Searching for the Higgs Boson in Pairs of Tau Leptons in Data from the ATLAS Experiment

Peter L. Rosendahl

Dissertation for the degree of Philosophiae Doctor (PhD)

Department of Physics and Technology
University of Bergen

October 2012

Abstract

One of the key questions in particle physics today, is the origin of the electroweak symmetry breaking. The answer to this question will most likely be solved with the data provided by the Large Hadron Collider which started colliding protons in 2008. Many ideas have been posed to how particles gain their masses. The most promising of these ideas is the Higgs mechanism which predicts the existence of a new massive scalar boson, the Higgs boson.

Since the discovery of a new particle consistent with a Standard Model Higgs boson was made on July 4 by the ATLAS and CMS experiment, the solution for the puzzle of the electroweak symmetry breaking might be very near. However, in order to fully claim a discovery of *the* Standard Model Higgs boson, the new particle has to be proven to be a scalar boson and its decay has to be observed in both bosonic and fermionic final states with the corrected branching ratios predicted by the Standard Model. So far the new boson has only been seen in the bosonic $\gamma\gamma$, ZZ and WW final states, and the new particle is yet to be observed in any fermionic final state. Among the fermionic final states, the largest sensitivity is expected in $\tau^+\tau^-$ and $b\bar{b}$.

In this thesis, a search for the Standard Model Higgs boson in the fully hadronic $\tau^+\tau^-$ final state is presented. In the selection of Higgs candidate events, a technique for recreating the rest frame of the τ-pairs is shown to be a powerful tool for discriminating between Higgs and background events. Using the reconstructed rest frame, a new mass estimation technique is developed and compared to the existing methods. Finally, an updated limit on the Higgs production cross section is derived using the data recorded by ATLAS during 2011, and compared to the existing ATLAS limit.

The thesis is structured as follows: In the first part a brief introduction to the Standard Model including the Higgs mechanism is given, followed by a description on the production of Higgs bosons at the LHC in the Standard Model. The second part describes the ATLAS experiment in general, but a special emphasis is put on the SCT prompt calibration loop, which a great deal of the work of this thesis has gone into developing. The last part describes in the analysis of τ-pairs using the ATLAS detector. First, general description of the newly developed method for the rest frame reconstruction is given, followed by several chapters on the search for the Higgs boson in the $\tau^+\tau^-$ final state. Here, the attention is put on mass estimation, background normalisation and evaluation of the theoretical uncertainties related

to categorisation of events based on jet kinematics.

Acknowledgements

Needless to say that I could not have completed the work for this thesis without the help and support for a number of people. It would not be possible to mention all of them here and for that I apologise.

First of all, I would like to thank my supervisor Prof. Bjarne Stugu for providing valuable guidance and support during this thesis, but also for giving me an incredible freedom in my work. Both the personal support and freedom have been highly appreciated during some stressful times.

It has been truly a pleasure working in the Particle Physics group in Bergen, and I want to thank the whole the group at IFT for creating a nice atmosphere. A special thanks should go to Thomas Burgess who has taught me many valuable things on various subjects from computing and physics, to good coffee. Thanks to Alex Kastanas for his assistance with the local computing facilities in Bergen.

I am also grateful to my office mates, Kristian Gregersen and Ask Emil Løvschall-Jensen, at NBI. Thanks for your friendship and the many long and fruitful discussions. Especially thanks for the enormous work on reading my thesis. I sincerely hope I will be able to pay it back. I have really enjoyed the many interesting discussions on tau subjects with the whole tau community at NBI. Thanks to Mogens Dam, Stefania Xella, Lotte Thomsen and Almut Pingel. Also, a big thanks to Jørgen Beck Hansen for help with multiple statistics topics.

I would like the thank Junji Tojo, Patricia Ward and Steve McMahon for their guidance with my work on the SCT, and to Luc Goossens for helping me with the TMS system. Thanks to Elisabetta Pianori for her instructions on how to tackle the theoretical uncertainties.

Finally, I would like to thank my family for their love and unconditional support. I cannot express enough gratitude to my wife, Sif. Thank you for your patience and love, without it I would truly not have been able to complete this work.

Contents

List of Figures

List of Tables

Part I

The Higgs Boson

CHAPTER 1

The Standard Model of Particle Physics

The present knowledge of Particle Physics has been put together in a model referred to as the Standard Model (SM). Although this model describes the data collected by current and past experiments to an extreme level of accuracy, it leaves many open questions. In this chapter, the features and principles in the SM is presented with special emphasis on the electroweak sector and Higgs mechanism.

1.1 Principles of the Standard Model.

One of the modern ways of describing quantum mechanical phenomena in high energy physics is using the path integral formulation. One of the major attractions with this formulation is that it gives an elegant and efficient way of handling the quantisation of gauge theories. The basic element in this theoretical framework is the Lagrangian, which contains all the known information on the system at hand. Normally knowing the Lagrangian will enable you to derive the equation of motion of the fields contained in the system and calculate any desired quantity of the system. Unfortunately, only solutions to the path integral of very idealised Lagrangians are known.

A way to get around this problem is perturbation theory. In perturbation theory, the Lagrangian is often split up into a free part and an interaction part where the interaction part is assumed to be weak and therefore is perturbed. If this is true, the interactions can be expanded in an infinite series, which might be solvable. If a system is solvable, a set of rules describing the transition amplitudes for the possible processes, can be constructed. These rules are known as the Feynman rules [1].

1.1.1 The gauge principle

Even before quantum mechanics, it was known that the physics of a given system was closely related to the symmetries of the system. The conservation of momentum, energy and angular momentum could be obtained by invariance of the Lagrangian under transformations such as translation, time displacement or rotation. In high energy physics the symmetries of a system are also important, e.g. the conservation of electric or colour

charge is connected to invariance under other types of transformations, which are called phase or gauge transformations.

Gauge theories are based on the Lagrangian being invariant under such transformations. An arbitrary phase transformation of a particle field ψ has the form [1]

$$\psi(x) \rightarrow e^{i\alpha}\psi(x) \tag{1.1}$$

Since the theory should be invariant under this type of transformation, the phase α must be immeasurable and can be chosen arbitrarily. However, once it has been chosen, it has to remain fixed over all space and time. Therefore such transformation are also referred to as *global* phase transformations. Allowing α to be a function of space-time, $\alpha = \alpha(x)$ would lead to a larger group of transformations. These types of transformations are called *local* phase transformations. It is obvious that invariance under such types of transformations is a stronger requirement.

1.1.2 Quantum Electrodynamics and U(1) Gauge Invariance

A successful example of such a gauge theory is Quantum electrodynamics (QED), which is the relativistic quantum field theory of electrodynamics and part of the electroweak theory, which describes the electromagnetic and the weak force in a combined theory (see section 1.2.1). Due to the confirmations of the extraordinarily precise QED predictions, it has been used as a template of the construction of many other gauge theories.

QED is based on the theory of non-interacting electrons and positrons. As will be shown in this section, the interactions of electrons and photons will arise from requiring a gauge invariant Lagrangian. The electrons and positrons are described by a common spinor field $\psi(x)$. The Lagrangian for the non-interacting electrons and photons satisfying the Dirac equation is given by[1]

$$\mathcal{L}_{\text{Dirac}}(x) = \bar{\psi}(x)(i\gamma^{\mu}\partial_{\mu} - m)\psi(x). \tag{1.2}$$

It is easily verified that this Lagrangian is invariant under global phase transformations. However, this is not true in the case of local phase transformations, also called a U(1) gauge transformation. This type of transformation can be expressed as

$$\psi(x) \rightarrow e^{i\alpha(x)}\psi(x), \quad \text{and} \quad \bar{\psi}(x) \rightarrow e^{-i\alpha(x)}\bar{\psi}(x), \tag{1.3}$$

Here the term $\partial_{\mu}\psi$ will not satisfy the invariance principle. However, the Lagrangian can be made invariant under U(1) gauge transformations by substituting our ordinary derivatives with new (covariant) derivatives, which are defined as

$$\partial_{\mu} \rightarrow D_{\mu} = \partial_{\mu} + ieQ\mathrm{A}_{\mu}, \tag{1.4}$$

This claims the existence of a new gauge field A_{μ}, which has to transform in a manner that ensures the gauge invariance of the Lagrangian given in equation 1.2. This will lead to the following set of gauge transformations

$$\psi(x) \rightarrow e^{i\alpha(x)}\psi(x) \tag{1.5}$$

$$\mathrm{A}_{\mu}(x) \rightarrow \mathrm{A}_{\mu}(x) - \frac{1}{eQ}\partial_{\mu}\alpha(x). \tag{1.6}$$

where eQ is the coupling constant between fermions and the gauge field. Hence, the claim of local gauge invariance introduced interactions between the fermions and the gauge field. By rewriting the Lagrangian from equation 1.2, the new part due to the substitution of the derivatives can be identified as the interaction term

$$\begin{aligned}\mathcal{L}_{\text{Int}}(x) &= \bar{\psi}(x)(i\gamma^\mu D_\mu - m)\psi(x) - \bar{\psi}(x)(i\gamma^\mu \partial_\mu - m)\psi(x) \\ &= eQ\bar{\psi}(x)\gamma^\mu \psi(x)\mathrm{A}_\mu(x) \end{aligned} \tag{1.7}$$

With our prior knowledge of particle physics, the gauge field A_μ is identified as the photon field. A part corresponding to the dynamics of this field has also to be added to our Lagrangian. Since this part has to be gauge invariant and can not contain second order derivatives, the part is determined to be

$$\mathcal{L}_{\mathrm{A}}(x) = -\frac{1}{4}F_{\mu\nu}^2(x), \tag{1.8}$$

where $F_{\mu\nu}(x) = \partial_\mu \mathrm{A}_\nu(x) - \partial_\nu \mathrm{A}_\mu(x)$.
Hence, the full Lagrangian for QED is given by

$$\mathcal{L}_{\text{QED}}(x) = \bar{\psi}(x)(i\gamma^\mu \partial_\mu - m)\psi(x) - eQ\bar{\psi}(x)\gamma^\mu \psi(x)\mathrm{A}_\mu(x) - \tfrac{1}{4}F_{\mu\nu}^2(x) \tag{1.9}$$

1.2 The Standard Model

Our present knowledge of particle physics has lead to a model, called the Standard Model (SM), which describes the elementary particles and the principal forces of nature, except for Gravity. The SM contains two types of particles; the matter particles and the force carriers.

The matter particles, which are point-like, spin one-half fermions, are divided into two groups; quarks and leptons. Both groups have three almost identical generations, that only differ in their masses. The heavier generations eventually decay into the lightest generation, leaving only this generation present in our every-day matter, e.g. atoms, molecules etc. The other generations have been discovered in cosmic rays studies or at collider experiments.

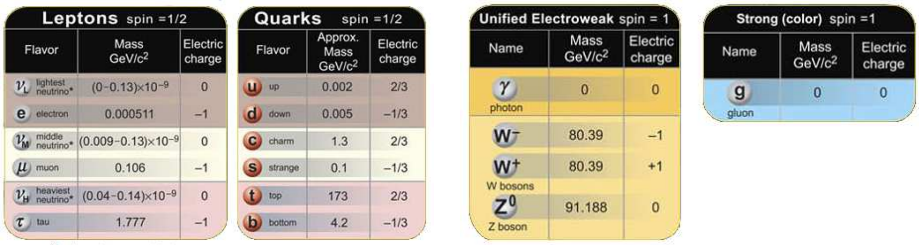
These particles interact due to three fundamental forces; the electromagnetic, the weak and the strong force. In quantum field theory (QFT), these forces mediate through point-like integer spin particles called the force carriers or gauge bosons. The large spread in the strengths and spacial extensions of these forces correspond to the fundamental quantities of their force carriers.

The force carrier of electromagnetism is the photon, which is massless and have neither electric nor colour charge. The photon couples to all electrically charged particles. Since the photon is massless, the electromagnetic force has an infinite range, which is why we are so familiar with this force in our every day live. The force carriers of the weak force are the W^+,W^- and Z^0 bosons, which couples to particles with flavour, i.e. all fermions. They are all massive particles without colour charge, but they differ in their charge. The W boson have electric charge of $\pm$ the electrons, while the Z^0 boson is neutral. Due to their masses, the weak force only works at very short ranges, e.g. inside the nucleus where it is responsible for radioactive decays.

The strong force have 8 force carries, called gluons, which couples to colour. This is the force that binds the nuclei together inside the nucleus, hence the name gluon.

