

# **Optimisation of the Hadronic Tau Identification Based on the Classification of Tau Decay Modes with the ATLAS Detector**

## **Diplomarbeit**

zur Erlangung des akademischen Grades

**Diplom-Physiker**  
(Dipl.-Phys.)

vorgelegt der

**Fakultät Mathematik und Naturwissenschaften der  
Technischen Universität Dresden**

von

**Stefanie Hanisch**

**geboren am 21.06.1990 in Freital**





Eingereicht am 02.02.2015

1. Gutachter: Prof. Dr. Arno Straessner
2. Gutachter: Dr. Frank Siegert



## Abstract

Hadronically decaying tau leptons play an essential role in the LHC physics program. Final states involving tau leptons are important to verify processes of the Standard Model of particle physics at the TeV scale, but are also of high interest for Higgs physics and beyond Standard Model studies, like Higgs CP measurements and  $A \rightarrow Zh$  searches. Due to the high production cross section of QCD jets which are the dominant background, efficient reconstruction and identification techniques are crucial to guarantee an excellent selection of interesting physics events. Therefore, sophisticated multivariate algorithms are used. This thesis presents an advanced concept exploiting the information of reconstructed neutral and charged pions in the ATLAS detector, to access the tau decay substructure, and thus enhance the applicability of the tau identification to a broader field of physics analyses. First, several updates of the general algorithms used within the tau identification are implemented in order to provide a more reliable performance. This thesis focuses on the investigation of a pure substructure based tau identification. Starting with the recalculation of the default identification variables exploiting the dedicated substructure algorithms CellBased and EflowRec, the respective performance is evaluated. Moreover, new variables are defined based on the kinematics of the tau decays. Their impact on the rejection rate of fake taus as well as the correlations between them are studied. Hence, it is possible to recover and for certain configurations even exceed the performance of the currently implemented standard strategy. It can be proven that a pure substructure approach for the identification of tau leptons is achievable, and hence might be featured by various physics analyses.

## Kurzfassung

Hadronisch zerfallende Tau-Leptonen spielen eine wesentliche Rolle im Physikprogramm der LHC-Experimente. Endzustände, in welche Tau-Leptonen involviert sind, haben eine besonderer Bedeutung für die Überprüfung von Standardmodell-Prozessen an der TeV-Skala. Weiterhin sind sie von hohem Interesse für Studien zur Erforschung des Higgs-Bosons, sowie von Modellen jenseits des Standardmodells, wie zum Beispiel Messungen der CP-Eigenschaften des Higgs-Bosons oder Suchen nach  $A \rightarrow Zh$  Zerfällen. Aufgrund der hohen Produktionswirkungsquerschnitte von QCD-Jets, welche den dominanten Untergrund darstellen, sind effiziente Rekonstruktions- und Identifikationsmethoden essentiell, um eine exzellente Auswahl von interessanten Physikereignissen zu garantieren. Deren Umsetzung basiert auf hochentwickelten multivariaten Algorithmen. In dieser Arbeit wird ein erweitertes Konzept präsentiert, welches die Informationen der mit dem ATLAS-Detektor rekonstruierten neutralen und geladenen Pionen verwendet, um die Substruktur von Tau-Zerfällen aufzulösen. Die Anwendbarkeit der Tau-Identifikation wird somit auf ein breiteres Anwendungsgebiet ausgeweitet. Dazu wurden im Vorfeld notwendige Aktualisierungen der allgemeinen Algorithmen vorgenommen. Der Fokus der vorliegenden Arbeit liegt jedoch im Studium einer reinen Substruktur-basierenden Tau-Identifikation. Beginnend mit der Neuberechnung der Standardidentifikationsvariablen mittels der dedizierten Substrukturalgorithmen CellBased und EflowRec werden deren Möglichkeiten zur Unterdrückung von QCD-jets analysiert. Des Weiteren werden neue Variablen basierend auf der Kinematik des Tau-Zerfalls definiert. Deren Einfluss auf die Untergrundunterdrückungsrate sowie deren Korrelationen werden untersucht. Infolgedessen ist es möglich, die Performanz der momentanen Standardidentifikation zu erreichen. Für bestimmte Konfigurationen kann diese auch übertroffen werden. Es kann gezeigt werden, dass eine reine Substruktur-basierte Identifikation von Tau-Leptonen realisiert werden kann, welche in diversen Physikanalysen Anwendungen finden kann.



# Contents

<b>Abstract</b>	<b>III</b>
<b>Kurzfassung</b>	<b>III</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Theoretical Foundations</b>	<b>3</b>
2.1 The Standard Model of Particle Physics . . . . .	3
2.2 Electroweak Theory . . . . .	5
2.2.1 Electroweak Symmetry Breaking and Higgs Mechanism . . . . .	6
2.2.2 Discovery of a Higgs Boson . . . . .	8
2.3 Physics Beyond the Standard Model . . . . .	9
2.3.1 Limitations of the Standard Model . . . . .	9
2.3.2 Theoretical Prospects . . . . .	10
2.4 Physics with Tau Leptons . . . . .	11
2.4.1 Standard Model Processes . . . . .	12
2.4.2 Searches for New Physics . . . . .	12
<b>3 The ATLAS Experiment</b>	<b>15</b>
3.1 The Large Hadron Collider . . . . .	15
3.2 The ATLAS Experiment . . . . .	17
3.2.1 Nomenclature . . . . .	17
3.2.2 Tracking System . . . . .	18
3.2.3 Calorimeter System . . . . .	20
3.2.4 Muon System . . . . .	21
3.2.5 Trigger System . . . . .	22
3.2.6 Forward Detectors . . . . .	23
<b>4 Simulation and Data</b>	<b>25</b>
4.1 Monte Carlo Simulation . . . . .	25
4.2 Data-taking in 2012 . . . . .	26
<b>5 Tau Reconstruction and Identification</b>	<b>29</b>
5.1 Reconstruction of Tau Leptons . . . . .	29
5.2 Decay Mode Classification of Tau Leptons . . . . .	32
5.2.1 CellBased Algorithm . . . . .	33
5.2.2 EflowRec Algorithm . . . . .	34
5.2.3 PanTau algorithm . . . . .	35
5.3 Identification of Tau Leptons . . . . .	35
5.3.1 Tau Identification Variables . . . . .	36
5.3.2 Discrimination Against Jets . . . . .	41
5.3.3 Discrimination against Light Leptons . . . . .	48
5.4 Tau Trigger . . . . .	49
<b>6 Optimisation of Tau Lepton Identification</b>	<b>51</b>
6.1 Optimisation of the Default Tau Identification . . . . .	51
6.1.1 Optimisation of the Boosted Decision Tree Based Tau Identification . . . . .	51
6.1.2 Optimisation of the Default Log-Likelihood Based Tau Identification . . . . .	53
6.2 Substructure Based Tau Identification . . . . .	57
6.2.1 Recalculation of Default Variables . . . . .	57
6.2.2 Investigation of Isolation Cone Variables . . . . .	58

6.2.3	Extended Variable Sets . . . . .	67
<b>7</b>	<b>Summary and Outlook</b>	<b>79</b>
<b>A</b>	<b>Auxiliary Information for the Theoretical Foundations</b>	<b>81</b>
A.1	Pauli Matrices . . . . .	81
A.2	Gamma Matrices . . . . .	81
<b>B</b>	<b>Samples</b>	<b>82</b>
<b>C</b>	<b>BDT Configuration</b>	<b>84</b>
<b>D</b>	<b>Auxiliary Material for Optimisation of Tau Identification</b>	<b>85</b>
D.1	Recalculated Variables . . . . .	85
D.2	Isolation Variables . . . . .	88
D.3	Further Variables . . . . .	90
D.4	Correlation Matrices . . . . .	94
	<b>Bibliography</b>	<b>103</b>
	<b>List of Figures</b>	<b>109</b>
	<b>List of Tables</b>	<b>113</b>



# 1 Introduction

The first initiative for modern particle physics was given by the discovery of the *electron* by J. J. Thomson in 1897. Though, the first concept of fundamental particles goes back to the ancient Greeks who introduced the *atom* as an indivisible building block of matter. Although, this theory turned out to be wrong by the measurement of a sub-division into a nuclei and a shell. It was finally disproved by the discovery of the *proton* and the *neutron* as the components of the nuclei which is surrounded by electrons. Nevertheless, the aim of a general concept, which describes the matter surrounding us, was kept. After the observation of the first elementary particle, many groundbreaking theoretical developments and experimental observations followed, e.g. the *general theory of relativity* (A. Einstein 1916), *Quantum Electrodynamics* (P. Dirac 1927) and the discovery of the partons (SLAC, 1968), as the composites of the proton, which later turned out to be the *up* and *down quark*. In the 1960's the concept of the *Standard Model of particle physics* (SM) [1–3] was formulated by S. Glashow, A. Salam and S. Weinberg, which does not only postulate the existence of 12 elementary particles but also provides a description of their interactions with each other mediated by so-called *gauge bosons*. The evidence for the predicted  $W^\pm$  and  $Z^0$  gauge bosons was succeeded in 1983 by the CERN experiments UA1 and UA2 [4, 5]. The discovery of the *gluon* followed in 1978, achieved by the PETRA experiments at DESY [6]. The SM describes our current knowledge of particle physics and is validated with an incredible precision by dozens of experiments. The final missing part, the *Higgs boson*, was first proposed in 1964 by Brout, Englert and Higgs within the theoretical description of the *Higgs mechanism* [7–11] which provides an explanation of the origin of fundamental particle masses. After a journey of almost 50 years, it was finally observed in 2012 [12, 13] at the *Large Hadron Collider* (LHC) [14] by the experiments ATLAS<sup>1</sup> [15] and CMS<sup>2</sup> [16].

The LHC is currently the largest collider experiment of particle physics providing the highest centre-of-mass energies of 8 TeV reached so far. It started operation in 2008, but an interruption short after the beginning delayed data-taking towards early 2009. During about three years protons and heavy ions were collided until late 2012. After a technical stop due to a phase of maintenance and upgrades, the LHC will restart operation again in early 2015, aiming for even higher centre-of-mass energies of 13 TeV and 14 TeV. This will give access to phase space regions where new physics might show up. Even though the SM is a very successful theory, it cannot explain all observed phenomena and open questions remain, like the existence of *Dark Matter* and *Dark Energy* [17–19], the *hierarchy problem* [20] or a *Grand Unification* [21] of all forces. The SM might be only an effective theory, hence concepts for *Beyond Standard Model* (BSM) physics aim to address these problems. A very popular approach is given by the *Minimal Supersymmetric extension of the Standard Model* (MSSM) [22–25] within the wide field of *Supersymmetry* (SUSY) [26, 27]. One consequence of those theories is a more complex Higgs sector which also includes the observed Higgs boson at 125 GeV [12, 13] which is similar but not exactly the one predicted by the SM. Therefore, it is essential to further determine the characteristics of the Higgs boson, e.g. the nature of its spin which can be accessed via the decay into two *tau leptons*.

Final states involving tau leptons are not only interesting for Higgs studies but also for other SM processes and various searches for new physics. In the latter case the tau lepton as the heaviest lepton, is assigned a leading role in several BSM theories since its final states are favoured over a wide phase space region. It was discovered in 1974 at the Stanford electron-positron collider as the third and last charged lepton covered within the SM. Furthermore, it is the only lepton which can decay both leptonically and hadronically. Its detection in proton-proton collision at the LHC is highly challenging due to the enormous QCD background, which yields similar signatures in the detector. Thus, efficient reconstruction and identification algorithms are required. Therefore, a continuous optimisation of the algorithms used is crucial to cope with changes e.g. in reconstruction or experimental set-up. Investigations of new approaches to improve the performance or to provide access to a wider phase space region are necessary to give room for various applications. Recent developments have brought up a new reconstruction strategy which allows to explore the proper decay mode of the hadronic tau decay. This might be in particular of

---

<sup>1</sup>A Toroidal LHC Apparatus

<sup>2</sup>Compact Muon Solenoid

interest for Higgs CP studies.

In this thesis a novel purely substructure based tau identification is developed. This will be able to serve as an alternative concept to the one currently featured by the ATLAS experiment. This approach is able to provide a better performance and further might even facilitate access to an extended phase space region not covered by the current identification technique. The analyses are based on *Monte Carlo* (MC) simulated events, serving as signal, and a selection of QCD di-jet events extracted from 8 TeV data taken with the ATLAS detector in 2012, used as background. The results are compared to the configuration applied by ATLAS in the 2011 and 2012 data analyses. Several improvements are applied to the standard algorithms to guarantee an ongoing excellent performance of the tau identification in general. This will also support the novel substructure based approach.

The theoretical foundations for the studies presented are outlined in Chapter 2 covering the SM as underlying theory including the Higgs sector, as well as the wide field of tau physics within BSM models. Chapter 3 gives an overview of the experimental framework, i.e. the LHC and the multipurpose detector ATLAS. A closer look on the events extracted from MC simulation as well as data-taking conditions is provided in Chapter 4. The tau reconstruction and identification as implemented in the current online and offline framework is discussed in detail in Chapter 5. This chapter also describes dedicated algorithms which provide access to the tau decay substructure. Chapter 6 focuses on the optimisation of the standard identification and introduces a new approach of tau identification based on the decay mode classification. The final results of those analyses as well as an outlook of further related studies are summarised in Chapter 7.

## 2 Theoretical Foundations

This chapter outlines the theoretical concepts of modern particle physics. The *Standard Model* (SM) [1–3], validated by many precision experiments, is the most successful theory describing the nature of elementary particles and their interactions. First in Section 2.1 the basic foundations of the SM will be introduced. As a part of the SM the electroweak theory is discussed in detail in Section 2.2 including the *Higgs Mechanism* [7–11]. Many theoretical extensions of the SM exist as there are many open questions which cannot be addressed by it. A selection of possible extensions dedicated to physics involving tau leptons will be discussed in Section 2.3. Finally, physical processes including tau leptons within and beyond the SM are covered by Section 2.4.

### 2.1 The Standard Model of Particle Physics

The Standard Model of particle physics is currently the most accurate description of elementary particles and their interactions. It is a relativistic *quantum field theory* (QFT) based on a  $SU(3)_C \times SU(2)_L \times U(1)_Y$  gauge group, whereby  $C$ ,  $L$  and  $Y$  denote colour, left chirality and weak hypercharge, respectively. The  $SU(3)_C$  term represents the strong interaction describing, e.g. the physics within the proton and the  $SU(2)_L \times U(1)_Y$  part stands for the electroweak interaction which is the unification of the electromagnetic and weak forces responsible for e.g. electromagnetism and  $\beta$ -decays, respectively. The latter will be described in detail in Section 2.2. A mathematical formulation is given by a gauge invariant Lagrangian. According to *Noether's Theorem* [28, 29] the invariance of an action under a certain symmetry transformation implies a conservation law. The local gauge invariance of the Lagrangian of the SM leads to a description of dynamics and interactions of elementary particles.

Each interaction is related to a specific charge and mediated by a gauge boson. The mediating particle of the electromagnetic interaction is the photon,  $\gamma$ , with the associated charge being the electromagnetic charge,  $Q$ . The photon is electrically neutral and massless. Hence, it cannot interact with itself. Contrary, the gauge bosons of the weak interaction,  $Z^0$  and  $W^\pm$ , can interact with each other as they carry a weak Isospin,  $\vec{I}$ , which is the charge of the weak interaction. Furthermore, the  $Z^0$  and  $W^\pm$  bosons are rather heavy with masses of  $m_Z = (91.1876 \pm 0.0021) \text{ GeV}$  and  $m_W = (80.385 \pm 0.015) \text{ GeV}$  [19], respectively. As a consequence of the *Heisenberg uncertainty principle*<sup>3</sup> [30], their effective interaction range is limited to a distance of the order of  $10^{-18} \text{ m}$ . For the same reason the mean life time of those particles is restricted to be in the order of  $10^{-25} \text{ s}$ . The gauge boson of the strong interaction is the gluon,  $g$ , with the corresponding colour charge,  $C$ . There are eight gluons which carry a mixture of the (anti-)colours (*anti*-)red,  $r(\bar{r})$ , (*anti*-)green,  $g(\bar{g})$ , and (*anti*-)blue,  $b(\bar{b})$ , and thus self-interaction is possible. Although they are massless, their interaction range is limited by *confinement*. Thus, colour charged particles, like quarks, can only be observed as bounded, colour neutral objects, so-called *hadrons*. Based on this phenomena a large potential energy rises if a single quark occurs and additional quark-anti-quark pairs emerge. This is called *hadronisation*. Hadrons can be categorised into *mesons* and *baryons* which are composites of two and three quarks, respectively. Nevertheless, quarks can be assumed as quasi-free particles at very small distances or very high energies, as can be reached in high energy proton-proton collisions (cf. Section 3). This is described as *asymptotic freedom*.

The strength of an interaction and its dependency on the transferred energy,  $q$ , are characterised by a coupling parameter,  $\alpha$ , which depends on the renormalisation scale. The coupling parameters of the three interactions are parameters of the SM and cannot be predicted by theory but have to be measured by experiments. However, their evolution with energy can be predicted by the *Renormalisation Group Equations* (RGEs) [21]. Hence, the coupling parameter of the strong force,  $\alpha_s$ , increases with decreasing  $|q|$ , and vice versa. Unlike the coupling parameter of the electromagnetic interaction,  $\alpha_{em}$ , which shows an opposite behaviour. The corresponding field theory of electromagnetism, *Quantum Electrodynamics*

---

<sup>3</sup>The Heisenberg uncertainty principle relates the uncertainty in simultaneous position and momentum measurements in quantum mechanics given by  $\Delta x \cdot \Delta p \geq \hbar/2$  or in term of the energy-time relation  $\Delta E \cdot \Delta t \geq \hbar/2$ .

(QED), is well understood and tested. Its coupling parameter can be determined to [31, 32]:

$$\alpha_{em}(q=0) = \frac{e^2}{4\pi} \approx \frac{1}{137} \approx 7.30 \cdot 10^{-3}, \quad (2.1)$$

where the electromagnetic charge magnitude is denoted by  $e$ .

The particle content of the SM can be classified by various *quantum numbers*, e.g. the *spin*, whereby integer values correspond to bosons and half-integer values to fermions. The group of bosons includes the already introduced gauge bosons  $\gamma$ ,  $g$ ,  $W^\pm$  and  $Z^0$ , which are summarised in Table 2.1.

interaction	gauge boson	mass [GeV]	effective range [m]
electromagnetic	photon ( $\gamma$ )	$< 1 \cdot 10^{-27}$	$\infty$
weak	W-boson ( $W^\pm$ )	80.385	$\sim 10^{-18}$
	Z-boson ( $Z^0$ )	91.188	
strong	gluon ( $g$ )	0 <sup>4</sup>	$< 10^{-15}$

**Table 2.1:** Fundamental interactions described by the Standard Model and mediating gauge bosons including their mass and the effective range of the interaction [19].

Fermions, which are also called matter particles, are further divided into leptons and quarks. They can be arranged in three generations which only differ in mass, while the coupling structure remains the same. Only particles of the lightest generation, namely electron, up and down quark, are constituents of ordinary matter. Heavy particles like the representatives of the third generation, decay into lighter particles unless the decay is forbidden by fundamental conservation laws. Neutrinos,  $\nu$ , as a sub-group of leptons, are nearly massless and interact only weakly. Hence, it is challenging to detect them. An overview of these particles and their characterising physical properties is given in Table 2.2. Furthermore, an anti-particle can be assigned to each particle whereby they have equal masses and life times but opposite charges.

generation	particle	electric charge, $Q$ [e]	weak isospin, $I_3$	colour, $C$	mass, $m$ [GeV]
I	$e$ electron	-1	$-\frac{1}{2}$	—	0.000511
	$\nu_e$ electron neutrino	0	$+\frac{1}{2}$	—	$< 2 \cdot 10^{-9}$
	$u$ up quark	$+\frac{2}{3}$	$+\frac{1}{2}$	r,g,b	0.0023
	$d$ down quark	$-\frac{1}{3}$	$-\frac{1}{2}$	r,g,b	0.0048
II	$\mu$ muon	-1	$-\frac{1}{2}$	—	0.106
	$\nu_\mu$ muon neutrino	0	$+\frac{1}{2}$	—	$< 0.19 \cdot 10^{-6}$
	$c$ charm quark	$+\frac{2}{3}$	$+\frac{1}{2}$	r,g,b	1.275
	$s$ strange quark	$-\frac{1}{3}$	$-\frac{1}{2}$	r,g,b	0.095
III	$\tau$ tau	-1	$-\frac{1}{2}$	—	1.777
	$\nu_\tau$ tau neutrino	0	$+\frac{1}{2}$	—	$< 18.2 \cdot 10^{-6}$
	$t$ top quark	$+\frac{2}{3}$	$+\frac{1}{2}$	r,g,b	173.2
	$b$ bottom quark	$-\frac{1}{3}$	$-\frac{1}{2}$	r,g,b	4.18

**Table 2.2:** Fermionic particle content of the Standard Model with corresponding charges and masses. [19].

<sup>4</sup>theoretical value

## 2.2 Electroweak Theory

The electromagnetic and weak forces are unified to the electroweak force based on the theoretical formulation by Glashow, Salam and Weinberg [1–3]. The underlying gauge group is a  $SU(2)_L \times U(1)_Y$  which decomposes under *spontaneous symmetry breaking* (SSB) to the electromagnetic  $U(1)_{em}$  and the weak gauge group  $SU(2)_L$ . The associated electromagnetic charge,  $Q$ , and the weak isospin,  $T_3$ , are related to the electroweak hypercharge,  $Y$ , by the *Gell-Mann-Nishijima relation* [33, 34] according to:

$$Y = Q - T_3 \quad \text{with} \quad T_a = \frac{\sigma_a}{2}, \quad (2.2)$$

with  $\sigma_a$  denoting the *Pauli matrices* which are defined in Appendix A.1.

The electroweak Lagrangian is invariant under a local gauge transformation:

$$U(x) = e^{i\alpha(x)Y} e^{i\sum_a \theta_a(x)T^a}, \quad (2.3)$$

where the two exponential terms are based on the  $U(1)_Y$  and  $SU(2)_L$  gauge groups with their respective generators  $Y$  and  $T^a$  as well as gauge functions  $\alpha(x)$  and  $\theta_a(x)$ . The two gauge groups provide the gauge fields  $B^\mu$  and  $W_a^\mu$  ( $a = 1, 2, 3$ ) which yields the respective field strength tensors:

$$B^{\mu\nu} = \partial^\mu B^\nu - \partial^\nu B^\mu \quad \text{and} \quad F_i^{\mu\nu} = \partial^\mu W_i^\nu - \partial^\nu W_i^\mu - g_w \epsilon^{ijk} W_j^\mu W_k^\nu \quad (2.4)$$

for  $SU(2)_L$  and  $U(1)_Y$ , respectively. Thus, the covariant derivative is given by:

$$D^\mu = \partial^\mu + ig_w T_a W_a^\mu + ig_Y Y B^\mu, \quad (2.5)$$

with the corresponding coupling constants  $g_w$  and  $g_Y$ . Given  $D^\mu$  the fermionic interactions can be described.

The gauge fields  $W_a^\mu$  and  $B^\mu$  transform under the unitary transformation of Equation 2.3 according to:

$$W_a^\mu \rightarrow W_a^\mu - \frac{1}{g_w} \partial^\mu \theta_a(x) + \epsilon_{abc} \theta^b(x) W_c^\mu, \quad (2.6)$$

$$B^\mu \rightarrow B^\mu - \frac{1}{g_Y} \partial^\mu \alpha(x). \quad (2.7)$$

Hence, the gauge groups  $SU(2)_L$  and  $U(1)_Y$  are represented by the gauge fields  $W_a^\mu$  and  $B^\mu$  and the coupling constants  $g_w$  and  $g_Y$ , respectively. Given the non-abelian character of the  $SU(2)_L$  gauge group the observed mass eigenstates are defined as a mixture of those gauge fields:

$$W^{\pm,\mu} = \frac{1}{\sqrt{2}} (W_1^\mu \mp iW_2^\mu), \quad (2.8)$$

$$\begin{pmatrix} A^\mu \\ Z^\mu \end{pmatrix} = \begin{pmatrix} \cos \vartheta_W & \sin \vartheta_W \\ -\sin \vartheta_W & \cos \vartheta_W \end{pmatrix} \begin{pmatrix} B^\mu \\ W_3^\mu \end{pmatrix}. \quad (2.9)$$

Thus, the mixed fields  $W^{\pm,\mu}$ ,  $Z^\mu$  and  $A^\mu$  can be identified with the physically observed bosons  $W^\pm$ ,  $Z^0$  and the photon,  $\gamma$ . In Equation 2.9,  $\vartheta_W$  denotes the *Weinberg mixing angle* which is given by the coupling strengths:

$$\tan \vartheta_W = \frac{g_w}{g_Y}, \quad (2.10)$$

and related to the electric charge by:

$$e = g_w \sin \vartheta_W = g_Y \cos \vartheta_W. \quad (2.11)$$

Low energy experiments, like the Wu-experiment [35], show that only left-handed particles and right-handed anti-particles participate in weak interactions. Thus, parity conservation is violated, which leads

to a chiral gauge group as implied by the index  $L$  referring to left chirality. The left- and right-handed states of a fermion field,  $\psi$ , are given by:

$$\psi_L = \frac{1}{2}(1 - \gamma_5)\psi \quad \text{and} \quad \psi_R = \frac{1}{2}(1 + \gamma_5)\psi, \quad (2.12)$$

with  $\gamma_5 = i\gamma_0\gamma_1\gamma_2\gamma_3$  (see Appendix A.2). As a result the left-handed fermions can be arranged in SU(2) doublets, whereas the right-handed fermions are represented by singlets. However, the down-type quark and neutrino flavour eigenstates,  $q' = (d', s', b')$  and  $\nu_f = (\nu_e, \nu_\mu, \nu_\tau)$ , are not the experimentally observed mass eigenstates,  $q = (d, s, b)$  and  $\nu = (\nu_1, \nu_2, \nu_3)$ , but are related by:

$$q'_i = \sum_j V_{ij}^{CKM} q_j \quad \text{and} \quad \nu_f = \sum_i U_{\alpha i}^{PMNS} \nu_i, \quad (2.13)$$

with the CKM<sup>5</sup> quark-mixing matrix [36],  $V^{CKM}$ , and the PMNS<sup>6</sup> neutrino-mixing matrix [37],  $U^{PMNS}$ . Finally the electroweak Lagrangian can be written as:

$$\mathcal{L}_{\text{ew}} = \underbrace{-\frac{1}{4}F_{\mu\nu}^i F_i^{\mu\nu} - \frac{1}{4}B_{\mu\nu} B^{\mu\nu}}_{\text{kinematic energies and self-interaction of gauge bosons}} + \underbrace{\sum_f i\bar{\psi}_f D_\mu \gamma^\mu \psi_f}_{\text{kinematic energies and interaction of fermions}}. \quad (2.14)$$

Although this Lagrangian provides an excellent description of electroweak interactions and the nature of the participating particles, it cannot account for the masses of the gauge bosons  $W^\pm$  and  $Z^0$  which are experimentally well confirmed. Including mass terms would violate the gauge invariance of  $\mathcal{L}_{\text{ew}}$ . One solution to generate mass terms is realised by the Higgs mechanism which is described in the next section.

### 2.2.1 Electroweak Symmetry Breaking and Higgs Mechanism

The Englert-Brout-Higgs mechanism [7–11] describes the generation of gauge invariant mass terms by spontaneous symmetry breaking. Therefore, a new scalar field with spin 0, the *Higgs field*,  $\phi$ , is introduced. It needs to have a non-vanishing hypercharge and weak isospin to guarantee the interaction with the gauge bosons. The simplest assumption for  $\phi$  is a complex isospin doublet:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \phi_3 + i\phi_4 \\ \phi_1 + i\phi_2 \end{pmatrix}, \quad (2.15)$$

with hypercharge  $Y = 1/2$  and the four real scalar fields  $\phi_i$ . The additional Higgs Lagrangian to the SM can then be written as:

$$\mathcal{L}_H = \underbrace{\sum_f c_f (\bar{\psi}_f^L \phi^\dagger \psi_f^R + \bar{\psi}_f^R \phi \psi_f^L)}_{\text{fermionic mass term}} + \underbrace{D_\mu \phi D^\mu \phi}_{\text{dynamic term}} - \underbrace{V(\phi)}_{\text{Higgs potential}}. \quad (2.16)$$

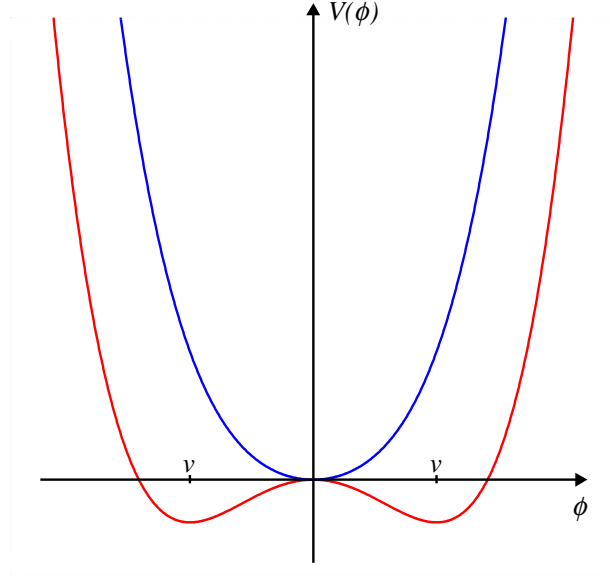
The Higgs potential,  $V(\phi)$ , is defined in a gauge invariant and renormalisable way as:

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2, \quad (2.17)$$

in which  $\mu$  and  $\lambda$  are free parameters. To ensure, that the potential is bounded from below,  $\lambda$  has to be positive. Spontaneous symmetry breaking, i.e. the Lagrangian being invariant under a certain symmetry but not the lowest energy state, requires  $\mu^2$  to be negative. Otherwise, only the trivial minimum  $\phi_i = 0$  exists. The Higgs potential for positive and negative  $\mu^2$  is depicted in Figure 2.1. Given that the Higgs

<sup>5</sup>Cabibbo-Kobayashi-Maskawa

<sup>6</sup>Pontecorvo-Maki-Nakagawa-Sakata



**Figure 2.1:** Higgs potential  $V(\phi)$  for arbitrary positive values for  $\lambda$  and  $\mu^2 < 0$  (blue) and  $\mu^2 > 0$  (red).

potential is invariant under rotation, an infinite number of stable ground states exist, but conveniently chosen to be:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \quad \text{with} \quad v = \sqrt{\frac{-\mu^2}{\lambda}}, \quad (2.18)$$

in which  $v$  denotes the *vacuum expectation value* (vev). The vev can be related to the *Fermi constant* by

$$v = \frac{1}{\sqrt{\sqrt{2}G_F}} \sim 246 \text{ GeV}. \quad (2.19)$$

The choice of the local minimum breaks the  $SU(2)_L \times U(1)_Y$  symmetry to  $U(1)_{em}$  and yields three massless *Goldstone bosons* [38]. These non-physical *degrees of freedom* can be absorbed by the  $W^\pm$  and  $Z^0$  gauge bosons via gauge transformation. Additional longitudinal polarised terms are added to the Lagrangian, which then lead to mass terms of the heavy gauge bosons. The remaining degree of freedom can be associated with a massive particle, the *Higgs boson*,  $H$ . Thus, the Higgs field can be written as:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H \end{pmatrix}. \quad (2.20)$$

Considering this and Equation 2.5, the kinematic component and the Higgs potential can be evaluated to:

$$\mathcal{L}^{kin} = \left(\frac{g_w v}{2}\right)^2 W_\mu^+ W^{-,\mu} + \frac{v^2}{8} \begin{pmatrix} W_\mu^3 & B_\mu \end{pmatrix} \begin{pmatrix} g_w^2 & -g_w g_Y \\ -g_w g_Y & g_Y^2 \end{pmatrix} \begin{pmatrix} W^{3,\mu} \\ B^\mu \end{pmatrix} + \lambda v^2 H^2 + \mathcal{O}(H, H^2), \quad (2.21)$$

$$V(\phi) = \frac{\mu^2}{2} \begin{pmatrix} 0 & v + H \end{pmatrix} \begin{pmatrix} 0 \\ v + H \end{pmatrix} + \frac{\lambda}{4} \left| \begin{pmatrix} 0 & v + H \end{pmatrix} \begin{pmatrix} 0 \\ v + H \end{pmatrix} \right|^2 = \lambda v^2 H^2 + \dots \quad (2.22)$$

Thus, the boson masses are defined as:

$$m_H = v \sqrt{2\lambda}, \quad m_Z = \frac{v}{2} \sqrt{g_w^2 + g_Y^2}, \quad m_W = \frac{v g_w}{2} \quad \text{and} \quad m_\gamma = 0. \quad (2.23)$$

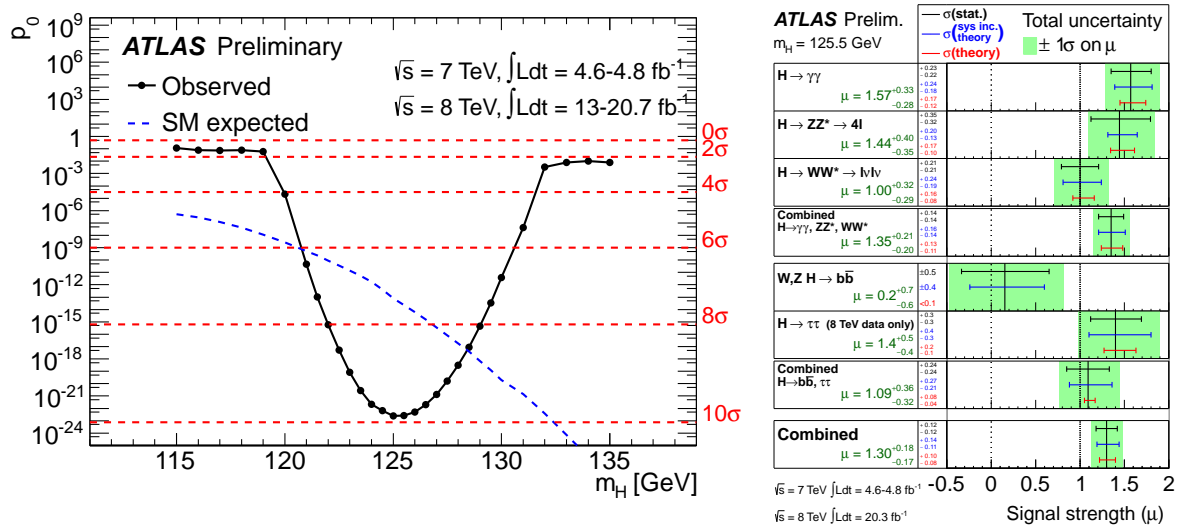
In the same way the couplings to bosons as well as self-coupling terms of the Higgs boson are calculated depending on the mass of the involved particles. The boson masses arise as a consequence of the SSB, while the fermion masses need to be introduced “by hand” via the *Yukawa coupling*, cf. Equation 2.16. The Higgs mass,  $m_H$ , remains the only free parameter in the SM which cannot be predicted by theory, and thus needs to be determined by experimental measurement.



### 2.2.2 Discovery of a Higgs Boson

After two years of data-taking, on July 4<sup>th</sup> 2012, the ATLAS and CMS collaborations announced the discovery of a new boson [12, 13], with properties consistent with the above mentioned SM Higgs boson.

The ATLAS experiment observed a  $5.9\sigma$  excess in a mass range of 122-131 GeV whereby the mass ranges 111-122 GeV and 131-559 GeV could be excluded at 95 % *confidence level* (C.L.). For this purpose, data taken in 2011 with an integrated luminosity corresponding to  $4.6\text{-}4.8\text{ fb}^{-1}$  and a centre-of-mass energy of 7 TeV and  $5.8\text{-}5.9\text{ fb}^{-1}$  of  $\sqrt{s} = 8\text{ TeV}$  data from 2012 were combined. To confirm an observation a standard deviation of at least  $5\sigma$  is required which could be achieved by considering the following decay channels  $H \rightarrow \gamma\gamma$ ,  $H \rightarrow ZZ^*$  and  $H \rightarrow WW^*$ . Afterwards analyses were extended to the full 2012 dataset of  $20.7\text{ fb}^{-1}$  at  $\sqrt{s} = 8\text{ TeV}$ , and thus the Higgs mass was determined to  $125.2 \pm 0.2(\text{stat}) \pm_{-0.6}^{+0.5}(\text{sys})\text{ GeV}$  with a local  $p$ -value corresponding to almost  $10\sigma$  in the high mass resolution channels  $H \rightarrow \gamma\gamma$  and  $H \rightarrow ZZ^*$  [39]. Furthermore, the  $H \rightarrow \tau\tau$ <sup>7</sup> and  $H \rightarrow b\bar{b}$  final states were considered for combination to extract a signal strength, defined as scaling parameter of the predicted standard model production cross section, of  $\mu = 1.3 \pm 0.12(\text{stat}) \pm_{0.11}^{+0.14}(\text{sys})$  for  $m_H = 125.5\text{ GeV}$  [40]. Both channels were excluded for the announcement of the discovery because they could not provide enough sensitivity by that time. For the  $H \rightarrow \tau\tau$  channel a  $4.5\sigma$  excess with  $\mu = 1.42_{-0.38}^{+0.44}$  for a Higgs mass of 125 GeV could be achieved [41]. These results are summarised in Figure 2.2 and are in agreement with the SM prediction.



**Figure 2.2:** Observed local  $p_0$ -value as a function of the Higgs mass [39] (left) and best-fit values for the signal strength  $\mu$  for individual decay channels [40] (right) based on the full dataset taken with the ATLAS detector at  $\sqrt{s}=7\text{ TeV}$  and  $\sqrt{s}=8\text{ TeV}$ .

At the announcement of the discovery of a new particle the CMS experiment published comparable values of  $125.3 \pm 0.4(\text{stat}) \pm 0.5(\text{sys})\text{ GeV}$  and  $5.0\sigma$  whereby the same Higgs decay channels but less data were analysed [13]. Exploiting to the full dataset led to a signal strength of  $\mu = 1.00 \pm 0.09(\text{stat}) \pm_{-0.07}^{+0.08}(\text{theo}) \pm 0.07(\text{sys})$  at  $m_H = 125.03 \pm_{-0.27}^{+0.26}(\text{stat}) \pm_{-0.15}^{+0.13}(\text{sys})\text{ GeV}$  [42], and thus the results are consistent with the SM prediction.

Nevertheless, it cannot be excluded that the observed state is not the one predicted by the SM but a boson suggested by *Beyond Standard Model* (BSM) theories. Hence, it needs to be characterised more precisely. Some properties are already defined, e.g. given the decay into  $W^\pm$ , its neutral electrical charge can be verified. Based on the observed decay into a pair of photons it can be excluded to be a spin 1 particle according to the *Landau-Yang-Theorem* [43, 44]. Two important physical quantities are the spin and parity of the Higgs which can be determined in decays like  $H \rightarrow \tau\tau$  [45, 46].

<sup>7</sup>abbreviated for  $H \rightarrow \tau^+\tau^-$



## 2.3 Physics Beyond the Standard Model

Although the Standard Model is a widely accepted theory, there are many open questions which cannot be addressed. Thus, a broad field of theoretical concepts beyond the SM has evolved. Until now no evidence to confirm any of these theories has been observed.

### 2.3.1 Limitations of the Standard Model

In the following, three prominent limitations of the SM closely related to the LHC physics program are briefly discussed. However, these are just a few among other open questions which need to be addressed by high energy physics, such as e.g. a quantised description of gravitation.

#### Hierarchy Problem

The *hierarchy problem* [20] describes the huge discrepancy of the strength of the gravitational and electroweak force which is of the order of 32 orders of magnitude. This then shows up in the theoretical treatment of the Higgs mass. If the SM is valid up to an energy scale in the order of the *Planck mass*, the Higgs mass would reach values of about  $10^{19}$  GeV caused by radiative corrections from higher order contribution in perturbative theory. The Higgs mass is given by the bare mass,  $m_{H,\text{bare}}$ , and additional contributions,  $\Delta m_H^2$ , from higher order calculations [20]:

$$m_H = m_{H,\text{bare}} + \Delta m_H^2. \quad (2.24)$$

The fermionic contribution to the quadratic term,  $\Delta m_{H,f}^2$ , is given by [20]:

$$\Delta m_{H,f}^2 = -\frac{|\lambda_f|^2}{8\pi^2} \left( \Lambda_{\text{UV}}^2 - 3m_f \ln \frac{\Lambda_{\text{UV}}^2}{m_f} + \dots \right), \quad (2.25)$$

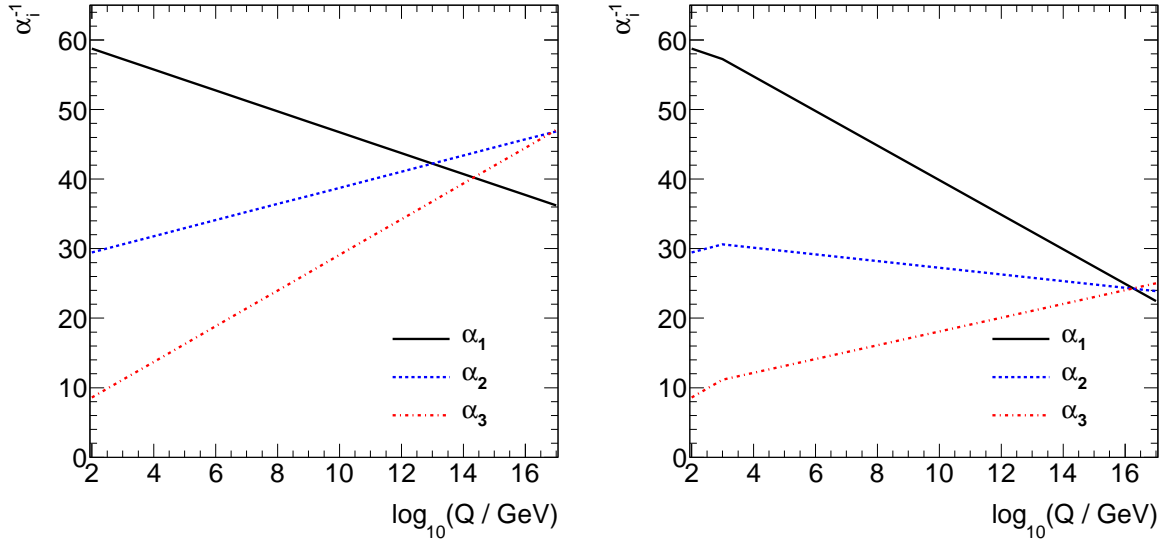
in which  $m_f$  denotes the fermion mass and  $\lambda_f$  parameterises the coupling of fermions to the Higgs boson. To regulate the ultraviolet-divergent behaviour of the loop integrals, a cut-off scale,  $\Lambda_{\text{UV}}^2$ , can be introduced, which defines the energy scale where new physics would occur. Given an observed Higgs mass of about 125 GeV a significant fine-tuning would be required, which apparently is possible but seems unnatural.

#### Dark Matter

Global fits of the *cosmic microwave background* (CMB) data, measured by WMAP [17] and Planck [18], reveal that the particle content of the SM only contributes to 5 % [17–19] of the total mass of the universe. However, cosmic effects like gravitational lensing or galactic rotation curves motivate the existence of *Dark Matter* (DM), whereby it only interacts weakly, and thus is difficult to detect. Dark Matter only can account for further 22 %, while the remaining 73 % are called *Dark Energy* (DE) [17–19]. The presence of Dark Energy is necessary to explain e.g. the accelerated expansion of the universe. For both, DM and DE, no candidate particles are found, yet.

#### Grand Unification

The unification of electromagnetic and weak interaction by the electroweak theory raises the question of the possibility to unify all forces. Within a *Grand Unified Theory* (GUT) only a unique interaction based on a simple gauge group exists. This requires the convergence of the coupling constants, which can be realised by the running of those constants defined by the RGEs [21]. In principle this is possible at high energy scales since the coupling constants of non-abelian gauge groups, i.e. strong and weak force, decrease with increasing energy while this behaviour is inverted for the electromagnetic force. However, Figure 2.3 demonstrates that the coupling strengths at the GUT scale still differ by more than eight standard deviations [21]. Hence, the SM cannot realise a unification of the fundamental forces.



**Figure 2.3:** Dependence of the gauge coupling constants,  $\alpha_i^{-1} = 4\pi/g_i^2$  ( $g_1 = \sqrt{5/3} = g_Y$ ,  $g_2 = g_w$  and  $g_3 = g_s$ ), on the energy scale  $Q$  in case of the SM (left) and MSSM (right). Due to contributions by superpartners unification is achieved at  $Q \sim 2 \cdot 10^{16}$  GeV, while in the SM they differ by about 8 standard deviations [21]. The calculations are based on [47] (p. 199 ff.) taking  $M_{\text{SUSY}} = 1$  TeV and  $\alpha_i^{-1}(m_Z)$  from [21]. Figures are taken from [48].

### 2.3.2 Theoretical Prospects

Several theoretical models aim to provide an explanation to the limitations of the SM. The most famous representative is *Supersymmetry* (SUSY) which offers itself various realisations. A short overview of SUSY will be given in the following.

#### Supersymmetry

SUSY describes a symmetry between bosons and fermions given by a supersymmetric transformation. Thus, a supersymmetric partner, a so-called *sparticle*, is introduced for each particle. Each fermion has a bosonic partner, *sfermion*, and each boson has a fermionic partner, *gaugino*. These superpartners can be arranged in supermultiplets containing fermionic and bosonic states which agree in electrical, weak and strong charge but differ by  $1/2$  in spin. The sfermions contribute equally to  $\Lambda_{\text{UV}}^2$  and  $\Delta m_H^2$  in Equation 2.25 like fermions but with different sign. Consequently, superpartners would cancel the divergent fermionic loop corrections and address the hierarchy problem. However, the correction occurs at an energy scale,  $M_{\text{SUSY}}$ , and thus a small fine-tuning remains (see below).

If SUSY is correct in its basic definition, particles and corresponding sparticles would have the same mass. Hence, the latter would have been already observed by various particle physics experiments. Since no evidence for SUSY was found so far it has to be a broken symmetry. This can be realised by *soft susy breaking* [20,21] which yields significant higher masses for the sparticles. The symmetry breaking occurs spontaneously at an energy scale in the order of 1 TeV and hides at low energies. Hence, the running of the coupling constants will change such that unification can be achieved, as shown in the right plot in Figure 2.3.

Furthermore, terms that violate the leptonic and baryonic number conservation are allowed within supersymmetric theories, whereby the decay of the proton within a short live time would be possible, which however is experimentally excluded. To prevent the proton decay  $R$ -parity can be introduced:

$$P_R = (-1)^{3(B-L)+S}, \quad (2.26)$$

in which  $S$  stands for the spin of the particle and  $L$  ( $B$ ) for the lepton (baryon) number. A  $R$ -parity of  $-1$  is

assigned to particles and +1 to sparticles. As a consequence of  $R$ -parity conservation sparticles can only be produced in pairs and their decay will always result in a final state with one remaining sparticle. The *lightest supersymmetric particle* (LSP) serves as a good candidate for DM since it is stable, electrical neutral and has a non-negligible mass.

### Minimal Supersymmetric Extension of the Standard Model

The simplest realisation to introduce SUSY is the *minimal supersymmetric extension of the standard model* (MSSM) [22–25] by which the smallest amount of new particles is added. Like the SM it contains three generations of sfermions and is based in a  $SU(3)_C \times SU(2)_L \times U(1)_Y$  gauge group. Consequently this leads to the same coupling structure, e.g. only left-handed sfermions and right-handed anti-sfermions can interact weakly. However, the Higgs sector is more complicated within the MSSM. Contrary to the SM it is not possible that the down- and up-type quarks couple to a single Higgs field and its complex conjugate, as this would cause gauge anomalies [49]. Thus, two scalar Higgs doublets are necessary which leads to five physical Higgs bosons: a light neutral Higgs,  $h^0$ , a heavy CP-even neutral Higgs,  $H^0$ , a heavy CP-odd neutral Higgs,  $A^0$ , and two charged Higgs,  $H^\pm$ . At tree-level the Higgs sector can be parametrised by two quantities being  $\tan\beta$ , the ratio of the two vacuum expectation values, and the mass of one of the Higgs bosons, commonly chosen to be  $m_A$ . However, tree-level calculations predict the mass of  $h^0$  to be smaller than the mass of the  $Z^0$  boson [49] which is excluded by the LEP experiments [50]. Hence, radiative corrections need to be taken into account which introduces further parameters. In doing so the mass can be extended up to 140 GeV [51] which is in agreement with the Higgs bosons observed at the LHC (cf. Section 2.2.2). On the other hand, the observed state gives stringent constraints on the MSSM parameter space.

### Heavy Gauge Bosons

Within various theoretical extensions of the SM like SUSY or GUT, heavy gauge bosons are predicted [52–55]. These arise from extended gauge sectors. The *Exceptional Supersymmetric Standard Model* [56] (ESSM) shall serve as an example. Starting from an  $E_6$  gauge group at the GUT scale it breaks down to  $SO(10)$  which further breaks down to  $SU(5) \times U(1)$ . The  $SU(5)$  gauge group then yields the SM and an additional  $U(1)$ . These two additional  $U(1)$  gauge groups can mix, and thus lead to a heavy  $Z'$  boson. However, the  $Z'$  has to be interpreted in a generic way. For instance, in the *Sequential Standard Model* (SSM) it behaves like a heavy partner of the Standard Model  $Z^0$  boson with same quantum numbers and coupling structure, while in other theoretical models it might be different. Hence, an additional unitary gauge group can be introduced such that the coupling structure can differ for the three generations of fermions. If the coupling to heavy leptons is preferred as in the *Topcolour Assisted Technicolour Models* [57, 58] (TC2) the tau final state becomes an essential probe for these type of models.

## 2.4 Physics with Tau Leptons

Given a mass of  $1776.82 \pm 0.16 \text{ MeV}$  [19] the tau lepton,  $\tau$ , is not only the heaviest charged lepton, but it is the only lepton which can decay both leptonically as well as hadronically. However, a direct detection of taus is very difficult because of their short mean life time of  $(290.3 \pm 0.5) \cdot 10^{-15} \text{ s}$  [19]. Although electrons and muons are in general easier to detect, it is more difficult in LHC experiments to distinguish between prompt leptons and those from leptonic tau decays. However, even for hadronic decays the discrimination against background is challenging which requires complex identification algorithms. A detailed description on the reconstruction and identification of tau leptons will be given in Chapter 5. Final states involving tau leptons are of high interest for studies to validate the SM at the TeV scale as well as for searches for new physics. The following sections will give a brief overview of physics analysis with tau leptons at the LHC.

### 2.4.1 Standard Model Processes

One of the aims of the LHC is to probe the SM at the TeV scale. Thus, SM processes involving tau leptons need to be considered, e.g. to validate the leptonic coupling of the third generation at these energies. Of special interest is the coupling to the weak gauge bosons,  $Z^0$  and  $W^\pm$ , which decay into taus with a *branching ratio* (BR) of  $11.38 \pm 0.21\%$  and  $3.370 \pm 0.008\%$  [19], respectively. In case of the decays  $W^+ \rightarrow \tau^+ \nu_\tau$  and  $W^- \rightarrow \tau^- \bar{\nu}_\tau$  an (anti-)neutrino is generated which cannot be detected directly, but can be reconstructed as *missing transverse energy*,  $E_T^{\text{miss}}$  (cf. Chapter 3). This is exploited by the trigger and reconstruction algorithms. The  $Z \rightarrow \tau\tau$  channel is perfectly qualified for cross section measurements. This process is further of high interest for various fields of tau physics e.g. as irreducible background for  $H \rightarrow \tau\tau$  searches or for tau identification performance measurements for which the semi-leptonic final state provides an outstanding probe. The latter will be discussed in detail in Chapter 5.

Furthermore, taus are important to study direct Higgs couplings to fermions in the di-tau final state because of the relatively high BR of 6.32% [59]. Its importance further arises from the fact that the dominating decay of  $H$  bosons into  $b\bar{b}$  pairs suffers from a high QCD multijet background. The combined measurements of the data taken with the ATLAS detector in the hadronic, full- and semi-leptonic tau decay channels show an excess at 125.36 GeV corresponding to 4.5 standard deviations at 95% C.L., whereby  $3.5\sigma$  were expected [41]. Data taken in 2011 and 2012 with an integrated luminosities of  $4.5 \text{ fb}^{-1}$  and  $20.3 \text{ fb}^{-1}$ , respectively, were analysed. A signal strength of  $\mu = 1.42^{+0.44}_{-0.38}$  [41] was observed which is summarised for the considered channel in Figure 2.4. Also the CMS experiment observed an excess around 125 GeV corresponding to  $3.2\sigma$  ( $3.7\sigma$  expected) whereas only an integrated luminosity of  $24.6 \text{ fb}^{-1}$  of combined 2011 and 2012 data were analysed [60]. These results are consistent with the Yukawa coupling predicted by the SM.

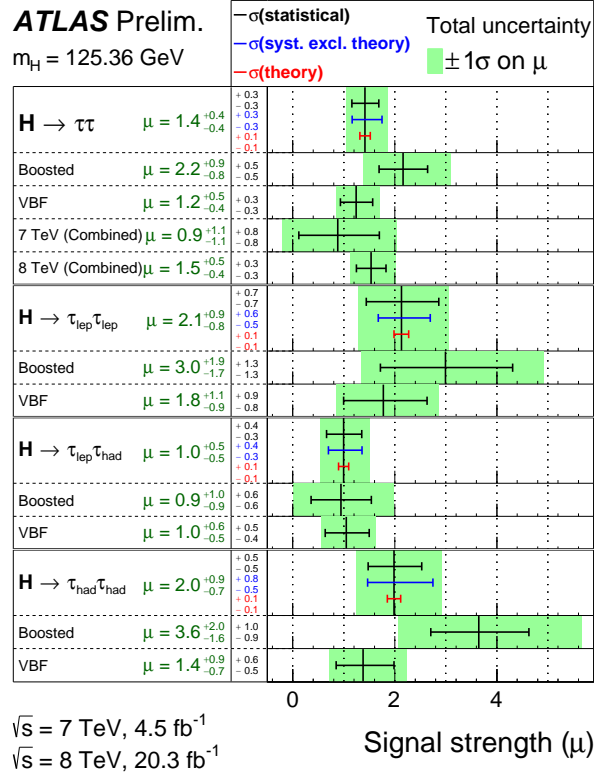
The fully hadronic  $H \rightarrow \tau\tau$  channel is not only important for the study of the expected coupling of the Higgs boson to fermions, but it can also provide further characterisations of the Higgs boson, e.g. spin and CP quantum numbers. The direct relation of the decay topology of the taus into pions to the spin of the Higgs boson can be exploited, e.g. by measuring the angle between the two decay planes. As a result it is possible to measure if the observed Higgs has a pure spin of 0 or is a mixture of spin states [45, 46].

