# Search for Semiboosted Higgs Pair Production From Vector Boson Fusion in the Single Lepton $b\bar{b}W^+W^-$ Final State Using the ATLAS Detector


Master Thesis at the Faculty of Physics
of
Ludwig-Maximilians-Universität München

submitted by
**Stefan Brückner**
born in Dachau

Munich, 11th May 2026

# Suche nach semiboosted Higgs Paarproduktion aus Vektorbosonfusion im Einzelleptonen-Endzustand $b\bar{b}W^+W^-$ mit dem ATLAS Detektor


Masterarbeit der Fakultät für Physik
der
Ludwig-Maximilians-Universität München

vorgelegt von
**Stefan Brückner**
geboren in Dachau

München, den 11.05.2026

# Abstract

The discovery of the Higgs boson, which is predicted by the Higgs mechanism, solved the problem of the generation of particle masses. However, many properties of the Higgs boson are still not well known and under ongoing investigations. One of these properties is the quartic coupling of a Higgs boson pair to a pair of electroweak gauge bosons ($g_{HHVV}$). A process involving this coupling is the di-Higgs production via vector boson fusion (VBF), which is the second most common production mode at LHC. The analysis in this thesis focuses on events in the $b\bar{b}W^+W^-$ final state with a single lepton, using a semiboosted approach. In this semiboosted scenario, one of the two objects, either the $H \to b\bar{b}$ or the W boson, which decays hadronically, is represented by a large radius jet.

The general kinematics of the VBF di-Higgs are studied for six different values of the coupling $g_{HHVV}$. In addition, the sensitivity of the semiboosted approach to possible deviations of the coupling to the Standard Model is tested. Based on that, an event selection strategy for the semiboosted approach for a non Standard Model coupling is derived and compared to $t\bar{t}$ events, which is the major background process.

# Contents

# Chapter 1

# Introduction

The ATLAS experiment is one of the major particle physics experiments located at the Large Hadron Collider (LHC) at CERN. The LHC collides bunches of up to about $10^{11}$ protons at very high energies, enabling detailed studies of the fundamental constituents of matter. A major achievement of the ATLAS experiment was the discovery of the Higgs boson in 2012, which confirmed the mechanism of electroweak symmetry breaking. The Higgs mechanism explains how particles aquire mass within the Standard Model, but key properties of the Higgs boson are still not exactly known. In particular, the exact shape of the Higgs potential, the trilinear Higgs self coupling and the coupling to vector bosons remain central topics of ongoing experimental investigations. Measuring these couplings is a key probe for the Higgs mechanism as predicted in the Standard Model, while also providing sensitivity to possible extensions beyond the Standard Model.

Vector Boson Fusion (VBF) di-Higgs production is of special interest, because it is directly sensitive to the Higgs self coupling and the quartic coupling between two vector bosons and two Higgs bosons ($g_{HHVV}$). The focus in this thesis is on investigating deviations of the quartic coupling $g_{HHVV}$ to the Standard Model in a semileptonic $b\bar{b}W^+W^-$ final state for the decay of the two Higgs bosons.

Chapters 2 and 3 will give a brief overview of the ATLAS detector and the relevant theoretical background. The truth studies in chapter 4 will investigate how different coupling strengths of the mentioned coupling $g_{HHVV}$ affect the kinematics of the process and if a semiboosted approach for an event selection would be sensitive to deviations of the coupling. Based on these studies, chapter 5 presents a possible event selection for the semiboosted approach for a non Standard Model value of the coupling. The event selection in this chapter is based on a semiboosted topology, in which one of the bosons ($H_{b\bar{b}}$ or $W_{had}$) is reconstructed as a large-radius jet, while the other is reconstructed as two resolved smallR jets. The performance of this selection, as well as the specific challenges of such an event selection are discussed in the end. Finally, chapter 6 gives a summary of the work presented in the thesis and provides an outlook of potential next steps and improvements for the selection.

# Chapter 2

# The ATLAS Detector

The ATLAS (A Toroidal LHC ApparatuS) detector is designed to study proton-proton collisions and heavy-ion collisions at the Large Hadron Collider (LHC) at CERN. This chapter provides a brief overview of the detector components and the coordinate system that is used to describe these collisons.

## 2.1 Overview of Detector Components

The ATLAS detector is 44 meters long and 25 meters high and has a overall mass of approximately 7000 tonnes. An overview of the detector and its components is given in Fig. 1.

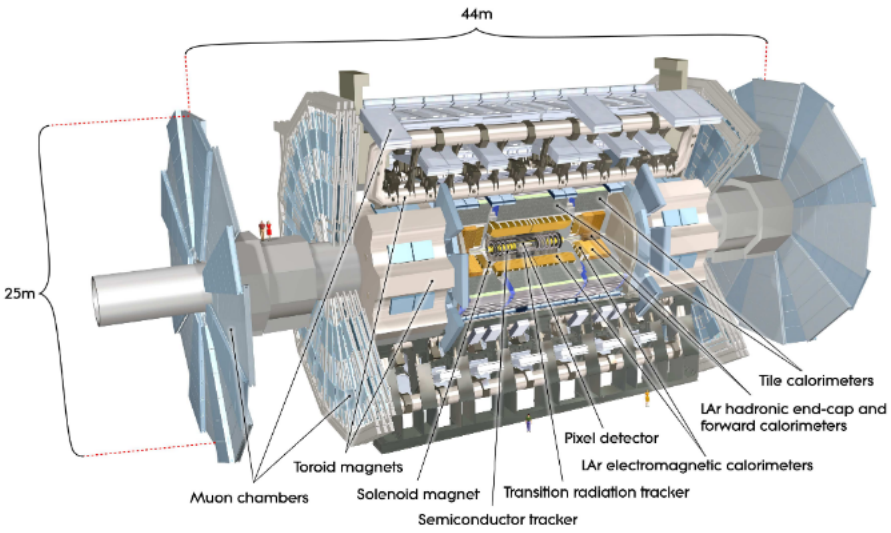



Figure 1: Overview of the ATLAS detector [3]

The inner detector consists of semiconductor pixel and strip detectors to measure momentum and vertices of particles with high resolution and straw tube tracking detectors to detect transition radiation. A strong axial magnetic field of 2 T is used in the inner detector to bend charged particle trajectories which enables momentum measurement. The liquid-argon (LAr) electromagnetic calorimeter measures electromagnetic particle showers in a pseudorapidity range of $|\eta| < 3.2$. The hadronic calorimeter consists of a scintilator tile calorimeter for $|\eta| < 1.7$ and a LAr calorimeter for $|\eta| > 1.5$ to match the pseudorapidity range of the

electromagnetic calorimeter. Additional LAr forward calorimeters for electromagnetic and hadronic measurements extend the coverage to $|\eta| = 4.9$. The muon spectrometer on the outer part of the detector consists of one barrel and two end-cap calorimeters which can measure charged particles in a pseudorapidity range of $|\eta| < 2.7$ and trigger on these particles in range of $|\eta| < 2.4$ [3] [5].

## 2.2 Coordinate System

The interaction point of the proton-proton collisions in the detector is defined as the origin of the coordinate system. The beamline is defined to be the z-axis and therefore the x-y-plane is perpendicular to the beam. The positive x-axis is defined to point to the centre of the accelarator ring and the positive y-axis is defined to point upwards.

The azimuthal angle $\phi$ is measured around the z-axis and the angle $\theta$ describes the angle to the beam axis. An important quantity to measure angular distances between objects is $\Delta R$ which is defined as $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ where the pseudorapidity is defined as $\eta = -ln \tan(\frac{\theta}{2})$ [3] [5].

# Chapter 3

# Theoretical Background

## 3.1 The Standard Model of Particle Physics

The Standard Model of particle physics (SM) describes the fundamental particles and their interactions through the electromagnetic, weak, and strong forces. Its particle content consists of fermions, which make up matter, and bosons, which mediate the fundamental interactions. Every fermion in the SM has a corresponding antiparticle with the same mass but opposite quantum numbers. An overview of the particles in the SM is given in Fig. 2.

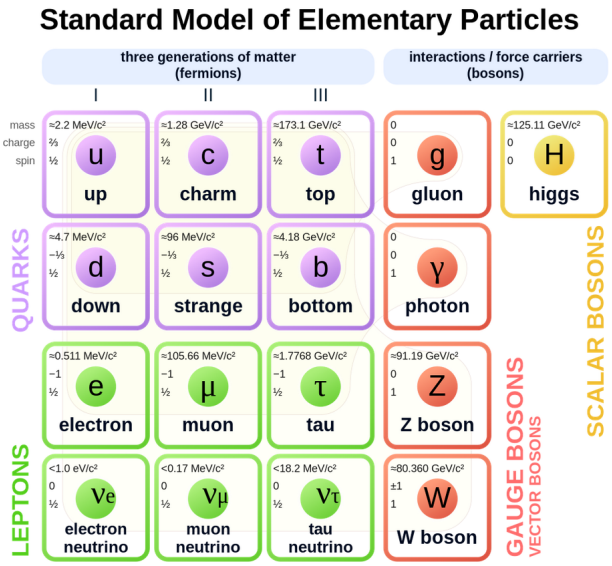



Figure 2: Elementary Particles of the Standard Model [9]

Fermions have spin 1/2 and can be categorized into quarks and leptons. Quarks and leptons can be further divided in 3 different generations, where the lightest and most stable particles make up the first generation and the heavier and more unstable particles belong to the second and third generations. Quarks have non integer electric charges and interact mostly via strong interaction which is described by Quantumchronodynamics (QCD). They also carry a color charge and hadronize in color neutral bound states. Leptons have integer electric charges and interact through the electromagnetic and weak force. Bosons have integer spin values and act as the force carriers of the fundamental forces. The massless and chargeless photon is the mediator of the electromagnetic force and couples to all charged particles. The W and Z bosons are the carriers of the electroweak interaction, which unifies the weak and

electromagnetic forces. The strong force is mediated by gluons and is responsible for the binding of quarks within hadrons. Gluons couple only to particles that carry color charge, which are quarks, antiquarks and other gluons. The Higgs boson has spin zero and arises from a scalar field, which gives mass to other particles through the Higgs mechanism, which will be explained in more detail in the next section.
In the SM the three fundamental forces mentioned are described by gauge theories, where each force corresponds to a certain gauge symmetry: U(1) for electromagnetism, SU(2) for the weak force, and SU(3) for the strong force. In the SM, gauge symmetry means that the Lagrangian remains invariant under local transformations of the fields.

Although the SM is very successful at describing the fundamental particles and their interactions there are still open problems that are not explained yet in the SM. Examples of such unsolved questions are the observed asymmetry between matter and antimatter or the description of gravity as a quantum field theory [9] [10].

## 3.2 Higgs Mechanism

While the Standard Model successfully describes particle interactions through gauge symmetries, the observed masses of 80.4 GeV for the W boson [14] and 91.2 GeV for the Z boson [12] require an additional mechanism. Directly adding mass terms to the SM Lagrangian for the gauge bosons would destroy the gauge symmetry since such terms are not invariant under local gauge transformations. The Higgs mechanism avoids this problem by generating masses after spontaneous symmetry breaking while keeping the gauge symmetry of the Lagrangian.

To achieve this, an additional term needs to be introduced in the Lagrangian:

$$\mathcal{L} = (D_\mu \Phi)^\dagger (D^\mu \Phi) - V(\Phi) \tag{3.1}$$

The Potential $V(\Phi)$ is given by:

$$V(\Phi) = -\mu^2 \Phi^\dagger \Phi + \frac{1}{4}\lambda^2 (\Phi^\dagger \Phi)^2 \tag{3.2}$$

$\mu$ and $\lambda$ are free parameters that need to be determined experimentally and $\Phi$ is a complex field: $\Phi = \Phi_1 + i\Phi_2$. The potential in equation 3.2 is shown in Fig. 3.

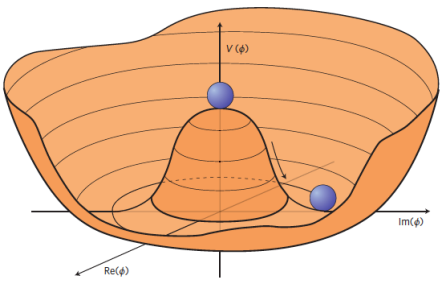



Figure 3: Illustration of the Higgs potential and spontaneous symmetry breaking. The symmetry is spontaneously broken by occupying a ground state [20].

The minima of the potential are on a cirlce of radius $\frac{\mu}{\lambda}$. Therefore to occupy a ground state the rotational symmetry of the system must be broken. This process is called spontaneous symmetry breaking. By defining two new fields

$$\eta = \phi_1 - \frac{\mu}{\lambda} \quad \xi = \phi_2 \tag{3.3}$$

and choosing a certain gauge transformation where $\xi = 0$, it can be shown that $\eta$ becomes a massive scalar, which correponds to the Higgs and a massive gauge field $A^\mu$, that explains the masses of the W and Z bosons [10]. The Higgs field is a scalar field that exists everywhere with a vacuum expectation value $\nu$ that is fixed by the Fermi coupling $G_F$.

$$\nu = (\sqrt{2}G_F)^{-1/2} \approx 246GeV \tag{3.4}$$

The vacuum expectation value can also be expressed by the parameters $\mu$ and $\lambda$ from the Higgs potential.

$$\nu = \sqrt{-\frac{\mu^2}{\lambda}} \tag{3.5}$$

In the SM the Higgs boson is a scalar particle with spin 0 and its mass is given by:

$$m_H = \sqrt{2\lambda}\nu \tag{3.6}$$

So in the SM the two free parameters $\mu$ and $\lambda$ are fixed by the vacuum expectation value and the mass of the Higgs boson. But it is important to note that the potential given in equation 3.2 is just the most general renormalizable potential that is invariant under local gauge transformations and allows for spontaneous symmetry breaking. The true shape of the Higgs potential is not known yet [1] [11]. The Higgs boson was first discovered in 2012 by the ATLAS [16] and CMS [19] collaborations. The experimentally measured mass up to date is $m_H = 125.20 \pm 0.11\ GeV$. Since the discovery of the Higgs boson the goal has been to determine its properties, especially its various couplings.

Measurements of the Higgs couplings are essential, since these couplings are directly connected to the mechanism of electroweak symmetry breaking. Any deviation from the predicted coupling strengths could therefore provide evidence that the Higgs potential differs from the Standard Model form and reveal signs of new physics.

## 3.3 Higgs Boson Couplings

As mentioned, the coupling strength of the Higgs boson coupling to other particles in the SM directly depends on their masses. For fermions the coupling strength is linearly proportional to the mass and for bosons the coupling strength is proportional to the square of the mass. This means that the coupling of the Higgs is weak for light particles like electrons or up and down quark, but strong for heavy particles like W and Z bosons. The couplings of the Higgs boson to fermions and gauge bosons aswell as the self couplings can be summarized in the following Lagrangian:

$$\mathcal{L} = -g_{Hf\bar{f}}\,\bar{f}f\,H + \frac{g_{HHH}}{6}\,H^3 + \frac{g_{HHHH}}{24}\,H^4 + \delta_V\,V_\mu V^\mu\left(g_{HVV}\,H + \frac{g_{HHVV}}{2}\,H^2\right) \tag{3.7}$$

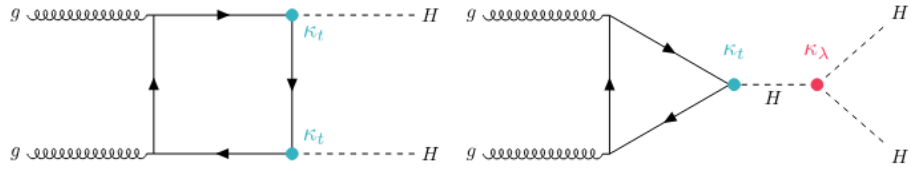
The specific couplings are given by:

$$g^{SM}_{Hf\bar{f}} = \frac{m_f}{v}, \quad g^{SM}_{HVV} = \frac{2m_V^2}{v}, \quad g^{SM}_{HHVV} = \frac{2m_V^2}{v^2}, \quad g^{SM}_{HHH} = \frac{3m_H^2}{v}, \quad g^{SM}_{HHHH} = \frac{3m_H^2}{v^2} \tag{3.8}$$

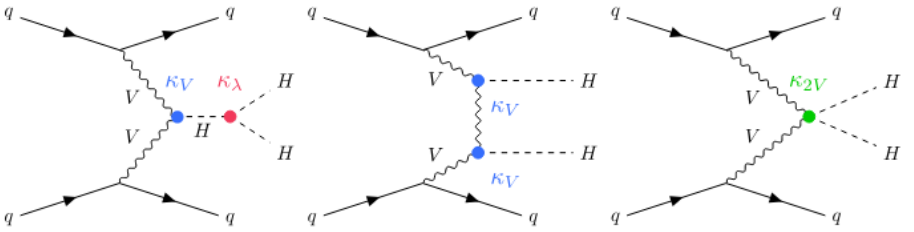
V equals to a $W^\pm$ or Z boson and $\delta_W = 1$ and $\delta_Z = 1/2$. Measuring and constraining these couplings experimentally is an important check of the Higgs mechanism and the shape of the Higgs potential in the SM. Deviations of these couplings to their predicted values would indicate physics beyond the SM [1].