The gluons are massless, but possess colour charge themselves, which makes them self interact. The massless gluons makes the strong force a long range force, but since it only couples to particles with colour charge and due to the colour confinement, the interactions between colours particles is only observed at very high energies or equivalent at very short distances.

The particle content of the SM and their properties is shown in figure 1.1.



Leptons spin =1/2

Flavor	Mass GeV/c²	Electric charge
ν_L lightest neutrino*	(0–0.13)×10⁻⁹	0
e electron	0.000511	−1
ν_M middle neutrino*	(0.009–0.13)×10⁻⁹	0
μ muon	0.106	−1
ν_H heaviest neutrino*	(0.04–0.14)×10⁻⁹	0
τ tau	1.777	−1

Quarks spin =1/2

Flavor	Approx. Mass GeV/c²	Electric charge
u up	0.002	2/3
d down	0.005	−1/3
c charm	1.3	2/3
s strange	0.1	−1/3
t top	173	2/3
b bottom	4.2	−1/3

Unified Electroweak spin = 1

Name	Mass GeV/c²	Electric charge
γ photon	0	0
W^-	80.39	−1
W^+ W bosons	80.39	+1
Z^0 Z boson	91.188	0

Strong (color) spin =1

Name	Mass GeV/c²	Electric charge
g gluon	0	0

Figure 1.1: Particle content of the Standard Model [2].

1.2.1 The Electroweak interaction and SU(2)×U(1) Gauge Theory

The standard theory of electroweak interactions is based on the Glashow-Weinberg-Salam model [1], in which four gauge bosons, the photon, the Wś and the Z^0 are introduced. Glashow unified the electromagnetic and the weak force by claiming that this theory could be derived from the SU(2)×U(1) gauge group. However, since the masses had to be put in by hand the gauge invariance was destroyed. The generation of masses by the Higgs mechanism[1] was introduced later on by Weinberg and Salam. However, the theory was first fully accepted when t'Hooft in 1971 showed that it was renormalisable.

The reason for choosing this particular gauge group to describe the electroweak interactions is not obvious. However, it has been shown to be extremely accurate in its predictions and it is the simplest group that reproduces the known features of the electroweak physics. Therefore it is widely accepted that the correct gauge group is $\mathrm{SU}(2)_T \times \mathrm{U}(1)_Y$, where T is the weak isospin and Y the hypercharge. Since the weak force acts differently on left and right-handed particles, the fermions fields are split into a left-handed ψ^{L} and a right-handed part ψ^{R}. The weak isospin is then defined as $\frac{1}{2}$ for left-handed fermions, which are grouped in SU(2) doublet, and 0 for right-handed fermions, grouped in SU(2) singlets. This transformation for the different field are given by

$$\begin{aligned} \psi^{\mathrm{L}}(x) &= \tfrac{1}{2}(1-\gamma_5)\psi(x) \rightarrow e^{i\,\tau_j\,\alpha^j(x)+i\,\mathrm{Y}\,\beta(x)}\psi^{\mathrm{L}}(x) \\ \psi^{\mathrm{R}}(x) &= \tfrac{1}{2}(1+\gamma_5)\psi(x) \rightarrow e^{i\,\mathrm{Y}\,\beta(x)}\psi^{\mathrm{R}}(x) \end{aligned} \tag{1.10}$$

where τ_j, $j = 1, 2, 3$, are the three generators of SU(2) and Y is the one generator of U(1).

The conservation of the weak isospin and the hypercharge is ensured by the gauge invariance and for a given fermion these quantum numbers determines uniquely how the fermion interacts with other particles in the electroweak theory. Yet, the quantum

Particles			T	T_3	Q	Y	B	L
$\begin{pmatrix}\nu_e\\ e^-\end{pmatrix}_{\rm L}$	$\begin{pmatrix}\nu_\mu\\ \mu^-\end{pmatrix}_{\rm L}$	$\begin{pmatrix}\nu_\tau\\ \tau^-\end{pmatrix}_{\rm L}$	$1/2$ $1/2$	$1/2$ $-1/2$	0 -1	-1 -1	0 0	1 1
$\begin{pmatrix}u\\ d'\end{pmatrix}_{\rm L}$	$\begin{pmatrix}c\\ s'\end{pmatrix}_{\rm L}$	$\begin{pmatrix}t\\ b'\end{pmatrix}_{\rm L}$	$1/2$ $1/2$	$1/2$ $-1/2$	$2/3$ $-1/3$	$1/3$ $1/3$	$1/3$ $1/3$	0 0
$e^-_{\rm R}$	$\mu^-_{\rm R}$	$\tau^-_{\rm R}$	0	0	-1	-2	0	1
$u_{\rm R}$	$c_{\rm R}$	$t_{\rm R}$	0	0	$2/3$	$4/3$	$1/3$	0
$d_{\rm R}$	$s_{\rm R}$	$b_{\rm R}$	0	0	$-1/3$	$-2/3$	$1/3$	0

Table 1.1: The fermion quantum number in the electroweak theory with hypercharge defined as $Y = 2(\mathrm{Q} - T_3)$.

number of electromagnetism, the electric charge Q, has to be related to the generator τ and Y in order to have a combined theory of the weak and electromagnetic force. Observations lead to the following relation

$$Q = T_3 + \frac{Y}{2}. \tag{1.11}$$

Furthermore, the fermions are given two additional quantum numbers, the baryon number B and lepton number L. The quarks are given $B = 1/3$ and $L = 0$, while leptons are given $B = 0$ and $L = 1$. Hence, this allows us to specify the quantum numbers in the electroweak theory for the fermions as shown in table 1.1. Just as it was the case for QED, the Lagrangian for the electroweak theory can now be constructed by requiring SU(2)×U(1) gauge invariance. The gauge invariance ensures the conservation of all the electroweak quantum numbers from electroweak interactions, which means that leptons and quarks can not disappear or emerge. The gauge invariance leads to new covariant derivatives of the form

$$\partial_\mu \to D_\mu = \partial_\mu + i\, g\, \tau_i \mathcal{W}^i_\mu + i\, g' \frac{\mathrm{Y}}{2} \mathcal{B}_\mu \tag{1.12}$$

where g, g' are the couplings between the fermions and the gauges fields. One gets the Lagrangian for the electroweak theory[1]

$$\begin{aligned} \mathcal{L}_{\rm EW} &= i\left(\bar\psi^{\rm L}_l \not{D} \psi^{\rm L}_l + \bar\psi^{\rm R}_l \not{D} \psi^{\rm R}_l\right) \\ &\quad -\frac{1}{4}\mathcal{W}^i_{\mu\nu}\mathcal{W}^{\mu\nu}_i - \frac{1}{4}\mathcal{B}_{\mu\nu}\mathcal{B}^{\mu\nu} \end{aligned} \tag{1.13}$$

where $\psi^{\rm L}$ and $\psi^{\rm R}$ denotes the left- and right-handed fermion fields, respectively, and $\not{D} = \gamma^\mu D_\mu$. Here, the first term is the kinetic energy and the interaction with the gauge field for left-handed fermions, while the second is the same for right-handed fermions. These two terms is henceforth denoted as $\mathcal{L}_{\rm L}$. The last two terms are the kinetic energy and the self-interaction of the gauges field, and will be denoted as $\mathcal{L}_{\rm B}$. The field strength tensors of these are given by

$$\begin{aligned} \mathcal{B}_{\mu\nu} &= \partial_\mu \mathcal{B}_\nu - \partial_\nu \mathcal{B}_\mu \\ \mathcal{W}^i_{\mu\nu} &= \mathcal{F}_{\mu\nu}{}^i - g\varepsilon_{ijk}\mathcal{W}^j_\mu \mathcal{W}^k_\nu \\ \mathcal{F}_{\mu\nu}{}^i &= \partial_\mu \mathcal{W}^i_\nu - \partial_\nu \mathcal{W}^i_\mu \end{aligned} \tag{1.14}$$

[1]Hereafter the (x) argument is omitting in the expressions.

However, since the $\mathcal{W}^i_{\mu\nu}$ and $\mathcal{B}_{\mu\nu}$ is introduced as massless, they still need to be related to the observed physical particles W, Z^0 and γ. This is done by the following linear transformation determined by the Higgs mechanism

$$\begin{aligned} \begin{pmatrix} W^+_\mu \\ W^-_\mu \end{pmatrix} &= \begin{pmatrix} 1/\sqrt{2} & -i/\sqrt{2} \\ 1/\sqrt{2} & i/\sqrt{2} \end{pmatrix} \cdot \begin{pmatrix} \mathcal{W}^1_\mu \\ \mathcal{W}^2_\mu \end{pmatrix}, \\ \begin{pmatrix} Z^0_\mu \\ \mathrm{A}_\mu \end{pmatrix} &= \begin{pmatrix} \cos\theta_\mathrm{W} & -\sin\theta_\mathrm{W} \\ \sin\theta_\mathrm{W} & \cos\theta_\mathrm{W} \end{pmatrix} \cdot \begin{pmatrix} \mathcal{W}^3_\mu \\ \mathcal{B}_\mu \end{pmatrix}, \end{aligned} \tag{1.15}$$

where θ_W, called the weak mixing angle or the Weinberg angle, is defined as

$$\cos\theta_\mathrm{W} = \frac{g}{\sqrt{g^2+g'^2}}, \quad \sin\theta_\mathrm{W} = \frac{g'}{\sqrt{g^2+g'^2}}. \tag{1.16}$$

Now, writing the Lagrangian from equation 1.13 as $\mathcal{L}_\mathrm{L} + \mathcal{L}_\mathrm{B}$, the boson interaction part can be rewritten as

$$\begin{aligned} \mathcal{L}_\mathrm{LB} &= -g\bar{\psi}^\mathrm{L}\gamma^\mu\tau_i\mathcal{W}^i_\mu\psi^\mathrm{L} - g'\bar{\psi}\gamma^\mu\frac{Y}{2}\mathcal{B}_\mu\psi \\ &= -\left(\frac{g}{\sqrt{2}}\left(J^\mu_\mathrm{cc}W^+_\mu + J^{\mu\dagger}_\mathrm{cc}W^-_\mu\right) + g\, J^\mu_\mathrm{em}\mathrm{A}_\mu \right. \\ &\qquad \left. + \sqrt{g^2+g'^2}J^\mu_{nc}Z^0_\mu\right), \end{aligned} \tag{1.17}$$

where the neutral current J_nc, the charged currents J_cc and the electromagnetic current J_em can be expressed as the following by using the Pauli spin step operators τ^+ (τ^-).

$$\begin{aligned} J^\mu_\mathrm{cc} &= \bar{\psi}^\mathrm{L}\gamma^\mu\tau^+\psi^\mathrm{L}, \\ J^\mu_\mathrm{em} &= \bar{\psi}\gamma^\mu Q\sin\theta_\mathrm{W}\psi, \\ J^\mu_\mathrm{nc} &= \bar{\psi}\gamma^\mu\left(\frac{1}{2}\left(1-\gamma_5\right)T_3 - Q\sin^2\theta_\mathrm{W}\right)\psi \end{aligned} \tag{1.18}$$

Hence, the gauge field A_μ is identified as the photon, since it only couples to electric charge[2] and does not distinguish between the left- and right-handed particles. Furthermore, the three remaining gauge fields couple to the fermions exactly the way they should for us to identify them as the experimentally discovered particles.

The Z^0 also couples to electric charge, but in addition it also couples to flavour, hence allowing it to decay into neutrinos. The W only couples to the left-handed fermions and is responsible for the flavour changing decays (seen by the spin step operator), such as $d' \to u$. These types of decays can only take place inside the doublets, but since the weak isospin doublets is made of gauge eigenstates, that are slightly rotated with respect to the mass eigenstates, the weak interactions can cause decays between all types of flavours.

The relation between the mass eigenstates and the gauge eigenstates is given by introducing a 3×3 unitary matrix, V_{CKM}, known as the Cabbibo-Kobayashi-Maskawa matrix, V_{CKM}. This matrix connects the gauge eigenstates, denoted d', s', and b', to the mass eigenstates, denoted d, s, and b.

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_\mathrm{CKM}\cdot\begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}\cdot\begin{pmatrix} d \\ s \\ b \end{pmatrix}, \tag{1.19}$$

[2]Provided that $e = g\sin\theta_\mathrm{W}$

where the u, c, and t quarks by convention are unmixed. An element V_{xy}, where $x \in \{u, c, t\}$ and $y \in \{d, s, b\}$ in the V_{CKM} matrix describes the transition amplitudes between the different quark eigenstates.

