Department of Physics and Astronomy
University of Heidelberg

Bachelor Thesis in Physics
submitted by

Sascha Simon Dreyer

born in Göttingen (Germany)

2016

Experimental Study of Jet Substructure Techniques with Tracks for Decays of Highly Energetic Massive Gauge Bosons, Higgs Bosons and Top Quarks

This Bachelor Thesis has been carried out by Sascha Simon Dreyer at the
Kirchhoff-Institute for Physics in Heidelberg
under the supervision of
Priv.-Doz. Dr. Oleg Brandt

Abstract

Predictions of new particle resonances around the TeV scale can be probed with the ATLAS detector via their decay to boosted, hadronically decaying massive gauge bosons, Higgs bosons and top quarks. Nevertheless, these channels are limited due to the tremendous amount of QCD background at the LHC.
Observables deduced from the Energy Correlation Functions and n-Subjettiness are used by the ATLAS collaboration in the highly energetic events to explore jet substructure in order to constrain these background by using calorimeter clusters.
The study presented in this Bachelor thesis evaluates the performance of these observables using tracks and the novel ansatz of subjet assisted tracks (TAS) in terms of rejecting QCD events for this scenarios. TAS is a scaling of tracks with matched calorimeter subjets to include information of the neutral part. A comparison with established methods is performed in six different regions of p_{T} from 250 to $\sim$2000 GeV.
For W/Z boson and top quark tagging, significant improvements over the commonly used calorimeter observables are found. Smaller advantages of tracks and TAS are observed for Higgs boson tagging. TAS is found to perform on an equal level as tracks at high energies and slightly better at lower energies.

Zusammenfassung

Vorhersagen neuer Teilchenresonanzen auf der TeV Skala können mit dem ATLAS Detektor durch ihre Zerfälle in hochenergetische, hadronisch zerfallende massive Eichbosonen, Higgs Bosonen und Top Quarks getestet werden. Allerdings sind diese Kanäle durch den hohen QCD Hintergrund am LHC limitiert.
Observablen, die aus den Energy Correlation Functions und n-Subjettiness abgeleitet sind, werden von der der ATLAS Kollaboration in hochenergetischen Ereignissen genutzt, um durch Untersuchung der Kalorimeter Jet Substruktur diesen Hintergrund zu verringern.
Die Studie, die in dieser Bachelorarbeit präsentiert wird, evaluiert die Leistung dieser Observablen mit Spuren und Subjet assistierten Spuren (TAS) bei der Unterdrückung von QCD Ereignissen in diesen Szenarien. TAS bedeutet das Skalieren der Spuren mithilfe von Kalorimeter Subjets, um Informationen des neutralen Anteils miteinzubeziehen. Ein Vergleich mit bereits etablierten Methoden wird in sechs verschieden p_{T} Regionen zwischen 250 und $\sim$2000 GeV vorgenommen.
Bei der Identifikation von W/Z Bosonen und Top Quarks werden signifikante Verbesserungen gegenüber den üblicherweise genutzten Kalorimeter Variablen festgestellt. Geringere Vorteile von Spuren und TAS werden bei der Higgs Boson Identifizierung beobachtet. Es hat sich gezeigt, dass TAS bei hohen Energien die gleiche Leistung wie Spuren und eine geringfügig bessere bei niedrigeren Energien erbringt.

Contents

1. Introduction and Motivation

The Standard Model of Particle Physics is a theoretical framework characterizing the most fundamental constituents of our universe and their interactions. Experimental data is well described, among others the great success of observing the Higgs boson with the ATLAS and CMS experiments at CERN in 2012. Although, the Standard Model features some problems as for example it does not provide a viable candidate for dark matter, does not include gravitation or offers a solution for the hierarchy problem.

Many models describing physics beyond the Standard Model predict new particle resonances around the TeV scale, which the LHC is designed to study. With experiments as ATLAS, new heavy particle resonances can be detected via their decay to Standard Model particles, e.g. the massive gauge bosons W/Z, the Higgs boson or the heaviest known elementary particle, the top quark. Due to the potentially high mass of the resonance, the decay products are not produced at rest, instead feature a high transverse momentum, called boosted. In the majority of cases, they decay further hadronically to light quarks, which hadronise and give rise to jets in the detector. The jets angular separation in turn, is inversely proportional to the boost of the decaying particle. Larger cone radii are used to combine the full e.g. W boson decay in one large radius jet. Since jets are very common at hadron colliders as the LHC, sophisticated methods are needed to distinguish jets originated by boosted W/Z and Higgs bosons as well as top quarks, from QCD jets emerging from a single quark or gluon.

Jet substructure techniques are used for differentiation, with the jet mass being the main discriminant since QCD jets are described by an exponentially falling mass spectrum. In addition, the substructure variables C2, D2 τ_{21}, and τ_{32}, deduced from the Energy Correlation Functions and n-Subjettiness, are used to explore the angular distribution of the individual signals inside the jet in order to further constrain the QCD background. This is done by evaluating how probable a given jet has originated from either three quarks for the top quark signal, two quarks for the W/Z and Higgs boson signals or one quark or gluon for the QCD background.

Currently, these substructure variables are calculated with calorimeter information only, which means by combinations of calorimeter clusters, representing the jet. Another possibility is to calculate the jet variables with tracks, whose angular coordinates are measured very precisely. This excellent angular resolution could improve substructure variables in terms of QCD rejection in the highly boosted regime due to further shrinking separation of the decay products.

However, tracks are exclusively produced by charged particles, thereby missing around one third of the jet since quarks predominantly hadronise to two third charged and one third uncharged pions. A relatively new ansatz within the ATLAS collaboration is to use assisted tracks for the jet mass calculation, which use the excellent angular resolution of

the tracking detector and are scaled via a matched calorimeter jet, thus regaining information about the previously missed neutral part. First results of this jet mass variable are promising, therefor, it is worthwhile to study the performance of the Energy Correlation Functions and n-Subjettiness, calculated with tracks and assisted tracks.

Subject of this thesis is to quantify the possible QCD rejection when tagging W/Z bosons, Higgs bosons and top quarks with substructure variables calculated with tracks and assisted tracks compared to the currently used calorimeter variables.

This Bachelor thesis is structured as follows: A brief introduction into the Standard Model is given in Chapter 2 with an emphasis on the W/Z bosons, the Higgs boson and the top quark since the discrimination of those from the QCD background is studied. The ATLAS detector at CERN is shortly presented afterwards in Chapter 3, focusing on calorimetry and the tracking detector. Subsequent, Chapter 4 describes the methodology of the performed study by introducing key concepts such as jets and track assisting, as well as specifying the technical implementation. Chapter 6 is dedicated to the evaluation of the results of this study, presented in terms of the achieved background rejections with track and TAS based substructure variables compared to calorimeter variables. A summary of the results is given in Chapter 7. Furthermore, an outlook to further studies can be found in the last part of this thesis, Chapter 8.

2. The Standard Model of Particle Physics

The following chapter gives a brief introduction into the Standard Model of Particle Physics. After a short description of the bosonic, Chapter 2.2, and fermionic, Chapter 2.3, sectors, the W and Z bosons as well as the Higgs boson and top quark are reviewed in more detail (with a focus on W/Z) in Chapter 2.4, since tagging these particles at high p_{T} is of major interest for this thesis. Subsequently, background processes for the considered hadronic decay channels are studied in Chapter 2.7. At the end of this part, two possible extension to the Standard Model, the Heavy Vector Triplet (or HVT) and the Randall-Sundrum (RS) model are presented concisely in Chapter 2.8 given that the substructure techniques studied in this thesis are well designated probing the predictions of the HVT, RS and other extensions.

2.1. Introduction to The Standard Model

In our current understanding of physics, particles interact via 4 different forces. Namely the strong force, the weak interaction, electromagnetism and gravity.

The Standard Model of Particle Physics (SM), see e.g. References [1, 2, 3], emerged during the $60's$ as a relativistic quantum field theory with major contributions from Weinberg, Glashow and Salam (see References [4, 5, 6]). The weak interaction and electromagnetism are combined into the electroweak interaction which, together with the strong interaction, provide the basis of the SM. There is no theory that fully includes gravity as a quantum field theory, though there are attempts at large scales such as String Theory.

The predictions of the SM are well describing experimental data in almost every scenario, see Reference [7]. However, there are certain observations, that are not explained by the SM. These involve the large matter - antimatter asymmetry in the universe (the SM does predict matter - antimatter asymmetry in the mixing of quark flavours but the magnitude is too small), neutrino oscillation and dark matter.

2.2. Bosonic Sector

The SM is a quantum field theory, relying on local gauge invariance under symmetry operations. Represented as groups, these are $\mathbb{SU}(3)_c \times \mathbb{SU}(2)_l \times \mathbb{U}(1)_Y$, where c stands for the colour-charge of the strong interaction, l for the left-handed nature of the weak interaction and Y for the hypercharge which is associated with the $\mathbb{U}(1)$ symmetry.

		Mass	Q	I_W^3	Discovery
Gauge	W^+	$80.385 \pm 0.015\,\text{GeV}$	$+1$	$+1$	1983
	W^-	$80.385 \pm 0.015\,\text{GeV}$	-1	-1	1983
	Z	$91.1876 \pm 0.0021\,\text{GeV}$	0	0	1983
	γ	0	-1	0	ca. 1900
	Glouns g	0	0	0	1979
Higgs	H	$125.09 \pm 0.24\,\text{GeV}$	0	0	2012

Table 2.1.: Table of elementary bosons in the SM, Q describes the electric charge and I_W^3 the weak isospin, values taken from the PDG, see Reference [7].

The generators of the symmetry groups enter observations as gauge bosons, or vector bosons as they are spin 1 particles. For $\mathbb{SU}(3)_c$, there are eight massless, differently colour charged gluons. The electroweak sector gives rise to four physical gauge bosons, the two electrically charged $W^\pm$ bosons and two neutral, the Z boson and the Photon, which is also called γ. While the $W^\pm$ bosons are linear combinations of entirely $\mathbb{SU}(2)_l$ generators, the neutral ones mix $\mathbb{SU}(2)_l$ and $\mathbb{U}(1)_Y$.

An important aspect of the electroweak sector is the mass of the $W^\pm$ and the Z bosons, while the Photon is massless. To achieve this, the symmetry is broken by adding an additional scalar field, the Higgs field, which is an example of Spontaneous Symmetry Breaking, see e.g References [8] and [9]. The commonly used term is electroweak symmetry breaking. Spontaneous symmetry breaking involves Goldstone bosons, in the case of the Higgs field there are four of them. The two electrically charged ones and the neutral Goldstone boson are interpreted as additional longitudinal degrees of freedom of the $W^\pm$ and Z bosons, which thus become massive. The fourth, neutral one, remains physically observable as the Higgs boson. The discovery of the corresponding spin zero Higgs particle with a mass of 125 GeV in 2012 by the ATLAS and CMS experiments at CERN was a great success for both, theory and experiments and was awarded with the Nobel Price for Physics in 2013, see Reference [10]. The bosonic sector of the SM is listed in Table 2.1.

2.3. Fermionic Sector

Alongside the gauge and the Higgs boson, there are 24 other elementary particles, which all feature spin $\frac{1}{2}$, thus are called fermions. Table 2.2 lists all elementary fermions described by the SM.

A distinction is drawn between Quarks and Leptons. The most important difference the colour charge. While quarks also carry colour charge, and therefore participate in all interactions including the strong interaction, leptons are not colour charged and therefore do not interact with gluons.

All particles have a left- and a right-handed chiral state, the left-handed part obtains a $\mathbb{SU}(2)$ quantum number, called weak isospin to which the $W^\pm$ as well as the $\mathbb{SU}(2)$ part of the Z and γ couple. Accordingly, only the left-handed chiral state of a particle can interact with $W^\pm$ bosons.

		Mass	Q	I_W^3	Interaction	Discovery
Leptons	ν_e	$< 2.2\,\mathrm{eV}$	0	$+\frac{1}{2}$	$\mathbb{SU}(2)_l$	1956
	e	$0.511 \pm 10^{-8}\,\mathrm{MeV}$	-1	$-\frac{1}{2}$	$\mathbb{SU}(2)_l$, $\mathbb{U}(1)$	1897
	ν_μ	$< 0.19\,\mathrm{MeV}$	0	$+\frac{1}{2}$	$\mathbb{SU}(2)_l$	1962
	μ	$105.7 \pm 4 \times 10^{-6}\,\mathrm{MeV}$	-1	$-\frac{1}{2}$	$\mathbb{SU}(2)_l$, $\mathbb{U}(1)_Y$	1936
	ν_τ	$< 18.2\,\mathrm{MeV}$	0	$+\frac{1}{2}$	$\mathbb{SU}(2)_l$	2000
	τ	$1776.86 \pm 0.12\,\mathrm{MeV}$	-1	$-\frac{1}{2}$	$\mathbb{SU}(2)_l$, $\mathbb{U}(1)_Y$	1975
Quarks	u	$2.3 \pm 0.7\,\mathrm{MeV}$	$+\frac{2}{3}$	$+\frac{1}{2}$	$\mathbb{SU}(3)_c$, $\mathbb{SU}(2)_l$, $\mathbb{U}(1)_Y$	1968
	d	$4.8 \pm 0.5\,\mathrm{MeV}$	$-\frac{1}{3}$	$-\frac{1}{2}$	$\mathbb{SU}(3)_c$, $\mathbb{SU}(2)_l$, $\mathbb{U}(1)_Y$	1968
	c	$1.275 \pm 0.025\,\mathrm{GeV}$	$+\frac{2}{3}$	$+\frac{1}{2}$	$\mathbb{SU}(3)_c$, $\mathbb{SU}(2)_l$, $\mathbb{U}(1)_Y$	1974
	s	$95 \pm 5\,\mathrm{MeV}$	$-\frac{1}{3}$	$-\frac{1}{2}$	$\mathbb{SU}(3)_c$, $\mathbb{SU}(2)_l$, $\mathbb{U}(1)_Y$	1968
	t	$173.2 \pm 0.9\,\mathrm{GeV}$	$+\frac{2}{3}$	$+\frac{1}{2}$	$\mathbb{SU}(3)_c$, $\mathbb{SU}(2)_l$, $\mathbb{U}(1)_Y$	1995
	b	$4.18 \pm 0.03\,\mathrm{GeV}$	$-\frac{1}{3}$	$+\frac{1}{2}$	$\mathbb{SU}(3)_c$, $\mathbb{SU}(2)_l$, $\mathbb{U}(1)_Y$	1977

Table 2.2.: Table of elementary fermions in the SM, Q describes the electric charge and I_W^3 the weak isospin, values taken from the PDG, see Reference [7].

In addition to that, there is an antiparticle to every particle with the same mass but inverse charge-like quantum numbers, referring e.g. to electric or colour charge and the weak-isospin.
Both, quarks and leptons, are grouped three families or generations, ordered by mass. The first generation in the quark sector is given by the two lightest quarks, the up (u) and the down (d) quark. They have counterparts with higher mass in the other generations, wherefore these are as well called up-type and down-type quarks. All up-type quarks feature electric charge $+2/3$ while the down-type quarks carry an electric charge of $-1/3$. Similarly, all up-type quarks and respectively down-types are assigned the same weak isospin, which means, they couple identical to $\mathbb{SU}(2)_l$. The up quark analogon in the second family is called charm (c) quark whereas the down-type is called strange (s) quark. In the last and heaviest family, they are called, in the same order, top (t) and bottom (b) quark. The leptonic side is structured similarly with three mass generations. Each generation has a member of electric charge -1, and a neutral one. The charged ones are called electron (e), muon (μ) and tau (τ) and again, their left-handed chiral states are assigned the same weak isospin. The neutral leptons are called neutrinos (ν) and, in the unextended SM, are massless (though neutrino oscillation indicates a small but finite mass). As neutrinos do only participate in the weak interaction, there is only a left-handed chiral state in the SM.

2.4. Physics of W and Z

The W and Z bosons were discovered 1983 at the CERN Super Proton Synchrotron, Sp$\bar{\text{p}}$S, by the UA1 and UA2 experiments. The Sp$\bar{\text{p}}$S operated as a hadron collider from 1981 to 1984, the accelerator is now used as final injector for the LHC. Carlo Rubbia and Simon van der Meer were awarded the Nobel Prize in Physics 1984 for the discovery of

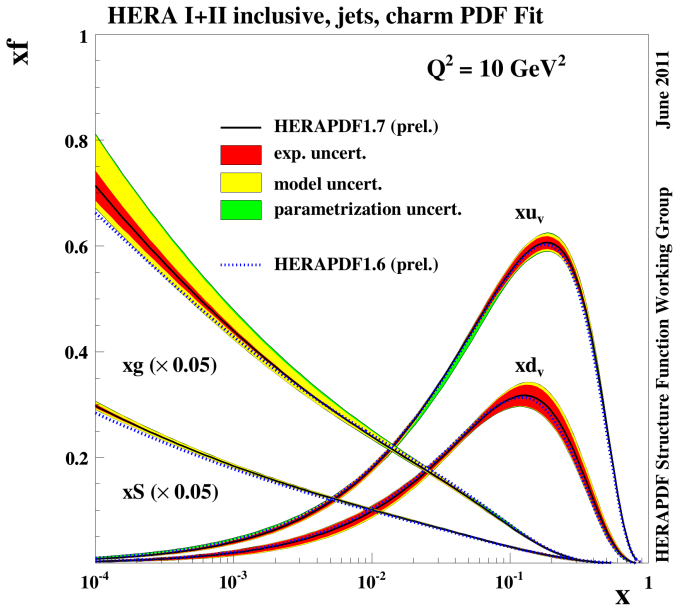

Figure 2.1.: Parton density function from HERAPDF1.7, taken from Reference [12].

the W and Z, see Reference [11].

Precision measurements of the W and Z boson masses as well as other properties, such as the Z line-shape or forward-backward asymmetry in the Z cross-section, were performed at LEPI and LEPII in electron - positron annihilation processes.
The world averages for the mass, given by the PDG (Particle Data Group), see Reference [7], are currently

$$M_W = 80.385 \pm 0.015\, GeV \tag{2.1}$$

$$M_Z = 91.1876 \pm 0.0021\, GeV \tag{2.2}$$

Massive gauge bosons as mediator of a force mean a finite range. That said, the weak interaction has a finite range of about 10^{-15} m in contrast to the infinitely ranged electromagnetic interaction or gravity whose mediators, the Photon respectively the predicted Graviton are massless. Gluons also are massless, but due to colour confinement, see Section 2.6, the range of the strong interaction is also finite.

2.4.1. Production and Decay

At tree level, that means in first order of perturbation theory, the production mechanism of W and Z bosons at hadron colliders such as the LHC, essentially is quark - antiquark annihilation, see Figure 2.4.

The proton primarily consists of three valence quarks, two up and one down quark, which determine the quantum numbers of the proton. Probed at large energy scales or small distances, the proton is more complex and features many gluons and sea quarks (quark - antiquark pairs) with a particularly low momentum fraction of the whole proton, called Bjorken x or -scale. Figure 2.1 gives an example of a proton Parton Density Function (PDF), which gives the probability to find a certain parton (constituent of a hadron) with a momentum fraction between $[x, x + dx]$ via $f(x)dx$. Because of the huge center of mass

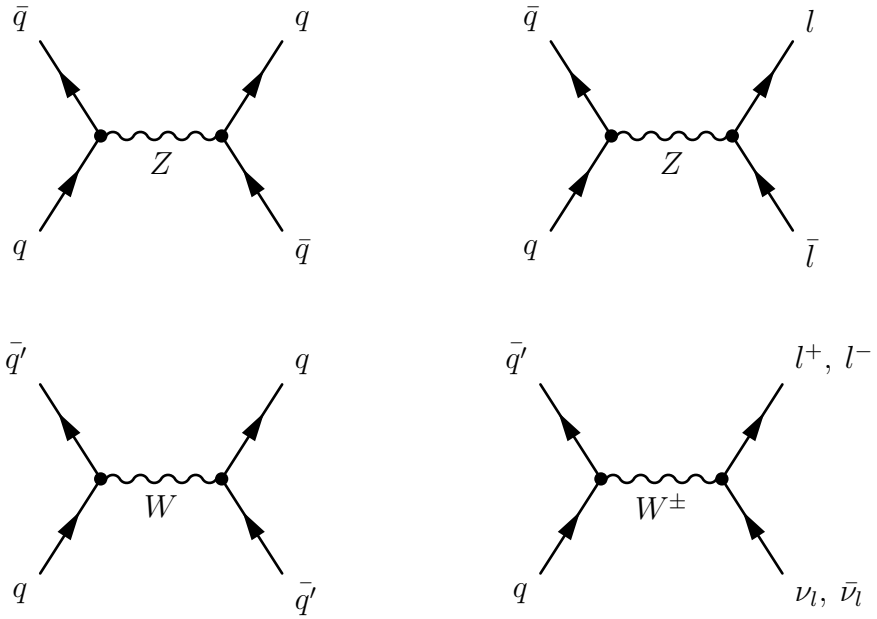

Figure 2.2.: Feynman diagrams for various W and Z boson decay channels.

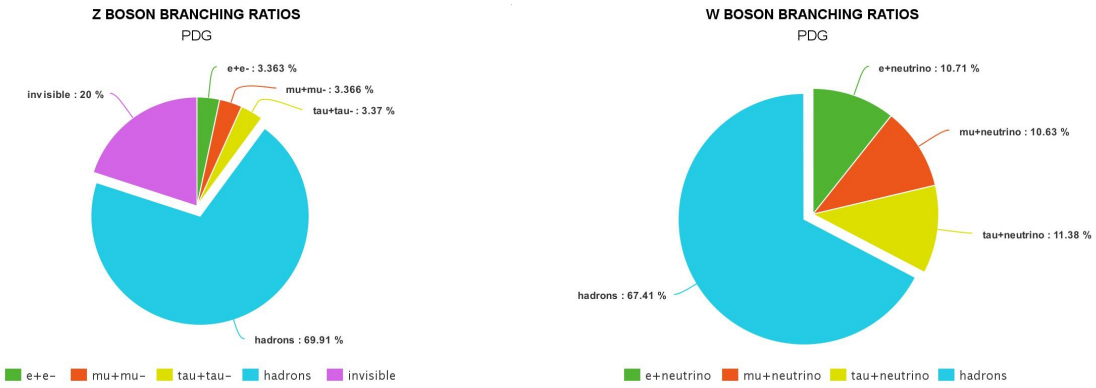

Figure 2.3.: Z boson (left) and W boson (right) decay channels with branching ratios, values taken from Reference [7].

energy of $\sqrt{s} = 13\,\text{TeV}$ at the LHC, sea quarks and gluons have enough energy to produce electroweak gauge bosons and heavier particles, despite their very small fraction of the overall momentum.

For the $W^\pm$ and Z bosons, possible are different decay channels. The branching ration (BR) is given by the ratio of the decay width Γ for a specific decay channel and the total decay width, thus describes how often the considered particle decays into a specific final state. For the $W^\pm$ and Z bosons, the decay channel with the highest branching ratio is to hadrons. This illustrates the importance of identifying hadronically decaying massive gauge bosons. In addition to that, the Z can also decay to pairs of leptons, while the $W^\pm$ decay to a charged lepton and the corresponding neutrino, given conservation of electric charge and lepton number. The main decay channels are shown in Feynman diagrams, see Figure 2.2, and the branching ratios are given in the pie chart, see Figure 2.3.

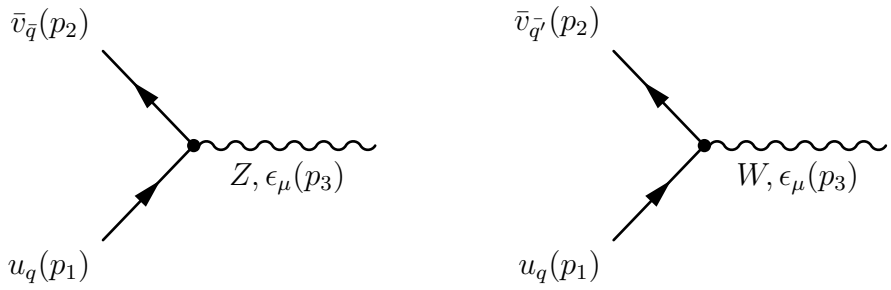

Figure 2.4.: Feynman diagrams for W/Z boson production.

2.4.2. Matrix Elements of the Weak Interaction

Needed for predictions of cross-sections is the transition probability from an initial state i to a final state f. Inside the framework of the SM, this is perturbatively calculable via the so called Matrix Element M_{fi}. The renowned Feynman rules are often used to determine M_{fi}. For a detailed treatment of this concept, the reader is referred to the standard literature, see e.g. Reference [1]. Briefly discussed in this part are the vertex factors of the weak gauge bosons on the basis of rather simple examples.

Given a real W production from quark - antiquark annihilation, see 2.4, needed to calculate M_{fi} are the fermion spinors $u_q(p_1)$, $\bar{v}_{\bar{q}'}(p_2)$, a polarisation 4-vector representing the W boson $\epsilon_\mu(p_3$ and the vertex factor:

$$V_W = -i\frac{e}{\sqrt{2}\sin(\Theta_W)}\gamma_\mu\frac{1-\gamma_5}{2}V_{CKM}^{ij} \tag{2.3}$$

For the same example with a Z, the vertex factor is defined as follows:

$$V_Z = -i\frac{e}{\sin(2\Theta_W)}\gamma_\mu(g_V^f - g_A^f\gamma_5) \tag{2.4}$$

Now, one can write down the Matrix Element (note that the i and possible signs are omitted, since only the absolute value squared is of physical interest):

$$M_{fi}^{q\bar{q}'\to W} = \frac{e}{\sqrt{2}\sin(\Theta_W)}\bar{v}_{\bar{q}'}(p_2)\gamma_\mu\frac{1-\gamma_5}{2}u_q(p_1)\epsilon^\mu(p_3)V_{CKM}^{ij} \tag{2.5}$$

And accordingly for the Z:

$$M_{fi}^{q\bar{q}\to Z} = \frac{e}{\sin(2\Theta_W)}\bar{v}_{\bar{q}}(p_2)\gamma_\mu(1-\gamma_5)u_q(p_1)\epsilon^\mu(p_3) \tag{2.6}$$

The fact that the weak interaction only couples to the left handed chiral state of a particle is given by the factor $\frac{1-\gamma_5}{2}$ which is a quantum mechanical operator and projects the left handed part of a Dirac spinor, $\Psi_L = \frac{1-\gamma_5}{2}\Psi$. This form of interaction is called V-A interaction, V stands for vector (γ_μ) and A for axial vector ($\gamma_\mu\gamma_5$), and is exactly given for the W boson. The Z boson, in contrast, features different weights to the V and A components (g_V and g_A) which represents the mixing of $\mathbb{SU}(2)_l$ and $\mathbb{U}(1)_Y$ and therefore different couplings to the right- and left-handed part of the interacting particle.

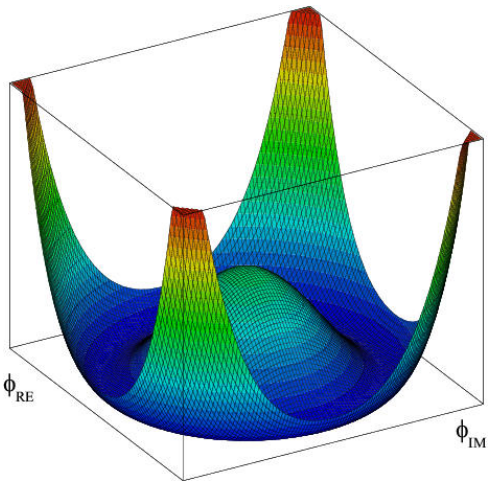

Figure 2.5.: Higgs potential, figure taken from Reference [13].

The exact coupling strength depends on the electric charge Q and weak isospin I_W^3 of the considered fermion.

$$g_V = \frac{I_W^3}{2} - Qsin^2(\Theta_W) \tag{2.7}$$

$$g_A = \frac{I_W^3}{2} \tag{2.8}$$

Another factor is the complex CKM matrix in the $Wq\bar{q}$ vertex factor. When interacting with a W boson, a quark can change its flavour, which is called flavour changing charged current. More specific, a given up-type quark of generation i can couple to a W along with a down-type quark of generation j, where i doesn't necessarily be the same as j. In particular, the coupling strength of quark with flavour i to one with flavour j is given by the matrix element V_{CKM}^{ij}. The case $i = j$, though, is the most probable as the diagonal elements of V_{CKM} are highest. This explains e.g. the predominant decay of the top quark to a bottom quark and a W boson.

2.5. Physics of the Higgs Boson and Top Quark

The Higgs boson was theoretically proposed by Higgs, Brout and Englert around 1964 to overcome the problems of generating mass for particles without violating gauge invariance. They proposed a new scalar field (Higgs field), which is invariant under the $\mathbb{SU}(2) \times \mathbb{U}(1)$ symmetry group. The Higgs field acquires a vacuum expectation value (vev) unequal to zero due to the special form of the potential, see Figure 2.5. The vacuum state with an expectation value breaks the symmetry and gives rise to 3 massless Goldstone bosons which mix with the $W^{\pm}$ and Z bosons to generate their mass, and a massive, physically observable scalar particle which, is known as the Higgs boson. Fermions acquire mass via Yukawa interactions with the Higgs field, whose strength is defined by the Higgs vev and the mass of the fermion. The Photon stays massless as the electromagnetic part of $\mathbb{SU}(2) \times \mathbb{U}(1)$ is unbroken. The vertex factors for an interaction of a Higgs boson with a fermion V_{hf}, and the massive vector bosons V_{hW} and V_{hZ} are given by:

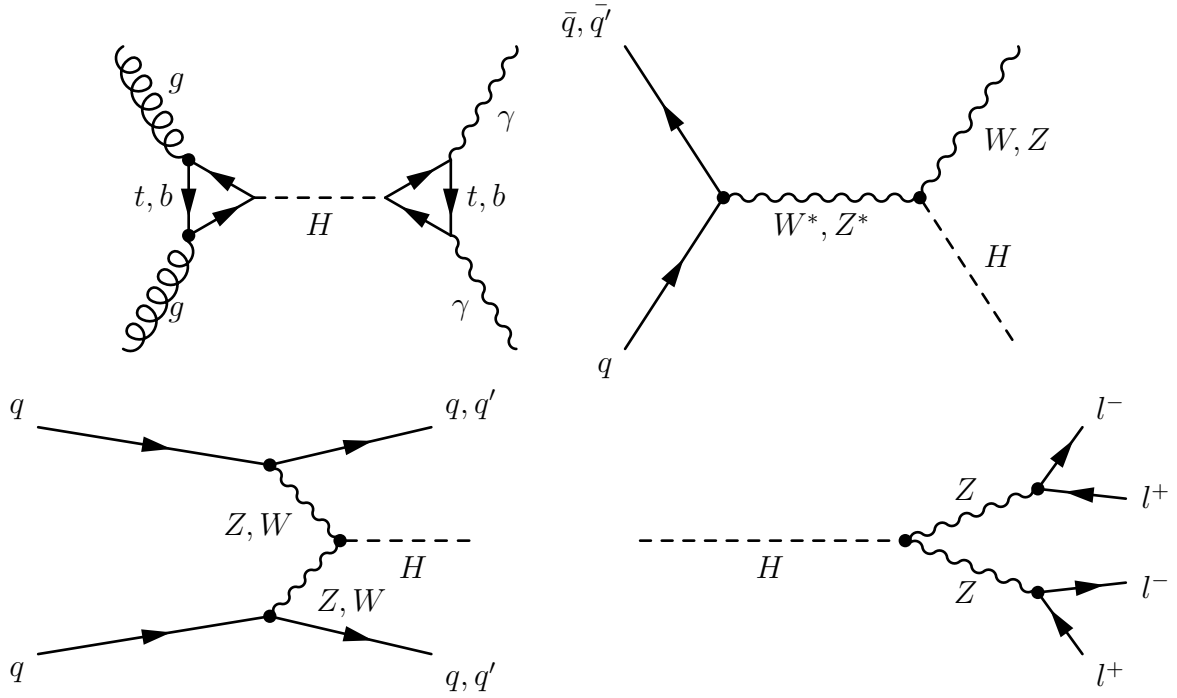

Figure 2.6.: Feynman diagram for Higgs production via gluon fusion and decay to two photons (upper left), production via Higgs-strahlung (upper right), production via VBF (lower left), decay to four leptons (lower right).

$$
\begin{aligned}
V_{hf} &= -i\frac{M_{fermion}}{v_0} \\
V_{hZ} &= 2i\frac{M_Z^2}{v_0}g_{\mu\nu} \\
V_{hW} &= 2i\frac{M_W^2}{v_0}g_{\mu\nu}
\end{aligned}
\tag{2.9}
$$

The top quark is the most massive elementary particle observed so far. Theoretically predicted in 1973 by Kobayashi and Maskawa to explain CP violation in the neutral Kaon system, it was first directly observed in 1995 by the CDF and DØ experiments at the Tevatron proton-antiproton collider located at Fermilab. Although, it was possible to set an indirect limit on the top mass via precise measurements of the W/Z bosons as their mass and couplings are sensitive to loop diagrams involving heavy particles as the top quark. With a mean lifetime of about $5 \cdot 10^{-25}$ s, the top quarks decays too fast to form bound state hadrons.

2.5.1. Production and Decay

At the LHC, the Higgs boson is dominantly produced by gluon fusion, Higgs-strahlung, vector boson fusion (VBF), and top fusion, see Figure 2.6. The predicted branching ratios are given in Figure 2.7.

A Higgs boson with a mass around 125 GeV, most often decays to a pair of bottom quarks. Unfortunately, this channel was not feasible for discovery due to the tremendous QCD background at hadron colliders, see Chapter 2.7, which indicates the necessity of

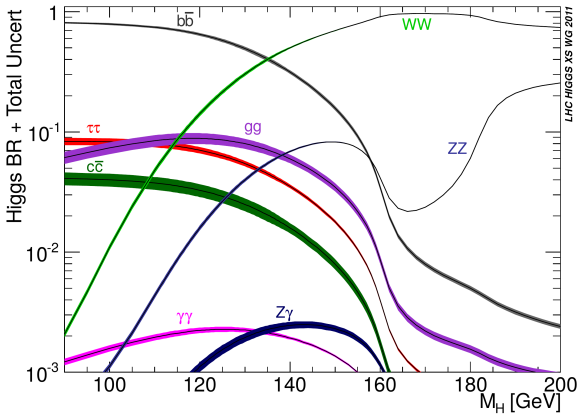

Figure 2.7.: Predicted Higgs branching ratios as a function of the Higgs mass m_H, taken from Reference [14].

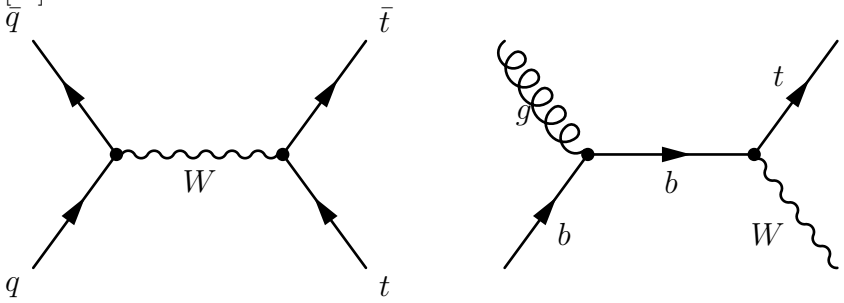

Figure 2.8.: Feynman diagram for top pair production via quark-antiquark annihilation (left) and tW-channel production (right).

effective methods to reduce QCD background. Instead, the ATLAS and CMS experiments especially focused on the rather clean $H \to \gamma\gamma$ and $H \to ZZ^* \to llll$ decay channels.

Top quarks are produced in pairs e.g. by gluon fusion or quark-antiquark annihilation, single tops can be produced in the t- and s-channel as well as together with a W boson, see Figure 2.8.

The top decays in the majority of cases (BR of around 0.998) into a b quark while radiating a W boson. This is reflected in the CKM matrix element $|V_{tb}|$ which is almost unity. Consequently, there are 3 different major decay channels for a pair of top quarks depending on the further decay of the W boson. Considering the branching fractions of the W, one can calculate the corresponding BR for top pairs.

- Full-hadronic: Both W bosons decay into quarks which results in 3 jets per top, BR ~ 0.44
- Semi-leptonic: One W boson decays hadronically, the other one decays to a charged lepton plus neutrino, BR ~ 0.44
- Full-leptonic: Characterized by both W bosons decaying leptonically. BR ~ 0.11

Interesting for this thesis are the hadronic decays of boosted tops since the discrimination of these jets from QCD jets is studied.

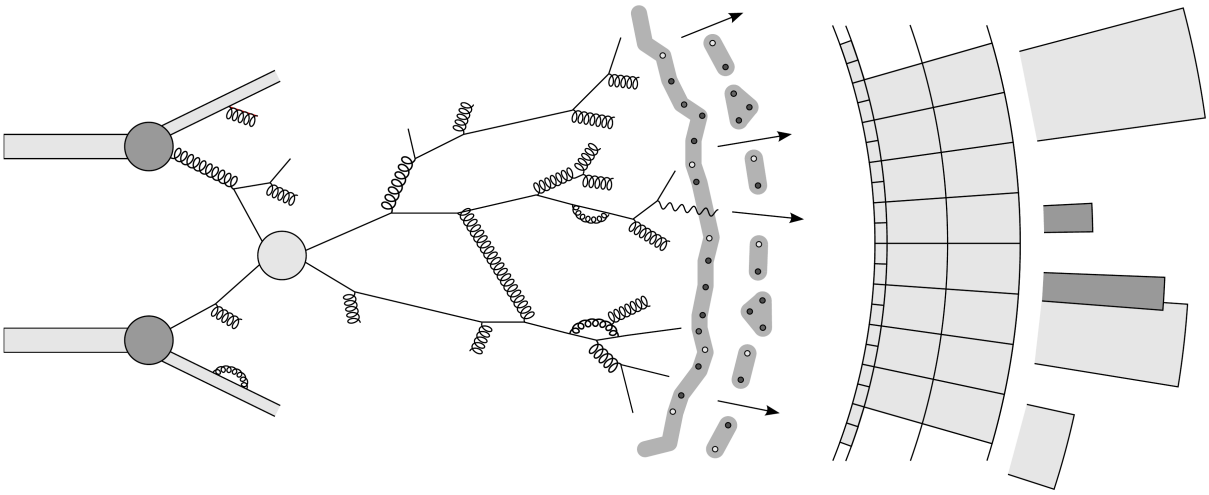

Figure 2.9.: Evolution of a hadronic shower, taken from Reference [15].

2.6. Introduction to Jets

The hadronic final state of an interaction in particle colliders consists mainly of strongly interacting quarks and/or gluons. Nevertheless, the nature of the strong force precludes the measurement of single, free partons inasmuch that it allows only for colourless or colour neutral states ($\mathbb{SU}(3)_c$ singlet), which is called confinement. At high energies, as achieved in collider experiments, the strong force becomes weaker and one can describe partons as asymptotically free, which is used e.g. when calculating cross-sections.

The evolution of a hadronic final state is as explained in the following. The final state partons, quarks and gluons, are in a state of asymptotic freedom. They emit gluons with a high cross-section which split further and further. This process is called fragmentation and showering. It continues until the particles carry so little energy that, with the now large strong coupling constant α_s, confinement becomes strong. At this point, the hadronisation will set in, which means that the particles form hadronic colourless bound states that may decay further to stable particles. Due to the large α_s, the hadronisation process is not calculable perturbatively.

This is illustrated in Figure 2.9, where the gray circle represents the hard interaction. Visible are the multiple radiations and splittings before bound states are formed on the right side. Jets are a definition as they depend on the actual choice of algorithm. Sketched further right is the detector with corresponding signals.
Pions are the predominant fraction of the hadronic shower, all three of them, π^+, π^- and π^0, are produced with similar probability and result in a mean charge to neutral fraction of 2/3 in the shower.

There are two mainly used models for the process of hadronisation, the cluster hadronisation, see Reference [16, 17], and the Lund string model, see e.g. Reference [18].
Each parton in the hadronic final state, called parton level, results, after the described evolution, in a collimated spray of hadrons, called particle level. These hadrons deposit energy in the detector, resulting in characteristic signals.

This stage is called detector level. Algorithms are used to cluster the measured signals at detector level to the jet observable. These clustering algorithms are describes in Chapter 4.1. Jets are a definition as they depend on the actual choice of algorithm. The idea is, that the jet describes the collimated spray of hadrons as exact as possible.

2.7. Background

The angular separation ΔR of decay products can be estimated via equation 2.7 with M and p_{T} describing the mass and transverse momentum of the decaying particle as well as p_i representing the four-momenta and $p_{T,i}$ the transverse momenta of the decay products.

$$\begin{aligned} M^2 = (p_i + p_j)^2 \;&\sim\; 2p_i \cdot p_j \;\sim\; p_{T,i} p_{T,j} \Delta R_{ij}^2 \\ p_{T,i} = p_{T,j} &= \frac{p_{\mathrm{T}}}{2} \\ \rightarrow\; \Delta R \;&\sim\; \frac{2M}{p_{\mathrm{T}}} \end{aligned} \tag{2.10}$$

The high boost of the decaying particle is responsible for a very small angular separation of the hadronic decay products. These are then combined into one large radius jet, see 4.1.

Without further investigation of the substructure, these jets will be indistinguishable from QCD jets with a mass comparable to the decaying particle, e.g. W/Z bosons, Higgs bosons or top quarks. A major contribution to QCD jets are dijet events. Several dijet production mechanisms are shown in Figure 2.10. Dijet events are very common at hadron colliders such as the LHC, which is why elaborated techniques to distinguish between QCD jets and jets containing decay products of heavier particles are of high importance.

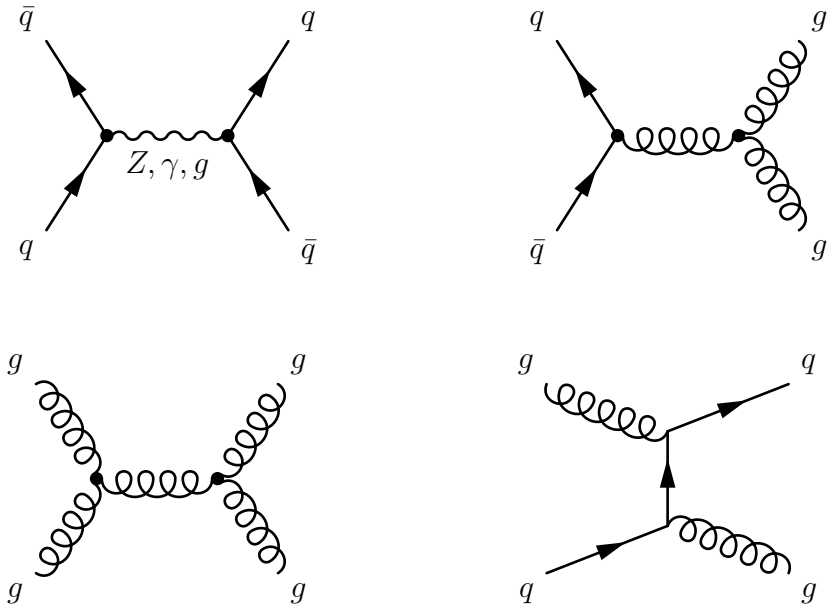

Figure 2.10.: Feynman diagrams for dijet production.

2.8. Possible SM Extensions

In this part, two possible extensions to the SM are presented briefly. The Heavy Vector Triplet models, Chapter 2.8.1, and the Randall-Sundrum model, Chapter 2.8.2. Predictions of those involve new heavy resonances which decay to known SM particles and are used as a source of boosted particles for this study. Moreover, searches for resonances like this would profit a lot from an improved background rejection in the hadronic decay channels of W/Z bosons, Higgs bosons and tops quarks.

2.8.1. Heavy Vector Triplet

Heavy Vector Triplet (HVT) models predict three new gauge bosons with similar properties as the massive gauge bosons present in the SM. These W' and Z' are much heavier than the known gauge bosons and can decay via $W' \to WZ$ (diboson resonance) or $Z' \to t\bar{t}$, see Figure 2.11. New physics approaches that predict heavy W' & Z' bosons can be found in Reference [19, 20]. The ATLAS and CMS collaborations search for diboson resonances in different channels, see e.g. Reference [21].

Often used is a simplified Lagrangian, see e.g. Reference [22], providing a phenomenological parametrization. Via measuring generalized couplings, a large amount of explicit theories can be compared with experimental results without the need for a large number of free parameters to which the experiments may not be sensitive to. The phenomenological Lagrangian provides a set of parameters, the experiment yields a likelihood function or confidence intervals for these parameters and model specific functions translate phenomenological to explicit parameters.

An example of those parameters would be the mass of the new heavy vector bosons, $m_{W'}$ and $m_{Z'}$, as well as couplings to SM particles. The coupling strength to the W and Z bosons is given by g_V, fermions couple via $g^2 c_F/g_V$ and the coupling to the Higgs is described by $g_V c_H$ where c_F and c_H are factors varying these couplings and g_V represents the usual $\mathbb{SU}(2)_{left}$ gauge coupling. A drawback of a phenomenological Lagrangian is its insensitivity to certain degrees of freedom which are integrated out, such as the weak isospin in Fermis theory of the weak interaction.

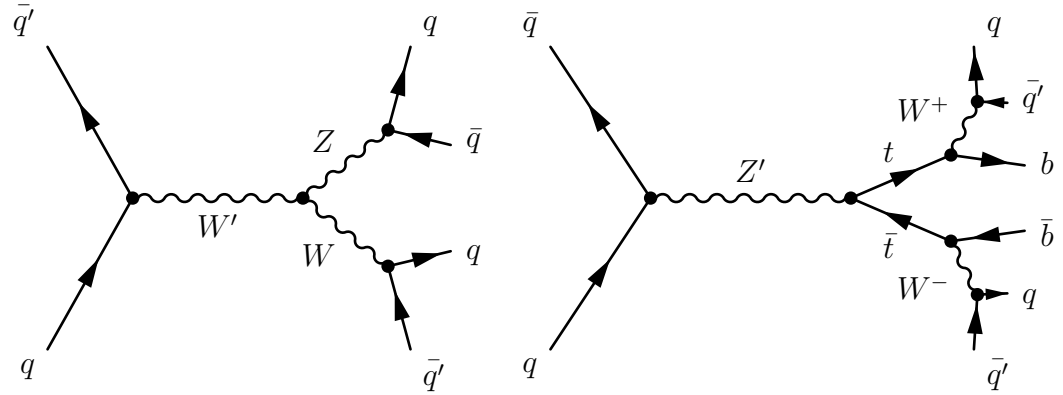

Figure 2.11.: Exemplary Feynman diagrams for production of W' and Z' bosons and their hadronic decay.

2.8.2. Randall-Sundrum Model

The Randall-Sundrum (RS) model was proposed 1999 by Randall and Sundrum, see Reference [23], to solve the hierachy problem in the SM. Large quadratic corrections to the Higgs field mass due to particles beyond the SM require a huge fine-tuning in order to fix the Higgs mass at the electroweak scale, where it was measured.

The RS approach considers a small 5th spacial dimension compactified on a circle with two 4-dimensional membranes at the endpoints, enclosing the 5-dimensional volume. One membrane is invisible and at the Plank scale. This setup solves the Einstein equations resulting in a 4-dimensional metric that is multiplied by an exponential factor dependent on the Planck scale k and the length scale of the additional dimension r_c, resulting e.g. in a exponentially increased vev of the Higgs field on the invisible membrane which allows for a Higgs mass at the right electroweak scale on the visible membrane, see equation 2.8.2.

$$m = e^{-kr_c\pi} m_0 \tag{2.11}$$

Furthermore, this model predicts visible excitation (gravitons) at the TeV scale, which can be tested at the LHC. One possibility is a RS-graviton G^* that decays to two Higgs bosons that are highly boosted due to the graviton mass m_{G^*}, see Figure 2.12.

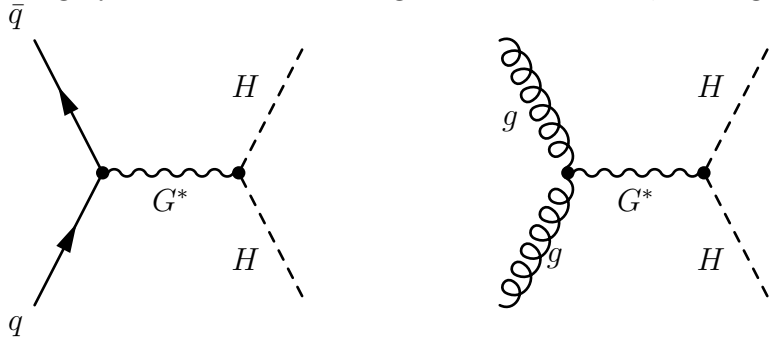

Figure 2.12.: Exemplary Feynman diagrams for RS-graviton production and decay to two Higgs bosons.

