GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737

Fakultät für Physik

## Master's Thesis

# Aspects of Jet Grooming in Theory and Experiment for Boosted Topologies in the Context of the ATLAS 1-lepton $X \to HH \to b\bar{b}WW^*$ Analysis


prepared by

**Leon Stöcker**

from Dormagen

at the Institut für Theoretische Physik and II. Physikalisches Institut



**Thesis number:** II.Physik-UniGö-MSc-2023/01

**Thesis period:** 16th May 2022 until 15th May 2023

**First referee:** Prof. Dr. Stanley Lai

**Second referee:** Prof. Dr. Steffen Schumann

# Abstract

In order to address strongly collimated final states in particle detectors, it is necessary to utilize jet objects possessing high spatial resolution. This requirement can be fulfilled by incorporating tracking information in the definition of jet objects. Examples of jet objects that perform this function include Track-Assisted-Reclustered (TAR) jets and Unified Flow Object (UFO) jets.
In this thesis, TAR jets and UFO jets are compared in semi-leptonic boosted $H \to WW^*$ decays on Monte Carlo (MC) truth level for the Higgs-Strahlung process $Z^* \to ZH \to b\bar{b}WW^*$, generated by Sherpa, and for samples of the ATLAS 1-lepton $X \to HH \to b\bar{b}WW^*$ analysis.
The jet objects are implemented in a Rivet analysis and validated. The performance of both jet objects in the discrimination of the Higgs-Strahlung process against QCD background is studied, showing miniscule differences in the separation power. The effect of clustering the lepton originating from the leptonically decaying $W$ boson into the jet of the hadronically decaying $W$ boson is observed, pointing out the grave deformation of observed distributions caused by an additional lepton in the jet, if no technique is applied to retroactively remove it. It is further shown that the jet objects are stable under systematic variations.
After providing similar performance of the jet objects for the Higgs-Strahlung process, the Sherpa samples are compared to ATLAS $X \to HH \to b\bar{b}WW^*$ samples, which include a detector simulation. The findings for the Higgs-Strahlung process are confirmed for the ATLAS samples, suggesting the analysis strategy built around TAR jets also works for UFO jets.
Furthermore, a new version of $V$+jets background samples is validated for the use in the $X \to HH(/SH) \to b\bar{b}WW^{(*)}$ analyses. Explorations of the MC modelling to data in a newly constructed $Z$+jets control region show that the background modelling would not benefit from a scaling of the $Z$+jets background.



# Zusammenfassung

Um stark kollimierte Endzustände in Teilchendetektoren untersuchen zu können, müssen Jet-Objekte mit hoher räumlicher Auflösung verwendet werden. Diese Anforderung kann durch die Einbeziehung von Tracking-Informationen in die Definition von Jet-Objekten erfüllt werden. Beispiele für Jet-Objekte, die diese Funktion erfüllen, sind Track-Assisted-Reclustered (TAR) Jets und Unified Flow Object (UFO) Jets.
In dieser Arbeit werden TAR Jets und UFO Jets in semi-leptonischen, geboosteten $H \to WW^*$ Zerfällen auf Monte Carlo (MC) Wahrheitsniveau für den Higgs-Strahlungsprozess $Z^* \to ZH \to b\bar{b}WW^*$, generiert durch Sherpa, und für Simulationen der ATLAS 1-Lepton $X \to HH \to b\bar{b}WW^*$ Analyse verglichen.
Die Jet-Objekte werden in einer Rivet-Analyse implementiert und validiert. Die Fähigkeit der beiden Jet-Objekte zur Unterscheidung des Higgs-Strahlungsprozesses vom QCD-Untergrund wird untersucht, wobei sich nur minimale Unterschiede in der Separationsfähigkeit zeigen. Der Effekt des Clusterns des Leptons, das aus dem leptonisch

zerfallenden $W$-Boson stammt, in den Jet des hadronisch zerfallenden $W$-Bosons wird beobachtet. Dabei werden gravierende Deformationen der beobachteten Verteilungen durch das zusätzliche Lepton sichtbar, solange keine Methoden angewendet werden, um es nachträglich zu entfernen. Es wird außerdem gezeigt, dass die Jet-Objekte unter systematischen Variationen stabil bleiben.
Nachdem eine ähnliche Leistung der Jet-Objekte für den Higgs-Strahlungsprozess gezeigt wurde, werden die SHERPA-Simulationen mit ATLAS $X \to HH \to b\bar{b}WW^*$ Simulationen, die eine Detektorsimulation enthalten, verglichen. Die Ergebnisse für den Higgs-Strahlungsprozess werden für die ATLAS Simulationen bestätigt, was darauf hindeutet, dass die für TAR Jets entwickelte Analysestrategie, auch für UFO Jets funktioniert.
Außerdem wird eine neue Version der $V$+Jets-Untergrundsimulation für die Verwendung in der $X \to HH(/SH) \to b\bar{b}WW^{(*)}$ Analysen validiert. Untersuchungen von MC Simulationen gegenüber Daten in einer neu konstruierten $Z$+Jets-Kontrollregion ergeben, dass die Untergrundmodellierung nicht von einer Skalierung des $Z$+Jets-Untergrundes profitiert.

# Contents

# 1. Introduction

> *It's a warm summer evening, circa 600 BC. You've finished your shopping at the local market, or agora ... and you look up at the night sky. There you notice some of the stars seem to move, so you name them planētēs or wanderer. [...] This is the beginning of a 2,600-year journey we're going to take together from the ancient Greeks through Isaac Newton to Niels Bohr to Erwin Schrödinger to the Dutch researchers that Leonard is currently ripping off.*
>
> Sheldon Lee Cooper (*The Big Bang Theory*)

Over the last century, physics has made a quantum leap with the discovery of an overwhelming amount of new particles and formulation of theories that would later make up the Standard Model [1, 2]. The Standard Model of particle physics is a widely accepted and experimentally tested theoretical framework [3]. Despite its success, the Standard Model is not a complete theory of the universe, as it only describes three out of four fundamental interactions. In addition, the Standard Model cannot account for the existence of dark matter, the imbalance between matter and antimatter, nor neutrino oscillations and other observed phenomena [4–6]. This indicates that there must be new physics beyond the Standard Model, which could include very heavy, currently unobserved particles. Physicists are actively searching for such new particles.

In the decay of heavy particles into much lighter decay products, the decay products are heavily boosted. Subsequent decays of the boosted particles create a highly collimated final state, for which small radius jets, seeded by individual quarks, cannot be resolved individually and are clustered into large radius jets. In experiments, purely calorimeter-based jets lack the spatial resolution to resolve the substructure of these large radius jets. For this reason, sophisticated jet objects need to be introduced that gain spatial resolution by taking Inner Detector tracks into account, which feature a much higher granularity than coarse calorimeter cells.

Track-Assisted-Reclustered (TAR) jets [7] and Unified Flow Object (UFO) jets [8] are examples of such jet objects, which also feature grooming techniques to clean them up from background noise and further improve their precision. As the ATLAS collaboration discourages the use of TAR jets in favour of UFO jets for Run 3, analyses built around

TAR jets have to verify the effectiveness of the currently deployed strategy for UFO jets. This thesis presents a comparison of TAR jets and UFO jets in semi-leptonic boosted $H \to WW^*$ decays on Monte Carlo (MC) truth level for the Higgs-Strahlung process $Z^* \to ZH \to b\bar{b}WW^*$ and in the context of the ATLAS 1-lepton $X \to HH \to b\bar{b}WW^*$ analysis [9]. In addition, other studies are performed on background samples of the ATLAS analysis.

An overview of the theoretical background of the Standard Model with two relevant extensions beyond the Standard Model, as well as principles of MC simulation and the topology of the analysis is given in Chapter 2. Chapter 3 describes the experimental setup of the LHC, the ATLAS detector, and the MC generator settings used for the production of the samples used in this thesis. The used observables and jet objects are introduced in Chapter 4. Chapter 5 presents the results obtained for the studies on truth level with the Higgs-Strahlung process, while Chapter 6 addresses the findings in the context of the ATLAS analysis. Finally, the results are summarized and discussed in Chapter 7.

# 2. Theoretical Background

This chapter describes the theoretical foundations of this thesis. It starts with the Standard Model (SM) [1, 2], a quantum field theory (QFT) describing three of the four known fundamental forces in the universe. The Higgs boson was the last experimentally observed particle in the SM [10, 11], and its mechanism, which gives mass to particles, as well as extensions of the Higgs sector will be treated with special care. As a concrete example resulting from possible Higgs sector extensions and context of the studies in this thesis, the topologies of $X \to HH(/SH) \to b\bar{b}WW^{(*)}$ will be discussed, after highlighting the concepts of MC event generation.

## 2.1. The Standard Model

There are four fundamental interactions in the universe: gravitational, electromagnetic, weak and strong interactions. Einstein's General Theory of Relativity [12] describes gravitational interactions, while the three other interactions were combined over the course of the last decades into the SM. The SM does not only describe three of the fundamental interactions, but also all known elementary particles. As the SM is a quantum field theory, its particles are described as excitations of such quantum fields. After describing the particle content of the SM, this section will discuss the associated SM Lagrangian densities. These Lagrangian densities already allow an interpretation of how the fields interact. They can further be used to derive the actual equations of motions of the fields. Richard Feynman came up with an elegant, graphical representation, that simplifies the calculation of cross sections of physical processes dramatically [13, 14]. A more detailed description of QFT, particle physics and the SM can be found in Refs. [1, 2].

### 2.1.1. Particle Content

The particles of the SM can be classified in two main categories, each containing subcategories: fermions with spin-1/2 and 5 bosons with integer spin. The twelve fermions make up all the surrounding matter, while the five bosons act as the mediators of the

three fundamental interactions and the Higgs mechanism. A depiction of all particles in the SM and some of their properties can be seen in Figure 2.1.

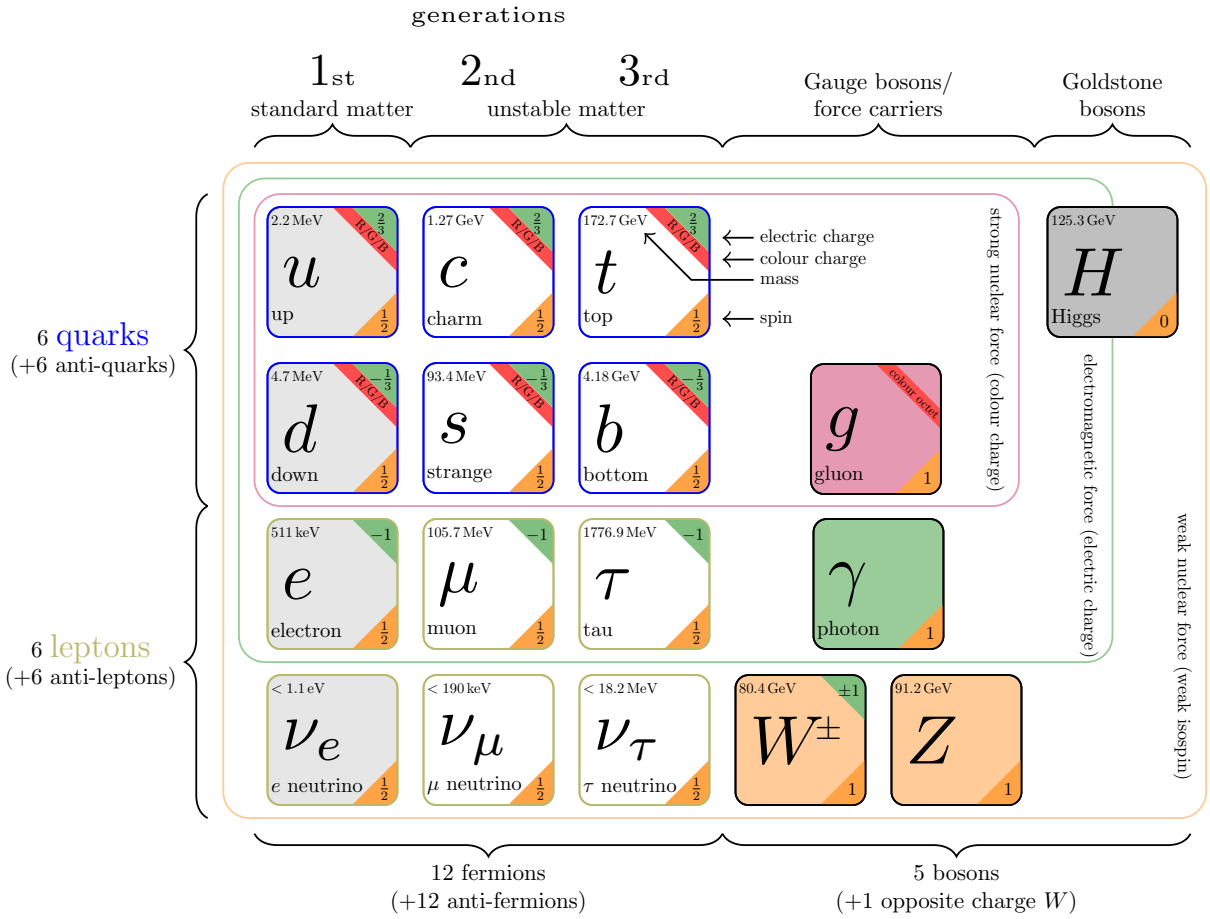


**Figure 2.1.:** Fundamental particles of the SM [1]. Mass values are taken from Ref. [3].

The fermions themselves can be divided into groups of leptons and quarks. In addition, the fermions are divided into three generations. Each generation contains two types of leptons and two types of quarks. One of the leptons has electric charge $-1$, while the other one remains neutrally charged and is therefore called the neutrino. The quarks on the other hand are classified into an up-type quark and a down-type quark. Up-type quarks have an electric charge of $+2/3$, while the down-type quarks have a charge of $-1/3$.
The quarks further carry the charge of the strong interaction, called colour. Colour charge comes in the three distinct values of red, green and blue. Despite the name of this charge, it has nothing to do with the colour our eyes perceive. The charges of the down-type quarks, up-type quarks, leptons and neutrinos remain the same across the

[1] Figure adapted from https://texample.net/tikz/examples/model-physics (visited on Mar. 5,2023) (CC BY 2.5)

three generations, however the mass and flavour changes.
The quarks are called up ($u$), down ($d$), charm ($c$), strange ($s$), top ($t$) and bottom ($b$) while the leptons ($\ell$) are called electron ($e$), muon ($\mu$) and tau ($\tau$), each with an associated neutrino ($\nu_\ell$). These fermions are arranged in generations of increasing mass such that particles from the second and third generation decay into particles of the first (and second) generation. This is the reason why the nucleons (protons and neutrons) of all atoms are made up of up and down quarks, surrounded by electrons.
The fermions form doublets and singlets of weak isospin. For left-chiral fermions, the fermion generations form doublets where up-type quarks and neutrinos have a third component of weak isospin $T_3 = 1/2$, while down-type quarks and charged leptons have a third component of weak isospin $T_3 = -1/2$. Right-chiral fermions have $T = T_3 = 0$ and form weak isospin singlets.
Aside from the fermions, the SM also describes antifermions, or more generally antiparticles. Antiparticles have the same mass as their associated particle, but the opposite charge. In particular right-chiral antifermions have $T = 1/2$, while left-chiral antifermions carry $T = 0$.
Four of the five bosons are spin-1 gauge bosons, they are needed to restore the invariance of the Lagrangian of the theory under local transformations (more details in the following subsections) and act as the mediators of their associated force. Namely, the gauge bosons are the gluon ($g$) (the exchange particle of strong interactions and appearing in eight colour combinations), the $W^-$ boson (exchange particle of charged weak interactions with its antiparticle $W^+$), the $Z$ boson (exchange particle of neutral weak interactions) and the photon (exchange particle of electromagnetic interactions). The fifth boson, the Higgs ($H$) boson, is spin-0 Goldstone boson. Goldstone bosons arise when a symmetry is spontaneously broken (see Section 2.2). The Higgs field gives the massive particles in the SM their mass, with neutrinos being the exception. The source of neutrino masses is a currently open question.

### 2.1.2. Quantum Electrodynamics

The theory of the SM can be divided and studied in parts. Although not describing the whole picture, it can be beneficial to look at such smaller parts to understand the techniques used in the complete theory. A relatively straightforward example is Quantum Electrodynamics (QED). QED describes the dynamics of a fermionic field $\psi$ with a photon $A_\mu$. The Lagrangian density of a free massive fermionic field with global U(1) symmetry

$$\mathscr{L} = \overline{\psi}(\mathrm{i}\not{\partial} - m)\psi \tag{2.1}$$

is invariant under the gauge transformation

$$\psi \to \psi' = e^{\mathrm{i}\alpha}\psi, \tag{2.2}$$

where $m$ is the fermion mass, $\not{\partial} = \gamma^\mu \partial_\mu$ and $\overline{\psi} = \psi^\dagger \gamma^0$ is the adjoint spinor with $\gamma^\mu$ denoting the gamma matrices. If one allows local transformations $\alpha \to \alpha(x)$, the Lagrangian density is no longer invariant. In order to restore the invariance, a massless gauge field $A_\mu(x)$, the photon, is added. The gauge field transforms as

$$A_\mu(x) \to A'_\mu(x) = A_\mu(x) - \frac{1}{g_e}\partial_\mu \alpha(x), \tag{2.3}$$

where the coupling strength $g_e$ is added for later convenience. By combining $A_\mu$ with the derivative into a covariant derivative $D_\mu$:

$$D_\mu \psi = (\partial_\mu + \mathrm{i} g_e A_\mu)\psi, \tag{2.4}$$

the Lagrangian density is again invariant under gauge transformations. This can be seen by evaluating the transformation of the differentiated object $D_\mu \psi$:

$$D_\mu \psi \to (D_\mu \psi)' = e^{\mathrm{i}\alpha(x)} D_\mu \psi, \tag{2.5}$$

where the phase $e^{i\alpha(x)}$ is cancelled by the transformed adjoint spinor. Finally, a kinetic term for the gauge field can be added, to obtain a Lagrangian density

$$\begin{aligned} \mathscr{L}_{\mathrm{QED}} &= -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \overline{\psi}(\mathrm{i}\not{D} - m)\psi, \\ &= \underbrace{-\frac{1}{4} F_{\mu\nu} F^{\mu\nu}}_{\text{kinetic gauge term}} + \underbrace{\overline{\psi}(\mathrm{i}\not{\partial} - m)\psi}_{\text{free fermion}} + \underbrace{g_e \overline{\psi} \not{A} \psi}_{\substack{\text{fermion-photon} \\ \text{interaction}}} \end{aligned} \tag{2.6}$$

where $F_{\mu\nu} = -\frac{\mathrm{i}}{g_e}[D_\mu, D_\nu] = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the field strength tensor. Gauge fields are needed to ensure invariance of the Lagrangian density under local gauge transformations. This results in interaction terms between the gauge field and the fermion field. For non-Abelian groups, whose elements do not commute, the commutator of the covariant derivative in Equation (2.4) introduces self-interactions terms of the gauge field.

### 2.1.3. Quantum Chromodynamics and Electroweak Unification

The example of QED does not describe the complete physics of the electromagnetic interaction in the SM and is therefore not directly included in the full theory. The SM contains the internal symmetries of the unitary product group

$$\underbrace{\mathrm{SU}(3)_C}_{\text{colour}} \times \underbrace{\mathrm{SU}(2)_W}_{\text{weak isospin}} \times \underbrace{\mathrm{U}(1)_Y}_{\text{hypercharge}} , \tag{2.7}$$

where the indices stand for the charges denoting the symmetry of the respective groups. Only (anti-)particles with a weak isospin of $T = 1/2$ can interact with the $W$ bosons. The weak hypercharge $Y_W$ is defined as

$$Y_W = 2(Q - T_3), \tag{2.8}$$

where $Q$ is the electric charge and $T_3$ is the third component of weak isospin. The gauge field corresponding to $\mathrm{U}(1)_Y$ is denoted by $B_\mu$, the three fields $W_\mu^{1,2,3}$ correspond to $\mathrm{SU}(2)_W$ and the octet of fields $G_\mu^{1,\dots,8}$ belong to $\mathrm{SU}(3)_C$. The $G_\mu^a$ are the eight gluon fields, transmitting the strong nuclear force, where each of the gluons carry a different colour charge. The field strength tensors of these fields are given by

$$\begin{aligned} G^a_{\mu\nu} &= \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g_s f^{abc} G^b_\mu G^c_\nu \\ W^a_{\mu\nu} &= \partial_\mu W^a_\nu - \partial_\nu W^a_\mu - g\varepsilon^{abc} W^b_\mu W^c_\nu \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu, \end{aligned} \tag{2.9}$$

where $g_s$ and $g$ are the coupling constants of the respective fields. The structure constants $f^{abc}$ and $\varepsilon^{abc}$ arise from the commutation relations of the generators of the respective group and introduce a self-coupling of gluons and $W_\mu^a$ fields. The generators of $\mathrm{SU}(2)_W$ are usually chosen to be proportional to the Pauli matrices $\sigma^a$, where the generators of $\mathrm{SU}(3)_C$ are proportional to the eight Gell-Mann matrices $\lambda^a$.

The fields $B_\mu$ and $W_\mu^a$ do not represent the physical fields of photons, $W$ and $Z$ bosons, rather the physical fields are a combination of the gauge fields. The photon and $Z$ boson fields are related to the $B$ and $W^3$ field via the weak mixing angle $\theta_W$

$$\begin{pmatrix} A \\ Z \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B \\ W^3 \end{pmatrix}. \tag{2.10}$$

The physical $W$ bosons are described by

$$W^{\pm} = \frac{1}{\sqrt{2}} \left( W^1 \mp \mathrm{i} W^2 \right). \tag{2.11}$$

The description of weak and electromagnetic interactions in the same theory is called electroweak (EW) unification. For the Lagrangian density of the EW sector, it is necessary to distinguish fermion fields of left- and right-chirality, since the $W$ boson only couples to left-chiral particles and right-chiral antiparticles. The covariant derivatives acting on the right-chiral fields therefore only include the field $B_\mu$

$$D_\mu \psi_R = \partial_\mu + \mathrm{i} \frac{g'}{2} Y_W B_\mu, \tag{2.12}$$

while the covariant derivatives acting on the left-chiral fields also include the $W^a_\mu$

$$D_\mu \psi_L = \partial_\mu + \mathrm{i} \frac{g}{2} W_\mu + \mathrm{i} \frac{g'}{2} Y_W B_\mu, \quad \text{where } W_\mu = \sum_{a=1}^{3} \sigma^a W^a_\mu. \tag{2.13}$$

The fermions can be grouped into weak isospin left-chiral doublets with $T = 1/2$: $q_L$ for the quarks and $l_L$ for the leptons, and isospin singlets with $T = 0$: $u_R$, $d_R$ for the up- and down-type quarks and $\ell_R$ for the electrically charged leptons. It is then possible to split the Lagrangian density of the EW sector into contributions with quarks, leptons and gauge kinetic terms

$$\mathscr{L}_{\mathrm{EW}} = \mathscr{L}_q + \mathscr{L}_l + \mathscr{L}_{\text{kinetic gauge}}. \tag{2.14}$$

The respective parts are

$$\mathscr{L}_q = \mathrm{i} \bar{q}_L \not{D} q_L + \mathrm{i} \bar{u}_R \not{D} u_R + \mathrm{i} \bar{d}_R \not{D} d_R, \tag{2.15}$$

$$\mathscr{L}_l = \mathrm{i} \bar{l}_L \not{D} l_L + \mathrm{i} \bar{\ell}_R \not{D} \ell_R, \tag{2.16}$$

$$\mathscr{L}_{\text{kinetic gauge}} = -\frac{1}{4} \sum_{a=1}^{3} W^{\mu\nu}_a W^a_{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}. \tag{2.17}$$

The remaining dynamics of the SM are those associated with the $\mathrm{SU}(3)_C$ group. As they are the dynamics of particles carrying a colour charge, the respective theory is called Quantum Chromodynamics (QCD). Very similar to the QED Lagrangian density in Equation (2.6), the Lagrangian density of QCD is given by

$$\mathscr{L}_{\mathrm{QCD}} = -\frac{1}{4} G^{\mu\nu} G_{\mu\nu} + \bar{q} \mathrm{i} \not{D}^{\mathrm{QCD}} q. \tag{2.18}$$

The components of the covariant derivatives are

$$\left(D_\mu^{\mathrm{QCD}}\right)_{ij} = \partial_\mu \delta_{ij} - i g_s \lambda_{ij}^a G_\mu^a. \tag{2.19}$$

So far the Lagrangian densities of QCD and the EW theory do not include any mass terms. For gauge bosons, any mass term of the form

$$-m_B^2 B_\mu B^\mu \tag{2.20}$$

would not be invariant under a gauge transformation. Such terms are therefore forbidden, and gauge bosons must be massless. Similarly, direct mass terms of fermions

$$-m_f \overline{\psi}\psi = -m_f(\overline{\psi}_R \psi_L + \overline{\psi}_L \psi_R) \tag{2.21}$$

are forbidden, since the SU(2) symmetry changes the flavour and therefore the mass of a fermion. Such mass terms would consequently not be invariant. The observed masses of all particles of the SM are gained through the Higgs mechanism, which is described in the next section.

## 2.2. The Higgs Mechanism

As mentioned in the prior section, none of the gauge bosons nor fermions "start" with a mass. Their mass is only acquired through the Higgs mechanism, which was developed by Brout, Englert, Guralnik, Hagen, Higgs and Kibble [15–20]. While the SM with the Higgs mechanism celebrates great successes in the description of experimentally observed processes, the lack of gravitation, explanation for dark matter, neutrino masses and matter-antimatter asymmetry make it very clear that the SM is not complete. As will be shown in Section 2.2.1, the Higgs boson couples to particles proportionally to their mass. It is therefore the perfect candidate in the search for new heavy particles. Such new particles could arise from an extended Higgs sector. A few possibilities of such an extended Higgs sector will be introduced in Section 2.2.2.

### 2.2.1. The Standard Model Higgs Boson

The Higgs boson is associated with a complex scalar field $\Phi$, which is a doublet of $\mathrm{SU}(2)_W \times \mathrm{U}(1)_Y$. The Lagrangian density of such a scalar is given by

$$\mathscr{L}_{\mathrm{Higgs}} = (D^\mu \Phi)^\dagger (D_\mu \Phi) - V(\Phi^\dagger \Phi), \tag{2.22}$$

where the covariant derivative takes the same form as Equation (2.13) and the Higgs potential is

$$V(\Phi^\dagger\Phi) = \mu^2\Phi^\dagger\Phi + \lambda\left(\Phi^\dagger\Phi\right)^2. \tag{2.23}$$

A simplified version of the potential can be seen in Figure 2.2, with $\lambda > 0$ and $\mu^2 < 0$.

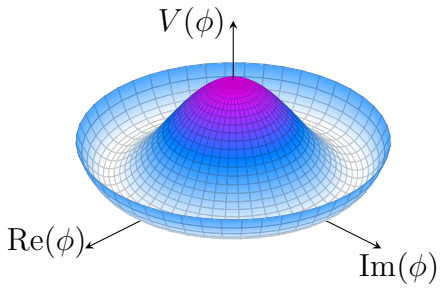


**Figure 2.2.:** Simplified Higgs Potential $V(\Phi) = \mu^2(\Phi^*\Phi) + \lambda(\Phi^*\Phi)^2$ for a complex scalar field $\Phi$ with $\lambda > 0$ and $\mu^2 < 0$.

For such a configuration the potential has an infinite set of minima around the origin at $\Phi^\dagger\Phi = v^2 \equiv -\mu^2/\lambda$, where $v$ is the vacuum expectation value. This vacuum expectation $v \neq 0$ is essentially the reason why the Higgs provides mass to particles. The Higgs field settles in this set of minima, and it is possible to arbitrarily choose

$$\langle\Phi\rangle = \begin{pmatrix} 0 \\ v \end{pmatrix}. \tag{2.24}$$

Expanding the field around this minimum

$$\Phi = \frac{1}{\sqrt{2}}\begin{pmatrix} \theta_1(x) + \mathrm{i}\theta_2(x) \\ v + H(x) + \mathrm{i}\theta_3(x) \end{pmatrix} \tag{2.25}$$

yields three massless Goldstone bosons $\theta_a(x)$ and one massive Higgs boson $H(x)$. The Goldstone bosons can be eliminated by a gauge transformation, referred to as unitary gauge:

$$\Phi = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}. \tag{2.26}$$

In the unitary gauge the Lagrangian density in Equation (2.22) becomes

$$\begin{aligned}\mathscr{L}_{\text{Higgs}} =& \frac{1}{2}\partial_\mu H \partial^\mu H - \lambda v^2 H^2 - \lambda v H^3 - \frac{1}{4}\lambda H^4 \\ &+ \frac{1}{8}(v+H)^2 \left(W_\mu^1, W_\mu^2, W_\mu^3, B_\mu\right) \begin{pmatrix} g^2 & 0 & & \\ 0 & g^2 & 0 & \\ & & g^2 & gg' \\ & 0 & gg' & g'^2 \end{pmatrix} \begin{pmatrix} W_\mu^1 \\ W_\mu^2 \\ W_\mu^3 \\ B_\mu \end{pmatrix}. \end{aligned} \tag{2.27}$$

Here the first line describes a free, massive scalar with cubic and quartic self-interactions. The interaction strength is proportional to $\lambda$. The last line contains the interaction of the $W_\mu^a$ and $B_\mu$ fields with the scalar field and mass terms for the bosons. Identifying the physical bosons as in Equation (2.10) and Equation (2.11), the Lagrangian density yields a massless photon and massive $W^\pm$, $Z$ and Higgs boson $H$, where the masses are

$$m_H = \sqrt{2\lambda v^2},\ m_W = \frac{gv}{2},\ m_Z = \frac{v}{2}\sqrt{g^2+g'^2} \text{ and } m_A = 0. \tag{2.28}$$

The couplings of the Higgs boson to the gauge bosons resulting from Equation (2.27) are proportional to the squared mass of the gauge bosons. The Higgs boson can couple, for instance, to the $W$ boson in interactions including only one Higgs boson $HW^+W^-$ or two Higgs bosons $HHW^+W^-$. The corresponding part of the Lagrangian density can be expressed as

$$\mathscr{L}_{HW} = \frac{g^2 v}{2} W_\mu^- W^{+\mu} H + \frac{g^2}{4} W_\mu^- W^{+\mu} H^2 = \frac{2m_W^2}{v} W_\mu^- W^{+\mu} H + \frac{m_W^2}{v^2} W_\mu^- W^{+\mu} H^2. \tag{2.29}$$

The Higgs mechanism described here is also called spontaneous symmetry breaking. The Lagrangian density in Equation (2.27) has only a $\mathrm{U}(1)_Y$ symmetry. It seems like three gauge symmetries from the $\mathrm{SU}(2)_W$ group were lost. In reality, the symmetries were only hidden by the choice of the unitary gauge. The degrees of freedom of the theory where simultaneously reshuffled. The three massless Goldstone bosons were absorbed to give the $W^\pm$ and $Z$ boson their mass.

The Higgs mechanism also generates masses for the fermions. Just like the fermions, the Higgs field is a doublet under SU(2) and can therefore fix the invariance of direct mass terms in Equation (2.21). Yukawa couplings of the Higgs doublet to two fermions turn once again into mass terms when the electroweak symmetry is spontaneously broken. Since the Higgs doublet in unitary gauge only projects onto the $T_3 = -1/2$ part of the fermions' weak isospin doublets, masses for down-type quarks together with neutral

leptons and up-type quarks together with charged leptons are generated in separate terms. For down-type quarks and charged leptons the Lagrangian density can be written as

$$\begin{aligned}\mathscr{L}_d &= -\lambda_f \left(\overline{\psi}_L \Phi \psi_R + \overline{\psi}_R \Phi^\dagger \psi_L\right) \\ &= -\frac{\lambda_f}{\sqrt{2}} \left( \begin{pmatrix} \overline{U} & \overline{D} \end{pmatrix}_L \begin{pmatrix} 0 \\ v+H \end{pmatrix} \begin{pmatrix} D \end{pmatrix}_R + \begin{pmatrix} \overline{D} \end{pmatrix}_R \begin{pmatrix} 0 & v+H \end{pmatrix} \begin{pmatrix} U \\ D \end{pmatrix}_L \right) \\ &= -\frac{\lambda_f v}{\sqrt{2}} \left(\overline{D}_L D_R + \overline{D}_R D_L\right) - \frac{\lambda_f}{\sqrt{2}} H \left(\overline{D}_L D_R + \overline{D}_R D_L\right).\end{aligned} \tag{2.30}$$

The first term can be identified as the mass term with $m_f = \lambda_f v/\sqrt{2}$, while the second term describes the interaction of fermionic and Higgs fields with an interaction strength proportional to $m_f$. Masses for up-type quarks and neutral leptons are generated similarly, where the conjugate scalar doublet is used to project onto the $T_3 = +1/2$ part of the fermions' weak isospin doublets. The Lagrangian density reads

$$\mathscr{L}_u = -\frac{\lambda_f v}{\sqrt{2}} \left(\overline{U}_L U_R + \overline{U}_R U_L\right) - \frac{\lambda_f}{\sqrt{2}} H \left(\overline{U}_L U_R + \overline{U}_R U_L\right). \tag{2.31}$$

### 2.2.2. Beyond the Standard Model Extensions

The analysis in this thesis studies resonant di-Higgs production, where the scalar resonance $X$ can also produce another scalar particle $S$ in its decay: $X \to HH$ or $X \to SH$. The addition of these scalar particles via two theories beyond the Standard Model (BSM) are briefly described. Since Higgs signal rate measurements are in good agreement with the SM predictions so far, theories leaving the SM Higgs boson basically unchanged are of great interest.

#### Two Real Singlet Model

The two real singlet model (TRSM) [21] extends the Higgs sector of the SM by introducing two additional real scalar singlet fields $S$ and $X$. They are symmetric under the $\mathbb{Z}_2$ symmetries

$$\begin{aligned}&\mathbb{Z}_2^S : \; S \to -S, \; X \to X, \; \text{SM} \to \text{SM} \quad \text{and} \\ &\mathbb{Z}_2^X : \; X \to -X, \; S \to S, \; \text{SM} \to \text{SM}.\end{aligned} \tag{2.32}$$

The most general scalar potential that is invariant under $\mathbb{Z}_2^S \otimes \mathbb{Z}_2^X$ and renormalizable is

$$\begin{aligned} V =& \mu_\Phi^2 \Phi^\dagger \Phi + \lambda_\Phi \left(\Phi^\dagger \Phi\right)^2 + \mu_S^2 S^2 + \lambda_S S^4 + \mu_X^2 X^2 + \lambda_X X^4 \\ &+ \lambda_{\Phi S} \Phi^\dagger \Phi S^2 + \lambda_{\Phi X} \Phi^\dagger \Phi X^2 + \lambda_{SX} S^2 X^2, \end{aligned} \tag{2.33}$$

with a total of nine real parameters. This potential extends the SM Higgs Potential in Equation (2.23). In the unitary gauge, the fields can be expressed as

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + \phi_H \end{pmatrix}, \quad S = \frac{v_S + \phi_S}{\sqrt{2}}, \quad X = \frac{v_X + \phi_X}{\sqrt{2}}, \tag{2.34}$$

where $v$ is the SM vacuum expectation value. For $v_S \neq 0$ and $v_X \neq 0$, the $\mathbb{Z}_2$ symmetries are broken and all fields $\phi_{H,S,X}$ mix into three physical scalar states. The physical mass eigenstates $h_{1,2,3}$ are related to the fields through a $3 \times 3$ mixing matrix $R$ parametrized by three mixing angles $\theta_{HS}$, $\theta_{HX}$ and $\theta_{SX}$

$$\begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix} = R \begin{pmatrix} \phi_H \\ \phi_S \\ \phi_X \end{pmatrix}. \tag{2.35}$$

Fixing one of the masses $M_a \equiv m_H$ to the SM Higgs mass leaves the TRSM being described by the seven free parameters

$$M_b,\ M_c,\ \theta_{HS},\ \theta_{HX},\ \theta_{SX},\ v_S,\ v_X, \tag{2.36}$$

with $a \neq b \neq c \in \{1, 2, 3\}$. The mixing of the fields $\phi_{H,S,X}$ results in the possibility of triple Higgs couplings, where any combination of the physical states $h_{1,2,3}$ are allowed.

### Two-Higgs-Doublet Model

A similar but different approach is taken by the two-Higgs-doublet model (2HDM) [22]. Instead of introducing two real scalar singlets, a second complex scalar doublet is added. To simplify the handling of the theory, it is assumed that charge conjugation parity (CP) symmetry is conserved in the Higgs sector, CP is not spontaneously broken and that quartic terms, that are odd in either of the doublets, are removed by discrete symmetries.

The most general potential for the two doublets $\Phi_1$ and $\Phi_2$ is

$$\begin{aligned} V =& m_{11}^2 \Phi_1^\dagger \Phi_1 + m_{22}^2 \Phi_2^\dagger \Phi_2 - m_{12}^2 \left(\Phi_1^\dagger \Phi_2 + \Phi_2^\dagger \Phi_1\right) + \frac{\lambda_1}{2}\left(\Phi_1^\dagger \Phi_1\right)^2 + \frac{\lambda_2}{2}\left(\Phi_2^\dagger \Phi_2\right)^2 \\ &+ \lambda_3 \Phi_1^\dagger \Phi_1 \Phi_2^\dagger \Phi_2 + \lambda_4 \Phi_1^\dagger \Phi_2 \Phi_2^\dagger \Phi_1 + \frac{\lambda_5}{2}\left[\left(\Phi_1^\dagger \Phi_2\right)^2 + \left(\Phi_2^\dagger \Phi_1\right)\right], \end{aligned} \tag{2.37}$$

which is minimal for

$$\langle \Phi_1 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v_1 \end{pmatrix}, \ \langle \Phi_2 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v_2 \end{pmatrix}. \tag{2.38}$$

Given that the number of complex doublets is twice that of the SM, the number of fields in the extension around the vacuum expectation values is also doubled, for a total of eight fields. Just like in the SM, three of these fields are "eaten" to give mass to the $W^\pm$ and $Z$ bosons. The remaining five are physical scalar fields. Two of these are neutral scalar particles (light and heavy Higgs $h$, $H$), another two are charged scalar particles ($H^\pm$) and the last one is a neutral pseudoscalar particle ($A$). The potential in Equation (2.37) allows all the scalar particles to couple to each other.
The coupling of the doublets to fermions can introduce flavour changing neutral currents (FCNC) at tree-level, which does not agree with current observations. It is therefore necessary to only consider couplings that forbid FCNCs.

## 2.3. Monte Carlo Event Generation

Monte Carlo (MC) event generators have numerous applications in particle physics. Experiments use simulated events that obey the rules of the SM to study the shape of background and signal events in their analysis separately. This can show if the proposed signal will be detectable behind the background. This works, because unlike real data, MC studies have the advantage of knowing exactly what process belongs to which final state or detector read-out. MC event generators can be furthermore extended to include BSM physics to allow the testing of such models. Furthermore, MC events allow easy tests of newly developed algorithms and objects before using them in experimental analyses.
As the working steps of MC event generators include perturbative methods, the running coupling of QCD is explained before giving a short overview of the stages of event generators. Finally, the concept of jets is introduced. A very detailed review of general-purpose event generators can be found in Ref. [23], while Ref. [24] provides an introduction into various QCD related concepts that are relevant for high energy particle physics.

### 2.3.1. Running Coupling Strengths

Feynman diagrams [13, 14] offer a great simplification of the QFT of the SM by providing a graphical representation of mathematical expressions of particle interactions. Every process with a given set of in- and output particles can occur in multiple ways. The complete process is therefore described by the sum of all contributing Feynman diagrams. Each additional term in the summation includes a higher order in the coupling as the preceding one. As the couplings are $\ll 1$, higher order terms are also less likely than lower order ones, but still contribute to the overall cross section. Feynman diagrams are a graphical representation of the terms in the perturbative expansion of QFT.
An example of such a summation is depicted in Figure 2.3. Here, the next-to-leading order (NLO) term is divergent.

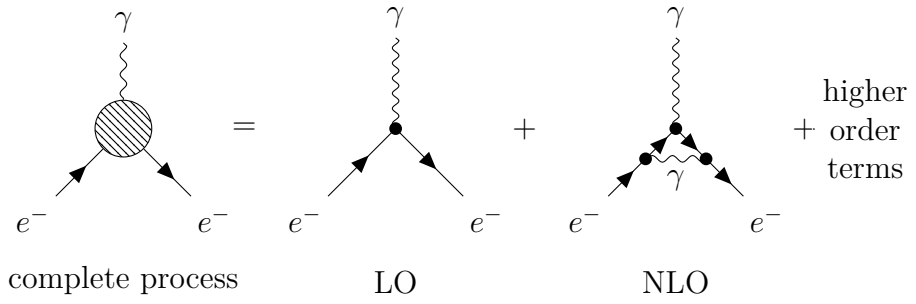


**Figure 2.3.:** Electron-photon interaction as described by Feynman diagrams. The complete process is the sum of terms of different orders in the coupling. The simplest possible order is referred to as leading order (LO), the next higher order term as next-to-leading order (NLO).

In order to obtain finite results from the divergent term, one can use regularization and renormalization. In dimensional regularization, the dimension of integrals needed in the calculation of such loop diagrams $d = 4$ is reduced to $d = 4 - \epsilon$, effectively splitting the divergences from finite terms. Renormalization introduces divergent counter-terms that cancel the divergences of the bare theory. In this series of steps, the divergences are essentially absorbed by renormalized parameters of the theory. In order to keep consistent dimensions for these parameters, an arbitrary renormalization scale $\mu$ is introduced, making the coupling dependent on this scale. For QCD the coupling $\alpha_s(\mu^2)$ is then

$$\alpha_s(\mu^2) = \frac{\alpha_s(\mu_0^2)}{1 + b_0 \alpha_s(\mu_0^2) \ln \frac{\mu^2}{\mu_0^2}} = \frac{1}{b_0 \ln \frac{\mu^2}{\Lambda^2}}, \text{ where } \quad b_0 = -\left(\frac{11}{3} N_c - \frac{4}{3} N_f\right). \tag{2.39}$$

Here $N_c$ is the number of colours, $N_f$ is the number of quark flavours and $\mu_0$ is a reference scale. The strength of QCD interactions for a process with momentum transfer $Q$ is given

by $\alpha_s(\mu)$ with $\mu \approx Q$. Measured values as well as theoretical predictions of $\alpha_s$ at different momentum transfer scales $Q$ can be seen in Figure 2.4. For increasing $Q$, the coupling decreases.

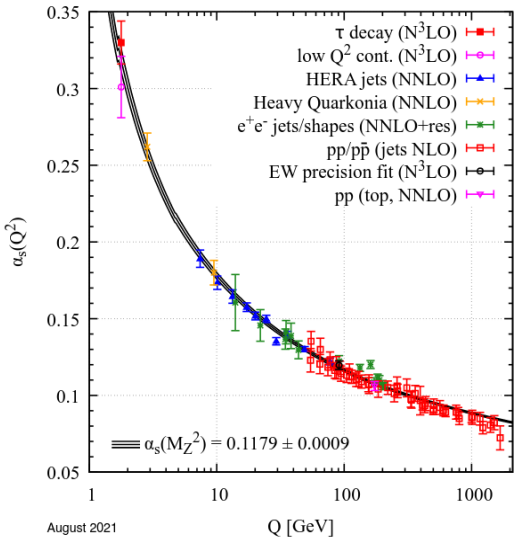


**Figure 2.4.:** Measured values for the QCD coupling $\alpha_s$ at different momentum transfer scales $Q$, together with theoretical predictions for $\alpha_s$ with the reference $\alpha_s(M_Z^2) = 0.1179$, from Ref. [3].

This results in the concepts of asymptotic freedom and confinement. Asymptotic freedom describes the phenomenon that at large energies (small distances) the coupling of gluons to quarks becomes almost negligible and the quarks can be seen as free particles. Confinement describes the inverse phenomenon. At small energies (large distances) quarks and gluons cannot exist in unbound states. They only exist in colourless objects called hadrons. Hadrons are composed of multiple partons, a general term for quarks and gluons, which each carry fractions of the hadrons total momentum [25]. At low energies, baryons consist of three valence quarks and mesons consist of two valence quarks. With increasing energies, sea partons become visible in addition to the valence quarks [26]. The inner structure of hadrons are described by parton distribution functions (PDFs).
The coupling of QED also depends on the renormalization scale. The behaviour is the opposite as for QCD. For increasing energy scales, the electromagnetic coupling also increases.

### 2.3.2. Principles of Monte Carlo Event Generation

A simulated event can be separated into different steps along the evolution of incoming hadrons towards the final state particles. The whole process is depicted in Figure 2.5.

The colours of the specific parts in this figure will be described in the following part, connecting the written explanation and the figure.

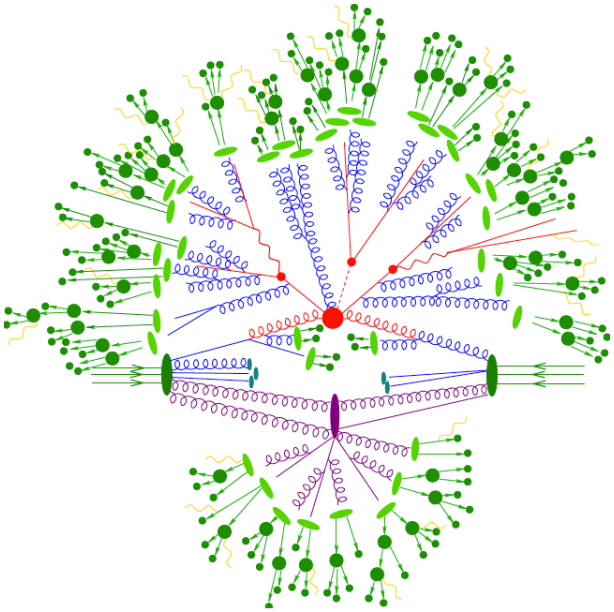

**Figure 2.5.:** Sketch of a hadron-hadron collision as simulated by MC event generators. The incoming hadrons enter from the left and the right. The hard process is depicted in red, initial and final state Bremsstrahlung in blue, parton-hadron transitions in light green, hadrons and hadron decays in dark green and electromagnetic radiation in yellow. The purple part represents multiple parton interactions. Figure adapted from Ref. [27].

