Universität Bonn

Physikalisches Institut

Development of a new reconstruction and classification method for Tau leptons and its application in the ATLAS detector at the LHC

Christian Limbach

This thesis documents the development of a new reconstruction and classification method for tau leptons in the ATLAS detector at the LHC. It also presents a possible application of the new reconstruction and classification method.
The new reconstruction method follows the *energy flow* approach and thus aims at reconstructing every decay product of the tau. This method utilises the different subdetectors of ATLAS, by combining their measurements, which allows to reconstruct the charged and neutral pions separately.
The new classification method combines the result of the new reconstruction method with kinematics of the tau decay. By calculating *kinematic tau variables*, for instance the angles between the tau decay products, additional information about the decay is made available. Following this approach, it is possible to reconstruct the decay mode and to improve the energy and spatial resolution of the tau.
A possible application of the new tau reconstruction and classification is the measurement of the mean tau polarisation. In a simulation-based analysis, it is shown that the new reconstruction and classification method allows for such measurements.

Physikalisches Institut der
Universität Bonn
Nussallee 12
D-53115 Bonn

BONN-IR-2015-03
May 2015
ISSN-0172-8741

Development of a new reconstruction and classification method for Tau leptons and its application in the ATLAS detector at the LHC

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

von
Christian Limbach
aus
Bonn

Bonn, 17.12.2014

Dieser Forschungsbericht wurde als Dissertation von der Mathematisch-Naturwissenschaftlichen Fakultät der Universität Bonn angenommen und ist auf dem Hochschulschriftenserver der ULB Bonn http://hss.ulb.uni-bonn.de/diss_online elektronisch publiziert.

1. Gutachter:	Prof. Dr. Klaus Desch
2. Gutachter:	Prof. Dr. Jochen Dingfelder
Tag der Promotion:	20.03.2015
Erscheinungsjahr:	2015

Overthinking, over-analysing,
separates the body from the mind.
Withering my intuition,
leaving opportunities behind.[*]

[*] From *Lateralus*, by *Tool*

Summary

This thesis documents the development of a new reconstruction and classification method for tau leptons in the ATLAS detector at the LHC. It also presents a possible application of the new reconstruction and classification method.

The tau lepton is the heaviest lepton known today. It is the only lepton that can decay hadronically. The tau decays before it can interact with the detector material, so that only its decay products can be reconstructed. The hadronic decays are of special interest for the tau reconstruction, because the leptonic decays into electrons and muons are hardly distinguishable from direct electron and muon production. The majority of hadronic tau decays involves charged pions and neutral pions.

The new reconstruction method follows the *energy flow* approach and thus aims at reconstructing every decay product of the tau. This method utilises the different subdetectors of ATLAS, by combining their measurements. By doing so, it is possible to disentangle the signatures of the charged and neutral pions, which overlap in parts in the detector. Thus, they can be reconstructed separately. As a result, also the precision, with which the tau itself is reconstructed, is improved. Also the decay mode of the tau can be determined, by counting the multiplicities of the reconstructed charged and neutral pions.

The new classification method combines the result of the new reconstruction method with kinematics of the tau decay. By calculating *kinematic tau variables*, for instance the angles between the tau decay products, additional information about the decay is made available. This information is exploited to significantly improve the classification of tau decays. In addition, based on the new classification, the calculation of the energy and direction of the tau lepton is improved as well.

A possible application of the new tau reconstruction and classification is the measurement of the mean tau polarisation. This measurement requires knowledge of the decay mode of the tau, and of the energy and direction of its decay products. In a simulation-based analysis, it is shown that the new reconstruction and classification method allows for such measurements. It is also discussed how the tau reconstruction impacts the shape of a polarisation sensitive variable, the charged energy asymmetry.

Contents

CHAPTER 1

Introduction

In particle physics, the elementary building blocks of matter and their interactions are studied. Over the course of the 20th century, a wide range of experiments have been carried out. The findings have been studied and the picture of the elementary particles and their interactions is described by the so-called *Standard Model*.

Many of the results were obtained in collision experiments, in which two beams of particles are accelerated and then brought to collision. Around this collision point, particle detectors are built, in order to measure the outcome of the reaction. The detectors are divided into several parts, which measure different particle types or different properties of a given particle type. Therefore, the detector response needs to be combined and converted from either analog or digital information into so-called higher level objects, which can be interpreted as particles. In physics analyses, which perform the actual measurements, the higher level objects are utilised to select interesting collisions and calculate properties of particles, for instance invariant masses.

Thus, there is an additional layer between the detector response and the physics analysis, which is the reconstruction of particles (or more generally, of objects) inside the detector. The object reconstruction directly feeds in the analyses that are carried out, and its performance will directly impact the performance of the analyses. For instance, if an invariant mass is to be measured, but the energies and positions of the particles in question have large uncertainties, because of the way they are reconstructed, the measurement will not be very significant.

A current example is the discovery of the Higgs boson in Summer 2012 [1, 2]. With its discovery, the Large Hadron Collider and its two main purpose detectors, ATLAS and CMS, reached one of their major physics goals. The Higgs boson was predicted nearly 50 years before its discovery [3] and the machines and experiments, that were finally able to discover it, are counted towards the largest and most complex machines ever built by mankind. First ideas for the Large Hadron Collider and letters of intent for ATLAS and CMS date back 30 years and 20 years respectively [4].

The main point of interest after the discovery is to measure properties of the Higgs boson as precisely as possible. One way to study Higgs boson properties is to study the decay of the Higgs boson into two tau leptons [5]. The success of this study will therefore largely depend on how precisely tau leptons are reconstructed in the detector, and how much information is available about them.

This thesis presents one integral part of a new tau lepton reconstruction method for the ATLAS exper-

iment. The new approach to tau lepton reconstruction makes use of the so-called *energy flow* concept, which allows to reconstruct the tau lepton in more detail and with a better precision. *More detail* specifically means, that individual decay products can be resolved. These decay products in turn can be used as a source of information about the tau decay itself, so that its decay mode can be reconstructed. Exploiting the information stored in the set of reconstructed decay products is the main topic of this thesis.

Not counting this introduction and the conclusions, this thesis is divided into six chapters. A coarse summary of the chapters is given below, a more detailed overview over the contents of each chapter can be found in the beginning of the respective chapter.

Chapter 2 will present the theoretical background. This includes a very brief overview over the so-called Standard Model and the parts of it that are important with respect to this thesis, especially the tau lepton and its decays.

Chapter 3 deals with the experimental setup of ATLAS and the LHC. The ATLAS experiment at the LHC is the experimental tool that allows to study the theory described in chapter two. The LHC accelerator setup and properties of proton-proton collisions at the LHC are explained in the first part. The second part gives information on the ATLAS experiment, i.e. the different sub-detectors and a short overview over simulations.

Chapter 4 is dedicated to the tau lepton reconstruction at ATLAS. The new approach to tau reconstruction which is presented in this thesis differs quite a bit from the approach which was pursued in the first data-taking period of ATLAS. This chapter aims at laying the ground for the discussion of the new substructure based tau reconstruction. To do so, it introduces the particle flow concept and its tau specific implementations in ATLAS. It also shows the setup of the different algorithms that form the new tau reconstruction.

The following two Chapters 5 and 6 build the core of this thesis and present the `PanTau` algorithm, i.e. the setup and the methods used therein and the results obtained in simulation of $Z \rightarrow \tau\tau$ events.

An application of the new substructure based tau lepton reconstruction in a polarisation study is discussed in Chapter 7. Here, the single pion reconstruction is used to calculate a tau polarisation sensitive variable, whose behaviour is studied in an $Z \rightarrow \tau\tau$ event selection. A method to extract the mean polarisation using a so-called fraction fit has been developed and is also presented in this chapter.

CHAPTER 2

Theoretical Background

This chapter provides the theoretical background for the topics presented later in this thesis. Although focusing on the theoretical background, emphasis is put on physics and comprehensive explanations instead of mathematical descriptions, which can be found in the references given.

The chapter is structured as follows. Section 2.1 introduces the *Standard Model*, its particles and their interactions. It also goes into a bit more detail describing the helicity of a particle and its role in the weak interaction, which is of importance in Chapter 7.

Section 2.2 is dedicated to the tau lepton. This includes its basic properties and its decay channels. This will be of importance in the main part of this thesis, Chapter 5, which presents an algorithm that significantly enhances the capability to correctly reconstruct the decay mode of a tau lepton.

Finally, Section 2.3 describes how hadronic decays of the tau lepton can be utilised to measure the tau polarisation. That approach is later on studied in Chapter 7 to develop a method of measuring tau polarisation in data using the new tau reconstruction and a fraction fit.

2.1 The Standard Model of Particle Physics

In order to describe phenomena that can be observed in nature, a reasonable approach is to build a (simplified) mathematical model. For instance, the model of the ideal gas, although purely theoretical, is able to describe real gases under specific circumstances and to make accurate predictions for real gases.

In particle physics the mathematical model is named "Standard Model". As the name may suggest, it turned out to be a very good model of particle physics, making it the one model that is studied and tested. The predictions made by the Standard Model have been tested experimentally, and only little tension has been observed. However, there are open questions, for instance concerning the corrections to the Higgs boson mass (as outlined in Section 1 in [6]) and non-vanishing neutrino masses implied by neutrino mixing [7, 8]. It is also worth mentioning that the Standard Model does not account for gravitational effects[1].

2.1.1 Particle Content

This section follows the outline given in [9]. The particle content of the Standard Model can be classified into two groups, matter particles (fermions) and force carrier particles (bosons).

- *Fermions:*
  Fermions have a spin of $1/2$, and fermions of the first generation (electron, electron neutrino[2], up-quark and down-quark), are the building blocks of existing matter. Experiments in the 20th century [10, 11] showed that there are copies of the electron with the same quantum numbers, but much higher masses, namely the muon and the tau lepton. Due to their weakly interacting nature and their very small mass, the corresponding neutrinos were found much (in each case $\approx$ 25 years) later [12, 13]. Also the up- and down-quarks have heavier copies in the second and third generation, the charm and strange quark in the second, and the top and bottom quark in the third generation.

- *Bosons:*
  Bosons have integer spin and the bosons with a spin of 1 are responsible for the different interactions between matter particles described by the Standard Model. There are no different generations of bosons.

It is convenient to subdivide the fermions into quarks and leptons:

- *Quarks:*
  Quarks are the building blocks of hadronic matter, which includes mesons (a bound quark-antiquark state) and baryons (a bound state consisting of three quarks). Each generation has two quarks, an *up-type* quark with an electric charge of $+2/3 \cdot |e|$ and a *down-type* quark with an electric charge of $-1/3 \cdot |e|$.

- *Leptons:*
  Each generation consists of an electrically charged particle (negative charge) and the neutral neutrino. The electron (charged lepton of the first generation) can form bound states with protons (i.e. baryons) to form atoms.

[1] These effects are negligibly small in current high energy experiments.
[2] The electron neutrino is no constituent of matter, but it is a fermion, so that it is included in this list.

Tables 2.1 and 2.2 summarise the fermions and bosons.

Electric charge	Generation I	Generation II	Generation III
Quarks			
$+\frac{2}{3} \cdot \lvert e \rvert$	Up (u) (1.5 – 3.3 MeV)	Charm (c) (1.27 GeV)	Top (t) (171 GeV)
$-\frac{1}{3} \cdot \lvert e \rvert$	Down (d) (3.5 – 66.0 MeV)	Strange (s) (104 MeV)	Bottom (b) (4.2 GeV)
Leptons			
$0 \cdot \lvert e \rvert$	e^--Neutrino (ν_e)	μ^--Neutrino (ν_μ)	τ^--Neutrino (ν_τ)
$-1 \cdot \lvert e \rvert$	Electron (e^-) (511 keV)	Muon (μ^-) (105 MeV)	Tau (τ^-) (1.78 GeV)

TABLE 2.1: *Fermions in the Standard Model. Only the particles are listed - the corresponding anti-particles can be obtained by inverting the sign of the electric charge. Numbers in brackets give the masses of the particles as found in [14]. Neutrinos are massless in the Standard Model, so that no mass is cited. However, experiments that measured neutrino oscillations [7, 8] indicate that at least two of them indeed have masses. Table taken from [9].*

Electric charge	Electromagnetic interaction	Weak interaction	Strong interaction
+1		W boson (W^+)	
0	Photon (γ)	Z boson (Z^0)	Gluon (g)
−1		W boson (W^-)	

TABLE 2.2: *Bosons in the Standard Model that carry forces. Taken from [9].*

2.1.2 Interactions

This section is based on the references [15–18] and its outline follows the one in [9].

The interactions in the Standard Model are described as gauge-interactions in a relativistic quantum field theory. There are four types of interactions, also called forces, which are the electromagnetic, weak, strong and gravitational force. The latter is not described by the Standard Model, as its strength is negligibly small in current experiments in high energy physics.

- *Electromagnetic force:*
  Each particle with a non-vanishing electric charge is interacting electromagnetically. This force is mediated by the massless and electrically neutral photon.

- *Strong force:*
  Particles with a colour-charge (red, green, blue and the corresponding anti-colours) take part in the strong interaction. In addition to quarks, also the force particle for the strong interaction, the gluon carries colour-charges.

- *Weak force:*
  Every fermion in the Standard Model is interacting weakly. The weak interaction can transform particles of one generation into particles of another generation by means of the charged W-bosons. A model to describe this was proposed by Cabibbo [19]. The weak force couples to the weak eigenstates of down-type quarks, which are a linear combination of the mass-eigenstates of the down-type quarks. These linear combinations can be expressed in a matrix and interpreted as a

rotation. I.e. the propagating down-quark, for example, is a mixture of the *weak down-quark*, the *weak strange-quark* and the *weak top-quark*.

The weak force shows maximal parity violation [16], i.e. it couples to left-handed particles and right-handed anti-particles only.

2.1.3 The Weak Interaction

The polarisation that can be measured in $Z \to \tau\tau$ decays (see Chapter 7) is because of the structure of the weak interaction, as outlined in this section. This section follows the argumentation in reference [20].

A quantum mechanical operator can be described as an operation, which transforms one state into another. Thus, there are operators for the interactions between particles, determining how states change under a given interaction.

The major feature of interest of the weak interaction in the context of this thesis is the maximal parity violation. The parity operator transforms the spatial coordinates into their negatives (like viewing the world through a mirror):

$$t \to t' = t \qquad x \to x' = -x \qquad y \to y' = -y \qquad z \to z' = -z \tag{2.1}$$

The parity operator is given as the matrix γ^0, which is $\gamma^0 = \mathrm{diag}(1, -1, -1, -1)$. It can be shown that the free Dirac equation is invariant under parity transformation, if the spinors transforms like $\psi \to \psi' = \gamma^0\psi$ under the parity transformation.

Experiments showed, however, that parity is violated in weak interactions [16]. Thus, the operator for weak interactions needs to be constructed such, that it is not invariant under parity transformations. The operator is a combination of γ matrices, and there are 5 bi-linear covariant combinations of γ matrices, each with a different behaviour under the parity transformation.

The important combinations are as follows.

- *Vector, V.* By using the 4-vector of gamma matrices γ^μ, the operator flips signs in its spatial part. The momentum is an example for this: If a particle travels to the right, it travels to the left in the mirrored world.
- *Axial vector, A.* Constructing the operator like $\gamma^\mu\gamma^5$ makes it invariant under parity transformations. For instance, the angular momentum $\vec{L} = \vec{r} \times \vec{p}$ behaves like this - both $\vec{r}$ and $\vec{p}$ flip sign under parity, thus $\vec{L}$ remains the same.

It turned out that the operator describing the charged weak interaction is a combination of a vector and the operator that projects out the left-chiral part of a particle:

$$\bar{\psi}\,\gamma^\mu\left(1 - \gamma^5\right)\phi = \bar{\psi}\left(\gamma^\mu - \gamma^\mu\gamma^5\right)\phi \tag{2.2}$$

Comparing to the combinations above, the structure of the interaction is $V - A$. Parity is thus violated, because the terms including odd powers of V will change their sign, while terms with even powers of V won't. Terms with even and odd powers of V and A arise when calculating the cross sections for weak interactions, in which squares of the operators enter.

In weak interactions including a Z^0-boson, the term becomes

$$\bar{\psi}\left(g_V^f\gamma^\mu - g_A^f\gamma^\mu\gamma^5\right)\phi \tag{2.3}$$

with g_V^f and g_A^f being the vector and axial vector couplings of the fermion f. For tau leptons and the up-type and down-type quarks, the values are given in Table 2.3.

Thus, for the weak interaction it is convenient to decompose particles into two chirality states. For instance the electron can be written as

$$|e>=|e_L>+|e_R>$$

Where $|e_L>=(1-\gamma^5)|e>$ is the left-chiral part which takes part in the charged weak interaction.

Fermion	g_V	g_A
τ^-	-0.04	-0.5
u, c, t quarks	0.19	-0.5
d, s, b quarks	-0.35	-0.5

TABLE 2.3: *Values of g_V and g_A for the tau lepton and the quarks. [20]*

2.1.4 Helicity

As a preparation for the study that investigates how tau lepton polarisation can be measured with the substructure based tau reconstruction, this section introduces the *helicity* of a particle. In order to define the helicity h of a particle, its spin and momentum is used as follows:

$$h = \frac{\vec{p} \cdot \vec{s}}{|\vec{p} \cdot \vec{s}|} \tag{2.4}$$

Here, $\vec{p}$ is the momentum vector of the particle and $\vec{s}$ is the spin vector. Thus, the helicity of a particle is the projection of its spin onto its momentum and it is $-1 \leq h \leq +1$. If the helicity is negative, the particle is called *left-handed*, if the helicity is positive, the particle is called *right-handed*. A graphical representation of this is given in Figure 2.1.

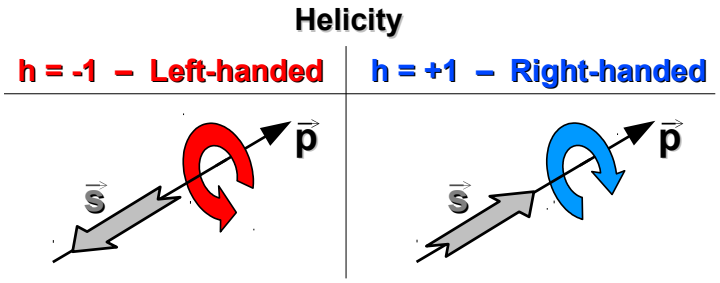


FIGURE 2.1: *Graphical representation of the helicity of a particle with momentum $\vec{p}$ and spin $\vec{s}$. If the spin is anti-parallel (parallel) to the momentum, the scalar product $h = \vec{p} \cdot \vec{s}$ is negative (positive), and the particle is called left-handed (right-handed). Making the thumb point along the momentum, and wrapping the other fingers around it will make the fingers point along the curved arrow when the right (left) hand is used for right-handed (left-handed) particles. The colour coding for left and right handed particles introduced here will be used throughout the thesis - left-handed particles are coloured red, right-handed particles are coloured blue.*

It is important to note that the helicity of a particle depends on the reference frame. If a right-handed particle is moving away from the observer, it appears indeed right-handed. However, if now the

observer is boosted to a reference frame, which moves faster than the particle, the particle will appear as left-handed (as it will now move towards the observer).

In the special case of a massless particle, helicity and chirality are identical. Since massless particles always move with the speed of light, there is no reference frame which has a higher velocity than the particle, so that its helicity is fixed.

For massive particles, this is also true in the limit of $E >> m$ (ultra-relativistic). However, when not in this limit, a particle needs to be considered as a superposition of both, left- and right-handed components.

2.2 The Tau Lepton

The tau lepton is the lepton of the third generation and the heaviest lepton known today. With its mass of $m_\tau = 1.777$ GeV[14], it is the only lepton that can decay both leptonically and hadronically. Because of this, the tau has a multitude of decay channels and the most important ones are listed in Table 2.4.

Decay products	Branching ratio / %
Leptonic modes	35.24
$e^- \nu_e$	17.83
$\mu^- \nu_\mu$	17.41
Hadronic 1-prong modes	46.68
π^-	10.81
$\pi^- \pi^0$	25.52
$\pi^- 2\pi^0$	9.30
$\pi^- 3\pi^0$	1.05
Hadronic 3-prong modes	13.93
$\pi^- \pi^+ \pi^-$	9.31
$\pi^- \pi^+ \pi^- \pi^0$	4.62
Other modes	4.15

TABLE 2.4: *Branching ratios of the most important tau decay modes [14]. The ν_τ is present in every channel and is omitted. Decay products are listed for the decay of a τ^-. Decay products of a τ^+ are the charge conjugates of the listed particles.*
Hadronic modes with pions make up 60.61% of all decay modes. In relation to these 60.61%, 77% (23%) are 1-prong (3-prong) decays, and the decay $\tau^- \rightarrow \nu_\tau \pi^- \pi^0$ is dominant with 42%.

The lifetime of the tau is $\tau_\tau = 290 \cdot 10^{-15}$ s, giving a $c\tau$ of 87.03 μm [14]. Its flight length l in μm in the detector before decaying as a function of its energy E_τ in GeV is roughly

$$l \approx 50\,\mu\text{m} \cdot \frac{E_\tau}{\text{GeV}} \tag{2.5}$$

Thus, tau leptons decay before interacting with the detector[3], so that only their decay products can be detected.

The Feynman diagram of leptonic and hadronic tau decays is shown in Figure 2.2. In both cases, a ν_τ and a W^- are emitted. The tau decay is classified as being hadronic or leptonic by the decay of the virtual W^-.

In case of a leptonic W^- decay, one light lepton (e^-, μ^-) and its anti-neutrino ($\bar{\nu}_e$, $\bar{\nu}_\mu$) are produced in addition to the ν_τ, rendering the majority of the decay products undetectable. Thus, these decays are experimentally not as accessible as the hadronic ones[4]. They can still be considered in physics analyses

[3] See Figure 3.4 - the first active layer of ATLAS has a distance of 5 cm from the centre of the beam pipe. A tau would need an energy of at least 1000 GeV (at $\eta = 0$) in order to reach the first layer.

[4] Because of the two neutrinos, the momentum of the $e^\pm$ or $\mu^\pm$ is smaller for a given τ energy, making an experimental

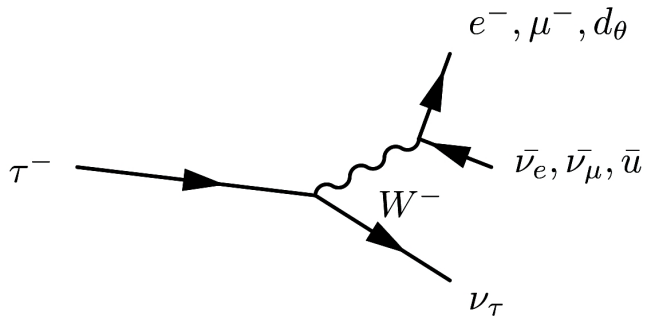


FIGURE 2.2: *Feynman diagram of the tau decay, taken from [21]. Here, it is* $d_\Theta = d \cdot \cos\Theta_C + s \cdot \sin\Theta_C$*, with* Θ_C *being the Cabibbo angle [22]. Leptonic tau decays are not considered in the tau reconstruction algorithms, because the the light lepton is the only detectable particle in the decay. Hadronic tau decays happen via mesons (*$\pi^\pm$*,* $\rho^\pm$ *and* $a_1^\pm$*) and ultimately produce* $\pi^\pm$ *and photons as detectable particles.*

by exploiting the small lifetime of the tau. However, they are not considered in the tau reconstruction algorithms at ATLAS. Therefore, the focus throughout this thesis will be on hadronically decaying taus.

In case of a hadronic $W^\pm$ decay, the quarks produced in the $W^\pm$ decay form a resonance. The $\pi^\pm$, $\rho^\pm$ and $a_1^\pm$ are the main mesons produced here, but mesons including strange quarks are also possible[5]. The $\pi^\pm$ has $c\tau = 8.7$ m [14] and thus will be measured in the detector. The $\rho^\pm$ and $a_1^\pm$ mesons decay instantly, mainly into $\pi^\pm$ and π^0, and the π^0 also mainly decays instantly into a pair of photons.

Hence, the particles that are seen by the detector are effectively the $\pi^\pm$ and photons, originating from π^0 that were produced by the $\rho^\pm$ and $a_1^\pm$.

The hadronic tau decays are thus classified by the number of charged and neutral pions in the decay. The number of $\pi^\pm$ is also referred to as *prong*.

2.3 Tau Leptons as a Tool to measure Polarisation

This section follows the outline in [23]. It describes the production of a Z^0-boson and the polarisation of its decay products (for $Z \to \tau^+\tau^-$). It further describes how the polarisation is (partially) retained in the subsequent decays of the tau leptons into ρ, and ρ decays into $\pi^\pm$ and π^0.

2.3.1 Z boson: Production & Decay into τ Leptons

The Z^0-Boson can be produced by a colliding fermion-anti-fermion pair. In the LHC, which collides protons with protons, this pair is a quark-anti-quark pair. The protons consist of only up- and down-quarks as valence quarks. However, these valence quarks interact with each other by exchanging gluons which can split into a quark-anti-quark pair (for instance in a strange and anti-strange quark). Thus, the proton indeed contains also anti-quarks and quark flavours other than up and down.

Because the annihilation of a quark-anti-quark pair and subsequent production of a tau-anti-tau pair can also be realised by creation of a virtual photon in the s-channel, the two processes (γ^*- and Z^0 production) interfere (c.f. Figure 2.3).

The Z^0-boson has a spin of 1, and in order to conserve the angular momentum, the incoming quark and incoming anti-quark need to have opposite helicities, i.e. their spins have to point in the same direction. This leads to two possible configurations, shown in the two columns of Figure 2.4.

selection less efficient and more biased. In addition, the leptonic tau decays can not be separated from prompt leptons (e.g. from $Z \to \mu\mu$) as well as hadronic tau decays can be separated from their most dominant background, QCD-induced jets [9].

[5] However, those are suppressed due to the weak mixing angle and are only a small fraction of the hadronic tau decays.

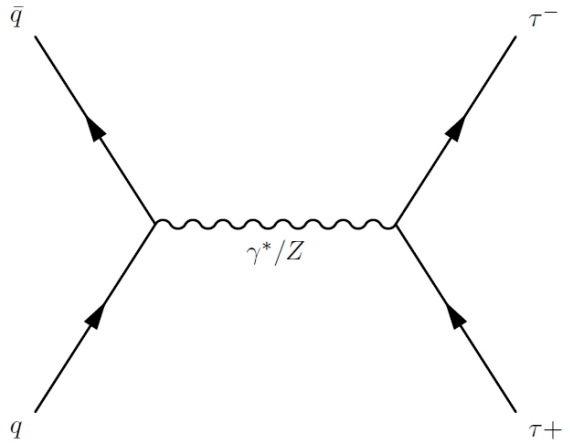


FIGURE 2.3: *Production of a Z^0 boson (or virtual photon) from a quark-anti-quark collision and decay into a tau-anti-tau pair. Both processes, $q\bar{q} \to Z \to \tau^+\tau^-$ and $q\bar{q} \to \gamma^* \to \tau^+\tau^-$, are possible and thus interfere with each other. Figure from [23].*

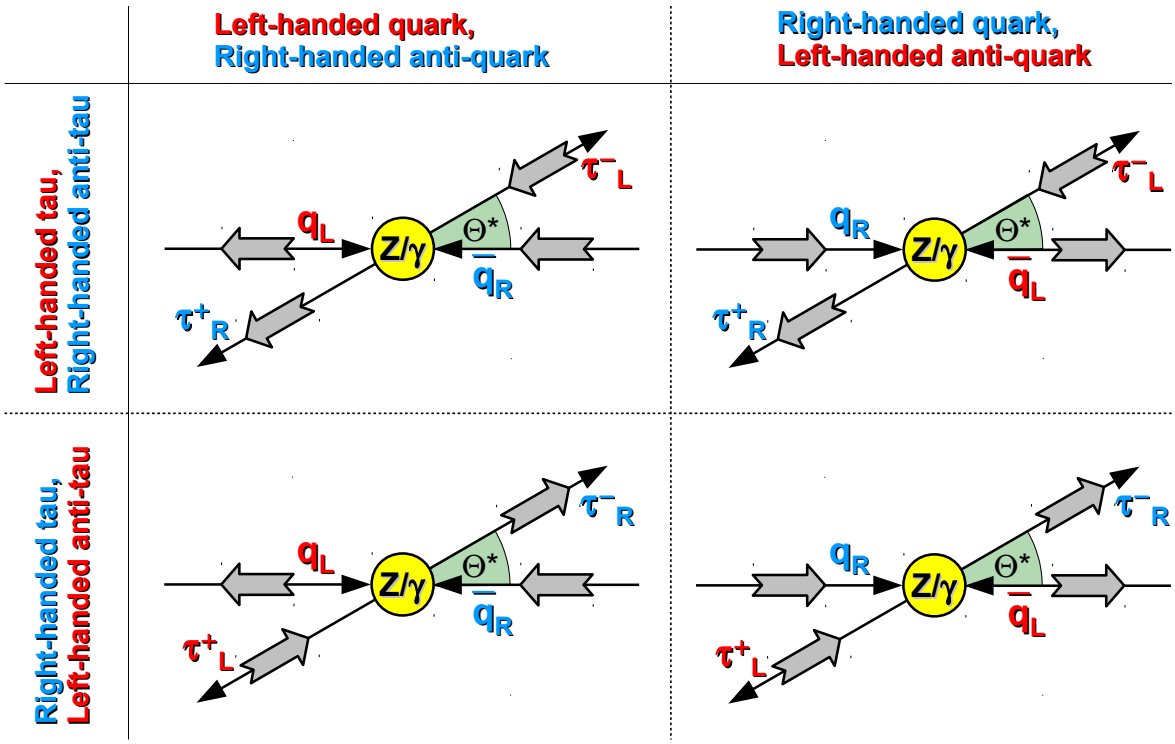


FIGURE 2.4: *Illustration of momentum and spin orientations in the process $q\bar{q} \to Z \to \tau^+\tau^-$. The direction of the momenta can be read off the thin black arrows, the orientation of the spin is denoted by the thick gray arrows. The angle Θ^* is measured between the flight direction of the incoming quark and the outgoing tau lepton.*
The left (right) column shows the possible configurations for left-handed (right-handed) quarks and right-handed (left-handed) anti-quarks. The top (bottom) row shows the possible configurations for left-handed (right-handed) tau leptons and right-handed (left-handed) anti-tau leptons.
Due to the conservation of angular momentum, the top left and bottom right configurations prefer small values of Θ^* and the top right and bottom left configurations favour $\Theta^* \to 180°$.
Inspired by Figure 3.3 from [23].

The same is true for the decay vertex, this time involving the tau and anti-tau. The two possible configurations here are shown in the two rows of Figure 2.4.

In addition to the spins, Figure 2.4 also shows the angle θ^*, which is measured between the incoming quark and outgoing tau. The cross-sections of the four different processes shown[6] depend on the angle θ^* because of the conservation of angular momentum.

For example, the upper left (and also lower right) process of Figure 2.4 will most likely happen with small values of θ^*, because for $\theta^* = 0$ angular momentum is conserved. In case of large values of θ^*, for instance 180°, the initial and final spin orientations differ, making the process less likely to happen.

For the other two processes of Figure 2.4, i.e. the upper right and lower left, θ^* favours large values (i.e. 180°). Explicit formulas for $\frac{d\sigma}{d\cos\theta^*}$ can be derived and can be found in equations 3.4 and 3.5 in [23]. For unpolarised beams (as it is the case in this thesis), only the superposition of the four processes is observed.

2.3.2 Z boson Polarisation

Because of the structure of the weak interaction and the different vector and axial vector couplings of the fermions, the Z^0-bosons that are produced show a non-vanishing average polarisation. This polarisation is transferred to the tau leptons, creating an imbalance of left- and right-handed taus. The average polarisation of the taus is given as

$$P_\tau = \frac{N(\tau_R) - N(\tau_L)}{N(\tau_R) + N(\tau_L)} \tag{2.6}$$

The numbers of left- and right-handed taus $N(\tau_R)$ and $N(\tau_L)$ are proportional to their corresponding cross-sections, which depend on the angle θ^* as discussed in 2.3.1. Because that angle is not accessible experimentally at the LHC, only the overall polarisation is measured, i.e. the polarisation as obtained when integrating over all values of θ^*. The mean polarisation takes the form [23]:

$$\langle P_\tau \rangle = -\frac{2 \cdot g_V^{\tau^-} \cdot g_A^{\tau^-}}{\left(g_V^{\tau^-}\right)^2 + \left(g_A^{\tau^-}\right)^2} = -0.159 \tag{2.7}$$

Thus, there is an asymmetry in the numbers of left- and right-handed taus. This asymmetry is preserved in the decay of the tau lepton into particles that can be measured in the detector.

2.3.3 τ Lepton Decays into ρ-Mesons

The helicity information of the tau is present in each of its decay modes. However, the information is best accessible experimentally in the decay $\tau \to \nu_\tau \rho$, which is described in the following.

The taus (and anti-taus) produced by the Z^0-boson decay can be left- or right-handed. The ρ meson is massive and has a spin of 1, so that it can be polarised transversely (helicity $h = \pm 1$) and longitudinally. Thus, there are six combinations, four of which are depicted in Figure 2.5. It is worth noticing that the transversely polarised ρ with $h = +1$ is not shown, because those configurations are suppressed due to the spins of the decaying $\tau-$ ($s = 1/2$) and the spins of ν_τ and ρ (parallel, adding up to $s = 3/2$).

When considering the spins only, then all four combinations in Figure 2.5 are equally likely, the top left and bottom right for $\theta \to 0°$ and the top right and bottom left for $\theta \to 180°$. However, due to the

[6] The process itself is the same, only the helicities are different. From top left to bottom right the processes are: $q_L\bar{q}_R \to \tau_L^-\tau_R^+$, $q_R\bar{q}_L \to \tau_L^-\tau_R^+$, $q_L\bar{q}_R \to \tau_R^-\tau_L^+$ and $q_R\bar{q}_L \to \tau_R^-\tau_L^+$.

high mass of the ρ, the negligibly small mass of the neutrino and the conservation of momentum, the ρ will be preferably emitted in direction of the tau lepton[7][8].

Therefore, the dominant combinations are the upper right and lower left, so that right-handed taus prefer to decay into longitudinally polarised ρ-mesons, and left-handed taus decay mostly into transversely polarised, left-handed ρ-mesons.

2.3.4 Decay of polarised ρ-Mesons

The ρ-meson itself decays further via $\rho^{\pm} \rightarrow \pi^{\pm}\pi^{0}$. The pions themselves have no spin, so that the spin of the ρ needs to impact the angle ψ between the flight direction of the ρ^{-} and the π^{-} in order for the total angular momentum to be conserved.

Because of the spinless pions, there are two configurations in this case (transversely and longitudinally polarised ρ), shown in Figure 2.6.

In the case of a longitudinally polarised ρ without angular momentum, $l = 0$, the main total angular momentum quantum number is $j = 1$ (due to the ρ spin $s = 1$) and the secondary total angular momentum quantum number is $m = 0$. Therefore, the angular distribution of the decay products follows the spherical harmonic $Y_{l=1,m=0}$[24]:

$$Y_{l=1,m=0} = \frac{1}{2}\sqrt{\frac{3}{\pi}}\cos\theta \propto \cos\theta \tag{2.8}$$

The function is maximal at $\theta = 0°$ and $\theta = 180°$, so that the pions are emitted preferably with these angles, i.e. parallel and anti-parallel to the flight direction of the rho. This holds true in the rest frame of the ρ. When boosting into the laboratory frame, this makes the pions have different energies, i.e. either the charged pion has a much higher energy ($\theta = 0°$), or a much lower energy ($\theta = 180°$) than the neutral pion.

For the transversely ($h = -1$) polarised ρ with $l = 0$, it is $j = 1$ and $m = -1$. The corresponding spherical harmonic is $Y_{l=1,m=-1}$[24]:

$$Y_{l=1,m=-1} = \frac{1}{2}\sqrt{\frac{3}{2\pi}}\sin\theta\, e^{i\phi} \propto \sin\theta \tag{2.9}$$

Here, the function has its maximum at $\theta = 90°$ and $\theta = 270°$. This means that both, the charged and the neutral pion, are emitted perpendicular to the flight direction of the ρ (in the rest frame of the ρ). Boosting into the laboratory frame, the energies of the neutral and charged pion will be similar for both cases, $\theta = 90°$ and $\theta = 270°$.

2.3.5 Polarisation Sensitive Variable: Υ

Combining the findings from Sections 2.3.2, 2.3.3 and 2.3.4, an experimentally measurable variable, which is sensitive to the tau polarisation P_τ, can be defined. The variable is called the charged energy

[7] To be more precise, the scalar product of $\vec{p_\rho}$ and $\vec{p_\tau}$ will be positive. The maximum angle between the two momenta will of course decrease with increasing tau momentum because of the boost.

[8] One way to see this, is to assume $p_\tau = (E_\tau, x_\tau, 0, 0)^T$, $p_\rho = (E_\rho, x_\rho, 0, 0)^T$ and $p_\nu = (E_\nu, x_\nu, 0, 0)^T$ and then solve the square of $p_\tau - p_\rho = p_\nu$ for x_ρ as a function of x_τ by using $E_\rho = E_\tau - E_\nu = E_\tau - x_\nu$. For x_τ above ≈ 1.65 GeV, the ρ can not be emitted backwards with respect to the tau.

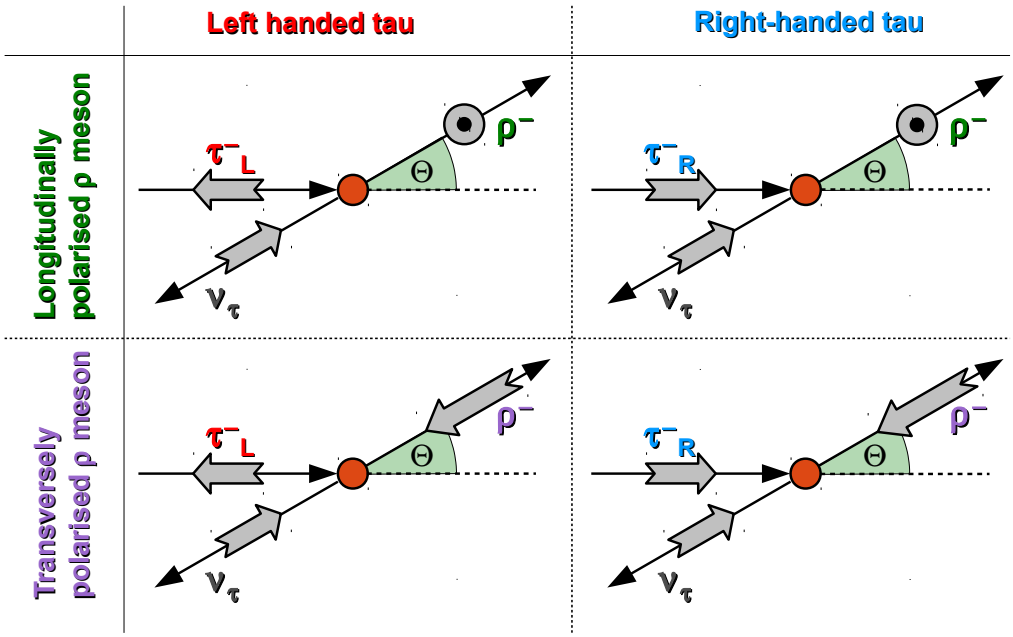


FIGURE 2.5: *Illustration of spin orientations in the decay $\tau \to \nu_\tau \rho$. The direction of the momenta is denoted by the thin black arrows, the thick gray arrows indicate the orientation of the spin. Θ is the angle between the flight direction of the tau and the flight direction of the ρ meson.*

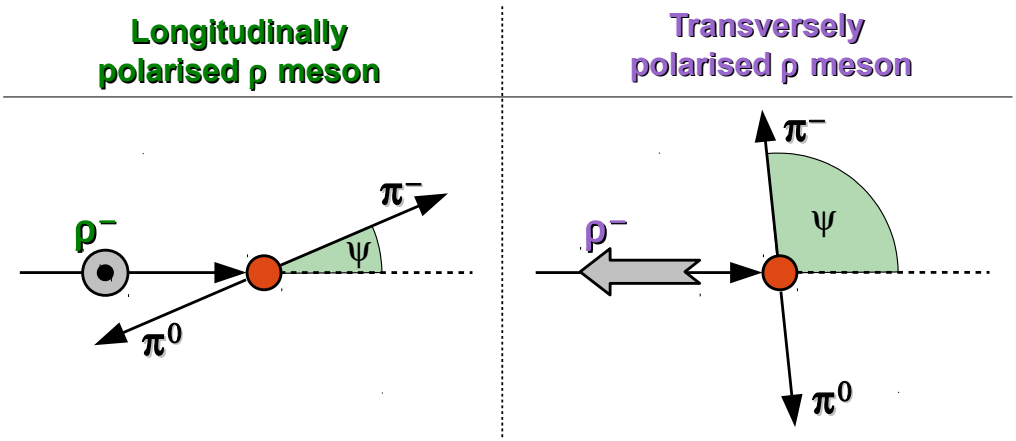


FIGURE 2.6: *Orientations of charged and neutral pion in the decays of longitudinally and transversely polarised ρ mesons.*
For longitudinally polarised taus, pions are emitted according to the spherical harmonic $Y_{1,0} \propto \cos(\psi)$. The preferred directions of the charged pion are therefore $\psi = 0°$ and $\psi = 180°$, making it either much harder or much softer than the neutral pion.
In the case of transversely polarised taus, the spherical harmonics $Y_{1,-1}$ is of interest, with $Y_{1,\pm1} \propto \sin(\psi)$. Thus, the pion is preferably emitted with $\psi = 90°$ and $\psi = 270°$, so that its energy is similar to that of the neutral pion.

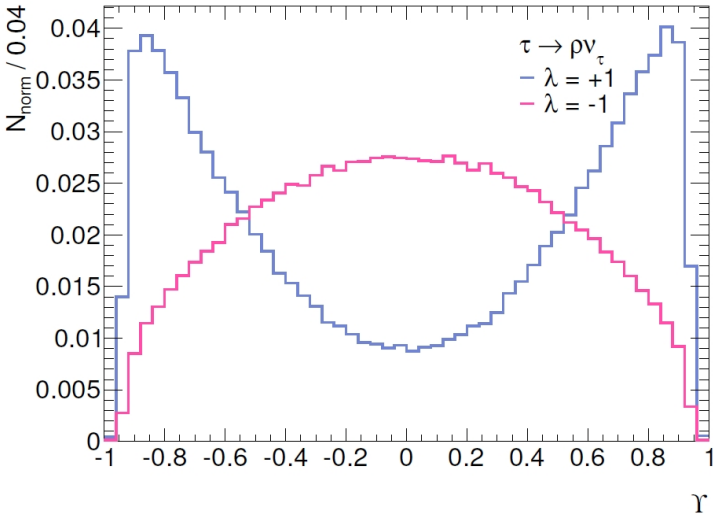


FIGURE 2.7: *Distribution of Υ for left-handed (red) and right-handed (blue) tau decays $\tau \to \rho\nu_\tau$ with $\rho \to \pi\pi^0$. Right-handed taus show a large asymmetry, while left-handed taus show a small asymmetry, as summarised in Section 2.3.5. Figure taken from [23].*

asymmetry, Υ:

$$\Upsilon = \frac{E(\pi^\pm) - E(\pi^0)}{E(\pi^\pm) + E(\pi^0)} \tag{2.10}$$

The codomain of Υ is $\Upsilon \in (-1, +1)$. Large values, $\Upsilon \to +1$, indicate that the charged pion has a much higher energy than the neutral pion. Small values, $\Upsilon \to -1$ indicate a low-energetic charged pion and a high-energetic neutral pion. Values around 0 point to equally energetic pions.

Going through the subsequent decays in reverse, the following is noted:

1. Large (small) absolute values of Υ indicate different (equal) energies of the $\pi^\pm$ and π^0.
2. A large (small) energy difference is due to a longitudinally (transversely) polarised ρ meson.
3. Longitudinally (transversely) polarised ρ mesons are produced preferably by right-handed (left-handed) tau leptons.

Thus, Υ provides access to the tau polarisation. The Υ-distributions for left- and right-handed tau decays $\tau \to \nu_\tau\rho \to \nu_\tau\pi\pi^0$ is shown in Figure 2.7.

However, in order for this quantity to be measured, it is necessary to correctly reconstruct and identify tau leptons and to classify them as decaying via $\tau \to \nu_\tau\rho$. It is also necessary to be able to reconstruct individual pions in a given tau decay.

CHAPTER 3

Experimental Setup

This chapter introduces the accelerator complex at CERN, where the LHC is located, in Section 3.1. The section then mainly focuses on the (instantaneous) luminosity, which plays an important role not only in reaching the physics analysis goals (the Higgs boson, searches for new physics), but also in (for instance) tau lepton reconstruction.

In Section 3.2, this chapter presents the ATLAS detector, one of the four large detectors at the LHC, and one of the two multi purpose detectors. This section also describes the different sub-detectors of ATLAS.

Finally, Section 3.3 briefly discusses simulations of high energy processes.

3.1 The Large Hadron Collider

An in-depth description of the LHC is given in [25]. The Large Hadron Collider (LHC) is a circular hadron accelerator and collider. It is located at CERN[1] near Geneva, Switzerland. The accelerator consists of two rings with a circumference of 26.7 km, which are installed in a tunnel in a depth of 40 m to 170 m. It accelerates two separate beams of protons to energies of 7 TeV[2] each, leading to proton-proton collisions at an unprecedented centre of mass energy of 14 TeV.

The LHC is the last stage of a chain of accelerators which is depicted schematically in Figure 3.1. The accelerator chain is as follows [26]. Initially, the protons are obtained by ionising hydrogen atoms. In a linear accelerator (the Linac 2) electric fields are applied to first accelerate the protons to an energy of 50 MeV. From here, the protons are injected into the Proton Synchrotron Booster (PSB), a circular accelerator which increases the energy up to 1.4 GeV. The next step is the Proton Synchrotron (PS), which accelerates the protons to energies of 25 GeV. In the PS, the protons are also separated spatially and accumulated in so-called *bunches*. The last pre-acceleration step is the Super Proton Synchrotron (SPS). Here, the protons are accelerated to an energy of 450 GeV, before they are extracted into the LHC rings. Once inside the LHC, they are accelerated even further to their target energy of 7 TeV. The number of proton bunches in the LHC is 2808 and the number of protons per bunch is $1.1 \cdot 10^{11}$.

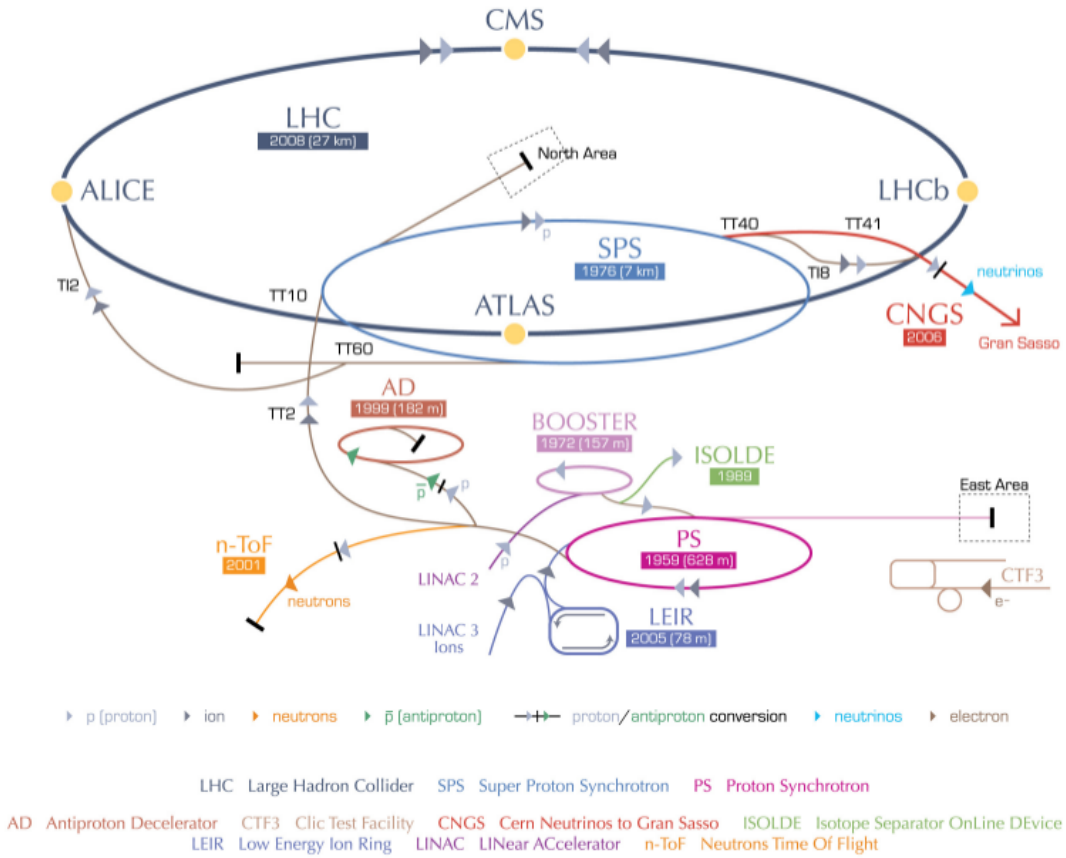


FIGURE 3.1: *LHC accelerator complex. The accelerator chain for the protons that are collided in the ATLAS detector is Linac2, Booster, PS, SPS and LHC. [27]*

[1] CERN: **C**onseil **E**uropean per la **R**echerche **N**ucleaire

[2] The design energy per beam is 7 TeV. However, in 2010 and 2011 the beam energy was 3.5 TeV; in 2012 it was 4 TeV.

The whole process takes roughly 30 minutes, 10 minutes to fill both rings and 20 minutes to increase the energy from 450 GeV up to 7 TeV. The protons are bend by superconducting magnets that provide a magnetic field of up to 8.33 T, working at a temperature of 1.9 K.

The two circulating proton beams are crossed and brought to collision at four interaction points. Around these interaction points, detectors are built to measure the proton-proton reactions. The four experiments are listed in alphabetical order.

- *ALICE*[3]: In addition to accelerating and colliding protons, the LHC can also work with fully ionised lead ions, $^{208}\mathrm{Pb}^{82+}$. The magnetic field of 8.33 T allows for energies of 2.76 TeV/nucleon. The ALICE detector is designed to study strongly interacting matter and the quark-gluon-plasma in nucleus-nucleus collisions [28].

- *ATLAS*[4]: ATLAS is one of the two general purpose detectors at the LHC, designed to cover a broad spectrum of proton-proton physics. It is described in more detail in Section 3.2.

- *CMS*[5]: CMS is the second multi purpose detector. Opposed to ATLAS, the CMS detector uses only one magnetic field to bend charged particles and both calorimeters are inside the solenoid magnet. The whole event reconstruction in CMS is based on an energy flow approach. In ATLAS, this approach is rather new and its benefits for tau leptons and jets are just starting to be explored. The way the energy flow approach is applied in tau lepton reconstruction in ATLAS is described in Sections 4.3 and 4.4.

- *LHCb*[6]: The LHCb detector is designed to study rare decays involving charm and bottom quarks. The aim is to understand the asymmetry between matter and antimatter, by studying these rare decays [29].

Figure 3.2 shows the total proton-proton cross section as a function of centre of mass energy together with the cross sections of other proton-proton processes, for example $pp \to H$. While the overall cross section is rather large ($\sigma_{\text{tot}} \approx 10^8$nb), the interesting processes like the production of a Higgs boson with a mass of 150 GeV have a cross section of only 10^{-3} nb, i.e. 11 orders of magnitude smaller.

Therefore, the challenge for the LHC and the detectors is twofold. For the LHC on the one hand, the challenge is to provide a high collision rate, so that a statistically significant amount of $pp \to H + X$ reactions is actually produced in a reasonable amount of time. By design, the rate of $pp \to H + X$ reactions is roughly 10^{-3} per second[7]. This rate is closely connected to the instantaneous luminosity, which is explained below.

For the detectors on the other hand, it is crucial to select these 10^{-3} interesting collisions per second out of the 10^8 collisions that are happening in total.

The quantity of interest for the production rate of any process is the instantaneous *luminosity* $\mathcal{L}$. It is given by [30, 31]:

$$\mathcal{L} = \frac{f_{\text{rev}} \cdot N_{\text{Bunches}} \cdot N_{\text{p},1} \cdot N_{\text{p},2}}{\sigma_x \cdot \sigma_y} \cdot F\left(\Phi, \sigma_x, \sigma_y\right) \tag{3.1}$$

[3] ALICE: **A** **L**arge **I**on **C**ollider **E**xperiment
[4] ATLAS: **A** **T**oroidal **L**HC **A**pparatu**S**
[5] CMS: **C**ompact **M**uon **S**olenoid
[6] LHCb: **L**arge **H**adron **C**ollider **b**eauty experiment
[7] The Higgs boson production merely serves as an example in this context. It stands as a placeholder for any interesting process to be studied.

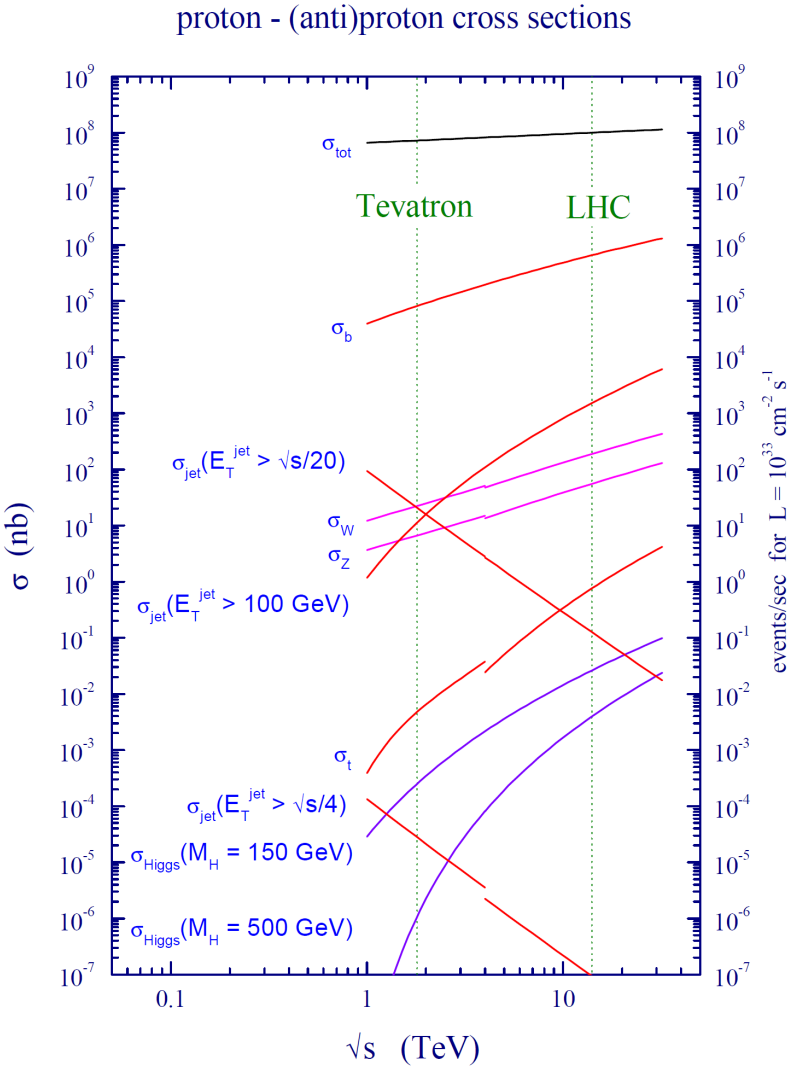


FIGURE 3.2: *Proton-proton cross sections as a function of centre of mass energy. Nine orders of magnitude is the difference between interesting processes ($pp \rightarrow H + X$) and the overall proton-proton cross section. Figure taken from [30]*

In the nominator, f_{rev} is the frequency with which the protons circulate in the LHC, N_{Bunches} is the total number of proton bunches and $N_{\text{p},1}$ and $N_{\text{p},2}$ are the numbers of protons in the two colliding proton bunches. The denominator contains the transverse width of the beams σ_x and σ_y. The additional factor $F\left(\Phi, \sigma_x, \sigma_y\right)$ is a geometrical factor that arises because the beams collide under a small crossing angle Φ. The design instantaneous luminosity of the LHC is 10^{34} cm^{-1} s^{-1}.

The rate $\frac{dN}{dt}(pp \to X)$ with which the reaction $pp \to X$ occurs depends directly on the instantaneous luminosity $\mathcal{L}$:

$$\frac{dN}{dt}(pp \to X) = \mathcal{L} \cdot \sigma(pp \to X) \tag{3.2}$$

where $\sigma(pp \to X)$ is the cross section of the process in question. Therefore, the absolute number of $pp \to X$ reactions can be calculated by integrating equation 3.2 over the time:

$$\begin{aligned} N(pp \to X) &= \left(\int \mathcal{L}\, dt\right) \cdot \sigma(pp \to X) \\ &= L \cdot \sigma(pp \to X) \end{aligned}$$

L is called the *integrated luminosity* and can be thought of as corresponding to the total amount of data that is collected.

Although the integrated luminosity is a measure of how much data is produced, it is the instantaneous luminosity that is important to acquire the data in a short amount of time. Therefore, it is the instantaneous luminosity that needs to be maximised if the goal is to produce as much Higgs bosons per second as possible. The instantaneous luminosity,

$$\mathcal{L} = \frac{f_{\text{rev}} \cdot N_{\text{Bunches}} \cdot N_{\text{p},1} \cdot N_{\text{p},2}}{\sigma_x \cdot \sigma_y} \cdot F\left(\Phi, \sigma_x, \sigma_y\right)$$

can be maximised in different ways.

- *Increasing the number of bunches, $N_{Bunches}$:*
  This leads to more bunches colliding per time, i.e. the time between two collisions in the detectors is reduced. The design bunch spacing of the LHC is 25 ns, leading to a collision rate of 40 MHz. Combined with the scales of the ATLAS detector (which is presented in Section 3.2), this results in an experimental challenge:

  If ATLAS is assumed to be of a cylindrical shape, it has a radius r of roughly 12 m. If particles produced in the pp collisions traverse the detector with the speed of light[8], c, then they need the time

  $$t = \frac{r}{c} \approx 40 \text{ ns}$$

  to leave the detector, if they travel perpendicular to the beam axis. In other words, they need at least these 40 ns, because the total width of the detector is ≈ 40 m. However, while these particles are still flying through the detector, the next bunches collide. This means, that at a given time, there are particles from different collisions in the detector. Because the particles originate from collisions that took place at a different time, this effect is called *out of time pile-up*.

- *Increasing the number of protons per bunch, $N_{p,1}$ and $N_{p,2}$:*
  This effectively increases the proton density in the bunch. Thus, it becomes more likely for two

[8] This is not possible for massive particles. However, for the sake of simplicity and the argument, c is used as a conservative approach.

protons to interact with each other. This is one of the sources of a major effect at the LHC, which is discussed below.

- *Decreasing the beam width, σ_x and σ_y:*
  Focusing the beam to a smaller area at the collision point is another method to increase the proton density and thus the chances for a pp interaction. However, reducing the beam width and increasing the number of protons per bunch lead to an additional effect.

Taking the last two measures listed above (increasing $N_{p,1}$ and $N_{p,2}$ or σ_x and σ_y) generally increases the chances for two protons to interact in a given bunch crossing[9]. That means it increases the probability for a process like $pp \to H$, but it also increases the chances of soft QCD reactions to happen. These soft QCD reactions have a lower momentum transfer and a higher cross section than the interesting processes. Therefore, there is more than one proton-proton collision in the same bunch crossing and the effect is called *in time pile-up*. Because this is the more important of the two pile-up effects for the studies presented in this thesis, it will later be referred to as *pile-up*. The additional particles produced in pile-up collisions typically have small energies, but they still overlay with particles from the interesting collision. Thus, they can either be mistaken as particles of interest (for instance neutral pions in the reconstruction of a tau decay), or they can alter the measured energy of a particle, if they overlap with it spatially. Combined with the high amount of pile-up collisions (up to more than 20 per bunch crossing), this effect is one of the major challenges in object reconstruction at LHC experiments.

3.2 The ATLAS Experiment

This section describes the ATLAS[10] detector, its general design, the different sub-detectors and the trigger system. Emphasis is put on sub-detectors that are of major importance for this thesis, i.e. the decay mode reconstruction and measurement of tau leptons. This section, including its subsections is largely based on [32].

The ATLAS detector is a general purpose detector, designed to cover a wide physics programme. One of the driving forces for the design is the Higgs boson. Depending on its mass, the decay modes of the Higgs range from isolated photons in $H \to \gamma\gamma$ over b-jets from $H \to bb$ to a multitude of final states from $H \to WW$ and $H \to ZZ$.

In theory, there can be additional heavy gauge bosons, W' and Z', which would decay (for instance) into high energetic leptons [33]. Finally, in R-parity conserving supersymmetry [6], supersymmetric particles decay via cascades into the lightest supersymmetric particle which is weakly interacting, thus producing a signature of missing transverse energy.

The discovery and measurement of these particles and phenomena can be translated into requirements that need to be fulfilled by the different detector parts [32].

Reconstruction and precise measurement, including the charge, of high energy leptons put demands on the inner detector. Of special importance is the capability to reconstruct secondary vertices in the case of tau leptons and jets containing b-quarks. In order to accurately measure the two photons from a Higgs decay, the electromagnetic calorimeter needs to have a good resolution. The hadronic calorimeter needs to cover the full solid angle, to provide a reliable measurement of the missing transverse energy.

[9] Bunch crossing is the crossing of two bunches at the collision point.
[10] ATLAS: **A** **T**oroidal **L**HC **A**pparatu**S**

The muon system needs to be capable of safely measuring the charge of high energetic muons from potential heavy gauge boson decays (or other sources).

Given the LHC conditions, all detectors need to be fast and radiation-hard. In addition, a high granularity is crucial because of the high particle fluxes.

3.2.1 Coordinate System and Overview

ATLAS uses a right-handed coordinate system. The origin of the system is set to be the nominal interaction point of the two proton beams. The positive x-axis points to the centre of the LHC, the y-axis points upwards to the earth surface and the direction of the z-axis coincides with the beam pipe. However, it is much more common to denote the momentum of a particle in the coordinates p_T, η and ϕ rather than p_x, p_y and p_z. p_T is the transverse momentum, i.e. the momentum in the x-y-plane,

$$p_T = \sqrt{p_x^2 + p_y^2}$$

and ϕ is the angle towards the x-axis in the x-y-plane.

$$\sin(\phi) = \frac{p_y}{p_x}$$

Using the polar angle θ, which measures the angle with respect to the beam pipe,

$$\sin(\theta) = \frac{p_y}{p_z}$$

the pseudorapidity η is defined:

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right)$$

Thus, $\eta = 0$ refers to $\theta = 90°$ (i.e. upwards, parallel to the y-axis) and $\eta \to \infty$ $(-\infty)$ corresponds to $\theta \to 0°$ $(90°)$. Distances between two directions are quantified by the measure ΔR,

$$\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$$

It is noteworthy that the actual angular distance that a ΔR value refers to depends on the pseudorapidity values that are involved, because of the natural logarithm that is involved in the calculation of η.

An overview of the complete ATLAS detector is given in Figure 3.3. Being a general purpose detector, i.e. designed to cover a wide range of physics analyses, it follows the typical onion-like structure. This means that the different sub-detectors are built around the interaction point in different layers, covering the largest feasible solid angle.

3.2.2 Inner Detector

Charged particles ionise the detector material they pass through, producing a trace through the inner detector. The trajectory of a charged particle is also called *track*. A magnetic field flows through the inner detector parallel to the z-axis, bending charged particles in the x-y-plane. Thus, due to the Lorentz force, the tracks show a curvature, which is used to determine the momentum of the particle. Because neutral particles do not ionise the detector material they can not be detected in this sub-detector.

The inner detector has a length of 6.2 meters and a diameter of 2.1 meters, and it covers pseudorapidities of $|\eta| < 2.5$. A solenoid magnet provides a magnetic field of 2 T. The inner detector is divided into

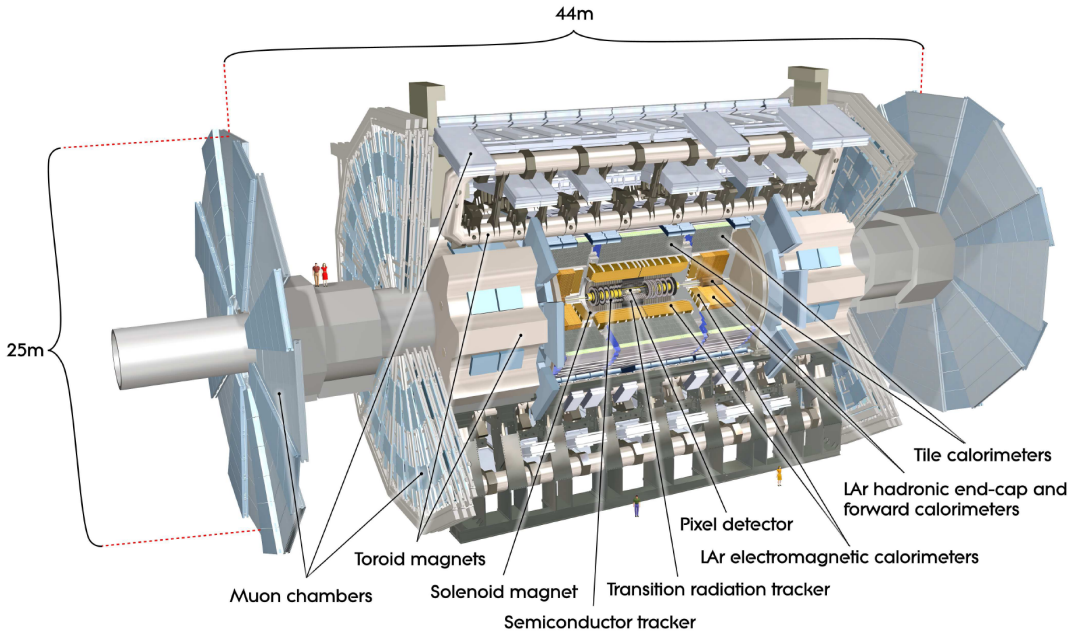


FIGURE 3.3: *The ATLAS Detector. Taken from [32]. The largest fraction of the volume is occupied by the muon chambers (light blue). In the central part, the tracking detectors, the electromagnetic calorimeter and the hadronic calorimeter are located.*

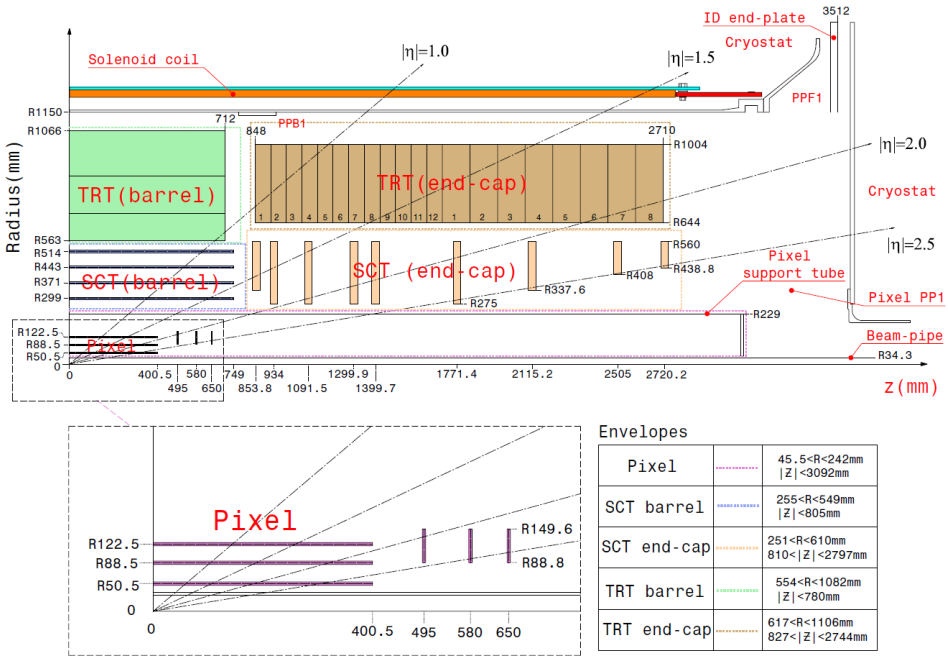


FIGURE 3.4: *The inner detector of ATLAS [32]. Visible are the three sub-detectors Pixel, SCT and TRT and their separation into barrel and end-cap parts. The inner detector can measure charged particles within $|\eta| < 2.5$. The number of readout channels exceeds 80 million.*

three parts, the Pixel, the semiconductor tracker (SCT) and the transition radiation tracker (TRT), and it is depicted in Figure 3.4.

