

## **Masterarbeit**

zur Erlangung des akademischen Grades  
Master of Science

# **Reconstruction and Identification of Boosted Tau Pair Topologies at ATLAS**

**David Kirchmeier**

geboren am 11.04.1990 in Cottbus

TECHNISCHE UNIVERSITÄT DRESDEN  
INSTITUT FÜR KERN- UND TEILCHENPHYSIK  
FACHRICHTUNG PHYSIK  
FAKULTÄT MATHEMATIK UND NATURWISSENSCHAFTEN

18. Mai 2015

Eingereicht am: 18. 05. 2015

1. Gutachter: Prof. Dr. A. Straessner
2. Gutachter: Prof. Dr. M. Kobel

## Abstract

Decays that involve a pair of tau leptons in the final state are important channels for the search of heavy resonances, which are predicted by theories that go beyond the Standard Model of particle physics. With the restart of LHC in 2015 higher energies and particle masses will be reachable for these processes. Thus, in particular the understanding of highly boosted tau pairs in the high energy region is essential for the search for new physics. With the current approach of tau reconstruction it is not possible to reconstruct di-tau topologies with low spatial separation. Due to the usage of anti- $k_t$ -4 seed jets, tau leptons with a sum of transverse momenta of  $p_T \gtrsim 500 \text{ GeV}$  or respectively an angular distance of  $\Delta R < 0.4$ , merge into the same jet. Therefore, in this thesis a new approach of di-tau reconstruction is introduced, which extends the sensitivity to tau pair decays to up to  $p_T \approx 1200 \text{ GeV}$  by reconstructing a di-tau topology in one anti- $k_t$ -10 seed jet. Additionally, with the di-tau identification a multi-variate background separation algorithm is developed, which is able to discriminate QCD di-jets with a high efficiency.

## Kurzfassung

Zerfälle mit zwei Tau-Leptonen im Endzustand sind wichtige Kanäle für die Suche nach schweren Resonanzen, die von Theorien jenseits des Standardmodells vorhergesagt werden. Mit dem Neustart des LHC im Jahr 2015 werden für solche Suchen noch höhere Energien und Massen als bisher erreichbar sein. Ein genaues Verständnis hochenergetischer, geboosteter Tau-Paare wird damit für die Suche nach neuer Physik immer wichtiger. Mit den bisherigen Methoden der Tau-Rekonstruktion ist es nicht möglich zwei nah beieinander liegende Tau-Leptonen getrennt zu rekonstruieren. Ab einer Summe der Tau-Impulse von  $p_T \approx 500 \text{ GeV}$ , was etwa einem räumlichen Abstand von  $\Delta R = 0.4$  entspricht, werden sie durch den zugrundeliegenden Anti- $k_t$ -4-Algorithmus nicht mehr getrennt erkannt, sondern als ein gemeinsamer Teilchenjet rekonstruiert. In dieser Arbeit wird daher ein neuer Di-Tau Algorithmus vorgestellt, der die bisherige Methode im Bereich hoher Energien, bzw. kleiner Tau-Abstände ergänzen soll. In einem ersten Rekonstruktionsschritt können Tau-Paare basierend auf einem Anti- $k_t$ -10 Jet mit hoher Effizienz mit Impulsen von bis zu  $p_T \approx 1200 \text{ GeV}$  rekonstruiert werden. Ein zusätzlicher Identifikationsschritt ermöglicht, basierend auf multivariaten Verfahren, eine effiziente Unterdrückung von QCD Di-Jet Untergrund.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>The Standard Model of Particle Physics and Heavy Higgs Bosons</b>	<b>3</b>
2.1	The Standard Model of Particle Physics . . . . .	3
2.1.1	The Particle Content of the Standard Model . . . . .	3
2.1.2	The Higgs Mechanism . . . . .	5
2.1.3	Problems of the Standard Model . . . . .	5
2.2	Two-Higgs-Doublet Models and Heavy Higgs Bosons . . . . .	6
2.2.1	The Four 2HDM Types . . . . .	6
2.2.2	Electroweak Symmetry Breaking . . . . .	7
2.2.3	Gauge Boson and Fermion Couplings to 2HDM Higgs Bosons	9
<b>3</b>	<b>The LHC and the ATLAS Detector</b>	<b>12</b>
3.1	The Large Hadron Collider . . . . .	12
3.2	The ATLAS Detector . . . . .	13
3.2.1	The ATLAS Coordinate System . . . . .	14
3.2.2	The Inner Detector . . . . .	15
3.2.3	The Calorimeter System . . . . .	17
3.2.4	The Muon Spectrometer . . . . .	19
3.2.5	The Forward Detectors . . . . .	19
3.2.6	The Trigger System . . . . .	19
<b>4</b>	<b>Single Tau Reconstruction and Identification</b>	<b>21</b>
4.1	Clustering . . . . .	22
4.2	Jet Algorithms . . . . .	22
4.3	Tau Reconstruction . . . . .	23
4.4	Tau Identification . . . . .	24
4.5	Reconstruction Results for Highly Boosted Di-Tau Topologies . . . . .	25
<b>5</b>	<b>Di-Tau Reconstruction and Identification</b>	<b>30</b>
5.1	The Di-Tau Decay Topology . . . . .	30
5.2	Technical Implementation . . . . .	33

5.3	Di-Tau Reconstruction . . . . .	34
5.3.1	Seed Jet Reconstruction . . . . .	34
5.3.2	Subjet Reconstruction . . . . .	34
5.3.3	Vertex Association . . . . .	35
5.3.4	Track Association . . . . .	36
5.3.5	Results . . . . .	37
5.4	Di-Tau Identification . . . . .	37
5.4.1	Discriminating Variables . . . . .	39
5.4.2	Boosted Decision Trees . . . . .	46
5.4.3	Results . . . . .	49
<b>6</b>	<b>Outlook</b>	<b>53</b>
<b>7</b>	<b>Summary</b>	<b>55</b>
<b>A</b>	<b>Units</b>	<b>57</b>
<b>B</b>	<b>Data Sets</b>	<b>58</b>
<b>C</b>	<b>BDT Configuration</b>	<b>60</b>
	<b>List of Figures</b>	<b>61</b>
	<b>List of Tables</b>	<b>63</b>
	<b>Bibliography</b>	<b>64</b>

# Chapter 1

## Introduction

The history of particle physics reaches back not even 120 years and is characterized by a rapid development of experimental observations, as well as their theoretical explanations. From the discovery of the first subatomic particle, the electron, which can be seen as the starting point of particle physics in 1897 by Joseph John Thomson [1], until the discovery of the muon [2], in studies of cosmic radiations, less than 40 years passed. The muon is a lepton, like the electron, with very similar characteristics, except a much higher mass. After nearly another 40 years, the third charged lepton could be observed at the Stanford electron-positron collider ring, the tau lepton [3].

In the meantime theorists developed a full description of particle interactions in the sub-atomic range, involving particles not even observed by then. Based on the work of theoretical physicists like Sin-Itiro Tomonaga, Julian Schwinger and Richard Feynman the theory of quantum electrodynamics was developed [5, 6, 7], which describes the electromagnetic interaction of particles. The quantum flavourdynamics is based on the work of Enrico Fermi [8], describing weak particle interactions and nuclear decays. By physicists like Murray Gell-Mann and George Zweig the third important force of particle interactions could be explained, the strong force, with the theory of quantum chromodynamics [9, 10]. These theories are incorporated into the Standard Model of particle physics [11]. In combination with the Higgs mechanism, which is essential to explain particle masses, this is the basis for particle physics descriptions. The Standard Model is the most successful theory in particle physics, thoroughly tested in the last decades and is showing a high accordance between its predictions and experimental observations.

The development of particle physics is also accompanied by the engineering of steadily larger and more complex experiments to study the rarest events at high energies. While Thomson started with accelerating electrons with cathodes in his laboratory, nowadays the Large Hadron Collider (LHC), with a circumference of 27 km, is the most complex particle collider ever built. The most important ob-

servation at the LHC so far was the discovery of the Higgs boson with a mass of  $125\text{ GeV}^1$  [12, 13] by both multi-purpose detectors at the LHC, ATLAS [14] and CMS [15]. With the discovery of the Higgs boson, the last missing particle of the Standard Model, the validity is confirmed furthermore.

Nevertheless there are some phenomena that can not be explained, which call for new physics explanations. The Two-Higgs-Doublet Models are one of these theories that go beyond the Standard Model. Those models predict the existence of four additional Higgs bosons, which are searched in high-energy decays at ATLAS. Promising decay channels for the search of heavy neutral Higgs bosons, like  $A \rightarrow Zh \rightarrow ll + \tau\tau$  and  $H \rightarrow hh \rightarrow \tau\tau + b\bar{b}$  involve a pair of boosted tau leptons.<sup>2</sup> In particular with the restart of LHC in 2015 with higher energies, such highly boosted topologies will become more likely and the current approaches of tau reconstruction and identification have to be reconsidered. In this thesis the current tau reconstruction is evaluated for highly boosted tau pairs. A new approach is introduced which is intended to extend the sensitivity for tau reconstruction to higher energies.

In Chapter 2 the theoretical foundations of the Standard Model and Two-Higgs-Doublet Models are introduced. In Chapter 3 an overview of the LHC and the ATLAS detector is given. The approach of single-tau reconstruction and identification is introduced in Chapter 4. There are also the results for tau reconstruction in the  $A \rightarrow Zh \rightarrow ll + \tau\tau$  decay channel discussed. In Chapter 5 the new di-tau reconstruction and identification method is introduced and evaluated for highly boosted tau pairs. And in Chapter 6 an outlook for the di-tau tagging is given and several future improvements are proposed.

---

<sup>1</sup> In this thesis natural units are used, see Appendix A.

<sup>2</sup> Here and in the remainder of this thesis, with  $l$  only the lighter leptons electron and muon are meant.

# Chapter 2

## The Standard Model of Particle Physics and Heavy Higgs Bosons

### 2.1 The Standard Model of Particle Physics

The Standard Model of particle physics (SM) is the most successful theory in explaining a wide range of phenomena in particle physics [11]. It has been tested in the last 40 years, convincing not only in its predictions of, before then, unknown particles like the gauge bosons,  $W$  and  $Z$ , and the Higgs boson  $h$ , but also in the high accordance of SM parameters with experimental observations.

This section serves as an overview of the particles predicted by the SM and introduces the important Higgs mechanism, which leads to the prediction of a Higgs boson. Beside its success, there are phenomena that can not be explained by the SM. These are addressed in the last part of this section.

#### 2.1.1 The Particle Content of the Standard Model

The Standard Model of particle physics describes the particles and their interactions via electromagnetic, weak and strong force, as a gauge quantum field theory. It describes quantum chromodynamics (QCD) and the unified electroweak interaction based on an overall gauge group of  $SU(3)_C \times SU(2)_L \times U(1)_Y$ .

The first group represents QCD. The index  $C$  stands for color, the charge in strong interactions. There are three color charges, called red, blue and green, with respectively one opposite anti-color. The strong interaction is mediated by eight different spin-one gauge bosons, called gluons, coupling to color-charged spin-one-half quarks. It is assumed that free particles have to be color-less, since quarks only appear in groups of two or three, with a vanishing net color-charge. No colored free particle have been observed so far.

$SU(2)_L \times U(1)_Y$  denotes the electroweak interaction between left-handed particles

with a hypercharge  $Y$ . It is the unification of quantum flavor dynamics (QFD) and quantum electrodynamics (QED). The hypercharge is the conserved quantity in the unified electroweak interaction.

$$Y = 2(Q - I_3^W) \quad (2.1)$$

The interacting fields are isospin-one-half doublets, where each component has a electromagnetic charge  $Q$  and respectively an opposite isospin  $I_3^W = \pm 1/2$ . This group is spontaneously broken in vacuum (see also Section 2.1.2), leading to three massive spin-one gauge bosons  $W^\pm$  and  $Z$  and a massless spin-one photon  $\gamma$ , which mediate the weak and electromagnetic force between spin-one-half fermions.

Besides the gauge bosons, which are summarized in Table 2.1, there are 12 different fermions, with respectively one anti-particle, known in the SM. They can be divided in quarks, which have the potential to interact strongly, and leptons, which can only interact via weak and potentially electromagnetic force. They can be grouped in weak isospin pairs and divided in three families. Corresponding particles of different families have the same characteristics, except the mass. As leptons there are electrons, muons and tauons, with respectively one neutrino in each isospin pair. As quarks, there is one up-type and one down-type quark in each family, namely up and down in the first family, charm and strange in the second one and top and bottom in the third family. The charges of fermions are summarized in Table 2.2.

There is one particle left, predicted by the SM: the spinless Higgs boson. The theoretical foundations of this particle are in the spontaneously broken electroweak interaction, which is discussed in the next section.

**Table 2.1:** The gauge bosons of the SM [16].

Interaction	Coupling to	Gauge Boson	Mass	Spin
strong	color	8 gluons $g$	0	1
electromagnetic	electric charge	photon $\gamma$	0	1
weak	weak isospin	$W^\pm$ $Z$	80.4 GeV 91.2 GeV	1 1

**Table 2.2:** The fermions of the SM [16].

Fermions	Family			Electric Charge	Color	Weak Isospin		Spin
	1	2	3			left-handed	right-handed	
Leptons	$\nu_e$	$\nu_\mu$	$\nu_\tau$	0	-	1/2	-	1/2
	$e$	$\mu$	$\tau$	-1	-		0	
Quarks	$u$	$c$	$t$	+2/3	r, b, g	1/2	0	1/2
	$d$	$s$	$b$	-1/3			0	

### 2.1.2 The Higgs Mechanism

The Higgs mechanism was initially developed by Robert Brout, François Englert, Peter Higgs, Gerald Guralnik, Carl Hagen and Tom Kibble [17, 18, 19, 20, 21]. Later it was incorporated by Abdus Salam and Steven Weinberg into the unified model of electroweak interactions, which was developed by Sheldon Lee Glashow [22, 23, 24].

An important characteristic of gauge theories like the SM is the invariance of the Lagrangian under local gauge transformations of fields  $\psi \rightarrow \psi' = \mathcal{U}(x^\mu)\psi$ . In  $SU(2)_L$  left-handed particle fields are represented by doublets like  $(\nu_e, e)_L$  or  $(u, d)_L$ . The problem is that it is not possible to introduce a mass term into the Lagrangian for such isospin doublets, which also fulfills local gauge invariance. This is in great divergence to experimental observations, where for example the tau lepton has a well measured mass of 1.78 GeV.

To keep the appealing symmetries of the gauge groups, but also to fulfill experimental reality, a spontaneous symmetry breaking is assumed in the Higgs mechanism. Consequently the unified electroweak gauge group  $SU(2)_L \times U(1)_Y$  does not refer to a state of lowest energy. In this vacuum state, or state of lowest energy, the electroweak symmetry is broken into a  $U(1)_Q$  symmetry, which only includes the conservation of charge in QED. Therefore an additional  $SU(2)$  Higgs doublet  $\Phi = (\phi_1, \phi_2)$  and a Higgs potential  $V(\Phi)$  is introduced into the SM Lagrangian. This extra term shifts the vacuum expectation value to an non-zero value  $v$  and adds couplings of the Higgs field to the gauge field, to fermions and to itself. In the new ground state this leads to mass terms for all particles the Higgs field couples to. Accordingly, the electroweak gauge field has to be redefined, to achieve a zero coupling to massless photons and a non-zero coupling to massive  $W^\pm$  and  $Z$  bosons. The mass of a particle is then determined by the strength of the coupling to the Higgs field. The couplings of the Higgs field to fermions, as well as to itself are free parameters of the SM. The masses of  $W^\pm$  and  $Z$  bosons are correlated via the weak angle  $\cos \theta_W = M_{W^\pm}/M_Z$  at tree level. The vacuum expectation value can be determined in muon decays to  $v = 246$  GeV [4].

### 2.1.3 Problems of the Standard Model

Despite the fact that the SM can explain most of the experimental observations of electroweak and strong interactions that are available today, it has its validity only in a certain energy range and there are theoretical challenges, which motivate extensions of the SM [25]. Examples for the most urgent problems that call for new physics explanations will be introduced here.

The SM describes strong and electroweak interactions as a direct product of three gauge groups  $SU(3)_C \times SU(2)_L \times U(1)_Y$ , but it lacks an explanation for the

origin of these three forces. Consequently it has several free parameters and can not predict Higgs couplings to fermions or the fermion masses. Moreover the fourth fundamental force, gravity, can also not be incorporated yet. Consequently it would be appealing to have a theory that unifies the forces in a more fundamental Grand Unified Theory.

Astrophysical observations have shown the presence of matter in the universe that does not interact via strong<sup>1</sup> or electromagnetic force, but makes up approximately one fourth of the energy of the universe, called Dark Matter. The SM can not provide any candidate particle for this phenomenon.

A third problem is the so called fine-tuning problem [26, 27, 28], that came with calculations of radiative corrections to the Higgs boson mass. A consistent theory, with Higgs boson masses at the electroweak scale, requires an extensive fine adjustment of parameters, which seems very unnatural. The problem can be solved with the assumption that there are additional scalar particles, like they are predicted by models of Supersymmetry [25].

## 2.2 Two-Higgs-Doublet Models and Heavy Higgs Bosons

In the SM it is assumed that there is one Higgs doublet field resulting in one Higgs boson. But in fact the spontaneous symmetry breaking and the explanations for particle masses can also be fulfilled with more than one Higgs doublet. Indeed several extensions of the SM, designed to explain shortcomings of the SM mentioned above, like the Minimal Supersymmetric Extension of the SM (MSSM) [29], twin Higgs models [30] and composite Higgs models [31] demand two Higgs doublets,

$$H_1 = \begin{pmatrix} H_1^0 \\ H_1^- \end{pmatrix}, \quad H_2 = \begin{pmatrix} H_2^+ \\ H_2^0 \end{pmatrix}, \quad (2.2)$$

with opposite hypercharges  $Y = \pm 1$  leading to three uncharged Higgs bosons  $h, H, A$  and two charged Higgs bosons  $H^\pm$ . The theoretical foundations of two-Higgs-doublet models (2HDMs) [32] are introduced in the following and it is shown which predictions can be made for heavy uncharged Higgs bosons.

### 2.2.1 The Four 2HDM Types

In principle 2HDMs have a large parameter space, though it can be decreased by several motivated assumptions [33]. For example it is reasonable to assume that the

---

<sup>1</sup> Precisely, the interaction is so weak, that it is not noticeable yet.

symmetry of charge and parity, CP, is not violated and that there are no flavor-changing neutral currents. This means that a coupling to only one of two doublets is allowed for each fermion type. Also it is assumed that the light Higgs boson,  $h$ , behaves SM like and has a mass of approximately 126 GeV to satisfy the observation of a Higgs boson, predicted by the Standard Model. If additionally the remaining Higgs bosons have masses at the same or a higher scale and are approximately similar  $m_H \approx m_A \approx m_{H^\pm}$ , it is assured that  $h$  decay modes only involve SM states.

The four types of 2HDMs (see Table 2.3) are distinguished by the coupling of the Higgs doublets to up- and down-type quarks and charged leptons. Type 1 demands all fermions to couple to the same doublet. By convention the up-type quarks always couple to  $H_2$  in all four 2HDM types. Type 2 corresponds to the MSSM, mentioned above. Here the up-type quarks couple to one doublet and the down-type quarks and charged leptons couple to the other. Type 3 is also called “lepton-specific” and demands up- and down-type quarks to couple to the same doublet and charged leptons to couple to the other. In Type 4, called “flipped”, up- and down-type quarks couple to separate doublets, and charged leptons couple to the same Higgs doublet as up-type quarks.

**Table 2.3:** Four types of 2HDMs.

Type	up-type quarks couple to	down-type quarks couple to	charged leptons couple to
1	$H_2$	$H_2$	$H_2$
2	$H_2$	$H_1$	$H_1$
3	$H_2$	$H_2$	$H_1$
4	$H_2$	$H_1$	$H_2$

This study is focused on the 2HDM type 2, the MSSM. In the MSSM as well as in 2HDM type 3, due to the coupling to  $H_1$  decays into charged leptons are enhanced, which is shown later. Since 2HDM type 1 and 3 as well as type 2 and 4 have respectively the same couplings to up- and down-type quarks, they show each a similar behavior in the main production processes.

### 2.2.2 Electroweak Symmetry Breaking

Like in the Standard Model the electroweak symmetry  $SU(3)_C \times SU(2)_L \times U(1)_Y$  is broken into  $U(1)_Q$ . With two scalar Higgs doublets,  $H_1$  and  $H_2$ , there are consequently two orthogonal vacuum states, where the symmetry is broken:

$$H_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} v_1 \\ 0 \end{pmatrix}, \quad H_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v_2 \end{pmatrix}, \quad (2.3)$$

leading to two vacuum expectation values  $v_1$  and  $v_2$ . To cover SM predictions that were already confirmed, the vacuum expectation values have to match the SM value via

$$v^2 = v_1^2 + v_2^2. \quad (2.4)$$

Though the ratio of 2HDM vacuum expectation values is still a free parameter. Commonly it is parameterized by

$$\tan \beta = \frac{v_2}{v_1}. \quad (2.5)$$

The two complex Higgs doublets, in Equation 2.2, have in total 8 degrees of freedom. Three of them are fixed for the appropriate polarization states of the gauge bosons  $W^\pm$  and  $Z$ . The remaining five degrees of freedom yield five Higgs bosons. To obtain the physical Higgs states,  $H_1$  and  $H_2$  have to be expanded around the vacuum state and divided in real parts,  $H_{1,2}^0$ , and in imaginary parts,  $P_{1,2}^0$ :

$$H_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} v_1 + H_1^0 + iP_1^0 \\ H_1^- \end{pmatrix}, \quad H_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} H_2^+ \\ v_2 + H_2^0 + iP_2^0 \end{pmatrix}. \quad (2.6)$$

The real parts in  $H_1$  and  $H_2$  represent CP-even Higgs bosons. The imaginary parts corresponds to a CP-odd Higgs boson plus a Goldstone boson, which does not represent a physical state. The actual observable mass eigenstates are given as a mixture of the Higgs field components:

$$\begin{pmatrix} h \\ H \end{pmatrix} = \begin{pmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{pmatrix} \begin{pmatrix} H_2^0 \\ H_1^0 \end{pmatrix}, \quad (2.7)$$

$$\begin{pmatrix} G^0 \\ A \end{pmatrix} = \begin{pmatrix} \cos \beta & \sin \beta \\ -\sin \beta & \cos \beta \end{pmatrix} \begin{pmatrix} P_1^0 \\ P_2^0 \end{pmatrix}, \quad (2.8)$$

$$\begin{pmatrix} G^\pm \\ H^\pm \end{pmatrix} = \begin{pmatrix} \cos \beta & \sin \beta \\ -\sin \beta & \cos \beta \end{pmatrix} \begin{pmatrix} H_1^\pm \\ H_2^\pm \end{pmatrix}. \quad (2.9)$$

This leads to five Higgs bosons: two CP even scalar neutral Higgs bosons  $h$  and  $H$ , where  $H$  is by convention the heavier one, a CP odd pseudoscalar neutral  $A$  boson and two charged Higgs bosons  $H^\pm$ . Additionally there are three Goldstone Bosons,  $G^0$  and  $G^\pm$ , which do not correspond to a physical state.