When simulating events, rather than waiting for the incoming protons to react in the (eventually very uncommon) type of way that is required for a specific analysis (e.g. $gg \to X \to HH$), the simulation is built around this hard subprocess (red) from the beginning. This saves an incredible amount of time and computing resources by not simulating unwanted events. In order to simulate events, the generator takes PDFs into account, which describe the inner quark-gluon-structure of protons, and parton level matrix elements, which have to be integrated over the corresponding phase space. Generators often rely on the Feynman rules to calculate these matrix elements. Since the integrations are performed over a high number of dimensions, MC integration becomes the only feasible numerical integration method, as its accuracy improves with $\sim 1/\sqrt{N}$, where $N$ is the number of integration points, regardless of the number of dimensions.
After the hard process has been simulated, the QCD particles that leave this hard process can radiate gluons, which again can radiate gluons or create $q\overline{q}$ pairs (blue). The initial outgoing QCD particle is the start of a parton shower which is simulated with a step-wise Markov chain. Parton showers are based on perturbation theory. They are only valid as

long as the coupling $\alpha_s \ll 1$. As more and more partons are added to the final state, the energy of each parton decreases from the initial hard parton down to energies as low as 1 GeV, where perturbation theory breaks down. Since the particles entering the hard process are typically QCD partons as well, parton shower algorithms can also be used to simulate initial state radiation (also blue, before the hard interaction).
The perturbative evolution has to be stopped and replaced by non-perturbative models that describe the confinement of the current coloured state into colourless hadrons (light green). Hadronization cannot be calculated from first principles. The most common model classes to describe the process are the string and cluster ones. The Lund string hadronization model [28, 29] is based on strings extending between $q\overline{q}$ pairs. As the quarks move away from each other, the strings break apart and form new $q\overline{q}$ pairs. The string model uses many parameters related to flavour properties which have to be tuned using data. The cluster hadronization model on the other hand uses fewer parameters. The model starts with the non-perturbative splitting of gluons that leave the parton shower into $q\overline{q}$ pairs. Clusters are then formed from colour-connected pairs which decay into hadrons. Both models are depicted in Figure 2.6.

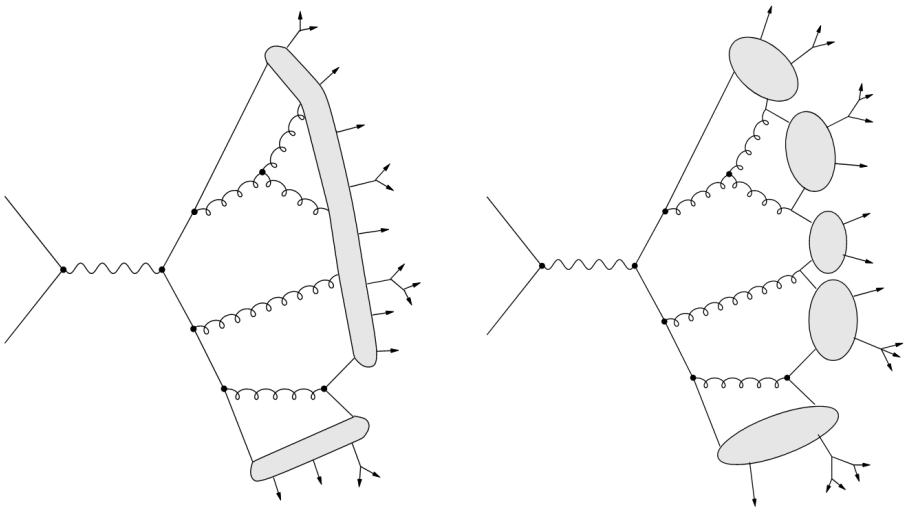

**Figure 2.6.:** Sketch of string (left) and cluster (right) hadronization model, taken from Ref. [30].

These hadrons, however, are not necessarily stable. Complex models, based on theory and experiment, exist to simulate their decay to lighter hadrons which can be considered stable on the timescales of particle physics detectors (dark green).
The incoming hadrons not only exclusively take part in the hard interaction, but additional partons from the incoming hadrons can also interact with each other (purple). These interactions are often a lot softer than the hard event but can overlay the struc-

ture of the event. The final state particles produced in these multiple parton interactions (MPI) is therefore part of the underlying event.

To make the output of MC generators directly comparable to the data taken by a detector, the detector response has to be simulated as well. Precise information of every part of the detector must be considered while calculating the detector simulation. The detector geometry and magnetic field strength as well as detector materials are evaluated while the particles are propagated through the detector. Neutrinos are entirely removed from the final state, since they do not interact with the detector. The ATLAS collaboration uses the GEANT 4 simulation toolkit [31] and ATLFAST-II [32] for their detector simulation. GEANT 4 simulates the entire interactions between the particles and the detector material, resulting in very accurately simulated events. However, the accuracy makes the simulation very CPU intensive. Since most of the computing time is used to simulate particles traversing the calorimeters, by default ATLFAST-II only uses the full simulation for the tracking systems and deploys a fast simulation of the calorimeters, meaning a parametrized shower model is used instead of calculating every single interaction. This improves the simulation time by a factor of 10.

### 2.3.3. Jets

The final state of MC event generators, as well as in particle physics experiments consist of great numbers of hadrons and leptons. Analysing every single particle is tedious and often not even useful. Particles that occur from arbitrary soft and/or collinear radiation in the parton shower do not lead to a different end result. To find a final state observable that is insensitive to soft and collinear emissions and allows comparing experimental data to various theoretical approaches, the concept of a jet is introduced.
Jets are transverse momentum weighted observables that do not depend on the detailed structure of the final state. Figure 2.7 sketches how a jet definition applied to several events that only differ through soft/collinear branching and hadronization can give identical jets.
Historically, jets were just geometrical cones, build around seed particles, which are the particles with the largest transverse momentum. Today, the standard algorithms are sequential clustering algorithms, which sequentially combine particles by adding their four-momentum, until the final jet is formed.
A widely used example for such sequential clustering algorithms is the **anti-$k_T$ jet clustering algorithm**. It introduces the distance measure $d_{ij}$ between any two particles or pseudojets $i$ and $j$, and the transverse momentum measure $d_{iB}$. The smallest measure

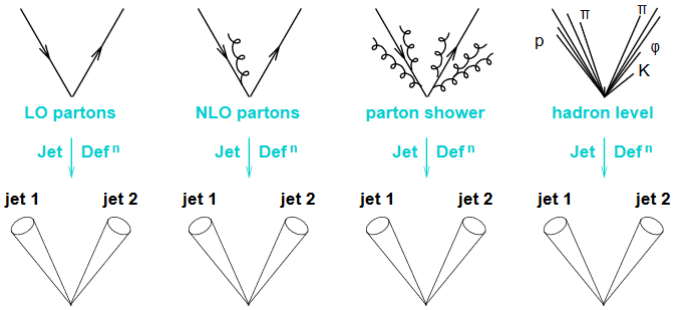


**Figure 2.7.:** Application of a jet definition to multiple events that differ just through soft/collinear branching and hadronization can yield the same jet, taken from Ref. [24].

between any of these objects is identified. If this distance is $d_{ij}$, the objects $i$ and $j$ are recombined into a pseudojet. If it is $d_{iB}$, then the object $i$ is called a jet and is removed from the list of objects. The procedure is repeated until no objects are left. The defining characteristic of the anti-$k_T$ algorithm lies in the definition of the distances:

$$d_{ij} = \min(k_{T,i}^{2p}, k_{T,j}^{2p}) \frac{\Delta_{ij}^2}{R^2}, \tag{2.40}$$

$$d_{iB} = k_{T,i}^{2p}, \tag{2.41}$$

where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ and $k_{T,i}$, $y_i$ and $\phi_i$ are the transverse momentum, rapidity and azimuth of particle $i$, respectively. For $p = -1$ the algorithm corresponds to the anti-$k_T$ algorithm [33]. The choice $p = 1$ identifies the $k_T$ algorithm [34], and $p = 0$ the Cambridge/Aachen algorithm [35] for $p = 0$.

## 2.4. $X \to HH(/SH) \to b\bar{b}WW^{(*)}$ Topologies

The studies presented in this thesis were performed in the context of the ATLAS $X \to HH(/SH) \to b\bar{b}WW^{(*)}$ analysis [9], searching for resonant Higgs boson pair production via BSM extensions, such as the ones described in Section 2.2.2. The two Higgs bosons are required to decay into a $b\bar{b}$ and $W^+W^-$ pair respectively, where the two $W$ bosons are required to decay semi-leptonically. For a high resonant mass $m_X$, the final state will be boosted, leading to small angular separations of the decay products, and the need to study and find techniques to resolve this signature. This section deals with the resonant and non-resonant production of Higgs pairs at hadron colliders, the possible decay channels of Higgs boson pairs, different event topologies for the resonant case and the main contributing background channels.

### 2.4.1. Higgs Boson Pair Production

Higgs boson pair production in the SM without any further particles in the final state occur dominantly in two ways at hadron colliders [36], both initiated by the fusion of gluons. The Feynman diagrams of these processes are drawn in Figure 2.8, where the production in Figure 2.8(a) is called a box diagram and the production in Figure 2.8(b) features the cubic self-interaction term already discussed in Section 2.2.1.

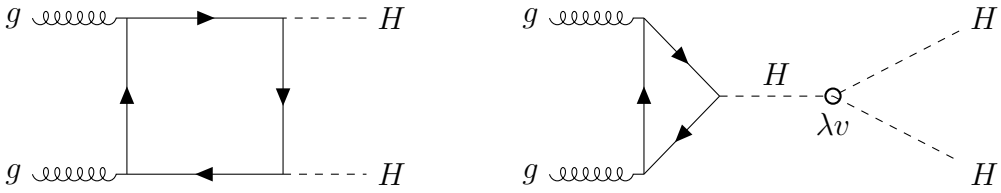


**Figure 2.8(a):** box diagram **Figure 2.8(b):** cubic self-coupling

**Figure 2.8.:** Dominant Feynman diagrams for SM Higgs pair production.

This process therefore allows to directly measure the parameter $\lambda$ of the SM Higgs potential. Both processes are induced by quark loops. As the Higgs coupling to fermions is proportional to the fermion's mass, light quark loops are suppressed. The one loop property of these processes already makes these processes less likely than tree level processes, but the two diagrams in Figure 2.8 also interfere destructively, lowering the cross section to

$$\sigma_{gg\to HH} = 31.05^{+2.2\%}_{-5.0\%}, \tag{2.42}$$

at a centre of mass energy of $\sqrt{s} = 13\,\text{TeV}$ [37].
When focusing on specific decay channels of the Higgs bosons and consecutive particles, the cross section of the desired process is much smaller. Applying cuts during the analysis of events reduces the event yield further, resulting in a very limited number of expected signal events.

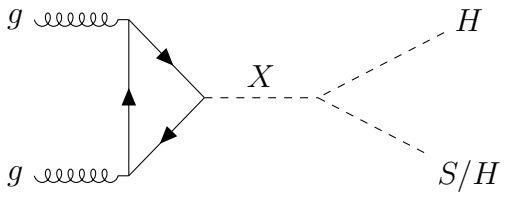


**Figure 2.9.:** Feynman diagram for resonant Higgs pair production.

However, with BSM extensions, the coupling and therefore cross section can be increased in resonant production. The relevant Feynman diagram can be seen in Figure 2.9, where

a heavy scalar particle $X$ decays into either $SH$ or a $HH$ pair. This additional production channel can increase the cross section of Higgs pair production, resulting in more signal events than the SM predicts.

### 2.4.2. Di-Higgs Decay Channels

The decay width of the SM Higgs boson is calculated in Ref. [38] to be $\Gamma_H = 4.097\,\text{MeV}$ and therefore has a lifetime of $\tau_H \approx 1.6 \cdot 10^{-22}\,\text{s}$. This lifetime is too short to be resolved in current particle detectors. Instead of the Higgs boson itself, its decay products are those that are visible in the detector. The branching ratios of Higgs pairs are listed in Table 2.2.

| $H_1 \rightarrow$ \ $H_2 \rightarrow$ | $b\bar{b}$ | $W^+W^-$ | $\tau^+\tau^-$ | $ZZ$ | $\gamma\gamma$ |
|---|---|---|---|---|---|
| $b\bar{b}$ | 34% | | | | |
| $W^+W^-$ | 25% | 4.6% | | | |
| $\tau^+\tau^-$ | 7.3% | 2.7% | 0.39% | | |
| $ZZ$ | 3.1% | 1.1% | 0.33% | 0.07% | |
| $\gamma\gamma$ | 0.26% | 0.1% | 0.02% | 0.01% | <0.001% |

**Table 2.2.:** Branching ratios of Higgs pairs with mass $m_H = 125\,\text{GeV}$. The combinations of the most relevant decay channels are given, where rows describe the decay mode of the first and columns describe the decay mode of the second Higgs boson. The colour coding emphasises the probability distribution over the different channels, where channels coded pink are more likely than (light-) blue channels. Branching ratios taken from Ref. [38].

Due to the Higgs coupling strength being proportional to the mass of particles, decays into heavy particles are generally preferred. But since the invariant mass of $W^+W^-$ and $ZZ$ systems are more massive than the mass of the Higgs boson, these decays are kinematically suppressed (even more so for $t\bar{t}$ pairs) and require one of the decay products to be off-shell. A Higgs boson decaying into a $b\bar{b}$ pair therefore has the highest probability, consequently the $b\bar{b}b\bar{b}$ final state has the highest probability for di-Higgs decays with 35%. This decay channel is however difficult to separate from the large amount of background, making it hard to be successfully identified. The second most likely decay channel is the $b\bar{b}WW^*$ channel, making up 25% of all di-Higgs decays. The W boson decays itself and different analysis strategies have to be applied for different decay channels. With a probability of about 32.72%, a $W$ boson will decay into a lepton and neutrino particle-antiparticle pair $W \rightarrow \ell\nu_\ell$ [3]. When the $W$ does not decay leptonically, it decays into hadrons, or more precisely at tree level into particle-antiparticle pairs of up- and down-type quarks, which

later form hadrons. Lepton decays produce a much cleaner final state than hadron decays. By targeting events where one $W$ decays leptonically, the identification of the di-Higgs decay becomes easier, while keeping one hadronically decaying $W$ does not reduce the event yield too much.

This decay channel is used in the analysis. To estimate the expected number of SM like $HH \to b\bar{b}W_{\text{had}}W_{\text{lep}}$ events, one can consider the cross section given in Equation (2.42), which represents the probability to find such an event before making any cuts on observables that are necessary in the analysis, and the integrated luminosity taken by the ATLAS detector in Run 2. The integrated luminosity $\mathcal{L}_{\text{tot}}$ is the ratio of the number of events $N$ detected over the cross section $\sigma$: $\mathcal{L}_{\text{tot}} = N/\sigma$. The luminosity taken by the ATLAS detector during the years 2015-2018 is $\mathcal{L}_{\text{tot}} = 140\,\text{fb}^{-1}$ [39]. By also taking into account the branching ratios of the di-Higgs and $W$ decay for the specific channel, the number of expected events are of order 100.

### 2.4.3. Event Topologies

Depending on the actual masses of the BSM $S$ and $X$ bosons, the topology of the events are vastly different. The spectrum of categories under which these topologies fall depending on the masses $m_S$ and $m_X$ are shown in Figure 2.10, the topologies of each category are depicted in Figure 2.11. The categories differ in how far the final state objects are separated from each other, with the two extrema being the resolved and boosted topologies.

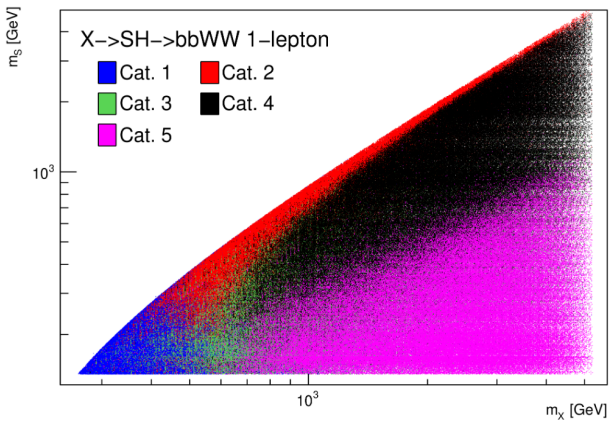


**Figure 2.10.:** Different event topology categories $X \to SH \to b\bar{b}WW$ in the single lepton channel depending on the masses $m_s$ and $m_X$. The categories are by ascending numerical order: resolved, split-resolved, semi-boosted, split-boosted and boosted. Figure taken from Ref. [40].

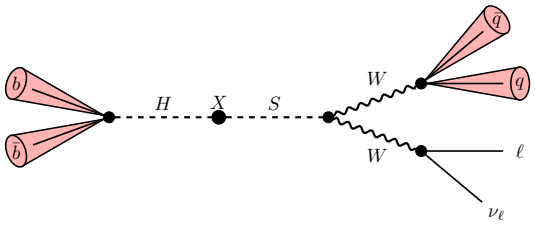


**Figure 2.11(a):** Cat. 1: resolved

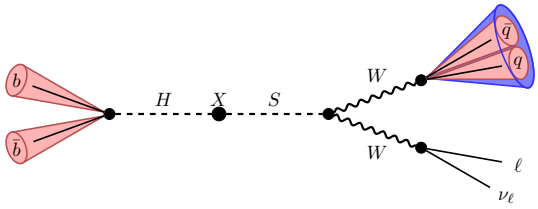


**Figure 2.11(b):** Cat. 2: split-resolved

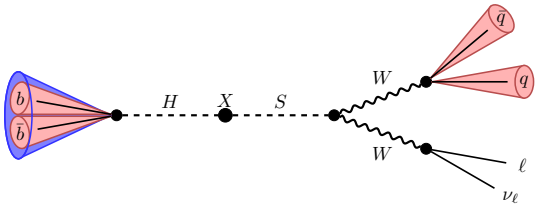


**Figure 2.11(c):** Cat. 3: semi-boosted

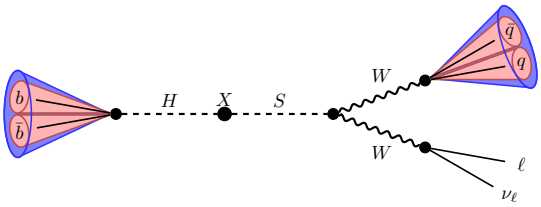


**Figure 2.11(d):** Cat. 4: split-boosted

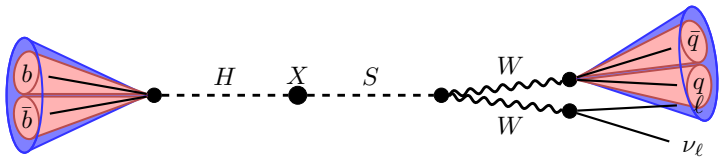


**Figure 2.11(e):** Cat. 5: boosted

**Figure 2.11.:** Different categories of the $X \to SH \to b\bar{b}WW$ topology in the single lepton channel.

The resolved topology is mostly sensitive to SM like $HH$ and resonant $SH$ decays, where both $m_X$ and $m_S$ are small. Most of the event's energy is used to create the $H$ and $S$ bosons, leaving not much energy for the boost of these particles. Further down in the decay chain, the particles receive most of their momentum from the mass of the decaying parent, separating all particles in the final state well from each other and allowing for the reconstruction of two $b$-jets from one Higgs, two jets from the hadronically decaying $W$, as well as a lepton neutrino pair from the leptonically decaying $W$ boson.
The boosted topology occurs for $HH$ and $SH$ decays, where the $X$ boson is heavy ($m_X \gtrsim 1\,\text{TeV}$) and the $S$ is significantly lighter ($m_S \leq 0.3 m_X$). The large mass difference means that a lot of the energy in the mass of the $X$ will be converted to kinematic energy and boosts the decay particles. The momentum transverse to this boost comes from the other decays in the decay chain, which will be much smaller than the initial boost, resulting in a very small spatial separation. The final state particles will be so close to each other that they cannot be resolved individually in a detector. Only a large radius (large-$R$) jet for the $H \to b\bar{b}$ decay and another large-$R$ jet for the hadronically decaying $W$ boson $W_{\text{had}}$, as well as a lepton neutrino pair from the leptonically decaying $W$ boson $W_{\text{lep}}$ can be detected. The final state is so collimated that the lepton can overlap

with the $W_{\text{had}}$ jet. Therefore, special care is needed to avoid double counting the lepton's energy in the detector.

To try and resolve the characteristics of signal events inside the large-$R$ jets, special jet objects, namely TAR and UFO jets (see Section 4.2) are used. These objects have increased resolution by making use of tracking information instead of only relying on the calorimeter (see Section 3.2). In addition, the use of substructure variables (see Section 4.1) helps to distinguish signal from background events. To study these objects independently of the ATLAS analysis framework and without having to implement the BSM extensions in the used event generator SHERPA [41], a mock signal is used for this thesis. The process $Z^* \to ZH \to b\bar{b}WW^*$ depicted in Figure 2.12 mimics the $H \to WW^*$ part of the $X \to HH \to b\bar{b}WW^*$ process. When the kinematics of the decaying Higgs boson are matched to eac other, the subprocesses are equivalent.

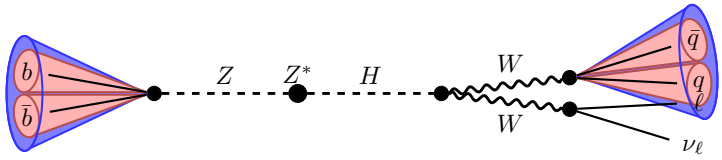


**Figure 2.12.:** Depiction of process used to mimic the $H \to WW^*$ part of the signal process.

# 3. Experimental Setup

This chapter deals with the machines that make physics analyses like the $b\bar{b}WW^{(*)}$ analysis possible. The Large Hadron Collider (LHC, see Section 3.1) provides accelerated protons that collide inside the ATLAS detector (see Section 3.2), which records the energies, charges, momenta and flight paths of the resulting spray of particles in multiple detection layers. In addition to the description of these apparatus, the detailed MC generator settings used in context of the mock signal as well as for the studies performed in context of the ATLAS analyses are summarized in Section 3.3.

## 3.1. The Large Hadron Collider

The Large Hadron Collider (LHC) is a synchrotron accelerator with a circumference of 27 km based at the European Organization for Nuclear Research (CERN) in Geneva. An in-depth description of the machine design can be found in Ref. [42]. The LHC is designed as a proton-proton collider but can also collide lead nuclei. It holds the record for the highest achieved energy in any particle collisions at a centre of mass energy of $\sqrt{s} = 13.6\,\text{TeV}$, and it is planned to increase the energy up to $\sqrt{s} = 14\,\text{TeV}$.
The protons cannot be accelerated straight away to the final centre of mass energy but rather undergo a series of acceleration steps. The first step in the series, also depicted in Figure 3.1, is the Linear Accelerator 4 (Linac4), which accelerates negative hydrogen ions $H^-$ to 160 MeV and then injects the ions into the Proton Synchrotron Booster (PSB). In the PSB, the electrons are stripped from the ions, leaving only the proton. The protons are then accelerated to 2 GeV and injected into the Proton Synchrotron (PS), where they are further accelerated to 26 GeV. The last acceleration step before the LHC is the performed in the Super Proton Synchrotron (SPS), where the protons reach an energy of 250 GeV. After several minutes the protons have reached the main ring, where they are accelerated over 20 min to their peak energy.
In order to accelerate particles with the same charge in opposite directions, the LHC requires separate rings with opposite magnetic dipole fields with a field strength of $\left|\vec{B}\right| = 8\,\text{T}$. To reach such high field strengths, the magnets are cooled down to below 2 K. Quadrupole

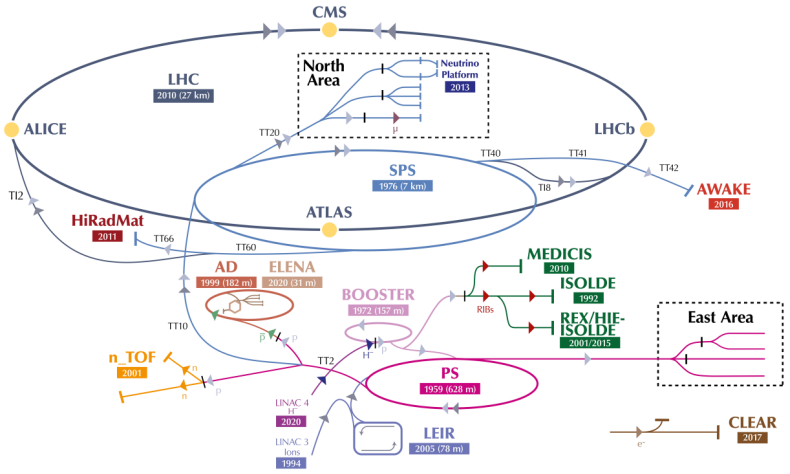


**Figure 3.1.:** Sketch of the accelerators at CERN, which eventually feed into the LHC, adapted from Ref. [43].

and higher order multipole magnets are used to perform minor corrections to the proton beams. The protons are accelerated in distinct packets, called bunches. There are up to $1.15 \cdot 10^{11}$ protons per bunch and up to 2808 bunches in total. When the protons have reached their target energy, they travel at almost the speed of light. The bunches follow each other with a time interval of 25 ns, colliding at a rate of 40 MHz at the four experiments situated at the LHC.

Every time the two proton beams cross, numerous interactions occur, with an average of 33.7 interactions per bunch crossing. The distribution of average interactions can be seen in Figure 3.2(a). The luminosity delivered by the LHC not only depends on the frequency $f$ of bunch crossings, the number of particles per bunch in both beams $N_1$, $N_2$, the number of bunches $n_b$, but also the transverse dimensions of the beams $\sigma_{x/y,i}$ and a luminosity reduction factor $S$, depending on the size of the bunches (also in the longitudinal direction) and the crossing angle of the bunches. The reduction by the factor $S$ essentially yields the effective beam size. The instantaneous luminosity can be expressed by [44]

$$\mathcal{L} = \frac{N_1 N_2 n_b f S}{2\pi\sqrt{\sigma_{x,1}^2 + \sigma_{x,2}^2}\sqrt{\sigma_{y,1}^2 + \sigma_{y,2}^2}}. \tag{3.1}$$

During the operation of the LHC the proton beams naturally lose protons to the interactions and smaller beam losses, such that a beam has a lifetime of approximately 15 h before the beams are dumped, and new beams have to be formed.

As mentioned earlier, there are four experiments at the LHC, each sitting at one of the four beam crossings. The two largest experiments at the LHC are ATLAS [45] and CMS [46], which both are general purpose experiments and together claimed the discovery of

the Higgs boson in 2012 [10, 11]. They focus on precision measurements to test the SM and the search for BSM evidence. The LHCb experiment [47] focuses on the research of $b$-physics, relying on the decay of $B$-mesons to test and measure parameters of the SM. In contrast to the detectors of ATLAS and CMS, the LHCb detector has an asymmetric structure. The ALICE experiment [48] is optimized to study lead-lead collisions, which produce quark-gluon plasma representing the state of the universe shortly after the Big Bang.

The LHC is not constantly running and started its operation in 2010 only with a reduced luminosity and a smaller centre of mass energy. Since 2010, the operation mode has switched between phases of collisions and data taking, referred to as a Run, and long shutdown (LS) phases. The accumulation of integrated luminosity during Run 2 from 2015 to 2018 can be seen in Figure 3.2(b). At the time of the writing of this thesis the LHC is in Run 3, targeting to reach an integrated luminosity of 250 $\mathrm{fb}^{-1}$ [49], almost double the amount of data taken in Run 2. The long-term goal is to repair and upgrade the LHC as well as the detectors in LS 3 to increase the instantaneous luminosity further and reach an integrated luminosity of 3000 $\mathrm{fb}^{-1}$ in Run 4.

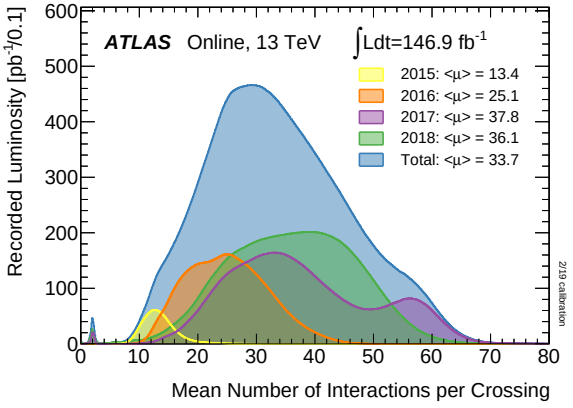


**Figure 3.2(a):** average number of interactions

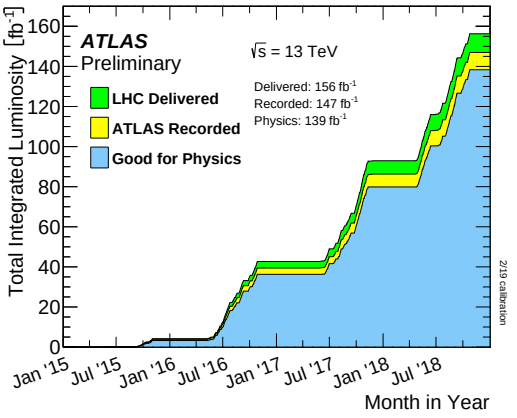


**Figure 3.2(b):** integrated luminosity

**Figure 3.2.:** The average number of interactions per bunch crossing during the years 2015-2018, as well as the luminosity delivered by the LHC, recorded by ATLAS and declared as usable for physics analyses over these years. Figure from Ref. [50].

## 3.2. ATLAS Detector

The ATLAS (A Toroidal LHC ApparatuS) detector [45] is the largest volume particle detector ever built with a length of 46 m, a diameter of 25 m and a mass of 7000 t.

Figure 3.3 shows a computer-generated image of the detector.

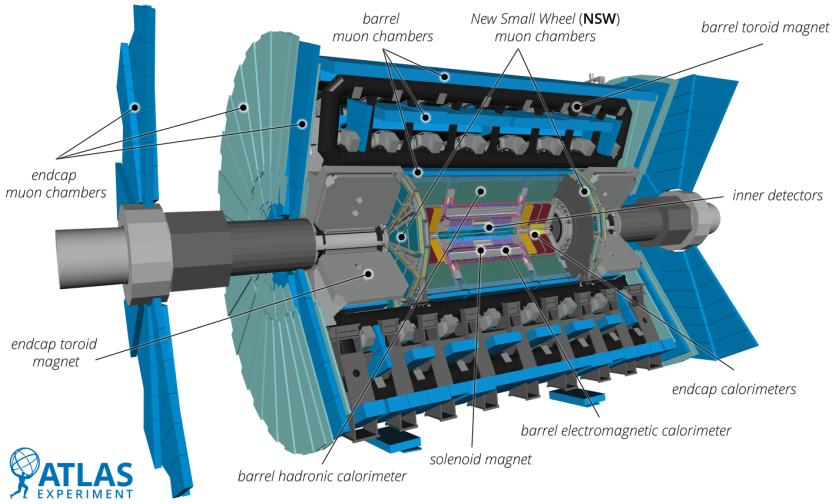


**Figure 3.3.:** Computer-generated illustration of the ATLAS detector, taken from Ref. [51].

It is built with cylindrical symmetry around the beampipe. The detector covers almost the full solid angle. As a general-purpose detector, several different detection layers exist to measure and identify (almost) all possible final states.

## Detector Layers

The first layer, the Inner Detector (ID), starts just 3.3 cm away from the beampipe. The ID lies in a 2 T magnetic field of a solenoid magnet parallel to the beam axis. Due to the Lorentz force, charged particles therefore propagate on a curved trajectory. They are detected where they interact with the individual detector cells, enabling the ID to track the flight path of charged particles. The curvature of the tracks can be used to calculate the momentum and charge of these particles. The ID itself consists of three subcomponents, the Pixel detector, the Semiconductor Tracker (SCT) and the Transition Radiation Tracker (TRT). The spatial resolution of the ID worsens with the distance to the beampipe, since charged particles can interact and scatter off of the detector materials. The pixel detector has a minimum cell size of $50 \times 400\,\mu\text{m}^2$ in $R\phi \times z$ space, where $R$ describes the distance to the beampipe, $\phi$ is the azimuthal angle around the $z$-axis, which points in the direction of the beampipe. An additional pixel layer, the Insertable B-Layer (IBL) [52] was installed during LS 1 in 2013-2014. The IBL has an even finer spatial resolution with pixel cells of size $50 \times 250\,\mu\text{m}^2$ ($R\phi \times z$). The accuracy of the SCT is $17 \times 580\mu\text{m}^2$ ($R\phi \times z$). The TRT is made out of long straws and only offers information

in $R\phi$, with a resolution of 130 $\mu$m per straw. The ID resolves the transverse momenta $p_T$ of particles with a resolution of $\sigma_{p_T}/p_T = 0.05\% \cdot p_T \oplus 1\%$.

After traversing the ID, the particles enter the calorimeters, which measure the energy of the particles. The calorimeters have alternating layers of high-density passive material, and active layers for measurements. The traversing particles interact with the passive layers, releasing bremsstrahlung and creating particle pairs, building a particle shower. The calorimeters are built long enough, such that the energy of most particles entering the calorimeters is eventually absorbed. The calorimeters are divided into an electromagnetic (ECAL) and a hadronic calorimeter (HCAL), where the ECAL measures the energy of electrons, positrons and photons, while the HCAL measures the energy of hadrons. The ECAL is surrounded by the HCAL. The ECAL does not measure the energy of tau leptons, since they decay before reaching the ECAL. Since electrons, positrons and photons only shower by electromagnetic processes and hadrons mainly interact via the strong force, different passive and active materials are required in the respective calorimeters. Lead is used as the passive layers in the barrel and end-cap part of the ECAL, while copper is used in the end-cap of the HCAL. An additional calorimeter is installed in the forward region of the detector, using copper and tungsten as passive materials. These systems share liquid Argon as the active material. The barrel part of the HCAL uses steel as the passive and plastic scintillating tiles as the active material. The ECAL is designed to achieve an energy $E$ resolution of $\sigma_E/E = 10\%/\sqrt{E} \oplus 0.7\%$ for electromagnetic showers. The barrel and end-cap segments of the HCAL are designed for an energy resolution of $\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$, while the forward segment is designed to achieve $\sigma_E/E = 100\%/\sqrt{E} \oplus 10\%$ for hadronic showers.

While neutrinos pass the ID and calorimeters without being detected at all, muons are minimal ionising particles and the measurement of muons in the ID are inaccurate, due to typical high momenta. The measurement of muons can however be improved by an additional tracking detector, the muon spectrometer. Neutrinos on the other hand stay undetected and can only be inferred from missing transverse energy, the energy needed not to violate energy conservation. The muon spectrometer is the outermost part of the ATLAS detector and enclosed by the fields of three toroidal magnets, bending the flight path of muons. One magnet each is located at the end-caps of the ATLAS detector, and the largest one surrounds the barrel. The muon spectrometer consists of thousands of individual muon chambers, which are wrapped around the ID and calorimeters, as well as the toroidal magnets. It is designed for a resolution of $\sigma_{p_T} = 10\%$ for $p_T$=1 TeV muons.

A schematic cross-section of the ATLAS detector, as well as a visualization of how particles behave in the detector and are detected, can be seen in Figure 3.4.

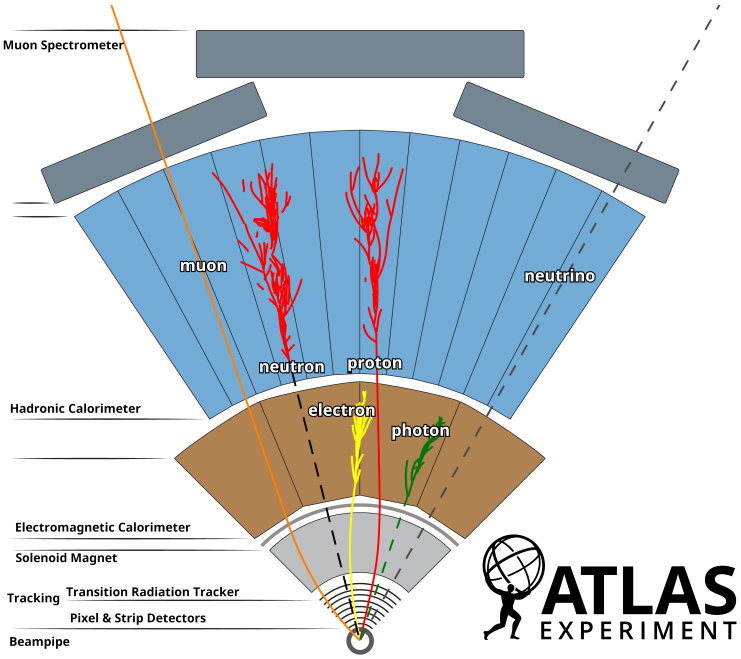


**Figure 3.4.:** Schematic cross-section of the ATLAS detector with a visualization of how the detector detects different particles, adapted from Ref. [53].

## Trigger System

The amount of events produced during the operation of the LHC is too big to store every event. On average there are 33.7 interactions per bunch crossing at a frequency of 40 MHz, yielding 1.348 billion interactions per second. In addition, due to the impossibility to save all the data at this rate, most of the interactions are also interactions at low energy scales, not featuring the targeted physics. It is therefore of great importance to filter the events for processes of interest in real time, while the data is being taken, and only store these interesting processes.

The ATLAS experiment uses a two-stage trigger system for this purpose [54, 55]. The two stages are the Level-1 (L1) and High-Level Trigger (HLT). The L1 trigger is a hardware-based system, using information of the calorimeters and muon spectrometer. The trigger may fire when event-level requirements are met, for example when there is a large amount of energy deposited in the calorimeters or a muon with high energy is detected. Other trigger requirements are placed on the multiplicity of objects above a certain energy threshold or when the combination of multiple clusters meet the trigger requirements, such as the invariant mass or angular distance between two objects. Multiple trigger paths exist to accommodate for diverse event signatures that are deemed to be interesting in various analyses. The triggers therefore function at different working points, for the criteria in question. The L1 trigger reduces the rate of bunch crossing from 40 MHz

down to the maximum detector read-out rate of 100 kHz. The HLT is software based and uses the whole granularity of the detector. The HLT runs through several stages, where each stage can veto the current event. To save CPU time, the stages are ordered by their computing intensity. The later stages of the HLT work similar to offline-like event construction algorithms, but since the trigger still needs to make decisions on the fly, it cannot reach the efficiency of offline procedures. The output of the HLT reduces the frequency of interesting events further down to 1.2 kHz, with an average physics throughput to permanent storage of 1.2 GB/s.

## 3.3. Monte Carlo Generator Setup

While MC generators all follow the same principle, each step in the simulation of events can be accomplished by different algorithms, making use of various physics models or approximations. The specific settings that were used in parts of this thesis are described in this section, divided into the truth studies performed using SHERPA [41] as the generator with RIVET [56] as the analysis tool, and the studies performed using ATLAS samples.

### 3.3.1. Truth Studies with Sherpa

The events in these studies are generated by SHERPA version 2.2.12 . The specific settings for each part of the truth studies are discussed in the following.

#### QCD Background

QCD jets are used to validate the implementation of different jet objects and as a background to the studied boosted decay. SHERPA is set up to collide two protons, each with an energy of 6.5 TeV. A $2 \to 2$ parton process is generated, where the two partons leaving the hard process are each the seed of a jet. The process is simulated at next-to-leading order (NLO) precision, where the tree-level matrix elements are generated by AMEGIC++ [57] and virtual corrections are provided by OPENLOOPS [58]. To combine the matrix elements with SHERPA's Catani-Seymour parton shower algorithm [59], the MEPS@NLO [60] formalism is used. The parton-to-hadron transition is handled by SHERPA's cluster fragmentation model implemented in the AHADIC++ module [61].
An excerpt of the used run card to start the event generation, showing relevant settings can be seen in Listing 1. Since very low-$p_T$ jets are of no importance in this study, as ultimately these QCD-jets are being compared against jets originating from a boosted decay, SHERPA explicitly checks the event for at least one anti-$k_T$ jet with $p_T > 75$ GeV.

This cut is applied at matrix element level, and therefore saves a lot of computation time on low-$p_T$ events.

```
(processes){
  Process 93 93 -> 93 93
  Order (*,0);
  NLO_QCD_Mode MC@NLO;
  ME_Generator Amegic;
  RS_ME_Generator Comix;
  Loop_Generator LOOPGEN;
  End process;
}(processes)
```

**Listing 1:** `(processes)` block in the SHERPA run card for QCD background generation. The $2 \to 2$ parton process is enabled by `Process 93 93 -> 93 93`, while the other lines specify the used modules.

### Higgs-Strahlung

To study boosted decays of Higgs bosons, the process of Higgs-Strahlung $pp \to Z^* \to ZH$, where an off-shell $Z$ boson decays into a $b\bar{b}$ pair and the Higgs decays into two $W$ bosons, is chosen. One of the $W$ bosons is forced to decay leptonically into either a muon or an electron and the corresponding neutrino. When not stated otherwise, the lepton was fixed to be a muon. The second $W$ boson is forced to decay hadronically into quarks of the first ($u\bar{d}$ or $\bar{u}d$) or second ($s\bar{c}$ or $\bar{s}c$) generation. The process is sketched in Figure 2.12. The tree-level matrix element is calculated by COMIX [62] and merged with SHERPA's Catani-Seymour parton shower by an extended CKKW approach [63] with $Q_{\text{cut}} = 20\,\text{GeV}$. The parton-to-hadron transition is handled by SHERPA's cluster fragmentation model implemented in the AHADIC++ module. An excerpt of the run card, describing the chosen decay channels and part of the generator settings used to generate the signal events, can be seen in Listing 2.
At matrix element level, SHERPA guarantees a transverse momentum of the Higgs boson of at least $p_T > 100\,\text{GeV}$. In addition to the signal process, the generator settings allow for up to one parton at the end of the hard process. The signal can therefore be contaminated by an additional QCD jet.

```
(run){
  % ...
  % decay information
  HARD_DECAYS On; HARD_MASS_SMEARING 1;
  STABLE[23] 0; STABLE[24] 0; STABLE[25] 0;
  HDH_STATUS[25,-24,2,-1] 2;
  HDH_STATUS[25,-24,4,-3] 2;
  HDH_STATUS[25,-24,12,-11] 2;
  HDH_STATUS[-24,-12,11] 2;
  HDH_STATUS[-24,-2,1] 2;
  HDH_STATUS[-24,-4,3] 2;
  HDH_STATUS[23,5,-5] 2;
}(run)
(processes){
  Process 93 93 -> 23 25 93{1};
  Order (*,2); CKKW sqr(QCUT/E_CMS);
  ME_Generator Comix;
  End process;
}(processes)
```

**Listing 2:** Run card excerpt for the signal generation. The hard process $pp \to Z^* \to ZH(+\text{jet})$ is enabled by `Process 93 93 -> 23 25 93{1}`. The different decay channels for the on- and off-shell $W$ bosons are described in the `HDH_STATUS` lines. The possibility of both $W$ bosons decaying hadronically/leptonically is vetoed by the event selection in the Rivet analysis.

### 3.3.2. ATLAS Samples

The various background and signal samples of the $H \to HH(/SH) \to b\bar{b}WW^{(*)}$ analysis are created using several MC generators and settings. The specific settings are mentioned separately for signal and background samples.

#### $X \to HH/SH$ Samples

The $X \to HH$ samples are generated by MadGraph 2.6.1 [64] at leading order using the NNPDF2.3 PDF set [65]. Parton shower simulation and hadronization are handled by Herwig 7.1.3 [66], where heavy quark flavour decays are taken over by EvtGen 1.6.0 [67]. The samples are produced for multiple mass points $m_X \in [0.8, 5.0]$ TeV. The branching ratio of the two Higgs bosons are set to 50% $b\bar{b}$ and 50% $WW^*$ or $ZZ^*$ (for the related 0-lepton final state analysis) decays. A filter ensures one $H \to VV^*$ and one $H \to b\bar{b}$ decay in the event.

The $X \to SH$ samples are generated by Pythia 8.244 [68], tuned to ATLAS data using the tune A14 [69], at leading order and with the NNPDF2.3 PDF set. The hadronization is performed by Pythia itself, where heavy quark flavour decays are taken over by EvtGen 1.7.0. Samples produced for combinations of the masses $m_S \in [0.17, 2.5]$ TeV and $m_X \in [0.35, 3.0]$ TeV that allow both $S$ and $H$ to be produced on-shell. The decays $H \to b\bar{b}$ and $S \to W^+W^-$ are enforced.
The detector simulation for both signals is performed by Atlfast-II. Furthermore, the events are filtered to only allow one lepton and one neutrino from the $W$ decays.

**Background Samples**

The following background processes were simulated: single top quark production (single Top), top quark pair production ($t\bar{t}$), massive vector boson production with additional jets ($V$+jets), massive vector boson pair production (diboson) and single Higgs boson production (single Higgs). The detector response is performed using the full simulation with Geant 4.
Top quarks that are produced together with a $W$ boson $tW$ are modelled with Powheg v2 [70–73] at NLO with the NNPDF3.0 set [74] where the diagram removal scheme [75] removes interference and overlaps with $t\bar{t}$ events. The parton shower and hadronization are executed by Pythia 8.230 in the A14 tune with the NNPDF2.3 PDF set.
Single top $t$-channel production is modelled by the Powheg v2 generator at NLO using the four-flavour scheme [76] and the NNPDF3.0 PDF set. The parton shower and hadronization are executed by Pythia 8.230 in the A14 tune with the NNPDF2.3 PDF set.
Single top $s$-channel production is modelled by the Powheg v2 generator at NLO using the NNPDF3.0 PDF set. Parton showers and hadronization are executed by Pythia 8.230 in the A14 tune with the NNPDF2.3 PDF set.
For the production of $t\bar{t}$ samples, PowhegBox v2 is used at NLO. The PDF set NNPDF3.0 is utilized. The $h_{\text{damp}}$ parameter is set to $1.5m_{\text{top}}$ [77], regulating the high-$p_T$ radiation against which the $t\bar{t}$ system recoils. Parton showers and hadronization are handled by Pythia 8.230 in the A14 tune with the NNPDF2.3 PDF set and heavy quark flavour decays being treated by EvtGen 1.6.0.
$V$+jets samples are produced in two different ways. Before the work of this thesis started, only samples produced with Sherpa 2.2.1 were used. In this thesis, new samples using Sherpa 2.2.11 are validated. In the 5 years development time between these versions, a lot of improvements were made. The rest of the settings are shared between the two sets of samples. The events are produced at NLO for up to two partons and at leading order for

up to four partons, where the matrix elements are calculated by Comix and OpenLoops. They are matched and merged with Sherpa's Catani-Seymour parton shower using the MEPS@NLO formalism with the NNPDF3.0 set.
Di-boson final states are either simulated with Sherpa 2.2.1 or 2.2.2 at NLO for up to one additional parton and at leading order for up to three additional parton emissions. Higgs boson contributions and off-shell effects are included.
Loop induced processes $gg \to VV$ are generated at leading order for up to one additional parton emission. Matrix elements are calculated by Comix with virtual QCD corrections provided by OpenLoops. The matrix elements are matched and merged with Sherpa's Catani-Seymour parton shower using the MEPS@NLO formalism with the NNPDF3.0 PDF set.
Single Higgs boson production is modelled by Powheg at NLO and interfaced with Pythia 8. Heavy quark flavour decays are treated by EvtGen 1.6.0. The production of single Higgs boson together with other massive bosons uses the NNPDF3.0 set. Single Higgs bosons that are produced together with a $t\bar{t}$ pair use the A14 tune and NNPDF2.3 PDF set.