## 3.4 Di-Higgs Boson Production

Experimental studies of the Higgs couplings require processes in which Higgs bosons are produced and the relevant interactions contribute directly. While single-Higgs production probes HVV couplings, access to the Higgs self-coupling and the quartic HHVV interaction requires processes involving Higgs boson pairs. At the LHC, the main di-Higgs production mode is gluon-gluon fusion (ggF) with a cross section of $30.8^{+2.0}_{-7.1} fb$, followed by vector-boson fusion (VBF) with a cross section of $1.69 \pm 0.05\ fb$ at a central mass energy of $\sqrt{s} = 13\ TeV$. For ATLAS Run3 at $\sqrt{s} = 13.6\ TeV$ the cross sections increase to $34.1^{+2.2}_{-7.9}\ fb$ for ggF and $1.87 \pm 0.05$ for VBF. The leading order Feynman diagrams for di-Higgs production are shown in Fig. 4.




(a) gluon gluon fusion




(b) vector boson fusion

Figure 4: Leading di-Higgs production Feynman diagrams of ggF and VBF

The coupling modifiers used in the diagrams are defined as the coupling strength divided by their predicted value from the SM.

$$\kappa_V = \frac{g_{HVV}}{g^{SM}_{HVV}} \quad \kappa_{2V} = \frac{g_{HHVV}}{g^{SM}_{HHVV}} \quad \kappa_\lambda = \frac{g_{HHH}}{g^{SM}_{HHH}} \tag{3.9}$$

The three VBF processes shown in Fig. 4b interfere destructively in the SM, which significantly reduces the cross section of the VBF di-Higgs process. This interference depends on the value of the coupling modifiers $\kappa_V$,$\kappa_{2V}$ and $\kappa_\lambda$. For example, a deviation of $\kappa_{2V}$ from its SM prediction weakens the destructive interference, leading to an increase in the cross section. The processes including $\kappa_V$ and $\kappa_{2V}$ have larger di-Higgs invariant masses compared to processes including $\kappa_\lambda$. Consequently, for non-SM values of $\kappa_{2V}$, an increased number of events is expected in the high-$m_{HH}$ region. In these events the Higgs bosons are produced with large transverse momenta their decay products tend to be collimated. These boosted topologies therefore provide good sensitivity to changes in $\kappa_{2V}$ which makes them particularly useful to probe $\kappa_{2V}$ experimentally. Constraining $\kappa_{2V}$ is a key test of the electroweak symmetry breaking mechanism in the SM and can also provide sensitivity to possible beyond the standard model scenarios [17] [7].

## 3.5 Semileptonic $b\bar{b}W^+W^-$ Final State of Di-Higgs Decay

To probe $\kappa_{2V}$ experimentally, specific di-Higgs decay channels must be considered. The branching ratios of the most common decays of the Higgs boson are shown in Fig. 5.

| Decay channel | Branching ratio | Rel. uncertainty |
|---|---|---|
| $H \to \gamma\gamma$ | $2.27 \times 10^{-3}$ | 2.1% |
| $H \to ZZ$ | $2.62 \times 10^{-2}$ | ±1.5% |
| $H \to W^+W^-$ | $2.14 \times 10^{-1}$ | ±1.5% |
| $H \to \tau^+\tau^-$ | $6.27 \times 10^{-2}$ | ±1.6% |
| $H \to b\bar{b}$ | $5.82 \times 10^{-1}$ | $^{+1.2\%}_{-1.3\%}$ |
| $H \to c\bar{c}$ | $2.89 \times 10^{-2}$ | $^{+5.5\%}_{-2.0\%}$ |
| $H \to Z\gamma$ | $1.53 \times 10^{-3}$ | ±5.8% |
| $H \to \mu^+\mu^-$ | $2.18 \times 10^{-4}$ | ±1.7% |

Figure 5: Most important branching ratios of a SM Higgs boson with relative uncertainties [1]

For the decay of two Higgs bosons there are many possible final states. The focus in this thesis is on a semiletonic $b\bar{b}W^+W^-$ final state in which one Higgs boson decays into a pair of b quark and antiquark ($H_{b\bar{b}}$) and the other Higgs boson decays into two W bosons. In addition one of the two W bosons decays hadronically ($W_{had}$) into a quark antiquark pair and the other W boson decays leptonically ($W_{lep}$) into a lepton and a antineutrino or a antilepton and a neutrino respectively. A diagram of the whole di-Higgs final state is shown in Fig. 6.

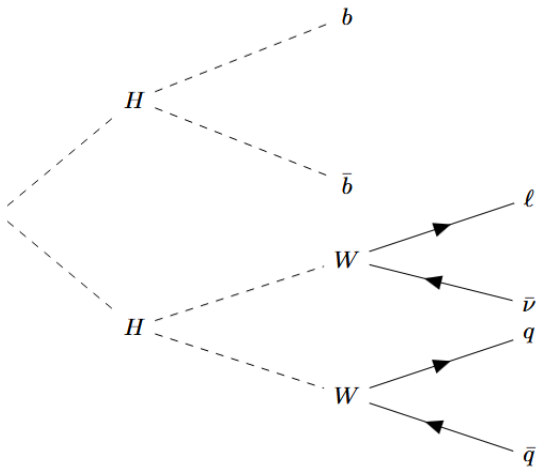



Figure 6: Schematic sketch of semileptonic final state of di-Higgs decay

One important detail of this final state is that since the mass of the Higgs boson is 125 GeV, it can not produce two on-shell W bosons which have a mass of 80.4 GeV. One of the two W bosons (either $W_{had}$ or $W_{lep}$) is produced off-shell which means it has a lower mass than the nominal 80.4 GeV.

## 3.6 $t\bar{t}$ Production as Background Process

A key challenge in selecting VBF di-Higgs events is to distinguish them from background processes that can produce similar final states. In general, the major background process for the VBF di-Higgs production is $t\bar{t}$ production, which at the LHC is mostly produced in gluon-gluon fusion. The semileptonic decay of a top-antitop pair is shown in Fig. 7.

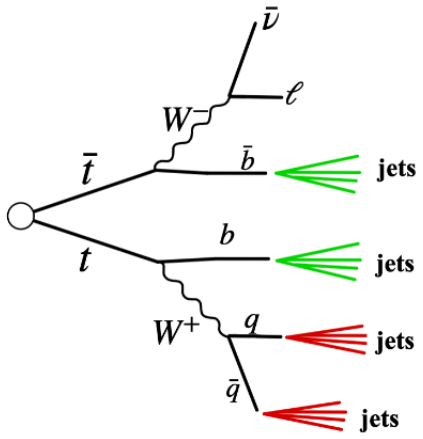



Figure 7: Schematic sketch of the semileptonic decay of a top-antitop pair. [21]

The most common decay mode of the top quark is into a W boson and a b quark, which occurs in approximately 99.8 % of cases. In the case where one W boson decays hadronically and one W decays leptonically, the particles in the final state are identical to the di-Higgs final state from Fig. 6. Although one key difference is that both b quarks decay from different top quarks in oppsite to the di-Higgs decay where both b quarks decay from the one of the Higgs. The same can be said for the W bosons which decay from different top quarks in the $t\bar{t}$ events, but the same particle in the di-Higgs decay.

The main challenge arises from the fact that $t\bar{t}$ production occurs far more frequently than VBF di-Higgs production. The $t\bar{t}$ cross section assuming a mass of 172.5 GeV is $\sigma = 833.9\,\text{pb}^{+20.5}_{-30.0}\,(\text{scale})^{+21.0}_{-21.0}\,(\text{PDF})$ at 13 TeV central mass energy and $\sigma = 923.6\,\text{pb}^{+22.6}_{-33.4}\,(\text{scale})^{+22.8}_{-22.8}\,(\text{PDF})$ at 13.6 TeV. For the selection of VBF di-Higgs events, it is therefore very important to cut off as many of these background events as possible during the event selection [1].

# Chapter 4

# Truth studies

## 4.1 Data and Simulation

To study VBF di-Higgs events in the semileptonic $b\bar{b}W^+W^-$ final state, simulated events in only this specific final state are used. The truth-level studies are based on Monte Carlo simulation samples generated within the ATLAS framework using PYTHIA8 [6]. For the analysis of deviations of $\kappa_{2V}$ to the SM, six different samples with each 50.000 events were used for six different values of the coupling modifier $\kappa_{2V} = [0, 0.5, 1, 1.5, 2, 3]$. The coupling modifiers $\kappa_\lambda$ and $\kappa_V$ are set to one in all six samples, to only focus on deviations of $\kappa_{2V}$.

The truth data used in this chapter only represents the generator level physics of the process without any simulation of the detector. The particles involved in the process and their properties can be directly accessed from the Monte Carlo data at generator level. This truth information is used for the analysis of the VBF di-Higgs process in this chapter. The truth jets that are used are formed with the anti-$k_t$ algorithm, which is a clustering algorithm that combines particles based on their relative distance in rapidity–azimuth space and their transverse momentum. The algorithm preferentially clusters high-$p_T$ particles first, resulting in approximately conical and stable jets [8].

## 4.2 Kinematics of the VBF Di-Higgs Process

The primary goal of the truth studies is to understand the general kinematics of the process and how deviations of $\kappa_{2V}$ from its SM value affect them. For this purpose, the most important properties of the two Higgs bosons and their decay products are studied. For investigating the kinematics of the VBF di-Higgs process, all events with an electron or a muon as the lepton in the final state are used, so approximately $\frac{2}{3}$ of the 50.000 events.

### 4.2.1 $\Delta R$ and $\Delta\phi$ of Higgs Bosons

One important property of the VBF di-Higgs process is the angular separation between the two Higgs bosons, since very characteristic differences can be observed for non-SM scenarios of $\kappa_{2V}$. Fig. 8 shows the $\Delta R$ and $\Delta\phi$ between the two Higgs bosons. For the non SM coupling values, $\Delta\phi$ peaks at $\Delta\phi = \pi$, which means the two Higgs bosons are boosted in opposite directions. This also explains the peak in the $\Delta R$ distribution, which primarily is caused by the $\Delta\phi$ contribution, rather than $\Delta\eta$. For the SM case of $\kappa_{2V} = 1$, both the $\Delta R$ and $\Delta\phi$ curves in black are much flatter and show only a very small peak at $\Delta\phi = \pi$.

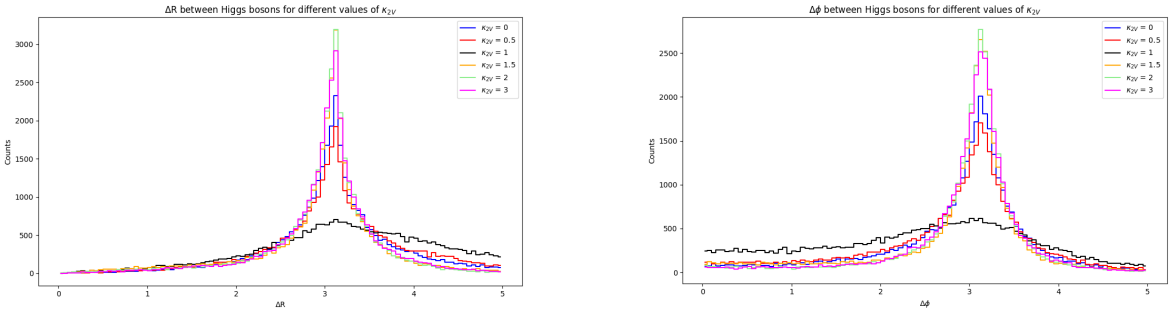



Figure 8: $\Delta R$ and $\Delta\phi$ between the two higgs bosons. The angular separation between the Higgs bosons shows a clear peak for the colored non-SM values of $\kappa_{2V}$ and a much flatter distribution for the black curve representing the SM value. The $\Delta\phi$ plot emphasizes that the peak in the $\Delta R$ plot is mostly caused by the separation in $\phi$

### 4.2.2 Signallepton from $W_{lep}$

The lepton originating from the $W_{lep}$ is a key characteristic of the process. In this studies the focus is on events, in which the lepton is an electron or muon, since tauons decay very fast and therefore are rather difficult to reconstruct. The lepton of the decay of the leptonic W boson is directly identified from the truth Monte Carlo information.

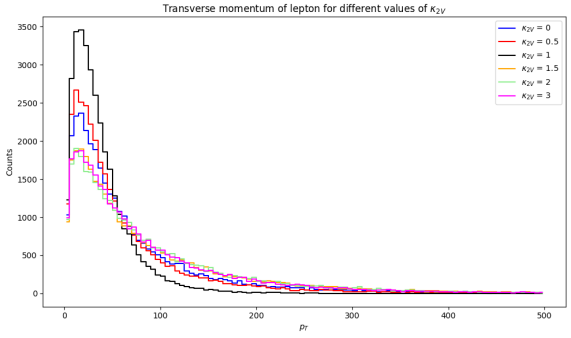



Figure 9: Transverse momentum of electrons and muons. Transverse momentum of lepton peaks at lower $p_T$ for SM in black and is shifted towards higher $p_T$ for non-SM cases

Fig. 9 shows the transverse momentum of the lepton, that directly come from the $W_{lep}$. In general these leptons have a rather high transverse momentum. For non SM values of the coupling modifier $\kappa_{2V}$ the lepton tends to have a higher transverse momentum compared to the SM case.

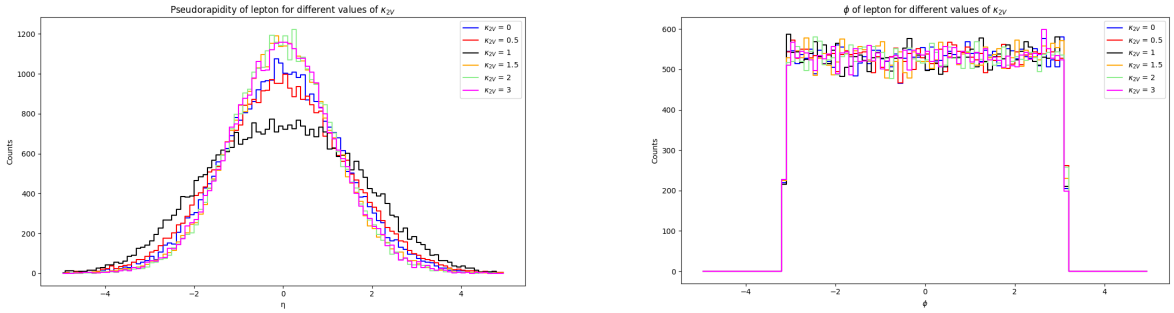



Figure 10: Pseudorapidity $\eta$ and azimuthal angle $\phi$ of electrons and muons. $\eta$ distribution is flatter for black SM curve and more peaked for the non SM values. $\phi$ is evenly distributed for all values of $\kappa_{2V}$

The pseudorapidity $\eta$ and the azimuthal angle $\phi$ of the lepton is shown in Fig. 10. The lepton in non SM scenarios tends to be more central in the detector, which corresponds to a $\eta$ value closer to zero. This also leads to a higher transverse momentum of the lepton in these scenarios as shown before, as $p_T = p/\cosh(\eta)$ increases for smaller values of $\eta$ at fixed momentum.

The angle $\phi$ of the leptons is evenly distributed between $-\pi$ and $\pi$ and the plots show very similar behaviour for all values of $\kappa_{2V}$. This is expected, since for all scenarios of $\kappa_{2V}$, there is no prefered orientation in regard to the angle around the beam axis.

### 4.2.3 $W_{had}$ and $H_{b\bar{b}}$

The $W_{had}$ and $H_{b\bar{b}}$ are identified using Monte Carlo truth information, which means all information about kinematic properties are known at generator level. The transverse momentum of the $W_{had}$ and $H_{b\bar{b}}$ is shown in Fig. 11.

