

# Improved Identification for $\tau$ Leptons in the ATLAS Experiment

ALBERT-LUDWIGS-UNIVERSITÄT  
FREIBURG

FAKULTÄT FÜR MATHEMATIK UND PHYSIK

LEHRSTUHL FÜR EXPERIMENTELLE TEILCHENPHYSIK  
PROF. DR. KARL JAKOBS

*Author:*  
Nico MEYER

*Supervisor:*  
Prof. Dr. Karl JAKOBS

June 12, 2008



# Contents

<b>Introduction</b>	<b>1</b>
<b>1 Theoretical Background</b>	<b>5</b>
1.1 The Standard Model . . . . .	5
1.2 The Electroweak Model . . . . .	6
1.3 Quantum Chromo Dynamics . . . . .	9
1.4 Physics at pp Colliders . . . . .	10
1.5 The $\tau$ Lepton . . . . .	11
1.5.1 Physics searches in $\tau$ Lepton Final States . . . . .	13
<b>2 The ATLAS Experiment</b>	<b>16</b>
2.1 The Large Hadron Collider . . . . .	16
2.1.1 Luminosity . . . . .	17
2.1.2 Pile Up Events . . . . .	18
2.2 The ATLAS Experiment . . . . .	18
2.2.1 The Inner Detector . . . . .	20
2.2.2 The Calorimeter . . . . .	22
2.2.3 Muon System . . . . .	25
2.2.4 The Trigger System . . . . .	26
<b>3 Classification Methods</b>	<b>28</b>
3.1 Multivariate Techniques . . . . .	28
3.2 Likelihood Method . . . . .	29
3.3 Artificial Neural Networks . . . . .	32
<b>4 Reconstruction and Identification of <math>\tau</math> Leptons</b>	<b>39</b>
4.1 Properties of Hadronic $\tau$ Decay Products . . . . .	39
4.2 Reconstruction of Hadronically Decaying $\tau$ Leptons . . . . .	41
4.3 The tauRec Package . . . . .	42
4.3.1 Seeding . . . . .	43
4.3.2 Identification Variables . . . . .	43
4.3.3 Identification . . . . .	55
4.4 The tau1P3P Package . . . . .	60
4.4.1 Seeding . . . . .	62
4.4.2 Identification Variables . . . . .	63
4.4.3 Identification . . . . .	67

<b>5</b>	<b>Improvements for tauRec</b>	<b>68</b>
5.1	Topological Clustering . . . . .	68
5.2	Implementation of TopoJet Seeds . . . . .	70
5.3	Shower Shape Variables with TopoClusters . . . . .	72
5.4	New Likelihood Distributions . . . . .	76
5.5	Performance . . . . .	79
<b>6</b>	<b><math>\tau</math> Algorithm Merging</b>	<b>83</b>
6.1	Merging of tauRec and tau1P3P . . . . .	83
6.2	Technical Implementation . . . . .	85
6.3	Optimisation of Identification Performance . . . . .	88
6.3.1	Training of the Neural Networks . . . . .	88
6.3.2	Training Categories and Data Preparation . . . . .	90
6.3.3	Optimisation for Overlap Candidates . . . . .	93
6.3.4	Optimisation for tauRec Candidates . . . . .	114
6.3.5	Combined Performance . . . . .	118
<b>7</b>	<b>Summary</b>	<b>123</b>
<b>A</b>	<b>Comparison of tauRec Identification Variables between Release 12 and 13</b>	<b>I</b>
	<b>List of Tables</b>	<b>VI</b>
	<b>List of Figures</b>	<b>VII</b>
	<b>Bibliography</b>	<b>XI</b>

# Introduction

The goal of particle physics is to describe the fundamental constituents of nature and the forces that govern their behaviour. Theoretical and experimental efforts in the second half of the last century have led to the development of the Standard Model, which is one of the most successful theories in the history of science. Its predictions have been tested with extremely high precision, and they are in vast agreement with the experimental data.<sup>1</sup>

The Standard Model is a gauge invariant quantum field theory that describes all known elementary particles and three of the four forces working between them – electromagnetism, the weak and the strong force. Electromagnetism and the weak force are unified into the electroweak force. The fourth force – gravity – could not be successfully included in the same theoretical framework.

Forces are carried by particles with integral spin called bosons. These are the photon, the Z-boson and the W-bosons as carriers of the electroweak force and the gluons as carriers of the strong force. These forces act upon the fermionic matter particles with spin  $\frac{1}{2}$ , which exist in three families and two main groups, quarks and leptons, of which only the quarks feel the strong force. All these particles have been observed and most of their properties have been measured. There is however one additional particle which is predicted by the Standard Model but has not been observed in any experiment yet, the Higgs boson. The theory predicts all the properties of the Higgs boson apart from its mass. A lower limit has been set by direct searches at LEP and an upper limit of about 1 TeV can be derived from unitarity arguments.<sup>2</sup>

In spite of its success in making accurate predictions, the Standard Model cannot be a complete theory of nature. Most importantly the standard model does not describe the gravitational force, which governs the large scale behaviour of the universe. It is assumed by many physicists that some outstanding problems in cosmology can only be solved once a theory is formulated that unifies all four forces of nature. A first step towards such a *Theory of Everything* (ToE) would be a *Grand Unified Theory* (GUT) that unifies the electroweak sector with the strong force at very high energies above  $10^{16}$  GeV, in analogy to the unification of the electromagnetic force with the weak force above the weak scale.

Another challenge to the Standard Model is the so called *hierarchy problem*. An extreme fine tuning is necessary in the Higgs sector in order to keep the standard model valid up to energies at the Planck scale of  $1.2 \cdot 10^{19}$  GeV. Otherwise the Higgs mass diverges when calculating higher order loop corrections to its propagator.

---

<sup>1</sup>The only assured observation contradicting the Standard Model is that of non vanishing neutrino masses, but they do not constitute a fundamental problem to the Standard Model.

<sup>2</sup>In units  $\hbar \equiv c \equiv 1$ .

It is also known from astronomical and cosmological observations that not all matter in the universe consists of the particles described by the standard model. In fact only about 4% of the energy density observed in the universe is provided by the known particles. About 23% consists of *dark matter* and 73% of *dark energy*. The nature of the dark matter and dark energy is not known yet.

These and other problems lead to the development of a multitude of extensions to the standard model which try to address those problems, though none of them can solve all problems yet. The concept of Supersymmetry lies at the heart of many of those theories consistent with all known data. Supersymmetry postulates the existence of a supersymmetric partner with opposite spin statistics for each of the Standard Model particles, and also additional types of Higgs bosons are required in most supersymmetric models. Supersymmetric models can solve the hierarchy problem and allow the strong and the electroweak force to become of equal strength at the GUT scale. Furthermore in many supersymmetric models the lightest supersymmetric particle is a natural candidate for the dark matter.

The Large Hadron Collider (LHC) which has been built in the last several years at CERN<sup>3</sup> near Geneva will start its operation this summer. With its unprecedented centre of mass energy of 14 TeV and a instantaneous luminosity of  $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  it will open up new frontiers in high energy physics. The LHC will accelerate protons, which will collide at four interaction points where four large detectors are situated. The two multipurpose experiments ATLAS<sup>4</sup> and CMS<sup>5</sup> were designed to investigate a large range of possible new physics scenarios and especially to confirm the existence of the Higgs boson, should it be realised in nature.

The data that will come out of the experiments at the LHC in the next years will help to narrow down the large list of proposed theoretical models beyond the standard model and will hopefully produce new ideas for the further development of those theories that remain.

However, before this can happen many experimental challenges must be overcome to find the few events that can give insight into new phenomena out of the many millions of collision taking place every second at each of the interaction points of the LHC. The reconstruction and identification of physical objects like electrons, muons and photons from the raw detector data is one of the fundamental tasks in the analysis of detector data from collider experiments. Further analyses work with those reconstructed objects to measure properties of the underlying theory or search for new particles.

One type of those reconstructed objects are  $\tau$  leptons. The  $\tau$  lepton as the heaviest lepton plays an important role in the search for the standard model Higgs boson in the mass range below about 135 GeV, where processes with  $\tau$  leptons in the final state are among the most promising channels for discovering the Higgs boson. In addition, many of

---

<sup>3</sup>Conseil Européen pour la Recherche Nucléaire

<sup>4</sup>A Toroidal LHC ApparatuS

<sup>5</sup>Compact Muon Solenoid

the extensions of the standard model, especially those involving supersymmetry, show an abundance of  $\tau$  leptons for certain regions of the parameter space. An efficient reconstruction and identification of  $\tau$  leptons therefore plays an important role in the investigation of these phenomena.

$\tau$  leptons decay 75% of the time into a number of hadrons. This makes their identification a challenge at a hadron collider, where jets produced by strong interactions have a very large cross section. Those jets have very similar signatures to hadronically decaying  $\tau$  leptons, especially at low energies. Certain well measured properties of the hadronic decays of the  $\tau$  lepton make a separation against QCD jets nevertheless possible.

This thesis is concerned with the ATLAS experiment, one of the two multipurpose detectors at the Large Hadron Collider. The reconstruction and identification of hadronically decaying  $\tau$  leptons in the ATLAS experiment is studied and new developments for the existing algorithms are examined for their potential to improve the discrimination between hadronically decaying  $\tau$  leptons and QCD jets.

In the ATLAS reconstruction software two competing algorithms for the reconstruction and identification of  $\tau$  leptons were available, following slightly different strategies. One of those algorithm, *tauRec*, is based on energy clusters reconstructed in the calorimeter, while the other algorithm, *tau1P3P*, is based on the reconstructed tracks of charged particles. These two algorithms will be described in Chapter 4 in more detail. In Chapter 5 improvements are developed for one of those algorithms, the calorimeter based tauRec. The tauRec algorithm had a very low efficiency to reconstruct hadronically decaying  $\tau$  leptons for whom the sum of their visible<sup>6</sup> decay products has a transverse momentum below 30 GeV. In many physics channels with  $\tau$  leptons in the final state a significant fraction of the  $\tau$  leptons populate this region of the phase space. It was therefore desirable to improve the reconstruction efficiency for soft  $\tau$  leptons. The presented improvements lead to a better reconstruction efficiency in this low energy region and at the same time to a better separation between  $\tau$  leptons and jets produced by QCD processes.

In Chapter 6 a first step towards a combination of the two  $\tau$  reconstruction algorithms is presented, which has been developed within the scope of this thesis. Combining the two algorithms offered the chance to benefit from the complementary strengths of the two  $\tau$  reconstruction algorithms to improve the identification performance compared to what is offered by each of them individually. Some studies are shown on the prospects for such improvement of the  $\tau$  identification performance for the ATLAS experiment.

In Chapter 1 an overview of the theoretical framework will be given, followed by a description of the LHC and the ATLAS detector in Chapter 2 and an introduction to the theory and application of classification methods in Chapter 3.

Throughout this thesis natural units will be used, where  $\hbar \equiv c \equiv 1$ . Therefore masses and momenta are given in units of eV.

---

<sup>6</sup>The visible decay products are the particles that can be detected, which excludes neutrinos.



# 1 Theoretical Background

In this chapter a short overview of the Standard Model of particle physics will be given in Section 1.1-1.3. Section 1.4 will describe some phenomenological details, with a focus on the physics especially relevant for hadron colliders.

## 1.1 The Standard Model

The Standard Model was developed in the 1960's and 70's to describe the large amount of data provided by particle accelerators with ever increasing energies, in a unified way. So far it has successfully described all experimental data.

The standard model is an effective quantum field theory of point like particles, which is based on the concept of local gauge symmetry. In this type of theory the force mediating bosons with spin 1 arise from the generators of the corresponding gauge symmetry group. In the case of the standard model this is the group  $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ , leading to three fundamental forces, the strong, the weak and the electromagnetic force. The standard model unifies two theories in a consistent framework; the *electroweak model*, describing the electromagnetic and the weak forces in a unified way, and the theory of the strong interactions, *Quantum ChromoDynamics* (QCD).

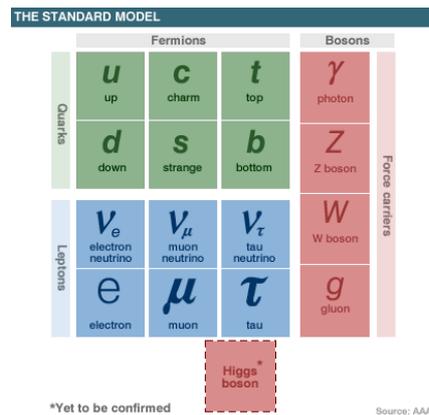


Figure 1.1: The particle content of the Standard Model [1]. Shown on the left side are the fermionic matter particles, with the quarks on the top and leptons on the bottom. The three columns represent the three families of matter. On the right side the force mediating bosons are shown. The particle on the bottom is the postulated Higgs boson, which has not been observed yet.

## 1 Theoretical Background

The particle content of the standard model can be seen in Figure 1.1. The particles representing matter with spin 1/2, called *fermions*, are divided into *quarks* and *leptons*. The quarks undergo strong interactions, in contrast to the leptons. The fermions are divided further into three families, with the second and third families being copies of the first one, differing only in the mass of the particles. The stable matter is made up only from the particles of the first family.

Both the electroweak model and QCD are based on a Lagrangian formalism, where the dynamics are derived by minimising the action  $S = \int d^4x \mathcal{L}$  of the Lagrange density  $\mathcal{L}$ . The first theory of this kind, which was able to give accurate predictions, was *quantum electrodynamics* (QED). QED is based on the simplest gauge symmetry, a local phase invariance. Because of its great success, it became the prototype for the development of the following theories.

### 1.2 The Electroweak Model

The electroweak model was developed mainly by Glashow, Salam and Weinberg in the late 60's [9, 28, 37]. It unified the electromagnetic force with the weak force at energies above the electroweak scale under the gauge group  $SU(2)_L \otimes U(1)_Y$ .  $SU(2)_L$  and  $U(1)_Y$  are the gauge groups of the *weak isospin* and the *weak hypercharge*. The bosons arising from this electroweak symmetry group have to be massless in order not to destroy the symmetry. This contradicted the observations of the weak W- and Z-bosons having masses of approximately 80.4 GeV and 91.2 GeV respectively. In addition the model also requires the fermions to be massless. Both the fermion and the boson masses can be introduced by breaking spontaneously the  $SU(2)_L \otimes U(1)_Y$  symmetry.

The electroweak symmetry group has one generator for  $U(1)_Y$  group and three generator for the  $SU(2)_L$  group. Therefore the theory contains four gauge boson fields  $B_\mu$  and  $W_\mu^i$ ,  $i = 1, 2, 3$ , corresponding to these generators. Experimental evidence shows that the four boson fields mix into mass eigenstates  $\gamma$ ,  $Z$  and  $W^\pm$  via the transformation

$$A_\mu = \cos \theta_W B_\mu + \sin \theta_W W_\mu^3 \quad (1.1)$$

$$Z_\mu = \sin \theta_W B_\mu - \cos \theta_W W_\mu^3 \quad (1.2)$$

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2) \quad (1.3)$$

with the weak mixing angle  $\theta_W$  as a free parameter, to be determined by experiment.

The fermion fields can be decomposed into a right and left handed component via the projection operators  $P^\pm = \frac{1}{2}(1 \pm \gamma^5)$ . To explain the parity violating nature of the weak force, only the left handed components of the fermions couple to the weak bosons. The right handed components are singlets of the  $SU(2)$  group, which means they do not possess weak isospin. The fermions and their quantum numbers are listed in table Table 1.1.

## 1 Theoretical Background

Leptonen	$I_3$	$Q$	$Y$	Quarks	$I_3$	$Q$	$Y$
$\nu_L$	1/2	0	-1	$u_L$	1/2	2/3	1/3
$l_L^-$	-1/2	-1	-1	$d_L$	-1/2	-1/3	1/3
$\nu_R$	0	0	0	$u_R$	0	2/3	4/3
$l_R^-$	0	-1	-2	$d_R$	0	-1/3	-2/3

Table 1.1: The particles for one family of the standard model with their quantum numbers. These are the third component of the weak isospin  $I_3$ , the weak hypercharge  $Y$  and the electric charge  $Q$ . The electric charge is connected to the other two quantum numbers by  $Q = Y/2 + I_3$  after the electroweak gauge group is spontaneously broken. The right handed neutrinos are sterile, and do not interact, but experimental evidence of a neutrino mass implies their existence.

When boson masses are added directly by introducing explicit mass terms like

$$m_W^2 W_\mu^\dagger W^\mu + \frac{1}{2} m_Z^2 Z_\mu Z^\mu, \quad (1.4)$$

the gauge symmetry of the theory is destroyed, and divergences arise at higher orders of perturbation theory. This problem is solved by spontaneous symmetry breaking via the Higgs mechanism [6, 13].

In the Higgs mechanism an isospin doublet is introduced consisting of two complex scalar fields

$$\Phi = \begin{pmatrix} \phi^0 \\ \phi^+ \end{pmatrix} \quad (1.5)$$

Furthermore, the new fields are subject to a potential

$$V(\Phi) = \frac{1}{2} \mu^2 \Phi^\dagger \Phi + \frac{1}{4} \lambda (\Phi^\dagger \Phi)^2, \quad (1.6)$$

so that the Higgs terms of the Lagrangian are

$$\mathcal{L}_{higgs} = (D^\mu \Phi)^\dagger (D_\mu \Phi) - V(\Phi), \quad (1.7)$$

with  $D_\mu$  being the covariant derivative

$$D_\mu = \partial_\mu + i \frac{g}{2} \tau_i W_\mu^i + i \frac{g'}{2} Y B_\mu, \quad (1.8)$$

so that the Lagrangian is invariant under  $SU(2)_L \otimes U(1)_Y$ . Here the  $\tau_i$  and  $Y$  are the generators of the  $SU(2)$  and  $U(1)$  group.  $g$  and  $g'$  are the coupling constants, describing the coupling of the fermion fields to the boson fields.

The  $\frac{\lambda}{4} (\Phi^\dagger \Phi)^2$  term is a self coupling term, and  $\frac{1}{2} \mu^2 \Phi^\dagger \Phi$  can be interpreted as a mass term if  $\mu^2 > 0$ . If, on the other hand,  $\mu^2$  is negative, the potential has a nontrivial degenerate minimum where

$$|\Phi_0| = \sqrt{\frac{-\mu^2}{\lambda}} \equiv \frac{\nu}{\sqrt{2}} \quad (1.9)$$

## 1 Theoretical Background

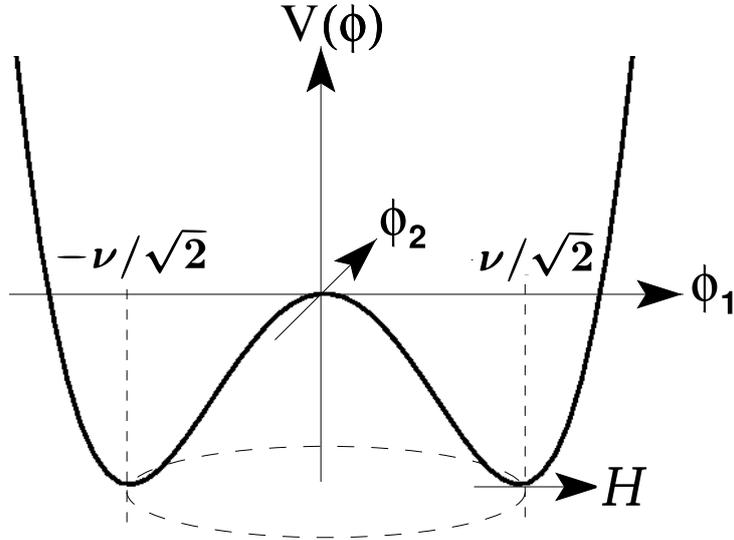


Figure 1.2: Illustration of the Higgs potential  $V(\Phi) = \frac{1}{2}\mu^2\Phi^\dagger\Phi + \frac{1}{4}\lambda(\Phi^\dagger\Phi)^2$  with  $\mu^2 < 0$  and  $\lambda > 0$ , leading to a degenerate minimum at  $|\Phi_0| = \frac{\nu}{\sqrt{2}}$ .

with  $\nu$  as the *vacuum expectation value* (VEV) of the Higgs field. By choosing without loss of generality a specific ground state

$$\langle \Phi_0 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu \end{pmatrix}, \quad (1.10)$$

the symmetry of the original Lagrangian is no longer manifest in this ground state.

Expanding  $\Phi$  around the ground state and substituting it into (1.7) results in a Lagrangian describing one massive scalar field  $H(x)$  with mass  $m_H = \sqrt{-2\mu^2}$ , corresponding to the **Higgs boson**, and three massless scalar **Goldstone bosons**. The Goldstone bosons are unphysical, but can be absorbed into the longitudinal polarisations of the  $W^+, W^-$  and  $Z$  bosons by choosing the so called ‘unitarity gauge’ where

$$\Phi'(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu + H(x) \end{pmatrix}. \quad (1.11)$$

By doing this the desired mass terms for the vector bosons arise without destroying the underlying gauge symmetry.

The Higgs Lagrangian takes the final form

$$\mathcal{L}_{higgs} = \frac{1}{2}\partial_\mu H \partial^\mu H + \frac{1}{4}g^2(\nu + H)^2 W_\mu W^\mu + \frac{1}{8}g_Z^2(\nu + H)^2 Z_\mu Z^\mu + V(\nu + H), \quad (1.12)$$

with

$$V(\nu + H) = \mu^2 H^2 - \lambda \nu H^3 - \frac{\lambda}{4} H^4 + const. \quad (1.13)$$

## 1 Theoretical Background

It now contains mass terms for the W and Z bosons, but no mass term for the photon, exactly as dictated by experimental evidence. The weak vector bosons obtain the masses

$$m_W = \frac{1}{2}g\nu \quad (1.14)$$

$$m_Z = \frac{1}{2}g_Z\nu = \frac{1}{2}\frac{g}{\cos\theta_W}\nu. \quad (1.15)$$

The Higgs mechanism also calls for Yukawa and quartic couplings between the Higgs and the weak vector bosons, as well as self interactions of the Higgs boson.

An appealing feature of the Higgs mechanism is that the same scalar doublet which gives masses to the bosons is also able to generate masses for the fermions via Yukawa couplings to the Higgs field. To introduce mass for a charged lepton  $\ell$  of the left handed isospin doublet  $\begin{pmatrix} \nu_\ell \\ \ell \end{pmatrix}_L$  the following term is added to the total Lagrangian:

$$L_{Yukawa} = -G_\ell \left[ (\bar{\nu}_\ell, \bar{\ell})_L \Phi \ell_R + \bar{\ell}_R \Phi^\dagger \begin{pmatrix} \nu_\ell \\ \ell \end{pmatrix}_L \right]. \quad (1.16)$$

By using again (1.11) for  $\Phi$ , one obtains

$$L_{Yukawa} = -\frac{G_\ell}{\sqrt{2}}\nu(\bar{\ell}_L\ell_R + \bar{\ell}_R\ell_L) - \frac{G_\ell}{\sqrt{2}}\nu(\bar{\ell}_L\ell_R + \bar{\ell}_R\ell_L)H, \quad (1.17)$$

which gives a lepton mass of  $m_\ell = \frac{G_\ell}{\sqrt{2}}\nu$  and at the same time produces a coupling of the lepton to the Higgs boson with a coupling strength proportional to  $m_\ell$ . The coupling  $G_\ell$  of each lepton to the Higgs field is a free parameter, which can be determined by an experimental measurement of the lepton mass. Quark masses can be introduced in an analogous way, but the weak eigenstates will differ from the mass eigenstates, since both members of an isospin doublet are massive.

The Higgs boson is the only particle of the standard model that has not been directly observed yet. The Higgs vacuum expectation value can be extracted from the muon lifetime as  $\nu \approx 246$  GeV, and so the couplings of the Higgs boson to the other particles of the standard model can be calculated. But the mass of the Higgs boson cannot be predicted without knowing the free parameter  $\lambda$ .

### 1.3 Quantum Chromo Dynamics

Quantum Chromo Dynamics, the theory describing the strong interaction between quarks and gluons, was developed in the early 1970's by Greenberg, Nambu, Wilczek, Gross and Politzer amongst others.

QCD is built upon a  $SU(3)$  colour symmetry group. Colour is the name given to the charges of the  $SU(3)$  symmetry group, of which there are three kinds. The quarks carry

## 1 Theoretical Background

one of these colour charge and in addition also an electrical charge of  $2/3$  (up type quarks) or  $-1/3$  (down type quarks). The strong force between the quarks is mediated by eight massless gauge bosons, the gluons, which are required to also carry colour charge in order to preserve the gauge symmetry. Therefore the gluons couple to each other, giving rise to the phenomenon of *confinement*. Confinement describes the experimental observation, that quarks only appear in colour neutral bound states called *hadrons*. The reason for the confinement is the long range behaviour of the strong force. For large distances, or low energies, the strong coupling constant becomes very large, and above a distance of about 1 fm the potential energy is large enough to create a quark-antiquark pair out of the vacuum.

The confinement is closely related to the opposite phenomenon of asymptotic freedom. At very short distances the strong coupling constant becomes very small, so that the quarks behave as quasi free particles. This means that QCD processes can be calculated using perturbation theory in the high energy regime.

Because of the confinement, isolated quarks cannot exist. If a quark inside a hadron gets scattered so the hadron breaks apart, one observes not the quark itself, but a ‘jet’ of numerous hadrons made up from the quark-antiquark pairs which are produced from the vacuum when the quark is moving away from the other quarks of the original hadron. This transition from particles with colour charge to colour neutral hadrons is called *fragmentation* or *hadronisation*. Because of the strong coupling involved, the detailed fragmentation process cannot be calculated perturbatively. Therefore one has to rely on phenomenological models to simulate the evolution of QCD jets.

### 1.4 Physics at pp Colliders

Compared to  $e^+ e^-$  colliders, the calculation of cross sections for proton-proton collisions is somewhat complicated by the fact that the colliding particles have a substructure, and the energy of the hadrons is shared between their constituents. Inside a hadron gluons are continuously exchanged between the *valence quarks*.<sup>1</sup> Also, quark-antiquark pairs (*sea quarks*) are constantly produced and annihilated within the hadron. All these processes cannot be calculated using perturbation theory, since their energy scale is very low and the strong coupling constant therefore becomes too large.

To calculate the hadronic cross section for a process

$$A + B \rightarrow c + X, \tag{1.18}$$

where  $c$  might be for example the Higgs boson and  $X$  is the remnant of the protons, associated with the partonic process<sup>2</sup>

$$a + b \rightarrow c, \tag{1.19}$$

---

<sup>1</sup>The valence quarks are the quarks that define the identity of a hadron.

<sup>2</sup>‘Parton’ is the generic name for quarks and gluons as the constituents of a hadron.

## 1 Theoretical Background

one needs to know the probability to encounter the parton  $a$  with energy fraction  $x_a$  inside hadron  $A$  and likewise for parton  $b$ . This probability density  $f_a^A(x, Q^2)$ , which depends on the energy scale  $Q^2$ , is called the *parton distribution function* (PDF). The hadronic cross section is given by convoluting the partonic cross section with the PDFs for  $a$  and  $b$

$$d\sigma(A + B \rightarrow cX) = \sum_{a,b} \int_0^1 dx_a \int_0^1 dx_b f_a^A(x_a, Q^2) f_b^B(x_b, Q^2) d\sigma(a + b \rightarrow c), \quad (1.20)$$

where the sum extends over all combinations of partons  $a$  and  $b$  that can produce the final state  $c$ .

This separation of the process into a so called hard scattering process, contained in the partonic cross section, and a soft non perturbative part with low  $Q^2$ , contained in the PDF, is called factorisation. This is only possible, because the PDFs are independent of the process, and can therefore be measured experimentally. The calculated cross sections depend on the chosen factorisation scale  $Q^2$ , but in higher order calculations the dependence becomes weaker and completely vanishes for calculations to all orders.

The PDFs measured at a certain energy scale can be extrapolated to a different scale using the Altarelli-Parisi equations. As an example, the PDF for a quark flavour  $q$ ,  $f_q(x, Q^2)$ , evolves according to

$$\frac{d}{d(\log Q^2)} f_q(x, Q^2) = \frac{\alpha_s}{2\pi} \int_x^1 f_q(y, Q^2) P_{qq}\left(\frac{x}{y}\right) \frac{dy}{y}. \quad (1.21)$$

Here  $\alpha_s$  is the strong coupling constant, and  $P_{qq}\left(\frac{x}{y}\right)$  is called the *splitting function*. The splitting function parametrises the probability, that a quark with momentum fraction  $x$  originates from a quark with momentum fraction  $y$  by emission of a gluon. A similar contribution from gluon splitting has to be taken into account in practise.

The splitting functions can be calculated using perturbation theory.

### 1.5 The $\tau$ Lepton

The  $\tau$  lepton is the charged lepton of the third family with a mass of 1.777 GeV [7], and was the first member of its family to be discovered. The  $\tau$  lepton decays via a W boson with a mean lifetime of 290 fs, either into a charged lepton and two neutrinos in about 35% of the cases, or into a quark-antiquark pair plus one neutrino about 65% of the time. The tree level diagrams for both decay modes are shown in Figure 1.3. The former decays are called *leptonic* and the latter are called *hadronic*.

The decay modes of the  $\tau$  are usually further classified by the number of charged particles amongst the decay products. For reasons of charge conservation this number must be odd. The leptonic decay modes produce one charged particle in the final state, whereas the hadronic decay modes have mostly one or three, rarely five or more, charged particles amongst their decay products.

## 1 Theoretical Background

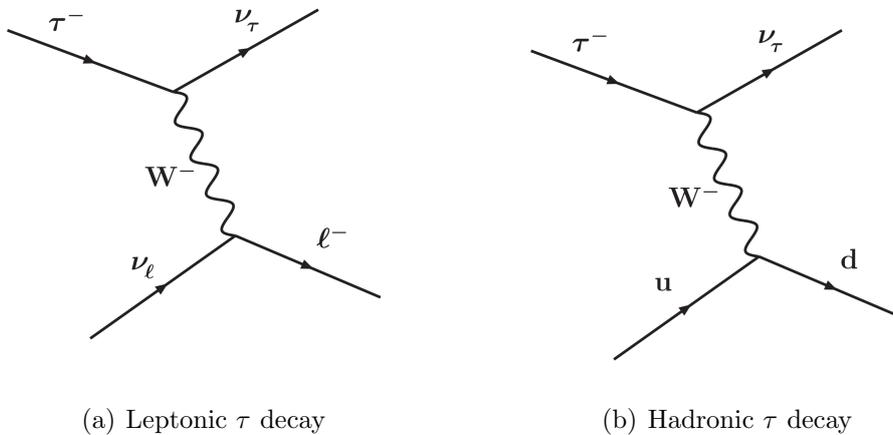


Figure 1.3: Feynman graphs of the leptonic and hadronic  $\tau$  decays. In the left diagram  $\ell$  can be an electron or a muon. In the right diagram the down quark can also be replaced by a strange quark.

Since the  $\tau$  decay products always contain neutrinos, which interact very weakly and cannot be detected in experimental setups, the kinematics of  $\tau$  final states can only be partially reconstructed. This specifically applies to the leptonic decay modes of the  $\tau$ , where two neutrinos are among the decay products. The energy of the decay products without the neutrinos is called the visible energy of the  $\tau$  lepton.

Because of the chiral structure of the charged weak current, the kinematics of the  $\tau$  decay products depend on the helicity of the decaying  $\tau$  lepton. Since the decay of the  $\tau$  is independent of its production, this can be exploited to measure the  $\tau$  polarisation. By convention, the polarisation of the  $\tau$  leptons in a process is defined by

$$P_{\tau^-} = -P_{\tau^+} = \frac{\sigma_R - \sigma_L}{\sigma_R + \sigma_L}, \quad (1.22)$$

where  $\sigma_R$  and  $\sigma_L$  denote the cross section of observing a right handed or left handed  $\tau$  lepton respectively. For example,  $\tau$  leptons produced by  $W$  decays have a polarisation of  $P_{\tau^-} = -1$  since the  $W$  boson only couples to left handed particles, whereas  $\tau$  leptons originating from the decay of a Higgs boson will have a vanishing polarisation, since the Higgs boson couples to left and right handed particles equally.

The two decay modes of the  $\tau$  lepton that are best suited for measuring the polarisation are

$$\tau^\pm \rightarrow \pi^\pm \nu_\tau \quad (1.23)$$

and

$$\tau^\pm \rightarrow \rho^\pm \nu_\tau \rightarrow \pi^\pm \pi^0 \nu_\tau, \quad (1.24)$$

which together make up about 50% of all hadronic decays.

## 1 Theoretical Background

In the rest frame of the  $\tau$  lepton, the differential decay widths into a  $\pi$  or a  $\rho$  vector meson are given by

$$\frac{1}{\Gamma_\pi} \frac{d\Gamma_\pi}{d\cos\theta} = \frac{1}{2}(1 + P_\tau \cos\theta), \quad (1.25)$$

$$\frac{1}{\Gamma_\rho} \frac{d\Gamma_{\rho,L}}{d\cos\theta} = \frac{\frac{1}{2}m_\tau^2}{m_\tau^2 + 2m_\rho^2}(1 + P_\tau \cos\theta), \quad (1.26)$$

$$\frac{1}{\Gamma_\rho} \frac{d\Gamma_{\rho,T}}{d\cos\theta} = \frac{m_\tau^2}{m_\tau^2 + 2m_\rho^2}(1 - P_\tau \cos\theta), \quad (1.27)$$

where  $L$  and  $T$  denote the longitudinal and transversal polarisation states of the  $\rho$ . The angle  $\theta$  measures the direction of the meson with respect to the  $\tau$  flight axis. With the collinear approximation  $m_\tau \ll p_\tau$ , where all the decay products are produced parallel to the  $\tau$  flight axis, one can relate  $\theta$  to the fraction  $x$  of the  $\tau$  energy carried by the meson in the laboratory frame. One has

$$\cos\theta = \frac{2x - 1 - m_{\pi,\rho}^2/m_\tau^2}{1 - m_{\pi,\rho}^2/m_\tau^2}. \quad (1.28)$$

Therefore the energy fraction carried by the visible  $\tau$  decay products will be smaller for  $\tau$  leptons with a negative polarisation. If the  $\tau^\pm \rightarrow \rho^\pm \nu_\tau$  decay mode can be isolated, the ratio  $E_\pi/E_{\tau,vis}$  of the pion energy to the total visible energy can be used to measure the polarisation. Left handed  $\tau$  leptons will favour an even sharing of the energy between the charged and the neutral hadron, whereas for right handed  $\tau$  leptons the charged pion will carry almost all or very little of the total energy. For more details see [8, 24].

### 1.5.1 Physics searches in $\tau$ Lepton Final States

$\tau$  leptons play an important role for the discovery of new phenomena at future particle accelerators, such as the LHC, which will be described in Section 2.1. One of the discoveries hopefully to be made at the LHC, is the direct observation of the Higgs boson. As already mentioned, the mass of the Higgs boson is not predicted by the standard model. But direct searches at the LEP collider lead to a lower limit of  $M_H > 114.4$  GeV at 95% confidence level [34] and precision measurements of electroweak observables at LEP and Tevatron, which are sensitive to the Higgs mass through loop corrections, set the upper limit for the Higgs mass of  $M_H < 190$  GeV at a confidence level of 95% [34].

As shown in Section 1.2 the coupling strength of fermions to the standard model Higgs boson is proportional to the mass of the fermion. Therefore the branching ratios of the Higgs boson into heavy fermions is much larger than for the lighter families. For a light Higgs boson with a mass of about 120 GeV, well below the threshold for decay into two vector bosons, the most prominent decay mode will be  $H \rightarrow b\bar{b}$  with a branching ratio of about 70% followed by  $H \rightarrow \tau^-\tau^+$  with a branching ratio of about 8%. The production

## 1 Theoretical Background

cross sections and branching ratios of the Higgs boson for different channels is shown in Figure 1.4 as a function of the Higgs mass.

The  $b\bar{b}$  decay channel will be experimentally very challenging in the two dominant Higgs production channels through gluon-gluon fusion and vector boson fusion(VBF). The VBF  $H \rightarrow \tau^-\tau^+$  channel, where one of the  $\tau$  leptons decays leptonically, offers a high sensitivity to observe a standard model like Higgs boson with a mass near the lower limit set by the LEP experiments. Additionally the  $\tau$  decay channel offers the possibility to probe the parity of the Higgs boson [39].

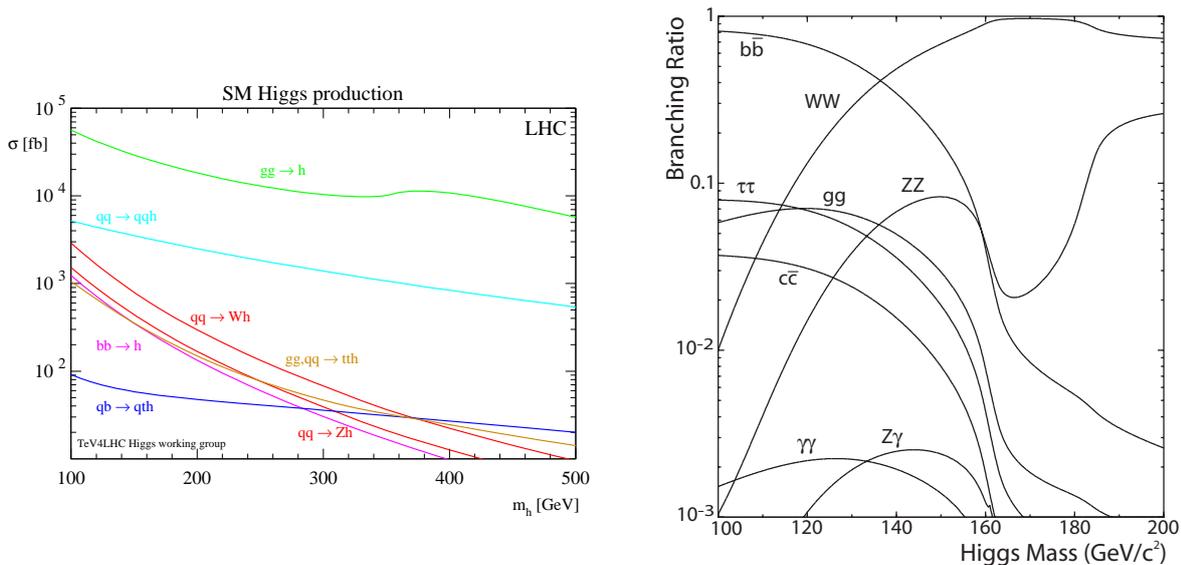


Figure 1.4: The production cross sections (left) and branching ratios (right) for a standard model Higgs boson as a function of its mass (from [2]).

Another area where  $\tau$  leptons might prove valuable is supersymmetry (SUSY). Supersymmetry is a proposed extension of the standard model, which introduces a new underlying symmetry between bosons and fermions, which is spontaneously broken at the TeV scale. In supersymmetric models every particle of the standard model has a heavy superpartner with a spin differing by half a unit. Also an additional Higgs doublet must be introduced to keep the theory consistent, so that instead of one Higgs boson there exist three neutral and two charged Higgs bosons. Supersymmetry solves some of the problems of the standard model, like the hierarchy problem, and under some conditions also offers a natural candidate for nonbaryonic dark matter.

The simplest supersymmetric extension of the standard model, the *minimal supersymmetric standard model* (MSSM), has over a hundred free parameters, but by introducing some theoretically motivated constraints, the parameter space can be reduced to few distinct parameters in addition to the standard model. There exist several different constraint scenarios of the MSSM, differing for example in the way supersymmetry is broken. One of the most extensively studied of these constrained supersymmetric models is called *minimal*

## 1 *Theoretical Background*

*supergravity* (mSUGRA), where the SUSY breaking is mediated by gravitational interactions.

In the mSUGRA model one of the parameters, the ratio of the vacuum expectation values of the two Higgs doublets,  $\tan\beta$ , plays a role in the the couplings of the supersymmetric particles as well as the Higgs bosons to the standard model fermions of the third family. For large values of  $\tan\beta$ , the couplings to  $\tau$  leptons are strongly enhanced. This is also the case for other MSSM scenarios. Because of this, an excess of  $\tau$  leptons can give a strong signal for the realisation of the MSSM in nature. In addition, the possibility to probe the spin structure of the resonances decaying to  $\tau$  leptons will be of importance to narrow down the underlying model, if a deviation from the standard model has been observed.

## 2 The ATLAS Experiment

This chapter gives an overview of the LHC accelerator, the ATLAS experiment and its various sub-detector systems.

The ATLAS<sup>1</sup> experiment is placed at one of the interaction points of the Large Hadron Collider (LHC) at CERN<sup>2</sup> near Geneva. Both the accelerator and the detector are nearing their completion and the data taking will start in summer 2008.

In 1992 the first concept for ATLAS was presented in a Letter of Intent [31] and its construction was approved 1995. The ATLAS collaboration has since grown to about 2100 members from 37 countries.

### 2.1 The Large Hadron Collider

The LHC is a proton-proton collider with a center of mass energy of 14 TeV and an design instantaneous luminosity of  $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ . Both the energy and the luminosity are an order of magnitude above the performance of current accelerators, and will open new possibilities for investigating physics at the TeV scale.

The tunnel housing the LHC – previously used for the electron positron collider LEP – has a circumference of 27 km and lies about 100 m beneath the surface. Inside the vacuum of the beam pipe the protons circulate in each direction in nearly 3000 *bunches* consisting of about  $10^{11}$  protons each with an energy of 7 TeV. At four different points around the ring, where detectors are positioned, the beams are focused and brought to collision. The bunches will meet every 25 ns and produce up to several hundred particles. This poses an enormous challenge for the detectors, which have to identify these particles and store the data at a very high rate. Such a high rate is needed to observe rare processes like the production of a Higgs boson.

Before the proton bunches are injected into the LHC, their energy is raised to 450 GeV by a series of smaller accelerators. They start in the linear accelerator LINAC2 with 50 MeV followed by the PS Booster with 1.4 GeV, the PS<sup>3</sup> with 26 GeV and finally the SPS<sup>4</sup> with 450 GeV. From the SPS the bunches are injected into the LHC in both directions and accelerated to their final energy of 7 TeV.

---

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

<sup>2</sup>Conseil Européen pour la Recherche Nucléaire

<sup>3</sup>Proton Synchrotron

<sup>4</sup>Super Proton Synchrotron

## 2 The ATLAS Experiment

center-of-mass energy	14	TeV
injection energy	450	GeV
instantaneous luminosity	$10^{34}$	$\frac{1}{\text{cm}^2\text{sec}}$
bunch spacing	24.95	ns
number of protons per bunch	$1.1 \cdot 10^{11}$	
beam-beam angle at the interaction point	300	$\mu\text{rad}$
half-life-period of the luminosity	10	h
stored energy per beam	350	MJ
number of dipole magnets	1232	
dipole field strength at 7 TeV per beam	8.33	T
magnet current	11796	A
number of quadrupole magnets	56	
number of cavities	8	
electric field gradient in the cavities	5	$\frac{\text{MV}}{\text{m}}$

Table 2.1: important characteristics of the LHC

To keep the protons on a circular trajectory, 1232 superconducting dipole magnets with a maximum field strength of 8.33 T bend their path around the ring. To reach these high field strengths, the magnets have to be kept at a temperature of 1.9 K. The proton beams will circulate for several hours at their nominal energy, during which the collisions take place at the interaction points. For the initial acceleration and to compensate for the energy loss through synchrotron radiation, eight superconducting radio frequency (RF) cavities with a field gradient of 5 MV/m are used.

Some key properties of the LHC are given in Table 2.1.

### 2.1.1 Luminosity

Besides the center of mass energy of the colliding particles, the other important performance measure for a collider is the *instantaneous luminosity*  $\mathcal{L}$ . The rate  $R = \frac{dN}{dt}$  of a certain physical process is determined by  $\mathcal{L}$  together with the cross section  $\sigma$  of that process by

$$R = \sigma \cdot \mathcal{L} \quad (2.1)$$

For simple geometries the luminosity can be calculated from machine parameters as

$$\mathcal{L} = f \frac{N_1 N_2}{A_{eff}} \quad (2.2)$$

with  $N_1, N_2$  being the number of particles in the two bunches,  $f$  the frequency of the collision and  $A_{eff}$  the effective collision area.

To calculate the expected number of events for a process  $X$  with cross section  $\sigma_X$ , the event rate  $R$  has to be integrated over time

$$N_X = \int \sigma_X \cdot \mathcal{L}(t) dt = \sigma_X \cdot \int \mathcal{L}(t) dt = \sigma_X \cdot \mathbf{L}, \quad (2.3)$$

with  $L$  being the *integrated luminosity*. The integrated luminosity has to be measured as precisely as possible, because its uncertainty introduces systematic errors on cross section measurements or exclusion limits.

### 2.1.2 Pile Up Events

Due to the high proton densities that are necessary to reach high luminosities, more than one proton pair will interact on average per bunch crossing. For the design instantaneous luminosity of the LHC ( $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ ) it is expected that there will be about 24 proton-proton interactions on average per bunch crossing.

Most of these additional interactions will not be hard scattering processes, but will typically consist of inelastic scattering with low momentum transfer. This follows from the much higher cross section of low  $Q^2$  processes. These events are commonly referred to as *minimum bias events* (MB). This kind of pile up is more specifically the so called *in-time pile up*.

Opposed to that is the pile up which originates from the limited time resolution of certain detector elements, the *out-of-time pile up*. The latter can lead to signals that are an overlay of several bunch crossings before and after the recorded event.

Both types of pile up have to be taken into account for an accurate simulation of the detector response and for a tuning of the reconstruction algorithms. In the studies presented in this thesis this was not possible, since suitable simulated data with pile up included were not available.

## 2.2 The ATLAS Experiment

The ATLAS detector is located at the southern interaction point of the LHC named *Point 1*. It is an all purpose detector in a layered design with nearly complete  $4\pi$  coverage around the interaction point as shown in Figure 2.1.

Two important goals of the experiment are to examine the details of electroweak symmetry breaking, by either confirming the existence of the Higgs boson or excluding it, and to search for physical phenomena beyond the standard model occurring at the TeV scale.

Some of the most important requirements influencing the design were:

- Fast and radiation hard components to cope with the high particle flux.
- High granularity to keep the occupancy at an acceptable level.
- Acceptance over the full azimuthal angle  $\phi$  and up to high pseudorapidities.<sup>5</sup>
- Good muon identification and momentum resolution and very good determination of the sign of the charge for high  $p_T$  muons.

---

<sup>5</sup>The pseudorapidity  $\eta$  is defined as  $\eta = -\ln(\tan\frac{\theta}{2})$ , with  $\theta$  being the polar angle with respect to the beam axis.

## 2 The ATLAS Experiment

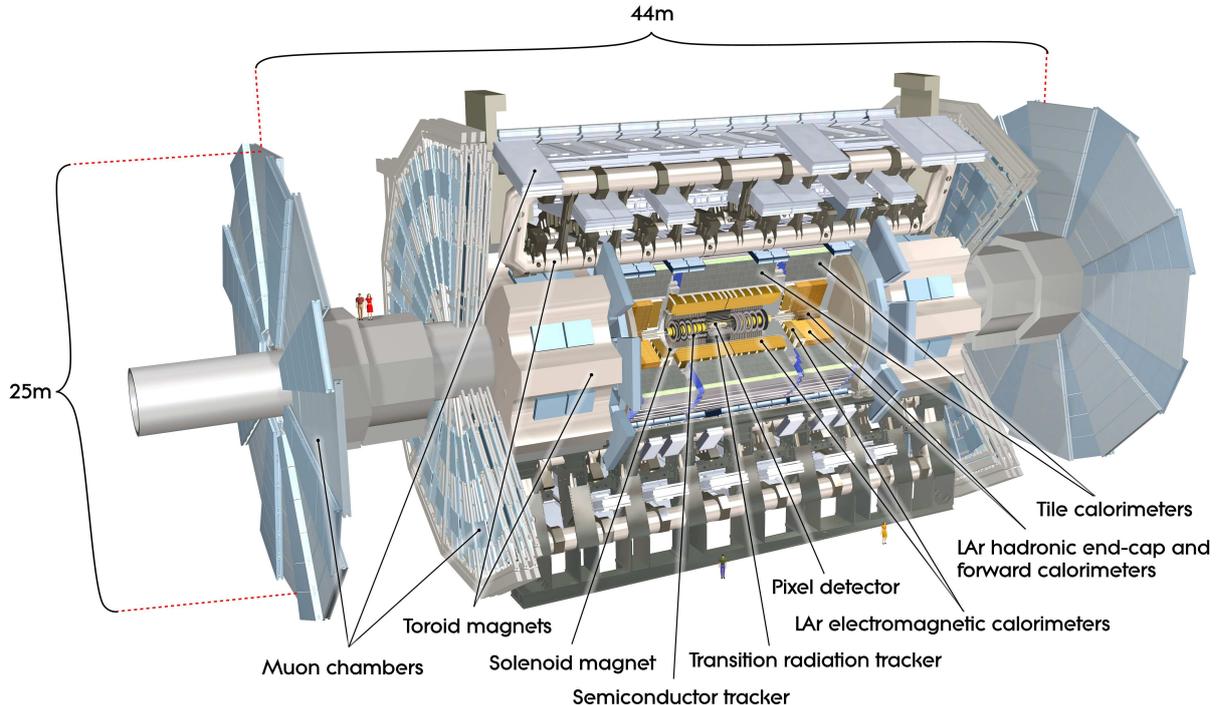


Figure 2.1: Cut-away view of the ATLAS detector [33]

- Efficient vertex reconstruction for the identification of  $\tau$ -leptons and jets from b-quarks and separation of pile up events.
- Good separation of  $\pi^0/\gamma$  and  $\pi^\pm/e^\pm$ .