# 4. Analysis Objects

In this chapter the observables utilized to study the substructure of jets, as well as the investigated jet objects, are introduced.

## 4.1. Observables

In order to compare and validate the different jet objects against each other, kinematic observables like the jet's transverse momentum $p_T$ or invariant mass $m$ are used. However, the invariant mass of the $HH/SH$ system

$$m^{HH/SH} = \sqrt{\left(p^{H\to b\bar{b}} + p^{W_{\text{had}}} + p^{\ell} + p^{\nu}\right)^2}, \tag{4.1}$$

where $p^i$ denotes the four momentum of the respective particle, cannot be simply reconstructed, due to missing information about the neutrino in the detector. The alternative observable, called visible+met mass, combines all the available information of the system by replacing the neutrino four momentum with the four momentum of the missing transverse energy $p^{\text{met}} = \left(E_T^{\text{miss}}, p_x^{\text{miss}}, p_y^{\text{miss}}, 0\right)$. It is defined as

$$m^{HH/SH}_{\text{vis+met}} = \sqrt{\left(p^{H\to b\bar{b}} + p^{W_{\text{had}}} + p^{\ell} + p^{\text{met}}\right)^2}. \tag{4.2}$$

In addition to the kinematic observables, the **generalized angularities** as defined in Ref. [78] are used. The general angularities are of the form

$$\lambda^{\kappa}_{\alpha} = \sum_{i\in\text{jet}} \left(\frac{p_{T,i}}{\sum_{j\in\text{jet}} p_{T,j}}\right)^{\kappa} \left(\frac{\Delta_i}{R_0}\right)^{\alpha}, \tag{4.3}$$

where

$$\Delta_i = \sqrt{(y_i - y_{\text{jet}})^2 + (\phi_i - \phi_{\text{jet}})^2}, \tag{4.4}$$

is the distance of particle $i$ from the jet axis in the rapidity-azimuth plane. Specifically, the angularities with parameters $(\kappa,\alpha)$=(0,0) and (2,0) are used. The multiplicity $\lambda^0_0$ counts the number of constituents inside a jet, where the constituents can either be all

final state particles in a jet, only the charged final state particles or the subjets inside a larger jet. The angularity $(p_T^D)^2 \equiv \lambda_0^2$ is defined in Ref. [79] and describes how the momentum of a jet is divided onto its constituents: whether it is carried by a few high momentum particles or equally spread over all constituents.

In order to also be able to evaluate the substructure of the jets, five substructure observables are used. The **N-subjettiness** [80] $\tau_N$ distinguishes between a jet consisting of less than or equal to $N$ subjets ($\tau_N \approx 0$) and a jet consisting of more than $N$ subjets ($1 \geq \tau_N \gg 0$). The N-subjettiness is calculated according to

$$\tau_N = \frac{1}{d_0} \sum_{i \in \text{jet}} p_{T,i} \min \left\{ \Delta R_{1,i}, \Delta R_{2,i}, \cdots \Delta R_{N,i} \right\}, \tag{4.5}$$

where $d_0$ is a normalization factor

$$d_0 = \sum_{i \in \text{jet}} p_{T,i} R_0. \tag{4.6}$$

$R_0$ is the characteristic jet radius used in the original jet clustering algorithm, $p_{T,i}$ is the transverse momentum of a jet constituent $i$ and $\Delta R_{j,i} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}$ is the distance (in the rapidity-azimuth plane) of the constituent $i$ to the jet axis of subjet $j$ of exactly $N$ candidate subjets that have been identified beforehand. A greater discrimination comes from taking the ratio of two observables with different values for $N$. A shorter expression for the ratio is introduced: $\tau_{NM} \coloneqq \tau_N / \tau_M$. In this thesis the observables $\tau_{21}$, $\tau_{32}$ and $\tau_{42}$ are considered.

A different substructure observable is the energy correlation function (ECF) introduced in Ref. [81]

$$\text{ECF}(N, \beta) = \sum_{i_1 < i_2 < \cdots < i_N \in \text{jet}} \left( \prod_{a=1}^{N} p_{T_{i_a}} \right) \left( \prod_{b=1}^{N-1} \prod_{c=b+1}^{N} \Delta R_{i_b, i_c} \right)^{\beta}, \tag{4.7}$$

from which the following ratios can be formed

$$C_N^\beta = \frac{\text{ECF}(N+1, \beta) \cdot \text{ECF}(N-1, \beta)}{\text{ECF}^2(N, \beta)} \quad \text{and} \tag{4.8}$$

$$D_N^\beta = \frac{\text{ECF}(N+1, \beta) \cdot \text{ECF}(N-1, \beta) \cdot \text{ECF}^N(1, \beta)}{\text{ECF}^3(N, \beta)}. \tag{4.9}$$

Since the observables $C_2^1$ and $D_2^1$ are found to discriminate 2-prong jets (like those from e.g. boosted $W$ decays) from 1-prong jets [82], these observables are investigated in this thesis. The number of prongs in a jet is related to the number of initial quarks or gluons that contributed to its formation.

## 4.2. Jet Objects

Three different jet objects each are implemented in a RIVET [83] analysis that obtains the SHERPA [41] generated event as input and processes it according to the jet objects' definitions to output histograms of chosen observables. In the following, the jet objects are described.

### 4.2.1. Standard Large-R Jets

As a baseline object, the standard large-R jet (LRJ) is implemented. The final state particles are clustered into an anti-$k_T$ jet with radius $R = 1.0$. The jet is then groomed via a grooming procedure called **Trimming** [84] to reduce soft radiation from initial state radiation, multiple parton interactions and event pile-up from multiple simultaneously occurring proton-proton interactions that overlap with the final state of a hard interaction. With a large-R jet as input, the algorithm reclusters the constituents into subjets with radius $R_{\text{trim}}$. In this specific analysis, the subjets are reclustered using the anti-$k_T$ algorithm with a radius of $R_{\text{trim}} = 0.2$. The subjets are only kept if their transverse momentum $p_T^{\text{subjet}} > f_{\text{cut}} \cdot p_T^{\text{jet}}$, where $f_{\text{cut}} = 0.05$ is a fixed chosen parameter and $p_T^{\text{jet}}$ is the large-R jet's transverse momentum. If the condition is not met, the subjet is discarded. After every subjet was checked for this condition, the remaining subjets are assembled into the trimmed large-R jet.
The anti-$k_T$ clustering algorithm as well as the Trimming algorithm are implemented in the RIVET analysis via the FASTJET package [85].

### 4.2.2. Track-Assisted Reclustered (TAR) Jets

In the context of particle detectors, the finite size of individual calorimeter detector cells limits the angular resolution of calorimeter-based jets. This means in heavily boosted decays, the substructure of large-R jets cannot be resolved reliably, as the detector reaches angular resolution limits. The ATLAS collaboration therefore proposes different algorithms to not only rely on information from the coarser calorimeter, but also the Inner

Detector with a much higher resolution. For this thesis, the TAR jet [7] (this subsection) as well as the Unified Flow Object (UFO) jet [8] (Section 4.2.3) are implemented in RIVET.
In Figure 4.1, the working steps of the TAR algorithm are shown for an explicit example event.

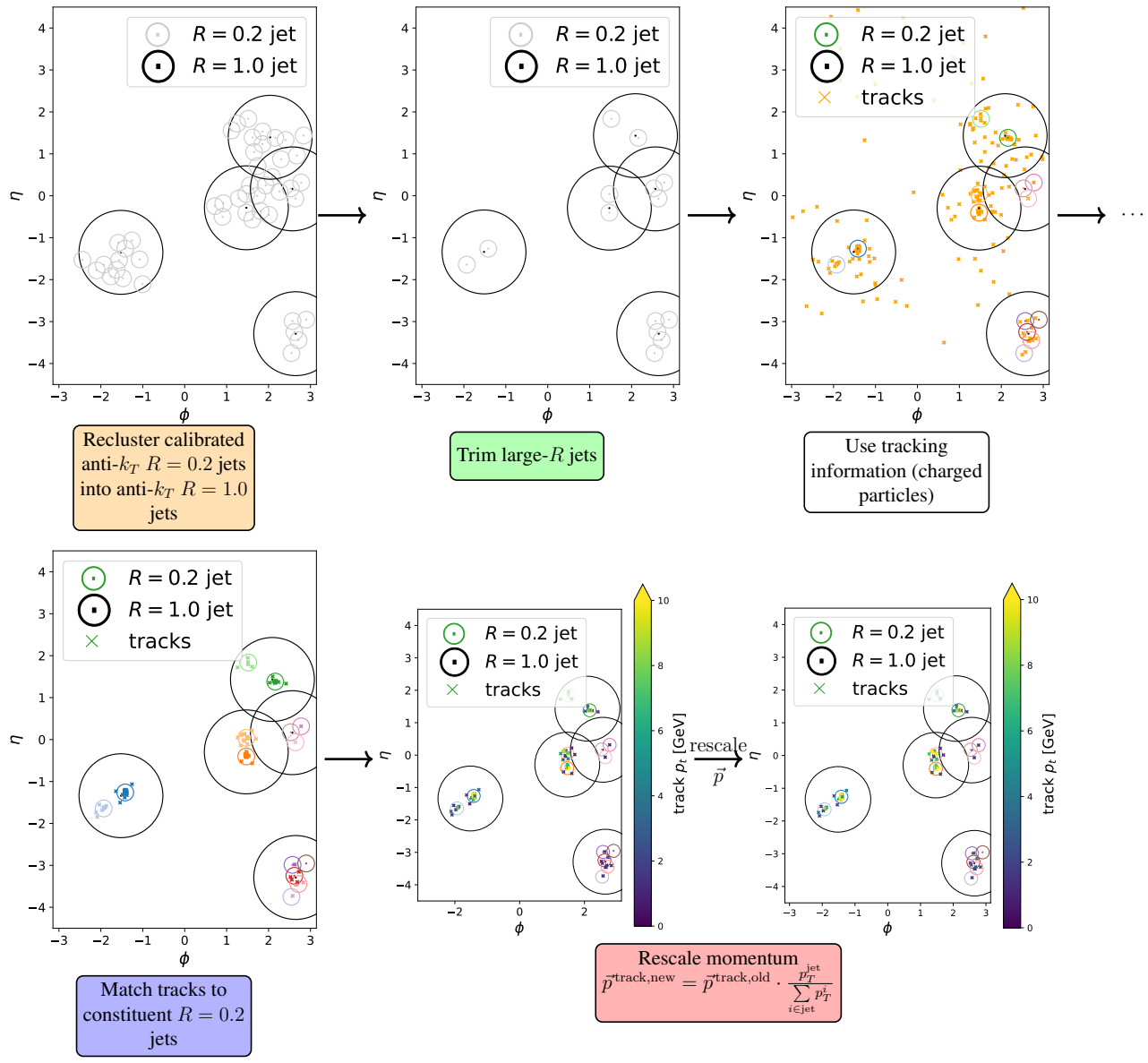


**Figure 4.1.:** Flow chart of TAR jet algorithm with an example event display, picturing each step of the algorithm.

The algorithm first clusters the final state particles into calibrated anti-$k_T$ $R = 0.2$ jets, which themselves are clustered into anti-$k_T$ $R = 1.0$ jets. These large-R jets then are trimmed using $f_{\text{cut}} = 0.05$ and $R_{\text{trim}} = 0.2$. Afterwards, the tracks of the event are

matched to the remaining $R = 0.2$ subjets using ghost association [86, 87]. Ghost association iterates through the list of all tracks to create a copy of the track in question with its transverse momentum set to $p_T = 1\,\text{eV}$. This ghost track is appended to the list of inputs for jet finding and the $R = 0.2$ jet clustering is repeated. If the ghost track is clustered into a $R = 0.2$ subjet that remains after the Trimming procedure, it is associated to this jet. Should any unassociated tracks remain, they are associated to the nearest subjet if the jet axis lies within $\Delta R_{\text{jet,track}} < 0.3$ of the track.

Since only charged particles leave tracks in the detector, the information of all neutral particles is lost at this point. To correct the TAR jet for this loss of information and keep the four-momentum of the large-R jet intact, the three-momentum of each track is scaled according to

$$\vec{p}^{\,\text{track,new}} = \kappa \cdot \vec{p}^{\,\text{track,old}}, \tag{4.10}$$

where $\kappa$ is calculated from the fraction of the subjet's $p_T$ over the sum of the tracks' $p_T$

$$\kappa = \frac{p_T^{\text{jet}}}{\sum\limits_{i \in \text{jet}} p_T^{\text{track},i}}. \tag{4.11}$$

The track's energy is consequentially set to $E^{\text{track}} = \sqrt{\kappa^2 \left(\vec{p}^{\,\text{track,old}}\right)^2 + \left(m^{\text{track}}\right)^2}$. The trimmed large-R jet is finally called TAR jet, whose constituents are the associated and rescaled tracks. Any substructure variable is calculated using these constituent tracks.

On truth level, the concept of tracks does not exist, since no detector simulation is deployed. In the truth level analysis in this thesis, tracks are equated to charged stable particles.

How the angularities $\lambda_0^0$ and $\lambda_0^2$ evolve over these steps is shown in Figure 4.2.

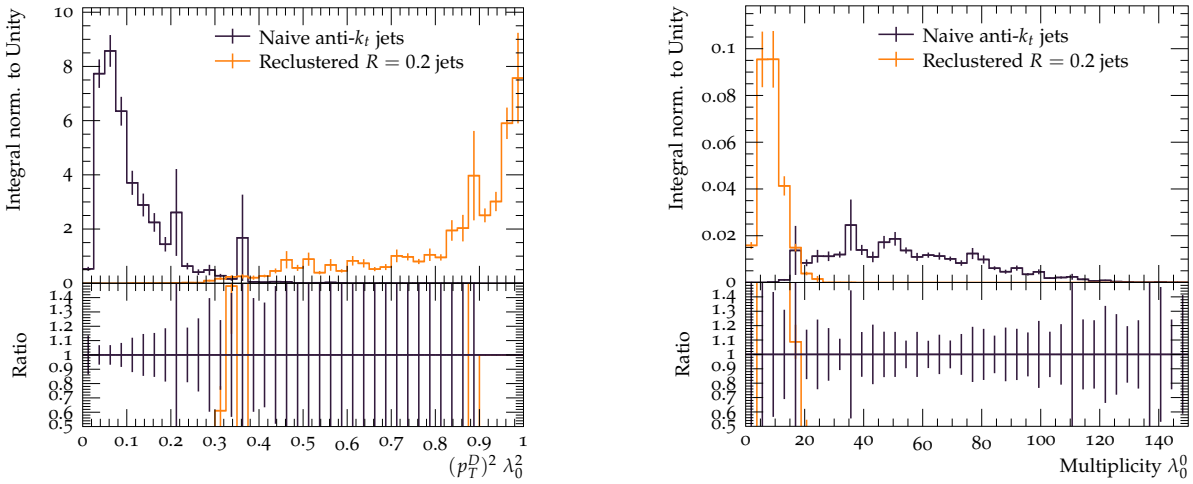


**Figure 4.2(a):** Naive anti-$k_T$ jets against reclustered jets from small-$R$ jets.

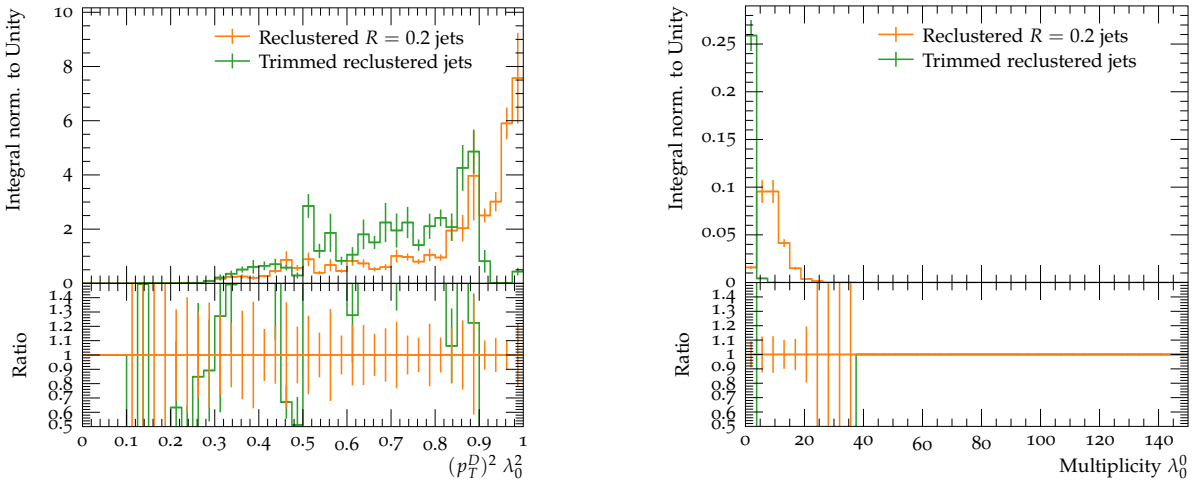


**Figure 4.2(b):** Reclustered jets from small-$R$ jets against reclustered and trimmed jets.

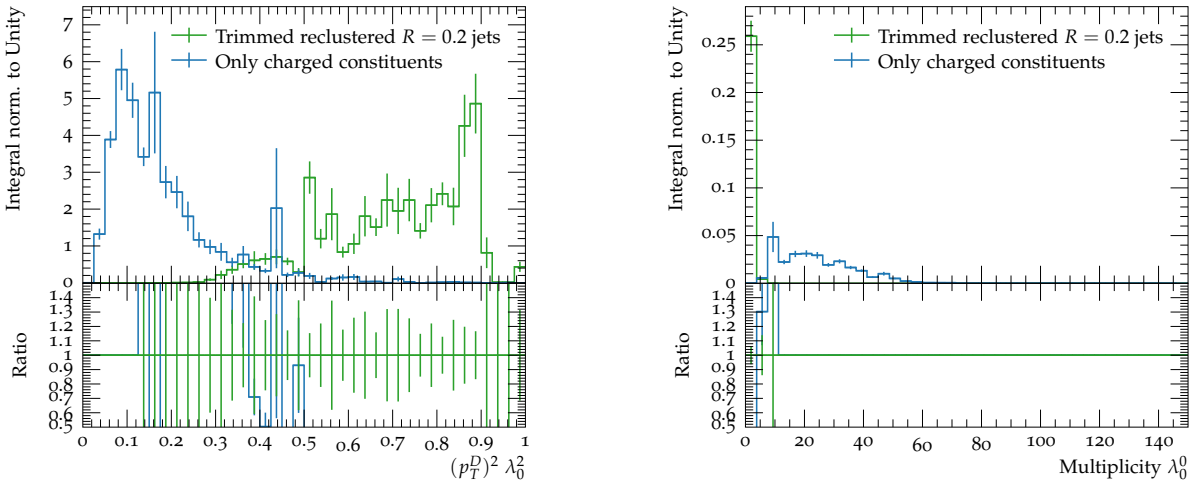


**Figure 4.2(c):** Reclustered and trimmed jets against jets built from ghost associated tracks.

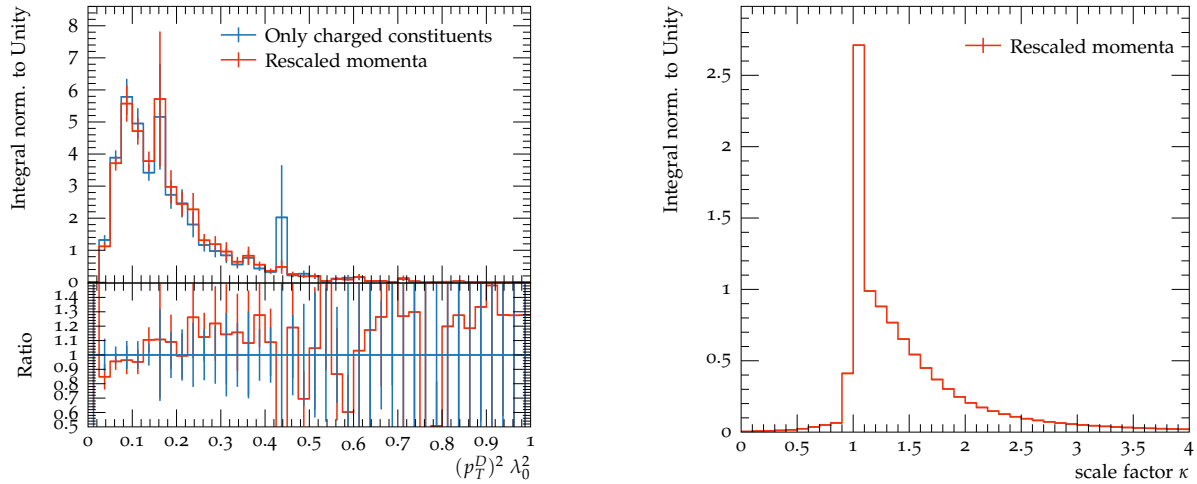


**Figure 4.2(d):** Jets built from ghost associated tracks against the jets with rescaled momenta.

**Figure 4.2.:** Evolution of $\lambda_0^0$ and $\lambda_0^2$ angularities over the steps of the TAR algorithm, as well as the scale factor $\kappa$ distribution. The algorithm is applied to QCD jets in a $p_T$ range between 1 TeV and 4 TeV.

The evolution starts from a naive anti-$k_T$ jet with radius $R = 1.0$ (black distributions), where numerous constituent final state particles carry a relatively low fraction of the jet's momentum individually.
This changes drastically when considering jets that are clustered from smaller anti-$k_T$ jets with radius $R = 0.2$ (orange distributions). The number of constituent subjets is of course lower than the number of constituent final state particles, which is very clear in the multiplicity distributions. The momentum of the jet is more concentrated in the subjets, causing the shift to higher $(p_T^D)^2$ values.
When these jets are trimmed (green distributions), the number of constituent subjets is further cut down drastically, since many of the subjets are soft. Because the surviving subjets are all of a higher $p_T$, narrowing down their $p_T$ distribution, the high values of the $(p_T^D)^2$ distribution are cut off.
Considering the tracks that are ghost associated to the subjets in the next step (blue distributions), the jet's momentum is divided onto these more numerous tracks, shifting the $(p_T^D)^2$ distribution to lower values again. The number of tracks is also higher than that of the subjets, increasing the multiplicity of constituent objects.
In the final step, the rescaling of the tracks' momenta (red distributions), the constituents and therefore structure of the jets are not changed any more, since only the tracks' momenta are rescaled by the scale factor $\kappa$. Because the rescaling happens for all tracks, the distribution of the jets' momenta onto their constituents is similar to the preceding step. The $(p_T^D)^2$ distribution therefore exhibits almost no change. In the shown $\kappa$ distribution, it is visible, that most of the momenta are rescaled with a factor of $\kappa \approx 1$. Values of $\kappa < 1$ are, due to its definition (Equation 4.11), very rare and only occur when tracks are matched to a subjet which itself did not contain the tracks originally.

### 4.2.3. Unified Flow Object (UFO) Jets

As a new and alternative object to simple calorimeter clusters that also uses tracking information, the ATLAS collaboration introduced the Unified Flow Object (UFO) [8]. An illustration of the algorithm can be seen in Figure 4.3.
The algorithm uses tracking information from the Inner Detector, Particle Flow Objects (PFOs) [88], Topological Clusters [89] and Track-CaloClusters [90] as input. All these objects try to obtain as much and as accurate information about the final state particles from the detector output as possible. For the truth level studies in this thesis, the specifics of combining the different inputs into one UFO, while avoiding double counting, can be ignored. The truth level relevant procedures are limited to pile-up mitigation and grooming of the clustered jet. The pile-up mitigation consists of two separate cases. For

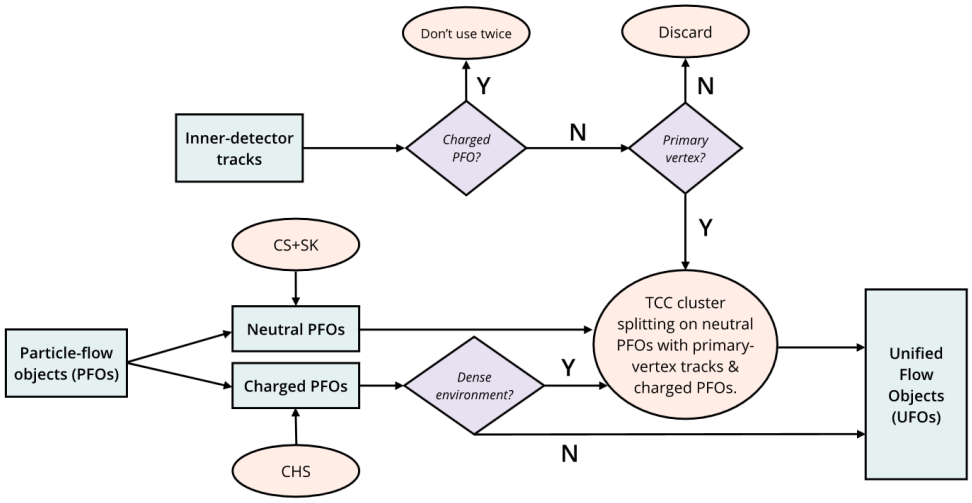


**Figure 4.3.:** Flow Chart of the unified flow object reconstruction algorithm, as implemented by the ATLAS collaboration. Figure taken from Ref. [8].

charged PFOs, Charged Hadron Subtraction (CHS) [91] is utilized. CHS removes the charged PFOs that are not matched to the primary vertex. For neutral PFOs, on the other hand, the Constituent Subtraction (CS) and Soft Killer (SK) algorithms are used. A nice feature of the CS algorithm is that it can directly be applied in the same way on Monte Carlo events as well as on detector objects like PFOs. The resulting jets can then be groomed. The Soft Drop algorithm is chosen for jet grooming for UFO studies in this thesis.

The aforementioned CHS does not make sense on truth level. The matching of particles to the primary vertex is 100% accurate on truth level and CHS would be equal to just turning off any radiation that does not come from the hard interaction. However, this does not reflect the real performance of CHS, due to detector inaccuracies. For this reason the pile-up mitigation method is limited to CS+SK in this thesis, which is applied to all the particles in the event that are used for jet finding.

The following grooming and pile-up mitigation algorithms are implemented in the Rivet analysis via the FastJet package [85].

### Constituent Subtraction and Soft Killer (CS+SK)

**Constituent Subtraction** [92] is a local per-particle subtraction of pile-up. The contamination due to pile-up is described by the transverse momentum density $\rho$ and mass density $\rho_m$. These quantities are estimated by forming patches, utilizing the $k_T$ jet clustering algorithm [93] with R=0.4, with a specific area $A_{\text{patch}}$ in $y$-$\phi$ space. The transverse

momentum and mass of the constituents in these patches are summed into $p_{T_{\text{patch}}}$ and $m_{\delta_{\text{patch}}}$, respectively. The densities are then formed by the median over the patches of the ratios $p_{T_{\text{patch}}}/A_{\text{patch}}$ and $m_{\delta_{\text{patch}}}/A_{\text{patch}}$, respectively. In the next step, the event is overlaid with ghost particles that cover a certain area $A_g$. A ghost's transverse momentum $p_T^g$ and mass $m_\delta^g$ is set to the product of the ghost's area with the respective densities $\rho$ and $\rho_m$:

$$\begin{aligned} p_T^g &= A_g \cdot \rho, \\ m_\delta^g &= A_g \cdot \rho_m. \end{aligned} \tag{4.12}$$

After defining a distance measure $\Delta R_{i,k}$ and creating a list of distances between each ghost $k$ and particle $i$, an iterative procedure subtracts the ghost's transverse momentum and mass from that of the particle. The subtraction begins with the particle-ghost pair with the lowest distance and works itself down to the pair with the highest distance. The algorithm can be stopped once there are only pairs left with $\Delta R_{i,k} > \Delta R_{\text{max}}$, where $\Delta R_{\text{max}}$ is a set threshold. The other parameters to be controlled are the ghost area $A_g$ and the parameter $\alpha$, which plays a role in the definition of the distance measure.

For the UFO objects in this thesis the parameters are set to be $A_g = 0.01$, $\alpha = 0$ and $\Delta R_{\text{max}} = 0.25$.

**Soft Killer** [94] is a very simple algorithm that further removes soft particles. A $p_T$-cut is applied and all particles with a transverse momentum lower than this threshold are removed. The characteristic feature of SK lies in the choice of this threshold. Similar to the CS algorithm, the event is divided into patches and the transverse momentum density $\rho$ is set by taking the median over these patches of the fraction $p_{T_{\text{patch}}}/A_{\text{patch}}$, where $p_{T_{\text{patch}}}$ is the sum of the particles' transverse momenta in each patch. The $p_T$-cut is then chosen such that $\rho$ becomes 0, after SK is applied. Thus, half of the patches are empty after the procedure has been applied. In contrast to the CS algorithm the patches consist of uniform squares in $y$-$\phi$ space. The length is here set to $\ell = 0.6$ in $y$-$\phi$.

### Soft Drop algorithm

The Soft Drop (SD) [95] algorithm works on an already clustered jet of radius $R_0$. The constituents of this jet are then reclustered using the Cambridge/Aachen jet clustering algorithm to form a pairwise clustering tree with an angular-ordered structure. The clustering steps are then undone, breaking the original jet $j$ into two subjets $j_1$ and $j_2$.

The subjets are tested in the Soft Drop condition

$$\frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{\text{cut}} \left( \frac{\Delta R_{12}}{R_0} \right)^{\beta}, \tag{4.13}$$

where $p_{T1}$ and $p_{T2}$ are the transverse momenta of $j_1$ and $j_2$, respectively, $\Delta R_{12}$ is their distance in the rapidity-azimuth plane, $\beta$ and $z_{\text{cut}}$ are parameters controlling the handling of angular and soft radiation, which are here set to $\beta = 1$ and $z_{\text{cut}} = 0.1$.
If the subjets pass this condition, $j$ is the final SD jet. If the condition fails, then the softer subjet with lower $p_T$ is dropped, and the subjet with higher $p_T$ is redefined to be the jet $j$ before the next clustering step is undone. If $j$ is a singleton and cannot be declustered, then $j$ is kept as the SD jet.
In contrast to Trimming, the SD algorithm not only removes soft, but also wide-angle radiation.
su

# 5. Results of Truth Level Study

After implementing the jet objects in the RIVET analysis, their behaviour is studied and compared against each other. Section 5.1 describes the steps taken in the analysis to select the jet of the hadronically decaying $W$ boson of the semi-leptonic boosted $H \to WW^*$ decay in the Higgs-Strahlung process $Z^* \to ZH \to b\bar{b}WW^*$. The selection steps are oriented on the preselection of the ATLAS $X \to HH(/SH) \to b\bar{b}WW^{(*)}$ analysis [9] (also described in Section 6.1), in order to compare the two signal processes in Chapter 6.
The implementation of the jet objects in the RIVET analysis is validated for QCD jets in Section 5.2, followed by the investigation of the jet objects performance in discriminating the Higgs-Strahlung process against QCD background in Section 5.3. Effects of clustering the lepton originating from the leptonically decaying $W$ boson into the jet of the hadronically decaying $W$ boson are observed in Section 5.4, while Section 5.5 shows the stability of the jet objects under systematic variations.

## 5.1. Event Selection

The main focus of the studies using Higgs-Strahlung is the jet corresponding to the hadronically decaying $W$ boson. To identify the jet successfully, the RIVET analysis goes through a series of steps. The first object that is identified is the lepton originating from the leptonically decaying $W$ boson, which will be called signal lepton in the following. As tau leptons are unstable, only electrons and muons are considered. Since the only leptons in the generated event that are not the signal lepton, can come from hadron decays, the signal lepton is required to not originate from a hadron. On truth level, this information is directly accessible, since the decay chains are available. The signal lepton also has to pass the conditions of $p_T > 10\,\text{GeV}$ and $|\eta| < 2.5$. The event is only accepted if exactly one signal lepton is found, otherwise the event is vetoed. The neutrino that is closest to this signal lepton is identified as the signal neutrino. Before the analysis continues with the object reconstruction, a trigger condition is applied. Anti-$k_T$ jets with radius $R = 1.0$ are clustered and trimmed with $f_{\text{cut}} = 0.05$. If at least one of the found jets has $p_T > 500\,\text{GeV}$, the event is kept, otherwise it is vetoed.

After the trigger condition is passed, the introduced jet objects are reconstructed, and at least two jet objects need to be found for the event to be selected. The jet originating from the hadronically decaying $W$ boson is identified as the jet that is closest to the lepton. In the following, it will be called the $W_{\text{had}}$ jet. The jet corresponding to the decaying $Z$ boson is the hardest jet in the event that has not been identified as the $W_{\text{had}}$ jet already. The final conditions that are checked for both the $W_{\text{had}}$ and $Z$ jet are that they satisfy $p_T > 100\,\text{GeV}$, $|\eta| < 2.0$ and have at least two jet constituents.
For QCD background events, the RIVET analysis selects the harder of the (typical) two jets in the event that also obeys the criteria of having a transverse momentum of $p_T > 100\,\text{GeV}$, being in a region of rapidity $|\eta| < 2.0$, and consisting of at least two constituents.
Because of the different selection schemes for QCD and signal jets, the $p_T$ distribution of these two processes will not be the same. This is not only caused by the events' kinematics, but is specifically amplified by the condition that there has to be a jet of $p_T > 500\,\text{GeV}$ in the signal event that does not necessarily have to be the $W_{\text{had}}$ jet. While the $W_{\text{had}}$ jet can have $p_T < 500\,\text{GeV}$, such jets are suppressed by the trigger condition. For jets with $p_T > 500\,\text{GeV}$, the QCD and signal jets become directly comparable. For this reason, the shown plots in this chapter are of jets with a $p_T$ between 500 and 1500 GeV. The distributions are normalized to be able to directly compare QCD jets against signal jets without worrying about different production rates.

## 5.2. Validation for QCD Jets

To validate the implementation of the jet objects and inspect their behaviour on typical background events, QCD events are generated and evaluated as described in Section 3.3.1. In Figure 5.1, the transverse momentum as well as the mass of the different QCD jet objects are depicted. In both histograms, the TAR jets differ more from UFO jets than from standard LRJs. While the deviations in the $p_T$ distribution are more subtle, the differences in the object's algorithms become clear in the mass distribution. The common factor between the TAR jets and standard LRJs is the jet grooming via Trimming, while the UFO jets use Soft Drop. The Soft Drop algorithm in combination with CS+SK pile-up mitigation cleans the jet in a more thorough way than Trimming, and therefore also reduces the mass of the jet more than Trimming.
Of more interest are the substructure observables. All three algorithms remove and rescale the constituents of the jets in different ways. By looking at the substructure observables, one can deduce if these procedures keep the substructure intact, worsen or even improve it. The subjettiness $\tau_{21}$ and ECF ratio $C_2^1$ are shown in Figure 5.2. While $\tau_{21}$, as well

as the other substructure observables (see Appendix A.1), agree with each other for the most part, $C_2^1$ exhibits more deviations for the different jet objects and is very sensitive to the constituents. In comparison to the standard LRJ, the other two objects flatten the distribution. How this affects the discrimination power of the observable is studied in the next subsection.

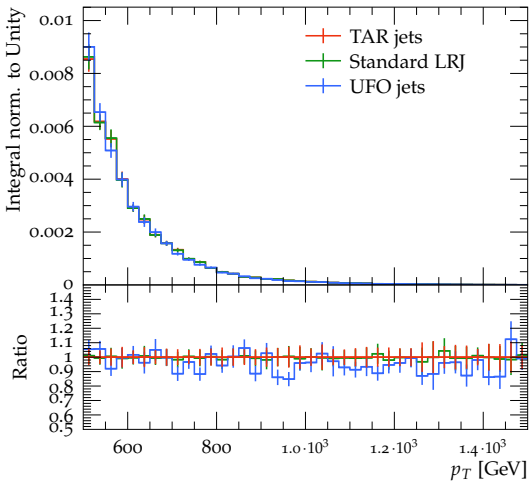


**Figure 5.1(a):** Jet $p_T$

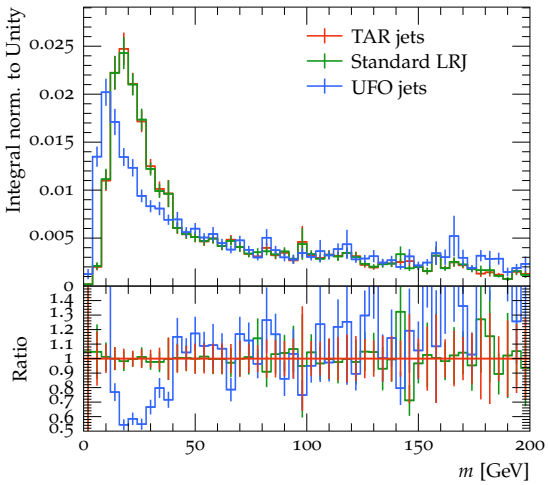


**Figure 5.1(b):** Jet mass

**Figure 5.1.:** Kinematic observables for TAR jet, UFO jet and standard LRJ objects on QCD samples.

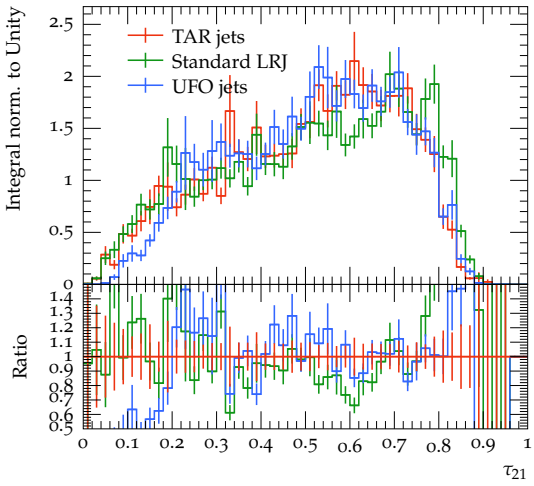


**Figure 5.2(a):** Subjettiness $\tau_{21}$

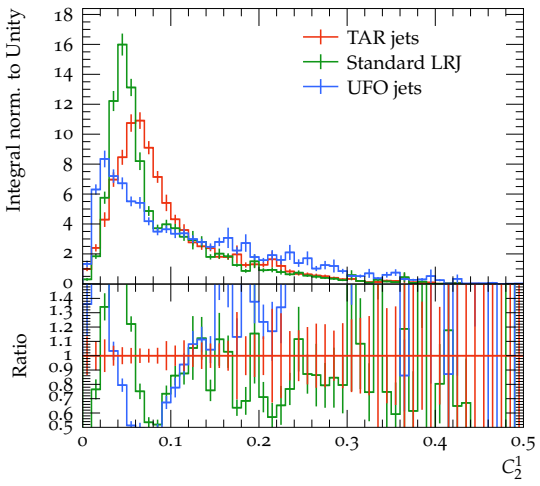


**Figure 5.2(b):** ECF ratio $C_2^1$

**Figure 5.2.:** Substructure observables for TAR jet, UFO jet and standard LRJ objects on QCD samples.

## 5.3. Discrimination of Boosted $W$-decays against QCD Background

The goal of sophisticated jet objects with high spatial resolution is to increase the ability to describe specific signals and distinguish it from any background processes that mimic the signal. To study this ability and any possible differences in the discrimination, the QCD samples are compared to the boosted $W$ decays in the Higgs-Strahlung process described in Section 3.3.1. The observables with the biggest differences between the signal and QCD jets are shown in Figure 5.3.

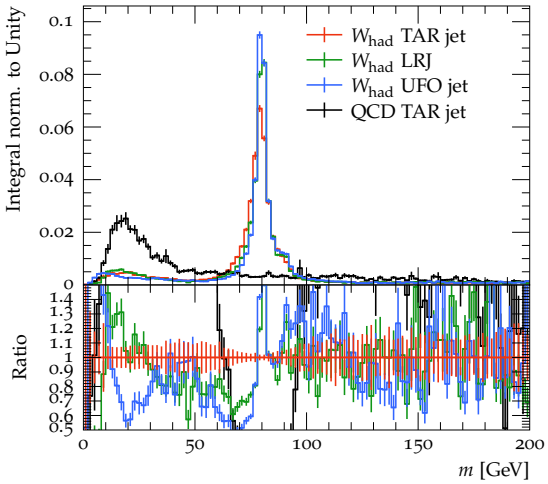


**Figure 5.3(a):** Jet mass

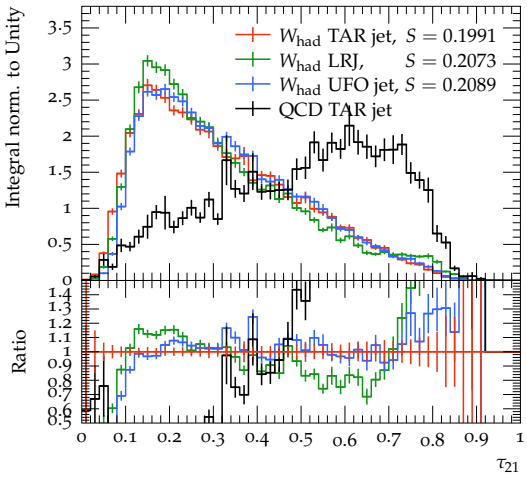


**Figure 5.3(b):** Subjettiness $\tau_{21}$

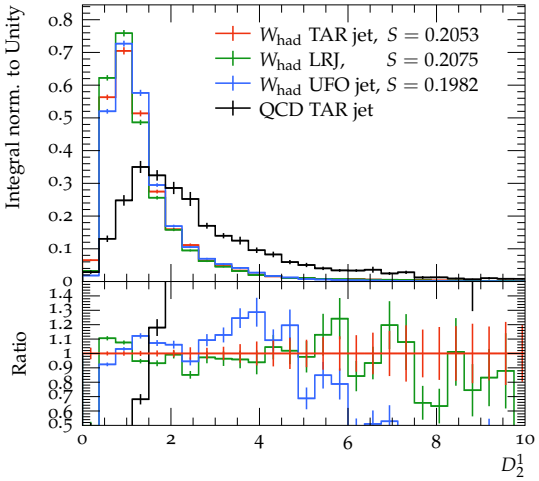


**Figure 5.3(c):** ECF ratio $D_2^1$

**Figure 5.3.:** Most discriminating observables for TAR jet, UFO jet and standard LRJ objects on Higgs-Strahlung samples compared against QCD TAR jets. The separation power $S$ is calculated for the signal distributions of the jet objects against their own respective QCD distributions.

The obvious difference between the QCD and signal process lies in the mass distribution shown in Figure 5.3(a). While the QCD jets favour masses around 20 GeV in a very broad distribution, the $W_{\text{had}}$ jets, produced by a resonance, are mostly centred around the $W$ mass of $m_W \approx 80\,\text{GeV}$. The UFO jets exhibit the sharpest distribution, while the TAR jets feature the broadest. A secondary and small bump in the distribution of the $W_{\text{had}}$ jets are located at light masses, similar to the QCD jets. These exist because the Higgs boson is lighter than the combined mass of the two $W$ bosons that are created in its decay. One of the $W$ bosons therefore cannot be on-shell. Because lighter particles receive a smaller boost than more massive particles in the same decay, less jets of off-shell $W$ bosons are present in the considered 500 GeV to 1500 GeV $p_T$ bin.

To quantify the ability of the jet objects to discriminate the Higgs-Strahlung process against the QCD background for different observables, the separation power $S$ is calculated

$$S = \frac{1}{2}\sum_i \frac{(S_i - B_i)^2}{S_i + B_i}. \tag{5.1}$$

Here, $S_i$ is the number of signal events and $B_i$ is the number of background events in bin $i$ of a histogram that is normalized to unity.

The subjettiness $\tau_{21}$ demonstrates the 2-prongness of the hadronic $W$ decay very well. For a two prong jet the subjettiness $\tau_1$ gravitates towards 1, while $\tau_2$ gravitates towards 0, resulting in small values for $\tau_{21}$. This is also seen for the three different jet objects in Figure 5.3(b). The QCD jets with a 1-prong substructure feature $\tau_{21}$ values closer to 1. This makes $\tau_{21}$ a good measure to discriminate the two processes from each other, independent of the studied jet objects. To create less visual noise in the plots, the signal jets are only compared against QCD TAR jets. The comparison of each signal jet object against its own QCD counterpart can be found in Appendix A.2.

For the subjettiness $\tau_{21}$, this yields a separation power of $S_{\text{LRJ},\tau_{21}} = 0.2073$ for standard LRJs, $S_{\text{UFO},\tau_{21}} = 0.2089$ for UFO jets and $S_{\text{TAR},\tau_{21}} = 0.1991$ for TAR jets. While the values are of similar magnitude, the separation power is largest for UFO jets.

