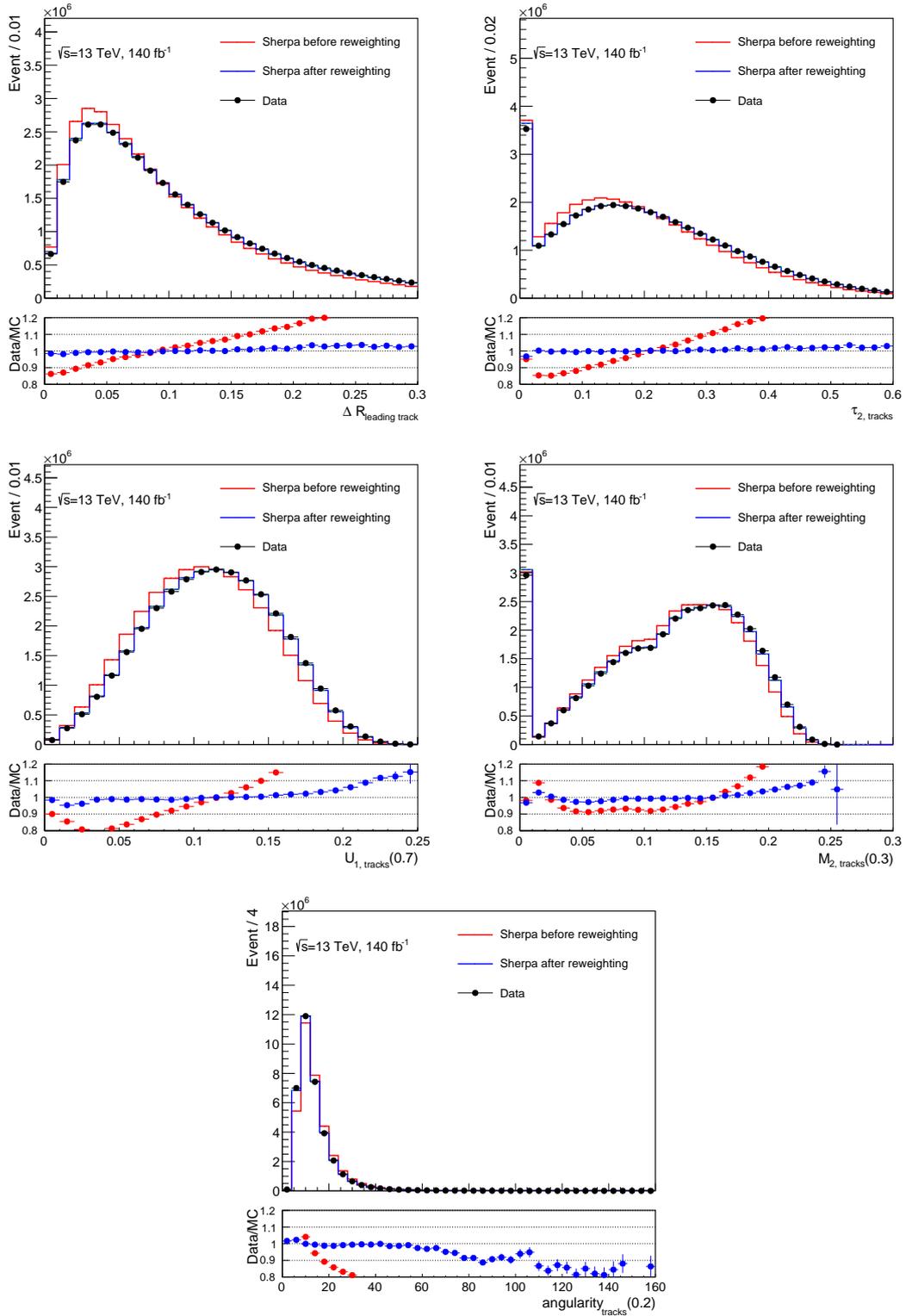


Figure 6.33: (Part 2) Distributions of 11 variables used to reweight the SHERPA background, after the full pre-selection. The red and blue histograms in the upper panels show the background distribution before and after the reweighting, respectively. In the lower panels, the red and blue dots represent the ratio of data to non-reweighted background and data to reweighted background, respectively.

The second modelling uncertainty is the theoretical uncertainty, coming from a different choice of the MC generator. As mentioned in Table 6.2, the nominal  $Z$ +jets background are POWHEG samples, the alternative choice is using SHERPA samples. A dedicated Reweighting NN is trained for SHERPA samples. The input variables and the training strategy are same as the nominal Powheg Reweighting NN in section 6.4.2. During the NN training, negative weighted events in SHERPA samples are ignored.

The comparison of POWHEG and SHERPA  $m_{llj}$  distributions before and after reweighting is shown in Figure 6.34. The reweighting NNs achieve similar levels of data-MC agreement on the  $m_{llj}$  distribution for two sets of background samples. After applying the Classification NN cut, as shown in Figure 6.35, the POWHEG-SHERPA disagreement is used as one of background modelling uncertainties. The results of all 11 reweighting variables for SHERPA samples is shown in Figure 6.33.

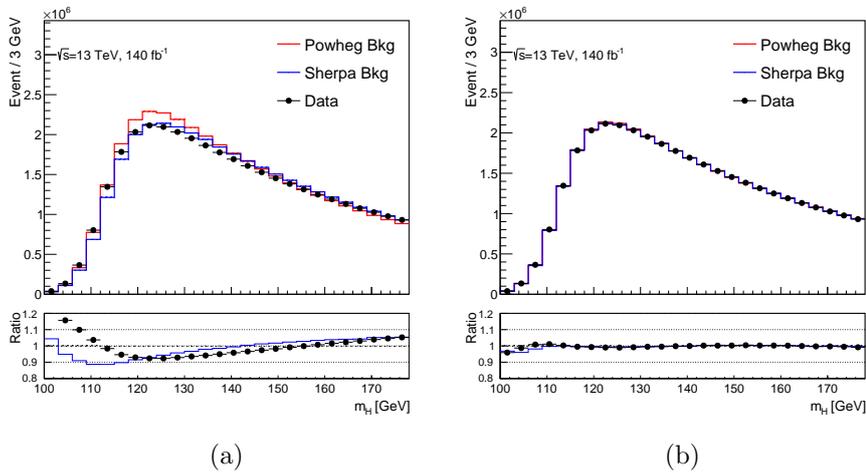


Figure 6.34: Distributions of  $m_{llj}$  of Powheg, Sherpa and data (a) before and (b) after reweighting.

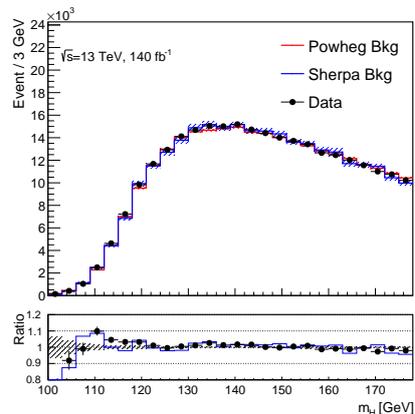


Figure 6.35: Distributions of  $m_{llj}$  of Powheg, Sherpa and data after the Classification NN cut.

The third is the JET experimental uncertainties. It has roughly 20 Jet Energy Scale (JES) uncertainty components and 13 Jet Energy Resolution (JER) uncertainty components, including  $\eta$ -calibration, pileup, flavor composition, flavor response,  $b$ -jets and punch-through jets. Table 6.21 lists these individual components.

All these uncertainty components are reweighted by the nominal reweighting NN, and then cut by classification NN output  $> 0.93$ . Figure 6.36 shows the comparison between the nominal background and the background including each of the 13 JER uncertainty components. The background including JER uncertainty components goes through the reweighting trained with the nominal MC.

The fourth is the bootstrapping uncertainty, describing the spread of reweighting NNs. To generate bootstrapping replicas, each event of background and data samples is assigned a new weight, which is randomly sampled from a Poisson distribution with mean value of 1. The reweighting NN is trained with these events for each replica. After being reweighted by those bootstrapping NN and the classification cut, the background distribution in the SR is shown in Figure 6.37 (left). The bootstrapping uncertainty is derived by calculating the standard derivation of bootstrapping NNs bin-by-bin, and then symmetrization. It is shown in Figure 6.37 (right).

---

<b>Small-<math>R</math> Jets</b>	
Systematic uncertainty name	Description
JET_EtaIntercal_Modelling	$\eta$ -intercalibration: MC generator modelling uncertainty
JET_EtaIntercal_NonClosure_2018data	$\eta$ -intercalibration: non-closure uncertainty of jet response, 2018 data
JET_EtaIntercal_NonClosure_highE	$\eta$ -intercalibration: non-closure uncertainty of jet response, high energy component
JET_EtaIntercal_NonClosure_negEta	$\eta$ -intercalibration: non-closure uncertainty of jet response, negative $\eta$ component
JET_EtaIntercal_NonClosure_posEta	$\eta$ -intercalibration: non-closure uncertainty of jet response, positive $\eta$ component
JET_PileUp_OffsetMu	Pileup: Offset, term for number of interactions per crossing $\mu$
JET_PileUp_OffsetNPV	Pileup: Offset, term for number of primary vertices
JET_PileUp_PtTerm	Pileup: Offset, $p_T$ term
JET_PileUp_RhoTopology	Pileup: Offset, $\rho$ topology uncertainty on jet areas
JET_PunchThrough_MC16	Punch-through correction uncertainty
JET_EffectiveNP{1-8}	Effective JES uncertainty
JET_Flavor_Composition	Flavor composition uncertainty
JET_Flavor_Response	Flavor response uncertainty (dominated by gluon response)
JET_BJES_Response	Jet energy scale uncertainty for $b$ -jets
JET_JER_DataVsMC_MC16	Data vs MC difference
JET_JER_EffectiveNP{1-12}	Effective JER uncertainty

---

Table 6.21: Summary of considered jet experimental systematic uncertainties.

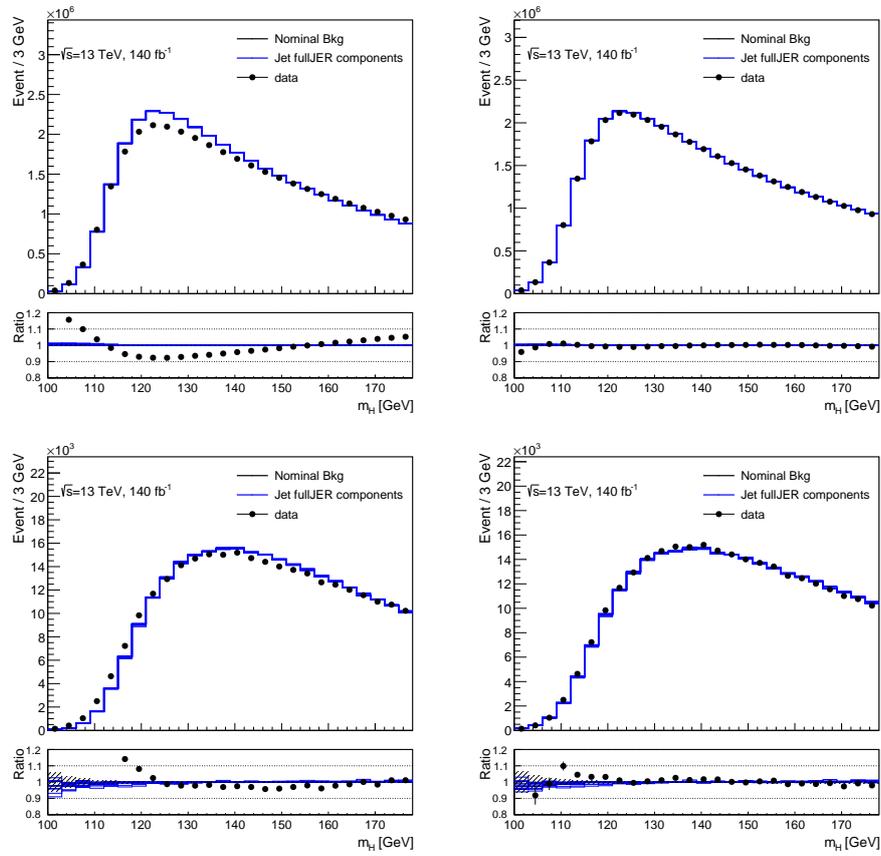


Figure 6.36:  $m_{lj}$  distributions of nominal background, data and the background including each of the 13 JER uncertainty components. (top-left) Before the reweighting and before the classification cut. (top-right) After the reweighting and before the classification cut. (bottom-left) Before the reweighting and after the classification cut. (bottom-right) After the reweighting and after the classification cut. The lower panels show the ratios to the nominal background.

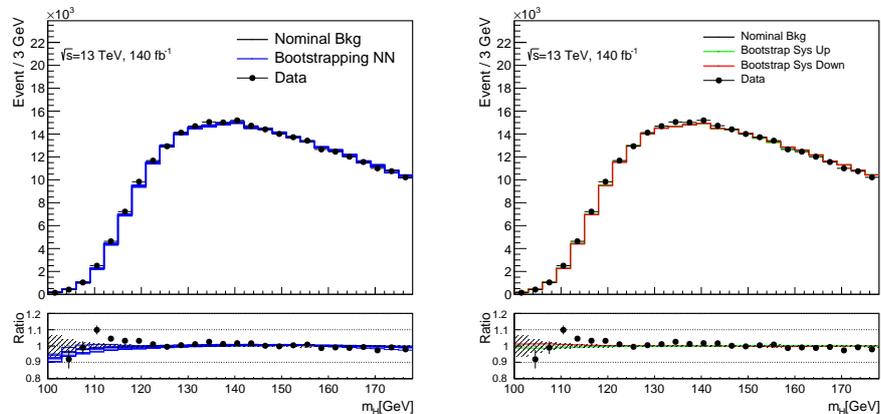


Figure 6.37: (left) Background bootstraps in the SR. (right) Calculated bootstrapping uncertainty. The lower panels show the ratio to the nominal background.