The mixing angle  $\alpha$  in Equation 2.7 is given by

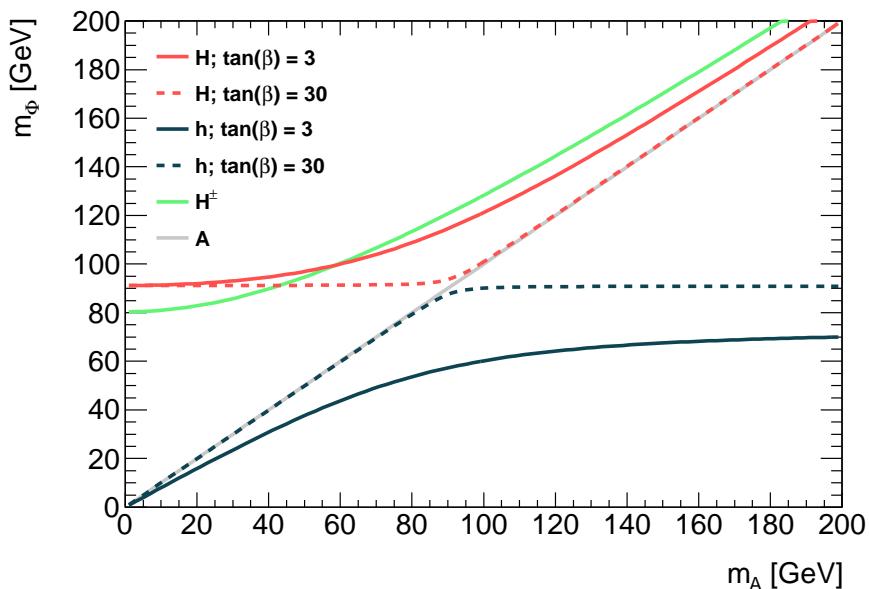
$$\tan 2\alpha = \tan 2\beta \frac{M_A^2 + M_Z^2}{M_A^2 - M_Z^2} \quad \text{with} \quad -\frac{\pi}{2} < \alpha < 0. \quad (2.10)$$

It can be shown that the boson masses are correlated via

$$M_{H^\pm}^2 = M_A^2 + M_W^2, \quad (2.11)$$

$$M_{H,h}^2 = \frac{1}{2} \left( M_A^2 + M_Z^2 \pm \sqrt{(M_A^2 + M_Z^2)^2 - 4M_A^2 M_Z^2 \cos^2 2\beta} \right). \quad (2.12)$$

Thus the masses of the Higgs bosons are at tree level fully determined by the parameters  $\alpha$  and  $\beta$ . Often the dependencies of the Higgs boson masses are parameterized in terms of  $\tan \beta$  and  $m_A$ , which is shown in Figure 2.1. Also, higher order corrections by virtual particle loops have an important contribution to the Higgs masses. These are dependent on the model; for the MSSM case this is discussed e.g. in [25].



**Figure 2.1:** Higgs bosons masses at tree level as a function of the  $A$  mass for  $\tan \beta = 3$  and  $\tan \beta = 30$ .

### 2.2.3 Gauge Boson and Fermion Couplings to 2HDM Higgs Bosons

Also the couplings of a single Higgs boson to vector bosons and to fermions are at tree level fully determined by the angles  $\alpha$  and  $\beta$  [33]. The coupling strengths compared to the SM predictions are shown in Table 2.4. The coupling strengths of  $h$  and  $H$  to vector bosons are respectively proportional to  $\sin(\beta - \alpha)$  and  $\cos(\beta - \alpha)$  in all four 2HDM types. The decay of an  $A$  boson into two vector bosons would violate CP conservation, therefore their coupling strength is zero.

Higgs-fermion couplings are dependent on the 2HDM type. Coupling strengths to fermions that coupled to the initial  $H_2$  field, are enhanced with  $\tan \beta$ . Experimental observations suggest a light Higgs boson  $h$  with couplings similar to the SM

predictions. This is the case in the alignment limit,  $\sin(\beta - \alpha) = 1$ .

**Table 2.4:** Coupling strengths of vector bosons, up- and down-type quarks, as well as charged leptons to a single Higgs boson, compared to the SM value [33]. Here,  $s_\theta$ ,  $c_\theta$  and  $t_\theta$  are abbreviations of  $\sin(\theta)$ ,  $\cos(\theta)$  and  $\tan(\theta)$ .

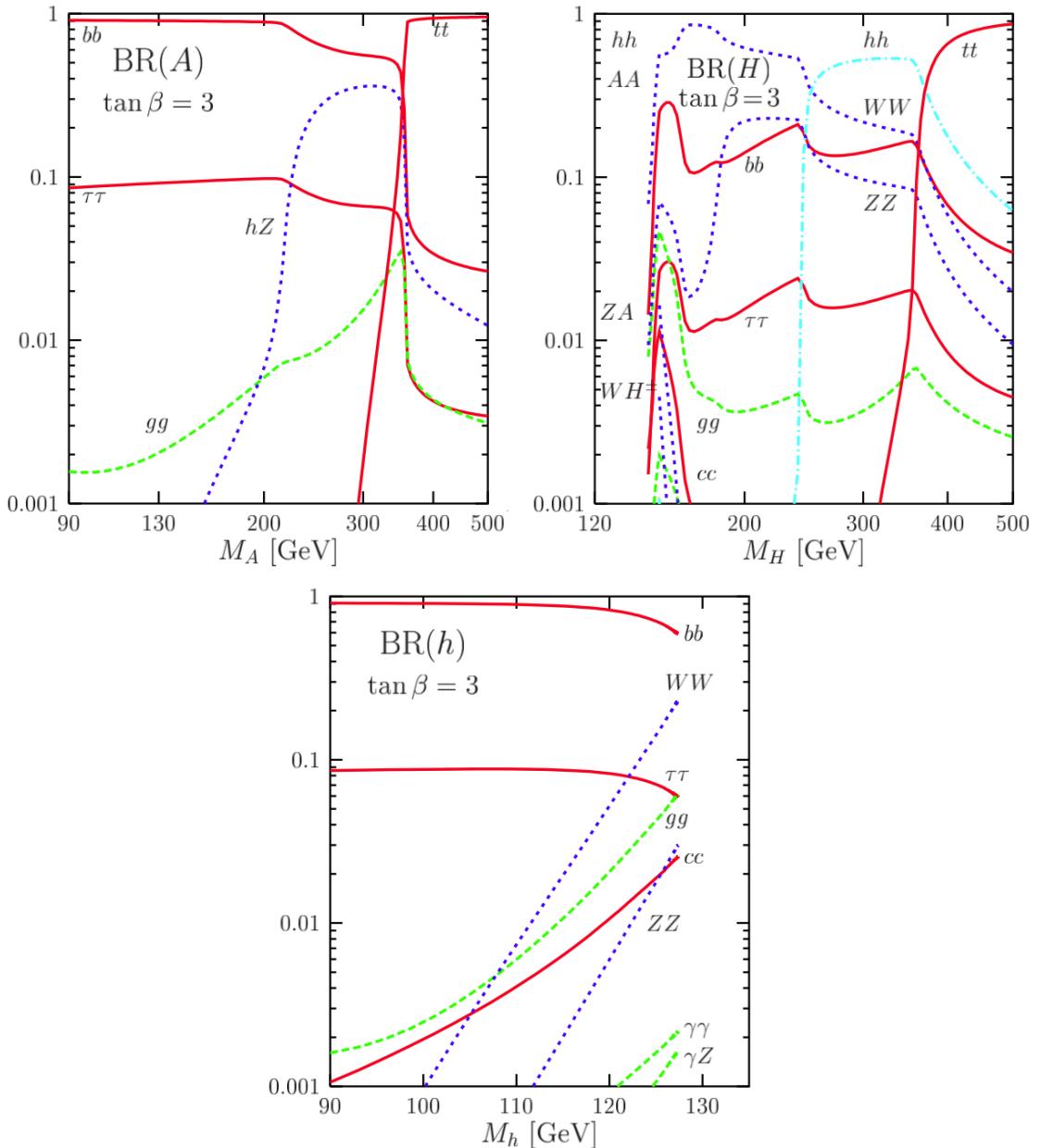
$y_{\text{2HDM}}/y_{\text{SM}}$	2HDM 1	2HDM 2	2HDM 3	2HDM 4
$hVV$	$s_{\beta-\alpha}$	$s_{\beta-\alpha}$	$s_{\beta-\alpha}$	$s_{\beta-\alpha}$
$hu\bar{u}$	$s_{\beta-\alpha} + c_{\beta-\alpha}/t_\beta$	$s_{\beta-\alpha} + c_{\beta-\alpha}/t_\beta$	$s_{\beta-\alpha} + c_{\beta-\alpha}/t_\beta$	$s_{\beta-\alpha} + c_{\beta-\alpha}/t_\beta$
$hdd\bar{d}$	$s_{\beta-\alpha} + c_{\beta-\alpha}/t_\beta$	$s_{\beta-\alpha} - t_\beta c_{\beta-\alpha}$	$s_{\beta-\alpha} + c_{\beta-\alpha}/t_\beta$	$s_{\beta-\alpha} - t_\beta c_{\beta-\alpha}$
$he\bar{e}$	$s_{\beta-\alpha} + c_{\beta-\alpha}/t_\beta$	$s_{\beta-\alpha} - t_\beta c_{\beta-\alpha}$	$s_{\beta-\alpha} - t_\beta c_{\beta-\alpha}$	$s_{\beta-\alpha} + c_{\beta-\alpha}/t_\beta$
$HVV$	$c_{\beta-\alpha}$	$c_{\beta-\alpha}$	$c_{\beta-\alpha}$	$c_{\beta-\alpha}$
$Hu\bar{u}$	$c_{\beta-\alpha} - s_{\beta-\alpha}/t_\beta$	$c_{\beta-\alpha} - s_{\beta-\alpha}/t_\beta$	$c_{\beta-\alpha} - s_{\beta-\alpha}/t_\beta$	$c_{\beta-\alpha} - s_{\beta-\alpha}/t_\beta$
$Hdd\bar{d}$	$c_{\beta-\alpha} - s_{\beta-\alpha}/t_\beta$	$c_{\beta-\alpha} + t_\beta s_{\beta-\alpha}$	$c_{\beta-\alpha} - s_{\beta-\alpha}/t_\beta$	$c_{\beta-\alpha} + t_\beta s_{\beta-\alpha}$
$He\bar{e}$	$c_{\beta-\alpha} - s_{\beta-\alpha}/t_\beta$	$c_{\beta-\alpha} + t_\beta s_{\beta-\alpha}$	$c_{\beta-\alpha} + t_\beta s_{\beta-\alpha}$	$c_{\beta-\alpha} - s_{\beta-\alpha}/t_\beta$
$AVV$	0	0	0	0
$Au\bar{u}$	$1/t_\beta$	$1/t_\beta$	$1/t_\beta$	$1/t_\beta$
$Ad\bar{d}$	$-1/t_\beta$	$t_\beta$	$-1/t_\beta$	$t_\beta$
$Ae\bar{e}$	$-1/t_\beta$	$t_\beta$	$t_\beta$	$-1/t_\beta$

Decays of heavy uncharged Higgs bosons often include one or two additional light Higgs bosons. When kinematically allowed and  $m_h < m_A \approx m_H$ , the processes  $A \rightarrow Zh$ ,  $H \rightarrow hh$  as well as  $H \rightarrow VV$  may be available. The coupling strengths in these three processes are proportional to  $\cos(\beta - \alpha)$  [33]. They are therefore forbidden in the exact alignment limit. Nevertheless these processes give an important contribution to all Higgs decays at even small deviations from the exact alignment limit. The coupling of an  $A$  to  $Z$  and  $h$  is given by

$$g_{hZA} = \frac{1}{2} \sqrt{g^2 + g'^2} \cos(\beta - \alpha). \quad (2.13)$$

The coupling strength  $g_{Hhh}$  is dependent on the 2HDM type, with terms not predicted by  $\alpha$  and  $\beta$ , but it can be noted that the branching ratio of  $H \rightarrow hh$  dominates over  $H \rightarrow VV$  in many cases. A more detailed discussion can be found in [33].

When kinematically available the decays into  $b\bar{b}$  and  $t\bar{t}$  give high contributions to the total widths of Higgs decays, but are experimentally hard to distinguish from prompt QCD jets. The branching ratios of Higgs boson decays are shown in Figure 2.2 for the 2HDM type 2. It can be seen that  $A \rightarrow Zh$  as well as  $H \rightarrow hh$  have significant contributions to the branching ratio in the high mass region. The resulting boosted light Higgs bosons can afterwards decay into a boosted pair of tau leptons in the final state, giving the event signature discussed in this study.



**Figure 2.2:** MSSM branching ratios of uncharged Higgs bosons,  $A$ ,  $H$  and  $h$ , versus their masses. It has to be noted that with a fixed  $\tan \beta = 3$  the mixing angle  $\alpha$  changes with the illustrated mass of the higgs bosons. [25]

# Chapter 3

## The LHC and the ATLAS Detector

### 3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [34] is a circular superconducting hadron accelerator and collider at the European Organization for Nuclear Research (CERN). It is situated at the Franco-Swiss border, near Geneva. The LHC is installed in a tunnel with a circumference of 26.7 km, which is located between 45 m and 170 m below the ground level and was formerly used by the Large Electron Positron Collider (LEP) [35]. It is designed to search for new physics and to test the Standard Model at high energies. To discover rare processes a high event rate  $dN/dt$  is needed. This depends on the physical cross section  $\sigma$  of the process and the detector parameter  $L$ , called luminosity:

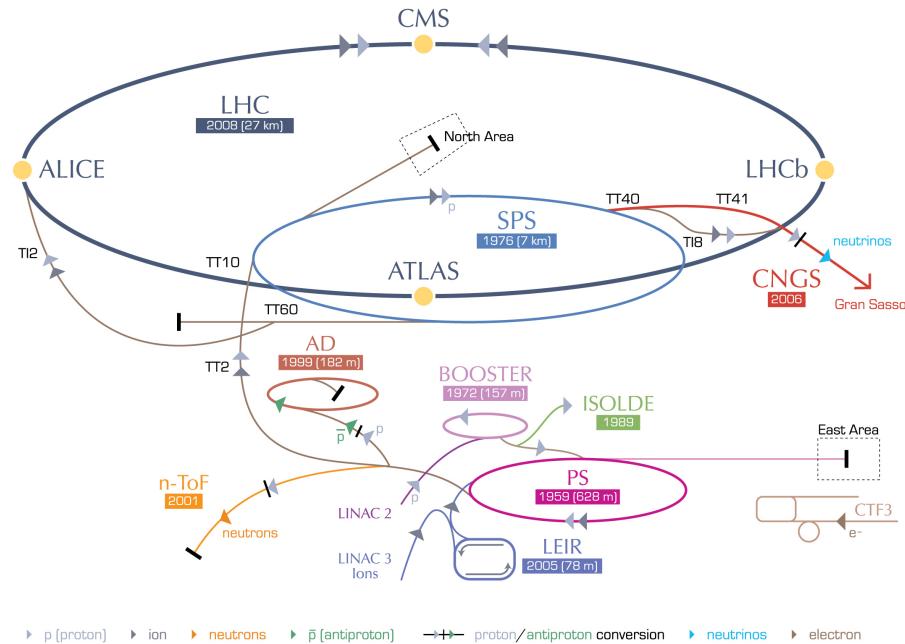
$$\frac{dN}{dt} = L \cdot \sigma. \quad (3.1)$$

At the interaction points proton beams can be collided with a center-of-mass energy of 14 TeV, and peak luminosities of  $10^{34} \text{ cm}^{-2}\text{s}^{-1}$ . Also heavy lead ions can be accelerated up to an energy of 2.8 TeV per nucleon with a peak luminosity of  $10^{27} \text{ cm}^{-2}\text{s}^{-1}$ .

The LHC is the largest one in a complex of accelerators at CERN. The accelerator chain is shown on Figure 3.1. Starting with hydrogen atoms, their orbiting electrons are stripped away, leaving single protons, which are then accelerated to an energy of 50 MeV by Linac2. Afterwards the protons are injected into the Proton Synchrotron Booster (PSB) and accelerated to an energy of 1.4 GeV. The next links in the chain are the Proton Synchrotron (PS) with an energy of 25 GeV and the Super Proton Synchrotron (SPS) with an energy of 450 GeV. Finally the protons are fed into the LHC, separated in two rings with clockwise and anticlockwise directions, and accelerated to energies of up to 7 TeV. For equally charged hadron beams, two rings

with separate magnetic bending systems are needed. With about 9600 magnets the particles' trajectories are optimized at LHC. With 1232 superconducting dipole magnets, kept at a temperature of 1.9 K the beams are bent with magnetic fields of up to 8.3 T. In particular the insertion quadrupoles play an important role in focusing the beam at the collision point, enhancing the probability of proton-proton interactions by decreasing the beam size.

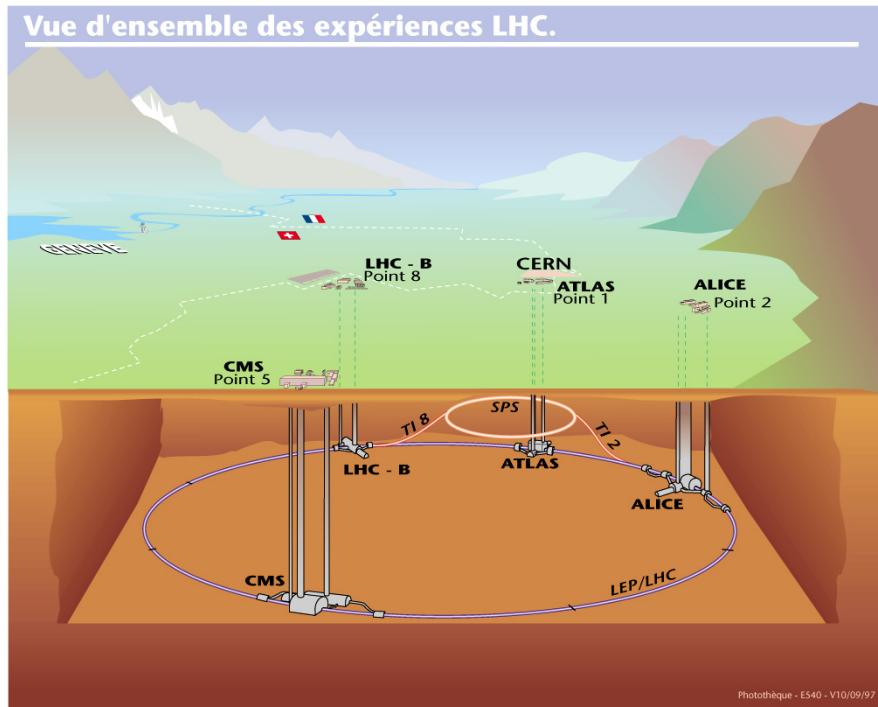
The LHC is separated in eight arcs, where the beams are bent by the dipole magnets, and eight long straight sections. On four of the straight sections the beams can be guided to collision. The four main experiments – ALICE [36], ATLAS [14], CMS [15] and LHCb [37] – are located at the beam crossing points. A schematic overview of the LHC experiments is provided in Figure 3.2. ATLAS and CMS are the two general-purpose detectors, designed to investigate the widest range of physics at high luminosities. LHCb is specialized in  $b$  quark physics and in determining CP-violations. ALICE analyzes heavy lead ion collisions and is focused on investigating properties of strongly interacting matter at high energies.



**Figure 3.1:** Schematic view of the CERN accelerator complex [38].

## 3.2 The ATLAS Detector

The ATLAS Detector is one of two multi-purpose detectors at the LHC. It weighs 7000 t and is with a length of 46 m and a diameter of 25 m the largest detector on a collider ever constructed. The ATLAS detector is build forward-backward symmetric, with layers of components located cylindrically around the beam axis



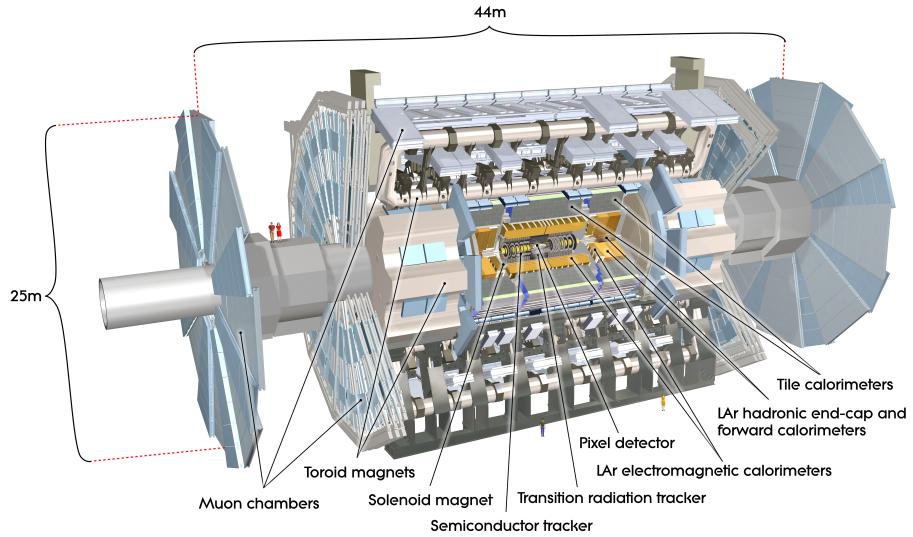
**Figure 3.2:** Schematic view of LHC experiments [39].

and two disk-shaped end-cap regions. A schematic overview of the ATLAS Detector is given in Figure 3.3. The ATLAS experiment consists of four major detector components: an inner detector, a calorimeter system, a forward detector and a muon spectrometer. A large magnet system is installed to bend the trajectories of charged particles and a trigger system to handle the enormous data flow.