The Pixel detector is divided into the barrel- and the end-cap-region. In the barrel part, it consists of three layers of cylindrical shape. The first layer is placed at a radial distance of 5.05cm to the proton beams. In the end-caps, the number of layers is also three, but they have the form of discs. The arrangement is such, that a particle traverses three layers of the Pixel detector in total, independent of its pseudorapidity. In $R-\phi$- (z-) direction, the resolution is around 10μm (115μm).

The second part of the inner detector is the semi conductor tracker (SCT). The resolution of the SCT, with values of 16μm ($R-\phi$) and 580μm (z) is slightly worse than the resolution of the Pixel. However, the SCT consists of four cylindrical layers in the barrel and 9 discs in the end-cap. They are arranged such, that at least 4 layers are traversed.

The third and last part of the inner detector is the transition radiation tracker (TRT). It consists of small cylindrical tubes filled with a gas mixture and with a wire located in the centre. Charged particles passing through the tubes will ionise the gas, producing electron-ion pairs. Because there is a voltage applied to the wire, the electrons and ions are accelerated towards the anode and cathode, producing a measurable signal after amplification. In between the tubes, there are radiators, in which transition radiation photons can be produced. The production of transition radiation photons depends on the γ-factor of the particles passing through the TRT. Thus, the TRT can be used as a tool to identify the particle that passed through it [34]. The TRT provides ≈ 30 spatial measurements of a particle, with a resolution of 170μm.

3.2.3 Calorimeters

The calorimeter system is used to measure the energy of neutral and charged particles. It is split into the electromagnetic (EM) and hadronic (HAD) calorimeter. Incident electrons will create bremsstrahlung photons when they traverse the calorimeter. The bremsstrahlung photons will convert into e^+e^- pairs in the electric field of the nuclei. These e^+ and e^- however will again produce bremsstrahlung photons and thus a shower of photons, electrons and positrons develops. This cycle repeats until the photon energy is below the pair production threshold. For an incoming photon, the process is the same, except that an e^+e^- pair is produced first. The length scale on which these processes happen is the radiation length χ_0.

Hadrons will produce showers too, but due to their mass and the strong interaction the length scale is much larger. Hadronic showers are typically more diffuse and suspect to fluctuations than electromagnetic showers. It is worth noticing that hadrons will deposit energy in all layers of the calorimeters, including the EM calorimeter.

Both the EM-calorimeter and the HAD-calorimeter are divided into a barrel part and two end-caps, as shown in Figure 3.5. The calorimeter system covers the full ϕ range, 0 to 2π and covers pseudorapidities up to $|\eta| = 4.9$.

The barrel part of the EM-calorimeter covers pseudorapidities of $|\eta| < 1.475$, the end-caps cover $1.375 < |\eta| < 3.2$. Its thickness exceeds 22 (24) radiation lengths in the barrel (end-caps). The electromagnetic calorimeter has a lateral and longitudinal segmentation. Figure 3.6 shows a part of the EM-calorimeter at $\eta = 0$.

Particles passing through the EM-calorimeter will first traverse the EM1-layer. EM1 is also called *Strip layer*, because the lateral segmentation is such, that the granularity in η is much higher than the

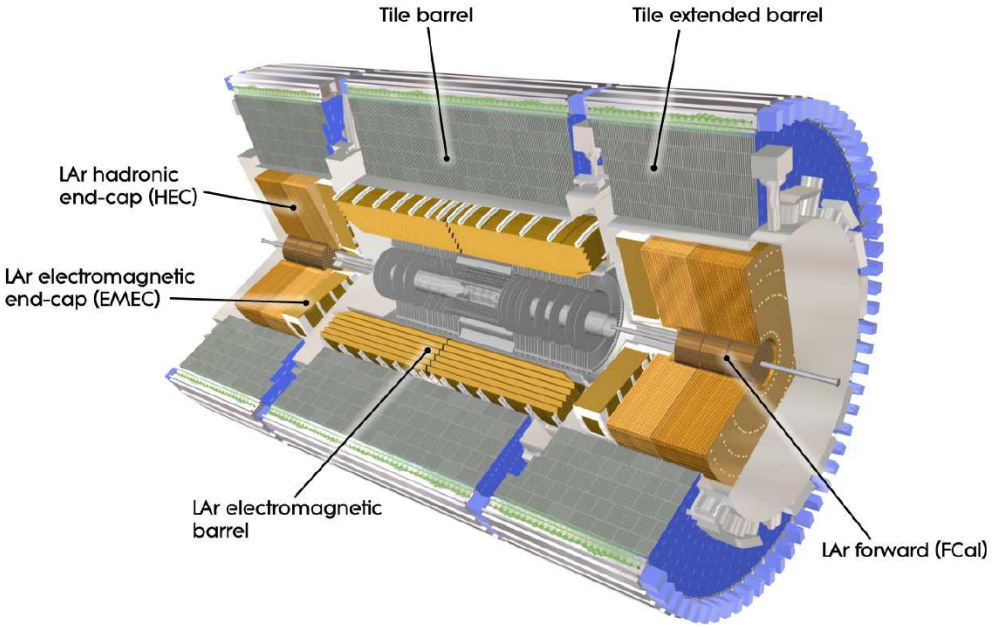


FIGURE 3.5: *Overview over the calorimeter system [32]. Both, the EM- and HAD-calorimeter have a longitudinal an lateral segmentation, with a granularity of* $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ *in most regions. The calorimeters reach out to* $|\eta| < 3.9$ *(in contrast to the inner detector, which is limited to* $|\eta| < 2.5$*).*

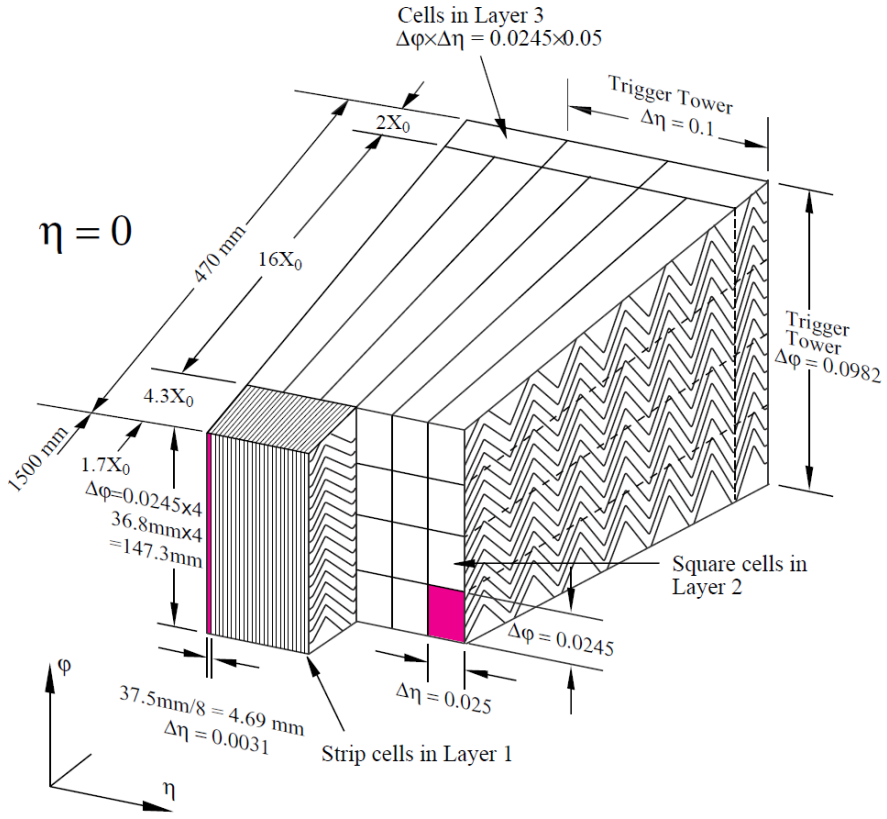


FIGURE 3.6: *Structure of the EM Calorimeter in the barrel region [32]. The longitudinal segmentation is exploited in the new tau reconstruction to estimate energies and therefore the presence of neutral pions. The first layer is particularly usefull when searching for photons from* $\pi^0 \to \gamma\gamma$ *decays in a tau decay, because of its very fine granularity along* η*.*

granularity in ϕ. In the barrel region, a cell in the strip layer has the size $\Delta\eta \times \Delta\phi = 0.003 \times 0.01$. The very fine granularity in this layer can be used to reconstruct individual photons from $\pi^0 \to \gamma\gamma$ processes in τ_{had} decays.

The second layer of the EM-calorimeter has cells of the size $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$. This layer absorbs the largest fraction of the energy of the an incoming π^0 (c.f. Figure 4.9).

With the bulk of energy from an EM shower deposited in EM2, the third layer contains only small energy fractions of those showers. It has a granularity of $\Delta\eta \times \Delta\phi = 0.050 \times 0.025$.

The hadronic calorimeter covers $|\eta| < 4.9$. It is also segmented longitudinally and the thicknesses in interaction lengths λ are $1.5 + 4.1 + 1.8 = 7.4$ ($1.5 + 2.6 + 3.3 = 7.4$) in the barrel (end-caps). The lateral segmentation is $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ for $|\eta| < 2.5$ for the first two layers.

3.2.4 Muon Chambers

With roughly 100 MeV, the muon mass is about 200 times larger than the electron mass ($\approx$ 500 keV). Thus, the muon is much less likely to radiate bremsstrahlung and it will pass the calorimeter without loosing a significant amount of energy. To measure muons, there is an additional tracking system outside the calorimeters, the muon system. The muon system provides another independent measurement for muons, which also produce a track in the inner detector.

Like all the other subsystems, the muon chambers are also divided into a barrel part and the end-caps. Also like the inner detector subsystems, the barrel region consists of cylindrical detector layers, while the end-caps have the form of discs. The muon chambers consist of the *monitored drift tube chambers* (MDTs), *cathode strip chambers* (CSCs), *resistive plate chambers* (RPCs) and *thin gap chambers* (TGCs). The former two are used to precisely measure the position of muons traversing the chambers, the latter two are used to quickly trigger on the presence of muons.

3.2.5 Trigger System

As outlined in Section 3.1, the ATLAS detector needs to select the physically interesting collisions out of all collisions. This is done by the trigger system, which basically flags each event as interesting or not interesting. If the event is flagged as interesting, i.e. it passed the trigger, it will be stored. Otherwise, it is not saved. The decision, whether to save a collision or discard it is based on three so-called trigger levels.

- *Level 1:* The design interaction rate is 40MHz, thus the first level of the trigger system has to make the decision within a very limited time. The desired output rate of the level 1 trigger is 75kHz. It has access to information from the calorimeter and the muon system and bases its decision on particle multiplicities and energy depositions.

- *Level 2:* The second level of the trigger is software-based and uses the full granularity of the detector in regions of the objects found by the first level. In addition to the calorimeter and muon systems, the level 2 trigger can also use information from the tracking system.

- *Event Filter:* Finally, the event filter uses the same algorithms to analyse the objects provided by the level 2 trigger that are also used in the offline event reconstruction. It has access to the complete detector information.

3.3 Simulations

In order to create or improve existing reconstruction algorithms, set out to reconstruct particles as precisely as possible, it is convenient to know how the particle to be reconstructed actually behaves in the detector. Thus, the particles are *simulated* and then passed to the ATLAS event reconstruction, dealing with them as if they were measured particles in an actual data-taking run. There are however differences between simulations and data.

In simulations, there is access to the generated (also called *true*) properties of the particles. These generated properties can then be compared to the properties as reconstructed by the ATLAS reconstruction algorithms. While this comes in very handy in studies into improvements of reconstruction algorithms and into behaviour of particles in the detector, the viability of this approach depends drastically on the accuracy of the simulations. If the simulations do not agree with the actual data, results obtained from simulations are practically worthless[11].

In data, there is no access to the real, true properties of the particles produced in a collision. However, measurements made here are correct in that sense, that they really describe nature[12], and simulations should be tuned such, that they match observations made in data.

The process of simulating data in ATLAS can be divided into two steps. The first step is to generate the hard process, which is done by *event generators*, such as `PYTHIA` [35]. They take into account parton distribution functions, initial state radiation, the matrix element of the hard process, final state radiation and modelling the particle showers due to hadronisation of colour charged particles. The outcome of the event generators is a list of particles produced in the simulated collision.

The second step is then to perform a detector simulation. The list of particles produced in the event generator is passed through a detailed model of the detector, using the `GEANT` toolkit [36]. Here the detector response is simulated, i.e. the interactions between the generated particles and the detector material. The output from this detector simulation ideally has exactly the same format as the data, so that simulation and data can be compared easily and consistently.

[11] This can also be a hint that there may be new physics involved in the dataset that is compared.
[12] Assuming there are no or only known biases introduced by the detector.

CHAPTER 4

τ Lepton Reconstruction

This chapter describes the tau lepton reconstruction in the ATLAS experiment, both the one used in run 1 and the new substructure based reconstruction.

In Section 4.1, which is of technical nature, the naming conventions and short hand notations that will be used to conveniently reference tau leptons and their decays are summarised.

Section 4.2 summarises the tau lepton reconstruction used in run 1, describing its reconstruction and identification step. The fundamental concept which is persued in the substructure based tau reconstruction, namely *energy flow*, is described in Section 4.3. The concept of energy flow needs to be implemented in reconstruction algorithms. In ATLAS reconstruction software, there are two implementations of the energy flow concept and they are presented in Section 4.5. A short summary giving the most important steps of the new approach is given in Section 4.6.

4.1 Notes on Nomenclature

Because of the multitude of decay channels, it is convenient to have a streamlined and standardised nomenclature. The nomenclature in this thesis follows the guideline given in [37]. To avoid the repetition of "visible part of a hadronically decaying τ lepton" and the like, the following abbreviations are used throughout the thesis. Although the list only states shorthand notations for taus with negative charge, they hold true for positively charged taus as well.

- The tau lepton itself is denoted as τ^+ and τ^-, or simply τ if the charge is of no importance.
- Tau leptons decaying via $\tau^- \to \nu_\tau e^- \bar{\nu}_e$ and $\tau^- \to \nu_\tau \mu^- \bar{\nu}_\mu$ are abbreviated by τ_e and τ_μ. A leptonically decaying tau is represented by τ_lep.
- τ_had means a hadronically decaying tau, including the ν_τ, while $\tau_\text{had-vis}$ specifically refers to the visible part of the hadronic decay.
- Since the hadronic tau decays can be subdivided in 1-prong and 3-prong decays, those are also abbreviated as $\tau_\text{1-prong}$ and $\tau_\text{3-prong}$.

In addition to the τ decay in general, the most important hadronic decay modes receive shorthand notations too. They are listed in Table 4.1. The notation combines the number of charged (prongs, p) and neutral (n) pions. Two of the shorthand notations refer to more than one τ_had decay, namely 1pXn and 3pXn. The former refers to all $\tau_\text{1-prong}$ decays with two or more π^0, the latter means all $\tau_\text{3-prong}$ decays with one or more π^0.

τ decay mode	Shorthand notation
$\tau^- \to \nu_\tau \pi^-$	1p0n
$\tau^- \to \nu_\tau \pi^- \pi^0$	1p1n
$\tau^- \to \nu_\tau \pi^- > 1\pi^0$	1pXn
$\tau^- \to \nu_\tau \pi^- \pi^- \pi^+$	3p0n
$\tau^- \to \nu_\tau \pi^- \pi^- \pi^+ > 0\pi^0$	3pXn

TABLE 4.1: *Shorthand notations for the most dominant τ decay modes. The first number indicates the number of $\pi^\pm$ while the second number marks the number of π^0. An X denotes an arbitrary number of π^0, excluding already defined decay modes. These shorthand notations do not change in case of the charge conjugated decays and will be used for the remainder of the thesis.*

4.2 ATLAS τ Reconstruction & Identification in LHC Run 1

The ATLAS tau reconstruction used in run 1 of the LHC is described in [38] and this section briefly summarises the main ideas and concepts given in that document. Because leptonic τ decays into electrons or muons can hardly be separated from direct electron or muon production [9], only hadronically decaying taus are considered. The main experimental challenge is to separate the latter from QCD induced jets. An illustration of a tau decay and a QCD jet with a short description of the differences is given in Figure 4.1.

The ATLAS tau reconstruction is divided into two parts. The first part is the reconstruction, which combines detector responses into a τ candidate. The second part is the identification step. Here, the τ candidate is separated from QCD jets, muons and electrons.

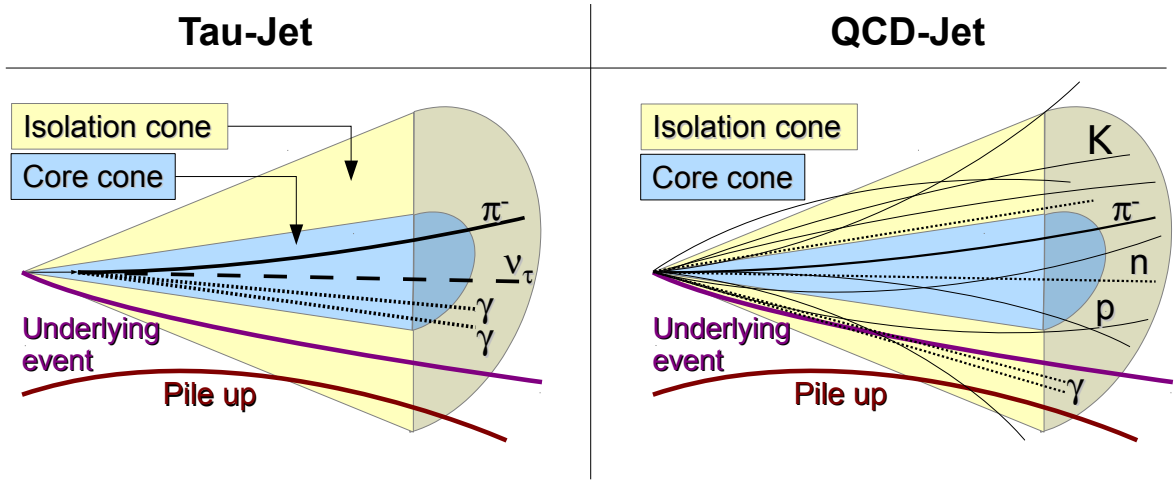


FIGURE 4.1: *Illustration of a hadronic tau decay and a QCD-induced jet. The blue cone is the cone in which the tau decay products are expected, the yellow cone is an area in which no particles are expected in case of a tau decay. This isolation, in addition to the collimation of the tau decay products form the main source of separation from QCD-induced jets.*

Reconstruction of hadronic τ decays

A more detailed description can be found in section 2 of [38], which this section is a summary of. The $\tau_{\text{had-vis}}$ reconstruction considers every jet-object with $E_T > 10\,\text{GeV}$ and $|\eta| < 2.5$ as a τ candidate. Jet-objects are found by combining topoligical clusters[1] using the Anti-Kt algorithm with a distance parameter of 0.4. Before associating tracks or clusters to the candidate, it is searched for the vertex the candidate originates from. The correct vertex association is important, because the track selection and directional information of clusters is calculated with respect to the associated vertex.

In order to find the correct vertex, a measure called the jet-vertex-fraction (JVF) is utilised. A jet has a set of tracks associated to it, and each of these tracks is associated to a vertex it originates from. The jet-vertex-fraction is defined for a given jet and a given vertex as:

$$f_{JVF}(\text{Jet, Vertex}) = \frac{\sum_{\text{Tracks in Jet}} p_T^{\text{Track, Vertex}}}{\sum_{\text{Tracks in Jet}} p_T^{\text{Track}}} \tag{4.1}$$

Thus, given an event with n jets and m vertices, the JVF can be calculated m times for each of the jets. A high JVF of a jet for a given vertex indicates that the majority of the summed track momenta of the jet comes from that one vertex. In other words, the jet is composed of tracks that originate from that vertex, i.e. the jet itself was produced in that vertex. A small JVF indicates that this is not likely to be the case. Thus, the correct tau vertex is assumed to be the vertex giving the highest JVF for the seed jet of the tau candidate.

Because the $\tau_{\text{had-vis}}$ reconstruction is seeded from jets, the clusters of the τ_{had} candidate are the same as the clusters of the seed jet. A first measure of the $\tau_{\text{had-vis}}$ four-momentum is obtained by calculating the

[1] Topological clusters: These are built from individual calorimeter cells in the following way. Seed for a cluster is the cell with the highest energy above 4-times the noise threshold (σ). Around this cell, it is searched for neighbouring cells that have energies exceeding $2 \cdot \sigma$. These cells are added to the cluster and it is again searched for neighbours with energies above $2 \cdot \sigma$. This is repeated until there are no more neighbouring cells fulfulling $E > 2 \cdot \sigma$. As a final step, all cells adjacent to the cluster are added to also account for tails of the energy deposition.

barycenter of the associated clusters. This four-momentum is then refined by excluding clusters outside a ΔR cone of 0.2 around the barycenter. To finally obtain the η and ϕ values of the four-momentum, the remaining clusters are corrected with respect to the associated vertex and summed up. The axis defined by this η and ϕ is called the *intermediate axis* and will also be referred to as the $\tau_{\text{had-vis}}$ axis. The final energy is given as the result of an additional calibration procedure which exploits the fact that τ_{had} decays only cover a limited range in $\pi^{\pm}$ and π^0 multiplicity. A final correction for η is applied to correct for poorly instrumented detector regions.

Thus, the four-momentum of the $\tau_{\text{had-vis}}$ is calculated by using calorimeter measurements.

In a ΔR cone of 0.2 around the intermediate axis, tracks are associated to the $\tau_{\text{had-vis}}$ candidate if they fulfill these requirements:

- $p_T \geq 1$ GeV
- $N_{\text{Pixel Hits}} \geq 2$
- $N_{\text{Pixel Hits}} + N_{\text{SCT Hits}} \geq 7$
- $d_0 \leq 1$mm, d_0 is the distance of closest approach between the track and the associated $\tau_{\text{had-vis}}$ vertex in the transverse plane.
- $|z_0 \cdot \sin\theta| \leq 1.5$mm, z_0 is the analogon to d_0 along the beam axis.

Tracks fulfilling these criteria are used to classify the $\tau_{\text{had-vis}}$ as one-prong (i.e. one such track is found) or multi-prong (i.e. two or more such tracks are found). In addition, tracks within $0.2 < \Delta R < 0.4$ passing the requirements listed above are saved. The region of $0.2 < \Delta R < 0.4$ is also called the *isolation annulus* and is important for $\tau_{\text{had-vis}}$ identification. Tracks found in this region are used to evaluate the isolation of a tau candidate.

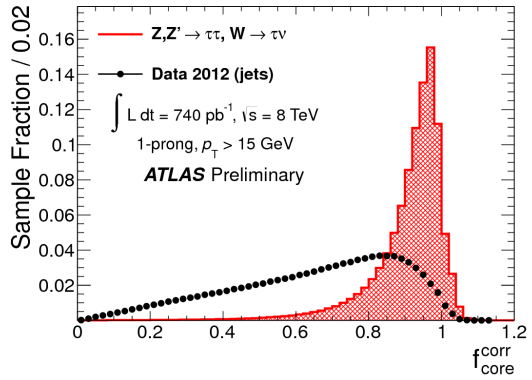
Based on the associated clusters and tracks it is possible to calculate variables that can separate $\tau_{\text{had-vis}}$ leptons from jets.

Tracking-based variables are quite pile-up robust, variables that use calorimeter-based information need to be corrected for pile-up effects[2]. Pile-up leads to an overestimation of the energy and to additional clusters.

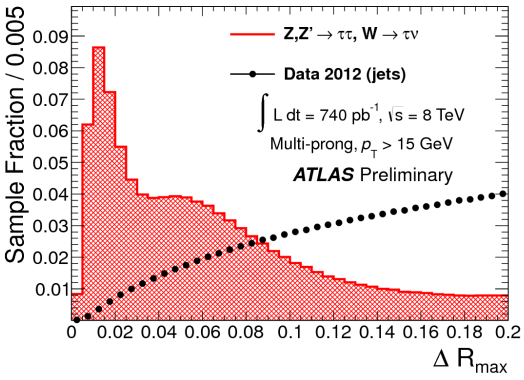
To correct for these pile-up effects, only clusters within $\Delta R \leq 0.2$ are considered in the calculation of calorimeter-based variables. These variables are also corrected for pile-up effects as a function of N_{Vertex}. This additional correction is only applied below a p_T threshold, above which the pile-up effects become less severe. The discrimination between taus and jet in the region $0.2 < \Delta R < 0.4$ is retained by the tracking-based variables.

Two examples of discriminating variables are shown in Figure 4.2. Figure 4.2a shows $f_{\text{core}}^{\text{corr.}}$, the fraction of the $\tau_{\text{had-vis}}$ energy in a ΔR cone of 0.1 around the $\tau_{\text{had-vis}}$ axis. Because on average the decay products of a $\tau_{\text{had-vis}}$ are confined to a smaller area than jets, they deposit their energy in a smaller cone. In other words, the $\Delta R = 0.1$ cone captures most clusters produced by a τ, but only a fraction of clusters produced by a jet.

[2] Pile-up: Due to the high luminosities at the LHC, there are more than one proton-proton interactions per bunch-crossing. These additional interactions are source of additional particles that overlap with the interesting (hard) collision. For more details, see the end of Section 3.1

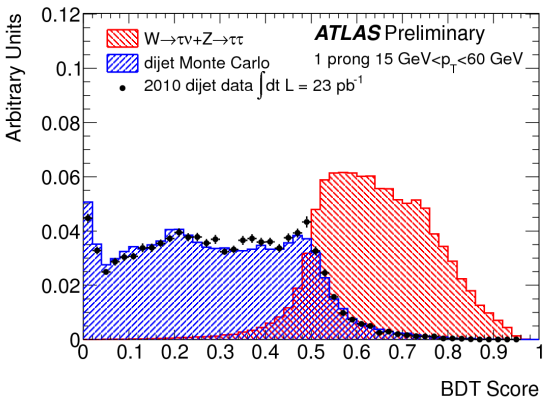


(A) $\tau_{had\text{-}vis}$ *energy fraction in* $\Delta R = 0.1$ *cone (pile-up corrected)*

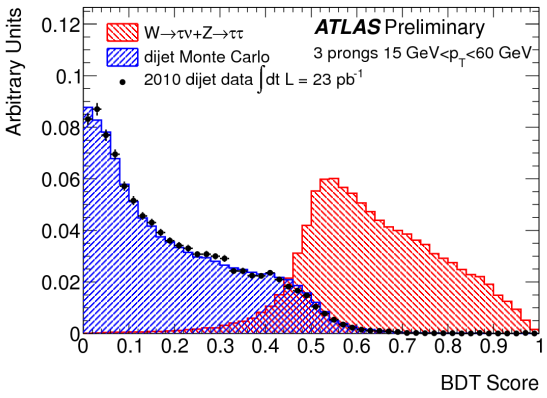


(B) *Largest distance in* $\eta - \phi$ *plane between a track and the* $\tau_{had\text{-}vis}$ *axis*

FIGURE 4.2: *Examples of calorimeter-based (left) and tracking-based (right) variables to separate* $\tau_{had\text{-}vis}$ *from jets. Histograms are normalised to unity. Both pictures taken from [38].*



(A) *1-prong* $\tau_{had\text{-}vis}$ *candidates*



(B) *Multi-prong* $\tau_{had\text{-}vis}$ *candidates*

FIGURE 4.3: *The BDT distributions as obtained from evaluation of the BDT on signal (red) and background (blue, black dots) test samples. All curves are normalised to unity to compare the shapes. By requiring a* $\tau_{had\text{-}vis}$ *candidate to have a BDT score larger than a given threshold, signal and background efficiencies can be determined. Because the description of* $\tau_{had\text{-}vis}$ *identification in [38] does not provide the BDT distributions, these figures are taken from [39]. As a result, the* p_T*-binning is different and the amount of data is more limited. However, the way in which these distributions are used remains unchanged.*

In Figure 4.2b, ΔR_{Max} is depicted. ΔR_{Max} is the largest ΔR value found in a multi-prong $\tau_{\text{had-vis}}$ candidate and is calculated as the distance in the $\eta - \phi$-plane between a track and the $\tau_{\text{had-vis}}$ axis. The narrow nature of τ leptons is also visible here, with $\tau_{\text{had-vis}}$ having smaller values of ΔR_{Max} than jets.

Identification of hadronic τ decays

This section is a summary of section 3 from [38]. The set of objects that is reconstructed as described above will contain a large amount of QCD-induced jets in addition to the taus that are of interest. Because QCD-induced jets are created at much higher rates at the LHC than processes that contain tau leptons (c.f. Figure 3.2), it is necessary to implement an identification, which every object from the set is put through and classified by as signal (i.e. real tau) or background (i.e. jet).

This identification is based on the discriminating variables (two examples of which are shown in Figure 4.2). There are several of these variables and to also account for correlations between them, they are combined into a boosted decision tree (BDT). Individual BDTs are trained for one-prong and three-prong candidates. The 1-prong BDT is trained and evaluated on true 1-prong $\tau_{\text{had-vis}}$, while the 3-prong BDT is trained and evaluated on true 3-prong $\tau_{\text{had-vis}}$ which are reconstructed with two or more tracks. Both BDTs are trained using the TMVA package [40].

The training is performed by using simulated signal $\tau_{\text{had-vis}}$ from $Z \to \tau\tau$, $W \to \tau\nu_\tau$ and $Z' \to \tau\tau$. In the Z' samples, the simulated Z' mass ranges from 250 GeV to 1250 GeV to increase the number of $\tau_{\text{had-vis}}$ candidates with high p_T.

Only $\tau_{\text{had-vis}}$ candidates with $|\eta| < 2.3$, $p_T > 15$ GeV and $\Delta R(\tau_{\text{had-vis}}, \text{true } \tau) < 0.2$ are used in the training process. The associated true τ needs to fulfill $|\eta^{\text{true}}_{\text{vis}}| < 2.5$, $p^{\text{true}}_{\text{T, vis}} > 10$ GeV and $N^{\text{true}}_{\text{prong}} \in \{1; 3\}$.

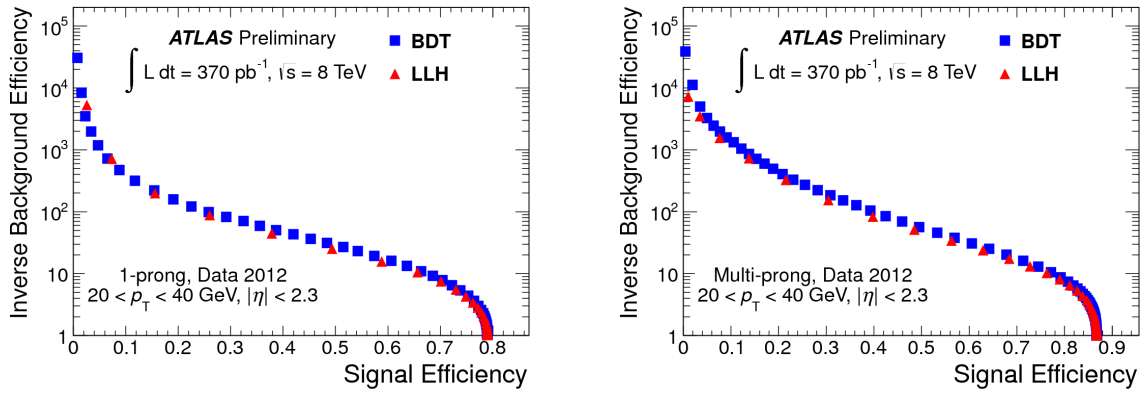
A data sample from 2011 is used to obtain QCD jets. These jets need to pass the same cuts as the $\tau_{\text{had-vis}}$ candidates.

After training and application to a test sample, BDT output distributions (also called BDT score) similar to the ones shown in figure 4.3 are obtained. By applying a cut on the BDT score, the signal (background) efficiencies can be calculated by using the fraction of $\tau_{\text{had-vis}}$ candidates in the signal (background) sample that pass the cut. The signal efficiency regarding reconstruction and identification is defined as the number of truth-matched $\tau_{\text{had-vis}}$ candidates that pass identification, divided by all truth-matched $\tau_{\text{had-vis}}$ candidates. The background efficiency is defined as the number of background $\tau_{\text{had-vis}}$ candidates that pass identification, divided by the number of all background $\tau_{\text{had-vis}}$ candidates. By applying a cut in every bin of the BDT score distributions and recording the resulting signal and background efficiencies, the ROC (Receiver Operation Characteristic) curve is retrieved. In this case however, the inverse background efficiency is plotted.

It is shown in Figure 4.4. Based on the ROC curves, three working points at fixed signal efficiencies are defined. The loose, medium and tight identification working points have signal efficiencies of 70%, 60% and 40% in the 1-prong case and 65%, 55% and 35% in the multi-prong case. The three working points have been shown to be stable with respect to the number of vertices in the event. While the signal efficiencies do not depend on the p_T of the true $\tau_{\text{had-vis}}$, background efficiencies tend to drop with increasing candidate p_T, but they are also dependent on the actual event selection in an analysis. This is shown in Figure 4.5

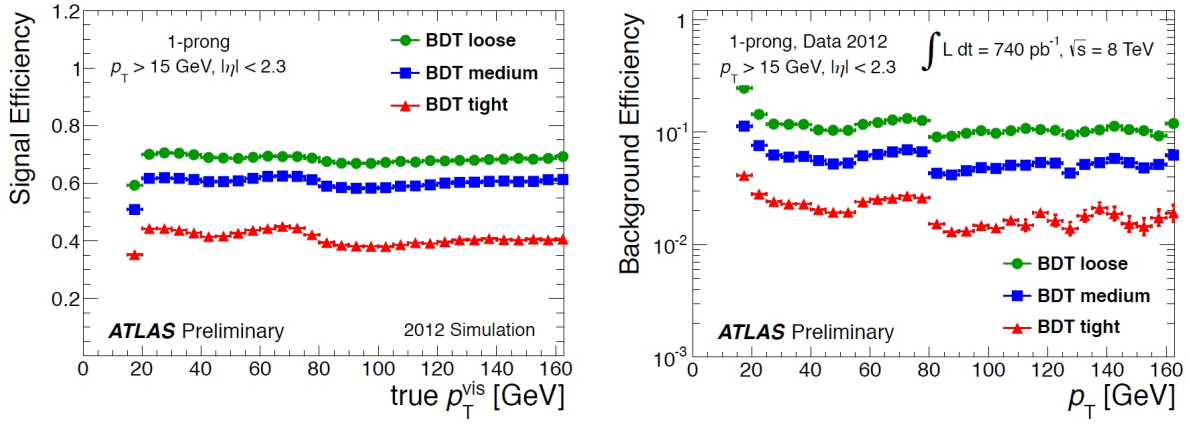
For completion it is mentioned that for 1-prong $\tau_{\text{had-vis}}$ candidates there is also a discrimination against electrons and muons.

Discrimination against electrons is achieved by exploiting the typical shower shapes of electrons



(A) *1-prong $\tau_{had\text{-}vis}$ candidates* (B) *Multi-prong $\tau_{had\text{-}vis}$ candidates*

FIGURE 4.4: *Inverse background efficiency as a function of signal efficiency for 1-prong (left) and multi-prong (right) $\tau_{had\text{-}vis}$ candidates in the lower p_T-bin. Figures are taken from [38].*



(A) *Signal efficiency* (B) *Background efficiency*

FIGURE 4.5: *Signal and background efficiencies as a function of tau p_T. Dips and jumps at $p_T = 80$ GeV are due to the fact that the pile-up correction is turned off when exceeding $p_T = 80$ GeV. Figures are taken from [38].*

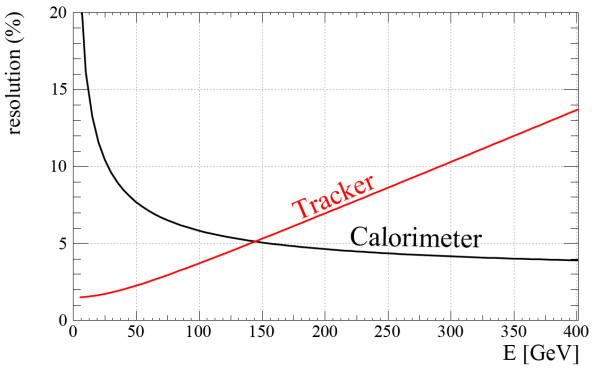


FIGURE 4.6: *Relative energy resolution of calorimeter and tracker as a function of the energy. For energies below $E \approx 140$ GeV, the resolution of the tracker is better than the resolution of the calorimeter. Figure taken from [41].*

(narrow) and charged pions (wider) and transition radiation which happens more likely for electrons due to the lower mass and thus a higher γ-factor. Similar to the discrimination against jets, a BDT which uses variables quantifying the differences is trained to separate τ_e from $\tau_{1\text{-prong}}$.

True muons, that are reconstructed as $\tau_{1\text{-prong}}$ fall into one of these categories.

- *Muons in ineffective regions of the Muon Spectrometer:*
  There is a gap in the muon coverage at $\eta = 0$ and $\eta = 1.3$, so that muons have a lower reconstruction efficiency in these regions.
- *Muons not reaching the muon system because of their low energy:*
  These muons either lost their energy due final state radiation or have a low energy to start with. This means they overlap with an dominantly electromagnetic shower, or the ratio of track momentum and calorimeter energy is smaller than expected for $\tau_{\text{had-vis}}$.
- *Muons depositing a large fraction of energy in the calorimeter*:
  The energy is mostly deposited in the hadronic calorimeter or in the electromagnetic calorimeter in case of bremsstrahlung. The ratio of energies in the electomagnetic and hadronic calorimeter is therefore different from what is expected for $\tau_{\text{had-vis}}$ decays, which have a larger electromagnetic component, or a lower component, in case of a muon with final state radiation.

4.3 The Particle Flow Concept

The $\tau_{\text{had-vis}}$ reconstruction measures the direction and energy of the $\tau_{\text{had-vis}}$ in the calorimeter and does not rely on track momenta (c.f. Section 4.2). In Figure 4.6, the energy resolution of the calorimeter is depicted as a function of particle energy. The figure also shows the momentum resolution of the tracker. The resolutions are obtained by following the parametrisations ([41] and [32], page 5, 304), with $\oplus$ denoting addition in quadrature:

$$\sigma(1/p_T) = \frac{0.34}{\text{TeV}}(1 \oplus \frac{44\ \text{GeV}}{p_T}) \qquad \text{Tracking} \tag{4.2}$$

$$\sigma(E)/E = \frac{50\%}{\sqrt{E/\ \text{GeV}}} \oplus 3\% \qquad \text{Had. calorimetry} \tag{4.3}$$

Up until $E \approx 140$ GeV, the resolution of the tracker is better than the resolution of the calorimeter. The lower the energy of the particle to be measured, the larger the gain is when using the tracker measurements. With energies rising above $E \approx 140$ GeV the calorimeter outperforms the tracker in terms of energy resolution. This is because the track has a smaller curvature, so that the calculation of the momentum becomes less precise. At the same time, with higher energies, more particles are produced in the shower in the calorimeter and the energy resolution gets better.

In τ_{had} decays, at least two particles are produced. However, most of the times, the τ decays into three (dominantly, because of the 1p1n branching ratio) or even four (in the 1pXn, 3p0n modes) particles. This makes the decay products quite soft, even though the τ_{had} may have had an energy of about (for instance) 300 GeV, because that energy is distributed among all the decay products. This renders the decay products rather soft, so that they fall into that part of Figure 4.6, in which the tracker performs significantly better than the calorimeter[3].

[3] This is specifically true in case of a Higgs decay into two taus, because of the rather small Higgs mass of $m_H \approx 125$ GeV, giving rise to two taus with $E_T \approx 60$ GeV, having decay product energies far below 140 GeV.

However, it is not possible to simply replace the energy measurement from the calorimeter with the energy measurement in the tracker (assuming tracks are due to $\pi^{\pm}$), because there may have been π^0 in the decay which contribute to the energy in the calorimeter. Thus the energy would be underestimated in the majority of τ_{had} decays, if only the tracks were considered.

To account for this, but still make use of the good tracker resolution, the $\tau_{\text{had-vis}}$ energy can be expressed as[4]

$$E(\tau_{\text{had-vis}}) = E(\text{EM-Calo}) + E(\text{HAD-Calo}) \tag{4.4}$$
$$= E(\text{Tracks}) + (E(\text{EM-Calo}) + E(\text{HAD-Calo}) - E(\text{Tracks})) \tag{4.5}$$
$$= E(\pi^{\pm}\text{ s}) + E(\text{Neutral objects}) \tag{4.6}$$

I.e. the energy due to neutral particles is estimated as the overshoot in calorimeter energy with respect to track energy. Obviously, from E(Neutral objects) it is very hard to tell how many particles contribute to the energy and whether those particles are in fact π^0, or whether they belong to the tau at all.

Thus, it is clear that a more sophisticated approach is needed to utilise the tracker resolution to improve the momentum resolution of $\tau_{\text{had-vis}}$ candidates.

What is missing in the idea above is an *association* of tracks and clusters. For simplicity, it is assumed that this association works perfectly. With such an association - a link between a track and clusters *belonging* to (or rather created by the same particle that produced) the track - it is possible to remove clusters from the calorimeter. After the removal of clusters belonging to tracks, all clusters that are left in the calorimeter are coming from neutral particles. Now, the number of clusters can be counted and by studying their shapes it is possible to identify them as coming from different particles.

This approach, subtracting the estimated energy deposits of charged particles from the calorimeters, is known as the *Energy Flow* concept and is not τ-specific, i.e. it can also be applied in jet reconstruction.

4.4 Substructure Based τ Reconstruction

Substructure Based τ Reconstruction describes the concept and the set of algorithms that will be used in ATLAS to reconstruct $\tau_{\text{had-vis}}$ in the upcoming data-taking period of the LHC. It applies the energy flow concept in a three step approach: The seeding of the tau candidate, the so-called substructure algorithm, which creates the basis of a substructure based reconstruction, and a tau-kinematics-based algorithm called *PanTau*.

1. *Seeding:*
   The seeding of τ_{had} remains basically unchanged. All reconstructed jets with $E_T > 10\,\text{GeV}$ and $|\eta| < 2.5$ are considered as a τ_{had} candidate. The addition is, that the number of tracks (N_{Tracks}) associated to the τ_{had}[5] needs to fullfill $1 \leq N_{\text{Tracks}} \leq 5$.

2. *Substructure Algorithm:*
   In the so-called *Substructure Algorithm*, the energy flow concept is applied. There are two implementations of the concept in the ATLAS software and they are described in the next section. While they differ in details, the major functionality is the same and can be divided into three parts:

[4] E(Tracks) is estimated by the product of the measured track momentum, and an E/p factor. This factor relates the track momentum with the energy that a track of a given momentum typically deposits in the calorimeter. However, the purpose of the equation is merely to illustrate how to include tracking measurements into the combined measurement of the visible tau.

[5] The selection and association of the tracks is performed as described in Section 4.2

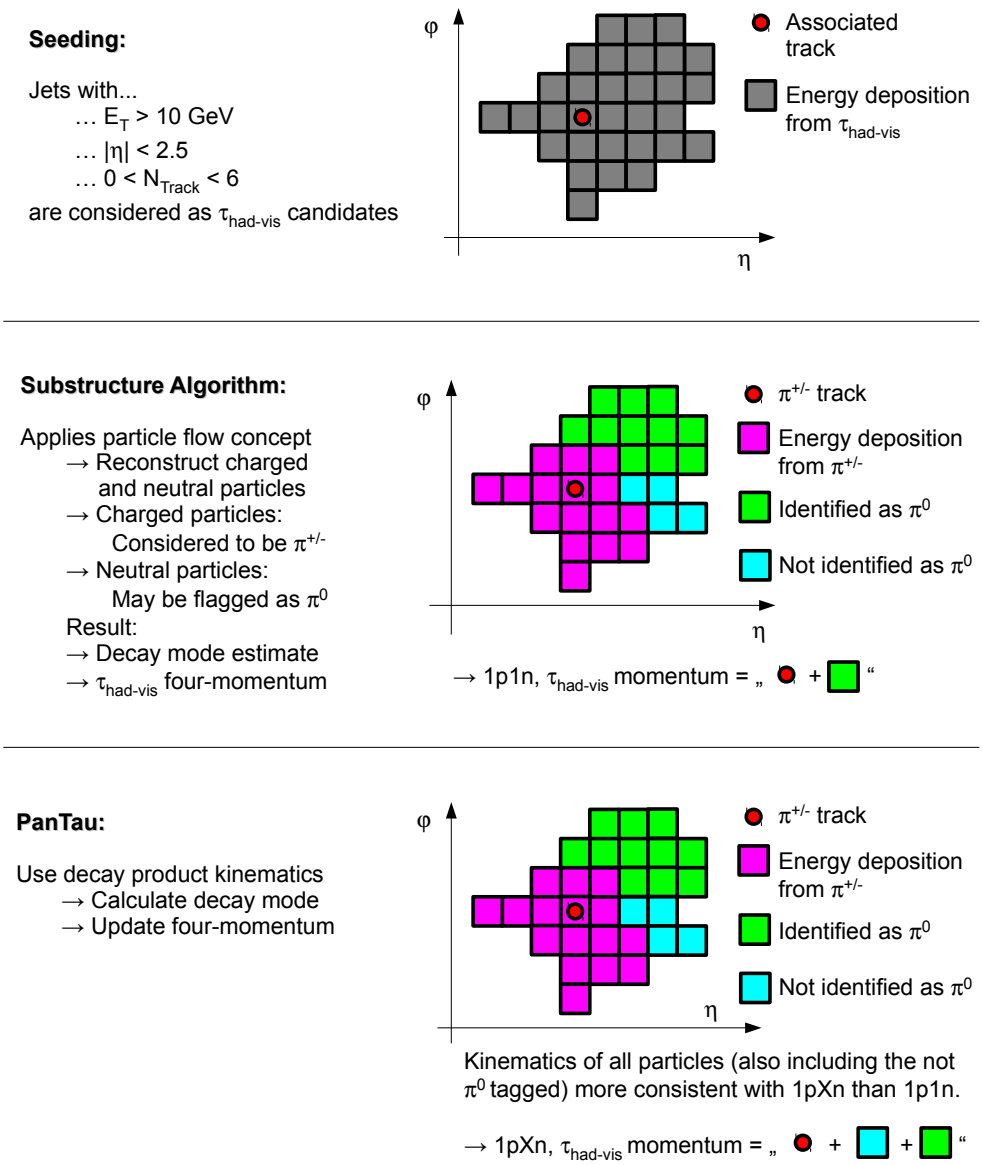


FIGURE 4.7: *Illustration of the workflow in substructure based τ reconstruction. Accuracy in every detail is not given, to conveniently convey the idea of the new reconstruction. The left column briefly recaps which tasks the three steps perform. The right column shows how the τ candidate can be thought of in the η − ϕ-plane in the second layer of the EM calorimeter.*
After the seeding, all energy depositions and tracks combined form the $\tau_{had\text{-}vis}$.
The substructure algorithms are capable of telling apart energy depositions from $\pi^{\pm}$ (magenta) and neutral particles. The latter can be identified as π^0 (green) or non-π^0 (blue). Counting the objects gives the decay mode of 1p1n (in this example) and the four-momentum.
PanTau investigates the kinematics of the decay and compares to what is expected from a 1p1n decay with one additional non-π^0 object. The result may be, that the kinematics are more consistent with a 1pXn decay. Hence the decay mode and four-momentum is updated accordingly.

The first step in the substructure algorithm is to reconstruct individual particles in a $\tau_{\text{had-vis}}$, by following the energy-flow concept. All charged particles are considered to be $\pi^{\pm}$.[6]

In a second step, the neutral particles are passed into a dedicated π^0 identification[7], which separates π^0 from noise, remnants of imperfect $\pi^{\pm}$ subtraction, and pile-up particles.

Finally, the third step is to calculate the four-momentum of the $\tau_{\text{had-vis}}$. It is calculated as the sum of the four-momenta of the charged particles and the neutral particles identified as π^0. By counting the number of $\pi^{\pm}$ and π^0, the decay mode is reconstructed too.

3. *Tau-kinematics-based algorithm (`PanTau`):*
   The `PanTau` algorithm is executed after the substructure algorithms to perform an analysis of the decay mode and the four-momentum of the $\tau_{\text{had-vis}}$. To do so, it uses the kinematics of all particles in the $\tau_{\text{had-vis}}$ decay found by the substructure algorithm in combination with the estimated decay mode based on counting the number of $\pi^{\pm}$ and π^0. More details are given in chapter 5.

A schematic description of the substructure based τ reconstruction is given in Figure 4.7.

4.5 Particle Flow Implementations in ATLAS

In the ATLAS software, there are currently two implementations of a energy flow approach to tau reconstruction. The `eflowRec` algorithm was developed as a generic implementation of the energy flow concept and has a special tune to also perform well in case of τ leptons. A τ-specific implementation of the concept is realised in the `CellBased` algorithm. This section will give an overview over these two approaches to energy flow based tau reconstruction. The `CellBased` part is based on [42], the `eflowRec` part is based on [43]. In the following two subsections, these two approaches are presented.

There is yet another approach to substructure based tau reconstrucion, called *cluster-based*. That approach however is not subject of this thesis and is only mentioned for completeness. Details about it can be found in [44].

4.5.1 The `CellBased`-Algorithm

The `CellBased` algorithm consists of two separate parts: The first part is responsible for removing the hadronic shower produced by charged pions and the second part identifies neutral pions in the remaining calorimeter deposits.

Subtraction of the hadronic shower

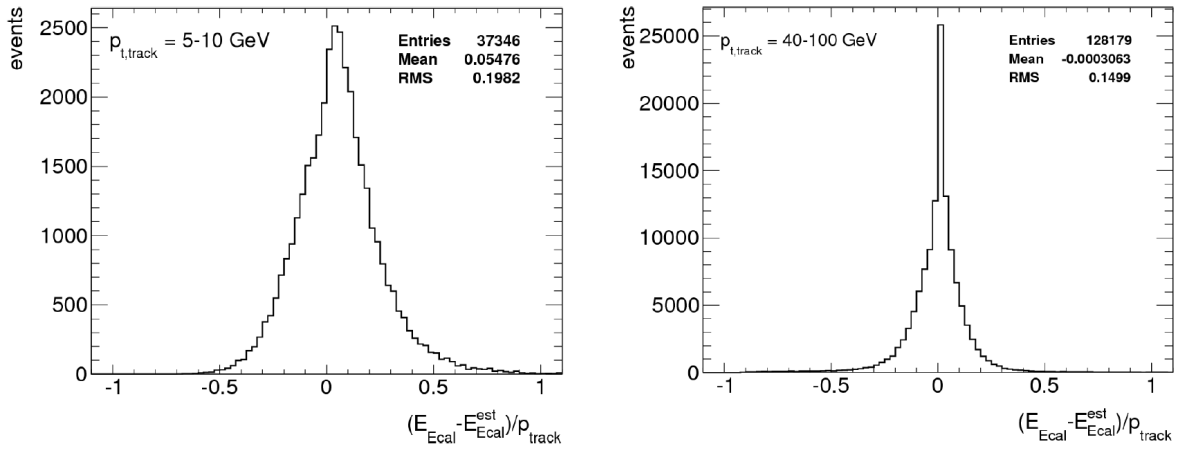
The subtraction is performed in three successive steps:

1. Estimate the total amount of energy deposited by the charged pion in the electromagnetic calorimeter.

2. Distribute the total energy to each layer of the electromagnetic calorimeter. This gives an estimate of how much energy needs to be subtracted in each layer.

3. For each layer, distribute the estimated energy to the individual cells.

[6] The branching ratios for decay modes involving Kaons are small enough to be neglected at first order.
[7] Again, decay modes with neutral particles other than π^0 are neglected.



(A) *Low charged pion momenta* (B) *High charged pion momenta*

FIGURE 4.8: *Resolution of the estimation of the total energy in the EM calorimeter for charged pions with low (left) and high (right) transverse momentum. The accuracy of the estimation increases because of the accuracy of the calorimeters increases with the deposited energy. Figures are taken from [42].*

In order to perform the longitudinal and lateral energy distribution in steps two and three, averaged hadronic shower shapes are used. In the barrel and in the endcap region, the subtraction is performed with an adjusted configuration.

To estimate the total amount of energy to be subtracted in the EM calorimeter, the formula ([42], Equation 4.1)

$$E_{\mathrm{ECal}}^{\mathrm{estimate}} = p_{\mathrm{Track}} - E_{\mathrm{HCal}} \tag{4.7}$$

is used. p_{Track} is the track momentum, and E_{HCal} is the energy in the hadronic calorimeter. The calculation of E_{HCal} takes into account all cells in a ΔR cone of 0.2 that belong to a topological cluster. By only considering cells that belong to a cluster, noise is surpressed. The resulting E_{HCal} is corrected for energy losses the barrel cryostat, which is positioned between the EM barrel and the tile hadronic calorimeter. Because hadronic showers are subject to rather large fluctuations, it may be the case that the estimated value of E_{HCal} is larger than the measured p_{Track}. This can happen for hadronic showers that deposit most of their energy in the hadronic calorimeter and would result in a negative energy to be subtracted. If this case occurs, $E_{\mathrm{ECal}}^{\mathrm{estimate}}$ is considered to be zero and no energy is subtracted.

Figure 4.8 shows the resolution of the energy estimation. For these plots, a sample of single $\pi^{\pm}$ was used, hence the estimated energy in the EM calorimeter should equal the deposited energy. For charged pions with high p_T, the estimation is more precise, because the calorimeter resolution improves. This is especially helpful in the case of high energetic τ_{had}, because in those decays the pions also have high p_T and tend to overlap with each other.

The distribution of the total estimated hadronic energy to the different layers of the EM calorimeter (second item in the enumeration above) is based on the so-called longitudinal weights. $E_{\mathrm{ECal}}^{\mathrm{estimate}}$ needs to be distributed to three layers: The presampler, the first and the second EM layer. The longitudinal weight of the presampler (for instance) gives the fraction of $E_{\mathrm{ECal}}^{\mathrm{estimate}}$ that is to be assigned to the presampler

layer. The weight w_l for a given layer l is calculated according to this formula:

$$w_l = \frac{\sum_{\pi^\pm} \sum_{i \in I_l} c_i E_i}{\sum_{\pi^\pm} E_{\mathrm{ECal}}} \tag{4.8}$$

E_i is the raw energy of cell i and c_i is its calibration constant. i runs over the set I_l which contains all cells in clusters in the layer l within $\Delta R < 0.2$ around the charged pion track. The outer sum runs over all $\pi^\pm$ in the sample.

The calculation of the w_l is done in each bin of a three dimensional binning which is introduced to reduce susceptibility to shower-by-shower fluctuations in the hadronic shower. The binning is done in $p_{\mathrm{T,\,Track}}$, η_{Track} and the hadronic energy fraction[8] f_{had} of the shower.

Examples of the longitudinal weights are shown in Figure 4.9. The p_T of the $\pi^\pm$ has the largest impact. With increasing p_T the showers become longer and hence have larger weights in EM layer 2. Showers with a higher hadronic fraction tend to start showering later, hence their weight in EM layer 2 is a bit larger.

To finally distribute the energy in a given layer to the cells, lateral weights are used. These describe how the energy is distributed around the track, in the $\phi - \eta$ plane. They are obtained similarly to the longitudinal weights and are calculated in the same three dimensional binning. A fit of two dimensional exponential functions is performed in each bin. To calculate the amount of energy to be subtracted for a given cell, the fitted two dimensional function is integrated over the corresponding cell area. Figure 4.10 shows an example of the lateral energy distribution and the fitted exponential functions.

At the time of writing reference [42], the `CellBased` algorithm did not subtract energy for $\tau_{\text{3-prong}}$ and the subtraction in the transition region of electromagnetic and hadronic calorimeter was in a preliminary state. However, in the beginning of 2013, the algorithm has been expanded to work properly in all detector regions ($|\eta| < 2.5$, because the subtraction needs a measured track) and on all τ_{had} decays with $N_{\mathrm{track}} \in [1, 5]$.

The estimation and distribution of the energy in the EM calorimeter to the layers and the cells is illustrated in Figure 4.11.

Identification of neutral pions

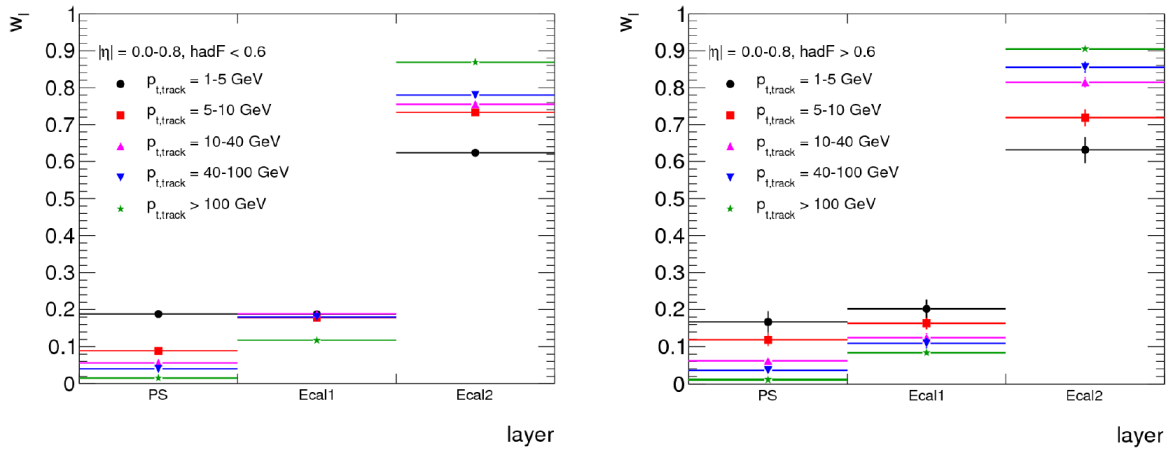
After the subtraction of the hadronic shower in the EM calorimeter, the remaining energy is reclustered using the ATLAS topological clustering algorithm, because the cell energies have been modified to correspond to the situation in which the $\pi^\pm$ did not shower in the EM calorimeter. In an ideal case, all the clusters left after subtraction are due to neutral pions. Neutral pions can produce three types of clusters:

- Clusters containing one of the photons from the π^0 decay, this can happen for low energy π^0.
- Clusters that contain both photons from $\pi^0 \to \gamma\gamma$.
- Clusters consisting of several photons from π^0 decays.

However, there are also other sources for clusters after the subtraction:

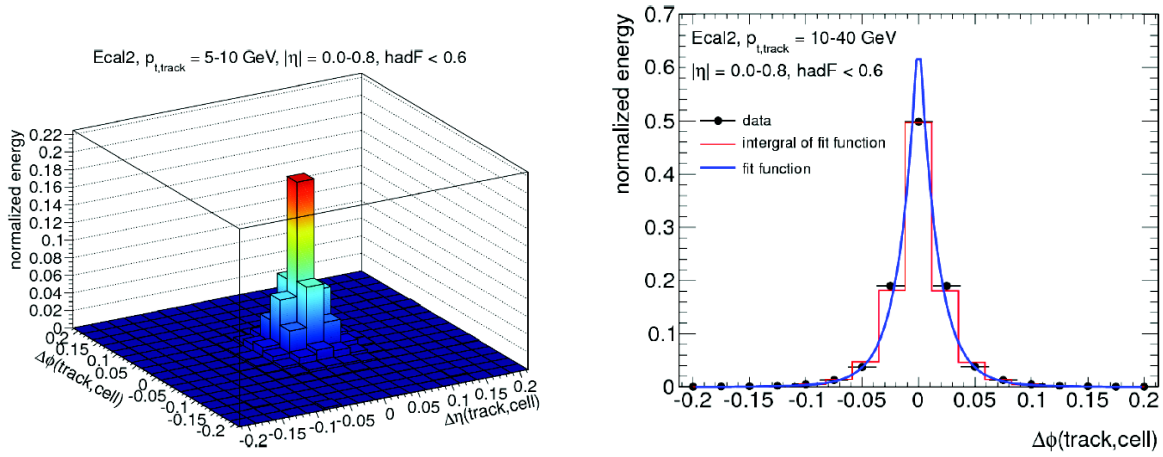
- Non ideal subtraction, i.e. the energy deposition of the $\pi^\pm$ in the EM calorimeter is underestimated and not enough energy is subtracted.

[8] The hadronic energy fraction is the ratio of energy of the $\pi^\pm$ that is deposited in the hadronic calorimeter and the momentum of the $\pi^\pm$ track.



(A) *Small hadronic energy fraction ($f_{had} < 0.6$)* (B) *High hadronic energy fraction ($f_{had} > 0.6$)*

FIGURE 4.9: *Longitudinal weights for the presampler, layer 1 and 2 of the electromagnetic calorimeter for $\eta \in [0.0, 0.8]$. The left (right) plot shows the longitudinal weights for a small (large) hadronic fraction. Figures taken from [42].*



(A) *Lateral energy distribution around track* (B) *Fitted exponential function in ϕ direction*

FIGURE 4.10: *The left figure shows an example for the lateral energy distribution around the $\pi^{\pm}$ track. These distributions are fitted with a sum of two dimensional exponentials. The ϕ projection of such a fit is depicted in the right figure. Both figures taken from [42].*

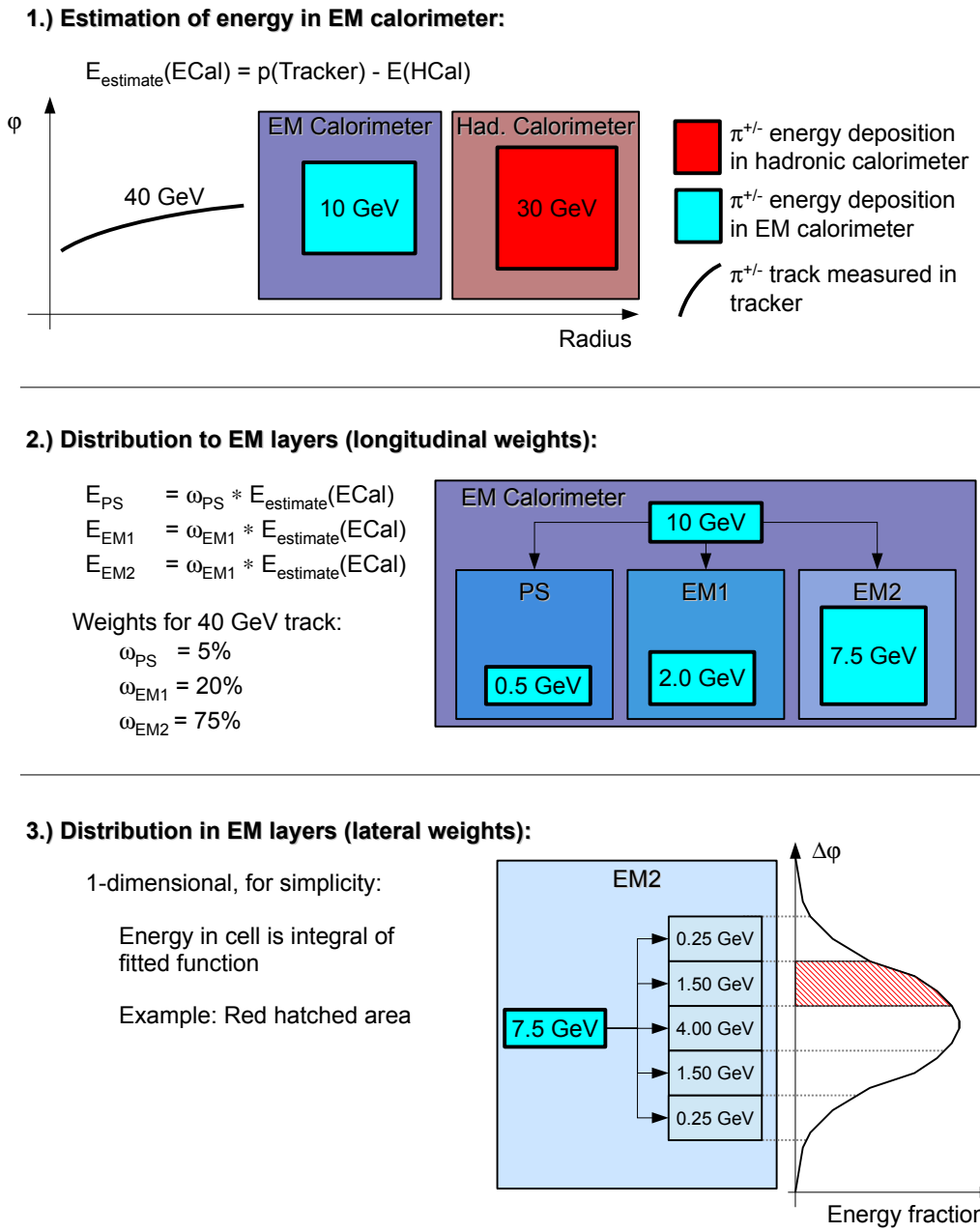


FIGURE 4.11: *Drawing to visualise the three steps that are performed in the* `CellBased` *algorithm to distribute the energy to the cells.*
First, the $\pi^{\pm}$ energy in the EM calorimeter is estimated by using the track measurement and the energy in the hadronic calorimeter. This estimate is then distributed to each layer of the EM calorimeter using the longitudinal weights ω. By integrating the 2-dimensional fit function (here: 1-dimensional for illustration purposes) over the cell area, the amount of energy to subtract is calculated.
Sizes of the EM and hadronic calorimeter, as well as the layers and cells are not to scale.

- Clusters due to other particles, coming from pile-up.
- Noise clusters.

The task of the identification is to separate clusters of true neutral pions from clusters due to the other sources (background clusters). Before the identification, a preselection is applied. Only clusters with an E_T value above an η-dependent threshold are considered as π^0 candidates[9]. Because most clusters due to pile-up are of low energy, this preselection also rejects a large fraction of pile-up clusters. Thus the `CellBased` algorithm provides a good pile-up robustness.

To optimize and study the identification, sets of signal[10] and background clusters[11] are obtained from simulations. The discrimination is based on the shape of the clusters, the energy of cells in the clusters, its angular properties and its position in η. There is a multitude of variables that provide discrimination between signal and background clusters, some of those are shown in Figure 4.12. The reason why shape variables are used is the following. The neutral pions that are searched for in the π^0 identification have a typical shower shape, resulting from the showers of the two photons from $\pi^0 \rightarrow \gamma\gamma$. In addition to those π^0 showers, there are also clusters that are due to noise, pile-up of imperfect subtraction. Those clusters will have a different shape, because they do not always originate from two photons. Thus variables describing the shape of the cluster (length, width, etc...) will look different for π^0 and the sources mentioned above.

These variables are combined in a boosted decision tree (to exploit their correlations), which provides a single output variable that is then utilised to flag a cluster as π^0 cluster or background cluster. The cut value that is applied in that distribution is optimized such that the counting performance of neutral pions is optimal. Figure 4.13 depicts the BDT output distribution.

The BDT thresholds that are used to determine whether a π^0 candidate passes the the identification are binned in pseudorapitidy and the number of charged pions. The values are listed in Table 4.2.

Table 4.3 lists the counting performance in $\tau_{\text{1-prong}}$ decays. These tables (with entries in percent) are also called efficiency matrices and are an important tool to measure the performance of the decay mode classification. Each column is normalised to 100, so that the diagonal entries indicate the efficiency to correctly count the number of π^0. This efficiency decreases with increasing number of true π^0, because of two reasons. The first reason is that with more π^0 produced in the $\tau_{\text{had-vis}}$ decay, their energy becomes smaller and they are more likely to be rejected by the preselection cut, or they are not found at all because of an oversubtraction.

In case of true 1p1n decays that are also reconstructed as such, the energy and spatial resolution of individual π^0 is shown in Figure 4.14. The RMS of the energy resolution is 32%, the RMS in the ΔR distribution is 0.02, which is roughly the width of one cell in layer 2 of the EM calorimeter.

4.5.2 The `eflowRec`-Algorithm

The `eflowRec` algorithm was developed as a generic energy flow algorithm. During the development of the substructure based τ reconstruction, the `eflowRec` algorithm has undergone several changes. The most important changes, whose documentation is not available publicly, is the addition of the ability to identify a neutral object as being a π^0 and the dedicated *τ-mode*. This section will describe `eflowRec` based on its documentation in [43], to give a basic overview of how it works. Important changes in the *τ-mode* and the π^0 identification are described at the end of this section.