3. LHC and the ATLAS Detector

The performance study presented in this thesis uses simulations from the ATLAS detector at the Large Hadron Collider located at CERN. After a short description of the collider in Chapter 3.1, the ATLAS detector and in particular the tracking system and the calorimeter will be presented in Chapter 3.2.

3.1. The Large Hadron Collider

The Large Hadron Collider (LHC) is located at CERN, Geneva, between 45 m and 170 m under the surface, in the tunnel of the former electron - positron collider LEP. Currently, it is the largest particle accelerator, built to collide protons and even heavy ions like lead. Superconducting Nb-Ti magnets, cooled with supra-fluid helium to 1.9 K, are used to achieve a magnetic field strength of 8.3 T, needed to hold the protons, accelerated via RF cavities, on their circular path.

There are four large scale experiments located at the LHC, each at a different interaction point in the 26.7 km long tunnel, see Figure 3.1. Two of them are multi-purpose experiments, CMS (Compact Muon Solenoid) and ATLAS. These were built for a variety of searches and measurements, including the Higgs Boson and Super Symmetry (SUSY). The smaller two have a more specific physics program. ALICE (A Large Ion Collider Experiment) was built to study the collisions of heavy ions like lead, aiming to produce a Quark-Gluon-Plasma, which describes the state of the very early universe. The LHCb (Large Hadron Collider beauty) is primarily designed to measure precisely the production and rare decays of hadrons including b quarks to study CP-violation and matter-antimatter asymmetry.

Two important parameters characterize a collider, the instantaneous luminosity $\mathscr{L}$ and the center-of-mass energy $\sqrt{s}$. The instantaneous luminosity is given by the rotation frequency of the beams in the tunnel, the number of bunches, the number of particles per bunch and the geometry of the beam at the interaction point. The integrated luminosity $\int \mathrm{d}t\, \mathscr{L}$ or L_{int} gives a measure that is directly connected to number of events, N, via the cross-section σ.

$$\begin{aligned} \dot{N} &= \mathscr{L} \cdot \sigma \\ N &= \int \mathrm{d}t\, \mathscr{L} \cdot \sigma \end{aligned} \tag{3.1}$$

The center-of-mass energy describes the maximum energy that can be transferred in an interaction. The LHC is currently operating with a center-of-mass energy of $\sqrt{s} = 13\,\mathrm{TeV}$ and a peak instantaneous luminosity of around $10^{34}\ \mathrm{cm}^{-2}\,\mathrm{s}^{-1}$.

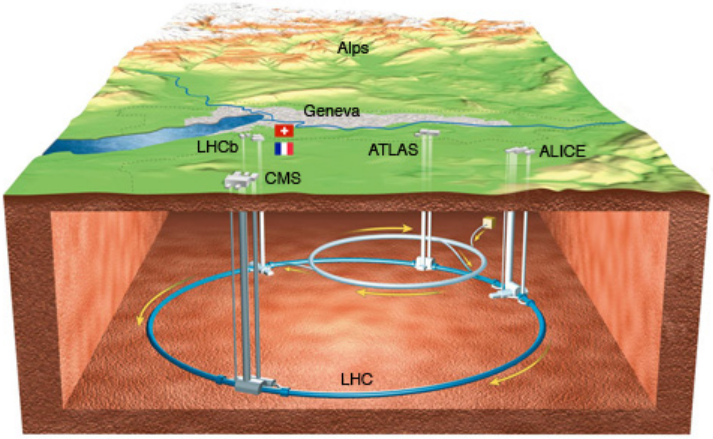

Figure 3.1.: The LHC complex, taken from Reference [24].

3.2. The ATLAS Detector

The ATLAS experiment is one of the four large experiments at the LHC, designed for a versatile physics program focussing on proton-proton collisions. It is a classically structured collider experiment with layers of different subdetectors around the interaction point (IP). The inner detector (ID) is formed by a tracking system, see Chapter 3.2.1. The tracker is followed by the calorimetry, Chapter 3.2.2, which consists of an electromagnetic (EMcal) and a hadronic (Hcal) part. The outer most layers are occupied by the muon system, see Chapter 3.2.3. In order to combine the information of the different subsystems and to be able to cope with the tremendous amount of interactions, an additional dedicated Trigger system is installed 3.2.4. Considering the study of this thesis, the main focus of the description will lie on the tracking system and the calorimeters. A comprehensive description of the ATLAS detector is given in Reference [25].

ATLAS uses a right-handed coordinate system with its origin at the interaction point. The positive z-axis is defined in beam direction while the transverse plain is described by the y-axis pointing upwards and the x-axis aiming to the center of the LHC ring. Given the symmetry of the ATLAS experiment, what is used are spherical coordinates with the azimuthal angle ϕ defined from the positive x-axis around the beam line and the polar angle θ, which gives the displacement from the z-axis.

Next to θ, there is the rapidity $y = \text{atanh}\,(\frac{p_z}{E})$ variable, with the advantage, that differences are invariant under Lorentz-boosts in beam direction. Commonly used is the pseudorapidity $\eta = \text{atanh}\,(\frac{p_z}{|p|}) = -\ln[\tan\frac{\theta}{2}]$, which is equivalent to the rapidity in the limit $v \to c$. The advantage of rapidity is the invariance of differences Lorentz-transformations in beam direction. Angular distances are measured in units of $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$. Since the partons incident momenta in z-direction are unknown and due to the momentum balance in the transverse plane to the beam direction, used is the transverse momentum, p_{T}, variable.

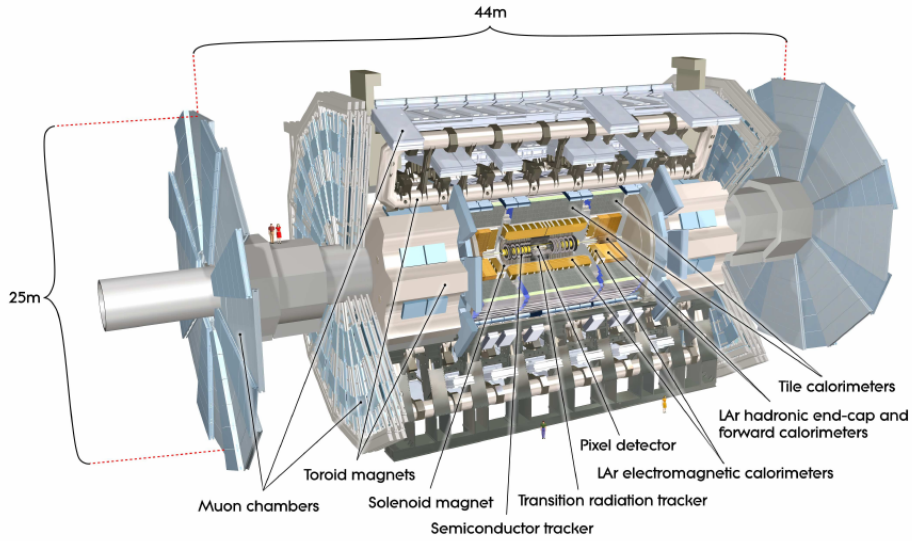

Figure 3.2.: Cut-away view of the ATLAS detector, taken from Reference [25]

3.2.1. Inner Detector

The inner part of the detector, see Figure 3.3 is used to track charged particles.
Due to the tremendous amount of scattering processes, a highly radiation hard setup is needed to maintain long-time stability. Simultanously, there should be as less as possible material in front of the calorimeter to minimize previous energy loss.

Charged particle trajectories are bent due to the solenoidal magnetic field of 2 T and the curvature provides information about their transverse momenta p_{T} and the sign of the charge. Energy depositions in the calorimeter can be linked to vertices and by resolving primary and secondary vertex locations it is possible to measure the life-time of certain particles and with this, tag jets to contain b-hadrons.

The tracking part is 6.2 m long, 2.1 m high and is divided into three different regions, see Figure 3.3. A high granularity silicon pixel detector is used in the inner layers, Chapter 3.2.1 with a new layer installed after Run 1, see Chapter 3.2.1. Further away of the IP are Silicon Strip Detectors installed, see Chapter 3.2.1. The outermost part is given by the Transition Radiation Tracker, Chapter 3.2.1, which works with gaseous straw tubes. The resolution of the measured transverse momentum is given by $\frac{\sigma_{p_{\mathrm{T}}}}{p_{\mathrm{T}}} = 0.05\% \cdot p_{\mathrm{T}}\,[\mathrm{GeV}] \oplus 1\%$

Silicon Pixeldetector

A highly accurate measurement of discrete space-points in the Pixel Detector is provided by 80.4 million single pixels, sized $50 \times 400\,\mu\mathrm{m}^2$ for 90% and $50 \times 600\,\mu\mathrm{m}^2$ for the remaining 10% of the pixels in the front-end electronic regions, grouped into 1744 sensor modules. The design of three barrel layers and two end-caps with three disc layers each, covers a region up to $|\eta| < 2.5$. The resolution of about 10 μm in R-ϕ and 115 μm in z (barrel layers) and 10 μm in R-ϕ and 115 μm in R (Disks) guarantees a good vertex recon-

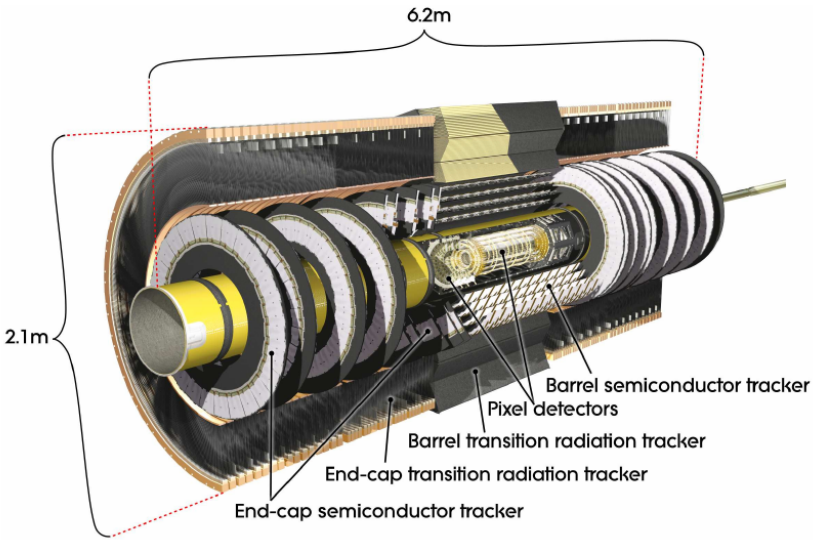

Figure 3.3.: The inner detector with its subsystems, taken from Reference [25]

struction. The 250 μm thick silicon sensor modules are using oxygenated n-type wafers with n-type readout which ensures radiation tolerance to charged hadrons.

Insertable B-Layer

The Insertable B-Layer (IBL), see Reference [26], was integrated into the ID during the shut down after Run 1 together with a smaller beam pipe. It is located between the new beam pipe and the first layer of the Pixel Detector, adding a fourth space point, mainly to improve b-tagging and to counter degrading effects due to the high radiation in the course of time. The layer consists of 26880 silicon pixels arranged in 80 columns and 336 rows with a size of $50 \times 250\,\mu\text{m}^2$, covering a range up to $|\eta| < 3.0$.

Silicon Strip Detector

The overall design of the Silicon Strip Detector (SCT) is similar to the pixel detector with four cylindrical layers and 9 disk layers per end-cap with a coverage of $|\eta| < 2.5$. There are 4088 modules, split into 2112 barrel and 1976 end-cap modules. Each is composed of two silicon strip sides glued back-to-back, producing stereo pairs of hits.

The resolution of the more cost-efficient SCT is not as good as in the pixel detector but with 17 μm in R-ϕ and 580 μm in z for the barrel, respectively R for disks it is is well suited for the main task of determining the transverse momentum. The technology used for the 285 μm thick sensors is a single-sided p-in-n one with AC-coupled readout strips.

Transition Radiation Tracker

In contrast to the former two subdetectors, the Transition Radiation Tracker (TRT) uses polyimide drift (straw) tubes to detect transition radiation produced by

charged particles passing through inhomogeneous media.

The tubes feature a diameter of 4 mm as well as a length of 144 cm in the barrel and 37 cm in the end-cap region and are filled with a mixture containing 70 % Xe, 27 % CO_2 and 3 % O_2 at a pressure of about $5 - 10$ mbar. They are arranged in parallel to the beam pipe and segmented into 32 read-out ϕ-sectors resulting in a reduced coverage of $|\eta| < 2.0$ and available information about only one spatial dimension. A gold plated tungsten wire serves as anode.

On average, 36 hits per track, used for pattern recognition, are measured by the TRT. Two different thresholds are used to distinguish between tracking signals and low energetic transition radiation photons as they produce a much larger amplitude. The resolution of the TRT is given by 130 μm in R-ϕ.

3.2.2. Calorimetry

The ATLAS calorimeters, see Chapter 3.4, measure the energy and direction of incident particles. Neutrinos and other particles that may have passed the system undetected are inferred by an energy imbalance transverse to the beam direction, called missing transverse energy or $E_{\mathrm{T,miss}}$.

Particles entering the calorimeter cause showers of secondary particles, which need to be stopped inside the detector to precisely measure the energy. The showers generated by electromagnetic objects like electrons and photons are measured in the first subsystem, the electromagnetic calorimeter, see Chapter 3.2.2. Hadronic showers usually are much larger and need to be stopped in the proceeding hadronic calorimeter, described in Chapter 3.2.2.

All calorimeters are of sampling type, alternating an absorbing material with an active material in which the secondary particles produce signals proportional to the energy of the incident particle by ionization or excitation.

Electromagnetic Calorimeter

The EMCal is thick enough (> 22 radiation lengths) to stop most electromagnetic showers and covers a region up to $\eta < 3.2$. Similar to the ID, it is split into a barrel part ($|\eta| < 1.475$) and two wheels per end-cap, which cover $1.375 < |\eta| < 2.5$ and $2.5 < |\eta| < 3.2$ consequently. The absorbing medium is lead while liquid-argon acts as active detecting material. Around $|\eta| = 1.4$, the transition of barrel and end-cap, there is a crack with lots of dead material (around 11 radiation lenghts) which makes a precise measurement difficult.

The kapton electrodes are accordion-shaped and feature a complete ϕ symmetry. The segmentation is given by 0.025×0.025 in $\eta \times \phi$ within the middle layer, where most of the energy is deposited. This region features the finest readout in contrast to the other parts with a coarser segmentation. In $|\eta|$, the spatial resolution is best for values smaller 2.5.
The energy resolution in the electromagnetic calorimeter is given by $\frac{\sigma_E}{E\,[\mathrm{GeV}]} = \frac{10\%}{\sqrt{E\,[\mathrm{GeV}]}} \oplus 0.7\%$

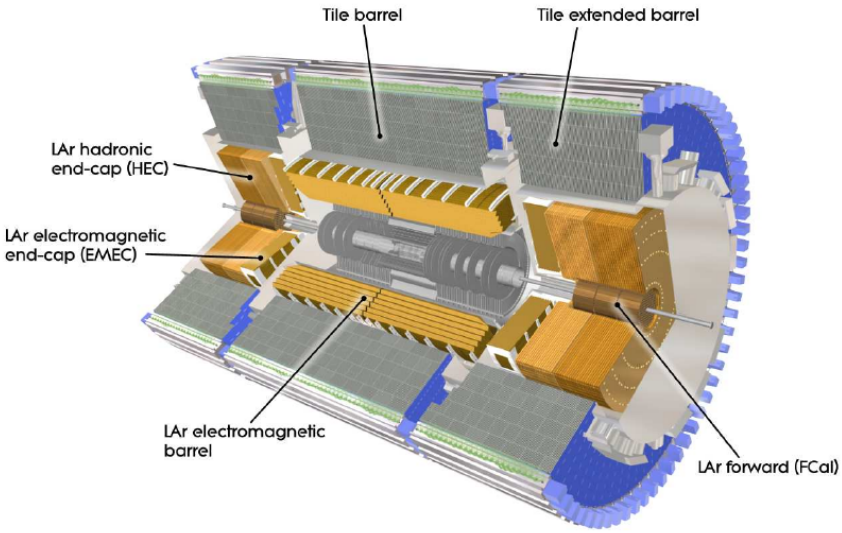

Figure 3.4.: The ATLAS calorimeter with its subsystems, taken from Reference [25]

Hadronic Calorimeter

The Hadronic Calorimeter (HCal) is split into a barrel part (coverage of $|\eta| < 1.7$), end-caps ($1.5 < |\eta| < 3.2$) and a Forward Calorimeter (FCal) which covers $3.1 < |\eta| < 4.9$. They are equipped with quite different absorbing and active materials. Liquid-argon is used for measuring in the end-caps and the FCal while the barrel part works with scintillating tiles. For absorption, steel was chosen for the barrel, copper for the end-caps and the part to the interaction point of the FCal and tungsten for the further FCal parts.

Given the thickness of 7.4 hadronic interaction lengths λ for the barrel, 9.7 for EMCal plus HCal and even more with dead material, the probability for a punch-through of a hadronic shower is very small. Here, the granularity of the read-out is given by 0.1×0.1 in $\eta \times \phi$ for $1.5 < |\eta| < 2.5$ and 0.2×0.2 for larger values of $|\eta|$.
The energy resolution of the hadronic calorimeter in the barrel region is given by $\frac{\sigma_E}{E\,[\mathrm{GeV}]} = \frac{50\%}{\sqrt{E\,[\mathrm{GeV}]}} \oplus 3\%$ and in the end-caps by $\frac{\sigma_E}{E\,[\mathrm{GeV}]} = \frac{100\%}{\sqrt{E\,[\mathrm{GeV}]}} \oplus 10\%$.

3.2.3. Muon Spectrometer

Due to their minimal ionizing nature, muons pass the calorimeters in the majority of cases. Therefore, the largest and outermost part of the ATLAS detector, the Muon Spectrometer (MS), is used for muon measurements.

The most important part are the toroidal magnetic fields generated by eight coils of large barrel toroids and two smaller end-cap toroids, see Figure 3.5. In conjunction with different tracking systems, this allows for a stand-alone p_{T} measurement up to 3 TeV. Other important functions of the MS are triggering (up to $|\eta| < 2.4$) and identification of bunch-crossings and background from cosmic rays.

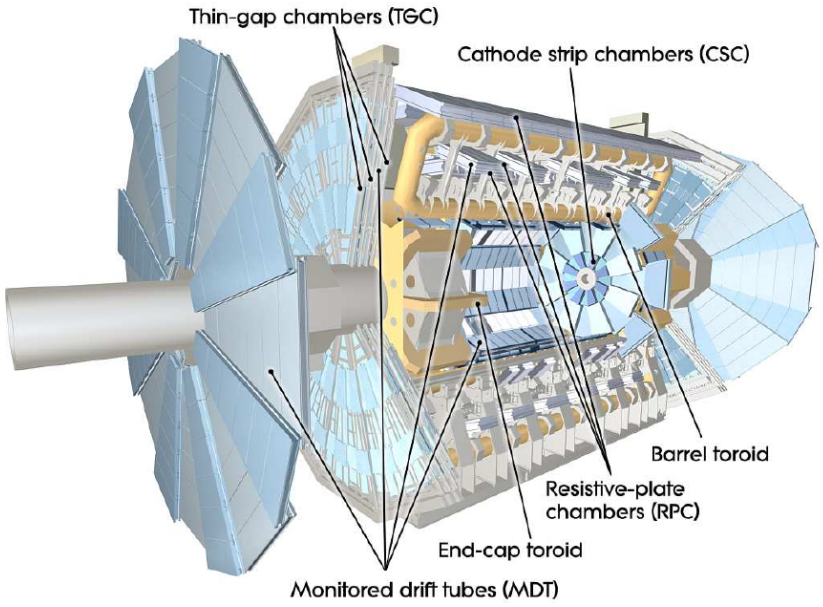

Figure 3.5.: Cut-away view of the muon system, taken from Reference [25]

The barrel section extends up to $|\eta| < 1.4$ and the end-caps cover a range of $1.6 < |\eta| < 2.7$. To fulfill the given tasks, four different types of tracking detectors with two main functions are used in the MS. Precise measurements of the p_{T} is accounted for by the Monitored Drift Tube Chambers (MDT) and Cathode Strip Chambers (CSC), whereas the Resistive-plate chambers (RPC) and Thin-Gap Chambers (TGC) are characterized by fast tracking, used e.g. for triggering, bunch-crossing identification or measuring the muon coordinate orthogonal to the direction measured by the MDT and CSC.

3.2.4. Trigger

The Trigger system aims to reduce the overwhelming amount of data produced by the ATLAS detector at an bunch-crossing rate of 40 MHz to a degree that is feasible for storage technologies. This is done via a system composed of a hardware and a software trigger by selecting events with a high momentum transfer or other characteristics that increase their relevance for certain studies.

The first part is the custom-hardware Level1Trigger (L1), which identifies of certain Regions-of-Interest (RoIs) and selects events with reduced granularity in $\eta \times \phi$ in view of signatures like high p_{T}, EM objects, muons, jets and missing E_T. The HighLevel-Trigger (HLT) is software based and forms the second part of the trigger system. It has access to the full detector information, including tracks, of events that pass L1. In the HLT, events are reconstructed step by step, running decisively algorithms as soon as the event building progressed sufficiently.

Events that pass L1 and the HLT are then stored permanently. L1 reduces the event rate from $\sim$ 1 GHz to $\sim$ 100 kHz with a decision time less than 2.5 μs, L2 has a decision time of 40 ms and reduces further to around 3.5 KHz and finally, the EF accounts for the

final step to approximately 1 kHz with a decision in several seconds.

4. Methodology

This chapter gives a description of the concepts that are used throughout this thesis. The first part, Chapter 4.1, describes the clustering of jets. Subsequently, the substructure variables studied in this thesis are defined, see Chapter 4.4. Track assisting substructure variables as well as the adaption for this thesis are described in Chapter 4.5.

4.1. Clustering Algorithms

Clustering algorithms can run on different inputs, e.g. four-momenta as measured by calorimeter clusters or tracks. Also possible are truth particles from Monte-Carlo, see as well Chapter 5.
Important properties of clustering algorithms are infrared and collinear safety or together IRC-SAFETY. Infrared safety addresses the infrared divergent nature of QCD and requires the resulting jet definition to be insensitive to low-energy QCD radiation. In addition, collinear safety requires insensitivity to a collinear splitting of jet constituents.

Three IRC-safe, sequential recombination algorithms are mainly used,
CAMBRIDGE/AACHEN, the k_{T}-ALGORITHM and ANTIk_{T}-ALGORITHM. Common for all three are an external radius parameter R and the general distance definition. At every step of the clustering, the distance d_{ij} between all pairs of constituents ij,

$$d_{ij} = min(k_{\mathrm{T}i}^{2p}, k_{\mathrm{T}j}^{2p}) \frac{\Delta R_{ij}^2}{R^2} \tag{4.1}$$

and the distance d_{ib} between a constituent and the beam line,

$$d_{ib} = k_{\mathrm{T}i}^{2p} \tag{4.2}$$

are determined. Where ΔR is the usual angular distance $\sqrt{\Delta\eta^2 + \Delta\phi^2}$ and k_T^i the transverse momentum of constituent i. When the smallest distance is d_{ij}, constituents i and j will be combined to a protojet and treated as a constituent in the next step. If d_{ib} is smallest, the entity i will be called a jet and removed from the list. This is repeated until all constituents are clustered to jets.

The characteristic of the Cambridge/Aachen algorithm is, that the parameter p in the distance measure is set to zero, only the geometric distance is considered.

Whereas the k_{T}-algorithm is defined by $p = 1$. This algorithm tends to cluster the low-momentum parts first and the high p_{T} regions last. Thus it is possible to explore the clustering history of the k_{T}-algorithm to gain information about hard substructure, which is done by k_{T}-splitting scales, see e.g. Reference [27]. The Cambridge/Aachen and k_{T}

algorithms feature rather irregular shapes, since they depend on the exact arrangement of soft radiation.

Additionally, $p = -1$ defines the antik_{T}-algorithm, see Reference [28]. This recombination algorithm is characterized by conical jet shapes of radius R, clustered around a hard structure. This results from the distance measure between a hard and a soft constituent, that is dominantly defined by the p_{T} of the harder constituent and is therefore smaller than the distance measure between two soft constituents of similar angular separation ΔR. The well defined, conical jet shape facilitates further studies and calibration of the jet. An example of the three sequential recombination algorithms as well as SISCone, which is not of importance for this study, is given in Figure 4.1.

An important concept in the context of clustering algorithms is Ghost Association. This is used to link given objects, e.g. tracks or track-jets, to a calorimeter jet. The mass of the objects of interest are set to zero and their p_{T} to an infinitesimal small value (ghosts). Only the information about the position in $\eta \times \phi$ is used. The ghosts are added to the list of constituents for the jet algorithm, that runs again. After clustering, it is possible to determine, which ghosts were clustered to a given calorimeter jet. The calorimeter jet is not affected in shape or energy scale due to the infinitesimal momentum of the ghosts.

This allows the definition of the active area of a jet by running the jet algorithm with a very large number of ghosts positioned into a grid in $\eta \times \phi$ and again selecting the ones that were clustered to the calorimeter jet of interest. This jet areas are as well shown in Figure 4.1.

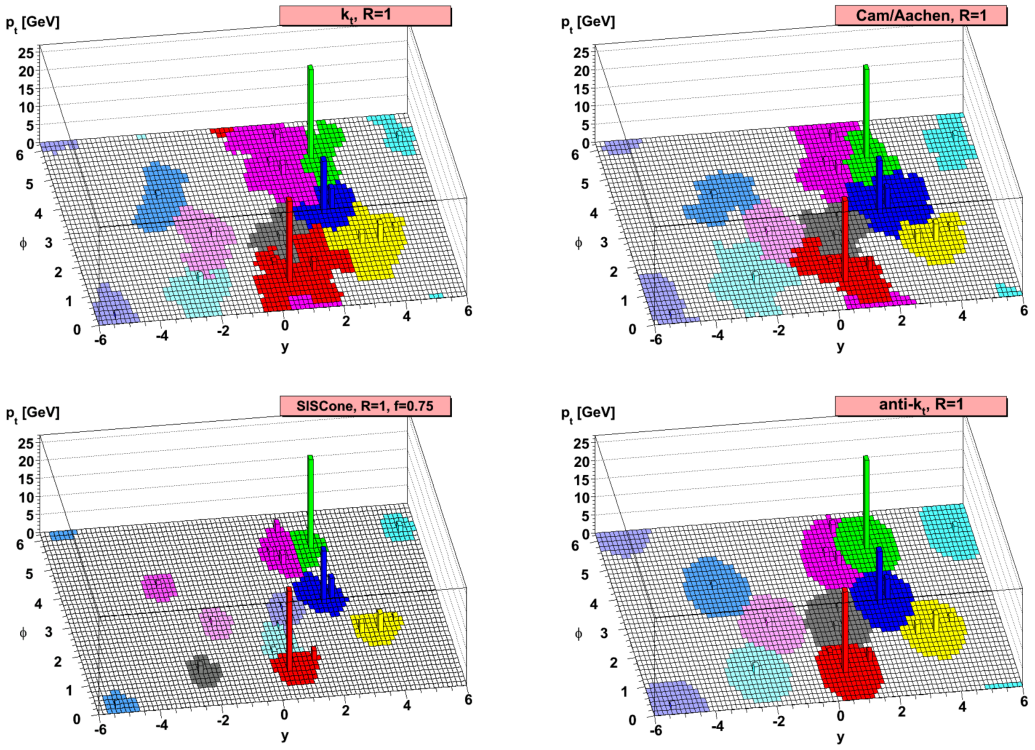

Figure 4.1.: Jet shapes and active areas of different recombination algorithms, taken from Reference [28].

For events with un- or just slightly boosted W/Z bosons, Higgs bosons or top quarks,

the decay products are sufficiently separated to distinguish between jets originated by the single quarks. A radius parameter of $R = 0.4$ is used in this energy regime, as every decay product can be easily matched to a certain jet of this size.
In the (highly) boosted case, it can be ambiguous which energy deposition should be assigned to a which quark, this is indicated for the W boson decay in Figure 4.2. The usual procedure is to cluster all decay products together in one jet with a radius parameter of $R = 1.0$ and to use substructure techniques to obtain additional information. The resulting jets are called LARGE-R or fat jet.

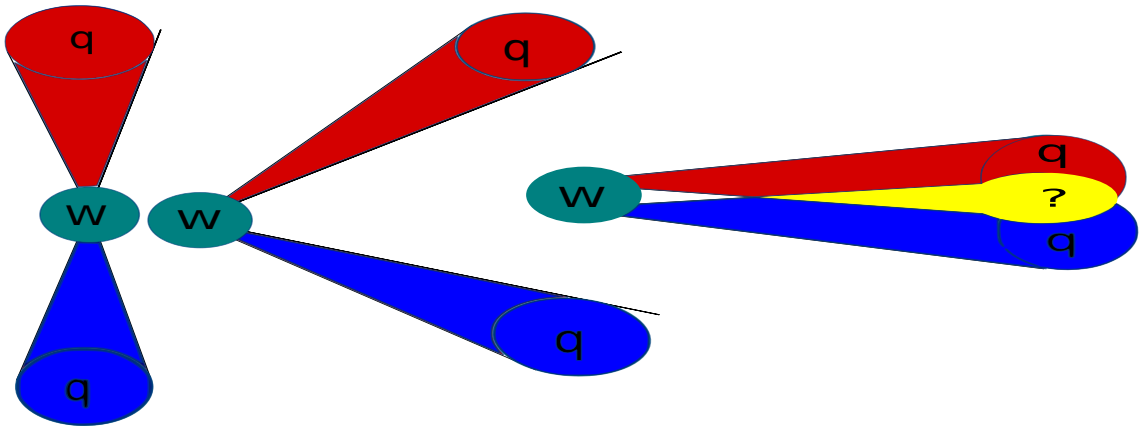

Figure 4.2.: W boson decaying to two quarks: no boost (left), intermediate boost (middle), highly boosted W boson (right).

4.2. Calibration

The energy of the reconstructed jet does not match the exact energy of the hadrons forming the jet because of detector imperfections, PILE-UP, UNDERLYING EVENT (UE) contamination and INITIAL STATE RADIATION (ISR).

In one bunch-crossing, many proton-proton collisions, μ, occur alongside the considered hard-scattering. These interactions predominantly feature a low momentum transfer and are called pile-up.
In addition, the UE is composed of the remnants, and their interactions, of the hard-scattering protons, which are colour connected to the event.
ISR are additional radiations from the initial state, also visible in Figure 2.9. The considered final state recoils against this ISR.
These effects are responsible for mostly low energetic signatures in the detector.

Dead material is instrumentation such as cables and support structures in which particles deposit energy without detection. Furthermore, the response, described by the ratio of the true energy value divided by the measured/reconstructed value, of the non-compensating calorimeters in the ATLAS detector to electromagnetic and hadronic interactions is different. A hadronic shower consists mainly of pions. The neutral part (π^0) decays to photons and imitates an electromagnetic shower.

In contrast, the charged part ($\pi^{\pm}$) forms a hadronic shower. Thereby, the response to the final jet is composed of variating hadronic and electromagnetic parts, which has to be corrected for.

The calibration of jets tries to account for these effects with a scheme of successive steps, see e.g. Reference [29]. The calorimeter clusters are calibrated to the electromagnetic scale, resulting in the energy response to EM interactions to be on average unity. The Local Cluster Weighting (LCW) calibration is applied on the EM constituents depending on the use case, correcting for energy fluctuations due to non-compensation. Up to this point, the calibration is done on jet constituent scale.
Further calibrations are applied on the clustered jet, shown in Figure 4.3. The final jet is then referred to as EM+JES (jet energy scale) or LCW+JES.

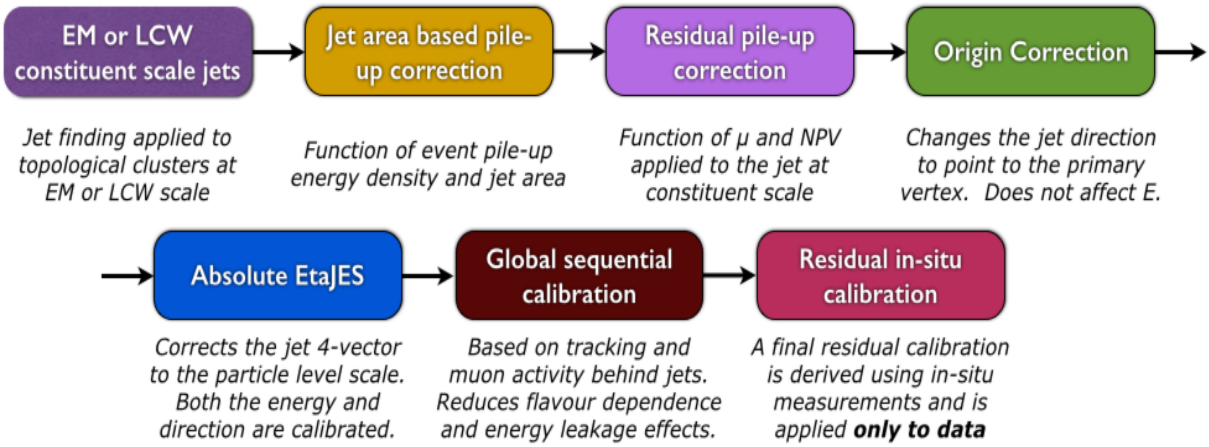

Figure 4.3.: ATLAS calibration scheme. NPV describes the number of primary vertices. The absolute EtaJES is derived from MC and the Global sequential calibration is only applied on jets with constituents on EM scale. Taken from Reference [30].

4.3. Grooming

Grooming describes the reduction of soft components in the jet from pile-up, UE and ISR, more details can be found e.g. in Reference [27]. For this, there are different methods such as Trimming, mass-drop filtering and pruning.

Important for this thesis is the trimming method, because it is the the standard recommendation for W/Z boson, as well as Higgs boson and top quark tagging.
Trimming addresses the softer components of the jet by reclustering the large jet with the k_{T}-algorithm and a small radius parameter, in this case $R = 0.2$. This results in a number of small jets, called Subjets, inside the jet. In the last step, all subjets with a momentum fraction, f_{cut}, of less than usually a few per-cent compared to the whole jet are removed. The trimming method is illustrated in Figure 4.4.

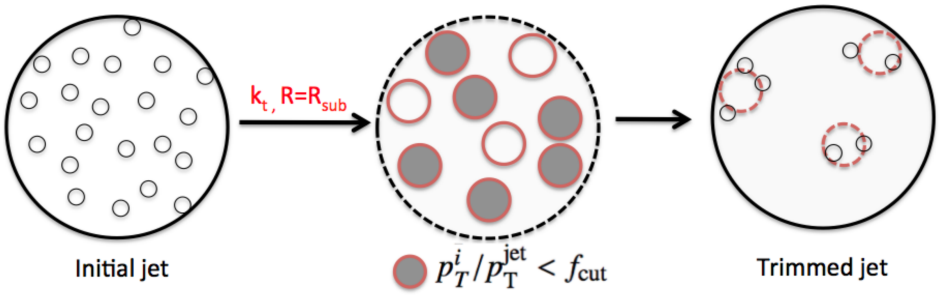

Figure 4.4.: The trimming procedure, taken from Reference [27].

4.4. Substructure Variables

In this section, important variables characterizing the substructure of a jet are introduced. The mass of gluons is zero and the mass of quarks (except for the top quark) is vanishing in comparison with the momentum transfer at as the LHC. The jet mass is deduced from energy depositions and the angle between them. Consequently, background jets, high p_{T} quarks or gluons, can acquire a mass similar to the one of the considered signal, W/Z bosons, Higgs bosons and top quarks. For this reason, the jet mass alone does not provide sufficient information about a given large-R jet to effectively distinguish signal and background.

All studied variables are designed to explore the substructure of large-R jets to discriminate between different event topologies. These are labeled one-, two- or three-prong like structures, arising from one, two and respectively three hard substructures inside the jet. A one-prong like jet originated from a single quark or gluon (QCD jet) and features a hard core of energy deposition with diffuse soft, wide angle emissions due to low energetic QCD splittings.
The two- and three-prong like jets are characterized by one and two hard energy splittings with fewer additional soft, wide angle emission inside the jet, which is illustrated in Figure 4.5. These figures represent two body decays such as $W \to q\bar{q}'$ and $H \to b\bar{b}$, three body decays such as $t \to Wb \to q\bar{q}'b$ and QCD jets.
This distinction helps to separate signal events, jets formed by hadronically decaying W/Z and Higgs bosons which are expected to be two-prong like, as well as top quarks jets with a three prong structure, and the QCD background from single quark/gluon jets, featuring an expected one-prong like structure.

4.4.1. N-Subjettiness

Historically, the first of the studied jet substructure variables was N-Subjettiness. A detailed description can be found in Reference [31].
The n-Subjettiness variable τ_N quantifies the level of agreement between a given jet and a certain number N of subjet axes defined.

There are several possibilities to define the subjet axes. Two often used axes definitions

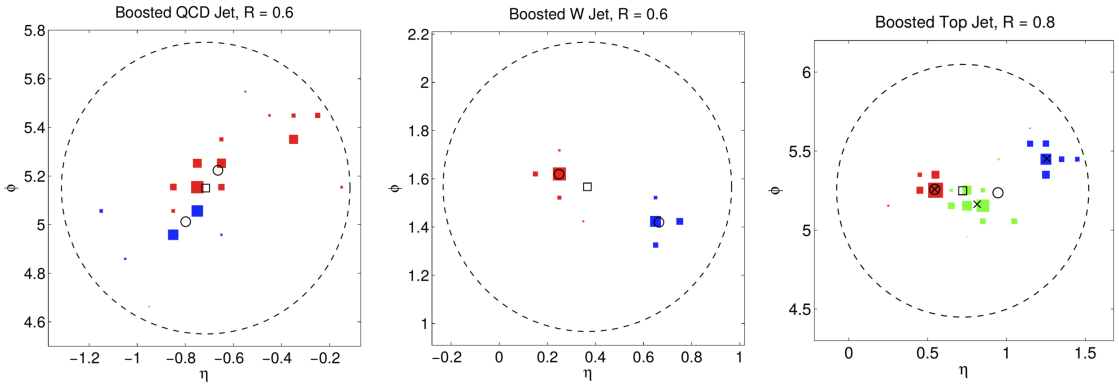

Figure 4.5.: Typical one-prong QCD jet (left), two-prong W boson jet (middle) and three-prong top quark jet (right) event displays, taken from Reference [31].

are k_T-axes and the k_T-WTA (Winner Takes All) definition. In both cases, the jet is reclustered with an exclusive k_T-algorithm, that means running the recombination just until N subjets are clustered. The k_T-axes are then defined by the four-momenta of the k_T-subjets while the k_T-WTA axes correspond to the four-momenta of the hardest constituent of each k_T-subjet.

N-Subjettiness is a measure for the whole jet and is calculated via a sum over the jets constituents (for example calorimeter clusters or tracks).

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k}\, min(\Delta R_{1,k}, \Delta R_{2,k}, ..., \Delta R_{N,k})^\beta \tag{4.3}$$

For each term, the constituents p_T is multiplied by the distance to the nearest subjet axes. The value is normalized with a sum over the constituents p_T times the characteristic radius parameter of the large jet R.

$$d_0 = \sum_k p_{T,k} R_0 \tag{4.4}$$

Furthermore, the weighting of the angular measure ΔR_{ij} relative to the p_T factor can be varied via the parameter β. N-Subjettiness is an IRC-safe variable for values of $\beta \geq 0$.

In general, it is possible to distinguish two areas of τ_N values. Small values of τ_N correspond to a jet where all constituents are more or less aligned or near to the given N subjet axes, hence the jet is compatible with the assumption to be composed of N or fewer subjets.
A higher value in contrast indicates a consistency with more than N subjets as a non negligible part is located further apart of the N subjet axes.
Consequently, W/Z or Higgs boson jets are likely to feature a small τ_2 and a high τ_1 value. For QCD jets with their one-prong structure, one can invert the argument, they result in a high τ_2 and a small $\tau 1$ value. While τ_1 and τ_2 alone provide only slightly separation, the ratio

$$\tau_{21} = \frac{\tau_2}{\tau_1} \tag{4.5}$$

is an effective discrimination variable, as can be seen in Figure 4.6.

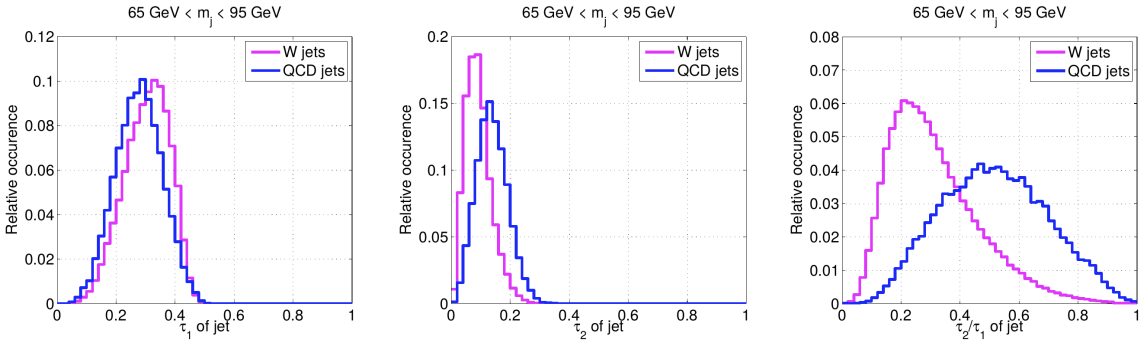

Figure 4.6.: τ_1 (left), τ_2 (middle) and τ_3 (right) distributions for QCD and W jets, taken from Reference [31]

The extension to three-prong like jet identification and discrimination from one and two-prong structures follows quite naturally by taking the ratio of τ_3 and τ_2.

$$\tau_{32} = \frac{\tau_3}{\tau_2} \tag{4.6}$$

Consequently, the hadronic decay of top quarks via $t \to Wb$ and the W decaying into two quarks can be tagged using the τ_{32} variable.

4.4.2. Energy Correlation Functions

The Energy Correlation Functions ECF(N,β) or N-point correlators, described in Reference [32], are similar to the n-Subjettiness variable. This variables as well explore the substructure of a jet by a sum over the constituents p_T, weighted by an angular factor that can be scaled via an exponent β.

In contrast, the correlation between pairs and triples of constituents is considered by the product of their p_T and the angular weighting, which is defined by the product of the pairwise angular distances of the considered constituents.

$$\begin{aligned}
\mathrm{ECF1} &= \sum_{constituents} p_\mathrm{T} \\
\mathrm{ECF2}(\beta) &= \sum_{i=1}^{n} \sum_{j=i+1}^{n} p_{\mathrm{T},i} p_{\mathrm{T},j} \Delta R_{ij}^{\beta} \\
\mathrm{ECF3}(\beta) &= \sum_{i=1}^{n} \sum_{j=i+1}^{n} \sum_{k=j+1}^{n} p_{\mathrm{T},i} p_{\mathrm{T},j} p_{\mathrm{T},k} (\Delta R_{ij} \Delta R_{ik} \Delta R_{jk})^{\beta}
\end{aligned} \tag{4.7}$$

The ECF variables can be expanded straightforwardly to larger values of N by considering this definition.

Like this, ECF2 uses pairwise correlation to be sensitive to two-prong structures and ECF3 relies on triple-wise correlations to identify three-prong structures. ECF1 however, represents the p_T of the overall jet by a summation over the constituents p_T, and serves as normalization for minimization of energy scale dependence.

The ECFN variable tends to very small values for collinear or soft configurations of N constituents and is defined as zero for less than N constituents. For ECF2, only pairs of constituents that are angular separated but not soft result in sum terms that are non-negligible, which directly leads to the picture of two hard substructures inside the jet. A similar conclusion can be made for ECF3 and three hard prongs.

Resulting from this, a jet with N or more hard substructures features a high ECFN value while a jet with fewer than N substructures has a lower ECFN value. Consequently, one can define, analogues to n-Subjettiness, ratios of Energy Correlation Functions. Two of them, called C2 and D2 are found to be very powerful to distinguish between one- and two-prong like jets, see e.g. Reference [33].

$$\begin{aligned} \mathrm{C2} &= \frac{\mathrm{ECF3} \cdot \mathrm{ECF1}}{\mathrm{ECF2}^2} \\ \mathrm{D2} &= \frac{\mathrm{ECF3} \cdot \mathrm{ECF1}^3}{\mathrm{ECF2}^3} \end{aligned} \tag{4.8}$$

A W boson jet for example has a small ECF3 value, but a high ECF2 value resulting in a very small C2/D2 variable, corresponding to a high agreement with the two-prong hypothesis. QCD jets as well feature a small ECF3 value, but in contrast a small ECF2. This results, considering the power of ECF2 in the definitions, in a higher C2/D2 value as for a W jet.

These variables are IRC-safe for $\beta > 0$ and theoretically very well understood, see Reference [34]. D2 was found to perform slightly better for tagging W jets as C2 in Reference [35], concerning background rejection and p_{T} robustness of the cut value.

Moreover, there is a corresponding ratio C3 build from ECF4, ECF3 and ECF2 to identify three prong like structures as produced by the top quark. Nevertheless, studies in Reference [32] have shown, that C3 does not perform as well as the n-Subjettiness ratio τ_{32}. Consequently, the effect of tracks and TAS is only studied on τ_{32} for this thesis.

4.5. Track Assisting Substructure Variables

Track assisting was first proposed for the jet mass variable, this concept is shown in Chapter 4.5.1. The used adaption for the studied observables deduced from the Energy Correlation Functions and n-Subjettiness is described in the following section 4.5.2.

4.5.1. Jet Mass

Usually, the mass of a jet is reconstructed via the p_{T} of the particles in the final state and the angle between them, corresponding to the energy measured by the calorimeter clusters and their angular distance. For a general two body decay such as a hadronically decaying W boson, neglecting quark masses:

$$M^2 \approx 2(E_1 E_2 - p_1 p_2 \cos\theta) \tag{4.9}$$

The hadronic decay of a Higgs boson, mainly to $b\bar{b}$, is topologically very similar to $W \to q\bar{q}'$ or $Z \to q\bar{q}$ since both feature a two-prong structure. The difference is the higher Higgs mass, which increases the angular separation of the $b\bar{b}$ pair in comparison with the W/Z boson decay products. In contrast, the hadronic decay of the top quark features a characteristic three-body structure.

In case of the exemplary two body decay, the quarks will be produced back to back, W/Z or Higgs boson at rest, while for increasing boost of the decaying particle, the angle θ between the decay products shrinks and the angular measure becomes increasingly important for the mass computation, see Figure 4.2.

Tracks are measured very precisely in η and ϕ and could thereby improve variables as the mass that rely on a angular measure in the highly boosted regime, when decay products are very close to each other. Nevertheless, tracks are only produced by charged particles, hence the neutral part of the hadronic shower is missed, as mentioned in Chapter 4.1. The missed part fluctuates from jet to jet and results in a degradation of the resolution and a lower energy scale of the track jet mass compared to the calorimeter jet mass.

A relatively new ansatz within the ATLAS collaboration is to use the excellent angular resolution of tracks and to scale tracks with the p_{T} ratio of the track jets and the corresponding calorimeter jets to include information about the neutral fraction. This procedure is called Track Assisting (TA).

$$m_{\mathrm{TA}} = m_{\mathrm{track-jet}} \cdot \frac{p_{\mathrm{T,calo-jet}}}{p_{\mathrm{T,track-jet}}} \tag{4.10}$$

Where $m_{\mathrm{track-jet}}$ describes the mass of track jet formed by all tracks ghost associated to the calorimeter subjets.

First results of calculating the mass using the track assisting method are promising but the response for W boson jets is worse compared to the calorimeter jet mass at p_{T} values lower than $\sim 1300\,\mathrm{GeV}$, as can be seen in Figure 4.7.

A possible improvement is to correct for the neutral part locally for each hard substructure. The two quarks of the W decay that are captured by the large-R jet hadronise differently and so the neutral fraction will vary. To account for this, single tracks, instead of track-jets, are ghost associated to the $R = 0.2$ k_{T}-subjets of the large-R calorimeter jet, which remain after the trimming procedure.
The p_{T} sum of all tracks ghost associated to a certain calorimeter subjet is denoted as $\sum_{gatracks} p_{\mathrm{T},track}$. Now, all tracks ghost associated to a certain calorimeter subjet are scaled with the ratio of this calorimeter subjet and the corresponding p_{T} sum of ghost associated tracks.

$$m_{\mathrm{TAS}} =' \sum_{\mathrm{tracks}}^{'} m_{\mathrm{track}} \cdot \frac{p_{\mathrm{T,subjet}}}{\sum_{\mathrm{gatracks}} p_{\mathrm{T,track}}} \tag{4.11}$$

The 'S' in the subscript TAS stands for the association to subjets and 'ga' for ghost association. $'\sum'_{\mathrm{tracks}}$ represents the deduction of the jet mass from the four-momenta of tracks with rescaled mass.

Currently, the performance of m_{TAS} is studied by the ATLAS collaboration, see again Figure 4.7, and a possible combination with the calorimeter mass is considered.

