GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Fakultät für
Physik

Master's Thesis

Performanzstudien zu erweiterten $H \to b\bar{b}$ Tagging-Methoden für Suchen nach resonanter Higgs-Boson-Paarproduktion im $b\bar{b}WW^*$ Zerfallskanal mit dem ATLAS-Detektor

Performance studies of advanced $H \to b\bar{b}$ tagging methods for searches for resonant Higgs boson pair production in the $b\bar{b}WW^*$ decay channel with the ATLAS detector

prepared by

Joshua Falco Beirer

from Schaffhausen

at the II. Physikalischen Institut

Thesis number: II.Physik-UniGö-MSc-2019/01

Thesis period: 14th May 2019 until 1st October 2019

First referee: Prof. Dr. Stan Lai

Second referee: Prof. Dr. Ariane Frey

Abstract

The Large Hadron Collider allows to conduct precision measurements of a large variety of processes and perform searches for beyond the Standard Model physics at unprecedented energy scales. Searches for Higgs boson pair production are of particular interest as measurements are expected to give access to the Higgs boson self-coupling, which would constitute a direct test of the Higgs mechanism. Moreover, many theories beyond the Standard Model predict heavy resonances, X, which decay into two Higgs bosons. In these topologies, the Higgs bosons are heavily boosted, such that they appear very closely in the detector and the decay products cannot be resolved individually any more. In order to gain sensitivity to resonant and non-resonant Higgs boson pair production, it is important to exploit the dominant Higgs boson decay to two b-quarks.

As a baseline procedure, boosted Higgs bosons decaying to two b-quarks are reconstructed from purely calorimeter-based large-R jets, while the b-quarks are identified as a pair of b-tagged track jets. For Higgs bosons with transverse momenta larger than 1 TeV, the performance of the standard $H \to b\bar{b}$ tagging approach deteriorates quickly as a result of overlapping hadronic decay products. The performance of three alternative subjet tagging techniques, namely variable-R track jets as well as exclusive-k_t and center-of-mass subjets was studied in the context of the resonant $HH \to b\bar{b}WW^*$ decay channel for three resonance masses $m_X = 1$, 2 and 3 TeV. Variable-R track jets were shown to provide the best performance with a significant improvement in efficiency and background rejection.

The finite size of topological clusters in the calorimeter limits the resolution of jet substructure variables, which can provide important aid in the identification of Higgs jets. Angular information from the inner detector can be exploited to overcome these limitations. For the first time, the $H \to b\bar{b}$ tagging performance of track-assisted reclustered jets used to reconstruct the Higgs bosons was studied. The performance was assessed for various different reclustering algorithms and jet sizes. Several substructure variables were studied, from which the energy correlation function ratio C_1 was found to provide the best separation power. Larger jet sizes were found to increase the rejection of top jets, while smaller jet sizes were found to be beneficial for the discrimination against QCD jets.

Contents

Chapter 1

Introduction

So that I may perceive whatever holds
The world together in its inmost folds
— J. W. von Goethe, *Faust I*

Humanity's knowledge about the innermost structure of the universe is the result of many decades of fundamental research. Today, the Standard Model (SM) of particle physics is the most precise and experimentally verified theory of nature. Its triumph began in 1961 when Sheldon Glashow discovered a way to combine the electromagnetic and weak interactions into a single electroweak theory [1]. In 1967, Steven Weinberg and Abdus Salam incorporated the Brout-Englert-Higgs mechanism [2–4] in Glashow's electroweak theory [5, 6] and shaped the SM as we know it today. The electroweak theory was confirmed in 1983 with the experimental discovery of the Z and W bosons [7, 8]. With the discovery of the Higgs boson in 2012 [9, 10], the full range of particles' predicted by the SM was observed for the first time. While this has been a major accomplishment and experimental results are consistent with the predictions of the SM so far, many questions still remain open today. Therefore, searches for extremely rare processes as well as for heavy particles predicted by extended models have become increasingly important with more and more high energy collision data available.

Higgs pair production processes are promising for searches of new heavy resonances such as the heavy CP-even scalar X which is predicted in two-Higgs-doublet models [11] or the spin-2 Kaluza-Klein excitations in the bulk Randall-Sundrum model [12–14]. Given the rare nature of the process, it is important to exploit the large $H \to b\bar{b}$ branching ratio of the single Higgs decay. Nevertheless, searches for heavy resonances in the multi-TeV range are very challenging, as the decay products are highly Lorentz boosted and appear very close to each other in the detector such that they cannot be resolved individually. As a consequence, the Higgs boson in such topologies is usually reconstructed as a large-R jet and the b-quarks are identified as subjets using small-R track jets from the inner detector.

In order to push the reach for new physics to higher mass scales, several alternative subjet reconstruction algorithms have been developed in recent years. This thesis presents performance studies of modern subjet reconstruction methods for boosted $H \to b\bar{b}$ tagging in the $X \to HH \to b\bar{b}WW^*$ channel. Moreover, the impact of using track-assisted reclustered jets [15] for $H \to b\bar{b}$ tagging instead of purely calorimeter-based large-R jets is investigated for the first time.

The thesis is structured as follows: In Chapter 2, a concise overview on the Standard Model and the BEH mechanism is given, followed by a short discussion of extended models and their implications. Furthermore, the process of event simulation as well as the current status of experimental limits on Higgs boson pair production is outlined. In Chapter 3, the structure and functioning of the ATLAS detector at the LHC is briefly described. In Chapter 4, experimental techniques for boosted jet reconstruction and identification relevant for this thesis are outlined. Finally, a truth-level study of the kinematic properties of the $b\bar{b}WW^*$ decay channel and the performance studies of $H \to b\bar{b}$ tagging are presented in Chapters 5 and 6 respectively.

Chapter 2

Theoretical Overview

The ultimate goal of particle physics is to understand the fundamental nature of energy, matter, space, and time by precisely studying the properties and the interactions of all existent particles in the universe. While currently the Standard Model represents the best theoretical framework at our disposal, it fails to provide some of the most long sought answers of our time. In this chapter, a concise theoretical overview of the SM is given and its shortcomings as well as some extension of the SM are briefly discussed. Furthermore, the relevant phenomenological results as well as the process of event simulation is described.

2.1. The Standard Model of Elementary Particles

The unification of quantum mechanics and special relativity led to the development of a basic theoretical framework known as quantum field theory (QFT). In quantum field theories such as the SM, particles emerge as a consequence of excitations of their underlying fields. In the SM, electroweak and strong interactions are described by gauge fields which preserve local gauge invariance under

$$\underbrace{SU(3)}_{strong} \otimes \underbrace{SU(2) \otimes U(1)}_{electroweak}$$

symmetry transformations. The interactions between elementary particles are mediated by quanta of these fields. While the strong interaction is mediated by eight massless gluons, the electroweak mediators resulting from the underlying $SU(2) \otimes U(1)$ symmetry are one massless photon γ and three massive gauge bosons W^+, W^- and Z. The masses of the gauge bosons are generated by a process known as Electroweak Symmetry Breaking (see Section 2.1.3) which predicts the existence of the scalar Higgs boson, H. A summary of the properties of the bosons in the SM is shown in Table 1.

Boson		J^P	Mass [GeV]
Photon	γ	1^-	0
W	$W^\pm$	1^-	80.4
Z	Z	1^-	91.2
Gluon ($\times 8$)	g	1^-	0
Higgs	H	0^+	125.1

Tab. 1: Summary of the properties of the bosons in the Standard Model [16]. J^P denotes the spin, J, and the parity, P, of the particle.

2.1.1. Particle Content

Noether's theorem [17] states that if a system has a continuous symmetry, then there exists an associated conserved quantity. Therefore, each symmetry in the SM can be linked to a conserved quantity. While the strong interaction $SU(3)_C$ can be associated with a quantum number known as colour charge, the electroweak symmetry groups $SU(2)_L$ and $U(1)_Y$ can be associated with the weak isospin I and the weak hypercharge Y respectively. The electromagnetic charge Q is connected to both through the relation $Q = I_3 + \frac{Y}{2}$, where I_3 is the third component of the weak isospin. Spin-$\frac{1}{2}$ fermions in the SM are classified into leptons and quarks. While quarks interact strongly and carry one of the three colour charges (red, green or blue), leptons only take part in electroweak interactions. As a consequence of the chiral structure of $SU(2)_L$, only left-handed fermions couple to the gauge bosons associated to the electroweak interactions. For this reason, fermions are furthermore categorized into left-handed doublets with isospin $I = \frac{1}{2}$ and right-handed singlets with isospin $I = 0$. Each fermion comes in a threefold copy with identical quantum numbers but increasing masses. While the first generation of leptons contains the electron e and its corresponding electron neutrino ν_e, the second and third generation consists of the muon μ and its corresponding muon neutrino ν_μ and the τ-lepton and its corresponding τ-neutrino ν_τ, respectively. Similarly, the first, second and third generation consists of the up-type quarks up u, charm c and top t in addition to the down-type quarks down d, strange s and bottom b respectively. While up-type quarks carry an electric charge of $+\frac{2}{3}\,e$, down-type quarks have $-\frac{1}{3}\,e$. Finally, every fermionic particle in the SM has a corresponding antiparticle with the same properties, except that the electrical charge as well as its handedness is flipped. A list of the fermions and a summary of their properties are shown in Table 2.

	Leptons				Quarks			
Generation	Particle		Charge [e]	Mass [GeV]	Particle		Charge [e]	Mass [GeV]
First	electron neutrino	ν_e	0	$< 2 \times 10^{-9}$	up	u	$+\frac{2}{3}$	0.002
	electron	e^-	-1	0.0005	down	d	$-\frac{1}{3}$	0.005
Second	muon neutrino	ν_μ	0	$< 1.9 \times 10^{-4}$	charm	c	$+\frac{2}{3}$	1.27
	muon	μ^-	-1	0.1057	strange	s	$-\frac{1}{3}$	0.096
Third	tau neutrino	ν_τ	0	$< 1.8 \times 10^{-2}$	top	t	$+\frac{2}{3}$	173.21
	tau	τ^-	-1	1.777	bottom	b	$-\frac{1}{3}$	4.66

Tab. 2: Summary of the properties of the fermions in the Standard Model. The values are taken from Ref. [16].

2.1.2. Gauge Interactions

In the Lagrangian formalism of QFT, a free fermion of mass m is described as a spinor ψ by the Lagrangian density

$$\mathcal{L}_D = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi, \tag{2.1}$$

where γ^μ are known as gamma matrices, which are defined by the anticommutation relation $\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}\,\mathbb{I}_4$. The underlying symmetry principles of the SM require the Lagrangian density to be invariant under local gauge transformation

$$\psi \to e^{i\vec{\epsilon}(x)\cdot\frac{\vec{\tau}}{2}}\psi, \tag{2.2}$$

where $\vec{\epsilon}(x)$ is a space-time dependent phase factor and $\vec{\tau}$ the group generators associated to the corresponding symmetry. While Equation 2.1 does not retain this symmetry principle, local gauge invariance can be established by introducing a covariant derivative

$$\partial_\mu \to \mathcal{D}_\mu = \partial_\mu + ig\frac{\vec{\tau}}{2}\vec{A}_\mu. \tag{2.3}$$

The local gauge invariance is preserved by introducing additional vector fields $\vec{A}_\mu$ and an arbitrary parameter g that will correspond to the interaction strength associated to the field. It can be shown that for a general transformation U, the components of the fields $\vec{A_\mu}$ are required to satisfy

$$U\tau_i A^i_\mu U^{-1} = \tau_i A'^i_\mu + \frac{2i}{g}\left(\partial_\mu U\right)U^{-1}. \tag{2.4}$$

In general, requiring a Lagrangian density to be invariant under some local gauge transformation will give rise to associated vector fields, which are known as gauge fields. In the SM, three of these symmetries are incorporated. While in the electroweak sector one gauge field B_μ is required by the $U(1)_Y$ symmetry, three gauge fields $W^i_\mu\,(i = 1, 2, 3)$ are associated to the $SU(2)_L$ group. For the strong interaction, the $SU(3)_C$ symmetry generates a total of eight gauge fields $G^a_\mu\,(a = 1,...,8)$, which correspond to the 8 color charge combinations of the gluons. While $U(1)_Y$ is an Abelian group, $SU(2)_L$ and $SU(3)_C$ are non-Abelian and therefore give rise to self-interactions of the associated gauge bosons. The covariant derivative introducing the couplings between fermionic and bosonic fields which guarantees the local gauge invariance of the SM is given by

$$\mathcal{D}_\mu = \partial_\mu + \underbrace{i\frac{g_1}{2}YB_\mu}_{U(1)_Y} + \underbrace{i\frac{g_2}{2}\tau_i W^i_\mu}_{SU(2)_L} + \underbrace{i\frac{g_3}{2}\lambda_a G^a_\mu}_{SU(3)_C}, \tag{2.5}$$

whereby Y and the matrices τ_i and λ_a are the generators corresponding to the $U(1)$, $SU(2)$ and $SU(3)$ groups, respectively. While the strong interaction can be associated directly to the gluons, a similar association of the electromagnetic and weak interactions with the gauge bosons is not possible. Instead, the experimentally accessible fields exhibit themselves as linear combinations of

the gauge fields:

$$\begin{aligned} A_\mu &= \frac{g_2 B_\mu + g_1 W_\mu^3}{\sqrt{g_1^2 + g_2^2}} & W_\mu^+ &= \frac{W_\mu^1 - iW_\mu^2}{\sqrt{2}} \\ Z_\mu &= \frac{g_2 W_\mu^3 - g_1 B_\mu}{\sqrt{g_1^2 + g_2^2}} & W_\mu^- &= \frac{W_\mu^1 + iW_\mu^2}{\sqrt{2}} \end{aligned} \tag{2.6}$$

Here, A_μ corresponds to the physical photon field, Z_μ to the Z boson field and $W_\mu^\pm$ to the $W^\pm$ boson fields. The Lagrangian density fully describing all fermions and their interactions can then be given by summing the Dirac terms in Equation 2.1 for all fermions and replacing the derivatives with the covariant derivatives 2.5 such that the resulting structure becomes

$$\boldsymbol{\mathcal{L}} = -\frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu} - \frac{1}{4} W_{\mu\nu}^i W_i^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} + \bar{\psi}_L^f i \mathcal{D}_\mu \gamma^u \psi_L^f + \bar{\psi}_R^f i \mathcal{D}_\mu \gamma^u \psi_R^f, \tag{2.7}$$

where $\psi_{L\,(R)}^f$ denotes all left (right) handed fermion isospin doublets (singlets). While the first three terms describe the kinematics of the gauge fields, the last two terms describe the kinematics of the fermions as well as the interactions between the fermionic and bosonic gauge fields. By construction, Equation 2.7 is invariant under local $U(1)_Y \otimes SU(2)_L \otimes SU(3)_C$ gauge transformations. However, the addition of fermionic mass terms

$$m\bar{\psi}\psi = m\left(\bar{\psi}_R \psi_L + \bar{\psi_L} \psi_R\right), \tag{2.8}$$

spoils the necessary $SU(2)_L$ gauge invariance, since right-handed isospin singlets transform differently than left-handed doublets. Furthermore, bosonic mass terms of the form $-m^2 A_\mu A^\mu$ must also be excluded from the theory in order to preserve the $SU(2)_L$ gauge invariance. With the exclusion of the respective terms, all fermions as well as the gauge bosons are predicted to be massless. As masses of the fermions, $W^\pm$ and Z bosons have been experimentally determined to be non-zero [7, 8], an alternative mechanism is required to explain the generation of masses in the SM. This is known as Electroweak Symmetry Breaking (EWSB) and is described in Section 2.1.3.

2.1.3. Electroweak Symmetry Breaking

In order to generate the masses of the gauge bosons in the electroweak sector, the $SU(2)$ symmetry of the vacuum needs to be broken while retaining the invariant nature of the Lagrangian density. In the 1960's, this spontanous symmetry breaking was incorporated into the SM and is referred to as the Brout-Englert-Higgs (BEH) mechanism [2–4]. The BEH mechanism introduces a Higgs field Φ as a complex scalar doublet

$$\Phi(\mathbf{x}) = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \phi_1(\mathbf{x}) + i\phi_2(\mathbf{x}) \\ \phi_3(\mathbf{x}) + i\phi_4(\mathbf{x}) \end{pmatrix}. \tag{2.9}$$

The resulting invariant Lagrangian density corresponding to the Higgs field in its most general form can be written as

$$\boldsymbol{\mathcal{L}}_H = \left(\mathcal{D}^\mu \Phi\right)^\dagger \mathcal{D}_\mu \Phi - V(\Phi), \tag{2.10}$$

where

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda \left(\Phi^\dagger \Phi\right)^2 \tag{2.11}$$

is known as the Higgs potential. Here, λ is a positive arbitrary parameter while the sign of the parameter μ^2 defines the form of the Higgs potential. While for $\mu^2 > 0$, the potential has one global minimum at $\Phi = 0$, the potential takes an infinite set of non-zero degenerate minima at

$$\Phi^\dagger \Phi = \frac{-\mu^2}{2\lambda} \equiv \frac{v^2}{2}, \tag{2.12}$$

for $\mu^2 < 0$. The parameter v is known as the vacuum expectation value (VEV) of the field Φ. In the two-dimensional case of a complex scalar field $\phi = \phi_{\mathbb{Re}} + i\phi_{\mathbb{Im}}$, the non-zero VEVs lie on a circle. The form of such a potential for $\mu^2 > 0$ as well as for $\mu^2 < 0$ is visualized in Figure 1a and 1b respectively.

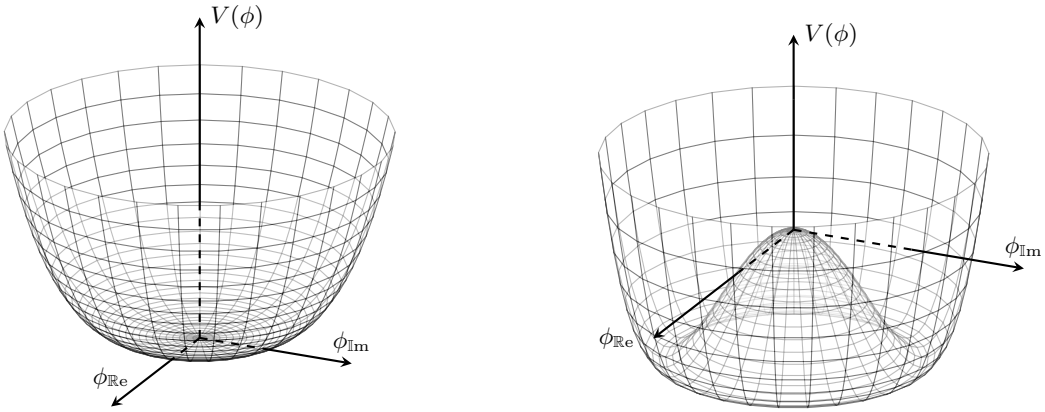


Fig. 1: The Higgs potential $V(\phi)$ for $\mu^2 > 0$ (a) and $\mu^2 < 0$ (b) for a complex scalar field ϕ.

While in the case of $\mu^2 > 0$ the theory is said to be unbroken, one of the infinite number of VEVs is chosen for $\mu^2 < 0$ such that the symmetry of the Lagrangian is spontaneously broken. When writing the field in unitary gauge and expanding the Higgs doublet around the electroweak symmetry breaking VEV

$$\Phi(\mathbf{x}) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(\mathbf{x}) \end{pmatrix}, \tag{2.13}$$

a new massive scalar boson H, which is known as the SM Higgs boson, as well as quadratic mass terms of the gauge bosons naturally emerge in the Lagrangian. The predicted masses are given by

$$\begin{aligned} m_H &= \sqrt{2\lambda v^2} & m_W &= \frac{1}{2} g_2 v \\ m_Z &= \frac{1}{2} v \sqrt{g_1^2 + g_2^2} & m_A &= 0 \end{aligned} \tag{2.14}$$

Furthermore, the expansion yields cubic and quartic Higgs boson self-coupling terms

$$\mathcal{L}_H \supset v\lambda H^3 + \frac{1}{4}\lambda H^4, \tag{2.15}$$

which predict trilinear and quartic Higgs boson self-couplings. The resulting interaction vertices and their respective coupling strengths are shown in Figure 2.

Fig. 2: Trilinear and quartic self-couplings of the SM Higgs boson.

While the spontanous symmetry breaking of the $U(1)_Y \otimes SU(2)_L$ gauge group generates the masses of the W and Z bosons, fermionic masses do not naturally emerge from the BEH mechanism. Nonetheless, the BEH mechanism can be expanded to also explain the fermion masses. This can be achieved by the addition of Yukawa coupling terms

$$\mathcal{L}_{\text{Yukawa}} = -g_f \left(\bar{\psi}_L^f \Phi \psi_R^f + \bar{\psi}_R^f \bar{\Phi} \psi_L^f \right). \tag{2.16}$$

The fermion masses arising through the coupling of the fermions to the Higgs field are given by

$$m_f = \frac{v g_f}{\sqrt{2}}. \tag{2.17}$$

2.2. Physics Beyond the Standard Model

While the SM offers a precise description of nature and elegantly describes the origin of mass, many unanswered questions remain. Physics beyond the SM (BSM) is expected to exist due to various reasons which are outlined in the following. Firstly, the SM excludes gravity in its theoretical framework, such that it cannot be considered a theory of everything (ToE). While there have been several attempts to quantize the gravitational interaction in the framework of QFT, resulting quantum theories of gravity are perturbatively non-renormalizable and therefore cannot be used to make meaningful physical predictions [18–21].

Furthermore, astrophysical observations show the existence of a significant amount of neutral, non-baryonic matter [22, 23] which is known as dark matter (DM). To explain these observations, many lines of reasoning suggest that DM consists of some new weakly interacting massive particles (WIMPs) [24], for which the SM cannot provide a suitable candidate. Also, the SM predicts a slight asymmetry between matter and antimatter due to CP violation, but there is currently no mechanism that can explain the overwhelming excess of particles to antiparticles in the universe [25].

From a theoretical perspective, some features of the SM might be considered unaesthetic. For instance, the bare mass of the Higgs boson is heavily fine-tuned with massive quantum loop corrections on the order of the Planck mass in order to explain the magnitude of the observable Higgs boson mass. This large discrepancy between the observable and the natural value of the Higgs boson mass, which is known as hierarchy problem [26], is considered unnatural and may indicate the existence of a more fundamental theory. Finally, the unification of the electromagnetic and weak force lead to the hope that all three forces in the SM unify at some energy scale. However, for the SM to be self-consistent, a convergence of all couplings in a single value is not possible [27–31].

2.2.1. Two-Higgs-Doublet Models

As a result of the problems mentioned in Section 2.2, many models try to extend the SM. As experimental observations are in harmony with the predictions of the SM, BSM extensions are required to reduce to the SM in a low-energy approximation. One of the simplest well-motivated extensions to the SM are known as two-Higgs-doublet models (2HDMs) [11]. While 2HDMs retain the $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ structure of the SM, the Higgs sector is extended by adding an additional $SU(2)_L$ doublet Φ_2, leading to two complex fields

$$\Phi_1 = \begin{pmatrix} \phi_1^+ \\ \phi_1^0 \end{pmatrix} \quad \text{and} \quad \Phi_2 = \begin{pmatrix} \phi_2^+ \\ \phi_2^0 \end{pmatrix}. \tag{2.18}$$

2HDMs offer very large parameter spaces such that phenomenological studies require additional simplifying assumptions. While 2HDMs allow for sources of CP violation in the Higgs sector in general, CP conservation is usually assumed. If one furthermore assumes that CP is not spontaneously broken and that discrete symmetries eliminate all quartic terms odd in either of the doublets from the potential, the most general scalar potential yields

$$\begin{aligned} V = {} & m_{11}^2 \Phi_1^\dagger \Phi_1 + m_{22}^2 \Phi_2^\dagger \Phi_2 - m_{12}^2 \left(\Phi_1^\dagger \Phi_2 + \Phi_2^\dagger \Phi_1 \right) + \frac{\lambda_1}{2} \left(\Phi_1^\dagger \Phi_1 \right)^2 + \frac{\lambda_2}{2} \left(\Phi_2^\dagger \Phi_2 \right)^2 \\ & + \lambda_3 \Phi_1^\dagger \Phi_1 \Phi_2^\dagger \Phi_2 + \lambda_4 \Phi_1^\dagger \Phi_2 \Phi_2^\dagger \Phi_1 + \frac{\lambda_5}{2} \left[\left(\Phi_1^\dagger \Phi_2 \right)^2 + \left(\Phi_2^\dagger \Phi_1 \right)^2 \right], \end{aligned} \tag{2.19}$$

for two complex doublets Φ_1 and Φ_2 and hypercharge $Y = +1$ [32]. The aforementioned assumptions allow to reduce the parameter space from 14 free variables to a smaller set of real parameters λ_i and m_{ij}. Furthermore it is useful to introduce the quantities $v^2 = v_1^2 + v_2^2$ and $\tan\beta = v_2/v_1$,

where v_1 and v_2 are the VEVs of the fields ϕ_1 and ϕ_2 respectively, as the phenomenology of the 2HDMs is heavily dependent on these parameters. With the introduction of two complex scalar doublets there are eight fields from with which three generate the masses of the Z and $W^\pm$ gauge bosons. After EWSB the remaining fields correspond to five physical Higgs bosons:

$$\begin{aligned} H &= -\sqrt{2}\left(\mathbb{R}\mathrm{e}\{\Phi_1^0\} - v_1\right)\sin\alpha + \sqrt{2}\left(\mathbb{R}\mathrm{e}\{\Phi_2^0\} - v_2\right)\cos\alpha \\ X &= \sqrt{2}\left(\mathbb{R}\mathrm{e}\{\Phi_1^0\} - v_1\right)\cos\alpha + \sqrt{2}\left(\mathbb{R}\mathrm{e}\{\Phi_2^0\} - v_2\right)\sin\alpha \\ A &= -\sqrt{2}\,\mathbb{I}\mathrm{m}\{\Phi_1^0\}\sin\beta + \sqrt{2}\,\mathbb{I}\mathrm{m}\{\Phi_2^0\}\cos\beta \\ H^\pm &= -\Phi_1^\pm\sin\beta + \Phi_2^\pm\cos\beta, \end{aligned} \tag{2.20}$$

where α is a mixing angle dependent on the various couplings of the potential. While H can be associated to the observed SM-like Higgs boson, X is a heavier neutral copy of H, A is a CP odd pseudoscalar and $H^\pm$ denotes two charged Higgs bosons. In comparison to the single free parameter m_H in the Higgs sector of the SM, this model has a total of six free parameters: the four masses of the Higgs bosons as well as the ratio of the VEVs and the Higgs mixing angle α.

In general, 2HDMs allow the existence of tree-level flavour-changing neutral currents (FCNCs). However, strict experimental limits on FCNCs as well as on CP violating transitions [33] disfavour the existence of these processes at leading order. Assuming natural flavour conservation, only four possible types of 2HDMs remain (see Table 3).

Model	u_R	d_R	ℓ_R
Type I	Φ_2	Φ_2	Φ_2
Type II	Φ_2	Φ_1	Φ_1
Lepton-specific	Φ_2	Φ_2	Φ_1
Flipped	Φ_2	Φ_1	Φ_2

Tab. 3: Fermion couplings to the Higgs doublets for 2HDM models which lead to natural flavour conservation. u denotes up-type quarks, d down-type quarks and ℓ charged leptons. The couplings are taken from Ref. [32].

The remaining models are categorised according to the type of fermion couplings they exhibit. Similar to the SM, the quark and leptons solely couple to one of the Higgs doublets in Type I 2HDMs, Φ_2 by convention. In Type II models, Φ_1 is assumed to couple to leptons and down-type quarks, while Φ_2 couples to up-type quarks and neutrinos. Other models include the lepton-specific 2HDM where all quarks couple to Φ_2, while leptons couple to Φ_1; and the flipped model, in which only the down-type quarks couple to Φ_2 and the remaining particles to Φ_1. In Table 4, the strength of the respective Yukawa couplings of up-type quarks, down-type quarks, and charged leptons to the neutral Higgs bosons H, X and A in the four different models are shown.

While two-Higgs-doublet models are considered to be promising due to various reasons, Supersymmetry (SUSY) [34–38] is the best-known motivation for 2HDMs. SUSY, at the very least, doubles the particle content of the SM by introducing an additional bosonic state for every fermion and vice versa. These additional particles are known as superpartners. As a result, it addresses most of the problems described in Section 2.2.

Fermion	Boson	Type I	Type II	Lepton-specific	Flipped
u	H	$\cos\alpha/\sin\beta$	$\cos\alpha/\sin\beta$	$\cos\alpha/\sin\beta$	$\cos\alpha/\sin\beta$
	X	$\sin\alpha/\sin\beta$	$\sin\alpha/\sin\beta$	$\sin\alpha/\sin\beta$	$\sin\alpha/\sin\beta$
	A	$\cot\beta$	$\cot\beta$	$\cot\beta$	$\cot\beta$
d	H	$\cos\alpha/\sin\beta$	$-\sin\alpha/\cos\beta$	$\cos\alpha/\sin\beta$	$-\sin\alpha/\cos\beta$
	X	$\sin\alpha/\sin\beta$	$\cos\alpha/\cos\beta$	$\sin\alpha/\sin\beta$	$\cos\alpha/\cos\beta$
	A	$-\cot\beta$	$\tan\beta$	$-\cot\beta$	$\tan\beta$
ℓ	H	$\cos\alpha/\sin\beta$	$-\sin\alpha/\cos\beta$	$-\sin\alpha/\cos\beta$	$\cos\alpha/\sin\beta$
	X	$\sin\alpha/\sin\beta$	$\cos\alpha/\cos\beta$	$\cos\alpha/\cos\beta$	$\sin\alpha/\sin\beta$
	A	$-\cot\beta$	$\tan\beta$	$\tan\beta$	$-\cot\beta$

Tab. 4: Yukawa couplings of up-type quarks u, down-type quarks d, and charged leptons ℓ, to the neutral Higgs bosons H, X and A in the four different models. The couplings are taken from Ref. [32].

For instance, it is able to solve the hierarchy problem by ensuring the exact cancellation of quadratic divergences with the introduction of aforementioned superpartners [39, 40]. Moreover, a local theory of Supersymmetry (SUGRA) [41] promises the integration of gravitational interactions and provides a natural candidate for a possible spin-2 graviton. SUSY also allows the inclusion of additional sources of CP violation as well as it may allow the three gauge couplings to unify at a scale near the Planck scale. As SUSY's parameter space is extremely large, Minimal Supersymmetric Models (MSSMs) [42], which only contain the minimum number of particles that are required for a supersymmetric theory, are often studied. In the MSSM, the quantum number known as R-parity $R = (-1)^{2S+3(B-L)}$, where B, L and S are the baryon and lepton number as well as the spin of particle respectively, is conserved. As for all particles in the SM, $R = +1$, while for their superpartners $R = -1$, conservation of R-parity implies the stability of the lightest supersymmetric particle (LSP). A neutral LSP would provide a suitable WIMP candidate for dark matter. As a single Higgs doublet fails to simultaneously generate the masses to the up- and down-type quarks, the introduction of a second doublet in the MSSM is always necessary. In additon to 2HDMs, some more exotic scenarios dealing with warped extra dimensions such as the Randall-Sundrum model [12–14], which predicts the existence of a Kaluza-Klein (KK) spin-2 graviton, are also considered as possible extensions.

2.3. Higgs Boson Production and Decay

In proton-proton (pp) collisions at high energies such as at the Lhc, the Higgs boson can be produced through various processes. The dominating process at the Lhc is gluon-gluon fusion (ggF) (see Figure 3a). In the ggF process, two gluons fuse predominantly via a top quark loop to give a single Higgs boson. The vector boson fusion (VBF) process is the second most dominant production mode at the Lhc. Two vector bosons radiated from incoming quarks fuse and result in a single Higgs boson. As the incoming quarks are scattered-off collinearly to the beamline, VBF

results in a unique signature in the detector. Furthermore, the associated production of a Higgs boson with a vector boson (VH) is also possible (see Figure 3c) and is known as Higgsstrahlung. Two quarks annihilate to produce a virtual vector boson which then decays to a Higgs boson and an associated W or Z boson. Finally, the associated production of a Higgs boson with a pair of top quarks (see Figure 3d) is also possible and provides a way to directly extract the top quark Yukawa coupling. In this case, two incoming gluons split into a $t\bar{t}$ pair each from which one $t\bar{t}$ pair fuses to give a single Higgs boson, while the other top quark pair exhibits as a typical $t\bar{t}$ signature in the detector. The Feynman diagrams for these four main single Higgs bosons production mechanisms in the SM and their respective cross sections at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ are given in Figure 3.

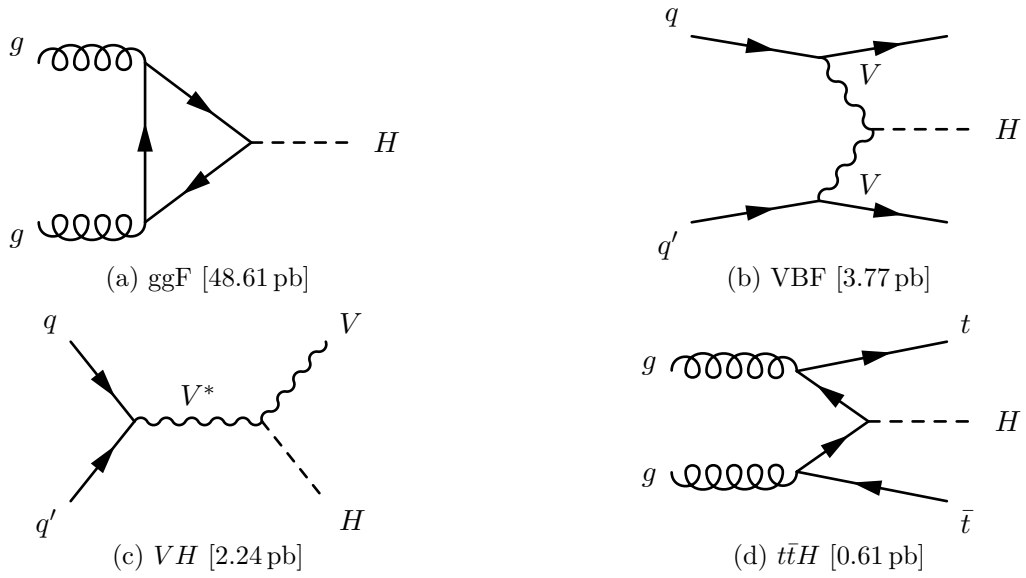


Fig. 3: The four main single Higgs boson production mechanisms in the SM. The values in brackets indicate the production cross section at $\sqrt{s} = 13\,\text{TeV}$ and are taken from Ref. [43].

As an unstable particle, the Higgs boson decays almost instantly after its production. While it can potentially decay to all massive particles i in the SM with $m_H > 2m_i$, the Higgs boson preferably couples to heavier particles. The largest branching ratio by far is for the decay into two b-quarks with 58%, followed by the decay into an on-shell and a virtual W-boson with a branching ratio of 22%. Moreover, the Higgs boson can also decay into massless particles via charged particle loops, which are dominated by virtual W bosons and top quarks. In Table 5, a list of branching ratios for the most common Higgs boson decays, assuming a Higgs boson mass $m_H = 125.09\,\text{GeV}$, is provided. Furthermore, the predicted production cross sections at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ and branching ratios as a function of the Higgs boson mass for a SM Higgs boson are shown in Figure 4.

In 2012, a new scalar particle was observed by the ATLAS and CMS collaborations with a significance of 5.9σ and 5.0σ, respectively. The mass of the Higgs boson was measured to $126.0 \pm 0.4\,(\text{stat.}) \pm 0.4\,(\text{syst.})$ GeV [9] by the ATLAS collaboration and $125.3 \pm 0.4\,(\text{stat.}) \pm 0.5\,(\text{syst.})$ GeV [10] by

Decay Mode						
$b\bar{b}$	WW^*	gg	$\tau^+\tau^-$	$c\bar{c}$	ZZ^*	$\gamma\gamma$
58.09	21.52	8.18	6.26	2.88	2.64	0.23

Tab. 5: Branching ratios $\mathcal{BR}(H \to xx')$ for the most common Higgs boson decays in %, assuming a Higgs boson mass $m_H = 125.09\,\text{GeV}$. The values are taken from Ref. [44].

the Cms collaboration. The current best value from a combined measurement using Run I data from the Atlas and Cms experiments is $m_H = 125.09 \pm 0.21\,(\text{stat.}) \pm 0.11\,(\text{syst.})$ GeV [45]. Even though the $H \to q\bar{q}$ channels provide large branching ratios, they were not of great importance in Higgs searches, as the signatures are subject to large QCD background. The two most important processes which led to the Higgs boson discovery were the $H \to ZZ^*(\to 4\ell)$ and $H \to \gamma\gamma$ decays. Even though they offer relatively low branching ratios (see Table 5), the background processes are well understood and the decay products provide a clear signature in the detector.

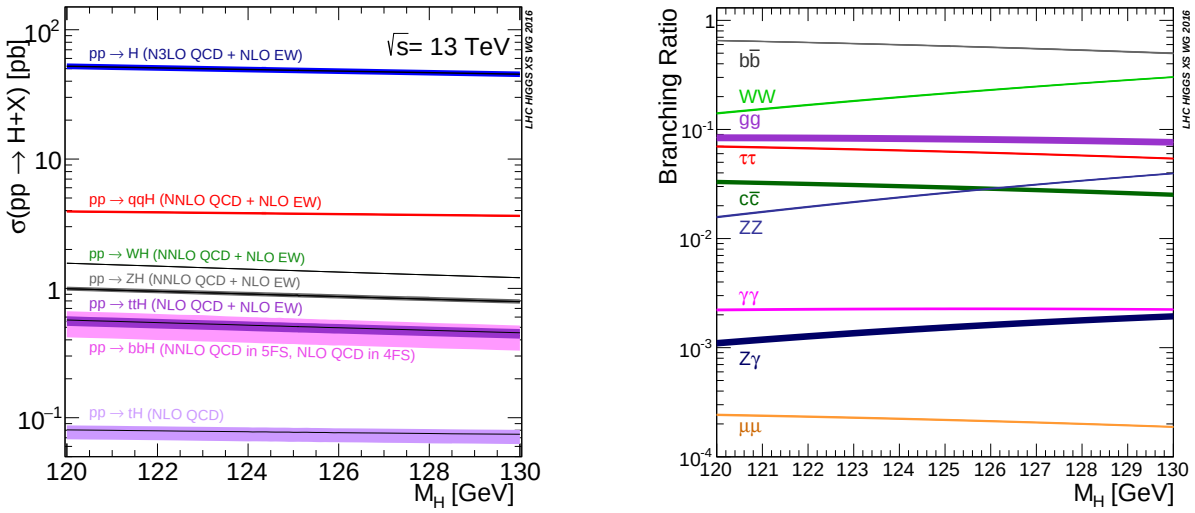


Fig. 4: Predicted production cross sections at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ and branching ratios as a function of the Higgs boson mass for a SM Higgs boson [44].

Higgs Boson Pair Production

In addition to single Higgs boson production, the production of Higgs boson pairs is also predicted in the SM. In analogy to single Higgs boson production, gluon gluon fusion via top quark loops are the dominant Higgs pair production processes. The leading-order diagrams of non-resonant Higgs boson pair production in the SM through the Higgs boson self-coupling and the Higgs-fermion Yukawa interactions are shown in Figure 5.

Fig. 5: Leading-order Feynman diagrams of non-resonant Higgs boson pair production in the SM through the Higgs boson self-coupling and the Higgs-fermion Yukawa interactions only.

The two diagrams interfere destructively, resulting in an extremely small predicted cross section of $\sigma(pp \to HH) = 31.05 \pm 0.93\,(\text{PDF} + \alpha_s) \pm 0.81(m_t)^{+0.68}_{-1.55}(\text{scale})\,\text{fb}$ at $\text{NNLO}_{\text{FTapprox}}$ for a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ [46]. Due to the presence of the self-coupling vertex HHH in the production process, a measurement of the production rate would give access to the trilinear coupling term in Equation 2.15 and therefore to the shape of the Higgs potential [47]. This would constitute a direct test of EWSB in the BEH mechanism. Even though non-resonant Higgs pair production predicted by the SM is out of reach with the current amount of data collected, the study of this process remains interesting as many BSM theories assume differing coupling strengths and predict an enlarged production cross section for Higgs pair production.

More relevant to this thesis, extensions to the Higgs sector of the SM such as 2HDM models predict heavy scalar particles X which can decay to two SM Higgs bosons (see Section 2.2.1). While non-resonant BSM enhancements would give an enlarged cross section, resonant production would result in a localized excess in the di-Higgs invariant mass spectrum m_{HH}. The Feynman diagram depicting the leading resonant production process for Higgs boson pair production is shown in Figure 6. Current experimental limits on Higgs boson pair production are discussed in Section 2.4.

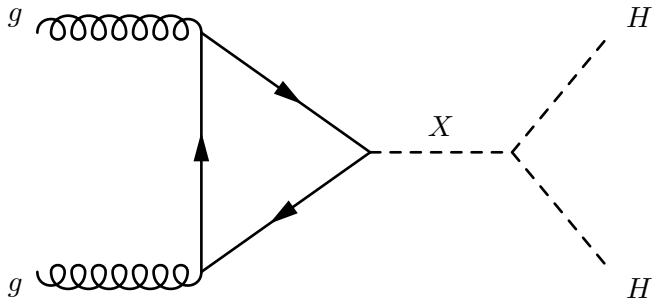


Fig. 6: Feynman diagram depicting the production of a heavy scalar resonance, X, as expected in extended 2HDM models such as the MSSM, which then decays into two lighter SM Higgs bosons.

