# The Higgs boson with the ATLAS experiment at the LHC: Discovery, measurement, and searches for new physics

A DISSERTATION PRESENTED
BY
TOMO LAZOVICH
TO
THE DEPARTMENT OF PHYSICS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN THE SUBJECT OF
PHYSICS

HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS
MAY 2016

Thesis advisor: Professor Joao Guimaraes da Costa and Melissa Franklin          Tomo Lazovich

# The Higgs boson with the ATLAS experiment at the LHC: Discovery, measurement, and searches for new physics

## ABSTRACT

We measured things. And searched for other things. Here is what we found, please let me graduate.

# Contents

# Listing of figures

# Listing of tables

This is the dedication.

# Acknowledgments

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Morbi commodo, ipsum sed pharetra gravida, orci magna rhoncus neque, id pulvinar odio lorem non turpis. Nullam sit amet enim. Suspendisse id velit vitae ligula volutpat condimentum. Aliquam erat volutpat. Sed quis velit. Nulla facilisi. Nulla libero. Vivamus pharetra posuere sapien. Nam consectetuer. Sed aliquam, nunc eget euismod ullamcorper, lectus nunc ullamcorper orci, fermentum bibendum enim nibh eget ipsum. Donec porttitor ligula eu dolor. Maecenas vitae nulla consequat libero cursus venenatis. Nam magna enim, accumsan eu, blandit sed, blandit a, eros.

# 0

# Introduction

# Part I

# Preliminaries

<span style="color:#8B0000">**1**</span>

# The Standard Model and beyond: a theoretical overview

*This is some random quote to start off the chapter.*

Firstname lastname

# 2

# The ATLAS detector and the Large Hadron Collider

# Part II

# Observation and measurement of Higgs boson decays to $WW*$ with the ATLAS detector in LHC Run 1 at $\sqrt{s} = 7$ and 8 TeV

*Basic research is what I am doing when I don't know what I am doing.*

Wernher von Braun

# 3

# $H \to WW^* \to \ell\nu\ell\nu$ Analysis Strategy

## 3.1 INTRODUCTION

This chapter will present an overview of the strategy for searching for a Higgs boson in the $H \to WW^* \to \ell\nu\ell\nu$ decay topology. Its purpose is to present in broad terms how the search and measurement are undertaken, before going into details on the specific sub-categories within the broader analysis.

First, the topology of the signal final state and corresponding backgrounds are presented. Next, an overview of the variables used to reduce the backgrounds and enhance the signal is given. These will be described in general, while specific values of selection cuts and background estimation will be provided in subsequent chapters. Finally, the parameters of interest in the search and measurement will be defined, and a brief overview of the statistical treatment of the final Higgs candidates is shown.
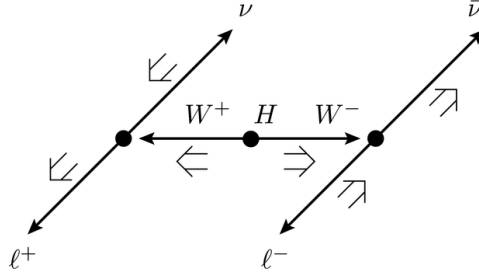
**Figure 3.1:** A cartoon of the WW final state. Momenta are represented with thin arrows, spins with thick arrows. [9]

## 3.2 Signal topology

The analysis presented here and in subsequent chapters is the study of the Higgs boson in the $WW$ final state, where each $W$ boson subsequently decays into a charged lepton and a neutrino. In its simplest form, the final state will then consist of two neutrinos and two charged leptons, each of which can be either an electron or a muon. If one or both of the $W$s decay to $\tau$ leptons, only leptonic decays of the $\tau$ are considered, leading to additional neutrinos in the final state but still giving two charged leptons as before. Neutrinos are not detected in ATLAS, so the final state ultimately consists of two reconstructed leptons and missing transverse momentum (denoted as $E_{\mathrm{T}}^{\mathrm{miss}}$). Final states where both of the charged leptons are electrons or muons are referred to as the "same flavor" final states, while those with one electron and one muon are referred to as "different flavor".

The final state leptons will also exhibit unique correlations due to the fact that they are arising from the decay of a spin zero resonance. In particular, the spins of the final state leptons and neutrinos must all cancel, as shown in figure 3.1. Because the neutrino has a left handed helicity and the anti-neutrino has a right handed helicity, the spin and momentum of the particles will be anti-aligned and aligned, respectively. In the transverse plane, the momenta of all four final state objects must cancel as well. With the constraint of having both the momenta and the spin alignments cancel, the final state kinematics strongly prefer having a small angle between the leptons in the transverse plane (low $\Delta\phi_{\ell\ell}$). This angular correlation will also lead to low values of the di-lepton invariant mass $m_{\ell\ell}$. These unique signal final state kinematic correlations will be exploited to define the ultimate signal region.

While the basic final state consists of two leptons and $E_\mathrm{T}^\mathrm{miss}$, there can be additional objects as well depending on the production mode of the Higgs. As described in detail in Chapter 1, if the Higgs is produced via vector boson fusion production, there will be two additional forward jets in the event. Even in gluon fusion, one or more jets can be produced through initial state radiation from the incoming gluons. The analysis is separated into different signal regions depending on the number of hard jets reconstructed in the final state as well.

## 3.3 Background processes

Many processes from the Standard Model can also produce a final state with two leptons and missing transverse momentum . This section lists the dominant backgrounds to Higgs production. It gives general descriptions of how the backgrounds mimic Higgs production and how they can be reduced. The details of background estimation and specific cuts are left for later sections. Table 3.1 summarizes the different processes.

### 3.3.1 Standard Model WW production
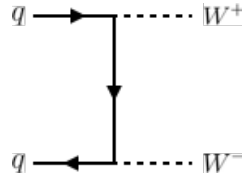


**Figure 3.2:** Feynman diagram for Standard Model WW production

Non-resonant Standard Model diboson production, as shown in figure 3.2, is an irreducible background to Higgs boson production in the WW final state. It produces the same exact final state objects, namely leptonically decaying W bosons. There are no additional objects in the final state that allow for background reduction. Therefore the analysis solely relies on the correlations between the leptons to reduce this background.

### 3.3.2 Top quark production

Production of top quarks, either in pairs ($t\bar{t}$ production) or singly (e.g. $Wt$ production), can also mimic Higgs production. Because top quarks decay via $t \rightarrow Wb$, top pair production can produce a final state with two W bosons that then decay leptonically. In this case, however, there are two additional jets from the bottom quarks in the final state. This allows the analysis to veto on the presence of jets identified as originating from a $b$ in order to reduce the size of the background.

Single top production can occur via $s$-channel, $t$-channel, or associated production ($Wt$). The mode which most closely resembles the Higgs final state is $Wt$. In this case, there are two real $W$ bosons produced, as with $t\bar{t}$. However, the decay of the single top quark will still also produce one $b$-jet, meaning a $b$ veto will reduce this background as well.

Figure 3.3 shows the Feynman diagrams for $t\bar{t}$ and $Wt$ production.
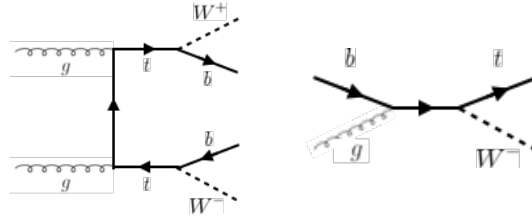


**Figure 3.3:** Feynman diagrams for top pair production (left) and $Wt$ production (right)

### 3.3.3 $W$+jets background

Single $W$ boson production, in association with jets, is a unique background. The other background considered so far have all included real leptons in the final state. In this case, however, only one real lepton from the decay of a $W$ exists in the final state. The second reconstructed lepton can arise from two different cases. First, the lepton may truly be an algorithm "fake", or a jet misidentified as a lepton by either the electron or muon reconstruction algorithms. Second, the lepton may be a real lepton but coming from semi-leptonic decays of particles inside the shower of the jet. This background can be reduced by requiring that the reconstructed lepton have little activity surrounding it in the calorimeter (also known as an "isolated" lepton). Figure 3.4 shows the Feynman diagram for $W$+jets production.
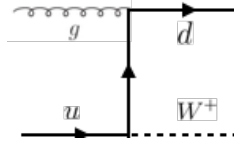
**Figure 3.4:** An example Feynman diagram of $W$+jets production

### 3.3.4 $Z/\gamma^*$+JETS BACKGROUND

Production of a $Z/\gamma^*$ in association with jets (also known as Drell-Yan) is also a background to Higgs production. In particular, the same flavor final states have a large $Z$+jets background, as the $Z$ decays into two leptons of the same flavor. (This background also enters the different flavor final state through the leptonic decays of $Z \to \tau\tau$). Figure 3.5 shows the production of a $Z$ in association with one jet. Because there are no neutrinos in this final state, variables like $E_{\mathrm{T}}^{\mathrm{miss}}$ can be used to reduce the background.
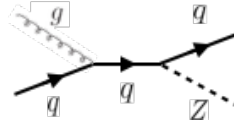


**Figure 3.5:** An example Feynman diagram of $Z$+jets production

### 3.3.5 OTHER (SUBDOMINANT) BACKGROUNDS

There are additional processes which contrinute to the background composition but are not produced as frequently as those listed already. The first of these are referred to as $VV$ or "Other diboson" processes and include multiple Standard Model diboson processes, including $WZ$, $ZZ$, $W\gamma$, $W\gamma^*$, and $Z\gamma$ production. Additionally, there is background from QCD multijet production, where two jets are misidentified as leptons.

## 3.4 ISOLATING AN $H \to WW^* \to \ell\nu\ell\nu$ SIGNAL

As presented in section 3.2, there are many different combinations of objects that can define a $H \to WW^* \to \ell\nu\ell\nu$ final state. The multiplicity of jets and the flavor combinations of the leptons both lead to a combinatorically large number of potential signal regions. Additionally, signal regions can be optimized separately to be sensitive to the distinct production modes of the Higgs. Gluon fusion, vector

| Category | Process | Description |
|---|---|---|
| SM $WW$ | $WW \to \ell\nu\ell\nu$ | Real leptons and neutrinos |
| Top quark production | $t\bar{t} \to WbW\bar{b} \to \ell\nu b\ell\nu\bar{b}$ | Real leptons, untagged $b$s |
| | $tW \to WbW \to \ell\nu\ell\nu b$ | Real leptons, untagged $b$ |
| | $t\bar{b}, tq\bar{b}$ | Untagged $b$, jet misidentified as lepton |
| Drell-Yan | $Z/\gamma^* \to ee, \mu\mu$ | "Fake" $E_T^{\text{miss}}$ |
| | $Z/\gamma^* \to \tau\tau \to \ell\nu\nu\ell\nu\nu$ | Real leptons and neutrinos |
| Other dibosons | $ZZ \to \ell\ell\nu\nu$ | Real leptons and neutrinos |
| | $W\gamma^*, WZ \to \ell\nu\ell\ell, ZZ \to \ell\ell\ell\ell$ | Unreconstructed leptons |
| | $W\gamma, Z\gamma$ | $\gamma$ reconstructed as $e$, unreconstructed lepton |
| $W$+jets | $Wj \to \ell\nu j$ | Jet reconstructed as lepton |
| QCD multijet | $jj$ | Jets reconstructed as leptons |

**Table 3.1:** A summary of backgrounds to the $H \to WW^* \to \ell\nu\ell\nu$ signal

boson fusion, and associated production of a Higgs all lead to unique final state topologies. Figure 3.6 delineates the different signal regions used in the gluon fusion and vector boson fusion $H \to WW^*$ analyses. While there are different optimizations possible in each signal region, there are also some commonly shared selections that will be described here.

### 3.4.1 Event pre-selection

Before being sorted into the distinct signal regions, basic cuts are applied on the reconstructed objects in the event to select Higgs-like event candidates. First, two oppositely charged leptons are required. The $p_T$ threshold on the leptons is a particularly important consideration for this signal. Because the second $W$ produced in the decay can be off-shell, it tends to produce lower momentum leptons. Thus, being able to lower the $p_T$ threshold while still maintaining a low background rate is critical. Figure 3.7 shows an example of the subleading lepton $p_T$ for a VBF $H \to WW^*$ signal compared to the corresponding $t\bar{t}$ background. Note that the lepton $p_T$ spectrum is considerably softer in the signal sample.