## 6.5.2 Signal

Both theoretical and experimental systematic uncertainties are considered for the signal model. The theoretical uncertainty comes from the parton shower and hadronization, which are evaluated by comparing the nominal signal MC samples (PYTHIA) with alternative samples (HERWIG). The experimental uncertainties are calculated from many components, including jet, tracking, pile-up, leptons, Trigger and vertex scale factors uncertainties.

The effect of systematic uncertainties on the signal model is estimated by fitting the  $m_{llj}$  distribution after the full event and NN selection for each systematic variation with a gaussian function, and calculating the difference of fit parameters: mean and sigma between the nominal and systematic variations. The difference in the number of the expected events is considered as a normalization uncertainty.

### Experimental

Signal experimental uncertainties come from jet, leptons, vertices, triggers, etc. The jet experimental uncertainties have many sub-components, same as listed in Table 6.21. The pile-up distribution affects MC modelling. The pile-up uncertainty is propagated through the MC mismodelling and affects the efficiency of the classification NN, and it is modelled by the reweighting [185]. The lepton uncertainties come from the reconstruction, identification and isolation of leptons [73], for the muon there are also uncertainties from TTVA [73]. The uncertainties from the primary vertex and the trigger efficiencies are calculated using dedicated tools. The full Run 2  $pp$  data sample corresponds to an integrated luminosity  $140.1 \pm 1.2 \text{ fb}^{-1}$ , which means a luminosity uncertainty of 0.83% [186].

For the tracking uncertainties, the HDBS3 derivation does not contain necessary information. The effect of tracking uncertainties is calculated from a 10k events 0.5 GeV signal sample produced locally. The impact on the normalization of the

0.5 GeV signal is expected to be the largest due to the low track multiplicity. The impact is less than 2% for any of the corresponding uncertainties.

The  $m_{llj}$  distributions of signal experimental systematic components are fit with Gaussian functions. Figure 6.38 shows an example of the distribution and the fit result. The fit parameters,  $\mu$  and  $\sigma$ , of nominal background (up, down systematic variation) are  $\mu_{\text{nominal}}$  ( $\mu_{\text{var,up}}$ ,  $\mu_{\text{var,down}}$ ) and  $\sigma_{\text{nominal}}$  ( $\sigma_{\text{var,up}}$ ,  $\sigma_{\text{var,down}}$ ).

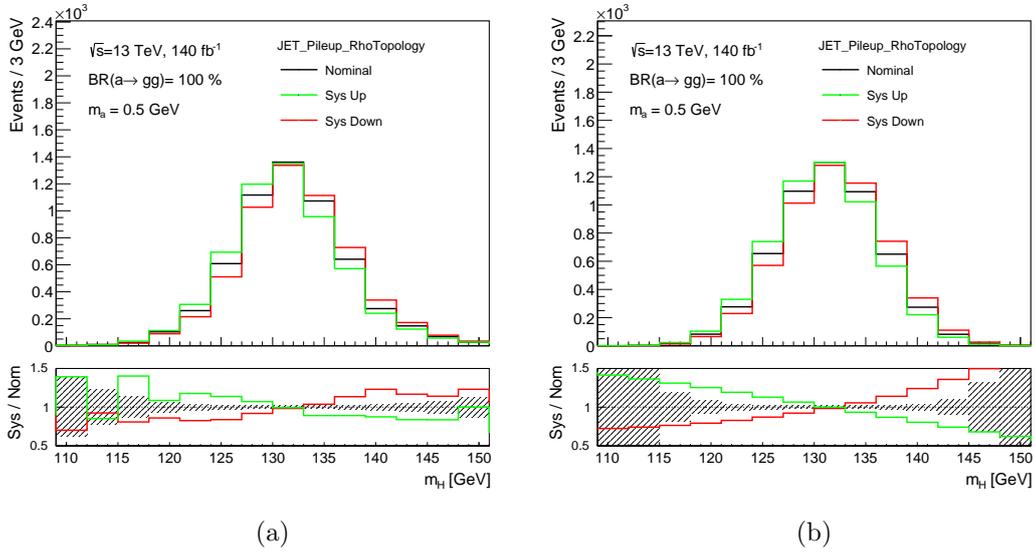


Figure 6.38: The jet experimental uncertainty component *JET Pileup RhoTopology* of 0.5 GeV signal. (a) Histograms before fitting. (b) Histograms generated from Gaussian fit functions.

The difference terms of fit parameters are defined by

$$\Delta\mu_{\text{var}} = \frac{|\mu_{\text{var,up}} - \mu_{\text{nominal}}| + |\mu_{\text{var,down}} - \mu_{\text{nominal}}|}{2} \quad (6.14)$$

and

$$\Delta\sigma_{\text{var}} = \frac{|\sigma_{\text{var,up}} - \sigma_{\text{nominal}}| + |\sigma_{\text{var,down}} - \sigma_{\text{nominal}}|}{2} \quad (6.15)$$

The 2 most significant uncertainty components in terms of  $\mu$  and  $\sigma$  for the 0.5 GeV signal are *JET Pileup RhoTopology* and *JET JER EffectiveNP 2* (more detail in Ref. [187]), their fit results are shown in Figure 6.39.

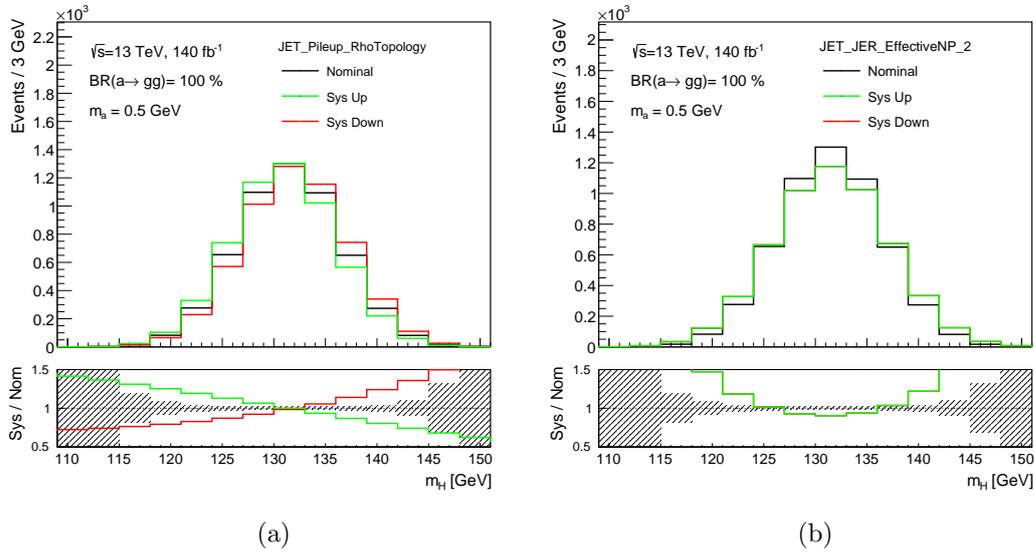


Figure 6.39: Effect of the most significant systematic uncertainty in terms of  $\mu$  and  $\sigma$  for the 0.5 GeV signal mass. (a) *JET Pileup RhoTopology* and (b) *JET JER EffectiveNP 2*.

The total signal experimental uncertainty is then estimated by adding these difference terms quadratically:

$$\Delta\mu = \sqrt{\sum_{\text{var}} \Delta\mu_{\text{var}}^2}, \quad \text{and} \quad \Delta\sigma = \sqrt{\sum_{\text{var}} \Delta\sigma_{\text{var}}^2}, \quad (6.16)$$

Signal mass (GeV)	$\Delta\mu$ (GeV)	$\Delta\sigma$ (GeV)
0.5	1.3	0.7
1.5	1.1	0.6
4.0	1.0	0.5

Table 6.22: Total experimental systematic uncertainties for 0.5, 1.5 and 4 GeV signal on  $\mu$  and  $\sigma$ .

The experimental uncertainties for 0.5, 1.5 and 4.0 GeV signals are calculated, the results are shown in Table 6.22. A conservative assumption of experimental uncertainties  $\sigma_{\mu,\text{exp}} = 1.5$  GeV and  $\sigma_{\sigma,\text{exp}} = 0.7$  GeV is used for all the mass points.

## Theoretical

Theoretical uncertainties mostly comes from the Parton Shower and Hadronization model, which is considered by comparing the signal samples made by another MC generator HERWIG 7.0 [172, 173]. The HERWIG cannot generate  $m_a < 2.0$  GeV samples because of its limitation, so HERWIG samples with masses 2.0, 2.5, 3.0, 3.5, 4.0 GeV for  $a \rightarrow gg$  and  $a \rightarrow qq$  decays are made. The final state invariant mass  $m_{uj}$  distributions in the Signal Region are shown in Figure 6.40 and 6.41 for both nominal PYTHIA and HERWIG samples. The HERWIG SR distributions are also fit with Gaussian functions, the  $\mu$  and  $\sigma$  after the fitting are shown in Figures 6.42 and 6.43 respectively. The numbers of expected events in the SR are also different between 2 generators, the comparison is shown in figure 6.44.

The differences on  $\mu$  and  $\sigma$  between two generators are considered as the theoretical uncertainties. The ratio of numbers of expected events in the SR of two generators is applied as a signal normalization uncertainty ( $\sigma_N$ ). They are collected for different mass points (2.0-4.0 GeV) and decay modes, as listed in the table 6.23.

Decay mode	Signal mass (GeV)	$\sigma_{\mu,\text{theo}}$ (GeV)	$\sigma_{\sigma,\text{theo}}$ (GeV)	$\sigma_{N,\text{theo}}$ (%)
$a \rightarrow gg$	2.0	2.4	0.1	61
	2.5	2.5	0.1	65
	3.0	1.6	0.2	65
	3.5	1.4	0.1	61
	4.0	1.4	0.1	56
$a \rightarrow qq$	2.0	1.3	0.1	23
	2.5	1.6	0.2	22
	3.0	1.3	0.1	25
	3.5	0.8	0.5	28
	4.0	0.3	0.6	5

Table 6.23: Parton Shower and Hadronization uncertainties for 2.0-4.0 GeV signals on  $\mu$ ,  $\sigma$  and normalization  $N$ , both  $a \rightarrow gg$  and  $a \rightarrow qq$  decay modes are shown.

For the signal models in section 6.6.2, the total uncertainties on  $\mu$  and  $\sigma$  are calculated by adding experimental and theoretical uncertainties quadratically:

$$\sigma_{\mu} = \sqrt{\sigma_{\mu,\text{exp}}^2 + \sigma_{\mu,\text{theo}}^2}, \quad \sigma_{\sigma} = \sqrt{\sigma_{\sigma,\text{exp}}^2 + \sigma_{\sigma,\text{theo}}^2}. \quad (6.17)$$

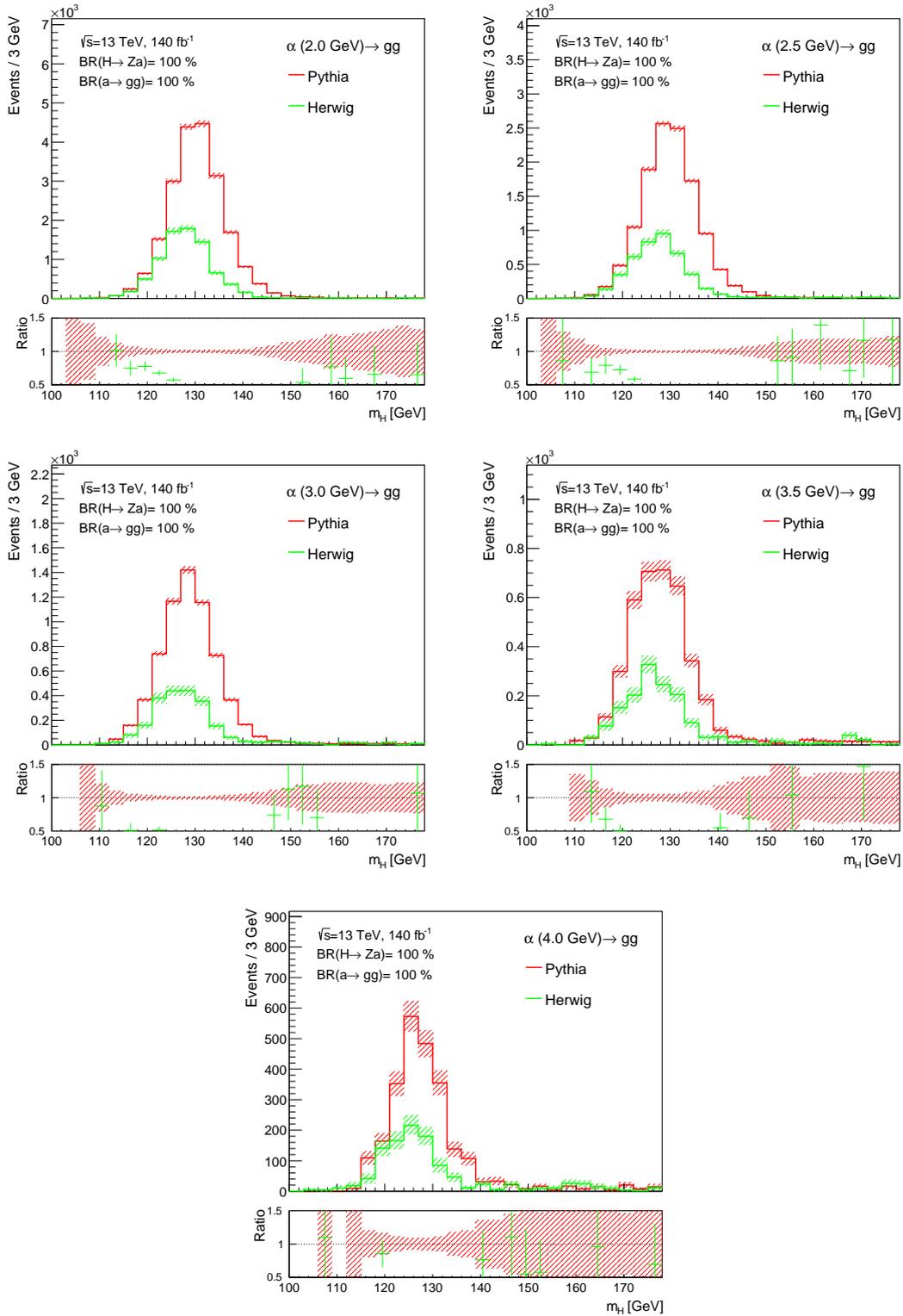


Figure 6.40:  $m_{H_j}$  distribution after full selection for PYTHIA and HERWIG for  $a \rightarrow gg$  decays for different signal masses, assuming  $\text{BR}(H \rightarrow Za) = 100\%$  and  $\text{BR}(a \rightarrow gg) = 100\%$ . The lower panels show the ratio of HERWIG to PYTHIA events.