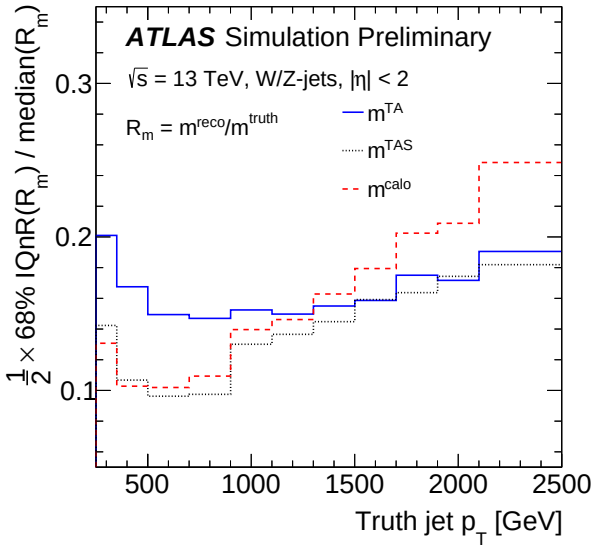

Figure 4.7.: Figure showing the performance of the calorimeter jet mass (calo) against the track assisted (TA) and the subjet track assisted (TAS) mass. The y-axis shows the IQnR (Inter Quantile Range), a figure of merit representing the width of the response function (reconstructed mass divided by truth mass) normalized to the median, taken from Reference [36].

Given the enhancements due to assisted tracks for the jet mass, it could be possible to improve other substructure variables, in particular n-Subjettiness and the Energy Correlation Double Ratios, which rely on the angular distance between constituents (and subjet axes), when calculating with assisted tracks rather than calorimeter clusters. Possible improvements for these discrimination variables would result in a higher QCD background rejection for cases such as W/Z boson, Higgs boson or top quark tagging.

4.5.2. Implementation for this Study

In contrast to the m_{TAS} variable introduced in Chapter 4.5.1, not the mass but the p_{T} of each track is scaled, since C2, D2, τ_{21} and τ_{32} are calculated with the constituents p_{T}.

The concept of track assisting with the p_{T} ratio of the whole jet is without effect for the studied substructure variables. This can be understood from the definitions of the weighted p_{T} sums. If corrected with only one ratio, all tracks are scaled by the same factor c, which then can be put in front of the sum and cancels as soon as the ratios τ_{21} and τ_{32}, respectively C2 and D2 are formed.

$$\begin{aligned} \tau_N &= \frac{1}{d_0} \sum_k p_{T,k}\; c\; min(\Delta R_{1,k}, \Delta R_{2,k}, ..., \Delta R_{N,k})^\beta \\ &= \frac{c}{d_0} \sum_k p_{T,k}\; min(\Delta R_{1,k}, \Delta R_{2,k}, ..., \Delta R_{N,k})^\beta \end{aligned} \tag{4.12}$$

Track assisting with ghost association to subjets (TAS) works with different scaling factors depending on the corresponding subjet c_k, which also affect ratios:

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k}\; c_k\; min(\Delta R_{1,k}, \Delta R_{2,k}, ..., \Delta R_{N,k})^\beta \tag{4.13}$$

This leads to the following used adaption of the TAS procedure:

$$\begin{pmatrix} m_{track} \\ p_{T,track} \\ \eta_{track} \\ \phi_{track} \end{pmatrix} \rightarrow \begin{pmatrix} m_{track} \\ p_{T,track} \frac{p_{T,subjet}}{\sum_{ga\ tracks} p_{T,track}} \\ \eta_{track} \\ \phi_{track} \end{pmatrix} \tag{4.14}$$

The standard calorimeter variables with an angular weighting $\beta = 1$ are directly taken from the LCW antik_{T} $R = 1.0$ jet collection of the used Monte Carlo samples. Variations of different β are calculated from the constituents on LCW with the help of the n-Subjettiness and Energy Correlator packages provided by FastJet Contrib, for the FatJet package see Reference [37].

N-Subjettiness is used with the k_{T}-WTA axes, which is the standard choice for the axes definition. In comparison, the variables are also computed with tracks and assisted tracks as input by using the same FastJet packages.

Tracks are selected with the TrackSelectionTool which applies quality criteria depending on the chosen cut level, used is the cut level 'loose'. Furthermore, a match to the primary vertex via the TightTrackVertexAssociationTool is required, where the primary vertex is defined as the vertex with the highest associated momentum transfer. Summarized:

- $p_{\mathrm{T}}^{track} > 400\,\mathrm{MeV}$
- $|\eta| < 2.5$
- Minimum 7 pixel + SCT hits
- Maximum pixel hole: 1
- Maximum silicon holes: 2
- No more than two shared modules
- Maximum z displacement to vertex: $z_0 = 2\,\mathrm{mm}$
- Distance in x-y between vertex and point of closest approach: $d_0 = 2.5\,\mathrm{mm}$

The p_{T} of the calorimeter jet is required to be greater than 250 GeV because the boosted regime is of interest and to minimize impact from ISR and pile-up. Furthermore, the calorimeter jet axis is restricted within $|\eta| < 2.0$ for comparison with tracks which are available up to $|\eta| < 2.5$.

Grooming is done via trimming with the k_{T}-algorithm, a radius parameter of 0.2 and a p_{T} fraction of 5% which the subjets have to exceed. The remaining subjets are the ones used for the ghost association of the tracks.

Truth identification is used to filter only W boson, respectively Higgs boson or top quark events. As the mass difference between Z and W is small and only the hadronic

decays are considered, the results of this study for W boson tagging can be transferred directly to the Z boson.

No additional b quark tagging information is used when studying the boosted Higgs or Top scenarios. Furthermore, all events with a top pair not decaying full-hadronically were exuded to prevent effects resulting from τ leptons generating a jet.

Tagging variables such as C2, D2, τ_{21} and τ_{32} are usually used after applying a mass cut around the interval that contains 68% of the signal events. Therefore, a cut is applied on the calibrated mass of the large-R calorimeter jet which is calculated to cover the smallest interval around the peak mass that contains 68% of the signal events.

4.6. Receiver Operator Characteristics

The separation power of discrimination variables can be studied quite intuitively by comparing the signal and background distributions of a certain variable. Another used figure of merit for the performance, especially for comparisons of different variables, is to use Receiver Operator Characteristics (ROCs) which show the achieved background rejection for different values of signal efficiency (signal fraction left after performing a cut).

Each point is calculated from the underlying signal and background distributions by integrating the background distribution from zero [1] to the point where the desired signal fraction is achieved. The fraction of background events contained in this region are kept when cutting at this signal efficiency, hence the inverse of this fraction, $\frac{1}{\epsilon_{background}}$ is an estimate for the background rejection. The lower the fraction of background events in the region, the better is the achieved exclusion. Accordingly, a good discrimination variable is represented by a ROC with preferably high values of background rejection up to high signal efficiencies.

ATLAS uses working points to compare variables at fixes signal efficiencies, which are then provided for use to the analysis. Studied here are the working points at 50% and 25% signal efficiency after applying the mass cut and tagging with a substructure variable. Depending on the amount of statistics and background predictions, an analysis would choose between the working points, 25 % e.g. results in only one fourth of signal events that are kept but usually features a very high background rejection.

The stated signal efficiencies are calculated after the mass cut plus tagging with n-Subjettiness or C2/D2. Therefore, the endpoint of the ROCs is at 68% signal efficiency, the fraction kept after the mass cut and before cutting on a tagging variable. Consequently, it is required to achieve a tagging only signal efficiency of $\frac{0.5}{0.68} \sim 0.74$ for a signal efficiency of 0.5 after mass cut and tagging. Similarly, the stated and compared background rejections result from the multiplication of the rejection provided by the mass cut and the rejection achieved by the considered tagging variable, thus representing the separation by the combination of both.

[1] If the signal distribution lies at lower values as the background.

5. Monte Carlo Simulation

Monte Carlo (MC) techniques are widely used to study problems with an underlying probability density function (pdf) that can not be solved analytically, by numerically computing and evaluating a large number of trials, called pseudo-experiments. For each trial, the problem or rather the corresponding function is evaluated at a random value of the pdf, which is varied within its uncertainties. This is repeated until sufficient information about the resulting pdf is available to deduce estimations for e.g. the mean value or other describing parameters.

In particle physics, MC is used to simulate events in different steps. First, the hard scatter is simulated at leading or next-to-leading order in perturbative QCD to obtain the matrix element of the interaction. The next step is to simulate the fragmentation and parton shower, which describes the process from final state partons to colourless bound state hadrons. Furthermore, pile-up events due to multiple interactions in one bunch crossing as well as the underlying event are simulated and overlayed. The knowledge about the state before any detector effects is stored as truth information inside the MC samples.

There are many different MC simulators for event generation available, including Sherpa, Pythia and Powheg. Sherpa uses the cluster fragmentation scheme mentioned above while Pythia for example relies on the Lund-String model. Some toolkits can be used to calculate the matrix element as well as simulating the hadronic shower, an example for this is Pythia. In contrast, e.g. AlpGen is used for simulation of the hard scatter and often combined with Herwig for the showering and hadronisation.

Within the ATLAS experiment, the detector effects respectively the interaction of the particles with the material is described with the Geant4 toolkit, see Reference [38]. This results in simulated energy depositions or hits inside the detector which are then translated via a simulation of the readout electronics into a signal (e.g. current or voltage) that can be analyzed with the same reconstruction as actual data. Additionally, a link between the truth information and the corresponding energy depositions respectively signals is stored.

The produced samples are used for comparison of data to theoretical predictions as well as calibration. Another important application, used here, is to analyze MC samples of signal and background predictions to study and improve techniques to distinguish between them.

5.1. Signal Samples

In this performance study, the signal samples with different resonance masses are predominantly used as a source of (highly) boosted W/Z bosons, Higgs bosons and top quarks,

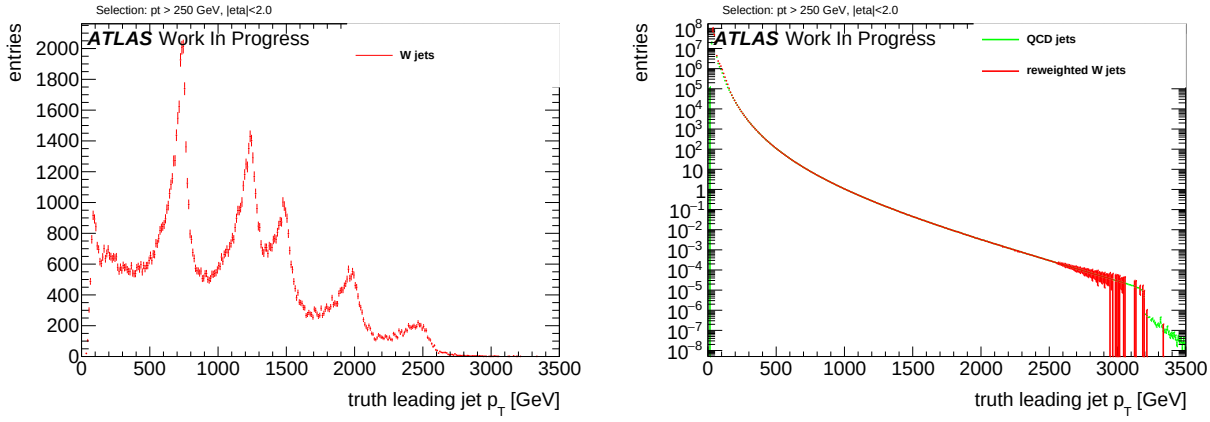

Figure 5.1.: p_{T} distributions of (left) W jets and (right) QCD jets from dijet events with reweighted W jets

whereas the actual resonance process is of lesser interest. For the W/Z bosons and top quarks, processes involving new HVT bosons are used, while samples with a Randall Sundrum graviton as resonance are used for the Higgs study. Detailed information about the used MC samples can be found Appendix A.

- $W' \to WZ$ with $m_{W'} = 1500, 2500, 3000, 4000, 5000\,\mathrm{GeV}$
- $G^* \to hh$ with $m_{G^*} = 500, 1000, 1500, 2000, 2500, 3000\,\mathrm{GeV}$
- $Z' \to t\bar{t}$ with $m_{Z'} = 1500, 1750, 2500, 3000, 4000, 5000\,\mathrm{GeV}$

5.2. Background Sample

The background which needs to be discriminated are QCD jets, formed by single quarks or gluons, which will hadronise and shower. A large source of QCD jets is the MC prediction of dijet events, occurring over the whole studied phase space.

5.3. Event weighting

While the p_{T} distribution of the dijet sample is exponentially falling, the p_{T} of the combined signal sample features characteristic peaks related to the different resonance masses, see Figure 5.1. To avoid bias in the substructure functions which are calculated from the constituents p_{T}, the combined signal sample is given weights sucht that the truth p_{T} distribution of the leading jet matches the one of the background sample. Furthermore, the spectrum is split into six different p_{T} regions to study the behavior with rising energy.

6. Results

This chapter presents the results of the performance study of the Energy-Correlation-Functions and n-Subjettiness variables with tracks and assisted tracks in comparison with the same variables calculated with calorimeter clusters. The influence of the topological track selection for the performance of substructure variables with tracks is studied in Section 6.1. The main results with an angular weight of $\beta = 1$ are presented thereafter in Section 6.2 through comparing signal and background distributions as well as Receiver Operator Characteristics for W/Z boson, Higgs boson and top quark tagging separately. In the last part of this Chapter, Section 6.3, the angular weight β is varied in order to identify and study the best tagging variables for these scenarios.

6.1. Track Selection

Tracks can be associated differently to a given calorimeter jet. Compared were tracks that are ghost associated to the ungroomed large-R calorimeter jet and tracks ghost associated to the k_{T}-subjets, which remain after the trimming of the large-R calorimeter jet.

The distributions showing the number of tracks associated to a calorimeter jet, see the left side of Figure 6.1, indicate, that on average around four tracks less are associated to the calorimeter subjets than associated to the whole large-R calorimeter jet. The right side of Figure 6.1 shows the topological difference between both collections with the angular distance ΔR between the single tracks, and the axis of the large-R calorimeter jet. Both distributions are aligned in the lower ΔR region while the histogram representing the tracks associated to the large-R jet shows an enhancement towards larger ΔR. Accordingly, tracks which are not associated to the subjets, but still associated to the ungroomed large-R jet feature an angular separation from the jet axis of more than 0.3, and are in consequence distributed primarily around the outer regions of the large-R calorimeter jet. Given the required primary vertex association, it is unlikely that these tracks originate from pile-up. Instead, the origin might be found in final- or initial state radiation.

Figure 6.2 shows the signal distributions of the Energy Correlation Double Ratios C2/D2, and the n-Subjettiness ratio τ_{21}, calculated with both selections of tracks for W jets. Especially for C2 and D2, but also for τ_{21}, the signal distributions calculated with tracks associated to the whole calorimeter jet are found to be broader and tending to more background like, higher values compared to the distributions calculated with tracks associated to the subjets. The large ΔR to the jet axis of the differing tracks push the substructure variables to higher, more background like values. Furthermore, the broader distributions are a result of the variating, large angular distance of the additional tracks. C2 and D2 are more sensitive to tracks with a large ΔR to the jet axis, since

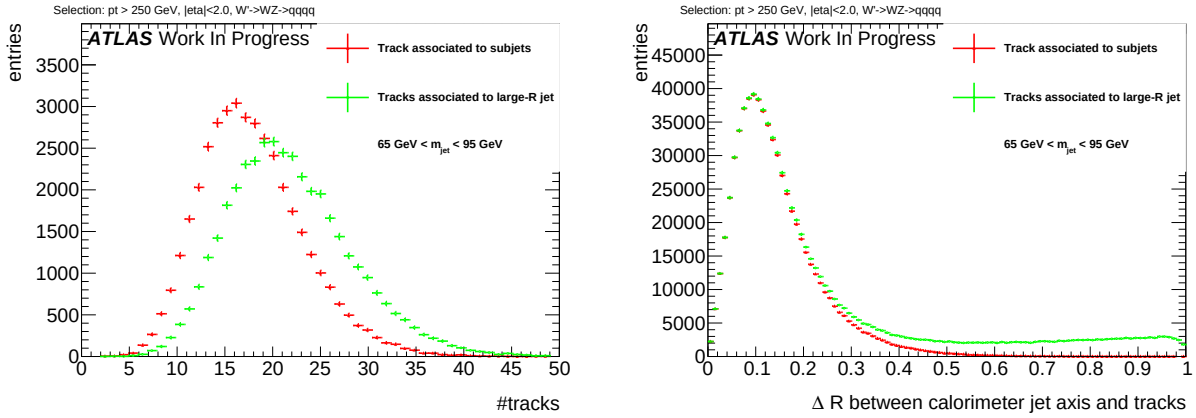

Figure 6.1.: The number of tracks ghost associated to the large-R jet and to the subjets (left) and angular distance of associated tracks to the large-R calorimeter jet axis (right). Signal events were not reweighted at this step.

the angular distance between all pairs and triples of tracks is considered For example correlations of tracks on opposite sides of the large-R jet, whereas τ_{21} uses distances to well defined k_{T}-WTA axes. For comparison, the signal and background distributions for the variables calculated with calorimeter clusters are shown as well. Through comparing the signal distributions calculated with tracks and clusters, it is possible to anticipate a performance of variables calculated with tracks, that is not worse compared to variables calculated with calorimeter clusters due to the similar width but lower mode of the track distributions.

In contrast, the jet mass variable performs considerably worse with tracks compared to the calorimeter, since the missing information about the fluctuating neutral fraction results in a smaller energy scale and a wider mass distribution. However, the substructure variables studied here are rather energy scale independent and are found to not be as sensitive to the missing fraction.

Starting from this observations, the performance of substructure techniques is compared with the following objects as input:

- Calorimeter clusters, labeled as 'calo'.
- Tracks that ghost associated to the calorimeter subjets remaining after trimming, labeled 'tracks'.
- The same collection of tracks, assisting as defined in Chapter 4.5, labeled 'TAS'.

6.2. General Performance

The p_{T} regions and corresponding mass windows containing 68% of the signal events are shown in section 6.2.1. Subsequent and as main part of this section, the performance of tracks and TAS with an angular weighting of $\beta = 1$ for tagging W bosons, see Chapter 6.2.2, Higgs bosons, see Chapter 6.2.3, and top quarks, see Chapter 6.2.4, is compared to the corresponding calorimeter variables.

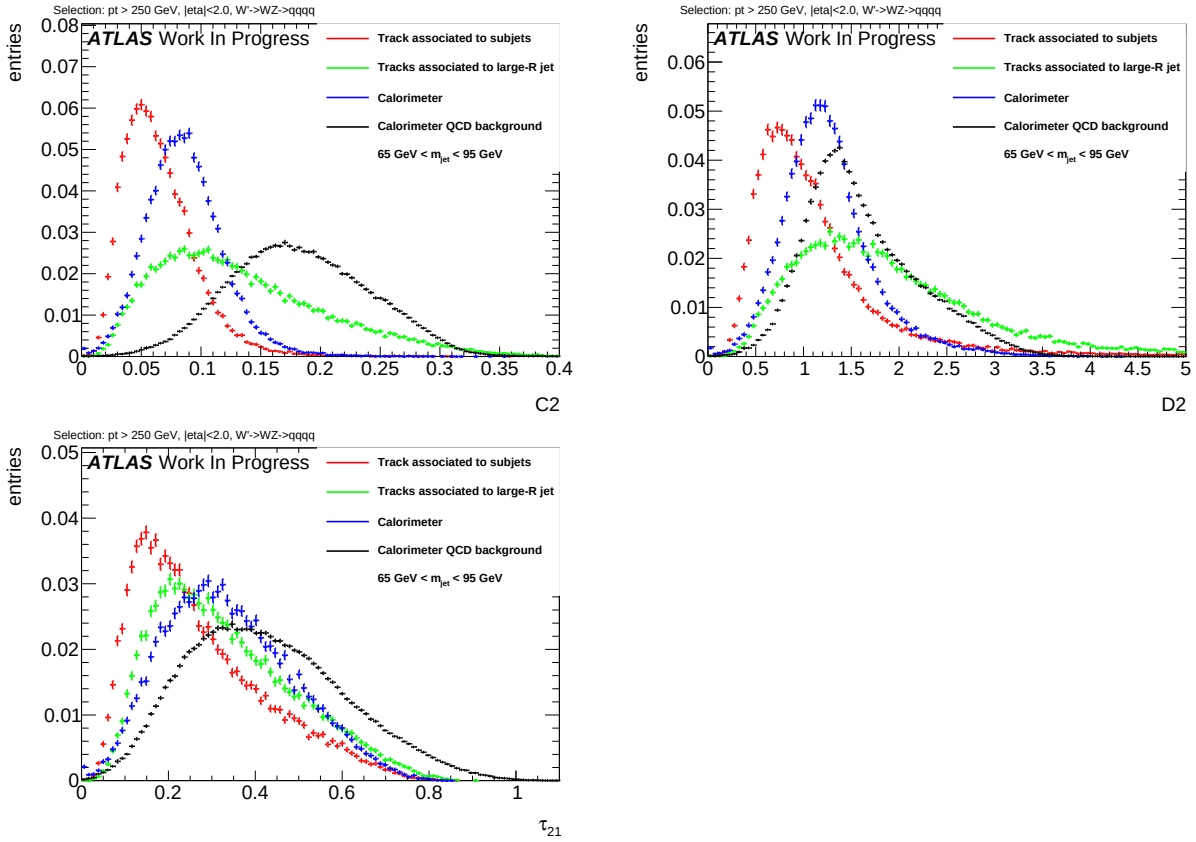

Figure 6.2.: Substructure variables (left) C2, (right) D2 and (below) τ_{21} calculated with calorimeter clusters as well as tracks associated to subjets and to the large-R jet. Signal events were not reweighted at this step.

6.2.1. Mass Windows and p_{T} Regions

The comparison is performed in six different p_{T} regions to study the behavior connected with rising energy of the decaying particle. These regions are presented in the left part of Table 6.1.

In case of the Higgs boson study, there is not enough statistics to derive a conclusive result for $p_{\mathrm{T}} > 2000\,\mathrm{GeV}$, since the highest resonance mass of the $G^* \to HH$ samples is 3000 GeV in contrast to 5000 GeV for the $Z' \to tt$ and $W' \to WZ$ samples. Hence this study is restricted to the five lower p_{T} bins.

Prior to tagging with the n-Subjettiness or C2/D2 variables, a cut on the calibrated calorimeter jet mass is applied, given that the mass is the main discriminant in QCD jet rejection. This cut is defined to choose the smallest interval around the peak mass containing 68% of the signal. However, the reconstructed mass depends on the p_{T} region, therefore a different cut was calculated for every region to meet the requirements.

Shown in Figure 6.3 are the calorimeter jet mass distributions in the lowest studied p_{T} region from $250 - 500\,\mathrm{GeV}$. For the Higgs boson, the low boost in this region results in a still high separation of the decay products, which results in Higgs boson tagged events with two large-R jets, each containing only one b quark. This produces a bump in the lower mass region and is less articulated for the lighter W boson as the mass of

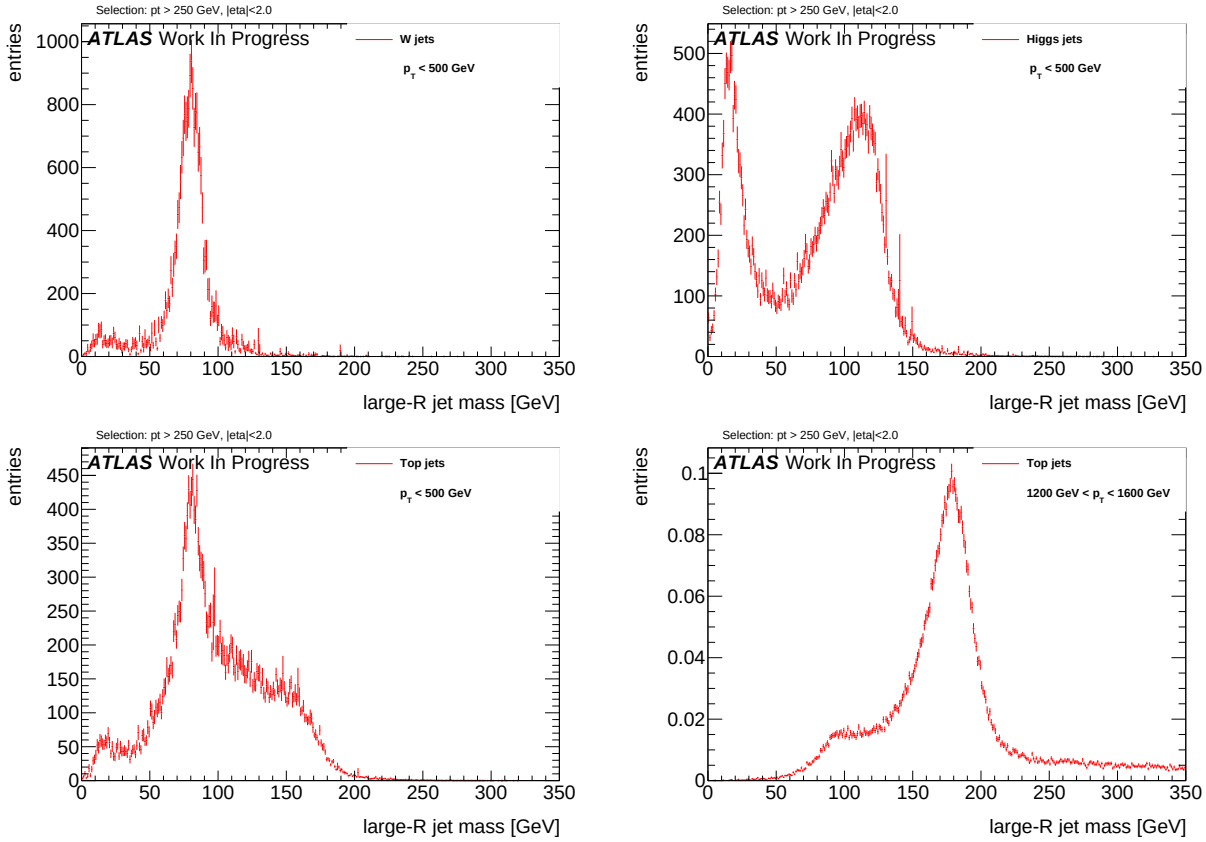

Figure 6.3.: Jet mass distributions in the first p_{T} bin for W boson jets (left) and Higgs boson jets (right). Below, shown are the top quark jet mass distributions for the first and the fourth p_{T} region, indicating the disappearance of the low mass shoulder with higher p_{T}.

the decaying particle is proportional to the maximal separation of its decay products, see section 2.7 in the theory chapter. In case of the top quark, it can as well be seen, that only a part of the top quark decays are reconstructed into a single large-R jet (right shoulder of the distribution), whereas in a large amount of events, one large-R jet contains only the W boson from the top quark decay (peak around the W boson mass) and the b quark is reconstructed into another jet. These cases become increasingly rare with rising boost, as can be seen in the lower right distribution of Figure 6.3, since the separation of decay products shrinks. The intervals were calculated with truth tagged events, allowing for cases just described. Consequently, the intervals for the first bins are larger and more pronounced to lower masses.

Furthermore, the background rejection achieved by the individual jet mass cuts alone were calculated and are presented in the right part of Table 6.1. The uncertainty on the jet mass window was estimated via variations of the calculated interval limits and evaluation of the covered signal fraction and lies around 1 GeV. This results in an estimation for the uncertainty on the corresponding background rejection due to the jet mass cut of about 2 %.

	W boson		**Higgs boson**		**Top quark**	
p_{T} [GeV]	Mass [GeV]	$\frac{1}{\epsilon_{bgr}}$	Mass [GeV]	$\frac{1}{\epsilon_{bgr}}$	Mass [GeV]	$\frac{1}{\epsilon_{bgr}}$
250 - 500	63 - 85	10.8	56 - 167	3.8	77 - 191	6.3
500 - 800	72 - 92	13.6	92 - 150	7.3	117 -205	6.9
800 - 1200	76 - 104	9.6	98 - 143	9.5	122 - 218	6.5
1200 - 1600	77 - 107	7.3	103 - 149	9.0	122 - 227	6.3
1600 - 2000	79 - 115	5.6	91 - 170	4.4	121 - 235	5.6
$>$ 2000	80 - 126	4.2	/	/	123 - 251	4.8

Table 6.1.: p_{T} regions and calculated 68% mass intervals along with the corresponding background rejections for W boson, Higgs boson and top quark jets.

6.2.2. Performance of W Boson Identification

First compared are the performance of calorimeter clusters, tracks and TAS for W tagging with the C2 variable, followed by D2 and the n-Subjettiness ratio τ_{21} using an angular weighting of $\beta = 1$. Distributions of variables calculated with tracks and TAS are quite similar and hard to compare, therefore distributions of track variables can be found in Appendix C.1. The actual performance comparison between tracks and TAS is done using ROCs. All distributions are scaled to unity in order to aid visibility and calculation of the ROCs.

Comparison of C2 for W Jets

Shown in Figure 6.4 are the distributions for C2 ($\beta = 1$) calculated with calorimeter clusters and TAS acting on W boson jets (signal) and QCD jets (background) in the lowest and and highest studied p_{T} region. The full set of distributions including the intermediate p_{T} regions can also be found in Appendix C.1. In the lowest p_{T} region, one can observe, that for TAS C2, the signal mode features a smaller value compared to the calorimeter C2 signal. The right handed tail is falling similarly and together with the steeper rising background for TAS C2, no articulated difference in separation power is expected.

With rising p_{T}, best visible in the highest p_{T} region, the signal distribution for TAS C2 rises steeper and moves to considerably smaller values compared to the signal distribution of calorimeter C2. The trend to lower values is also visible, but not as articulated for the mode of the TAS C2 background distribution, which features a more pronounced (slower falling) tail to higher values compared to the background of calculated with calorimeter C2. These observations indicate an enhancement of separation power for TAS C2 with rising p_{T}.

The ROCs in Figure 6.5 show the background rejection as function of the signal efficiency for TAS, track and calorimeter variables in the two lowest and the two highest studied p_{T} regions. The numerical value for the background rejection achieved at the 50% working point is displayed in the legend. In the lowest p_{T} region, the difference in rejection power is indeed small, with the calorimeter C2 variable performing slightly better

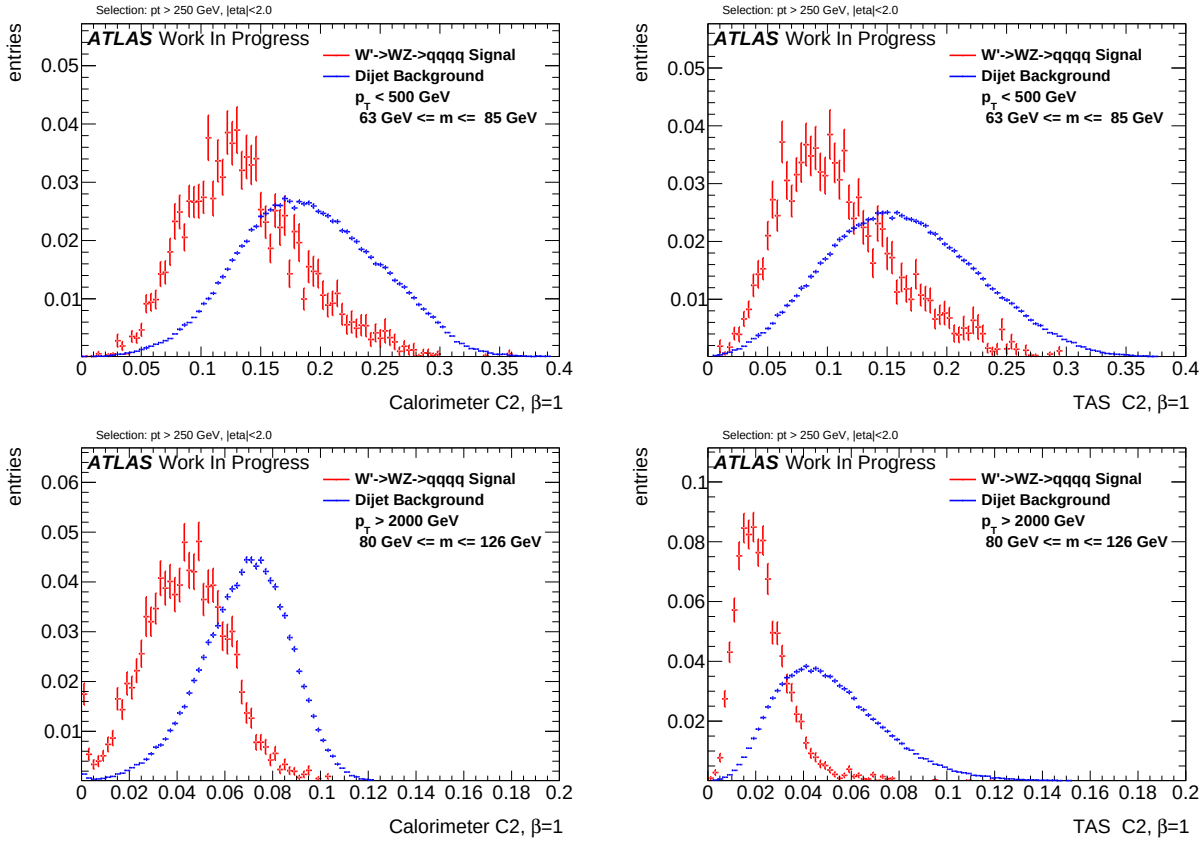

Figure 6.4.: W boson signal and QCD background distributions for calorimeter C2 (left) and TAS C2 (right) at $\beta = 1$ for the lowest (above) and highest (below) p_{T} bins.

than C2 calculated with tracks or TAS. This is expected, since both quarks are still well separated and the angular resolution in the calorimeter does not limit the performance of the calorimeter C2 variable. With rising p_{T}, the separation shrinks and C2 starts to benefit from the better angular resolution of tracks, resulting in track and TAS C2 to outperform calorimeter C2 in terms of QCD rejection. For very large boosts, the background rejection of tracks and TAS is about 50 % higher than achieved with calorimeter clusters. For C2, the performance difference between tracks and TAS is only slightly visible in the lowest p_{T} bin, for all other regions they discriminate equally well.

Comparison of D2 for W Jets

The distributions for D2 ($\beta = 1$), shown in Figure 6.6, can be studied similarly to the ones for C2. Inspecting the lowest p_{T} region, the signal distributions of D2 calculated with TAS rises steeper and therefore peaks at a smaller value compared to the calorimeter D2 signal counterpart. The background distribution of D2 TAS is shifted only very slightly to smaller values, predicting a higher separation power of TAS D2 over D2 with calorimeter clusters already in the first studied p_{T} region.

Moving to higher values of p_{T}, the signal distribution for calorimeter D2 is more symmetric as the TAS D2 signal distributions, which narrows due to the sharp rising left hand side. The QCD background is represented with TAS D2 by a distribution with

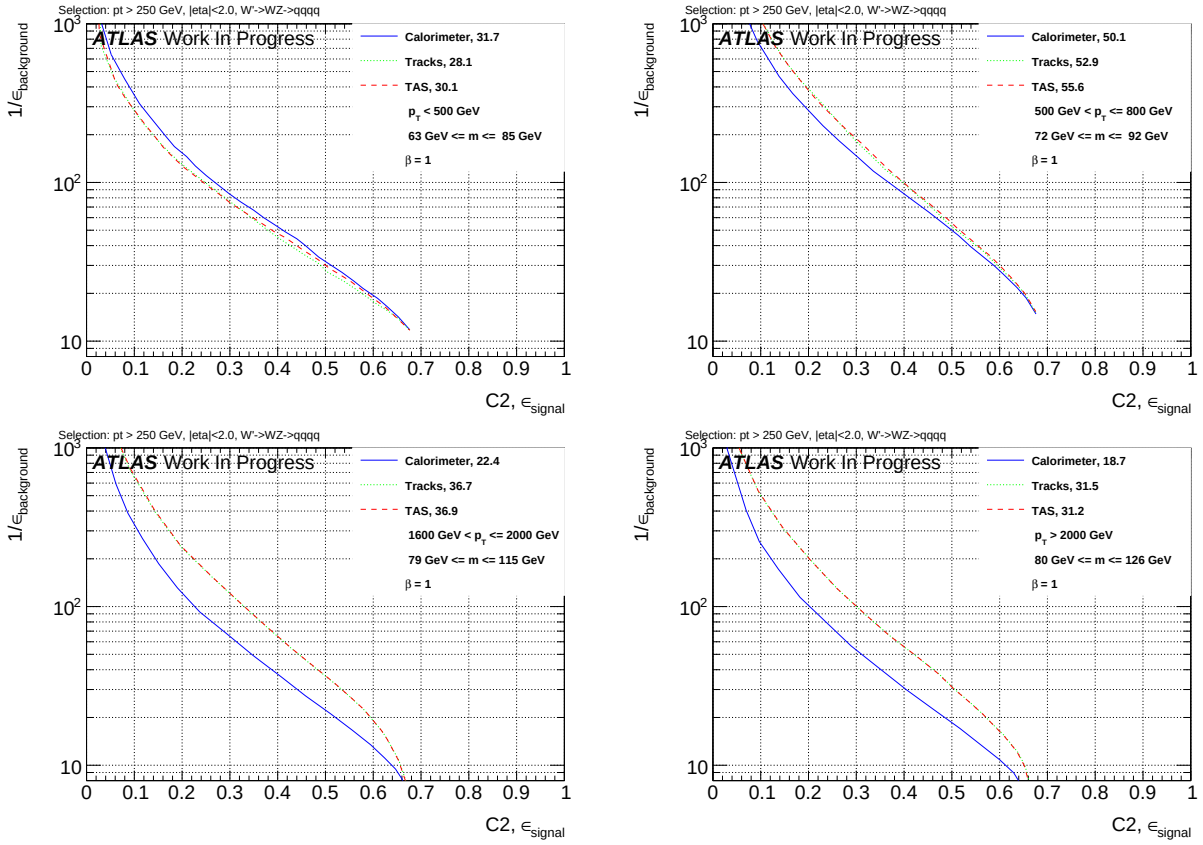

Figure 6.5.: ROCs showing QCD rejection against W boson efficiency for tracks, TAS and colorimeter C2 at $\beta = 1$, p_{T} ordering from upper left to lower right.

a smaller mode, D2 ≈ 2 compared to D2 ≈ 3 for the background distributions calculated with calorimeter clusters. Nevertheless, the tail of the D2 TAS background distribution becomes much more pronounced to larger values. This development intensifies with larger boost, promising a very good discrimination performance of TAS D2 at high transverse momenta.

Furthermore, it is possible to observe a higher energy scale robustness for D2 in comparison with C2, given that the distributions of D2 do not shift as distinctly as the ones of the C2 variable with varying p_{T}.

The ROCS in Figure 6.7, show that D2 TAS already performs slightly better compared to calorimeter D2 in the first p_{T} bin as anticipated by the signal and background distributions. Moreover, this enhancement increases with rising p_{T} due to the higher impact of the excellent angular resolution of the tracker. The background rejection of D2 calculated with calorimeter clusters is slightly higher for very high signal efficiencies and values of p_{T}. Nevertheless, since the regions around 25 % and 50 % signal efficiency are of interest, TAS D2 can be concluded to outperform the calorimeter D2 variable for all studied energies of the W boson.

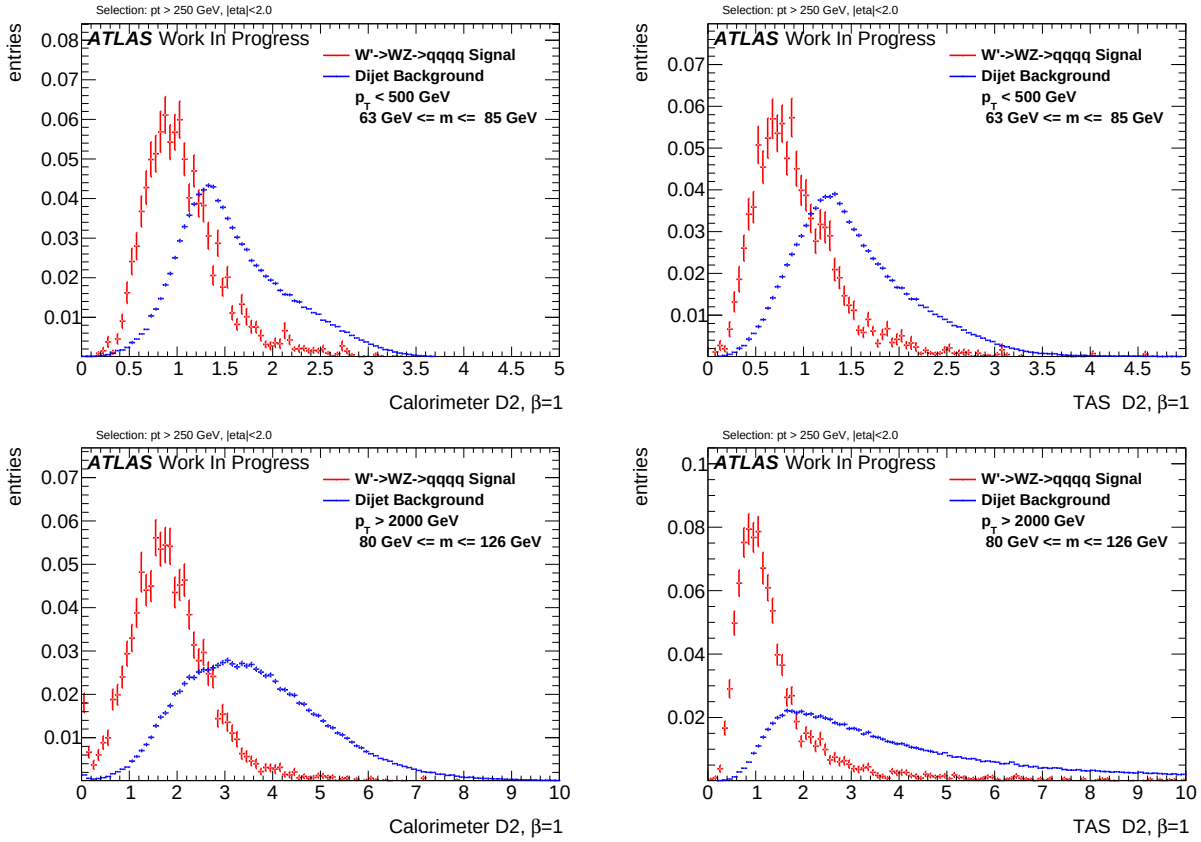

Figure 6.6.: W boson signal and QCD background distributions for calorimeter D2 (left) and TAS D2 (right) at $\beta = 1$ for the lowest (above) and highest (below) p_{T} bins.

The enhancement in the highest p_{T} bins due to the use of TAS/tracks as input is of order 50 %, similarly to C2. Tracks are performing distinctly worse than TAS and calorimeter D2 in the first p_{T} bin but enhance with rising p_{T}, slowly passing the calorimeter and then match the TAS distributions for very high values of p_{T}.

Comparison of τ_{21} for W Jets

The signal distributions for τ_{21}, shown in Figure 6.8, computed with TAS again feature a mode at lower values due to steeper rising left side in comparison with the calorimeter τ_{21} variable. Furthermore, they fall faster for higher values, which increases separability since background tends to higher values. This effect again becomes more evident with rising p_{T}. On the other hand, the background for TAS rises slightly faster and drops earlier as well, therefore becoming more signal like. Nevertheless, one can anticipate a better separation for τ_{21} calculated with TAS due to the distinct observations found for the the signal distributions.

Similarly to the D2 variable, the ROCs in Figure 6.9 show that TAS τ_{21} outperforms the calorimeter counterpart in every studied p_{T} bin and for all values of signal efficiency.

The separation with tracks is somewhat worse compared to TAS but still better as with calorimeter clusters for lower p_{T}. For high boosts, the separation of τ_{21} with tracks matches the performance of τ_{21} with TAS.

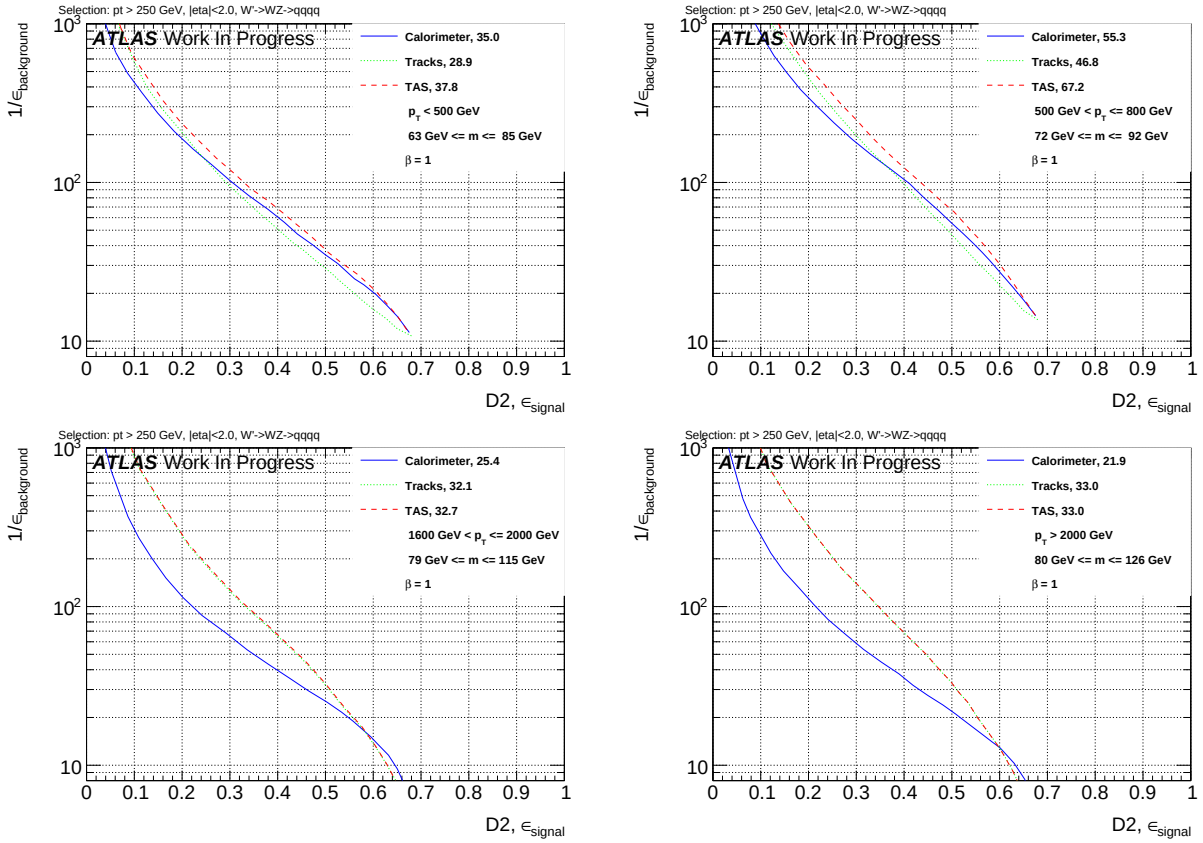

Figure 6.7.: ROCs showing QCD rejection against W boson efficiency for tracks, TAS and colorimeter D2 at $\beta = 1$, p_{T} ordering from upper left to lower right.

As expected due to the findings with C2 and D2, the improvements in background rejections with TAS and tracks over calorimeter clusters are for τ_{21} highest at large energies and can reach up to 50 %. Moreover, the background rejection with τ_{21} is noticeably lower as with C2 and D2. This behaviour is expected, since former studies, e.g. in Reference [35], already identified C2 and D2 to be the most discriminative variables for W boson tagging. Nevertheless, is is found that TAS τ_{21} performs comparable to C2 calculated with calorimeter clusters.

Tracks and TAS at very high p_{T}

The C2 variable was found to perform equally well with tracks and TAS as input, except for a small difference in the lowest p_{T} bin. This variable seems to be relative insensitive to the track assisting and tracks alone already perform well.

In contrast, D2 and τ_{21} calculated with tracks feature a visibly worse separation of QCD and signal events than the corresponding TAS variables. In these cases, the scale difference due to the missing neutral fraction seems to have a greater influence.