The ATLAS detector has an overall length of 44 m, a height of 25 m and weighs approximately 7000 t. The detector system closest to the beam line is the inner detector or ‘Tracker’ and is described in Section 2.2.1. The inner detector is surrounded the electromagnetic calorimeter followed by the hadronic calorimeter (see Section 2.2.2). Most of the detector volume is occupied by the muon spectrometer, which forms the outermost layer and will be described in more detail in Section 2.2.3. Each of these components provides crucial information about the scattering processes that take place inside the detector.

The coordinate system used for the ATLAS detector has the origin defined by the nominal interaction point. The  $z$ -axis is defined by the beam direction and the  $x$ - $y$  plane is transverse to the beam direction. The positive  $x$ -axis is defined as pointing from the interaction point toward the center of the LHC ring and the positive  $y$ -axis is pointing upwards. Furthermore a polar coordinate system is defined. The azimuthal angle  $\phi$  is defined as the angle to the positive  $x$ -axis in the  $x$ - $y$  plane with  $|\phi| \leq \pi$ , and the polar angle  $\theta$  is defined as the angle to the beam axis in the range  $[0, \pi]$ . The  $R$  coordinate is the distance from the beamline in the transverse plane.

In the following, a short description of the detector components will be given. Much more details about the general detector design as well as all the sub components can be found in [32, 33].

### 2.2.1 The Inner Detector

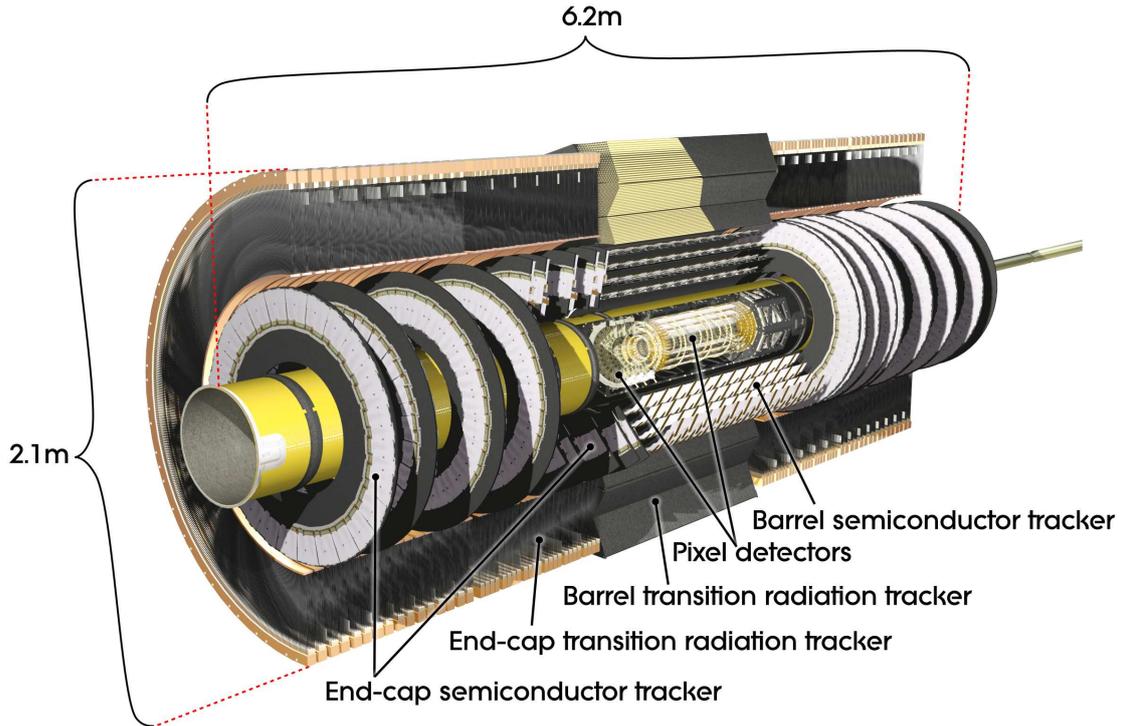


Figure 2.2: Cut-away view of the ATLAS Inner Detector, from [33]

The *inner detector* (ID) is the part of the detector closest to the interaction point (Figure 2.2). Its function is to record the path of ionising particles emanating from the interaction point with high efficiency and spatial resolution. This has to be accomplished within 25 ns before the next bunch crossing occurs and in a high radiation environment. The reconstructed paths of the particles are called *Tracks*.

The inner detector is placed inside a superconducting solenoid magnet, which provides a magnetic field of approximately 2 T parallel to the beam line over the volume of the ID. This allows the measurement of the ratio of charge to momentum of charged particles by measuring the curvature of their tracks in the plane orthogonal to the beam line.

Another important reason to reconstruct the trajectory of charged particles, is the reconstruction of vertices. Vertices are space-points where two or more particles are created, either inside or outside the beam pipe. The vertex with the highest sum of transverse momenta of the tracks originating from it is reconstructed as the *primary vertex*. Other vertices usually come from pile up events, and the vertexing helps to identify these. But

## 2 The ATLAS Experiment

vertices can also originate from long-lived particles, decaying inside the detector volume. These vertices are called *secondary vertices*. The distance of the secondary vertex to the primary vertex<sup>6</sup> can help to identify those long-lived particles. The most important examples of such particles are  $\tau$ -leptons and B-hadrons.

To accomplish its function, the inner detector uses three different components corresponding to different distances to the interaction point. They are each arranged in several layers. In the central *barrel* region ( $|\eta| \leq 1.5$ ), they are placed on concentric cylinders around the beam axis, while in the forward and backward *end-cap* region they are arranged on disks perpendicular to the beam axis. The total coverage is up to  $|\eta| \leq 2.5$ , matching the range of the calorimeters where they have a high granularity.

The innermost layers are equipped with radiation hard silicon pixel detectors with a pixel size of  $50 \times 400 \mu\text{m}$ , which results in an intrinsic resolution of about  $10 \mu\text{m}$  in the  $R - \phi$  plane and  $115 \mu\text{m}$  in  $z$  direction for the barrel and in  $R$  direction for the end-cap. There are three layers of pixel modules in the barrel, and also three layers in both of the end-caps, so that each particle typically has three 3-dimensional measurements (also called space-points) assigned from the pixel layers. The innermost pixel layer in the barrel has a distance of only about 5 cm from the beam axis, which together with the high granularity is very important for an accurate reconstruction of impact parameters<sup>7</sup> and vertices.

The second type of component is called **SemiConductor Tracker** (SCT). The SCT is made of silicon microstrip detectors with a strip pitch of  $80 \mu\text{m}$  in the  $R - \phi$  direction. Two sensors are glued back-to-back with a small stereo angle of 40 mrad to provide a measurement along the strip axis as well. The intrinsic resolution per module is about  $17 \mu\text{m}$  in the  $R - \phi$  plane and  $580 \mu\text{m}$  for the  $z$  coordinate in the barrel and for the  $R$  coordinate in the endcap. The SCT modules are arranged in four layers in the barrel region and nine disks in the end-caps, providing at least four additional potential space-points per track.

The outermost layer of the inner detector covering a pseudorapidity up to  $|\eta| \leq 2.0$  is the **Transition Radiation Tracker** (TRT). It uses straw drift tubes with a diameter of 4 mm filled with xenon gas, each with an intrinsic resolution of about  $130 \mu\text{m}$ . The straw tubes are oriented parallel to the beam axis in the barrel region and radially in the end-caps, and thus the TRT delivers only a measurement of the azimuthal angle  $\phi$ . The number of straw tubes crossed by each track is very high – usually 36. Combined with the large lever arm this contributes greatly to the precise reconstruction of the track momentum. An additional feature of the TRT is the ability to identify particles by the detection of transition radiation. Highly relativistic charged particles will radiate low energy photons when crossing the interface of media with different dielectric constants. The number of emitted photons increases with the  $\gamma$  factor of the particle. These photons will be absorbed in the xenon gas inside the straw tubes via the photoeffect, and will thus lead to a stronger signal than the traversal of a minimum ionizing particle. To make use of this effect, the space between the straw tubes is filled with radiating material, and the readout electronics

---

<sup>6</sup>This distance is referred to as the *flight length*.

<sup>7</sup>The impact parameter of a track is the minimal distance between the track and the primary vertex.

## 2 The ATLAS Experiment

Track Parameter	$p_T$ of Track			
	1 GeV		100 GeV	
	$\eta = 0$	$\eta = 2.1$	$\eta = 0$	$\eta = 2.1$
$\sigma(p_T)/p_T$	1.3%	5%	3.8%	11%
$\sigma(d_0)$	135 $\mu\text{m}$	320 $\mu\text{m}$	11 $\mu\text{m}$	15 $\mu\text{m}$
$\sigma(z_0) \cdot \sin(\theta)$	250 $\mu\text{m}$	420 $\mu\text{m}$	110 $\mu\text{m}$	110 $\mu\text{m}$

Table 2.2: Expected resolutions for the reconstruction of important track parameters. The resolution of the transverse momentum has been obtained from a muon sample, the resolution of the other parameters are obtained from a pion sample. The numbers are taken from [33].

is capable of distinguishing photon signals from pure ionisation signals from heavy particles. There are two signal thresholds implemented for this purpose. Straw tubes showing a signal above one or both thresholds are counted as *low* or *high threshold hits* respectively. Electrons will typically produce a higher ratio of high to low threshold hits. The separation of electrons and charged pions works best at energies between 2 and 20 GeV, achieving pion rejection factors (inverse efficiencies) of more than 50 for electron efficiencies of about 90%.

The expected performance of the reconstruction of particle tracks, can be summarised by a few key aspects. Important performance measures are the resolutions of the momentum and the the transverse and longitudinal impact parameters for the reconstructed tracks. The expected resolutions for different kinematic regions are shown in table Table 2.2. Closely related to the momentum resolution is the probability to assign the wrong sign of the charge to a track. For particles with a  $p_T$  less than 200 GeV this probability is below 0.1% and is dominated by multiple scattering and hadronic interactions. The efficiency to reconstruct the track of a particle depends on the energy and the pseudorapidity. The efficiency for charged pions is shown in Figure 2.3. It is at the order of 90% for particles with a  $p_T$  of several GeV.

### 2.2.2 The Calorimeter

After traversing the inner detector the particles reach the calorimeter system (Figure 2.4), which measures the energy of particles and is also the only component that detects neutral particles. The requirements on the calorimeter are demanding. It should accurately and uniformly quantify the energy of different types of particles over an energy range spanning three orders of magnitude from 1 GeV up to 1 TeV. Also it should have a high granularity to determine the direction of neutral particles with high precision and resolve single particles within highly boosted and therefore collimated jets. At the same time the calorimeter needs to have fast response and low dead time to distinguish between bunch crossings.

Like the inner detector, the calorimeter is divided into the central barrel and the two end-cap parts. It also consists of three bigger subsystems with different functions. The *Electromagnetic Calorimeter* (EMCAL), which is responsible for detecting mainly electrons

## 2 The ATLAS Experiment

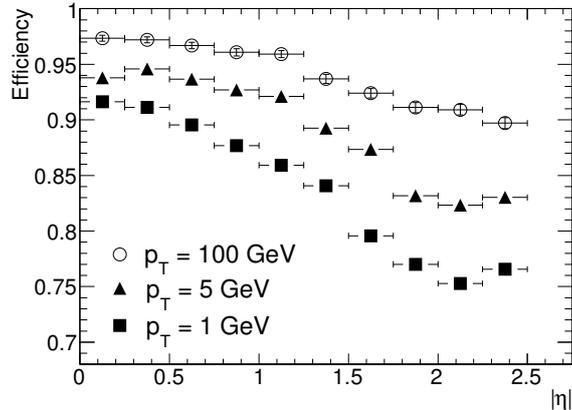


Figure 2.3: Track reconstruction efficiencies as a function of  $\eta$  for charged pions with different transverse momenta. This plot is taken from [33].

and photons, the *Hadronic Calorimeter* (HCAL), which surrounds the electromagnetic calorimeter and is responsible for detecting charged and neutral hadrons, and the *Forward Calorimeter* (FCAL), measuring particles with small angles to the beam axis ( $|\eta| > 3.1$ ), which is particularly important for the reconstruction of the missing transverse energy.

The EMCAL is a sampling calorimeter with liquid argon as active material and a lead absorber. The EMCAL is designed in an accordion shape with three layers and cells projective to the nominal interaction point. The accordion shape avoids gaps in  $\phi$ -direction between modules and adds to the mechanical stability. A detailed part of the EM barrel can be seen in Figure 2.5. One additional layer, placed before the actual calorimeter directly behind the solenoid magnet, is the *Presampler*. It is used to correct for the energy lost through early showering in the dead material in front of the calorimeter. In contrast to the other layers it consists of a thin layer of instrumented liquid argon and only covers pseudorapidities up to  $|\eta| < 1.8$ .

The first layer the calorimeter is the so called  $\eta$ -strip layer. It is very finely segmented in  $\eta$ -direction and is able to resolve the two photons from  $\pi^0$  decays to distinguish them from prompt photons.

The next layer, Layer 2 in Figure 2.5, is called the ‘middle layer’, which is the thickest part of the EM calorimeter (16 radiation lengths at  $\eta = 0$ ) and absorbs the largest part of the electromagnetic showers. It is followed by the third layer (back layer), which absorbs the longitudinal tails of the showers. The total depth of the EMCAL amounts to 24 radiation lengths at  $\eta = 0$ .

For pseudorapidities above 2.5 the EMCAL has only two layers and coarser readout cells. Thus, the range for precision measurements is limited to  $|\eta| < 2.5$ , as in the case of the silicon tracking detectors. Also in the transition region between the barrel and the end-cap ( $1.3 < |\eta| < 1.7$ , also called *gap region*) where the services (cables, cooling, etc.)

## 2 The ATLAS Experiment

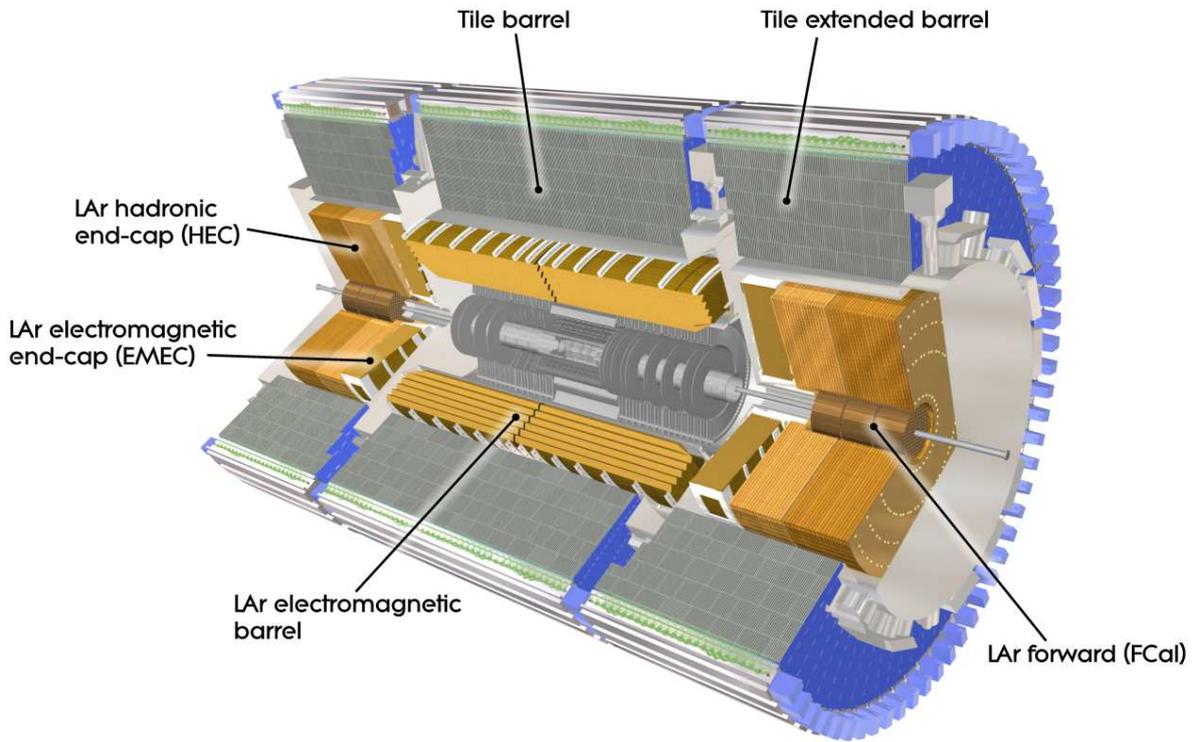


Figure 2.4: Cut-away view of the ATLAS Calorimeter System, from [33]

are routed out of the detector, the performance of the calorimeter is somewhat degraded.

The second big part of the calorimeter is the HCAL, again separated into the barrel part and the end-cap parts. The barrel is further divided into the central barrel and two extended barrels. The hadronic end-caps are similar in design to the EMCAL, being liquid argon sampling calorimeters, but use copper as the absorber material and have a planar geometry. In the barrel region the HCAL is designed as a steel structure with inserted scintillating plastic tiles (TileCal), which are read out by photomultipliers.

The FCAL is also a liquid argon calorimeter with copper absorbers in the first layer and tungsten in the second and third layer. It consists of rods embedded in a matrix of the absorber material with a liquid argon filled gap between them.

Detailed information about the resolution and the granularity of the different calorimeter subsystems are given in Table 2.3

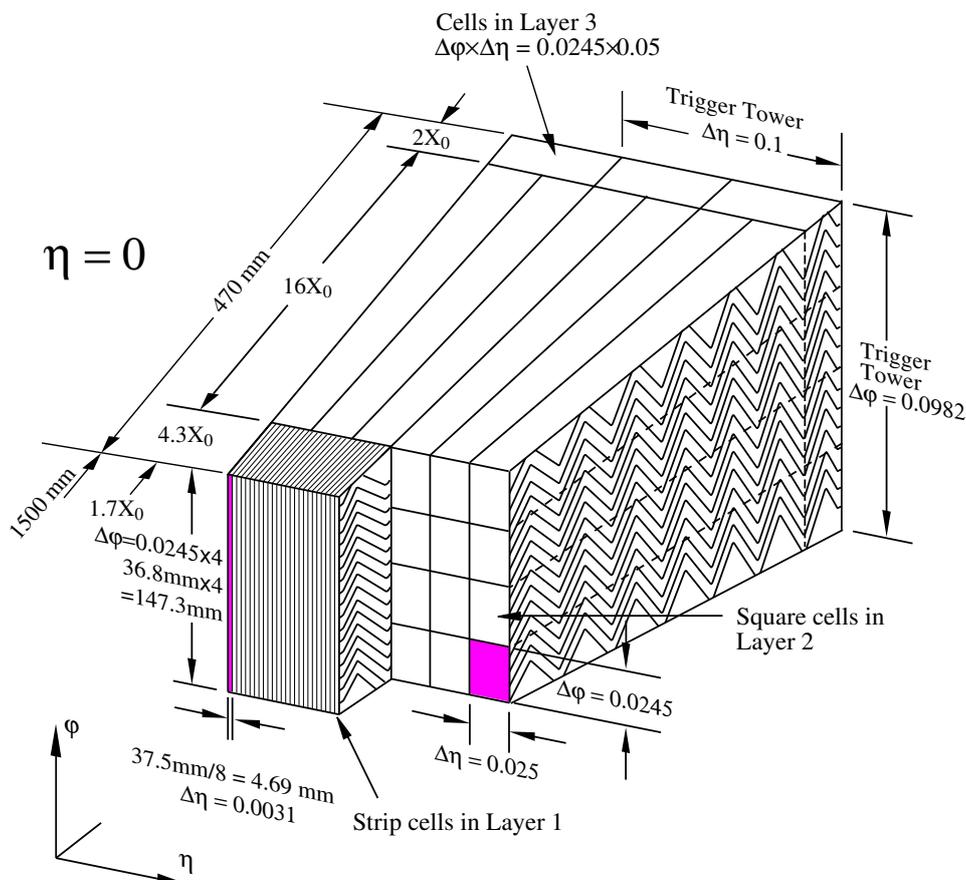


Figure 2.5: Schematic drawing of a part of the EM barrel calorimeter, taken from [33]

### 2.2.3 The Muon System

If properly reconstructed, muons can give a very clean signal for many physics channels. A major design goal for the ATLAS detector was therefore an excellent efficiency and resolution for the muon reconstruction independent of the inner detector.

The muon spectrometer is the biggest part of the detector, and determines the overall envelope size of ATLAS. A dedicated magnet system for the muon spectrometer – eight air core toroids in the barrel and two end-cap toroids – make an independent momentum measurement possible. The bending plane of the magnets is the  $R - z$  plane.

Several different technologies are used inside the muon system, but they can be divided in two functional groups: the triggering chambers consisting of *Thin Gap Chambers* (TGC) and *Resistive Plate Chambers* (RPC) and the precision chambers consisting of *Monitored Drift Tubes* (MDT) and *Cathode Strip Chambers* (CSC). The trigger chambers have a fast response time but low spatial resolution of several mm, the precision chambers have a resolution in the bending plane of 35 – 40  $\mu\text{m}$ . Together with the large lever arm this makes a momentum resolution of less than 10% for muons with an energy of 1 TeV possible. By

## 2 The ATLAS Experiment

Subsystem/Layer	Energy Resolution $\sigma(E)/E$	Granularity $\Delta\eta \times \Delta\phi$
<hr/>		
ECAL		
<hr/>		
Presampler	$\frac{10\%}{\sqrt{E}} \oplus 0.2\%$ in barrel	$0.025 \times 0.1$
$\eta$ -strip layer		$0.003 \times 0.1$
Middle	$\frac{12\%}{\sqrt{E}} \oplus 0.35\%$ in end-cap	$0.025 \times 0.025$
Back		$0.05 \times 0.025$
<hr/>		
HCAL		
<hr/>		
Barrel	$\frac{56\%}{\sqrt{E}} \oplus 5.5\%$	$0.1-0.2 \times 0.1$
End-Cap	$\frac{71\%}{\sqrt{E}} \oplus 5.8\%$	$0.1-0.2 \times 0.1-0.2$
<hr/>		
FCAL	$\frac{70\%}{\sqrt{E}} \oplus 3.0\%$	$0.15 \times 0.15$
<hr/>		

Table 2.3: Energy resolution and granularity of the different calorimeter subsystems.  $\oplus$  denotes that the two values are to be added in quadrature.

combining the track from the muon spectrometer with the tracks from the inner detector the resolution can be further improved.

### 2.2.4 The Trigger System

Due to budgetary and technological constraints, the maximum rate of events that can be recorded and stored is about 200 Hz, while the bunch crossings occur with a rate of 40 MHz. The number of recorded events thus must be reduced by a factor of  $2 \cdot 10^5$ , while ensuring that the events of interest are not discarded. The system responsible for this is called *Trigger*.

The Trigger system breaks down the decision to store or reject an event into three distinct levels named *Level 1* (L1), *Level 2* (L2) and *Event Filter* (EF). The L2 and EF together constitute the *High Level Trigger* (HLT). Each level further refines the trigger decision and tightens the criteria for selecting an event.

L1 makes the first decision to retain the data associated with a bunch crossing and passes the data on to the L2 trigger. Already at the L1 trigger a rejection factor of 400-500 is necessary, since the maximum data readout rate is limited to 75 kHz (upgradeable to 100 kHz), and the latency of the L1 trigger is designed to be smaller than  $2.5 \mu\text{s}$ , to keep the pipeline buffers on the front end electronics reasonably short. To reach these design goals, the L1 trigger is based on customised ASIC and FPGA hardware and is placed close to the detector to minimise signal latency in the cables. It also uses very fast but simple algorithms. They work with strongly reduced calorimeter granularity of  $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$  and information from the inner detector is not used at all, since the reconstruction of tracks is a complex and therefore time consuming task. The L1

## 2 The ATLAS Experiment

trigger mainly selects events which contain high  $p_T$  objects like electrons, photons, muons, hadronically decaying  $\tau$ 's or jets, with optional isolation criteria. It also can trigger on the missing  $E_T$  or the sum of the transverse energy.

The HLT is software based and runs on commercially available hardware. This hardware can be located on the surface, since latency issues are less pressing. The L2 trigger further refines the decision of the L1 trigger by using the full detector granularity and also tracking information and energy calibration. The algorithms used however are still simplified compared to the offline reconstruction. The L2 trigger is seeded by so called *Regions of Interest* (RoI), which are the coordinates and already reconstructed properties of interesting objects as defined by the L1 trigger. Each RoI can be examined on a different computation node in parallel, so the throughput is scalable. After the L2 trigger the rate is reduced to about 3.5 kHz.

In the last stage the EF reduces the event rate to the final goal of roughly 200 Hz by use of the same algorithms as the offline reconstruction. An identification as loose as possible within the 200 Hz budget is applied in this last step. If an event gets selected all the detector data of about 1.3 MB/s is written to dedicated output nodes and from there transferred to the CERN computing center for long time storage and further analysis. The total average storage bandwidth during a full luminosity run amounts to nearly 300 MB/s with a peak bandwidth of 600 MB/s.

## 3 Classification Methods

One of the biggest challenges for modern high energy experiments is to obtain a sizable number of events from a particular physical process, called *signal*, while minimising the number of events from unwanted processes, called *background*. Most of the time, when searching for new phenomena or measuring the values of certain model parameters, one needs a clean sample of a selected small number of specific processes.

Unwanted processes with the same signature as the signal are called *irreducible background* and nothing can be done to suppress them. They can for example be taken into account by modelling their contribution as precisely as possible. The background events with a final state different from the signal events are called *reducible*. They can in principle be distinguished from the signal, but might still contribute to the background if the signature is very similar to the signal.

The process of assigning each event to either the signal or the background is called *classification*. Classification problems are not unique to high energy physics, so there exists a large body of work on this topic. The general problem of classification can be defined as follows:

given a vector  $\mathbf{x} = (x_1, \dots, x_N)$  of characterising variables (pattern), determine the *a posteriori* probability  $P(C_i|\mathbf{x})$  for the pattern to belong to class  $i$ .

The  $N$ -dimensional space spanned by the patterns is called *feature space*. For  $N = 1$  the problem is called *univariate*, otherwise it is a *multivariate* problem. In *separable* problems there is no overlap in the feature space between the classes, so a  $(N-1)$ -dimensional hypersurface exists, which partitions the feature space into volumes containing only patterns of one class. In general, problems are not separable, but still one can construct a hypersurface by defining an appropriate error measure for misclassified patterns and minimising this error.

### 3.1 Multivariate Techniques

Very often the analytical form of  $P(C_i|\mathbf{x})$  is unknown. A lot of methods have been devised to approximate this probability from data, which belong to a known class. A central theorem of probability theory is Bayes' theorem:

$$P(C_i|\mathbf{x}) = P(C_i) \frac{P(\mathbf{x}|C_i)}{P(\mathbf{x})} \quad (3.1)$$

### 3 Classification Methods

Here  $P(C_i)$  is the *a priori* probability of an event belonging to class  $i$ ,  $P(\mathbf{x}|C_i)$  is the probability of the pattern  $\mathbf{x}$  if the event belongs to class  $i$  and  $P(\mathbf{x}) = \sum_i P(\mathbf{x}|C_i)$  is the *a priori* probability of pattern  $\mathbf{x}$ . In high energy physics applications,  $P(C_i)$  is known from the cross sections of the event classes, typically after applying some analysis specific event selection criteria. The class specific probabilities  $P(\mathbf{x}|C_i)$  are often independent of this event selection. Therefore arbitrary prior probabilities can be assigned to all classes, and the resulting probability has to be matched to the true probability from analysis to analysis.

By modelling  $P(\mathbf{x}|C_i)$  for all classes from a known *training sample*, one can calculate the *a posteriori* probability for data not contained in the training sample. Strictly speaking, this is only possible in the limit of a infinite number of training patterns. All the different multivariate classifiers therefore try to approximate  $P(\mathbf{x}|C_i)$  with a finite number of training patterns. A common property of all such estimators is the so called *bias-variance-tradeoff*, which is a trade-off between a model with too little degrees of freedom, which cannot approximate the training data very well (bias), and a model with too much flexibility, which is very sensitive to statistical fluctuations within the training sample (variance) and will not describe independent data very well. The latter is also called overtraining. One common method to deal with this problem, is to split the data into a training and a test sample. The test data can be used to judge the performance of a method on ‘new’ data.

In the following the number of classes is restricted to two, corresponding to a signal and background class. In this case  $P(C_{background}|\mathbf{x}) = 1 - P(C_{signal}|\mathbf{x})$  holds, and a binary classification is sufficient.

## 3.2 Likelihood Method

A commonly used approach to classification is the *Likelihood Method*. It is based on the densities of data points in the feature space of the signal and background samples. The density  $L_{S/B}(\mathbf{x}) = P(\mathbf{x}|C_{S/B})$  is called *likelihood*, alternatively the term *probability density function* (PDF) is frequently used.

A PDF can be extracted from the training sample by various density estimation methods. With unlimited statistics it would contain the maximum information available and hence a classifier build on the full PDF will be optimal. But in general it will be impossible to get a good approximation for the full N-dimensional form. This follows from the fact, that the volume of the feature space grows exponentially with the dimension, and statistical fluctuations will be overwhelming with practically attainable sample sizes. This problem is dubbed *the curse of dimensionality* in the statistical literature.

To avoid this obstacle, one can make a factorisation ansatz for the likelihood by assuming the different variables to be statistically independent:

$$L_{S/B}(\mathbf{x}) = \prod_{i=1}^N L_{i,S/B}(x_i) \tag{3.2}$$

### 3 Classification Methods

where  $L_{i,S/B}(x_i)$  is the likelihood for the signal or background when projecting onto the  $i$ th dimension. Therefore this method is called *projective Likelihood Method*. This method discards all correlations between the variables, and as a consequence a degraded performance is expected if such correlations exist. Even though this is the case in many applications, the disadvantage incurred by ignoring the correlations is often compensated by a better estimation of the likelihoods. This is on one hand a consequence of the larger density of data points when projecting out one dimension, and on the other hand the existing methods for density estimation in one dimension are more robust and easier to optimise. The factorisation can also be easily extended to project onto subspaces with dimensionality greater than one, to include the correlations between certain subsets of variables.

If one has approximated the PDFs, classification is done by mapping each vector in the feature space to a real value  $d_L$  via the logarithm of the likelihood ratio:

$$d_L = \log \frac{L_S(\mathbf{x})}{L_B(\mathbf{x})} = \log \prod_{i=1}^N \frac{L_{i,S}(x_i)}{L_{i,B}(x_i)} = \sum_{i=1}^N \log \frac{L_{i,S}(x_i)}{L_{i,B}(x_i)} \quad (3.3)$$

The ratio of the likelihoods is called *odds*. It gives the probability of one hypothesis over another, where a value of 1 means equal probability and values smaller or larger than unity signify a lower or higher probability for the tested hypothesis respectively. Taking the logarithm of the odds provides for better numerical behaviour when calculating  $d_L$  on limited precision hardware. By requiring a minimum value for  $d_L$ , above which an event is classified as a signal event, a discrimination from the background can be achieved.

As already mentioned, there exists a big variety of different techniques for approximating the PDFs from training data. If the underlying analytic function of the PDF is known, one can do a parametric estimation to find the appropriate parameters. Since this is rarely the case, non-parametric methods for density estimation are widely used, the simplest of which is creating a histogram and normalising it. Other methods, such as fitting of splines, can be used on top of a histogram to give a smooth PDF. But methods relying on histograms are sensitive to the bin size, as well as to the position of the bins, so the resulting estimate is not always very reliable.

Another family of methods is based on kernel density estimation (KDE). The simplest form of KDE for a density  $p(x)$  given the 1-dimensional training samples  $d_1, \dots, d_N$  is of the form [20, 21]:

$$p(x) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{x - d_i}{h}\right), \quad (3.4)$$

where  $K(\cdot)$  is the kernel function and  $h$  is called *bandwidth*. The bandwidth controls how close the estimate will be to the training data. Smaller bandwidths lead to less bias and more variance, and vice versa. A kernel function is a non-negative integrable function which satisfies:

$$\int_{-\infty}^{+\infty} K(u) du = 1, \quad (3.5)$$

### 3 Classification Methods

$$K(u) = K(-u) \quad \forall u \in \mathbb{R}, \quad (3.6)$$

$$K(0) \geq K(u) \quad \forall u \in \mathbb{R}, \quad (3.7)$$

and

$$K(u_1) \geq K(u_2) \quad \text{for } |u_1| > |u_2|. \quad (3.8)$$

Often used kernels are for example the Gaussian kernel

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}, \quad (3.9)$$

or the Epanechnikov kernel

$$K(u) = \begin{cases} \frac{3}{4}(1 - u^2), & \text{for } |u| \leq 1, \\ 0, & \text{for } |u| > 1. \end{cases} \quad (3.10)$$

It has been shown, that the performance of the density estimation is not very dependent on the actual form of the kernel [36], rather the optimal bandwidth depends on the choice of the kernel function. The optimal bandwidth also depends on the form of the unknown density like

$$h = C \left( \int (p''(x))^2 dx \right)^{-1/5} N^{-1/5}. \quad (3.11)$$

where  $C$  depends only on  $K(u)$ . If one assumes a normal distribution for  $p(x)$  and a Gaussian kernel is used, a rule-of-thumb for the optimal bandwidth is

$$h_{opt} \approx 1.06 \sigma_x N^{-1/5}, \quad (3.12)$$

where  $\sigma_x$  is taken to be the RMS of the data.

Even the optimal fixed bandwidth does not always produce very good results. If the true density contains sharp peaks and long tails, it is often impossible to find a single bandwidth, which handles both regions equally well. Therefore it is beneficial to vary the bandwidth over the training data, based on the local features of the data. One approach is to start from a fixed bandwidth  $h_{na}$  and define the adaptive bandwidth as

$$h_a(x) = \frac{h_{na}}{\sqrt{\tilde{p}(x)}}, \quad (3.13)$$

so the bandwidth gets smaller in regions with high density, where the approximation is more reliable, and bigger in the tails. For the *pilot density*  $\tilde{p}(x)$ , one can use the result of a first pass with the fixed bandwidth  $h_{na}$ .

Another approach is to determine the adaptive bandwidth with a nearest-neighbour method. Here the bandwidth at each point of the training sample is determined by the average distance to a fixed number of nearest neighbours in the sample. This has the advantage that near steep rises of the density – so in areas with high curvature – the

### 3 Classification Methods

bandwidth is also reduced, and these ‘edges’ get smeared out less. The parameter that controls the bias-variance-tradeoff in this case is the number of neighbours to include in the bandwidth selection.

The method presented so far has some problems with distributions common to high energy physics applications. One problem is the boundary problem, which arises if the real density is not continuous. An example might be an energy fraction, which has valid values only between zero and one. In this case the distribution gets smeared out near the boundary and unphysical values get a nonzero probability. Also, if only the physical region is used, the density function is not normalised anymore, and simply renormalising it can introduce bias away from the boundaries. Another related problem is the systematic underestimation of peak heights. Since naturally a large fraction of events will have a value in the vicinity of the peak, a good modelling of the peaks is especially important.

The simple method, by means of the kernel functions, defines a weighted neighbourhood of the point to be evaluated and fits a constant to the training points in this neighbourhood [15]. An obvious extension of this scheme is to fit more complicated functions to this neighbourhood, to locally approximate also the derivatives of the density [4]. The easiest choice is to locally fit low order polynomials, which is motivated by a Taylor expansion of the density at each point. This method reduces some of the problems of the simple methods, like the boundary problem and strong bias in regions with high curvature.

Despite its simplicity and its disregard of the correlations between the identifying variables, the Likelihood method often gives surprisingly good results [17]. Apart from that, it has the advantage of being very easy to interpret and robust against badly estimated variables.

## 3.3 Artificial Neural Networks

A somewhat different approach to classification are *artificial neural networks* (ANN). They were developed in an attempt to mimic the way biological nervous systems work. The first mathematical models for neural networks were proposed in the 1940’s, but with the technology of the time, they were not much of practical use. Renewed interest in the 1980’s led to a rapid development of ANNs into a powerful method for ‘machine learning’ problems. Nowadays the applications of neural networks are manifold, not only in academic, but also in commercial contexts. The way ANNs make the problem of modelling the high dimensional full probability density more tractable is by modelling only the hypersurface mentioned before, which maximally separates the different classes under some error measure.

ANNs are based on connections between independent entities. More precisely a neural network is a directed and weighted graph. The knots of the graph are the *neurons* and the edges are called *connections*. Connections transfer data from one neuron to another. Each neuron can have multiple incoming connections and also multiple outgoing connections to

### 3 Classification Methods

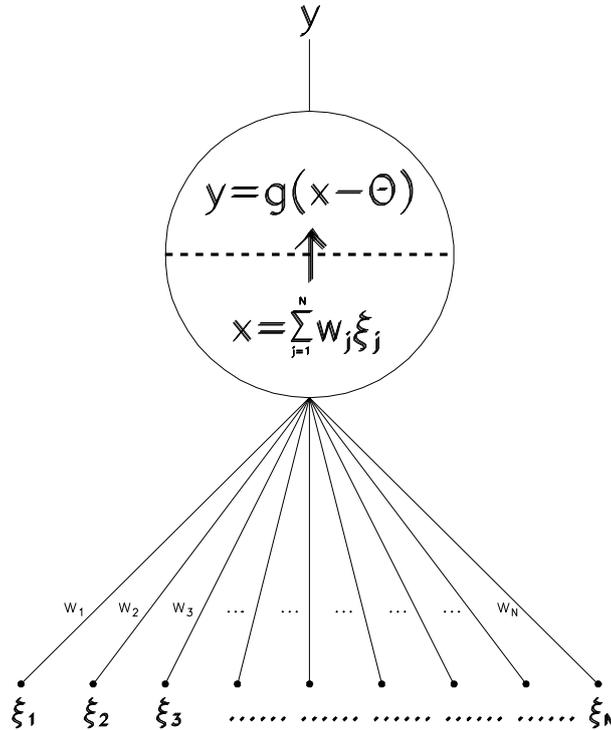


Figure 3.1: graphical representation of a neuron. The inputs  $\xi$  and the connection weights  $w_i$  are used to calculate the effective input  $x = \sum_i w_i \xi_i$  to the activation function  $g(\cdot)$ . The threshold  $\Theta$  is used to select a ‘working point’ of the activation function. The figure is taken from [38].

other neurons. The outgoing connections all carry the same data. Neurons compute their output as a function of the inputs in two steps. The first step is to combine the inputs into one value, usually by taking the sum of the inputs weighted by the corresponding connection weights. The second step is to apply an *activation function* to the result of the first step. So each neuron does the following computation when the inputs  $\mathbf{x} = (x_1, \dots, x_i)$  with the connection weights  $w_1, \dots, w_i$  are applied to it

$$o(\mathbf{x}) = g\left(\sum_i w_i x_i - \Theta\right). \quad (3.14)$$

Here  $g(x)$  is the activation function and  $\Theta$  is called the *threshold*. A graphical representation of a neuron is shown in Figure 3.1. The threshold is an adjustable parameter of each neuron, and is often taken into account by adding an additional connection, which is connected to a node with a constant output of one. The negative weight of this connection takes the role of the threshold.

### 3 Classification Methods

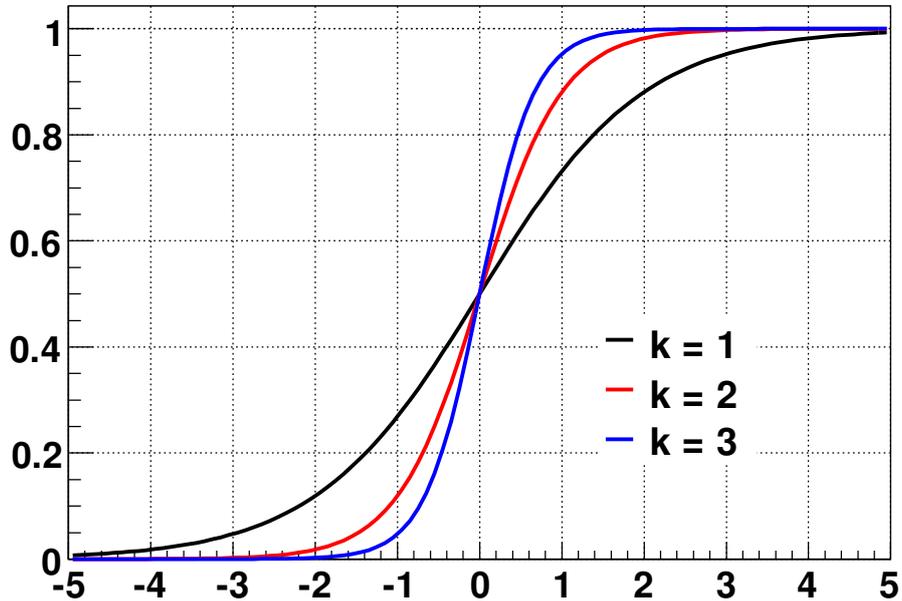


Figure 3.2: The sigmoid function  $g(x) = \frac{1}{2}(1 + \tanh(kx))$  for different values of the *inverse temperature*  $k$ .

The activation function is often the *sigmoid function* as shown in Figure 3.2, which can be understood as a softened step function. A neural network consisting of only one such neuron and multiple input connections is called *perceptron*.

A perceptron can describe a  $(n-1)$ -dimensional hyperplane in the feature space, and is only viable for linear problems. Many problems in high energy physics do not fall in that category. A simple and most often used extension of the perceptron is the *multi layer perceptron* (MLP) as shown in Figure 3.3. Here neurons are arranged in several ordered layers and each node is connected to all nodes in the next layer. The first layer is called the *input layer* and its nodes simply represent the input data. The last layer is called *output layer* and for classification problems it often has as many nodes, as there are different classes. All the layers between the input and output layer are called *hidden layers*. Since in an MLP data always flows in the forward direction from one layer to the next, it is an example of a feed-forward network.

Mathematically an MLP with one hidden layer with  $n_h$  nodes and one output node is represented by the expression

$$o(\mathbf{x}) = g \left[ \sum_{j=1}^{n_h} w_{o,j} \cdot g \left( \sum_{i=1}^n w_{h,ij} \cdot x_i - \Theta_{h,j} \right) - \Theta_o \right], \quad (3.15)$$

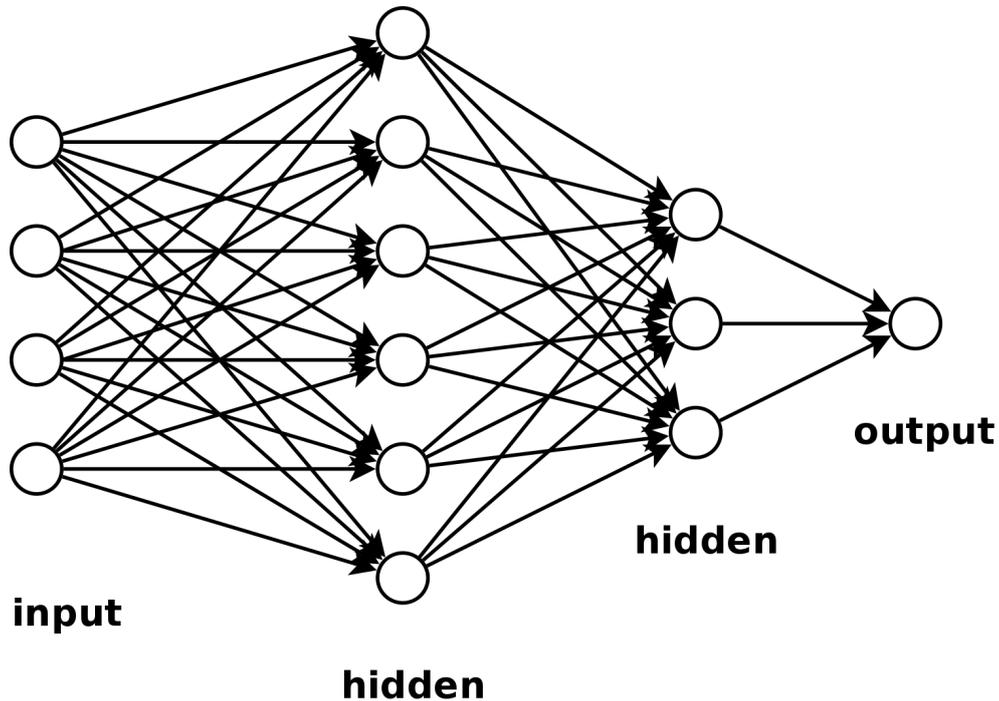


Figure 3.3: a multi layer perceptron with four input variables, one output node and two hidden layers. The signals are propagated from the input layer to the output layer as indicated by the arrows.

where  $w_{h,ij}$  are the weights of the connections going from input node  $i$  to node  $j$  in the hidden layer,  $w_{o,j}$  are the weights of the connection going from hidden node  $j$  to the output node and  $\Theta_{h,j}$  and  $\Theta_o$  are the threshold of the hidden and output nodes. One can see that this expression is not linear in  $\mathbf{x}$  anymore if  $g(\cdot)$  is in turn not linear. Given enough nodes in the hidden layer, an MLP with one hidden layer can approximate any convex boundary in the feature space. With two hidden layers any boundary can be described [14].

To prepare an MLP for a specific problem, one has to choose a topology and find the correct values of the weights and thresholds to describe the data. The threshold value is not explicitly mentioned anymore. It can instead be handled like a weight, by creating a special node, which has a fixed output of one and is connected to all nodes. The weights of these connections then take the role of the threshold values.

Training the network with training data works similar to fitting a function to data. An error function, quantifying the ability of the parameters to describe the data, has to be minimised numerically. A unique output vector is assigned to every class. The error is calculated by comparing the actual output of the network for each training pattern to the target value of the class it belongs to. In a classification setting with two classes, signal and background, and only one output node the target output value of the network is set to 1 for signal and to 0 for background patterns. The actual output of the neural network

### 3 Classification Methods

will have some distribution between the two target values. The error function commonly used is the summed square error

$$E(\mathbf{w}) = \frac{1}{2} \sum_i [t_i - o_i(\mathbf{w})]^2, \quad (3.16)$$

where  $\mathbf{w}$  is the vector of all the weights, the sum goes over all training events,  $t_i$  is the target value for training event  $i$  and  $o_i(\mathbf{w})$  is the output value of the network for event  $i$ .

Training of the network is done iteratively in several steps, since it is not possible to find the optimal weights analytically in all but the simplest cases. The steps are

- propagate all the training vectors through the network
- compare the outputs to the target values and calculate the error
- update the weights to minimise the error

One such cycle is called an *epoch*. It should be repeated until the network converges and the error does not decrease anymore.

The question remains how to minimise the error. The simplest algorithm for doing this is called *backpropagation*. Backpropagation is an algorithm to find the minimum of the error function by changing the weights in small steps in the opposite direction of the gradient of the error with respect to the weights. It is important for this algorithm, that the transfer function is differentiable, which is the case for the aforementioned sigmoid function  $g(x) = \frac{1}{2} (1 + \tanh(kx))$ . The weights in epoch  $t + 1$  are computed as

$$\mathbf{w}^{t+1} = \mathbf{w}^t - \eta \nabla_{\mathbf{w}} E \quad (3.17)$$

with the learning rate  $\eta$ . The learning rate has to be chosen carefully, since values that are too small can lead to very slow convergence and make it more likely to converge into local minima, and learning rates that are too large can lead to oscillations around the minimum or even to divergence. Moreover there is no fixed rule for determining the optimal learning rate, since it is very dependent on the specific problem to be trained. Another problem in the training can arise from regions of the error surface with very small slope, where the backpropagation will make almost no progress. This will happen if a lot of neurons have a very high or very low input value, where the derivative of the sigmoid function goes to zero. Therefore, at the beginning of the training the weights need to be initialised with random values in a small interval around zero, so that several attempts with different starting weights are often enough to avoid difficult regions during the training. Slightly different network topologies might also help to alleviate convergence problems.

Several extensions of the basic backpropagation algorithm have been proposed, in order to improve the speed of convergence and to reduce the need for trial and error in finding good parameters for the learning algorithm. One extension, that is used very often, is to introduce a momentum term in (3.17)

$$\mathbf{w}^{t+1} = \mathbf{w}^t - \eta \nabla_{\mathbf{w}} E + \alpha \Delta \mathbf{w}^t, \quad (3.18)$$

### 3 Classification Methods

where the momentum factor  $\alpha$  should be chosen between 0 and 1. The momentum term dampens oscillations around the minimum and also helps to bridge plateaus and shallow local minima. Another simple extension is the *stochastic learning*, where the weights are not updated after the full training sample has been propagated through the network, but after a small number of training vectors. In essence this uses a different error surface for each update and adds some noise to the learning procedure, which also helps to escape problematic areas of the error surface. The assumption is that the average effect is still to converge towards the global minimum of the full error, which often holds in practise for large enough training samples.