1.2.2 The Higgs Mechanism

Although the Lagrangian in equation 1.13 describes all the electroweak interactions some of which has been confirmed by observations to an extremely high precision, it is clearly not satisfactory since all particles are massless – in contrast to what has been observed by experiments. Adding mass terms to the Lagrangian by hand turns out to break gauge invariance. Furthermore, masses for the gauge bosons have to be added in a way which give masses to the W and Z^0 while the photon remains massless.

A way to achieve this, is the *Higgs mechanism*. This consists of adding new *Higgs* fields, Φ, arranged in a SU(2) doublet,

$$\Phi(x) \;=\; \begin{pmatrix} \phi_a(x) \\ \phi_b(x) \end{pmatrix} \tag{1.20}$$

with a non-vanishing vacuum expectation value $\langle 0|\Phi|0\rangle \neq 0$. The Φ doublet is invariant under SU(2)×U(1) gauge transformations, and transforms as the weak isospin doublet, ψ^{L} following the transformations given in equation 1.10.

The Φ field will add an additional terms, $\mathcal{L}_{\mathrm{H}}$, to the electroweak Lagrangian given in equation 1.13 of the form

$$\mathcal{L}_{\mathrm{H}} = (D_\mu \Phi)^\dagger (D^\mu \Phi) - \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2 \tag{1.21}$$

The parameter λ has to be positive for Higgs potential to be bounded from below. For $\lambda > 0$ and $\mu^2 < 0$, the ground state of Φ will be degenerate with a potential minimum at $\Phi^\dagger \Phi = \frac{\mu^2}{2\lambda}$. Hence, the ground state can be chosen as

$$\Phi_0 = \begin{pmatrix} \phi_a^0 \\ \phi_b^0 \end{pmatrix} = \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix}, \tag{1.22}$$

where $v = \sqrt{-\mu^2/\lambda}$, since any other choice of Φ_0 can be transformed into Φ_0 by a global phase transformation.

The ground state, Φ_0, is not invariant under SU(2)×U(1), and hence the gauge symmetry is broken spontaneously. However, Φ_0 has to be invariant under U(1) in order to preserve conservation of electric charge and keep the photon massless. This can be done by assigning weak hypercharge 1 to Φ, thus from equation 1.11 it follows that the lower component, ϕ_b, has to be neutral.

The Higgs fields can now be expressed as expansion around the ground state in four real fields, σ, η_1, η_2 and η_3 as

$$\Phi(x) \;=\; {}^1/\sqrt{2} \begin{pmatrix} \eta_1(x) + i\,\eta_2(x) \\ v + \sigma(x) + i\,\eta_3(x) \end{pmatrix} \tag{1.23}$$

By substituting equation 1.23 into equation 1.21, its seems that upon quantisation of the Lagrangian we are left with four real fields. However, by studying the Lagrangian in the unitary gauge [3], it can be shown the only the field, σ will give rise to a real massive

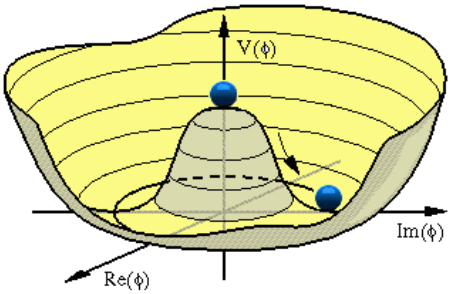


Figure 1.2: The Higgs potential: The *Mexican hat*, $\mu^2 > 0$.

scalar whereas η_1, η_2 and η_3 are unphysical fields and will disappear. The massive scalar created by σ is called the *Higgs boson*.

In the unitary gauge, $\mathcal{L}_{\mathrm{B}} + \mathcal{L}_{\mathrm{H}}$ expressed in the physical fields Z and W, can now be written as

$$\begin{aligned}
\mathcal{L}_{\mathrm{B}} + \mathcal{L}_{\mathrm{H}} &= -\tfrac{1}{4} F_{\mu\nu} F^{\mu\nu} \\
&\quad -\tfrac{1}{2} \mathcal{F}_{\mu\nu} \mathcal{F}^{\mu\nu} + m_W^2 \mathrm{W}_\mu^\dagger \mathrm{W}^\mu \\
&\quad -\tfrac{1}{4} \mathrm{Z}_{\mu\nu} \mathrm{Z}^{\mu\nu} + m_Z^2 \mathrm{Z}_\mu^\dagger \mathrm{Z}^\mu \\
&\quad +\tfrac{1}{2} (\partial^\mu \sigma)(\partial_\mu \sigma) - \tfrac{1}{2} m_H^2 (\sigma)^2 \\
&\quad +\mathcal{L}_{\mathrm{BB}} + \mathcal{L}_{\mathrm{HH}} + \mathcal{L}_{\mathrm{HB}} \qquad (1.24)
\end{aligned}$$

$\mathcal{L}_{\mathrm{BB}}$, $\mathcal{L}_{\mathrm{HH}}$ and $\mathcal{L}_{\mathrm{HB}}$ terms incorporates the interactions among the bosons and $g = {}^{2}/_{\sin\theta_{\mathrm{W}}}$. Most notable are the Higgs self couplings given in

$$\mathcal{L}_{\mathrm{HH}} = -\tfrac{1}{4} \lambda \sigma^4 - \lambda v \sigma^3 \qquad (1.25)$$

and the Higgs coupling to the Z^0 and W bosons given by

$$\begin{aligned}
\mathcal{L}_{\mathrm{HB}} &= \tfrac{1}{2} \tfrac{m_W^2}{v} \mathrm{W}_\mu^\dagger \mathrm{W}^\mu \sigma + \tfrac{1}{4} g^2 \mathrm{W}_\mu^\dagger \mathrm{W}^\mu \sigma^2 \\
&\quad \tfrac{1}{2} \tfrac{m_Z^2}{v} \mathrm{Z}_\mu^\dagger \mathrm{Z}^\mu \sigma + \tfrac{1}{4} g^2 \mathrm{Z}_\mu^\dagger \mathrm{Z}^\mu \sigma^2 \qquad (1.26)
\end{aligned}$$

which states that the Higgs coupling to the gauge bosons is proportional to the boson mass squared.

The mass parameters given in equation 1.24 are defined as

$$m_W = \frac{1}{2} v g, \quad m_Z = \tfrac{m_W}{\cos\theta_{\mathrm{W}}}, \quad m_H = 2 v^2 \lambda \qquad (1.27)$$

Hence, the non-zero vacuum expectation value of Φ generates masses to the $W^\pm$ and Z^0 and leaves the photon massless. However, an additional term, $\mathcal{L}_{\mathrm{LH}}$, has to added to the Lagrangian in order to generate masses for the leptons.

$$\begin{aligned}
\mathcal{L}_{\mathrm{LH}} &= -g_l \left(\bar{\psi}_l^{\mathrm{L}} \psi_l^{\mathrm{R}} \Phi + \Phi^\dagger \bar{\psi}_l^{\mathrm{R}} \psi_l^{\mathrm{L}} \right) \qquad (1.28) \\
&\quad -g_{\nu_l} \left(\bar{\psi}_l^{\mathrm{L}} \psi_{\nu_l}^{\mathrm{R}} \tilde{\Phi} + \tilde{\Phi}^\dagger \bar{\psi}_{\nu_l}^{\mathrm{R}} \psi_l^{\mathrm{L}} \right) \qquad (1.29)
\end{aligned}$$

where g_l and g_{ν_l} are dimensionless coupling constants and $\tilde{\Phi}$ is defined as

$$\tilde{\Phi} = \begin{pmatrix} \phi_b^* \\ -\phi_a^* \end{pmatrix} \tag{1.30}$$

This way of generating masses to the fermions is called Yukawa interactions.

In the unitary gauge, the part of the Lagrangian describing interactions between the Higgs boson and fermions, can now be expressed in terms of the physical fields as

$$\mathcal{L}_{\mathrm{HL}} = -\tfrac{1}{v} m_l \bar{\psi}_l \psi_l \sigma - \tfrac{1}{v} m_{\nu_l} \bar{\psi}_{\nu_l} \psi_{\nu_l} \tag{1.31}$$

where m_l and m_{ν_l} denotes the fermion masses. Hence the Higgs couplings to the fermions are proportional to the fermion mass.

The final Lagrangian of the EW theory including the Higgs mechanism can now be expressed as a free part $\mathcal{L}_0$ and a series of interaction terms

$$\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_{\mathrm{LB}} + \mathcal{L}_{\mathrm{BB}} + \mathcal{L}_{\mathrm{HH}} + \mathcal{L}_{\mathrm{LH}} \tag{1.32}$$

The existence of the Higgs boson is crucial to the success of the EW theory, since higher order Feynman diagrams containing Higgs bosons ensures the renormalisability of the theory. The size of these corrections depends on the Higgs mass, and therefore a theoretical upper limit on m_H can be set around 1 TeV to explain the success of perturbative calculations of EW processes.

1.2.3 Quantum Chromo Dynamics

The basics of the EW theory has now been demonstrated, but we have not yet introduced the strong force. The gauge group of the strong force is SU(3), which means that the strong force have 8 gauge bosons, since there are 8 generators of SU(3). Hence, the gauge group of the full Standard Model Lagrangian is U(1)×SU(2)×SU(3).

As mentioned earlier the gauge bosons of the strong force, called gluons, couples to colour. The quarks possess one of the three colours red, blue or green, while the anti-quarks possess anti-red, anti-blue or anti-green). The gluons possess both a colour and a anti-colour and cause the colour changing of the quarks. This also means that the gluons couple to themselves.

The theory of the strong force is called Quantum Chromo Dynamics (QCD). In QCD, the coupling constant that controls the interactions between coloured particles, α_{S} runs, i.e. α_{S} changes value as a function of the momentum transfer Q^2 in the interaction. For large values of Q^2 the coupling constant becomes weak, which means the quarks behaves like free particles, while for small values of Q^2 α_{S} diverges due to the self coupling of the gluon. The large value of α_{S} at low Q^2 causes perturbation theory to break down, hence a lower cut-off scale, Λ has to be introduced. Below this cut-off perturbation theory is no longer valid. Experimental measurements show that $\Lambda \approx 200$ MeV, so perturbation theory will only be valid well above this scale. A typical limit for Λ used in most generators is 1 GeV at which $\alpha_{\mathrm{S}} \approx 0.4$.

Another feature of the strong force is colour confinement, which states that only colour neutral particles can be observed. Therefore no free quarks can be observed, but only bound states such as mesons, consisting of a coloured quark and an anti-quark with the corresponding anti-colour, or baryons, consisting of three quark (or anti-quarks) with different colours.

1.3 Open questions in the Standard Model.

In this chapter the SM has been presented. However, the SM still leaves many open questions to be examined:

- Of the four fundamental forces of nature, only three are included in the SM. In order for gravity in be properly included in the SM, it has to be described by QFT and hence quantised. This has so far not been done successfully. Luckily, gravity is so much weaker than the other forces that it can be neglected for all experimental purposes. Although going to increasingly higher energies and hence smaller distances, gravity will eventually become important and hence has to be included in our calculations.

- This leads to the question: Why is gravity so much weaker than the other forces of nature? From the scale of the electroweak symmetry breaking ($M_{EW} \leq 1$ TeV) to the scale where gravity is expected to become as strong as the other tree forces, the Planck scale ($M_{Pl} \sim 10^{19}$ GeV), is a huge gap of many orders of magnitude. Moreover, if we assume the SM to be valid up to the Planck scale, we would expect the mass of the Higgs boson to receive large radiative corrections from vacuum polarisation. This would make the Higgs mass many orders of magnitude larger than the currently restricted value, unless there is an tremendous amount of fine-tuning cancellations among the quantum corrections. It should be noted that extensive fine-tuning is not prohibited by any fundamental principle, although thought be to unnatural. This problem is known as the Hierarchy problem.

- Another question is: Why do there exist exactly three generations of matter particles?