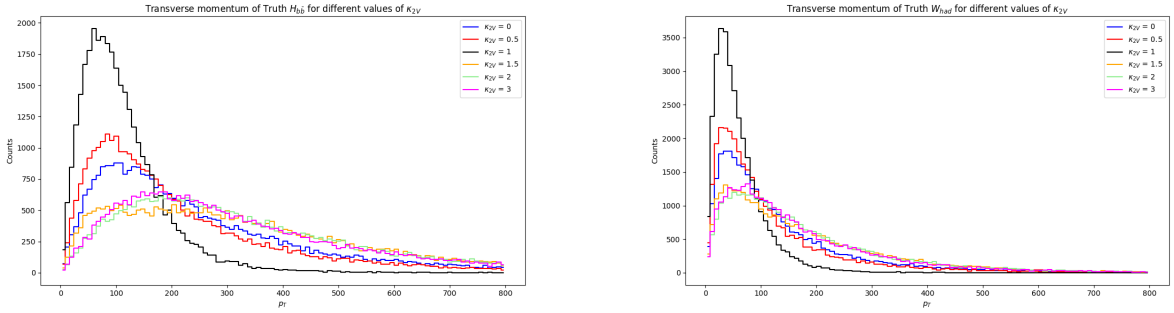



Figure 11: Transverse momentum of $H_{b\bar{b}}$ and $W_{had}$ for different values of $\kappa_{2V}$. Transverse momentum of $H_{b\bar{b}}$ and $W_{had}$ peaks at lower $p_T$ for SM in black and is shifted towards higher $p_T$ for non-SM cases.

For the transverse momentum of the $H_{b\bar{b}}$ and $W_{had}$, the behaviour between the different coupling modifiers is similar compared to the lepton in Fig. 9. In the SM case, the transverse momentum peaks at lower momentum and has only a small tail up to higher momenta. In comparison, for the non SM scenarios, the peak at lower momentum is much smaller and there are much more entries at higher momenta.

The plots also show, that in general the $H_{b\bar{b}}$ has a much higher transverse momentum compared to the $W_{had}$, which is expected since the $W_{had}$ is produced in the decay $H \to WW^*$ and therefore only carries a part of the momentum of the parent Higgs boson.

The plots of $\eta$ and $\phi$ of the $H_{b\bar{b}}$ and $W_{had}$ presented in Fig. 12 show the same tendencies for the different $\kappa_{2V}$ as already seen for the lepton. In the non SM cases, the $H_{b\bar{b}}$ and $W_{had}$ tend have $\eta$ values closer to zero and be more central in the detector. Both the $H_{b\bar{b}}$ and the $W_{\text{had}}$ are uniformly distributed in $\phi$, which reflects the same reason as for the lepton: there is no preferred orientation of the process around the beam axis.

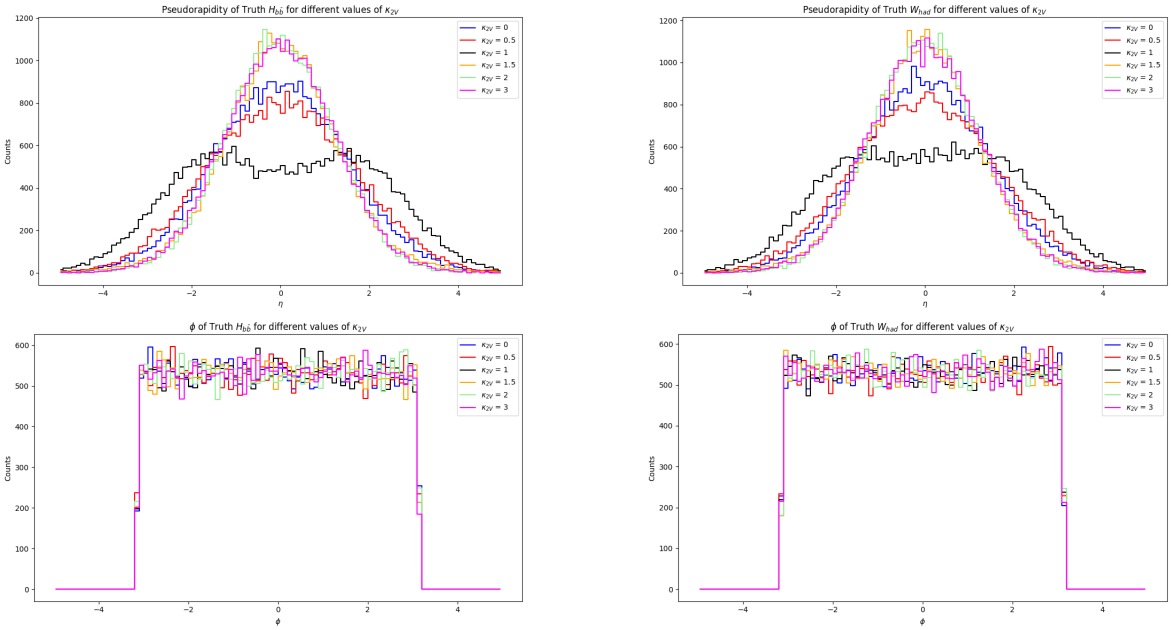
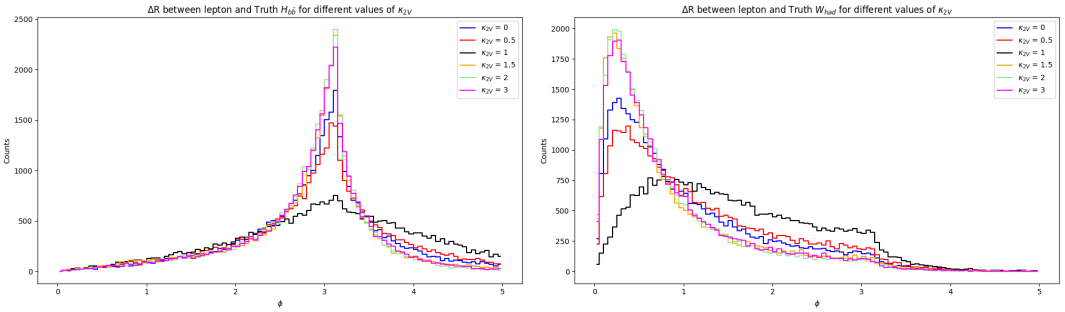



Figure 12: Pseudorapidity and $\phi$ of $H_{b\bar{b}}$ and $W_{had}$. For both $H_{b\bar{b}}$ and $W_{had}$, the $\eta$ distribution is flatter for the black SM curve and more peaked for the non-SM values. $\phi$ is evenly distributed for all values of $\kappa_{2V}$

.

Another relevant property is the angular distance between the lepton and $H_{b\bar{b}}$, aswell as between the lepton and $W_{had}$. The $W_{had}$ originates from the same Higgs boson as the $W_{lep}$, so in general the $W_{had}$ is expected rather close to the lepton, while the $H_{b\bar{b}}$ is expected to have a higher $\Delta R$ to the lepton. The $\Delta R$ of the $H_{b\bar{b}}$ and $W_{had}$ to the lepton is shown in Fig. 13a and Fig. 13b.




(a) $\Delta R$ between $H_{b\bar{b}}$ and lepton

(b) $\Delta R$ between $W_{had}$ and lepton

The $\Delta R$ between the $H_{b\bar{b}}$ and the lepton in Fig. 13a peaks around 3 similar to the $\Delta R$ between the two Higgs bosons in Fig. 8. But the peaks are flatter compared to the di-Higgs $\Delta R$, since the lepton does not necessarily go in the same direction as the Higgs it originates

from. The peak at around $\Delta R = 3$ is much higher for non SM values of $\kappa_{2V}$, which shows that in these scenarios there are indeeed more events where the lepton is boosted in the same direction as the $W_{had}$ and the Higgs, which decays to $W^+W^-$. This can also be observed in Fig. 13b, which shows that the $\Delta R$ between the lepton and the $W_{had}$ tends towards smaller valures for non SM scenarios.

### 4.2.4 Quarks from $W_{had}$ and $H_{bb}$

In Truth data it is possible to directly access information about the quarks that decay from the $W_{had}$ and $H_{b\bar{b}}$. The $\Delta R$ between these quarks is shown in Figure 14.

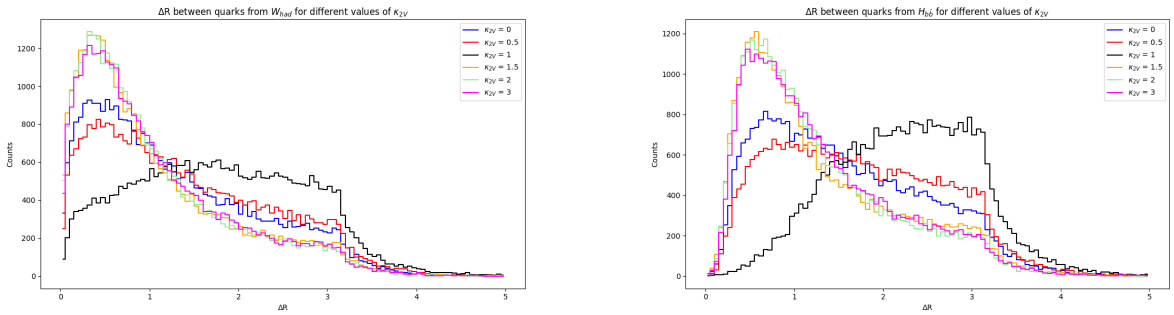



Figure 14: $\Delta R$ between quarks from $W_{had}$ and $H_{b\bar{b}}$ for different values of $\kappa_{2V}$. The angular separation tends to be smaller for non SM scenarios and shifted towards higher values for the SM curves in black.

For both $W_{had}$ and $H_{b\bar{b}}$ the quarks tend to have a much smaller angular distance in the non SM scenarios where the plots show peaks between zero and one. In the SM case the quarks are much more separated and especially in case of the two b quarks from the $H_{b\bar{b}}$ have much less entries for $\Delta R < 1$.

Fig. 14 also shows that the b quarks from the $H_{b\bar{b}}$ in general have a higher angular separation compared to the quarks that decay from the $W_{had}$, which is caused by the different masses of the Higgs boson compared to the W boson and the higher transverse momentum of the b quarks.

### 4.2.5 Conclusion of Kinematics of the VBF Di-Higgs Process

Overall it was shown that in non SM scenarios of $\kappa_{2V}$, the $H_{b\bar{b}}$ and $W_{had}$ have higher transverse momentum, are more central in the detector and their decay products have a smaller angular separation. This demonstrates that a selection in which the $H_{b\bar{b}}$ and $W_{had}$ are reconstructed as large radius (LR) jets is particularly sensitive to deviations of $\kappa_{2V}$ from its Standard Model value, as an increased number of events is expected in this selection for non-Standard Model scenarios. Such a selection will be presented in section 4.4.

## 4.3 VBF jets

The two outgoing quarks of the scattering process produce two smallR jets also called VBF jets. These jets have very characteristic features and are therefore very useful to distinguish the VBF process from other backgrounds. The quarks that take part in the scattering process originate from opposite bunches and are expected to have a high separation in $\eta$ because of that.

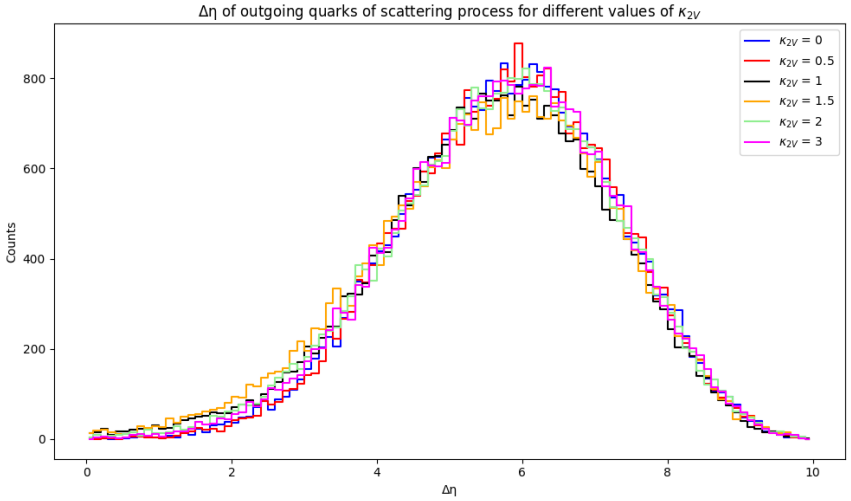



Figure 15: $\Delta\eta$ between quarks from scattering process. Distributions show high $\eta$ separation for all values of $\kappa_{2V}$

The plots in Fig. 15 show the $\Delta\eta$ between the outgoing quarks of the VBF process. The two quarks have a high separation in eta, which does not depend on the value of $\kappa_{2V}$. This is also expected and can be explained by the fact, that the coupling modifier only changes the coupling between the two vector bosons and the two Higgs bosons, but does not directly impact the scattering process.

Another important characteristic of the VBF jets is their high invariant mass, which is expected, because of the fact, that the incoming quarks of the scattering process have a big part of the central mass energy of the collision.

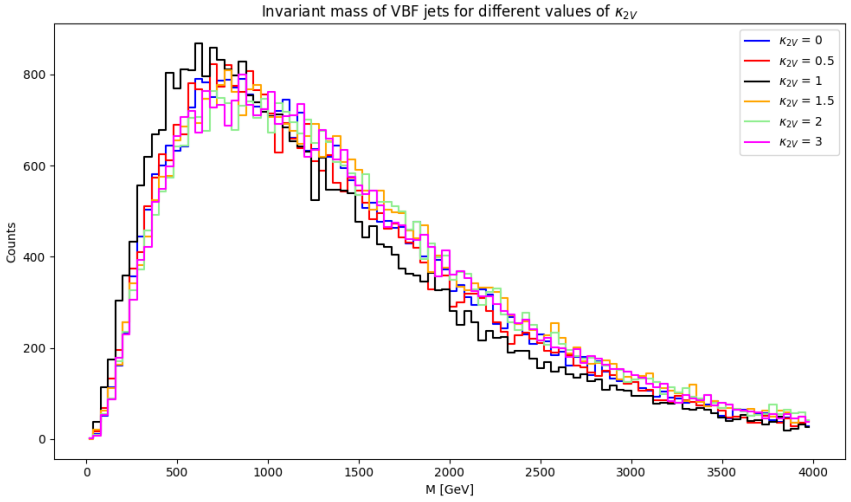



Figure 16: Invariant mass of VBF jets combined. Plots show high invariant mass of VBF jet combination for all values of $\kappa_{2V}$

The invariant mass of the two VBF jets is shown in Figure 16. In the Truth studies the VBF jets are found by matching two smallR jets (R = 0.4) to the two quarks from scattering process with a $\Delta R < 0.4$. The invariant mass of the VBF jets combination is then defined as:

$$m_{\text{inv}}^2 = (p_1 + p_2)^2 \tag{4.1}$$

$$m_{\text{inv}}^2 = (E_1 + E_2)^2 - (p_{x1} + p_{x2})^2 - (p_{y1} + p_{y2})^2 - (p_{z1} + p_{z2})^2 \tag{4.2}$$

where $p_1$ and $p_2$ are the fourvectors of the matched VBF jets. Similar to the $\Delta\eta$ curves in Fig. 15, the plots of the invarinat mass of the VBF jets also look very similar for the different $\kappa_{2V}$ values. The explanation for this is the same as before. The coupling modifier does not directly impact the scattering process.

## 4.4 LR Jet Matching

In general, the quarks from the $W_{had}$ decay and the b quarks from the $H_{b\bar{b}}$ decay can be each reconstructed as smallR jets. These smallR jets are formed by clustering particles within a cone of radius $R = 0.4$ and contain the particles produced during the parton shower and hadronization of the quarks in the detector.

For events, in which the Higgs bosons have high transverse momentum and their decay products are very collimated like it was shown in section 4.2, it is hard to find two resolved jets, but it is possible to reconstruct the quarks coming from the $W_{had}$ and $H_{b\bar{b}}$ as large radius (LR) jets with a radius of R = 1.

For events where both the $H_{b\bar{b}}$ and $W_{had}$ can be reconstructed as LR jets, their mass distributions are shown in Fig. 17.

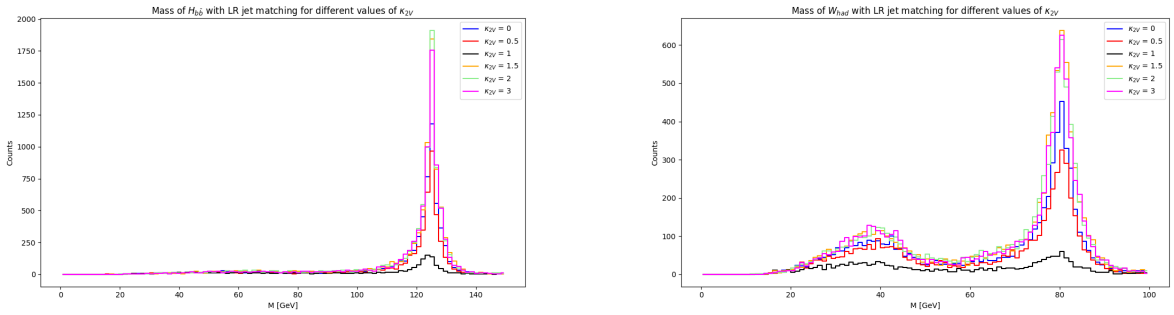



Figure 17: Mass of $H_{b\bar{b}}$ and $W_{had}$ reconstructed with LR jet matching. Plots show more events for non-SM $\kappa_{2V}$ and peak at the expected masses of the Higgs and W boson. The mass plot of the $W_{had}$ shows a second peak at lower masses corresponding to offshell W bosons.