For very high p_{T} values however, it is often the case, that the large-R calorimeter jet features only one $R = 0.2$ subjet after trimming, due to the now very small distance between the decay products of the W boson. A single subjet results in the TAS procedure to be same as normal track assisting since all tracks are now scaled with the same factor. As

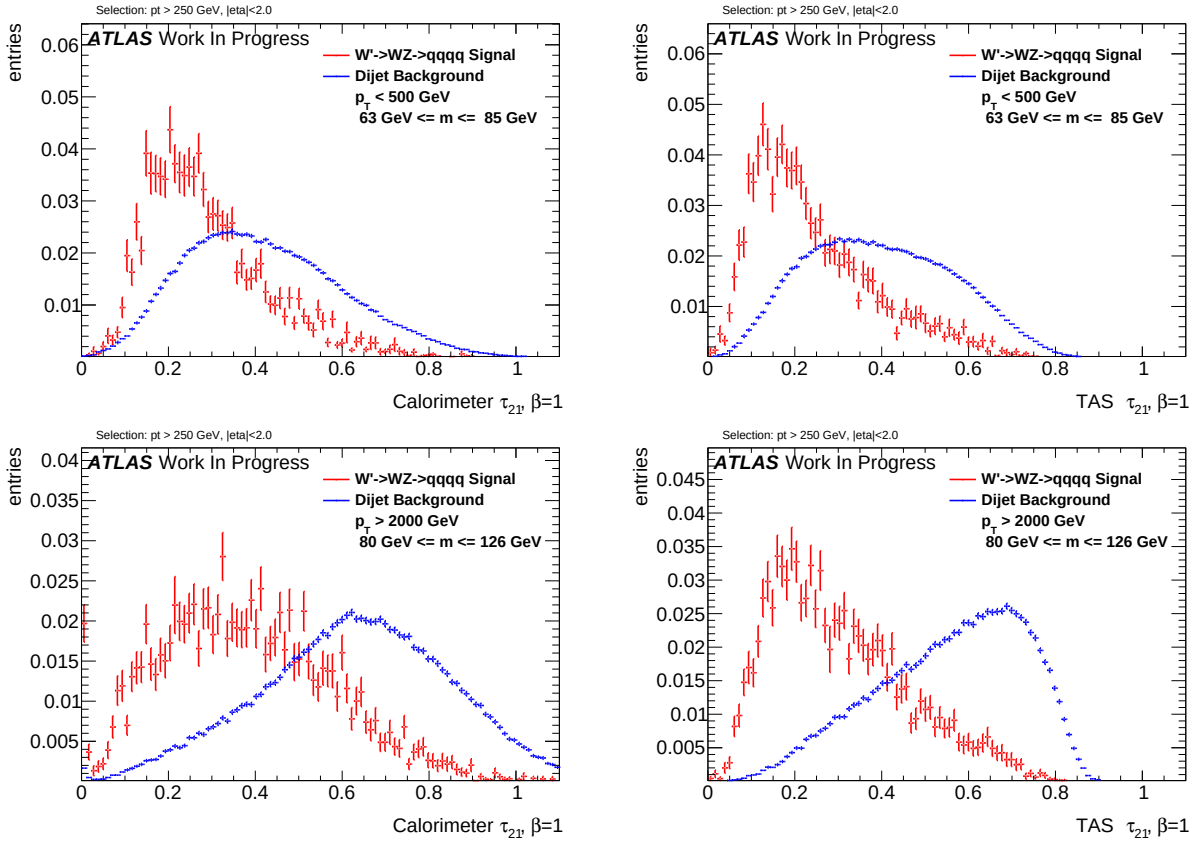

Figure 6.8.: W boson signal and QCD background distributions for calorimeter τ_{21} (left) and TAS τ_{21} (right) at $\beta = 1$ for the lowest (above) and highest (below) p_{T} bins.

stated in Chapter 4.5.2, normal track assisting has no impact on the studied substructure variables. Therefore, the studied substructure variables perform equally well if calculated with tracks or TAS for events with only one subjet and thereby the difference between both decreases for very large boosts.

Correlation with p_{T}

A way to study the p_{T} dependence of the discrimination variables is to count the events with a certain value of the considered variable against the p_{T} of the corresponding jet in a correlation plot.

Due to the rapidly falling p_{T} spectrum and hence very low weights for high p_{T}, it is not possible to show the whole range at once. Therefore, the plots are divided into the six different p_{T} regions and the colour code can (f) be applied for events in the same region.

For C2, see Figure 6.10, one can observe a strong trend to lower values for signal and background with calorimeter clusters as well as TAS. Furthermore, it is possible to verify the previous observation, that the TAS distributions concentrate at lower values compared to calorimeter counterparts.

In the cases of D2, Figure 6.11, and τ_{21}, Figure 6.12, there is a small upward trend of the calorimeter variables visible in the lower p_{T} regions which, with rising boost, slows down for D2 and τ_{21} and ends for τ_{21} in a broader distribution. This verifies the higher p_{T}

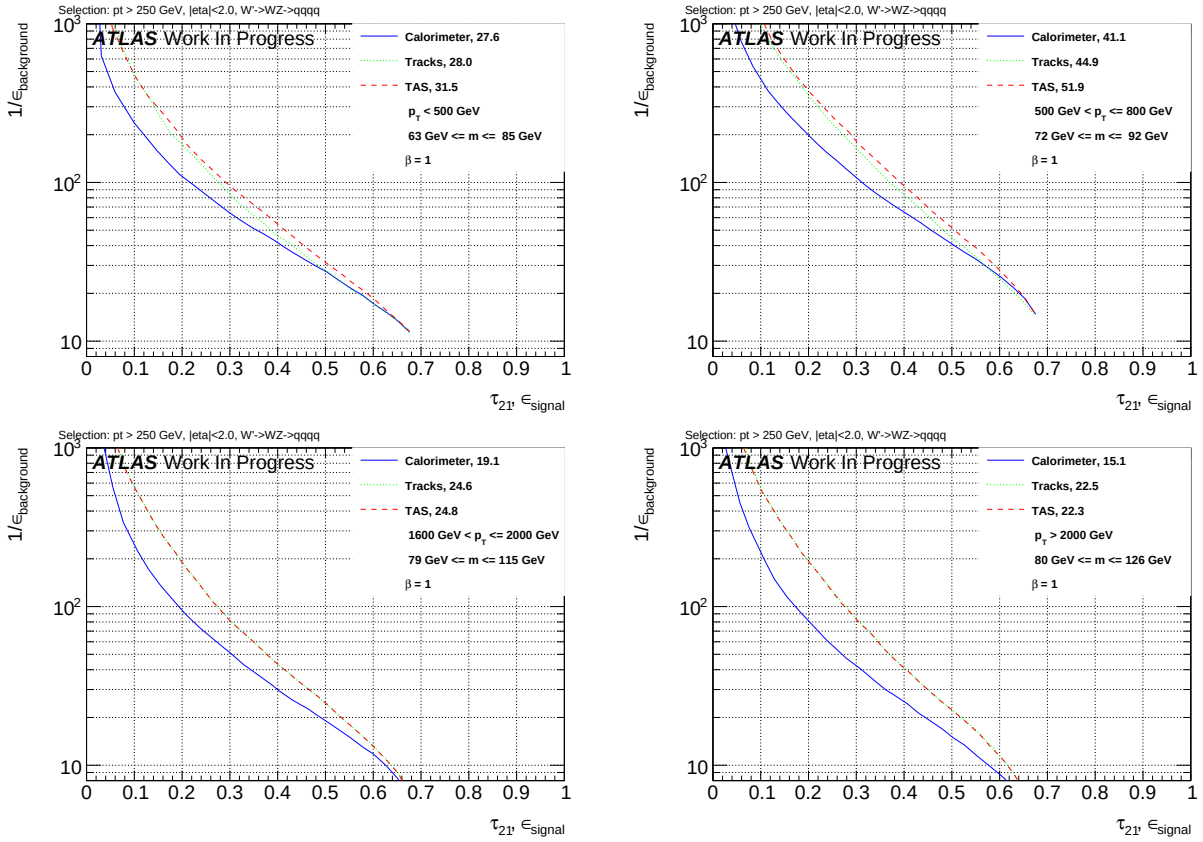

Figure 6.9.: ROCs showing QCD rejection against W boson efficiency for tracks, TAS and colorimeter τ_{21} at $\beta = 1$, p_{T} ordering from upper left to lower right.

dependence of the C2 variable in comparison to D2 and τ_{21}. The TAS counterparts feature an even more robust signal with the background moving to higher values, hence improving separation. The p_{T} dependence of variables calculated with tracks is very similar to the ones with TAS, therefore they are omitted.

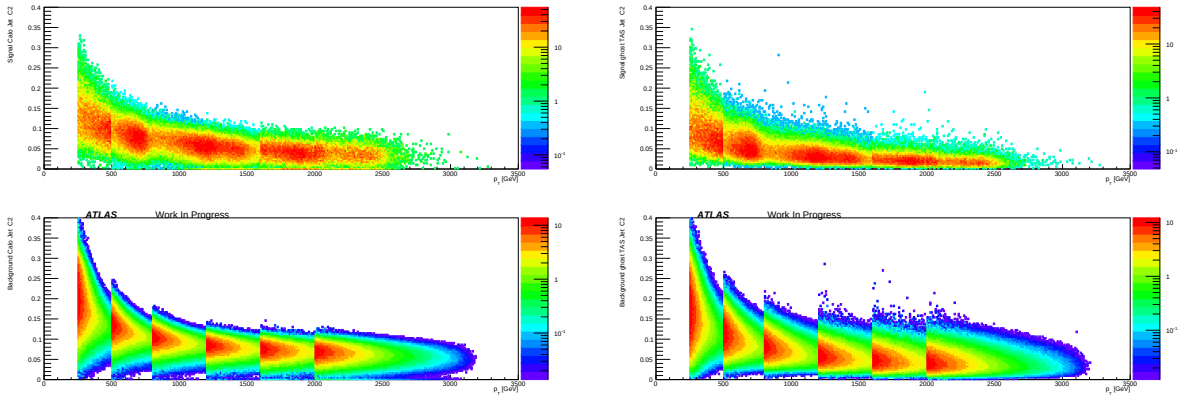

Figure 6.10.: Correlation between C2 at $\beta = 1$ and p_{T} applied on W boson signal (above) and QCD background (below) for calorimeter (left) and TAS (right).

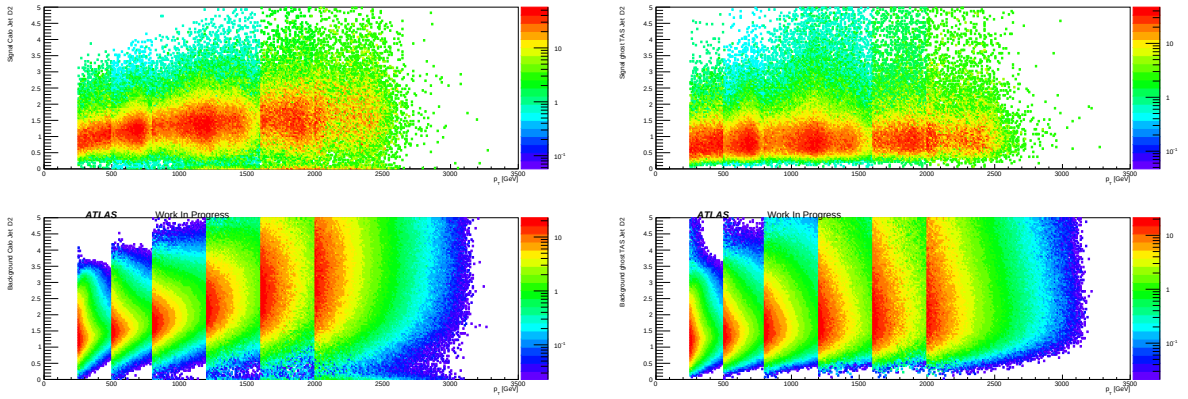

Figure 6.11.: Correlation between D2 at $\beta = 1$ and p_{T} applied on W boson signal (above) and QCD background (below) for calorimeter (left) and TAS (right).

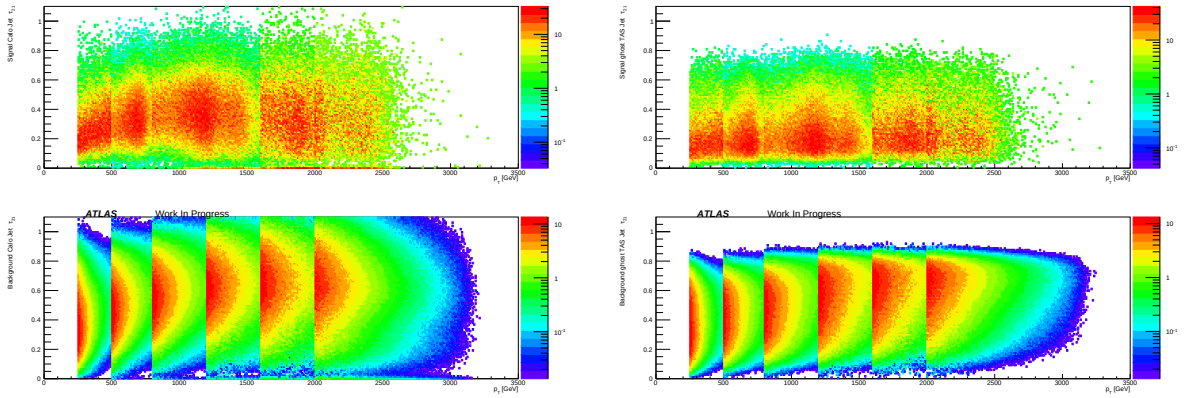

Figure 6.12.: Correlation between τ_{21} at $\beta = 1$ and p_{T} applied on W boson signal (above) and QCD background (below) for calorimeter (left) and TAS (right).

Summary of the W Boson study with $\beta = 1$

With an angular weight of $\beta = 1$, it was found for W boson tagging, that all studied substructure variables can increase in performance when calculated with tracks or TAS instead of calorimeter clusters. The improvements were shown to increase with the transverse momentum of the decaying W bosons. Furthermore, it was verified, that τ_{21} performs worse in QCD rejection as C2 or D2. Tracks perform equally well as TAS for the C2 variable. In case of τ_{21} and D2, TAS yields a higher QCD rejection for lower p_{T} regions. In the very high p_{T} limit, TAS and tracks become identical due to large-R calorimeter jets with only one subjet.

6.2.3. Performance of Higgs Boson Identification

The full set of distributions can again be found in Appendix C.1. Shown here is the second p_{T} bin of the signal and background distributions and the second to the last p_{T} region for the ROCs. The separation for p_{T} lower as 500 GeV is not suited for comparison of

tracks/TAS and calorimeter clusters due to the amount of events with the Higgs boson not being reconstructed into a single large-R jet. The Higgs boson features a larger mass as the W boson, 125 GeV compared to 84 GeV. Therefore, smaller or no possible improvements with TAS and tracks over calorimeter clusters are expected, as the separation of the $b\bar{b}$ pair still high compared to the W boson decay products.

Comparison of C2 for Higgs jets

The signal and background distributions in the Higgs scenario are shown in Figure 6.13. As already observed in the W boson case, the TAS C2 signal distributions rises sharper and tends to lower values as the calorimeter C2 signal. However, this effect is not as articulated for Higgs jets as for W jets. Furthermore, the C2 background distributions are also moving left when going from calorimeter clusters to TAS as C2 input. With these observations, no distinct differences in separation power between calorimeter C2 and TAS C2 are expected for the p_{T} region between 500 and 800 GeV. These conclusions

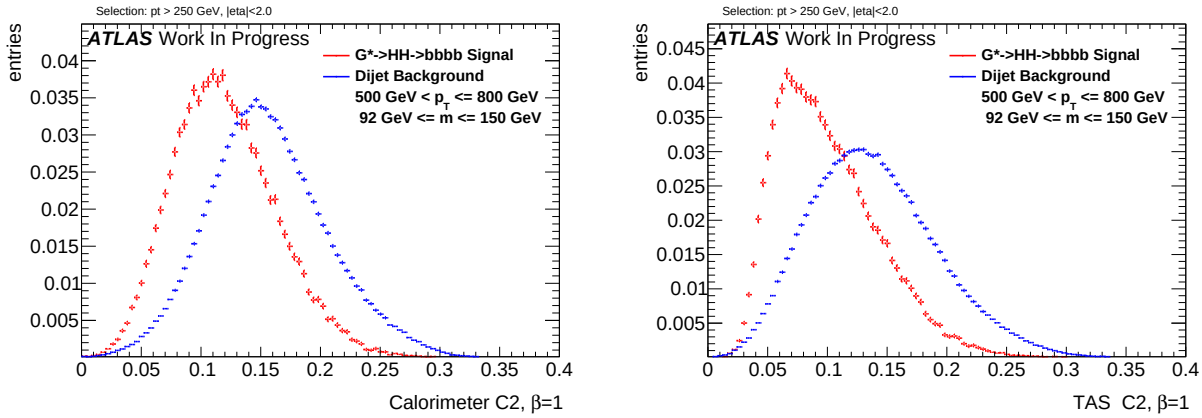

Figure 6.13.: Higgs boson signal and QCD background distributions for calorimeter D2 (left) and TAS C2 (right) at $\beta = 1$ for the second p_{T} bin.

are confirmed by the ROCs in Figure 6.14, which show an equal performance of TAS and the calorimeter C2 in the this region. Additionally, an advantage of the calorimeter clusters for C2 at high p_{T} is observed. Tracks again feature a separation very similar or equal to TAS, verifying the low sensitivity of the C2 variable to rescaling of the tracks p_{T} .

Comparison of D2 for Higgs jets

Signal and background distributions shown in Figure 6.15, as well as the corresponding ROCs, see Figure 6.16, show no possible improvements in discrimination of Higgs boson and QCD jets for D2 ($\beta = 1$) due to tracks or TAS compared to calorimeter clusters. As for C2, the shift to lower values of the signal distribution calculated with TAS instead of calorimeter clusters and the associated enhancements are small and therefore compensated by the also left moving TAS D2 background distribution.

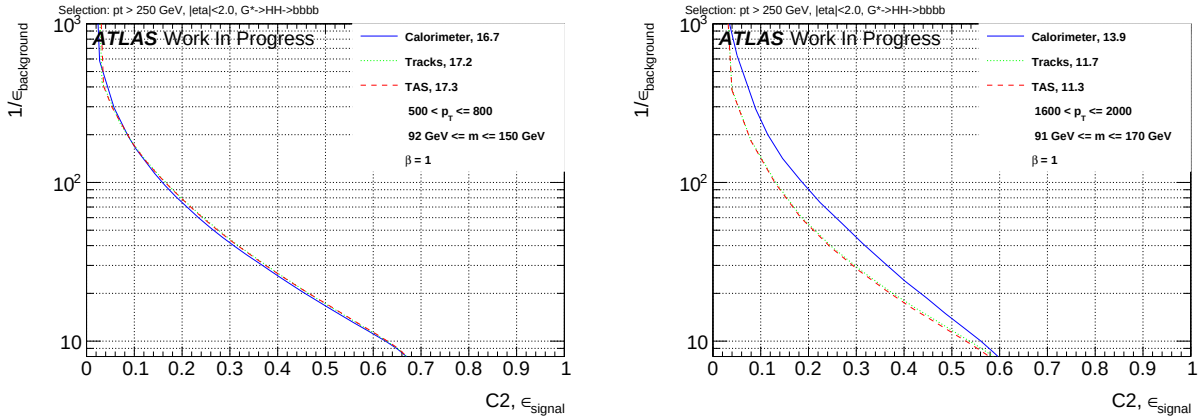

Figure 6.14.: ROCs showing QCD rejection against Higgs boson efficiency for tracks, TAS and calorimeter C2 at $\beta = 1$, second p_{T} bin (left) and highest bin (right).

With only minor differences, depending on the p_{T} region, TAS D2 and calorimeter D2 are observed to perform equally well over the studied p_{T} range. Found are no distinct differences between TAS D2 and track D2.

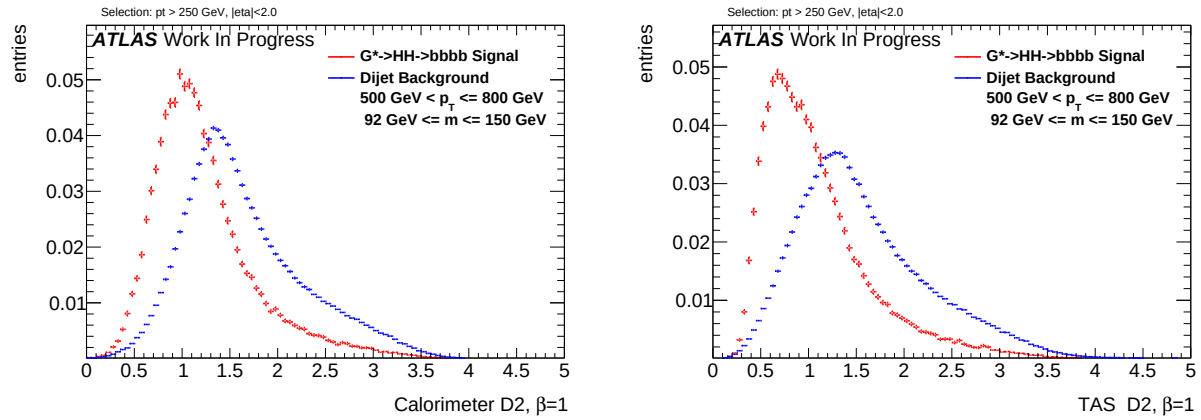

Figure 6.15.: Higgs boson signal and QCD background distributions for calorimeter D2 (left) and TAS C2 (right) at $\beta = 1$ in the second p_{T} bin.

Comparison of τ_{21} for Higgs jets

The distributions for $\tau_{21}(\beta = 1)$ in Figure 6.17 indicate no articulated difference between TAS and the calorimeter for separating Higgs boson and QCD jets, confirmed by the ROCs shown in Figure 6.18. Analogous to C2 and D2, the characteristics of the TAS signal and background distributions already seen in the W boson study are also present for Higgs boson jets but not as pronounced, leading to a comparable background rejection of TAS τ_{21} and calorimeter τ_{21}. Tracks and TAS perform equally well for τ_{21} for Higgs boson identification over the whole studied p_{T} range.

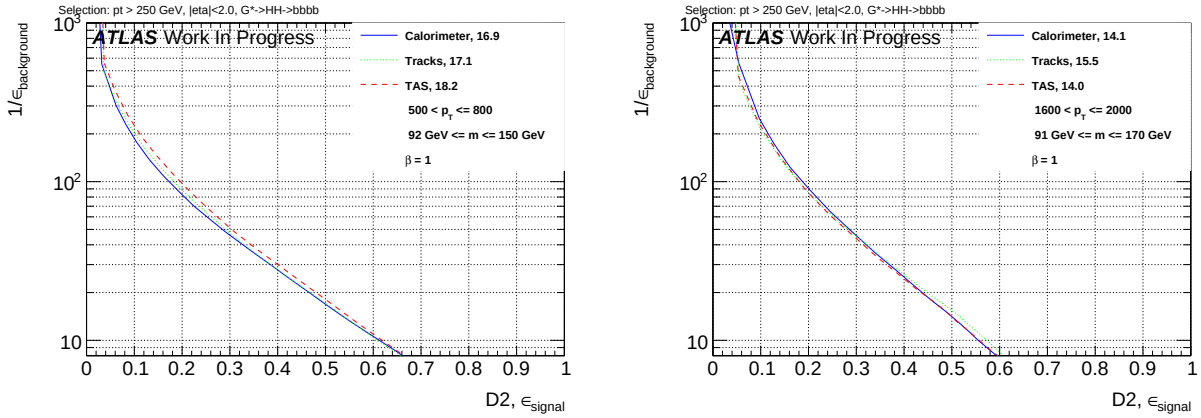

Figure 6.16.: ROCs showing QCD rejection against Higgs boson efficiency for tracks, TAS and calorimeter D2 at $\beta = 1$, second p_{T} bin (left) and highest bin (right).

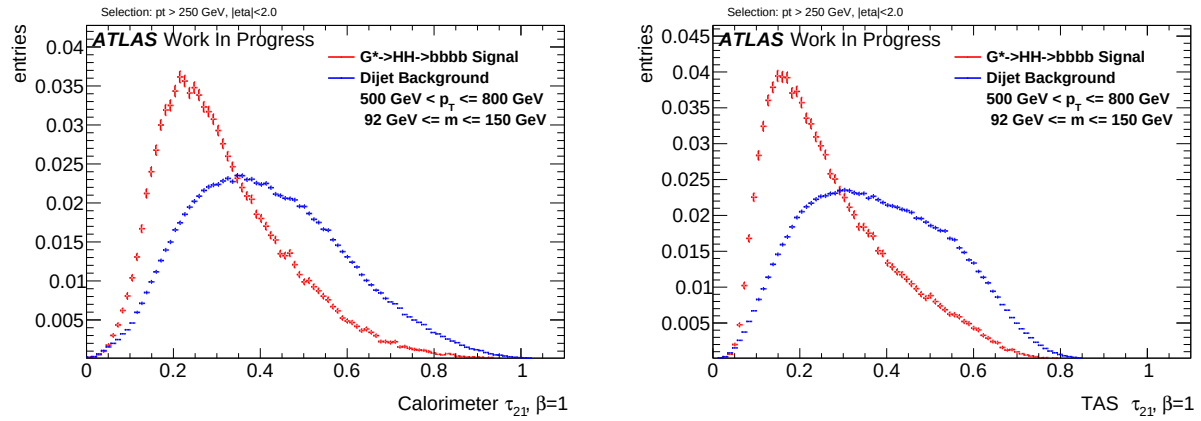

Figure 6.17.: Higgs boson signal and QCD background distributions for calorimeter D2 (left) and TAS τ_{21} (right) at $\beta = 1$ for the second p_{T} bin.

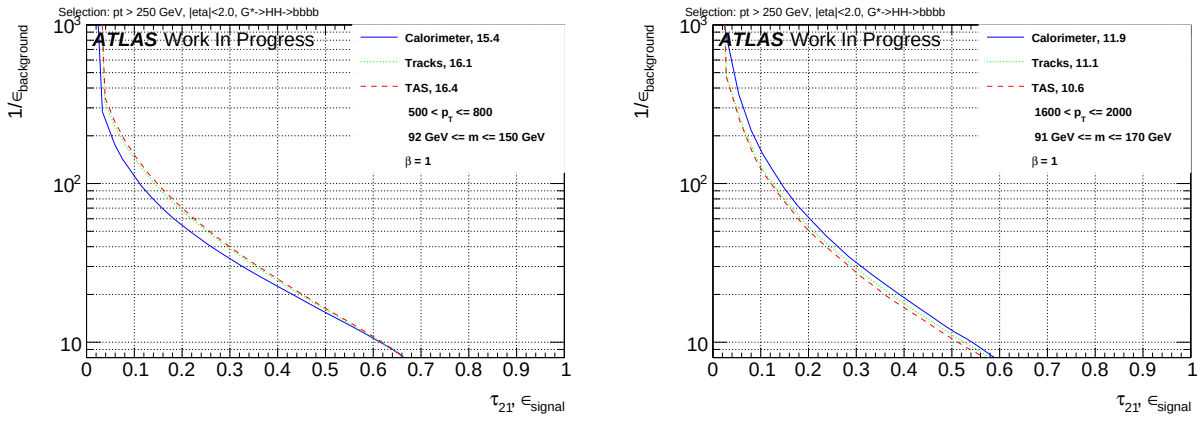

Figure 6.18.: ROCs showing QCD rejection against Higgs boson efficiency for tracks, TAS and calorimeter τ_{21} at $\beta = 1$, second p_{T} bin (left) and highest bin (right).

Summary of Higgs Boson Identification with $\beta = 1$

For Higgs boson tagging and an angular weight of $\beta = 1$, found were no distinct improvements with TAS or tracks compared to calorimeter clusters. The C2 variable performs better with calorimeter clusters, D2 yields an equal QCD discrimination with TAS and calorimeter clusters. The n-Subjettiness ratio τ_{21} benefits from TAS in some p_{T} regions, while the calorimeter pendant performs better in the other regions. Furthermore, tracks and TAS perform comparable over the whole studied p_{T} range.

6.2.4. Performance of Top Quark Identification

The top quark features a characteristic three body decay and a very high mass around 173 GeV.

Studied here is the n-Subjettiness ratio τ_{32} to distinguish the three prong like top quark jets and QCD background jets. The full set of distributions can be found in Appendix C.1.

Shown in Figure 6.19 are the τ_{32} signal and background distributions for the lowest (250 to 500 GeV) and the highest (> 2000 GeV) p_{T} region. For the lowest region, a large number of top quark events is not reconstructed into a single large-R jet. Therefore, observed are characteristic shoulders of the τ_{32} signal distributions. The left shoulder indicates the top quark events reconstructed into one large-R jet. The right shoulder at background like τ_{32} values represents events with large-R jets containing only part of the top quark decay.

Considering the higher p_{T} region, the TAS τ_{32} background distribution is observed to rise at higher values as the calorimeter τ_{32} background distribution (0.2 compared to 0), but also to fall very steep around 0.9. Furthermore, the narrower TAS signal distributions feature a sharper falling tail as the calorimeter τ_{32} signal distribution.

The ROCs in Figure 6.20 show the accompanying improvements in the separation power of τ_{32} possible with TAS. Tagging tops quark events with τ_{32} is found to greatly benefit from the excellent angular resolution of tracks. This is especially the case for high p_{T} where the limitation of the calorimeter cell size clearly diminishes the possible identification of three distinct substructures inside a large-R jet. The enhancements are not as articulated for the low p_{T} regions, nevertheless TAS τ_{32} performs here at least equally well as calorimeter τ_{32}. Furthermore, tracks are observed to perform slightly worse in comparison with TAS for the lower p_{T} regions, but match the TAS performance for very large boosts as expected.

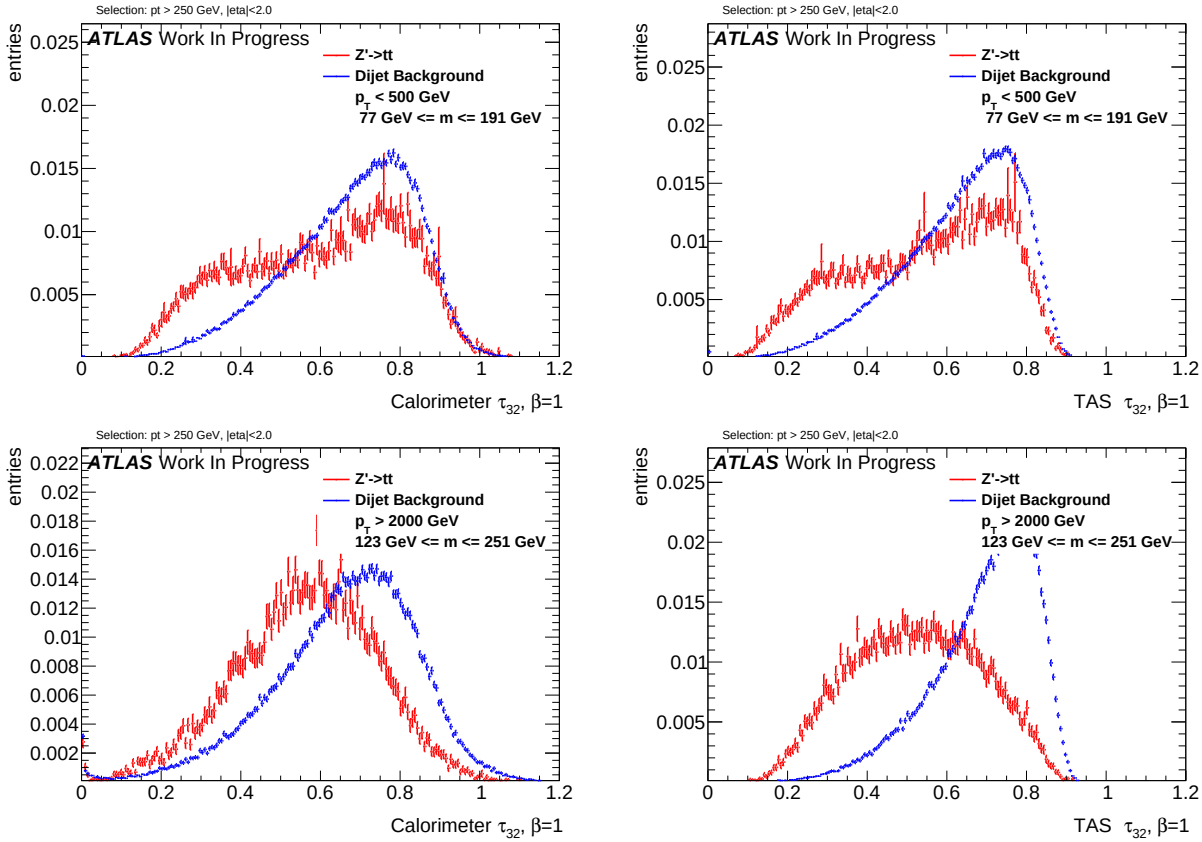

Figure 6.19.: Top quark signal and QCD background distributions for calorimeter τ_{32} (left) and TAS τ_{32} (right) at $\beta = 1$ for the lowest (above) and highest (below) p_{T} bin.

6.3. Variation of the Angular Weight β and Identification of Best Performing Variables

The observed background rejection of variables calculated with TAS and tracks is at least as high as of calorimeter cluster based variables due to the high angular resolution of tracks. Therefore, studied are the effects of a higher weighting of the angular part of the substructure variables. Considered as well is a lower weighting, expected is a decrease in separation power for TAS and tracks. Previous studies of calorimeter variables for W boson tagging, see e.g. Reference [35], found $\beta = 1$ to maximize the separation power of calorimeter variables.

A scan over the values $\beta = 0.5, 1, 1.7, 2, 3$ is performed in order to identify the best variables for the specific scenarios of tagging W boson, Higgs boson or top quark jets. The background rejections and their uncertainties, in brackets, achieved at the 50 % working point after mass cut and tagging are summarized in tables.

The previously shown signal and background distributions feature a statistical uncertainty due to the finite size of the Monte Carlo samples. To propagate these uncertainties to the achieved background rejections, used are pseudo-experiments. The distributions are varied within their statistical uncertainties via a Poissonian pdf 1000 times and at each step, the background rejection is calculated. An estimate of the statistical uncer-

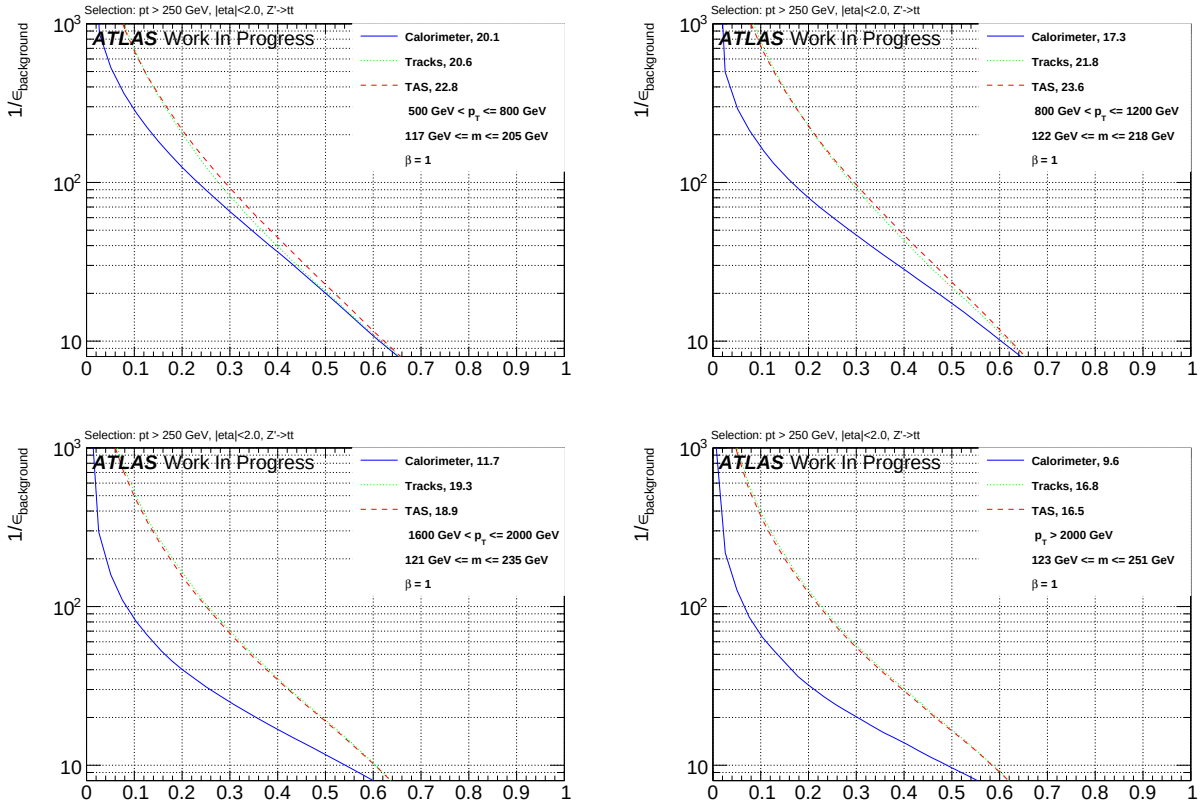

Figure 6.20.: ROCs showing QCD rejection against top quark efficiency for tracks, TAS and colorimeter τ_{32} at $\beta = 1$, p_{T} ordering from upper left to lower right.

tainty is given by the fluctuation of the pseudo-experiments, respectively the calculated minimum and maximum of background rejections. Stated as uncertainty is then the half of the difference between the minimal and maximal value,

$$\Delta\frac{1}{\epsilon_{bgr}} = \frac{(\max_{pseudo-exp}[\frac{1}{\epsilon_{bgr}}] - \min_{pseudo-exp}[\frac{1}{\epsilon_{bgr}}])}{2} \oplus 2\,\% \tag{6.1}$$

quadratically combined with the 2% uncertainty on the background rejection from the mass cut. ROCs were observed to cross rarely and if, only for very high signal efficiencies. Therefore, it is sufficient to compare the background rejections at 50 % signal efficiency to identify variables with the highest background rejection at the 50 % and 25 % working points.

Shown for these variables are the cut value to achieve 50% respectively 25% signal efficiency as function of p_{T} and a comparison with the best corresponding calorimeter variable. A final view on the best variables with ROCs can be found in Appendix B.

6.3.1. Tagging W jets

The background rejections for W boson jets achieved after performing the mass cut and tagging with the discrimination variables for different values of β are summarized in Table 6.2. Tracks and TAS are observed to perform worse with a lower angular weighting, in this case $\beta = 0.5$, except for a minor difference with τ_{21} at lower p_{T} values. Calorimeter cluster based variables separate slightly worse for $\beta = 0.5$ as well. However the performance loss is smaller, especially at higher p_{T}, where the background rejection with tracks and TAS at $\beta = 0.5$ is poor.

Moving to higher values of β, it is found, that tracks and TAS gain in separation power with a higher weighting of the angular part. Nevertheless, the separation is observed to degrade if the weighting of the angular part compared to the p_{T} part is too high (here $\beta = 3$).

Therefore, an angular weighting of around 2 maximizes the overall separation power of tracks and TAS. The enhancements for $\beta = 2$ in comparison with $\beta = 1$ are found at higher p_{T} values, minor losses are visible in the lowest p_{T} regions. This indicates, that a slightly lower value for β could retain the great performance of a higher angular weighting at high p_{T} while loosing as less separation power as possible for smaller boosts. A value of $\beta = 1.7$ is found to accomplish these expectations.

The results verify as well, that the calorimeter based variables perform best with an angular weighting of $\beta = 1$, except for minor improvements with larger β values for calorimeter τ_{21}. Overall, variables calculated with calorimeter clusters are not as sensitive to a variation of the angular weighting as variables with tracks and TAS. This proves the the high influence of the angular measure for the performance of track and TAS based discrimination variables.

The differences between tracks and TAS are as expected by observations in previous sections. While C2 is found to be quite insensitive to a rescaling of the tracks p_{T}, separation with τ_{21} visibly degrades without subjet assisting of tracks. D2 as well looses a significant part of its discrimination power with tracks compared to TAS at lower p_{T}. Starting around a p_{T} of 1200 GeV, tracks are performing similarly, and for even higher p_{T} equally well as TAS, which is expected due to the rising amount of events with only one subjet.

Furthermore, τ_{21} was found to perform worse in comparison to C2 and D2, independent of the angular weighting and tracks, TAS or calorimeter clusters as input.

With an angular weighting $\beta > 1$, track and TAS based C2 and D2 are found to yield a higher QCD rejection as the calorimeter cluster based counterparts.

These observations lead to the conclusion, that the best variables for distinguishing W jets from QCD background are C2 and D2 with TAS and an angular weighting of $\beta = 1.7$. TAS C2 ($\beta = 1.7$) achieves a lower background rejection in the lower both p_{T} regions as the corresponding D2. In the third region, the separation is equal and in the highest p_{T} bins, C2 again considerably exceeds the performance of D2. Moreover, D2 was shown to deliver cut values that are more robust over p_{T} but C2 could also be used with tracks instead of TAS, which saves the track assisting. Track assisting complicates the derivation of systematic uncertainties on the variables since uncertainties on the subjets,

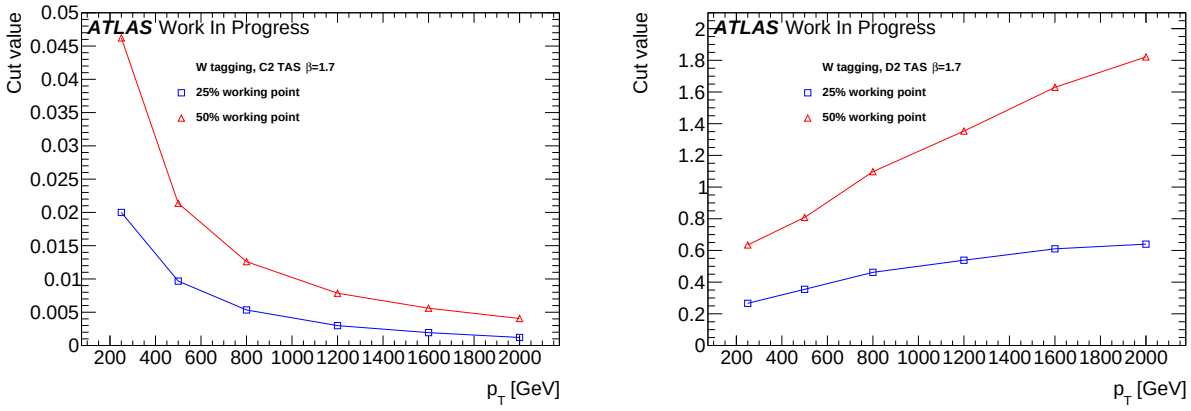

Figure 6.21.: Cut values for TAS (left) C2 and (right) D2 at $\beta = 1.7$ to achieve 50 % and 25 % W boson efficiency.

whose derivation is being worked at, need to be propagated while tracks feature already well-known uncertainties. Hence C2 with tracks could be used on a shorter timescale.

TAS C2 and D2 with $\beta = 1.7$

Shown in Figure 6.21 are the actual cut values against the p_{T} of the large-R calorimeter jet for TAS C2 and D2, both with $\beta = 1.7$, in order to achieve the working points at 50 % and 25 % signal efficiency.

D2 TAS ($\beta = 1.7$) is observed to be slightly more robust against a p_{T} variation compared to the corresponding C2 variable. This conclusion especially holds for the lower 25 % working point. As already seen in the p_{T} correlation plots, D2 moves to higher values while C2 moves into the opposite direction with rising energy.

The much steeper 50 % working point curves can be explained in the view of the actual signal distributions, see Figure 6.22. For both variables, the signal distributions steeply rise at low values, which does not chance considerably with rising p_{T}. The right sides of the distributions in contrast, become increasingly narrow for C2 TAS ($\beta = 1.7$) and broader for D2 TAS ($\beta = 1.7$). Therefore, the curves representing the 50 % working point feature a higher slope as the 25 % working point is already defined by the sharp rise. This is particularly observed for D2 TAS ($\beta = 1.7$) while for C2 TAS ($\beta = 1.7$) also the 25 % working point is distinctly p_{T} dependent.

Furthermore, Table 6.3 lists the background rejections after applying the mass cut plus TAS C2 and D2 at $\beta = 1.7$ and for the currently used calorimeter D2 ($\beta = 1$) with the corresponding improvements due to the use of TAS instead of calorimeter clusters. The values verify, that for lower p_{T}, D2 TAS is the best choice. For very high boosts, C2 TAS is shown to performs superior, especially for the 25 % working point, at which the background rejection with C2 TAS is around 3.5 times as large as the QCD rejection with calorimeter D2 variable.

These enormous improvements at the 25 % working point compared to higher signal efficiencies are a result of the signal distributions which, as already seen, rise for TAS and tracks much earlier than for calorimeter clusters. The tail of the TAS signal dis-

Calorimeter	C2					D2					τ_{21}				
p_T [GeV]	$\beta = 0.5$	1	1.7	2	3	$\beta = 0.5$	1	1.7	2	3	$\beta = 0.5$	1	1.7	2	3
250 - 500	29.7(1.5)	31.7(1.9)	31.4(1.6)	30.7(1.9)	28.5(1.4)	27.2(2.0)	35.0(2.0)	33.0(1.8)	31.3(1.7)	25.7(1.2)	33.1(1.8)	27.6(1.3)	26.2(1.4)	25.1(1.2)	22.4(0.8)
500 - 800	44.2(1.8)	50.1(2.0)	49.6(1.9)	48.6(1.8)	42.6(1.9)	40.3(2.2)	55.3(2.6)	56.3(2.4)	52.5(2.1)	39.3(1.3)	49.4(2.0)	41.1(1.4)	43.3(1.7)	41.3(1.6)	36.1(1.2)
800 - 1200	32.0(1.5)	37.5(1.7)	35.4(1.5)	33.4(1.5)	26.8(0.9)	34.0(2.1)	41.1(2.0)	38.5(1.6)	34.9(1.3)	25.4(0.7)	30.5(1.2)	30.9(1.2)	33.8(1.4)	32.5(1.3)	28.1(0.9)
1200 - 1600	30.1(1.3)	34.4(1.8)	29.4(1.3)	26.8(1.0)	20.7(0.8)	34.1(1.8)	38.1(1.9)	31.4(1.4)	27.6(1.2)	19.3(0.5)	23.1(0.9)	27.3(1.)	31.1(1.2)	29.9(1.3)	24.8(0.9)
1600 - 2000	20.9(1.3)	22.4(1.5)	18.2(1.2)	16.5(0.9)	12.9(0.6)	26.4(1.7)	25.4(1.3)	19.3(1.1)	16.9(0.9)	11.9(0.5)	16.4(1.0)	19.1(1.1)	21.1(1.1)	19.9(1.0)	16.0(0.9)
> 2000	16.9(1.4)	18.7(1.4)	14.1(0.9)	12.6(0.8)	9.9(0.7)	23.3(1.9)	21.9(1.7)	15.7(1.1)	13.5(0.9)	9.2(0.4)	12.3(1.1)	15.5(1.1)	17.2(1.2)	15.7(1.1)	11.9(0.8)

TAS	C2					D2					τ_{21}				
p_T [GeV]	$\beta = 0.5$	1	1.7	2	3	$\beta = 0.5$	1	1.7	2	3	$\beta = 0.5$	1	1.7	2	3
250 - 500	29.4(1.9)	30.1(1.9)	28.9(1.5)	28.5(1.3)	27.7(1.3)	28.6(2.0)	37.7(2.1)	35.4(2.3)	33.4(2.0)	29.4(1.2)	36.2(2.2)	31.5(1.6)	26.8(1.3)	25.4(1.4)	24.0(1.0)
500 - 800	48.2(2.0)	55.5(2.7)	58.6(2.6)	59.1(2.7)	56.8(2.0)	42.8(2.3)	67.2(3.1)	67.6(3.2)	63.7(3.0)	52.6(2.3)	55.7(2.6)	51.9(2.1)	45.5(2.0)	44.0(1.9)	41.3(1.5)
800 - 1200	31.0(1.2)	44.6(1.9)	54.6(2.8)	55.2(2.8)	53.0(3.2)	26.1(1.3)	47.6(2.3)	54.9(2.4)	52.6(2.8)	43.1(1.5)	36.4(1.8)	37.3(1.7)	36.2(1.8)	36.2(1.6)	35.5(1.6)
1200 - 1600	20.9(0.7)	39.1(1.9)	53.8(2.6)	55.1(3.0)	50.1(1.6)	22.7(1.4)	42.1(2.4)	50.8(1.8)	49.6(2.3)	41.1(1.2)	27.9(1.3)	31.4(1.5)	33.4(1.6)	34.0(2.0)	33.0(1.8)
1600 - 2000	16.7(0.7)	36.9(2.9)	50.9(4.3)	50.3(4.4)	42.2(2.4)	18.7(1.7)	32.7(3.3)	37.8(2.0)	36.1(2.4)	28.7(1.2)	20.5(1.2)	24.8(1.6)	26.1(2.0)	26.5(2.0)	25.4(2.0)
> 2000	11.6(0.6)	31.2(3.2)	46.1(4.7)	45.5(5.2)	35.5(3.8)	17.8(2.0)	33.0(4.0)	36.3(2.0)	34.0(2.5)	27.4(1.3)	16.4(1.3)	22.3(2.0)	24.2(2.2)	24.4(2.5)	21.8(2.4)

Tracks	C2					D2					τ_{21}				
p_T [GeV]	$\beta = 0.5$	1	1.7	2	3	$\beta = 0.5$	1	1.7	2	3	$\beta = 0.5$	1	1.7	2	3
250 - 500	27.1(1.2)	28.1(1.5)	28.7(1.9)	28.7(1.9)	28.2(1.7)	21.6(1.2)	28.9(2.0)	29.5(1.8)	29.1(1.6)	28.1(1.3)	28.7(1.8)	28.0(1.7)	25.6(1.3)	25.1(1.3)	24.2(0.9)
500 - 800	46.5(1.9)	52.9(2.4)	57.7(2.6)	58.1(2.7)	55.8(2.5)	30.1(1.8)	46.8(2.4)	53.4(2.2)	52.1(2.3)	46.6(1.7)	46.1(2.3)	44.9(1.8)	41.7(2.1)	40.6(1.8)	39.2(1.5)
800 - 1200	30.3(1.1)	44.5(2.2)	54.8(2.8)	56.4(3.0)	53.7(3.6)	24.5(1.5)	42.3(2.3)	48.6(2.5)	47.5(1.2)	42.4(1.2)	34.5(1.6)	36.2(1.8)	36.0(1.8)	36.2(1.8)	35.7(1.5)
1200 - 1600	20.7(0.6)	39.0(1.9)	54.2(2.7)	55.5(3.3)	50.9(1.7)	22.7(1.3)	41.0(2.2)	50.0(1.6)	47.6(2.2)	41.4(1.2)	27.7(1.2)	31.3(1.4)	33.3(1.6)	33.9(1.7)	33.2(1.8)
1600 - 2000	16.6(0.7)	36.7(2.3)	51.7(5.2)	51.6(4.0)	43.1(2.3)	18.5(1.7)	32.1(3.0)	37.0(1.9)	35.9(2.3)	29.3(1.2)	20.5(1.3)	24.6(1.7)	26.2(1.8)	26.7(2.0)	25.9(2.2)
> 2000	11.6(0.5)	31.5(3.0)	46.8(5.7)	46.0(4.2)	36.1(4.3)	17.8(2.2)	33.0(3.3)	35.9(2.1)	34.2(2.6)	28.1(1.0)	16.4(1.4)	22.5(1.8)	24.5(2.4)	24.7(2.6)	22.2(2.6)

Table 6.2.: Listing of the QCD background rejection for W boson signal achieved with C2, D2 and τ_{21} together with different angular weightings β and for calorimeter cluster, tracks and TAS.

