國 立 清 華 大 學

碩士論文

利用 ATLAS 探測器的 MonoHbb 分析之背景估計研究

Studies of the Background Estimation in the MonoHbb

Analysis with the ATLAS Detector

系　　　　所：物理研究所

學　　　　號：105022511

研 究 生：韓正忻　　　　(Cheng-Hsin Han)

指導教授：徐百嫻　博士　(Prof. Pai-Hsien Jennifer Hsu)

中 華 民 國 一○八 年 一 月

國 立 清 華 大 學

碩士論文

利用 ATLAS 探測器的 MonoHbb 分析之背景估計研究

Studies of the Background Estimation in the MonoHbb

Analysis with the ATLAS Detector

系　　　所：物理研究所

學　　　號：105022511

研 究 生：韓正忻　　　　　(Cheng-Hsin Han)

指導教授：徐百嫻　博士　(Prof. Pai-Hsien Jennifer Hsu)

中 華 民 國 一〇八 年 一 月

利用 ATLAS 探測器的 MonoHbb 分析之背景估計研究

摘要

我們透過大強子對撞機中的 ATLAS 探測器量測與一般物質有交互作用的物質的特性，分析某模型預測中，較不能與一般物質有交互作用之物質的存在與否，對於其背景事件的研究將在本篇文章中提及。

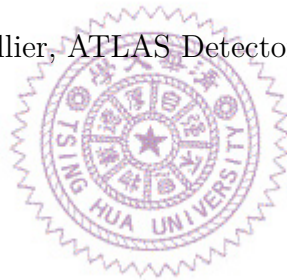關鍵字：大型強子對撞機, ATLAS 探測器, 希格斯玻色子, 暗物質

Studies of the Background Estimation in the MonoHbb Analysis with the ATLAS Detector

Abstract

We make use of ATLAS detector in the Large Hadron Collider to measure the physical quantity of matters that interact with other matters. This measurement are used to analyze the existence of a predicted invisible particle by a certain model. The studies of the background estimation are stated in this article.

Keywords: Large Hadron Collier, ATLAS Detector, Higgs boson, dark matter

Acknowledgement

Contents

Chapter 1

Introduction

It is generally acknowledged that the universe does not only contain matter. Dark matter (DM) [1] comprises more percentage in the universe than matter, and its nature remains one of the biggest unknown question in physics. A striking hypothesis on the nature of DM [2, 3] is that it is a electrically neutral, and thus being dark, stable particle, denoted as χ. It has weak interactions with Standard Model (SM) particles in addition to gravitational interactions. Its mass is predicted to range from the order of GeV to TeV. In this hypothesis, ways of its detection are presented and carried out as shown in figure 1.1. A collider way of detection for the DM [4] is discussed in this thesis.
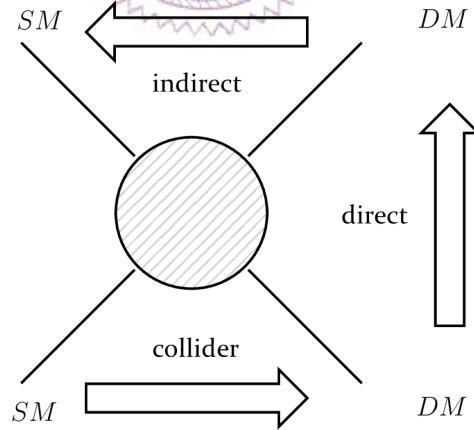
Figure 1.1: Different ways of dark matter search. Studies on the SM particles that are predicted to be the products of DM interactions are the indirect search. Researches on the SM particles which are assumed to have interaction with DM are the direct way of search. Colliding the SM particles and investigating the undetected parts of the products is the collider way of detection.

The DM signature in the collider under the hypothesis that they weakly interact with SM particles is the missing momentum. It can be observed by a detectable particle X that produced in associated with the DM. Such X+missing momentum signature is explored by the LHC (Large Hadron Collider) experiments with X being a jet (would be discussed in section 3.1) [5, 6], a heavy quark [7–10], a vector boson [6, 11–15], or a Higgs boson [16–19]. This thesis presents the search with X being a Higgs boson h that further decays into a pair of b-quarks, $h \to b\bar{b}$, which is the most frequent decay channel of h. The analyzed dataset of proton-proton collisions has the center-of-mass energy $\sqrt{s} = 13$ TeV and was recorded by ATLAS (**A T**oroidal **L**HC **A**pparatu**S**, would be discussed in chapter 2) during 2015 to 2017, corresponding to an integrated luminosity of 79.8 fb^{-1}.

Another progress in this analysis is the introduction of variable-radius jet, which would result in improvements in object reconstruction and performance. More information would be discussed in section 3.1.3 and its effect compared to the previous iteration is covered in chapter 6.
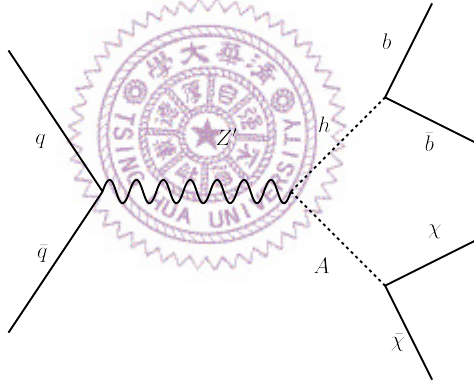
Figure 1.2: The Feynman diagram of the Z'-2HDM model in leading order. The collison of two SM particles produces the mediator Z', which further decays into a SM Higgs h and a pseudo-scalar Higgs A. The pseudo-scalar Higgs A is assumed to decay into a pair of DM candidate particles. The decay channel of the SM Higgs is a pair of b-quarks for this analysis.

The siganl model used in this iteration is a Type-II two-Higgs-doublet model (2HDM). The 2HDM model is one of the assumption to extend the beyond SM predictions. This prediction leads to a result of five Higgs bosons. The charge conjugation parity symmetry (CP) even h and H, where H is heavier than h, CP odd pseudo-scalar A and two charged $H^{\pm}$.

An additional mediator Z' is introduced in this signal model and thus the model is referred to as the Z'-2HDM model [17]. In this, one scalar Higgs boson h and a pseudo-

scalar Higgs A are found. The whole process reads $pp \rightarrow Z' \rightarrow Ah \rightarrow \chi\bar{\chi}b\bar{b}$, in which the DM pair decay from A and the b-quarks pair decay from h. The Feynman diagram is shown in figure 1.2. The relevant parameters are the masses m_{A}, $m_{\mathrm{Z'}}$, m_{χ}, the coupling of Z', $g_{Z'}$ and $\tan(\beta)$, the ratio of the vacuum expectation values of the two Higgs fields coupling to up-type and down-type quarks.

The main SM backgrounds are the top-quark pairs ($t\bar{t}$) and vector bosons (W or Z boson) with b-jets. The signal region and control region are used to constrain the background. More details would be covered in chapter 4.

Chapter 2

The ATLAS detector

2.1 Coordinates

The ATLAS experiment [20] is one of the seven detector in LHC at CERN (European Organization for Nuclear Research). Its cylindrical symmetry and end caps covers nearly 4π in solid angle.

A coordinate system is used to describe every recorded signals nearby. The origin is set at the center of the detector, or the interaction point (IP). The x-axis points toward the center of the LHC ring; the y-axis points vertically upward; the z-axis points along one of the beam pipe direction such that a right-handed coordinate sysetem is created.

A modified version of cylindrical coordinate is more commonly used in the experiment. The pseudorapidity $\eta \equiv -\ln\tan(\theta/2)$, in which θ is the polar angle in cylindrical coordinate, is used to decribe the angle between the z-axis and the direction of interest. (r, ϕ) is the same system to describe the tranverse plane, with ϕ being the azimuthal angle. In addition, the cone size variable, which is used in object selection and reconstruction (see chapter 3), is defined as $\Delta R \equiv \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$.