The ECF ratio $D_2^1$ in Figure 5.3(c) is the other substructure observable with a large difference between the signal and QCD background process. The signal process has most of its events at very low values of $D_2^1$, while the QCD process exhibits a broader distribution that is shifted to the right. The peak of the QCD and signal distributions are very close to each other, making the discrimination using a simple cut very difficult. Just like for $\tau_{21}$, the direct comparison between signal and QCD background with their separation power can be found in Appendix A.2. It shows that again the separation power does not differ greatly for the different jet objects, but for $D_2^1$, the standard LRJs have the largest

separation power with $S_{\mathrm{LRJ},D_2^1} = 0.2075$.

In Figure 5.4, the signal ECF ratio $C_2^1$ for the TAR jet, UFO jet and standard LRJ objects are compared against the corresponding QCD distributions. The differences between the jet objects, seen in the QCD distributions in Section 5.2 for $C_2^1$, also apply to the $W_{\mathrm{had}}$ distributions. The signal and background distributions are almost identical, making the observable virtually useless to distinguish the two processes. Only the signal distribution for the UFO jets is shifted to the right of the QCD distribution and shows the largest separation power with $S_{\mathrm{UFO},C_2^1} = 0.0707$. For the ECF ratio $C_2^1$, the separation power of the UFO jets is more than three times as large as for the other jet objects.

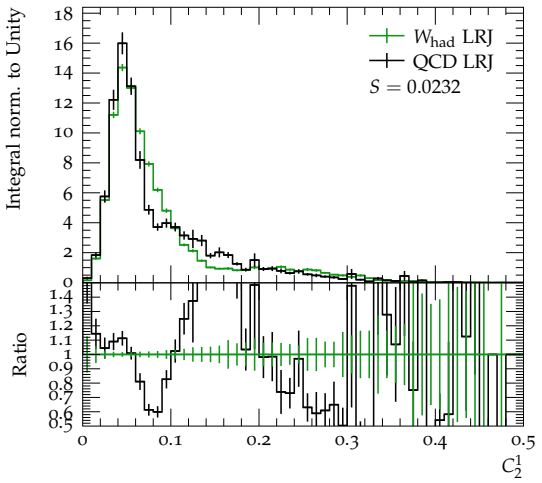


**Figure 5.4(a):** Standard LRJ

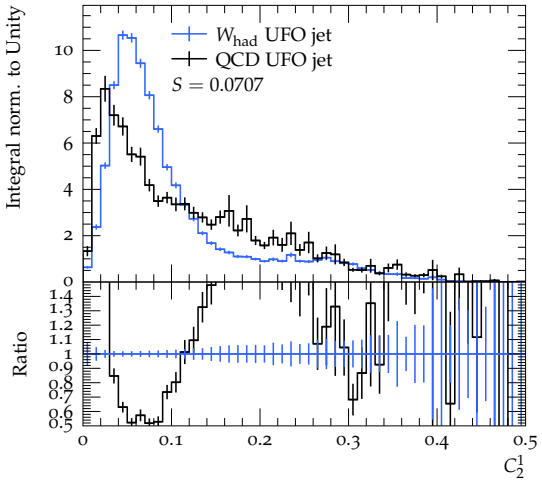


**Figure 5.4(b):** UFO jet

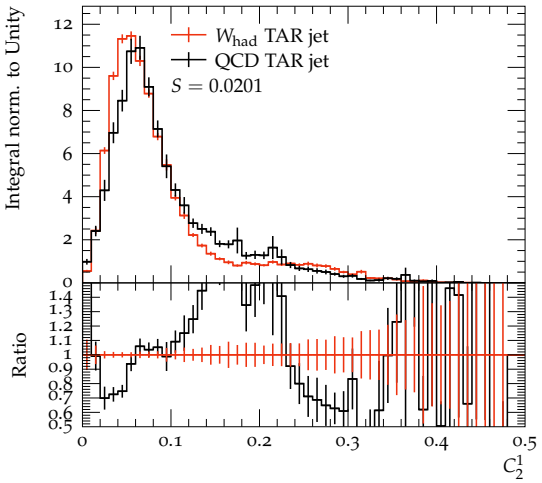


**Figure 5.4(c):** TAR jet

**Figure 5.4.:** ECF ratio $C_2^1$ for TAR jet, UFO jet and standard LRJ objects on Higgs-Strahlung samples compared against their QCD counterpart.

## 5.4. Lepton Clustering in Higgs Decay

One difficulty of reconstructing the chosen signal topology is that the signal lepton often overlaps with the $W_{\text{had}}$ jet, as shown in Figure 2.12. If the leptons are carelessly included in the jet clustering algorithms, they will change the signature of the event drastically. Experiments must use procedures to remove the signal lepton from the jet finding or the jets. In this section, performance differences are studied for when the signal lepton is included in jet finding compared to when it is not included in jet finding. It is further assumed that, for TAR jets, the lepton can and will be removed from the jet, once clustered inside.

For the case where the signal lepton is a muon, it will directly be subtracted from the TAR jet's constituents after reclustering the $R = 0.2$ jets, making use of truth information. This basically transforms the TAR jet into a version where the muon was never included in the jet finding beforehand, making the study of the muon removal on truth level redundant and will not be considered further.

For processes, where the leptonically decaying $W$ boson decays into an electron and a neutrino, the subjet that overlaps geometrically with the electron will be removed. This will remove more energy from the TAR jet than what was added by the electron, raising the question if this technique is still appropriate to identify the signal process.

This procedure removes the electron 100% of the time from the TAR jets, as can be seen in Table 5.1, while the UFO jets and standard LRJs remove the electron around 3.5% of the time during the grooming procedure. The Soft Drop algorithm in UFO jets removes the electron as often as Trimming in standard LRJs.

| Jet object | Probability |
|---|---|
| TAR | 100% |
| UFO | 3.5% |
| Standard LRJ | 3.5% |

**Table 5.1.:** Probability for the signal electron to be removed from the jet, either by grooming techniques (UFO jets and standard LRJs) or explicit overlap removal (TAR jets).

The following plots are produced for a signal electron that is either included or ignored in the jet finding, respectively removed together with the overlapping subjet for TAR jets. The separation power $S$ is calculated between the $W_{\text{had}}$ distributions and the QCD distribution.

Accidentally clustering the signal lepton into the jet completely ruins the structure of the signal, as can be seen in Figures 5.5–5.7.

The mass distributions (Figure 5.5) for UFO and standard LRJs are shifted to the right and smeared out, as the energy of the signal electron is added to the jet.

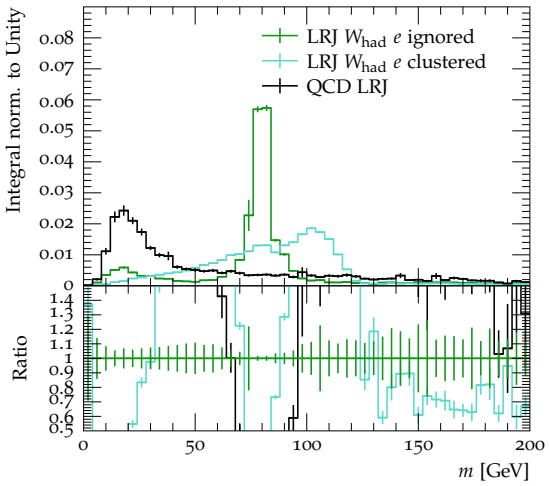


**Figure 5.5(a):** Standard LRJ

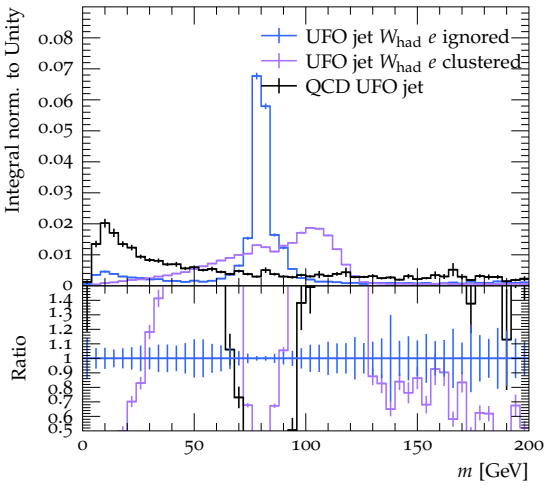


**Figure 5.5(b):** UFO jet

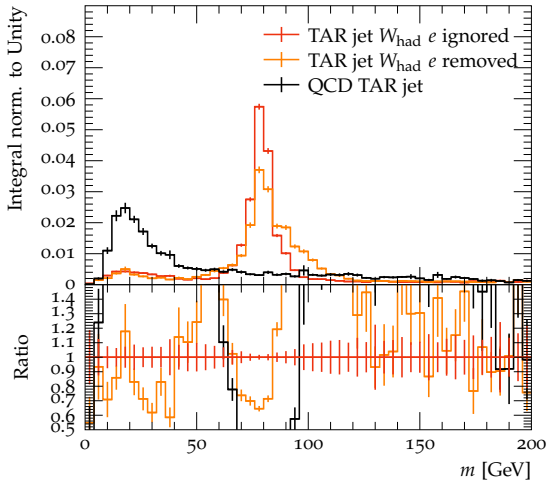


**Figure 5.5(c):** TAR jet

**Figure 5.5.:** Mass distribution for UFO, TAR and standard LRJ on signal samples compared to respective QCD jets. The signal electron is either ignored in the jet finding or included, respectively removed for TAR jets. The separation power $S$ is calculated between the $W_{\text{had}}$ distributions and the QCD distribution.

The shift of the distributions in Figure 5.6 reduce the separation power of $\tau_{21}$ by less than 3% for standard LRJs and less than 19% for UFO jets. While the separation power shrinks, it still can be used to distinguish the process from QCD background.
However, $\tau_{42}$ in Figure 5.7 gains a lot of separation power, as the jet has essentially gained an additional prong. While $\tau_{21}$ is sensitive to 2-prong jets, $\tau_{42}$ is sensitive to 3-prong jets. With a separation power of up to $S_{\text{UFO},\tau_{42}} = 0.4616$ for UFO jets, $\tau_{42}$ provides the largest

separation power with the clustered signal lepton.

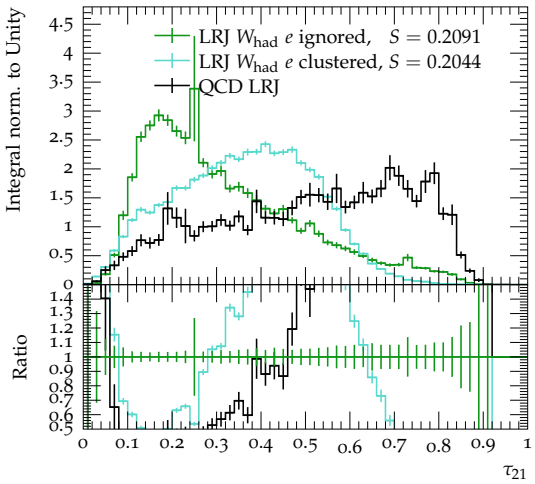


**Figure 5.6(a):** Standard LRJ

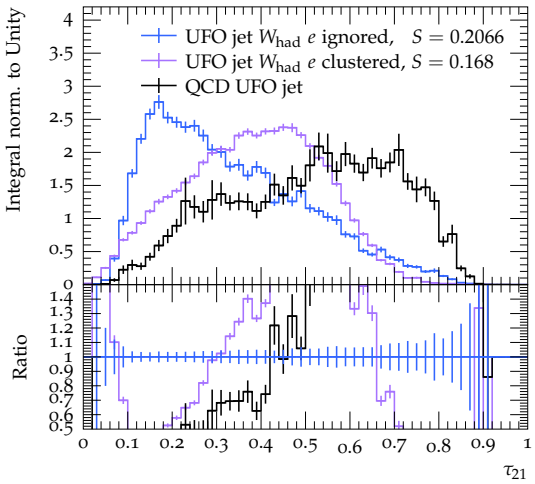


**Figure 5.6(b):** UFO jet

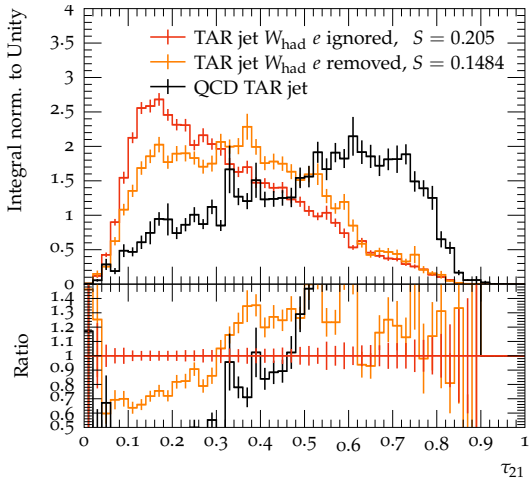


**Figure 5.6(c):** TAR jet

**Figure 5.6.:** Subjettiness $\tau_{21}$ for UFO, TAR and standard LRJ on signal samples compared to respective QCD jets. The signal electron is either ignored in the jet finding or included, respectively removed for TAR jets. The separation power $S$ is calculated between the $W_{\text{had}}$ distributions and the QCD distribution.

For TAR jets, the removal of the subjet overlapped by the signal electron works quite well to regain the original, best case distributions. While some resolution in the mass reconstruction is lost, there remains a clear peak at the $W$ mass. Similar changes can be seen for the subjettiness distributions. The features become smeared out, but stay mainly intact and still provide separation power. Like for the UFO jet and standard LRJ objects, the separation power for TAR jets and the subjettiness $\tau_{21}$ is reduced, for $\tau_{42}$ it

is increased. The reduction for $\tau_{21}$ is the largest for the three jet objects with more than 27%, while $\tau_{42}$ only reaches $S_{\mathrm{TAR},\tau_{42}} = 0.1269$.
Additional observables can be found in Appendix A.3.

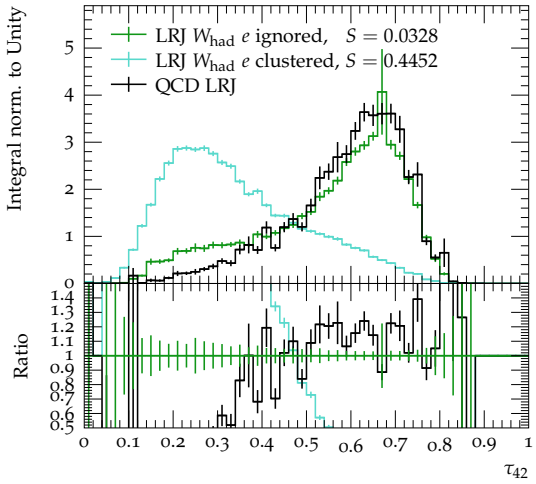


**Figure 5.7(a):** Standard LRJ

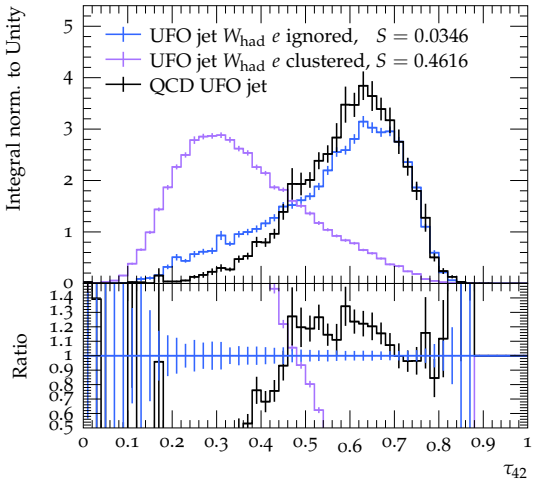


**Figure 5.7(b):** UFO jet

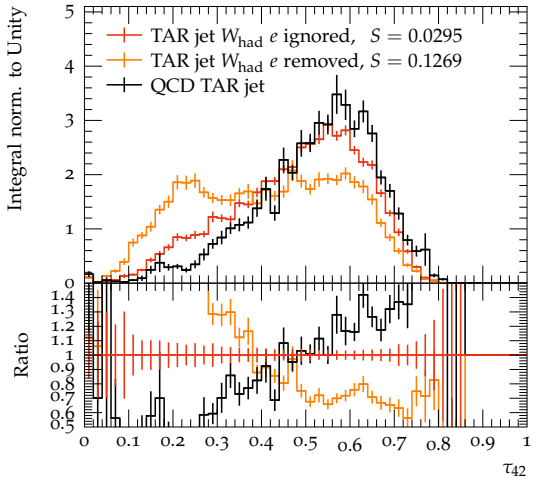


**Figure 5.7(c):** TAR jet

**Figure 5.7.:** Subjettiness $\tau_{42}$ for UFO, TAR and standard LRJ on signal samples compared to respective QCD jets. The signal electron is either ignored in the jet finding or included, respectively removed for TAR jets. The separation power $S$ is calculated between the $W_{\mathrm{had}}$ distributions and the QCD distribution.

## 5.5. Systematic Variations

To study the stability of the objects under systematic variations, the SHERPA generator settings are varied. While most settings remain the same as for the Higgs-Strahlung generation, the settings for the MPI, that were enabled before, and the hadronization module is modified.
In particular the MPI are toggled between on and off and the hadronization module is changed between the cluster fragmentation model implemented in AHADIC++ [61] and the Lund string hadronization model [28, 29], creating four different settings in total.
The toggling of the MPI is of special interest, since it shows the objects' ability to reduce background noise. To demonstrate this feature, a very minimalistic fourth jet object gets reconstructed in the RIVET analysis. The **naive jet** is an $R = 1.0$ anti-$k_T$ jet without any jet grooming or altering of the constituents. In Figure 5.8 the ECF ratio $C_2^1$ for UFO, TAR, standard LRJ and naive jets with the four different SHERPA settings are shown. This observable shows the largest separation for the different distributions of the naive jet.
For the naive jets, all the distributions are changed by the MPI, the distributions without MPI look similar to those of the other three jet objects. However, all added noise from the MPI is absorbed into the naive jet, resulting in big differences in the distributions with and without MPI. The different hadronization models have no significant influence on the distributions.
For the more sophisticated jet objects, there is virtually no difference for all four generator settings. The MPI is completely removed by the grooming and pile-up mitigation techniques, whether it is Soft Drop with CS+SK or Trimming. More observables can be found in Appendix A.4.

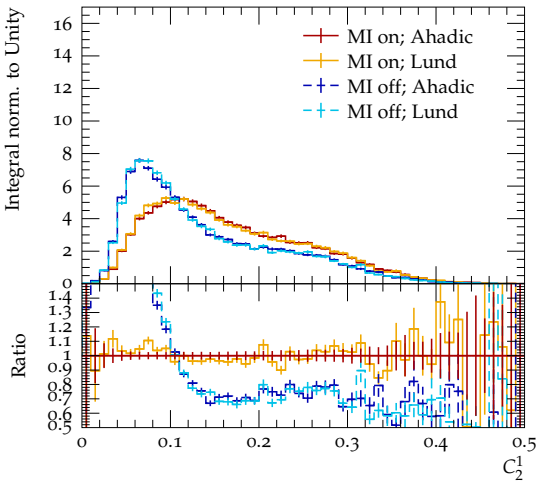


**Figure 5.8(a):** Naive jet

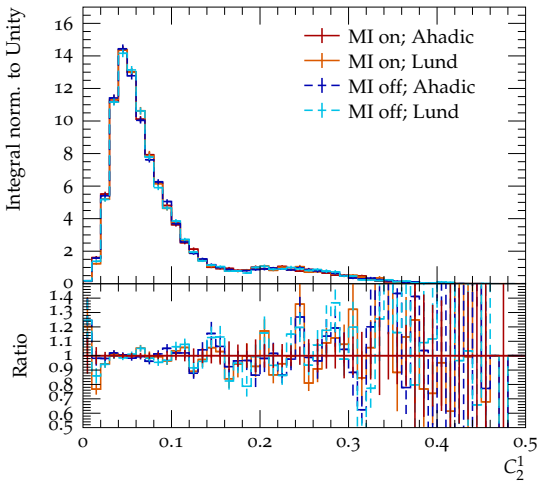


**Figure 5.8(b):** Standard LRJ

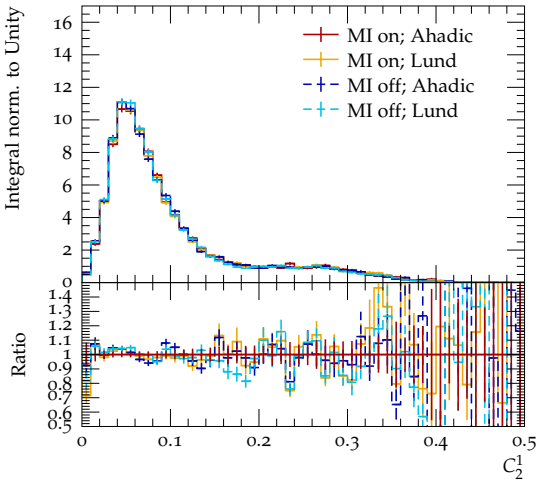


**Figure 5.8(c):** UFO jet

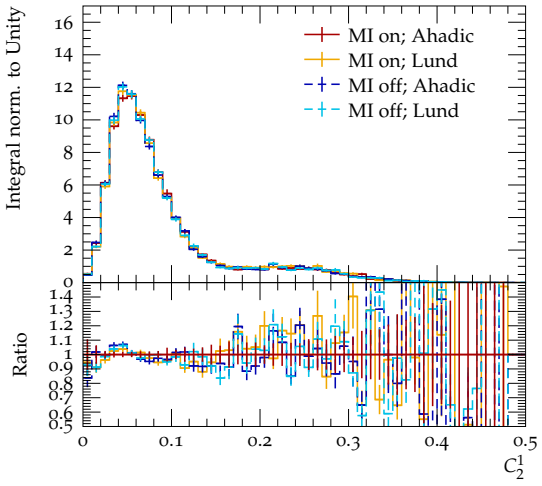


**Figure 5.8(d):** TAR jet

**Figure 5.8.:** ECF ratio $C_2^1$ for UFO, TAR, standard LRJ and naive jets on signal samples. The SHERPA settings for the MPI and hadronization are varied.

# 6. Studies for the ATLAS $b\bar{b}WW^*$ Analysis

The results of studies performed using data taken by the ATLAS detector and samples used in the ATLAS $X \to HH(/SH) \to b\bar{b}WW^{(*)}$ analysis are presented in this chapter. Relevant technicalities of the analysis are described in Section 6.1. The jet object comparison on truth level from Chapter 5 is continued in Section 6.2 using ATLAS samples. During the validation of new $V$+jets samples produced by Sherpa 2.2.11 in Section 6.3, the question arose whether the background modelling could benefit from normalization of the $Z$+jets background. This question is addressed in Section 6.4.

## 6.1. Event Selection

Events recorded by the detector and produced with MC generators are processed to select only those relevant to the analysis. Selection steps are taken in such a way that only interesting events are filtered out without making too strict cuts to not reduce the number of recorded events below a sensible threshold. To choose the events of interest, a preselection is applied. The selected events are then split into different phase-space regions, used to analyse data (signal region) and tuning MC (control region).
The preselection and phase-space regions used in this thesis are also introduced in Ref. [9].

### 6.1.1. Preselection

The preselection (or boosted selection) is applied to all events in the analysis and selects those that have the signature of boosted 1-lepton $X \to HH(/SH) \to b\bar{b}WW^{(*)}$ events. It works analogous to the selection steps described in Chapter 5.
The preselection begins with a single large-$R$ jet trigger that fires when at least one trimmed anti-$k_T$ jet with radius $R = 1.0$ and Trimming parameter $f_{\text{cut}} = 0.05$ with a sufficient $p_T$ is found. The $p_T$ threshold depends on the LHC run period and varies from 360 GeV to 480 GeV. The trigger is accompanied by the required on the leading jet to carry $p_T > 500$ GeV.

To include the signature of the leptonically decaying $W$ boson, exactly one signal lepton is required in the event. The kinematics of signal leptons must fulfil $p_T^\ell > 10\,\text{GeV}$ with $|\eta^\mu| < 2.5$ for muons and $|\eta^e| < 2.47$ for electrons with an exclusion region of $1.37 < |\eta^e| < 1.52$ corresponding to a poorly instrumented between calorimeter subsystems. The promptness of leptons is determined by requiring their tracks to be close to the beamspot position, which is defined as the average position of the $pp$ interactions. A lepton is classified as originating from the hard interaction when its track's product of the longitudinal impact parameter $z_0$ with the sine of the polar angle $\theta$ satisfies $|z_0 \sin\theta| < 0.5\,\text{mm}$, and its track's transverse impact parameter $d_0$ satisfies $|d_0^e/\sigma_{d_0^e}| < 5.0$ for electrons and $|d_0^\mu/\sigma_{d_0^\mu}| < 3.0$ for muons. In addition to these kinematic selection cuts, identification and isolation criteria are combined to form loose and tight selection working points of leptons. Only tight leptons are accepted as signal leptons.
The boosted topology is ensured by demanding the lepton to be very close to the $W_{\text{had}}$ jet with $\Delta R(\ell, \text{closest jet}) < 1.0$. Furthermore, there have to be at least two jets in the event, to ensure that the reconstruction of both the $H \to b\bar{b}$ and the $W_{\text{had}}$ jet is possible. The $W_{\text{had}}$ jet is chosen to be the closest jet to the signal lepton. The jet corresponding to the $H \to b\bar{b}$ decay is identified as the hardest remaining jet in the event.
Requiring $p_T^{H\to b\bar{b}} > 500\,\text{GeV}$ is found to increase the signal sensitivity for $m_X \gtrsim 1\,\text{TeV}$. Combined with the trigger condition, the leading jet is almost always classified as the $H \to b\bar{b}$ candidate.

### 6.1.2. Phase Space Regions and Background Channels

Signal regions (SR) are chosen to increase the purity of the signal and efficiency of the signal selection, while control regions (CR) are chosen to enhance the background contributions of the process of interest, in order to control and, if necessary, to correct backgrounds to fit data. The normalization is performed in a single bin in the control regions, since the shape of distributions are assumed to be well modelled by MC.

#### Signal Region

The considered signal region is defined by a window on the $H \to b\bar{b}$ mass, evaluated to contain 70% of $X \to HH \to b\bar{b}WW^*$ events and depicted in Figure 6.1(a). In addition, the $H \to b\bar{b}$ jet is required to be double $b$-tagged while no other jet in the event is $b$-tagged. $b$-tagging makes use of the long lifetime of $B$-hadrons and utilizes the resulting displaced, secondary vertices as an indication of $b$-jets. The signal region is named SRp2 after the passing of the mass window and the number of $b$-tags. Additional signal regions exist in

the analysis, but are not considered in this thesis, because SRp2 offers the highest signal efficiency of all signal regions.

The events are further classified into $SH$-like or $HH$-like signatures by making use of the ECF ratio $C_2^1 \coloneqq C_2$ and the distance between the $W_{\text{had}}$ jet and the lepton $\Delta R(W_{\text{had}}, \ell)$. Events are labelled $SH$-like if the $H \to b\bar{b}$ jet passes a $C_2$ window, evaluated to contain 80% of $X \to SH \to b\bar{b}WW$ events. This selection window ensures the jet is more 2-prong-like than 1-prong-like and shown in Figure 6.1(b). Events are labelled $HH$-like when additionally passing the $\Delta R(W_{\text{had}}, \ell)$ window displayed in Figure 6.1(c), evaluated to contain 80% of $X \to HH \to b\bar{b}WW^{(*)}$ events.

The $m^{H\to b\bar{b}}$ and $C_2^{H\to b\bar{b}}$ windows depend on $p_T^{H\to b\bar{b}}$, while the $\Delta R(W_{\text{had}}, \ell)$ window is dependent on the $p_T$ of the visible $H \to WW$ decay, $H_{\text{vis}} = W_{\text{had}} + \ell$.

The cuts made in the signal and control regions are summarized in Table 6.1.

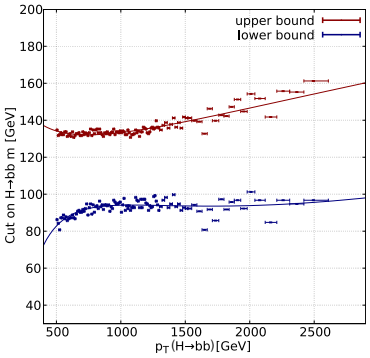


**Figure 6.1(a):** $m^{H\to b\bar{b}}$

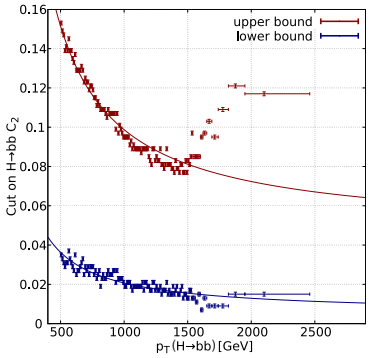


**Figure 6.1(b):** $C_2^{H\to b\bar{b}}$

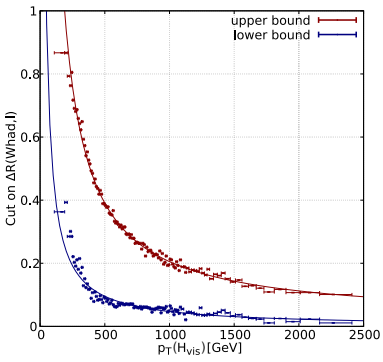


**Figure 6.1(c):** $\Delta R(W_{\text{had}}, \ell)$

**Figure 6.1.:** $m^{H\to b\bar{b}}$ window containing 70% of $HH$ events, $C_2^{H\to b\bar{b}}$ window containing 80% of $SH$ events, $\Delta R(W_{\text{had}}, \ell)$ window containing 80% of $HH$ events and corresponding fits. Figures adapted from Ref. [40].

### Control Regions

Three control regions for the three dominating background processes are constructed. The details of the $t\bar{t}$ CR are neglected here, since only the $W$+jets CR and QCD CR are relevant for this thesis. Each control region is orthogonal to the other control regions and the signal region by construction. Typical processes for the control regions of interest are sketched in Figure 6.2.

The $W$+jets CR is built around the kinematics of prompt lepton $W$ boson decays with additional jets in the final state. An absence of $b$-tagged jets in the event, and the transverse mass of the leptonically decaying $W$ boson, $W_{\text{lep}} = \ell + E_T^{\text{miss}}$, $m_T^{W_{\text{lep}}}$ satisfying $60\,\text{GeV} < m_T^{W_{\text{lep}}} < 120\,\text{GeV}$ are required.

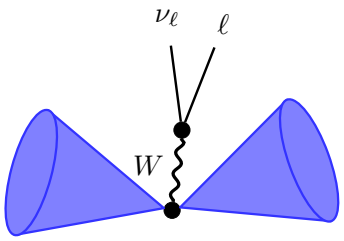


**Figure 6.2(a):** $W$+jets

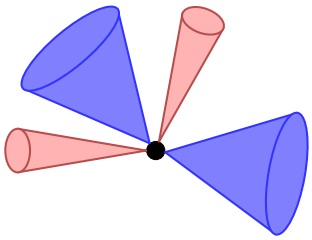

**Figure 6.2(b):** multijet

**Figure 6.2.:** Sketches of the main background processes.

The QCD CR is constructed to control the modelling of non-prompt lepton backgrounds. Non-prompt backgrounds are not simulated by MC, but estimated in a data-driven approach described in Ref. [9]. Since non-prompt backgrounds are dominated by QCD multijet events, this control region maximises the contribution of multijet events. An absence of $b$-tagged jets and $m_T^{W_{\text{lep}}} < 60\,\text{GeV}$ or $m_T^{W_{\text{lep}}} > 120\,\text{GeV}$ are required.

The cuts made in the signal and control regions are summarized in Table 6.1.

| Region | Cuts |
|---|---|
| $W$+jets CR | 0 $b$-tagged jets, $60\,\text{GeV} < m_T^{W_{\text{lep}}} < 120\,\text{GeV}$ |
| QCD CR | 0 $b$-tagged jets, $m_T^{W_{\text{lep}}} < 60\,\text{GeV}$ or $m_T^{W_{\text{lep}}} > 120\,\text{GeV}$ |
| SRp2 | 2 $b$-tagged jets, pass 70% $m^{H\to b\bar{b}}$ mass window |

**Table 6.1.:** Cuts chosen for the definition of the control regions and signal region considered in this thesis. All regions pass the preselection.

## 6.2. Jet Object Comparison

In this section, the ATLAS independent Sherpa samples are compared against the ATLAS samples. The biggest differences in the production and selection of these events are removed in order to gain insight into detector effects present in the ATLAS samples. Finally, UFO and TAR jets are compared against each other for ATLAS samples, similarly to the comparison in Chapter 5.

The best agreement to the kinematics of the mock signal, described in Section 3.3.1, is achieved with ATLAS $HH$-samples with $m_X = 1.2\,\text{TeV}$, which are used exclusively in this section.

### 6.2.1. Overlap Removal

In order to avoid energy double-counting, an overlap removal (OLR) is used. Tracks are rejected from TAR jet inputs if they are matched to either an electron or a muon. Since electrons deposit their energy in the calorimeters, they cannot be isolated in the detector readout. To remove the energy deposited in the calorimeters by electrons for TAR jets, a small-$R$ jet with radius $R = 0.2$ is rejected, if it overlaps with an electron, as this strongly suggests that the jet is actually an electron. For muons, the rejection of small-$R$ jets is not necessary, since they only deposit a small energy fraction in the calorimeters.
This part of the OLR has been already implemented in the SHERPA TAR jets and studied in Section 5.4.
However, the comparison between SHERPA and ATLAS samples in Figure 6.3 shows large differences in the distributions of the hadronically decaying $W$ boson's mass and $p_T$.

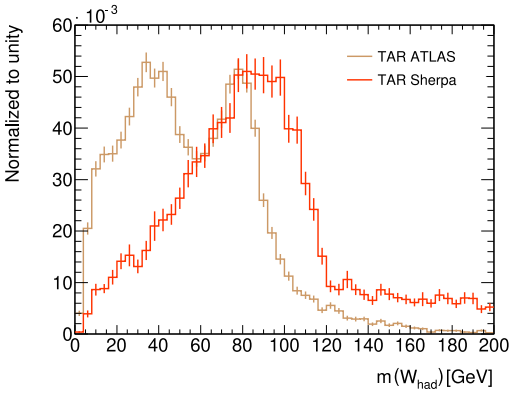


**Figure 6.3(a):** $m(W_{\text{had}})$

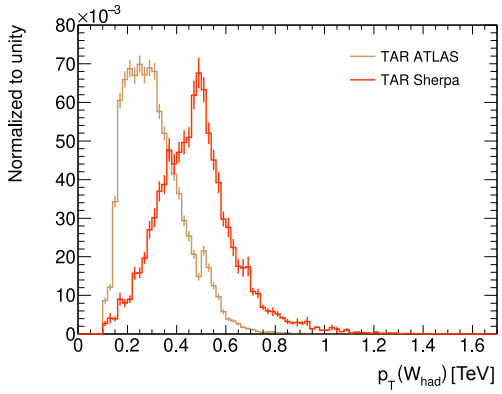


**Figure 6.3(b):** $p_T(W_{\text{had}})$

**Figure 6.3.:** Comparison of the mass and $p_T$ of the $W_{\text{had}}$ jet for TAR jet objects between the ATLAS and SHERPA samples.

The SHERPA samples feature almost exclusively on-shell $W_{\text{had}}$ bosons and a sharp peak at 500 GeV in the $p_T$ distribution, while the ATLAS samples feature more off-shell than on-shell $W_{\text{had}}$ bosons and a $p_T$ distribution that is shifted to lower values. The ATLAS samples also feature a noticeable peak at $p_T \approx 500$ GeV. This is the effect of the trigger condition for the leading TAR jet to satisfy $p_T > 500$ GeV. Most of the $W_{\text{had}}$ bosons in the ATLAS samples have $p_T < 500$ GeV, since the preselection almost always assigns the $W_{\text{had}}$ jet to be the subleading jet. Up until this point, such a condition is not present in the SHERPA samples. In addition, the mass imbalance of the $Z$ and Higgs boson in the mock signal used in the SHERPA samples enables the $H \to WW^*$ side of the process to have a higher $p_T$ than the $Z \to b\bar{b}$ jet. The $W_{\text{had}}$ jet is therefore almost always the leading jet in the SHERPA samples. To compensate for this, the conditions of $p_T^Z > 500$ GeV and

$p_T^Z > 0.9 p_T^H$ are imposed on the $Z$ boson on matrix element level.
The difference of on-shell and off-shell $W_{\text{had}}$ bosons between the two samples can be further explained by an extra step of the OLR. Efforts are made to remove leptons that were emitted during the formation of the jet but are falsely identified as a signal lepton, since they are already accounted for in the energy of the jet. In the topology of the analysis, real signal leptons can also be in proximity to the jet, making the distinction between real signal leptons and leptons emitted by the jet important. The distinction is based on the respective energy spectra of the leptons. Leptons emitted by a jet generally have lower energy, while signal leptons have higher energy and are thus allowed to be in closer proximity to the jet. The jets used in this step are particle flow jets [88] of radius $R = 0.4$. Leptons are removed if the distance between a lepton and a jet fulfils

$$\Delta R(\ell, R = 0.4 \text{ jet}) < \min\left(0.4,\ 0.04 + \frac{10\,\text{GeV}}{p_T^\ell}\right). \tag{6.1}$$

This condition selects high-$p_T$ leptons preferentially. If the $W_{\text{lep}}$ boson has higher $p_T$, the $W_{\text{had}}$ boson tends to have lower $p_T$. The correlation between objects of higher mass receiving a higher boost results in this OLR step favouring heavy (on-shell) $W_{\text{lep}}$ bosons and thus light (off-shell) $W_{\text{had}}$ bosons. This OLR step is implemented in the Rivet analysis for the Sherpa samples by considering standard anti-$k_T$ jets with radius $R = 0.4$. After the implementation of both mentioned corrections in the Rivet analysis, the comparison to ATLAS samples in Figure 6.4 shows much better agreement.

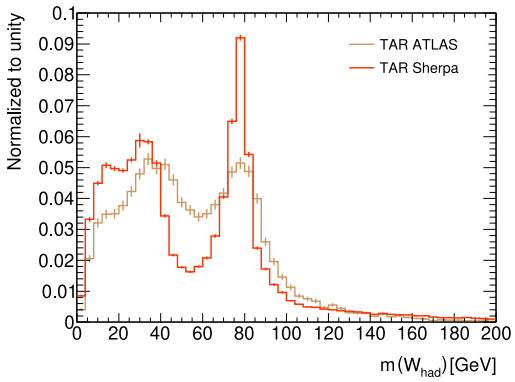


**Figure 6.4(a):** $m(W_{\text{had}})$

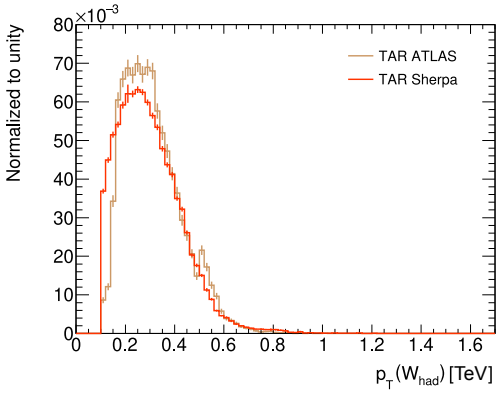


**Figure 6.4(b):** $p_T(W_{\text{had}})$

**Figure 6.4.:** Comparison of the mass and $p_T$ of the $W_{\text{had}}$ jet for TAR jet objects between the ATLAS and Sherpa samples. The Sherpa samples include the full OLR as well as the conditions of $p_T^Z > 500\,\text{GeV}$ and $p_T^Z > 0.9 p_T^H$ on matrix element level.

### 6.2.2. Reweighting of the $H \to WW$ Boson $p_T$

In order to ensure kinematic agreement of the mock and ATLAS signals, the particle level truth information about the $H \to WW$ $p_T$ is compared for the ATLAS and SHERPA samples in Figure 6.5.

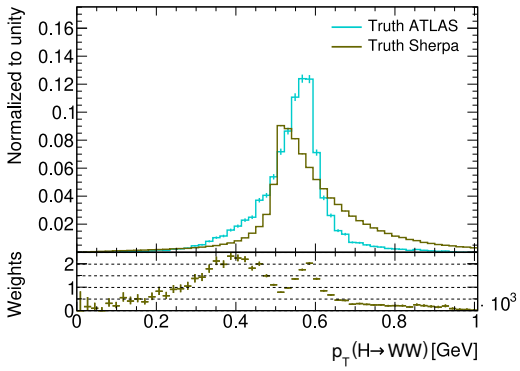


**Figure 6.5.:** Comparison of ATLAS and SHERPA samples for the $H \to WW$ $p_T$ on particle level and their ratio, which is applied as additional weights for the SHERPA samples.

The ratio of the distributions is applied as an additional weight $w_i$ per $p_T(H \to WW)$-bin with uncertainty $\sigma_{w_i}$ on the SHERPA samples to reduce the visible shape differences and to improve the agreement on particle level. The uncertainty $\sigma_{w_i}$ of weight $w_i$ is propagated by drawing the weight randomly from a normal distribution $\mathcal{N}\left(w_i, \sigma_{w_i}^2\right)$ with mean $w_i$ and width $\sigma_{w_i}$.
The $p_T$ of the Higgs boson and mass of the $W_{\text{had}}$ boson after reweighting can be seen in Figure 6.6.
The agreement of the distributions is significantly improved. However, the peaks of the $p_T$ distributions of the Higgs bosons still vary from each other. A total agreement is not expected, since the signals are fundamentally different. Comparisons for additional kinematic observables can be found in Appendix A.5.1.

### 6.2.3. Comparison of Sherpa Samples against ATLAS Samples

After the reweighting, ATLAS and SHERPA samples are compared for TAR and UFO jet objects in the muon channel. Because of challenges in the OLR of electrons for UFO jets that will be addressed in Section 6.2.4, the comparison is restricted solely to the muon channel. In addition, the fraction of on-shell $W_{\text{had}}$ bosons is calculated as the fraction of $W_{\text{had}}$ jets with masses $m_{W_{\text{had}}} > 60\,\text{GeV}$. The comparison of the $W_{\text{had}}$ mass in Figure 6.7 shows a much better mass resolution for the SHERPA samples.

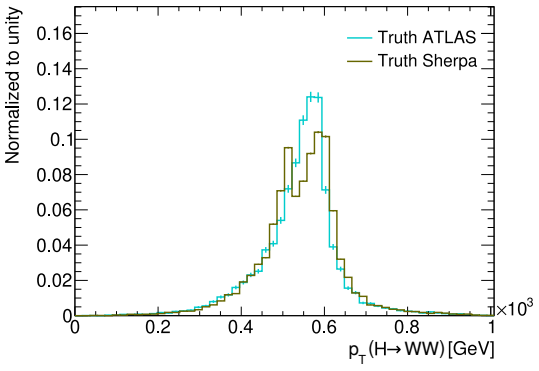


**Figure 6.6(a):** $p_T(H \to WW)$

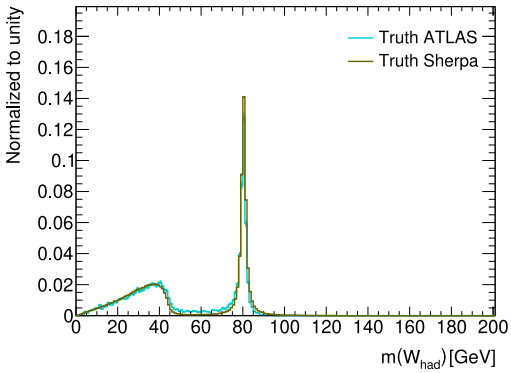


**Figure 6.6(b):** $m(W_{\text{had}})$

**Figure 6.6.:** Comparison of ATLAS and SHERPA samples for the $H \to WW$ boson $p_T$ and $W_{\text{had}}$ boson mass on particle level, after reweighting of the SHERPA samples.

In contrast to the SHERPA samples, the ATLAS samples include a detector simulation. Exposing the signal to smearing caused by a finite resolution and other detector effects, this difference is expected. Aside from the differences in the resolution, the different samples exhibit a similar mass distribution. The fraction of on-shell $W_{\text{had}}$ bosons $f_{\text{on-shell}} = 0.44$ is the same for TAR jets for SHERPA and ATLAS samples, demonstrating a similar performance of both implementations of the TAR jet algorithm. The TAR jet algorithm slightly favours off-shell $W_{\text{had}}$ bosons. The fraction of on-shell $W_{\text{had}}$ bosons for UFO jets differ for the two implementations. While the SHERPA samples show an equal split of on-shell and off-shell bosons with $f_{\text{on-shell}} = 0.5$, the ATLAS samples slightly favour on-shell $W_{\text{had}}$ bosons with $f_{\text{on-shell}} = 0.57$. The differences in the two implementations could originate from steps in the algorithm that cannot be implemented on truth level objects, as described in Section 4.2.3.

In Figure 6.8, it is further visible that the two implementations agree reasonably well with each other with respect to jet substructure observables, as shown for the example of the $W_{\text{had}}$ bosons subjettiness $\tau_{21}$. Additional observables can be found in Appendix A.5.2.

To demonstrate and quantify the detector effects seen in the mass distribution $h(m)$, the on-shell peak of the SHERPA samples with masses $60\,\text{GeV} < m_{W_{\text{had}}} < 100\,\text{GeV}$ is smeared by convolution with a normal distribution $g \coloneqq \mathcal{N}(0, \sigma^2)$

$$(h * g)(m) \coloneqq \int h(m)\, g(m - \tau)\, \mathrm{d}\tau\,. \tag{6.2}$$

The width of the normal distribution $\sigma$ is varied to minimize the chi-square per degree of freedom $\chi^2/n$. The on-shell peak with and without smearing can be seen in Figure 6.9

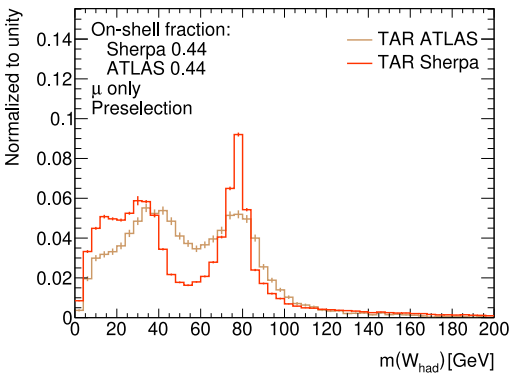


**Figure 6.7(a):** TAR jet

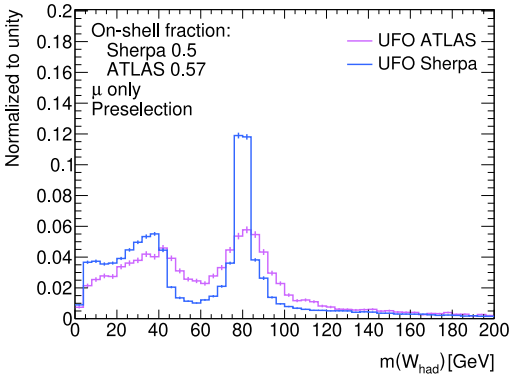


**Figure 6.7(b):** UFO jet

**Figure 6.7.:** Comparison of ATLAS and SHERPA samples for the $m(W_{\text{had}})$ distribution for TAR and UFO jet objects in the muon channel. The fraction of events with $m_{W_{\text{had}}} > 60\,\text{GeV}$ is calculated.