2.4. Experimental Limits on Higgs Boson Pair Production

While no evidence for non-resonant Higgs boson pair production has been observed so far, upper limits on the cross-section have been set. Figure 7 shows upper limits set by ATLAS at 95% CL on the cross section of the gluon-gluon fusion SM HH production normalised to its SM expectation $\sigma_{\mathrm{SM}}^{\mathrm{ggF}}(pp \to HH)$ from searches in the $b\bar{b}\tau\tau$, $b\bar{b}b\bar{b}$, $b\bar{b}\gamma\gamma$, $WWWW$, $WW\gamma\gamma$ and $b\bar{b}WW^*$ decay channels. The most stringent limits were set in the $b\bar{b}\tau^+\tau^-$ channel. The combined observed limit on non-resonant Higgs boson pair production cross section is 6.9 times the predicted SM cross section. For resonant Higgs boson pair production, upper limits on the cross section for a spin-0

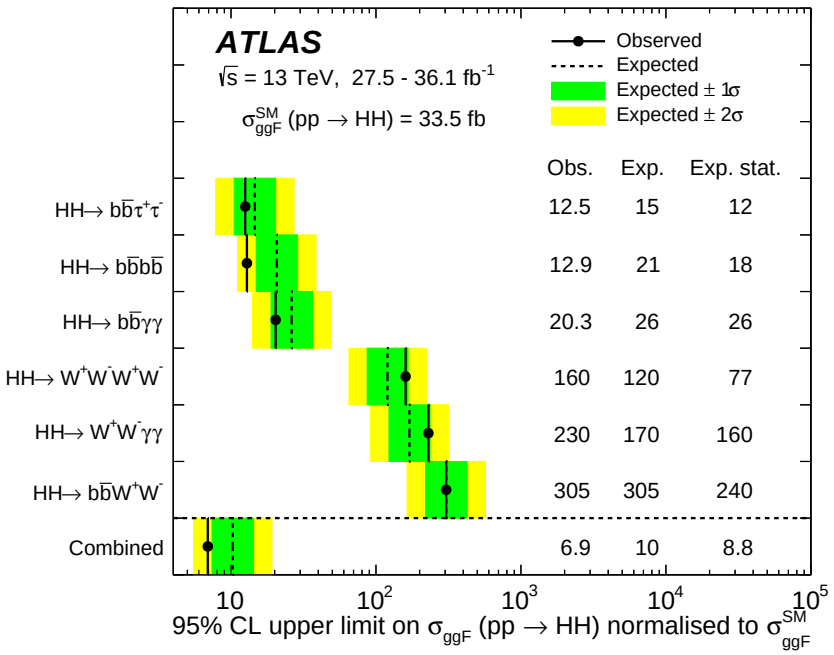


Fig. 7: Upper limits at 95% CL on the cross section of the gluon-gluon fusion SM HH production normalised to its SM expectation $\sigma_{\mathrm{SM}}^{\mathrm{ggF}}(pp \to HH)$ from the searches in the several non-resonant Higgs boson pair decay channels, and their statistical combination [48].

heavy scalar and a spin-2 KK graviton have been set by the ATLAS and CMS collaborations. The combined limits on the cross section of the resonant Higgs boson pair production for a spin-0 heavy scalar and a spin-2 KK graviton are shown in Figure 8 for ATLAS and in Figure 9 for CMS. Note that CMS combined the results from searches in the $b\bar{b}\gamma\gamma$, $b\bar{b}\tau\tau$, $b\bar{b}b\bar{b}$ and $b\bar{b}VV$ decay channels, while ATLAS additionally used results from searches in the $WWWW$ and $WW\gamma\gamma$ channels. The limits range between ~ 1 pb for low resonance masses ($m_X = 300$ GeV) and ~ 5 fb for high resonance masses ($m_X = 3$ TeV).

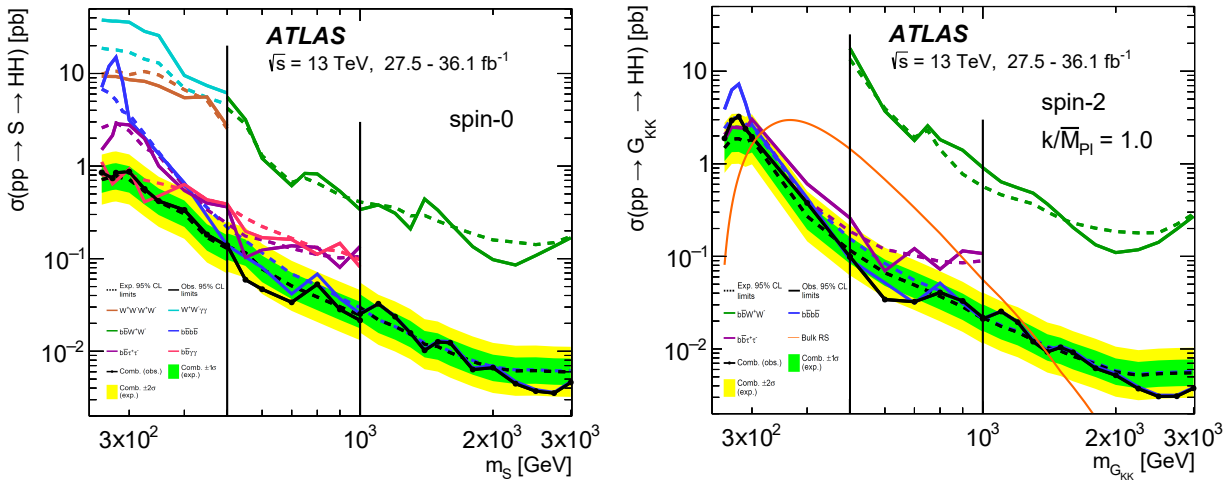


Fig. 8: Expected (dashed) and observed (solid) upper limits at 95% CL on the cross section of the resonant Higgs boson pair production for a spin-0 heavy scalar (left) and a spin-2 KK graviton with $k/\bar{M}_{\mathrm{Pl}} = 1.0$ (right). The green and yellow bands indicate the $\pm 1\sigma$ and $\pm 2\sigma$ regions respectively for the combination. Only the $b\bar{b}b\bar{b}$, $b\bar{b}W^+W^-$ and $b\bar{b}\tau^+\tau^-$ search results are used in the spin-2 resonant combination. The vertical black lines in each pannel indicate mass intervals where different final states are combined [48].

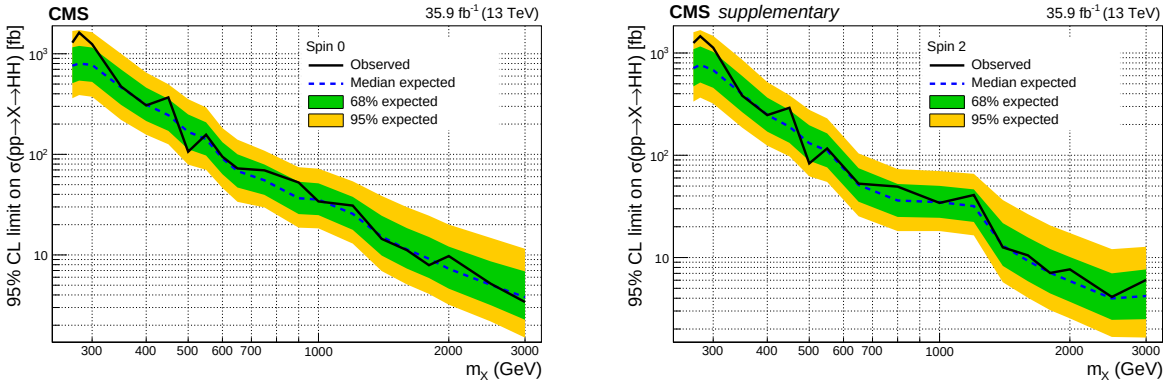


Fig. 9: Expected (dashed) and observed (solid) upper limits at 95% CL on the cross section of the resonant Higgs boson pair production for a spin-0 heavy scalar (left) and a spin-2 KK graviton with $k/\bar{M}_{\mathrm{Pl}} = 1.0$ (right). The green and yellow bands indicate the $\pm 1\sigma$ and $\pm 2\sigma$ regions respectively. [49].

2.5. Proton-Proton Collisions and Event Simulation

In pp colliders such as the LHC, multiple collisions take place when two particle bunches cross. Most of these collisions are low-energy (soft) collisions which are not of interest. Mitigating these pile-up (PU) contributions is a crucial experimental challenge. At high energies, the interesting fundamental collisions do not take place between the protons themselves. Instead, the constituents of the protons, the gluons and quarks, interact in hard collisions. Before the primary hard scattering, the partons can undergo a number of different interactions such as gluon splitting $g \to q\bar{q}$ or

the radiation of a gluon from a quark $q \to qg$ or via gluon self-interaction $g \to gg$. The resulting partons form what is known as the initial state parton shower. The initial state radiation and all subsequent interactions as well as interactions from other partons in the proton which do not take part in the hard collision constitute the underlying event (UE). Products of the hard collision form the final state parton shower in which short-lived resonances are produced and rapidly decay. Finally, as a consequence of confinement, particles carrying color charge will fragment and form jets of stable hadrons. A schematic depiction of the stages of a pp collision is given in Figure 10.

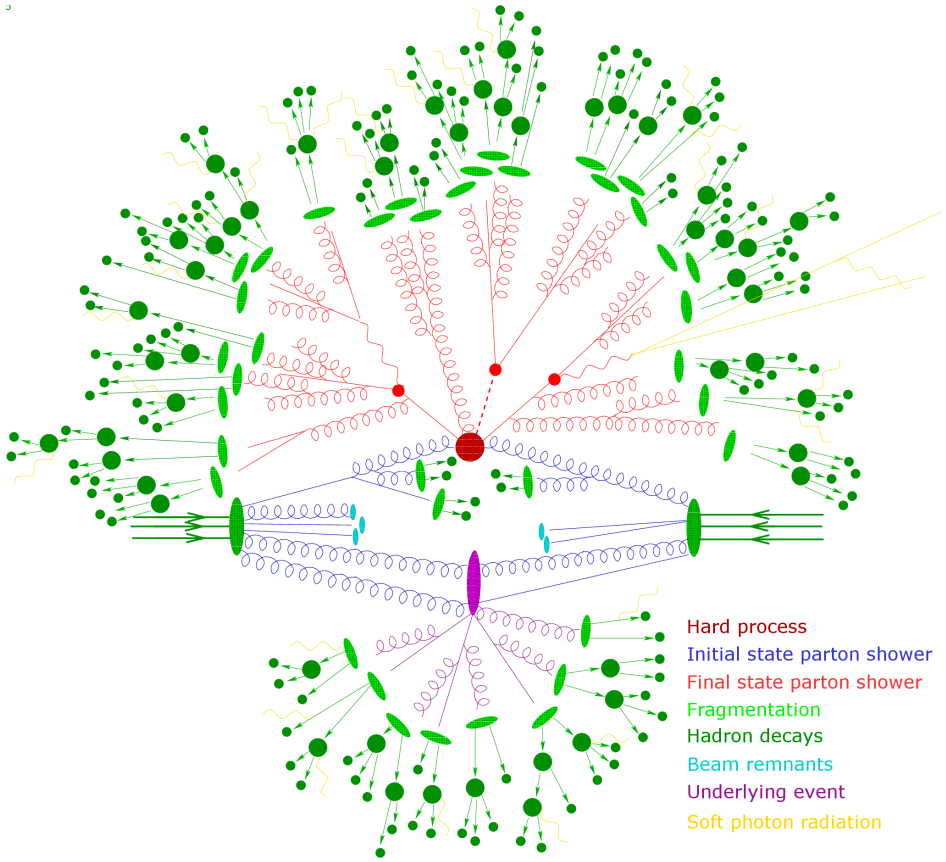


Fig. 10: Schematic depiction of the stages of simulation of a hadron-hadron collision by a parton shower Monte Carlo event generator [50].

The Monte Carlo (MC) simulation of signal and background processes is one of the most essential tools for all types of measurements and searches at hadron colliders and roughly follows the steps outlined above. The cross sections of the hard partonic interactions are extracted from scattering amplitudes $|\mathcal{M}|^2$ which can be calculated from the respective Lagrangian using a perturbative expansion. The differential cross section of an interaction is given by

$$\mathrm{d}\sigma = \frac{|\mathcal{M}|^2}{F}\mathrm{d}Q, \tag{2.21}$$

where F is known as the Møller flux and $\mathrm{d}Q$ is a Lorentz-invariant phase space factor. While the energy of the protons corresponds to the beam energy E_{beam}, the energy of the proton constituents,

which determine the hard scatter, is unknown. Experimentally derived parton density functions (PDFs) $f_i(x,Q^2)$ are used to encode the distributions of their momentum fractions x which they carry inside the proton at a fixed momentum transfer Q_0^2. The evolution of the PDFs with Q^2 is described by the DGLAP [51–53] equations. The total hadronic cross-section of a process $pp \to X$ is then given by the factorisation theorem [54]:

$$\sigma_{pp\to X} = \int \mathrm{d}x_1 \mathrm{d}x_2 \sum_{q_1,q_2} f_{q_1}(x_1,\mu_F^2) f_{q_2}(x_2,\mu_F^2) \hat{\sigma}_{q_1 q_2 \to X}(\hat{s},\mu_F^2). \tag{2.22}$$

The sum is over all parton combinations q_1, q_2 in the proton and $\hat{\sigma}_{q_1 q_2 \to X}(\hat{s},\mu_F^2)$ is the partonic cross section for the process $q_1 q_2 \to X$ at $\hat{s} = x_1 x_2 s$. The parameter μ_F^2 is known as factorization scale and marks a cut-off point at which perturbative calculations of the partonic cross sections at low momentum transfer scales become infeasible such that they start being encoded in the PDFs. The aforementioned parton showers manifest as a sequence of gluon radiations with ever decreasing momentum scales. At some point, perturbative calculations fail and phenomenological hadronisation models are required. One widely used model for hadronisation is the Lund model [55].

Several programs have been developed in recent years for matrix-element calculations, parton showering, hadronisation as well as underlying event simulation. In this thesis, the matrix elements for the signal processes are simulated using the MADGRAPH5_aMC@NLO [56] event generator while for parton showering, hadronisation and underlying event, HERWIG++ [57] is used. As for the PDFs, the NNPDF23 PDF [58] set was applied. For the different simulation steps of background processes, PYTHIA8 [59], POWHEG [60] as well as SHERPA [61] are used. A complete list of the samples which are used for producing the results presented in this thesis can be found in Appendix D.

In addition to the pure event simulation, a precise simulation of the detector response is equally important. A detailed simulation of the ATLAS detector is possible with the GEANT4 [62] program, which simulates the passage of each single particle through the detector material as well as the detectors response. However, the GEANT4 simulation is very CPU intensive and the limited amount of MC events is one of the largest sources of systematic uncertainties. To fulfil the MC needs, ATLAS developed the GEANT4-based fast simulation ATLFAST-II (AFII) [63], which is used to produce all signal samples in this thesis. While the AFII simulation of the inner detector is identical to the full simulation of GEANT4, approximate methods are used to simulate the electromagnetic and hadronic calorimeter response.

Chapter 3

The ATLAS detector at the LHC

To explore the current energy frontier, large accelerators and sophisticated particle detectors are required. In circular colliders like the LHC, high particle density bunches are accelerated and then brought to collision at the intersection points of the beam lines. The luminosity of an accelerator with two colliding bunches containing N_1 and N_2 particles is given by

$$\mathscr{L} = f \frac{N_1 N_2}{4\pi\sigma_x\sigma_y}, \tag{3.1}$$

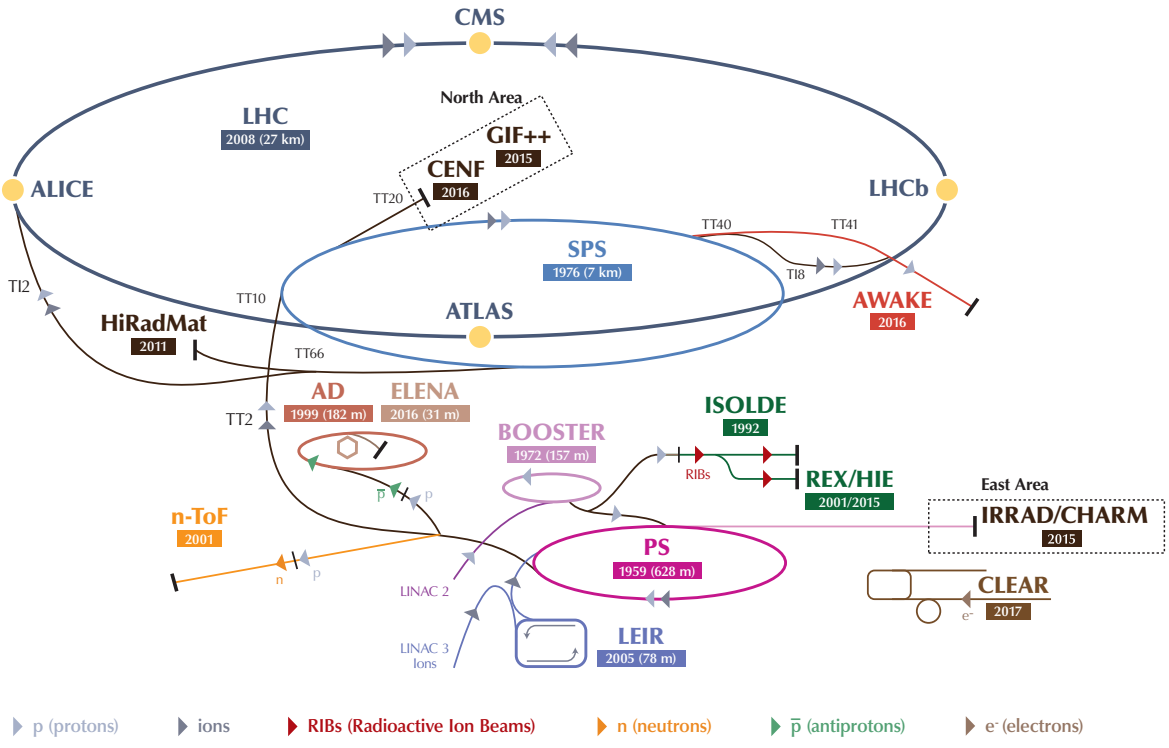
where f is the frequency at which the bunches cross, and σ_x and σ_y the horizontal and vertical beam dimensions, respectively. The luminosity is proportional to the interaction rate

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \mathscr{L} \cdot \sigma \tag{3.2}$$

of a process with cross section σ. Hence, in addition to the center-of-mass energy, the luminosity is the single most important benchmark value used to assess the performance of a collider. A related quantity is the integrated luminosity, $L = \int \mathscr{L}\,\mathrm{d}t$, which is used to characterize the amount of data collected over a specific time period. In the following, the LHC and the main components of the ATLAS detector are described.

3.1. The Large Hadron Collider

The LHC [64] is situated 175 m below earth's surface at the European Organization for Nuclear Research (CERN) in Geneva, Switzerland and is the largest particle accelerator ever constructed by mankind. With a circumference of 26.7 km, it can accelerate and collide protons as well as heavy ions. Particles are injected into two adjacent parallel beam pipes of the LHC after they pass a series of pre-accelerators which successively increase the particles energy. They are then brought to collision at four intersection points where the main experiments are located: the ATLAS [65], CMS [66], LHCb [67] and ALICE [68] experiments. While ATLAS and CMS are general purpose detectors, ALICE is specialised in heavy ion collisions and LHCb focusses on the study of b-physics. The beams are kept on their circular path by superconducting dipole magnets which are cooled down by liquid helium and operate at a temperature of 1.9 K, while higher order magnets are used to keep the beam focussed. A sketch of the CERN accelerator complex showing the various pre-accelerators and detectors is shown in Figure 11.



LHC - Large Hadron Collider // SPS - Super Proton Synchrotron // PS - Proton Synchrotron // AD - Antiproton Decelerator // CLEAR - CERN Linear Electron Accelerator for Research // AWAKE - Advanced WAKefield Experiment // ISOLDE - Isotope Separator OnLine // REX/HIE - Radioactive EXperiment/High Intensity and Energy ISOLDE // LEIR - Low Energy Ion Ring // LINAC - LINear ACcelerator // n-ToF - Neutrons Time Of Flight // HiRadMat - High-Radiation to Materials // CHARM - Cern High energy AcceleRator Mixed field facility // IRRAD - proton IRRADiation facility // GIF++ - Gamma Irradiation Facility // CENF - CErn Neutrino platForm

Fig. 11: Sketch of the CERN acclerator complex showing the various pre-accelerators and detectors [69].

For pp collisions, the LHC is designed to provide a center-of-mass energy of $\sqrt{s} = 14\,\text{TeV}$ and a peak luminosity of $\mathscr{L} = 10^{34}\text{cm}^{-2}\,\text{s}^{-1}$. It can hold up to 2808 proton bunches per beam with a bunch spacing of 25 ns and up to $1.15 \cdot 10^{11}$ protons per bunch. In 2010, the LHC was commissioned for pp collisions with a $\sqrt{s} = 7\,\text{TeV}$ center-of-mass energy which was later increased to $\sqrt{s} = 8\,\text{TeV}$, concluding Run 1 in 2012. In Run 1, the LHC delivered a total integrated luminosity of $5.46\,\text{fb}^{-1}$ at $\sqrt{s} = 7\,\text{TeV}$ and $22.8\,\text{fb}^{-1}$ at $\sqrt{s} = 8\,\text{TeV}$. After a first long shutdown (LS1) period used for maintaining and upgrading the detector as well as the accelerator components, the LHC restarted in 2015 with an increased center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$. Run 2 concluded in late 2018 and delivered a total integrated luminosity of $156\,\text{fb}^{-1}$. The machine is currently undergoing a second upgrade phase LS2 and is expected to restart its physics programme in 2021, expectedly reaching the design value of $\sqrt{s} = 14\,\text{TeV}$ for the center-of-mass energy. In 2024, the LHC will undergo a third upgrade phase LS3 after which the machine will start in the HL-LHC [70] configuration, aiming to provide a peak luminosity of up to $\mathscr{L} = 5 \cdot 10^{34}\text{cm}^{-2}\,\text{s}^{-1}$ and allowing an integrated luminosity of $250\,\text{fb}^{-1}$ per year with the goal of 3000 fb^{-1} in about 12 years after the upgrade.

3.2. The ATLAS Detector

The ATLAS detector is a multi-purpose detector located at the LHC designed to measure the properties of as many particles originating from the collision as precisely as possible. It consists of end-caps and a series of concentric layers around the collision point which act as various detector types and almost cover the full 4π solid angle around the interaction point. The detector has a diameter of 25 m, a length of 44 m and weighs approximately 7,000 tonnes. A sketch of the components and subsystems of the ATLAS detector is provided in Figure 12. The three main parts of the detector, namely the inner detector, the calorimeters and the muon spectrometer are briefly described in the following.

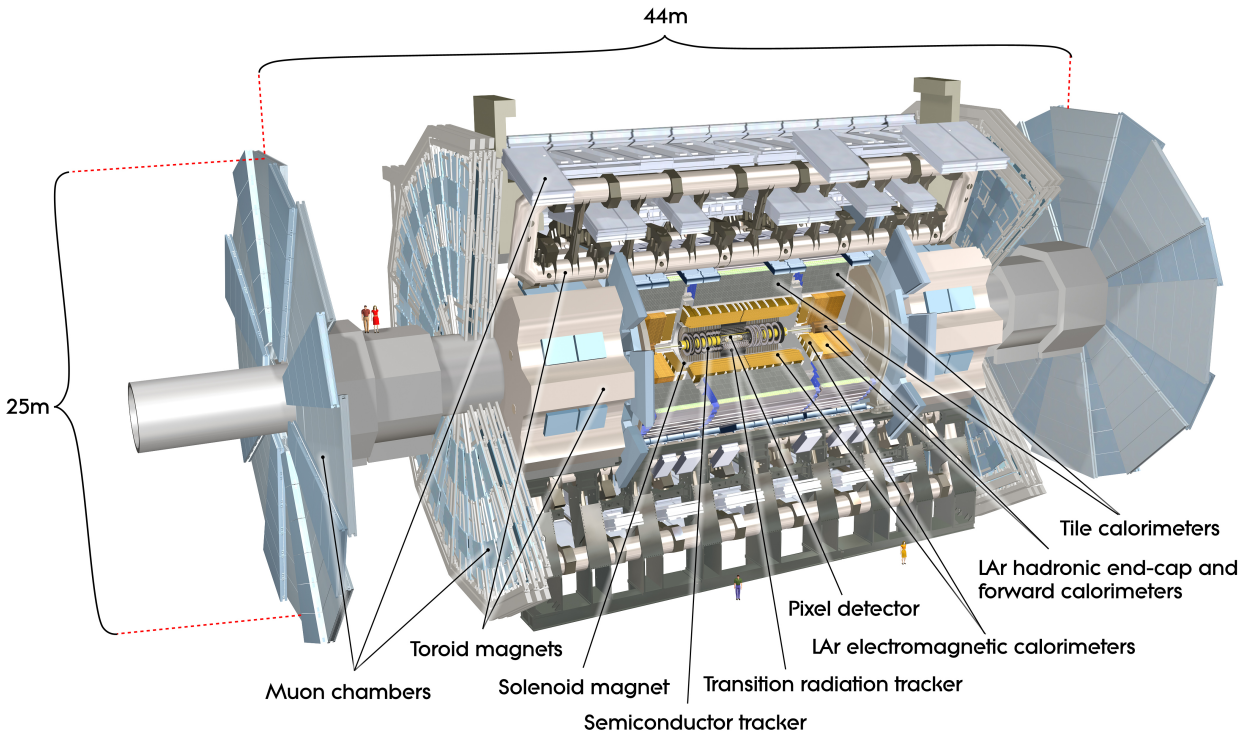


Fig. 12: Sketch of the ATLAS detector and its subsystems [65].

3.2.1. Coordinate System

As a result of the cylindrical shape of the detector, ATLAS uses a right-handed cylindrical coordinate system. The interaction point is defined as the origin of the coordinate system. The beam line corresponds to the z-axis while the $x-y$ plane is defined as the transverse plane. The azimuthal angle ϕ is defined as the angle around the beam line, while θ denotes the polar angle with respect to the beam line. The pseudorapidity $\eta \equiv -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]$ is usually used instead of the polar angle. In the relativistic limit, pseudorapidity is equal to the rapidity $y \equiv \frac{1}{2}\cdot\ln\left(\frac{E+p_L}{E-p_L}\right)$. Differences in rapidity

are Lorentz invariant under boosts along the longitudinal axis. The separation between objects in the detector is usually measured by defining the angular separation $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ in the $\eta - \phi$ plane.

3.2.2. Inner Detector

The inner detector (ID) is the innermost layer of the ATLAS detector and is designed to measure the tracks from charged particles. It starts 3.3 cm from the beam axis and extends to a length of 6.2 m and a radius of 1.2 m. The ID consists of four components: the insertable B-Layer (IBL), the pixel detector, the semi-conductor tracker (SCT) and the transition radiation tracker (TRT) (see Figure 12). The IBL was installed during LS1 and uses silicon pixels in order to assist the pixel detector in measuring the tracks of the particles and improve the tracking robustness as well as the stability in high luminosity conditions. Furthermore, the improved tracking performance from the IBL also helps to reconstruct the primary interaction vertex as well as secondary vertices, which occur in the decay of B-hadrons and are a important tool for b-jet identification (see Section 4.5). The pixel detector itself consists of 1,740 modules, each containing approximately 74,000 pixels (50 μm $\times$ 400 μm) and therefore provides extremely precise tracking. The SCT resembles the pixel detector, but employs narrow strips instead of pixels. The last layer, the TRT, uses drift tubes which are filled with an ionizing gas mixture. The space between the straws consists of materials of differing refraction indices such that traversing particles generate transition radiation (TR). As lighter particles such as electrons are more likely to produce TR compared to heavy hadrons, the TRT provides aid to particle identification. The complete ID structure is immersed in a 2 T solenoidal magnetic field which causes the path of charged particles to bend and allows the measurement of their transverse momentum from the curvature of their tracks. The design resolution and coverage angle of the ID is shown in Table 6.

3.2.3. Calorimeters

Outside of the ID, calorimeters are used to measure the energy of the traversing particles. In general, calorimeters consist of highly dense material with which the incoming particles interact, causing particle showers which are then absorbed by the material such that ideally all of the energy is deposited in the calorimeter. ATLAS uses sampling calorimeters which consist of alternating layers of active material in which the energy is measured and passive layers in which the absorber material is situated. The calorimeter system consists of an inner electromagnetic calorimeter (ECAL), which is designed to measure the energy of photons and electrons and an outer hadronic calorimeter (HCAL), which is designed to measure the energy of hadrons. The ECAL is finely segmented and is composed of lead as the absorber and liquid argon (LAr) as the active medium. In contrast, the barrel HCAL has lower granularity and consists of steel absorbers and scintillating tiles. The coverage of the HCAL is extended with forward calorimeters and hadronic end-cap calorimeters (HEC) consisting of two wheels per end-cap which use LAr as active and copper as absorbing

material. In order to contain a large fraction of the electromagnetic shower, the thickness of the ECAL is $> 22X_0$ in the barrel and $> 24X_0$ in the end-caps, where X_0 is known as the radiation length and defined as the mean distance over which electromagnetically interacting particles lose all but $1/e$ of their original energy. The HCAL consists of three layers of thickness 1.5λ, 4.1λ and 1.8λ in the central barrel region and three layers of thickness 1.5λ, 2.6λ and 3.3λ in the extended barrel region, where λ is known as the interaction length and is defined as the mean distance over which hadronically interacting particles lose all but $1/e$ of their original energy. The design resolutions and coverage angles of the ECAL and HCAL are shown in Table 6.

3.2.4. Muon Spectrometer

The outermost layer of the ATLAS detector which begins at a radius of 4.25 m and extends to the full 11 m radius of the ATLAS detector is the muon spectrometer. As compared to the electron, the muon is approximately 200 times heavier, the emission of Bremsstrahlung is greatly reduced and consequently, muons barely interact in the detector and leave little energy in the calorimeter systems. Immersed in a large toroidal magnetic field in the barrel and end-cap regions, drift tubes as well as cathode strip chambers provide transverse momentum measurements through the amount of bending of the muon tracks. The design resolution and coverage angle of the muon spectrometer is shown in Table 6.

3.2.5. Trigger System

In principle, it is desirable to record all data produced by the collisions at the LHC for offline analysis. However, with a bunch crossing space of 25 ns, 600 TB of data would need to be processed and stored every second. With the current technology, data processing rates at these magnitudes are infeasible. Instead, ATLAS uses a multilevel trigger system [71] which is designed to efficiently discard uninteresting events while permanently saving potentially interesting events for future analysis. For Run 2, ATLAS employed a two-level trigger system consisting of the Level 1 (L1) hardware trigger and the software-based Higher Level Trigger (HLT). The L1 trigger selects events with a latency of 2.5 μs using only information from the calorimeters and muon spectrometers, effectively reducing the initial event rate from 40 MHz to 100 kHz. In addition, the L1 trigger identifies Regions of Interest (RoI) in which it identifies potentially interesting features such as high-p_T jets, leptons or photons. The RoI are passed to the HLT which uses all the available reconstructed detector data within the RoI to further reduce to event rate to 1 kHz. Thereafter, the remaining events are permanently stored for further analysis. The coverage angles for the trigger systems in the various detector components are shown in Table 6.

Detector component	Resolution	η coverage	
		Measurement	Trigger
Tracking	$\sigma_{p_T}/p_T = 0.05\% \cdot p_T \oplus 1\%$	$\lvert\eta\rvert < 2.5$	
TRT		$\lvert\eta\rvert < 2.0$	
ECAL	$\sigma_E/E = 10\%/\sqrt{E} \oplus 0.7\%$	$\lvert\eta\rvert < 3.2$	$\lvert\eta\rvert < 2.5$
HCAL			
barrel and end-cap	$\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$	$\lvert\eta\rvert < 3.2$	$\lvert\eta\rvert < 3.2$
forward	$\sigma_E/E = 100\%/\sqrt{E} \oplus 10\%$	$3.1 < \lvert\eta\rvert < 4.9$	$3.1 < \lvert\eta\rvert < 4.9$
Muon spectrometer	$\sigma_{p_T}/p_T = 10\%$ at $p_T = 1\,\text{TeV}$	$\lvert\eta\rvert < 2.7$	$\lvert\eta\rvert < 2.4$

Tab. 6: General performance of the ATLAS detector showing the resolutions and coverage angles of the various components. E and p_T are in GeV. The values are taken from Ref. [65].

Chapter 4

Jets in Boosted Topologies

With an unprecedented center-of-mass energy available at the LHC, jets with transverse momenta of up to a few TeV are produced. High-p_T jets originating from the hadronic decay products of Higgs bosons, W/Z bosons and top quarks open an important window for searches for new physics. The efficient reconstruction and identification of these jets is crucial in order to gain sensitivity to hypothetical heavy particles predicted by BSM theories. The mean angular separation of the decay products from a resonance with mass m and transverse momentum p_T can be estimated by [72]

$$\Delta R \approx \frac{2m}{p_T}. \tag{4.1}$$

In the case of jets originating from low-p_T resonances, the hadronic decay products are well-separated in the detector. In this *resolved* regime, the decay products can be individually reconstructed. In contrast, the decay products originating from high-p_T resonances are *boosted* and tend to appear very close in the detector such that an individual resolution is no longer possible. In recent years, several techniques have been developed to reconstruct and tag boosted Higgs boson, W/Z boson and top quark jets. In this chapter, jet reconstruction and tagging techniques with focus on the boosted $H \to b\bar{b}$ decay are described.

4.1. Jet Reconstruction

As mentioned in Section 2.5, partons are not directly observable but fragment and form a collimated spray of hadrons in the detector. While all hadrons deposit energy in the calorimeter, charged hadrons additionally leave tracks in the ID. This information is then used to group together stable hadrons, ideally reconstructing all hadrons originating from the original parton into one single object known as jet. This procedure allows to infer properties from the original parton and therefore constitutes the link between the most fundamental interactions and the observable particles in the detector. While jet reconstruction algorithms are well-motivated *ad-hoc* methods, they do not follow an underlying physical principle. However, there is a broad consensus on the requirements these algorithms need to fulfil. For instance, jet reconstruction algorithms should be infrared (IR) and collinear safe (IRC safe) [73]. In IR safe algorithms, the addition of soft particles does not significantly distort the output of the algorithm such that sensitivity to UE and PU is greatly reduced. In collinear safe algorithms, the invariance of the algorithm's output after the splitting of a high-p_T particle into two collinear particles is ensured. At the present day, IRC safe sequential

jet clustering algorithms are most widely used. The most common is known as the generalized-k_t algorithm and follows the following procedure:

Step 1: Based on the inputs provided, create a list of proto-jets and compute the distance defined by a certain metric d_{ij} between every pair of proto-jets i and j.

Step 2: Compute the distance $d_{i,B} = p_{T,i}^{2p}$ between each proto-jet i and the beam B.

Step 3: Find the minimum distance.

Step 4: If the minimum distance is between the proto-jets, merge both objects to form a new proto-jet and remove the individual proto-jets from the list. If the minimum distance is between the proto-jet and the beam, remove the proto-jet from the list and call it a jet.

Step 5: Repeat the procedure until no proto-jets are left.

The metric d_{ij} is given by

$$d_{ij} = \min\left(p_{T,i}^{2p}, p_{T,j}^{2p}\right) \frac{\Delta R_{ij}^2}{R^2} \quad \text{with} \quad p = \begin{cases} +1, & k_t \\ 0, & \text{Cambridge/Aachen} \\ -1, & \text{anti-}k_t \end{cases}, \tag{4.2}$$

where p defines the type of the clustering algorithm and R is known as the size parameter of the jet.[1] A value of $p = +1$ corresponds to the k_t (KT) algorithm [74], which begins by clustering low-p_T proto-jets first. In contrast, the anti-k_t (AKT) algorithm [75], defined by $p = -1$, begins by clustering high-p_T proto-jets first. In contrast, the Cambridge/Aachen (CA) algorithm [76] offers a p_T independent jet clustering. A comparison between the output of the AKT and the output of the CA algorithm using the same MC event with a size parameter $R = 1.0$ is shown in Figure 13.

Note that while the jets produced using the AKT algorithm have regular and cone-like shapes, the same event reconstructed with the CA algorithm yields more irregular jet shapes. The size parameter R can be tuned to provide optimal performance. While a large jet radius is beneficial to ensure full containment of the hadrons resulting from the original parton, a smaller radius helps to reduce the PU and UE event contribution and hence prevents an overestimation of the jet energy and mass.

In recent years, extensions to the aforementioned jet clustering algorithms have been developed, in particular in the context of boosted analysis. The variable-R (VR) [77] modification introduces a jet p_T dependence in the size parameter by replacing it with an effective size parameter

$$R \to R_{\text{eff}}(p_T) = \frac{\rho}{p_T}, \tag{4.3}$$

[1] In this case, the angular separation ΔR_{ij} is defined by $\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)$ with the rapidity y as opposed to the more common definition introduced in Section 3.2.1 which uses the pseudorapidity η.

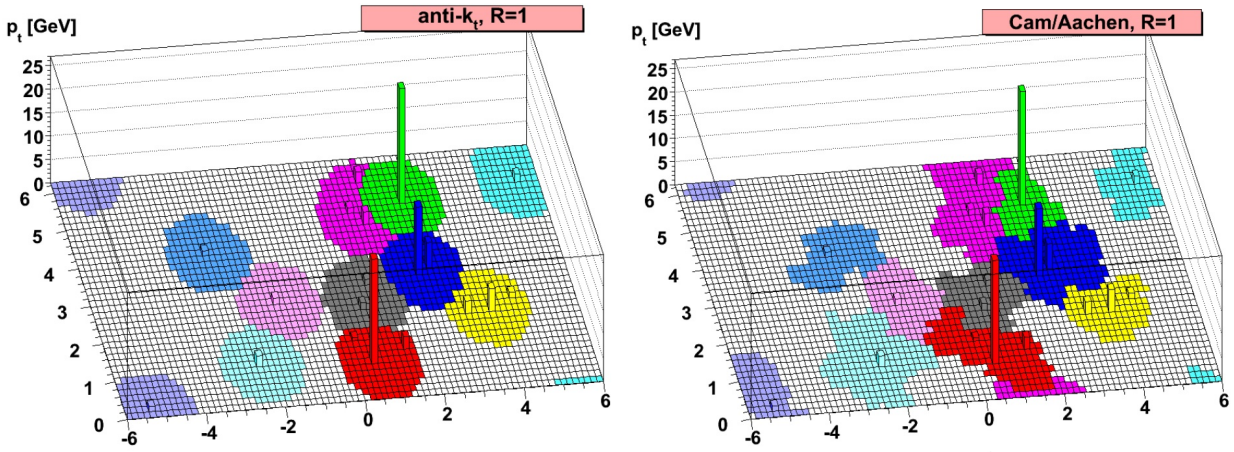


Fig. 13: The same MC event reconstructed using the AKT (left) and the CA (right) algorithm. Energy depositions are depicted in the y-ϕ plane and the respective transverse momentum p_T is shown on the z-axis. Each color corresponds to a reconstructed jet [75].

where ρ is a free dimensionful parameter that determines the absolute scaling. In addition to ρ, minimum and maximum cut-off values R_{min} and R_{max} for R_{eff} are defined. These prevent jets shrinking below detector resolution at high p_T and jets becoming too large at low p_T. All three parameters ρ, R_{min} and R_{max} can be tuned to give the best performance for the specific use case. In the following, the VR extensions to the KT, AKT and CA algorithms are denoted KT-VR, AKT-VR and CA-VR, while the standard fixed-R algorithms are denoted KT-FR, AKT-FR and CA-FR, respectively. A comparison between the output of the AKT-FR algorithm and its VR modification using the same MC event is shown in Figure 14. Note that the high-p_T jets (dark

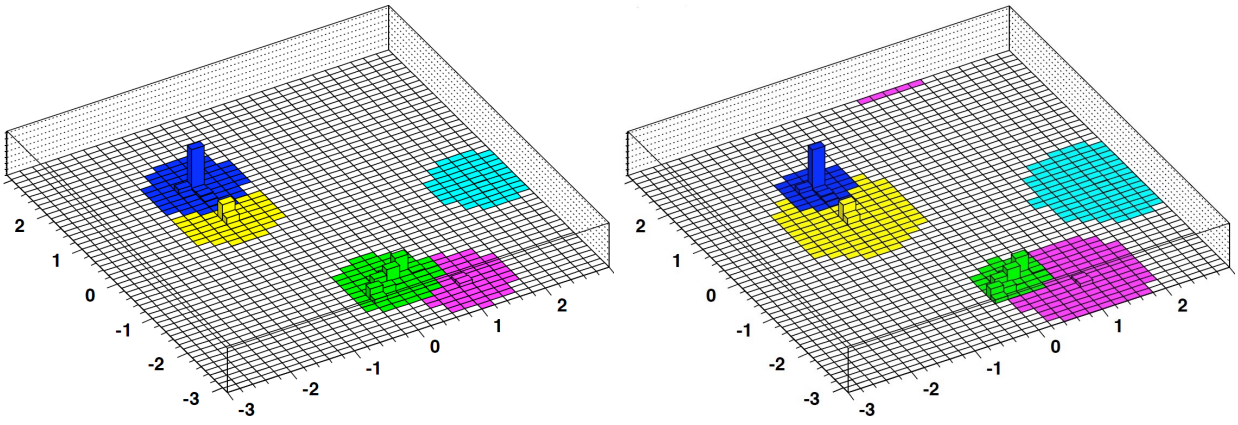


Fig. 14: The same event reconstructed by AKT-FR (left) and its VR modification (right). Energy depositions are depicted in the y-ϕ plane and the respective transverse momentum p_T is shown on the z-axis. Each color corresponds to a reconstructed jet [77].

blue, green) have been reduced in size while softer jets (yellow, purple, light blue) have grown. The use of VR jets is further motivated in Section 4.6. In ATLAS, the jet reconstruction is performed using the FASTJET [78] clustering package.