Once the leptons are selected, the last requirement for event pre-selection is the presence of neutrinos. As neutrinos cannot be detected directly in ATLAS, $E_T^{\text{miss}}$ can be used as a proxy for the combined neutrino momentum in the transverse plane. In general, it is expected that the signal should have a harder
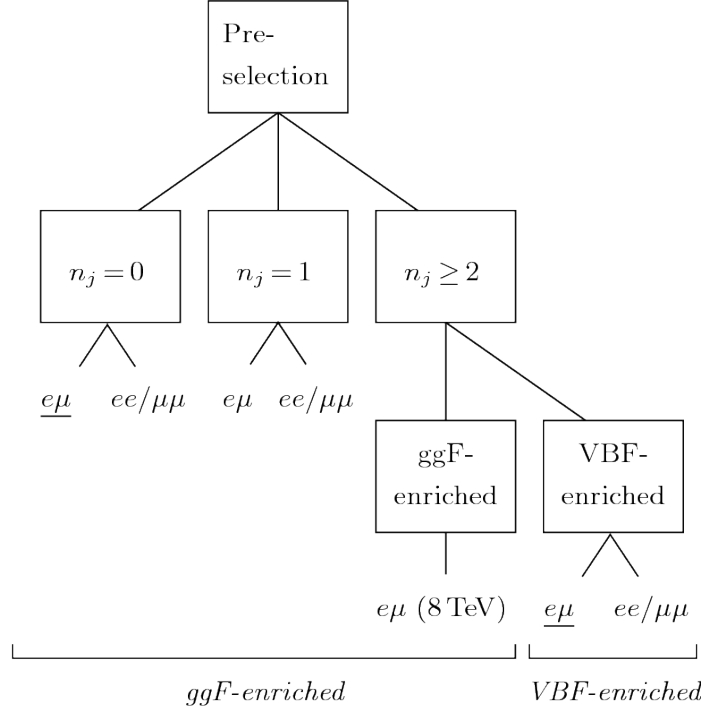
**Figure 3.6:** An illustration of the unique analysis signal regions [9]
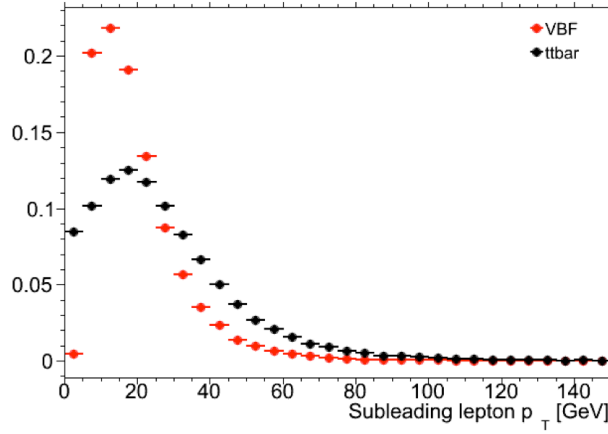


**Figure 3.7:** A comparison of the subleading lepton $p_\mathrm{T}$ spectrum between VBF $H \to WW^*$ production and $t\bar{t}$ background

$E_\mathrm{T}^\mathrm{miss}$ spectrum than backgrounds, especially if those backgrounds did not contain neutrinos. One additional consideration when using $E_\mathrm{T}^\mathrm{miss}$ is the fact that mis-measurements of objects in the detector can lead to imbalances in the transverse plane that are not due to real particles escaping the detector. One indicator that this is the case is that the $E_\mathrm{T}^\mathrm{miss}$ vector in the transverse plane will be pointing in the

same direction as the mis-measured object. Therefore, a new variable, $E_{\mathrm{T,rel}}^{\mathrm{miss}}$, is used in the pre-selection. $E_{\mathrm{T,rel}}^{\mathrm{miss}}$ is defined in equation 3.1.

$$E_{\mathrm{T,rel}}^{\mathrm{miss}} = \begin{cases} E_{\mathrm{T}}^{\mathrm{miss}} \, \sin \Delta\phi_{\mathrm{near}} & \text{if } \Delta\phi_{\mathrm{near}} < \pi/2 \\ E_{\mathrm{T}}^{\mathrm{miss}} & \text{otherwise,} \end{cases} \tag{3.1}$$

If the closest object to the $E_{\mathrm{T}}^{\mathrm{miss}}$ vector is within $\pi/2$ radians in the transverse plane, the $E_{\mathrm{T}}^{\mathrm{miss}}$ is projected away from this object. Otherwise, the normal $E_{\mathrm{T}}^{\mathrm{miss}}$ vector is used. Figure 3.8 shows a graphical illustration of this concept.
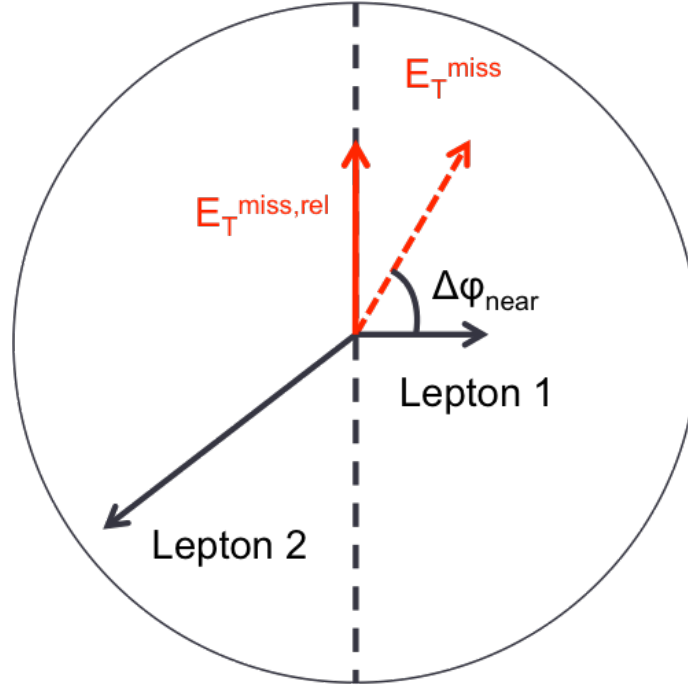


**Figure 3.8:** A graphical illustration of the $E_{\mathrm{T,rel}}^{\mathrm{miss}}$ calculation

Once both the lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$ pre-selections are made, the analysis can be divided into different regions according to jet multiplicity.

### 3.4.2 Jet multiplicity

Jet multiplicity, denoted as $n_j$, is used to sub-divide the analysis into its distinct signal regions. The reason for this is twofold. First, different jet multiplicity bins will be more or less sensitive to different Higgs

production modes. For example, the $n_j \geq 2$ region is more sensitive to VBF production because of the two hard jets produced at matrix element level. For gluon fusion production to enter this bin, two initial state radiation jets must be emitted. Second, background composition varies greatly in different bins of $n_j$. Figure 3.9 shows the jet multiplicity in both the different flavor and same flavor regions. It also shows the background composition in the bins of $n_b$. There are a few clear trends from this distribution. The first is that the Drell-Yan background dominates in the same flavor channels for $n_j \leq 1$. Second, the top background becomes a clear contributor to the total background for $n_j \geq 1$. Lastly, the SM WW production dominates in the $n_j = 0$ bin, as it is an irreducible background to $H \to WW^*$ production. Because of these distinct features, each jet multiplicity bin is treated separately.
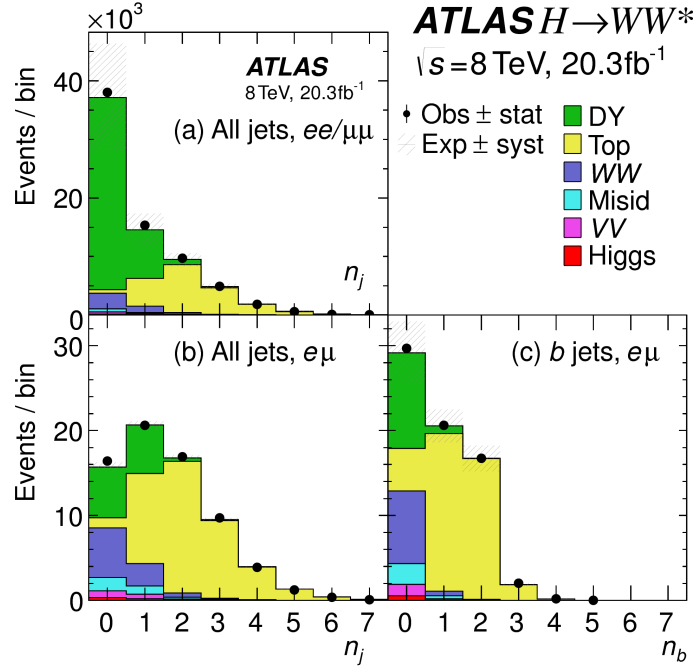


**Figure 3.9:** Predicted backgrounds (compared with data) as a function of $n_j$ (a and b) and $n_b$ (c)

## 3.5 BACKGROUND REDUCTION IN SAME-FLAVOR FINAL STATES

As described in section 3.4.2, the background composition of the same flavor final states is unique to that of the different flavor states. In particular, Drell Yan processes play a much larger role because the $Z/\gamma^*$ decays to same flavor leptons. Because real neutrinos are absent in the $Z/\gamma^*$ decays to $ee$ and $\mu\mu$,

a cut on $E_T^{miss}$ should largely reduce the background. However, as this section will demonstrate, with increasing pileup conditions the resolution of the calorimeter-based $E_T^{miss}$ degrades greatly. Therefore, two new variables for $Z/\gamma^*$ background reduction are constructed and described in this section.

### 3.5.1  PILEUP AND $E_T^{miss}$ RESOLUTION

Secondary interactions of protons in the colliding bunches of the LHC (known as pileup interactions, described in detail in Chapter 2) deposit energy into the ATLAS calorimeter on top of the energy that comes from the hard scatter process that is being searched for or analyzed. The calculation of $E_T^{miss}$ is fundamentally Poissonian, as summing up all of the energy deposits in individual calorimeter cells or clusters is similar to a counting experiment. Thus, the energy resolution scales as $\sqrt{E}$, just as the error on a mean of $N$ in a Poisson distribution is $\sqrt{N}$. As more energy is deposited in the calorimter, the $E_T^{miss}$ resolution degrades, meaning that the $E_T^{miss}$ resolution is particularly sensitive to LHC instantaneous luminosity conditions.

Figure 3.10 shows an event display of a $Z/\gamma^*$ + jets event candidate with the twenty-five reconstructed primary vertices. This display illustrates that while the interaction of interest only has tracks coming from the hardest primary vertex, all of the secondary interactions will deposit energy in the calorimeter as well.

Figure 3.11 shows the RMS of the $E_T^{miss}$ distribution in $Z \to \mu\mu$ events (where there are no real neutrinos) as a function of the number of the average number of interactions. Under 2011 LHC conditions, this RMS was approximately 9 GeV, while under 2012 running conditions the resolution worsened to 12 GeV. This worsening dilutes the efficacy of a cut on $E_T^{miss}$ to reduce the $Z/\gamma^*$ background.

### 3.5.2  TRACK-BASED DEFINITIONS OF MISSING TRANSVERSE MOMENTUM

Because the increasing number of secondary proton-proton interactions degrades calorimeter-based $E_T^{miss}$ resolution, a new variable using only contributions from the primary interaction vertex is necessary to further reduce the $Z/\gamma^*$ background. While it is not possible to associate calorimeter energy deposits with a particular vertex, individual charged particle tracks in the Inner Detector are associated to
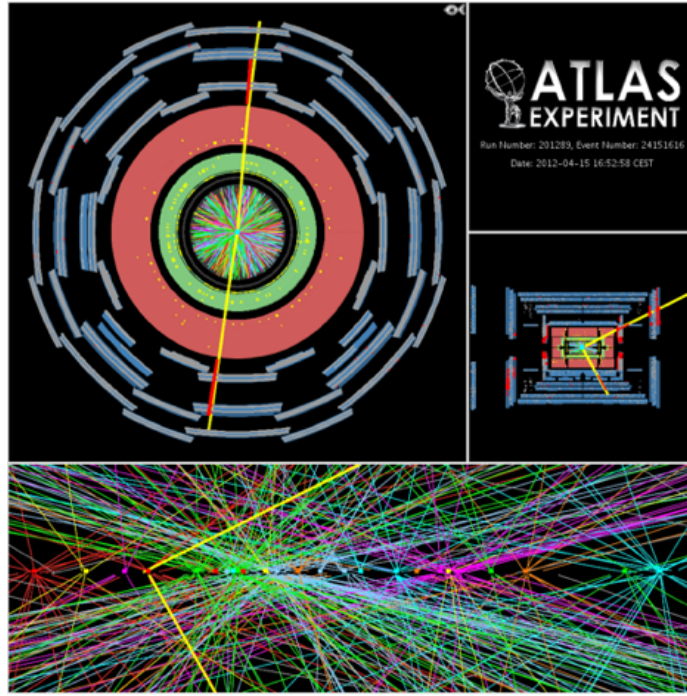
**Figure 3.10:** An event display of a $Z/\gamma^*$ + jets event illustrating the effect of pileup interactions
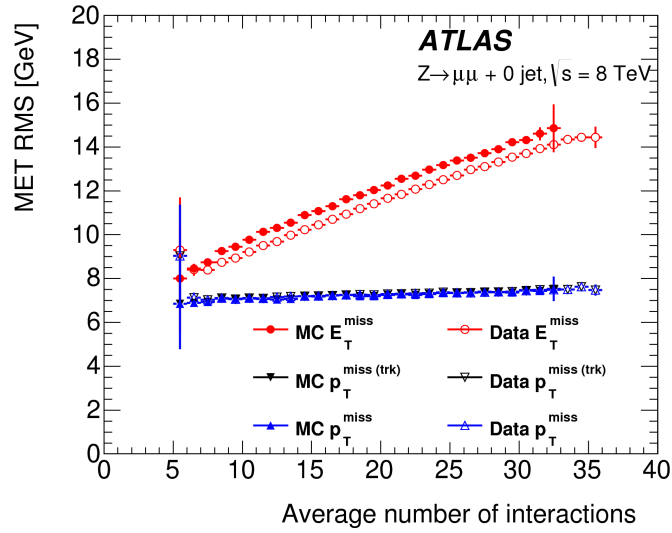


**Figure 3.11:** The RMS of different missing transverse momentum definitions as a function of the average number of interactions per bunch crossing

unique vertices. Thus, two track-based definitions of missing transverse momentum , using only tracks coming from the primary vertex in the event, are used in the analysis. The simplest variable, $p_{\mathrm{T}}^{\mathrm{miss\,(trk)}}$, is the vectorial sum of the $p_{\mathrm{T}}$ of all of the tracks from the primary vertex and the selected leptons (exclud-

ing the tracks associated with the selected leptons to avoid double counting). This is defined in equation 3.2.

$$p_T^{\text{miss (trk)}} = -\left( \sum_{\substack{\text{selected} \\ \text{leptons}}} p_T + \sum_{\substack{\text{other} \\ \text{tracks}}} p_T \right), \qquad (3.2)$$

In events with hard jets, a better resolution on the missing transverse momentum is obtained by including the calorimeter based measurement of the hard jets rather than the track based measurements. Thus, another variable, $p_T^{\text{miss}}$, is defined, using the nominal measurements of $p_T$ for the selected leptons and jets and using tracks rather than calorimeter clusters for the soft component of the missing transverse momentum . This is defined in equation 3.3.

$$p_T^{\text{miss}} = -\left( \sum_{\substack{\text{selected} \\ \text{leptons}}} p_T + \sum_{\substack{\text{selected} \\ \text{jets}}} p_T + \sum_{\substack{\text{other} \\ \text{tracks}}} p_T \right), \qquad (3.3)$$

Figure 3.11 illustrates that these two new variables accomplish their intended purpose. The resolution as a function of mean number of interactions for both $p_T^{\text{miss (trk)}}$ and $p_T^{\text{miss}}$ is much flatter compared to the dependence for $E_T^{\text{miss}}$.

Figure 3.12a shows the difference between the true and reconstructed values of missing transverse momentum using both the track-based $p_T^{\text{miss}}$ and calorimeter based $E_T^{\text{miss}}$. The RMS of the distribution improves by 3.5 GeVwhen using $p_T^{\text{miss}}$.