The number of events in both plots heavily depends on the value of $\kappa_{2V}$, which is expected from the kinematics that were shown. To be reconstructed as a LR jet, the $\Delta R$ between the quarks from the $H_{b\bar{b}}$ or $W_{had}$ needs to be small. Like shown in Figure 14, this is much more often the case in the non SM scenarios of $\kappa_{2V}$, which results in more events and higher peaks in the mass plots for the LR jet matching.

The mass of the Higgs boson decaying to two b quarks peaks at the expected Higgs mass of 125 GeV. The mass of the hadronically decaying W boson has a peak at the expected nominal mass at 80 GeV and a second broader peak at lower mass. This second peak corresponds to events, in which the $W_{lep}$ has the nominal mass of 80 GeV and the $W_{had}$ is produced off-shell and has a much lower mass.

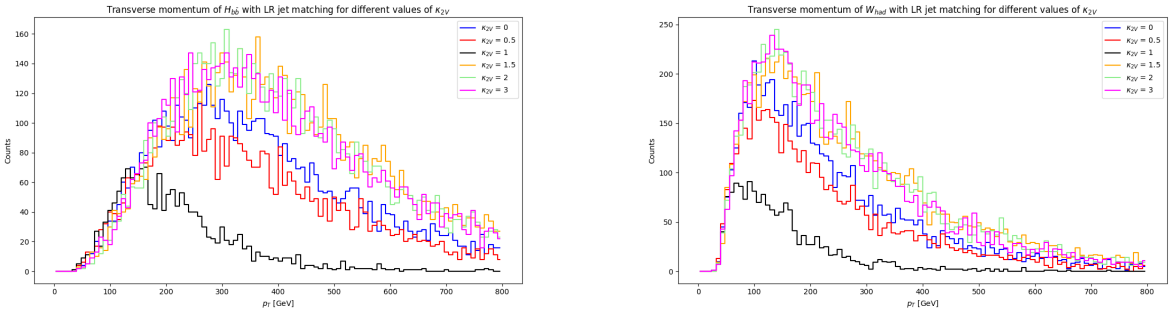



Figure 18: Transverse momentum of $H_{b\bar{b}}$ and $W_{had}$ reconstructed with LR jet matching

The diagrams in Fig. 18 show the transverse momentum of the $H_{b\bar{b}}$ or $W_{had}$, that were reconstructed with LR jet matching. The transverse momentum of the events selected with LR jets is much higher compared to the $p_T$ of all events in general, which was shown in Fig. 11. This also verifies, that LR jet matching is very sensitive to boosted events, which are expected to have a high transverse momentum.

## 4.5 Semiboosted Approach

As in the analyses of the kinematics, all events with an electron or a muon as the lepton in the final state are used for investigating the semiboosted approach in this section. In section 4.4 it was shown, that in boosted events the $H_{b\bar{b}}$ and $W_{had}$ can be reconstructed as LR jets. For an event selection of a fully boosted event two LR jets would be required, one corresponding to the $H_{b\bar{b}}$ and one corresponding to the $W_{had}$. Althogh requiring both $H_{b\bar{b}}$ and $W_{had}$ to be reconstructed as LR jets provides a selection which is highly sensitive to $\kappa_{2V}$, the requirement may reject events in which only one of the two Higgs bosons is boosted. A semiboosted approach of an event selection would select such events while possibly still be sensitive to anomalous $\kappa_{2V}$.

The idea of the semiboosted approach is to search for events where only either the $H_{b\bar{b}}$ or the $W_{had}$ is boosted and reconstructed as a LR jet. The goal in the truth studies is to show how often such semiboosted events occur compared to fully boosted events, and if the semiboosted events have the same sensitivity to $\kappa_{2V}$.

One possible definition for a boosted Truth $H_{b\bar{b}}$ and a Truth $W_{had}$ is given by:

$$\text{Truth } H_{b\bar{b}} : \Delta R(b,\bar{b}) < 1 \quad p_T > 250GeV \tag{4.3}$$

$$\text{Truth } W_{had} : \Delta R(q,\bar{q}) < 1 \quad p_T > 150GeV \tag{4.4}$$

The requirement $\Delta R < 1$ is motivated by the LR jet radius R = 1, ensuring both quarks can be captured within a single LR jet. The transverse momentum thresholds can be motivated by the following approximate relation for boosted to body decays:

$$\Delta R \sim \frac{2m}{p_T} \tag{4.5}$$

Requiring the quarks to fit inside a LR jet with R = 1 implies $p_T \gtrsim \frac{2m}{R}$, which for the given masses of the Higgs and W boson corresponds roughly to the given $p_T$ thresholds [13].

Fig. 19 shows the percentages of fully boosted and semiboosted events for the given definition of boosted truth $H_{b\bar{b}}$ and $W_{had}$. As expected from the kinematics discussed in 4.2, the amount of fully boosted events depends heavily on $\kappa_{2V}$ resulting in much more fully boosted events

for the non SM values of $\kappa_{2V}$. But also in the semiboosted case, the percentage of events for the different coupling strengths looks very similar compared to the fully boosted case, although there are a bit more events for the SM scenario. So overall the semiboosted case also shows high sensitivity to deviations of $\kappa_{2V}$ to the SM.

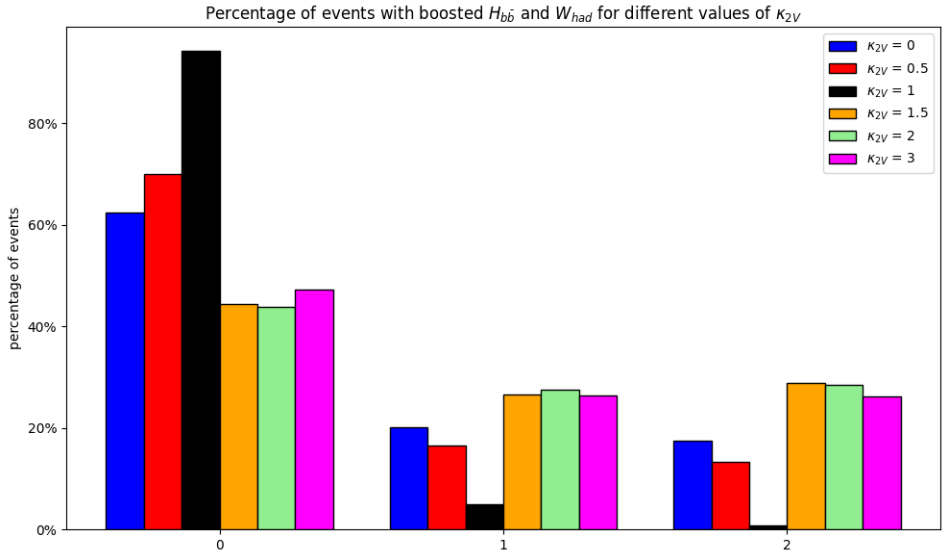
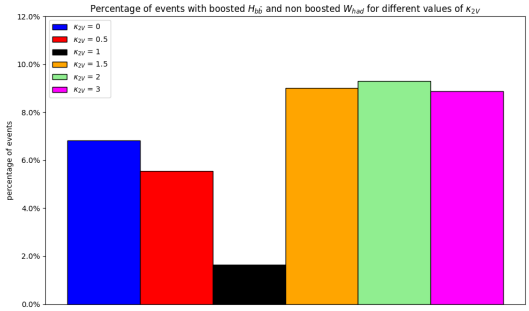
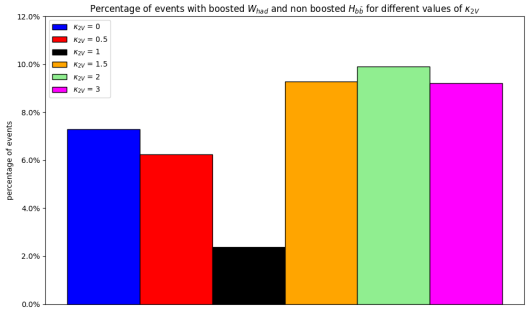



Figure 19: percentage of events with boosted Truth $H_{b\bar{b}}/W_{had}$: enries in 0 represent events where neither $H_{b\bar{b}}$ or $W_{had}$ are boosted according to the defined conditions. For entries in 1, either the $H_{b\bar{b}}$ or the $W_{had}$ is boosted and for entries in 2 both are boosted.

As mentioned before, in the semiboosted approach there are two different scenarios: events with a boosted $H_{b\bar{b}}$ and a non boosted $W_{had}$ or events with a boosted $W_{had}$ and a non boosted $H_{b\bar{b}}$. Since the two Higgs bosons are produced symmetrically in the process, no strong preference is expected for either semiboosted topology. To verify this, Fig. 20 shows the percentage of events for both semiboosted cases separately.




(a) percentage of events with boosted Truth $H_{b\bar{b}}$ and non boosted Truth $W_{had}$




(b) percentage of events with boosted Truth $W_{had}$ and non boosted Truth $H_{b\bar{b}}$

Figure 20: Percentage of events for both semiboosted cases

As expected, the percentages of events is very similar for the two different semiboosted cases. In addition, the behaviour of the different coupling modifiers is the same for both semiboosted cases. Overall, these truth-level studies show that an event selection for semiboosted topologies provides sensitivity to $\kappa_{2V}$.

This motivates an explicit derivation of the event selection for the semiboosted approach, which is presented for a non-SM scenario in the next chapter.

# Chapter 5

# Event Selection for Semiboosted Approach with ATLAS Detector Simulation Data

As the truth studies have shown that the semiboosted approach is sensitive to $\kappa_{2V}$, the goal is to derive an event selection for a non SM scenario of $\kappa_{2V} = 1.5$. The analysis is performed using Monte Carlo simulation, which now includes a full simulation of the ATLAS detector as presented in 2. The simulation of the detector is done using Geant4 [2]. Overall the data used for the analysis includes 118795 di-Higgs VBF signal events in the semileptonic final state and $4.16 \times 10^8$ semileptonic $t\bar{t}$ background events. The plots shown in this chapter do not show the number of overall events on the y axis, but a number of expected events, which is weighted with the cross section of the according process and the integrated luminosity achieved in 2018.

The objective is to construct an event selection that exploits the semiboosted topology, in which only either the $H_{b\bar{b}}$ or the $W_{had}$ is reconstructed as a LR jet, while the other is reconstructed as two resolved smallR jets. An overview of the selection strategy is given in Fig. 21.

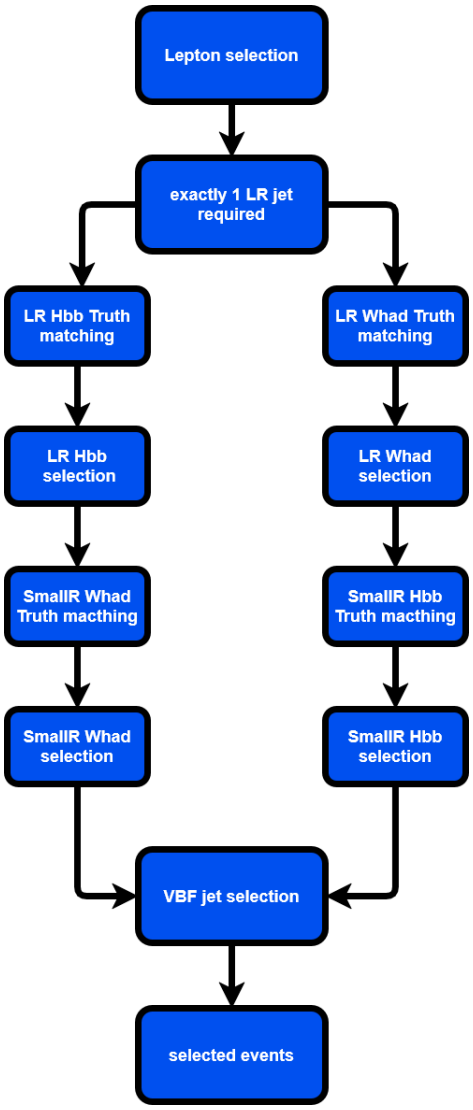



Figure 21: Schematic of the seletion steps for the semiboosted approach

First the lepton originating from the decay of the $W_{lep}$ is selected, which is explained in section 5.1. After the lepton selection, exactly 1 LR jet is required in the event, representing the semiboosted topology in which only the $H_{b\bar{b}}$ or the $W_{had}$ is boosted and can be reconstructed as a LR jet. In the next step the selection is divided into the two separate mentioned cases. The LR jet is matched either to the truth $H_{b\bar{b}}$ or the truth $W_{had}$, which is exlpained in section 5.2 and defines the two different selections. In the case in which the LR jet is matched to the truth $H_{b\bar{b}}$, a selection of the LR jet will be derived followed by a truth matching and selection of two smallR jets to reconstruct the non-boosted $W_{had}$. In the case of a boosted $W_{had}$, the LR jet is therefore matched to the truth $W_{had}$ and a selection criteria for the selection of the LR jet coming from the $W_{had}$ is derived. This is followed by the truth matching and selection of two smallR jets to reconstruct the Higgs decaying into two b quarks. In the end a selection of the two VBF jets is derived, which is identical for both of the two boosted cases, since the VBF jets should not be affected by which of the $H_{b\bar{b}}$ or the $W_{had}$ is boosted. Events which pass all of the steps are then selected.

The efficiencies of the selection criteria derived for signal events are compared to $t\bar{t}$ events for each step. The truth matching steps are only applied in signal to define the two different selection cases and develop separate selection criteria. The selection efficiencies for each step

are defined by the number of events that pass the selection divided by the number of events before the selection step. An overview of the cumulative selection efficiencies for the whole selection is shown in the last section of the chapter.

## 5.1 Lepton Selection

The first step of the event selection is to select the lepton in the final state, which originates from the decay of the $W_{lep}$. As mentioned, only events with an electron or muon are considered in this selection. The lepton is required to have a minimum transverse momentum of $p_T >$ 10 GeV. Additionally, it is required that either an electron or a muon working point (WP) is passed, which defines a set of identification and isolation criteria to distinguish the final state lepton from the $W_{lep}$ decay from background contributions, such as misidentified jets. For the selection a Tight WP is chosen, which refers to the strongest restrictive selection in regards to the identification and isolation criteria. Overall, exactly 1 lepton with a transverse momentum of at least 10 GeV is selected, that passes either the Tight electron WP or the Tight muon WP [4] [15]. An overview of the selection efficiencies of the lepton selection is given in Table 1. Since the lepton selection is the first step of the selection the efficiency is defined by the events that pass the lepton selection divided by all events in the dataset. The efficiency of the WP in Table 1 is defined relative to the $p_T$ requirement. It is given by the fraction of events that pass both the $p_T$ cut and the WP, divided by the number of events that pass the $p_T$ cut.

| selection cut | signal efficiency | $t\bar{t}$ efficiency |
|---|---|---|
| lepton $p_T > 10\ GeV$ | 46.36 % | 55.44 % |
| lepton passes WP | 88.44 % | 77.37 % |
| lepton selection combined ($p_T$ cut and WP) | 41.00 % | 42.89 % |

Table 1: selection efficiencies of lepton selection for signal and $t\bar{t}$ background

About one third of both signal and background events are rejected because the lepton is a tau. Apart from that, in most of the rejected signal events the lepton does not pass the $p_T$ threshold of 10 GeV. The overall efficiency for the lepton selection is very similar for signal and $t\bar{t}$ events, as expected since both final states contain a lepton originating from the decay of a W boson.