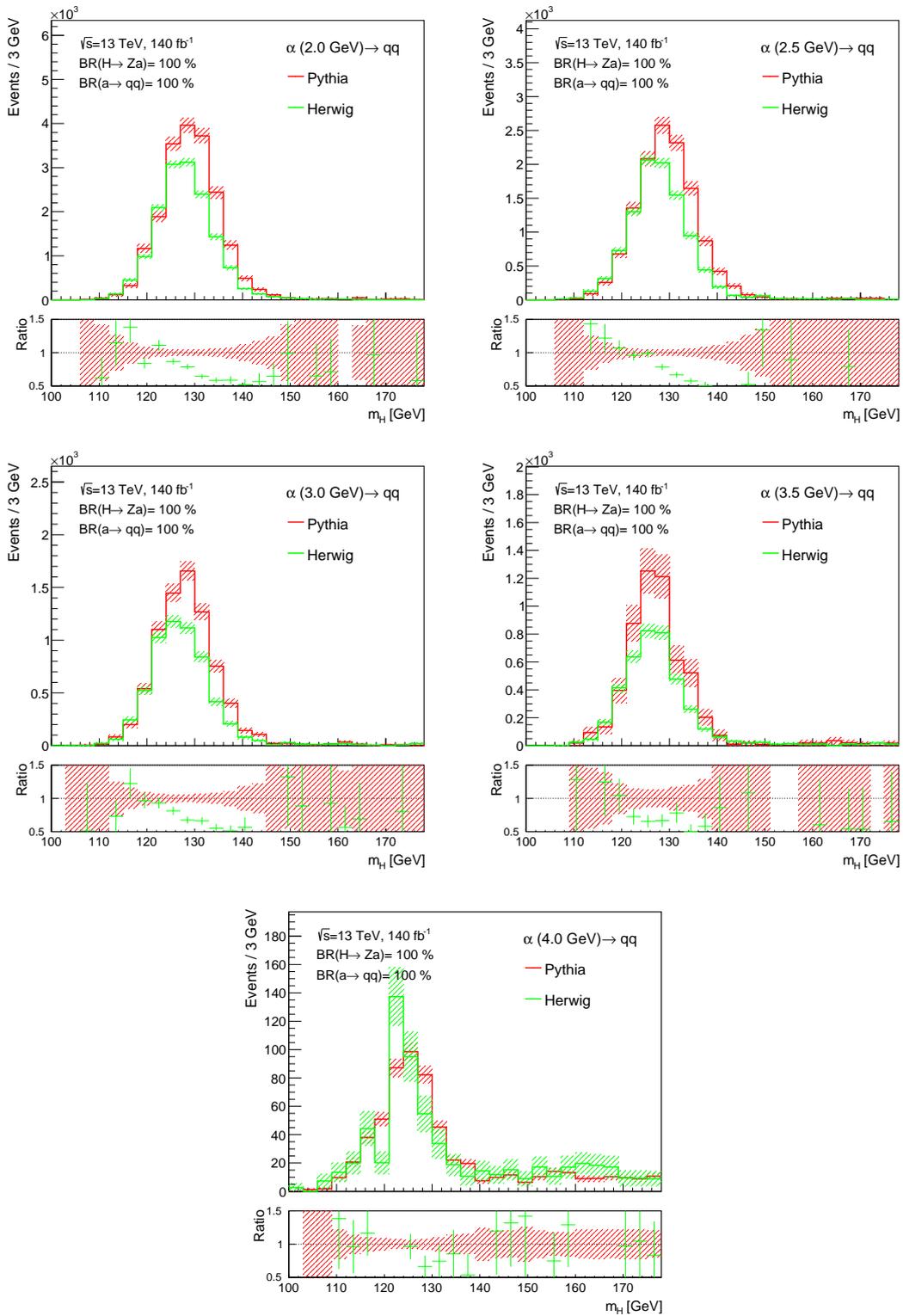


Figure 6.41:  $m_{Hj}$  distribution after full selection for PYTHIA and HERWIG for  $a \rightarrow qq$  decays for different signal masses, assuming  $\text{BR}(H \rightarrow Za) = 100\%$  and  $\text{BR}(a \rightarrow qq) = 100\%$ . The lower panels show the ratio of HERWIG to PYTHIA events.

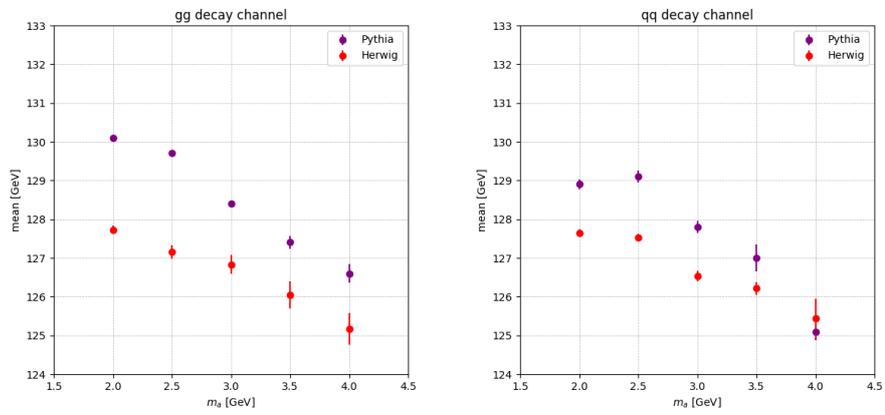


Figure 6.42: Mean after fitting the  $m_{Uj}$  distribution in the SR with a gaussian function for both PYTHIA and HERWIG samples.

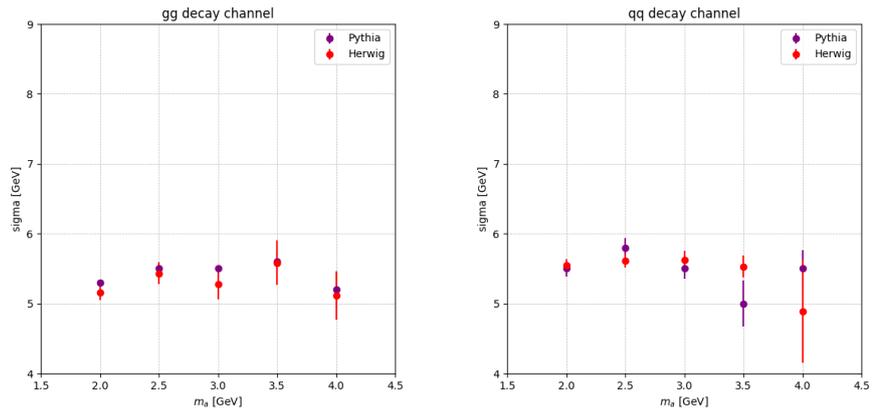


Figure 6.43: Sigma after fitting the  $m_{Uj}$  distribution in the SR with a gaussian function for both PYTHIA and HERWIG samples.

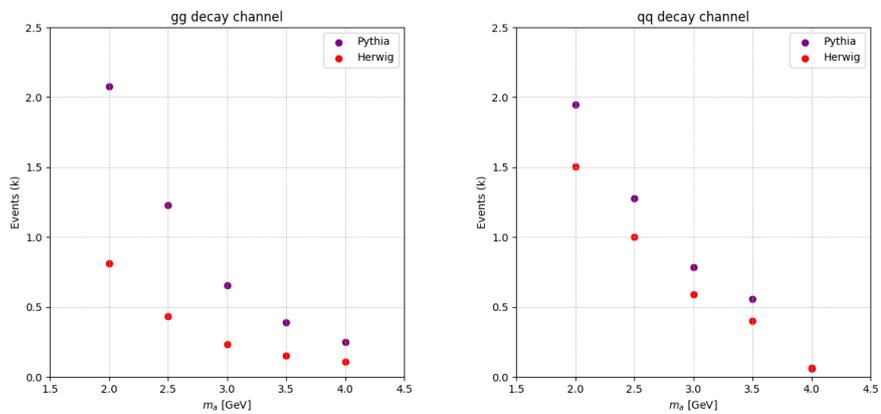


Figure 6.44: Number of expected events in the SR for both PYTHIA and HERWIG samples.

For the signals less than 2 GeV, the assumptions are made based on 2.0 GeV samples.

Additionally, a total Higgs production cross section uncertainty ( $\delta\sigma =_{-7.4\%}^{+5.6\%}$ ) is included [29], it is added quadratically to the signal normalization uncertainty.

## 6.6 Statistical Model

The final state invariant mass  $m_{llj}$  is used as the fitting variable. To extract any possible signal, the data is compared to the SM background. Any differences are quantified in terms of a hypothesis test, performed by evaluating a binned likelihood function. To interpret the results, a profile likelihood fit [188] is applied on a model including signal and background events to the data.

### 6.6.1 Statistical methods

The binned likelihood function is

$$\mathcal{L}(x; \mu, b, \alpha) = \prod_i \text{Pois}(n_i; \mu S_i(\alpha) + b B_i(\alpha)) \prod_k G(\alpha_k), \quad (6.18)$$

where  $i$  indicates the bin number of the histogram, the signal strength parameter  $\mu$  multiplies the expected signal yield  $S_i$  in each bin,  $b$  denotes the normalization of the background yield  $B_i$  in each bin, and finally  $\alpha$  represents additional nuisance parameters (NPs). These NPs quantify the effect of systematic uncertainties on the expected number of events  $S_i$  and  $B_i$ . There are two types of uncertainties. Those determined only from data are referred to as unconstrained (background normalization  $b$ ). The constrained systematic uncertainties described in section 6.5 are considered in the fit model, and are added in the fit as Gaussian priors with mean 0 and standard deviation 1, i.e. the likelihood is multiplied by a Gaussian term  $G(\alpha_k)$  that constrains the value of  $\alpha_k$  around its estimated value within the

uncertainty of the estimation.

To estimate  $\mu$ , the parameter of interest (POI) of this analysis, the negative log-likelihood function is minimized with respect to all parameters. The upper limits on  $\mu$  are calculated using a  $q_\mu$ -test based on the profile likelihood ratio:

$$q_\mu = -2 \ln(\mathcal{L}(\mu, \hat{\alpha}) / \mathcal{L}(\hat{\mu}, \hat{\alpha})), \quad (6.19)$$

where  $\hat{\mu}$  and  $\hat{\alpha}$  are the values which provide the global maximum of the likelihood  $\mathcal{L}$ , and  $\hat{\alpha}$  are the values which maximise  $\mathcal{L}$  with a given  $\mu$ . Larger value of  $q_\mu$  means larger disagreement between data and  $\mu$  hypothesis, which is quantified by  $p$ -value:

$$p_\mu = \int_{q_{\mu, \text{obs}}}^{\infty} f(q_\mu | \mu) dq_\mu, \quad (6.20)$$

where  $q_{\mu, \text{obs}}$  is the observed value of  $q_\mu$  from the data, and  $f(q_\mu | \mu)$  is the probability density function of  $q_\mu$  under the assumption of  $\mu$ . By measuring the incompatibility of the background-only model with the data, the exclusion intervals are derived using the  $CL_s$  method [188, 189]. Usually a 95% confidence level (CL) is used, which means

$$CL_s \equiv p'_\mu = \frac{p_\mu}{1 - p_b} = 5\%, \quad (6.21)$$

where  $p_b$  is the  $p$ -value under the background-only ( $\mu = 0$ ) hypothesis. While increasing the value of  $\mu$ , the first value that satisfies the equation (6.21) is the smallest  $\mu$  that can be excluded at 95% CL, i.e. the upper limit for  $\mu$  at 95% CL. Before applying the real observed data, the  $q_{\mu, \text{obs}}$  in the  $p$ -value calculation (equation (6.20)) is replaced by the median value of  $q_\mu$  under the background-only assumption, which can be derived from Asimov datasets [188]. The upper limits produced through this  $p$ -value is called the expected limits.