### 3.5.3 Distinguishing $Z/\gamma^*$ +jets and $H \to WW^*$ topologies

The track-based definitions of missing transverse momentum were constructed to mitigate degrading performance as a function of pileup. However, an additional variable can be constructed to exploit kinematic and topological differences between the $Z/\gamma^*$ background and $H \to WW^*$ signal. Because there are no real neutrinos in the final state (in the case of $Z/\gamma^* \to ee, \mu\mu$ decays), the dilepton system of a $Z/\gamma^*$ will be balanced with the jets produced in the hard scatter. A new variable, $f_{\text{recoil}}$, is constructed to estimate the balance between the dilepton system and the jets in the quadrant opposite the dilepton vector in the transverse plane. It is defined in equation 3.4. The numerator of $f_{\text{recoil}}$ is the magnitude of

17
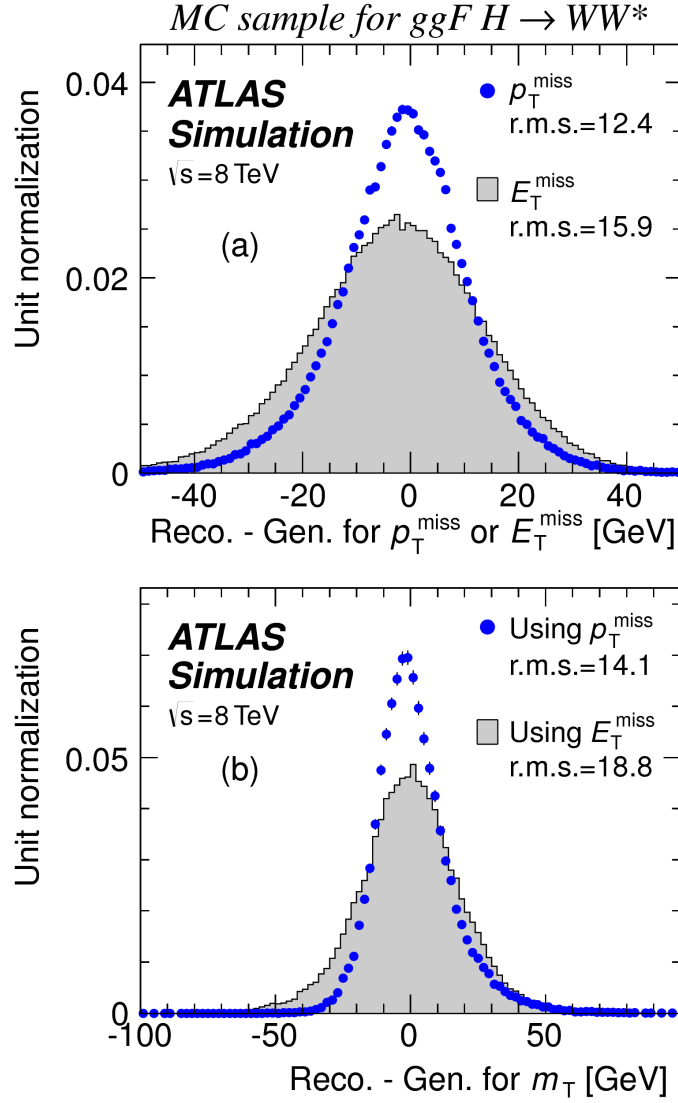
**Figure 3.12:** The difference between the true and reconstructed values of the missing transverse momentum (a) and $m_T$ (b) in a gluon fusion signal sample

the vectorial sum of the $p_T$ of jets in the quadrant opposite the dilepton system, weighted by each jet's Jet Vertex Fraction (JVF, described in chapter 2). The denominator is the magnitude of the dilepton $p_T$.

$$f_{\text{recoil}} = \left| \sum_{\text{jets } j \text{ in } \wedge} \text{JVF}_j \cdot \boldsymbol{p}_T^j \right| \Big/ p_T^{\ell\ell}. \tag{3.4}$$

Figure 3.13 shows a shape comparison of the distribution of $f_{\text{recoil}}$ in a simulated $Z/\gamma^*$ + jets sample,

a $H \rightarrow WW^*$ signal sample, and other backgrounds that contain real neutrinos. The $Z/\gamma^*$ + jets events tend to be more balanced between the dilepton system and recoiling jets, while the processes containing real neutrinos are less balanced in the transverse plane. Thus, a cut on $f_{\mathrm{recoil}}$ will also reduce the $Z/\gamma^*$ + jets background while maintaining a good signal efficiency.
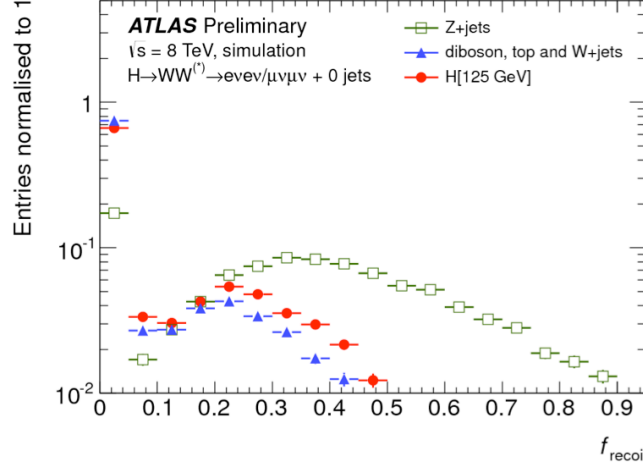


**Figure 3.13:** Comparison of $f_{\mathrm{recoil}}$ distributions for $Z/\gamma^*$+jets, $H \rightarrow WW^*$, and other backgrounds with real neutrinos.

### 3.5.4 OPTIMIZING BACKGROUND REDUCTION CUTS

The cuts on $p_{\mathrm{T}}^{\mathrm{miss\,(trk)}}$ and $f_{\mathrm{recoil}}$ used to reduce the Z+jets background must be optimized to maximize their efficacy. Figure 3.14 shows an early attempt to optimize the combination of the two cuts in the gluon fusion zero jet bin. Each bin shows the expected signal significance if the $p_{\mathrm{T,rel}}^{\mathrm{miss\,(trk)}}$ is required to be greater than the left edge of the bin and the $f_{\mathrm{recoil}}$ is required to be less than the top edge of the bin. The figure shows that the best signal significance comes from requiring low values of $f_{\mathrm{recoil}}$ ($<\,0.05$) and $p_{\mathrm{T,rel}}^{\mathrm{miss\,(trk)}}$ values greater than 45 GeV.

### 3.6 PARAMETERS OF INTEREST AND STATISTICAL TREATMENT

As with any search or measurement, there are particular parameters of the Higgs that the $H \rightarrow WW^*$ analysis is interested in measuring. In this case, the parameters of interest are the mass of the Higgs boson and its production cross section. Because the $H \rightarrow WW^* \rightarrow \ell\nu\ell\nu$ process does not have a closed
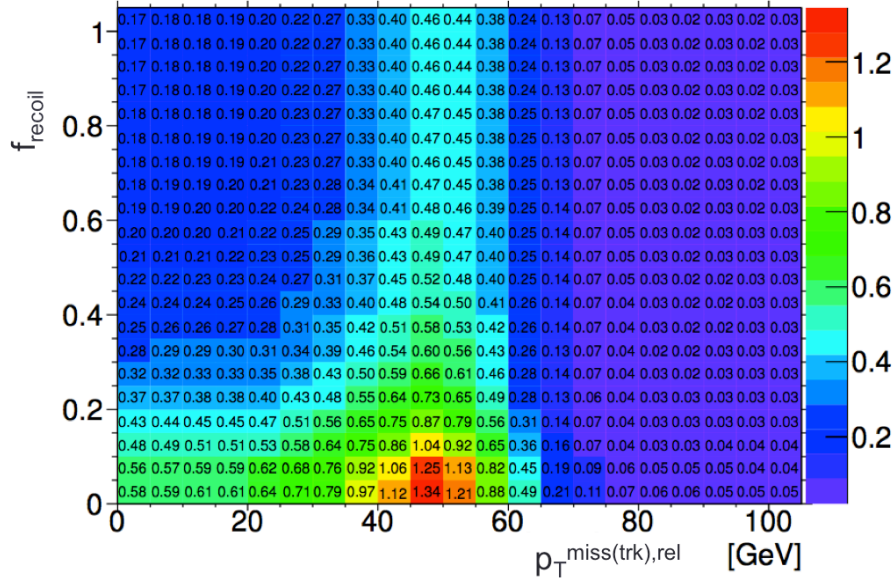
Figure 3.14: Signal significance as a function of cut value in the ggF $H \to WW^*$ with $n_j = 0$

final state, it is not possible to measure the full invariant mass of the particle that may have produced the final state. However, a proxy for the invariant mass using transverse plane information can be defined. This is described in more detail in section 3.6.1. The second parameter of interest is the ratio of the measured cross section to that expected from the Standard Model Higgs, which is denoted a $\mu$. This is defined in equation 3.5.

$$\mu = \frac{\sigma}{\sigma_{\text{SM}}} \tag{3.5}$$

All of the likelihoods used in the statistical analysis of the final signal region events are paramaterized as a function of $\mu$. $\mu$ is a natural variable for hypothesis testing, as $\mu = 0$ corresponds to a background only hypothesis and $\mu = 1$ corresponds exactly to a Standard Model Higgs.

### 3.6.1 TRANSVERSE MASS

Because the longitudinal information about the neutrinos is not attainable, the $H \to WW^* \to \ell\nu\ell\nu$ analysis uses a mass variable, the transverse mass, that exploits information in the transverse plane as a

proxy for the full invariant mass. The transverse mass is defined in equation 3.6.

$$m_{\mathrm{T}} = \sqrt{\left(E_{\mathrm{T}}^{\ell\ell} + p_{\mathrm{T}}^{\mathrm{miss}}\right)^2 - \left| \boldsymbol{p}_{\mathrm{T}}^{\ell\ell} + \boldsymbol{p}_{\mathrm{T}}^{\mathrm{miss}} \right|^2},$$ (3.6)

Here the $E_{\mathrm{T}}^{\ell\ell}$ and $p_{\mathrm{T}}^{\ell\ell}$ are the transverse energy and momentum of the dilepton system, while $p_{\mathrm{T}}^{\mathrm{miss}}$ is a proxy for the transverse momentum of the di-neutrino system. The track-based $p_{\mathrm{T}}^{\mathrm{miss}}$ is used in the $m_{\mathrm{T}}$ rather than the calorimeter based $E_{\mathrm{T}}^{\mathrm{miss}}$ because it has a better resolution on the true transverse mass. Figure 3.12b shows the improvement in the RMS of the difference between the true and reconstructed transverse mass in a ggF signal sample. The RMS improves by 4.7 GeV using $p_{\mathrm{T}}^{\mathrm{miss}}$ in the $m_{\mathrm{T}}$ calculation.

### 3.6.2   STATISTICAL TREATMENT[*]

#### LIKELIHOOD FUNCTION

The statistical analysis of final event candidates is framed as a hypothesis test, where the null hypothesis is background-only (no Standard Model Higgs). The first step in the analysis is to form a likelihood function for the data. In its simplest form, this likelihood is the probability of observing the number of events seen in the final signal region given knowledge of the signal strength. Because observation of events is fundamentally a Poisson counting experiment, this simple likelihood can be expressed as a Poisson probability of observing $N$ events given a total number of predicted signal and background events. This basic likelihood is shown in equation 3.7.

$$\mathcal{L}(\mu) = P\left(N | \mu S + B\right)$$ (3.7)

Here, $P$ is the Poisson probability density function, $N$ is the total number of observed events, $\mu$ is the signal strength, $S$ is the predicted number of signal events, and $B$ is the predicted number of background events.

In particle physics, certain background estimates are commonly normalized in so-called "control" re-

---

[*]Many thanks to Aaron Armbruster, whose thesis[4] inspired parts of this section.

gions and those predictions are scaled by the same normalization factor in the signal region. This leads to a slightly more complicated likelihood, which is a function of both the signal strength and the background normalization. This is shown in equation 3.8.

$$\mathcal{L}(\mu, \theta) = P\left(N | \mu S + \theta B\right) P\left(N_{\mathrm{CR}} | \theta B_{\mathrm{CR}}\right) \tag{3.8}$$

Here, $\theta$ is a so-called "nuisance parameter", a parameter that is not a primary parameter of interest but still enters the likelihood. The second Poisson term adds an extra term to the likelihood, enforcing the fact that the background normalization must be consistent with the number of observed events in data in the control region, $N_{\mathrm{CR}}$.