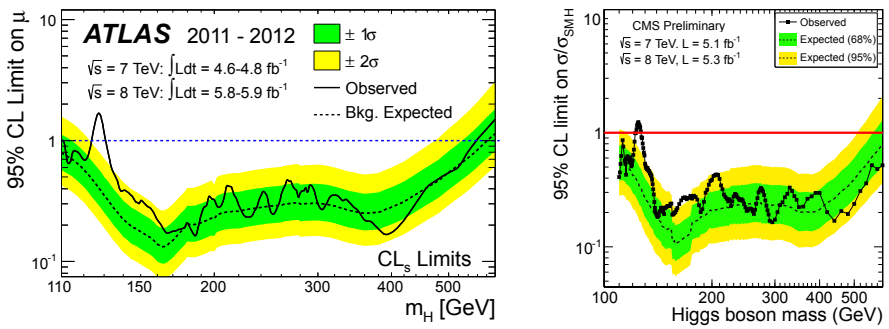
- The masses of the fermions are generated in the SM by the Yukawa couplings, but their values differ by many orders of magnitude and have to be put in by hand. The SM does not give any physical explanation why the masses of the fermions have such a large span.

- The mass of the neutrinos pose another problem. In the SM, the neutrinos are massless as shown in figure 1.1, however experiments [4] have discovered the neutrinos to indeed be massive particles by showing a mixing between the neutrino types. Yet, these types of measurements only determine the mass difference between the neutrinos with different flavour and not the precise mass. Severals ways exist to give masses to the neutrinos, however which of them is the correct only experiments can tell.

CHAPTER 2

Higgs production at the LHC

The discovery or exclusion of the Higgs boson has been named as one of the main goals of the LHC project [5]. At the startup of the LHC, the possible mass of the Higgs had a lower bound at 114 GeV set by the LEP experiments [6] and a theoretical upper limit around 1 TeV as mentioned in section 1.2.2. Recently, these bounds have been tighten significantly, leaving open only a small range in which a new particle consistent with the SM Higgs has been discovered as will be mentioned later. Figure 2.1 shows the recent exclusion plots from both ATLAS [7] and CMS [8] where the Higgs boson has been excluding for all mass points where the solid lines are below 1. The construction of these plots is described in section 9.1.



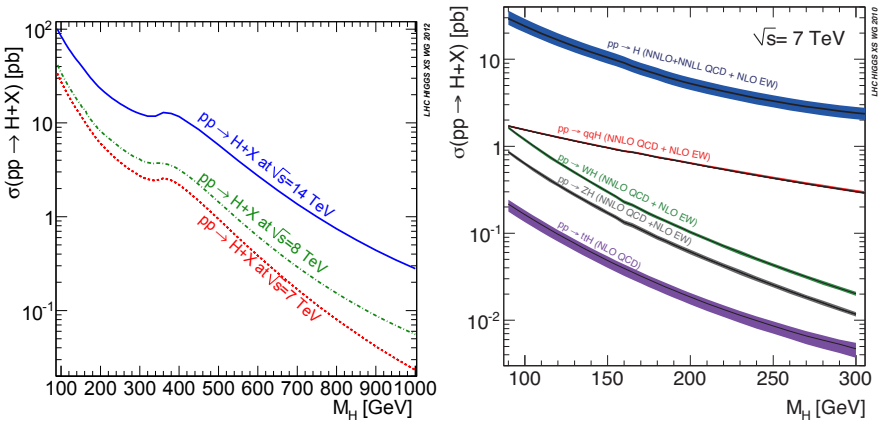
(a) ATLAS [7]

(b) CMS [8]

Figure 2.1: Higgs exclusion limits combined across channels. Observed (solid) and expected (dashed) 95% CL upper limit on the Higgs cross section as a function of m_H, scaled to the SM Higgs cross section. The expected limits are derived under the background-only hypothesis, e.g. without the SM Higgs. The filled bands shows the 1 (2) σ uncertainty bands on the expected limit.

2.1 Higgs production mechanisms

The production mechanism and decay of the Higgs boson is strongly dependent on its mass. As illustrated in figure 2.2(a) the inclusive cross section for $pp \rightarrow H + X$ is rapidly decreasing with increasing m_H. The inclusive cross section is also very dependent on the centre-of-mass energy in the collisions. In figure 2.2(b), the inclusive Higgs production is split into the most significant production mechanisms. The largest contribution comes from the gluon-fusion process, ggH, where two gluons produces a Higgs boson through a quark-loop. Since the Higgs coupling to the quarks is proportional to the mass of the quark as illustrated in equation 1.31, the only significant contribution comes from the diagram with the top quark in the quark loop.



(a) Higgs cross section for different centre-of-mass energies

(b) Higgs cross section for different production modes

Figure 2.2: Inclusive Higgs production cross section as function of the Higgs mass, m_H.

2.1.1 Gluon Fusion production

The leading-order Feynman diagram for ggH is shown in figure 2.3(a). Since this process is controlled by the strong dynamics, the next-to-leading-order contributions to this process are large which means that the Higgs is often produced in association with one or more jets when produced through gluon-fusion. It also means that the theoretical uncertainty on the cross section from QCD scales are large. This will be discussed in detail in chapter 8. Figure 2.3(b)-(d) shows the Feynman diagrams for Higgs+1jet production through ggH.

2.1.2 Weak production

The second most dominant production mechanism for Higgs bosons at the LHC is the Vector-Boson-Fusion, VBF [9, 10]. The VBF is defined as the t- and u-channel Feynman diagrams shown in figure 2.4(a) and figure 2.4(b) where the Higgs boson is produced

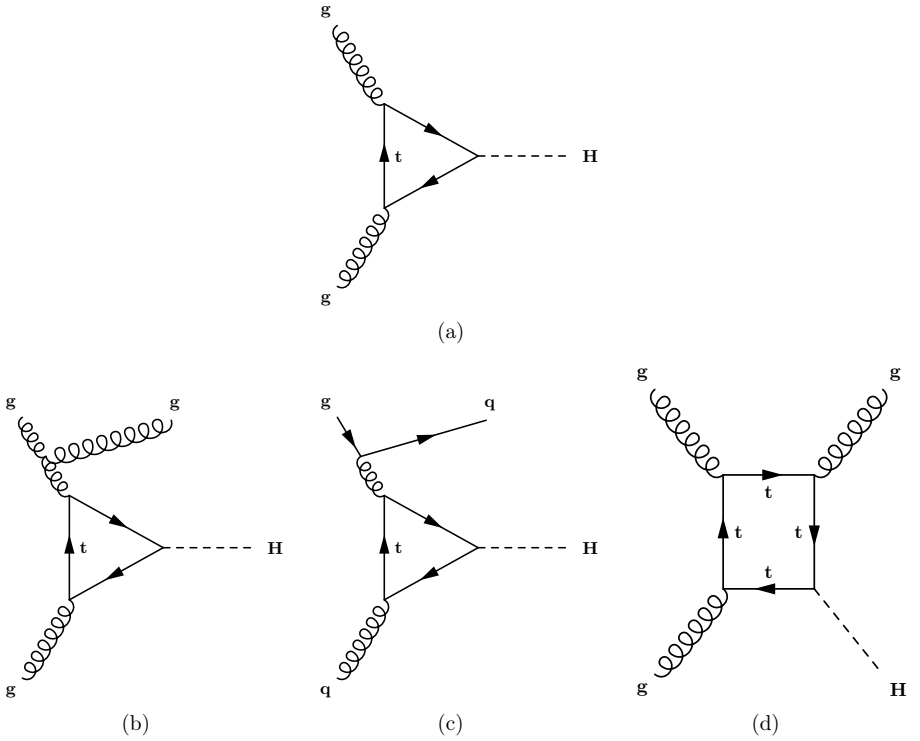


Figure 2.3: Feynman diagrams for Higgs production through gluon-fusion. The leading-order diagram is shown in (a) and the next-to-leading-order diagrams with real emission are shown in (b)-(d).

in association with two hard forward jets through a pair of W or Z^0-bosons. However this process is only gauge invariant if the s-channel diagram, shown in figure 2.4(c), is included as well.

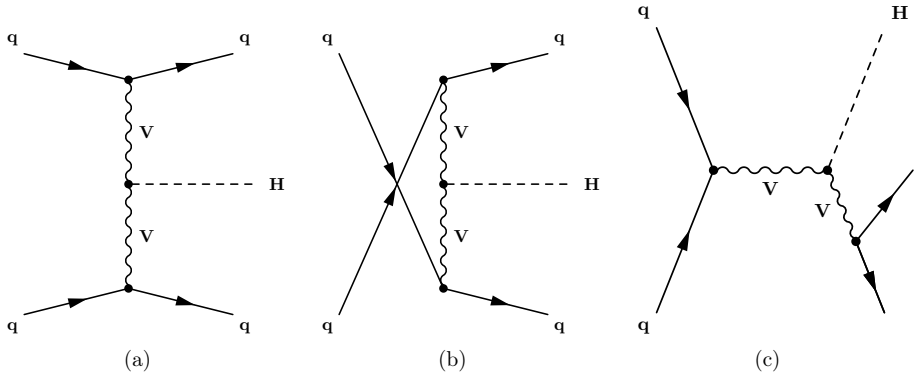


Figure 2.4: Feynman diagrams for Higgs production through weak interactions.

The two forwards jets in the VBF mode allow for a tagging signature of these events and can help discriminate VBF production from other processes [11]. Furthermore, the theoretical uncertainties associated with this process is much smaller than for ggH.

The VBF process is unique due to the colour-flow in the process as illustrated in figure 2.5. Both of the two associated jets will tend to be very forward in the detectors and each jet will be colour-connected with the beams remnants travelling in the same direction. This means that additional jets will tend to be produced in the forward regions and that the events will have a rapidity-gap in the central part of the detector without any hard jet activity.

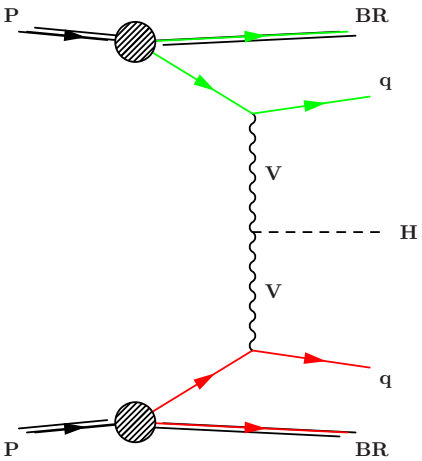


Figure 2.5: Example on colour-flow in VBF production. The colour-flows that appear between the upper and lower part of the diagram are separated which will result in a rapidity gap Each of the two forward jets will be colour-connected with the beam remnants.

The final diagram in figure 2.4(c) where the Higgs boson is produced in association with a Z^0 or W-boson is referred to as *associated production*, VH. Since VH and VBF production leaves two distinct signatures in the detector they are usually separated in analyses as two production mechanisms, although they are closely connected. As seen from figure 2.2(b) the VH cross section, is well below that of VBF, but the possibility of reconstructing the weak boson creates a tagging signature for this channel as well.

2.2 Higgs branching ratios

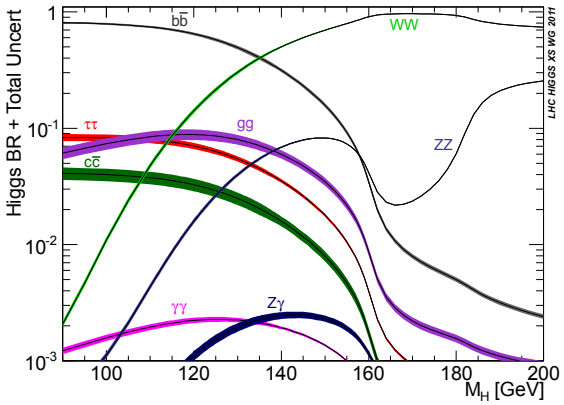


Figure 2.6: Higgs branching ratios, BR, to different final states as a function of the Higgs mass, m_H. The filled areas around the curves indicate the total uncertainties on each BR. [12]

In figure 2.6 the Higgs branching ratios, BR, are shown as a function of m_H in the low mass region. At higher values of m_H, only the decays to Z^0Z^0 and WW are non-negligible. Both these channels have relatively low backgrounds in the leptonic final states. Furthermore, the Z^0Z^0 offers a very good mass resolution which enhances the discovery potential and is crucial for determining the Higgs mass in case an excess is observed.

In the low mass region, the decay to a pair of τ-leptons or photons, $\gamma\gamma$, becomes important [13, 14, 11]. The $\gamma\gamma$ offers a very clean signature in the detector and like the Z^0Z^0 it has a good mass resolution. In the $\tau^+\tau^-$ channel the mass distribution is smeared due to the neutrinos, however the resolution is still better than in the WW channel. The Higgs decay into $b\bar{b}$ also has a high BR in the low mass region, however this signature is very hard to separate from the huge swamp of QCD-jet background events. Figure 2.7 shows the expected sensitivity from ATLAS [7] for the different decay channels.