2.2 Components of ATLAS

Depending on its function, the components are categorized into four parts - inner detector, calorimeters, muon spectrometer, and the magnetic system. Apart from these, there are three levels of triggers which are designed to reduce the amount of data and also keep the signals of interest. Figure 2.1 shows the schematic positions and 2.2 shows

the side view of each components of ATLAS. The solenoidal magnets surround the inner detector while the toroidal magnets affects the signals in the muon spectrometer. These two magnets form the magnetic system. The others consist of smaller layers or components which are described in the following.
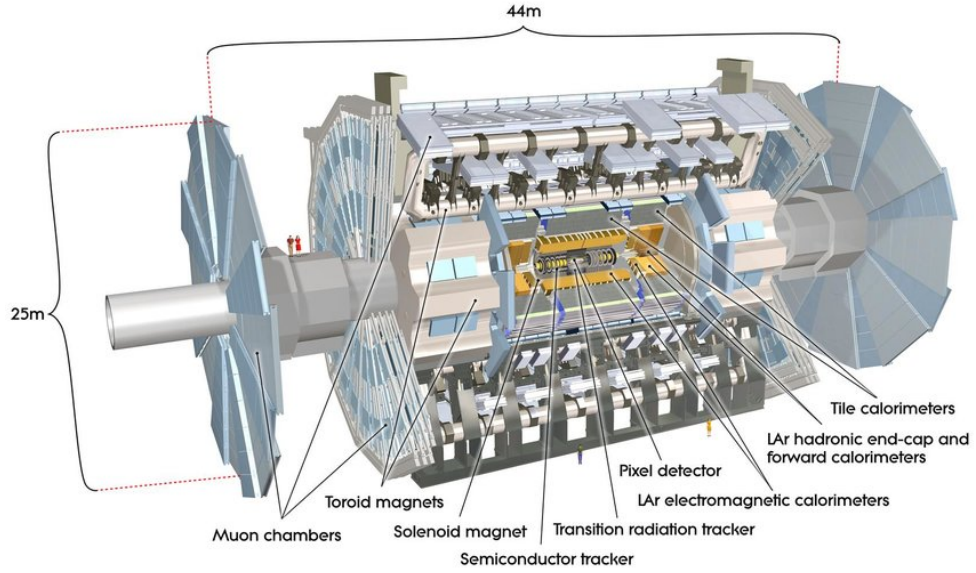

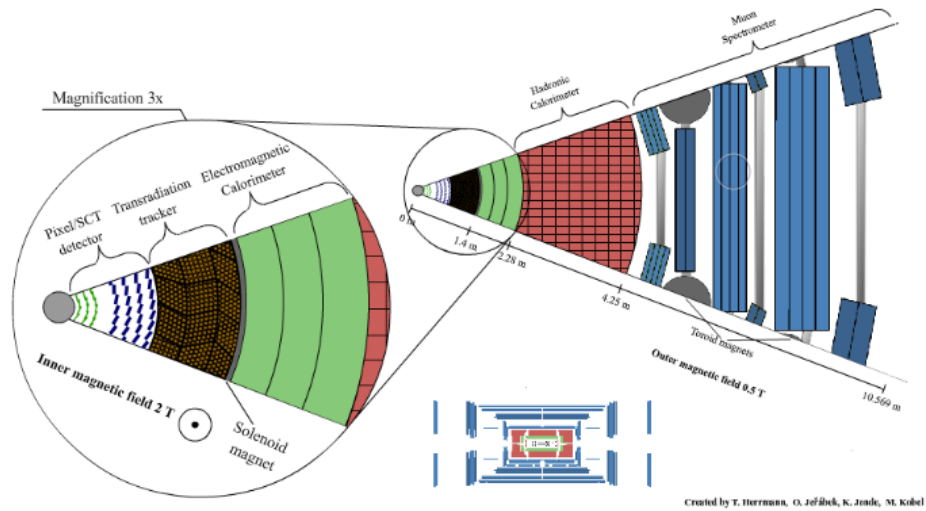Figure 2.1: Schematic plot of the ATLAS detector as well as the positions of its components.

Figure 2.2: Schematic plot of the side view of the ATLAS detector.

2.2.1 Inner Detector

The inner detector begins few centimeters from the IP. Its main function is to track the trace of charged particles by their interations with the materials. A 2T magnetic field, which is generated from the solenoidal magnets surrounding the whole inner detector, causes the charged ones to bend. Based on the directions and the curvatures, one can determine their charges and momenta preliminarily. The inner detector comprises three parts - the pixel detector, the semi-conductor tracker (SCT), and the transition rediation tracker (TRT).

The pixel detector is located at the innermost part of the inner detector. It contains four layers of modules, which is made up of silicon, in the direction perpendicular to the beam. It covers pseudorapidity range $|\eta| < 2.5$ and its proximity to the IP is meant to measure extremely precise trace of the charged particles. Three disks, which are made up of similar material, are at each end cap of the detector.

The semi-conductor tracker has a similar concept and function to the pixel detector. It lies in the middle part of the inner detector. The SCT is in a long and narrow strip-shape rather than small pixels and covers the perpendicular directions to the beam instead of nearly full coverage. The SCT overlays a larger area than the pixel detector does. Therefore, it has more sampled points and is of great importance on tracking the transverse directions with roughly the same accuracy compared to the pixel detector.

TRT is the outermost component. It includes straw tube trackers and transition radiation detectors. Its precision in tracking is not as high and its coverage in pseudorapidity, about $|\eta| < 2.0$, is not as wide as those of the other two components. However, TRT possesses transition radiation detection capability, which is useful for identifying charged particles. Lighter particles tend to have higher speed, which generates greater transition radiation. Thus, electrons and positrons, the lightest charged particles, would leave strong signals in TRT.

2.2.2 Calorimeters

Outside the solenoidal magnet, which envelops the inner detector, are the calorimeters. By absorbing the particles, the calorimeters measure the energies of them. Two layers of components compose the calorimeter systems, the inner electromagnetic (EM) calorimeter

and the outer hadronic calorimeter.

As its name suggests, the EM calorimeter absorbs energies from particles that interact electromagnetically, including photons and charged particles. High-granularity lead/liquid argon (LAr) EM calorimeter covers a range of $|\eta| < 3.2$, which includes the barrel and end cap. In addition, a LAr persampler which is meant to correct the energy loss in materials of the calorimeters covers $|\eta| < 1.8$. For the forward region, which has the range $3.1 < |\eta| < 4.9$, a LAr EM calorimeter with copper is also deployed.

Hadronic calorimeter is less precise in both energy magnitude and localization than EM calorimeter. It absorbs energies from the particles that interact via strong force. Hadrons and τ leptons, which are identified as jets (see section 3.2.3), are the targeted particles of hadronic calorimeter. Steel/scintillator-tile covering $|\eta| < 1.7$, two copper/LAr end cap calorimeters overlaying $1.5 < |\eta| < 3.2$, and a forward-regional ($3.1 < |\eta| < 4.9$) tungsten absorbers constitute the hadronic calorimeter.

2.2.3 Muon Spectrometer

Muon spectrometer, which is meant to provide more precise measurement of muon momentum and tracks, surrounds the calorimeters. Due to the fact that almost only muons reach it, muon spectrometer also has a function of identifying the muons. A magnetic field, provided by three toroidal magnets and thus is not uniform, creates a curve in muon tracks. These tracks can be made use of measuring the momentum. Detectors with triggers provide the identification and momentum measurements of the muons within the range $|\eta| < 2.4$; over a thousand precision tracking chambers covering $|\eta| < 2.7$ serve the muon spatial measurements.

2.2.4 Trigger System

A trigger is a set of device which sets thresholds on some physical quantities such as momenta and positions. If the threshold of one event is met, one keeps it; otherwise one abandons it. The ATLAS triggers consist of three levels. The first level is hardware based while the other two are software based. From roughly 1 billion events per second, these three levels of triggers combined select about few hundreds interesting ones. Namely, the interaction rate is reduced from 1 GHz to few hundreds Hz.

Chapter 3

Object Selection and Reconstruction

Signals recorded in the ATLAS detector are categorized or reconstructed as pysical objects, which could be further used in the analyses. Besides the head-to-head (or hard-scattered) events which have high transverse momentum (p_T) there are also additional collisions with lower p_T. These are referred to as the pile-up events, which one often wants to exclude in the analyes. The reconstruction and definition of objects used in this study are listed in the following.

3.1   Jets

Jets are the reconstruction of collimated bunches of hadrons. In ATLAS experiment, the anti-k_T algorithm [21] is usually used to recombine hadrons into cone-sized shapes. In short, the anti-k_T algorithm makes use of p_T of each given entities and reconstructs jets whose R is the given radius parameter and which center on hard-scattered particles. How the jets are reconstructed and defined in this study is tabulated in table.3.1 and explained as follows.

Table 3.1: Summary of the jet selection and reconstruction.

type	(*central*) small-R jets	(*forward*) small-R jets	large-R jets	VR track jets
p_T (GeV)	> 20	> 30	> 200	> 0.5
$|\eta|$	(0, 2.5)	(2.5, 4.5)	(0, 2.0)	(0, 2.5)
R	0.4		1.0	$\rho/p_T \in (0.02,\ 0.4)$ with $\rho = 30$ GeV
additional	if $|\eta| < 2.4$ then $p_T < 60$ GeV	-		$|z_0 \sin(\theta)| < 3$ mm

3.1.1 Small-radius Jets

Small-radius (small-R) jets have a relative smaller size compared to large-radius (large-R) jets, which would be discussed in subsection 3.1.2. They are main feature when the missing transverse momentum (E_T^{miss}, discussed in section 3.3) is below a certain value (see section 4.1).

Using the anti-k_T algorithm with a raidus parameter of 0.4, small-R jets can be reconstructed from the energy deposits in the calorimeter. Small-R jets are further categorized into two types, the central jets and the forward jets. Central small-R jets are those within $|\eta| < 2.5$ and whose $p_T > 20$ GeV. Forward small-R jets are within $2.5 < |\eta| < 4.5$ and are required of a p_T threshold of greater than 30 GeV. For jets in $|\eta| < 2.4$, an additional threshold of $p_T < 60$ GeV is require. To further suppress these jets from the pile-up interactions, they are required to be originated from the reconstructed location of the collision, or the primary vertex [22]. In addition, jets containing b-quark are referred to as b-jets. The method used to identify them as b-jets are called b-tagging. Because b-quarks have a relative longer lifetime compared to other particles, they would emit from a secondary vertex rather than the primary one, which can be made use of performing the b-tagging. A working point of 77% on averge for the b-tagging efficiency is used in this analysis.