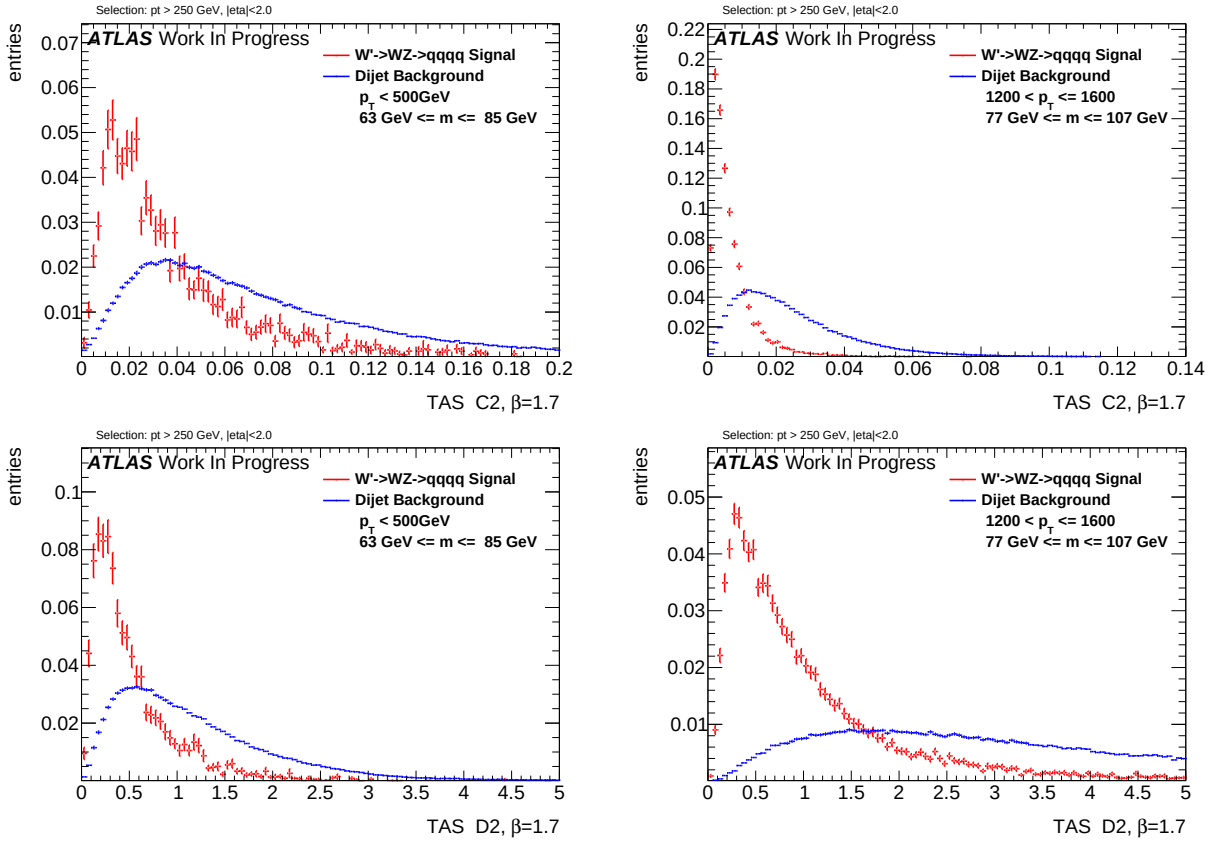

Figure 6.22.: W boson signal and QCD background distributions for C2 TAS at $\beta = 1.7$ (above) and D2 TAS at $\beta = 1.7$ (below) for the first (left) and fourth (right) p_{T} bin.

tributions to higher, background like values in contrast, is more comparable to the tail of the calorimeter variables. This results in enhancements over the calorimeter variables especially at lower signal efficiencies.

The improvements due to TAS are found to lie around 50 % for D2 and up to a 100 % for C2 in the upper p_{T} regions and 50 % W boson efficiency. For the lower working point, they can reach 200 % for D2 and around 250 % for C2, again in the higher p_{T} regions where the W boson is highly boosted.

50% ϵ_{signal}	**W tagging**		
p_{T} [GeV]	Calo D2, $\beta = 1$	TAS D2, $\beta = 1.7$	TAS C2, $\beta = 1.7$
250 - 500	35.0 ± 2.0	35.4 ± 2.3 (+1 ± 9%)	28.9 ± 1.5 (-17 ± 6%)
500 - 800	55.3 ± 2.6	67.6 ± 3.2 (+22 ± 8%)	58.6 ± 2.6 (+6 ± 7%)
800 - 1200	41.1 ± 2.0	54.9 ± 2.4 (+34 ± 9%)	54.6 ± 2.8 (+33 ± 9%)
1200 - 1600	38.1 ± 1.9	50.8 ± 1.8 (+33 ± 8%)	53.8 ± 2.7 (+41 ± 10%)
1600 - 2000	25.4 ± 1.3	37.8 ± 2.0 (+49 ± 11%)	50.9 ± 4.3 (+100 ± 20%)
$>$ 2000	21.9 ± 1.7	36.3 ± 2.0 (+66 ± 16%)	46.1 ± 4.7 (+111 ± 27%)

25% ϵ_{signal}	**W tagging**		
p_{T} [GeV]	Calo D2, $\beta = 1$	TAS D2, $\beta = 1.7$	TAS C2, $\beta = 1.7$
250 - 500	139.6 ± 9.8	146.0 ± 12.4 (+5 ± 12%)	108.2 ± 7.5 (-22 ± 8%)
500 - 800	243.7 ± 13.2	360.1 ± 21.1 (+48 ± 12%)	298.4 ± 15.9 (+22 ± 9%)
800 -1200	181.0 ± 8.8	308.5 ± 19.3 (+70 ± 14%)	313.2 ± 24.4 (+78 ± 16%)
1200 - 1600	156.9 ± 8.3	295.4 ± 17.8 (+88 ± 15%)	354.6 ± 25.6 (+126 ± 20%)
1600 - 2000	84.6 ± 5.7	219.6 ± 10.9 (+160 ± 22%)	320.5 ± 31.4 (+279 ± 45%)
$>$ 2000	78.9 ± 7.6	233.5 ± 14.7 (+196 ± 34%)	288.4 ± 33.3 (+266 ± 55%)

Table 6.3.: Listing of the background rejections after the jet mass cut and tagging at 50% and 25% W signal efficiency for the identified best variables TAS D2/C2 ($\beta = 1.7$) together with the improvements over the standard choice calorimeter D2 at $\beta = 1$.

6.3.2. Tagging Higgs Jets

The study of $\beta = 1$ in the Higgs case, see section 6.2.3, indicated no improvements in separation of QCD background due to tracks and TAS. This is a result of the higher Higgs mass which in turn, increases the separation of the decay products whereby angular resolution effects are not relevant. Note, that because of limited statistics, conclusions can only be drawn up to the fifth p_{T} region between 1600 and 2000 GeV.

Presented in Figure 6.4 are the achieved QCD rejections for Higgs tagging with C2, D2 and τ_{21} together with different angular weightings β. As already seen for the W boson, the performance of tracks and TAS also diminishes considerably for Higgs tagging with an angular weighting of $\beta = 0.5$.

No improvement of τ_{21} is observed with tracks or TAS, calorimeter clusters perform equally well for lower p_{T} and slightly better at very high boosts. However, the QCD rejection achieved with the n-Subjettiness ratio τ_{21} is again found to be exceeded by C2 and D2.

The discrimination with calorimeter clusters profits from a higher angular weighting, although the gain is not as significant as for tracks and TAS, which consistently shows the lower sensitivity to a variation of the angular weight. The small gain in performance up to $\beta = 1.7$ for the calorimeter based C2 and D2 variables are connected to the higher separation of the $b\bar{b}$ in comparison with the decay products of the W boson.

For boosted Higgs tagging, D2 is observed to perform considerably better as C2 over the whole studied p_{T} range. Values of 1.7 and 2 for β are found to yield the highest

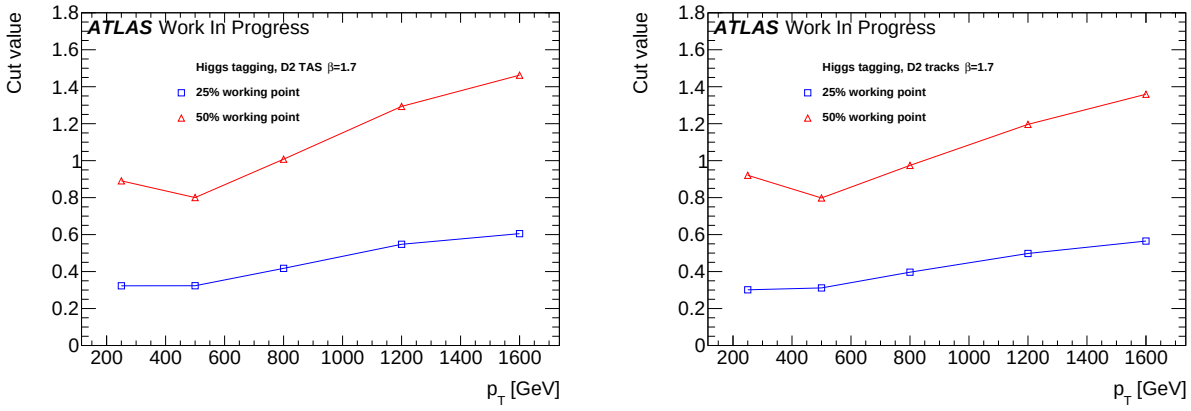

Figure 6.23.: Cut values for TAS (left) and track D2 (right) at $\beta = 1.7$ to achieve 50% and 25% Higgs boson efficiency.

background rejection for the track and TAS based D2. For both values, tracks and TAS D2 are performing equally well as the calorimeter D2 at low p_{T} and superior with rising boost of the Higgs.

The differences between $\beta = 1.7$ and $\beta = 2$ are inconclusive with minor advantages at high and slight inferiorities at low p_{T} for $\beta = 2$.

Comparing tracks and TAS for these cases, tracks perform slightly worse for lower energies but similarly better in the two highest studied p_{T} regions. Therefore, it is not of highest importance to choose between tracks and TAS. Chosen for further examination are TAS and track D2 with $\beta = 1.7$.

TAS and track D2 with $\beta = 1.7$

Shown in Table 6.23 are the cut values against the p_{T} of the large-R calorimeter jet, in order to achieve the working points at 50 % and 25 % signal efficiency, for TAS and track D2 with $\beta = 1.7$.

Visible is the slight p_{T} dependence of the rising cut value. Moreover, the cut values for the first p_{T} bin are higher as for the second p_{T} bin, in contrast to the observed upward trend of D2. This is the result of the low boost in the first p_{T} region which results in a left shoulder of the mass distributions representing large-R jets that contain only part of the Higgs boson decay. These jets feature a high, background like D2 value due to their one-prong structure. Therefore, the cut value to achieve a certain percentage of the signal distribution is higher for a small Higgs boson boost.

Observed are cut values for tracks that are marginal lower as for TAS, which is expected, since TAS are the same tracks with rescaled, higher p_{T} and the D2 variable features a rising tendency for higher p_{T} values.

Furthermore, Table 6.5 lists the background rejections for TAS and track D2 with $\beta = 1.7$, and for the best calorimeter variable, which is D2 ($\beta = 1$), with the corresponding improvements due to the use of TAS respectively tracks instead of calorimeter clusters.

The improvements for the Higgs boson case are lower as for W boson tagging. Nevertheless, considerably improvements are found at very high p_{T}, where the angle between

Calorimeter	C2					D2					τ_{21}				
$p_T\,[GeV]$	$\beta = 0.5$	1	1.7	2	3	$\beta = 0.5$	1	1.7	2	3	$\beta = 0.5$	1	1.7	2	3
250 - 500	4.6(0.1)	5.0(0.1)	5.2(0.1)	5.3(0.1)	5.5(0.1)	5.7(0.1)	7.3(0.2)	8.4(0.2)	8.4(0.2)	8.4(0.2)	7.6(0.2)	8.0(0.2)	7.9(0.2)	7.8(0.2)	7.5(0.2)
500 - 800	15.7(0.3)	16.7(0.4)	17.0(0.4)	16.9(0.4)	16.2(0.4)	13.6(0.3)	16.9(0.4)	17.7(0.4)	17.2(0.4)	15.2(0.3)	16.7(0.4)	15.4(0.3)	15.2(0.3)	14.8(0.3)	14.0(0.3)
800 - 1200	22.1(0.5)	23.8(0.5)	25.0(0.6)	25.0(0.6)	23.4(0.5)	18.4(0.4)	23.7(0.6)	26.3(0.6)	25.6(0.6)	22.3(0.5)	22.8(0.5)	21.9(0.5)	22.6(0.5)	22.1(0.5)	20.9(0.5)
1200 - 1600	24.0(0.6)	26.0(0.8)	26.4(0.8)	25.9(0.7)	23.0(0.6)	19.3(0.6)	24.9(0.7)	27.0(0.8)	26.1(0.7)	21.9(0.5)	21.3(0.5)	22.6(0.6)	24.0(0.6)	23.7(0.6)	22.2(0.5)
1600 - 2000	12.1(0.7)	13.9(0.8)	14.3(0.7)	14.0(0.7)	12.3(0.6)	11.1(0.7)	14.1(0.9)	14.9(0.8)	14.2(0.6)	11.8(0.5)	10.3(0.5)	11.9(0.5)	13.1(0.6)	13.1(0.7)	12.3(0.7)

TAS	C2					D2					τ_{21}				
$p_T\,[GeV]$	$\beta = 0.5$	1	1.7	2	3	$\beta = 0.5$	1	1.7	2	3	$\beta = 0.5$	1	1.7	2	3
250 - 500	4.8(0.1)	5.2(0.1)	5.5(0.1)	5.6(0.1)	5.8(0.1)	5.9(0.1)	7.6(0.2)	8.5(0.2)	8.6(0.2)	8.5(0.2)	7.6(0.2)	8.0(0.2)	7.7(0.2)	7.6(0.2)	7.4(0.2)
500 - 800	16.1(0.4)	17.3(0.4)	17.7(0.4)	17.7(0.4)	17.6(0.4)	14.0(0.3)	18.2(0.4)	18.7(0.4)	18.3(0.4)	16.9(0.4)	16.2(0.4)	16.4(0.4)	15.4(0.4)	15.1(0.3)	14.6(0.3)
800 - 1200	20.6(0.5)	23.5(0.5)	26.2(0.6)	26.9(0.7)	27.7(0.6)	18.8(0.4)	25.6(0.6)	28.5(0.7)	28.4(0.7)	26.8(0.6)	21.7(0.5)	22.4(0.5)	22.1(0.5)	22.0(0.5)	21.8(0.5)
1200 - 1600	18.6(0.4)	22.6(0.6)	27.4(0.7)	28.7(0.8)	30.0(0.7)	17.9(0.4)	24.3(0.7)	28.9(0.7)	29.3(0.6)	28.1(0.7)	19.3(0.5)	20.0(0.5)	20.7(0.5)	21.0(0.6)	21.9(0.5)
1600 - 2000	8.0(0.3)	11.3(0.5)	15.4(0.9)	16.5(1.0)	17.8(0.7)	10.0(0.5)	14.0(0.8)	17.7(0.8)	18.1(.9)	17.9(0.6)	9.8(0.4)	10.6(0.5)	11.4(0.6)	11.8(0.6)	12.6(0.6)

Tracks	C2					D2					τ_{21}				
$p_T\,[GeV]$	$\beta = 0.5$	1	1.7	2	3	$\beta = 0.5$	1	1.7	2	3	$\beta = 0.5$	1	1.7	2	3
250 - 500	4.91(0.1)	5.2(0.1)	5.5(0.1)	5.6(0.1)	5.9(0.1)	5.8(0.1)	7.4(0.2)	8.3(0.2)	8.3(0.2)	8.5(0.2)	7.4(0.2)	7.9(0.2)	7.8(0.2)	7.7(0.2)	7.6(0.2)
500 - 800	15.6(0.3)	17.2(0.4)	17.8(0.4)	17.9(0.4)	17.7(0.4)	13.5(0.3)	17.1(0.4)	17.9(0.4)	17.7(0.4)	16.8(0.4)	15.7(0.3)	16.1(0.4)	15.5(0.3)	15.3(0.3)	14.8(0.1)
800 - 1200	20.1(0.5)	24.0(0.5)	26.9(0.6)	27.7(0.7)	28.4(0.6)	18.8(0.4)	25.3(0.6)	28.0(0.7)	28.0(0.7)	26.9(0.6)	22.0(0.5)	22.7(0.5)	22.5(0.5)	22.4(0.5)	22.4(0.3)
1200 - 1600	18.5(0.5)	23.8(0.6)	28.8(0.8)	30.0(0.8)	31.1(0.7)	19.4(0.5)	26.3(0.7)	30.0(0.8)	30.3(0.8)	29.2(0.7)	20.8(0.5)	21.4(0.5)	21.9(0.6)	22.3(0.6)	23.0(0.5)
1600 - 2000	8.0(0.3)	11.7(0.5)	16.1(0.9)	17.1(0.9)	18.3(0.9)	11.0(0.7)	15.5(0.7)	18.5(0.8)	18.7(0.8)	18.4(0.6)	10.4(0.5)	11.1(0.5)	12.0(0.6)	12.4(0.7)	13.2(0.6)

Table 6.4.: Listing of the QCD background rejection for Higgs signal achieved with C2, D2 and τ_{21} together with different angular weightings β and for calorimeter cluster, tracks and TAS.

50% ϵ_{signal}	**Higgs tagging**		
p_{T} [GeV]	Calo D2, $\beta = 1$	TAS D2, $\beta = 1.7$	track D2, $\beta = 1.7$
250 - 500	8.4 ± 0.2	8.5 ± 0.2 (+1 ± 4%)	8.3 ± 0.2 (-1 ± 3%)
500 - 800	17.7 ± 0.4	18.7 ± 0.4 (+6 ± 3%)	17.9 ± 0.4 (+1 ± 3%)
800 - 1200	26.3 ± 0.6	28.5 ± 0.7 (+8 ± 4%)	28.0 ± 0.7 (+6 ± 4%)
1200 - 1600	27.0 ± 0.8	28.9 ± 0.7 (+7 ± 4%)	30.0 ± 0.8 (+11 ± 4%)
1600 - 2000	14.9 ± 0.8	17.7 ± 0.8 (+19 ± 8%)	18.5 ± 0.8 (+24 ± 9%)

25% ϵ_{signal}	**Higgs tagging**		
p_{T} [GeV]	Calo D2, $\beta = 1$	TAS D2, $\beta = 1.7$	track D2, $\beta = 1.7$
250 - 500	25.1 ± 0.6	28.9 ± 0.7 (+15 ± 4%)	30.5 ± 0.8 (+22 ± 4%)
500 - 800	54.1 ± 1.4	69.6 ± 1.9 (+29 ± 5%)	64.9 ± 1.8 (+20 ± 5%)
800 -1200	90.8 ± 2.5	121.3 ± 3.4 (+34 ± 5%)	117.9 ± 3.2 (+30 ± 5%)
1200 - 1600	97.6 ± 3.1	117.7 ± 3.8 (+21 ± 5%)	122.4 ± 4.2 (+25 ± 6%)
1600 - 2000	54.6 ± 3.5	74.0 ± 5.7 (+36 ± 14%)	75.0 ± 5.1 (+37 ± 13%)

Table 6.5.: Listing of the background rejections after the jet mass cut and tagging at 50% and 25% Higgs signal efficiency for the identified best variables TAS/track D2($\beta = 1.7$) together with the improvements over the best variable with calorimeter clusters which is D2 at $\beta = 1.7$.

the $b\bar{b}$ pair is small despite the high Higgs boson mass and the effect of the calorimeter cell size becomes visible.

Moreover, significant improvements are observed with TAS and tracks at lower signal efficiencies, e.g. the 25 % working point. The improvements due to TAS as input for D2 are found to be single-digit percentages for low p_{T} and up to 20 % for the highest studied p_{T} bin at 50 % Higgs boson efficiency. For the lower working point, they reach around 30 % of the QCD rejection achieved with calorimeter cluster based D2.

6.3.3. Tagging Top Jets

Studied for top quark tagging is the n-Subjettiness ratio τ_{32}. In section 6.2.4, it was found, that tracks and TAS can improve τ_{32} for discrimination between QCD and top jets. Presented for τ_{32} in Table 6.6, are values of $\beta \geq 1$, since the W boson and Higgs boson parts affirmed the expected lower performance of track and TAS based variables with an angular weighting of $\beta = 0.5$.

The calorimeter τ_{32} variable is observed to profit from a higher angular weighting up to around $\beta = 2$, but degrades in performance for $\beta = 3$. Nevertheless, since the involved three prong structure of the top quark decay requires a good angular separation of the jet constituents to be resolved, tracks and TAS perform superior to calorimeter clusters for QCD rejection with τ_{32}.

A higher angular weighting is found not to significantly improve the separation of track and TAS variables, a value of $\beta = 2$ already diminishes the performance. This can be understood by comparing the distributions in Figure 6.24. The TAS signal distribution

improves by moving further left compared to the calorimeter τ_{32} distribution. Nevertheless, the background distribution for TAS is found to start rising at considerably lower values, resulting in a comparable to slightly worse background rejection at 50% signal efficiency compared to the variable involving calorimeter clusters.

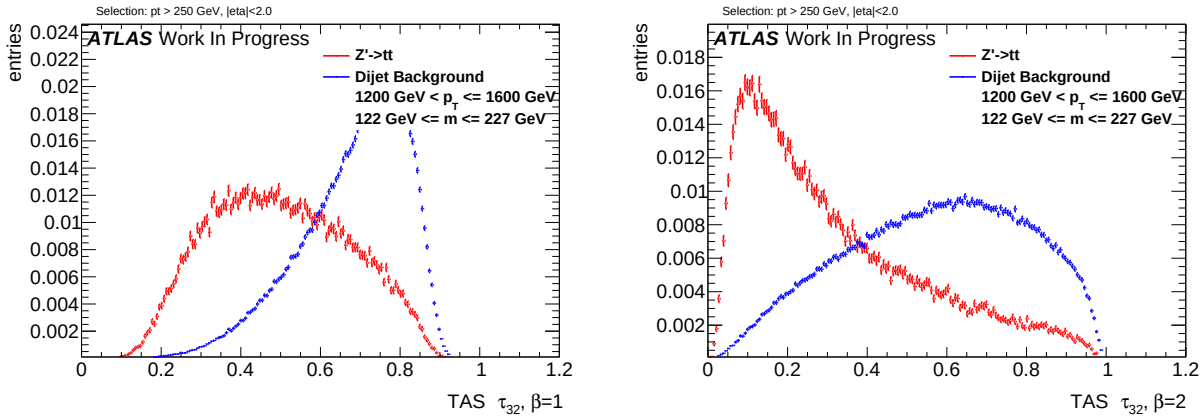

Figure 6.24.: Top signal and QCD background distributions for $\beta = 1$ (left) and $\beta = 2$ (right) with a p_{T} between 1200 and 1600 GeV.

From these observations it can be concluded, that the best discrimination is achieved with TAS with a value of $\beta = 1$ and $\beta = 1.7$. The marginal differences between both values of β depend on the considered p_{T} region. Regarding the effect of rescaling the p_{T} of the tracks, track τ_{32} achieves a lower separation as TAS, except for regions with very high boosts, but as well outperforms the calorimeter variable.

TAS τ_{32} with $\beta = 1$ and $\beta = 1.7$

Shown in Figure 6.25 are the cut values against the p_{T} of the large-R calorimeter jet for TAS τ_{32} with $\beta = 1$ and $\beta = 1.7$, in order to achieve the working points at 50 % and 25 % signal efficiency. The crack between the first and second p_{T} bin is more evident since the top quark with its much higher mass is here very unlikely to be reconstructed into a single large-R jet, resulting in background like signal events.

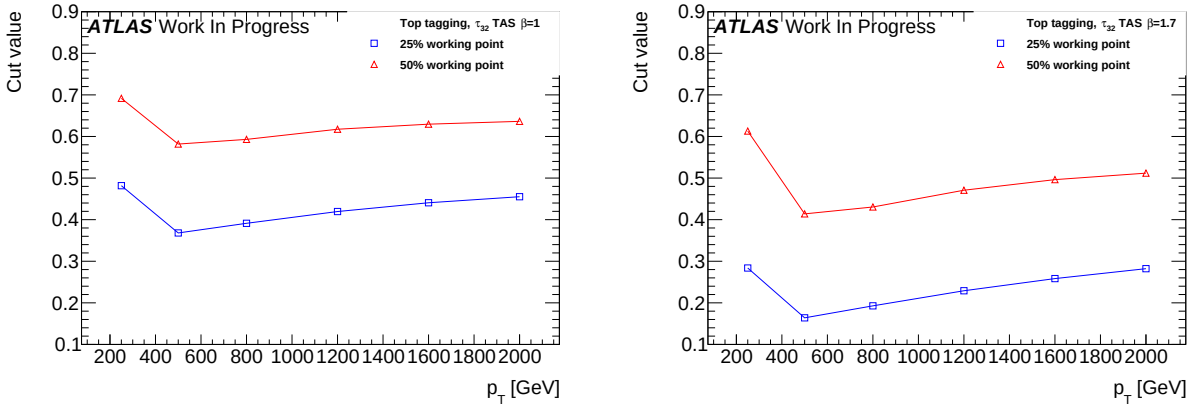

Figure 6.25.: Cut values for TAS τ_{32} at $\beta = 1$ (left) and $\beta = 1.7$ (right) to achieve 50% and 25% top quark efficiency

Calorimeter	τ_{32}			
$p_T\,[GeV]$	$\beta = 1$	1.7	2	3
250 - 500	9.7 ± 0.2	9.5 ± 0.2	9.5 ± 0.4	9.4 ± 0.2
500 - 800	20.1 ± 0.5	22.2 ± 0.6	22.4 ± 0.6	22.0 ± 0.6
800 - 1200	17.3 ± 0.4	20.3 ± 0.5	20.6 ± 0.5	20.3 ± 0.5
1200 - 1600	14.3 ± 0.3	16.4 ± 0.4	16.6 ± 0.5	16.1 ± 0.5
1600 - 2000	11.7 ± 0.3	13.3 ± 0.4	13.3 ± 0.4	12.6 ± 0.3
> 2000	9.6 ± 0.3	11.0 ± 0.4	10.9 ± 0.4	10.1 ± 0.3

TAS	τ_{32}			
$p_T\,[GeV]$	$\beta = 1$	1.7	2	3
250 - 500	10.7 ± 0.2	10.1 ± 0.2	9.9 ± 0.2	9.6 ± 0.2
500 - 800	22.8 ± 0.6	22.8 ± 0.6	22.5 ± 0.6	21.6 ± 0.6
800 - 1200	23.6 ± 0.6	24.1 ± 0.6	23.6 ± 0.6	22.2 ± 0.5
1200 - 1600	22.0 ± 0.6	22.3 ± 0.6	21.7 ± 0.6	19.8 ± 0.6
1600 - 2000	18.9 ± 0.6	18.8 ± 0.6	17.9 ± 0.5	16.0 ± 0.5
> 2000	16.5 ± 0.7	15.7 ± 0.7	15.2 ± 0.7	13.1 ± 0.6

Tracks	τ_{32}			
$p_T\,[GeV]$	$\beta = 1$	1.7	2	3
250 - 500	10.5 ± 0.2	9.8 ± 0.2	9.6 ± 0.2	9.4 ± 0.2
500 - 800	20.6 ± 0.5	21.3 ± 0.6	21.1 ± 0.5	20.3 ± 0.5
800 - 1200	21.8 ± 0.6	22.9 ± 0.6	22.6 ± 0.6	21.4 ± 0.6
1200 - 1600	21.7 ± 0.6	22.1 ± 0.6	21.6 ± 0.6	19.5 ± 0.6
1600 - 2000	19.3 ± 0.6	19.0 ± 0.6	18.2 ± 0.6	16.0 ± 0.5
> 2000	16.8 ± 0.7	15.8 ± 0.7	15.1 ± 0.7	13.0 ± 0.5

Table 6.6.: Listing of the QCD background rejection for top signal achieved with τ_{32} together with different angular weightings β and for calorimeter cluster, tracks and TAS.

Furthermore, τ_{32} ($\beta = 1.7$) needs to be cut at lower values as τ_{32} ($\beta = 1$) to achieve a certain signal efficiency. This is the result of the higher angular weighting that shifts the overall distributions to lower values, because the angular distance between two constituents inside a (highly) boosted large-R jet is in the majority of cases lower than one. Thus, the angular part of τ_{32} decreases with $\beta > 1$.

The TAS τ_{32} cut value is observed to be robust against variations of p_T, in accordance to the results of the p_T correlation plots, see 6.10.

50% ϵ_{signal}	**Top Tagging**		
p_{T} [GeV]	Calo τ_{32}, $\beta = 2$	TAS τ_{32}, $\beta = 1$	TAS τ_{32}, $\beta = 1.7$
250 - 500	9.5 ± 0.2	10.7 ± 0.2 (+13 ± 3 %)	10.1 ± 0.2 (+6 ± 3 %)
500 - 800	22.4 ± 0.6	22.8 ± 0.6 (+2 ± 4 %)	22.8 ± 0.6 (+2 ± 4 %)
800 - 1200	20.6 ± 0.5	23.6 ± 0.6 (+15 ± 4 %)	24.1 ± 0.6 (+17 ± 4 %)
1200 - 1600	16.6 ± 0.4	22.0 ± 0.6 (+33 ± 5 %)	22.3 ± 0.6 (+34 ± 5 %)
1600 - 2000	13.3 ± 0.4	18.9 ± 0.6 (+42 ± 6 %)	18.8 ± 0.6 (+41 ± 6 %)
> 2000	10.9 ± 0.4	16.5 ± 0.7 (+51 ± 8 %)	15.7 ± 0.7 (+44 ± 8 %)

25% ϵ_{signal}	**Top Tagging**		
p_{T} [GeV]	Calo τ_{32}, $\beta = 2$	TAS τ_{32}, $\beta = 1$	TAS τ_{32}, $\beta = 1.7$
250 - 500	33.7 ± 1.0	37.6 ± 1.4 (+12 ± 5 %)	36.7 ± 1.2 (+9 ± 5 %)
500 - 800	114.7 ± 3.3	138.0 ± 4.3 (+20 ± 5 %)	139.1 ± 4.2 (+21 ± 5 %
800 -1200	97.0 ± 2.7	144.6 ± 4.9 (+49 ± 7 %)	149.6 ± 5.2 (+54 ± 7 %)
1200 - 1600	68.6 ± 2.1	133.2 ± 4.6 (+94 ± 9 %)	134.7 ± 5.1 (+96 ± 10 %)
1600 - 2000	47.5 ± 1.6	100.3 ± 4.2 (+111 ± 11 %)	99.9 ± 4.4 (+110 ± 12 %)
> 2000	36.3 ± 1.6	80.2 ± 5.0 (+121 ± 17 %)	75.5 ± 4.9 (+108 ± 16 %)

Table 6.7.: Listing of the background rejections after the jet mass cut and tagging at 50% and 25% top signal efficiency for the identified best variables TAS τ_{32} ($\beta = 1$, 1.7) together with the improvements over the best variable with calorimeter clusters which is D2 at $\beta = 2$.

Moreover, Table 6.7 lists the background rejections for TAS τ_{32} at $\beta = 1$ and $\beta = 1.7$ as well as for the best calorimeter variable, τ_{32} with $\beta = 2$. The differences between TAS τ_{32} with $\beta = 1$ and $\beta = 1.7$ are found to be marginal, as well for lower signal efficiencies. Furthermore, improvements due to the use of TAS instead of calorimeter clusters are possible for top quark tagging over the whole studied p_{T} range. These enhancements are, as expected, rising with the boost of the top quark and can reach around 50 % for the 50 % working point and even 100 % for 25 % top efficiency.

7. Conclusions

The efficient differentiation of QCD background jets and signal jets originated by boosted, hadronically decaying massive gauge bosons, Higgs bosons or top quarks is important for the potential significance of searches for physics beyond the Standard Model at the TeV scale.

Studied in this thesis was the background rejection of τ_{21}, τ_{32}, C2 and D2 calculated with tracks and assisted tracks (TAS) compared to the currently used calorimeter clusters for W boson, Higgs boson and top quark tagging. The performance was studied in six different p_{T} regions from 250 to $\sim 2000\,\mathrm{GeV}$ with signal and background distributions as well as with ROCs. Furthermore, different angular weightings were studied to identify the best tagging variables.

The ghost association of tracks to $R = 0.2$ k_{T}-subjets of the large-R calorimeter jet was found to yield a better separation of signal and background for track and TAS based variables, as the association of tracks to the ungroomed large-R jet.

For W boson tagging, TAS C2 and D2, both with $\beta = 1.7$, were identified as the best performing discrimination variables in the studied phase space. At the 50 % (25 %) signal efficiency working point, improvements in QCD rejection over the currently used calorimeter D2 ($\beta = 1$) were found to reach 50 % (200 %) for D2 TAS, and up to 100 % (250 %) for C2 TAS in the highest analyzed energy regions.

In case of Higgs boson tagging, the observed advantages of tracks and TAS are smaller due to the large mass, resulting in a higher separation of the $b\bar{b}$ pair. Here, best performing was D2 at $\beta = 1.7$ with TAS and tracks, with improvements over the best calorimeter variable, D2 at $\beta = 1.7$, of up to 20 % (30 %) at the 50 % (25 %) working point which again show a rising behaviour with the Higgs boost.

For top tagging, the τ_{32} variable was examined. Despite the high top quark mass pushing the decay products apart, the identification of the involved three-prong structure benefits from the angular resolution of tracks, resulting in TAS τ_{32} ($\beta = 1$ and 1.7) found to be the best discriminating variables. The improvements in comparison with calorimeter clusters, which perform best with $\beta = 2$, are found to reach around 50 % (100 %) for the higher p_{T} regions at the 50 % (25 %) working point.

In every case, improvements were observed in all studied p_{T} regions. The advantages of tracks and TAS are found to be most distinct for the higher p_{T} regions. The transverse momentum further decreases the separation of decay products, which increases the influence of the excellent angular resolution of tracks.

Furthermore, it was found that tracks and TAS do perform comparable in most cases with slight advantages for TAS. The differences were predominantly observed in the lower p_{T} regions, since the assisting was shown to be increasingly ineffective for higher energies.

8. Outlook

A difference between tracks and TAS is the time scale for the derivation of systematic uncertainties on the substructure variables, which is needed, before the here introduced variables with tracks or TAS can be used. TAS requires a derivation of uncertainties on the calorimeter subjets, which are being worked at this moment, before these can be propagated to a possible $D2^{\mathrm{TAS}}$ or similar. On the other hand, tracks already offer well known uncertainties which need to be propagated.

Another important study which needs to be performed is the robustness of the track selection to the definition of calorimeter subjets. As mentioned, the ghost association to calorimeter subjets is of utmost importance for the performance of tracks and TAS. The calorimeter subjets are the result of the trimming procedure, which commonly uses the k_{T}-algorithm to define the subjets. Other possibilities are to use the antik_{T} or Cambridge/Aachen algorithms, to define the subjets. The performance of tracks associated to those needs to be evaluated, whereas the author expects only marginal changes in performance, because the most important function of the ghost association is to exclude the tracks, that were shown to lie around the borders of the large-R jet. In contrast, changing the subjet algorithm would result in a minor fluctuation of tracks that are used or excluded around the different subjet areas.

Bibliography

[1] M. E. Peskin and D. V. Schröder. *An Introduction to Quantum Field Theory*. Westview Press, 1995.

[2] D. J. Griffith. *An Introduction to Elementary Particles*. Wiley, 1997.

[3] Mark Thomson. *Modern Particle Physics*. Cambridge University Press, 2013.

[4] Steven Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19 (1967), pp. 1264–1266.

[5] Sheldon Lee Glashow. "Partial Symmetries of Weak Interactions". In: *Nucl. Phys.* 22 (1961), pp. 579–588.

[6] Abdus Salam and John Clive Ward. "Electromagnetic and Weak Interactions". In: *Phys. Lett.* 13 (1964), pp. 168–171.

[7] K.A. Olive and Particle Data Group. "Review of Particle Physics". In: *Chinese Physics C* 38.9 (2014), p. 090001. URL: http://stacks.iop.org/1674-1137/38/i=9/a=090001.

[8] P. W. Higgs. "Broken Symmetries, Massless Particles and Gauge Fields". In: *Phys. Lett.* 12 (1964), pp. 132–133.

[9] V. L Ginzburg and L. D. Landau. "On the Theory of Superconductivity". In: *Zh. Eksp. Teor. Fizz.* 20 (1950), p. 1064.

[10] *Nobel prize 2013*. URL: http://www.nobelprize.org/nobel_prizes/physics/laureates/2013/.

[11] *Nobel prize 1984*. URL: http://www.nobelprize.org/nobel_prizes/physics/laureates/1984/.

[12] A. M. Cooper-Sarkar. "PDF Fits at HERA". In: *PoS* EPS-HEP2011 (2011), p. 320. arXiv: 1112.2107 [hep-ph].

[13] *Higgs potential, wikipedia page*. URL: https://en.wikipedia.org/wiki/Higgs_mechanism#Abelian_Higgs_mechanism.

[14] S. Dittmaier et al. "Handbook of LHC Higgs Cross Sections: 2. Differential Distributions". In: (2012). DOI: 10.5170/CERN-2012-002. arXiv: 1201.3084 [hep-ph].

[15] Felix Müller. "Jet production measurements at the ATLAS experiment". dissertation. Universität Heidelberg, 2013. URL: http://www.kip.uni-heidelberg.de/Veroeffentlichungen/details.php?id=2966.

[16] B.R. Webber. "A QCD model for jet fragmentation including soft gluon interference". In: *Nuclear Physics B* 238.3 (1984), pp. 492–528. ISSN: 0550-3213. DOI: http://dx.doi.org/10.1016/0550-3213(84)90333-X. URL: http://www.sciencedirect.com/science/article/pii/055032138490333X.

[17] Thomas D. Gottschalk. "An improved description of hadronization in the QCD cluster model for e^+e^- annihilation". In: *Nuclear Physics B* 239.2 (1984), pp. 349–381. ISSN: 0550-3213. DOI: http://dx.doi.org/10.1016/0550-3213(84)90253-0. URL: http://www.sciencedirect.com/science/article/pii/0550321384902530.

[18] Bo Andersson. *The Lund Model*. Cambridge Books Online. Cambridge University Press, 1998. ISBN: 9780511524363. URL: http://dx.doi.org/10.1017/CBO9780511524363.

[19] Paul Langacker. "The Physics of Heavy Z' Gauge Bosons". In: *Rev. Mod. Phys.* 81 (2009), pp. 1199–1228. DOI: 10.1103/RevModPhys.81.1199. arXiv: 0801.1345 [hep-ph].

[20] Martin Schmaltz and Christian Spethmann. "Two Simple W' Models for the Early LHC". In: *JHEP* 07 (2011), p. 046. DOI: 10.1007/JHEP07(2011)046. arXiv: 1011.5918 [hep-ph].

[21] Georges Aad et al. "Combination of searches for WW, WZ, and ZZ resonances in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector". In: *Phys. Lett.* B755 (2016), pp. 285–305. DOI: 10.1016/j.physletb.2016.02.015. arXiv: 1512.05099 [hep-ex].

[22] Duccio Pappadopulo et al. "Heavy Vector Triplets: Bridging Theory and Data". In: *JHEP* 09 (2014), p. 060. DOI: 10.1007/JHEP09(2014)060. arXiv: 1402.4431 [hep-ph].

[23] Lisa Randall and Raman Sundrum. "A Large mass hierarchy from a small extra dimension". In: *Phys. Rev. Lett.* 83 (1999), pp. 3370–3373. DOI: 10.1103/PhysRevLett.83.3370. arXiv: hep-ph/9905221 [hep-ph].

[24] URL: http://lhcb-public.web.cern.ch/lhcb-public/en/detector/Detector-en.html.

[25] The ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *Journal of Instrumentation* 3.08 (2008), S08003. URL: http://stacks.iop.org/1748-0221/3/i=08/a=S08003.

[26] M Capeans et al. *ATLAS Insertable B-Layer Technical Design Report*. Tech. rep. CERN-LHCC-2010-013. ATLAS-TDR-19. Sept. 2010. URL: https://cds.cern.ch/record/1291633.

[27] Georges Aad et al. "Performance of jet substructure techniques for large-R jets in proton-proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector". In: *JHEP* 09 (2013), p. 076. DOI: 10.1007/JHEP09(2013)076. arXiv: 1306.4945 [hep-ex].

[28] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The Anti-k(t) jet clustering algorithm". In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

[29] *Determination of the jet energy scale and resolution at ATLAS using Z/γ–jet events in data at $\sqrt{s} = 8$ TeV*. Tech. rep. ATLAS-CONF-2015-057. Geneva: CERN, Oct. 2015. URL: https://cds.cern.ch/record/2059846.

[30] Ariel Schwartzman. "Jet energy calibration at the LHC". In: *Int. J. Mod. Phys.* A30.31 (2015), p. 1546002. DOI: 10.1142/S0217751X15460021. arXiv: 1509.05459 [hep-ex].

[31] Jesse Thaler and Ken Van Tilburg. "Identifying Boosted Objects with N-subjettiness". In: *JHEP* 03 (2011), p. 015. DOI: 10.1007/JHEP03(2011)015. arXiv: 1011.2268 [hep-ph].

[32] Andrew J. Larkoski, Gavin P. Salam, and Jesse Thaler. "Energy Correlation Functions for Jet Substructure". In: *JHEP* 06 (2013), p. 108. DOI: 10.1007/JHEP06(2013)108. arXiv: 1305.0007 [hep-ph].

[33] Andrew J. Larkoski, Ian Moult, and Duff Neill. "Power Counting to Better Jet Observables". In: *JHEP* 12 (2014), p. 009. DOI: 10.1007/JHEP12(2014)009. arXiv: 1409.6298 [hep-ph].

[34] Andrew J. Larkoski, Ian Moult, and Duff Neill. "Analytic Boosted Boson Discrimination". In: *JHEP* 05 (2016), p. 117. DOI: 10.1007/JHEP05(2016)117. arXiv: 1507.03018 [hep-ph].

[35] Georges Aad et al. "Identification of boosted, hadronically decaying W bosons and comparisons with ATLAS data taken at $\sqrt{s} = 8$ TeV". In: *Eur. Phys. J.* C76.3 (2016), p. 154. DOI: 10.1140/epjc/s10052-016-3978-z. arXiv: 1510.05821 [hep-ex].

[36] *Jet mass reconstruction with the ATLAS Detector in early Run 2 data*. Tech. rep. ATLAS-CONF-2016-035. Geneva: CERN, July 2016. URL: http://cds.cern.ch/record/2200211.

[37] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "FastJet User Manual". In: *Eur. Phys. J.* C72 (2012), p. 1896. DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph].

[38] S. Agostinelli et al. "GEANT4: A Simulation toolkit". In: *Nucl. Instrum. Meth.* A506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8.

A. Monte Carlo Samples

Listed here are the information about the used Monte-Carlo generators, tunings and number of simulated events.

A.1. Signal

W Boson Tagging Study

Used as source for boosted W bosons are W' resonances: $W' \to WZ \to qqqq$, see Table A.1. The MC samples are based on simulations with a center-of-mass energy of 13 TeV.

$m_{W'}$ [GeV]	Simulator	Data format	p tag	tuning	#events
1500	Pythia8 & EvtGen	DAOD_JETM8	p2425	A14NNPDF23LO	29894
2500	Pythia8 & EvtGen	DAOD_JETM8	p2425	A14NNPDF23LO	29818
3000	Pythia8 & EvtGen	DAOD_JETM8	p2425	A14NNPDF23LO	29538
4000	Pythia8 & EvtGen	DAOD_JETM8	p2425	A14NNPDF23LO	29505
5000	Pythia8 & EvtGen	DAOD_JETM8	p2425	A14NNPDF23LO	29068

Table A.1.: MC samples used for the W boson study.

Higgs Boson Tagging Study

Used as source for boosted Higgs bosons are G^* resonances: $G^* \to HH \to 4b$, see Table A.2. The MC samples are based on simulations with a center-of-mass energy of 13 TeV.

m_{G^*} [GeV]	Simulator	Data format	p tag	tuning	#events
500	MadGraph & Pythia8 & EvtGen	DAOD_JETM8	p2425	A14NNPDF23LO	93281
1000	MadGraph & Pythia8 & EvtGen	DAOD_JETM8	p2425	A14NNPDF23LO	99999
1500	MadGraph & Pythia8 & EvtGen	DAOD_JETM8	p2425	A14NNPDF23LO	98999
2000	MadGraph & Pythia8 & EvtGen	DAOD_JETM8	p2425	A14NNPDF23LO	88800
2500	MadGraph & Pythia8 & EvtGen	DAOD_JETM8	p2425	A14NNPDF23LO	60000
3000	MadGraph & Pythia8 & EvtGen	DAOD_JETM8	p2425	A14NNPDF23LO	78000

Table A.2.: MC samples used for the Higgs boson study.

Top Quark Tagging Study

Used as source for boosted top quarks are Z' resonances: $Z' \to t\bar{t}$, see Table A.3. The MC samples are based on simulations with a center-of-mass energy of 13 TeV.

$m_{Z'}$ [GeV]	Simulator	Data format	p tag	tuning	#events
1500	Pythia8 & EvtGen	DAOD_JETM8	p2375	A14NNPDF23LO	197466
1750	Pythia8 & EvtGen	DAOD_JETM8	p2375	A14NNPDF23LO	198765
2500	Pythia8 & EvtGen	DAOD_JETM8	p2375	A14NNPDF23LO	197490
3000	Pythia8 & EvtGen	DAOD_JETM8	p2425	A14NNPDF23LO	195309
4000	Pythia8 & EvtGen	DAOD_JETM8	p2375	A14NNPDF23LO	198611
4000	Pythia8 & EvtGen	DAOD_JETM8	p2425	A14NNPDF23LO	198958
5000	Pythia8 & EvtGen	DAOD_JETM8	p2425	A14NNPDF23LO	175841

Table A.3.: MC samples used for the top quark study.

A.2. Background

Used as source for QCD jets are dijet events, see Table A.4. The MC samples are based on simulations with a center-of-mass energy of 13 TeV.