### 2.4.2 Searches for New Physics

Besides confirming the SM, tau final states can give access to a wide field of new physics, e.g. searches for BSM Higgs bosons. Within the MSSM three neutral Higgs bosons are predicted with a BR for the decay into two taus of around 10% [61–64] on a wide mass range covering masses up to the TeV scale. Furthermore, the decay into taus is the dominating channel for the two charged Higgs bosons,  $H^\pm$ , if their mass is below the mass of the top-quark. The data analysed in 2012 by the ATLAS detector at  $\sqrt{s} = 8 \text{ TeV}$  corresponding to an integrated luminosity of  $19.5\text{--}20.3 \text{ fb}^{-1}$  for the hadronic and (semi-)leptonic decay, respectively, show no excess over the expected SM background [65]. However, exclusion limits on the production cross section of neutral MSSM Higgs boson are determined within the scope of various benchmark scenarios. The latter are a fixed parameter set defined to constrain the MSSM parameter space interesting for physics applicable at the LHC. For instance in the case of the  $m_h^{\text{max}}$  benchmark scenario [66] a heavy neutral Higgs boson with a mass of  $m_A < 160 \text{ GeV}$  is excluded at 95% C.L. [65] assuming that the lightest neutral MSSM Higgs boson is equal to the observed SM Higgs boson at 125 GeV.

Recent data analysis in  $t\bar{t}$  events by the ATLAS experiment exploiting the full dataset of 2012 have not shown any significant excess with respect to the standard model prediction. Hence, limits on the branching ratio  $BR(t \rightarrow bH^\pm) \times BR(H^\pm \rightarrow \tau^\pm \nu)$  have been set in the mass range between 80 and 160 GeV to 0.23–1.3% at 95% C.L. [67].

Also for the purpose of searches for high mass resonances, like  $Z'$  bosons, the tau lepton plays an important role. As mentioned in Section 2.3 lepton universality is not necessarily guaranteed, and thus a  $Z'$  could couple stronger to third generation fermions. Until now no significant excess above the SM



**Figure 2.4:** Signal strength,  $\mu$ , for the individual  $H \rightarrow \tau\tau$  decay channels for the full ATLAS dataset at centre-of-mass energy of 7 TeV and 8 TeV. For the combined analysis it was possible to measure  $\mu = 1.4$  for  $m_H = 125.36 \text{ GeV}$  [41].

background was observed in the hadronic di-tau final state. With the data taken with the ATLAS detector in 2012 at  $\sqrt{s} = 8 \text{ TeV}$  and a collected integrated luminosity of  $19.5 \text{ fb}^{-1}$  it was possible to exclude  $Z'$  bosons with a mass of  $M_{Z'} < 1.9 \text{ TeV}$  with 95 % credibility [68].



### 3 The ATLAS Experiment

The *Large Hadron Collider* (LHC) [14] is the currently largest particle physics experiment in the world, located at the *European Organisation for Nuclear Research* (CERN) near Geneva, Switzerland. The LHC accelerates and collides protons and ions at highest energies to test the SM at the TeV scale and search for new physics beyond the Standard Model.

After a ten-year-long construction phase, the first beam was launched in 2008. However, an incident caused by a magnetic connection failure postponed the first run to November 2009. Data-taking started with an injection energy of 450 GeV per proton beam at end of 2009. Until autumn 2012 the beam energy was increased stepwise to 4 TeV which corresponds to a centre-of-mass energy,  $\sqrt{s}$ , of 8 TeV. During the data-taking period in 2012 an integrated luminosity of  $22.8 \text{ fb}^{-1}$  [69] was collected. At the moment the LHC is in the *Long Shutdown 1* (LS1) phase and will restart operation in spring 2015. In the meantime, maintenance and upgrade services are performed with the purpose to reach the nominal beam energy of 7 TeV and at least the design luminosity of  $10^{34} \text{ cm}^{-2}\text{s}^{-1}$ . Besides accelerating protons, it is also possible to run with heavy ions, mainly lead nuclei, where energies up to 5.5 TeV per nucleon can be achieved.

There are four main experiments installed along the LHC where collisions take place. Two of them are multipurpose detectors, the ATLAS (*A Toroidal LHC ApparatuS*) [15] and CMS (*Compact Muon Solenoid*) [16] experiments which aim for high precision measurements of the SM and the search for new physics. The forward spectrometer LHCb (*Large Hadron Collider beauty*) [70] investigates CP violation via b-quark physics. ALICE (*A Large Ion Collider Experiment*) [71] is designed e.g. for the research of the quark-gluon plasma in lead-lead collisions. The main LHC experiments are accompanied by three smaller experiments, namely LHCf [72], MoEDAL [73] and TOTEM [74]. LHCf measures neutral particles in the very forward region of LHC collisions to calibrate the hadron interaction models used in extremely high energy cosmic ray research. The purpose of TOTEM is to determine the total proton-proton cross section and to study elastic scattering and diffractive dissociation. MoEDAL searches for magnetic monopoles and other highly ionising (pseudo-)stable massive particles.

#### 3.1 The Large Hadron Collider

Installed in the former LEP (*Large Electron-Positron*) tunnel, the LHC has a circumference of 26.5 km and crosses the Franco-Swiss border 100 m underground. Contrary to its predecessor LEP where electrons and positrons collided, the LHC operates with protons or heavy ions. The operation of a circular accelerator with hadrons instead of leptons has the advantage that the energy loss,  $\Delta E$ , due to synchrotron radiation is greatly reduced. For a single proton this is given per revolution by:

$$\Delta E = \frac{e^2 \beta^3 \gamma^4}{\epsilon_0 3R}, \quad (3.1)$$

in which  $e$  denotes the elementary charge,  $\epsilon_0$  the dielectric constant and  $R$  the radius of the ring. Furthermore,  $\beta$  defines the relativistic velocity and  $\gamma$  the relativistic factor which is inversely proportional to the mass of the accelerated particle. Consequently, the energy loss for protons is about eleven orders of magnitude smaller than for electrons, such that higher energies can be reached by a hadron collider.

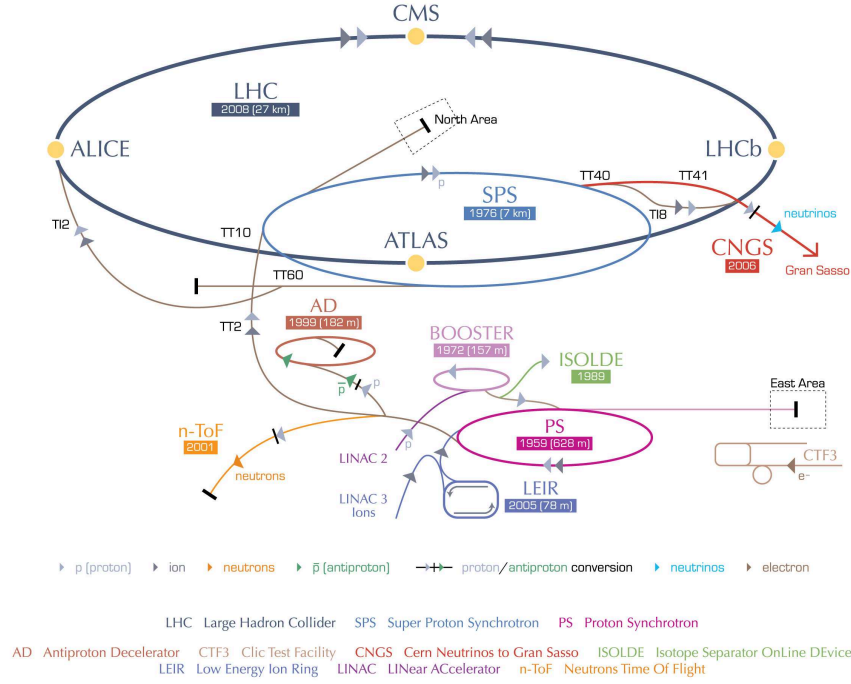
As two particle beams of the same charge are accelerated in opposite directions, a separate magnetic bending system is necessary to keep them on a circular trajectory. This is realised by space-saving twin-bore magnets which yields the disadvantage of a magnetic coupling of the two beam pipes. Nevertheless, the helium filled cryogenic system to cool down the superconducting magnets to a temperature of 1.7 K can be shared. The beamline has eight arcs where 1232 niobium-titanium magnets generate a magnetic field of 8.33 T to bend the proton beams. The eight straight accelerating sections are equipped by detectors and setups to measure the quality of the beam profile. Additionally, 500 quadrupol magnets are installed for focusing the beam.

Before the proton beams can enter the LHC, they need to run through the pre-accelerator system which is depicted in Figure 3.1. After the extraction by ionisation from a hydrogen source, the protons are accelerated to 50 MeV by the linear accelerator LINAC2. After that they enter the *Proton Synchrotron*



*Booster* (PSB) where they gain an energy of 1.4 GeV. Next, they are injected in the *Proton Synchrotron* (PS) which the protons leave with an energy of 25 GeV. Afterwards, the protons are fed in as packages in the *Super Proton Synchrotron* (SPS) where they are ramped up to 450 GeV. Finally they are split into two beams and enter the LHC via two transfer tunnels.

### CERN's accelerator complex



European Organization for Nuclear Research | Organisation européenne pour la recherche nucléaire

© CERN 2008

**Figure 3.1:** Schematic overview of the accelerator complex at CERN [75].

Considering a physics process with a cross section,  $\sigma_{process}$ , the mean number of events per second,  $N_{event}$ , is given by:

$$N_{event} = L \sigma_{process} , \quad (3.2)$$

with  $L$  denoting the luminosity which is defined for a Gaussian beam profile as:

$$L = \frac{N_b^2 n_b f_{rev} \gamma}{4\pi \epsilon \beta^*} F . \quad (3.3)$$

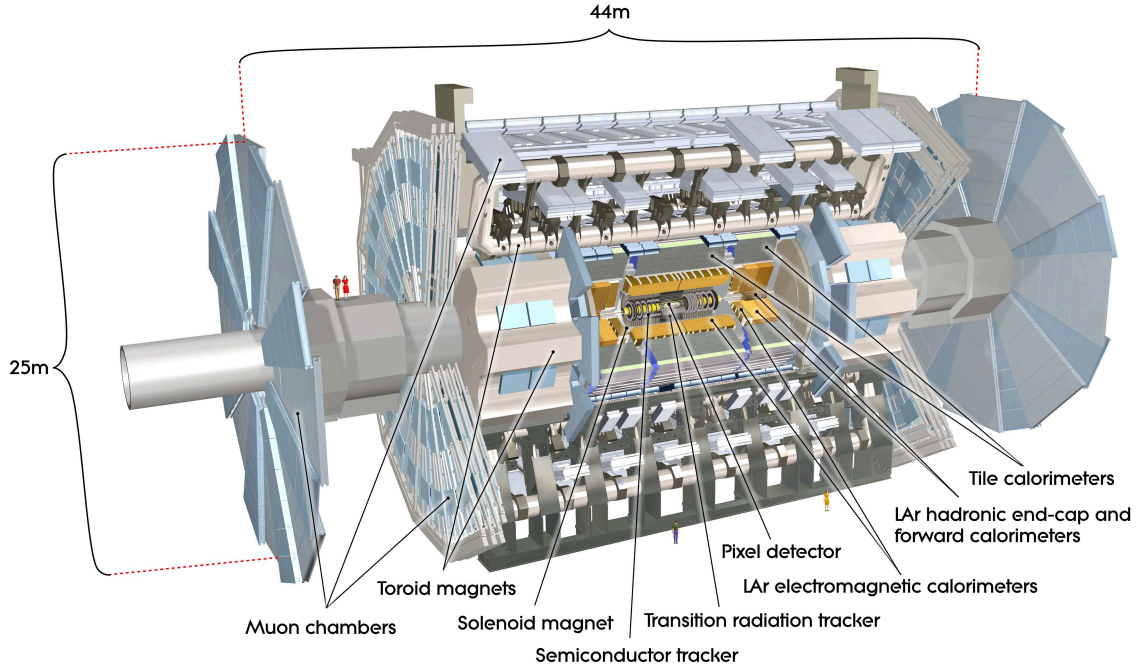
Where the number of particles per bunch,  $N_b$ , and the number of bunches per beam,  $n_b$ , are considered as well as the revolution frequency,  $f_{rev}$ . Furthermore, the luminosity depends on the beam parameters  $\epsilon$  and  $\beta^*$  which represent the normalised transverse beam emittance and the beta function at the collision point, respectively. Also the geometrical luminosity reduction factor, which is related to the crossing angle of the colliding beams at the interaction point, is taken into account by  $F$ . Considering a beam consisting of 2802 bunches with up to  $1.6 \cdot 10^{11}$  protons each and a bunch spacing of 50 ns, the design luminosity of  $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  can be reached. As a nominal beam energy of 7 TeV this corresponds to an average number of 23 proton-proton collision per bunch crossing. However, as mentioned above the design parameters could not be reached so far, and thus data-taking conditions of 2012 will be discussed in Section 4.2. The two high luminosity experiments, ATLAS and CMS, require such high luminosities to search for new physics beyond the Standard Model, e.g. SUSY. Furthermore, high precision and cross section measurements as well as studies of the SM at high energies are part of the LHC's physics program. A



first milestone has been achieved in 2012 when the Higgs boson was discovered by the ATLAS and CMS experiment [12, 13]. The following section will introduce the ATLAS experiments.

### 3.2 The ATLAS Experiment

The ATLAS detector is the largest experiment at the LHC with a length of 44 m and a diameter of 25 m. Its purpose is to study physics, both within and beyond the SM. ATLAS consists of three high precision sub-detector systems which are arranged concentrically around the beam pipe as can be seen in Figure 3.2. The innermost part of the detector aims for precise reconstruction of charged particles as well



**Figure 3.2:** Cut-away view of the ATLAS detector [15].

as primary and secondary decay vertices. It is followed by the calorimeter system to measure particle energies. Finally, the muon system was built to perform identification and measurement of momenta of muons. Calorimeter and muon spectrometer further serves for triggering purposes. Given the high bunch crossing rate of 40 MHz and limited datastorage, an outstanding trigger system is crucial. A description of each sub-system is given in the following sections after an introduction of the ATLAS coordinate system and important physical quantities.

#### 3.2.1 Nomenclature

The ATLAS coordinate system is right handed and has its origin in the nominal interaction point which is equal to the centre of the detector. The  $z$  axis points in beam direction, hence the  $x$ - $y$  plane is transverse to the beamline. Thereby, the positive  $x$  axis points towards the centre of the LHC and the positive  $y$  axis upwards. Along the  $z$  axis the detector is divided into an A-side (positive  $z$ ) and a C-side (negative  $z$ ).

Based on the concentric design of the detector around the beamline, many measured quantities are expressed in cylindrical coordinates, with the azimuthal angle,  $\phi$ , in  $x$ - $y$  plane around the beam axis and the polar angle,  $\theta$ , with respect to the beam axis. An important quantity in hadron collider physics is the rapidity,  $y$ , because of the invariance of particle production as a function of  $y$  under longitudinal boosts. The rapidity is defined as:

$$y = \frac{1}{2} \cdot \ln \frac{E + p_z}{E - p_z}, \quad (3.4)$$

where  $E$  denotes the particle energy and  $p_z$  its momentum component parallel to the beamline. For massless particles it reduces to the pseudorapidity,  $\eta$ , which depends on the polar angle through:

$$\eta = -\ln \tan \frac{\theta}{2}. \quad (3.5)$$

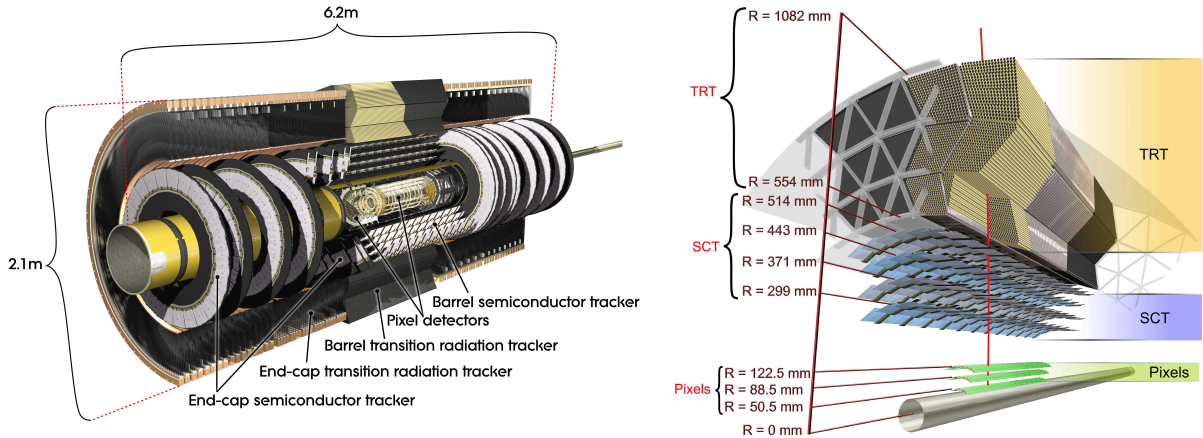
Further the spacial distance,  $\Delta R$ , of two objects in the  $\eta$ - $\phi$  plane is given by:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \quad (3.6)$$

Particular attention is given to transverse quantities since the longitudinal momentum of the initial partons cannot be determined, but the vector sum of momenta in the transverse plane is zero at vanishing beam crossing angles. Highly important for physics analyses are the transverse momentum,  $p_T$ , the transverse energy,  $E_T$ , and the missing transverse energy,  $E_T^{\text{miss}}$ , which are defined in the  $x$ - $y$  plane.

### 3.2.2 Tracking System

The detector part closest to the beampipe is the tracking system. It extends to 7 m in length and 2.3 m in diameter and covers a range of  $|\eta| < 2.5$ . The inner detector is composed of three sub-detectors and surrounded by a 2 T solenoid magnet as depicted in Figure 3.3. The main purpose of the inner detector is



**Figure 3.3:** Cut-away view of the inner detector of ATLAS (left) and sketch of its structure traversed by a charged particle (right) [15].

high precision tracking of charged objects, which further allows the extrapolation of secondary vertices. For instance in the case of a tau lepton decaying into charged particles like pions, the decay vertex and the tracks of the decay products can be reconstructed. Furthermore, an accurate measurement of impact parameters can be used for tagging of jets containing B hadrons. The longitudinal,  $z_0$ , and transverse,  $d_0$ , impact parameter resolution for high  $p_T$  tracks ( $p_T > 30 \text{ GeV}$ ) has been evaluated by cosmic ray data taken in 2008 to [76]:

$$\sigma(d_0) = (22.1 \pm 0.9) \mu\text{m} \quad \text{and} \quad \sigma(z_0) = (122 \pm 4) \mu\text{m}. \quad (3.7)$$

Additionally, a transverse momentum resolution of [76]

$$\frac{\Delta p}{p} = (4.83 \pm 0.16) \cdot 10^{-4} \frac{p_T}{\text{GeV}} \quad (3.8)$$

has been determined for a single charged particle with  $p_T > 30 \text{ GeV}$ . Those resolutions are valid for the barrel region as cosmic muons only pass from the top to the bottom of the detector. Besides the barrel region, two end-caps are part of the inner detector.

### Pixel Detector

As innermost part of ATLAS the pixel detector has to resist very high radiation doses and is therefore built with hybrid silicon pixel sensors. It is composed of three barrel layers and three end-cap disks, with an radial extension of 5.05-12.25 cm and 49.5-65 cm in  $z$  direction, respectively. Thus a range of  $\eta < 2.5$  is covered. The pixel sensors which combine a junction diode depleted by a bias voltage as active sensor and a read-out chip. 1744 sensors each with 47232 pixels lead to 80.4 million channels to be read out. With an intrinsic resolution of  $10(10)\mu\text{m}$  in  $r$ - $\phi$  plane and  $115(105)\mu\text{m}$  in  $z$  direction in the barrel (end-cap) region the pixel detector provides the highest precision of all detector parts. The matrix layout allows a highly granular measurement in both coordinates in the layer plane. Hence, an excellent impact parameter resolution and vertex reconstruction can be achieved which supports the classification of particles. Approximately 1000 charged particles propagate through the detector per bunch crossing, whereby one particle hits on average three sensors. If a charge particle hits a pixel, electron-hole pairs emerge in the pixel sensor via ionisation along the track. The electrons drift to a cathode where they are collected by tiny bump bonds and are converted to a measurable current.

### Semi Conductor Tracker

The *Semi Conductor Tracker* (SCT) exploits the same physical concept like the pixel detector whereas the 6.3 million silicon sensors are arranged in 4088 microstrips. Each module is built of four active sensors, whereby two of them are wire bound, such that a two dimensional track measurement can be realised. The SCT is divided in eight barrel layers, covering radii of 30-52 cm and a range of  $|\eta| < 1.4$ , and nine disks in forward region providing coverage to  $|\eta| < 2.5$ . Although its granularity is coarser, a precise measurement of vertex positions, momenta and impact parameters of charged particles can be guaranteed. The intrinsic resolution in the transverse plane,  $r$ - $\phi$ , is determined to  $17\mu\text{m}$  and  $580\mu\text{m}$  in  $z$  direction.

### Transition Radiation Tracker

The outermost part of the inner detector is the *Transition Radiation Tracker* (TRT) which is the least precise one. It has a length of 5.3 m and a diameter of 2.5 m. Strawtubes of 4 mm in diameter, filled with a gas mixture of 70 % xenon, 27 % carbon oxide and 3 % oxygen, are deployed. In the barrel region they are surrounded by polypropylene fibres and by foils in the end-caps. Each of the about 50000 straws has a  $30\mu\text{m}$  thin gold-plated tungsten-rhenium wire in its centre, which is set under high voltage. The TRT has in total 420000 read-out channels and an intrinsic resolution of  $130\mu\text{m}$  per strawtube. Though, measurements are only possible for  $\eta < 2.0$ .

Once a charged particle enters a strawtube, the gas volume gets ionised and the freed electrons drift with a constant velocity to the cathode. Due to the high electric field, avalanching occurs close to the cathode which can be measured as electric signal. Since the drift time is well known, the location where the particle entered the tube can be calculated. Furthermore, a particle has to pass through various boundaries with different refraction indices. In the case of an ultra-relativistic particle with  $\gamma > 1000$ , transition radiation is emitted in a small angle with respect to the particle direction, whereby this effect only occurs on layer boundaries. Because of the high atomic number of xenon, the photons will be absorbed efficiently and contribute to the measured current. Consequently a distinction of light electrons with  $\gamma > 1000$  and heavier charged pions with  $\gamma < 1000$  with same momentum, is possible which can be exploited in tau identification.

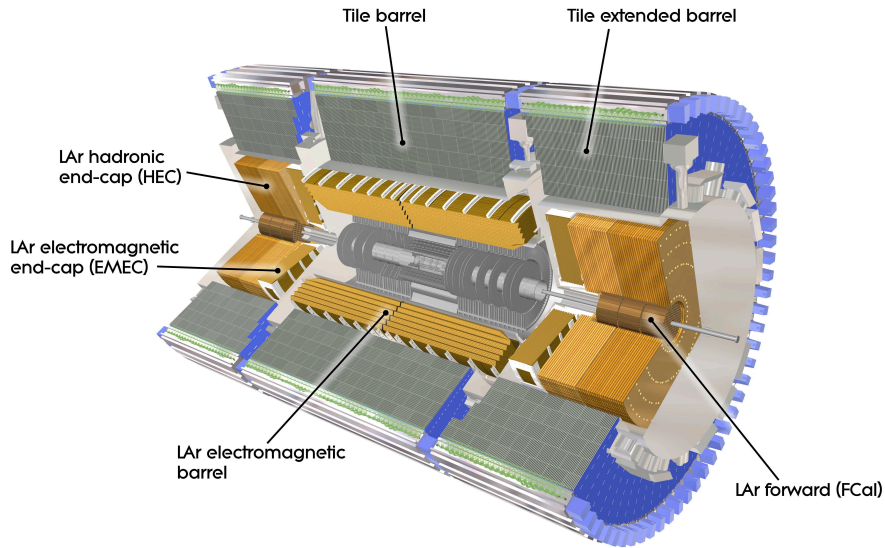
### Solenoid Magnet

To bend the traversing charged particles on curved trajectories a superconducting solenoid encloses the inner detector. Aluminium-cooper-niobium-titanium wires are rolled up to a 2.46 m long and 5.3 m diameter magnet. To reach a magnetic field of 2 T with a current of 7.73 kA it is necessary to cool the entire setup. The cryogenic system can be shared with the liquid argon calorimeter, thus the amount

of material in front of the calorimeter is minimised which leads to a reduced deterioration of energy measurement.

### 3.2.3 Calorimeter System

Within the calorimeter system all particles are stopped by absorbing their energy, except of heavy muons and the weakly interacting neutrinos which pass through the entire detector without being stopped. The calorimeter is separated into two parts: The electromagnetic calorimeter which is dedicated for electrons and photons, and the hadronic calorimeter focusing on hadronically interacting particles. Likewise, the leptonic and hadronic decay products of tau leptons are stopped in the electromagnetic and hadronic calorimeter, respectively. A layout of the calorimeter system is shown in Figure 3.4. Thanks to its high hermiticity the missing energy can be reconstructed precisely as negative vectorial sum of all energy deposits.



**Figure 3.4:** Sketch of the ATLAS calorimeter system [15].

### Electromagnetic Calorimeter

The *electromagnetic calorimeter* (ECAL) is based on a sampling technology with liquid argon as active material and lead as absorber. To allow a full coverage in  $\phi$  without any cracks an accordion shaped design was chosen. The *electromagnetic barrel* (EMB) has a gap of 6 mm at  $z = 0$  and comprises a region of  $|\eta| < 1.52$ . An overlap with the two *electromagnetic end-caps* (EMECs) guarantees a smooth transition. Each EMEC consists of two wheels covering a pseudorapidity range of 1.375-2.5 (inner wheel) and 2.5-3.2 (outer wheel). After an electron or photon has entered the calorimeter their energy is reduced by alternating bremsstrahlung and pair production which leads to an electromagnetic particle shower. Energy deposits of charged particles ionise the liquid argon and electrons are collected by copper electrodes. This yields a signal proportional to the deposited energy.

The sensitive region lies between  $\eta$  of  $-2.5$  and  $2.5$  where three layers have a thickness of at least 24 radiation lengths,  $X_0$ . Most of the energy of electrons and positrons is deposited in the second layer. The first layer is the one with the highest granularity with segments corresponding to  $\Delta\eta = 0.0031$  and  $\Delta\phi = 0.1$ . This, for instance, allows the separation of photons originating from a neutral pion decay, which is of high interest for tau decay mode classification.

Because of the inner detector and its read-out electronic the electromagnetic showering already starts before the ECAL is reached. However, to reconstruct the entire shower energy a pre-sampler has been

installed in front of the calorimeter. It consists of one liquid argon calorimeter layer of 1.1(0.5) cm thickness in the barrel (end-cap) region and extends up to  $|\eta| < 1.8$ . Although the granularity is much coarser for larger pseudorapidities, a sufficient energy resolution can be provided. The latter can be parametrised by a noise term  $a$ , a sampling term  $b$  and a constant term  $c$ , given by:

$$\frac{\sigma_E}{E} = \frac{a}{E} \oplus \frac{b}{\sqrt{E}} \oplus c. \quad (3.9)$$

### Hadronic Calorimeter

Behind the ECAL the *hadronic calorimeter* (HCAL) is equipped. Its purpose is to determine the energy of hadronically interacting particles, like charged pions from tau decays. Therefore, it is composed of three sub-systems. The *tile barrel* is a sampling calorimeter with steel absorbers and plastic scintillators as active material. It covers a range of  $|\eta| < 1.7$  and extends up to 9.7 hadronic interaction lengths in the transverse central region. Thus, a high energy resolution is gained and a punch through to the muon system is significantly suppressed. If a particle traverses the detector a hadronic shower is formed which involves an electromagnetic component. The produced secondary particles then stimulate the scintillating material. The emitted relaxation light is collected by *photo multiplier tubes* (PMTs) via wavelength shifters to increase the efficiency. The measured signal is proportional to the deposited energy. A pseudorapidity range of  $|\eta| = \pm(1.5 - 3.2)$  is covered by the *hadronic end-cap*, and thus it overlaps with the tile barrel. To cope with the higher radiation densities liquid argon and copper are arranged in a sandwich structure. The third device, the *forward calorimeter* (FCAL), is the closest to the beam pipe, extending over a range of  $3.1 < |\eta| < 4.9$ . It is composed of one copper layer to measure electromagnetic interactions, and two tungsten layers for hadronically interacting particles. Longitudinal channels with grounded tubes and high voltage-fed rods are distributed over each layer. The 0.25-0.5 mm wide gaps are filled with liquid argon, thus the same physical principle as by the ECAL is exploited.

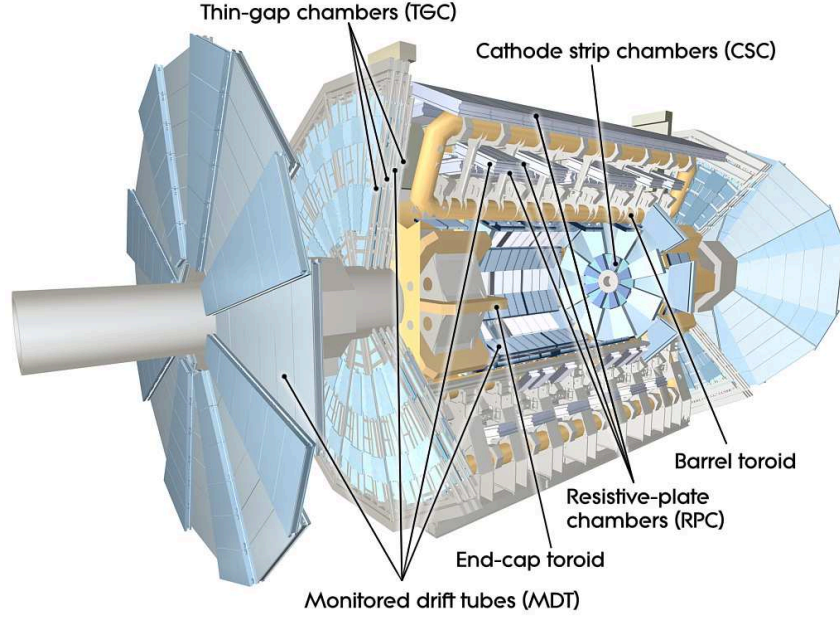
#### 3.2.4 Muon System

The outermost part of the ATLAS detector is equipped by the *muon spectrometer* (MS) which concentrates on the measurement of muons. Due to the fact, that muons are minimal ionising at LHC energies, they traverse the detector without only little interactions. Only muons with an energy of less than 5 GeV are stopped before the MS. Nevertheless, to determine the muon momentum with high precision a specialised detector is necessary. Therefore, a toroid magnet system composed of eight superconducting coils which generates an almost circular field is installed radial symmetrically around the beam pipe. The muons are bent on curved trajectories whereby the bending power of the 0.5 T barrel field ( $|\eta| < 1.4$ ) is 1.5-5.5 Tm and 1-7.5 Tm in the end-caps ( $1.6 < |\eta| < 2.4$ ). The latter is provided by a magnetic field of 1 T. In the intermediate region both fields overlap. A sketch of the MS and its three sub-detectors is depicted in Figure 3.5.

The middle and the outer layer of the *Monitored Drift Tube Chambers* (MDTs) extend over a full range of  $|\eta| < 2.7$ , while the inner layer covers only pseudorapidities smaller than 2.0. The MDT consists of cylindrical drift tubes, filled with a gas mixture of argon and carbon dioxide. A tungsten-rhenium alloyed aluminium wire in the centre of each tube collects the electrons freed by ionisation of the gas volume by traversing muons. Within a region of  $2.0 < |\eta| < 2.7$  *Cathode Strip Chambers* (CSCs) are applied to provide a better time resolution. Multi-wire proportional chambers with tungsten-rhenium anode wires are filled with a gas mixture as well, but the composition differs from the MDTs, i.e. 50 % carbon dioxide, 30 % argon and 20 % carbon tetra-fluoride is used.

Both detectors are too slow to be used by the trigger. Only the *Restrictive Plate Chambers* (RPCs) and the *Thin Gap Chambers* (TGCs) have a sufficient intrinsic time resolution of 1.5 ns and 4 ns, respectively. A high electric field is applied in the 2 mm gap between the two parallel arranged restrictive plates of the RPC, such that a muon which enters the detector ionises the volume and causes avalanching. The collected signal can be exploited to veto fake muons, e.g. muons from cosmic rays, in the barrel region



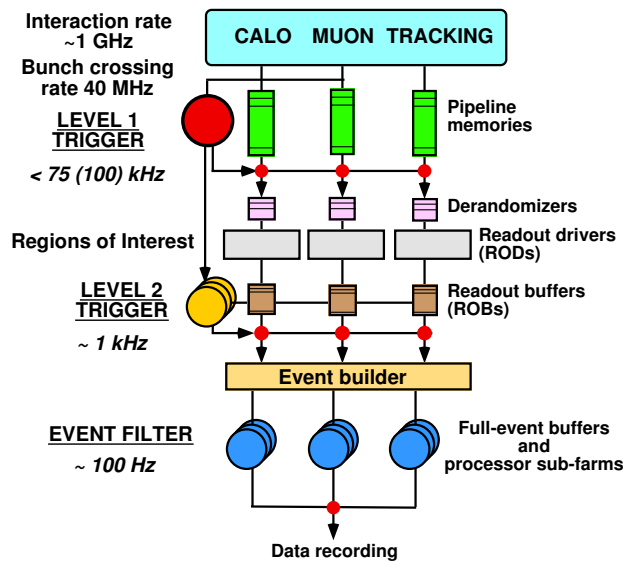


**Figure 3.5:** Sketch of the muon system of the ATLAS detector [15].

by a coincidence measurement. The end-caps are equipped by TGCs which operate similar to the multi-wire proportional chamber but with a radial and azimuthal segmentation of the wires to provide a good trigger signal.

### 3.2.5 Trigger System

Given an event rate of 40 MHz and a typical event size of 1.3 MB a data amount of 50 TB per second is generated by the ATLAS detector. Hence, an efficient trigger system is required to reduce it to a storable size. Therefore, a three level trigger system is set up at ATLAS. It uses both, hardware and software algorithms to reject background events while preserving interesting events for physics analyses. Figure 3.6 summarises the outline of the ATLAS trigger system.

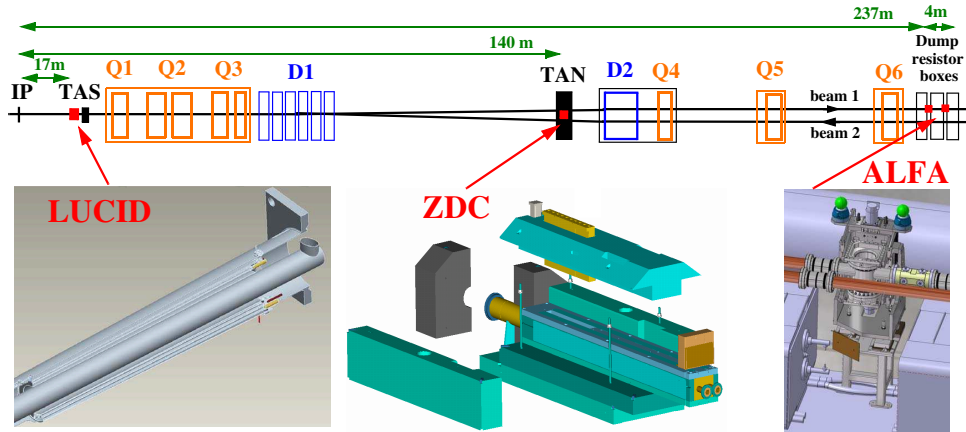


**Figure 3.6:** Overview of the ATLAS trigger system [77].

The *Level 1 Trigger* (L1) is fully hardware based and shrinks the rate to about 75 kHz with a latency of  $2.5 \mu\text{s}$ . It exploits the calorimeter information as well as the signals given by the RPCs and TGCs of the muon spectrometer, whereby single objects with large transverse momenta, like hadronically decaying tau leptons, muons, electrons, photons and jets are taken into account. The decision is based on a minimal isolation of energy deposits and various thresholds, e.g. on the transverse energy of a particle. Furthermore, the negative sum of the transverse energy can be used to trigger on  $E_T^{\text{miss}}$ . Finally, a *Region of Interest* (ROI), which includes information about  $\eta$  and  $\phi$  of the accepted object, is selected by a pass-fail decision. Afterwards the data is written to the *Read-Out Driver* (ROD) and the *Read-Out Buffer* (ROB). The data rate is further downsized by the *Level 2 Trigger* (L2), exploiting the full detector granularity in the ROIs to apply additional cuts. With a latency of 40 ms this stage of the *High Level Trigger* (HLT) leads to a rate of about 3.5 kHz. Afterwards, the *Event Filter* provides a reduction to about 200 Hz. Within 4 s per event a full event reconstruction is performed where the used algorithms are close to those applied offline.

### 3.2.6 Forward Detectors

Additional four devices to measure luminosity and beam conditions are installed in forward direction of the ATLAS detector. In Figure 3.7 the locations of three of them is shown. At  $z$  and  $\eta$  locations of



**Figure 3.7:** Alignment of ATLAS forward detectors along the beamline [15].

$\pm 184$  cm and  $\pm 4.2$ , respectively, the *Beam Condition Monitors* (BCMs) are the closest to the ATLAS detector. It allows a bunch-per-bunch luminosity measurement to react on occurring beam anomalies in order to prevent the ATLAS detector from a potential damage. LUCID (*Luminosity measurement using Cherenkov Integrating Detector*) is installed at  $z = \pm 17$  m and  $|\eta| \sim 5.8$ , providing a relative luminosity measurement by inelastic proton-proton scattering. As ATLAS main luminosity detector it is used for online monitoring. The absolute luminosity is determined at  $z = \pm 240$  m by ALFA (*Absolute Luminosity for ATLAS*) which is needed to calibrate LUCID. Scintillating fibre trackers inside of roman pots measure elastic proton-proton scattering. The resulting amplitude is linked with the total cross section and absolute luminosity via the optical theorem. In between the two luminosity detectors the *Zero Degree Calorimeter* (ZDC) is located at  $z = \pm 140$  m and  $|\eta| > 8.3$ , which analyses the centrality of collisions observed by neutral particles generated within the collisions.





## 4 Simulation and Data

The results presented within this thesis are based on data collected by the ATLAS detector in 2012 as well as *Monte Carlo* (MC) simulation. Therefore, an overview of MC event generation and detector simulation is given in Section 4.1. The data-taking conditions in 2012 for the ATLAS detector are summarised in Section 4.2.

### 4.1 Monte Carlo Simulation

The simulation of physics processes is a crucial method for particle physics analyses. On the one hand it serves for testing and optimisation of analysis setups, on the other hand it provides the possibility to study the discovery potential for new physics and estimation of background processes.

As protons are collided at the LHC a precise simulation of hadronic interactions is required. This involves two main components, the description of the hard process, like the production of a  $Z$  boson decaying into a tau-anti-tau pair, and the parton shower. Starting from a dynamic description given by the SM Lagrangian, the cross section of the hard scattering process can be calculated by the underlying theories, i.e. *perturbative QCD* (pQCD) and electroweak theory. At high energies and short distances the actual interactions take place between the partons of the proton which can be assumed as free particles. Their momentum distribution is described by *parton density functions* (PDFs) which have to be measured experimentally as they cannot be predicted by theory, though their evolution with the momentum transfer  $q$  can be modelled by the *DGLAP equation* [78–80]. The cross section of the hadronic interaction can then be calculated from the differential cross section  $\hat{\sigma}_{ij}$  for scattering of two partons  $i$  and  $j$  as:

$$\sigma = \sum_{i,j} \iint dx_1 dx_2 (f_i(x_1, \mu_F^2) f_j(x_2, \mu_F^2)) \hat{\sigma}_{ij}, \quad (4.1)$$

whereby  $x_1$  and  $x_2$  denote the momentum fraction of the involved partons and  $\mu_F$  the factorisation scale. Additionally, soft processes summarised as *initial* (ISR) and *final state radiation* (FSR) need to be taken into account. These arise from gluon radiation or QED bremsstrahlung, i.e. photons. The gluons will produce quark-anti-quark pairs and eventually hadrons due to confinement (cf. 2) while the photon will convert most likely into electron-positron pairs. Thus, a parton shower emerges. In the detector these will be observed as a so-called *hadronic jet*. The measured jet energy approximates the energy of the initial parton. Various jet clustering algorithms exist in order to associate single hadrons to the jet. The assignment is either based on a geometrical cone matching [81] or sequential clustering [82–85] which performs the matching depending on the relative difference in  $p_T$ . By now only the interaction of two partons per bunch crossing were discussed. Actually, multiple interactions taking place at each bunch crossing need to be taken into account. In addition many proton-proton collisions take place in the same bunch crossing at high luminosities. This is called *pile-up*.

All steps are considered by Monte Carlo event generators which aim to simulate the interaction system. The calculation of the hard scattering process is based on the Lagrangian and *Feynman rules* of the underlying theory. Hence, the *Matrix Element* (ME) can be calculated on tree level without any loop correction. Although this provides an effective way to determine the hard interaction at high energies, it cannot serve for the calculation of the soft part. This needs to be addressed by the so-called *Parton Shower* (PS). The PS models the evolution from the energy scale of the hard interaction to the scale where hadronisation enters. Finally, complementary concepts of the ME providing the description of separated jets at fixed order and the soft and collinear contribution at all orders describing the structure inside of the jet given by the PS, have to be merged. For this purpose several matching algorithms, which avoid possible double counting, are available. Furthermore pile-up contributions need to be simulated. There are two types of pile-up: *In-time pile-up* which occurs during a single bunch crossing as overlay of the hard process with additional inelastic proton-proton collisions and *out-of-time pile-up* describing the overlap of detector signals of sequential bunch crossings. Various MC generators, like Pythia [86] and Sherpa [87], are available. The datasets used in this thesis were generated by Pythia8 [88]. A special

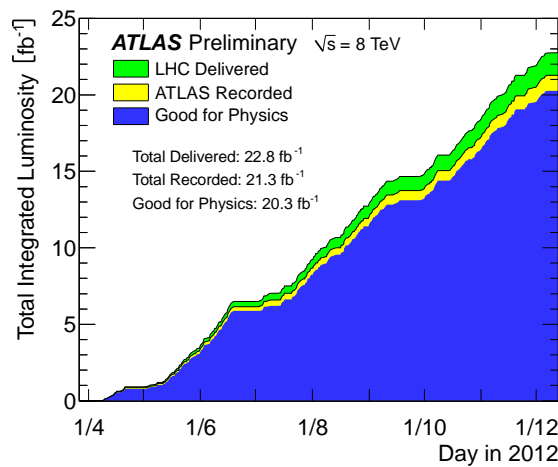
treatment is applied for tau decays in order to e.g. model intermediate resonances in the decay. Pythia8 can do this by itself while other generators feature a dedicated tool called TAUOLA [89]. It is specialised on the simulation of hadronic and leptonic tau decays considering neutrinos, intermediate resonances as well as spin correlations.

The final step in the simulation chain is the application of the detector simulation, resulting in an output which is similar to actual data. MC generators provide a four-vector for each particle for the underlying process. Then the detector simulation is added using a dedicated framework called ATHENA [90] which was developed by the ATLAS collaboration. It simulates the trajectory of a particle traversing the detector and its interaction with the detector material. The description of hadronic as well as electromagnetic showers and displaced vertices of long lived particles is considered, too, resulting in a complete event description. The *full simulation* is based on the toolkit GEANT4 [91] which exploits the full detector description to simulate the interaction of particles and matter. Therefore, run conditions, the detector model and alignment are stored in a database. Essentially three steps are performed. First, the hits for each sub-detector including energy deposits and merging of hits due to pile-up, are calculated. Afterwards, they are digitised such that one digit corresponds to an energy deposition exceeding a predefined voltage or current threshold in a certain time window. This leads to *raw data objects* (RDOs) which contain the information of the data event as well as of the generator which is called *truth information*. Finally the digits are used to reconstruct final state particles. Here, the same trigger and reconstruction algorithms as used for data processing are applied. This is a very complex procedure and demands a huge amount of CPU hours, e.g. the full simulation chain takes about 15 minutes per event, which is far too long for processes with high production cross sections.

For the analysis represented in this thesis mainly  $Z \rightarrow \tau\tau$  MC generated events were used, but also simulated  $Z' \rightarrow \tau\tau$  samples with resonance masses between 250 and 1250 GeV are considered to enhance the available statistic for high- $p_T$  tau candidates.

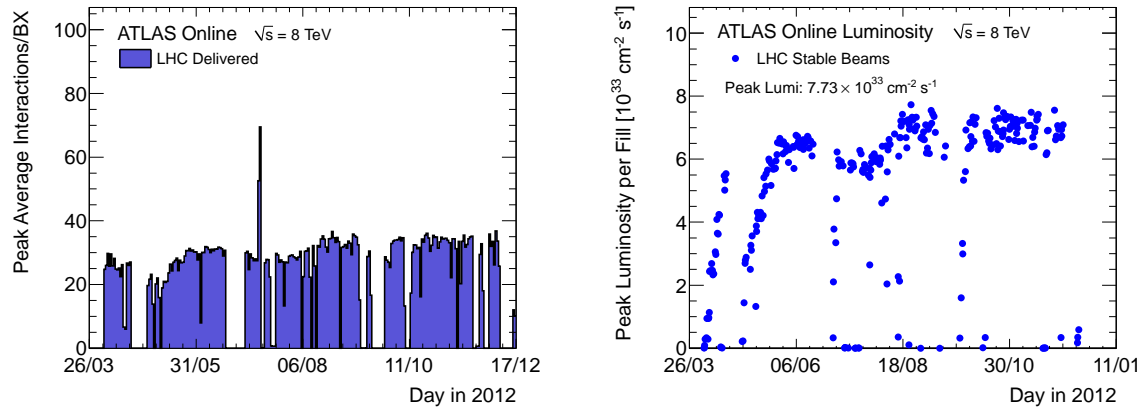
## 4.2 Data-taking in 2012

In 2012 most of the runs at the LHC were performed with 1374 proton bunches per beam and a bunch spacing of 50 ns at a centre-of-mass energy of  $\sqrt{s} = 8$  TeV [92]. An integrated luminosity of  $22.8 \text{ fb}^{-1}$  [69] was delivered. Given a data-taking efficiency of 93 % for the ATLAS detector this corresponds to  $21.3 \text{ fb}^{-1}$  of data taken. The actual amount of data usable for physics analysis accounts to  $20.3 \text{ fb}^{-1}$  as depicted in Figure 4.1. Furthermore, the total integrated luminosity provided by the LHC and collected by the ATLAS detector is shown. On average, almost 40 interactions took place per bunch crossing which



**Figure 4.1:** Total integrated luminosity delivered by the LHC (green), collected by the ATLAS detector (yellow) and utilisable for physics analysis (blue) based on the entire data-taking period in 2012 [69].

is already the from the design parameters expected amount. Finally, the obtained peak luminosity was  $7.73 \cdot 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ . Figure 4.2 shows the evolution of the peak luminosity per fill and the peak average number of interactions per bunch crossing,  $\mu_{\text{max}}$ , over the entire data-taking period in 2012. An overview of all data-taking periods and respective run conditions of proton-proton collision data collection in 2012 is given in Table 4.1. Within the scope of this thesis only the periods A, B and I were used since the needed information of the tau decay mode classification is not provided for all periods. Within run-II, starting in 2015, the number of bunches will be increased to 2808 and the bunch spacing will be halved to 25 ns. Hence, the design luminosity of  $10^{34} \text{ cm}^{-2}\text{s}^{-1}$  [15] will be reached and even higher values will be accessible.



**Figure 4.2:** Evolution of the peak average number of interactions per bunch crossing (left) and of the peak luminosity per fill over the entire data-taking period in 2012 at the ATLAS detector [69].

period	start	end	run-number range	recorded lumi [ $\text{pb}^{-1}$ ]	efficiency [%]	$\mu_{\text{max}}$
A	Apr-04	Apr-20	200804–201556	910	98	30
B	May-01	Jun-18	202660–205113	5594	98	31
C	Jul-01	Jul-24	206248–207397	1643	98	34
D	Jul-24	Aug-23	207447–209025	3598	98	34
E	Aug-23	Sep-17	209074–210308	2863	98	36
G	Sep-26	Oct-08	211522–212272	1404	98	34
H	Oct-13	Oct-26	212619–213359	1655	98	35
I	Oct-26	Nov-02	213431–213819	1149	98	34
J	Nov-02	Nov-26	213900–215091	2941	98	35
L	Nov-30	Dec-06	215414–215643	983	98	36
M	Dec-15	Dec-16	216399–216432	14	97	12

**Table 4.1:** Overview of proton-proton collision data-taking periods of the ATLAS detector in 2012 [93].



## 5 Tau Reconstruction and Identification

An efficient reconstruction and identification of tau leptons is highly important for many physics analyses at the LHC, like the verification of the SM at the TeV scale, e.g. via  $Z \rightarrow \tau\tau$  decays [94], the measurement of SM-like  $H \rightarrow \tau\tau$  decays or the search for MSSM Higgs bosons in the di-tau final state (cf. 2.4). Given the high mass of 1.77 GeV [19] taus can decay both leptonically as well as hadronically. In 35.2 % [19] of the time tau leptons decay in lighter leptons, i.e. electron or muon, and two corresponding neutrinos. The remaining 64.8 % are of hadronic nature, with decays mainly into charged and neutral pions,  $\pi^\pm$  and  $\pi^0$ , accompanied by a tau neutrino. For the reconstruction and identification only hadronic decays are considered, since the decay products of leptonic decays are almost impossible to distinguish from prompt electrons and muons. In 72 % of the cases taus decay into a single charged pion, while in 23 % they decay into three charged pions [19]. According to this, hadronic tau decays are usually categorised into *1-prong* and *3-prong* decays. The remaining decay modes with higher multiplicities are not taken into account because of the low branching ratios and high background contribution. The various hadronic and leptonic decay modes and their BRs are summarised in Table 5.1. Depending on the hadronic decay mode, between zero and three neutral pions accompany the decay, while higher numbers of  $\pi^0$  are unlikely.

	decay mode	branching ratio $\Gamma_i/\Gamma$
leptonic decays	$\tau^\pm \rightarrow e^\pm + \nu_e + \nu_\tau$	17.83 %
	$\tau^\pm \rightarrow \mu^\pm + \nu_\mu + \nu_\tau$	17.41 %
hadronic decays	$\tau^\pm \rightarrow \pi^\pm + \nu_\tau$	10.83 %
	$\tau^\pm \rightarrow \pi^\pm + \pi^0 + \nu_\tau$	25.52 %
	$\tau^\pm \rightarrow \pi^\pm + 2\pi^0 + \nu_\tau$	9.30 %
	$\tau^\pm \rightarrow \pi^\pm + \pi^\pm + \pi^\mp + \nu_\tau$	8.99 %
	$\tau^\pm \rightarrow \pi^\pm + \pi^\pm + \pi^\mp + \pi^0 + \nu_\tau$	2.70 %

**Table 5.1:** Decay modes and branching ratios of leptonic and hadronic tau decays [19].

A crucial property of 1- and 3-prong tau decays is the typical shape of the showers emerging in the calorimeter system. Further, specific fractions of energy deposits in the electromagnetic and hadronic calorimeter characterise the tau decay, and hence can be exploited in the reconstruction chain. Both features can be mimicked by QCD jets and electrons, and thus they can pass the reconstruction and identification as so-called *fake taus*. To maximise the fake rejection as well as the tau acceptance, a chain of trigger, reconstruction and identification algorithms based on multivariate techniques is exploited. The two advanced multivariate approaches utilise *Boosted Decision Trees* (BDTs) and *projective log-Likelihoods* (LLHs) [95] which are essential for the identification of taus and will be described in detail in Section 5.3. A BDT is also chosen to differentiate taus from electrons which show narrower shower profiles and lower activity in the hadronic calorimeter while a higher transition radiation is measured. For the discrimination against jets a BDT as well as a LLH are used. In some cases also muons deposit energy in the calorimeter, and thus can mimic tau decays. A cut-based approach is chosen to reject them.

The next section will give an overview of the tau reconstruction with the ATLAS detector followed by a detailed discussion on the specific decay substructure in Section 5.2. In Section 5.3 the focus will be laid on the tau identification including a detailed description of the multivariate analysis techniques used.