## 5.2 LR Jet Truth Matching

Following the lepton selection, for the semiboosted approach it is required that exactly one LR jet exists per event, which can either come from the $H_{b\bar{b}}$ or the $W_{had}$. The efficiency in signal and $t\bar{t}$ for this requirement is shown in Table 2. The efficiency is defined by the number of events that pass the lepton selection and have exactly 1 LR jet divided by the overall number of events that pass the lepton selection.

| | signal | $t\bar{t}$ background |
|---|---|---|
| exactly 1 LR jet | 34.75 % | 9.47 % |

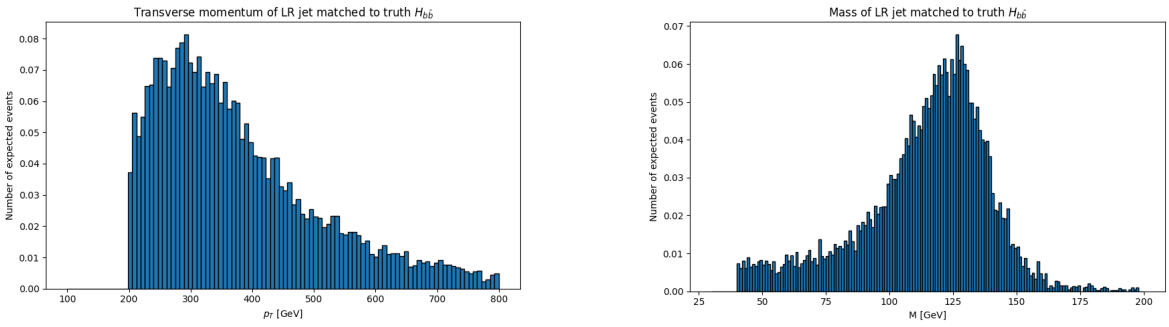
Table 2: efficiency of requirirng exactly 1 LR jet in signal and $t\bar{t}$ background events

More of the signal events have 1 LR jet in the event compared to $t\bar{t}$ background events. The LR jet in semileptonic $t\bar{t}$ production can potentially arise from the $W_{had}$ similar as in signal or even possibly from the $W_{had}$ in combination with tracks from the b quark that decays from the same top quark as the $W_{had}$. However, LR jets originating from two b quarks are more characterisic of signal events, since in $t\bar{t}$ production the b quarks originate from different top quarks. Therefore, the higher efficiency in signal compared to $t\bar{t}$ can be expected when requiring exactly 1 LR jet.

To distinguish the two cases and derive separate event selections, the LR jet is matched either to the truth $H_{b\bar{b}}$ or the truth $W_{had}$ in the next step. For the LR jet truth matching, the angular separation between the LR jet and the truth level $H_{b\bar{b}}$ or $W_{had}$ is required to satisfy $\Delta R < 0.5$.

### 5.2.1 LR $H_{b\bar{b}}$ Truth Matching

The transverse momentum and mass of the LR jet, that passes the matching to the truth $H_{b\bar{b}}$ is shown in Fig. 22. The minimum transverse momentum threshod for a LR jet is 200 GeV, which causes the cutoff at this $p_T$ value in the diagram. Apart from that, the transverse momentum of the reconstructed LR jet looks similar to the $p_T$ distribution of the $H_{b\bar{b}}$ from the LR jet matching in the truth studies in Fig. 18. The mass of the LR jet shows a broad peak around the expected Higgs mass of 125 GeV and a small tail towards lower masses. Since the LR jet just includes all tracks of particles inside a cone of radius R = 1, any missing or incorrectly assigned tracks can lead to deviations from the expected jet mass.




(a) Transverse momentum of LR jet

(b) Mass of LR jet

Figure 22: Transverse momentum and mass of LR jet matched to truth $H_{b\bar{b}}$

In the truth studies it was discussed, that for $\kappa_{2V} = 1.5$ the $\Delta R$ between the lepton and the $H_{b\bar{b}}$ has a clear peak around 3, which corresponds to the fact that also the $\Delta R$ between the two Higgs has a peak around 3 in Fig.8 and the decay products of the Higgs have small angular separation. The $\Delta R$ between the lepton and the reconstructed LR jet that is matched to the truth $H_{b\bar{b}}$ also demonstrate this tendency, which is shown in Fig. 23.

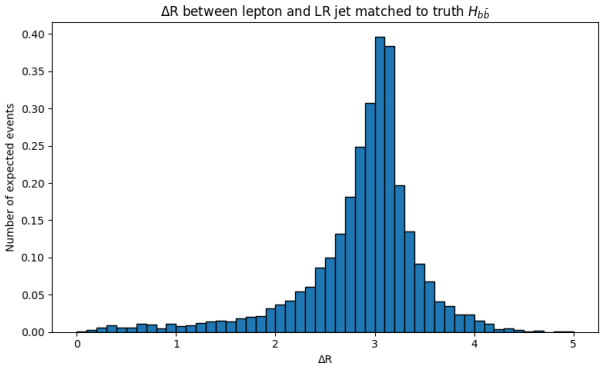
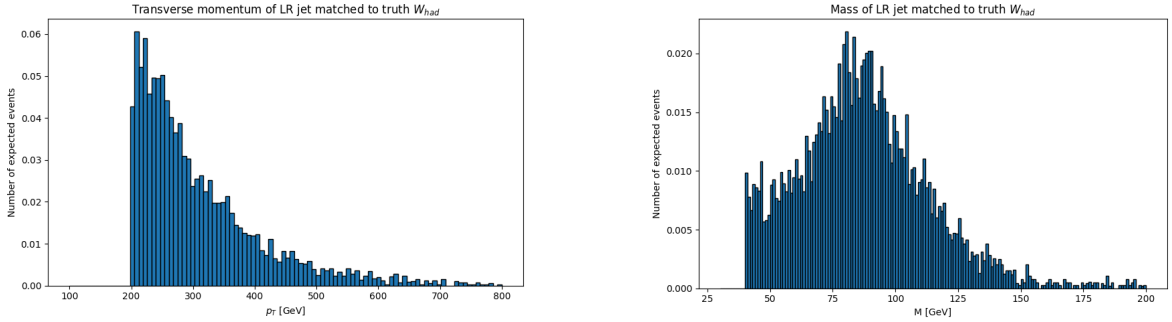



Figure 23: $\Delta R$ between lepton and LR jet for LR jet matched to truth $H_{b\bar{b}}$

In overall 68.63 % of the events with 1 LR jet, the LR jet is matched to the truth $H_{b\bar{b}}$. This is in contrast to the results from the truth studies, which show comparable percentages of events for boosted $H_{b\bar{b}}$ and $W_{had}$ in semiboosted events. A possible reason for this might be the requirements for the reconstruction of an LR jet. This will be further explained when looking at the LR jets matched to the truth $W_{had}$ in the next section.

### 5.2.2 LR $W_{had}$ Truth Matching

In 25.76 % of the events the LR jet is matched to the truth $W_{had}$. The transverse momentum and mass of the LR jet in these events is shown in Fig. 24.




(a) Transverse momentum of LR jet

(b) Mass of LR jet

Figure 24: Transverse momentum and mass of LR jet matched to truth $W_{had}$

The transverse momentum of the LR jets matched to the $W_{had}$ in Fig. 24a is generally lower compared to the LR jets that were matched to the $H_{b\bar{b}}$ in Fig. 22a, which is expected, as it was observed in truth studies that the $H_{b\bar{b}}$ has in general a higher $p_T$ than the $W_{had}$. This could also explain why for events with exactly one LR jet, the majority of them is matched to the truth $H_{b\bar{b}}$, since the required $p_T$ threshold of 200 GeV for LR jets seems to cut off events where the $W_{had}$ could be reconstructed by a LR jet, but does not have the required $p_T$. The mass of the LR jet in Fig. 24b has a broad peak at around 80 GeV, which is the expected mass of the W boson. A potential mass peak of the offshell $W_{had}$ is only barely visible, as the mass is cut off at the minimum required mass for a LR jet at 40 GeV.

In the case of a boosted $W_{had}$, which is reconstructed as LR jet, the selected lepton from the leptonically decaying W is also expected to be close to the LR jet, potentially even in the LR jet, as both W bosons come from the same Higgs. This can be verified in Fig. 25, which shows the $\Delta R$ between the selected lepton and the LR jet that is matched to the truth $W_{had}$.

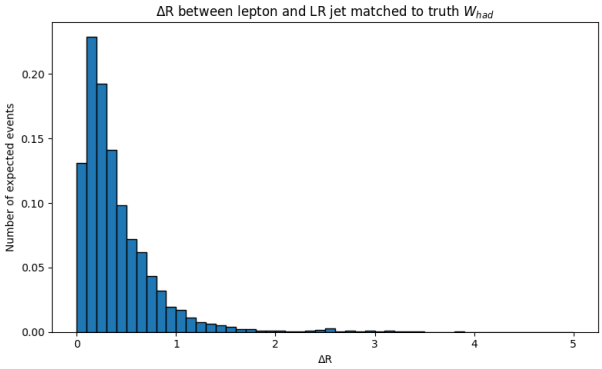



Figure 25: $\Delta R$ between lepton and LR jet for LR jet matched to truth $W_{had}$

For most events where the LR jet comes from the $W_{had}$, the $\Delta R$ between the lepton and the LR jet is smaller than one. A comparison of Fig. 23 and Fig. 25 indicates that the $\Delta R$ between the lepton and the LR jet is very characteristic and can be used to distinguish between LR jets originating from the $H_{b\bar{b}}$ or the $W_{had}$.

## 5.3 LR Jet Selection

After the truth matching of the LR jet, the next step is to find a criteria to select the LR jet for both cases. The goal is to select as many of the truth matched signal events as possible, while also rejecting as much $t\bar{t}$ background as possible.

### 5.3.1 LR $H_{b\bar{b}}$ Selection

To select the LR jet from the $H_{b\bar{b}}$, one possible selection criterion is to require a certian $\Delta R$ interval for the $\Delta R$ between the lepton and the LR jet. For $2 < \Delta R < 4$ most of the truth matched events would be selected like shown in Fig. 23. This $\Delta R$ criterion would perform well for the selection on signal, but it is also impotant to consider how the $\Delta R$ between the LR jet and the lepton is distributed in $t\bar{t}$ events with exactly one LR jet, which is shown in Fig. 26.

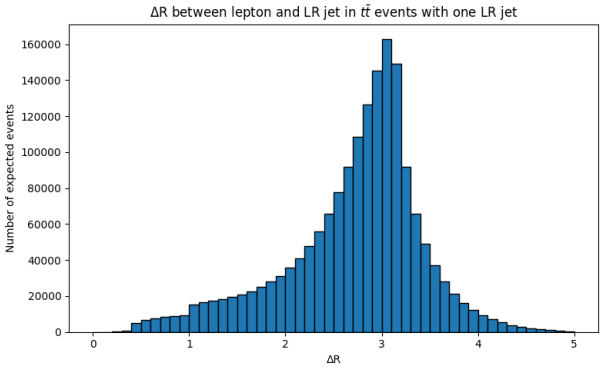



Figure 26: $\Delta R$ between LR jet and lepton for $t\bar{t}$ events with exactly one LR jet

In $t\bar{t}$ events the LR jet can potentially arise from the $W_{had}$ alone or in combination with the b quark, which decays from the same top quark as the $W_{had}$. The $W_{had}$ and $W_{lep}$ decay from different top quarks, which means LR jets coming from the $W_{had}$ alone or in a combination

with the b quark have a rather high angular separation to the lepton, which can be observed in Fig. 26. Comparing Fig. 26 and Fig. 23 the $\Delta R$ between the LR jet and the lepton peaks within a similar range for $t\bar{t}$ events and signal events where the LR jet is truth matched to the $H_{b\bar{b}}$. This impilies selecting the $H_{b\bar{b}}$ LR jet just based on a $\Delta R$ criterion works for signal, but also results in selecting most of the $t\bar{t}$ background events.

Another option to select the $H_{b\bar{b}}$ LR jet is to use a graph neural network classifier, which is called GN2X working point (WP). The GN2X algorithm is specifically trained to identify LR jets that come boosted Higgs bosons decaying into a pair of bottom quarks. The algorithm uses typical features of b-hadron decays like displaced secondary vertices which occur because b-hadrons travel a measurable distance in the detector before they decay. The GN2X WP can be defined at different efficiencies where higher efficiency corresponds to looser tagging requirements and therefore higher acceptances of the LR jets. The efficiency here refers to the fraction of QCD background LR jets that pass the GN2X WP, with the 1.25 % WP being the loosest available WP [18].

The mass of the LR jet for signal events that pass the LR $H_{b\bar{b}}$ truth matching and the GN2X WP at 1.25 % efficiency is shown in Fig. 27.

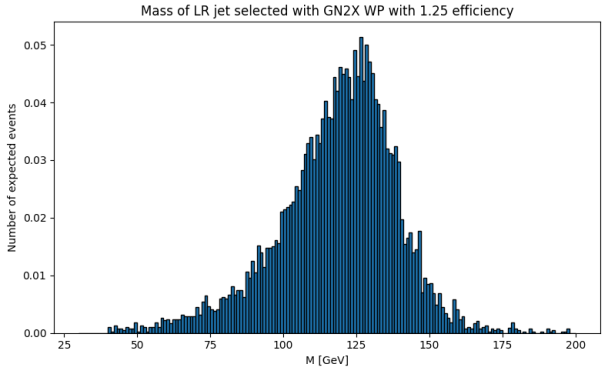



Figure 27: Mass of LR jet selected with GN2X WP

In comparison to the truth matched events in Fig. 22b, the GN2X WP cuts off many events in small mass tail of the LR jet, but also the overall peak decreases. The signal selection efficiencies of the $H_{b\bar{b}}$ LR jet selection relative to the truth matched events for the $\Delta R$ criterion described before and two efficiencies for the GN2X WP are shown in Table 3. The efficiencies are defined by the number of events that pass the LR $H_{b\bar{b}}$ selection criterion divided by the number of events that pass the selection including the LR $H_{b\bar{b}}$ truth matching.

| selection cut | absolute number of events | percentage of events |
|---|---|---|
| $\Delta R$ criterion | 10516 | 90.55 % |
| GN2X WP 0.94 | 7444 | 64.10 % |
| GN2X WP 1.25 | 8003 | 68.91 % |

Table 3: Efficiencies of LR $H_{b\bar{b}}$ signal selection comparing GN2X WP with $\Delta R$ criterion relative to events that pass the LR $H_{b\bar{b}}$ truth matching

The comparison of these signal efficiency to the selection effficiency of the criteria for $t\bar{t}$ events is illustrated in Fig. 28.

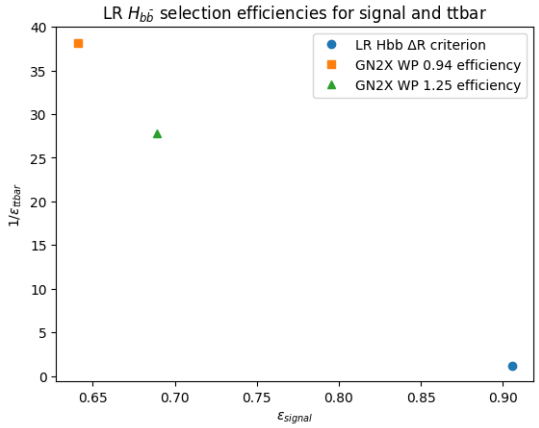



Figure 28: Selection efficiencies of LR $H_{b\bar{b}}$ comparing selection criteria in signal and $t\bar{t}$ background events. The signal efficiencies from table 3 are on the x-axis and the inverse of the $t\bar{t}$ selection efficiency are on the y-axis.

The GN2X WPs reject much more of the $t\bar{t}$ events compared to the $\Delta R$ criterion, because they specifically try to identify LR jets from a pair of b quarks, which is very unlikely to happen in a $t\bar{t}$ event, since the b quarks originate from different top quarks. The 1.25 GN2X WP accepts more signal events, but also rejets less $t\bar{t}$ background events compared to the 0.94 GN2X WP. Because the signal process is already very rare, the GN2X WP with an efficiency of 1.25 % is chosen for the selection of the $H_{b\bar{b}}$ LR jet, which has a higher signal efficiency, taking into account less background rejection.

### 5.3.2 LR $W_{had}$ Selection

As illustrated in Fig. 25, one possible strategy is to identify LR jets originating the $W_{had}$ is to require a small angular separation $\Delta R$ between the LR jet and the lepton. Fig. 26 also shows that such a $\Delta R$ criterion would reject most of the $t\bar{t}$ events. To achieve a good tradeoff between signal efficiency and backround rejection, five different values for a $\Delta R$ threshold were tested for signal and $t\bar{t}$ backround events. The $\Delta R$ thresholds were chosen to select most of the truth matched events based on Fig. 25.

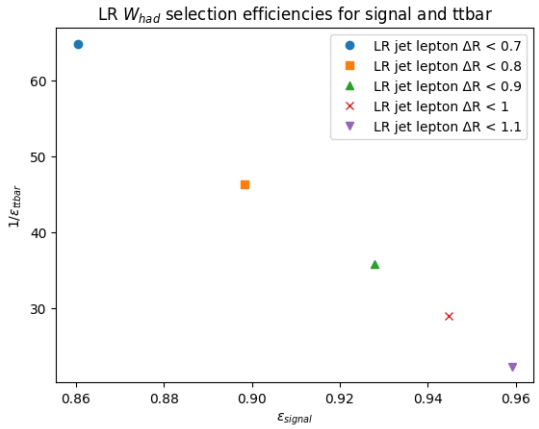



Figure 29: Selection efficiencies of LR $W_{had}$ comparing selection criteria in signal and $t\bar{t}$ background events. The signal efficiencies from table 3 are on the x-axis and the inverse of the $t\bar{t}$ selection efficiency are on the y-axis.