[9] In [42], the E_T threshold is 2.5 GeV and does not depend on η. Futher studies showed that the performance increases when applying an η-dependent cut on E_T.
[10] A cluster is defined as signal, if it has a true π^0 within $\Delta R < 0.1$.
[11] Clusters failing the signal cluster selection are considered background.

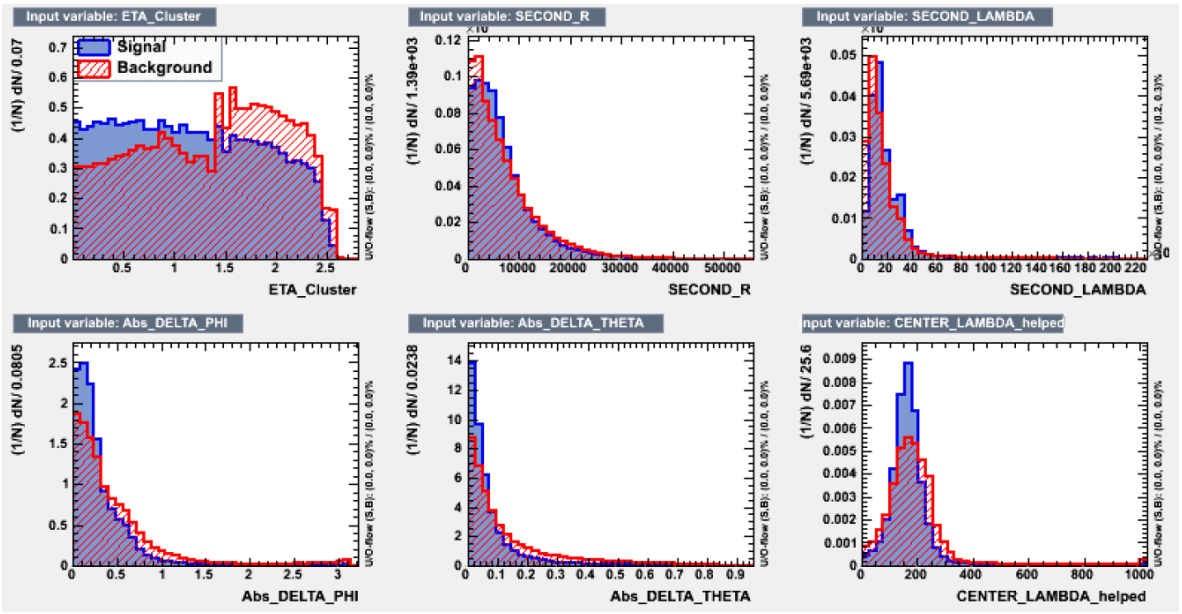


FIGURE 4.12: Top left: *η of the cluster.*
Top center: *Second moment in R. R is the distance to the shower axis.*
Top right: *Second moment in λ, with λ being the distance to the shower center along the shower axis.*
Bottom left: *Absolute value of $\Delta\phi$, which is the difference in ϕ between the shower axis and the shower center when calculated using the reconstructed primary vertex.*
Bottom center: *Absolute value of $\Delta\theta$. This is the analogon to $\Delta\phi$, just using the polar angle.*
Bottom right: *Distance between shower center and calorimeter front face along the shower axis.*
Figure taken from [42].

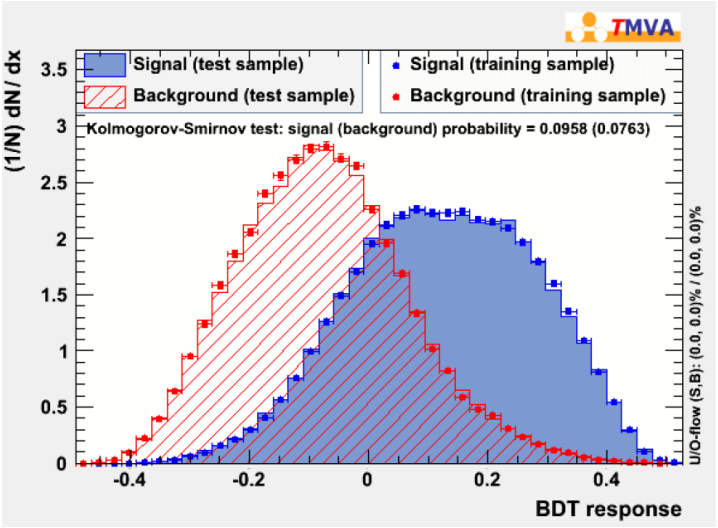


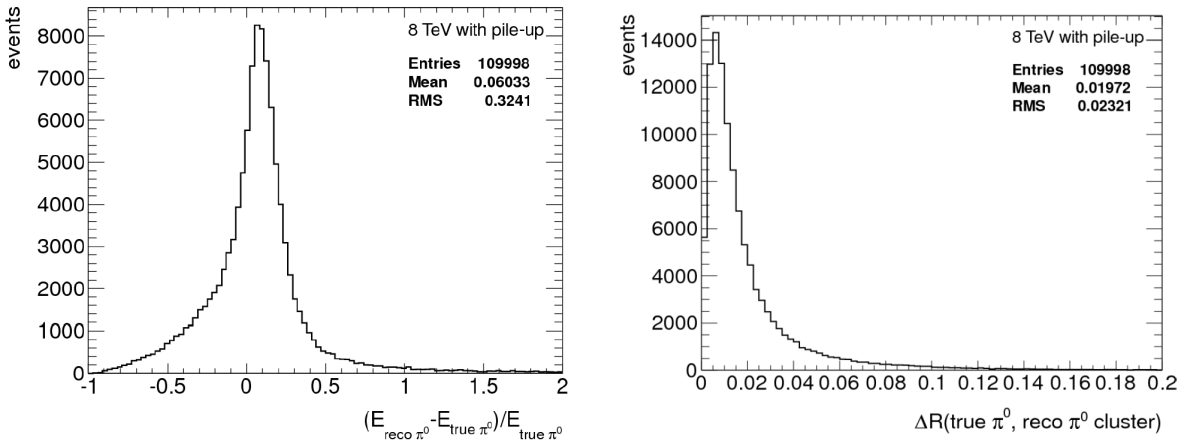
FIGURE 4.13: *BDT output distribution to separate signal clusters originating from π^0 (blue) and background clusters (red, due to pile-up, noise, imperfect subtraction) from each other. Figure taken from [42].*

π^0 identification	BDT threshold				
BDT thresholds	$\lvert\eta\rvert \leq 0.8$	$0.8 < \lvert\eta\rvert \leq 1.4$	$1.4 < \lvert\eta\rvert \leq 1.5$	$1.5 < \lvert\eta\rvert \leq 1.9$	$1.9 < \lvert\eta\rvert$
1-prong taus	-0.06	-0.14	+0.01	-0.10	-0.01
3-prong taus	0.43	0.51	0.48	0.66	0.65

TABLE 4.2: *π^0-BDT thresholds to determine whether a neutral PFO should be classified as a π^0-PFO or a non-π^0-PFO. Neutral PFOs with π^0-BDT values larger than the given values are classified as π^0-PFOs.*

	$N_{\text{true }\pi^0} = 0$	$N_{\text{true }\pi^0} = 1$	$N_{\text{true }\pi^0} = 2$	$N_{\text{true }\pi^0} > 2$
$N_{\text{reco }\pi^0} = 0$	81.3	16.6	6.5	3.3
$N_{\text{reco }\pi^0} = 1$	16.3	68.2	52.3	40.0
$N_{\text{reco }\pi^0} = 2$	2.2	13.8	35.2	43.2
$N_{\text{reco }\pi^0} > 2$	0.3	1.4	6.1	13.5

TABLE 4.3: *Counting performance of the `CellBased` algorithm in $\tau_{1\text{-prong}}$ decays in a $\sqrt{s} = 7$ TeV sample. Entries are given in percent, normalisation is along columns. Statistical errors are $\leq 0.6\%$. Numbers taken from [42].*



(A) *Energy resolution.* (B) *Distance in ΔR between true and reconstructed π^0*

FIGURE 4.14: *Energy and spatial resolution of π^0 in true 1p1n $\tau_{\text{had-vis}}$ decays that are reconstructed as such. According to [42], the slight shift towards higher values is an indication, that the shift is either due to the local hadron calibration, which scales up the energy of low energetic cells, or due to pile-up. Figures taken from [42].*

`eflowRec` Workflow

Input to `eflowRec` are the tracks and uncorrected topological clusters[12] of an event. All clusters are taken into account and no selection is performed. For the tracks, it is required that they are reasonably close to the primary vertex and have at least one hit in the Pixel detector and 6 hits in the SCT. Also, it is assumed that all tracks originate from $\pi^{\pm}$.

Firstly, the tracks are extrapolated to the second layer of the electromagnetic calorimeter. Around the extrapolated position, it is searched for the cluster which minimizes the quantity

$$\Delta R = \sqrt{\frac{(\eta_{\text{Track}} - \eta_{\text{Cluster}})^2}{\sigma_\eta^2} + \frac{(\phi_{\text{Track}} - \phi_{\text{Cluster}})^2}{\sigma_\phi^2}} \tag{4.9}$$

Here, η_{Track} and ϕ_{Track} are the η and ϕ values of the extrapolated track position. σ_η^2 and σ_ϕ^2 are the standard deviations of the cells contributing to the cluster in η and ϕ direction. Hence, the cluster which is closest to the track in units of its η-ϕ size is chosen.

Once a cluster is matched to a track, it is checked whether the condition

$$E_{\text{clus}} > E_{\text{expected}} - k_2 \cdot \sigma_{\text{expected}} \tag{4.10}$$

is fulfilled. E_{clus} is the uncalibrated cluster energy, E_{expected} is the energy that is expected to be found based on charged pions with the energy and η of the matched track, σ_{expected} is the width of E_{expected} and k_2 is a free parameter. E_{expected} is obtained from single pion simulations, where the energy deposit in the calorimeter is studied as a function of $p_{T,\text{Track}}$ and η_{Track}. The distribution of E_{expected} is also referred to as the E/P distribution.

If the condition 4.10 is not fulfilled, the charged pion most likely deposited its energy in several clusters. This is a common effect and to properly treat its occurrence, clusters failing the conidtion 4.10 are flagged and passed to an additional algorithm which is explained later on. If, however, condition 4.10 is fulfilled, the expected energy is removed from the cluster. This is the main difference to the `CellBased` algorithm and the rules by which this removal is done are described later. This subtraction is performed for every track that is matched to the cluster.

Once all tracks for a given cluster have been processed, another condition is checked:

$$E_{\text{clus}} < k_1 \cdot \sigma_{\text{expected}} \tag{4.11}$$

If a cluster passes this condition, then the remaining energy is consistent with noise and the cluster is discarded. Clusters failing this condition have a significant amount of energy left, even after subtraction of all associated tracks. This indicates that another particle (a neutral one) deposited energy in the cluster, hence the cluster is kept.

After processing all tracks and clusters in this way, an additional algorithm is run to deal with clusters that did not pass condition 4.10. Failing that condition indicates a *split shower*, which means that the $\pi^{\pm}$ created several distinct clusters instead of one, hence the additional algorithm is called *split shower recovery*. The split shower recovery collects all clusters in a cone of $\Delta R = 0.2$ around the initial cluster that failed condition 4.10. Then, it goes on to subtract the expected energy from this list of clusters.

Finally, once the split shower recovery is done, the clusters are energy calibrated and the list of output objects is complete.

[12] Topological clusters: See footnote in Section 4.2

A schematic overview over the algorithm is given in Figure 4.15.

Subtraction procedure: Layer of first interaction

The subtraction method in `eflowRec` differs significantly from the method that is applied in the `CellBased` algorithm. In the `CellBased` algorithm, the cells are modified by altering their energy. However, `eflowRec` removes cells as a whole in a specific order (the so-called *cell-ordering*). The order depends on the layer, in which the shower started and is such that cells with higher energies are subtracted first.

To determine this layer, also called *layer of first interaction* (LFI), characteristics of hadronic showers are utilised. These typically start with a prompt electromagnetic shower due to π^0 production which defines the core of the shower. Around this comparibly well defined core, the shower is characterised by diffuse hadronic interactions. Because electromagnetic showers are less susceptible to fluctuations (compared to hadronic showers), the idea is to first subtract this part of the shower. The beginning of the shower, the electromagnetic core, is marked by the largest increase of the energy density along the shower axis. The layer in which this happens is the LFI and the cell-ordering can be derived.

It is assumed that the longitudinal shower axis is the same as the extrapolated track axis. Then, initially, without any assumptions, each cell in the cell-ordering can be described using cylindric coordinates, with the track direction being the z-axis. However, it is assumed that the showers show a radial symmetry[13], thus reducing the dimensions of the ordering to two (position along the shower axis and distance to shower axis). The cell-ordering then becomes an ordering of rings of cells.

The energy density ρ_{ij} of cell j in layer i of a given cluster is calculated via

$$\rho_{ij} = \frac{E_{ij}}{V_{ij}} \cdot \frac{1}{X_{0i}^3} \tag{4.12}$$

The raw energy density is given by the energy in the cell E_{ij} divided by its volume V_{ij}. A correction of $\frac{1}{X_{0i}^3}$ is applied to account for different materials in the different calorimeter layers.

The energy of the cells is weighted by an additional factor, which depends on the distance between the cell and the track axis. Cells that are farther away from the track receive a smaller weight in comparison to cells that are close to the track. If a cluster is large, it may contain showers from other particles. Weighting the contribution of cells based on their distance to the track of interest reduces the influence of those other showers. The distance is measured in units of the Moliere radius [45] and only cells within $3R'_{\text{Moliere}}$ - where R'_{Moliere} is set to be slightly larger than the expected Moliere radius - are considered.

With the distance-weights ω and the cut-off at $3R'_{\text{Moliere}}$, the formula for the energy density in layer i becomes

$$\langle\rho\rangle_i = \frac{\sum_j \omega_{ij}\rho_{ij}}{\sum_j \omega_{ij}} \qquad \left(\forall\, j \mid \Delta R_{ij} < 3R'_{\text{Moliere}}\right) \tag{4.13}$$

Here, j runs over all cells that are closer than $\Delta R = 3R'_{\text{Moliere}}$ to the track.

Finally, the depth of the shower in the calorimeter needs to be expressed. This can be done in units of the nuclear interaction length λ. Along lines of constant η and ϕ, the thickness of the materials of the different layers is divided by the corresponding value for λ to obtain the mean depths for each layer d_i. Thus, combined with the energy density in each layer, the longitudinal density profile can be calculated

[13] C.f. section 4.2 of [43]

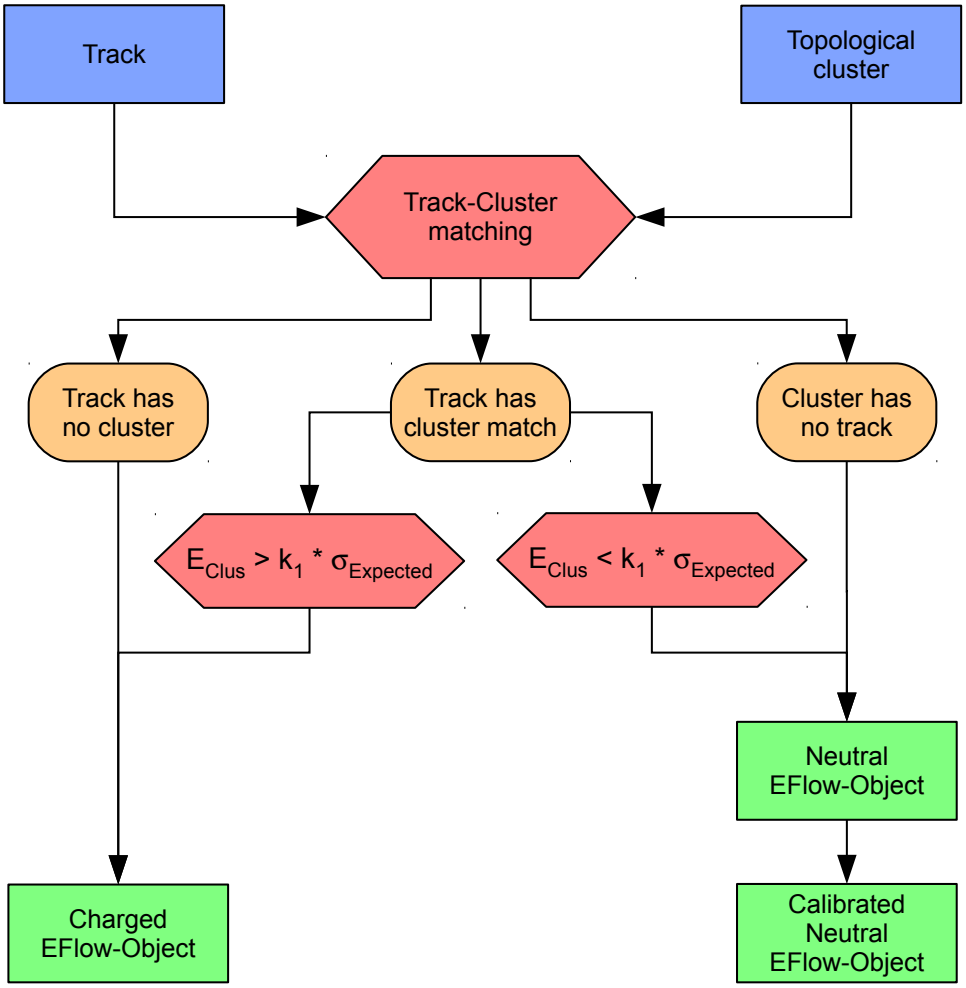


FIGURE 4.15: *Graphical explanation of the* `eflowRec` *workflow. Tracks and clusters of interest are handed to the track-cluster-matching. If no matching cluster (track) is found for a track (cluster), a charged (neutral) energy flow object is created. In case of a match, the track energy is subtracted from the cluster. The remaining cluster energy is tested to be consistent with noise. If the remaining energy is smaller than the noise threshold, a charged energy flow object is created and the cluster is discarded. A neutral energy flow object is created from the cluster if the remaining energy is significantly large. Graphic is based on Figure 16 from [43].*

by evaluating the expression

$$Y_i = \frac{\langle\rho\rangle_i - \langle\rho\rangle_{i-1}}{d_i - d_{i-1}}$$

for each layer i.
The layer of first interaction is then given by the layer with the highest value of Y_i.

Subtraction procedure: Cell ordering

The rings of cells are ordered by their energy density, starting with the most energetic rings. The substraction then follows the list and removes the most energetic rings first, followed by the ring with the next highest energies and so on, until the expected amount of energy is subtracted. When a ring is subtracted, all cells belonging to the ring are *removed*[14] from the cluster, and its kinematics are recalculated. The order of the rings depend on the layer of first interaction.

τ-mode and π^0 identification

The description of the changes in τ-mode with respect to the default setup of `eflowRec` is not documented yet and is based on [46]. The explanation of how the π^0 identification works in `eflowRec` is based on [47].

The changes in the τ-mode are as follows.

- *Hadronic energy constraint:*
  In $\tau_{\text{1-prong}}$ decays, i.e. τ_{had} candidates with one track, it is assumed that all energy in the hadronic calorimeter comes from the $\pi^{\pm}$. Thus, the energy that the $\pi^{\pm}$ deposited in the electromagnetic calorimeter is the difference between the track momentum, corrected with the appropriate E/p factor, and the energy in the hadronic calorimeter. The track energy is therefore replaced with this difference. Consequently, in the subtraction procedure, only the remaining track energy (i.e. initial track momentum times E/p minus the energy found in the hadronic calorimeter) is subtracted in the electromagnetic calorimeter.

- *Energy left overs:*
  After subtracting the track energy from a matched cluster, the condition 4.11 from above

  $$E_{\text{clus}} < k_1 \cdot \sigma_{\text{expected}}$$

  is not checked in τ-mode. Instead, these clusters are passed to the π^0-BDT (see below), which will classify them as π^0-like or background-like.

Aim of the π^0 identification is to provide a score for each neutral energy flow object that tells how likely it is that the neutral energy flow object was created by a π^0. The idea behind this is the same as the idea behind the π^0 identification in `CellBased`.

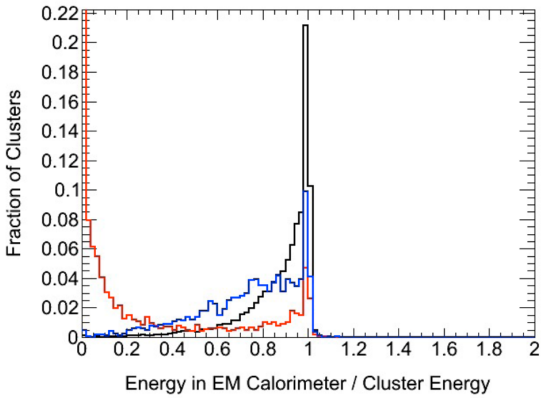
The selection of clusters and the definition of signal and background differs from what is done in `CellBased`. Only clusters with $E_T > 2$ GeV are selected to be classified as signal or background, to surpress clusters from pile-up. Furthermore, the fraction of energy in the cluster from true π^0 or true $\pi^{\pm}$ needs to be larger than 0.2. This cut also reduces contributions from pile-up and the underlying event.

[14] In comparison to `CellBased`, which only changes the energies of the cells, `eflowRec` alters the set of cells that a given cluster is composed of.

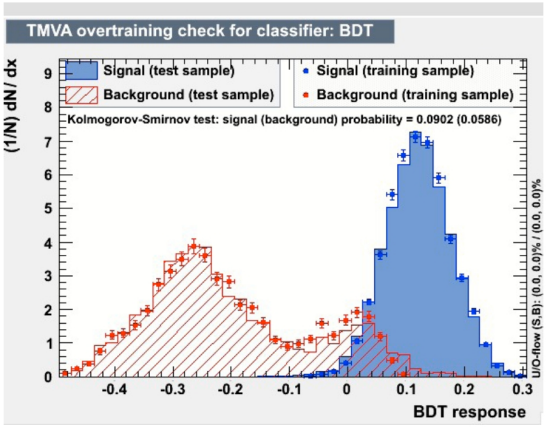
A cluster is defined as a signal cluster, if at least 80% of its energy originates from one or more π^0. If the π^0 contribution is below 10%, the cluster is considered a background cluster. Clusters with values between 10% and 80% are called *mixed* and are not included in the BDT training.

Analogously to the approach in the `CellBased` algorithm, various cluster variables are calculated and compared for signal and background clusters. Variables that show different behaviour for signal and background clusters are combined in a boosted decision tree. An example of a discriminating variable, together with the BDT output distribution is shown in Figure 4.16.

To flag a neutral energy flow object as a π^0-object requires a BDT value above a certain threshold. The threshold depends on the η value, at which the neutral energy flow object is found, as well as the number of tracks in the τ decay.



(A) *Electromagnetic energy fraction: Fraction of cluster energy deposited in Presampler and the three EM-layers, divided by the cluster energy. Signal (black) peaks at values close to one, since the two γ from the π^0 shower in the EM calorimeter. The background (red) is located at small EM-fractions.*



(B) *BDT output distribution with overtraining test. Only clusters classified as signal or background are included in the training and testing.*

FIGURE 4.16: *Input variable for the π^0-BDT (left) and the BDT output distribution (right) for* `eflowRec`.

4.6 Summary

The new approach to τ lepton reconstruction presented in this chapter is based on the energy flow concept. It aims at reconstructing every single particle in τ_{had} candidates that have 1 to 5 tracks. From the individual reconstructed particles, the τ-momentum and the decay mode of the τ_{had} is determined. Reconstruction of the decay products is taken care of by a so-called *substructure algorithm*, which also gives a first estimate of the decay mode and the τ-momentum. The final determination of the decay mode and τ-momentum is based on more information than exploited by the algorithms described in this chapter and is done by the `PanTau` algorithm. `PanTau` has not been discussed in this section, because it is the main accomplishment of this thesis and is discussed in detail in the next chapter.

The `CellBased` algorithm was chosen to be the default substructure algorithm, based on its performance in terms of decay mode classification, neutral pion resolution and resolution of the hadronic tau decay itself. Based on the tracks and cells in clusters associated to a τ_{had} candidate, it performs a

cell-based subtraction of averaged hadronic shower shapes to remove energy depositions of $\pi^{\pm}$ in the EM calorimeter. Remaining energy depositions in the EM calorimeter are passed into a BDT which was trained to identify showers of π^0. By counting the tracks and the number of clusters that have been identified as π^0, a first estimate of the decay mode is derived. By adding the momenta of the tracks and π^0-clusters, the τ-momentum is calculated.

All this information, i.e. the individual τ decay products ($\pi^{\pm}$, π^0-clusters and clusters not identified as π^0), is passed to the `PanTau` algorithm. The next chapter explains how the information is exploited in `PanTau`.

CHAPTER 5

PanTau - Kinematics-Driven Decay Mode Classification

This chapter is dedicated to the PanTau algorithm. Section 5.1 motivates why a second step in tau decay mode identification is necessary and beneficial. In Section 5.2, the classification of particle flow objects in a tau is presented. The basic idea and details of the decay mode reconstruction in PanTau is explained in Section 5.3. Tightly related to the decay mode reconstruction is the selection of those particles in the tau decay that should be considered as tau decay products (in Section 5.4). This chapter concludes in Section 5.5 with a summary and outlook on possible improvements to the algorithm. The actual performance of PanTau will be presented in Chapter 6.

5.1 Motivation

The τ decay products are reconstructed by the substructure algorithm. For the neutral objects, a π^0 identification is applied. From the set of neutral objects identified as π^0 and the charged objects (i.e. the tracks), the decay mode is read-off and the four momentum of the $\tau_{\text{had-vis}}$ is calculated.

In general, this approach gives good results. However, there is additional information which is not used that can help in situations where a decay mode is not reconstructed correctly.

The information of the tau lepton in the detector that has been used so far is the shape information of the clusters to identify them as coming from a π^0 or background (noise, pile-up, non-ideal subtraction). Additional information about the tau decay is stored in the kinematics of the decay products, for instance the angle between the $\pi^\pm$ and the π^0 in 1p1n decays, the invariant mass of the tracks in a 3p0n decay, or the estimated number of photons in the complete tau lepton (estimated from energy depositions in EM1).

It is this sort of information that can further improve the decay mode classification and also the resolution of the $\tau_{\text{had-vis}}$ momentum. This is best illustrated with an example.

Figure 5.1 shows a simplified picture of a 1p1n tau decay. In this specific example, let the energy deposition of the charged pion in the EM calorimeter be subtracted correctly so that there are two clusters left. Both are passed to the π^0 identification of the substructure algorithm and it finds that they both look like clusters a π^0 may produce. However, only one of these clusters is because of the π^0 from the tau decay, the other one comes from a pile-up collision and does not belong to the tau at all. Because both clusters are tagged as π^0, the substructure algorithm classifies the tau as a 1pXn tau, although it is a true 1p1n decay. The substructure algorithm also uses the track and both π^0 clusters to calculate the four momentum, which leads to an overestimation of the energy.

The kinematics of the decay products can remedy this situation, i.e. correct the decay mode classification and correct the energy measurement by adapting it to the correct decay mode.

One such kinematic variable is the fraction of the total τ energy in neutral clusters. This variable is on average higher for true 1pXn decays than for 1p1n decays in the scenario in which these decays are both reconstructed with two neutral objects. Therefore, by combining several such variables, they can *repair* the decay mode to be 1p1n. With the decay mode being 1p1n, it is clear that only one cluster should be considered in the calculation of the tau momentum. This means that also the reconstructed momentum is corrected to be closer to its true value.

The calculation of the kinematic variables, their application in different decay modes and the adaptation of the decay mode is done in the `PanTau` algorithm.

5.2 Classification of Tau Constituents

It is convenient to classify the particle flow objects (PFOs) in a tau. The classification described here is of major importance for the remainder of this chapter.

As pointed out earlier (c.f. Section 2.2 and Figure 4.1), the tau decay is rather confined spatially. Therefore, after its calculation, the intermediate tau axis (c.f. section 4.2) is used as the centre of a circle of radius $\Delta R = 0.2$ in $\eta - \phi$ space. This circle defines the *core region* of the tau decay. It is assumed that all decay products lie within this circle and Figure 5.2 shows that this is a reasonable assumption[1].

For completeness it is mentioned that an isolation annulus is defined, with ΔR values of $0.2 < \Delta R < 0.4$, i.e. around the core region. PFOs reconstructed here are not taken into account when calculating the decay mode or the $\tau_{\text{had-vis}}$ momentum, because it is very unlikely that tau decay products have such

[1] The figure also shows that the cone is too large for taus with increasing momentum, c.f. Section 5.5.2.

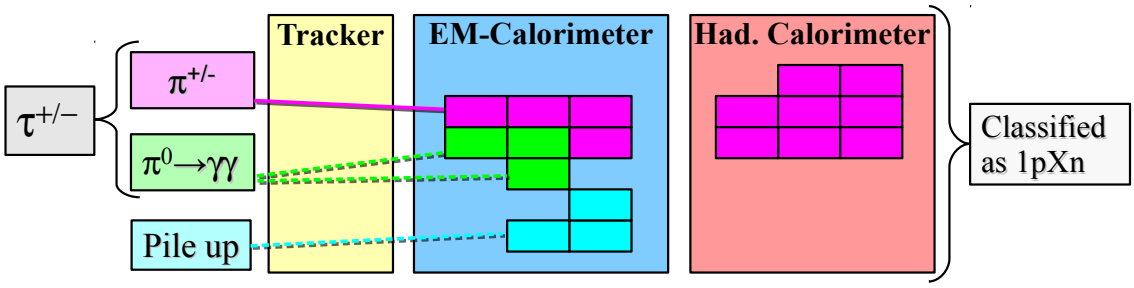
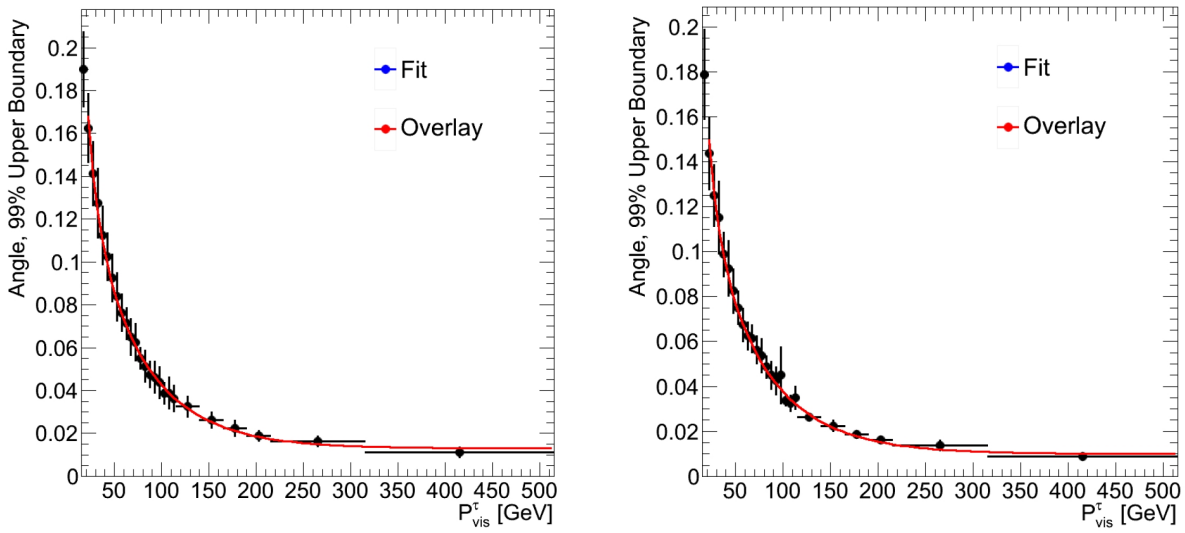


FIGURE 5.1: *Example 1p1n τ decay misreconstructed as 1pXn. The charged pion (magenta) in the tau decay, leaves a track in the tracker, a cluster in the EM calorimeter and a cluster in the hadronic calorimeter. The signature of the neutral pion (light green) is a cluster in the EM calorimeter, which is the same as the signature of the pile-up particle (light blue). Both neutral particles, the neutral pion and the pile-up particle have been identified as π^0 by the π^0 identification of the substructure algorithm, leading to a wrong decay mode and wrong energy measurement.*
Key to correctly identifying the 1p1n decay is the exploit of the kinematics of the decay products. In this specific case, the ratio of energy in neutral clusters and energy of the tau lepton can help to disentangle the 1pXn looking decay from a 1p1n decay.
For illustration purposes, the sizes of the tracker, the calorimeter and the cells are not to scale.



(A) *1-prong tau decays.*

(B) *3-prong tau decays.*

FIGURE 5.2: *Solid angle with respect to the tau axis that contains all neutral pions in 99% of the cases, for the central region of the detector $0 < |\eta| < 0.25$. Both figures taken from [48].*

PFO type	p_T cut in GeV	E_T cut in GeV				
		$\lvert\eta\rvert \leq 0.8$	$0.8 < \lvert\eta\rvert \leq 1.4$	$1.4 < \lvert\eta\rvert \leq 1.5$	$1.5 < \lvert\eta\rvert \leq 1.9$	$1.9 < \lvert\eta\rvert$
Charged PFO	1.0					
Neutral PFO		1.5	1.7	1.8	1.5	1.3
Outer charged PFO	1.0					
Outer neutral PFO		1.0	1.0	1.0	1.0	1.0

TABLE 5.1: *Values of the p_T and E_T cuts that are used in the PFO classification. The E_T cuts have been tuned in different η bins to account for the η dependent behaviour of the backgrounds to neutral pion identification.*

a large distance to the tau axis. However, they are still kept as part of the tau, because they can be used to separate real tau leptons from fake tau leptons due to quark or gluon initiated jets.

In order to suppress particles from pile-up collisions, which are typically rather soft, the charged and neutral PFOs in the two regions are required to pass p_T and E_T cuts. If a PFO does not pass the required p_T or E_T cut, it is discarded. The cut values are actually the same as in the `CellBased` substructure algorithm and are listed in Table 5.1.

The following list gives an overview of the definition of the different PFO types, a graphical representation of the classification is given in Figure 5.3. To get a feeling for how a $\tau_{\text{had-vis}}$ can be thought of with this classification, Figure 5.4 illustrates the two regions and the different PFO objects.

- *Charged PFOs: $PFO^{\pm}$*
  A PFO is classified as a charged PFO (short: $PFO^{\pm}$), if it has a non vanishing charge and it is closer than $\Delta R = 0.2$ to the intermediate tau axis. In addition, the track that it was created from needs to have a transverse momentum of $p_T > 1$ GeV. Because all charged PFOs are assumed to originate from charged pions, they will also be called $\pi^{\pm}$-PFOs. In the substructure algorithm, charged PFOs are used to determine whether a tau is 1-prong or 3-prong.

- *Neutral PFOs:*
  If a PFO has no charge and is closer than $\Delta R = 0.2$ to the intermediate tau axis, it is called a neutral PFO. An η-dependent cut on E_T is also applied to reject neutral particles from pile-up collisions. The values are given in Table 5.1.

- *π^0-PFOs:*
  π^0-PFOs are neutral PFOs that passed the π^0 identification of the substructure algorithm. Hence, the set of π^0-PFOs in a tau decay is a subset of the set of neutral PFOs, i.e. every π^0-PFO is a neutral PFO, but not every neutral PFO is a π^0-PFO. To determine whether a neutral PFO passes the π^0 identification, its π^0-BDT score is compared against an η dependent threshold. If its BDT score exceeds the threshold, the neutral PFO is also classified as a π^0-PFO. If its BDT score does not exceed the threshold, the neutral PFO is tagged as non-π^0-PFO (see next item). The values of the thresholds also depend on whether the tau has one or three tracks. Table 4.2 lists the cut values.

  In the substructure algorithms, π^0-PFOs are counted to retrieve the number of neutral pions in the tau decay.

- *non-π^0-PFOs:*
  A neutral PFO that does not pass the π^0 identification of the substructure algorithm is referred to as non-π^0-PFO.

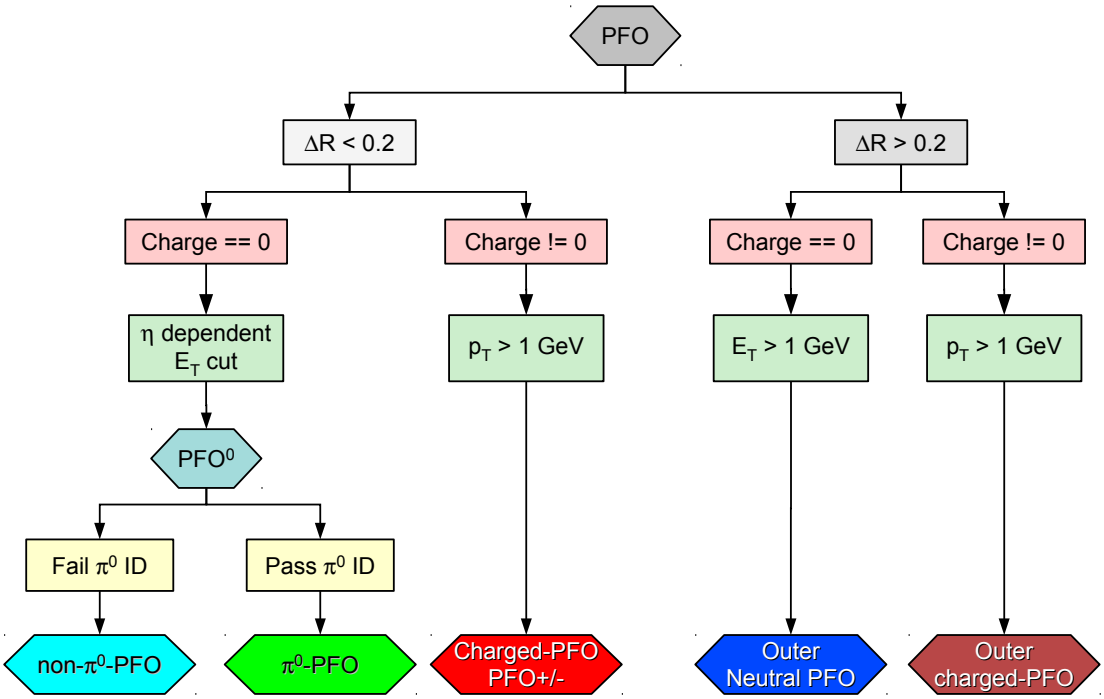


FIGURE 5.3: *Classification of particle flow objects based on their distance in ΔR towards the intermediate tau axis, their charge, their momentum or energy and π^0 identification of the substructure algorithm.*

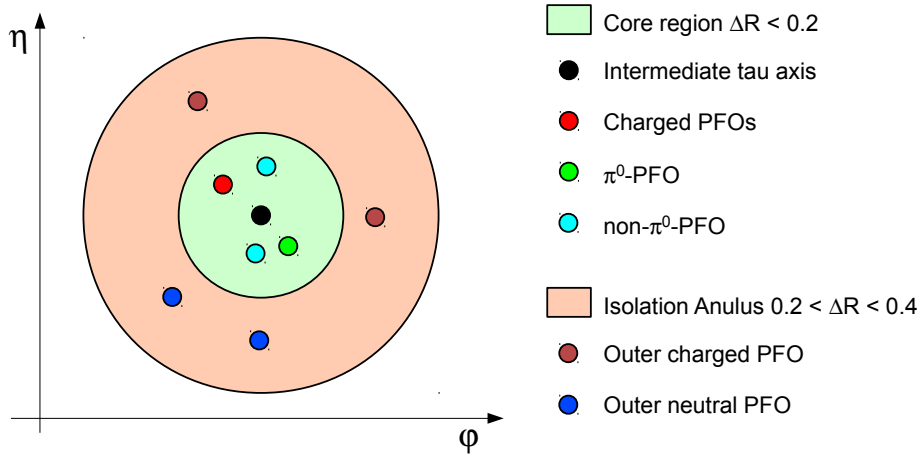


FIGURE 5.4: *Classification of PFOs in a tau decay. The $\tau_{had\text{-}vis}$ decay is divided into the core region, a circle with radius 0.2 in $\eta - \phi$ space, and the isolation annulus, which spans from $\Delta R = 0.2$ to $\Delta R = 0.4$ in $\eta - \phi$ space. Because tau decays are narrow, only PFOs in the core region are considered in the decay mode reconstruction.*

These PFOs are not used in the decay mode determination in the substructure algorithm. However, they play a special role in PanTau, which is explained in Section 5.3.2.

The set of non-π^0-PFOs and the set of π^0-PFOs are disjoint. Their union is the set of neutral PFOs.

- *Outer charged PFOs* and *outer neutral PFOs:*
  If a PFO is not inside the $\Delta R = 0.2$ circle but in the isolation annulus, it falls into one of the *outer* categories. Here, it is only separated between PFOs with and without charge, and they are referred to as outer charged and outer neutral PFOs. The outer neutral PFOs are not subdivided into outer-π^0-PFOs and outer-non-π^0-PFOs. The transverse momentum (energy) of outer charged (neutral) PFOs needs to be larger than 1 GeV.

  Neither outer charged PFOs nor outer neutral PFOs are considered when reconstructing the decay mode or when calculating the tau momentum. They are only mentioned for completeness and play no further role in the decay mode reconstruction or the tau momentum calculation.

5.3 Decay Mode Reconstruction

In the introduction subsection 5.3.1, this section describes the idea behind the decay mode reconstruction in PanTau. In subsection 5.3.2, it goes on to describe the *decay mode tests*, in which the actual decay mode reclassification is performed. Subsections 5.3.3 and 5.3.4 document the method of selecting kinematic variables that form the basis on which the reclassification is evaluated. The Subsections 5.3.5, 5.3.6 and 5.3.7 describe the decay mode tests in detail. The last Subsection (5.3.8) discusses the results and improvements that are achieved by the reclassification.

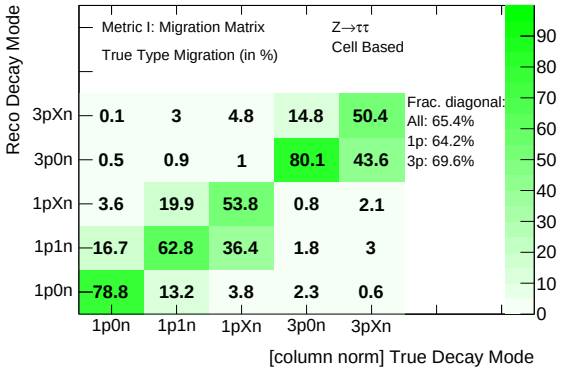
5.3.1 Introduction

The starting point of the decay mode reconstruction in PanTau is the decay mode classification performed by the substructure algorithm. The decay mode classification is quantified in *migration matrices* and they are shown in Figure 5.5. The migration matrices are a core performance metric in the substructure based tau reconstruction. They show the true tau decay mode on the x-axis, and the reconstructed tau decay mode on the y-axis.

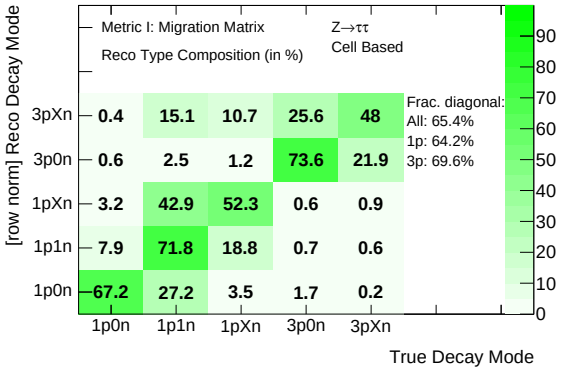
If the normalisation is such, that each column adds to 100%, it is called the *efficiency matrix*. This is indicated by the text *[column norm]* in front of the x-axis label and by the text *True Type Migration* above the actual matrix. With this normalisation, the entry in column c and row r tells the efficiency with which the true mode belonging to column c is reconstructed as the mode belonging to row r. For example, the efficiency of a true 1p1n to be reconstructed as a 1p0n is 13.2% based on Figure 5.5a. The efficiency is an important measure, because it says what fraction of taus with a given decay mode are actually reconstructed.

Another way to perform the normalisation is to make the entries of each row add up to 100%, in which case the matrix is called the *purity matrix*. These matrices are marked by the prefix *[row norm]* in front of the y-axis label and by the label *Reco Type Composition* at the top of the plot. Here, the entry in column c and row r gives the admixture of the true mode belonging to column c in the reconstructed mode belonging to row r. In Figure 5.5b, this means that 27.2% of reconstructed 1p0n taus are true 1p1n taus. Like the efficiency, also the purity is an important measure, because it reflects what fraction of the reconstructed taus of a given mode are actually truly of that mode.

The diagonal entries have a special role, because they indicate the fraction of correctly migrated tau leptons. This fraction is also given at the right side of the matrix, below "Frac. diagonal". "All" gives



(A) Efficiency Matrix *of the* `CellBased` *algorithm.*



(B) Purity Matrix *of the* `CellBased` *algorithm.*

FIGURE 5.5: Migration Matrices *of the* `CellBased` *algorithm. Explanations of how the matrices are read are in the text. The diagonal entries are the dominant entries, which indicates a good classification performance. However, significant contributions from next-to-diagonal elements lead to misclassifications and flawed energy measurements. Other entries are negligibly small at first order.*

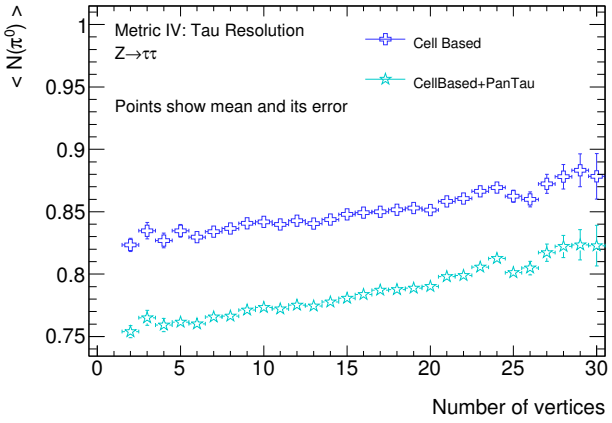


FIGURE 5.6: *Mean number of π^0 found in a tau candidate as a function of primary vertices in the event. The mean increases by 0.07 over 30 vertices.*

the percentage of all taus that are classified correctly, "1p" and "3p" state the same number for the 3x5 1-prong sub-matrix and 2x5 3-prong sub-matrix respectively.

In Figure 5.5a, the diagonal entries are already quite dominant. However, the entries directly next to the diagonal are not negligible. They arise because of these main reasons for misclassifications, which can increase or decrease the number of reconstructed neutral pions by one:

1. *Errors in the subtraction procedure:*
   The subtraction does not always work properly because of large fluctuations in the charged pion shower. As shown in Figure 4.8, the estimated energy to be subtracted in the EM calorimeter is not always exactly the energy that was really deposited by the charged pion. The precision increases with the $\pi^{\pm}$ energy, but the RMS remains above $\approx$ 10%. Thus it can happen that not enough energy is subtracted in the electromagnetic calorimeter, so that the cluster produced by the charged pion is not removed. If this cluster then passes the π^0 identification by accident (c.f. Figure 4.13, separation between signal and background is not perfect), the reconstructed tau has

more π^0 than the true tau decay.

The fluctuations can also be such, that the estimated energy to be subtracted is much larger than the energy that was actually deposited in the EM calorimeter. This has severe consequences in the case in which the charged pion cluster overlaps with the cluster from a neutral pion[2], because over-subtracting here leads to an underestimation of the π^0 cluster energy. Thus, it can happen that the energy falls below the preselection cut and the neutral cluster is not considered at all in the π^0 identification. $\tau_{\text{had-vis}}$ candidates in which this happens will have less neutral pions than the true decay.

2. *Small energies:*
   A rather subtle reason for misclassification is that - especially in taus with small transverse momenta - a neutral pion may simply not have enough energy to be considered as belonging to the tau lepton. It fails the preselection cut for neutral pions and is not reconstructed. Thus, the number of reconstructed π^0 is too small. Figure 7.9 shows that this indeed can be the case, especially if the charged pion carries more energy than the neutral pion.

3. *Pile-up particles:*
   Other proton proton collisions taking place in the same bunch crossing as the hard collision in which the tau is produced can create neutral particles that overlap with the tau. A fraction of these particles passes the π^0 identification, increasing the number of reconstructed π^0, so that their number is overestimated (Figure 5.6).

4. *Cluster merging:*
   Intrinsically, the tau decay happens in a narrow cone, i.e. the distance between the decay products in the calorimeter is very small. With increasing transverse momentum of the tau lepton, this effect becomes more pronounced. Therefore, in tau decays with two (or more in rare cases) neutral pions, the neutral pions can merge into one cluster[3]. The number of reconstructed π^0 is underestimated in this case.

The idea behind the decay mode calculation in `PanTau` is to build on the already quite good classification performance of the substructure algorithm. The migration matrix in Figure 5.5a tells which decay modes are likely to be confused with one another. The three pairs of decay modes that benefit most from additional separation power are:

- *1p0n ↔ 1p1n migration*:
  16.7% of true 1p0n taus are reconstructed as 1p1n and 13.2% of true 1p1n are reconstructed as 1p0n. The latter has a very significant impact on the 1p0n purity, because the 1p1n branching ratio is roughly 2.5 times higher than the 1p0n branching ratio. This leads to the 27.2% contamination of 1p1n taus in the reconstructed 1p0n mode (see Figure 5.5b). Pile-up (1p0n → 1p1n) and imperfect subtraction or small π^0 energies (1p1n → 1p0n) are the reasons for this misclassification.

[2] As mentioned before, the decay products of the tau are boosted. Figure 5.2 showed the angle that contains all neutral pions in 99% of the cases. Figure 5.12c shows the ΔR between the charged and the neutral PFO in 1p1n decays. Combined with the cell width of $\Delta\eta \times \Delta\phi \approx 0.025 \times 0.025$ in the EM2 layer, it is reasonable to assume that an overlap between the two clusters may happen.

[3] The angle containing all π^0 in 99% of the cases is shown in Figure 5.2. Above a tau energy of 50 GeV, this angle spans over roughly 3x3 cells in the EM2 layer, so that there is a non vanishing probability for the two π^0 to be found in the same cluster.

- *1p1n ↔ 1pXn migration*:
  Separation of these modes is incredibly challenging, as the numbers from Figure 5.5a indicate. 19.9% of true 1p1n taus are reconstructed as 1pXn, and 36.4% of true 1pXn taus are wrongly classified as 1p1n. A finer transverse segmentation in the calorimeter would help to disentangle the modes, because with increasing number of neutral pions in the decay, the more likely it is for them to merge into one cluster.

- *3p0n ↔ 3pXn migration*:
  In the three prong category, the mixing between modes with and without neutral pions is as follows. 14.8% of true 3p0n taus are reconstructed as 3pXn, while 43.6% of true 3pXn are reconstructed as 3p0n. In the three prong case, three charged pions need to be subtracted from the EM calorimeter. For a tau with given p_T, this means that the subtraction works worse than in a one prong decay, because the energy needs to be distributed among three $\pi^{\pm}$, making them lower energetic and the subtraction less accurate (compare to Figure 4.8). In addition neutral pions may fail the preselection cut on the energy, because in a 3pXn decay, the tau energy is distributed among at least five particles (the ν_τ, three $\pi^{\pm}$, at least one π^0). Further more, the particle density in the tau decay is higher than for 1-prong decays, so that the probability of the neutral pion to overlap with one of the charged pions is higher.

  However, while these are the physically relevant reasons, there is a technical reason [49] that the 3pXn→3p0n migration is around 43%. The reason is that in this version of the `CellBased` algorithm, no subtraction was performed in 3-prong taus. This leads to a much larger background coming from imperfect subtraction (because there was no subtraction) and thus the necessity to apply a harder cut in the π^0 BDT score, reducing the efficiency to reconstruct π^0.

The misclassification can be tracked down to real π^0 accidentally failing (fake π^0 accidentally passing) the π^0 identification, decreasing (increasing) the number of reconstruction π^0 by one. The goal for `PanTau` is to minimise migration in these decay mode pairs.

5.3.2 Decay Mode Tests

A *decay mode test* is a test that is performed in `PanTau` to check which of two possible decay modes is more likely. This means that in each test, the likeliness of two possible decay modes is compared and the one with the higher likeliness is chosen as the reconstructed decay mode. Three tests are implemented in `PanTau`, namely tests for 1p0n-vs-1p1n, 1p1n-vs-1pXn and the 3p0n-vs-3pXn. Migration in these cases has been discussed in the previous section. These modes are chosen, because they offer the largest possibilities for improvements.

The decision, which of the two possible true modes is chosen in a decay mode test is made by evaluating a boosted decision tree (BDT). The reasons for choosing a BDT over a cut-based approach are:

- The BDT takes correlations between variables into account. While this can also be done in a cut-based approach, the number of variable pairs to take into account increases quadratically with the number of variables to base the decay mode test on[4].

[4] For n variables, there are $n \cdot (n-1) \propto n^2$ combinations.

- Using a BDT allows to tune the result continuously by varying the cut on the BDT output score. While this is in principle also possible in a cut-based approach, it is much more involved to set up technically.
- The BDT output is quite robust with respect to over-training (to be discussed later), as it is merely a large set of rectangular cuts.

Thus, in order to execute the decay mode tests, three BDTs (one for each test) need to be set up, i.e. variables need to be selected, the BDTs need to be trained and their performances need to be evaluated. Training and evaluation is done using the TMVA package [40].

However, before the variable selection is discussed, it is important to correctly classify the reconstructed taus into the different decay mode tests and BDTs. Therefore, it is important to introduce a convenient nomenclature to refer to the different modes, that are considered in the decay mode tests. The *complete tau composition* as reconstructed by the substructure algorithm determines the test that is to be performed. Here, *complete tau composition* means the set of all charged PFOs and all neutral PFOs, both π^0-PFOs and non-π^0-PFOs. It is convenient to introduce a shorthand notation for the complete composition of a reconstructed tau, to easily denote the number of charged PFOs, π^0-PFOs and non-π^0-PFOs in a τ_{had}. The notation is such that it lists the multiplicities of the different PFO types as a three digit number, prefixed by an R to denote that it is a reconstructed tau. For instance, *Rijk* refers to reconstructed taus with i charged PFOs, j π^0-PFOs and additional k non-π^0-PFOs - i.e. the total number of neutral PFOs is $j + k$. Table 5.2 gives an overview over the relation between the shorthand notation for the important decay modes of the tau from Tables 2.4 and 4.1 and the shorthand notation for the complete tau composition.

It is noteworthy that Table 5.2 only lists the complete tau compositions for ideal situations, in which all π^0 from the τ_{had} decay are identified as π^0-PFOs, i.e. only the real daughters of the tau lepton are reconstructed and identified correctly. In a real tau decay however, the algorithms may reconstruct, for instance, one charged pion, no π^0-PFOs and one or more non-π^0-PFOs (i.e. a R10X tau). Such a tau is classified as 1p0n by the substructure algorithm, because there the decay mode is obtained by counting charged and π^0-PFOs. The true decay of this example tau could be 1p0n (then the neutral PFO comes from pile-up or noise), or 1p1n (then the neutral pion accidentally failed the π^0 identification, or was not reconstructed at all).

If the non-π^0-PFO was counted as π^0-PFO - i.e. the decay mode was changed to be 1p1n - in those cases in which the true decay was 1p1n, the misclassification would be resolved. Based on kinematic variables of the charged PFO and the non-π^0-PFO, it is indeed possible to identify those taus, in which the non-π^0-PFO more likely was a π^0. Generalised to all possible complete tau compositions, this captures the main idea behind the three decay mode tests in `PanTau`.

Table 5.3 lists all complete tau compositions that are taken into account in `PanTau`, together with the two most likely true modes they belong to.

A detailed discussion of the complete tau compositions listed in Table 5.3 follows.

- *R100 taus:*
  These are taus that only consist of a charged pion, because no neutral component was reconstructed at all. This is perfectly fine for isolated 1p0n decays. This composition is also retrieved for 1p1n decays, in which the neutral pion was too soft to pass the preselection cut.

  It is not possible at all to recover true 1p1n taus that are reconstructed as R100, because there is no neutral object to take the place of the π^0 from the 1p1n decay.

True Decay Mode	Shorthand notations	
	Decay Mode	Complete tau composition
$\tau^- \rightarrow \nu_\tau \pi^-$	1p0n	R100
$\tau^- \rightarrow \nu_\tau \pi^- \pi^0$	1p1n	R110
$\tau^- \rightarrow \nu_\tau \pi^- > 1\pi^0$	1pXn	R1X0
$\tau^- \rightarrow \nu_\tau \pi^- \pi^-\pi^+$	3p0n	R300
$\tau^- \rightarrow \nu_\tau \pi^- \pi^-\pi^+ > 0\pi^0$	3pXn	R3X0

TABLE 5.2: *Relation between decay modes, their shorthand notation and the shorthand notation for the complete tau composition in an ideal scenario. The "R" in the complete tau composition shorthand notation denotes that it is a reconstructed tau. The first digit marks the number of charged pions reconstructed. The second digit refers to the multiplicity of π^0-PFOs. Similarly to the decay mode shorthand notation, an "X" is used to inclusively refer to all multiplicities not categorised before in the respective prong category. The third digit states the number of non-π^0-PFOs. This number is zero in this table, because in an ideal scenario all π^0 from the true τ_{had} decay are identified as such and there are no other particles in the core region, i.e. only the real daughters of the tau lepton are reconstructed and identified correctly.*

Complete tau composition	Number of $\pi^\pm$-PFOs	Number of π^0-PFOs	Number of non-π^0-PFOs	Most likely true modes	`CellBased` classification
R100	1	0	0	1p0n, 1p1n	1p0n
R10X	1	0	≥ 1	1p0n, 1p1n	1p0n
R110	1	1	0	1p0n, 1p1n	1p1n
R11X	1	1	≥ 1	1p1n, 1pXn	1p1n
R1XX	1	≥ 2	≥ 0	1p1n, 1pXn	1pXn
R300	3	0	0	3p0n, 3pXn	3p0n
R30X	3	0	≥ 1	3p0n, 3pXn	3p0n
R3XX	3	≥ 1	≥ 0	3p0n, 3pXn	3pXn

TABLE 5.3: *All complete tau compositions that are considered in `PanTau`, alongside their most probable origin and the classification by `CellBased`. Although the R100 and R300 compositions do not contain any reconstructed π^0, their most probably true mode (besides 1p0n and 3p0n respectively) are modes that have the same number of charged particles and one neutral pion (c.f. argumentation in 5.3.1). The `CellBased` classification is based on the number of $\pi^\pm$- and π^0-PFOs, as described in Section 4.6.*

Because there are no options regarding the decay mode, other than sticking to 1p0n, these taus do not enter a decay mode test in `PanTau` and they also do not have their momentum recalculated.

- *R10X taus:*
  Taus with one $\pi^\pm$ and one or more non-π^0-PFOs are classified as R10X taus. The non-π^0-PFOs are due to either ...
  - ... noise, pile-up or imperfect subtraction, if the decay was 1p0n,
  - or a not identified π^0, if the decay was 1p1n.

  Taus of the R10X type enter the 1p0n-vs-1p1n decay mode test because of this.

  Of course, a smaller fraction of R10X taus is coming from 1pXn taus and an even smaller fraction is coming from 3p0n and 3pXn (track merging, inefficiencies). However, they are not as dominant as 1p0n and 1p1n and thus not taken into account (c.f. Figure 5.25).

- *R110 taus:*
  In contrast to what the reconstructed composition suggests, namely a 1p1n tau, it is also possible for a true 1p0n tau to fall into this category. This can happen if a cluster not produced by the tau ends up being classified as π^0-PFO. Thus, R110 taus also enter the 1p0n-vs-1p1n decay mode test.

  Contributions from other true modes are neglected (c.f. Figure 5.25).

- *R11X taus:*
  If there is at least one non-π^0-PFO in an R110 tau, it is classified as R11X. True modes contributing the most to R11X are:

  – 1p1n taus, with additional neutral PFOs from pile-up.

  – 1pXn taus, with one π^0-PFO failing the identification.

  Therefore, R11X taus are put into the 1p1n-vs-1pXn test.

  Contributions from 1p0n are unlikely, because there need to be two additional neutral PFOs, one of which needs to be tagged as π^0-PFO, c.f. Figure 5.25. Contamination from $\tau_{\text{3-prong}}$ is possible if tracks merge or are not reconstructed, but is suppressed because of the small 3pXn branching ratio (also in Figure 5.25).

- *R1XX taus:*
  Taus with at least two π^0-PFOs are not divided any further, so that R1XX refers to all tau decays with one $\pi^\pm$, two or more π^0-PFOs and any number of additional non-π^0-PFOs. Thus, analogously to the discussion of the other complete tau compositions, the two most likely true decays to produce an R1XX tau are:

  – 1p1n taus, in which a neutral PFO accidentally passed the π^0 identification.

  – 1pXn taus, either correctly reconstructed or with additional non-π^0-PFOs from pile-up.

  Hence, R1XX taus also end up in the 1p1n-vs-1pXn test.

  Because of the rather large amount of neutral particles that is needed to be classified as R1XX, contributions from other true decay modes can be neglected, this is shown in Figure 5.25, too.

- *R300 taus:*
  Treatment of this case is completely in line with what is done in the R100 case. True 3pXn modes will contribute to R300 taus, but because no neutral particle is reconstructed at all, nothing can be done to correctly migrate R300 to the 3pXn mode.

  Therefore, R300 taus do not enter any of `PanTau`'s decay mode tests, and they also do not have their momentum recalculated.

- *R30X taus:*
  The R30X category is the 3-prong version of the R10X category. The non-π^0-PFOs are due to noise, pile-up or imperfect subtraction (more likely than in R10X, because of lower decay product energies) if the true mode is 3p0n. If the true mode is 3pXn, the neutral pion failed the π^0 identification.

  R30X taus are put into the 3p0n-vs-3pXn decay mode test.

  Because of the fact that the `CellBased` algorithm did not perform the subtraction at the time the data samples were produced, the background due to imperfect subtraction is extraordinarily large.

Hence, neutral pions are likely to be classified as non-π^0-PFOs because of the harder cut in the π^0 identification.

- *R3XX taus:*
  Finally, a tau reconstructed with three charged PFOs, at least one π^0-PFO and any number of non-π^0-PFOs is put into the R3XX category. This category of reconstructed taus mainly consists of:
  - 3p0n taus, with a neutral PFO accidentally passing the π^0 identification.
  - 3pXn taus, either reconstructed correctly or with additional non-π^0-PFOs from pile-up.

  Taus classified as R3XX are considered in the 3p0n-vs-3pXn decay mode test.

  Because of the rather large conversion probability of photons into a e^+e^- pair in the ATLAS detector, there is a contribution from $\tau_{\text{1-prong}}$, as shown in Figure 5.25. However, this fraction is smaller than the contribution from 3p0n to 3pXn and so it is neglected in the current version of `PanTau`.

A special note concerning R110 and R11X taus:
If there are three or more reconstructed photons in a tau with one of these compositions, it is not passed to the decay mode tests, but directly reclassified as 1pXn, because the (at least) three photons are assumed to come from the (at least) two $\pi^0 \to \gamma\gamma$ decays in the 1pXn tau.

Table 5.4 summarises the discussion of the different complete tau compositions. It shows, that each complete tau composition enters exactly one decay mode test (if the composition can be recovered). It also shows, that in each decay mode test, two categories of taus are present, for instance R10X and R110 in 1p0n-vs-1p1n. Those complete tau compositions that enter the same decay mode test also have the same most likely true modes.

Complete tau composition	Most likely true modes	Can be recovered	Decay Mode test
R100	1p0n, 1p1n		
R10X	1p0n, 1p1n	✓	1p0n-vs-1p1n
R110	1p0n, 1p1n	✓	1p0n-vs-1p1n
R11X	1p1n, 1pXn	✓	1p1n-vs-1pXn
R1XX	1p1n, 1pXn	✓	1p1n-vs-1pXn
R300	3p0n, 3pXn		
R30X	3p0n, 3pXn	✓	3p0n-vs-3pXn
R3XX	3p0n, 3pXn	✓	3p0n-vs-3pXn

TABLE 5.4: *All complete tau compositions that are considered in `PanTau`, their most probable origin and the decay mode test they are passed to. There are two complete tau compositions in each decay mode test, and they share the same most likely true modes.*

In each decay mode test, one of the two most likely true decay modes is the correct one, while the other one is not[5]. The correct true decay mode in a decay mode test is called signal mode, the other one is called background mode. Which mode is signal and background depends on the complete tau composition that is reconstructed, see Table 5.5. In principle it is possible to now divide each decay mode test into two parts, based on the complete tau composition (for example an R10X-1p0n-vs-1p1n and an R110-1p0n-vs-1p1n version of the 1p0n-vs-1p1n test). However, this would require to split

[5] In an ideal world, this is true. As pointed out in the description of the complete tau compositions, only the two most likely true modes are considered. It can be the case that a true 3p0n is reconstructed as a R10X due to track merging. Then of course neither 1p0n, nor 1p1n is the correct true mode.

Complete tau composition	Decay Mode Test	Signal	Background
R10X	1p0n-vs-1p1n	1p0n	1p1n
R110	1p0n-vs-1p1n	1p1n	1p0n
R11X	1p1n-vs-1pXn	1p1n	1pXn
R1XX	1p1n-vs-1pXn	1pXn	1p1n
R30X	3p0n-vs-3pXn	3p0n	3pXn
R3XX	3p0n-vs-3pXn	3pXn	3p0n

TABLE 5.5: *Connection between the complete tau compositions, the decay mode test they enter, and the resulting signal and background modes.*

the set of all taus into 12 parts (two sets for each complete tau composition) and thus further limit the statistics in the training process of the boosted decision tree (which is explained later on). It would also lead to the introduction of six BDTs in total. Even without the subdivision, the statistics is rather limited, because of the splitting of the complete set of taus into one subset for each decay mode[6]. Thus, all complete tau compositions that share the same most likely true modes enter the same test (i.e. two per test).

Figure 5.7 gives a graphical representation of the process of assigning the different complete tau compositions into the three decay mode tests.

This example illustrates how the reclassification in `PanTau` is done.

1. `CellBased` reconstructs a tau with one $\pi^{\pm}$-PFO, no π^0-PFO and one additional non-π^0-PFO, i.e. with a complete tau composition R*ijk*. Therefore, the tau is classified as a 1p0n (c.f. Section 4.6). This tau is passed to `PanTau`.

2. `PanTau` looks at the complete tau composition, which is R10X in this example (Table 5.3).

3. Based on Table 5.4, the tau is passed to the corresponding decay mode test, in this case the 1p0n-vs-1p1n decay mode test.

4. The complete tau composition and the decay mode test determine whether the tau is part of the signal or background in the BDT (Table 5.5).

5. Based on the BDT-score, the classification is (not) changed, so that the decay mode is altered.

5.3.3 Variable Ranking

In an automated fashion, `PanTau` calculates a vast number of kinematic and other variables for each tau. These variables will also be referenced as *global tau variables*, *features* or *tau features*. Variables that are calculated are based on single PFOs or types of PFOs (i.e. $\pi^{\pm}$-PFOs, π^0-PFOs and neutral PFOs) and include multiplicities, energy ratios, angles, averaged angles, standard deviations, invariant masses, π^0-BDT values and generic jet shape variables. In order to identify variables that show separation between signal and background in a given decay mode test, the *overlap* between the signal and background distributions is calculated for each tau feature.

The overlap of the signal and background distribution can be thought of as the shared area, or the shared integral of the two distributions. If both distributions are normalised to unity, i.e. their integral is

[6] There only around $0.5 \cdot 10^5$ to $1.8 \cdot 10^5$ taus available for the training process.