3.1.2 Large-radius Jets

Large-R jets have a relative larger size compared to small-R jets, which is covered in the last subsection. They are the characteristic objects when the E_T^{miss} exceeds a threshold value (see section 4.1). Usually they are associated with variable-radius (VR) jets, which will be covered in subsection 3.1.3.

The large-R jets are reconstructed via the anti-k_T algorithm with a radius parameter of 1.0. The reconstruction highly depends on the calorimeter and the tracking system. Certain calibration method is appiled for the energy deposits in the calorimeter. The calibrated jets have a p_T threshold of 200 GeV and are required to be within the range of $|\eta| < 2.0$.

3.1.3 Variable-radius Jets

The large-R jets are a result of two "merged" small jets. In order to identify the small jets inside the reconstructed large-R jets, the VR track jets are introduced.

The VR track jets are also reconstructed using the anti-k_T algorithm. Those jets whose p_T does not exceed 0.5 GeV or out of the range of $|\eta| < 2.5$ are not considered. To suppress the pile-up jets, a creiteria on the longitudinal impact parameter, $|z_0 \sin(\theta)| < 3$ mm, is required. In this, z_0 is the point closet to the vertex along the logitudinal axis and θ is the polar angle of the track. The main feature of the VR track jets is that the radius parameter depends on the value of p_T [23] rather than a constant value:

$$R \rightarrow R_{\text{eff}}(p_T) \approx \frac{\rho}{p_T}, \tag{3.1}$$

in which ρ is set at 30 GeV, which is the optimal value ragarding the efficiency for b-tagging. The upper limit and lower limit, $R_{\max}$ and $R_{\min}$ are set at 0.4 and 0.02 in this regard.

3.2 Leptons

Leptons are used to categorized the regions selected in this analysis, which would be covered in chapter 4. Requirements of each flavor are explained in the following and summarized in table.3.2.

Table 3.2: Summary of lepton selection and reconstruction. The rightmost column are the requirements for the reconstructed small-R jet that decays from a τ-lepton candidate. The parentheses in the cell of μ ID would be covered in the next chapter.

flavor	e		μ		τ
categorization	baseline	signal	baseline	signal	-
p_T (GeV)	> 7	> 27	> 7	> 25	> 20
$|\eta|$	(0, 2.47)		(0, 2.7)	(0, 2.5)	(0, 1.37) $\cup$ (1.52, 2.5)
ID	Loose		Loose (0/2-lepton) Medium (1-muon)		Loose
transverse impact parameter	$d_0/\sigma(d_0) < 5$		$d_0/\sigma(d_0) < 3$		$|d_0| < 1$ mm
$|z_0 \sin(\theta)|$ (mm)	< 0.5				< 1.5
additional	-				one to four track-jets $\Delta\phi(\tau, E_T^{\text{miss}}) < \frac{\pi}{8}$

3.2.1 Electrons

Electron candidates are reconstructed from the energy deposits in EM calorimeter that match a track recorded in the inner detector. All candidates within $|\eta| < 2.47$ are considered. In addition, there is a likelihood (LLH)-based algorithm, which further makes use of multivariable analysis (MVA), applied for the electron ID. Three levels of ID operating point, loose, medium and tight, are povided; among them, loose ID is used in this study for the electrons. Requirements of the impact parameter are examined, both in the transverse and logitudinal directions. For the former, the relative resolution, which is the fraction of the transverse impact parameter d_0 and its resolution $\sigma(d_0)$, has an upper bound of 5. For the latter, the value $|z_0 \sin(\theta)| < 0.5$ mm is set.

Electrons are divided into two groups. The baseline electrons are those whose p_T exceed 7 GeV. The signal electrons, which require a tighter threshold of $p_T > 27$ GeV, are the other groups. The purpose of this categorization is to label which electrons decay more likely from heavier particles. Therefore, it can be used as a sign of different analysis regions. More details of these regions can be found in chapter 4.

3.2.2 Muons

Muon candidates, also divided into baseline and signal muons, are reconstructed with high dependence of inner detector and the muon spectrometer. The purpose of the categorization is the same as that of the electrons. Signal muons are those whose p_T exceed 27 GeV. Also, one needs to have more accurate identification of the signal muons compared to the baseline ones. They are required to leave tracks in the inner detector, and thus shall be found in $|\eta| < 2.5$. For baseline muons, which have a looser p_T threshold of 7 GeV, an acurate identification is not sufficient. Therefore, leaving signals in the inner detector is not required. Their pseudorapidity range are set within $|\eta| < 2.7$, which is the range of the muon spectrometer.

On top of that, the impact parameter must be consistent with the the primary vertex. $d_0/\sigma(d_0) < 3$ and $|z_0 \sin(\theta)| < 0.5$ mm are set. Finally, a loose ID is used for the zero-lepton and two-lepton channel whilst the one-muon channel makes use of a medium ID for the muons in this analysis. These channels would be covered in the upcoming chapter.

3.2.3 Taus

τ-leptons, whose decay length is few μm, barely reach the ATLAS detector and thus are mainly reconstructed from their decay products. Due to the fact that most of the τ-leptons decay into hadrons, the τ-lepton candidates are reconstructed from jets. The transverse impact parameter, $|d_0| < 1$ mm, and the logitudinal one, $|z_0 \sin(\theta)| < 1.5$ mm, are set for the jet tracks and the τ vertex [24]. The threshold on the p_T of the jets is set at 20 GeV; the range of $|\eta| < 2.5$, excluding $1.37 < |\eta| < 1.52$. The excluded region is those between the barrel and the forward region, or the crack region. Furthermore, the ID is built on a boosted decision tree (BDT) that makes use of the information from the tracks and the calorimeter. The loose working point on the τ-leptons is used. Finally, the small-R jet is required to contain one to four track-jets and within a range of $\Delta\phi < \frac{\pi}{8}$ with the missing transverse energy (E_T^{miss}, see section 3.3) in order to suppress the W-boson-decayed τ-leptons.

3.3 Missing Transverse Momentum

The missing transverse momentum is the imbalance in p_T of the reconstructed objects. Thus, it can be reconstructed as the negative vector transverse momentum sum. This includes the hard-scattered objects and the soft terms [25]. The momentum of hard-scattered reconstructed particles construct the former. As for the soft term, tracks that are not associated to any reconstructed hard objects are considered. The vetoed leptons, which are the leptons that needs to removed from the analysis in some channels, which would be covered in the next chapter, also count as the soft term. Jets that failed the selection of small-R jets also contributes the soft term, with the p_T threshold on the forward small-R jet being modified to 20 GeV. The absolute value of the missing transverse momentum is denoted as E_T^{miss}.

An estimation variable E_T^{miss} significance, S, is used to check the genuinity in whether E_T^{miss} comes from undectable particles instead of mismeasurements or any inefficiencies. There are two types of E_T^{miss} significance, the event-based one and the object-based one. The former one is the ratio of the E_T^{miss} and the square root of the transverse momentum

of hard-scattered particles, H_T, i.e.,

$$S_{\text{event}} = \frac{E_T^{\text{miss}}}{\sqrt{H_T}}, \tag{3.2}$$

where

$$H_T = \Sigma p_T^{\mu} + \Sigma p_T^{e} + \Sigma p_T^{\gamma} + \Sigma p_T^{\tau} + p_T^{\text{jets}}. \tag{3.3}$$

If the soft term contributes little in the transverse momentum, the value of H_T would be approximately that of E_T^{miss}, and thus the value of S_{event} tends to be small. It would be used in the background estimation mainly and would be used as a selection in the next chapter.

The other significance value, the object-based one, is a log-likelihood ratio value. It is used to test the hypothesis that the total momentum of the invisible particle is zero against the hypothesis that it is not. In short, if the value of the object-based significance, S_{object} which is non-negative, is close to zero, then it implies high possibilities that E_T^{miss} does not come from the invisible particles; if the value is large, then it implies a possibility of existing invisible particles.

3.4 Overlap Removal

Object ambiguities happen when objects match multiple reconstruction criteria. In the following listed specific steps of object reconstruction that are required to solve the problem.

1. If two electron candidates share the same track, remove the one with lower p_T.

2. If a τ-lepton candidate lies within $\Delta R = 0.2$ of an electron or muon, it is removed.

3. Reject the electron candidates whose track is shared with a muon candidate.

4. Small-R jets are removed if they are within $\Delta R = 0.2$ of an electron.

5. Remove the electrons that are within $\Delta R = \min(0.4, 0.04 + 10\text{GeV}/p_T^{\text{electron}})$ of a small-R jet.

6. If the separation between a small-R jet and a muon is within $\Delta R = 0.2$, the small-R jet is removed provided that it has fewer than three tracks or that the muon p_T is

greater than 50% of the jet p_T and is greater than 70% of the p_T sum of the tracks associated to the jets.