### 5.1 Reconstruction of Tau Leptons

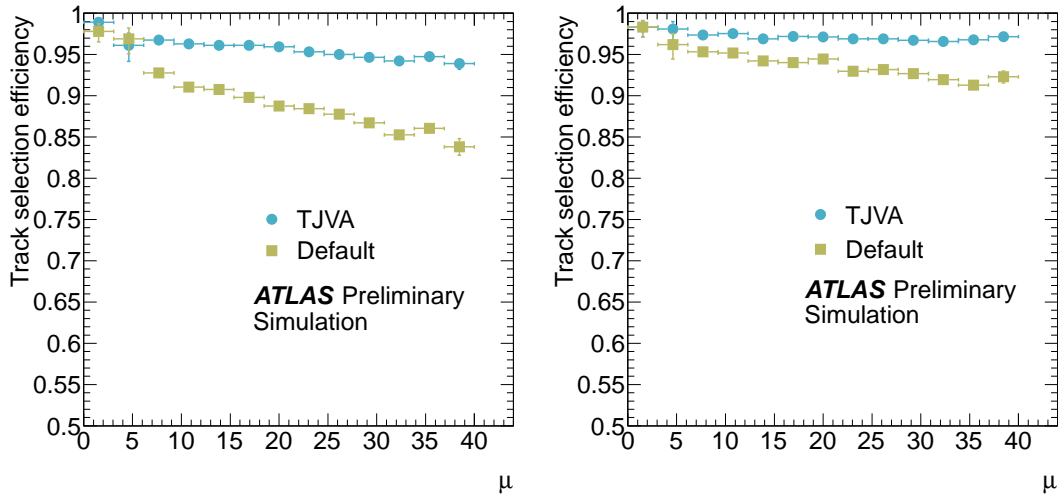
The aim of the tau reconstruction is to build tau candidates based on the information given by the tracking and calorimeter system [96]. Since the signature of a hadronically decaying tau is similar to the signature

of jets and lighter leptons it is crucial to combine several sub-detector signals to get a sufficient performance. Hadronic tau candidates,  $\tau_{\text{had}}$ , are seeded from jets reconstructed by the anti- $k_t$  algorithm [97] with a distance parameter  $R = 0.4$ . The addressed calorimeter cells are grouped to topological clusters [98] and can be associated to a tau candidate. The *Local Hadron Calibration* (LC) [99] is applied, i.e. weights depending on the reconstructed shower profile are assigned to the energy of each topo cluster to take non-compensation of the calorimeter, dead material and energy deposits outside of the cluster into account.

The vertex candidate with the highest squared sum of track  $p_T$ 's is taken as the primary vertex [100]. The probability for misidentification with pile-up vertices increases with the average number of interactions per bunch crossing. Thus, a dedicated algorithm, called TJVA (*Tau Jet Vertex Association*) [101], is used as a more pile-up robust approach to define the tau vertex. It is based on the *Jet Vertex Association* [102] algorithm which assigns to each jet the vertex candidate with the highest jet vertex fraction,  $f_{\text{JVF}}$ . For a given jet and vertex candidate it is defined as:

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

where  $p_T^{\text{trk}|\text{vtx}}$  is the transverse momentum of a track assigned to a given vertex and  $p_T^{\text{trk}}$  denotes this quantity for all tracks within the jet cone. Further track quality criteria specialised on tau decays are required. In Figure 5.1 the track selection efficiency against the average number of interactions per bunch crossing,  $\mu$ , is depicted for the default track selection algorithm and TJVA. The comparison shows that the pile-up robustness is significantly increased by applying TJVA.



**Figure 5.1:** Track selection efficiency against the average number of interactions per bunch crossing,  $\mu$ , for reconstructed 1-prong (left) and 3-prong (right) tau decays in MC generated  $Z \rightarrow \tau\tau$  events. The green boxes represent the performance of the default track selection and the blue circles show the efficiency which can be obtained by applying TJVA [101].

Based on the topo cluster measurement in the calorimeter and the vertex assignment the intermediate axis and direction of the tau candidate can be determined. Therefore, the topological cluster of the constituents within  $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.2$  around their barycenter are taken into account. The direction, i.e.  $\eta$  and  $\phi$ , of the tau candidate is calculated from the sum of the four-momenta of the jet constituents. The mass of the jet constituents, and thus the tau candidate is defined to be zero [103], such that its energy and momenta are identically in the transverse plane,  $p_T = E_T = E \sin \theta$ . The coordinate system of the tau vertex is adjusted to calculate the four-vectors of the clusters whereby the sum results in the intermediate tau axis. Given this axis the direction of the tau candidate is set. Afterwards tracks within the *core cone*, i.e. a region of  $\Delta R \leq 0.2$  around the intermediate tau axis, are associated to the tau candidate, on the condition they fulfil the following quality criteria:

- transverse momentum,  $p_T \geq 1 \text{ GeV}$ ,
- number of pixel hits,  $N_{\text{pixel}} \geq 2$ ,
- number of pixel and SCT hits,  $N_{\text{pixel+SCT}} \geq 7$ ,
- distance of closest approach to tau vertex in transverse plane,  $|d_0| \leq 1.0 \text{ mm}$ ,
- distance of closest approach to tau vertex in longitudinal direction,  $|z_0 \sin \theta| \leq 1.5 \text{ mm}$ .

According to the number of tracks within the core cone a prongness of one or three is assigned to each tau candidate. Nevertheless, tracks inside of the *isolation annulus* are considered for variable calculations to increase the suppression of fake taus in the identification algorithms. The isolation annulus surrounds the core cone and is defined in a region of  $0.2 < \Delta R \leq 0.4$  to the intermediate tau axis.

The calorimeter response defined by the jet energy scale is not able to reflect the characteristics of the tau decay products, hence a dedicated *tau energy scale* (TES) [104] is defined. Therefore, it is essential to determine the number of  $\pi^0$ s and their deposited energy. A dedicated algorithm, called *neutral pion finder* is applied based on the measurements of the calorimeter clusters [105]. Multivariate methods, namely BDTs, are used to exploit the information provided through five variables which are sensitive to the typical signature of neutral pion decays. The first BDT determines if any neutral pion occurred, while the second BDT does the decay mode classification. Thereby only the main decay modes, i.e.  $\tau^\pm \rightarrow \pi^\pm + \pi^0$ ,  $\tau^\pm \rightarrow \pi^\pm + 2\pi^0$  and  $\tau^\pm \rightarrow 2\pi^\pm + \pi^\mp + \pi^0$ , are considered. Additionally, several corrections have to be applied, to account for noise, underlying event and pile-up. The latter is estimated from the isolation annulus. Further a “pi0 likeness score” is defined as the energy fraction of the presampler and the strip layer w.r.t. the charged pions, and thus indicates the probability that the related clusters are caused by a neutral pion. If a neutral pion was found by the BDT the four-vector is calculated from the assigned calorimeter clusters and serves as a base for the calculation of variables used in tau identification (cf. Section 5.3). Further, a correction of the hadronic energy contamination is applied to avoid double counting, and hence improve the energy resolution for neutral pions.

With the TES a dedicated calibration on the tau energy is applied which introduces corrections on the tau momentum and pseudorapidity as well as pile-up corrections. The tau momentum resolution w.r.t. the electromagnetic energy scale is already improved by the local hadron calibration. Nevertheless, discrepancies caused by the energy loss in front of the calorimeter, underlying event, pile-up and out-of-cone effects are not considered. Hence, the tau energy calibration is used to account for those effects. It is derived from reconstructed hadronic tau candidates fulfilling a medium BDT-based jet discrimination criteria taken from MC generated  $W \rightarrow \tau\nu_\tau$ ,  $Z/\gamma^* \rightarrow \tau\tau$  and  $Z' \rightarrow \tau\tau$  events, and is applicable to tau candidates with  $p_T = 15 \text{ GeV}$ . The calibrated momentum,  $p^{\text{cal}}$ , of the tau candidate is given by:

$$p^{\text{cal}} = \frac{p^{\text{LC}}}{R(p^{\text{LC}}, |\eta_{\text{reco}}|, n_{\text{prong}})}, \quad (5.2)$$

where  $R$  is the calibration term which depends on the momentum at the LC scale,  $p^{\text{LC}}$ , the reconstructed pseudorapidity,  $\eta_{\text{reco}}$ , and the number of assigned tracks,  $n_{\text{prong}}$ . The calibration constants are defined by the ratio of the momentum at the LC scale,  $p^{\text{LC}}$ , and the generated visible momentum,  $p_{\text{vis}}^{\text{gen}}$ , for 1-prong and 3-prong decays. The response is fitted by a Gaussian function in various bins of  $p_{\text{vis}}^{\text{gen}}$  and  $|\eta_{\text{reco}}|$ . The momentum dependency of the calibration constants is obtained from the empiric fit function of the mean of the Gaussian fit. Further it has to be taken into account that the poor instrumentation in some  $|\eta|$  regions of the calorimeter still lead to deviations of more than 2 %. Hence, the reconstructed energy of the affected clusters is underestimated, such that it is necessary to apply a pseudorapidity correction which results in a corrected pseudorapidity,  $|\eta_{\text{corr}}|$ , given as:

$$|\eta_{\text{corr}}| = |\eta_{\text{reco}}| - \eta_{\text{bias}}. \quad (5.3)$$

Within this  $\eta_{\text{bias}}$  denotes the average deviation of the reconstructed pseudorapidity,  $|\eta_{\text{reco}}|$  and the generated pseudorapidity,  $|\eta_{\text{gen}}|$ . However, a pile-up dependence is still observed which raises the necessity



of a correction accounting for the effect of pile-up events. This results in the following impact on the momentum of the tau candidate:

$$p_{\text{pile-up}} = A(|\eta_{\text{reco}}|, n_{\text{prong}})(N_{\text{PV}} - \langle N_{\text{PV}} \rangle), \quad (5.4)$$

whereby  $A$  is a constant derived from a linear fit in  $|\eta_{\text{reco}}|$  and  $n_{\text{prong}}$ . Further  $N_{\text{PV}}$  and  $\langle N_{\text{PV}} \rangle$  stand for the number of reconstructed and average number of primary vertices, respectively. Finally, the momentum of the tau candidate, considering all calibrations and corrections, is determined as:

$$p^{\text{final}} = \frac{p^{\text{LC}} - p_{\text{pile-up}}}{R(p^{\text{LC}}, |\eta_{\text{reco}}|, n_{\text{prong}})}. \quad (5.5)$$

Nevertheless, the various effects, like the calorimeter response, choice of the MC generator which includes the chosen underlying event model and detector description, as well as the non-closure of the calibration method, which are considered in the TES, result in non-negligible uncertainties, which differ significantly in  $\eta$  and  $p_{\text{T}}$ . Given this, the tau axis and direction as well as all basic tau variables are recalculated and prepared for the identification step.

## 5.2 Decay Mode Classification of Tau Leptons

Recent developments have been pushed towards a dedicated reconstruction of charged and neutral pions and kaons to reveal the substructure of hadronic tau decay.

Charged pions and kaons of high momentum have both a decay length of a few meters [19], and thus they can pass the tracking detector and reach the calorimeter system. In both detector components they leave characteristic signatures. In the tracker a signal evaluated to a curved trajectory depending on the charge and kinematic energy is measured. The energy deposits in the calorimeter are dominated by event level fluctuations which overcome the  $p_{\text{T}}$  and  $\eta$  dependence. Usually charged pions and kaons from tau decays pass several calorimeter layers before the first hadronic interaction takes place and a hadronic shower emerges. The number and four-momenta of pions and kaons with  $p_{\text{T}} \lesssim 150 \text{ GeV}$  are determined from the measurements of the tracker, as it leads to a more precise result than being extracted from the calorimeter signal. The momentum and direction is fixed by a fit on the tracks reconstructed in the tau core cone. The energy of the kaons is typically underestimated by construction due to the choice of the pion mass hypothesis. Yet, the resulting discrepancy is small compared to a calorimeter based calculation, and thus the track based determination still leads to a gain in performance. Likewise, the bias for the reconstructed tau energy is negligibly small with less than 2 %. Nevertheless, a significant fraction of charged tau decay products fail to be reconstructed in the tracking system caused by hadronic interactions with the material of this detector part, which is the leading cause of inefficiency in the track reconstruction.

Neutral pions show a smaller decay length than the charged pions and decay almost exclusively in two photons before they reach the calorimeter. There, electromagnetic showers are induced which are less sensitive to fluctuations on event level. Most of the showers start in the presampler or first layer of the ECAL and deposit a significant fraction of energy. In the second layer the photons from the pion decay already lost at least half of the energy of the initial pion, such that only a small signal is measured in the hadronic calorimeter. Since photons are electrically neutral they are invisible for the inner detector. However, some of the photons convert in an electron-positron-pair in the tracker or beam pipe, and thus leave measurable tracks. Those conversion tracks are only rarely associated to tau decays due to the strict quality criteria on the reconstructed tracks. Typically, photons as well as the conversion electrons and positrons are very close to each other due to the high boost of the initial pions. Hence, very narrow shower profiles evolve and only one topo cluster is assigned to the neutral pions. Almost the same applies for neutral kaons while they have a far longer decay length of some meters depending on the CP eigenstate. In about 50 % of the decays into kaons a  $K_{\text{L}}^0$  occurs which passes the tracker without any interaction, but initiate a hadronic shower similar to those of pions. Whereas the remaining decays are into  $K_{\text{S}}^0$ , their decay length is shorter than the decay length of  $K_{\text{L}}^0$  but larger than for  $\pi^0$ . However, if



their transverse momentum is less than 20 GeV they decay in 69.2 % [19] of the cases into a  $\pi^+\pi^-$ -pair within the tracking system. Consequently, tracks appear in the outer layers of the tracker and a hadronic shower arises in the calorimeter. The remaining 30.7 % [19] of  $K_S^0$  decays are into a pair of neutral pions which turn into two photon pairs and result in an electromagnetic shower.

Typically, the energy deposits of the charged and neutral decay products of tau decays are very close to each other, such that the calorimeter clusters overlap. Hence, the discrimination of them is not trivial and relies on specific energy fractions and shower shapes which are inherent for those hadrons. Several dedicated approaches to cope with this challenge were developed, which especially aim for a good reconstruction of neutral pions. The two algorithms which were used within the scope of this thesis, the CellBased and EflowRec algorithm, are introduced in the following sections. Further detailed discussion of the substructure reconstruction of hadronic tau decays can be found in [106].

### 5.2.1 CellBased Algorithm

The CellBased algorithm [106] is dedicated to reconstruct and identify neutral and charged pions in hadronic tau decays. It runs after the intermediate axis is determined and tracks are associated to the tau candidate by the standard tau reconstruction. Only tau candidates with less than six core tracks are taken into account due to CPU and storage consumption issues. Tau candidates with more or no core tracks will not be considered further. They only serve for reconstruction studies.

The applied procedure is based on detailed studies of hadronic tau decays and their signature in the detector. Neutral pions, i.e. the two photons from the  $\pi^0$  decay, deposit their energy mainly in the electromagnetic calorimeter, while the charged pions leave only a small fraction of their total energy there and deposit a large part of their energy in the hadronic calorimeter. Given that tau decays are highly collimated, also their decay products, i.e. the neutral and charged pions, are very close together, and thus their energy deposits in the calorimeter often overlap and the charged fraction needs to be subtracted to get access to the neutral part. Hence, a good estimation of the energy deposition of charged pions is crucial. Therefore, the track momentum measurement from the inner detector and the assigned energy deposition in the HCAL are used to estimate the total energy deposited by each charged pion in the ECAL. An important method to determine the energy deposition in each cell are averaged shower shapes, which are based on MC simulated single  $\pi^\pm$  samples. A weight is applied on each cell and enters the calculation of the energy which need to be subtracted. Within the CellBased algorithm the subtraction of the charged fraction is performed cell-by-cell to achieve the best performance. Afterwards the remaining electromagnetic cells are re-clustered based on the standard topo clustering. The re-clustering operates on unweighted cells of the presampler and the first two layers of the ECAL. After that, variables sensitive to cluster shapes are calculated and exploited in a preliminary BDT based identification of the neutral pion clusters. Finally, the decay mode of the tau candidate is determined by simply counting the amount of charged and neutral pions for each tau candidate. Whereby it is assumed that each core track originates from only one charged pion and each cluster in the ECAL from one neutral pion. The four-momentum of the tau candidate is given by the sum of the four-momenta of the assigned charged and neutral pions.

Five decay modes are defined based on the amount of neutral and charged decay products. The standard notation for these decay modes follows the structure  $XpYn$  which refers to  $X$  charged pions and  $Y$  neutral pions. Table 5.2 summarises the reconstruction efficiencies of the CellBased algorithm for a single decay mode as well as for the correct number of charged decay product. Further the relative amount of decays reconstructed with too few or too many neutral decay products is presented, whereby those values are determined without regarding the correct reconstruction of prongness. The reconstruction efficiency for the each individual decay mode suffers mainly from migrations to decay modes with lower or higher numbers of neutral decay product. In the case of too few neutral decay products the calorimeter cells of them merge with those of the charged or other neutral decay products. Too many neutrals occur if the reconstruction misses a  $\pi^\pm$  and its energy deposition in the calorimeter is assigned to a neutral pion or kaon. Further a very high energy deposition of a single neutral particle can be mistakenly interpreted as originating from two or even more particles. In addition, all decay modes are affected

by contributions from pile-up and underlying event. As can be seen, the reconstruction efficiency of the exact prongness is almost 100 %. The decay modes with zero neutral particles shows with more than 80 % also a quite good performance, whereas the modes 1pXn and 3pXn migrate to 46.4 % and 39.3 % to lower number of neutrals or a incorrect prongness.

The reconstruction efficiencies of individual decay modes are not only important for the optimisation of the individual substructure algorithm, but further serve to compare the different substructure algorithms. The CellBased algorithm shows one of the best performance on 8 TeV data in terms of the decay mode classification as well as the resolution in  $E_T$ ,  $\eta$  and  $\phi$  in the 1p1n decay mode.

algorithm	reconstruction category	decay mode				
		1p0n	1p1n	1pXn	3p0n	3pXn
CellBased	correct reconstructed	87.1	62.9	53.6	84.7	60.7
	correct reconstructed prongness	98.9	95.8	93.9	93.7	93.2
	too few reconstructed neutrals	—	15.2	40.8	—	36.9
	too many reconstructed neutrals	12.1	21.6	—	11.6	—
EflowRec	correct reconstructed	77.8	56.2	55.8	59.8	72.8
	correct reconstructed prongness	98.9	95.8	94.0	93.7	92.8
	too few reconstructed neutrals	—	13.8	38.3	—	30.0
	too many reconstructed neutrals	21.5	29.8	—	37.8	—
CellBased+PanTau	correct reconstructed	85.5	74.9	40.8	89.0	62.0
	correct reconstructed prongness	98.9	86.6	94.0	93.7	93.2
	too few reconstructed neutrals	—	12.4	53.5	—	36.3
	too many reconstructed neutrals	13.7	12.5	—	8.0	—
EflowRec+PanTau	correct reconstructed	84.2	68.5	38.8	84.0	58.3
	correct reconstructed prongness	99.0	95.8	93.8	93.7	93.1
	too few reconstructed neutrals	—	15.2	55.3	—	39.1
	too many reconstructed neutrals	15.1	16.1	—	13.7	—

**Table 5.2:** Efficiencies of correct reconstruction of the decay mode and the prongness as well as the efficiencies for too few and too many reconstructed neutral decay product for the CellBased, EflowRec, CellBased+PanTau and EflowRec+PanTau algorithm. The values are based on MC generated  $Z \rightarrow \tau\tau$  events [106].

### 5.2.2 EflowRec Algorithm

Another approach to access the substructure of hadronic tau decays is provided by the EflowRec algorithm [106]. It is a general purpose particle flow algorithm which exploits the information of the tracking and calorimeter system to reconstruct individual charged and neutral pions in hadronic tau decays. As in the CellBased algorithm the energy deposition of charged pions is subtracted to get the neutral part of the decay, although here the subtraction is based on so-called *energy flow objects* (eflow objects).

Within the EflowRec algorithm the cells of a topo cluster are ordered in rings according to the decreasing energy density around the extrapolated track to which the cluster is assigned. Subsequently, cells are removed from the rings until the expected energy of the charged pion is reached. Whereby the subtraction starts in the hadronic calorimeter and continuous in the electromagnetic calorimeter after the expected energy is updated. The expected energy of charged pions is extracted from MC generated single  $\pi^\pm$  samples. If several tracks are assigned to a set of close-by clusters, the tracks are ordered in  $p_T$ , such

that the cells from the shower associated to the highest- $p_T$  track are subtracted first. The resulting output eflow object can be charged or neutral. To classify overlapping or close-by clusters as being charged or neutral eflow objects, a set of shower shape variables is exploited by a BDT which is trained on a MC generated  $Z \rightarrow \tau\tau$  events. Afterwards the estimated  $\pi^0$  energy is calculated for each neutral eflow object. Based on the BDT output the final decay mode is determined by counting the neutral and charged objects like in the CellBased algorithm.

Table 5.2 summarises the reconstruction efficiency of the EflowRec algorithm for a single decay mode as well as for the correct number of charged decay products. As can be seen the overall performance of the EflowRec algorithm is slightly worse than achieved by the CellBased algorithm. For instance for the 1p0n decay mode the CellBased algorithm yields a 10 % higher purity.

### 5.2.3 PanTau algorithm

The PanTau algorithm [106] runs on top of each substructure algorithm to improve the decay mode classification, especially the determined number of neutral pions. Therefore, the kinematics of a hadronic tau decay is further exploited as well as correlations between the decay products. The base algorithms, e.g. CellBased and EflowRec, provide a set of reconstructed charged and neutral pions for each tau candidate on which the determination of the decay mode relies. A BDT approach is chosen which exploits variables sensitive to the tau decay kinematics and the correlations between the charged and neutral pions. Thus, the decay mode classification is improved significantly because it is less affected by the mis-tagging of  $\pi^0$ s. Two kinds of mis-tagging can occur: too many or too few  $\pi^0$ s are reconstructed by the base algorithm. In the first case additional clusters caused by charged remnants, pile-up, underlying event or, unlikely, a splitting of the two photons are reconstructed as neutral pions. On the other hand, neutral pions are missed if a reconstructed  $\pi^0$  fails the  $\pi^0$ -BDT identification or a  $p_T$  threshold, respectively. Another reason can be that two very close  $\pi^0$ s merge and are reconstructed as one  $\pi^0$ .

In particular three different BDTs are used whereby each of them tests whether one of two decay modes is more likely. The considered decay mode pairs are 1p0n vs. 1p1n, 1p1n vs. 1pXn and 3p0n vs. 3pXn where for each only the relevant reconstructed sub-samples including additional neutral hadron not tagged as  $\pi^0$  (denoted by +0/X), are taken, i.e. 1p0n+X and 1p1n+0, 1p1n+X and 1p1n+0/X, 3p0n+X and 3pXn+0/X, respectively. The variable set used differs for the CellBased and EflowRec algorithm, but for both it is based on angular distances and  $E_T$  ratios of the decay products and  $\pi^0$ -BDT information. The BDT is optimised individually for each base algorithm and each sub-sample. Thus, the fraction of correctly reconstructed tau decay modes is maximised. The results for the two considered algorithms are shown in Table 5.2. It can be seen that the migration between the various decay modes are minimised. In particular the reconstruction efficiency for the 1p1n decay mode is improved significantly for both base algorithms by about 12 %. Though, the reconstruction efficiency for the 1pXn decay mode decreases due to migrations to the 1p1n decay mode caused by the additional requirements. However, these are in general less important for physics analyses.

## 5.3 Identification of Tau Leptons

Within the tau reconstruction almost all jets arising from QCD multi-jet production are accepted, such that a high performance identification algorithms are required to achieve an efficient rejection of fake taus, while providing a high acceptance for real taus. The two successfully operating multivariate techniques are a boosted decision tree and a log-Likelihood which further exploit the properties of hadronic tau decays, e.g. narrower shower profiles. A continuous optimisation and adjustment of both algorithms is necessary to push the performance, but also coping with changing circumstances like increasing pile-up. Therefore, MC generated events with tau final states as signal and a di-jet data selection enriched with fake taus as background are used.

### 5.3.1 Tau Identification Variables

The tau identification variables are already calculated during the reconstruction step and are tuned to provide an outstanding separation power of signal and background, i.e. real taus and fake taus, respectively. Jets raise in general wider showers than taus for a given momentum which can be extracted from the measurement of the tracking and calorimeter system. Also the amount of reconstructed decay products differs for jets and taus, and hence is used for a further discrimination. Events accepted for tau identification studies have to pass a preselection to guarantee clean signatures. The following quality criteria are applied:

- Good Runs List is applied as a pre-selection to ensure stable detector and trigger conditions and a high quality of the considered proton-proton collision data (data only)
- at least one reconstructed PV with a minimum of four assigned tracks
- at least one reconstructed tau candidate with  $p_T > 15 \text{ GeV}$  and  $|\eta| < 2.3$
- $p_{T,\text{vis}}^{\text{gen}} > 10 \text{ GeV}$ ,  $|\eta_{T,\text{vis}}^{\text{gen}}| < 2.5$  (signal only)

Calorimeter based variables are not defined outside of the core cone ( $\Delta R < 0.2$  w.r.t. the intermediate tau axis) due to their strong pile-up sensitivity caused by additional energy deposits very close to the tau candidate. Whereas track-based variables are less affected by pile-up events, and thus are partly extended to the isolation annulus ( $0.2 < \Delta R < 0.4$  w.r.t. the intermediate tau axis). The following list summarises the set of variables for 1-prong and 3-prong tau decays for the current tau identification [96, 101]:

**Track radius,  $R_{\text{track}}$ :**  $p_T$ -weighted track width:

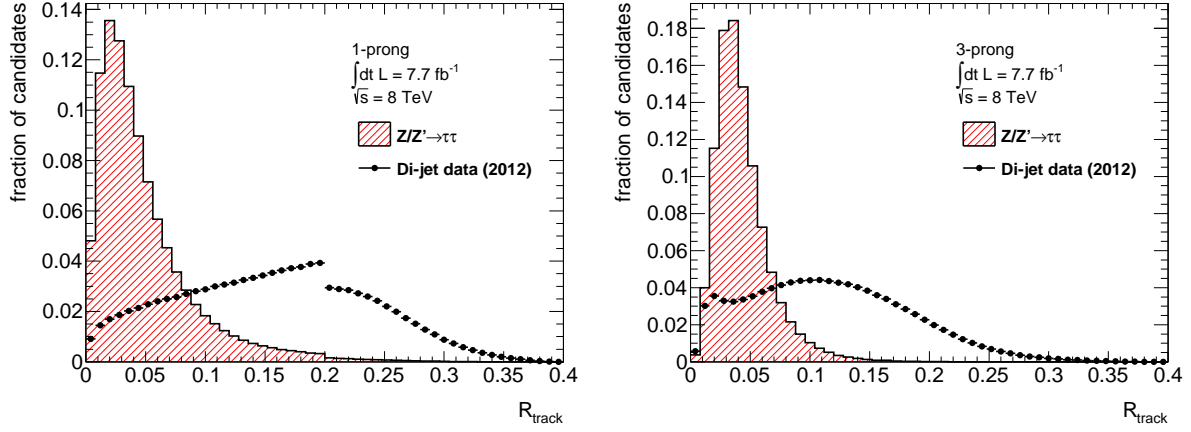
$$R_{\text{track}} = \frac{\sum_i^{\Delta R_i < 0.4} p_{T,i} \cdot \Delta R_i}{\sum_i^{\Delta R_i < 0.4} p_{T,i}}, \quad (5.6)$$

where  $i$  runs over all tracks of the tau candidate within the isolation cone ( $\Delta R < 0.4$  w.r.t. the intermediate tau axis). The transverse momentum of the  $i$ -th track is denoted by  $p_{T,i}$  and  $\Delta R_i$  refers to the distance of the  $i$ -th track to the intermediate tau axis. Real taus have lower track multiplicities which result in smaller values of  $R_{\text{track}}$ . Whereas QCD jets tend to larger values due to the inherent higher track multiplicities and wider shower profiles. Figure 5.2 presents the signal and background distributions for reconstructed 1-prong and 3-prong tau candidates.

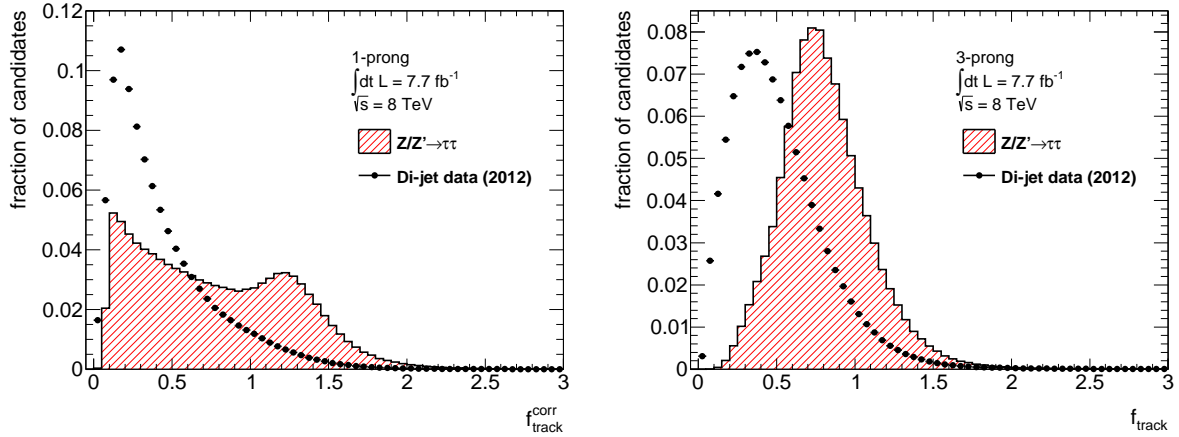
**Leading track momentum fraction,  $f_{\text{track}}$ :** Transverse momentum of the highest- $p_T$  track,  $p_{T,l}^{\text{track}}$ , w.r.t. the transverse energy of the tau candidate,  $E_T^\tau$ :

$$f_{\text{track}} = \frac{p_{T,l}^{\text{track}}}{E_T^\tau}, \quad (5.7)$$

where only tracks reconstructed within the core cone are taken into account and  $E_T^\tau$  is calibrated at the electromagnetic energy scale. Since real taus have a low number of decay products they tend to have  $f_{\text{track}}$  values close to one. Although this can differ especially in the case of 1-prong tau candidates due to additional  $\pi^0$ s which are not considered in the calculation. This quantity is pile-up sensitive, and thus a correction depending on the number of reconstructed vertices,  $N_{\text{vtx}}$ , is applied. The resulting corrected leading track momentum fraction is given by:  $f_{\text{track}}^{\text{corr}} = f_{\text{track}} + 0.003 \cdot N_{\text{vtx}}$  [101]. The distributions for reconstructed 1-prong and 3-prong tau candidates are summarised in Figure 5.3.



**Figure 5.2:** Track radius,  $R_{\text{track}}$ , for reconstructed 1-prong (left) and 3-prong (right) tau candidates. The red dashed histogram illustrates signal events obtained from MC simulation with tau candidates required to match to generated 1-/3-prong decays. The black dots represent the background events extracted from a QCD di-jet selection in data.



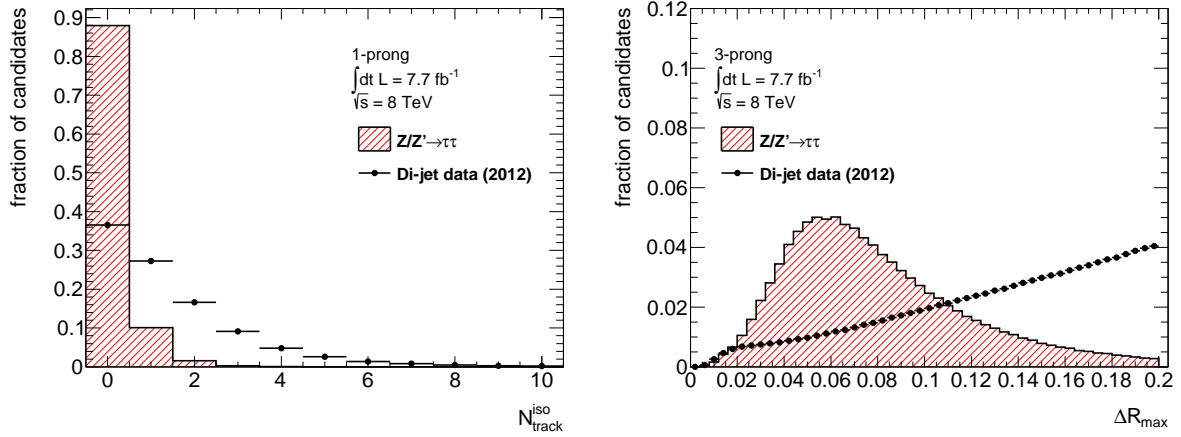
**Figure 5.3:** Leading track momentum fraction,  $f_{\text{track}}^{\text{corr}}$ , for reconstructed 1-prong (left) and 3-prong (right) tau candidates. The red dashed histogram illustrates signal events obtained from MC simulation with tau candidates required to match to generated 1-/3-prong decays. The black dots represent the background events extracted from a QCD di-jet selection in data.

**Number of isolation tracks,  $N_{\text{track}}^{\text{iso}}$ :** Number of reconstructed tracks within the isolation annulus. For real tau decays less tracks are expected than for QCD jets which contain additional tracks originating from gluon splitting. This quantity is only used for reconstructed 1-prong tau candidates. The according distributions of signal and background event can be seen in Figure 5.4 (left).

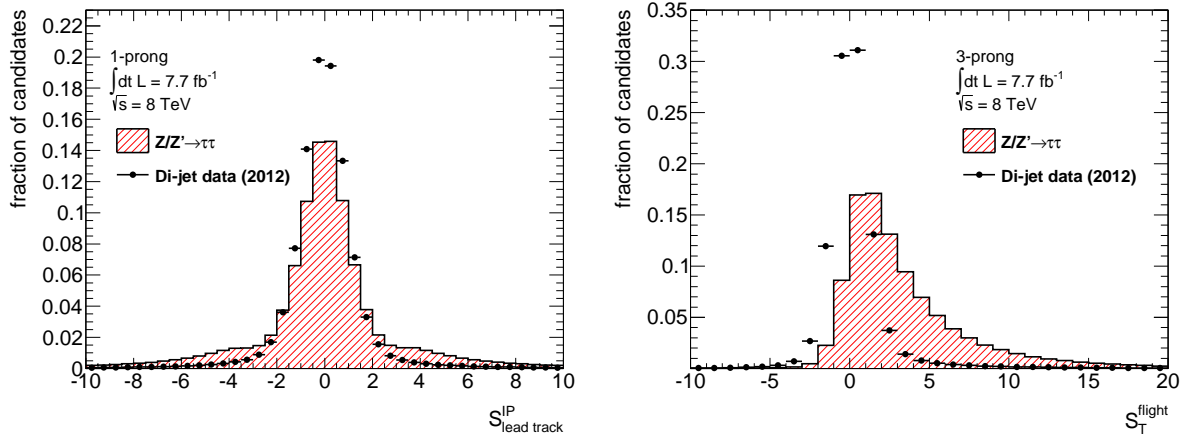
**Maximum  $\Delta R$ ,  $\Delta R_{\text{max}}$ :** Maximal distance between a track associated to the tau candidate and the intermediate tau axis. It is only defined within the core cone. Real tau decays tend to small values of  $\Delta R_{\text{max}}$  given that they are highly collimated compared to QCD jets which are widespread, and thus higher values are observed. This quantity is only used for 3-prong tau decays, as for 1-prong tau candidates it is equivalent to  $R_{\text{track}}$ . Figure 5.4 (right) shows the resulting signal and background distributions.

**Leading track IP significance,  $S_{\text{lead track}}^{\text{IP}}$ :** Impact parameter significance of the highest- $p_T$  track of the tau candidate:

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



**Figure 5.4:** Number of isolation tracks,  $N_{\text{track}}^{\text{iso}}$ , for reconstructed 1-prong (left) and Maximum  $\Delta R$ ,  $\Delta R_{\text{max}}$ , for 3-prong (right) tau candidates. The red dashed histogram illustrates signal events obtained from MC simulation with tau candidates required to match to generated 1-/3-prong decays. The black dots represent the background events extracted from a QCD di-jet selection in data.



**Figure 5.5:** Leading track IP significance,  $S_{\text{lead track}}^{\text{IP}}$ , for reconstructed 1-prong (left) and transverse flight path significance,  $S_T^{\text{flight}}$ , for 3-prong (right) tau candidates. The red dashed histogram illustrates signal events obtained from MC simulation with tau candidates required to match to generated 1-/3-prong decays. The black dots represent the background events extracted from a QCD di-jet selection in data.

with  $d_0$  denoting the closest approach of the leading track to the reconstructed PV in the transverse plane and  $\delta d_0$  stands for its estimated uncertainty. Figure 5.5 shows the distribution only for 1-prong tau decays, as this quantity is only considered for tau candidates with one reconstructed track.

**Transverse flight path significance,  $S_T^{\text{flight}}$ :** Decay length significance of the reconstructed secondary vertex in the transverse plane:

$$S_T^{\text{flight}} = \frac{L_T^{\text{flight}}}{\delta L_T^{\text{flight}}}, \quad (5.9)$$

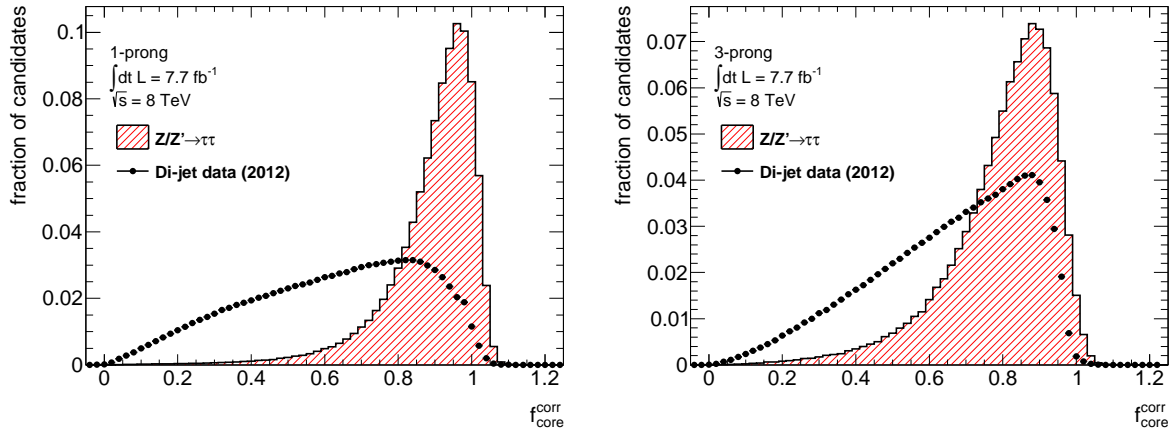
where  $L_T^{\text{flight}}$  stands for the reconstructed signed decay length and  $\delta L_T^{\text{flight}}$  for its estimated uncertainty. All tracks reconstructed within the core cone are used for the vertex fit. Figure 5.5 presents the distributions of this quantity which is only defined for 3-prong tau decays.



**Core energy fraction,  $f_{\text{core}}$ :** Fraction of the transverse energy,  $E_T$ , in the central region ( $\Delta R < 0.1$  w.r.t. the intermediate tau axis) and the core region of the tau candidate:

$$f_{\text{core}} = \frac{\sum_{i \in \{\text{all}\}}^{\Delta R_i < 0.1} E_{T,i}^{\text{EM}}}{\sum_{j \in \{\text{all}\}}^{\Delta R_j < 0.2} E_{T,j}^{\text{EM}}}, \quad (5.10)$$

where  $i$  and  $j$  run over all calorimeter cells associated to the tau candidate within  $\Delta R_{i(j)} < 0.1(0.2)$  around the intermediate tau axis, calibrated at the electromagnetic energy scale. This quantity gives access to the narrowness of the shower of the tau candidate. Real tau decays are more collimated than QCD jets, and thus tend to higher values while smaller values are observed for jets. Further, this quantity is sensitive to energy deposits caused by pile-up events. Hence, a correction depending on  $N_{\text{vtx}}$  is applied to increase the pile-up robustness. Whereby only tau candidates with  $p_T < 80 \text{ GeV}$  are affected which results in the corrected core energy fraction:  $f_{\text{core}}^{\text{corr}} = f_{\text{core}} + 0.003 \cdot N_{\text{vtx}}$  [101]. The distribution for 1-prong and 3-prong tau candidates are depicted in Figure 5.6.



**Figure 5.6:** Core energy fraction,  $f_{\text{core}}^{\text{corr}}$ , for reconstructed 1-prong (left) and 3-prong (right) tau candidates. The red dashed histogram illustrates signal events obtained from MC simulation with tau candidates required to match to generated 1-/3-prong decays. The black dots represent the background events extracted from a QCD di-jet selection in data.

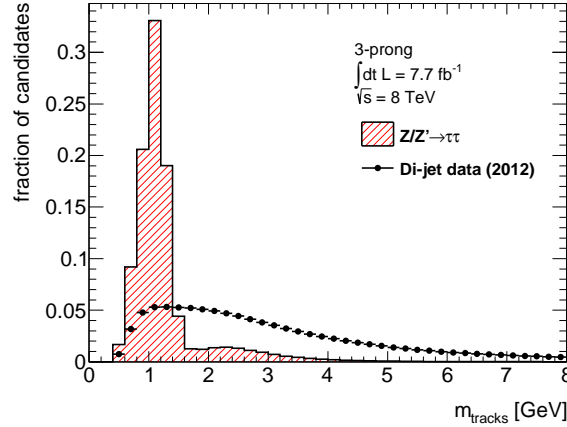
**Mass of the track system,  $m_{\text{tracks}}$ :** Invariant mass of the track system:

$$m_{\text{tracks}} = \sqrt{\left( \sum_{i \in \{\text{tracks}\}}^{\Delta R_i < 0.4} E_i \right)^2 - \left( \sum_{i \in \{\text{tracks}\}}^{\Delta R_i < 0.4} \vec{p}_i \right)^2}, \quad (5.11)$$

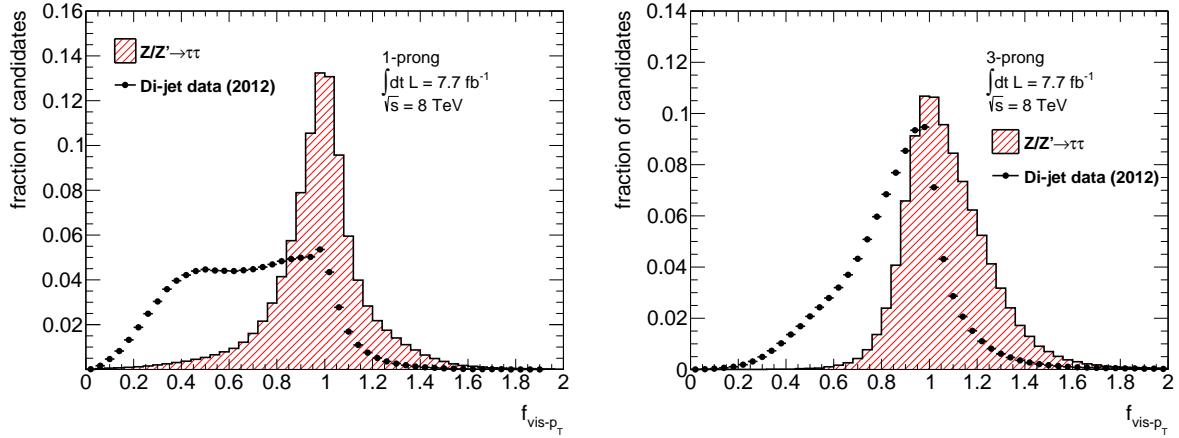
where all tracks within the isolation cone enter the calculation. Considering the occurring not-counted neutral pions in real tau decays values close-by but smaller than the tau mass are expected, while QCD jets lead to a widespread distribution. Reasonably this quantity is only used for reconstructed 3-prong tau candidates, the according distributions are shown in Figure 5.7.

**Transverse momentum ratio of track and  $\pi^0$  system w.r.t. the tau candidate,  $f_{\text{vis-}p_T}$ :** Fraction of the sum of the transverse momenta of charged and neutral pions and the transverse momentum of the tau candidate,  $p_T^\tau$ :

$$f_{\text{vis-}p_T} = \frac{\sum_{i \in \{\text{tracks}, \pi^0\}}^{\Delta R_i < 0.2} p_{T,i}}{p_T^\tau}, \quad (5.12)$$



**Figure 5.7:** Mass of the track system,  $m_{\text{tracks}}$ , for reconstructed 3-prong (right) tau candidates. The red dashed histogram illustrates signal events obtained from MC simulation with tau candidates required to match to generated 1-/3-prong decays. The black dots represent the background events extracted from a QCD di-jet selection in data.

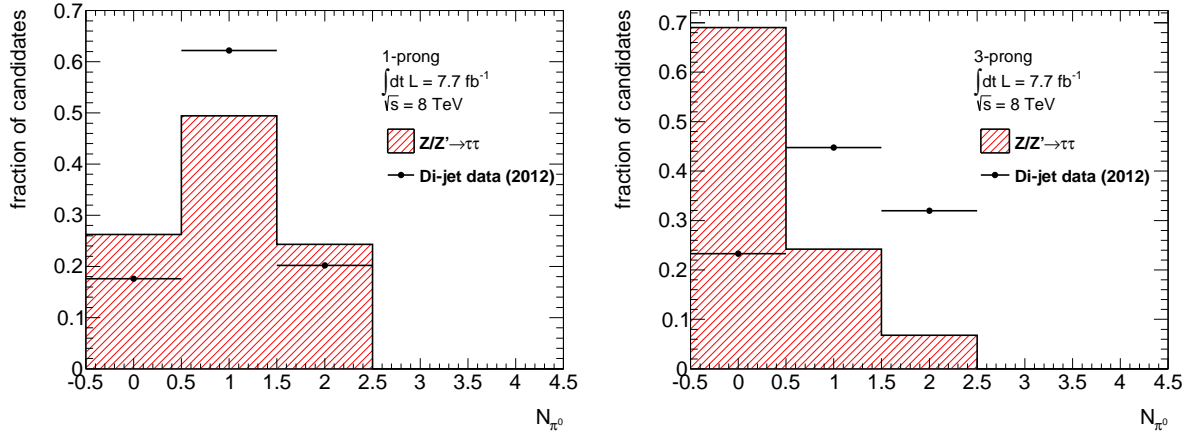


**Figure 5.8:** Transverse momentum ratio of track and neutral pion system w.r.t. the tau candidate,  $f_{\text{vis-p}_T}$ , for reconstructed 1-prong (left) and 3-prong (right) tau candidates. The red dashed histogram illustrates signal events obtained from MC simulation with tau candidates required to match to generated 1-/3-prong decays. The black dots represent the background events extracted from a QCD di-jet selection in data.

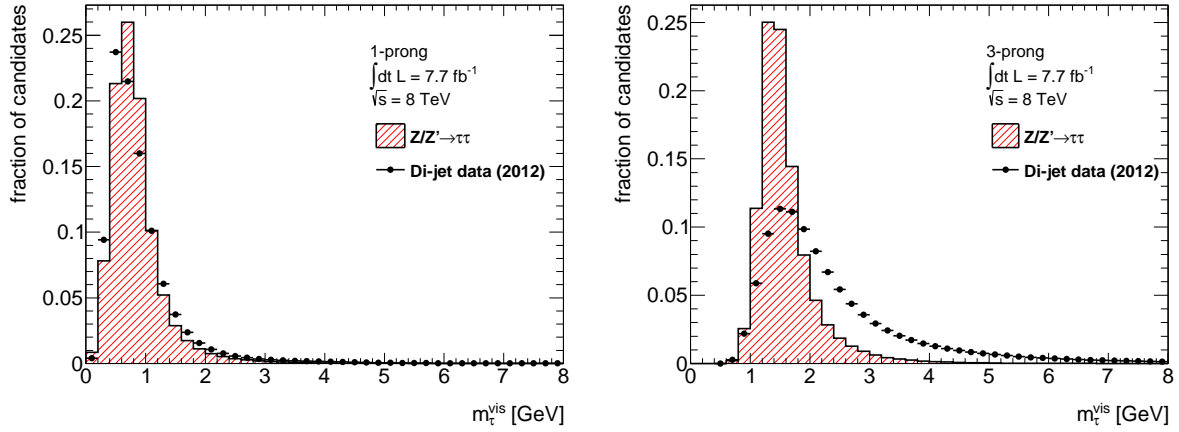
where  $i$  runs over all tracks and neutral pions reconstructed by the  $\pi^0$  finder within the core cone.  $f_{\text{vis-p}_T}$  is close to one for real tau decays, whereas additional neutral hadrons lead to a shift towards smaller values for QCD jets. The distributions for reconstructed 1-prong and 3-prong decays are represented by Figure 5.8.

**Number of reconstructed neutral pions,  $N_{\pi^0}$ :** Number of neutral pions reconstructed by the  $\pi^0$  finder within the core cone. The spectrum of neutral pions originating from real tau decays is well defined, whereas QCD jets tend to higher values. In Figure 5.9 the distributions for reconstructed 1-prong and 3-prong tau candidates are depicted.





**Figure 5.9:** Number of reconstructed neutral pions,  $N_{\pi^0}$ , for reconstructed 1-prong (left) and 3-prong (right) tau candidates. The red dashed histogram illustrates signal events obtained from MC simulation with tau candidates required to match to generated 1-/3-prong decays. The black dots represent the background events extracted from a QCD di-jet selection in data.



**Figure 5.10:** Visible mass of tau candidate,  $m_{\tau}^{\text{vis}}$ , for reconstructed 1-prong (left) and 3-prong (right) tau candidates. The red dashed histogram illustrates signal events obtained from MC simulation with tau candidates required to match to generated 1-/3-prong decays. The black dots represent the background events extracted from a QCD di-jet selection in data.

**Visible mass of tau candidate,  $m_{\tau}^{\text{vis}}$ :** Invariant mass of the track system and the reconstructed neutral pions:

$$m_{\tau}^{\text{vis}} = \sqrt{\left( \sum_{i \in \{\text{tracks}, \pi^0\}} E_i \right)^2 - \left( \sum_{i \in \{\text{tracks}, \pi^0\}} \vec{p}_i \right)^2}, \quad (5.13)$$

whereby only the tracks and clusters within the core cone are considered. Real taus have well defined mass, contrary to QCD jets which lead to arbitrary mass values. Hence, the values for real taus peak slightly below the actual tau mass since only the visible fraction is taken into account but not the invisible neutrino component. This can be seen in Figure 5.10 for reconstructed 1-prong and 3-prong tau candidates.

### 5.3.2 Discrimination Against Jets

Two different multivariate approaches are pursued in the current tau identification [101] to discriminate real taus against QCD jets. On the one hand, an identification based on a boosted decision tree is pro-

vided using a dedicated ROOT software toolkit, TMVA [107], on the other hand, a strategy exploiting a projective log-Likelihood was developed. For training and optimisation of both algorithms strict requirements on the number of reconstructed tracks is applied, such that only tau candidates with one or three reconstructed tracks are taken into account. Separate BDT and LLH classifiers were trained for 1-prong and 3-prong hadronic tau decays differing in the chosen variable sets which are summarised in Table 5.3.

variable	BDT		LLH	
	1-prong	3-prong	1-prong	3-prong
$R_{\text{track}}$	•	•	•	•
$f_{\text{track}}^{\text{corr}}$	•	•	•	•
$N_{\text{track}}^{\text{iso}}$	•		•	
$\Delta R_{\text{max}}$		•		•
$S_{\text{lead track}}^{\text{IP}}$	•		•	
$S_{\text{T}}^{\text{flight}}$		•		•
$f_{\text{core}}^{\text{corr}}$	•	•	•	•
$m_{\text{tracks}}$		•		•
$f_{\text{vis-}p_{\text{T}}}$	•	•		
$N_{\pi^0}$	•	•		
$m_{\tau}^{\text{vis}}$	•	•		

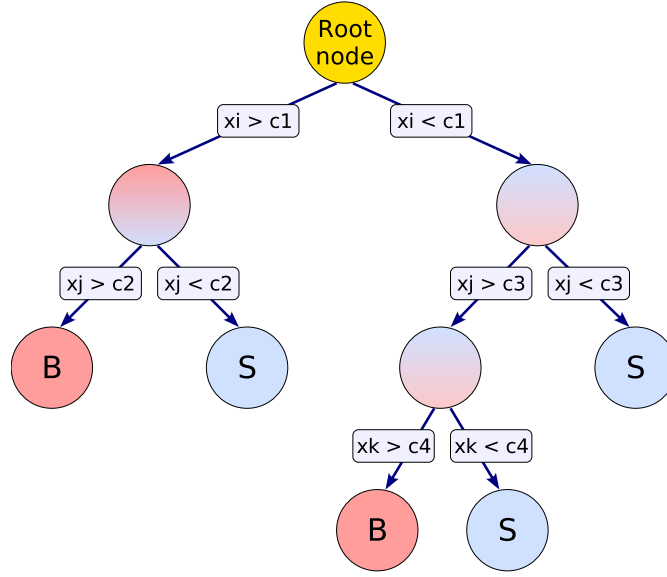
**Table 5.3:** Summary of identification variables used for BDT- and log-Likelihood-based jet discrimination. The black bullets indicate the usage of a particular variable. Different variables are used for each classifier as well as for 1- and 3-prong hadronic tau decays [101].

### Boosted Decision Tree Based Tau Identification

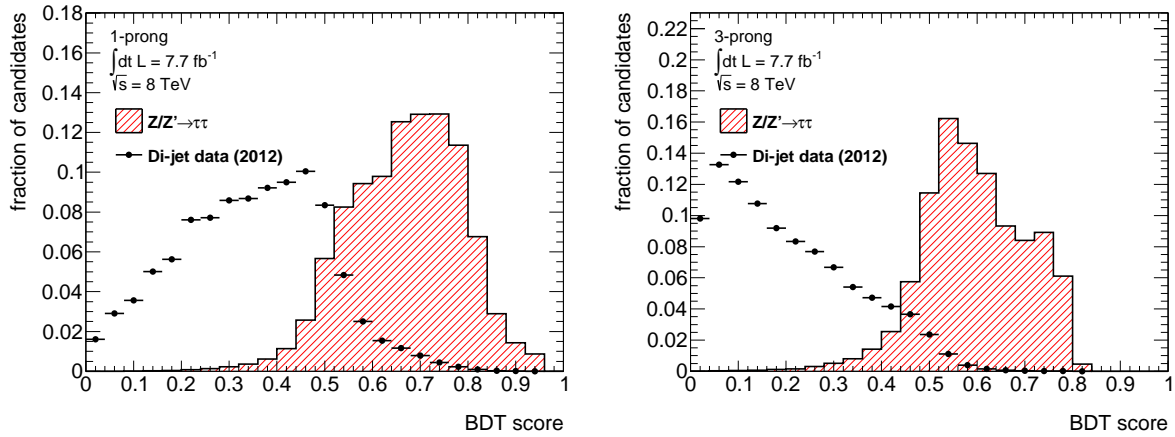
The classifier yielding the best performance in the current tau identification is based on a Boosted Decision Tree [95]. The underlying principle for this approach is a simple *decision tree* (DT) which build a binary tree structure by a series of pass-fail-decisions. Figure 5.11 shows a schematic outline of a simple DT.