The Higgs boson discovery in July 2012 [12] was one important benchmark to demonstrate the performance of the ATLAS experiment. Now its parameters are to be measured at a high level of precision. Furthermore with a increase of the center-of-mass energy at LHC to 13 TeV in 2015 and a data taking with higher integrated luminosities, the ATLAS experiment has the potential to observe new phenomena at the TeV scale. For example heavy neutral Higgs bosons,  $A$  and  $H$ , and charged Higgs bosons,  $H^\pm$ , predicted by Two-Higgs-Doublet models, could be accessible (see Chapter 2). New heavy gauge bosons  $W'$  and  $Z'$  could be discovered and QCD expectations can be studied with high- $p_T$  jet measurements. The detector components of the ATLAS experiment are designed for such measurements from the beginning and will be introduced in the following.

### 3.2.1 The ATLAS Coordinate System

The origin of the ATLAS coordinate system is set to the proton-proton collision point in the center of the detector. The  $z$ -axis of a right handed cartesian coordinate system is defined along the beam axis. The positive  $x$ -axis is pointing in the direction



**Figure 3.3:** Schematic overview of the full ATLAS Detector and its sub-systems [14].

of the the center of the LHC ring and the positive  $y$ -axis is orientated upwards to the earth's surface. The  $x$ - $y$ -plane is transverse to the beam pipe. Transverse variables, like the transverse momentum  $p_T$  are projections into this plane. The azimuthal angle  $\phi$  is the angle around the beam axis and the polar angle  $\theta$  is measured from the beam axis. Instead of  $\theta$ , the pseudorapidity  $\eta$ , defined as

$$\eta = -\ln \left( \tan \frac{\theta}{2} \right), \quad (3.2)$$

is commonly used as a particle coordinate. In the case of negligible particle masses,  $\eta$  converges to the Lorentz invariant rapidity  $y$

$$y = \frac{1}{2} \ln \left( \frac{E + p_z}{E - p_z} \right). \quad (3.3)$$

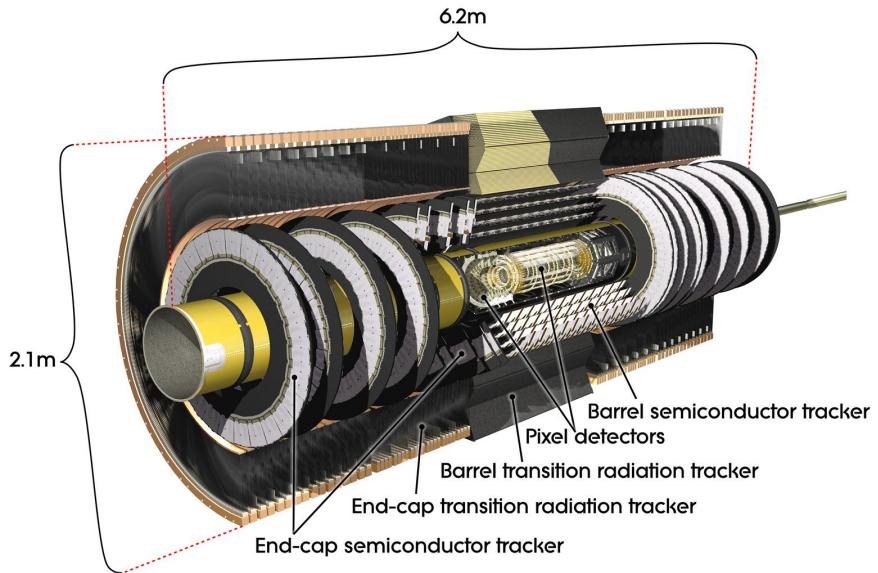
To measure the distance between two particles, the pseudorapidity-azimuthal angle distance  $\Delta R$  is used, defined as

$$\Delta R = \sqrt{(\eta_1 - \eta_2)^2 + (\phi_1 - \phi_2)^2}. \quad (3.4)$$

### 3.2.2 The Inner Detector

The inner detector is the innermost component of the ATLAS detector, with the closest distance of 5 cm to the beam axis. It has a total length of 6.2 m and a diameter of 2.1 m [14]. The inner detector and its sub-systems are illustrated in Figure 3.4. Its tracking system is designed to deliver a very efficient momentum and vertex

reconstruction. A high primary and secondary vertex measurement performance in a range of  $|\eta| < 2.5$  enhances for example the efficiency of tau lepton and bottom quark recognition. In a range of  $|\eta| < 2.0$  a reliable transition radiation detection is performed, which is essential for electron identification. This is achieved by a combination of high granularity and precision measurements close to the interaction region and robust pattern recognition and momentum resolutions with sub-detectors in its outer part. The inner detector consists of three independent components: the silicon pixel detector, the semiconductor tracker (SCT) and the transition radiation tracker (TRT). The central solenoid, surrounding the inner detector, immerses the sub-detectors in a 2 T solenoidal magnetic field, allowing measurements of the charge and momentum of charged particles.



**Figure 3.4:** Overview of the ATLAS Inner Detector [14].

### The Pixel Detector

The pixel detector is the system closest to the interaction region. It shows the highest granularity and covers  $|\eta| < 2.5$ . Due to the small distance to the proton-proton collision point, the pixel sensors are manufactured with a high radiation hardness. Nevertheless they provide a good charge-collection efficiency at the same time. 1744 pixel modules, with a thickness of 250  $\mu\text{m}$ , are arranged in three layers of coaxial cylinders in the barrel region and in parallel disks in the end-cap regions. Generally all three layers are crossed by each charged particle. The modules can measure  $R - \phi$  and  $z$  or  $R$  independently with accuracies of 10  $\mu\text{m}$  ( $R - \phi$ ) and 115  $\mu\text{m}$  ( $z$ ) in the barrel region and of 10  $\mu\text{m}$  ( $R - \phi$ ) and 115  $\mu\text{m}$  ( $R$ ) in the end-caps region. The pixel detector provides a total number of approximately 80.4 million readout channels.

### The Semiconductor Tracker

The SCT uses 4088 modules of long and narrow sensor strips to cover a large area of  $63\text{ m}^2$ . It consists of four concentric layers in the barrel region and nine disk layers in the end-caps region. Each particle crosses typically eight strip layers, allowing the measurement of four space points with intrinsic accuracies of  $17\text{ }\mu\text{m}$  ( $R - \phi$ ) and  $580\text{ }\mu\text{m}$  ( $z$ ) per module in the barrel region and  $17\text{ }\mu\text{m}$  ( $R - \phi$ ) and  $580\text{ }\mu\text{m}$  ( $R$ ) per module in the end-caps. The SCT provides approximately 6.3 million readout channels.

### The Transition Radiation Tracker

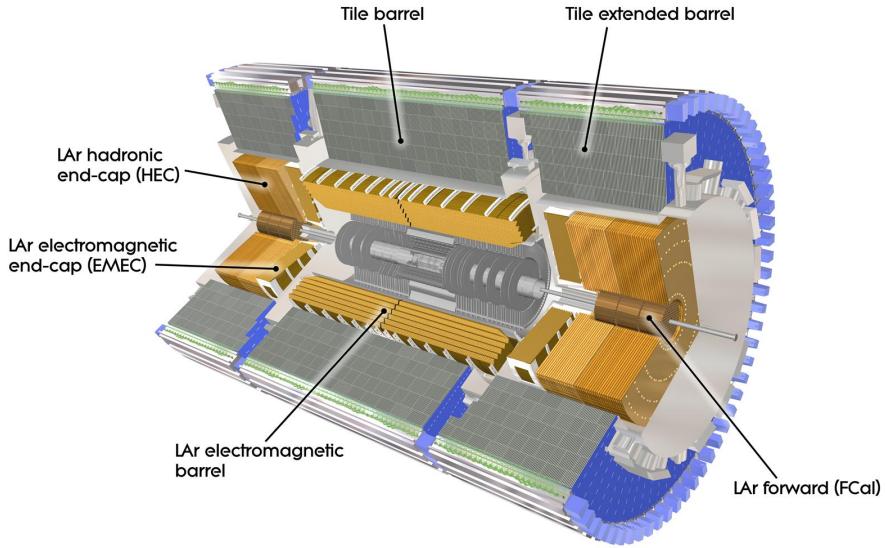
The outermost component of the inner detector, the TRT, covers a region of  $|\eta| < 2.0$ . It consists of 73 layers of drift tube straws in the barrel region and 160 straw planes in the end-caps. To provoke transition radiations of highly relativistic particles, the straws are interleaved with fibers in the barrel region and with foils in the end-caps, allowing transition radiation measurements for electron identification. The 4 mm straw tubes allow a track following in  $R - \phi$ , with typically 36 hits per charged particle and accuracies of  $130\text{ }\mu\text{m}$  per straw, on a total number of 351,000 readout channels.

### 3.2.3 The Calorimeter System

The calorimeter system, surrounding the inner detector components, is illustrated in Figure 3.5. It consists of two sampling calorimeter systems – the inner electromagnetic calorimeter and the outer hadronic calorimeter – measuring the amount and position of absorbed energy in a range of  $|\eta| < 4.9$ . Different techniques are used over the large range of  $|\eta|$ , to adapt to physical requirements in the region of interest. In the region  $|\eta| < 2.5$ , which is also covered by the inner detector, a high granularity is continued, ensuring high precision electron and photon measurements. The high  $\eta$ -coverage with a coarser granularity at  $|\eta| > 2.5$  allows a reliable jet reconstruction and missing transverse momentum measurement. A good containment for electromagnetic and hadronic showers, as well as a punch-through into the muon system is controlled by the calorimeters thickness.

### The Electromagnetic Calorimeter

The electromagnetic calorimeter absorbs electromagnetically interacting particles and consists of a barrel component, covering a region of  $|\eta| < 1.475$  and two end-cap parts, covering  $1.375 < |\eta| < 3.2$ . The barrel component is divided into two halves, with a small gap of 4 mm at  $z = 0$ . Each end-cap component consists of



**Figure 3.5:** Overview of the full calorimeter system [14].

two coaxial wheels. Lead is used as energy-absorbing material and liquid argon (LAr) as active sampling material. It has accordion-shaped electrodes and absorber plates, providing a complete  $\phi$  symmetry, with a precedence of high accuracy in the precision measurement area  $|\eta| < 2.5$ .

### The Hadronic Calorimeter

The hadronic calorimeter absorbs the energy of hadrons, like protons, neutrons or pions. It consists of three sub-components: the tile calorimeter, the LAr hadronic end-cap calorimeter and the LAr forward calorimeter.

The tile calorimeter is divided into a barrel region, covering  $|\eta| < 1.0$  and two extended barrel regions, covering  $0.8 < |\eta| < 1.7$ . Incoming particles interact with steel as absorbing material by producing a particle shower. Scintillating tiles are used to sample the deposited energy.

The hadronic end-cap calorimeter is divided into two wheels in each end-cap. With a coverage of  $1.5 < |\eta| < 3.2$  it slightly overlaps with the tile calorimeter and the forward calorimeter. Copper plates are used as absorbing material, with LAr gaps as sensitive medium for sampling.

### The Forward Calorimeter

The forward calorimeter provides the measurement of electromagnetic interactions, as well as hadronic interactions. It consists of three modules per end-cap. The first one is optimized for electromagnetic interactions and the other two measure hadronic interactions. Each one is using LAr as active material plus copper as absorber in

the first module and tungsten in the other two modules.

### 3.2.4 The Muon Spectrometer

Muons are the only charged particles known, where a relevant amount can pass the calorimeter systems with just a few interactions. Therefore the muon spectrometer is installed as the outermost component of the ATLAS detector. It consists of large superconducting air-core magnets to bend the muon trajectories, tracking chambers with high precision and a separate trigger system. The muon spectrometer is designed with a light and open structure, to avoid multiple-scattering. To achieve a high momentum resolution, three layers of tracking chambers are installed over high distances in the barrel and end-cap region.

Four different detector components are installed to track the muons coordinates. Monitored Drift Tubes provide a precision muon measurement over most of the  $|\eta|$  range. Cathode Strip Chambers are used at high pseudorapidities  $2.0 < |\eta| < 2.7$ , since they are able to measure with a higher granularity in conditions with more background events. For the trigger system, covering  $|\eta| < 2.4$ , Resistive Plate Chambers and Thin Gap Chambers provide high timing resolutions between 1.5 ns and 4 ns, respectively in the barrel region and in the end-caps.

### 3.2.5 The Forward Detectors

Three additional smaller detectors are associated as ATLAS forward detectors. Two of them – LUCID (Luminosity measurement using Cerenkov Integrating Detector) and ALFA (Absolute Luminosity for ATLAS) – are measuring the luminosity delivered to ATLAS. The system of the Zero-Degree Calorimeter (ZDC) determines heavy-ion collisions close to the beam line. LUCID monitors the relative luminosity via proton scatterings in a 17 m distance in both beam line directions. ALFA is situated at  $\pm 140$  m from the collision point and measures the absolute luminosity with scintillating fiber trackers inside Roman Pots. ZDC measures neutral particles at pseudorapidities of  $|\eta| < 8.2$  and distances of  $\pm 140$  m using layers of quartz rods and tungsten plates.

### 3.2.6 The Trigger System

The design luminosity of  $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  at ATLAS detector corresponds to an event data rate of 1 GHz, while current data recording techniques require rates of about 200 Hz. Therefore three levels of triggers are applied to filter physically interesting signatures in real time.

The Level-1 (L1) trigger reduces the data rate to approximately 75 kHz by searching for events with a high missing transverse momentum or total transverse momentum, as well as signatures with high momentum electrons, muons, photons, jets or tau leptons decaying into hadrons. The L1 trigger uses a sub set of information of the muon spectrometer and the calorimeter systems with a coarser granularity to make a decision and to define one or more Regions-of-Interest (ROI) passed to the next trigger level. The software-based Level-2 (L2) trigger uses full detector information in the ROI to reduce the event rate to approximately 3.5 kHz. Finally the Event Filter reduces the data rate to approximately 200 Hz by applying procedures similar to offline analyses.

# Chapter 4

## Single Tau Reconstruction and Identification

Tau Leptons are the heaviest leptons known, with a mass of 1.77682 GeV [4]. They decay almost immediately with a proper life time of 290.3 fs [4]. With a proper decay length of  $c\tau = 87.03 \mu\text{m}$  [4] most of the tau decays in ATLAS happen inside the beam pipe, before reaching any detector components and can only be reconstructed by their decay products. It is the only lepton heavy enough to decay in both lighter leptons and hadrons.

Tau leptons decay with a branching ratio of 35% in the leptonic mode via  $\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$  or  $\tau^- \rightarrow e^- \bar{\nu}_e \nu_\tau$ .<sup>1</sup> An interaction of neutrinos with the detector material is very unlikely. Only the charged particles and probably missing transverse momentum can be observed. Therefore the leptonic tau decay is hard to distinguish from prompt electrons or muons. This study will concentrate on the hadronic decay mode<sup>2</sup>, which has a branching ratio of 65%. Here the tau lepton decays into a tau neutrino and at least one charged meson, like a pion,  $\pi^\pm$ , or kaon,  $K^\pm$ , and possibly uncharged mesons. Tau leptons in the hadronic mode decaying into one charged particle are called *1-prong*, which has a probability of 85.3%. With a probability of 14.6% they decay into three charged particles, called *3-prong*.

A hadron collider experiment like ATLAS at LHC is dominated by quark and gluon jets (QCD jets). It is therefore important to differentiate these QCD jets from taus in the hadronic mode, which also form jets in the calorimeter. The most important difference is that the decay products of a  $\tau_{\text{had}}$  have a more collimated shape. This is due to a boosting via the intermediate  $W$  boson. QCD jets have in principle a wider shape with more particles, which, among other characteristics, is used to differentiate tau decays from QCD jets.

<sup>1</sup> The opposite charged  $\tau^+$  lepton decays analogously via  $\tau^+ \rightarrow \mu^+ \nu_\mu \bar{\nu}_\tau$  or  $\tau^+ \rightarrow e^+ \nu_e \bar{\nu}_\tau$ .

<sup>2</sup> A tau in hadronic decay mode and its decay products will also be referred to as  $\tau_{\text{had}}$  in the remainder of this thesis.

In the next sections the basic principles of tau reconstruction and background separation in the ATLAS experiment are introduced. Also the results of tau reconstruction are shown for the  $A \rightarrow Zh \rightarrow ll + \tau\tau$  topology.

## 4.1 Clustering

The basis of tau reconstruction are hadronic showers, formed in the calorimeter. When a particle deposits energy in the calorimeter, it does this usually in many cells. With clustering algorithms it is possible to group these cells into clusters and to sum up and calibrate their energy. A particle type specific energy calibration is needed to compensate for energy deposits out of the detector's active areas. Energy deposits have to be differentiated from background noise. Noise can arise from two different sources: directly from the readout electronics or from pile-up. Pile-up comes from either secondary interactions in the same proton bunch crossing or from primary interactions of previous proton bunch crossings.

Clustering in ATLAS is done by using a topological algorithm [40]. Starting with a seed cell, every neighboring cell is clustered iteratively on condition that it has a significantly higher energy than the expected noise. This algorithm performs very well in noise suppression and in handling a large number of cells. Thus it is used in jet and missing transverse energy reconstruction.

## 4.2 Jet Algorithms

There are two approaches for reconstructing jets: cone algorithms and sequential clustering algorithms. Cone algorithms follow the idea that particles of a jet deposit their energy within a circle of radius  $R$  in the plane of rapidity and azimuth  $y - \phi$ . Therefore such algorithms search for cones, where the summed up 4-momenta of its constituents are along the cone's axis (*stable cone*). For overlapping cones an additional splitting and merging procedure has to be implemented. Getting circular shaped jets is favorable in order to have a predictable outcome and makes it easy to subtract a uniform background. But usually jets from cone algorithms loose their circular area after the splitting and merging step. Another constraint jets should fulfill is infrared and collinear safety (IRC safety): an emission of a soft or collinear gluon should not change the jet. But most naive cone algorithms search only for cones around constituents with highest  $p_T$ , which is not a collinear safe concept. Although there are cone algorithms like SISCone (Seedless Infrared Safe Cone) [41], where all possible stable cones are considered, at the cost of higher calculation times.

Sequential clustering algorithms are reconstructing jets by a pairwise merging of constituents into proto-jets by comparing their distances and their energies. The

most common are:  $k_t$  [42, 43], Cambridge/Aachen [44, 45] and anti- $k_t$  [46]. These algorithms are IRC-safe by construction. Two measures of distances are defined:  $d_{ij}$  between two entities  $i$  and  $j$  and  $d_{iB}$  between one entity  $i$  and the beam  $B$ . If the smallest distance is a  $d_{ij}$  it recombines the entities  $i$  and  $j$ . If it is a  $d_{iB}$ , it removes  $i$  from the pool of available entities and calls it a jet. This procedure is continued until no particles are left or a stop criterion is fulfilled. The difference between the three algorithms is in how the distances  $d_{ij}$  and  $d_{iB}$  are defined. They can be generalized the with following formulas [46]:

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta^2}{R^2}, \quad (4.1)$$

$$d_{iB} = k_{ti}^{2p}, \quad (4.2)$$

where  $\Delta^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$  is the rapidity-azimuthal angle space between the particles,  $k_t$  is the transverse momentum and  $R$  a distance parameter, forcing every final pair of jets to have a minimum distance of  $\Delta_{ij} = R$ . The relative power of energy versus geometrical scales is determined by the parameter  $p$  and refers to a dedicated algorithm:  $p = 1$  for  $k_t$ ,  $p = 0$  for Cambridge/Aachen and  $p = -1$  for the anti- $k_t$  clustering algorithm.

With the positive momentum exponents of the  $k_t$  algorithm, it can be seen as the inversion of QCD branching processes. Cambridge/Aachen is based only on the spacial separation of the particles. It is used in  $b$  jet studies and shows good results in resolving jet substructure. The anti- $k_t$  algorithm with a negative power in transverse momentum behaves like an idealized cone algorithm, leading to conical shaped jets<sup>3</sup>. Such a predictable shape makes it easier to subtract a uniform particle background. Consequently it is well suited for tau reconstruction, where the jets are later separated in subcones as described in the next section.

### 4.3 Tau Reconstruction

The tau reconstruction [47] in ATLAS searches for signatures of hadronically decaying tau leptons. The very specific decay characteristics of a  $\tau_{\text{had}}$  is reconstructed by combining both tracking and calorimeter information provided by the detector. The energy information originates from calorimeter cells, merged to topological clusters. Jets are formed out of these clusters with an anti- $k_t$  algorithm with a distance parameter  $R = 0.4$ . These jets are used as seeds for tau reconstruction. Afterwards several corrections have to be applied. The primary vertex associated to the seed jet is not necessarily the origin vertex of the tau lepton, hence a new vertex hypothesis

<sup>3</sup> Conical jets are an idealization which is only fulfilled for jets with only soft fragmentation.

is calculated. The seed axis and 4-momentum are recalculated in the new tau vertex coordinate system. Also an energy calibration is performed by taking into account the specific mixture of charged and uncharged particles [48]. At last, tracks within a cone of  $\Delta R < 0.2$ , which fulfill various selection criteria [47], are associated to the tau candidate. They are used to classify the tau candidate as 1-prong (with one track), 3-prong (with 3 tracks), or multi-prong (with 2 or 3 tracks). Depending on the number of tracks, tau candidates are handled differently in the identification procedure later. Several variables are calculated for background rejection [47], referring to subcones in the tau candidate signature. The *centermost cone* is defined as the region within  $\Delta R < 0.1$ , where most of the tracks and energy deposits are expected. The *core cone*, defined by  $\Delta R < 0.2$ , describes the collimated tau signature. In the *isolation annulus* with  $0.2 < \Delta R < 0.4$ , only a few interactions are expected in contrast to QCD jets.