7. Remove the muons that are within $\Delta R = \min(0.4, 0.04 + 10\text{GeV}/p_T^{\text{muon}})$ of a small-R jet.

8. Large-R jets whose track is within $\Delta R = 0.1$ to that of an electron are removed.

Chapter 4

Event Selection

A number of selection is applied on the events in this study. Because the final state of the Z'-2HDM model contains no leptons (see fig. 1.2), one can define this zero-lepton region as the signal region (SR). However, there are other processes having the same final products as that of this model, which are recognized as the background. To further estimate the amount of the background events, one analyzes other processes that are similar to the background. These processes are called the control region (CR). Both SR and CR are discussed in this chapter. Some comparisons and characteristic selections of SR and CRs are summarized in table 4.1.

4.1 Signal Region

First, no baseline leptons candidates, including the τ-leptons, are considered. The value of E_T^{miss} is required to be at least 150 GeV. Multijet backgrounds may pass the selection criteria. To further suppress this background, one adds two selections. The first one requires the azimuthal angle between E_T^{miss} and any of the three highest-p_T jets $(\min(\Delta\phi(E_T^{\mathrm{miss}}, \mathrm{jet})))$ greater than 20°. The forward jets are only considered if the number of central jets is less than 3. The other selection requires the azimuthal angle between E_T^{miss} and p_T^{miss} smaller than 90°, where p_T^{miss} is the negative sum of object momentum measured by the inner detector.

After these selections, one devides these events into two regions based on the value of E_T^{miss}. Those whose E_T^{miss} is smaller than 500 GeV are defined as the resolved region, whilst the merged region contains those events with $E_T^{\mathrm{miss}} > 500$ GeV. The resolved region is further divided into three regions, whose E_T^{miss} are within $(150, 200]$ GeV, $(200, 350]$

GeV, and $(350, 500]$ GeV. Different sets of selections are applied in these two regions and tabulated in table 4.2.

For the resolved region, the events are required to have at least two small-R jets, which is the reason of the naming of this region. In these jets, exactly two b-tagged jets are also required. The p_T of one of the jets has to exceed 45 GeV. For the events with two (three or more) jets, the scalar sum of the p_T of the highest two (three) jets is also required to be at least 120 (150) GeV. A requirement of $S_{\text{object}} > 16$ is also applied. To further suppress the multijet backgrounds, a selection on the separation are required. Due to the conservation of the momentum, one expects the tracks of the E_T^{miss} and the Higgs candidate are back-to-back. The azimuthal angle between E_T^{miss} and the Higgs candidate is required to be greater than 120°. The azimuthal angle between the two jets from the Higgs candidate ($\Delta\phi(\text{jet}_1, \text{jet}_2)$) tends to be large in multijet background events because of the topology. Thus, it is required to be smaller than 140° in this study. Furthermore, in order to reject $t\bar{t}$ background events, two more selections are added. One of them is $\Delta R(\text{jet}_1, \text{jet}_2) < 1.8$. The scalar sum of the first three highest p_T of the jets is required to be greater than 63% of the scalar sum of all jets is considered as the other one. At last, one adds a Higgs mass window selection on the two b-tagged jets, 50 GeV $< m_{\text{jj}} < 280$ GeV.

As for the merged region, the events which contain at least one large-R jet are considered, which is the reason of the name "merged". The large-R (VR track) jet with the highest p_T are referred to as the leading large-R (VR track) jet. Two leading VR track jets associated with the leading large-R jet are required be to b-tagged. Events with any VR track jets outside the large-R jets are rejected. To suppress the $t\bar{t}$ background events, the value of p_T of the leading large-R jet is required to exceed 43% of the scalar sum of the p_T of the leading large-R jets and all small-R jets. Finally, the invariant mass of the large-R jet is also required to be within a Higgs mass window, 50 GeV $< m_{\text{J}} < 270$ GeV.

4.2 Control Region

Events containing leptons are used to define the CRs. One makes use of the CRs to estimate the main backgrounds of the Z'-2HDM model shown in fig. 1.2. These main backgrounds are shown in 4.1. Based on the number of leptons of the analyzed events, they

can be further categorized into two CRs. The $t\bar{t}$ and W+jets production are estimated from the 1-muon CR. The 2-lepton CR is used for the Z-jets events analyses.
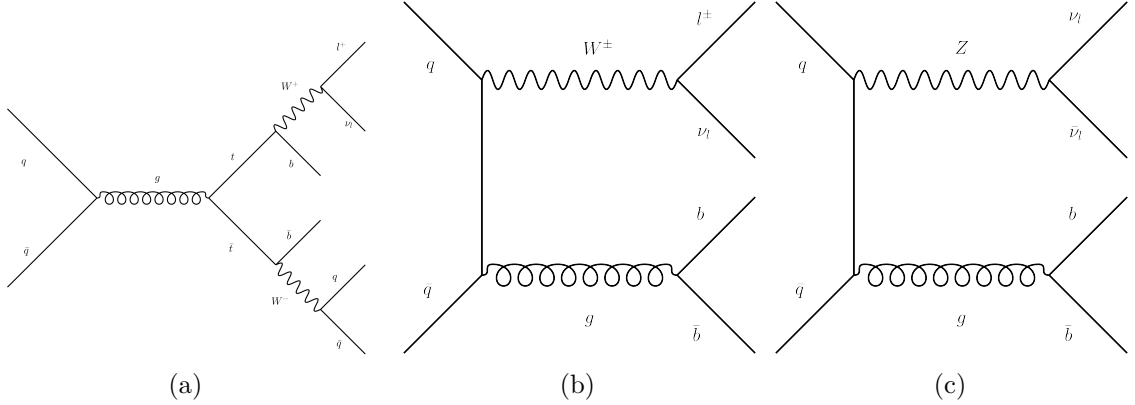

Figure 4.1: The Feynamn diagram of the main background events of the Z'-2HDM model. From left to right, (a) shows the $t\bar{t}$; (b) shows the W+jets events; (c) shows the Z+jets events. For the $t\bar{t}$ events, the lepton could also decay from the W^{-} boson, which is not shown in this figure.

4.2.1   1-muon Control Region

Although there are leptons in figure 4.1 (a) and (b), these events become the backgrounds because of misidentification or missed detections. As a result, the misidentified or undetected leptons would contribute to E_T^{miss}. The 1-muon CR is made use of estimating the backgrounds of these $t\bar{t}$ and W+jets events.

Because of the relatively higher fake rate of identification, single electron events are not considered in this background estimation. Therefore, no baseline electrons and exactly one signal muon are required. To mimic how these background events in SR behave, a "proxy" missing transverse momentum is defined as the sum of E_T^{miss} and the transverse momenta of the muon, or $E_T^{\text{miss,no } \mu}$. The resolved and merged region are also defined using the threshold of 500 GeV on $E_T^{\text{miss,no } \mu}$. Other selections are the same as those mentioned in section 4.1 except the S_{object} selection in the resolved region.

4.2.2   2-lepton Control Region

The neutrinos in figure 4.1 (c) cannot be detected and thus are part of the contributions to E_T^{miss}. The 2-lepton CR serves as the estimation of this background.

Two same-flavor leptons are required first. For both two-electron or two-muon cases,

one lepton is required to be the signal one while the other has to pass the criteria of the baseline ones. Additional requirement for the two-muon case is to have opposite charge. In order to make sure that the lepton pair truly decay from the Z boson candidate, a dilepton mass window is applied. The requirement on invariant mass of the two-electron (muon) system is 83 GeV $< m_{\text{ee}} < 99$ GeV (71 GeV $< m_{\mu\mu} < 106$ GeV). The "proxy" missing transverse momentum is the absoulte value of the sum of the transverse momentum of the two-leptons, or p_T^{ll}. One makes use of p_T^{ll} to define the resolved merged region based on the 500 GeV boundary. Those selections stated in section 4.1 except the S_{object} selection in the resolved region are applied. Apart from that, in order to reduce the $t\bar{t}$ and single top background in this CR, the S_{event} is required to be less than 3.5 GeV$^{\frac{1}{2}}$.