Another group of methods tries to adapt the speed of the weight change during the training with heuristic methods. One such method is called *resilient propagation* (RPROP) [25], which is also the method used for the present work. RPROP uses only the sign of the partial derivative for each weight and also adapts the step size for each weight individually. For the first epoch, the size of the update step  $\Delta_0$  has to be chosen by the user, but its value is not critical, so a conservative value of 0.1 – 0.5 can be used. After each epoch the following value is added to each weight  $w_i$ :

$$\Delta w_i^{(t)} = \begin{cases} -\Delta_i^{(t)}, & \text{if } \frac{\partial E}{\partial w_i}^{(t)} > 0, \\ +\Delta_i^{(t)}, & \text{if } \frac{\partial E}{\partial w_i}^{(t)} < 0, \\ 0 & \text{otherwise.} \end{cases} \quad (3.19)$$

The value of  $\Delta_i^{(t)}$  is determined by the value for the previous epoch by

$$\Delta_i^{(t)} = \begin{cases} \eta^+ \cdot \Delta_i^{(t-1)}, & \text{if } \frac{\partial E}{\partial w_i}^{(t-1)} \cdot \frac{\partial E}{\partial w_i}^{(t)} > 0, \\ \eta^- \cdot \Delta_i^{(t-1)}, & \text{if } \frac{\partial E}{\partial w_i}^{(t-1)} \cdot \frac{\partial E}{\partial w_i}^{(t)} < 0, \\ \Delta_i^{(t-1)} & \text{otherwise,} \end{cases} \quad (3.20)$$

where

$$0 < \eta^- < 1 < \eta^+.$$

$\eta^+$  and  $\eta^-$  are parameters that can be set by the user, but the recommended values of  $\eta^+ = 1.2$  and  $\eta^- = 0.5$  are suitable for many problems.

Expressed in words the algorithm proceeds as follows: after each epoch, if the sign of the partial derivative of the error with respect to the weight stayed the same compared to the previous epoch, increase the step size to speed up the convergence when away from a minimum. If the sign of the derivative changed, decrease the step size for this weight, since this means a minimum was crossed.

The RPROP algorithm has several characteristics that make it a good choice for training neural networks:

- it is very fast both in terms of update steps needed as well as in terms of computation time

### 3 Classification Methods

- it adapts to problems where the input variables have different scales, through the use of individual learning rates per weight
- it has few parameters, which are easy to tune

There exists no easy rule, to determine the number of hidden nodes needed for an MLP to faithfully describe the underlying model of the training data. Therefore choosing a topology is often a tedious trial-and-error process. A network with too little hidden nodes cannot describe the training data very well, so several configurations might exist that have a similar (big) error. In other words there tend to be a lot of local minima. Such networks are often hard to train consistently. If on the other hand, the network is too large it often ‘memorises’ only the training data, and it will have a poor performance on new data, hence the network is overtrained. The susceptibility to overtraining depends very much on the amount of training data. A big number of training samples can constrain the weights of the network better, even if it is more complex than necessary.

One simple way to control for overtraining, is to split the available sample data into a training set and a statistically independent validation set. Typically one uses about one third of the data for validation. Only the training set is presented to the network during the training. One can then compare the performance of the network for the training and the validation data, and if it is very different the network is probably overtrained. Based on this, an ad-hoc method to get to a network that gives a good performance on the training data as well as on new data, called *early stopping*, is very often used. For this method one should start with a network with slightly too many weights, so that overtraining will occur if trained to the minimal error. During the training the error for the validation sample is periodically calculated, and if the error starts to rise the training is stopped. Although it is not obvious that this method leads to better generalisation, experience shows that it often does.

# 4 Reconstruction and Identification of $\tau$ Leptons

The reconstruction and identification of all lepton flavours is important for the physics program at the LHC, since they allow to distinguish certain rare processes from the large QCD background.  $\tau$  leptons are especially difficult in this respect, due to the variety of possible decay modes, some of them containing hadrons in the final state. Therefore the goal of this work was to improve the existing algorithms for the reconstruction and identification of  $\tau$  leptons.

The leptonic decay modes of the  $\tau$  lepton cannot be efficiently distinguished from electrons and muons produced in the hard interaction. They are usually treated with the dedicated electron and muon reconstruction algorithms. For the purpose of  $\tau$  identification, only the hadronic decay modes are thus considered. In the following sections some properties of the hadronic  $\tau$  decay products will be discussed, and the existing techniques to reconstruct and identify them are introduced.

## 4.1 Properties of Hadronic $\tau$ Decay Products

The  $\tau$  lepton can decay into an odd number of charged hadrons, a neutrino, and eventually one or more neutral hadrons with a total branching ratio of about 65%. The different hadronic decay modes can be grouped by their charged multiplicity into so called *n-prong* decays, where n is the number of charged hadrons. One mainly considers 1-prong and 3-prong decays for inclusive  $\tau$  identification, as the other decay modes have a combined branching ratio of well under 1 percent. The charged hadrons are mainly  $\pi^\pm$ , in a few percent of the cases also  $K^\pm$ . The neutral hadrons consist mainly of  $\pi^0$ , which will decay into two photons inside the detector, and sometimes also neutral kaons (in  $\approx 1 - 2\%$  of the cases). The absolute and relative branching ratios for the most important decays are shown in Table 4.1. At the LHC  $\tau$  leptons are typically produced with a boost of at least several GeV, and thus their decay products will form a collimated jet of particles, approximately collinear with the original  $\tau$  lepton. The combined visible decay products of the  $\tau$  lepton are referred to as a  $\tau$  *jet*. For the relevant energies  $\gtrsim 5$  GeV all the visible decay products are contained in a cone<sup>1</sup> of radius 0.4 around the direction of the  $\tau$  jet.

Due to the proper lifetime of the  $\tau$  lepton of  $290.6 \pm 1.0$  fs [7]) the charged particles from the decay will be produced at a mean distance of  $\beta\gamma c\tau = 87.11$   $\mu\text{m}$ , resulting in

---

<sup>1</sup>a cone is defined by all the points with  $\sqrt{\Delta\eta^2 + \Delta\phi^2}$  smaller than a fixed value

## 4 Reconstruction and Identification of $\tau$ Leptons

$\tau$ decay mode	BR <sub>abs</sub> [%]	BR <sub>lep/had</sub> [%]
leptonic	35.2	100
$e^- \bar{\nu}_e \nu_\tau$	17.8	50.6
$\mu^- \bar{\nu}_\mu \nu_\tau$	17.4	49.4
hadronic	64.8	100.0
1-prong	49.5	76.4
$\pi^- \nu_\tau$	10.9	16.8
$\pi^- \pi^0 \nu_\tau$	25.5	39.4
$\pi^- 2 \pi^0 \nu_\tau$	9.3	14.4
$\pi^- 3 \pi^0 \nu_\tau$	1.0	1.5
$\pi^- \bar{K}^0 \nu_\tau$	0.9	1.4
$K^- \geq 0 \pi^0 \nu_\tau$	1.6	2.5
others	0.4	0.5
3-prong	15.2	23.5
$\pi^- \pi^+ \pi^- \nu_\tau$	9.3	14.4
$\pi^- \pi^+ \pi^- \pi^0 \nu_\tau$	4.6	7.1
$\geq 1 K^\pm \geq 1 \pi^0 \nu_\tau$	0.9	1.4
others	0.4	0.5
5-prong	0.1	0.2

Table 4.1: Relevant decay modes of the  $\tau$  lepton with their branching ratios. The given numbers are experimental values taken from [7]. Errors are not shown for clarity, but are at the percent level for the most prominent decay modes. The last column gives the branching ratio relative to all leptonic (rows 2-4) or all hadronic decay modes (rows 5-19).

## 4 Reconstruction and Identification of $\tau$ Leptons

reconstructed tracks that usually have an impact parameter measurably different from zero. In the case of 3-prong decays the reconstruction of a secondary vertex is often possible.

Together, all these features of hadronic  $\tau$  decays lead to the following signatures of  $\tau$  jets:

- one, three or five charged tracks with a total charge of one
- a hadronic energy deposition in the calorimeter with a narrow shower shape
- the possible existence of neutral electromagnetic clusters from photons coming from  $\pi^0$  decays
- the possible existence of electron tracks and clusters from photon conversion and to a lesser extent dalitz pairs.
- tracks consistent with coming from a secondary vertex for multi prong decays

The detailed properties of this signature can be used to separate  $\tau$  jets from jets which are initiated by strongly interacting particles. The latter ones will be called *QCD jets* and constitute the dominant background for the identification of hadronically decaying  $\tau$  leptons.

### 4.2 Reconstruction of Hadronically Decaying $\tau$ Leptons

The reconstruction of  $\tau$  leptons in the ATLAS software framework *ATHENA* is a so called combined reconstruction algorithm, because it is built upon several reconstructed objects from lower level algorithms. These are mainly tracks reconstructed from the inner detector information and cells and energy clusters reconstructed from the calorimeter data. The low level objects are combined to decide whether their properties are compatible with originating from a hadronic  $\tau$  decay.

When the work on this thesis started, there were two loosely integrated  $\tau$  reconstruction and identification algorithms available in the *ATHENA* software. They are called ‘tauRec’ and ‘tau1P3P’. Each follows a somewhat different strategy, but, as would be expected, there is also an overlap in certain aspects. Both algorithms are integrated into the same package and share the basic infrastructure and the data representation of the common properties of the reconstructed  $\tau$  object, like kinematics and the association of lower level objects. The common C++ class representing  $\tau$  jets is called *TauJet*. Objects of this classtype can have further algorithm specific information attached to them.

To assess the performance of the  $\tau$  identification algorithms, the important figures of merit are the *signal efficiency* and the *background rejection*.

## 4 Reconstruction and Identification of $\tau$ Leptons

The signal efficiency quantifies how many real  $\tau$  jets are identified as such. It is defined by

$$\epsilon_\tau = \frac{\text{number of identified } \tau \text{ jets}}{\text{total number of true } \tau \text{ jets}}. \quad (4.1)$$

The true  $\tau$  jets are the  $\tau$  jets formed by the visible decay products of hadronically decaying  $\tau$  leptons, which have  $p_T > 10$  GeV and  $|\eta| \leq 2.5$ . A  $\tau$  jets is counted as identified, if a reconstructed  $\tau$  object, which passes some identification criteria, is found within a 0.2 radius around the  $\tau$  jet direction.

The background rejection is a measure of how many QCD jets are wrongly identified as  $\tau$  jets. The definition of the rejection used in this study, if not stated otherwise, is

$$R_{QCD} = 1/\epsilon_{QCD} = \frac{\text{number of true jets}}{\text{number of jets identified as } \tau \text{ jets}}. \quad (4.2)$$

The true jets are the ‘Cone4TruthJets’, which are reconstructed by a jet algorithm running over the stable<sup>2</sup> truth particles. Only true jets with  $p_T > 10$  GeV and  $\eta \leq 2.5$  are considered. A true jet is considered as identified, if it is within a 0.2 radius of a  $\tau$  object that passes some identification criteria. In the rest of this thesis rejections will be given as function of the true visible transverse momentum. For  $\tau$  jets this is the transverse momentum of the  $\tau$  jet, and for QCD jets this is the transverse momentum of the *Cone4TruthJet* matched to the  $\tau$  candidate.

Signal efficiency and background rejection have to be considered together, since a higher signal efficiency leads to a higher fraction of misidentified background. If a method shows a higher rejection at a given signal efficiency compared to another method, than the former can better discriminate between  $\tau$  jets and QCD jets for this specific signal efficiency. Most classifiers allow to choose an efficiency working point by choosing a minimum value of the classifier output above which a  $\tau$  object is considered to be a true  $\tau$ . By plotting the background rejection against the signal efficiency for different methods, one can directly compare their performance.

In the following both tauRec and tau1P3P will be described the way they were at the time the presented work on them started. Of course both of them evolved in the intervening time in different respects, and not all of the active ongoing work can be fully presented here.

### 4.3 The tauRec Package

The first of the two  $\tau$  reconstruction algorithms is strongly based on calorimetric information, in analogy to the reconstruction of QCD jets. Candidates for hadronically decayed  $\tau$  leptons are seeded by calorimeter clusters<sup>3</sup>. Around each seed the calorimeter cells are used with the full granularity of the calorimeter to calculate several quantities describing the shape of the shower. The same cells are also used to reconstruct the energy and

<sup>2</sup>In this context a particle is considered stable if it does not decay before reaching the calorimeter

<sup>3</sup>a cluster is a spatially localised energy deposition

direction of the  $\tau$  jet after applying a calibration procedure. In addition nearby tracks are associated to the  $\tau$  candidate and are used to calculate further quantities used for identification.

The discriminating variables are combined into a likelihood ratio, where the PDFs have been trained separately for different  $E_T$  ranges.

### 4.3.1 Seeding

The clusters used as input to tauRec are created by a sliding window algorithm, which creates the so called ‘combined clusters’. The sliding window algorithm works on calorimeter *towers*, which are summed up cells of all calorimeter layers over a grid of  $\Delta\eta \times \Delta\phi = 0.1 \times 2\pi/64$ . A  $5 \times 5$  window is put at each point of this grid, and a cluster is created if the summed up transverse energy exceeds the threshold of 15 GeV. If two clusters overlap, only the cluster with the higher transverse energy is reconstructed. Finally the position of the cluster is determined by the central  $3 \times 3$  towers of the cluster.

Only combined clusters with  $|\eta| < 2.5$  are used as seed for  $\tau$  candidates, which coincides with the coverage of the precision calorimeters. The efficiency to reconstruct a true  $\tau$  jet as a combined cluster is shown in Figure 4.1 as a function of the transverse momentum that the visible  $\tau$  decay products carry. A true  $\tau$  jet is considered as reconstructed, if a cluster is found with a barycentre within  $\Delta R < 0.2$  of the true  $\tau$  jet direction. The reconstruction efficiency rises from 12% at 15 GeV to 90% at 25 GeV and saturates at about 99% for transverse momenta above 30 GeV.

### 4.3.2 Identification Variables

After a cluster has been reconstructed, all calorimeter cells around the cluster barycentre within a cone with  $\Delta R < 0.4$  are collected. Cells that have a measured energy less than two times their expected noise are removed from the list of cells. Noise in the energy measurement arises from electronic noise and fluctuations of the pile up energy. The cut on the cell energy is applied to suppress this noise.

Tracks are associated to the  $\tau$  candidate by looking for tracks inside a cone of  $\Delta R < 0.3$  around the cluster centre with a transverse momentum of at least 2 GeV. The considered tracks are reconstructed by the default track reconstruction packet of the ATHENA software, called ‘new Tracking’ [30]. At the time of implementation of the original version of tauRec, the default track reconstruction package was called ‘xKalman’, where tracks were required to meet significantly tighter criteria with regard to the quality of the reconstructed track.<sup>4</sup> These criteria are given in [11]. This should be kept in mind, when comparing results presented here to the reference.

Several variables are calculated from the calorimeter cells and the associated tracks. These variables try to exploit some characteristic properties of  $\tau$  jets to distinguish them

---

<sup>4</sup>The quality of the track reconstruction is determined with criteria like the number of hits in different detector systems and the compatibility of the fitted trajectory with these hits.

#### 4 Reconstruction and Identification of $\tau$ Leptons

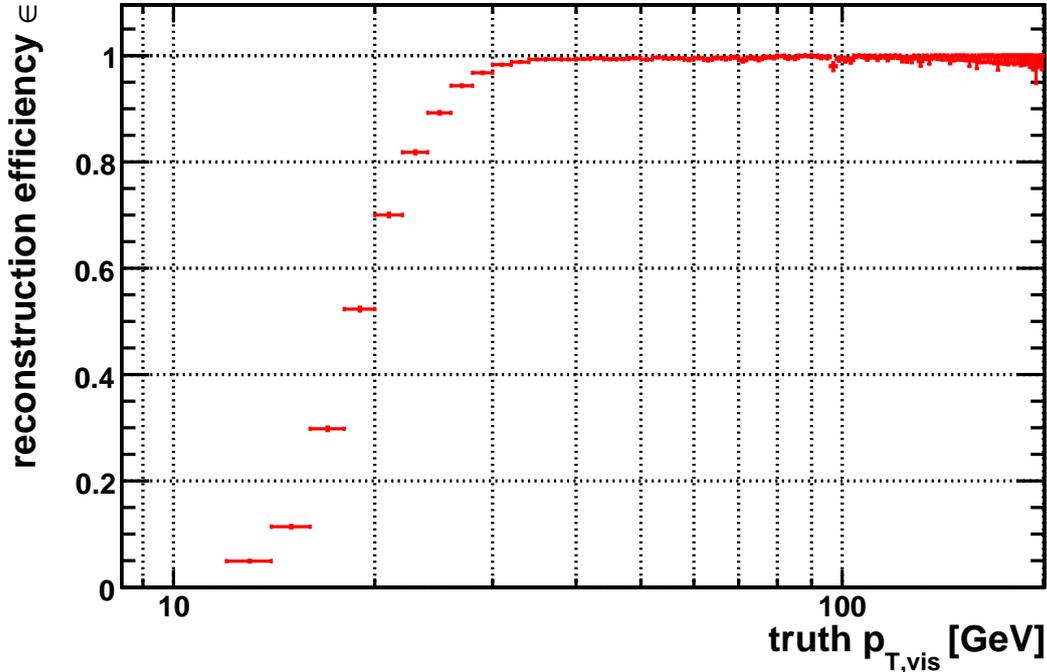


Figure 4.1: Reconstruction efficiency for  $\tau$  jets as a function of  $p_T$ . Note that the used sample has a minimum  $p_T$  cut of 12 GeV applied on truth level.

from QCD jets. Eight variables are constructed to be used for the identification procedure. These are described in the following.

##### - EM-Radius

The electromagnetic radius is exploring the smaller transverse shower profile of  $\tau$  jets compared to QCD jets with similar energy. It is defined as

$$R_{em} = \frac{\sum_i E_T^i \sqrt{(\eta_i - \eta_{cluster})^2 + (\phi_i - \phi_{cluster})^2}}{\sum_i E_T^i}. \quad (4.3)$$

The sums run over all cells in the electromagnetic calorimeter with  $\Delta R < 0.4$  with respect to the cluster center,  $\eta_i$  and  $\phi_i$  are the geometric centres of each cell.

The distributions of  $R_{em}$  for  $\tau$  jets and QCD jets is shown for several  $E_T$  ranges in Figure 4.2. This variable provides a good separation for low transverse energies. For higher energies the separation becomes less effective, as signal and background are both more collimated due to higher boosts.

##### - Calorimeter isolation fraction

Due to the strong collimation and low multiplicity of  $\tau$  jets they deposit nearly all their

#### 4 Reconstruction and Identification of $\tau$ Leptons

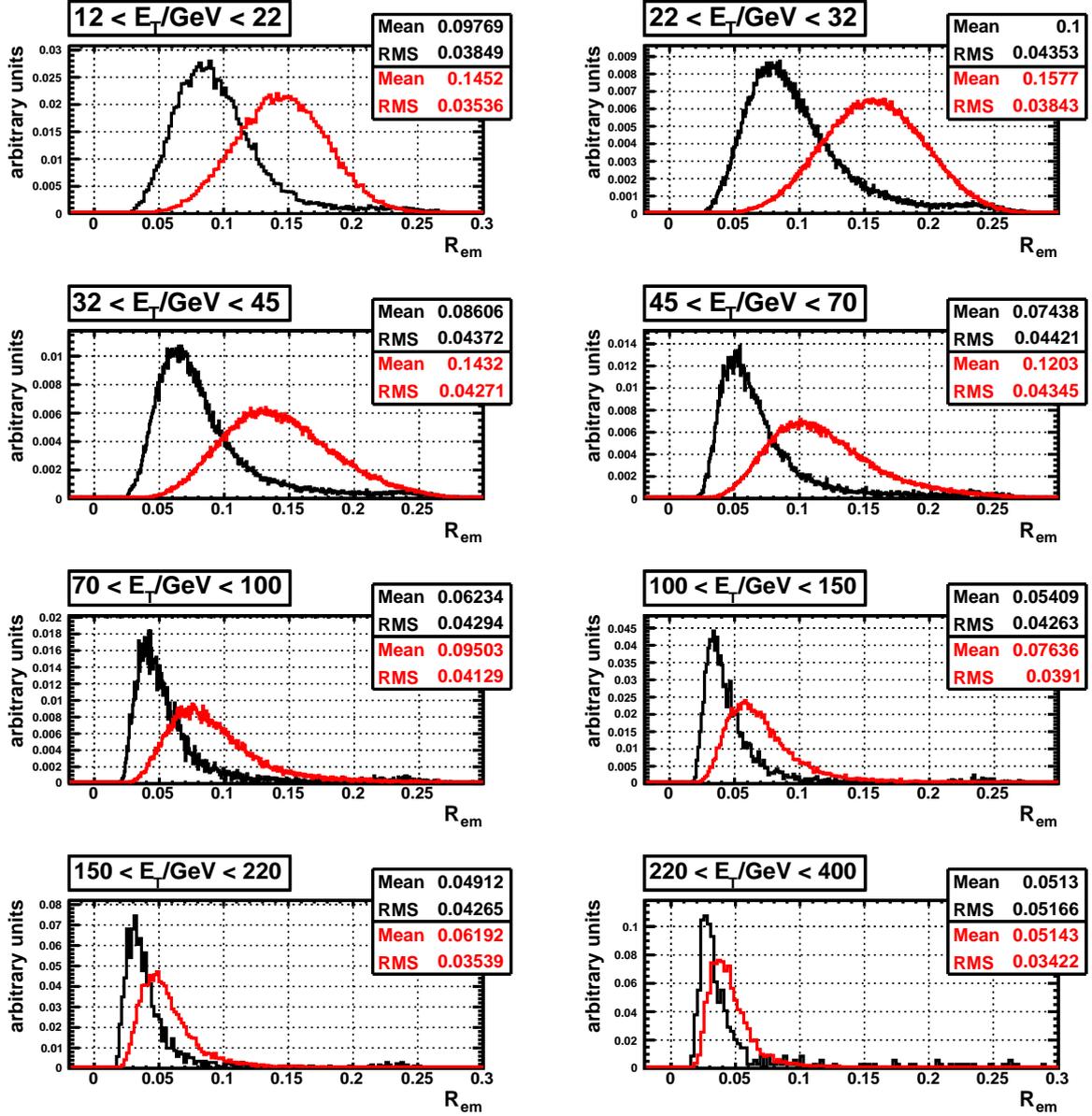


Figure 4.2: EM-Radius for different  $E_T$  ranges of the  $\tau$  candidates after requiring 1 to 3 associated tracks. The black distributions are for  $\tau$  jets and the red ones are for QCD jets. The distributions are normalised to a unit area.

## 4 Reconstruction and Identification of $\tau$ Leptons

energy in a small region around the  $\tau$  jet axis. QCD jets on the other hand spread out their energy more evenly. The calorimeter isolation fraction is designed to capture this difference. It is calculated as the ratio between the transverse energy in a ring of  $0.1 < \Delta R < 0.2$  and the total transverse energy in the electromagnetic calorimeter

$$\Delta E_T^{12} = \frac{\sum_j E_T^j}{\sum_i E_T^i}, \quad (4.4)$$

where  $j$  runs over all cells with  $0.1 < \Delta R < 0.2$  in the EM calorimeter, and  $i$  runs over all cells with  $\Delta R < 0.4$ .

Figure 4.3 shows the the distributions of the isolation fraction for  $\tau$  jets and QCD jets in bins of  $E_T$ .  $\tau$  jets tend to exhibit smaller values for this variable. As for the EM-Radius the distributions get narrower with higher boosts and the separation diminishes. Also, the performance of this variable depends on the event type. In events with a higher level of activity like  $t\bar{t}$  production, it is expected to provide less separation.

### - Energy width in the $\eta$ -strip layer

To make use of the lower number of  $\pi^0$  mesons in  $\tau$  jets and their subsequent higher boosts than in comparable QCD jets, the width in the  $\eta$ -strip layer ( $\eta$ -strip width) is calculated. It is defined by

$$\Delta\eta_{strip} = \frac{\sum_i (\eta_i)^2 E_T^i}{\sum_i E_T^i} - \frac{(\sum_i \eta_i E_T^i)^2}{(\sum_i E_T^i)^2} \quad (4.5)$$

The sums are to be taken over all associated cells belonging to the first layer of the electromagnetic calorimeter.

The distributions for this variable are shown in Figure 4.4. Again this variable shows a good separation for low  $E_T$ , but loses separation power with higher boosts.

### - Number of hits in the $\eta$ -strip layer

A similar criterion is the number of hits in the  $\eta$ -strip layer. It is defined by the number of cells in the first layer of the electromagnetic calorimeter with  $\Delta R < 0.4$  and containing an energy of at least 200 MeV.

The distributions for this variable are shown in Figure 4.5. A significant fraction of  $\tau$  jets have no hit in the  $\eta$ -strip layer, owing mostly to the single pion decay mode of the  $\tau$ . This happens less frequently for QCD jets. For high  $E_T$ , QCD jets tend to produce more hits than  $\tau$  jets.

### - Number of associated tracks

The naive expectation from the  $\tau$  decay modes, would be to see mostly one or three associated tracks for each  $\tau$  jet. However, a significant fraction of  $\tau$  jets have two associated tracks, and a smaller fraction has zero or more than three tracks.

$\tau$  jets with no associated tracks originate mainly from 1-prong decays, where the pion track is not reconstructed due to inefficiencies in the inner detector or where it does not have the required transverse momentum of at least 2 GeV. In addition, hadronic interactions

#### 4 Reconstruction and Identification of $\tau$ Leptons

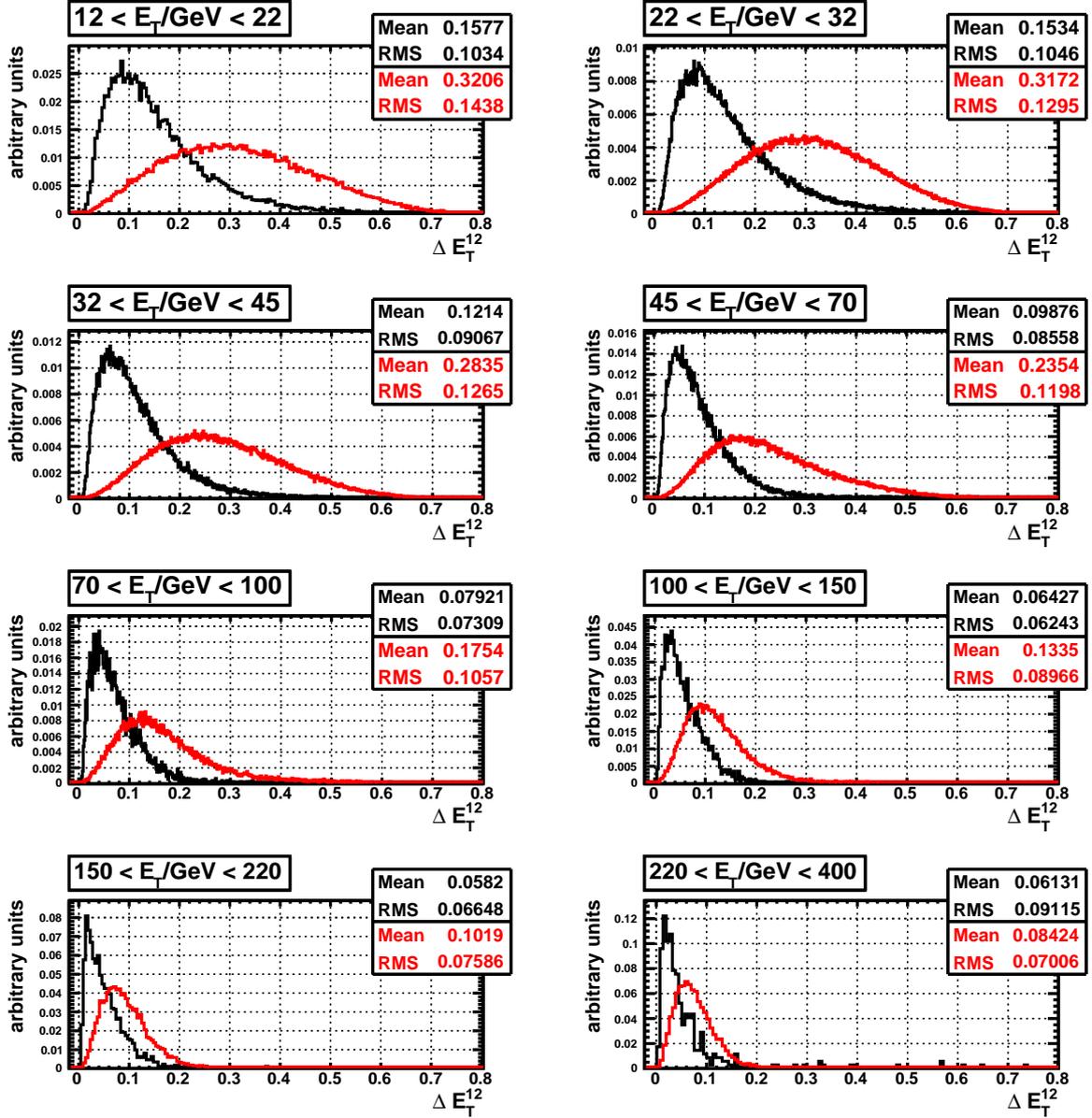


Figure 4.3: Calorimeter isolation for different  $E_T$  ranges of the  $\tau$  candidates after requiring 1 to 3 associated tracks. The black distributions are for  $\tau$  jets and the red ones are for QCD jets. The distributions are normalised to a unit area.

#### 4 Reconstruction and Identification of $\tau$ Leptons

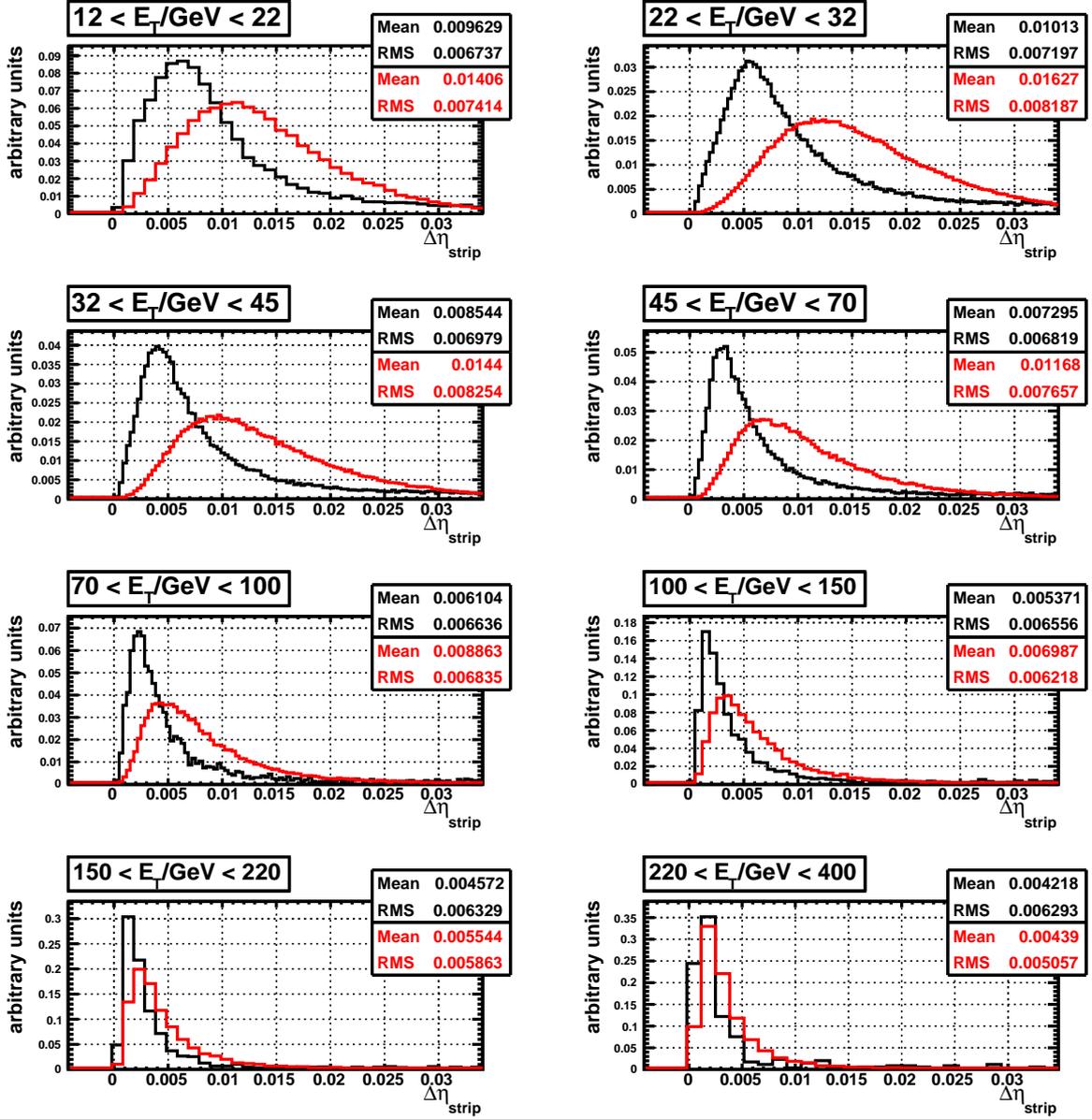


Figure 4.4:  $\eta$ -strip width for different  $E_T$  ranges of the  $\tau$  candidates after requiring 1 to 3 associated tracks. The black distributions are for  $\tau$  jets and the red ones are for QCD jets. The distributions are normalised to a unit area.

#### 4 Reconstruction and Identification of $\tau$ Leptons

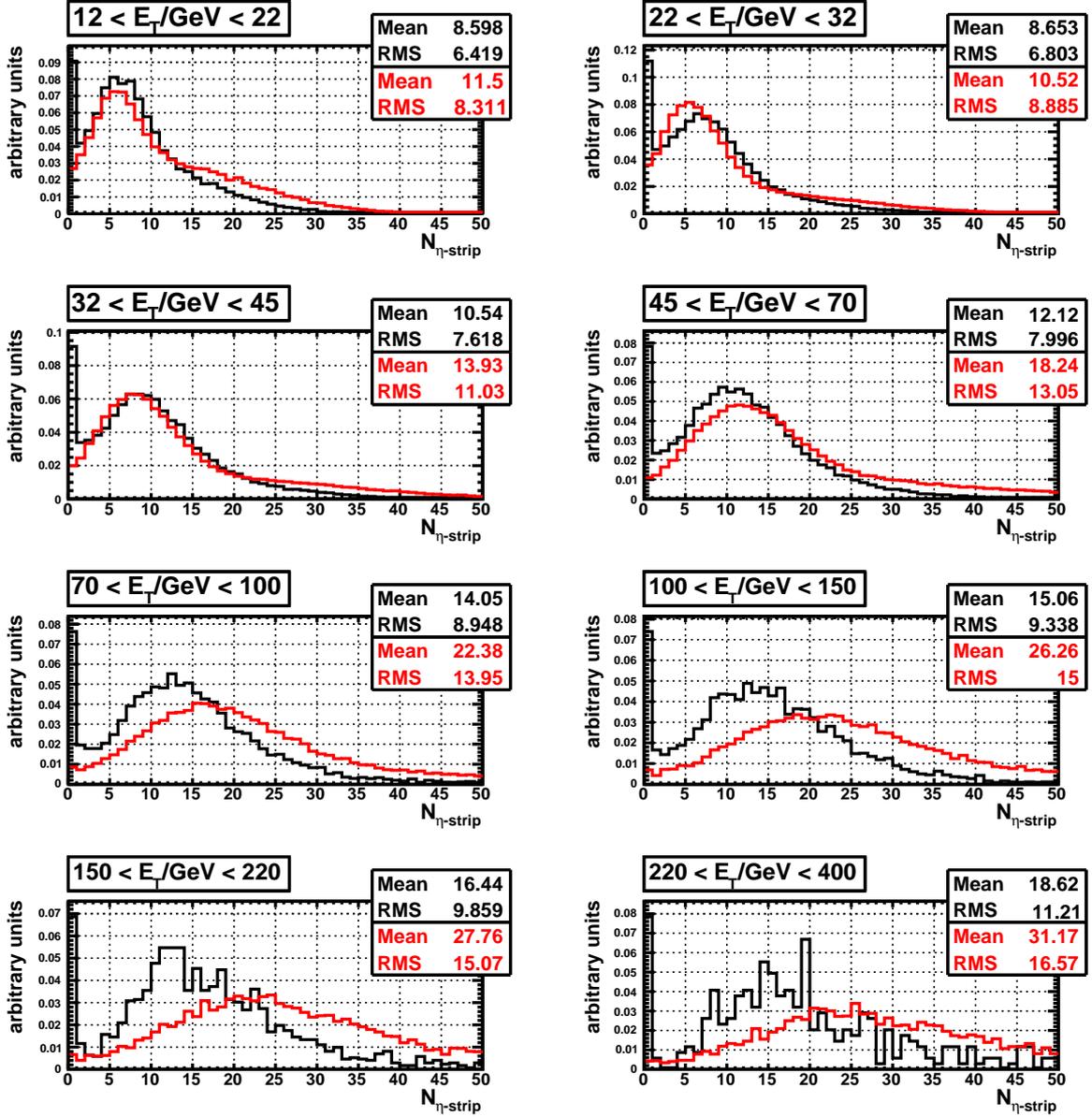


Figure 4.5: Number of hits in the  $\eta$ -strip layer after requiring 1 to 3 associated tracks. The black distributions are for  $\tau$  jets and the red ones are for QCD jets. The distributions are normalised to a unit area.

## 4 Reconstruction and Identification of $\tau$ Leptons

of the pion early in the detector can lead to a number of low  $p_T$  tracks, which are not reconstructed.

Several effects can lead to candidates with two associated tracks. For 3-prong decays, one track can be lost for similar reasons as discussed for 1-prong decays. In one prong decay modes containing  $\pi^0$  mesons, additional tracks can arise if one of the photons from the  $\pi^0$  decay has converted in the inner detector and one of the resulting electron/positron tracks is not reconstructed or does not pass the selection criteria. Additional tracks can also originate from multiple interactions and pile up events or can be picked up from nearby objects belonging to the same final state as the  $\tau$  lepton.

As a first step in the identification, a preselection of the  $\tau$  candidates is done, requiring one to three associated tracks. About 90% of real  $\tau$  jets pass this preselection, as shown in Figure 4.6 as a function of the transverse momentum of the true  $\tau$  jets. The efficiency of the preselection for selecting QCD jets is shown in Figure 4.7 separately for jets initiated by gluons (gluon jets), for jets initiated by b quarks (b jets) and for jets initiated by other quarks (quark jets).<sup>5</sup> One can see that less gluon jets and b jets pass the preselection for comparable jet energies than light quark jets. The reason for this are differences in the composition of the different types of jets. Gluon jets and b jets consist of more particles than quark jets for different reasons. Gluons have twice the colour charge of quarks and therefore have a higher probability to split up into two gluons, which leads to a higher multiplicity in gluon jets. In b jets the decay of the heavy B-hadron produces on average about five charged particles [3], so that b jets also show a higher track multiplicity.

The distributions for the number of tracks after applying the preselection are shown in Figure 4.8. Apart from some threshold effects for the lowest  $E_T$  bins, the dependence on  $E_T$  is not very strong for candidates passing the preselection.

### - Absolute Charge

The charge of the  $\tau$  candidate is defined by the sum of the charge of all associated tracks. The absolute value is used for identification purposes rather than the signed one. The sign of the charge is however useful for certain analyses where the final state contains a  $\tau^- \tau^+$  pair. Events which contain two  $\tau$  leptons with the same sign can be used to estimate the background contribution in the selected sample.

For  $\tau$  jets a charge of one could be always expected if the number of associated tracks is odd. But about 3% of the  $\tau$  jets with three associated tracks have a charge of three. The main source of a wrongly assigned charge are electron tracks from converted photons.

The distribution of the absolute charge is shown in Figure 4.9. A larger fraction of QCD jets are reconstructed with a charge of zero or three. The charge is strongly correlated with the number of associated tracks.

### - Lifetime signed transverse impact parameter significance

A variable designed to exploit the lifetime of the  $\tau$  lepton is the ‘lifetime signed transverse

---

<sup>5</sup>The particle that initiated a jet is defined as the particle with the highest transverse momentum inside a cone of  $\Delta R < 0.4$  around the  $\tau$  candidate direction.

#### 4 Reconstruction and Identification of $\tau$ Leptons

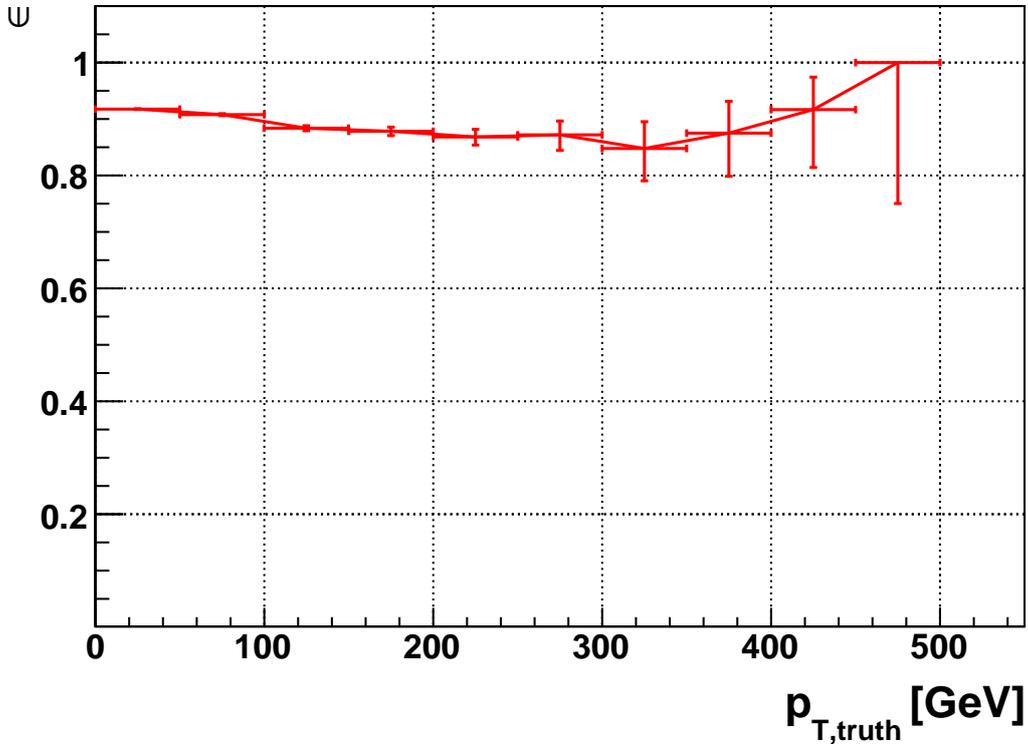


Figure 4.6: Efficiency of requiring 1 to 3 associated tracks for  $\tau$  jets as a function of the true  $p_T$  of the visible tau decay products.

impact parameter significance’. For conciseness ‘signed impact parameter’ will be used synonymously where there is no danger of ambiguity.

The signed impact parameter is based on the 2-dimensional impact parameter in the transverse plane with respect to the beam direction. The transverse impact parameter of a track is the minimum distance of the track to the nominal beam line when projected onto the  $x$ - $y$  plane. The geometric quantities are clarified in Figure 4.10. Tracks originating from the decay of particles with a lifetime will on average have a non vanishing impact parameter.

Using the direction of the  $\tau$  candidate reconstructed from the calorimeter, one can assign a sign to each track. The direction of the  $\tau$  candidate is assumed to reproduce the direction of the  $\tau$  lepton. This approximation is valid for boosted  $\tau$  leptons, where the neutrino is emitted almost collinear with the  $\tau$  flight direction. A positive sign is given to tracks that intersect the axis defined by the  $\tau$  direction in ‘front of’ the primary vertex, a negative sign is given to tracks that intersect the axis ‘behind’ the primary vertex. This is illustrated in Figure 4.10. For particles coming from the primary vertex, the assigned lifetime sign is random due to the limited impact parameter resolution. For particles coming from a real

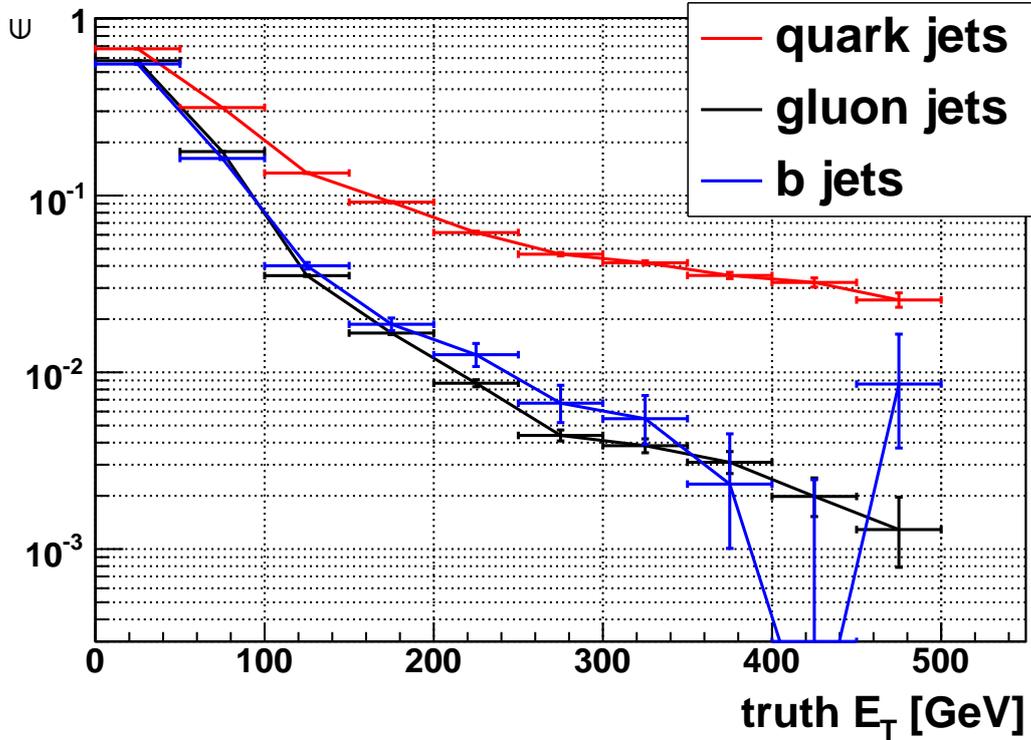


Figure 4.7: Efficiency of requiring 1 to 3 associated tracks for QCD jets as a function of the true  $E_T$ . The efficiencies are shown separately for different types of jets.

$\tau$  decay, negative signs can be assigned due to a badly reconstructed impact parameter or due to a badly reconstructed flight axis of the  $\tau$  lepton.

Since the resolution of the impact parameter measurement varies strongly with the track momentum and the quality of the track reconstruction, the significance of the impact parameter is used, that is the impact parameter divided by its error. The error can be obtained from the track fitting procedure. The final definition of the lifetime signed transverse impact parameter significance is

$$\sigma_{IP} = d_0/\sigma_{d_0} \cdot \text{sign}(\sin(\phi_{jet} - \phi_{tr})). \quad (4.6)$$

Here  $d_0$  is the impact parameter with a sign according to the direction of the angular momentum of the track with respect to the primary vertex, and  $\sigma_{d_0}$  its error.  $\phi_{jet}$  is the azimuthal angle of the  $\tau$  candidate and  $\phi_{tr}$  is the azimuthal angle of the track at the point of closest approach to the primary vertex in the transverse plane. The distributions of the signed impact parameter significance for  $\tau$  jets and QCD jets are shown in Figure 4.11.

The average distance the  $\tau$  lepton travels before it decays is proportional to the boost  $\gamma$ , while at the same time the opening angle of the decay products is proportional to  $1/\gamma$ , so the average impact parameter is approximately independent of the boost. A bias towards

#### 4 Reconstruction and Identification of $\tau$ Leptons

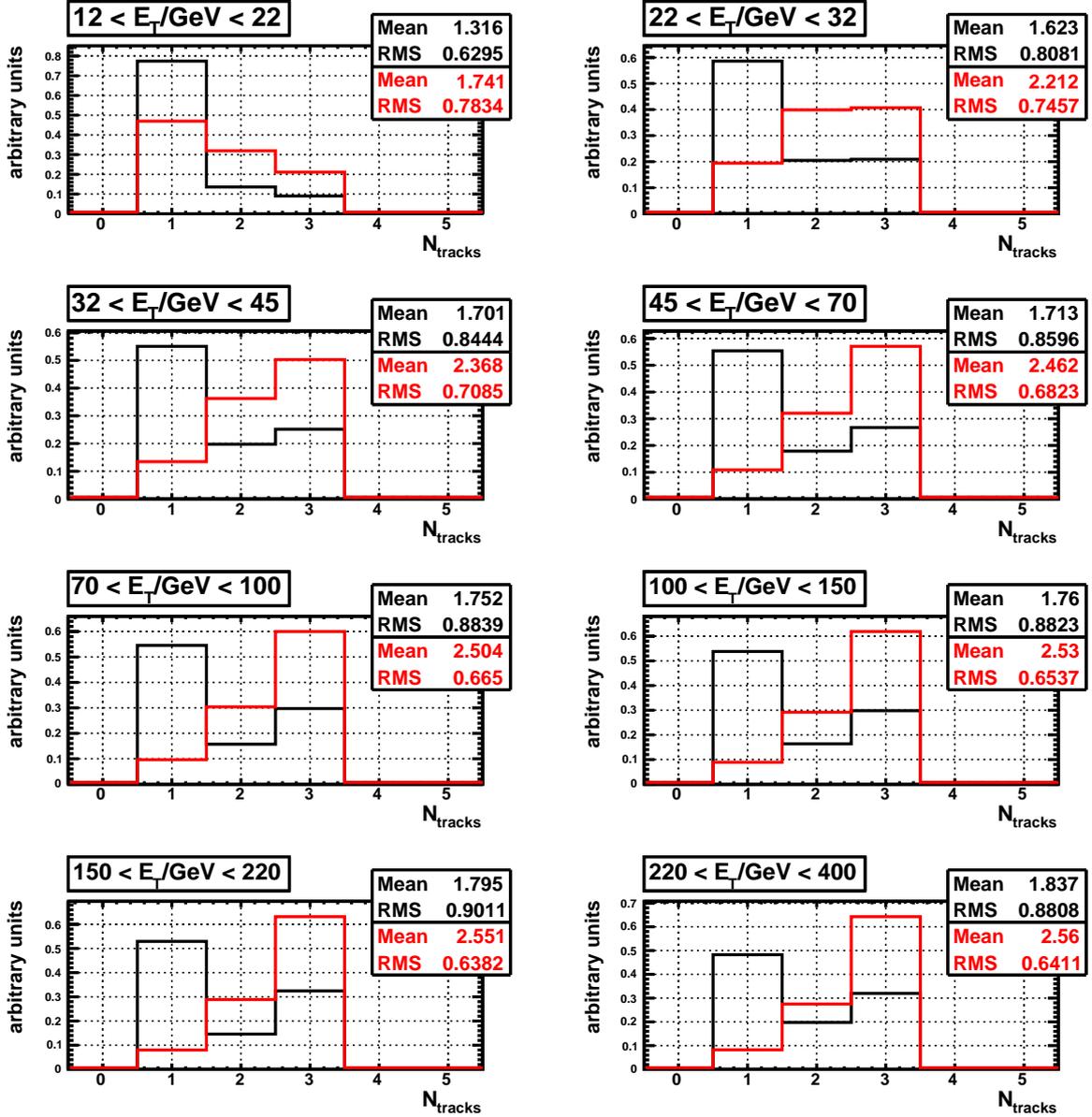


Figure 4.8: Number of associated tracks in the range 1 to 3. The black distributions are for  $\tau$  jets and the red ones are for QCD jets. All distributions are normalised to a unit area. See page 46 for a detailed discussion.