Tags	Simulator	Data format	tuning	#events
e3569_s2576_s2132_r7773_r7676_p2613	Pythia8 & EvtGen	DAOD_JETM8	A14NNPDF23LO	41793
e3569_s2576_s2132_r7773_r7676_p2613	Pythia8 & EvtGen	DAOD_JETM8	A14NNPDF23LO	110610
e3668_s2576_s2132_r7773_r7676_p2613	Pythia8 & EvtGen	DAOD_JETM8	A14NNPDF23LO	1536718
e3668_s2576_s2132_r7773_r7676_p2613	Pythia8 & EvtGen	DAOD_JETM8	A14NNPDF23LO	5891992
e3668_s2576_s2132_r7773_r7676_p2613	Pythia8 & EvtGen	DAOD_JETM8	A14NNPDF23LO	1989195
e3668_s2576_s2132_r7773_r7676_p2613	Pythia8 & EvtGen	DAOD_JETM8	A14NNPDF23LO	5986897
e3569_s2608_s2183_r7773_r7676_p2613	Pythia8 & EvtGen	DAOD_JETM8	A14NNPDF23LO	1893398
e3668_s2608_s2183_r7773_r7676_p2613	Pythia8 & EvtGen	DAOD_JETM8	A14NNPDF23LO	1786199
e3569_s2576_s2132_r7773_r7676_p2613	Pythia8 & EvtGen	DAOD_JETM8	A14NNPDF23LO	1743198

Table A.4.: MC samples used for the QCD background prediction.

B. ROCs for the Best Variables

B.1. W boson Tagging

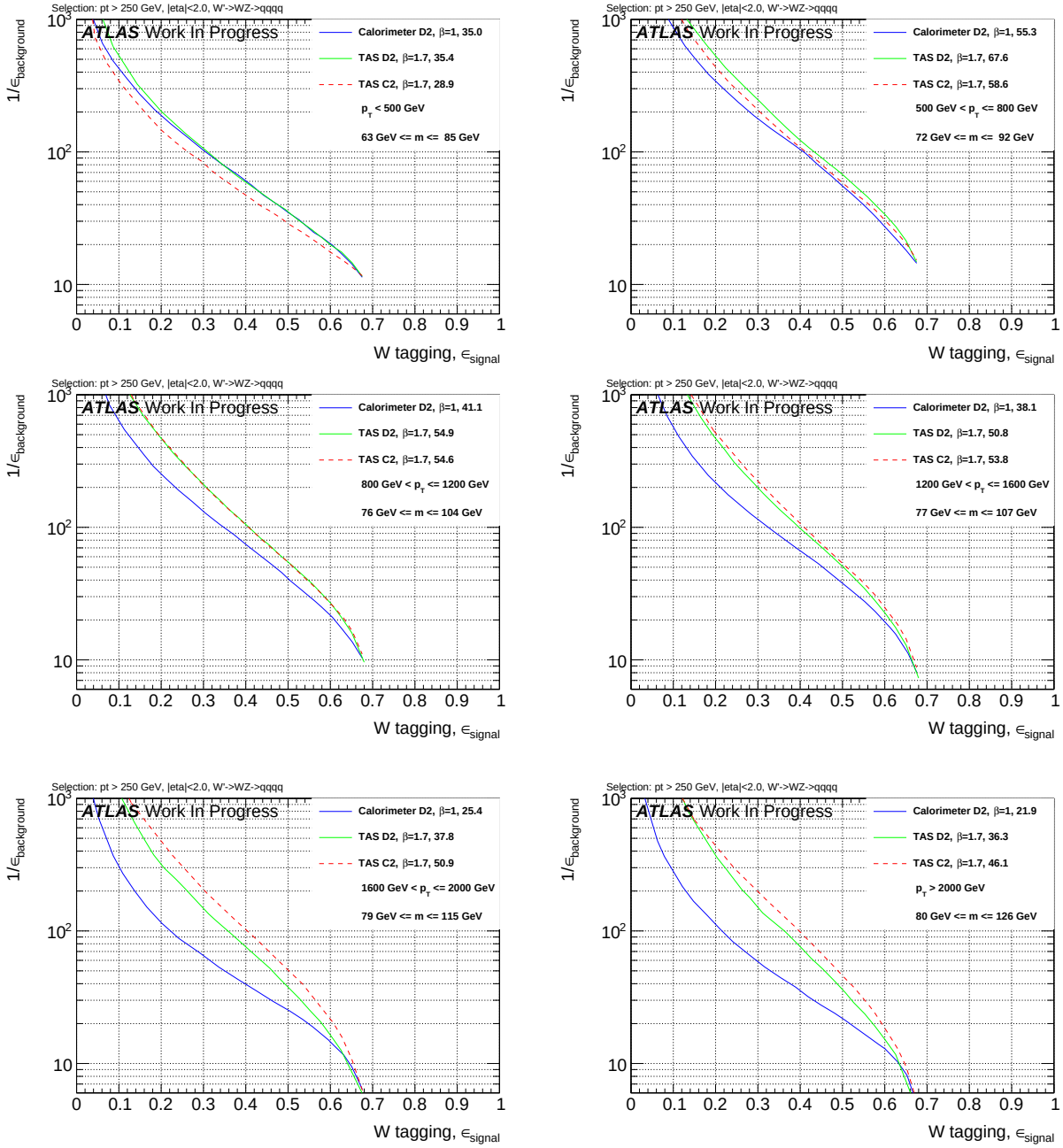

Figure B.1.: ROCs showing QCD rejection against W jet efficiency for TAS C2 and D2 ($\beta = 1.7$) compared to colorimeter D2 at $\beta = 1$

B.2. Higgs Boson Tagging

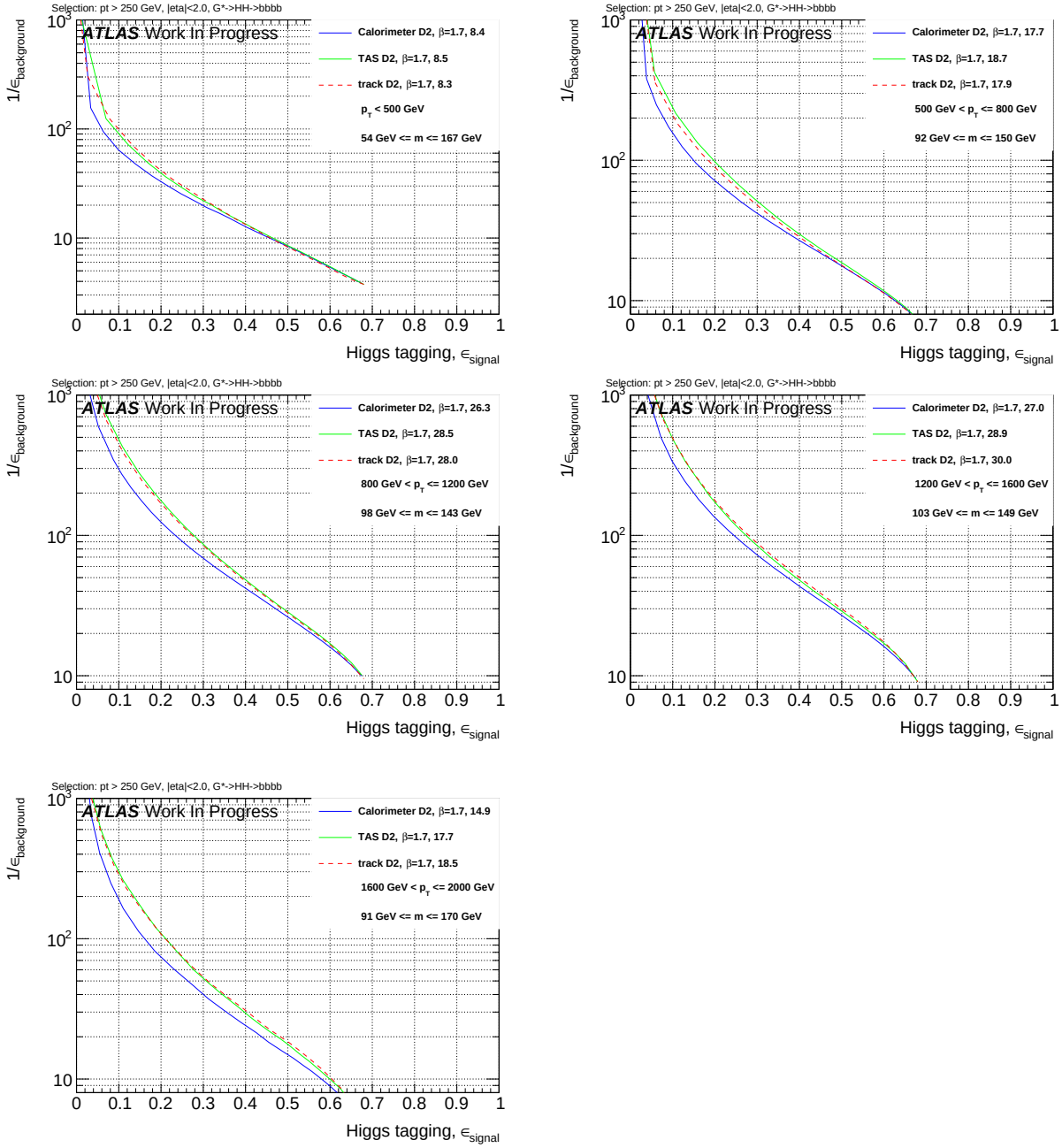

Figure B.2.: ROCs showing QCD rejection against Higgs jet efficiency for TAS and track D2 ($\beta = 1.7$) compared to colorimeter D2 at $\beta = 1$

B.3. Top Quark Tagging

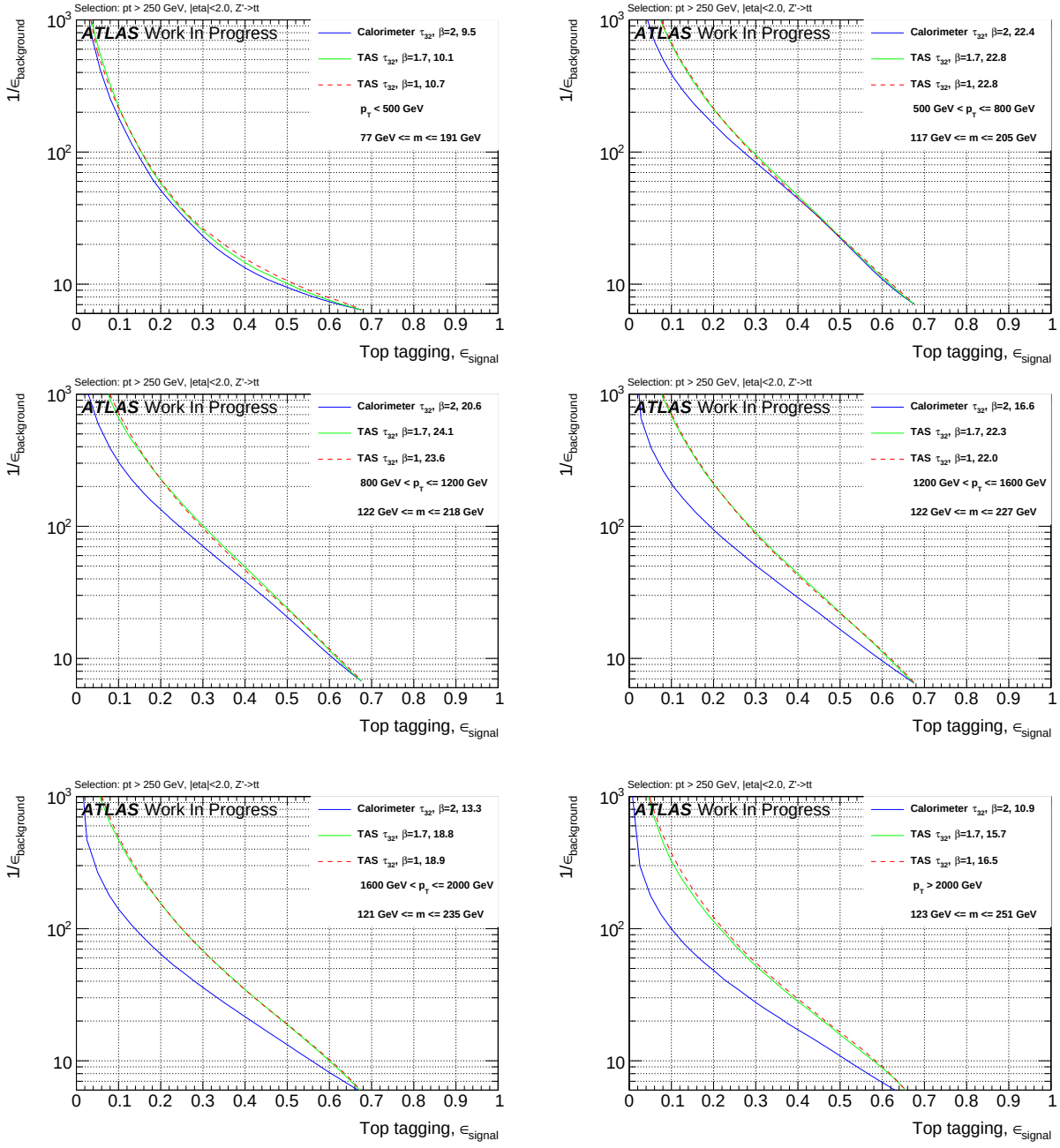

Figure B.3.: ROCs showing QCD rejection against Top jet efficiency for TAS τ_{32} ($\beta = 1, 1.7$) compared to colorimeter τ_{32} at $\beta = 2$

C. Signal and Background Distributions

C.1. W Distributions

$\beta = 0.5$

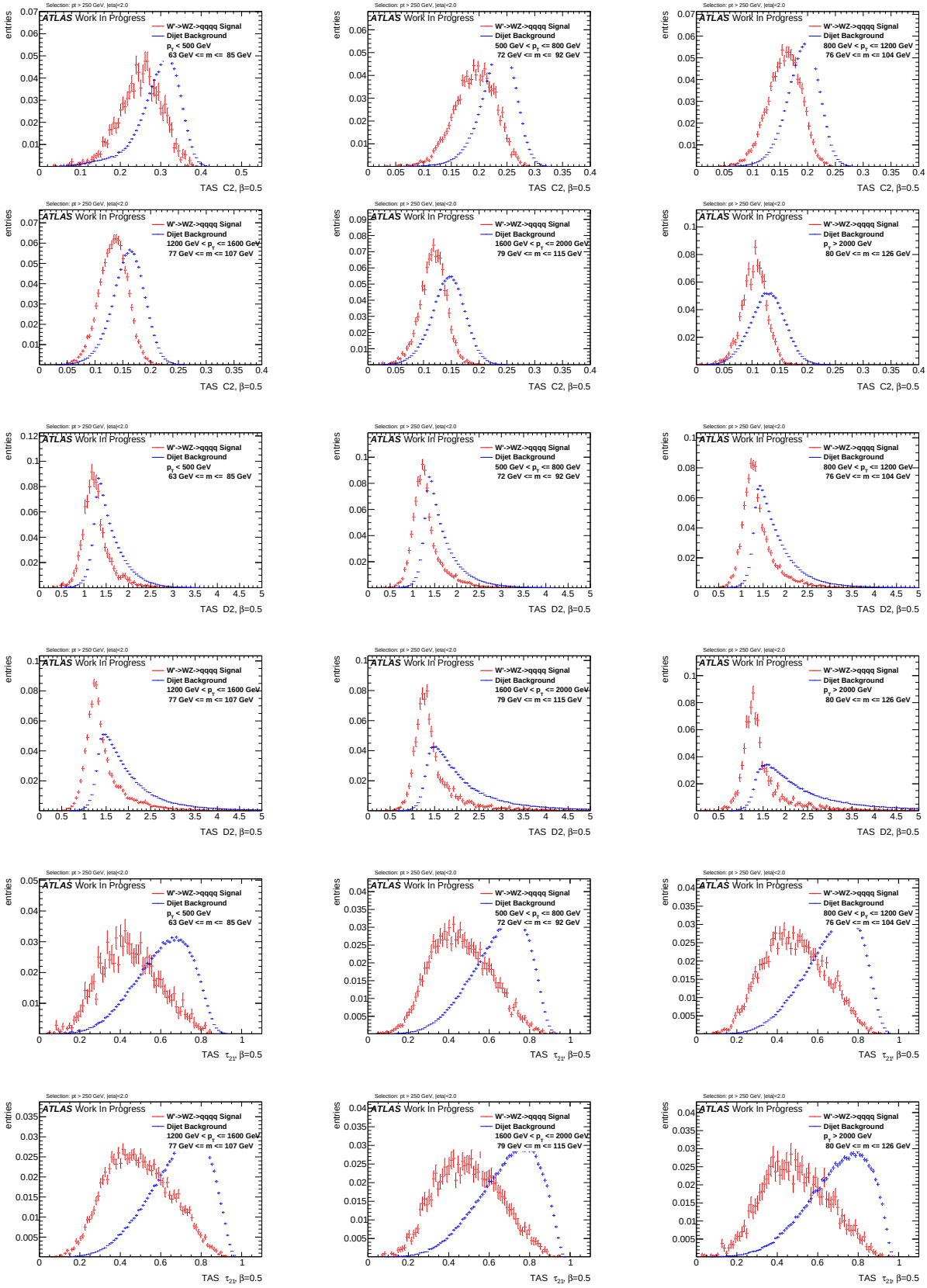

Figure C.1.: Distributions for W boson tagging using TAS $\beta = 0.5$. C2, D2, τ_{21} top down.

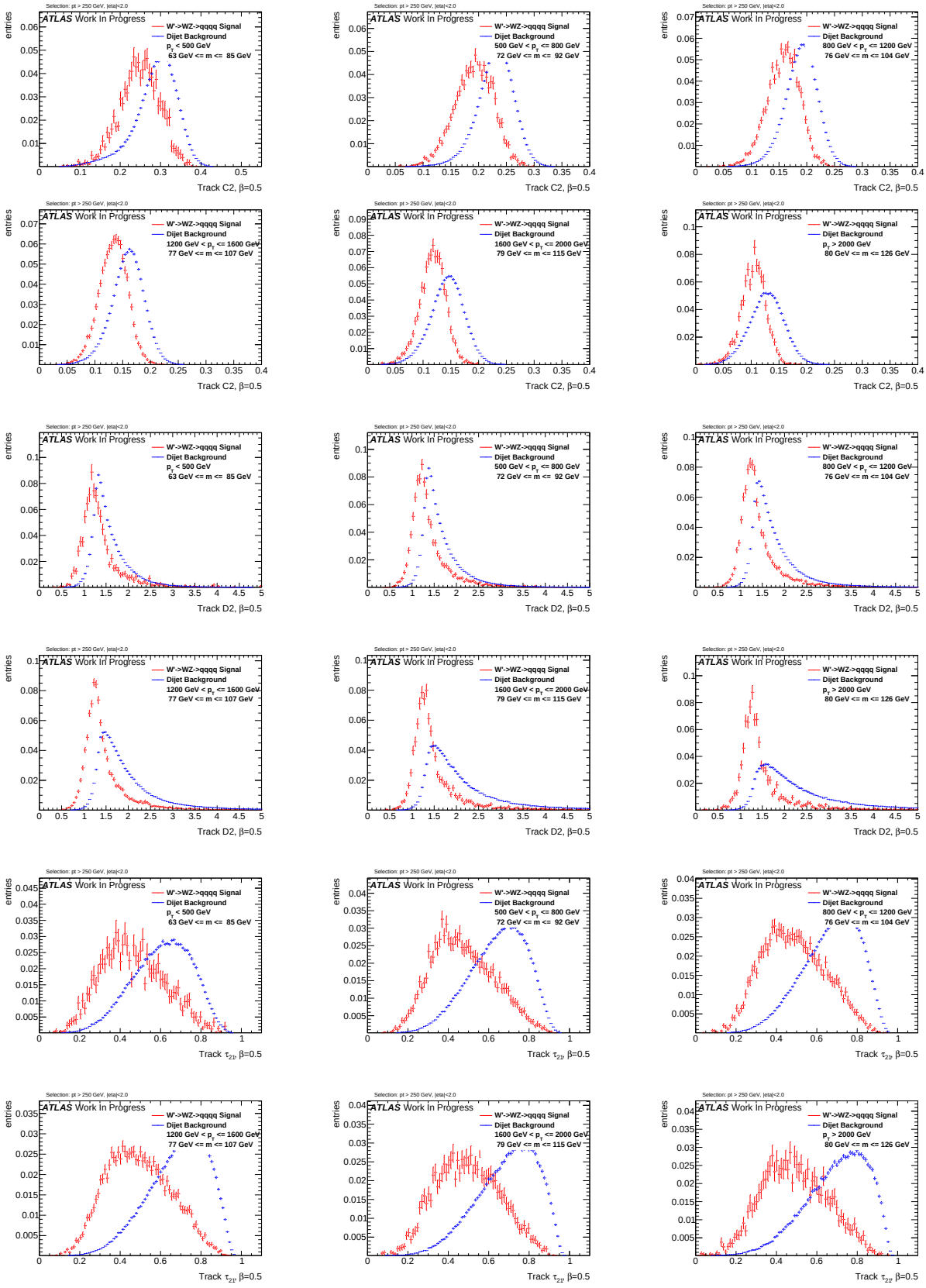

Figure C.2.: Distributions for W boson tagging using tracks $\beta = 0.5$. C2, D2, τ_{21} top down.

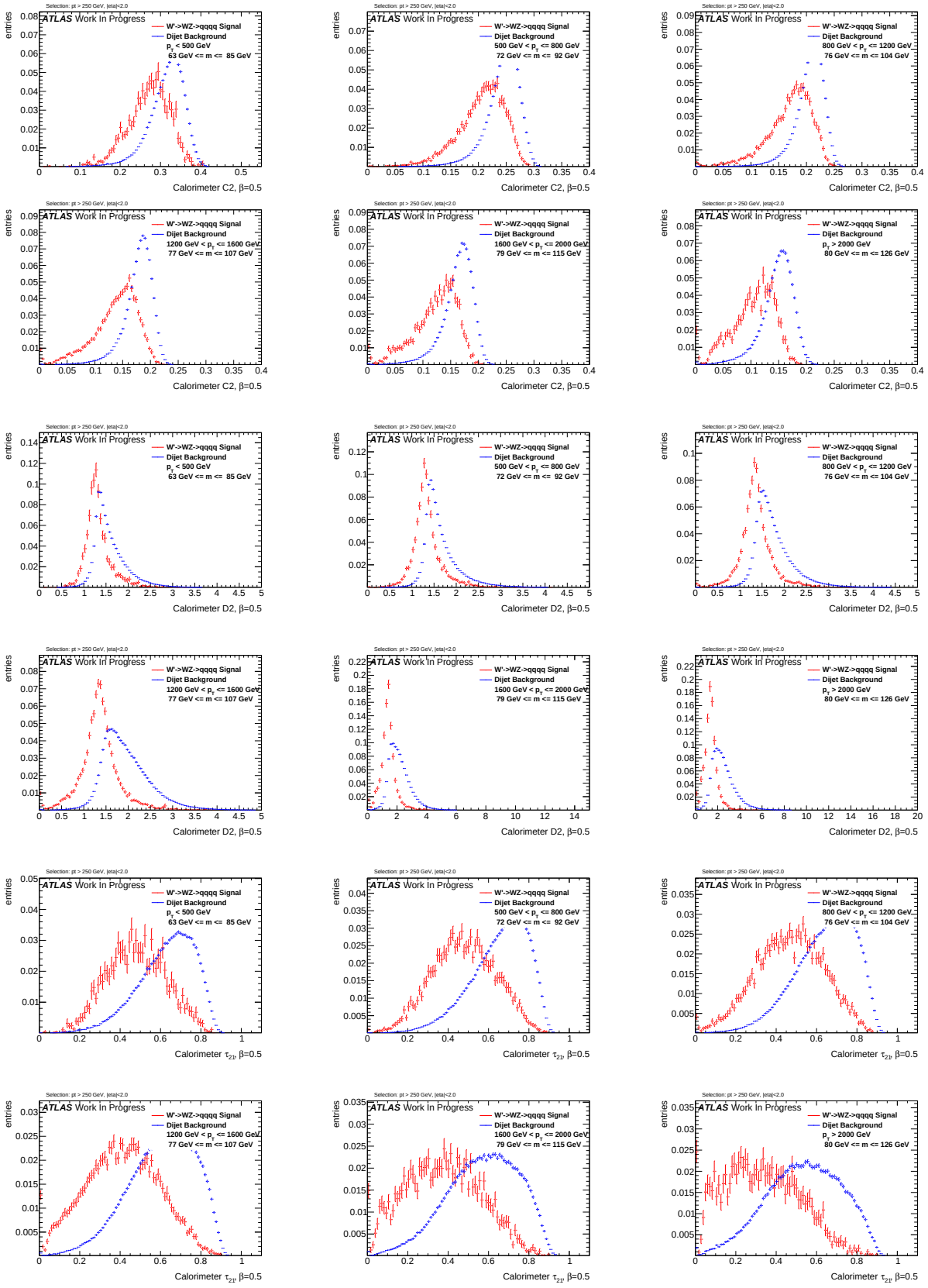

Figure C.3.: Distributions for W boson tagging using calorimeter clusters $\beta = 0.5$. C2, D2, τ_{21} top down.

$\beta = 1$

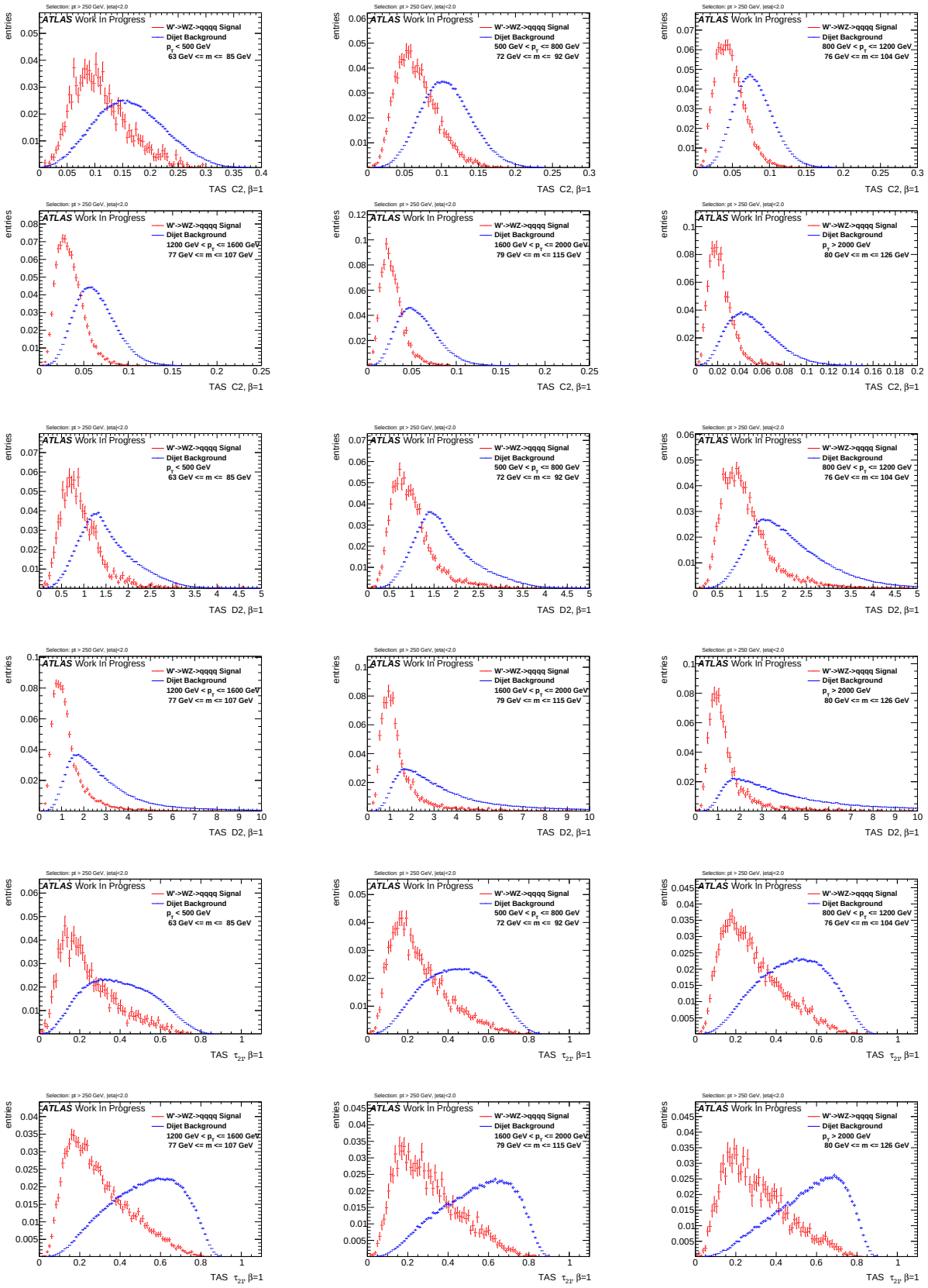

Figure C.4.: Distributions for W boson tagging using TAS $\beta = 1$. C2, D2, τ_{21} top down.

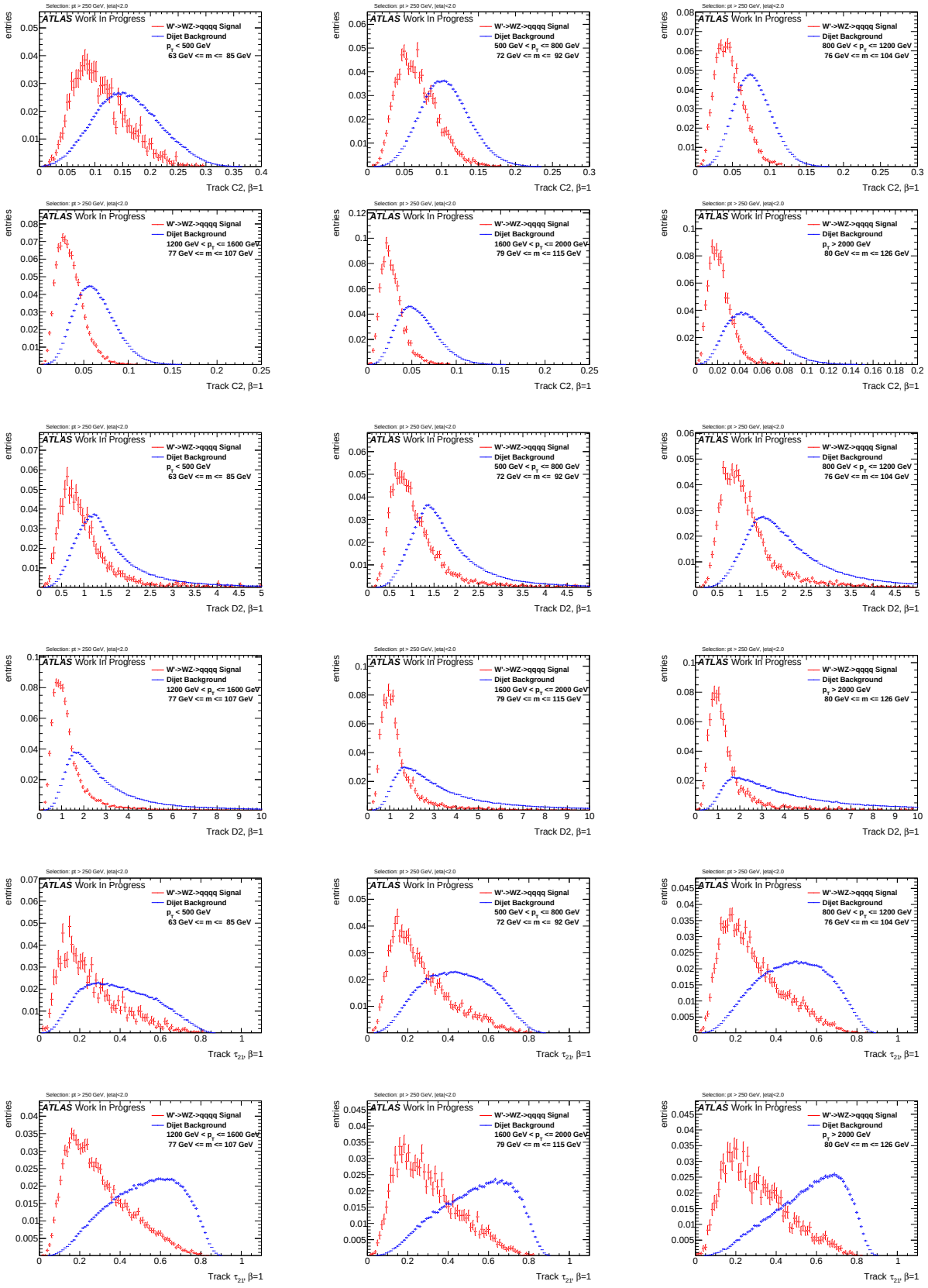

Figure C.5.: Distributions for W boson tagging using tracks $\beta = 1$. C2, D2, τ_{21} top down.

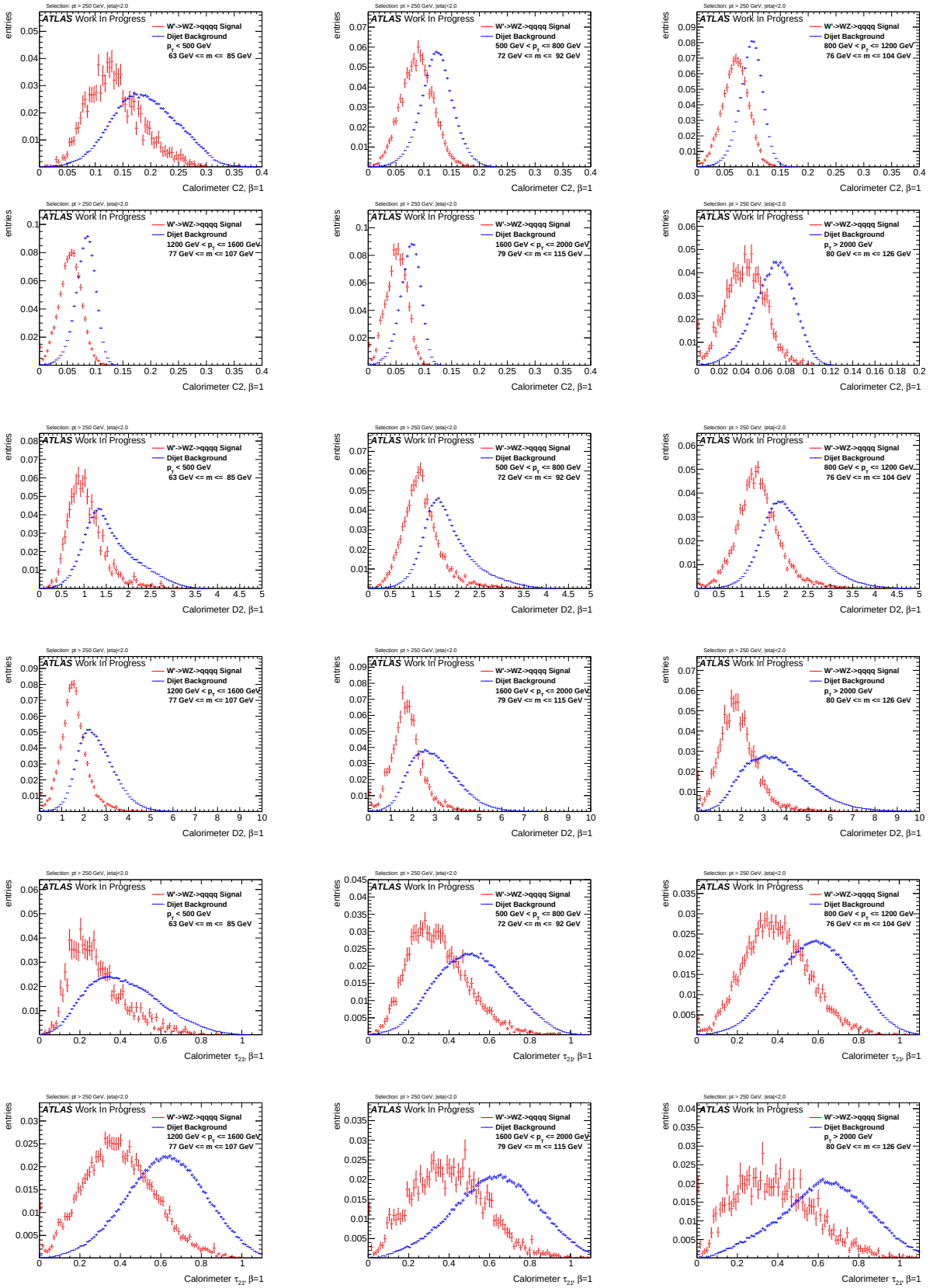

Figure C.6.: Distributions for W boson tagging using calorimeter clusters $\beta = 1$. C2, D2, τ_{21} top down.

$\beta = 1.7$

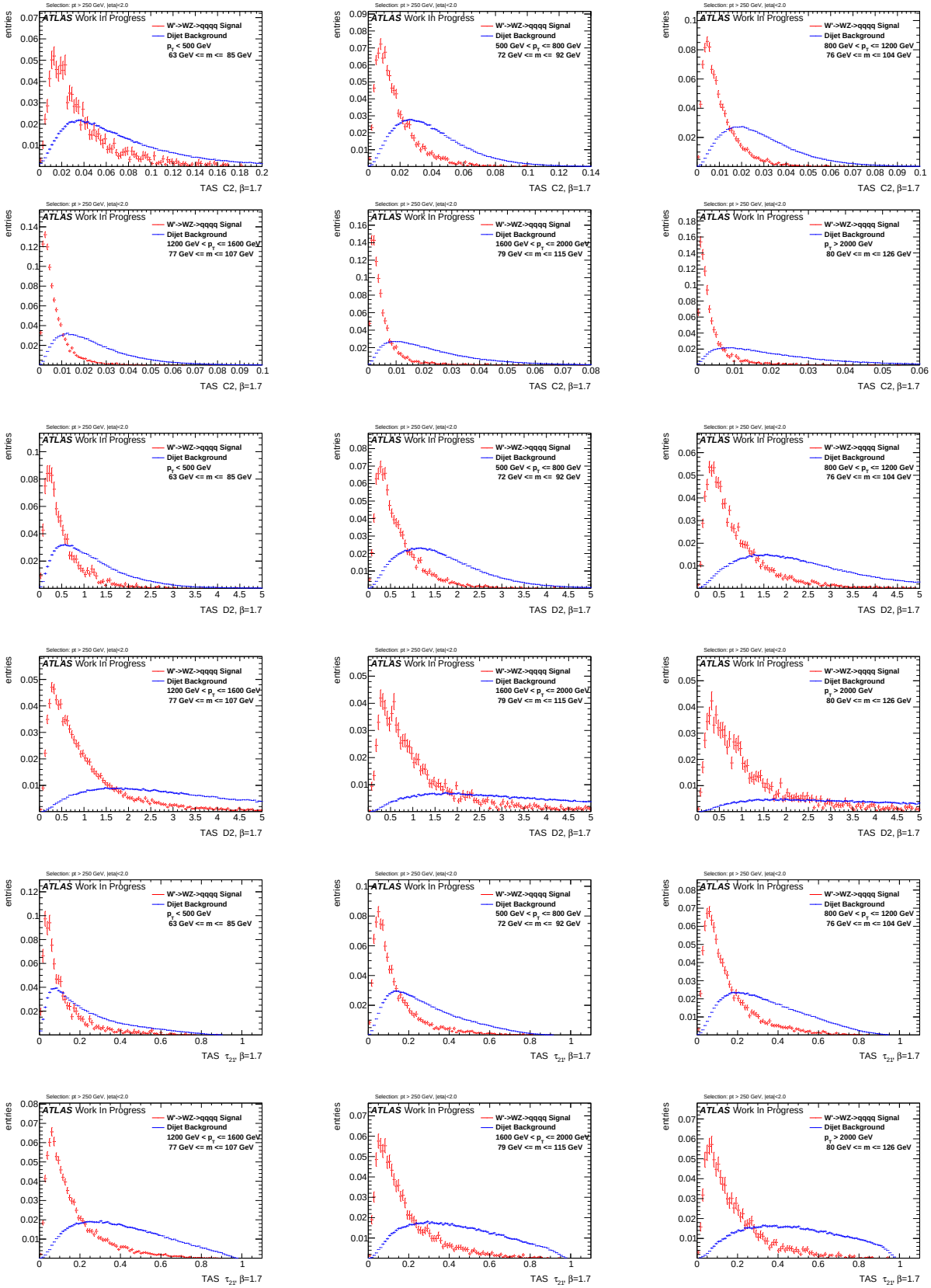

Figure C.7.: Distributions for W boson tagging using TAS $\beta = 1.7$. C2, D2, τ_{21} top down.

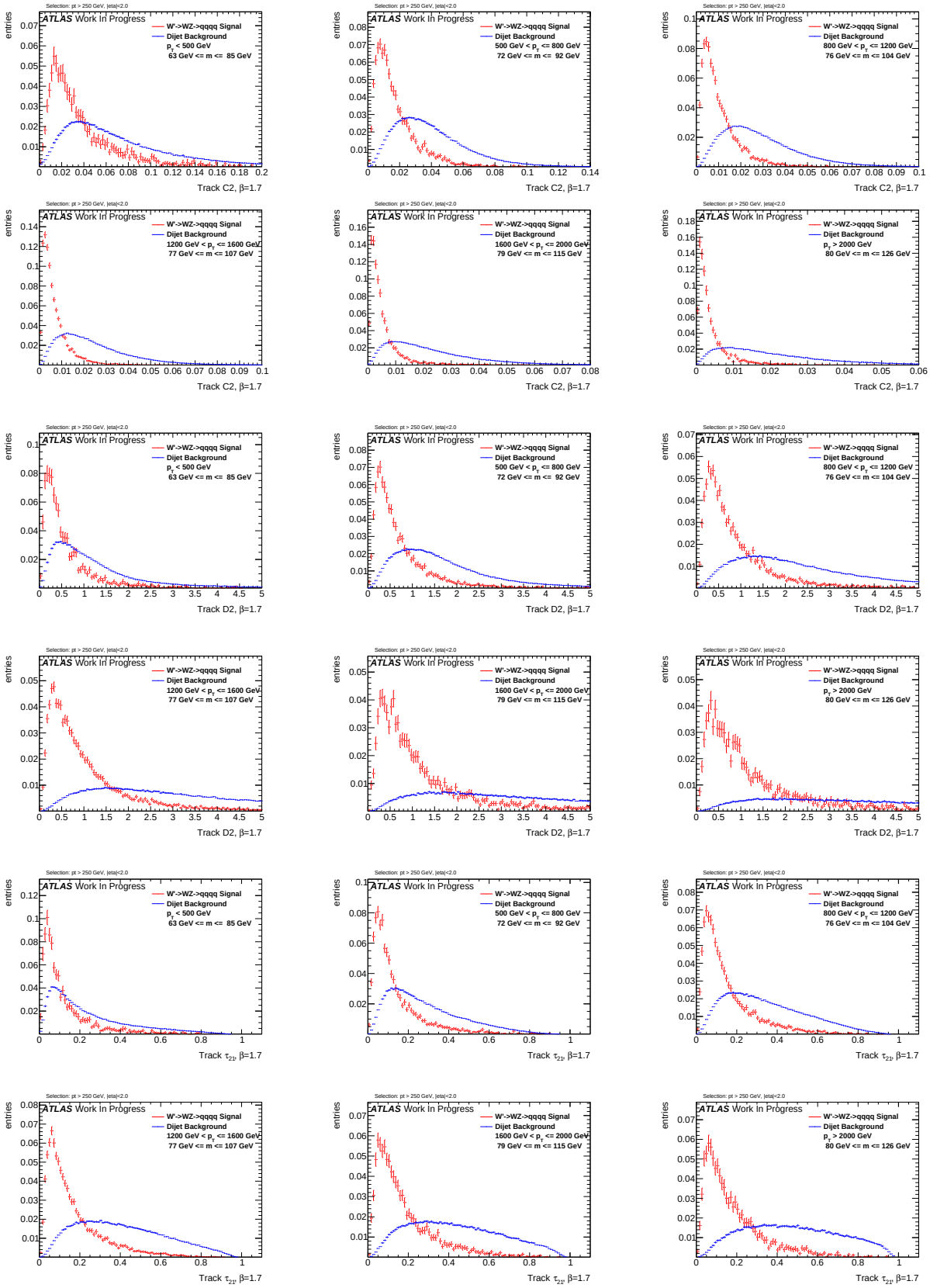

Figure C.8.: Distributions for W boson tagging using tracks $\beta = 1.7$. C2, D2, τ_{21} top down.

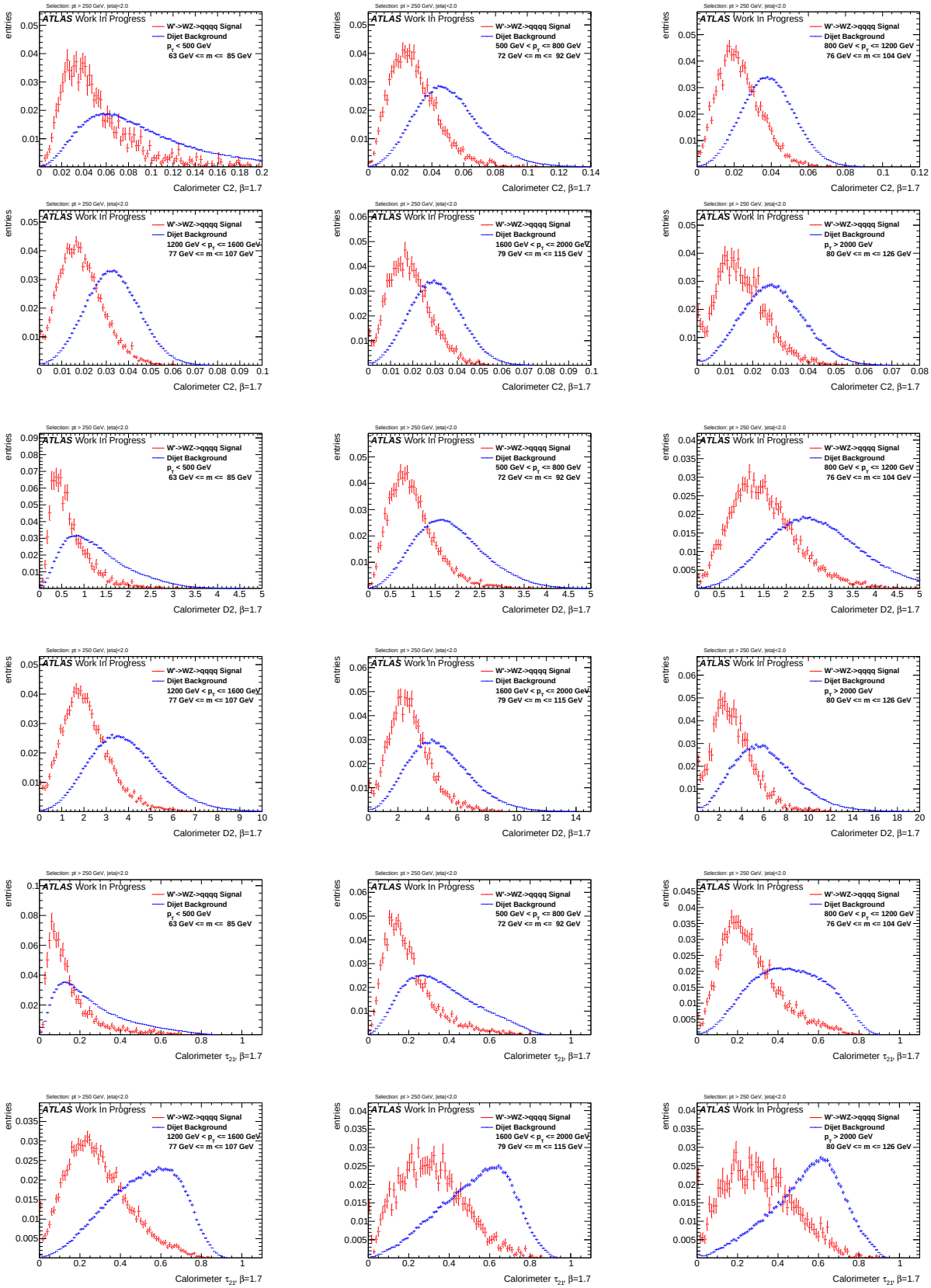

Figure C.9.: Distributions for W boson tagging using calorimeter clusters $\beta = 1.7$. C2, D2, τ_{21} top down.