Table 4.1: Comparison of signal region and two control regions. This table only shows the characteristic selections and different definitions of (proxy) missing transverse momentum in different regions. One should combine the selections summarized in table 4.2 to get the whole set of selections in a certain region.

region	signal region	1-muon control region	2-lepton control region
lepton selection	no loose leptons	one signal muon no baseline electrons	two baseline muons or electrons one of them being signal
(proxy) missing transverse momentum	E_T^{miss}	$E_T^{\text{miss,no }\mu} = E_T^{\text{miss}} + p_T^{\mu}$	p_T^{ll}
E_T^{miss} significance selection (resolved only)	$S_{\text{object}} > 16$	-	$S_{\text{event}} < 3.5$ GeV$^{\frac{1}{2}}$
additional	-	-	83(99) GeV $< m_{\text{ee}(\mu\mu)} < 99(106)$ G opposite muon charge

Table 4.2: Summary of common selections of resolved and merged region for all SR and CRs. One should combine the characteristic selections stated in 4.1 to get the whole set of selections in a certain region.

resolved	merged
τ-veto	
$\min(\Delta\phi(E_T^{\mathrm{miss}}, \mathrm{jet})) > 20°$	
$\Delta\phi(E_T^{\mathrm{miss}}, p_T^{\mathrm{miss}}) < 90°$	
(proxy) $E_T^{\mathrm{miss}} \in (150, 500]$ GeV	(proxy) $E_T^{\mathrm{miss}} \in (500, \infty)$ GeV
≥ 2 b-tagged small-R jets	≥ 1 large-R jet
$\Delta R(\mathrm{jet}_1, \mathrm{jet}_2) < 1.8$	$\Delta R(\mathrm{VR}_1, \mathrm{VR}_2) > R_{\min}$
$p_T^{\mathrm{jet}_1} > 45$ GeV or $p_T^{\mathrm{jet}_2} > 45$ GeV	-
$\sum_{i=1}^{2(3)} p_T^{\mathrm{jet}_i} > 120(150)$ GeV	-
$\Delta\phi(E_T^{\mathrm{miss}}, h) > 120°$	-
$\Delta\phi(\mathrm{jet}_1, \mathrm{jet}_2) < 140°$	-
$\sum_{i=1}^{3} p_T^{\mathrm{jet}_i} / \sum_{i=1}^{\infty} p_T^{\mathrm{jet}_i} > 0.63$	$p_T^{\mathrm{large-Rjet}} / (p_T^{\mathrm{large-R\ jet}} + \sum p_T^{\mathrm{small-R\ jet}}) > 0.43$
50 GeV $< m_{\mathrm{jj}} < 280$ GeV	50 GeV $< m_{\mathrm{J}} < 270$ GeV

Chapter 5

Study on the Background Estimation

As mentioned in the last chapter, the CRs are used to estimate the backgrounds in the SR. Because W+jets events and $t\bar{t}$ events in both SR and one-muon CR are the same processes, one assumes the kinematics are the same. As for the two-lepton CR, first note that the mass of the charged leptons and neutrinos are negligible compared to that of the Z boson. Thus, one can argue that the kinematics of the Z(ll)+jets and Z($\nu\nu$)+jets events are roughly the same, where the Z(ll) means the charged-leptonic decay of the Z boson.

During the analysis, the author and Po-Shan Shih, partner of the author, are both working on the estimation of the two-lepton CR. Due to the fact that our group changed the framework to a unfamiliar one, validation is needed. In this chapter, the problems that the author participated or fixed during the study are stated. Also, a shape difference of W($\mu\nu$)+jets events in SR and one-muon CR is carried out as an additional work.

5.1 Overlap Removal

The comparison plots of Monte Carlo (MC) to data are always reviewed by analyzers. Those plots are referred to as the pre-fit plots. The pre-fit plots contain the data and MC samples that pass the selection criteria. Those events are then filled in the plot based on the value of the physical quantity of interest.

An unusual peak was spotted in one of the Z+jets CR plot during the analysis, as shown in figure 5.1. The plot is the distribution of the leading large-R jet mass and the peak is within 90 to 105 GeV. It is unusual because the statistics of the mass spectum should generally become less as the energy goes up. However, the peak reveals that many events assemble around a high energy value and thus create a discontinuity.
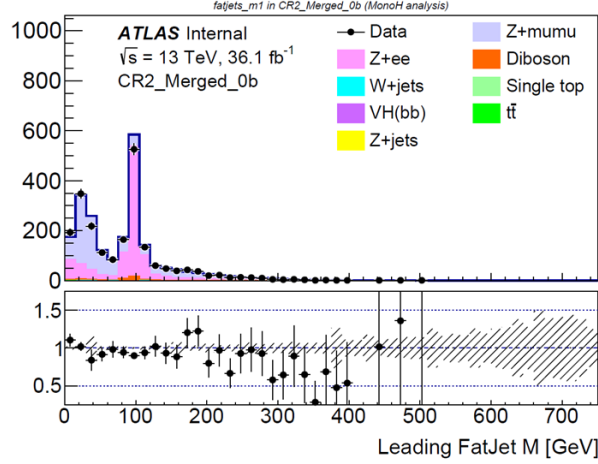
Figure 5.1: The plot which shows the issue of overlap removal. It shows the distribution of leading large-R jet mass. The dotted points represent the data while the solid histogram shows the MC simulations. Both of them pass the selection criteria of two-lepton CR. The spotted unusual peak lies in the bin of 90 to 105 GeV. The legends of "Z+mumu" and "Z+ee" in the upper-right of the plots are typos which weren't fixed at that time. They refer to the MC simulations of Z(ee)+jets and $Z(\mu\mu)$+jets events respectively.

The bin includes the Z boson mass, which is about 91 GeV. From the Feynman diagram of the Z+jets background (figure 4.1c) with the neutrinos replaced with leptons, one can tell that the invariant mass of the lepton pair would be close to that of a Z boson. In figure 5.1, nevertheless, it is large-R jet that has the invariant mass around it.

Because all these events pass the two-lepton CR selections as stated in section 4.2.2, one of the possible explanation to this peak is misidentification of leptons and large-R jets. It happens when they share the same track or overlap. From figure 5.1, one recognizes that the Z(ee)+jets events contributes the most in that bin. Recall the last reconstruction strategy from section 3.4 that an electron is removed if it it within $\Delta R = 0.1$ with a large-R jet.

As a result, a check in the codes on the overlap removal (OR) was conducted, and the same plot after adding it is shown in figure 5.2. From the plot, one can see that the peak around the Z boson mass is removed. This discovery and study reveal the fact that OR wasn't properly applied at that time. After a careful check on the codes with the developer of the framework we used, the OR problem was officailly fixed.
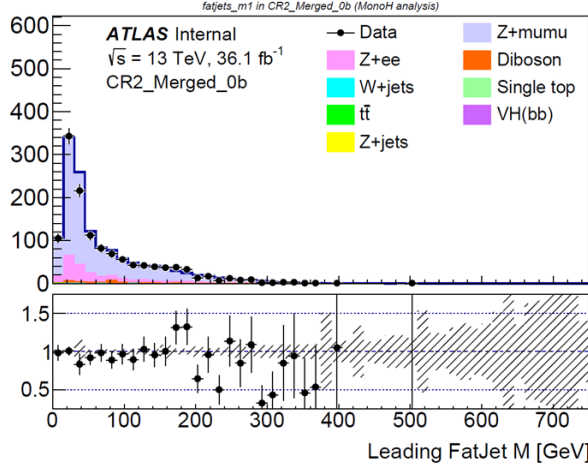
Figure 5.2: The plot which shows the result after overlap removal applies properly. The dotted points represent the data while the solid histogram shows the MC simulations.

5.2   Dealing with Large Weights

MC simulations use the energy density as the phase space to calculate the cross section of the sample in the leading order. However, when including higher orders, parton showers which result from strong interactions are not well covered. To further include these situations in the simulations, weights are applied. After the calculations done by MC, the raw number of events would be multiplied by these weights.

Another peak was discovered in the Z+jets CR plot during the analysis, as shown in figure 5.3. The shown plot is also the distribution of leading large-R jet mass, but it also appeared in other different plots. Following the logic in the last section, one can tell that it is unusual. However, because of the fact that OR had fixed this time and also because of this spike-like behavior existed in other kinematic variable plots, it must be a result of a different issue.

As it turned out, an event was multiplied by a really high generator weight which resulted in a large contribution to the MC. The reason of this is a result of a random generation, which happens in the the type of MC generator the analyzers used. Another similar weird behavior could happen when the generator weight is a great negative value, as shown in figure 5.5, which was another plot during the analysis. A clear pit shows up in the distribution where it should be continuous.

The solution to these weights in this iteration is to remove those events, because the number of events with these extreme value of generator weight is small and negligible.
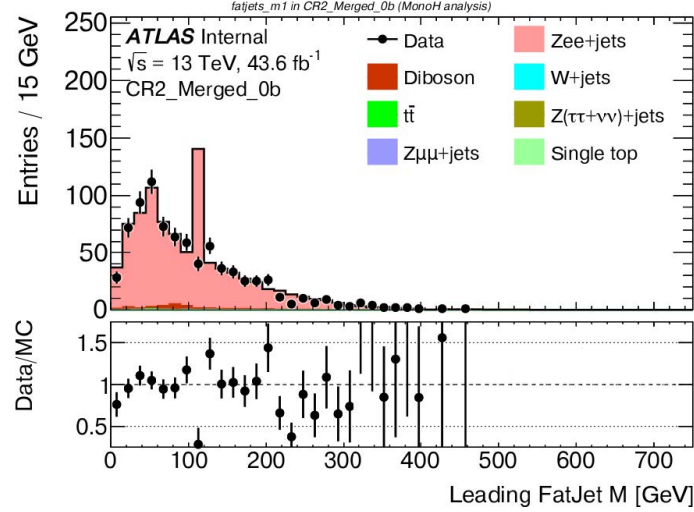
Figure 5.3: The plot which shows the unusual peak resulting from the generator weight. It shows the distribution of leading large-R jet mass. The dotted points represent the data while the solid histogram shows the MC simulations. The legends of "Z+mumu" and "Z+ee" in the upper-right of the plots are typos which weren't fixed at that time. They refer to the MC simulations of Z(ee)+jets and Z($\mu\mu$)+jets events respectively.
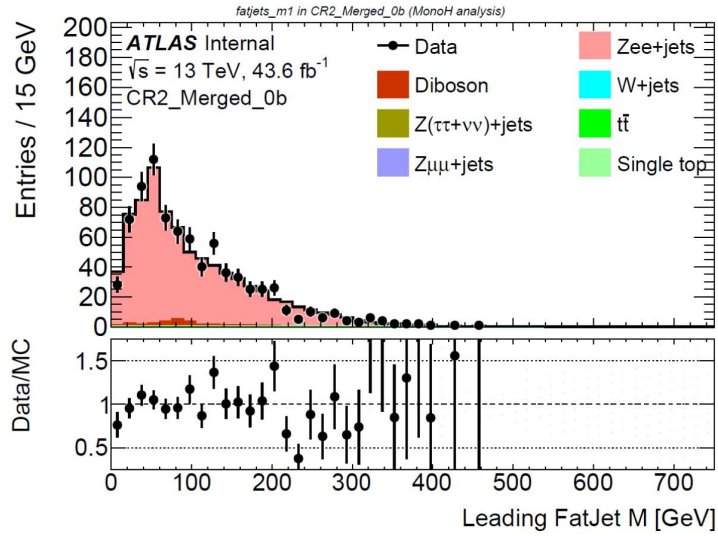