So far, these two formulations of likelihoods have assumed a single signal region and do not take into account any shape information of potential discriminating variables. The $H \to WW^*$ analysis is divided into many different categories, and we can perform the same counting experiment described above in each individual category. As mentioned in section 3.6.1, the transverse mass is used as the primary discriminating variable in many of the $H \to WW^*$ sub-analyses, so additionally we can perform the same counting experiment in each bin of the $m_{\mathrm{T}}$ distribution to incorporate some shape information. Thus, the total likelihood becomes a product over signal regions and bins of the $m_{\mathrm{T}}$ distribution. Finally, there are usually many backgrounds that are normalized in control regions, so the new formulation of the likelihood takes this into account as well by including a product over control regions in the second Poisson term. All of these modifications are shown in equation 3.9.

$$\mathcal{L}(\mu, \boldsymbol{\theta}) = \prod_{\substack{\mathrm{SRs\ i} \\ \mathrm{bins\ b}}} P\left(N_{ib} \middle| \mu S_{ib} + \sum_{\mathrm{bkg\ k}} \theta_k B_{kib}\right) \prod_{\mathrm{CRs\ l}} P\left(N_l \middle| \sum_{\mathrm{bkg\ k}} \theta_k B_{kl}\right) \tag{3.9}$$

The final step to get the full likelihood used in the analysis is to add nuisance parameters for the systematic uncertainties. In cases where the uncertainty does not affect the shape of $m_{\mathrm{T}}$ bin-by-bin, each systematic uncertainty $\epsilon$ is allowed to affect the expected event yields through a linear response function of the nuisance parameter, namely $\nu(\theta) = (1 + \epsilon)^{\theta}$. If instead the uncertainty does affect the shape, the effect is instead parameterized by $\nu_b(\theta) = 1 + \epsilon_b \theta$. The value of the nuisance parameters for the

systematic uncertainty are constrained with a Gaussian term that is added to the likelihood as well. This is of the form $g(\delta|\theta) = e^{-(\delta-\theta)^2/2}/\sqrt{2\pi}$, where $\delta$ is the central value and $\theta$ is a nuisance parameter. Finally, a last term is added to account for the statistical uncertainty in the Monte Carlo samples used, which adds an additional poisson term. The full likelihood used in the final statistical analysis is defined in equation 3.10.

$$
\begin{aligned}
\mathcal{L}(\mu, \boldsymbol{\theta}) = \prod_{\substack{\text{SRs } i \\ \text{bins } b}} P\left(N_{ib} \middle| \mu S_{ib} \cdot \prod_{\substack{\text{sig.} \\ \text{syst.} \\ r}} \nu_{br}(\theta_r) + \sum_{\text{bkg } k} \theta_k B_{kib} \cdot \prod_{\substack{\text{bkg.} \\ \text{syst.} \\ s}} \nu_{bs}(\theta_s)\right) \\
\cdot \prod_{\text{CRs } l} P\left(N_l \middle| \sum_{\text{bkg } k} \theta_k B_{kl}\right) \\
\cdot \prod_{\substack{\text{syst} \\ t}} g(\delta_t|\theta_t) \cdot \prod_{\text{bkg } k} P(\xi_k|\zeta_k \theta_k)
\end{aligned}
\tag{3.10}
$$

In the fourth term of the equation, quantifying uncertainty due to finite Monte Carlo sample size, $\xi$ represents the central value of the background prediction, $\theta$ is the associated nuisance parameter, $\zeta = (B/\delta B)^2$, where $\delta B$ is the statistical uncertainty of $B$.

The best fit value of the signal strength $\mu$ is determined by finding the values of $\mu$ and $\boldsymbol{\theta}$ that maximize the likelihood, while setting $\delta = 0$ and $\xi = \zeta$.

Once the likelihood is defined, a test statistic must be built for use in hypothesis testing.

### Test statistic

To distinguish whether the data match a background only or background and signal hypothesis, a test statistic must be used. The $H \rightarrow WW^*$ analysis used the profile likelihood technique[11]. The first step in formulating this test statistic is to define the profile likelihood ratio, shown in equation 3.11.

$$
\lambda(\mu) = \frac{\mathcal{L}\left(\mu, \hat{\hat{\theta}}_\mu\right)}{\mathcal{L}\left(\hat{\mu}, \hat{\theta}\right)}
\tag{3.11}
$$

Here $\hat{\hat{\theta}}_\mu$ is the value of $\theta$ that maximizes the likelihood for the choice of $\mu$ being tested. Additionally,

$\hat{\theta}$ and $\hat{\mu}$ represent the values of $\theta$ and $\mu$ that gives the overall maximum value of the likelihood.

Once this is defined, a test statistic $q_\mu$ is constructed. This is shown in equation 3.12.

$$q_\mu = -2 \ln \lambda(\mu) \tag{3.12}$$

A higher value of $q_\mu$ means that the data are more incompatible with the hypothesized value of $\mu$, and $q_0$ then corresponds to the value of the test statistic for the background only hypothesis. A $p_0$ value is then defined to quantify the compatibility between the data and the null hypothesis. The $p_0$ value is the probability of obtaining a value of $q_0$ larger than the observed value, and this is shown in equation 3.13.

$$p_0 = \int_{q_0^{\text{obs}}}^{\infty} f(q_\mu | \mu = 0) dq_\mu \tag{3.13}$$

Here $f(q_\mu)$ is the probability distribution function of the test statistic. Finally, the $p_0$ value can be converted into a signal significance, using the formula in equation3.14, or the one-sided tail of the Gaussian distribution.

$$Z_0 = \sqrt{2} \text{erf}^{-1}(1 - 2p_0) \tag{3.14}$$

The threshold for discovery used in particle physics is $Z_0 \geq 5$, more commonly known as a value of $5\sigma$.

# 4

The discovery of the Higgs boson and the role of the $H \rightarrow WW^* \rightarrow \ell\nu\ell\nu$ channel

*The imagination of nature is far, far greater than the imagination of man.*

Richard Feynman

# 5

# Observation of Vector Boson Fusion production of $H \rightarrow WW^* \rightarrow \ell\nu\ell\nu$

## 5.1  Introduction

After the discovery of a particle consistent with the Higgs boson, the $H \rightarrow WW^*$ analysis had two main goals. The first goal was to increase the sensitivity of the analysis to fully confirm that the $H \rightarrow WW^*$ process did indeed exist. The second goal was to characterize the particle as much as possible, including searching for the lower cross-section production modes, in order to confirm that it was indeed a Higgs boson. This chapter presents a dedicated search for Vector Boson Fusion (VBF) production of a Higgs boson decaying via the $H \rightarrow WW^* \rightarrow \ell\nu\ell\nu$ mode. First, basics of the topology of VBF production are presented. Then, the details of the analysis are shown, including signal region definition, background estimation techniques, and systematic uncertainties. Finally, the results of the analysis are shown. As will
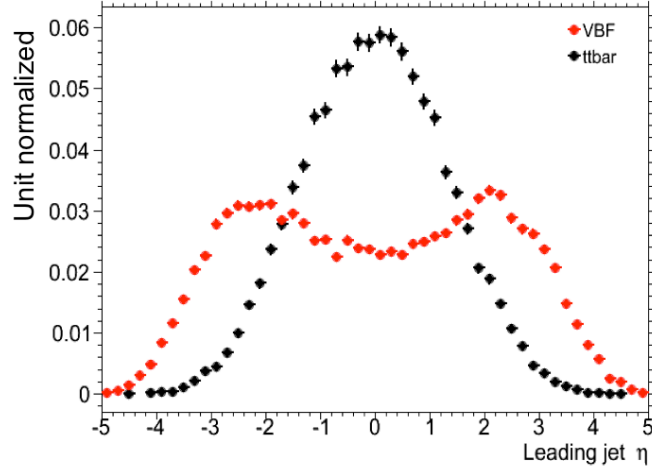
**Figure 5.1:** Leading jet $\eta$ in VBF $H \to WW^*$ (red) and $t\bar{t}$ (black)

be shown, this analysis is the first and most sensitive observation of the VBF production mode of the Higgs on ATLAS.

## 5.2  Topology of VBF $H \to WW^*$ production

As discussed in Chapter 1, the characteristic feature of VBF production of the Higgs is the presence of two additional forward jets coming from the incoming partons which radiate the vector bosons that make the Higgs. These jets are forward because the outgoing partons still carry the longitudinal momentum of the incoming partons. Figure 5.1 shows the distribution of the $\eta$ for the leading jet in a VBF event compared to a background top pair production event. As can be seen, the VBF jets tend to be more forward in $\eta$, while the $t\bar{t}$ jets are more central.

Because the cross section for VBF production is about an order of magnitude smaller than gluon fusion production, these forward jets must be used in order to better reduce background and achieve a good signal to background ratio. The analysis selection is constructed to maximally exploit the features of the unique VBF topology.

27

## 5.3 Data and simulation samples

The results presented here are with 20.3 fb$^{-1}$ taken at $\sqrt{s} = 8$ TeV and 4.5 fb$^{-1}$ taken at $\sqrt{s} = 7$ TeV. The details of the LHC and detector conditions during this period are given in Chapter 2. The trigger selection defining the dataset is discussed in section 5.3.1. The simulation samples used for signal and background modeling are given in section 5.3.2.

### 5.3.1 Triggers

The analysis uses a combination of single lepton and dilepton triggers to allow lowering of the $p_T$ thresholds and increased signal acceptance. As discussed in Chapter 2, there are multiple levels in the ATLAS trigger system, and there are different $p_T$ thresholds imposed for the leptons at each level. Additionally, some triggers have a loose selection on the isolation of the lepton (looser than that applied offline in the analysis object selection). Table 5.1 shows the thresholds used for single lepton triggers, while table 5.2 shows the thresholds coming from di-lepton triggers. The single lepton trigger efficiency for muons that pass the analysis object selection is 70% for muons in the barrel region ($|\eta| < 1.05$) and 90% in the endcap region. The electron trigger efficiency increases with electron $p_T$ but the average is approximately 90%. These efficiencies are measured by combined performance and trigger signature groups[7,8].

| | Level-1 threshold | High-level threshold |
|---|---|---|
| Electron | 18 | $24i$ |
| | 30 | 60 |
| Muon | 15 | $24i$ |
| | | 36 |

Table 5.1: Single lepton triggers used for electrons and muons. A logical "or" of the triggers listed for each lepton type is taken. Units are in GeV, and the $i$ denotes an isolation requirement in the trigger.

The combination of all triggers shown gives good efficiency for signal events. This efficiency is summarized in table 5.3. The relative improvement in efficiency by adding the dilepton triggers is also shown in the same table. The largest gain comes in the $\mu\mu$ channel. Overall the trigger selection shows a good efficiency for $H \to WW^*$ signal events.

| | Level-1 threshold | High-level threshold |
|---|---|---|
| $ee$ | 10 and 10 | 12 and 12 |
| $\mu\mu$ | 15 | 18 and 8 |
| $e\mu$ | 10 and 6 | 12 and 8 |

**Table 5.2:** Di-lepton triggers used for different flavor combinations. The two thresholds listed refer to leading and sub-leading leptons, respectively. The di-muon trigger only requires a single lepton at level-1.

| Channel | Trigger efficiency | Gain from $2\ell$ trigger |
|---|---|---|
| $ee$ | 97% | 9.1% |
| $\mu\mu$ | 89% | 18.5% |
| $e\mu$ | 95% | 8.3% |
| $\mu e$ | 81% | 8.2% |

**Table 5.3:** Trigger efficiency for signal events and relative gain of adding a dilepton trigger on top of the single lepton trigger selection. The first lepton is the leading, while the second is the sub-leading. Efficiencies shown here are for the ggF signal in the $n_j = 0$ category but are comparable for the VBF signal.

### 5.3.2 MONTE CARLO SAMPLES

Modeling of signal and background processes in the signal region, in particular for the $m_\mathrm{T}$ distribution, is an important consideration for the final interpretation of the analysis. Therefore, careful consideration must be paid to which Monte Carlo (MC) generators are used for specific processes. With the exception of the $W$+jet and multijet backgrounds, the $m_\mathrm{T}$ shape used as the final discriminant is taken from simulation. (Many backgrounds are normalized from data, as described in section 5.6).

Table 5.4 shows the MC generators used for the signal and background processes, as well as their cross sections. In order to include corrections up to next-to-leading order (NLO) in the QCD coupling constant $\alpha_s$, the POWHEG[16] generator is often used. In some cases, only leading order generators like AC-ERMC[5] and GG2VV[14] are available for the process in question. If the process requires good modeling for very high parton multiplicities, the SHERPA[13] and ALPGEN[15] generators are used to provide merged calculations for five or fewer additional partons. These matrix element level calculations must then be additionally matched to models of the underlying event, hadronization, and parton shower. There are four possible generators for this: SHERPA , PYTHIA 6[20], PYTHIA 8[21], or HERWIG[10] + JIMMY[6]. The simulation additionally requires an input parton distribution function (PDF). The CT10[12] PDFs are used for SHERPA and POWHEG simulated samples, while CTEQ6L1[17] is used for ALPGEN +HERWIG and ACERMC

| Process | | MC generator | $\sigma \cdot \mathcal{B}$ (pb) |
|---|---|---|---|
| **Signal** | | | |
| ggF | $H \to WW^*$ | POWHEG +PYTHIA 8 | 0.435 |
| VBF | $H \to WW^*$ | POWHEG +PYTHIA 8 | 0.0356 |
| VH | $H \to WW^*$ | PYTHIA 8 | 0.0253 |
| *WW* | | | |
| $q\bar{q} \to WW$ and $qg \to WW$ | | POWHEG +PYTHIA 6 | 5.68 |
| $gg \to WW$ | | GG2VV +HERWIG | 0.196 |
| $(q\bar{q} \to W) + (q\bar{q} \to W)$ | | PYTHIA 8 | 0.480 |
| $q\bar{q} \to WW$ | | SHERPA | 5.68 |
| VBS $WW + 2$ jets | | SHERPA | 0.0397 |
| Top quarks | | | |
| $t\bar{t}$ | | POWHEG +PYTHIA 6 | 26.6 |
| $Wt$ | | POWHEG +PYTHIA 6 | 2.35 |
| $tq\bar{b}$ | | ACERMC +PYTHIA 6 | 28.4 |
| $t\bar{b}$ | | POWHEG +PYTHIA 6 | 1.82 |
| Other dibosons $(VV)$ | | | |
| $W\gamma$ | $(p_T^\gamma > 8\,\mathrm{GeV})$ | ALPGEN +HERWIG | 369 |
| $W\gamma^*$ | $(m_{\ell\ell} \le 7\,\mathrm{GeV})$ | SHERPA | 12.2 |
| $WZ$ | $(m_{\ell\ell} > 7\,\mathrm{GeV})$ | POWHEG +PYTHIA 8 | 12.7 |
| VBS $WZ + 2$ jets | $(m_{\ell\ell} > 7\,\mathrm{GeV})$ | SHERPA | 0.0126 |
| $Z\gamma$ | $(p_T^\gamma > 8\,\mathrm{GeV})$ | SHERPA | 163 |
| $Z\gamma^*$ | $(\mathrm{min.}\ m_{\ell\ell} \le 4\,\mathrm{GeV})$ | SHERPA | 7.31 |
| $ZZ$ | $(m_{\ell\ell} > 4\,\mathrm{GeV})$ | POWHEG +PYTHIA 8 | 0.733 |
| $ZZ \to \ell\ell\,\nu\nu$ | $(m_{\ell\ell} > 4\,\mathrm{GeV})$ | POWHEG +PYTHIA 8 | 0.504 |
| Drell-Yan | | | |
| $Z$ | $(m_{\ell\ell} > 10\,\mathrm{GeV})$ | ALPGEN +HERWIG | 16500 |
| VBF $Z + 2$ jets | $(m_{\ell\ell} > 7\,\mathrm{GeV})$ | SHERPA | 5.36 |

**Table 5.4:** Monte Carlo samples used to model the signal and background processes [9].

simulations. The Drell-Yan samples are reweighted to the MRST[19] PDFs, as these are found to give the best agreement between data and simulation.