#### 4 Reconstruction and Identification of $\tau$ Leptons

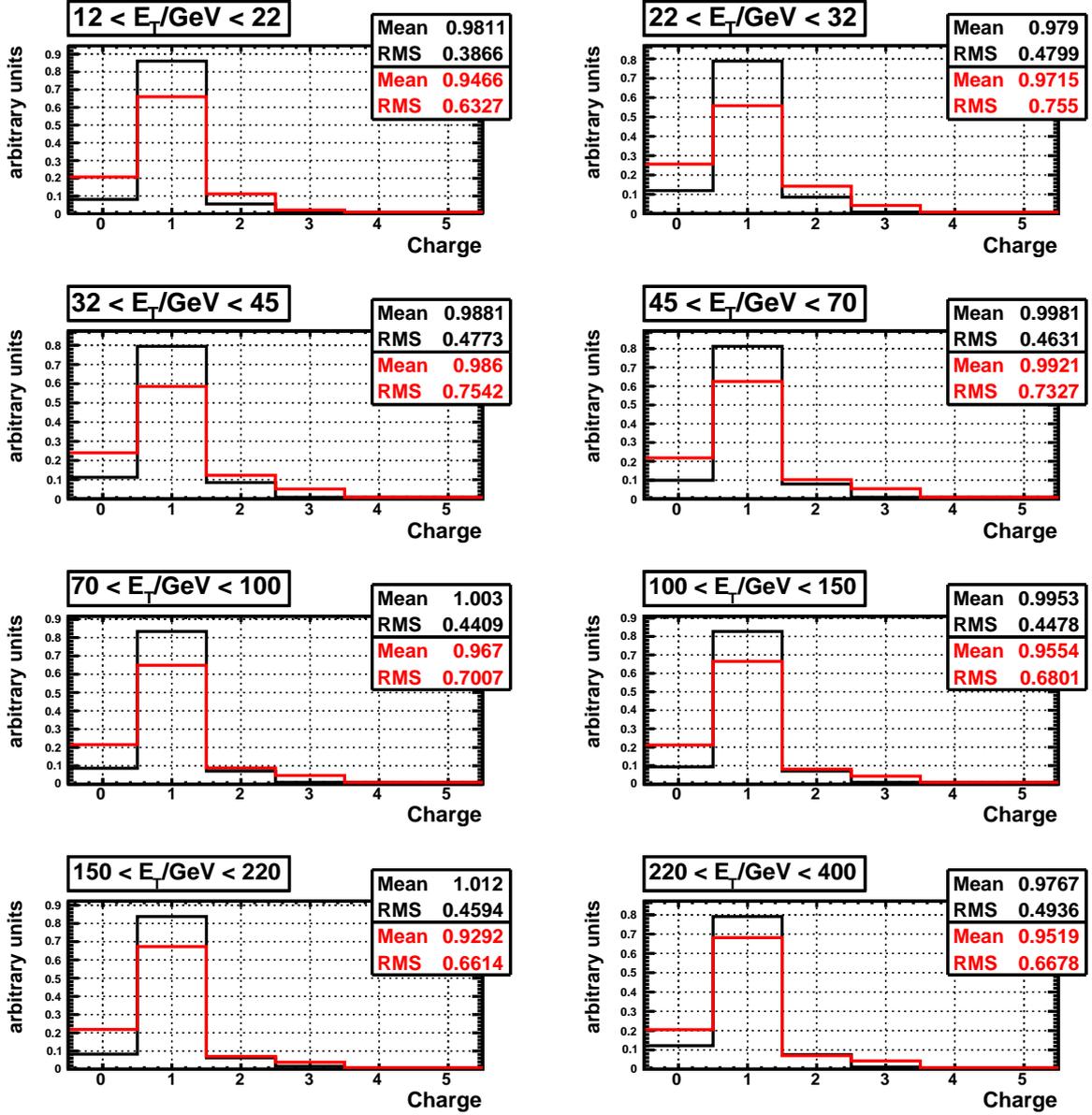


Figure 4.9: Absolute charge of the  $\tau$  candidate after requiring 1 to 3 associated tracks. The black distributions are for  $\tau$  jets and the red ones are for QCD jets. All distributions are normalised to a unit area.

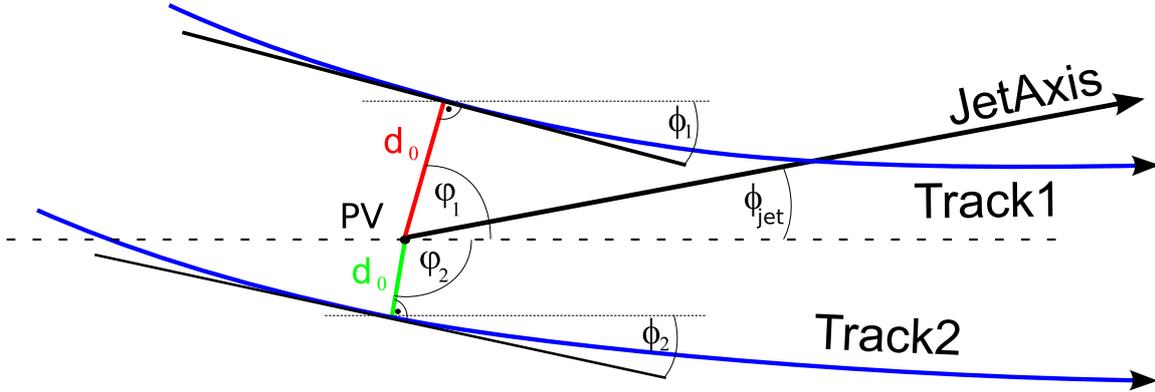


Figure 4.10: The lifetime signed transverse impact parameter significance of a track is defined in a way to be positive if the decay from which it originated happened in flight direction of the reference jet. An example for this is Track1 above. Track2 could not have originated from the decay of a particle within the jet. The sign of the impact parameter is defined as positive, if for the azimuthal angle of the point of closest approach to the primary vertex  $\varphi$  and the azimuthal angle of the tangent to the track at this point  $\phi$  the relation  $\varphi - \phi = (\pi/2) + n \cdot 2\pi$  holds, where  $n$  is any integer. Otherwise it is negative. Here  $d_0$  is positive and  $d_0$  is negative. Both tracks have a similar  $\phi$ , so  $\text{sign}(\sin(\phi_{jet} - \phi_{tr}))$  has also the same sign for both tracks. Hence, the resulting  $\sigma_{IP}$  is positive for Track1 and negative for Track2.

a positive  $\sigma_{IP}$  is expected for  $\tau$  jets, which can be seen in Figure 4.11. Since the impact parameter resolution improves with the track  $p_T$ , a slightly increased excess towards positive values can be observed for  $\tau$  jets with higher transverse energies.

#### - $E_T$ over $p_T$ of the leading track

On average the most energetic particle of a  $\tau$  jet carries around 50% of the total visible energy, and 100% for the single pion decay mode. In QCD jets, where the energy is more evenly distributed between the decay products, this fraction tends to be lower. This becomes more pronounced for higher jet energies. The ratio of the calibrated reconstructed  $E_T$  of the  $\tau$  candidate to the  $p_T$  of the leading track aims to exploit this difference.<sup>6</sup>

The distributions for this variable for  $\tau$  jets and QCD jets are shown in Figure 4.12. One can see a peak around one for  $\tau$  jets, compatible with the single pion decay mode. For high  $E_T$  bins, where the average charged multiplicity of QCD jets is high, the preselection cut introduces a bias towards jets with a generally low particle multiplicity, leading to lower values of the  $E_T/p_T$  variable.

### 4.3.3 Identification

To distinguish real  $\tau$  jets from QCD jets, the eight variables described above are combined with a likelihood method. The PDFs (see Section 3.2) are created separately for eleven

<sup>6</sup>The leading track is defined as the track with the highest transverse momentum of the associated tracks.

#### 4 Reconstruction and Identification of $\tau$ Leptons

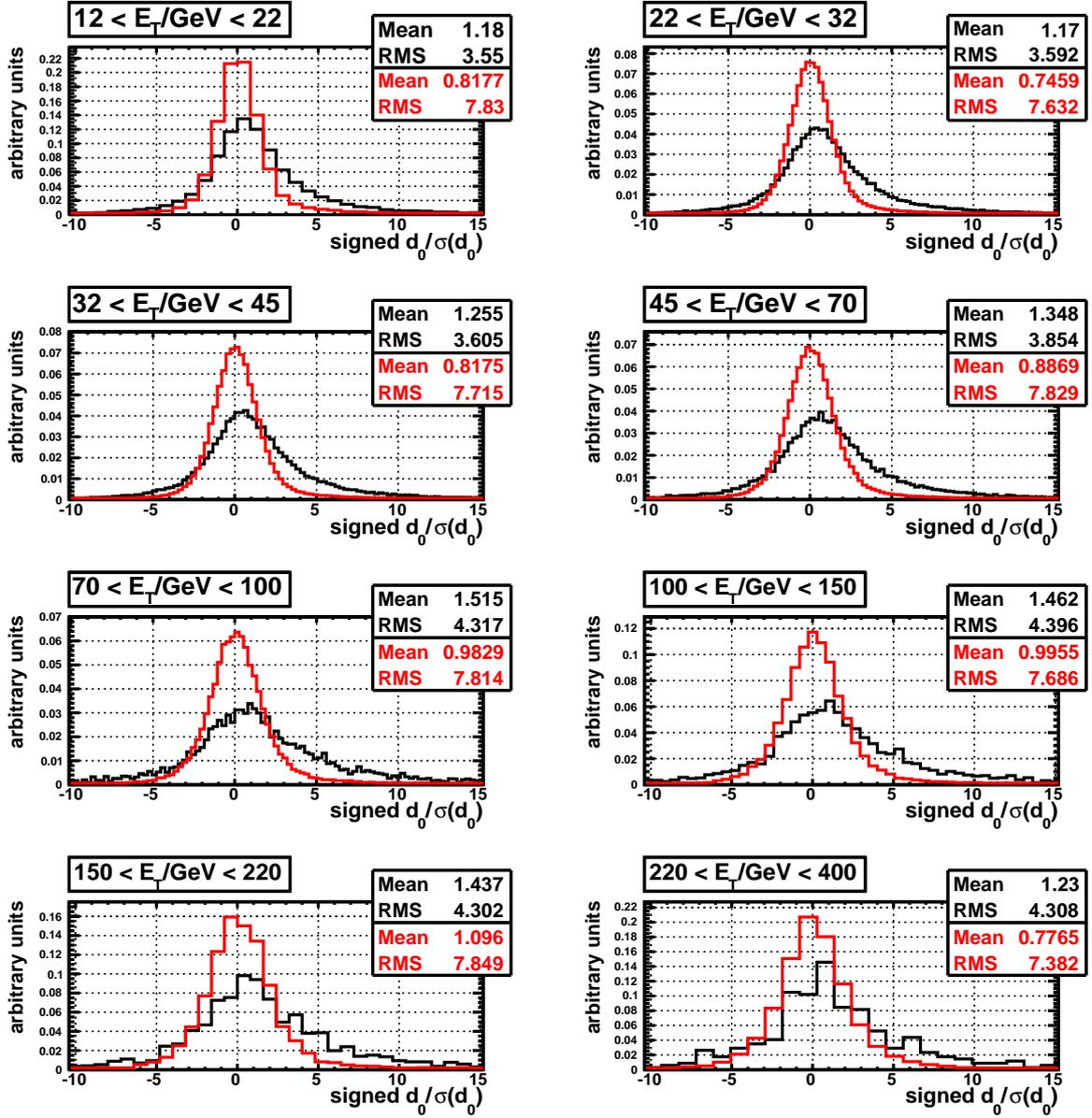


Figure 4.11: Lifetime signed transverse impact parameter significance of the  $\tau$  candidate after requiring 1 to 3 associated tracks. The black distributions are for  $\tau$  jets and the red ones are for QCD jets. All distributions are normalised to a unit area.

#### 4 Reconstruction and Identification of $\tau$ Leptons

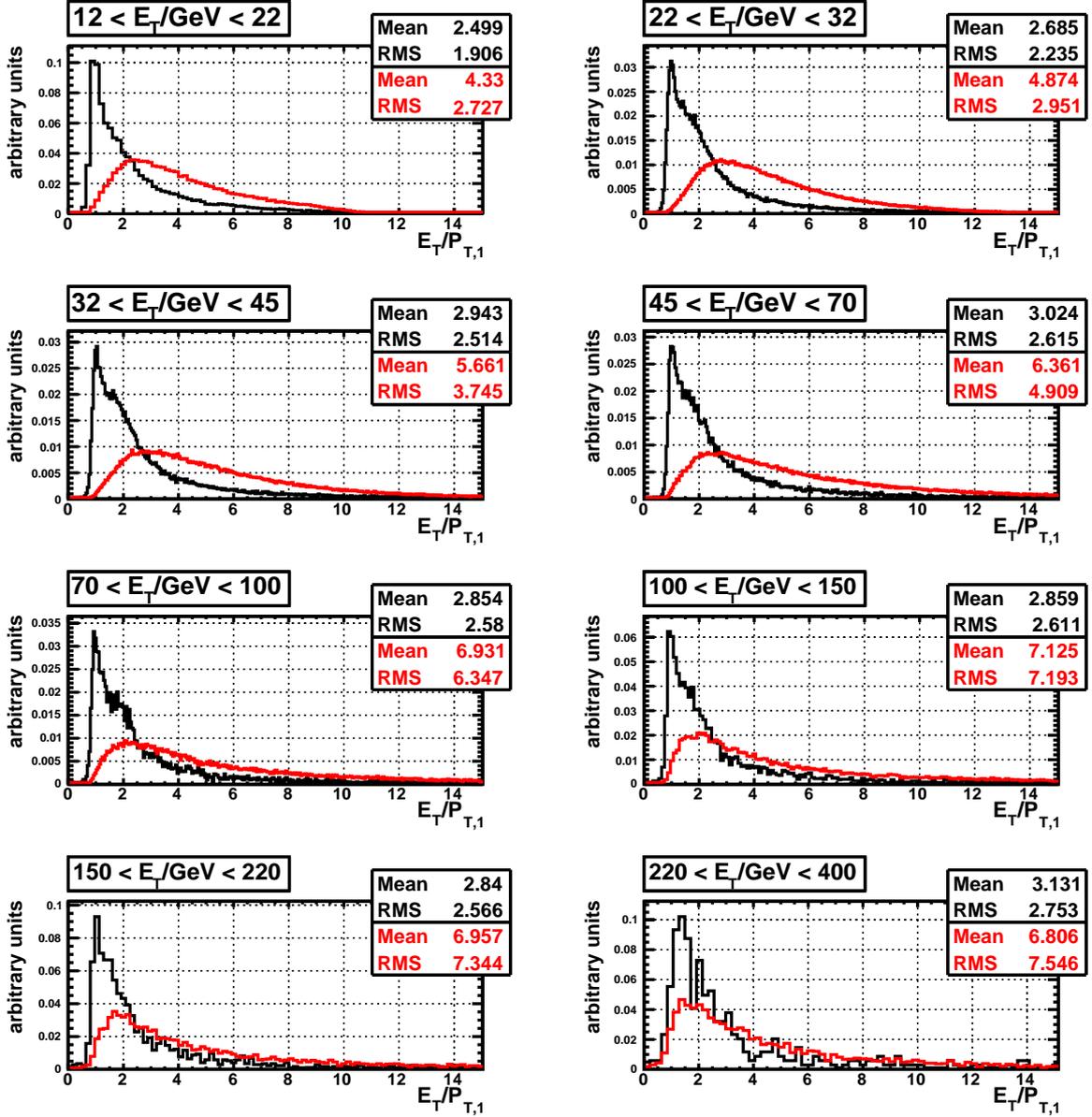


Figure 4.12:  $E_T$  over  $p_T$  of the leading track of the  $\tau$  candidate after requiring 1 to 3 associated tracks. The black distributions are for  $\tau$  jets and the red ones are for QCD jets. All distributions are normalised to a unit area.

#### 4 Reconstruction and Identification of $\tau$ Leptons

different regions of the transverse energy of the  $\tau$  candidate to approximately take into account the correlations of the identification variables with the transverse energy of the  $\tau$  candidate. The different  $E_T$  bin boundaries as given in [12] are 15, 28, 44, 62, 88, 134, 218, 334, 434, 600 and  $>600$  GeV. Of the eight identification variables three are discrete and five are continuous.

- The discrete variables are:
  - $N_{Tr}$  Number of associated tracks
  - $N_{\eta\text{-hits}}$  Number of hits in the  $\eta$ -strip layer
  - **Absolute Charge** of the  $\tau$  candidate
- The continuous variables are:
  - $R_{em}$  Electromagnetic radius
  - $\Delta E_T^{12}$  Calorimeter Isolation fraction
  - $\Delta\eta_{strip}$  Energy width in the  $\eta$ -strip layer
  - $d_0/\sigma_{d_0}$  Lifetime signed impact parameter significance
  - $E_T/p_{T,1}$  Ratio of transverse energy to leading track transverse momentum

For the discrete variables the normalised histograms of the distributions are directly used as PDFs. The PDFs of the continuous variables were obtained by fitting appropriate polynomials to finely binned histograms of the distributions. It has to be remarked, that the PDFs have been obtained from older, so called ‘rome’ data sets. The detector description used to simulate the samples from which the results in this section were produced was more realistic with regard to the amount of passive material in the detector and the simulation of noise, than the one used in [12]. This leads to a diminished rejection of QCD jets for the same identification efficiency for  $\tau$  jets of about a factor of two.

Although the likelihood method discards all correlations between the identification variables, the method nevertheless provides good separation power between  $\tau$  jets and QCD jets. Typical distributions for the likelihood discriminant are shown in Figure 4.13 for  $\tau$  jets in black and QCD jets in red. The  $\tau$  jets are obtained from  $Z \rightarrow \tau\tau$  and  $A \rightarrow \tau\tau$  events, and the QCD jets come from QCD dijets samples. The A boson is the CP-odd Higgs boson, arising in supersymmetric models with two Higgs doublets, and constitutes a source for  $\tau$  leptons with large transverse momenta. The mass of the A boson is varied between 150 GeV and 800 GeV by choosing the model parameters appropriately. It should be noted, that events in which a heavy particle like the A boson is produced exhibit in general a higher amount of activity than for example  $Z \rightarrow \tau\tau$  events, due to the larger momentum transfer in the partonic interaction. This can have an influence on the identification performance of the  $\tau$  reconstruction algorithms. All samples were simulated and reconstructed with ATHENA release 12 as part of the central ATLAS Monte Carlo production.

The background rejections for the signal efficiencies that result from different cuts on the likelihood value are shown in Figure 4.14 separately for different  $p_T$  ranges. In the lowest

#### 4 Reconstruction and Identification of $\tau$ Leptons

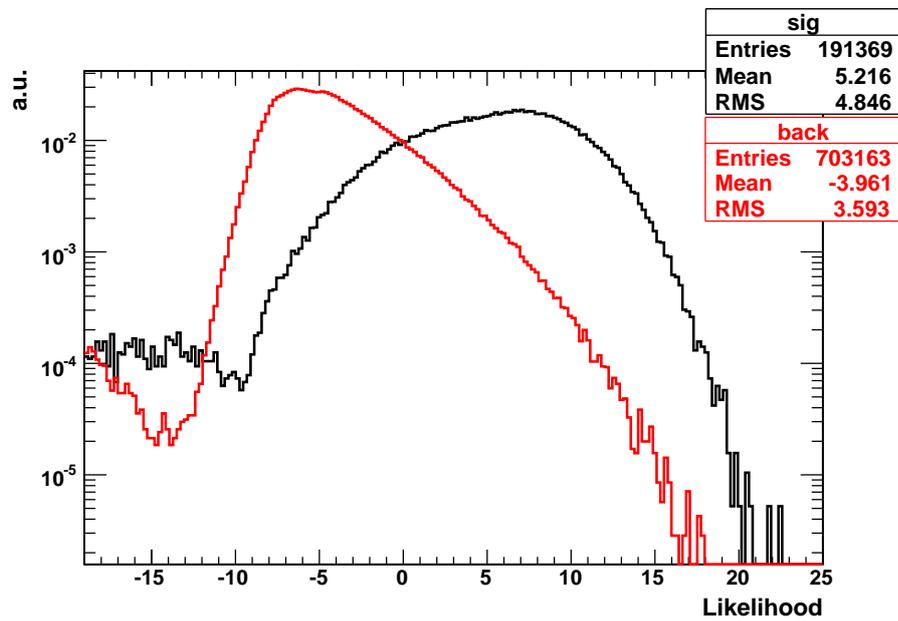


Figure 4.13: Distribution of the likelihood value for  $\tau$  jets from  $Z/A \rightarrow \tau\tau$  events in black and QCD jets from QCD dijet events in red after requiring 1 to 3 associated tracks. The rise at large negative likelihood values is due to a penalty of -10 for values outside the fitted region of the probability density function.

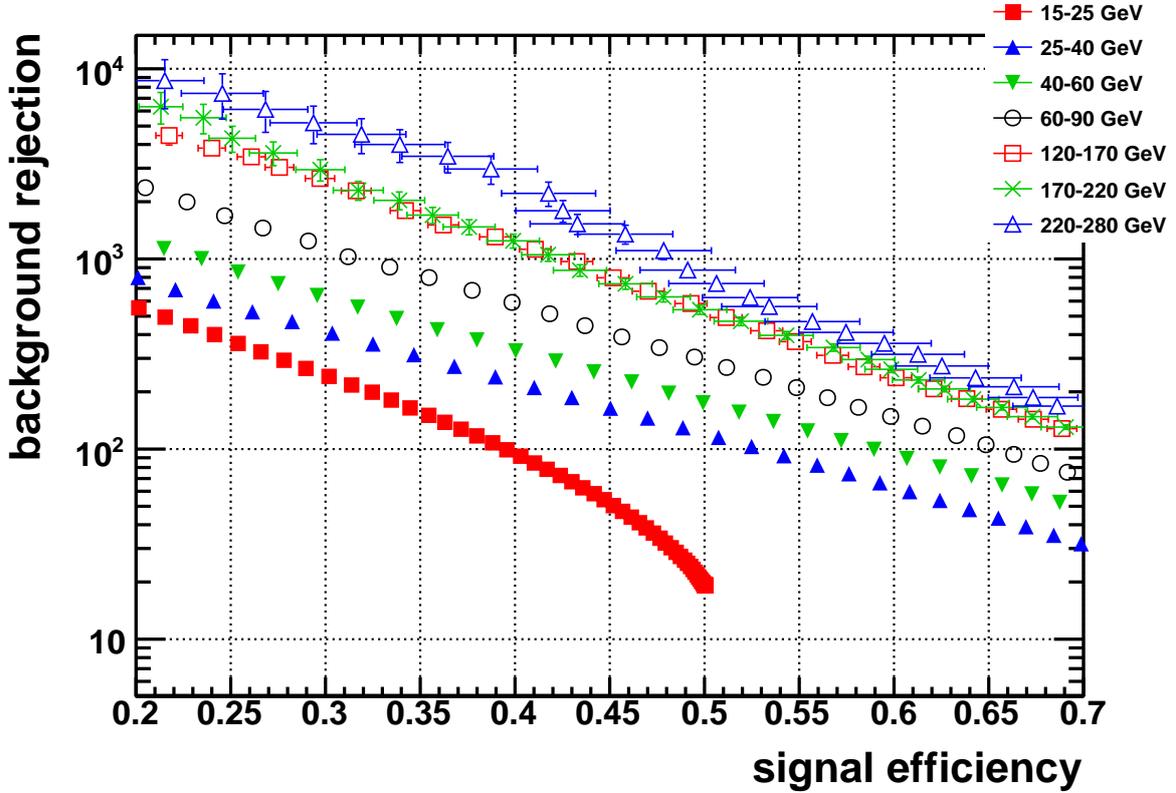


Figure 4.14: Rejection of QCD jets as a function of the efficiency to select true  $\tau$  jets. The rejections are shown for different bins of the true visible  $p_T$ .

$p_T$  bin shown, between 15 and 25 GeV, one can see the effect of the low reconstruction efficiency for soft  $\tau$  leptons.

In Figure 4.15 the rejection is shown as a function of the true transverse momentum of the  $\tau$  candidates for two typical signal efficiencies of 30 and 50 percent. The separation between  $\tau$  jets and QCD jets becomes better at higher transverse momenta, owing in part to the growing track multiplicity of the QCD jets, so less of them pass the preselection. The values of the rejection factors for a few selected points are shown in Table 4.2.

## 4.4 The tau1P3P Package

Tau1P3P, The second of the two  $\tau$  reconstruction algorithms more strongly focuses on information from the inner detector. The seeds for  $\tau$  candidates are tracks with a  $p_T$  above a certain configurable threshold, which fulfil some minimal requirements on the quality of the track reconstruction. Additional nearby tracks are then associated and the direction of the  $\tau$  candidate is defined by the  $p_T$  weighted barycentre of all associated tracks. Calorimeter cells in a cone around the  $\tau$  direction are collected, as in the tauRec

#### 4 Reconstruction and Identification of $\tau$ Leptons

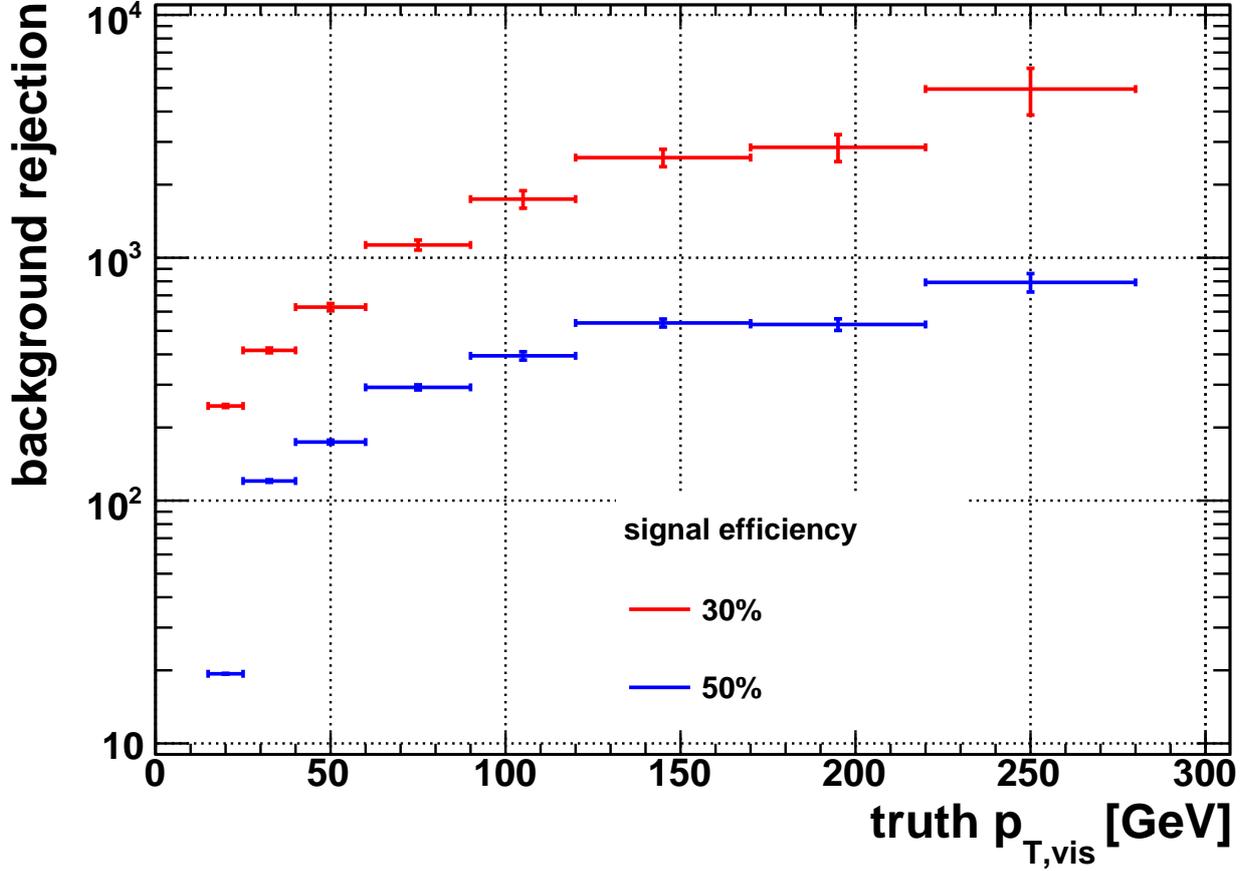


Figure 4.15: Rejection against QCD jets as a function of  $p_T$ . The rejections are shown for signal efficiencies 30% (red) and 50% (blue) for each  $p_T$  bin.

truth $p_{T,vis}$	signal efficiency	
	30%	50%
15 – 25 GeV	245.3 ± 3.5	19.35 ± 0.08
25 – 40 GeV	415.5 ± 9.4	120.4 ± 1.5
40 – 60 GeV	626 ± 22	174.2 ± 3.2
60 – 90 GeV	1128 ± 53	293 ± 7
90 – 120 GeV	1744 ± 145	395 ± 16
170 – 220 GeV	2850 ± 362	531 ± 29

Table 4.2: Rejection factors for some selected signal efficiencies and  $p_T$  ranges.

## 4 Reconstruction and Identification of $\tau$ Leptons

algorithm, and several identification variables are calculated from them as well as from the properties of the tracks.

To define the energy of the reconstructed  $\tau$  candidate, an *energy flow* approach is applied. The energy flow algorithm attempts to subtract the energy deposited by charged pions from the calorimeter, in order to replace it with the momentum of the corresponding tracks. For the typical pion transverse momenta in  $\tau$  jets the momentum resolution of the track reconstruction is about an order of magnitude better than the energy resolution of the hadronic calorimeter. The energy flow approach therefore has the potential to determine the energy of the  $\tau$  jets more precisely, especially for low energies. The energy measured in the hadronic calorimeter is assumed to arise only from charged pions, and is completely replaced by the track momenta. This property of the energy flow algorithm will be of importance in Section 6.3.3.

The tau1P3P algorithm offers several discriminant values, obtained from the variables calculated before by different analysis techniques, ranging from a simple cut based approach to a more sophisticated neural network analysis. Both the identification and the energy reconstruction have been mainly optimised for transverse energies of the  $\tau$  jets below 100 GeV.

### 4.4.1 Seeding

To be considered as a seed for a  $\tau$  candidate, a track is required to fulfil:

- $p_T$  greater than 6 GeV.
- transverse impact parameter less than 1 mm.
- at least eight hits in the silicon detectors and ten hits in the TRT.
- reduced  $\chi^2$  of the track fit smaller than 1.7.

The tracks that fulfil these requirements are ordered by  $p_T$ , starting with the highest transverse momentum. For each  $\tau$  candidate defined by these seed tracks other tracks are searched for around the direction of the seed track in a cone of  $\Delta R < 0.4$ . Those additional tracks are required to have a  $p_T$  greater than 1 GeV. A loose rejection against electrons from photon conversions is achieved by requiring less than 20% of the TRT hits to be high threshold hits and at least one hit in the innermost pixel layer. Of those tracks, the ones with  $\Delta R < 0.2$  are associated to the candidate as *core tracks*, and are used subsequently to calculate the direction and the charge of the  $\tau$  object. The tracks with  $0.2 < \Delta R < 0.4$  are counted as *isolation tracks*, and their number is used in the identification step later on. Candidates with up to seven additional tracks in the core region are reconstructed.

Finally all calorimeter cells in a cone of  $\Delta R < 0.4$  are collected. Only cells which pass a cut of  $|E| > 2\sigma_{noise}$  on the cell energy are used for the following steps, in order to suppress the contribution of noise.

#### 4.4.2 Identification Variables

The tau1P3P algorithm builds several identification variables from the calorimeter cells and associated tracks, some of which are quite similar to the variables calculated by the tauRec algorithm. Only a short overview of the different variables will be given here.

The purely calorimeter based variables are:

**- EM-Radius**

The definition of the electromagnetic radius is analogous to the tauRec variable of the same name, but only cells inside a cone of  $\Delta R < 0.2$  around the  $\tau$  jet direction are considered. This leads on average to a smaller value of the variable, especially for QCD jets, which will usually deposit energy outside the cone of radius 0.2.

**- Number of hits in the  $\eta$ -strip layer**

This variable is again analogous to the corresponding variable in the tauRec algorithm, also only considering cells within a cone of radius 0.2. This leads to a lower number of hits for both signal and background compared to the tauRec definition.

**- Width in the  $\eta$ -strip layer**

The width in the  $\eta$ -strip layer for the tau1P3P algorithm is defined in the same way as the corresponding variable for the tauRec algorithm, but only cells in the  $\eta$ -strip layer with  $\Delta R < 0.2$  are taken into account.

**- Annular isolation fraction**

The annular isolation fraction is defined as the ratio of the energy deposited in a ring with  $0.1 < \Delta R < 0.2$  to the total energy in a cone of  $\Delta R < 0.2$ . All calorimeter layers are taken into account.

**- Isolation ratio**

This variable is very similar to the annular isolation fraction. It is defined as the ratio of the energy in the *isolation region* ( $0.2 < \Delta R < 0.4$ ) to the energy in the *core region* ( $\Delta R < 0.2$ ).

**- Number of  $\pi^0$  clusters**

In  $\tau$  jets, significant electromagnetic energy depositions arise mainly from  $\pi^0$  mesons, which decay into two photons. If the clusters in the electromagnetic calorimeter produced by these pions can be reconstructed, the additional information can constrain other variables that are correlated with the  $\tau$  decay mode. At high energies the number of  $\pi^0$  mesons in QCD jets becomes much higher than for  $\tau$  jets, so the number of reconstructed electromagnetic clusters might also provide for additional discrimination at these energies.

The  $\pi^0$  clusters are reconstructed by the topological clustering algorithm, which will be described in Section 5.1. The clustering algorithm is configured to only consider cells in

## 4 Reconstruction and Identification of $\tau$ Leptons

the first two layers of the electromagnetic calorimeter and in the presampler and to use seed, neighbour and cell thresholds of 4, 2 and 0  $\sigma_{noise}$ , respectively.<sup>7</sup> To be considered as a  $\pi^0$  cluster, the electromagnetic clusters need to have a transverse energy of at least 1 GeV, at least 10% of the cluster energy must be deposited in the presampler and the  $\eta$ -strip layer and none of the associated tracks must be close by.

The number of  $\pi^0$  clusters is shown in Figure 4.16 for  $\tau$  candidates with a true transverse momentum between 12 GeV and 100 GeV. On the left the distributions for the most important 1-prong decay modes of  $\tau$  leptons are shown. On the right the distribution for QCD jets is shown.

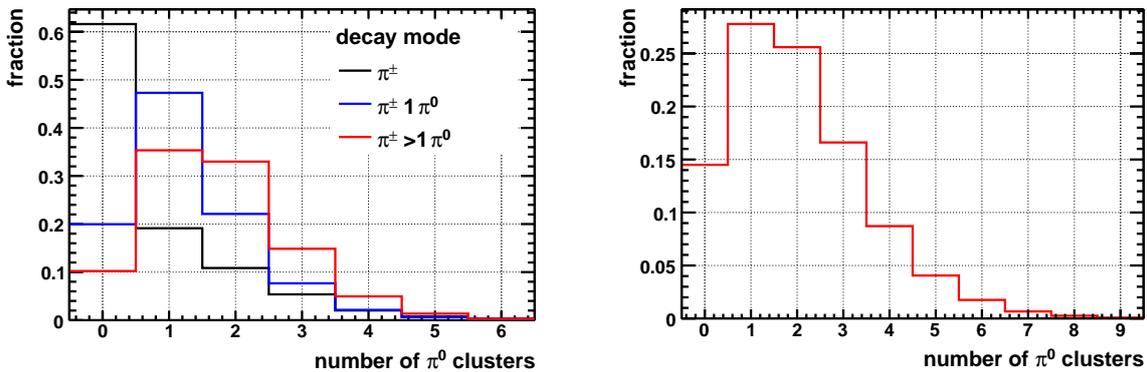


Figure 4.16: Number of  $\pi^0$  clusters for  $\tau$  candidates with a true visible transverse momentum between 12 GeV and 100 GeV. On the left the number of  $\pi^0$  clusters is shown for different decay modes of  $\tau$  jets. On the right number of  $\pi^0$  clusters is shown for QCD jets. The distributions are each normalised to one.

The variables calculated from the associated tracks are:

### - Number of Isolation Tracks

This is the number of isolation tracks in a region with  $0.2 < \Delta R < 0.4$  around the seed. This variable is quite powerful, since  $\tau$  jets are almost entirely contained within a cone of radius 0.2, and tracks in the isolation region should therefore only arise from conversions or nearby tracks from the underlying event. Consequently around 90% of the  $\tau$  jets show zero and about 5% show one isolation track. From the QCD jets that are reconstructed by tau1P3P with up to eight tracks within a cone of  $\Delta R < 0.2$ , about 50% have zero isolation tracks for very low transverse energies, settling at about 20% for high energies.

### - Track width

The width of the track system is only defined for  $\tau$  candidates with more than one associ-

<sup>7</sup>The meaning of the different thresholds will be explained in Section 5.1.

#### 4 Reconstruction and Identification of $\tau$ Leptons

ated track. The definition is, similar to the width in the  $\eta$ -strip layer, the variance of the pseudorapidity of the tracks weighted by the transverse momentum of the track

$$\Delta\eta_{tracks} = \frac{\sum_i (\eta_i)^2 p_T^i{}^2}{\sum_i p_T^i{}^2} - \frac{(\sum_i \eta_i p_T^i)^2}{(\sum_i p_T^i)^2}, \quad (4.7)$$

where  $i$  labels the associated tracks. The variable shows a good separation, especially at higher energies, where the  $\tau$  jets become much more collimated than comparable QCD jets.

##### - Invariant mass of the track system

Hadronic  $\tau$  decays proceed mainly through a small number of intermediate resonances with masses below the  $\tau$  mass. The invariant mass of the tracks associated to the  $\tau$  jet is therefore expected to show several peaks in the range up to 1 GeV, which are heavily smeared by the width of the intermediate particles and the detector resolution. For QCD jets a continuum of invariant masses can be expected, and the invariant mass is strongly correlated with the energy of the jet. This can indeed be observed. The invariant mass of the associated tracks provides some separation for very low transverse momenta and becomes more discriminating with growing momentum. The distribution of the invariant track mass is shown in Figure 4.17 for  $\tau$  candidates with a true transverse momentum between 12 GeV and 100 GeV.

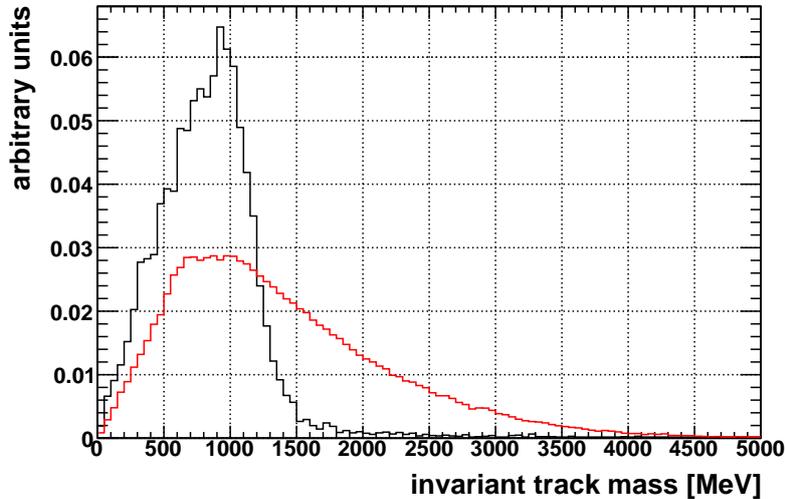


Figure 4.17: Invariant mass of the track system for  $\tau$  jets and QCD jets with a true transverse momentum between 12 GeV and 100 GeV and more than one associated track. The distribution for  $\tau$  jets is shown in black, the one for QCD jets in red.

##### - Transverse flight path significance

If a  $\tau$  candidate has more than one associated track, it is often possible to reconstruct a displaced vertex using those tracks. The significance of the distance between this vertex

## 4 Reconstruction and Identification of $\tau$ Leptons

and the primary vertex of the event in the transverse plane is calculated, and a sign is assigned in analogy to the lifetime signed impact parameter. The sign is positive if the distance vector between the primary vertex and the reconstructed vertex from the  $\tau$  tracks and the direction of the  $\tau$  candidate enclose an angle of less than  $90^\circ$  in the transverse plane, negative otherwise.

For  $\tau$  jets with higher transverse energies larger values of the flight path significance are observed on average, owing to the better resolution of the relevant track parameters for tracks with larger transverse momenta. This leads to a better discrimination against quark and gluon jets at larger transverse energies.

Although b jets contain long lived particles with a lifetime which is about five times larger than the lifetime of the  $\tau$  lepton, they tend to show similar distributions for the transverse flight path significance [27]. Therefore this variable does not provide a discrimination between  $\tau$  jets and b jets.

### - Transverse and longitudinal impact parameter significance

For  $\tau$  candidates with only one associated track, the unsigned transverse impact parameter significance is used as an identification variable. Since the direction of the  $\tau$  candidate is defined from the direction of the track, a lifetime sign cannot be defined and hence the unsigned impact parameter is used. The longitudinal impact parameter of the track is also used. The longitudinal impact parameter is the distance along the beam axis between the point of closest approach and the primary vertex.

To determine the transverse and longitudinal impact parameters of the associated track, the primary vertex is refitted without including this track in the fitting procedure. The point of closest approach is redefined by extrapolating the track to the point with the smallest transverse distance to the refitted primary vertex along the track. For calculating the significances the errors of the impact parameters as well as the error of the primary vertex position are taken into account. Higher mean values than for QCD jets are expected for  $\tau$  jets for both variables. They are not calculated for  $\tau$  candidates with more than one associated track.

Additionally some variables are calculated that combine calorimetric and tracking information:

### - Invariant mass between tracks and electromagnetic cluster

In addition to the invariant mass of the track system, one can also calculate the invariant mass of the track system and the electromagnetic component in the calorimeter. The EM component is taken from the cells in the first two layers of the EM calorimeter, removing the cells close to a track.

This variable shows a similar behaviour to the invariant mass of the track system, but can also be applied to  $\tau$  candidates with only one associated track.

### - Ratio of hadronic energy to transverse momentum of track system

## 4 Reconstruction and Identification of $\tau$ Leptons

The ratio of the uncalibrated transverse energy deposited in the hadronic calorimeter in a cone of  $\Delta R < 0.2$  to the sum of the transverse momenta of the associated tracks is also used as a variable. It shows only a very modest separation power. This is to be expected, as in both  $\tau$  jets and QCD jets most tracks arise from charged hadrons, which deposit most of their energy in the hadronic calorimeter.

### 4.4.3 Identification

For the identification step there are several methods implemented in the tau1P3P package. The method offering the best performance in terms of rejection of QCD jets is based on a neural network. The  $\tau$  candidates are assigned to four different categories, and separate networks have been trained for each category. The categories are:

- one associated track, no  $\pi^0$  cluster
- one associated tracks,  $\geq 1$   $\pi^0$  clusters
- two associated tracks
- more than two associated tracks

For each category a neural network with eleven input variables and two hidden layers with 30 nodes each is trained with hadronically decaying  $\tau$  leptons from  $Z^0 \rightarrow \tau\tau$  and  $W \rightarrow \tau\nu_\tau$  events and QCD jets from QCD dijet events. Nine of the input variables are the same for all categories. These are

- $E_T$  calculated from energy flow
- EM-Radius
- Isolation fraction
- Isolation ratio
- $\eta$ -strip width
- Hits in the  $\eta$ -strip layer
- Hadronic  $E_T$  over sum of  $p_T$
- Invariant mass between tracks and EM cluster
- Number of isolation tracks.

For the categories with one associated tracks the last two input variables are the transverse and the longitudinal impact parameter significance of the track. For the categories with more than one track these are the width of the track system and the transverse flight path significance.

## 5 Improvements for tauRec

As shown in Section 4.3.1, the tauRec algorithm seeded by the sliding window clusters shows a very low reconstruction efficiency for  $\tau$  jets with small transverse momenta below 20 GeV. Since this energy region is interesting for different physics scenarios, it was desirable to improve the reconstruction efficiency for soft  $\tau$  leptons. Also the sliding window clusters in the configuration used for the  $\tau$  reconstruction were no longer actively maintained, which provided additional motivation to switch to a different type of seeding.

There are two clustering algorithms implemented in the ATLAS reconstruction software. One is the sliding window algorithm that was used for seeding the existing version of the tauRec package. The other clustering algorithm is the *topological clustering* (TopoCluster). In a preliminary study it was shown that the topological clustering can provide a high reconstruction efficiency down to transverse momenta of about 10 GeV.

The topological clustering, in addition to the higher reconstruction efficiency, has the beneficial feature of a better noise suppression than the simple fixed cut on the significance of the cell energy used so far. This feature promises to make improvements to the identification possible, especially in the very low energy region, where the contribution of calorimeter noise on the shower shape variables is more prominent.

In the following a short overview over the TopoCluster algorithm is given and the integration of the TopoClusters in the tauRec package is explained. Using a large data sample reconstructed with this modified version of tauRec, the probability density functions for the likelihood discrimination were updated to make full use of the good noise suppression capabilities of the topological clustering algorithm. At the end of the chapter the performance of this updated identification procedure is discussed.

### 5.1 Topological Clustering

The topological clustering algorithm exploits the neighbourhood relations between calorimeter cells and the significances of their energy measurements, to reconstruct calorimeter clusters consisting of a variable number of cells and showing low sensitivity to noise. The topological clustering algorithm makes it possible to use a relatively low threshold on the cluster energy while keeping the number of fake clusters low.

The algorithm makes the assumption that the noise of the energy measurement is uncorrelated between nearby calorimeter cells. On the other hand, the real deposited energy from a particle obviously is correlated with the distance from the shower maximum. A cell neighbouring a cell with a highly significant energy content is therefore more likely to

## 5 Improvements for tauRec

contain a real energy deposition than the same cell would be without the neighbour. The topological clustering algorithm exploits this fact.

The algorithm has three adjustable thresholds on the significance  $|E_{cell}|/\sigma_{noise}$  of the energy measured by a cell. These are the *seed threshold*, the *neighbour threshold* and the *cell threshold*. They are used in the three steps of the cluster building. In the first step the cell with the highest significance above the seed threshold, which is not a member of any cluster so far, is searched for and a new cluster is created containing this cell. In the second step the neighbouring cells of the cells that are already part of the cluster are added if they have a significance above the neighbour threshold. This step is repeated until there are no new cells added to the cluster. If a cell to be added is already a member of another cluster, both clusters are merged into one. In a final step, one more iteration of adding neighbour cells is performed with the lowest threshold on the significance, the cell threshold. The neighbour relations are defined between cells in all space directions crossing different calorimeter layers and subsystems if applicable.