Figure 5.4: The plot which shows the result after the event with large generator weight removed. Comparing to 5.3 reveals that the peak disappears.

The results after removing the events are shown in figure 5.4 and 5.6.
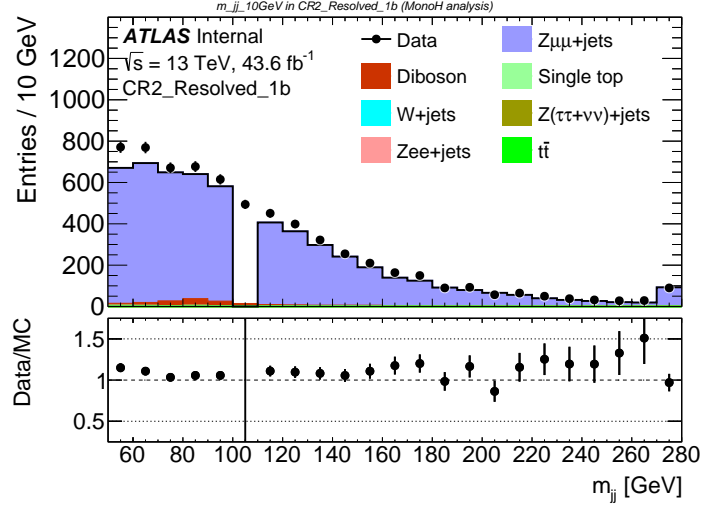
Figure 5.5: The plot which shows the unusual pit. It shows the distribution of dijet mass. The dotted points represent the data while the solid histogram shows the MC simulations.
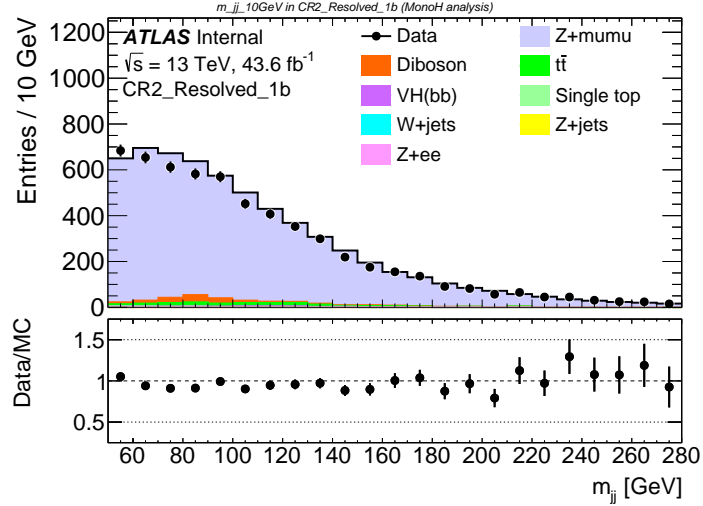

Figure 5.6: The plot which shows the result after the events with large negative value of generator weight are removed. Comparing to 5.5 reveals that the pit disappears.

5.3 W(μν) +jets Events Shape Comparison

As mentioned in the beginning of this chapter, the W+jets and $t\bar{t}$ events in either SR or one-muon CR are assumed to have the same kinematics. Thus, one-muon CR can be made use of estimating the background in SR. However, the author wonders whether the comparison of the kinematics between these two regions matches. Therefore, a study on this is performed.

A set of W(μν)+jets MC samples is used. In these samples, the selection for one-muon CR and SR mentioned in the last chapter are applied. The samples after the selections

of SR is used as a comparison to the one-muon CR.

The reason why a comparison between W($\mu\nu$)+jets samples and data is not used is because one cannot know the truth-level comparison via it. The data after the one-muon selections would contain the $t\bar{t}$ events since they are also one of the estimation target of one-muon CR. If one adds the $t\bar{t}$ MC samples in the comparison, data and MC comparison is apple-to-apple, but one would lose the information of the shape difference on W($\mu\nu$) +jets events only.

Because one doesn't know the yields of both samples after the selection, a normalization is used on them. After normalization, one can deduce the shape difference without any adjustment on the yields of one of them. The events are put in a histogram plot for comparison. The shape comparison for (proxy) E_T^{miss} is shown in figure 5.7. Besides, a yields uncertainty of 5% is assigned to this SR and one-muon CR extrapolation. The reason of 5% is due to the fact that some uncertainties used in this analysis follow the results of another analysis group, the Vh(bb) group, including this uncertainty value.
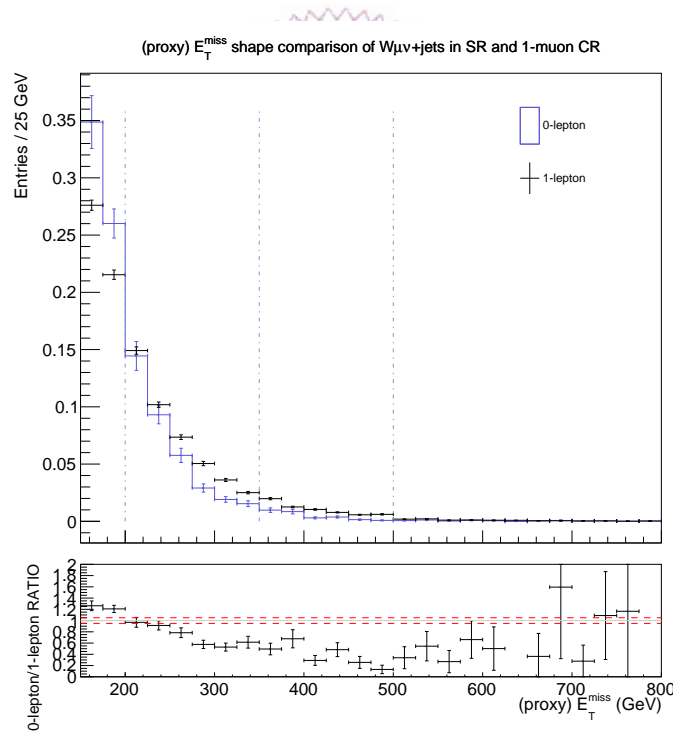

Figure 5.7: The comparison of the distribution of (proxy) E_T^{miss} between SR and one-muon CR for the W($\mu\nu$)+jets events. The upper pannel shows the normalized distribution and the lower pannel shows the ratio of SR to CR. Three auxiliary vertical lines at 200, 350, 500 GeV are shown on the upper one to distinguish different bins. Two auxiliary horizontal lines at 0.95 and 1.05 are shown on the lower one to indicate a 5% yields uncertainty.

From figure 5.7, two interseting things are found. First, the shape distribution has the same trend. Second, the ratios between the SR and one-muon CR are generally away from the 5% difference as shown. It is noticed that at lower energy, the ratios are larger than the upper limit and, at higher energy, the ratios are less than the lower limit, except for the last few bins, where the statistics are low. That is to say, at lower (higher) energy, W($\mu\nu$)+jets events underestimate (overestimate) the backgrounds in SR.

In author's opinion, although the W($\mu\nu$)+jets events are the same in both SR and one-muon CR, the situation is a little bit different. These events would appear in the SR only when the muon is misidentified. If the muon is soft, which means it has low p_T, it's more likely to be misidentified and thus be reconstructed as E_T^{miss}. On the other hand, if the p_T is higher, misidentification are less likely to happen and the muon are rejected by the selection of SR. Therefore, more events whose muons have higher p_T are removed in the SR but events with more low p_T muons stay in the SR. This means that the true simulation, which is expected in one-muon CR, underestimates (overestimates) the yields at lower (higher) energy. Due to this, one sees the severely flucuating ratio values.