Once the basic hard scattering process is simulated, it must be passed through a detector simulation

and additional pile-up events must be overlaid. The pile-up events are modeled with PYTHIA 8, and the ATLAS detector is simulated with GEANT4 [18]. Because of the unique phase space of the $H \to WW^*$ analysis, events are sometimes filtered at generator level to allow for more efficient generation of relevant events. The efficiency of the trigger in MC simulation does not always match the measured efficiency in data, so trigger scale factors are applied to correct the MC efficiency to the data. These are derived by the combined performance groups [7,8].

## 5.4   OBJECT SELECTION

In order to define the signal region, the analysis must first select the objects to be considered. The details of the object reconstruction algorithms are discussed in Chapter 2, while this section gives specific selection cuts used in the $H \to WW^*$ analysis.

The first step in this process is to select a primary vertex candidates. The event's primary vertex is the vertex with the largest sum of $p_T^2$ for associated tracks and is required to have at least three tracks with $p_T > 450$ MeV. Many of the object selection cuts are then made relative to this chosen primary vertex.

### 5.4.1   MUONS

The analysis uses combined muon candidates, where a track in the Inner Detector has been matched to a standalone track in the Muon Spectrometer. The track parameters are combined statistically in the muon reconstruction algorithm [3]. The muons are required to be within $|\eta| < 2.5$ and have a $p_T > 10$ GeV. To reduce backgrounds coming from mis-reconstructed leptons, there are requirements on the impact parameter of the muon relative to the primary vertex. The transverse impact parameter $d_0$ is required to be small relative to its estimated uncertainty, the exact cut value being $d_0/\sigma_{d_0} < 3$. The longitudinal impact parameter $z_0$ must satisfy $|z_0 \sin \theta| < 1$ mm.

As discussed previously, the muons must also be isolated. There are two types of lepton isolations that are calculated: track-based and calorimeter-based. For muons, the track-based isolation is defined using the scalar sum $\sum p_T$ for tracks with $p_T > 1$ GeV (excluding the muon's track) within a cone of $\Delta R = 0.3 \, (0.4)$ for muon with $p_T > 15$ GeV $(10 < p_T < 15$ GeV$)$. The final isolation requirement

is made my requiring that this scalar sum be no more than a certain fraction of the muon's $p_T$. This requirement varies with muon $p_T$ and the exact cuts are defined in table 5.5.

The calorimeter-based muon isolation is defined using as a $\sum E_T$ calculated from calorimeter cells using the same cone size as the track-based isolation but excluding cells with $\Delta R < 0.05$ around the muon. This requirement is also definedas a cut on the ratio of the sum to the muon $p_T$ and varies with muon $p_T$. The cut values are also given in table 5.5.

The isolation requirements loosen as a function of $p_T$ to allow for larger signal acceptance. At low $p_T$, the isolation is tightened to reduce the $W$+jets background which arises from a misidentified lepton.

| $p_T$ range (GeV) | Calorimeter isolation | Track isolation |
|:---:|:---:|:---:|
| $10 - 15$ | 0.06 | 0.06 |
| $15 - 20$ | 0.12 | 0.08 |
| $20 - 25$ | 0.18 | 0.12 |
| $> 25$ | 0.30 | 0.12 |

**Table 5.5:** $p_T$ dependent isolation requirements for muons. Muons are required to have the amount of calorimeter or track based cone sums be less than this fraction of their $p_T$.

### 5.4.2 ELECTRONS

Electrons are identified by matching reconstructed clusters in the electromagnetic calorimeter with tracks in the inner detector. The electrons are identified using a likehood based method [2,1] which takes into account the shower shapes in the calorimeter, the matching of tracks to clusters, and the amount of transition radiation in the TRT. The electrons are required to have $|\eta| < 2.47$, and candidates in the transition region between the barrel and endcap ($1.37 < |\eta| < 1.52$) are excluded. As the muons, the electrons are required to have transverse impact parameter significance $< 3$, while in the longitudinal direction they must have $|z_0 \sin \theta| < 0.4$ mm. Some electron requirements also vary with electron $E_T$, and these requirements are summarized in table 5.6.

The isolation for electrons are defined similarly to the muons but with unique cuts on the objects included. The track-based isolation is defined using tracks with $p_T > 400$ MeV with cone sizes as defined previously. The calorimeter-based isolation also uses the same cone size as the muon, but here the cells

within a $0.125 \times 0.175$ area in $\eta \times \phi$ around the electron cluster's barycenter are excluded. The other difference with respect to muons is that the denominator of the isolation ratio is the electron's $E_T$ rather than $p_T$. The isolation cuts very with electron $E_T$ and are defined in table 5.6.

The electron is also required to not be consistent with a vertex coming from a photon conversion.

| $p_T$ range (GeV) | Quality cut | Calorimeter isolation | Track isolation |
|---|---|---|---|
| $10 - 15$ | Very tight LH | 0.20 | 0.06 |
| $15 - 20$ | Very tight LH | 0.24 | 0.08 |
| $20 - 25$ | Very tight LH | 0.28 | 0.10 |
| $> 25$ | Medium | 0.28 | 0.10 |

Table 5.6: $p_T$ dependent requirements for electrons. Electrons are required to have the amount of calorimeter or track based cone sums be less than this fraction of their $E_T$.

### 5.4.3 JETS

Jets are clustered with the anti-$k_T$ reconstruction algorithm using a radius parameter of $R = 0.4$. They are required to have a jet vertex fraction (JVF) of at least 50%, meaning that half of the tracks associated with the jet originated from the primary vertex. Jets with no tracks associated (i.e. those outside the acceptance of the ID) do not have this requirement applied. Jets are required to have $p_T > 25$ GeV if they are within the tracking acceptance ($|\eta| < 2.4$). Jets with $2.4 < |\eta| < 4.5$ are required to have $p_T > 30$ GeV. This tighter requirement reduces jets from pileup in the region where JVF requirements cannot be applied. The two highest $p_T$ jets in the event are referred to as the "VBF" jets and used to compute various analysis selections later.

Identification of $b$-jets is done using the MV1 algorithm and is limited to the acceptance of the ID ($|\eta| < 2.5$). The operating point of MV1 that is used is the one that is 85% efficient for identifying true $b$-jets. This operating point has a 10.3% of mis-tagging a light quark jet as a $b$-jet. In order to improve the rejection of $b$-jets, a lower threshold than the nominal $p_T$ threshold described above is used. For the purposes of counting the number of $b$-jets, jets with $p_T$ down to 20 GeV are used.

### 5.4.4 Overlap removal

There are some cases where certain reconstructed objects will overlap and one will have to be chosen (for example, an electron and a jet in the calorimeter). First, the case of lepton overlap is dealt with. If an electron candidate extends into the muon spectrometer, it is removed. If a muon or electron have a $\Delta R < 0.1$, the electron is removed and the muon is kept. If two electron candidates overlap within the same radius, then the higher $E_T$ electron is kept. Next, the overlap between leptons and jets is considered. If an electron and jet are within $\Delta R < 0.3$ of one another, the electron is kept and the jet is removed. However, if a muon and jet overlap within $\Delta R < 0.3$, the jet is kept (as it is likely that the muon is the result of a semileptonic decay inside the jet).

Once the overlap removal is complete, the final set of objects used in the analysis is defined.

### 5.5 Analysis selection

The VBF analysis uses two distinct selections. The first is a looser selection that uses a Boosted Decision Tree (BDT) score as the final discriminator in order to take advantage of the detailed correlations between the VBF variables. The second is a more standard selection, referred to as "cut-based", that applies cuts on the VBF variables and uses $m_T$ as the final discriminating variable. While the BDT analysis is ultimately more sensitive, the cut-based serves as an important component of the analysis. First, the cut-based allows for confirming the modeling and validity of many variables used as input to the BDT. Second, because this is the first use of such an MVA technique in the $H \to WW^*$ analysis, the cut-based selection allows confirmation of the final BDT result with a more traditional analysis. Both analyses will be discussed here.

One important note is that because this analysis is dedicated to the measurement of the VBF production mode of the Higs, events coming from gluon fusion production with the Higgs decaying via $H \to WW^* \to \ell\nu\ell\nu$ are treated as background events. This will be seen throughout the vargious cutflow tables and yields shown.

34

Both the BDT and cut-based analyses have a common pre-selection that is applied before their main cuts. The cuts on leptons are common to all $n_j$ bins. The analysis requires two oppositely charged leptons, with the leading lepton required to have $p_\mathrm{T} > 22$ GeVwhile the subleading lepton must have $p_\mathrm{T} > 10$ GeV. Next, to cut out low mass $Z/\gamma^*$ events, a cut on the dilepton mass $m_{\ell\ell} > 10\,(12)$ GeVis applied in the different (same) flavor channel. In the same flavor channels, there is an additional veto placed on the region around the Z peak, requiring that $|m_{\ell\ell} - m_Z| > 15$ GeV.

There are also requirements on the amount of missing transverse momentum in the event. These are only applied in the same flavor channels, as in the different flavor channels $t\bar{t}$ is the dominant background in $n_j \geq 2$. The BDT analysis requires $p_\mathrm{T}^\mathrm{miss} > 40\,\mathrm{GeV}$ and $E_\mathrm{T}^\mathrm{miss} > 45\,\mathrm{GeV}$. The cut-based analysis must cut tighter on these variables to have maximal sensitivity and thus requires $p_\mathrm{T}^\mathrm{miss} > 50\,\mathrm{GeV}$ and $E_\mathrm{T}^\mathrm{miss} > 55\,\mathrm{GeV}$.

Finally, because this analysis is focused on VBF, a requirement on the jet multiplicity is placed, with $n_j \geq 2$. Additionally, the analysis requires that there are no jets identified as b-quarks in the event, or $n_b = 0$.

## 5.5.2   CUT-BASED SELECTION

The cut-based selection places sequential requirements on variables reconstructed from the VBF jets in order to increase the signal to background ratio.

### GENERAL BACKGROUND REDUCTION

Top pair production is the primary background in the $n_j \geq 2$ bin. Even though $n_b = 0$ is required, an additional variable is constructed to further suppress the top background. There is often additional QCD radiation that accompanies the $t\bar{t}$ system when it is produced. Therefore, a variable which tests for the presence of this additional radiation, $p_\mathrm{T}^\mathrm{sum}$, is constructed. It is defined in equation 5.1.

$$\boldsymbol{p}_\mathrm{T}^\mathrm{sum} = \boldsymbol{p}_\mathrm{T}^{\ell\ell} + \boldsymbol{p}_\mathrm{T}^\mathrm{miss} + \sum \boldsymbol{p}_\mathrm{T}^j \qquad (5.1)$$

The first cut after pre-selection in the cut-based analysis requires $p_{\mathrm{T}}^{\mathrm{sum}} < 15\,\mathrm{GeV}$ to further suppress $t\bar{t}$ production.

In the different flavor channels, a cut is made to reduce the contamination from $Z \to \tau\tau$ decays. The di-$\tau$ invariant mass, $m_{\tau\tau}$, is constructed by assuming that the neutrinos from the $\tau$ decays were collinear with the leptons[3] . The analysis requires that this mass not be consistent with a $Z$ by requiring $m_{\tau\tau} < m_Z - 25$ GeV.

## VBF TOPOLOGICAL CUTS

After the cuts mentioned above, the analysis applies specific cuts to take advantage of the unique VBF topology, collectively referred to as the "VBF topological cuts". First, a requirement on the dijet invariant mass of the VBF jets, $m_{jj}$, is placed, requiring $m_{jj} > 600\,\mathrm{GeV}$. Next, the event is required to have a large gap in rapidity between the two VBF jets, or $\Delta y_{jj} > 3.6$. Both of these cuts put tight requirements on the presence of two forward, high $p_{\mathrm{T}}$ jets moving in opposite directions in the longitudinal plane.

Beyond requiring the presence of the two forward VBF jets, the analysis also vetoes on the presence of any additional jets that fall between the two VBF jets. This cut is referred to as the central jet veto, or CJV. Any events with a third jet with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ whose rapidity is between the region defined by the two VBF jets are vetoed. This can be expressed in terms of a variable called the jet centrality, defined in equation 5.2.

$$C_{j3} = \left| \eta_{j3} - \frac{\eta_{j1} + \eta_{j2}}{2} \right| \bigg/ \frac{|\eta_{j1} - \eta_{j2}|}{2}, \tag{5.2}$$

Here, $\eta_{j1}$ and $\eta_{j2}$ are the pseudorapidities of the leading and subleading jets, respectively, while $\eta_{j3}$ is the pseudorapidity of the extra jet in the event (if one exists). Intuitively, $C_{j3}$ is zero when $\eta_{j3}$ is directly centered between the two jets and unity when $\eta_{j3}$ is aligned with either of the VBF jets. Thus, the CJV can be expressed as a requirement that $C_{j3} > 1$.