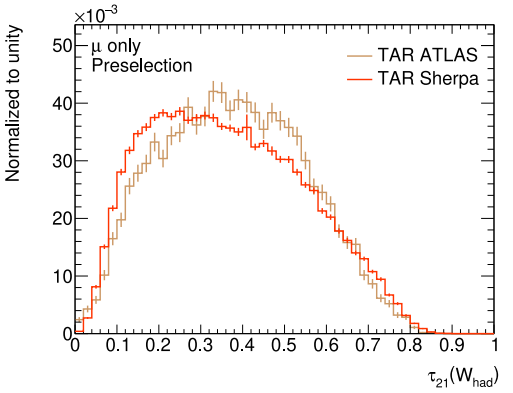


**Figure 6.8(a):** TAR jet

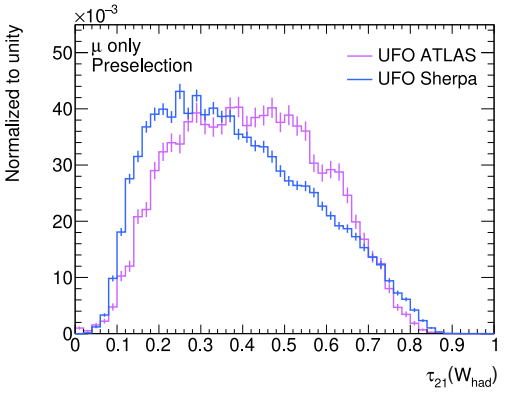


**Figure 6.8(b):** UFO jet

**Figure 6.8.:** Comparison of ATLAS and SHERPA samples for the $\tau_{21}(W_{\text{had}})$ distribution for TAR and UFO jet objects in the muon channel.

for TAR jets and Figure 6.10 for UFO jets.

The smearing creates a strong agreement between the SHERPA and ATLAS samples. With a width of $\sigma = 10.36\,\text{GeV}$ of the smearing distribution, the reduced chi-square between the TAR jet distributions reaches $\chi^2/n = 3.68$. For the UFO jet distributions the reduced chi-square reaches $\chi^2/n = 3.48$ for a width of $\sigma = 11.42\,\text{GeV}$. For the studied topology, the ATLAS detector causes a smearing effect on the mass distribution of the $W_{\text{had}}$ jet of 10.36 GeV for TAR jets and 11.42 GeV for UFO jets.

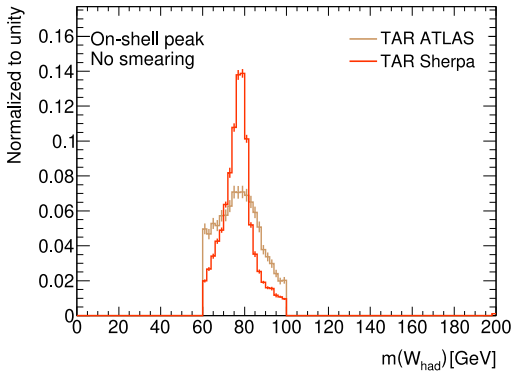


**Figure 6.9(a):** Without smearing

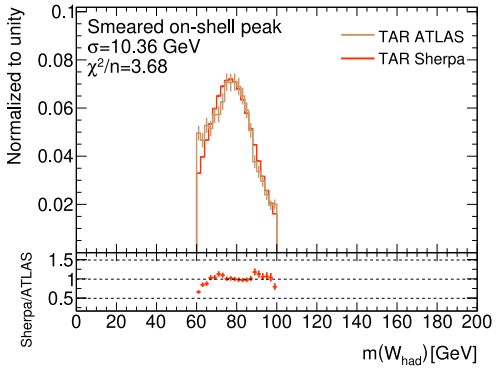


**Figure 6.9(b):** With smearing

**Figure 6.9.:** Comparison of ATLAS and SHERPA samples for TAR jets at the on-shell peak of the $W_{\text{had}}$ jet mass distribution, with and without smearing the SHERPA signal.

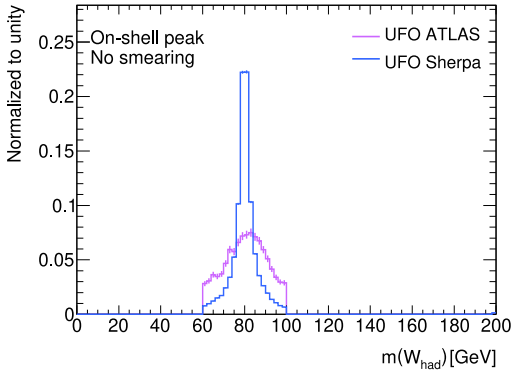


**Figure 6.10(a):** Without smearing

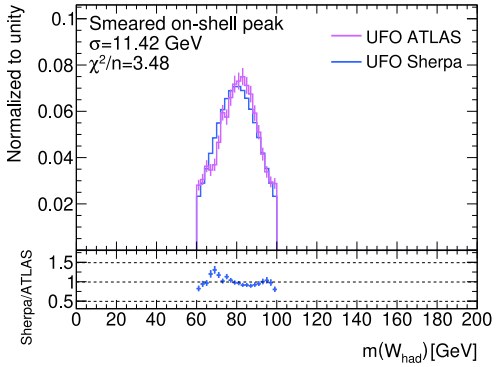


**Figure 6.10(b):** With smearing

**Figure 6.10.:** Comparison of ATLAS and SHERPA samples for UFO jets at the on-shell peak of the $W_{\text{had}}$ jet mass distribution, with and without smearing the SHERPA signal.

### 6.2.4. Comparison of TAR Jets against UFO Jets for ATLAS Samples

The performance differences between TAR jets and UFO jets are studied using only ATLAS samples. Next to kinematic observables, observables that are of special interest are those used to define cuts for signal and control regions. In Figure 6.11, the final discriminant $m_{\text{vis+met}}(HH)$ and the minimal distance between the signal lepton and the jets $\min \Delta R(\ell, \text{jet})$ are compared for TAR jets and UFO jets at different mass points $m_X = 1.0,\ 2.0$ and $3.0\,\text{TeV}$.

For lower mass points, the distributions for $\min \Delta R(\ell, \text{jet})$ are nearly identical. At higher

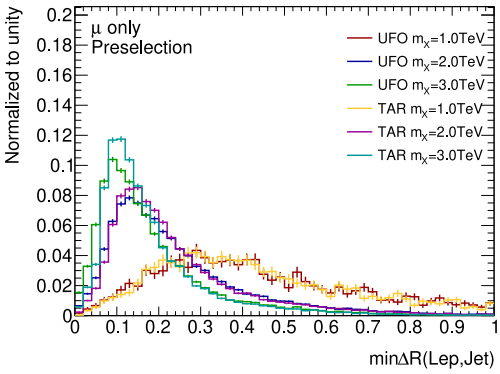


**Figure 6.11(a):** $\min \Delta R(\ell, \text{jet})$

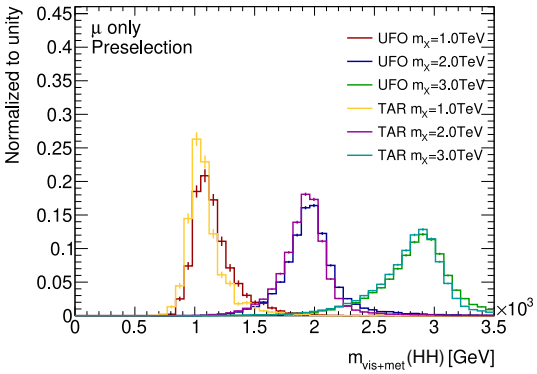


**Figure 6.11(b):** $HH$ $m_{\text{vis+met}}$

**Figure 6.11.:** Comparison between TAR and UFO jet objects for $HH$ samples at different mass points $m_X$ in the muon channel for the minimal distance between the signal lepton and jets $\min \Delta R(\ell, \text{jet})$ and the final discriminant $m_{\text{vis+met}}(HH)$.

masses, the TAR jets exhibit a slightly larger separation of the signal lepton and the nearest jet. A possible explanation for this behaviour is the fact that UFO jets favour on-shell $W_{\text{had}}$ bosons, while TAR jets favour off-shell ones. Therefore, TAR jets typically feature a higher boosted, and therefore more collimated, $W_{\text{lep}}$ boson, resulting in the lepton not coming as close to the axis of the $W_{\text{had}}$ jet, as for UFO jets.

For $m_{\text{vis+met}}(HH)$, a shift to lower values is noticeable for TAR jets, with respect to UFO jets. This is a general effect of the different mass definitions for UFO and TAR jets. The effect is even more visible when looking at the mass of the $W_{\text{had}}$ boson in Figure 6.12. The constituents of the two jet objects, which make up for the total mass, differ. For the TAR jets, they are the ghost matched tracks with rescaled momenta, while for the UFO jets they are a combination of PFOs and tracks. The masses of the two jet objects are therefore not directly comparable.

As described in Section 6.2.1, to avoid double-counting of the electron energy, small-$R$ jets with radius $R = 0.2$ are removed when overlapping with the electron. These small-$R$ jets are only used in the TAR jet algorithm but not in the UFO jet algorithm. Therefore, there does not exist any technique to avoid double-counting of the electron energy for UFO jets. Just like for the comparison of Sherpa samples in Section 5.4, this becomes visible when comparing TAR jets and UFO jets in the electron channel. Figure 6.13 shows the extra energy of the electron in UFO jets shifting $m_{\text{vis+met}}$ mass to higher values in comparison to TAR jets. The current analysis strategy would not work for UFO jets in the electron channel. To facilitate the use of the electron channel for UFO jets, it would be necessary to implement an electron OLR technique within the UFO jets, or to modify

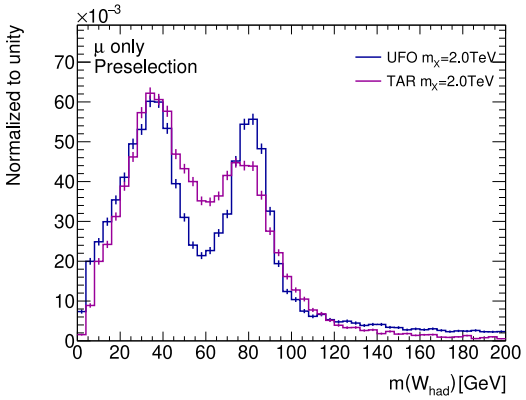


**Figure 6.12.:** Comparison between TAR and UFO jet objects for $HH$ samples with $m_X = 2\,\text{TeV}$ in the muon channel for the mass of the $W_{\text{had}}$ jet.

the analysis strategy to account for the additional energy contribution from the electron in the jets.

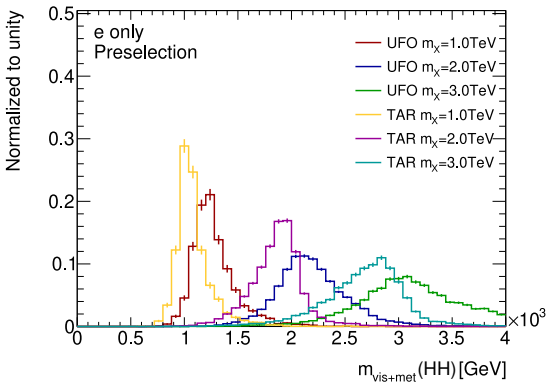


**Figure 6.13.:** Comparison between TAR and UFO jet objects for $HH$ samples at different mass points $m_X$ in the electron channel for $m_{\text{vis+met}}(HH)$.

To evaluate and compare the discrimination performance of substructure observables for the two jet objects, it is necessary to inspect the behaviour of signal jets in front of background events. In Figure 6.14 the ECF ratio $C_2$ is displayed along with the main contributing backgrounds, $W$+jets and $t\bar{t}$ events, for TAR jets and UFO jets.
In the analysis, the cuts on the $C_2$ observable are made in the signal region SRp2 according to a selection window evaluated on $SH$ samples. For this reason, $SH$ signals in the signal region are displayed at different mass points $(m_X, m_S) = (2\,\text{TeV}, 400\,\text{GeV})$, $(3\,\text{Tev}, 400\,\text{GeV})$ and $(3\,\text{TeV}, 1\,\text{TeV})$. The comparison shows similar behaviour for TAR jets and UFO jets, suggesting only small changes to the $C_2$ selection window, if any at all, are necessary when using UFO jets.
In Appendix A.5.3, more observables for $HH$ and $SH$ samples, with and without back-

grounds and in both the electron and muon channel are portrayed.

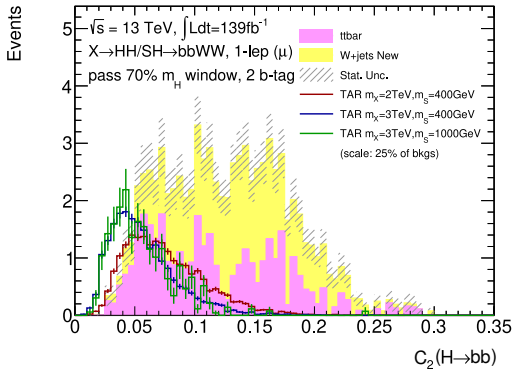


**Figure 6.14(a):** TAR jet

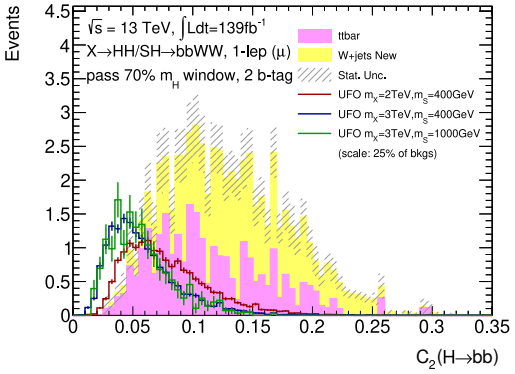


**Figure 6.14(b):** UFO jet

**Figure 6.14.:** $SH$ samples over the most contributing backgrounds at different mass points $m_X$ in the SRp2 signal region for the ECF ratio $C_2$ of the $H \to b\bar{b}$ jet. The signal samples are scaled to match 25% of the background integral.

## 6.3. Validation of new $V$+jets Background Samples

The $V$+jets samples used in the analysis are produced by Sherpa 2.2.1 and feature a mismodelling in the phase space of the analysis. Sherpa version 2.2.1 was released in 2016. Since then, several improvements and new features have been added by the Sherpa collaboration in newer releases [96, 97]. Newer $V$+jets samples, produced with Sherpa 2.2.11 exist in ATLAS, but have not been used before in the analysis. This section highlights the mismodelling present in the old $V$+jets samples, compares them to the new samples, and presents the improved modelling achieved with the use of the new samples.

The normalization factors obtained with the old samples in a single bin, based upon the ratios of backgrounds to data, are listed in Table 6.2. The data is fitted simultaneously in the three control regions, varying scale factors for the $t\bar{t}$, $W$+jets and non-prompt lepton background. The normalization factor for $W$+jets background $\mu_{W+\text{jets}}$ is a drastic correction of the background with $\mu_{W+\text{jets}} = 0.53$.

In addition, shape differences between combined backgrounds and data are visible. Figure 6.15 shows the event yields and the MC prediction as a function of the final discriminant $m_{\text{vis+met}}(HH)$ in the $W$+jets CR. The largest deviation is in the peak, where the MC prediction underestimates data. In the tail of the distribution, the deviations are smaller and the MC prediction overestimates data.

| Background | Normalization factor |
|---|---|
| $W$+jets | 0.53 |
| $t\bar{t}$ | 0.71 |
| non-prompt lepton | 1.18 |

**Table 6.2.:** Normalization factors for the $t\bar{t}$, $W$+jets and non-prompt lepton backgrounds with $V$+jets samples produced with SHERPA 2.2.1 [9].

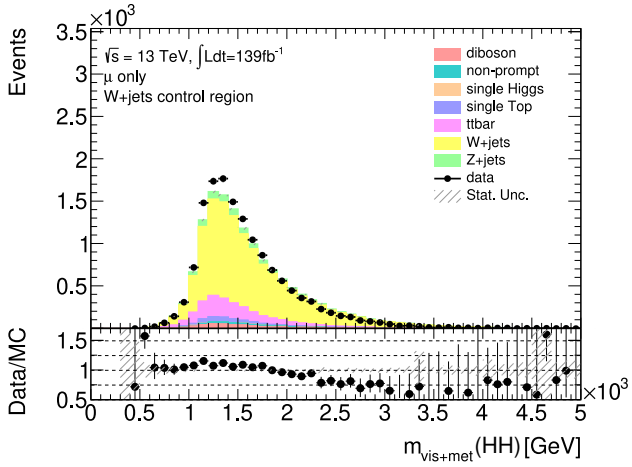


**Figure 6.15.:** ATLAS data against background including $V$+jets samples produced with SHERPA 2.2.1 for $m_{\text{vis+met}}(HH)$ in the $W$+jets CR.

To compare both sample versions, the $W$+jets samples are assessed in the $W$+jets CR, while the QCD CR is used for evaluating the $Z$+jets samples. These regions are chosen as the respective background processes of interest contribute maximally there, as can be seen in Figure 6.16.

Figure 6.17 shows the comparison of both versions of $V$+jets samples. $W$+jets and $Z$+jets samples are compared separately from each other in the respective control regions. The new samples feature approximately half of the weighted number of events as the old samples, across the complete depicted mass range.

This already indicates the potential to improve the normalization factor by bringing it closer to 1 for the $W$+jets background, and the improvement of the described differences for the combined MC backgrounds and data distributions. More importantly, the new samples do not exhibit any unexpected behaviour and major shape differences. This is highlighted by the normalized $p_T(W_{\text{had}})$ distributions in Figure 6.18. Placing both sample distributions on the same graphic reveals only minor discrepancies.

Additional comparisons for various observables and normalized distributions can be found in Appendix A.6.1.

New normalization factors are computed by fitting MC backgrounds including the SHERPA

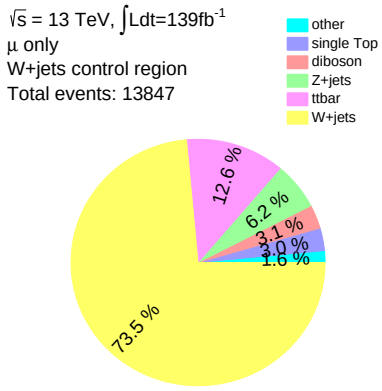


**Figure 6.16(a):** $W$+jets CR

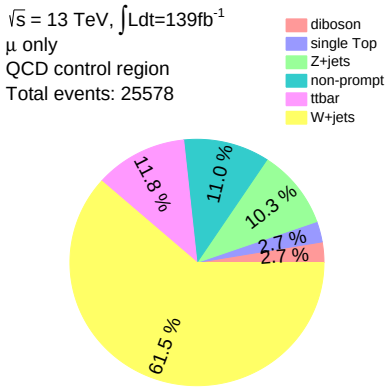


**Figure 6.16(b):** QCD CR

**Figure 6.16.:** Background composition in the $W$+jets and QCD CR for Sherpa 2.2.1 $V$+jets samples. Backgrounds contributing less than 2.2% are collected in the "other" category.

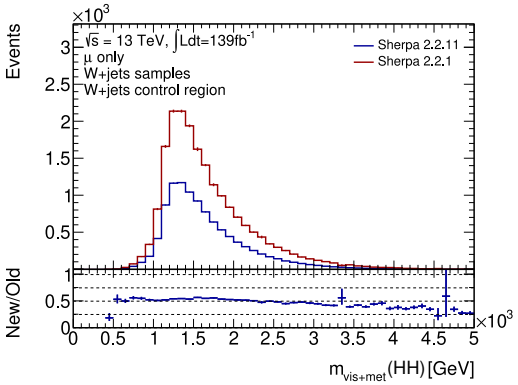


**Figure 6.17(a):** $W$+jets

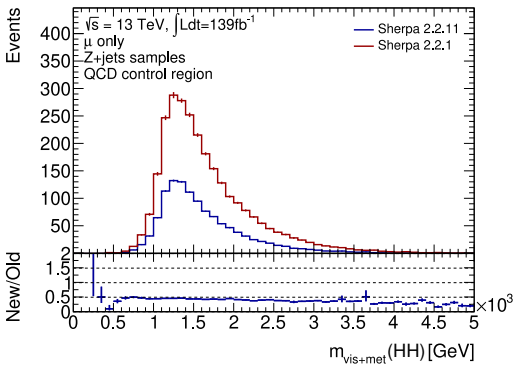


**Figure 6.17(b):** $Z$+jets

**Figure 6.17.:** Comparison between $V$+jets samples produced with Sherpa version 2.2.1 and 2.2.11 for $m_{\text{vis+met}}(HH)$. $W$+jets samples are compared in the $W$+jets CR, while $Z$+jets samples are compared in the QCD CR.

2.2.11 $V$+jets to data and are listed in Table 6.3. The normalization factor for $W$+jets samples improves from $\mu_{W+\text{jets,old}} = 0.53$ to $\mu_{W+\text{jets,new}} = 1.074$, evidence of a much better modelling of the background in the Sherpa 2.2.11 samples.

Figure 6.19 shows data and MC background predictions for the final discriminant $m_{\text{vis+met}}(HH)$, in the $W$+jets CR including the Sherpa 2.2.11 $V$+jets with the new normalization factors. The distributions feature a slightly better agreement with the new $V$+jets samples. Most noticeable are the smaller differences in the peak of the distribution, where the MC prediction's underestimation of data was reduced, but does still exist. Similar improvements are seen for the MC prediction's overestimation of data in the tail of the distribution.

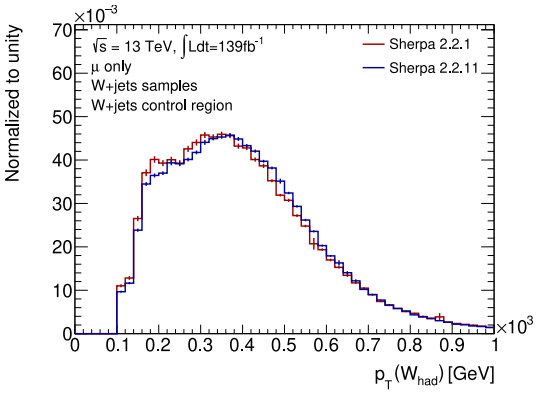


**Figure 6.18(a):** $W$+jets

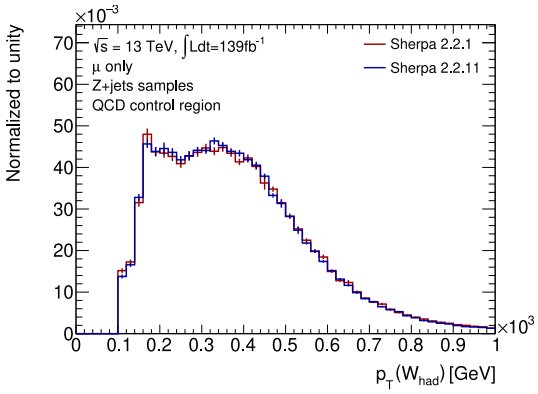


**Figure 6.18(b):** $Z$+jets

**Figure 6.18.:** Comparison between normalized $V$+jets samples produced with Sherpa version 2.2.1 and 2.2.11 for $p_T(W_{\text{had}})$. $W$+jets samples are compared in the $W$+jets CR, while $Z$+jets samples are compared in the QCD CR.

| Background | Normalization factor |
|---|---|
| $W$+jets | 1.074 |
| $t\bar{t}$ | 0.775 |
| non-prompt lepton | 0.86 |

**Table 6.3.:** Normalization factors for the $t\bar{t}$, $W$+jets and non-prompt lepton backgrounds with $V$+jets samples produced with Sherpa 2.2.11 [40].

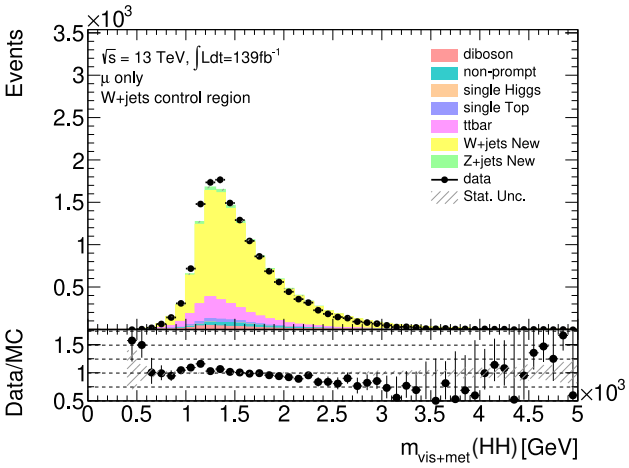


**Figure 6.19.:** ATLAS data against background for $V$+jets samples produced with Sherpa 2.2.11 for $m_{\text{vis+met}}(HH)$ in the $W$+jets CR.

Due to the substantial enhancements in both the shape of distributions and normalization achieved with the new samples, it is recommended that samples produced with Sherpa 2.2.11 are utilized in the ATLAS analysis.

Additional comparisons for data against backgrounds for various observables and the two

$V$+jets sample versions are collected in Appendix A.6.2.

## 6.4. Normalization of $Z$+jets Background

Comparing the event yields and MC prediction for SHERPA 2.2.1 $V$+jets samples in Figure 6.15 with those for SHERPA 2.2.11 $V$+jets samples in Figure 6.19, a strong reduction in $Z$+jets events can be noted, which are mainly present in the peak of the distribution. As the discrepancy between backgrounds and data is most pronounced in this peak, the suitability of scaling up $Z$+jets samples is explored. The decision to do so can only be supported by background-to-data fits, which also adjust the number of $Z$+jets events. For this, a control region has to be constructed, which enhances the contribution of $Z$+jets events. The basis of this new control region is the combination of the $W$+jets CR and QCD CR, where the only requirement is 0 $b$-tags in the event. The characteristic signature of the leptonic decay of a $Z$ boson $Z \to \ell^+\ell^-$ is the occurrence of two leptons with opposite signs. However, the MC background samples of the analysis feature only a single lepton in the event that passes the tight selection working point, as required by the preselection. The second lepton therefore has to be chosen from those that only pass the less strict loose selection working point. The second lepton is accepted if it has the opposite charge of the tight lepton.
The background composition of the $Z$+jets CR is shown in Figure 6.20.

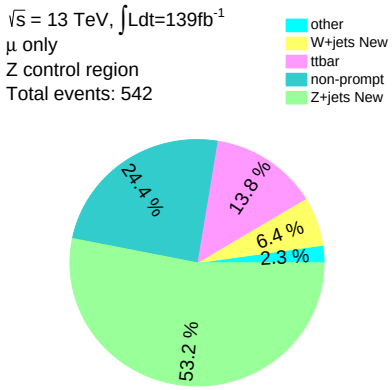


**Figure 6.20.:** Background composition in the $Z$+jets CR. Backgrounds contributing less than 2.2% are collected in the "other" category.

The chosen cuts are very effective in enhancing the contribution of $Z$+jets events, with a fraction of 53.2% of all SM processes. However, the amount of $Z$+jets events passing the preselection are very limited, resulting in a control region of very low statistics. The $Z$+jets CR consists of less than 550 events, while the $W$+jets CR and QCD CR consist of

over 13000 and 25000 events respectively, at an integrated luminosity of $\mathcal{L}_{\text{tot}} = 139\,\text{fb}^{-1}$. The comparison of data against background in the $Z$+jets CR can be seen for the reconstructed $Z$ boson mass and the (falsely) reconstructed mass of the $W_{\text{lep}}$ boson in Figure 6.21. It becomes evident that the fit of MC backgrounds to data would not benefit from scaling of the $Z$+jet background.

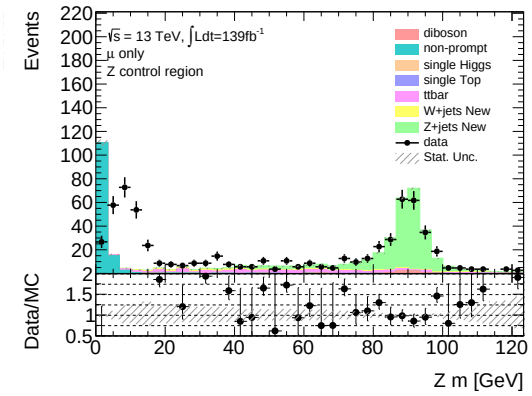


**Figure 6.21(a):** $m(Z)$

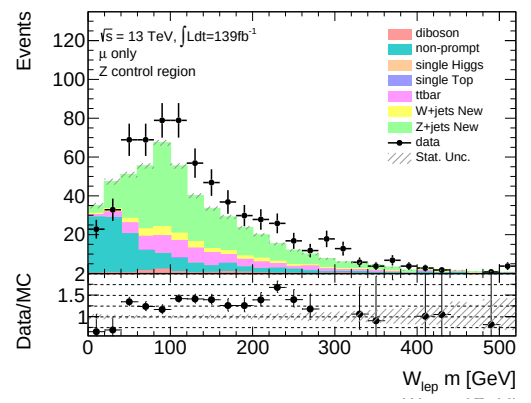


**Figure 6.21(b):** $m(W_{\text{lep}})$

**Figure 6.21.:** ATLAS data against background for $m(Z)$ and $m(W_{\text{lep}})$ in the $Z$+jets CR.

The apparent mismodelling of non-prompt lepton backgrounds to the first peak in Figure 6.21(a) is explained by the derivation of these samples. The non-prompt leptons are estimated in a data-driven approach which uses the efficiencies of loose (non-)prompt leptons passing the tight lepton selection. Altering these efficiencies by requiring a stricter selection than the loose lepton selection, such as requiring a specific charge, biases the non-prompt lepton background estimate. Therefore, it is expected that the estimate does not function correctly in the $Z$+jets control region.
Although the constructed $Z$+jets control region functions as intended, the $Z$+jets samples demonstrate good modelling. For this reason, there is no need to scale the background predictions.
Additional comparisons for data against backgrounds for various observables in the $Z$+jets CR can be found in Appendix A.7.

# 7. Conclusion

*The universe is full of magical things patiently waiting for our wits to grow sharper.*

Eden Phillpotts

The final state of heavy particles decaying into much lighter decay products features a unique, collimated signature. For searches involving heavy particles, it is essential that the reconstructed jets have precise spatial resolution. Traditional jets reconstructed using calorimeters face challenges due to their large cell sizes, necessitating innovative methods for jet reconstruction in experiments.

This thesis investigated and compared the performance of sophisticated jet objects that utilize tracking information and deploy different jet grooming techniques in the decay of semi-leptonic boosted $H \to WW^*$ decays. The signature of the decays features a high-energy jet ($W_{\text{had}}$) and a close by, or even overlapping, lepton. Specifically, the focus was placed on TAR jets and UFO jets, which were implemented in a Rivet analysis, allowing for the flexible adjustment of a broad range of parameters. A third jet object, the standard LRJ, was used as a control object.

The successful implementation of the jet objects was demonstrated for QCD jets, showing mostly small differences between the jet objects. The common use of Trimming in TAR jets and standard LRJs became apparent because the differences between these objects were smaller than for the UFO jets, which use the Soft Drop grooming algorithm and CS+SK pile-up mitigation. The energy correlation function ratio $C_2^1$ proved itself to be especially sensitive to the differences in the jet reconstruction algorithms.

When comparing the $W_{\text{had}}$ jets from the Higgs-Strahlung process $Z^* \to ZH \to b\bar{b}WW^*$ to QCD jets, the subjettiness $\tau_{21}$, as well as the energy correlation function ratio $D_2^1$, were able to discriminate the signal process from the background. While the separation power was very similar for the different jet objects, for most observables the UFO jets possessed a small advantage over the other jet object definitions.

It was shown that Trimming and Soft Drop with CS+SK remove the lepton originating from the leptonically decaying $W$ boson equally rare (in 3.5% of all events) from the

jet of the hadronically decaying $W$ boson. It is a huge advantage of TAR jets to be able to remove such a lepton retroactively. The clustering of the lepton showed that the subjettiness $\tau_{32}$ and $\tau_{42}$ become important to distinguish the Higgs-Strahlung process from QCD background with the additional prong of the lepton, while the subjettiness $\tau_{21}$ loses separation power. However, kinematic observables such as the mass become heavily influenced by the additional lepton energy in the jet and become unreliable for the identification of the object of interest. While the additional energy removed with the electron in TAR jets shows that the overlap removal method for electrons is imperfect, it is a large improvement over ignoring the clustered leptons.

Under systematic variations, TAR and UFO jet objects proved to be stable, and the used grooming techniques are observed to be able to remove soft radiation from MPI.

Similarly, the ATLAS implementation of TAR jets and UFO jets was studied in the $X \to HH \to b\bar{b}WW^*$ analysis, aiming to closely replicate and study their application in experiments. To verify whether the findings for the Sherpa Higgs-Strahlung sample would be transferable to the ATLAS signal, the two samples were compared. The Higgs-Strahlung samples were modified for a better kinematic agreement, resulting in a stronger bias to select off-shell $W_{\text{had}}$ bosons and matching distributions. The comparison also demonstrated the impact of detector effects on the optimal resolution achievable on truth level.

Comparing TAR jets and UFO jets for the ATLAS samples demonstrated only minor differences, even for the energy correlation function ratio $C_2^1$, which was very sensitive to the different jet objects on truth level. Larger deviations were seen for mass observables, caused by the differences in inputs of the jet objects. However, the distributions of the final discriminant, the visible+met mass of the $HH$ system, were only slightly influenced by this.

Considering all evidence available until this point, the thesis showed that the jet objects are similar enough to maintain the current analysis strategy for UFO jets in the muon channel. In this case, the selection cuts have to be checked and possibly reevaluated to yield the same efficiencies as for TAR jets. However, there are no major differences tgat showed any need for a completely new strategy. For the inclusion of the electron channel, a technique has to be developed to mitigate the clustering of signal electrons into the jets, as this is currently not possible for UFO jets.

The studies in the ATLAS $X \to HH \to b\bar{b}WW^*$ analysis were extended to the validation of a new version $V$+jets samples, which were shown to improve the scale factors of the background samples drastically, while also providing small improvements to the shape of distributions. As a result, the new samples are now used in the ATLAS analysis.

The construction of a $Z$+jets control region proved to be very effective in enhancing the background contribution of $Z$+jets events. The low statistics in this region, and the agreement between MC events and data suggest that a scaling of the $Z$+jets event normalization would not improve the modelling.

Jets are an essential probe for the investigation of new physics, as novel particles are anticipated to form jets during their decay. The application of jet algorithms exploiting tracking information plays a crucial role in improving the resolution and purity of the reconstructed objects. Thus, these algorithms have a direct impact on the accuracy of measurements, contributing to the process of setting exclusion limits on theoretical models and eventually leading to the discovery of new physics. Ultimately, the understanding and refinement of track-assisted jet algorithms is essential for uncovering new physics phenomena and advancing our understanding of the universe at the smallest scales.

# Bibliography


[1] M. E. Peskin and D. V. Schroeder: *An Introduction To Quantum Field Theory*. (1995). DOI: 10.1201/9780429503559.

[2] D. Griffiths: *Introduction to Elementary Particles*. (1987). DOI: 10.1002/9783527618460.

[3] Particle Data Group: *Review of Particle Physics*. PTEP 083C01 (8 2022). DOI: 10.1093/ptep/ptac097.

[4] G. Bertone et al.: *Particle dark matter: evidence, candidates and constraints*. Phys. Rep. 405.5 (2005). DOI: 10.1016/j.physrep.2004.08.031.

[5] A. D. Sakharov: *Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe*. Sov. Phys. Usp. 34 (1991). DOI: 10.1070/PU1991v034n05ABEH002497.

[6] LSND Collaboration: *Evidence for neutrino oscillations from the observation of $\overline{\nu}_e$ appearance in a $\overline{\nu}_\mu$ beam*. Phys. Rev. D 64 (11 2001). DOI: 10.1103/PhysRevD.64.112007.

[7] ATLAS Collaboration: *Track assisted techniques for jet substructure*. ATL-PHYS-PUB- 2018-012. (2018). URL: https://cds.cern.ch/record/2630864.

[8] ATLAS Collaboration: *Optimisation of large-radius jet reconstruction for the ATLAS detector in 13 TeV proton-proton collisions*. Eur. Phys. J. C 81.4 (2021). DOI: 10.1140/epjc/s10052-021-09054-3.

[9] K. Abeling: *Search for resonant Higgs boson pair production in the $b\bar{b}WW^*$ decay channel in the boosted 1-lepton final state using the full Run 2 ATLAS dataset*. Reference: II.Physik-UniGö-Diss-2022/01. PhD thesis. (2022). DOI: 10.53846/goediss-9203.

[10] ATLAS Collaboration: *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*. Phys. Lett. B 716.1 (2012). DOI: 10.1016/j.physletb.2012.08.020.

[11] CMS Collaboration: *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*. Phys. Lett. B 716.1 (2012). DOI: 10.1016/j.physletb.2012.08.021.

[12] A. Einstein: *Die Grundlage der allgemeinen Relativitätstheorie*. Ann. Phys. 354.7 (1916). DOI: 10.1002/andp.19163540702.

[13] R. P. Feynman: *The Theory of Positrons*. Phys. Rev. 76 (6 1949). DOI: 10.1103/PhysRev.76.749.

[14] R. P. Feynman: *Space-Time Approach to Quantum Electrodynamics*. Phys. Rev. 76 (6 1949). DOI: 10.1103/PhysRev.76.769.

[15] P. W. Higgs: *Broken symmetries, massless particles and gauge fields*. Phys. Lett. 12.2 (1964). DOI: 10.1016/0031-9163(64)91136-9.

[16] F. Englert and R. Brout: *Broken Symmetry and the Mass of Gauge Vector Mesons*. Phys. Rev. Lett. 13 (9 1964). DOI: 10.1103/PhysRevLett.13.321.

[17] P. W. Higgs: *Broken Symmetries and the Masses of Gauge Bosons*. Phys. Rev. Lett. 13 (16 1964). DOI: 10.1103/PhysRevLett.13.508.

[18] G. S. Guralnik et al.: *Global Conservation Laws and Massless Particles*. Phys. Rev. Lett. 13 (20 1964). DOI: 10.1103/PhysRevLett.13.585.

[19] P. W. Higgs: *Spontaneous Symmetry Breakdown without Massless Bosons*. Phys. Rev. 145 (4 1966). DOI: 10.1103/PhysRev.145.1156.

[20] T. W. B. Kibble: *Symmetry Breaking in Non-Abelian Gauge Theories*. Phys. Rev. 155 (5 1967). DOI: 10.1103/PhysRev.155.1554.

[21] T. Robens et al.: *Two-real-scalar-singlet extension of the SM: LHC phenomenology and benchmark scenarios*. Eur. Phys. J. C 80.2 (2020). DOI: 10.1140/epjc/s10052-020-7655-x.

[22] G. C. Branco et al.: *Theory and phenomenology of two-Higgs-doublet models*. Phys. Rep. 516.1-2 (2012). DOI: 10.1016/j.physrep.2012.02.002.

[23] A. Buckley et al.: *General-purpose event generators for LHC physics*. Phys. Rep. 504 (2011). DOI: 10.1016/j.physrep.2011.03.005.

[24] G. P. Salam: *Elements of QCD for hadron colliders*. (2010). DOI: 10.48550/ARXIV.1011.5131.

[25] R. P. Feynman: *The Behavior of Hadron Collisions at Extreme Energies*. In: *Special Relativity and Quantum Theory: A Collection of Papers on the Poincaré Group*. (1988). DOI: 10.1007/978-94-009-3051-3_25.

[26] S. D. Drell and T.-M. Yan: *Massive Lepton-Pair Production in Hadron-Hadron Collisions at High Energies*. Phys. Rev. Lett. 25 (5 1970). DOI: 10.1103/PhysRevLett.25.316.

[27] S. Höche: *Introduction to parton-shower event generators*. (2014). DOI: 10.48550/ARXIV.1411.4085.

[28] B. Andersson et al.: *Parton fragmentation and string dynamics*. Phys. Rep. 97.2 (1983). DOI: 10.1016/0370-1573(83)90080-7.

[29] T. Sjöstrand: *Jet fragmentation of multiparton configurations in a string framework*. Nucl. Phys. B 248.2 (1984). DOI: 10.1016/0550-3213(84)90607-2.

[30] R. K. Ellis et al.: *QCD and collider physics*. Cambridge University Press, (1996). ISBN: 0 521 58189 3.

[31] S. Agostinelli et al.: *Geant4 — a simulation toolkit*. 506.3 (2003). DOI: 10.1016/S0168-9002(03)01368-8.

[32] ATLAS Collaboration: *Performance of the Fast ATLAS Tracking Simulation (FATRAS) and the ATLAS Fast Calorimeter Simulation (FastCaloSim) with single particles*. ATL-SOFT-PUB-2014-001. (2014). URL: https://cds.cern.ch/record/1669341.

[33] M. Cacciari et al.: *The anti-$k_t$ jet clustering algorithm*. JHEP 2008.04 (2008). DOI: 10.1088/1126-6708/2008/04/063.

[34] S. Catani et al.: *New clustering algorithm for multijet cross sections in e+e- annihilation*. Phys. Lett. B 269.3 (1991). DOI: 10.1016/0370-2693(91)90196-W.

[35] Yu. L. Dokshitzer et al.: *Better jet clustering algorithms*. JHEP 1997.08 (1997). DOI: 10.1088/1126-6708/1997/08/001.

[36] T. Plehn et al.: *Pair production of neutral Higgs particles in gluon-gluon collisions*. Nucl. Phys. B 479.1 (1996). DOI: 10.1016/0550-3213(96)00418-X.

[37] M. Grazzini et al.: *Higgs boson pair production at NNLO with top quark mass effects*. JHEP 2018.5 (2018). DOI: 10.1007/jhep05(2018)059.

[38] LHC Higgs Cross Section Working Group: *Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector*. (2017). DOI: 10.23731/CYRM-2017-002.

[39] ATLAS Collaboration: *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*. (2022). DOI: 10.48550/ARXIV.2212.09379.

[40] K. Abeling et al.: *Search for resonant boosted HH and SH production in the bbVV decay channel with 0 or 1 lepton in the final state using the full Run 2 ATLAS data*. ATL-COM-PHYS- 2022-020. (2022). URL: https://cds.cern.ch/record/2799424.

[41] Sherpa Collaboration: *Event generation with Sherpa 2.2*. SciPost Phys. 7.3 (2019). DOI: 10.21468/scipostphys.7.3.034.

[42] L. Evans and P. Bryant: *LHC Machine*. JINST 3 (2008). DOI: 10.1088/1748-0221/3/08/S08001.

[43] E. Mobs: *The CERN accelerator complex in 2019. Complexe des accélérateurs du CERN en 2019* (2019). URL: https://cds.cern.ch/record/2684277 (visited on Mar. 12, 2023).

[44] W. Herr and B. Muratori: *Concept of luminosity*. In: *CERN Accelerator School: Intermediate Course on Accelerator Physics, Zeuthen, Germany, 15-26 Sep 2003*. Ed. by D. Brandt. (2006). DOI: 10.5170/CERN-2006-002.

[45] ATLAS Collaboration: *The ATLAS Experiment at the CERN Large Hadron Collider*. JINST 3.08 (2008). DOI: 10.1088/1748-0221/3/08/S08003.

[46] CMS Collaboration: *The CMS experiment at the CERN LHC*. JINST 3.08 (2008). DOI: 10.1088/1748-0221/3/08/S08004.

[47] LHCb Collaboration: *The LHCb Detector at the LHC*. JINST 3.08 (2008). DOI: 10.1088/1748-0221/3/08/S08005.

[48] ALICE Collaboration: *The ALICE experiment at the CERN LHC*. JINST 3.08 (2008). DOI: 10.1088/1748-0221/3/08/S08002.

[49] *Run 3 Performance*. URL: https://lhc-commissioning.web.cern.ch/performance/Run-3-targets.htm (visited on Mar. 20, 2023).

[50] ATLAS Collaboration: *Luminosity Public Results Run 2*. URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2 (visited on Mar. 20, 2023).

[51] R. M. Bianchi: *ATLAS experiment schematic illustration* (2022). URL: https://cds.cern.ch/record/2837191 (visited on Mar. 12, 2023).

[52] B. Abbott et al.: *Production and integration of the ATLAS Insertable B-Layer*. JINST 13.05 (2018). DOI: 10.1088/1748-0221/13/05/T05008.

[53] S. Mehlhase: *ATLAS detector slice (and particle visualisations)* (2021). URL: https://cds.cern.ch/record/2770815 (visited on Mar. 12, 2023).

[54] ATLAS Collaboration: *Operation of the ATLAS trigger system in Run 2*. JINST 15.10 (2020). DOI: 10.1088/1748-0221/15/10/P10004.

[55] ATLAS Collaboration: *Performance of the ATLAS trigger system in 2015*. Eur. Phys. J. C 77.317 (2017). DOI: 10.1140/epjc/s10052-017-4852-3.

[56] C. Bierlich et al.: *Robust Independent Validation of Experiment and Theory: Rivet version 3*. SciPost Phys. 8.2 (2020). DOI: 10.21468/scipostphys.8.2.026.

[57] F. Krauss et al.: *AMEGIC++ 1.0, A Matrix Element Generator In C++*. JHEP 2002.02 (2002). DOI: 10.1088/1126-6708/2002/02/044.