## 4.4 Tau Identification

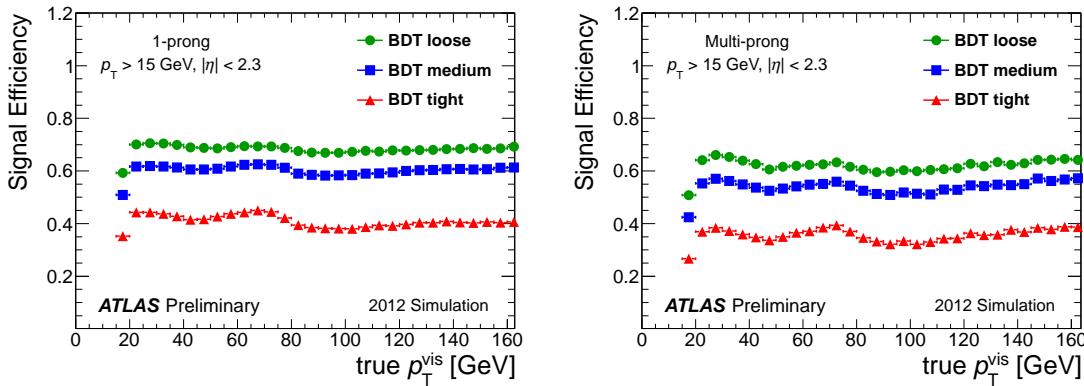
The tau identification performs the distinction of actual tau decays against background events. Due to the tuning of the reconstruction step to find as many taus as possible also many background events are passing. These are handled in this separate step. A  $\tau_{\text{had}}$  can be faked by electrons or muons, that where misinterpreted as jets or are inside a QCD jet. They can be treated via electron- and muon-vetos or an overlap removal, which excludes events with signatures of electrons or muons near jets.

The main source of background are quark- or gluon-initiated jets. These are produced at high rates at the LHC, with signatures, similar to a tau in hadronic decay mode. For discrimination a set of variables is calculated in the tau reconstruction, describing e.g. the width of energy deposits in the calorimeter, the number and position of tracks, as well as decay lengths of tau candidates. A full list of separation variables is given in [47]. These variables are used for a training of multivariate techniques, like a projective likelihood method [49] or boosted decision trees (BDTs) [50].<sup>4</sup> As  $\tau_{\text{had}}$  signal, events of simulated  $Z \rightarrow \tau\tau$ ,  $W \rightarrow \tau\nu_{\tau}$  and  $Z' \rightarrow \tau\tau$  decays are used. The last decay channel enhances the number of high- $p_{\text{T}}$  hadronic taus by simulating  $Z'$  masses between 250 GeV and 1250 GeV. Only correctly reconstructed tau candidates, which have a maximum angular distance of  $\Delta R = 0.2$  to the axis of the simulated visible tau, are called truth-matched and are considered for training. Jet enriched data events are taken for QCD background. The set of variables for training differs, depending on the number of tracks associated to the tau candidate (1-prong, 3-prong, multi-prong) to enlarge the sensitivity. The results

---

<sup>4</sup> The functionality of BDTs is also described in Chapter 5.

of the identification are evaluated by their signal and background efficiencies. The signal efficiency is the number of correctly identified  $\tau_{\text{had}}$  divided by all true tau decays. Analogously, the background efficiency is the number of background events mis-identified as  $\tau_{\text{had}}$  divided by the number of all background events used for testing. Next to providing high signal efficiencies and low background efficiencies, the sets of variables are as well designed to deliver efficiencies with a low pile-up and  $p_T$  dependency. This is supported by weighting the training events according to the number of primary vertices and the tau transverse momentum. The performance of the tau identification in Run-I is shown in Figure 4.1.



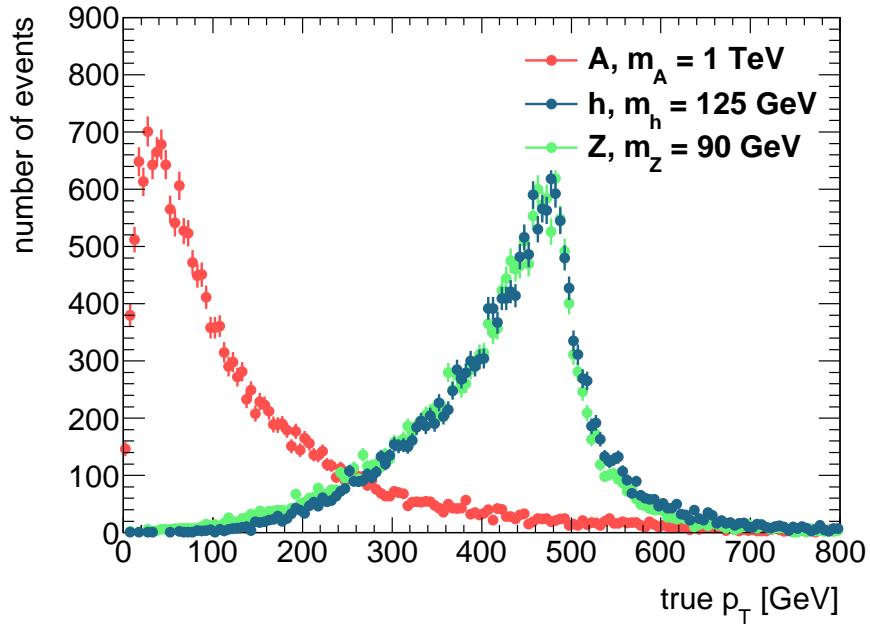
**Figure 4.1:** Signal efficiencies in Run-I for 1-prong (left) and multi-prong (right) tau candidates. The tau reconstruction and the tau identification is applied, with boosted decision trees for three different working points, called loose, medium and tight [51].

## 4.5 Reconstruction Results for Highly Boosted Di-Tau Topologies

Before coming to reconstruction results for highly boosted tau pair topologies, some general characteristics of  $A \rightarrow Z h$  decays have to be discussed. The events are generated with POWHEG BOX [52]. The parton showering is done by PYTHIA 8.1 [53]. In these simulated events the  $A$  boson is produced via gluon fusion with a mass of 1 TeV. Due to the high mass, the sample is enriched with high- $p_T$  tau pairs. The results of this chapter rely on 40.000 simulated events.

In Figure 4.2 the simulated true transverse momentum distributions for a heavy Higgs boson and its prompt decay products  $Z$  and  $h$  are shown. It can be seen that  $A$  bosons are not only produced with a high mass, resulting in boosted tau pairs, but also are produced with a high transverse momentum itself. This is due to initial state radiation. In gluon initiated production the likelihood for radiating gluons is enhanced in contrast to production processes with quarks, because of a higher

possibility of matching color charges. Also the gluon-gluon radiation depends on their relative momentum, making extra recoil jets likely for high energies. The decay to  $Z$  and  $h$  is equivalent in leading order except a small mass difference of both bosons. This results in a Jacobian peak like distribution for both  $Z$  and  $h$ , where approximately half of the energy is carried away by each boson.



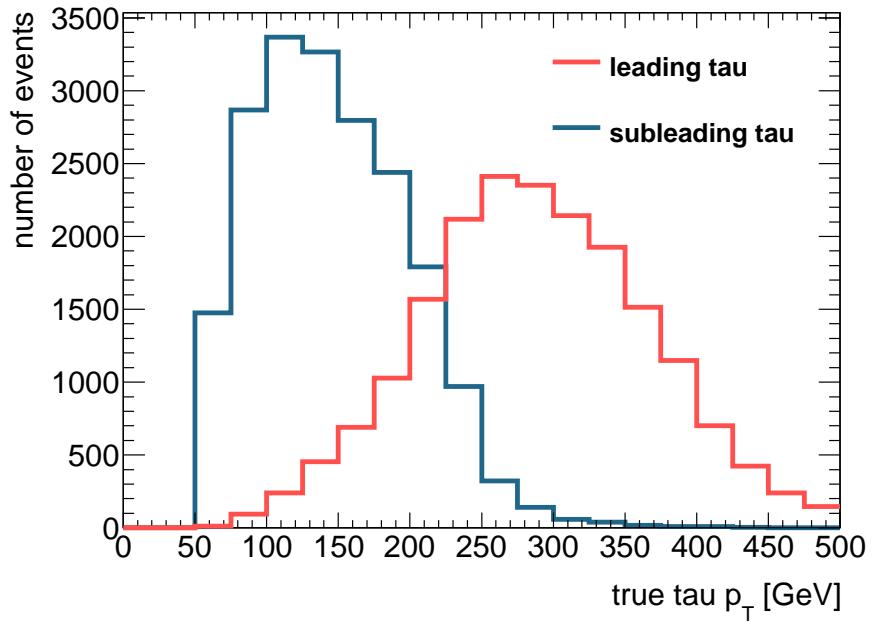
**Figure 4.2:** Distributions of true  $p_T$  for a heavy Higgs boson  $A$  and its prompt decay products  $Z$  and  $h$ .

A boosted tau pair is characterized by the high energies of its constituents, where with higher boosts the distance between them gets smaller. In Figure 4.3 the  $p_T$  distributions of tau leptons are shown, reaching momenta of up to 500 GeV. Each pair is divided in a leading tau, with the higher transverse momentum and a subleading tau with the lower transverse momentum. The  $p_T$  distributions of leading and subleading tau are peaking at 260 GeV and 110 GeV respectively. In Figure 4.4 the angular distance  $\Delta R$  between both tau leptons is shown. With higher values of Higgs  $p_T$  they get stronger boosted and their distances are shrinking.

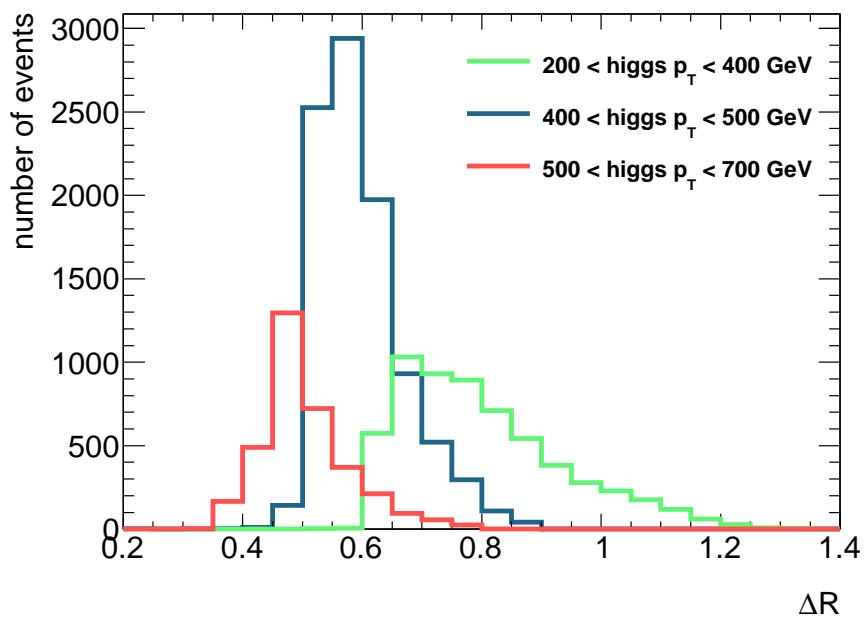
For investigating the results of tau reconstruction, just a minimal set of quality criteria is applied:

- $p_T$  of seed jets from calorimeter  $> 15$  GeV and  $|\eta| < 2.5$ ,
- absolute value of combined charges of associated tracks  $= 1$ ,
- $|\eta|$  of the associated leading track  $< 2.47$ .

In heavy Higgs boson search, which is a use-case for boosted tau identification, both taus have to be reconstructed at high energies. For that purpose a di-tau



**Figure 4.3:** Distributions of true tau  $p_T$ , for leading taus and subleading taus.

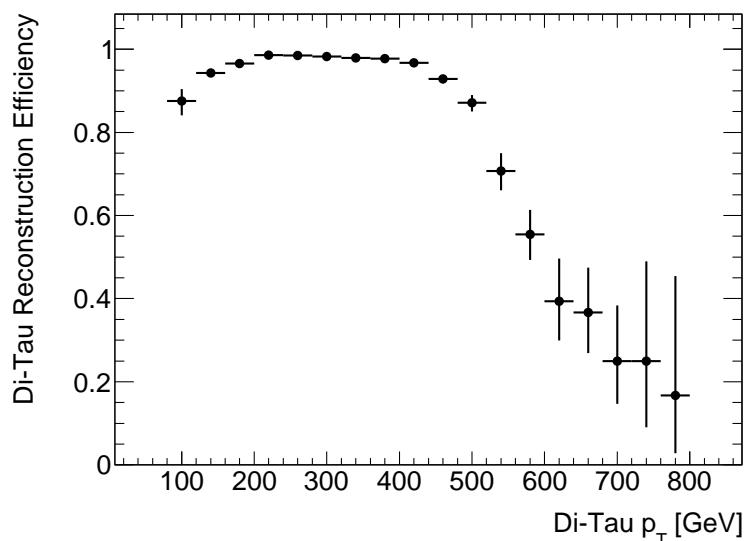


**Figure 4.4:** Angular distance  $\Delta R$  between both tau leptons, for different ranges of Higgs  $p_T$ .

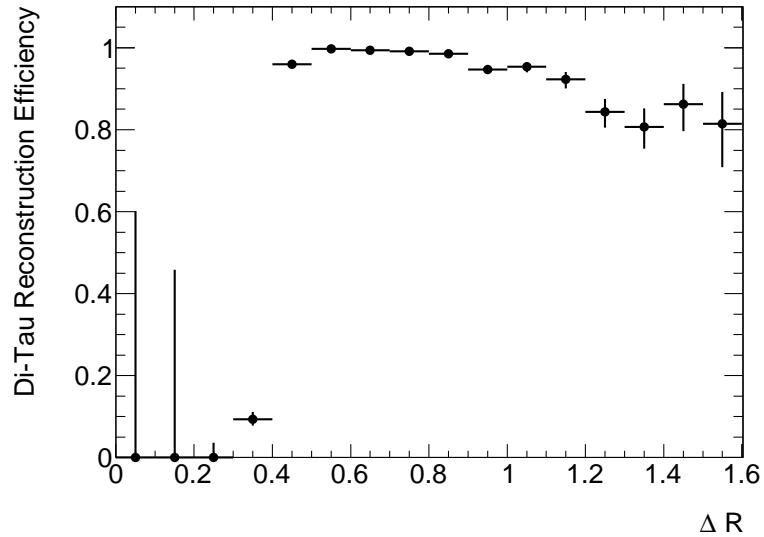
reconstruction efficiency is defined. This is the number of events where both tau leptons are reconstructed correctly and truth-matched, divided by all events. In Figure 4.5 the di-tau reconstruction efficiency is shown in bins of di-tau  $p_T$ , which is the combined transverse momentum of the two visible  $\tau_{\text{had}}$ . Here, a large plateau in a range of 150 GeV up to 500 GeV can be seen, where nearly all tau pairs can be reconstructed. But for higher values of di-tau  $p_T$  the efficiency drops significantly.

Since there is no tau identification applied this is exclusively an effect of the reconstruction method. The drop of efficiency is due to mis-reconstructions of highly boosted taus with small angular distances, as can be seen in Figure 4.6, where the reconstruction efficiencies are shown as a function of the angular distance  $\Delta R$ . Again there is a plateau of large reconstruction efficiencies for large angular distances  $\Delta R$ , corresponding to low energies, but with a sharp drop of efficiency at 0.4. As described in section 4.3 the tau reconstruction is seeded by anti- $k_t$ -jets with a distance parameter of  $R = 0.4$ , corresponding to the maximal distance between jet axes. Therefore tau pairs with smaller distances  $\Delta R < 0.4$  are merged into the same jet and can not be reconstructed separately. This is illustrated in Figure 4.7. To reconstruct highly boosted tau pairs a new approach is necessary, since they can not be handled with existing methods by construction.

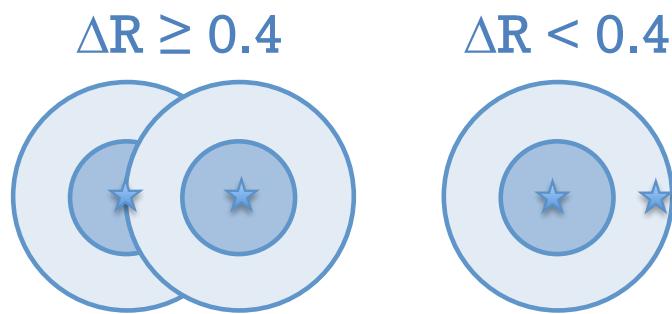
At this point, there are two possible solutions: to reduce the jets distance parameter until both  $\tau_{\text{had}}$  can be reconstructed again, or to implement a reconstruction where both taus are reconstructed as one object. It was the main aim of this study to find such an reconstruction method for boosted tau topologies. The results are discussed in Chapter 5.



**Figure 4.5:** Di-tau reconstruction efficiency in bins of di-tau  $p_T$ . After a large plateau where nearly all tau pairs can be reconstructed the efficiency drops significantly at  $p_T \approx 500$  GeV.



**Figure 4.6:** Di-tau reconstruction efficiency in bins of the angular distance  $\Delta R$  between both taus. Highly boosted tau pairs with  $\Delta R < 0.4$  can not be reconstructed with the standard tau reconstruction method.



**Figure 4.7:** Illustration of anti- $k_t$  seed jet merging for boosted tau lepton pairs. In the standard tau reconstruction, tau pairs with distances of  $\Delta R < 0.4$  merge into the same seed jet.

# Chapter 5

## Di-Tau Reconstruction and Identification

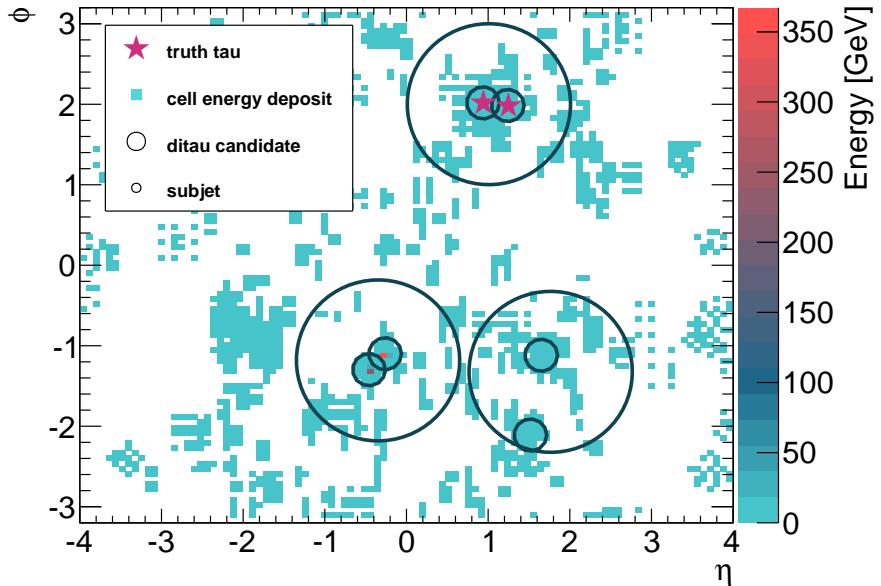
The search for new particle physics phenomena was always driven by a combination of an accurate detector hardware design as well as by efficient software algorithms to interpret the detected signatures. With the ATLAS detector, which is able to detect new physics at the TeV scale, and the restart of the LHC with higher event rates and center-of-mass energies of up to 14 TeV, the available approaches of particle reconstruction have to be re-considered. In this study, this is done in the field of tau reconstruction. A new approach for reconstructing boosted pairs of hadronically decaying taus is developed, which is considered to be an extension of the standard single-tau approach in the high-energy region. Decay channels of heavy Higgs bosons, decaying with SM-like Higgs bosons as mediators, like  $H \rightarrow hh$  and  $A \rightarrow Zh$  are an important field of application.

In this chapter the new approach of boosted di-tau tagging is introduced. The first section serves as a motivation and overview. The technical implementation of the boosted di-tau tagging is discussed in the second section. Afterwards the di-tau reconstruction and identification (ID) as well as their results are discussed in detail in the third and fourth section of this chapter.

### 5.1 The Di-Tau Decay Topology

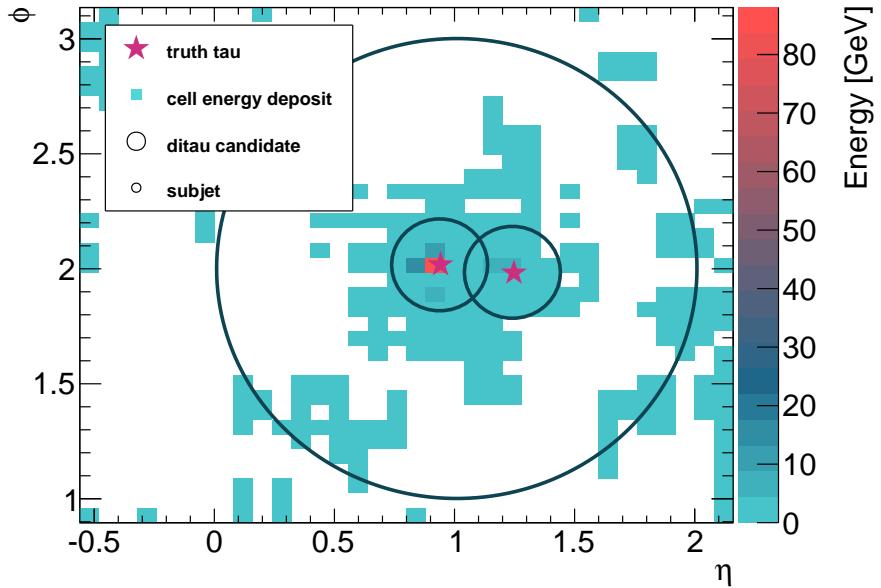
In Figure 5.1 an example for a simulated  $A \rightarrow Zh$  decay is shown with  $m_A = 2$  TeV. This includes a boosted pair of hadronically decaying tau leptons, originating from the SM-like Higgs boson  $h$ , a pair of electrons, originating from the  $Z$  boson, and at least one QCD jet. In this figure, the energy, deposited in calorimeter cells, is shown in bins of the pseudorapidity,  $\eta$ , and the azimuthal angle,  $\phi$ . A closer look at the di-tau decay topology of the same event can be taken in Figure 5.2. The illustrated

topology of a boosted tau pair decay is meant to motivate the further di-tau tagging approach.