These three steps are repeated until there are no more unclustered seed cells left. The algorithm will often merge the energy depositions from different particles into one cluster, so a splitting procedure is performed on all clusters. Cell forming local maxima are searched for in the second and third layer of the electromagnetic calorimeter and, if more than one is found in a cluster, each will form a new cluster. A cell is considered to form a local maximum if it has an energy of at least 500 MeV, no neighbouring cell has a larger energy and it has at least 4 neighbours in the clusters. The border cells between two split clusters will be shared amongst them, with the two weights  $w_1$  and  $w_2$  calculated from the energies  $E_1$  and  $E_2$  of the two clusters and the distances  $d_1$  and  $d_2$  of the cells from the cluster centres as

$$w_1 = \frac{E_1}{E_1 + rE_2}, \quad w_2 = 1 - w_1, \quad \text{with } r = \exp(d_1 - d_2). \quad (5.1)$$

The clusters created by different configurations of the TopoCluster algorithm are characterised by the three cell energy thresholds and the calorimeter layers, that are considered for the selection of cells. Two collections of clusters are created in the ATLAS default reconstruction. The first is called ‘CaloTopoClusters’ and uses the cells from all calorimeter layers with the thresholds on the cell significance in a 4/2/0 configuration for seed/neighbour/cell thresholds. The cell threshold of zero means, that in the last step of the cluster building one extra layer of neighbour cells will always be added, regardless of their energy. The CaloTopoClusters are used inside the ATLAS reconstruction software to reconstruct calorimeter jets with various algorithms and configurations and also to reconstruct the missing transverse energy of the event. The second collection is called ‘EMTopoClusters’ and uses only the electromagnetic barrel and end-cap calorimeters with a 6/3/3 configuration for the thresholds. In the following, when speaking of TopoClusters it is implied that they come from the CaloTopoCluster collection.

## 5.2 Implementation of TopoJet Seeds

The topological clustering described above typically produces one cluster per particle. A  $\tau$  jet or QCD jet therefore gives rise to several clusters, which need to be identified as belonging to the same jet. This is done with a jet reconstruction algorithm, of which several have been developed [5] and are available in the Athena software [29]. One of the most often used jet reconstruction algorithms is the *seeded cone* algorithm, which combines clusters within a radius  $R$  into one object.<sup>1</sup> Typical radii used for inclusive jet reconstruction in ATLAS are 0.4 and 0.7.

To define the jet axis in the seeded cone algorithm, the axis is first centered on the cluster with the highest energy (the seed) and then it is iteratively recalculated by combining all the four-momenta of the clusters<sup>2</sup> inside the radius  $R$ . The algorithm stops when the axis has become stable. Overlapping jets are subject to a splitting/merging step. The shared constituents of the jets are split up between the jets according to their distance to both of the jets and the ratio of the transverse jet energies if both jets have at least half the transverse energy of the other jet. Otherwise the two jets are merged into one.

In the course of the default event reconstruction inside ATHENA, different jet reconstruction algorithms with different parameters are run, each with calorimeter towers as well as *TopoClusters* as input. In addition, for simulated events also the simulated stable particles are grouped into jets with the same algorithms. The cone jets with a radius of 0.4 created from the *TopoJets* of the event offer the functionality that is needed to seed the  $\tau$  reconstruction. The 0.4 radius coincides with the cone size used for the cell selection in the original tauRec package, which makes it a robust choice to base the switch to TopoJets on. In ATHENA these jets are called ‘Cone4TopoJets’.

The *Cone4TopoJets* are only accepted as seeds if they have a transverse energy of more than 10 GeV and an absolute pseudorapidity of  $|\eta| < 2.5$ . The efficiency to reconstruct a true  $\tau$  jet with  $|\eta| < 2.5$  as a *Cone4TopoJet* is shown in Figure 5.1 as function of the transverse momentum of the true  $\tau$  jet for a combined  $Z \rightarrow \tau\tau$  and  $A \rightarrow \tau\tau$  sample. The samples have a lower cut of 12 GeV applied to the visible transverse energy on generator level. The reconstruction efficiency is above 96% for the lowest energies and quickly saturates at nearly 99%.

The jet objects in ATHENA give access to the underlying constituents down to the detector level. In the present case this allows to get the list of calorimeter cells contained in the TopoClusters which in turn make up the jet. This feature makes it possible to directly save a link to the seed jet inside the  $\tau$  candidate and subsequently access the calorimeter cells – used to calculate identification variables and calibrate the energy of the  $\tau$  jet – through that jet object. The advantage of this approach is on the one hand execution speed, as unnecessary copying of data is avoided and already reconstructed objects can be utilised, and on the other hand it provides for a one-to-one relationship

<sup>1</sup>The radius  $R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$  is defined as the distance in the  $\eta$ - $\phi$  plane.

<sup>2</sup>The four-momentum of a cluster is calculated by treating each cell as a massless particle and adding their four-vectors.

## 5 Improvements for tauRec

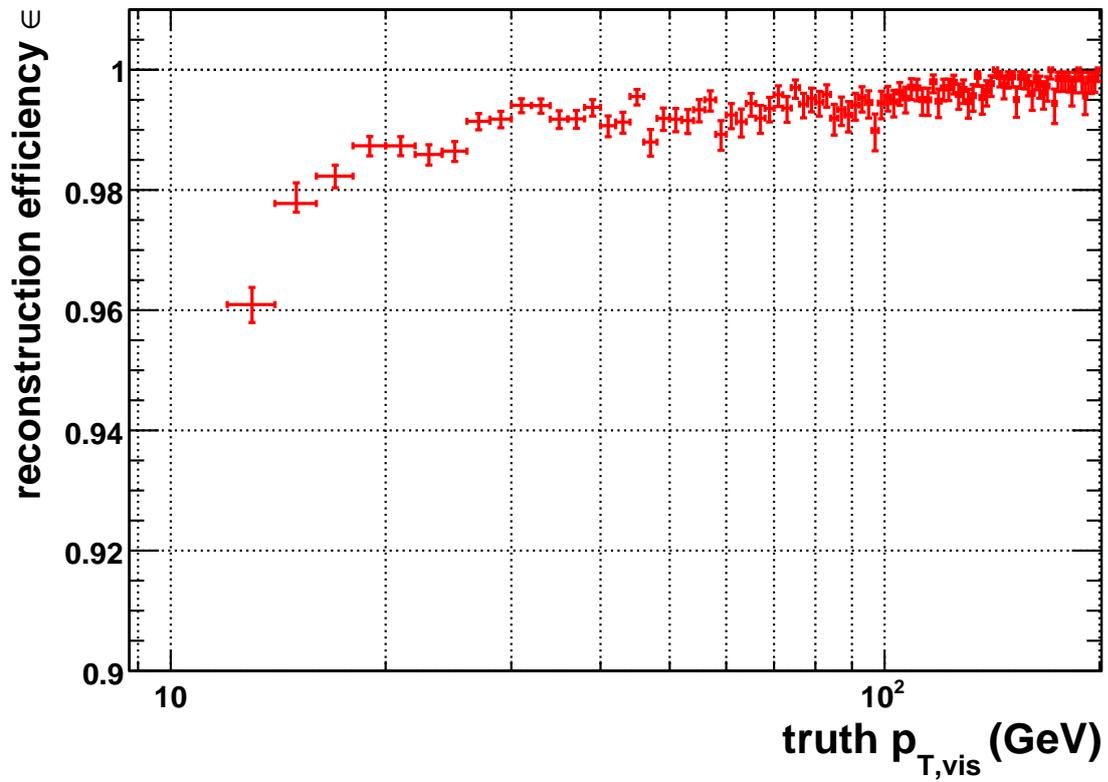


Figure 5.1: Probability to reconstruct a true  $\tau$  jet as a Cone4TopoJet as function of  $p_T$ . Note that the used sample has a minimum  $p_T$  cut of 12 GeV applied on the truth level.

between reconstructed jets and  $\tau$  candidates, making a matching between the two kind of objects simple and unambiguous.

In order to use the seed jet in the above way, the  $\tau$  object was modified to include a link to the jet object and methods were added to access the associated cells inside the  $\tau$  object. The reconstruction tools, which make use of the calorimeter cells were changed to ask the  $\tau$  object for the list of cells, without the need to know the source of the cells. These changes made it possible to switch between the default behaviour and the new seeding through the runtime configuration without changes to the C++ code.

### 5.3 Shower Shape Variables with TopoClusters

It is expected, that the distributions for variables calculated from energy depositions in the calorimeter will change when the cells are taken from the clusters reconstructed by the topological clustering algorithm as opposed to the default cell selection by applying a fixed cut on the significance of the energy deposition in each cell. In order to compare the distributions with and without the use of the topological clustering, about 135.000  $Z \rightarrow \tau\tau$  and  $A \rightarrow \tau\tau$  events as signal and about 140.000 dijet events for the QCD background have been reconstructed on a local cluster in Freiburg with a modified version of the ATHENA release 13. This version included the changes described above. The simulated detector data used for these samples came from the official ATLAS Monte Carlo production, and were produced with release 12 of the ATHENA software.

The more efficient suppression of noise by the topological clustering leads in most cases to smaller mean values for variables describing the size of a shower. As expected, the low energy range is affected more strongly, because the relative contribution of noise to the measured energy is larger for smaller real energy depositions. Also one should keep in mind that the mean transverse energy of the lowest bin is shifted to smaller values because of the higher reconstruction efficiency in this range.

The discrimination of a variable can be quantified by the separation power  $S$ , which is defined by [10]

$$S = \frac{1}{2} \int \frac{(P_{sig}(y) - P_{back}(y))^2}{P_{sig}(y) + P_{back}(y)} dy. \quad (5.2)$$

$P_{sig}$  and  $P_{back}$  are the probability densities of the considered variable for signal (*sig*) and background (*back*), respectively. The separation power has a value of one if the distributions for signal and background do not overlap, hence the variable discriminates perfectly between signal and background. If the distributions of the variable are the same for signal and background the separation power has a value of zero. The separation powers of some of the variables used for the  $\tau$  identification are shown for two intervals of the transverse energy in Table 5.1, separately for the two samples seed with and without topological clusters. One can see that most variables have a better separation power in the sample that was reconstructed with TopoJet seeds, especially in those cases where a significant difference can be observed.

## 5 Improvements for *tauRec*

variable	E <sub>T</sub> range	separation power $S$	
		default	TopoJet
EM radius	17 – 32 GeV	0.4214	0.5411
	45 – 70 GeV	0.3861	0.5317
Isolation	17 – 32 GeV	0.3647	0.3477
	45 – 70 GeV	0.4235	0.4811
$\eta$ -strip width	17 – 32 GeV	0.1768	0.3883
	45 – 70 GeV	0.2026	0.4338
no. of $\eta$ -strip hits	17 – 32 GeV	0.040	0.035
	45 – 70 GeV	0.076	0.066
E <sub>T</sub> /p <sub>T,leading</sub>	17 – 32 GeV	0.2730	0.2855
	45 – 70 GeV	0.2206	0.2363

Table 5.1: Separation power of the calorimeter based variables for default reconstruction and seeding by *TopoJets*. Most variables have a better separation power when reconstructed from topological clusters.

The variables that changed the most are the *EM-Radius*, the *isolation fraction*, the  *$\eta$ -strip width* and the number of  $\eta$ -strip hits. Comparisons of those four variable distributions between the original and the modified version of the algorithm are shown in the following for different E<sub>T</sub> bins. The other variables are not affected much and are therefore not shown here but in Appendix A.

The change of the EM-Radius distribution, which is shown in Figure 5.2, increases the discrimination of the variable, especially in the very low energy region. In the lowest E<sub>T</sub> bin shown in Figure 5.2 the effect of the adaptive noise suppression of the topological clustering can be observed. The EM-Radius for  $\tau$  jets is shifted to lower values while it becomes much larger for QCD jets.  $\tau$  jets deposit their energy in a small cone and therefore a lot of cells in the cone of radius 0.4 used for calculating the EM-Radius have only very little energy deposited in them and are more efficiently suppressed by the topological clustering than by a fixed cut on the significance. For QCD jets the situation is reversed. They spread out their energy over a large region, so that on average only a small amount of energy is deposited in each cell. In the default reconstruction many cells with real energy were therefore suppressed by the fixed cut on the significance. The topological clustering retains these cells if they are close enough to the shower maximum of one of the particles in the jet, leading to a broader reconstructed shower shape. At the higher transverse energies

## 5 Improvements for tauRec

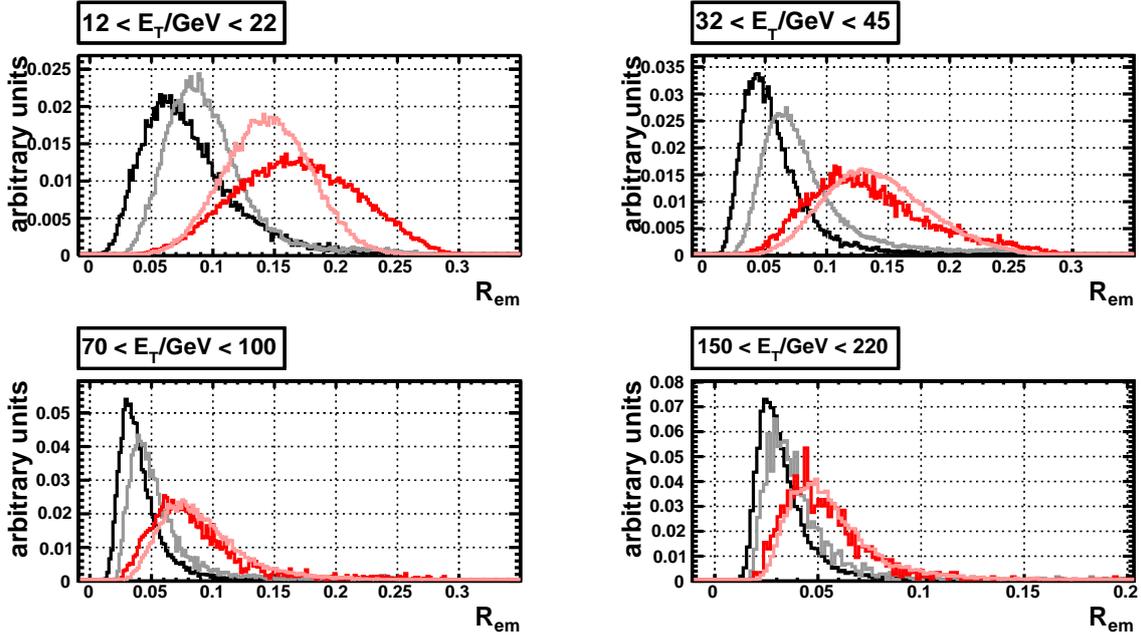


Figure 5.2: Comparison of the EM-Radius between default tauRec and *TopoJet* seeded tauRec for different  $E_T$  ranges. Light coloured distributions are for the default version and the dark coloured ones for *TopoJets*. Black/grey distributions are for  $\tau$  jets and the red ones are for QCD jets.

both signal and background tend to be shifted to smaller values, but the effect is more pronounced for  $\tau$  jets for the same reasons, also leading to a stronger separation.

The isolation fraction, which is shown in Figure 5.3, also shows some small differences in the shape of the distributions, mainly in the lowest  $E_T$  bin. But the discrimination does not change much.

The  $\eta$ -strip width shows the most pronounced change in the shape of the distribution, as shown in Figure 5.4. The distribution generally peaks at lower values and especially for the signal has a very sharp peak at zero. Also now negative values can be observed, which is due to the cut on the absolute cell energy applied by the *TopoCluster* algorithm, whereas in the default algorithm cells with negative energy were discarded. This might also explain the dramatic change in the shape, since cells with positive and negative fluctuations now cancel each other out. The discrimination of the variable is somewhat larger than in the default reconstruction.

The distribution for the number of  $\eta$ -strip hits is practically unchanged other than in the lowest  $E_T$  bin, as can be seen in Figure 5.5. This is expected, since the 200 MeV requirement on the cell energy already corresponds to an energy of about  $4\sigma_{noise}$ . So, cells that were counted as hits in the default algorithm will be able to start a *TopoCluster* most of the time, and are therefore also contained in the list of cells for the new algorithm. The change in the lowest bin can be attributed to the lower mean energy of reconstructed objects, due to the higher efficiency at low energies.

## 5 Improvements for tauRec

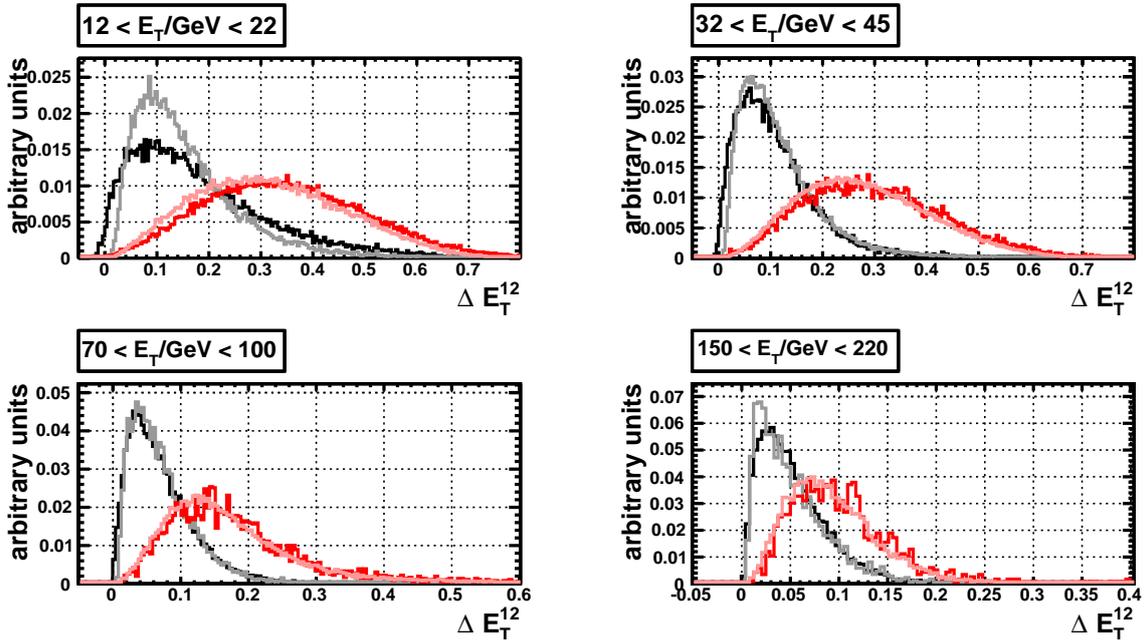


Figure 5.3: Comparison of the *isolation fraction* between default tauRec and *TopoJet* seeded tauRec for different  $E_T$  ranges. Light coloured distributions are for the default version and the dark coloured ones for *TopoJets*. Black/grey distributions are for  $\tau$  jets and the red ones are for QCD jets.

## 5 Improvements for tauRec

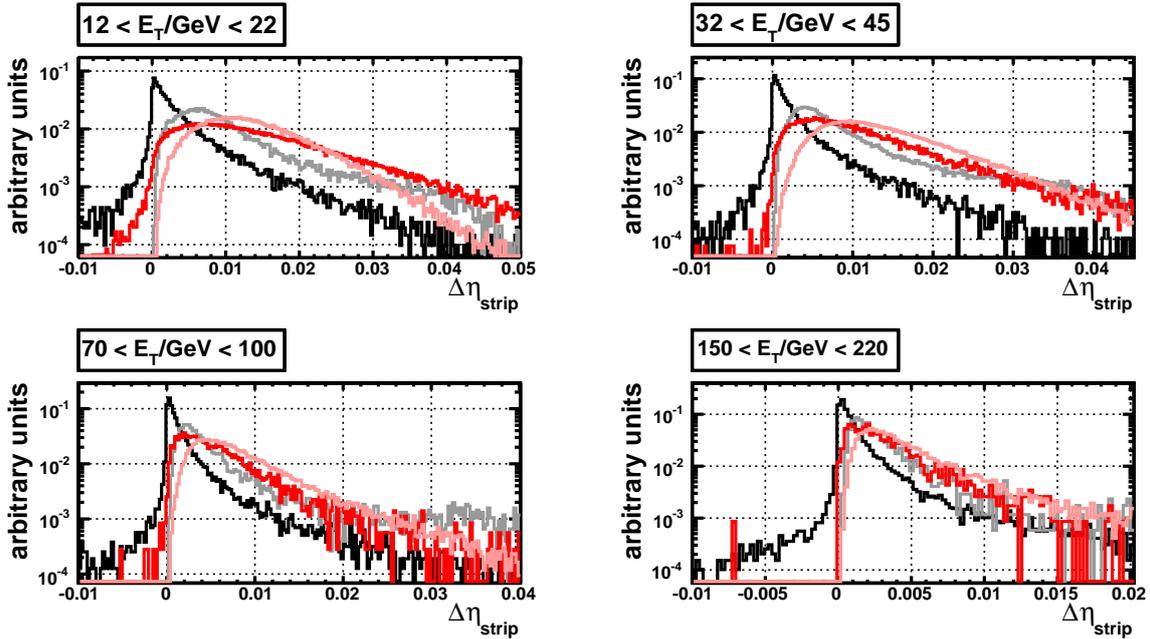


Figure 5.4: Comparison of the  $\eta$ -strip width between default tauRec and *TopoJet* seeded tauRec for different  $E_T$  ranges. Light coloured distributions are for the default version and the dark coloured ones for *TopoJets*. Black/grey distributions are for  $\tau$  jets and the red ones are for QCD jets.

### 5.4 New Likelihood Distributions

To make use of the better discrimination offered by the identification variables when calculated from TopoCluster cells, it was necessary to create new probability density functions to be used in the calculation of the likelihood ratio.

To obtain the PDFs from the available samples, all reconstructed  $\tau$  objects from the signal samples that could be matched to a true  $\tau$  jet within a cone of 0.2 were used as the signal. All unmatched  $\tau$  objects from the QCD dijet samples were used as the background. Only candidates having one to three associated tracks were used to create the probability density functions, corresponding to the preselection cut later applied during the  $\tau$  reconstruction inside ATHENA.

The probability density functions of the eight variables were created for nine different  $E_T$  bins. Their bin boundaries were chosen as 10, 17, 32, 45, 70, 100, 150, 220, 500 and  $> 500$  GeV, to have roughly the same number of  $\tau$  jets in each bin. The distributions for the five continuous variables were smoothed using the ‘Locfit’ software package [18, 19]. The Locfit package implements various methods for density estimation based on the kernel techniques described in Section 3.2 in a computationally efficient way. The specific method used was the local likelihood method described at the end of Section 3.2, where polynomials are fitted to weighted neighbourhood of each point. A nearest neighbour bandwidth and third order polynomials were used in the local fitting procedure. The number of nearest

## 5 Improvements for tauRec

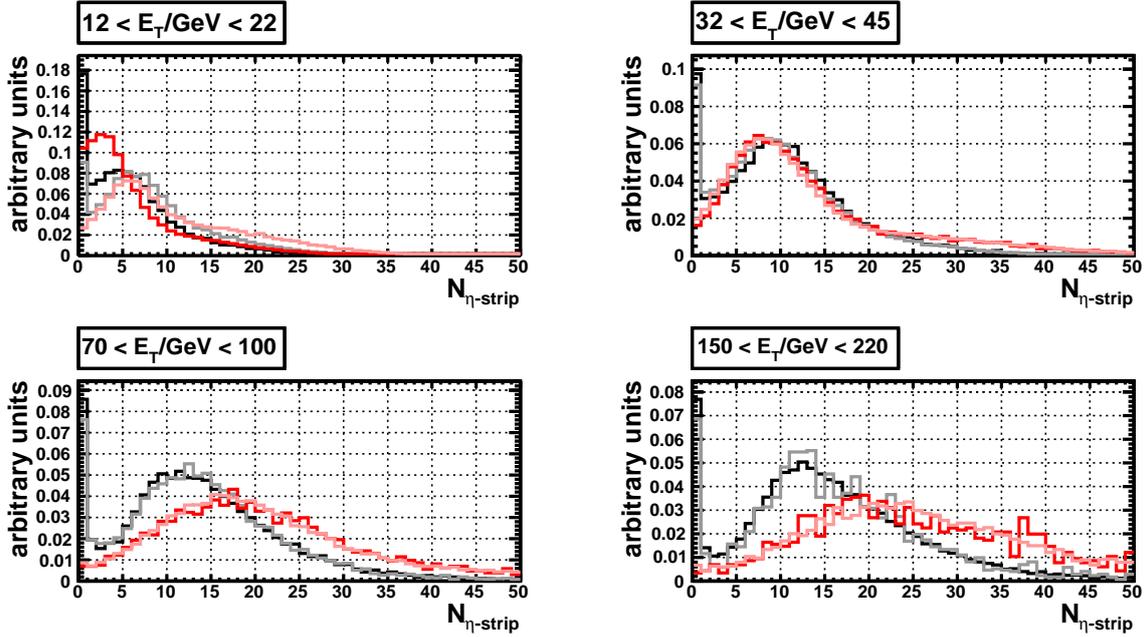


Figure 5.5: Comparison of the number of  $\eta$ -strips hits between default tauRec and *TopoJet* seeded tauRec for different  $E_T$  ranges. Light coloured distributions are for the default version and the dark coloured ones for *TopoJets*. Black/grey distributions are for  $\tau$  jets and the red ones are for QCD jets.

neighbours that were used to determine the bandwidth was adapted to the available statistics in each  $E_T$  bin to produce a smooth result and still reproduce the gross features of the distributions. An example for the result of this estimation procedure can be seen in Figure 5.6 for the  $E_T$  bin between 45 GeV and 70 GeV.

The discrete distributions for the number of  $\eta$ -strip hits were smoothed using a moving average smoothing algorithm described in [16], which evens out bin by bin fluctuations. The first bin was excluded from the smoothing procedure. An example of this is also included in Figure 5.6. For the number of tracks and the charge, the histograms were directly used as the PDFs.

As already mentioned, the PDFs used in the default version of tauRec were created from older monte carlo samples, based on a simpler detector geometry and simulation compared to the current version of ATHENA. To disentangle the effect on the identification performance due to a more up to date training from the effect due to the switch to TopoJet seeds, the same procedure as described above was also used to extract PDFs from samples, that were reconstructed with an unmodified version of the ATHENA software.

## 5 Improvements for tauRec

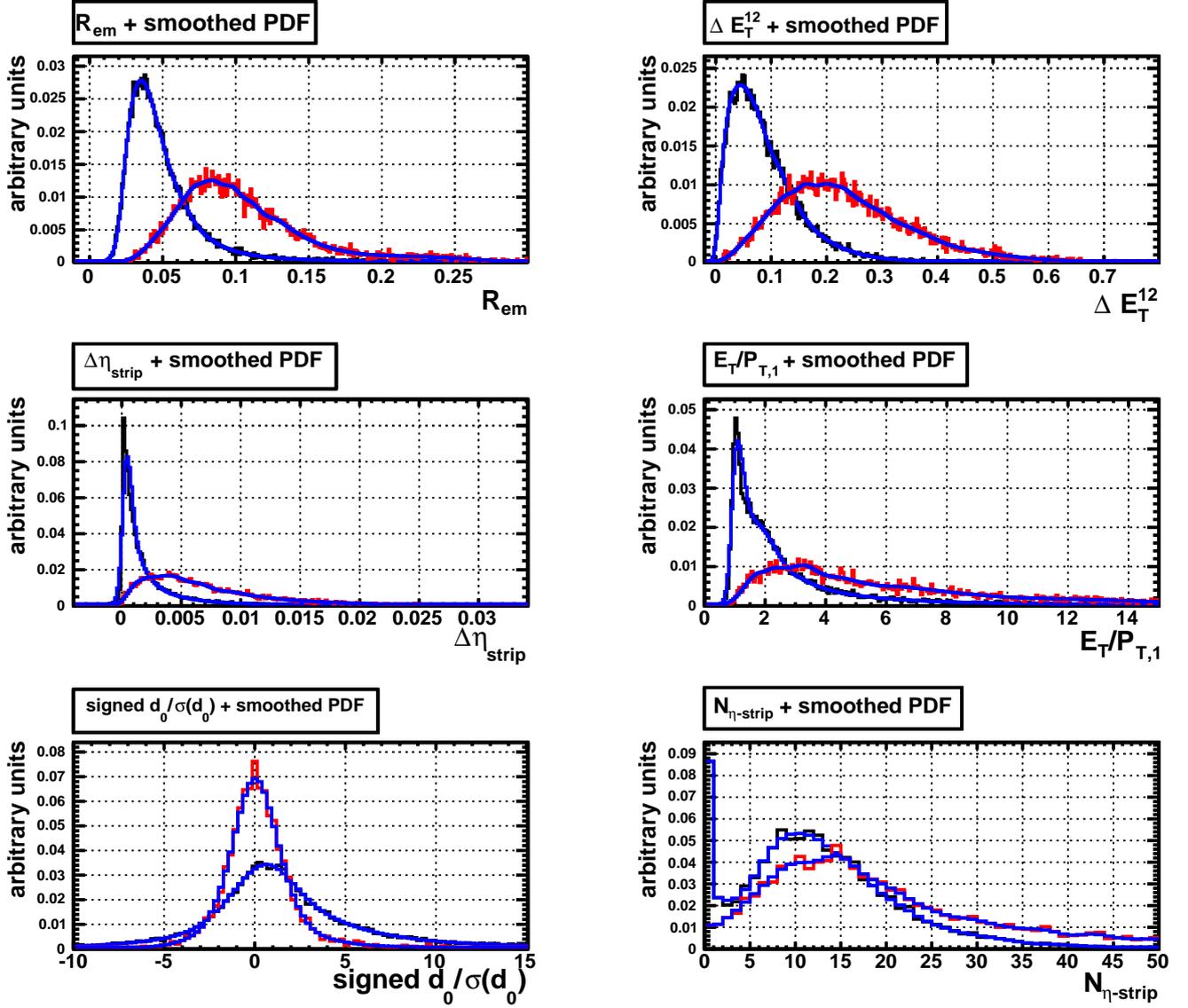


Figure 5.6: The smoothed PDFs overlaid with the unsmoothed distributions for the  $E_T$  bin from 45 GeV to 70 GeV. The black distributions are for  $\tau$  jets and the red ones are for QCD jets. The smoothed distributions are shown in blue. The Ntrack and Charge variables are not shown, since their distributions are not smoothed.

## 5.5 Performance

In order to assess the performance of the tauRec algorithm based on TopoClusters, the same samples were used, from which the PDFs have been extracted. Ideally, statistically independent data samples should be used, but it was necessary to use the same samples because of the limited available statistics. Due to the high achievable rejection factors a large number of QCD jets is needed to reach a useful error on the rejection. A slightly higher rejection can be expected compared to a statistically independent sample. But the effect should be fairly small, since the likelihood method is generally not very prone to overtraining, and the applied density estimation procedure further helps to reduce the effects of fluctuations in the training sample on the probability densities.

The distributions of the likelihood discriminant were calculated for several  $p_T$  bins using the different probability density functions, which were extracted from the official samples used for the results shown in Section 4.3.3, and the privately created samples with TopoJet seeding. By varying the cut on the discriminant the background rejection for different signal efficiencies can be determined.

The resulting rejections for the TopoJet samples when applying the likelihood with PDFs created from the same samples are shown in Figure 5.7 as a function of the signal efficiency for different  $p_T$  bins and in Figure 5.8 as a function of the true visible  $p_T$  for two selected signal efficiencies. Here again the true visible The values for some selected  $p_T$  bins and signal efficiencies are given in Table 5.2 together with the rejections achieved by the default version of tauRec. The efficiencies and rejections are calculated with respect to the true  $\tau$  jets and Cone4Truth jets respectively, both are required to have  $|\eta| < 2.5$ .

truth $p_{T,vis}$ [GeV]	signal efficiency TopoJet		signal efficiency default		gain	
	30%	50%	30%	50%	30%	50%
12 – 15	$139.5 \pm 7.1$	$41.5 \pm 1.1$	-	-	-	-
15 – 25	$319 \pm 18$	$77.5 \pm 2.1$	$245.3 \pm 3.5$	$19.35 \pm 0.08$	+30%	+300%
25 – 40	$630 \pm 61$	$170.4 \pm 8.6$	$415.5 \pm 9.4$	$120.4 \pm 1.5$	+52%	+42%
40 – 60	$1116 \pm 174$	$267 \pm 20$	$626 \pm 22$	$174.2 \pm 3.2$	+78%	+53%
60 – 90	$1299 \pm 213$	$389 \pm 35$	$1128 \pm 53$	$293 \pm 7$	+15%	+33%
90 – 120	$5290 \pm 2366$	$635 \pm 98$	$1744 \pm 145$	$395 \pm 16$	+203%	+61%

Table 5.2: Rejection factors for some selected signal efficiencies and  $p_T$  ranges. For even higher transverse momenta the errors on the rejection become extremely large due to limited statistics. For comparison the rejections which can be achieved with the default version of tauRec are given (see Section 4.3.3) and the also the relative gain of the TopoJet seeded reconstruction over the default reconstruction is shown in the last two columns.

Compared to the rejections achieved with the default tauRec package, given in Section 4.3.3, improvements of 15-78% can be observed in the  $p_T$  range below 100 GeV. For low transverse momenta (between 15 GeV and 25 GeV) and 50% efficiency the TopoJet seeded reconstruction can even improve the rejection by a factor of four. Below 15 GeV

## 5 Improvements for tauRec

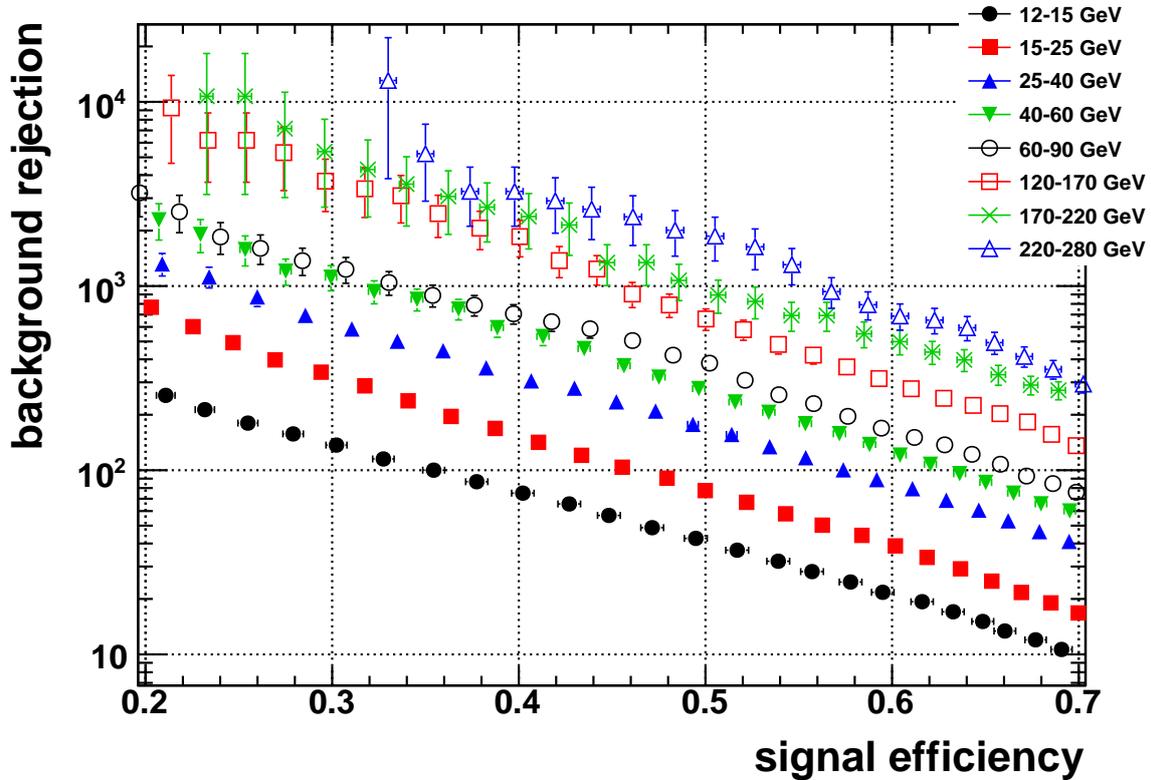


Figure 5.7: Rejection of QCD jets as function of the efficiency to select true  $\tau$  jets. The rejections are shown for different bins of the true visible  $p_T$ .

the new algorithm is now able to reach an efficiency of 50% after identification, while still providing some modest rejection against QCD jets.

It should be noted, that the sample from which the above results have been obtained was reconstructed with release 13 of the ATLAS reconstruction software ATHENA, while the numbers shown in Section 4.3.3 were obtained from a sample which was reconstructed with release 12 of ATHENA. Both samples were created from simulated detector data based on the same simulation code and the same detector geometry. No significant changes in the reconstruction affecting the performance of the tauRec package were introduced between the two releases. There was not enough statistic available for a direct comparison of the identification performance between release 12 and release 13, but a small sample reconstructed with release 13 was available for a comparison of the identification variables. This comparison is shown in Appendix A. The distributions of the identification variables agree quite well, so no significant difference in performance is to be expected between the two releases.

It is interesting to know how much of the performance gain between the default tauRec version and the TopoJet seeded one can be attributed to an intrinsically better discrimi-

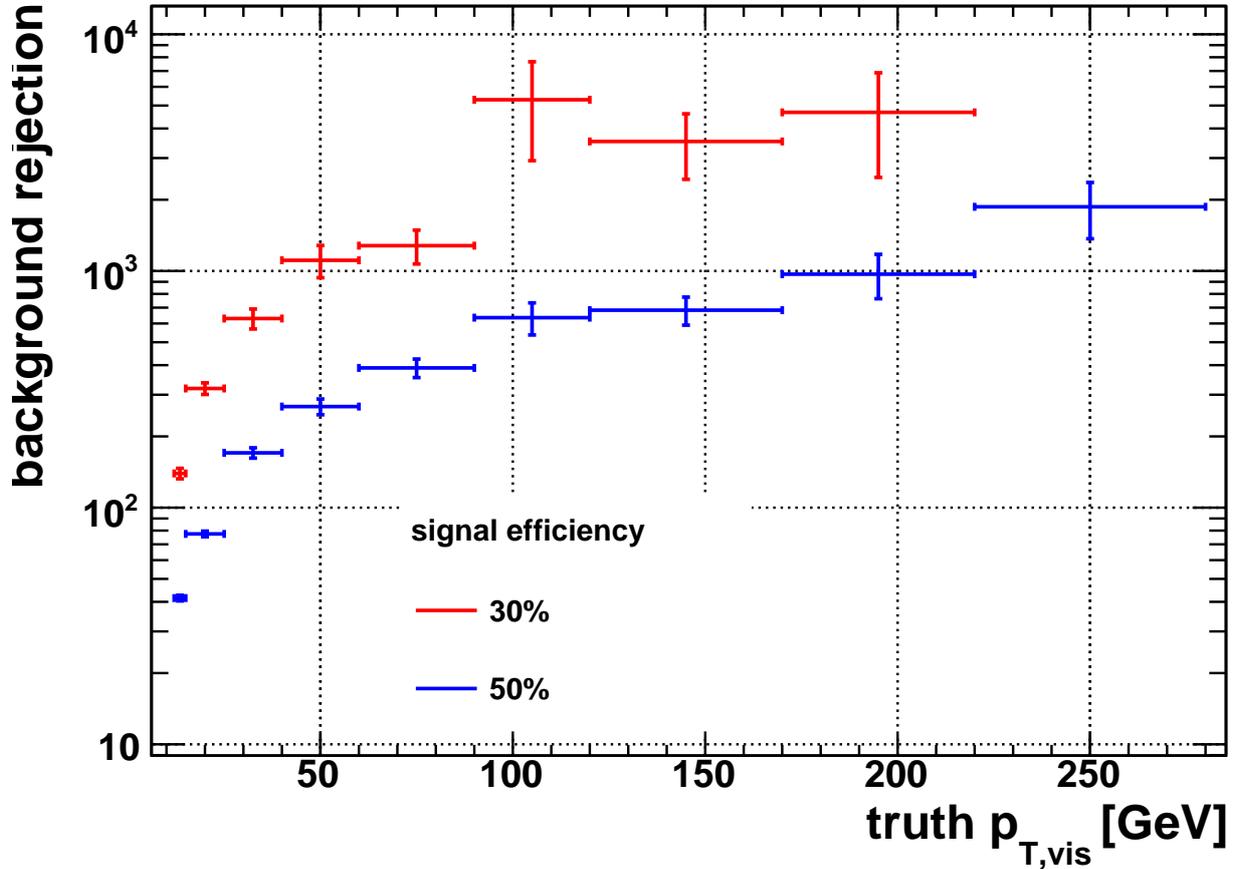


Figure 5.8: Rejection against QCD jets as a function of the visible  $p_T$  for a signal efficiency of 30% (red) and 50% (blue). For the highest  $p_T$  bin at 30% efficiency no QCD jets passed the selection.

nation power of the identification variables when calculated from TopoCluster cells, and how much arises due to fact that the PDFs were created from data simulated with the same detector description as the data used to determine the rejection of the new method.

To disentangle those two effects, new PDFs were also created from the sample that was reconstructed with release 12 of ATHENA and was used for the results shown in Section 4.3. The likelihood method was applied to this same sample, to compare the identification performance to the one achieved on the TopoJet sample with the appropriate likelihood training. The resulting rejections are shown in Figure 5.9 as a function of the transverse momentum of the  $\tau$  jets.

It is evident that for small transverse momenta up to about 60 GeV the rejection against QCD jets that can be reached with the default tauRec implementation can be only marginally increased by updating the likelihood with data, which were simulated with an up to date detector description. In the intermediate  $p_T$  region roughly half of the

## 5 Improvements for tauRec

increase in the rejection can be reproduced by using the updated likelihood on top of the default tauRec reconstruction.

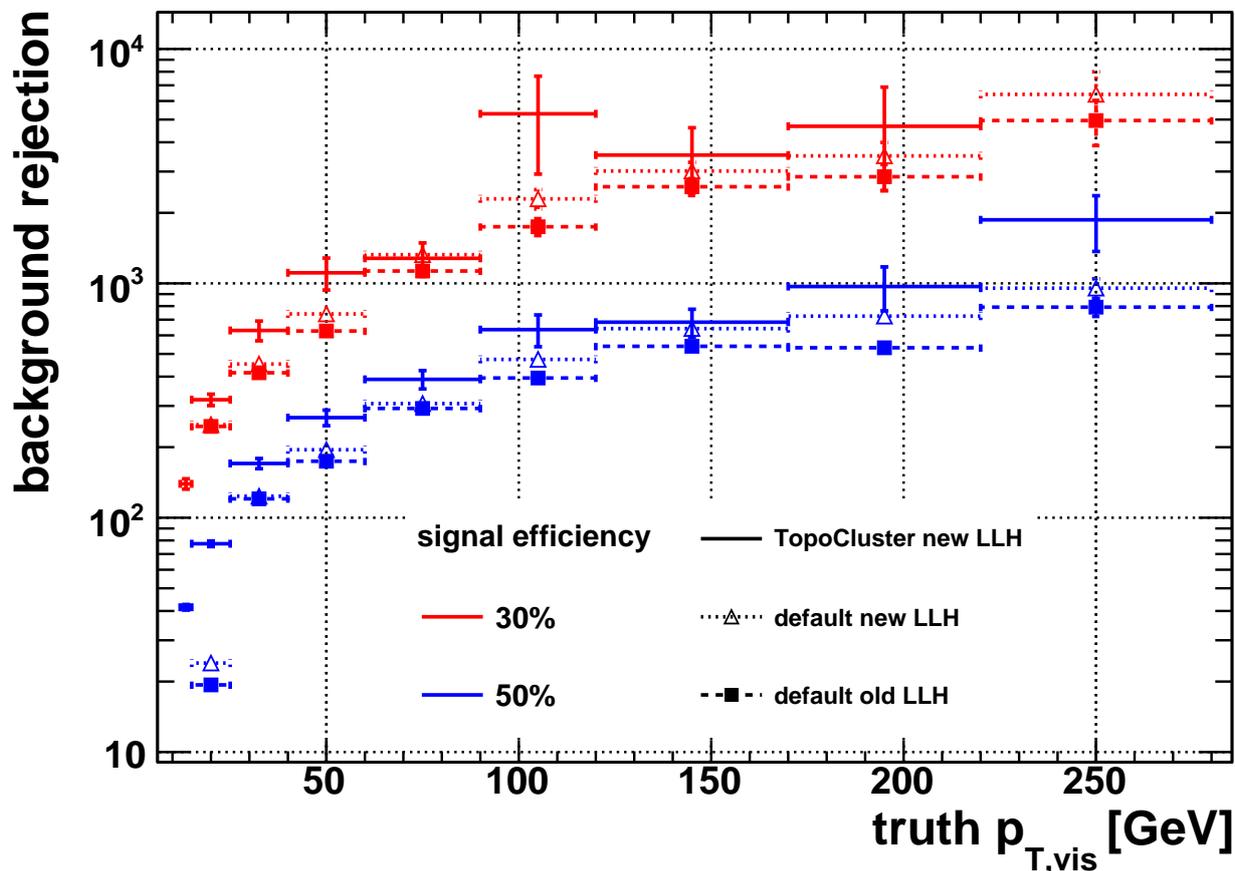


Figure 5.9: Comparison of the rejection against QCD jets as a function of the visible  $p_T$  between the reconstruction with TopoJet seeds and the default reconstruction with two different likelihood discriminants. These are the default likelihood (old LLH), as implemented in ATHENA, and a likelihood where the PDFs have been retrained from the same sample used to calculate the rejection (new LLH). The TopoJet seeded sample and the sample using the default tauRec version were reconstructed with release 13 and release 12 of ATHENA, respectively. Details are given in the text.

## 6 $\tau$ Algorithm Merging

The two distinct  $\tau$  reconstruction algorithms, tauRec and tau1P3P, have some complementary strengths and weaknesses. The ‘end user’ doing an analysis involving  $\tau$  leptons needs to decide which algorithm to use when selecting  $\tau$  leptons. In an experiment as complex as ATLAS this means a lot of time has to be spent to obtain the detailed knowledge that is needed to make an optimal choice. It was therefore decided to merge both algorithms into a single default algorithm, providing the best features of both algorithms and hopefully eliminating some of the weaknesses.

A first step towards a merged algorithm was implemented in the scope of this thesis, by combining the  $\tau$  objects created by both algorithms into one collection. This was done by matching the seeds of both algorithms at reconstruction level, and adding functionality which enables the user to check by which of the two algorithms a specific  $\tau$  object was reconstructed. This can also include  $\tau$  jets reconstructed by both algorithms, and each  $\tau$  object can have the detailed information reconstructed by each of the two existing  $\tau$  algorithms attached to it.

### 6.1 Merging of tauRec and tau1P3P

The two  $\tau$  reconstruction algorithms tauRec and tau1P3P were essentially implemented as two independent algorithms sharing some common code and data structures. During the event reconstruction in ATHENA both algorithms were called in sequence, each looping over all appropriate seeds in the event. In the end two separate collections of TauJet objects were produced, one each for tauRec and tau1P3P, with no connection between corresponding objects in both of them.

A necessary condition for combining both algorithms is to have a common definition of a reconstructed  $\tau$  object. The decision was made to use the union of the  $\tau$  candidates reconstructed by both algorithms. In Figure 6.1 the number of true  $\tau$  jets reconstructed by either both algorithms or only one of them is shown for a  $Z \rightarrow \tau\tau$  sample as a function of the visible transverse momentum of the  $\tau$  jets. The sample was reconstructed with a version of ATHENA which already included a version of tauRec seeded by topological clusters.

The fraction of  $\tau$  jets reconstructed by both algorithms is about 75% for the whole sample. At a transverse momentum of 12 GeV this fraction is 50% and reaches about 90% for high transverse momenta above 80 GeV. This fraction is dominated by the reconstruction efficiency of the tau1P3P algorithm. Due to the high reconstruction efficiency of tauRec only a small fraction of  $\tau$  jets is reconstructed by tau1P3P only. The fractions of  $\tau$  jets

## 6 $\tau$ Algorithm Merging

truth $p_{T,vis}$	both	only tauRec	only tau1P3P
12 – 20 GeV	60%	37%	1%
20 – 30 GeV	74%	26%	0.5%
30 – 40 GeV	80%	19%	0.2%
40 – 50 GeV	83%	16%	0.1%
50 – 60 GeV	86%	13%	0.1%
60 – 70 GeV	87%	13%	0.1%
70 – 80 GeV	90%	10%	0.1%
80 – 90 GeV	91%	9%	0.3%

Table 6.1: Fraction of  $\tau$  jets that is reconstructed by each algorithm or by both for different bins of the true visible transverse momentum.

reconstructed by both algorithms or only one of them are listed in Table 6.1 for different ranges of the true visible transverse momentum.

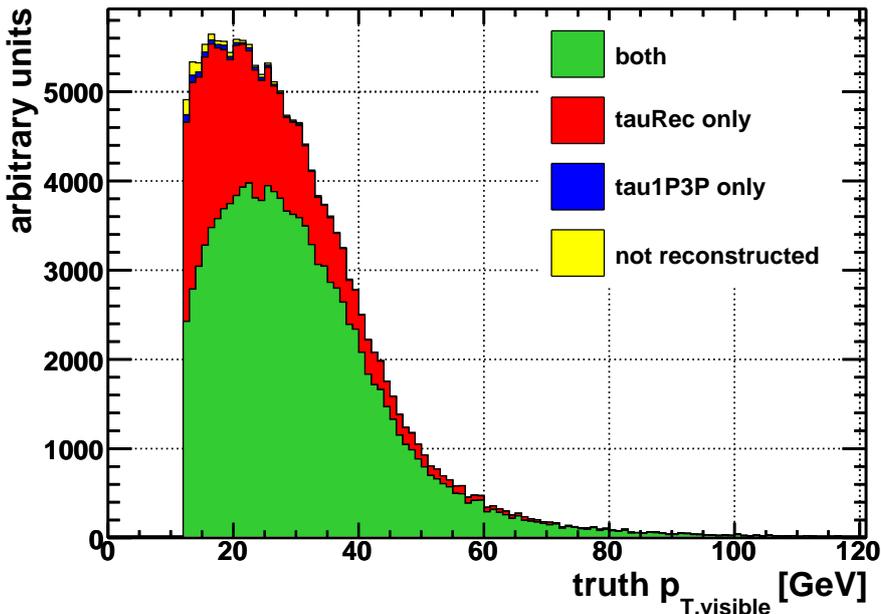


Figure 6.1: Transverse momentum of  $\tau$  jets reconstructed by either of the two algorithms for a  $Z \rightarrow \tau\tau$  sample. Shown is the sum of the distributions for  $\tau$  jets reconstructed by both algorithms (green), only tauRec (red), only tau1P3P (blue) or none of the two (yellow) algorithms. For the whole sample about 75% of all  $\tau$  jets are reconstructed by both algorithms, saturating at about 85% for higher  $p_T$ .