$\beta = 2$

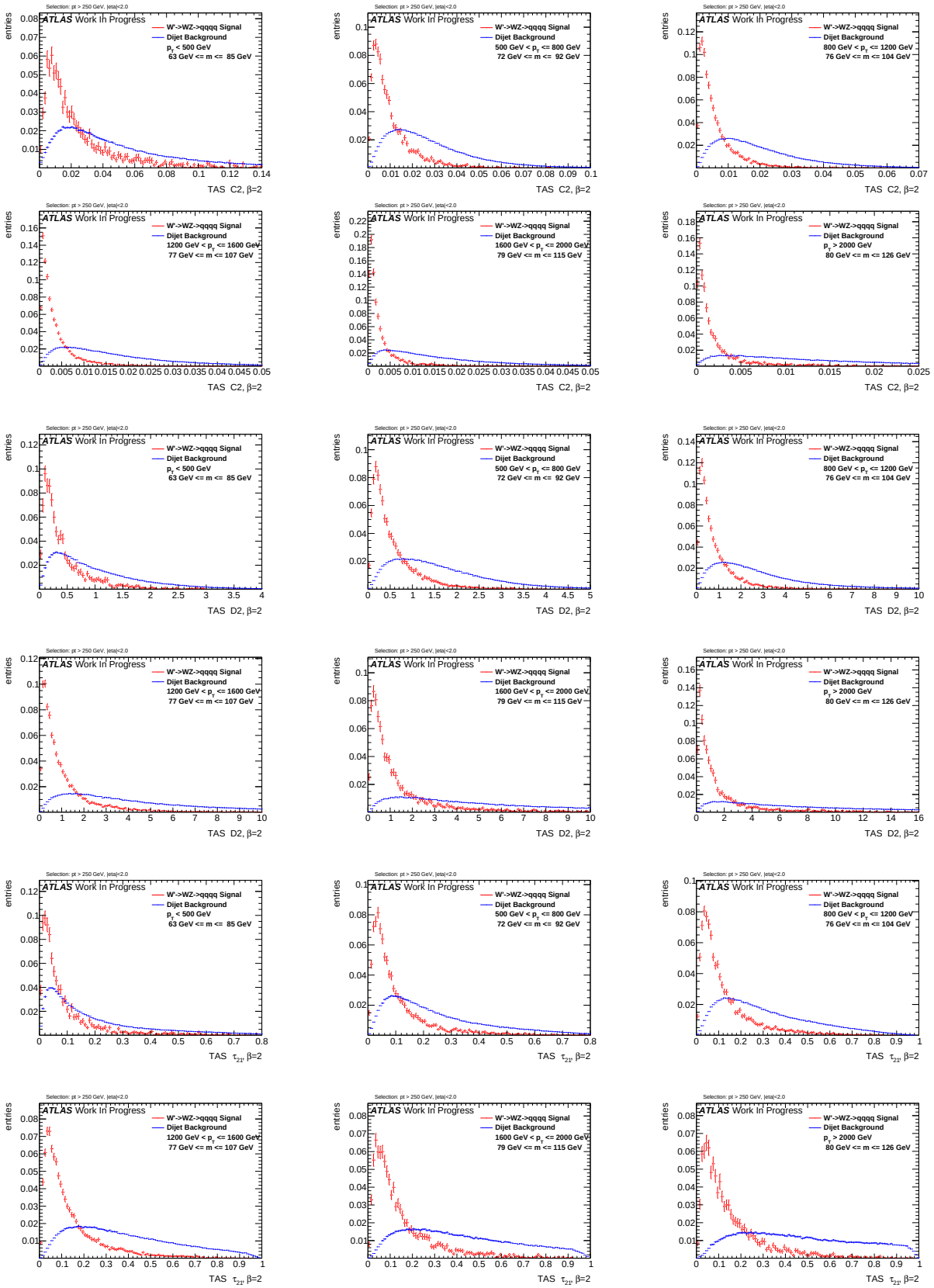

Figure C.10.: Distributions for W boson tagging using TAS $\beta = 2$. C2, D2, τ_{21} top down.

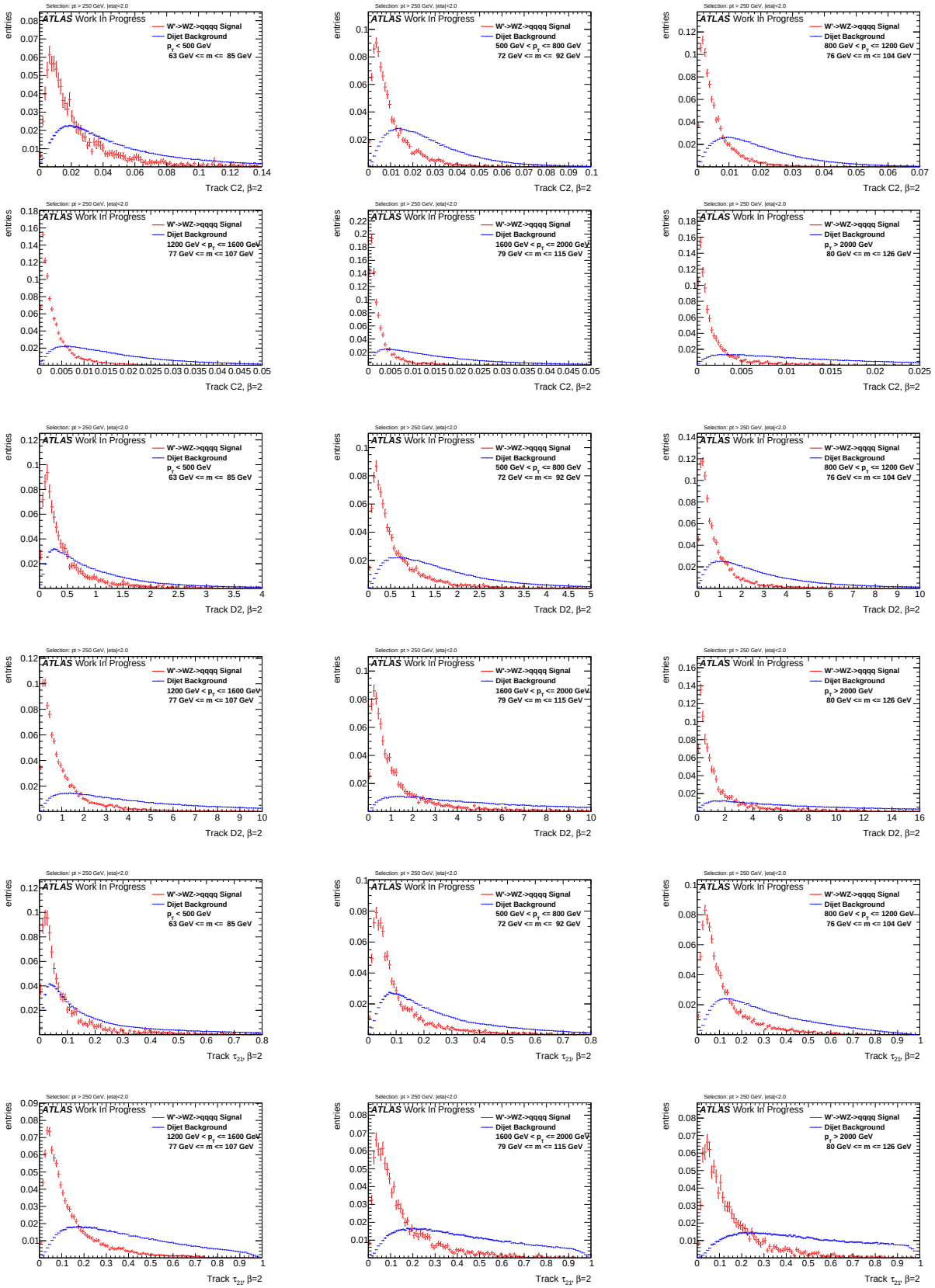

Figure C.11.: Distributions for W boson tagging using tracks $\beta = 2$. C2, D2, τ_{21} top down.

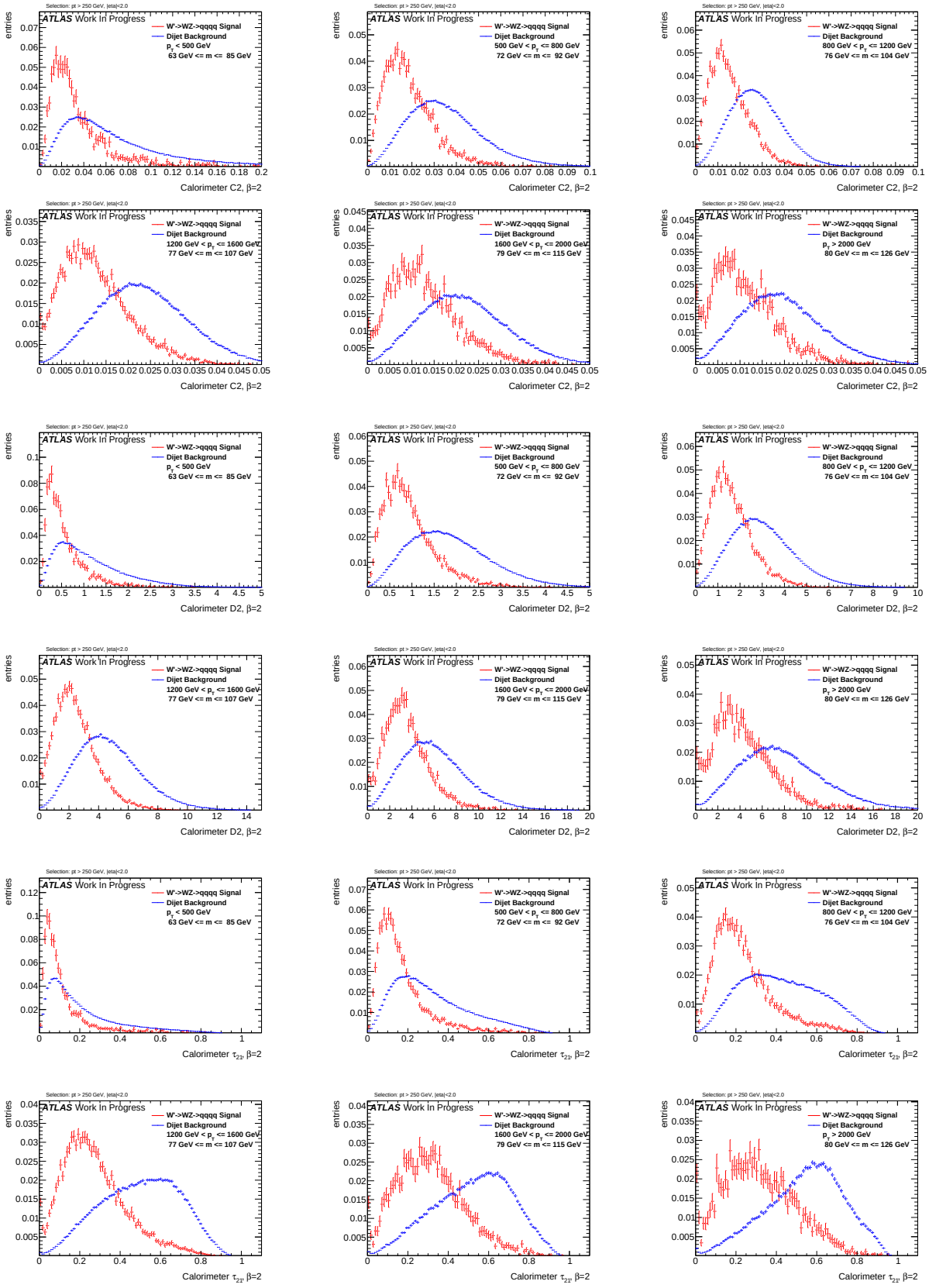

Figure C.12.: Distributions for W boson tagging using calorimeter clusters $\beta = 2$. C2, D2, τ_{21} top down.

$\beta = 3$

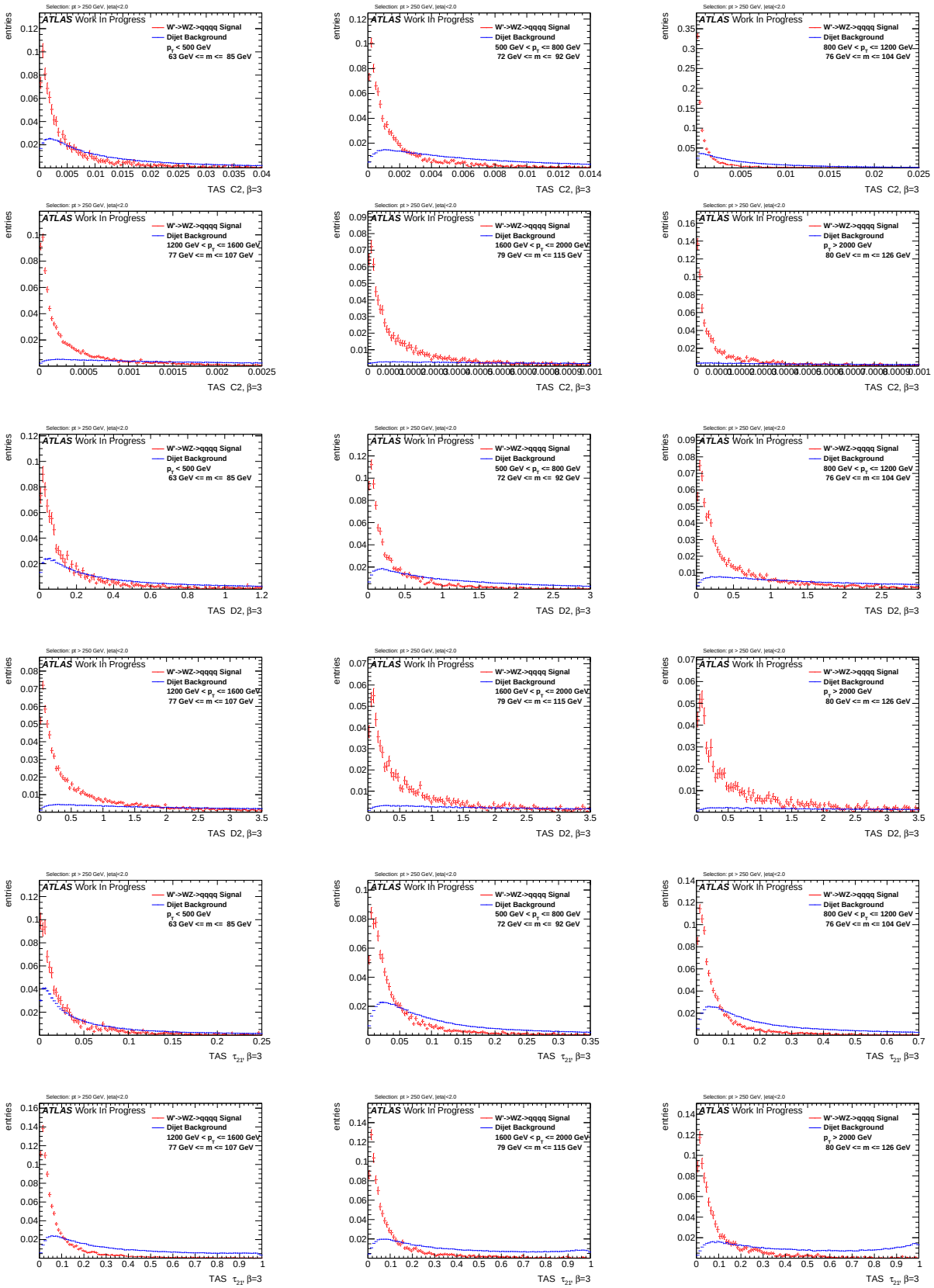

Figure C.13.: Distributions for W boson tagging using TAS $\beta = 3$. C2, D2, τ_{21} top down.

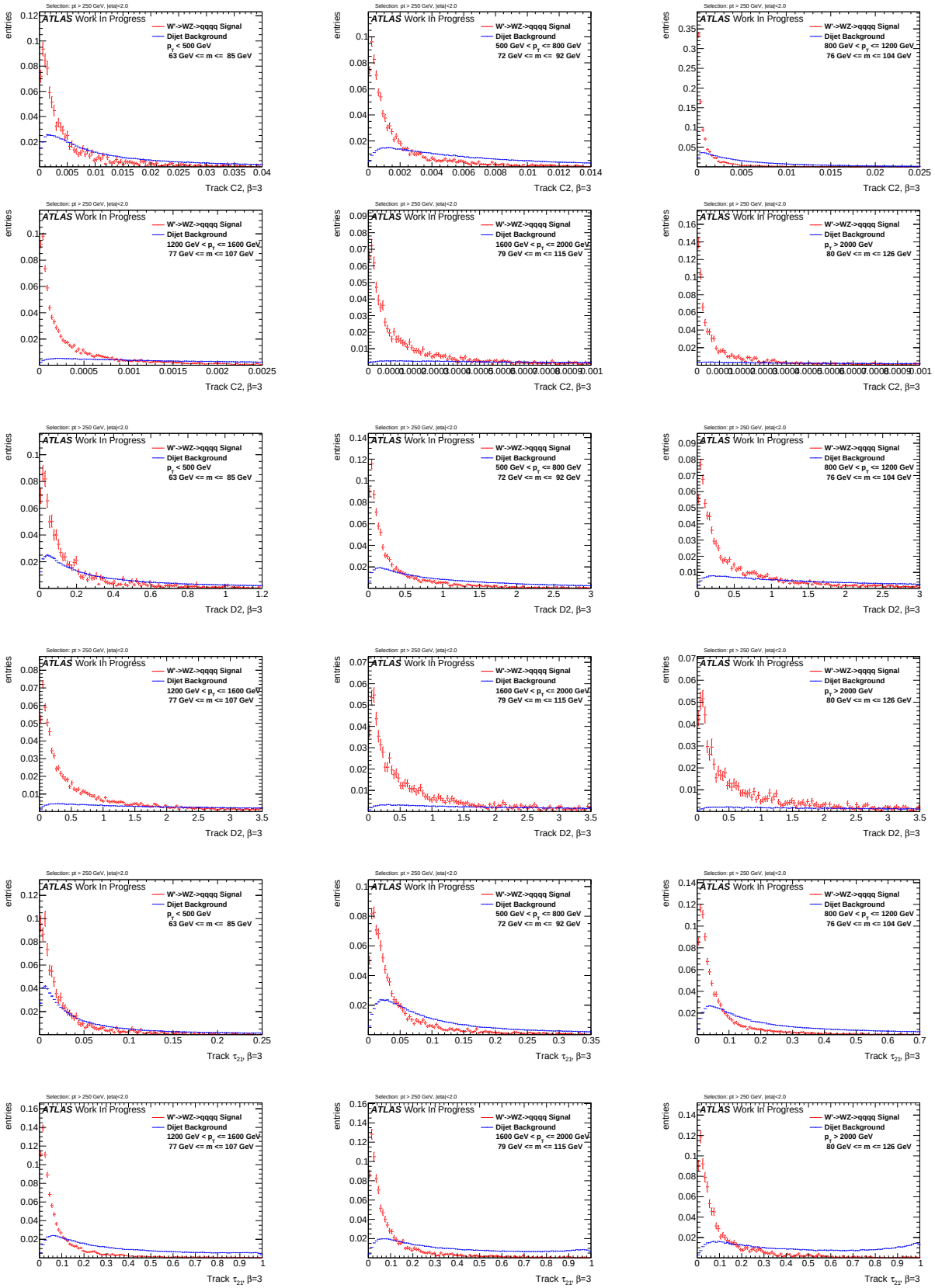

Figure C.14.: Distributions for W boson tagging using tracks $\beta = 3$. C2, D2, τ_{21} top down.

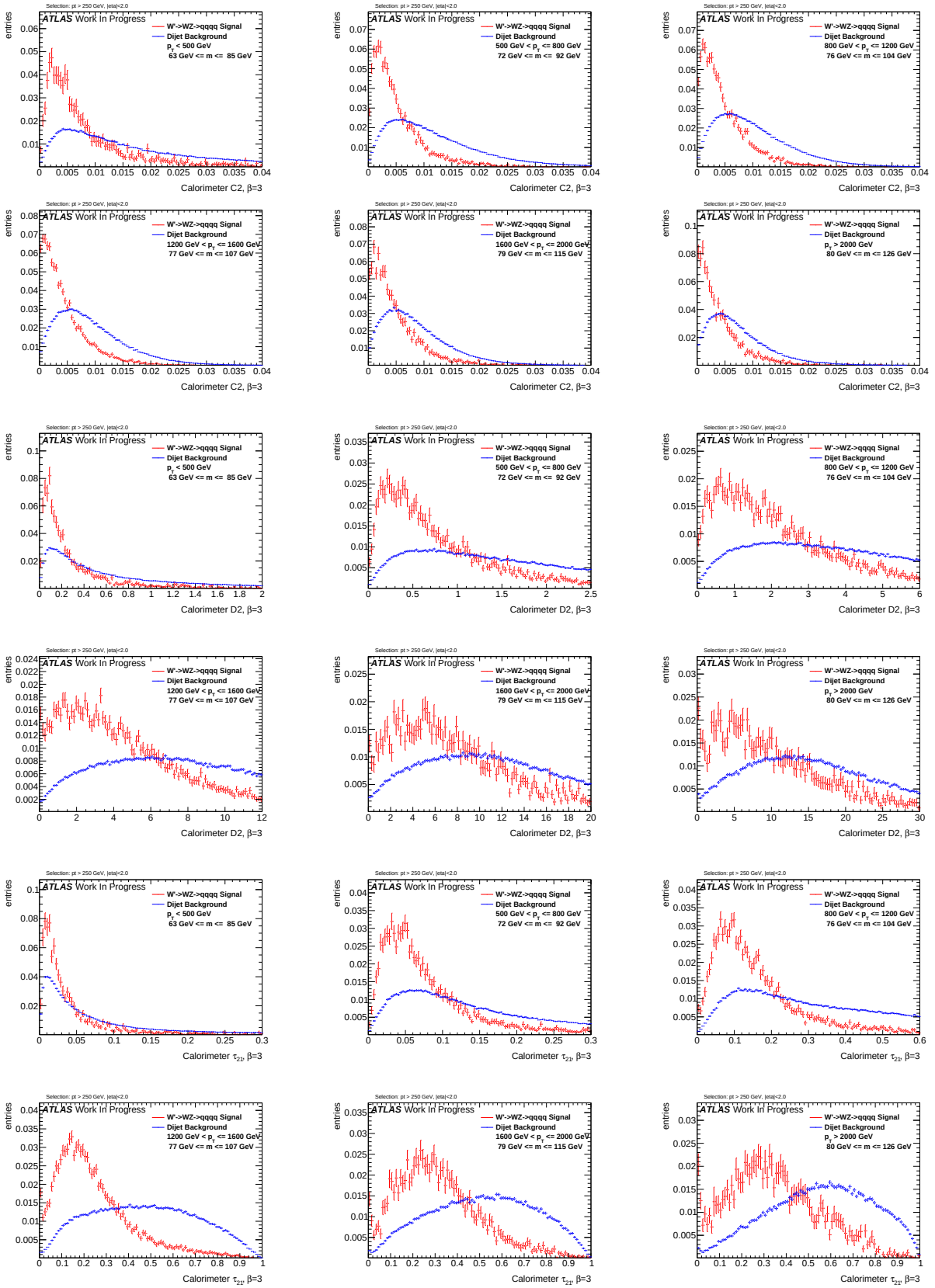

Figure C.15.: Distributions for W boson tagging using calorimeter clusters $\beta = 3$. C2, D2, τ_{21} top down.

C.2. Higgs Distributions

$\beta = 0.5$

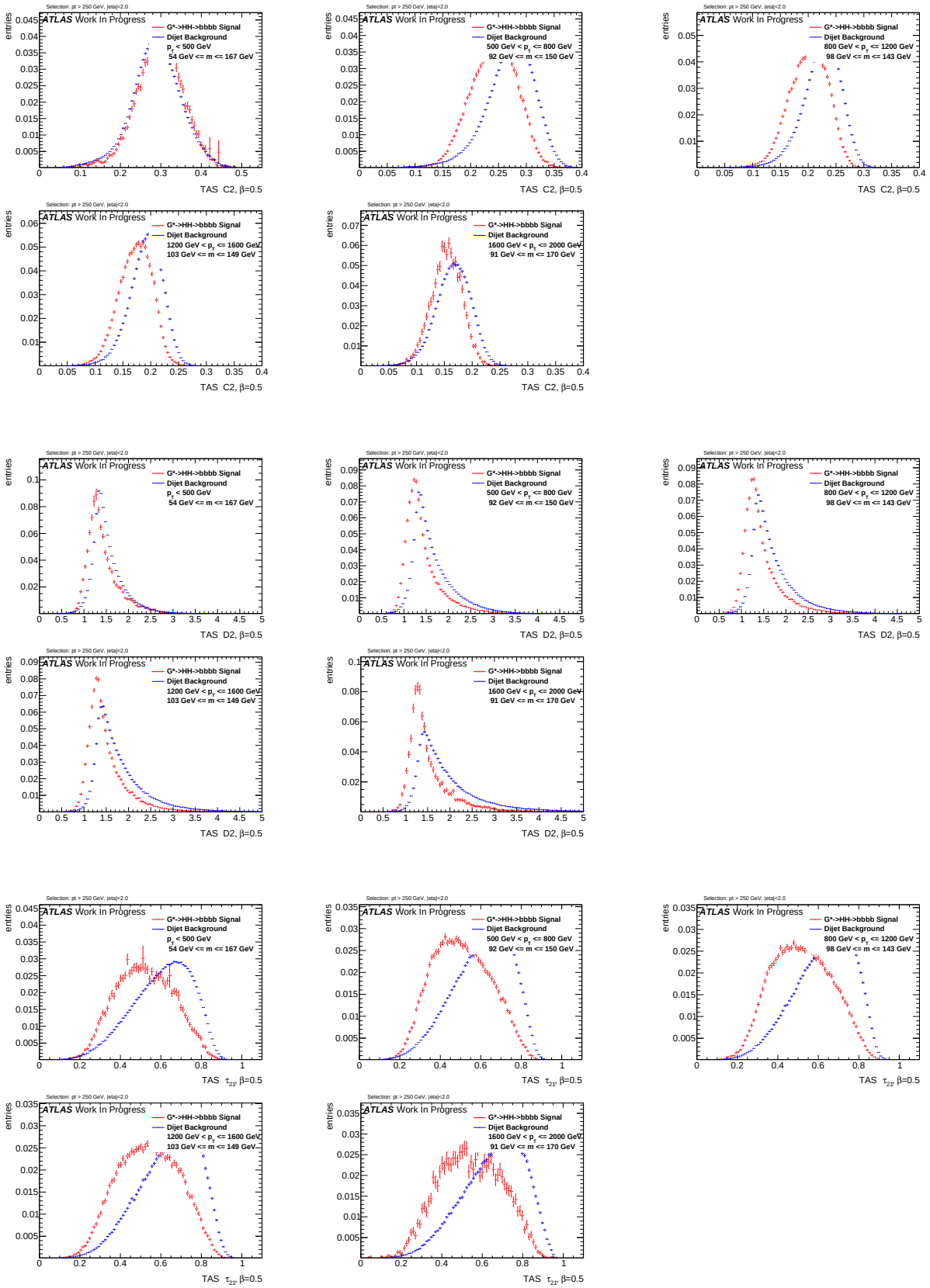

Figure C.16.: Distributions for Higgs boson tagging using TAS $\beta = 0.5$. C2, D2, τ_{21} top down.

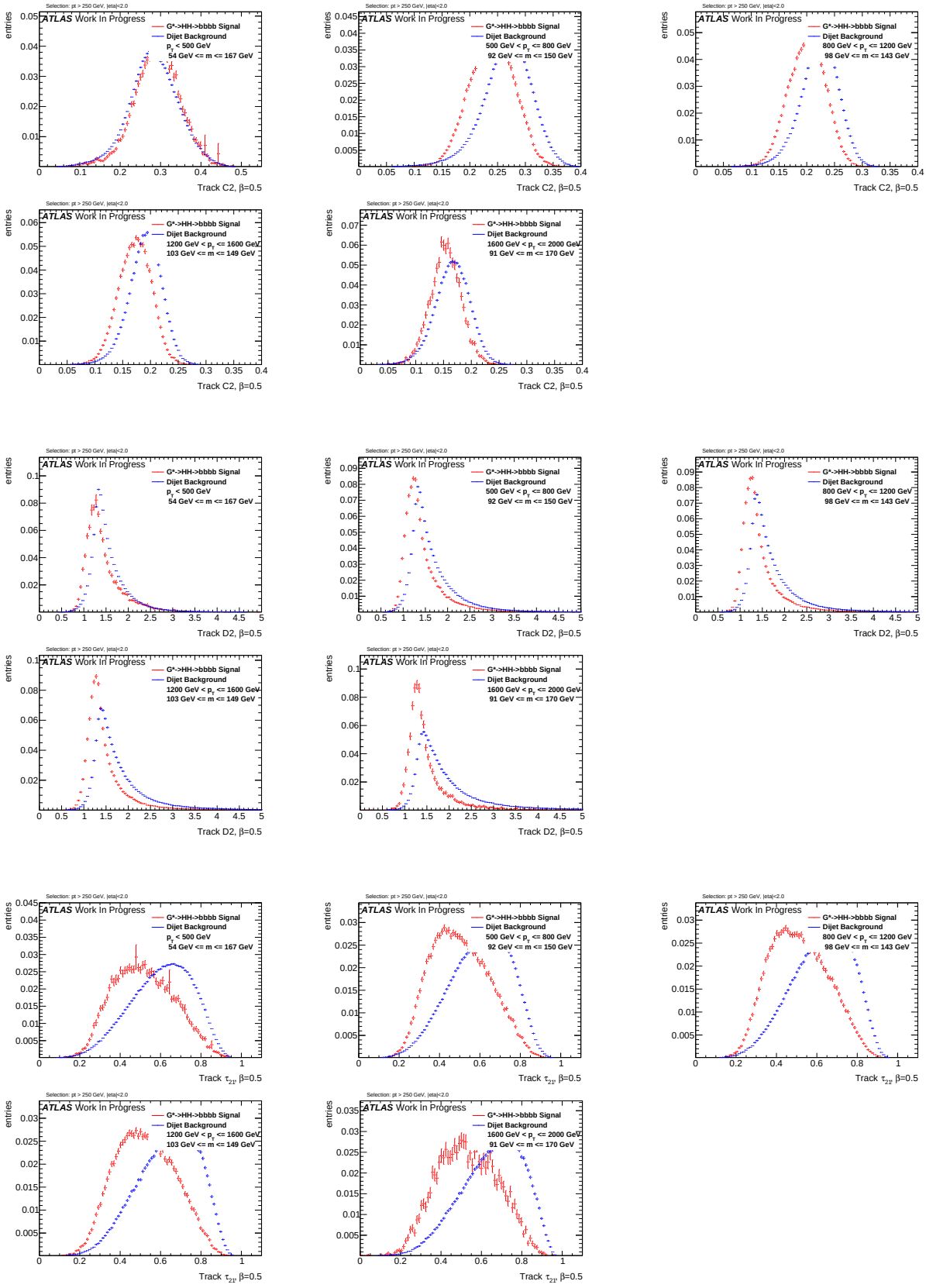

Figure C.17.: Distributions for Higgs boson tagging using tracks $\beta = 0.5$. C2, D2, τ_{21} top down.

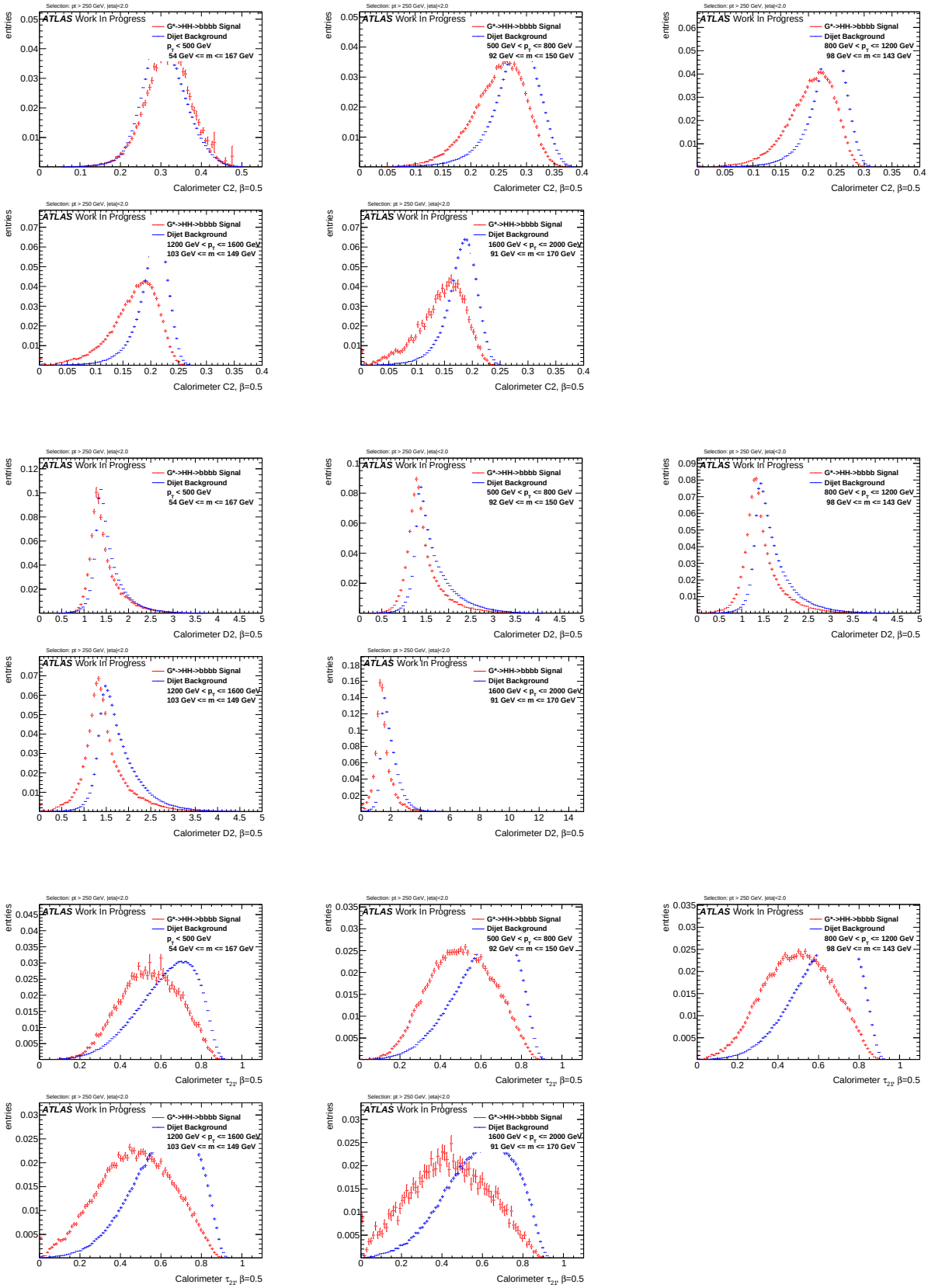

Figure C.18.: Distributions for Higgs boson tagging using calorimeter clusters $\beta = 0.5$. C2, D2, τ_{21} top down.

$\beta = 1$

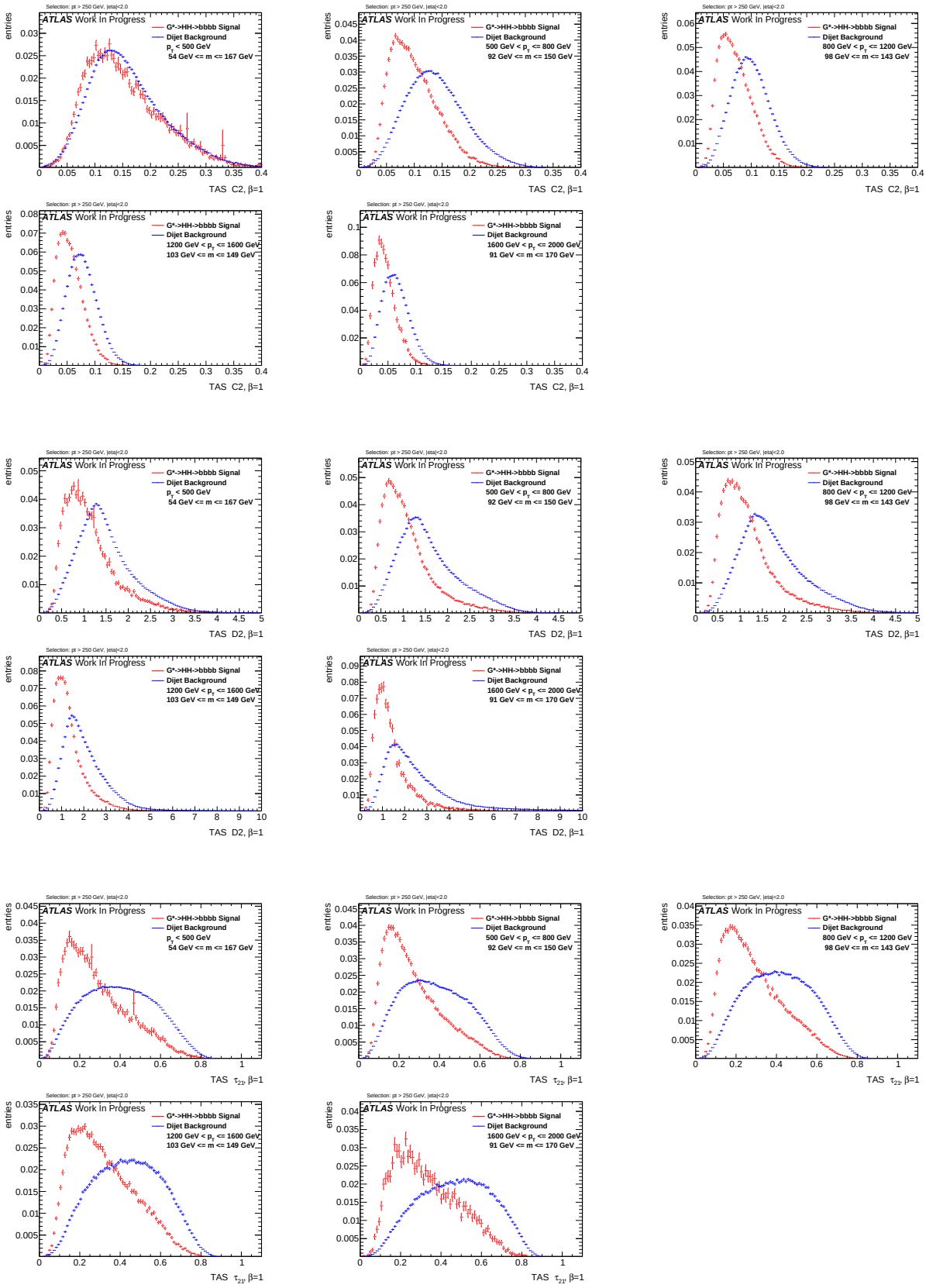

Figure C.19.: Distributions for Higgs boson tagging using TAS $\beta = 1$. C2, D2, τ_{21} top down.

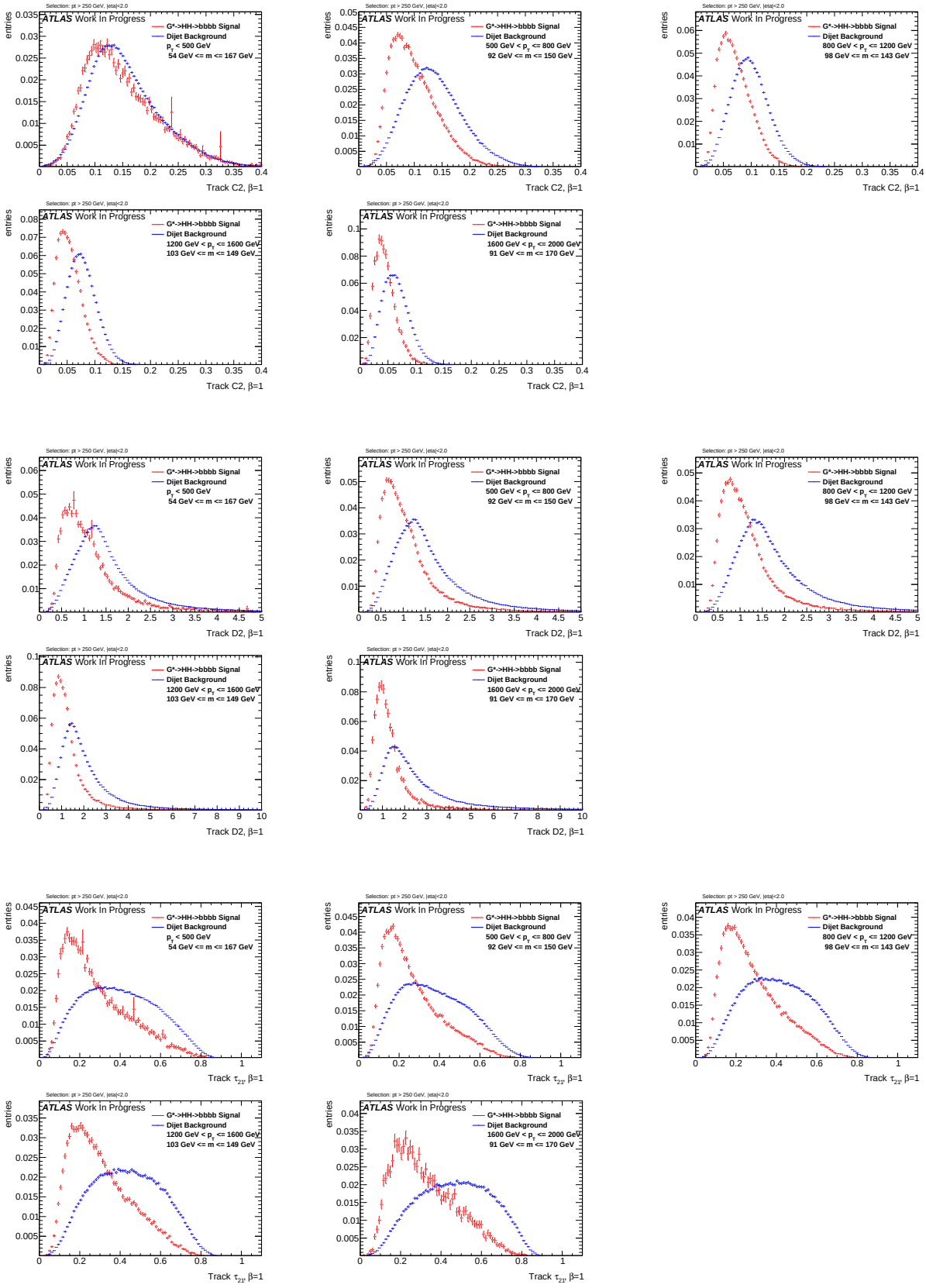

Figure C.20.: Distributions for Higgs boson tagging using tracks $\beta = 1$. C2, D2, τ_{21} top down.

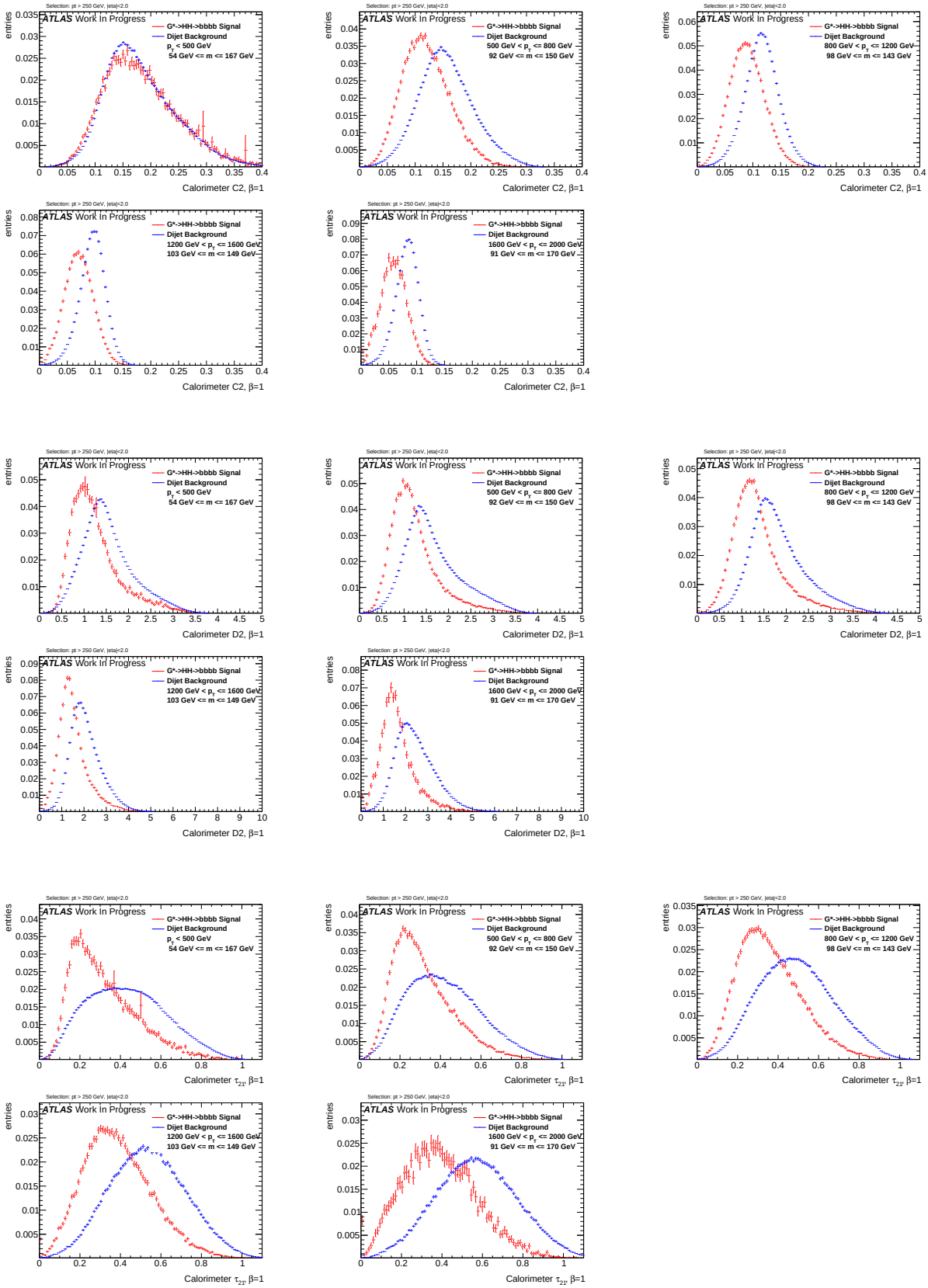

Figure C.21.: Distributions for Higgs boson tagging using calorimeter clusters $\beta = 1$. C2, D2, τ_{21} top down.

$\beta = 1.7$

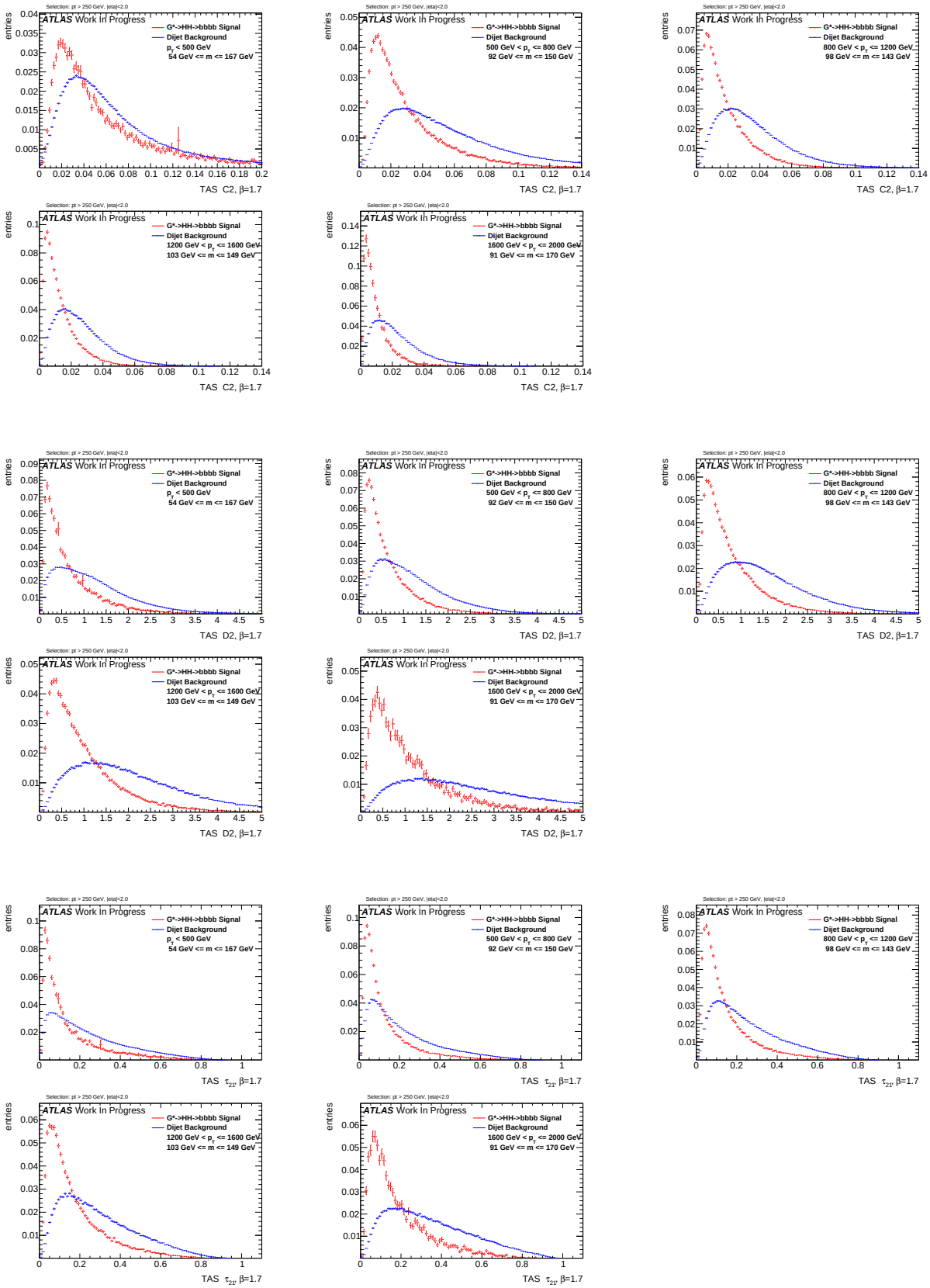

Figure C.22.: Distributions for Higgs boson tagging using TAS $\beta = 1.7$. C2, D2, τ_{21} top down.