**Figure 5.1:** Simulated  $A \rightarrow Z h$  example decay with  $m_A = 2$  TeV. The energy deposits in the calorimeter cells are shown in bins of the pseudorapidity,  $\eta$ , and the azimuthal angle,  $\phi$ .

The simulated tau leptons in this event have an angular distance of approximately  $\Delta R = 0.3$ . Therefore they can not be reconstructed by the single-tau reconstruction method, using anti- $k_t$  seed jets with a distance parameter of  $R = 0.4$ . As already noted at the end of Chapter 4, there are two possible solutions: reducing the anti- $k_t$  distance parameter, until both  $\tau_{\text{had}}$  can be reconstructed separately again, or to reconstruct the tau pair as one object. For the first suggestion it has to be assumed that the tau leptons have significantly higher energies than the energies expected in the current single tau reconstruction, leading to a more collimated  $\tau_{\text{had}}$  decay topology. Shrinking cone approaches may indeed improve the efficiency of a single-tau reconstruction in a high-energy area, but in the field of boosted tau pairs it may lead to complications, since not necessarily both taus have to have a high momentum. This was already illustrated in Figure 4.3, where the average leading tau lepton has more than twice the energy of the subleading tau. This can also be seen in the example decay in Figure 5.2. The energy deposited in the core region of the leading  $\tau_{\text{had}}$  is approximately 80 GeV, while the subleading  $\tau_{\text{had}}$  deposits an energy of approximately 10 GeV. For reasons of combinatorics the tau lepton, tagged as subleading, is likely to have a significantly lower energy as the leading tau, even if both taus originate from the same  $p_T$  distribution. Additionally, in the case of a spin-zero Higgs boson as parent particle, one tau lepton is likely to give the majority

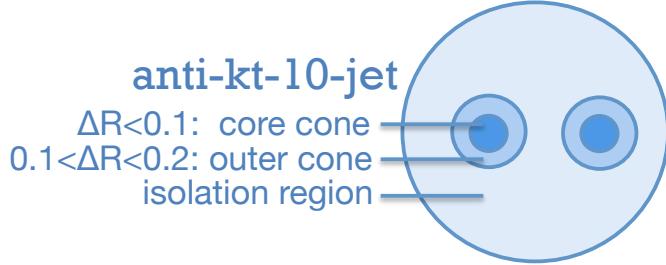


**Figure 5.2:** Example of a di-tau decay topology from a simulated  $A \rightarrow Z h$  example decay with  $m_A = 2$  TeV. The energy deposits in the calorimeter cells are shown in bins of the pseudorapidity,  $\eta$ , and the azimuthal angle,  $\phi$ .

of its energy to the tau neutrino, for reasons of spin and  $CP$  conservations. This enhances the probability of topologies with one high-energy  $\tau_{\text{had}}$  and one lower energy  $\tau_{\text{had}}$ . Applying a smaller seed jet cone to the tau lepton, which does not necessarily have a more collimated signature than in the case of single-tau reconstruction, leads to a tau candidate, that may occupy the whole jet area. Consequently, the main characteristic for QCD jet rejection is lost, the isolation criteria. This may lead to acceptable reconstruction results, but makes a distinction between signal and background very challenging.

A more promising approach, in terms of background discrimination, is to tag a boosted  $\tau_{\text{had}}$  pair as one di-tau object. Here, inside a bigger seed jet subjets are identified. The area around the subjets is defined as isolation region, where low energy deposits are expected. A schematic view of the di-tau topology is illustrated in Figure 5.3. Similar to the single-tau approach, the boosted di-tau tagging approach of this study is separated in a reconstruction step, that searches for di-tau candidates, and a later ID step for background rejection. The results of the implemented di-tau reconstruction for the example event are also shown in Figure 5.1 and Figure 5.2. Three di-tau candidates could be reconstructed in the example event. One candidate contains the actual  $\tau_{\text{had}}$  pair. It has two subjets, matching to the cell energy peaks in its centers, surrounded by an isolation region, with low energy deposits. The other two di-tau candidates are containing a  $Z \rightarrow ee$  decay topology and a QCD jet, requiring a later background discrimination. After a brief overview over

of the technical implementation of the di-tau tagging approach in the next section, the di-tau reconstruction and identification are discussed in detail in Section 5.3 and Section 5.4.



**Figure 5.3:** Schematic view of the expected topology in di-tau reconstruction.

## 5.2 Technical Implementation

In this study, nearly the full detector information is needed to reconstruct di-tau candidates and to calculate the ID variables for background discrimination. Cluster information is needed over the full  $\eta - \phi$  range to reconstruct custom seed jets. For the calculation of efficient ID variables, the information of energy deposits is needed with a high granularity, provided by the calorimeter cells. Also the information about primary and secondary vertices is needed to apply track quality criteria and to calculate tau decay lengths. All this information is provided in the ESD format. ESD files are a slimmed version of the raw detector information file, and provide a baseline event reconstruction and digitization.

For di-tau signal,  $A \rightarrow Zh \rightarrow ll + \tau\tau$  decays with  $m_A = 2 \text{ TeV}$  are generated with POWHEG BOX [52]. The parton showering is done by PYTHIA 8.1 [53]. Afterwards a detector simulation and digitization is done, with ESD output. Because of the high  $A$  mass, the sample, which includes 50,000 events, is enriched with highly boosted  $\tau_{\text{had}}$  pairs. For QCD background, di-jet data is taken. Here, the available amount of data was limited, since only a subset of the 2012 ATLAS data is reprocessed into the newest ESD release format. A high- $p_T$  filter was implemented, which preprocesses QCD samples and tags rare high-energy di-tau candidates. In these samples, 6164 di-tau candidates were found with at least one seed jet with  $p_T > 400 \text{ GeV}$ . A full list of the used data samples can be found in Appendix B.

Like the single-tau reconstruction, the di-tau reconstruction is implemented as an ATHENA package running on ESD files. It is written in software release 19, which is ready for Run 2. For di-tau seed jet reconstruction the FastJet Package [54] is used in version 3.0. The output of the ATHENA package contains the location and energy information of di-tau candidates for each event, its associated subjets and

tracks, as well as a set of variables for a later background rejection. The results are plotted with ROOT [55] in version 5.34. Afterwards, the output of the ATHENA package is evaluated with several Python [56] classes. These have the ability to define new identification variables, which were not calculated in the reconstruction step before. They also prepare the ATHENA output for the ID. The signal and background samples are split into one training sample and one testing sample. The training and testing samples have a tree structure, with one entry for each candidate, containing only information needed for the background separation. The ID is done by multi-variate techniques using the TMVA [57] package in ROOT. Afterwards the evaluation of the background separation is done again with Python and ROOT.

## 5.3 Di-Tau Reconstruction

### 5.3.1 Seed Jet Reconstruction

Like in the single-tau reconstruction, anti- $k_t$  jets serve as seeds for the reconstruction, though with a larger distance parameter of  $R = 1.0$ . These jets will also be referred as anti- $k_t$ -10 jets in the remainder of this thesis<sup>1</sup>. The larger seed jets are meant to contain the signatures of both  $\tau_{\text{had}}$  decays. As already noted in Section 4.2, the anti- $k_t$  algorithm provides the most predictable jet shapes, and is therefore very well suited for example for later area definitions, like the isolation region, or for pile-up reduction. The wide jet area can contain not only highly boosted tau pairs with a sufficient isolation region, but also tau leptons that could be reconstructed by the single-tau tagging. This gives the opportunity for cross-checks between both methods and an uninterrupted transition region between them. Also, the additional di-tau approach may improve tau tagging in the transition region of close tau leptons ( $0.4 < \Delta R < 1.0$ ). In this study, custom jets are calculated, using topological clusters, which were reconstructed from calorimeter cells. In an analysis application, anti- $k_t$ -10 jets are commonly provided by the standard ATLAS event reconstruction software. Only jets with a transverse momentum of  $p_T > 15 \text{ GeV}$  are considered as di-tau seeds.

### 5.3.2 Subjet Reconstruction

After the seed jet reconstruction, each jet is filtered for subjets. For subjet finding the Cambridge/Aachen algorithm is used with a distance parameter of  $R = 0.2$ . Such a technique was initially introduced for Higgs boson searches with a boosted

---

<sup>1</sup> This is a common notation: for example anti- $k_t$  jets with a distance parameter of  $R = 0.4$  are also referred as anti- $k_t$ -4 jets.

$b\bar{b}$  pair in the decay products [58]. There, the analysis of jet substructure could enhance the background separation.

The subjet with the highest momentum is referred to as leading subjet, the subjet with the second highest momentum is referred to as subleading subjet. In di-tau signal events, these subjets are expected to contain the tau lepton decay signatures. For a first QCD jet rejection, subjets are excluded that contain more than 4 tracks. A di-tau candidate has to fulfill the following criteria: the jet should include at least two subjets, and both, the leading and the subleading subjet, should have at least one track. The di-tau four-momentum is at this point defined as the sum of the four-momenta of these two subjets.

### 5.3.3 Vertex Association

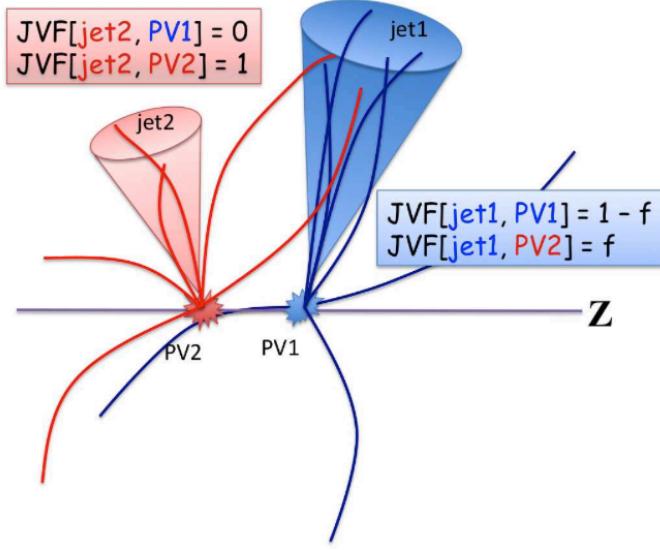
The determination of vertices is important for the calculation of track quality criteria, and background rejecting variables. There are two types of vertices of interest in this study: primary and secondary vertices.

A primary vertex is the point of an initial proton-proton collision. A vertex can not be measured directly, but can be extrapolated via the charged particle's trajectories. Since there are many proton collisions in each bunch crossing, multiple vertices are aligned along the beam axis in each event. Usually the vertex with the highest sum of squared track momenta  $\sum(p_T^{\text{trk}})^2$  is defined as primary vertex. But this is not necessarily the vertex, the tau lepton emerges from. Therefore an algorithm is applied to find the primary tau vertex for each subjet. This approach, called Tau Jet Vertex Association (TJVA) [59], is also implemented in the single-tau reconstruction. The TJVA functionality is illustrated in Figure 5.4. Here, for each jet and each vertex the jet vertex fraction,  $f_{\text{JVF}}$ , is calculated via

$$f_{\text{JVF}}(\text{jet}|\text{vtx}) = \frac{\sum p_T^{\text{trk}|\text{vtx}}}{\sum p_T^{\text{trk}}}. \quad (5.1)$$

In the denominator the sum includes all tracks with a distance of  $\Delta R < 0.2$  to the subjet axis. In the numerator, just the subset of tracks, that emerge from the same vertex, are included. For each subjet, the vertex with the highest  $f_{\text{JVF}}$  serves as tau vertex.

Since tau leptons have a finite proper life time, there is a secondary vertex, slightly displaced from the beam axis. It can be reconstructed if there are two or more tracks associated to a subjet. A secondary vertex is associated to each multi-prong subjet.



**Figure 5.4:** Illustration of the tau jet vertex association [59]. The vertex, that provides the higher jet vertex fraction is assumed to be the tau vertex.

### 5.3.4 Track Association

Tracks are associated to each di-tau candidate, if they have a distance of  $\Delta R < 1.0$  to the seed jet axis, as well as to each subjet, if they have a distance of  $\Delta R < 0.2$  to the subjet axis. Tracks are needed for background discrimination and for the primary and secondary vertex fitting. For the vertex association just a minimal set of track cuts is applied to exclude pile-up tracks:

- $p_T \geq 1 \text{ GeV}$ ,
- Number of pixel hits  $\geq 2$ ,
- Number of pixel hits + number of SCT hits  $\geq 7$ .

Additionally, the final track cuts include impact parameter cuts, which are dependent on the primary vertex:

- $|d_0| \leq 1.0 \text{ mm}$ ,
- $|z_0 \sin \theta| \leq 1.5 \text{ mm}$ .

The parameter  $d_0$  is the closest distance of the track to the primary vertex in the transverse plane and  $z_0$  is the longitudinal distance of closest approach. The impact parameter cuts for tracks within a subjet are calculated with respect to the associated tau vertex. For the other tracks in the di-tau isolation region, the vertex with the highest sum of squared track momenta  $\sum(p_T^{\text{trk}})^2$  is used for the impact parameter calculations.

### 5.3.5 Results

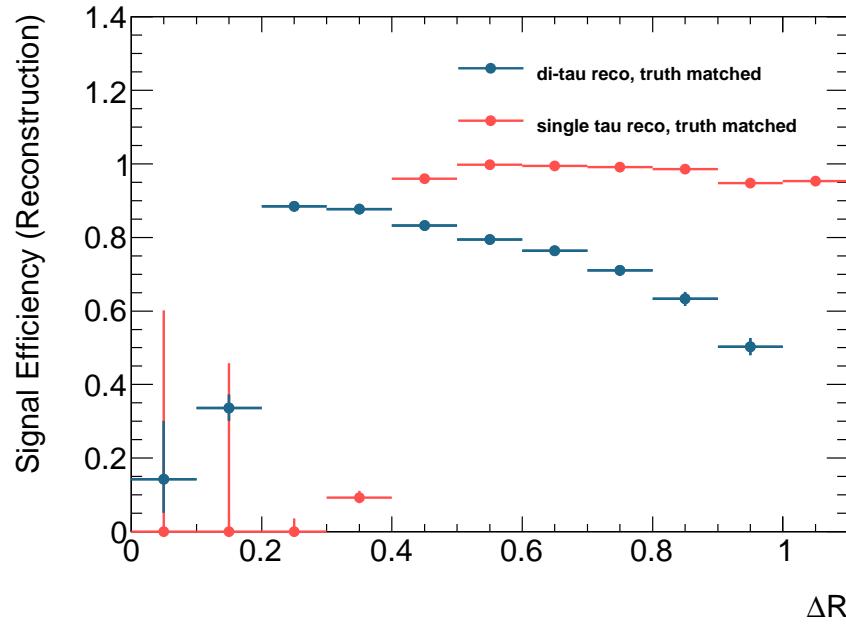
In this section, the results of the di-tau reconstruction are presented in terms of signal efficiencies in comparison to the single-tau reconstruction. Here, a tau pair is referred to as correctly reconstructed, if each visible truth  $\tau_{\text{had}}$ , matches to the leading or subleading subjet of a di-tau candidate with a distance of  $\Delta R < 0.2$  respectively. In Figure 5.5 the reconstruction efficiency is presented as a function of the tau angular distance  $\Delta R$ . In Figure 5.6 it is presented as a function of di-tau  $p_T$ . Depending on the distance and momentum, high reconstruction efficiencies of up to 90% can be achieved. In particular in the low distance and high-energy areas, the di-tau reconstruction shows largely improved reconstruction efficiencies compared to the standard single-tau reconstruction. Di-taus can be reconstructed with distances between  $0.2 < \Delta R < 1.0$ . At higher distances, the tau leptons do not fit into a single anti- $k_t$ -10 seed jet. At lower distances, less than 0.2, both  $\tau_{\text{had}}$  merge into the same subjet.

The newly accessible area  $0.2 < \Delta R < 0.4$  allows the reconstruction of highly boosted tau pairs at much higher energies. While with the single-tau reconstruction, tau pairs with a momentum of up to 500 GeV could be reconstructed, the new di-tau approach reconstructs tau pairs with a momentum of up to 1200 GeV with high efficiency. In Figure 5.6 it can be seen that the di-tau reconstruction is a valuable addition to the single-tau reconstruction, in the high-energy area. Tau pairs with an angular distance higher than 1.0 are excluded in this plot, since they can not be reconstructed by the di-tau reconstruction by definition and their number is dependent on the  $p_T$  distributions of the input sample.

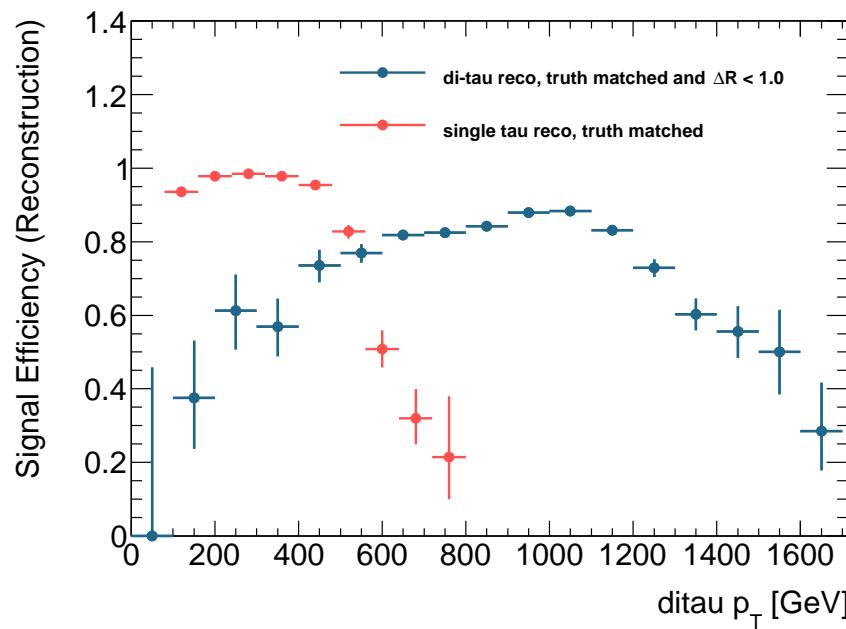
## 5.4 Di-Tau Identification

The main source of background, for the di-tau tagging, are QCD jets. Except for the requiring of two subjets, with at least one track, and the upper cut on the number of tracks in a subjet, the di-tau reconstruction provides a very little discrimination against QCD background. Depending on the decay channel, there might be more background signatures, like electron or muon pairs, that have to be handled by the dedicated analysis. Another possible source of background events, which has to be tested by further studies, are  $t\bar{t}$  decays.

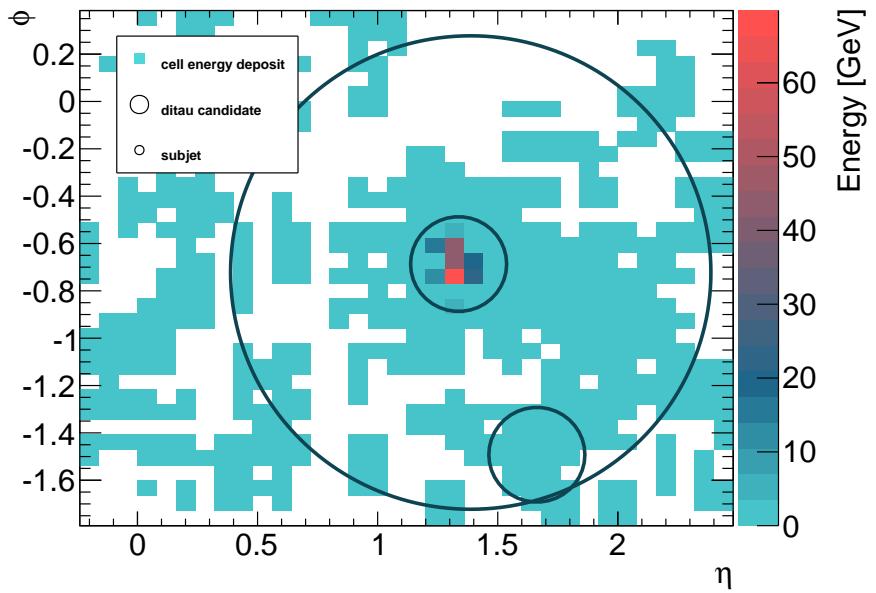
This study intends to show the possibility of QCD background discrimination with high efficiencies. This is done with a multi-variate evaluation of a set of variables, which combines tracker information of the inner detector with information from the calorimeters. In Figure 5.7 an example of a QCD jet from ATLAS data is shown, which was reconstructed as a di-tau candidate. In contrast to di-tau signal



**Figure 5.5:** Di-tau and single-tau reconstruction efficiencies as a function of the tau lepton distance  $\Delta R$ . Between  $\Delta R = 0.2$  and  $\Delta R = 0.4$  the di-tau reconstruction shows largely improved reconstruction efficiencies compared to the standard single-tau reconstruction



**Figure 5.6:** Di-tau and single-tau reconstruction efficiencies as a function of di-tau  $p_T$ . The new di-tau algorithm extends the sensitivity for tau pair reconstruction to up to 1200 GeV.