### 6.6.2 Fit model

The nuisance parameters considered in the fit are the background normalisation  $b$ , which is free-floating, and the background and signal systematics uncertainties as described in section 6.5.2. Tables 6.24, 6.25 and 6.26 show the parameter values of models for  $a \rightarrow gg$ ,  $a \rightarrow qq$  decay modes and SM charmonium states ( $Z\eta_c$  and  $J/\psi$ ).

### 6.6.3 Asimov fits

The Asimov datasets are defined such that the best-fit parameter values are equal to their true values [188], and the event count in each bin is set to the expected event yield for the chosen model. Asimov datasets are generated based on the background-only hypothesis, and also with different signals. These are used to produce expected limits.

A binned likelihood fit is used with 26 bins over the 100-178 GeV  $m_{llj}$  range. For the signal+background test, the 0.5 GeV and the 4.0 GeV  $a \rightarrow gg$  events are injected with a  $\text{BR}(H \rightarrow Za)=1\%$ . Fit distributions for these 2 cases and the background-only fit are shown in Figure 6.45.

### 6.6.4 Fit to data

The real data is used for both background+signal and background only fit. Fit distributions are shown in Figure 6.46.

The fit values for the signal strength  $\mu$ , the background normalization  $B$  and the systematic uncertainties are summarized in the pull plots. The pull for each nuisance parameter  $\theta$  is defined as

$$\text{pull}(\theta) = \frac{\hat{\theta} - \theta_0}{\sigma_\theta}, \quad (6.22)$$

where  $\hat{\theta}$  is the NP obtained from the fit,  $\theta_0$  is the expectation value of the NP, and

Mass (GeV)	Events $N$ ( $\cdot 10^3$ )	$\mu$ (GeV)	$\sigma$ (GeV)	$\sigma_N$ (%)	$\sigma_\mu$ (GeV)	$\sigma_\sigma$ (GeV)
0.5	58	131.5	5.0	+61.3 -61.7	2.7	0.7
1.0	33	130.5	5.1	+61.3 -61.7	2.7	0.7
1.5	24	130.7	5.1	+61.3 -61.7	2.7	0.7
2.0	21	130.1	5.3	+61.3 -61.7	2.8	0.7
2.5	12	129.7	5.5	+65.3 -65.7	2.9	0.7
3.0	6.5	128.4	5.5	+65.3 -65.7	2.2	0.7
3.5	3.9	127.4	5.6	+61.3 -61.7	2.0	0.7
4.0	2.5	126.6	5.2	+56.3 -56.8	2.1	0.7

Table 6.24: Parameter values for signal  $a \rightarrow gg$  models with different mass points.

Mass (GeV)	Events $N$ ( $\cdot 10^3$ )	$\mu$ (GeV)	$\sigma$ (GeV)	$\sigma_N$ (%)	$\sigma_\mu$ (GeV)	$\sigma_\sigma$ (GeV)
1.5	28	129.4	5.4	+23.7 -24.8	2.0	0.7
2.0	19	128.9	5.5	+23.7 -24.8	2.0	0.7
2.5	13	129.1	5.8	+22.8 -23.9	2.1	0.7
3.0	7.9	127.8	5.5	+25.7 -26.7	2.1	0.7
3.5	5.6	127.0	5.0	+28.6 -29.5	1.6	0.9
4.0	0.61	125.1	5.5	+7.7 -10.6	1.5	0.9

Table 6.25: Parameter values for signal  $a \rightarrow qq$  models with different mass points.

	Events $N$ ( $\cdot 10^3$ )	$\mu$ (GeV)	$\sigma$ (GeV)	$\sigma_N$ (%)	$\sigma_\mu$ (GeV)	$\sigma_\sigma$ (GeV)
$\eta_c$	4.6	128.4	5.5	+65.3 -65.7	2.2	0.7
$J/\psi$	4.6	128.9	5.6	+65.3 -65.7	2.2	0.7

Table 6.26: Parameter values for  $H \rightarrow Z\eta_c$  and  $H \rightarrow ZJ/\psi$ .

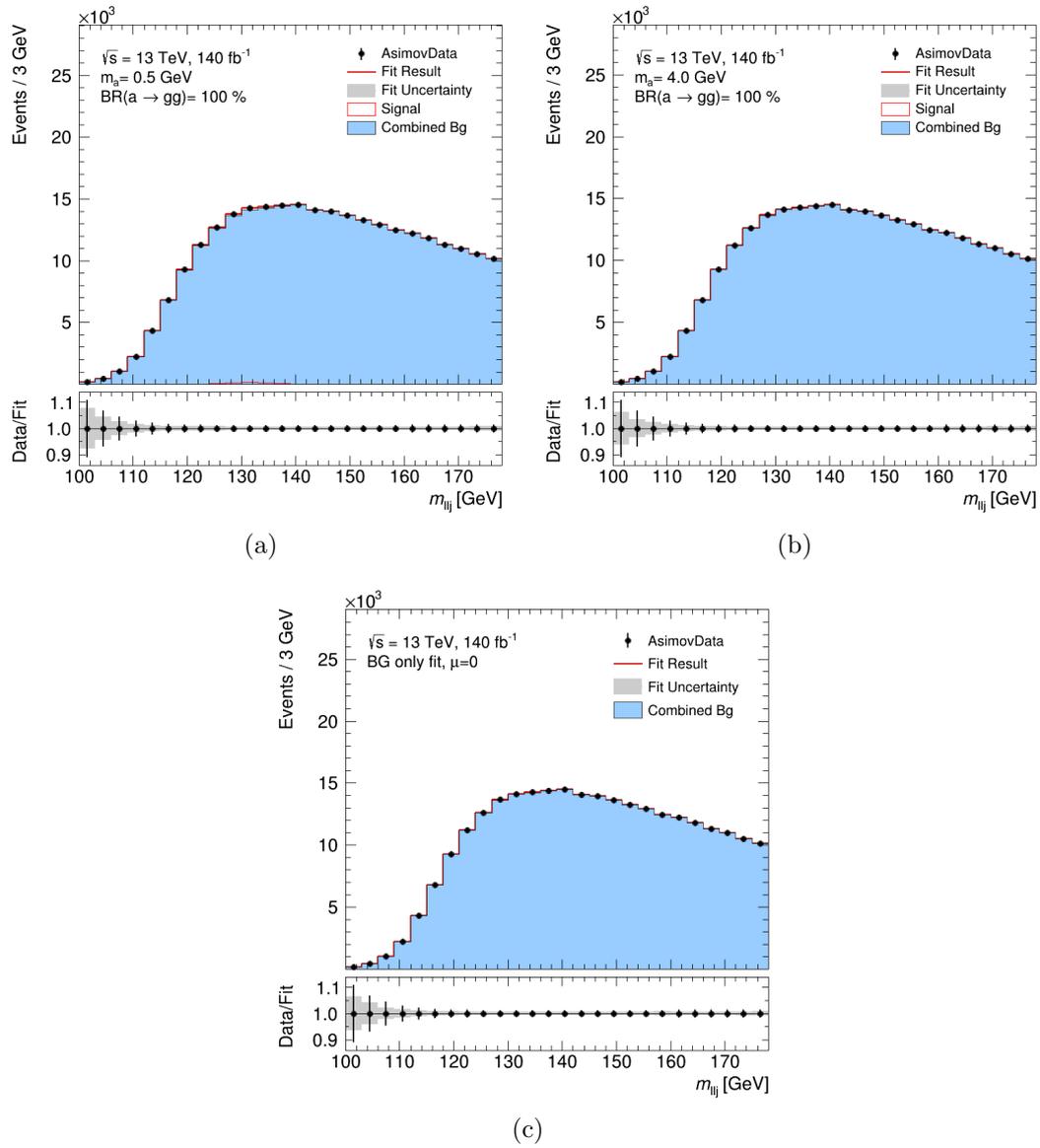


Figure 6.45: Asimov fits for (a) the 0.5 GeV and (b) 4.0 GeV signal model and for (c)  $\mu = 0$  background-only model.

$\sigma_\theta$  is the standard deviation. The NPs are expected to have an error interval of  $[-1, +1]$  with a Gaussian constraint. After the fit, NPs with errors smaller than 1 are called ‘constrained’, which means the fit measurement is more accurate than the auxiliary measurement, i.e. the fit results are closer to the true value.

Correlation matrices are shown in figure 6.48, describing the relationship between systematics.

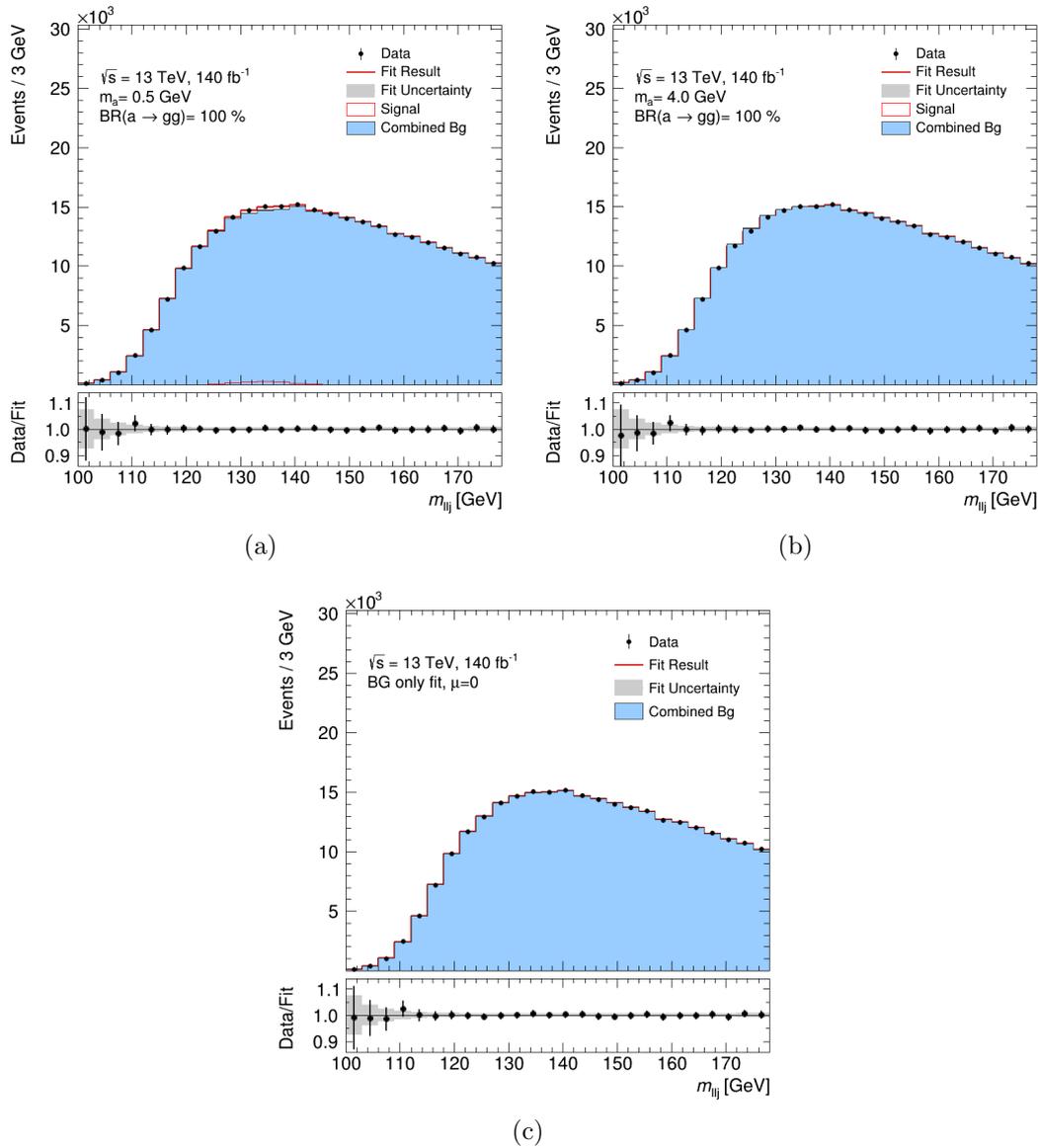


Figure 6.46: Observed data fits for (a) the 0.5 GeV and (b) 4.0 GeV signal model and for (c)  $\mu = 0$  background-only model.

### 6.6.5 Nuisance parameter impact on the signal strength

The impact of a specific NP on the signal strength  $\mu$  is estimated by fixing the NP to the largest (or lowest) constrained value and performing the signal+background fit. The impact is then given by  $\Delta\mu = \mu' - \hat{\mu}$ , where  $\mu'$  is the fit value from the fit with the fixed NP, and  $\hat{\mu}$  is the best-fit value. The impacts of NPs for 0.5 GeV and 4 GeV signals are shown in Figure 6.49.