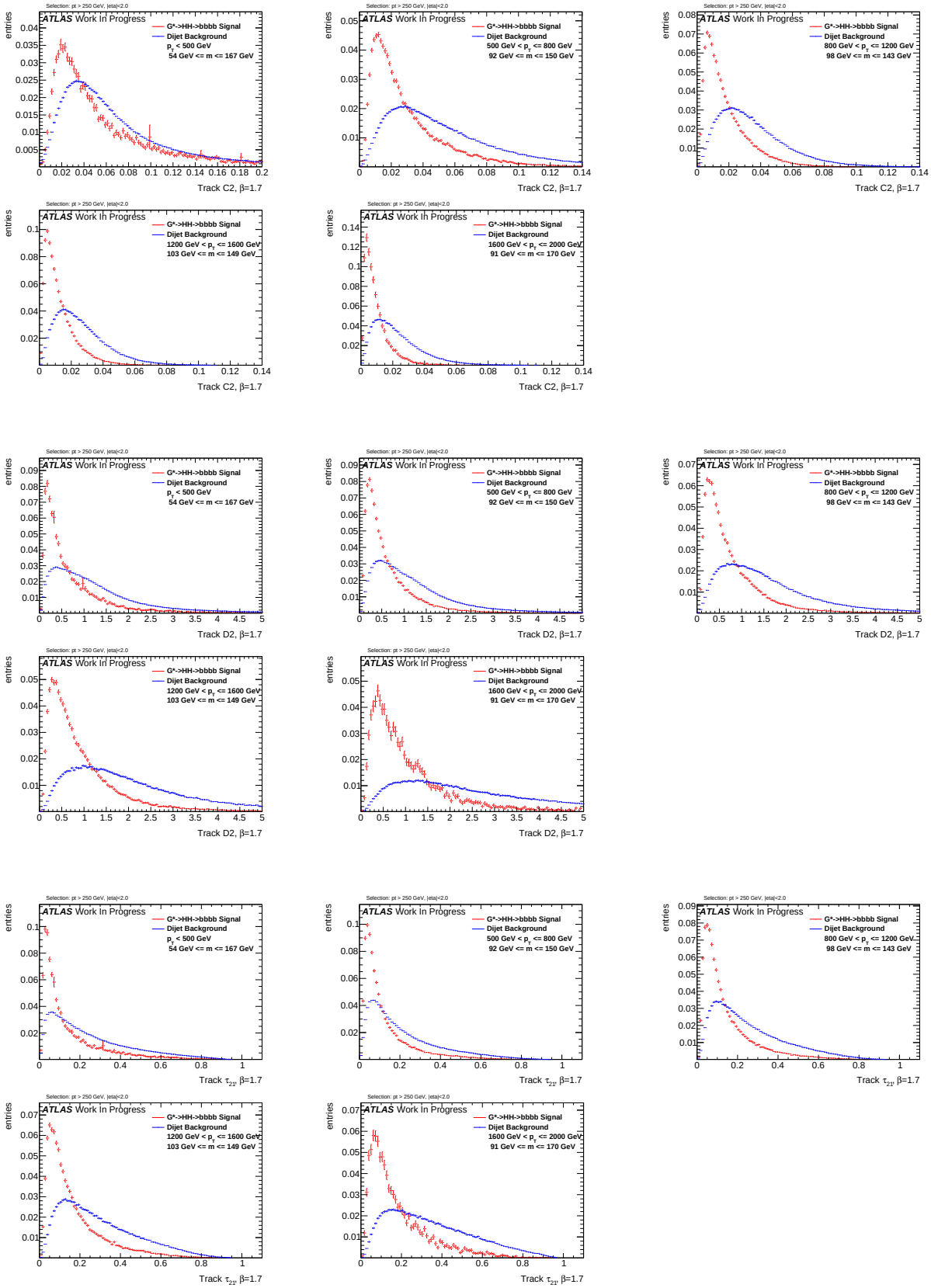

Figure C.23.: Distributions for Higgs boson tagging using tracks $\beta = 1.7$. C2, D2, τ_{21} top down.

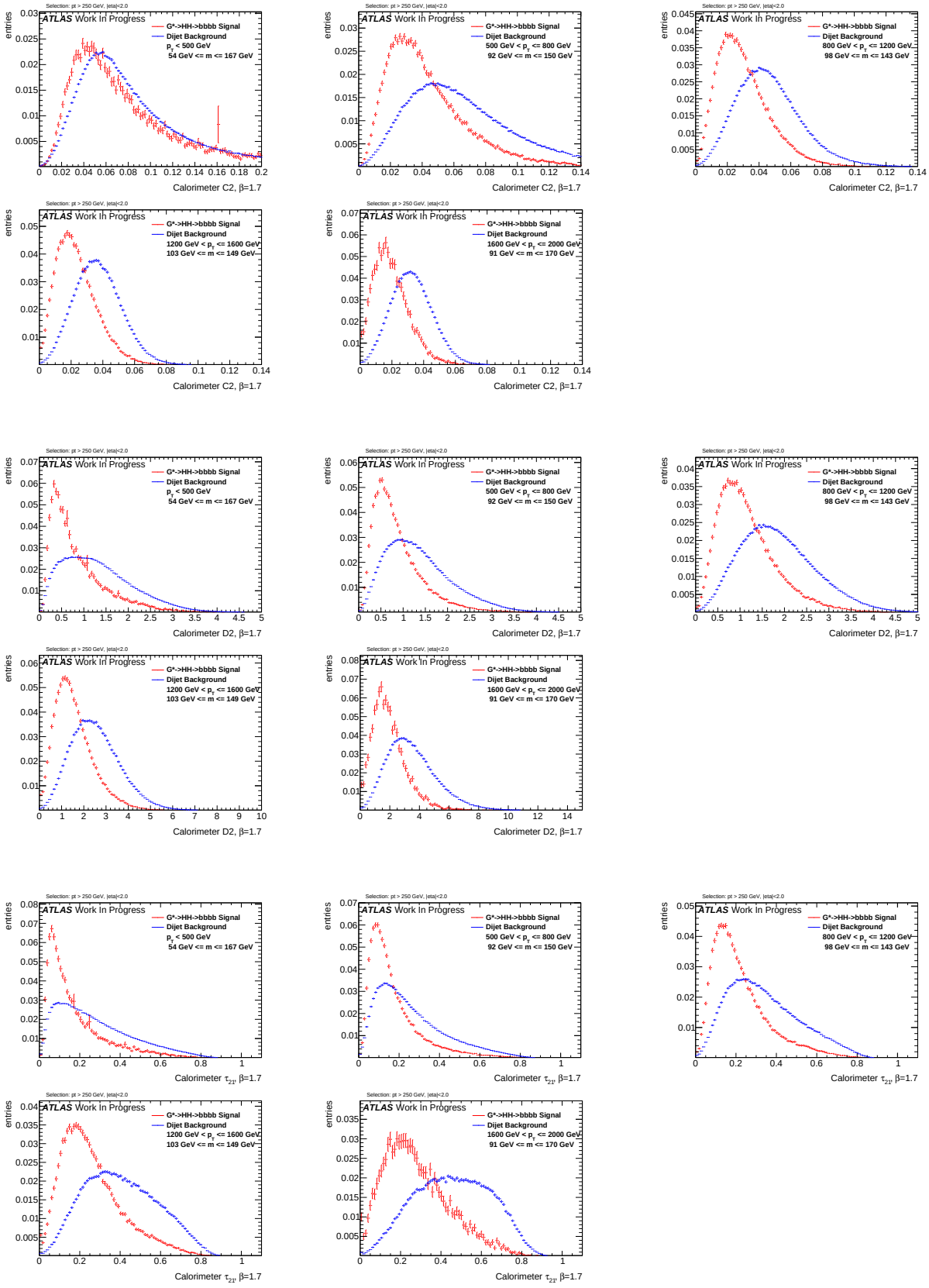

Figure C.24.: Distributions for Higgs boson tagging using calorimeter clusters $\beta = 1.7$. C2, D2, τ_{21} top down.

$\beta = 2$

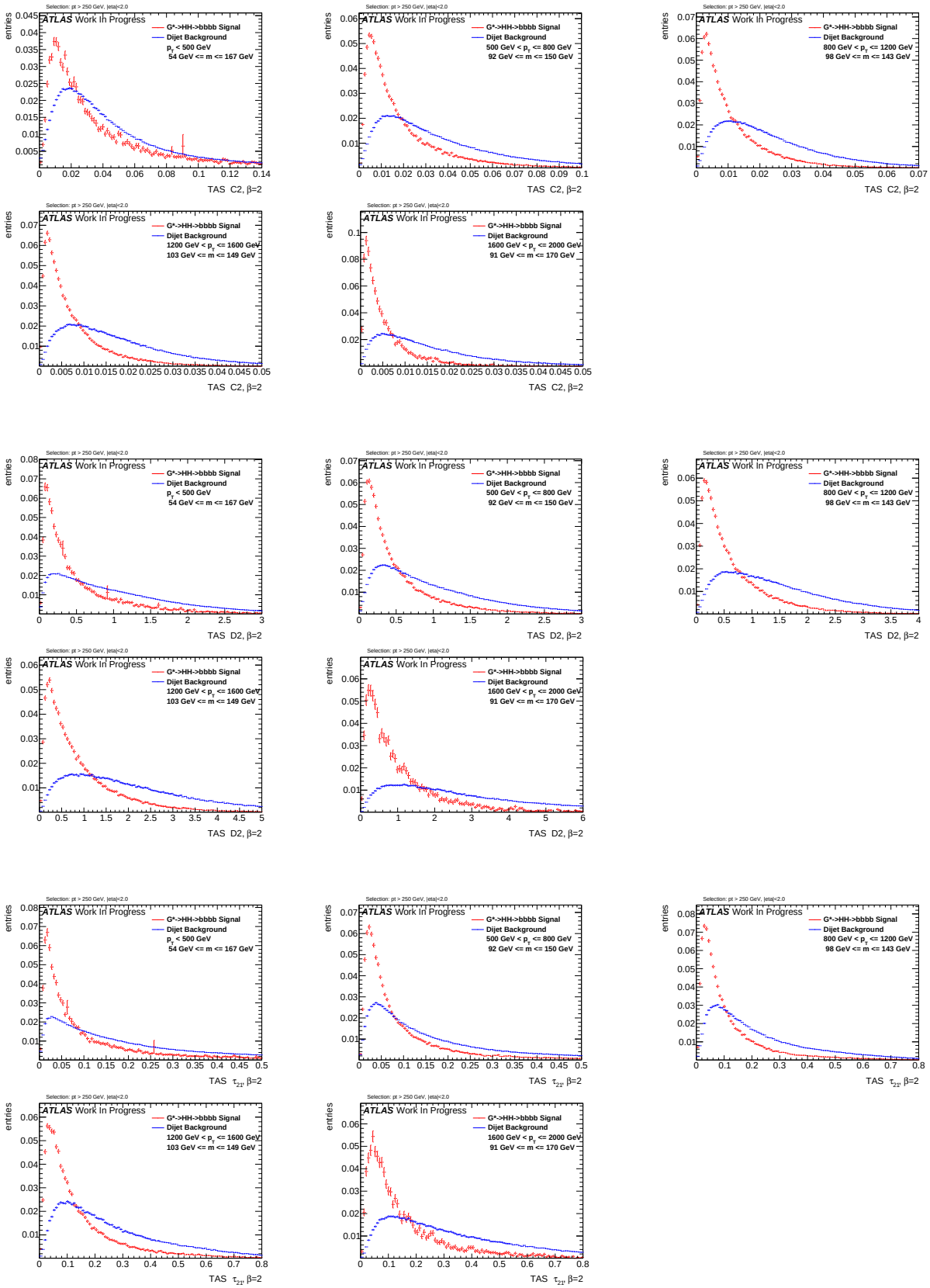

Figure C.25.: Distributions for Higgs boson tagging using TAS $\beta = 2$. C2, D2, τ_{21} top down.

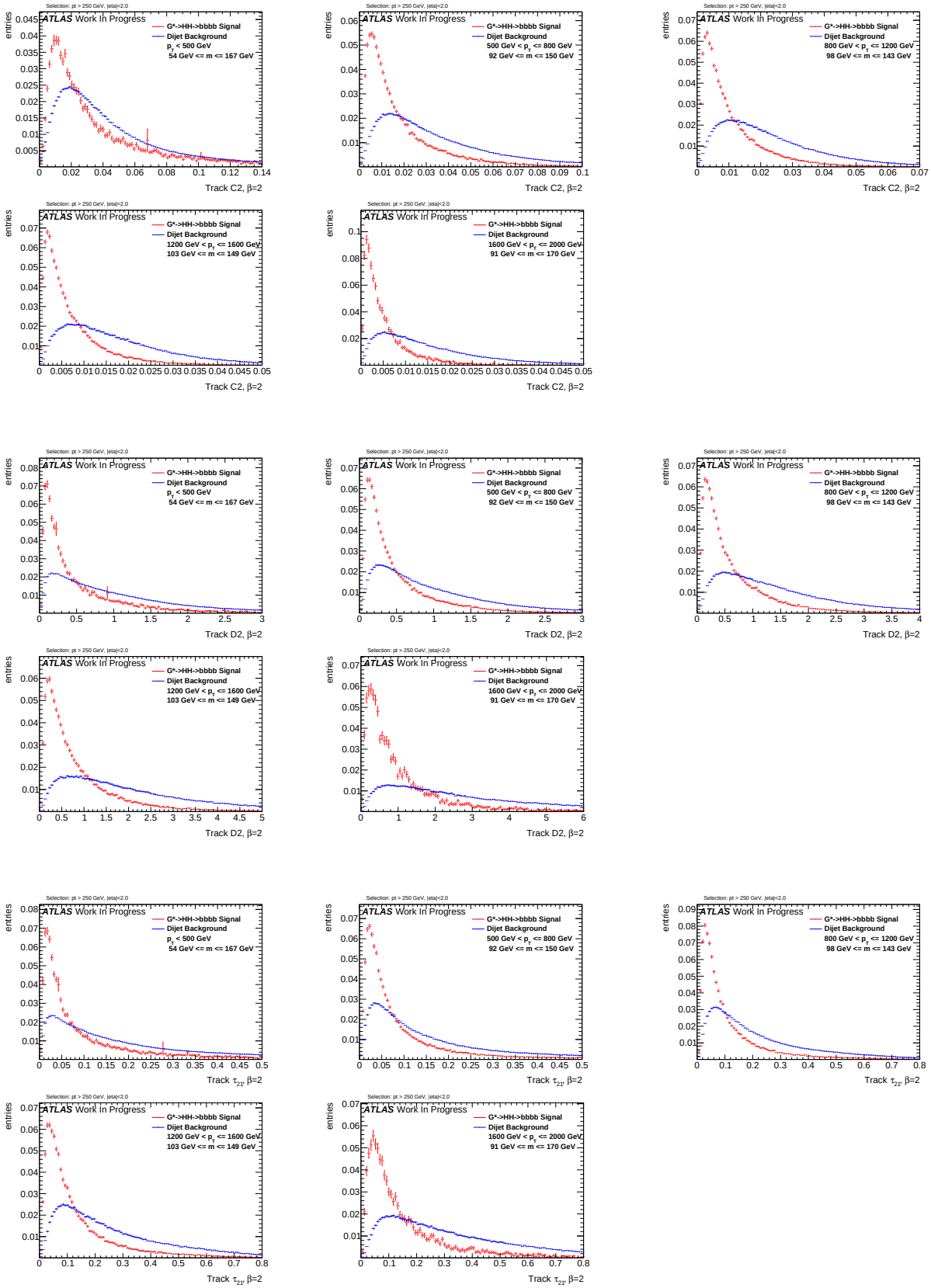

Figure C.26.: Distributions for Higgs boson tagging using tracks $\beta = 2$. C2, D2, τ_{21} top down.

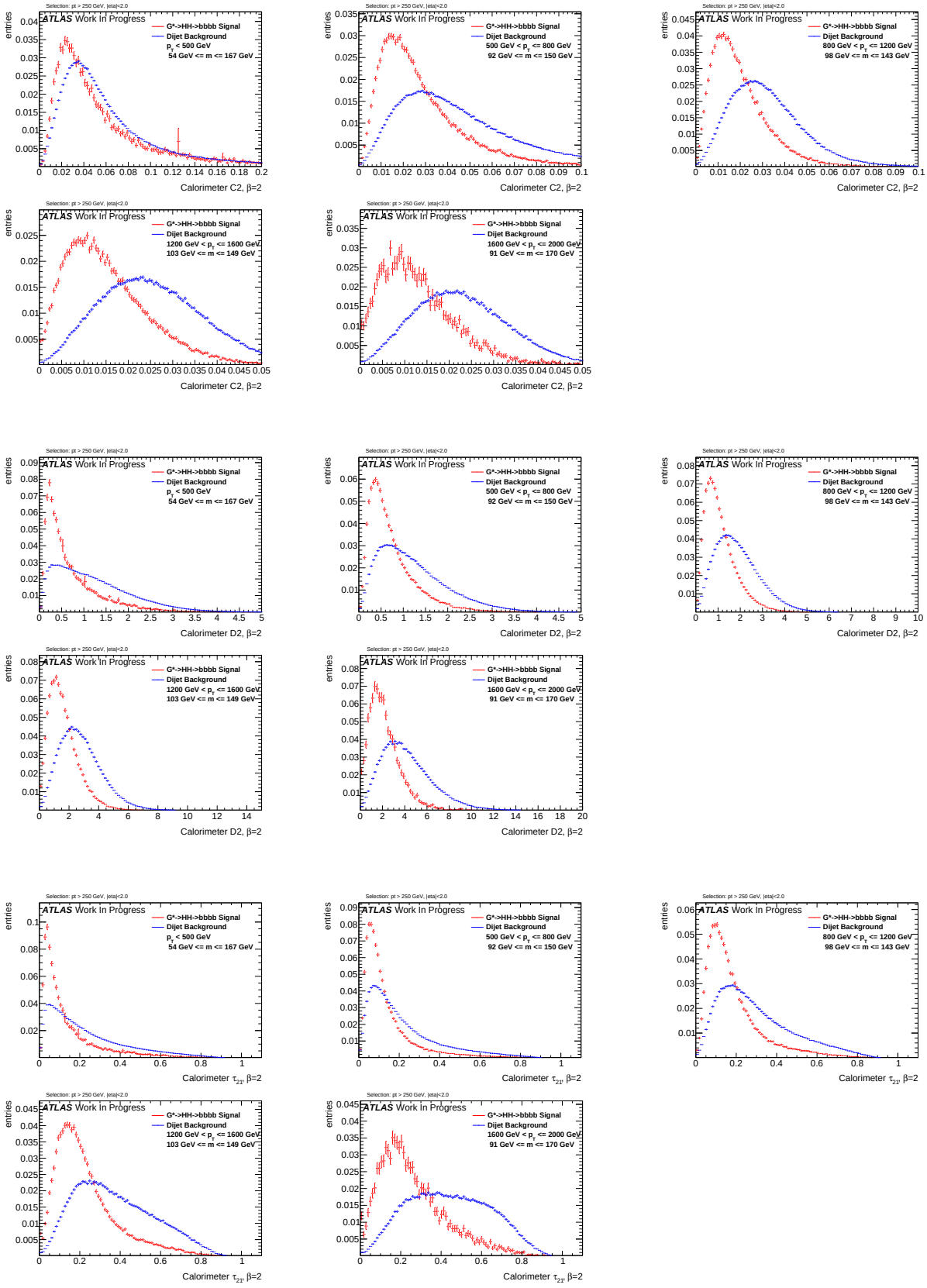

Figure C.27.: Distributions for Higgs boson tagging using calorimeter clusters $\beta = 2$. C2, D2, τ_{21} top down.

$\beta = 3$

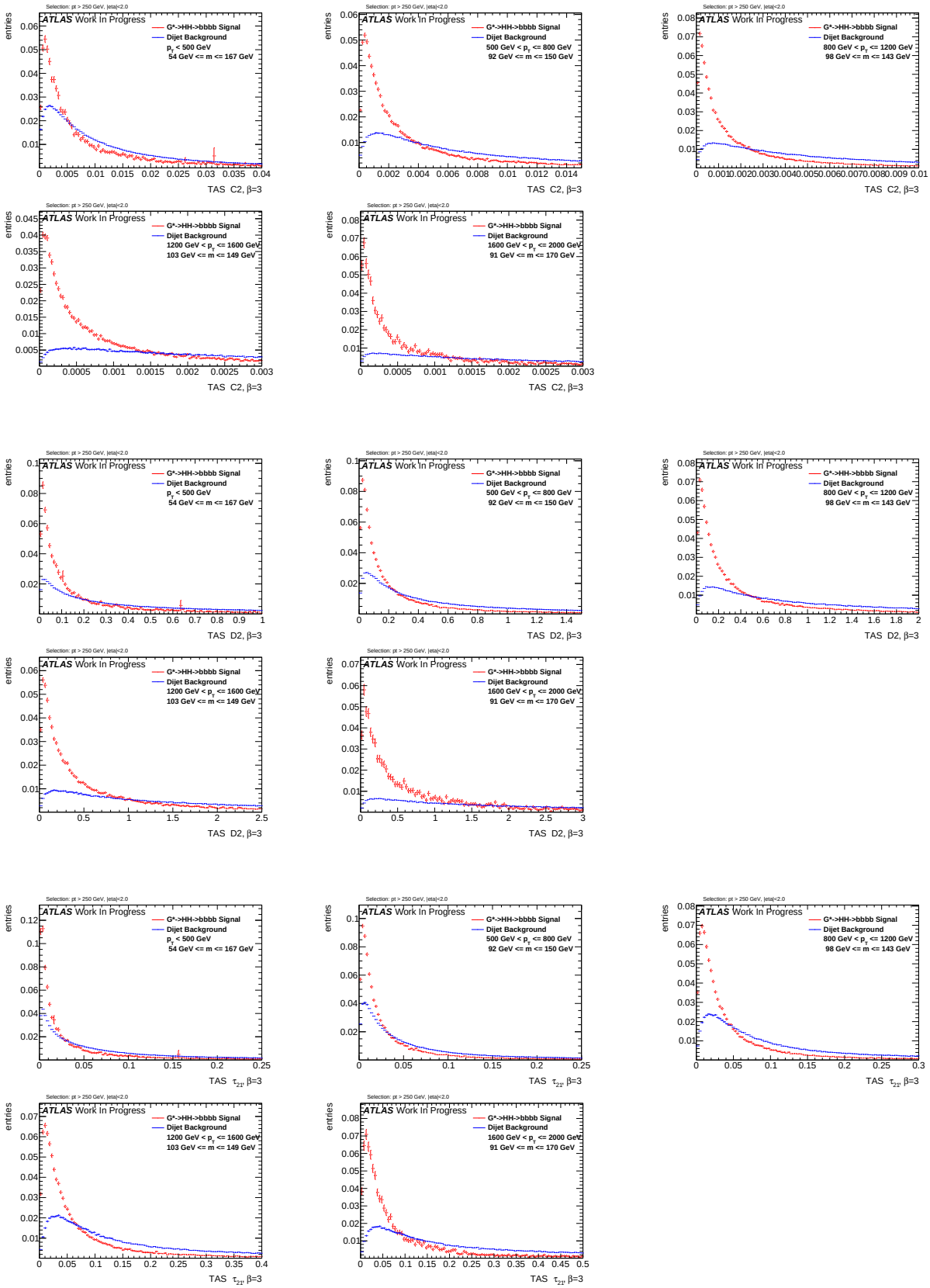

Figure C.28.: Distributions for Higgs boson tagging using TAS $\beta = 3$. C2, D2, τ_{21} top down.

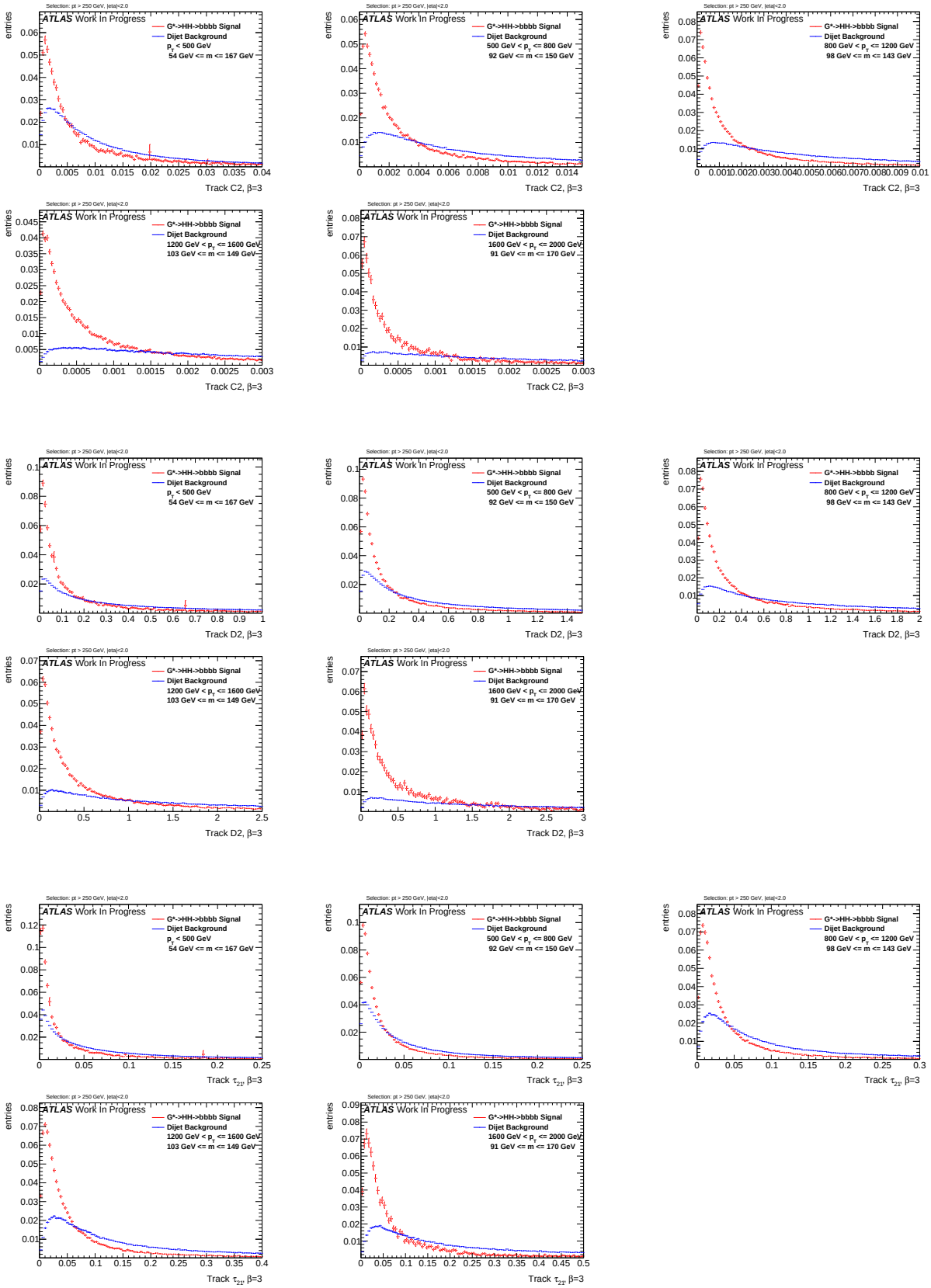

Figure C.29.: Distributions for Higgs boson tagging using tracks $\beta = 3$. C2, D2, τ_{21} top down.

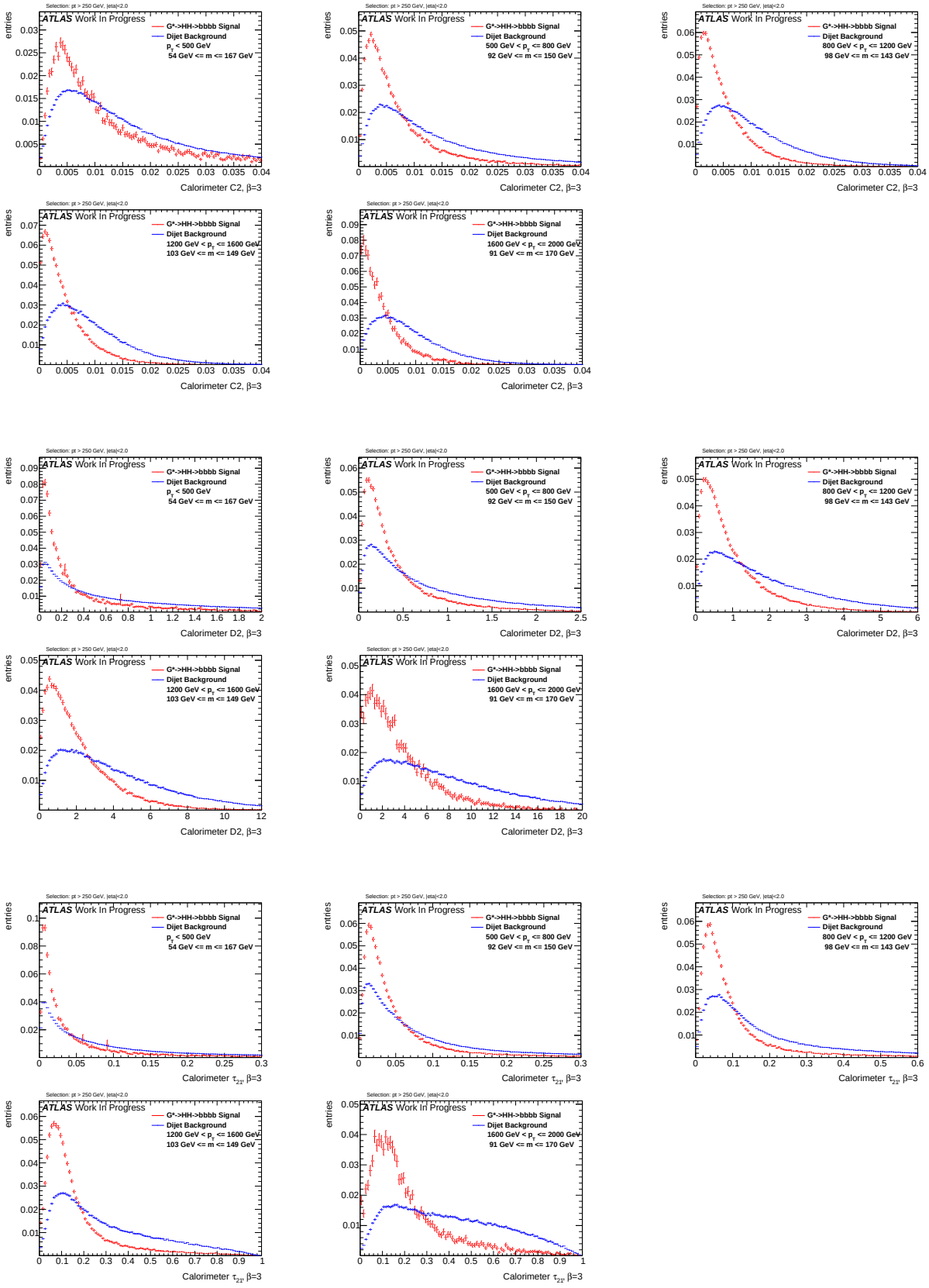

Figure C.30.: Distributions for Higgs boson tagging using calorimeter clusters $\beta = 3$. C2, D2, τ_{21} top down.

C.3. Top Distributions

$\beta = 1$

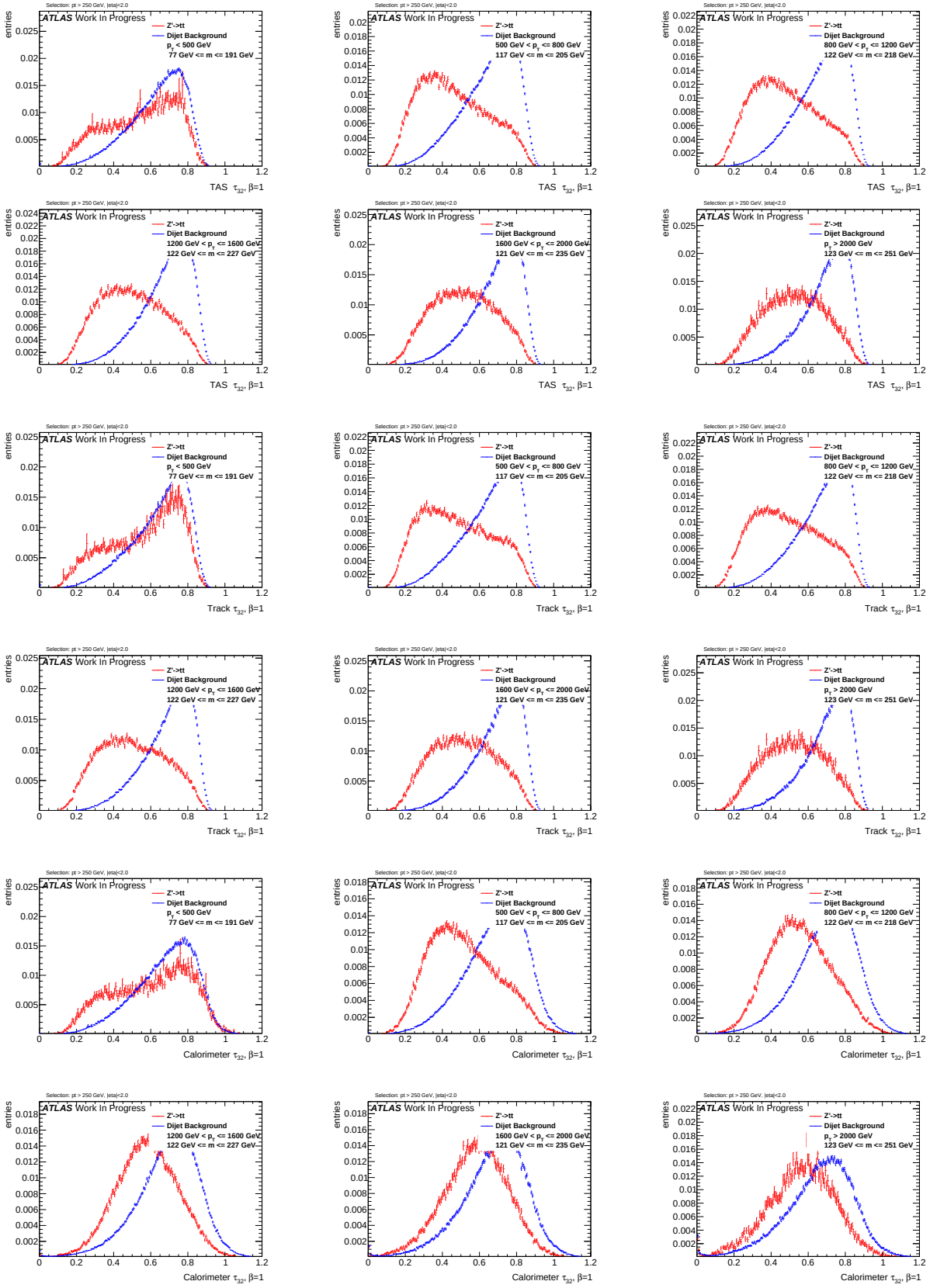

Figure C.31.: Distributions for Top tagging using τ_{32} ($\beta = 1$) using TAS, tracks and calorimeter clusters top down.

$\beta = 1.7$

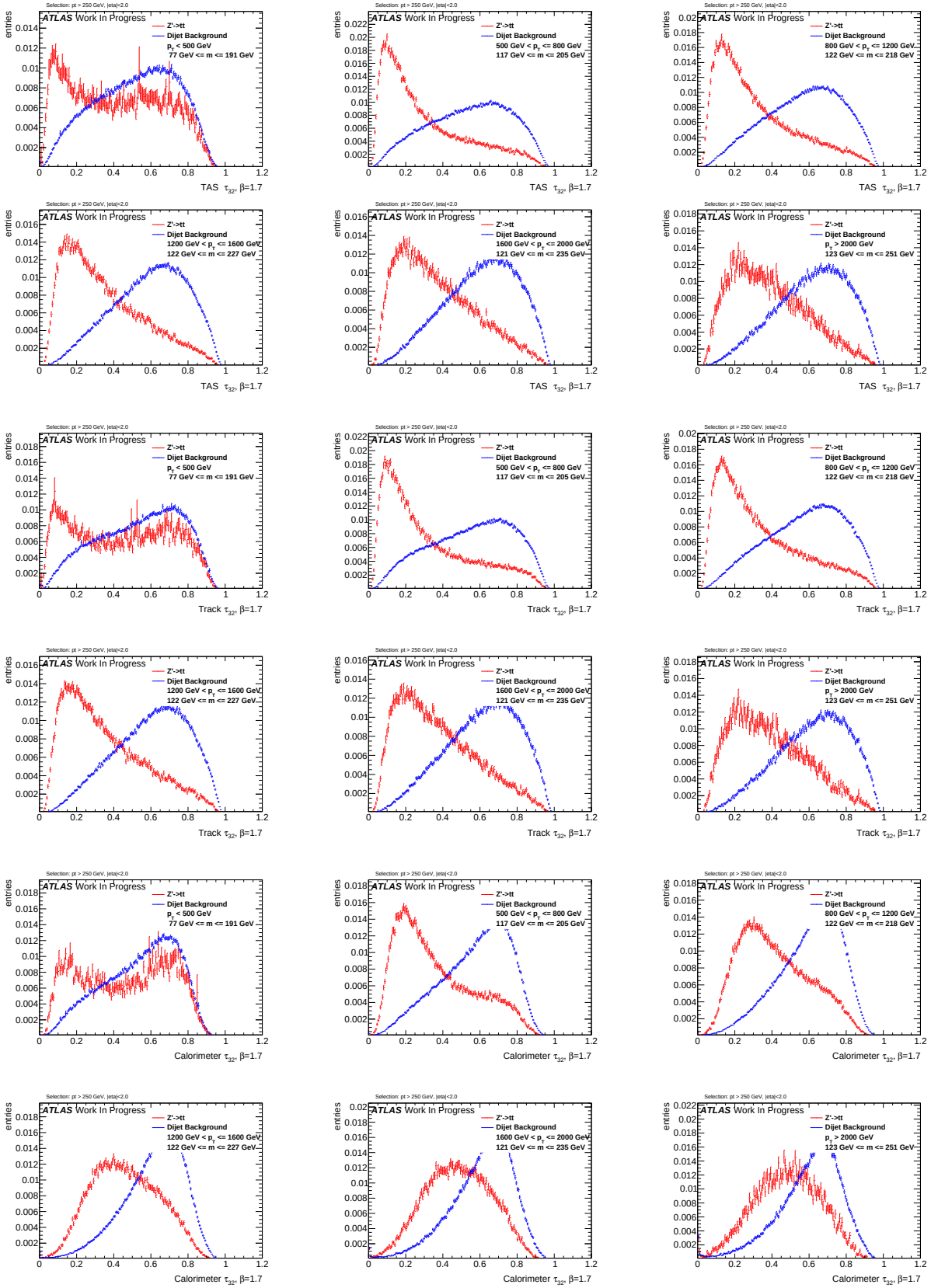

Figure C.32.: Distributions for Top tagging using τ_{32} ($\beta = 1.7$) using TAS, tracks and calorimeter clusters top down.

$\beta = 2$

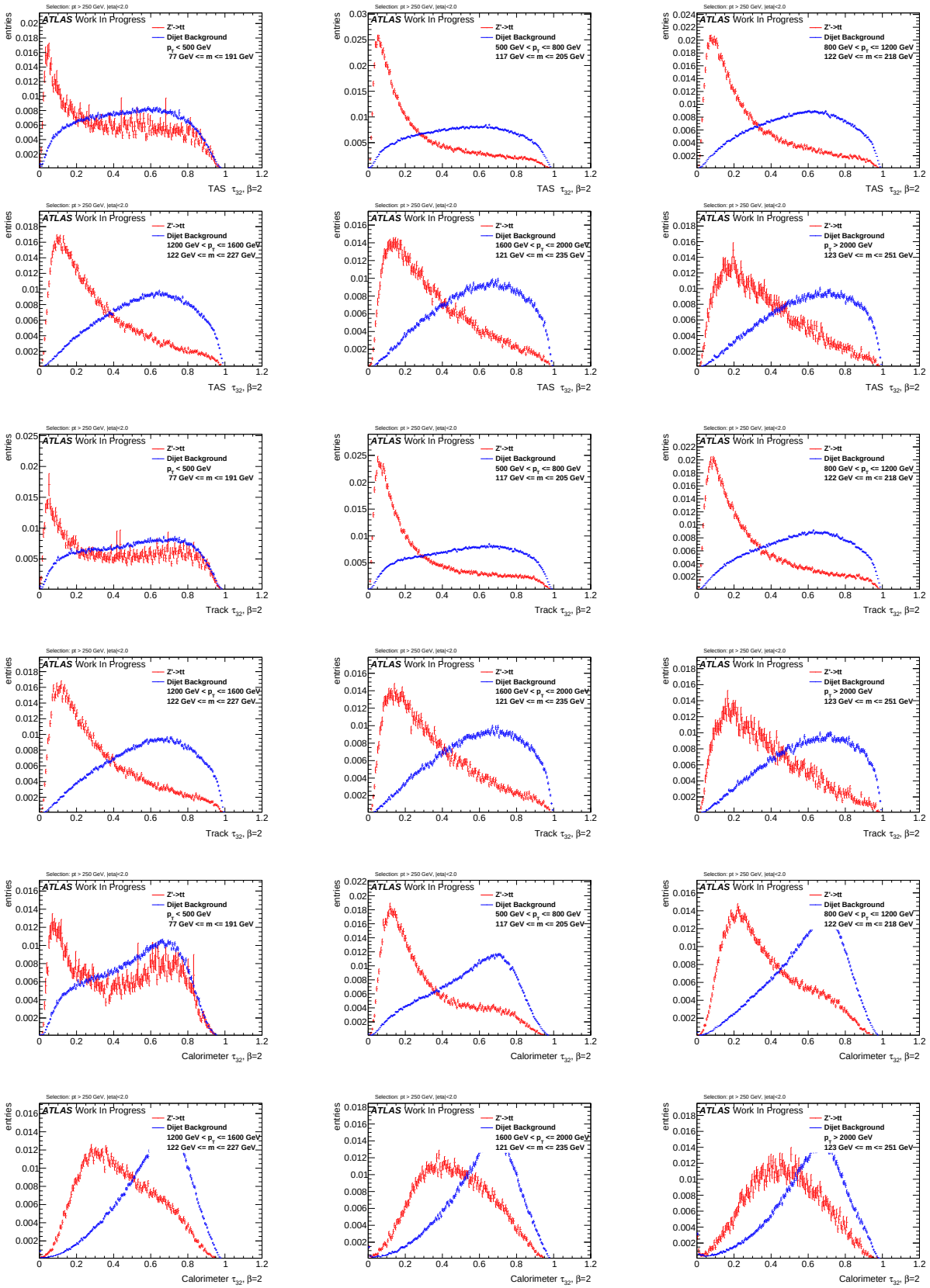

Figure C.33.: Distributions for Top tagging using τ_{32} ($\beta = 2$) using TAS, tracks and calorimeter clusters top down.

$\beta = 3$

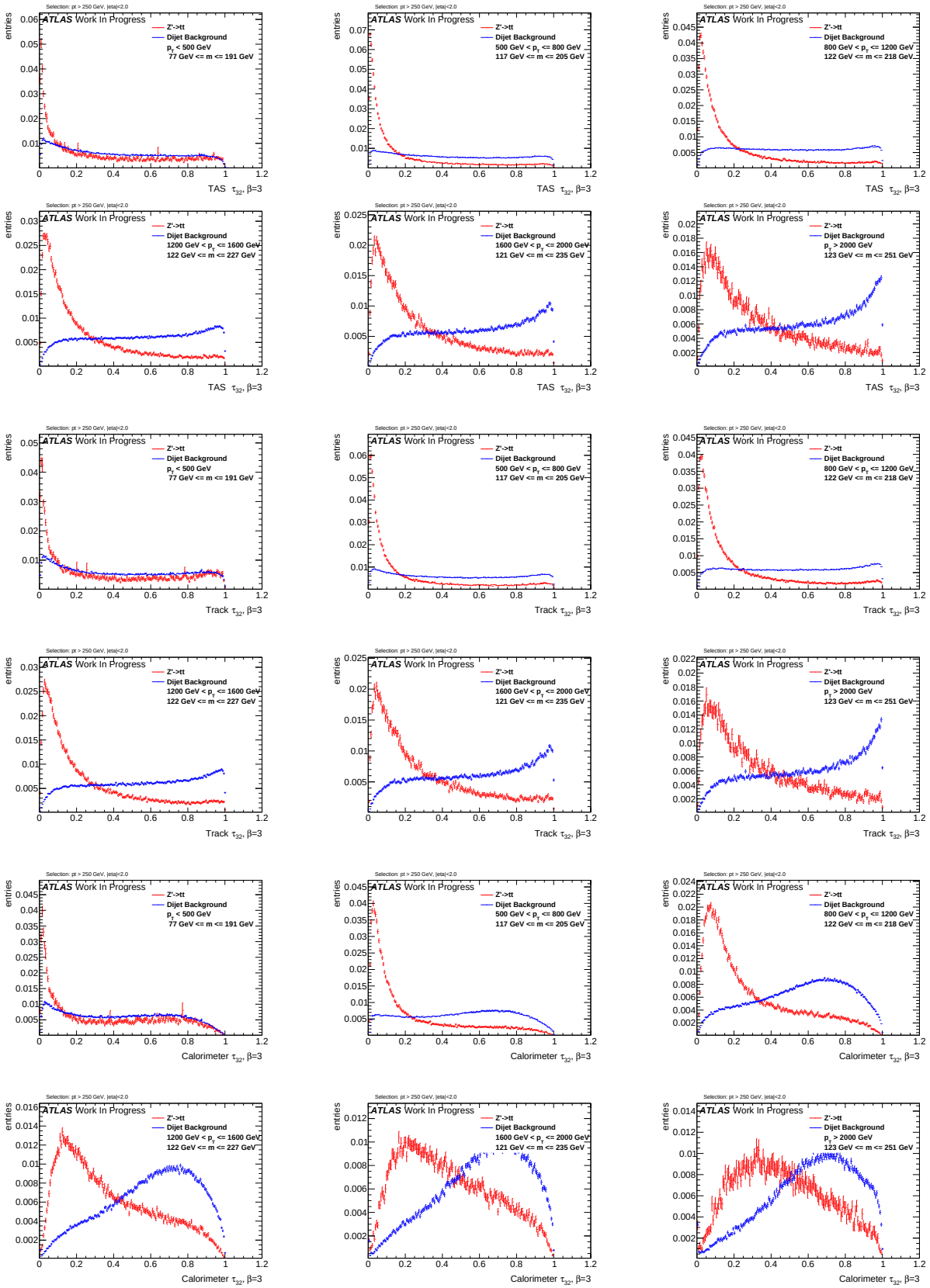

Figure C.34.: Distributions for Top tagging using τ_{32} ($\beta = 3$) using TAS, tracks and calorimeter clusters top down.

Acknowledgements

Foremost, I would like to thank my supervisor, Dr. Oleg Brandt for spending an extraordinary amount of time, energy and thoughts on supporting my first approach to scientific studies.

I also would like to thank Prof. Dr. Hans-Christian Schultz-Coulon, whose lectures teached me the basic principles of physics, decisively influenced my physical intuition and first woke my interest for elementary particle physics.

Of course, many thanks are going to the people of the ATLAS group at the KIP, especially to Merve, Stanislav, Daniel and Fabrizio, who always helped me with technical but also other problems. I remember lots of fruitful discussion about current problems and principle particle physics with Christof, thanks for always having time for coffee and chat.

Last but not least, thanks to my family and my partner, Natascha, who never questioned decisions and constantly support me.

Erklärung

Ich versichere, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, den 29.08.2016,