[58] F. Buccioni et al.: *OpenLoops 2*. Eur. Phys. J. C 79.10 (2019). DOI: 10.1140/epjc/s10052-019-7306-2.

[59] S. Schumann and F. Krauss: *A parton shower algorithm based on Catani-Seymour dipole factorisation*. JHEP 2008.03 (2008). DOI: 10.1088/1126-6708/2008/03/038.

[60] S. Höche et al.: *QCD matrix elements + parton shower:. The NLO case*. JHEP 2013.4 (2013). DOI: 10.1007/jhep04(2013)027.

[61] J.-C. Winter et al.: *A modified cluster-hadronisation model*. Eur. Phys. J. C 36.3 (2004). DOI: 10.1140/epjc/s2004-01960-8.

[62] T. Gleisberg and S. Höche: *Comix, a new matrix element generator*. JHEP 2008.12 (2008). DOI: 10.1088/1126-6708/2008/12/039.

[63] S. Höche et al.: *QCD matrix elements and truncated showers*. JHEP 2009.05 (2009). DOI: 10.1088/1126-6708/2009/05/053.

[64] J. Alwall et al.: *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*. JHEP 2014.7 (2014). DOI: 10.1007/jhep07(2014)079.

[65] NNPDF Collaboration: *Parton distributions with LHC data*. Nucl. Phys. B 867.2 (2013). DOI: 10.1016/j.nuclphysb.2012.10.003.

[66] J. Bellm et al.: *Herwig 7.0/Herwig++ 3.0 release note*. Eur. Phys. J. C 76.196 (2016). DOI: 10.1140/epjc/s10052-016-4018-8.

[67] D. J. Lange: *The EvtGen particle decay simulation package*. Nucl. Instrum. Meth. A 462.1 (2001). DOI: 10.1016/S0168-9002(01)00089-4.

[68] T. Sjöstrand et al.: *An introduction to PYTHIA 8.2*. Comput. Phys. Commun. 191 (2015). DOI: 10.1016/j.cpc.2015.01.024.

[69] ATLAS Collaboration: *ATLAS Pythia 8 tunes to 7 TeV data*. ATL-PHYS-PUB- 2014-021. CERN, (2014). URL: https://cds.cern.ch/record/1966419.

[70] S. Frixione et al.: *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*. JHEP 2007.09 (2007). DOI: 10.1088/1126-6708/2007/09/126.

[71] P. Nason: *A new method for combining NLO QCD with shower Monte Carlo algorithms*. JHEP 2004.11 (2004). DOI: 10.1088/1126-6708/2004/11/040.

[72] S. Frixione et al.: *Matching NLO QCD computations with parton shower simulations: the POWHEG method*. JHEP 2007.11 (2007). DOI: 10.1088/1126-6708/2007/11/070.

[73] S. Alioli et al.: *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*. JHEP 2010.43 (2010). DOI: 10.1007/JHEP06(2010)043.

[74] NNPDF Collaboration: *Parton distributions for the LHC run II*. JHEP 2015.40 (2015). DOI: 10.1007/JHEP04(2015)040.

[75] S. Frixione et al.: *Single-top hadroproduction in association with a W boson*. JHEP 2008.07 (2008). DOI: 10.1088/1126-6708/2008/07/029.

[76] R. Frederix et al.: *Single-top t-channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO*. JHEP 2012.130 (2012). DOI: 10.1007/JHEP09(2012)130.

[77] ATLAS Collaboration: *Studies on top-quark Monte Carlo modelling for Top2016*. ATL-PHYS-PUB- 2016-020. (2016). URL: https://cds.cern.ch/record/2216168.

[78] A. J. Larkoski et al.: *Gaining (mutual) information about quark/gluon discrimination*. JHEP 2014.11 (2014). DOI: 10.1007/jhep11(2014)129.

[79] CMS Collaboration: *Search for a Higgs boson in the decay channel $H \to ZZ^{(*)} \to q\bar{q}\ell^-\ell^+$ in pp collisions at $\sqrt{s} = 7$ TeV*. JHEP 2012.4 (2012). DOI: 10.1007/jhep04(2012)036.

[80] J. Thaler and K. van Tilburg: *Identifying boosted objects with N-subjettiness*. JHEP 2011.3 (2011). DOI: 10.1007/jhep03(2011)015.

[81] A. J. Larkoski et al.: *Energy correlation functions for jet substructure*. JHEP 2013.6 (2013). DOI: 10.1007/jhep06(2013)108.

[82] A. J. Larkoski et al.: *Power counting to better jet observables*. JHEP 2014.12 (2014). DOI: 10.1007/jhep12(2014)009.

[83] C. Bierlich et al.: *Robust Independent Validation of Experiment and Theory: Rivet version 3*. SciPost Phys. 8.2 (2020). DOI: 10.21468/scipostphys.8.2.026.

[84] D. Krohn et al.: *Jet trimming*. JHEP 2010.2 (2010). DOI: 10.1007/jhep02(2010)084.

[85] M. Cacciari et al.: *FastJet user manual*. Eur. Phys. J. C 72.3 (2012). DOI: 10.1140/epjc/s10052-012-1896-2.

[86] M. Cacciari and G. P. Salam: *Pileup subtraction using jet areas*. Phys. Lett. B 659.1-2 (2008). DOI: 10.1016/j.physletb.2007.09.077.

[87] M. Cacciari et al.: *The catchment area of jets*. JHEP 2008.04 (2008). DOI: 10.1088/1126-6708/2008/04/005.

[88] ATLAS Collaboration: *Jet reconstruction and performance using particle flow with the ATLAS Detector*. Eur. Phys. J. C 77.7 (2017). DOI: 10.1140/epjc/s10052-017-5031-2.

[89] ATLAS Collaboration: *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*. Eur. Phys. J. C 77.7 (2017). DOI: 10.1140/epjc/s10052-017-5004-5.

[90] ATLAS Collaboration: *Improving jet substructure performance in ATLAS using Track-CaloClusters*. ATL-PHYS-PUB- 2017-015. (2017). URL: https://cds.cern.ch/record/2275636.

[91] A. M. Sirunyan et al.: *Particle-flow reconstruction and global event description with the CMS detector*. J. Instrum. 12.10 (2017). DOI: 10.1088/1748-0221/12/10/p10003.

[92] P. Berta et al.: *Particle-level pileup subtraction for jets and jet shapes*. JHEP 2014.6 (2014). DOI: 10.1007/jhep06(2014)092.

[93] S. D. Ellis and D. E. Soper: *Successive combination jet algorithm for hadron collisions*. Phys. Rev. D 48.7 (1993). DOI: 10.1103/physrevd.48.3160.

[94] M. Cacciari et al.: *SoftKiller, a particle-level pileup removal method*. Eur. Phys. J. C 75.2 (2015). DOI: 10.1140/epjc/s10052-015-3267-2.

[95] A. J. Larkoski et al.: *Soft drop*. JHEP 2014.5 (2014). DOI: 10.1007/jhep05(2014)146.

[96] Sherpa Collaboration: *SHERPA GitLab*. URL: https://gitlab.com/sherpa-team/sherpa (visited on Apr. 20, 2023).

[97] Sherpa Collaboration: *SHERPA Homepage*. URL: https://sherpa-team.gitlab.io (visited on Apr. 20, 2023).

# A. Additional Figures

## A.1. QCD Background

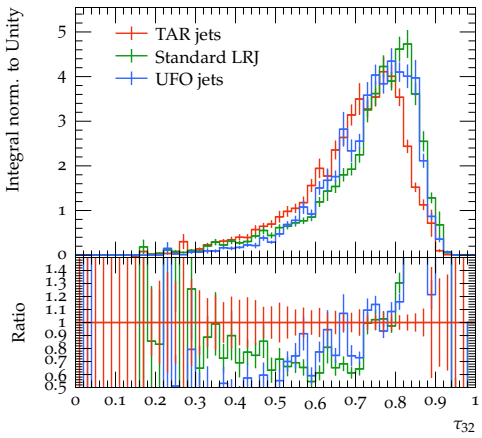


**Figure A.1(a):** Subjettiness $\tau_{32}$

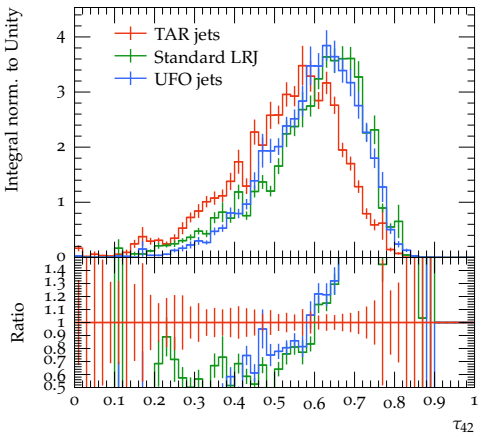


**Figure A.1(b):** Subjettiness $\tau_{42}$

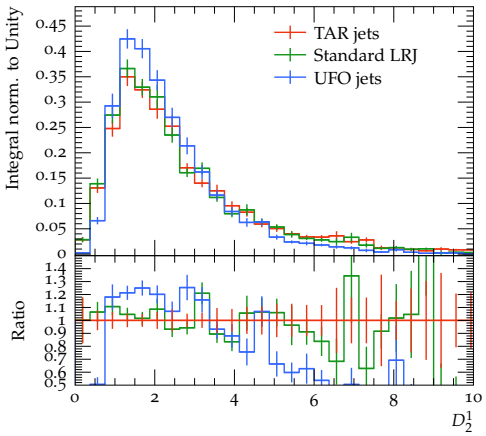


**Figure A.1(c):** ECF ratio $D_2^1$

**Figure A.1.:** Additional substructure observables for TAR jet, UFO jet and standard LRJ objects on QCD samples.

# A.2. Signal against QCD background

## Standard LRJ

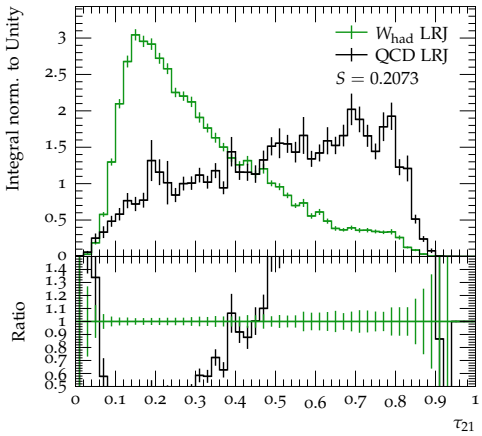


**Figure A.2(a):** Subjettiness $\tau_{21}$

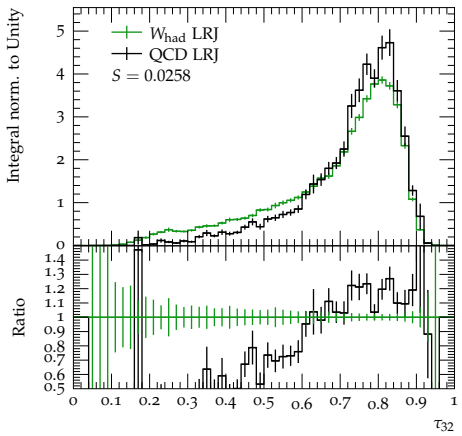


**Figure A.2(b):** Subjettiness $\tau_{32}$

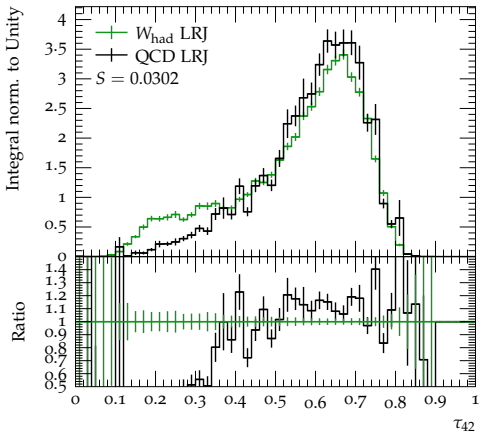


**Figure A.2(c):** Subjettiness $\tau_{42}$

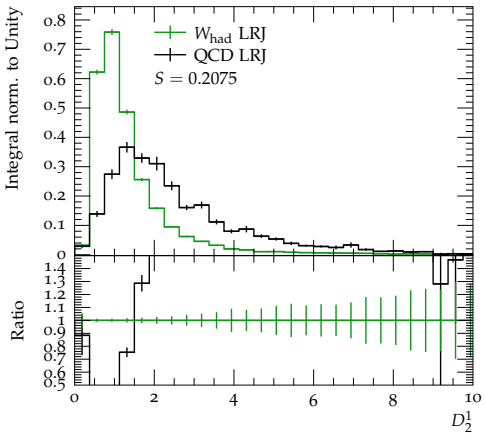


**Figure A.2(d):** ECF ratio $D_2^1$

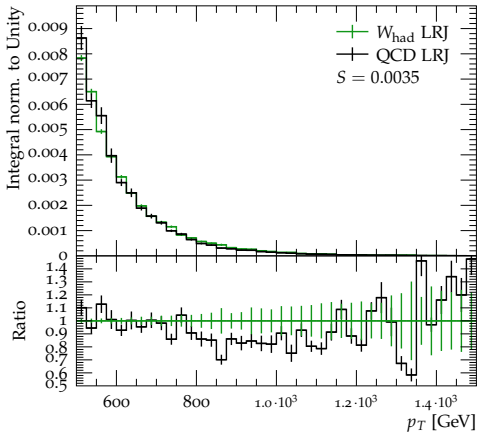


**Figure A.2(e):** Jet $p_T$

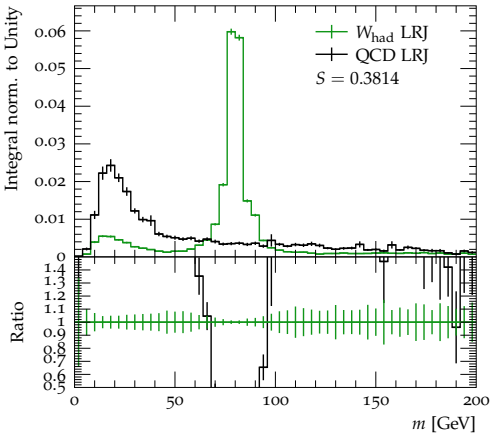


**Figure A.2(f):** Jet mass

**Figure A.2.:** Additional observables for the $W_{\text{had}}$ jet with standard LRJ objects on Higgs-Strahlung samples against QCD standard LRJ objects and their separation power $S$.

## UFO jets

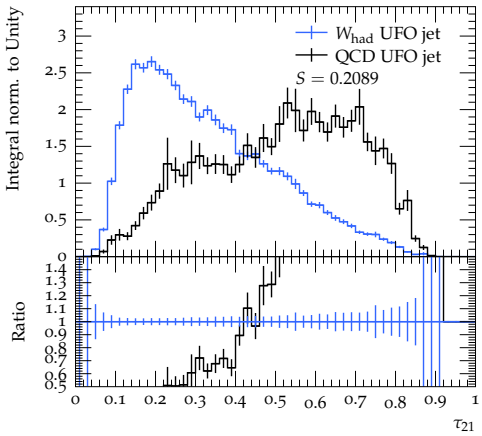


**Figure A.3(a):** Subjettiness $\tau_{21}$

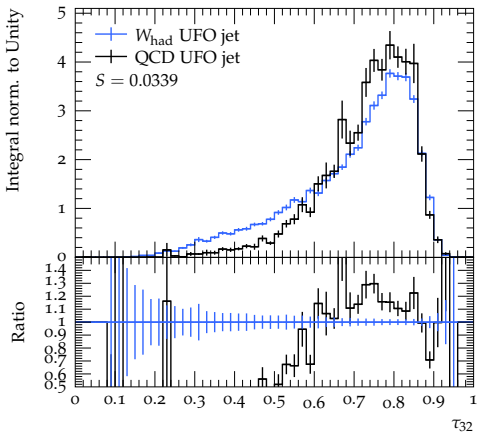


**Figure A.3(b):** Subjettiness $\tau_{32}$

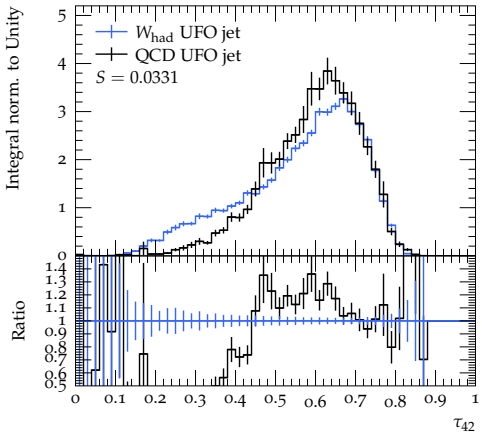


**Figure A.3(c):** Subjettiness $\tau_{42}$

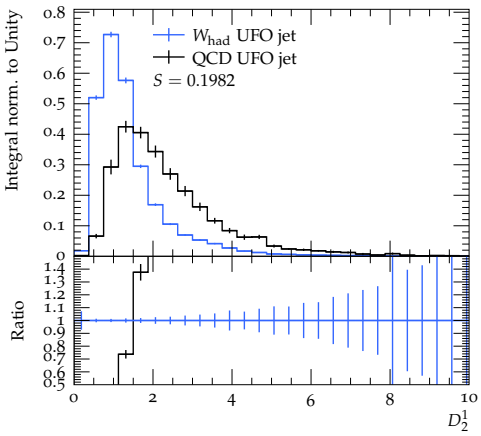


**Figure A.3(d):** ECF ratio $D_2^1$

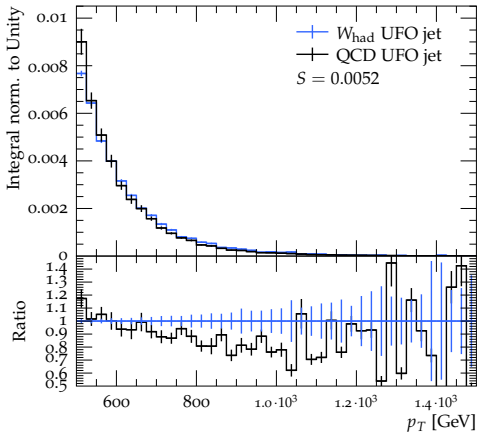


**Figure A.3(e):** Jet $p_T$

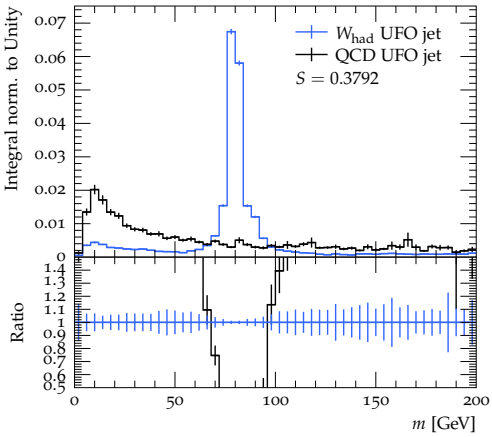


**Figure A.3(f):** Jet mass

**Figure A.3.:** Additional observables for the $W_{\text{had}}$ jet with UFO jet objects on Hiss-Strahlung samples against QCD UFO jet objects and their separation power $S$.

## TAR jets

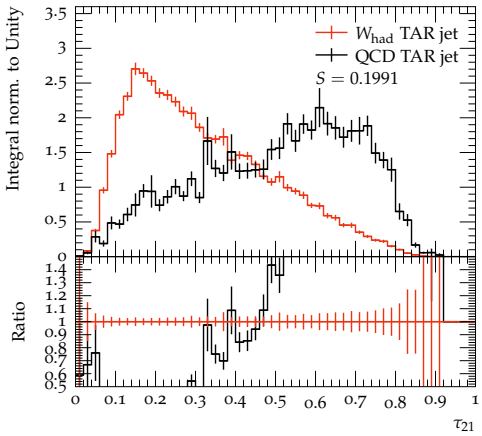


**Figure A.4(a):** Subjettiness $\tau_{21}$

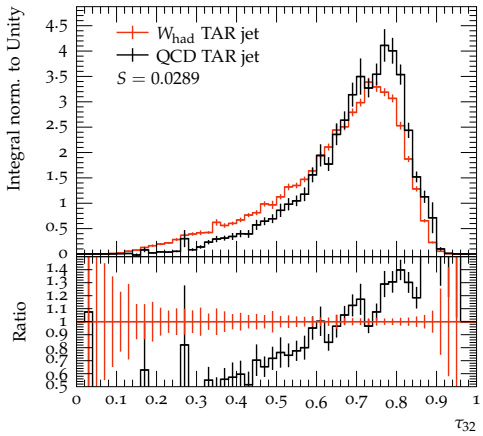


**Figure A.4(b):** Subjettiness $\tau_{32}$

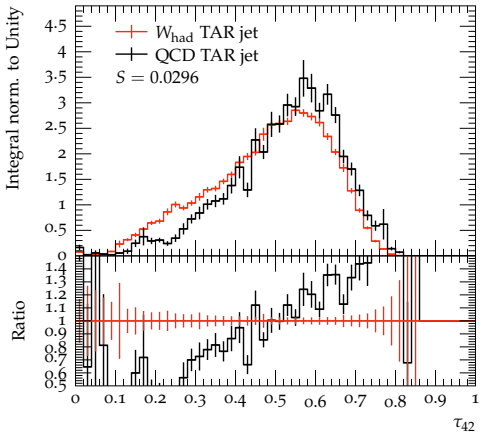


**Figure A.4(c):** Subjettiness $\tau_{42}$

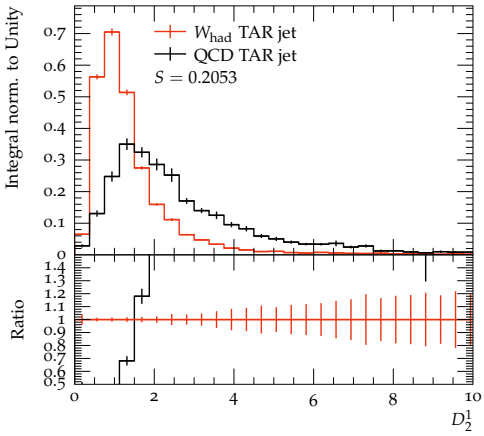


**Figure A.4(d):** ECF ratio $D_2^1$

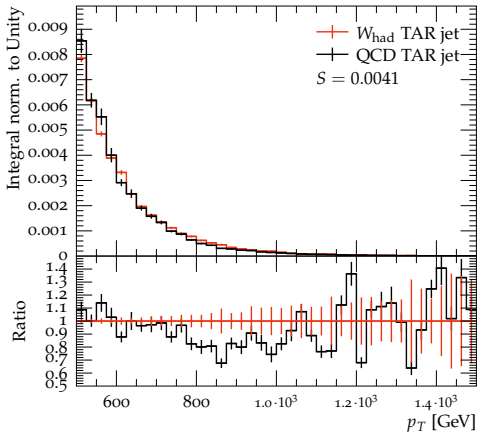


**Figure A.4(e):** Jet $p_T$

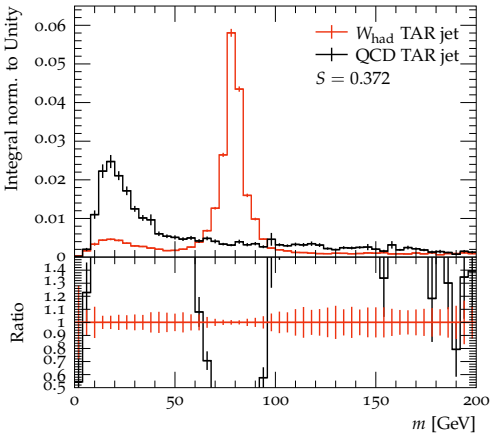


**Figure A.4(f):** Jet mass

**Figure A.4.:** Additional observables for the $W_{had}$ jet with TAR jet objects on Higgs-Strahlung samples against QCD TAR jet objects and their separation power $S$.

# A.3. Lepton Clustering in Higgs decay

## Standard LRJ

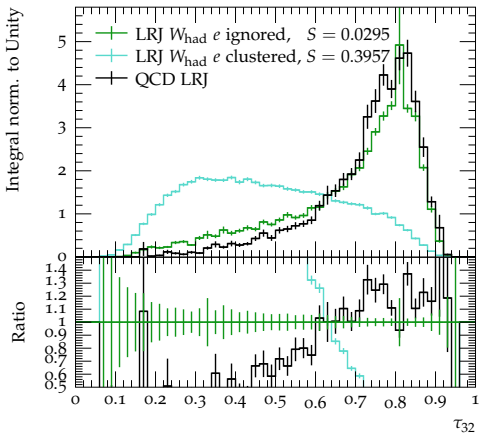


**Figure A.5(a):** Subjettiness $\tau_{32}$

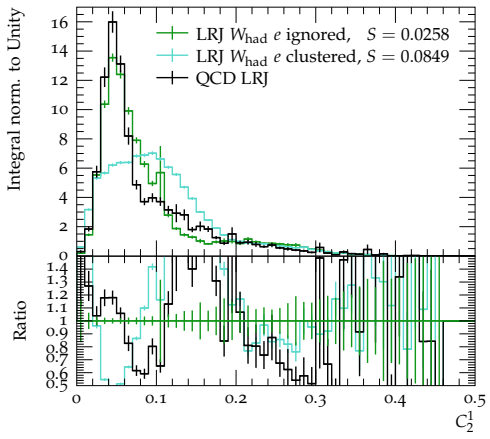


**Figure A.5(b):** ECF ratio $C_2^1$

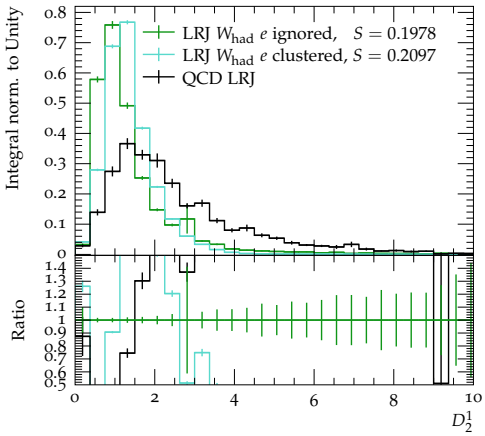


**Figure A.5(c):** ECF ratio $D_2^1$

**Figure A.5.:** Subjettiness $\tau_{32}$ and ECF ratios $C_2^1$ and $D_2^1$ for the $W_{\text{had}}$ jets with standard LRJ objects on Higgs-Strahlung samples compared to QCD standard LRJ. The signal electron was either ignored in the jet finding or included. The separation power $S$ is calculated between the $W_{\text{had}}$ distributions and the QCD distribution.

## UFO jets

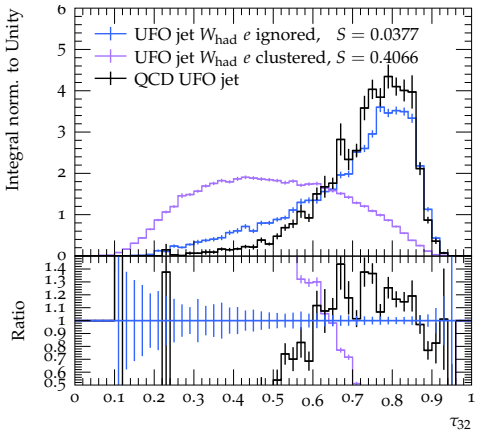


**Figure A.6(a):** Subjettiness $\tau_{32}$

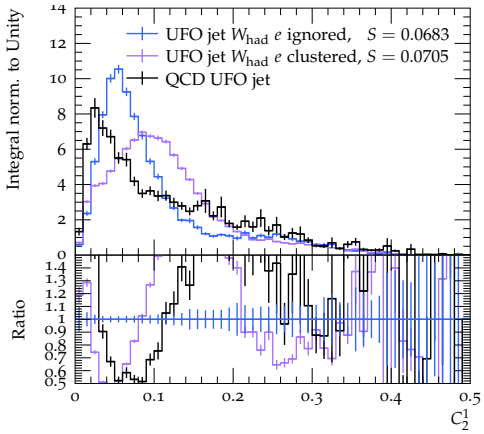


**Figure A.6(b):** ECF ratio $C_2^1$

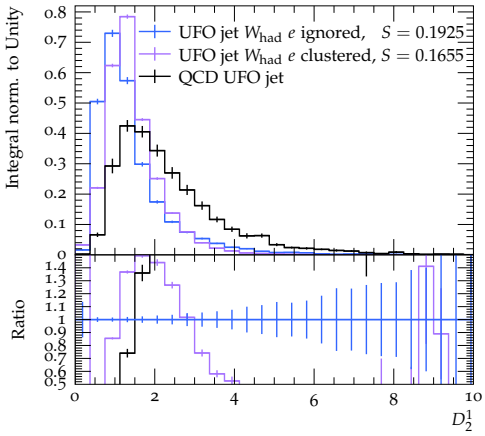


**Figure A.6(c):** ECF ratio $D_2^1$

**Figure A.6.:** Subjettiness $\tau_{32}$ and ECF ratios $C_2^1$ and $D_2^1$ for the $W_{\text{had}}$ jets with UFO jet objects on Higgs-Strahlung samples compared to QCD UFO jets. The signal electron was either ignored in the jet finding or included. The separation power $S$ is calculated between the $W_{\text{had}}$ distributions and the QCD distribution.

## TAR jets

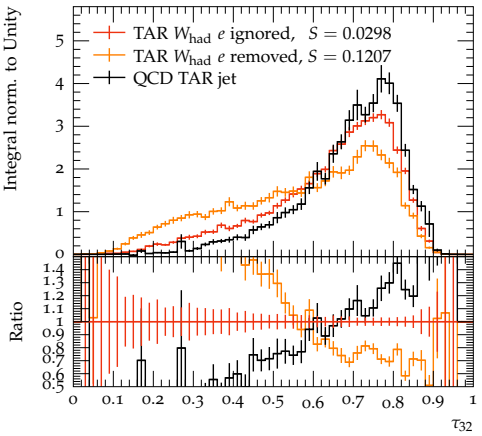


**Figure A.7(a):** Subjettiness $\tau_{32}$

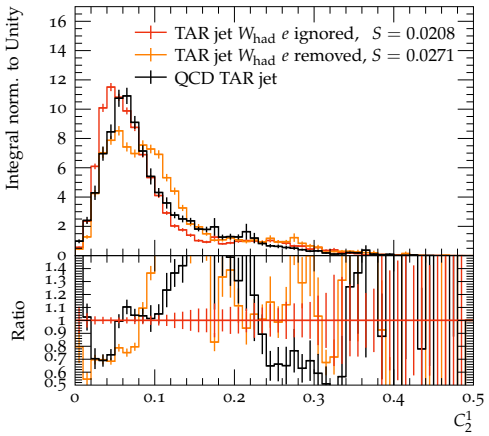


**Figure A.7(b):** ECF ratio $C_2^1$

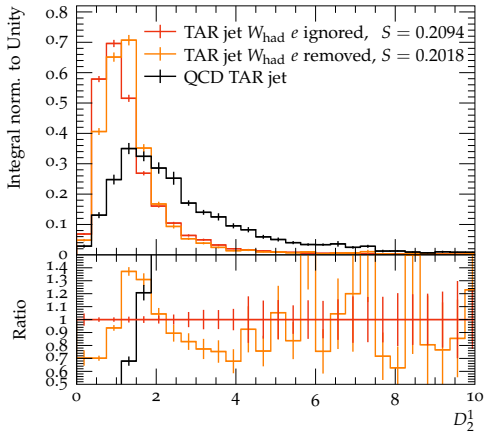


**Figure A.7(c):** ECF ratio $D_2^1$

**Figure A.7.:** Subjettiness $\tau_{32}$ and ECF ratios $C_2^1$ and $D_2^1$ for $W_{\text{had}}$ jets with TAR jet objects on Higgs-Strahlung samples compared to QCD TAR jets. The signal electron was either ignored in the jet finding or included and retroactively removed. The separation power $S$ is calculated between the $W_{\text{had}}$ distributions and the QCD distribution.

## A.4. Systematic Variations

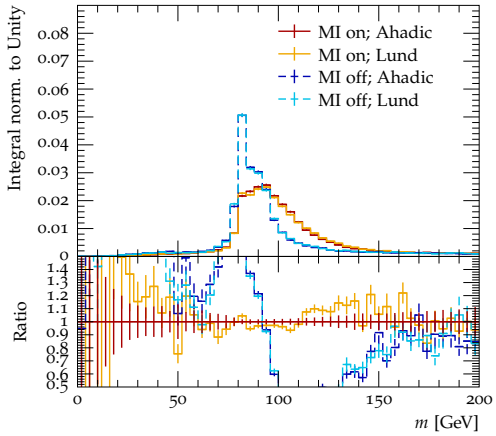


**Figure A.8(a):** Naive jet

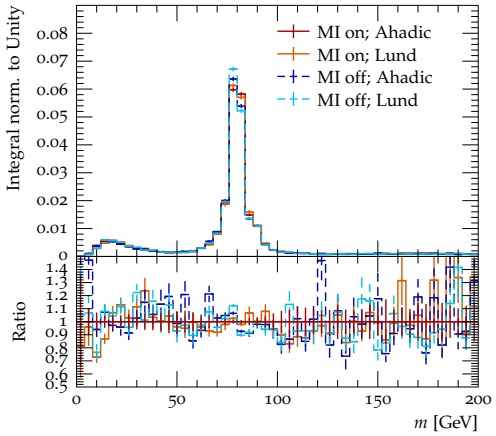


**Figure A.8(b):** Standard LRJ

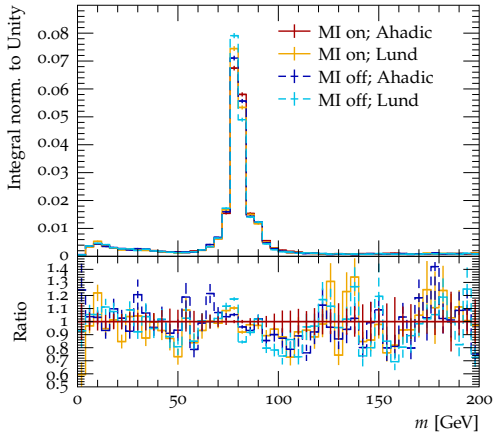


**Figure A.8(c):** UFO jet

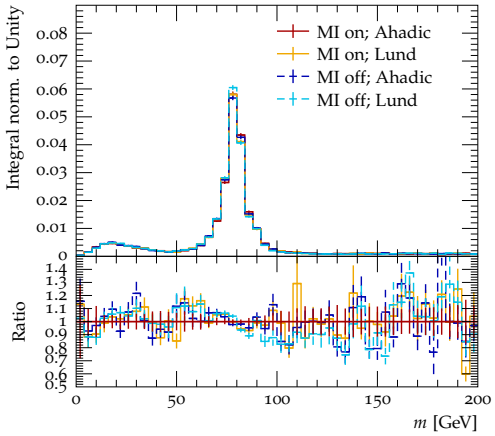


**Figure A.8(d):** TAR jet

**Figure A.8.:** Mass distribution for UFO jets, TAR jets, standard LRJ and naive jets on signal samples. The SHERPA settings for the MPI and hadronization are varied.

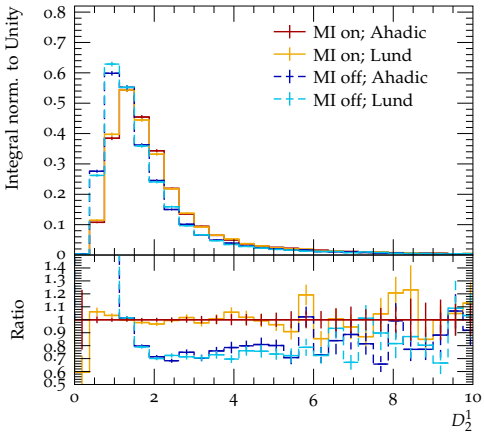


**Figure A.9(a):** Naive jet

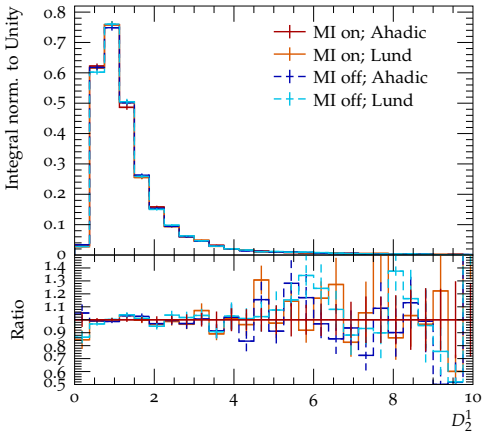


**Figure A.9(b):** Standard LRJ

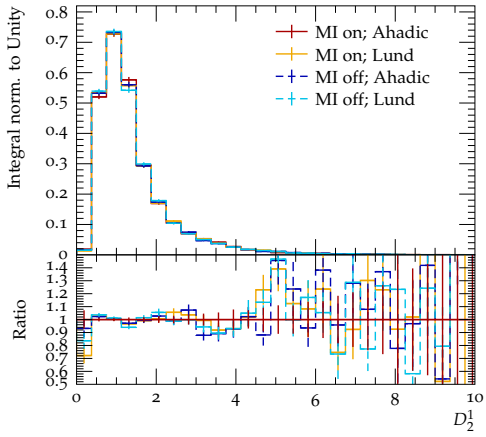


**Figure A.9(c):** UFO jet

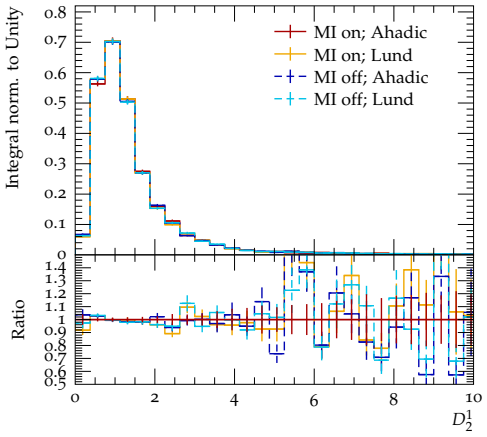


**Figure A.9(d):** TAR jet

**Figure A.9.:** ECF ratio $D_2^1$ for UFO jets, TAR jets, standard LRJ and naive jets on signal samples. The SHERPA settings for the MPI and hadronization are varied.

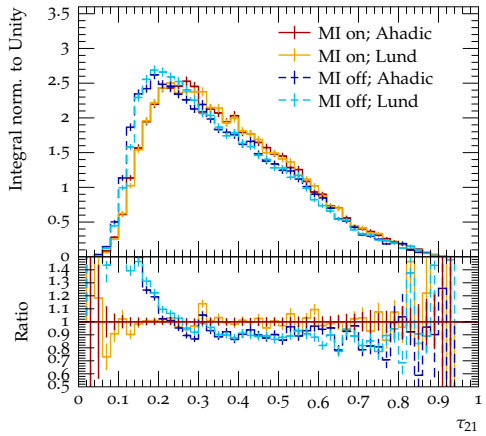


**Figure A.10(a):** Naive jet

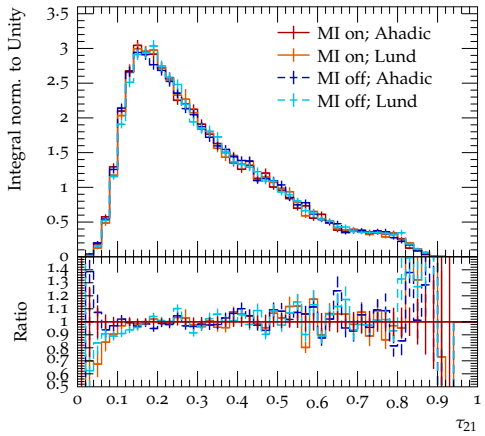


**Figure A.10(b):** Standard LRJ

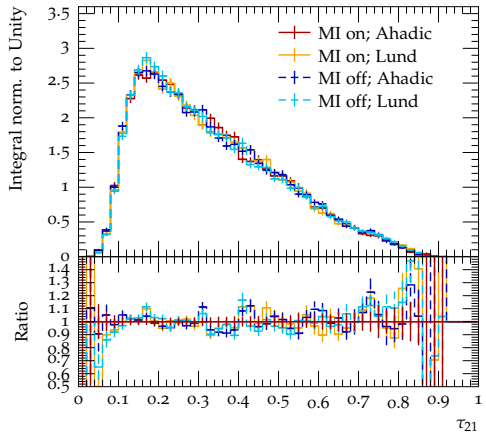


**Figure A.10(c):** UFO jet

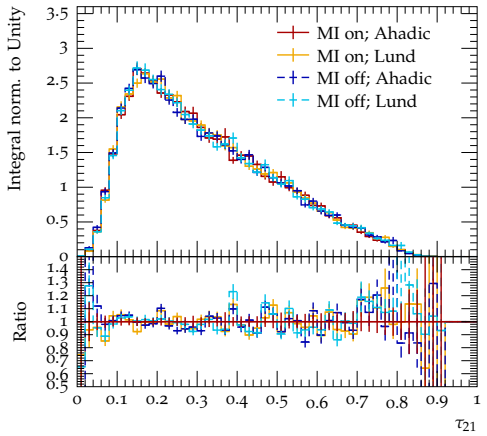


**Figure A.10(d):** TAR jet

**Figure A.10.:** Subjettiness $\tau_{21}$ for UFO jets, TAR jets, standard LRJ and naive jets on signal samples. The SHERPA settings for the MPI and hadronization are varied.

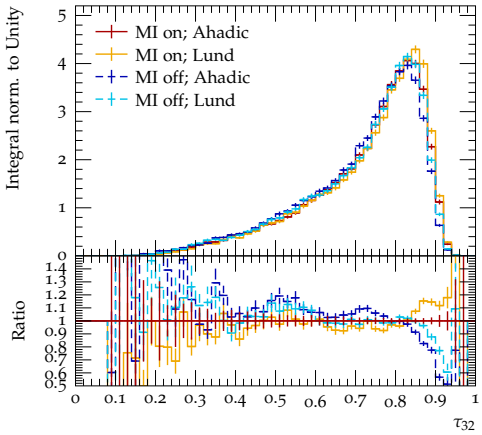


**Figure A.11(a):** Naive jet

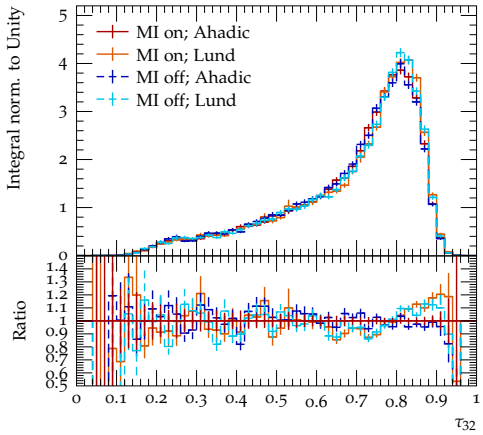


**Figure A.11(b):** Standard LRJ

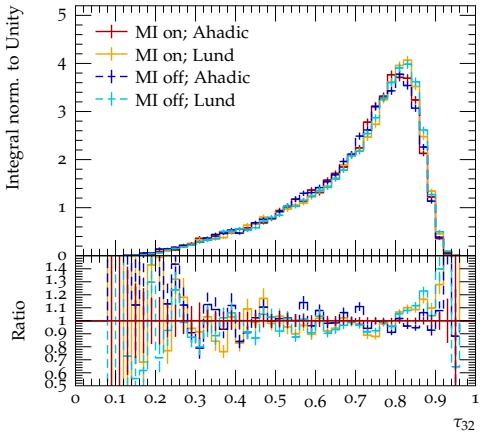


**Figure A.11(c):** UFO jet

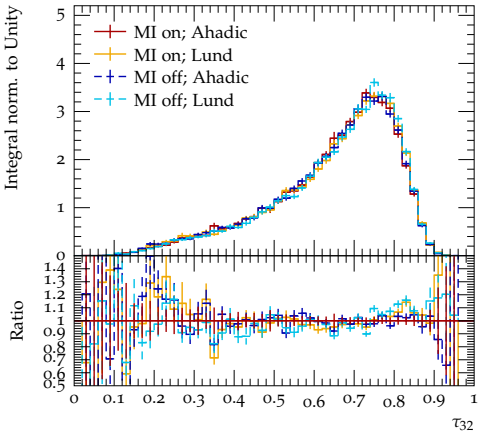


**Figure A.11(d):** TAR jet

**Figure A.11.:** Subjettiness $\tau_{32}$ for UFO jets, TAR jets, standard LRJ and naive jets on signal samples. The SHERPA settings for the MPI and hadronization are varied.

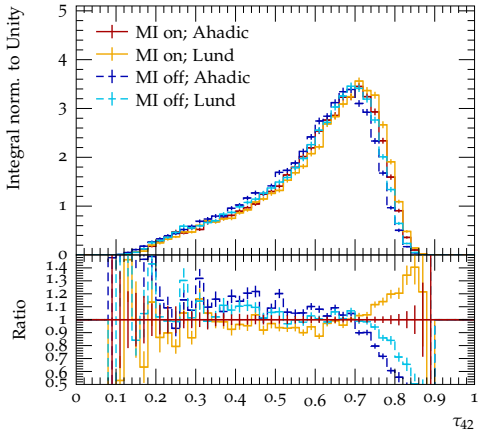


**Figure A.12(a):** Naive jet

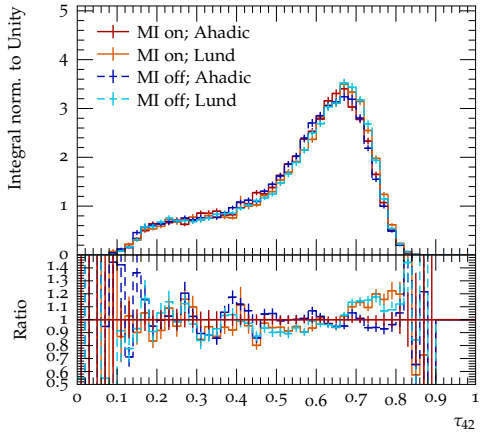


**Figure A.12(b):** Standard LRJ

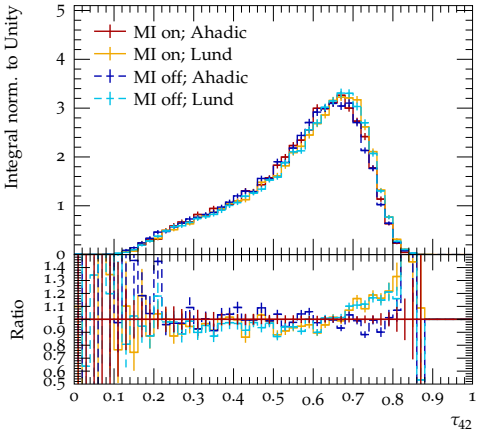


**Figure A.12(c):** UFO jet

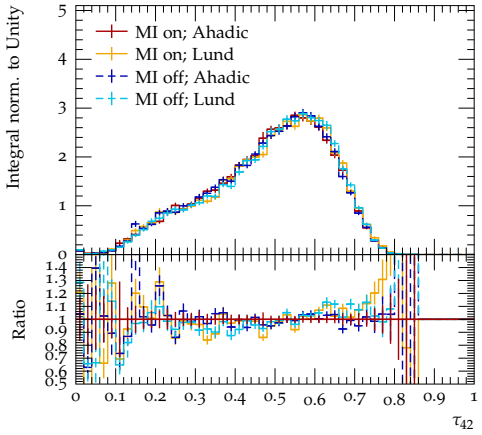


**Figure A.12(d):** TAR jet

**Figure A.12.:** Subjettiness $\tau_{42}$ for UFO jets, TAR jets, standard LRJ and naive jets on signal samples. The SHERPA settings for the MPI and hadronization are varied.