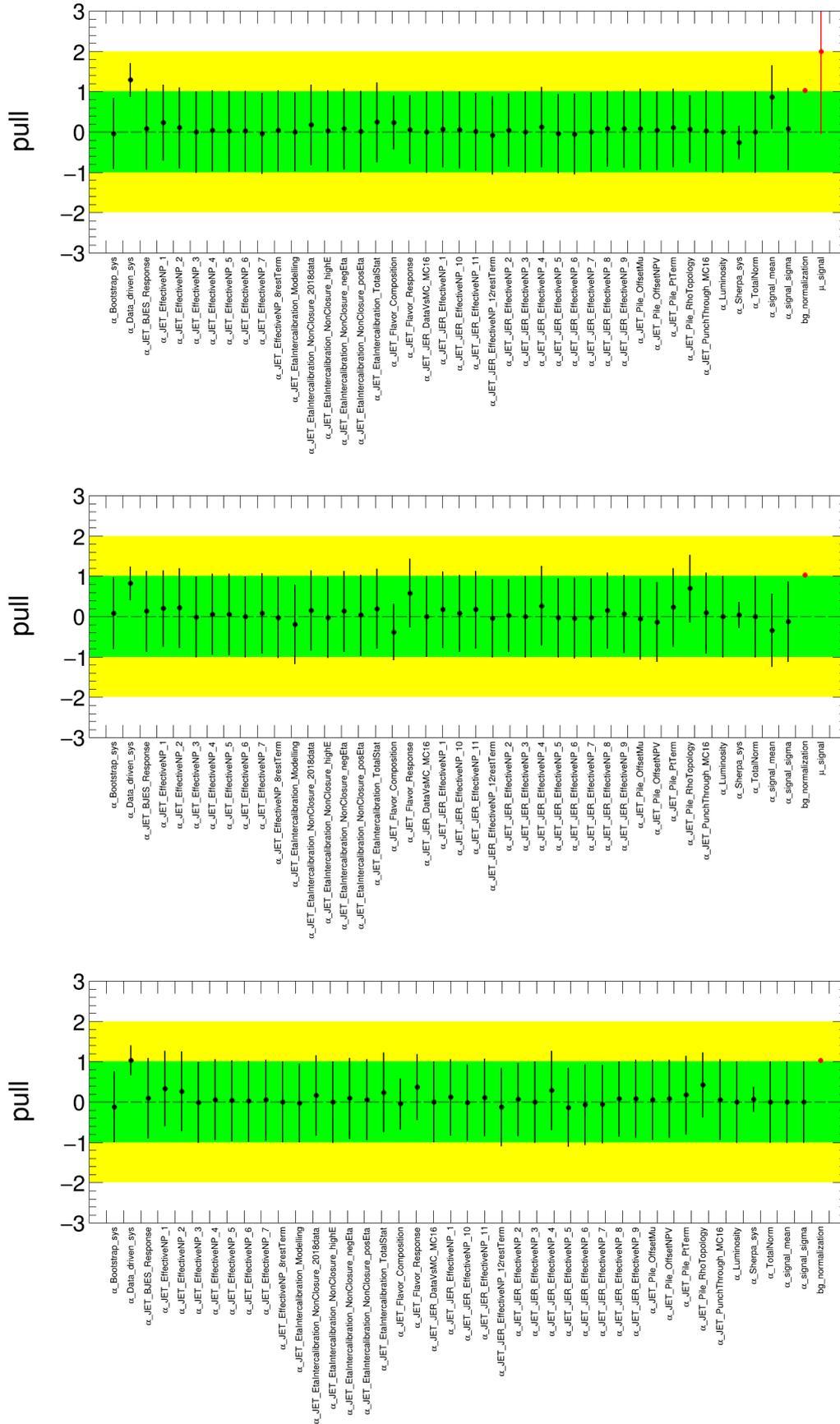


Figure 6.47: Pull plots for (top) 0.5 GeV, (middle) 4 GeV  $a \rightarrow gg$  signal and (bottom)  $\mu = 0$  background-only.



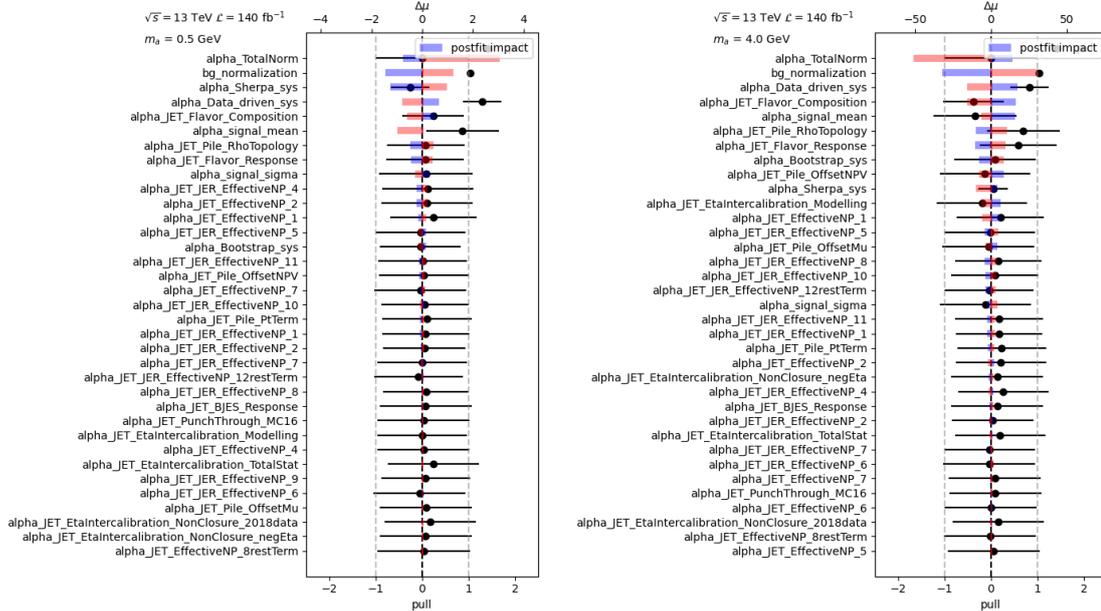


Figure 6.49: Ranking of the NPs according to their impact on the measured signal strength  $\mu$  for (left) 0.5 GeV and (right) 4 GeV  $a \rightarrow gg$  samples. The purple and red rectangles are the post-fit impact and the dots with error bars show the pulls. The background normalization is an unconstrained, free parameter. The ‘pulled’ of 1 indicates that it remains unchanged from its value before the fit.

### 6.6.6 Alternative strategy

An alternative simpler test is also made, where the background modelling uncertainties only come from data-driven. The uncertainty from the reweighting NN performance is calculated from the standard deviation of bootstrapping trials bin-by-bin, and added to the statistical uncertainty quadratically. With this strategy, the upper limits set on  $\text{BR}(H \rightarrow Za)$  at 95% CL for  $a \rightarrow gg$  and  $a \rightarrow qq$  are shown in Table 6.27.

Signal mass (GeV)	$a \rightarrow gg$		$a \rightarrow qq$	
	Exp.(%)	Obs.(%)	Exp.(%)	Obs.(%)
0.5	$5.1^{+2.0}_{-1.4}$	8.7	—	—
1.0	$9.1^{+3.6}_{-2.5}$	14.1	—	—
1.5	$12.5^{+4.9}_{-3.5}$	19.9	$4.9^{+1.9}_{-1.4}$	5.5
2.0	$15.0^{+5.9}_{-4.2}$	22.6	$7.2^{+2.8}_{-2.0}$	7.6
2.5	$31.6^{+12.4}_{-8.8}$	44.9	$11.0^{+4.3}_{-3.1}$	12.0
3.0	$58.0^{+22.7}_{-16.2}$	55.1	$17.8^{+7.0}_{-5.0}$	16.3
3.5	$82.9^{+32.5}_{-23.2}$	66.4	$24.0^{+9.4}_{-6.7}$	19.1
4.0	$101.6^{+39.8}_{-28.4}$	72.4	$204.6^{+80.2}_{-57.2}$	153.6

Table 6.27: Upper limits on  $\text{BR}(H \rightarrow Za)$  at 95% CL for  $a \rightarrow gg$  and  $a \rightarrow qq$  decay modes with the alternative strategy. Both expected and observed limits are shown.

Signal mass (GeV)	$a \rightarrow gg$		$a \rightarrow qq$	
	Exp.(%)	Obs.(%)	Exp.(%)	Obs.(%)
0.5	$7.5^{+2.9}_{-2.1}$	13.9	—	—
1.0	$13.7^{+5.4}_{-3.8}$	23.8	—	—
1.5	$18.7^{+7.3}_{-5.2}$	33.1	$7.6^{+3.0}_{-2.1}$	9.2
2.0	$23.3^{+9.1}_{-6.5}$	39.4	$11.6^{+4.6}_{-3.3}$	12.7
2.5	$50.3^{+19.7}_{-14.1}$	83.2	$17.9^{+7.0}_{-5.0}$	20.7
3.0	$93.7^{+36.7}_{-26.2}$	87.7	$29.4^{+11.5}_{-8.2}$	25.9
3.5	$137.4^{+53.8}_{-38.4}$	100.7	$38.4^{+15.0}_{-10.7}$	28.2
4.0	$165.4^{+64.8}_{-46.2}$	103.6	$364.7^{+142.8}_{-101.9}$	251.9

Table 6.28: Upper limits on  $\text{BR}(H \rightarrow Za)$  at 95% CL for  $a \rightarrow gg$  and  $a \rightarrow qq$  decay modes. Both expected and observed limits are shown.

## 6.7 Results and Interpretations

### 6.7.1 Model Independent Interpretation

Exclusion limits for the  $a \rightarrow gg$  and  $a \rightarrow qq$  are produced. For both cases  $a \rightarrow gg/qq = 100\%$  is assumed. Upper limits are set on  $\text{BR}(H \rightarrow Za)$  at 95% CL. The values of expected and observed 95% CL<sub>s</sub> upper limits on  $\text{BR}(H \rightarrow Za)$  are listed in table 6.28. For 1.5-3.5 GeV signals, the quarks in  $a \rightarrow qq$  decay modes are  $s\bar{s}$ , while for 4.0 GeV signal these are  $c\bar{c}$ .

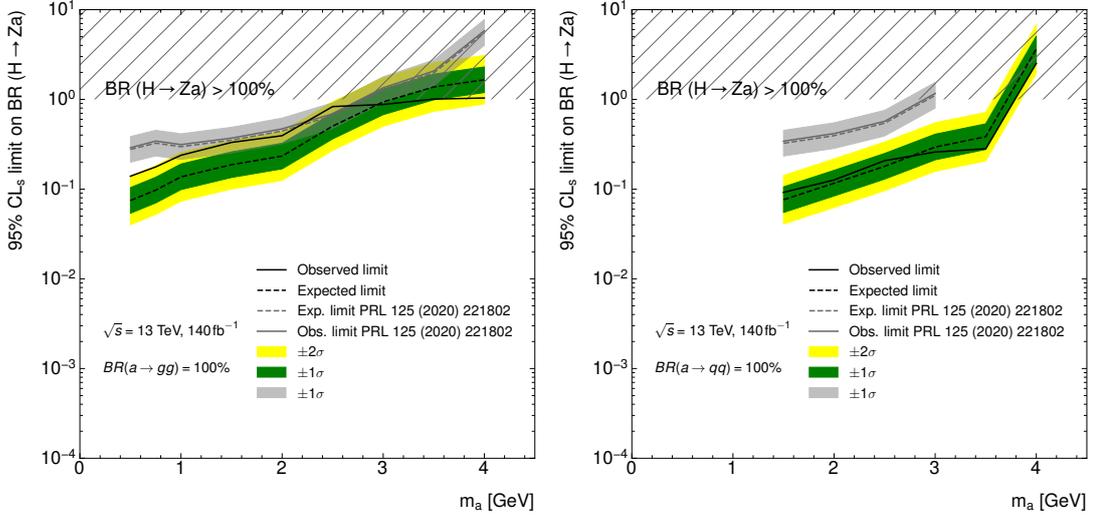


Figure 6.50: Upper limits on  $\text{BR}(H \rightarrow Za)$  95% CL and  $1\sigma$  and  $2\sigma$  bands. The previous result from Ref. [50] is also shown.

Comparing to previous results [50], as shown in Figure 6.50, the observed limits are overall improved. For the low mass ( $< 2$  GeV)  $a \rightarrow gg$  signals, the upper limits are improved 1.2-2.2 times. For the  $a \rightarrow qq$  signals, this analysis gets significant improvement by a factor of 2.9-4.7. Meanwhile, a small excess (less than  $2\sigma$ ) is observed for low mass  $a \rightarrow gg$  decays.

### 6.7.2 $\eta_c$ and $J/\psi$

The modelling of  $H \rightarrow Z\eta_c(J/\psi)$  is derived from PYTHIA8 samples, while the associated uncertainties are taken from the 3 GeV  $a \rightarrow gg$  model, as most of  $\eta_c$  or  $J/\psi$  decay hadronically. Upper limits are set on  $\text{BR}(H \rightarrow Z\eta_c(J/\psi))$  at 95% CL. The limits are shown in Table 6.29.

$H \rightarrow ZQ$	Exp.(%)	Obs.(%)
$\eta_c$	$135.4^{+53.0}_{-37.8}$	126.7
$J/\Psi$	$136.7^{+53.5}_{-38.2}$	148.1

Table 6.29: Upper limits of  $\text{BR}(H \rightarrow Z\eta_c)$  and  $\text{BR}(H \rightarrow ZJ/\psi)$  at 95% CL. Both expected and observed limits are shown.