4.2. Input to Jet Reconstruction

As mentioned in Section 4.1, jets can be clustered using the four-momenta of different input quantities. Calorimeter jets (CALO jets) are built from topological clusters of calorimeter cell energy depositions [79]. In ATLAS, CALO jets are built with the AKT algorithm with size parameter $R = 1.0$ ("large-R CALO jets"), with size parameter $R = 0.4$ ("small-R CALO jets") or with size parameter $R = 0.2$. Compared to small-R jets, large-R jets have a larger catchment area such that they are more susceptible to UE and PU contributions. Several grooming techniques which aim to remove soft objects from the jets have been developed in order to mitigate these contributions. As standard grooming procedure, ATLAS adopted the trimming algorithm [80] for which a schematic depiction is provided in Figure 15. In a first step, the constituents of the large-R jet are reclustered with the KT algorithm to form subjets with a smaller size parameter $R_{\text{sub}} < R$. Subjets are then discarded if the ratio between the transverse momentum of the subjet and the transverse momentum of the large-R jet is smaller than a fixed cut-off value f_{cut}. The remaining subjets are then merged and form what is known as the trimmed jet. In ATLAS, the standard values $R_{\text{sub}} = 0.2$ and $f_{\text{cut}} = 0.05$ are used. After the trimming procedure, the mass and energy of the jets are calibrated

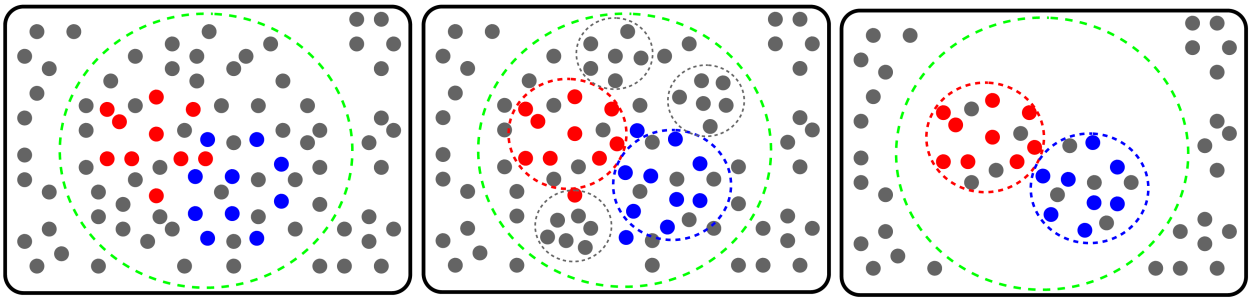


Fig. 15: Schematic depiction of the grooming method known as trimming algorithm in the η-ϕ plane. The grey dots show jet constituents from UE and PU contributions, while the blue and red dots show constituents from the hard interaction. The green dotted circle illustrates the boundary of the large-R jet.

by applying correction factors derived using simulated events [81, 82]. This is required in order to account for detector effects such as the underestimation of the jets energy as a result of energy depositions in the passive layers of the sampling calorimeters.

Furthermore, standard track jets in ATLAS are reconstructed from ID tracks with the AKT algorithm using a size parameter of $R = 0.2$. In order to reduce the number of tracks originating from PU, the tracks are required to fulfil the following quality requirements:

- $p_T > 0.4\,\text{GeV}$ and $|\eta| < 2.5$,
- at least 7 hits in the pixel and SCT detectors,
- not more than one missing expected hit in the pixel detector and not more than two missing expected hits in the combined pixel and SCT detectors,
- not more than one hit in the pixel detector that is shared by multiple tracks,

- the track is required to be a constituent of the primary vertex or satisfies $|z_o \sin\theta| < 3.0\,\text{mm}$,

where z_0 is the longitudinal impact parameter with respect to the primary vertex and θ is the polar angle of the track momentum. Note that the primary vertex is defined as the reconstructed vertex with the highest scalar sum of the transverse momenta of the associated tracks. In addition, the track jets can be also reconstructed using the AKT-VR algorithm which is expected to considerably improve the jet identification performance in the boosted regime (see Section 4.6).

Finally, truth jets can be constructed with the identical clustering algorithms used for the reconstruction of the CALO jets. These jets are built in simulated events from stable truth particles with $c\tau > 10\,\text{mm}$, excluding muons and neutrinos, and are useful as a reference for detector-level jets.

4.3. Jet Substructure

Even though it is not possible to individually resolve the decay products in heavily boosted topologies, information about the original jet-initiating partons can be gained using jet substructure variables. One widely used variable is known as N-subjettiness τ_N [83]. To calculate this variable, one first reclusters the jet constituents from the candidate jet into exactly N KT subjets. Using these subjets, τ_N is defined as

$$\tau_N = \frac{\sum\limits_{i \in J} p_{T,\,i} \min\left(\Delta R_{k,\,i} \,|\, k = 1,..., N\right)}{\sum\limits_{i \in J} p_{T,\,i}\, R}, \tag{4.4}$$

where the sum goes over the constituent particles of a given jet and $\Delta R_{k,\,i}$ is the angular separation between a candidate subjet k and the constituent particle i. N-subjettiness is a measure of how well a given jet is described by being composed of at least N subjets. In practice, the ratio $\tau_{NM} = \tau_N/\tau_{N-1}$ is used. For a particular jet containing N-subjets ("N-prong jet"), τ_N is expected to yield small values, while τ_{N-1} is expected to give a sharp rise such that $\tau_{NM} \ll 1$. A graphic illustration of the properties of N-subjettiness is shown in Figure 16. N-subjettiness ratios have

Fig. 16: Graphical illustration of a large-R jet with low τ_2 values (left) and a large-R jet with high τ_2 values (right) in the η-ϕ plane. The green circles show clustered subjets while the red circles represent additional energy depositions within the jet.

been shown to be effective in identifying top [84], W [85] and $H \to WW^*$ [86] jets. While the

subjet axis for calculating the N-subjettiness variables is usually defined by the vector sum of all its constituents, the axis can also be defined by its hardest constituent and in this case is known as the Winner Takes All (WTA) axis [87]. N-subjettiness variables computed using the WTA axis are denoted as τ_{NM}^{WTA}.

A further set of interesting variables are known as energy correlation functions (ECFs) [88] and are defined as

$$\mathrm{ECF}\,(N,\beta) = \sum_{i_1<i_2<\ldots<i_N\in J} \left(\prod_{a=1}^{N} p_{T_{i_a}}\right) \left(\prod_{b=1}^{N-1}\prod_{c=b+1}^{N} \Delta R_{i_b i_c}\right)^{\beta}, \tag{4.5}$$

where the sum goes over all constituents i of a jet J and β is a positive tunable parameter. In contrast to N-subjettiness, ECFs are only based on energy and angular information of particles within the jet and do not require computationally intensive subjet finding procedures. In analogy to the ratio τ_{NM} one defines the potentially discriminating dimensionless ratios

$$r_N^{(\beta)} = \frac{\mathrm{ECF}(N+1,\beta)}{\mathrm{ECF}(N,\beta)}, \tag{4.6}$$

as well as

$$C_N^{(\beta)} = \frac{r_N^{(\beta)}}{r_{N-1}^{(\beta)}} = \frac{\mathrm{ECF}_{N+1}\times\mathrm{ECF}_{N-1}}{\mathrm{ECF}_N^2} \tag{4.7}$$

and

$$D_N^{(\beta)} = \frac{\mathrm{ECF}_{N+1}\times\mathrm{ECF}_{N-1}\times\mathrm{ECF}_1^N}{\mathrm{ECF}_N^3}. \tag{4.8}$$

The variables C_2^{β} and D_2^{β} are a measure for the transverse energy distribution within a jet and it has been observed that $C_2^{\beta=1}$ is an effective variable to discriminate 2-prong decays, such as jets originating from W-boson decays [85]. In recent years, jet substructure variables based on more flexible, generalized forms of the standards ECFs have been developed [89] . These variables include the generalized ECF ratios M_2, N_2, N_3 [89] and the L-series $L_1 - L_5$ [90].

4.4. Track-Assisted Reclustered Jets

As a result of the finite angular resolution of the Atlas calorimeter, the reconstruction of jet substructure variables from large-R CALO jets becomes increasingly difficult in boosted topologies, where jets appear very close in the detector. To help overcome this limitation it is possible to assist the jet reconstruction using track information from the ID, which offers excellent angular resolution. As a standard procedure, Atlas already employs the track-assisted (TA) jet mass for standard large-R CALO jets. Conventionally, the mass of a jet J reconstructed in the calorimeter is defined as

$$m^{\mathrm{calo}} = \sqrt{\left(\sum_{i\in J} E_i\right)^2 - \left(\sum_{i\in J} \vec{p_i}\right)^2}, \tag{4.9}$$

where the sum goes over all cluster constituents i of the jet J. In order to include track information, the particle tracks in the ID are associated with the calorimeter clusters through a procedure known as ghost-association [91]. For this, the p_T of the tracks is set to an infinitesimal value. Then, the scaled tracks are added to the input list of the corresponding clustering algorithm. The scaling ensures that only angular information is added and the jet reconstruction is not distorted. The TA jet mass is then defined as [82]

$$m^{\mathrm{TA}} = \frac{p_T^{\mathrm{calo}}}{p_T^{\mathrm{track}}} \cdot m^{\mathrm{track}}, \tag{4.10}$$

where m_{track} is the mass and p_T^{track} the transverse momentum of the associated tracks. The reweighting $p_T^{\mathrm{calo}}/p_T^{\mathrm{track}}$ is required to ensure the inclusion of the energy of neutral hadrons to which the ID is not sensitive. As a further motivation, m^{TA} is particularly useful when using the AFII detector simulation as the parametrized detector response of the calorimeter results in a poorly modelled calorimeter-based mass.

While this approach improves the resolution of jet mass measurements and in principle could be extended to further jet substructure variables, it does not take into account local differences in energy deposited from neutral and charged hadrons within the jet. A novel jet collection known as track-assisted reclustered (TAR) jets [15] aims to circumvent this problem by applying local corrections to the p_T of close-by tracks. TAR jets are built using tracks and calibrated AKT-FR jets with a size parameter of $R = 0.2$ reconstructed from energy clusters in the calorimeters. The AKT-FR jets are reclustered into large-R jets and then trimmed using the parameter $f_{\mathrm{cut}} = 0.05$ such that soft components are removed. The remaining AKT-FR ($R = 0.2$) jets are assigned to tracks using ghost-association. If tracks are not associated to any of the jets, they are matched to the closest $R = 0.2$ jet in case the jet axis lies within $\Delta R < 0.3$ of the track. If a jet axis does not lie within this value, the non-associated track is discarded. After the matching procedure, the p_T of the associated tracks is rescaled according to

$$p_T^{\text{track, new}} = p_T^{\text{track, old}} \times \frac{p_{T,j}^{\text{subjet}}}{\sum\limits_{i \in j} p_{T,i}^{\text{track, old}}}, \tag{4.11}$$

where the sum runs over all tracks i matched to the subjet j. Similar to the track-assisted mass, the rescaling aims to account for the energy depositions of neutral hadronic components. A large-R jet is then formed using the jets with the rescaled associated tracks and is in the following called TAR jet. The jet substructure variables including the jet mass, m^{TAR}, are then computed using the TAR jets rescaled tracks. A schematic depiction of this procedure is shown in Figure 17. As opposed to the calculation of the track-assisted mass, m^{TA}, the procedure offers the advantage that missing energy from neutral hadrons is corrected on a subjet-by-subjet rather than on a jet-by-jet basis. Hence, local effects are taken into account such that the resolution of the jet substructure variables can be significantly improved. Figure 18 shows the resolution[2] of τ_{21}^{WTA} for W and QCD jets as a

[2]The resolution is defined as the width of the response distribution $\mathcal{R}(p_T^{\mathrm{jet}}) = \tau_{21}^{\mathrm{reco}} - \tau_{21}^{\mathrm{true}}$, where the width is defined as one half of the 50% interquantile range.

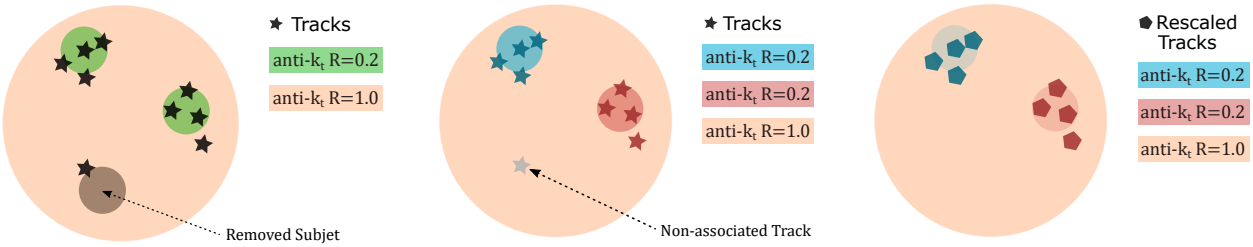


Fig. 17: Schematic depiction of the track-assisted jet reclustering algorithm in the $\eta - \phi$ plane.

function of truth jet p_T for TAR jets and for jets reconstructed using the calorimeter-only based standard reconstruction algorithm. A significant improvement in the resolution is observed for TAR jets, with increasing relative gain at higher jet p_T. TAR jets additionally offer the advantage that calibrations and uncertainties can be propagated from the constituent $R = 0.2$ jets and tracks such that no additional calibration procedures are required. As a result, TAR jets are not limited to AKT-FR reclustering with a fixed size parameter $R = 1.0$, but provide the flexibility of using whichever jet definition that provides the best performance for the specific use case.

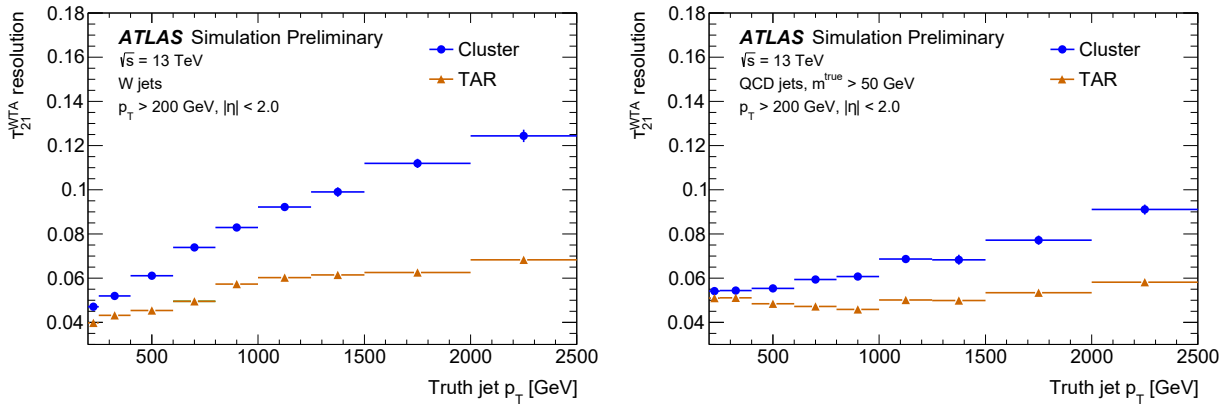


Fig. 18: Resolution of τ_{21}^{WTA} for W (left) and QCD jets (right) as a function of truth jet p_T. The TAR jet reconstruction algorithm is compared to the standard algorithm labelled as "Cluster" [15].

4.5. b-Jet Identification

The identification of jets containing B-hadrons is a task of utmost importance in many physics analyses and is based on several unique features of b-jets. In ATLAS, b-tagging is performed by providing ID tracks as input to b-jet identification algorithms, which individually aim for three different strategies:

- **Impact parameter information (IP2D, IP3D):** Due to the long lifetime of hadrons containing a b-quark ($c\tau \sim 450\,\mu$m), a typical b-jet is characterized by a secondary vertex being displaced from the collision point. A sketch showing the typical displaced secondary vertex in B-hadron decays is shown in Figure 19. Two different taggers, IP2D and IP3D [92], make use of the displacement in order to discriminate between light jets and b-jets. Whereas

IP2D takes advantage of the transverse impact parameter d_0, defined as the closest distance between the track and the primary vertex in the $r-\phi$ plane, IP3D uses the transverse as well as the longitudinal impact parameter, $z_0 \sin\theta$, exploiting the correlations between the two variables.

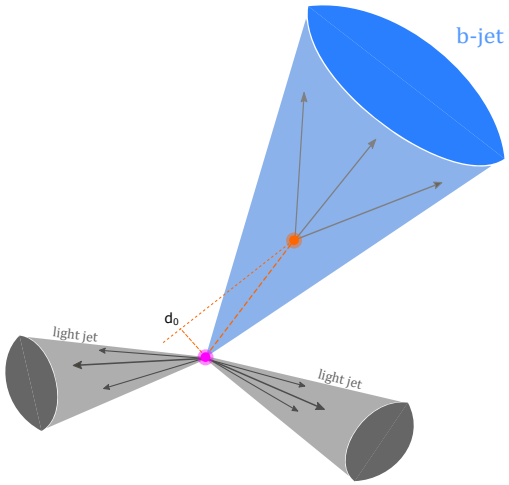


Fig. 19: Sketch showing the typical displaced secondary vertex in B-hadron decays.

- **Secondary vertex information (SV1):** The secondary vertex finding algorithm [93] explicitly reconstructs a displaced vertex within a jet by testing all track pairs within a jet for a two-track vertex hypothesis. From the reconstructed vertex, the number of two-track vertices, the invariant mass of all tracks associated to the vertex and the ratio of the sum of the energies of the tracks in the vertex to the sum of the energies of all tracks in the jet are exploited in order to separate b-jets from c- and light jets.
- **B- to C- hadron decay chain information (JetFitter):** At a parton level, B-hadrons will mostly undergo $b \to c$ transitions and form C-hadrons which will result in a tertiary vertex. The JetFitter [94] algorithm aims to find an approximate B-hadron flight path by using a Kalman filter [95] in order to find a common line on which the $b-$, c- and primary vertices lie. The discrimination between b-, c- and light jets is then based on similar variables as for the SV1 tagger.

The MV2c10 [92] algorithm then combines all of this information into a multivariate technique known as Boosted Decision Tree (BDT). The training of the BDT is performed using small-R CALO jets from $t\bar{t}$ events containing approximately 93% light flavour jets and 7% c-jets, but is also applied to large-R CALO jets and track jets. The light-flavour jet and c-jet rejections versus the b-jet tagging efficiency for the IP3D, SV1, JetFitter and MV2 b-tagging algorithms in simulated $t\bar{t}$ events is shown in Figure 20. For a jet to be considered b-tagged, the output of the BDT is required to be over a certain threshold value. One defines several working points (WP), which correspond to an average b-tagging efficiency in $t\bar{t}$ events. This thesis makes mostly use of the 77% and 70% WPs.

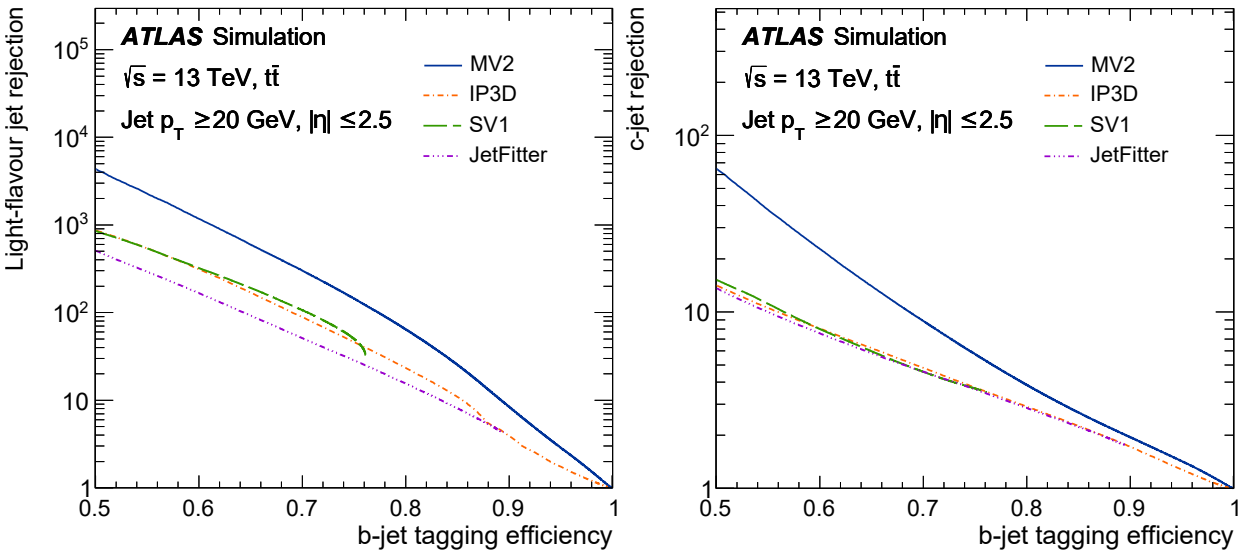


Fig. 20: The light-flavour jet (left) and c-jet (right) rejections versus the b-jet tagging efficiency for the IP3D, SV1, JetFitter and MV2 and b-tagging algorithms in simulated $t\bar{t}$ events [96].

4.6. $H \to b\bar{b}$ Jet Tagging Procedure

As a standard procedure, boosted Higgs bosons decaying into two b-quarks are identified in three steps. First, hadronic decay products are clustered into a single large-R CALO jet. Then, subjets within the large-R CALO jet are reconstructed in order to identify the B-hadrons within the jet. Subjets are then ghost associated to the untrimmed large-R CALO jet. As a baseline procedure, ATLAS uses AKT-FR track jets to reconstruct the subjets within the jet. A cartoon illustrating the baseline subjet reconstruction using FR track jets is shown in Figure 21. As a final step, b-

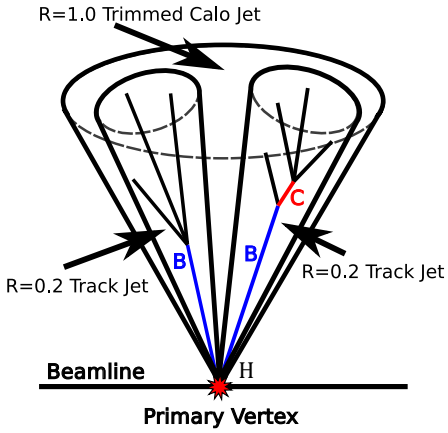


Fig. 21: Cartoon illustrating the baseline subjet reconstruction using fixed radius $R = 0.2$ track jets for $H \to b\bar{b}$ tagging [97].

jets are identified by providing tracks exclusively associated to one of the subjets as input to the MV2c10 algorithm described in Section 4.5. A track is considered to be associated to a subjet with transverse momentum p_T if it satisfies

$$\Delta R(\text{track, subjet}) < 0.239 + e^{-1.220-1.64\times10^{-5}\cdot p_T\,[\text{MeV}]}. \tag{4.12}$$

In the case that a track fulfils this requirement for several subjets, it is matched to the subjet with the smallest angular separation with respect to the track. The ΔR requirement plateaus at a value of 0.239 for high p_T subjets. As an alternative to large-R CALO jets, one can also use TAR jets for which the subjet reconstruction and association procedure is identical to the one described above.

While the identification of $H \to b\bar{b}$ jets using FR track jets works well for Higgs jet transverse momenta of up to $p_T \sim 1\,\text{TeV}$, the hadronisation products from the two b-quarks start to overlap for higher boosts such that the two subjets can not be resolved individually any more [98]. Alternatives to the subjet reconstruction using FR track jets have been proposed in recent years in order to improve the Higgs jet tagging performance at high jet p_T. One obvious approach to tackle this problem is to use AKT-VR instead of AKT-FR subjets. This allows the subjets to take different size parameters such that an individual resolution is possible, even though the hadronisation products begin to overlap. A cartoon illustrating the subjet reconstruction using VR track jets is shown in Figure 22. VR subjets have been shown to perform significantly better at high-p_T and the

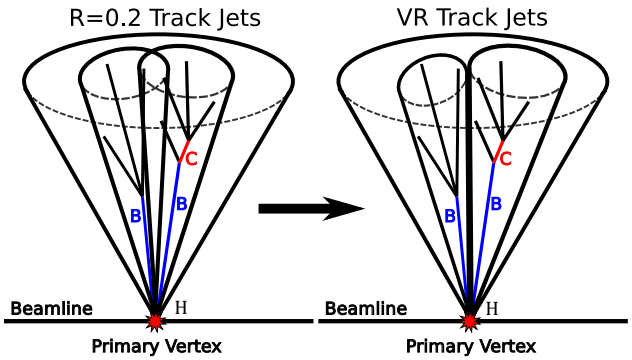


Fig. 22: Cartoon illustrating subjet reconstruction using variable radius track jets [97].

parameters $\rho = 30\,\text{GeV}$, $R_{\text{max}} = 0.4$ and $R_{\text{min}} = 0.06$ have been determined to be most effective for $H \to b\bar{b}$ tagging [97] and hence will be used in the following. The association of tracks to subjets is performed in the same way as for AKT-FR track jets.

Another alternative subjet reconstruction method is given by the exclusive-k_t (ExKT) algorithm. The ExKT algorithm is a simple variation of the KT algorithm and uses the same distance metric defined by Equation 4.2 but only stops clustering constituents when exactly two subjets are obtained or until all d_{ij} exceeds a certain cut-off value. For this, calorimeter subjets are reclustered from the trimmed large-R jet such that calorimeter cell clusters originating from soft processes such as PU and UE activities are effectively removed. The ExKT algorithm does not have an intrinsic parameter R defining the sizes of the subjets, but instead splits the trimmed large-R CALO jet into two parts. An illustration of the subjet reconstruction using exclusive-k_T subjets is shown in Figure 23. This technique is particularly suitable for high-p_T jets as it ensures that exactly two subjets, each expected to contain one B-hadron, are reconstructed. Hence, compared to the reconstruction of a single subjet as a consequence of overlapping hadronisation products in the case of FR-AKT subjets, the ExKT reconstruction provides a higher resemblance to the $H \to b\bar{b}$ topology. The

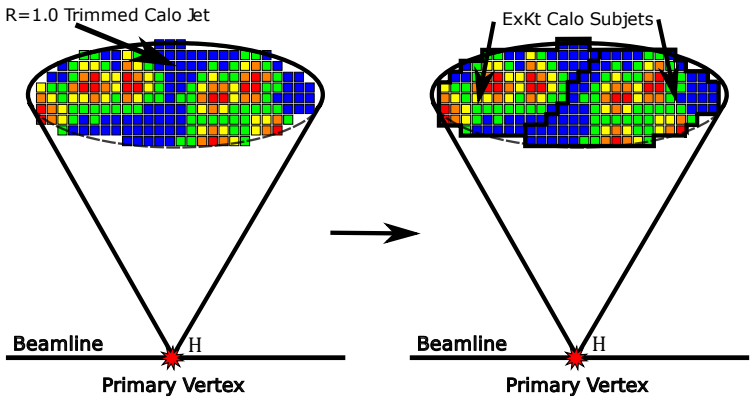


Fig. 23: Cartoon illustrating subjet reconstruction using exclusive-k_T subjets [97].

association of tracks to subjets is performed in the same way as for AKT-FR and AKT-VR track jets.

It has been demonstrated that studying jet substructure in the center-of-mass (CoM) frame of jets is useful to discriminate W/Z [99] from QCD jets. Furthermore, it has been shown that this approach can also be used to identify high-p_T $H \to b\bar{b}$ jets [100]. Thus, as a third alternative subjet reconstruction algorithm, the constituents of the groomed large-R jet are boosted to the jet rest frame defined as the reference frame in which the four-momentum of the large-R CALO jet takes the form $p_\mu = (m_{\text{jet}}, \vec{0})$, where m_{jet} is the invariant mass of the large-R CALO jet. The constituents of the jet are then reclustered with a modification of the CA algorithm to form subjets known as CoM jets. Similar to the ExKT algorithm, the clustering is performed until exactly two subjets are found or a fixed cut-off value d_{ij} is exceeded. In the CoM frame, the tracks are associated to the large-R jet through the procedure described in Ref. [97]. After this matching procedure, the subjets and tracks are boosted back to the laboratory frame. The CoM subjet reconstruction aims to exploit the fact that the subjets should be easier to resolve in the jet rest frame where they are back-to-back. The tracks associated to each of the subjets are then used as inputs for the MV2c10 algorithm in order to identify the presence of B-hadrons. A cartoon illustrating the subjet reconstruction using CoM subjets is shown in Figure 24.

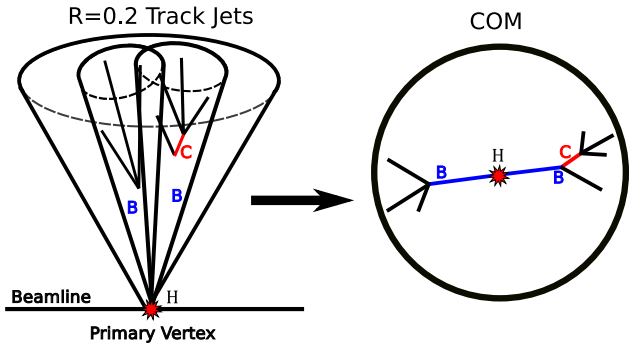


Fig. 24: Cartoon illustrating subjet reconstruction using CoM subjets [97].

Chapter 5

The $b\bar{b}WW^*$ decay channel

The studies presented in this thesis are performed in the resonant $b\bar{b}WW^*$ decay channel. In this channel, a heavy scalar resonance, X, decays into two SM Higgs bosons H, where one Higgs boson decays via $H \to b\bar{b}$ and the other via $H \to WW^*$. With a branching ratio of $\mathcal{BR}(H \to WW^*) \approx 22\%$, the decay into two W bosons is the second most common after the $H \to b\bar{b}$ process with $\mathcal{BR}(H \to b\bar{b}) \approx 58\%$. Altogether, the $4b$ channel offers the highest Higgs boson pair branching ratio with $\sim 33\%$, closely followed by the $bbWW^*$ channel with $\sim 25\%$. Analyses in the $4b$ channel must overcome the overwhelming QCD background in the corresponding final states. In contrast, sensitivity to HH production in analyses in the $bbWW^*$ channel can be gained if the dominant $t\bar{t}$ background can be well-separated from the signal processes. As $\mathcal{BR}(t \to bW) \approx 99\%$, the $t\bar{t}$ process results in a irreducible background with the same final state. The W bosons originating from the single Higgs boson decay consequently decay leptonically $W \to \ell\nu_\ell$ ($\ell = e, \mu$) in $\sim 21\%$ of the cases or hadronically $W \to qq'$ with a branching ratio of $\sim 67\%$. The branching ratios for the several possible decay channels of Higgs boson pair production and the Feynman diagram depicting the resonant Higgs boson pair production in the $b\bar{b}WW^*$ decay mode are shown in Figure 25.

	bb	WW	ττ	ZZ	γγ
bb	33%				
WW	25%	4.6%			
ττ	7.4%	2.5%	0.39%		
ZZ	3.1%	1.2%	0.34%	0.076%	
γγ	0.26%	0.10%	0.029%	0.013%	0.0005%

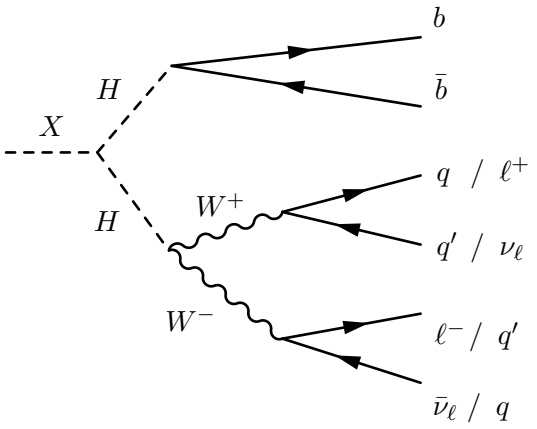


Fig. 25: Branching ratios for the various decay channels of Higgs boson pairs (left) and Feynman diagram depicting the resonant Higgs boson pair production in the $b\bar{b}WW^*$ decay mode (right).

The $b\bar{b}WW^*$ decay channel is further split according to the subsequent decay of the W bosons

- 0-lepton ("0ℓ") final state - both W bosons decay hadronically resulting in the largest branching ratio $\mathcal{BR}(HH \to b\bar{b}WW^* \to b\bar{b}4q) \approx 11.4\%$

- 1-lepton ("1ℓ") final state - one W boson decays leptonically, while the other decays hadronically, resulting in a branching ratio $\mathcal{BR}(HH \to b\bar{b}WW^* \to b\bar{b}\ell\nu_\ell qq') \approx 7.19\%$
- 2-lepton ("2ℓ") final state - both W bosons decay leptonically resulting in the smallest branching ratio $\mathcal{BR}(HH \to b\bar{b}WW^* \to b\bar{b}\ell\nu_\ell\ell\bar{\nu}_\ell) \approx 1.14\%$

While the 0ℓ channel profits from the relatively large branching ratio, the presence of a lepton in the 1ℓ channel is useful to reduce the contamination from QCD contributions.

As mentioned before, this thesis focusses on the resonant Higgs boson pair production, where the two Higgs bosons originate from the decay of a heavy resonance. In boosted topologies, the Higgs bosons are expected to be produced back-to-back and the decay products appear very close to each other in the detector (see Chapter 4). A sketch showing the boosted topology expected in the resonant $b\bar{b}WW^*$ channel is shown in Figure 26.

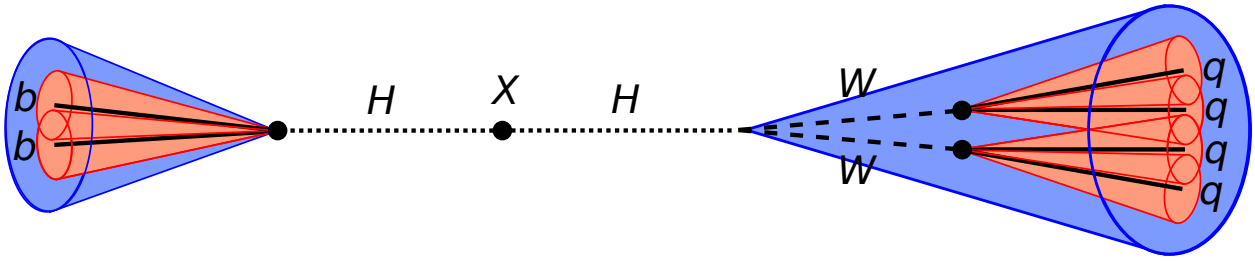


Fig. 26: Sketch showing the boosted topology of the resonant $b\bar{b}WW^*$ Higgs pair production mode in the 0ℓ channel. The Higgs bosons are produced back-to-back and the hadronic decay products appear close to each other in the detector such that the jets seeded by the partons cannot be resolved individually any more.

In order to understand the kinematic properties of the resonant $b\bar{b}WW^*$ decay channel, a truth-level study of the decay kinematics for three different heavy resonance masses $m_X = 1\,\text{TeV}$, $m_X = 2\,\text{TeV}$ and $m_X = 3\,\text{TeV}$ is presented in the following.

5.1. Truth-Level Study of Decay Kinematics

Truth particles from simulation are used to study the kinematic properties of the WW^* and $b\bar{b}$ system. For this study, the 1ℓ final state is used. The results presented are expected to be independent from the type of the subsequent W boson decay.

Figure 27 shows the truth Higgs boson transverse momentum distribution, $p_{T,H}$, for the three mass points. As expected, the Higgs boson momentum distributions are shifted with respect to each other and peak at $p_{T,H} \sim m_X/2$ and then decrease rapidly for higher momentum values. Note that while the figure shows the p_T spectrum of the Higgs boson that then decays to the two bottom quarks, there is no significant difference compared to the p_T spectrum of the Higgs boson that decays to WW^*. It is possible to study the kinematic properties of the on- and off-shell W bosons individually. For this, the on-shell W boson is identified as the truth W boson with mass closest to

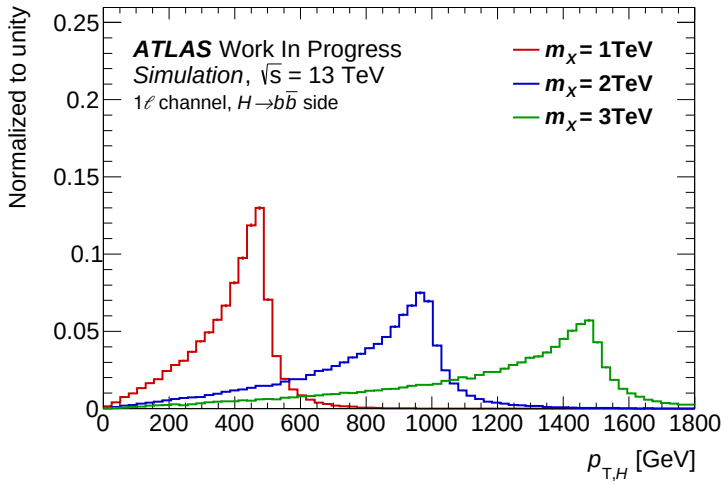
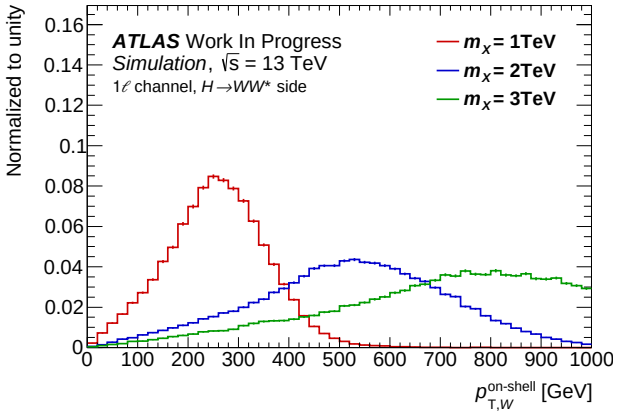
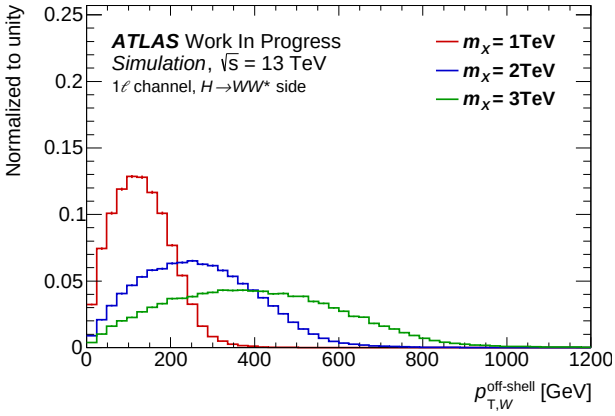


Fig. 27: Truth Higgs boson transverse momentum distributions for several resonance mass points.

the W mass, $m_W \sim 80\,\text{GeV}$, while the other one is identified as the off-shell W boson. Figure 28a shows the transverse momentum spectrum of the on-shell W boson, $p_{T,W}^{\text{on-shell}}$ and Figure 28b shows the transverse momentum spectrum of the off-shell W boson, $p_{T,W}^{\text{off-shell}}$. While both spectra show a broader distribution for higher resonance masses, the off-shell W boson p_T is considerably softer than the on-shell W boson p_T. This an expected feature of the Lorentz-boost which is directly related to the boosted particle's energy. In the case of $H \to WW^*$, the energy of the particle is, to first order, the mass of the particle such that higher mass particles becomes more boosted. The complete W boson p_T spectrum is a superposition of off-shell and on-shell contributions and



(a) On-shell

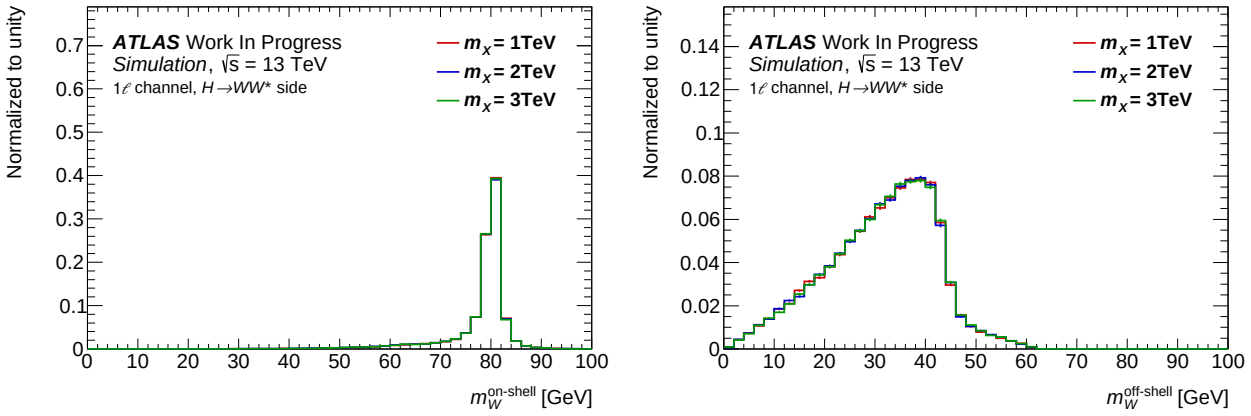


(b) Off-shell

Fig. 28: On- and off-shell truth W boson transverse momentum spectra for several resonance mass points.

is shown in Figure 51 in Appendix A.1. Note that what is shown is the p_T spectrum of the hadronically decaying $W(\to 2q)$ boson, but there is no difference with respect to the p_T spectrum of the leptonically decaying W boson. Similarly, the mass spectrum of the on-shell W boson and the

off-shell W boson are shown in Figure 29a and Figure 29b respectively. Note that the distributions for all three mass points almost overlap completely. While the on-shell distribution shows the expected mass peak at the W boson mass $m_W \approx 80\,\text{GeV}$, Figure 29b exhibits the expected off-shell invariant mass shape resulting from virtual W bosons. Figure 51 in Appendix A.1 shows the mass spectrum $m_{W(\to 2q)}$ of the hadronically decaying W boson. Both, on-shell and off-shell W boson contributions to the mass spectrum are clearly visible.