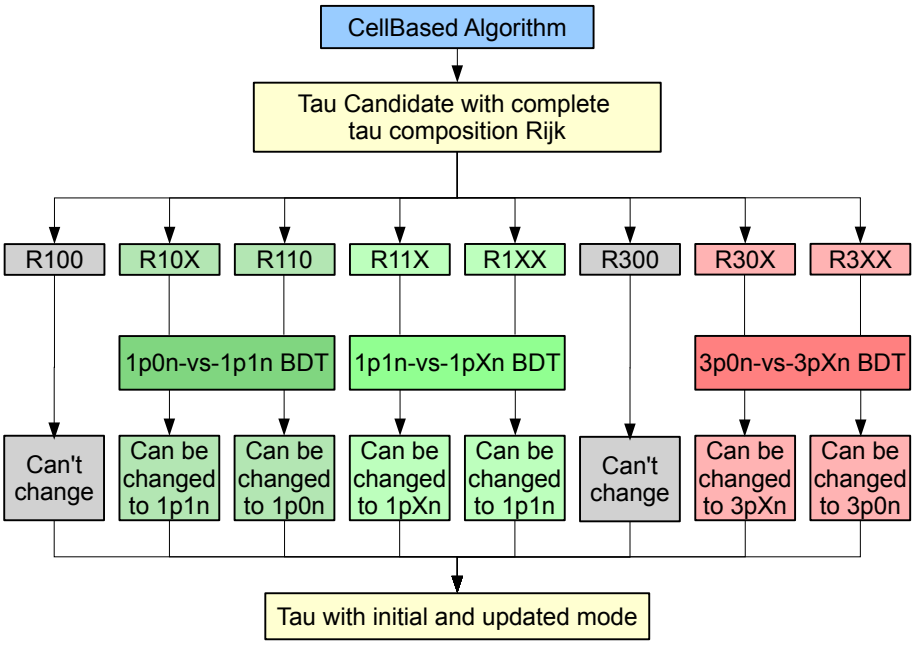
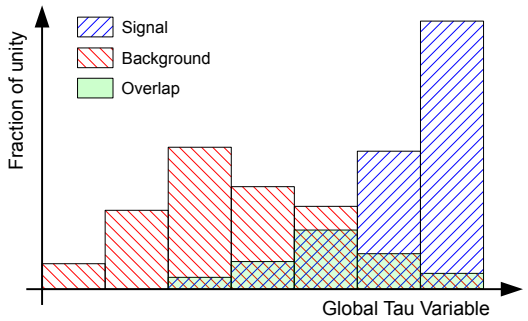
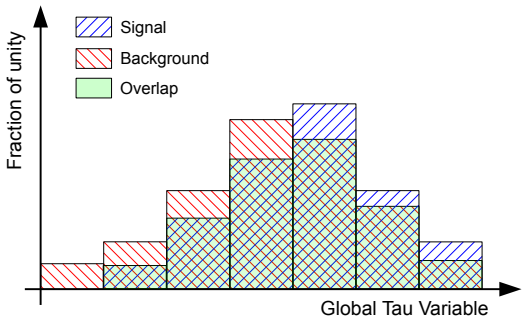


FIGURE 5.7: *Illustration of the decay mode reclassification in* `PanTau`*. The* `CellBased` *algorithm provides a tau candidate with a number of $\pi^{\pm}$-PFOs, π^0-PFOs and non-π^0-PFOs, i.e. a complete tau composition Rijk. Based on this composition, the tau is passed into one of the decay mode tests, except for R100 and R300 taus. The tests can then change the decay mode, based on the evaluation of the corresponding BDT. The result of the procedure is a tau that has an initial decay mode (the one from counting the $\pi^{\pm}$- and π^0-PFOs) and an updated mode (the one set by* `PanTau`*).*



(A) *Overlap in a variable with good separation between signal and background.*



(B) *Overlap in a variable with rather bad separation between signal and background.*

FIGURE 5.8: *Illustration of overlap between signal and background distribution for small overlap (left) and big overlap (right). The overlap is the sum over all bins, where the summands are the smaller value of the two values found in signal and background.*

equal to one, then the overlap is zero in case of completely separated signal and background and one if the two distributions are the same. The calculation of the overlap O in a variable x follows the formula

$$O(x) = \sum_{i \in \text{Bins}} \min\left(x_i(\text{Signal}), x_i(\text{Background})\right)$$

where i runs over all bins in the histogram and x_i(Signal) (x_i(Background)) is the value of the variable x in bin i for the signal (background). An illustration of the overlap is given in Figure 5.8.

For given signal and background distributions, the overlap depends on the binning of the distributions. On the one hand, if there is only one bin, the overlap will always be one, because both distributions are normalised to one and are in the only possible bin. On the other hand, with the number of bins tending to infinity, the overlap will tend to zero, because the variable will never have the exact same real value for signal and background[7]. However, the overlap value has no need to be known precisely. It is merely a mean to have a coarse sorting of the large number of variables that are calculated for each tau.

The overlap calculation is performed in each of the three decay mode tests separately and the resulting ranking is the first step towards selecting a set of variables capable of separating signal and background.

5.3.4 Variable Selection

The variable ranking obtained by the overlap calculation is only a rough guideline to which variables may be important. There are two reasons for this. The first reason is, that the overlap is calculated for all variables. That means that it is calculated, for instance, for a variable involving π^0-PFOs in R10X decays, in which there are no π^0-PFOs. While the overlap calculation can be executed in this case, its result can not be trusted, because the supposed separation comes from cases in which the variable cannot be calculated. Whether a variable, which is based on a set of types of PFOs, can be calculated or not solely depends on the presence of PFOs of the corresponding types. Therefore, instead of using a variable V with default values, it is more reasonable to simply use the multiplicity of the corresponding PFO types, that go into the calculation variable V. An example for this is shown in Figure 5.9.

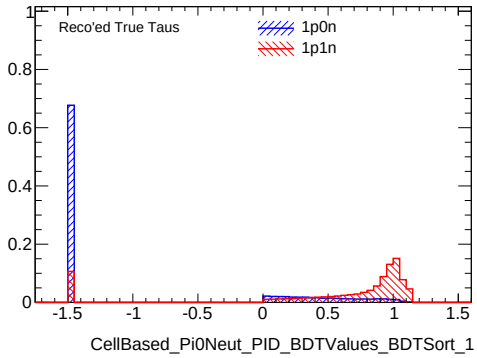
Figure 5.9a shows the highest π^0-BDT score found in the set of π^0-PFOs in the tau. For true 1p0n taus, it can happen, that no π^0-PFO is reconstructed (see enumeration in section 5.3.1, items 1 and 2). In these cases, because there is no π^0-PFO, the largest π^0-BDT score in π^0-PFOs is not defined, and a default value of -1.5 is set. This value is set exactly as many times, as there are no π^0-PFOs in a tau, i.e. it corresponds to the zero-bin in the multiplicity distribution shown in 5.9b.

Therefore, all variables that can have default peaks are excluded and will not be used in the decay mode tests.

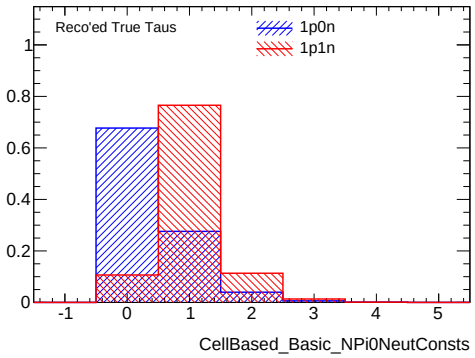
The second reason why the overlap is just a rough guideline is that it does not take into account correlations between variables. Figure 5.10 shows two variables with small overlap between signal and background and a two-dimensional plot in which the two variables are plotted against each other. The variables are $f_{\text{Neutral, All}}$ and $f_{\text{Charged, All}}$, they are calculated by

$$f_{\text{Neutral, All}} = \frac{E_T(\text{Neutral PFOs})}{E_T(\text{Core PFOs})} \qquad f_{\text{Charged, All}} = \frac{E_T(\text{Charged PFOs})}{E_T(\text{Core PFOs})}$$

[7] While this is true mathematically, it technically can happen that a variable yields the same number for signal and background, because the variables are stored as `float` numbers in C++, which have an accuracy of only 10^{-8}...10^{-7}.[50]

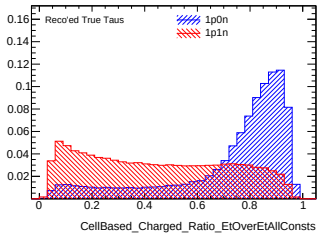


(A) *Highest BDT score found in π^0-PFOs. If no π^0-PFO was found, the default value -1.5 is filled. Because this variable can have a default value, it is discarded from the variable selection.*

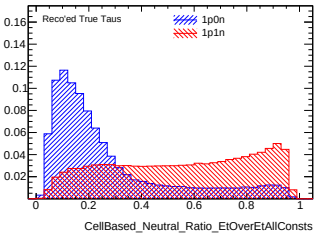


(B) *Multiplicity of π^0-PFOs. Instead of using a variable that can have default values, the multiplicity of the corresponding PFO type (here π^0-PFOs) is included, if it shows a good separation.*

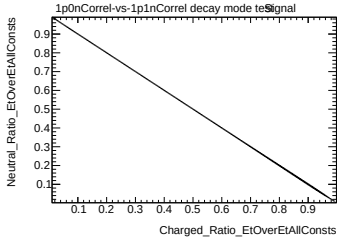
FIGURE 5.9: *Example for a variable with default peak in the 1p0n-vs-1p1n decay mode test. The overlap value in the BDT distribution (left) is highly correlated to the multiplicity of π^0-PFOs (right). Indeed, the heights of the default peaks are identical to the heights of the bin for taus with zero π^0-PFOs.*



(A) *Fraction of core energy in charged PFOs $f_{\text{Charged, All}}$.*



(B) *Fraction of core energy in neutral PFOs $f_{\text{Neutral, All}}$.*



(C) *The two variables $f_{\text{Charged, All}}$ and $f_{\text{Neutral, All}}$ plotted against each other.*

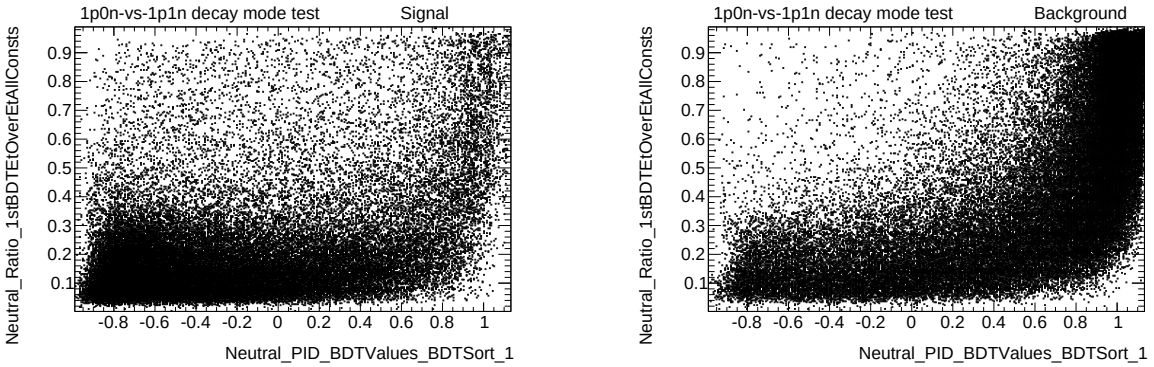
FIGURE 5.10: *The distributions of $f_{\text{Charged, All}}$ and $f_{\text{Neutral, All}}$ both indicate good separation power because of a small overlap. However, they are correlated, which eliminates the usefulness of either of them, if the other one is already included in the BDT.*

The correlation of these two variables is 100%, because $f_{\text{Neutral, All}}$ can be expressed in terms of $f_{\text{Charged, All}}$. This is due to the energy of all core PFOs being the sum of all charged and all neutral PFOs:

$$
\begin{aligned}
1 - f_{\text{Neutral, All}} &= 1 - \frac{E_T(\text{Neutral PFOs})}{E_T(\text{Core PFOs})} \\
&= \frac{E_T(\text{Neutral PFOs}) + E_T(\text{Charged PFOs})}{E_T(\text{Neutral PFOs}) + E_T(\text{Charged PFOs})} - \frac{E_T(\text{Neutral PFOs})}{E_T(\text{Neutral PFOs}) + E_T(\text{Charged PFOs})} \\
&= \frac{E_T(\text{Charged PFOs})}{E_T(\text{Neutral PFOs}) + E_T(\text{Charged PFOs})} \\
&= f_{\text{Charged, All}}
\end{aligned}
$$

Thus, using both variables in the BDTs does not give any additional separation power.

However, there are cases in which correlated variables do provide additional separation power. This



(A) *Correlation in signal sample (true 1p0n taus with one non-π^0-PFO)*

(B) *Correlation in background sample (true 1p1n taus without non-π^0-PFO)*

FIGURE 5.11: *Correlation between the highest π^0-BDT score found in neutral PFOs and the E_T fraction of the highest π^0-BDT score neutral PFO, for the signal (left) and background (right) sample. Because the correlation is much more pronounced in the background sample, the usage of both variables further improves the separation compared to using only one of the two variables.*

is the case if the correlation between the two variables is different in the signal and background sample. An example for this is shown in Figure 5.11.

It shows the correlation of two variables for the signal (i.e. true 1p0n taus with one non-π^0-PFO, Figure 5.11a) and background (i.e. true 1p1n taus without non-π^0-PFOs, Figure 5.11b) sample. The variables are V_x on the x-axis and V_y on the y-axis:

$$V_x = \text{Highest } \pi^0\text{-BDT score found in neutral PFOs}$$
$$V_y = \frac{E_T(\text{PFO}_{\text{1st BDT}})}{E_T(\text{All charged and neutral PFOs})}$$

The linear correlation coefficients obtained by TMVA are 42% (signal) and 69% (background). In an ideal case, each variable is completely uncorrelated to all other variables for both signal and background. Thus, the variable would contribute to both, the ability of the BDT to identify a tau as signal or background. If a variable shows high correlation in signal and background (c.f. the example above with $f_{\text{Charged, All}}$ and $f_{\text{Neutral, All}}$), it does not add anything to the BDT for signal or background identification capabilities. The case in which the correlations for signal and background are different sits in between these two extremes. A small correlation in the signal sample and high correlation in the background sample contributes to the performance in identifying signal taus, but not as much as it helps to identify background taus. Therefore, variables showing a different correlation in signal and background are kept, even though the correlation might be rather high in either signal or background.

The variables that are selected to enter the different decay mode tests are given in Table 5.7, their overlap values are listed in Table 5.6. The sets are obtained by selecting variables with the smallest overlap, following the guidelines described above:

- From the list of overlap values, variables that have default peaks are removed.
- From the remaining list, variables with the smallest overlap are used in the training.
- The linear correlation coefficients are evaluated. If two variables show a large correlation, the one

1p0n-vs-1p1n test	
Highest π^0-BDT score found in all neutral PFOs	0.392
Ratio of E_T in highest π^0-BDT score neutral and E_T of all core constituents	0.511
Number of photons expected in tau candidate	0.529
Sum of E_T weighted distance of charged PFOs to the tau axis	0.550
Distance in ΔR between leading neutral and leading charged PFO	0.641
1p1n-vs-1pXn test	
Second-highest π^0-BDT score found in all neutral PFOs	0.739
Number of photons expected in tau candidate	0.750
Ratio of energy in neutral PFOs divided by energy of all core PFOs	0.814
Invariant mass of all neutral PFOs	0.823
Number of neutral PFOs	0.830
3p0n-vs-3pXn test	
Ratio of energy in neutral PFOs divided by energy of all core PFOs	0.539
Highest π^0-BDT score found in all neutral PFOs	0.597
Number of photons expected in tau candidate	0.612
Ratio of standard deviation of $PFO^{\pm}$ E_T values and E_T of all core PFOs	0.747

TABLE 5.6: *Overlap values of the variables that are selected for the three BDTs. Overlap values in the 1p1n-vs-1pXn test are generally worse than in the other two tests.*

with the bigger overlap is removed. A variable is generally kept if its correlation to other variables is below $\approx 60\%$. If it shows a larger correlation in the signal (background), it is only kept if its correlation in the background (signal) differs by $\approx 20\%$.

- New variables that were not included in the first set are added to replace variables that were removed because of their correlation. However, variables are skipped if they provide the same kind of information as variables that are already in the set, for instance energy ratios (c.f. the example in Figure 5.10).

The following subsections discuss the variables for each decay mode test. The distributions of global tau variables, as well as the correlation matrices and BDT output distributions shown there are obtained from the training sample (c.f. Table A.1). The distributions from which the optimal `PanTau`-BDT thresholds are obtained are taken from the evaluation samples, which are also listed in Table A.1. Adjusting the `PanTau`-BDT thresholds corresponds to calibrating the algorithm and it is possible to have different calibrations for different samples or different purposes.

Only hadronically decaying tau leptons with a reconstructed $E_T > 15$ GeV, that lie within $\Delta R < 0.2$ of a true tau are considered in the training processes. In addition, the true taus that are reconstructed are required to have the corresponding decay mode and particle content that is required for the different BDTs. For example, for the 1p0n taus in the 1p0n-vs-1p1n decay mode test, only true 1p0n taus that are reconstructed with at least one neutral PFO are considered.

A note concerning the π^0-BDT score variables:
In `CellBased`, the π^0-BDT distribution ranges from -1.0 to $+1.0$. The threshold in the π^0-BDT score is dependent on the number of charged PFOs and the η-value of the neutral PFO in question. In `PanTau`, the π^0-BDT score that is used in the decay mode tests is corrected for the threshold, to include the η dependency of the π^0 identification. For example, if the threshold in the `CellBased` π^0-BDT for a given neutral PFO with a π^0-BDT score of 0.9 was -0.2, its π^0-BDT score in `PanTau` would be stored as

Name	Formula	Short description
Neutral_PID_BDTValues_BDTSort_1	-	Highest π^0-BDT score found in all neutral PFOs
Neutral_Ratio_1stBDTEtOverEtAllConsts	$\frac{E_T(PFO_{1.\text{BDT}})}{E_T(\text{Core PFOs})}$	Ratio of E_T in highest π^0-BDT score neutral and E_T of all core constituents
Combined_DeltaR1stNeutralTo1stCharged	$\Delta R(PFO^0_{1.E_T}, PFO^\pm_{1.E_T})$	Distance in ΔR between the leading neutral and leading charged PFO
Charged_JetMoment_EtDRxTotalEt	$\sum_{i \in PFO^\pm} E_{i,T} \cdot \Delta R(PFO^\pm_i, \tau)$	Sum of E_T weighted distance of charged PFOs to the tau axis
Neutral_Shots_NPhotonsInSeed	$\sum_{i \in \text{Shots}} N_i(\text{Photons})$	Number of photons expected in tau candidate

(A) Selected variables for the decay mode test 1p0n-vs-1p1n

Name	Formula	Short description
Neutral_PID_BDTValues_BDTSort_2	-	Second-highest π^0-BDT score found in all neutral PFOs
Neutral_HLV_SumM	$\sqrt{((\sum_{PFO^0} p(PFO^0))^2)}$	Invariant mass of all neutral PFOs
Neutral_Ratio_EtOverEtAllConsts	$\frac{E_T(\text{Neutral } PFOs)}{E_T(\text{Core PFOs})}$	Ratio of energy in neutral PFOs divided by energy of all core PFOs
Basic_NNeutralConsts	-	Number of neutral PFOs
Neutral_Shots_NPhotonsInSeed	$\sum_{i \in \text{Shots}} N_i(\text{Photons})$	Number of photons expected in tau candidate.

(B) Selected variables for the decay mode test 1p1n-vs-1pXn

Name	Formula	Short description
Neutral_PID_BDTValues_BDTSort_1	-	Highest π^0-BDT score found in all neutral PFOs
Neutral_Ratio_EtOverEtAllConsts	$\frac{E_T(\text{Neutral } PFOs)}{E_T(\text{Core PFOs})}$	Ratio of energy in neutral PFOs divided by energy of all core PFOs
Charged_StdDev_Et_WrtEtAllConsts	$E_T(\text{Core PFOs})^{-1} \cdot \left[\frac{\sum_{\text{PFO}^\pm} E_T^2}{N(\text{PFO}^\pm)} - (\frac{\sum_{\text{PFO}^\pm} E_T}{N(\text{PFO}^\pm)})^2\right]$	Ratio of standard deviation of $PFO^\pm$ E_T values and E_T of all core PFOs
Neutral_Shots_NPhotonsInSeed	$\sum_{i \in \text{Shots}} N_i(\text{Photons})$	Number of photons expected in tau candidate.

(C) Selected variables for the decay mode test 3p0n-vs-3pXn

TABLE 5.7: *Variables that enter the different decay mode tests in* PanTau. *Detailed explanations for all variables can be found in the corresponding decay mode test section, i.e. Sections 5.3.5, 5.3.6 and 5.3.7.*

1.1. This is the reason why the π^0-BDT scores in `PanTau` are not ranging from -1.0 to $+1.0$ and have sharp edges, which correspond to the BDT thresholds that are applied in `CellBased` (most prominent in Figure 5.20a).

5.3.5 1p0n-vs-1p1n Decay Mode Test

This test deals with taus that have one charged and one neutral PFO. The variables that have been chosen for the BDT are depicted in Figure 5.12. In the figures, 1p0n refers to true 1p0n taus, that are reconstructed with at least one neutral PFO. Thus, the neutral PFO does not belong to the tau. 1p1n refers to true 1p1n taus that are reconstructed with not more than one neutral PFO in total. Here, the neutral PFO is a decay product of the tau. It should be noted that R110 taus, that contain three or more reconstructed photons are excluded from this test and are reclassified as 1pXn (c.f. Section 5.3.2).

The first variable to look at is the complete distribution of the π^0-BDT score of the neutral PFO in Figure 5.12a. The substructure algorithm places a fixed cut in the distribution, which maximises the number of correctly classified taus. The signal part of the `CellBased` π^0-BDT distribution has tails reaching out to roughly -0.4 (c.f. Figure 4.13). The threshold that is applied however is much higher than that to suppress the background. By using the complete π^0-BDT distribution in conjunction with other kinematic tau variables, `PanTau` may find that a neutral may not have passed the π^0-BDT cut in `CellBased`, but that the tau otherwise is much more consistent with a 1p1n than with a 1p0n decay.

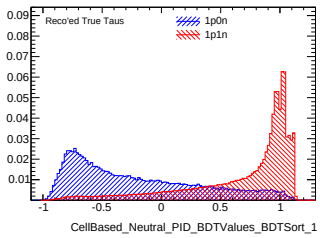
A very good kinematic variable is the E_T fraction of the neutral PFO with highest π^0-BDT score (Figure 5.12b). In true 1p0n taus, the neutral PFO comes from pile-up, noise or imperfect subtraction, rendering it rather soft. This is very different for true 1p1n decays in which it can carry a large fraction of the tau momentum because it really is a tau decay product.

Another good variable is the distance in ΔR between the charged and neutral PFO, shown in Figure 5.12c. For 1p0n taus, the distribution has two components. The first component, peaking at ΔR values below 0.05 is due to neutrals from imperfect subtraction or pile-up particles hitting the same cell as the charged PFO. The second component, ranging from 0 to 0.2 and being rather flat is coming from pile-up particles that are not associated to the tau and thus have no preferred distance to the tau axis. The distribution for true 1p1n taus is the ΔR distribution for the π^0 in a 1p1n tau decay, because here the reconstructed mode equals the true one.

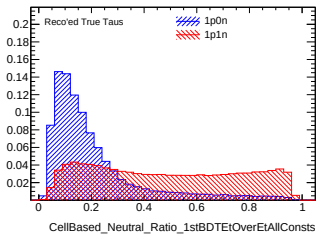
The E_T weighted ΔR between the charged PFO to the tau axis is depicted in Figure 5.12d. The tau axis is calculated by summing the 4-vectors of all charged and neutral PFOs in the tau. This makes the tau axis lie very close to the charged PFO direction for 1p0n taus, because here the neutral PFOs are of low energy. In 1p1n taus, the neutral PFO has a higher energy and is farther away from the charged PFO, making the distribution peak at higher values. The variable can be thought of as a combination of the two previous variables, which also explains its medium correlation with them (see Figure 5.13).

The variable list in the 1p0n-vs-1p1n test is concluded by the expected number of photons in the tau. This number is calculated from the energy depositions in the first layer of the EM calorimeter, also called *shots*. If the energy of a shot is smaller than ≈ 350 MeV[8], it is assumed to have no photons. If the

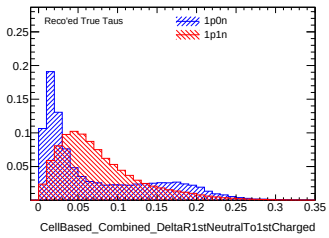
[8] The threshold is dependent on $|\eta|$. In the transition region between barrel and end-caps, $1.39 < |\eta| < 1.51$, no photons are counted.



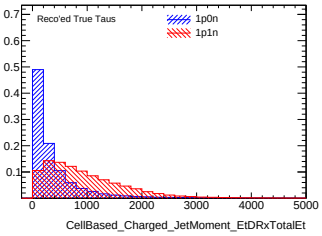
(A) *Highest π^0-BDT value found in neutral PFOs.*



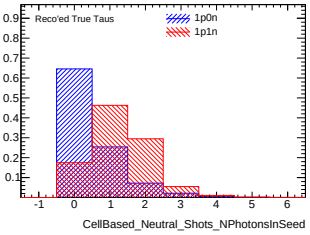
(B) *Ratio of the E_T of the neutral PFO with highest π^0-BDT score and the E_T of all neutral and charged PFOs.*



(C) *The distance in ΔR between the most energetic neutral and most energetic charged PFO.*



(D) *E_T weighted distance in $\eta - \phi$-space of charged PFO to tau axis.*



(E) *The expected total number of photons in the tau.*

FIGURE 5.12: *The set of tau variables that is used in the BDT for the 1p0n-vs-1p1n decay mode test.*

shot energy is larger than 10 GeV, two photons are counted for this shot[9]. For energies between these two values, one photon is counted per shot. The number of photons per tau is the sum of photons per shot over all neutral PFOs in the tau.

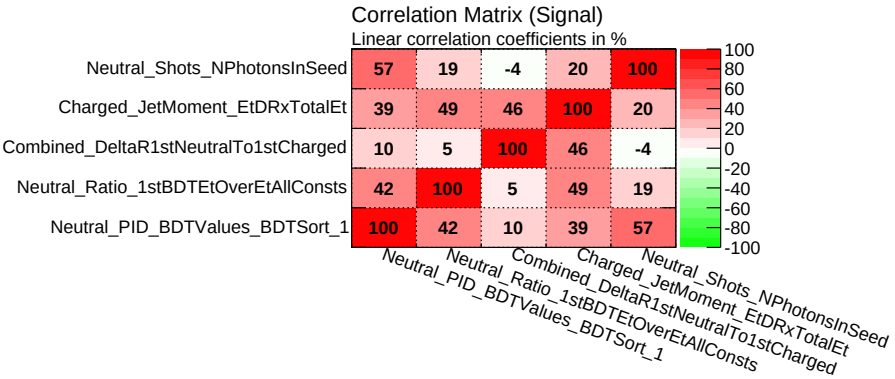
It can happen that the charged pion produces a shot which is not subtracted. The two photons can merge and not be counted as two because the energy was too small. However, referring to Figure 5.12e, most of the time there are no shots in a 1p0n tau and one or even two shots in 1p1n taus, making this variable a good addition to the 1p0n-vs-1p1n decay mode test[10].

This list of variables is utilised to train the 1p0n-vs-1p1n BDT using the TMVA software package [40]. Figure 5.13 shows the linear correlation coefficients of the variables in the list. The highest π^0-BDT score of neutral PFOs is the main variable. The next variable, the E_T fraction of the highest π^0-BDT score neutral, shows different correlations in signal and background so that it is kept. The distance between the neutral and charged PFO shows close to no correlation. Medium (small) correlation is observed in the E_T weighted ΔR of the charged PFO to the tau axis and the number of photons in the tau. The correlations are not large enough to justify dropping them so they are kept in the current version.

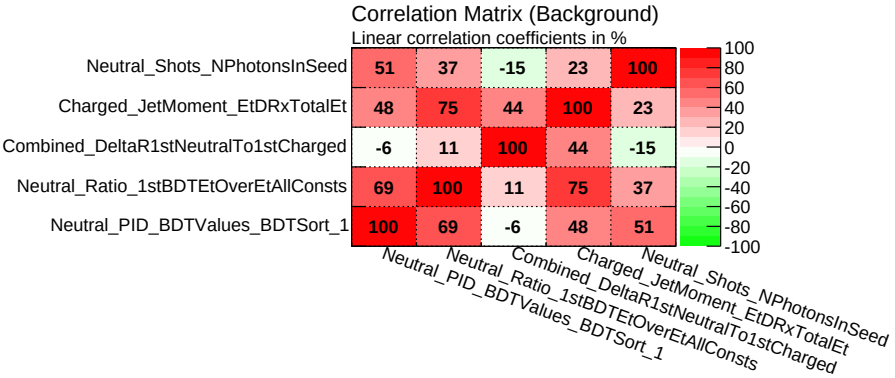
After training the 1p0n-vs-1p1n BDT on the training sample and evaluating it on a statistically inde-

[9] The assumption is that the two photons merged into one shot because the π^0 was of high energy. It is unlikely that a single photon would deposit this amount of energy in EM1.

[10] A note concerning the usage of shots: As long as shot information is not validated to be well described in simulations, only basic shat information is used. There are many variables that can be built with shots to increase the separation power. However, if their simulation does not match what is observed in real data, they won't serve their purpose.

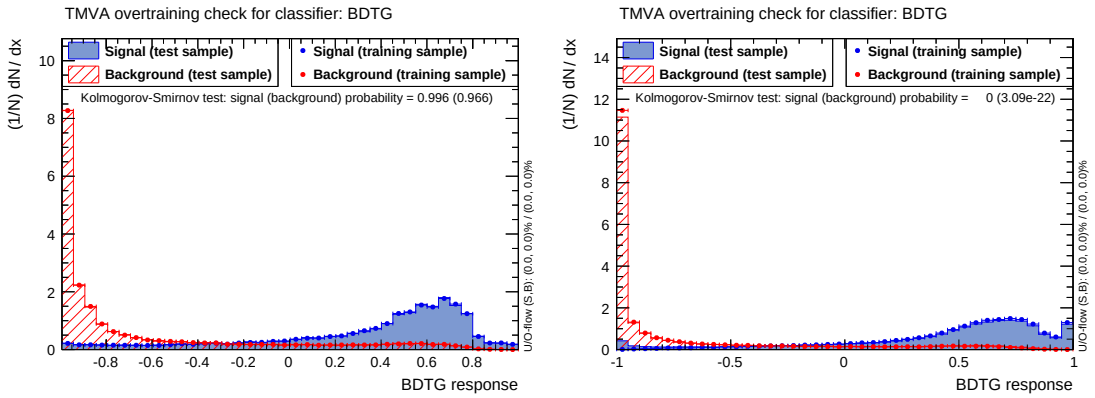


(A) *Signal taus (i.e. true 1p0n taus reconstructed with at least one neutral PFO).*



(B) *Background taus (i.e. true 1p1n taus reconstructed not more than one neutral PFO).*

FIGURE 5.13: *Linear correlation coefficients for signal (top) and background (bottom) of the variables in the 1p0n-vs-1p1n BDT.*



(A) *Nominal BDT setup, no over-training is observed.*

(B) *Artificial setup to enforce over-training.*

FIGURE 5.14: *1p0n-vs-1p1n BDT distribution for signal (true 1p0n, reconstructed with at least one neutral PFO, blue) and background (true 1p1n, reconstructed with not more than one neutral PFOs, red) taus, as obtained in the training (dots) and test (filled) samples. The nominal training setup is shown in the left Figure 5.14a. An artificial setup that forces the BDT to be over-trained is shown in the right plot (5.14b). The differences are not very significant, the effect is more pronounced in the artificially over-trained BDTs in Figures 5.18b and 5.22b.*

Configuration parameter	Nominal	Over-training
Number of trees	500	500
Boost type	Gradient boosting	Gradient boosting
Shrinkage	0.3	0.3
Use bagged gradient	True	True
Gradient bagging fraction	0.6	0.6
Separation Type	Gini index	Gini index
Number of cuts	20	20
Prune method	Cost complexity	None
Prune Strength	50	None
Max number of nodes	5	100
Max depth	3	100

TABLE 5.8: *Nominal and over-training configuration parameters of the three BDTs used in* `PanTau`. *The highly increased number of nodes and depth in the over-training configuration will make the BDT subdivide the sample into very small parts, in which statistical fluctuations become significant. Because these fluctuations are different in an independent test sample, the performance is different here.*

pendent test sample[11], the distributions shown in Figure 5.14a are obtained. It is important to note that there is no *over-training* in the BDT distribution 5.14a, because the training and test distributions agree within statistical uncertainties.

The nominal BDT configuration is listed in Table 5.8 and corresponds to the default configuration for gradient BDTs in the TMVA package.

To illustrate the effect of *over-training*, the training was repeated with a different BDT configuration. In this configuration[12], the depth of the BDT and number of nodes is increased to 100 (up from 3 and 5), and the removal of insignificant nodes (pruning) is switched off (it is on in the nominal configuration). As a result, the BDT will segment the variable space with a much higher granularity, becoming sensitive to individual tau decays and statistical fluctuations. Those fluctuations are properties of the training sample, but not of tau decays in general. This leads to a much better classification on the training sample, than on a statistically independent testing sample. Thus, as constructed, the resulting BDT distributions in Figure 5.14b show a disagreement between training and testing samples.

The differences are seen in both signal and background taus, and are most prominent at small and high BDT values. For this BDT, the largest difference between training and testing samples is observed at a BDT value of -1. Here, the background training sample has more entries than the testing sample and the signal training sample has less entries than the testing sample, indicating an overestimation of the BDT performance.

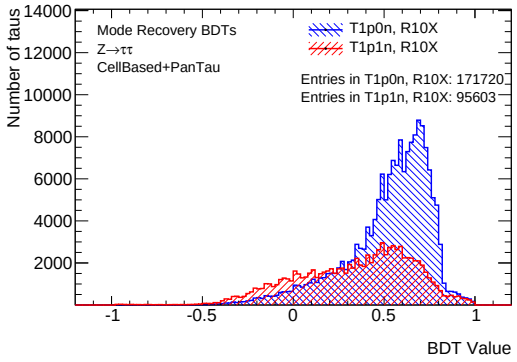
For illustration purposes, the additional configuration with enforced over-training is also done for the two other BDTs, where the effects are more pronounced (Sections 5.3.6 and 5.3.7).

As mentioned before, the 1p0n-vs-1p1n BDT is evaluated for two complete tau compositions, the R10X and the R110 composition. For both compositions, the BDT score distribution is plotted for the corresponding signal and background decay modes[13]. The distributions are shown in Figures 5.15a and 5.15c. There is no normalisation applied, because the figure of merit to be maximised is the absolute number of correctly classified taus. Thus, the branching ratios play an important role. If the absolute number of complete taus is to be maximised, it is most important to correctly reconstruct the most dominant mode, i.e. the 1p1n decay mode. However, the branching ratios are reflected in the abundance

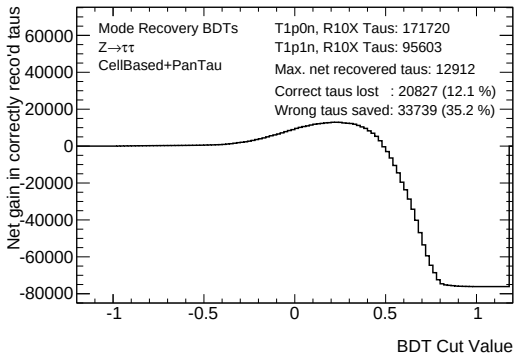
[11] The sample that is used for training is given in Table A.1. The testing sample is also listed in that table and is the one containing `tid00999073`.

[12] This configuration is used for illustration purposes only.

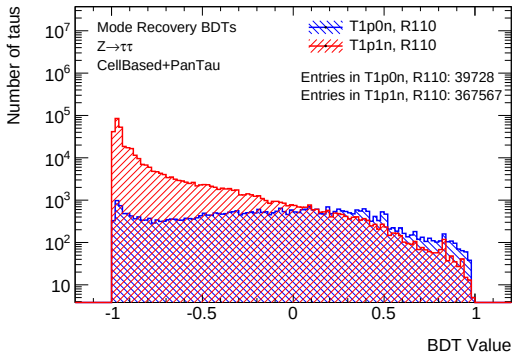
[13] These distributions are extracted from the evaluation samples, c.f. Table A.1, to calibrate the algorithm to the sample at hand.



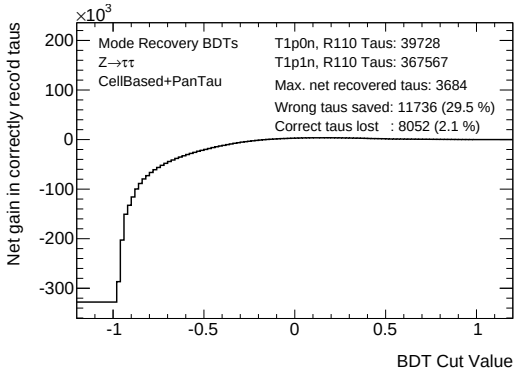
(A) *BDT distribution for signal (true 1p0n) and background (true 1p1n) in the R10X part of the 1p0n-vs-1p1n decay mode test*



(B) *Scan through the R10X part of the 1p0n-vs-1p1n BDT, to find the optimal cut value.*



(C) *BDT distribution for signal (true 1p1n) and background (true 1p0n) in the R110 part of the 1p0n-vs-1p1n decay mode test. The y-axis is logarithmic.*



(D) *Scan through the R110 part of the 1p0n-vs-1p1n BDT, to find the optimal cut value.*

FIGURE 5.15: *BDT distributions and BDT scans for the 1p0n-vs-1p1n decay mode test.*

Complete tau composition	R10X	R110
Correct taus from `CellBased`	171,720	367,587
Wrong taus from `CellBased`	95,603	39,728
Reclassification	1p0n→1p1n	1p1n→1p0n
BDT threshold to reclassify	< 0.22	> 0.14
Correctly reclassified	35.2%	29.5%
Wrongly reclassified	12.1%	2.1%
Net gain in correct taus	12,912	3,684

TABLE 5.9: *Optimal BDT thresholds for reclassifications and fractions of correctly (wrongly) reclassified taus. The fraction of correctly reclassified taus is the fraction of (wrong) background taus that are reclassified. The fraction of wrongly reclassified taus is the fraction of (correct) signal taus that are reclassified.*

of taus of a given decay mode, i.e. their number.

In Figure 5.15a, only R10X taus are considered and they are split into signal (blue, true 1p0n taus with at least one non-π^0-PFO) and background (red, true 1p1n with not more than one neutral PFO). Every tau that enters this distribution was classified as a 1p0n tau by `CellBased`. Although the two distributions overlap in large parts, it is clearly visible that R10X taus with a BDT value smaller than a threshold of ≈ 0.2 are more likely to be true 1p1n taus. Therefore, R10X taus that fall below that threshold have their decay mode *changed* from 1p0n to 1p1n by `PanTau`. Because there are more true 1p1n taus than 1p0n taus below the threshold, `PanTau` (on average) classifies more taus correctly.

To exactly determine the threshold and to count how many taus are migrated from a wrong `CellBased`-mode to a correct mode (and vice versa), the BDT distribution in Figure 5.15a is scanned (compare to Fig. 5.15b). For each bin, it is counted how many taus would be classified as 1p0n and 1p1n, if the threshold was set in the corresponding bin. With a threshold of -1.0 for reclassification to 1p1n, all taus remain 1p0n taus. Increasing the threshold from -1.0 to 0.2, more and more taus are classified as 1p1n. In this range, there are more true 1p1n than there are true 1p0n, which leads to the number of correctly reclassified taus growing faster than the number of wrongly reclassified taus. The difference between those numbers is the net gain in correctly reconstructed taus. Moving from a threshold of 0.2 to $+1.0$, the number of wrongly reclassified taus grows much faster than the number of correctly reclassified taus, decreasing the net gain and making it negative in the case of reclassifying every R10X tau to 1p1n. The threshold for 1p0n→1p1n reclassification corresponds to the BDT score of the maximum of the distribution shown in Figure 5.15b.

The same procedure is repeated for the R110 complete tau composition (Figures 5.15c and 5.15d) A difference to the R10X part is that signal and background modes change their roles, so that the reclassification 1p1n→1p0n of a R110 tau is performed if its BDT score is larger than the threshold. Table 5.9 lists the BDT thresholds and the fractions of correctly (wrongly) reclassified taus.

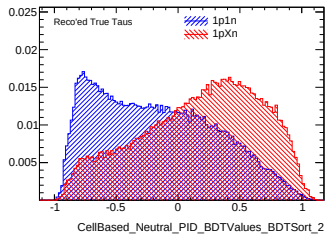
5.3.6 1p1n-vs-1pXn Decay Mode Test

The second test takes care of 1-prong taus with one π^0-PFO and at least one additional neutral PFO. Here, 1p1n taus refer to true 1p1n taus reconstructed with one charged PFO, and two or more neutral PFOs. 1pXn taus mean true 1pXn taus reconstructed with one charged PFO and two or more neutral PFOs. For 1p1n taus, the second neutral PFO does not belong to the tau decay, whereas the two neutral PFOs in 1pXn taus do.

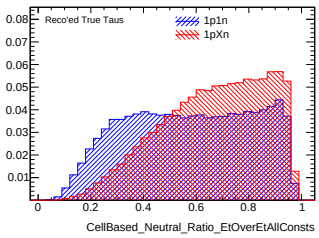
R11X taus that contain three or more reconstructed photons are migrated to 1pXn and are not passed to this decay mode test (c.f. Section 5.3.2). The set of variables used in this decay mode test is shown in Figure 5.16.

Similar to the 1p0n-vs-1p1n decay mode test, the main variable is the π^0-BDT score (Figure 5.16a). However, in this decay mode test, the most interesting π^0-BDT score is the score of the second neutral PFO, i.e. the neutral PFO with the second highest π^0-BDT score. This neutral PFO is the non-π^0-PFO in the 1p1n decays and one of the two π^0-PFOs in the 1pXn decays. The cut to classify a neutral PFO as π^0-PFO is optimal on average (like in the 1p0n-vs-1p1n case), but not optimal when including information from other kinematic variables.

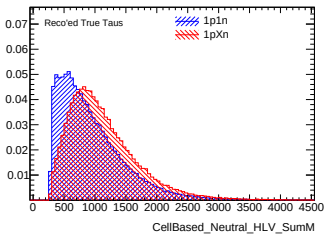
The E_T fraction of neutral PFOs in Figure 5.16b shows some separation between the two modes. This is because for 1pXn, both neutrals belong to the tau decay, while the second neutral in 1p1n decays does



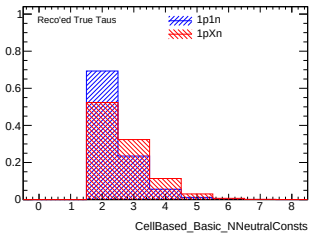
(A) *Second highest π^0-BDT score found in the set of neutral PFOs of a tau.*



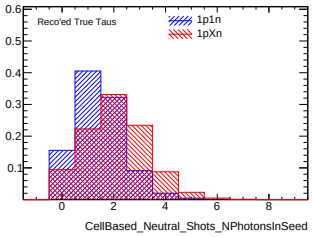
(B) *Ratio of transverse energy of all neutral PFOs and transverse energy of all neutral and charged PFOs.*



(C) *Invariant mass of the 4-vector sum of neutral PFOs.*



(D) *Number of neutral PFOs in the tau.*



(E) *The expected total number of photons in the tau.*

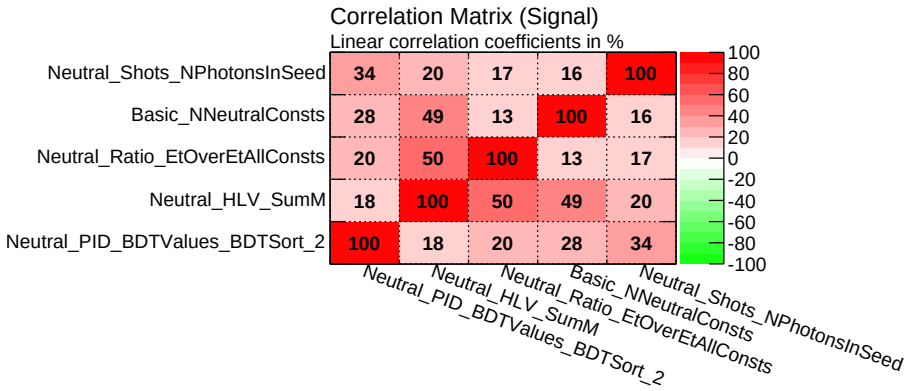
FIGURE 5.16: *Kinematic variables that are exploited in the 1p1n-vs-1pXn BDT. Explanations are in the text.*

not. Hence following the arguments given in the 1p0n-vs-1p1n decay mode test, lower fractions are expected for 1p1n taus. The separation is not as good as in 1p0n-vs-1p1n, because of the one neutral PFO that belongs to the decay in both cases. This harder neutral PFO very roughly has an E_T fraction as shown in Figure 5.12b, i.e. covering huge parts of the possible spectrum. On top of that, there is an unassociated neutral PFO with preferably lower energy because it comes from pile-up or imperfect subtraction in the 1p1n taus. However, in 1pXn taus there also is a second neutral PFO with less energy than the other neutral PFO, making a separation very difficult.

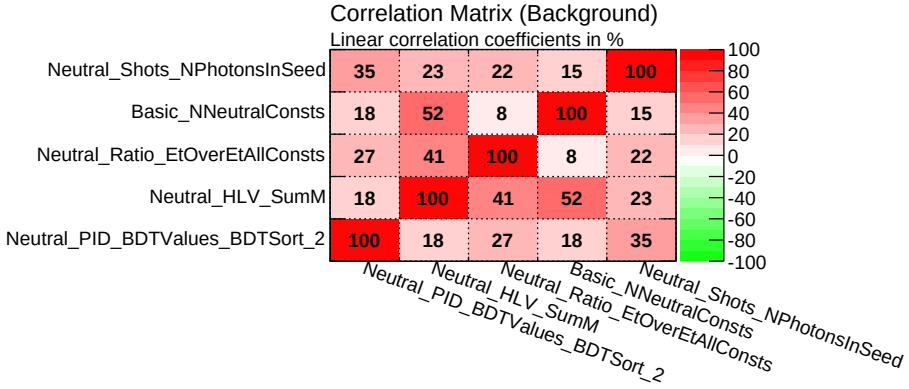
The invariant mass of all neutral PFOs (Figure 5.16c) exploits the overall difference in the four momenta between the additional neutral PFO in 1p1n and the π^0-PFO in 1pXn decays. It is expected to be lower if one of the particles is less energetic or the distance between the neutral particles is larger. While this behaviour is observed, the difference is not very significant.

The number of neutral PFOs provides a bit separation power between 1p1n and 1pXn, see Figure 5.16d. This, combined with the low correlation to most other variables is the motivation to include this variable.

Figure 5.16e shows the total number of photons in the tau candidate. The calculation of this variable is explained in 5.3.5. Ideally, there are at least 4 photons in 1pXn decays, if there is no photon merging. If there is photon merging of two, three or a pair-wise merging of photons into the same cluster, the number can decrease to 2 photons. This means that on average, there should be more photons in 1pXn taus than in 1p1n taus, because the latter have only one π^0.

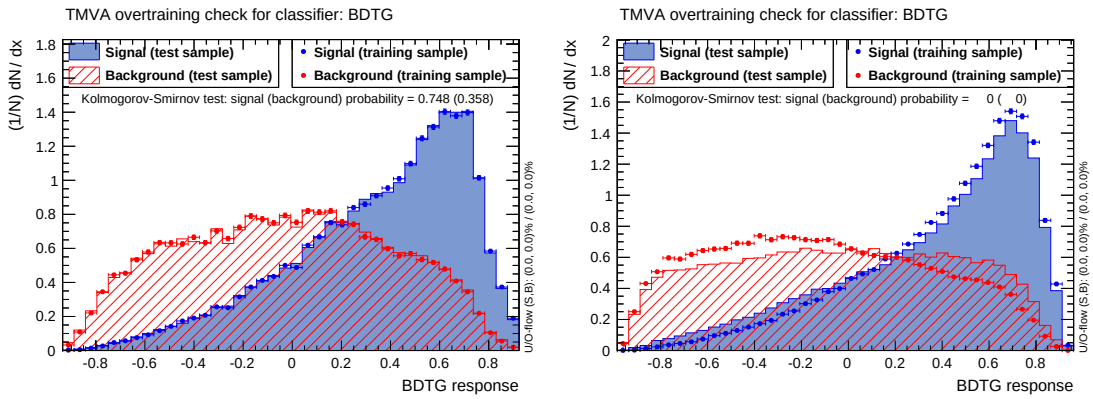


(A) *Signal taus (i.e. true 1p1n taus reconstructed with at least two neutral PFOs).*



(B) *Background taus (i.e. true 1pXn taus reconstructed with at least two neutral PFOs).*

FIGURE 5.17: *Linear correlation coefficients for signal (top) and background (bottom) of the variables in the 1p1n-vs-1pXn BDT.*



(A) *Nominal BDT configuration, no over-training is observed.*

(B) *BDT configuration enforcing over-training.*

FIGURE 5.18: *1p1n-vs-1pXn BDT distribution for signal (true 1p1n, reconstructed with at least two neutral PFOs, blue) and background (true 1pXn, reconstructed with at least two neutral PFOs, red) taus, as obtained in the training (dots) and test (filled) samples. The left Figure 5.18a shows the nominal BDT configuration. Figure 5.18b represents the configuration that over-trains the BDT.*

The linear correlation coefficients for the variables used in the 1p1n-vs-1pXn BDT are shown in Figure 5.17. The main variable is the second highest π^0-BDT score found in neutral PFOs. Hardly correlated to that is the invariant mass of all neutral PFOs in the tau. Small to medium correlations are observed in the neutral E_T fraction and the multiplicity of neutral PFOs. The number of photons in the seed shows small correlations.

In general, correlations in this case are smaller than in the 1p0n-vs-1p1n case. This is due to the fact that taus in this test contain more PFOs to create variables from, making it easier to construct uncorrelated variables. This is countered by the difficulties to create variables with high separation power, because there are more PFOs in the core region of the tau, making them overlap and harder to reconstruct and identify. A finer granularity of the calorimeter could help to disentangle 1p1n from 1pXn taus.

Training the BDT on the training sample and evaluating it on a test sample with TMVA yields the BDT distributions shown in Figure 5.18a. No over-training is observed in the nominal configuration.

The BDT setup that is designed to produce over-training nicely shows the effect. For signal taus, the BDT distribution of the training sample is shifted towards higher BDT values in comparison to the testing sample. Comparing the training sample of background taus to their testing sample, a shift towards lower BDT values is observed. Hence, using the over-training configuration, the separation power between the two modes is overestimated.

As listed in Table 5.5, there are two complete tau compositions that enter the 1p1n-vs-1pXn decay mode test, namely R11X and R1XX. In these two 1p1n-vs-1pXn decay mode tests, the same procedure as in the 1p0n-vs-1p1n decay mode tests is applied. The BDT score is evaluated for both complete tau compositions and the distributions for signal and background are shown in Figures 5.19a (R11X part) and 5.19c (R1XX part). All taus in the R11X (R1XX) distribution have been classified as 1p1n (1pXn) by `CellBased`. In the R11X case, while there is quite a significant amount of wrongly classified true 1pXn in there ($\approx$ 26%), it cannot be recovered because it is hidden under the majority of correctly classified 1p1n taus. The picture changes in the R1XX case; the contribution of wrongly classified taus by `CellBased` is $\approx$ 46%[14], of which 61% can be recovered. By doing so, 31% of the correctly classified taus are migrated to 1p1n.

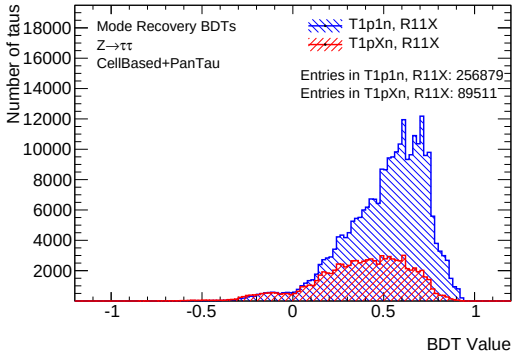
The reclassifications and thresholds are listed in Table 5.10.

5.3.7 3p0n-vs-3pXn Decay Mode Test

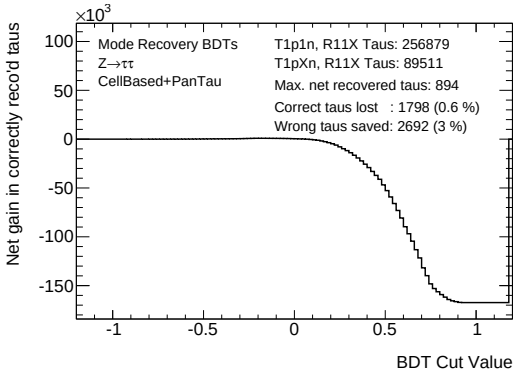
Finally, the 3p0n-vs-3pXn decay mode test deals with three prong taus. Only three prong taus that have at least one neutral PFO enter this test. Hence, the 3p0n refers to true 3p0n taus reconstructed with three charged PFOs and at least neutral PFO and 3pXn refers to true 3pXn taus reconstructed with three charged and at least one neutral PFO. Very similar to the 1p0n-vs-1p1n decay mode test, the neutral PFO does not belong to the tau for 3p0n, while it does belong to the tau in the case of 3pXn. Figure 5.20 shows all the variables that are used in the 3p0n-vs-3pXn decay mode test.

Just like in the two other tests, the main variable is the π^0-BDT score of the neutral with the highest π^0-BDT score. The idea here is the same as in the two other tests, namely that the π^0-BDT cuts in `CellBased` are optimal on average, but may be improved when taking other kinematic tau variables into consideration. Also, the sharp edges due to the shift of the BDT-scores by their respective threshold in `CellBased` are visible (at $\approx$ 0.3).

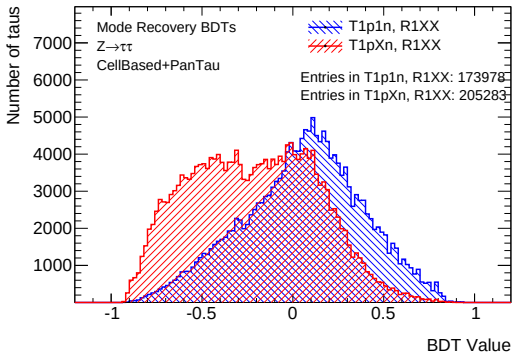
[14] The $\approx$ 46% are obtained by relating the $\approx$ 174000 wrongly classified R1XX to the $\approx$ 205000 correctly classified R11X.



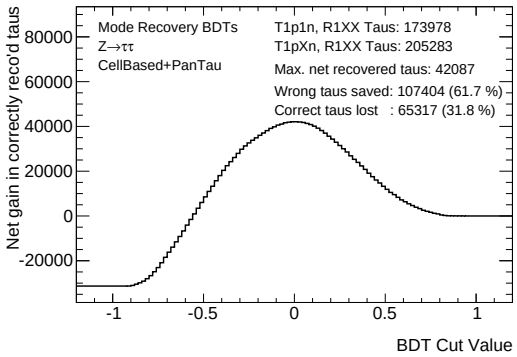
(A) *BDT distribution for signal (true 1p1n) and background (true 1pXn) in the R11X part of the 1p1n-vs-1pXn decay mode test*



(B) *Scan through the R11X part of the 1p1n-vs-1pXn BDT, to find the optimal cut value.*



(C) *BDT distribution for signal (true 1pXn) and background (true 1p1n) in the R1XX part of the 1p1n-vs-1pXn decay mode test*

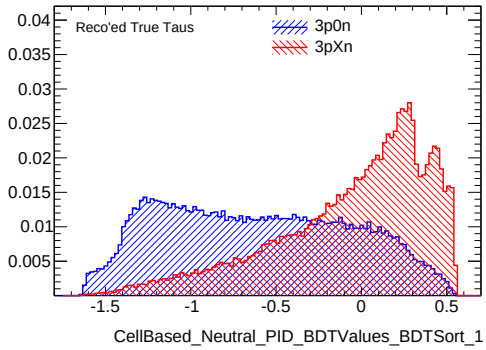


(D) *Scan through the R1XX part of the 1p1n-vs-1pXn BDT, to find the optimal cut value.*

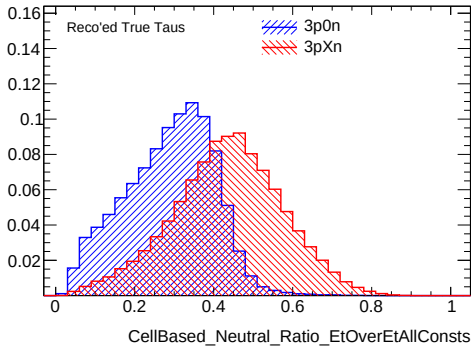
FIGURE 5.19: *BDT distributions and BDT scans for the 1p1n-vs-1pXn decay mode test.*

Complete tau composition	R11X	R1XX
Correct taus from `CellBased`	256,879	205,283
Wrong taus from `CellBased`	89,511	173,978
Reclassification	1p1n→1pXn	1pXn→1p1n
BDT threshold to reclassify	< -0.18	> 0.00
Correctly reclassified	3.0%	61.7%
Wrongly reclassified	0.6%	31.8%
Net gain in correct taus	894	42,087

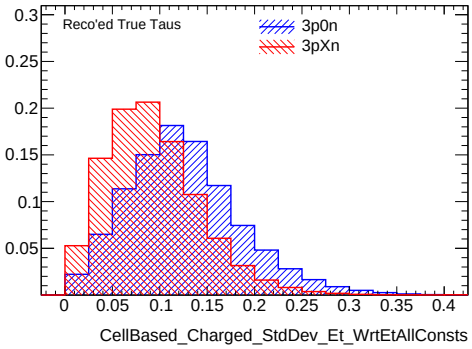
TABLE 5.10: *Optimal BDT thresholds for reclassifications and fractions of correctly (wrongly) reclassified taus. The fraction of correctly reclassified taus is the fraction of (wrong) background taus that are reclassified. The fraction of wrongly reclassified taus is the fraction of (correct) signal taus that are reclassified.*



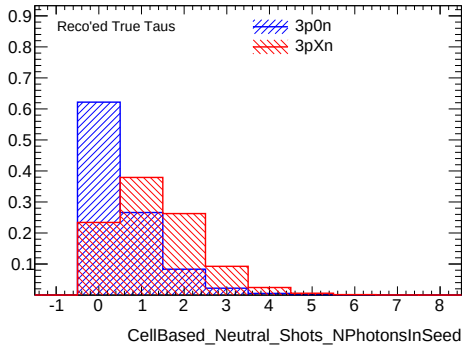
(A) *Highest π^0-BDT value found in neutral PFOs.*



(B) *Ratio of transverse energy of all neutral PFOs and transverse energy of all neutral and charged PFOs.*



(C) *Standard deviation of the E_T values of the three charged PFOs, divided by the E_T of all charged and neutral PFOs.*



(D) *The expected total number of photons in the tau.*

FIGURE 5.20: *The 3p0n-vs-3pXn BDT makes use of these variables. They are explained in the text.*

The first kinematic variable is the ratio of the energy in the non-π^0-PFOs and the energy in all PFOs (Figure 5.20b). Similarly to the two other tests, the non-π^0-PFO in true 3p0n taus does not belong to the decay, so that its energy tends to be smaller than the energy of a π^0 in a 3pXn decay which was not flagged as such.

Figure 5.20c shows the standard deviation of the set of E_T values of charged PFOs. It is obtained by putting the E_T value of every single charged PFO into a set (i.e. the set will have three numbers in it) and then calculating the standard deviation of that set. It is lower for 3pXn taus, because the additional neutral PFO belongs to the tau and uses up available phase space on its own, effectively reducing the available phase space for the charged PFOs. This results in more confined E_T values, making the standard deviation smaller than in the case in which the neutral does not belong to the decay.

Finally, as in the two other BDTs, the last variable is the number of photons in the seed. It is shown in Figure 5.20d. Because of the non-π^0-PFO being a real π^0, 3pXn decays on average have more photons in the decay than 3p0n decays.

Another input variable, which improves the 3p0n-vs-3pXn performance, is the invariant mass of the charged PFOs. For true 3p0n decays, the invariant mass spectrum should be the one of the a_1 resonance, while for true 3pXn decays, the invariant mass of the charged subsystem is lower because of the additional neutral. However, at the time of creating the variable and training setup, usage of invariant masses

in the decay mode classification was not allowed, because they have been studied in tau identification, to reduce jet backgrounds. It turned out, that the invariant masses did not significantly improve the tau identification performance, so that they could be used in PanTau. Thus, studies have been carried out to look into how including the masses would improve the classification. The result was, that the mass in $\tau_{\text{1-prong}}$ only improved the performance slightly, so that it is also not used there. For $\tau_{\text{3-prong}}$, as mentioned, the performance increases and the mass is used in a newer version of the PanTau algorithms, which runs in a newer release of the ATLAS event reconstruction.

The linear correlation coefficients for the variable set discussed above are shown in Figures 5.21a and 5.21b for signal (3p0n) and background (3pXn) taus. The coefficients are very similar for signal and background, only real exception being the correlation between the standard deviation and the E_T ratio.

The correlation between the E_T ratio and the BDT value is due to real neutral pions having a higher E_T than their background. Both of these variables are rather uncorrelated with the standard deviation of the charged PFOs. A negative correlation between the standard deviation and the E_T ratio can be understood such that with increasing energy of the neutral PFO, the fewer phase space is left for the charged PFOs, leaving little room for large deviations in their energy. While this holds true for the background taus (3pXn), for 3p0n taus it cannot be the reason. For 3p0n taus, it could be due to the E_T cut on the tau candidates that enter the training. Only taus above 15 GeV are considered in the training, so that for a 15 GeV tau with a neutral fraction of 40% for example limits the total energy of the charged system to 9 GeV. Thus, the higher the neutral E_T ratio, the lower the variance in the charged system. The number of photons is positively correlated to the BDT score of the non-π^0-PFO, because with increasing BDT values, the more likely it is that the neutral PFO indeed was a π^0, which ideally is reconstructed with two photons. The smaller correlation between the number of photons and the E_T ratio is due to the E_T dependent counting of photons.

After training and testing the BDTs with the variables listed above in the nominal and over-training configuration, the Figures 5.22a (nominal) and 5.22b (over-training) are obtained. There is no to little over-training in the nominal configuration.

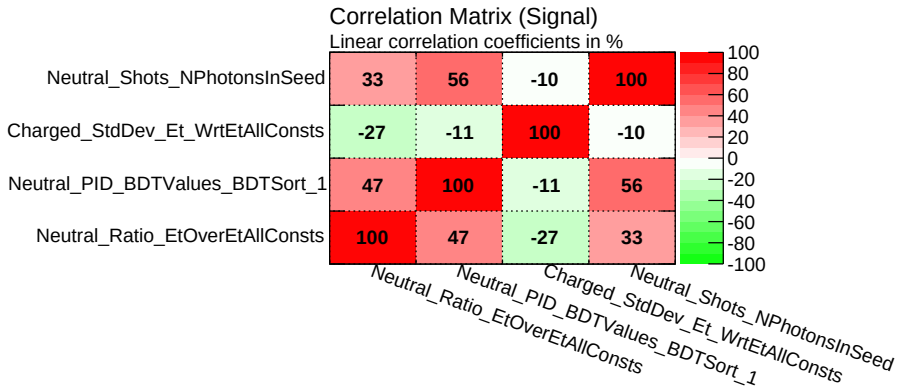
Over-training in the over-training configuration is much more pronounced and subsequent bins of the training sample lie above the testing sample, for signal and for background taus.

Since there are only single and not subsequent bins that do not agree between training and testing sample in the nominal configuration, it is used as it is.

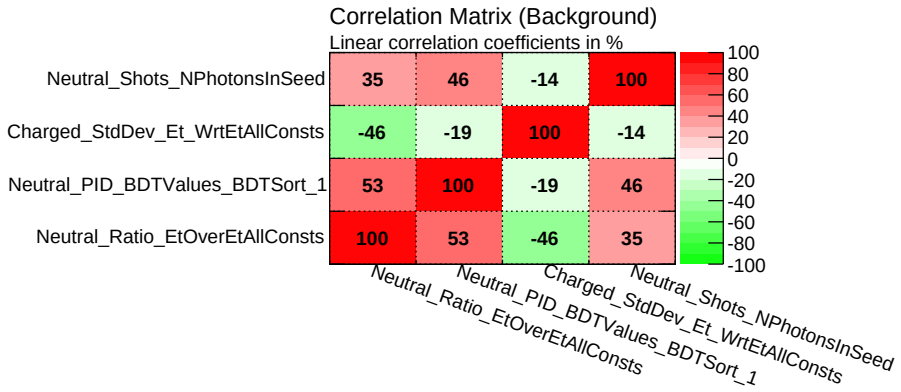
The 3p0n-vs-3pXn BDT is evaluated for the two complete tau compositions R30X and R3XX. Figure 5.23a shows the BDT distribution for the R30X case. All taus in this distribution are considered as 3p0n by CellBased, although roughly one third are true 3pXn taus. A scan through the BDT distribution (completely analogous to the 1p0n-vs-1p1n and 1p1n-vs-1pXn case) gives the contents of Figure 5.23b. Using a BDT value of -0.05 as the threshold below which to reclassify a tau to be 3pXn, the recovery is as follows. 30.3% of the 3pXn taus in the R30X composition can be recovered to be classified as 3pXn, while at the same time 3.8% of the correctly classified 3p0n are migrated to 3pXn. The net gain is $15 \cdot 10^3$ taus.

The same information is depicted for the R3XX complete tau composition, in Figures 5.23c and 5.23d. In this case, all taus have been classified as 3pXn by CellBased, albeit $\approx$ 35% are true 3p0n taus. If all R3XX taus with a BDT value above -0.02 are reclassified as 3p0n, 69.7% of the wrongly classified taus are corrected, and 21.7% of the correctly classified 3pXn taus are migrated to 3p0n.

These numbers are also summarised in Table 5.11.