The decay products of the Higgs tend to be central as well. Thus, the analysis also requires that both leptons in the analysis fall within the rapidity gap defined by the jets. This cut is referred to as the outside
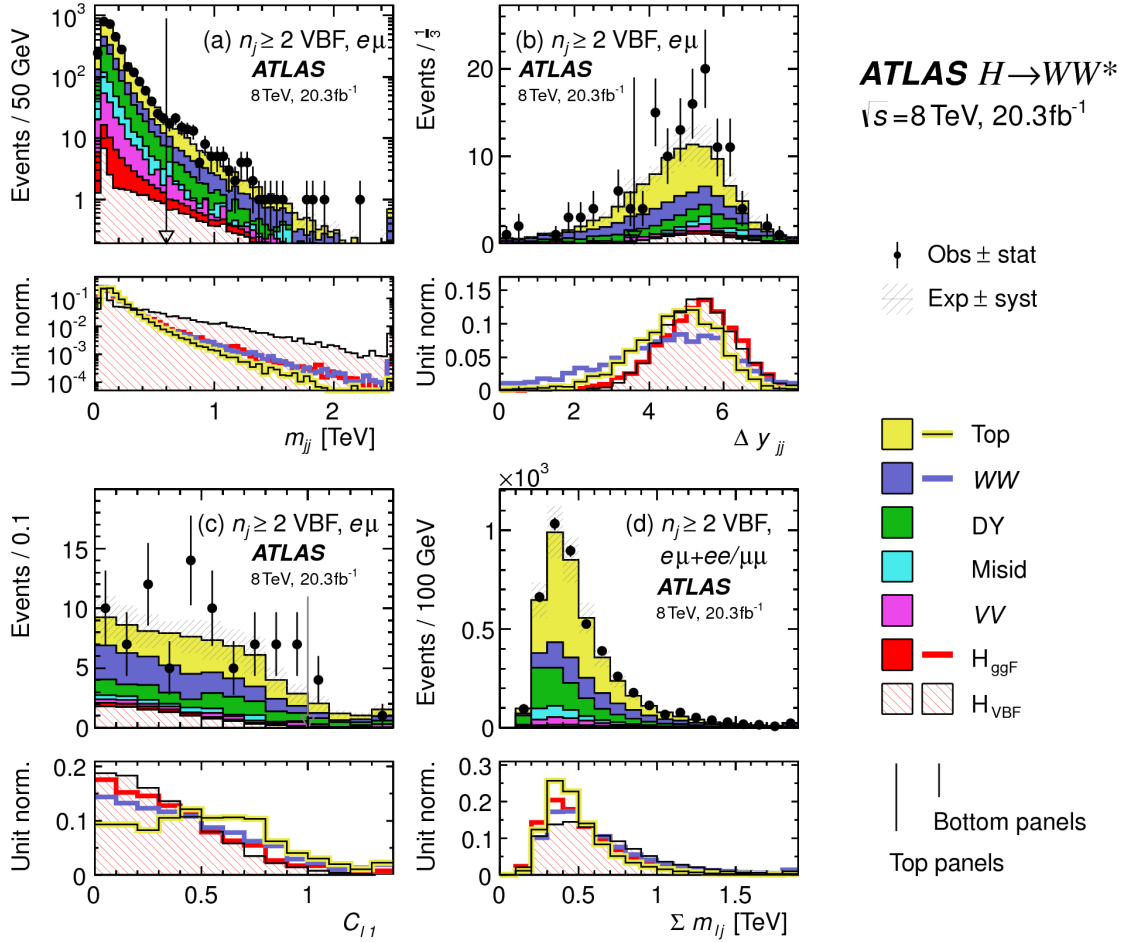
**Figure 5.2:** Distributions of (a) $m_{jj}$, (b) $\Delta y_{jj}$, (c) $C_{\ell 1}$, and (d) $\sum m_{\ell j}$, for the VBF analysis. The top panels compare simulation and data, while the bottom panels show normalized distributions for all background processes and signal [9].

lepton veto, or OLV. A quantitative way to define the cut is to require that the centrality of each lepton (defined analogously to that of the third jet in equation 5.2) correspond to the lepton being within the jet rapidity gap, or $C_\ell < 1$ for both leptons.

Figure 5.2a-c shows the $m_{jj}$, $\Delta y_{jj}$, and $C_{\ell 1}$ variables at the stage where all previous cuts in the sequence have been made. The agreement between data and Monte Carlo is good, and the bottom panels show their power in discriminating the VBF signal from the background processes.

The final signal region is also split into two bins of $m_{jj}$, with the first bin corresponding to $600\,\mathrm{GeV} < m_{jj} < 1\,\mathrm{TeV}$ and the second bin corresponding to $m_{jj} > 1\,\mathrm{TeV}$. The first bin has more statistics but also a larger contribution from background, while the second bin has lower statistics but a 1:1 signal to

background ratio.

## Higgs topological cuts

The analysis also places additional requirements on the final state leptons, as discussed in general in section 3.2. The leptons coming from the Higgs decay should be close in the transverse plane due to the spin-0 nature of the Higgs. As a result, two requirements on dilepton kinematics are made that are common with lower multiplicity jet bins as well. The angle between leptons in the transverse plane, $\Delta\phi_{\ell\ell}$, is required to be less than 1.8 radians. Additionally, the dilepton mass $m_{\ell\ell}$ is required to be less than 50 GeV.

The cut-based analysis uses $m_\text{T}$ as the final discriminating variable as in the ggF focused analysis. The optimal number of bins in $m_\text{T}$ was found to be three bins, with the bin boundaries at 80 and 130 GeV.

Table 5.7 shows the data and estimated signal and background yields from simulation as each cut described above is made. The table shows how each cut reduces specific backgrounds and how the overall signal to background ratio grows through the cutflow.

Figure 5.3 shows an ATLAS event display of a candidate event in the final signal region.

### 5.5.3   BDT based selection

The boosted decision tree based analysis takes a different philosophy compared to the cut-based. Rather than cutting sequentially on many variables, the BDT analysis uses many of these variables as inputs to the BDT and the output BDT score as the final discriminant. The BDT is trained with the VBF $H \rightarrow WW^*$ simulation as the signal samples and all other processes as background, including ggF $H \rightarrow WW^*$ production.

### Pre-training selection and BDT inputs

Before training, the common preselection cuts described in section 5.5.1 are applied. Additionally, the central jet veto and outside lepton veto described in section 5.5.2 are applied. The BDT has eight input variables, six of which are also variables that are used in the cut-based analysis. The six shared variables are $p_\text{T}^\text{sum}$, $m_{jj}$, $\Delta y_{jj}$, $m_{\ell\ell}$, $\Delta\phi_{\ell\ell}$, and $m_\text{T}$. The seventh variable input in the BDT is a combination of

**Table 5.7:** Event selection for the $n_j \geq 2$ VBF analysis in the 8 TeV cut-based analysis [9].

| | Summary | | | | | | Composition of $N_{bkg}$ | | | | | | | | | |
| Selection | $N_{obs}/N_{bkg}$ | $N_{obs}$ | $N_{bkg}$ | $N_{signal}$ $N_{ggF}$ | $N_{VBF}$ | $N_{VH}$ | $N_{WW}$ $N_{WW}^{QCD}$ | $N_{WW}^{EW}$ | $N_{top}$ $N_{t\bar{t}}$ | $N_t$ | $N_{misid}$ $N_{Wj}$ | $N_{jj}$ | $N_{VV}$ | $N_{Drell\text{-}Yan}$ $N_{ee/\mu\mu}$ | $N_{\tau\tau}^{QCD}$ | $N_{\tau\tau}^{EW}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $e\mu$ sample | $1.00 \pm 0.00$ | 61434 | 61180 | 85 | 32 | 26 | 1350 | 68 | 51810 | 2970 | 847 | 308 | 380 | 51 | 3260 | 46 |
| $n_b = 0$ | $1.02 \pm 0.01$ | 7818 | 7700 | 63 | 26 | 16 | 993 | 43 | 3000 | 367 | 313 | 193 | 273 | 35 | 2400 | 29 |
| $p_T^{sum} < 15$ | $1.03 \pm 0.01$ | 5787 | 5630 | 46 | 23 | 13 | 781 | 38 | 1910 | 270 | 216 | 107 | 201 | 27 | 2010 | 23 |
| $m_{\tau\tau} < m_Z - 25$ | $1.05 \pm 0.02$ | 3129 | 2970 | 40 | 20 | 9.9 | 484 | 22 | 1270 | 177 | 141 | 66 | 132 | 7.6 | 627 | 5.8 |
| $m_{jj} > 600$ | $1.31 \pm 0.12$ | 131 | 100 | 2.3 | 8.2 | — | 18 | 8.9 | 40 | 5.3 | 1.8 | 2.4 | 5.1 | 0.1 | 15 | 1.0 |
| $\Delta y_{jj} > 3.6$ | $1.33 \pm 0.13$ | 107 | 80 | 2.1 | 7.9 | — | 11.7 | 6.9 | 35 | 5.0 | 1.6 | 2.3 | 3.3 | — | 11.6 | 0.8 |
| $C_{j3} > 1$ | $1.36 \pm 0.18$ | 58 | 43 | 1.3 | 6.6 | — | 6.9 | 5.6 | 14 | 3.0 | 1.3 | 1.3 | 2.0 | — | 6.8 | 0.6 |
| $C_{\ell 1} < 1, C_{\ell 2} < 1$ | $1.42 \pm 0.20$ | 51 | 36 | 1.2 | 6.4 | — | 5.9 | 5.2 | 10.8 | 2.5 | 1.3 | 1.3 | 1.6 | — | 5.7 | 0.6 |
| $m_{\ell\ell}, \Delta\phi_{\ell\ell}, m_T$ | $2.53 \pm 0.71$ | 14 | 5.5 | 0.8 | 4.7 | — | 1.0 | 0.5 | 1.1 | 0.3 | 0.3 | 0.3 | 0.6 | — | 0.5 | 0.2 |
| $ee/\mu\mu$ sample | $0.99 \pm 0.01$ | 26949 | 27190 | 31 | 14 | 10.1 | 594 | 37 | 23440 | 1320 | 230 | 8.6 | 137 | 690 | 679 | 16 |
| $n_b, p_T^{sum}, m_{\tau\tau}$ | $1.03 \pm 0.03$ | 1344 | 1310 | 13 | 8.0 | 4.0 | 229 | 12.0 | 633 | 86 | 26 | 0.9 | 45 | 187 | 76 | 1.5 |
| $m_{jj}, \Delta y_{jj}, C_{j3}, C_\ell$ | $1.39 \pm 0.28$ | 26 | 19 | 0.4 | 2.9 | 0.0 | 3.1 | 3.1 | 5.5 | 1.0 | 0.2 | 0.0 | 0.7 | 3.8 | 0.7 | 0.1 |
| $m_{\ell\ell}, \Delta\phi_{\ell\ell}, m_T$ | $1.63 \pm 0.69$ | 6 | 3.7 | 0.3 | 2.2 | 0.0 | 0.4 | 0.2 | 0.6 | 0.2 | 0.2 | 0.0 | 0.1 | 1.5 | 0.3 | 0.1 |
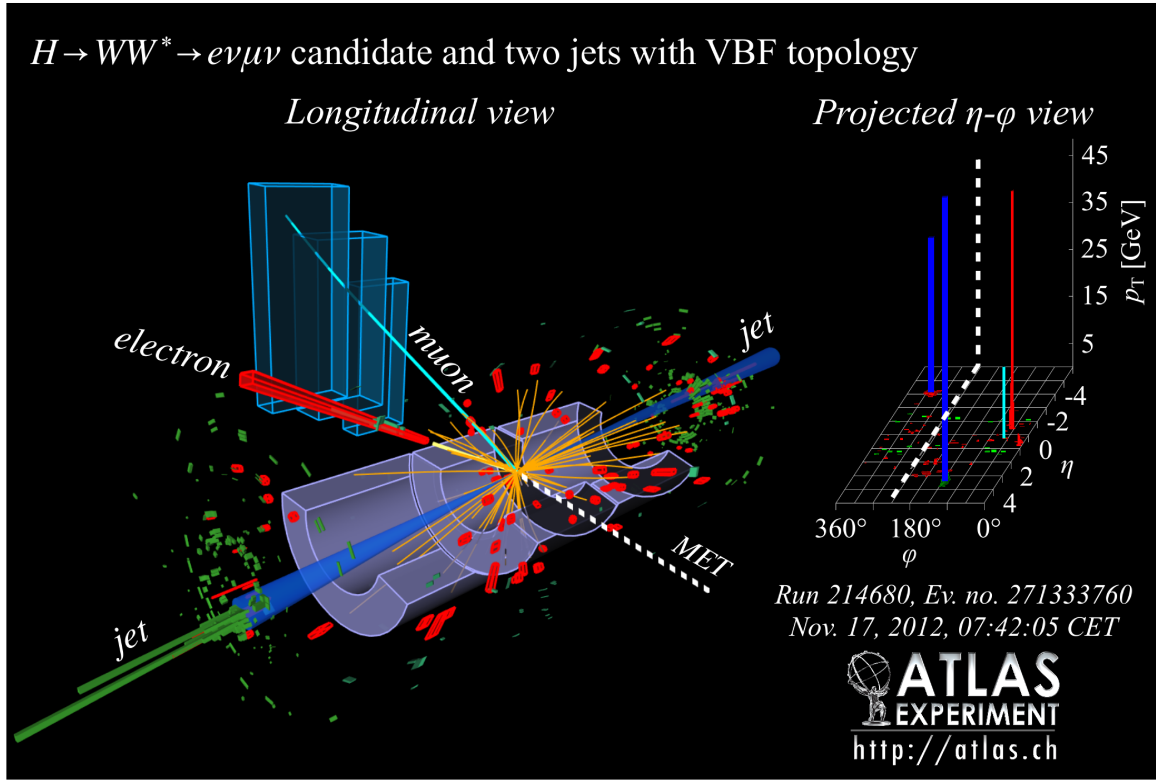
**Figure 5.3:** Event display of a VBF candidate event[9].

the variables used to do the OLV in the cut-based analysis. The BDT uses as input the sum of lepton centralities, or $\sum C_\ell = C_{\ell 1} + C_{\ell 2}$. The final BDT input variable, $\Sigma\, m_{\ell j}$, is constructed to account for the correlations between the jets and leptons in the event. It is the sum of the invariant masses of all four possible lepton-jet combinations, shown in equation 5.3.

$$\Sigma\, m_{\ell j} = m_{\ell 1, j1} + m_{\ell 1, j2} + m_{\ell 2, j1} + m_{\ell 2, j2} \tag{5.3}$$

Figure 5.2d shows the agreement between data and simulation for the $\Sigma\, m_{\ell j}$ variable, as well as showing its discriminating power. Figure 5.4 shows the distributions of the Higgs topological variables used as BDT inputs for VBF signal and corresponding backgrounds. Figure 5.5 shows the distributions of the VBF topological variables used as BDT inputs. In both cases, the VBF yield has been scaled by a factor of 50 to better show the shape difference compared to the backgrounds.