**Figure 5.7:** Example of a QCD jet decay topology, from ATLAS data. The energy deposits in the calorimeter cells are shown in bins of the pseudorapidity,  $\eta$ , and the azimuthal angle,  $\phi$ .

signatures (see Figure 5.2), QCD jets are expected to have less pronounced subjets, with a less collimated signature. Also QCD jets are likely to have more tracks, than di-tau signal events, with accordingly less energy in each track. Additionally, for di-tau signal events it is expected, that a noticeable decay length can be measured for multi-prong taus. These characteristics are quantified in the set of variables, described in Section 5.4.1. A subset of those variables is used for background discrimination with BDT, which is described in Section 5.4.2. In Section 5.4.3 the results of the di-tau ID are shown and combined with the results of the di-tau reconstruction.

#### 5.4.1 Discriminating Variables

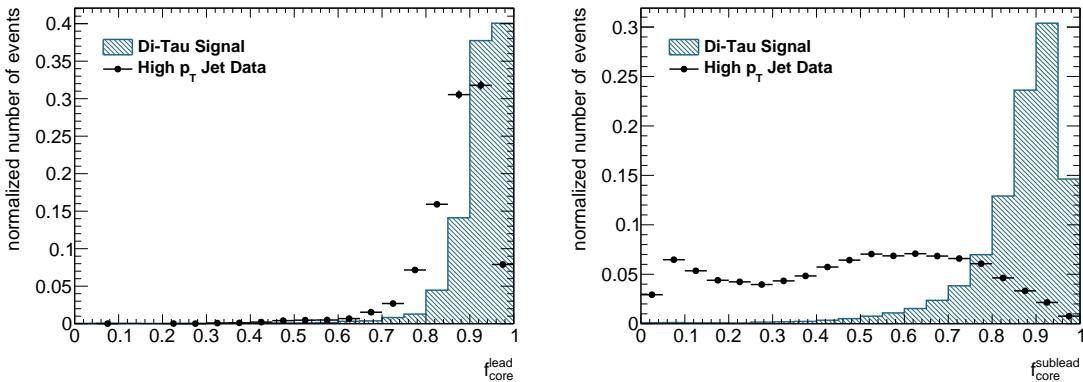
In this section, a set of potential variables for background discrimination is introduced. The plots of this section show the normalized distributions for QCD background and di-tau signal. For QCD background, 10,000 di-tau candidates are taken, that were reconstructed in di-jet ATLAS data. 10,000 truth matched di-tau candidates from simulated  $A \rightarrow Zh$  decays are taken for signal. Most of these variables have an single-tau equivalent [47] and are defined for the leading subjet, as well as for the subleading subjet of the di-tau candidate.

## Core Energy Fractions

The core energy fraction is the fraction of energy deposited in the core region,  $\Delta R < 0.1$ , of a subjet. This variable is defined for the leading and the subleading subjet separately. Here, the transverse momentum, registered in each cell of the core region is summed up, and divided by the transverse momenta of the full subjet area,  $\Delta R < 0.2$ :

$$f_{\text{core}}^{(\text{sub})\text{lead}} = \frac{\sum_{\text{cells}}^{\Delta R < 0.1} p_{\text{T}}^{\text{cell}}}{\sum_{\text{cells}}^{\Delta R < 0.2} p_{\text{T}}^{\text{cell}}}. \quad (5.2)$$

The normalized signal and background distributions for the leading and subleading subjet are shown in Figure 5.8. It can be seen that for the leading subjets of both QCD background and di-tau signal, almost the full energy is deposited in the core region of the subjet, although leading subjets of QCD jets have a slightly wider signature. The subleading  $\tau_{\text{had}}$  also forms a collimated subjet, resulting in a core energy fraction distribution similar to that of the leading subjet. This is in high contrast to the distribution of subleading subjets of QCD background. Here, often just a small fraction of energy is deposited in the core region.



**Figure 5.8:** Normalized distributions of the core energy fraction of the leading subjet  $f_{\text{core}}^{\text{lead}}$  (left) and of the subleading subjet  $f_{\text{core}}^{\text{sublead}}$  (right). The markers represent QCD jet background and the hatched histograms represent di-tau signal.

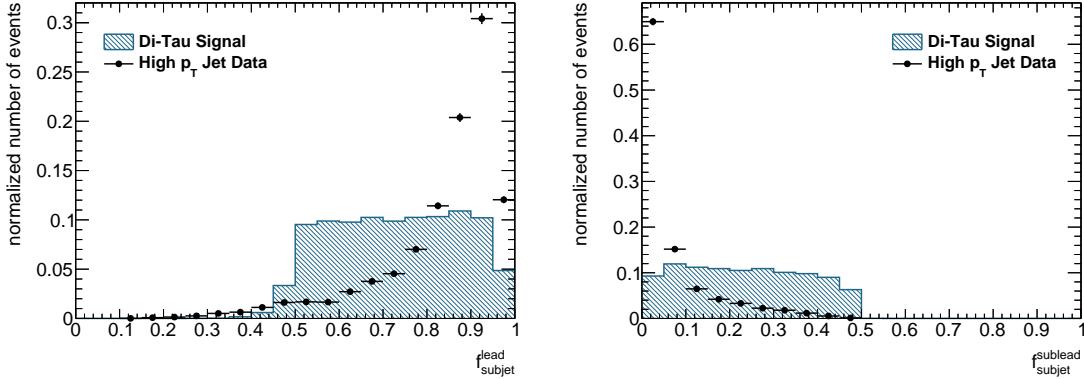
## Subjet Energy Fractions

The subjet energy fractions of the leading and subleading subjet are the fractions of energy, deposited in a subjet, compared to that of the anti- $k_t$ -10 seed jet:

$$f_{\text{subjet}}^{(\text{sub})\text{lead}} = \frac{p_{\text{T}}^{\text{subjet}}}{p_{\text{T}}^{\text{jet}}}. \quad (5.3)$$

The  $f_{\text{subjet}}$  distributions for the leading and subleading subjets are shown in Figure 5.9 for di-tau signal and QCD background. The di-tau signal shows nearly

uniform distributions for the leading and subleading subjets, with a drop of the subjet energy fraction at 0.5. This is in contrast to the QCD background, where the majority of the energy is deposited in the leading subjet, and accordingly nearly no energy is deposited in the subleading subjet.



**Figure 5.9:** Normalized distributions of the subjet energy fraction of the leading subjet  $f_{\text{subjet}}^{\text{lead}}$  (left) and of the subleading subjet  $f_{\text{subjet}}^{\text{sublead}}$  (right). The markers represent QCD jet background and the hatched histograms represent di-tau signal.

Another possible definition of a subjet energy fraction is the following:

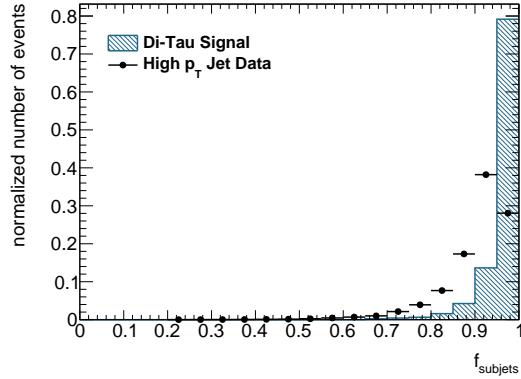
$$f_{\text{subjets}} = \frac{p_{\text{T}}^{\text{leading subjet}} + p_{\text{T}}^{\text{subleading subjet}}}{p_{\text{T}}^{\text{jet}}}, \quad (5.4)$$

which is illustrated in Figure 5.10, and describes the fraction of energy deposited in the two leading subjets. The inverse of this variable quantifies the fraction of energy deposited in the isolation region. It can be seen that in di-tau decays nearly the complete energy is deposited in the subjets. The  $f_{\text{subjets}}$  distribution for QCD jets is slightly broader, with more energy deposited in the isolation region. Nevertheless  $f_{\text{subjets}}$  is highly correlated with the above variables, since it can be expressed as the sum of both. The lack of information about the individual subjets, causes a lower discrimination power.

### Leading Track Momentum Fractions

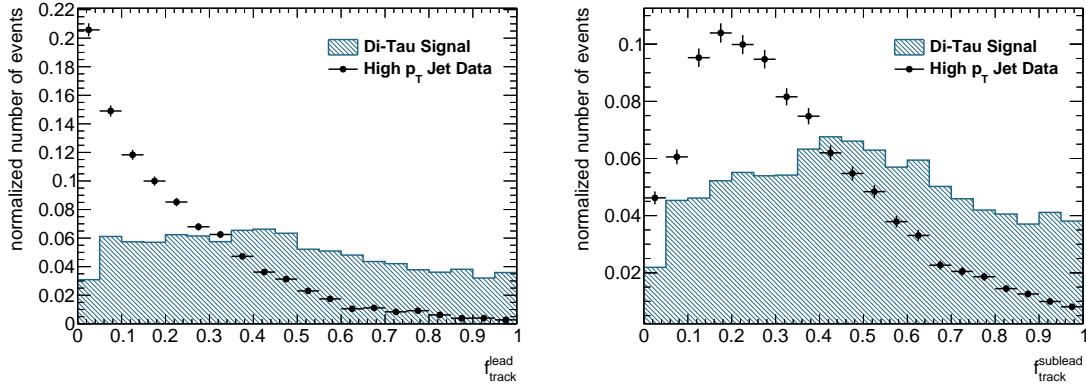
The leading track momentum fraction is the transverse momentum of the highest  $p_{\text{T}}$  track within a subjet divided by the subjet's transverse momentum. This is calculated for the leading and subleading subjet via

$$f_{\text{track}}^{(\text{sub})\text{lead}} = \frac{p_{\text{T}}^{\text{leadtrk}}}{p_{\text{T}}^{\text{subjet}}}. \quad (5.5)$$



**Figure 5.10:** Normalized distributions of the subjet energy fraction  $f_{\text{subjets}}$  for di-tau signal and QCD background. The markers represent QCD jet background and the hatched histograms represent di-tau signal.

The distributions of this ID variable are shown in Figure 5.11. It can be seen that QCD jets tend to have lower energy tracks. In the di-tau signal, more often a high amount of energy is included in one track.

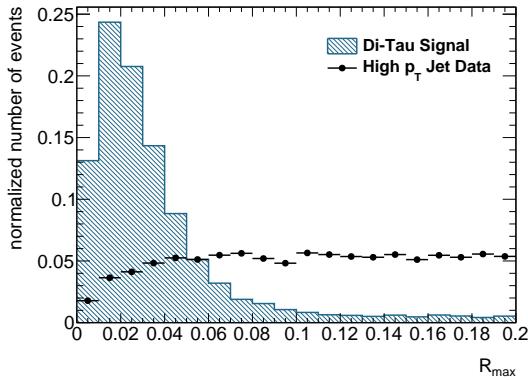


**Figure 5.11:** Normalized distributions of the leading track momentum fraction in the leading subjet  $f_{\text{lead track}}$  (left) and in the subleading subjet  $f_{\text{sublead track}}$  (right). The markers represent QCD jet background and the hatched histograms represent di-tau signal.

### Track Distances

Di-tau signal is expected to have more collimated subjets, not only in the signature of energy deposits, but also in the distribution of tracks. For this purpose, the track distance variables  $R_{\max}$  and  $R_{\text{track}}$  are defined.  $R_{\max}$  measures the maximum distance of an associated track to the dedicated subjet axis. This is illustrated in Figure 5.12. It can be seen that an average di-tau track is very close to the subjet axis, while QCD jet tracks can have a broad local distribution.

$R_{\text{track}}$  is the  $p_{\text{T}}$ -weighted sum of all track distances to a subjet axis, divided by

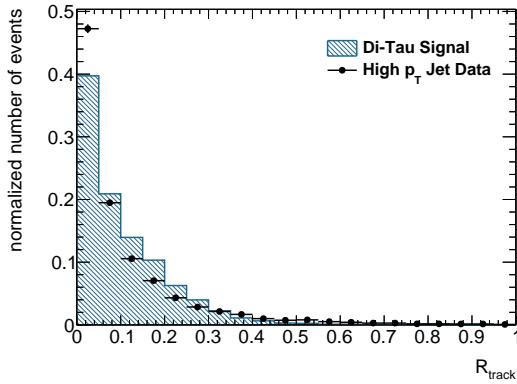


**Figure 5.12:** Normalized distributions of the maximum track distance  $R_{\max}$  for di-tau signal and QCD background. The markers represent QCD jet background and the hatched histograms represent di-tau signal.

the overall track  $p_T$ :

$$R_{\text{track}} = \frac{\sum_i^{\Delta R_i < 0.2} p_{T,i}^{\text{trk}} \Delta R_i}{\sum_i^{\Delta R_i < 0.2} p_{T,i}^{\text{trk}}}. \quad (5.6)$$

The  $R_{\text{track}}$  distributions are illustrated in Figure 5.13. It can be seen that the distribution for QCD background is slightly narrower. This means that in QCD background tracks with a high momentum are very close to the subjet axis. Di-tau signal, in particular 3-prong tau decays, has tracks with almost identical momenta, which are slightly off-centered compared to the QCD background.

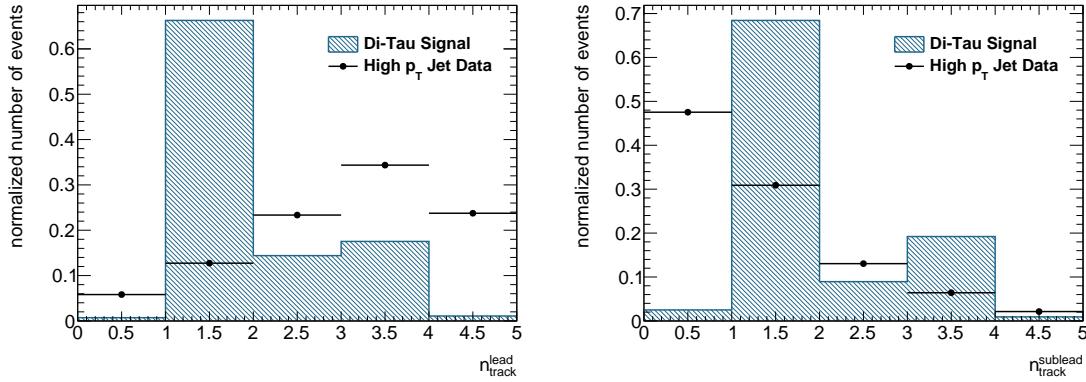


**Figure 5.13:** Normalized distributions of the track distance  $R_{\text{track}}$  for di-tau signal and QCD background. The markers represent QCD jet background and the hatched histograms represent di-tau signal.

### Number of Tracks

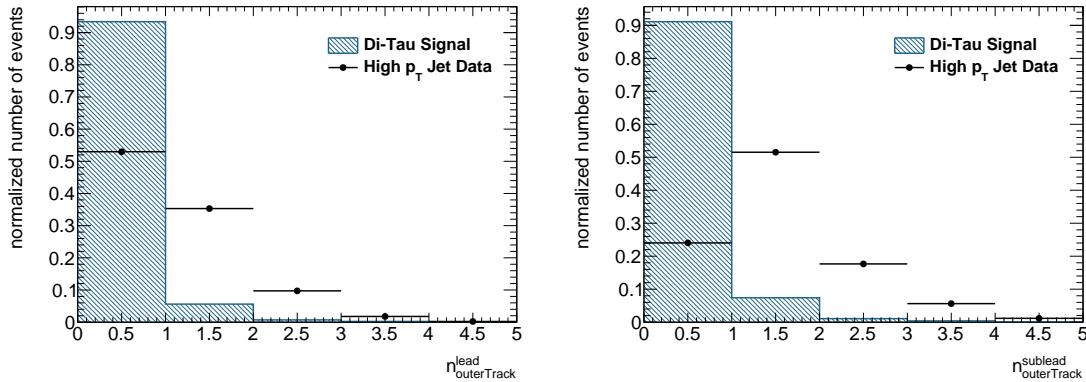
The number-of-track variable,  $n_{\text{track}}$ , counts the tracks in the core cones,  $\Delta R < 0.1$ , of the leading and subleading subjet. This is the equivalent to the single-tau prongness. In the single-tau tagging, different discriminators are trained for one and three prong taus. Here,  $n_{\text{track}}$  itself enters as one variable into the ID to consider

dependencies of the ID variables on the number of tracks. This gives the ID the opportunity to handle variables depending on the number of tracks, but also to handle mis-identification in the number of tracks. Additionally,  $n_{\text{track}}$  has a discriminating power itself, as shown in Figure 5.14. Most of the di-taus have one, two or three tracks in their subjets. Subjets from QCD background tend to have more often zero or more than three tracks in their core cone.



**Figure 5.14:** Normalized distributions of the number of tracks in the leading subjet  $n_{\text{track}}^{\text{lead}}$  (left) and in the subleading subjet  $n_{\text{track}}^{\text{sublead}}$  (right). The markers represent QCD jet background and the hatched histograms represent di-tau signal.

Also the number of tracks  $n_{\text{outerTrack}}$  in the outer cone,  $0.1 < \Delta R < 0.2$ , of a subjet is counted. In contrast to di-tau signal, QCD subjets often have at least one track in their outer cone. This is illustrated in Figure 5.15.



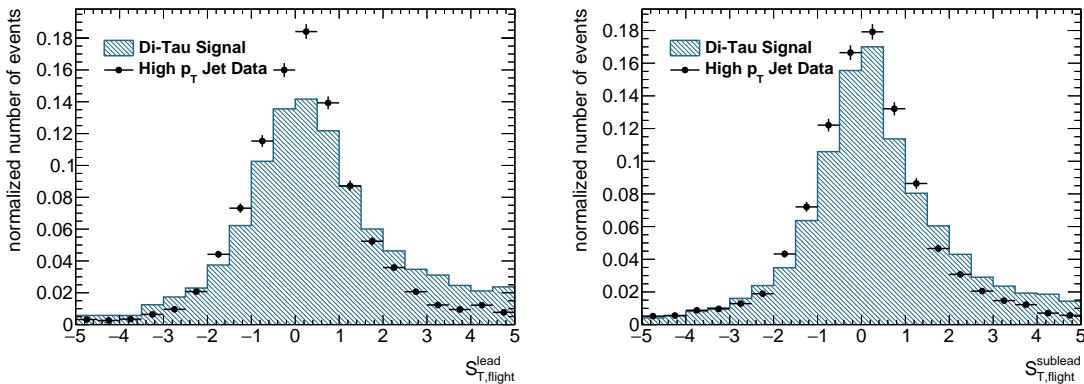
**Figure 5.15:** Normalized distributions of the number of tracks in the outer cone of the leading subjet  $n_{\text{outerTrack}}^{\text{lead}}$  (left) and of the subleading subjet  $n_{\text{outerTrack}}^{\text{sublead}}$  (right). The markers represent QCD jet background and the hatched histograms represent di-tau signal.

## Transverse Flight Path Significances

The transverse flight path significance describes the  $\tau_{\text{had}}$  decay length. It is defined only for subjets with more than one associated track via:

$$S_{\text{T,flight}}^{(\text{sub})\text{lead}} = \frac{L_{\text{T,flight}}}{\delta L_{\text{T,flight}}}, \quad (5.7)$$

where  $L_{\text{T,flight}}$  is the signed distance of the associated primary vertex to the secondary vertex in the transverse plain and  $\delta L_{\text{T,flight}}$  is its estimated uncertainty. The transverse flight path significances of the leading and subleading subjets are illustrated in Figure 5.16. As expected the di-tau signal shows broader distributions, with noticeably more taus with larger decay lengths than in QCD background. Nevertheless it shows a low discrimination power in contrast to the single-tau case, where this is an important discriminating variable for the ID. In the high-energy area, the decay lengths have larger uncertainties and the reconstruction of a secondary vertex seems to fail more often.



**Figure 5.16:** Normalized distributions of the transverse flight path significance of the leading subjet  $S_{\text{T,flight}}^{\text{lead}}$  (left) and of the subleading subjet  $S_{\text{T,flight}}^{\text{sublead}}$  (right). The markers represent QCD jet background and the hatched histograms represent di-tau signal.

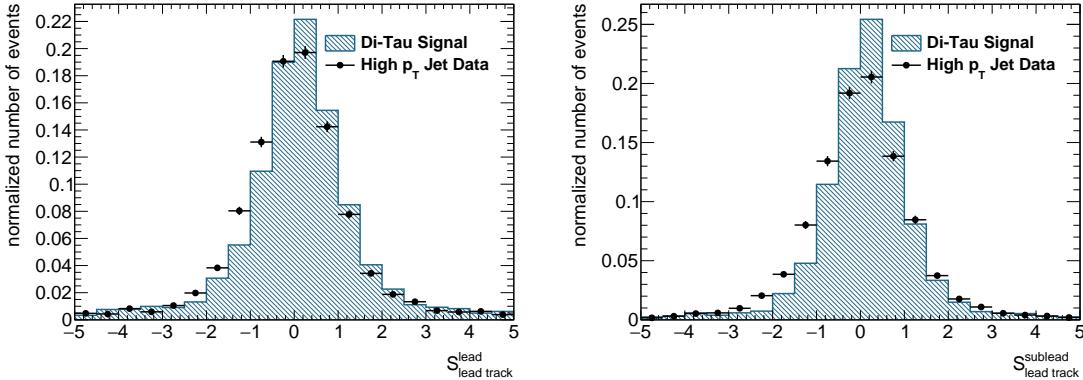
## Leading Track Impact Parameter Significances

The leading track impact parameter significance can be seen as an addition to the impact parameter cut on  $d_0$  for the leading track. Here,  $d_0$  is divided by its estimated uncertainty  $\delta d_0$ :

$$S_{\text{lead track}}^{(\text{sub})\text{lead}} = \frac{d_0}{\delta d_0}. \quad (5.8)$$

The distributions for this variable is shown in Figure 5.17. It also can be defined for the impact parameter  $z_0 \sin \theta$ , resulting in similar distributions. It can be seen that the distributions of di-tau signal and QCD background are almost identical.