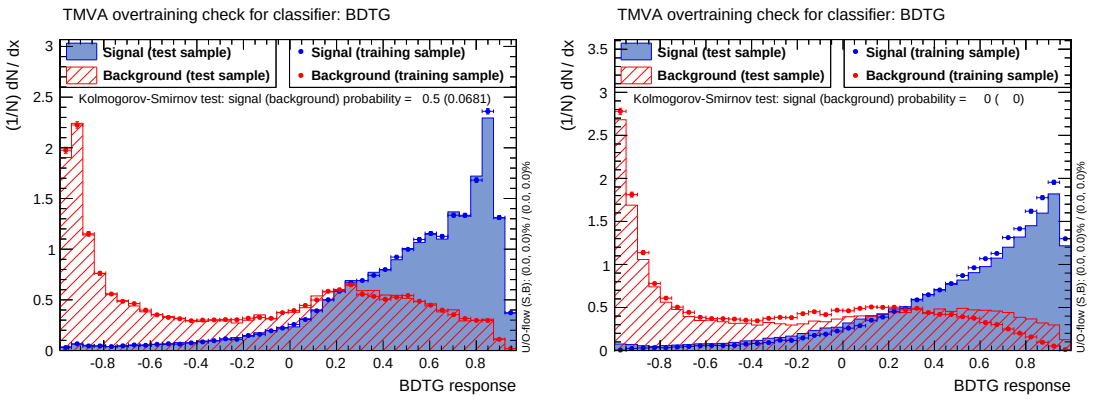


(A) *Signal taus (i.e. true 3p0n taus reconstructed with at least one neutral PFO).*



(B) *Background taus (i.e. true 3pXn taus reconstructed with at least one neutral PFO).*

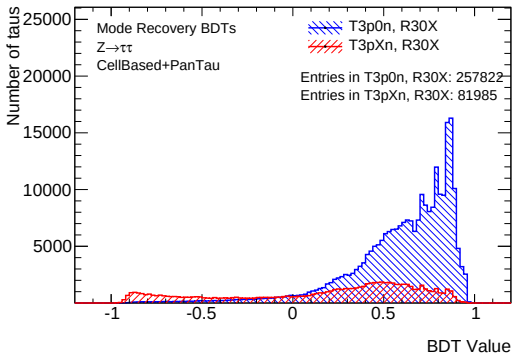
FIGURE 5.21: *Linear correlation coefficients for signal (top) and background (bottom) of the variables in the 3p0n-vs-3pXn BDT.*



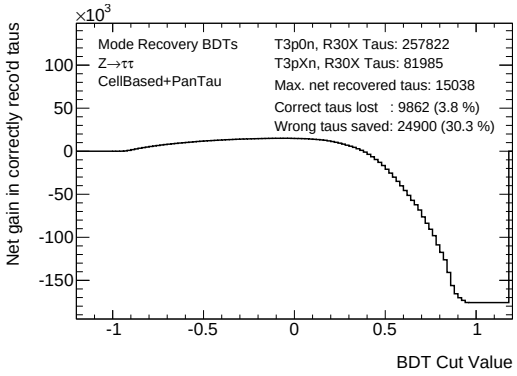
(A) *Nominal BDT configuration, no to very little over-training is observed.*

(B) *BDT configuration enforcing over-training.*

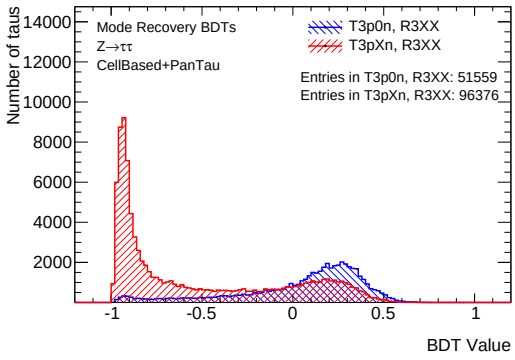
FIGURE 5.22: *3p0n-vs-3pXn BDT distribution for signal (true 3p0n, reconstructed with at least one neutral PFO, blue) and background (true 3pXn, reconstructed with at least one neutral PFO, red) taus, as obtained in the training (dots) and test (filled) samples. The two figures show the nominal (Fig. 5.22a) and over-training (Fig. 5.22b) configuration of this BDT.*



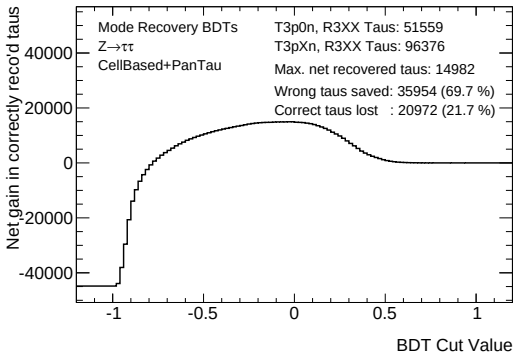
(A) *BDT distribution for signal (true 3p0n) and background (true 3pXn) in the R30X part of the 3p0n-vs-3pXn decay mode test*



(B) *Scan through the R30X part of the 3p0n-vs-3pXn BDT, to find the optimal cut value.*



(C) *BDT distribution for signal (true 3pXn) and background (true 3p0n) in the R3XX part of the 3p0n-vs-3pXn decay mode test*



(D) *Scan through the R3XX part of the 3p0n-vs-3pXn BDT, to find the optimal cut value.*

FIGURE 5.23: *BDT distributions and BDT scans for the 3p0n-vs-3pXn decay mode test.*

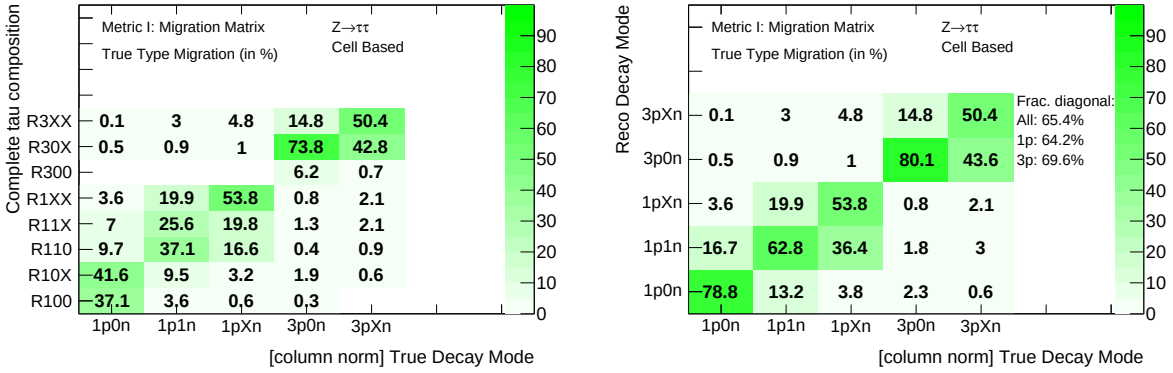
Complete tau composition	R30X	R3XX
Correct taus from `CellBased`	257,822	96,376
Wrong taus from `CellBased`	81,985	51,559
Reclassification	3p0n→3pXn	3pXn→3p0n
BDT threshold to reclassify	< -0.05	> -0.02
Correctly reclassified	30.3%	69.7%
Wrongly reclassified	3.8%	21.7%
Net gain in correct taus	15,038	14,982

TABLE 5.11: *Optimal BDT thresholds for reclassifications and fractions of correctly (wrongly) reclassified taus. The fraction of correctly reclassified taus is the fraction of (wrong) background taus that are reclassified. The fraction of wrongly reclassified taus is the fraction of (correct) signal taus that are reclassified.*

5.3.8 Improvements To Decay Mode Reconstruction

The last section explained how the different BDT tests work and how they are evaluated. This section will discuss the combination and calculate the improvement that can be expected in the classification performance.

It is helpful to look at a different kind of migration matrix, which shows the true mode on the x-axis and the complete tau compositions reconstructed by `CellBased` on the y-axis. That migration matrix is shown in Figure 5.24a. For comparison, the default `CellBased` efficiency matrix is shown in Figure 5.24b.



(A) `CellBased` migration matrix comparing the true decay mode to the complete tau composition.

(B) Default `CellBased` efficiency matrix, with the reconstructed decay mode on the y-axis.

FIGURE 5.24: *Left: Migration matrix showing the migration from true modes (x-axis) to the different complete tau compositions in* `CellBased` *(y-axis). The normalisation is such that the sum of entries in each column equals to 100, entries are given in percent. For instance, in the R100 row and the 1p0n column, it can be read off, that* 37.1% *of true 1p0n are classified as R100 taus.*
Right: Default `CellBased` *efficiency matrix. The pairs of the rows R100 and R10X, R110 and R11X, R300 and R30X are combined in this matrix, as both components of a given pair refers to the same decay mode in terms of PFO-counting. For example, the 19.8% and 16.6% from true 1pXn→R110 and true 1pXn→R110 are combined to the* 36.4% *in 1pXn→1p1n. Mismatches are due to rounding errors.*

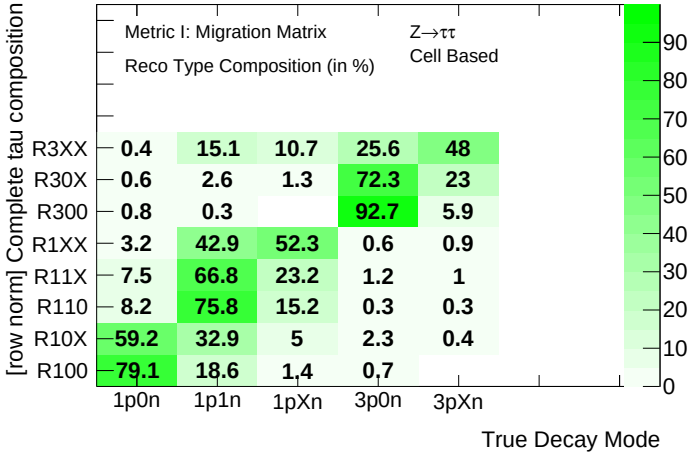


FIGURE 5.25: *Compositions of the complete tau compositions. The admixture of 1pXn and 3p taus in the R10X (R110) composition is* 5% *and* 2.7% *(*15.2% *and* 0.6%*). In R11X (R1XX) the 1p0n and 3p admixtures are* 7.5% *and* 2.2% *(*3.2% *and* 1.5%*). The admixtures of 1p in R30X and R3XX are* 4.5% *and* 26.2%*. The true modes tested for are the dominant, although there are sizable contributions from 1p1n, 1pXn in R3XX and from 1pXn in R11X to some extend.*

To understand how the reclassification in PanTau will work, it is important to note that the rows R100 and R300 are not changed, because those are $\tau_{\text{had-vis}}$, which do not have a neutral PFO. Hence, the mode cannot be changed. This happens for 37.1% of true 1p0n taus, 3.6% of true 1p1n taus, and 6.2% of true 3p0n taus, the other fractions are below 1% and can be neglected.

Every other remaining complete tau composition on the y-axis of Figure 5.24a does enter exactly one of the PanTau decay mode tests. In each of the tests, one complete tau composition is tested against two true modes, for instance the R10X composition is tested against true 1p0n and true 1p1n.

As can be seen from Figure 5.25, there are also taus with true decay modes 1pXn, 3p0n and 3pXn classified as R10X, so that they enter the 1p0n-vs-1p1n test. For instance, there is a contribution of 5% from true 1pXn taus in the R10X composition (which will then be classified as 1p0n or 1p1n). It is noteworthy that because of this, the reclassification affects all true decay modes. For example, such a true 1pXn, reconstructed as R10X (i.e. 1p0n) by CellBased, will enter the 1p0n-vs-1p1n decay mode test and it is possible that it is reclassified by PanTau to be 1p1n. As a result, the far off-diagonal entries can differ when comparing the CellBased and PanTau migration matrices - even though there is no 1p0n-vs-1pXn decay mode test. However, these are not the dominant cases and are neglected in the further discussion[15].

Moving to the dominant true mode contributions in a given complete tau composition, the way in which they are reclassified can be taken from the PanTau decay mode tests. The numbers are given in Table 5.12, which summarises the Tables 5.9, 5.10 and 5.11. It lists all complete tau compositions, the associated reclassification that can happen, the number of correctly and wrongly classified taus by CellBased and the fraction of those two contributions that are reclassified. Here, "correctly reclassified" refers to the fraction of wrongly classified taus by CellBased, that are reclassified to the correct mode by PanTau. "wrongly reclassified" denotes the fraction of correctly classified taus by CellBased, that are reclassified to the wrong mode by PanTau.

So far, there have not been detailed studies looking into the systematic error of the classification in CellBased or PanTau. In the following, a systematic uncertainty of 5% is used [51][16]. This uncertainty is implemented such, that it is assumed that the number of correct and wrong $\tau_{\text{had-vis}}$ in the different complete tau compositions as determined by CellBased vary by 5%. It is assumed further that also the actual performance of PanTau on those taus is subject to an uncertainty, assumed to be 10%. This uncertainty is applied to the fraction of correctly and wrongly reclassified taus, including a correlation of -100%, because every tau that is not reclassified correctly is reclassified wrongly.

The statistical error on the correctly and wrongly reconstructed taus by CellBased is less than 0.5%, when a Gaussian error is assumed[17]. The statistical uncertainties on the reclassification fractions are also much smaller (factor $\approx$ 40 and more) than the estimated 10% systematic uncertainty[18]. Thus, the contribution from the estimated systematic uncertainty is dominant and the statistical errors are neglected in the following.

[15] The admixture of undesired true modes in a decay mode test can be circumvented by following another approach to the idea of an algorithm like PanTau, which is described briefly in Section 5.5.2

[16] Because there have not been detailed studies yet, this uncertainty is not very well known and rather a guess. On the one hand, studies into agreement between simulation and data indicate that the substructure variables are modelled reasonably well (c.f. [52], also see Fig. 6.32 in Sec. 6.6). On the other hand, the differences found in [52] have not been translated into uncertainties on the actual classification.

[17] The 0.5% statistical uncertainty is obtained for the $\approx$ 40.000 wrongly classified R110 taus by CellBased. The other contributions are all larger (up to a factor of 9 for correct R110 taus, c.f. Tab. 5.12), so that the statistical uncertainty is lower than 0.5% in most cases.

[18] These uncertainties are obtained by applying a Gaussian error to the number of taus reconstructed by CellBased and to the number of taus reclassified by PanTau and then propagating these errors to the reclassification fraction which is calculated as the ratio of these two numbers.

Complete tau composition	Reclassification	Correct $\tau_{\text{had-vis}}$ from `CellBased`	Wrong $\tau_{\text{had-vis}}$ from `CellBased`	Correctly reclassified	Wrongly reclassified
R10X	1p0n→1p1n	171,720	95,603	35.2%	12.1%
R110	1p1n→1p0n	367,587	39,728	29.5%	2.1%
R11X	1p1n→1pXn	256,879	89,511	3.0%	0.6%
R1XX	1pXn→1p1n	205,283	173,978	61.7%	31.8%
R30X	3p0n→3pXn	257,822	81,985	30.3%	3.8%
R3XX	3pXn→3p0n	96,376	51,559	69.7%	21.7%

TABLE 5.12: *Fractions of correctly and wrongly reconstructed taus in the different complete tau compositions as obtained from the `CellBased` algorithm. The numbers are taken from the tables 5.9, 5.10 and 5.11. The last column refers to the fraction of taus in a complete tau composition that initially had the correct mode, but were then reclassified by `PanTau`, i.e. the reclassification is wrong. The next to last column denotes the fraction of taus with a wrong classification, that get reclassified to have the correct mode.*

With these numbers at hand, the actual changes in the migration matrix can be calculated (rather than just observed). This is done to achieve a cross check between what is seen in the reclassifications shown in Figures 5.15, 5.19 & 5.23 and in the migration matrix that shows the results of `CellBased` and `PanTau` combined (Figure 6.3). It is important to understand that the net gains in correctly classified taus (listed in the Tables 5.9, 5.10 and 5.11) are referring to the change in the corresponding complete tau composition. The changes in the migration matrix entries may indeed decrease, although all net gains are positive. In the following, the actual changes of the migration matrix entries are calculated.

In order to perform the calculation, it is very convenient to formulate a notation.

- *Number of taus:*
  The number of taus with a given true mode αpβn (for example 1p0n) and complete tau composition Rijk (for example R10X) is denoted as

  $$N(T\,\alpha\text{p}\beta\text{n}, \text{Rijk})$$

  If instead of the complete tau composition the reconstructed mode γpδn (for instance 1p1n) is to be noted, the expression is

  $$N(T\,\alpha\text{p}\beta\text{n}, \text{R}\gamma\text{p}\delta\text{n})$$

  where the R is kept to explicitly refer to the fact that this is the reconstructed mode. It is assumed that these numbers are subject to a systematic uncertainty of 5%.

- *Reclassification fraction:*
  The reclassification is defined by two parameters. The complete tau composition that is reclassified and the direction of the reclassification, i.e. from a wrong to the correct mode or vice versa. Thus, the fraction of taus reconstructed as Rijk that are reclassified in the wrong (correct) way is written as

  $$r(\text{Rijk}, \text{wrong/correct})$$

  To give an example, r(R10X, correct) would correspond to the 35.2% from the R10X line, next to last column in Table 5.12. Likewise, r(R1XX, wrong) is 31.8% and r(R1XX, correct) is 61.7%. As mentioned above, an uncertainty of 10% is applied to these fractions, with a correlation of −100% for a given complete tau composition.

With this notation, the reclassification by PanTau can be expressed (neglecting taus that have been misreconstructed by CellBased, for instance true 3p0n $\tau_{\text{had-vis}}$ in the R110 composition). The following equations describe the change in true 1p0n taus due to the R10X and R110 decay mode test:

$$\begin{aligned} N(\text{T1p0n}, \text{R10X})_{\texttt{PanTau}} = \; & N(\text{T1p0n}, \text{R10X}) \\ & + N(\text{T1p0n}, \text{R110}) \cdot r(\text{R110}, \text{correct}) \\ & - N(\text{T1p0n}, \text{R10X}) \cdot r(\text{R10X}, \text{wrong}) \\ = \; & N(\text{T1p0n}, \text{R10X}) + \Delta\tau_{\text{1p0n}} \end{aligned} \tag{5.1}$$

Here, the first line is the number of true 1p0n taus that have been classified as R10X by CellBased. The following two lines describe the migration *into* and *out of* the R10X composition. Taus that are migrated *into* R10X stem from R110 taus, and the amount that is migrated is the fraction of correctly migrated R110 taus. Taus migrating *out of* R10X are R10X taus that are reclassified as 1p1n, so their number is the number of available R10X taus times the fraction of wrongly reclassified R10X taus.

The difference in $N(\text{T1p0n}, \text{R10X})$ when going from CellBased to PanTau has been defined with a positive sign as $+\Delta\tau_{\text{1p0n}}$:

$$\begin{aligned} \Delta\tau_{\text{1p0n}} = & + N(\text{T1p0n}, \text{R110}) \cdot r(\text{R110}, \text{correct}) \\ & - N(\text{T1p0n}, \text{R10X}) \cdot r(\text{R10X}, \text{wrong}) \end{aligned} \tag{5.2}$$

The sign is chosen to be + because a change here would increase the overall fraction of correctly classified taus, because $N(\text{T1p0n}, \text{R10X})_{\texttt{PanTau}}$ counts the number of true 1p0n, reconstructed as R10X (i.e. 1p0n).

The difference $\Delta\tau_{\text{1p0n}}$ also appears in the calculation of $N(\text{T1p0n}, \text{R110})_{\texttt{PanTau}}$:

$$\begin{aligned} N(\text{T1p0n}, \text{R110})_{\texttt{PanTau}} = & N(\text{T1p0n}, \text{R10X}) \\ & - N(\text{T1p0n}, \text{R110}) \cdot r(\text{R110}, \text{correct}) \\ & + N(\text{T1p0n}, \text{R10X}) \cdot r(\text{R10X}, \text{wrong}) \\ = \; & N(\text{T1p0n}, \text{R10X}) - \Delta\tau_{\text{1p0n}} \end{aligned} \tag{5.3}$$

The appearance of $-\Delta\tau_{\text{1p0n}}$ here is not surprising, because PanTau does not reject (or create) $\tau_{\text{had-vis}}$ candidates, it just reclassifies them.

So far, only the correct R110 and the wrong R10X reclassification has been discussed. In the 1p0n-vs-1p1n BDT, there also is wrong R110 and correct R10X reclassification. These happen for true 1p1n taus and can be used to formulate similar equations for true 1p1n taus, reconstructed as R10X and R110.

$$\begin{aligned} N(\text{T1p1n}, \text{R10X})_{\texttt{PanTau}} = \; & N(\text{T1p1n}, \text{R10X}) \\ & - N(\text{T1p1n}, \text{R10X}) \cdot r(\text{R10X}, \text{correct}) \\ & + N(\text{T1p1n}, \text{R110}) \cdot r(\text{R110}, \text{wrong}) \\ = \; & N(\text{T1p1n}, \text{R10X}) - \Delta\tau_{\text{1p1n,a}} \end{aligned} \tag{5.4}$$

$$\begin{aligned} N(\text{T1p1n, R110})_{\texttt{PanTau}} &= N(\text{T1p1n, R10X}) \\ &\quad + N(\text{T1p1n, R10X}) \cdot r(\text{R10X, correct}) \\ &\quad - N(\text{T1p1n, R110}) \cdot r(\text{R110, wrong}) \\ &= \quad N(\text{T1p1n, R10X}) + \Delta\tau_{\text{1p1n,a}} \end{aligned} \tag{5.5}$$

Analogously to the case above, the difference is denoted as $\Delta\tau_{\text{1p1n,a}}$:

$$\begin{aligned} \Delta\tau_{\text{1p1n,a}} &= + N(\text{T1p1n, R10X}) \cdot r(\text{R10X, correct}) \\ &\quad - N(\text{T1p1n, R110}) \cdot r(\text{R110, wrong}) \end{aligned} \tag{5.6}$$

The additional index a is necessary to separate this migration from the other migration in true 1p1n taus, which is due to the 1p1n-vs-1pXn BDT.

Concluding the calculation of the migration for the 1p0n-vs-1p1n BDT, the gain of correctly classified taus can be expressed as

$$\begin{aligned} \Delta\tau_{\text{1p0n-vs-1p1n}} &= \Delta\tau_{\text{1p0n}} + \Delta\tau_{\text{1p1n,a}} \\ &= (-9091 \pm 2673) + (25687 \pm 3878) = 16596 \pm 6175 \end{aligned} \tag{5.7}$$

The errors are obtained by Gaussian error propagation, the error on the sum includes the anti-correlation between the reclassification fractions R10X and R110.

For convenience, a graphical illustration of the migrations between complete tau compositions together with the $\Delta\tau$ variables is given in Figure 5.26.

Following this prescription to calculate the differences in the number of classified taus, all $\Delta\tau$ values can be calculated.

$$\Delta\tau_{\text{1p1n-vs-1pXn}} = \Delta\tau_{\text{1p1n,b}} + \Delta\tau_{\text{1pXn}} \tag{5.8}$$

$$\begin{aligned} &= \quad N(\text{T1p1n, R1XX}) \cdot r(\text{R1XX, correct}) \\ &\quad - N(\text{T1p1n, R11X}) \cdot r(\text{R11X, wrong}) \\ &\quad + N(\text{T1pXn, R11X}) \cdot r(\text{R11X, correct}) \\ &\quad - N(\text{T1pXn, R1XX}) \cdot r(\text{R1XX, wrong}) \end{aligned} \tag{5.9}$$

$$= (105606 \pm 12010) - (62625 \pm 7309) = 42981 \pm 7567$$

$$\Delta\tau_{\text{3p0n-vs-3pXn}} = \Delta\tau_{\text{3p0n}} + \Delta\tau_{\text{3pXn}} \tag{5.10}$$

$$\begin{aligned} &= \quad N(\text{T3p0n, R3XX}) \cdot r(\text{R3XX, correct}) \\ &\quad - N(\text{T3p0n, R30X}) \cdot r(\text{R30X, wrong}) \\ &\quad + N(\text{T3pXn, R30X}) \cdot r(\text{R30X, correct}) \\ &\quad - N(\text{T3pXn, R3XX}) \cdot r(\text{R3XX, wrong}) \end{aligned} \tag{5.11}$$

$$= (26092 \pm 4168) + (3928 \pm 3640) = 30020 \pm 3260$$

Thus, the overall absolute gain in correctly classified taus in the sample studied is

$$\Delta\tau_{\text{Total}} = \Delta\tau_{\text{1p0n-vs-1p1n}} + \Delta\tau_{\text{1p1n-vs-1pXn}} + \Delta\tau_{\text{3p0n-vs-3pXn}} = 89597 \pm 10296 \tag{5.12}$$

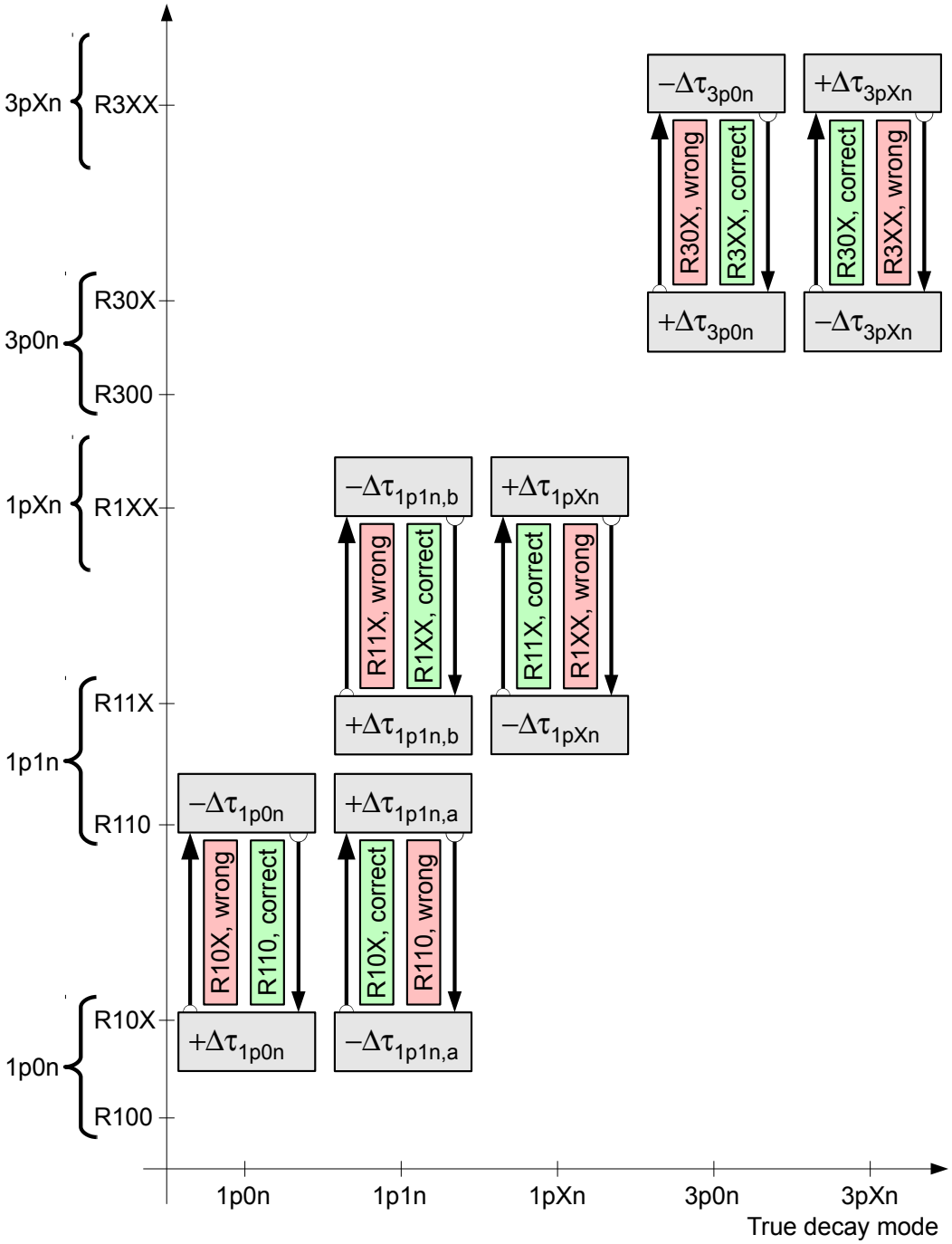


FIGURE 5.26: *Illustration of where the reclassifications happen in the migration matrix. Migrations in the 1p0n-vs-1p1n BDT involving 1pXn, 3p0n or 3pXn true modes are neglected, similar for the 1p1n-vs-1pXn and 3p0n-vs-3pXn BDT.*

Decay mode	Total	1p0n	1p1n	1pXn	3p0n	3pXn
Number of taus	2,391,392	410,590	993,562	448,045	348,308	190,887
Absolute change ($\Delta\tau$)	89,597	-9,091	+131,293	-62,625	+26,092	+3,928
Uncertainty on $\Delta\tau$	±10,296	±2,673	±12,620	±7,309	±4,168	±3,640
Change in efficiency	+3.75%	-2.2%	+13.2%	-14.0%	+7.5%	+2.1%
Uncertainty on efficiency	±0.43%	±0.7%	±1.3%	±1.6%	±1.2%	±1.9%

TABLE 5.13: *Number of taus in the different decay modes, and expected absolute and relative changes including their errors after applying the* `PanTau` *decay mode tests to the* `CellBased` *classification. The losses in 1p0n and 1pXn efficiencies are countered by large gains in the 1p1n classification. The calculation does not take into account the existence of wrong decay modes in the decay mode tests, for instance the existence of true 1pXn taus in the 1p0n-vs-1p1n decay mode test. The errors are obtained assuming a 5% uncertainty on the correct and wrong classification prior to* `PanTau` *by* `CellBased`*, and another 10% uncertainty on the* `PanTau` *reclassification.*

The expected changes in the different diagonal entries of the efficiency matrix are listed in Table 5.13. Note however, that these were obtained by neglecting the migration and possible reclassification of true decay modes that are not considered in the decay mode tests (for instance a true 1p1n, reconstructed as R30X by `CellBased` and then reclassified to 3pXn by `PanTau`).

The overall improvement that is expected is 3.75%. While this seems rather small in terms of a percentage, it should be considered that this should be seen in relation to the improvement that `CellBased` gives with respect to a naive decay mode classification. The naive classification is to classify every $\tau_{\text{1-prong}}$ as a 1p1n and every $\tau_{\text{3-prong}}$ as a 3p0n, because those have the highest branching ratios in the prong categories. As a result, 57.5%[19] of all taus are classified correctly. Thus, `CellBased` increases this by 8.0% and `PanTau` by another 3.75%, so that the gain by `PanTau` is sizable.

In this naive classification, the efficiency is of course 0% for all modes, except 1p1n and 3p0n, for which it is 100%. The purities therefore only exist for 1p1n and 3p0n, and they are 54.7% and 66.8% respectively.

5.4 Decay Product Selection

The decay product selection defines which PFOs are used in the calculation of the 4-momentum of the $\tau_{\text{had-vis}}$. In substructure based tau reconstruction, the 4-momentum of the $\tau_{\text{had-vis}}$ is the sum of the 4-momenta of its constituents, i.e. the $\pi^{\pm}$-PFOs and the π^0-PFOs. For the substructure algorithm, the calculation is rather simple, because its output is the set of $\pi^{\pm}$ and π^0-PFOs. From this set, the decay mode is obtained by counting the objects, and the 4-momentum is obtained by adding the corresponding 4-momenta. For `PanTau`, things are a bit more involved.

After the reclassification of the decay mode, there are actually two decay modes attached to a given tau. The first decay mode is the mode that was obtained by counting the $\pi^{\pm}$-PFOs and the π^0-PFOs in the `CellBased` algorithm. The second decay mode is the mode that is set by `PanTau`, following the decay mode tests described in the previous section 5.3.

Hence, it can happen that the `PanTau` decay mode is inconsistent with the number of objects from `CellBased`, for instance in a 1p0n→1p1n reclassification. Here, the `CellBased` version of the $\tau_{\text{had-vis}}$ contains one $\pi^{\pm}$-PFO, no π^0-PFOs and at least one non-π^0-PFO. While this is expected for a 1p0n mode, is is inconsistent with the 1p1n mode, because here a π^0-PFO should exist. Thus the goal in the decay

[19] Obtained by $\frac{\text{BR}(1p1n)+\text{BR}(3p0n)}{\text{BR(Any hadronic mode)}}$

product selection is to make the PanTau decay mode consistent with the number of PFOs in the $\tau_{\text{had-vis}}$. A combination, or rather comparison, of the two decay modes is the starting point of the decay product selection in PanTau. A schematic overview of the process (not including the exceptions discussed further below) is depicted in Figure 5.27.

There are three possibilities in the comparison of the decay modes.

1. The tau is not reclassified by PanTau, so that the two decay modes are the same. This is always the case for R100 and R300 taus, but can also happen for every other complete tau composition.
2. The tau is reclassified such, that the PanTau decay mode contains *more* neutral pions than the CellBased mode. This happens in the reclassifications 1p0n→1p1n, 1p1n→1pXn and 3p0n→3pXn.
3. Finally, the last possibility is that the PanTau mode contains *less* neutral pions than the CellBased mode. If the reclassification is 1p1n→1p0n, 1pXn→1p1n and 3pXn→3p0n, this is the case.

If the two modes are equal, the calculation of the 4-momentum of the $\tau_{\text{had-vis}}$ is exactly the same as in CellBased, i.e. the 4-momenta of the $\pi^{\pm}$- and π^0-PFOs are added. However, there are two exceptions to this rule.

- The first exception is that if the decay mode is 3pXn, only up to one π^0-PFO is used to calculate the $\tau_{\text{had-vis}}$ 4-momentum. This is because the branching ratio of the mode 3p2n is roughly 0.5%, i.e. roughly 10 times smaller than the 3p1n branching ratio. Thus it is more likely that the second π^0 comes from a misclassification rather than the tau decay.
- The second exception affects $\tau_{\text{had-vis}}$, in which the CellBased decay mode is 1p1n and the number of photons from π^0 candidates is equal to or greater than three. In these cases, the PanTau mode is set to be 1pXn, because the three photons are unlikely to come from only one π^0. Instead, it is assumed that there have been at least two π^0, like in a 1pXn decay, producing at least four photons, three of which have been reconstructed. In these cases, the actual number of π^0-PFOs does not need to be two.

If the two modes differ and the number of neutral pions in the PanTau mode is less that the number of neutrals in the CellBased mode, the calculation is as follows.

- *CellBased 1p1n → PanTau 1p0n:*
  The decay mode test indicates that the π^0-PFO was tagged as such by accident and does not belong to the tau decay. Thus, the π^0-PFO from the 1p1n mode is not considered in the 4-momentum calculation. The 4-momentum of the $\tau_{\text{had-vis}}$ is set to be the 4-momentum of the $\pi^{\pm}$-PFO.
- *CellBased 1pXn → PanTau 1p1n:*
  Based on the decay mode test, the PFOs are more consistent with a 1p1n decay than a 1pXn decay. At the moment, the 4-momenta of the two π^0-PFOs have their mass set to zero and are combined to form the 4-momentum of the one neutral pion in 1p1n. This is done based on the assumption that the two π^0 stem from the two photons from $\pi^0 \to \gamma\gamma$. The two photons have a higher chance of being resolved individually with decreasing $\tau_{\text{had-vis}}$ and π^0 p_T.

  It might be worth to discard the π^0-PFO with the lower π^0-BDT score if the momentum of the tau itself and the momentum carried by the $\pi^{\pm}$ exceed a certain threshold, because then it is unlikely that the two clusters indeed stem from two resolved photons.

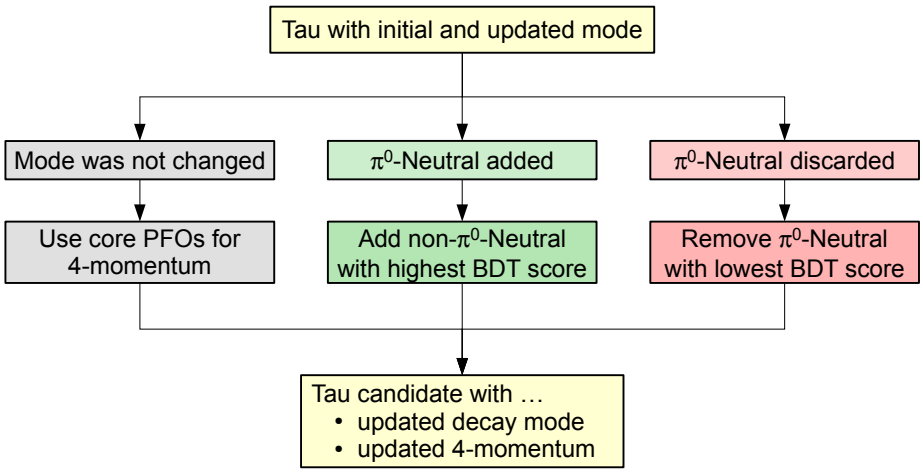


FIGURE 5.27: *Illustration of the decay product selection in* PanTau. *There are exceptions to these basic rules depicted here, which are discussed in the text. The exceptions affect the case in which the mode is 3pXn and is not changed, and the case in which a 1pXn is reclassified to be a 1p1n.*

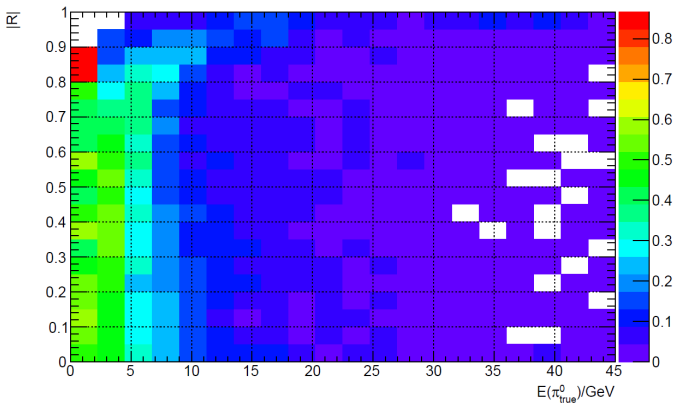


FIGURE 5.28: *This plot shows the fraction of photons in separate clusters (z-Axis) as a function of the neutral pion energy (x-axis) and the ratio of photon energies* $|R| = \frac{E_1 - E_2}{E_1 + E_2}$ *(y-axis). Above* $\approx$ *5 GeV, the fraction appears to be negligible. Thus, above 5 GeV, the merging of the two clusters into one should give worse results in terms of energy resolution, and instead the cluster with the lower* π^0 *BDT score should be neglected. Taken from [53].*

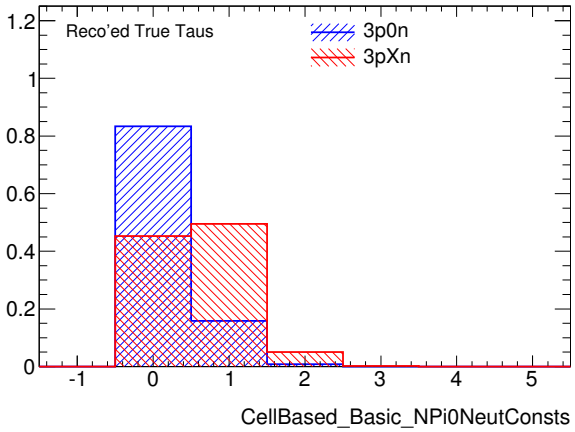


FIGURE 5.29: *Number of* π^0*-PFOs in true 3p0n and 3pXn decays, reconstructed as R30X or R3XX. 95% of the true 3pXn decays (i.e. combined 3p1n, 3p2n, etc.) are reconstructed with less than two* π^0*-PFOs. This is more than expected (90%) for a ratio of branching ratios of* $3p1n : 3p2n = 10 : 1$*, but this may be due to* π^0 *clusters merging (either with each other or with a* $\pi^\pm$ *cluster, being susceptible to imperfect subtraction) in the 3pXn decay.*

In the context of recent studies [54] it was found that this effect is only relevant for π^0 energies below roughly 5 GeV, see Figure 5.28. Thus, above these $\approx$ 5 GeV, the energy and mass resolution might profit from discarding the PFO with the lower BDT score.

- *CellBased 3pXn → PanTau 3p0n:*
  The set of PFOs in the tau was reclassified to 3p0n, so that the π^0-PFOs do not belong to the decay. Like in the 1p1n→1p0n case, the π^0-PFOs are not used in the 4-momentum calculation. The $\tau_{\text{had-vis}}$ 4-momentum is calculated by adding the 4-momenta of the three $\pi^\pm$-PFOs.

The third possibility in the comparison of the CellBased and PanTau decay modes is that they differ, and that the PanTau mode contains more π^0 than the CellBased mode. In this case, a non-π^0-PFO needs to be selected to play the role of the additional π^0 that is needed in the PanTau mode.

- *CellBased 1p0n → PanTau 1p1n:*
  Here, the list of neutral PFOs to choose from contains no π^0-PFO. Hence the non-π^0-PFO with the highest π^0-BDT score is picked and counted towards the $\tau_{\text{had-vis}}$ 4-momentum.

- *CellBased 1p1n → PanTau 1pXn:*
  In this case, there is one π^0-PFO among the neutral PFOs and it therefore has the highest π^0-BDT value. This PFO, together with the non-π^0-PFO having the second highest π^0-BDT value is considered in the $\tau_{\text{had-vis}}$ 4-momentum calculation.

- *CellBased 3p0n → PanTau 3pXn:*
  Although 3pXn refers to $\tau_{\text{3-prong}}$ with at least one neutral pion, PanTau will only add one π^0-PFO to the calculation of the 4-momentum. The reason is twofold. First, the branching ratio of 3p2n is roughly a factor of 10 smaller than the branching ratio of 3p1n, so that a priori, setting the mode to 3p1n is more likely to be correct. Including more than one π^0-PFO is therefore likely to lead to an overestimation of the energy. Second, the 3p0n-vs-3pXn BDT uses no variables that are sensitive to a second π^0-PFO. This is because those variables may contain default values for decays which only contain one π^0-PFO. In the current setup, a new decay mode test would need to be introduced in order to separate 3p1n from 3p2n taus.

  Figure 5.29 shows the number of reconstructed π^0-PFOs in true 3pXn decays, reconstructed as R30X or R3XX. It shows, that only in 5% of all cases, two π^0-PFOs are reconstructed. Thus, including only one π^0-PFO in the momentum calculation is a reasonable approach.

  Calculation of the $\tau_{\text{had-vis}}$ 4-momentum is done by adding all $\pi^\pm$-PFOs and the non-π^0-PFO with the highest π^0-BDT-score.

A more streamlined and consistent approach to decay product selection is possible in the alternative PanTau-like algorithm proposed in Section 5.5.2.

5.5 Summary & Possible Improvements

5.5.1 Summary

A short visual summary of the PanTau algorithm is shown in Figure 5.30, alongside its description.

The improvement in decay mode classification to be expected by PanTau is 3.75% in total. The expected losses of −2.2% and −14.0% in the 1p0n and 1pXn efficiencies are countered by an expected gain of 13.2% in the 1p1n mode. Because of the large branching ratio of 1p1n, these changes effectively

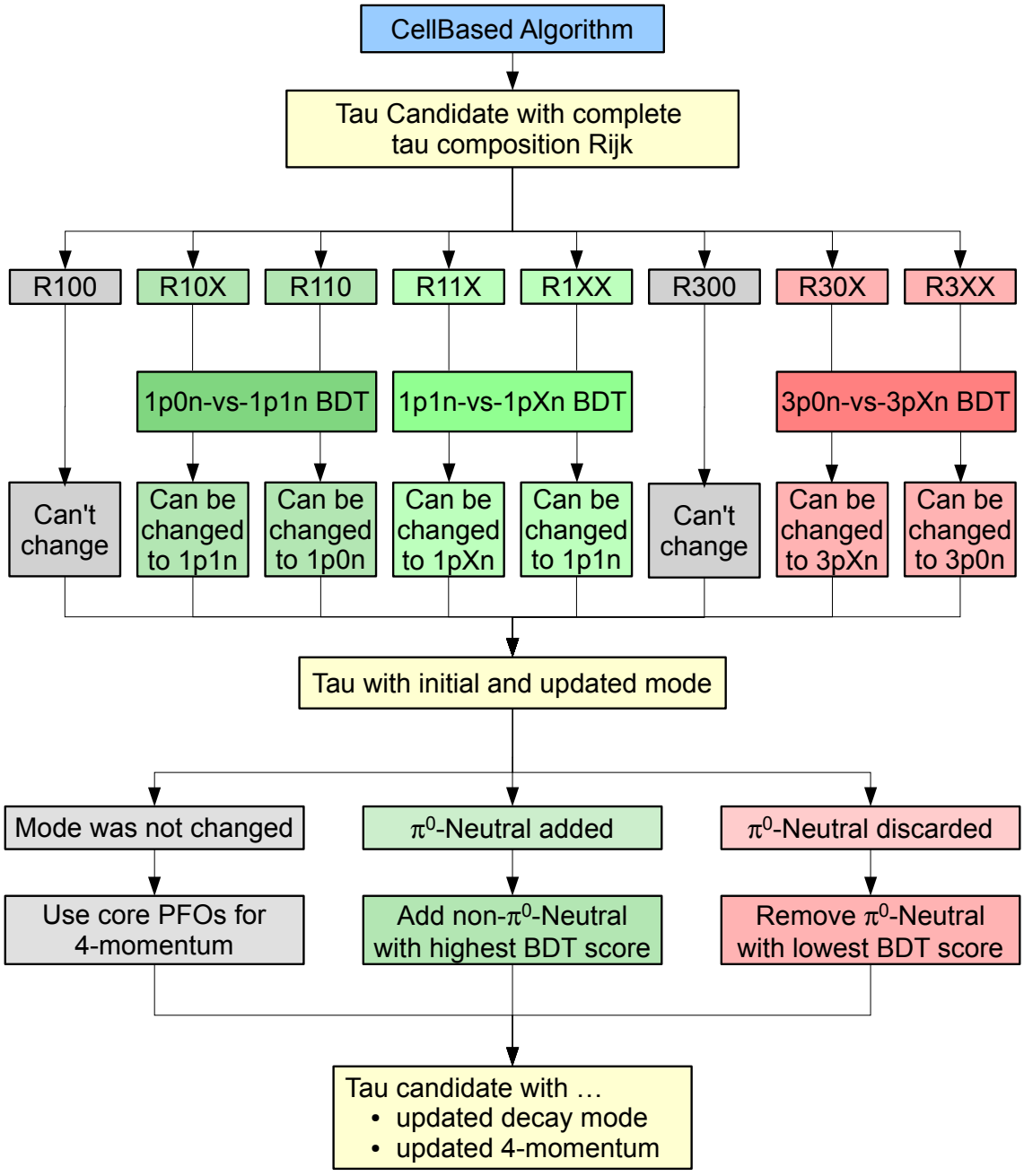


FIGURE 5.30: *Visual illustration of the work-flow in* `PanTau`*. Input are reconstructed taus from the* `CellBased` *algorithm. These taus contain a list of charged, π^0- and non-π^0-PFOs. From this list, the complete tau composition can be derived. The notation is Rijk, where i, j and k denote the number of charged, π^0- and non-π^0-PFOs. If there are no neutral PFOs at all in the tau, no action is taken and the decay mode is kept. Otherwise, the tau enters one of three BDTs that can distinguish between two modes that are likely to be confused. Thus, the misclassification is reduced, for instance by reclassifying a R1XX to a 1p1n.*
Based on the initial decay mode set by `CellBased` *and the mode set by* `PanTau`*, the 4-momentum is updated. If the mode is not changed, the charged and π^0-PFOs in the tau are used for its calculation. If the mode is changed, non-π^0-PFOs are added or π^0-PFOs are removed from the calculation. Note that this is a bit simplified, there are exceptions that are explained in section 5.4.*
Finally, `PanTau` *outputs a modified tau, which contains a refined decay mode and a refined 4-momentum.*

increase the number of correctly classified taus[20]. A large fraction of correctly reconstructed 1p1n taus is of special interest in studies like the one presented in Chapter 7, which rely on a large and pure set of 1p1n taus.

In $\tau_{\text{3-prong}}$, the efficiencies are increased for both 3p0n and 3pXn. In particular, the 3p0n efficiency is expected to be increased by 7.5%.

A detailed discussion of the performance, which includes more than the decay mode classification, is given in Chapter 6.

5.5.2 Possible Improvements

This section lists some improvements that can be done in order to further enhance the performance, or to add utility to the algorithm.

Inclusion of a shrinking-cone

Currently `PanTau` uses a fixed cone of $\Delta R = 0.2$ to classify PFOs as belonging to the tau or not. However, as shown in Figure 5.2, the maximal angle between the tau direction and its decay products decreases with increasing momentum. Thus, a core cone of $\Delta R = 0.2$ is too large for the majority of tau momenta.

Once a so-called shrinking cone algorithm is implemented in the reconstruction software, the `PanTau` algorithm needs to be adapted to it. The shrinking-cone should not increase classification performance, but it should increase the pile-up stability. This is because with a decreasing cone size, the probability to find the same amount of pile-up particles decreases (assuming an isotropic pile-up density in vicinity of the tau candidate).

Different `PanTau` tunes

The thresholds in the decay mode tests have been chosen such, that the number of correctly classified taus is maximised. However, there may be situations, in which it is not of interest to have the overall number of correctly classified taus as large as possible. For instance if an analysis is interested in 1p1n taus only (c.f. to chapter 7), it might be better to increase the efficiency and purity of that specific mode.

Thus, one possibility to enhance the `PanTau` performance would be to tailor the thresholds in the decay mode tests to the application at hand.

BDT Tuning

The BDT configuration has an impact on the performance of the different decay mode tests. An example was given in the discussion of the over-training, where (drastically) changing the BDT configuration lead to over-training. While those exotic configurations have been created for illustration purposes, there most likely are configurations between the nominal one and the over-training configuration, which lead to increased performance without over-training.

For example, in the over-training configuration, the maximal number of nodes and the depth were increased by factors of 20 and ≈ 33, respectively. An increase by a factor of 2 (for instance) may improve

[20] For testing purposes, the BDT thresholds have been adjusted such, that there are no losses in any decay mode. This special setup, which is not the default one discussed in this thesis and is only mentioned here for completeness, also showed an increase in the absolute number of correctly classified taus. However, it was not as large as the one that is obtained when losses are allowed.

the performance without introducing over-training.

Ideally, a study aimed at optimising the BDT configuration would create plots showing the ROC integral as a function of all BDT configuration parameters that may influence the BDT performance. Then, parameters that do not or only slightly increase the performance would be filtered out. The remaining set of parameters should then be used to create an n-dimensional grid, where each point corresponds to a BDT configuration. At each point, a figure of merit is evaluated and the point with the highest figure of merit would correspond to the optimal BDT configuration. The figure of merit may be the ROC integral, but ideally (increasing the extent of the study even further) it should be the net gain in correctly identified taus.

However, this requires large amounts of coding and time. Optimising the parameters individually (instead of simultaneously as described above) is less time-consuming and might give results close to the approach outlined above.

p_T corrections

So far, the BDTs have only been trained on taus originating from $Z \rightarrow \tau\tau$. Because of the Z mass of ≈ 90 GeV (despite a potential boost of the Z-Boson), the p_T spectrum of taus that enter the training is fairly limited, i.e. there are only few taus exceeding 100 GeV.

However, in searches for new and heavier particles decaying into τ-leptons, the taus can have p_T values well above what was used in the training. Without including high-p_T taus into the training, the classification of those high-p_T taus is not guaranteed to work properly.

Pile-up correction

In the complete substructure based tau reconstruction algorithm chain, there is no dedicated pile-up suppression or correction, except for the E_T threshold applied to neutral PFOs and the p_T threshold applied to charged PFOs. As will be shown in Chapter 6, the algorithms are quite pile-up robust, with some slight dependencies over 30 vertices.

A correction for this dependency may be implemented in several ways.

- *Pile-up binning:*
  The training process could be split into several bins of in-time pile-up, by using the number of vertices in the event. This would allow to individually set the BDT thresholds in the decay mode tests, depending on the pile-up situation.

  Because of the splitting into different pile-up bins, the available statistics in the training process is reduced, so that the BDTs are more likely to be over-trained.

- *Pile-up dependent change of E_T threshold:*
  Another way to reduce the pile-up dependency is to make the E_T threshold for neutral PFOs dependent on the number of vertices in the event. The η-dependency is kept, but the cut is raised by (for instance) 25 MeV per vertex. A nominal threshold of 2.0 GeV would then become 2.5 GeV in an event with 20 pile-up vertices.

  Such a cut would have a good and a bad effect though. On the one hand, it would decrease contamination from pile-up neutrals, but on the other hand, it would lower the efficiency to reconstruct decay modes with neutral pions, in which the neutral pions are of low energy. This means that with an increasing number of neutral pions failing to be reconstructed because of the increased

E_T threshold, there will be more taus which cannot be recovered by PanTau, because particles belonging to the decay are discarded.

Migration across $\pi^{\pm}$ multiplicities

None of the decay mode tests in PanTau will migrate a $\tau_{\text{1-prong}}$ into a $\tau_{\text{3-prong}}$ or vice versa[21]. In addition, taus with two or four charged pions are not considered in the decay mode tests.

Hence, an obvious improvement to PanTau is to implement migration across $\pi^{\pm}$ multiplicities. The main taus that needs to be taken care of are two prong taus.

The main source of these taus are $\tau_{\text{3-prong}}$ in which two tracks merged into one track (because they were close-by), or in which one track failed the quality criteria. Other sources are 1p1n and 1pXn taus, in which at least one of the photons from $\pi^0 \rightarrow \gamma\gamma$ converted and the resulting e^+e^- pairs lead to the reconstruction of two tracks.

Thus, it is required to be able to flag the tracks from the charged pions to come from a photon conversion. Ideally, this flag is a real number with upper and lower bounds, so that it can be reinterpreted as a probability.

The approach in the substructure based tau reconstruction would then be as follows:

Firstly in CellBased, a cut is applied on the conversion probability of each track. This gives the number of charged pions in the decay and is optimised such, that the overall classification performance is optimal[22]. Then, the neutral PFOs are flagged as π^0 or non-π^0, which leads to new complete tau compositions, R200 and R2XX.

In PanTau, the setup of BDTs would need to be changed such, that the $\pi^{\pm}$ multiplicity is fixed to one or three before any other BDT (i.e. the ones described in section 5.3) is evaluated. Fixing the $\pi^{\pm}$ multiplicity can be done using the following approach.

Yet another BDT is trained with two different sets of taus. Set A contains true 1-prong taus, reconstructed with two tracks, and set B contains true 3-prong taus, reconstructed with two tracks. Both sets do not place cuts on the number of neutral PFOs. Only taus with the composition R200 and R2XX will be evaluated by this BDT. One important input variables to the BDT will of course be the conversion probability. Other variables could be the distance between the two tracks and their p_T ratio. It is worth noticing that none of the variables may involve neutral PFOs, because there may be none in the input tau. Based on the BDT score, a tau is carried over to one of the existing 1-prong or 3-prong compositions, i.e. an R212 tau will become an R112 (for instance). Then PanTau can proceed as it is described in this chapter.

Depending on the level of detail that is required, the decision may be to have more than one BDT for this. The two input sets will remain unchanged in the sense that they stem from true 1-prong and true 3-prong taus, reconstructed with two tracks. However, they will have different cuts on the neutral PFOs they may contain.

The first 2-prong BDT will require the taus to have no neutrals at all, i.e. R200 taus. It is then limited to the track based variables mentioned above.

The second BDT only looks at taus with neutrals, i.e. R2XX. This BDT can use the tracking based variables above. It can also use variables that contain up to one neutral object, ideally in conjunction

[21] Except, of course, for true $\tau_{\text{3-prong}}$ that are reconstructed as $\tau_{\text{1-prong}}$- those may enter one of the two 1-prong BDT tests. However, there are no PanTau decay mode tests, that are *designed* to perform a $\tau_{\text{1-prong}}$- $\tau_{\text{3-prong}}$ migration.

[22] This is completely analogous to what is done with the cut on the π^0 BDT.

with properties of the second track (i.e. the track in question). For example the E_T ratio of and distance between the second track and the first neutral PFO.

This approach trusts that the cut based classification in `CellBased` is correct by up to ±1 track[23]. It does not take into account an unlikely (but yet possible) migration from a true 1pXn with two conversion tracks, reconstructed as an R3XX back to an R1XX.

Additional decay mode tests

At the moment, `PanTau` cannot recover true 1pXn taus, that fall into the R10X or R110 composition. `PanTau` also cannot recover true 1p0n taus, that are fall into the R11X or R1XX compositions.

Thus, the idea here would be to revisit the setup and evaluation chain of the 1-prong BDTs. The composition R10X needs to be split into R101 and R10X, where the latter now implicitly means $X \geq 2$. An R101 tau still enters the 1p0n-vs-1p1n decay mode test, but an R10X would need to enter the 1p0n-vs-1p1n *and* the new 1p0n-vs-1pXn test.

Also, both complete tau compositions R11X and R1XX would need to enter a 1p0n-vs-1pXn *and* a 1p1n-vs-1pXn test.

The order of tests could be such, that the tests containing the mode with the higher branching ratios are considered first, so that in the case of doubt, the mode with the highest a priori probability is chosen. I.e. R11X and R1XX would first enter the 1p1n-vs-1pXn test. If the result is 1pXn, 1p0n-vs-1pXn is tested.

However, although the idea is rather simple, the implications on the code are rather large, as this additional BDT introduces the chaining of two BDT decisions, which is not foreseen in the current design. Thus, the recommendation would be to re-code the `PanTau` algorithms and tools based on the new setup.

Alternative approach to a `PanTau`-like algorithm

Reviewing what `PanTau` actually does, it appears that this can be achieved in a probably more consistent and streamlined fashion. The `PanTau`-algorithm looks at neutral objects in the tau decay and associates them with the rest of the tau (c.f. most of the input variables of the BDTs, Table 5.7, for instance `Neutral_Ratio_1stBDTEtOverEtAllConsts`.). Based on these variables, a decision affecting the complete tau is made, which may or may not change the decay mode. As discussed in Section 5.3.8 and to be shown in Chapter 6, this indeed improves the decay mode classification.

However, there is a more consistent approach, which (in parts) captures the idea of `PanTau` better than its current version. This approach is such, that `PanTau` would make a decision *for each* neutral PFO in the tau, whether it should be counted in the decay mode classification or not. Currently, `PanTau` makes decisions for each tau, taking into account its composition. This eliminates the admittedly a bit cumbersome setup of BDTs and classification of taus using their constituents. A description of how the new version, making a decision for each neutral PFO, could work is given in the following.

Based on the `CellBased` algorithm, each neutral PFO is assigned a π^0-BDT score. This BDT score can be used in conjunction with other properties *of that very* neutral PFO, to evaluate a new BDT-score, based on more information. The main difference in this new proposed setup is, that each variable must be calculable for each neutral PFO - and not for every tau, as it is in the current version. Thus the variables

[23] c.f. the current situation with tagging of π^0-PFOs by `CellBased`.

would move from being *kinematic tau variables* to *neutral PFO - tau* variables. A few examples of these *neutral PFO - tau* variables follow.

- *Energy ratio to charged PFOs:* The energy of the neutral PFO in question is related to the energy of all charged PFOs in the tau (or to the leading one).
- *Distance to charged PFO(s):* The angular distance (or distance in the $\eta - \phi$-plane) between the neutral PFO in question and either the charged subsystem of the tau, or one of the charged PFOs.
- *Energy fraction of neutral PFOs:* The energy of the neutral PFO in question is related to the total energy of neutral PFOs in the tau. It is noteworthy, that no variable can explicitly reference a second neutral - because there may be none.
- *Distance to neutral subsystem:* The distance between the neutral PFO (angular or ΔR) and a charged component of the tau.

With this setup, it would be sufficient to have two BDTs (at most), one for $\tau_{\text{1-prong}}$ and one for $\tau_{\text{3-prong}}$. This refined BDT-score should be stored as a property of the neutral PFO. When calculating the `PanTau`-decay mode, neutral PFOs exceeding a given refined BDT-score threshold are counted as π^0-PFOs and determine the decay mode together with the $\pi^{\pm}$-PFOs. Thus, this setup would also include a 1p0n-vs-1pXn test, because *every* neutral PFO is assigned a refined BDT-score and can be included or excluded from the tau.

There is however the drawback, that no global tau variable, like an invariant mass, can be used to classify the decay mode. The invariant mass of neutral PFOs in a tau can not be calculated for a single neutral PFO. Also the number of photons in the tau, which was helpful in all of the three decay mode tests, would no longer be accessible.

In conclusion, it may be worth to split the algorithm in two parts. One part following the approach outlined above, producing a more refined π^0-BDT-score, based on *neutral PFO - tau* variables, and a second part (maybe cut-based), taking into account the global tau variables that can not be handled in this approach.

CHAPTER 6

Performance in $Z \to \tau\tau$ events

In this chapter, the performance of the substructure based tau reconstruction is discussed. The performance is measured in four different metrics in simulated $Z \to \tau\tau$ events. Section 6.1 introduces the four metrics and the object selection. In the following Sections 6.2, 6.3, 6.4 and 6.5, the performance in the individual metrics is discussed in detail.

6.1 Introduction

During the development of the substructure based $\tau_{\text{had-vis}}$ reconstruction, four metrics have been created in order to measure, judge and compare the performance of the new reconstruction methods. Before discussing the results, these four metrics will be explained. The metrics were first documented in [55] but have evolved since then. After describing the metrics, the object selection is presented. The actual performance results will be discussed in Sections 6.2, 6.4 and 6.5.

6.1.1 Performance Metrics

The performance metrics are as follows.

- *Metric I: Decay Mode Classification*
  This metric aims at quantifying how well the decay mode classification works. The core tool for this are the migration matrices, which have been described in Section 5.3.1. It has been decided to use the fraction of correctly classified $\tau_{\text{had-vis}}$ as the figure of merit.

  This metric allows to compare the overall classification performance. It also shows the efficiencies with which the decay modes are reconstructed and the purity of the reconstructed modes. It can be considered the most important metric, because imperfections in the decay mode classification directly feed into the momentum reconstruction (c.f. Section 6.5).

- *Metric II: Single Neutral Pion Reconstruction*
  The second metric is only evaluated in true and reconstructed 1p1n decays. In these decays, it evaluates the quality of the π^0 reconstruction. The resolution of the π^0 4-momentum (energy and spatial resolution) is the basis for this metric.

  Although it is evaluated in true and reconstructed 1p1n decays only, the classification performance is relevant. This is because all correctly reconstructed 1p1n decays are used. Thus, if an algorithm only reconstructs *easily* classifiable 1p1n decays, it may show a better π^0 resolution than another algorithm, which also reconstructed 1p1n decays in which the $\pi^\pm$ and π^0 overlap (for instance).

- *Metric III: Resonance Reconstruction*
  As outlined in Section 2.2, the 1p1n, 1pXn and 3p0n τ_{had} decays happen via intermediate resonances. These are the ρ and a_1 mesons in 1p1n and 1pXn & 3p0n decays respectively. This metric investigates the capability of reconstructing the invariant mass of the mesons in those cases, in which the decay mode was classified correctly.

- *Metric IV: Global Tau Resolution*
  The last metric scrutinises the resolution of the $\tau_{\text{had-vis}}$ momentum. The resolutions are studied for each decay mode exclusively, as well as inclusively for comparison with the default method for tau reconstruction.

For all metrics, the performance is also studied in dependence of p_T (τ_{had}) (π^0 p_T in case of Metric II and resonance p_T in case of Metric III), pile-up and detector region.

In order to evaluate these metrics, a dedicated software package was developed[1]. The complete evaluation of all metrics covers roughly 500 pages and is available via [56]. The most important results will be discussed in the following sections.

[1] C.f. notes on software in Section A.1.

6.1.2 Object Selection

When evaluating the metrics, a loose object selection is applied. The different selection cuts are listed in Table 6.1.

- *True τ match:*
  All reconstructed taus that are considered in the performance evaluation need to be matched to a true $\tau_{\text{had-vis}}$. This is done in order to be able to compare the reconstructed $\tau_{\text{had-vis}}$ properties to the generated ones, i.e. the true decay mode, the true energy etc. A reconstructed $\tau_{\text{had-vis}}$ is matched to a true $\tau_{\text{had-vis}}$, if their distance in ΔR is less than 0.2.

- *Trigger:*
  The trigger cut is only applied in dedicated studies to measure the impact of the preselection that happens at trigger level. In the plots shown here, it is not applied.

- *Transverse Energy:*
  A cut on the transverse energy requiring $E_T > 15$ GeV.

- *In tracker region:*
  The inner detector of ATLAS ends at η values of -2.5 and $+2.5$. Because the new $\tau_{\text{had-vis}}$ reconstruction heavily depends on the tracks, no $\tau_{\text{had-vis}}$ leptons outside the tracker region are studied.

- *Number of prongs:*
  To avoid complications that arise from conversion of photons and track-merging, the metric evaluation is restricted to τ leptons with either one or three tracks[2].

- *Tau Identification:*
  This cut is only applied in studies looking at a bias introduced by requiring the $\tau_{\text{had-vis}}$ to have passed identification criteria. The criteria are documented in Section 4.2. In the plots shown in this thesis, this cut is not applied.

Based on this object selection (also summarised in Table 6.1), the cut-flow depicted in Figure 6.1 is obtained. The samples contain simulated data of $Z \rightarrow \tau\tau$ events, the dataset names are listed in the appendix A.1. The simulated data samples contain 6.11 million $\tau_{\text{had-vis}}$ in total, out of which 2.4 million pass the default selection without a trigger or identification requirement. Table 6.2 lists the number of taus in the 5 decay modes considered.

As mentioned above, all metrics are also evaluated as a function of a corresponding p_T value, the number of vertices in the event and detector region. While the taus are distributed roughly isotropically across the detector, the distributions in p_T and the number of vertices have a bulk and tails. These are shown in Figure 6.2.

The $p_{T,\text{vis}}$ distribution peaks around 20-30 GeV and has tails up to 100 GeV. The number of vertices has a reasonable amount of entries up to 25 vertices, after which the statistics drop very fast, so that there are less than 20 taus from events with more than 35 vertices. It is worth noticing that for some performance plots, taus from each of the p_T and number of vertices bins are further subdivided into the different decay modes. Thus, results in these regions (if shown in the plots[3]) should be taken with caution, due to the limited statistics and the resulting fluctuations.

[2] A possible improvement to take conversions into account in `PanTau` is given in Section 5.5.2.

[3] The performance as a function of the number of primary vertices is truncated at 30 vertices. At this point, the bins contain less then 1000 taus in total. Including the additional 10 bins affects the axis scaling so that readability is reduced, without adding more information to the plot because of the small statistics.

Index	Selection name	Selection cut	Taus	Comment
(0)	All true taus	None	$6.11 \cdot 10^6$	
(1)	True τ match		$3.76 \cdot 10^6$	
(2)	Trigger	EF_tau20Ti_medium1	$3.76 \cdot 10^6$	Not applied
(3)	Transverse Energy	$E_T > 15$ GeV	$3.07 \cdot 10^6$	
(4)	In tracker region	$\lvert\eta\rvert < 2.5$	$3.07 \cdot 10^6$	
(5)	Number of prongs	$n_{\text{Tracks}} \in 1; 3$	$2.39 \cdot 10^6$	
(6)	Tau Identification			
(6a)	Electron Veto	Pass Medium Veto	$2.39 \cdot 10^6$	Not applied
(6b)	Jet Discrimination	Pass Medium Identification	$2.39 \cdot 10^6$	Not applied

TABLE 6.1: *This table lists the object selection cuts that are applied to select $\tau_{\text{had-vis}}$ leptons that are to be evaluated in the four metrics. Cuts (2), (6a) and (6b) are not applied in this discussion of the results. They have been included and switched on in dedicated studies in order to see how trigger and $\tau_{\text{had-vis}}$ identification bias the performance of substructure reconstruction.*

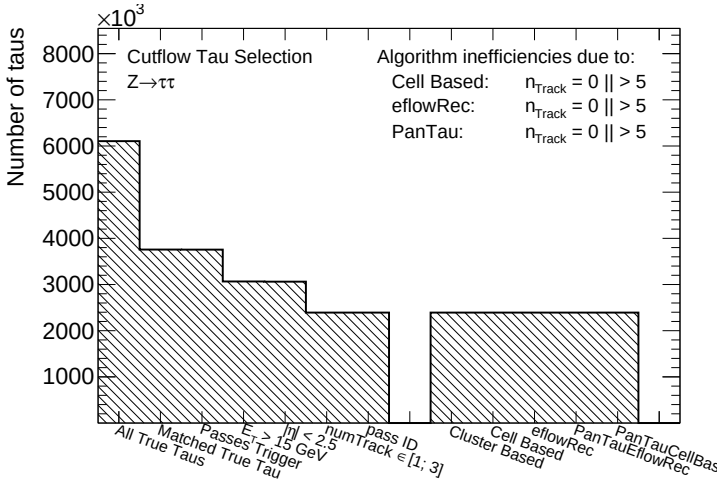


FIGURE 6.1: *Cut-flow in the $Z \to \tau\tau$ sample. Out of the 6.2 million simulated $\tau_{\text{had-vis}}$, approximately 2.4 million $\tau_{\text{had-vis}}$ are considered in the performance evaluation. These consist of more than $410 \cdot 10^3$ 1p0n and 1pXn, $990 \cdot 10^3$ 1p1n, $340 \cdot 10^3$ 3p0n and $340 \cdot 10^3$ 3pXn taus.*

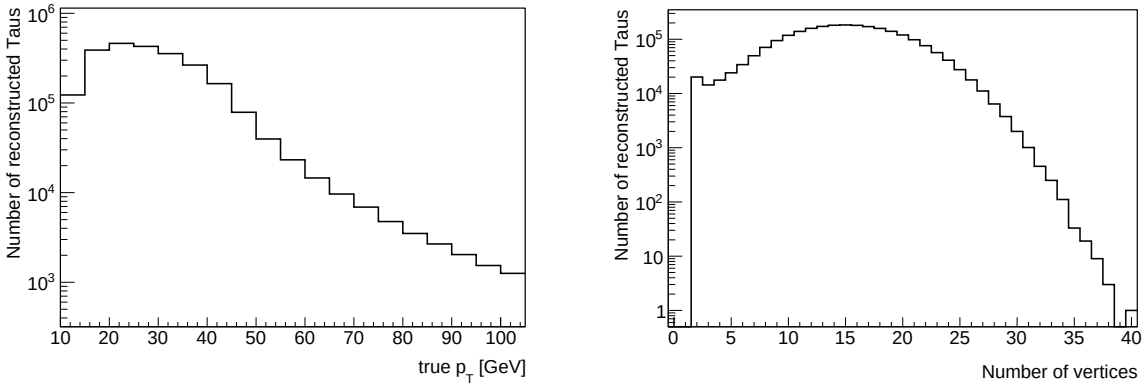


FIGURE 6.2: *Distribution of tau p_T (left) and the number of vertices in the event (right). Because the taus are coming from (potentially boosted) $Z \to \tau\tau$, their visible transverse momentum is most likely below $\approx 40-50$ GeV. The pile-up that was simulated does not exceed 40 vertices. The majority of events has less than $\approx 25-30$ vertices. Above this, statistics are very limited and the results obtained here become unreliable because of this.*

Total	1p0n	1p1n	1pXn	3p0n	3pXn
2,391,392	410,590	993,562	448,045	348,308	190,887

TABLE 6.2: *Total number of reconstructed true $\tau_{\text{had-vis}}$ of considered decay modes in the performance evaluation samples.*

6.2 Metric I - Decay Mode Classification

This section will discuss the migration matrices, both the efficiency and purity matrix and the dependencies of the reconstruction efficiencies on the true $\tau_{\text{had-vis}}$ p_T, detector region and pile-up.