Even though the sensitivity across the different channels has a large variation, observing a possible Higgs in several channels is crucial for determining the properties of the Higgs boson. Especially, observing the Higgs boson in both vector boson and fermion final states is important for measuring the Higgs couplings [15] and hence for concluding that an observed particle is indeed a SM Higgs boson.

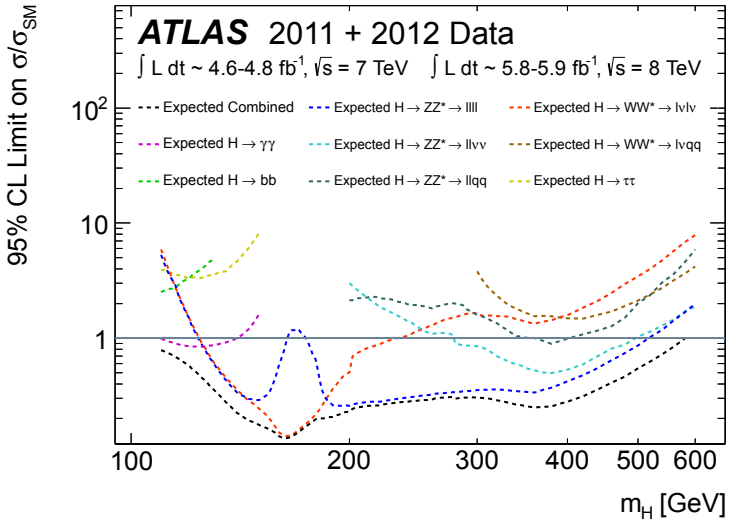


Figure 2.7: The expected 95% CL cross section upper limits for the different decay channels in ATLAS, normalised to the SM Higgs boson production cross section, as a function of m_H. An expected limit below one means that the experiment is sensitive to excluding the Higgs at 95% CL [7]

2.3 Recent discovery of a new Higgs-like particle

Recently, both CMS [16] and ATLAS [7] have announced the discovery of a new Higgs-like boson with a mass around 125 GeV. Figure 2.8 shows *local* p_0 *value*, i.e. the probability that the background through random fluctuations can produce larger signal[1] than the excess observed in data. The local p_0 value is shown as a function of the m_H, as it has been obtained by ATLAS over time. Clearly, the effect of the new particle is presently well above the 5σ value that by convention is needed to claim a discovery.

The statement that the observed new particle is a *Higgs-like boson* refers to following. First of all, it must be a boson since it has been seen decaying into two bosons (So far the WW,Z^0Z^0 and $\gamma\gamma$ have been used in the combination). Furthermore, it decays into two photons and therefore it must have spin 0 or 2. Secondly, the new boson has been observed in all the *sensitive* final states where a *sensitive* final state refers to a final state with an expected μ below 1.

However, so far the new Higgs-like boson have only been observed bosonic final states[2], and in order to conclude that the new particle indeed is the SM Higgs, we need to see it decays into fermions. And the most promising fermionic final states is the $H \to \tau\tau$ search channel.

[1] Here a larger signal means a larger likelihood

[2] Although, the new Higgs-like boson have only been observed in bosonic final states, the expected decay mode for $\gamma\gamma$ is through a quark loop. Hence, one could argue that a fermionic decay has already been observed.

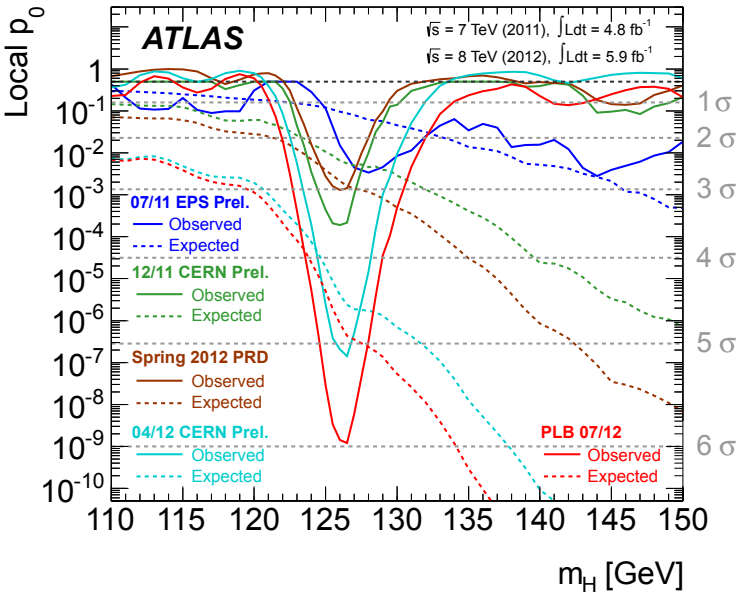


Figure 2.8: Combined search results: The observed (solid) local p0 as a function of mH and the expectation (dashed) for a SM Higgs boson signal hypothesis at the given mass. [7]

Part II

The ATLAS Experiment

CHAPTER 3

The ATLAS Detector

The ATLAS (A Toriodal Lhc ApparatuS) experiment is the largest of the four main experiments located at the Large Hadron Collider (LHC) at European Organisation for Nuclear Research CERN. It is designed to be a multi-purposed experiment and therefore the requirements for the detector also have a wide range. In this chapter the layout if the ATLAS experiment is described. However a special emphasis is put on the Semi Conductor Tracker and its prompt calibration, since a large part of the thesis work has been implementing and operating this procedure. Also the ATLAS computer model is described. Here, part of the thesis work has been acting as Production Manager of the SCT.

3.1 The Large Hadron Collider

The LHC [17] has been designed to collide bunches of up to 10^{11} of protons (p) at a rate of 40 million times per second generating pp-collisions with a centre-of-mass energy of 14 TeV and a instantaneous luminosity of $10^{34}\text{cm}^{-2}\text{s}^{-1}$ extending the energy and luminosity frontier with a factor of $\sim$ 10 and $\sim$ 100 in comparison to the previous hadron colliders [18]. In addition to colliding protons, the LHC also has a heavy ions program which will not be covered here.

Currently, the LHC is not at the stage in its running program where the peak design criteria are reached. Instead the LHC is currently running with a centre-of-mass energy of 8 TeV and a peak luminosity $\sim 8 \times 10^{33}$ $\text{cm}^{-2}\text{s}^{-1}$. Below, the major mile stones from the start up of the LHC until now are sketched.

10 Sep 2008:	First colliding protons at the LHC.
23 Nov 2009:	First collisions at 450 GeV in all 4 detectors.
30 Nov 2009:	First collisions at 1.18 TeV per beam[1].
30 Mar 2010:	First collisions at 7 TeV.
5 Apr 2012:	First collisions at 8 TeV.
4 July 2012:	First announcement on the observation of a new boson consistent with the Higgs boson.

In figure 3.1 the integrated luminosity of the LHC in the three years of data taking is shown as measured by the ATLAS online monitoring.

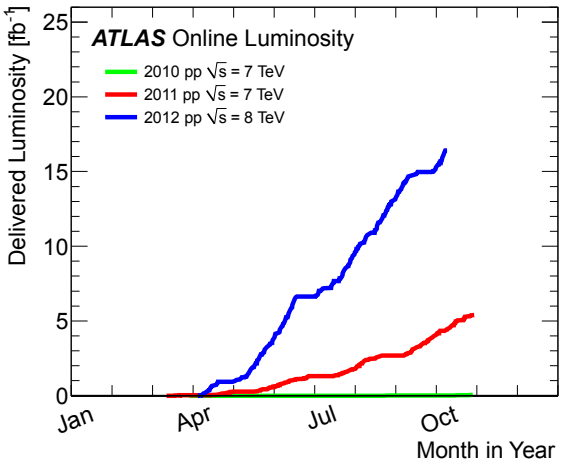


Figure 3.1: Integrated luminosity during the years 2010, 2011 and 2012 as measured in ATLAS by the online luminosity monitoring.[2]

The high interaction rates and the high beam energies at the LHC lead to a great potential for discovering new physics, but these also set the requirements to the detector in order to extract the few important events from the huge amount of background events. These requirement are:

- A high precision electromagnetic calorimeter for identification and reconstruction of electrons and photons
- A fully covering high precision hadronic calorimeter for identification and reconstruction of jets and missing transverse energy.
- A muon spectrometer for high-precision muon momentum measurements outside the calorimeters.
- A strong magnetic field to be increase the curvature of tracks from charge particles – especially muons.
- An efficient tracking for measuring the momentum and charge of high-p_{T} particles and enable reconstruction of particle vertices.
- Fast and radiation-hard electronics.
- A fast triggering and data-acquisition system.

The overall layout of the ATLAS detector – shown in figure 3.2 – consists of an Inner Detector (ID) responsible for the tracking of charged particles, two calorimeters responsible for measuring energies of particles and an outside muon system responsible for tracking of muons. The ID is surrounded by a thin super-conducting solenoid magnet generating the inner magnetic field while a large superconducting air-core toroid generates the magnet field in the outer part of the detector. A detailed description of the ATLAS detector is given in [19]

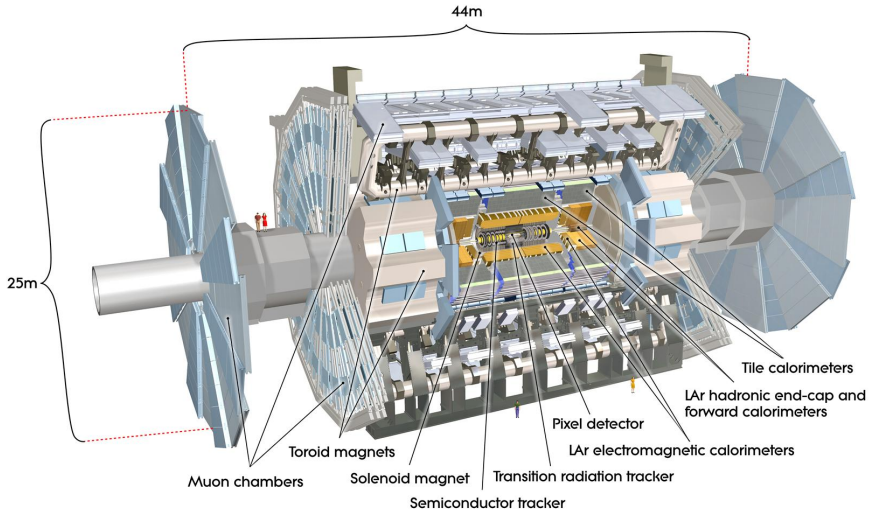


Figure 3.2: Overview of the ATLAS detector [19].

3.2 Coordinate System

In order to describe the detector in a consistent way, a common coordinate system is defined within the experiment. The origin of this coordinate system is taken to be the nominal interaction point. In cartesian coordinates, the z-axis is defined to be along the beam direction, the x-axis is pointing towards the centre of the LHC ring and the y-axis is pointing upwards. In spherical coordinates instead of the polar angle, θ, it is often convenient to use the pseudo-rapidity, η, defined as $-\ln(\tan\theta/2)$. Unless stated otherwise, the transverse variables, e.g. transverse momentum, p_{T}, are defined as the component in the $x-y$ plane. The measure of object separation ΔR is defined in the pseudorapidity-azimuthal angle space as $\Delta R = \sqrt{\phi^2 + \eta^2}$.

3.3 The Magnet System

As mentioned earlier the ATLAS magnet system is divided in two parts, a central solenoid and an outer air-core toroid. The central solenoid, surrounding the ID, generates a magnetic field parallel to the beam axis with a slightly varying field strength of 2 T at the centre of detector and a peak value of 2.6 T at the solenoid. The solenoid has a length of 5.3 m, an inner diameter of 2.44 m and an outer diameter of 2.63 m. Due to its placement in front of the calorimeters, it has to contain as little material as possible to get as precise an energy measurement as possible.

The air-core toroid consists of two end-cap toroid magnets and the barrel toroid magnet. The barrel toriod part is placed in as large coils outside the calorimeters, while the end-cap toroids are placed at each end of the central solenoid magnet and inserted into the barrel part at each end. Both end-cap and barrel part consist of 8 coils placed in symmetrically around the beam pipe, therefore generating a strongly non-uniform field

which has to be monitored and taken into account for the reconstruction of tracks.

The choice of a solenoid magnet in the central part of the detector ensures a small magnetic field in the calorimeter compared to the outer field, which is important for measuring the shower shapes in the calorimeter. The strong field in the outer part ensures a good momentum resolution in the muon spectrometer.