# A.5. Jet Object Comparison with ATLAS Samples

## A.5.1. Reweighting of Sherpa Samples

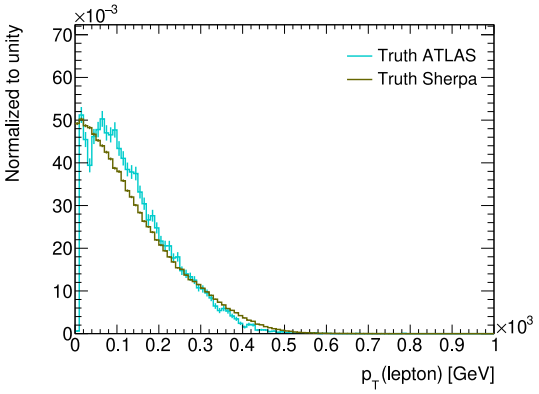


**Figure A.13(a):** $p_T(\ell)$

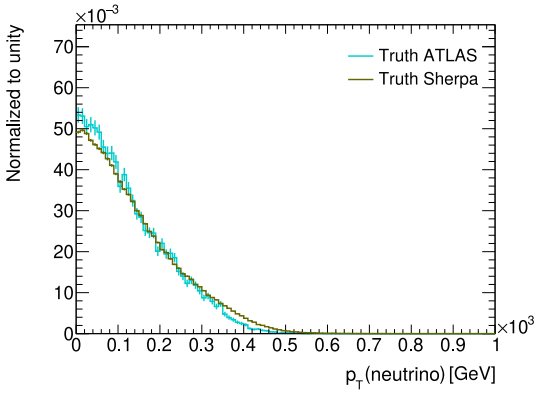


**Figure A.13(b):** $p_T(\nu)$

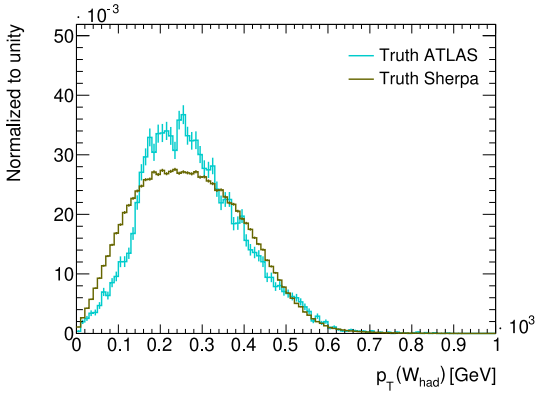


**Figure A.13(c):** $p_T(W_{\text{had}})$

**Figure A.13.:** Comparison of ATLAS and SHERPA samples for the neutrino, lepton and $W_{\text{had}}$ $p_T$ on particle level, after reweighting of the SHERPA samples.

### A.5.2. Sherpa Samples against ATLAS Samples

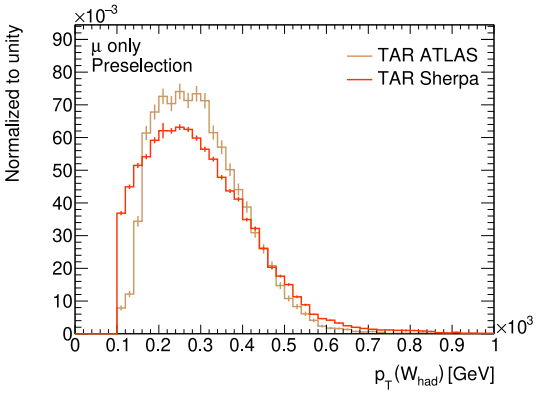


**Figure A.14(a):** TAR jets

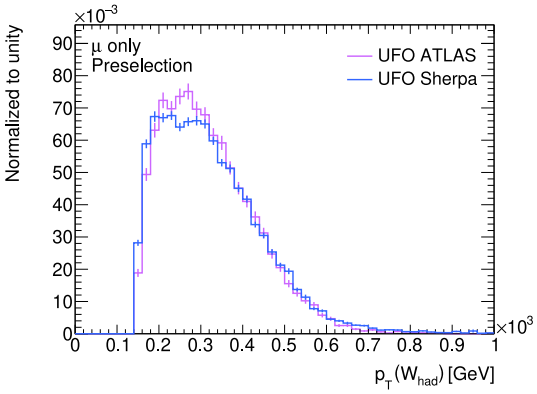


**Figure A.14(b):** UFO jets

**Figure A.14.:** Comparison of ATLAS and SHERPA samples for the $p_T(W_{\text{had}})$ distribution for TAR jets and UFO jets in the muon channel.

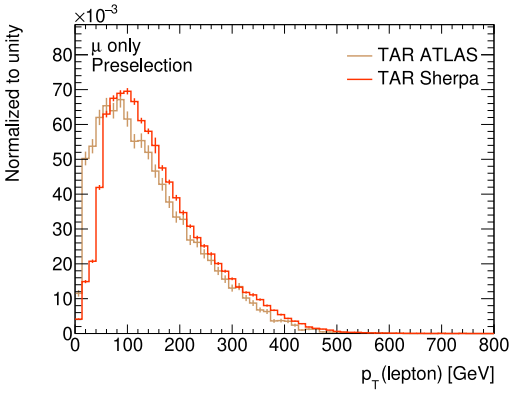


**Figure A.15(a):** TAR jets

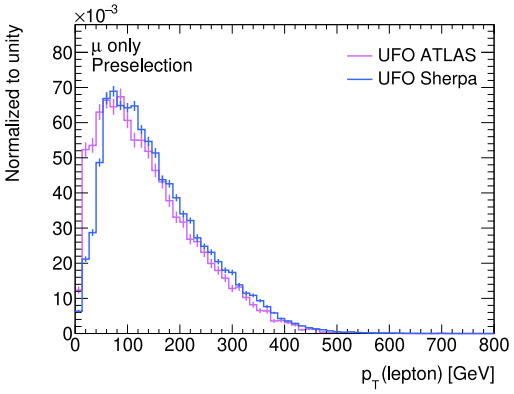


**Figure A.15(b):** UFO jets

**Figure A.15.:** Comparison of ATLAS and SHERPA samples for the $p_T(\ell)$ distribution for TAR jets and UFO jets in the muon channel.

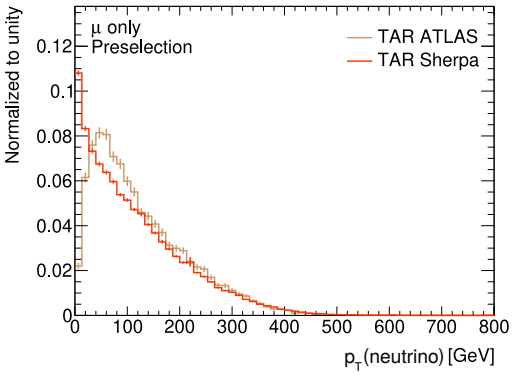


**Figure A.16(a):** TAR jets

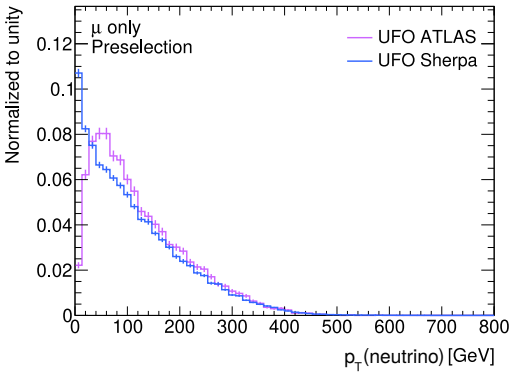


**Figure A.16(b):** UFO jets

**Figure A.16.:** Comparison of ATLAS and SHERPA samples for the $p_T(\nu)$ distribution for TAR jets and UFO jets in the muon channel.

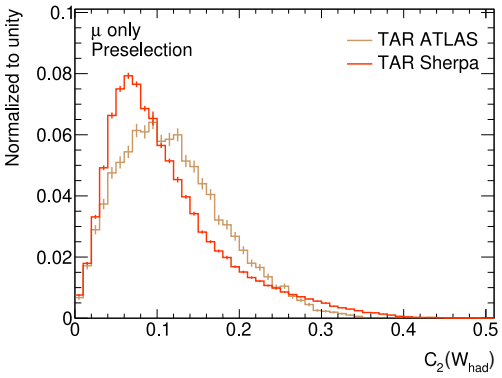


**Figure A.17(a):** TAR jets

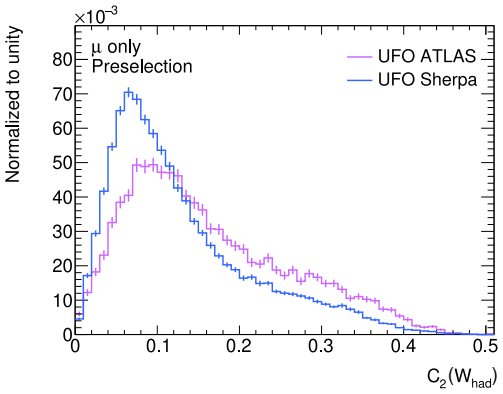


**Figure A.17(b):** UFO jets

**Figure A.17.:** Comparison of ATLAS and SHERPA samples for the $C_2(W_{\text{had}})$ distribution for TAR jets and UFO jets in the muon channel.

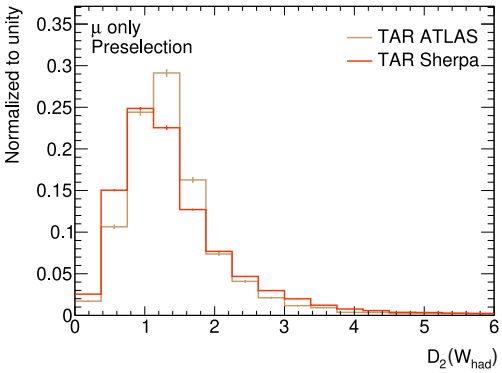


**Figure A.18(a):** TAR jets

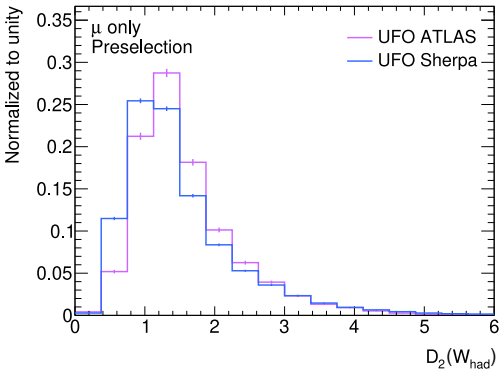


**Figure A.18(b):** UFO jets

**Figure A.18.:** Comparison of ATLAS and SHERPA samples for the $D_2(W_{\text{had}})$ distribution for TAR jets and UFO jets in the muon channel.

### A.5.3. TAR against UFO Jets for ATLAS Samples

#### Muon Channel

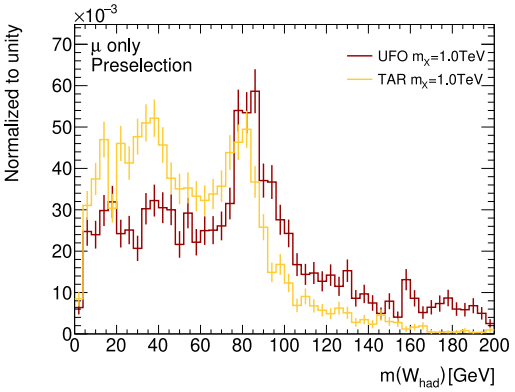


**Figure A.19(a):** $m_X = 1\,\text{TeV}$, $m(W_{\text{had}})$

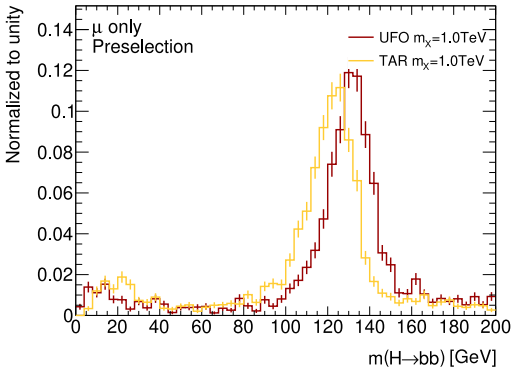


**Figure A.19(b):** $m_X = 1\,\text{TeV}$, $m(H \to b\bar{b})$

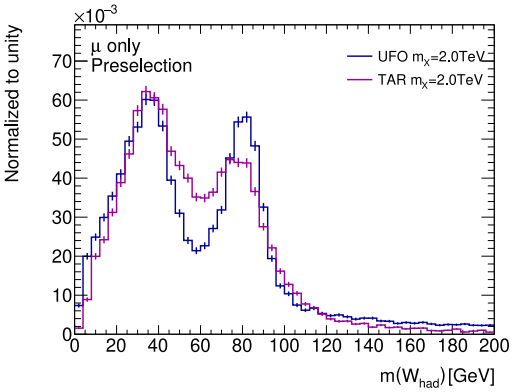


**Figure A.19(c):** $m_X = 2\,\text{TeV}$, $m(W_{\text{had}})$

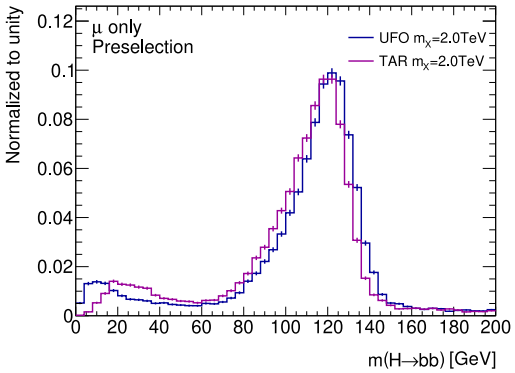


**Figure A.19(d):** $m_X = 2\,\text{TeV}$, $m(H \to b\bar{b})$

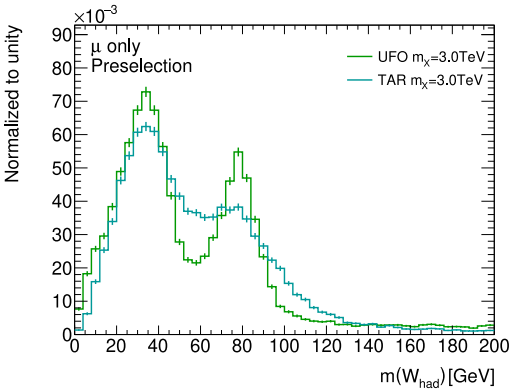


**Figure A.19(e):** $m_X = 3\,\text{TeV}$, $m(W_{\text{had}})$

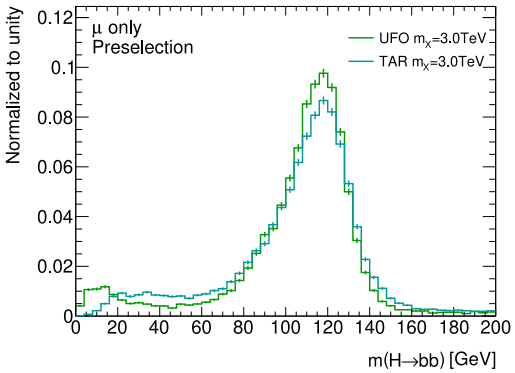


**Figure A.19(f):** $m_X = 3\,\text{TeV}$, $m(H \to b\bar{b})$

**Figure A.19.:** Comparison between TAR jets and UFO jets for $HH$ samples for different mass points $m_X$ in the muon channel for the $W_{\text{had}}$ and $H \to b\bar{b}$ jet mass.

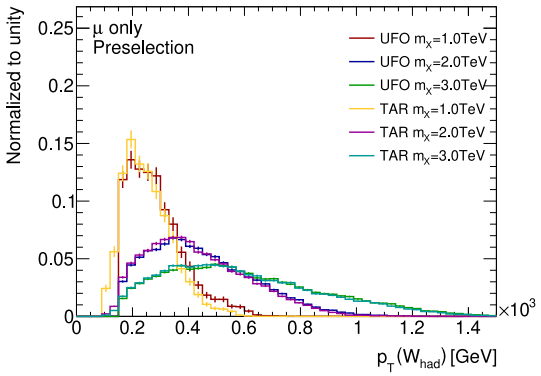


**Figure A.20(a):** $p_T(W_{\text{had}})$

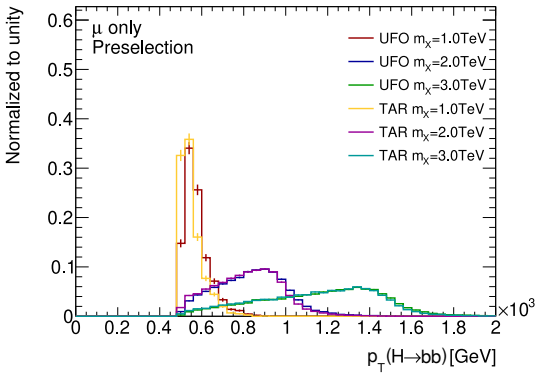


**Figure A.20(b):** $p_T(H \to b\bar{b})$

**Figure A.20.:** Comparison between TAR jets and UFO jets for $HH$ samples for different mass points $m_X$ in the muon channel for $p_T(W_{\text{had}})$ and $p_T(H \to b\bar{b})$.

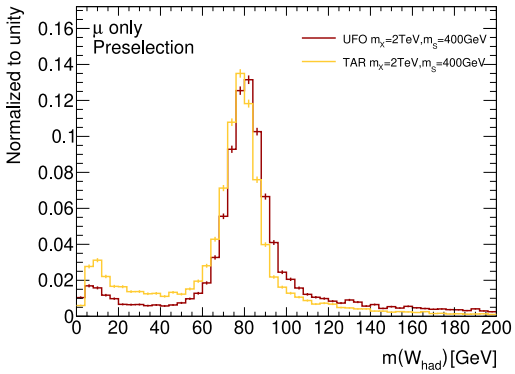


**Figure A.21(a):** $m_X = 2\,\text{TeV}$, $m_S = 0.4\,\text{TeV}$ $m(W_{\text{had}})$

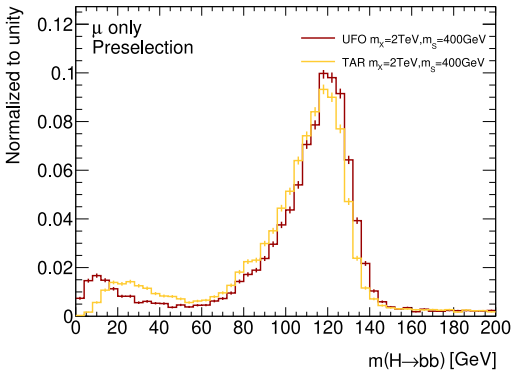


**Figure A.21(b):** $m_X = 2\,\text{TeV}$, $m_S = 0.4\,\text{TeV}$ $m(H \to b\bar{b})$

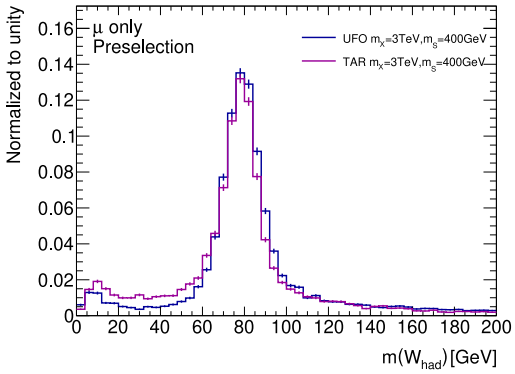


**Figure A.21(c):** $m_X = 3\,\text{TeV}$, $m_S = 0.4\,\text{TeV}$ $m(W_{\text{had}})$

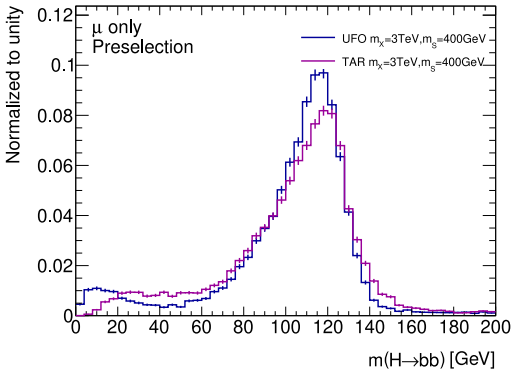


**Figure A.21(d):** $m_X = 3\,\text{TeV}$, $m_S = 0.4\,\text{TeV}$ $m(H \to b\bar{b})$

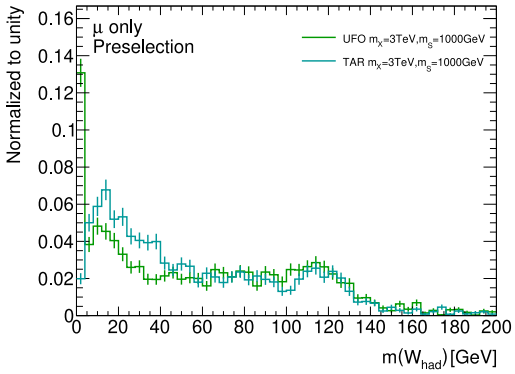


**Figure A.21(e):** $m_X = 3\,\text{TeV}$, $m_S = 1\,\text{TeV}$ $m(W_{\text{had}})$

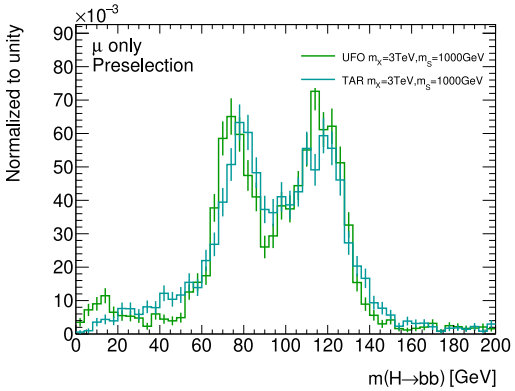


**Figure A.21(f):** $m_X = 3\,\text{TeV}$, $m_S = 1\,\text{TeV}$ $m(H \to b\bar{b})$

**Figure A.21.:** Comparison between TAR and UFO jets for $SH$ samples for different mass points $m_X$, $m_S$ in the muon channel for the $W_{\text{had}}$ and $H \to b\bar{b}$ jet mass.

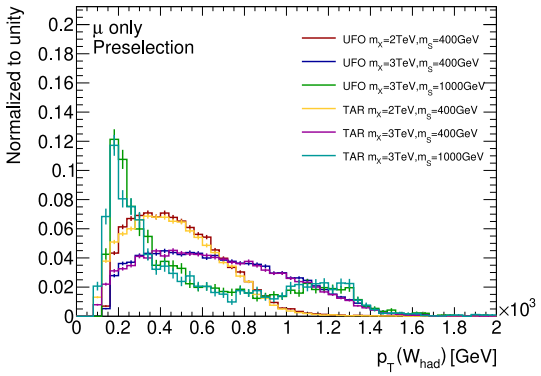


**Figure A.22(a):** $p_T(W_{\text{had}})$

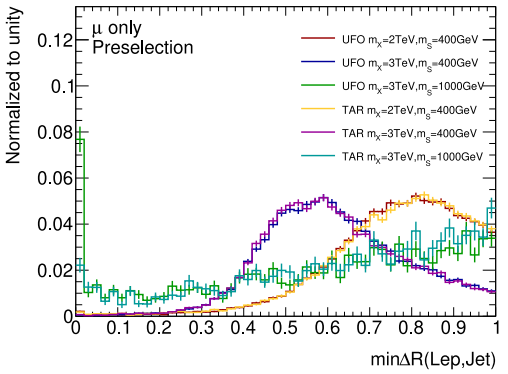


**Figure A.22(b):** $\min \Delta R(\ell, \text{jet})$

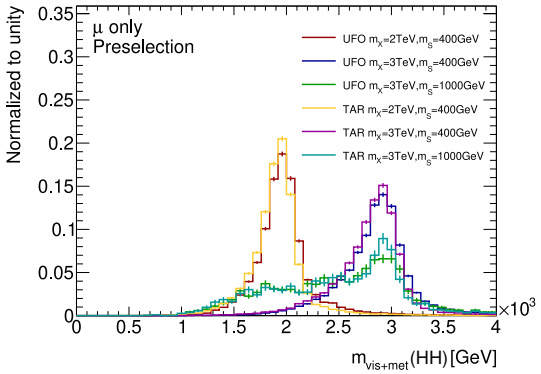


**Figure A.22(c):** $m_{\text{vis+met}}(SH)$

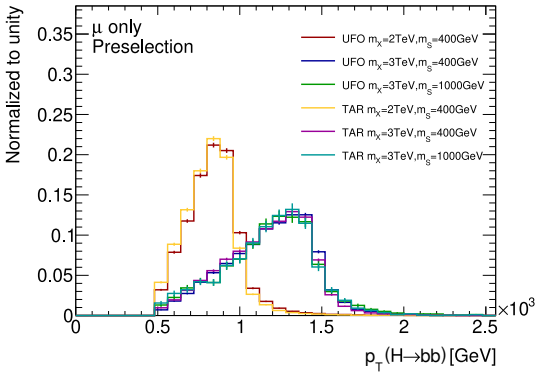


**Figure A.22(d):** $p_T(H \to b\bar{b})$

**Figure A.22.:** Comparison between TAR and UFO jets for $SH$ samples for different mass points $m_X$, $m_S$ in the muon channel for various observables.

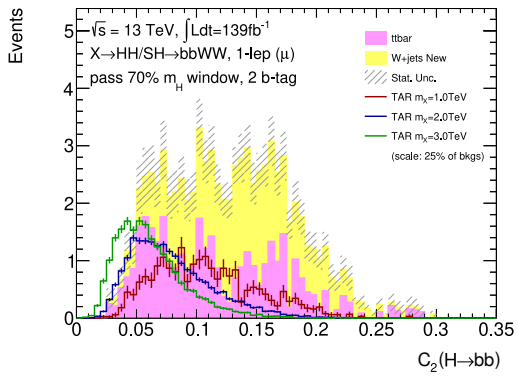


**Figure A.23(a):** TAR jet

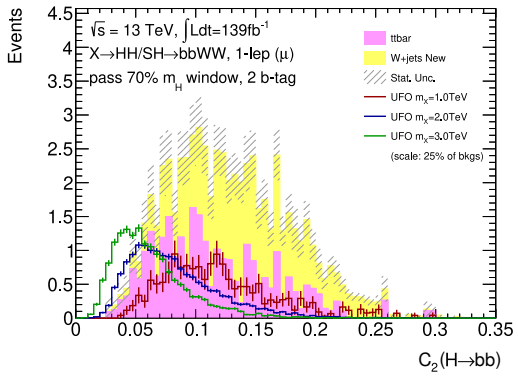


**Figure A.23(b):** UFO jet

**Figure A.23.:** $HH$ samples over the most contributing backgrounds at different mass points $m_X$ in the SRp2 signal region for $C_2(H \to b\bar{b})$. The signal samples are scaled to equal 25% of the summed backgrounds.

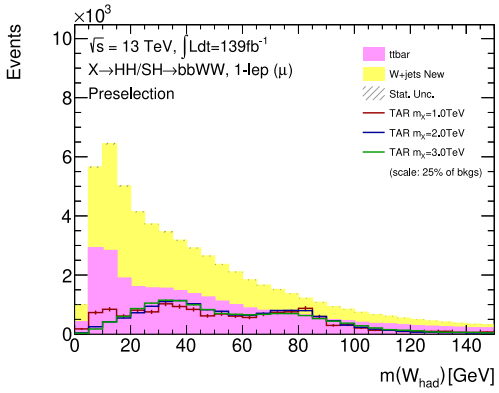


**Figure A.24(a):** TAR jet, $m(W_{\text{had}})$

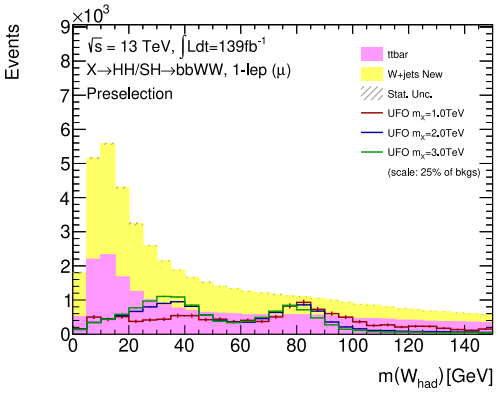


**Figure A.24(b):** UFO jet, $m(W_{\text{had}})$

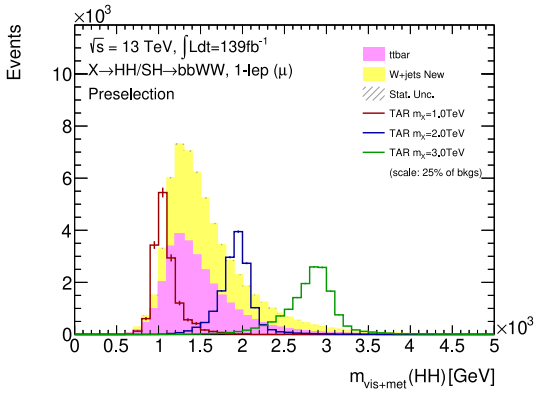


**Figure A.24(c):** TAR jet, $m_{\text{vis+met}}(HH)$

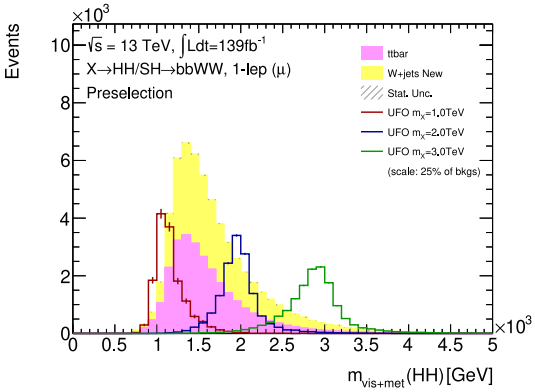


**Figure A.24(d):** UFO jet, $m_{\text{vis+met}}(HH)$

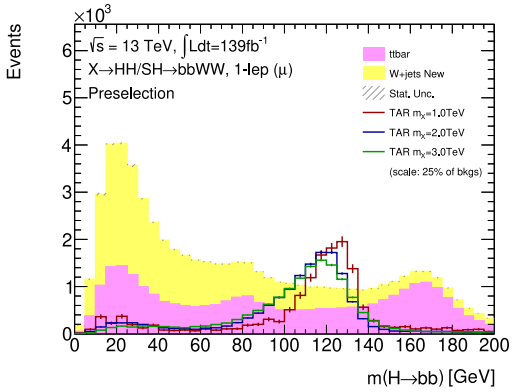


**Figure A.24(e):** TAR jet, $m(H \to b\bar{b})$

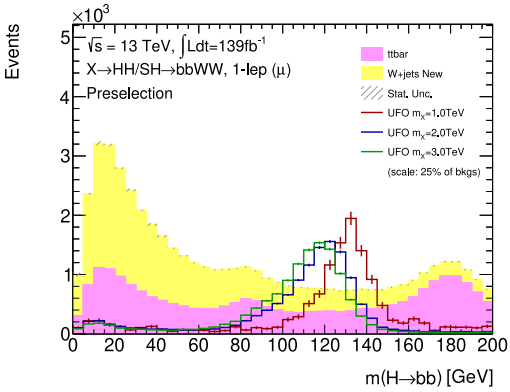


**Figure A.24(f):** UFO jet, $m(H \to b\bar{b})$

**Figure A.24.:** $HH$ samples over the most contributing backgrounds at different mass points $m_X$ after the boosted selection for various observables. The signal samples are scaled to equal 25% of the summed backgrounds.

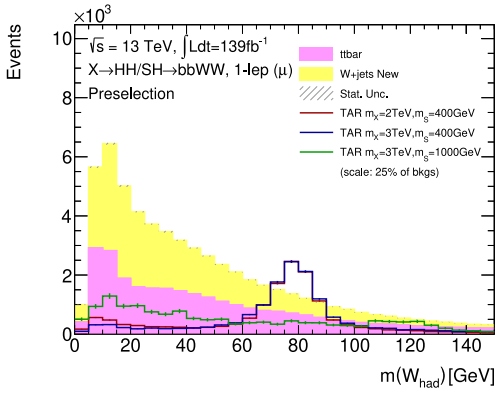


**Figure A.25(a):** TAR jet, $m(W_{\text{had}})$

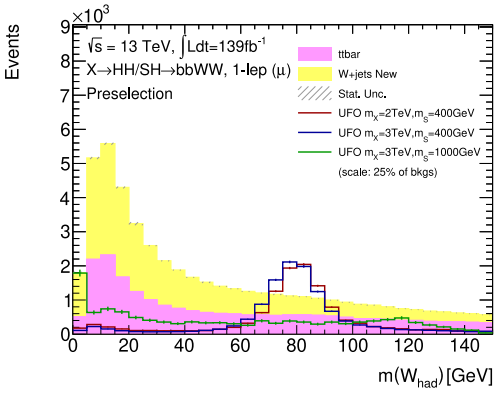


**Figure A.25(b):** UFO jet, $m(W_{\text{had}})$

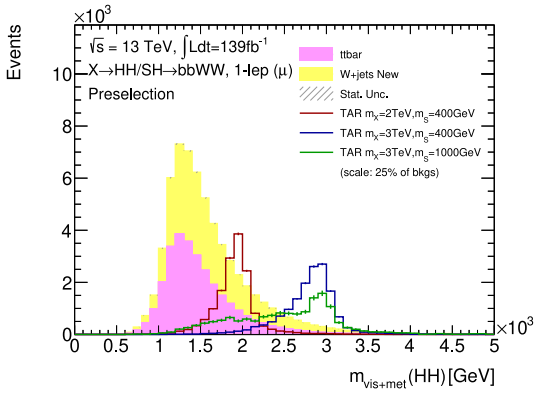


**Figure A.25(c):** TAR jet, $m_{\text{vis+met}}(HH)$

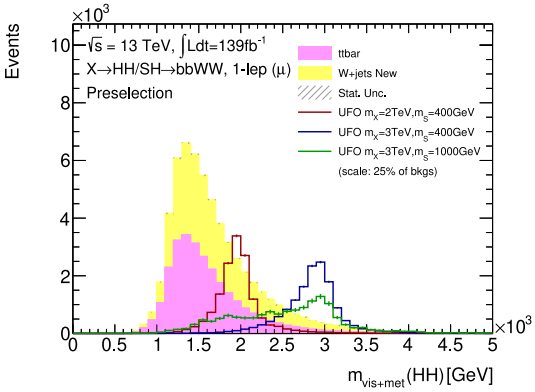


**Figure A.25(d):** UFO jet, $m_{\text{vis+met}}(HH)$

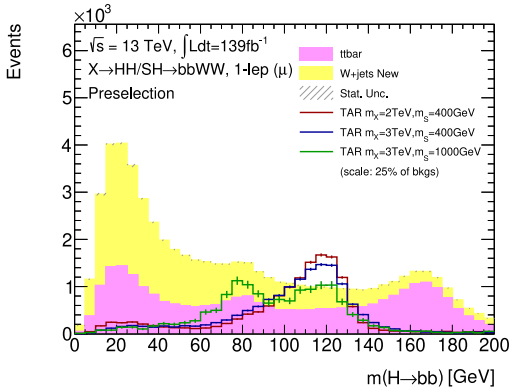


**Figure A.25(e):** TAR jet, $m(H \to b\bar{b})$

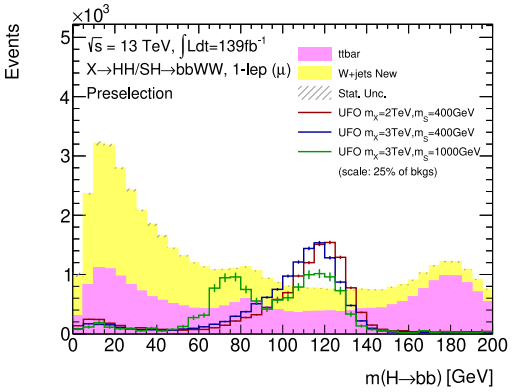


**Figure A.25(f):** UFO jet, $m(H \to b\bar{b})$

**Figure A.25.:** $SH$ samples over the most contributing backgrounds at different mass points $m_X$, $m_S$ after the boosted selection for various observables. The signal samples are scaled to equal 25% of the summed backgrounds.

### Electron Channel

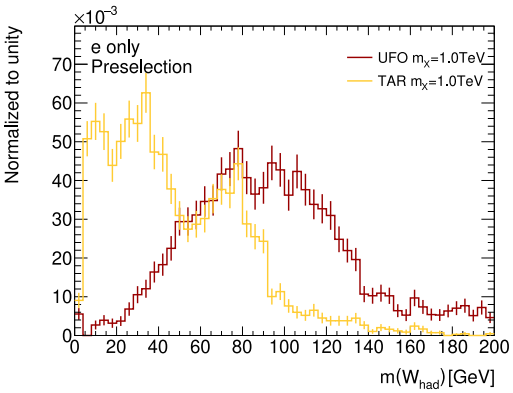


**Figure A.26(a):** $m_X = 1\,\text{TeV}$, $m(W_{\text{had}})$

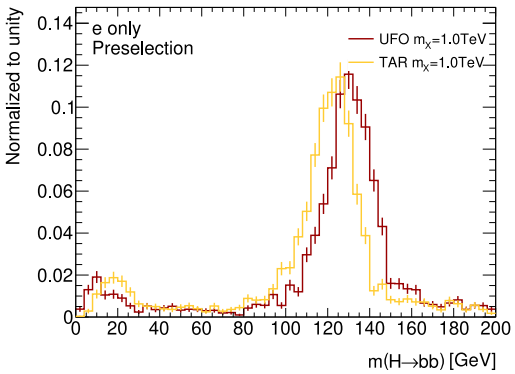


**Figure A.26(b):** $m_X = 1\,\text{TeV}$, $m(H \to b\bar{b})$

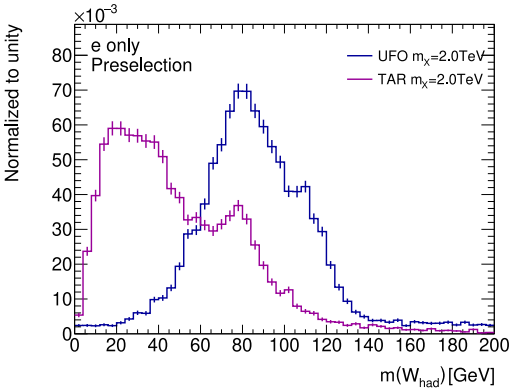


**Figure A.26(c):** $m_X = 2\,\text{TeV}$, $m(W_{\text{had}})$

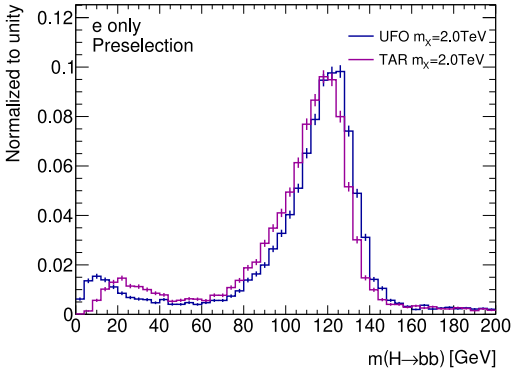


**Figure A.26(d):** $m_X = 2\,\text{TeV}$, $m(H \to b\bar{b})$

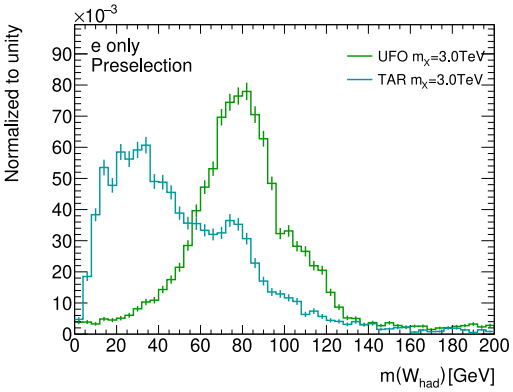


**Figure A.26(e):** $m_X = 3\,\text{TeV}$, $m(W_{\text{had}})$

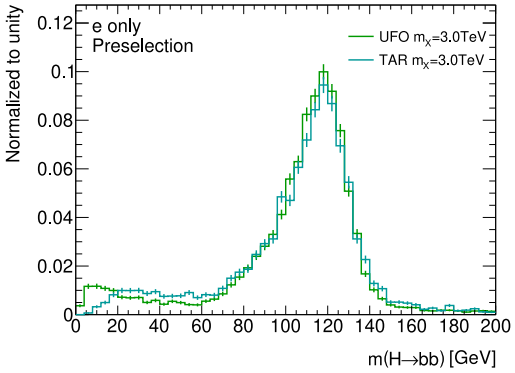


**Figure A.26(f):** $m_X = 3\,\text{TeV}$, $m(H \to b\bar{b})$

**Figure A.26.:** Comparison between TAR jets and UFO jets for $HH$ samples for different mass points $m_X$ in the electron channel for the $W_{\text{had}}$ and $H \to b\bar{b}$ jet mass.

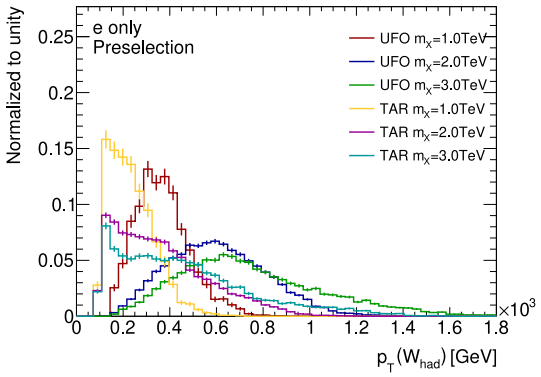


**Figure A.27(a):** $p_T(W_{\text{had}})$

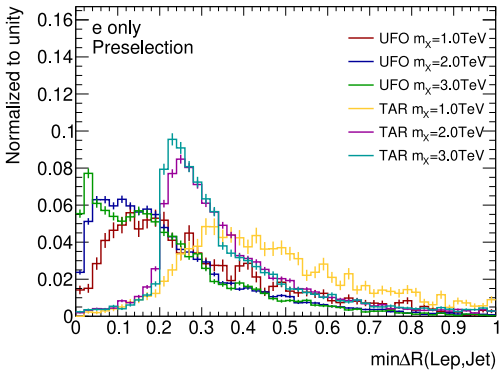


**Figure A.27(b):** $\min \Delta R(\ell, \text{jet})$

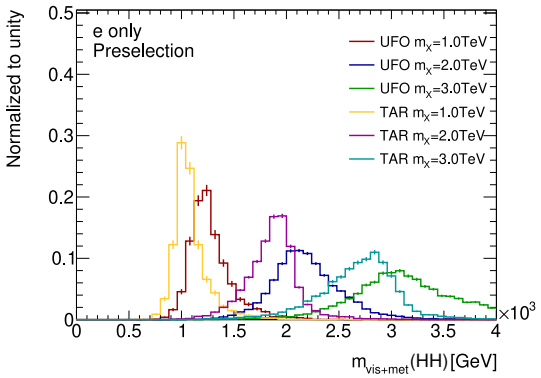


**Figure A.27(c):** $m_{\text{vis+met}}(HH)$

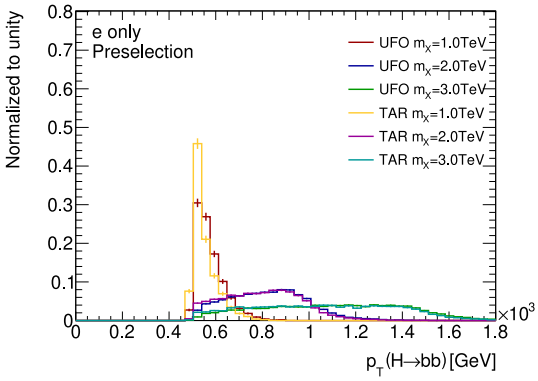


**Figure A.27(d):** $p_T(H \to b\bar{b})$

**Figure A.27.:** Comparison between TAR jets and UFO jets for $HH$ samples for different mass points $m_X$ in the electron channel for various observables.