### 6.7.3 ALP Interpretation

The width of the  $H \rightarrow Za$  decay for an ALP  $a$  depends on the strength of the effective coupling  $C_{Zh}^{\text{eff}}$  [139]:

$$\Gamma(H \rightarrow Za) = \frac{m_H^3}{16\pi\Lambda^2} |C_{Zh}^{\text{eff}}|^2 \lambda^{3/2} \left( \frac{m_Z^2}{m_H^2}, \frac{m_a^2}{m_H^2} \right), \quad (6.23)$$

where

$$\lambda(x, y) = (1 - x - y)^2 - 4xy. \quad (6.24)$$

If this decay is possible it will contribute to the Higgs width and the BR will be

$$BR(H \rightarrow Za) = \frac{\Gamma(H \rightarrow Za)}{\Gamma_{SM}(H) + \Gamma(H \rightarrow Za)} \quad (6.25)$$

Then the limit on  $C_{Zh}^{\text{eff}}/\Lambda$  can be set from the limit on  $BR(H \rightarrow Za)$ . Assuming a ALP with relevant effective Wilson coefficients equal to 1 [139], the  $a \rightarrow 3\pi$  is one of the main decay modes. Since  $3\pi^0$  events lack inner tracks, only  $\pi^+\pi^-\pi^0$  events are considered here. For the ALP  $m_a = 0.5$  and  $1.0$  GeV,  $BR(a \rightarrow \pi^+\pi^-\pi^0)$  are 10% and 20% respectively, the corresponding limits on  $BR(H \rightarrow Za)$  are set on 45% and 81%. Therefore, the upper limits on  $C_{Zh}^{\text{eff}}/\Lambda$  at 95% CL are set on  $2.0 \text{ TeV}^{-1}$  for  $m_a = 0.5$  GeV and  $0.89 \text{ TeV}^{-1}$  for  $m_a = 1.0$  GeV.

---

## Conclusion

---

Since the Higgs boson being discovered in 2012, both ATLAS and CMS experiments have performed a variety of measurements of Higgs properties and Higgs decays. One of recent hotspots is to search for additional (pseudo)scalars predicted by BSM and exotic Higgs decays.

In this thesis, a search for the decays of Higgs boson to a  $Z$  boson and a light hadronically decaying resonance ( $a$ ) is performed, while the full Run 2 data recorded by the ATLAS detector is used. 95%  $\text{CL}_s$  upper limits on  $\text{BR}(H \rightarrow Za)$  are set for  $a \rightarrow gg$  with  $m_a \leq 4$  GeV, and  $a \rightarrow qq$  with  $m_a$  in the range of 1.5-4 GeV. The limits of  $\text{BR}(H \rightarrow Za)$  start from 13.9% for the  $a \rightarrow gg$  channel, and start from 9.2% for  $a \rightarrow qq$  channel. A slight excess with a significance of  $< 2\sigma$  is found for an  $a$  mass hypothesis of less than 2 GeV for the  $a \rightarrow gg$  decay channel. Comparing to previous ATLAS results, these upper limits show an improvement by a factor of 1.2-2.2 for the lower mass  $a \rightarrow gg$  decays, and an overall improvement by a factor of 2.9-4.7 for  $a \rightarrow qq$  decays.

Despite the significant progress achieved, there is still room for future improvements. The limits set in this thesis are affected significantly by the difference signal modelling between PYTHIA and HERWIG samples, which provides a huge theoretical uncertainty on the normalization. This is because the efficiency of Classification NN is much lower for HERWIG samples than nominal PYTHIA samples, so that the yields of signal are much lower for HERWIG samples. The main reason is that the PYTHIA and HERWIG signal samples have different complexities in ghost-associated tracks, while the Classification NN is only trained with jet substructure variables of PYTHIA signal events. Additionally, the limits set for higher mass  $a$  are not as good as lower mass. This is mainly caused by the low Classification NN significance for higher mass signals. To make the analysis simpler, a single Classification cut is used for every signal mass. To cut at a good significance for lower mass signals which we are interested in more, the significance for higher mass signals is more or less sacrificed. Therefore, the Classification NN cut does not work very well for higher mass signals.

In conclusion, this thesis makes an improvement on the search for  $H \rightarrow Za \rightarrow ll + \text{jet}$  for low mass  $a$ . Further improvement on the results can be made in the future, if the strategy is further optimized.



---

## REFERENCES

---

- [1] F. Englert and R. Brout, “Broken Symmetry and the Mass of Gauge Vector Mesons,” *Phys. Rev. Lett.*, vol. 13, pp. 321–323, Aug 1964.
- [2] P. W. Higgs, “Broken Symmetries and the Masses of Gauge Bosons,” *Phys. Rev. Lett.*, vol. 13, pp. 508–509, 1964.
- [3] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, “Global conservation laws and massless particles,” *Phys. Rev. Lett.*, vol. 13, pp. 585–587, Nov 1964.
- [4] ATLAS Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,” *Physics Letters B*, vol. 716, no. 1, pp. 1–29, 2012.
- [5] CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,” *Physics Letters B*, vol. 716, no. 1, pp. 30–61, 2012.
- [6] “The Nobel Prize in Physics 2013.” <https://www.nobelprize.org/prizes/physics/2013/summary/>. Accessed: 2024-08-30.
- [7] ATLAS Collaboration, “A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery,” *Nature*, vol. 607, no. 7917, pp. 52–59, 2022.
- [8] CMS Collaboration, “A portrait of the Higgs boson by the CMS experiment ten years after the discovery,” *Nature*, vol. 607, no. 7917, pp. 60–68, 2022.
- [9] D. Curtin *et al.*, “Exotic decays of the 125 GeV higgs boson,” *Physical Review D*, vol. 90, oct 2014.

- 
- [10] N. Turok and J. Zadrozny, “Electroweak baryogenesis in the two-doublet model,” *Nuclear Physics B*, vol. 358, no. 2, pp. 471–493, 1991.
- [11] J. E. Kim, “Light pseudoscalars, particle physics and cosmology,” *Physics Reports*, vol. 150, no. 1, pp. 1–177, 1987.
- [12] H. Haber and G. Kane, “The search for supersymmetry: Probing physics beyond the standard model,” *Physics Reports*, vol. 117, no. 2, pp. 75–263, 1985.
- [13] G. Branco *et al.*, “Theory and phenomenology of two-higgs-doublet models,” *Physics Reports*, vol. 516, no. 1, pp. 1–102, 2012.
- [14] Wikipedia, “Standard model.” [https://en.wikipedia.org/wiki/Standard\\_Model](https://en.wikipedia.org/wiki/Standard_Model). Accessed: 2024-08-30.
- [15] T. Kajita, “Discovery of neutrino oscillations,” *Reports on Progress in Physics*, vol. 69, p. 1607, may 2006.
- [16] M. Thomson, *Modern Particle Physics*. Cambridge University Press, 2013.
- [17] S. L. Glashow, “Partial-symmetries of weak interactions,” *Nuclear Physics*, vol. 22, no. 4, pp. 579–588, 1961.
- [18] S. Weinberg, “A Model of Leptons,” *Phys. Rev. Lett.*, vol. 19, pp. 1264–1266, Nov 1967.
- [19] A. Salam, “Weak and Electromagnetic Interactions,” *Conf. Proc. C*, vol. 680519, pp. 367–377, 1968.
- [20] A. Pich, “The Standard Model of Electroweak Interactions,” 2012. ArXiv:1201.0537.
- [21] D. Griffiths, *Introduction to elementary particles*. John Wiley & Sons, 2020.
- [22] F. Hanzel and A. D. Martin, *Quarks and Leptons: An introductory course in modern particle physics*. John Wiley & Sons, 1984.
- [23] J. Goldstone *et al.*, “Broken Symmetries,” *Phys. Rev.*, vol. 127, pp. 965–970, Aug 1962.
- [24] ATLAS Collaboration, “Combined Measurement of the Higgs Boson Mass from the  $H \rightarrow \gamma\gamma$  and  $H \rightarrow ZZ^* \rightarrow 4l$  Decay Channels with the ATLAS Detector Using  $\sqrt{s} = 7, 8,$  and  $13$  TeV  $pp$  Collision Data,” *Phys. Rev. Lett.*, vol. 131, p. 251802, Dec 2023.
- [25] H. M. Georgi *et al.*, “Higgs Bosons from Two-Gluon Annihilation in Proton-Proton Collisions,” *Phys. Rev. Lett.*, vol. 40, pp. 692–694, Mar 1978.
- [26] M. Ciccolini, A. Denner, and S. Dittmaier, “Strong and Electroweak Corrections to the Production of a Higgs Boson + 2 Jets via Weak Interactions at the Large Hadron Collider,” *Phys. Rev. Lett.*, vol. 99, p. 161803, Oct 2007.

- [27] A. Denner *et al.*, “Electroweak corrections to Higgs-strahlung off W/Z bosons at the Tevatron and the LHC with Hawk,” *Journal of High Energy Physics*, vol. 2012, Mar. 2012.
- [28] ATLAS Collaboration, “Observation of  $H \rightarrow bb$  decays and VH production with the ATLAS detector,” *Physics Letters B*, vol. 786, pp. 59–86, 2018.
- [29] D. de Florian *et al.*, “Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector,” *CERN Yellow Rep. Monogr.*, 2017.
- [30] ATLAS Collaboration, “ATLAS searches for additional scalars and exotic Higgs boson decays with the LHC Run 2 dataset,” 2024. <https://arxiv.org/abs/2405.04914>.
- [31] G. Carratta, “Exotic Higgs decays, including new scalars (ATLAS+CMS),” *PoS*, vol. LHCP2022, p. 223, 2023.
- [32] A. Albert *et al.*, “Recommendations of the LHC Dark Matter Working Group: Comparing LHC searches for heavy mediators of dark matter production in visible and invisible decay channels,” 2017. ArXiv:1703.05703.
- [33] S. Baum and N. R. Shah, “Two Higgs doublets and a complex singlet: disentangling the decay topologies and associated phenomenology,” *Journal of High Energy Physics*, vol. 2018, Dec. 2018.
- [34] J. F. Gunion *et al.*, *The Higgs Hunter’s Guide*, vol. 80. CRC Press, 2000.
- [35] V. Keus, S. F. King, and S. Moretti, “Three-Higgs-doublet models: symmetries, potentials and Higgs boson masses,” *Journal of High Energy Physics*, vol. 2014, Jan. 2014.
- [36] H. Georgi and M. Machacek, “Doubly charged Higgs bosons,” *Nuclear Physics B*, vol. 262, no. 3, pp. 463–477, 1985.
- [37] H. E. Haber and R. Hempfling, “Renormalization-group-improved Higgs sector of the minimal supersymmetric model,” *Physical Review D*, vol. 48, p. 4280–4309, Nov. 1993.
- [38] ATLAS Collaboration, “Combined measurements of Higgs boson production and decay using up to  $139 \text{ fb}^{-1}$  of proton-proton collision data at  $\sqrt{s} = 13$  TeV collected with the ATLAS experiment,” 2021. ATLAS-CONF-2021-053.
- [39] U. Ellwanger, C. Hugonie, and A. M. Teixeira, “The Next-to-Minimal Supersymmetric Standard Model,” *Physics Reports*, vol. 496, p. 1–77, Nov. 2010.
- [40] ATLAS Collaboration, “Search for Higgs boson decays into a pair of pseudoscalar particles in the  $bb\mu\mu$  final state with the ATLAS detector in  $pp$  collisions at  $\sqrt{s} = 13$  TeV,” *Phys. Rev. D*, vol. 105, p. 012006, Jan 2022.

- 
- [41] ATLAS Collaboration, “Search for Higgs boson decays into two new low-mass spin-0 particles in the  $4b$  channel with the ATLAS detector using  $pp$  collisions at  $\sqrt{s} = 13$  TeV,” *Phys. Rev. D*, vol. 102, p. 112006, Dec 2020.
- [42] ATLAS Collaboration, “Search for the Higgs boson produced in association with a  $W$  boson and decaying to four  $b$ -quarks via two spin-zero particles in  $pp$  collisions at 13 TeV with the ATLAS detector,” *Eur. Phys. J.*, vol. C76, no. 11, p. 605, 2016.
- [43] ATLAS Collaboration, “Search for Higgs boson decays into a pair of light bosons in the  $bb\mu\mu$  final state in  $pp$  collision at  $\sqrt{s} = 13$  TeV with the ATLAS detector,” *Phys. Lett.*, vol. B790, pp. 1–21, 2019.
- [44] ATLAS Collaboration, “Search for Higgs bosons decaying into new spin-0 or spin-1 particles in four-lepton final states with the ATLAS detector with 139  $fb^{-1}$  of  $pp$  collision data at  $\sqrt{s} = 13$  TeV,” *Journal of High Energy Physics*, vol. 2022, Mar. 2022.
- [45] CMS Collaboration, “Search for light pseudoscalar boson pairs produced from decays of the 125 GeV Higgs boson in final states with two muons and two nearby tracks in  $pp$  collisions at  $\sqrt{s} = 13$  TeV,” *Phys. Lett. B*, vol. 800, p. 135087, 2020.
- [46] CMS Collaboration, “Search for exotic Higgs boson decays to a pair of pseudoscalars in the  $\mu\mu bb$  and  $\tau\tau bb$  final states in proton-proton collisions with the CMS experiment,” *The European Physical Journal C*, vol. 84, no. 5, p. 493, 2024.
- [47] M. Cepeda *et al.*, “Exotic Higgs Decays,” *Annual Review of Nuclear and Particle Science*, vol. 72, p. 119–149, Sept. 2022.
- [48] ATLAS Collaboration, “Search for the decay of the Higgs boson to a  $Z$  boson and a light pseudoscalar particle decaying to two photons,” *Physics Letters B*, vol. 850, p. 138536, 2024.
- [49] CMS Collaboration, “Search for an exotic decay of the Higgs boson into a  $Z$  boson and a pseudoscalar particle in proton-proton collisions at  $\sqrt{s} = 13$  TeV,” *Physics Letters B*, vol. 852, p. 138582, 2024.
- [50] ATLAS Collaboration, “Search for Higgs Boson Decays into a  $Z$  Boson and a Light Hadronically Decaying Resonance Using 13 TeV  $pp$  Collision Data from the ATLAS Detector,” *Phys. Rev. Lett.*, vol. 125, p. 221802, Nov 2020.
- [51] L. Evans and P. Bryant, “LHC Machine,” *Journal of Instrumentation*, vol. 3, p. S08001, Aug 2008.
- [52] ATLAS Collaboration, “The ATLAS Experiment at the CERN Large Hadron Collider,” *Journal of Instrumentation*, vol. 3, p. S08003, Aug 2008.