Fig. 29 compares the selection efficiency of the different $\Delta R$ cuts for signal and $t\bar{t}$ background. Similar to the $H_{b\bar{b}}$ seletion, for choosing an appropiate cut requires a tradeoff between signal efficiency and backround rejection. Since the focus is on keeping most of the signal events, the $\Delta R < 1$ cut was chosen for the selection of the LR $W_{had}$ jet, which still provides good background rejection while selecting 94.49 % of the truth matched signal events.

## 5.4 SmallR Jet Truth Matching

The next step after the selection of the LR jet is to select two smallR jets corresponding to the b quarks from the $H_{b\bar{b}}$ or the two quarks from the $W_{had}$, depending on the selected LR jet. Before the actual selection a truth matching is applied for both cases to study the characteristics of the smallR jets for possible selection criteria. To avoid any overlap of the smallR jets with the selected LR jet, all smallR jets are required to have $\Delta R > 1.4$ to the LR jet, in order to be considered for the smallR jet selection or the VBF jet selection and therefore also for the smallR jet truth matching steps.

In the whole event selection for the semiboosted approach, overall four smallR jets need to be selected, two VBF jets and two smallR jets corresponding either to the $H_{b\bar{b}}$ or the $W_{had}$. To account for this, an additional cut is applied before the truth matching and selection of smallR jets, which requires at least four smallR jets in the event. Therefore all events with three or less smallR jets get rejected. This requirement is very useful to reject a high percentage of $t\bar{t}$ background events while keeping most of the signal events since there are no VBF jets in $t\bar{t}$ events and therefore overall less smallR jets. The efficeincies of requiring four smallR jets in the event are shown in Table 4. The efficiency here is defined by the number of events after the LR jet selection that pass the cut of having four smallR jets divided by the total amount of events after the LR jet selection.

| | signal efficiency | $t\bar{t}$ background efficiency |
|---|---|---|
| four smallR jets required after LR $H_{b\bar{b}}$ selection | 89.91 % | 11.78 % |
| four smallR jets required after LR $W_{had}$ selection | 93.54 % | 10.97 % |

Table 4: signal and background efficiencies of requiring four smallR jets in the event after the LR jet selection.

For both selections almost 90 % of background events get rejected while keeping most of the signal events. The requirement of at least four smallR in an event is therefore a very useful selection step before moving on to the smallR jet truth matching and the smallR jet selection.

### 5.4.1 SmallR Jet Truth Matching of $H_{b\bar{b}}$

The truth matching of smallR jets to the b quarks from the $H_{b\bar{b}}$ is performed for events that pass the LR $W_{had}$ selection. The data includes Truth information about the Higgs bosons and their children, so the two b quarks in the case of the $H_{b\bar{b}}$. In the truth matching step, for each b quark, the smallR jet closest to the b quark is selected and the $\Delta R$ between the b quark and the smallR jet is required to be smaller than 0.5. So for an event to pass the truth matching there needs to exist a smallR jet for each b quark with $\Delta R < 0.5$. For events that pass the truth matching, the fourvectors of the two smallR jets can be added to reconstruct the $H_{b\bar{b}}$. The mass and transverse momentum of this from smallR jets reconstructed $H_{b\bar{b}}$ is shown in Fig. 30.

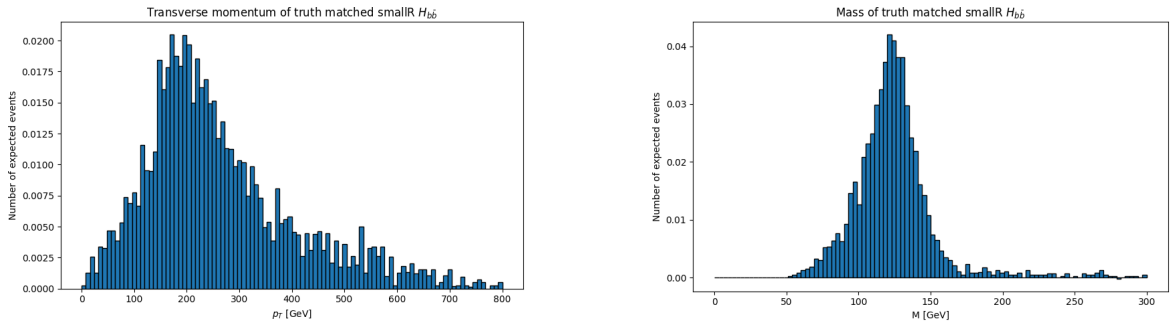



Figure 30: Transverse momentum and mass of $H_{b\bar{b}}$ reconstructed with two truth matched smallR jets

The mass of the reconstructed $H_{b\bar{b}}$ peaks at the expected Higgs mass of 125 GeV. The transverse momentum of the non-boosted $H_{b\bar{b}}$ reconstructed by smallR jets is much smaller compared to the boosted case where the $H_{b\bar{b}}$ can be reconstructed with one LR jet, which was shown in Fig. 22. Another important property for the selection later on is the $\Delta R$ between the truth matched smallR jets, presented in Fig. 31.

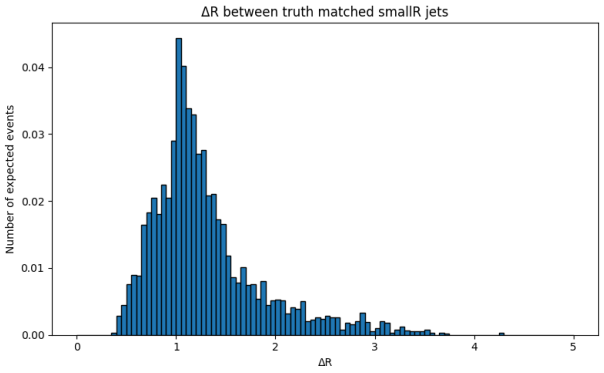
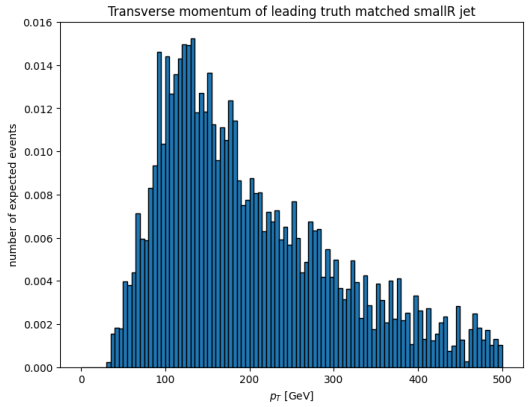


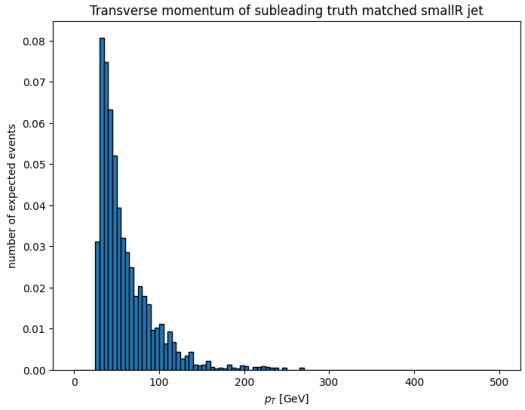

Figure 31: $\Delta R$ between smallR jets truth matched to $H_{b\bar{b}}$

The $\Delta R$ distribution peaks around one, indicating that even when reconstructed as two resolved jets, the jets remain relatively close in angular separation in the semiboosted events.

Another relevant property for the selection is the transverse momentum of the individual smallR jets. For two resolved jets originating from two b quarks one would expect a $p_T$ in a similar range for both jets. The transverse momentum of the two truth matched smallR jets is shown in Fig. 32.




(a) Transverse momentum of leading $p_T$ smallR jet



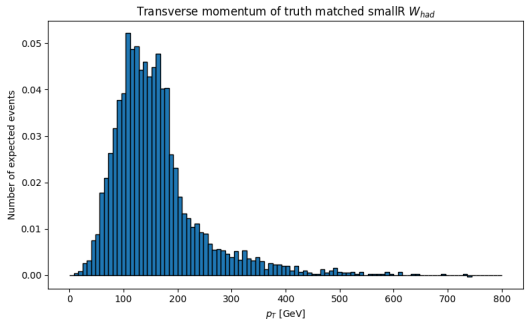

(b) Transverse momentum of subleading $p_T$ smallR jet

Figure 32: Transverse momentum of smallR jets truth matched to b quarks of $H_{b\bar{b}}$. The leading $p_T$ smallR jet has significantly higher transverse momentum compared to the subleading matched smallR jet

The leading smallR jet with higher $p_T$ shown in Fig. **??** tends to have much more transverse momentum compared to the second matched jet with lower $p_T$ in Fig. **??**. The transverse momentum of the subleading jet is significantly lower and, for many events, lies close to the minimum threshold of $p_T = 20$ GeV. This can explain cases in which no pair of small-R jets can be matched to the two b-quarks. The second jet may fail to be reconstructed if its transverse momentum falls below the required threshold.

### 5.4.2 SmallR Jet Truth Matching of $W_{had}$

The truth matching of smallR jets to the quarks of the $W_{had}$ can not be done in the same way as for the $H_{b\bar{b}}$, because the truth infomrmation is only provided for the children of the Higgs bosons ($W_{had}$ and $W_{lep}$ in this case), but not for the quarks from the $W_{had}$. So instead of directly matching the smallR jets to the quarks, the approach is to find the combination of smallR jets that matches best to the truth $W_{had}$. For this, out of all possible combinations of two smallR jets the one with the smallest $\Delta R$ to the truth $W_{had}$ is selected. In order to pass the truth matching this $\Delta R$ is then required to be smaller than 0.5. The transverse momentum and mass of the combination of smallR jets for events that pass the truth matching requirement is shown in Fig. 30.



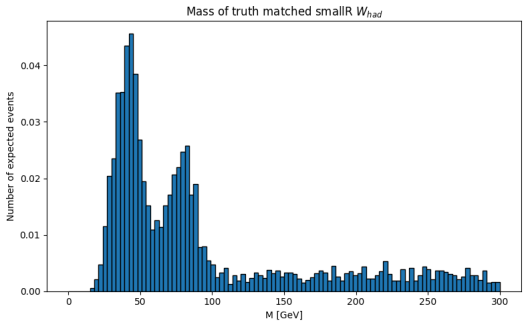




Figure 33: Transverse momentum and mass of $W_{had}$ reconstructed with two truth matched smallR jets

Similar to the case of the smallR $H_{b\bar{b}}$, the transverse momentum of the $W_{had}$ reconstructed by two smallR jets is lower compared to the boosted $W_{had}$ reconstructed by a single LR jet in Fig. 24. The mass plot has two peaks: one at the nominal mass of a W boson around 80 GeV and one higher peak for W bosons produced at a offshell mass at around 40 GeV. There are also a few entries in the tail of the mass distribution, which have a much higher mass than expected for a W boson. The likely reason for these entries is the fact that the matching was done with the combination of smallR jets to the truth $W_{had}$ instead of a direct matching to the quarks. Since combinations of smallR jets that do not actually originate from the quarks of the $W_{had}$, but randomly match to the truth $W_{had}$ when adding the fourvectors of the jets also pass the truth matching. These combinations then usually have a high angular separation, which is demonstrated in Fig. 34. High $\Delta R$ values between the smallR jets result in high invariant masses for the combination of the jets, which corresponds to the tail of the mass distribution.

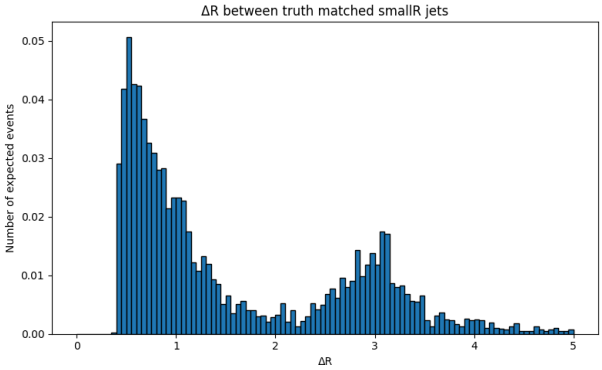



Figure 34: $\Delta R$ between smallR jets truth matched to $W_{had}$

To verify that the second peak in the $\Delta R$ distribution actually is caused by the matching as jet combinations, the matching for the truth $W_{had}$ can be compared to the truth matching for smallR jets of the $H_{b\bar{b}}$, where the matching is now done in the same way for the jet combination. The $\Delta R$ between the two matched jets in this case is shown in Fig. 35.

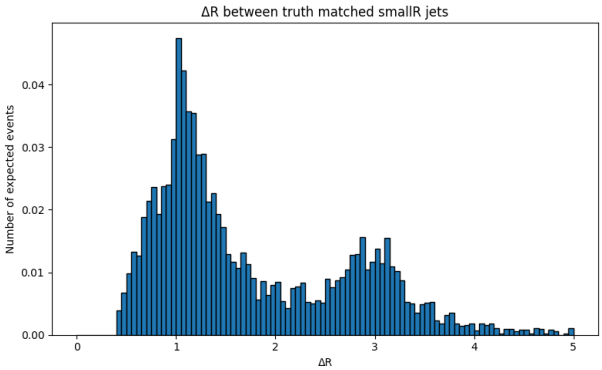



Figure 35: $\Delta R$ between smallR jets truth matched to $H_{b\bar{b}}$

A second peak in the plot of the $\Delta R$ between the matched jets shows up, which was not the case when matching the jets directly to the b quarks as in 32. A comparison between Fig. 35 and fig. 31 demonstrates that the second peak in the plot is only caused by the matching as a jet combination instead of matching the smallR jet directly to the quarks like it was originally done for the $H_{b\bar{b}}$. This indicates that the $\Delta R$ between the smallR jets originating from the quarks of the $W_{had}$ is expected to be rather small as reflected in the left peak of the

plot in Fig. 34.

The transverse momenta of the two smallR jets truth matched as a combination to the $W_{had}$ are shown in Fig. 36.

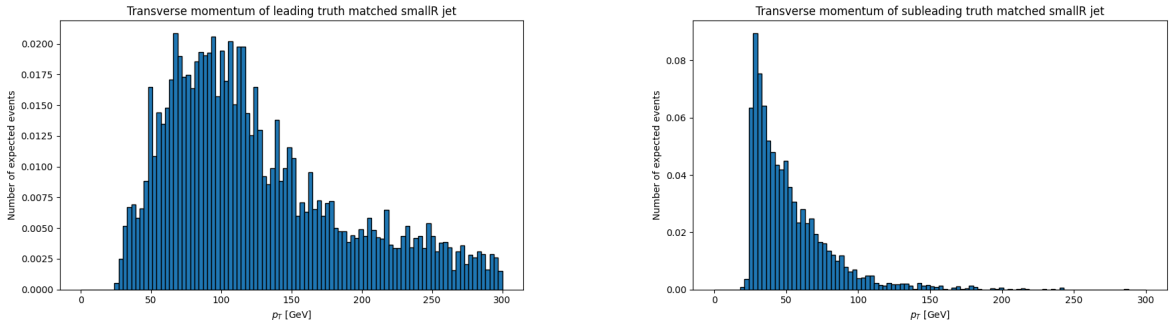
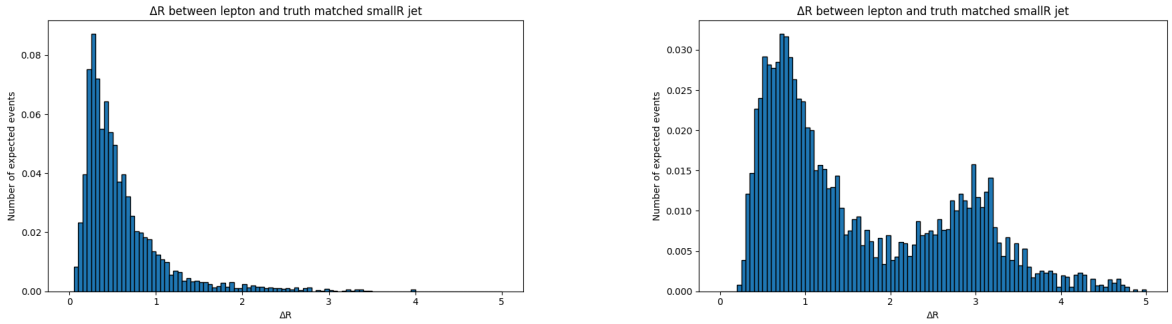



Figure 36: Transverse momentum truth matched smallR jets ordered by $p_T$.