As a minor conclusion, the W($\mu\nu$)+jets comparison between SR and CR doesn't quite fit in the 5% limit. One of the assumption on this is that the over- and under- estimation on different value of p_T are not well covered by the uncertainty. The yield uncertainty shall be recalculated for this issue.

Chapter 6

Results of monoHbb Analysis

6.1 Main Uncertainties

The systematic uncertainties dominate the main uncertainty in this analysis. Among them, the b-tagging efficiency, the integrated luminosity, jet energy scale (JES) and jet energy resolution (JER) contribute the most to the experimental uncertainty. The SM predictions are the sources of the theoretical uncertainties. $t\bar{t}$ events, W/Z +jets events, associated production of SM Higgs boson decay into $b\bar{b}$ (Vh($b\bar{b}$), where V is the vector boson) and diboson events dominate the theoretical uncertainties.

The uncertainty on the b-tagging efficiency originate from the flavor tagging efficiency in the $t\bar{t}$ events. A calibration, which highly makes use of beam testing in LHC, on the integrated luminosity is performed with a value a 2.0%.

The theoretical uncertainties originate from the SM modeling. Normalizations, acceptance difference and differential distributions of important kinematic variables derive the uncertainties.

6.2 Results

A fit to the invariant mass of Higgs candidate m_h is used to search for the signal. For resolved region, m_{jj} represents the Higgs mass while m_J is used in the merged region.

The fit is based on a LLH based approach. The systematic uncertainties are used in the likelihood function as nuisance parameters. The data in SR and two CRs are fit simultaneously for all four different (proxy) E_T^{miss} bins: $[150, 200)$ GeV, $[200, 350)$ GeV,

$[350, 500)$ GeV, and $[500, \infty)$ GeV. m_h is the fit variable in the SR. The fit variable used in the one-muon CR is the μ lepton charge. $t\bar{t}$ processes tend to produce the same amount of μ^+ and μ^- leptons but the W+jets events produce more μ^+ than μ^- leptons, which originates from proton-proton collision in LHC and from the conservation of electric charges. μ-charge can thus be made use of as a differentiation of $t\bar{t}$ and W+jets events. In the two-lepton CR, the event yields serve as the fit variable because of limited data statistics.

Figure 6.1 shows the E_T^{miss} distribution in SR. Figure 6.2 shows the distribution of m_{jj} or m_{J} in SR for resolved and merged region respectively. The data yields agree with the SM predictions. That is to say, no significant excess of the signal is found.

An exclusion limit at 95% confidence level (CL) is used for the interpretation of this analysis. The exclusion contour in $(m_{\mathrm{A}}, m_{\mathrm{Z'}})$ phase space is shown in figure 6.3. It also shows the result in the previous iteration. As it suggests, more region are excluded compared to the previous result. Figure 6.4 shows the comparison of the upper limit of the signal strength between this result and the previous result, in which the fixed-radius (FR) track jets are used. In contrast to VR, FR track jets have a fixed cone size of $R = 0.2$. The plot also suggests more region in the phase space is excluded.
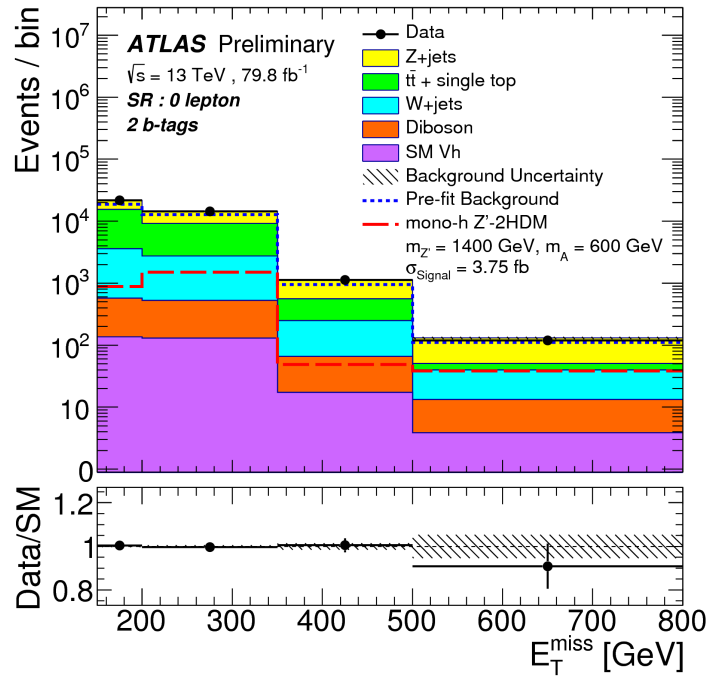

Figure 6.1: The E_T^{miss} distribution for resolved and merged combined in SR. The dashed blue lines are the expectation yields before fits. The solid histograms are the simulations after fit. The dashed red lines are the expected signal from Z'-2HDM model.
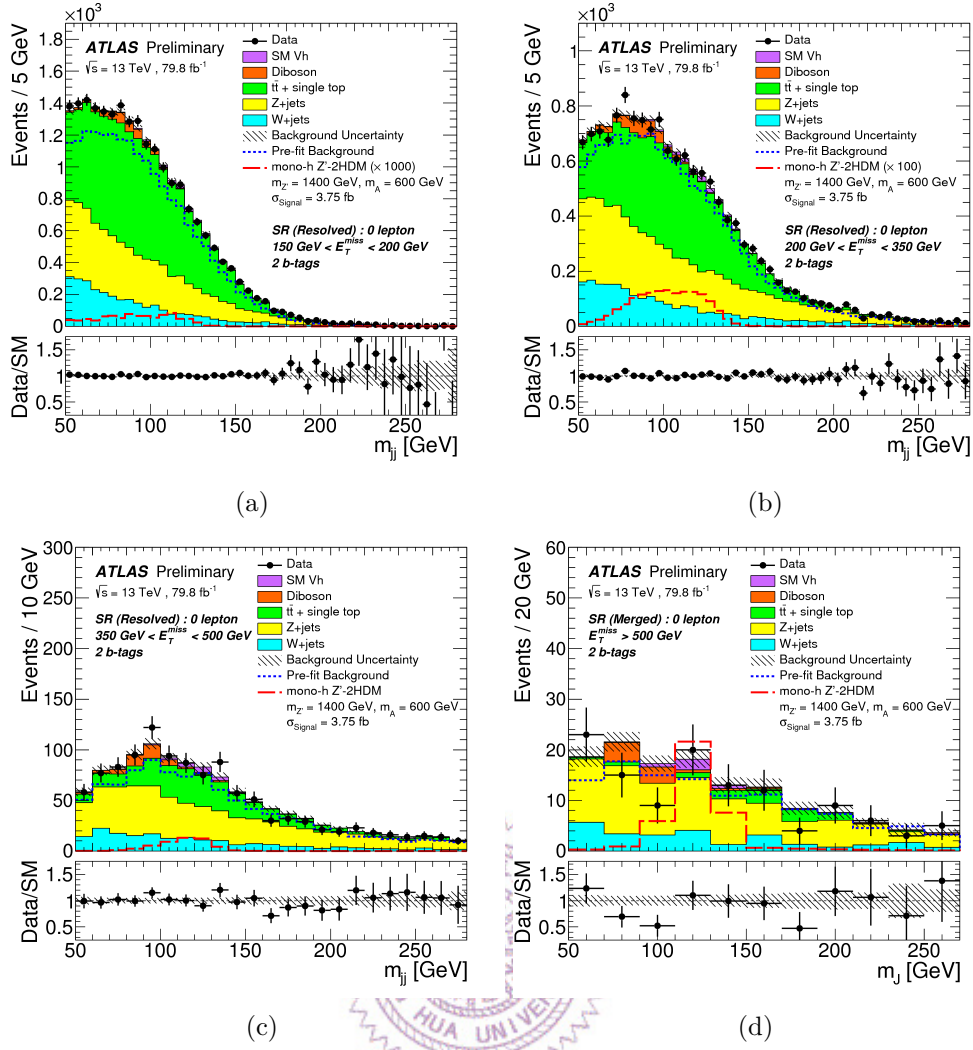
(a)

(b)

(c)

(d)