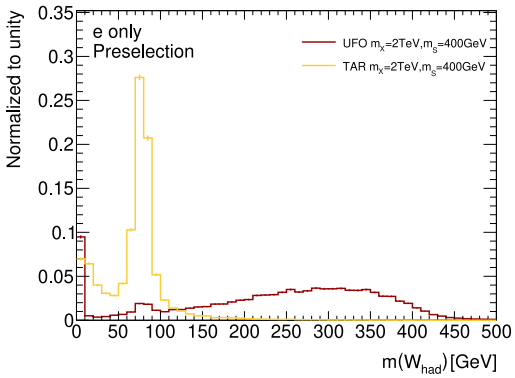


**Figure A.28(a):** $m_X = 2\,\text{TeV}$, $m_S = 0.4\,\text{TeV}$ $m(W_{\text{had}})$

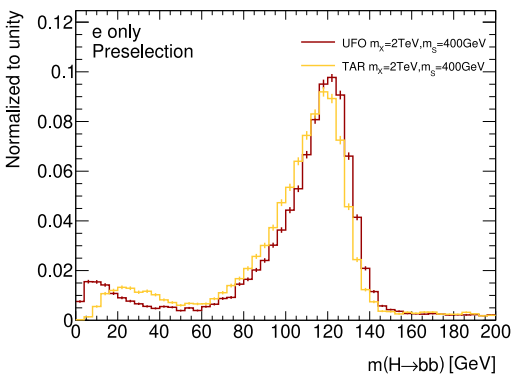


**Figure A.28(b):** $m_X = 2\,\text{TeV}$, $m_S = 0.4\,\text{TeV}$ $m(H \to b\bar{b})$

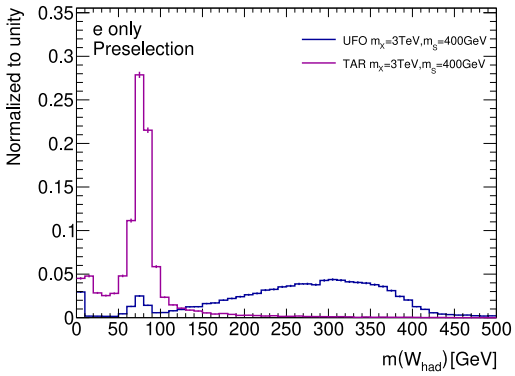


**Figure A.28(c):** $m_X = 3\,\text{TeV}$, $m_S = 0.4\,\text{TeV}$ $m(W_{\text{had}})$

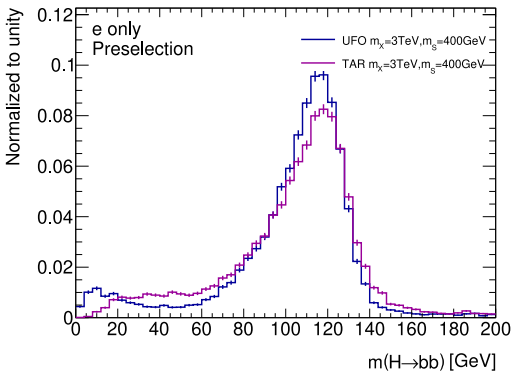


**Figure A.28(d):** $m_X = 3\,\text{TeV}$, $m_S = 0.4\,\text{TeV}$ $m(H \to b\bar{b})$

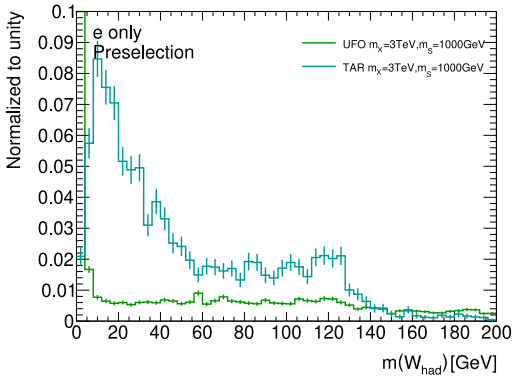


**Figure A.28(e):** $m_X = 3\,\text{TeV}$, $m_S = 1\,\text{TeV}$ $m(W_{\text{had}})$

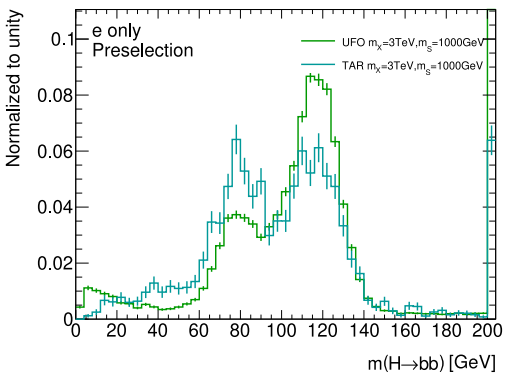


**Figure A.28(f):** $m_X = 3\,\text{TeV}$, $m_S = 1\,\text{TeV}$ $m(H \to b\bar{b})$

**Figure A.28.:** Comparison between TAR jets and UFO jets for $SH$ samples for different mass points $m_X$, $m_S$ in the electron channel for the $W_{\text{had}}$ and $H \to b\bar{b}$ jet mass.

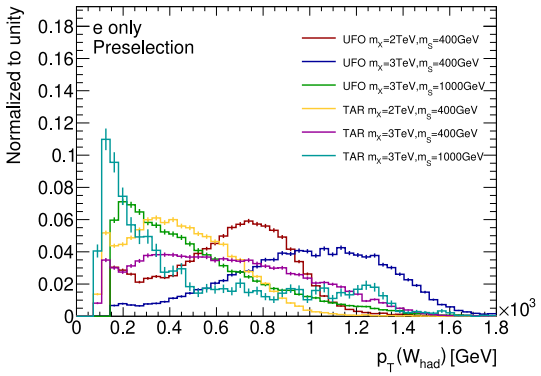


**Figure A.29(a):** $p_T(W_{\text{had}})$

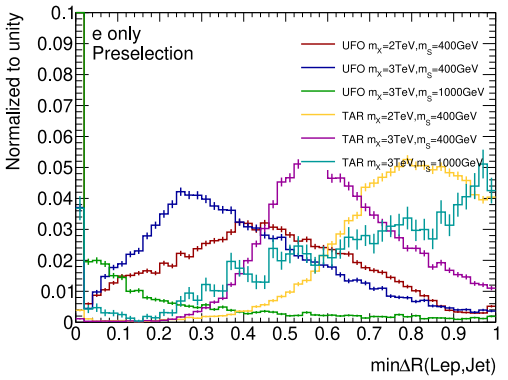


**Figure A.29(b):** $\min \Delta R(\ell, \text{jet})$

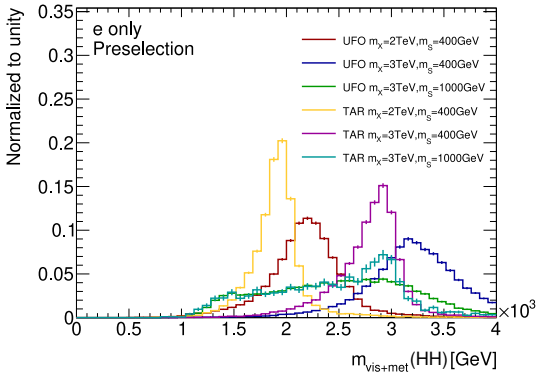


**Figure A.29(c):** $m_{\text{vis+met}}(SH)$

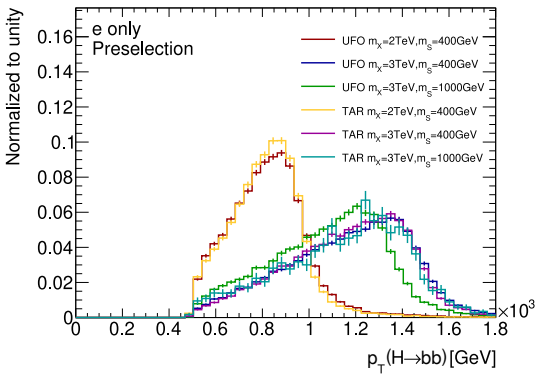


**Figure A.29(d):** $p_T(H \to b\bar{b})$

**Figure A.29.:** Comparison between TAR jets and UFO jets for $SH$ samples for different mass points $m_X$, $m_S$ in the electron channel for various observables.

# A.6. Validation of new $V$+jets Background Samples

## A.6.1. Comparison of $V$+jets Versions

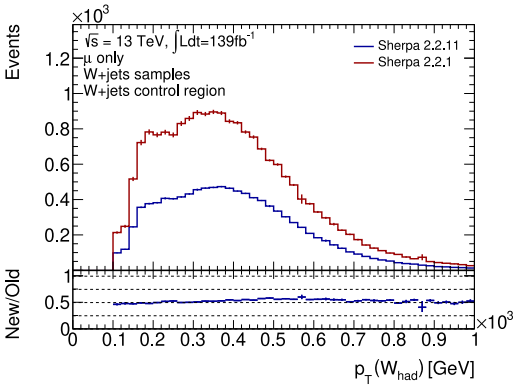


**Figure A.30(a):** $p_T(W_{\text{had}})$

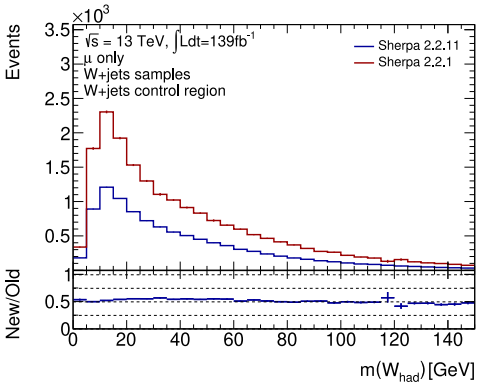


**Figure A.30(b):** $m(W_{\text{had}})$

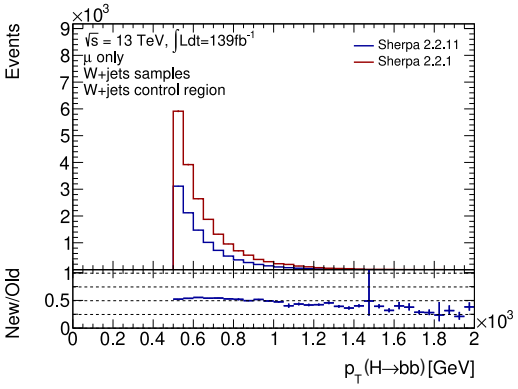


**Figure A.30(c):** $p_T(H \to b\bar{b})$

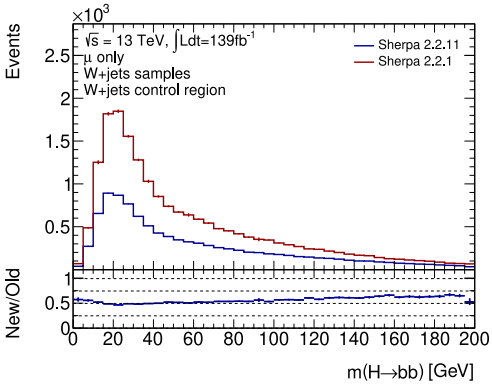


**Figure A.30(d):** $m(H \to b\bar{b})$

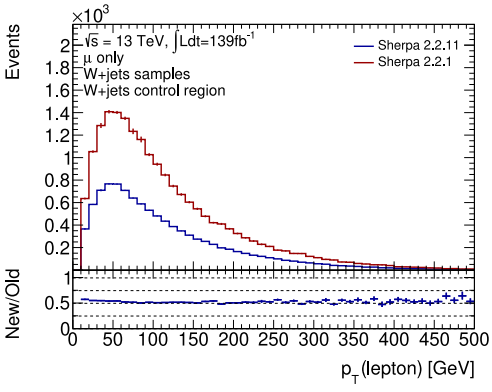


**Figure A.30(e):** $p_T(\ell)$

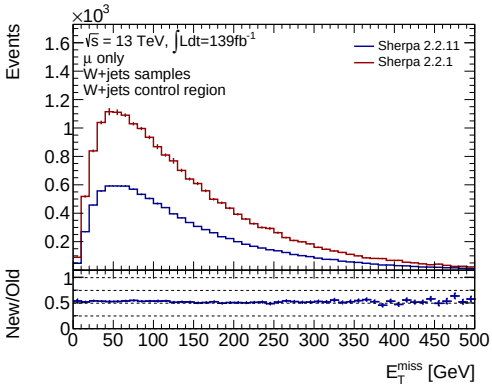


**Figure A.30(f):** $E_T^{\text{miss}}$

**Figure A.30.:** Comparison between $W$+jets samples produced with Sherpa version 2.2.1 and 2.2.11 for various observables in the $W$+jets control region.

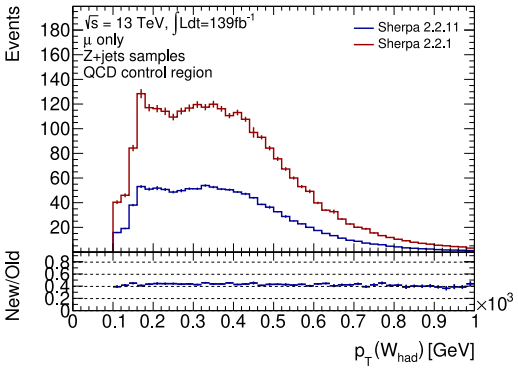


**Figure A.31(a):** $p_T(W_{\text{had}})$

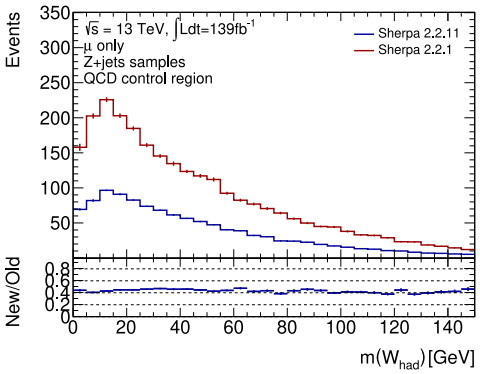


**Figure A.31(b):** $m(W_{\text{had}})$

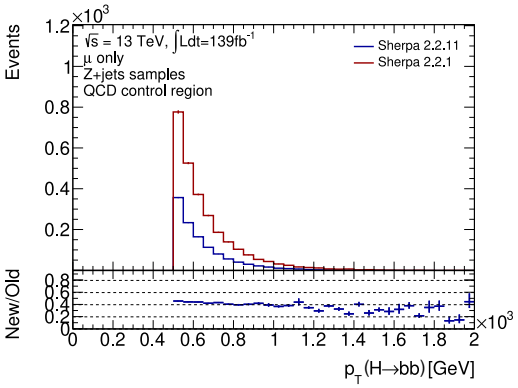


**Figure A.31(c):** $p_T(H \to b\bar{b})$

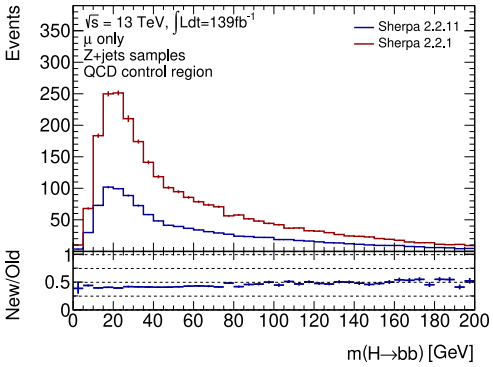


**Figure A.31(d):** $m(H \to b\bar{b})$

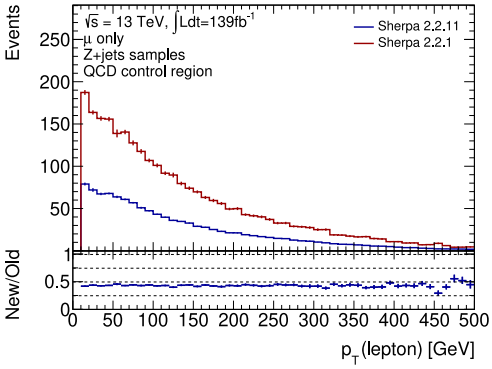


**Figure A.31(e):** $p_T(\ell)$

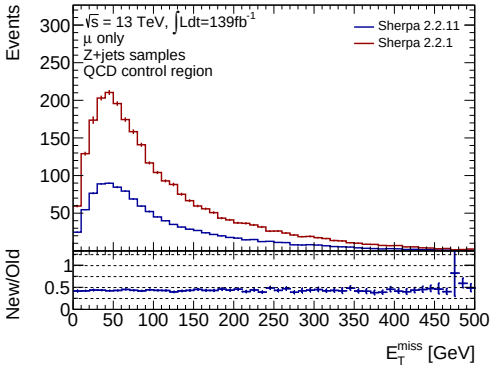


**Figure A.31(f):** $E_T^{\text{miss}}$

**Figure A.31.:** Comparison between $Z$+jets samples produced with Sherpa version 2.2.1 and 2.2.11 for various observables in the QCD control region.

### Normalized Distributions

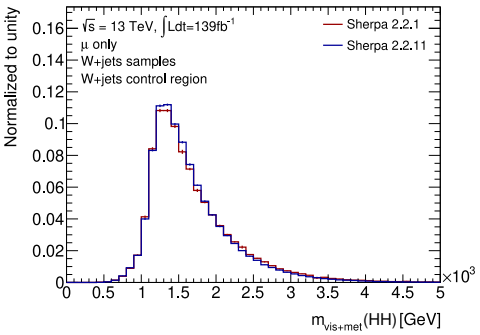


**Figure A.32(a):** $m_{\mathrm{vis+met}}(HH)$

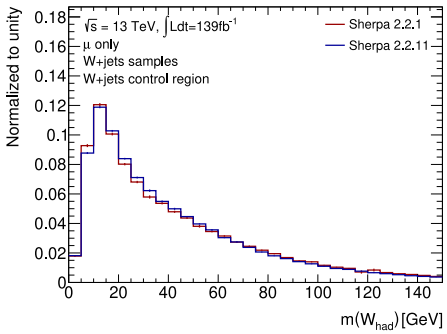


**Figure A.32(b):** $m(W_{\mathrm{had}})$

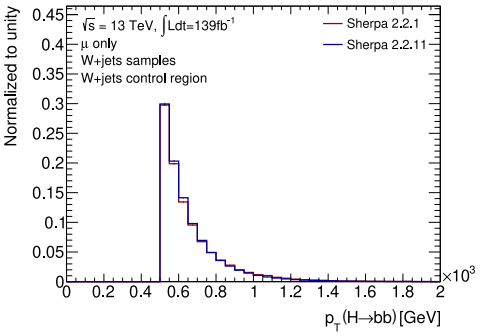


**Figure A.32(c):** $p_T(H \to b\bar{b})$

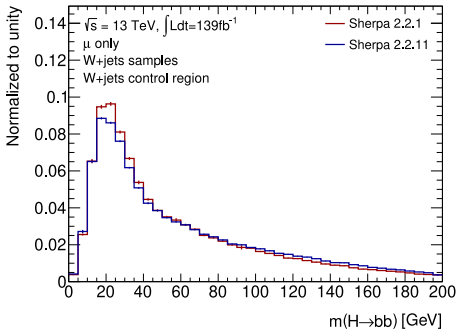


**Figure A.32(d):** $m(H \to b\bar{b})$

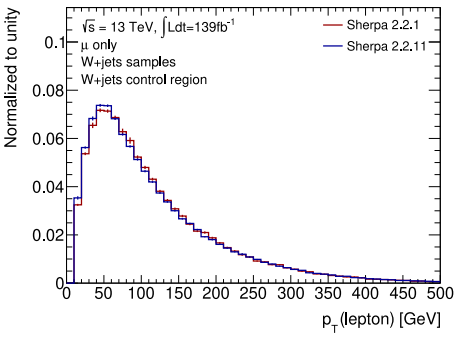


**Figure A.32(e):** $p_T(\ell)$

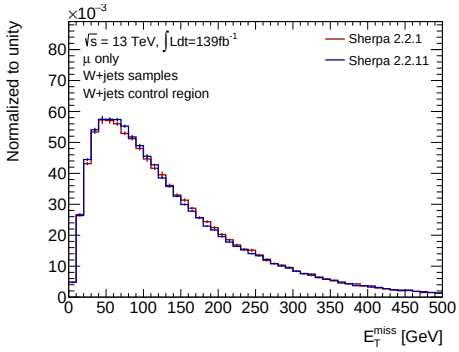


**Figure A.32(f):** $E_T^{\mathrm{miss}}$

**Figure A.32.:** Comparison between normalized $W$+jets samples produced with Sherpa version 2.2.1 and 2.2.11 for various observables in the $W$+jets control region.

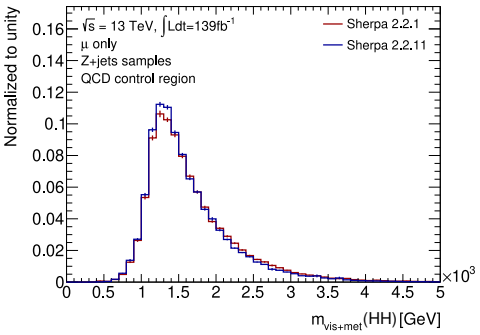


**Figure A.33(a):** $m_{\text{vis+met}}(HH)$

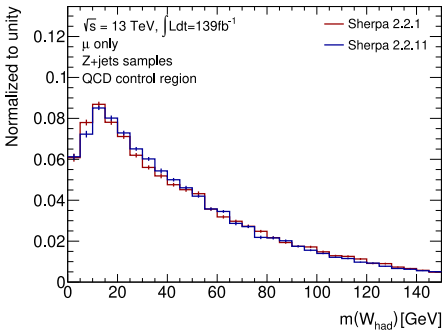


**Figure A.33(b):** $m(W_{\text{had}})$

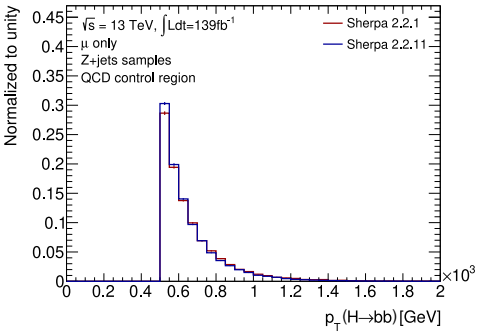


**Figure A.33(c):** $p_T(H \to b\bar{b})$

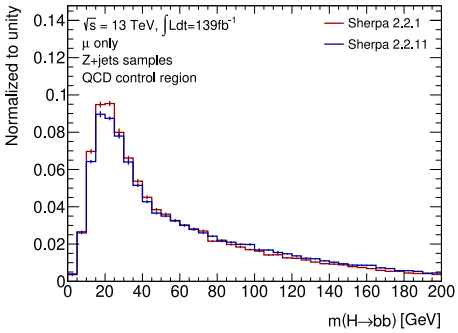


**Figure A.33(d):** $m(H \to b\bar{b})$

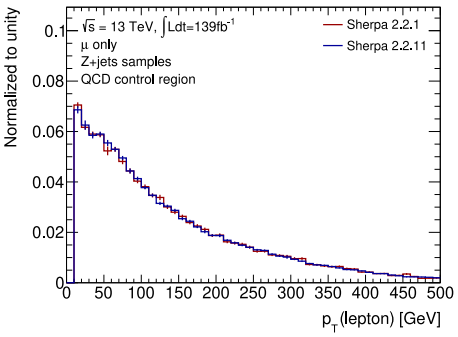


**Figure A.33(e):** $p_T(\ell)$

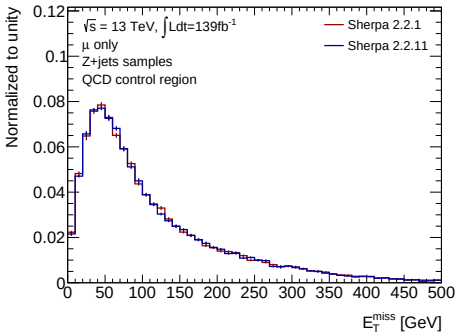


**Figure A.33(f):** $E_T^{\text{miss}}$

**Figure A.33.:** Comparison between normalized $Z$+jets samples produced with Sherpa version 2.2.1 and 2.2.11 for various observables in the QCD control region.

## A.6.2. Data against MC Background

### Sherpa 2.2.1 $V$+jets Samples

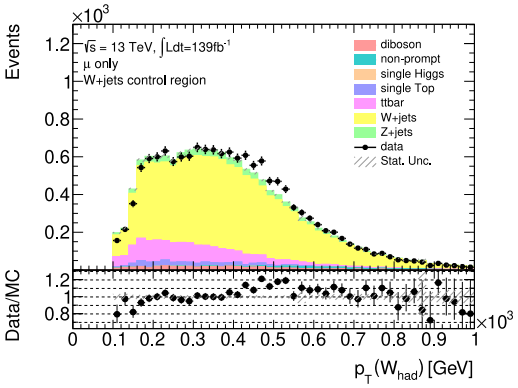


**Figure A.34(a):** $p_T(W_{\text{had}})$

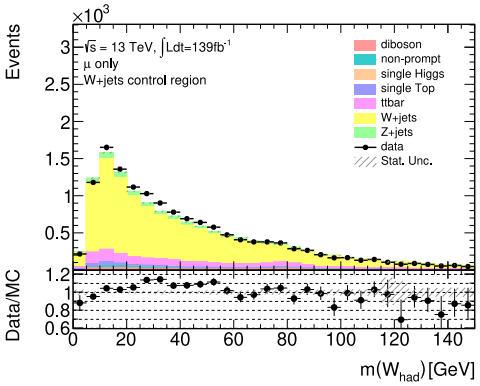


**Figure A.34(b):** $m(W_{\text{had}})$

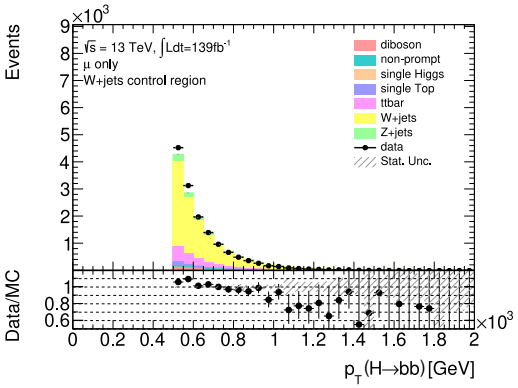


**Figure A.34(c):** $p_T(H \to b\bar{b})$

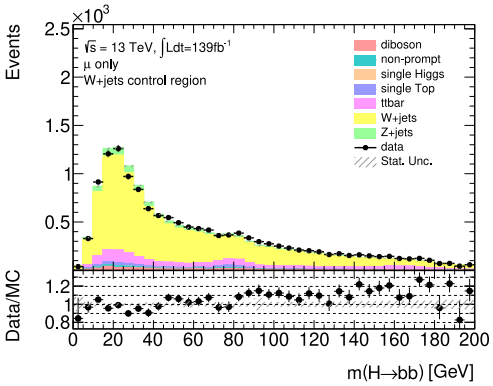


**Figure A.34(d):** $m(H \to b\bar{b})$

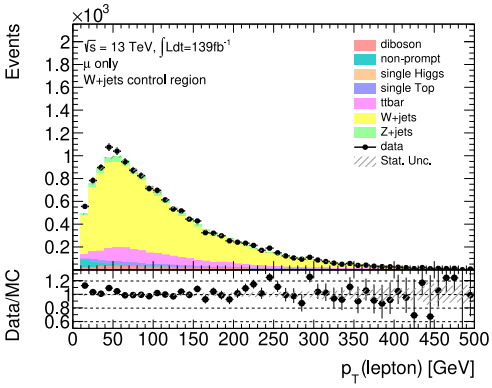


**Figure A.34(e):** $p_T(\ell)$

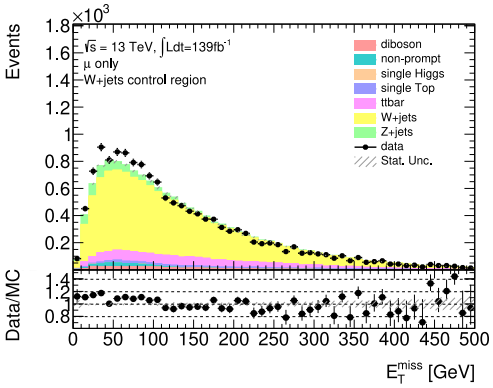


**Figure A.34(f):** $E_T^{\text{miss}}$

**Figure A.34.:** ATLAS data against backgrounds for $V$+jets samples produced with Sherpa 2.2.1 for various observables in the $W$+jets control region.

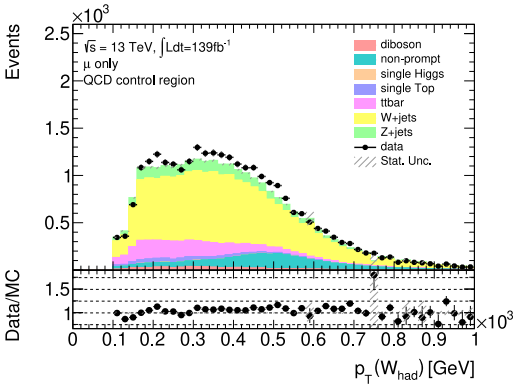


**Figure A.35(a):** $p_T(W_{\text{had}})$

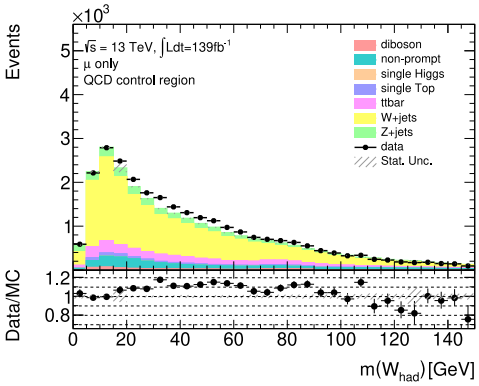


**Figure A.35(b):** $m(W_{\text{had}})$

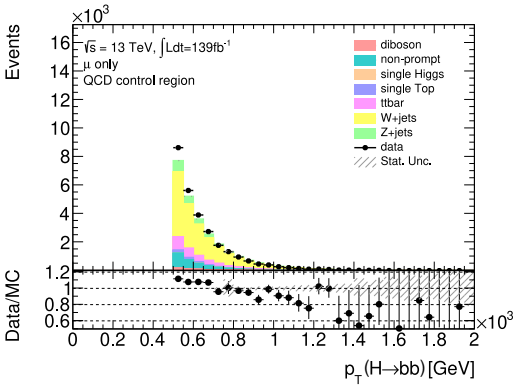


**Figure A.35(c):** $p_T(H \to b\bar{b})$

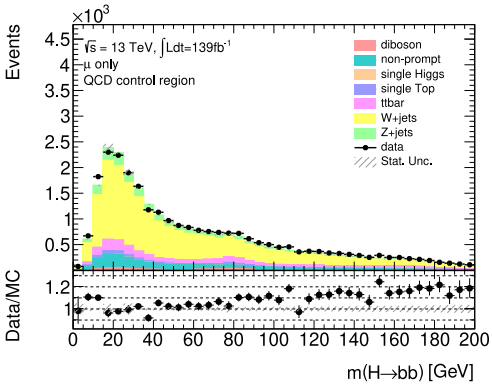


**Figure A.35(d):** $m(H \to b\bar{b})$

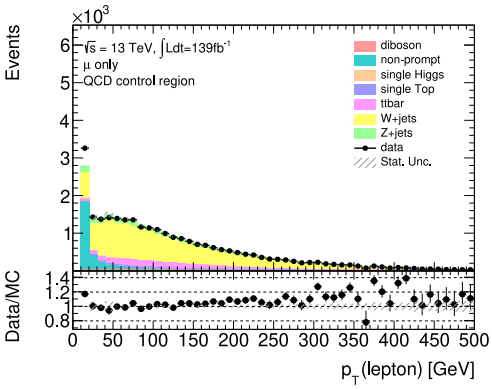


**Figure A.35(e):** $p_T(\ell)$

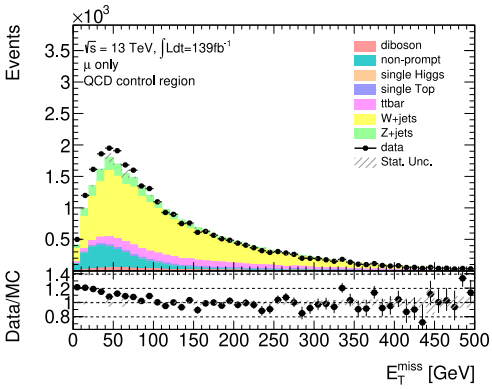


**Figure A.35(f):** $E_T^{\text{miss}}$

**Figure A.35.:** ATLAS data against backgrounds for $V$+jets samples produced with Sherpa 2.2.1 for various observables in the QCD control region.

### Sherpa 2.2.11 $V$+jets Samples

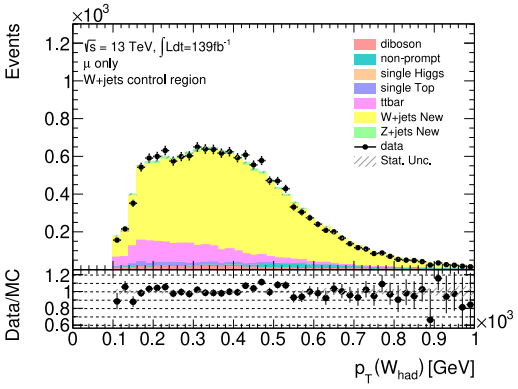


**Figure A.36(a):** $p_T(W_{\text{had}})$

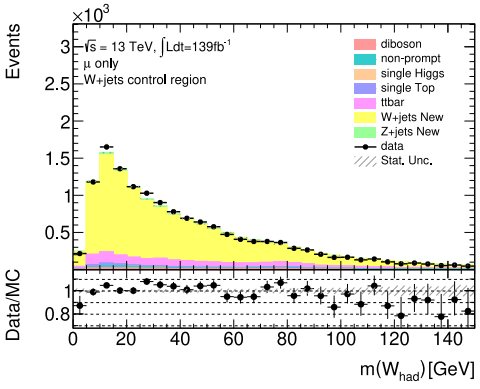


**Figure A.36(b):** $m(W_{\text{had}})$

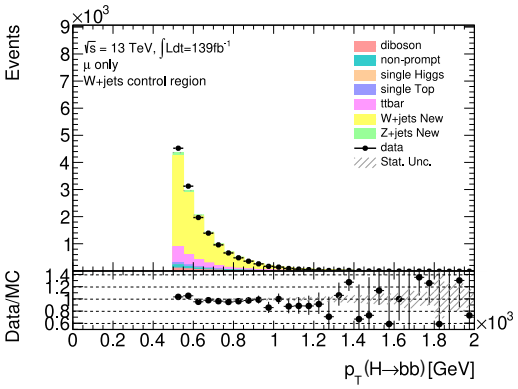


**Figure A.36(c):** $p_T(H \to b\bar{b})$

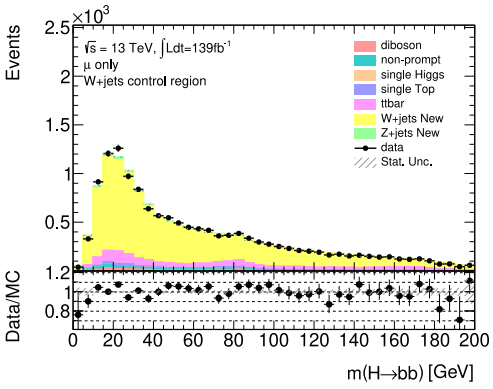


**Figure A.36(d):** $m(H \to b\bar{b})$

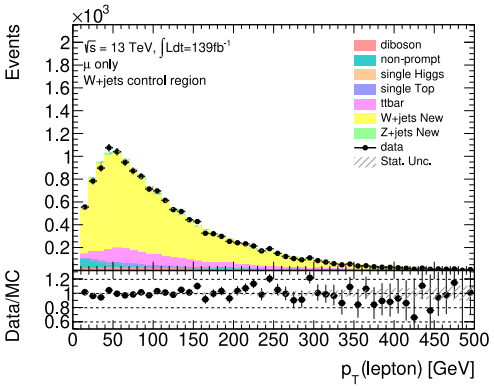


**Figure A.36(e):** $p_T(\ell)$

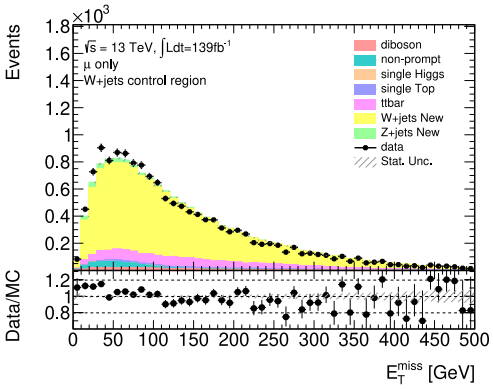


**Figure A.36(f):** $E_T^{\text{miss}}$

**Figure A.36.:** ATLAS data against backgrounds for $V$+jets samples produced with Sherpa 2.2.11 for various observables in the $W$+jets control region.

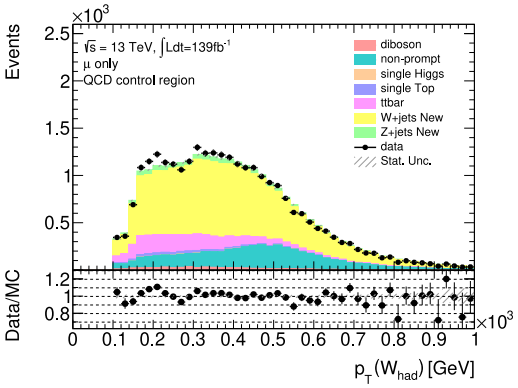


**Figure A.37(a):** $p_T(W_{\text{had}})$

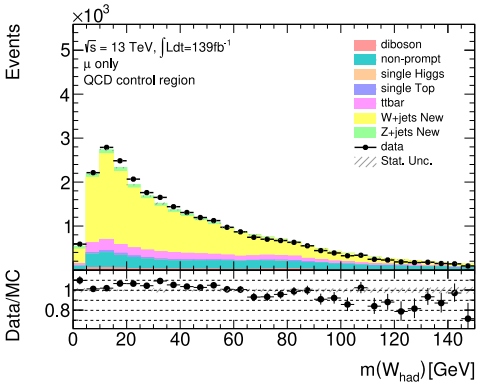


**Figure A.37(b):** $m(W_{\text{had}})$

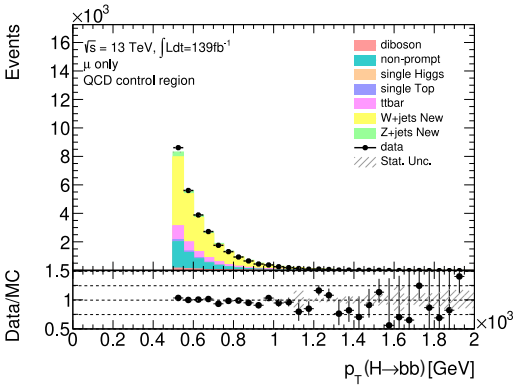


**Figure A.37(c):** $p_T(H \to b\bar{b})$

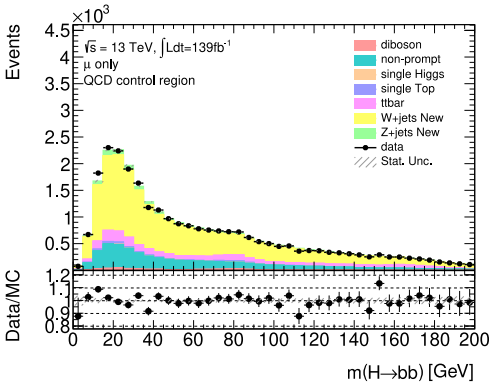


**Figure A.37(d):** $m(H \to b\bar{b})$

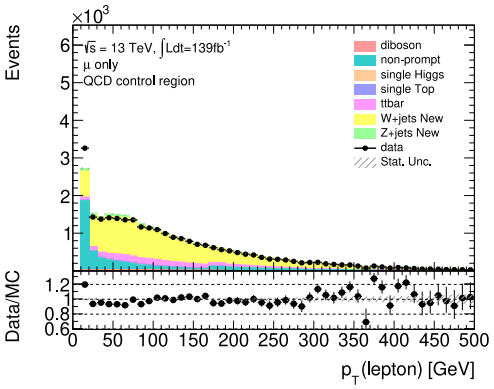


**Figure A.37(e):** $p_T(\ell)$

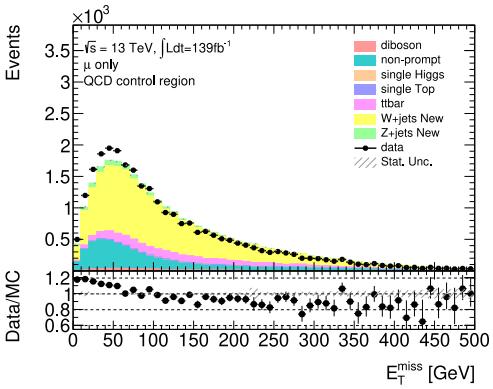


**Figure A.37(f):** $E_T^{\text{miss}}$

**Figure A.37.:** ATLAS data against backgrounds for $V$+jets samples produced with Sherpa 2.2.11 for various observables in the QCD control region.

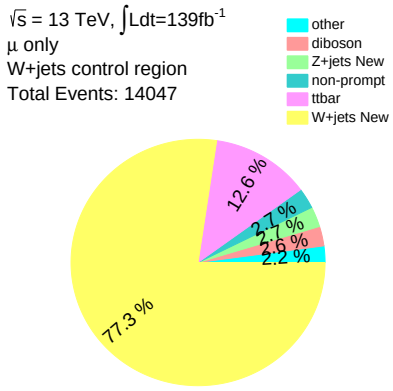


**Figure A.38(a):** $W$+jets CR

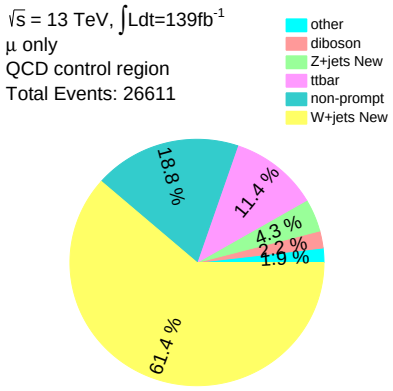


**Figure A.38(b):** QCD CR

**Figure A.38.:** Background composition in the $W$+jets and QCD control region for Sherpa 2.2.11 $V$+jets samples. Backgrounds contributing less than 2.2% are collected in the "other" category.

# A.7. Normalization of $Z$+jets Background

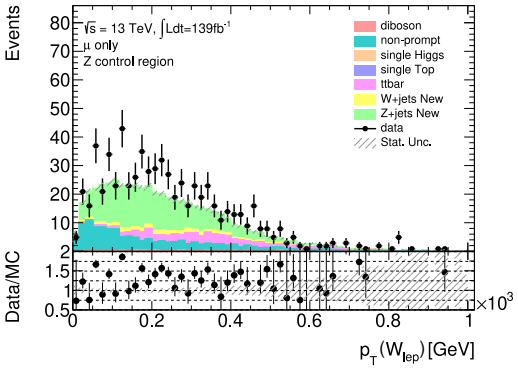


**Figure A.39(a):** $p_T(W_{\text{lep}})$

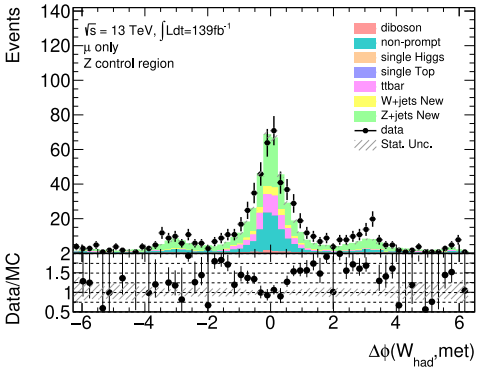


**Figure A.39(b):** $\Delta\phi(W_{\text{had}}, E_T^{\text{miss}})$

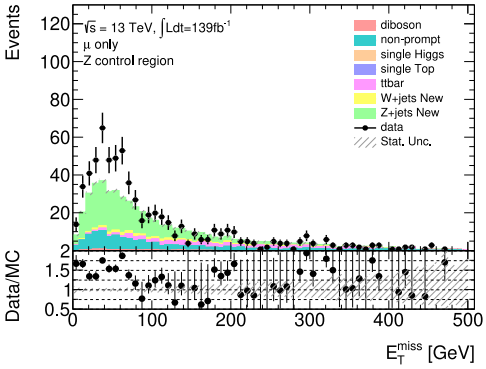


**Figure A.39(c):** $E_T^{\text{miss}}$

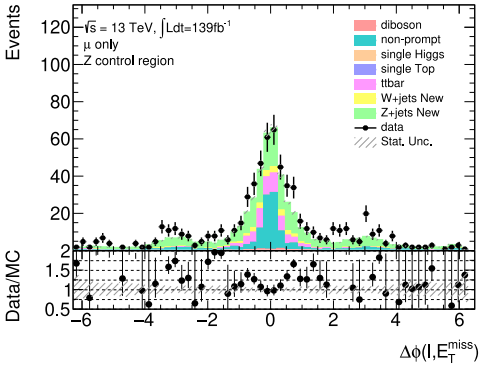


**Figure A.39(d):** $\Delta\phi(\ell, E_T^{\text{miss}})$

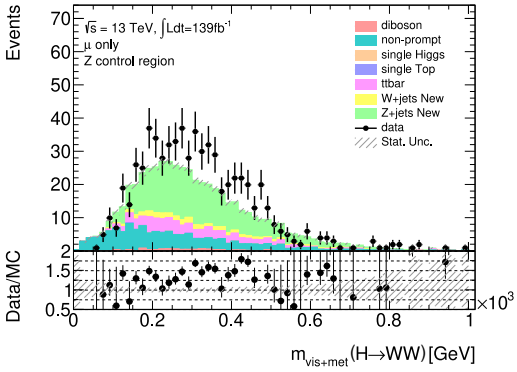


**Figure A.39(e):** $m_{\text{vis+met}}(H \to WW)$

**Figure A.39.:** ATLAS data against backgrounds in the possible $Z$+jets control region for various observables.

**Erklärung** nach §12(9) der Prüfungsordnung für den Master-Studiengang Physics an der Universität Göttingen:

Hiermit erkläre ich, dass ich diese Abschlussarbeit selbständig verfasst habe, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen wurden, als solche kenntlich gemacht habe.

Darüberhinaus erkläre ich, dass diese Abschlussarbeit nicht, auch nicht auszugsweise, im Rahmen einer nichtbestandenen Prüfung an dieser oder einer anderen Hochschule eingereicht wurde.

Göttingen, den 15. Juni 2023

(Leon Stöcker)