The smallR jets matched for the $W_{had}$ have in general lower $p_T$ compared to the jets that were matched to the b quarks in the smallR $H_{b\bar{b}}$ truth matching, which can be expected since it was sown that the $W_{had}$ in general has lower $p_T$ compared to the $H_{b\bar{b}}$. But similar to the smallR jets of the $H_{b\bar{b}}$, the leading $p_T$ smallR jet here also has significantly higher transverse momentum compared to the subleading jet.

A key property to potentially select the smallR jets produced by the quarks of the $W_{had}$ is their $\Delta R$ to the lepton. Even in the semiboosted topology the smallR jets are expected to be rather close to the lepton since both $W_{had}$ and $W_{lep}$ come from the same Higgs. This is illustrated in Fig. 37 which shows the $\Delta R$ of both truth matched smallR jets to the lepton.




(a) $\Delta R$ between lepton and smallR jet closest to the lepton

(b) $\Delta R$ between lepton and smallR jet second closest to the lepton

Figure 37: $\Delta R$ between the matched smallR jets and the lepton

The distributions show clear peaks at smallR $\Delta R$ between the jets and the lepton. In some the events, the second matched jet has a very high $\Delta R$ to the lepton which is likely due to incorrect jet assignment caused by the truth matching as jet combination as discussed before. Overall this suggests the smallR jets originating from the $W_{had}$ are typically found within $\Delta R \lesssim 2$ to the lepton, which correspond to the peaks in Fig. 37.

## 5.5 SmallR Jet Selection

### 5.5.1 SmallR $H_{b\bar{b}}$ Selection

To select the two smallR jets coming from the two b quarks, b-tagging can be used. Similar to the GN2X WP, there also exist different efficiencies for the b tagging of smallR jets. In order to maximize signal efficiency a b-tagging WP with a high efficiency of 85 % was chosen. To identify smallR jets originating from b quarks, the b tagger uses typical features of b-jets like displaced secondary vertices and large impact parameter. For the selection of the non boosted $H_{b\bar{b}}$ two b tagged smallR jets would be expected, one for each b quark. The number of b-tagged smallR jets after the truth matching is shown in Fig. 38.

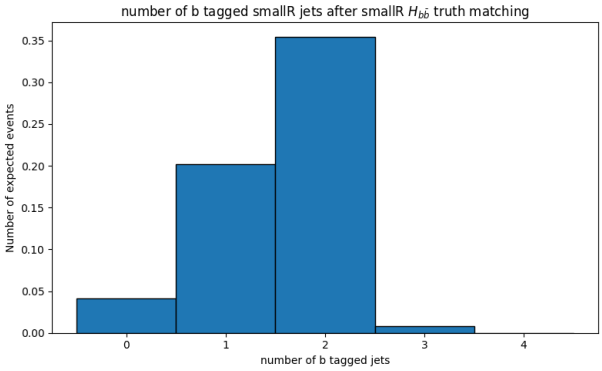
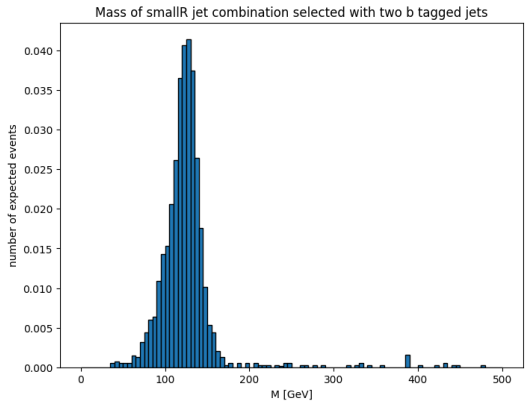



Figure 38: Number of b-tagged smallR jets in signal events after smallR $H_{b\bar{b}}$ truth matching

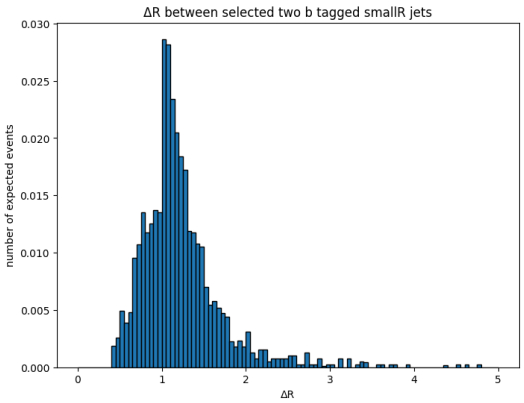
The expected two b tagged smallR jets do not exist in all of the truth matched events. There are some events with only one b tagged jet and very few events with zero or three b tagged smallR jets. Since the number of events with zero or three b tagged jets is very small, only events with one or two b tagged jets are considered for the selection.

In events containing two b-tagged jets, these jets can be directly identified as the candidates for the $H_{b\bar{b}}$ reconstruction. For events with only one b-tagged jet, additional criteria must be applied to identify the second smallR jet. Since it was shown in Fig. 31 that for the truth matched events the $\Delta R$ between the smallR jets is rather small, one approach is to select the smallR jet closest to the b-tagged jet.

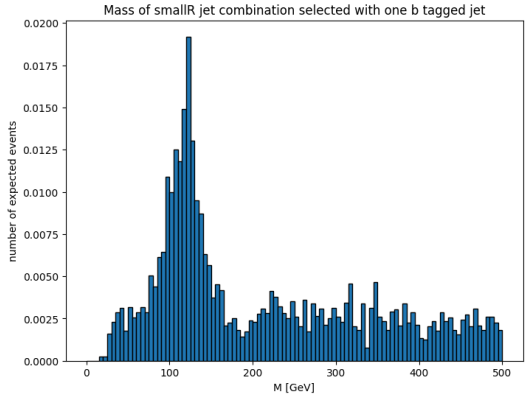
To validate the selection and reject events in which the $H_{b\bar{b}}$ can not be reconstructed it is also necessary to study events that fail the truth macthing. For this purpose, the mass and $\Delta R$ between the selected jets for events with one and two b-tagged jets before the truth matching step is shown in Fig. 39.



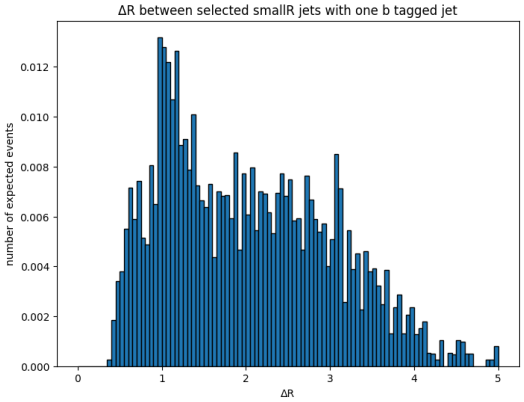

(a) Invariant mass of two b tagged smallR jets




(b) $\Delta R$ between two b tagged smallR jets




(c) Invariant mass of selected b tagedd jet combined with closest smallR jet




(d) $\Delta R$ between selected b tagged jet and closest smallR jet

Figure 39: Mass of $\Delta R$ of selected jet combinations: events with two b tagged jets on top and events with one b tagged jet on bottom. The selections are without the smallR $H_{b\bar{b}}$ truth matching step, so the selection is applied after the LR jet selection. In the case of two b-tagged jet the distributions look like expected from truth matching. For the case of one b-tagged jet some additional criteria must be applied, since the second selected jet might not originate from a b quark

For events with two b-tagged jets the distributions look very similar to the truth matched events in figure 31 and 30. This should be the case, because two the b-tagged jets are expected to come from the two b quarks. For events with one b-tagged jet, a tail for higher masses appears in the mass plot in Fig. 39c. Considering the $\Delta R$ distribution for these events in figure 39d, it indicates that the higher reconstructed masses are associated with a larger angular separation between the selected jets than expected from truth matching. So for some events with just one b-tagged jet, the second selected jet closest to the b-tagged jet might not actually come from a b quark, but from another smallR jet in the event.

One approach for an additional selection criterion is to require the $\Delta R$ between the two selected jets to be below a certain threshold. For the case of two b-tagged jets it looks like such a $\Delta R$ criterion is not needed for signal. But since $t\bar{t}$ events also have two b quarks in the final state, which come from different top quarks, the $\Delta R$ criterion should be applied for events with one and two b-tagged jets. To decide on the exact threshold for the $\Delta R$ cut, different values were tested for signal and $t\bar{t}$ background. The selection efficiencies for $\Delta R$ cuts between 2.0 and 3.0 for signal and background events are plotted in Fig. 40 and an overview of the exact values is given in the Table 5. Similar to the LR jet selections, the signal

efficiency is defined by the number of truth matched events that pass the selection criterion divided by the overall number of truth matched events. The $t\bar{t}$ efficiency is defined as the number of events that pass the selection steps including the smallR $H_{b\bar{b}}$ selection divided by the number of events that only pass the selection up to the requirement of four smallR jets.

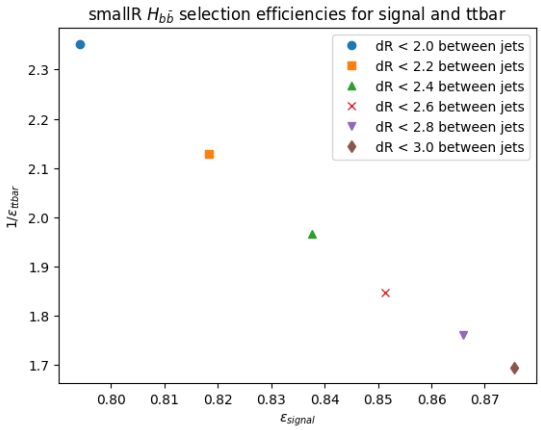



Figure 40: selection efficiencies for smallR $H_{b\bar{b}}$ selection for signal and $t\bar{t}$. The $\Delta R$ between the selected jets is required to be smaller than a certain threshold and different values for the $\Delta R$ were tested. The two selected are either the two b-tagged jets or the one b-tagged jet and jet closest to the b-tagged jet. The signal efficiency is on the x-axis and the inverse of the background efficiency is on the y-axis.

| $\Delta R$ cut | signal efficiency | $t\bar{t}$ background efficiency |
|---|---|---|
| $\Delta R < 2.0$ | 79.42 % | 42.51 % |
| $\Delta R < 2.2$ | 81.84 % | 46.98 % |
| $\Delta R < 2.4$ | 83.77 % | 50.88 % |
| $\Delta R < 2.6$ | 85.14 % | 54.14 % |
| $\Delta R < 2.8$ | 86.59 % | 56.79 % |
| $\Delta R < 3.0$ | 87.56 % | 59.01 % |

Table 5: selection efficiency of smallR $H_{b\bar{b}}$ selection for signal and background for different $\Delta R$ cuts

The selection efficiencies increase more strongly for background compared to signal for higher $\Delta R$ thresholds. This favors a smaller threshold, although no single cut turns out to be a clear optimum. The $\Delta R < 2.2$ was chosen in the end, since it has better background supression compared to higher $\Delta R$ cuts. In addition, it was shown that larger $\Delta R$ thresholds for events with one b-tagged jet increasingly select signal events in which the second jet does not originate from a b quark.

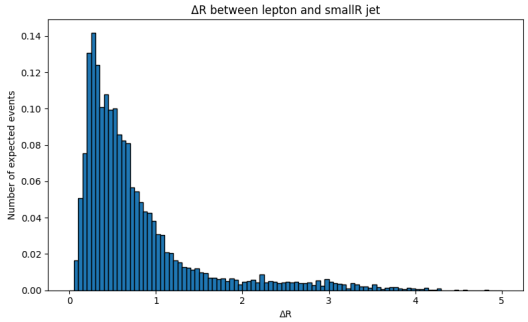
So overall the selection of the smallR $H_{b\bar{b}}$ is done as follows: If two b-tagged smallR jets exist in the event, these are selected if the $\Delta R$ is smaller than 2.2 between the jets. If one b-tagged smallR jet exists, the b-tagged jet and the smallR jet closest to the b-tagged jet are selected if the $\Delta R$ is smaller than 2.2 between the jets.

### 5.5.2 SmallR $W_{had}$ Selection

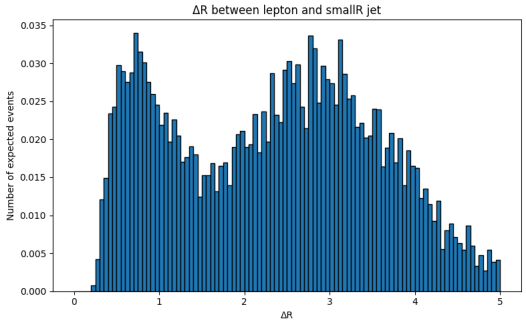
For the selection of two smallR jet that come from the quarks of the $W_{had}$, one needs to make use of the properties that are known from the truth matched events. Since it is expected that

the jets from the $W_{had}$ are relatively close to the lepton, one possible approach is to select two smallR jets that are closest to the lepton.

Although two small-R jets closest to the lepton can be identified in every event, they do not necessarily originate from the $W_{had}$. Therefore it is necessary to also study events that are not truth matched to derive an additional criterion for the selection of the two smallR jets. Fig. 41 shows the $\Delta R$ between the lepton and the two selected jets and Fig. 42 presents the $\Delta R$ between the two selected jets without the truth matching step.




(a) $\Delta R$ between lepton and smallR jet closest to the lepton




(b) $\Delta R$ between lepton and smallR jet second closest to the lepton

Figure 41: $\Delta R$ between lepton and two closest smallR jets to the lepton

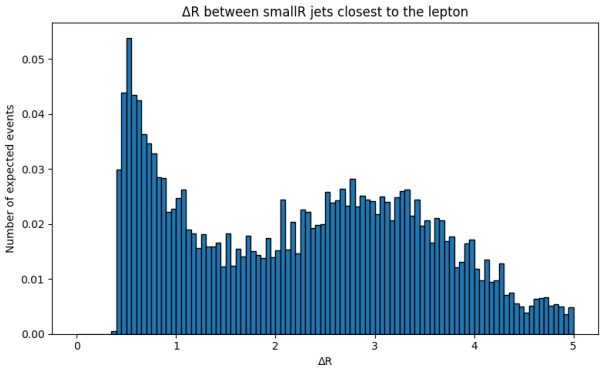



Figure 42: $\Delta R$ between smallR jets closest to the lepton

The $\Delta R$ between the jets has a peak for low values of $\Delta R$ similar to the expected truth matched events in Fig. 34. In a lot of events the $\Delta R$ between the jets closest to the lepton is much higher than expected for two smallR jets coming from the $W_{had}$. The reason for this can be seen when considering the $\Delta R$ of both jets to the lepton separately. In most events it exists one jet relatively close to the lepton, but only some events have a second smallR jet within a similar range to the lepton. The broad second peak for higher values of $\Delta R$ in Fig. 41b shows that the second closest smallR jet to the lepton already has a much higher angular separation to the lepton than expected for the truth matched events.

This suggests to require maximum $\Delta R$ between the selcted jets similar to the case of the smallR $H_{b\bar{b}}$. Six different potential values for a $\Delta R$ threshold were tested for signal and $t\bar{t}$ background. The efficiencies illustrated in Fig. 43 and Table 6 are defined in the same way as in the smallR $W_{had}$ selection.

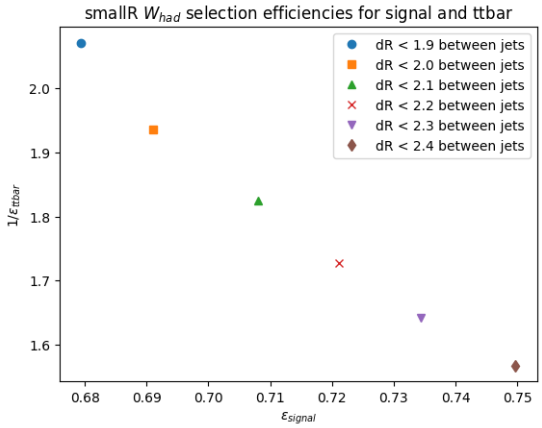



Figure 43: selection efficiencies for smallR $W_{had}$ selection for signal and $t\bar{t}$. The $\Delta R$ between the selected jets is required to be smaller than a certain threshold and different values for the $\Delta R$ were tested. The two selected jets are the two jets closest to the lepton. The signal efficiency is on the x-axis and the inverse of the background efficiency is on the y-axis.

| $\Delta R$ cut | signal efficiency | $t\bar{t}$ background efficiency |
|---|---|---|
| $\Delta R < 1.9$ | 67.94 % | 48.27 % |
| $\Delta R < 2.0$ | 69.10 % | 51.66 % |
| $\Delta R < 2.1$ | 70.80 % | 54.79 % |
| $\Delta R < 2.2$ | 72.11 % | 57.88 % |
| $\Delta R < 2.3$ | 73.44 % | 60.90 % |
| $\Delta R < 2.4$ | 74.97 % | 63.79 % |

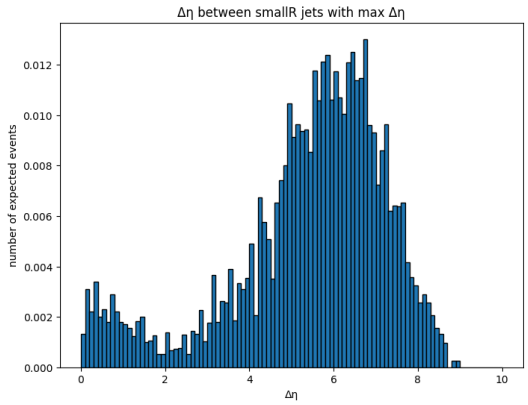
Table 6: selection efficiency of smallR $W_{had}$ selection for signal and background for different $\Delta R$ cuts

Similar to the smallR $H_{b\bar{b}}$ selection, there is no single cut which shows a clear optimum. Considering the $\Delta R$ plot from truth matching 34 and the signal-to-background ratio, a $\Delta R$ cut of $\Delta R < 2.1$ was chosen for the selection.