Table ?? summarizes the cuts applied for the BDT and cut-based analyses, as well as which variables

are used as input to the BDT.
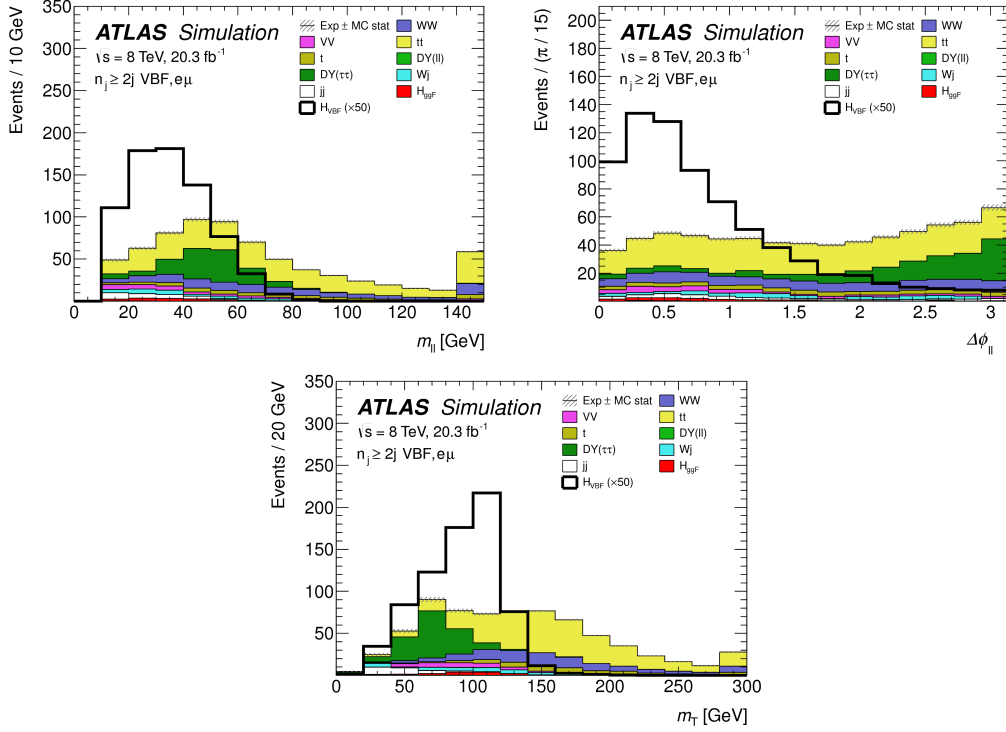


**Figure 5.4:** Distributions of $m_{\ell\ell}$ (top left), $\Delta\phi_{\ell\ell}$ (top right), and $m_{\mathrm{T}}$ (bottom) for the VBF BDT analysis. These are plotted after all of the BDT pre-training selection cuts [9].

## BDT output

After training, the BDT outputs a score ($O_{\mathrm{BDT}}$) which is in the range $[-1, 1]$, where $-1$ corresponds to background-like events and $+1$ corresponds to signal-like events. Figure 5.6 shows the output BDT distribution in both the different flavor and same flavor channels. For the final discriminant analysis, the $O_{\mathrm{BDT}}$ distribution is divided into four bins, with boundaries at $[-1, -0.48, -0.3, 0.78, 1]$. The bins are numbered from $0, 1, 2, 3$ respectively. Because bin 0 is predominantly background, it is excluded from the likelihood analysis.
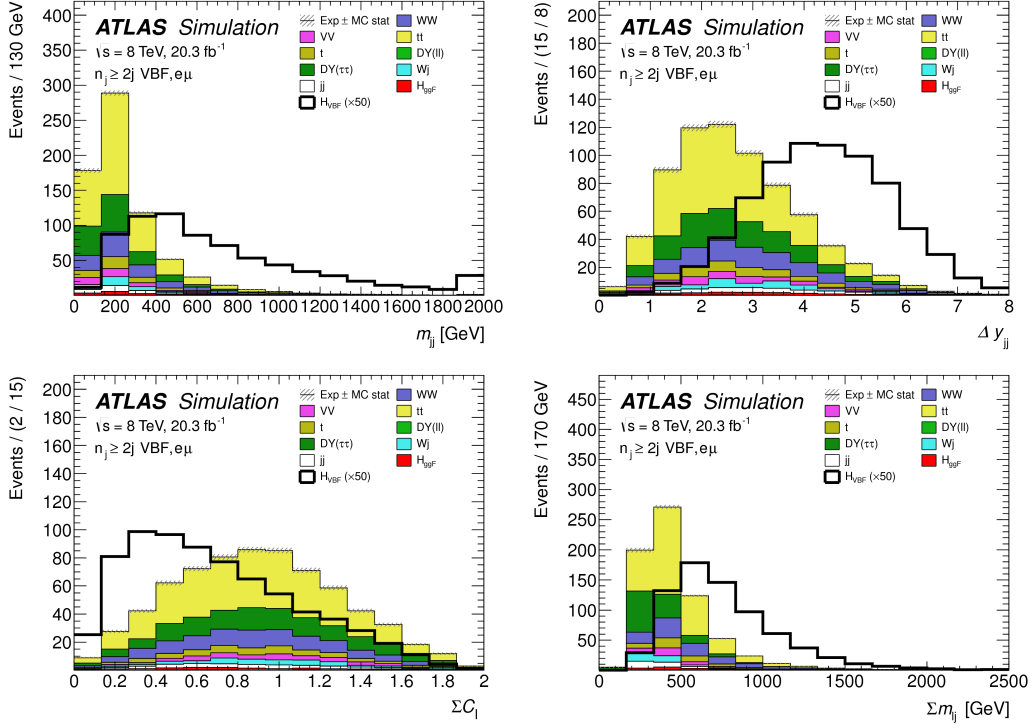
**Figure 5.5:** Distributions of $m_{jj}$ (top left), $\Delta y_{jj}$ (top right), $\sum C_\ell$ (bottom left), and $\sum m_{\ell j}$ (bottom right) for the VBF BDT analysis. These are plotted after all of the BDT pre-training selection cuts[9].

## 5.6 Background estimation

This section describes the procedures used to estimate backgrounds for the VBF analysis in both the cut-based and BDT analyses.

### 5.6.1 General strategy

Most of the backgrounds in the VBf analysis have shapes estimated from Monte Carlo simulation but normalizations derived from control regions in data. In essence, a normalization factor (denoted with $\beta$ or abbreviated as NF) is derived by scaling the MC yield in the control region to the corresponding yield in data. Once this factor is derived, it can be used to scale the MC estimate of the background in the signal region. This is illustrated in equation 5.4.

$$B_{\mathrm{SR}}^{\mathrm{est}} = B_{\mathrm{SR}} \times \frac{N_{\mathrm{CR}}}{B_{\mathrm{CR}}} \equiv B_{\mathrm{SR}} \times \beta \tag{5.4}$$
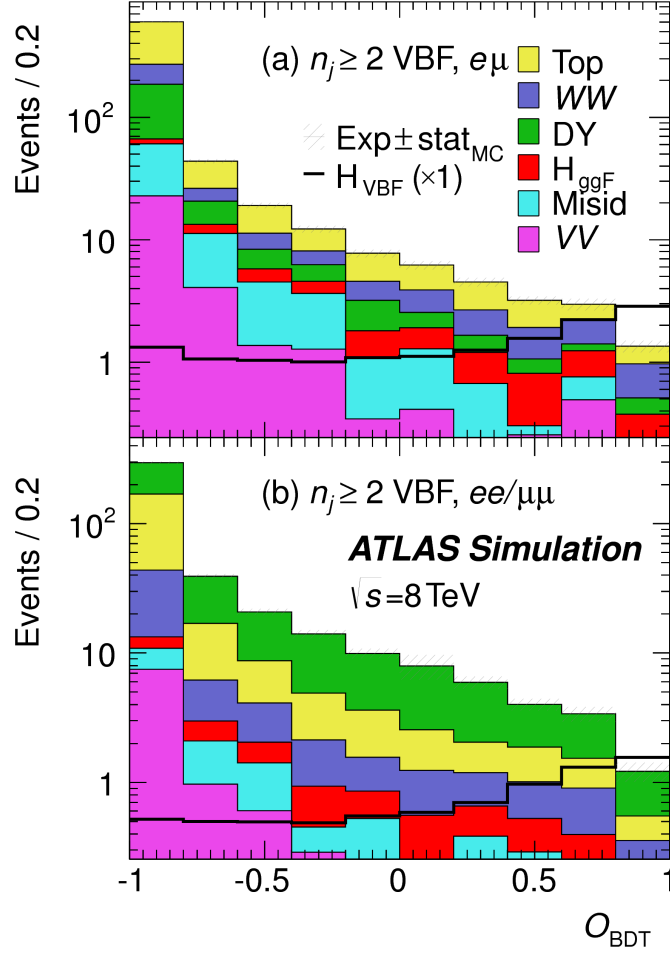
**Figure 5.6:** Distributions of $O_{\mathrm{BDT}}$ for the VBF signal and associated backgrounds after the VBF pre-training selection[9].

Here, $B$ denotes the MC yield prediction in the denoted region, while $N$ denotes the observed number of events in data in the denoted region.

Another way of writing the same equation, in terms of an extrapolation factor $\alpha$ rather than a normalization factor $\beta$. The overall calculation is exactly the same. However, when phrased in this way, it shows hwo the uncertainty on the background estimation can be reduced. This is shown in equation 5.5.

$$B_{\mathrm{SR}}^{\mathrm{est}} = N_{\mathrm{CR}} \times \frac{B_{\mathrm{SR}}}{B_{\mathrm{CR}}} \equiv N_{\mathrm{CR}} \times \alpha \qquad (5.5)$$

Phrased this way, the equation shows that with enough statistics in the control region, a large theoretical uncertainty on the overall background yield in the signal region can be replaced by a small statistical

uncertainty coming from the number of data events in the CR and a smaller theoretical uncertainty on the extrapolation from the control region to the signal region.

## 5.6.2 Top background

The normalization factor $\beta_t$ for the top background in the VBF analysis is derived in a region required to have one b-tagged jet, or $n_b = 1$. In the cut-based analysis, normalization factors are computed at every stage of the cutflow by applying the appropriate cuts in the CR. These NF are then applied to the $t\bar{t}$ and single top event yields in the SR. In the BDT analysis, a single normalization factor is computed for each bin of $O_{\mathrm{BDT}}$ after applying the BDT pre-training cuts described previously. The computed normalization factors are derived with all flavor combinations combined in order to decrease statistical uncertainty. Additionally, in the BDT analysis, BDT bins 2 and 3 are merged for the same reason.

Table 5.8 shows the evolution of the $\beta_t$ through the cut-based selection. Table 5.9 shows the value of the $\beta_t$ in each bin of $O_{\mathrm{BDT}}$. In all cases, the computed factors are relatively consistent with unity, with the largest discrepancy coming in bin 1 of $O_{\mathrm{BDT}}$.

| Cut | $\beta_t$ |
|:---:|:---:|
| $p_{\mathrm{T}}^{\mathrm{sum}} < 15\,\mathrm{GeV}$ | $1.03 \pm 0.01$ |
| $m_{\tau\tau} < m_Z - 25$ | $1.05 \pm 0.01$ |
| $m_{jj} > 600\,\mathrm{GeV}$ | $0.96 \pm 0.06$ |
| $\Delta y_{jj} > 3.6$ | $1.02 \pm 0.08$ |
| CJV | $1.13 \pm 0.16$ |
| OLV | $1.01 \pm 0.19$ |
| $m_{jj} < 1\,\mathrm{TeV}$ | $0.94 \pm 0.19$ |
| $m_{jj} > 1\,\mathrm{TeV}$ | $1.48 \pm 0.66$ |

**Table 5.8:** Top normalization factors computed at each stage of the cut-based selection. Uncertainties are statistical only.

| $O_{\mathrm{BDT}}$ | $\beta_t$ |
|:---:|:---:|
| Bin0 | $1.09 \pm 0.02$ |
| Bin1 | $1.58 \pm 0.15$ |
| Bin2 | $0.95 \pm 0.31$ |
| Bin3 | $0.95 \pm 0.31$ |

**Table 5.9:** Top normalization factors computed for each bin of $O_{\mathrm{BDT}}$. Uncertainties are statistical only.

Figure 5.7 shows the $m_{jj}$ and $O_{\mathrm{BDT}}$ distributions in the top control region. Overall the modeling looks consistent with the data.



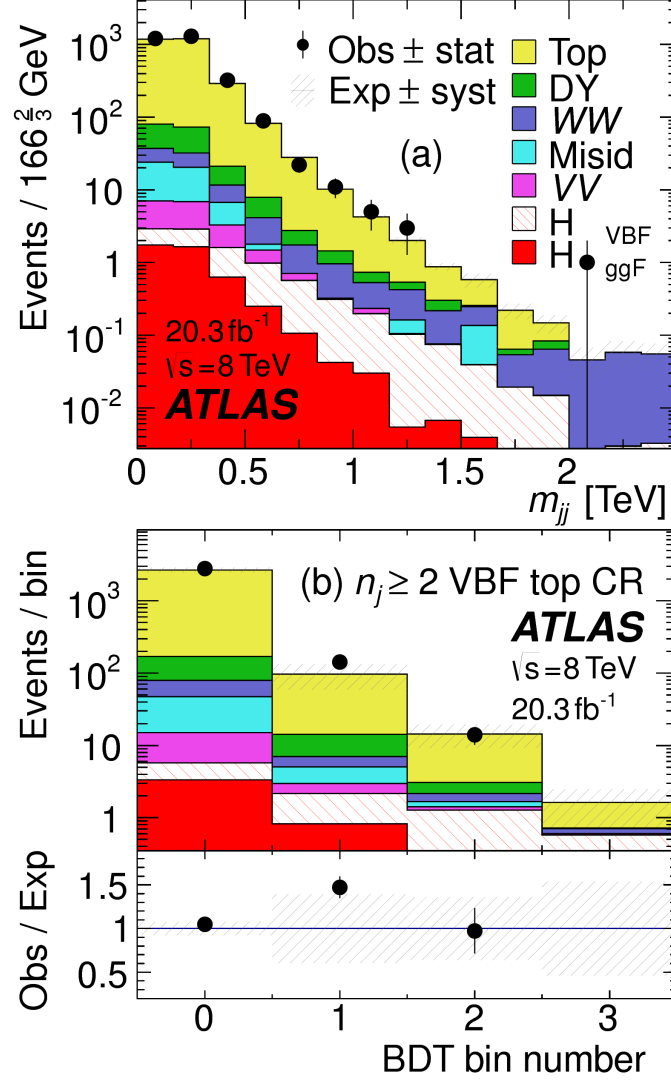**Figure 5.7:** Distributions of $m_{jj}$ (a) and $O_{\mathrm{BDT}}$ (b) in the VBF $n_b = 1$ top CR [9].