Starting from a root node a cut on one variable  $i$  is applied on the entire sample of objects to split it into two parts. Both child nodes still contain signal and background objects but one shows a higher signal purity while the other one is enriched with background events. Further cuts on different discrimination variables ( $j, k, \dots$ ) subsequently split the datasets of the leaf nodes until a certain stopping criterion is reached. A typical stopping criterion is e.g. the minimal number of objects in a single leaf. The final nodes should contain pure samples of signal or background events. Since this is impossible in practice it aims for the highest possible achievable purity for signal and background events in each node.

The advantage of this technique is that objects are not immediately discarded like in a cut-based approach. Nevertheless, a simple decision tree is still a weak classifier and only slightly better than a random decision. To improve the performance a general technique called *boosting* is exploited which is not restricted to DTs. The *adaptive boosting* (AdaBoost) [108] is used in the scope of tau identification. AdaBoost combines several decision trees to yield a gain in performance. Therefore, a weight is applied on the datasets depending on the performance of the previous DT. Thus, higher weights are assigned to misclassified objects, such that they are higher profiled in the next training. The results of all decision trees are summarised exploiting a *weighted majority scheme*, to the final discriminator, a so-called BDT score. The BDT score is conveniently defined in the way that an object with a score value close to 1 is more signal-like than an object which tends to values close to 0, i.e. being more background-like.



**Figure 5.11:** Sketch of a decision tree taken from [107]. Starting from a root node the dataset is split by several cuts on selected variables into signal (blue) and background (red) leaves.



**Figure 5.12:** BDT score distributions for 1-prong (left) and 3-prong (right) decays. The red dashed histograms illustrate signal events obtained from MC simulation with tau candidates required to match to generated 1-/3-prong decays. The black dots represent the background events extracted from a QCD di-jet selection in data.

The currently exploited variables are listed in Table 5.3 separately for 1- and 3-prong tau decays. In Figure 5.12 the resulting BDT score for signal distributions and background events are depicted for reconstructed 1- and 3-prong tau candidates.

As mentioned above a dedicated ROOT toolkit, called TMVA [107], is used for training, optimisation and evaluation of the BDT-based tau identification. For further information on multivariate methods and the boosting technique see e.g. [95].

### Log-Likelihood Based Tau Identification

The second multivariate approach featured for tau identification exploits a log-Likelihood method. A log-Likelihood is based on *probability density functions* (p.d.f.s) of several discriminating variables. For

a given set of variables ( $x_i$ ) the likelihood for being signal- or background-like is defined by:

$$L_{S(B)} = \prod_{i=1}^N p_i^{S(B)}(x_i), \quad (5.14)$$

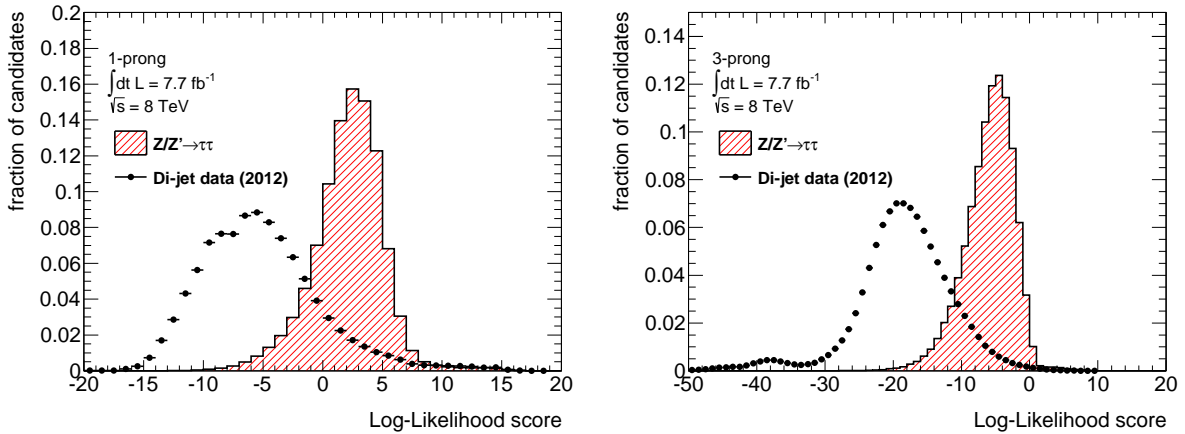
where  $p_i^{S(B)}(x_i)$  denotes the signal (background) p.d.f.s for the  $i$ -th variable. The p.d.f.s are determined separately for 1- and 3-prong tau candidates. Variables used by the current log-Likelihood approach are summarised in Table 5.3. A strong dependency on the kinematics of the tau decay are observed for the p.d.f.s. Thus, they are binned in the reconstructed transverse momentum of the tau candidate. Three  $p_T$  bins are chosen:  $p_T < 45$  GeV,  $45 \text{ GeV} \leq p_T < 100$  GeV and  $p_T \geq 100$  GeV. Discontinuities can occur along the bin borders, hence an interpolation procedure is applied. The affected values correspond to  $p_T$  entries in a symmetric window of  $\pm 10$  GeV around the according bin boarder. Given this, the final likelihood value,  $L_{\text{final}}$ , can be written as:

$$L_{\text{final}} = L \cdot \frac{10 + \delta[\text{GeV}]}{20} + L' \cdot \frac{10 - \delta[\text{GeV}]}{20}, \quad (5.15)$$

where  $L$  is the likelihood value corresponding to the bin into which the tau candidate falls and  $L'$  the one of the adjacent bin. Further,  $\delta$  denotes the distance of the transverse momentum of the tau candidate,  $p_T^\tau$ , to the bin border,  $p_T^{\text{bin edge}}$ , i.e.  $\delta = |p_T^\tau - p_T^{\text{bin edge}}|$ . An exception is made in the case of 1-prong tau candidates for which an asymmetric window of  $[-30, +60]$  GeV is chosen for the last  $p_T$  bin. The combination of the signal and background likelihood values,  $L_S$  and  $L_B$ , for the same variable set results in the final discriminator, the log-likelihood score,  $S_{\text{LLH}}$ , given by:

$$S_{\text{LLH}} = \ln\left(\frac{L_S}{L_B}\right) = \sum_{i=1}^N \ln\left(\frac{p_i^S(x_i)}{p_i^B(x_i)}\right). \quad (5.16)$$

Figure 5.13 presents the signal and background distributions of the log-Likelihood scores for 1- and 3-prong tau candidates.



**Figure 5.13:** Log-Likelihood score distributions for 1-prong (left) and 3-prong decays. The red dashed histograms illustrate signal events obtained from MC simulation with tau candidates required to match to generated 1-/3-prong decays. The black dots represent the background events extracted from a QCD di-jet selection in data.

### Evaluation and Comparison of a BDT-based and a Log-Likelihood-based Tau Identification

To quantify the performance of the tau identification algorithm an evaluation on a statistical independent dataset is performed. Looser requirements on the selected events are set to match the possible physics

needs. Hence, also two reconstructed tracks for one tau candidate are taken into account and are summarised with reconstructed 3-prong decays to so-called *multi-prong* decays. Given this, three types of tau candidates are defined:

- **1-prong tau candidates:** Tau candidates with one reconstructed track which matches to a generated tau decaying into one charged hadron.
- **3-prong tau candidates:** Tau candidates with three reconstructed tracks which match to a generated tau decaying into three charged hadrons.
- **multi-prong tau candidates:** Tau candidates with two or three reconstructed tracks which match to a generated tau decaying into three charged hadrons.

The first two types of tau candidates are solely used for training and optimisation of the classifiers while within the evaluation also the last type is considered. Those definitions are only valid for MC generated events used as signal since a truth matching is a required which is not needed for events extracted from data.

Three figures of merit are defined to evaluate the performance of the individual classifier but also to allow for comparison between them. One is the signal efficiency which is defined as the ratio of reconstructed 1(multi)-prong tau candidates passing the identification criteria and generated 1(3)-prong tau decays as following:

$$\epsilon_{1(\text{multi})\text{-prong}}^{\text{sig}} = \frac{N_{\text{passed}}^{1(\text{multi})\text{-prong}}}{N_{\text{generated}}^{1(3)\text{-prong}}} . \quad (5.17)$$

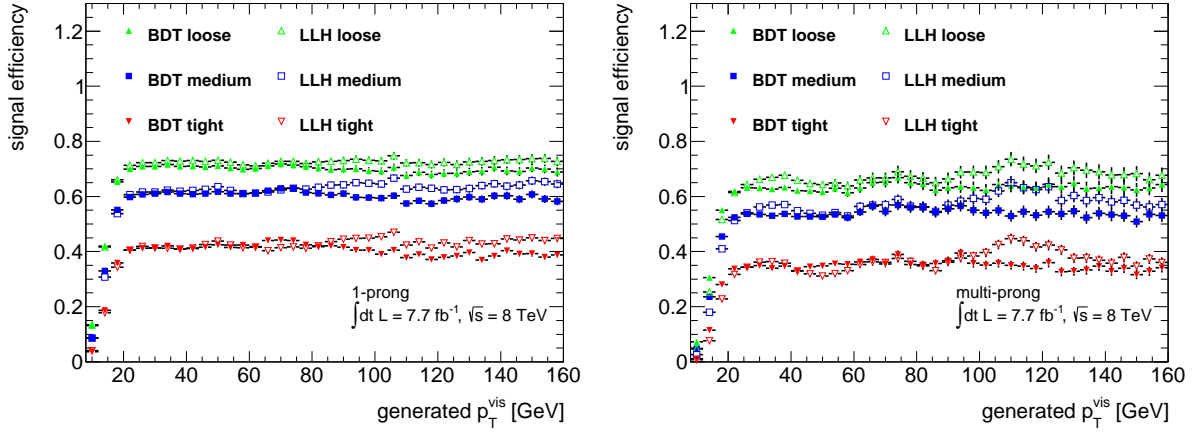
Three working points, *loose*, *medium* and *tight*, corresponding to predefined target signal efficiencies of 70 % (65 %), 60 % (55 %) and 40 % (35 %) are defined for 1(multi)-prong tau decays. The working points are assigned to cuts on the signal score of the classifier, conveniently those cuts are determined, such that the signal efficiency is flat against the generated  $p_T$  of the tau candidate. This is presented in Figure 5.14 for the BDT- and log-Likelihood-based tau identification for 1- and multi-prong tau decays, whereby not the pure identification efficiency but the combined reconstruction and identification efficiency is shown. For the log-Likelihood-based classifier an unstable behaviour is observed for multi-prong tau candidates, while in all other cases the expected flat behaviour is observed. The issue for multi-prong decays will be discussed in more detail in Section 6.1. The sudden shift at  $p_T = 80 \text{ GeV}$  can be tracked down to the change in the pile-up correction of the discriminating variables. To ensure that the signal efficiency is unaffected by pile-up a flat behaviour w.r.t. the average number of interactions per bunch crossing,  $\mu$ , is expected. This is fulfilled for both classifiers thanks to the pile-up corrections applied on the input variables and can be seen in Figure 5.15.

The background efficiency, i.e. the acceptance for fake taus is given by:

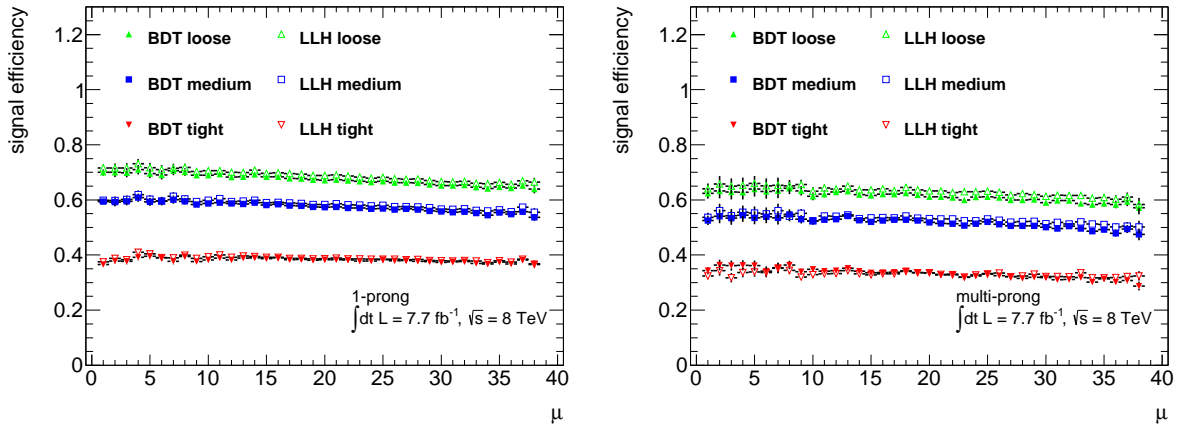
$$\epsilon_{1(\text{multi})\text{-prong}}^{\text{bkg}} = \frac{N_{\text{passed}}^{1(\text{multi})\text{-prong}}}{N_{\text{reconstructed}}^{1(\text{multi})\text{-prong}}} , \quad (5.18)$$

where apparently no match of the reconstructed tau candidate to a generated tau is required. Since fake taus exhibit a different  $p_T$  dependence than real tau decays the background efficiency will not be flat against  $p_T$ . Further, it strongly depends on the pre-selection and the fake composition, i.e. the ratio of gluons and quarks. Figure 5.16 presents the background efficiency against the  $p_T$  of reconstructed taus for the BDT-based and log-Likelihood-based identification algorithms. They differ by up to a factor 2, whereby the log-Likelihood approach shows a higher probability of misclassification. The huge difference between the BDT-based and log-Likelihood-based identification algorithms also reflects in the background efficiency w.r.t. the average number of interactions per bunch crossing. Nevertheless, a flat behaviour is achieved as depicted in Figure 5.17.

The signal and background efficiencies are combined to the final figure of merit which is suitable to compare the performance of different classifiers: the background rejection rate for a given signal

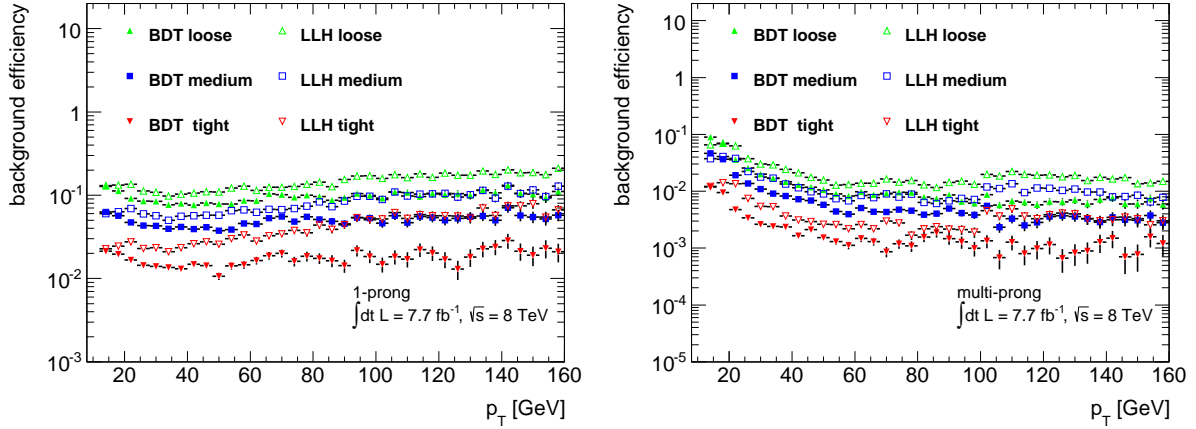


**Figure 5.14:** Signal efficiency against generated tau  $p_T$  for 1-prong (left) and multi-prong (right) tau decays. The BDT-based tau identification is represented by the filled markers and the log-Likelihood-based approach by the empty markers. The three working points loose, medium and tight are depicted for both classifiers as green triangles, blue squares and red triangles, respectively.

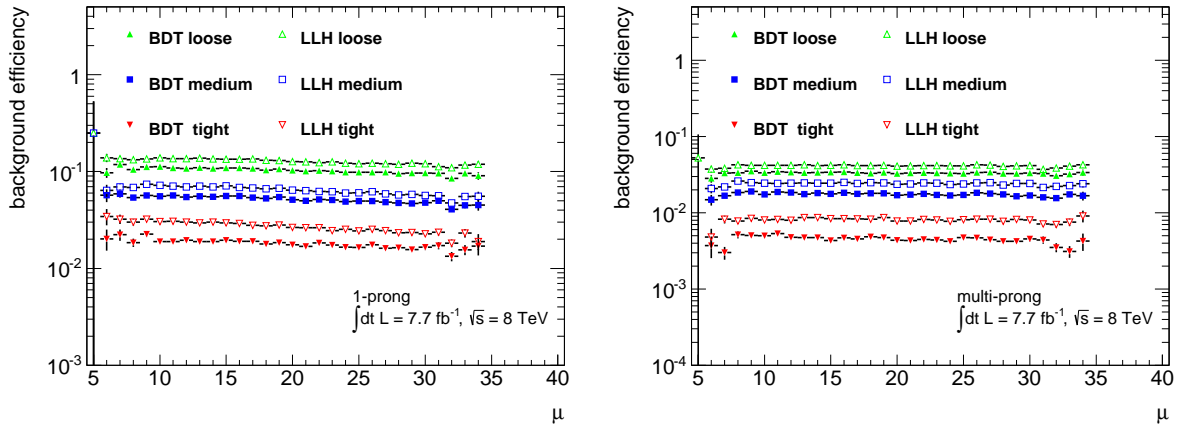


**Figure 5.15:** Signal efficiency against the average number of interactions per bunch crossing for 1-prong (left) and multi-prong (right) tau candidates. The BDT-based tau identification is represented by the filled markers and the log-Likelihood-based approach by the empty markers. The three working points loose, medium and tight are depicted for both classifiers as green triangles, blue squares and red triangles, respectively.

efficiency. The background rejection rate is defined as the inverse of the background efficiency. Since the  $p_T$  dependence of the classifiers are different for signal and background events the cuts on the scores are scanned over the entire signal efficiency range. The resulting distributions are illustrated in Figure 5.18 for the BDT- and log-Likelihood-based 1- and multi-prong tau identification. The upper bound on the signal efficiency is explained by the loss of tracks in the reconstruction step. It is observed that the BDT approach leads to a significantly better performance as the background rejection rate is higher than for a given signal efficiency. Nevertheless, it is useful to consider the log-Likelihood-based algorithm for cross validation and to provide an alternative approach for physics analysis groups since it might be more suitable in some phase space regions.



**Figure 5.16:** Background efficiency against reconstructed tau  $p_T$  of 1-prong (left) and multi-prong (right) tau decays. The BDT-based tau identification is represented by the filled markers and the log-Likelihood-based approach by the empty markers. The three working points loose, medium and tight are depicted for both classifiers as green triangles, blue squares and red triangles, respectively.

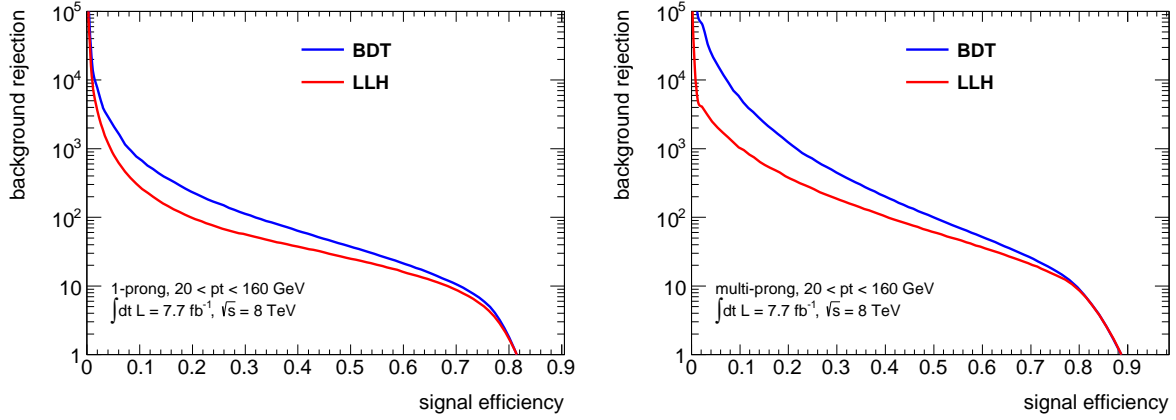


**Figure 5.17:** Background efficiency against the average number of interactions per bunch crossing for 1-prong (left) and multi-prong (right) tau candidates. The BDT-based tau identification is represented by the filled markers and the log-Likelihood-based approach by the empty markers. The three working points loose, medium and tight are depicted for both classifiers as green triangles, blue squares and red triangles, respectively.

### Performance of Identification Algorithms

The tau identification algorithms are based on MC generated events, and thus an evaluation on data is essential to verify the correct modelling of the input variables and their correlations. Hence, a pure tau sample extracted from data is needed, which is practically challenging and not completely achievable. The remaining background has a QCD component of multi-jet events as well as an electroweak part caused by quarks and gluons in association with vector bosons. To account for the background contamination a *template fit method* based on an *extended track multiplicity distribution* [101] is performed, before and after the identification criteria is applied, to parameterise the yields of the classifier. Therefore, the *tag-and-probe method* [109] is exploited on  $Z \rightarrow \tau_l \tau_h$  events where the leptonically decaying tau is tagged. Thereby, both the electron and the muon final state are considered in the current analysis. The hadronically decaying tau is used for the performance measurement of the identification algorithm. The resulting identification efficiency is given by the ratio of taus passing the identification criteria and



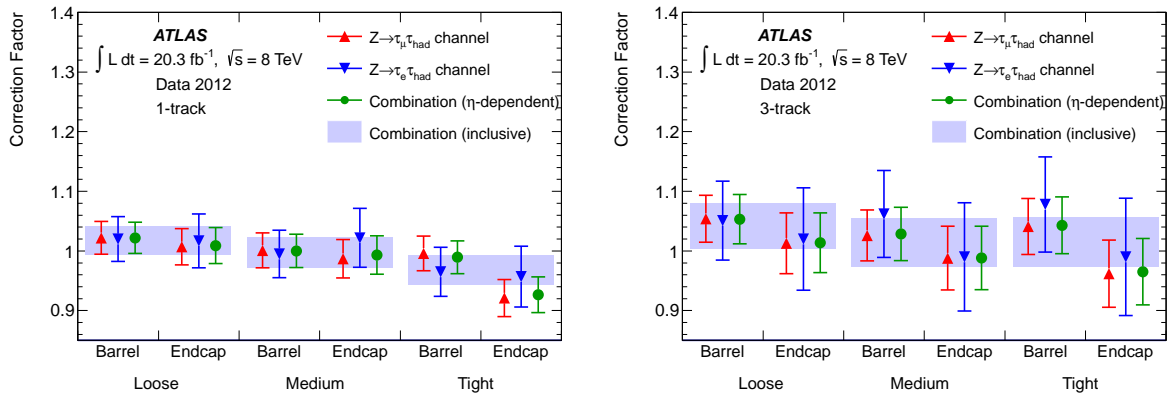


**Figure 5.18:** Background rejection against signal efficiency for 1-prong (left) and multi-prong (right) tau candidates. The BDT-based tau identification is represented by the blue line and the log-Likelihood-based approach by the red line. The distributions are shown in the  $p_T$  range of [20, 160] GeV.

all considered taus. Consequently, the efficiency is measured in both, the events obtained from a selection in data and from MC simulation. A comparison between both yields the quality of the performance prediction expressed in so-called *scale factor* (SF) defined as:

$$SF = \frac{\epsilon_{\text{data}}}{\epsilon_{\text{MC}}} . \quad (5.19)$$

One scale factor is determined for each working point and separately for 1- and multi-prong hadronic tau decays. The scale factors are applied as further correction factors in physics analysis. The most recent scale factors valid for the official tau identification based on a BDT, are summarised in Figure 5.19. It is expected that they are close to one, i.e. the training and optimisation based on MC events yield a good prediction of data, which can be achieved quite well.



**Figure 5.19:** Scale factors for BDT-based tau identification for 1-prong (left) and multi-prong (right) tau candidates [96].

### 5.3.3 Discrimination against Light Leptons

The two tau identification approaches described in Section 5.3.2 provide a good discrimination against jets but cannot account for the acceptance of fake taus originating from lighter leptons, i.e. electrons and muons. Hence, individual algorithms to reject those fake taus are developed.



### Electron Veto

Electrons raise a track in the inner detector and a cluster in the electromagnetic calorimeter while traversing the detector, and thus can mimic 1-prong hadronic tau decays. A BDT is chosen to reduce the number of electron fakes. MC generated  $Z \rightarrow \tau\tau$  and  $Z \rightarrow ee$  events are used for signal and background, respectively. Thereby they have to pass a  $p_T$  threshold of 20 GeV and fulfil a loose BDT-based tau identification criterion. The variables entering the classifier are motivated by the different shower profiles of taus and electrons. The showers emerging from hadronically decaying taus are typically wider and longer than those of electrons. Moreover, a higher probability for the occurrence of transition radiation originating from electrons is exploited. Like for the jet discrimination three working points, loose, medium and tight, corresponding to signal efficiencies of 95 %, 85 % and 75 %, are defined. Only reconstructed 1-prong taus matching to generated tau decays with one track within  $\Delta R < 0.2$  are taken for the determination of the signal efficiency. Whereas the background efficiency is based on reconstructed electrons which match to a generated electron within  $\Delta R < 0.2$ . The electron veto is divided in four  $\eta$  regions: barrel ( $|\eta| < 1.37$ ), crack ( $1.37 < |\eta| < 1.52$ ), end-cap ( $1.52 < |\eta| < 2.0$ ) and forward end-cap region ( $|\eta| > 2.0$ ). For each a dedicated variable set is optimised to obtain the best performance. Scale factors on the electron discrimination are extracted by exploiting the tag-and-probe method from simulated  $Z \rightarrow ee$  events and reconstructed hadronic tau decays. They are considered as correction factors in the individual analysis. The remaining background is formed by QCD multi-jets,  $W \rightarrow \tau\nu_\tau$ ,  $Z \rightarrow \tau\tau$  and  $t\bar{t}$  events. The uncertainties on the scale factor is higher than for jet discrimination algorithms ranging from 20-77 %, depending on the  $\eta$  region.

### Muon Veto

Muons are minimal ionising particles at LHC energies, and thus may be reconstructed as 1-prong tau candidates in rare cases. They leave a track in the inner detector and mostly traversing the calorimeter without any interaction. Nevertheless, high energetic muons can cause calorimeter clusters, and hence mimic tau decays given that the measurement in the muon spectrometer failed. Low energy muons can be stopped already in the calorimeter, which eventually leads to a misclassification as tau. Photon radiation of muons can also lead to energy deposits in the ECAL with slight leakage into the HCAL. To discriminate against muons a cut-based technique which relies on the leading track momentum fraction,  $f_{\text{track}}$ , and the energy fraction measured in the electromagnetic calorimeter,  $f_{\text{EM}}$ , is used. The latter quantity is typically lower for muons than for hadronic taus decays. To optimise the classifier MC generated  $Z \rightarrow \tau\tau$  and  $Z \rightarrow \mu\mu$  events are used as signal and background, respectively. As a consequence of this discrimination method the muon fake rate is reduced by 40 %, while 96 % of real hadronically decaying taus are preserved.

## 5.4 Tau Trigger

Finally, it should be mentioned that a dedicated tau trigger [15, 110, 111] exists to cope with the high bunch crossing rates. As important the triggering on taus is for many analysis as challenging it is due to the high multi-jet production and the consequent fake rates.

The L1 tau trigger is hardware based and uses the information provided by the calorimeter and muon system with coarser granularity. Electromagnetic as well as hadronic trigger towers with a size of  $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$  are used. A core and an isolation region are defined on a set of  $2 \times 2$  ( $4 \times 4$ ) trigger towers corresponding to a size of  $\Delta\eta \times \Delta\phi = 0.2 \times 0.2$  ( $\Delta\eta \times \Delta\phi = 0.4 \times 0.4$ ). Different cuts on  $E_T$  are available, corresponding to 8, 11, 15, 20 and 40 GeV. Further it can be combined with other L1 triggers like triggers on electrons or  $E_T^{\text{miss}}$ . At this stage of the trigger chain a region of interest is defined on which the L2 trigger is executed.

The aim for the L2 trigger is to reject as much multi-jet background as possible without losing real taus. Therefore, signatures with low track multiplicities and narrow shapes are selected based on the measurements of the inner detector and calorimeter system, respectively. The calorimeter cells within a

set of  $0.8 \times 0.8$  trigger towers are used to refine the position of the ROI. Afterwards,  $E_T$  thresholds and noise suppression are applied on each cell and the variables describing the shower shapes are calculated, similar to the ones used in the offline identification. The track-based variables are determined by track algorithms. Cuts on the set of variables depend on the chosen trigger item and the prongness. They are optimised w.r.t. the offline identification algorithms to contribute to an optimal efficiency for physics analysis. The already in Section 5.3.1 introduced identification variables  $f_{\text{track}}$ ,  $R_{\text{track}}$  and  $f_{\text{core}}$ , but also the transverse energy, sum of the transverse momentum in the core and isolation region and the number of tracks, are exploited.

The final step in the trigger chain is the event filter which is configured very close to the offline reconstruction and identification. Although it is not updated as often as the offline algorithms, it is ensured that the selected phase space does not change significantly. Thus, the set of discriminating variables is similar to those summarised in Table 5.3 and contains the following quantities:  $f_{\text{core}}$ ,  $f_{\text{track}}$ ,  $R_{\text{track}}$ ,  $S_{\text{lead track}}^{\text{IP}}$ ,  $N_{\text{track}}^{\text{iso}}$ ,  $S_T^{\text{flight}}$ ,  $m_{\text{tracks}}$  and  $\Delta R_{\text{max}}$ . Only pile-up uncorrected variables enter the BDT and log-Likelihood algorithms since the provided time is not long enough to perform a vertex reconstruction. Nevertheless, a cut on  $\Delta z_0$  can increase the pile-up robustness sufficiently. Further, no tau energy calibration is applied which leads to a shift of the calorimeter based variables. Since no dedicated data is available for the training of the classifiers MC simulated events are used for signal as well as background. For 2012 data-taking only the BDT-based EF was online as the time constraints did not allow to run both algorithms in parallel, such that the better performing one was chosen.

## 6 Optimisation of Tau Lepton Identification

A continuous optimisation of the tau identification algorithms is crucial to guarantee an outstanding performance and cope with changing circumstances, like increasing contribution from pile-up events, but also to profit from developments of the software used. Section 6.1 will focus on this topic.

However, the default strategies might not always be the best choice for some physics analyses or phase space regions. Thus, further investigations are essential to satisfy physics analyses needs. As discussed in Section 6.2 one important field of physics studies in the next data taking run will be focused on measurements concerning the exploration of the nature of the observed Higgs boson. A key ingredient for these studies will be the CP measurement in the di-tau channel. The current tau identification might not be an optimal choice for this. Hence, a new identification based on detailed information of the tau decay substructure has been developed. It exploits the recently developed reconstruction algorithms, namely CellBased and EflowRec (cf. Section 5.2), aiming to reach or even exceed the performance of the current standard tau identification. This will be discussed in detail in Section 6.2.

### 6.1 Optimisation of the Default Tau Identification

Both tau identification techniques introduced in Section 5.3.2 are optimised within the scope of this thesis. The log-Likelihood-based tau identification was abandoned in the last round of ATLAS analysis updates, and thus needs to be updated to the state of the BDT strategy. This includes physics but also technically motivated modifications of the algorithm. The latter was raised for the BDT approach as well, due to changes in the underlying software chain. All performed updates and their results are the topic of the following sections.

#### 6.1.1 Optimisation of the Boosted Decision Tree Based Tau Identification

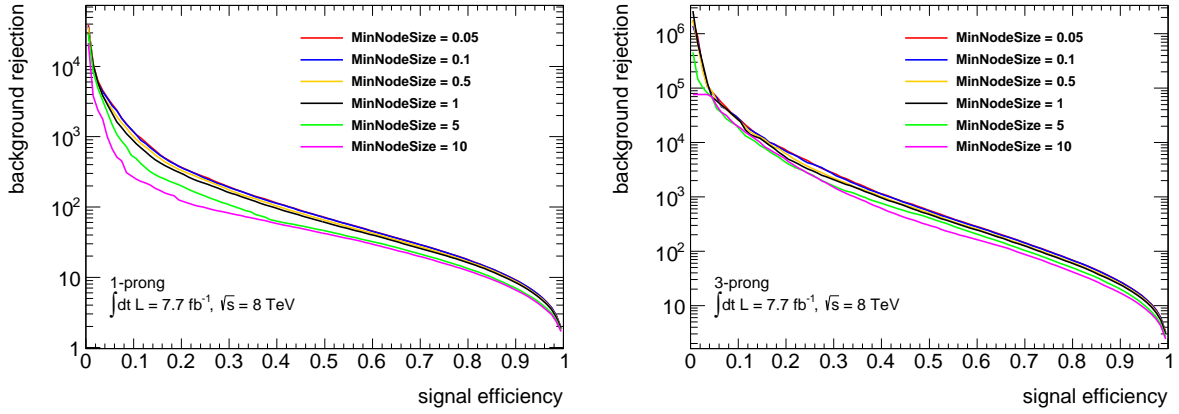
As mentioned in Section 5.3.2 the ROOT toolkit TMVA provides the technical framework for the BDT-based tau identification. A recent change of the configuration options of the BDT classifier due to optimisation raised the need of a closer investigation of this modification.

#### Optimisation of the BDT Configuration

The way to set the option to restrict the minimal number of objects per leaf was changed from an absolute to a relative choice. This quantity is very important as it defines the stopping criterion of the DT training. Within TMVA this option is now called `MinNodeSize`<sup>8</sup>. The default value is currently defined as `MinNodeSize = 5` corresponding to 5 % of the events in the starting dataset at the root node. An amount of events in the order of one million is available for the training of the BDT for both, 1- and 3-prong tau candidates, which leads to a quite large number of roughly 50000 events in the final node. Thus, the selected hyper cube in the n-dimensional grid is very coarse, resulting in a loss of information. The specific cell will have too poor purity, such that the full potential of the variable set is not exhausted. To investigate whether a more suitable value can be chosen for `MinNodeSize`, various BDTs for several smaller and larger values were trained. The background rejection against signal efficiency is used to verify the performance of each individual configuration. The resulting curves are depicted in Figure 6.1. Thereby the pure identification efficiency is presented. It can be seen that the default value of 5 % does not provide the best obtainable performance. Also higher values like 10 % lead to a significantly worse performance as the node size becomes even bigger, and thus the purity decreases further. In contrast, for smaller values a gain can be observed. 0.1 % turned out to be the best choice to adjust the minimal number of events per node. Although, the configuration with `MinNodeSize=0.05` shows a slightly better performance, the gain is negligible and does not justify the higher CPU consumption.

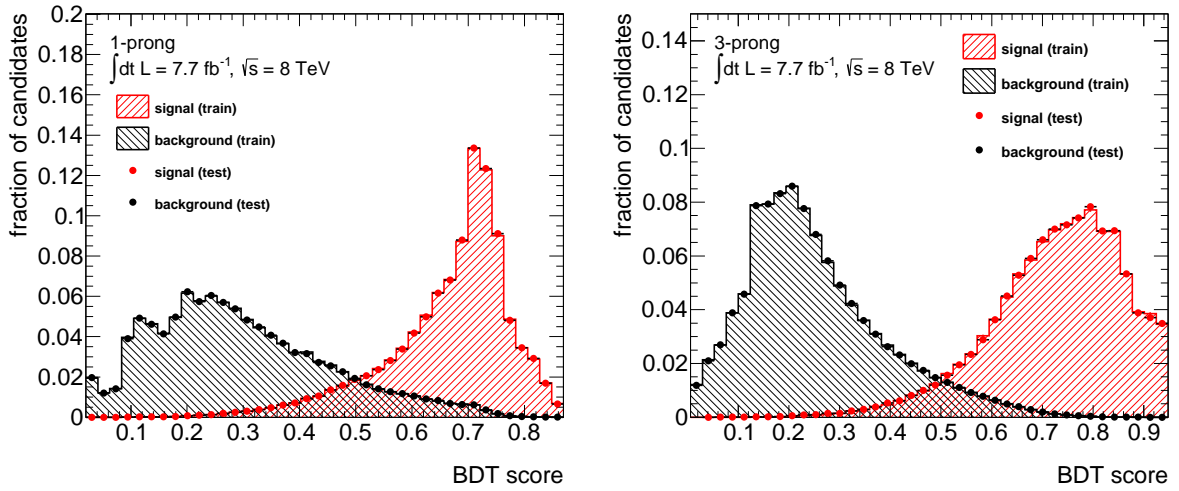
---

<sup>8</sup>The TMVA version used for the studies presented within this thesis is TMVA-v4.2.0. The minimal number of objects per leaf was previously (TMVA-v4.1.4) adjusted with the option `nEventsMin`.



**Figure 6.1:** Background rejection against signal efficiency for  $\text{MinNodeSize} = [0.05, 0.1, 0.5, 1, 5, 10]$  for 1-prong (left) and 3-prong (right) tau candidates. The pure identification efficiency is presented.

The configuration corresponding to a  $\text{MinNodeSize}$  of 0.1 % needs to be evaluated on a statistically independent dataset. A so-called overtraining test is performed to verify whether the observed performance of this specific BDT set-up is able to yield an equivalent performing classification of signal and background events on a statistically independent dataset. The signal and background score distributions for the training and testing dataset of this configuration are depicted in Figure 6.2. Both signal as well as background distributions agree very well for the training and testing dataset. Hence, no overtraining occurs and the performance will not degrade once the BDT is applied to a similar event topology. A BDT configuration for which this agreement is not observed, cannot serve as a basis of an identification as its performance will be biased. Further, the impact of the change in the BDT configuration on the other



**Figure 6.2:** Overtraining test for a BDT configuration with  $\text{MinNodeSize} = 0.1$  % for 1-prong (left) and 3-prong (right) tau candidates. The signal (red) and background (black) distributions for the dataset used for training and evaluation are illustrated as dashed histograms and filled dots, respectively.

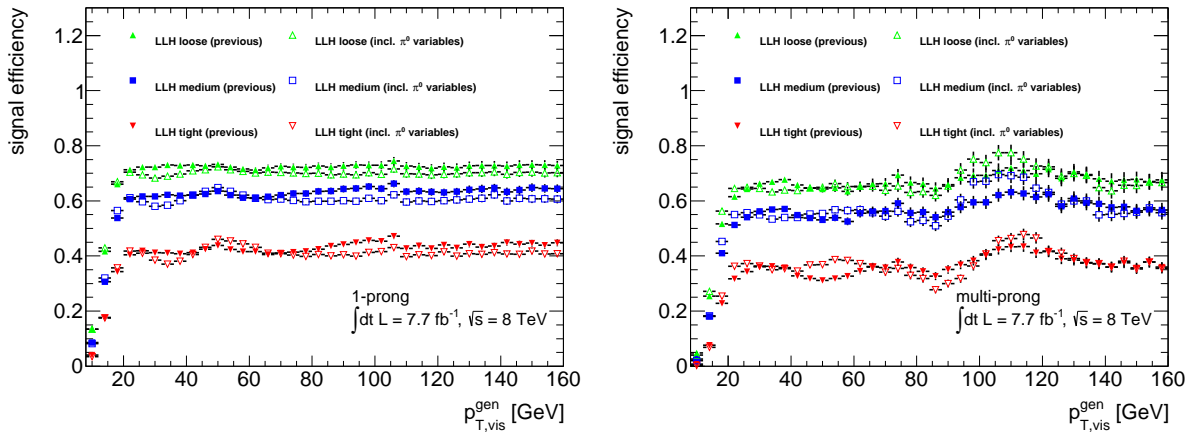
options, like the number of trained trees or the number of applied cuts, was studied and has not shown any deviations, and hence the adjustment for those options is preserved.

### 6.1.2 Optimisation of the Default Log-Likelihood Based Tau Identification

Three updates are carried out for the LLH-based tau identification. First, the considered variable set is extended to those exploited for the BDT approach. A reweighting of the used events extracted from MC simulation and data is applied according to the strategy used for the BDT which already utilises this. Additionally, a technical correction of the p.d.f. calculation is implemented. The updates are applied successively, in order that their individual impact can be analysed.

#### Extension of the Variable Set

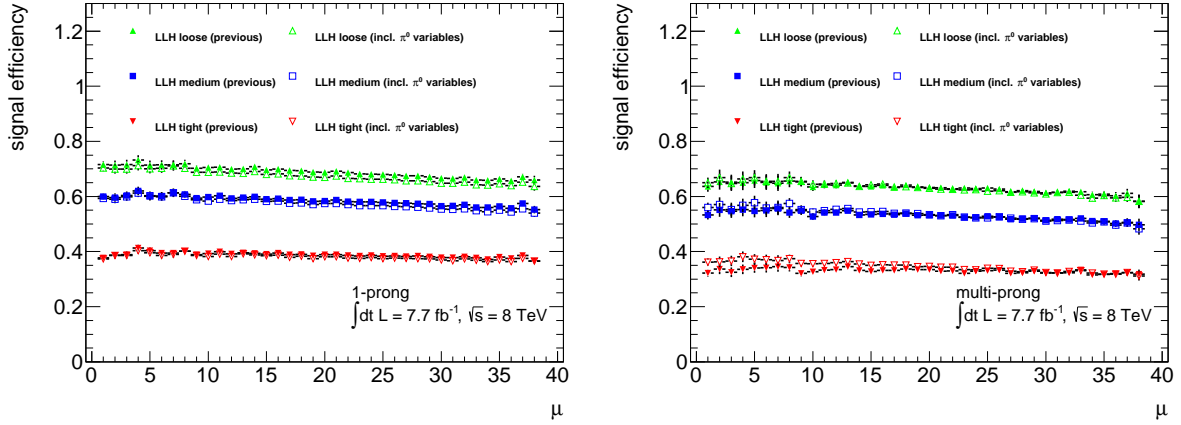
The variable sets used previously are summarised in Table 5.3 and do not include quantities based on reconstructed neutral pions, i.e.  $f_{\text{vis-}p_T}$ ,  $N_{\pi^0}$  and  $m_{\tau}^{\text{vis}}$ . In the case of the BDT-based identification a remarkable performance gain was observed by consideration of those quantities and an up to 100 % [112] more efficient background rejection was achieved. Thus, it is reasonable to add the  $\pi^0$  variables also to the LLH variable sets. Their p.d.f.s are determined and considered in the calculation of the log-Likelihood discriminator. The extracted signal efficiencies follow the expected flat behaviour for 1-prong tau candidates against the generated visible transverse momentum,  $p_{T,\text{vis}}^{\text{gen}}$ , and the average number of interactions per bunch crossing,  $\mu$ . This is shown in Figure 6.3 (left) and 6.4 (left) in combination with the results based on the old variable set. In the case of multi-prong tau decays this also can be confirmed



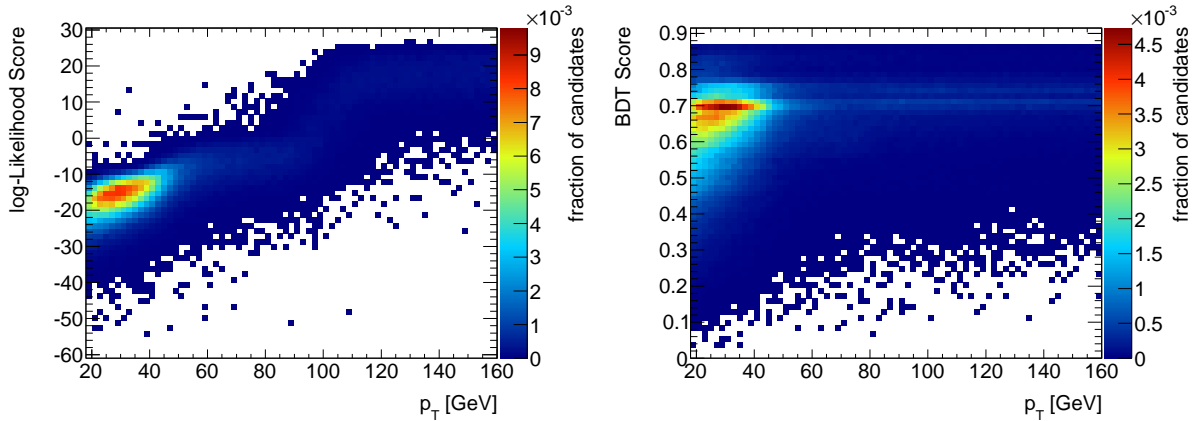
**Figure 6.3:** Signal efficiency against generated visible transverse tau momentum,  $p_{T,\text{vis}}^{\text{gen}}$ , based on a LLH-based tau identification for 1-prong (left) and multi-prong (right) tau decays. The results for the variable set without  $\pi^0$  variables is represented by the filled markers while the empty markers stand for the variable set including those variables. The three working points loose, medium and tight are depicted for both classifiers as green triangles, blue squares and red triangles, respectively.

w.r.t.  $\mu$  as can be seen in Figure 6.4 (right). Whereas the instabilities against  $p_{T,\text{vis}}^{\text{gen}}$  are still present as already observed for the variable set without  $\pi^0$  quantities (cf. Section 5.3.2). Those deviations arise from the strong  $p_T$  dependence of the log-Likelihood score as Figure 6.5 shows. Whereas the BDT score does not suffer that much from this issue, and thus the signal efficiency is not harmed.

Figure 6.6 presents the background rejection against signal efficiency of the identification with the log-Likelihood method exploiting the variable set with and without quantities based on reconstructed neutral pions. As can be seen the consideration of the  $\pi^0$  variables does not provide any remarkable gain in performance. Both distributions are very close to each other, only for 1-prong tau decays a small gain of about 5 % is observed for signal efficiencies above 60 %. In the case of multi-prong tau candidates the updated algorithm even leads to a slight performance loss of 5 % to 10 % for the interesting efficiency range between 30 % and 80 %. The reason that the LLH-based identification cannot profit from considering the  $\pi^0$  variables can be tracked down to their small separation power between taus and jets.



**Figure 6.4:** Signal efficiency against the average number of interactions of bunch crossing,  $\mu$ , based on a LLH-based tau identification for 1-prong (left) and multi-prong (right) tau decays. The results for the variable set without  $\pi^0$  variables is represented by the filled markers while the empty markers stand for the variable set including those variables. The three working points loose, medium and tight are depicted for both classifiers as green triangles, blue squares and red triangles, respectively.



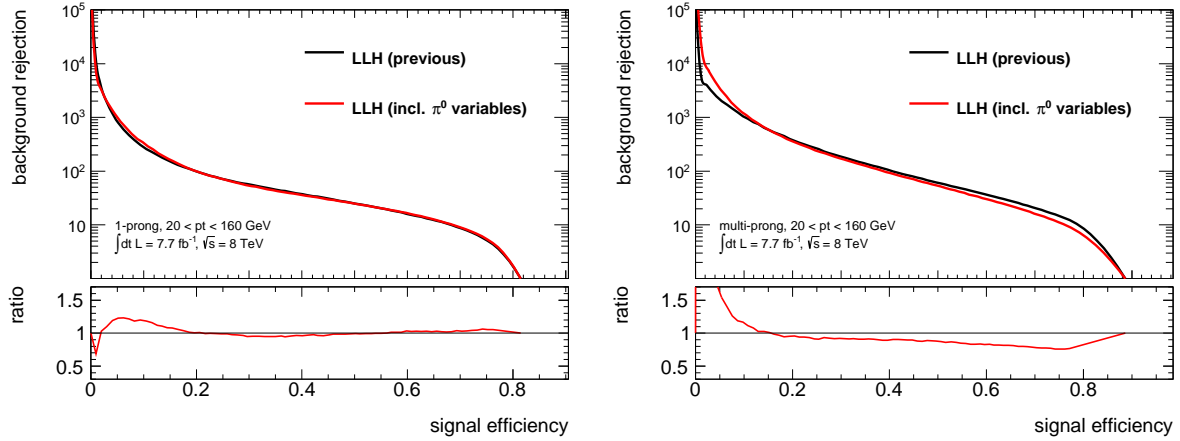
**Figure 6.5:** LLH (left) and BDT (right) signal scores against  $p_T$  for multi-prong tau decays. The fraction of tau candidates for a certain score- $p_T$  pair is indicated by a colour scale.

Exemplary, the LLH values for the variables core energy fraction,  $f_{\text{core}}^{\text{corr}}$ , and the visible mass of the tau candidate,  $m_{\tau}^{\text{vis}}$ , are depicted in Figure 6.7 for 1-prong tau candidates.  $m_{\tau}^{\text{vis}}$ , as one of the  $\pi^0$  variables, leads to an almost complete overlapping distribution of signal and background events, which results finally in no gain in performance and even a loss is possible. Whereas,  $f_{\text{core}}^{\text{corr}}$  is one of the variables already included in the previous variable set and shows a good separation as supposed. The BDT on the other hand has still the possibility to gain advantage from the correlations of the  $\pi^0$  variables with the others, and hence might gain from this aspect. Nevertheless, it is reasonable to consider the neutral pion based variables to allow a comparison of the BDT and LLH approach.

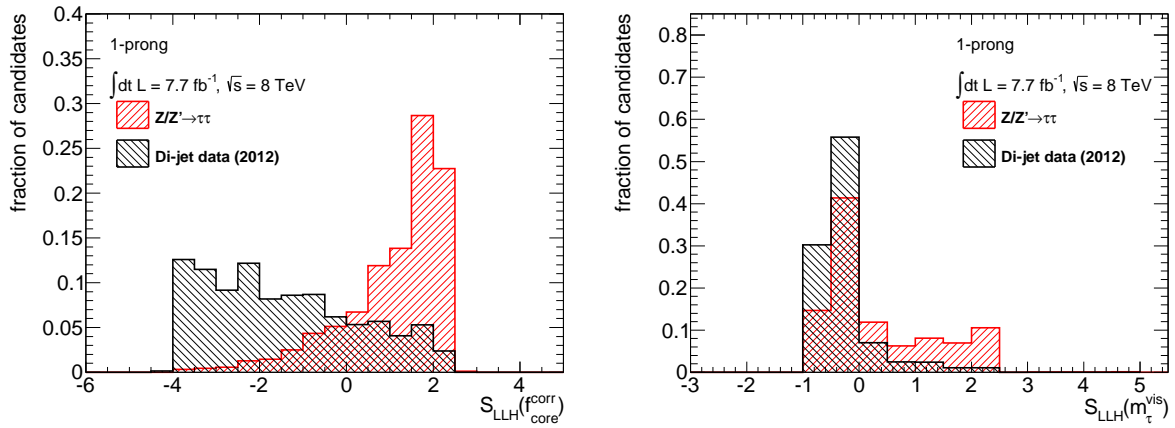
### Event Reweighting

The used signal and background events, extracted from MC simulations and a QCD di-jet selection in 2012 data, do not cover the same kinematic region, and thus initiate the implementation of weights on the transverse momentum,  $p_T$ , and the average number of interactions per bunch crossing,  $\mu$ . The corresponding distributions are presented in Figure 6.8 for 1-prong tau decays. The distribution for QCD





**Figure 6.6:** Background rejection against signal efficiency for a LLH-based tau identification exploiting a set of discriminating variables with (red) and without (black)  $\pi^0$  quantities. The distributions are shown for 1-prong (left) and multi-prong (right) tau candidates. The ratios below each plot depict the according deviations of the background rejection for a given signal efficiency w.r.t. the default approach.

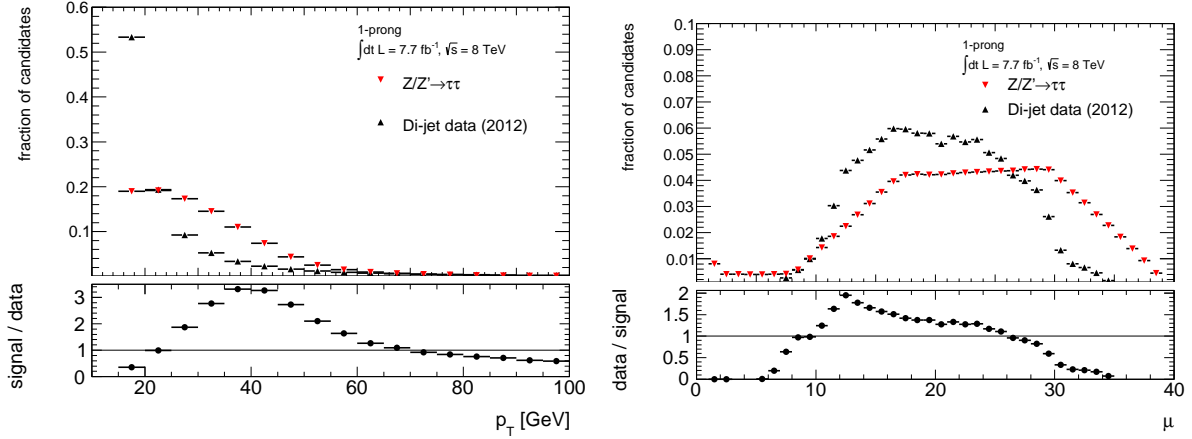


**Figure 6.7:** LLH values of 1-prong tau candidates for the variables  $f_{\text{core}}^{\text{corr}}$  (left) and  $m_{\tau}^{\text{vis}}$  (right). The distributions of signal events obtained from MC simulation with tau candidates required to match to generated 1-prong decays are illustrated as red dashed histograms. The black dashed histograms represent the distribution of background events extracted from a QCD di-jet selection in data.