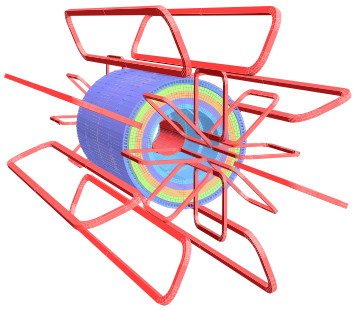


Figure 3.3: Overview of the magnet system [19].

3.4 The Inner Detector

The ID [20], shown in figure 3.4, is responsible for the tracking of charged particles and consists of three sub-detectors using different technologies: closest to the beams the silicon pixel detector, next the Semi-Conductor Tracker (SCT) and farthest away from the beam axis the Transition Radiation Tracker (TRT). The outer dimensions of the ID are 5.6m (length) and 2.1m (diameter).

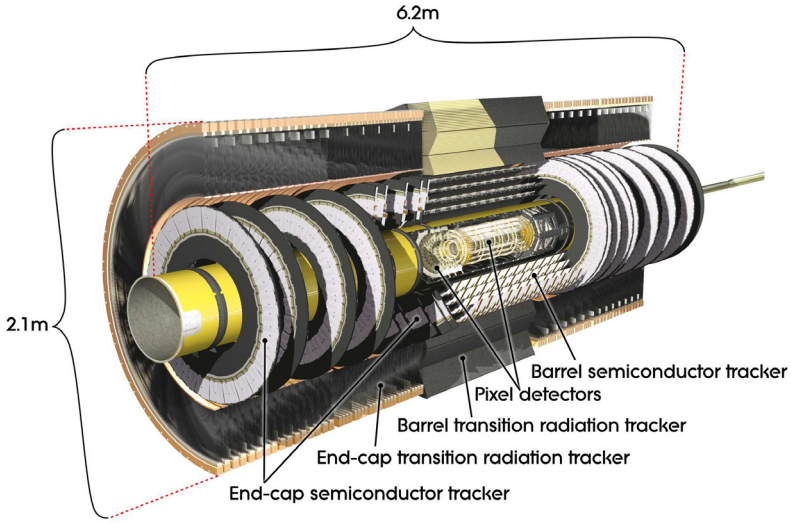


Figure 3.4: Overview of the parts and dimension of the ATLAS The Inner Detector [19].

The main purpose of the ID is high precision tracking of charged particles providing a good reconstruction of the 4-momentum and the vertex of the tracks. The material budget in of the ID had to be minimised to reduce multiple scattering, worsening the momentum resolution, and to reduce energy loss before particles reaching the calorimeters. All the three sub-detectors of the ID, described in the following sections, are divided into barrel and end-cap parts.

3.4.1 The Pixel Detector

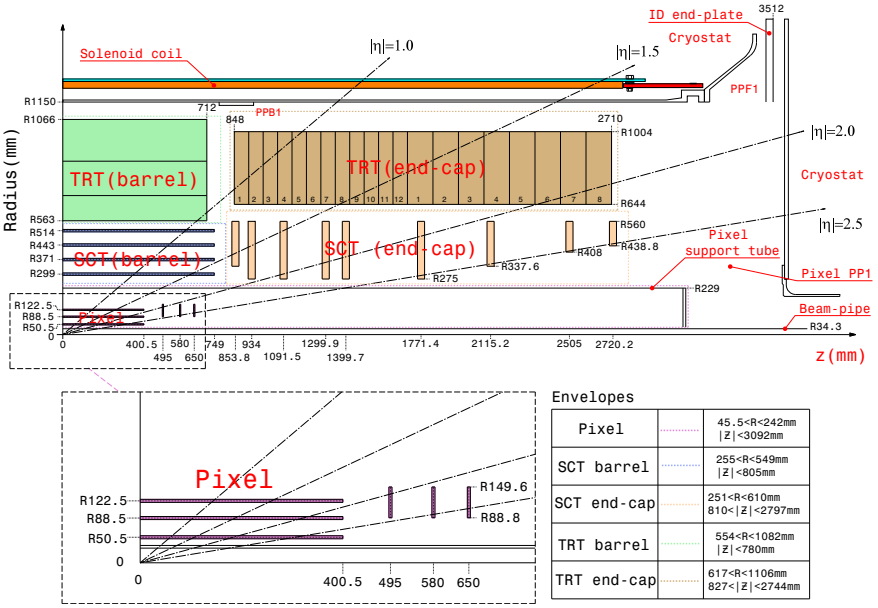


Figure 3.5: Close-up of the different types the detectors in the Inner Detector. In the region closest to the beam axis we see the Pixel detectors followed by the Semiconductor Strip Trackers and the Transition Radiation Trackers [19].

The pixel detector is the innermost detector in ATLAS providing the highest precision measurements for the tracking and vertex determination. It contains 80 million channels, placed in three concentric cylinders around the beam axis in the barrel part and in three disks perpendicular to the beam axis in each end-cap. The sensor size of 50×400 μm^2 generates a hit precision of 10 μm(r-ϕ) and 115 μm(z) in the barrel regions and 10 μm(r-ϕ) and 115 μm(r) in the end-caps. The pixel detector covers an η range up to 2.5 as shown in figure 3.5.

3.4.2 The Semiconductor Trackers

The SCT, also depicted in figure 3.5, consists of 6.3 million silicon strip channels grouped in 4088 modules with 1536 channels per module [21]. In each module, the strips are placed in two back-to-back layers which are slightly rotated by a 40 mrad stereo angle wrt. each other. In this way, information on a measured hit is obtained along the strip

direction and a full 3D space point measurement is obtained. The modules are placed in 4 barrel cylinders and 18 end-cap disks, designed so particles will pass 8 layers on their way from the interaction point throughout the detector. Due to this the end-cap and barrel modules have different geometries. For the barrel modules the strip pitch is 80 μm while the pitch varies from $56.9 - 90.4$ μm in the end-cap disks. The SCT covers a pseudo-rapidity range of $|\eta| < 2.5$ with a hit precision of 17 μm(r-ϕ) and 580 μm(z) in the barrel modules and 17 μm(r-ϕ) and 580 μm(r) in the end-caps.

3.4.3 Transition Radiation Tracker

The outermost part of the ID, the TRT has approximately 351,000 readout channels. Instead of few high precision measurements, the TRT, provides a series of tracking measurements by hits generated in 4 diameter straws tubes with a 30 μm tungsten wire in the centre and filled with a mixture of 70% Xe, 30% CO_2 and a small amount of O_2. A typical track has approximately 30 hits in the TRT, which allow for an almost continuous tracking of the particles in the region of $|\eta| < 2.0$. The TRT provides r-ϕ information with an intrinsic accuracy of 130 μm per straw. A third coordinate cannot be obtained since the space point coordinate along the wire direction can not be determined. In the barrel part, covering $|\eta| < 0.8$, the straws are placed parallel to the beam axis with the wires divided into two halves at $\eta = 0$, while in the end-caps, covering from $|\eta| = 0.8$ and up, the straws are arranged radially to maximise the numbers of straws passed by a particle coming from the interaction point.

The combination of the inner very precise trackers with the TRT at a larger radius gives very robust pattern recognition and high precision with a full set of track coordinates. Especially for high p_{T} tracks the outer measurements by the TRT contributes significantly to the track resolution.

The TRT also provides a good electron identification. Due to the transition radiation produced by highly relativistic particles ($\gamma \geq 1000$) passed the material between in the TRT straws, good separation between electrons and charged pions is achieved.

3.5 Calorimeter

Outside the ID, the calorimeter system is placed as shown in figure 3.6. It consists of a fine granularity liquid-argon electromagnetic (EM) calorimeter and a coarser hadron calorimeter. The EM has a pseudo-rapidity range of $|\eta| < 3.2$, hereby fully covering the of the ID and providing valuable information on photons and electrons, while the hadronic calorimeter extents its range up to $|\eta| = 4.9$. Both calorimeters have an almost complete coverage in ϕ which together with the large η-range is crucial for the ability of reconstruct missing transverse energy, $E_{\mathrm{T}}^{\mathrm{miss}}$.

Calorimeters have to provide good containment for electromagnetic and hadronic showers and minimise the amount of particles faking muons by reaching the muon system. Therefore, the depth of each of the calorimeters has been an important design criteria.

3.5.1 Electromagnetic calorimeter

The EM calorimeter consists of a barrel part, covering $|\eta| < 1.475$, and two end-cap parts, covering $1.375 < |\eta| < 3.2$. The barrel is divided into two half-barrels, which

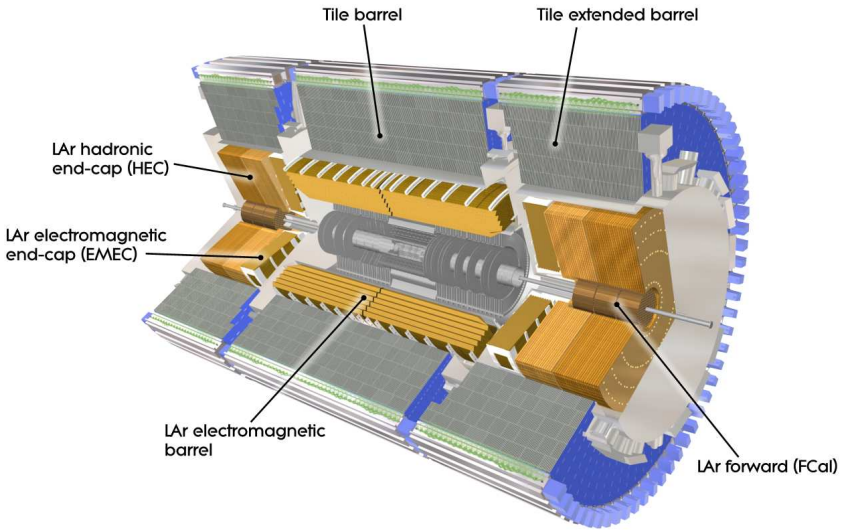


Figure 3.6: Overview of the different sub-detectors in the calorimeter system in ATLAS [19].

are separated by a small gap of 6 mm in the middle. Each end-cap is divided into an inner wheel, covering $2.5 < |\eta| < 3.2$, and an outer wheel, covering $1.375 < |\eta| < 2.5$. This splitting is done to optimise the granularity and dimensions in each region. In the EM calorimeter the energies are measured by having alternating layers of absorbing material of high density (lead), which will make the particles start showering, and layers of active materials (LAr), that will detect the showering particles. To get a homogeneous response in ϕ, each LAr cell is made in an accordion shape. The multilayer structure of the calorimeter allows for determination of the lateral and longitudinal shower shapes.

The layout of the magnet system and the material used in the calorimeter ensures that the magnetic field inside the EM calorimeter is small, which is important when determining the shape of showers made by incoming particles. A strong magnetic field inside the calorimeter would have caused an artificial broad shower shape. Unfortunately the solenoid adds extra material in front of the calorimeter, which can cause the particles to start showering before they reach the calorimeter and thereby make the energy measurements less precise.

The presampler: In order to correct for the energy loss in the ID and the solenoid, a single layer of LAr has been put in front of the EM barrel. This layer, called the presampler is not shown in figure 3.6. It only covers the region $|\eta| < 1.8$, since this is the region where particles pass most material in the ID due to the solenoid magnet.

The first sampling: The first layer in the longitudinal direction in the barrel and the end-caps has the finest granularity ($\Delta\eta \times \Delta\phi \approx 0.025 \times 0.1$) in order to best determine the impact point in the calorimeter and in order to separate neutral pions from single photons. The presence of additional neutral pions is important

for reconstruction of τ-lepton which decays into hadrons. The depth of the first sampling is 4.3 radiation lengths (X_0).

The second sampling: The second layer in the EM calorimeter is 16 X_0 deep, and is where most of the energy is deposited. In order to get a precise direction measurements in the EM calorimeter a fine granularity in both η and ϕ in the second layer is needed. Although, the granularity in the second sampling is a factor $\approx 5 - 10$ worse in η and a factor ≈ 2 in ϕ than in the first sampling.

The third sampling: The third layer in the EM calorimeter only exists in the barrel and the outer wheels of the end-caps. It mainly has to measure leakage from the second sampling generated by very energetic particles. The resolution in ϕ is the same as for the second sampling, but is reduced by a factor 2 in η. The depth of the third sampling is 2 X_0.