The reconstruction efficiencies of the two algorithms for QCD jets are shown in Figure 6.2. Tau1P3P reconstructs QCD jets with a lower probability than  $\tau$  jets at comparable momenta, which provides already a very modest rejection against QCD jets up to

transverse momenta of about 60 GeV. This can mostly be attributed to the requirement of at least one track with a transverse momentum of more than 6 GeV, which many QCD jets do not fulfil.

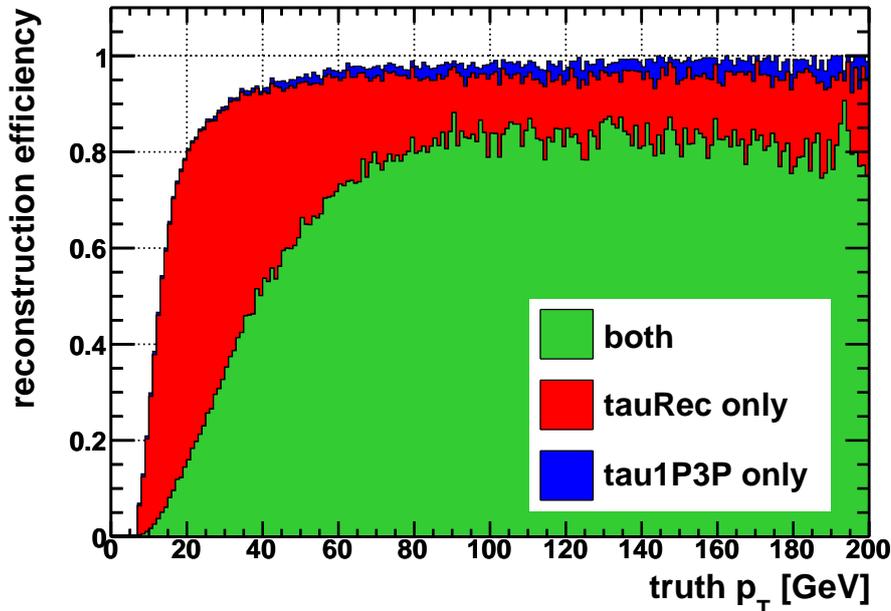


Figure 6.2: Reconstruction efficiency for QCD jets for a combination of several QCD dijet samples. Shown is the sum of the efficiencies for QCD jets to be reconstructed by both algorithms (green), only tauRec (red), or only tau1P3P (blue).

The goal of combining the algorithms was to reconstruct exactly one  $\tau$  object for each true  $\tau$  jet, which offers the specific information reconstructed by each of the two algorithms. By separately saving the data reconstructed by tauRec and tau1P3P in the  $\tau$  object, it is possible to use them as if the algorithms were still completely separated. This is especially important for validating the implemented changes by checking them for consistency with the known behaviour of the two original algorithms.

## 6.2 Technical Implementation

At the time when the implementation of the merged  $\tau$  algorithm started, the version of the ATHENA software was already considered to be close to the final version used for the first data taking. It was therefore important to take the first steps towards this merging in a way to minimise the amount of code that had to be changed and validated. It was also important to keep a high degree of ‘backward compatibility’, to make it possible to perform previously established analyses on data that will be reconstructed with the merged algorithm without the need for extensive changes to the analysis code.

## 6 $\tau$ Algorithm Merging

When using the combination of the two different seed definitions, the cone jets made from topological clusters and the tracks in the inner detector with a transverse momentum above a certain threshold, it is desirable to combine two seeds which originate from the same particle or jet into one  $\tau$  object, in order to avoid the reconstruction of duplicate  $\tau$  objects. In this first step towards a combined algorithm this was, for reasons of simplicity mentioned above, implemented as a simple cone matching between the  $\tau$  objects reconstructed by the tau1P3P package and the *Cone4TopoJets* which are the seeds for the tauRec package. For every  $\tau$  object that is reconstructed by tau1P3P, it is checked if any eligible seed for tauRec is inside a cone of radius 0.2 around the direction of this  $\tau$  object and has not been used before.

The cone size has been chosen to be large enough to match nearly all candidates originating from the same  $\tau$  jet. Figure 6.3 shows the distribution of the distance in the  $\eta$ - $\phi$  plane between tau1P3P candidates and tauRec candidates that are both matched to each other and where the tau1P3P candidate is matched to a true  $\tau$  jet within a cone of radius 0.2. The plot was created from a mixed  $Z \rightarrow \tau\tau$  and  $A \rightarrow \tau\tau$  sample with various masses of the A boson, in which the  $\tau$  jets are mostly isolated from the rest of the event. Within this sample, the applied matching procedure shows an efficiency of nearly 100%. The situation will most likely be different for QCD jets or for events in which many jets and other objects are produced and hence combinatorial effects will play a bigger role. In such events the axis of the tau1P3P or tauRec candidates reconstructed from a  $\tau$  jet can be shifted by nearby tracks or energy depositions in the calorimeter, respectively. This can lead to duplicate reconstructed  $\tau$  jets or to two different close by true objects being matched to the same  $\tau$  candidate. Therefore the matching procedure should be studied in more detail in the future, to determine the probability of such occurrences in different types of events.

After no more seeds for tau1P3P are left, all *Cone4TopoJets* that are not yet matched to a tau1P3P candidate are used to seed a  $\tau$  candidate which is only reconstructed by tauRec.

As already mentioned, the C++ class `TauJet` allows to attach algorithm specific information of different content in order to use the same class for  $\tau$  objects reconstructed by either tau1P3P or tauRec. This is done by attaching specialised objects with class names like `TauRecDetails` and `Tau1P3PDetails`. The user can retrieve the objects with a template method using the class name as an identifier. It was therefore straightforward to attach more than one of these detail objects to the `TauJet` object, coming from different algorithms. Furthermore, a method was added to check whether a given  $\tau$  object was reconstructed by tau1P3P, by tauRec or by both of them. This scheme has the advantage that the interface could be left unchanged, so the migration is straightforward for users of the  $\tau$  objects.

Some basic properties of the  $\tau$  objects, like the four vector and the list of associated tracks, are contained in the `TauJet` class itself and have conflicting definitions in both algorithms. In this cases it was decided upon one definition to be used as the default in case a  $\tau$  candidate is reconstructed by both algorithms and the conflicting definition is

## 6 $\tau$ Algorithm Merging

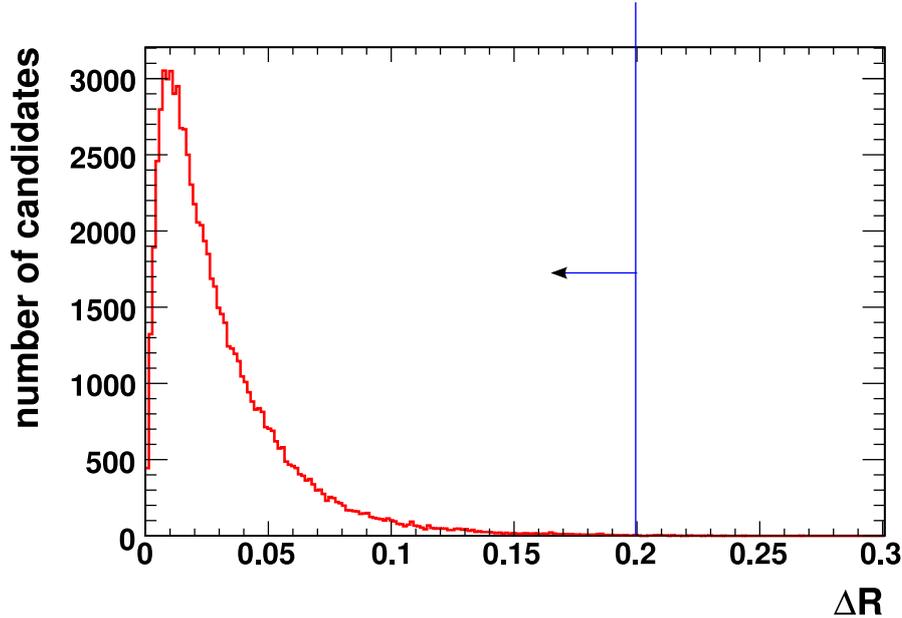


Figure 6.3: The distance in the  $\eta$ - $\phi$  plane between matched tau1P3P and tauRec candidates. The tau1P3P candidate is also required to be matched to a true  $\tau$  jet within a cone of radius 0.2.

stored in the algorithm specific `TauRecDetails` or `Tau1P3PDetails` object. Specifically, the following definitions are used for overlapping candidates:

- transverse energy: tauRec
- $\eta$ ,  $\phi$ : tau1P3P
- charge: tau1P3P
- associated tracks: tau1P3P

This choice was based on the quality of the reconstruction by both algorithms for those properties. The transverse energy was taken from the tauRec algorithm, although the transverse energy reconstructed by tau1P3P has a better resolution in the range below about 100 GeV. However, the energy flow calculation of tau1P3P is not calibrated for higher energies and does not have a linear response in this region, whereas the transverse energy reconstructed by tauRec has the correct scale over the whole energy range. The pseudorapidity and the azimuthal angle are reconstructed with a slightly better resolution by tau1P3P, where they are determined from the track system, owing to the higher granularity of the inner tracker compared to the calorimeter. Also the percentage of  $\tau$  jets with correctly assigned charge and track multiplicity is higher for tau1P3P, due to the more stringent requirements on the track quality.

All these modifications, though often trivial on their own, can have unwanted side effects in a complex software like ATHENA, where a lot of sometimes hidden dependencies exist between different parts of the code. Therefore a great deal of attention was put into validating the modifications at each step, to make sure the produced results are consistent with the ones before the merged reconstruction was implemented.

## 6.3 Optimisation of Identification Performance

The merged TauJet objects make it possible to use identification variables coming from both tauRec and tau1P3P in case the  $\tau$  jet is reconstructed by both algorithms. In order to see how this can be exploited to further improve the performance of the identification offered by the individual algorithms, a study was done to gauge the potential gain when combining variables from both algorithms. It was decided to base the new identification procedure on a neural network discriminant.

The  $\tau$  candidates fall into three categories: reconstructed by both algorithms (overlap candidates), reconstructed only by tauRec (tauRec candidates) and reconstructed only by tau1P3P (tau1P3P candidates). Only the overlap candidates in the first category have all the identification variables from both algorithms available. To make use of the maximum available information for each category, different neural networks are trained for overlap candidates and tauRec candidates. No special training was done for tau1P3P candidates, since they constitute only a relatively small fraction of the signal candidates, mainly in the very low  $p_T$  region, and an optimised neural network discriminant is already available for them.

The newly trained discriminant is later compared to the recommended discriminant for each of the two standalone algorithms, which is the likelihood developed in Section 5.4 for tauRec and the neural network discriminant described in Section 4.4.3 for tau1P3P. Compared to what is described in Section 4.3 and Chapter 5, an additional development took place for the tauRec package, which has an impact on the identification performance. The criteria for associating tracks were changed, lowering the minimal required transverse momentum to 1 GeV and introducing some loose requirements on the quality of the track reconstruction. The probability density functions were updated to reflect this change and the rejection against QCD jets was improved by about 10-20% compared to the numbers shown in Section 5.5.

### 6.3.1 Training of the Neural Networks

For the training and evaluation of the neural networks the FORTRAN package JETNET was used in the version 3.5 [22], and was interfaced through an enhanced version the C++ class TJetNet [23, 35]. The JETNET package offers a wide variety of options for training multilayer feed-forward networks, such as different learning algorithms, activation functions and error measures, and is a well established software package in high energy physics. To train the neural networks, a program was written, which implements the early stopping

## 6 $\tau$ Algorithm Merging

described in Section 3.3 in an automated way, so that a large number of different neural networks can be efficiently trained to compare their performance.

To become familiar with the properties of the training procedure and the preparation of the training data influencing the classification performance of the resulting neural network, the configuration of the neural network inside the tau1P3P package was used as a template for further optimisation. A neural network with a fixed topology of two hidden layers with 30 nodes each and a fixed set of input variables given in Section 4.4.3 was trained using different options for the training procedure and the composition of the training samples (splitting of the data into a training and a validation sample, scaling of the input variables onto a fixed range of values) until comparable results to the tau1P3P neural network discriminant could be reached. This training was done separately with tau1P3P candidates with one associated track and more than one track.

It showed that the performance of the trained network was not very sensitive to the different training parameters, as long as the convergence of the training is guaranteed. But the necessary time to train a network was dependent on the applied learning procedure. The RPROP algorithm described in Section 3.3 proved to lead to shorter training times and better convergence than the Backpropagation algorithm in most cases, and was therefore used for further training. Also some standard parameters have been chosen for all further trainings. Some important of these where:

- a sigmoid activation function for all hidden nodes and the output node
- a quadratic error function
- an initial random Gaussian distribution of the connection weights with a width of 0.8
- an initial learning rate of 2 to speed up the early learning phase
- weight update parameters of the RPROP algorithm of  $\eta^+ = 1.2$  and  $\eta^- = 0.5$
- the input variables were scaled to have the bulk of the values to lie between  $-1$  and  $1$  for signal and background<sup>1</sup>

Also it was determined, that in order to get a consistently good training result it is important to chose a sensible fraction of the available data to be used as the validation set. If the validation set is too small, the fluctuations of the validation error are too large for the training program to find the training epoch with the minimal error, and the network is often over- or undertrained. If, on the other hand, the fraction of the training events used for the validation is too large, then the remaining events available for training the network may not be able to constrain the weights well enough during the training and the optimal performance is often not reached. For the available statistics, good results

---

<sup>1</sup>Scaling of the input variables avoids extremely large or small values for the weights, which can lead to numerical instabilities and longer training times.

could be achieved when using between a quarter to a third of the training samples for the validation set.

### 6.3.2 Training Categories and Data Preparation

For the training of the neural networks and the evaluation of their performance, about 290.000 reconstructed  $\tau$  candidates matched to true  $\tau$  jets as the signal and about 1.4 million reconstructed  $\tau$  candidates from QCD dijet events as the background where available from several productions done for validation purposes inside the ATLAS collaboration and a private production on a local cluster. Of these candidates around 80% of the signal and 40% of the background were overlap candidates. It should be noted, that a mix of samples reconstructed with different versions of ATHENA has been used, mostly from version 13.0.40 and version 14.0.0, but it was ensured that no significant changes to the reconstruction have been introduced between those versions. No samples were reconstructed twice from the same simulated data. Other than that, all samples are based on data which was simulated with the same version of the simulation code and the same detector geometry.

The neural network discriminant of the tau1P3P package uses four different categories for the  $\tau$  candidates as described in Section 4.4.3. It has not been demonstrated that this complex setup can significantly improve the identification performance. After some experimentation with the tau1P3P categories and a simple classification into candidates with one track and candidates with more than one track, it was decided for the present study to use the latter, since no significant difference in performance could be observed.

One third of the signal sample and two thirds of the background were set aside to be later used as a statistically independent sample to evaluate the performance of the trained neural networks. The rest of the data was split up into four different training sets:

- overlap candidates with exactly one associated track as defined by tau1P3P and at least one track as defined by tauRec
- overlap candidates with more than one and less than nine associated tracks as defined by tau1P3P and at least one track as defined by tauRec
- candidates reconstructed at least by tauRec with one associated associated track as defined by tauRec
- candidates reconstructed at least by tauRec with more than one and less than nine associated tracks as defined by tauRec.

Requiring at least one associated track as defined by tauRec for the overlap candidates makes sure that all tauRec variables are defined, and can be later used in the neural network discriminant. A negligible fraction of events is removed by this additional requirement.

These four training sets are used to train neural networks for each of the four respective categories. The network topology and the used subset of identification variables are opti-

## 6 $\tau$ Algorithm Merging

mised independently for each of the categories.

Since some of the most separating identification variables are strongly correlated with the transverse energy, it is highly desirable to include the  $E_T$  of the  $\tau$  candidate in the identification procedure in some way, especially if the goal is to offer a very generic  $\tau$  identification over a wide kinematic region. In the existing discriminants this is already done. The tau1P3P neural network discriminant explicitly uses the transverse energy calibrated by the energy flow approach as an input variable to the neural network, whereas the likelihood discriminant of the tauRec package does so implicitly by binning the probability density functions into different ranges of the transverse energy.

For a general purpose particle identification, like the  $\tau$  identification considered here, one must be careful not to introduce a bias towards the specific signature of the sample used for training the identification procedure. This is especially relevant when incorporating kinematic variables into the discriminant, since the kinematical properties will usually be the ones that show the largest variation between different physical processes. This means that the kinematic variables itself should have no discrimination power between signal and background candidates, but should only contribute through their correlations with other variables used in the identification procedure.

One way to achieve such a behaviour, is to modify the samples used for the training of the discrimination method in such a way as to have the same kinematic properties. At the very least those properties that are used as inputs to the identification need to have approximately the same distribution.

To equalise the distributions of a variable for the signal and background sample the following procedure was applied:

- scale the distributions for signal and background to an area of one
- take the ratio of signal to background in an appropriate number of bins
- optionally smooth the histogram of the ratio to avoid large discontinuities between adjacent bins
- loop over the events and look up the ratio  $r$  in the bin given by the variable
- draw a random number  $x$  between 0 and 1
- if it is a signal event, remove the event from the sample if  $r > 1$  and  $x > \frac{1}{r}$
- if it is a background event, remove the event from the sample if  $r < 1$  and  $x > r$

This procedure will be called *normalisation* in the following. Since events are removed from the sample this normalisation reduces the available statistics.

The normalisation was first applied to each of the four training samples on the transverse energy  $E_T$  as defined by tauRec. The samples resulting from this step were again normalised to have roughly the same distribution for the pseudorapidity  $\eta$ . It was checked

## 6 $\tau$ Algorithm Merging

that the normalisation of  $\eta$  did not affect the one applied on the transverse energy. The resulting distribution of the transverse energy and the pseudorapidity in the training samples containing only overlap candidates is shown in Figure 6.4. For the sample containing candidates with only one associated track not many candidates above about 100 GeV survive the normalisation, since only a very small fraction of QCD jets fall into this category at high energies. A similar effect can be observed for the tauRec training samples, but it is more pronounced due to the looser requirements on the quality of the associated tracks and a therefore even smaller fraction of QCD jets with only one associated track. The trained networks might therefore not develop the full potential for single track candidates with high transverse energies. Since the fraction of QCD jets with one associated track becomes very small for high energies this should not have a large effect on the overall rejection in real applications, since it is dominated by the rejection of QCD jets with a higher track multiplicity.

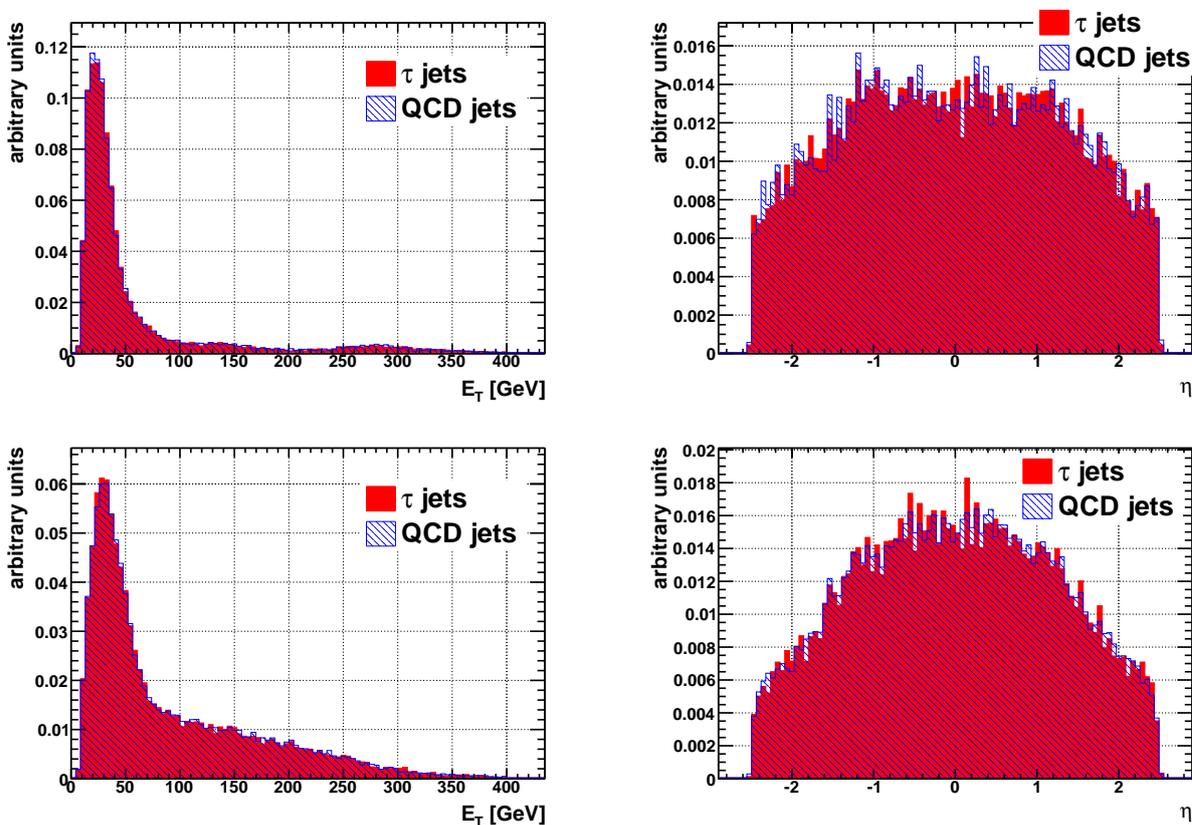


Figure 6.4: Comparison of the  $E_T$  (left column) and  $\eta$  (right column) distribution between  $\tau$  jets and QCD jets in the normalised training sample for overlap candidates. In the top row the candidates with only one associated track are shown. In the bottom row the ones with more than one associated track are shown.

The remaining number of  $\tau$  candidates in each category after normalising the training samples are listed in Table 6.2.

## 6 $\tau$ Algorithm Merging

category	no. of $\tau$ jets	no. of QCD jets
overlap, 1 track	31324	15612
overlap, >1 track	32443	84353
tauRec, 1 track	32850	28924
tauRec, >1 track	52606	141933

Table 6.2: Number of  $\tau$  jets and QCD jets in each training sample.

### 6.3.3 Optimisation for Overlap Candidates

The optimisation of the neural network discriminant for the overlap candidates was based on the configuration of the neural network discriminant used in the tau1P3P package, except for the smaller number of different classes of  $\tau$  candidates as mentioned above. The tau1P3P neural network discriminant offers a superior performance to the tauRec likelihood discriminant for very small transverse momenta below about 40 GeV, where the need for an improved rejection is most pressing. Therefore the tau1P3P neural network was considered to be a good basis for further optimisation.

It must be mentioned, that the tau1P3P neural network has been trained without applying any normalisation to the training sample as it was for example shown in Section 6.3.2. The exact composition of the training sample used is not known, but it consisted of  $\tau$  jets from  $Z \rightarrow \tau\tau$  and  $W \rightarrow \tau\nu_\tau$  events and QCD jets from QCD dijet events [26]. This seems to lead to a higher performance for transverse momenta below about 40 GeV at the cost of strongly distorting the  $p_T$  spectrum of identified  $\tau$  candidates. Also in the available samples, the neural network discriminant is not defined for  $\tau$  candidates with  $E_{T,eflow} > 100$  GeV. When comparing the performance of other discriminants to the tau1P3P neural network, therefore only  $\tau$  candidates with  $E_{T,eflow} \leq 100$  GeV are considered if no explicit dependency on the transverse energy or momentum is shown.

The eleven input variables used for the neural networks in the tau1P3P package are listed again in Table 6.3. Nine variables are the same for candidates with one associated track and those with more than one track. The other two variables in the multi track class, the transverse flight path significance and the width of the track system, are only defined for more than one track and are replaced by the transverse and longitudinal impact parameter significance in the networks used for single track candidates.

For some of the identification variables a very similarly defined variable is reconstructed by tauRec, which exploits the same property of  $\tau$  jets but uses for example a different selection of the cells from which it is calculated. The EM-Radius is an example for such a variable. Some of those variables will later be replaced by their tauRec counterparts to determine if the latter offer a better discrimination.

One additional variable that has been devised, combining information from tauRec and tau1P3P, is the ratio of the transverse energy reconstructed by the energy flow approach in the tau1P3P to the transverse energy reconstructed purely from the calorimeter as in

## 6 $\tau$ Algorithm Merging

1 track candidates	multi track candidates
E <sub>T</sub> reconstructed by energy flow	
EM-Radius	
Isolation fraction	
Isolation ratio	
$\eta$ -strip width	
Hits in the $\eta$ -strip layer	
Hadronic E <sub>T</sub> over sum of $p_{T,tracks}$	
Invariant mass between tracks and EM cluster	
Number of isolation tracks	
Transverse impact parameter significance	Transverse flight path
Longitudinal impact parameter significance	Width of track system

Table 6.3: The list of variables which are used as input for the tau1P3P neural network discriminant. The first nine variables are used both for one and multi track  $\tau$  candidates.

tauRec. This variable will be referred to as ‘E<sub>T</sub> ratio’. The reason why the E<sub>T</sub> ratio offers any discrimination power lies in the fact, that the energy flow algorithm is optimised for  $\tau$  jets and therefore severely underestimates the energy of QCD jets which have a different composition and shower shape than  $\tau$  jets. For low transverse energies this effect is enhanced, because only tracks and calorimeter cells in a cone of radius 0.2 are taken into account by the energy flow algorithm. QCD jets with low transverse energies, which are very broad on average, often deposit energy outside of this cone. At higher energies more and more of the energy of QCD jets is carried by neutral hadrons. Their contributions to the total jet energy are discarded by the energy flow approach, since the energy measured in the hadronic calorimeter is not included in the reconstruction.

The distributions of the E<sub>T</sub> ratio for  $\tau$  jets and QCD jets in the training samples are shown in Figure 6.5 for candidates with true transverse momentum between 15 GeV and 100 GeV. One can see that this variable provides quite some separation if considered in isolation, but it can be expected that underestimation of the energy is much smaller for QCD jets that have a signature which is more similar to a  $\tau$  jet than the average QCD jet. This leads to strong correlations of the E<sub>T</sub> ratio with other identification variables.

## 6 $\tau$ Algorithm Merging

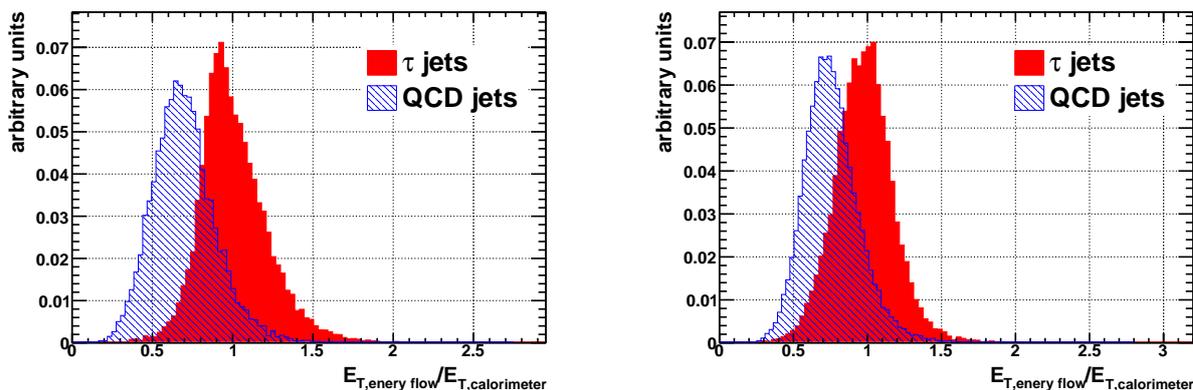


Figure 6.5: Distribution of the  $E_T$  ratio in the training samples for single track candidates on the left and multi track candidates on the right for candidates with a true  $p_T$  between 15 GeV and 100 GeV.

### Single Track Candidates

First the optimisation of the neural network for overlap candidates in the single track category will be discussed. In Figure 6.6 the rejection against QCD jets is shown as function of the efficiency for  $\tau$  jets for several neural networks with one hidden layer and different numbers of hidden nodes. Contrary to plots shown before, the efficiency and the rejection are calculated with respect to the number of reconstructed overlap signal and background candidates with one associated track. Those candidates additionally have to be matched to a  $\tau$  jet or a *Cone4TruthJet* respectively, which are required to have a true transverse momentum between 12 GeV and 100 GeV and an absolute pseudorapidity of less than 2.5.

To relate the efficiencies and rejections shown here to the efficiencies and rejections shown in other plots before, one needs to take into account the reconstruction efficiencies for  $\tau$  jets and QCD jets as well as the probabilities to reconstruct a candidate with one associated track. It should therefore be kept in mind, that the numbers shown here cannot be directly compared to the ones shown in earlier plots. Nevertheless the numbers for different networks are calculated from the same sample, so it is still valid to compare the relative performance between those networks.

The performance of the neural network for single track candidates with one hidden layer depends little on the number of hidden nodes at high efficiencies above about 70%, but for lower efficiencies the rejection rises when the number of hidden nodes is raised from 20 to 30. When increasing the number of nodes further, no improvement is observed. In Figure 6.7 the same networks are compared for a fixed signal efficiency of 50% at different  $p_T$  bins. The rejections for the tau1P3P neural network discriminant and the tauRec likelihood are shown for comparison. The conclusion on the number of hidden nodes also holds when comparing individual  $p_T$  bins.

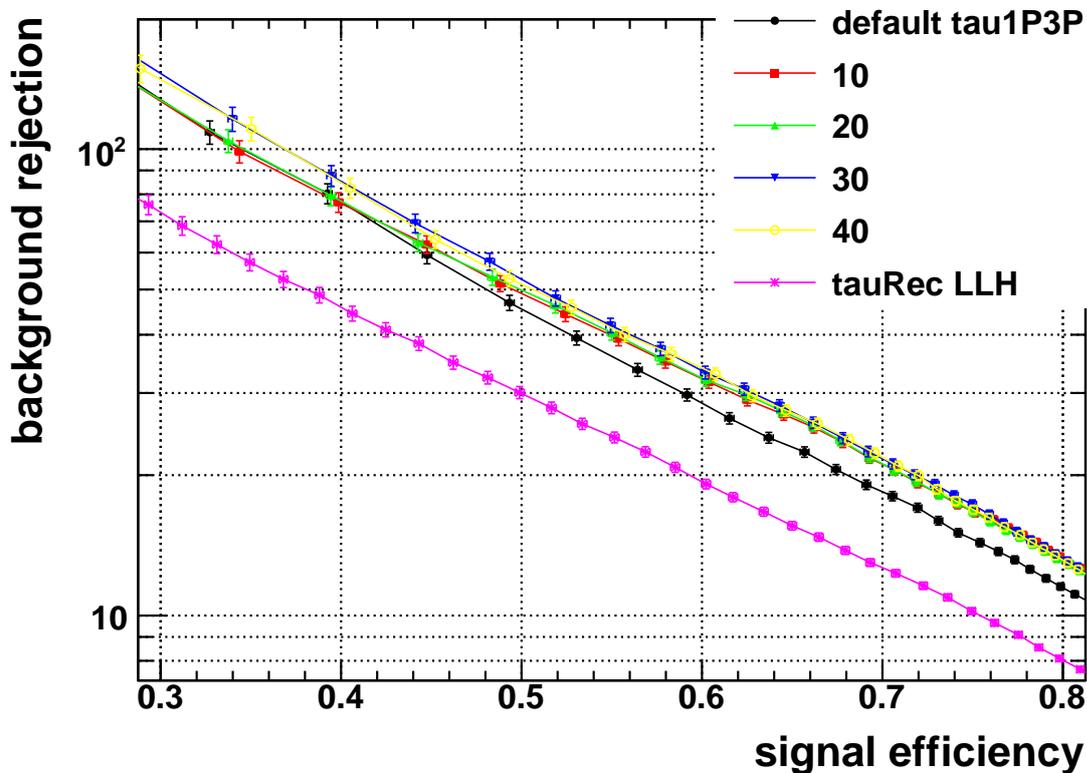


Figure 6.6: Rejection against QCD jets as function of the signal efficiency for overlap candidates with one associated track and a true transverse momentum between 12 GeV and 100 GeV. Different network topologies with one hidden layer are compared. The legend denotes the number of hidden nodes. Also the rejection of the default tau1P3P network and the tauRec likelihood are shown for comparison. Note that, contrary to plots shown before, the efficiencies and the rejections are calculated with respect to reconstructed  $\tau$  candidates.

In Figure 6.7 one can also see the effect of the training with  $E_T$  normalised samples compared to the tau1P3P network, which uses the same input variables. The tau1P3P network has a much better performance at small transverse momenta below about 35 GeV, but quickly loses rejection power for higher momenta, whereas the newly trained neural networks and the tauRec likelihood stay at decent level up to transverse momenta of 400 GeV.

Also networks with two hidden layers were tested for their performance. In Figure 6.8 the rejection against QCD jets is shown as function of the signal efficiency for some neural network topologies with two hidden layers. The best performing neural network with one hidden layer, having 30 hidden nodes, is shown for comparison. The performance could not be improved any further by adding a second hidden layer.

## 6 $\tau$ Algorithm Merging

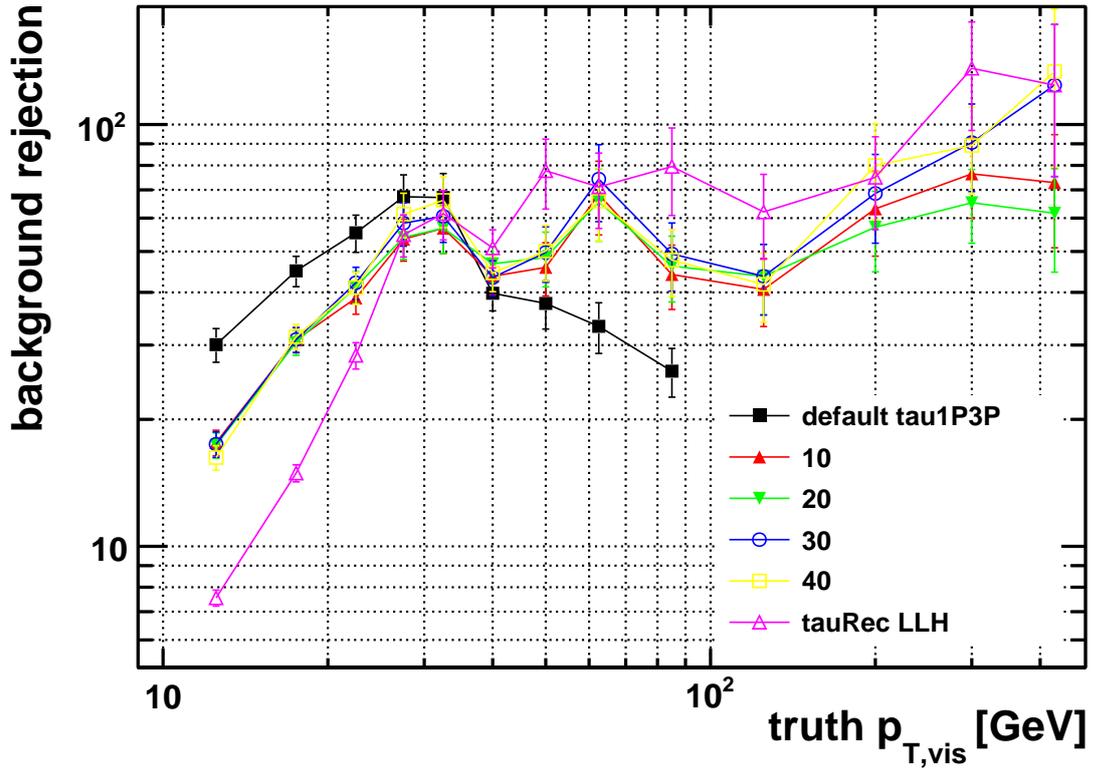


Figure 6.7: Rejection against QCD jets as function of the true transverse momentum at a signal efficiency of 50%. Different network topologies with one hidden layer are shown. The efficiency and rejection are calculated with respect to reconstructed overlap candidates with one associated track. The rejection of the default tau1P3P network and the tauRec likelihood are shown for comparison.

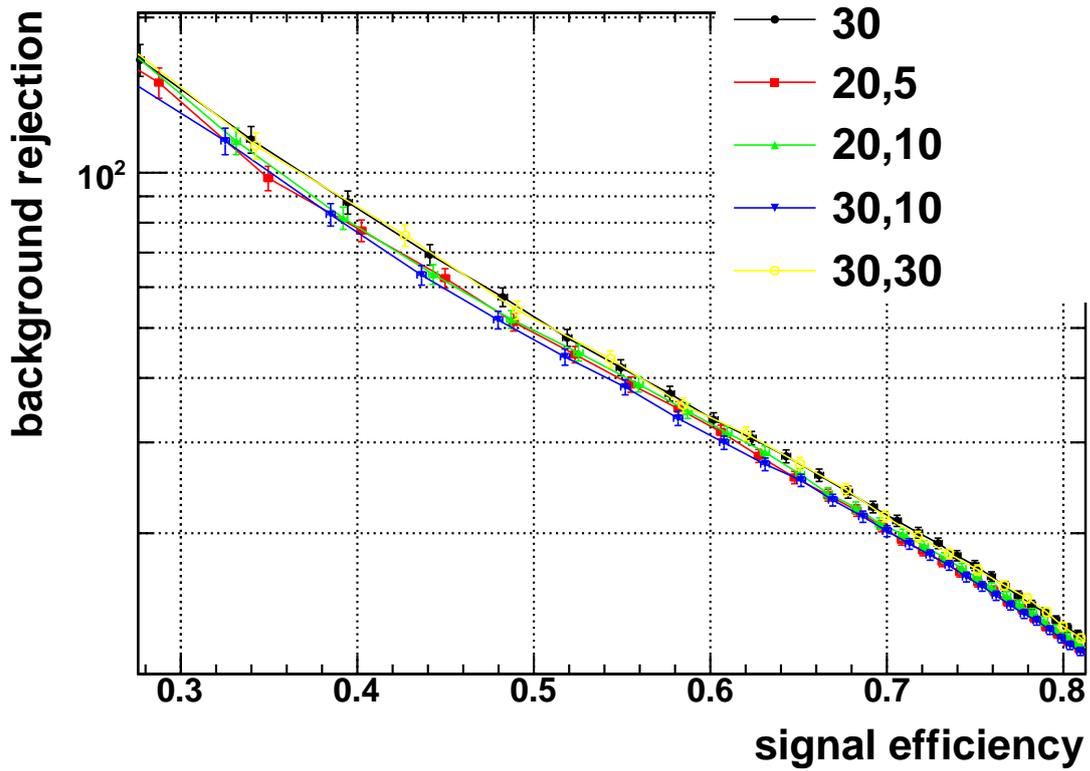


Figure 6.8: Rejection against QCD jets as function of the efficiency for real  $\tau$  jets for overlap candidates with one associated track and  $12 \text{ GeV} < p_{T,vis} < 100 \text{ GeV}$ . Different network topologies with two hidden layers are compared. The legend denotes the number of hidden nodes in the first and the second hidden layer. Also the rejection of the best performing network with one hidden layer (30) is shown for comparison. The efficiency and the rejection is calculated with respect to reconstructed  $\tau$  candidates.

## 6 $\tau$ Algorithm Merging

For the following comparison of different set of input variables for the category of single track overlap candidates the network topology was fixed to one hidden layer with 30 hidden nodes.

The transverse energy reconstructed with the energy flow algorithm, which is used as a variable in the neural networks shown so far, has the disadvantageous property of possessing complicated correlations with other properties of the  $\tau$  candidate. These are expected to be quite different for  $\tau$  jets and QCD jets and also between jets initiated by quarks and jets initiated by gluons. This makes the  $E_T$  defined by the energy flow algorithm sensitive to the differences between quark jets and gluon jets. Therefore a certain amount of discrimination on the specific kinematic composition of the training sample is indirectly trained into the network, even when normalising the sample with respect to  $E_{T,\text{eflow}}$ . It was therefore deemed more natural to use the transverse energy as reconstructed in the calorimeter as an input variable to the network instead, since it is much less sensitive to the exact jet composition. The discrimination power offered by the energy flow approach can later be reintroduced via the  $E_T$  ratio defined before, which is only weakly correlated with the true transverse energy.

In the following therefore a network where the transverse energy defined by tau1P3P is replaced by the transverse energy as defined by tauRec is used as the basis for further optimisation. This leads to a small loss in performance compared to the networks considered above, which is expected, since the training sample was not normalised with respect to the transverse energy from tau1P3P.

On top of that, the following changes were made to the set of input variables used by this base network

- adding the number of  $\pi^0$  clusters from tau1P3P (denoted by +num  $\pi^0$  in plots),
- adding the  $E_T$  ratio (denoted by + $E_{T,\text{Eflow}}/E_T$  in plots) and
- replacing  $E_{T,\text{had}}/\sum_{\text{tracks}} p_T$  from tau1P3P by  $E_T/p_{T,\text{leading}}$  from tauRec (denoted by + $E_T/p_T$  in plots),

as well as combinations thereof. The rejections as function of the signal efficiency for some combinations of the above variations are shown in Figure 6.9. The default tau1P3P neural network is shown as a reference, as well as the base network where the tau1P3P  $E_{T,\text{eflow}}$  variable has been replaced by the tauRec  $E_T$ . The latter is denoted by ‘base’ in the legend of Figure 6.9.

The rejection can be improved by adding each of the above three variables alone, but the biggest improvement can be observed when adding the ratio between the transverse energy and the transverse momentum of the leading track. Since this variable is sensitive to the the presence of neutral particles one can expect a strong correlation with the number of  $\pi^0$  clusters. In fact one can observe a further improvement of the same magnitude in the performance when adding both  $E_T/p_{T,\text{leading}}$  and the number of  $\pi^0$  clusters together. Adding also the  $E_T$  ratio actually diminishes the performance. This can happen when a variable is added to a neural network which is very strongly correlated with the existing

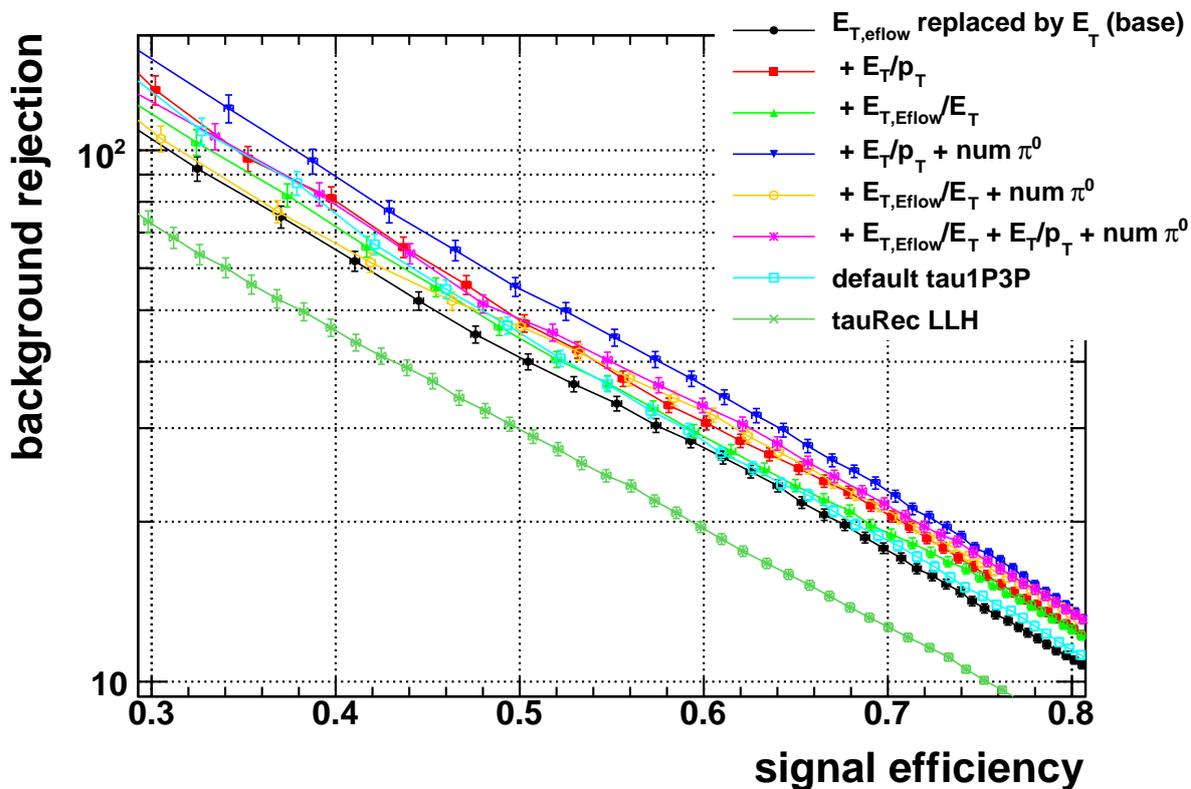


Figure 6.9: Rejection against QCD jets as function of the efficiency for real  $\tau$  jets for overlap candidates with one associated track and  $12 \text{ GeV} < p_{T,vis} < 100 \text{ GeV}$ . Different combinations of input variables are compared. In the legend the addition of variables is denoted by a plus sign. See the text for a more detailed description. The efficiency and rejection are calculated with respect to reconstructed overlap candidates with one associated track.

input variables. Such a variable essentially adds no new information but the dimensionality of the problem is enlarged, making it harder for the neural network to discriminate signal an background due to the ‘curse of dimensionality’.<sup>2</sup> The combination of  $E_T/p_{T,leading}$  and the number of  $\pi^0$  clusters clearly offers the best performance.

In addition to the changes just shown, also four shower shape variables where replaced by their counterparts in the tauRec package which exploit the same properties of  $\tau$  jets but are defined slightly different than in the tau1P3P package. These four variables are the EM-Radius, the isolation fraction, the  $\eta$  strip width and the number of  $\eta$  strip hits. All four variables differ between tauRec and tau1P3P only in the list of calorimeter cells that are used for calculating them. In the tau1P3P package only cells within a cone of radius 0.2 around the  $\tau$  direction are used, whereas in the tauRec package a cone of radius 0.4 is used. One would expect a better discrimination if a larger radius is used, because  $\tau$

<sup>2</sup>The curse of dimensionality is explained in Section 3.2.

## 6 $\tau$ Algorithm Merging

jets are nearly always contained within a cone of radius 0.2, while QCD jets deposit also energy outside this cone. Therefore the identification variables would be unaffected for  $\tau$  jets but are shifted to larger values for QCD jets when using a larger cone size.

Another difference between the shower shape variables of the two algorithms arises, because the tauRec package was already based on topological clusters, which means the cuts on the significance of the cell energies differ between the two algorithms. As has been shown in Chapter 5, the topological clustering improves the separation power of the shower shape variables.

In Figure 6.10 the rejection for several combinations of input variables when also replacing the four shower shape by their tauRec equivalents ( +tauRec shape ) are shown as function of the signal efficiency. For comparison the base network without additional variables and the best network with shower shape variables taken from tau1P3P ( + $E_T/p_T$  +num  $\pi^0$  ) are also shown. Using the tauRec shower shape variables can improve the rejection of the previous best network by roughly another 10% for signal efficiencies of more than 40%. For lower efficiencies it does not add much to the performance.

In Figure 6.11, where the rejection for different combinations of input variables is shown at a signal efficiency of 50% as function of  $p_T$ , one observes an improvement by adding  $E_T/p_{T,\text{leading}}$  to the input variables or using tauRec shower shape variables only for transverse momenta below about 50 GeV. Combining both of those on other hand can improve the rejection power over the whole range momenta. The addition of the number of  $\pi^0$  clusters can improve the rejection by roughly another 10% in the  $p_T$  region below 25 GeV at the cost of reducing the rejection around 100 GeV by a similar factor. Since ratio of QCD jets with one associated track is extremely small at 100 GeV the loss of performance in this region is outweighed by the better rejection at very low transverse momenta.