- [53] CMS Collaboration, “The CMS experiment at the CERN LHC,” *Journal of Instrumentation*, vol. 3, p. S08004, Aug 2008.
- [54] LHCb Collaboration, “The LHCb Detector at the LHC,” *Journal of Instrumentation*, vol. 3, p. S08005, Aug 2008.
- [55] ALICE Collaboration, “The ALICE experiment at the CERN LHC,” *Journal of Instrumentation*, vol. 3, p. S08002, Aug 2008.
- [56] E. Lopienska, “The CERN accelerator complex, layout in 2022. Complexe des accélérateurs du CERN en janvier 2022,” 2022. CERN-GRAPHICS-2022-001. <https://cds.cern.ch/record/2800984>.
- [57] ATLAS Collaboration, “ATLAS data quality operations and performance for 2015–2018 data-taking,” *JINST*, vol. 15, no. 04, p. P04003, 2020.
- [58] ATLAS Collaboration, “Public ATLAS Luminosity Results for Run-3 of the LHC.” <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun3>.
- [59] ATLAS Collaboration, “The ATLAS Experiment at the CERN Large Hadron Collider: A Description of the Detector Configuration for Run 3,” *Journal of Instrumentation*, vol. 19, p. P05063, may 2024.
- [60] ATLAS Collaboration, *ATLAS inner detector: Technical Design Report, 1*. Technical design report. ATLAS, Geneva: CERN, 1997. CERN-LHCC-97-016.
- [61] ATLAS Collaboration, *ATLAS inner detector: Technical Design Report, 2*. Geneva: CERN, 1997. CERN-LHCC-97-017, ATLAS-TDR-5.
- [62] ATLAS Collaboration, “ATLAS pixel detector electronics and sensors,” *Journal of Instrumentation*, vol. 3, p. P07007, jul 2008.
- [63] The ATLAS IBL Collaboration, “Production and integration of the ATLAS Insertable B-Layer,” *Journal of Instrumentation*, vol. 13, p. T05008–T05008, May 2018.
- [64] ATLAS Collaboration, “Operation and performance of the ATLAS semiconductor tracker,” *Journal of Instrumentation*, vol. 9, p. P08009, Aug 2014.
- [65] ATLAS TRT collaboration , “The ATLAS Transition Radiation Tracker (TRT) proportional drift tube: design and performance,” *Journal of Instrumentation*, vol. 3, p. P02013, Feb 2008.
- [66] ATLAS Collaboration, “Performance of the ATLAS Transition Radiation Tracker in Run 1 of the LHC: tracker properties,” *Journal of Instrumentation*, vol. 12, p. P05002–P05002, May 2017.

- 
- [67] ATLAS collaboration, *ATLAS liquid-argon calorimeter: Technical Design Report*. Technical design report. ATLAS, Geneva: CERN, 1996. CERN-LHCC-96-41.
- [68] ATLAS collaboration, *ATLAS Tile calorimeter: Technical Design Report*. Technical design report. ATLAS, Geneva: CERN, 1996. CERN-LHCC-96-42.
- [69] D. Mahon, “ATLAS LAr calorimeter performance in LHC Run 2,” *Journal of Instrumentation*, vol. 15, p. C06045, Jun 2020.
- [70] ATLAS Collaboration, “Operation and performance of the ATLAS Tile calorimeter in LHC Run 2,” 2024. ArXiv:2401.16034.
- [71] *ATLAS muon spectrometer: Technical Design Report*. Technical Design Report. ATLAS, Geneva: CERN, 1997. CERN-LHCC-97-022, ATLAS-TDR-10.
- [72] T. Kawamoto *et al.*, “ATLAS New Small Wheel Technical Design Report,” tech. rep., 2013. CERN-LHCC-2013-006, ATLAS-TDR-020.
- [73] ATLAS Collaboration, “Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at  $\sqrt{s} = 13$  TeV,” *The European Physical Journal C*, vol. 81, July 2021.
- [74] ATLAS Collaboration, “Technical Design Report for the Phase-I Upgrade of the ATLAS TDAQ System,” tech. rep., 2013. CERN-LHCC-2013-018, ATLAS-TDR-023.
- [75] ATLAS collaboration, “The ATLAS Fast Tracker system,” *Journal of Instrumentation*, vol. 16, p. P07006, jul 2021.
- [76] ATLAS collaboration, “Operation of the ATLAS trigger system in Run 2,” *JINST*, vol. 15, no. 10, p. P10004, 2020.
- [77] ATLAS collaboration, “Performance of electron and photon triggers in ATLAS during LHC Run 2,” *The European Physical Journal C*, vol. 80, Jan. 2020.
- [78] ATLAS collaboration, “Performance of the ATLAS muon triggers in Run 2,” *Journal of Instrumentation*, vol. 15, p. P09015–P09015, Sept. 2020.
- [79] ATLAS Collaboration, “Alignment of the ATLAS Inner Detector in Run 2,” *The European Physical Journal C*, vol. 80, Dec. 2020.
- [80] ATLAS Software Documentation, “Atlas tracking software tutorial.” <https://atlassoftwaredocs.web.cern.ch/trackingTutorial/index.html>.
- [81] T. Cornelissen *et al.*, “The new ATLAS track reconstruction (NEWT),” *J. Phys. Conf. Ser.*, vol. 119, p. 032014, 2008.
- [82] ATLAS Collaboration, “Early Inner Detector Tracking Performance in the 2015 data at  $\sqrt{s} = 13$  TeV,” tech. rep., CERN, Geneva, 2015. ATL-PHYS-PUB-2015-051.

- [83] F. Meloni, “Primary vertex reconstruction with the ATLAS detector,” *Journal of Instrumentation*, vol. 11, p. C12060, dec 2016.
- [84] ATLAS Collaboration, “Primary vertex reconstruction at the ATLAS experiment,” *Journal of Physics: Conference Series*, vol. 898, p. 042056, oct 2017.
- [85] ATLAS Collaboration, “Vertex Reconstruction Performance of the ATLAS Detector at  $\sqrt{s} = 13$  TeV,” tech. rep., CERN, Geneva, 2015. ATL-PHYS-PUB-2015-026.
- [86] W. Lampl *et al.*, “Calorimeter clustering algorithms: Description and performance,” 5 2008. ATL-LARG-PUB-2008-002, ATL-COM-LARG-2008-003.
- [87] ATLAS Collaboration, “Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at  $\sqrt{s} = 13$  TeV,” *The European Physical Journal C*, vol. 79, Aug. 2019.
- [88] ATLAS Collaboration, “Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung,” tech. rep., CERN, Geneva, 2012. ATLAS-CONF-2012-047.
- [89] ATLAS Collaboration, “Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data,” *Journal of Instrumentation*, vol. 14, p. P12006, dec 2019.
- [90] ATLAS Collaboration, “Electron and photon efficiencies in LHC Run 2 with the ATLAS experiment,” *Journal of High Energy Physics*, vol. 2024, no. 5.
- [91] ATLAS Collaboration, “Muon reconstruction performance of the ATLAS detector in proton–proton collision data at  $\sqrt{s} = 13$  TeV,” *The European Physical Journal C*, vol. 76, May 2016.
- [92] ATLAS Collaboration, “Search for high-mass dilepton resonances in  $pp$  collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector,” *Phys. Rev. D*, vol. 90, p. 052005, Sep 2014.
- [93] ATLAS Collaboration, “Search for new particles in events with one lepton and missing transverse momentum in  $pp$  collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector,” *Journal of High Energy Physics*, vol. 2014, Sept. 2014.
- [94] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- $k_t$  jet clustering algorithm,” *Journal of High Energy Physics*, vol. 2008, p. 063–063, Apr. 2008.
- [95] ATLAS Collaboration, “Jet energy scale and resolution measured in proton–proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector,” *The European Physical Journal C*, vol. 81, Aug. 2021.
- [96] ATLAS Collaboration, “Jet reconstruction and performance using particle flow with the ATLAS Detector,” *Eur. Phys. J. C*, vol. 77, no. 7, p. 466, 2017.

- 
- [97] ATLAS Collaboration, “Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at  $\sqrt{s} = 13$  TeV,” *Eur. Phys. J. C*, vol. 78, no. 11, p. 903, 2018.
- [98] ATLAS Collaboration, “Tagging and suppression of pileup jets,” 5 2014. ATLAS-CONF-2014-018.
- [99] ATLAS Collaboration, “EM+JES and PFlow Pile-Up jet rate - LHCP 2017.” <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/JETM-2017-006/>.
- [100] M. Cacciari and G. P. Salam, “Pileup subtraction using jet areas,” *Physics Letters B*, vol. 659, no. 1, pp. 119–126, 2008.
- [101] ATLAS Collaboration, “Performance of jet substructure techniques for large-R jets in proton-proton collisions at  $\sqrt{s} = 7$  TeV using the ATLAS detector,” *Journal of High Energy Physics*, vol. 2013, Sept. 2013.
- [102] A. Collaboration, “The performance of missing transverse momentum reconstruction and its significance with the ATLAS detector using  $140 \text{ fb}^{-1}$  of  $\sqrt{s} = 13$  TeV  $pp$  collisions,” 2024. ArXiv:2402.05858.
- [103] T. Mkrtchyan, “The Phase-I Upgrade of the ATLAS Level-1 Calorimeter Trigger,” tech. rep., CERN, Geneva, 2022. ATL-DAQ-PROC-2022-027.
- [104] W. Qian, “Design and test performance of the ATLAS Feature Extractor trigger boards for the Phase-1 Upgrade,” tech. rep., CERN, Geneva, 2017. ATL-DAQ-PROC-2016-024.
- [105] ATLAS Collaboration, “ATLAS Liquid Argon Calorimeter Phase-II Upgrade: Technical Design Report,” tech. rep., CERN, Geneva, 2017. CERN-LHCC-2017-018, ATLAS-TDR-027.
- [106] W. Qian, “eFEX modules manufacturing and test at RAL — Phase-1 Upgrade.” ATLAS-PHOTO-2021-079. <https://cds.cern.ch/record/2778165>, 2021.
- [107] W. Qian, “Design and test of the electron Feature EXtractor (eFEX) preproduction module for the ATLAS Phase-I Upgrade,” 2019. ATL-DAQ-SLIDE-2019-642.
- [108] ATLAS Collaboration, “Level-1 calorimeter trigger public results.” <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/L1CaloTriggerPublicResults>. Accessed: 2024-08-30.
- [109] W. Buttinger, “The ATLAS Level-1 Trigger System,” 2012. ATL-DAQ-PROC-2012-024, ATL-COM-DAQ-2012-080.
- [110] W. Cleland and E. Stern, “Signal processing considerations for liquid ionization calorimeters in a high rate environment,” *NIMA: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 338, no. 2, pp. 467–497, 1994.

- [111] ATLAS Collaboration, “ATLAS Liquid Argon Calorimeter Phase-I Upgrade: Technical Design Report,” tech. rep., 2013. CERN-LHCC-2013-017, ATLAS-TDR-022.
- [112] A. M. TURING, “COMPUTING MACHINERY AND INTELLIGENCE,” *Mind*, vol. LIX, pp. 433–460, 10 1950.
- [113] B. Mahesh, “Machine learning algorithms-a review,” *International Journal of Science and Research (IJSR)*, vol. 9, no. 1, pp. 381–386, 2020.
- [114] K. Albertsson *et al.*, “Machine Learning in High Energy Physics Community White Paper,” *Journal of Physics: Conference Series*, vol. 1085, p. 022008, sep 2018.
- [115] A. Radovic *et al.*, “Machine learning at the energy and intensity frontiers of particle physics,” *Nature*, vol. 560, no. 7716, pp. 41–48, 2018.
- [116] F. Rosenblatt, “The perceptron - A perceiving and recognizing automaton,” Tech. Rep. 85-460-1, Cornell Aeronautical Laboratory, Ithaca, New York, January 1957.
- [117] A. LeNail, “NN-SVG: Publication-Ready Neural Network Architecture Schematics,” *Journal of Open Source Software*, vol. 4, no. 33, p. 747, 2019.
- [118] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [119] J. Gu *et al.*, “Recent advances in convolutional neural networks,” *Pattern recognition*, vol. 77, pp. 354–377, 2018.
- [120] L. R. Medsker *et al.*, “Recurrent neural networks,” *Design and Applications*, vol. 5, no. 64-67, p. 2, 2001.
- [121] A. Vaswani, “Attention is all you need,” *Advances in Neural Information Processing Systems*, 2017.
- [122] T. Wolf *et al.*, “Transformers: State-of-the-Art Natural Language Processing,” (Online), pp. 38–45, Association for Computational Linguistics, Oct. 2020.
- [123] P. J. Huber, “Robust Estimation of a Location Parameter,” *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73 – 101, 1964.
- [124] L. Rosasco *et al.*, “Are Loss Functions All the Same?,” *Neural computation*, vol. 16, pp. 1063–76, 06 2004.
- [125] D. E. Rumelhart *et al.*, “Learning representations by back-propagating errors,” *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.
- [126] H. Robbins and S. Monro, “A stochastic approximation method,” *The Annals of Mathematical Statistics*, vol. 22, no. 3, pp. 400–407, 1951.