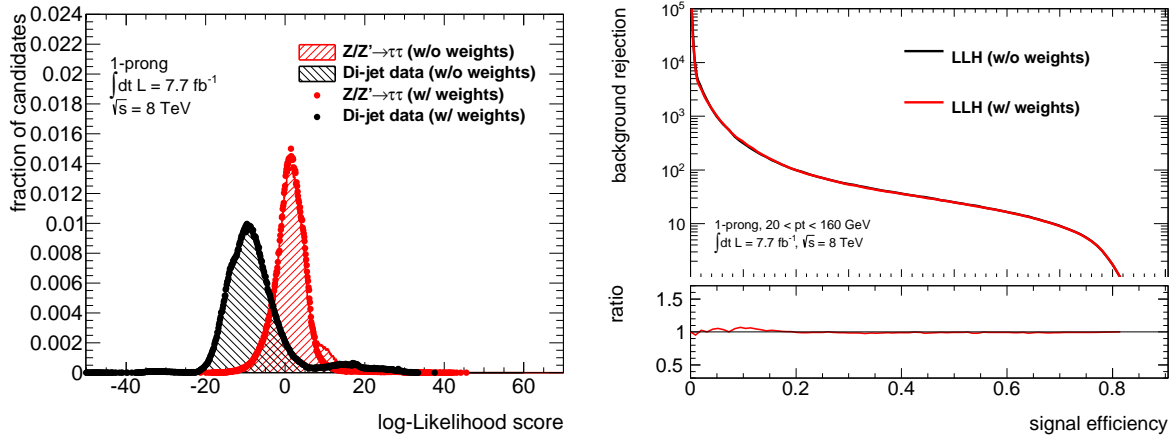
di-jets decrease exponentially with increasing  $p_T$ , while the spectra extracted from MC generated  $Z \rightarrow \tau\tau$  and  $Z' \rightarrow \tau\tau$  events reflects the resonance structure. The  $p_T$  dependence is not explicitly integrated in the identification algorithms, such that fake tau candidates in data with low  $p_T$  are enhanced leading to a bias in the considered variable space. To prevent a potential bias a weight harmonising the  $p_T$  dependencies needs to be applied. Moreover, the  $\mu$  distribution in MC events is generated in a generic way beforehand of data-taking. Consequently, the particular 2012 run conditions were not considered. Hence, the dependence differ for both and a correction on MC based events is necessary to account for this. On the one hand this guarantees a more realistic reflection of the dependence on the average number of interactions per bunch crossing. On the other hand this requires the application of the  $p_T$  weights on the background sample, such that the determination of cuts on the score yielding a flat signal efficiency against the generated visible  $p_T$  is not distorted. Accordingly, the correction on the transverse momentum distribution is performed on the events extracted from data. The calculated  $p_T$  and  $\mu$  weights are presented in Figure 6.8. The effects on the LLH-based tau identification are illustrated in Figure 6.9



for the score distribution and the resulting background rejection against signal efficiency for 1-prong tau candidates. The observed deviations by applying the event reweighting are almost negligible.



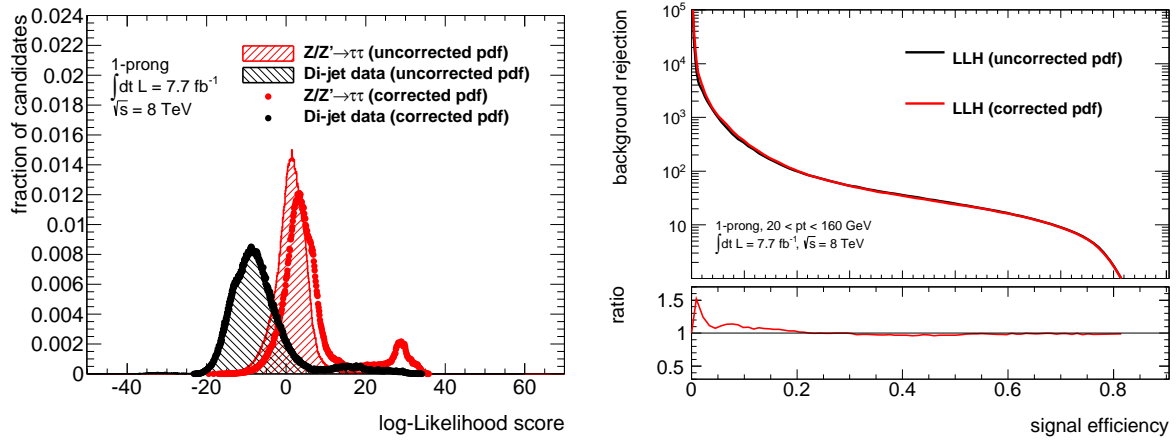
**Figure 6.8:**  $p_T$  (left) and  $\mu$  (right) distributions of 1-prong tau candidates extracted from  $Z \rightarrow \tau\tau$  and  $Z' \rightarrow \tau\tau$  MC simulation (red) and a QCD di-jet selection in data (black). The calculated weights are represented below the corresponding distributions.



**Figure 6.9:** The LLH score distribution (left) for 1-prong tau candidates before and after considering  $p_T$  and  $\mu$  corrections is illustrated as dashed histograms and filled dots, respectively. The distributions of signal events obtained from MC simulation with tau candidates required to match to generated 1-prong decays are illustrated in red and the background events extracted from a QCD di-jet selection in data in black. The resulting background rejection against signal efficiency (right) is depicted for an identification before (black) and after (red) an application of the  $p_T$  and  $\mu$  reweighting. The ratio below depicts the according deviations of the background rejection for a given signal efficiency w.r.t. the default approach.

### Correction of the Calculation of Probability Density Functions

Within the scope of the detailed study of the log-Likelihood approach it was revealed that the calculation of the p.d.f.s was incorrect. They have not been normalised properly, and hence the definition of a p.d.f. was not fulfilled. Consequently, the resulting LLH score is sensitive to statistical differences between the events taken for signal and background. Nevertheless, there was no significant change observed since the considered statistics of signal and data events is comparable. This is confirmed e.g. by the distributions for the LLH scores and the background rejection against the signal efficiency, which are



**Figure 6.10:** The LLH score distributions (left) for 1-prong tau candidates before and after correction of the p.d.f. calculation are illustrated as dashed histograms and filled dots, respectively. The distribution of signal events obtained from MC simulation with tau candidates required to match to generated 1-prong decays is illustrated in red and the background events extracted from a QCD di-jet selection in data in black. The resulting background rejection against signal efficiency (right) is depicted for an identification before (black) and after (red) the correction of the p.d.f. calculation. The ratio below depicts the according deviations of the background rejection for a given signal efficiency w.r.t. the default approach.

presented in Figure 6.10. Thus, analyses which exploited the incorrect LLH-based tau identification, were not affected by this issue. Nevertheless, this cannot be guaranteed in the future, and hence a fix for this issue has been implemented and will be used throughout the rest of this chapter.

## 6.2 Substructure Based Tau Identification

One approach to extend the tau identification for future applications, e.g. Higgs CP studies, is to develop an identification based on the tau decay mode classification. This is the main topic of this thesis and will be discussed in the following sections. Several variable sets based on the tau decay substructure are explored. This includes a variable set based on the default one, cf. Section 6.2.1, but also extensions by new defined variables to analyse the nature of tau decays covered in Section 6.2.2 and 6.2.3. Thus, various variable set are exploited by a BDT and LLH classifiers and their performance is evaluated. The investigation and development of a pure substructure based identification is performed exploiting the CellBased algorithm, but for completeness the EflowRec algorithm is considered in addition. Both are combined with the PanTau algorithm introduced in Section 5.2 but are conveniently referred here with the name of the base algorithms.

### 6.2.1 Recalculation of Default Variables

A first step on the way to a pure substructure based tau identification, is to recalculate the default variable set defined in Section 5.3.2 and 5.3.1, exploiting the substructure algorithms. Calorimeter based variables serve as input for the charged and neutral components of the identification variables. Since those are only defined within  $\Delta R \leq 0.2$  w.r.t. the intermediate tau axis, the tracks from charged particles as well as calorimeter entries from neutral particles have to be considered. In the latter case it has to be taken into account that not only neutral pions but also neutral kaons are included. Consequently, it is unavoidable to deviate from the original definition of isolation variables. A representative selection of three of the eleven standard variables is depicted in Figure 6.11 for the recalculation based on the CellBased and EflowRec algorithm. The distributions of all relevant variables can be found in Appendix D.1. The variables  $S_{\text{lead track}}^{\text{IP}}$  and  $S_{\text{T}}^{\text{flight}}$  are also considered as defined in Section 5.3.1 since the required information is

neither provided by the CellBased nor by the EflowRec algorithm.

In general, depending on the substructure algorithm, the signal and background distributions differ more or less from each other in comparison to the default distribution. Some variables, like  $\Delta R_{\max}$ , can be recalculated with the output of the substructure algorithms without any or only very little deviations from the distributions of the default variables calculated from cells and clusters in the reconstruction step. However, variables which are defined as energy fractions suffer from the coarser granularity given by the fact that not cells, but reconstructed pions and kaons enter the calculation. Thus, a loss in resolution can be observed and strong peaking distributions are the result, as depicted for  $f_{\text{core}}^{\text{corr}}$  in Figure 6.11 (top right). Such effects provoke that signal and background distributions are more similar for the respective variable and finally could lead to negative influences in the performance of the actual identification. Also for variables with a neutral component a substructure based recalculation turned out to be difficult, since the spectra of the number of neutral pions expands to higher values (cf. Figure 6.11 (bottom left)) which influences various other variables. One of the affected variables is  $f_{\text{vis-}p_T}$ , also shown in Figure 6.11 (bottom right). The impact of the additionally reconstructed pions is two-fold. On the one hand, neutral pions from tau decays which the default  $\pi^0$  reconstruction misses get recovered, resulting in  $f_{\text{vis-}p_T}$  values closer to one. On the other hand, fake neutral pions might be reconstructed by the substructure algorithms resulting in  $f_{\text{vis-}p_T}$  values larger than one. For QCD jets the CellBased algorithm reconstructs more jets with zero neutral pions resulting in an enhancement at low  $f_{\text{vis-}p_T}$ . This leads to a larger separation, which is likely to improve the performance. The EflowRec algorithm reconstructs more neutral pions than the default  $\pi^0$  reconstruction. This reflects in a shift in  $f_{\text{vis-}p_T}$  towards larger values, and thus lower separation between signal and background. Depending on the importance of  $N_{\pi^0}$  and  $f_{\text{vis-}p_T}$  this will either lead to a better or worse performance.

The resulting variable set is taken as input for the BDT- as well as the LLH-based tau identification described in Section 5.3.2, to verify the impact of the deviations in the calculation on the performance of the identification algorithms. Figure 6.12 and 6.13 summarise the background rejections against the signal efficiency for 1-prong and multi-prong tau candidates after the evaluation of the two identification techniques exploiting the three variable sets corresponding to the different  $\pi^0$  reconstruction algorithms.

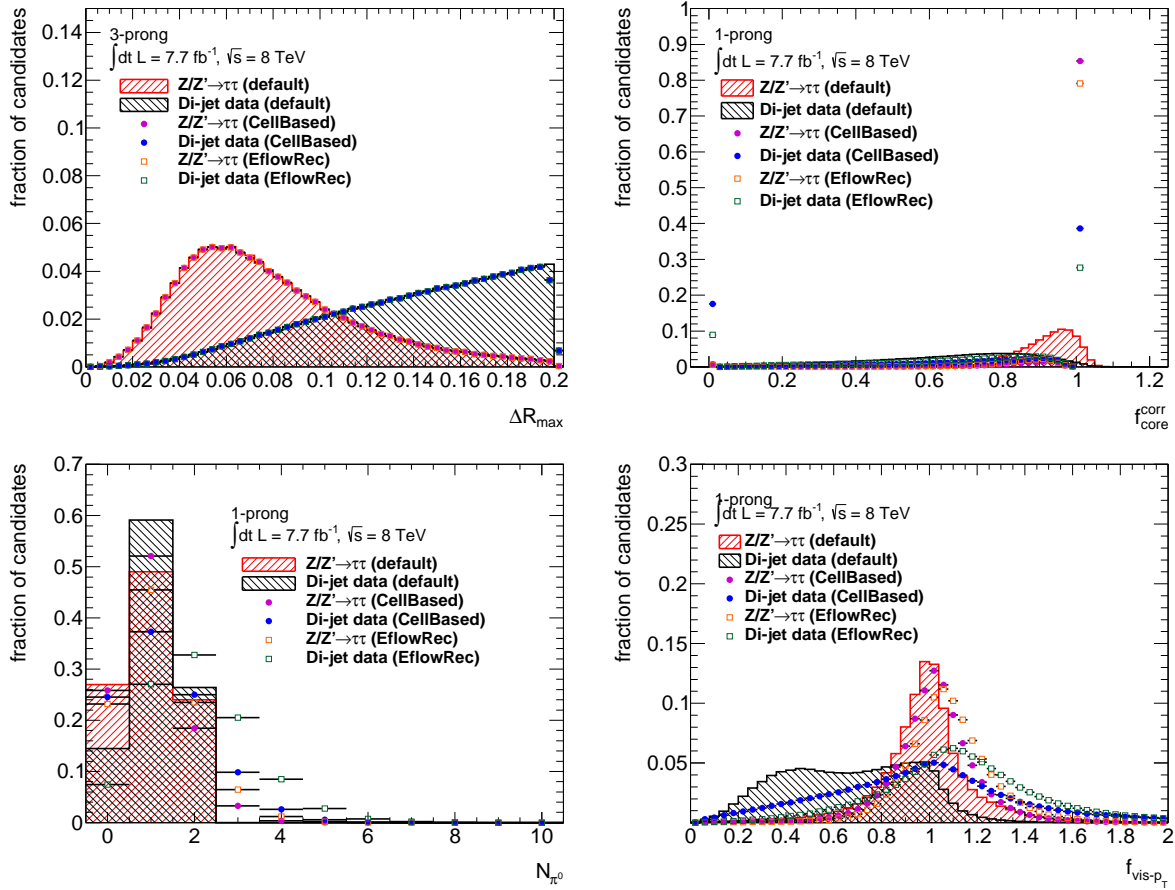
For the BDT as well as for the LLH a huge performance loss in order of a factor of two to three in the interesting signal efficiency range for the approaches considering the variable set based on the substructure algorithm w.r.t. the default variable set can be stated. As mentioned above this can be explained by the coarser granularity which degrades the separation of the very important energy fractions. Comparing the  $\pi^0$  reconstruction algorithms, one can see that the EflowRec outperforms the CellBased tau identification, though the difference is rather small. Hence, it is essential to find a new strategy to construct powerful variables based on the tau decay substructure which can address these issues.

### 6.2.2 Investigation of Isolation Cone Variables

The previous section showed that the loss in granularity results in a loss in performance for the tau identification exploiting the variable set recalculated with the output of the CellBased and EflowRec algorithms. In order to recover the performance the affected variables need to be replaced by similar quantities which do not suffer from this issue. With this aspiration the variable calorimeter radius,  $R_{\text{cal}}$ , is defined as the sum of  $p_T$  weighted distances w.r.t. the tau axis of the reconstructed pions according to:

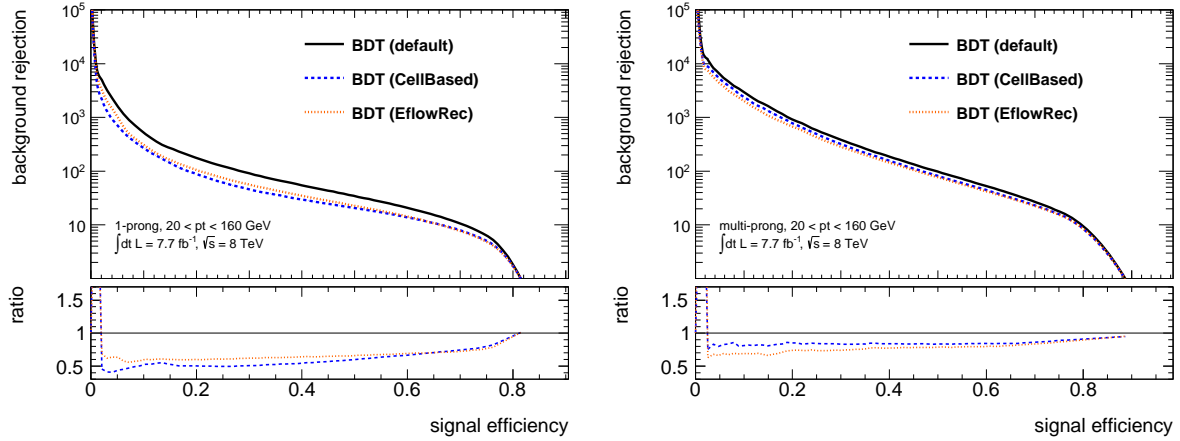
$$R_{\text{cal}}^{\text{core(iso)}} = \frac{\sum_{i \in \{\pi^\pm, \pi^0\}}^{\Delta R < 0.2(0.4)} p_{T,i} \cdot \Delta R_i}{\sum_{i \in \{\pi^\pm, \pi^0\}}^{\Delta R < 0.2(0.4)} p_{T,i}}, \quad (6.1)$$

where  $i$  runs over all charged and neutral pions within the core (isolation) cone. The information provided by this variable is similar to the core energy fraction, and thus could serve as a replacement. Figure 6.14 and 6.15 present the distributions of  $R_{\text{cal}}$  for 1- and 3-prong tau candidates within the core and isolation cone, respectively. Since the intermediate tau axis is calculated from all cells within  $\Delta R < 0.2$ , while the pion reconstruction only considers the cells near by the energy deposit, the distribution does not peak at zero but is slightly shifted toward higher values for 1p0n decays.  $R_{\text{cal}}$  provides for both, core and

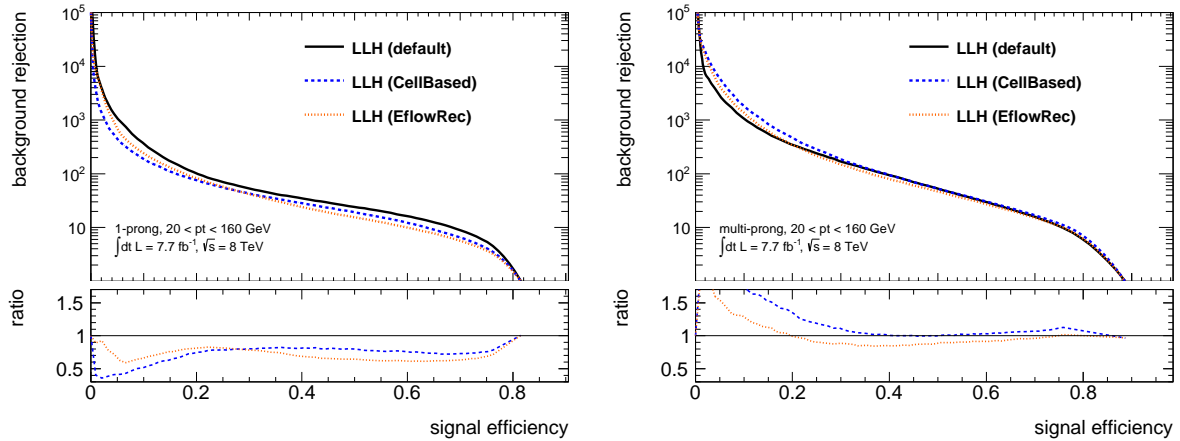


**Figure 6.11:**  $\Delta R_{\max}$  (top left) for reconstructed 3-prong tau candidates and  $f_{\text{core}}^{\text{corr}}$  (top right),  $N_{\pi^0}$  (bottom left) and  $f_{\text{vis-p}_T}$  (bottom right) for reconstructed 1-prong tau candidates. The red and black dashed histograms illustrate signal events obtained from MC simulation with tau candidates required to match to generated 1-/3-prong decays and background events extracted from a QCD di-jet selection in data, respectively, based on the default calculation. Signal and background distribution are indicated by the magenta and blue filled dots for the recalculation based on the CellBased algorithm and as orange and green empty squares for the EflowRec algorithm.

isolation cone, a good separation of signal and background events and does not have a strong peaking behaviour like  $f_{\text{core}}^{\text{corr}}$ . Hence, the core energy fraction is replaced by the calorimeter radius and the new variable set serves as baseline for the tau identification, where  $R_{\text{cal}}^{\text{core}}$  and  $R_{\text{cal}}^{\text{iso}}$  are considered separately. The results of this, i.e. the background efficiency against signal efficiency, are summarised in Figure 6.16 and 6.17 for the BDT approach based on the CellBased and EflowRec algorithm. The consideration of  $R_{\text{cal}}$  clearly leads to an improvement of the performance for both featured substructure algorithms as well as for the two considered tau decay prong modes, in comparison to the one obtainable with the recalculated variable set. By exploiting the EflowRec algorithm the gain is even higher than achievable by the CellBased algorithm as already implied in Section 6.2.1. Taking the core calorimeter radius into account results in a significant enhancement of up to 10 % w.r.t. to the one of the recalculated variable set. Extending the calorimeter radius to the isolation cone leads to a performance close to the one based on the default variable set, especially in the region of moderate signal efficiencies. Nevertheless, in the 1-prong case a reduction in background rejection of up to 20 % for the CellBased approach and about 10 % for the EflowRec algorithm is still present for signal efficiencies between 30 % and 80 %. Whereas for multi-prong tau decays the results for both substructure algorithms are slightly better than the default strategy.



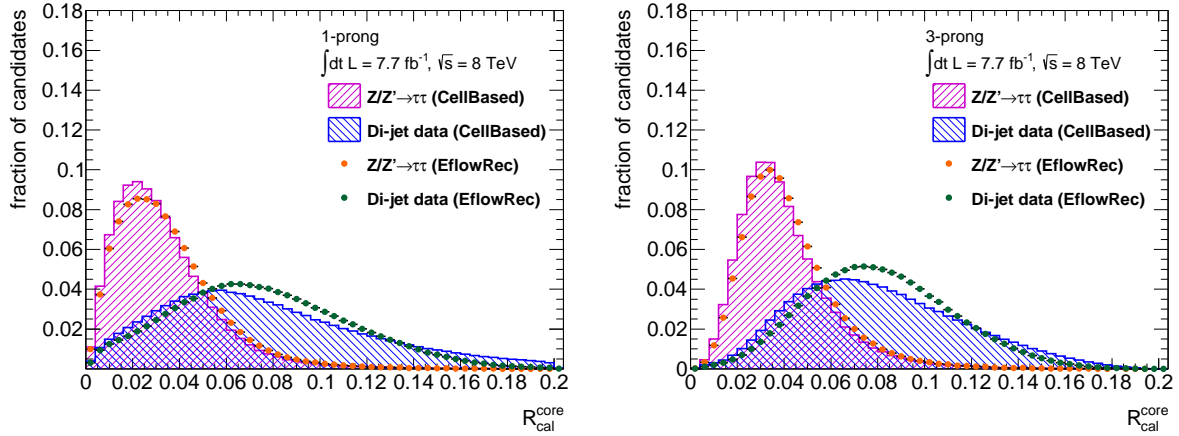
**Figure 6.12:** Background rejection against signal efficiency for a BDT-based tau identification exploiting the default variable set (black) and variable sets recalculated with the substructure algorithms, CellBased (blue) and EflowRec (orange). The performance of the evaluation on 1-prong and multi-prong tau decays is shown in the left and right figure, respectively. The ratios below each plot depict the according deviations of the background rejection for a given signal efficiency w.r.t. the default approach. The same colour scheme is applied.



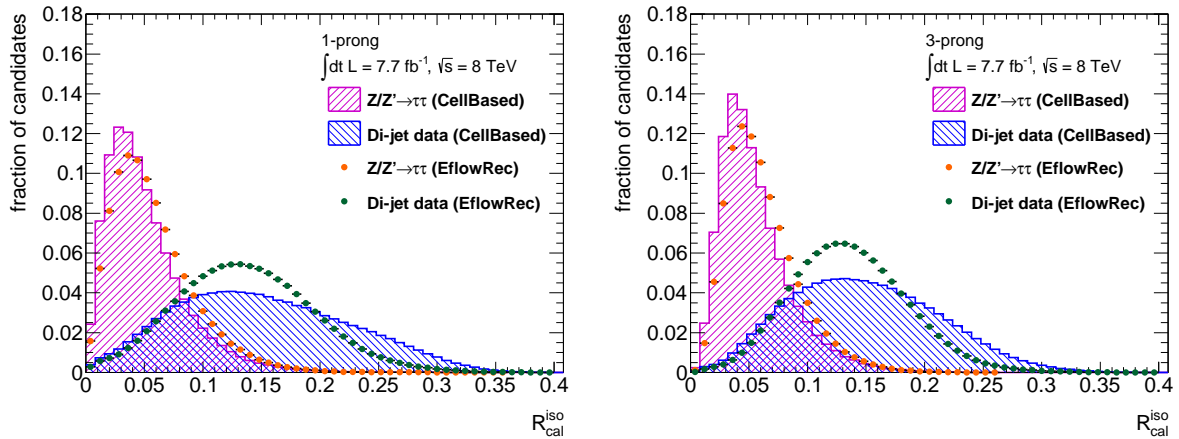
**Figure 6.13:** Background rejection against signal efficiency for a LLH-based tau identification exploiting the default variable set (black) and variable sets recalculated with the substructure algorithms, CellBased (blue) and EflowRec (orange). The performance of the evaluation on 1-prong and multi-prong tau decays is shown in the left and right figure, respectively. The ratios below each plot depict the according deviations of the background rejection for a given signal efficiency w.r.t. the default approach. The same colour scheme is applied.

Also for the LLH-based tau identification an improvement in performance can be observed thanks to the information provided by  $R_{\text{cal}}$ . However, the deviations between the default and substructure based recalculated variable sets are not as high as for the BDT strategy. On the one hand the information loss due to the coarser granularity affects both approaches equally. On the other hand only the boosted decision tree uses the correlations between the discriminating variables, such that it has an effect on the BDT but not on the log-Likelihood. Figure 6.18 and 6.19 summarise the performance of the LLH-based tau identification exploiting the CellBased and the EflowRec algorithm. As already implied by the results of the BDT approach, it is not possible to reach or outperform the performance of the default variable set for reconstructed 1-prong tau candidates by consideration  $R_{\text{cal}}^{\text{core}}$  or  $R_{\text{cal}}^{\text{iso}}$ . Only in the case of the CellBased variable set including  $R_{\text{cal}}^{\text{iso}}$  the performance is close to the default one and exceed it for the loose working point (40 %) and differs only by about 5 % for higher signal efficiencies. With the





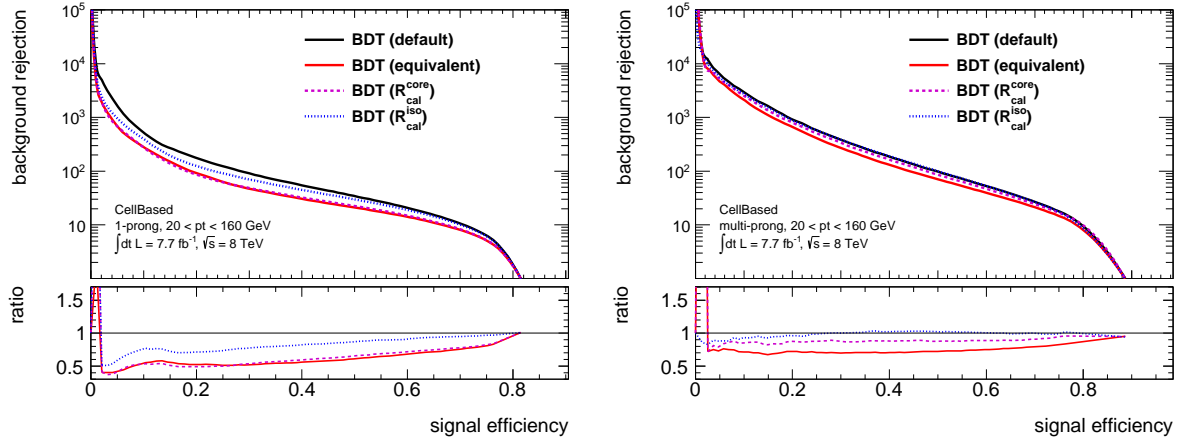
**Figure 6.14:** Core calorimeter radius,  $R_{\text{cal}}^{\text{core}}$ , for reconstructed 1-prong (left) and 3-prong (right) tau candidates. For a calculation based on the CellBased algorithm the distribution of signal events obtained from MC simulation with tau candidates required to match to generated 1-/3-prong decays is illustrated as dashed magenta histogram and the distribution of background events extracted from a QCD di-jet selection in data as dashed blue histogram. This is indicated in orange and green for the EflowRec algorithm.



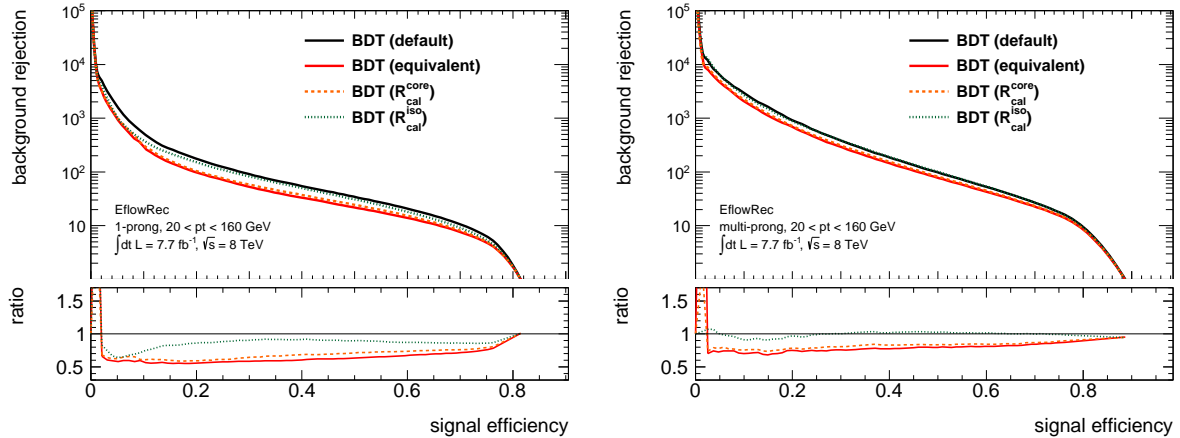
**Figure 6.15:** Isolation calorimeter radius,  $R_{\text{cal}}^{\text{iso}}$ , for reconstructed 1-prong (left) and 3-prong (right) tau candidates. For a calculation based on the CellBased algorithm the distribution of signal events obtained from MC simulation with tau candidates required to match to generated 1-/3-prong decays is illustrated as dashed magenta histogram and the distribution of background events extracted from a QCD di-jet selection in data as dashed blue histogram. This is indicated in orange and green for the EflowRec algorithm.

EflowRec variable set this can be achieved, too, though only in the low signal efficiencies region around 20 %, which is of low interest for typical application scenarios of tau identification. For multi-prong tau decays it was already possible to recover the performance of the performance of the default variable set by exploiting both substructure based variable set including  $R_{\text{cal}}^{\text{iso}}$ . This is even exceeded in the case of the LLH-based tau identification. Hence, a performance boost of more than 20 % can be achieved for the interesting signal efficiency range between 30 % and 80 %.

Following those results, a consideration of the isolation variable  $R_{\text{cal}}^{\text{iso}}$  is unavoidable, though this comes at the cost of an increased dependence of the efficiency on the number of pile-up events. This is reflected in Figure 6.20 which illustrates the signal efficiency w.r.t. the average number of interactions per bunch crossing for the BDT and LLH algorithm, respectively. The signal efficiencies for the three working points are determined via a  $p_T$  dependent cut on the individual score, such that the extracted



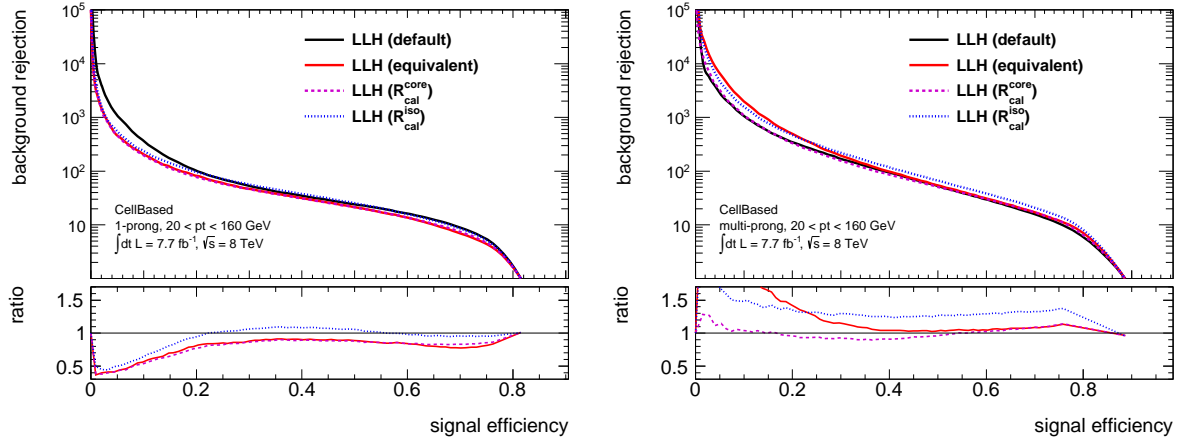
**Figure 6.16:** Background rejection against signal efficiency for a BDT-based tau identification exploiting the default (black), recalculated (red) variable sets and a recalculated variable set including  $R_{\text{cal}}^{\text{core}}$  (magenta dashed) or  $R_{\text{cal}}^{\text{iso}}$  (blue dotted) based on the substructure algorithm CellBased. The performance for the evaluation on 1-prong and multi-prong tau decays are shown in the left and the right figure, respectively. The ratios below each plot depict the according deviations of the background rejection for a given signal efficiency w.r.t. the default approach. The same colour scheme is applied.



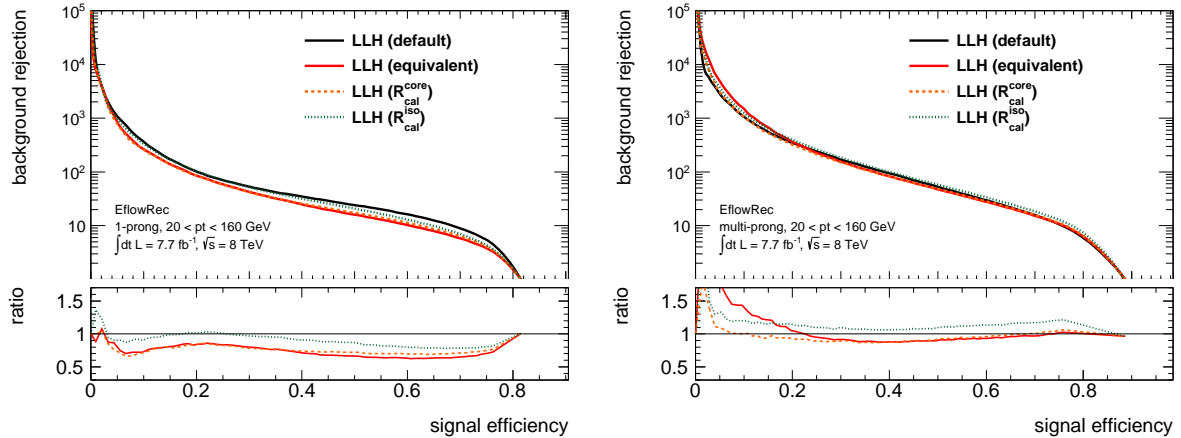
**Figure 6.17:** Background rejection against signal efficiency for a BDT-based tau identification exploiting the default (black), recalculated (red) variable sets and a recalculated variable set including  $R_{\text{cal}}^{\text{core}}$  (orange dashed) or  $R_{\text{cal}}^{\text{iso}}$  (green dotted) based on the substructure algorithm EflowRec. The performance for the evaluation on 1-prong and multi-prong tau decays are shown in the left and the right figure, respectively. The ratios below each plot depict the according deviations of the background rejection for a given signal efficiency w.r.t. the default approach. The same colour scheme is applied.

signal efficiencies are flat w.r.t.  $p_{T,\text{vis}}^{\text{gen}}$  (cf. Section 5.3.2). Only the efficiency evaluated based on the CellBased algorithm is shown. The one for the EflowRec algorithm can be found in Appendix D.2. For both underlying multivariate techniques a large drop with increasing  $\mu$  can be observed. To prevent this pile-up dependence a so-called pile-up correction is developed and applied on the variable. Therefore, a linear fit is performed on the mean of the signal variable distribution against  $\mu$  over the entire  $\mu$  range. Its slope serves as a correction factor  $f$ . A linear correction term including this factor of the form  $f \cdot \mu$  is added to the variable distribution. The outcome is depicted in Figure 6.21 for the calorimeter radius of reconstructed 1-prong tau decays within the isolation cone. The utilisation of pile-up correction flattens the signal distribution of  $R_{\text{cal}}$  for the two considered tau decay prong modes. This is also reflected



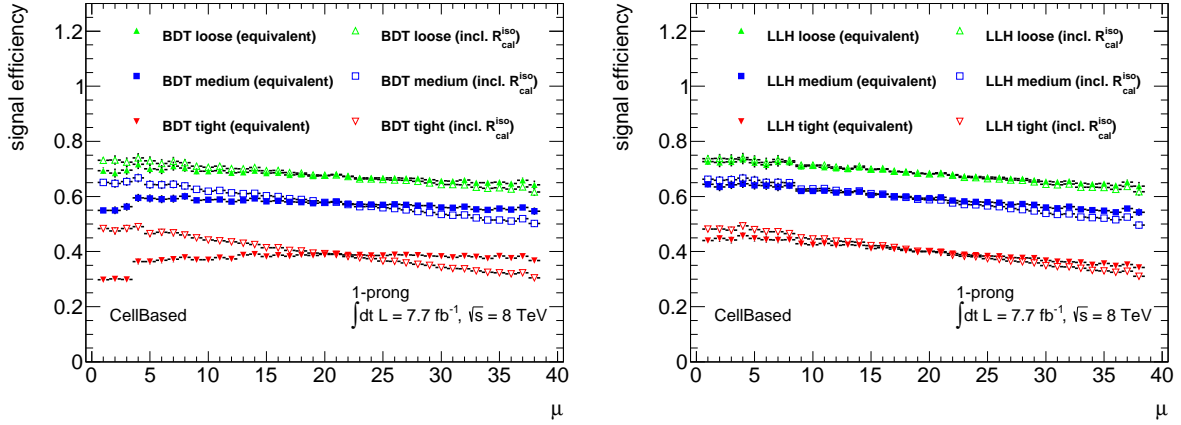


**Figure 6.18:** Background rejection against signal efficiency for a LLH-based tau identification exploiting the default (black), recalculated (red) variable sets and a recalculated variable set including  $R_{\text{cal}}^{\text{core}}$  (magenta dashed) or  $R_{\text{cal}}^{\text{iso}}$  (blue dotted) based on the substructure algorithm CellBased. The performance for the evaluation on 1-prong and multi-prong tau decays are shown in the left and right figure, respectively. The ratios below each plot depict the according deviations of the background rejection for a given signal efficiency w.r.t. the default approach. The same colour scheme is applied.

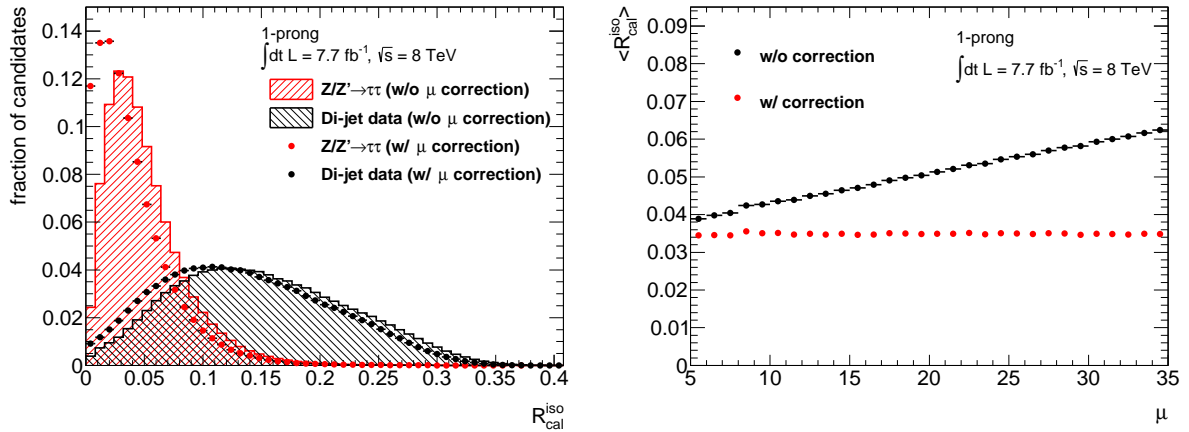


**Figure 6.19:** Background rejection against signal efficiency for a LLH-based tau identification exploiting the default (black), recalculated (red) variable sets and a recalculated variable set including  $R_{\text{cal}}^{\text{core}}$  (orange dashed) or  $R_{\text{cal}}^{\text{iso}}$  (green dotted) based on the substructure algorithm EflowRec. The performance for the evaluation on 1-prong and multi-prong tau decays are shown in the left and right figure, respectively. The ratios below each plot depict the according deviations of the background rejection for a given signal efficiency w.r.t. the default approach. The same colour scheme is applied.

in the distribution of the signal efficiency against  $\mu$  as confirmed in Figure 6.22 for both identification algorithms for 1-prong tau candidates. The flattening w.r.t. the average number of interactions per bunch crossing has a negative impact on the performance for the identification of 1-prong taus as confirmed in the left plots of Figure 6.23 and 6.24. Especially for the variable set including  $R_{\text{cal}}^{\text{iso}}$  a performance loss can be stated. Hence, the background rejection for the three considered substructure variable sets can only achieve about 70 % for the BDT approach and 80 % for the LLH strategy over the relevant signal efficiency range. For multi-prong tau decays the resulting performance is slightly better for the variable sets including  $R_{\text{cal}}$  after the reweighting is applied, and thus also more efficient than the default strategy. The background rejection resulting from a LLH-based tau identification exploiting a variable



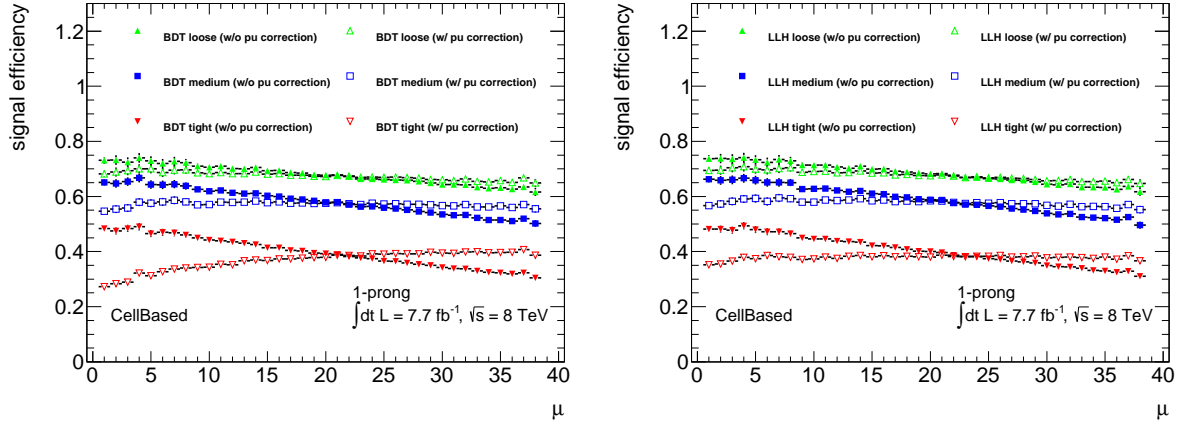
**Figure 6.20:** Signal efficiency against the average number of interactions per bunch crossing for 1-prong tau candidates. The BDT- (left) and the LLH-based (right) tau identification exploiting the default variable set is represented by the filled and for the CellBased algorithm based variable set including  $R_{\text{cal}}^{\text{iso}}$ , by empty markers. The three working points loose, medium and tight are depicted for both classifiers as green triangles, blue squares and red triangles, respectively.



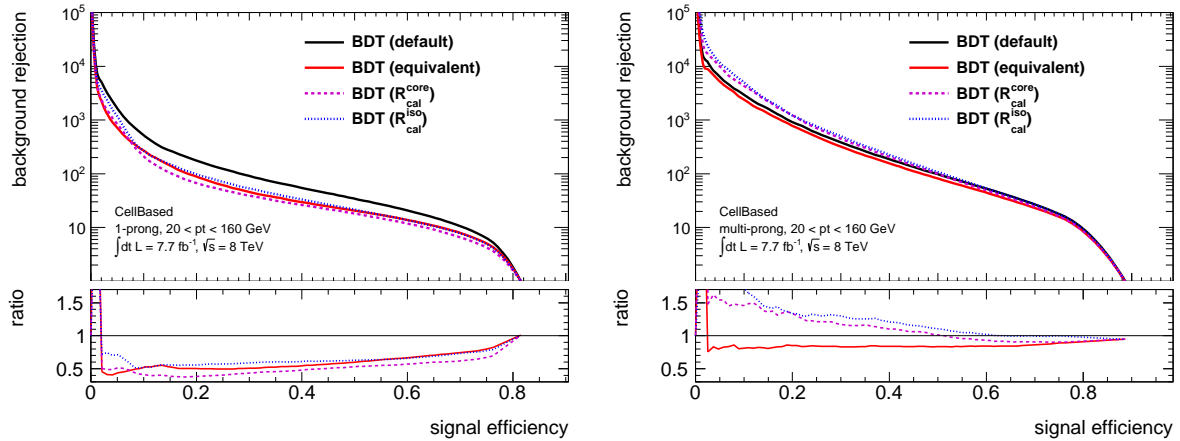
**Figure 6.21:** Calorimeter radius (left) for reconstructed 1-prong tau candidates within the isolation cone. The distribution without pile-up correction is depicted as dashed histograms and with correction by filled dots for signal events obtained from MC simulation with tau candidates required to match to generated 1-prong decays (red) and background events extracted from a QCD di-jet selection in data (black). The corresponding distribution of  $\langle R_{\text{cal}}^{\text{iso}} \rangle$  against  $\mu$  (right) with and without pile-up corrections is illustrated by red and black dots, respectively.

set with  $R_{\text{cal}}^{\text{iso}}$ , is 50 % better than the default one over the interesting signal efficiency range between 30 % and 80 %. In the case of the BDT approach the performance of the same variable set is close to the one obtained by the default strategy for signal efficiencies above 60 %, while an increasing background rejection is observed for low signal efficiencies.

The core energy fraction is not the only variable suffering from the coarser granularity of the substructure calculation, a second variable is the leading track momentum fraction. However,  $f_{\text{track}}$  is not affected as much as  $f_{\text{core}}$  and an expansion to the isolation cone already reduces this effect, such that it might be sufficient to extend instead of replacing it. Figure 6.25 shows the distributions of the leading track momentum in the core,  $f_{\text{lead,chr}}^{\text{core}}$ , and the isolation cone,  $f_{\text{lead,chr}}^{\text{iso}}$ , calculated using reconstructed charged hadrons provided by the CellBased and EflowRec algorithm. It can be seen that the single peak at one is strongly reduced by the expansion to the isolation cone and contains almost no back-

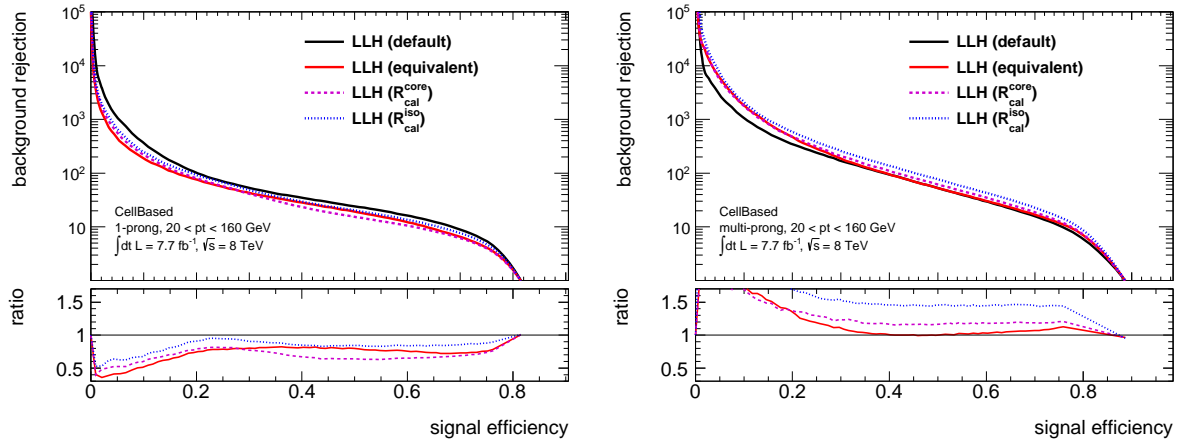


**Figure 6.22:** Signal efficiency against the average number of interactions per bunch crossing for 1-prong tau candidates. The BDT- (left) and the LLH-based (right) tau identification exploiting the CellBased algorithm based variable set including  $R_{\text{cal}}^{\text{iso}}$  without pile-up corrections illustrated by filled markers and with by empty markers. The three working points loose, medium and tight are depicted for both classifiers as green triangles, blue squares and red triangles, respectively.

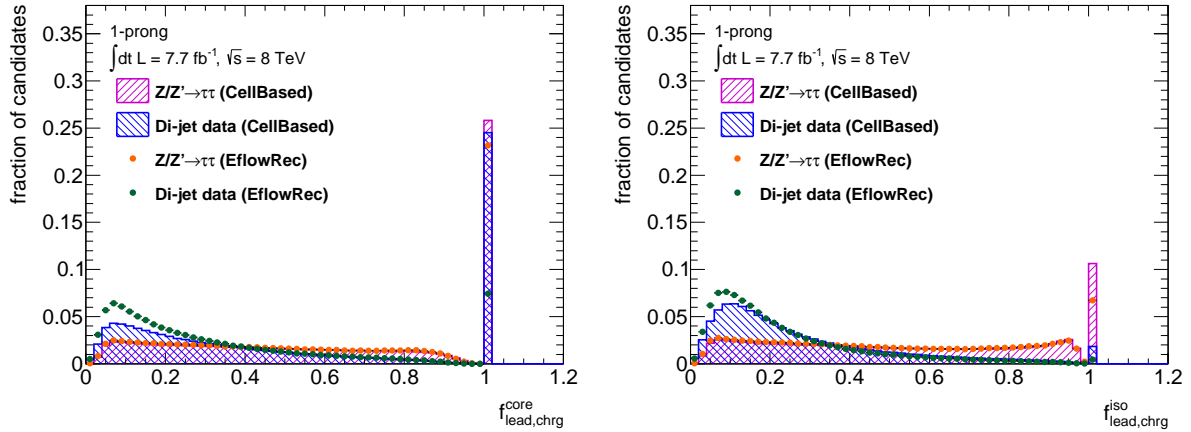


**Figure 6.23:** Background rejection against signal efficiency for a BDT-based tau identification exploiting the default (black), recalculated (red) variable sets and a recalculated variable set including  $R_{\text{cal}}^{\text{core}}$  (magenta dashed) or  $R_{\text{cal}}^{\text{iso}}$  (blue dotted) based on the substructure algorithm CellBased. Whereby, pile-up corrections are taken into account for the variable calculation. The performance for the evaluation on 1-prong and multi-prong tau decays are shown in the left and right figure, respectively. The ratios below each plot depict the according deviations of the background rejection for a given signal efficiency w.r.t. the default approach. The same colour scheme is applied.

ground events, and thus could provide a better separation of real and fake taus. The background rejection against signal efficiency resulting from the variable sets including  $R_{\text{cal}}^{\text{iso}}$  and  $f_{\text{lead, chrg}}^{\text{iso}}$  is shown w.r.t. the default variable set and the one including only  $R_{\text{cal}}^{\text{iso}}$ , in Figure 6.26 and 6.27 for the BDT- and LLH-based identification. Additionally, the performance of a pure isolation variable set is included. This means, all variables recalculated by exploiting the CellBased algorithm are extended to the isolation cone, whereby the energy fraction is replaced again by the calorimeter radius. It can be assumed that the separation power of each individual variable is improved by this extension. For each of the presented variable sets the procedure of pile-up correction introduced above is applied. The pure isolation variable set as

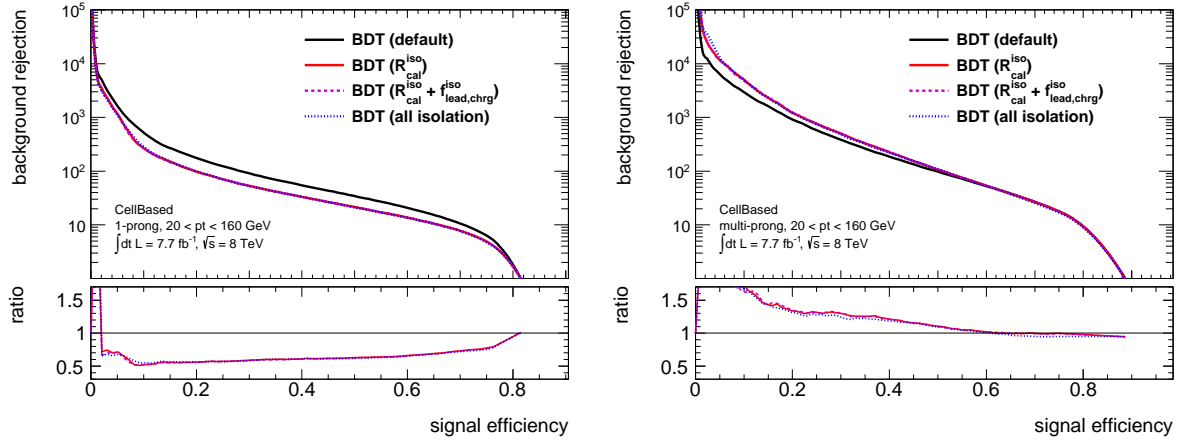


**Figure 6.24:** Background rejection against signal efficiency for a LLH-based tau identification exploiting the default (black), recalculated (red) variable sets and a recalculated variable set including  $R_{\text{cal}}^{\text{core}}$  (magenta dashed) or  $R_{\text{cal}}^{\text{iso}}$  (blue dotted) based on the substructure algorithm CellBased. Whereby, pile-up corrections are taken into account for the variable calculation. The performance for the evaluation on 1-prong and multi-prong tau decays are shown in the left and right figure, respectively. The ratios below each plot depict the according deviations of the background rejection for a given signal efficiency w.r.t. the default approach. The same colour scheme is applied.