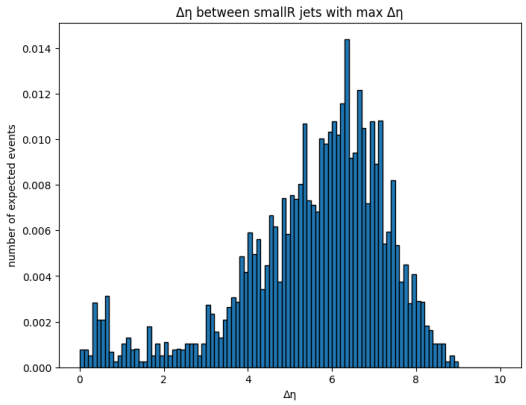
In summary, the reconstruction of the smallR $W_{had}$ is performed by selecting the two smallR jets with the smallest angular separation to the lepton and requiring their separation to satisfy $\Delta R < 2.1$.

## 5.6 VBF Jet Selection

The final step of the event selection is the VBF jet selection. SmallR jets that were previously selected in the smallR jet selection are excluded from the VBF jet selection. Like mentioned in the truth studies the VBF jets are produced by the quarks and are expected to have a high eta separation, which was shown in Fig. 15. Therefore to select these jets the approach is to select the jet combination with the maximum $\Delta\eta$ and require a certain minimal $\Delta\eta$. In order to choose a suitable threshold, the $\Delta\eta$ distributions for signal is shown in Fig. 44 and $t\bar{t}$ background in Fig. 45. The plots show both selection regions, meaning either events that pass the LR $H_{b\bar{b}}$ and smallR $W_{had}$ selection or the evnts that pass the LR $W_{had}$ and smallR $H_{b\bar{b}}$ selection, including the truh matching steps for signal events.

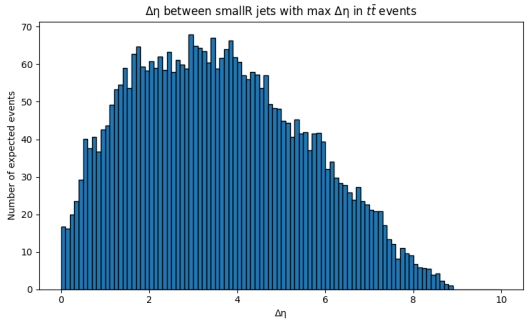



(a) $\Delta\eta$ of smallR jets with maximum $\Delta\eta$ for signal events with boosted $H_{b\bar{b}}$

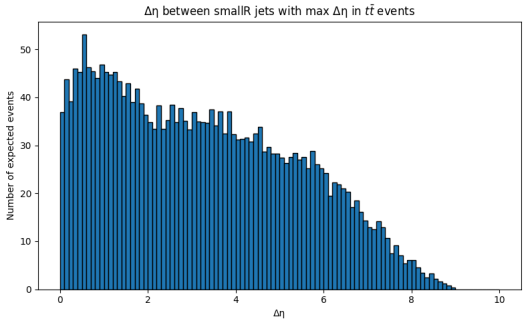



(b) $\Delta\eta$ of smallR jets with maximum $\Delta\eta$ for signal events with boosted $W_{had}$

Figure 44: $\Delta\eta$ of smallR jets with maximum $\Delta\eta$ for signal




(a) $\Delta\eta$ of smallR jets with maximum $\Delta\eta$ for $t\bar{t}$ events that pass the selection for boosted $H_{b\bar{b}}$ including the smallR $W_{had}$ selection




(b) $\Delta\eta$ of smallR jets with maximum $\Delta\eta$ for $t\bar{t}$ events that pass the selection for boosted $W_{had}$ including the smallR $H_{b\bar{b}}$ selection

Figure 45: $\Delta\eta$ of smallR jets with maximum $\Delta\eta$ for $t\bar{t}$ background

The $\Delta\eta$ for signal events show the expected peaks for high for eta separation with a smaller tail at lower values. Possible reasons for not finding two jets with the expected eta separation are that one of the VBF jets is not reconstructed correctly or that on eof the VBf jets is falsely selected in the smallR jet selection before and can therefore not get selected as VBF jet.

For $t\bar{t}$ events the curves are more shifted towards lower values of $\Delta\eta$, which is consistent with the fact that no VBF jets are expected in $t\bar{t}$ production. However, because the smallR jet pair with maximum $\Delta\eta$ is chosen it shows that also in $t\bar{t}$ production some events have two jets with high eta separation.

Comparing the signal and $t\bar{t}$ distributions, there is no clear cut that separates signal from background events. To keep most of the signal events, a $\Delta\eta$ threshold of 3.5 is chosen, which is roughly the value where the peaks in the signal plots starts to arise.

The selection efficiencies for the VBF jet selection are shown in 7 for signal and background for both selection regions. The efficiency is defined by the number of events that pass the VBF jet selection and all previous selection steps divided by the number of events that passed the previous steps but not the VBF selection.

The selection efficiencies for the VBF jet selection are shown in Table 7 for signal and background for both selection regions.

| | signal | $t\bar{t}$ background |
|---|---|---|
| boosted $H_{b\bar{b}}$ | 59.79 % | 50.12 % |
| boosted $W_{had}$ | 69.00 % | 39.84 % |

Table 7: efficiencies of VBF jet selection in both selection region for signal and $t\bar{t}$ background

The efficiencies for both selection regions are higher in signal compared to background like expected from the $\Delta\eta$ plots shown before. However the the discrepancy in the efficiencies for signal compared to background is not as high as expected with the fact that there should not exist VBF jets in the $t\bar{t}$ events. This is because a big part of the background rejection of the VBF jets is already done by the requirement of 4 smallR jets before the smallR jet selections. This cut provided very strong background rejection with high signal efficiency, because of the presence of VBF jets, 2 additional smallR jets are expected in signal events, but not in $t\bar{t}$ events.

## 5.7 Full Event Selection

After applying the full event selection, the overall efficiencies for signal and background can be compared for both boosted reconstruction scenarios. The cumulative efficiencies at each selection step are summarized in Table 8 for the boosted $H_{b\bar{b}}$ selection and in Table 9 for the boosted $W_{had}$ selection. The truth matching steps that were performed in signal are included in the according selection step for signal events.

| selection cut | cumulative signal | cumulative $t\bar{t}$ background |
|---|---|---|
| lepton selection | 40.997 % | 42.894 % |
| exactly 1 LR jet | 14.246 % | 4.062 % |
| LR jet selection | 6.737 % | 0.146 % |
| at least 4 smallR jets | 6.302 % | 0.016 % |
| smallR jet selection | 2.060 % | 0.009 % |
| VBF jet selection | 1.232 % | 0.004 % |

Table 8: cumulative selection efficiencies for signal and background in case of boosted $H_{b\bar{b}}$.

| selection cut | cumulative signal | cumulative $t\bar{t}$ background |
|---|---|---|
| lepton selection | 40.997 % | 42.894 % |
| exactly 1 LR jet | 14.246 % | 4.062 % |
| LR jet selection | 3.435 % | 0.140 % |
| at least 4 smallR jets | 3.080 % | 0.016 % |
| smallR jet selection | 1.648 % | 0.008 % |
| VBF jet selection | 1.137 % | 0.003 % |

Table 9: cumulative selection efficiencies for signal and background in case of boosted $W_{had}$.

The cumulative efficiencies for the whole selection are very similar for both semiboosted selection paths, in both signal and background events. Since the two selection regions are distinct and defined by the LR jet assignment, the total efficiency is obtained by combining the contributions from both selections. Therefore 2.369 % of signal events and 0.007 % of $t\bar{t}$ events are selected by the selection overall. The LR jet selection efficiencies may appear low when considered individually. However, the LR jet selection divides the analysis into two

complementary regions, which means the corresponding efficiencies must also be combined to obtain the overall performance.

The largest losses in signal efficiency are therefore caused by the smallR jet selection and also VBF jets selection. The reconstruction of two smallR jets associated with either the $H_{b\bar{b}}$ or the $W_{had}$ proves to be challenging, especially in the case of the $W_{had}$. One possible reason for this is that one smallR jet is not reconstructed because it does not satisfy the transverse momentum threshold of $p_T > 20GeV$. As a result, the two quarks may be too close to be reconstructed as two distinct smallR jets, while at the same time the overall transverse momentum is insufficient to form a LR jet. This effect is expected to happen more often for the $W_{had}$, which in general has a lower $p_T$ than the Higgs boson. This was also observed in the LR jet selection of the $W_{had}$, which implies that potential LR jet candidates of the $W_{had}$ are not reconstructed as LR jet since they fail the minimum transverse momentum requirement of $p_T > 200$ GeV. In summary, the smallR jet selection might be inefficient in semiboosted events because failing the LR jet criteria does not necessarily imply that the decay products can be successfully reconstructed as two resolved jets.

The selection reduces the $t\bar{t}$ background significantly, but since the discrepancy in cross section of the two processes is so high, further improvements are necessary to be able to distinguish between signal an background events. One possibility would be to train a boosted decision tree (BDT) based on the selected events to distinguish between the remaining signal and $t\bar{t}$ events. Another approach to improve the selection performance could be to use neural networks, in particular for the smallR jet and VBF selection which show a significant loss in signal efficiency. For example, the selection including the requirement of 4 smallR jets could be used as a preselection for a neural network. This preselection already improves the signal-to-background ratio quite a lot while maintaining a manageable event sample. The neural network potentially performs better than the cut based approach for the smallR jets, since it can exploit more variables of the smallR jets, which should help with some of the problems that were mentioned for the smallR jet matching.

# Chapter 6

# Summary

This thesis investigates Vector Boson Fusion di-Higgs production in the semileptonic $b\bar{b}W^+W^-$ final state, with a particular focus on deviations of the quartic coupling of two vector bosons to two Higgs bosons, represented by the coupling modifier $\kappa_{2V}$. The truth studies showed how deviations of $\kappa_{2V}$ to the SM case impact the kinematics. Compared to the SM case, non-SM scenarios favor a more boosted topology of the Higgs bosons, resulting in higher transverse momenta and more collimated decay products. In addition, it was shown that a semiboosted approach in which only one of the bosons is boosted is sensitive to deviation of $\kappa_{2V}$ to the SM.

Based on the truth studies, an event selection for an semiboosted approach for $\kappa_{2V} = 1.5$ is derived using the full ATLAS detector simulation. The selection is divided into two complementary reconstruction paths for the two semiboosted cases: one targeting events with a boosted $H_{b\bar{b}}$ and a resolved $W_{had}$, and the other targeting a boosted $W_{had}$ and a resolved $H_{b\bar{b}}$. In both cases the boosted particle is reconstructed as an LR jet and the resolved particle by two smallR jets. The key steps of the selection are the identification of the lepton, the selection of the LR and smallR jets and the VBF jet selection in the end. The efficiency of the selection steps are compared for signal and $t\bar{t}$ events, which is the most dominant background. The cumulative efficiency for the whole selection is 2.369 % for signal events and 0.007 % for $t\bar{t}$ events. The signal efficiency suffers mostly from losses in the smallR jet and VBF jet selection.

The main limitation of the analysis is that in the semiboosted approach the boson not reconstructed as an LR jet often cannot be reliably reconstructed as two resolved smallR jets. Potential improvements could include the use of machine learning techniques to better distinguish between the remaining signal and background events, as well as to optimize the event selection itself. For instance, the selection up to the stage of the smallR jet reconstruction could serve as a preselection for a neural network, which may outperform the cut-based approach.

# Bibliography


[1] S. Navas et al. (Particle Data Group). "Review of Particle Physics". In: *Physical Review D* 110.3 (Aug. 2024), p. 030001. ISSN: 2470-0029. DOI: 10.1103/PhysRevD.110.030001.

[2] S. Agostinelli et al. "Geant4—a simulation toolkit". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(03)01368-8.

[3] The ATLAS Collaboration et al. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08003–S08003. ISSN: 1748-0221. DOI: 10.1088/1748-0221/3/08/S08003.

[4] ATLAS Collaboration. "Electron and photon efficiencies in LHC Run 2 with the ATLAS experiment". In: (2023). DOI: https://doi.org/10.48550/arXiv.2308.13362.

[5] ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider: A Description of the Detector Configuration for Run 3". In: (2023). DOI: https://doi.org/10.48550/arXiv.2305.16623.

[6] Christian Bierlich et al. "A comprehensive guide to the physics and usage of PYTHIA 8.3". In: (2022). DOI: https://doi.org/10.48550/arXiv.2203.11601.

[7] Fady Bishara, Roberto Contino, and Juan Rojo. "Higgs pair production in vector-boson fusion at the LHC and beyond". In: *The European Physical Journal C* 77.7 (July 2017). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-017-5037-9.

[8] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The anti-$k_t$ jet clustering algorithm". In: (2008). DOI: https://doi.org/10.48550/arXiv.0802.1189.

[9] ATLAS Open Data. URL: https://opendata.atlas.cern/docs/documentation/introduction/SM_and_beyond.

[10] David J Griffiths. *Introduction to elementary particles, 2nd rev. version*. Wiley, 2008.

[11] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Physical Review Letters* 13.16 (Oct. 1964), pp. 508–509. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.13.508.

[12] The ALEPH Collaboration et al. "Precision Electroweak Measurements on the Z Resonance". In: (2005). DOI: https://doi.org/10.48550/arXiv.hep-ex/0509008.

[13] The ATLAS Collaboration. "Identification of boosted, hadronically decaying W bosons and comparisons with ATLAS data taken at $\sqrt{s} = 8$ TeV". In: *The European Physical Journal C* 76.3 (Mar. 2016). ISSN: 1434-6052. DOI: https://doi.org/10.1140/epjc/s10052-016-3978-z.

[14] The ATLAS Collaboration. "Measurement of the W-boson mass and width with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 7TeV$". In: *The European Physical Journal C* 84.12 (Dec. 2024). ISSN: 1434-6052. DOI: https://doi.org/10.1140/epjc/s10052-024-13190-x.

[15] The ATLAS Collaboration. "Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s}$ = 13 TeV". In: *The European Physical Journal C* 81.7 (2021). ISSN: 1434-6052. DOI: https://doi.org/10.1140/epjc/s10052-021-09233-2.

[16] The ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: (2012). DOI: https://doi.org/10.48550/arXiv.1207.7214.

[17] The ATLAS Collaboration. "Search for pair production of boosted Higgs bosons via vector-boson fusion in the $b\bar{b}b\bar{b}$ final state using pp collisions at s=13TeV with the ATLAS detector". In: *Physics Letters B* 858 (Nov. 2024), p. 139007. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2024.139007.

[18] The ATLAS Collaboration. *Transformer Neural Networks for Identifying Boosted Higgs Bosons decaying into $b\bar{b}$ and $c\bar{c}$ in ATLAS*. Tech. rep. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2023-021. Geneva: CERN, 2023. URL: https://cds.cern.ch/record/2866601.

[19] The CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: (2012). DOI: https://doi.org/10.48550/arXiv.1207.7235.

[20] Department of Physics University of Oxford. URL: https://www.physics.ox.ac.uk/news/why-two-higgs-are-better-one.

[21] Mira Varma and O. K. Baker. "Quantum Entanglement in Top Quark Pair Production". In: (2023). DOI: https://doi.org/10.48550/arXiv.2306.07788.

# Declaration

I hereby declare that this thesis is my own work, and that I have not used any sources and aids other than those stated in the thesis.

München, den 11. Mai 2026

Stefan Brückner