(a) On-shell (b) Off-shell

Fig. 29: On- and off-shell truth W boson invariant mass spectrum for several resonance mass points.

Angular distributions

Geometric distributions such as the angular distribution between the partons have been studied as well. Figure 30 shows the angular separation, $\Delta R(H,H)$, between the two Higgs bosons that result from the decay of the heavy resonance. As expected, the Higgs bosons are produced back-to-back, i.e. the angular distribution peaks at $\Delta R \approx \pi$ with a tail to larger values.

Of more interest is the angular separation of the Higgs boson decay products. The angular separation, $\Delta R(W,W^*)$, between the two W bosons as well as the angular separation, $\Delta R(b,\bar{b})$, between the two b-quarks is shown in Figure 31a and Figure 31b respectively. For higher resonance masses the Higgs boson is more boosted, such that the angular separation distribution is shifted to lower values. While for the $m_X = 3\,\text{TeV}$ sample, the angular distribution between the W bosons results in a relatively narrow peak with a maximum at $\Delta R(W,W^*) \approx 0.06$, the $m_X = 1\,\text{TeV}$ sample results in a very broad spectrum peaking at $\Delta R(W,W^*) \approx 0.2$. The b-quarks are rather well-separated for a $m_X = 1\,\text{TeV}$ resonance mass, resulting in a broad spectrum peaking at $\Delta R(b,b) \approx 0.55$. For the $m_X = 3\,\text{TeV}$ resonance mass, most events are below $\Delta R(b,\bar{b}) = 0.2$, which represents a critical value as the hadronic decay products of b-jets are no longer resolvable with standard $R = 0.2$ track jets. On the other hand, a notable fraction of b-quarks have $\Delta R(b,\bar{b}) > 1$ for the $m_X = 1\,\text{TeV}$ resonance mass such that they cannot be collected using standard large-R CALO jets.

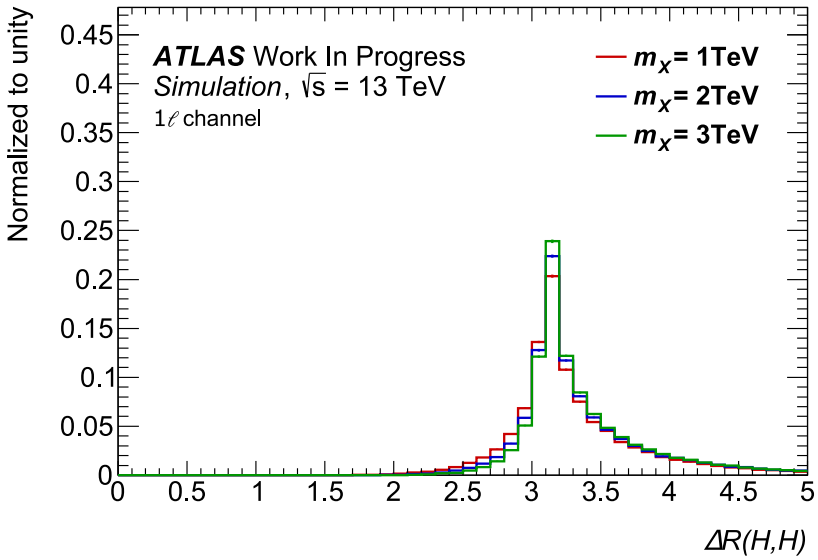
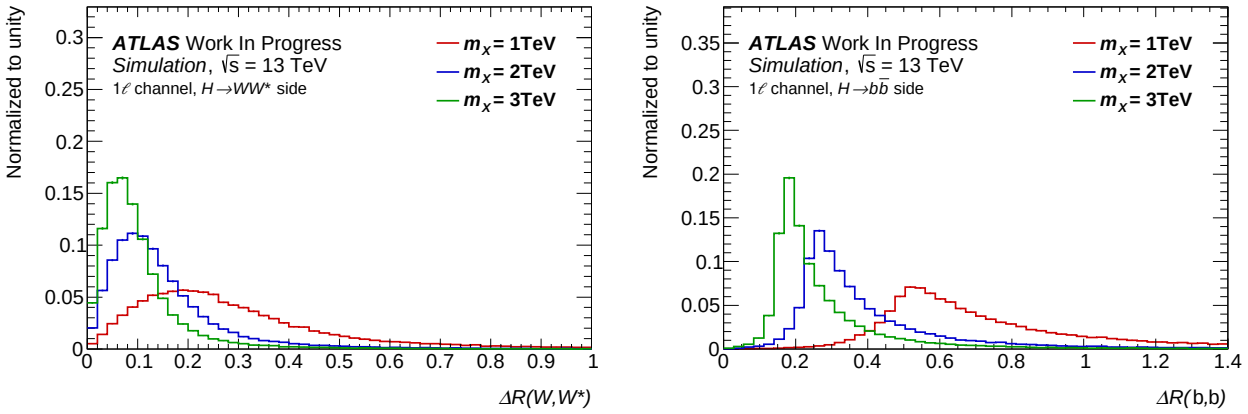


Fig. 30: The distribution of the angular separation between the two truth Higgs bosons for several resonance mass points.



(a) W bosons (b) b-quarks

Fig. 31: The distribution of the angular separation between the two truth W bosons and between the two truth b-quarks resulting from the boosted Higgs boson decay for several resonance mass points.

Chapter 6

Performance Studies of $H \to b\bar{b}$ Tagging

In the following, performance studies of boosted $H \to b\bar{b}$ tagging techniques are presented. As a general approach, a $H \to b\bar{b}$ jet tagger can be constructed using three steps:

- b-tagging of the subjets
- Jet mass window cut around the Higgs boson mass peak
- Cuts on jet substructure variables

These steps aim to exploit the heavy flavour content, the mass and the internal structure of the reconstructed Higgs jet to discriminate from QCD jets and hadronic top decays. The metrics used to study the performance of such a Higgs tagger are the

$$\text{Higgs jet efficiency} = \frac{N(b\text{-tag} \mid \text{Higgs jet})}{N(\text{Higgs jet})}, \tag{6.1}$$

where N is the number of the jets, and the QCD and top jet rejections defined as the inverse QCD and top jet efficiencies. As a general measure of tagging performance, Receiver Operatic Characteristics (ROC) curves, which show the background rejection as a function of the signal efficiency, are used. ROC curves have the advantage that they show the performance of a tagger across all signal efficiencies and are not dependent on the used WP. In this thesis, Higgs, top and QCD jets are reconstructed using standard large-R CALO jets or TAR jets, both of which are required to satisfy $p_T > 250\,\text{GeV}$ and $|\eta| < 2.0$. Furthermore, the subjets are required to have $p_T > 10\,\text{GeV}$ and $|\eta| < 2.5$. FR and VR track jets are moreover required to have at least two track constituents.

A Higgs jet is defined as the reconstructed jet in simulated $b\bar{b}WW^*$ events which is closest to the large-R Higgs truth jet. The large-R Higgs truth jet is determined by labelling its flavour by spatially matching the jet to generator-level B-hadrons with $p_T > 5\,\text{GeV}$ which are within $\Delta R < 1.0$ of the jet axis. In case of ambiguities, the hadron is only matched to the closest jet. A Higgs truth jet is required to be double B-labelled and fulfil the identical p_T and η requirement as the reconstructed Higgs jet. Events which do not contain a Higgs truth jet are discarded. Top and QCD jets are defined as the p_T-leading reconstructed large-R CALO or TAR jet in $t\bar{t}$ and QCD di-jet events, respectively.

The MC signal and $t\bar{t}$ events have been weighted according to the procedure described in Appendix C to yield the physical predictions expected for the 2015-2017 data taking period. QCD di-jet events have been weighted to give a physical, smoothly falling p_T spectrum (see Appendix C).

For QCD, only relative event numbers are discussed, as a full QCD background estimate, which was not available during the time of writing, is required to compute absolute event yields. For the production of a $m_X = 1\,\text{TeV}$ resonance mass, a cross section of 15 pb was assumed, while the cross section for the production of a $m_X = 2$ and 3 TeV heavy resonance mass was assumed to be 2 pb. The assumed cross sections roughly correspond to the current experimental limits on resonant Higgs-boson pair production for a spin-0 scalar which are presented in Section 2.4. For this study, the 0ℓ final state is used, but the results presented are expected to be independent from the type of the subsequent W boson decay.

For the b-tagging of the subjets, the 77% WP is used. Three categories have been studied:

- Double b-tagging ("2btag"): the two p_T-leading subjets are required to pass the b-tagging requirement
- Inclusive single b-tagging ("≥1btag"): at least one of the two p_T-leading subjets are required to pass the b-tagging requirement
- Exclusive single b-tagging ("1btag"): exactly one of the two p_T-leading subjets are required to pass the b-tagging requirement

In Section 6.1, a brief truth-level study aiming to mimic the standard tagging procedure for AKT-FR track jets and to motivate the use of alternative subjet reconstruction algorithms is presented. In Section 6.2, the performance of $H \to b\bar{b}$ tagging using AKT-FR, AKT-VR, CoM and ExKT subjets with standard large-R CALO jets is explored. In Section 6.3, the use of TAR jets for $H \to b\bar{b}$ tagging is studied for the first time. The performance of various TAR jet definitions with VR track jets is presented. Furthermore, different TAR jet mass window cuts are applied and substructure distributions are used to further increase the discrimination against QCD and top jets.

6.1. Truth $H \to b\bar{b}$ Labelling Studies

In order to understand and motivate the need for advanced boosted $H \to b\bar{b}$ tagging procedures, it is useful to mimic the baseline tagging employed by ATLAS at truth level. For this, stable truth particles after hadronisation are used as input to the respective jet algorithms to create truth jets (see Section 4.2). The truth particles are clustered into truth large-R AKT-FR jets (in the following just "large-R jets") and to AKT-FR ($R = 0.2$) jets, aiming to mimic the large-R CALO jets and the AKT-FR track jets used for the standard boosted $H \to b\bar{b}$ tagging procedure at detector level. The truth large-R jets undergo the same grooming procedure as the reconstructed jets. In the following, AKT FR (R=0.2) jets which lie within $\Delta R = 0.5$ of the closest large-R jet, are called subjets of the respective large-R jet. The $H \to b\bar{b}$ large-R jets can then be selected by requiring them to be double B-labelled. However, in order to mimic the tagging procedure at reconstruction level, this is compared to selecting the large-R jet by requiring the association with exactly two subjets which each have been individually B-labelled ("B-associated"). The labelling follows the same procedure described above, with the difference that B-hadrons are matched to the subjets

if they lie within $\Delta R < 0.2$ of the subjet axis. Note that while this study uses the identical jet reconstruction algorithms as described in Section 4.1, the inputs are different and only aim to mimic the tagging procedures at truth level as precisely as possible.

Figure 32 shows the invariant mass of the double B-labelled large-R jet, $m_{b\bar{b}}^{\mathrm{LRJ}}$, and the difference, $m_{b\bar{b}}^{\mathrm{Parton}} - m_{b\bar{b}}^{\mathrm{LRJ}}$, between the invariant $b\bar{b}$ mass at parton level, $m_{b\bar{b}}^{\mathrm{Parton}}$, and the large-$R$ jet mass. The $m_{b\bar{b}}^{\mathrm{LRJ}}$ distribution shows a peak around the Higgs boson mass, a slow decrease towards low large-R jet masses as well as a very sharp decrease towards higher masses. The sharp drop is an expected feature and is the result of the grooming procedure, which removes soft components from PU and UE contributions. As a result, the mass distribution of the large-R jet tends to slightly lower values compared to the invariant mass at parton level.

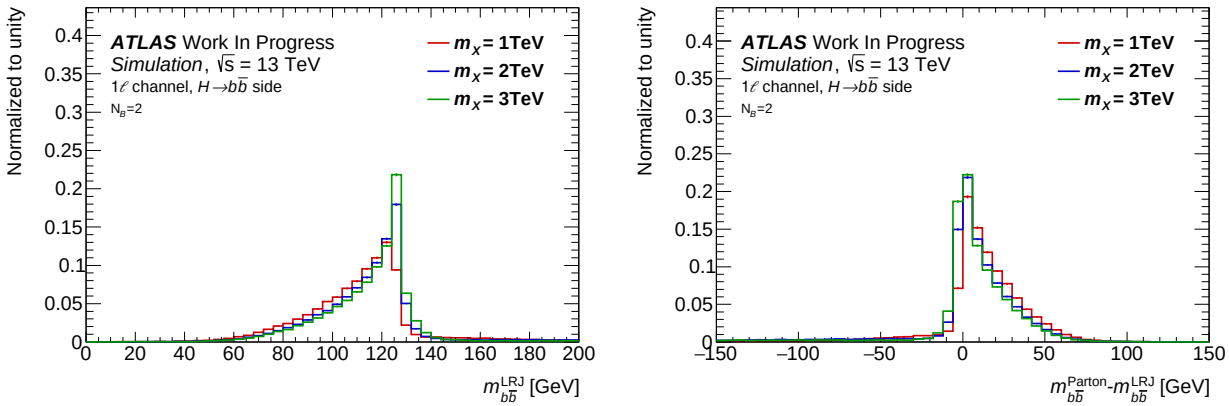


Fig. 32: Invariant mass spectrum of the double B-labelled large-R jet (left) and difference between the invariant mass at parton level and the invariant mass of the double B-labelled large-R jet (right).

In Figure 33, the angular separation, $\Delta R(\mathrm{SJ}, \mathrm{SJ})$, between both B-labelled subjets (SJs) of the large-R jet is shown for the several mass points. As expected, the shape compares well to the angular separation seen at parton level (see Figure 31b). The distributions show a hard cut-off in the spectrum at the critical value $\Delta R = 0.2$, which corresponds to the size parameter of the subjet reconstruction algorithm. The inputs to the subjet reconstruction algorithm start to overlap such that one instead of two subjets is reconstructed ("subjet merging"). While this affects a significant number of events in the case of a $m_X = 3\,\mathrm{TeV}$ heavy resonance, the Higgs bosons originating from lower resonance masses are less boosted and therefore these events seldomly have an angular separation below the threshold of $\Delta R = 0.2$. As a consequence, improvements in subjet reconstruction methods are expected to primarily benefit the $H \to b\bar{b}$ tagging for resonance masses with $m_X \gtrsim 3\,\mathrm{TeV}$.

The effect of subjet merging is also observable in the p_T spectrum of the large-R jet. Figure 34 shows the jet transverse momentum, $p_{\mathrm{T},b\bar{b}}^{\mathrm{LRJ}}$, of the B-associated large-R truth jets and of large-R truth jets which are selected by requiring one or two subjets with one or two B-labels each, respectively. While the distribution for the B-associated jets remain mostly unaffected for the 1 TeV and 2 TeV

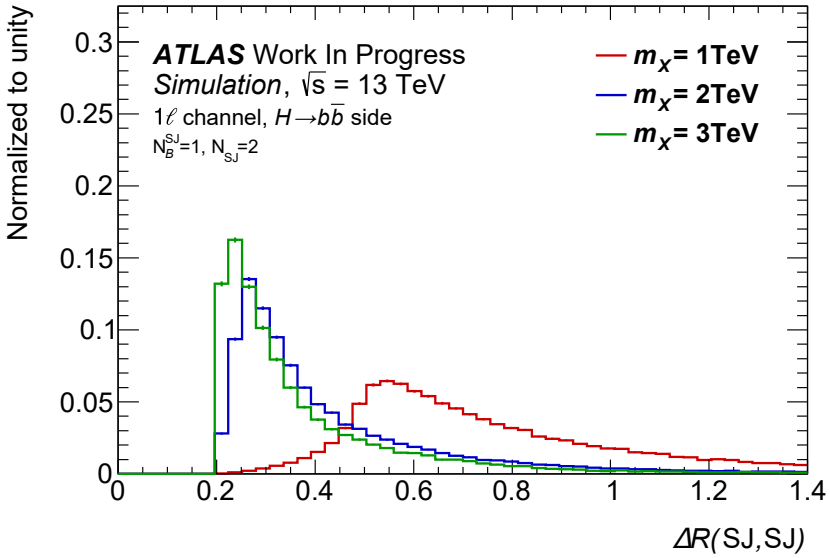


Fig. 33: Angular separation between the two B-labelled subjets which have been associated to the large-R jet. A hard cut-off is observed at $\Delta R = 0.2$.

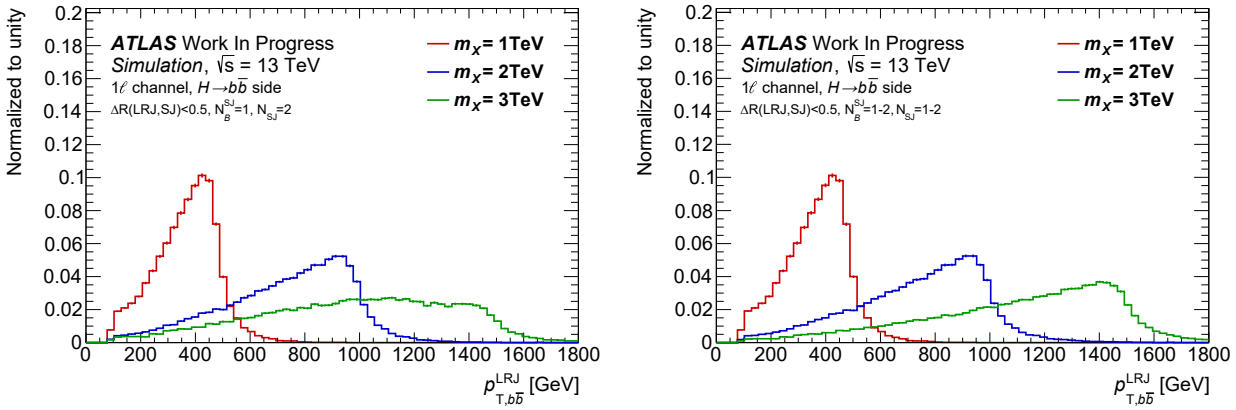


Fig. 34: Transverse momenta of the B-associated large-R jets (left) and for large-R jets which are selected by requiring the association to one or two B-labelled subjets (right). The original p_T spectrum for double B-labelled large-R jets is recovered.

resonance masses, events in the high-p_T region are lost in the $m_X = 3\,\text{TeV}$ case, which causes a shape distortion in the p_T spectrum. When including the case of only one B-labelled subjet being present, the original p_T spectrum obtained from double B-labelled large-R jets (see Figure 35) is recovered.

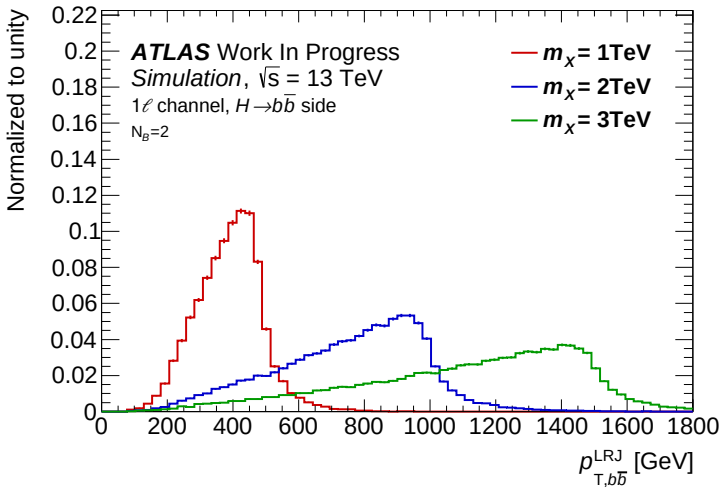


Fig. 35: Transverse momentum of the double B-labelled large-R jets.

6.2. Performance of Subjet Reconstruction Techniques

In this section, the performance of the standard $H \to b\bar{b}$ tagging procedure employed by ATLAS is compared to the use of the alternative subjet reconstruction techniques presented in Section 4.6. Standard large-R CALO jets are used for the reconstruction of the $H \to b\bar{b}$ and top jets. As a baseline, the use of FR track jets is compared to the use of VR track jets as well as CoM and ExKT subjets for b-tagging. Figure 36 shows the Higgs jet efficiencies in the double b-tag category for the various subjet reconstruction algorithms as a function of the transverse momentum, $p_{T,\text{Jet}}$, and as a function of the pseudorapidity, η_{Jet}, of the reconstructed Higgs jet for the three mass points $m_X = 1$, 2 and 3 TeV. As the p_T spectrum of the Higgs boson peaks at $p_{T,H} \approx m_X/2$ and then decays rapidly for higher mass values (see Figure 27), the Higgs jet efficiencies as a function of the jet p_T start to become statistically limited for $p_{T,\text{Jet}} \geq m_X/2$. In the low-$p_T$ region up until $\sim$1 TeV a Higgs jet efficiency between $\sim$0.4-0.6 is observed, where FR jets give a slightly higher efficiency. However, in the high-p_T region, the efficiency for FR jets drops significantly to values even below 0.1 at $p_{T,\text{Jet}} \approx 1.8$ TeV. VR jets also show a small decrease in efficiency in the high-p_T region, but allow the recovery of most of the efficiency loss seen with FR jets. The CoM and ExKT subjets also show some improvements in the high-p_T region, but to a lesser extent. ExKT subjets seem to give marginally higher efficiencies compared to CoM subjets. The loss of efficiency seen in the p_T spectrum of the jet translates into a dip for central values in the η distribution of the Higgs jet. While the 1 TeV resonance mass results in a flat η spectrum with little hierarchy between the subjets, the efficiencies exhibit a large drop around $\eta \approx 0$ for the 3 TeV resonance mass. Only VR jets are able to almost completely preserve the flat shape seen for less boosted heavy resonances with $m_X = 1$ TeV. The Higgs jet efficiencies in the inclusive single b-tagging category are shown in Figure 52 in Appendix A.2 and show a mostly flat distribution across all mass points and subjet reconstruction algorithms at a $\sim$90 % efficiency. The overall Higgs jet efficiencies and the efficiencies in the high-p_T region in the double and inclusive single b-tag category are shown in Table 7. These

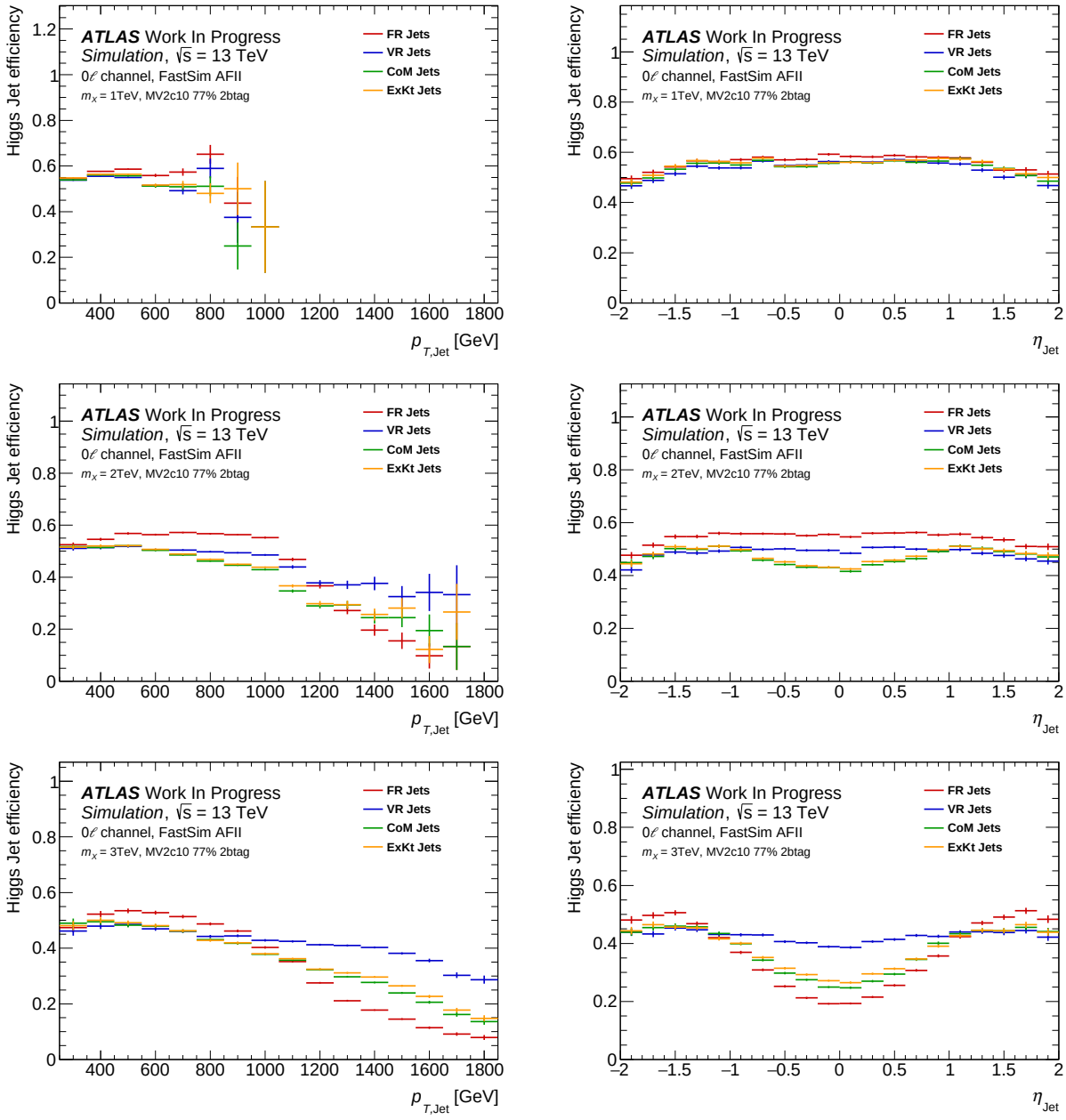


Fig. 36: Higgs jet efficiencies as a function of p_T (left) and η (right) in the double b-tag category for the various subjet reconstruction algorithms for the three mass points $m_X = 1$, 2 and 3 TeV.

measurements are compatible with previous observations [97] and the truth-level study presented in Section 6.1. Subjet merging effects start to gain relevance for boosted jets with a p_T in the order of ~ 1 TeV, such that the use of advanced subjet reconstruction techniques can improve the $H \to b\bar{b}$ tagging for resonance masses of $m_X \gtrsim 3$ TeV. Compared to FR track jets, VR track jets show the best improvement with an efficiency gain of up to $\sim 77\%$ in the high-p_T region and an overall efficiency gain of $\sim 37\%$ in the case of a 3 TeV resonance mass. Only marginal losses are observed for lower resonance masses.

Jet Type	Signal mass point m_X				
	1 TeV	2 TeV		3 TeV	
	Overall	Overall	High-p_T	Overall	High-p_T
FR	0.57 (0.94)	0.55 (0.94)	0.43 (0.89)	0.30 (0.89)	0.22 (0.87)
VR	0.55 (0.93)	0.49 (0.92)	0.42 (0.87)	0.41 (0.89)	0.39 (0.87)
ExKT	0.56 (0.93)	0.47 (0.90)	0.35 (0.85)	0.35 (0.85)	0.30 (0.82)
CoM	0.55 (0.92)	0.46 (0.90)	0.33 (0.84)	0.34 (0.85)	0.28 (0.81)

Tab. 7: Higgs jet efficiences for VR, FR, ExKT and CoM subjets in the double (inclusive single) b-tag category for the 77% WP. Efficiencies are shown for the standard jet selection ("Overall") and for the high-p_T region where $p_{T,\text{Jet}} > 1\,\text{TeV}$.

In addition to the Higgs jet efficiencies, it is also important to study the background rejection for the various subjet reconstruction algorithms, in particular for the dominant $t\bar{t}$ background. In Figure 37, the top jet rejection as a function of the top jet p_T is shown for the various subjet reconstruction algorithms. The top jet rejection is nearly constant throughout the p_T bins for all

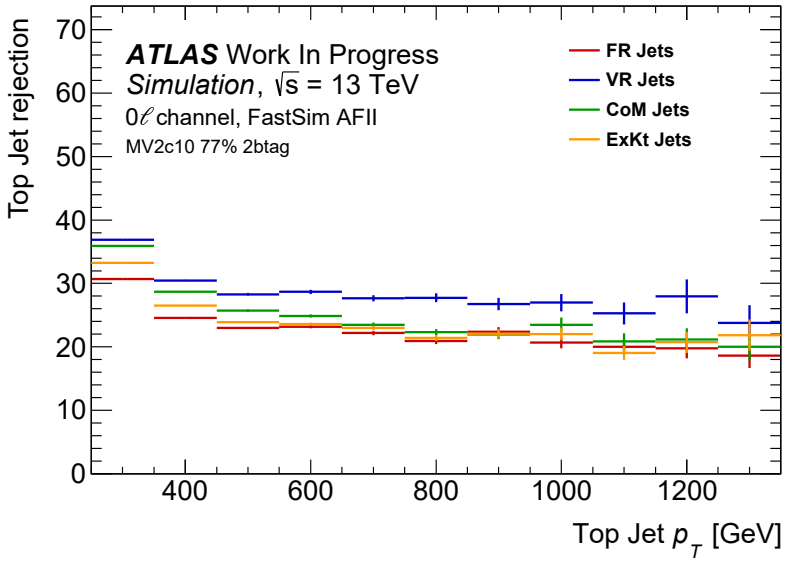


Fig. 37: Top jet rejection as a function of the top jet p_T in the double b-tag category for the various subjet reconstruction algorithms.

subjet reconstruction algorithms. While FR, CoM and ExKT jets exhibit a very similar rejection, the top jet rejection is best for VR track jets. The top jet rejection as a function of the top jet p_T in the single inclusive b-tag category is shown in Figure 53 in Appendix A.2 and exhibits a significantly worse top jet rejection of ~ 2 throughout the p_T bins. As the b-quarks resulting from the two individual top decays are expected to have a higher angular separation compared to those resulting from Higgs decays, it is unlikely that both will be reconstructed within subjets which are associated to one single large-R CALO jet. In contrast, $t \to bW$ jets are expected to contain

exactly one B-hadron such that the discrimination of top jets from Higgs jets becomes significantly more difficult in the (inclusive) single b-tag category.

The ROC curves comparing the performance of various subjet reconstruction techniques for a heavy resonance mass $m_X = 3\,\text{TeV}$ and the ROC curves comparing the FR and VR track jet performance for the various mass points are shown in Figure 38. As expected, VR track jets perform significantly better compared to FR track jets, while the CoM and ExKT Jets provide some improvement. As a result of subjet merging in the case of a 3 TeV resonance, FR track jets reach a maximum signal efficiency of ~ 0.7, while CoM and ExKT jets reach ~ 0.8 and VR jets are not affected at all. For lower resonance masses, all subjet reconstruction algorithms reach the full signal efficiency and only small differences are seen in the performance. The ROC curves comparing the performance of the four subjet reconstruction algorithms for resonances with $m_X = 1$ and 2 TeV are shown in Figure 54 in Appendix A.2. It is noteworthy to mention that VR track jets can overlap ("jet-in-jet topology"). The interplay of overlapping track jets with the track association for b-tagging is questionable and currently under investigation. Possible solutions to this problem, such as vetoing highly collimated VR track jets, could reduce the efficiencies but are outside of the scope of this thesis.

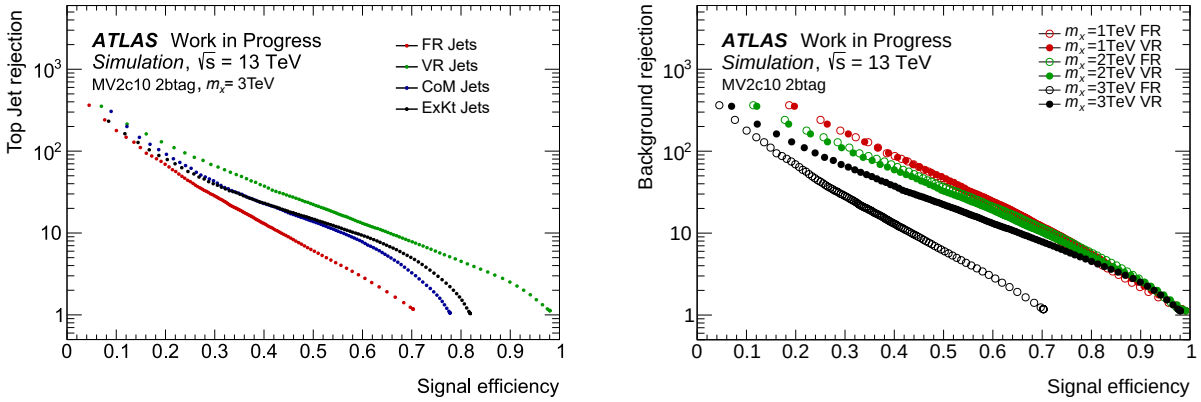


Fig. 38: ROC curves comparing the performance of various subjet reconstruction techniques for a heavy resonance mass $m_X = 3\,\text{TeV}$ (left) and ROC curves comparing FR and VR track jet performance for the $m_X = 1$, 2 and 3 TeV resonance masses (right).

6.3. Performance of Track-Assisted Reclustered Jets

In the following, $H \to b\bar{b}$ tagging performance studies using TAR jets instead of large-R CALO jets for the reconstruction of the $H \to b\bar{b}$ jet are presented. Since in Section 6.2 VR track jets have been shown to offer the best performance for standard large-R CALO jets, VR track jets are used in the following for the flavour tagging. The impact on $H \to b\bar{b}$ tagging is studied for a variety of TAR jet definitions. The reclustering is performed using the AKT-FR, AKT-VR as well as the CA-FR and CA-VR algorithms. For the FR algorithms, the size parameter R is varied in steps of 0.25 from $R = 0.50$ to $R = 1.50$. For the VR algorithms, the energy scale ρ is varied in steps of

50 GeV from $\rho = 150\,\text{GeV}$ to $\rho = 350\,\text{GeV}$. The minimum size parameter is set to $R_{\text{min}} = 0.4$ and the maximum size parameter to $R_{\text{max}} = 1.6$. No significant differences are observed between AKT and CA reclustering, such that for the sake of brevity, most results presented in the following make use of the AKT algorithm.

Figure 39 shows the Higgs jet efficiencies as a function of the TAR jet p_T in the double b-tag category for AKT-FR and AKT-VR TAR jets for the $m_X = 3\,\text{TeV}$ resonance mass. The efficiencies for all resonance masses, including the $m_X = 1\,\text{TeV}$ and $m_X = 2\,\text{TeV}$ resonance masses are shown in Figure 55 in Appendix A.3. The efficiencies for AKT-FR ($R = 1.0$) TAR jets compare well

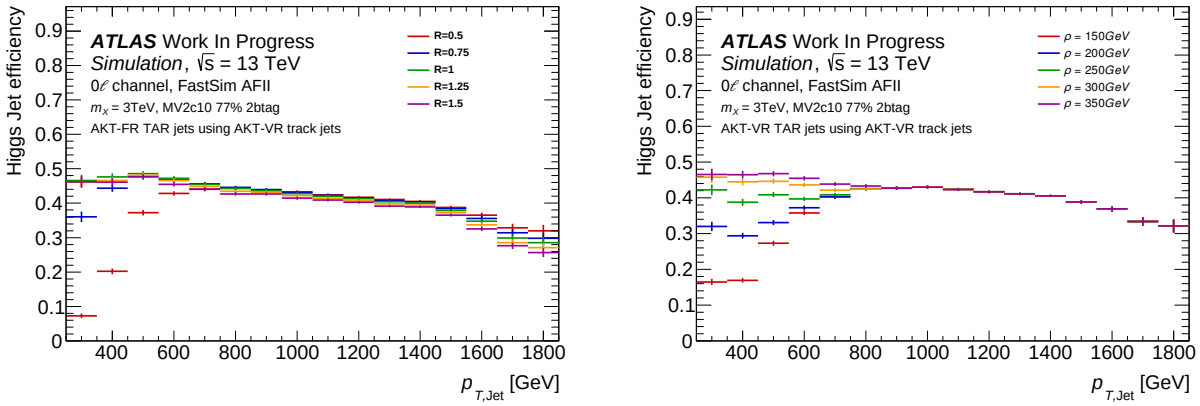


Fig. 39: Higgs jet efficiencies as a function of the TAR jet p_T in the double b-tag category for AKT-FR (left) and AKT-VR (right) TAR jets for the $m_X = 3\,\text{TeV}$ resonance mass. For AKT-FR TAR jets, the size parameter is varied, while the energy scale is varied for AKT-VR TAR jets.

to the results shown for standard large-R CALO jets and VR track jets in Section 6.2. For both AKT-FR and AKT-VR TAR jets, a gradual decrease in efficiency is seen in the low-p_T region for smaller size parameters and energy scales. For higher TAR jet p_T, the efficiencies converge such that no difference is observed between the different parameter values. In the low-p_T region, the Higgs bosons are less boosted such that the decay products appear relatively widely distributed. Small jet sizes fail to collect both B-hadrons from the $H \to b\bar{b}$ jet such that the tagging performance decreases rapidly in this region. According to Equation 4.1, the decay products of a Higgs boson with $p_T \geq 500\,\text{GeV}$ have a mean angular separation of ~ 0.5. Hence, $R = 0.5$ jets begin being relatively efficient at a Higgs jet p_T of $\geq 500\,\text{GeV}$. After exceeding a jet p_T threshold of $\sim 700\,\text{GeV}$, no significant differences between the parameter values are observed for AKT-FR TAR jets. For AKT-VR TAR jets a similar behaviour is observed. The Higgs jet efficiencies as a function of the TAR jet p_T in the single inclusive and exclusive b-tag category for AKT-FR and AKT-VR TAR jets for the various mass points are shown in Figure 56 and Figure 57 in Appendix A.3, respectively. As expected, the inclusive single b-tag category exhibits high efficiencies throughout the p_T bins with a minor decrease in the very low-p_T region for the lower values of the parameters. As the exclusive single b-tag category is orthogonal to the double b-tag category, it exhibits the opposite tendency, i.e. smaller values of the parameters perform best at low-p_T, while the same convergence is seen in the high-p_T region. The total Higgs jet efficiencies for the various TAR jet definitions and signal

mass points for the 77% (70%) WP are shown in Table 8. As mentioned earlier, only marginal

TAR jet definition			Signal mass point m_X		
Type		R | ρ [GeV]	1 TeV	2 TeV	3 TeV
AKT	FR	0.50	0.14 (0.11)	0.45 (0.36)	0.41 (0.31)
		0.75	0.47 (0.40)	0.49 (0.40)	0.42 (0.32)
		1.00	0.55 (0.46)	0.50 (0.40)	0.42 (0.32)
		1.25	0.54 (0.46)	0.49 (0.39)	0.41 (0.32)
		1.50	0.53 (0.45)	0.48 (0.38)	0.40 (0.31)
	VR	150	0.24 (0.20)	0.41 (0.33)	0.40 (0.31)
		200	0.39 (0.33)	0.44 (0.35)	0.41 (0.31)
		250	0.48 (0.41)	0.46 (0.37)	0.41 (0.32)
		300	0.52 (0.44)	0.48 (0.38)	0.42 (0.32)
		350	0.54 (0.46)	0.49 (0.39)	0.42 (0.32)
CA	FR	0.50	0.13 (0.11)	0.45 (0.36)	0.41 (0.31)
		0.75	0.47 (0.40)	0.50 (0.40)	0.42 (0.32)
		1.00	0.55 (0.46)	0.50 (0.40)	0.42 (0.32)
		1.25	0.54 (0.46)	0.49 (0.39)	0.41 (0.31)
		1.50	0.53 (0.45)	0.48 (0.38)	0.40 (0.31)
	VR	150	0.24 (0.20)	0.41 (0.33)	0.40 (0.30)
		200	0.39 (0.33)	0.43 (0.35)	0.40 (0.31)
		250	0.48 (0.40)	0.46 (0.37)	0.41 (0.31)
		300	0.52 (0.44)	0.47 (0.38)	0.42 (0.32)
		350	0.54 (0.45)	0.49 (0.39)	0.42 (0.32)

Tab. 8: Higgs jet efficiencies in the double b-tag category for the various TAR jet definitions and signal mass points for the 77% (70%) WP.

differences are observed between AKT and CA reclustering. Efficiencies remain almost unchanged for the different parameters for the $m_X = 2\,\text{TeV}$ and $m_X = 3\,\text{TeV}$ resonance masses. A large drop in efficiency is seen for the smallest parameters $R = 0.5$ and $\rho = 150\,\text{GeV}$. Up to a factor of ~ 3.4 in efficiency is lost with respect to the second lowest parameter values.