The best configuration of the input variables for the category of overlap candidates with one associated track uses therefore the following input variables:

- $E_T$  defined by tauRec
- EM-Radius, calorimeter isolation fraction,  $\eta$ -strip width and number of  $\eta$ -strip hits defined by tauRec
- Isolation ratio
- $E_T/p_{T,\text{leading}}$
- number of  $\pi^0$  clusters
- invariant mass between track and EM cluster
- number of isolation tracks
- unsigned transverse and longitudinal impact parameter

In Figure 6.12 the performance of this network is compared to the default discriminants of tau1P3P and tauRec and to the base network where the transverse energy as defined

## 6 $\tau$ Algorithm Merging

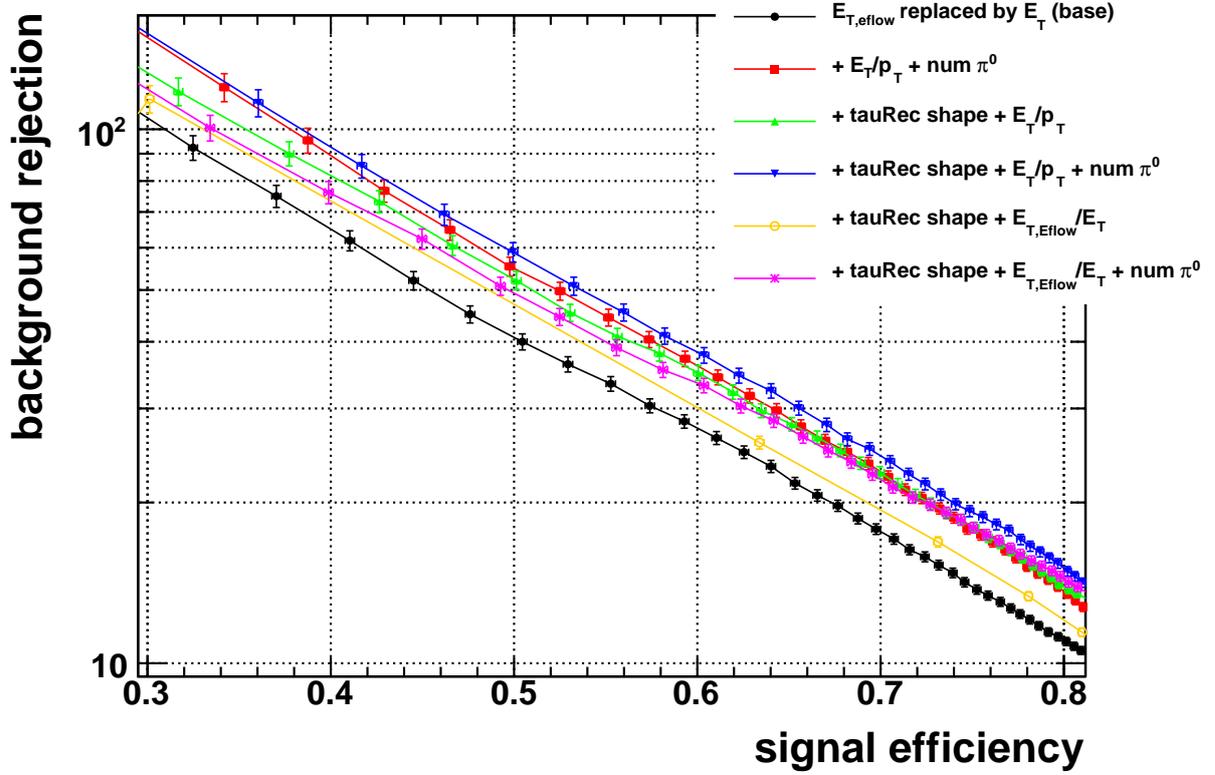


Figure 6.10: Rejection against QCD jets as function of the efficiency for real  $\tau$  jets for overlap candidates with one associated track and  $12 \text{ GeV} < p_{T,vis} < 100 \text{ GeV}$ . Different combinations of input variables are compared. In the legend the addition of variables is denoted by a plus sign, and ‘+tauRec shape’ denotes replaced variables. See the text for a more detailed description. The efficiency and rejection are calculated with respect to reconstructed overlap candidates with one associated track.

by tauRec is used. The rejection at an  $\tau$  efficiency of 50% is shown in bins of the true transverse momentum. The rejections are also listed in Table 6.4. The new network offers the best rejection of the compared discriminants for transverse momenta above 25 GeV.

## 6 $\tau$ Algorithm Merging

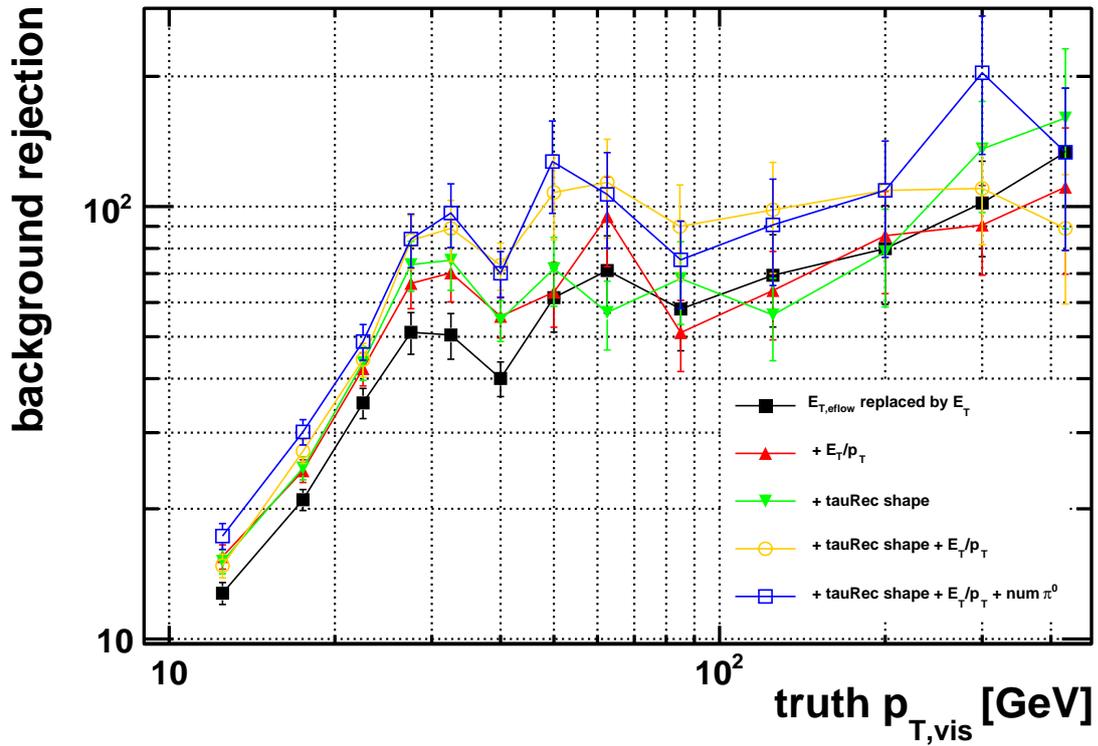


Figure 6.11: Rejection against QCD jets as function of the true transverse momentum at a signal efficiency of 50%. Different combinations of input variables are shown. The efficiency and rejection are calculated with respect to reconstructed overlap candidates with one associated track.

## 6 $\tau$ Algorithm Merging

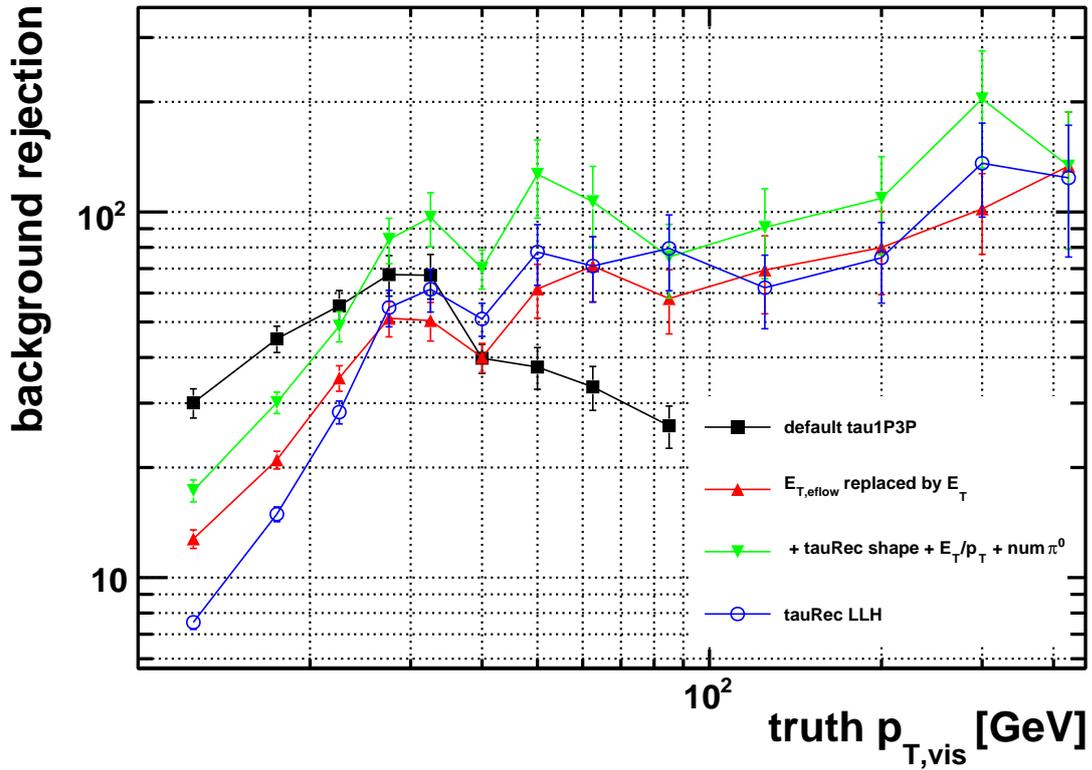


Figure 6.12: Rejection against QCD jets as function of the true transverse momentum at a signal efficiency of 50%. Shown are the default discriminant of tau1P3P and tauRec, the base network and the network with the best performing combination of input variables. The efficiency and rejection are calculated with respect to reconstructed overlap candidates with one associated track.

## 6 $\tau$ Algorithm Merging

discriminant	default tau1P3P	default tauRec	base network	best configuration
<i>true <math>p_{T,vis}</math></i>				
12 – 15 GeV	$30.1 \pm 2.8$	$7.5 \pm 0.3$	$12.8 \pm 0.7$	$17.3 \pm 1.2$
15 – 20 GeV	$45.0 \pm 3.8$	$14.9 \pm 0.7$	$21.0 \pm 1.2$	$30.1 \pm 2.0$
20 – 25 GeV	$55.4 \pm 5.6$	$28.4 \pm 2.0$	$35.2 \pm 2.8$	$48.7 \pm 4.6$
25 – 30 GeV	$67.4 \pm 8.6$	$54.8 \pm 6.3$	$51.2 \pm 5.7$	$84 \pm 12$
30 – 35 GeV	$67.1 \pm 9.4$	$61.5 \pm 8.2$	$50.5 \pm 6.1$	$97 \pm 16$
35 – 45 GeV	$39.8 \pm 3.6$	$51.0 \pm 5.3$	$40.0 \pm 3.7$	$70 \pm 9$
45 – 55 GeV	$37.7 \pm 4.9$	$78 \pm 15$	$62 \pm 10$	$127 \pm 31$
55 – 70 GeV	$33.2 \pm 4.6$	$71 \pm 14$	$71 \pm 14$	$107 \pm 27$
70 – 100 GeV	$26.0 \pm 3.4$	$80 \pm 19$	$58 \pm 12$	$75 \pm 17$
100 – 150 GeV	-	$62 \pm 14$	$69 \pm 17$	$91 \pm 25$
150 – 250 GeV	-	$75 \pm 19$	$80 \pm 21$	$109 \pm 33$

Table 6.4: Rejections at an efficiency of 50% for true  $\tau$  jets for the default discriminants of tau1P3P and tauRec, the base network and the optimal network. The efficiency and rejection are calculated with respect to reconstructed overlap candidates with one associated track.

### Multi Track Candidates

For the multi track category the same procedure was used to find the optimal network configuration as in the single track category. First the impact of different network topologies was checked with the input variables as used by the default neural network in the tau1P3P package given in Table 6.3. One fixed network topology was then chosen to optimise the list of input variables used for the neural network.

Figure 6.13 shows the rejections for a number of different network topologies as function of the signal efficiency for  $\tau$  candidates with a true transverse momentum between 12 GeV and 100 GeV. The performance of the different network topologies is quite similar as long as there are more than ten nodes in the first hidden layer. Also no significant improvement over the default tau1P3P network can be observed when averaging over the efficiency. A network topology with two hidden layers with 20 and 10 hidden nodes respectively has been chosen for further optimisation, since it provides a slightly better performance at transverse momenta above 100 GeV.

Like in the single track category, the transverse energy calculated by the energy flow algorithm was replaced by the transverse energy reconstructed in the calorimeter for all networks for reasons given above. A neural network with just this modification, again called base network, will be used as a reference. The different changes to the list of input variables that were considered are:

- adding the  $E_T$  ratio (denoted by  $+E_{T,\text{Eflow}}/E_T$  in plots),
- replacing  $E_{T,\text{had}}/\sum_{\text{tracks}} p_T$  from tau1P3P by  $E_T/p_{T,\text{leading}}$  from tauRec (denoted by  $+E_T/p_T$  in plots)
- adding the invariant mass of the track system (+inv track mass in plots)
- adding the absolute value of the charge ( $+|\text{charge}|$  in plots)
- adding the number of associated tracks as defined by tauRec (+no of tracks in plots)
- replacing the four shower shape variables EM-Radius, calorimeter isolation fraction,  $\eta$ -strip width and the number of  $\eta$ -strip hits with the same variables from the tauRec package (+tauRec shape in plots)

Various combinations of these changes were again evaluated.

In Figure 6.14 the performance for most combinations of the first three modifications are shown as function of the  $\tau$  efficiency for true transverse momenta between 12 GeV and 100 GeV. The largest improvement can be observed when adding the  $E_T$  ratio to the input variables. The performance can be further improved when the invariant track mass is also added to the list of variables. The rejection decreases slightly for low efficiencies when the  $E_T/p_{T,\text{leading}}$  variable is also introduced, but rises for high efficiencies. Typical efficiency working points for  $\tau$  candidates with more than one track will be rather low, since QCD jets most of the time have more than one track at the typical transverse momenta for  $\tau$  jets. In Figure 6.15 some configurations which include also the last three modifications of the

## 6 $\tau$ Algorithm Merging

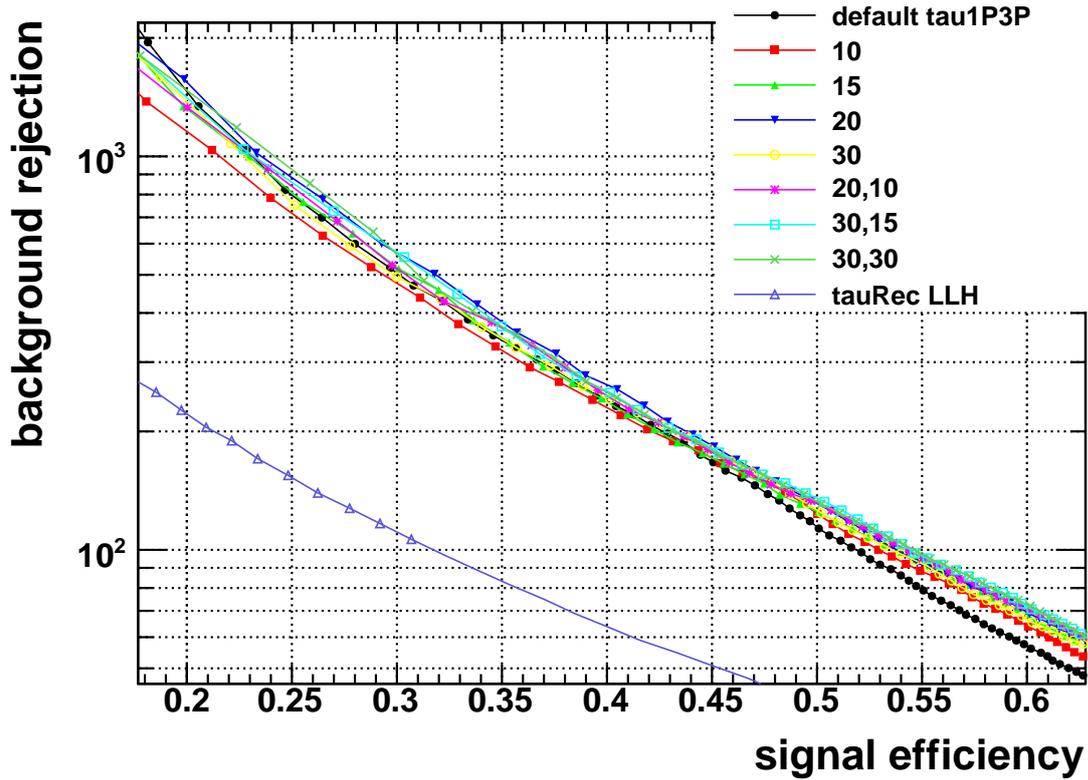


Figure 6.13: Rejection against QCD jets as function of the efficiency for real  $\tau$  jets for overlap candidates with more than one associated track and  $12 \text{ GeV} < p_{T,vis} < 100 \text{ GeV}$ . Different network topologies with one and two hidden layers are compared. The legend denotes the number of hidden nodes in the first and the second hidden layer. Also the rejection of the default discriminants are shown. The efficiency and the rejection is calculated with respect to reconstructed  $\tau$  candidates. The error bars are suppressed.

## 6 $\tau$ Algorithm Merging

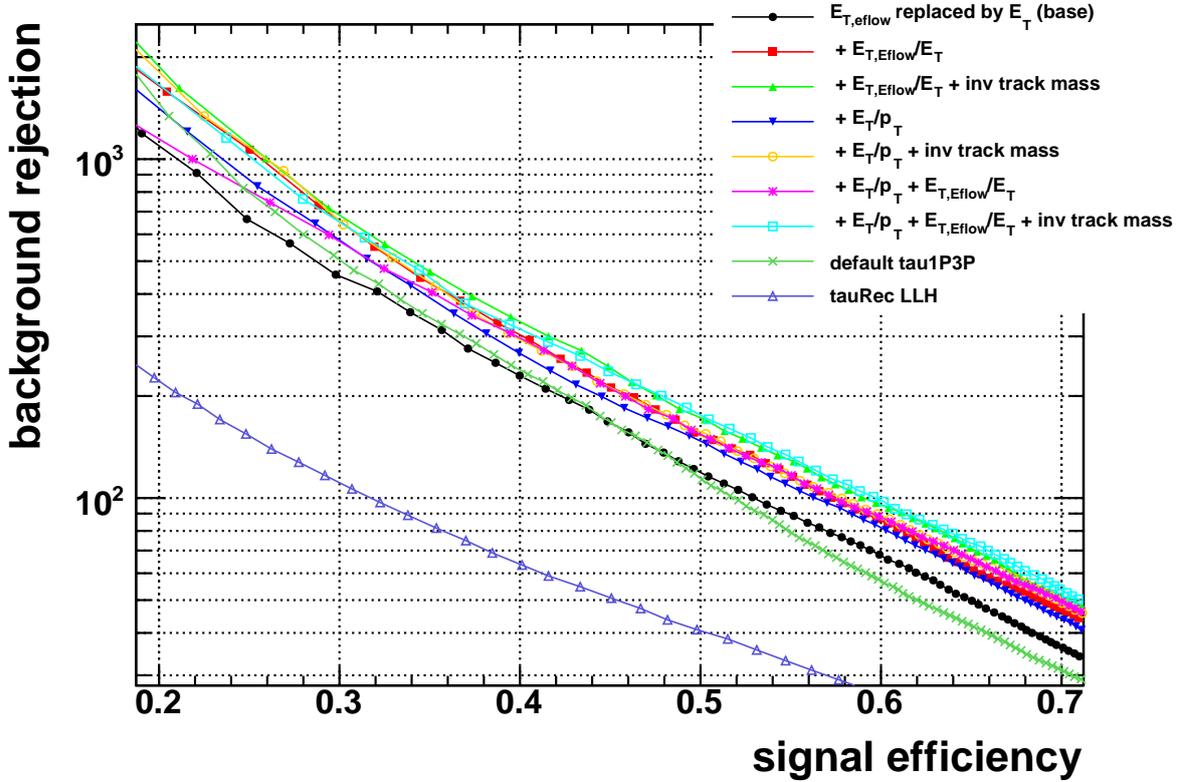


Figure 6.14: Rejection against QCD jets as function of the efficiency for real  $\tau$  jets for overlap candidates with more than one associated track. Different combinations of input variables are shown. The efficiency and rejection are calculated with respect to reconstructed overlap candidates with more than one associated track. The error bars are suppressed.

above list are shown. The network using the  $E_T$  ratio and the invariant track mass as input variables, which was one of the best so far, is shown as a reference in dark blue. Although the tauRec shower shape variables decrease the rejection when included on their own, they boost the performance if they are used in combination with the invariant track mass. Also, the tauRec shower shape variables provide a better rejection for large transverse momenta above 60 GeV, as can be seen in Figure 6.16. Another large gain in performance can be achieved by adding the absolute charge, providing a better separation for large efficiencies, and the number of associated tracks as defined by tauRec, which improves the performance for small efficiencies. Also the number of tracks as defined by tau1P3P has been tested as a variable, but the performance was much worse for low efficiencies. Altogether the rejection can be increased by a factor of about two compared to the base network.

The rejection for some configurations, successively adding or replacing more variables, is shown for a signal efficiency of 50% as function of the transverse momentum in Figure 6.16. The improvement of the rejection is quite uniform over all  $p_T$  bins for the best performing network.

## 6 $\tau$ Algorithm Merging

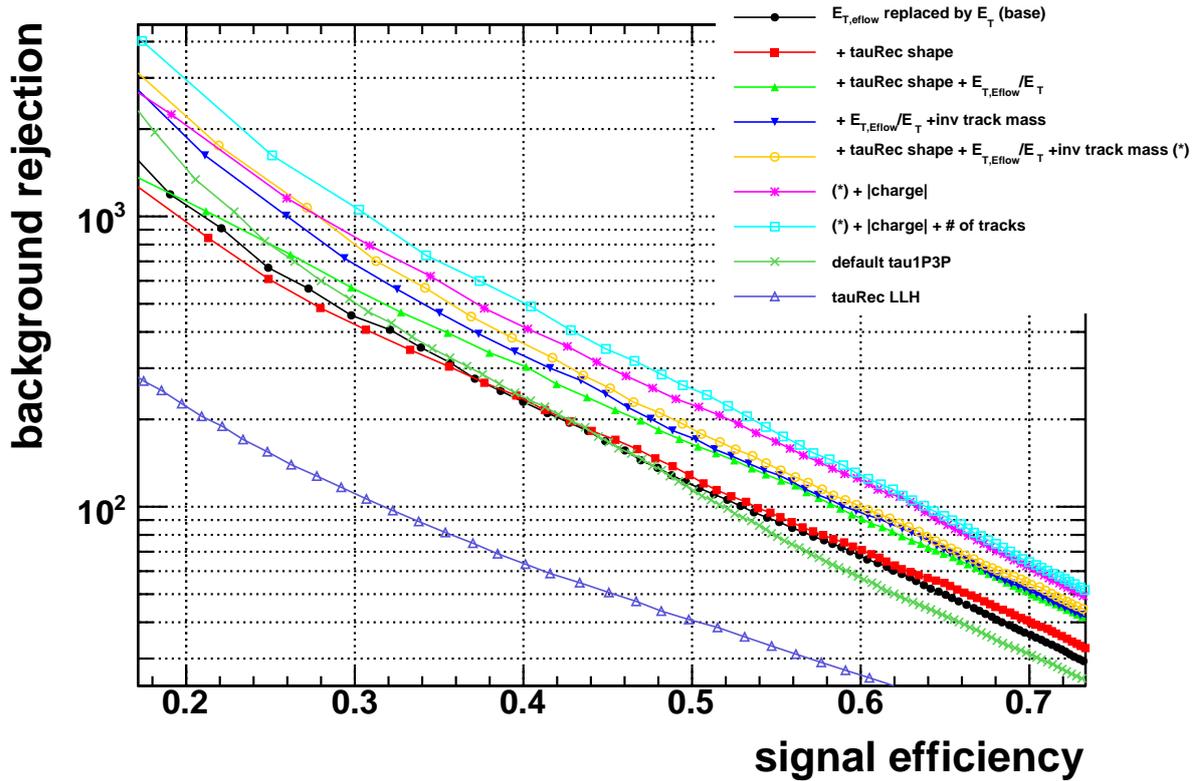


Figure 6.15: Rejection against QCD jets as function of the efficiency for real  $\tau$  jets for overlap candidates with more than one associated track. Different combinations of input variables are shown. The efficiency and rejection are calculated with respect to reconstructed overlap candidates with more than one associated track. The error bars are suppressed.

## 6 $\tau$ Algorithm Merging

The neural network with the strongest discrimination power of the tested configurations has the following input variables:

- $E_T$  defined by tauRec
- EM-Radius, calorimeter isolation fraction,  $\eta$ -strip width and number of  $\eta$ -strip hits defined by tauRec
- Isolation ratio
- $E_{T,\text{had}} / \sum_{\text{tracks}} p_T$
- invariant mass between track and EM cluster
- number of isolation tracks
- lifetime signed transverse flight path significance
- $E_T$  ratio
- invariant mass of the track system defined by tau1P3P
- absolute value of the charge defined by tau1P3P
- number of associated tracks defined by tauRec

In Figure 6.17 the performance of this network is compared to the default discriminants of tau1P3P and tauRec and to the base network where the transverse energy as defined by tauRec is used. The background rejection at a  $\tau$  efficiency of 50% is shown in bins of the true transverse momentum. The rejections are also listed in Table 6.5. The best new network outperforms the default tau1P3P neural network and the base network by a factor of more than two for transverse momenta above 20 GeV and by more than 20% for lower transverse momenta.

## 6 $\tau$ Algorithm Merging

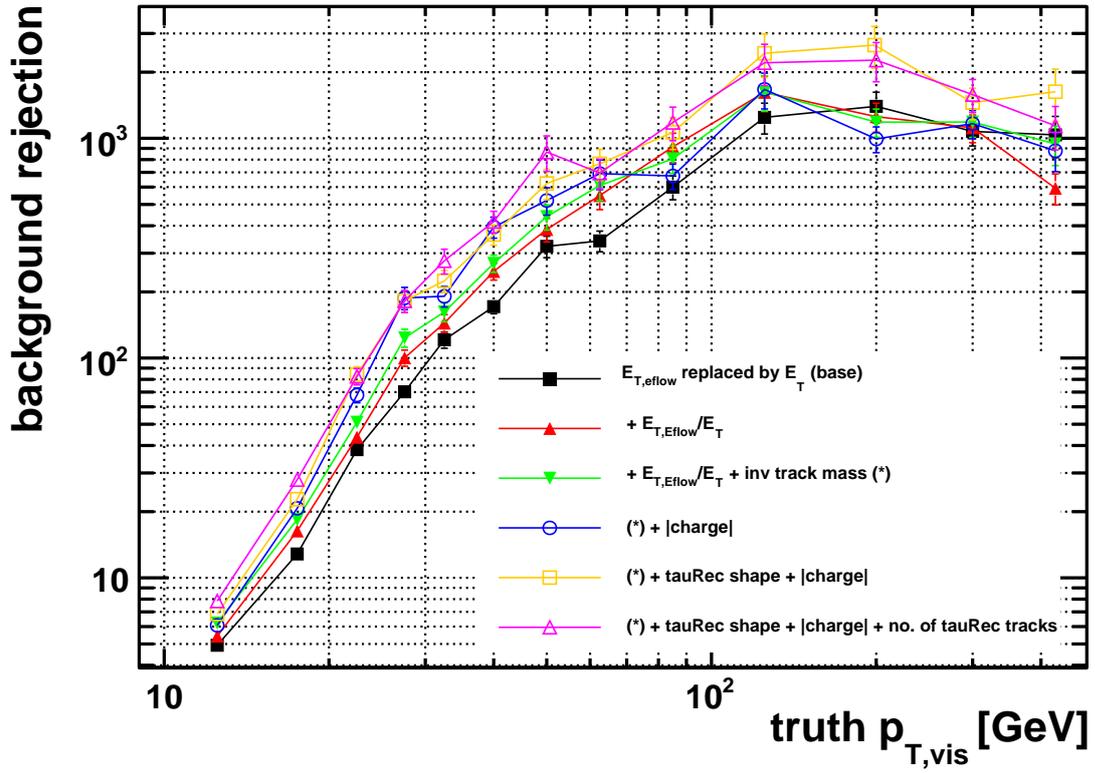


Figure 6.16: Rejection against QCD jets as function of the true transverse momentum at a signal efficiency of 50%. Different network with successively added variables are shown. The efficiency and rejection are calculated with respect to reconstructed overlap candidates with more than one associated track.

## 6 $\tau$ Algorithm Merging

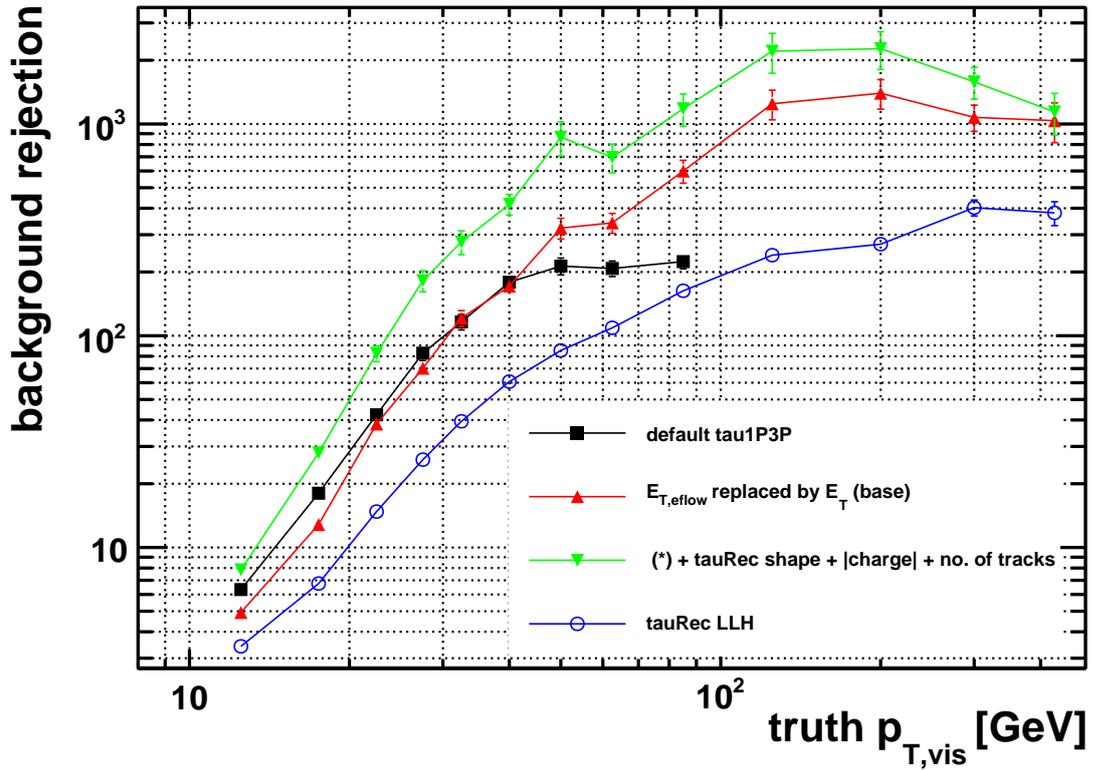


Figure 6.17: Rejection against QCD jets as function of the true transverse momentum at a signal efficiency of 50%. Shown are the default discriminant of tau1P3P and tauRec, the base network and the network with the best performing combination of input variables. The efficiency and rejection are calculated with respect to reconstructed overlap candidates with more than one associated track.

## 6 $\tau$ Algorithm Merging

discriminant	default tau1P3P	default tauRec	base network	best configuration
true $p_{T,vis}$				
12 – 15 GeV	$6.3 \pm 0.3$	$3.4 \pm 0.1$	$4.9 \pm 0.2$	$7.8 \pm 0.4$
15 – 20 GeV	$18.0 \pm 0.8$	$6.8 \pm 0.2$	$12.8 \pm 0.5$	$28.0 \pm 1.6$
20 – 25 GeV	$42.3 \pm 2.6$	$14.8 \pm 0.5$	$38.3 \pm 2.2$	$82.5 \pm 7.0$
25 – 30 GeV	$82.7 \pm 6.4$	$26.0 \pm 1.1$	$70.2 \pm 5.0$	$182 \pm 20.8$
30 – 35 GeV	$116 \pm 10$	$39.4 \pm 1.9$	$121 \pm 10$	$277 \pm 35.7$
35 – 45 GeV	$179 \pm 13$	$60.7 \pm 2.6$	$171 \pm 12$	$418 \pm 47.2$
45 – 55 GeV	$213 \pm 19$	$85.2 \pm 4.9$	$322 \pm 36$	$866 \pm 159$
55 – 70 GeV	$208 \pm 18$	$109 \pm 6.6$	$341 \pm 37$	$695 \pm 107$
70 – 100 GeV	$224 \pm 17$	$163 \pm 11$	$599 \pm 74$	$1181 \pm 205$
100 – 150 GeV	-	$240 \pm 17.$	$1245 \pm 199$	$2208 \pm 471$
150 – 250 GeV	-	$271 \pm 19$	$1398 \pm 222$	$2268 \pm 460$

Table 6.5: Rejections at an efficiency of 50% for true  $\tau$  jets for the default discriminants of tau1P3P and tauRec, the base network and the optimal network. The efficiency and rejection are calculated with respect to reconstructed overlap candidates with more than one associated track.

### 6.3.4 Optimisation for tauRec Candidates

In the category for tauRec candidates no optimisation of the input variables was performed, since no new variables became available by the merging of the algorithms. Further optimisation is surely possible but is outside the scope of this study. The neural networks in this category are trained mainly to have a discriminant value with a similar distribution to the ones for overlap candidates, so that one discriminant value can be used to cut on when using overlap candidates and also candidates reconstructed only by tauRec. The higher  $\tau$  reconstruction efficiency when using candidates from both categories allows to use a tighter cut on the discriminant value for the same overall signal efficiency. At the same time a larger fraction of QCD jets is reconstructed. It remains to be seen, if the tighter identification criterion can outweigh the larger background efficiency.

#### Single Track Candidates

The input variables used for the neural networks in the single track category were taken to be the same as the variables used for the likelihood discriminant in the tauRec package:

- EM-Radius
- Isolation fraction
- $\eta$ -strip width
- number of  $\eta$ -strip hits
- $E_T/p_{T,\text{leading}}$
- lifetime signed transverse impact parameter significance
- $E_T$  from the calorimeter

The transverse energy has been added to account for correlations with the transverse boost of the  $\tau$  candidate. The number of associated tracks and the absolute charge were not used, since those variables always have a value of one in this category with one associated track.

Neural networks with different topologies were trained to compare their relative performance. The rejection achieved by these networks is shown in Figure 6.18 as function of the true transverse momentum for a  $\tau$  jet efficiency of 50%. Only some prominent configurations are shown. The performance of the networks with one hidden layer is quite similar, provided that at least ten hidden nodes are used. By adding a second hidden layer the performance could be further improved for transverse momenta above 25 GeV by about 10%. Only the best of those networks is shown, which has two hidden layers with 20 and 10 nodes, respectively. This network topology was used in the following for the category of single track tauRec candidates. In general, the performance can be improved slightly by 20-50% for transverse momenta below 100 GeV compared to the tauRec likelihood. For higher momenta the number of background events was too small to draw any conclusions.

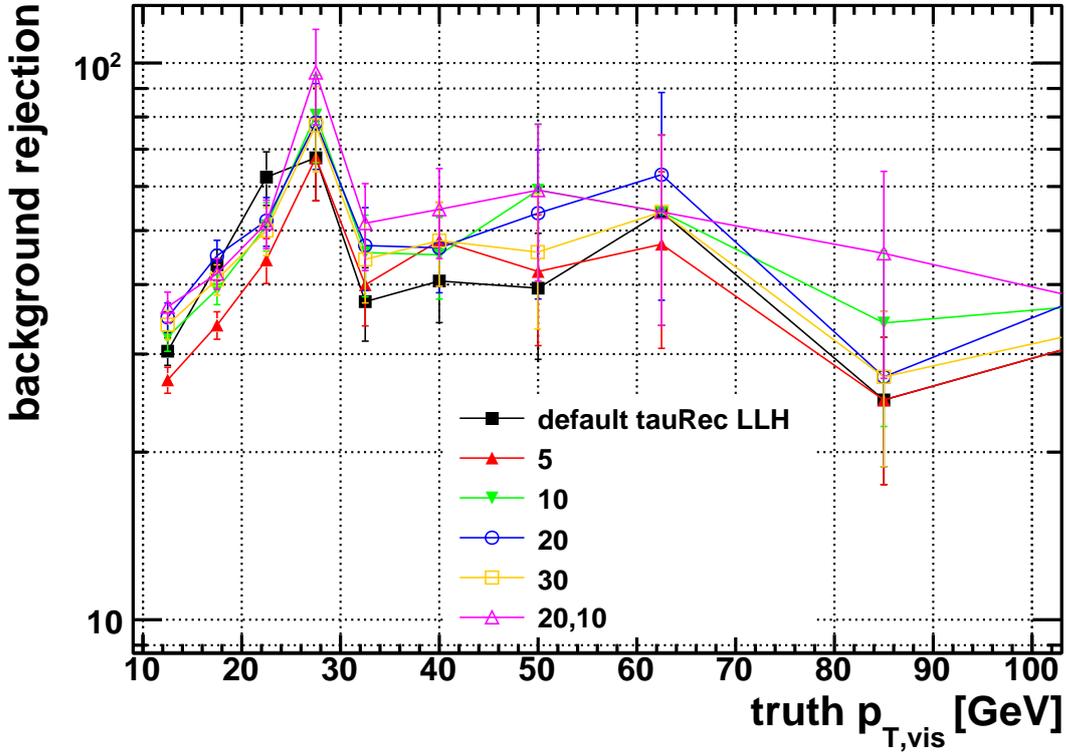


Figure 6.18: Rejection against QCD jets as function of the true transverse momentum at a signal efficiency of 50%. Different network topologies are shown. The default tauRec likelihood is included as a reference. The efficiency and rejection are calculated with respect to reconstructed tauRec candidates with one associated track.

The rejections for the tauRec likelihood and the best performing network are listed in Table 6.6.

### Multi Track Candidates

For the category of tauRec candidates with more than one associated track two new identification variables were used compared to the tauRec likelihood discriminant. The first of those variables, the lifetime signed transverse flight path significance, became available for tauRec candidates some time before the merging of the algorithms was implemented. The same code is shared between tauRec and tau1P3P to reconstruct a secondary vertex from the associated tracks and calculate the transverse flight path significance. The second variable, the invariant track mass of the track system defined by tauRec, has been calculated in the offline scripts, which were used to write out all  $\tau$  candidates from the full event data. It has been checked, that those two variables actually improve the discrimination power. A network trained without them will be shown later for comparison.

## 6 $\tau$ Algorithm Merging

discriminant	default tauRec	best configuration
true $p_{T,vis}$		
12 – 15 GeV	$30.4 \pm 1.7$	$36.5 \pm 2.3$
15 – 20 GeV	$43.4 \pm 2.8$	$41.9 \pm 2.7$
20 – 25 GeV	$62.4 \pm 6.9$	$51.6 \pm 5.2$
25 – 30 GeV	$68 \pm 11$	$96 \pm 19$
30 – 35 GeV	$37.3 \pm 5.6$	$51.5 \pm 9.2$
35 – 45 GeV	$40.6 \pm 6.4$	$54.6 \pm 10$
45 – 55 GeV	$39.4 \pm 10$	$59 \pm 19$
55 – 70 GeV	$54 \pm 20$	$54 \pm 20$
70 – 100 GeV	$24.8 \pm 7.3$	$46 \pm 18$

Table 6.6: Rejections at an efficiency of 50% for true  $\tau$  jets for the default discriminant of tauRec and the best performing network with two hidden layers with 20 and 10 nodes, respectively. The efficiency and rejection are calculated with respect to reconstructed tauRec candidates with one associated track.

The number of hidden layers and hidden nodes have been optimised using the following input variables:

- EM-Radius
- Isolation fraction
- $\eta$ -strip width
- number of  $\eta$ -strip hits
- $E_T/p_{T,leading}$
- absolute charge
- $E_T$  from the calorimeter
- invariant track mass using tracks as defined by tauRec
- lifetime signed transverse flight path significance

The rejections for different network topologies are shown in Figure 6.19 for a  $\tau$  jet efficiency of 50% as function of the true transverse momentum. As before, more than ten hidden nodes must be used to reach a good performance, but improvements can still be observed up to 40 hidden nodes. Also neural networks with two hidden layers can slightly improve the performance by a few percent over a network with 40 hidden nodes in one hidden layer. A network with two hidden layers with 30 and 10 nodes has been chosen, since it provided the best performance for small transverse momenta, which is usually the region populated by the most QCD jets.

## 6 $\tau$ Algorithm Merging

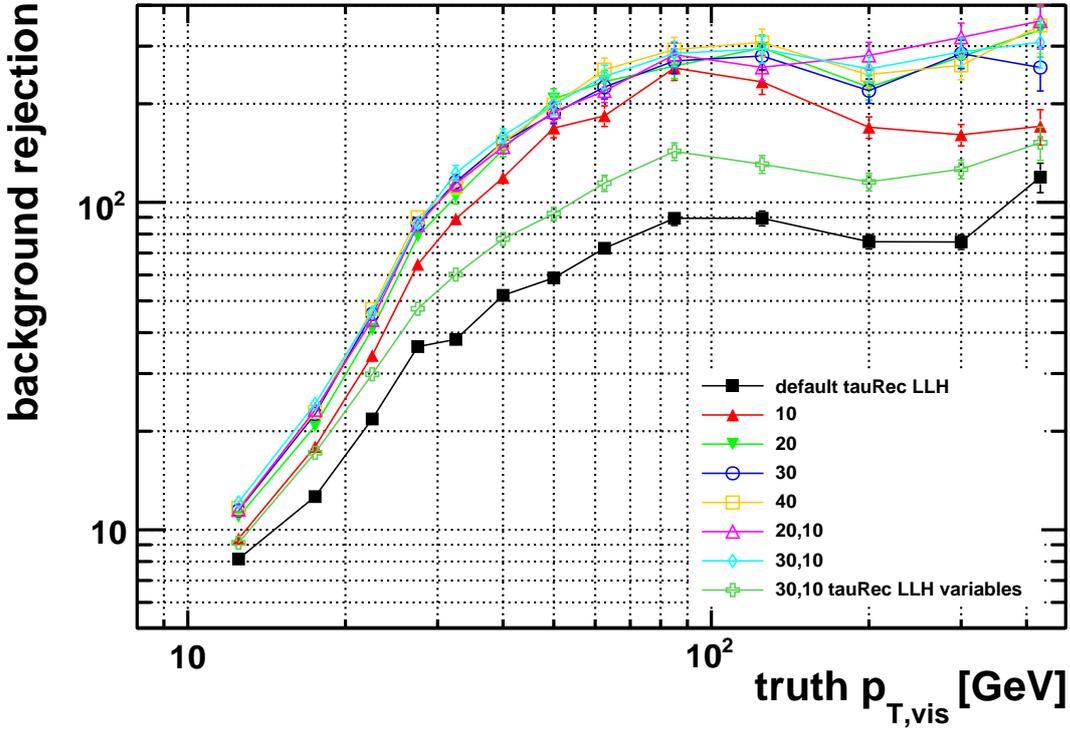


Figure 6.19: Rejection against QCD jets as function of the true transverse momentum at a signal efficiency of 50%. Different network topologies are shown. The default tauRec likelihood is included as a reference, as well as a network which uses the same variables. See the text for a detailed description. The efficiency and rejection are calculated with respect to reconstructed tauRec candidates with one associated track.

In Figure 6.19 the rejections achieved by the default tauRec likelihood discriminant are shown. Also the performance for a network is shown, which uses only the variables used in the likelihood and the transverse energy and which uses a topology of two hidden layers with 30 and 10 nodes denoted by ‘30,10 tauRec LLH variables’ in the legend. In this network the lifetime signed transverse flight path significance and the invariant track mass are not used and instead the number of associated tracks and the lifetime signed transverse impact parameter significance are added to the list of variables. The rejections for these two networks and for the optimal network determined above are given in Table 6.7.

In this category a large improvement of the discrimination of a factor of two to three is achieved by the optimal neural network discriminant compared to the tauRec likelihood. The neural network discriminant might therefore prove useful when using only tauRec candidates, without making use of the merged algorithm.

## 6 $\tau$ Algorithm Merging

discriminant	default tauRec	tauRec LLH variables	best configuration
true $p_{T,vis}$			
12 – 15 GeV	$8.1 \pm 0.1$	$9.1 \pm 0.1$	$12.1 \pm 0.2$
15 – 20 GeV	$12.6 \pm 0.2$	$17.2 \pm 0.3$	$24.2 \pm 0.5$
20 – 25 GeV	$21.8 \pm 0.4$	$29.8 \pm 0.7$	$46.1 \pm 1.3$
25 – 30 GeV	$36.3 \pm 1.0$	$47.3 \pm 1.5$	$86.3 \pm 3.7$
30 – 35 GeV	$38.1 \pm 1.1$	$60.2 \pm 2.3$	$123 \pm 6.6$
35 – 45 GeV	$52.0 \pm 1.5$	$77.3 \pm 2.7$	$160 \pm 8.0$
45 – 55 GeV	$58.8 \pm 2.3$	$92.4 \pm 4.5$	$199 \pm 14$
55 – 70 GeV	$72.4 \pm 3.3$	$114 \pm 6.5$	$243 \pm 20$
70 – 100 GeV	$89.4 \pm 4.4$	$143 \pm 8.9$	$285 \pm 25$
100 – 150 GeV	$89.4 \pm 4.7$	$131 \pm 8.2$	$294 \pm 28$
150 – 250 GeV	$75.8 \pm 3.8$	$115 \pm 7.1$	$256 \pm 24$

Table 6.7: Rejections at a  $\tau$  jet efficiency of 50% for the default likelihood discriminant of tauRec, a neural network using the same variables as the likelihood and the best performing network using the transverse flight path and the invariant track mass. Both neural networks consist of two hidden layers with 30 and 10 nodes, respectively. The efficiency and rejection are calculated with respect to reconstructed tauRec candidates with more than one associated track.

### 6.3.5 Combined Performance

In Section 6.3.3 it has been shown that the identification performance can be improved significantly for  $\tau$  candidates that are reconstructed by both  $\tau$  identification algorithms, by making use of the additional information available for those  $\tau$  candidates. Especially for overlap candidates with more than one associated tracks a large improvement of a factor of more than two could be observed above a transverse momentum of 25 GeV.

A better identification performance could be also attained for the  $\tau$  candidates which are only reconstructed by tauRec. Compared to the likelihood method implemented in the tauRec package by using a neural network discriminant which is trained separately for candidates with one associated track and candidates with more than one track. In the latter category the rejection against QCD jets was increased by a factor more than two.

However, the numbers shown so far cannot be readily used to compare the rejection against true jets that can be expected between the different discriminants, since they are calculated with respect to a different number of total signal and background candidates in each category. Therefore the sample that was set aside at the beginning of the training, was used to assess the performance of the new discriminants on a sample containing a ‘natural’ mixture of  $\tau$  candidates, where no special selection has been applied.

A discriminant value is calculated for every  $\tau$  candidate using the best neural network that has been determined above for its respective category.  $\tau$  candidates which do not belong to any of the four categories, like tau1P3P only candidates, get assigned a value of -1, so that they are never selected as a  $\tau$  jet.

## 6 $\tau$ Algorithm Merging

In Figure 6.20 the rejection for the the tau1P3P neural network discriminant and the tauRec likelihood discriminant are compared with the rejection for the neural networks that were optimised in Section 6.3.3 and Section 6.3.4 for a  $\tau$  efficiency of 50%. Here the efficiency and rejection is given with respect to the true  $\tau$  jets and the *Cone4TopoJets*, respectively. The new discriminant provides the best performance for transverse momenta below 60 GeV. For higher momenta the performance of the tauRec likelihood becomes better, because it has a much lower efficiency for  $\tau$  candidates with more than one track. In this high transverse momentum region the background is dominated by such candidates. For higher efficiencies the new discriminant subsequently gains performance relative to the two default discriminants. This can be seen in Figure 6.21 and Figure 6.22 where the rejections for 60% and 70% signal efficiency are shown.

Even though the new discriminant performs better in each category than the two default discriminants, the combined discriminant still shows a worse rejection against QCD jets in some  $p_T$  ranges for low signal efficiencies. Clearly the performance of the new discriminant for low  $\tau$  efficiencies could be further improved by optimising the cut on the discriminant value independently for single track and multitrack candidates. The optimal relative signal efficiencies for the the different categories will depend on the specific analysis, and should be optimised from case to case.

## 6 $\tau$ Algorithm Merging

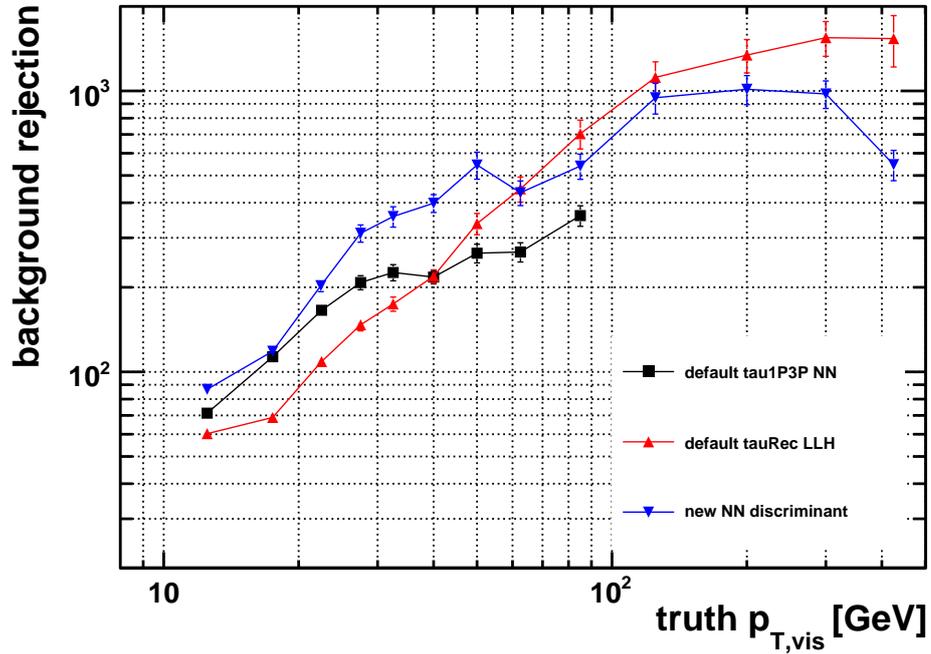


Figure 6.20: Rejection against QCD jets as function of the true transverse momentum at a signal efficiency of 50%. The rejection of the tau1P3P neural network discriminant and the tauRec likelihood discriminant are compared to rejection of the optimised neural network discriminant. The efficiency and rejection are calculated with respect to true  $\tau$  jets and Cone4TruthJets, respectively.