In contrast to the single-tau ID, with these variables nearly no discrimination power can be achieved in the case of boosted tau pairs.



**Figure 5.17:** Normalized distributions of the leading track impact parameter significance of the leading subjet  $S_{\text{lead}}^{\text{lead track}}$  (left) and of the subleading subjet  $S_{\text{lead}}^{\text{sublead}}$  (right). The markers represent QCD jet background and the hatched histograms represent di-tau signal.

### 5.4.2 Boosted Decision Trees

For background discrimination in the di-tau ID, the multi-variate technique of Boosted Decision Trees (BDTs) [60] is used. BDTs are the state of the art technique in the single-tau ID and are known for a good separation performance among the separation methods. In contrast to a simple cut based approach, an event is not rejected in the BDT approach when it fails a certain cut criterion. Instead it is tested on all other discriminating variables, and though may appear as signal-like in the end. Rather than a binary output like in a cut based approach, BDTs deliver a continuous BDT score between zero, for background-like events and one for signal-like events. An analysis can choose a BDT cut score for the background separation based on the desired signal efficiency or background efficiency.

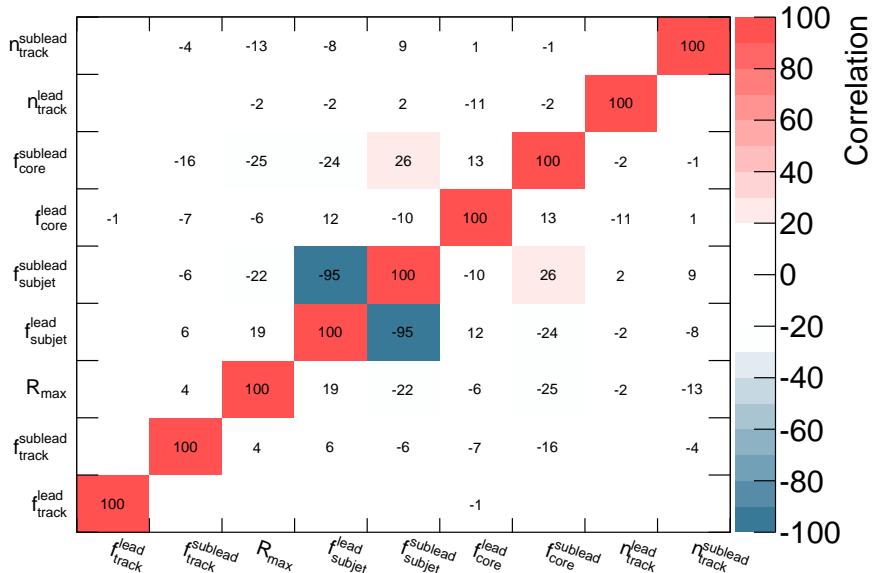
Before applying BDTs on a separation problem, they have to be trained with events, pre-classified as background or signal. Beginning in the root node of one decision tree, an optimal variable cut is identified that leads to the highest separation between signal and background in all events. Each event that passes this cut criterion is passed into one node, each event that fails is passed into another. This procedure is continued recursively for every node until a stop condition is fulfilled, for example when a node contains a minimal number of events or a maximal tree depth is reached. Each leaf node at the ends of a decision tree has a dedicated signal purity, that let events of this node appear as more signal-like or more background-like. The idea of boosting is to enhance the separation performance by combining multiple decision trees, where each tree is optimized to re-classify events, that were misclassified by

the previous tree. This is done by an increased weight of previously misclassified events.

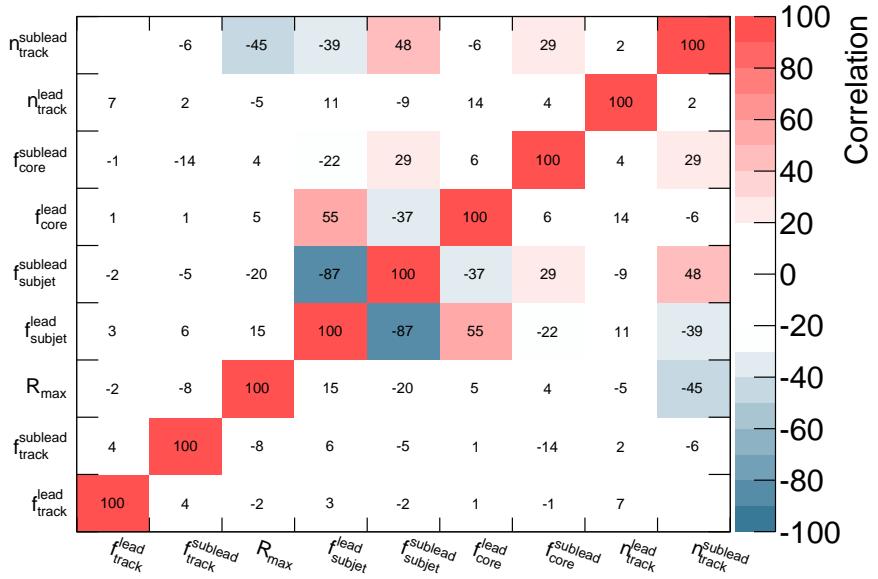
For the training of the di-tau ID only correctly reconstructed true di-tau events are used for signal, and high- $p_T$  di-jet data is used for QCD background. To get the same  $p_T$  distributions for signal and background and thus to compensate  $p_T$  dependencies of ID variables in the training, the background events are  $p_T$  re-weighted. A set of nine variables is chosen, regarding their separation power, their performance in previous BDT training tests and regarding their correlations:

- the core energy fractions  $f_{\text{core}}^{(\text{sub})\text{leading}}$ ,
- the subjet energy fractions  $f_{\text{subjet}}^{(\text{sub})\text{leading}}$ ,
- the leading track momentum fractions  $f_{\text{track}}^{(\text{sub})\text{leading}}$ ,
- the maximum track distance  $R_{\text{max}}$ ,
- the number of tracks  $n_{\text{track}}^{(\text{sub})\text{leading}}$ .

The correlation of these variables is illustrated in Figure 5.18 for the di-tau signal and in Figure 5.19 for the QCD jet background. Correlated variables usually do not provide additional information in the BDT training. For the signal, very low correlations are achieved with this set of variables, except for the subjet energy fractions. Nevertheless both variables are included, since their combination describes the important isolation criteria, noted before. For the background there are slight correlations for  $f_{\text{core}}^{\text{lead}}$ ,  $n_{\text{track}}^{\text{sublead}}$  and  $R_{\text{max}}$ .

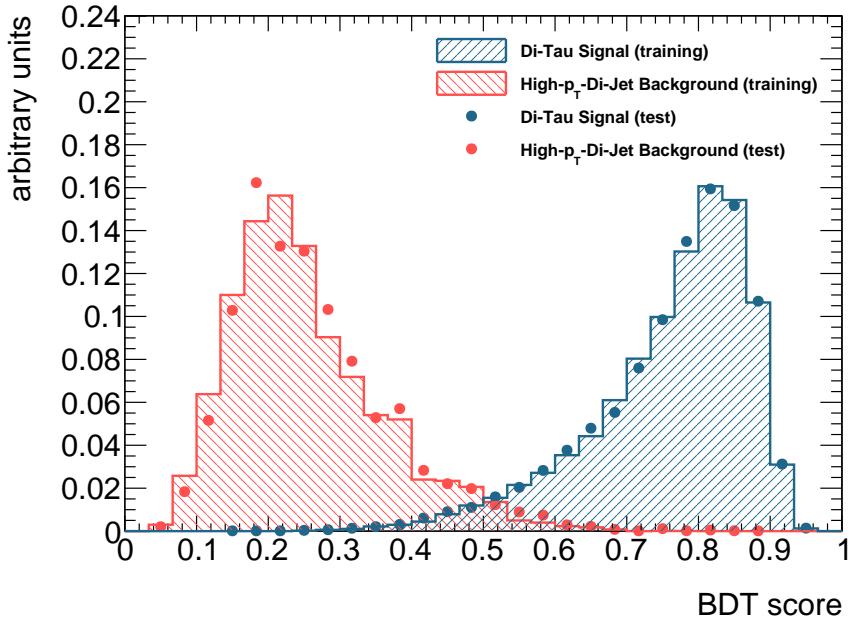


**Figure 5.18:** Correlation of the nine ID variables for di-tau signal events.



**Figure 5.19:** Correlation of the nine ID variables for QCD background events.

For boosting the AdaBoost (Adaptive Boosting) [61] algorithm was used, with maximum 50 trees, each with a depth of maximal 8 nodes. 200 cuts are applied on each ID variable, maximum. The stop criterion is a minimal node size of 1% of the number of input events. The full BDT configuration can be found in Appendix C. These parameters are tuned for good separation results without running into over-training. The effect of over-training appears when the trained decision trees are dominated by statistical fluctuations due to a too detailed training. This is indicated by significant differences between the BDT score distributions of the training and testing sample. Studies like this, which have access to only a rather small number of background events, are very sensitive to over-training. This is taken into account by the choice of parameters, leading to an early stop criterion and avoiding deep decision trees. Also a high number of ID variables can favor over-training. Two thirds of the available signal and background samples are used for training, and one third is used for testing afterwards. The BDT score output that could be achieved by this training is given in Figure 5.20. Two peaks can be seen, for QCD background and di-tau signal, which are well separated. For the ID only the BDT score area, which includes a significant number of di-tau signal events is of further interest. In this case this is approximately 0.4. There, the distributions of the training and testing samples do not provide significant differences.

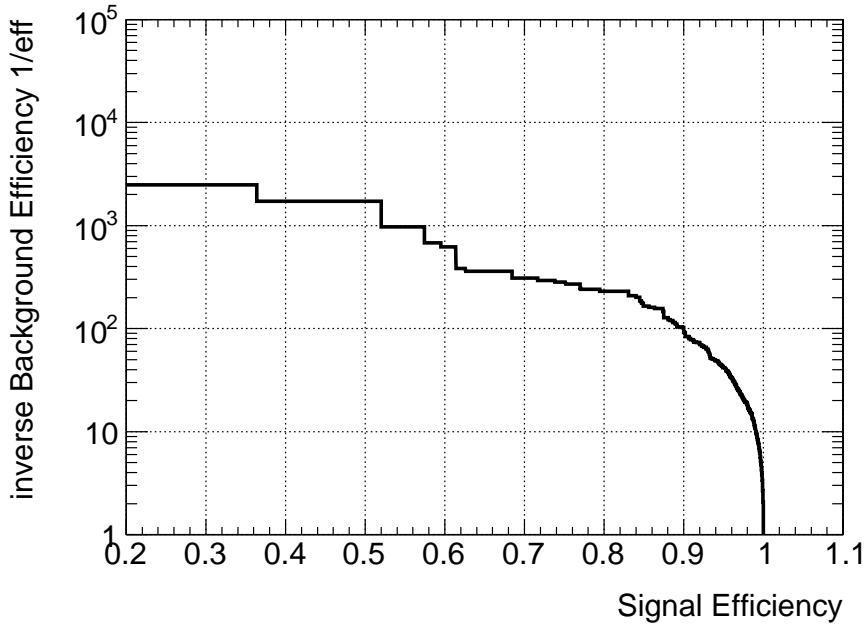


**Figure 5.20:** BDT score distributions for di-tau signal and QCD background. The hatched histograms represent the BDT score distributions of the training samples and the markers represent the BDT score distributions of the testing samples.

### 5.4.3 Results

Already the good separation of the signal and background BDT score distributions in Figure 5.20 indicates a high separation power of the di-tau ID. For a more detailed look the so called ROC curve (Receiver Operating Characteristic) is calculated for the BDT score distributions of the testing samples. Here, the inverse background efficiency versus signal efficiency is illustrated. Each point in a ROC curve is related to a dedicated BDT cut. Events getting a higher BDT score than the cut score are considered signal, events with a lower score are considered background. The signal efficiency is the fraction of testing events, that pass the BDT cut. The background efficiency is the fraction of background events, that are mis-classified as signal. The ROC curve is shown in Figure 5.21. Depending on the desired signal efficiency and background rejection a working point can be defined in the ROC curve, which is then related to a dedicated BDT cut. Here, with a working point at a signal efficiency of for example 90% an inverse background efficiency of 100 can be achieved, meaning that one background event out of 100 is mis-identified as signal.

Next to a high separation power, another important characteristic for the ID is a possibly low dependence on the transverse momentum. This can be achieved with ID variables with preferably low  $p_T$  dependencies, as well as with a  $p_T$  re-weighting of the background or signal sample, to get identical transverse momentum distributions in the training. Usually a low  $p_T$  dependence of the ID is realized by a

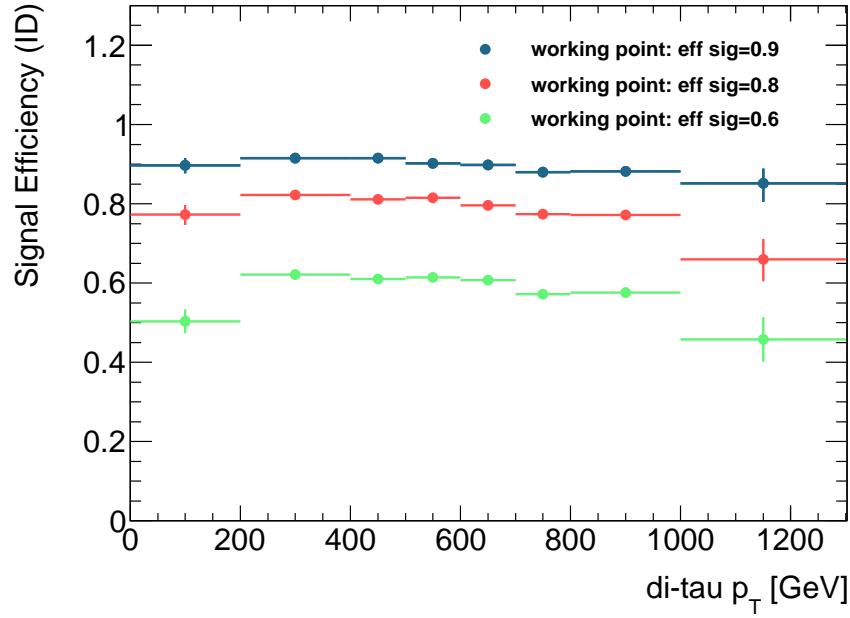


**Figure 5.21:** ROC curve for signal and QCD background discrimination with the di-tau ID.

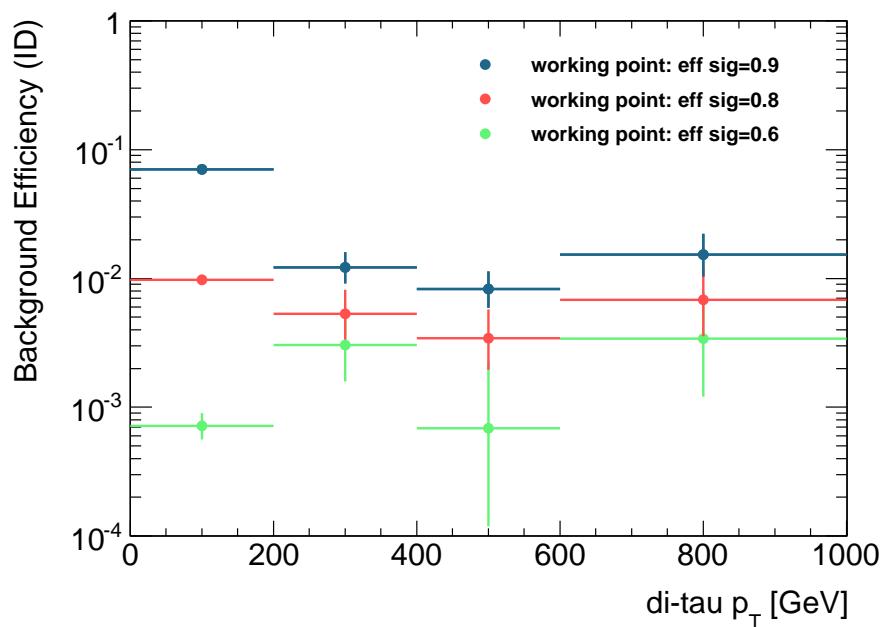
$p_T$  dependent BDT cut. Although here it will be shown that the di-tau ID has a low  $p_T$  dependence already by itself. For that purpose three working points are chosen, related to overall signal efficiencies of 60%, 80% and 90%. For these working points respectively one global BDT cut can be determined. Then the change of signal efficiency for different  $p_T$  areas can be analyzed. The signal efficiency as a function of the transverse momentum is shown in Figure 5.22. The signal efficiency changes with  $p_T$  by approximately 5% at maximum, apart from the first and the last bin. These bins of highest deviations are the bins with the lowest statistics in the training sample, resulting in a less robust training. In the first  $p_T$  bin, only a small number of di-tau events is available, since already the di-tau reconstruction is not sensitive in this area. In the last bin the amount of background is very small, due to the limited amount of ATLAS data available during this study.

In Figure 5.23 the background efficiency versus  $p_T$  is illustrated. Depending on the working point, small background efficiencies between  $10^{-1}$  to  $10^{-3}$  can be achieved. Here, the working point related to a signal efficiency of 0.8 shows the smallest dependence on  $p_T$ .

Finally the signal efficiencies of the di-tau reconstruction and the di-tau ID can be combined. This is illustrated in Figure 5.24 for the three working points at identification efficiencies of 0.6, 0.8 and 0.9. In an area of di-tau  $p_T$  between 400 GeV and 1200 GeV high signal efficiencies can be achieved with efficiencies between 0.5 and 0.8, depending on the desired background rejection. In the  $p_T$  region between 400 GeV and 500 GeV the single-tau tagging, as well as the di-tau approach can

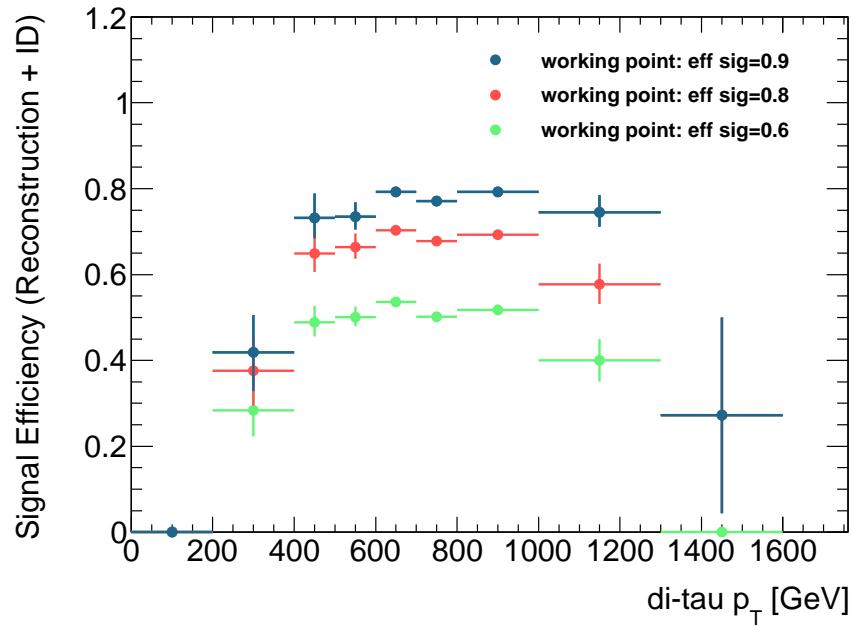


**Figure 5.22:** Signal efficiency as a function of the di-tau transverse momentum for three working points, dedicated to signal efficiencies of 0.6, 0.8 and 0.9.



**Figure 5.23:** QCD background efficiency as a function of the di-tau candidate's transverse momentum for three working points, dedicated to signal efficiencies of 0.6, 0.8 and 0.9.

be applied. A combination of the high reconstruction efficiency of the single-tau reconstruction with the high QCD background discrimination power of the di-tau ID has the potential to improve the boosted tau pair tagging remarkably in this region. In particular the  $p_T$  area above 500 GeV were not accessible for the tau pair tagging before. The di-tau tagging extends the sensitivity for analyses searching for boosted tau leptons up to an energy of 1200 GeV.



**Figure 5.24:** Combined reconstruction and ID signal efficiency as a function of the di-tau transverse momentum for three working points, dedicated to signal efficiencies of 0.6, 0.8 and 0.9.

# Chapter 6

## Outlook

Although a very efficient di-tau tagging method has been presented, there are several sources for potential improvements, that could not be implemented or investigated in the limited time of a master thesis. The comparison to the single-tau approach shows that a highly performant tagging method takes years of enduring work and fine tuning and is nevertheless still under investigation for improvements. Several areas for further improvements of the di-tau tagging method occurred during this study. These are introduced in this chapter. Also an overview of future plans with di-tau tagging is given, as well as a summary of requirements needed to get the di-tau tagging into analysis application.

At the moment, the efficiency of the di-tau reconstruction is mostly limited by the track reconstruction. A subjet that inherits no tracks is not considered for di-tau reconstruction. For reasons of comparability, the same track cuts are used for the di-tau reconstruction as in the single-tau approach. Nevertheless it is advisable to investigate, whether a track cut optimization for high-energy tracks can improve the di-tau reconstruction efficiency.

Also the pile-up removal can be improved in the current di-tau reconstruction. At the moment, di-tau candidates, emerging from pile-up, are rejected with cuts on the seed jet  $p_T$  and with track criteria. But pile-up signatures inside a di-tau candidate are not handled. The influence of these pile-up signatures on the di-tau reconstruction could be investigated. On the one hand pile-up events have a nearly negligible amount of energy compared to the high energy of the observed tau pairs. On the other hand, the amount of pile-up will increase with higher collision rates after the restart of LHC, and the large jet areas of the anti- $k_T$ -10 seed jets are more susceptible to pile-up signatures. For this purpose, the FastJet package, that was already used for seed jet reconstruction, provides algorithms for noise-removal and jet transformation [54]. Such approaches, may let the two-subjet topologies of the di-tau decay appear more significant.