Figure 6.3 shows the two migration matrices for `CellBased` (Fig. 6.3a and 6.3b) and `PanTau` in conjunction with `CellBased` (Fig. 6.3c and 6.3d). The entries of the matrices are given in percent and the statistical error on the diagonal ranges from 0.1% (1p1n) to 0.2% (3pXn). The relative statistical error in the off-diagonal elements is below 3.5%. Within the estimated errors, the observed differences meet the expectations formulated in Table 5.13.

When comparing the efficiency matrices 6.3a and 6.3c, the main difference in the 1-prong sub-matrix happens around the 1p1n efficiency.

The efficiency to correctly reconstruct a 1p1n decay improves by 12.5% (or 19.9%, relatively). This is due to the correct R10X and R1XX reclassifications (c.f. Table 5.12). These reclassifications are responsible for the reduction of 10.2% in the 1p1n→1pXn migration and 2.4% in the 1p1n→1p0n migration[4]. Due to the large 1p1n branching ratio, the gain here outweighs the loss in 1p0n and 1pXn efficiency by far, as can also be seen in the 1-prong figure of merit, which improves by 3.2%.

The loss in 1p0n→1p0n efficiency comes from falsely migrated R10X ($\approx$ 12.1% of R10X), which cannot be countered by correctly reclassified R110 ($\approx$ 29.5% of R110), because only few true 1p0n are reconstructed as R110[5]. These falsely migrated R10X also in turn increase the 1p0n→1p1n efficiency by 3.3%.

The 1pXn→1pXn efficiency drops significantly, too. This is due to the fact that only 3% of R11X are reclassified as 1pXn, while $\approx$ 31.8% of R1XX are migrated to 1p1n, enriching the 1pXn→1p1n migration.

In the 3-prong sub-matrix, both diagonal entries, the 3p0n and 3pXn efficiencies are increased. The 3p0n efficiency increases by 7.5% and the 3pXn efficiency by 1.9%. Overall, this improves the figure of merit in the 3-prong sub-matrix by 5.5%

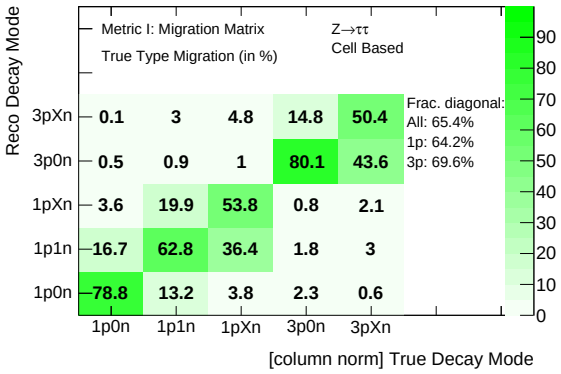
Differences in the purity matrices, Figure 6.3b and 6.3d, can be deduced by analysing the changes in the efficiency matrix.

The 1p0n purity increases, because a reasonable fraction of the true 1p1n polluting the reconstructed 1p0n taus has been migrated to the 1p1n mode, while the 1p0n efficiency stayed roughly constant.

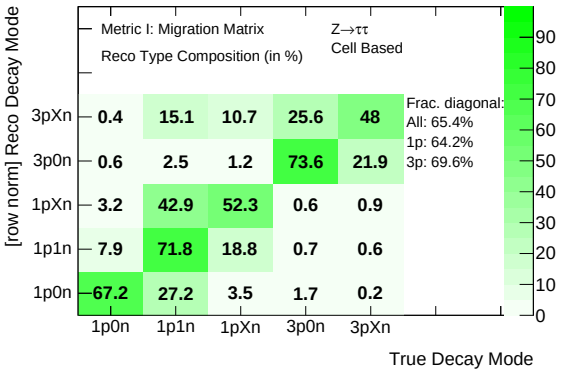
For the 1pXn purity, a very similar thing happens. Although the 1pXn efficiency is reduced significantly (17%), the main contribution to the composition of reconstructed 1pXn, namely 1p1n, has been reduced by 8.7%. Because of the very different branching ratios of 1p1n and 1pXn, this results in an increase of 8.5% in purity.

In case of 1p1n decays, the purity takes a slight hit. This is because not only the 1p1n→1p1n efficiency increased, the 1p0n→1p1n and 1pXn→1p1n efficiency increased as well. Thus, the purity is reduced by 2.1%, but drops only slightly below 70%, as a result of this overall increased reconstruction

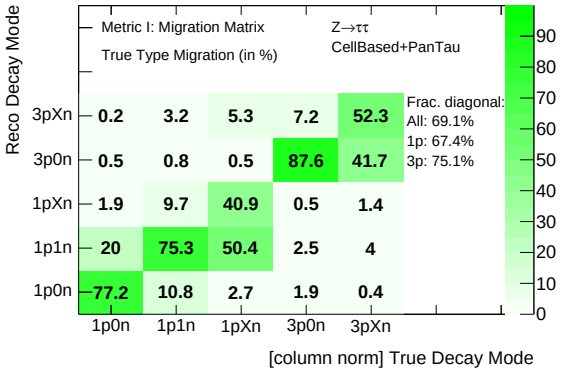
[4] The numbers 10.2% and 2.4% do not add up exactly to 12.5% because the numbers in the migration matrix are rounded.
[5] C.f. Table 5.12: The ratio of correctly classified to wrongly classified taus from `CellBased` in R110 is 9.25 : 1.



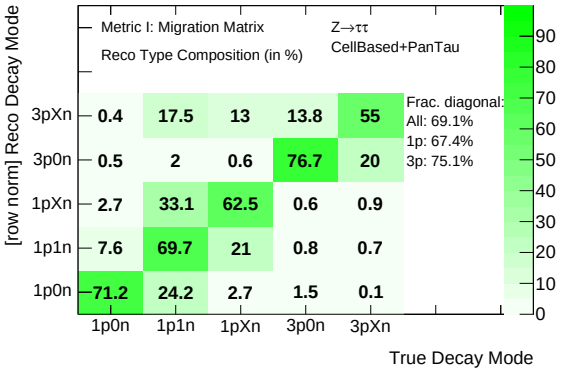
(A) *Efficiency matrix for* `CellBased`.



(B) *Purity matrix for* `CellBased`

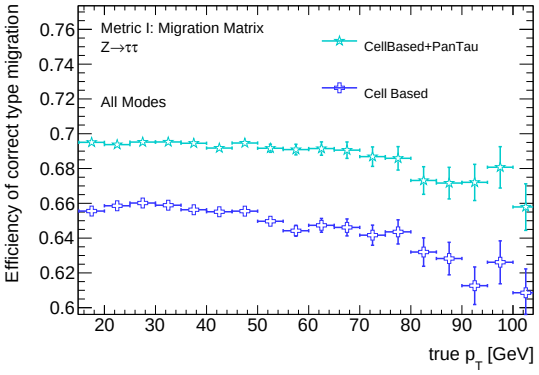


(C) *Efficiency matrix for* `PanTau`.

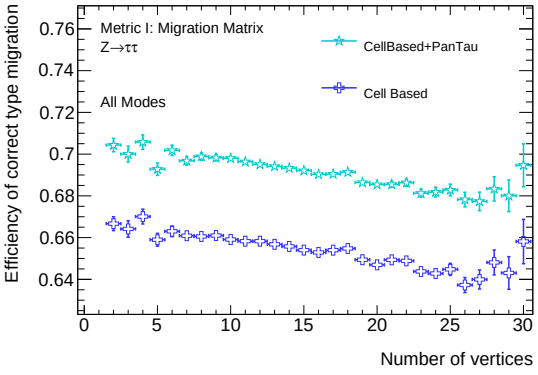


(D) *Purity matrix for* `PanTau`.

FIGURE 6.3: *Efficiency (left) and purity (right) matrices for the* `CellBased` *algorithm (top) and the combination of* `CellBased` *and* `PanTau` *(bottom). Major benefit from* `PanTau` *is the increased 1p1n efficiency,* $\approx 12.5\%$*, and the increased 3p0n efficiency (7.5%). More details are discussed in the text.*



(A) *Overall efficiency to reconstruct the correct decay mode of a tau as a function of true* $\tau_{\text{had-vis}}$ p_T.



(B) *Overall classification efficiency as a function of reconstructed vertices.*

FIGURE 6.4: *Overall efficiencies of the decay mode reconstruction as a function of true* $\tau_{\text{had-vis}}$ p_T *and the number of reconstructed vertices. All decay modes combined, note the zero suppression on the y-axis. The losses are roughly 2.5% in the range from 15 GeV to 100 GeV and* $\approx 2\%$ *over 30 vertices, the same is true for the purity*

of the 1p1n mode.

The overall efficiency to reconstruct the correct decay mode of a tau is 69.1%. Figure 6.4a shows how this fraction depends on the transverse momentum of the true simulated $\tau_{\text{had-vis}}$.

The figure shows that the efficiency drops by about 2.5% with the true $\tau_{\text{had-vis}}$ p_T increasing from 15 to 100 GeV. As can be seen from Figure 6.5, the loss comes mainly from 1p0n decays (in Figure 6.5a). There, the drop is around 7%, from 80% at 20 GeV down to $\approx$ 73% at 90 GeV. The efficiency drops with increasing p_T, because if the subtraction of a high energetic charged pion shower in the EM-Calorimeter is not done perfectly, the remnants will easily pass the E_T threshold for neutral clusters, and might be misidentified as a π^0.

The 1p1n efficiency shows a drop from 78% at 20 GeV to 73% at which it stays flat within ±1% (c.f. Figure 6.5b). The 1pXn efficiency increases up to 60% at 40 GeV and then drops to roughly 50% (Figure 6.5c). This is likely due to the fact that with small transverse momenta, one of the π^0 does not pass the E_T threshold in `CellBased`. This can happen, because in a 1pXn decay, the $\tau_{\text{had-vis}}$ energy has to be distributed among the three pions. In addition, with all particles being rather soft, subtraction of the charged pion is less precise (c.f. Figure 4.8), so a very soft π^0 might be lost due to over-subtraction. At high transverse momenta, the efficiency drops because the decay products are boosted and the distance between them becomes smaller (c.g. Figure 5.2). This increases the chance of the two clusters overlapping with each other.

From Figure 6.5d it can be seen that the 3p0n efficiency is flat at $\approx$ 88%. In this mode, the `PanTau` dependence on the $\tau_{\text{had-vis}}$ p_T is very different than the dependence of `CellBased`. This is most likely due to the technical reason outlined in Section 5.3.1. Because there is no subtraction performed, the energy depositions of charged pions in the EM-Calorimeter are not removed by `CellBased`. As can be seen from Figure 4.9, increasing $\pi^{\pm}$ p_T increases the fraction of energy that is deposited in the EM-Calorimeters. Thus, the 3p0n efficiency drops with increasing $\tau_{\text{had-vis}}$ p_T, because the increasing amounts of $\pi^{\pm}$ energy in the EM-Calorimeter is not removed, but considered as π^0 candidates.

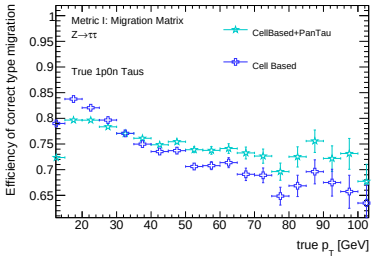
The opposite happens in the 3pXn efficiency, shown in Figure 6.5e. Here, because of the technical reason and its implications outlined above, the 3pXn efficiency in `CellBased` increases with $\tau_{\text{had-vis}}$ p_T. Although the `PanTau` efficiency is well below the `CellBased` efficiency for most of the p_T values, the overall 3pXn efficiencies are about the same, because most taus from $Z \rightarrow \tau\tau$ have p_T values around 20 – 25 GeV (c.f. Figure 6.2).

It is important to understand the behaviour of the decay mode reconstruction in different pile-up scenarios. Figure 6.4b shows the inclusive efficiency and purity of the decay mode reconstruction as a function of reconstructed vertices in the event in which the tau was reconstructed.

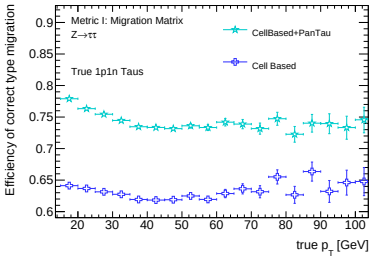
They are rather stable, with a decrease of 2% over 30 vertices. Figure 6.6 shows the classification efficiency of every single mode as a function of the number of vertices. Thus it can be deduced that the loss is mainly driven by 1p0n, 1p1n (small efficiency drop, but large branching ratio) and 3p0n. Ideas on how to counter this dependency are given in Section 5.5.2.

For 1p0n taus (see Figure 6.6a), there is a clear dependency. The efficiency to correctly reconstruct this mode drops from above 80% down to around 70% over 30 vertices. Reason for this is that with increasing pile-up collisions, the probability that a pile-up particle is picked up as a π^0 increases. The individual probability for this to happen may be small, but if the number of trials is large enough, the effect is visible.

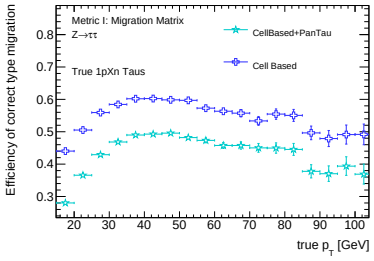
The dependency is much less pronounced in 1p1n decays (Fig. 6.6b), which are practically stable within 2% around an efficiency of 75%. A reason for the more stable behaviour may be that in these decays, the pile-up particles have a chance to overlap in parts or completely with the π^0 of the decay.



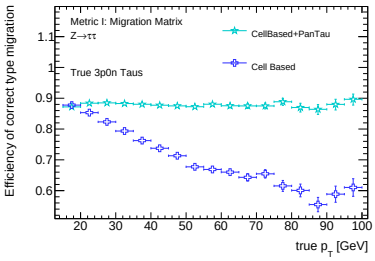
(A) *1p0n efficiency versus $\tau_{\text{had-vis}}$ p_T*



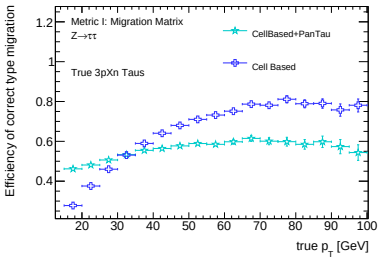
(B) *1p1n efficiency versus $\tau_{\text{had-vis}}$ p_T*



(C) *1pXn efficiency versus $\tau_{\text{had-vis}}$ p_T*

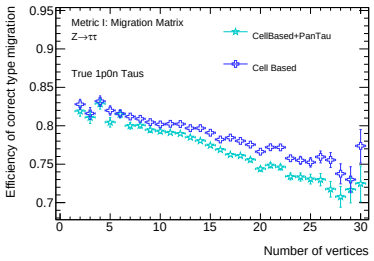


(D) *3p0n efficiency versus $\tau_{\text{had-vis}}$ p_T*

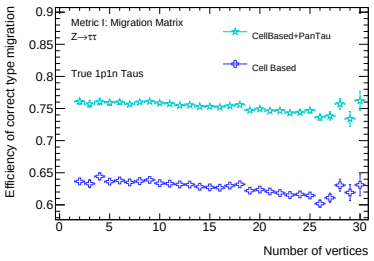


(E) *3pXn efficiency versus $\tau_{\text{had-vis}}$ p_T*

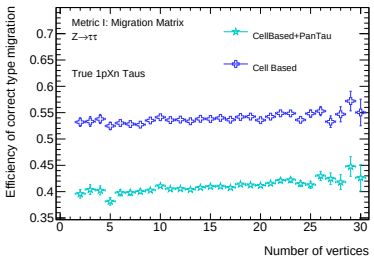
FIGURE 6.5: *Single decay mode efficiencies as a function of true $\tau_{\text{had-vis}}$ p_T. More details are given in the text. The 1p0n efficiency drops because small fluctuations in the $\pi^{\pm}$ subtraction can lead to fake π^0 more easily for higher $\pi^{\pm}$ p_T. The 1pXn efficiency rises and falls because of π^0 with very small p_T values and merging of π^0. In case of 3-prongs, it is mentioned that this version of* `CellBased` *did not perform the subtraction for 3-prongs, hence π^0 are created more easily.*



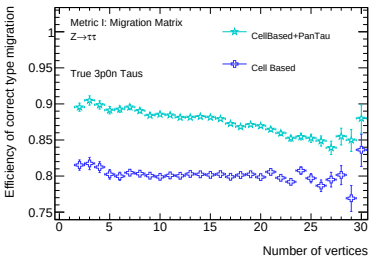
(A) *1p0n efficiency versus N_{Vertex}*



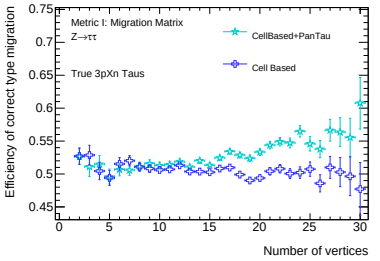
(B) *1p1n efficiency versus N_{Vertex}*



(C) *1pXn efficiency versus N_{Vertex}*



(D) *3p0n efficiency versus N_{Vertex}*



(E) *3pXn efficiency versus N_{Vertex}*

FIGURE 6.6: *Single decay mode efficiencies as a function of the number of primary vertices. The efficiency loss in 1p0n and 3p0n is due to additional neutral particles faking π^0. Efficiencies of 1pXn and 3pXn show a slight increase with number of vertices, because additional pile-up particles increase the neutral energy fraction, increasing 1pXn and 3pXn likeliness. More details are given in the text.*

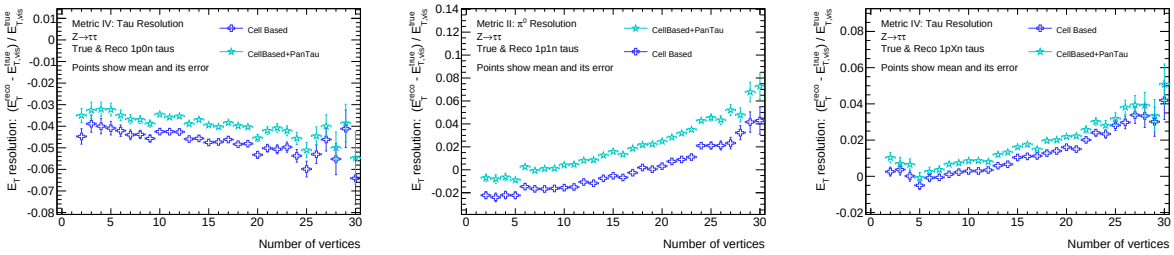
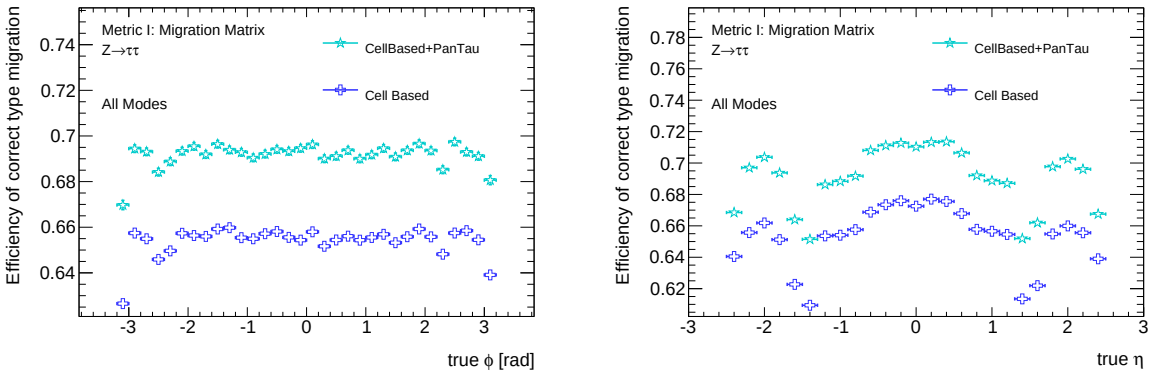


FIGURE 6.7: *Mean and its error of the E_T resolution as a function of vertices in the event for correctly classified taus:* Left: *For 1p0n taus, there is a slight underestimation of the $\tau_{had\text{-}vis}$ energy with increasing pile-up events.* Middle: *In 1p1n taus, the $E_T(\pi^0)$ increases by 7% over 30 vertices.* Right: *The $\tau_{had\text{-}vis}$ energy increases by 3 – 4% over 30 vertices.*
The charged component is not affected by pile-up in the way that the neutral component of a tau is affected. Pile-up increases the energy of the reconstructed neutral pions. Thus, pile-up in the cases of 1p1n and 1pXn taus does not necessarily lead to misclassification, but distorts the energy measurement so that the energy is overestimated.



(A) *Efficiency of correct decay mode classification as a function of ϕ.*

(B) *Efficiency of correct decay mode classification as a function of η.*

FIGURE 6.8: *Efficiency and purity of correctly classified tau leptons as a function of ϕ and η. They are flat as a function of ϕ within 1%. They do depend on η, in the transition region between the barrel and the end-cap part of the calorimeter, see text.*

Thus, no extra particle is reconstructed, but the energy of the already (and correctly) reconstructed π^0 is increased, leading to an overestimation of the $\tau_{\text{had-vis}}$ energy (c.f. Figure 6.7).

A combination of these two effects can be responsible for the slightly rising ($\approx 2\%$) efficiency of the 1pXn mode in Figure 6.6c. Inclusion of a pile-up particle as a π^0 in a prior to that reconstructed 1p1n may recover a true 1pXn. Either the missing second π^0 is replaced by the pile-up particle, or the pile-up particle overlaps with the true π^0, which would not pass the E_T threshold. In addition, each pile-up particle reconstructed in this tau increases the neutral energy fraction (c.f. Figure 5.16b), making the tau more likely to be (re-)classified as 1pXn. Either way, while this helps the raw efficiency, it does not help $\tau_{\text{had-vis}}$ reconstruction, because the energy is overestimated (see Figure 6.7).

For 3p0n taus (Figure 6.6d), the same effect as in 1p0n taus is responsible for the drop in classification efficiency from 90% to 85%.

Finally, there is again a slight increase in the 3pXn efficiency, see Figure 6.6e. Like for the 1pXn case, every neutral PFO associated to the tau, independent of whether it belongs to the tau decay or other collisions, increases the neutral energy fraction, making a reclassification as 3pXn more likely (see Figure 5.20b).

For completeness, Figure 6.8 shows the dependency of the overall efficiencies and purities as a function of the azimuthal angle ϕ and the pseudorapidity η of the $\tau_{\text{had-vis}}$. The efficiency and purity versus the azimuthal angle is flat within 1%. Calculating the efficiency in bins of the pseudorapidity, it shows dips around $|\eta| = 1.4$ and a general decrease in efficiency at $|\eta| > 1.5$ This is the transition region between the barrel and the end-cap region of the calorimeter, in which the π^0 identification provided by `CellBased` seems to have an efficiency drop.

6.3 Metric II - Single Pion Resolution

Although `PanTau` does not reconstruct neutral pions directly, `PanTau` still has an impact on the energy and spatial resolution of neutral pions, because it alters the particle that plays the role of a neutral pion in a tau. This happens in two scenarios.

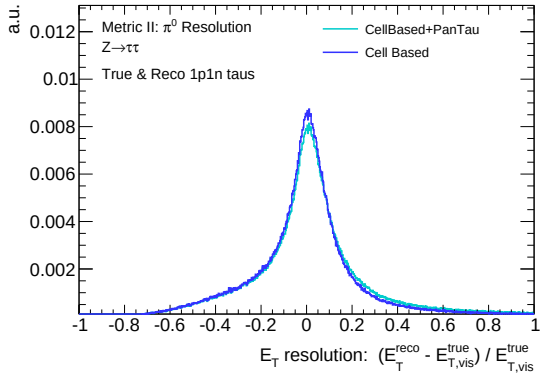
- If a decay is classified as 1p0n by `CellBased` and is reclassified by `PanTau` to be 1p1n, then a non-π^0-PFO takes the place of the π^0 in $\tau \to \nu_\tau \pi^\pm \pi^0$.
- If `CellBased` finds a 1pXn mode, which is reclassified by `PanTau`, then the two π^0-PFOs found by `CellBased` are made massless and are merged into one new PFO[6], which is used as the one π^0 in 1p1n.

Therefore, the single π^0 resolution in `PanTau` is based on more π^0 and on ones that are a bit more difficult to reconstruct, because the mode was classified incorrectly in the first place by `CellBased`.

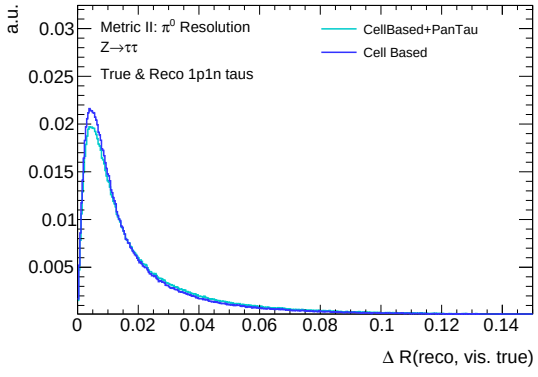
Figure 6.9 shows the E_T and spatial resolution of reconstructed π^0 in correctly classified 1p1n decays. In both cases, `CellBased` performs slightly better than `CellBased` + `PanTau`. However, the actual performance is very similar.

The E_T resolution, centered at 0 for both algorithm combinations, is asymmetric with a high energy tail which leads to an RMS of 0.23 (`CellBased`) and 0.25 (`PanTau`). `PanTau` shows more entries in the high energy tail, which may be correlated to what was found in Figure 5.28. For π^0 energies above 5 GeV, combining the two PFOs from 1pXn into one may not be optimal, because it is less likely that

[6] The assumption is that the two PFOs are the two photons from $\pi^0 \to \gamma\gamma$, c.f. Section 5.4, Figure 5.28

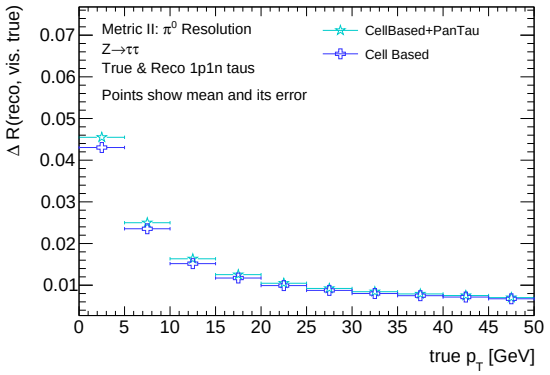


(A) *Transverse energy resolution.*

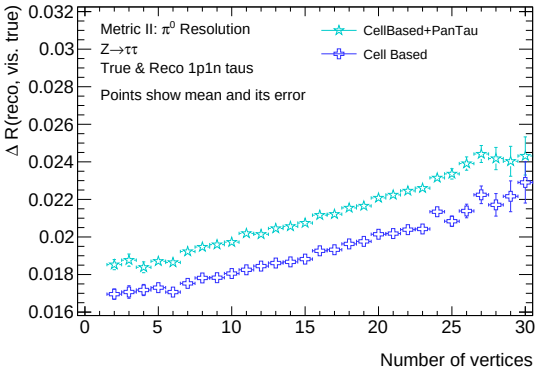


(B) *Distance in ΔR between true and reconstructed π^0.*

FIGURE 6.9: *Transverse energy resolution (left) and distance in ΔR between true and reconstructed π^0 (right) for π^0 in correctly classified 1p1n decays.* `CellBased` *in conjunction with* `PanTau` *performs slightly worse than* `CellBased` *alone. This may be due to the fact that the* `PanTau` *distributions include decays that are classified wrongly by* `CellBased` *and due to the combination of the two π^0-PFOs in 1pXn→1p1n reclassification, regardless of the tau momentum.*



(A) *Spatial resolution as a function of π^0 p_T.*



(B) *Spatial resolution as a function of the number of vertices in the event.*

FIGURE 6.10: *Distance in ΔR between true and reconstructed π^0 as a function of π^0 p_T (left) and number of vertices (right) for π^0 in correctly classified 1p1n decays.*

the two PFOs indeed stem from two photons from $\pi^0 \to \gamma\gamma$. Thus neglecting one of the two would reduce the energy and may remove the high energy tail.

The spatial resolution in ΔR shows a similar feature, a slightly more pronounced high ΔR tail for `PanTau`. The mean for `CellBased` here is $18.4 \cdot 10^{-3}$, for `PanTau` it is $19.6 \cdot 10^{-3}$.

The dependence of the mean of the ΔR between the true and reconstructed π^0 on $p_T(\pi^0)$ and the number of vertices is shown in Figure 6.10. The slight shift between `CellBased` and `PanTau` is also visible here, however the trends are exactly the same. The mean of the ΔR decreases with increasing π^0 p_T, because the boost of the π^0 increases and the shower is thus more collimated. As a function of the number of vertices, the mean of ΔR increases by $6 \cdot 10^{-3}$ over 30 vertices.

Figure 6.11 shows the E_T resolution as a function of π^0 p_T and number of vertices in the event.

The mean of the E_T resolution is stable for π^0 p_T between 5 and 50 GeV. For true π^0 p_T below 5 GeV, the mean is shifted up by $\approx 10\%$. This may be due to the subtraction procedure, but it may also be due to the fact that this bin, 0 to 5 GeV, also contains true π^0 below the actual E_T thresholds. If these decays are classified as 1p1n, the π^0 energy must be overestimated (otherwise the tau would not be classified as 1p1n). The RMS of the E_T resolution decreases with increasing π^0 p_T, as it is expected for calorimeter measurements. Starting at ≈ 0.37, the RMS decreases down to 0.1 for π^0 with $p_T > 30$ GeV.

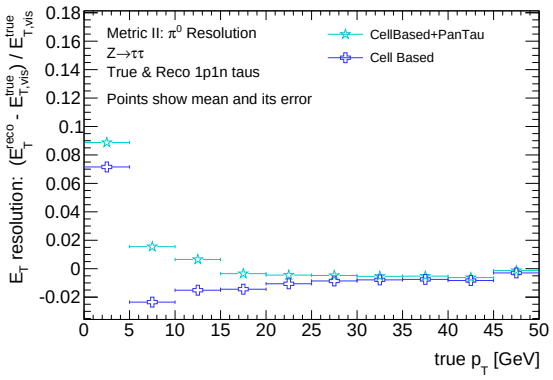
As a function of the number of vertices in the event, the mean of the E_T resolution increases by 7% over 30 vertices. The RMS shows a smaller dependency, increasing from 0.22 up to 0.26 over 30 vertices for `PanTau`. The RMS increase comes from a more pronounced energy overestimation with increased pile-up, as Figure 6.12 shows.

6.4 Metric III - Resonance Reconstruction

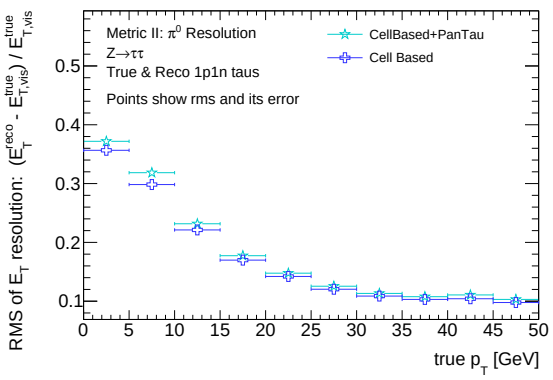
As outlined in Section 2.2, the decay of a tau lepton happens via intermediate resonances. The resonances are the ρ (1p1n) and the a_1 (1pXn, 3p0n). With the reconstruction of individual decay products, it is possible to reconstruct the resonances. Thus, calculating the invariant mass of the decay products should recover the resonance masses in the corresponding decay channels.

	True visible Tau		Reconstructed by `CellBased` + `PanTau`	
Resonance	Mean / MeV	RMS / MeV	Mean / MeV	RMS / MeV
ρ	$823 \pm 188 \cdot 10^{-3}$	$188 \pm 132 \cdot 10^{-3}$	$831 \pm 294 \cdot 10^{-3}$	$288 \pm 208 \cdot 10^{-3}$
a_1(1pXn)	$1.19 \cdot 10^3 \pm 261 \cdot 10^{-3}$	$175 \pm 184 \cdot 10^{-3}$	$1.28 \cdot 10^3 \pm 757 \cdot 10^{-3}$	$417 \pm 535 \cdot 10^{-3}$
a_1(3p0n)	$1.17 \cdot 10^3 \pm 294 \cdot 10^{-3}$	$174 \pm 208 \cdot 10^{-3}$	$1.14 \cdot 10^3 \pm 328 \cdot 10^{-3}$	$186 \pm 232 \cdot 10^{-3}$

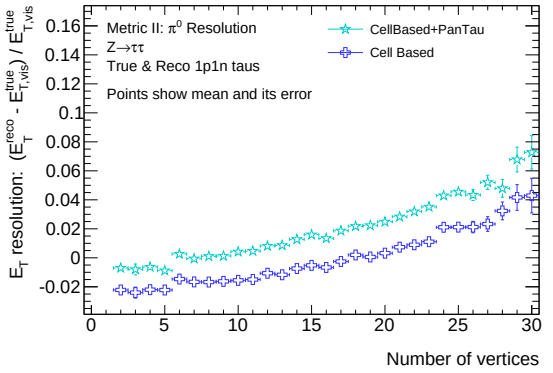
TABLE 6.3: *Mean and RMS values alongside their errors for the different invariant resonance masses. The table shows the values obtained from the generated tau (second and third column) and the ones obtained in the simulated reconstruction. Values are obtained in the cases in which the reconstructed decay mode is also the true simulated decay mode. The mean agrees reasonably well in all cases, the RMS is subject to the calorimeter energy resolution, as can be seen in the different RMS values for a_1(1p2n).*



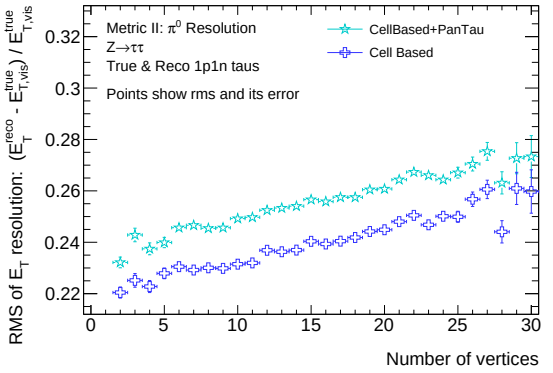
(A) *Mean and its error of the transverse energy resolution as a function of π^0 p_T.*



(B) *RMS and its error of the transverse energy resolution as a function of π^0 p_T.*



(C) *Mean and its error of the transverse energy resolution as a function of number of vertices.*



(D) *RMS and its error of the transverse energy resolution as a function of number of vertices.*

FIGURE 6.11: *The mean (left), RMS (right) and their respective errors of the transverse energy resolution as a function of π^0 p_T (top) and the number of vertices in the event (bottom).*

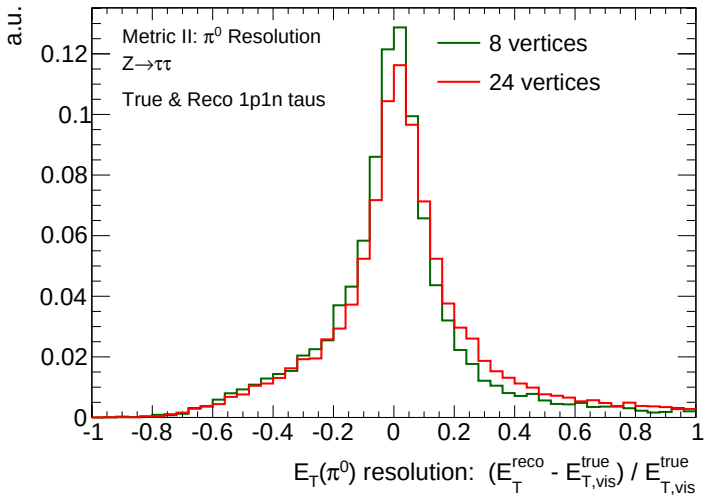


FIGURE 6.12: *E_T resolution of π^0 in correctly classified 1p1n decays for two bins of the number of vertices. Both distributions are normalised to unity. The resolution in events with 24 vertices shows a bigger high energy tail than the resolution for events with 8 vertices. The low energy tail stays the same. As a result, both the mean and the RMS increase with pile-up.*

ρ Resonance

Figure 6.13 shows the invariant mass distribution in true and reconstructed 1p1n decays and the invariant mass for all reconstructed 1p1n decays[7]. The mean and RMS of the invariant mass distributions for all resonances (ρ, a_1(1p2n), a_1(3p0n)) are listed in Table 6.3. In case of the ρ resonance, the reconstructed mean of 831 MeV agrees very well with the simulated one (823 MeV), and the RMS is $\approx$ 50% larger, because of the energy resolution of the EM calorimeter (c.f. the π^0 resolution in Figure 6.9).

The mean as a function of the resonance p_T is shown in Figure 6.14, which shows that it is constant at 800 − 900 MeV for true $\tau_{\text{had-vis}}$ p_T between 15 GeV and 100 GeV. The PanTau curve follows the curve of the generated true visible tau in both the mean and the RMS, the only difference is that the RMS of PanTau is higher than the generated one.

Figure 6.15 shows the dependence of the mean and RMS of the invariant 1p1n visible mass on the number of vertices in the event. The mean increases by roughly 100 MeV over 30 vertices, and the RMS by $\approx$ 80 MeV. Because the composition of the tau is fixed to one charged PFO and one π^0-PFO in 1p1n decays, the increase of energy has to come from pile-up particles that partially or completely overlap with the π^0-PFO, increasing its energy and thus the mass of the complete tau. As can be seen from Figure 6.12, the contribution of pile-up to the overestimation of the energy is one source of the increasing RMS.

a_1 Resonance

The a_1 resonance decays into both 1p2n and 3p0n. Decays into three charged pions are easier to reconstruct, because the charged pions produce a track in the inner detector, which can be resolved very well. Decays into one charged and two neutral pions however need to rely on the subtraction of the charged pion in the EM calorimeter and the correct measurement of the two neutral pions, making the reconstruction less accurate.

Figure 6.16 shows the invariant mass in 3p0n decays. In part 6.16a, only correctly reconstructed 3p0n taus are considered, while part 6.16b shows the invariant mass distribution of all reconstructed 3p0n taus. For correctly reconstructed 3p0n taus, the mean of simulated and reconstructed mass agrees within 30 MeV, the RMS also agrees well within 12 MeV (c.f. Table 6.3).

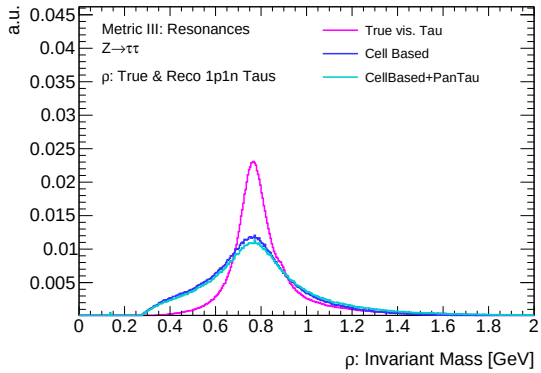
The dependence of the reconstructed invariant mass on the true $\tau_{\text{had-vis}}$ p_T is depicted in 6.17. While the mean is constant slightly below 1.2 GeV, the RMS increases by 40 MeV when raising the true $\tau_{\text{had-vis}}$ p_T from 15 GeV to 100 GeV. It is consistent with the behaviour of the generated mass.

Because the reconstruction of 3p0n decays depends solely on tracks, the pile-up stability is excellent (c.f. Figure 6.18). There is no visible shift in the mean of the reconstructed mass (Fig. 6.18a) and no clear trend in the RMS, shown in Figure 6.18b.

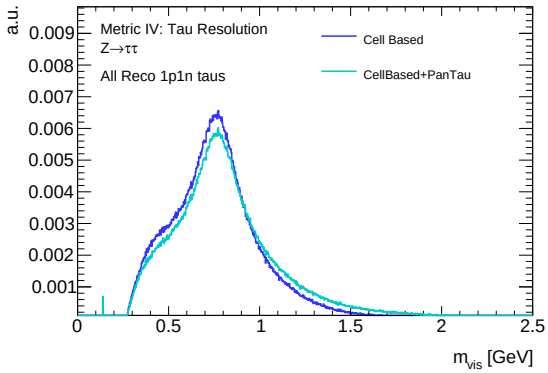
The situation is less optimal in $a_1 \to \pi^\pm\pi^0\pi^0$ decays. The reconstructed invariant mass spectrum is shown in Figure 6.19. Clearly visible is the bad resolution compared to the 3p0n case, which is due to two of the $\pi^\pm$ being replaced with π^0, measured in the calorimeter with a worse resolution. This results in the reconstructed RMS being more than twice as large as the generated one as Table 6.3 shows. However, the shift in the mean of the distribution is rather small (overestimation < 10%).

The dependency on the true $\tau_{\text{had-vis}}$ p_T is depicted in Figure 6.20. The reconstructed mass as a function of p_T is stable within $\approx$ 150 MeV, the RMS increases from 300 MeV to 600 MeV. This is due

[7] The appearence of entries at the $\pi^\pm$-mass is understood and corrected in newer versions of the code that produces these figures (pages 53 and 200 in [57]).

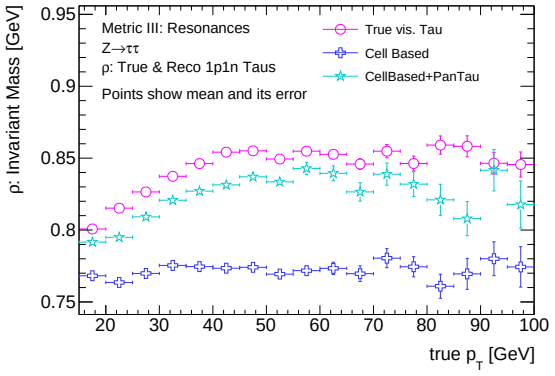


(A) *Invariant mass in true and reconstructed 1p1n decays.*

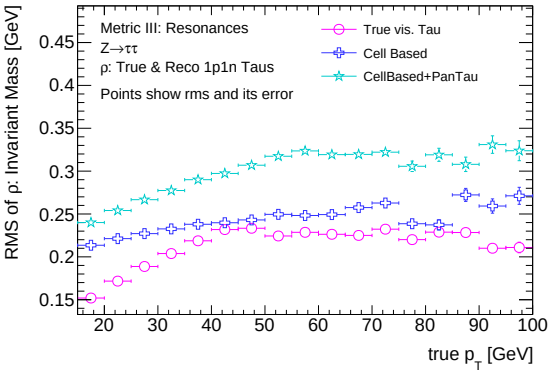


(B) *Invariant mass in reconstructed 1p1n decays.*

FIGURE 6.13: *Invariant mass in 1p1n decays. The mean of the distributions agrees very well, the peak position is also correct. The RMS is larger because of the E_T resolution of the π^0. Whether the reconstructed mode is also the true mode has no big effect on the shape of the mass distribution (right).*

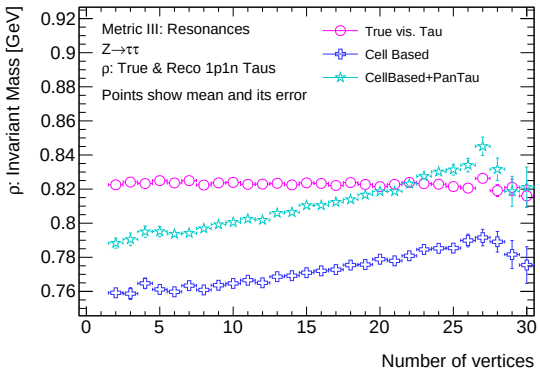


(A) *Mean and its error of true and reconstructed 1p1n decays as a function of true $\tau_{had\text{-}vis}$ p_T.*

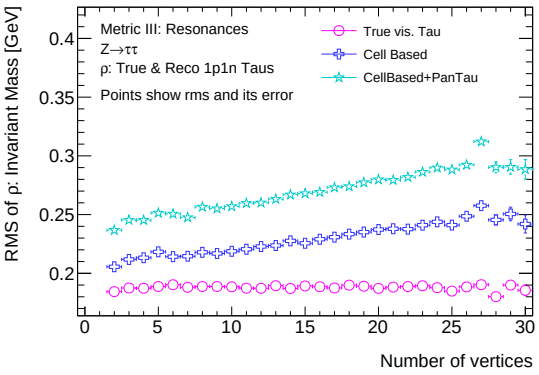


(B) *RMS and its error of true and reconstructed 1p1n decays as a function of true $\tau_{had\text{-}vis}$ p_T.*

FIGURE 6.14: *Mean, RMS and their errors in true and reconstructed 1p1n decays as a function of true $\tau_{had\text{-}vis}$ p_T.*

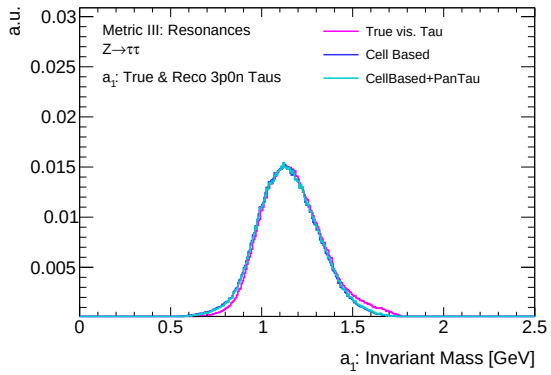


(A) *Mean and its error of the invariant mass of true and reconstructed 1p1n decays.*

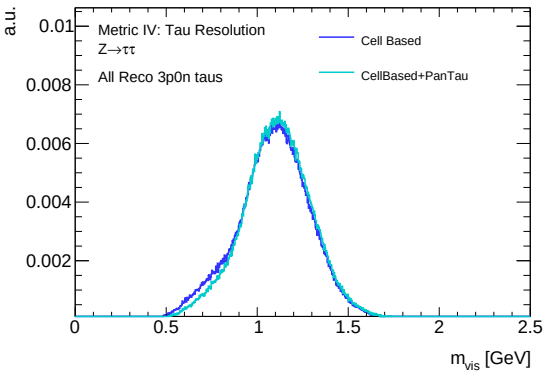


(B) *RMS and its error of the invariant mass of true and reconstructed 1p1n decays.*

FIGURE 6.15: *Mean, RMS of the invariant $\tau_{had\text{-}vis}$ mass and their errors as a function of the number of vertices.*

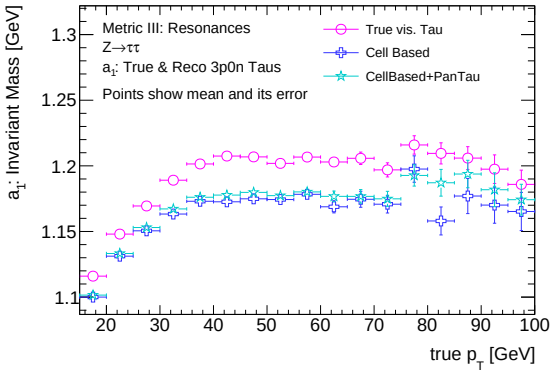


(A) *Invariant mass in correct 3p0n decays.*

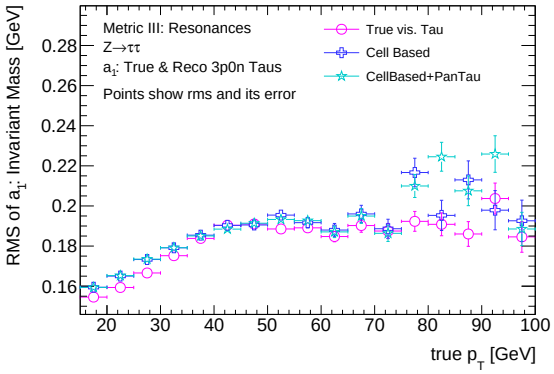


(B) *Invariant mass in reconstructed 3p0n decays.*

FIGURE 6.16: *Invariant mass of 3p0n decays. The reconstructed masses are very close to the generated ones, thanks to the excellent tracker resolution.*

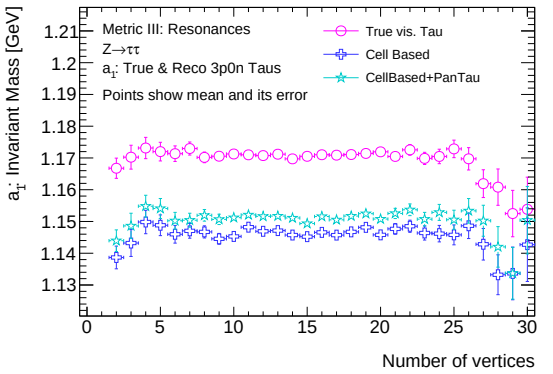


(A) *Mean and its error as a function of true $\tau_{had\text{-}vis}$ p_T.*

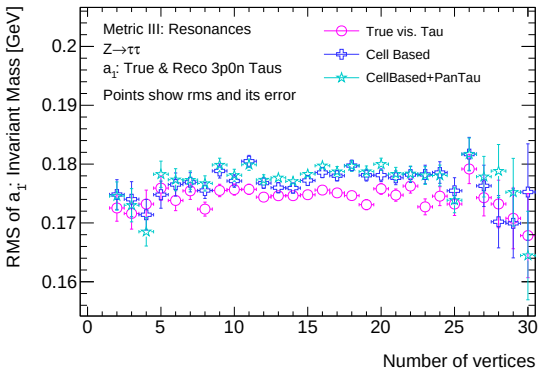


(B) *RMS and its error as a function of true $\tau_{had\text{-}vis}$ p_T.*

FIGURE 6.17: *Mean and RMS of the invariant mass in 3p0n decays as a function of true $\tau_{had\text{-}vis}$ p_T.*

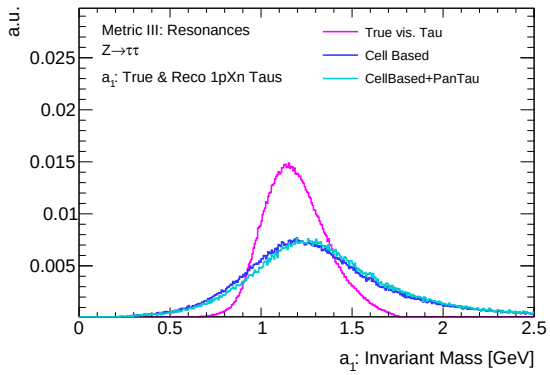


(A) *Mean and its error in bins of reconstructed vertices.*

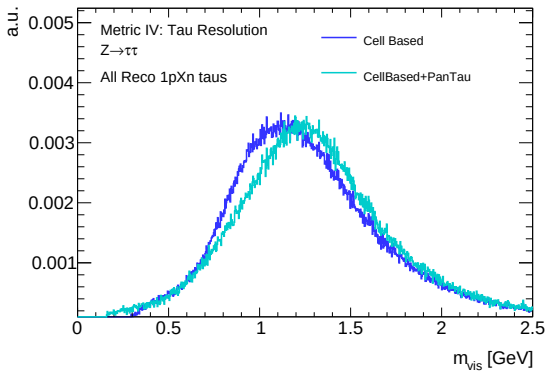


(B) *RMS and its error in bins of reconstructed vertices.*

FIGURE 6.18: *Mean and RMS of the invariant mass of 3p0n decays in bins of reconstructed vertices in the event.*

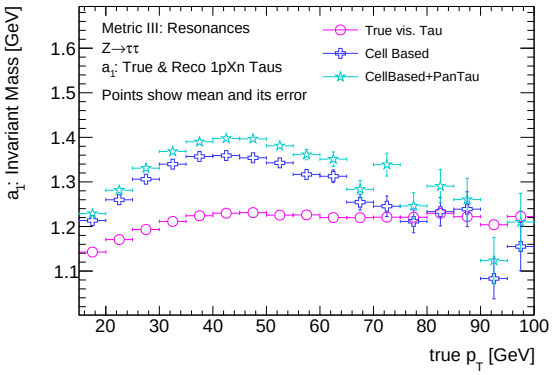


(A) *Invariant mass in correct 1p2n decays.*

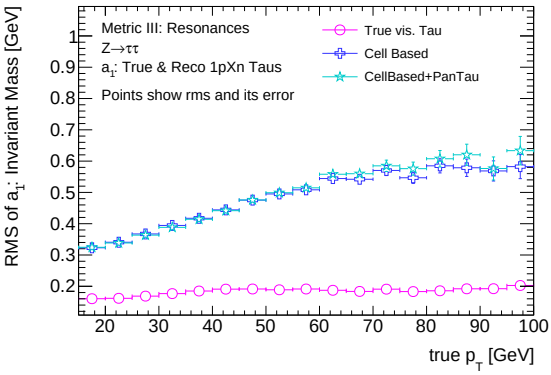


(B) *Invariant mass in reconstructed 1p2n decays.*

FIGURE 6.19: *Invariant mass of 1pXn decays. Because of the presence of two π^0, the mass resolution is worse than in 1p1n decays and 3p0n decays. The peak position is not affected as much as the width of the distributions.*

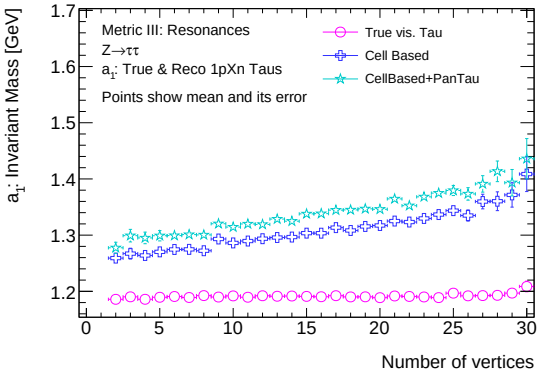


(A) *Mean and its error of the invariant mass in true and reconstructed 1pXn decays as a function of $p_T(a_1)$.*

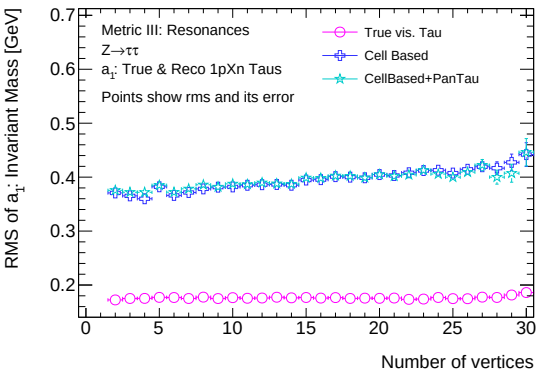


(B) *RMS and its error of the invariant mass in true and reconstructed 1pXn decays as a function of $p_T(a_1)$.*

FIGURE 6.20: *Mean, RMS and the corresponding errors in the reconstruction of the a_1 resonance in true and reconstructed 1p2n decays as a function of the resonance p_T.*



(A) *Mean and its error of the invariant mass in true and reconstructed 1pXn decays as a function of N_{Vertex}.*



(B) *RMS and its error of the invariant mass in true and reconstructed 1pXn decays as a function of N_{Vertex}.*

FIGURE 6.21: *Mean, RMS and the corresponding errors in the reconstruction of the a_1 resonance in true and reconstructed 1p2n decays as a function of vertices in the event.*

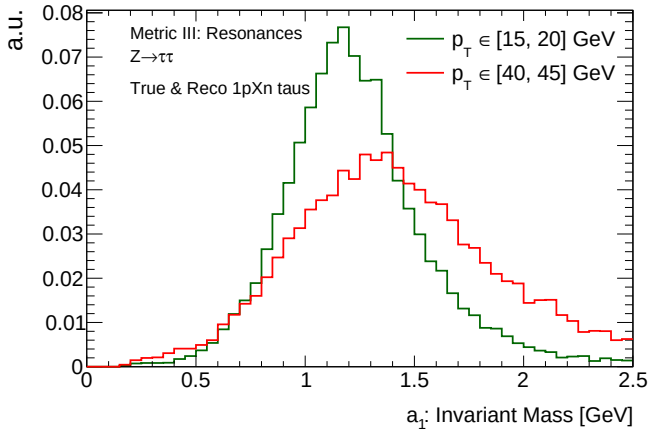
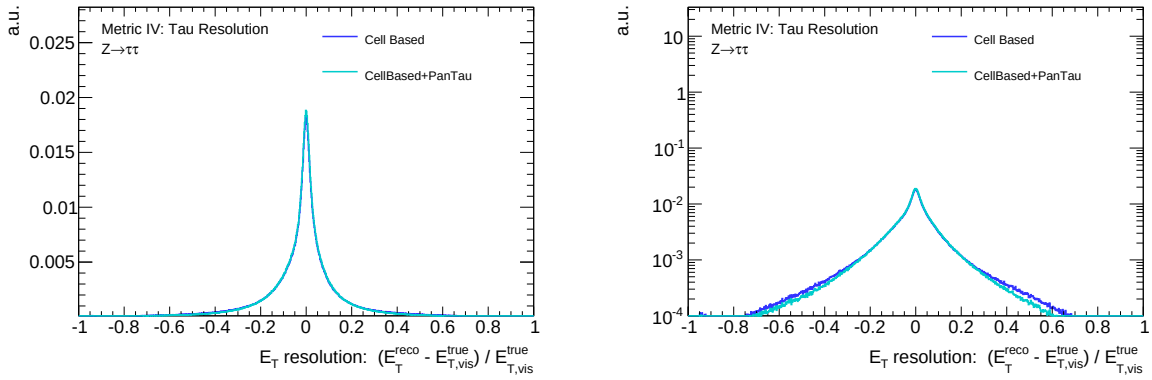


FIGURE 6.22: *Invariant mass from* `PanTau` *in correctly classified 1pXn decays, for low (green) and higher (red) resonance momenta. It is visible, that the mass is overestimated for higher resonance momenta and that the distribution is broadened.*



(A) *Linear y-axis scale.* (B) *Logarithmic y-axis scale.*

FIGURE 6.23: *Inclusive transverse energy resolution, with linear (left) and logarithmic (right) y-axis in order to see the core and the tails. The* `CellBased` *and* `PanTau` *resolutions are very similar, especially in the core region. In the tails,* `PanTau` *performs a bit better than* `CellBased`*.*

to an overestimation and broadening of the mass, shown in Figure 6.22. The effect is visible for both `CellBased` and `PanTau`, so that it is not because of a wrong selection of π^0. One reason for this can be that with increasing p_T, the decay products are more boosted, so that it is harder to accurately resolve individual π^0-PFOs, which deteriorates the mass resolution.

As a function of the number of vertices in the event, the a_1 mass reconstruction performs as shown in Figure 6.21. The mean of the mass increases from 1.2 GeV to 1.4 GeV over 30 vertices, the RMS increases from 400 MeV to 500 MeV.

6.5 Metric IV - Global Tau Resolution

Finally, this section discusses the resolution of the $\tau_{\text{had-vis}}$ object as a whole, as reconstructed by substructure based tau reconstruction. The discussion will focus around the energy resolution, because the energy measurement is quite sensitive to the reconstruction of the $\tau_{\text{had-vis}}$. The spatial resolution, $\Delta\phi$ and $\Delta\eta$ will be discussed briefly at the end of the section.

Reconstructed Decay Mode	Inclusive	1p0n	1p1n	1pXn	3p0n	3pXn
	Mean of E_T resolution / %					
1p0n	-7.4	-3.9	-12.6	-25.7	-47.4	-56.6
1p1n	-1.7	$634 \cdot 10^{-3}$	$-174.5 \cdot 10^{-3}$	-5	-36.3	-30.6
1pXn	1.4	-10.5	3.7	1.5	-35.6	-26.9
3p0n	-3.3	13	-7.7	-17.9	$-175.5 \cdot 10^{-3}$	-14.6
3pXn	15	12	25	5.6	38	8.2
	Error of mean of E_T resolution / %					
1p0n	$30 \cdot 10^{-3}$	$33 \cdot 10^{-3}$	$53 \cdot 10^{-3}$	$208 \cdot 10^{-3}$	$324 \cdot 10^{-3}$	$923 \cdot 10^{-3}$
1p1n	$16 \cdot 10^{-3}$	$110 \cdot 10^{-3}$	$16 \cdot 10^{-3}$	$30 \cdot 10^{-3}$	$267 \cdot 10^{-3}$	$250 \cdot 10^{-3}$
1pXn	$31 \cdot 10^{-3}$	$316 \cdot 10^{-3}$	$60 \cdot 10^{-3}$	$31 \cdot 10^{-3}$	$502 \cdot 10^{-3}$	$390 \cdot 10^{-3}$
3p0n	$16 \cdot 10^{-3}$	$582 \cdot 10^{-3}$	$190 \cdot 10^{-3}$	$353 \cdot 10^{-3}$	$12 \cdot 10^{-3}$	$37 \cdot 10^{-3}$
3pXn	$61 \cdot 10^{-3}$	1.8	$149 \cdot 10^{-3}$	$160 \cdot 10^{-3}$	$180 \cdot 10^{-3}$	$63 \cdot 10^{-3}$
	RMS of E_T resolution / %					
1p0n	20	18	17	22	26	26
1p1n	16	31	13	14	25	21
1pXn	16	28	18	13	21	20
3p0n	10	27	17	17	6.8	10
3pXn	26	44	26	24	28	20
	Error of RMS of E_T resolution / %					
1p0n	$21 \cdot 10^{-3}$	$23 \cdot 10^{-3}$	$38 \cdot 10^{-3}$	$147 \cdot 10^{-3}$	$229 \cdot 10^{-3}$	$653 \cdot 10^{-3}$
1p1n	$11 \cdot 10^{-3}$	$78 \cdot 10^{-3}$	$11 \cdot 10^{-3}$	$21 \cdot 10^{-3}$	$189 \cdot 10^{-3}$	$177 \cdot 10^{-3}$
1pXn	$21 \cdot 10^{-3}$	$224 \cdot 10^{-3}$	$43 \cdot 10^{-3}$	$22 \cdot 10^{-3}$	$355 \cdot 10^{-3}$	$276 \cdot 10^{-3}$
3p0n	$11 \cdot 10^{-3}$	$412 \cdot 10^{-3}$	$134 \cdot 10^{-3}$	$249 \cdot 10^{-3}$	$8.8 \cdot 10^{-3}$	$26 \cdot 10^{-3}$
3pXn	$43 \cdot 10^{-3}$	1.3	$105 \cdot 10^{-3}$	$113 \cdot 10^{-3}$	$127 \cdot 10^{-3}$	$45 \cdot 10^{-3}$

TABLE 6.4: *Mean, its error and RMS and its error of the different contributions to the transverse energy resolution of the reconstructed decay modes. Inclusive refers to the distribution that contains every τ_{had} reconstructed as a given decay mode. Values are given in %, i.e. a mean of -4.4% in true and reconstructed 1p0n decays means that the energy is underestimated by 4.4%. Analogously, an RMS of 13% in true and reconstructed 1p1n means that the RMS is 13% of the tau E_T.*

The inclusive transverse energy resolution is shown in Figure 6.23. The mean of the PanTau distribution is -0.0135, and the RMS is 0.1831. Because the inclusive resolution is the sum of the individual resolutions of the different reconstructed modes, the shape of the resolution is not Gaussian. There are two major contributions to the shape, the core region between -0.05 and +0.05 and the low energy and high energy tails between -1.0 and -0.05 and above 0.05, respectively.

The core is dominated by the reconstructed $\tau_{\text{had-vis}}$ decaying into 1p0n and 3p0n, see Figures 6.25a and 6.25d. In these decays, the resolution is only limited by the tracking resolution, which is $< 3\%$[8] in $Z \to \tau\tau$ processes in which the tau decay products have momenta below 50 GeV. The tails have several contributions, which are discussed in the following paragraphs. A quantitative summary of the resolutions is given in Table 6.4, which summarises the mean and the RMS of the reconstructed modes alongside their different contributions.

The low energy tail between 0.05 and -0.4 in the 1p0n resolution (6.25a) is dominated by true 1p1n decays, in which the neutral pion is not reconstructed. Thus, the energy of the $\tau_{\text{had-vis}}$ that is carried away in the π^0 is missing in the reconstructed energy.

The low energy tail below -0.4 is in fact dominated by true 1p0n decays. This is due to a technical reason. In the computing step creating the simulated data samples, only $\pi^\pm$ and π^0 are used in the determination of the true decay mode, i.e. decays $\tau^\pm \to \pi^\pm K^0 \nu_\tau$ are classified as 1p0n. The neutral kaons tend to shower in the hadronic calorimeter, so no neutral energy is reconstructed in those decays. For the charged energy, the track of the $\pi^\pm$ is used. Thus, there are $\tau^\pm \to \pi^\pm K^0 \nu_\tau$, depositing large parts of their energy in the hadronic calorimeter, which are not reconstructed, that are read-out as 1p0n decays. This leads to a significant underestimation of the energy, leading to the large low energy tail in true and reconstructed 1p0n decays.

Contributions from other modes are located at negative resolution values as well, because at least one particle of the true decay is not reconstructed.

In the energy resolution of the reconstructed 1p1n decays (6.25b), the correctly reconstructed 1p1n are centered at 0 and the shape is symmetric up to ≈ 0.2.

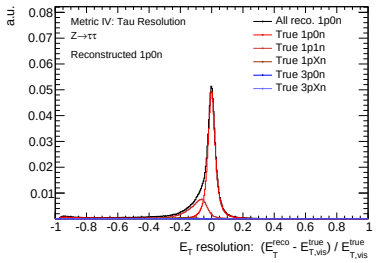
At this point it is worth mentioning that the x-axis is inherently asymmetric. The co-domain c of this measure of resolution is $c = (-1, \infty)$ with a perfect resolution being a delta function at 0. An underestimation of the energy by a factor of 2 is located at -0.5, while an overestimation by a factor of 2 is located at +2.0. Thus, equal under- & overestimation is not located at equal distances from the optimal value of the resolution measure.

The contribution of true 1p1n to the reconstructed 1p1n energy resolution is dominant. Only in the far tails, below -0.6 and above 0.5 the misreconstructed 1p0n decays become equally important. True 1pXn decays, in which one π^0 is not reconstructed contribute at resolutions around -0.2. 3-prong decays barely make a contribution to the resolution.

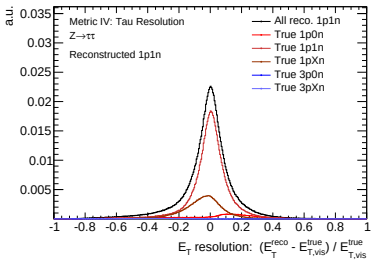
True 1p0n decays are important at below -0.6 and above 0.5. Below -0.6, it is due to the neutral kaons which are not considered in the decay mode calculation. The effect is not as prominent, because the tau has been reconstructed as 1p1n, i.e. an additional cluster, not belonging to the tau, contributes to the energy. The contribution above 0.5, peaking at 0.2, may come from picking up a random cluster that does not belong to the decay at all.

True 1pXn decays misclassified as 1p1n are slightly shifted towards negative resolutions. This may be due to one of the two π^0 being rather soft, making it hard to be reconstructed and not shifting the energy too much.

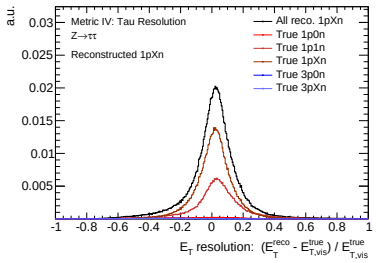
[8] According to the design goal of $\sigma(p_T)/p_T = 0.05\% \cdot p_T / \text{GeV} \oplus 1\%$.