## 6 $\tau$ Algorithm Merging

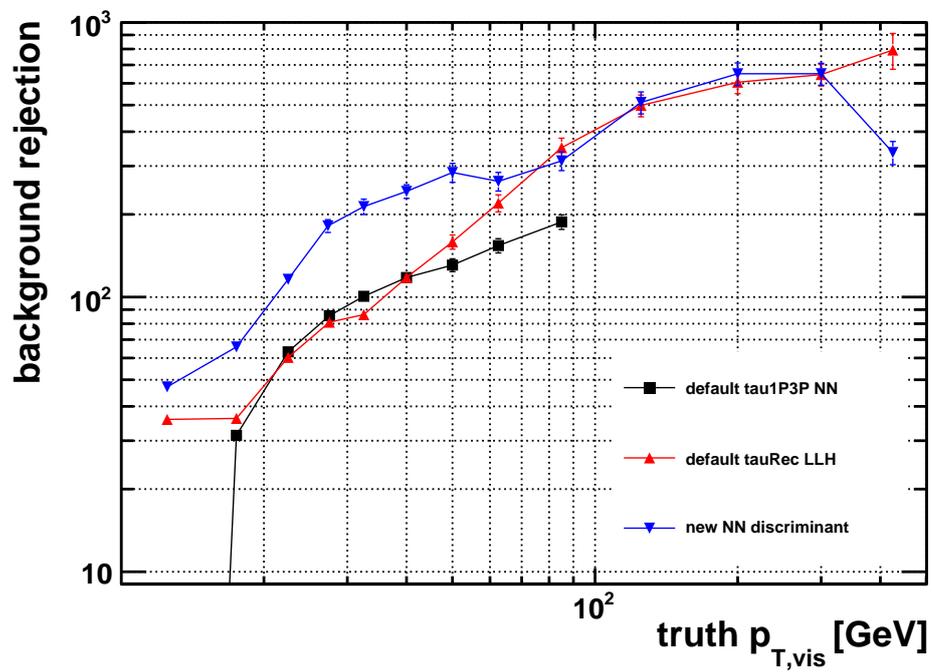


Figure 6.21: Rejection against QCD jets as function of the true transverse momentum at a signal efficiency of 60%. The rejection of the tau1P3P neural network discriminant and the tauRec likelihood discriminant are compared to rejection of the optimised neural network discriminant.

## 6 $\tau$ Algorithm Merging

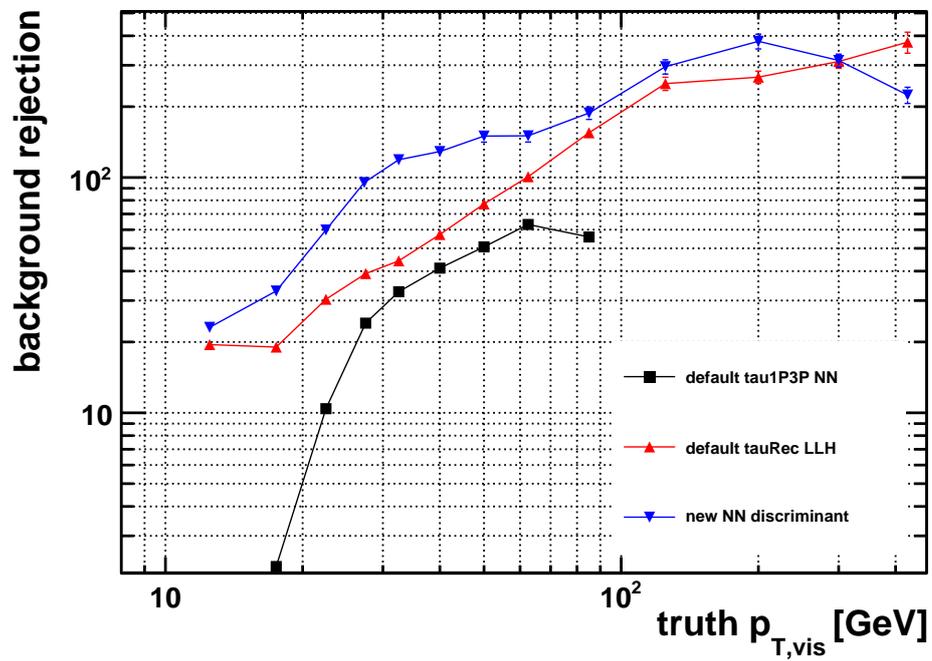


Figure 6.22: Rejection against QCD jets as function of the true transverse momentum at a signal efficiency of 70%. The rejection of the tau1P3P neural network discriminant and the tauRec likelihood discriminant are compared to rejection of the optimised neural network discriminant.

# 7 Summary

At the LHC, final states containing hadronically decaying  $\tau$  leptons will be an important signature in the search for a standard model Higgs boson in the low mass region as well for new physics beyond the standard model. Once a signal for new physics can be observed,  $\tau$  leptons might be able to provide further insight into the spin structure and couplings of the new phenomena. Therefore, the efficient reconstruction and identification of hadronically decaying  $\tau$  leptons is an important task for the detectors at the LHC.

In this thesis the existing algorithms for the reconstruction of  $\tau$  leptons in the ATLAS experiment have been studied in view of potential improvements, which lead to a cleaner selection of  $\tau$  jets. Two complementary  $\tau$  reconstruction algorithms have been developed for the ATLAS experiment, following somewhat different strategies.

The older of the two algorithms, *tauRec*, is based on energy depositions in the calorimeter and reconstructs hadronically decaying  $\tau$  leptons with a high efficiency above a visible transverse momentum of about 30 GeV. It provides a very strong discrimination against QCD jets with high transverse momenta.

The second algorithm, *tau1P3P*, is based on tracks reconstructed in the inner detector. It is aimed at low transverse momenta below 100 GeV and reconstructs specifically  $\tau$  jets with a charged particle carrying a sizable fraction of the visible transverse momentum of the  $\tau$ . This provides some suppression of QCD jets already at the reconstruction stage but also limits the maximal reconstruction efficiency, especially for low transverse momenta.

Modifications to the *tauRec* package have been presented, which were developed in the scope of this thesis with the goal to improve the efficiency for reconstructing  $\tau$  leptons with small transverse momenta. To reach this goal, use was made of a more sophisticated topological clustering algorithm to reconstruct energy clusters in the calorimeter. It was shown that the topological clustering algorithm provides a higher reconstruction efficiency for  $\tau$  jets with small visible transverse momenta. At the same time the clustering algorithm is able to reduce the contribution of noise to shower shape variables calculated from the cells of the calorimeter. The reconstruction efficiency was considerably improved in the transverse momentum region below 30 GeV, reaching a level of more than 90% down to transverse momenta of 10 GeV. The reconstruction of  $\tau$  leptons in this low momentum regime might prove important in certain scenarios for new physics, like MSSM scenarios with a large value of the  $\tan\beta$  parameter.

Since the cross section for QCD jets becomes very high at such low transverse momenta, the ability to discriminate hadronically decaying  $\tau$  leptons from QCD jets had to be improved. This has been aided by the mentioned property of the topological clustering, to provide an excellent suppression of noise in the energy measurement of the calorimeter

## 7 Summary

while retaining real energy depositions. The identification performance has been improved by optimising the existing likelihood discriminant of the tauRec package. This led to a significantly better rejection against QCD jets for a wide kinematic region with transverse momenta between 15 GeV and 120 GeV, where the rejection factors for a typical  $\tau$  efficiency of 30%, as used in many analyses, could be improved by up to 78% percent.

The modifications to the tauRec package have since become the default in the ATLAS reconstruction software.

A first step towards a merged algorithm has been developed, combining the two complementary approaches to the reconstruction of hadronically decaying  $\tau$  leptons implemented in the ATLAS reconstruction framework. The implementation of the merged algorithm was documented and the choices that were made during this work have been motivated. One motivation for combining the two approaches into a single algorithm was to concentrate development efforts and to facilitate the usage of the reconstructed  $\tau$  leptons in high level analyses by eliminating the burden of choice from the user. Apart from that, this development also offered the chance to investigate possible synergy effects between the two approaches and their potential to further improve the discrimination between hadronically decaying  $\tau$  leptons and QCD jets.

A study of this potential was shown, which combined different properties of  $\tau$  leptons and QCD jets reconstructed by the calorimeter based and the track based algorithm in a neural network discriminant. The reconstructed  $\tau$  candidates were grouped in four categories depending on whether they are reconstructed by both algorithms or not and on their number of associated tracks.

It was shown for the overlap candidates reconstructed by both algorithms, that the rejection against QCD jets can be significantly improved when certain identification variables defined by only one of the two algorithms are used in conjunction. Especially for 3-prong decays of the  $\tau$  lepton the rejection against QCD jets with more than one reconstructed charged particle could be improved by more than a factor of two over a wide range of transverse momenta between 12 GeV and 250 GeV and a wide range of signal efficiencies. For 1-prong decays the rejection against QCD jets with one reconstructed charged particle could be improved by 50% on average over the same kinematic region and wide range of efficiencies. These results can provide insight for the further development of a merged algorithm, where the reconstruction of certain strongly discriminating properties should be extended to all  $\tau$  candidates.

# A Comparison of tauRec Identification Variables between Release 12 and 13

In the following the identification variables used by tauRec are compared between ATHENA release 12 and release 13. This is done to justify the comparison of the identification performance between these two releases in Chapter 5. In all the following plots the distributions for release 12 will be shown in light colours (light gray/light red) and the distributions for release 13 in dark colours. The black/gray distributions are for  $\tau$  jets and the red/light red distributions are for QCD jets. Please note that the errors are only statistical. The samples do not consist of the same events and therefore have a somewhat different kinematic composition.

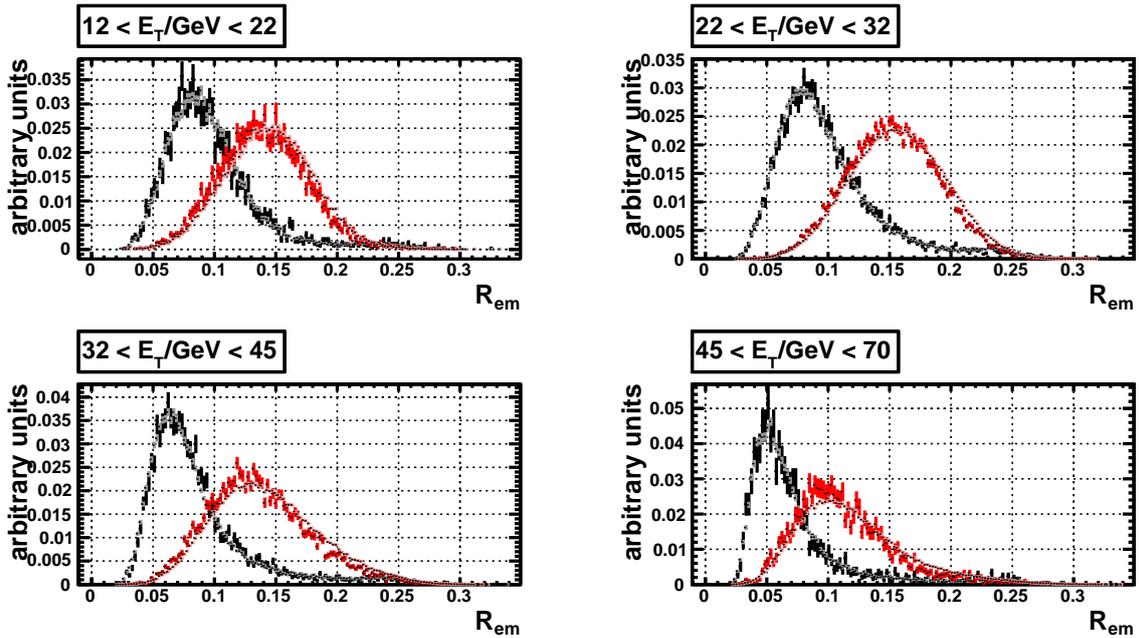


Figure A.1: Comparison of the EM-Radius between ATHENA release 12 and release 13 using the default tauRec reconstruction for different  $E_T$  ranges

A Comparison of tauRec Identification Variables between Release 12 and 13

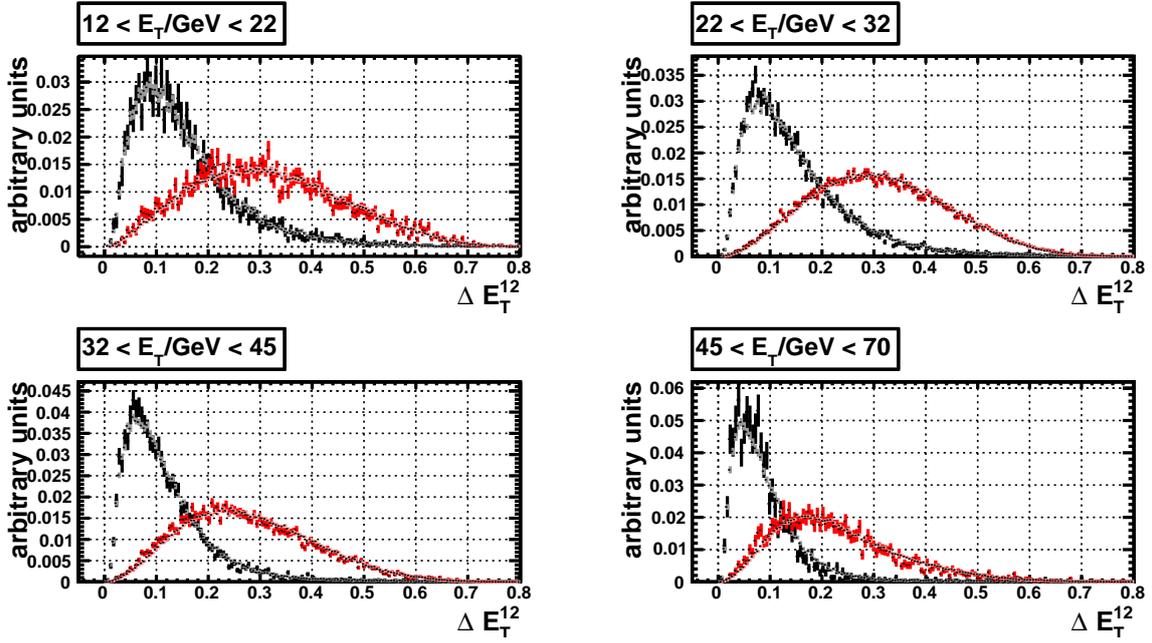


Figure A.2: Comparison of the calorimeter isolation between ATHENA release 12 and release 13 using the default tauRec reconstruction for different  $E_T$  ranges

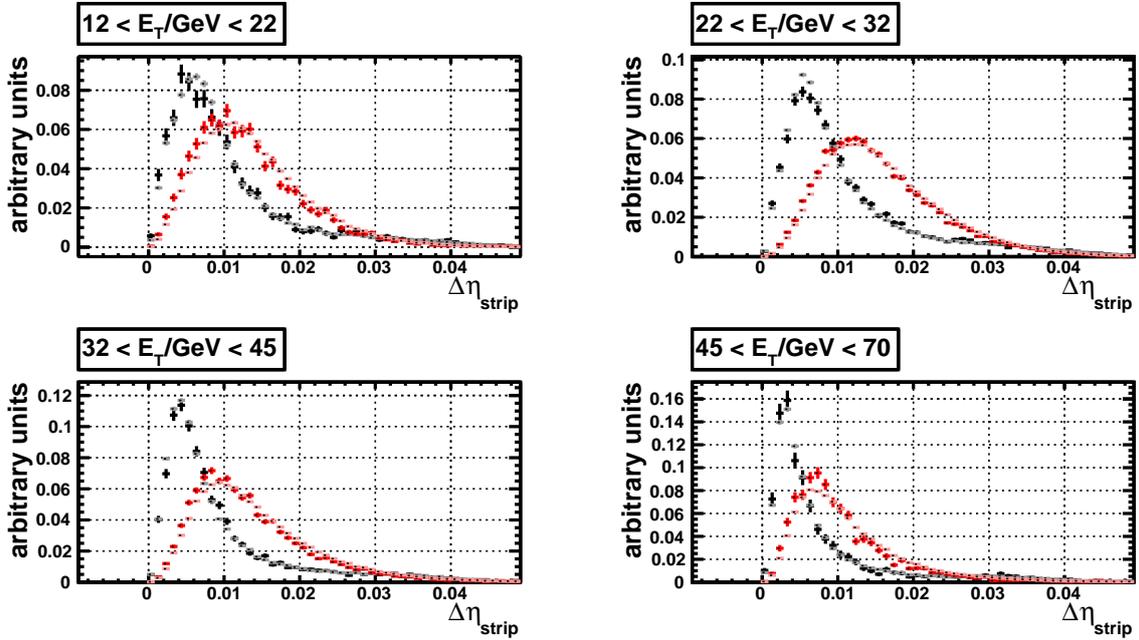


Figure A.3: Comparison of the  $\eta$  strip width between ATHENA release 12 and release 13 using the default tauRec reconstruction for different  $E_T$  ranges

A Comparison of tauRec Identification Variables between Release 12 and 13

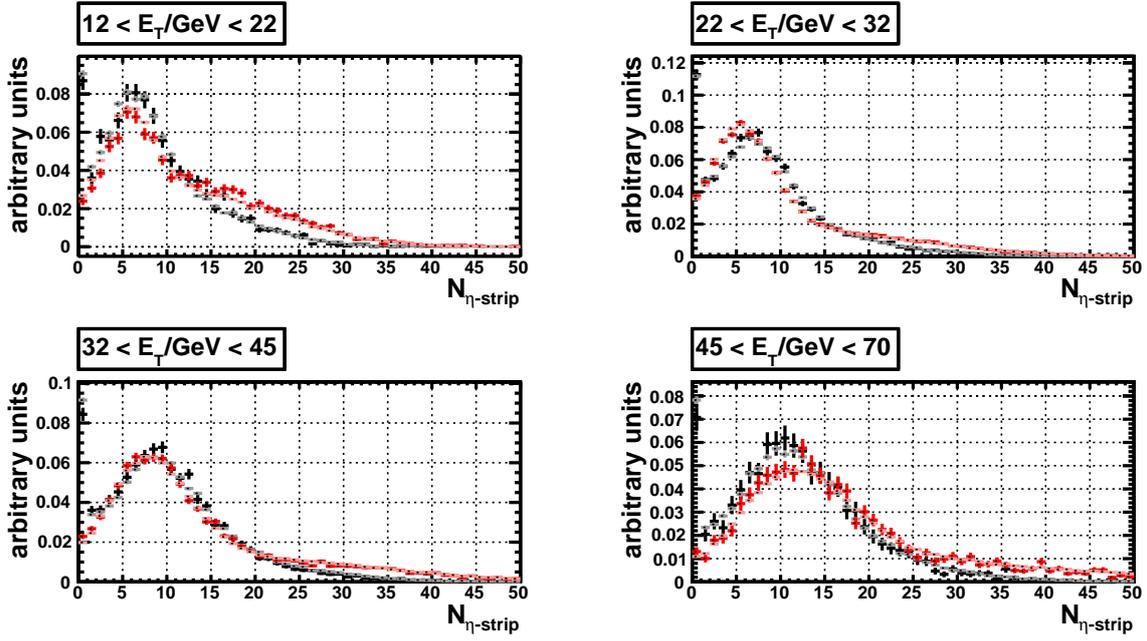


Figure A.4: Comparison of the number of  $\eta$  strip hits between ATHENA release 12 and release 13 using the default tauRec reconstruction for different  $E_T$  ranges

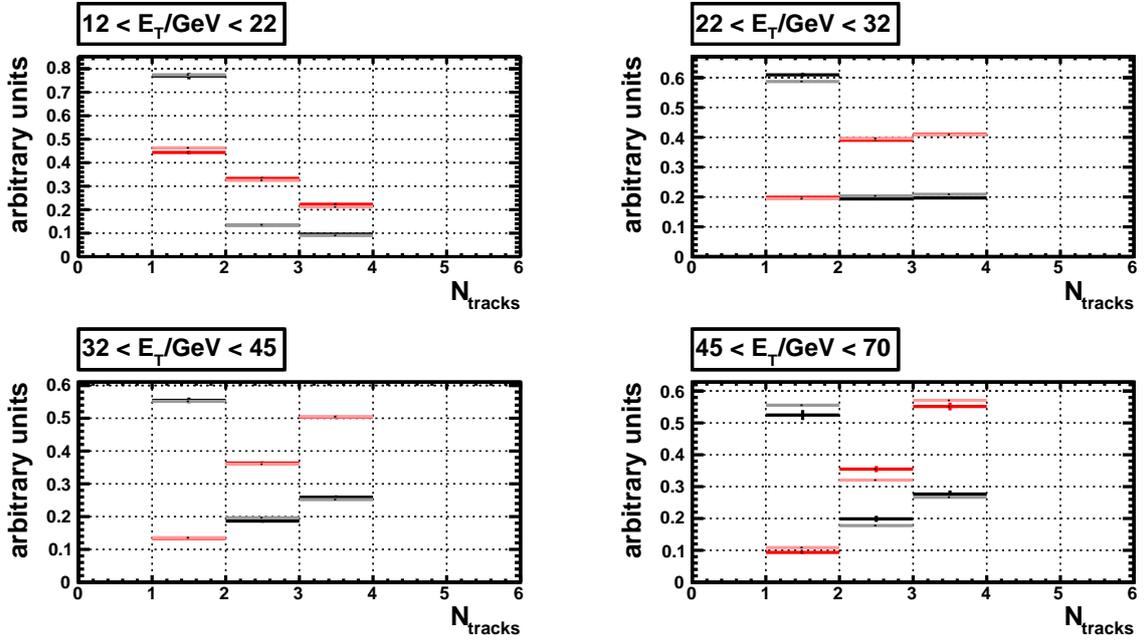


Figure A.5: Comparison of the number of associated tracks between ATHENA release 12 and release 13 using the default tauRec reconstruction for different  $E_T$  ranges

A Comparison of tauRec Identification Variables between Release 12 and 13

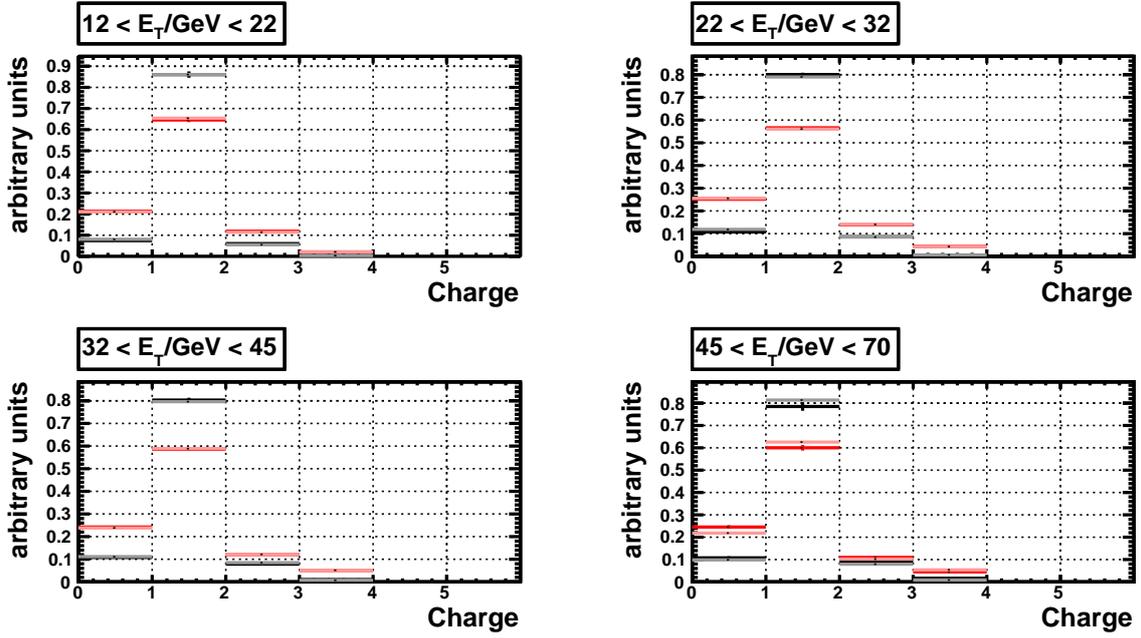


Figure A.6: Comparison of the absolute value of the  $\tau$  charge between ATHENA release 12 and release 13 using the default tauRec reconstruction for different  $E_T$  ranges

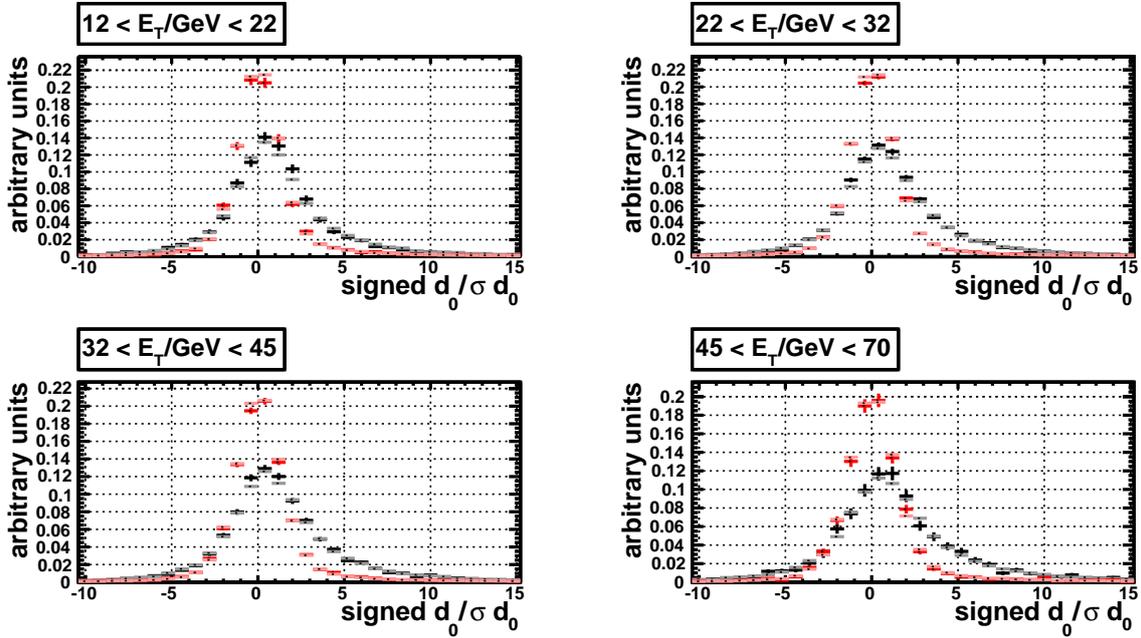


Figure A.7: Comparison of the signed transverse impact parameter significance between ATHENA release 12 and release 13 using the default tauRec reconstruction for different  $E_T$  ranges

A Comparison of tauRec Identification Variables between Release 12 and 13

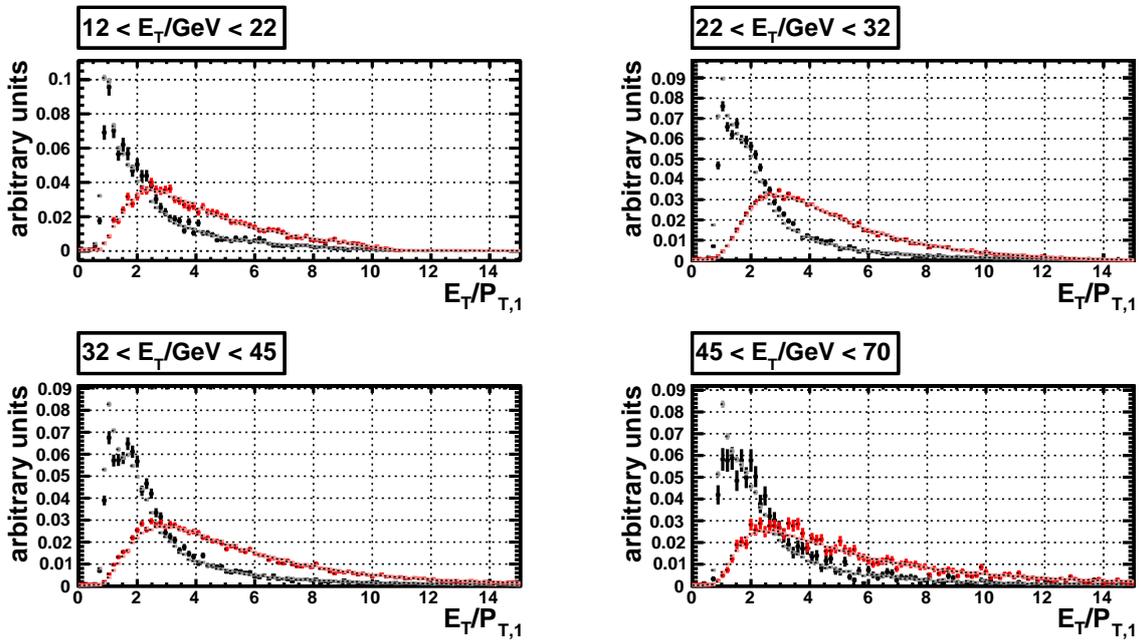


Figure A.8: Comparison of  $E_T/p_{T,\text{leading}}$  between ATHENA release 12 and release 13 using the default tauRec reconstruction for different  $E_T$  ranges

# List of Tables

1.1	The particles of the first family of the standard model with their quantum numbers . . . . .	7
2.1	important characteristics of the LHC . . . . .	17
2.2	Expected resolution for the reconstruction of track parameters . . . . .	22
2.3	Energy resolution and granularity of the different calorimeter subsystems .	26
4.1	$\tau$ branching ratios . . . . .	40
4.2	Rejection factors for some selected signal efficiencies and $p_T$ ranges for default tauRec . . . . .	61
5.1	Separation power of the calorimeter based variables for default reconstruction and seeding by <i>TopoJets</i> . . . . .	73
5.2	Rejection factors for some selected signal efficiencies and $p_T$ ranges for <i>TopoJet</i> seeded tauRec . . . . .	79
6.1	Fraction of $\tau$ jets that is reconstructed by each algorithm or by both for different bins of the true visible transverse momentum. . . . .	84
6.2	Number of $\tau$ jets and QCD jets in each training sample. . . . .	93
6.3	The list of variables which are used as input for the tau1P3P neural network discriminant . . . . .	94
6.4	Rejections for overlap candidates with one associated track . . . . .	105
6.5	Rejections for overlap candidates with more than one associated track . . .	113
6.6	Rejections for tauRec candidates with one associated track . . . . .	116
6.7	Rejections for tauRec candidates with more than one associated track . . .	118

# List of Figures

1.1	The particle content of the Standard Model . . . . .	5
1.2	Higgs potential . . . . .	8
1.3	Feynman graphs of the leptonic and hadronic $\tau$ decays . . . . .	12
1.4	The production cross sections (left) and branching ratios (right) for a standard model Higgs boson as a function of its mass (from [2]). . . . .	14
2.1	Cut-away view of the ATLAS detector [33] . . . . .	19
2.2	Cut-away view of the ATLAS Inner Detector, from [33] . . . . .	20
2.3	Track reconstruction efficiencies as a function of $\eta$ for charged pions with different transverse momenta. This plot is taken from [33]. . . . .	23
2.4	Cut-away view of the ATLAS Calorimeter System, from [33] . . . . .	24
2.5	Schematic drawing of a part of the EM barrel calorimeter, taken from [33]	25
3.1	graphical representation of a neuron . . . . .	33
3.2	The sigmoid function for different values of the inverse temperature . . . . .	34
3.3	a multi layer perceptron with four input variables, one output node and two hidden layers . . . . .	35
4.1	Reconstruction efficiency for $\tau$ jets as a function of $p_T$ for <i>CombinedClusters</i>	44
4.2	EM-Radius for different $E_T$ ranges of the $\tau$ candidates after requiring 1 to 3 associated tracks . . . . .	45
4.3	Calorimeter isolation for different $E_T$ ranges of the $\tau$ candidates after requiring 1 to 3 associated tracks . . . . .	47
4.4	$\eta$ -strip width for different $E_T$ ranges of the $\tau$ candidates after requiring 1 to 3 associated tracks . . . . .	48
4.5	Number of hits in the $\eta$ -strip layer after requiring 1 to 3 associated tracks .	49
4.6	Efficiency of requiring 1 to 3 associated tracks for $\tau$ jets as a function of the true $p_T$ of the visible tau decay products. . . . .	51
4.7	Efficiency of requiring 1 to 3 associated tracks for QCD jets as a function of the true $E_T$ . . . . .	52
4.8	Number of associated tracks in the range 1 to 3 . . . . .	53
4.9	Absolute charge of the $\tau$ candidate after requiring 1 to 3 associated tracks	54
4.10	The lifetime signed transverse impact parameter significance of a track . .	55
4.11	Lifetime signed transverse impact parameter significance of the $\tau$ candidate after requiring 1 to 3 associated tracks . . . . .	56

*List of Figures*

4.12	E <sub>T</sub> over p <sub>T</sub> of the leading track of the $\tau$ candidate after requiring 1 to 3 associated tracks . . . . .	57
4.13	Distribution of the likelihood value for $\tau$ jets from Z/A $\rightarrow \tau\tau$ events in black and QCD jets from QCD dijet events in red after requiring 1 to 3 associated tracks . . . . .	59
4.14	Rejection against QCD jets as a function of the efficiency to select true $\tau$ jets for default tauRec . . . . .	60
4.15	Rejection against QCD jets as a function of p <sub>T</sub> for default tauRec . . . . .	61
4.16	Number of $\pi^0$ clusters . . . . .	64
4.17	Invariant mass of the track system . . . . .	65
5.1	Probability to reconstruct a true $\tau$ jet as a Cone4TopoJet as function of p <sub>T</sub>	71
5.2	Comparison of the <i>EM-Radius</i> between default tauRec and <i>TopoJet</i> seeded tauRec for different E <sub>T</sub> ranges . . . . .	74
5.3	Comparison of the <i>isolation fraction</i> between default tauRec and <i>TopoJet</i> seeded tauRec for different E <sub>T</sub> ranges . . . . .	75
5.4	Comparison of the <i><math>\eta</math>-strip width</i> between default tauRec and <i>TopoJet</i> seeded tauRec for different E <sub>T</sub> ranges . . . . .	76
5.5	Comparison of the number of $\eta$ -strips hits between default tauRec and <i>TopoJet</i> seeded tauRec for different E <sub>T</sub> ranges . . . . .	77
5.6	The smoothed PDFs overlaid with the unsmoothed distributions for the E <sub>T</sub> bin from 45 GeV to 70 GeV . . . . .	78
5.7	Rejection of QCD jets as function of the efficiency to select true $\tau$ jets . . .	80
5.8	Rejection against QCD jets as a function of the visible p <sub>T</sub> for a signal efficiency of 30% and 50% . . . . .	81
5.9	Comparison of the rejection against QCD jets as a function of the visible p <sub>T</sub> between the reconstruction with TopoJet seeds and the default reconstruction with two different likelihood discriminants . . . . .	82
6.1	Transverse momentum of $\tau$ jets reconstructed by either tauRec or tau1P3P	84
6.2	Reconstruction efficiency for QCD jets for a combination of several QCD dijet samples . . . . .	85
6.3	The distance in the $\eta$ - $\phi$ plane between matched tau1P3P and tauRec candidates . . . . .	87
6.4	Comparison of the E <sub>T</sub> and $\eta$ distribution between $\tau$ jets and QCD jets in the normalised training sample for overlap candidates . . . . .	92
6.5	Distribution of the E <sub>T</sub> ratio in the training samples for single track candidates and multi track candidates . . . . .	95
6.6	Rejection for overlap candidates with one associated track as function of the signal efficiency for network topologies with one hidden layer . . . . .	96
6.7	Rejection for overlap candidates with one associated track as function of p <sub>T</sub> for network topologies with one hidden layer . . . . .	97

*List of Figures*

6.8	Rejection for overlap candidates with one associated track as function of the signal efficiency for network topologies with two hidden layers . . . . .	98
6.9	Rejection for overlap candidates with one associated track as function of the signal efficiency for different input variables . . . . .	100
6.10	Rejection for overlap candidates with one associated track as function of the signal efficiency using tauRec shower shape variables . . . . .	102
6.11	Rejection for overlap candidates with one associated track as function of $p_T$ using tauRec shower shape variables . . . . .	103
6.12	Rejection for overlap candidates with one associated track as function of $p_T$ , comparing the default discriminants with the optimised network . . . . .	104
6.13	Rejection for overlap candidates with more than one associated track as function of the signal efficiency for different network topologies . . . . .	107
6.14	Rejection against QCD jets as function of the efficiency for real $\tau$ jets for overlap candidates with more than one associated track . . . . .	108
6.15	Rejection against QCD jets as function of the efficiency for real $\tau$ jets for overlap candidates with more than one associated track . . . . .	109
6.16	Comparison of rejection between default discriminants and best network for overlap candidates with more than one associated track as function of $p_T$ , successively adding variables . . . . .	111
6.17	Rejection for overlap candidates with more than one associated track as function of $p_T$ . . . . .	112
6.18	Rejection for tauRec candidates with one associated track for different network topologies as function of $p_T$ . . . . .	115
6.19	Rejection for tauRec candidates with more than one associated track for different network topologies as function of $p_T$ . . . . .	117
6.20	Rejection for overlap and tauRec candidates as function of $p_T$ for 50% efficiency	120
6.21	Rejection for overlap and tauRec candidates as function of $p_T$ for 60% efficiency	121
6.22	Rejection for overlap and tauRec candidates as function of $p_T$ for 70% efficiency	122
A.1	Comparison of the EM-Radius between ATHENA release 12 and release 13 using the default tauRec reconstruction for different $E_T$ ranges . . . . .	I
A.2	Comparison of the calorimeter isolation between ATHENA release 12 and release 13 using the default tauRec reconstruction for different $E_T$ ranges . . . . .	II
A.3	Comparison of the $\eta$ strip width between ATHENA release 12 and release 13 using the default tauRec reconstruction for different $E_T$ ranges . . . . .	II
A.4	Comparison of the number of $\eta$ strip hits between ATHENA release 12 and release 13 using the default tauRec reconstruction for different $E_T$ ranges . . . . .	III
A.5	Comparison of the number of associated tracks between ATHENA release 12 and release 13 using the default tauRec reconstruction for different $E_T$ ranges . . . . .	III
A.6	Comparison of the absolute value of the $\tau$ charge between ATHENA release 12 and release 13 using the default tauRec reconstruction for different $E_T$ ranges . . . . .	IV

A.7	Comparison of the signed transverse impact parameter significance between ATHENA release 12 and release 13 using the default tauRec reconstruction for different $E_T$ ranges . . . . .	IV
A.8	Comparison of $E_T/p_{T,\text{leading}}$ between ATHENA release 12 and release 13 using the default tauRec reconstruction for different $E_T$ ranges . . . . .	V

# References

- [1] American Association for the Advancement of Science. <http://www.aaas.org>.
- [2] Sara Bolognesi, Giuseppe Bozzi, and Andrea Di Simone. Higgs at LHC, 2008.
- [3] G. Brandenburg and CLEO Collaboration. Charged Track Multiplicity in B Meson Decay. *Physical Review D*, 61:072002, 2000.
- [4] W. Cleveland and C. Loader. Smoothing by local regression: Principles and methods.
- [5] S. D. Ellis, J. Huston, K. Hatakeyama, P. Loch, and M. Toennesmann. Jets in hadron-hadron collisions, 2007.
- [6] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13(9):321–323, Aug 1964.
- [7] W.-M. Yao et al. (Particle Data Group). Review of Particle Physics. *J. Phys. G*, 33(1), 2006. and 2007 partial update for the 2008 edition.
- [8] Jonathan L. Feng and Mihoko M. Nojiri. Supersymmetry and the Linear Collider, 2002.
- [9] S. L. Glashow. Partial Symmetries of Weak Interactions. *Nucl. Phys.*, 22:579–588, 1961.
- [10] ed. Harrison, P. F. and ed. Quinn, Helen R. The BaBar physics book: Physics at an asymmetric B factory. Papers from Workshop on Physics at an Asymmetric B Factory (BaBar Collaboration Meeting), Rome, Italy, 11-14 Nov 1996, Princeton, NJ, 17-20 Mar 1997, Orsay, France, 16-19 Jun 1997 and Pasadena, CA, 22-24 Sep 1997.
- [11] M. Heldmann and D. Cavalli. An Improved  $\tau$ -Identification for the ATLAS-Experiment, 2006. ATL-PHYS-PUB-2006-008.
- [12] Michael Heldmann. *The Identification of Hadronically Decaying  $\tau$  Leptons in the ATLAS Experiment and Investigation of the Discovery Potential for MSSM Higgs Bosons in  $\tau$  Final States*. PhD thesis, Universität Freiburg, 2007.
- [13] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13(16):508–509, Oct 1964.

## REFERENCES

- [14] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989.
- [15] W. Häderle. *Applied Nonparametric Regression*. Number 19 in Econometric Society Monographs. Cambridge University Press, 1992.
- [16] J. Friedman. 353QH twice smoothing algorithm. In *1974 CERN School of Computing*. Norway, Aug. 11-24 1974.
- [17] Kuncheva L.I. On the optimality of naïve bayes with dependent binary features. *Pattern Recognition Letters*, 27:830–837, 2006.
- [18] C. Loader. Locfit - Local Regression and Likelihood. <http://locfit.herine.net>.
- [19] C. Loader. *Local Regression and Likelihood*. Springer New York, 1999.
- [20] E. A. Nadaraya. On estimating regression. *Theory of Probability and its Applications*, 9(1).
- [21] E. Parzen. On the estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33:1065–1076, 1962.
- [22] Carsten Peterson, Thorsteinn Rögnvaldsson, and Leif Lönnblad. Jetnet 3.0 - a versatile artificial neural network package.
- [23] Giacinto Piacquiadio. private communications.
- [24] Sreerup Raychaudhuri and D. P. Roy. Sharpening up the Charged Higgs Boson Signature using Tau Polarization at LHC. *Physical Review D*, 53:4902, 1996.
- [25] M Riedmiller and H Braun. A direct adaptive method for faster backpropagation learning: The RPROP algorithm. In *Proceedings of the IEEE International Conference on Neural Networks*, pages 586–591, 1993.
- [26] Christoph Ruwiedel. private communications.
- [27] Christoph Ruwiedel. Vertexing for taus. Talk given at ATLAS Tau Working Group phone meeting, <http://indico.cern.ch/contributionDisplay.py?contribId=5&confId=17514>.
- [28] Abdus Salam. Weak and electromagnetic interactions. In N. Svartholm, editor, *Proceedings of the 8th Nobel Symposium*, page 367. Almqvist and Wiksell, Stockholm, 1968.
- [29] The ATLAS Collaboration. Jet Reconstruction. <https://twiki.cern.ch/twiki/bin/view/Atlas/JetAlgorithms>.

## REFERENCES

- [30] The ATLAS Collaboration. New Tracking. <https://twiki.cern.ch/twiki/bin/view/Atlas/NewIdOfflineTracking>.
- [31] The ATLAS Collaboration. ATLAS letter of intent for a general-purpose pp experiment at the large hadron collider at cern, 1992. CERN/LHCC/92-4, LHCC/I2.
- [32] The ATLAS Collaboration. ATLAS Detector and Physics Performance Technical Design Report. Technical report, CERN/LHCC-99-14, 1999. ATLAS TDR-014, Volume I.
- [33] The ATLAS Collaboration. The ATLAS experiment at the CERN large hadron collider, 2008.
- [34] The LEP Electroweak Working Group. <http://lepewwg.web.cern.ch/LEPEWWG/>.
- [35] Vassil Verguilov. TJETNET - jetnet wrapper for ROOT. <http://vasoto.web.cern.ch/vasoto/>.
- [36] M.P. Wand and M.C. Jones. *Kernel Smoothing*. Chapman Hall/CRC, 1995. Vol. 60 of Monographs on Statistics and Applied Probability.
- [37] Steven Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266, 1967.
- [38] Christian Weiser. Separation von Quark- und Gluonjets und Bestimmung des relativen  $K_s^0$ - und  $\Lambda^0$ -Produktionsverhältnisses mit dem DELPHI-Detektor bei LEP. Diplomarbeit Universität Karlsruhe, 1994.
- [39] Malgorzata Worek. Higgs CP from H/A  $\rightarrow$  tau tau decay. *ACTA PHYS.POLON.B*, 34:4549, 2003.

# Danksagung

Ich möchte allen danken, die mich auf dem Weg zur Erstellung dieser Arbeit unterstützt haben.

Ich danke Prof. Dr. Jakobs für die Möglichkeit an einem interessanten Thema zu arbeiten.

Mein besonderer Dank gilt auch Dr. Christian Weiser, der in der Endphase der Arbeit trotz chronischer Überarbeitung stets bemüht war mit Rat und Tat zur Seite zu stehen.

Ebenfalls danken möchte ich Dr. Stan Lai für die nette Zusammenarbeit und die oft lustigen Gespräche sowie für das Korrekturlesen von großen Teilen dieser Arbeit.

Matti, Ralf und Inga danke ich für das last minute Korrekturlesen.

Bei meinen Freunden und Bekannten entschuldige ich mich für mein Eremitentum in letzter Zeit.

Danken möchte ich auch Marcus dafür, daß er mir den Rücken freigehalten hat so gut es ihm möglich war.

Ich danke auch meiner Mutter und meiner Schwester für ihr Verständnis - ich verspreche mich bald wieder öfter zu melden!

Und zuletzt muß ich dem liebsten Menschen auf der ganzen Welt danken, für ihr unerschöpfliches Verständnis, ihre fortwährende Unterstützung in den kleinen und großen Dingen des Alltags und ihre mir entgegengebrachte Liebe - Monika, ich liebe dich!

# Zusammenfassung

Im Sommer dieses Jahres wird der Large Hadron Collider (LHC) am CERN Forschungszentrum nahe Genf nach langjähriger Bauzeit den Betrieb aufnehmen, und Protonen bei einer Schwerpunktsenergie von 14 TeV/c mit instantanen Luminositäten von bis zu  $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  zur Kollision bringen. Vier Detektoren befinden sich an den Wechselwirkungspunkten, um die bei den Kollisionen entstehenden Teilchen zu vermessen und dadurch Rückschlüsse auf die bei der Kollision abgelaufenen Prozesse zu ermöglichen. Einer dieser Detektoren ist der ATLAS Detektor, der als Allzweckdetektor konzipiert ist um einen möglichst großen Bereich von physikalischen Fragestellungen abdecken zu können.

Zu den wichtigen Fragen, die man durch die Experimente am LHC hofft klären zu können, gehört die Frage, ob das Higgs Boson wie es im Standardmodell der Teilchenphysik vorhergesagt wird in der Natur existiert. Die bisherigen experimentellen Daten deuten hierbei auf ein Higgs Boson mit einer Masse unterhalb von  $160 \text{ GeV}/c^2$ .

Auch wird die Suche nach neuer Physik, welche über das Standardmodell hinausgeht einen wichtigen Aspekt der Arbeit darstellen. Verschiedene supersymmetrische Modelle sind vielversprechende Kandidaten für solch neue Physik.

In beiden Fällen können Signaturen mit  $\tau$  Leptonen im Endzustand eine wichtige Rolle spielen, weshalb eine effiziente Rekonstruktion und Identifikation von  $\tau$  Leptonen einen entscheidenden Beitrag zur Untersuchung dieser Phänomene liefert.

In dieser Arbeit wurden die existierenden Methoden und Algorithmen zur Rekonstruktion und Identifikation von semileptonisch zerfallenden  $\tau$  Leptonen im ATLAS Experiment hinsichtlich ihrer Leistungsfähigkeit untersucht. Es wurden Verbesserungen an den Algorithmen vorgenommen, die zu einer effektiveren Rekonstruktion von  $\tau$  Leptonen mit kleinen transversalen Impulsen geführt haben. Desweiteren konnte auch die Fähigkeit zur Unterscheidung von  $\tau$  Leptonen und Jets in diesem Bereich gesteigert werden. Jets werden durch die starke Wechselwirkung produziert und stellen den größten Untergrund bei der Rekonstruktion von  $\tau$  Leptonen dar.

# Erklärung

Hiermit erkläre ich, dass ich die vorgelegte Arbeit selbständig verfasst und nur die angegebenen Quellen und Hilfsmittel verwendet habe.

Freiburg, den 12. Juni 2008

Nico Meyer