One thing that is not implemented into the di-tau reconstruction yet is an energy

calibration. Although calibrated clusters are used for jet and subjet reconstruction, these are not optimized for the special particle mixture of hadronic tau decays. One possible approach for energy calibration is to calculate  $p_T$  dependent energy correction factors, by comparing the subjet energies with the energies of simulated visible tau decay products.

Another field for possible improvements is the tau vertex association. In the single-tau tagging, the coordinates of jets and tracks are recalculated after the vertex association, in the coordinate system of the new tau vertex. For the di-tau reconstruction and identification it is in question, whether a recalculation can give noticeable improvements, but also how a consistent coordinate system in an overlap region between two subjets can be defined, in cases where the two taus originate from different vertices.

For the di-tau identification, the choice of ID variables should be consistently checked for improvements. In particular the tau decay length is an important characteristic of the tau decay topology. But it was not considered in the current di-tau ID, because of the low discrimination power of the transverse flight path significances  $S_{T,\text{flight}}^{(\text{sub})\text{lead}}$  in the high-energy region. For that purpose, new variable definitions for the tau decay lengths might improve the background discrimination.

Also, for future improvements of the di-tau ID, more background data is required. As long as there is not enough ATLAS data available, simulated QCD jet events can be taken for background. In addition the important  $t\bar{t}$  and di-boson background should be tested. Afterwards, with enough statistics, the number of ID variables can be extended without running into over-training. Higher statistics also allow a closer look into  $p_T$  dependencies of the ID and its variables in the high-energy region.

Avoiding pile-up dependencies in the lower  $p_T$  region may also improve the background discrimination. Similar to the  $p_T$  re-weighting of background events, an additional weighting could be introduced, which depends on the number of vertices in each event.

Finally, it is intended to implement the di-tau reconstruction method into the standard ATLAS event reconstruction chain, which makes the di-tau tagging easily accessible to analysis teams studying  $A \rightarrow Zh \rightarrow ll + \tau\tau$  and  $H \rightarrow hh \rightarrow bb + \tau\tau$  topologies. This requires not only several modifications to the code basis, but also a more robust background separation training on more background topologies, as mentioned above. Also, for an efficient boosted di-tau search in ATLAS data, an algorithm to trigger on di-tau topologies should be developed. In addition, an extension of the di-tau reconstruction for semi-hadronic tau pair decays, where one tau lepton decays hadronically, and one tau decays into an electron or muon plus a neutrino could increase the sensitivity to di-tau signatures.

# Chapter 7

## Summary

The recent discovery of the Higgs boson has once again confirmed the validity of the Standard Model of particle physics. Nevertheless it is limited on a certain range of energy. Several observations have been made, that call for new physics explanations, like the existence of Dark Matter or the fine-tuning problem. One of the extensions of the SM are the Two-Higgs-Doublet models, which predict four additional Higgs bosons. Depending on the model, decay channels involving a boosted pair of tau leptons are promising for the search for new neutral heavy Higgs bosons.

The ATLAS detector at the LHC is well prepared for the search for new physics at center-of-mass energies of up to 14 TeV. The combination of a highly granular inner detector with the precise measurement of energy deposits in the calorimeter gives the opportunity to discover rare processes at the TeV scale, not only for decays involving tau leptons.

The single-tau reconstruction method is a valuable technique for reconstructing hadronically decaying tau leptons in a large energy region. But in cases of topologies with highly boosted tau pairs the reconstruction efficiency drops significantly. Due to the anti- $k_t$ -4 seed jets basis, taus which are closer than  $\Delta R = 0.4$  can not be reconstructed individually. In cases where a SM-like Higgs boson decays into a pair of tau leptons this corresponds to a di-tau  $p_T$  of approximately 500 GeV.

The new di-tau tagging approach is a dedicated reconstruction and identification algorithm for highly boosted tau pairs and is intended to be an addition to the single-tau reconstruction in the high-energy region. By using anti- $k_t$ -10 seed jets, that include the decay topology of both tau leptons, and a later subjet filtering up to 90% of the boosted tau pairs can be reconstructed. High reconstruction efficiencies can be achieved for di-tau transverse momenta of up to 1200 GeV. With the training of the multi-variate di-tau identification, based on Boosted Decision Trees, a high QCD jet rejection was achieved.



# Appendix A

## Units

In this thesis, natural units are chosen. Here, the vacuum speed of light, the reduced Planck constant, and the Boltzmann constant are defined to be all equal to one:

$$c = \hbar = k_B = 1. \quad (\text{A.1})$$

As a consequence, all units of physical quantities can be expressed as powers of the energy. In particle physics, the energy is usually measured in electron volt, which can be converted to joules via

$$1 \text{ eV} = 1.602176565 \times 10^{-19} \text{ J.} \quad (\text{A.2})$$

# Appendix B

## Data Sets

The data sets, that were used for this study, are listed in Table B.2. ATLAS data from 2012 with  $\sqrt{s} = 8$  TeV is taken for QCD jet background. Custom Monte Carlo simulated  $A \rightarrow Zh \rightarrow ll\tau\tau$  decays are taken for signal. In the Monte Carlo sets a  $p_T$  filter is applied, which requires a transverse momentum of the SM-like Higgs boson greater than 600 GeV, to exclude a high majority of not boosted  $\tau$  pairs. The ATLAS data is pre-processed with a high- $p_T$  jet tagger, which searches for anti- $k_T$ -10 jets with a  $p_T > 400$  GeV. In Table B.1 the numbers of events are listed before and after the di-tau reconstruction for di-tau signal and QCD background. Afterwards the reconstructed di-tau candidates are divided into a BDT training and a BDT testing sample, with a ration of 2:1.

**Table B.1:** Numbers of events before and after the di-tau reconstruction for di-tau signal and QCD background.

	Di-tau signal	QCD background
Events used for reconstruction	49800	127739
Events, containing a di-tau candidate	40422	102897
Di-tau candidates with $p_T > 400$ GeV	32281	6164
Candidates in BDT training tree	26948	68598
Candidates in BDT testing tree	13474	34299

**Table B.2:** ATLAS data and Monte Carlo sets, that were used in this study.

---

data:	
	data12_8TeV.00205112.physics.JetTauEtmiss.recon.ESD.r5723
	data12_8TeV.00204910.physics.JetTauEtmiss.recon.ESD.r5723
	data12_8TeV.00204265.physics.JetTauEtmiss.recon.ESD.r5723
	data12_8TeV.00204240.physics.JetTauEtmiss.recon.ESD.r5723
	data12_8TeV.00202740.physics.JetTauEtmiss.recon.ESD.r5723
Monte Carlo samples:	
	user.lhauswal.8TeV_MadGraphPythia8_AU2CTEQ6L1_ggA2000_Zh125_lltautau_hh.filter_h600.ESD.v12bb_s1930_r5676_EXT0.13707648
	user.lhauswal.8TeV_MadGraphPythia8_AU2CTEQ6L1_ggA2000_Zh125_lltautau_hh.filter_h600.ESD.v12bb_s1930_r5676_EXT0.13711321

---

# Appendix C

## BDT Configuration

The BDT training algorithms are provided by TMVA [57] version 4.2.0, which is part of the ROOT [55] 5.34 toolkit. The parameters for BDT training in this study are summarized in Table C.1. These parameters refer to BDTs, with a relatively low depth, to avoid over-training, since there is a low number of background events in the high  $p_T$  region available (see Appendix B).

**Table C.1:** BDT configuration.

---

NTrees	=	50
MaxDepth	=	8
BoostType	=	AdaBoost
AdaBoostBeta	=	0.2
UseYesNoLeaf	=	FALSE
SeparationType	=	GiniIndex
nCuts	=	200
PruneMethod	=	NoPruning
DoBoostMonitor:MinNodeSize	=	1

---

# List of Figures

2.1	Higgs bosons masses at tree level	9
2.2	MSSM branching ratios of uncharged Higgs boson decays	11
3.1	Schematic view of the CERN accelerator complex	13
3.2	Schematic view of LHC experiments	14
3.3	Schematic overview of the full ATLAS Detector and its sub-systems	15
3.4	Overview of the ATLAS Inner Detector	16
3.5	Overview of the full calorimeter system	18
4.1	Signal efficiencies in Run-I for 1-prong and multi-prong tau candidates	25
4.2	Distributions of true $p_T$ for a heavy Higgs boson $A$ and its prompt decay products $Z$ and $h$	26
4.3	Distributions of true tau $p_T$ , for leading taus and subleading taus	27
4.4	Angular distances $\Delta R$ between tau leptons	27
4.5	Di-tau reconstruction efficiency versus di-tau $p_T$	28
4.6	Di-tau reconstruction efficiency versus tau angular distances $\Delta R$	29
4.7	Illustration of the anti- $k_t$ seed jet merging for boosted $\tau_{\text{had}}$ pairs	29
5.1	$A \rightarrow Zh$ example decay topology	31
5.2	Example of a di-tau decay topology	32
5.3	Schematic view of the expected topology in di-tau reconstruction	33
5.4	Illustration of the tau jet vertex association	36
5.5	Di-tau and single-tau reconstruction efficiencies versus tau distance $\Delta R$	38
5.6	Di-tau and single-tau reconstruction efficiencies versus di-tau $p_T$	38
5.7	Example of a QCD jet decay topology	39
5.8	Normalized distributions of the core energy fractions $f_{\text{core}}^{(\text{sub})\text{lead}}$	40
5.9	Normalized distributions of the subjet energy fractions $f_{\text{subjet}}^{(\text{sub})\text{lead}}$	41
5.10	Normalized distributions of the subjet energy fraction $f_{\text{subjets}}$	42
5.11	Normalized distributions of the leading track momentum fractions $f_{\text{track}}^{(\text{sub})\text{lead}}$	42
5.12	Normalized distributions of the maximum track distance $R_{\text{max}}$	43

5.13	Normalized distributions of the track distance $R_{\text{track}}$	43
5.14	Normalized distributions of the number of inner tracks $n_{\text{track}}^{(\text{sub})\text{lead}}$	44
5.15	Normalized distributions of the number of outer tracks $n_{\text{outerTrack}}^{\text{sublead}}$	44
5.16	Normalized distributions of the transverse flight path significances $S_{\text{T,flight}}^{\text{sublead}}$	45
5.17	Normalized distributions of the leading track impact parameter significances $S_{\text{lead track}}^{(\text{sub})\text{lead}}$	46
5.18	Correlation of the ID variables for di-tau signal events	47
5.19	Correlation of the ID variables for QCD background events	48
5.20	BDT score distributions for di-tau signal and QCD background	49
5.21	ROC curve for signal and QCD background discrimination	50
5.22	Signal efficiency versus di-tau $p_{\text{T}}$	51
5.23	QCD background efficiency versus di-tau candidate $p_{\text{T}}$	51
5.24	Combined reconstruction and ID signal efficiency versus di-tau $p_{\text{T}}$	52

# List of Tables

2.1	The gauge bosons of the SM . . . . .	4
2.2	The fermions of the SM . . . . .	4
2.3	Four types of 2HDMs . . . . .	7
2.4	Coupling strengths to a single Higgs boson . . . . .	10
B.1	Numbers of events in used data . . . . .	58
B.2	Used data sets . . . . .	59
C.1	BDT configuration . . . . .	60

# Bibliography

- [1] J. J. Thomson, *Cathode Rays*, Philosophical Magazine 44.293 (1897).
- [2] S. H. Neddermeyer and C. D. Anderson, *Note on the Nature of Cosmic-Ray Particles*, Phys. Rev., 51 (1937) 884.
- [3] M. L. Perl, G. Abrams, A. Boyarski, M. Breidenbach, D. Briggs, F. Bulos, W. Chinowsky, J. Dakin et al., *Evidence for Anomalous Lepton Production in  $e^+e^-$  Annihilation*, Phys. Rev. Lett. 35 (1975) 1489 – 1492.
- [4] K. A. Olive et al. [Particle Data Group Collaboration], *Review of Particle Physics*, Chin. Phys. C 38, 090001 (2014).
- [5] R. Feynman, *Space-Time Approach to Non-Relativistic Quantum Mechanics*, Reviews of Modern Physics 20.2 (1948), 367 – 387.
- [6] J. Schwinger, *On Gauge Invariance and Vacuum Polarization*, Physical Review 82.5 (1951), 664 – 679.
- [7] S.-I. Tomonaga and J. Oppenheimer, *On Infinite Field Reactions in Quantum Field Theory*, Physical Review 74.2 (1948), 224 – 225.
- [8] E. Fermi, *An attempt of a theory of beta radiation. 1.*, Z. Phys. 88 (1934), 161 – 177.
- [9] M. Gell-Mann, *A schematic model of baryons and mesons*, Physics Letters 8.3 (1964), 214 – 215.
- [10] G. Zweig, *An  $SU_3$  model for strong interaction symmetry and its breaking; Version 2*, CERN preprint CERN-TH-412, 1964, <http://cds.cern.ch/record/570209>.
- [11] G. D. Coughlan, J. E. Dodd and B. M. Gripaios, *The Ideas of Particle Physics – An Introduction for Scientists*, Cambridge University Press, New York (2006).
- [12] The ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Physics Letters B 716.1 (2012), 1 – 29. arXiv: 1207.7214 [hep-ex].

- [13] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Physics Letters B 716.1 (2012), 30 – 61. arXiv: 1207.7235 [hep-ex].
- [14] ATLAS Collaboration: *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST 3 S08003 (2008).
- [15] CMS Collaboration, *The CMS experiment at the CERN LHC*, Journal of Instrumentation 3.08 (2008), S08004–S08004.
- [16] B. Povh, K. Rith, C. Scholz and F. Zetsche, *Teilchen und Kerne – Eine Einführung in die physikalischen Konzepte*, Springer, Berlin Heidelberg New York (2004), 172 – 174.
- [17] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. 13 (1964) 321 – 323.
- [18] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. 12 (1964) 132 – 133.
- [19] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. 13 (1964) 508 – 509.
- [20] P. W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, Phys. Rev. 145 (1966) 1156 – 1163.
- [21] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. 13 (1964) 585 – 587.
- [22] A Model of Leptons *Steven Weinberg* Phys. Rev. Lett. 19, 1264 (1967).
- [23] S. L. Glashow, *Partial-Symmetries of Weak Interactions*, Nucl. Phys. 22 (1961) 579.
- [24] A. Salam *Elementary Particle Theory, Relativistic Groups and Analyticity* N. Svartholm (Ed.), Proc. 8th Nobel Symposium, Almqvist and Wiksell, Stockholm (1968), 367.
- [25] A. Djouadi, *The Anatomy of Electro-Weak Symmetry Breaking. II: The Higgs bosons in the Minimal Supersymmetric Model*, Phys. Rept. 459: 1 – 241, 2008, arXiv: hep-ph/0503173.
- [26] S. Weinberg, *Gauge hierarchies*, Phys. Lett. 82B (1979) 387.
- [27] M. Veltman, *The Infrared – Ultraviolet Connection*, Acta. Phys. Polon. B12 (1981) 437.

- [28] C.H. Llewellyn Smith and G.G. Ross, *The real gauge hierarchy problem*, Phys. Lett. 105B (1981) 38.
- [29] S. P. Martin, *A Supersymmetry Primer*, arXiv: hep-ph/9709356.
- [30] Z. Chacko, Y. Nomura, M. Papucci and G. Perez, *Natural little hierarchy from a partially goldstone twin Higgs*, JHEP 0601, 126 (2006), arXiv: hep-ph/0510273.
- [31] J. Mrazek, A. Pomarol, R. Rattazzi, M. Redi, J. Serra, et al., *The Other Natural Two Higgs Doublet Model*, Nucl. Phys. B853 (2011) 1 – 48, arXiv: 1105.5403 [hep-ph].
- [32] J.F. Gunion, H.E. Haber, G. Kane and S. Dawson, *The Higgs Hunter’s Guide*, Addison-Wesley, Reading, (1990).
- [33] N. Craig, J. Galloway and S. Thomas, *Searching for Signs of the Second Higgs Doublet*, arXiv: 1305.2424 [hep-ph].
- [34] L. Evans and P. Bryant, *LHC Machine*, JINST 3 (2008), S08001.
- [35] CERN, *LEP design report*, Report CERN-LEP-84-01. Geneva, 1984.
- [36] The ALICE Collaboration, *The ALICE experiment at the CERN LHC*, Journal of Instrumentation 3.08 (2008), S08002–S08002.
- [37] The LHCb Collaboration, *The LHCb Detector at the LHC*, Journal of Instrumentation 3.08 (2008), S08005–S08005.
- [38] C. Lefèvre, *The CERN accelerator complex*, CERN-DI-0812015 (2008).
- [39] Jean-Luc Caron, *Overall view of LHC experiments*, LHC Project Illustration LHC-PHO-1998-349 (1998).
- [40] W. Lampl et al., *Calorimeter Clustering Algorithms: Description and Performance*, ATL-LARG-PUB-2008-002.
- [41] G. P. Salam and G. Soyez, *A practical Seedless Infrared-Safe Cone jet algorithm*, JHEP 0705 (2007) 086 arXiv: 0704.0292 [hep-ph].
- [42] S. Catani, Y. L. Dokshitzer, M. H. Seymour and B. R. Webber, *Longitudinally-invariant  $k_\perp$ -clustering algorithms for hadron-hadron collisions*, Nucl. Phys. B 406 (1993) 187.
- [43] S. D. Ellis and D. E. Soper, *Successive Combination Jet Algorithm For Hadron Collisions*, Phys. Rev. D 48 (1993) 3160 arXiv: hep-ph/9305266.

- [44] Y. L. Dokshitzer, G. D. Leder, S. Moretti and B. R. Webber, *Better Jet Clustering Algorithms*, JHEP 9708, 001 (1997) arXiv: hep-ph/9707323.
- [45] M. Wobisch and T. Wengler, *Hadronization Corrections to Jet Cross Sections in Deep-Inelastic Scattering* arXiv: hep-ph/9907280.
- [46] M. Cacciari, G. P. Salam and G. Soyez, *The anti- $k_t$  jet clustering algorithm*, JHEP 0804 (2008) 063 arXiv: 0802.1189 [hep-ph].
- [47] ATLAS Collaboration, *Identification and energy calibration of hadronically decaying tau leptons with the ATLAS experiment in pp collisions at  $\sqrt{s} = 8$  TeV*, arXiv: 1412.7086 [hep-ex].
- [48] ATLAS Collaboration, *Determination of the tau energy scale and the associated systematic uncertainty in proton-proton collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector at the LHC in 2012* ATLAS-CONF-2013-044, <http://cds.cern.ch/record/1544036>.
- [49] ATLAS Collaboration, *Performance of the Reconstruction and Identification of Hadronic Tau Decays with ATLAS*, ATLAS-CONF-2011-152, <http://cds.cern.ch/record/1398195>.
- [50] ATLAS Collaboration, *Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons*, ATLAS-CONF-2011-077, <http://cds.cern.ch/record/1353226>.
- [51] ATLAS Collaboration, *Identification of Hadronic Decays of Tau Leptons in 2012 Data with the ATLAS Detector*, ATLAS-CONF-2013-064, <http://cds.cern.ch/record/1562839>.
- [52] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, Journal of High Energy Physics 6 (2010). arXiv: 1002.2581 [hep-ph].
- [53] T. Sjöstrand, S. Mrenna and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. 178 (2008) 852 arXiv: 0710.3820 [hep-ph].
- [54] M. Cacciari, G. P. Salam and G. Soyez, *FastJet user manual*, arXiv: 1111.6097 [hep-ph].
- [55] R. Brun and F. Rademakers, *ROOT – An Object Oriented Data Analysis Framework*, Proceedings AIHENP'96 Workshop, Lausanne, Sep. 1996, Nucl. Inst. & Meth. in Phys. Res. A 389 (1997) 81–86. See also <http://root.cern.ch/>.

- [56] G. van Rossum and F.L. Drake, *Python Reference Manual*, PythonLabs, Virginia, USA, 2001. Available at <http://www.python.org>.
- [57] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag et al. *TMVA: Toolkit for Multivariate Data Analysis*, PoS A CAT 040 (2007) arXiv:physics/0703039 [physics.data-an].
- [58] J. M. Butterworth, A. R. Davison, M. Rubin and G. P. Salam, *Jet substructure as a new Higgs search channel at the LHC*, Phys. Rev. Lett. 100 (2008) 242001 arXiv: 0802.2470 [hep-ph].
- [59] ATLAS Collaboration, *Performance of the Reconstruction and Identification of Hadronic  $\tau$  Decays in ATLAS with 2011 Data*, ATLAS-CONF-2012-142, <http://cds.cern.ch/record/1485531>.
- [60] Y. Freund and R. Shapire, *Experiments with a New Boosting Algorithm*, in *Proceedings 13th International Conference on Machine Learning* (1996).
- [61] Y. Freund and R. E. Schapire, *A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting*, Journal of Computer and System Sciences 55 (1997) 119 – 139.

# Danksagung

Ich möchte mich noch herzlich bei allen bedanken, die zum Gelingen dieser Arbeit beigetragen haben. Allen voran gilt mein Dank Prof. Dr. Arno Straessner und Dr. Wolfgang Mader für die hervorragende Betreuung. Für die viele Unterstützung bei physikalischen und technischen Fragen danke ich Lorenz Hauswald, Sebastian Wahrmund und Dirk Duschinger. Es war mir eine große Freude mit euch zusammenzuarbeiten. Außerdem bedanke ich mich bei allen Mitarbeiterinnen und Mitarbeitern des IKTP und besonders Prof. Dr. Michael Kobel, als Institutsleiter, für die angenehme Arbeitsatmosphäre.

Ein großer Dank gilt auch meinen Freunden, die mir den nötigen Ausgleich zum Physikstudium gegeben haben. Von ihnen danke ich besonders Theresa Werner und Carsten Bittrich, auch für die vielen, manchmal aufreibenden, Lerntage in der Bibliothek. Die Zeit mit euch war toll. Mein größter Dank gilt meinen Eltern und meiner Familie, dafür dass sie mich stets bei meinen Entscheidungen unterstützt haben, auch wenn ich es ihnen vielleicht nicht immer leicht machte.

# Selbstständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

David Kirchmeier  
Dresden, 18. Mai 2015