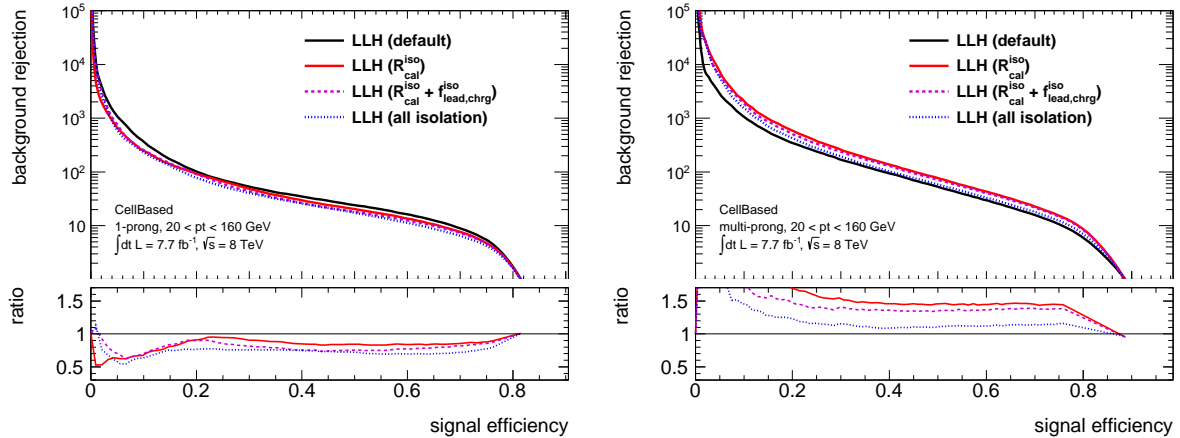


**Figure 6.25:** Leading track momentum fraction within the core,  $f_{\text{lead,charg}}^{\text{core}}$ , (left) and the isolation cone,  $f_{\text{lead,charg}}^{\text{iso}}$ , (right) for reconstructed 1-prong tau candidates. For a calculation based on the CellBased algorithm the distribution of signal events obtained from MC simulation with tau candidates required to match to generated 1-prong decays is illustrated as dashed magenta histogram and the distribution of background events extracted from a QCD di-jet selection in data as dashed blue histogram. This is indicated as orange and green dots for the EflowRec algorithm.

well as the one including  $R_{\text{cal}}^{\text{iso}}$  and  $f_{\text{lead,charg}}^{\text{iso}}$  lead to almost the same background rejection as the variable set considering only  $R_{\text{cal}}^{\text{iso}}$ , over the entire signal efficiency range for the BDT technique. Consequently, it is not possible to improve the performance for 1-prong tau candidates but the one for multi-prong decays is not harmed, such that it is still more efficient than the default one. The LLH approach even results in a worse performance for the two new variable sets in comparison to the one including  $R_{\text{cal}}^{\text{iso}}$ , although their performance is still better than the one obtainable by the default strategy for multi-prong tau decays. The loss can be explained by the increase of pile-up contribution and related necessity of pile-up correction.



**Figure 6.26:** Background rejection against signal efficiency for a BDT-based tau identification exploiting the default (black) variable sets, recalculated variable sets including  $R_{\text{cal}}^{\text{iso}}$  (red),  $R_{\text{cal}}^{\text{iso}}$  and  $f_{\text{lead,chg}}^{\text{iso}}$  (magenta dashed) and a set of exclusively isolation variables (blue dotted) based on the substructure algorithm CellBased. The performance for the evaluation on 1-prong and multi-prong tau decays are shown in the left and right figure, respectively. The ratios below each plot depict the according deviations of the background rejection for a given signal efficiency w.r.t. the default approach. The same colour scheme is applied.



**Figure 6.27:** Background rejection against signal efficiency for a LLH-based tau identification exploiting the default (black) variable sets, recalculated variable sets including  $R_{\text{cal}}^{\text{iso}}$  (red),  $R_{\text{cal}}^{\text{iso}}$  and  $f_{\text{lead,chg}}^{\text{iso}}$  (magenta dashed) and a set of exclusively isolation variables (blue dotted) based on the substructure algorithm CellBased. The performance for the evaluation on 1-prong and multi-prong tau decays are shown in the left and right figure, respectively. The ratios below each plot depict the according deviations of the background rejection for a given signal efficiency w.r.t. the default approach. The same colour scheme is applied.

By using the BDT this can be revoked by exploring the correlations between the discriminating variables.

The studies concerning the pile-up correction and the impact of the partly and complete extension of the variables to the isolation cone is also studied for the EflowRec algorithm and shows similar results, which are presented in Appendix D.2.

### 6.2.3 Extended Variable Sets

A remarkable performance gain is achieved for the substructure based tau identification for 3-prong tau decays by introducing the isolation calorimeter radius, nevertheless not all possibilities are exhausted,

yet. Hence, new variables are defined which are inspired by the characteristics of the tau decay and aim to further exploit its substructure. The following quantities are defined and used for searches for a more powerful set of discriminating variables.

**Number of decay products,  $N$ :** Number of reconstructed decay products in a specific cone around the intermediate tau axis. It will be distinguished between the amount of charged ( $N_{\text{chrg}}^{\text{cent}}, N_{\text{chrg}}^{\text{core}}, N_{\text{chrg}}^{\text{wide}}$ ), neutral ( $N_{\text{neut}}^{\text{cent}}, N_{\text{neut}}^{\text{core}}, N_{\text{neut}}^{\text{wide}}$ ) and all decay products ( $N_{\text{all}}^{\text{cent}}, N_{\text{all}}^{\text{core}}, N_{\text{all}}^{\text{wide}}$ ). The central and core cone, but also the isolation annulus are considered, referred by the indices “cent”, “core” and “wide”, respectively. For real tau decays the number of decay products, i.e. pions and kaons, is well defined and expected to be smaller than for QCD jets. The variable  $N_{\text{all}}^{\text{wide}}$  is equally defined as the number of isolation tracks,  $N_{\text{track}}^{\text{iso}}$ , and  $N_{\text{neut}}^{\text{core}}$  to the number of reconstructed neutral pions,  $N_{\pi^0}$ , both already introduced in Section 5.3.1.

**Mass of the decay products,  $m$ :** Invariant mass of the decay products:

$$m_{\text{chrg/neut/all}}^{\text{cent/core/iso}} = \sqrt{\left( \sum_{i \in \{\pi^\pm\}/\{\pi^0\}/\{\pi^\pm, \pi^0\}}^{\Delta R_i < 0.1/0.2/0.4} E_i \right)^2 - \left( \sum_{i \in \{\pi^\pm\}/\{\pi^0\}/\{\pi^\pm, \pi^0\}}^{\Delta R_i < 0.1/0.2/0.4} \vec{p}_i \right)^2}, \quad (6.2)$$

where  $i$  runs over charged, neutral or all decay products within the central, core or isolation cone as mentioned in Section 5.2. In the latter case also neutral kaons are included. Real taus have a well defined mass, contrary to QCD jets which lead to arbitrary mass values. Hence, the values for real taus peak slightly below the actual tau mass, since only the visible fraction is taken into account, but not the invisible neutrino component. Depending on the actual considered decay products the peak of the distribution of real taus is shifted towards lower values, but nevertheless distinct in contrast to the one of fake taus. The variable  $m_{\text{chrg}}^{\text{iso}}$  is equal to the mass of the track system,  $m_{\text{tracks}}$ , and  $m_{\text{all}}^{\text{core}}$  to the visible mass of the tau candidate,  $m_\tau^{\text{vis}}$  (cf. Section 5.3.1).

**Maximal  $\Delta R$ ,  $\Delta R_{\text{max}}$ :** Maximal distance between the reconstructed charged pions and the intermediate tau axis, referred as  $\Delta R_{\text{max, chrg}}^{\text{core}}$  and  $\Delta R_{\text{max, chrg}}^{\text{iso}}$  for tracks within the core and isolation cone. Real tau decays tend to small values of  $\Delta R_{\text{max}}$  given that they are highly collimated compared to QCD jets which are widespread, and thus higher values are observed. The first is equal to the  $\Delta R_{\text{max}}$  introduced in Section 5.3.1.

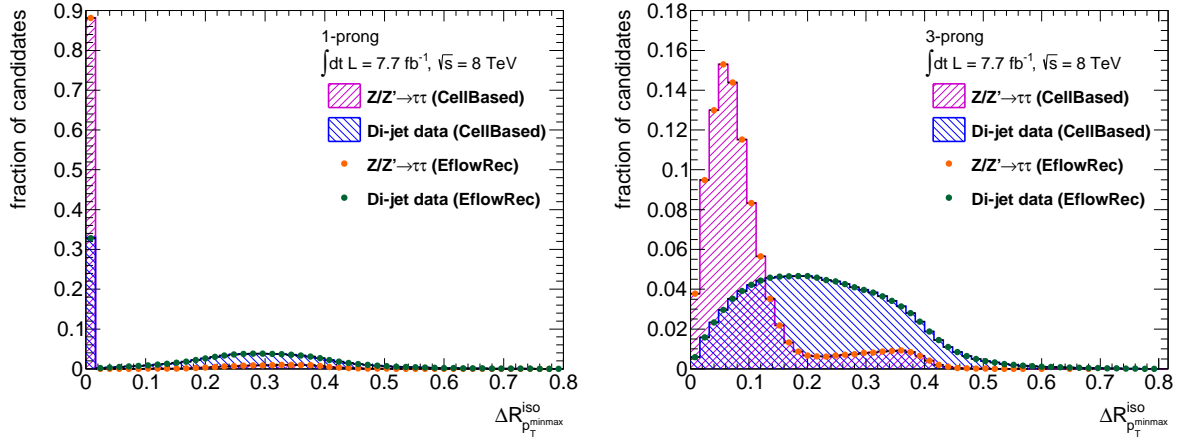
**Distance between low- and high- $p_T$  track,  $\Delta R_{p_T^{\text{minmax}}}^{\text{iso}}$ :** Distance between the two reconstructed tracks with lowest and highest  $p_T$  within the isolation cone. For real 3-prong tau decays a peaking structure is observed at low values while fake taus result in a widespread distribution with the maximum shifted towards higher values. In the 1-prong case only entries for zero should occur for real taus by definition, but contributions originating from pile-up events lead to values differing from zero. Fake taus in general show larger values due to a higher activity in the isolation annulus. Figure 6.28 represents the distributions for 1- and 3-prong tau decays.

**Transverse momentum ratio of the decay products and the tau candidate,  $f_{p_T}$ :** Fraction of the sum of the transverse momenta of the charged, neutral or all decay products and the transverse momenta of the tau candidate,  $p_T^\tau$ :

$$f_{p_T, \text{chrg/neut/all}}^{\text{cent/core/iso}} = \frac{\sum_{i \in \{\pi^\pm\}/\{\pi^0\}/\{\pi^\pm, \pi^0\}}^{\Delta R_i < 0.1/0.2/0.4} p_{T,i}}{p_T^\tau}, \quad (6.3)$$

where  $i$  runs over all considered reconstructed decay products within the central, core or isolation cone.  $f_{p_T, \text{all}}^{\text{core}}$  is similar to  $f_{\text{vis-}p_T}$  defined in Section 5.3.1, but differs slightly because of neutral kaons entering the calculation. These quantities tend to values close to one for real taus, while fake taus result in a broader distribution shifted towards lower values caused by the additional neutral hadrons.





**Figure 6.28:** Distance between the tracks with lowest and highest  $p_T$ ,  $\Delta R_{p_{Tminmax}}^{iso}$ , for reconstructed 1-prong (left) and 3-prong (right) tau candidates. For a calculation based on the CellBased algorithm the distribution of signal events obtained from MC simulation with tau candidates required to match to generated 1-/3-prong decays is illustrated as dashed magenta histogram and the distribution of background events extracted from a QCD di-jet selection in data as dashed blue histogram. This is indicated in orange and green for the EflowRec algorithm.

**$p_T$  weighted  $\Delta R$ ,  $R_{cal}$ :**  $p_T$  weighted distance of the charged, neutral and all decay products w.r.t. the intermediate tau axis,

$$R_{cal,chg/neut/all}^{core/iso} = \frac{\sum_{i \in \{\pi^\pm\}/\{\pi^0\}/\{\pi^\pm, \pi^0\}} \Delta R < 0.2/0.4} \sum_{i \in \{\pi^\pm\}/\{\pi^0\}/\{\pi^\pm, \pi^0\}} p_{T,i} \cdot \Delta R_i}{\sum_{i \in \{\pi^\pm\}/\{\pi^0\}/\{\pi^\pm, \pi^0\}} p_{T,i}}, \quad (6.4)$$

where  $i$  runs over all considered decay products reconstructed within the core or isolation cone. The signal peaks at low values, while the background distribution is spread over a wide range since real tau decays are highly collimated compared to QCD jets.  $R_{cal,all}^{core}$  and  $R_{cal,all}^{iso}$  were already defined in Section 6.2.2 as  $R_{cal}^{core}$  and  $R_{cal}^{iso}$ . Also  $R_{cal,chg}^{iso}$  was already introduced as the default variable  $R_{track}$  in Section 5.3.1.

**Energy fraction,  $f_E$ :** Fraction of the transverse energy,  $E_T$ , within the central and core cone, core and isolation cone or central and isolation cone,

$$f_{E_T,chg/neut/all}^{cent-core/core-iso/cent-iso} = \frac{\sum_{i \in \{\pi^\pm\}/\{\pi^0\}/\{\pi^\pm, \pi^0\}} \Delta R_i < 0.1/0.2/0.1} E_{T,i}}{\sum_{j \in \{\pi^\pm\}/\{\pi^0\}/\{\pi^\pm, \pi^0\}} \Delta R_j < 0.2/0.4/0.4} E_{T,j}}, \quad (6.5)$$

where  $i$  and  $j$  run over charged, neutral or all decay products within the respective cones. As the ratio of  $E_T$  in the central and isolation cone it is equal to the core energy fraction,  $f_{core}$  (cf. Section 5.3.1).

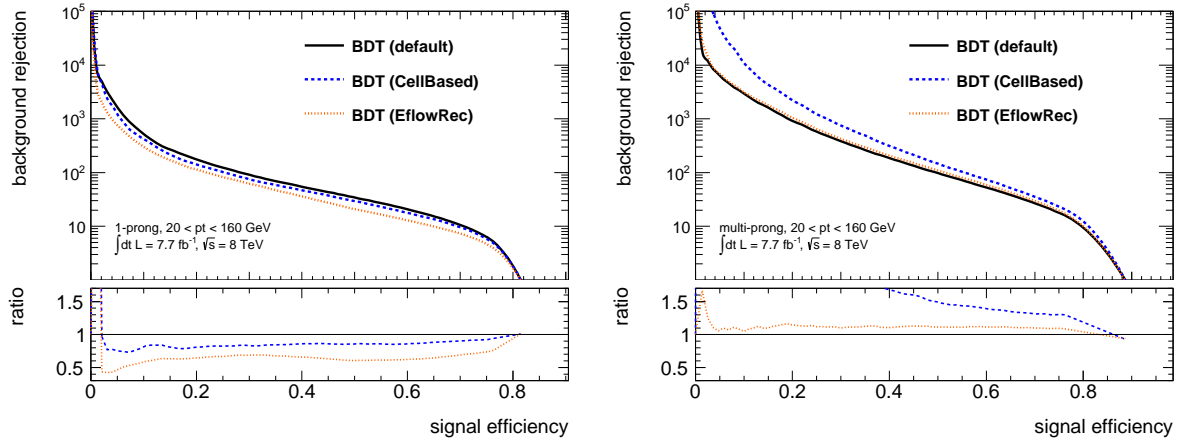
**Leading charged pion momentum fraction,  $f_{lead}$ :** Transverse momentum of the reconstructed charged pion with the highest- $p_T$ ,  $p_{T,lead,chg}$ , w.r.t. the sum of the transverse energy of all decay products,

$$f_{lead,chg}^{cent-core/iso} = \frac{p_{T,lead,chg}^{\Delta R < cent/core/iso}}{\sum_{i \in \{\pi^\pm\}/\{\pi^0\}/\{\pi^\pm, \pi^0\}} E_T^{\Delta R < cent/core/iso}}, \quad (6.6)$$

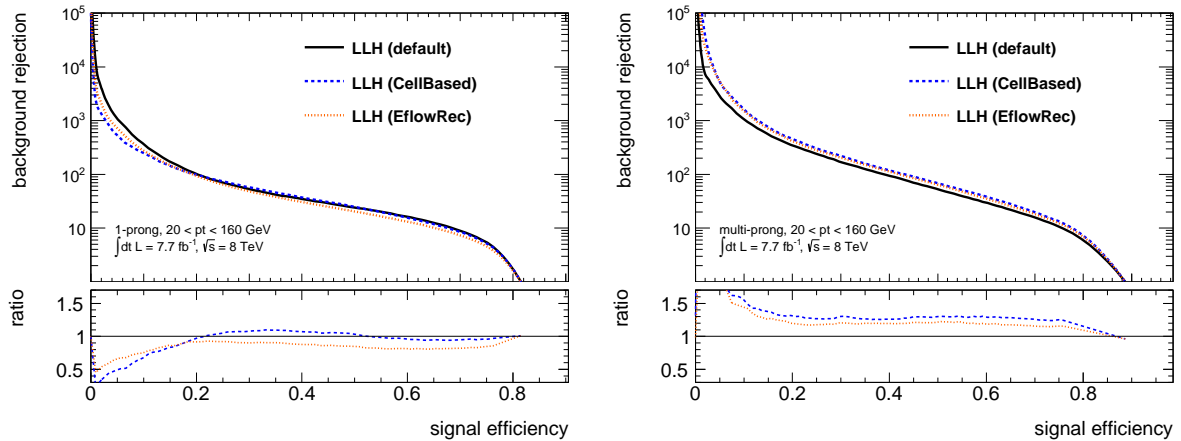
where  $i$  runs over charged, neutral or all decays products within the central, core or isolation cone. Real taus tend to higher values since the amount of decay products is comparatively small. Due to the coarse granularity, a distinct peak at one occurs.  $f_{lead,chg}^{core}$  is equal to the leading track momentum fraction,  $f_{track}$  (cf. Section 5.3.1).



Based on those definitions it is possible to compose a variable set of 38 (43) variables for 1-prong (3-prong) tau candidates. The two variable sets are exploited by the BDT and LLH approach for both substructure algorithms. Figure 6.29 and 6.30 present the corresponding results in terms of the background rejection against signal efficiency. For 1-prong tau decays a better performance w.r.t. the default strategy was not achievable in the case of the BDT-based identification. Furthermore, the background rejection is still about 15 % (30 %) worse for the same signal efficiency for the CellBased (EflowRec) based variable set. However, for the LLH technique a slight improvement of up to 10 % for a signal efficiency range between 25 % and 50 % can be stated in the case of a CellBased calculation.



**Figure 6.29:** Background rejection against signal efficiency for a BDT-based tau identification exploiting the default variable set (black) and a variable set of all defined substructure based variables for the evaluation on 1-prong (left) and multi-prong (right) tau candidates. The approach based on the CellBased algorithm is illustrated as blue dashed line and for EflowRec as orange dotted line. The ratios below each plot depict the according deviations of the background rejection for a given signal efficiency w.r.t. the default approach. The same colour scheme is applied.



**Figure 6.30:** Background rejection against signal efficiency for a LLH-based tau identification exploiting the default variable set (black) and a variable set of all defined substructure based variables for the evaluation on 1-prong (left) and multi-prong (right) tau candidates. The approach based on the CellBased algorithm is illustrated as blue dashed line and for EflowRec as orange dotted line. The ratios below each plot depict the according deviations of the background rejection for a given signal efficiency w.r.t. the default approach. The same colour scheme is applied.

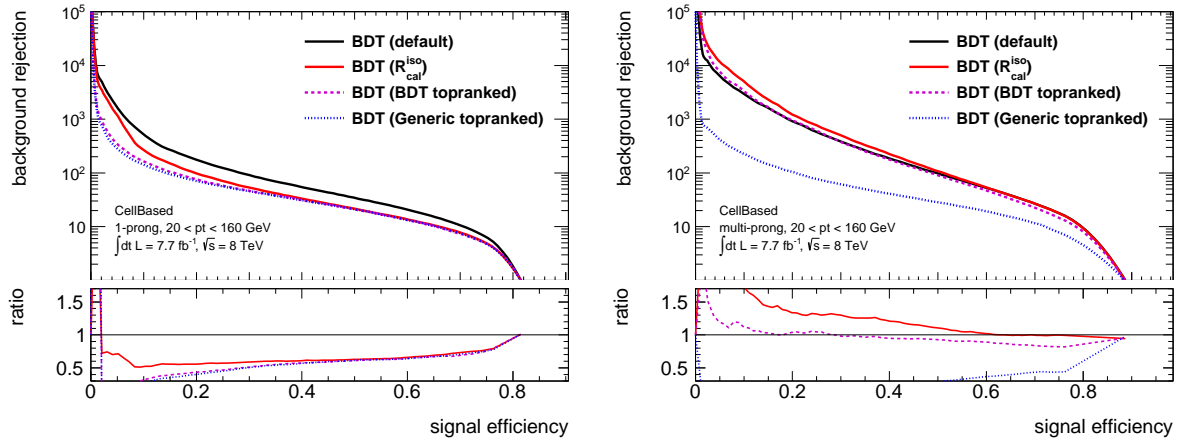
Although the set of all substructure-based variables provide a lot of information to discriminate real taus against QCD jets, it is not possible to improve the performance for 1-prong decays significantly. However, the default performance is almost recovered, which is a huge achievement given the much coarser granularity of the input variables whereas for multi-prong decays a clear gain in performance is observed for both multivariate approaches as well as for both substructure algorithms. A profit for the LLH technique is seen almost over the entire signal efficiency range about 30 % (20 %) for the CellBased (EflowRec) variables. A similar result of an about 10 % higher background rejection for the EflowRec based BDT strategy is achieved. For the calculation by the CellBased algorithm it was even possible to exceed the performance of the default strategy by roughly 50 % for the medium working point (corresponding to a signal efficiency of 55 %).

One has to take into account that the constructed boosted decision tree cannot serve for a practical identification approach, because the improvement obtained through the information provided by such a large variable set comes at the cost of an enlargement of systematic uncertainties. Each variable increases the overall uncertainty due to a slightly incorrect simulation and a non-perfect determination of pile-up corrections and  $p_T/\mu$  weights, which is expected to show up in increased scale factors. Additionally, this will be increased by uncertainties raised within the reconstruction step. Nevertheless, the rankings of the discriminating variables based on the BDT classification can be used for further investigations. The rankings are summarised in Table 6.1 for the highest-ranked variables. The remaining variables are not considered.

	generic ranking				BDT specific ranking			
	CellBased		EflowRec		CellBased		EflowRec	
	1-prong	3-prong	1-prong	3-prong	1-prong	3-prong	1-prong	3-prong
1	$R_{\text{cal,all}}^{\text{iso}}$	$R_{\text{cal,chg}}^{\text{iso}}$	$R_{\text{cal,chg}}^{\text{iso}}$	$R_{\text{cal,chg}}^{\text{iso}}$	$R_{\text{cal,all}}^{\text{iso}}$	$R_{\text{cal,chg}}^{\text{iso}}$	$f_{E,\text{chg}}^{\text{cent-iso}}$	$R_{\text{cal,chg}}^{\text{iso}}$
2	$f_{E,\text{all}}^{\text{cent-iso}}$	$f_{E,\text{chg}}^{\text{cent-iso}}$	$f_{E,\text{chg}}^{\text{cent-iso}}$	$f_{E,\text{chg}}^{\text{cent-iso}}$	$f_{E,\text{all}}^{\text{cent-iso}}$	$f_{E,\text{chg}}^{\text{core-iso}}$	$f_{E,\text{chg}}^{\text{core-iso}}$	$f_{E,\text{chg}}^{\text{core-iso}}$
3	$R_{\text{cal,chg}}^{\text{iso}}$	$f_{E,\text{all}}^{\text{cent-iso}}$	$R_{\text{cal,all}}^{\text{iso}}$	$f_{E,\text{all}}^{\text{cent-iso}}$	$f_{E,\text{all}}^{\text{core-iso}}$	$f_{E,\text{chg}}^{\text{core-iso}}$	$f_{E,\text{all}}^{\text{cent-iso}}$	$S_{\text{T}}^{\text{flight}}$
4	$f_{E,\text{chg}}^{\text{cent-iso}}$	$R_{\text{cal,all}}^{\text{iso}}$	$f_{E,\text{chg}}^{\text{cent-iso}}$	$\Delta R_{\text{max,all}}^{\text{iso}}$	$N_{\text{chg}}^{\text{wide}}$	$R_{\text{cal,all}}^{\text{iso}}$	$\Delta R_{\text{max,all}}^{\text{iso}}$	$f_{E,\text{chg}}^{\text{cent-iso}}$
5	$f_{E,\text{all}}^{\text{core-iso}}$	$\Delta R_{\text{max,all}}^{\text{iso}}$	$\Delta R_{\text{max,all}}^{\text{iso}}$	$R_{\text{cal,all}}^{\text{iso}}$	$R_{\text{cal,all}}^{\text{core}}$	$S_{\text{T}}^{\text{flight}}$	$R_{\text{cal,all}}^{\text{iso}}$	$f_{E,\text{all}}^{\text{core-iso}}$
6	$\Delta R_{\text{max,all}}^{\text{iso}}$	$f_{E,\text{all}}^{\text{core-iso}}$	$R_{\text{cal,all}}^{\text{core}}$	$f_{E,\text{chg}}^{\text{core-iso}}$	$\Delta R_{\text{max,all}}^{\text{iso}}$	$f_{E,\text{all}}^{\text{cent-iso}}$	$R_{\text{cal,all}}^{\text{core}}$	$f_{E,\text{all}}^{\text{core-iso}}$
7	$R_{\text{cal,all}}^{\text{core}}$	$f_{E,\text{chg}}^{\text{core-iso}}$	$f_{E,\text{all}}^{\text{core-iso}}$	$R_{\text{cal,chg}}^{\text{core}}$	$f_{E,\text{all}}^{\text{core-iso}}$	$N_{\text{chg}}^{\text{wide}}$	$R_{\text{cal,chg}}^{\text{iso}}$	$R_{\text{cal,all}}^{\text{iso}}$
8	$f_{E,\text{chg}}^{\text{core-iso}}$	$R_{\text{cal,chg}}^{\text{core}}$	$f_{E,\text{chg}}^{\text{core-iso}}$	$\Delta R_{\text{max,all}}^{\text{core}}$	$f_{E,\text{chg}}^{\text{cent-iso}}$	$f_{\text{pt,chg}}^{\text{core}}$	$N_{\text{chg}}^{\text{wide}}$	$N_{\text{chg}}^{\text{wide}}$
9	$\Delta R_{\text{max,all}}^{\text{core}}$	$\Delta R_{\text{max,all}}^{\text{core}}$	$R_{\text{cal,chg}}^{\text{core}}$	$R_{\text{cal,all}}^{\text{core}}$	$S_{\text{lead track}}^{\text{IP}}$	$f_{\text{pt,neut}}^{\text{iso}}$	$S_{\text{lead track}}^{\text{IP}}$	$m_{\text{chg}}^{\text{core}}$
10	$R_{\text{cal,chg}}^{\text{core}}$	$R_{\text{cal,all}}^{\text{core}}$	$\Delta R_{\text{max,all}}^{\text{core}}$	$f_{E,\text{all}}^{\text{core-iso}}$	$f_{\text{pt,all}}^{\text{core}}$	$m_{\text{chg}}^{\text{core}}$	$f_{\text{pt,all}}^{\text{core}}$	$f_{\text{pt,chg}}^{\text{core}}$
11	$\Delta R_{\text{pt,minmax}}^{\text{iso}}$	$m_{\text{chg}}^{\text{iso}}$	$N_{\text{chg}}^{\text{wide}}$	$m_{\text{chg}}^{\text{iso}}$	$f_{\text{pt,all}}^{\text{cent}}$	$m_{\text{all}}^{\text{core}}$	$f_{\text{pt,all}}^{\text{cent}}$	$\Delta R_{\text{max,all}}^{\text{iso}}$
12	$f_{\text{pt,all}}^{\text{cent}}$	$\Delta R_{\text{pt,minmax}}^{\text{iso}}$	$\Delta R_{\text{pt,minmax}}^{\text{iso}}$	$\Delta R_{\text{pt,minmax}}^{\text{iso}}$	$f_{\text{pt,all}}^{\text{iso}}$	$\Delta R_{\text{max,all}}^{\text{iso}}$	$m_{\text{all}}^{\text{iso}}$	$f_{\text{pt,chg}}^{\text{cent}}$
13	$m_{\text{chg}}^{\text{iso}}$	$N_{\text{chg}}^{\text{cent}}$	$m_{\text{chg}}^{\text{iso}}$	$N_{\text{chg}}^{\text{cent}}$	$m_{\text{all}}^{\text{iso}}$	$\Delta R_{\text{max,all}}^{\text{core}}$	$f_{\text{pt,all}}^{\text{iso}}$	$m_{\text{all}}^{\text{core}}$
14	$N_{\text{chg}}^{\text{wide}}$	$m_{\text{chg}}^{\text{cent}}$	$f_{\text{lead,chg}}^{\text{iso}}$	$m_{\text{chg}}^{\text{cent}}$	$m_{\text{neut}}^{\text{core}}$	$f_{\text{pt,all}}^{\text{core}}$	$m_{\text{all}}^{\text{core}}$	$R_{\text{cal,all}}^{\text{core}}$
15	$N_{\text{chg}}^{\text{cent}}$	$m_{\text{all}}^{\text{cent}}$	$f_{\text{pt,all}}^{\text{cent}}$	$m_{\text{all}}^{\text{iso}}$	$m_{\text{neut}}^{\text{cent}}$	$S_{\text{lead track}}^{\text{IP}}$	$f_{\text{lead,chg}}^{\text{iso}}$	$S_{\text{lead track}}^{\text{IP}}$
16	$f_{\text{pt,all}}^{\text{core}}$	$f_{\text{pt,all}}^{\text{cent}}$	$m_{\text{all}}^{\text{iso}}$	$f_{\text{lead,chg}}^{\text{iso}}$	$f_{\text{pt,neut}}^{\text{iso}}$	$f_{\text{pt,chg}}^{\text{cent}}$	$m_{\text{neut}}^{\text{cent}}$	$f_{\text{lead,chg}}^{\text{iso}}$
17	$m_{\text{all}}^{\text{iso}}$	$m_{\text{all}}^{\text{iso}}$	$N_{\text{chg}}^{\text{cent}}$	$S_{\text{T}}^{\text{flight}}$	$N_{\text{neut}}^{\text{wide}}$	$m_{\text{chg}}^{\text{cent}}$	$m_{\text{chg}}^{\text{iso}}$	$m_{\text{chg}}^{\text{cent}}$
18	$N_{\text{neut}}^{\text{wide}}$	$S_{\text{T}}^{\text{flight}}$	$f_{\text{pt,chg}}^{\text{cent}}$	$f_{\text{pt,all}}^{\text{cent}}$	$f_{\text{lead,chg}}^{\text{core}}$	$m_{\text{chg}}^{\text{iso}}$	$f_{\text{pt,chg}}^{\text{core}}$	$m_{\text{chg}}^{\text{iso}}$
19	$f_{\text{lead,chg}}^{\text{iso}}$	$f_{\text{pt,chg}}^{\text{cent}}$	$f_{\text{pt,chg}}^{\text{core}}$	$m_{\text{all}}^{\text{core}}$	$m_{\text{chg}}^{\text{iso}}$	$f_{\text{pt,chg}}^{\text{iso}}$	$\Delta R_{\text{pt,minmax}}^{\text{iso}}$	$\Delta R_{\text{max,all}}^{\text{core}}$
20	$f_{\text{pt,chg}}^{\text{cent}}$	$f_{\text{lead,chg}}^{\text{iso}}$	$N_{\text{neut}}^{\text{core}}$	$f_{\text{pt,chg}}^{\text{cent}}$	$f_{\text{pt,chg}}^{\text{core}}$	$f_{\text{lead,chg}}^{\text{core}}$	$f_{E,\text{all}}^{\text{core-iso}}$	$m_{\text{all}}^{\text{iso}}$

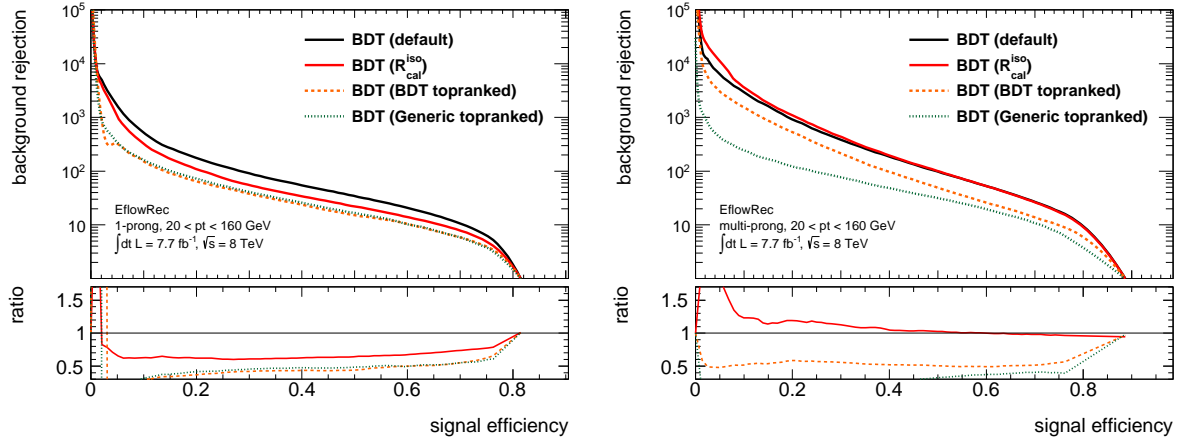
**Table 6.1:** Summary of the top ranked variables for 1- and 3-prong tau candidates based on a generic and a BDT specific ranking.

A general as well as a BDT specific ranking is generated by TMVA, which differ for CellBased and EflowRec, and thus are considered separately for the individual substructure algorithms. As a first approach only the eight and nine top ranked variables are taken into account for 1- and 3-prong tau candidates, respectively. This is motivated by the size of the variable sets exploited in the default strategy, which was validated to be a reasonable choice to address the complexity of hadronic tau decays. The resulting background rejection against signal efficiency is shown in Figure 6.31 and 6.32 for the BDT-based identification for both substructure algorithms. For 1-prong tau candidates the generic as well as the BDT specific top ranked variable set yields almost the same background rejection. In the case of the CellBased approach the performance is similar to the one obtained with the substructure variable set including  $R_{\text{cal}}^{\text{iso}}$  for signal efficiencies above 40 %, but decrease towards lower signal efficiencies. Both rankings based on the EflowRec algorithm perform even worse compared to the variable set containing  $R_{\text{cal}}^{\text{iso}}$ , and thus result in a significant performance loss of about 50 % w.r.t. the default strategy. The set of the top ranked variables lead to a huge loss in background rejection for 3-prong tau candidates. Only the CellBased approach is not affected that much, but performs nevertheless worse than the variable set including  $R_{\text{cal}}^{\text{iso}}$ .

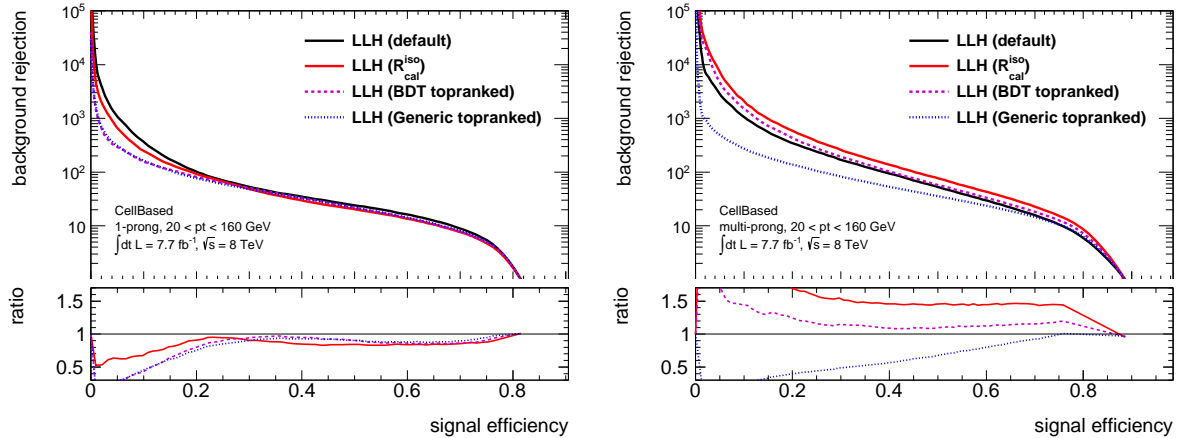


**Figure 6.31:** Background rejection against signal efficiency for a BDT-based tau identification exploiting the default variable set (black), a substructure equivalent variable set including  $R_{\text{cal}}^{\text{iso}}$  (red) and a set of the eight (nine) top ranked substructure based variables for the evaluation on 1-prong (left) and multi-prong (right) tau candidates. The approaches based on the general and BDT specific ranking for the CellBased algorithm is illustrated by the dotted blue and dashed magenta line. The ratios below each plot depict the according deviations of the background rejection for a given signal efficiency w.r.t. the default approach. The same colour scheme is applied.

Since the generic ranking is calculated based on the separation power of each single variable it is also suitable for the LLH technique, as it represents the information accessible for the LLH approach. However, the BDT specific ranking considers correlations in addition, and thus might provide an even better performing selection. Hence, this ranking is investigated for the LLH as well. Figure 6.33 and 6.34 presents the background rejection against signal efficiency for the generic and BDT specific ranking for both substructure algorithms. Also for the LLH-based identification of 1-prong tau candidates the performance is similar for both rankings. Furthermore, the obtained background rejection is slightly better and differs only by about 5 % (10 %) from the default strategy for the CellBased (EflowRec) algorithm. For multi-prong tau decays the performance achievable with the BDT specific top ranked variables calculated on the output of the CellBased algorithm is about 10 % higher than the one obtainable with the default variable set. The same variable set based on the EflowRec algorithm results in a performance fluctuating around the one of the default approach. While the best variables of the generic ranking provide in general a less efficient background rejection.

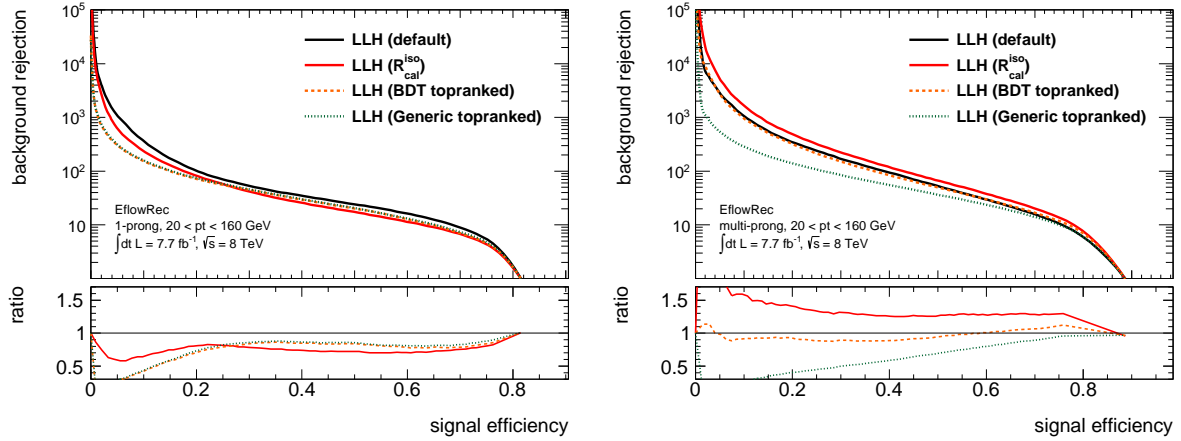


**Figure 6.32:** Background rejection against signal efficiency for a BDT-based tau identification exploiting the default variable set (black), a substructure equivalent variable set including  $R_{\text{cal}}^{\text{iso}}$  (red) and a set of the eight (nine) top ranked substructure based variables for the evaluation on 1-prong (left) and multi-prong (right) tau candidates. The approaches based on the general and BDT specific ranking for the EflowRec algorithm is illustrated by the dotted green and dashed orange line. The ratios below each plot depict the according deviations of the background rejection for a given signal efficiency w.r.t. the default approach. The same colour scheme is applied.



**Figure 6.33:** Background rejection against signal efficiency for a LLH-based tau identification exploiting the default variable set (black), a substructure equivalent variable set including  $R_{\text{cal}}^{\text{iso}}$  (red) and a set of the eight (nine) top ranked substructure based variables for the evaluation on 1-prong (left) and multi-prong (right) tau candidates. The approaches based on the general and BDT specific ranking for the CellBased algorithm is illustrated by the dotted blue and dashed magenta line. The ratios below each plot depict the according deviations of the background rejection for a given signal efficiency w.r.t. the default approach. The same colour scheme is applied.

For the BDT as well as for the LLH technique the variable set extracted from the BDT specific ranking separates real and fake taus better than the one of the generic ranking. However, in both cases the performance of the substructure variable set including  $R_{\text{cal}}^{\text{iso}}$  cannot be exceeded for multi-prong decays. The same applies for the BDT approach for 1-prong tau candidates. Only with the LLH technique exploiting the eight top ranked variables it is possible to obtain a small gain in background rejection within the relevant signal efficiency range.

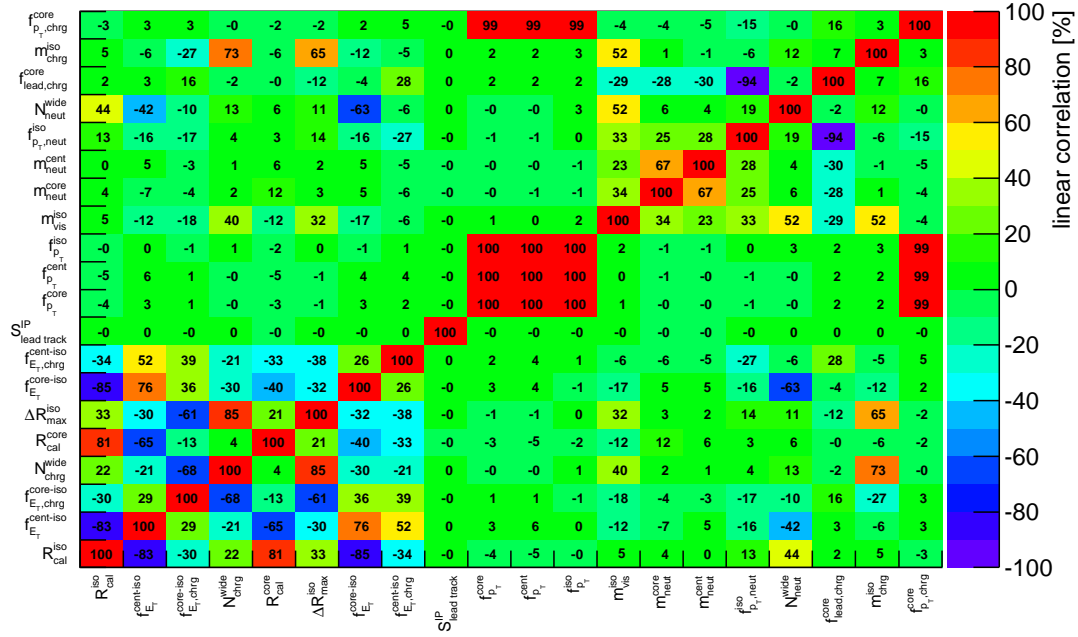


**Figure 6.34:** Background rejection against signal efficiency for a LLH-based tau identification exploiting the default variable set (black), a substructure equivalent variable set including  $R_{\text{cal}}^{\text{iso}}$  (red) and a set of the eight (nine) top ranked substructure based variables for the evaluation on 1-prong (left) and multi-prong (right) tau candidates. The approaches based on the general and BDT specific ranking for the EflowRec algorithm is illustrated by the dotted green and dashed orange line. The ratios below each plot depict the according deviations of the background rejection for a given signal efficiency w.r.t. the default approach. The same colour scheme is applied.

Considering those results the information extracted from the rankings is not the best choice and needs to be further investigated. In addition to the ranking the correlations between the variables are analysed to optimise the variable set. In particular for the BDT specific ranking it might occur that highly correlated variables are ranked at top, as it is based on the choice of each single decision tree, which can take either of two well separating but strongly correlated variables. Therefore, linear correlation matrices are calculated separately for 1- and 3-prong signal events as depicted in Figure 6.35 exemplary for the BDT specific top ranked 1-prong variables. The corresponding correlation matrices for background events are very similar to those obtained from signal events, and thus do not contribute any further information here. Only the 20 highest ranked variables are considered since it can be assumed that the lower ranked variables provide only a very low separation of real and fake taus.

Starting from the first variable of the ranking a variable is kept if its correlation to any previous one is less than 50%, otherwise it is rejected. This procedure is based on the assumption, that two highly correlated quantities provide similar information, and thus one of them is sufficient and a consideration of the second will not lead to a significant improvement of the performance. Hence, the size of the resulting variable sets, summarised in Table 6.2, can differ depending on the correlations.

Figures 6.36-6.39 summarise the background rejection against signal efficiency for the BDT- and LLH-based tau identification exploiting sets of top ranked variables which are optimised by considering their correlations between each other. Both substructure algorithms and categories of tau decays are depicted. The constructed variable sets exploiting the ranking as well as the variable correlations, lead to an improvement in the background rejection w.r.t. those only based on the rankings. For the BDT-based identification of 1-prong tau decays the obtainable performance is still about 15 % worse than the one extracted from the default strategy. As observed before, an identification exploiting the generic ranking does not perform as well as those based on the BDT specific one, even though the correlations are considered. Only in the case of the BDT (LLH) approach, exploiting the according set of CellBased (EflowRec) variables the background rejection is about 10 % higher, but cannot outperform the substructure variable set including  $R_{\text{cal}}^{\text{iso}}$ . Nevertheless, the performance of the default variable set can be reached for the BDT-based multi-prong identification as well as for the 1-prong LLH-based identification for both underlying substructure algorithms. For the LLH approach the background rejection is even up to 20 % higher than achievable by the default strategy for multi-prong tau decays. Similar results can be stated



**Figure 6.35:** Linear correlation coefficients of the BDT specific top ranked CellBased variables for re-constructed 1-prong tau candidates extracted from MC generated events required to match to generated 1-prong decays.

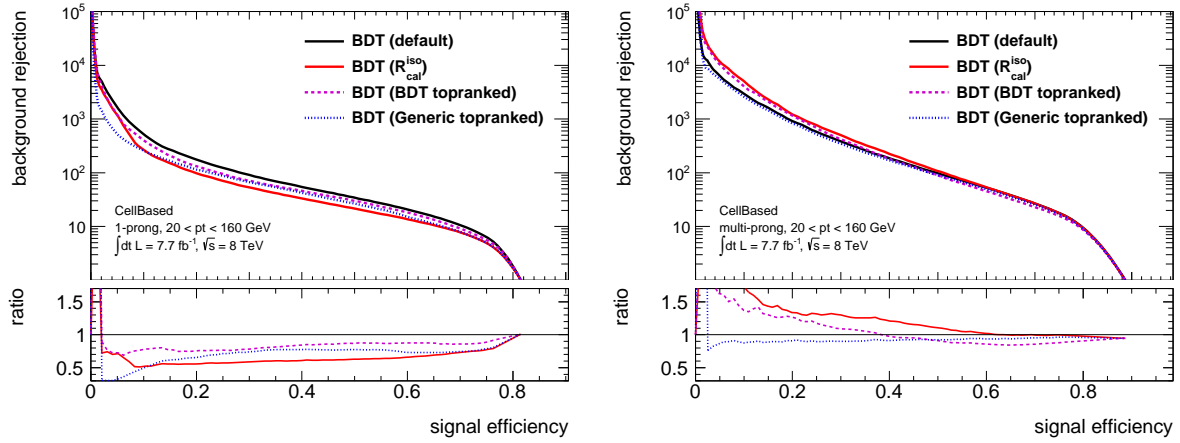
generic ranking				BDT specific ranking			
CellBased		EflowRec		CellBased		EflowRec	
1-prong	3-prong	1-prong	3-prong	1-prong	3-prong	1-prong	3-prong
$R_{\text{cal,all}}^{\text{iso}}$	$R_{\text{cal,chg}}^{\text{iso}}$	$R_{\text{cal,chg}}^{\text{iso}}$	$R_{\text{cal,chg}}^{\text{iso}}$	$R_{\text{cal,all}}^{\text{iso}}$	$R_{\text{cal,chg}}^{\text{iso}}$	$f_{E,\text{chg}}^{\text{cent-iso}}$	$R_{\text{cal,chg}}^{\text{iso}}$
$f_{E,\text{chg}}^{\text{cent-iso}}$	$\Delta R_{\text{max,all}}^{\text{iso}}$	$f_{E,\text{all}}^{\text{core-iso}}$	$\Delta R_{\text{max,all}}^{\text{iso}}$	$N_{\text{chg}}^{\text{wide}}$	$f_{E,\text{chg}}^{\text{core-iso}}$	$f_{E,\text{chg}}^{\text{core-iso}}$	$f_{E,\text{chg}}^{\text{core-iso}}$
$\Delta R_{\text{max,all}}^{\text{iso}}$	$m_{\text{chg}}^{\text{iso}}$	$N_{\text{chg}}^{\text{wide}}$	$f_{E,\text{all}}^{\text{core-iso}}$	$\Delta R_{\text{max,all}}^{\text{iso}}$	$S_{\text{T}}^{\text{flight}}$	$f_{E,\text{all}}^{\text{cent-iso}}$	$S_{\text{T}}^{\text{flight}}$
$f_{p_{\text{T}},\text{all}}^{\text{cent}}$	$m_{\text{chg}}^{\text{cent}}$	$\Delta R_{p_{\text{T}}^{\text{minmax}}}^{\text{iso}}$	$m_{\text{chg}}^{\text{cent}}$	$f_{E,\text{chg}}^{\text{cent-iso}}$	$N_{\text{chg}}^{\text{wide}}$	$S_{\text{lead track}}^{\text{IP}}$	$f_{E,\text{all}}^{\text{core-iso}}$
$m_{\text{all}}^{\text{iso}}$	$m_{\text{all}}^{\text{cent}}$	$f_{\text{lead,chg}}^{\text{iso}}$	$m_{\text{all}}^{\text{iso}}$	$S_{\text{lead track}}^{\text{IP}}$	$f_{p_{\text{T}},\text{chg}}^{\text{core}}$	$f_{p_{\text{T}},\text{all}}^{\text{core}}$	$m_{\text{chg}}^{\text{core}}$
$f_{\text{lead,chg}}^{\text{iso}}$	$f_{p_{\text{T}},\text{all}}^{\text{cent}}$	$f_{p_{\text{T}},\text{all}}^{\text{cent}}$	$f_{\text{lead,chg}}^{\text{iso}}$	$f_{p_{\text{T}},\text{all}}^{\text{core}}$	$f_{p_{\text{T}},\text{neut}}^{\text{iso}}$	$m_{\text{all}}^{\text{iso}}$	$m_{\text{all}}^{\text{core}}$
	$S_{\text{T}}^{\text{flight}}$	$m_{\text{all}}^{\text{iso}}$	$S_{\text{T}}^{\text{flight}}$	$m_{\text{all}}^{\text{iso}}$	$m_{\text{all}}^{\text{core}}$	$f_{\text{lead,chg}}^{\text{iso}}$	$S_{\text{lead track}}^{\text{IP}}$
	$f_{\text{lead,chg}}^{\text{iso}}$		$f_{p_{\text{T}},\text{all}}^{\text{cent}}$	$m_{\text{neut}}^{\text{core}}$	$S_{\text{lead track}}^{\text{IP}}$	$m_{\text{neut}}^{\text{cent}}$	$f_{\text{lead,chg}}^{\text{iso}}$
				$f_{p_{\text{T}},\text{neut}}^{\text{iso}}$		$m_{\text{chg}}^{\text{iso}}$	$m_{\text{chg}}^{\text{iso}}$

**Table 6.2:** Summary of the top ranked variables reduced according to their correlations, for 1- and 3-prong tau candidates based on a general and a BDT specific ranking.

for the variable sets based on the BDT specific ranking, optimised by the consideration of correlations, for the LLH technique for 1-prong and for the BDT approach for multi-prong tau candidates. In the latter case a significant increase of the background rejection is achieved towards lower signal efficiency values including also interesting values between 30 % and 40 %, by using an EflowRec based variable calculation. A similar behaviour is observed for the identification of 1-prong taus, although here the performance drops strongly. The performance gain w.r.t. the default approach of about 50 % for signal

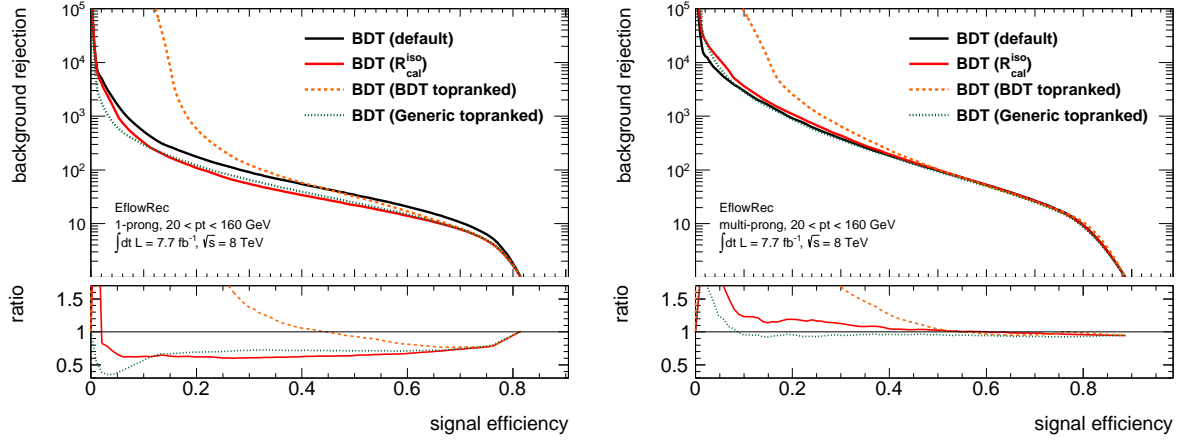


efficiencies smaller than 45 %, turns in to a loss of about 20 % for higher values. The background rejection resulting from the LLH-based algorithm for multi-prong tau candidates is clearly better for the correlation optimised variable set for both substructure algorithms compared to the default variable set. It is even possible to exceed the performance obtainable by the variable set including  $R_{\text{cal}}^{\text{iso}}$  for signal efficiencies lower than 65 % (42 %) in the case of the CellBased (EflowRec) as underlying substructure algorithm. The achieved performance is up to a factor of 2 more efficient than observed for the default strategy. Considering those results it is possible to reach or even exceed the default performance by well chosen sets of variables providing a good separation power itself, but also through the correlations with each other. Further investigation are necessary to find the optimal variable set which provides a good background rejection over the entire signal efficiency range. These might be different for the BDT and LLH approach.