The corresponding top jet rejections and QCD jet rejections as a function of the TAR jet p_T in the double b-tag category for AKT-FR and AKT-VR TAR jets are shown in Figure 40. The hierarchy seen for the Higgs jet efficiencies is reversed for the background rejection. While in the low-p_T region, smaller size parameters and energy scales lead to lower efficiencies, the background rejection is enhanced. In particular, a jet size of $R = 0.5$ increases the top jet rejection by a factor of up to 3.5 in the low-p_T region. The probability of the two B-hadrons being contained in one single top jet is very small, such that at least one light track jet needs to be mistagged to get a double b-tagged TAR jet. Reducing the radius of the TAR jet, lowers the probability to find one B-hadron within the p_T-leading TAR jet, such that two track jets need to be mistagged and the rejection increases. In the high-p_T region, the B-hadron can be collected even by smaller jet sizes such that a similar convergence for the rejection is seen as for the efficiencies.

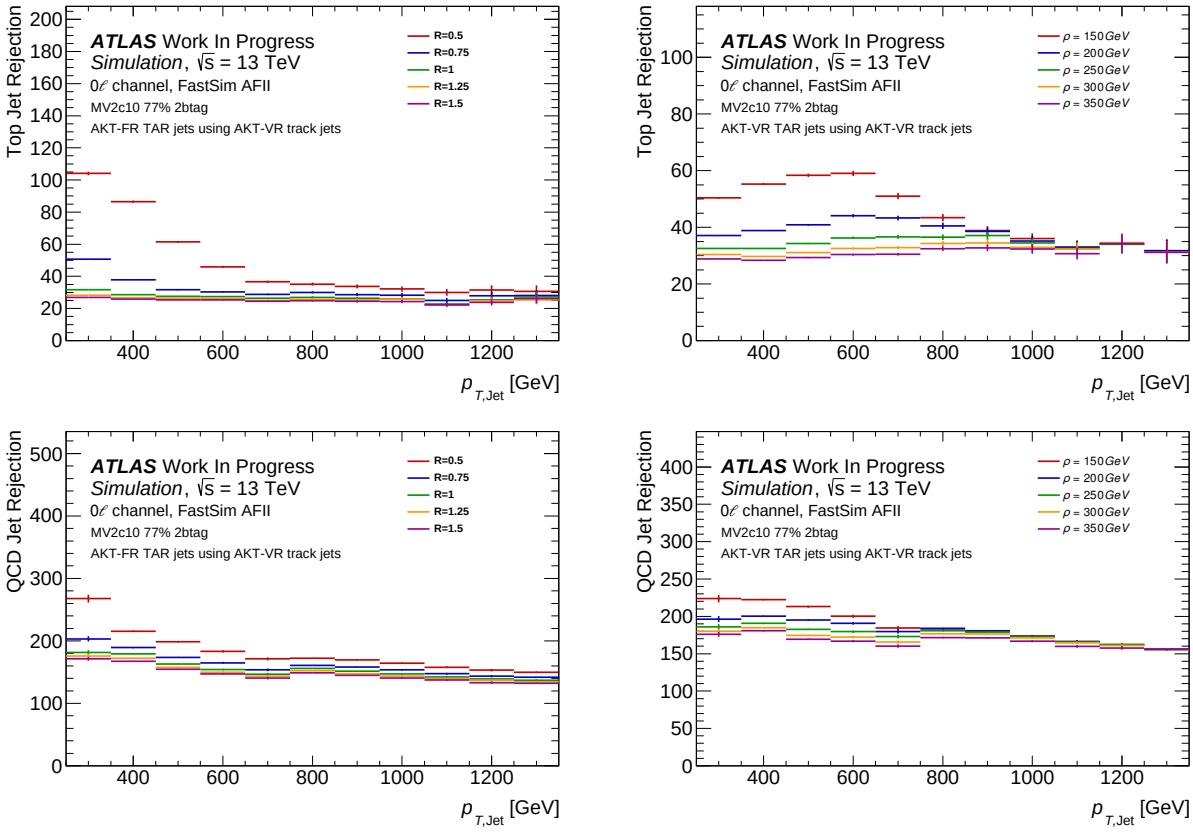


Fig. 40: Top jet rejection (top) and QCD rejection (bottom) as a function of the TAR jet p_T in the double b-tag category for AKT-FR (left) and AKT-VR (right) TAR jets. For AKT-FR TAR jets, the size parameter is varied, while the energy scale is varied for AKT-VR TAR jets.

Figure 41 shows the ROC curves comparing the Higgs jet tagging performance against top jets in the double b-tagging category for AKT-FR TAR jets and AKT-VR TAR jets for the various mass points and jet parameters. In the case of a $m_X = 3\,\text{TeV}$ resonance mass, lower values of R and ρ perform best throughout all signal efficiencies. For all parameters, the full efficiency is reached. In the case of a $m_X = 1\,\text{TeV}$ resonance mass, a large drop in performance is seen for $R = 0.5$ and $\rho = 150\,\text{GeV}$. As a result of the loss in efficiency in the low-p_T region seen in Figure 39, the full signal efficiency is not reached. In this case, the $R = 1.0$ and $\rho = 350\,\text{GeV}$ values perform best. In the intermediate case of a $m_X = 2\,\text{TeV}$ resonance mass, lower values of R and ρ seem to perform best at high signal efficiencies, while larger parameter values are beneficial for high signal efficiencies. Due to the massive drop in Higgs jet tagging performance for $m_X = 1\,\text{TeV}$ resonance masses, the parameters $R = 0.5$ and $\rho = 150\,\text{GeV}$ should not be used. A very similar behaviour is observed for QCD jets. Figure 60 in Appendix A.3 shows the ROC curves comparing the Higgs jet tagging performance against QCD jets in the double b-tag category for AKT-FR TAR jets and AKT-VR TAR jets for the various mass points and jet parameters.

Summarising, only marginal differences in performance between AKT and CA reclustering are observed for double b-tagged TAR jets. The AKT-FR and AKT-VR algorithms provide similar

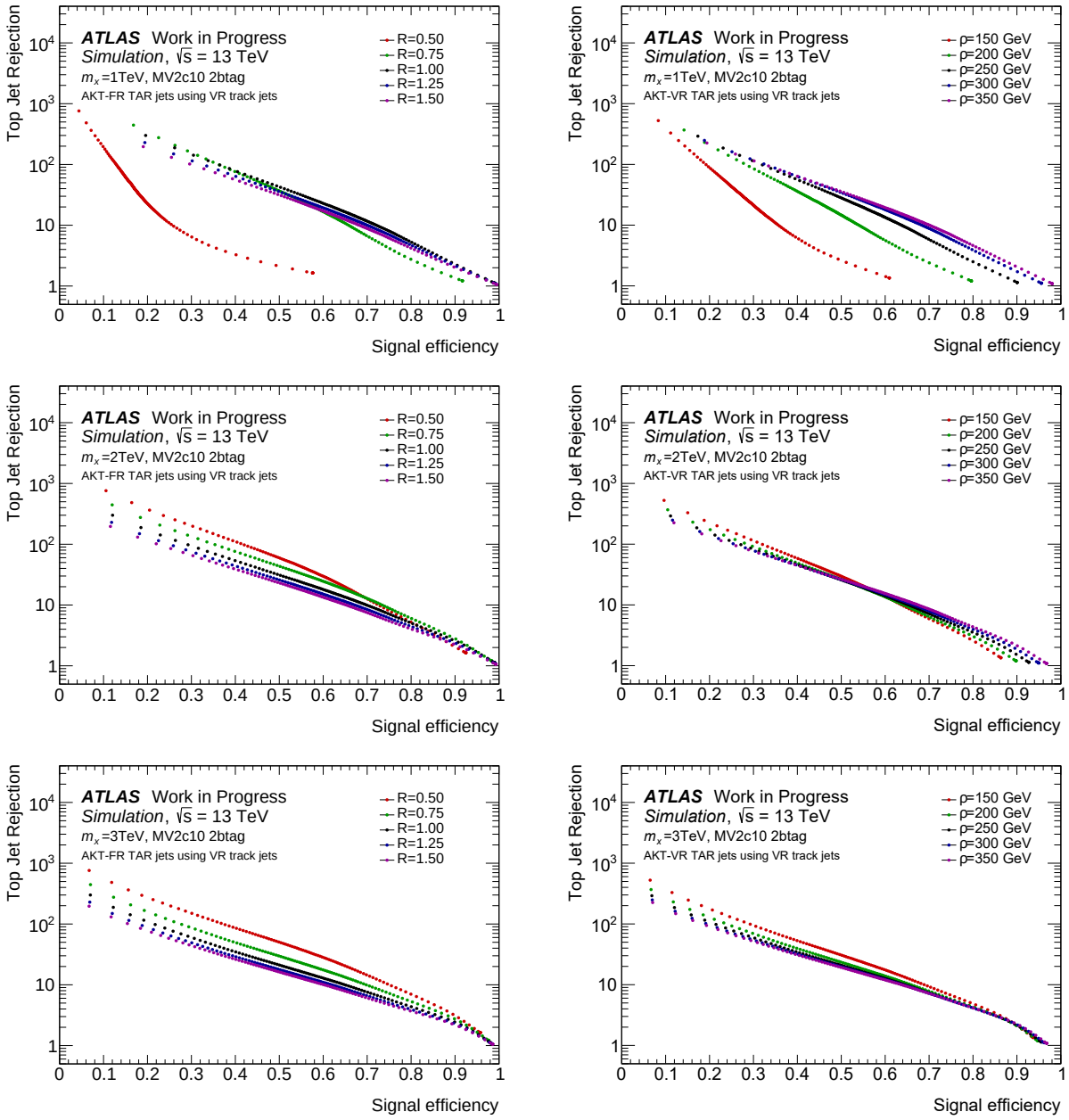
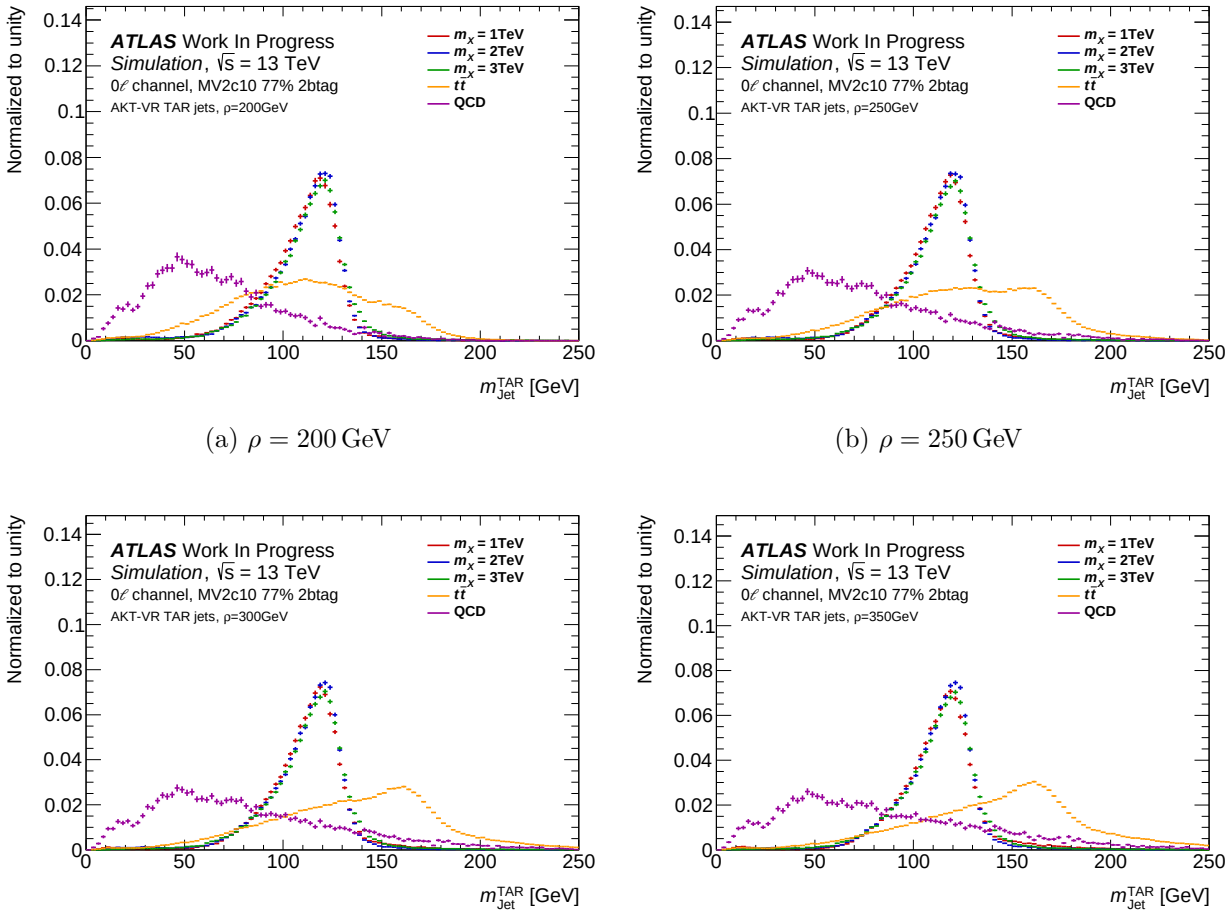


Fig. 41: ROC curves comparing the Higgs jet tagging performance against top jets in the double b-tag category for AKT-FR TAR jets (left) and AKT-VR TAR jets (right) for the $m_X = 1\,\text{TeV}$ (first row), $m_X = 2\,\text{TeV}$ (second row) and $m_X = 1\,\text{TeV}$ (third row) resonance mass. For AKT-FR TAR jets, the size parameter is varied, while the energy scale is varied for AKT-VR TAR jets.

performance, while smaller jet sizes perform better for heavier resonance masses and larger jet sizes perform better for lighter resonance masses. The smallest energy scale $\rho = 150\,\text{GeV}$ and size parameter $R = 0.5$ can be excluded, as compared to other parameter values, the Higgs jet tagging performance drops significantly for a $m_X = 1\,\text{TeV}$ resonance mass. In the following, the Higgs jet tagger is extended to include jet substructure information and its performance is further studied.

6.3.1. Track-Assisted Reclustered Jet Mass Cut

As mentioned in the introduction to this chapter, the flavour tagging of the track jets is only one part of constructing a $H \to b\bar{b}$ jet tagger. A mass window cut on the TAR jet mass, $m_{\text{Jet}}^{\text{TAR}}$, can significantly help to increase the discrimination against QCD and top jets. Figure 42 shows $m_{\text{Jet}}^{\text{TAR}}$ of top, QCD and Higgs jets reconstructed using the AKT-VR TAR jet algorithm for various values of the energy scale in the double b-tag category. The respective distributions for AKT-FR TAR jets for several size parameters are shown in Figure 62 in Appendix A.4. As expected, a smooth peak



(a) $\rho = 200\,\text{GeV}$ (b) $\rho = 250\,\text{GeV}$

(c) $\rho = 300\,\text{GeV}$ (d) $\rho = 350\,\text{GeV}$

Fig. 42: Masses of top, QCD and Higgs jets originating from heavy resonances of mass m_X, reconstructed using the AKT-VR TAR jet algorithm. Distributions are shown for various values of the energy scale in the double b-tag category.

around the Higgs boson mass $m_H \approx 125\,\text{GeV}$ is observed for all signal samples. For $\rho = 200\,\text{GeV}$, top jets exhibit a wide distribution around the Higgs boson mass. For higher values of ρ, the distributions shifts to larger mass values, giving a peak close to the top mass $m_t \approx 173\,\text{GeV}$ and a long tail towards lower masses. Full containment of the top decay products in the TAR jet seems to be already reached at $\rho = 300\,\text{GeV}$, as the distribution is not considerably altered for higher values. In contrast, QCD jets exhibit a broad peak around $\sim 50\,\text{GeV}$ for $\rho = 200\,\text{GeV}$. For higher

energy scales, the distribution widens such that more jets fall into the region around the Higgs boson mass peak.

In order to reduce the QCD and top contributions, three different $m_{\text{Jet}}^{\text{TAR}}$ window cuts are studied. The mass window cuts are defined to give a 90%, 80% and 68% signal efficiency in the case of a $m_X = 3\,\text{TeV}$ resonance for the various TAR jet definitions and are then applied to the masses of the other two mass points as well as to the top and QCD jets. As an example, the resulting $m_{\text{Jet}}^{\text{TAR}}$ mass window cuts applied to AKT-VR TAR jets with $\rho = 250\,\text{GeV}$ are visualized in Figure 43.

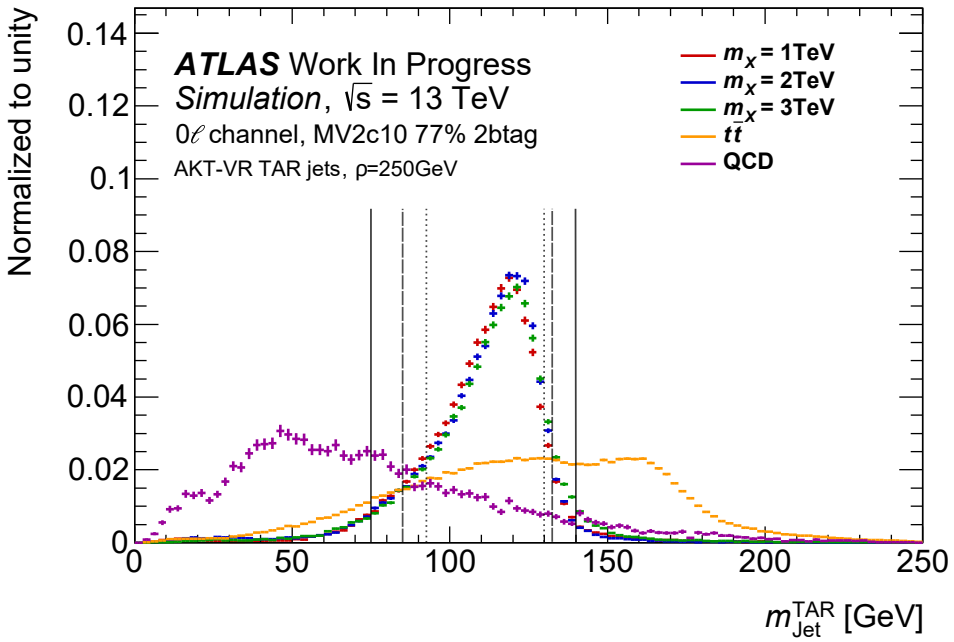


Fig. 43: Masses of top, QCD and Higgs jets originating from heavy resonances of mass m_X, reconstructed using the AKT-VR TAR jet algorithm with $\rho = 250\,\text{GeV}$. The vertical lines show the various Higgs jet mass window cuts corresponding to a 90% (solid), to an 80% (dashed) and to a 68% (dotted) signal efficiency.

A cutflow showing the event yields after every step of the selection for the signal and background contributions for the various energy scales of AKT-VR TAR jets is shown in Table 9. The identical table for AKT-FR jets is shown in Tab 11 in Appendix B. Note that the number of initial events, N_{ini}, is not identical for all parameters in $t\bar{t}$ as a result of events which have failed to be processed. However, the maximal difference is $\sim 0.05\%$, which can be neglected in the following. As for the $m_X = 1\,\text{TeV}$ resonance mass, a cross section of 15 pb was assumed, the initial number of events is higher compared to the $m_X = 2\,\text{TeV}$ and $m_X = 3\,\text{TeV}$ resonance masses, where a cross section of 2 pb was assumed. For the central value $\rho = 250\,\text{GeV}$ and an 80% mass cut, approximately 76%, 70% and 73% of the signal events are lost for a $m_X = 1$, 2 and 3 TeV resonance mass, respectively. At the same time, approximately 99.84% of $t\bar{t}$ events and $\sim 99.99\%$ of QCD di-jet events are rejected. In addition to the yields, the signal significance, $\Sigma = s/\sqrt{b}$, for the $t\bar{t}$ background was computed in order to assess which mass cut and jet definition gives the best separation between

ρ [GeV]	Selection	1 TeV	2 TeV	3 TeV	$t\bar{t}$	QCD [10^{-6}]
150	N_{ini}	989.54	137.56	138.62	33258526	1000000
	jet selection	592.51	107.48	113.78	3449667	244149
	double b-tag	139.34	44.41	45.25	64293	1087
	mass cut 68%	95.77	31.61	30.82	23605	93
	mass cut 80%	112.09	36.78	36.21	31296	139
	mass cut 90%	125.91	40.74	40.74	41506	245
200	N_{ini}	989.54	137.56	138.62	33271372	1000000
	jet selection	602.16	107.71	113.86	4102003	256934
	double b-tag	236.47	47.30	46.09	106648	1310
	mass cut 68%	169.18	33.68	31.41	37032	175
	mass cut 80%	196.48	39.24	36.88	48261	250
	mass cut 90%	218.06	43.46	41.50	63203	395
250	N_{ini}	989.54	137.56	138.62	33271372	1000000
	jet selection	604.52	107.77	113.88	4560982	267141
	double b-tag	289.91	49.61	46.81	140349	1476
	mass cut 68%	208.65	35.40	31.88	40558	236
	mass cut 80%	241.98	41.26	37.46	52588	324
	mass cut 90%	267.94	45.68	42.15	69055	489
300	N_{ini}	989.54	137.56	138.62	33254288	1000000
	jet selection	604.80	107.77	113.88	4854070	275068
	double b-tag	316.10	51.31	47.31	163290	1589
	mass cut 68%	225.32	36.69	32.21	38290	271
	mass cut 80%	261.28	42.75	37.87	49762	371
	mass cut 90%	289.50	47.28	42.60	66059	538
350	N_{ini}	989.54	137.56	138.62	33271372	1000000
	jet selection	604.87	107.77	113.88	4978317	278741
	double b-tag	326.59	52.57	47.65	176363	1656
	mass cut 68%	227.93	37.59	32.45	34535	288
	mass cut 80%	264.87	43.80	38.16	45155	382
	mass cut 90%	293.56	48.38	42.91	60851	552

Tab. 9: Cutflow showing the event yields after every step of the selection in the double b-tag category (77% WP) for the signal and background contributions for the various energy scales of AKT-VR TAR jets. As a full background estimate was not available during the time of the writing, the yields for QCD are normalized to one.

signal and background. Here, s is the number of signal events passing the selection and b is the number of $t\bar{t}$ events passing the selection. The computed significances for the various energy scales and mass cuts for AKT-VR TAR jets as well as for the various size parameters and mass cuts for AKT-FR TAR jets are shown in Table 10.

As a result of the higher cross section assumption for $m_X = 1\,\text{TeV}$ resonances, significances are higher compared to heavier resonance masses. For the $m_X = 1\,\text{TeV}$ resonance, $\rho = 350\,\text{GeV}$ with a mass cut of 80% gives the largest significance of ~ 1.25 for AKT-VR jets. This is not

R	Mass cut	1 TeV	2 TeV	3 TeV
0.50	68%	0.78	0.55	0.50
	80%	0.76	0.53	0.49
	90%	0.68	0.46	0.44
0.75	68%	1.37	0.26	0.22
	80%	1.36	0.25	0.22
	90%	1.26	0.23	0.21
1.00	68%	1.22	0.19	0.16
	80%	1.22	0.19	0.16
	90%	1.03	0.16	0.14
1.25	68%	1.18	0.19	0.16
	80%	1.16	0.18	0.16
	90%	0.85	0.13	0.11
1.50	68%	1.15	0.18	0.15
	80%	1.11	0.17	0.15
	90%	0.80	0.12	0.10

ρ [GeV]	Mass cut	1 TeV	2 TeV	3 TeV
150	68%	0.62	0.21	0.20
	80%	0.63	0.21	0.21
	90%	0.62	0.20	0.20
200	68%	0.88	0.18	0.16
	80%	0.90	0.18	0.17
	90%	0.87	0.17	0.17
250	68%	1.04	0.18	0.16
	80%	1.06	0.18	0.16
	90%	1.02	0.17	0.16
300	68%	1.15	0.19	0.16
	80%	1.17	0.19	0.17
	90%	1.13	0.18	0.17
350	68%	1.23	0.20	0.17
	80%	1.25	0.21	0.18
	90%	1.19	0.20	0.17

Tab. 10: The significance Σ in the double b-tag category (77% WP) for the various size parameters for AKT-FR TAR jets (left) and for the various energy scales of AKT-VR TAR jets (right) for the several resonance masses.

surprising as $\rho = 350\,\text{GeV}$ has been shown to perform best for a this mass point (see Figure 41). Similarly, the maximum significance for the $m_X = 3\,\text{TeV}$ resonance is ~ 0.21 for the lowest energy scale $\rho = 150\,\text{GeV}$ and an 80% central mass cut. For the $m_X = 2\,\text{TeV}$ resonance, the lowest and largest parameter yield the best significances. Also note that, as shown for the performance, the significance for a $m_X = 1\,\text{TeV}$ resonance drops from 1.23 to 0.62 with respect to the largest and lowest values of the size parameter R, while for the other heavy resonances the significances remain relatively constant. As for the mass cut, a central 80% mass cut seems to give the best significances.

6.3.2. Track-Assisted Reclustered Jet Substructure Variables

After applying a TAR jet mass cut, additional discrimination against QCD and top jets can be achieved by exploiting the internal structure of the Higgs jets through jet substructure variables. Several variables have been studied for this purpose:

- The N-subjettiness ratios τ_{21}, τ_{32} and τ_{42}
- The ECF ratios C_1, C_2, C_3 and D_2
- The generalized ECF ratios M_2, N_2, N_3, L_1, L_2, L_3, L_4 and L_5

This section aims to study the separation power of these variables and to find the variable which gives the optimal discrimination between Higgs jets, QCD and top jets for the use of TAR jets. Furthermore, it is studied whether different jet definitions lead to better performance when including jet substructure information. A large number of parameters can be varied for this study,

including the various signal samples, jet definitions and mass cuts. For reasons of simplicity, only AKT-VR TAR jets with the central value $\rho = 250\,\text{GeV}$ are used in the following. Moreover, the central $m_X = 2\,\text{TeV}$ heavy resonance mass is used for all further studies. The 80% mass cut is applied. From the resulting distributions, five variables are chosen which provide good discrimination against top and QCD jets. Using ROC curves, a scan through these variables is performed to show the signal efficiency against the top and QCD jet rejections. The performance differences between the jet definitions is then assessed using a two-step procedure. First, the size parameters and energy scales are fixed to the central values, while the type of reclustering algorithm is varied for the best variables. Then, the AKT-VR TAR jet algorithm is fixed, while ρ is varied for the best variables.

Figure 44 shows the N-subjettiness ratio τ_{42} in the double b-tag category for QCD and top jets as well as for Higgs jets with and without an 80% mass cut applied for $\rho = 200\,\text{GeV}$ and $\rho = 300\,\text{GeV}$. Without the mass cut, very similarly shaped unimodal distributions peaking at ~ 0.6 are seen for

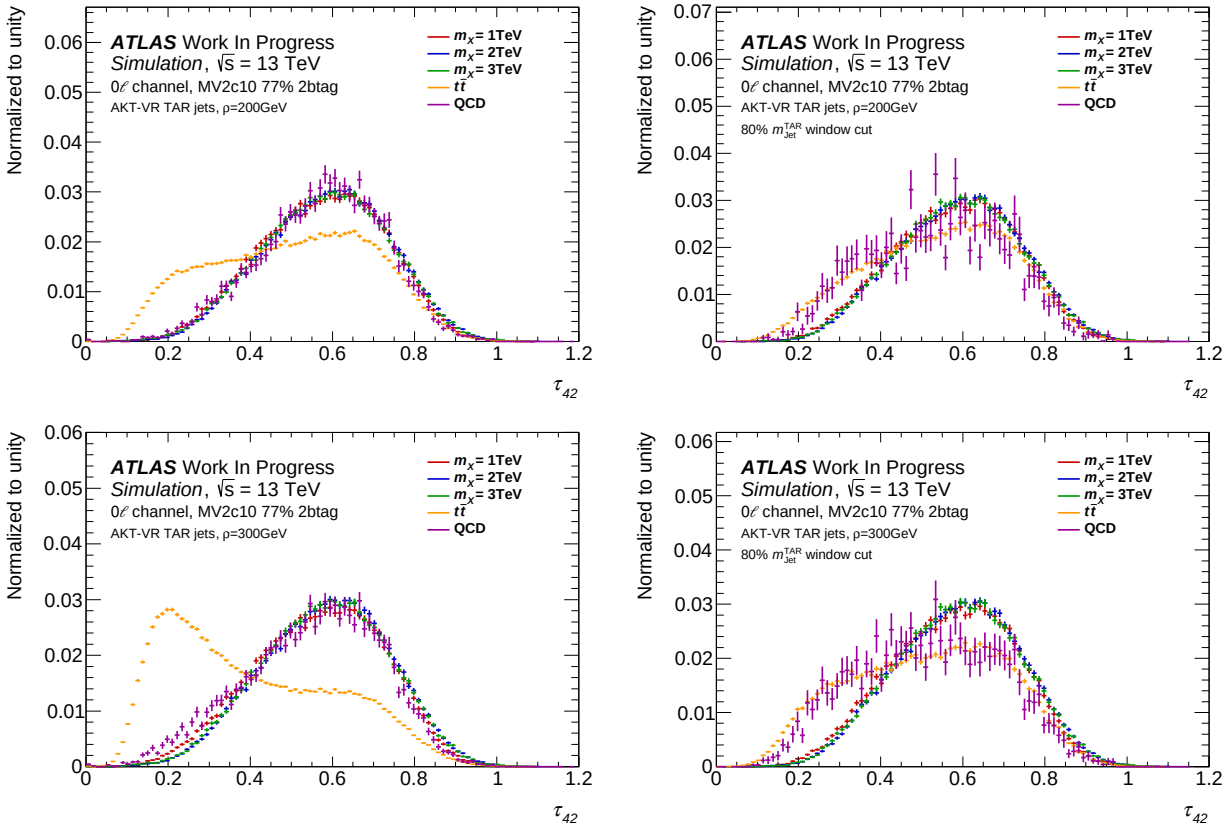


Fig. 44: N-subjettiness ratio τ_{42} in the double b-tag category for QCD and top jets as well as for Higgs jets for various values of m_X reconstructed using AKT-VR TAR jets. Distributions are shown without (left) and with an 80% $m_{\text{Jet}}^{\text{TAR}}$ cut (right) for the mass scale values $\rho = 200\,\text{GeV}$ (top) and $\rho = 300\,\text{GeV}$ (bottom).

Higgs and QCD jets. For top jets, a bimodal distribution is observed. For $\rho = 200\,\text{GeV}$, the larger peak of the bimodal distribution is within the signal distribution, while the smaller maximum peaks

outside at ~ 0.2. This peak grows considerably for $\rho = 300\,\text{GeV}$, such that the separation power is greatly increased. This is, however, the result of a higher fraction of top jets passing the selection for larger values of ρ (see Table 9). While the separation power is larger, this does not necessarily mean that larger ρ values lead to overall higher significances. After applying the mass cut, most of the separation power is lost and higher values of the energy scale only provide marginally more discrimination against top jets. However, the distribution for QCD jets changes after the mass cut and closely follows the same distribution as for top jets, such that separation power between Higgs and QCD jets is increased. This feature has been observed for all studied jet substructure variables: QCD jets tend to become more top-like for a 90% mass cut and almost completely follow the top jet distributions for tighter mass cuts. Further distributions of N-subjettiness ratios, including τ_{21} and τ_{32}, are shown in Appendix A.5.

Figure 45 shows the ECF ratio C_1 in the double b-tag category for QCD and top jets as well as for Higgs jets with and without an 80% mass cut applied. The distributions are heavily dependent

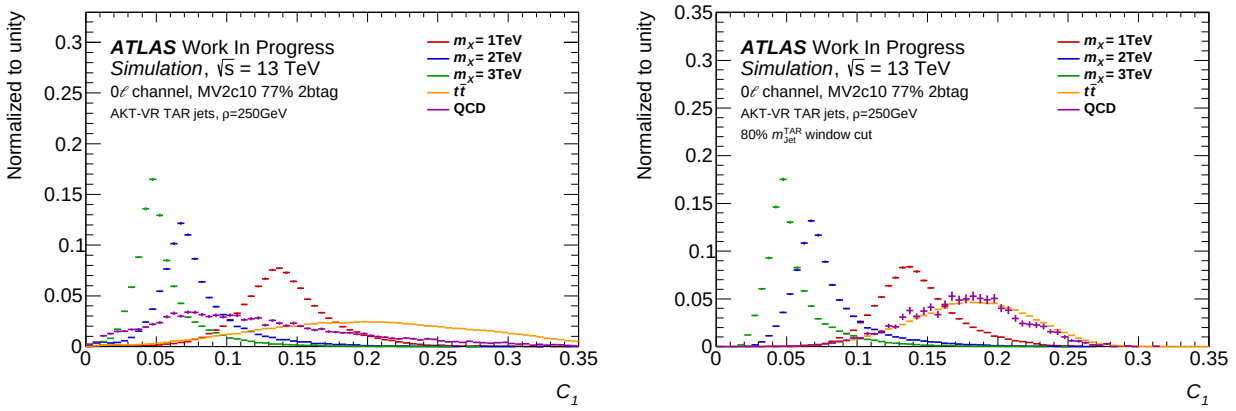


Fig. 45: ECF ratio C_1 in the double b-tag category for QCD and top jets as well as for Higgs jets for various values of m_X reconstructed using AKT-VR TAR jets with $\rho = 250\,\text{GeV}$. Distributions are shown without (left) and with an 80% $m_{\text{Jet}}^{\text{TAR}}$ cut (right)

on the resonance mass. For a $m_X = 1\,\text{TeV}$ resonance, C_1 peaks at ~ 0.14, while the peak of the $m_X = 2\,\text{TeV}$ and $m_X = 3\,\text{TeV}$ resonances are shifted towards $C_1 \approx 0.07$ and $C_1 \approx 0.05$, respectively. For top and QCD jets, very wide distributions are seen, where top jets tend to higher and QCD jets to lower C_1 values. After applying an 80% mass cut, the top jet distribution becomes narrower and exhibits a prominent maximum at ~ 0.19. As mentioned before, the distribution for QCD jets closely follows this shape. With application of the mass cut, a large separation power is gained for both background jets. Almost a complete separation between the background distributions and the $m_X = 2\,\text{TeV}$ and $m_X = 3\,\text{TeV}$ resonance masses is reached. For the $m_X = 1\,\text{TeV}$ resonance mass, a good discrimination power is still seen. C_1 distributions for further energy scales are shown in Figure 66 in Appendix A.2.

As a last example, Figure 46 shows the ECF ratio D_2 in the double b-tag category for QCD and top jets as well as for Higgs jets with and without an 80% mass cut applied. For signal jets, the distributions show a single peak at ~ 0.8 with a small dependency on the resonance mass, while for

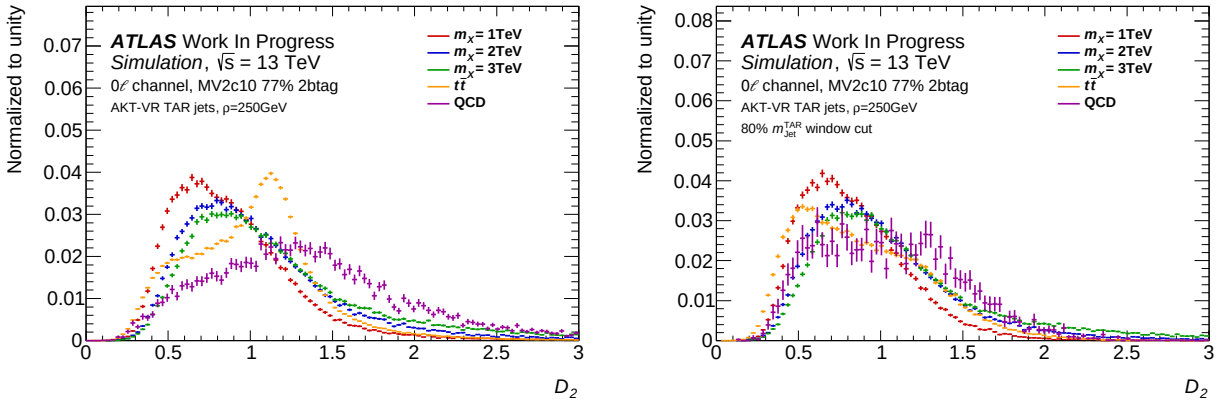


Fig. 46: ECF ratio D_2 in the double b-tag category for QCD and top jets as well as for Higgs jets for various values of m_X reconstructed using AKT-VR TAR jets with $\rho = 250\,\text{GeV}$. Distributions are shown without (left) and with an 80% $m_{\text{Jet}}^{\text{TAR}}$ cut (right).

top jets the distribution exhibits a sharp peak at ~ 1.1 with a smaller bump at ~ 0.5. QCD jets lead to a wide distribution which tends to larger values, thus providing some separation power. After the mass cut, most of the discrimination power against QCD and top jets is lost. D_2 distributions for further energy scales are shown in Figure 67 in Appendix A.5.

As mentioned before, a total of five variables with good separation power are selected for further studies. These include the variables presented before, τ_{42}, C_1, D_2 as well as L_2 and N_2. Distributions for L_2 and N_2 are shown in Figure 69 and Figure 68 in Appendix A.5, respectively. Figure 47 shows the ROC curves comparing the performance of lower ("<": for a variable X_i and a certain threshold Y, all events with $X_i < Y$ are kept) and upper cuts (">": for a variable X_i and a certain threshold Y, all events with $X_i > Y$ are kept) on these substructure variables. Note that the efficiencies in these ROC curves are calculated based on jets passing the jet selection, b-tagging requirement and mass cut in order to only focus on the separation power of the respective substructure variable. A lower cut on C_1 significantly outperforms all other cuts on the studied variables in the case of top jets as well as in the case for QCD jets. For top jets, an upper cut on τ_{42} provides the second best separation while lower cuts on L_2 and N_2 perform second-best for QCD jets. For D_2, a upper cut gives some separation power against top jets, but a lower cut is required to discriminate against QCD jets for a $m_X = 2\,\text{TeV}$ resonance. For this reason, a single-sided cut on D_2 can be discarded. Note that C_1 shows a significant dependence on the resonance mass (see Figure 45). However, even in the case of a $m_X = 1\,\text{TeV}$ resonance mass, which offers the worst separation power, C_1 still considerably outperforms all other variables (see Figure 48). Furthermore, it is seen that a lower cut on D_2 performs better compared to an upper cut for QCD and top jets in this case. Figure 70 in Appendix A.5 shows the respective ROC curves for all heavy resonances, including for a $m_X = 3\,\text{TeV}$ resonance mass.

Next, studies on the impact of the jet reclustering algorithm choice are presented. The ROC curves shown in the following are normalized to the number of Higgs jets passing the jet selection, such that the b-tagging and mass cut performance is included. In Figure 49, the ROC curves comparing

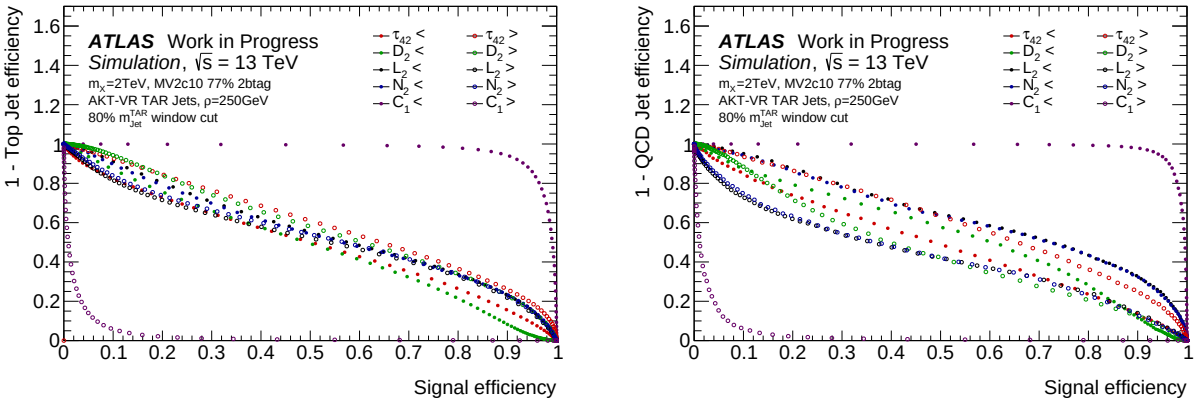


Fig. 47: ROC curves comparing the performance of lower and upper cuts on various substructure variables for AKT-VR TAR jets with $\rho = 250\,\text{GeV}$ in the case of top jets (left) and QCD jets (right) for a $m_X = 2\,\text{TeV}$ resonance. An 80% TAR jet mass window cut is applied.