While these normalization factors can be computed and applied to the expected background yields listed in tables like table 5.7, in the end the normalization of the top background is profiled (meaning there is a dedicated Poisson constraint) and allowed to float in the final statistical fit.

### 5.6.3   $Z/\gamma^* \to \tau\tau$ BACKGROUND

In the different flavor channels, the $Z/\gamma^* \to \tau\tau$ background is an important one. Di-tau production can produce an $e\mu$ final state if each $\tau$ lepton decays to a different flavor lepton.

In the BDT analysis, a single normalization factor for the background is derived. A control region is defined using the pre-training selection cuts, except requiring that $|m_{\tau\tau} - m_Z| < 25\,\mathrm{GeV}$ so that the region is enriched in $Z/\gamma^* \to \tau\tau$ background. Additional requirements of $m_{\ell\ell} < 80(75)\,\mathrm{GeV}$ in the different (same) flavor channel, as well as $O_{\mathrm{BDT}} > -0.48$ are applied to increase the purity of the region. The final $\beta_{Z/\gamma^* \to \tau\tau}$ is calculated to be $0.9 \pm 0.3$ (statistical uncertainty only). Because of the small contribution of this background in the BDT analysis and the large statistical uncertainty, no additional systematics are calculated. The final SR estimate is scaled by this $\beta$ and not allowed to float in the fit.

The cut-based corrections are a bit more involved because they need to be applied cut by cut through the cutflow, as well as in the final signal region for the fit. The region is defined including all SR cuts up to the $Z/\gamma^* \to \tau\tau$ veto, which is instead made into a Z mass peak requirement as for the BDT region. The $m_{\ell\ell}$ cut from the BDT region is included as well. The cut-based approach aims to correct the normalization of the $Z/\gamma^* \to \tau\tau$ background in two ways. First, an overall normalization factor is computed from the control region. However, the VBF topological cuts are not included in this region, and applying them as is done in the top CR is not feasible due to limited statistics. So, instead, correction factors (CF) to the cut efficiencies of the VBF cuts are derived in a same flavor $Z \to \ell\ell$ control region, which has significantly more statistics. The CF is simply the ratio of the cut efficiencies in data and MC derived in this region. In the end, the overall background estimate is given by equation 5.6.

$$N_{Z/\gamma^* \to \tau\tau}^{\mathrm{est}} = B_{Z/\gamma^* \to \tau\tau}^{\mathrm{SR}} \times \beta_{\tau\tau} \times \frac{\epsilon_{\mathrm{VBF\,cuts}}^{\mathrm{data}}}{\epsilon_{\mathrm{VBF\,cuts}}^{\mathrm{MC}}} \tag{5.6}$$

The hypothesis is that while the normalization correction must be derived in a dedicated region, the efficiency of the VBF cuts should not be sensitive to the type of $Z/\gamma^*$ process and thus the larger control region can be exploited to derive the CF. Figure 5.8 shows a shape comparison for the $m_{jj}$ variable in

$Z \to \tau\tau$ events in the signal region and $Z \to \ell\ell$ events in the control region. The figure shows that the shapes are indeed comparable and thus any CF derived in the same flavor control region can reliably be applied in the signal region.
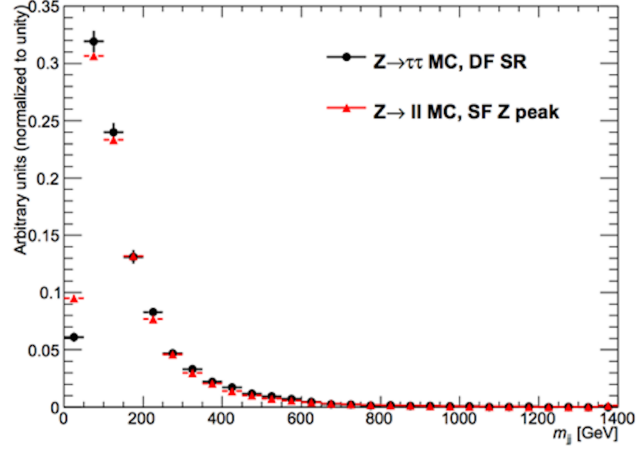


**Figure 5.8:** Comparison of $m_{jj}$ shape in a same flavor $Z \to \ell\ell$ control region and the VBF cut-based signal region.

Table 5.10 shows the overall normalization factor $\beta_{\tau\tau}$ and the efficiency correction factors for the various VBF topological cuts. In general, the statistical uncertainties on the cut efficiency corrections are quite good, and the MC tends to underestimate the efficiency of the VBF cuts for the $Z/\gamma^* \to \tau\tau$ background.

| $\beta_{\tau\tau}$ | $0.97 \pm 0.04$ |
|:---:|:---:|
| Cut | Correction factors |
| $m_{jj} > 600\,\mathrm{GeV}$ | $1.09 \pm 0.01$ |
| $\Delta y_{jj} > 3.6$ | $1.14 \pm 0.02$ |
| CJV | $1.20 \pm 0.02$ |
| OLV | $1.17 \pm 0.03$ |
| $m_{jj} < 1\,\mathrm{TeV}$ | $1.17 \pm 0.06$ |
| $m_{jj} > 1\,\mathrm{TeV}$ | $1.18 \pm 0.13$ |

**Table 5.10:** $Z/\gamma^* \to \tau\tau$ correction factors for the VBF cut-based analysis. Uncertainties are statistical only.

### 5.6.4 $Z/\gamma^* \to \ell\ell$ background

In the same flavor channels, the $Z/\gamma^* \to \ell\ell$ background is dominant and thus must be estimated correctly. In both the BDT and cut-based analyses, the background is estimated using the so-called "ABCD" method. The ABCD method creates four different regions by defining cuts on two variables. One of the regions (A) is the signal region, while the other regions are defined by inverting one of both of the cuts. in this case, the two variables used are $m_{\ell\ell}$ and $E_{\mathrm{T}}^{\mathrm{miss}}$, because inverting either of the SR cuts on these variables will give regions rich in the $Z/\gamma^* \to \ell\ell$ background. Figure 5.9 illustrates the general strategy for each region.

| **Region A (SR)** | **Region C** |
|---|---|
| High $E_T^{\mathrm{miss}}$ | High $E_T^{\mathrm{miss}}$ |
| Low $m_{\ell\ell}$ | Z peak |
| **Region B** | **Region D** |
| Low $E_T^{\mathrm{miss}}$ | Low $E_T^{\mathrm{miss}}$ |
| Low $m_{\ell\ell}$ | Z peak |

**Figure 5.9:** General illustration of the ABCD region definitions for $Z/\gamma^* \to \ell\ell$ background estimation.

In both of the BDT and cut-based analyses, the Z peak region is defined with $|m_{\ell\ell} - m_Z| < 15\,\mathrm{GeV}$. In the BDT analysis, low $m_{\ell\ell}$ corresponds to $m_{\ell\ell} < 75\,\mathrm{GeV}$ while in the cut-based it is $m_{\ell\ell} < 50\,\mathrm{GeV}$ (as this defines the cut-based SR). The BDT low $E_{\mathrm{T}}^{\mathrm{miss}}$ region is between 25 and 45 GeV, while the high $E_{\mathrm{T}}^{\mathrm{miss}}$ region is $E_{\mathrm{T}}^{\mathrm{miss}} > 45\,\mathrm{GeV}$. In the cut-based, high and low $E_{\mathrm{T}}^{\mathrm{miss}}$ are defined as opposite ends of the 55 GeV cut applied for the signal region definition.

Once the regions are defined, the final signal region background estimate is done by taking the estimate in region B and extrapolating it to the signal region (A) by multiplying it by the ratio of regions C and D. Effectively, the $Z$ peak region is used to estimate the efficiency of the $E_{\mathrm{T}}^{\mathrm{miss}}$ cut in data, and then this efficiency is applied in the low $m_{\ell\ell}$ region. An additional correction is also applied for the non-

closure of the method in MC. This is summarized in equations 5.7 and 5.8.

$$N_{Z/\gamma^*\to\ell\ell}^{\mathrm{SR}} = N_{Z/\gamma^*\to\ell\ell}^{\mathrm{B}} \times \frac{N_{Z/\gamma^*\to\ell\ell}^{\mathrm{C}}}{N_{Z/\gamma^*\to\ell\ell}^{\mathrm{D}}} \times f_{\mathrm{corr}} \tag{5.7}$$

$$f_{\mathrm{corr}} = \frac{B_{\mathrm{MC}}^{\mathrm{A}}/B_{\mathrm{MC}}^{\mathrm{B}}}{B_{\mathrm{MC}}^{\mathrm{C}}/B_{\mathrm{MC}}^{\mathrm{D}}} \tag{5.8}$$

Here, the $N$ refer to data yields in each region with the non $Z/\gamma^*$ backgrounds subtracted, while $B$ refer to the $Z/\gamma^*$ yields in MC in each region.

A normalization factor $\beta_{\ell\ell}$ is computed for each analysis as the ratio of the predicted data yield to the MC yield in the SR. The shape of the BDT distribution is taken from data region B, while the shape of the $m_{\mathrm{T}}$ distribution in the cut-based analysis is taken from $Z/\gamma^*$ MC in the SR. The values of the $\beta_{\ell\ell}$ in the BDT and cut-based analyses from this method are summarized in table 5.11. In the cut-based analysis, the same cut efficiency correction factors shown in table 5.10 are also applied (in product with the $\beta_{\ell\ell}$) in the same flavor channels to this background, as they were derived in the $Z$ peak region.

|  | $\beta_t$ |
|---|---|
| BDT Bin 1 | $1.01 \pm 0.15$ |
| BDT Bin 2 | $0.89 \pm 0.28$ |
| Cut-based | $0.81 \pm 0.21$ |

**Table 5.11:** $Z/\gamma^* \to \ell\ell$ normalization factors for cut-based and BDT analyses. Uncertainties are statistical only.

## 5.7 Systematic uncertainties

## 5.8 Results

# 6

# Combined Run 1 $H \to WW^* \to \ell\nu\ell\nu$ results

# Part III

# Search for Higgs pair production in the $HH \rightarrow b\bar{b}b\bar{b}$ channel in LHC Run 2 at $\sqrt{s} = 13$ TeV

# 7

# Search overview

# 8

# Search for Higgs pair production in boosted

# final states

# 9

# Results with Run 2 2015 dataset

# Part IV

# Looking ahead

# 10
## Conclusion

We found the Higgs. Then measured it. Then used it to look for new physics. What a time to be alive!

# References

[1] (2012). *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung.* Technical Report ATLAS-CONF-2012-047, CERN, Geneva.

[2] (2014). *Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton-proton collision data.* Technical Report ATLAS-CONF-2014-032, CERN, Geneva.

[3] Aad, G. et al. (2014). Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton–proton collision data. *Eur. Phys. J.*, C74(11), 3130.

[4] Armbruster, A. J. (2013). Discovery of a higgs boson with the atlas detector. CERN-THESIS-2013-047.

[5] B. P. Kersevan and E. Richter-Was (2004). The Monte Carlo event generator AcerMC version 2.0 with interfaces to PYTHIA 6.2 and HERWIG 6.5.

[6] Butterworth, J. M., Forshaw, J. R., & Seymour, M. H. (1996). Multiparton interactions in photo-production at HERA. *Z. Phys.*, C 72, 637.

[7] Collaboration, A. (2014). Performance of the ATLAS muon trigger in pp collisions at $\sqrt{s} = 8$ TeV. *Eur. Phys. J. C*, (arXiv:1408.3179. CERN-PH-EP-2014-154), 75. 19 p.

[8] collaboration, A. (2015). Electron trigger performance in 2012 ATLAS data. ATLAS-COM-DAQ-2015-091.

[9] Collaboration, A. (2015). Observation and measurement of higgs boson decays to ww* with the atlas detector. *Phys. Rev. D*, 92(012006).

[10] Corcella, G. et al. (2001). HERWIG 6: An event generator for hadron emission reactions with interfering gluons (including super-symmetric processes) . *JHEP*, 01, 010.

[11] G. Cowan, K. Cranmer, E. Gross, and O. Vitells (2011). Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J.*, C 71, 1554.

[12] Gao, J., Guzzi, M., Huston, J., Lai, H.-L., Li, Z., et al. (2014). The CT10 NNLO Global Analysis of QCD. *Phys.Rev.*, D89, 033009.

[13] Gleisberg, T., Hoeche, S., Krauss, F., Schonherr, M., Schumann, S., et al. (2009). Event generation with SHERPA 1.1. *JHEP*, 0902, 007.

[14] Kauer, N. & Passarino, G. (2012). Inadequacy of zero-width approximation for a light Higgs boson signal.

[15] Mangano, M. L. et al. (2003). ALPGEN, a generator for hard multiparton processes in hadronic collisions. *JHEP*, 0307, 001.

[16] Nason, P. (2004). A new method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11, 040.

[17] P. M. Nadolsky (2008). Implications of CTEQ global analysis for collider observables. *Phys. Rev.*, D 78, 013004.

[18] S. Agostinelli et al. (2003). Geant4, a simulation toolkit. *Nucl. Instrum. Meth.*, A 506, 250.

[19] Sherstnev, A. & Thorne, R. S. (2009). Parton distributions for the LHC. *Eur. Phys. J*, C 55, 553.

[20] Sjostrand, T., Mrenna, S., & Skands, P. Z. (2006). PYTHIA 6.4 Physics and Manual. *JHEP*, 0605, 026.

[21] Sjostrand, T., Mrenna, S., & Skands, P. Z. (2008). A Brief Introduction to PYTHIA 8.1. *Comput.Phys.Commun.*, 178, 852–867.

THIS THESIS WAS TYPESET using LaTeX, originally developed by Leslie Lamport and based on Donald Knuth's TeX. The body text is set in 11 point Egenolff-Berner Garamond, a revival of Claude Garamont's humanist typeface. The above illustration, *Science Experiment 02*, was created by Ben Schlitter and released under CC BY-NC-ND 3.0. A template that can be used to format a PhD dissertation with this look & feel has been released under the permissive AGPL license, and can be found online at github.com/asm-products/Dissertate or from its lead author, Jordan Suchow, at suchow@post.harvard.edu.