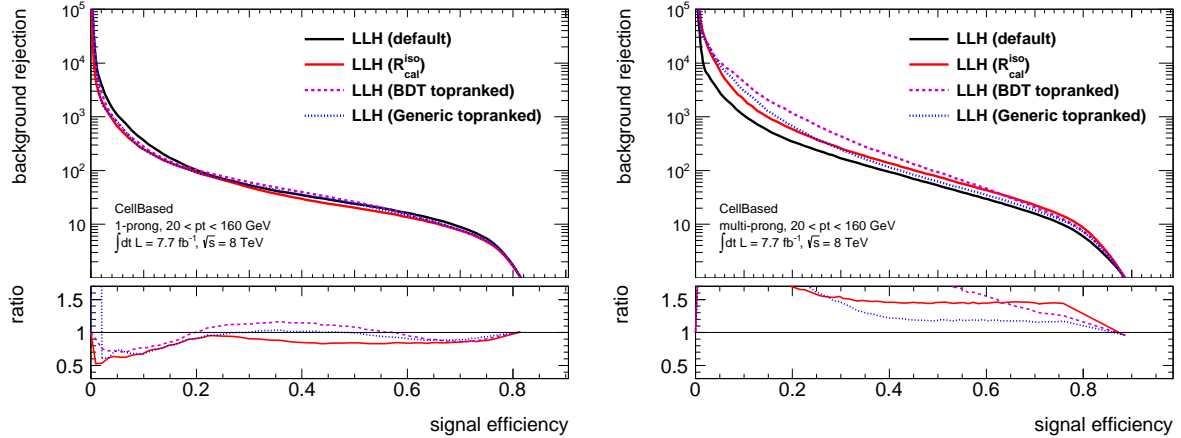


**Figure 6.36:** Background rejection against signal efficiency for a BDT-based tau identification exploiting the default variable set (black), a substructure equivalent variable set including  $R_{\text{cal}}^{\text{iso}}$  (red) and a set of the top ranked substructure based variables considering correlation, for the evaluation on 1-prong (left) and multi-prong (right) tau candidates. The approaches based on the general and BDT specific ranking for the CellBased algorithm is illustrated by the dotted blue and dashed magenta line. The ratios below each plot depict the according deviations of the background rejection for a given signal efficiency w.r.t. the default approach. The same colour scheme is applied.

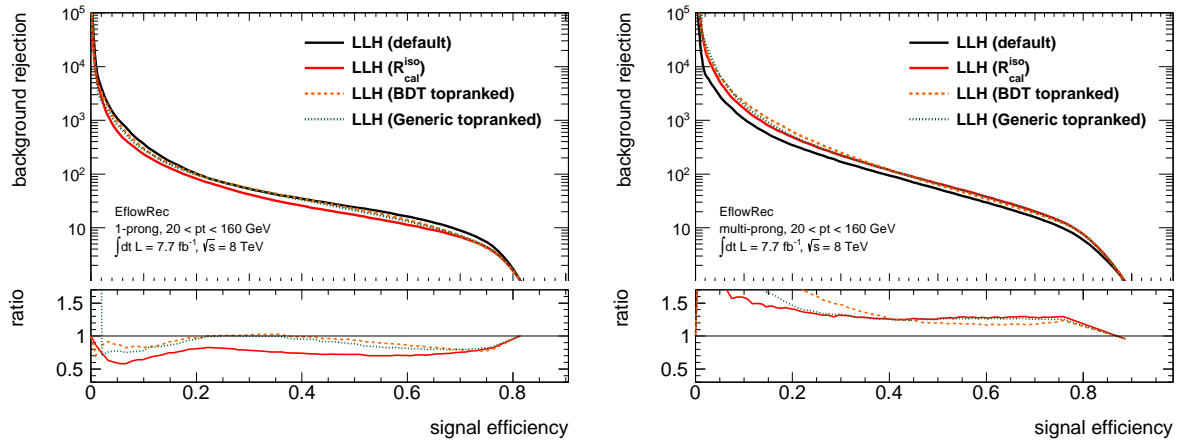




**Figure 6.37:** Background rejection against signal efficiency for a BDT-based tau identification exploiting the default variable set (black), a substructure equivalent variable set including  $R_{\text{cal}}^{\text{iso}}$  (red) and a set of the top ranked substructure based variables considering correlation, for the evaluation on 1-prong (left) and multi-prong (right) tau candidates. The approaches based on the general and BDT specific ranking for the EflowRec algorithm is illustrated by the dotted green and dashed orange line. The ratios below each plot depict the according deviations of the background rejection for a given signal efficiency w.r.t. the default approach. The same colour scheme is applied.



**Figure 6.38:** Background rejection against signal efficiency for a LLH-based tau identification exploiting the default variable set (black), a substructure equivalent variable set including  $R_{\text{cal}}^{\text{iso}}$  (red) and a set of the top ranked substructure based variables considering correlation, for the evaluation on 1-prong (left) and multi-prong (right) tau candidates. The approaches based on the general and BDT specific ranking for the CellBased algorithm is illustrated by the dotted blue and dashed magenta line. The ratios below each plot depict the according deviations of the background rejection for a given signal efficiency w.r.t. the default approach. The same colour scheme is applied.



**Figure 6.39:** Background rejection against signal efficiency for a LLH-based tau identification exploiting the default variable set (black), a substructure equivalent variable set including  $R_{\text{cal}}^{\text{iso}}$  (red) and a set of the top ranked substructure based variables considering correlation, for the evaluation on 1-prong (left) and multi-prong (right) tau candidates. The approaches based on the general and BDT specific ranking for the EflowRec algorithm is illustrated by the dotted green and dashed orange line. The ratios below each plot depict the according deviations of the background rejection for a given signal efficiency w.r.t. the default approach. The same colour scheme is applied.

## 7 Summary and Outlook

Efficient tau identification algorithms are crucial to cover a broad spectrum of studies within and beyond the Standard Model. Therefore, a continuous optimisation of the exploited techniques is essential, but also new approaches need to be studied to extend the spectra of applicability.

The application and impact of several updates on the two approaches used within the tau identification were presented. A change in the technical framework TMVA used for the BDT approach raised the necessity for validation of the configuration of the BDT. Hence, the way to set the minimal number of objects per leaf is changed from an absolute to a relative criterion. The implemented default value of 5 % turned out to be not reasonable, thus configurations differing in this adjustment were studied. Those studies showed that a value of 0.1 % is the best choice to set the minimal number of events per leave, in terms of performance and CPU consumption. The LLH-approach has been updated to the state of the BDT, since it was abandoned in the last round of updates. This includes the expansion of the discriminating variable set by the three  $\pi^0$  quantities,  $N_{\pi^0}$ ,  $m_{\tau}^{\text{vis}}$  and  $f_{\text{vis-}p_T}$ . No performance loss was observed by considering these variables, but no enhancement could be achieved either. The reason for this was found to be in the low separation power of those variables and the fact that the LLH technique cannot profit from the correlations between the variables, while the BDT is able to do so. A second update was the application of a  $p_T$  and  $\mu$  reweighting technique. This is necessary to guarantee that the datasets used for signal and background events have the same kinematic behaviour and no bias due to different kinematics of the datasets can occur. The reweighting works reasonably well and did not harm the performance of the log-Likelihood tau identification. Further, the calculation of the p.d.f.s was corrected, such that the performance is independent of the amount of used signal and background events.

A new approach of a pure substructure based tau identification was developed to extend its application to a broader field of physics analyses. In the first step within the scope of these studies, the standard set of discriminating variables is recalculated based on the dedicated substructure algorithms CellBased and EflowRec in combination with the PanTau algorithm. Hence, the tau decay products, i.e. neutral and charged pions and kaons, are considered instead of calorimeter cells, which leads to a coarser granularity, such that e.g. variables defined as energy fractions suffer from this issue. Apparently, this affects both identification algorithms and results in a performance loss w.r.t. the default strategy. The negative impact was even higher for the BDT-based approach, since not only the information provided by a single variable gets lost, but also the one extracted from the correlations between the variables.

A new variable, the calorimeter radius defined in the isolation cone, was introduced as replacement for the affected variable in the default set to recover the performance. Hence, it was possible to achieve almost 90 % of the performance of the default variable set for an identification of 1-prong tau decays for both underlying multivariate techniques. This was even exceeded up to 20 % w.r.t. the standard strategy for an identification of 3-prong taus. The performance gain comes at the cost of a stronger dependence on contributions of pile-up events. A pile-up correction was applied to counteract this negative effect. Though, this decreases the performance of the BDT approach, while it increases the one obtainable with the LLH technique. This can be explained by the fact that the pile-up corrections do not only influence the actual distributions of the variables but also the correlations between them, and thus only the BDT-based tau identification algorithm is affected. A further optimisation of the recalculated standard variable set is the expansion to the isolation cone of the quantity describing the leading track momentum fraction. This variable does also suffer from the coarser granularity, but the consideration of the information provided by the isolation annulus reduces this significantly. Exploiting the substructure equivalent variable set including  $R_{\text{cal}}^{\text{iso}}$  and  $f_{\text{lead,chg}}^{\text{iso}}$  almost the same or only a slightly worse background rejection as for the one only considering  $R_{\text{cal}}^{\text{iso}}$ , was achievable. Based on this an expansion to the isolation annulus for all discriminating variables was studied and similar results were observed. The increased contribution of pile-up events raised the necessity of  $\mu$  dependent corrections which lower the obtainable efficiency in terms of background rejection w.r.t. to the variable set including only  $R_{\text{cal}}$  and  $f_{\text{track}}$  in the isolation cone, for the LLH approach. The BDT-based tau identification algorithm was not affected by this since the information accessible through the variable correlation remains the same.

Furthermore, the extension of the set of identification variables which includes the introduction of new variables exploiting the tau decay kinematics, was investigated. About 40 substructure based variables and their separation power as well as their correlations between each other were studied for both identification approaches and substructure algorithms. It was possible to almost recover the performance of the default strategy in terms of background rejection. Especially for the identification of 3-prong taus an up to a factor of 2 higher background rejection w.r.t. the one for the default strategy, was achievable. The presented analyses showed that the approach of a pure substructure based tau identification is successful and that it is possible to recover and partly also to exceed the performance obtainable by the default variable set. This is a huge achievement given the coarser granularity of the substructure based variables.

Nevertheless, it is still possible to further optimise these results. Thus, continuing investigation of the separation power and correlations of the variables might lead to an even better rejection of fake taus, whereby the resulting variable sets could differ for the identification technique as well as for the underlying substructure algorithms. It would also be interesting to determine the corresponding scale factors to quantify the reliability in data. A further distinction of the decay modes also based on the amount of neutral decay products instead of only the number of tracks, could yield a more suitable and effective identification. Especially the performance for the decay modes with one or three charged and zero or one neutral decay product might be enhanced. Yet, a lack of data statistics to study such categories individually, might be the limiting factor. Moreover, an optimisation of the technical concept could increase the performance of the LLH approach. Hence, considering the correlations between the discriminating variables would presumably lead to an improvement. Another option is to refine the  $p_T$  binning or introduce a binning in  $\eta$ , but also here the amount of available events will be an issue. Nevertheless, the outlined studies show that it is possible to construct a pure substructure based tau identification which may serve to explore e.g. the nature of the Higgs boson during tun-II data-taking, though the performance and reliability of the newly developed algorithms will have to be validated on 13/14 TeV MC simulated data.

## A Auxiliary Information for the Theoretical Foundations

This appendix summarises auxiliary information for the theoretical foundations used to describe high energy physics. [A.1](#) and [A.2](#) introduce the gamma and Pauli matrices, respectively.

### A.1 Pauli Matrices

The *Pauli matrices* are  $2 \times 2$  matrices given by:

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \text{and} \quad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (\text{A.1})$$

### A.2 Gamma Matrices

The *gamma matrices* are a set of  $4 \times 4$  matrices given in the *Dirac basis* as:

$$\gamma^0 = \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{1} \end{pmatrix} \quad \text{and} \quad \gamma^i = \begin{pmatrix} \mathbf{0} & \sigma^i \\ \sigma^i & \mathbf{0} \end{pmatrix}. \quad (\text{A.2})$$

Furthermore, the following anticommutation relation is guaranteed:

$$\{\gamma^\mu, \gamma^\nu\} = \gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu = 2\eta^{\mu\nu} \mathbf{1}_{4 \times 4}, \quad (\text{A.3})$$

with  $\eta^{\mu\nu}$  being the *Minkowski metric* with signature  $(+ - - -)$ .

## **B Samples**

The following samples of events generated by MC or extracted from a QCD-di-jet selection in 2012 data were used within this thesis.

MC samples:
group.perf-tau.TauPi0Rec.D3PD.147818.Pythia8_AU2CTEQ6L1_Ztautau.recon.ESD.e1176_s1479_s1470_r3553_tid00999073_00.v05-01
group.perf-tau.TauPi0Rec.D3PD.147818.Pythia8_AU2CTEQ6L1_Ztautau.recon.ESD.e1176_s1479_s1470_r3553_tid00999074_00.v05-01
group.perf-tau.TauPi0Rec.D3PD.147818.Pythia8_AU2CTEQ6L1_Ztautau.recon.ESD.e1176_s1479_s1470_r3553_tid00999075_00.v05-01
group.perf-tau.TauPi0Rec.D3PD.147818.Pythia8_AU2CTEQ6L1_Ztautau.recon.ESD.e1176_s1479_s1470_r3553_tid00999076_00.v05-01
group.perf-tau.TauPi0Rec.D3PD.170201.Pythia8_AU2CTEQ6L1_Zprime250tautau.recon.ESD.e1176_s1479_s1470_r3553.v05-01
group.perf-tau.TauPi0Rec.D3PD.170202.Pythia8_AU2CTEQ6L1_Zprime500tautau.recon.ESD.e1176_s1479_s1470_r3553.v05-01
group.perf-tau.TauPi0Rec.D3PD.170203.Pythia8_AU2CTEQ6L1_Zprime750tautau.recon.ESD.e1176_s1479_s1470_r3553.v05-01
group.perf-tau.TauPi0Rec.D3PD.170204.Pythia8_AU2CTEQ6L1_Zprime1000tautau.recon.ESD.e1176_s1479_s1470_r3553.v05-01
group.perf-tau.TauPi0Rec.D3PD.170205.Pythia8_AU2CTEQ6L1_Zprime1250tautau.recon.ESD.e1176_s1479_s1470_r3553.v05-01
data samples:
group.perf-tau.TauPi0Rec.D3PD.periodA.physics_JetTauEtmisss.PhysCont.DESD_CALJET.repro14_v01.v05-01
group.perf-tau.TauPi0Rec.D3PD.periodB.physics_JetTauEtmisss.PhysCont.DESD_CALJET.repro14_v01.v05-01
group.perf-tau.TauPi0Rec.D3PD.periodI.physics_JetTauEtmisss.PhysCont.DESD_CALJET.repro14_v01.v05-01

**Table B.1:** Samples of events generated by MC or extracted from a QCD-di-jet selection in 2012 data were used within this thesis.



## C BDT Configuration

The options chosen to configure the BDT exploited in the tau identification, are summarised in Table C.1. The underlying framework is TMVA-v4.2.0.

NormMode	EqualNumEvents
NTrees	100
option	value
MaxDepth	8
nCuts	200
MinNodeSize	0.1
BoostType	AdaBoost
AdaBoostBeta	0.2
UseYesNoLeaf	FALSE
SeparationType	GiniIndex
PrunMethod	CostComplexity
PruneStrength	60
PruningValFraction	0.5

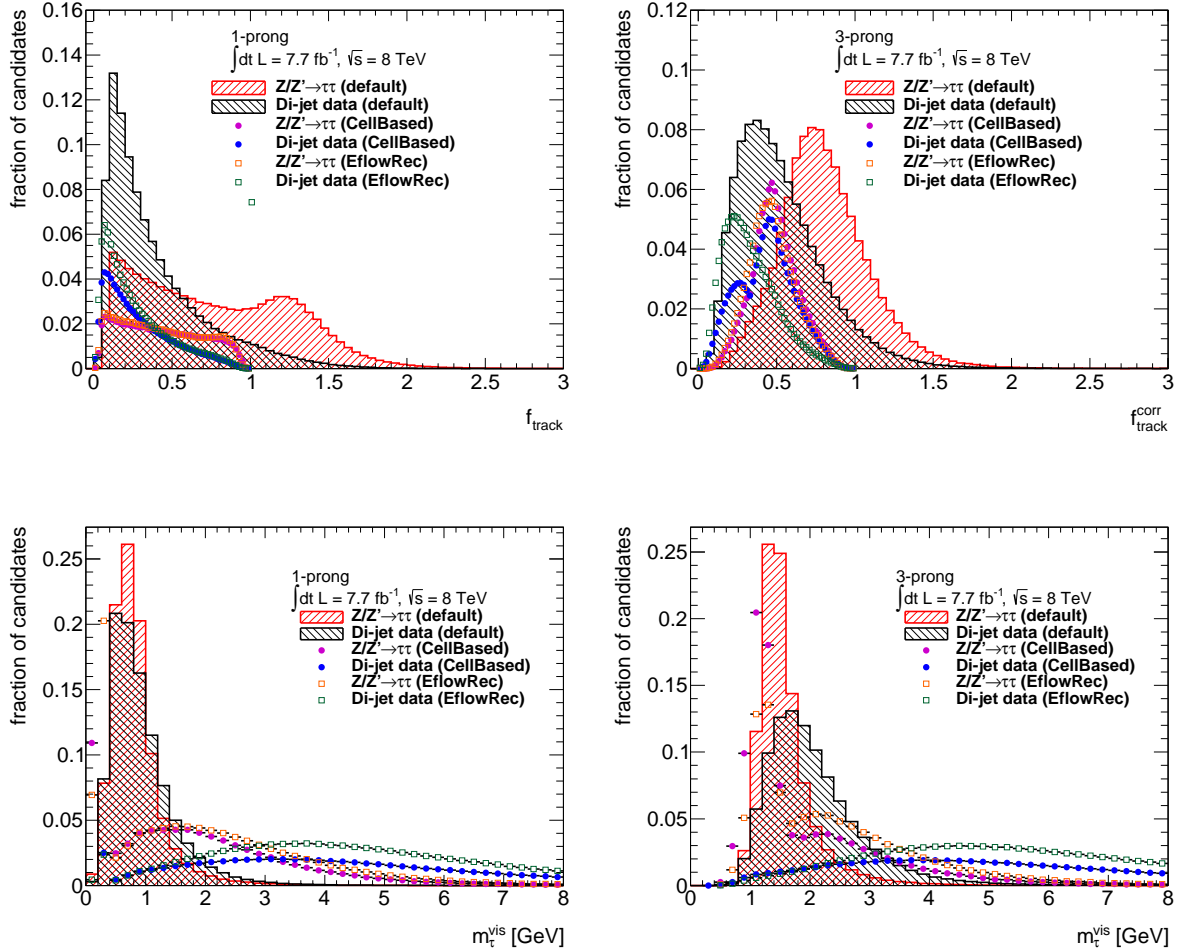
**Table C.1:** BDT configuration used for tau identification.

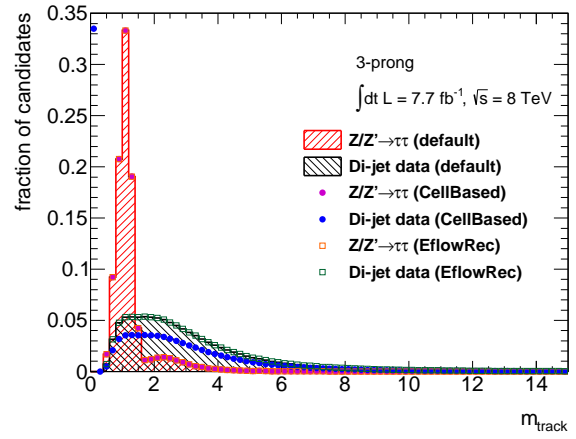
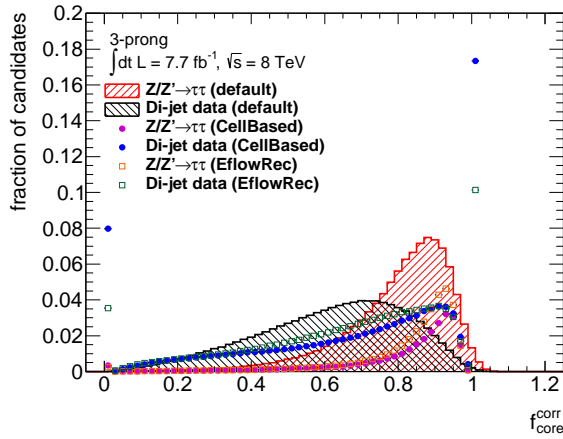
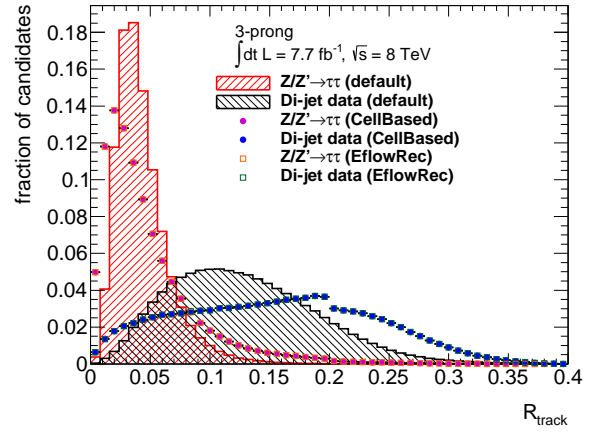
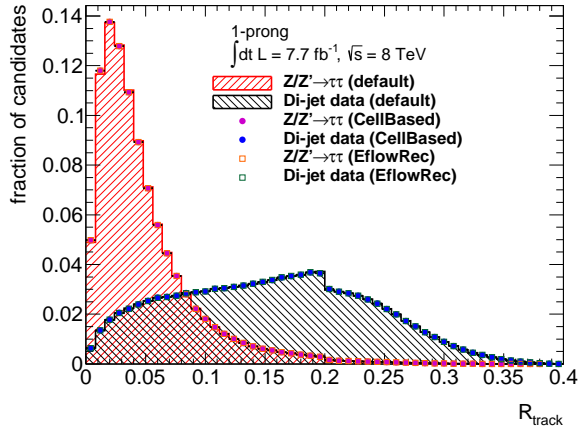
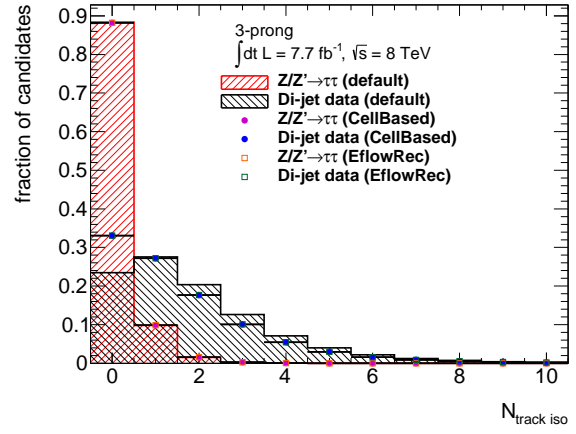
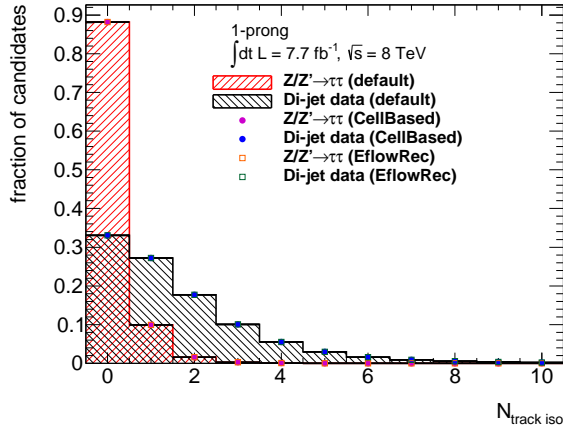
## D Auxiliary Material for Optimisation of Tau Identification

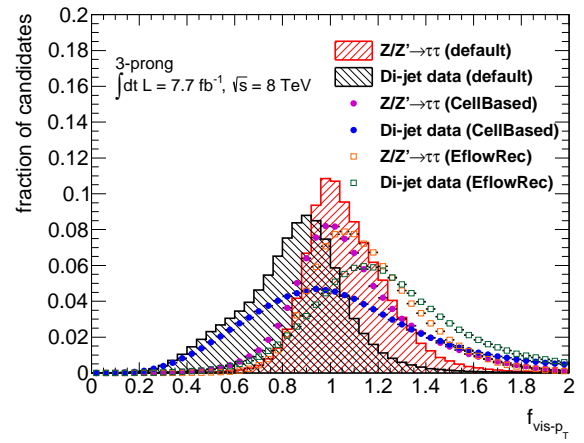
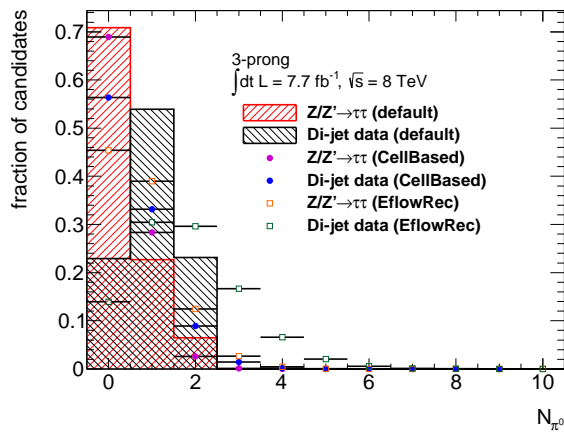
This appendix provides auxiliary material for the optimisation of tau identification. Section D.1 presents the distributions of the default identification variables recalculated with the substructure algorithms. The performance plots for the extension to the isolation cone of the variable sets based on the EflowRec algorithm, not shown in Section 6.2.2, are summarised in Appendix D.2. All distributions of variables defined in Section 6.2.3 and the correlation matrices of the variables included in the corresponding top 20 ranking are presented in Section D.3 and D.4, respectively.

### D.1 Recalculated Variables

In the following The distribution of the identification variables recalculated with the substructure algorithms CellBased and EflowRec, are presented. The distributions already shown in Section 6.2.2 are not depicted here. The distributions of signal events obtained from MC simulation with tau candidates required to match to generated 1-/3-prong decays and background events extracted from a QCD di-jet selection in data based on the default calculation are illustrated as red and black dashed histograms, respectively. This is indicated by the magenta and blue filled dots for the recalculation based on the CellBased algorithm and as orange and green empty squares for the EflowRec algorithm.

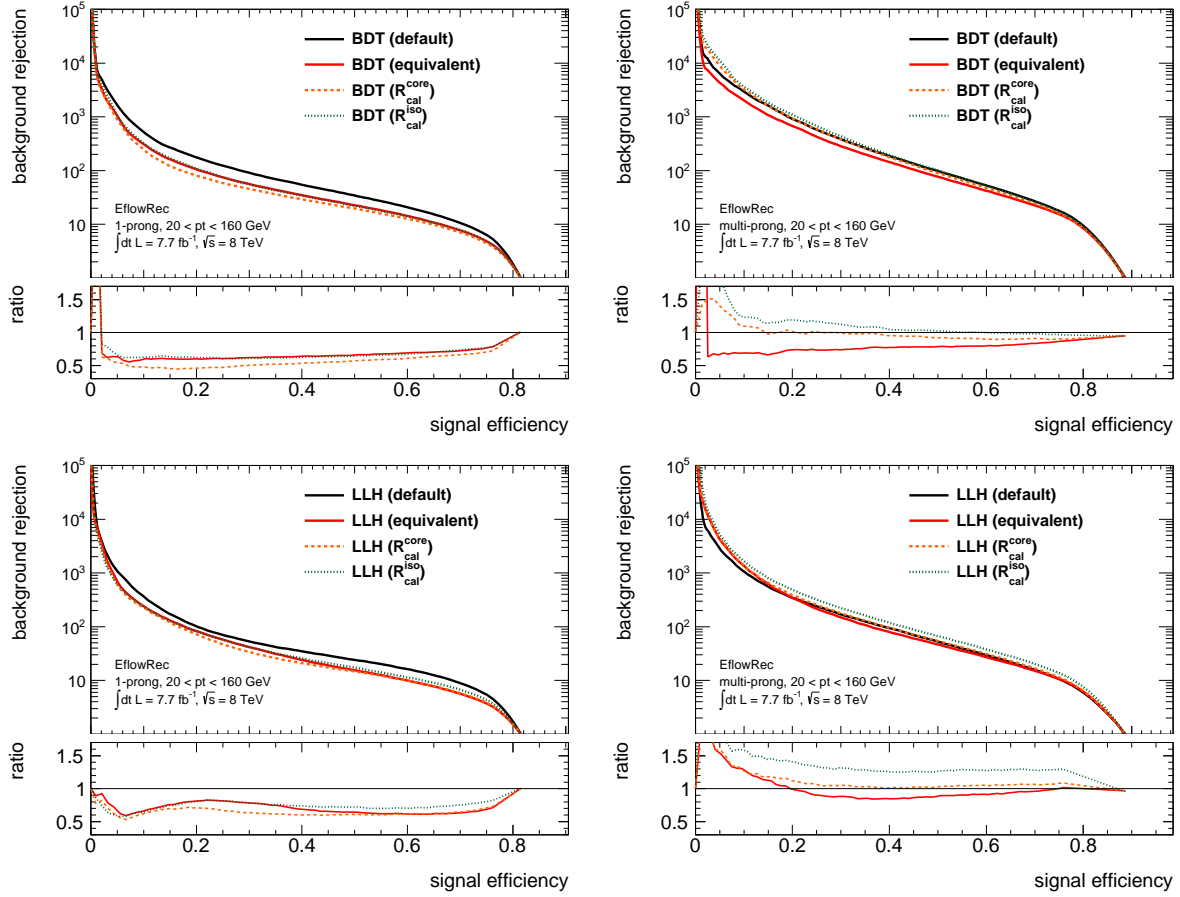






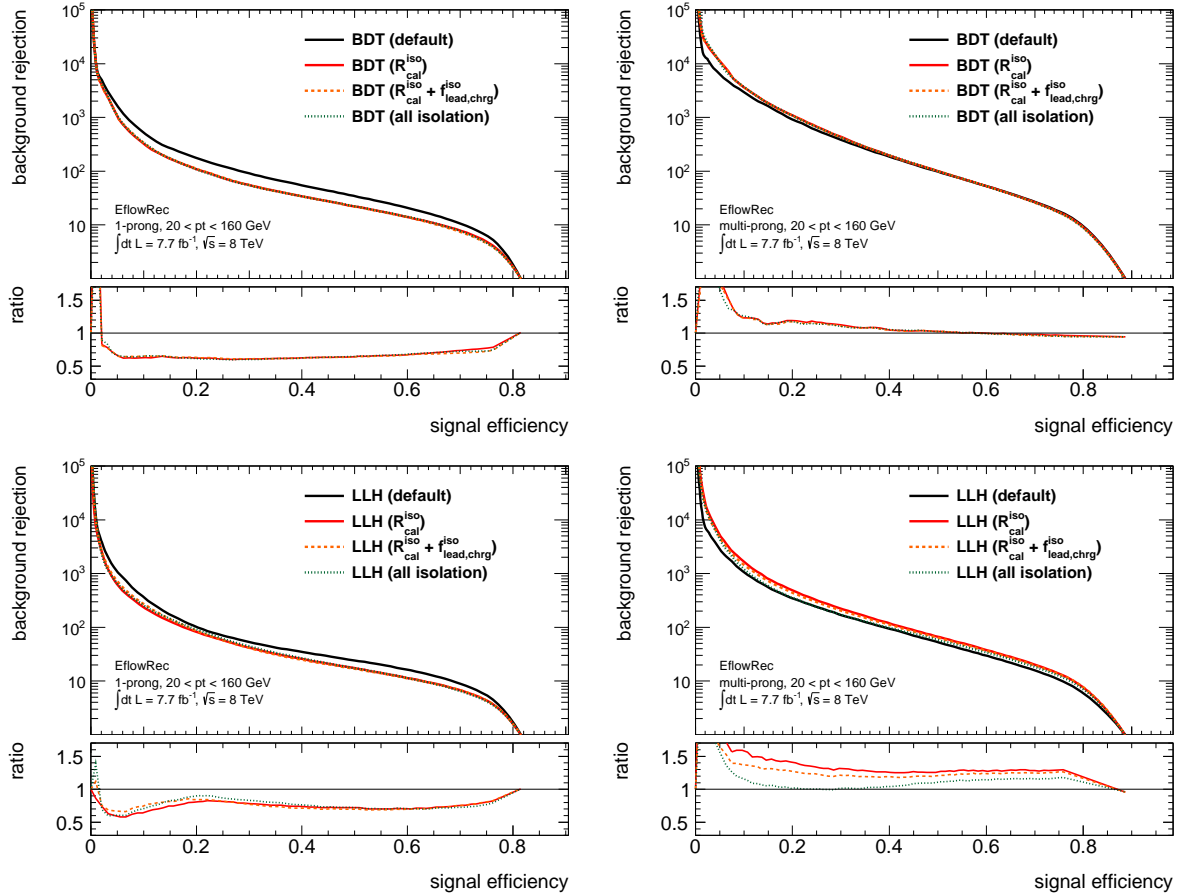
## D.2 Isolation Variables

This section summarises the background rejection against signal efficiency for an identification variable set including the quantities  $R_{\text{cal}}^{\text{core}}$  or  $R_{\text{cal}}^{\text{iso}}$  instead of  $f_{\text{core}}$  based on the EflowRec algorithm after the application of a pile-up correction introduced in Section 6.2.2. Figure D.1 depicts this for the BDT and LLH approach.



**Figure D.1:** Background rejection against signal efficiency for a BDT- (top) and LLH-based (bottom) tau identification exploiting the default (black), recalculated (red) variable sets and a recalculated variable set including  $R_{\text{cal}}^{\text{core}}$  (orange dashed) or  $R_{\text{cal}}^{\text{iso}}$  (green dotted) based on the substructure algorithm EflowRec. Whereby, pile-up corrections are taken into account for the variable calculation. The performance for the evaluation on 1-prong and multi-prong tau decays are shown in the left and right figure, respectively. The ratios below each plot depict the according deviations of the background rejection for a given signal efficiency w.r.t. the default approach. The same colour scheme is applied.

The quantity  $f_{\text{track}}$  was expanded to the isolation cone, referred as  $f_{\text{lead,chg}}^{\text{iso}}$ . The resulting background rejection w.r.t. the signal efficiency for a BDT- and LLH-based tau identification is shown in Figure D.2. Both approaches exploit a variable set which is recalculated with the EflowRec algorithm, whereby  $f_{\text{core}}$  is replaced by  $R_{\text{cal}}^{\text{iso}}$  and  $f_{\text{track}}$  is expanded to  $f_{\text{lead,chg}}^{\text{iso}}$ . In addition those results are also presented for a variable set exclusively based on variables defined in the isolation cone.

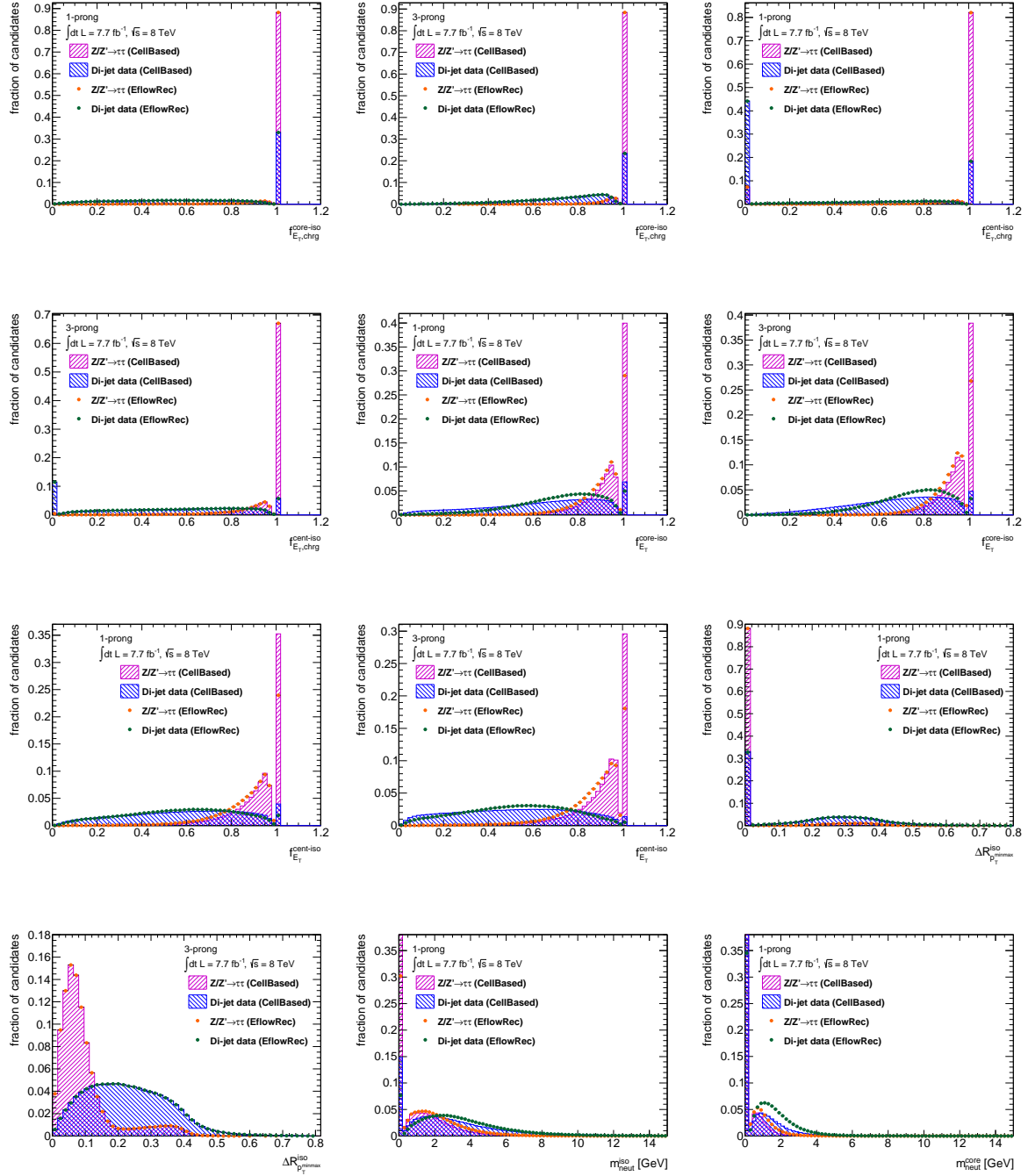


**Figure D.2:** Background rejection against signal efficiency for a BDT- (top) and LLH-based (bottom) tau identification exploiting the default (black) variable sets, recalculated variable sets including  $R_{\text{cal}}^{\text{iso}}$  (red),  $R_{\text{cal}}^{\text{iso}}$  and  $f_{\text{lead,chg}}^{\text{iso}}$  (orange dashed) and a set of exclusively isolation variables (green dotted) based on the substructure algorithm EflowRec. The performance for the evaluation on 1-prong and multi-prong tau decays are shown in the left and right figure, respectively. The ratios below each plot depict the according deviations of the background rejection for a given signal efficiency w.r.t. the default approach. The same colour scheme is applied.

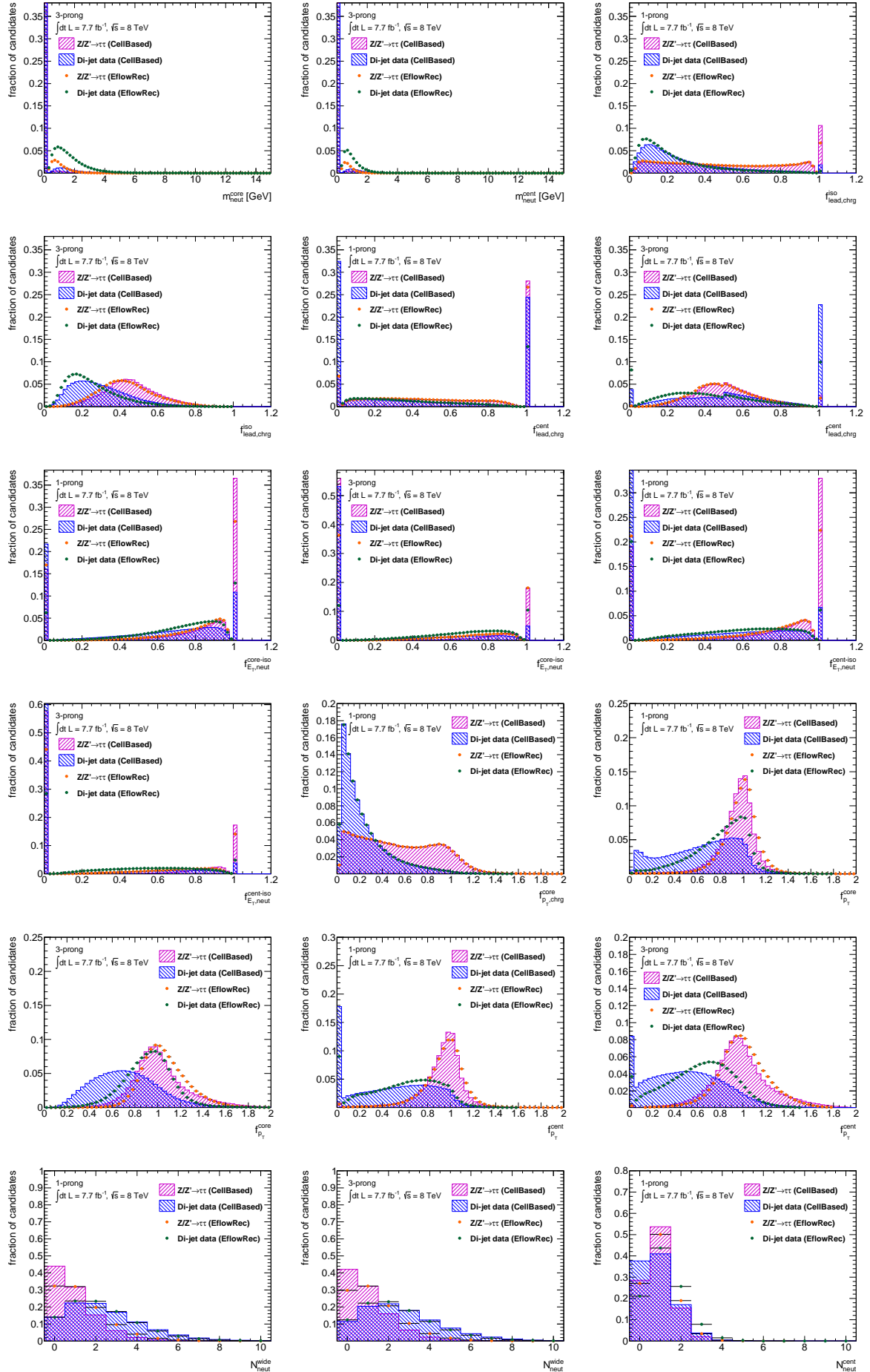
### D.3 Further Variables

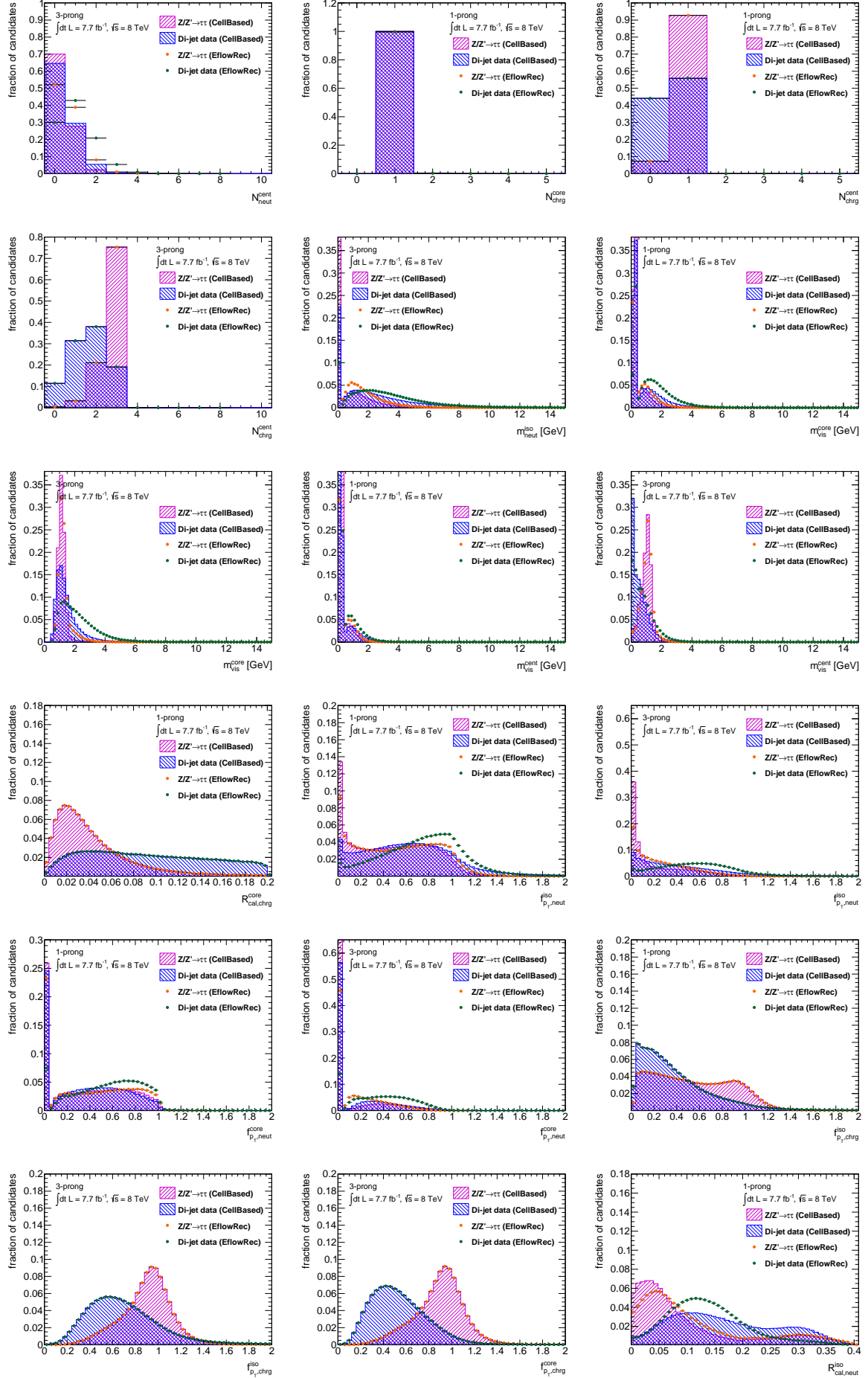
The following figures depict a meaningful selection of distributions of all variables introduced in Section 6.2.3 which are not already shown in Section 6.2.1–6.2.3 and Appendix D.1.

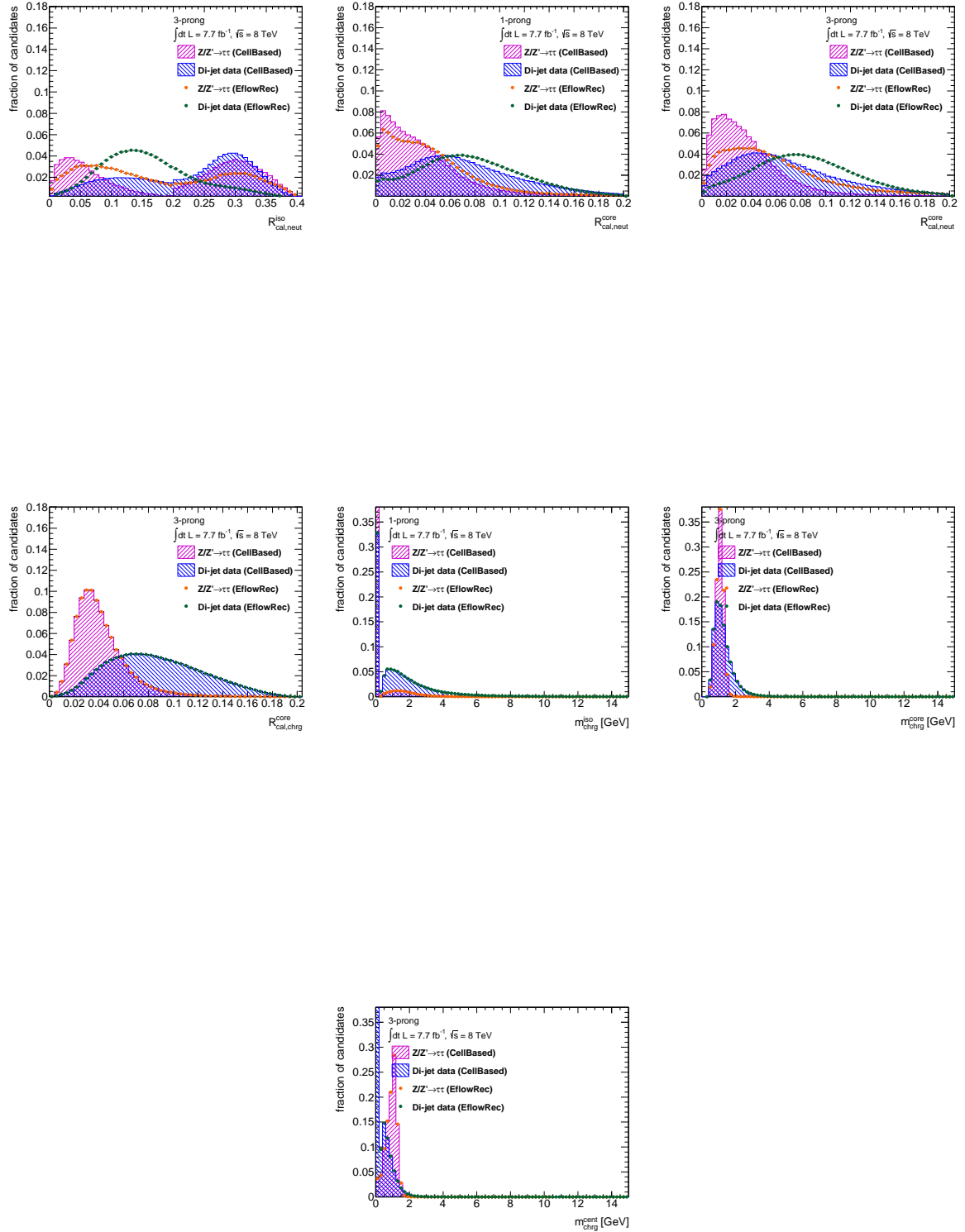
The distribution of signal events is obtained from MC simulation with tau candidates required to match to generated 1-/3-prong decays and the on of background events extracted from a QCD di-jet selection in data. The first is illustrated as magenta and the second as blue dashed histogram for the calculation with the CellBased algorithm. This is indicated in orange and green for the EflowRec algorithm.







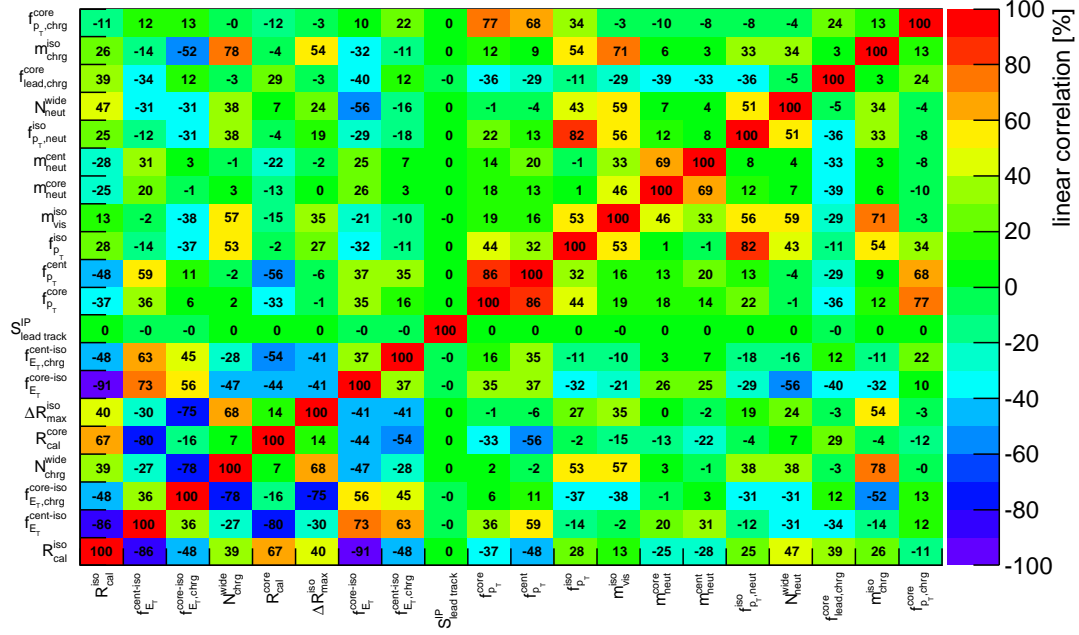




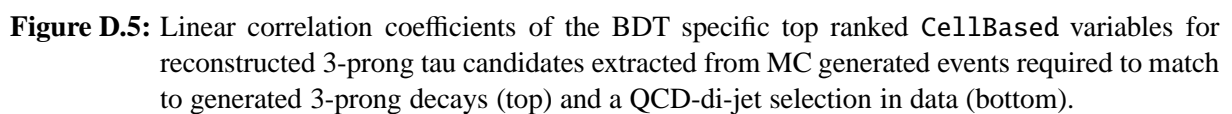
**Figure D.3:** Selection of distributions of investigated substructure based variables.