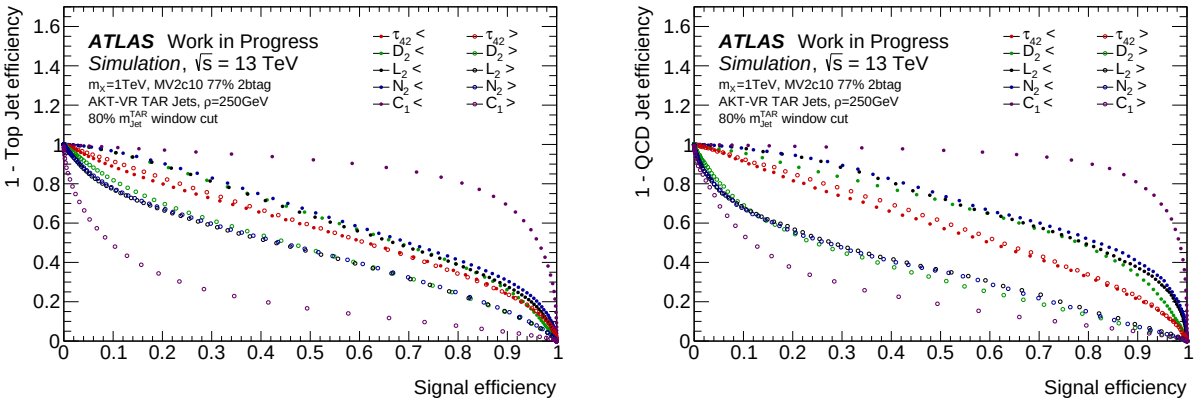


Fig. 48: ROC curves comparing the performance of lower and upper cuts on various substructure variables for AKT-VR TAR jets with $\rho = 250\,\text{GeV}$ in the case of top jets (left) and QCD jets (right) for a $m_X = 2\,\text{TeV}$ resonance. An 80% TAR jet mass window cut is applied.

the performance of the FR ($R = 1.0$) and VR ($\rho = 250\,\text{GeV}$) TAR jet collections for a lower cut on C_1 and an upper cut on τ_{42} in the case of top jets and QCD jets are shown. VR TAR jets reach a signal efficiency of up to ~ 0.4, while FR TAR jets reach slightly lower efficiency values. Note that there is not a one-to-one correspondence between the central values $R = 1.0$ and $\rho = 250\,\text{GeV}$ for FR and VR jets, respectively, such that a direct comparison is only possible between the AKT and CA reclustering algorithms. While the CA algorithm performs marginally better, no significant differences are observed between the two algorithms for cuts on τ_{42} and C_1. The ROC curves for the D_2, L_2 and N_2 variables are shown in Figure 71 in Appendix A.5 and show an identical behaviour.

Finally, the energy scale of the AKT-VR jets is varied for scans over the five variables in order to assess which energy scale ρ performs best. Figure 50 shows the ROC curves comparing the performance of the various energy scales for a lower cut on C_1 and an upper cut on τ_{42} in the case of top and QCD jets. As expected, higher efficiencies are reached for larger values of ρ as more

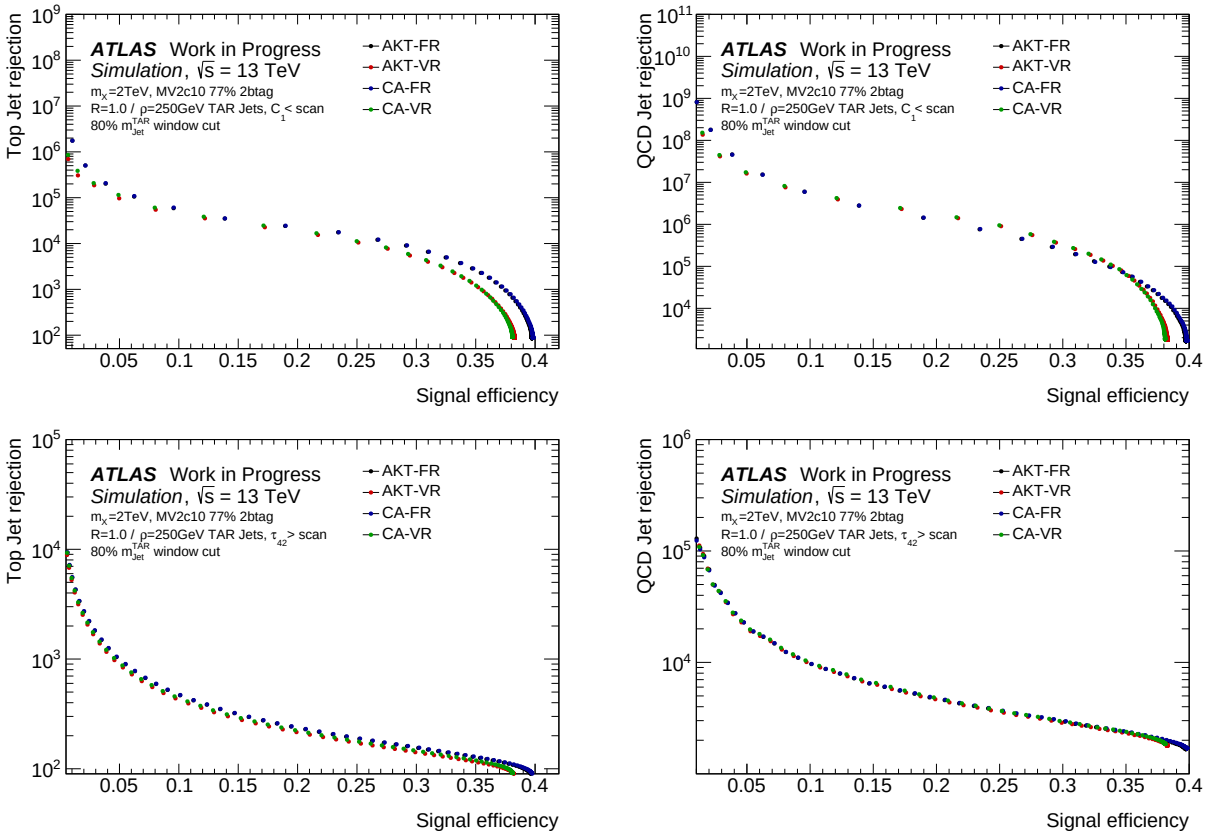


Fig. 49: ROC curves comparing the performance of the FR ($R = 1.0$) and VR ($\rho = 250\,\text{GeV}$) TAR jet collections for a lower cut on C_1 (top) and an upper cut on τ_{42} (bottom) in the case of top jets (left) and QCD jets (right). An 80% TAR jet mass cut is applied.

hadronic decay products are collected, increasing the probability of collecting both B-hadrons from the Higgs boson decay. For top jets, higher values of the energy scale perform best, while for QCD jets, the opposite behaviour is seen. While this property is less pronounced for C_1, the N-subjettiness ratio τ_{42} exhibits a significant dependence on the choice of ρ. For top jets, the parameter choice impacts the background rejection by up to a factor of ~ 1.5, while for QCD jets the choice impacts the rejection by up to a factor of ~ 3. Similar observations can be made for the other variables. The respective ROC curves are shown in Figure 72 in Appendix A.5.

Summarising, the ECF ratio C_1 is by far the most promising jet substructure variable and can help to significantly reduce QCD and top jet contributions, in particular, after a TAR jet mass cut is applied. Furthermore, the N-subjettiness ratio τ_{42} is also a promising variable for distinction between Higgs and background jets. While the CA reclustering algorithm provides marginally better performance, no significant differences are found between the CA and AKT algorithms. For the rejection of top jets, higher energy scales and size parameters are expected to perform better, while for QCD jets, lower values of R and ρ seem to give better results. For the best performing variable, C_1, the choice of parameter only plays a minor role. In order to give a final recommendation on the choice of the jet definition, a full QCD background estimate is required.

Moreover, the final decision on which jet definition provides the optimal performance in the $b\bar{b}WW^*$ channel is also dependent on the $H \to WW^*$ system, for which separate studies are required.

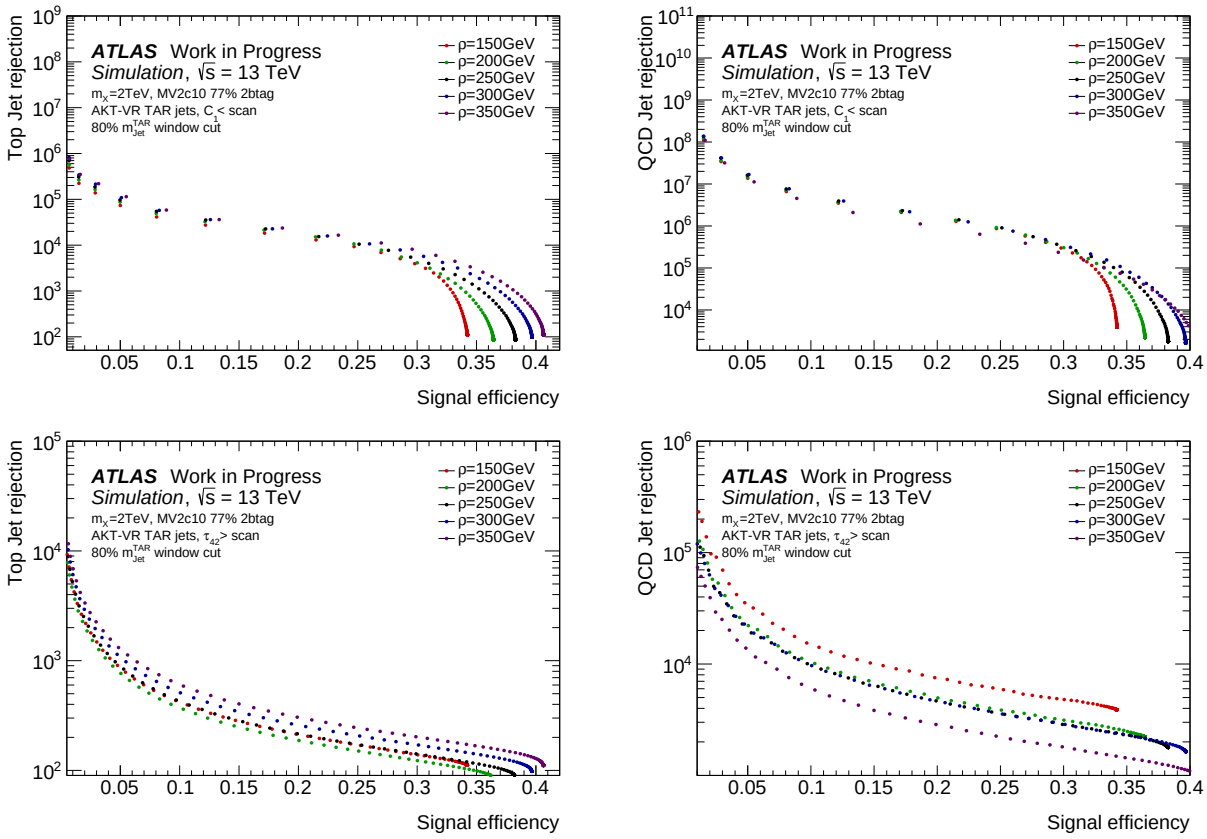


Fig. 50: ROC curves comparing the performance of the various energy scales for AKT-VR TAR jets for a lower cut on C_1 (top) and an upper cut on τ_{42} (bottom) in the case of top jets (left) and QCD jets (right). An 80% TAR jet mass cut is applied.

Chapter 7

Conclusions

Performance studies on boosted $H \to b\bar{b}$ tagging in the $HH \to b\bar{b}WW^*$ decay channel were presented. First, a brief truth-level study on the decay kinematics of the $b\bar{b}WW^*$ channel was shown. Then, the need for advanced subjet reconstruction techniques for heavily boosted topologies was motivated in a truth-level study aiming to mimic the standard $H \to b\bar{b}$ tagging procedure implemented by ATLAS. For large-R CALO jets, the performance of various new subjet reconstruction techniques, which aim to improve the tagging for high-p_T jets, was studied and compared to the use of standard FR track jets. It was found that the Higgs jet efficiency decreases rapidly for FR track jets starting from a jet p_T threshold of $\sim 1\,\text{TeV}$ as a result of an increasing overlap of the hadronic decay products. The loss in efficiency can be effectively recovered by using VR track jets, which allow the B-hadrons to be collected in differently sized track jets. Compared to FR track jets, VR jets were found to give an efficiency gain of up to $\sim 77\%$ for jets with large transverse momenta ($p_T > 1\,\text{TeV}$) and an overall efficiency gain of $\sim 37\%$ in the case of a $m_X = 3\,\text{TeV}$ resonance mass. To a lesser extent, CoM and ExKT subjets are also able to recover some of the efficiency loss, where ExKT subjets offer a slightly better performance. With regard to the top jet rejection, VR track jets also offer the best performance, such that their use should be preferred over the standard FR, CoM and ExKT subjets.

Furthermore, TAR jets, which provide a variety of advantages compared to large-R CALO jets, were studied for boosted $H \to b\bar{b}$ tagging for the first time. The large flexibility offered by TAR jets was exploited to compare the performance of different reclustering parameters and algorithms. For low p_T jets, hadronic decay products are relatively widely distributed within the detector. The reconstruction with smaller energy scales ρ and size parameters R fails to collect both of the B-hadrons from the $H \to b\bar{b}$ decay, such that the Higgs jet efficiencies decrease significantly. In the high-p_T region, the decay products are collimated. Smaller parameter values suffice to collect both B-hadrons, such that no differences in efficiencies were observed. In turn, smaller sizes are beneficial to reject $t \to bW$ jets in the low-p_T region as the probability of collecting the B-hadron decreases such that both subjets are required to be mistagged for the TAR jet to be double b-tagged. For QCD jets, a similar but less pronounced behaviour was observed. As the smallest value of the size parameter, $R = 0.5$, and energy scale, $\rho = 150\,\text{GeV}$, exhibits a massive decrease in performance for the $m_X = 1\,\text{TeV}$ resonance, their use is not advised.

After exploitation of the heavy flavour content of the TAR jets, the mass and the internal structure of the reconstructed Higgs jets was used to further discriminate Higgs jets against QCD and top jets. Three TAR jet mass window cuts corresponding to a 90%, 80% and 68% signal efficiency were

studied. For top jets, the 80% mass window cut was shown to give the best significance for most of the studied parameters. In addition, TAR jet substructure variables were studied and five variables which provide good separation power were selected for further studies. It was shown that after applying an 80% mass cut, most of the separation power of the jet substructure variables is lost. The distributions for QCD jets are also seen to closely follow those for top jets after applying a TAR jet mass cut. It can be inferred that the QCD jets become top-like by adopting a very similar internal structure as the top jets. The ECF ratio C_1 was found to give the best separation power by a great amount, followed by the N-subjettiness ratio τ_{42}. While the CA algorithm provides marginally better performance, no significant differences were found between the CA and AKT reclustering algorithms. Larger jet sizes were shown to be beneficial for the rejection of top jets, while lower energy scales and size parameters were shown to perform better for QCD jets. This behaviour is observed in all of the studied variables, but is less pronounced in the case of the best-separating variable C_1. A final answer on the optimal jet definition for $H \to b\bar{b}$ tagging requires a full QCD background estimate, which was not available during the time of the writing. Moreover, separate studies are required on the $H \to WW^*$ system to make a final decision on which jet definition provides the optimal performance in the $b\bar{b}WW^*$ channel.

The LHC is expected to provide approximately $3000\,\text{fb}^{-1}$ of integrated luminosity by the year 2035. With this amount of data collected, sensitivity to the Higgs boson self-coupling in non-resonant Higgs boson pair production is expected to be gained. Moreover, searches for heavy resonances in boosted topologies will gain more importance. In particular, advanced methods designed for heavy-flavour tagging for boosted resonances will need to be employed to ensure an effective rejection of the large background contributions expected in Run 3 and beyond.

Acknowledgements

Firstly, I would like to thank Stan for giving me the opportunity to conduct this research work in his group and providing all the necessary support throughout my studies. Furthermore, I would like to thank Jason for the many physics discussions as well as for the technical support. Special thanks go to Kira for her endless support in all aspects. Without your guidance and persistent help this thesis would not have been possible. Many thanks also to the whole group, which I always perceived as very welcoming and supportive. Finally, I wish to thank my parents and friends for their endless encouragement throughout my studies. I am happy that I will continue being part of this group and I am looking forward to new and valuable experiences at CERN in the coming years.

Appendix A

Additional Figures

A.1. Truth-Level Study of Decay Kinematics

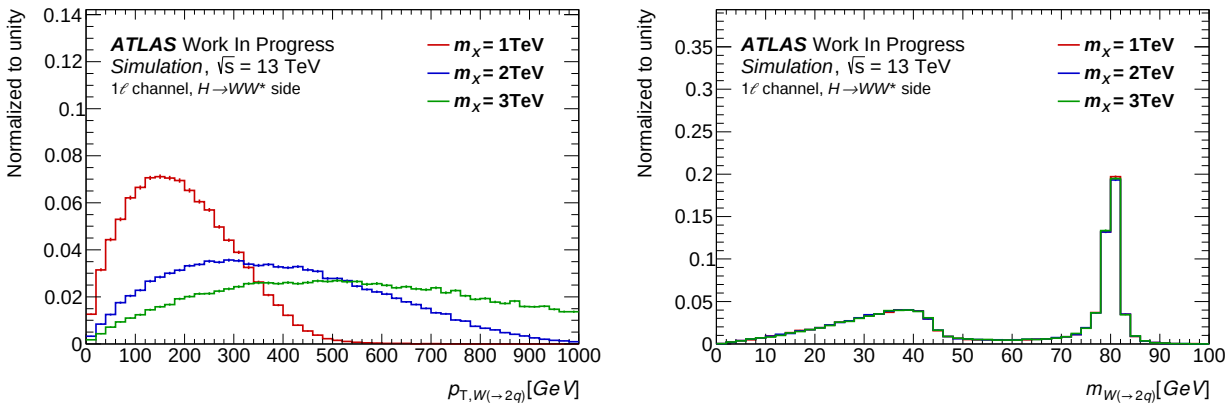


Fig. 51: Transverse momentum (left) and invariant mass spectrum (right) of the hadronically decaying truth W boson for several resonance mass points.

A.2. Performance of Subjet Reconstruction Techniques

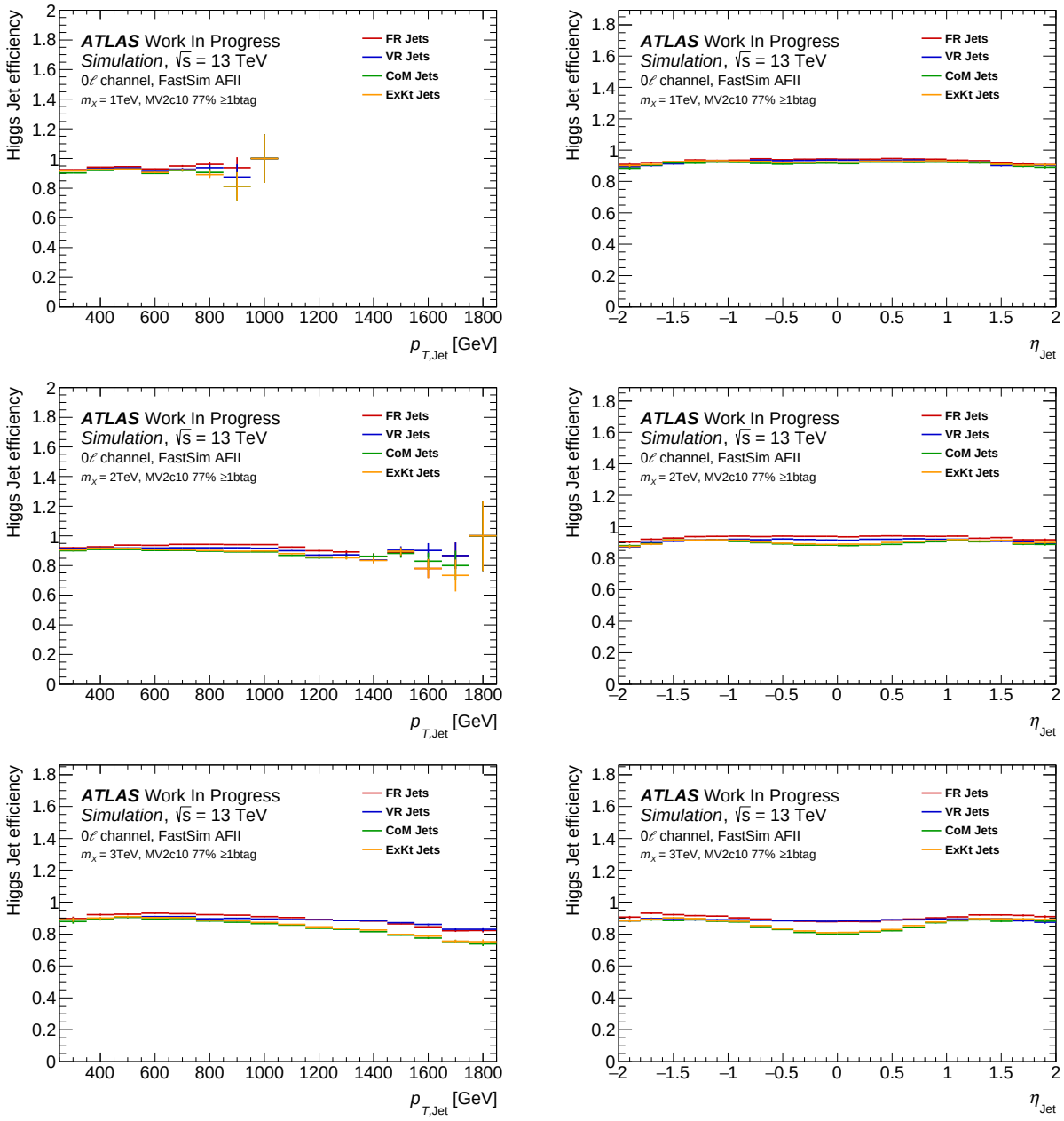


Fig. 52: Higgs jet efficiencies as a function of p_T (left) and η (right) in the inclusive single b-tag category for the various subjet reconstruction algorithms for the three mass points $m_X = 1$, 2 and 3 TeV.

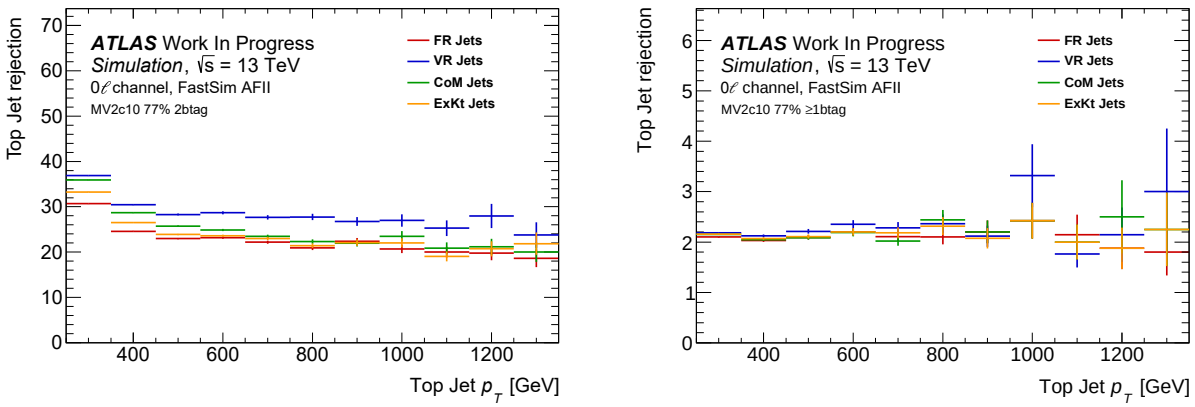


Fig. 53: Top jet rejection as a function of the top jet p_T in the double (left) and inclusive single (right) b-tag category for the various subjet reconstruction algorithms.

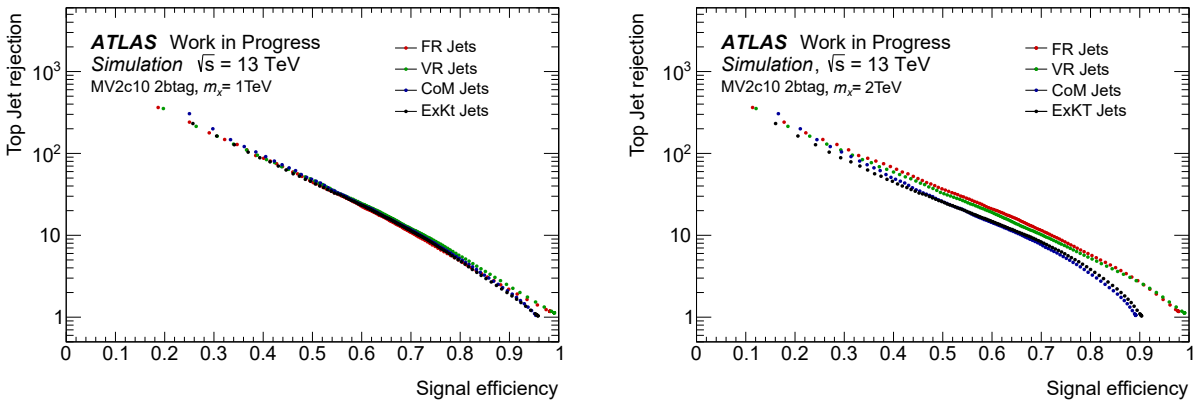


Fig. 54: ROC curves comparing the performance of various subjet reconstruction techniques for heavy resonance masses $m_X = 1\,\text{TeV}$ (left) and $m_X = 2\,\text{TeV}$ (right).

A.3. Performance of Track-Assisted Reclustered Jets

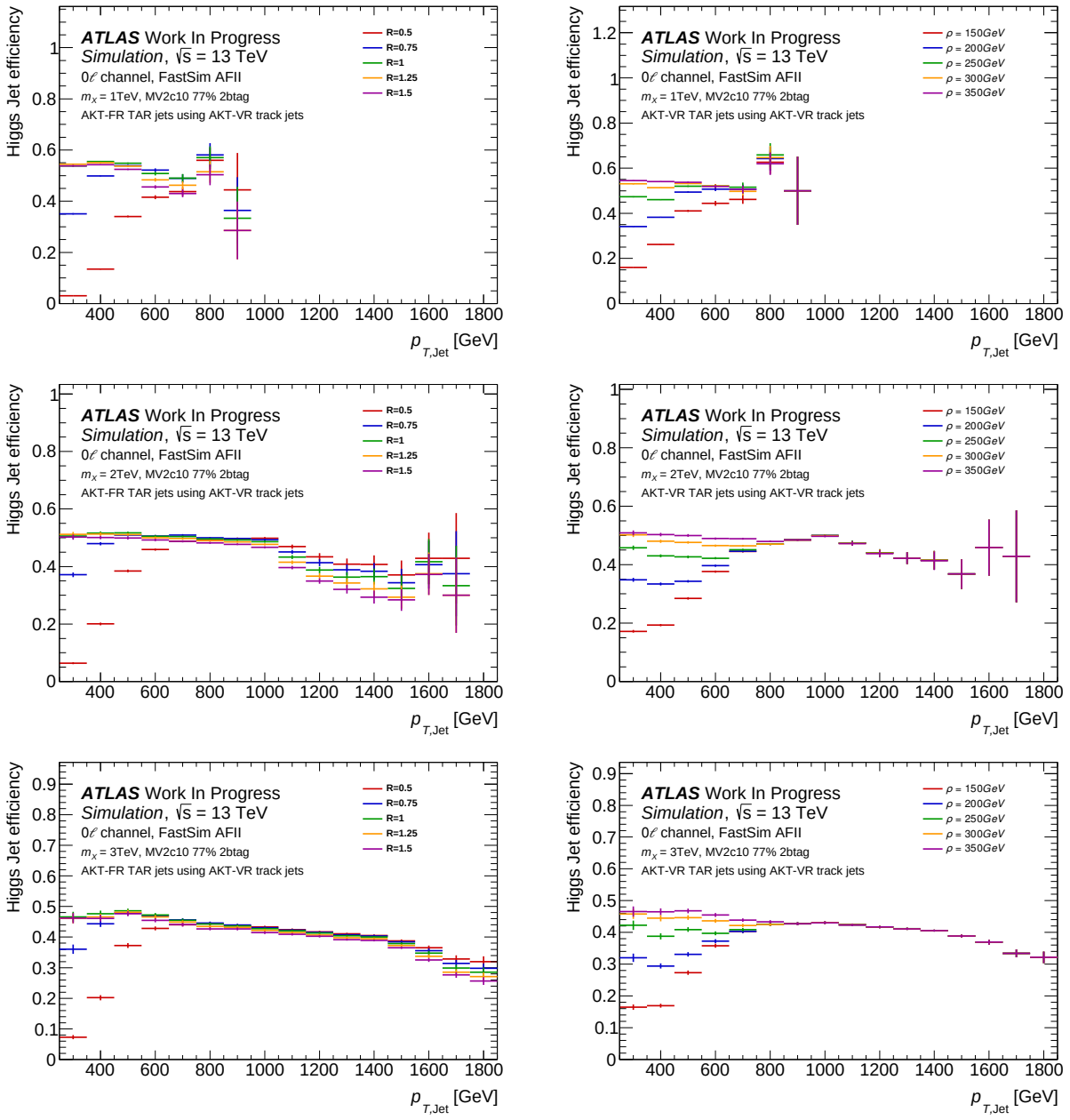


Fig. 55: Higgs jet efficiencies as a function of the TAR jet p_T in the double b-tag category for AKT-FR (left) and AKT-VR (right) TAR jets for the three mass points $m_X = 1\,\text{TeV}$ (first row), $m_X = 2\,\text{TeV}$ (second row) and $m_X = 3\,\text{TeV}$ (third row). For AKT-FR TAR jets, the size parameter is varied, while the energy scale is varied for AKT-VR TAR jets.

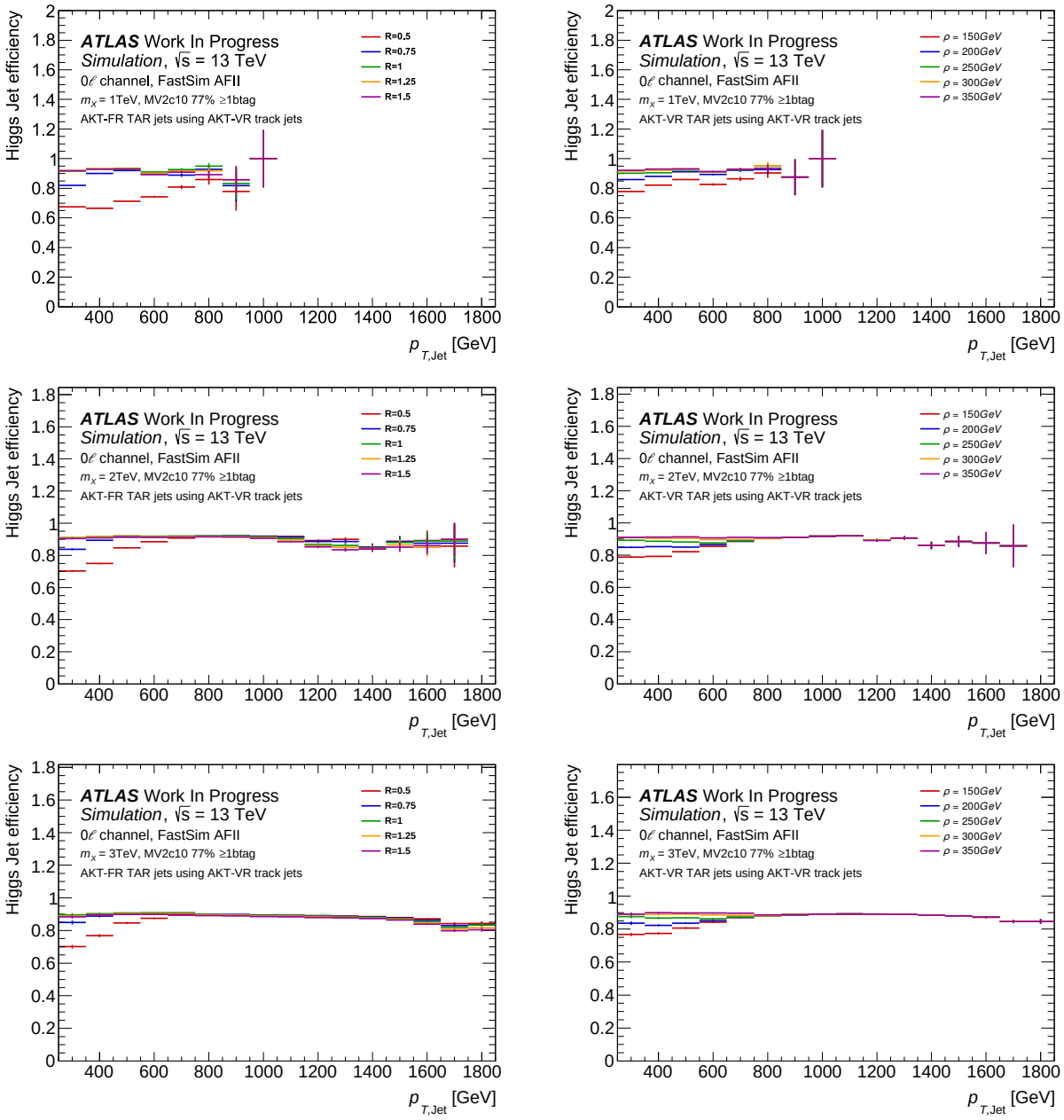


Fig. 56: Higgs jet efficiencies as a function of the TAR jet p_T in the inclusive single b-tag category for AKT-FR (left) and AKT-VR (right) TAR jets for the three mass points $m_X = 1\,\mathrm{TeV}$ (first row), $m_X = 2\,\mathrm{TeV}$ (second row) and $m_X = 3\,\mathrm{TeV}$ (third row). For AKT-FR TAR jets, the size parameter is varied, while the energy scale is varied for AKT-VR TAR jets.

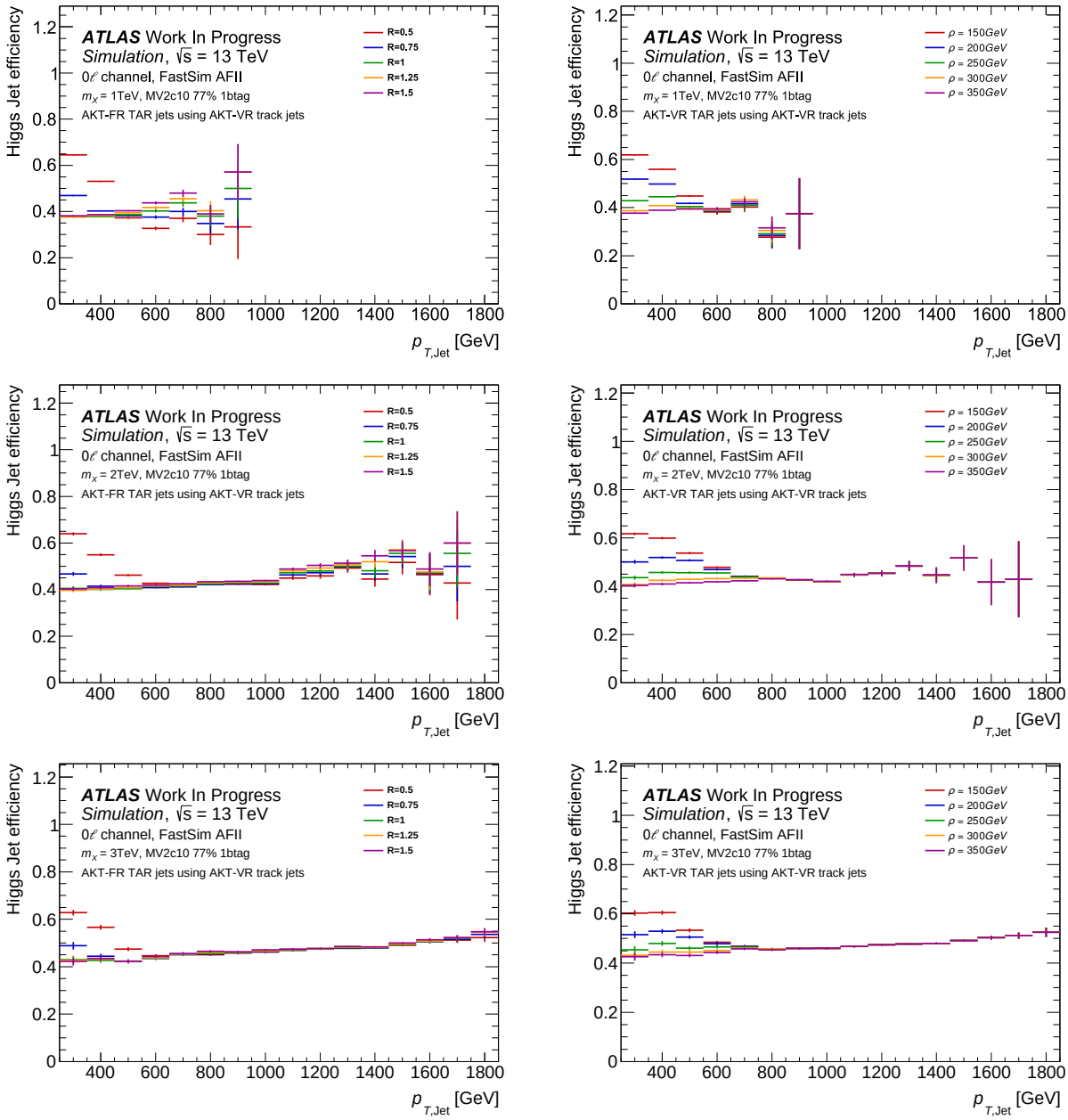


Fig. 57: Higgs jet efficiencies as a function of the TAR jet p_T in the exclusive single b-tag category for AKT-FR (left) and AKT-VR (right) TAR jets for the three mass points $m_X = 1\,\mathrm{TeV}$ (first row), $m_X = 2\,\mathrm{TeV}$ (second row) and $m_X = 3\,\mathrm{TeV}$ (third row). For AKT-FR TAR jets, the size parameter is varied, while the energy scale is varied for AKT-VR TAR jets.

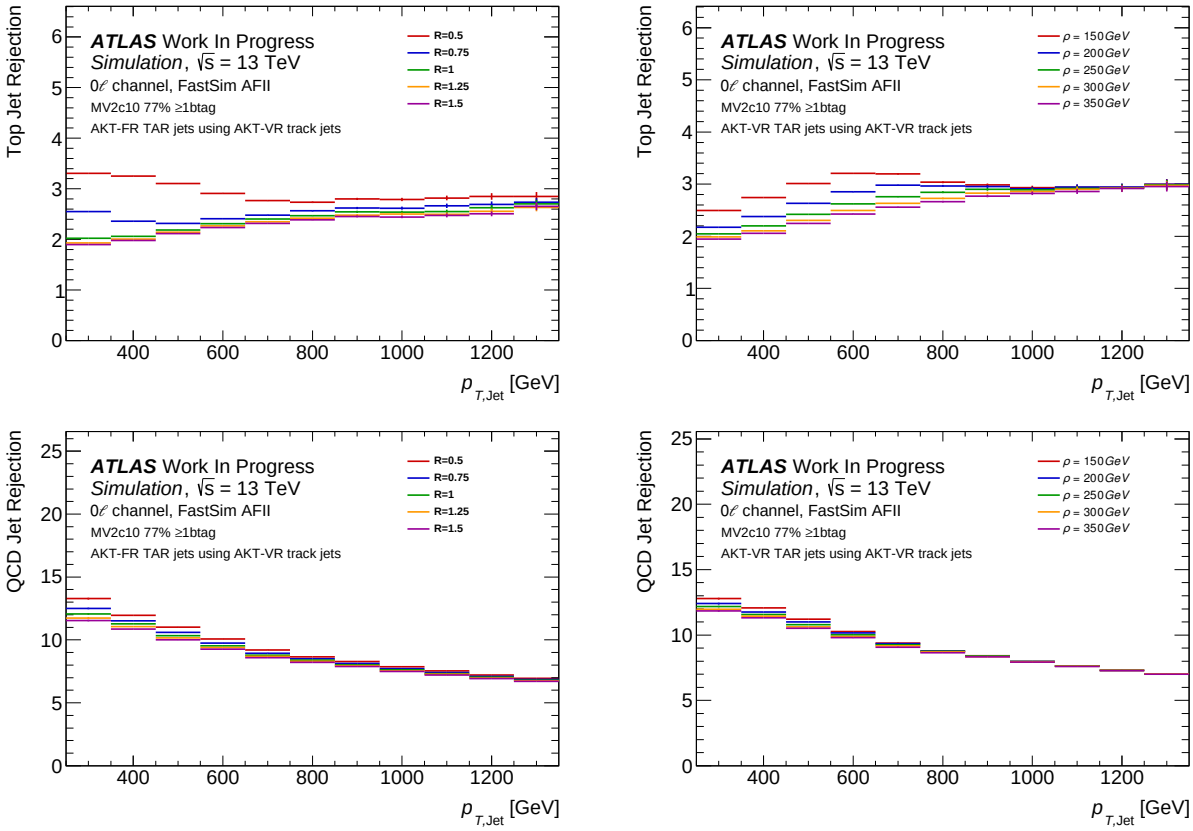


Fig. 58: Top jet rejection (top) and QCD rejection (bottom) as a function of the TAR jet p_T in the inclusive single b-tag category for AKT-FR (left) and AKT-VR (right) TAR jets. For AKT-FR TAR jets, the size parameter, is varied, while the energy scale is varied for AKT-VR TAR jets.