- 
- [127] Y. Liu *et al.*, “An improved analysis of stochastic gradient descent with momentum,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 18261–18271, 2020.
- [128] J. Duchi *et al.*, “Adaptive Subgradient Methods for Online Learning and Stochastic Optimization,” *Journal of Machine Learning Research*, vol. 12, no. 61, pp. 2121–2159, 2011.
- [129] G. Hinton *et al.*, “Neural networks for machine learning lecture 6a overview of mini-batch gradient descent,” Available at [https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture\\_slides\\_lec6.pdf](https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf).
- [130] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” 2017. ArXiv:1412.6980.
- [131] R. Tibshirani, “Regression Shrinkage and Selection via the Lasso,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.
- [132] A. E. Hoerl and R. W. Kennard, “Ridge Regression: Biased Estimation for Nonorthogonal Problems,” *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.
- [133] N. Srivastava *et al.*, “Dropout: a simple way to prevent neural networks from overfitting,” *J. Mach. Learn. Res.*, vol. 15, p. 1929–1958, Jan 2014.
- [134] J. Bergstra and Y. Bengio, “Random search for hyper-parameter optimization,” *Journal of Machine Learning Research*, vol. 13, p. 281–305, Feb 2012.
- [135] J. Snoek, H. Larochelle, and R. P. Adams, “Practical Bayesian Optimization of Machine Learning Algorithms,” 2012. ArXiv:1206.2944.
- [136] J. Bergstra *et al.*, “Algorithms for hyper-parameter optimization,” in *Proceedings of the 24th International Conference on Neural Information Processing Systems*, NIPS’11, p. 2546–2554, Curran Associates Inc., 2011.
- [137] F. Hutter *et al.*, “Sequential Model-Based Optimization for General Algorithm Configuration,” in *Learning and Intelligent Optimization*, pp. 507–523, Springer Berlin Heidelberg, 2011.
- [138] ATLAS Collaboration, “Search for NMSSM Higgs bosons in the  $H \rightarrow aa \rightarrow \mu\mu\mu\mu, \mu\mu\tau\tau$  channels using  $pp$  collisions at  $\sqrt{s} = 1.96$  TeV,” *Phys. Rev. Lett.*, vol. 103, p. 061801, 2009.
- [139] M. Bauer *et al.*, “Collider probes of axion-like particles,” *Journal of High Energy Physics*, vol. 2017, dec 2017.
- [140] CMS Collaboration, “Search for the decay of the Higgs boson to a pair of light pseudoscalar bosons in the final state with four bottom quarks in proton-proton collisions at  $\sqrt{s} = 13$  TeV,” *Journal of High Energy Physics*, vol. 2024, June 2024.

- [141] CMS Collaboration, “Search for an exotic decay of the Higgs boson to a pair of light pseudoscalars in the final state with two b quarks and two  $\tau$  leptons in proton-proton collisions at  $\sqrt{s} = 13$  TeV,” *Phys. Lett. B*, vol. 785, p. 462, 2018.
- [142] CMS Collaboration, “Search for an exotic decay of the Higgs boson to a pair of light pseudoscalars in the final state with two muons and two b quarks in pp collisions at 13 TeV,” *Phys. Lett. B*, vol. 795, pp. 398–423, 2019.
- [143] CMS Collaboration, “Search for exotic decays of the Higgs boson to a pair of new light bosons with two muon and two b jets in final states,” tech. rep., CERN, Geneva, 2016. CMS-PAS-HIG-14-041.
- [144] CMS Collaboration, “Search for a very light NMSSM Higgs boson produced in decays of the 125 GeV scalar boson and decaying into  $\tau$  leptons in pp collisions at  $\sqrt{s} = 8$  TeV,” *JHEP*, vol. 01, p. 079, 2016.
- [145] CMS Collaboration, “Search for Higgs Decays to New Light Bosons in Boosted Tau Final States,” Tech. Rep. CMS-PAS-HIG-14-022, CERN, Geneva, 2015.
- [146] ATLAS Collaboration, “Search for Higgs bosons decaying to  $aa$  in the  $\mu\mu\tau\tau$  final state in  $pp$  collisions at  $\sqrt{s} = 8$  TeV with the ATLAS experiment,” *Phys. Rev. D*, vol. 92, p. 052002, Sep 2015.
- [147] CMS Collaboration, “Search for an exotic decay of the Higgs boson to a pair of light pseudoscalars in the final state of two muons and two  $\tau$  leptons in proton-proton collisions at  $\sqrt{s} = 13$  TeV,” *JHEP*, vol. 11, p. 018, 2018.
- [148] CMS Collaboration, “Search for the exotic decay of the Higgs boson to two light pseudoscalar bosons with two taus and two muons in the final state at  $\sqrt{s} = 8$  TeV,” Tech. Rep. CMS-PAS-HIG-15-011, CERN, Geneva, 2016.
- [149] CMS Collaboration, “A search for pair production of new light bosons decaying into muons,” *Phys. Lett.*, vol. B752, pp. 146–168, 2016.
- [150] ATLAS Collaboration, “Search for Higgs boson decays into pairs of light (pseudo)scalar particles in the  $\gamma\gamma jj$  final state in pp collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector,” *Physics Letters B*, vol. 782, p. 750–767, July 2018.
- [151] CMS Collaboration, “Search for exotic Higgs boson decays  $H \rightarrow \mathcal{A}\mathcal{A} \rightarrow 4\gamma$  with events containing two merged diphotons in proton-proton collisions at  $\sqrt{s} = 13$  TeV,” *Phys. Rev. Lett.*, vol. 131, p. 101801, Sep 2023.
- [152] ATLAS Collaboration, “Search for new phenomena in events with at least three photons collected in  $pp$  collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector,” *Eur. Phys. J.*, vol. C76, no. 4, p. 210, 2016.
- [153] CMS Collaboration, “Search for the exotic decay of the Higgs boson into two light pseudoscalars with four photons in the final state in proton-proton collisions at  $\sqrt{s} = 13$  TeV,” *JHEP*, vol. 07, p. 148, 2023.

- 
- [154] ATLAS Collaboration, “Search for Higgs boson decays to beyond-the-Standard-Model light bosons in four-lepton events with the ATLAS detector at  $\sqrt{s} = 13$  TeV,” *JHEP*, vol. 06, p. 166, 2018.
- [155] CMS Collaboration, “Search for new neutral Higgs bosons through the  $H \rightarrow ZA \rightarrow \ell^+ \ell^- b \bar{b}$  process in pp collisions at  $\sqrt{s} = 13$  TeV,” *JHEP*, vol. 03, p. 055, 2020.
- [156] E. T. J. Reynolds, *Searching for New Physics in Decays of the Higgs Boson with the ATLAS Detector*. PhD thesis, University of Birmingham, September 2019. Available at <https://etheses.bham.ac.uk/id/eprint/9807/1/Reynolds2019PhD.pdf>.
- [157] J. F. Gunion and H. E. Haber, “The CP conserving two Higgs doublet model: The Approach to the decoupling limit,” *Phys. Rev.*, vol. D67, p. 075019, 2003.
- [158] S. Alte, M. König, and M. Neubert, “Exclusive Weak Radiative Higgs Decays in the Standard Model and Beyond,” *JHEP*, vol. 12, p. 037, 2016.
- [159] ATLAS Collaboration, “Observation of  $H \rightarrow b \bar{b}$  decays and  $VH$  production with the ATLAS detector,” *Phys. Lett.*, vol. B786, pp. 59–86, 2018.
- [160] ATLAS Collaboration, “Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector,” *Phys. Lett.*, vol. B784, pp. 173–191, 2018.
- [161] A. L. Kagan *et al.*, “Exclusive Window onto Higgs Yukawa Couplings,” *Phys. Rev. Lett.*, vol. 114, no. 10, p. 101802, 2015.
- [162] A. S. Chisholm, S. Kuttimalai, K. Nikolopoulos, and M. Spannowsky, “Measuring rare and exclusive Higgs boson decays into light resonances,” *Eur. Phys. J.*, vol. C76, no. 9, p. 501, 2016.
- [163] G. Isidori, A. V. Manohar, and M. Trott, “Probing the nature of the Higgs-like Boson via  $h \rightarrow V\mathcal{F}$  decays,” *Phys. Lett.*, vol. B728, pp. 131–135, 2014.
- [164] R. D. Peccei and H. R. Quinn, “CP Conservation in the Presence of Pseudoparticles,” *Phys. Rev. Lett.*, vol. 38, pp. 1440–1443, Jun 1977.
- [165] P. Nason, “A New method for combining NLO QCD with shower Monte Carlo algorithms,” *JHEP*, vol. 11, p. 040, 2004.
- [166] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method,” *JHEP*, vol. 11, p. 070, 2007.
- [167] S. Alioli *et al.*, “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX,” *JHEP*, vol. 06, p. 043, 2010.

- [168] ATLAS Collaboration, “Measurement of the  $Z/\gamma^*$  boson transverse momentum distribution in  $pp$  collisions at  $\sqrt{s} = 7$  TeV with the ATLAS detector,” *JHEP*, vol. 09, p. 145, 2014.
- [169] T. Sjostrand *et al.*, “A Brief Introduction to PYTHIA 8.1,” *Comput. Phys. Commun.*, vol. 178, pp. 852–867, 2008.
- [170] A. Ryd *et al.*, “EvtGen: A Monte Carlo Generator for B-Physics,” 2005. EVTGEN-V00-11-07. <https://evtgen.hepforge.org/>.
- [171] H.-L. Lai *et al.*, “New parton distributions for collider physics,” *Phys. Rev.*, vol. D82, p. 074024, 2010.
- [172] M. Bähr *et al.*, “Herwig++ physics and manual,” *Eur. Phys. J. C*, vol. 58, no. 4, pp. 639–707, 2008.
- [173] J. Bellm *et al.*, “Herwig 7.0/Herwig++ 3.0 release note,” *The European Physical Journal C*, vol. 76, Apr. 2016.
- [174] T. Gleisberg *et al.*, “Event generation with SHERPA 1.1,” *Journal of High Energy Physics*, vol. 2009, p. 007–007, Feb. 2009.
- [175] R. D. Ball *et al.*, “Parton distributions with LHC data,” *Nucl. Phys.*, vol. B867, pp. 244–289, 2013.
- [176] J. Butterworth *et al.*, “Single Boson and Diboson Production Cross Sections in  $pp$  Collisions at  $\sqrt{s} = 7$  TeV,” Tech. Rep. ATL-COM-PHYS-2010-695, CERN, Geneva, Aug 2010.
- [177] GEANT4 collaboration, “Geant4—a simulation toolkit,” *NIMA: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, no. 3, pp. 250 – 303, 2003.
- [178] ATLAS collaboration, “The ATLAS Simulation Infrastructure,” *The European Physical Journal C*, vol. 70, pp. 823–874, Dec 2010.
- [179] D. Krohn *et al.*, “Jet cleansing: Separating data from secondary collision induced radiation at high luminosity,” *Physical Review D*, vol. 90, Sept. 2014.
- [180] D. Adams *et al.*, “Recommendations of the Physics Objects and Analysis Harmonisation Study Groups 2014,” tech. rep., CERN, Geneva, 2014.
- [181] J. Thaler and K. Van Tilburg, “Identifying Boosted Objects with N-subjettiness,” *JHEP*, vol. 03, p. 015, 2011.
- [182] I. Moutl, L. Necib, and J. Thaler, “New Angles on Energy Correlation Functions,” *JHEP*, vol. 12, p. 153, 2016.
- [183] Almeida *et al.*, “Substructure of high- $p_T$  Jets at the LHC,” *Phys. Rev.*, vol. D79, p. 074017, 2009.

- [184] G. V. Moustakides and K. Basioti, “Training Neural Networks for Likelihood/Density Ratio Estimation,” 2019. [arxiv.org/abs/1911.00405](https://arxiv.org/abs/1911.00405).
- [185] W. Buttinger, “Using Event Weights to account for differences in Instantaneous Luminosity and Trigger Prescale in Monte Carlo and Data,” tech. rep., CERN, Geneva, 2015.
- [186] ATLAS collaboration, “Luminosity determination in pp collisions at  $\sqrt{s} = 13$  TeV using the ATLAS detector at the LHC,” *The European Physical Journal C*, vol. 83, Oct. 2023.
- [187] ATLAS Collaboration, “Monte Carlo Calibration and Combination of In-situ Measurements of Jet Energy Scale, Jet Energy Resolution and Jet Mass in ATLAS,” tech. rep., CERN, Geneva, 2015.
- [188] G. Cowan *et al.*, “Asymptotic formulae for likelihood-based tests of new physics,” *The European Physical Journal C*, vol. 71, Feb. 2011.
- [189] A. L. Read, “Presentation of search results: the CLs technique,” *Journal of Physics G: Nuclear and Particle Physics*, vol. 28, p. 2693, sep 2002.