3.5.2 Hadronic calorimeter

Surrounding the EM calorimeter is the hadronic calorimeter, divided into a barrel part, two extended-barrels and two end-caps. In order to achieve the required stopping power of particles and radiation hardness, the hadronic calorimeter consists of different types. The barrel, covering the region $|\eta| < 0.8$, and the two extended-barrels, covering the range $0.8 < |\eta| < 1.7$, are made up of samplings of steel and scintillating tiles. Both barrel and extended-barrel are divided azimuthally into 64 modules and in depth into three layers with an approximate thickness of 1.5, 4.1 and 1.8 interaction lengths in the barrel and 1.5, 2.6, and 3.3 in the extended barrels. The outer radius of the tile calorimeter is 4.25 m.

The two end-cap calorimeters are LAr calorimeters located directly in extension of the EM end-caps. Each of the end-caps consist of two wheels covering a range of $1.5 < |\eta| < 3.2$ and hereby creating a small overlap with both the forward calorimeter, starting at $|\eta| = 3.1$, and the barrels of the tile calorimeter. The wheels are divided into 32 modules in ϕ and two segments in depth. The thickness of the steel layers are doubled in the outer wheels (50 mm) compared to the inner wheels (25 mm). Each layer of LAr has a depth of 8.5 mm in both wheels.

In addition to the Hadronic end-caps calorimeter a forward calorimeter covers the high density region $3.1 < |\eta| < 4.9$. This calorimeter is also LAr based, but uses tungsten in two of its three layers as in absorbing material to make it more dense and more radiation hard. The depth of the forward calorimeter is approximately 10 interaction lengths.

3.6 Muon Spectrometer

The muon spectrometer is the outermost part of the ATLAS detector, and also the largest sub-detector. Due to the strong magnetic field, it enables a high-precision tracking of muons. Since mostly only muons reach the muon spectrometer, it also provides a high trigger efficiency on high-p_{T} muons.

The muon system consists of four different sub-detectors. The Monitored Drift Tubes (MDT), covering the η-range of up to 2.7 (2.0 for the innermost layer), are responsible for the precision tracking. The MDT's are aluminium tubes of 3 cm in diameter with an

anode-wire in the middle. The tubes are filled with a gas consisting of 93% Ar and 7% CO_2. The average drift time in the MDT's is 300 ns and the expected resolution on a single wire is ~ 80 μm.

In η-range of $2.0 - 2.7$, the precision tracking is done by the Cathode Strip Chambers (CSC), which are multi-wire proportional chambers, and provides an even higher granularity than the MDT to sustain the high radiation level expected in this region.

In order to provide extremely fast feedback for the trigger system, special subdetectors are needed for the muon system . In the barrel part, this is provided by the Resistive Plate Chambers (RPC), covering $|\eta| < 1.05$, while the Thin Gap Chambers (TGC) are covering in the end-cap region, $|\eta| < 2.7$[3]. The RPC's and TGC's also measures the coordinate perpendicular to the wires in the MDT's and CSC's, hence providing an additional coordinate to the measured space points. The layout of the muon spectrometer with the different types of chambers can be seen in figure 3.7.

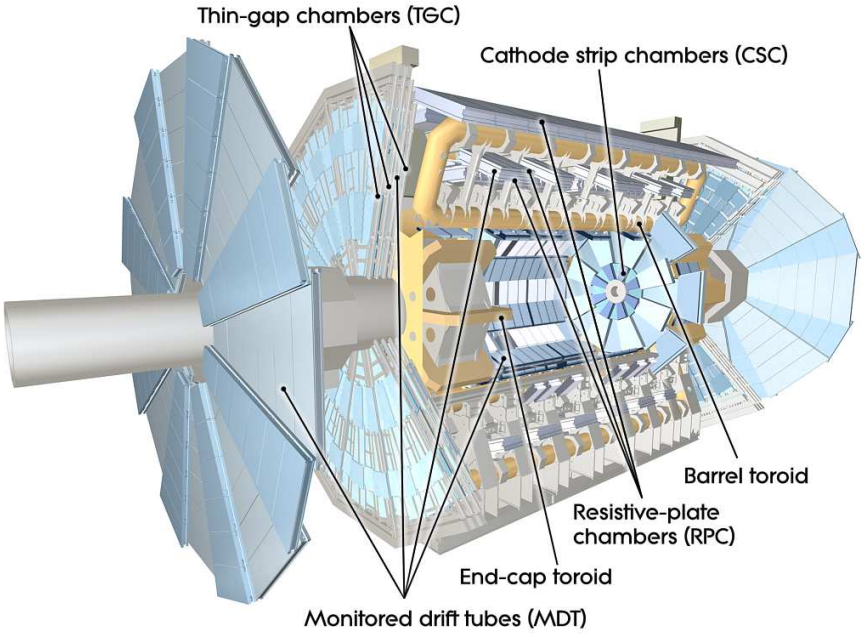


Figure 3.7: An overview of the muon spectrometer with the different types of muon chamber [19]. For precision measurements the Monitoring Drift Tubes and Cathode Strip Chambers are used, while the triggering uses the Resistive Plate Chambers and Thin Gap Chambers.

3.7 Trigger System

Even for the low luminosity runs in the first years of data-taking, the event rates are too high for all events to be recorded. Therefore, a coarse selection of which events to keep is needed. This triggering on events is designed to select potentially interesting events by

[3]The TGC covers an η-range up to 2.7, but only provides trigger information in the region up to $|\eta| = 2.4$.

using increasing levels of information details. The triggering is divided into categories dependent on its signature, e.g. high-p_{T}-lepton or jet, or large $E_{\mathrm{T}}^{\mathrm{miss}}$, and later combined into menus.

At design peak luminosity, $\sim 10^9$ collisions will happen in the ATLAS detector each second. However, the maximal event rate which can be stored due to limitations in bandwidth, cpu and storage capacity, is $\sim 10^2$ events per second. This means that a reduction of 10^7 is needed. Since the main part of the collisions will be soft interaction without interest for most studies at the LHC, most of the trigger menus are based on high-p_{T} signatures of leptons. While most of the bandwidth in the trigger is used for these *physics* streams, dedicated calibration and random triggers also exist for performance studies.

Since getting clean signatures in most cases requires some computations with increasing cpu requirement dependent on the purity of the signature, the ATLAS trigger system is divided into three trigger levels, with increased event information available for processing at each level and increasing demands on the purity which each level lets through. The trigger and data-acquisition system can be seen in figure 3.8. Level 1

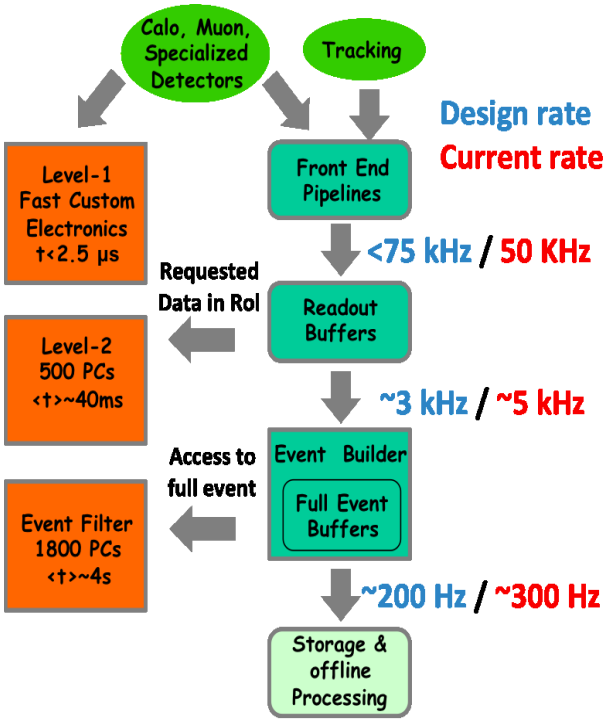


Figure 3.8: Overview of the trigger design in ATLAS. Both design and present rates at each level of the triggering system is indicated.[22]

(LVL1) of the trigger system is designed to reduce the event rate to < 75 kHz. This is part of the trigger system is hardware-based due to the high demand on the decision time ($\sim 2.5\mu$s). The LVL1 makes a decision on whether to skip the event or send it to the

next level for further processing. This decision is based on the muon information from the muon trigger chambers as described in section 3.6 and a reduced information from the calorimeter [4]. In the muon chambers the decision is based on candidates for high-p_{T} muons, while in the calorimeter the signatures are high-p_{T} objects, e.g. electrons, photon or jets, as well as large $E_{\mathrm{T}}^{\mathrm{miss}}$ or large total transverse energies.

During the LVL1 processing time, the full event information from all the sub-detectors are kept in the trigger readout pipelines. If an object passes the LVL1 trigger requirements, a region-of-interest (RoI), is defined as a suitable region in the detector around this object. When an event passes the LVL1 trigger, full detector informations from all RoI's are send to the second trigger level (LVL2).

At LVL2, information from the RoI's is processed using more sophisticated cuts, and the event rate is reduced to $\sim$ 3 kHz. Normally, only a small fraction of the detector is selected as RoI. For the muons tracking information from the precision chambers is combined with the tracking information from the ID, hereby improving the p_{T} measurement, and isolation information from the calorimeter is processed. For the electrons, the LVL2 matches calorimeter information to tracks in the ID as well as processing particle identification from the TRT.

Events that fulfil the requirement of the LVL2 are passed on the third level, the Event Filter (EF). At this level, the algorithms are very similar to the offline reconstruction algorithms. The designed output event rate of the EF is $\sim$ 200 Hz, which corresponds to $\sim$ 200 MB/s of stored data.

Figure 3.9 shows the output rate in ATLAS for different τ-trigger menus in 2012. Section 5.3.2 describes the trigger menus used in the $H \to \tau\tau$ analysis in more detail.

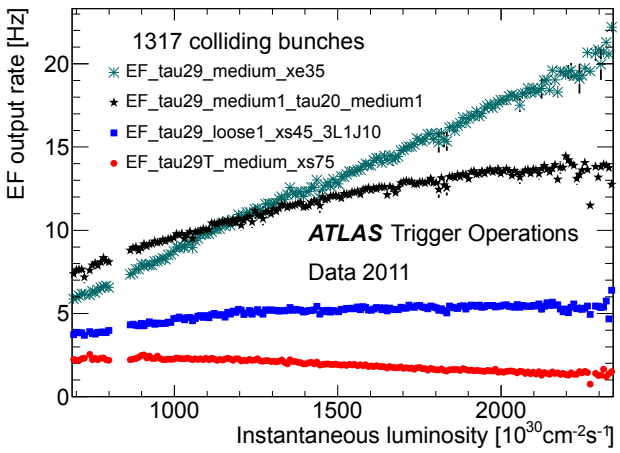


Figure 3.9: Output rate of different tau trigger menus during 2011 data taking [23].

[4] The energies in the calorimeter at LVL1 is calculated in towers by adding up the energy in all the samplings.

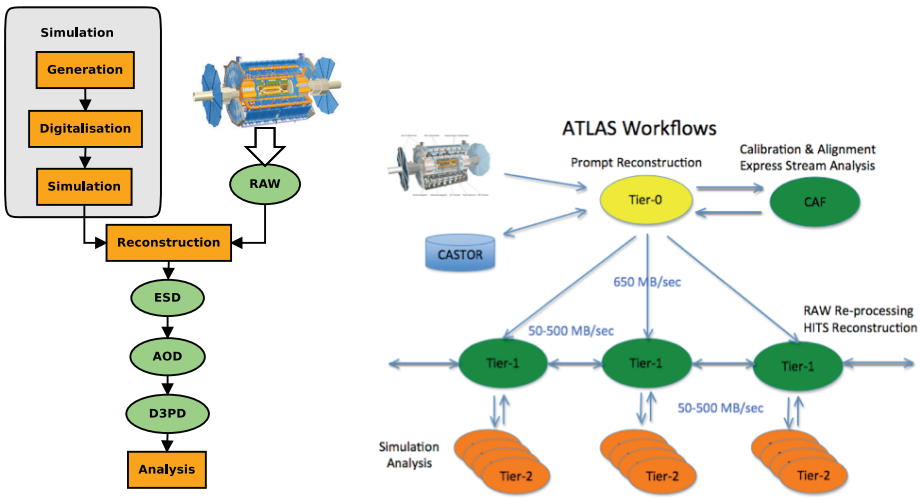
3.8 ATLAS Computing Model

The design of the ATLAS Computing Model is described in the Computing Technical Design Report, C-TDR [24, 25]. The computing challenges foreseen in the C-TDR consist of the following

- An enormous amount of data. The foreseen size of the data from the ATLAS experiment amounts to many petabytes of raw and processed data every year.
- A diversity of different data formats with big variation of files sizes.
- A world-wide user community.