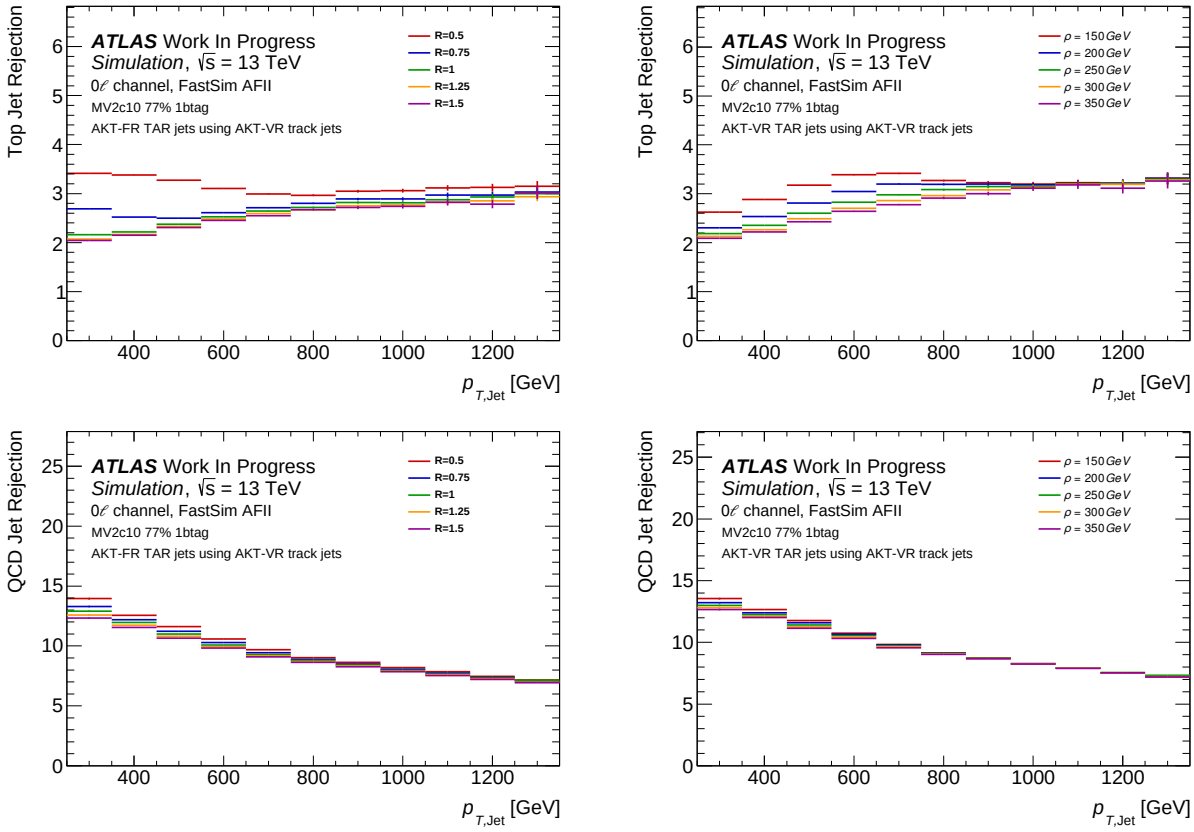


Fig. 59: Top jet rejection (top) and QCD rejection (bottom) as a function of the TAR jet p_T in the exclusive single b-tag category for AKT-FR (left) and AKT-VR (right) TAR jets. For AKT-FR TAR jets, the size parameter is varied, while the energy scale is varied for AKT-VR TAR jets.

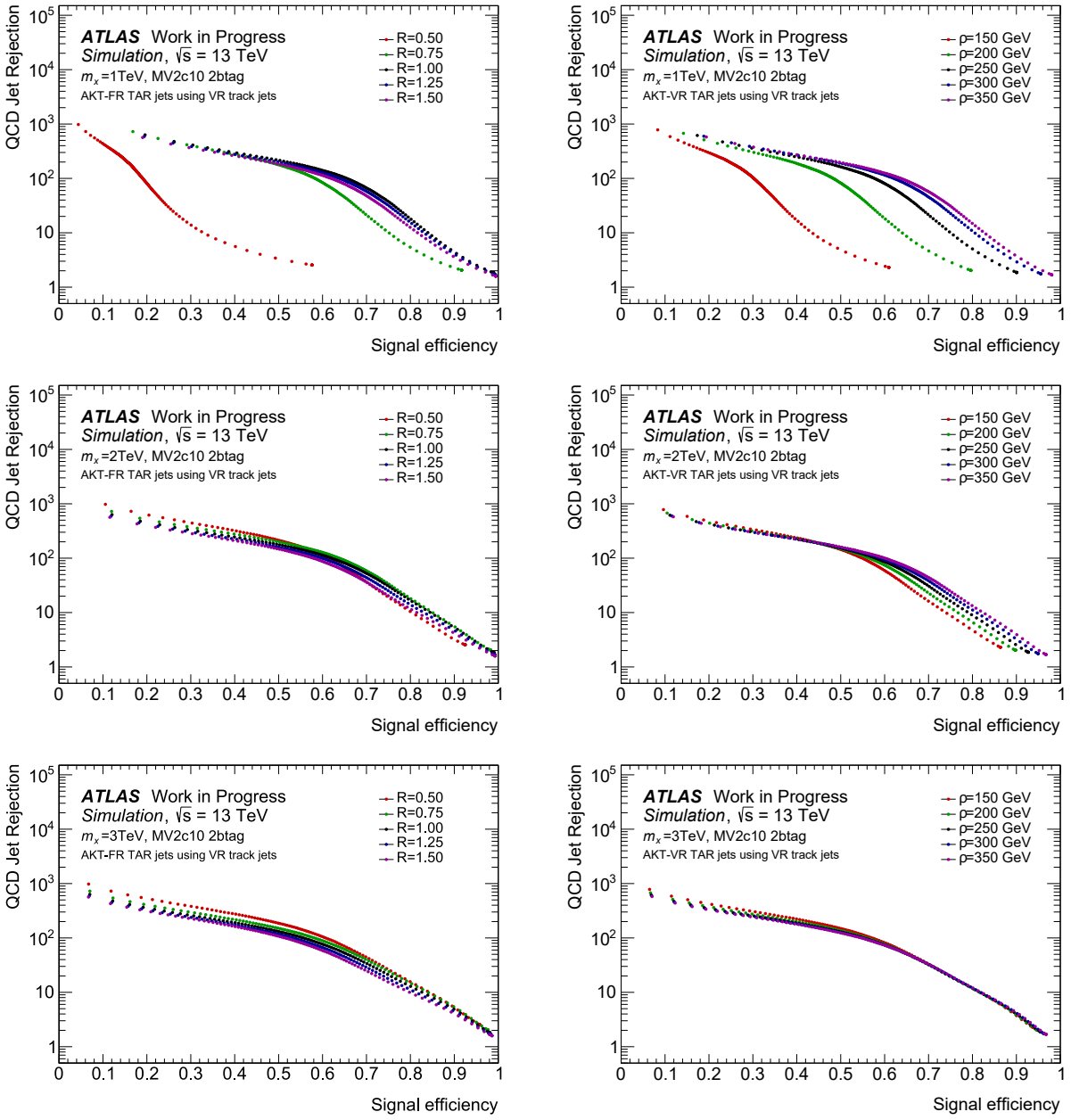
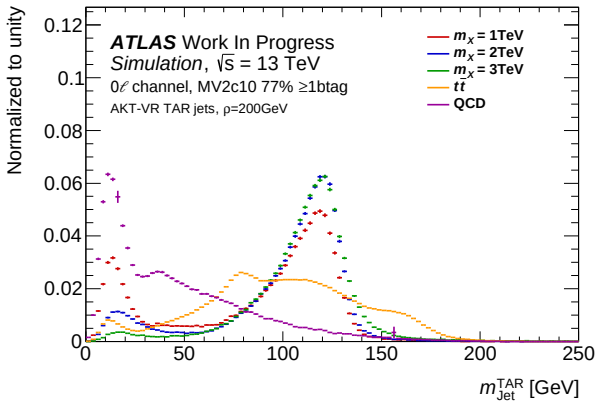
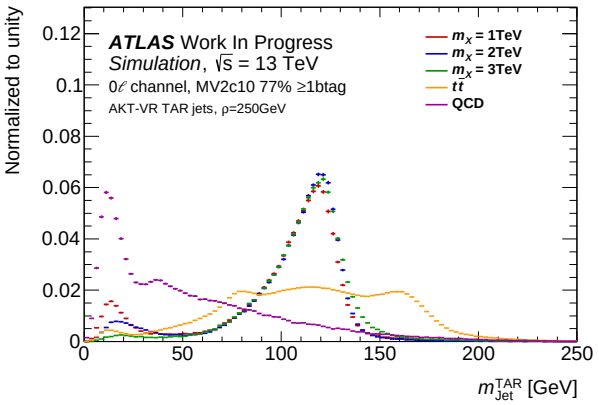


Fig. 60: ROC curves comparing the Higgs jet tagging performance against QCD jets in the double b-tag category for AKT-FR TAR jets (left) and AKT-VR TAR jets (right) for the $m_X = 1\,\text{TeV}$ (first row), $m_X = 2\,\text{TeV}$ (second row) and $m_X = 1\,\text{TeV}$ (third row) resonance mass. For AKT-FR TAR jets, the size parameter is varied, while the energy scale is varied for AKT-VR TAR jets.

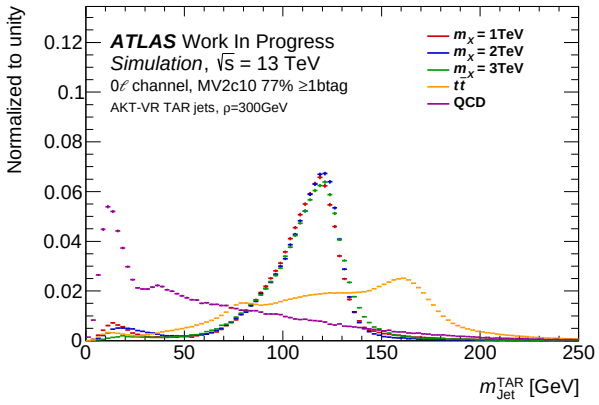
A.4. Track-Assisted Reclustered Jet Mass Cut



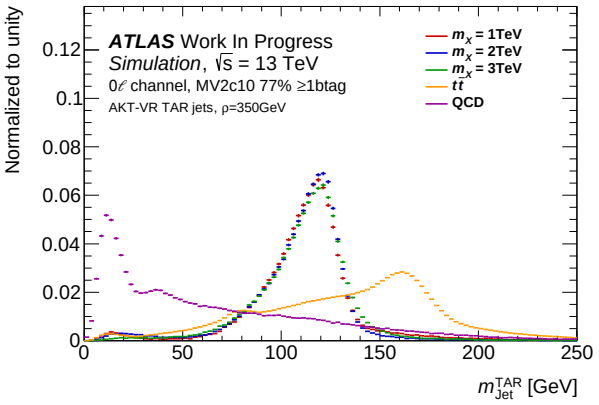
(a) $\rho = 200\,\text{GeV}$



(b) $\rho = 250\,\text{GeV}$

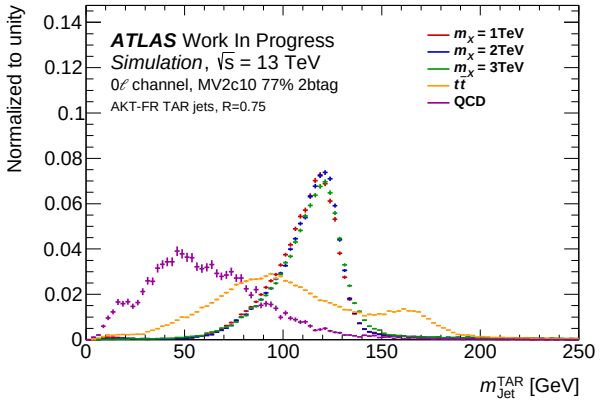


(c) $\rho = 300\,\text{GeV}$

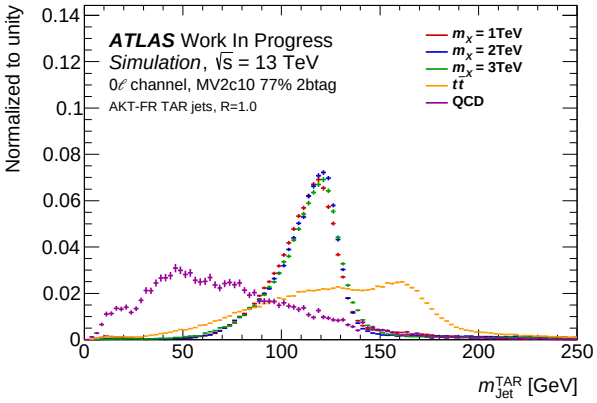


(d) $\rho = 350\,\text{GeV}$

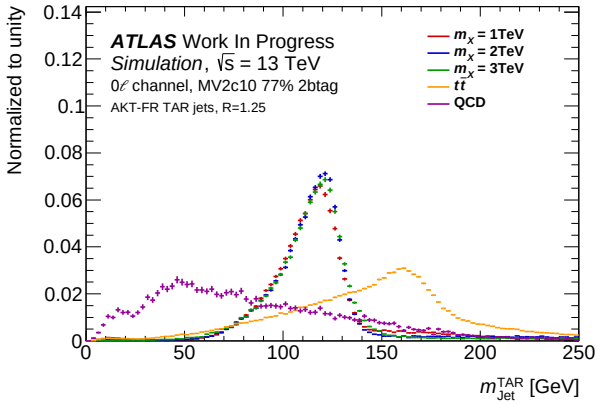
Fig. 61: Masses of top, QCD and Higgs jets originating from heavy resonances of mass m_X, reconstructed using the AKT-VR TAR jet algorithm. Distributions are shown for various values of the energy scale in the inclusive single b-tag category.



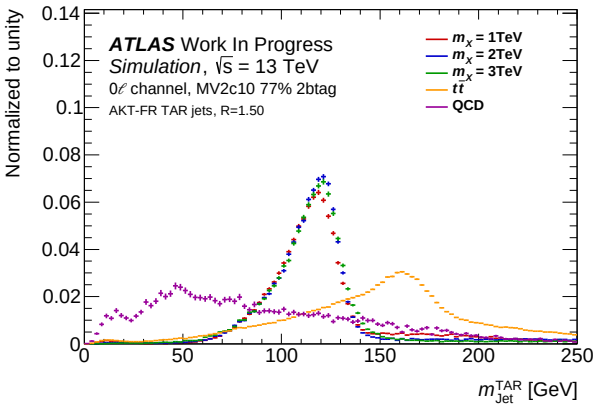
(a) $R = 0.75$



(b) $R = 1.0$



(c) $R = 1.25$



(d) $R = 1.50$

Fig. 62: Masses of top, QCD and Higgs jets originating from heavy resonances of mass m_X, reconstructed using the AKT-FR TAR jet algorithm. Distributions are shown for various values of the size parameter in the double b-tag category.

A.5. Track-Assisted Reclustered Jet Substructure Variables

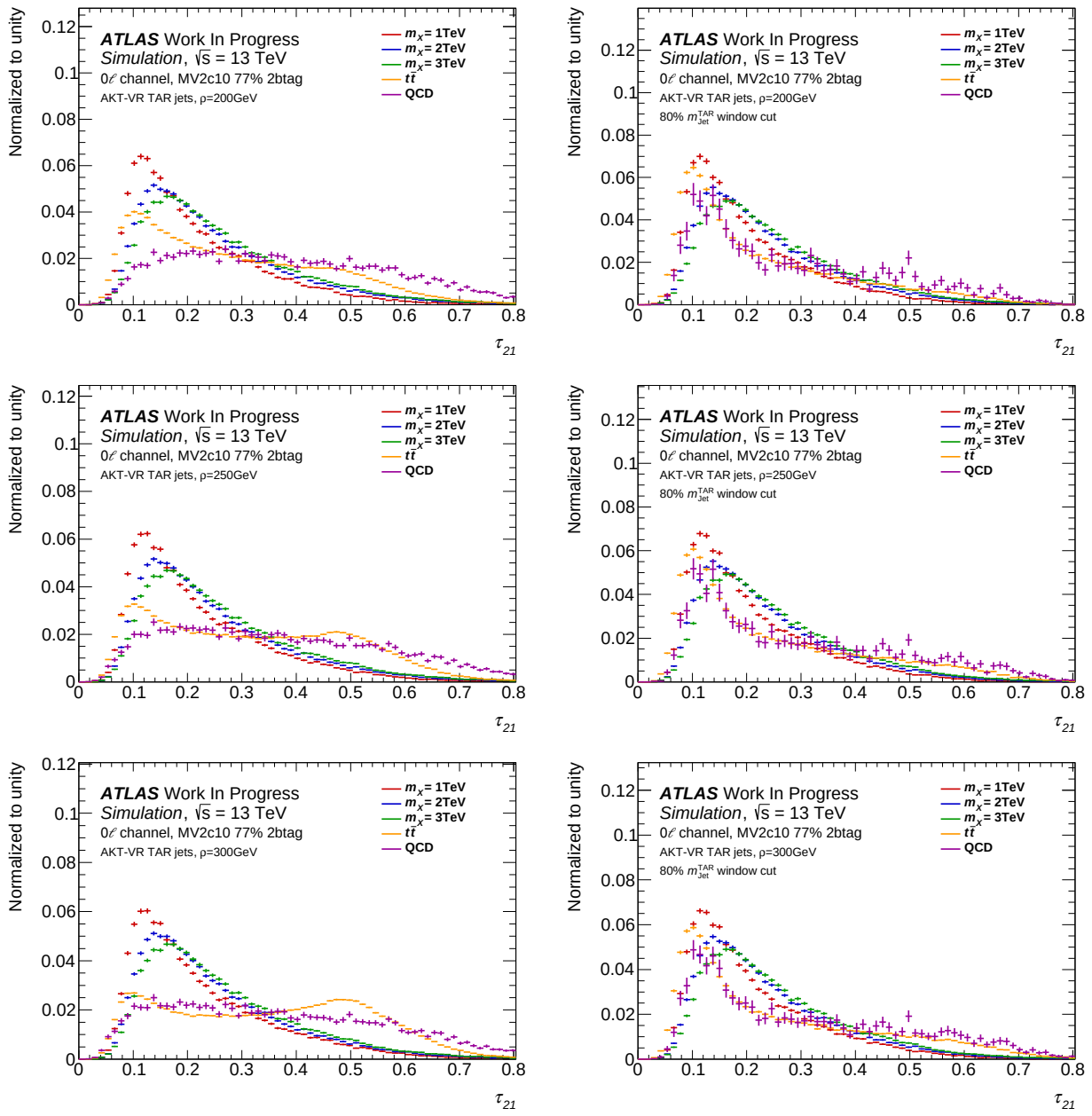


Fig. 63: N-subjettiness ratio τ_{21} in the double b-tag category for QCD and top jets as well as for Higgs jets for various values of m_X reconstructed using AKT-VR TAR jets. Distributions are shown without (left) and with an 80% $m_{\text{Jet}}^{\text{TAR}}$ cut (right) for the mass scale values $\rho = 200\,\text{GeV}$ (first row), $\rho = 250\,\text{GeV}$ (second row) and $\rho = 300\,\text{GeV}$ (third row).

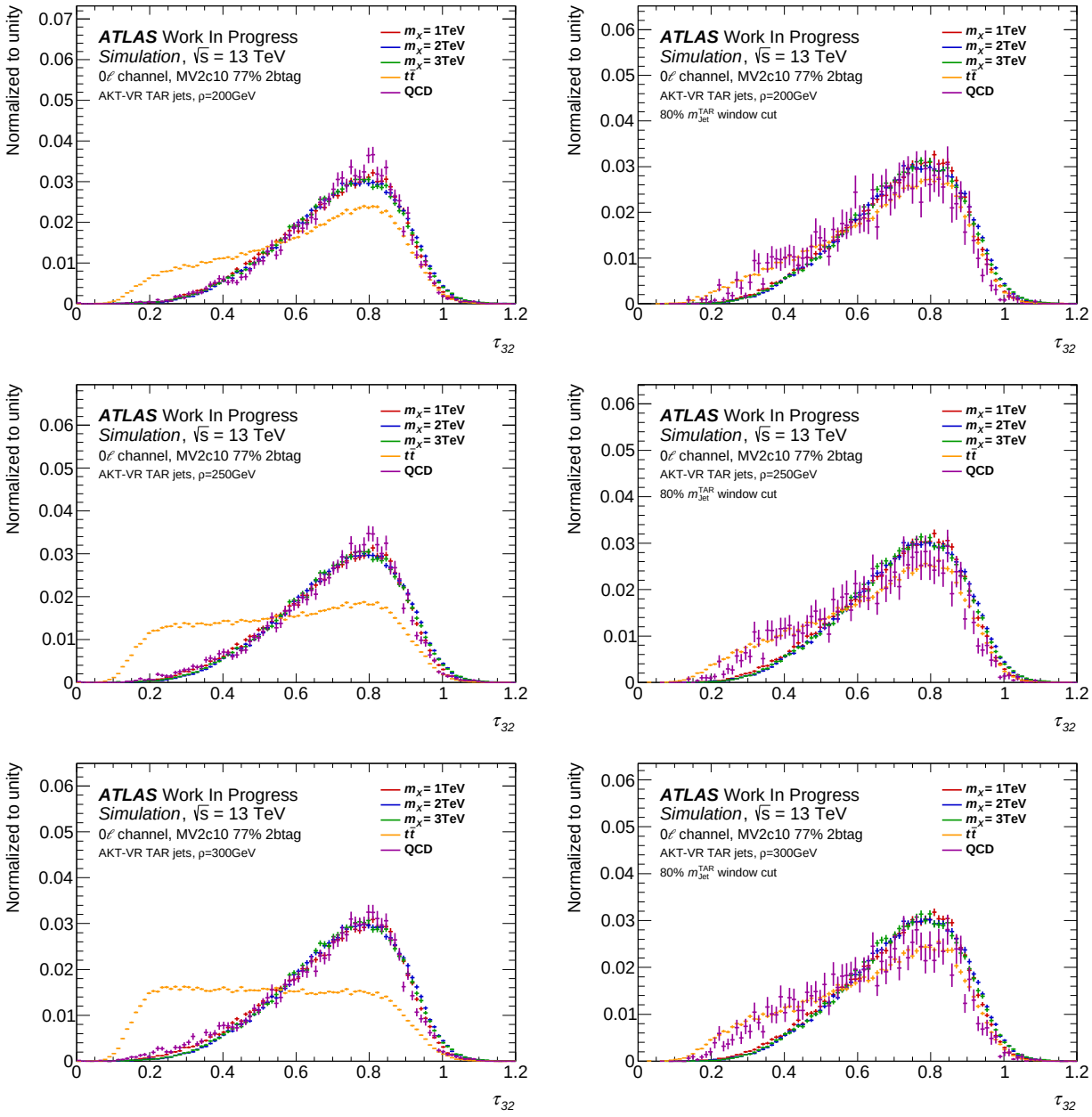


Fig. 64: N-subjettiness ratio τ_{32} in the double b-tag category for QCD and top jets as well as for Higgs jets for various values of m_X reconstructed using AKT-VR TAR jets. Distributions are shown without (left) and with an 80% $m_{\mathrm{Jet}}^{\mathrm{TAR}}$ cut (right) for the mass scale values $\rho = 200\,\mathrm{GeV}$ (first row), $\rho = 250\,\mathrm{GeV}$ (second row) and $\rho = 300\,\mathrm{GeV}$ (third row).

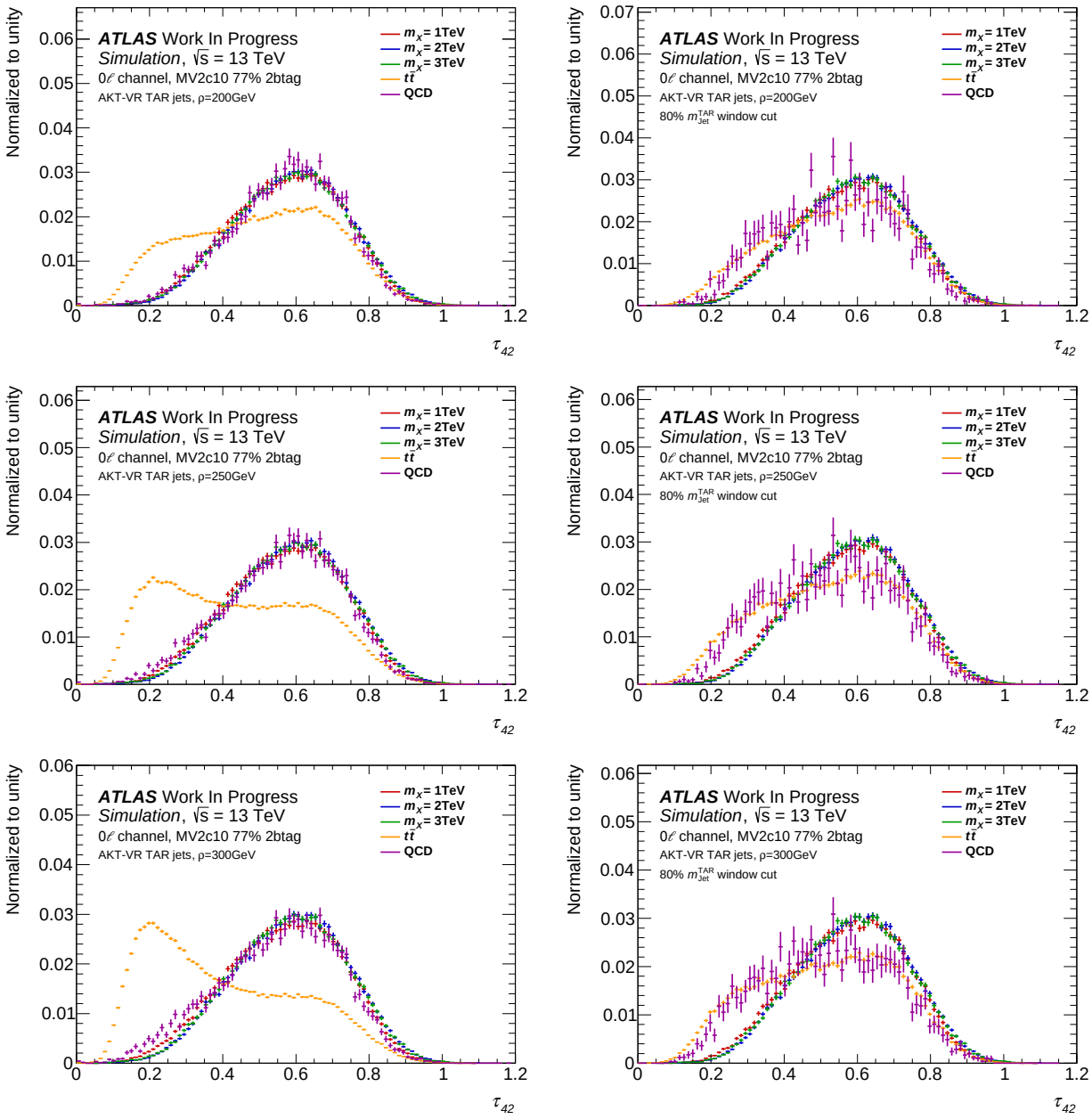


Fig. 65: N-subjettiness ratio τ_{42} in the double b-tag category for QCD and top jets as well as for Higgs jets for various values of m_X reconstructed using AKT-VR TAR jets. Distributions are shown without (left) and with an 80% $m_{\text{Jet}}^{\text{TAR}}$ cut (right) for the mass scale values $\rho = 200\,\text{GeV}$ (first row), $\rho = 250\,\text{GeV}$ (second row) and $\rho = 300\,\text{GeV}$ (third row).

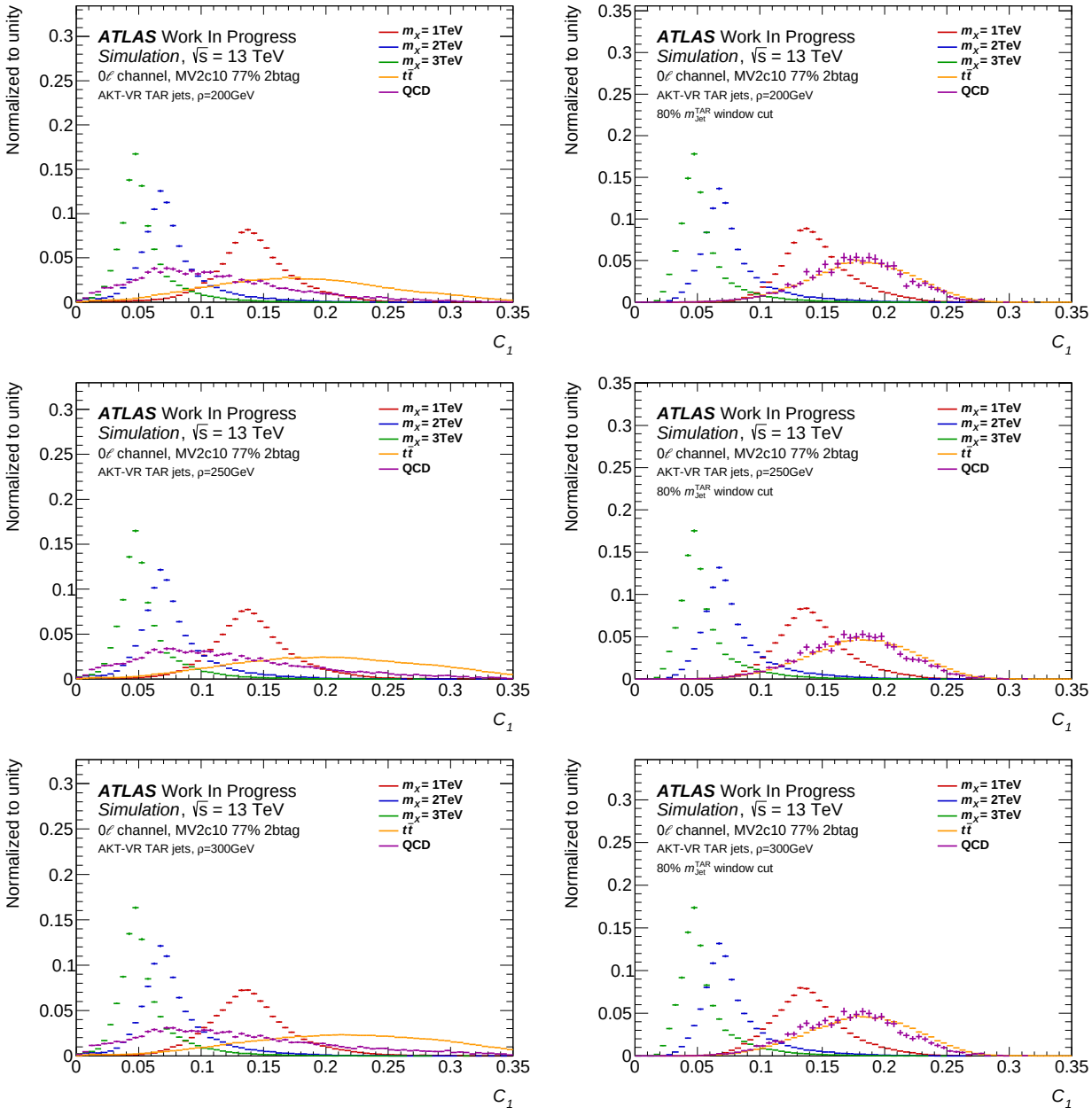


Fig. 66: ECF ratio C_1 in the double b-tag category for QCD and top jets as well as for Higgs jets for various values of m_X reconstructed using AKT-VR TAR jets. Distributions are shown without (left) and with an 80% $m_{\mathrm{Jet}}^{\mathrm{TAR}}$ cut (right) for the mass scale values $\rho = 200\,\mathrm{GeV}$ (first row), $\rho = 250\,\mathrm{GeV}$ (second row) and $\rho = 300\,\mathrm{GeV}$ (third row).

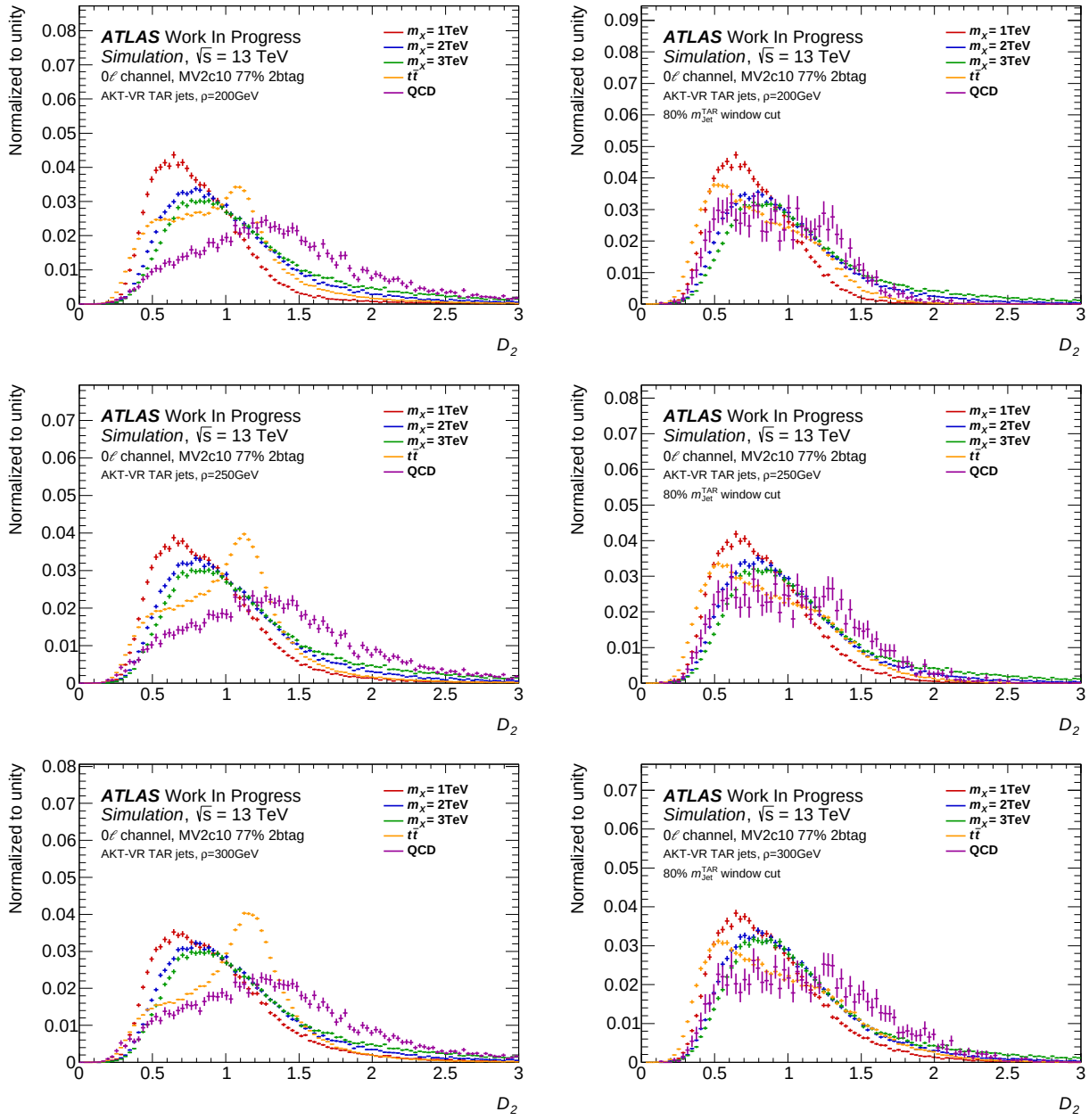


Fig. 67: ECF ratio D_2 in the double b-tag category for QCD and top jets as well as for Higgs jets for various values of m_X reconstructed using AKT-VR TAR jets. Distributions are shown without (left) and with an 80% $m_{\text{Jet}}^{\text{TAR}}$ cut (right) for the mass scale values $\rho = 200\,\text{GeV}$ (first row), $\rho = 250\,\text{GeV}$ (second row) and $\rho = 300\,\text{GeV}$ (third row).

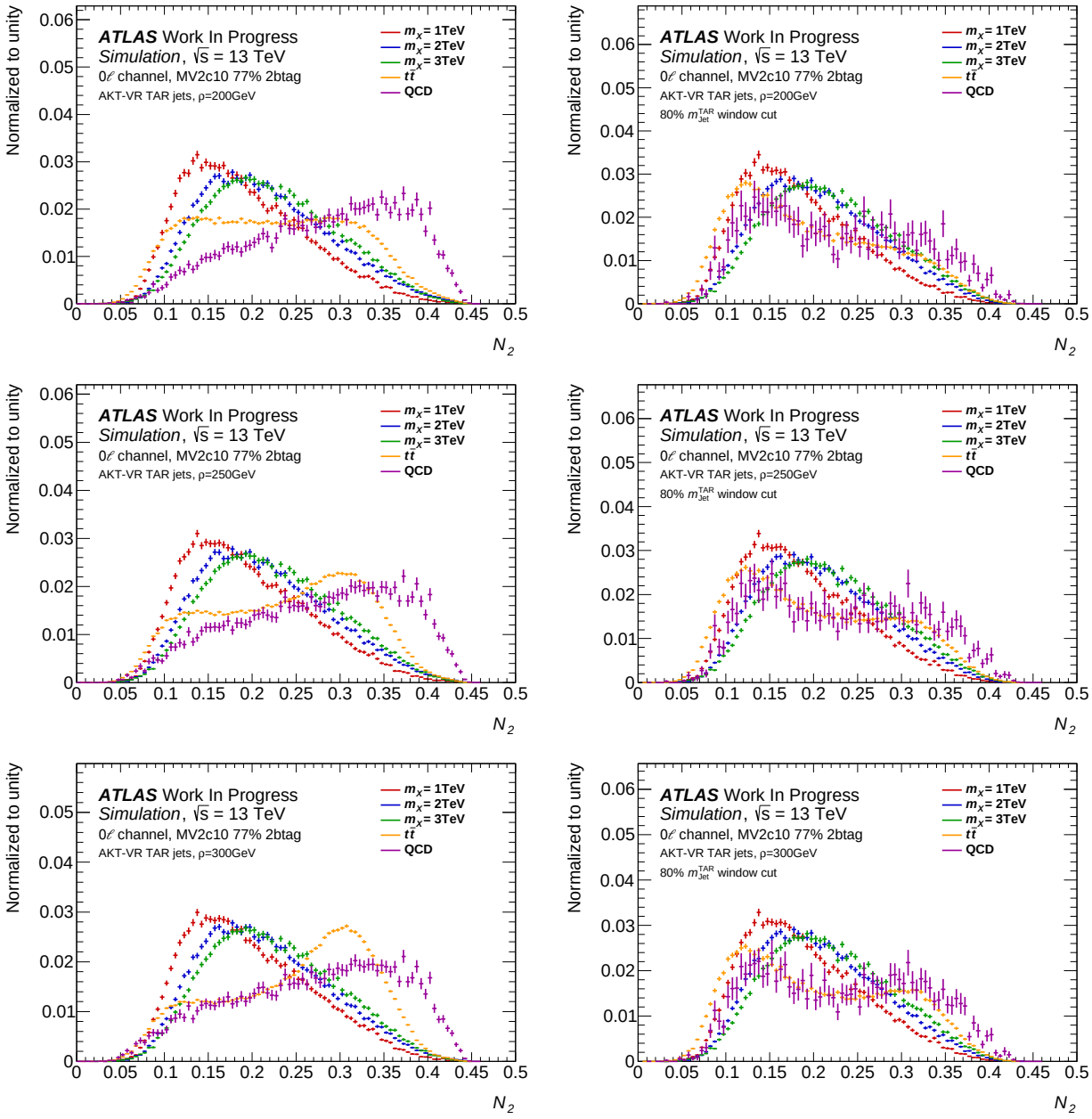


Fig. 68: N_2 distribution in the double b-tag category for QCD and top jets as well as for Higgs jets for various values of m_X reconstructed using AKT-VR TAR jets. Distributions are shown without (left) and with an 80% $m_{\mathrm{Jet}}^{\mathrm{TAR}}$ cut (right) for the mass scale values $\rho = 200\,\mathrm{GeV}$ (first row), $\rho = 250\,\mathrm{GeV}$ (second row) and $\rho = 300\,\mathrm{GeV}$ (third row).