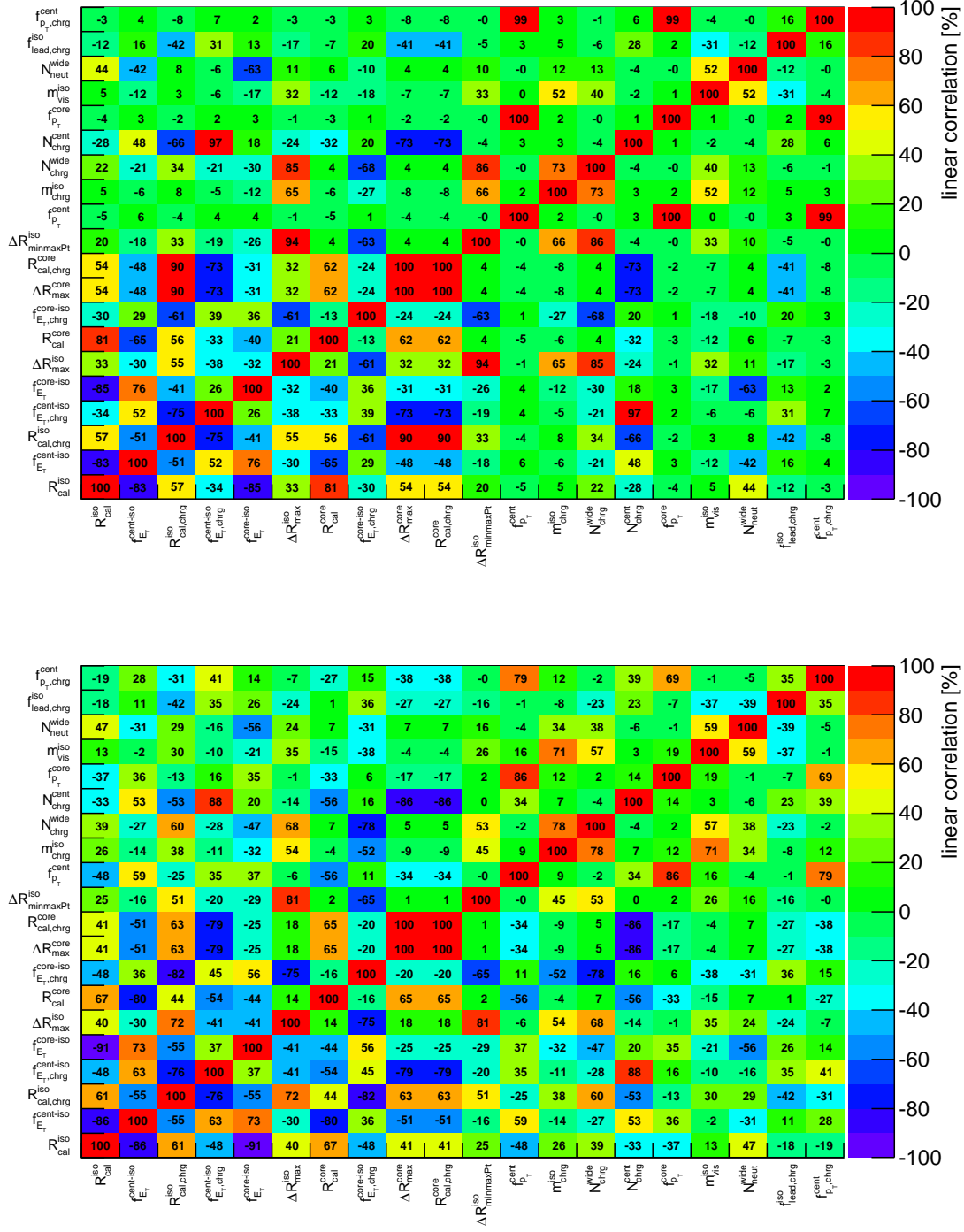
## D.4 Correlation Matrices

Figure D.4 – D.11 show the correlation matrices for the 20 highest-ranked variables based on the calculation with the substructure algorithms CellBased and EflowRec. A general and a BDT specific ranking are used to select those variables. The matrices are shown separately for 1- and 3-prong extracted from MC generated events and a QCD-di-jet selection in data.

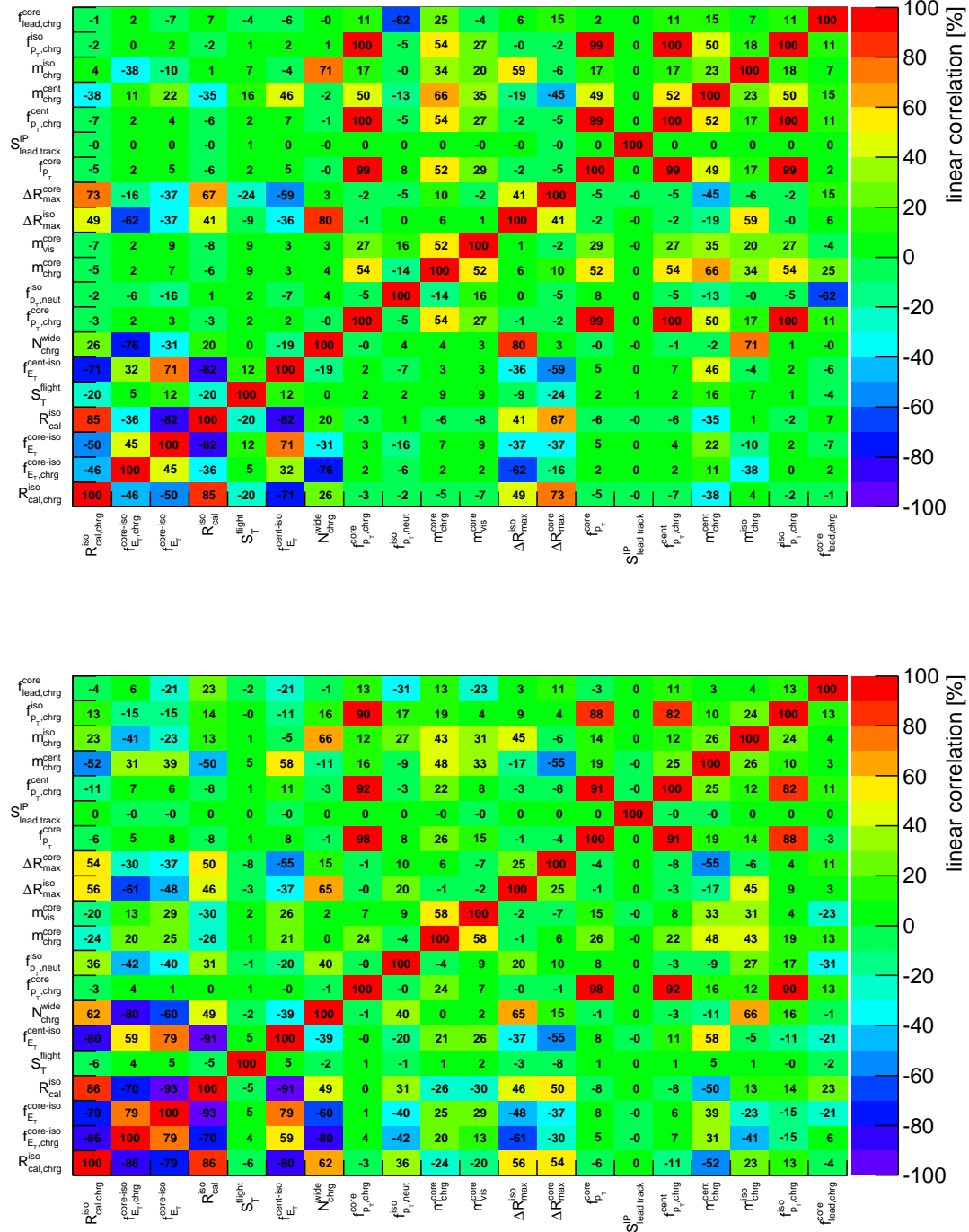


**Figure D.4:** Linear correlation coefficients of the BDT specific top ranked CellBased variables for reconstructed 1-prong tau candidates extracted from a QCD-di-jet selection in data.



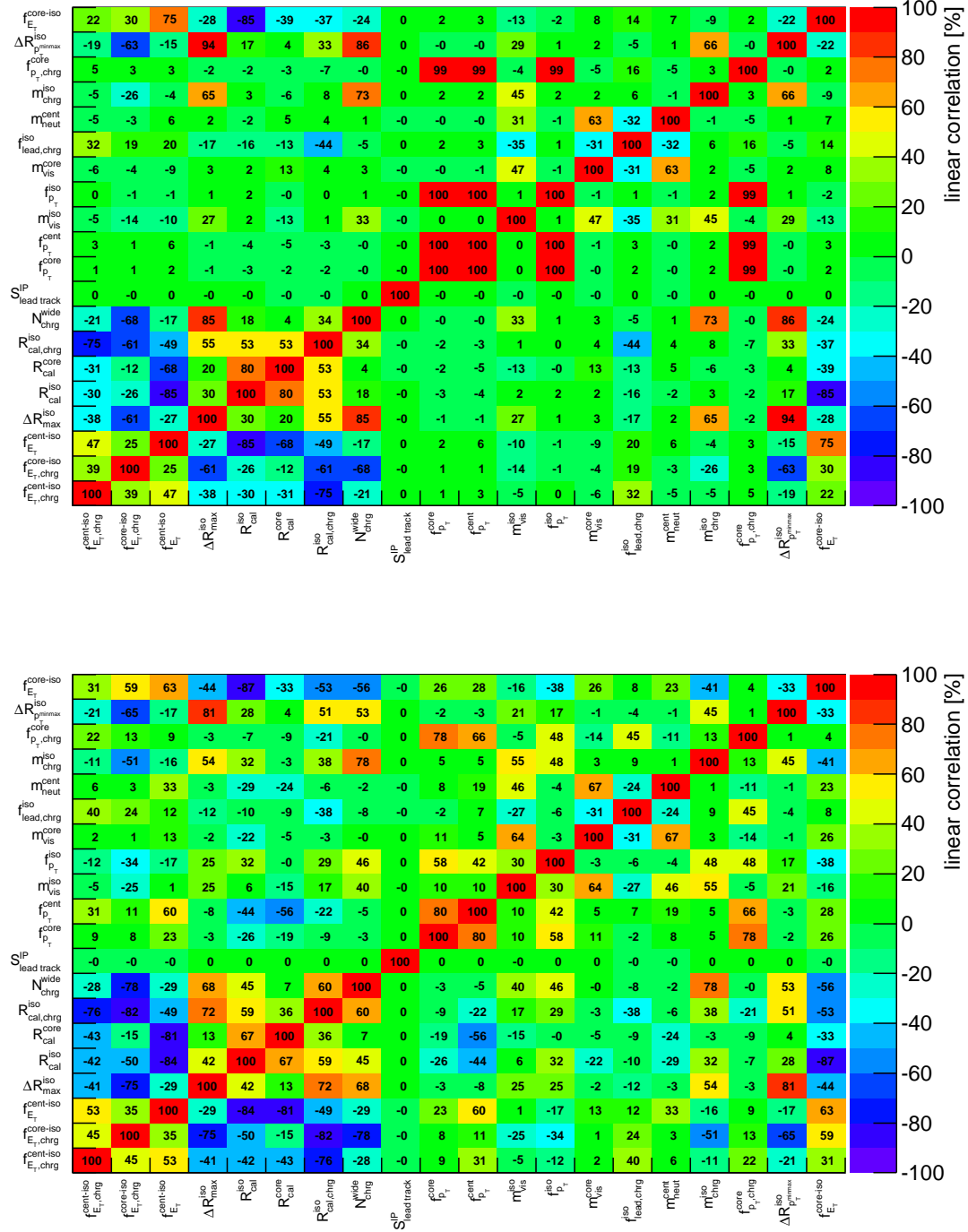


**Figure D.6:** Linear correlation coefficients of the generic top ranked CellBased variables for reconstructed 1-prong tau candidates extracted from MC generated events required to match to generated 1-prong decays (top) and a QCD-di-jet selection in data (bottom).

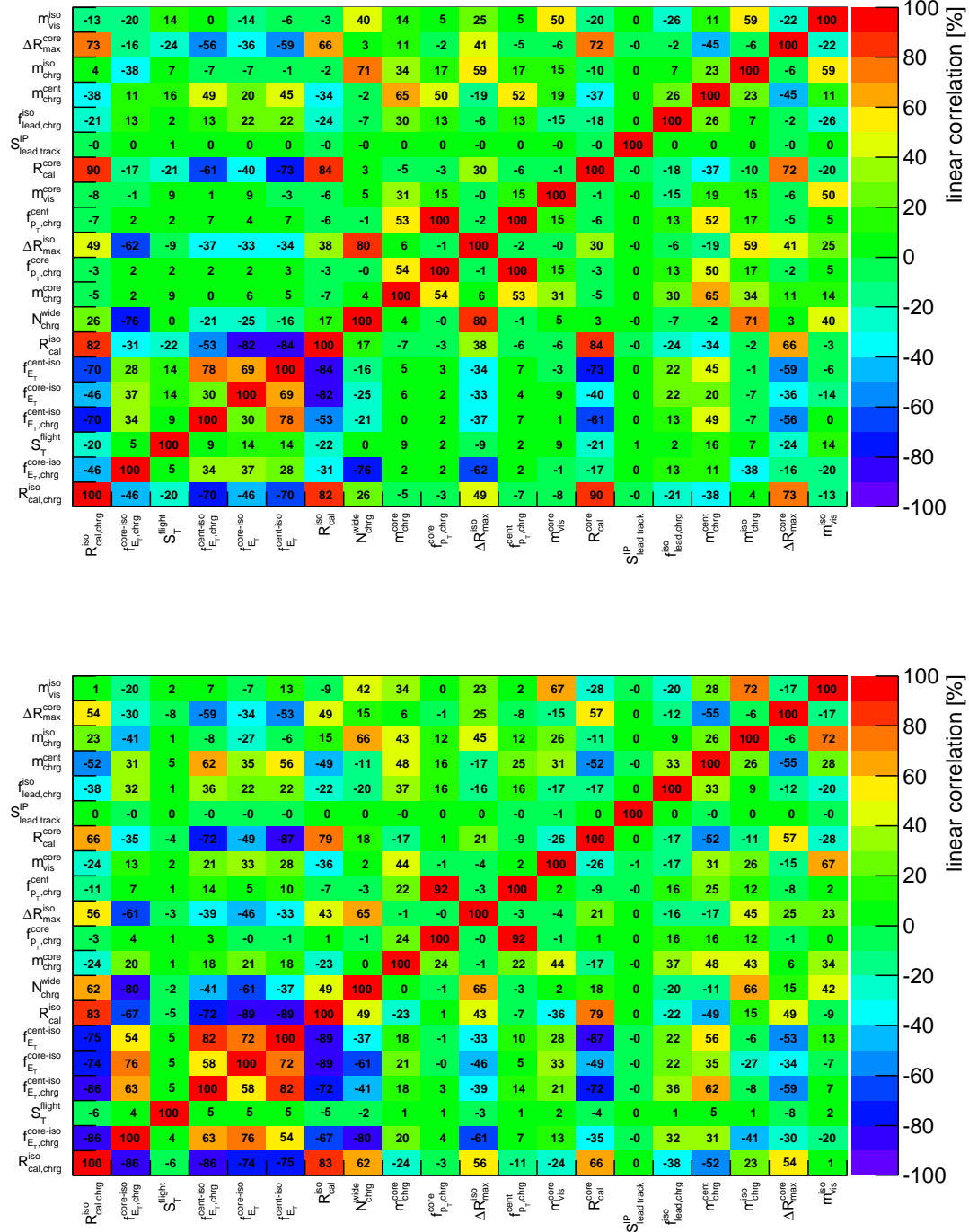


**Figure D.7:** Linear correlation coefficients of the generic top ranked CellBased variables for reconstructed 3-prong tau candidates extracted from MC generated events required to match to generated 3-prong decays (top) and a QCD-di-jet selection in data (bottom).

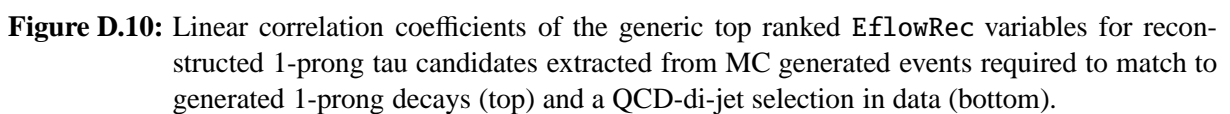


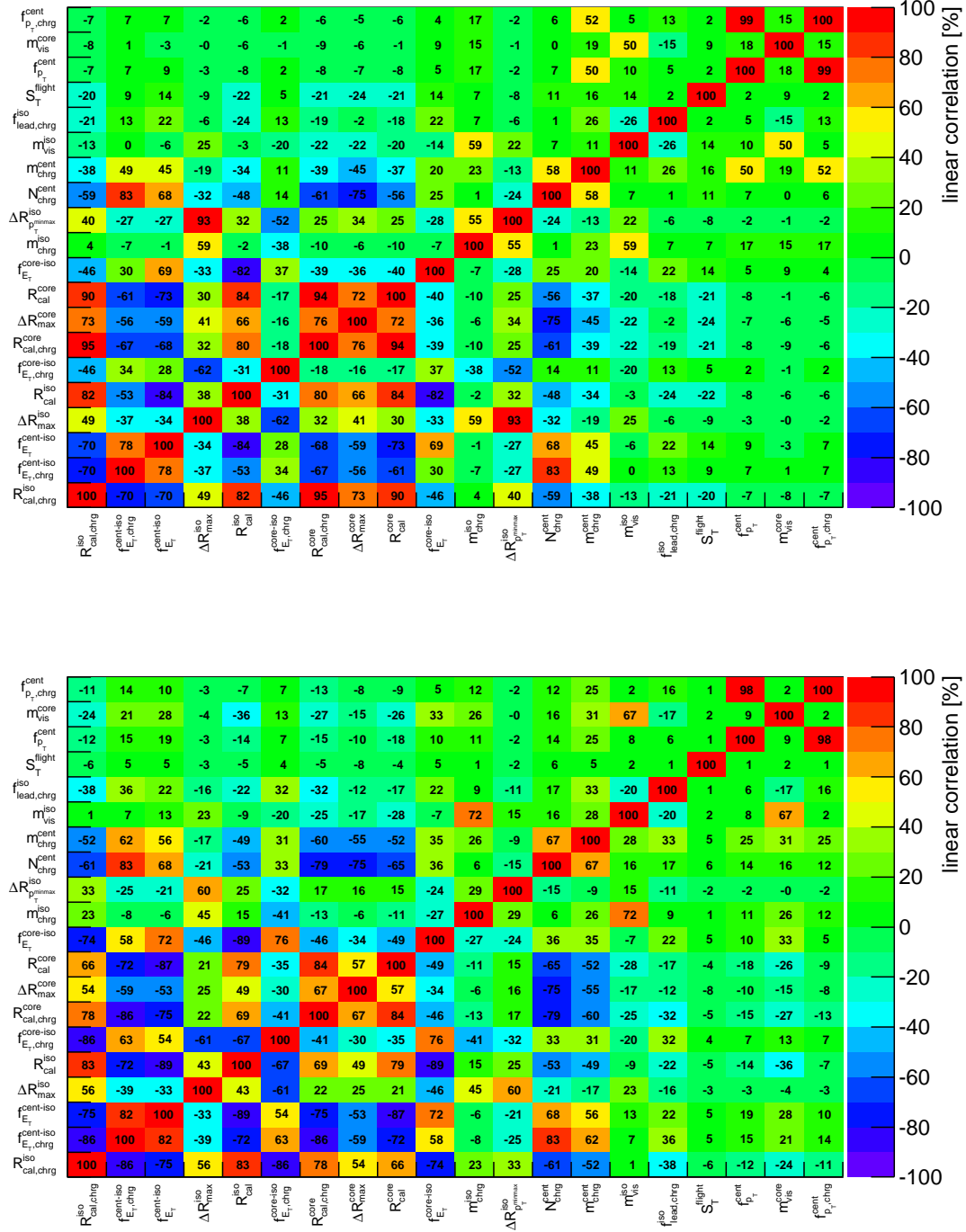


**Figure D.8:** Linear correlation coefficients of the BDT specific top ranked EflowRec variables for re-constructed 1-prong tau candidates extracted from MC generated events required to match to generated 3-prong decays (top) and a QCD-di-jet selection in data (bottom).



**Figure D.9:** Linear correlation coefficients of the BDT specific top ranked EflowRec variables for re-constructed 3-prong tau candidates extracted from MC generated events required to match to generated 3-prong decays (top) and a QCD-di-jet selection in data (bottom).





**Figure D.11:** Linear correlation coefficients of the generic top ranked EflowRec variables for reconstructed 3-prong tau candidates extracted from MC generated events required to match to generated 3-prong decays (top) and a QCD-di-jet selection in data (bottom).



## Bibliography

ATLAS internal documents are listed for completeness, although not publically available.

- [1] S. L. Glashow, *Partial Symmetries of Weak Interactions*, [Nucl. Phys. \*\*22\*\* \(1961\) 579 – 588](#).
- [2] S. Weinberg, *A Model of Leptons*, [Phys. Rev. Lett. \*\*19\*\* \(1967\) no. 21, 1264 – 1266](#).
- [3] A. Salam, *Weak and Electromagnetic Interactions*, [Conf. Proc. \*\*C680519\*\* \(1968\) 367–377](#).
- [4] UA1 Collaboration, *Experimental observation of lepton pairs of invariant mass around 95 GeV/c<sup>2</sup> at the CERN SPS collider*, [Phys. Lett. \*\*B126\*\* \(1983\) no. 5, 398–410](#).
- [5] UA2 Collaboration, *Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN  $\bar{p}p$  collider*, [Phys. Lett. \*\*B122\*\* \(1983\) no. 5–6, 476–485](#).
- [6] R. Brandelik et al., *Evidence for planar events in  $e^+e^-$  annihilation at high energies*, [Phys. Lett. \*\*B86\*\* \(1979\) no. 2, 243–249](#).
- [7] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, [Phys. Rev. Lett. \*\*13\*\* \(1964\) 321 – 323](#).
- [8] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, [Phys. Lett. \*\*12\*\* \(1964\) 132 – 133](#).
- [9] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, [Phys. Rev. Lett. \*\*13\*\* \(1964\) 508 – 509](#).
- [10] P. W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, [Phys. Rev. \*\*145\*\* \(1966\) 1156 – 1163](#).
- [11] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, [Phys. Rev. Lett. \*\*13\*\* \(1964\) 585 – 587](#).
- [12] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, [Phys. Lett. \*\*B716\*\* \(2012\) no. 1, 1–29](#).
- [13] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, [Phys. Lett. \*\*B716\*\* \(2012\) no. 1, 30–61](#), [arXiv:1207.7235 \[hep-ex\]](#).
- [14] L. R. Evans and P. Bryant, *LHC Machine*, [JINST \*\*3\*\* \(2008\) S08001](#).
- [15] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, [JINST \*\*3\*\* \(2008\) S08003](#).
- [16] CMS Collaboration, *The CMS Experiment at the CERN LHC*, [JINST \*\*3\*\* \(2008\) S08004](#).
- [17] WMAP Collaboration, *Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results*, [arXiv:1212.5225 \[astro-ph\]](#).
- [18] Planck Collaboration, *Planck 2013 results. I. Overview of products and scientific results*, [arXiv:1303.5062 \[astro-ph\]](#).
- [19] K. A. Olive et al. (Particle Data Group), *Review of Particle Physics, 2014-2015*, [Chin. Phys. \*\*C38\*\* \(2014\) 090001](#).
- [20] S. P. Martin, *A Supersymmetry Primer*, [arXiv:9709356 \[hep-ph\]](#).

- [21] D. Kazakov, *Beyond the Standard Model (In Search of Supersymmetry)*, [arXiv:0012288 \[hep-ph\]](#).
- [22] P. Fayet, *Supersymmetry and Weak, Electromagnetic and Strong Interactions*, *Phys. Lett.* **B64** (1976) 159.
- [23] P. Fayet, *Spontaneously Broken Supersymmetric Theories of Weak, Electromagnetic and Strong Interactions*, *Phys. Lett.* **B69** (1977) 489.
- [24] P. Fayet, *Relations Between the Masses of the Superpartners of Leptons and Quarks, the Goldstino Couplings and the Neutral Currents*, *Phys. Lett.* **B84** (1979) 416.
- [25] G. Farrar and P. Fayet, *Phenomenology of the Production, Decay, and Detection of New Hadronic States Associated with Supersymmetry*, *Phys. Lett.* **B76** (1978) 575–579.
- [26] Y. A. Golfand and E. P. Likhtman, *Extension of the Algebra of Poincare Group Generators and Violation of  $p$  Invariance*, *JETP Lett.* **13** (1971) 323–326.
- [27] J. Wess and B. Zumino, *Supergauge Transformations in Four-Dimensions*, *Nucl. Phys.* **B70** (1974) 39–50.
- [28] E. Noether, *Invarianten beliebiger Differentialausdrücke*, *Gött. Nachr.* (1918) 37–44.
- [29] E. Noether, *Invariante Variationsprobleme*, *Gött. Nachr.* (1918) 235–357.
- [30] W. Heisenberg, *Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik*, *Zeitschrift für Physik* **43** (1927) 3–4.
- [31] D. Griffiths, *Introduction to Elementary Particles*. Wiley-VCH Verlag, 2008.
- [32] P. J. Mohr et al., *CODATA Recommended Values of the Fundamental Physical Constants: 2010*, *Rev.Mod.Phys.* **84** (2012) 1527–1605, [arXiv:1203.5425 \[physics.atom-ph\]](#).
- [33] T. Nakano and K. Nishijima, *Charge Independence for V-particles*, *Prog. Theor. Phys.* **10** (1953) no. 5, 581–582.
- [34] M. Gell-Mann, *The interpretation of the new particles as displaced charge multiplets*, *Il Nuovo Cimento* **4** (1956) no. 2, 848–866.
- [35] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, *Experimental Test of Parity Conservation in Beta Decay*, *Phys. Rev.* **105** (1957) no. 4, 1413–1415.
- [36] M. Kobayashi and T. Maskawa, *CP-Violation in the Renormalizable Theory of Weak Interaction*, *Prog. Theor. Phys.* **49** (1973) 652–657.
- [37] Z. Maki, M. Nakagawa, and S. Sakata, *Remarks on the Unified Model of Elementary Particles*, *Progress of Theoretical Physics* **28** (1962) 870–880.
- [38] G. Bhattacharyya, *A Pedagogical Review of Electroweak Symmetry Breaking Scenarios*, *Rept. Prog. Phys.* **74** (2011) 026201, [arXiv:0910.5095 \[hep-ph\]](#).
- [39] ATLAS Collaboration, *Combined coupling measurements of the Higgs-like boson with the ATLAS detector using up to  $25\text{ fb}^{-1}$  of proton-proton collision data*, [ATLAS-CONF-2013-034](#).
- [40] ATLAS Collaboration, *Updated coupling measurements of the Higgs boson with the ATLAS detector using up to  $25\text{ fb}^{-1}$  of proton-proton collision data*, [ATLAS-CONF-2014-009](#).
- [41] ATLAS Collaboration, *Evidence for Higgs boson Yukawa couplings in the  $H \rightarrow \tau\tau$  decay mode with the ATLAS detector*, [ATLAS-CONF-2014-061](#).



- [42] CMS Collaboration, *Precise determination of the mass of the Higgs boson and studies of the compatibility of its couplings with the standard model*, [CMS-PAS-HIG-14-009](#), 2014.
- [43] L. Landau, *On the angular momentum of a two-photon system*, Dokl. Akad. Nauk. Ser. Fiz. **60** (1948) 207–209.
- [44] C.-N. Yang, *On the angular momentum of a two-photon system*, [Phys. Rev.](#) **77** (1950) 242–245.
- [45] K. Desch et al., *Probing the CP nature of the Higgs boson at linear colliders with tau spin correlations: The Case of mixed scalar-pseudoscalar couplings*, [Phys. Lett.](#) **B579** (2003) 157–164, [0307331 \[hep-ph\]](#).
- [46] S. Berge et al., *Determination of the Higgs CP mixing angle in the tau decay channels at the LHC including the Drell-Yan background*, [Eur. Phys. J.](#) **C74** (2014) 3164, [1408.0798 \[hep-ph\]](#).
- [47] H. Baer and X. Tata, *Weak Scale Supersymmetry: From Superfields to Scattering Events*. Cambridge University Press, 2006.
- [48] M. Morgenstern, *Search for heavy resonances decaying into the fully hadronic di-tau final state with the ATLAS detector*. PhD thesis, TU Dresden, 2014.
- [49] M. Gomez-Bock, M. Mondragón, M. Mühlleitner, M. Spira, and P. M. Zerwas, *Concepts of Electroweak Symmetry Breaking and Higgs Physics*, [arXiv:0712.2419 \[hep-ph\]](#).
- [50] LEP Working Group for Higgs boson searches, ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, *Search for Neutral MSSM Higgs Bosons at LEP*, [Eur. Phys. J.](#) **C47** (2006) 547–587, [arXiv:0602042 \[hep-ex\]](#).
- [51] A. Djouadi, *The anatomy of electroweak symmetry breaking Tome II: The Higgs bosons in the Minimal Supersymmetric Model*, [Phys. Rept.](#) **459** (2008) no. 1-6, 1–241.
- [52] P. Langacker, *The Physics of Heavy Z' Gauge Bosons*, [Rev. Mod. Phys.](#) **81** (2009) 1199–1228, [arXiv:0801.1345 \[hep-ph\]](#).
- [53] T. G. Rizzo, *Z' phenomenology and the LHC*, [arXiv:0610104 \[hep-ph\]](#). Published in Boulder, 2006, Colliders and Neutrinos (TASI 2006).
- [54] R. Diener, S. Godfrey, and T. A. Martin, *Unravelling an Extra Neutral Gauge Boson at the LHC using Third Generation Fermions*, [Phys. Rev.](#) **D83** (2011) 115008, [arXiv:1006.2845 \[hep-ph\]](#).
- [55] A. Leike, *The Phenomenology of extra neutral gauge bosons*, [Phys. Rept.](#) **317** (1999) 143–250, [arXiv:9805494 \[hep-ph\]](#).
- [56] S. King, S. Moretti, and R. Nevzorov, *Theory and Phenomenology of an Exceptional Supersymmetric Standard Model*, [Phys. Rev.](#) **D73** (2006) 035009, [arXiv:0510419 \[hep-ph\]](#).
- [57] R. S. Chivukula, M. Dugan, and M. Golden, *Electroweak Corrections in Technicolor Reconsidered*, [Phys. Lett.](#) **B292** (1992) 435–441, [arXiv:9207249 \[hep-ph\]](#).
- [58] G. Buchalla, G. Burdman, C. Hill, and D. Kominis, *GIM Violation and New Dynamics of the Third Generation*, [Phys. Rev.](#) **D53** (1996) 5185–5200, [arXiv:9510376 \[hep-ph\]](#).
- [59] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, [arXiv:1307.1347 \[hep-ph\]](#).
- [60] CMS Collaboration, *Evidence for the 125 GeV Higgs boson decaying to a pair of tau leptons*, [JHEP](#) **05** (2014) 104, [arXiv:1401.5041 \[hep-ex\]](#).

- [61] S. Heinemeyer, W. Hollik, and G. Weiglein, *The Masses of the Neutral CP-even Higgs Bosons in the MSSM: Accurate Analysis at the Two-Loop Level*, *Eur. Phys. J. C* **9** (1999) 343–366, [arXiv:9812472 \[hep-ph\]](#).
- [62] S. Heinemeyer, W. Hollik, and G. Weiglein, *FeynHiggs: a program for the calculation of the masses of the neutral CP-even Higgs bosons in the MSSM*, *Comput. Phys. Commun.* **124** (2000) 76–89, [arXiv:9812320 \[hep-ph\]](#).
- [63] G. G. Degrandi, S. Heinemeyer, W. Hollik, P. Slavich, and G. Weiglein, *Towards High-Precision Predictions for the MSSM Higgs Sector*, *Eur. Phys. J. C* **28** (2003) 133–143, [arXiv:0212020 \[hep-ph\]](#).
- [64] M. Frank, T. Hahn, S. Heinemeyer, W. Hollik, H. Rzehak, and G. Weiglein, *The Higgs Boson Masses and Mixings of the Complex MSSM in the Feynman-Diagrammatic Approach*, *JHEP* **02** (2007) 047, [arXiv:0611326 \[hep-ph\]](#).
- [65] ATLAS Collaboration, *Search for neutral Higgs bosons of the minimal supersymmetric standard model in pp collisions at  $\sqrt{s}=8$  TeV with the ATLAS detector*, *JHEP* **1411** (2014) 056, [arXiv:1409.6064 \[hep-ex\]](#).
- [66] M. Carena, S. Heinemeyer, O. Stål, C. Wagner, and G. Weiglein, *MSSM Higgs Boson Searches at the LHC: Benchmark Scenarios after the Discovery of a Higgs-like Particle*, *Eur. Phys. J. C* **73** (2013) 2552, [arXiv:1302.7033 \[hep-ph\]](#).
- [67] ATLAS Collaboration, *Search for charged Higgs bosons decaying via  $H^\pm \rightarrow \tau^\pm \nu$  in hadronic final states using pp collision data at  $\sqrt{s}=8$  TeV with the ATLAS detector*, *ATLAS-CONF-2014-050*.
- [68] ATLAS Collaboration, *A search for high-mass ditau resonances decaying in the fully hadronic final state in pp collisions at  $\sqrt{s}=8$  TeV with the ATLAS detector*, *ATLAS-CONF-2013-066*.
- [69] ATLAS public luminosity results, <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults>.
- [70] LHCb Collaboration, *The LHCb Experiment at the CERN LHC*, *JINST* **3** (2008) S08005.
- [71] ALICE Collaboration, *The ALICE Experiment at the CERN LHC*, *JINST* **3** (2008) S08002.
- [72] LHCf Collaboration, *The LHCf detector at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08006.
- [73] MoEDAL Collaboration, *Technical Design Report of the MoEDAL Experiment*, *CERN-LHCC-2009-006*, *MoEDAL-TDR-001*, 2009.
- [74] TOTEM Collaboration, *The TOTEM Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08007.
- [75] E. Lefèvre, *The CERN accelerator complex*, *CERN-DI-0812015*.
- [76] ATLAS Collaboration, *The ATLAS Inner Detector commissioning and calibration*, *Eur. Phys. J. C* **70** (2010) 787–821, [arXiv:1004.5293 \[hep-ex\]](#).
- [77] ATLAS Collaboration, *ATLAS Detector and Physics Performance: Technical Design Report*, *ATLAS-TDR-014*, *CERN/LHCC 99-14*, 1999.
- [78] V. Gribov and L. Lipatov,  *$e^+ e^-$  pair annihilation and deep inelastic  $e p$  scattering in perturbation theory*, *Sov. J. Nucl. Phys.* **15** (1972) 675–684.

- [79] Y. L. Dokshitzer, *Calculation of the structure functions for deep inelastic scattering and  $e^+e^-$  annihilation by perturbation theory in quantum chromodynamics (in Russian)*, Sov. Phys. JETP **46** (1977) 641–653.
- [80] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language*, Nucl. Phys. **B126** (1977) 298.
- [81] G. Soyez, *The SISCone and anti- $k_t$  jet algorithms*, [arXiv:0807.0021 \[hep-ph\]](#).
- [82] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, *Better Jet Clustering Algorithms*, JHEP **9708** (1997) 001, [arXiv:9707323 \[hep-ph\]](#).
- [83] M. Wobisch and T. Wengler, *Hadronization Corrections to Jet Cross Sections in Deep-Inelastic Scattering*, [arXiv:9907280 \[hep-ph\]](#).
- [84] S. D. D. Ellis and D. E. Soper, *Longitudinally invariant  $K_t$  clustering algorithms for hadron hadron collisions*, Nucl. Phys. **B406** (1993) 187–224.
- [85] S. D. D. Ellis and D. E. Soper, *Successive Combination Jet Algorithm For Hadron Collisions*, Phys. Rev. **D48** (1993) 3160–3166, [arXiv:9305266 \[hep-ph\]](#).
- [86] T. Sjöstrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 physics and manual*, JHEP **05** (2006) 026, [arXiv:0603175 \[hep-ph\]](#).
- [87] T. Gleisberg et al., *Event generation with SHERPA 1.1*, JHEP **02** (2009) 007, [arXiv:0811.4622 \[hep-ph\]](#).
- [88] T. Sjöstrand, S. Mrenna, and P. Skands, *A Brief Introduction to Pythia 8.1*, [arXiv:0710.3820 \[hep-ph\]](#).
- [89] N. Davidson, G. Nanava, T. Przydzinski, E. Richter-Was, and Z. Was, *Universal Interface of TAUOLA Technical and Physics Documentation*, Comput. Phys. Commun. **183** (2012) 821–843, [arXiv:1002.0543 \[hep-ph\]](#).
- [90] The ATLAS Computing Group, *ATLAS Computing*, ATLAS-TDR-017, CERN-LHCC-2005-022, 2005.
- [91] S. Agostinelli [GEANT4 Collaboration], *GEANT4: A simulation toolkit*, Nucl. Instr. and Meth. A **506** (2003) 250.
- [92] *LHC Performance and Statistics*, <http://lhc-statistics.web.cern.ch/LHC-Statistics/index.php>.
- [93] *COMA Period Description Report*, [https://atlas-tagservices.cern.ch/tagservices/RunBrowser/rBR\\_Period\\_Report.php](https://atlas-tagservices.cern.ch/tagservices/RunBrowser/rBR_Period_Report.php).
- [94] ATLAS Collaboration,  *$Z \rightarrow \tau\tau$  cross section measurement in  $pp$  collisions at 7 TeV with the ATLAS experiment*, ATLAS-CONF-2012-006.
- [95] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2007.
- [96] 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 \[hep-ex\]](#).
- [97] M. Cacciari, G. P. Salam, and G. Soyez, *The anti- $k_t$  jet clustering algorithm*, JHEP **04** (2008) 063, [arXiv:0802.1189 \[hep-ph\]](#).
- [98] W. Lampl, S. Laplace, D. Lelas, P. Loch, H. Ma, S. Menke, S. Rajagopalan, D. Rousseau, S. Snyder, and G. Unal, *Calorimeter Clustering Algorithms: Description and Performance*, ATL-LARG-PUB-2008-002.

- [99] T. Barillari et al., *Local Hadronic Calibration*, [ATL-LARG-PUB-2009-001](#).
- [100] ATLAS Collaboration, *Performance of the ATLAS Inner Detector Track and Vertex Reconstruction in the High Pile-Up LHC Environment*, [ATLAS-CONF-2012-042](#), 2012.
- [101] ATLAS Collaboration, *Performance of the Reconstruction and Identification of Hadronic Tau Decays in ATLAS with 2011 Data*, [ATLAS-CONF-2012-142](#).
- [102] D. W. Miller, A. Schwartzman, and D. Su, *Jet-Vertex Association Algorithm*, [ATL-COM-PHYS-2008-008](#).
- [103] ATLAS Collaboration, *Measurement of the jet fragmentation function and transverse profile in proton-proton collisions at a center-of-mass energy of 7 TeV with the ATLAS detector*, [Eur.Phys.J.C. 71 \(2011\) 1795](#), [arXiv:1109.5816 \[hep-ex\]](#).
- [104] 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](#).
- [105] M. Trottier-McDonald, *A Cluster-based Approach to Reconstructing  $\pi^0$ s in  $\tau$  Decays*, Talk given at the Tau Substructure mini-workshop, 2012. <https://indico.cern.ch/getFile.py/access?contribId=2&resId=0&materialId=slides&confId=194235>.
- [106] ATLAS Collaboration, *Performance of the Substructure Reconstruction of Hadronic Tau Decays with ATLAS*, 2015. To be published.
- [107] A. Hoecker et al., *TMVA - Toolkit for Multivariate Data Analysis*, ArXiv Physics e-prints (2007) , [arXiv:0703039 \[physics\]](#).
- [108] 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](#).
- [109] ATLAS Collaboration, *Identification of Hadronic Decays of Tau Leptons in 2012 Data with the ATLAS Detector*, [ATLAS-CONF-2013-064](#).
- [110] P. Rados, *Tau Trigger Efficiency Measurement from  $Z \rightarrow \tau_{\text{had}}\tau_{\mu}$  Events*, Talk given in informal Tau Trigger Slice meeting, 2013. <https://indico.cern.ch/getFile.py/access?contribId=6&resId=0&materialId=slides&confId=239543>.
- [111] P. Rados, *2013 Tau Trigger Scale Factors ( $Z \rightarrow \tau\tau \rightarrow \mu\tau_{\text{had}}$  tag-and-probe method) Emulated Trigger Menu Reference Slides*, Talk given in informal Tau Trigger Slice meeting, 2013. <https://indico.cern.ch/getFile.py/access?contribId=3&resId=2&materialId=slides&confId=242060>.
- [112] P. Sales de Bruin et al., *Tau ID recent developments*, Talk given at the Tau Working Group Workshop 2013, 2013. <https://indico.cern.ch/event/238609/session/1/contribution/25/material/slides/0.pdf>.

## List of Figures

2.1	Higgs potential $V(\phi)$	7
2.2	Observed local $p_0$ -value and best-fit values for the signal strength $\mu$	8
2.3	Dependence of the gauge coupling constants on the energy scale	10
2.4	Signal strength for the individual $H \rightarrow \tau\tau$ decay channels	13
3.1	Schematic overview of the accelerator complex at CERN	16
3.2	Cut-away view of the ATLAS detector.	17
3.3	Cut-away view of the inner detector of ATLAS and sketch of its structure traversed by a charged particle	18
3.4	Sketch of the ATLAS calorimeter system	20
3.5	Sketch of the muon system of the ATLAS detector	22
3.6	Overview of the ATLAS trigger system	22
3.7	Alignment of ATLAS forward detectors along the beamline	23
4.1	Total integrated luminosity delivered by the LHC, collected by the ATLAS detector and utilisable for physics analysis based on the entire data-taking period in 2012	26
4.2	Evolution of the peak average number of interactions per bunch crossing and of the peak luminosity per fill over the entire data-taking period in 2012 at the ATLAS detector	27
5.1	Comparison of the track selection efficiency versus the average number of interactions per bunch crossing for the default track selection and TJVA	30
5.2	Track radius	37
5.3	Leading track momentum fraction	37
5.4	Number of isolation tracks and maximum $\Delta R$	38
5.5	Leading track IP significance and transverse flight path significance	38
5.6	Core energy fraction	39
5.7	Mass of the track system	40
5.8	Transverse momentum ratio of track and neutral pion system w.r.t. the tau candidate	40
5.9	Number of reconstructed neutral pions	41
5.10	Visible mass of tau candidate	41
5.11	Sketch of a decision tree	43
5.12	BDT score distributions	43
5.13	Log-Likelihood score distributions	44
5.14	Signal efficiency vs. $p_{T,vis}^{gen}$ for the default tau ID	46
5.15	Signal efficiency vs. $\mu$ for the default tau ID	46
5.16	Background efficiency vs. $p_T$ for the default tau ID	47
5.17	Background efficiency vs. $\mu$ for the default tau ID	47
5.18	Background rejection vs. signal efficiency for the default tau ID	48
5.19	Scale factors	48
6.1	Background rejection against signal efficiency for several values of MinNodeSize	52
6.2	Overtraining test for a BDT configuration with MinNodeSize = 0.1 %	52
6.3	Signal efficiency vs. $p_{T,vis}^{gen}$ for the LLH including $\pi^0$ variables	53
6.4	Signal efficiency vs. $\mu$ for the LLH including $\pi^0$ variables	54
6.5	LLH and BDT signal scores vs. $p_T$	54
6.6	Background rejection vs. signal efficiency for the LLH with and without $\pi^0$ variables	55
6.7	LLH values of 1-prong tau candidates for the variables $f_{core}^{corr}$ and $m_{\tau}^{vis}$	55
6.8	Distributions and weights for $p_T$ and $\mu$ dependence	56
6.9	Effect of a kinematic reweighting	56
6.10	Impact of the correction of the p.d.f. calculation	57
6.11	$\Delta R_{max}$ , $f_{core}^{corr}$ , $N_{\pi^0}$ and $f_{vis-p_T}$ based on the default calculation, CellBased and EflowRec algorithm	59



6.12	Background rejection vs. signal efficiency for a BDT-based tau identification exploiting the default and recalculated variables set based on the substructure algorithm CellBased and EflowRec . . . . .	60
6.13	Background rejection vs. signal efficiency for a LLH-based tau identification exploiting the default and recalculated variables set based on the substructure algorithm CellBased and EflowRec . . . . .	60
6.14	Core calorimeter radius . . . . .	61
6.15	Isolation calorimeter radius . . . . .	61
6.16	Background rejection vs. signal efficiency for a BDT-based tau identification exploiting a variable set including $R_{\text{cal}}$ based on the CellBased algorithm . . . . .	62
6.17	Background rejection vs. signal efficiency for a BDT-based tau identification exploiting a variable set including $R_{\text{cal}}$ based on the EflowRec algorithm . . . . .	62
6.18	Background rejection vs. signal efficiency for a LLH-based tau identification exploiting a variable set including $R_{\text{cal}}$ based on the CellBased algorithm . . . . .	63
6.19	Background rejection vs. signal efficiency for a LLH-based tau identification exploiting a variable set including $R_{\text{cal}}$ based on the EflowRec algorithm . . . . .	63
6.20	Signal efficiency vs. $\mu$ for a BDT- and LLH-based tau identification for 1-prong tau candidates exploiting a variable set including $R_{\text{cal}}^{\text{iso}}$ based on the CellBased algorithm .	64
6.21	$R_{\text{cal}}^{\text{iso}}$ with and without pile-up corrections . . . . .	64
6.22	Signal efficiency vs. $\mu$ with and without pile-up corrections for a BDT- and LLH-based tau identification for 1-prong tau candidates exploiting a variable set including $R_{\text{cal}}^{\text{iso}}$ based on the CellBased algorithm . . . . .	65
6.23	Background rejection vs. signal efficiency for a BDT-based tau identification exploiting a variable set including $R_{\text{cal}}^{\text{iso}}$ based on the CellBased algorithm considering pile-up corrections . . . . .	65
6.24	Background rejection vs. signal efficiency for a LLH-based tau identification exploiting a variable set including $R_{\text{cal}}^{\text{iso}}$ based on the CellBased algorithm considering pile-up corrections . . . . .	66
6.25	Leading track momentum in core and isolation cone . . . . .	66
6.26	Background rejection vs. signal efficiency for a BDT-based tau identification exploiting a variable set including $R_{\text{cal}}^{\text{iso}}$ , $f_{\text{lead,chg}}^{\text{iso}}$ and all isolation variables based on the CellBased algorithm for 1- and 3-prong tau candidates . . . . .	67
6.27	Background rejection vs. signal efficiency for a LLH-based tau identification exploiting a variable set including $R_{\text{cal}}^{\text{iso}}$ , $f_{\text{lead,chg}}^{\text{iso}}$ and all isolation variables based on the CellBased algorithm for 1- and 3-prong tau candidates . . . . .	67
6.28	Distance between low- and high- $p_T$ track . . . . .	69
6.29	Background rejection vs. signal efficiency for a BDT-based tau identification exploiting a set of all substructure based variables for 1- and multi-prong tau candidates . . . . .	70
6.30	Background rejection vs. signal efficiency for a LLH-based tau identification exploiting a set of all substructure based variables for 1- and multi-prong tau candidates . . . . .	70
6.31	Background rejection vs. signal efficiency for a BDT-based tau identification exploiting a set top ranked variables based on the CellBased algorithm for 1- and multi-prong tau candidates . . . . .	72
6.32	Background rejection vs. signal efficiency for a BDT-based tau identification exploiting a set top ranked variables based on the EflowRec algorithm for 1- and multi-prong tau candidates . . . . .	73
6.33	Background rejection vs. signal efficiency for a LLH-based tau identification exploiting a set top ranked variables based on the CellBased algorithm for 1- and multi-prong tau candidates . . . . .	73

6.34	Background rejection vs. signal efficiency for a LLH-based tau identification exploiting a set top ranked variables based on the EflowRec algorithm for 1- and multi-prong tau candidates . . . . .	74
6.35	Linear correlation coefficient for the top ranked CellBased variables . . . . .	75
6.36	Background rejection vs. signal efficiency for a BDT-based tau identification exploiting a set top ranked variables considering a ranking, based on the CellBased algorithm for 1- and multi-prong tau candidates . . . . .	76
6.37	Background rejection vs. signal efficiency for a BDT-based tau identification exploiting a set top ranked variables considering a ranking, based on the EflowRec algorithm for 1- and multi-prong tau candidates . . . . .	77
6.38	Background rejection vs. signal efficiency for a LLH-based tau identification exploiting a set top ranked variables considering a ranking, based on the CellBased algorithm for 1- and multi-prong tau candidates . . . . .	77
6.39	Background rejection vs. signal efficiency for a LLH-based tau identification exploiting a set top ranked variables considering a ranking, based on the EflowRec algorithm for 1- and multi-prong tau candidates . . . . .	78
D.1	Background rejection vs. signal efficiency for a BDT- and LLH-based tau identification exploiting a variable set including $R_{cal}^{iso}$ based on the EflowRec algorithm considering pile-up corrections . . . . .	88
D.2	Background rejection vs. signal efficiency for a BDT- and LLH-based tau identification exploiting a variable set including $R_{cal}^{iso}$ , $f_{lead,chg}^{iso}$ and all isolation variables based on the EflowRec algorithm for 1- and 3-prong tau candidates . . . . .	89
D.3	Selection of distributions of investigated substructure based variables . . . . .	93
D.4	Linear correlation coefficients for the BDT specific top ranked CellBased variables for 1-prong decays . . . . .	94
D.5	Linear correlation coefficients for the BDT specific top ranked CellBased variables for 3-prong decays . . . . .	95
D.6	Linear correlation coefficients for the generic top ranked CellBased variables for 1-prong decays . . . . .	96
D.7	Linear correlation coefficients for the generic top ranked CellBased variables for 3-prong decays . . . . .	97
D.8	Linear correlation coefficients for the BDT specific top ranked EflowRec variables for 1-prong decays . . . . .	98
D.9	Linear correlation coefficients for the BDT specific top ranked EflowRec variables for 3-prong decays . . . . .	99
D.10	Linear correlation coefficients for the generic top ranked EflowRec variables for 1-prong decays . . . . .	100
D.11	Linear correlation coefficients for the generic top ranked EflowRec variables for 3-prong decays . . . . .	101





## List of Tables

2.1	Fundamental interactions of the SM and mediating gauge bosons . . . . .	4
2.2	Fermionic particle content of the Standard Model . . . . .	4
4.1	Overview of $pp$ collision data-taking periods of the ATLAS detector in 2012 . . . . .	27
5.1	Decay modes and branching ratios of leptonic and hadronic tau decays . . . . .	29
5.2	Reconstruction efficiencies of the CellBased, EflowRec, CellBased+PanTau and EflowRec+PanTau algorithm . . . . .	34
5.3	Summary of tau identification variables . . . . .	42
6.1	Summary of the top ranked variables . . . . .	71
6.2	Summary of the top ranked variables considering correlations . . . . .	75
B.1	Data and MC samples . . . . .	83
C.1	BDT configuration . . . . .	84



## Danksagung

Abschließend möchte ich mich bei all jenen bedanken, die mir während der Anfertigung meiner Diplomarbeit hilfreich zur Seite standen.

Zunächst möchte ich Prof. Dr. Arno Straessner danken, welcher als mein Betreuer, diese Arbeit ermöglichte und mich mit vielen hilfreichen Kommentaren und Ratschläge unterstützte. Des Weiteren bedanke ich mich bei Prof. Dr. Michael Kobel, welcher als Institutsdirektor des IKTP, angenehme Rahmenbedingungen schaffte. Insbesondere danke ich den Mitarbeitern der Dresdener Tau-Gruppe, Dirk Duschinger, Felix Friedrich, Lorenz Hauswald, Dr. Wolfgang Mader, Dr. Marcus Morgenstern und Sebastian Wahrmond, für die vielen hilfreichen Diskussionen physikalischer und programmiertechnischer Natur.

Weiterhin möchte ich allen Mitgliedern und besonders den Convenern der Tau-Arbeitsgruppe, Attilio Andreazza und Will Davey, für die angenehme Zusammenarbeit danken. Besonderer Dank geht an die Tau-ID-Experten, Pedro Sales de Bruin, Alexey Soloshenko und Almut Pingel für ihre schnelle und kompetente Hilfe bei jeglichen Fragen. Außerdem bedanke ich mich bei der Bonner Tau-Gruppe, speziell Christian Limbach, Peter Wagner und Benedict Winter für die zahlreichen Hinweise bei Fragen und Problemen. Spezieller Dank gilt auch den Entwicklern von TMVA.

Ganz besonders möchte ich meiner Familie danken, die mich immer unterstützte und immer an mich glaubte. Ohne eure finanzielle und moralische Unterstützung in den vergangenen Jahren, wären weder Physikstudium, noch diese Diplomarbeit möglich gewesen.

Mein größter Dank jedoch, gilt dir, Marcus. Du standest mir nicht nur mit deiner fachlichen Kompetenz bei zahlreichen physikalischen und programmiertechnischen Diskussionen konstruktiv zur Seite, sondern warst vor Allem als mein Freund immer für mich da und wohl die größte Unterstützung.



## **Erklärung**

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.

Stefanie Hanisch  
Dresden, 02.02.2014