Figure 6.2: Distirbution of the invariant mass of the Higgs boson candidate m_{h} with two b-tagged jets. The upper two plots are for $E_T^{\mathrm{miss}} \in [150, 200)$ GeV and $[200, 350)$ GeV bins. The lower two ones are for $E_T^{\mathrm{miss}} \in [350, 500)$ GeV and $[500, \infty)$ GeV. The dashed blue lines are the expectation yields before fits. The solid histograms are the simulations after fits. The dashed red lines are the expected signal from Z'-2HDM model. Its yields for the upper two plots are scaled up by a factor of 100 and 1000 from left to right respectively.
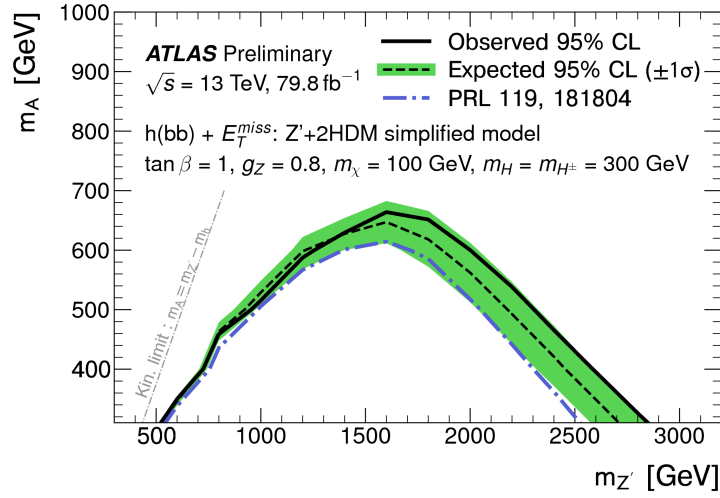
Figure 6.3: Exclusion contour in $(m_A, m_{Z'})$ phase space. Regions under the curve are excluded. The solid line shows consistency with SM-only hypothesis. The dashed blue line are the results from previous ATLAS results of $\sqrt{s} = 13$ TeV.
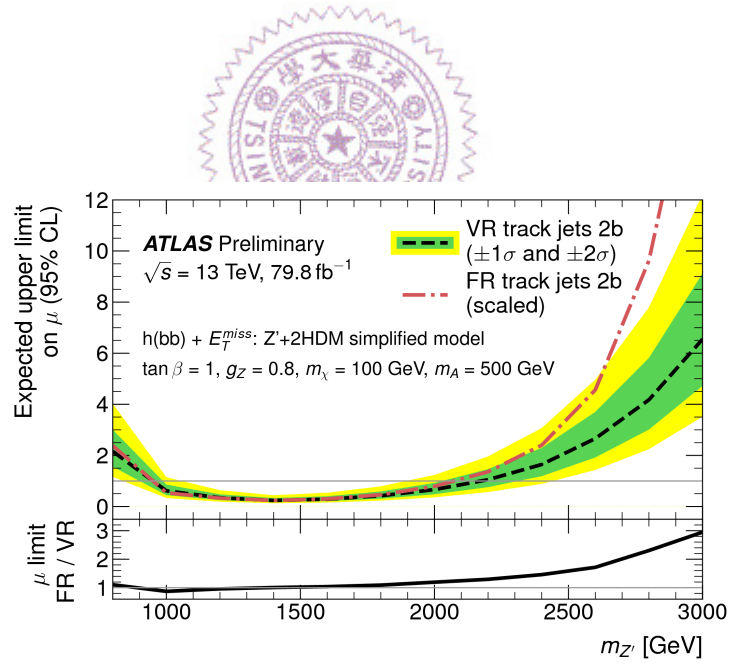

Figure 6.4: The upper limit of the signal strength of this model with m_A fixed at 500 GeV. The dashed red line is the result of previous iteration, which made use of FR track jets. The dashed black line in the result of this iteration, where VR track jets are used. Signals, if exist, would appear in the region where the signal strength is greater than one. As a result, all regions whose upper limit is smaller than one is excluded. As the plot shows, making use of VR track jets extends the excluded region.

Chapter 7

Conclusion

To sum up, a collider search for the dark matter in association with a final state of E_T^{miss} and $b\bar{b}$, which decay from the Higgs boson candidate, was performed using 79.8 fb^{-1} of proton-proton collision at $\sqrt{s} = 13$ TeV recorded in the ATLAS at LHC. The results are in agreement with SM predictions. An exclusion on the parameter space of Z'-2HDM model is excluded, up to $m_{Z'} = 2.8$ TeV and $m_{A} = 300$ GeV with other parameters set at $\tan\beta = 1$, $g_{Z'} = 0.8$ and $m_{\chi} = 100$ GeV.
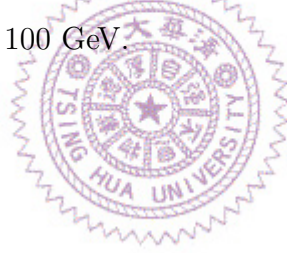
Bibliography

1. Bertone, G., Hooper, D. & Silk, J. Particle Dark Matter: Evidence, Candidates and Constraints. eprint: `arXiv:hep-ph/0404175` (2004).

2. Steigman, G. & Turner, M. S. Cosmological constraints on the properties of weakly interacting massive particles. *Nuclear Physics B* **253,** 375–386. ISSN: 0550-3213. `http://www.sciencedirect.com/science/article/pii/0550321385905371` (1985).

3. Scherrer, R. J. & Turner, M. S. On the relic, cosmic abundance of stable, weakly interacting massive particles. *Phys. Rev. D* **33,** 1585–1589. `https://link.aps.org/doi/10.1103/PhysRevD.33.1585` (6 Mar. 1986).

4. Abercrombie, D. *et al. Dark Matter Benchmark Models for Early LHC Run-2 Searches: Report of the ATLAS/CMS Dark Matter Forum* 2015. eprint: `arXiv:1507.00966`.

5. ATLAS Collaboration. Search for dark matter and other new phenomena in events with an energetic jet and large missing transverse momentum using the ATLAS detector. eprint: `arXiv:1711.03301` (2017).

6. CMS Collaboration. Search for new physics in final states with an energetic jet or a hadronically decaying W or Z boson and transverse momentum imbalance at $\sqrt{s} = 13$ TeV. eprint: `arXiv:1712.02345` (2017).

7. ATLAS Collaboration. Search for dark matter produced in association with bottom or top quarks in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector. eprint: `arXiv:1710.11412` (2017).

8. ATLAS Collaboration. Search for top-squark pair production in final states with one lepton, jets, and missing transverse momentum using 36 fb^{-1} of $\sqrt{s} = 13$ TeV *pp* collision data with the ATLAS detector. eprint: `arXiv:1711.11520` (2017).

9. CMS Collaboration. Search for dark matter in events with energetic, hadronically decaying top quarks and missing transverse momentum at $\sqrt{s} = 13$ TeV. eprint: `arXiv:1801.08427` (2018).

10. CMS Collaboration. Search for dark matter produced in association with heavy-flavor quarks in proton-proton collisions at sqrt(s) = 13 TeV. eprint: `arXiv:1706.02581` (2017).

11. ATLAS Collaboration. Search for dark matter at $\sqrt{s} = 13$ TeV in final states containing an energetic photon and large missing transverse momentum with the ATLAS detector. eprint: `arXiv:1704.03848` (2017).

12. ATLAS Collaboration. Search for an invisibly decaying Higgs boson or dark matter candidates produced in association with a Z boson in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. eprint: `arXiv:1708.09624` (2017).

13. CMS Collaboration. Search for new physics in events with a leptonically decaying Z boson and a large transverse momentum imbalance in proton-proton collisions at $\sqrt{s} = 13$ TeV. eprint: `arXiv:1711.00431` (2017).

14. CMS Collaboration. Search for new physics in the monophoton final state in proton-proton collisions at sqrt(s) = 13 TeV. eprint: `arXiv:1706.03794` (2017).

15. CMS Collaboration. Search for dark matter and unparticles in events with a Z boson and missing transverse momentum in proton-proton collisions at sqrt(s) = 13 TeV. eprint: `arXiv:1701.02042` (2017).

16. ATLAS Collaboration. Search for dark matter in association with a Higgs boson decaying to two photons at $\sqrt{s}=$ 13 TeV with the ATLAS detector. eprint: `arXiv:1706.03948` (2017).

17. ATLAS Collaboration. Search for Dark Matter Produced in Association with a Higgs Boson Decaying to $b\bar{b}$ using 36 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector. eprint: `arXiv:1707.01302` (2017).

18. CMS Collaboration. Search for dark matter produced in association with a Higgs boson decaying to $\gamma\,\gamma$ or $\tau^{+}\tau^{-}$ at $\sqrt{s} = 13$ TeV. eprint: `arXiv:1806.04771` (2018).

19. CMS Collaboration. Search for associated production of dark matter with a Higgs boson decaying to b b-bar or gamma-gamma at sqrt(s) = 13 TeV. eprint: `arXiv:1703.05236` (2017).

20. The ATLAS Collaboration *et al.* The ATLAS Experiment at the CERN Large Hadron Collider. *Journal of Instrumentation* **3,** S08003. `http://stacks.iop.org/1748-0221/3/i=08/a=S08003` (2008).

21. Cacciari, M., Salam, G. P. & Soyez, G. The anti-k_t jet clustering algorithm. eprint: `arXiv:0802.1189` (2008).

22. Collaboration, A. Performance of b-Jet Identification in the ATLAS Experiment. eprint: `arXiv:1512.01094` (2015).

23. Krohn, D., Thaler, J. & Wang, L.-T. Jets with Variable R. eprint: `arXiv:0903.0392` (2009).

24. ATLAS Collaboration. Identification and energy calibration of hadronically decaying tau leptons with the ATLAS experiment in pp collisions at $\sqrt{s}=$8 TeV. eprint: `arXiv:1412.7086` (2014).

25. ATLAS Collaboration. Performance of algorithms that reconstruct missing transverse momentum in $\sqrt{s} = 8$ TeV proton-proton collisions in the ATLAS detector. eprint: `arXiv:1609.09324` (2016).