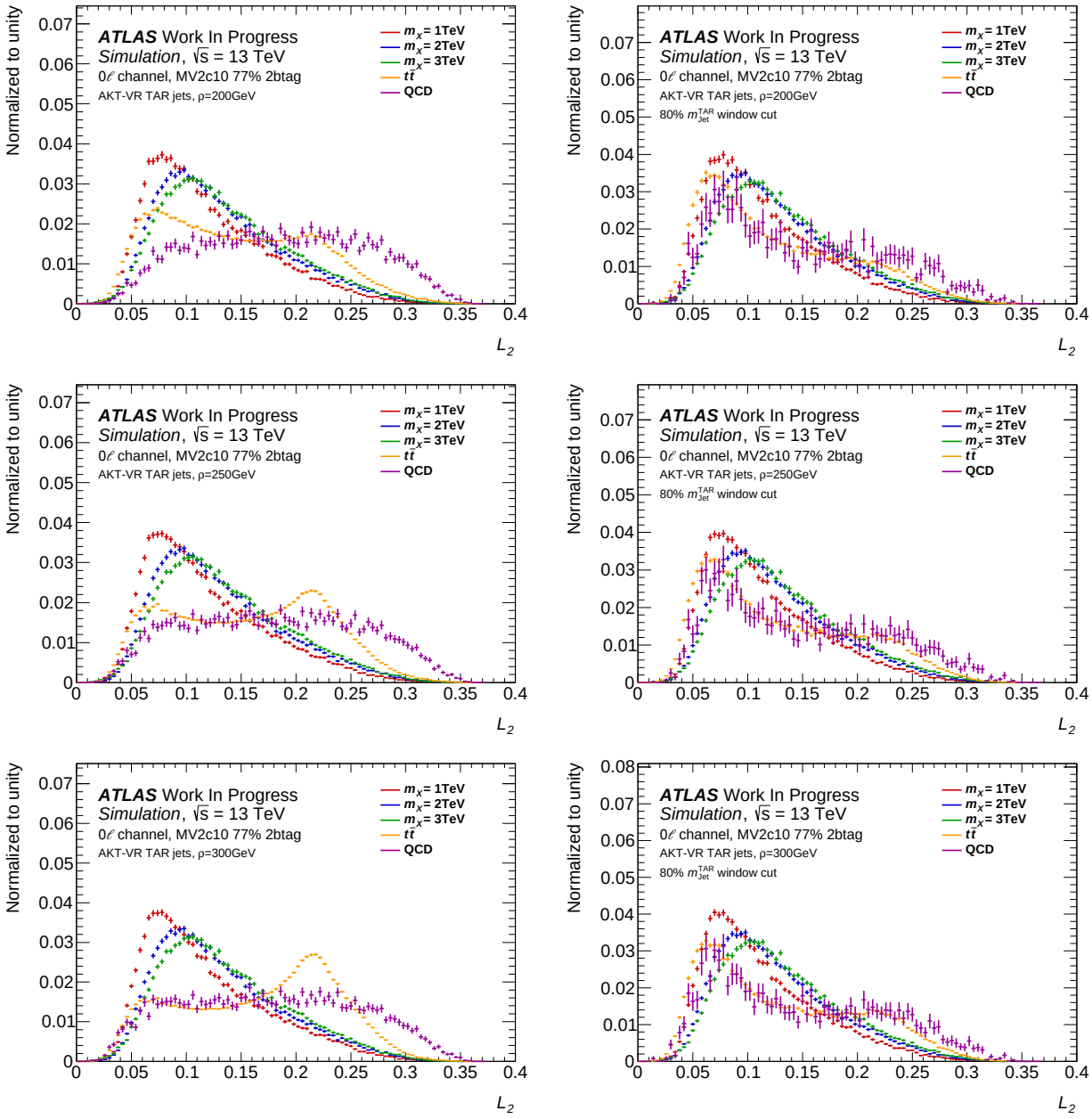


Fig. 69: L_2 distribution in the double b-tag category for QCD and top jets as well as for Higgs jets for various values of m_X reconstructed using AKT-VR TAR jets. Distributions are shown without (left) and with an 80% $m_{\text{Jet}}^{\text{TAR}}$ cut (right) for the mass scale values $\rho = 200\,\text{GeV}$ (first row), $\rho = 250\,\text{GeV}$ (second row) and $\rho = 300\,\text{GeV}$ (third row).

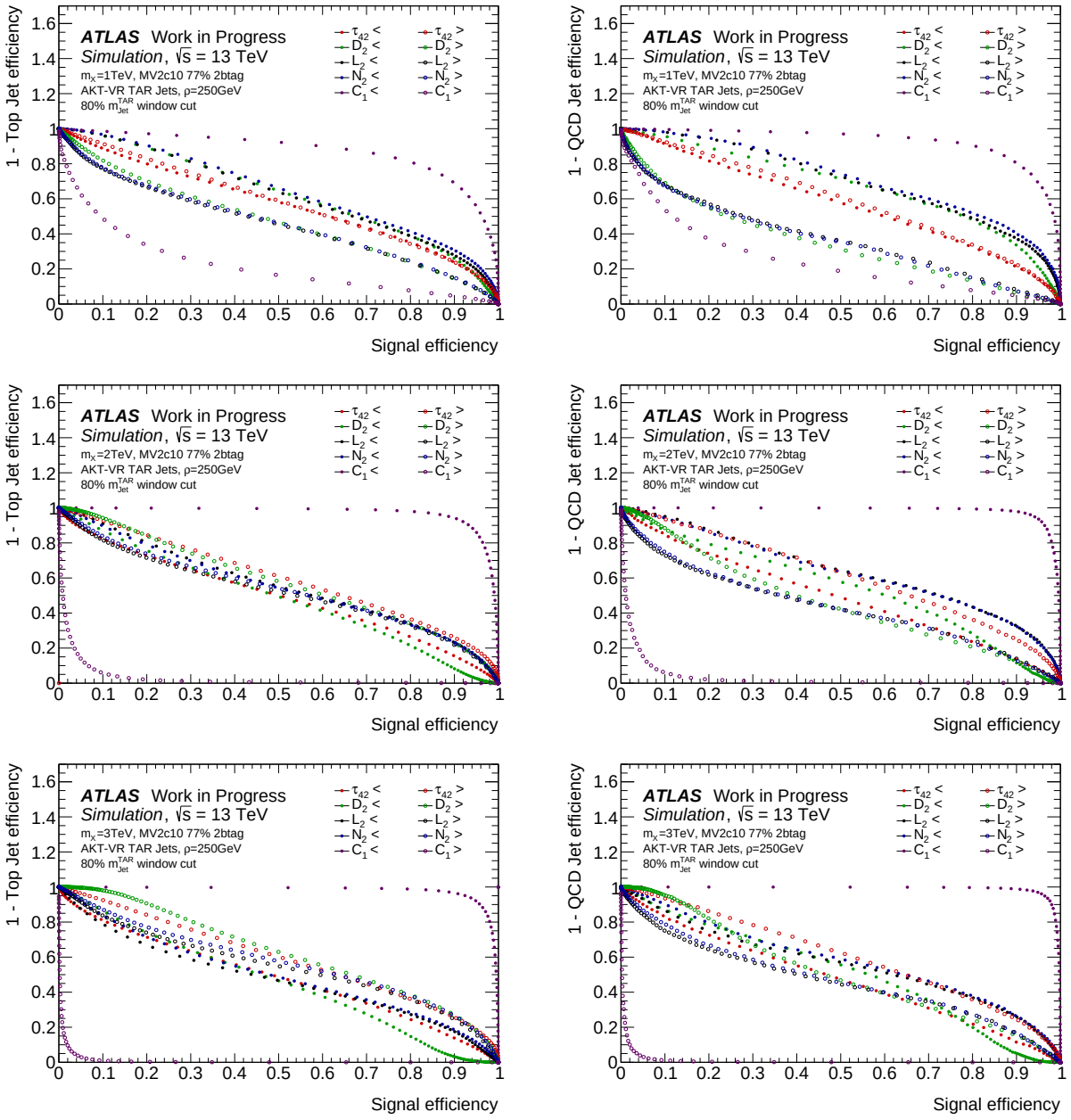


Fig. 70: ROC curves comparing the performance of lower and upper cuts on various substructure variables for AKT-VR TAR jets with $\rho = 250\,\text{GeV}$ in the case of top jets (left) and QCD jets (right) for a $m_X = 1\,\text{TeV}$ resonance (first row), a $m_X = 2\,\text{TeV}$ resonance (second row) and a $m_X = 3\,\text{TeV}$ resonance (third row). An 80% TAR jet mass window cut is applied.

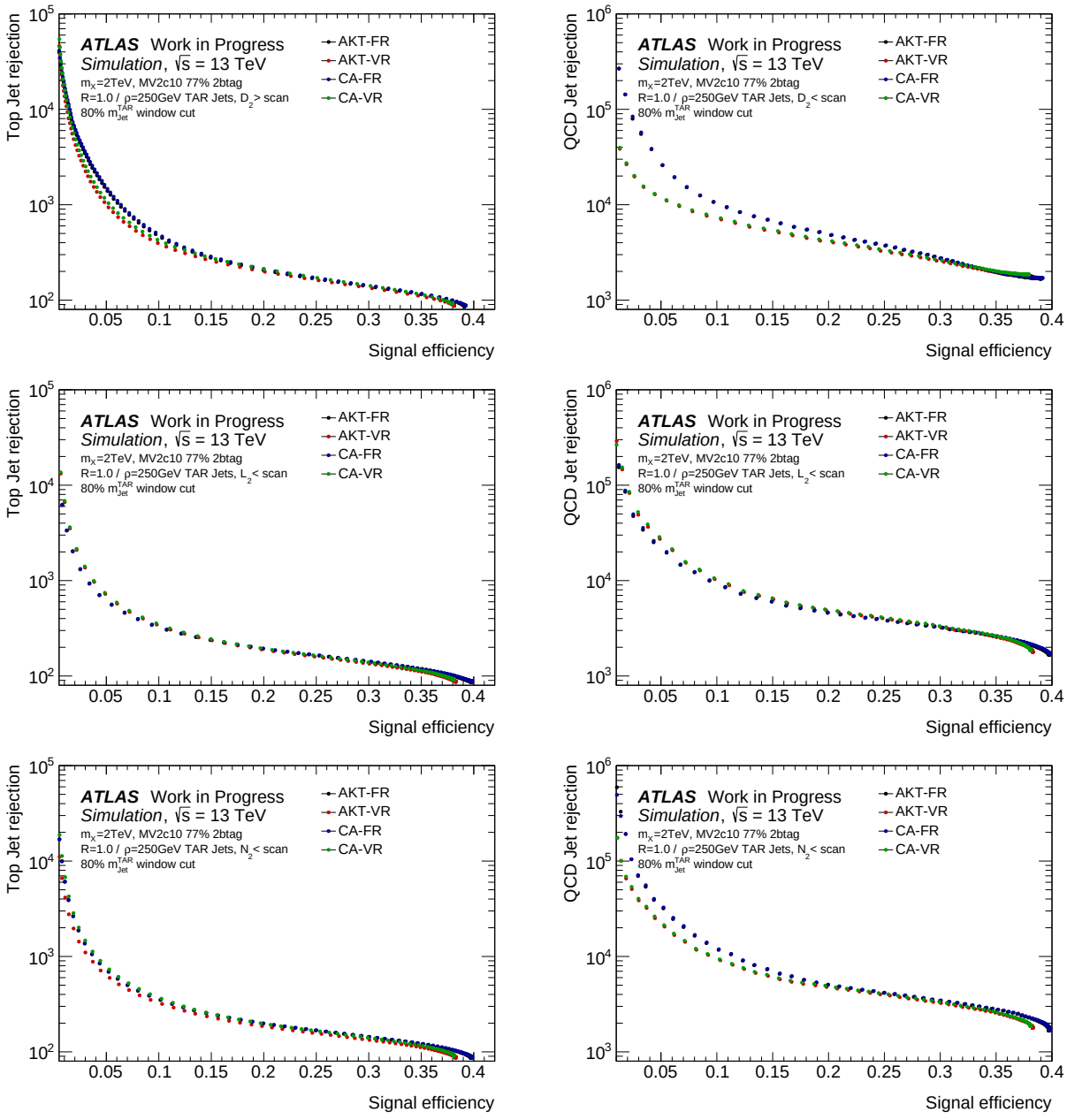


Fig. 71: ROC curves comparing the performance of the FR ($R = 1.0$) and VR ($\rho = 250\,\text{GeV}$) TAR jet collections for a cut on D_2 (first row), L_2 (second row) and N_2 (third row) in the case of top jets (left) and QCD jets (right). An 80% TAR jet mass cut is applied.

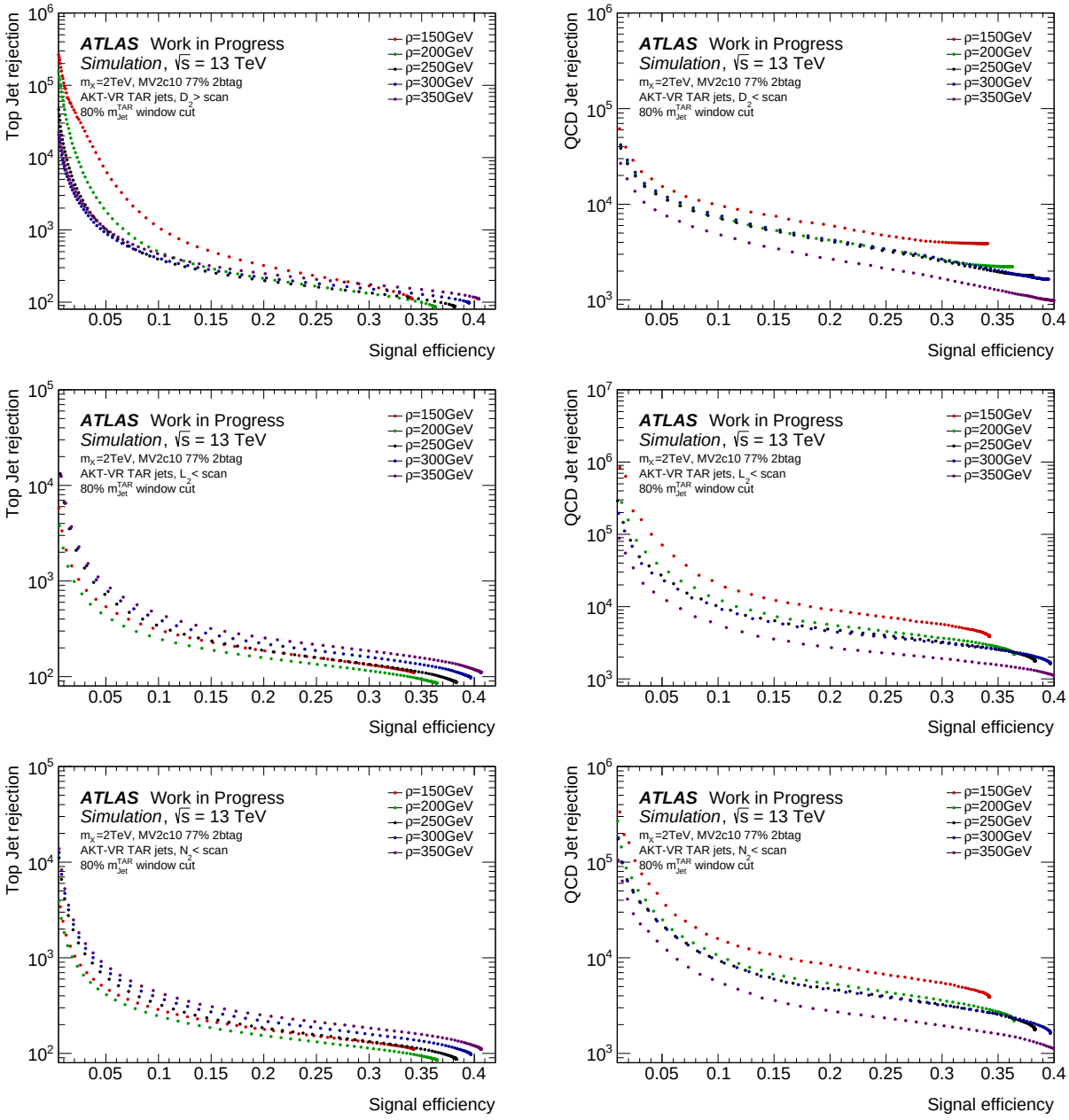


Fig. 72: ROC curves comparing the performance of the various energy scales for AKT-VR TAR jets for a cut on D_2 (first row), L_2 (second row) and N_2 (third row) in the case of top jets (left) and QCD jets (right). An 80% TAR jet mass cut is applied. Note that the cut on D_2 is inverted for top and QCD jets.

Appendix B

Additional Tables

R	Selection		1 TeV	2 TeV	3 TeV	$t\bar{t}$	QCD [10^{-6}]
0.50	N_{ini}		989.54	137.56	138.62	33258526	1000000
	jet selection		563.94	106.90	113.55	2364135	227609
	double b-tag		77.71	47.99	46.40	28784	849
	mass cut	68%	48.96	34.52	31.55	3959	19
		80%	57.77	40.18	37.14	5850	32
		90%	65.39	44.24	41.73	9197	71
0.75	N_{ini}		989.54	137.56	138.62	33271372	1000000
	jet selection		593.11	107.51	113.80	3425011	246401
	double b-tag		280.22	53.06	47.56	75787	12046
	mass cut	68%	202.07	38.03	32.10	21781	117
		80%	234.96	43.99	37.68	30082	176
		90%	259.61	48.30	42.37	42360	294
1.00	N_{ini}		989.54	137.56	138.62	33271372	1000000
	jet selection		604.01	107.76	113.88	4219907	260672
	double b-tag		330.05	53.24	47.27	130776	1435
	mass cut	68%	231.95	37.12	31.11	36393	243
		80%	269.98	42.82	36.55	49359	331
		90%	302.31	47.08	41.06	85376	518
1.25	N_{ini}		989.54	137.56	138.62	33254289	1000000
	jet selection		604.77	107.77	113.88	4763263	272205
	double b-tag		327.12	52.37	46.41	170288	1605
	mass cut	68%	219.08	34.80	29.21	34200	285
		80%	255.16	40.05	34.33	48456	398
		90%	293.20	44.40	38.59	118194	672
1.50	N_{ini}		989.54	137.56	138.62	33271372	1000000
	jet selection		604.87	107.77	113.87	4973644	278459
	double b-tag		320.83	51.20	45.49	188762	17081
	mass cut	68%	207.09	32.69	27.36	32337	290
		80%	240.91	37.55	32.18	46866	405
		90%	281.13	41.79	36.17	125720	737

Tab. 11: Cutflow showing the event yields after every step of the selection in the double b-tag category (77% WP) for the signal and background contributions for the various size parameters of AKT-FR TAR jets. As a full background estimate was not available during the time of the writing, the yields for QCD are normalized to one.

Appendix C

MC Weighting Procedure

In order to obtain yields corresponding to the physical expected event counts, events in the MC samples are weighted with

$$w = \frac{w_{\mathrm{MC}} \cdot \sigma \cdot \epsilon_f \cdot k \cdot L}{N}, \tag{C.1}$$

where w_{MC} is the MC event weight, σ the cross section of the simulated sample, ϵ_f the filter efficiency, L the integrated luminosity and $k = \sigma_{\mathrm{NLO}}/\sigma_{\mathrm{LO}}$ the k-factor. The definition of N is generator dependent and can be either the number of generated (unweighted) events or the sum of weights (w_{MC}) in the corresponding sample. The filter efficiency encapsulates the efficiency for generator-level cuts. The k-factor encodes some information from NLO calculations and is only applied to the LO $t\bar{t}$ sample. The di-jet QCD samples are split into p_T bins in order to enhance the statistics in the high p_T region of the p_T spectrum of the di-jet system. The cross sections are therefore dependent on the p_T bins such that the weighting yields a smoothly falling, physical p_T spectrum for the QCD jets. The transverse momentum spectrum of the p_T-leading AKT-FR ($R = 1.0$) TAR jet in the weighted QCD di-jet sample is shown in Figure 73.

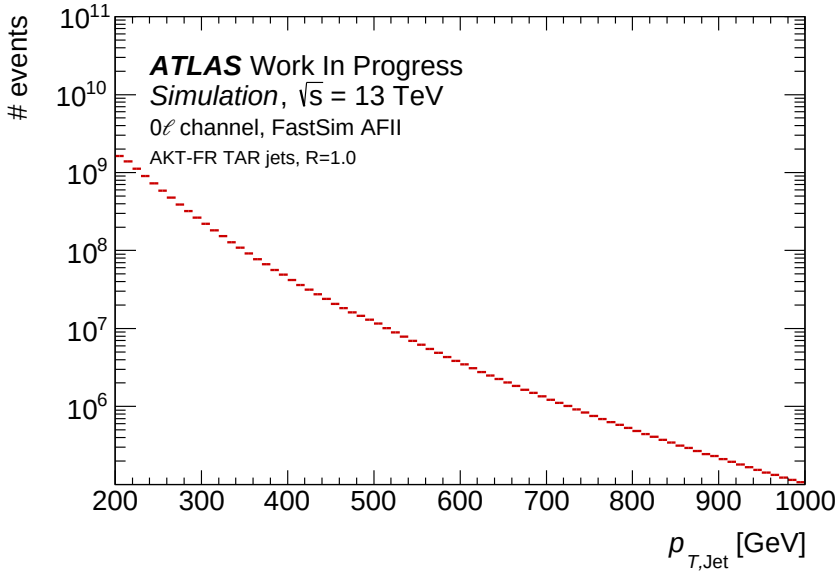


Fig. 73: Transverse momentum spectrum of the p_T-leading AKT-FR TAR jet in the weighted QCD di-jet sample.

Appendix D

MC Samples

m_X	Sample name
1 TeV	mc16_13TeV.450190.MadGraphHerwig7EvtGen_PDF23LO_X1000tohh_WWbb_allhad.deriv.DAOD_EXOT8.e7201_a875_r9364_p3895
	mc16_13TeV.450190.MadGraphHerwig7EvtGen_PDF23LO_X1000tohh_WWbb_allhad.deriv.DAOD_EXOT8.e7201_a875_r10201_p3895
	mc16_13TeV.450190.MadGraphHerwig7EvtGen_PDF23LO_X1000tohh_WWbb_allhad.deriv.DAOD_EXOT8.e7201_a875_r10724_p3895
2 TeV	mc16_13TeV.450200.MadGraphHerwig7EvtGen_PDF23LO_X2000tohh_WWbb_allhad.deriv.DAOD_EXOT8.e7201_a875_r9364_p3895
	mc16_13TeV.450200.MadGraphHerwig7EvtGen_PDF23LO_X2000tohh_WWbb_allhad.deriv.DAOD_EXOT8.e7201_a875_r10201_p3895
	mc16_13TeV.450200.MadGraphHerwig7EvtGen_PDF23LO_X2000tohh_WWbb_allhad.deriv.DAOD_EXOT8.e7201_a875_r10724_p3895
3 TeV	mc16_13TeV.450210.MadGraphHerwig7EvtGen_PDF23LO_X3000tohh_WWbb_allhad.deriv.DAOD_EXOT8.e7201_a875_r9364_p3895
	mc16_13TeV.450210.MadGraphHerwig7EvtGen_PDF23LO_X3000tohh_WWbb_allhad.deriv.DAOD_EXOT8.e7201_a875_r10201_p3895
	mc16_13TeV.450210.MadGraphHerwig7EvtGen_PDF23LO_X3000tohh_WWbb_allhad.deriv.DAOD_EXOT8.e7201_a875_r10724_p3895

Tab. 12: Names of the signal samples used for the various mass points m_X.

Background	Sample name
QCD	mc16_13TeV.426131.Sherpa_CT10_jets_JZ1.deriv.DAOD_EXOT8.e4355_s3126_r9364_p3895
	mc16_13TeV.426132.Sherpa_CT10_jets_JZ2.deriv.DAOD_EXOT8.e4355_s3126_r9364_p3895
	mc16_13TeV.426133.Sherpa_CT10_jets_JZ3.deriv.DAOD_EXOT8.e4355_s3126_r9364_p3895
	mc16_13TeV.426134.Sherpa_CT10_jets_JZ4.deriv.DAOD_EXOT8.e4355_s3126_r9364_p3895
	mc16_13TeV.426135.Sherpa_CT10_jets_JZ5.deriv.DAOD_EXOT8.e4355_s3126_r9364_p3895
	mc16_13TeV.426136.Sherpa_CT10_jets_JZ6.deriv.DAOD_EXOT8.e4355_s3126_r9364_p3895
	mc16_13TeV.426137.Sherpa_CT10_jets_JZ7.deriv.DAOD_EXOT8.e4635_s3126_r9364_p3895
	mc16_13TeV.426138.Sherpa_CT10_jets_JZ8.deriv.DAOD_EXOT8.e4635_s3126_r9364_p3895
	mc16_13TeV.426139.Sherpa_CT10_jets_JZ9.deriv.DAOD_EXOT8.e4635_s3126_r9364_p3895
	mc16_13TeV.426140.Sherpa_CT10_jets_JZ10.deriv.DAOD_EXOT8.e4635_s3126_r9364_p3895
	mc16_13TeV.426141.Sherpa_CT10_jets_JZ11.deriv.DAOD_EXOT8.e4635_s3126_r9364_p3895
	mc16_13TeV.426142.Sherpa_CT10_jets_JZ12.deriv.DAOD_EXOT8.e4635_s3126_r9364_p3895
	mc16_13TeV.426131.Sherpa_CT10_jets_JZ1.deriv.DAOD_EXOT8.e4355_s3126_r10201_p3895
	mc16_13TeV.426132.Sherpa_CT10_jets_JZ2.deriv.DAOD_EXOT8.e4355_s3126_r10201_p3895
	mc16_13TeV.426133.Sherpa_CT10_jets_JZ3.deriv.DAOD_EXOT8.e4355_s3126_r10201_p3895
	mc16_13TeV.426134.Sherpa_CT10_jets_JZ4.deriv.DAOD_EXOT8.e4355_s3126_r10201_p3895
	mc16_13TeV.426135.Sherpa_CT10_jets_JZ5.deriv.DAOD_EXOT8.e4355_s3126_r10201_p3895
	mc16_13TeV.426136.Sherpa_CT10_jets_JZ6.deriv.DAOD_EXOT8.e4355_s3126_r10201_p3895
	mc16_13TeV.426137.Sherpa_CT10_jets_JZ7.deriv.DAOD_EXOT8.e4635_s3126_r10201_p3895
	mc16_13TeV.426138.Sherpa_CT10_jets_JZ8.deriv.DAOD_EXOT8.e4635_s3126_r10201_p3895
	mc16_13TeV.426139.Sherpa_CT10_jets_JZ9.deriv.DAOD_EXOT8.e4635_s3126_r10201_p3895
	mc16_13TeV.426140.Sherpa_CT10_jets_JZ10.deriv.DAOD_EXOT8.e4635_s3126_r10201_p3895
	mc16_13TeV.426141.Sherpa_CT10_jets_JZ11.deriv.DAOD_EXOT8.e4635_s3126_r10201_p3895
	mc16_13TeV.426142.Sherpa_CT10_jets_JZ12.deriv.DAOD_EXOT8.e4635_s3126_r10201_p3895
	mc16_13TeV.426131.Sherpa_CT10_jets_JZ1.deriv.DAOD_EXOT8.e4355_e5984_s3126_r10724_r10726_p3895
	mc16_13TeV.426132.Sherpa_CT10_jets_JZ2.deriv.DAOD_EXOT8.e4355_e5984_s3126_r10724_r10726_p3895
	mc16_13TeV.426133.Sherpa_CT10_jets_JZ3.deriv.DAOD_EXOT8.e4355_e5984_s3126_r10724_r10726_p3895
	mc16_13TeV.426134.Sherpa_CT10_jets_JZ4.deriv.DAOD_EXOT8.e4355_e5984_s3126_r10724_r10726_p3895
	mc16_13TeV.426135.Sherpa_CT10_jets_JZ5.deriv.DAOD_EXOT8.e4355_e5984_s3126_r10724_r10726_p3895
	mc16_13TeV.426136.Sherpa_CT10_jets_JZ6.deriv.DAOD_EXOT8.e4355_e5984_s3126_r10724_r10726_p3895
	mc16_13TeV.426137.Sherpa_CT10_jets_JZ7.deriv.DAOD_EXOT8.e4635_e5984_s3126_r10724_r10726_p3895
	mc16_13TeV.426138.Sherpa_CT10_jets_JZ8.deriv.DAOD_EXOT8.e4635_e5984_s3126_r10724_r10726_p3895
	mc16_13TeV.426139.Sherpa_CT10_jets_JZ9.deriv.DAOD_EXOT8.e4635_e5984_s3126_r10724_r10726_p3895
	mc16_13TeV.426140.Sherpa_CT10_jets_JZ10.deriv.DAOD_EXOT8.e4635_e5984_s3126_r10724_r10726_p3895
	mc16_13TeV.426141.Sherpa_CT10_jets_JZ11.deriv.DAOD_EXOT8.e4635_e5984_s3126_r10724_r10726_p3895
	mc16_13TeV.426142.Sherpa_CT10_jets_JZ12.deriv.DAOD_EXOT8.e4635_e5984_s3126_r10724_r10726_p3895
$t\bar{t}$	mc16_13TeV.410471.PhPy8EG_A14_ttbar_hdamp258p75_allhad.deriv.DAOD_EXOT8.e6337_e5984_s3126_r9364_r9315_p3895
	mc16_13TeV.410471.PhPy8EG_A14_ttbar_hdamp258p75_allhad.deriv.DAOD_EXOT8.e6337_e5984_s3126_r10201_r10210_p3895
	mc16_13TeV.410471.PhPy8EG_A14_ttbar_hdamp258p75_allhad.deriv.DAOD_EXOT8.e6337_e5984_s3126_s3136_r10724_r10726_p3895

Tab. 13: Names of the used $t\bar{t}$ and QCD background samples.

Bibliography

[1] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961)

[2] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964)

[3] F. Englert, R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964)

[4] G. S. Guralnik, C. R. Hagen, T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13** (1964)

[5] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967)

[6] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. **C680519** (1968)

[7] UA1 Collaboration, *Experimental Observation of Lepton Pairs of Invariant Mass Around* 95 GeV/c^2 *at the CERN SPS Collider*, Phys. Lett. **126B** (1983)

[8] UA1 Collaboration, *Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at* $\sqrt{s} = 540$ *540 GeV*, Phys. Lett. **122B** (1983)

[9] 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. **B716** (2012)

[10] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. **B716** (2012)

[11] T. D. Lee, *A Theory of Spontaneous T Violation*, Phys. Rev. **D8** (1973)

[12] L. Randall, R. Sundrum, *A Large mass hierarchy from a small extra dimension*, Phys. Rev. Lett. **83** (1999)

[13] K. Agashe, et al., *Warped Gravitons at the LHC and Beyond*, Phys. Rev. **D76** (2007)

[14] A. L. Fitzpatrick, et al., *Searching for the Kaluza-Klein Graviton in Bulk RS Models*, JHEP **09** (2007)

[15] ATLAS Collaboration, *Track assisted techniques for jet substructure*, ATL-PHYS-PUB-2018-012 (2018)

[16] M. Tanabashi, et al. (Particle Data Group), *Review of Particle Physics*, Phys. Rev. **D98** (2018)

[17] E. Noether, *Invariant variation problems*, Transport Theory and Statistical Physics **1** (1971)

[18] K. G. Wilson, J. Kogut, *The renormalization group and the epsilon expansion*, Physics reports **12** (1974)

[19] K. G. Wilson, M. E. Fisher, *Critical exponents in 3.99 dimensions*, Phys. Rev. Lett. **28** (1972)

[20] K. G. Wilson, *Renormalization group and critical phenomena. I. Renormalization group and the Kadanoff scaling picture*, Phys. Rev. B **4** (1971)

[21] F. J. Wegner, A. Houghton, *Renormalization group equation for critical phenomena*, Phys. Rev. **A8** (1973)

[22] V. Trimble, *Existence and Nature of Dark Matter in the Universe*, Ann. Rev. Astron. Astrophys. **25** (1987)

[23] D. Clowe, M. Bradac, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, D. Zaritsky, *A direct empirical proof of the existence of dark matter*, Astrophys. J. **648** (2006)

[24] G. Jungman, M. Kamionkowski, K. Griest, *Supersymmetric dark matter*, Phys. Rept. **267** (1996)

[25] R. H. Cyburt, B. D. Fields, K. A. Olive, *Primordial nucleosynthesis in light of WMAP*, Phys. Lett. **B567** (2003)

[26] C. H. Llewellyn Smith, G. G. Ross, *The Real Gauge Hierarchy Problem*, Phys. Lett. **105B** (1981)

[27] J. Ellis, S. Kelley, D. V. Nanopoulos, *Precision LEP data, supersymmetric GUTs and string unification*, Phys. Lett. B **249** (1990)

[28] U. Amaldi, W. de Boer, H. Fürstenau, *Comparison of grand unified theories with electroweak and strong coupling constants measured at LEP*, Phys. Lett. B **260** (1991)

[29] J. Ellis, S. Kelley, D. V. Nanopoulos, *Probing the desert using gauge coupling unification*, Phys. Lett. B **260** (1991)

[30] C. Giunti, C. Kim, U. Lee, *Running coupling constants and grand unification models*, Mod. Phys. Lett A **6** (1991)

[31] P. Langacker, M. Luo, *Implications of precision electroweak experiments for M_t, ρ_0, $\sin^2\theta_W$, and grand unification*, Phys. Rev. D **44** (1991)

[32] G. C. Branco, et al., *Theory and phenomenology of two-Higgs-doublet models*, Phys. Rept. **516** (2012)

[33] M. Raidal, et al., *Flavour physics of leptons and dipole moments*, Eur. Phys. J. **C57** (2008)

[34] Yu. A. Golfand, E. P. Likhtman, *Extension of the Algebra of Poincare Group Generators and Violation of p Invariance*, JETP Lett. **13** (1971)

[35] J. Wess, B. Zumino, *Supergauge Transformations in Four-Dimensions*, Nucl. Phys. **B70** (1974)

[36] J. Wess, B. Zumino, *Supergauge Invariant Extension of Quantum Electrodynamics*, Nucl. Phys. **B78** (1974)

[37] S. Ferrara, B. Zumino, *Supergauge Invariant Yang-Mills Theories*, Nucl. Phys. **B79** (1974)

[38] A. Salam, J. A. Strathdee, *Supersymmetry and Nonabelian Gauges*, Phys. Lett. **51B** (1974)

[39] M. J. G. Veltman, *The Infrared - Ultraviolet Connection*, Acta Phys. Polon. **B12** (1981)

[40] R. K. Kaul, P. Majumdar, *Cancellation of Quadratically Divergent Mass Corrections in Globally Supersymmetric Spontaneously Broken Gauge Theories*, Nucl. Phys. **B199** (1982)

[41] D. Z. Freedman, P. van Nieuwenhuizen, S. Ferrara, *Progress Toward a Theory of Supergravity*, Phys. Rev. **D13** (1976)

[42] C. Csaki, *The Minimal supersymmetric standard model (MSSM)*, Mod. Phys. Lett. **A11** (1996)

[43] M. Cepeda, et al. (HL/HE WG2 group), *Higgs Physics at the HL-LHC and HE-LHC*, CERN-LPCC-2018-04 (2019)

[44] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, FERMILAB-FN-1025-T, CERN-2017-002-M (2016)

[45] The ATLAS and CMS Collaborations, *Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments*, Phys. Rev. Lett. **114** (2015)

[46] M. Grazzini, et al., *Higgs boson pair production at NNLO with top quark mass effects*, JHEP **05** (2018)

[47] J. Baglio, et al., *The measurement of the Higgs self-coupling at the LHC: theoretical status*, JHEP **04** (2013)

[48] ATLAS Collaboration, *Combination of searches for Higgs boson pairs in pp collisions at $\sqrt{s}$ =13 TeV with the ATLAS detector*, CERN-EP-2019-099 (2019)

[49] CMS Collaboration, *Combination of searches for Higgs boson pair production in proton-proton collisions at $\sqrt{s} = 13$ TeV*, Phys. Rev. Lett. **122** (2019)

[50] S. Höche, *Introduction to parton-shower event generators*, in *Journeys Through the Precision Frontier: Amplitudes for Colliders (TASI 2014)* (2015)

[51] Y. L. Dokshitzer, *Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics.*, Sov. Phys. JETP **46** (1977)

[52] V. N. Gribov, L. N. Lipatov, *Deep inelastic ep scattering in perturbation theory*, Sov. J. Nucl. Phys. **15** (1972)

[53] G. Altarelli, G. Parisi, *Asymptotic Freedom in Parton Language*, Nucl. Phys. **B126** (1977)

[54] J. Benecke, et al., *Hypothesis of limiting fragmentation in high-energy collisions*, Phys. Rev. **188** (1969)

[55] B. Andersson, et al., *Parton Fragmentation and String Dynamics*, Phys. Rept. **97** (1983)

[56] 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 **07** (2014)

[57] M. Bahr, et al., *Herwig++ Physics and Manual*, Eur. Phys. J. **C58** (2008)

[58] R. D. Ball, et al., *Parton distributions with LHC data*, Nucl. Phys. **B867** (2013)

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

[60] S. Alioli, et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010)

[61] T. Gleisberg, et al., *Event generation with SHERPA 1.1*, JHEP **02** (2009)

[62] S. Agostinelli, et al., *GEANT4: A Simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003)

[63] W. Lukas, *Fast Simulation for ATLAS: Atlfast-II and ISF*, J. Phys. Conf. Ser **396** (2012)

[64] O. Bruning, et al., *LHC Design Report Vol.1: The LHC Main Ring* (2004)

[65] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008)

[66] C. Collaboration, *The CMS Experiment at the CERN LHC*, JINST **3** (2008)

[67] LHCb Collaboration, *The LHCb detector at the LHC*, JINST **3** (2008)

[68] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, JINST **3** (2008)

[69] E. Mobs, *The CERN accelerator complex. Complexe des accélérateurs du CERN*, OPEN-PHO-ACCEL-2018-005 (2018)

[70] G. Apollinari, et al., *High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report*, CERN Yellow Reports: Monographs (2017)

[71] M. zur Nedden, *The LHC Run 2 ATLAS trigger system: design, performance and plans*, JINST **12** (2017)

[72] J. M. Butterworth, A. R. Davison, M. Rubin, G. P. Salam, *Jet substructure as a new Higgs search channel at the LHC*, Phys. Rev. Lett. **100** (2008)

[73] G. P. Salam, *Towards Jetography*, Eur. Phys. J. **C67** (2010)

[74] S. Catani, Y. L. Dokshitzer, M. H. Seymour, B. R. Webber, *Longitudinally invariant K_t clustering algorithms for hadron hadron collisions*, Nucl. Phys. **B406** (1993)

[75] M. Cacciari, G. P. Salam, G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008)

[76] Y. L. Dokshitzer, et al., *Better jet clustering algorithms*, JHEP **08** (1997)

[77] D. Krohn, J. Thaler, L.-T. Wang, *Jets with Variable R*, JHEP **06** (2009)

[78] M. Cacciari, G. P. Salam, G. Soyez, *FastJet User Manual*, Eur. Phys. J. **C72** (2012)

[79] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. C **77** (2017)

[80] D. Krohn, J. Thaler, L.-T. Wang, *Jet trimming*, JHEP **2010** (2010)

[81] ATLAS Collaboration, *Performance of jet substructure techniques for large-R jets in proton-proton collisions at $\sqrt{s}$ = 7 TeV using the ATLAS detector*, JHEP **09** (2013)

[82] ATLAS collaboration, *Jet mass reconstruction with the ATLAS Detector in early Run 2 data*, ATLAS-CONF-2016-035 (2016)

[83] J. Thaler, K. Van Tilburg, *Identifying boosted objects with N-subjettiness*, JHEP **2011** (2011)

[84] ATLAS Collaboration, *Identification of high transverse momentum top quarks in pp collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector*, JHEP **2016** (2016)

[85] ATLAS Collaboration, *Identification of boosted, hadronically decaying W bosons and comparisons with ATLAS data taken at $\sqrt{s}$ = 8 TeV*, Eur. Phys. J. C **76** (2016)

[86] ATLAS Collaboration, *Search for a massive resonance decaying into a Higgs boson and a W or Z boson in hadronic final states in proton-proton collisions at $\sqrt{s}$ = 8 TeV*, JHEP **2016** (2016)

[87] A. J. Larkoski, D. Neill, J. Thaler, *Jet Shapes with the Broadening Axis*, JHEP **04** (2014)

[88] A. J. Larkoski, G. P. Salam, J. Thaler, *Energy correlation functions for jet substructure*, JHEP **2013** (2013)

[89] I. Moult, L. Necib, J. Thaler, *New Angles on Energy Correlation Functions*, JHEP **12** (2016)

[90] M. LeBlanc, *3- vs. 2-prong jet discrimination*, ATLAS Hadronic Calibration Workshop (2018)

[91] M. Cacciari, G. P. Salam, *Pileup subtraction using jet areas*, Phys. Lett. **B659** (2008)

[92] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*, ATL-PHYS-PUB-2017-013 (2017)

[93] ATLAS Collaboration, *Secondary vertex finding for jet flavour identification with the ATLAS detector*, ATL-PHYS-PUB-2017-011 (2017)

[94] G. Piacquadio, C. Weiser, *A new inclusive secondary vertex algorithm for b-jet tagging in ATLAS*, J. Phys. Conf. Ser. **119** (2008)

[95] R. E. Kalman, *A new approach to linear filtering and prediction problems*, J. Basic Eng. **82** (1960)

[96] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV*, CERN-EP-2019-132 (2019)

[97] ATLAS Collaboration, *Variable Radius, Exclusive-k_T, and Center-of-Mass Subjet Reconstruction for Higgs($\rightarrow b\bar{b}$) Tagging in ATLAS*, ATL-PHYS-PUB-2017-010 (2017)

[98] ATLAS Collaboration, *Flavor Tagging with Track Jets in Boosted Topologies with the ATLAS Detector*, ATL-PHYS-PUB-2014-013 (2014)

[99] C. Chen, *New approach to identifying boosted hadronically-decaying particle using jet substructure in its center-of-mass frame*, Phys. Rev. **D85** (2012)

[100] C. Chen, *Identification of a bottom quark-antiquark pair in a single jet with high transverse momentum and its application*, Phys. Rev. **D92** (2015)

Erklärung nach §17(9) der Prüfungsordnung für den Bachelor-Studiengang Physik und den Master-Studiengang Physik 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 1. November 2019

(Joshua Falco Beirer)