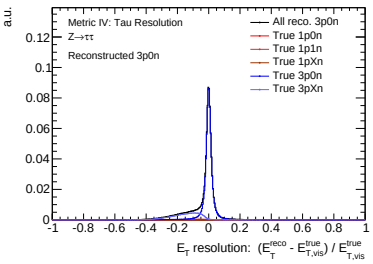
(A) *Reconstructed 1p0n $\tau_{had\text{-}vis}$*



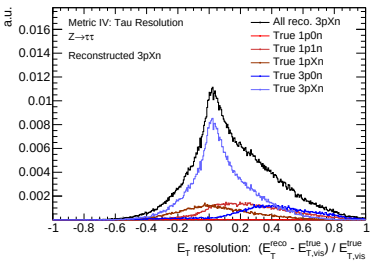
(B) *Reconstructed 1p1n $\tau_{had\text{-}vis}$*



(C) *Reconstructed 1pXn $\tau_{had\text{-}vis}$*

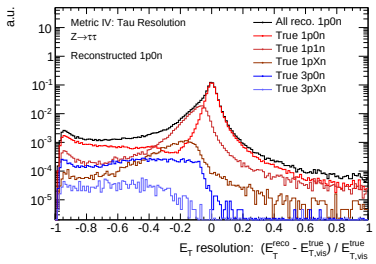


(D) *Reconstructed 3p0n $\tau_{had\text{-}vis}$*

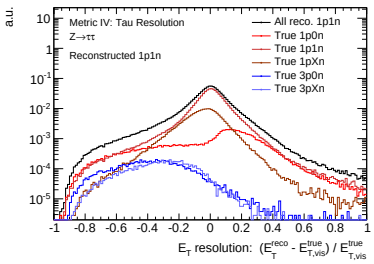


(E) *Reconstructed 3pXn $\tau_{had\text{-}vis}$*

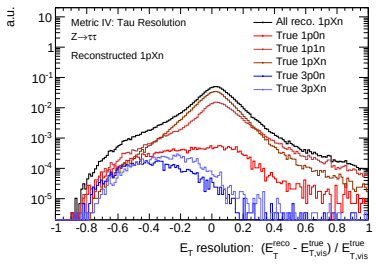
FIGURE 6.24: *Transverse energy resolution in reconstructed modes. The plots are the same as the ones in Figure 6.25, except for a slightly finer binning and the linear y-axis.*



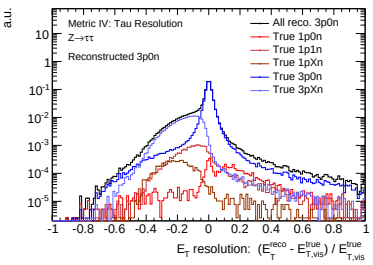
(A) *Reconstructed 1p0n $\tau_{had\text{-}vis}$*



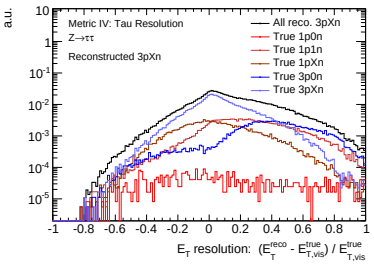
(B) *Reconstructed 1p1n $\tau_{had\text{-}vis}$*



(C) *Reconstructed 1pXn $\tau_{had\text{-}vis}$*



(D) *Reconstructed 3p0n $\tau_{had\text{-}vis}$*



(E) *Reconstructed 3pXn $\tau_{had\text{-}vis}$*

FIGURE 6.25: *Transverse energy resolution in reconstructed modes. Note that the plots show indeed the resolution of the reconstructed modes, i.e. there are different true decay modes in each distribution and the relative abundances are as given in the purity matrix in Figure 6.3d.*

For reconstructed 1pXn decays, the core resolution is dominated by true 1p1n and true 1pXn decays (c.f. Figure 6.25c and 6.24c). While the 1pXn contribution is centered at zero and is rather symmetric, the contribution from true 1p1n decays is shifted towards higher energies, because an additional cluster not belonging to the tau is picked up.

The true 1p1n reconstructed as 1pXn are the main contribution in both tails. In the high energy tail because of an additional cluster, and in the low energy tail presumably due to the same effect as the 1p0n contribution in the 1p0n resolution, namely neutral kaons.

True 1p0n and 3-prong taus are almost negligible and only contribute to the very low energy tail.

The core of the 3p0n resolution (Figures 6.25d & 6.24d) consists purely of true 3p0n decays. The low energy shoulder comes from true 3pXn, in which the neutral cluster was not picked up, hence their energy is underestimated. The tails are also dominated by 3p0n decays.

Finally, Figures 6.25e and 6.24e show the energy resolution of reconstructed 3pXn taus. While the high energy tail of true 3pXn reconstructed as such is not very pronounced, it is augmented by basically every other decay mode, except for 1p0n.

Energies of true 1p1n and 1pXn decays reconstructed as 3p0n are overestimated. The reason they are classified as 3pXn could be that at least one of the at least two photons in the decay converts into a e^+e^- pair, so that there are additional tracks and EM-clusters in the core region of the tau. If the conversion happens rather early, it might happen that the tracks produced by the electrons and positrons fulfil the selection criteria, so that a $\tau_{\text{1-prong}}$ is reconstructed as a $\tau_{\text{3-prong}}$. The migration matrix in Figure 6.3c shows, that the probability to reconstruct a $\tau_{\text{1-prong}}$ as a $\tau_{\text{3-prong}}$ increases with increasing number of neutrals.

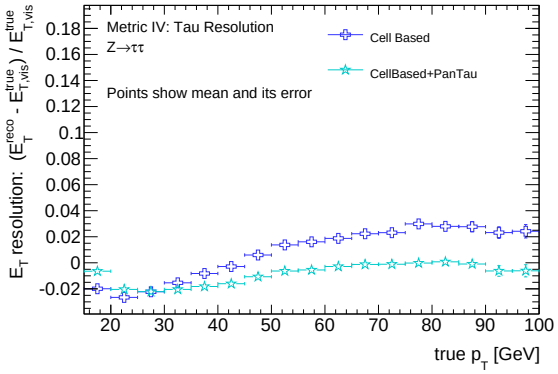
Generally, if a τ_{had} which did not decay over 3pXn is classified as such, it must be due to clusters and/or tracks that do not belong to the tau, making an overestimation of the energy inevitable. The opposite is true for the 1p0n decay mode.

In Figure 6.26, the dependence of the mean (Fig. 6.26a) and the RMS (Fig. 6.26b) of the inclusive transverse energy resolution on $p_T(\tau_{\text{had-vis}})$ is shown. The mean of the E_T resolution is stable around $\approx -0.01 \pm 0.01$ up to 100 GeV, and the RMS is also stable around 0.18 ± 0.01.

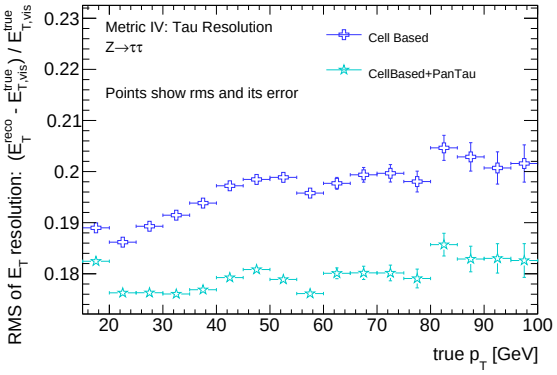
The inclusive E_T resolution is the sum of the individual E_T resolutions from the different decay modes. As discussed above, each of those has contributions from all true modes, with relative abundances given by the migration matrix in Figure 6.3d. Thus, the inclusive distributions are highly non-Gaussian, and the RMS may not be the best figure of merit. However, the RMS still is a measure for the width of the distribution, so it is used as such for the remainder of this chapter.

Figure 6.27 shows the pile-up dependency of the E_T resolution, both for its mean and RMS, in Figures 6.27a and 6.27b. With increasing pile-up, the mean of the resolution is expected to be shifted to higher values, because the probability of reconstructing a pile-up cluster as a neutral pion and thus including it in the E_T calculation increases. Over a range of $\approx$ 30 vertices, this is indeed the case for `PanTau`, the mean increases from -2% at two vertices to 0% at 30 vertices.

The resolution of the transverse energy as a function of the azimuthal angle ϕ is shown in Figure 6.28, both for the mean and the RMS. The mean of the E_T resolution (Fig. 6.28a) fluctuates between -0.02 and -0.01, so that there is a slight underestimation of the energy. The RMS of the E_T resolution (Fig. 6.28b) does not depend on ϕ.

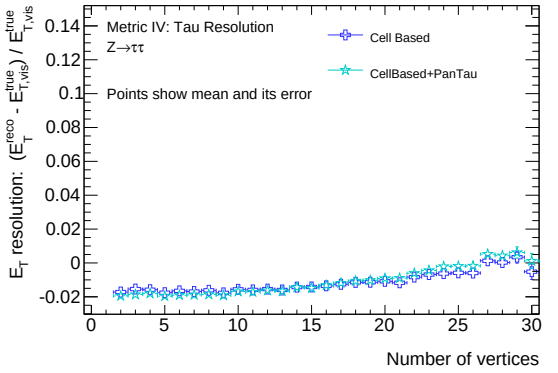


(A) *Mean and its error of the inclusive E_T-resolution as a function of true $\tau_{had\text{-}vis}$ p_T.*

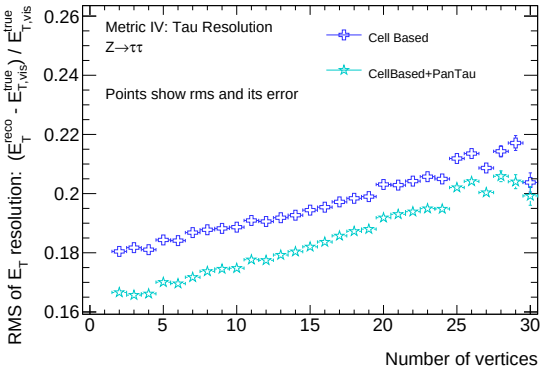


(B) *RMS and its error of the inclusive E_T-resolution as a function of true $\tau_{had\text{-}vis}$ p_T.*

FIGURE 6.26: *Inclusive transverse energy resolution as a function of the p_T of the true $\tau_{had\text{-}vis}$.*

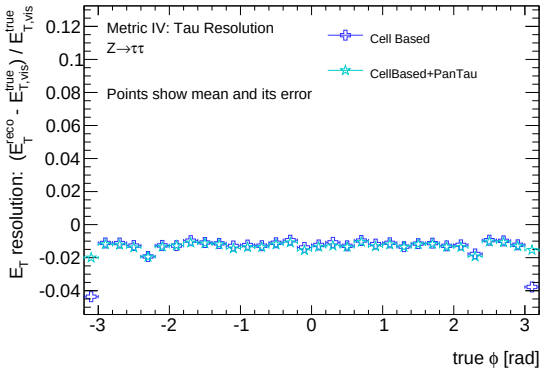


(A) *Mean and its error of the inclusive E_T-resolution as a function of the number of reconstructed vertices.*

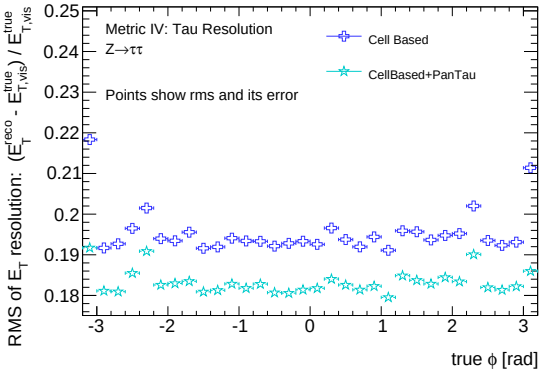


(B) *RMS and its error of the inclusive E_T-resolution as a function of the number of reconstructed vertices.*

FIGURE 6.27: *Inclusive transverse energy resolution as a function of the number of vertices in the event.*



(A) *Mean and its error of the inclusive E_T-resolution as a function of the azimuthal angle ϕ.*



(B) *RMS and its error of the inclusive E_T-resolution as a function of the azimuthal angle ϕ.*

FIGURE 6.28: *Inclusive transverse energy resolution as a function of the azimuthal angle ϕ*

Concluding the discussion of the E_T resolution, Figure 6.29 shows it as a function of the pseudorapidity η. In the transition region and in the end-cap region, the energy is underestimated by 3% - 4% (Fig. 6.29a), likely because of the poorer instrumentation in the transition region and the end-caps in comparison to the barrel. There, the energy is overestimated by 1%. This could be corrected for by applying a method similar to the one that is used in the tau reconstruction during the first data-taking period [58]. The method described there applies an correction factor, which is binned in η, the number of tracks and tau p_T, to scale the measured momentum to the true momentum. The basic idea of that method is also applicable in substructure based tau reconstruction. Figure 6.29b shows the RMS of the E_T resolution as a function of η, and except for the transition region, it is flat.

The difference between the generated and the reconstructed azimuthal angle ϕ, as well as the difference in the corresponding pseudorapidities of tau candidates is shown in Figure 6.30. Both distributions are sums of the distributions for the different decay modes. As a result, they have two components, the core and the tails. The former is because of the fact that the 1p0n and 3p0n modes are measured solely with the tracker, which provides an excellent position resolution. The latter is mainly due to measurements in the EM calorimeter, which have a greater impact on the resolution with increasing number of π^0 in the decay.

Taking the tails into account, the position precision in ϕ and η is better than 0.04 in both cases, corresponding to $\Delta\phi < 2.3^\circ$ and $\Delta\theta < 2.0^\circ$ at $\eta = 0$.

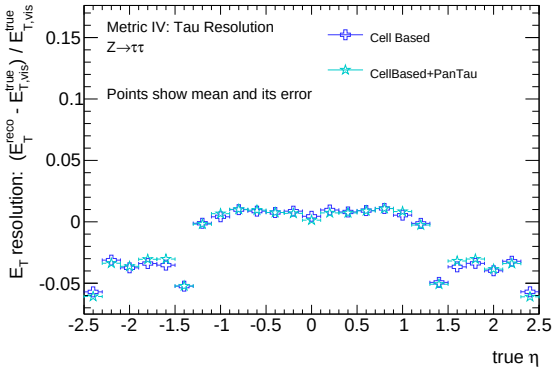
6.6 Summary

In summary the substructure based tau reconstruction discussed in this and the last chapter provides a robust and precise method to reconstruct tau leptons in the ATLAS detector.

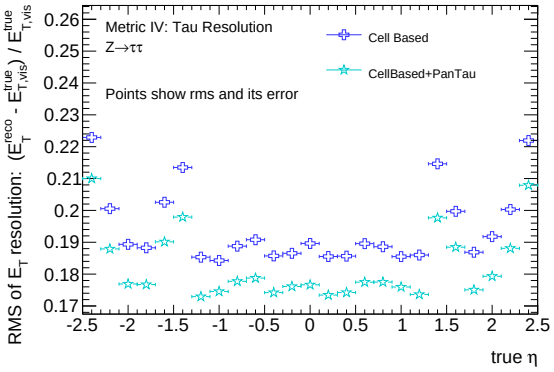
In comparison to the default method, which can only classify taus as $\tau_{\text{1-prong}}$ or $\tau_{\text{3-prong}}$, the substructure based tau reconstruction significantly improves the energy and spatial resolution, as shown in Figure 6.31. The default method is based on calorimeter measurements only and does not differentiate between different decay modes, so that the width is larger and the shape is Gaussian. In the substructure based tau reconstruction, the core is very sharp and dominated by pure track measurements in 1p0n and 3p0n decays. The tails come from misclassified decays, as discussed in Section 6.5. Thus, the new method excels in the core, but has some weaknesses in the far tails, in which the old method performs better. However, the new method is still being developed, and not all possible improvements have been made (c.f. Section 5.5.2), so that there is potential to improve the tails.

The spatial resolution is improved drastically. Distances between the generated and the reconstructed tau are below 0.002 in $\eta - \phi$ (1p0n and 3p0n taus, measured in the tracker), outliers come from the other modes, which still utilises the tracker measurements for the $\pi^\pm$.

In addition to these results, which are based purely on simulations, first comparisons between data and simulations have been carried out [52] and they show good agreement. The study looked into fake tau leptons, obtained from $Z \to \mu\mu$ + Jets events. Here, the 4-momentum of the Z-Boson is obtained by combining the well measured 4-momenta of the two muons, which also provide a very clean signature to trigger on and select these events. Opposite to the Z-Boson in the transverse plane, it is searched for a tau candidate. This tau candidate (although not a real tau) is passed to the substructure based tau reconstruction, which will handle it as if it was a tau and thus also try to determine the decay mode, for instance. The simulated and measured abundances of decay modes in this sample of fake taus are shown in Figure 6.32.

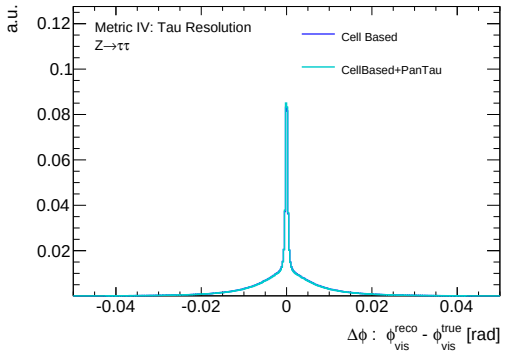


(A) *Mean and its error of the inclusive E_T-resolution as a function of pseudorapidity η.*

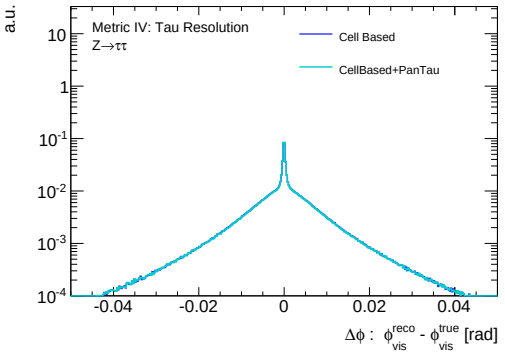


(B) *RMS and its error of the inclusive E_T-resolution as a function of pseudorapidity η.*

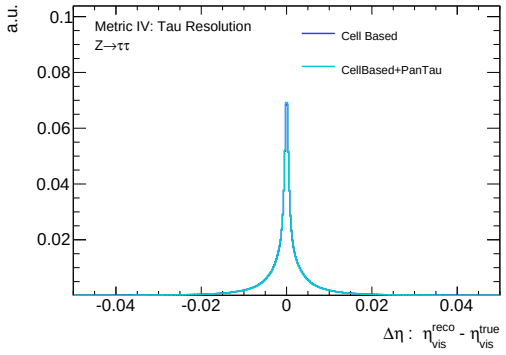
FIGURE 6.29: *Inclusive transverse energy resolution as a function of pseudorapidity η*



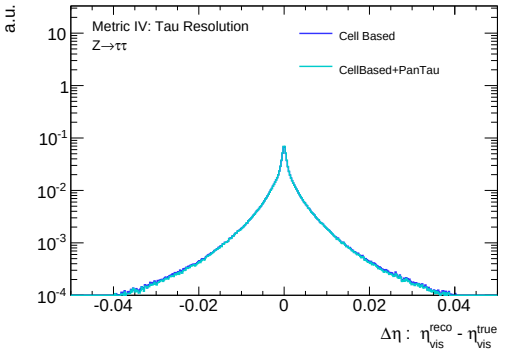
(A) *Difference in generated and reconstructed ϕ on a linear scale.*



(B) *Difference in generated and reconstructed ϕ on a logarithmic scale.*

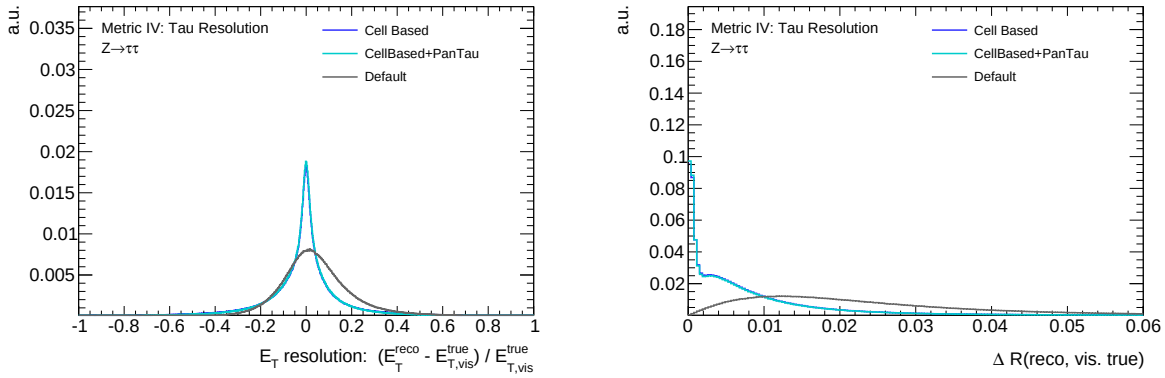


(C) *Difference in generated and reconstructed η on a linear scale.*



(D) *Difference in generated and reconstructed η on a logarithmic scale.*

FIGURE 6.30: *Difference between generated and reconstructed ϕ (top) and η (bottom) on a linear (left) and logarithmic (right) scale. The distributions are inclusive over all reconstructed modes. The sharp peaks at 0 are due to the excellent tracker resolution of only a few percent, which is exploited in to its full extend in 1p0n and 3p0n modes ($\approx$ 33% of all hadronic modes). The tails stem from the other modes, which need to rely on the EM-calorimeter for π^0 reconstruction. Like outlined in the discussion of the E_T resolution, the distributions are highly non-Gaussian.*



(A) *Inclusive transverse energy resolution.* (B) *Inclusive spatial resolution.*

FIGURE 6.31: *Transverse energy resolution and spatial resolution of all reconstructed tau leptons passing the selection cuts. The transverse energy resolution has a much more pronounced core, while the far tails are worse than in the default reconstruction. The accuracy in the reconstruction of the position increases drastically.*

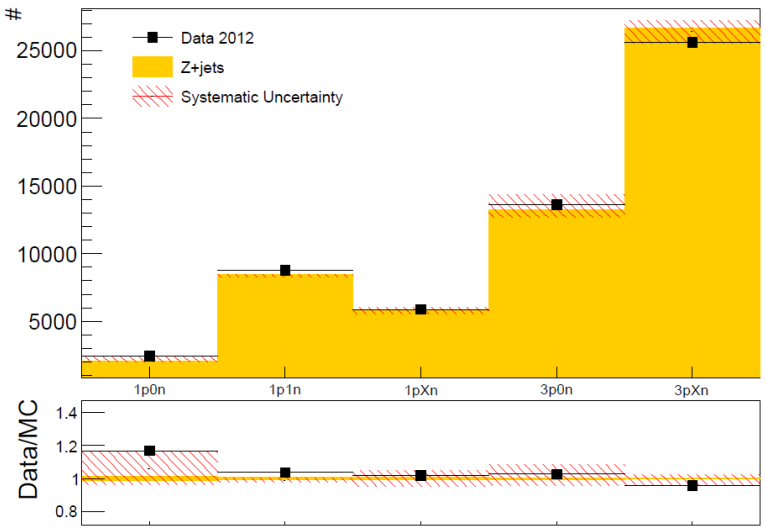


FIGURE 6.32: *Abundances of the five decay modes of fake taus as obtained in $Z \to \mu\mu$ + Jets events from simulations (yellow) and data (black). The red shaded area is the systematic uncertainty, which is obtained by varying the energy of neutral PFOs by 20% and then subsequently adjusting the decay mode to account for neutral PFOs passing or failing the E_T cut (c.f. Table 5.1). Data and simulations agree within the expected uncertainties. Figure taken from [52].*

CHAPTER 7

Application: Polarisation Studies in $Z \to \tau\tau$

This final chapter presents a possible application of the new substructure based tau reconstruction method. It discusses a short simulation-based study of how well the tau polarisation in $Z \to \tau\tau$ decays can be measured with the new tau reconstruction. The outline is as follows.
Section 7.1 motivates, why a polarisation measurement is of interest. In Sections 7.2 and 7.3, the setup of the analysis and the event- and object-selection are described. Section 7.4 presents the polarisation sensitive variable, Υ, that is used to measure the polarisation. The behaviour of Υ under the event selection is discussed in detail in Section 7.5. Finally, Section 7.6 describes how the actual measurement of the polarisation is performed. A short summary is given in Section 7.7.

7.1 Motivation

Discovering the Higgs boson in summer 2012 opened the doors for new types of physics analyses, namely those that try to measure Higgs boson properties. These kind of analyses require a clean sample of Higgs bosons to study, so that all other processes, for instance Z boson production, are considered background processes. When investigating the channel of a Higgs decaying into two tau leptons[1], one of the few things that separates the Z boson background from the Higgs signal is the invariant mass of the di-tau system. For Higgs decays, the invariant di-tau mass is a bit higher than for Z bosons, because of the Higgs mass of $m_H \approx 125$ GeV ($m_Z \approx 91$ GeV). However, the Z boson is a spin-1 particle while the Higgs boson is a spin-0 particle.

Thus, being able to measure the polarisation in tau decays allows for additional separation between Z- and Higgs bosons.

In Section 2.3, a method to gain access to the tau polarisation is described. The method makes use of the energy asymmetries in the tau decay products. Chapters 4 and 5 introduced the concepts of a new tau reconstruction and discussed the `PanTau`-algorithm in detail. Using the new approach to tau reconstruction and the `PanTau` algorithm, it is possible to follow the method described in 2.3 and study whether a measurement of the polarisation is indeed possible.

It should be noted that the studies presented in the following are by no means a full-fledged analysis. The study should be seen as a proof of principle, that the substructure based tau reconstruction allows for a polarisation measurement. It was performed to gain insight on these two subjects:

- How well does the substructure based tau reconstruction retain polarisation sensitive variables? How are they affected by a typical event & object selection for $Z \to \tau\tau$ events?
- Is it possible to measure the mean tau polarisation by using the benefits of the new tau reconstruction method?

7.2 Analysis Setup

To measure the mean tau polarisation, it is inevitable to have access to the spin of the tau leptons. While access to spin information of a single tau is not possible in data, it is indeed possible to access the information in simulation. However, in the standard ATLAS simulation procedure, the spin information of the tau is not saved. It can be reconstructed by using an external program package called `TauSpinner`.

A detailed description of how `TauSpinner` works is given in [59], a short description following [60] follows. Given that the origin of the tau lepton, i.e. the particle that produced the tau in its decay, is known, `TauSpinner` can recalculate the longitudinal tau polarisation on a statistical basis. Inputs for this calculation are the four-momenta of the mother particle (i.e. a Z boson in this study) and the four momenta of all tau decay products (including the neutrino). By combining the information from all four-momenta, weights for the two possible Z-decay spin configurations (c.f. Figure 2.4) $Z \to \tau_R^- \tau_L^+$ and $Z \to \tau_L^- \tau_R^+$ are calculated. These can be used to calculate a probability for the two configurations, for instance $p(\tau_L^-, \tau_R^+)$ for $Z \to \tau_L^- \tau_R^+$. At this point, `TauSpinner` creates a uniformly distributed random number r between 0 and 1 and checks whether r is smaller than $p(\tau_L^-, \tau_R^+)$. If it is, the decay is assumed to have happened via this configuration and the helicities are set accordingly. If r is larger than $p(\tau_L^-, \tau_R^+)$, the other configuration is chosen.

[1] One reason to do this is given in [5].

Thus, in order to be able to study how a reconstructed tau is affected by its spin, it is matched to a true tau. Its helicity is set to the helicity of the matched true tau, which had its helicity calculated by `TauSpinner`. The differences in reconstruction between left- and right-handed taus are of major interest in this study (and this information is only available easily for true taus). Thus, a match between true and reconstructed taus is the first cut that is applied in the selection.

Another point worth mentioning beforehand is that no data is used in this study, it is based on simulation only. To emulate the process of measuring the polarisation, the simulated set of $Z \to \tau\tau$ events is divided into two parts. The first part is considered to be simulated, i.e. in this part the information of true taus is used in histograms and calculations. The second part is treated as if it was data, i.e. here no separation between left- and right-handed taus is performed. However, the same event- & object-selection as in the first part is deployed[2]. Thus, this study only takes signal events into account. Effects due to jets misreconstructed as taus in a true $Z \to \tau\tau$ event, or effects due to a non-pure data sample in which the measurement is to be performed are not accounted for.

The simulated events that this study is based on are the ones labelled *evaluation sample* in Table A.1. Thus, the study is based on the same events as the performance evaluation of the `PanTau` algorithm presented in Chapter 6. The same `PanTau` configuration as discussed in the last two chapters is used.

In summary, this study should be considered a proof of principle of extracting the mean tau polarisation by using the new substructure based tau reconstruction in an ideal scenario.

7.3 Event & Object Selection

The event and object selection is a very basic form of a typical tag-and-probe selection to isolate $Z \to \tau_\mu\tau_{\text{had}}$ events. In this kind of selection, the leptonic tau is used to trigger the events, because muons have a very clean signature. The other tau is required to decay hadronically, because only hadronic tau decays provide access to polarisation sensitive variables which are reconstructable in data. In principle it is possible to increase the number of hadronic taus in the sample by requiring both taus to decay hadronically. However, it is very difficult to trigger $Z \to \tau_{\text{had}}\tau_{\text{had}}$ events and to distinguish them from di-jet events, because the cross section of processes leading to a di-jet signature is much larger (c.f. Figure 3.2).

The list of cuts can be divided into two parts. While the first part places cuts on the hadronic tau from $Z \to \tau_\mu\tau_{\text{had}}$, the second part selects the leptonically decaying tau. A graphical representation of the cut flow is shown in Figure 7.1. That figure shows the cut flow for the set of all true taus in Figure 7.1a and for the set of all true taus, matched to a reconstructed tau classified as 1p1n 7.1b. A more detailed discussion of the cut flow is given after the explanation of the selection cuts.

The list of cuts dealing with the hadronic tau is as follows:

- *No cut*:
  At this cut stage, no selection is performed. However, when filling histograms or calculating properties that require truth matched reconstructed taus at this cut stage, the requirement of the reconstructed tau to be truth matched is applied regardless.

[2] It therefore also applies the match between true and reconstructed taus. Thus, the assumption is, that in a study involving data, the amount of fake taus is either negligible or can be estimated and corrected for.

In addition, a very strict true tau selection was implemented as a result of reoccuring problems when trying to access the datasets via the computing grid (c.f. Section A.1). With hindsight, it may have been better to change the technical setup of the analysis, so that it could be performed on the local computing cluster and maybe drop some of the requirements as they also veto valid tau decays[3]. The strict true tau selection is performed before the actual event selection and any event containing one or more true taus that do not pass it, are excluded from the analysis[4]. This results in reduced statistics of this analysis when compared to the analysis of the `PanTau` performance in Chapter 6.

- *Truth matching (τ_{had} match)*
  This cut performs a matching between true (i.e. generated) and reconstructed taus. A true tau from a $Z \to \tau\tau$ decay may have any η value, for example greater than 2.5. This true tau will not be reconstructed, because the substructure based tau reconstruction needs track measurements, which are only possible at $|\eta| < 2.5$ where the tracking detectors are located (c.f. Section 3.2.2). Thus, not all true taus will pass this cut.

  The event passes this cut if it contains exactly one true tau fulfilling these conditions:

  1. The true tau is matched, i.e. there is a reconstructed tau within $\Delta R < 0.2$ with respect to the true tau.
  2. The matched true tau decayed hadronically, i.e. it did not decay leptonically.
  3. The hadronically decaying matched true tau has a Z boson as mother.

  At every later cut stage, the hadronic true tau will have a match to a reconstructed tau. Thus, the hadronic tau in the event can be thought of as one single object, which has two different parts: The true, generated part and the reconstructed part. Because of that, this hadronic tau will be referred to as *the hadronic tau*. The terms *true* and *reconstructed* will only be used to indicate from which part of the hadronic tau the information is taken.

- *Transverse momentum (τ p_T):*
  The hadronic tau is required to have a reconstructed p_T larger than 20 GeV. The p_T value is obtained by using the substructure based reconstruction, as it is more precise than the default p_T value associated to the hadronic tau (c.f. Figure 6.31a). The cut is placed at 20 GeV as recommended by the tau working group in ATLAS [61].

- *Transition region veto (Crack Veto):*
  The transition region is located at $1.37 < |\eta| < 1.52$. The energy reconstruction is known to be bad in this part of the detector (c.f. Figure 6.29). Thus, if the hadronic tau falls into this η-region, it is not considered in the analysis.

- *1-prong or 3-prong (1p or 3p):*
  Because the substructure based tau reconstruction only considered 1-prong and 3-prong taus at the time of performing this analysis, only these are accepted. For this analysis, it would be sufficient to select 1-prong taus, because the 1p1n decay mode provides the best polarisation sensitive variables.

[3] Later it turned out that files, which caused crashes on the computing grid actually could be processed locally without any problems.

[4] The selection vetoes leptonic tau decays that do not have a corresponding neutrino, hadronic tau decays that do not contain a $\pi^{\pm}$, $K^{\pm}$ or $K^{*\pm}$ and tau decays with flawed entries in the Monte Carlo record.

- *Charge of 1 ($|q| = 1$):*
  In order to avoid picking up wrongly associated tracks in the case of 3-prong taus, the absolute value of the charge of the hadronic tau is forced to be one. Similarly to what is stated above, this cut does not have an effect on this analysis, because it only considers 1-prong taus.

- *Electron Veto (Ele Veto):*
  The electron veto is part of the tau identification and is explained briefly at the end of Section 4.2. The strength of the veto applied here is the *loose* veto, corresponding to a cut in the electron veto BDT at 0.3.

- *Tau Identification (Jet Score):*
  The last cut on the hadronic side of the Z boson is the tau identification cut. The tau identification is explained in Section 4.2. The hadronic tau is required to pass the *loose* tau identification, which translates into a cut on the tau identification BDT at 0.3.

The cuts that the leptonic tau needs to pass are:

- *Truth matching (μ match):*
  This cut performs the matching between a reconstructed muon and the true leptonic tau. An event passes this cut if it has exactly one true muon fulfilling these conditions:

  1. The true muon has a reconstructed μ within $\Delta R < 0.2$.
  2. The true muon is the daughter of a tau and that tau originates from a Z boson.

  Similarly to the hadronic tau described above, this tau will be referred to as the *leptonic tau*, or simply the *muon*.

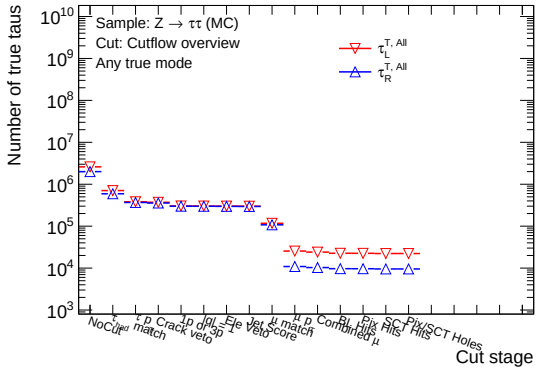
- *Transverse momentum (μ p_T):*
  The muon needs to have a p_T of at least 26 GeV and is higher than the p_T cut on the hadronic tau. As mentioned before, the muon is used to trigger the event. The muon trigger with the lowest p_T threshold that is not prescaled[5] requires a muon with a p_T of 24 GeV at trigger level. Because the efficiency of the trigger depends on the p_T of the muon, and the distribution has no sharp turn-on at 24 GeV, an additional cut is required to make sure that the trigger efficiency can be assumed to be constant[6].

- *Combined muon (Combined μ):*
  This cut makes sure that the muon was reconstructed in the inner detector and in the muon chambers, and that the two independent measurements are consistent with one another. Muons that are only reconstructed in the inner detector or in the muon system would fail this cut.

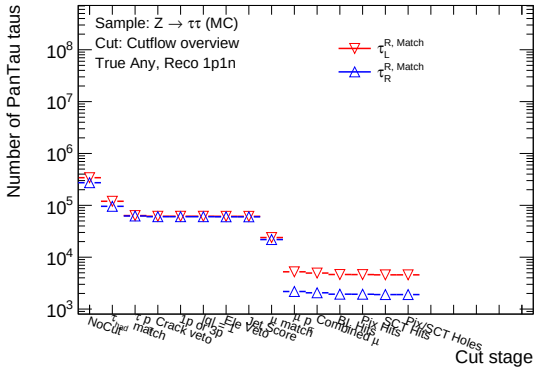
- *Number of Beam Layer Hits (BL Hits):*
  To pass this cut, the muon is required to have a hit in the beam layer.

[5] Two of the defining properties of a trigger are its threshold and its prescale. The threshold defines the p_T value that a muon needs to have in order for the trigger to fire. The prescale is the inverse of the fraction of events passing the trigger that are accepted. For example a prescale of 1000 would mean that out of 1000 events passing the trigger, only 1 event is stored on disk. This mechanism is necessary because cross sections rise with decreasing p_T and the bandwidth with which events can be written to disk is limited.

[6] It is of course possible to neglect this cut and use all triggered muons. However, this would require a muon efficiency correction which depends on the muon p_T, accounting for inefficiencies of the trigger at lower muon p_T values.



(A) *Cut-flow for true and reconstructed taus, all tau decay modes considered.*



(B) *Cut flow for matched reconstructed 1p1n taus.*

FIGURE 7.1: Left: *Cut-flow of all and matched true taus, as well as matched reconstructed taus, divided in to left-handed and right-handed taus. All decay modes are considered.* Right: *Cut-flow of matched true and matched reconstructed taus, reconstructed as 1p1n.*

- *Number of Pixel Hits (Pix Hits):*
  In addition to the hit in the beam layer, the muon also needs to have two or more hits in the pixel detector.
- *Number of SCT Hits (SCT Hits):*
  The muon track in the inner detector has to have 6 or more hits in the SCT sub-detector.
- *Number of Pixel and SCT holes (Pix/SCT Holes):*
  The sum of holes in the pixel and SCT detector has to be two or smaller.

The cut flow consisting of the cuts described above is visualised in Figure 7.1. Figure 7.1a shows the amount of left- and right-handed true taus of any decay mode. The amount of left- and right-handed hadronic taus that are classified as 1p1n in the reconstruction is shown in 7.1b. This subset of taus is of major importance for the polarisation measurement. There is no selection cut in the cut flow itself, that requires the hadronic tau to be of a specific decay mode, because of technical reasons[7]. The separation into left- and right-handed taus will become important later on in the extraction of the mean tau polarisation. Table 7.1 lists the actual numbers of taus after each cut stage.

What can be seen from the figures is that the main cuts, i.e. cuts that significantly change the number of taus, are the following ones.

- *Truth matching:*
  The truth matching has a rather large impact on the number of taus, because it implicitly fixes the type of decays in $Z \to \tau\tau$.
- *Transverse momentum:*
  For the hadronic tau, this cut is less severe than for the leptonic tau. If the Z boson is at rest, both

[7] The main reason is to not limit the available information to one decay mode. Without placing a cut on the decay mode, all combination of true and reconstructed modes are available and can be studied at each cut stage. This greatly simplifies the setup of the C++ code that is used to perform the analysis.

Cut	True taus passing cuts			Truth matched 1p1n taus passing cuts		
	τ^{True}	$\tau_{\text{L}}^{\text{True}}$	$\tau_{\text{R}}^{\text{True}}$	$\tau^{\text{Matched 1p1n}}$	$\tau_{\text{L}}^{\text{Matched 1p1n}}$	$\tau_{\text{R}}^{\text{Matched 1p1n}}$
No cuts	$4.59 \cdot 10^6$	$2.59 \cdot 10^6$	$1.99 \cdot 10^6$	$614 \cdot 10^3$	$341 \cdot 10^3$	$273 \cdot 10^3$
τ^{reco} matched to τ_{had}^{true}	$1.29 \cdot 10^6$	$705 \cdot 10^3$	$592 \cdot 10^3$	$215 \cdot 10^3$	$119 \cdot 10^3$	$95.7 \cdot 10^3$
$p_T(\tau^{\text{reco}}) > 20$ GeV	$748 \cdot 10^3$	$381 \cdot 10^3$	$366 \cdot 10^3$	$125 \cdot 10^3$	$63.4 \cdot 10^3$	$62.3 \cdot 10^3$
$1.37 > \lvert\eta(\tau^{\text{reco}})\rvert > 1.52$	$727 \cdot 10^3$	$371 \cdot 10^3$	$356 \cdot 10^3$	$121 \cdot 10^3$	$61.5 \cdot 10^3$	$60.3 \cdot 10^3$
$n_{\text{Track}}(\tau^{\text{reco}}) = 1$ or 3	$602 \cdot 10^3$	$303 \cdot 10^3$	$299 \cdot 10^3$	$121 \cdot 10^3$	$61.5 \cdot 10^3$	$60.3 \cdot 10^3$
$\lvert\text{Charge}(\tau^{\text{reco}})\rvert = 1$	$600 \cdot 10^3$	$302 \cdot 10^3$	$298 \cdot 10^3$	$121 \cdot 10^3$	$61.5 \cdot 10^3$	$60.3 \cdot 10^3$
BDT Ele. Score > 0.3	$597 \cdot 10^3$	$301 \cdot 10^3$	$296 \cdot 10^3$	$121 \cdot 10^3$	$61.4 \cdot 10^3$	$60.1 \cdot 10^3$
BDT Jet Score > 0.3	$594 \cdot 10^3$	$299 \cdot 10^3$	$294 \cdot 10^3$	$121 \cdot 10^3$	$61.3 \cdot 10^3$	$60 \cdot 10^3$
μ^{reco} matched to $\mu_{\text{from }\tau}^{\text{true}}$	$224 \cdot 10^3$	$116 \cdot 10^3$	$107 \cdot 10^3$	$45.8 \cdot 10^3$	$23.9 \cdot 10^3$	$21.8 \cdot 10^3$
$p_T(\mu^{\text{reco}}) > 26$ GeV	$36.3 \cdot 10^3$	$25.4 \cdot 10^3$	$10.8 \cdot 10^3$	$7.4 \cdot 10^3$	$5.22 \cdot 10^3$	$2.18 \cdot 10^3$
μ is combined	$34.3 \cdot 10^3$	$24 \cdot 10^3$	$10.3 \cdot 10^3$	$7 \cdot 10^3$	$4.94 \cdot 10^3$	$2.05 \cdot 10^3$
N(BL Hits) ≥ 1	$32.2 \cdot 10^3$	$22.5 \cdot 10^3$	$9.68 \cdot 10^3$	$6.57 \cdot 10^3$	$4.63 \cdot 10^3$	$1.93 \cdot 10^3$
N(Pix Hits) ≥ 2	$32.1 \cdot 10^3$	$22.5 \cdot 10^3$	$9.65 \cdot 10^3$	$6.55 \cdot 10^3$	$4.62 \cdot 10^3$	$1.93 \cdot 10^3$
N(SCT Hits) ≥ 6	$31.7 \cdot 10^3$	$22.2 \cdot 10^3$	$9.53 \cdot 10^3$	$6.47 \cdot 10^3$	$4.56 \cdot 10^3$	$1.9 \cdot 10^3$
N(SCT Holes) ≤ 2	$31.7 \cdot 10^3$	$22.2 \cdot 10^3$	$9.53 \cdot 10^3$	$6.47 \cdot 10^3$	$4.56 \cdot 10^3$	$1.9 \cdot 10^3$

TABLE 7.1: *Number of taus that pass the successive selection cuts. The first column lists the cuts, which are explained in the text. The following two triplets of columns show the numbers of all true taus and all reconstructed truth matched taus classified as 1p1n.*

Cut	Efficiency times acceptance	
	$\tau_{\text{L}}^{\text{Matched 1p1n}}$	$\tau_{\text{R}}^{\text{Matched 1p1n}}$
No cuts	$131 \cdot 10^{-3} \pm 239 \cdot 10^{-6}$	$136 \cdot 10^{-3} \pm 278 \cdot 10^{-6}$
τ^{reco} matched to τ_{had}^{true}	$46.1 \cdot 10^{-3} \pm 136 \cdot 10^{-6}$	$47.8 \cdot 10^{-3} \pm 158 \cdot 10^{-6}$
$p_T(\tau^{\text{reco}}) > 20$ GeV	$24.4 \cdot 10^{-3} \pm 98.1 \cdot 10^{-6}$	$31.1 \cdot 10^{-3} \pm 126 \cdot 10^{-6}$
$1.37 > \lvert\eta(\tau^{\text{reco}})\rvert > 1.52$	$23.7 \cdot 10^{-3} \pm 96.6 \cdot 10^{-6}$	$30.1 \cdot 10^{-3} \pm 124 \cdot 10^{-6}$
$n_{\text{Track}}(\tau^{\text{reco}}) = 1$ or 3	$23.7 \cdot 10^{-3} \pm 96.6 \cdot 10^{-6}$	$30.1 \cdot 10^{-3} \pm 124 \cdot 10^{-6}$
$\lvert\text{Charge}(\tau^{\text{reco}})\rvert = 1$	$23.7 \cdot 10^{-3} \pm 96.6 \cdot 10^{-6}$	$30.1 \cdot 10^{-3} \pm 124 \cdot 10^{-6}$
BDT Ele. Score > 0.3	$23.6 \cdot 10^{-3} \pm 96.5 \cdot 10^{-6}$	$30 \cdot 10^{-3} \pm 124 \cdot 10^{-6}$
BDT Jet Score > 0.3	$23.6 \cdot 10^{-3} \pm 96.5 \cdot 10^{-6}$	$30 \cdot 10^{-3} \pm 124 \cdot 10^{-6}$
μ^{reco} matched to $\mu_{\text{from }\tau}^{\text{true}}$	$9.21 \cdot 10^{-3} \pm 59.8 \cdot 10^{-6}$	$10.9 \cdot 10^{-3} \pm 74.3 \cdot 10^{-6}$
$p_T(\mu^{\text{reco}}) > 26$ GeV	$2.01 \cdot 10^{-3} \pm 27.8 \cdot 10^{-6}$	$1.09 \cdot 10^{-3} \pm 23.3 \cdot 10^{-6}$
μ is combined	$1.9 \cdot 10^{-3} \pm 27.1 \cdot 10^{-6}$	$1.02 \cdot 10^{-3} \pm 22.7 \cdot 10^{-6}$
N(BL Hits) ≥ 1	$1.78 \cdot 10^{-3} \pm 26.2 \cdot 10^{-6}$	$968 \cdot 10^{-6} \pm 22 \cdot 10^{-6}$
N(Pix Hits) ≥ 2	$1.77 \cdot 10^{-3} \pm 26.1 \cdot 10^{-6}$	$965 \cdot 10^{-6} \pm 21.9 \cdot 10^{-6}$
N(SCT Hits) ≥ 6	$1.75 \cdot 10^{-3} \pm 26 \cdot 10^{-6}$	$951 \cdot 10^{-6} \pm 21.8 \cdot 10^{-6}$
N(SCT Holes) ≤ 2	$1.75 \cdot 10^{-3} \pm 26 \cdot 10^{-6}$	$951 \cdot 10^{-6} \pm 21.8 \cdot 10^{-6}$

TABLE 7.2: *Efficiency times acceptance for left- and right-handed taus for each of the cut stages explained in the text. The numbers for left-handed (right-handed) taus in this table are obtained by dividing $\tau_L^{Matched\ 1p1n}$ ($\tau_R^{Matched\ 1p1n}$) by τ_L^{True} (τ_R^{True}) from Table 7.1.*

taus receive an energy of roughly 45 GeV. Neglecting the mass, the p_T is then less than 45 GeV for any tau not emitted with $\eta = 0$. In case of the hadronic tau, part of the p_T is distributed to the ν_τ, making the tau softer than $p_T = 20$ GeV in some cases. About 54% of the left-handed hadronic taus pass this cut, for right-handed hadronic taus, the fraction is 62%, due to the different energy fractions that are passed to the ρ in the tau decay, see Figure 7.2. Thus, this cut actually biases the sample that is studied towards right-handed taus. When extracting the polarisation, this effect needs to be accounted for.

For the leptonic tau, there are two neutrinos involved. Thus, on average, more energy is distributed to neutrinos than for the hadronic tau, reducing the p_T further. In addition, the p_T cut for the leptonic tau is $p_T(\mu) > 26$ GeV. Here, the cut has an even larger impact on the relative abundances of left- and right-handed taus. Its efficiency for left-handed taus is 20%, twice as large as for right-handed taus. This is because of the different energy fractions carried by a muon in left- and right-handed taus, see Figure 7.2. For a decaying left-handed tau the fraction of the tau energy that it carries is larger than in the case of a right-handed tau decay. Thus, requiring a large muon momentum shifts the sample towards left-handed taus. Like above, this effect needs to be corrected for when calculating the polarisation.

In order to be able to perform the correction for different cut efficiencies for left- and right-handed taus, Table 7.2 shows the values of efficiency times acceptance for left- and right-handed taus classified as 1p1n. In this analysis, the acceptance a and efficiency ϵ at cut stage i for taus of helicity h are defined by the number of taus N:

$$a_i = \frac{N(\text{True } \tau_{\text{had}} \text{ after cut i})}{N(\text{True } \tau_{\text{had}} \text{ after cut 0})}$$
$$\epsilon_i = \frac{N(\text{Reconstructed } \tau_{\text{had}} \text{ classified as 1p1n after cut i})}{N(\text{True } \tau_{\text{had}} \text{ after cut i})}$$

where *cut 0* corresponds to the first cut stage, at which no selection cut is applied. Acceptances and efficiencies for left-handed and right-handed taus are defined the same way, by replacing the number of taus with the number of left- and right-handed taus respectively. The product $a_i \cdot \epsilon_i$ simplifies to

$$a_i \cdot \epsilon_i = \frac{N(\text{Reconstructed } \tau_{\text{had}} \text{ classified as 1p1n after cut i})}{N(\text{True } \tau_{\text{had}} \text{ after cut 0})}$$

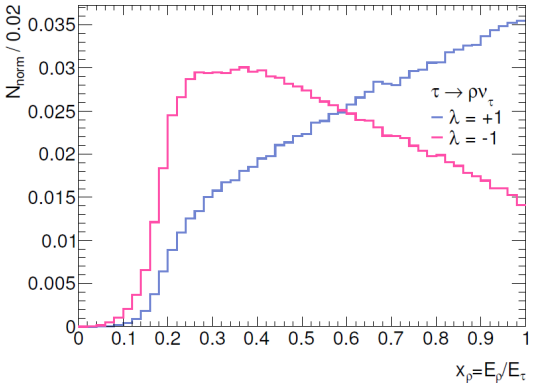
The efficiency only takes into account the efficiency of the process to reconstruct a true tau and then classify it as a 1p1n. Anything else that reduces the number of taus (i.e. the selection cuts) is put into the acceptance, because the analysis deals with truth matched taus only. Finally, it is only the product $a_i \cdot \epsilon_i$ that matters in this analysis (see section 7.6).

The changes in the acceptance throughout the cut flow is the same as in Figure 7.1a, because of its definition[8]. The product $a_i \cdot \epsilon_i$ and the efficiency itself on the other hand are shown as a function of the selection cuts in Figure 7.3. While the efficiencies for left- and right-handed taus are very close to each other, the situation is different in the acceptance.

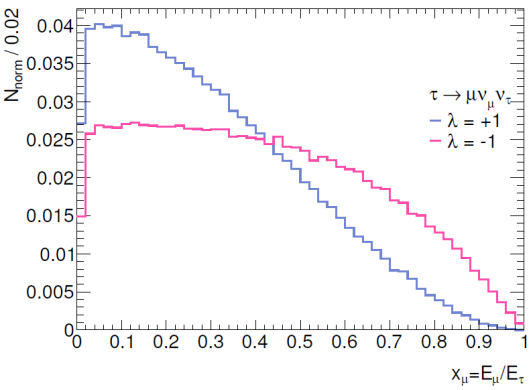
How these numbers are used to correct the polarisation measurement is explained in the corresponding section, Section 7.6.

Moving away from the numbers of taus to the number of events, there are 15,541 events that pass all

[8] The plot showing the acceptance as a function of the cuts will look the same as Figure 7.1a, the only difference is that the values are divided by the number of true taus at the first cut.

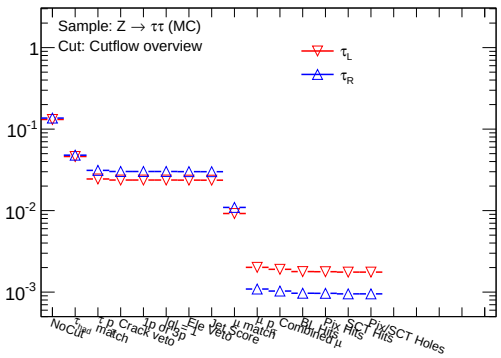


(A) *Fraction of the tau energy that is passed to the ρ meson. It is higher for right-handed taus.*

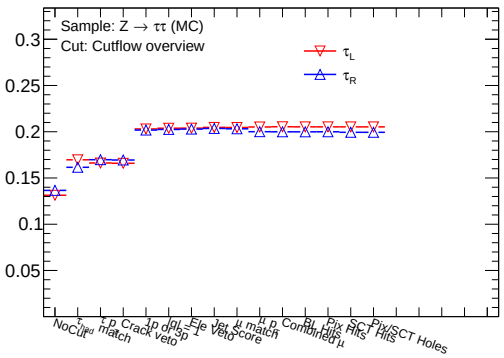


(B) *Fraction of the tau energy that is passed to the muon. It is higher for left-handed taus.*

FIGURE 7.2: *Fraction of the tau energy carried by the muon (left) and ρ meson (right) in tau decays for left-handed (red) and right-handed (blue) taus (both figures taken from [23]).*
Left: *When decaying into a ρ meson, right-handed taus produce higher energetic ρ mesons. This shifts the sample towards right-handed taus.*
Right: *In left-handed tau decays, the muon carries a larger fraction of the tau energy on average, so that a requirement of large muon momenta will preferable select left-handed taus.*



(A) *Acceptance times efficiency for every cut stage and left- and right-handed taus.*



(B) *Efficiency to reconstruct a true tau and classify it as 1p1n.*

FIGURE 7.3: Left: *Acceptance times efficiency to correctly reconstruct a truth matched 1p1n tau. These numbers (also given in Table 7.2) are used to correct for different cut efficiencies for left- and right-handed taus.*
Right: *The efficiency to correctly reconstruct a 1p1n tau decay.*

selection cuts. The total size of the pseudo-data sample of $Z \to \tau\tau$ events is $N = 2,242,486$, so that the event selection efficiency is $6.93 \cdot 10^{-3}$. The cross section of $Z \to \tau\tau$ at the LHC was measured to be $\sigma(Z \to \tau\tau, 66 < m_{\text{inv}} < 116\,\text{GeV}) = 0.92\,\text{nb}$ [62]. Thus, the sample corresponds to an integrated luminosity of

$$L_{\text{int}} = N/\sigma(Z \to \tau\tau, 66 < m_{\text{inv}} < 116\,\text{GeV}) = 2.44\,\text{fb}^{-1}$$

This is much less than the LHC delivered and the ATLAS detector recorded. Thus, the statistical uncertainties that are attached to results from the pseudo-data part of the sample can be expected to be smaller in an actual measurement. Since it is one sample that is divided into two parts, increasing the pseudo-data part would result in a smaller part that is used for simulation studies. Moving events from the simulation into the pseudo-data part would have a negative impact on the templates that are used in the fit to extract the polarisation (c.f. Section 7.6), so that the splitting of roughly 50% pseudo-data and 50% simulation is kept. However, since the recorded data corresponds to roughly 20 fb^{-1}, a sample four times as large as the one available would be required to reproduce the statistics that are available in data. Another sample would then be required to perform the simulation based studies.

7.4 Polarisation Sensitive Variable

As pointed out in Section 2.3, in particular Subsection 2.3.5, the charged energy asymmetry, Υ, is sensitive to the tau polarisation. Υ can be calculated in the 1p1n and 1pXn decay modes. Taking into account the efficiencies and purities of the 1p1n and 1pXn modes given in Figure 6.3, the best results should be achievable in the 1p1n mode.

Figure 7.4 shows the distribution of the true charged energy asymmetry Υ for all true taus decaying into 1p1n. The charged energy asymmetry is defined as

$$\Upsilon = \frac{E(\pi^{\pm}) - E(\pi^0)}{E(\pi^{\pm}) + E(\pi^0)}$$

Thus, large positive values of Υ correspond to decays in which the charged (neutral) pion is hard (soft), and large negative values of Υ correspond to decays with a soft $\pi^{\pm}$ and a hard π^0.

Figure 7.5 shows the distribution of Υ from true 1p1n decays of left-handed (Fig. 7.5a) and right-handed (Fig. 7.5b) taus for different cut stages. Not all cut stages are overlayed due to visual reasons. The cut stages shown are the very first stage, at which no selection is performed. The other cut stages shown are the truth match and the last object selection cut of the hadronic and leptonic tau selection.

The shape of the Υ distribution for left-handed taus does not change significantly. For right-handed taus, the M-like shape of the distribution becomes a bit more pronounced. There are less entries at $\Upsilon = 0$ and more entries at $|\Upsilon| > 0.6$ at the later cut stages. A feature that is visible in both distributions is that the statistics are reduced significantly at the final event selection stage compared to the other selection stages. This is mainly because of the high $p_T(\mu)$ cut of 26 GeV, c.f. Table 7.1.

In summary, the shapes of the true Υ distributions are not altered significantly by the selection cuts. The sensitivity to the tau polarisation in a measurement is given by the very different shapes of Υ for left-handed and right-handed taus. Thus, it is important to understand the shape differences that occur when calculating Υ in reconstructed tau leptons. The next section discusses the impact of the selection cuts for the reconstructed version of the Υ-variable.

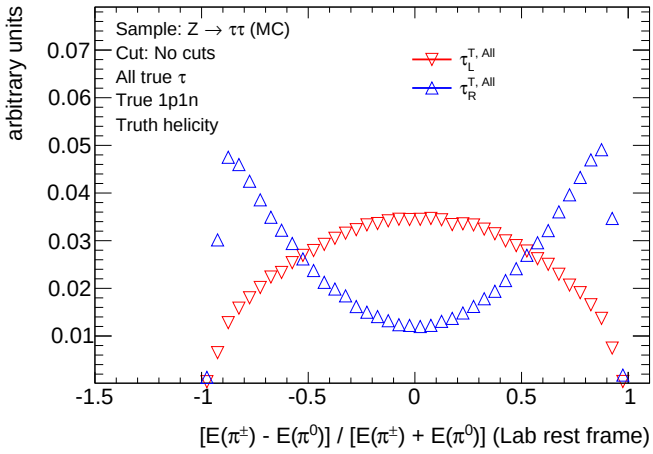
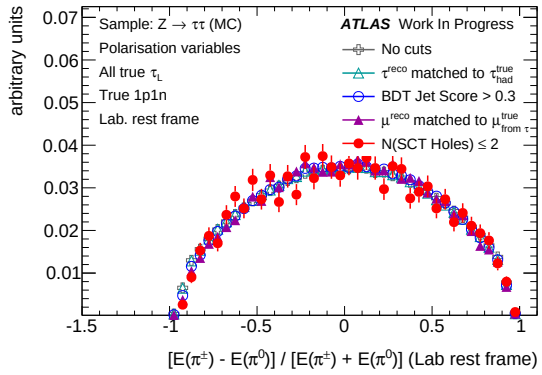
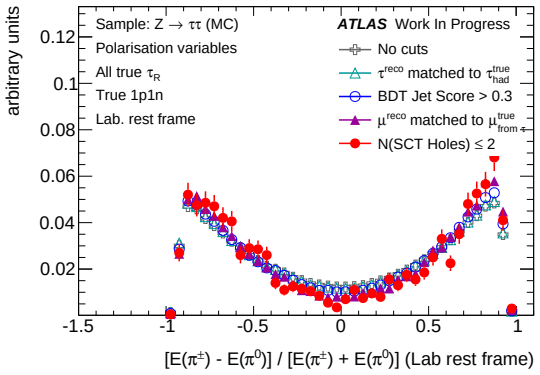


FIGURE 7.4: *Distribution of the charged energy asymmetry ϒ as obtained from the set of all true taus generated in the $Z \rightarrow \tau\tau$ sample, decaying into 1p1n.*



(A) *Changes in the shape for left-handed taus.*



(B) *Changes in the shape for right-handed taus.*

FIGURE 7.5: *Shapes of the ϒ distribution for all generated taus decaying into 1p1n. Each plot shows an overlay of the shapes as obtained at the different cut stages. The left (right) plot shows the distribution for left-handed (right-handed) taus.*

7.5 Reconstruction-Induced Effects in the ϒ-Shape

This section focuses on discussing the changes in the ϒ distribution when moving from the set of all true 1p1n taus to the reconstructed 1p1n taus, that pass all selection cuts. The steps, which move from the true taus to the reconstructed 1p1n taus step-by-step are the following ones:

- Start with the set of all true 1p1n taus and move to the set of true taus that are also reconstructed (without requirements on the reconstructed mode).
- Require that the reconstructed true 1p1n taus are also of mode 1p1n.
- Make the calculation of ϒ use the reconstructed pions, and not the generated ones.
- Drop the requirement of the true mode being 1p1n, i.e. look at reconstructed 1p1n taus, that are matched to a true tau of any mode.

- Apply the event selection to the set of reconstructed 1p1n taus that have a truth match.

Each of these steps will be described in its own subsection.

7.5.1 From true 1p1n Taus to reconstructed true 1p1n Taus

In Figure 7.6, the Υ variable is calculated for all true taus (red) and for matched true taus (blue). There are smaller deviations for the left-handed taus at Υ values near one. True taus with this charged energy asymmetry are slightly less likely to be matched to a reconstructed tau. For right handed taus, a similar sized inefficiency can be found of Υ values of zero. However, these are relatively small effects and are not followed up further.

7.5.2 Requiring the true 1p1n Taus to be reconstructed as 1p1n

A larger effect is visible in the distributions shown in Figure 7.7. The blue triangles in this figures are the same as the blue triangles from Figure 7.6. In this set of matched true 1p1n taus, Υ is still calculated by using the true $\pi^{\pm}$ and π^0.

Now it is required that the reconstructed tau that is matched to the true tau is also reconstructed as 1p1n, while still calculating Υ with the true π^0 and $\pi^{\pm}$ (red triangles). A dent in both distributions 7.7a and 7.7b at large Υ values is clearly visible.

Figure 7.8 shows the contributions of all reconstructed modes to the Υ shape. The true 1p1n taus that are misclassified as 1p0n are located at high Υ values and are responsible for the change in shape of both τ_{Left} and τ_{Right}. The 1p1n→1pXn misclassified τ_{Left} have roughly the same shape as the correctly classified ones, thus they do not alter the shape much. τ_{Right} misclassified like 1p1n→1pXn are distributed equally over all possible Υ values, with a tendency towards small Υ values.

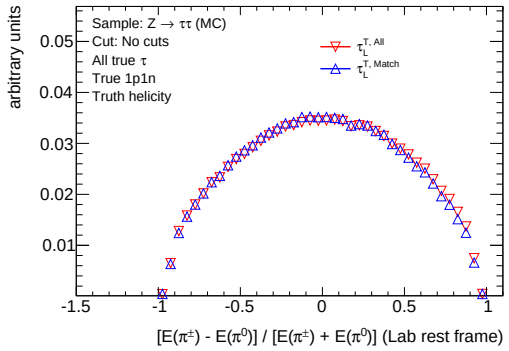
The reason why these misclassifications show up in the observed regions of the Υ distribution can be understood by recalling the definition of Υ:

$$\Upsilon = \frac{E(\pi^{\pm}) - E(\pi^0)}{E(\pi^{\pm}) + E(\pi^0)}$$

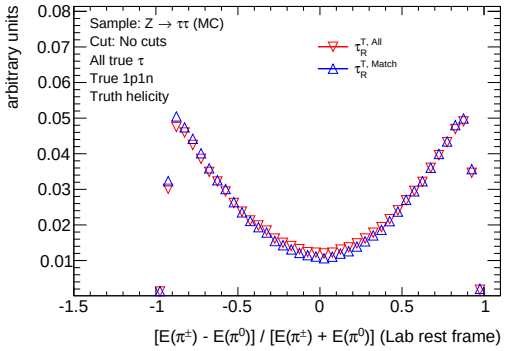
In order to create large (small) Υ values, the neutral pion (charged pion) needs to be of very low energy. For example, to enforce $\Upsilon \geq 0.9$ with a charged pion of $E_T = 20$ GeV, the neutral pion needs to have $E_T \leq 1.1$ GeV, which is below the E_T threshold applied in the π^0 identification. Thus, a true 1p1n tau with this configuration can not be classified correctly, because the neutral pion fails a basic selection criterion and therefore the tau appears as a 1p0n in the reconstruction. Even if the π^0 has an energy above the threshold, say $E_T = 5$ GeV for example, the charged pion needs to have $E_T \geq 40$ GeV to achieve $\Upsilon \geq 0.8$, making the π^0 very soft in comparison to the $\pi^{\pm}$. Because the energy difference is so large, it is more likely to be reconstructed with a wrong energy, or not reconstructed at all, due to fluctuations in the charged pion shower.

These two effects are also visualised in Figure 7.9. Only left- (red) and right-handed (blue) hadronic 1p1n taus, classified as 1p0n with $\Upsilon > 0.7$ enter the distributions shown there, hence the $\pi^{\pm}$ always has a higher energy than the π^0.

Part 7.9a shows the difference in E_T between the charged and neutral pion. For right handed taus, the energy difference peaks at around 15 GeV, but has a rather large tail, reaching out to approximately

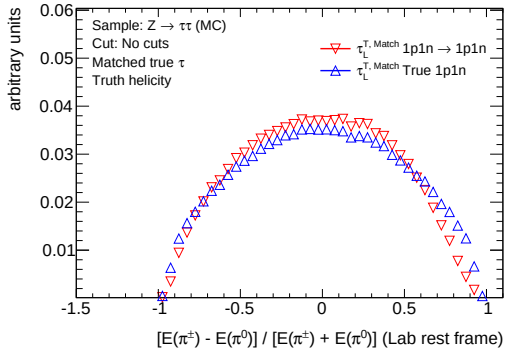


(A) *Left-handed taus.*

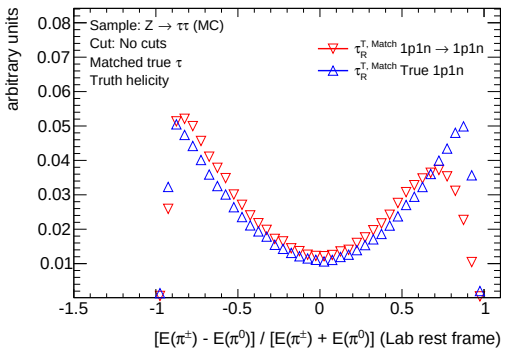


(B) *Right-handed taus.*

FIGURE 7.6: *ϒ distributions for all true 1p1n taus (red) and all true 1p1n taus that are matched to reconstructed taus (blue).*

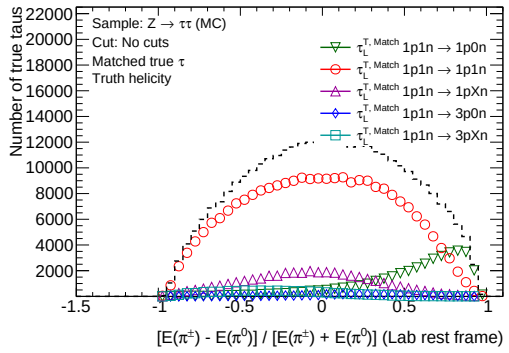


(A) *Left-handed taus.*

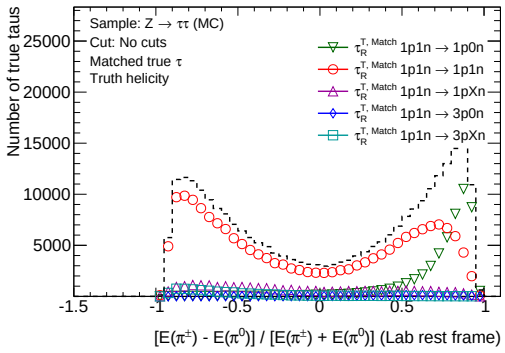


(B) *Right-handed taus.*

FIGURE 7.7: Blue: *Shows the same taus as the blue triangles in Figure* **??**, *i.e. ϒ calculated from true 1p1n taus, that are matched to a reconstructed tau.* Red: *The same set of taus as used in the blue distribution, with the additional requirement that the matched reconstructed tau was reconstructed with the 1p1n decay mode. The changes in the distribution are discussed in the text and in Figures 7.8 and 7.9.*



(A) *Composition of left-handed true 1p1n taus.*



(B) *Composition of right-handed true 1p1n taus.*

FIGURE 7.8: *The ϒ distribution of matched true taus, decomposed into the different reconstructed modes. Relative amounts correspond to the 1p1n column in 6.3c. The true 1p1n taus reconstructed as 1p0n taus are located at ϒ values above 0.6 indicating a small π^0 energy, making a correct classification very difficult. It is noted that there are also true taus that are reconstructed as 2-prong taus, which are not considered in this version of substructure based tau reconstruction. More details in the text and in Figure 7.9.*

40 GeV. Thus, for a large fraction of these true 1p1n taus with $\Upsilon > 0.7$, the energy difference between the $\pi^\pm$ and the π^0 is large, making the π^0 susceptible to misreconstruction in terms of E_T estimation.

The second part, Fig. 7.9b shows the E_T value of the π^0 in those taus. For τ_{Right}, it peaks at 1.5 GeV, so that the majority of the π^0 are below the preselection cut of $\approx$ 1.6 GeV (c.f. Table 5.1). Because of the more steeply falling tail, the probability of a π^0 to have an E_T just slightly above the E_T threshold is large. Combined with the difficulties induced by the large E_T differences, a misreconstruction is likely to put the π^0 below the E_T threshold.

The third part, Figure 7.9c, shows the E_T difference as a function of the $E_T(\pi^0)$. The sharp edge is because of the $\Upsilon > 0.7$ requirement, which connects $E(\pi^\pm)$ with $E(\pi^0)$[9] as discussed above. In this figure, every entry below $E_T(\pi^0) \approx 1.6$ GeV is rejected because of the preselection in π^0-ID. The remaining entries extend to large values of $E_T(\pi^\pm) - E_T(\pi^0)$, so that a correct reconstruction of the π^0 is difficult.

For τ_{Left}, the effects are the same, but they are less pronounced, because the tail in the E_T difference (7.9a) is smaller and the distribution itself peaks at smaller values. In addition, the tail in Figure 7.9b is larger, so that a misreconstruction is not as likely to put the π^0 below the E_T threshold as it is the case for τ_{Right}.

7.5.3 Calculating Υ with reconstructed Pions

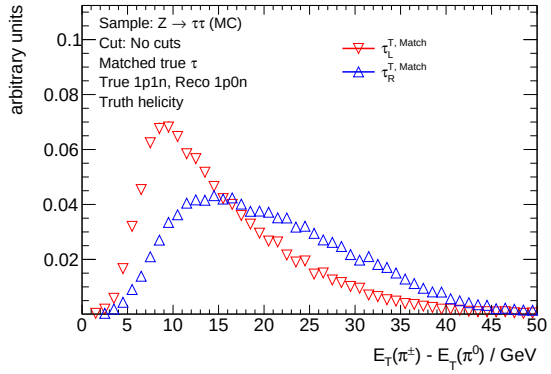
The next change to be discussed is the differences in the Υ shape when calculating Υ with the energies of reconstructed pions, instead of true pions. Figure 7.10 shows how the Υ distributions looks like for every combination of true and reconstructed pions. The magenta stars in this figure show the same distribution as the red triangles in Figure 7.7. The figure shows that there barely is a change when switching from true $\pi^\pm$ to reconstructed $\pi^\pm$ (magenta stars and blue triangles). However, when using the reconstructed π^0 instead of the true π^0 (magenta stars and red triangles), the change is more pronounced and has the same effect as the one observed in Figure 7.7. This behaviour is expected, because the momentum resolution of the charged pion in the inner detector is far better than the energy resolution of the neutral pion (see Equations 4.2 & 4.3 and Figure 6.9).

7.5.4 From correctly reconstructed 1p1n to all reconstructed 1p1n Taus

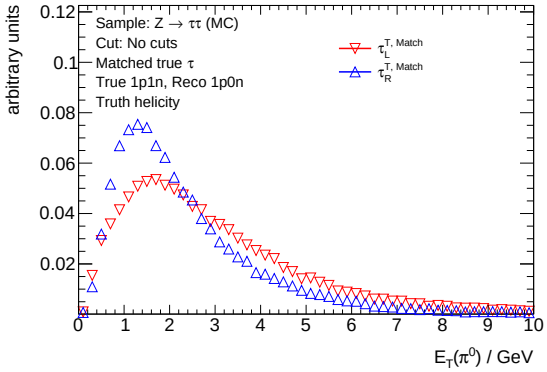
The final step is to move from the set of correctly reconstructed 1p1n taus to all reconstructed 1p1n taus, as they are available in data. The reconstructed 1p1n taus produce the Υ distributions shown in Figure 7.11. In that figure, the blue triangles are the same distribution as the black dots from Figure 7.10. The red triangles correspond to Υ distributions obtained from the (larger) set of all true taus, reconstructed as 1p1n.

The change for left-handed taus is such that the distribution is broader than the one for correctly classified 1p1n taus. For right-handed taus, the left peak is less pronounced when considering all true modes that are reconstructed as 1p1n. The contributions of the other true modes, mainly 1p0n and 1pXn, are shown in Figure 7.12. The black dashes (red circles) in this figure correspond to the red (blue) triangles in Figure 7.11. For both helicities, the misclassification 1pXn→1p1n (1p0n→1p1n) can be found at negative values and values around zero (at large positive values).

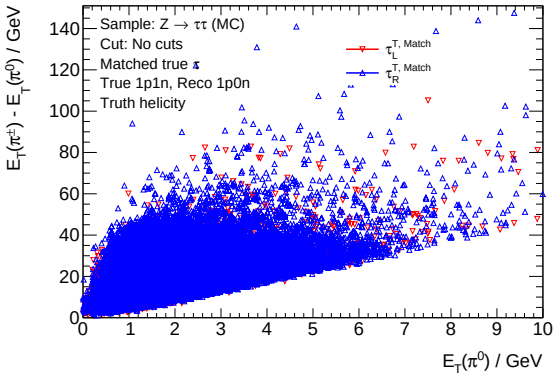
[9] Although Υ uses the energies of the $\pi^\pm$ and π^0, it also connects $E_T(\pi^\pm)$ and $E_T(\pi^0)$, because the η values of the $\pi^\pm$ and π^0 are very similar. The distribution of $\Delta R(\pi^\pm, \pi^0)$ is shown in Figure 5.12c.



(A) *Difference in transverse energy of the $\pi^{\pm}$ and π^0. The large tail in the distribution for τ_{Right} makes a correct reconstruction of the π^0 harder.*

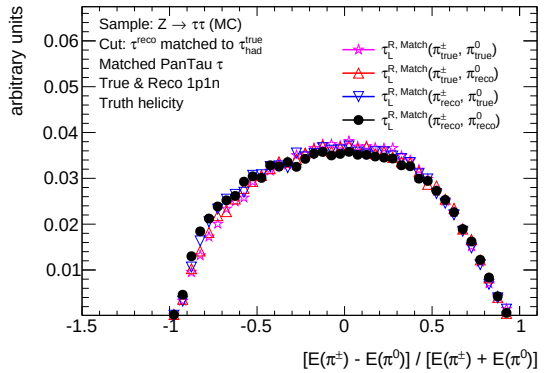


(B) *Transverse energy of the π^0. The majority of true 1p1n taus with $\Upsilon > 0.7$ from $Z \to \tau\tau$ decays have a π^0 that falls below the E_T threshold in π^0 reconstruction.*

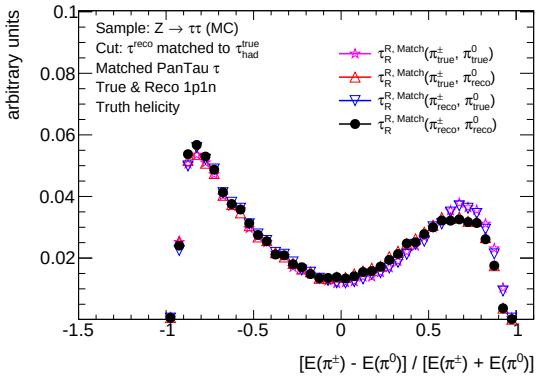


(C) *Energy difference as a function of $E_T(\pi^0)$. The sharp edge is due to the requirement $\Upsilon > 0.7$.*

FIGURE 7.9: *Transverse energy difference of $\pi^{\pm}$ and π^0 (7.9a), transverse energy of π^0 (7.9b) and the former as a function of the latter (7.9c) for true 1p1n taus, reconstructed as 1p0n with $\Upsilon > 0.7$ to investigate the dent in Figure 7.7.*
In summary, the dent arises from low energetic π^0 that fail the preselection cut, or π^0 that are badly reconstructed because of a much higher energy $\pi^{\pm}$ that is hard to subtract correctly. The bad reconstruction most likely leads to a reconstructed energy below the threshold, so that the π^0 is not reconstructed at all. More details are given in the text.

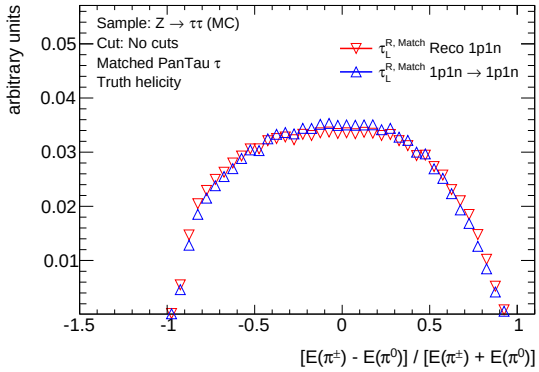


(A) *Left-handed, correctly reconstructed 1p1n taus.*

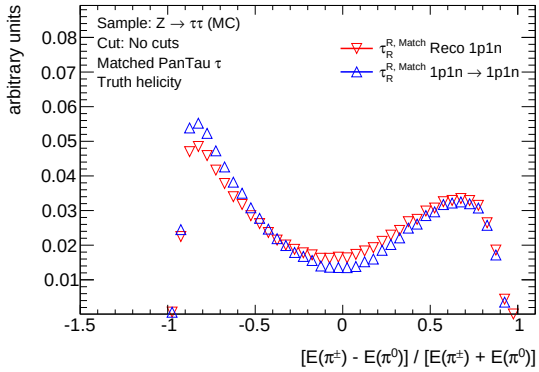


(B) *Right-handed, correctly reconstructed 1p1n taus.*

FIGURE 7.10: *ϒ distributions for left- and right-handed, correctly reconstructed 1p1n taus, using different variants of the $\pi^{\pm}$ and π^0. The magenta stars calculate ϒ using the true $\pi^{\pm}$ and π^0. The red (blue) triangles exchange the true π^0 ($\pi^{\pm}$) with the reconstructed one. Finally, the black dots represent the ϒ distribution calculated when using both reconstructed $\pi^{\pm}$ and π^0.*
Because of the good momentum resolution, exchanging the true $\pi^{\pm}$ with the reconstructed ones does not make a large difference. Analogously, due to the worse energy resolution of π^0, moving from true π^0 to reconstructed π^0 changes the distribution.

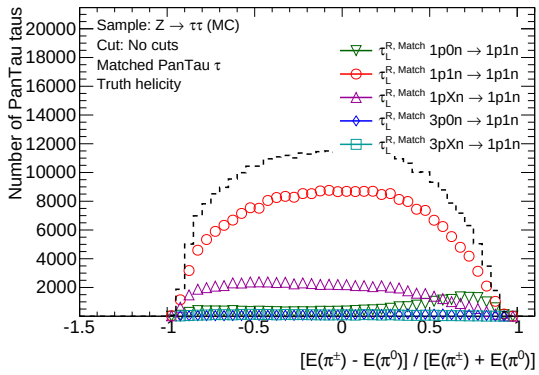


(A) *Left-handed taus.*

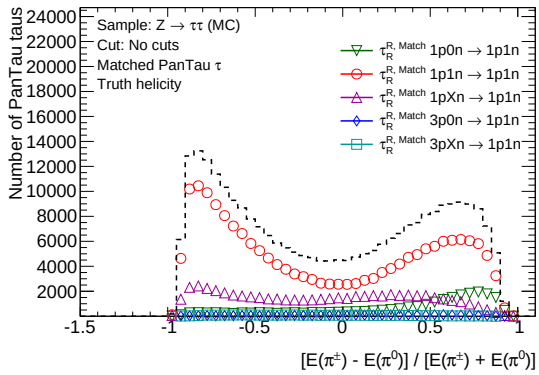


(B) *Right-handed taus.*

FIGURE 7.11: *Υ distributions for correctly reconstructed 1p1n true taus (blue) and all true taus that are reconstructed as 1p1n (red). The Υ for reconstructed 1p1n left-handed taus is a bit broadened compared to the one for correctly reconstructed 1p1n taus. In the set of right-handed taus, the negative Υ values are shifted towards zero.*



(A) *Left-handed taus.*



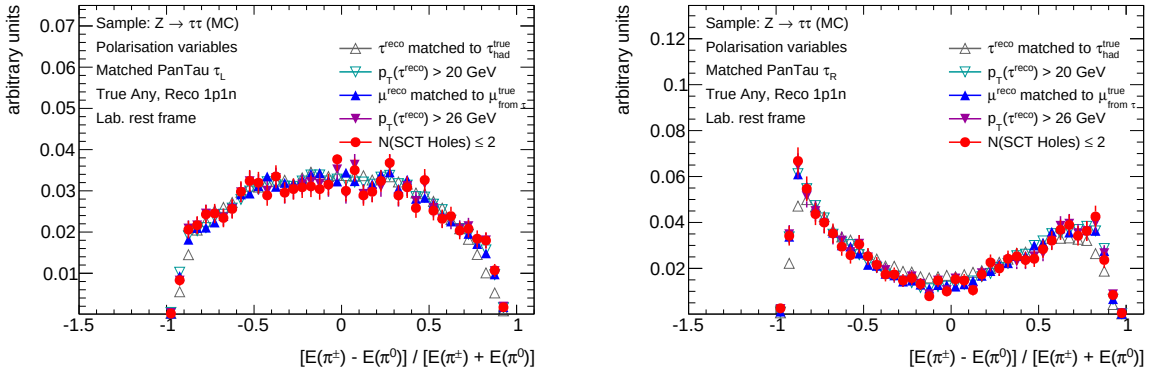
(B) *Right-handed taus.*

FIGURE 7.12: *Breakdown of the Υ shape for true taus reconstructed as 1p1n into the different true modes that contribute. The black dashed line is the sum of all the other curves. The true 1pXn and true 1p0n that are reconstructed as 1p1n are the main sources for the changes observed here. Contributions from $\tau_{3\text{-prong}}$ are negligible.*

This is somewhat expected, as negative Υ values indicate a high energetic π^0. Such a π^0 can be produced if the two π^0 from a 1pXn decay are close by and are not resolved[10], so that they appear as one neutral pion. This pion then has the energy of the two pions from 1pXn, lowering the resulting Υ value.

For 1p0n→1p1n it is the other way around. The π^0 in the 1p1n is not a real π^0, but stems from either noise, pile-up or imperfect subtraction, making it very low energetic in comparison to the $\pi^\pm$ which really originates from the tau decay.

[10] There are such taus which are classified as 1pXn - although they only have one reconstructed π^0- due to the number of photons found in the shots in EM1, c.f. Section 5.3.2. However, not all 1pXn taus are correctly classified because of this, and many 1pXn taus are migrated to 1p1n by `PanTau` (c.f. Table 5.12).



(A) ϒ for reconstructed 1p1n taus, matched to a left-handed true tau.

(B) ϒ for reconstructed 1p1n taus, matched to a right-handed true tau.

FIGURE 7.13: ϒ shapes of reconstructed 1p1n taus that have a truth match at selected stages of the event selection. The main differences can be observed for right-handed taus when applying the $p_T(\tau)$ cut, as discussed in the text in Section 7.3. The no cut stage is not shown because the set of taus shown here implicitly has a cut applied to it, namely the match to a true tau.

7.5.5 Reconstructed 1p1n Taus in the Event Selection

Finally, Figure 7.13 shows how the ϒ shapes are affected by the rest of the event selection cuts, other than moving from all true 1p1n taus to the set of all reconstructed 1p1n taus that have a truth match. The gray triangles in this figure correspond to the red triangles from Figure 7.11. There are two main effects when applying the rest of the event selection.

The first effect appears when applying the $p_T(\tau^{\text{reco}})$ cut of 20 GeV. For left- and right-handed taus, this cut shifts the ϒ distribution towards larger absolute values. The second effect is introduced by the cut on $p_T(\mu^{\text{reco}}) > 26$ GeV. While the distributions only change slightly, the statistics are reduced significantly, by a factor of 5 (10) for left-handed (right-handed) taus (c.f. Table 7.2).

7.6 Fraction Fit in ϒ

This section discusses the method that is deployed to extract the mean polarisation. As mentioned in section 7.2, no data is used in this study. To emulate the process of extracting the mean polarisation, the simulated sample of $Z \to \tau\tau$ events is split into two parts. The first part (the simulation part) is used to study the ϒ distribution and to extract selection efficiencies, as described in the previous sections. The second part (labelled *pseudo-data*) however, is only used in this part of the study and is treated as if it did not contain any truth information.

The idea behind the fraction fit is the following. In data, only one ϒ distribution is accessible, namely the one for all reconstructed 1p1n taus, i.e. it includes an unknown fraction of left-handed taus f_L, and an unknown fraction of right-handed taus, f_R. However, from the simulation part it is known how the left- and right-handed ϒ distributions for reconstructed 1p1n taus look like. They are shown in Figure 7.13.

In a so-called *fraction fit*, two sets of histograms are used. The one set, the observation, consists of one histogram only, in this case the ϒ distribution for all reconstructed 1p1n taus. The other set contains several histograms, the templates, which are supposed to contribute to the observation. Here, there are

two templates, the left- and right-handed Υ distributions for reconstructed 1p1n taus.

The templates can be added with different relative weights f_L and f_R, each combination giving a different agreement with the observation. The agreement in this case is quantified by calculating a χ^2 between the observation and the sum of the templates. The combination of f_L and f_R giving the minimal χ^2 - and therefore the best agreement - is the result of the fraction fit.

A more visual explanation is given in Figures 7.14 and 7.15. The first figure shows how the observation and the templates are obtained, the second figure illustrates how the fitting is performed. The result of the template fit depends on the shape differences between the templates. If the two templates shown in Figure 7.15 were both constants, then the weights could be chosen arbitrarily, so that no meaningful result can be obtained. Therefore it is important that the reconstructed version of Υ still retains the main features for its left- and right-handed parts.

An implementation of such a fit is provided in the ROOT software package [63], which was used to perform the actual fitting procedure. Technically the fit is set up such, that it is performed at every cut stage. The templates entering the fit are the left- and right-handed Υ distributions for reconstructed 1p1n taus with a truth match from the simulation part of the sample. The observation entering the fit is the Υ distribution for reconstructed 1p1n taus with a truth match from the pseudo-data part of the sample. Both the templates and the observation are taken from the corresponding cut stage. Figure 7.16 shows the graphical results of the fraction fit at four selected stages of the event selection. At all stages, the fit result (green) matches the observation in the pseudo-data (black). Also shown are the two templates, weighted by the corresponding fractions f_L and f_R that the fit provides. The numerical results of the fits are summarised in Table 7.3.

The *raw* fit results f_L^{Raw} and f_R^{Raw}, i.e. the fractions of the τ_{Left} and τ_{Right} contributions, need to be corrected for different selection efficiencies for left- and right-handed taus. This is done by using the values for efficiency times acceptance $\epsilon \cdot a$ that are listed in Table 7.2. The corrected fit results, $f_L^{\text{Corr.}}$ and $f_R^{\text{Corr.}}$, are calculated as follows.

$$f_L^{\text{Corr.}} = f_L^{\text{Raw}}/(\epsilon \cdot a) \tag{7.1}$$
$$f_R^{\text{Corr.}} = f_R^{\text{Raw}}/(\epsilon \cdot a) \tag{7.2}$$

Following the formulas

$$\left\langle p^{\text{Raw}} \right\rangle = \frac{f_R^{\text{Raw}} - f_L^{\text{Raw}}}{f_R^{\text{Raw}} + f_L^{\text{Raw}}} \tag{7.3}$$
$$\left\langle p^{\text{Corr.}} \right\rangle = \frac{f_R^{\text{Corr.}} - f_L^{\text{Corr.}}}{f_R^{\text{Corr.}} + f_L^{\text{Corr.}}} \tag{7.4}$$

the mean polarisation is obtained.

To test, whether the correction works properly, the observation is taken from the simulation part of the sample too. Thus, the two templates are *guaranteed* to exactly add up to the observation. The results of this test are shown in Figure 7.17. It shows the raw mean polarisation $\left\langle p^{\text{Raw}} \right\rangle$ without corrections and the corrected mean polarisation $\left\langle p^{\text{Corr.}} \right\rangle$. After the correction, the mean polarisation does not change throughout the complete event selection, except for its uncertainty, which increases due to the decreasing statistics in the templates and the observation. Thus, the correction is assumed to work properly.

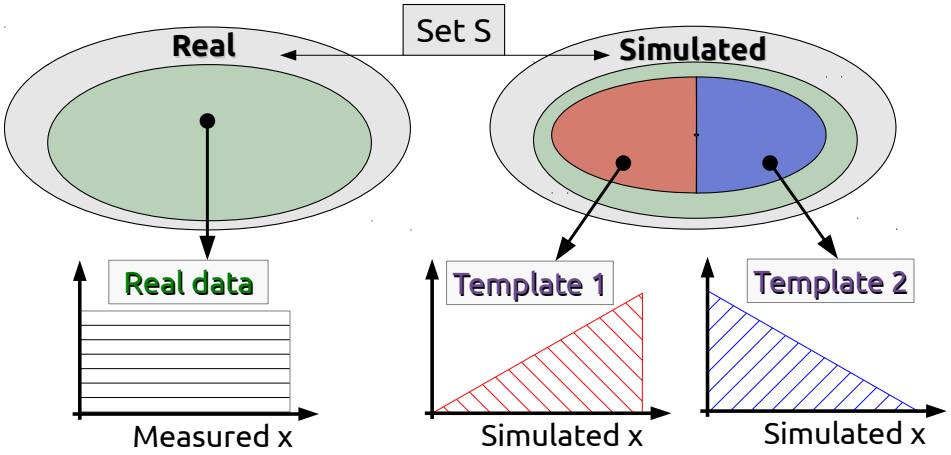


FIGURE 7.14: *This figure illustrates how the observation and the templates that enter a fraction fit are obtained. A given set S of events is measured in an experiment and is also simulated, so that there are two versions of it. In the study at hand, S is the set of $Z \to \tau\tau$ events. Within set S, there is a subset, coloured green, which is of interest and is known to be composed of two different contributions. Here, the subset corresponds to $Z \to \tau\tau \to \tau_\mu \rho\nu_\tau$, which is composed of left-handed and right-handed taus. In this special case, the two contributions make up for the complete subset, but that does not need to be the case. In the real version of set S, no subdivision is possible, so that the complete subset corresponds to the observation. The simulated version of S however, allows for a decomposition into the two contributions, coloured red and blue. These can be extracted from the simulation and are used as the templates.*

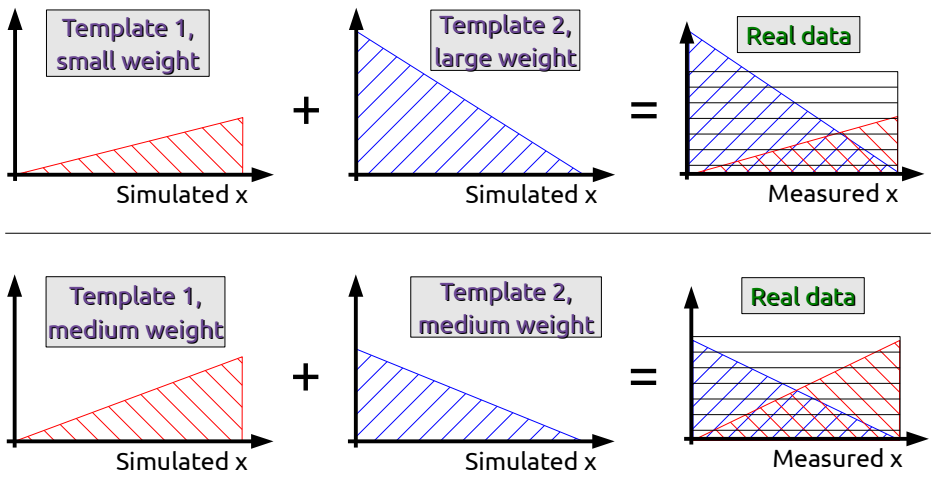
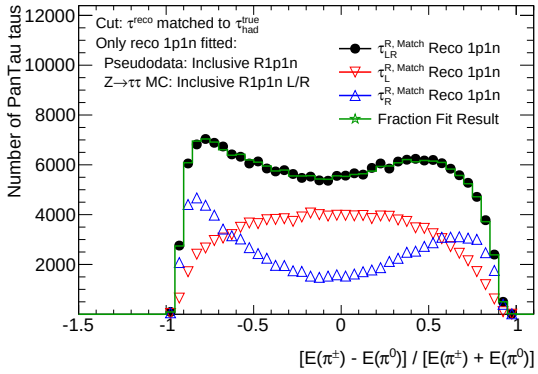
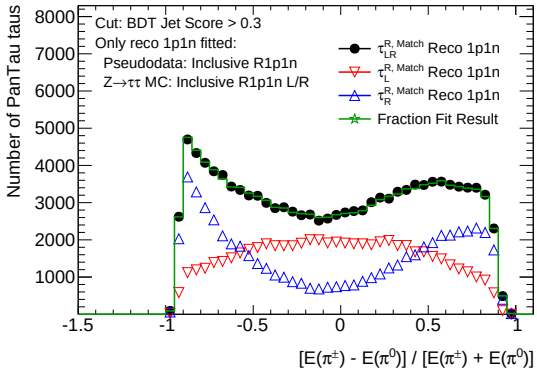


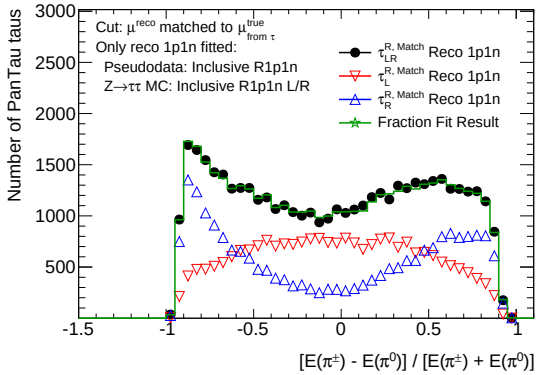
FIGURE 7.15: *Illustration of how the optimal weights for the templates are obtained in the example of two templates. Each template is assigned a weight and the weighted templates are summed up bin-by-bin. The sum is then compared to the observation by means of a χ^2 test. In this figure, two examples are given. The upper case uses a small weight for template 1 and a large fraction for template 2, resulting in a bad modelling of the observation. The sum of the templates overshoots the data at small x, and underestimates it at large x, resulting in a large χ^2. For other weights in the lower case, the sum of the weighted templates agrees much better with the observation, so that these weights give a smaller χ^2. The result of the fit corresponds to those weights, which give the minimal χ^2.*



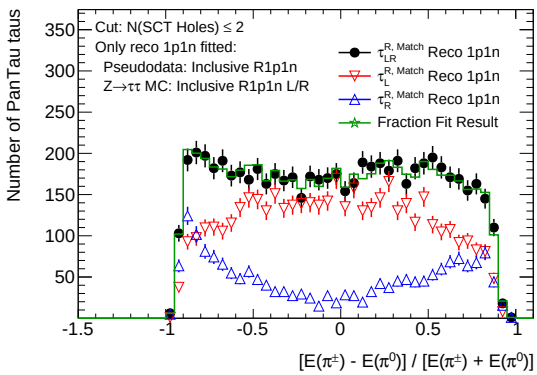
(A) *Template fit after the truth matching of the $\tau_{had\text{-}vis}$.*



(B) *Template fit after the last cut of the $\tau_{had\text{-}vis}$ selection.*



(C) *Template fit when requiring a truth matched μ.*



(D) *Template fit after the last muon selection cut.*

FIGURE 7.16: *Graphical results of the template fit using reconstructed 1p1n and neglecting the misreconstructed modes. The simulated templates are the red and blue distributions of left- and right-handed taus, the dots correspond to the pseudo-data.*
The fit is performed after the cut requiring the reconstructed tau to be truth matched (7.16a), after the last tau selection cut (7.16b), after the selecting truth matched reconstructed muons (7.16c), and after the last muon selection cut (7.16d). Numerical results of the fit are listed in Table 7.3.

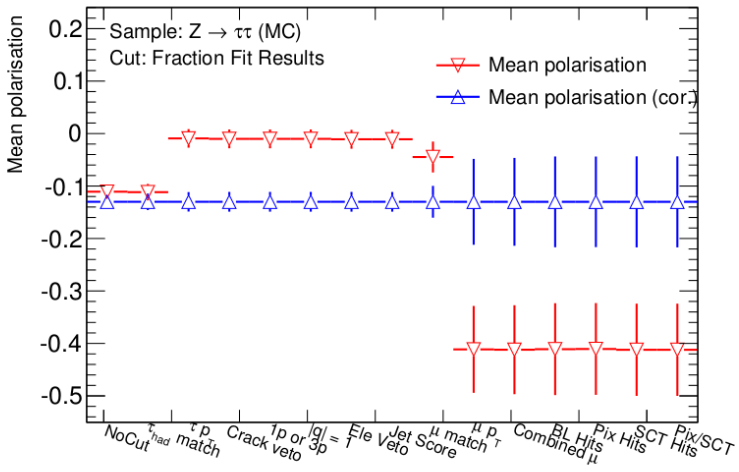


FIGURE 7.17: *Raw (red) and corrected (blue) mean polarisation as obtained in the simulation part of the sample. This test is done to see whether the correction for helicity dependent cut efficiencies works. The raw polarisation changes significantly throughout the event selection, while the corrected polarisation does not change, indicating that the correction works as intended.*

Cutstage	$\frac{\chi^2}{\text{NDF}}$	f_L^{Raw} in %	f_R^{Raw} in %	$\left\langle p^{\text{Raw}} \right\rangle$ in %	$f_L^{\text{Corr.}}$ in a.u.	$f_R^{\text{Corr.}}$ in a.u.	$\left\langle p^{\text{Corr.}} \right\rangle$ in %
No cuts	54.5/50	55.4 ± 0.33	44.5 ± 0.32	-10.9 ± 0.94	$4.21 \pm 25.4 \cdot 10^{-3}$	$3.26 \pm 24 \cdot 10^{-3}$	−12.8 ± 0.9
τ^{reco} matched to τ_{had}^{true}	56.2/50	55.7 ± 0.55	44.2 ± 0.54	-11.5 ± 1.57	$12 \pm 120 \cdot 10^{-3}$	$9.24 \pm 114 \cdot 10^{-3}$	−13.4 ± 1.5
$p_T(\tau^{\text{reco}}) > 20$ GeV	51/50	49.4 ± 0.63	50.5 ± 0.63	1 ± 1.78	$20.2 \pm 258 \cdot 10^{-3}$	$16.2 \pm 203 \cdot 10^{-3}$	−11.1 ± 1.8
$1.37 > \lvert\eta(\tau^{\text{reco}})\rvert > 1.52$	51.9/50	49.3 ± 0.63	50.6 ± 0.64	1.2 ± 1.81	$20.8 \pm 269 \cdot 10^{-3}$	$16.7 \pm 212 \cdot 10^{-3}$	−10.8 ± 1.8
$n_{\text{Track}}(\tau^{\text{reco}}) = 1$ or 3	51.9/50	49.3 ± 0.63	50.6 ± 0.64	1.2 ± 1.81	$20.8 \pm 269 \cdot 10^{-3}$	$16.7 \pm 212 \cdot 10^{-3}$	−10.8 ± 1.8
$\lvert\text{Charge}(\tau^{\text{reco}})\rvert = 1$	51.9/50	49.3 ± 0.63	50.6 ± 0.64	1.2 ± 1.81	$20.8 \pm 269 \cdot 10^{-3}$	$16.7 \pm 212 \cdot 10^{-3}$	−10.8 ± 1.8
BDT Ele. Score > 0.3	52.3/50	49.3 ± 0.63	50.6 ± 0.64	1.2 ± 1.81	$20.8 \pm 270 \cdot 10^{-3}$	$16.8 \pm 213 \cdot 10^{-3}$	−10.7 ± 1.8
BDT Jet Score > 0.3	52.7/50	49.3 ± 0.63	50.6 ± 0.64	1.3 ± 1.81	$20.8 \pm 270 \cdot 10^{-3}$	$16.8 \pm 214 \cdot 10^{-3}$	−10.6 ± 1.8
μ^{reco} matched to $\mu^{\text{true}}_{\text{from }\tau}$	48.6/50	50.6 ± 1.05	49.3 ± 1.05	-1.3 ± 2.98	54.9 ± 1.14	$45.1 \pm 965 \cdot 10^{-3}$	−9.9 ± 3
$p_T(\mu^{\text{reco}}) > 26$ GeV	34.8/50	70.4 ± 2.83	29.5 ± 2.64	-41 ± 8.58	350 ± 14	270 ± 24.2	−12.8 ± 8.5
μ is combined	29.8/50	71.6 ± 2.89	28.3 ± 2.67	-43.3 ± 8.83	376 ± 15.2	275 ± 25.9	−15.5 ± 8.6
N(BL Hits) ≥ 1	30.4/50	71.5 ± 2.98	28.4 ± 2.76	-43.1 ± 9.11	400 ± 16.7	294 ± 28.5	−15.4 ± 8.9
N(Pix Hits) ≥ 2	30.8/50	71 ± 2.98	28.9 ± 2.76	-42.2 ± 9.07	399 ± 16.7	299 ± 28.6	−14.3 ± 8.9
N(SCT Hits) ≥ 6	29.2/50	70.7 ± 2.99	29.2 ± 2.77	-41.6 ± 9.07	402 ± 17	306 ± 29.1	−13.5 ± 8.9
N(SCT Holes) ≤ 2	29.2/50	70.7 ± 2.99	29.2 ± 2.77	-41.6 ± 9.07	402 ± 17	306 ± 29.1	−13.5 ± 8.9

TABLE 7.3: *Numerical results of the template fit taking into account reconstructed 1p1n taus only.*
The first column states the cut stage, at which the fit is performed. The second column lists the χ^2/NDF of the fit.
Columns three and four contain the fractions that were extracted from the fit, i.e. the fractions of left-handed and right-handed 1p1n taus in the set of reconstructed 1p1n taus. In column five the mean polarisation is calculated from the fractions given in columns three and four.
The last triplet of columns contains the same information as columns three to five, except that the corrections for different selection efficiencies for left- and right-handed taus are taken into account. The correction is described in the text and the numerical values of the corrections are listed in Table 7.2.

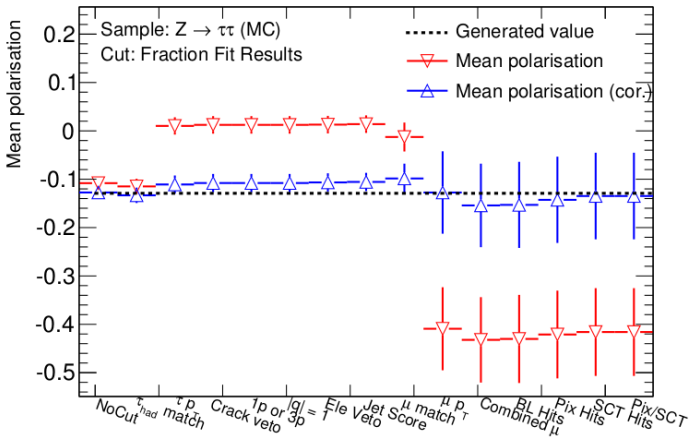


FIGURE 7.18: *Raw (red) and corrected (blue) mean polarisation as obtained in the pseudo-data part of the sample. Also shown is the generated value of* $\left\langle p^{\text{Gen.}} \right\rangle = -0.1303$ *as the black dashed line. The corrected mean polarisation matches the generated value roughly within the statistical error. At the event selection stages that select the hadronic tau, the polarisation is overestimated by a bit more than once the statistical uncertainty. Because the pseudo-data and the simulation part are statistically independent, it is expected that the measured value fluctuates around the generated one.*

When taking the observation from the pseudo-data part, in order to perform the measurement, the raw and corrected mean polarisation behave as shown in Figure 7.18. It is visible that the uncorrected polarisation follows the same pattern as in the simulation part of the sample. The cuts on the hadronic tau bias the remaining events to contain more right-handed taus, while the cuts on the muon from the leptonic tau bias the sample to contain more left-handed taus. The reasons for this have been discussed in Section 7.3, especially Figure 7.2.

The corrected mean polarisation now does not stay at a single value, because the two parts, pseudo-data and simulation, are statistically independent, so that it is expected that the mean polarisation fluctuates around the generated one within about the statistical uncertainty.

The generated and measured value of the mean polarisation, $\left\langle p^{\text{Gen.}} \right\rangle$ and $\left\langle p^{\text{Corr.}} \right\rangle$, are as follows. The number of left- and right-handed true taus, $N_{L,\text{true}}$ and $N_{R,\text{true}}$ in the pseudo-data part of the sample are:

$$N_{L,\text{true}} = 2534786 \qquad N_{R,\text{true}} = 1950259$$

Thus, the generated mean polarisation, $\left\langle p^{\text{Gen.}} \right\rangle$, is

$$\left\langle p^{\text{Gen.}} \right\rangle = \frac{N_{R,\text{true}} - N_{L,\text{true}}}{N_{R,\text{true}} + N_{L,\text{true}}} = \frac{1950259 - 2534786}{1950259 + 2534786} = -0.1303 \tag{7.5}$$

From the numeric results of the fraction fit, listed in Table 7.3, it can be read off that the corrected mean polarisation, $\left\langle p^{\text{Corr.}} \right\rangle$, as measured in the pseudo-data set is

$$\left\langle p^{\text{Corr.}} \right\rangle = \frac{f_R^{\text{Corr.}} - f_L^{\text{Corr.}}}{f_R^{\text{Corr.}} + f_L^{\text{Corr.}}} = \frac{306 - 402}{306 + 402} = -0.135 \pm 0.089^{\text{stat}} \tag{7.6}$$

Thus, the measured value matches the generated one within the (large) statistical uncertainty. As mentioned in section 7.3, the pseudo-data sample corresponds to an integrated luminosity of roughly 2.5fb^{-1}. By the end of 2012, however, the ATLAS detector recorded more than 20fb^{-1}, nearly 9 times as much as available in this study. Hence, projecting the results towards higher luminosity and assuming that the statistical error falls with the number of events N like $\frac{1}{\sqrt{N}}$, the statistical error for the complete

ATLAS data would be around 0.03.

However, in an actual analysis, also systematic uncertainties would need to be taken into consideration. For this kind of study, the main systematic uncertainties related to the fraction fit are:

- *Accuracy of decay mode classification:* As already mentioned in Section 5.3.8, there are no systematic uncertainties on the decay mode classification available so far. However, as the decay mode classification is crucial in this study when selecting reconstructed 1p1n taus, the systematic uncertainty on doing so will impact the result of the fraction fit.
- *Impact of misreconstructed modes, 1p0n and 1pXn:* The set of reconstructed 1p1n taus contains admixtures of falsely reconstructed 1p0n and 1pXn taus, which change the Υ shape, as shown in Figure 7.12. Uncertainties on those admixtures directly translate into uncertainties on the Υ shape.
- *Mismeasurements of the π^0 and $\pi^\pm$ energy:* Figure 7.10 showed that the Υ shape is altered when switching from the true π^0 and $\pi^\pm$ energies to the reconstructed ones. While the effect is small for $\pi^\pm$, the measurement of π^0 energies has a larger impact. Thus, the accuracy of the $E(\pi^0)$ measurement will have an impact on the fraction fit, which is sensitive to the shapes of the templates and the observation.
- *Contributions from background events:* In this simplified study, it was assumed that there is no background in the Υ distribution. In a real analysis, fake tau leptons will enter the distribution. There are methods to estimate their contribution, but that estimation will have an uncertainty applied to it, which will need to be propagated to the fraction fit result.

7.7 Summary

This chapter presented a small study of a possible application of the new substructure based tau reconstruction, which was described in the chapters 4 and 5. With the new reconstruction, it is possible to easily access polarisation sensitive variables in hadronic tau decays, especially in 1p1n decays.

A simplified event selection to select $Z \to \tau\tau \to \tau_\mu \rho\nu_\tau$ events has been performed on two sets of simulated $Z \to \tau\tau$ events. In one of the sets, the behaviour of the charged energy asymmetry Υ, a variable sensitive to the tau polarisation, has been studied. The effects that lead to changes in the shape of Υ have been discussed in Section 7.5. The main effects are due to soft π^0, which either escape reconstruction because of their low energy, or have their energy not measured very precisely (in comparison to the $\pi^\pm$) due to the energy resolution of the calorimeter.

Finally, a fraction fit has been performed in the distribution of all reconstructed 1p1n taus that are matched to a true tau in the pseudo-data part, with templates taken from the simulation part of the sample. The fit results have been corrected for helicity dependent cut efficiencies. The corrected pseudo-measurement of the mean Z boson polarisation of $\left\langle p^{\text{Corr.}} \right\rangle = -0.135 \pm 0.089^{\text{stat}}$ agrees with the generated value $\left\langle p^{\text{Gen.}} \right\rangle = -0.1303$ within its statistical uncertainties.

The rather large statistical uncertainties are expected to be reduced by a factor of ≈ 3 when performing this analysis on the actual data recorded by ATLAS, which corresponds to a much larger integrated luminosity than the pseudo-data part available for this study.

The main sources of systematic uncertainties, that need to be considered in an actual measurement are the accuracy of the decay mode classification, the contribution from misreconstructed decay modes, shifts in the Υ distribution due to the energy measurement of the π^0 and contributions from background events.

CHAPTER 8

Conclusions

A new approach to tau lepton reconstruction and one of its possible applications were presented in this thesis.

In the application of the new reconstruction method, the behaviour of a tau polarisation sensitive variable, the charged energy asymmetry Υ, was studied. It was found that although there are changes induced by the reconstruction of the taus, the main features of the variable shape remain. The differences between the generated and reconstructed distribution have been discussed and are mostly due to soft objects, which do not pass the E_T requirements imposed by the pile-up conditions. A usage of the Υ-variable in a fraction fit to extract the mean tau polarisation was presented and shown to work on a set of simulated $Z \to \tau\tau$ events. The mean Z polarisation was measured in the pseudo-data part and found to be $\left\langle p^{\text{Corr.}} \right\rangle = -0.135 \pm 0.089^{\text{stat}}$, in agreement with the generated value of $\left\langle p^{\text{Gen.}} \right\rangle = -0.1303$.

The new tau lepton reconstruction is based on the *energy flow* principle and consists of two major steps. The first step is to deploy the *energy flow* principle in the `CellBased` algorithm, which creates particle objects, referred to as PFOs. The second step is performed by `PanTau`. `PanTau` combines information from all reconstructed PFOs in order to (re-) classify the decay mode of the tau and adapt the 4-momentum.

In this first such version of `PanTau`, the reclassification is based on the number of PFOs in the tau and on the multiplicities and types of PFOs found by `CellBased`. As pointed out (c.f. Section 5.5.2), other approaches exist and may prove to be either more consistent, adaptive and flexible, or show a better performance.

The classification performance is measured in the fraction of correctly classified taus. Given the branching ratios of the tau, this fraction is 57.5% in a naive classification. The `CellBased` algorithm improves this fraction by 8%, and `PanTau` increases it by another 3.75%, resulting in close to 70% of correctly classified hadronically decaying tau leptons. The increase mainly comes from improving the 1p1n efficiency by $\approx$ 13%, while its purity is decreased by only 2%.

Because `PanTau` alters the decay mode, it also needs to add and remove particles from the tau, so that the 4-momentum reflects the (possible) changes in the decay mode. The approach taken here is rather simple and has room for improvements, because it just removes neutral pions or adds the neutral PFO with the highest π^0-BDT score, with only one exception. A more refined approach, whose actions depend on the tau kinematics, like E_T, could prove to give better results in terms of energy and spatial

resolution.

In comparison with the default reconstruction method, the new method drastically improves the energy and spatial resolution and opens the possibility to classify the tau decay into one of 5 categories, instead of two. The cost for this is a highly non-Gaussian energy and spatial resolution and a higher dependence on simulations, because the π^0-BDT and the `PanTau` BDTs rely on a proper simulation of tau decays. However, recent studies showed, that the agreement between the simulations and data is reasonable [52].

The utility of this new reconstruction method in tau identification was not found to be extraordinary [21, 41], which is because it deals with *coarser* objects. *Coarse* means, that variables based on the new reconstruction are calculated on a particle level, while variables in the current tau identification have access to the cell level, which is much finer. It needs to be seen whether a combination - however it may look like - of these information, cell-level information with a finer granularity and object-level information with combined detector measurements from tracker and calorimeter, is able to improve the current tau identification performance.

The new reconstruction method and its future development - if the method continues to work as shown in this thesis - may also benefit other areas of object reconstruction in ATLAS and inspire them to follow the same principle. A unified object reconstruction, based on the *energy flow* approach - which is much closer to what especially students beginning to work in this field are used to from lectures - may serve the consistency and clearness of the complete event reconstruction. The other multi-purpose detector at the LHC, CMS, already follows this approach.

On a personal note, it is pleasing to see that after a lot of re-evaluation and rethinking, `PanTau` now finally seems to have found its spot, not too far away from what has been stated by Robindra Prabhu and Sebastian Fleischmann, who started the work on `PanTau`:

"It was also shown that PanTau is sensitive to the decay modes of the τ-lepton and generally provides an improved transverse energy resolution. As all steps in the algorithm, including the building of discriminating features, rely squarely on the resolved objects of the decay, PanTau is distinguished insofar as it maintains a clear and natural separation between detector effects and the physics of the τ-lepton decay." - Robindra Prabhu [21]

"PanTau will be an important contribution to the aim of including more information about the substructure of tau jets and the identification of individual charged and neutral pions in the jets." - Sebastian Fleischmann [41]

Acknowledgements

At first I would like to thank *Prof. Klaus Desch* for giving me the opportunity to work on a PHD in his group with everything that this entails. The friendly and open atmosphere in the group was what convinced me to write my diploma thesis in his group and is also the reason I stayed for a PHD when I received the invitation.

I would like to thank *Philip Bechtle* for his motivation in times when it was urgently needed. Thanks to *Peter Wagner* for the discussions on the topics of this thesis and for proof reading this thesis when there was enough else to do. I'd also like to thank *Prof. Ian Brock*, for providing the LaTeX-template that is used by this thesis.

The day-to-day work would have not been the same without the awesome other diploma-, master- and PHD students I had the pleasure to share an office with. From making the cactus between Carolin and myself a Protection Force to helping each other out with technical and physical problems - I think we had a fun but also an efficient and productive time.

It should be noted that nothing in this thesis would have been possible without *Peter Wienemann*, *Sebastian Fleischmann* and *Robindra Prabhu*. As far as I am informed, the idea for `PanTau` came to Peter Wienemann, when he listened to a talk about the `Pandora` particle flow algorithm to be used at the International Linear Collider. The idea was to try particle flow in the tau reconstruction of the ATLAS detector. Combining `Pandora` and Tau, `PanTau` was born, initially to improve jet rejection by taking decay mode classification into account. Together with Sebastian Fleischmann and Robindra Prabhu, they began to write up the `PanTau` algorithm, which I then took over. Thus, if it hasn't been for them, this algorithm wouldn't exist.

There are, of course, more people to be mentioned with respect to `PanTau`. Therefore, I would like to thank *Sebastian Matberg* and *Robert Beckmann* for their comparisons of `PanTau`-variables in data and simulation. Robert Beckmann also invested a reasonable amount of his time into adapting the code that produces the performance plots to the new data format during the spring and summer of 2014 - thank you for that. A thank you is appropriate for *Christian Kurtscheid*, for providing insights into π^0 behaviour in tau decays. I'd also like to thank *Maike Hansen*, for looking into a possible p_T dependent training of the `PanTau` BDTs and rewriting the code that creates the input files for the `PanTau` training, when I did not have time for it. Finally, another thank with respect to `PanTau` efforts is due to *Peter Wagner*, who took over the creation of performance plots from Robert and recently took over taking care of the `PanTau` packages in the ATLAS reconstruction software framework.

It should also be mentioned, that `PanTau` in its first - but also current - form is only possible with algorithms that implement the energy flow principle. Thus I'd like to thank *Mark Hodgkinson*, for his help with the `eflowRec` algorithm, which provided the input for `PanTau` in the early days. I also like to thank - of course - *Benedict Winter*, who created the `CellBased` algorithm in the context of his master thesis. Closely related to this are *Stephanie Yuen* - she added the shot-finding to `CellBased`, improving

the `CellBased` performance and adding more variables to be exploited by `PanTau` - and *Will Davey* - with whom I think we had the first discussions about making `PanTau` more general so that it would work on any input algorithm.

Furthermore, I would like to thank *Felix Friedrich* and *Michel Janus*, for their help on technical issues, like the unwilling 'TC2' web pages and discussions about software implementations.

A special thank deserve *Simone Zimmermann*, *Sebastian Matberg*, *Mathias Uhlenbrock* and *Martin Schultens*. I'm not so sure about our audience, but I think we sure had fun practising and performing various songs at the yearly Brock-Desch-Christmas-Parties.

Another special thank goes out to *Mathias Uhlenbrock* and *Steffen Schaepe*. To Mathias, because he is the spiritual father of awesome 'Doktorwagen' - designing, building and celebrating the various 'Doktorwagen' always was a highlight. To Steffen, because of his motivation and his willingness, to take on new methods in creating 'Doktorwagen' - one day they will fly.

Abschliessend bedanke ich mich bei meiner Familie für die Unterstützung in den ganzen Jahren, insbesondere während des Studiums, der Diplomarbeit und der Anfangszeit der Doktorarbeit: Birgit, Jürgen, Karl und nicht zu vergessen Lilly, Franka und Fiby.

APPENDIX A

Additional information

A.1 Note about software, code and tools

In order to produce the histogram figures and numerical results presented in this thesis, for instance the figures in Chapter 6, dedicated C++ programs and software packages have been written, using a graphical user interface with an integrated development environment like `kDevelop`.

Due to the way that data is stored in the ATLAS experiment, these programs make extensive use of the `ROOT` libraries [63]. For the training and testing of the `PanTau`-BDTs, the `TMVA` package [40] has been used in conjunction with `ROOT` libraries and custom C++ code. Basic information on C++ has been taken from the C++ reference webpages [64], specific problems have been solved by referring to the webpages of *stackoverflow* [65]. The `TMVA` package also provides `ROOT` macros to extract the figures for overtraining tests and correlation matrices. These macros have been modified (altering the appearence of the figures) and utilised to produce the corresponding plots in this thesis. Documentation on `TMVA` has been taken from the TMVA homepage [66], the documentation for `ROOT`-classes was found in the `ROOT` class index [67]. In rare cases, information could also be found on the web pages of the ATLAS Experiment, the so-called *TWiki*, [68].

The analysis code for the polarisation analysis is written within the `SFrame` software framework [69], in conjunction with a D3PD interface developed in Bonn, called *SFM2* [70].

The code for the `PanTau` algorithm itself is also written in C++, within the ATLAS event reconstruction software framewok `ATHENA`. Configuration of the algorithm is done in `python`. Information on other software packages, which the `PanTau` software package depends on or makes use of, has been taken from the corresponding source code, available via the *trac* code browser [71] and the *acode-browser* [72]. Documentation for `python` was also taken from *stackoverflow* [65], other `python` source code and [73].

In order to access the many simulated and recorded data events and not have the need to store samples occupying terabytes of disk space at each institute that studies them, samples are stored with only a few replicas at computing centres (sites) around the world (the *computing grid* [74]). At these sites, there are also so-called worker nodes, i.e. computers, to which the analysis code can be sent. The analysis code is then executed on these remote computers, which have direct access to the sample that is to be analysed. The output, which is of much smaller size than the sample on which it is executed, can be downloaded to a local computer or laptop to be studied.

A.2 Samples used in Tau Reconstruction Studies & Polarisation studies

The samples that are used in this thesis are listed in Table A.1.

A.3 Software versions

Simulation, digitisation and reconstruction of the $Z \rightarrow \tau\tau$ events was performed with the ATLAS event reconstruction software, `ATHENA`, release 17.2.7.5. Conversion into the D3PD format, which is easily readable by the ROOT analysis framework [63], was done in `ATHENA` release 17.2.7.5.3.

The so-called 'patch-packages', which do not belong the the `ATHENA` releases mentioned above, but are needed as they provided parts of the new tau reconstruction software, are listed in Table A.2.

Events	Sample (short name)	Comment
$1.50 \cdot 10^6$	tid778651_00.v06-00/	Training sample (PanTau, Cluster-Based)
$1.25 \cdot 10^6$	tid00999072_00.v06-00/	Training sample (Cell-Based , eflowRec)
$1.25 \cdot 10^6$	tid00999073_00.v06-00_VersionA/	Evaluation sample
$1.25 \cdot 10^6$	tid00999074_00.v06-00/	Evaluation sample
$1.25 \cdot 10^6$	tid00999075_00.v06-00/	Evaluation sample
$1.00 \cdot 10^6$	tid00999076_00.v06-00/	Evaluation sample

TABLE A.1: *List of samples used in the studies using τ leptons from $Z \rightarrow \tau\tau$ events with pile up conditions as in 2012. All samples simulate $Z \rightarrow \tau\tau$ decays, and all samples are in the D3PD format.*
The full names of the samples can be obtained, by prefixing the short names given in the table with 'user.mhodgkin.TauPi0Rec_D3PD.', '147818.Pythia8_AU2CTEQ6L1_Ztautau.recon.ESD.' and 'e1176_s1479_s1470_r3553_' (in this order). The entries labelled with 'Evaluation sample' have been used in both the performance studies in Chapter 6, as well as in the polarisation studies in Chapter 7. The sample containing tid00999073 also served as the sample to test the PanTau-BDTs, that are trained on sample containing tid778651, for overtraining. In the latter, they have been subdivided, so that half of each of the four parts contribute equally to the pseudo-data and simulation part of the polarisation study.

Package name and version	
Reconstruction Step	D3PD Creation
TrigTauRec-00-08-16	
RecTPCnv-00-12-19	
RecAthenaPool-00-06-18	
RecEvent-00-00-02	
tauEventTPCnv-00-00-03	
tauEventAthenaPool-00-00-01	
tauRec-04-03-11	
TauDiscriminant-01-07-43	
eflowRec-00-02-98-Pi0InformationBranch-01	
eflowEvent-00-02-67-Pi0InformationBranch-01	
eflowEventTPCnv-00-01-08-Pi0InformationBranch-01	
eflowAthenaPool-00-00-09-Pi0InformationBranch-01	
tauEvent-00-07-06	tauEvent-00-07-06-01
PanTauAlgs-00-04-08	PanTauAlgs-00-04-12
PanTauEvent-00-04-07	PanTauEvent-00-04-08
PanTauExample-00-04-00	PanTauExample-00-04-00
PanTauInterfaces-00-04-05	PanTauInterfaces-00-04-06
PanTauUtils-00-04-00	PanTauUtils-00-04-00
PanTauAnalysis-00-04-06	PanTauAnalysis-00-04-09
CaloRec-02-12-91	TauD3PDMaker-01-08-33-branch
	TrigTauD3PDMaker-00-02-08

TABLE A.2: *Patch packages applied to ATHENA release 17.2.7.5 (reconstruction step) and 17.2.7.5.3 (D3PD creation). At the time of producing the samples, PanTau was also executed in the D3PD creation step, so that the important PanTau versions are the ones in the right column.*

Bibliography

[1] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", CERN-PH-EP-2012-218 (Aug. 2012), DOI: 10.1016/j.physletb.2012.08.020, URL: https://cds.cern.ch/record/1471031.

[2] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", CERN-PH-EP-2012-220 (July 2012), DOI: 10.1016/j.physletb.2012.08.021, URL: https://cds.cern.ch/record/1471016.

[3] P. W. Higgs, "Broken symmetries and the masses of gauge bosons", *Phys. Rev. Lett. 13 (1964) 508-509* (1964).

[4] *Building the LHC*, URL: http://timeline.web.cern.ch/timelines/Building-the-LHC.

[5] Stefan Berge, Werner Bernreuther and Sebastian Kirchner, "Determination of the Higgs CP-mixing angle in the tau decay channels" (2014), URL: http://arxiv.org/abs/1410.6362.

[6] Stephen P. Martin, "A Supersymmetry Primer", arXiv:hep-ph/9709356 (2011), URL: http://arxiv.org/abs/hepph/9709356.

[7] Super-Kamiokande Collaboration, "A Measurement of atmospheric neutrino oscillation parameters by SUPER-KAMIOKANDE I", *Phys.Rev. D71 (2005) 112005* (2005), URL: http://arxiv.org/abs/hep-ex/0501064.

[8] The SNO Collaboration, "Electron energy spectra, Fluxes, and day-night asymmetries of B-8 solar neutrinos from the 391-day salt phase SNO data set", *Phys. Rev. C72 (2005) 055502* (2005), URL: http://arxiv.org/abs/nucl-ex/0502021.

[9] C. Limbach, "Leptonische Tau-Zerfalle in supersymmetrischen Prozessen im ATLAS-Detektor", 2009, URL: http://www.lhc-ilc.physik.uni-bonn.de/ergebnisse/abschlussarbeiten.

[10] S. H. Neddermeyer and C. D. Anderson, *Note on the nature of cosmic-ray particles*, 1937, URL: http://timeline.web.cern.ch/anderson-and-neddermeyer-discover-the-muon.

[11] Perl, M. L. et al., *Evidence for Anomalous Lepton Production in e+e- Annihilation*, 1975, URL: http://journals.aps.org/prl/abstract/10.1103/PhysRevLett.35.1489.

[12] G. Danby et al., "Observation of High-Energy Neutrino Reactions and the Existence of Two Kinds of Neutrinos", *Physical Review Letters, vol. 9, Issue 1, pp. 36-44* (1962), URL: http://inspirehep.net/record/28187.

[13] DONUT Collaboration, "Observation of Tau Neutrino Interactions", *Phys.Lett.B504:218-224* (2001), URL: http://arxiv.org/abs/hep-ex/0012035.

[14] K.A. Olive et al. (Particle Data Group), *Chin. Phys. C, 38, 090001* (2014), URL: http://pdg.lbl.gov/2014/tables/contents_tables.html.

[15] Halzen, F. and Martin, A. D., *Quarks & Leptons: An Introductory Course in Modern Particle Physics*, John Wiley & Sons, 1984.

[16] Wu, C. S., *Experimental Test of Parity Conservation in β-Decay*, 1957, URL: http://journals.aps.org/pr/abstract/10.1103/PhysRev.105.1413.

[17] Griffith, D., *Introduction to Elementary Particles*, John Wilez & Sons, 1987.

[18] Peskin, M. E., Schroeder, D. V., *An Introduction to Quantum Field Theory*, Westview Press, 1995.

[19] Particle Data Group, "The CKM quark-mixing matrix", *Physics Letters B. Band 667, 2008, p. 1* (2008), URL: pdg.lbl.gov/2008/reviews/kmmixrpp.pdf.

[20] S. Boyd, *The Weak Interaction*, 2014, URL: http://www2.warwick.ac.uk/fac/sci/physics/staff/academic/boyd/stuff/neutrinolectures.

[21] R. P. Prabhu, "Studies into tau reconstruction, missing transverse energy and photon induced processes with the ATLAS detector at the LHC", BONN-IR-2011-10, Universität Bonn, 2011, URL: http://www.lhc-ilc.physik.uni-bonn.de/research-groups/experimental-physics/prof.-k.-desch/results/theses.

[22] A. Pich, "Tau Physics", *Adv.Ser.Direct.High Energy Phys. 15 (1998) 453-492* (), URL: http://arxiv.org/abs/hep-ph/9704453.

[23] I. Deigard, "Measurement of the Tau Polarization in Z-tautau Decays with the ATLAS Detector", 2012, URL: https://cds.cern.ch/record/1464097/?ln=en.

[24] *Particle Physics Booklet*, 2014, URL: http://pdg.lbl.gov/2014/reviews/rpp2014-rev-clebsch-gordan-coefs.pdf.

[25] P. B. Lyndon Evans, "LHC Machine", JINST 3 S08001 (2008), DOI: doi:10.1088/1748-0221/3/08/S08001, URL: http://iopscience.iop.org/1748-0221/3/08/S08001.

[26] C. Lefevre, "LHC: the guide" (2009), URL: http://cds.cern.ch/record/1165534/.

[27] *Cern Accelerator Complex*, URL: http://te-epc-lpc.web.cern.ch/te-epc-lpc/machines/pagesources/Cern-Accelerator-Complex.jpg.

[28] ALICE Collaboration, *J. Phys. G: Nucl. Part. Phys.* 32 (2006) 1295–2040, DOI: doi:10.1088/0954-3899/32/10/001, URL: http://iopscience.iop.org/0954-3899/32/10/001/pdf/g6_10_001.pdf.

[29] *The science behind LHCb*, URL: http://lhcb-public.web.cern.ch/lhcb-public/en/Physics/Physics-en.html.

[30] S. Catani, "Aspects of QCD, from the Tevatron to the LHC", CERN-TH-2000-073 (May 2000), URL: https://cds.cern.ch/record/439454?ln=en.

[31] O. Bruning, "LHC challenges and upgrade options", 110.J. Phys. Conf. Ser. (2008) 112002, DOI: 10.1088/1742-6596/110/11/112002, URL: http://iopscience.iop.org/1742-6596/110/11/112002/.

[32] The ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", JINST 3 S08003 (Aug. 2008).

[33] P. Langacker, "The Physics of Heavy Z' Gauge Bosons", arXiv:0801.1345 [hep-ph] (Nov. 2009), DOI: 10.1103/RevModPhys.81.1199, URL: http://arxiv.org/abs/0801.1345.

[34] S. Zimmermann, "A Search for Stable Massive Particles Carrying Electric Charges of 2e to 6e with the ATLAS Detector at the LHC", BONN-IR-2013-12, Universität Bonn, 2013, URL: http://www.lhc-ilc.physik.uni-bonn.de/ergebnisse/abschlussarbeiten.

[35] Torbjorn Sjostrand, Stephen Mrenna, Peter Skands, *PYTHIA 6.4 Physics and Manual*, 2006, URL: http://arxiv.org/abs/hep-ph/0603175.

[36] S. Agostinelli et al., "Geant4 - A Simulation Toolkit", *Nuclear Instruments and Methods A 506 (2003) 250-303* (), URL: http://geant4.cern.ch/results/publications.shtml.

[37] *Standardized Notation for tau Leptons at ATLAS*, URL: https://twiki.cern.ch/twiki/pub/AtlasProtected/EditorialBoardGuidelines/tau-notation-v1.0.pdf.

[38] The ATLAS collaboration, "Identification of the Hadronic Decays of Tau Leptons in 2012 Data with the ATLAS Detector", ATLAS-CONF-2013-064 (2013), URL: https://cds.cern.ch/record/1562839?ln=en.

[39] The ATLAS collaboration, "Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons", ATLAS-CONF-2011-077 (2011), URL: https://cds.cern.ch/record/1353226?ln=en.

[40] A. Hoecker et al., "TMVA: Toolkit for Multivariate Data Analysis", *PoS* ACAT (2007) 040, arXiv: physics/0703039.

[41] S. Fleischmann, "Tau Lepton Reconstruction with Energy Flow and the Search for R-Parity-Violating Supersymmetry at the ATLAS Experiment", BONN-IR-2012-09, Universität Bonn, 2011, URL: http://cds.cern.ch/record/1504815?ln=de.

[42] B. T. Winter, "Reconstruction of neutral pions in hadronic tau lepton decays in the ATLAS detector", BONN-IB-2013-04, Universität Bonn, 2013, URL: http://hep1.physik.uni-bonn.de/atlas_physics_pub.html#c297.

[43] M. Hodgkinson, D. Tovey, R. Duxfield, "Energy Flow Reconstruction with the eflowRec Combined Reconstruction Software in Athena 15.6.9.8", tech. rep. ATL-PHYS-INT-2011-031, CERN, Apr. 2011, URL: %7Bhttps://cds.cern.ch/record/1342125%7D.

[44] M. Trottier-McDonald, "Identifying Hadronic Tau Decays Using Calorimeter Topological Clusters at ATLAS", CERN-THESIS-2010-243, Simon Fraser University, 2010, URL: http://cds.cern.ch/record/1390497.

[45] W.-M. Yao et al., *J. Phys. G: Nucl. Part. Phys.* 33 (2006) 267.

[46] M. Hodgkinson, *Private e-mail exchange*.

[47] M. Hodgkinson, *Pi0 Tagging with Neutral EFO Objects in eflowRec*, URL: https://indico.cern.ch/event/242890/material/slides/0?contribId=6.

[48] R. N. Clarke, *Pi0 Shrinking Cone Parametrizations*, 2013, URL: https://indico.cern.ch/event/247745/.

[49] B. Winter, *Private e-mail exchange, May 16th, 2014*.

[50] P. Böhme, *Programmiersprache C/C++, 6.2.3. Datentyp float*, 1995, URL: http://www2.informatik.uni-halle.de/lehre/c/c623.html.

[51] B. Winter, *Private e-mail exchange, November 4th, 2014*.

[52] R. Beckmann, "Data/Monte-Carlo Comparison for Tau Lepton Reconstruction with Decay Mode Identification in the ATLAS Detector", University of Bonn, 2014, URL: http://www.lhc-ilc.physik.uni-bonn.de/ergebnisse/abschlussarbeiten.

[53] C. Kurtscheid, Private eMail exchange, code producing the plot to be found on: baf.physik.uni-bonn.de/lustre/atlas/data/share/tau/ckurtscheid/SFM2/ZtautauPolarizationStudy/src/Pi0Study.cxx, 2014.

[54] C. Kurtscheid, "Rekonstruktionsstudien neutraler Pionen in hadronischen Tau-Lepton-Zerfällen im ATLAS Detektor", 2014, URL: http://pi.physik.uni-bonn.de/pi_plone/lhc-ilc/view.php?lang=de&c=t&id=51.

[55] S. Lai and P. Wagner, *Tau Substructure Metrics*, URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/TauSubstructureMetrics.

[56] Christian Limbach, for the ATLAS Tau Combined Performance Group, *Substructure Plot Book*, 2013, URL: https://indico.cern.ch/event/272214/,https://indico.cern.ch/event/252804/.

[57] Peter Wagner, *Plotbook November 2014*, URL: https://pwagner-afs.web.cern.ch/pwagner-afs/tmp/PlotBook_noID_v36.pdf.

[58] The ATLAS collaboration, "Determination of the tau energy scale and the associated systematic uncertainty in proton-proton collisions at $\sqrt{s}$=8TeV with the ATLAS detector at the LHC in 2012" (2013), ATLAS-CONF-2013-044, URL: https://cds.cern.ch/record/1544036.

[59] Z. Czyczula, T. Przedzinski, and Z. Was, "TauSpinner Program for Studies on Spin Effect in tau Production at the LHC", CERN-PH-TH/20011-307 (2012), URL: http://arxiv.org/abs/1201.0117.

[60] M. Roggendorf, "Studies of Tau-Lepton Polarisation in Decays of Higgs and Z Bosons with the ATLAS Experiment", 2014, URL: http://hep1.physik.uni-bonn.de/diploma.html?&no_cache=1.

[61] *TauRecommendationsWinterConf2013*, URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/TauRecommendationsWinterConf2013.

[62] The ATLAS Collaboration, "Z to tau tau cross section measurement in proton-proton collisions at 7 TeV with the ATLAS experiment", ATLAS-CONF-2012-006 (2012), URL: http://cds.cern.ch/record/1426991.

[63] Rene Brun et al., *ROOT - Data Analysis Framework*, URL: http://root.cern.ch/drupal/.

[64] *Standard C++ Library reference*, URL: http://www.cplusplus.com/reference/.

[65] *stackoverflow*, URL: http://stackoverflow.com/.

[66] *TMVA Toolkit for Multivariate Data Analysis with ROOT*,
URL: http://tmva.sourceforge.net/.

[67] *ROOT Class Index*, URL: http://root.cern.ch/root/html/ClassIndex.html.

[68] *ATLAS TWiki Portal*, Limited to ATLAS members.,
URL: https://twiki.cern.ch/twiki/bin/viewauth/Atlas/WebHome.

[69] David Berge, Johannes Haller, Attila Krasznahorkay Jr.,
SFrame - A ROOT data analysis framework,
URL: http://sourceforge.net/projects/sframe/.

[70] Sebastian Matberg, *SFM2*,
URL: https://wiki.physik.uni-bonn.de/atlas/public/index.php/SFM2.

[71] *trac code browser*, Limited to ATLAS members.,
URL: https://svnweb.cern.ch/trac/atlasoff/browser.

[72] *ATLAS Software Code Browsing*, URL: http://acode-browser2.usatlas.bnl.gov/.

[73] *Python Documentation contents*, URL: https://docs.python.org/2/contents.html.

[74] *Worldwide LHC Computing GRID*, URL: http://wlcg-public.web.cern.ch/.

List of Figures

List of Tables