In order to cope with these challenges, a very hierarchical computing model is defined in order to ensure a stable and robust running. This structure should guarantee both that official data cannot be compromised by user interaction and that sufficient resources remain available for the offline reconstruction and processing of data and no *bottlenecks* occurs while constant resources also are accessible for single user analysis. The solution for this has been *the Grid* [26] and the Tiers-of-ATLAS structure which allows to maintain the both strict and yet flexible structure needed to meet the continuously updated requirements from the experiment.

3.8.1 Data formats



(a) Data formats

(b) Tiers-of-Atlas structure [25]. The Tier-3's are left out.

Figure 3.10: ATLAS Computing Model.

The processing chain in ATLAS is depicted in figure 3.10(a). For full chain simulations, events from MC generators are converted to an Event-Summary-Data, ESD, through a GEANT4-simulation[27] of the particles interaction with the detector followed

by a digitalisation modelling of the detector response. This format is identical to the format produced by the offline reconstruction software for real data – except the truth record from the MC generator.

The ESD format contains almost all information from the reconstruction – including hits in the tracker and cell information in the calorimeters – and has a size of ~ 1 MB/event. From this format a chain of slimmed formats are derived keeping only a smaller and smaller fraction of information needed in the physics analyses. The final format used in most physics analysis is the D3PD format which is a ROOT ntuple.

For real data, events are split into blocks of a fixed integrated luminosity where each RAW file stored offline matches a single luminosity block. The block size is defined sufficiently small that detector and accelerator conditions can be assumed constant with one block. As data leaves the detector, it is split up into streams defined by their triggers, e.g. e-gamma stream, muon stream or ETmiss-Jet stream, which are kept separate throughout the full processing. The event overlap between different streams is $\sim 10\%$. A few special streams are also defined for detector calibration and monitoring purposes, but also an express stream which contains mixture of all physics streams.

3.8.2 The Tiers-of-ATLAS

The distributed computing model of ATLAS is divided into four level[5], referred to as Tiers, as illustrated in figure 3.10(b).

Tier-0 and AF The first level in the data processing model is the Tier-0 hosted at CERN. This centre is responsible for the first pass processing of the express streams and calibration streams from the RAW format into ESD's and AOD's. The Tier-0 is also responsible for exporting the RAW and derived data to the Tier-1 for improved reconstruction. Therefore large disk capacity is required at Tier-0. An integrated part of the Tier-0, is the Calibration and Alignment Facility, CAF, which takes care of the calibration and alignment tasks performed on the express and calibration streams.
The Tier-0 also processes a fraction of the data from the physics streams in the bulk processing of data, while the rest of the data is shipped to the Tier-1's.

Tier-1 There exist 10 Tier-1 centres around the world. Each receive a portion of the RAW data and is responsible for the reconstruction of this portion using the updated calibration and alignment constants. The Tier-1 is also responsible for the long-term storage of the derived data formats from its portion of data. In addition to the processing of real data, the Tier-1's are also performing the main part of the processing of simulated data.
Furthermore, the Tier-1's are providing a connection to the Tier-2's. This means that all data transfers between Tier-2 centres goes through the Tier-1's.

Tier-2 and Tier-3 The Tier-2 and Tier-3 are where the user analysis tasks are performed. All formats until D3PD production are derived by the ATLAS central production system. The Tier-2's and Tier-3's provide dedicated resources for user analysis.

[5]Originally, the Tier-structure of ATLAS was defined as a three-level structure. However as the need for distributed storage of output from user analysis increased, a fourth level (Tier-3's) where added

Since long-term storage for user generated data was not foreseen in the original computing model, a distinction is made between Tier-2's and Tier-3's depending on whether long-term storage for user data is available or not.

3.9 Calibration and Alignment

The ATLAS calibration and alignment tasks cover a wide variety of processes generating non-event data to be used for the reconstruction or monitoring of the detector [28]. The input data can come from one or more sub-detectors and the format ranges from dedicated readout on the individual sub-detectors to fully reconstructed event data. The output from all calibration and monitoring processing is stored in a conditions COOL database which can be accessed both by the online system as well in the reconstruction or analysis steps later on.
The calibrations tasks can be categorised in the following ways:

Online calibration: The purpose of the online calibration is to provide immediate feedback. Therefore these tasks are performed on resources at Point 1 very close to or on the detector. The online category is further divided depending on whether the processes are performed on subsystem readout (ROD) or in the Event Filter(EF) system. Both types of calibration need special triggers or dedicated calibration runs as well as dedicated resources in terms of CPU and disk spaces. However, whereas the ROD calibration tasks are limited by only having access to a partial event information, the EF system has access to the full event. The online calibration is very limited by both the time and resources available.

Offline calibration: The offline calibration runs on reconstructed events and its purpose is to further improve the reconstruction as much as possible. These tasks run either on Tier-0 at CERN or at the Tier-1's outside CERN.

Prompt calibration: The prompt calibration sits in between the online and offline calibration. The purpose of these tasks are to process the express stream of the RAW data as fast as possible to provide the best possible detector status to the reconstruction before the reconstruction of the full physics data starts. These tasks are run on a dedicated part of the Tier-0 in order to optimise both the performance power and the feedback time.

3.10 Prompt Calibration Model

In figure 3.11, an overview of the Data Quality workflow is depicted. Once data is recorded the process of all physics streams, the *bulk* reconstruction, is delayed for 36 hours[6]. During this period the express stream and calibration streams are processed by the Prompt Calibration, and before the period has ended the new calibration and alignment data is uploaded to the central database. Once the bulk processing start, the updated detector conditions can be used.

[6]The initial delay period for the bulk reconstruction was 24 hours, but since some calibration task took longer than this, the period was extended to 36 hours

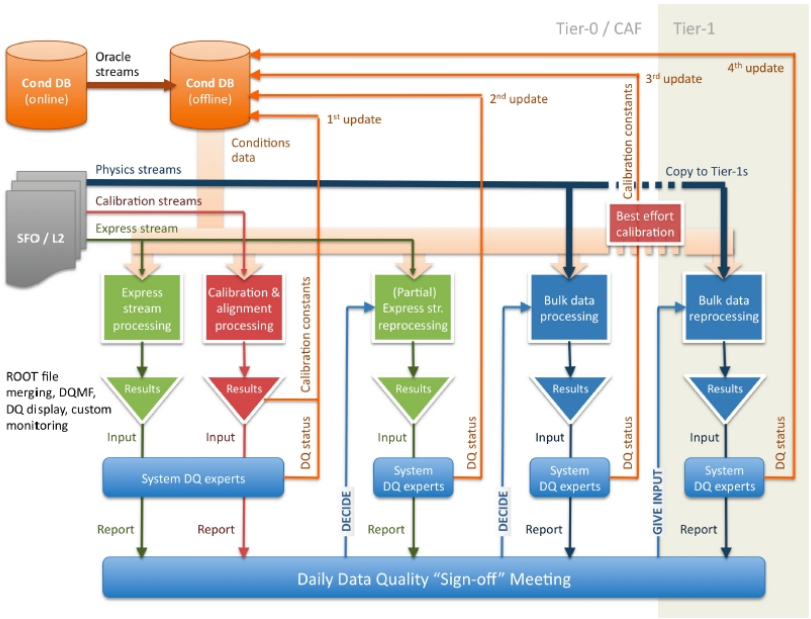


Figure 3.11: Overview of the ATLAS Prompt Data Model [24]

3.10.1 Task Management System

The Task Management System(TMS) is responsible for creating, monitoring, and controlling tasks running on Tier-0 as well as the staging of data to and from these tasks. The system was originally developed to handle bulk reconstruction tasks running at Tier-0, but as the need for more robust and automatic handling of the calibration tasks increased, its use was expanded to the prompt calibration tasks.

The SCT was the first sub-detector to deploy this system, but the use of TMS has later been deployed to several other sub-detectors including the Pixel and TRT.

In TMS, a task is defined as an entry in the Task Database(DB). Such an entry contains information on the task description, input and output datasets, associated jobs at Tier-0 and a global status. A single task will most commonly contain several jobs. An example from the reconstruction is the `ESD2AOD` task which converts ESD datasets into an AOD dataset. For each new ESD dataset a task will be created, however each ESD dataset contains many files and therefore each task will most likely contain several jobs, e.g. one job for each file in the dataset.

The typical workflow of the TMS system, sketched in figure 3.12, consists of following steps:

- The TOM daemon scans the central database for new datasets. In case a new dataset matches the task criterions defined in any of the tasks, a new task will be inserted in the Task DB. An example criterion for task creation could be the number of events in the dataset. Until the jobs associated to this task are defined, the task will have status `DEFINED`.

- The central Eowyn daemon scans the TMS DB for tasks. If a task with status

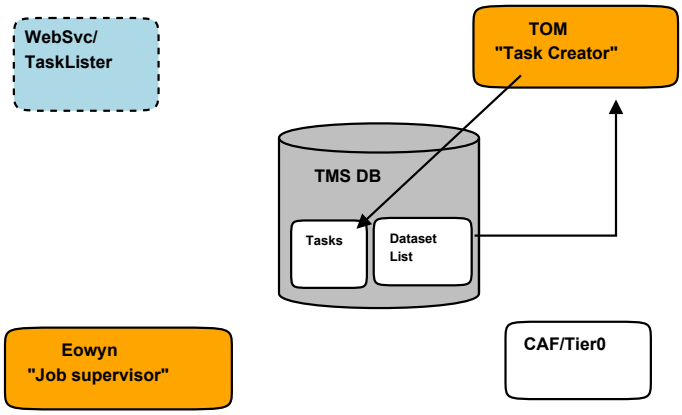


Figure 3.12: Overview of the Task Management System workflow. The TOM deamon handles all task management, while the central Eowyn deamon is responsible for the handling of individual jobs inside each task.

`DEFINED` is found, jobs are created according to the task description and submitted to Tier-0. The task status is then changed to `RUNNING`

- The Eowyn regularly scans the jobs status and update the task status accordingly. If jobs have failed they are automatically resubmitted. The Eowyn also handles the staging of input files and output files produced by each job.

- Once all jobs belonging to a task have finished, the TOM deamon changes the status of the task to `FINISHED`.

The TMS also provides a web service that allows user interaction with the Task DB. This enables users to manually abort or create tasks which can be useful both for testing purposes and for cases of special runs.

The webpage called the TaskLister, shown in figure 3.13, enables the user to easily monitor both current and past tasks. For each task the task name and type is listed together with the number of failed, running and successful jobs. In case a task has failing jobs, a link is provided to a copy of the log files which make quick debugging possible for the user.

Run No.	Task Name	Task Information			Job Statistics					
		ID	Type	Status	#Done	#R	#Pr	#T	#Al	#T.
149175	data10_cos.00149175.physics_RNDM.merge.NTUP_TRKVALID.f214_m370_c0.sctcal...	349586	sctcalibhv	FINISHED	1	0	0	0	0	0
149175	data10_cos.00149175.physics_RNDM.merge.NTUP_TRKVALID.f214_m370_c0.sctcal...	349585	sctcalibns	FINISHED	1	0	0	0	0	0
149172	data10_cos.00149172.physics_RNDM.merge.NTUP_TRKVALID.f214_m370_c0.sctcal...	349578	sctcalibhv	FINISHED	1	0	0	0	0	0
149172	data10_cos.00149172.physics_RNDM.merge.NTUP_TRKVALID.f214_m370_c0.sctcal...	349577	sctcalibns	FINISHED	1	0	0	0	0	0
149166	data10_1beam.00149166.physics_RNDM.merge.NTUP_TRKVALID.f214_m370_c0.sctc...	349568	sctcalibhv	FINISHED	1	0	0	0	0	0
149166	data10_1beam.00149166.physics_RNDM.merge.NTUP_TRKVALID.f214_m370_c0.sctc...	349567	sctcalibns	FINISHED	1 (1)	0	0	0	0	0
149158	data10_1beam.00149158.physics_RNDM.merge.NTUP_TRKVALID.f214_m370_c0.sctc...	349563	sctcalibhv	FINISHED	1	0	0	0	0	0
149158	data10_1beam.00149158.physics_RNDM.merge.NTUP_TRKVALID.f214_m370_c0.sctc...	349562	sctcalibns	FINISHED	1 (1)	0	0	0	0	0
149146	data10_1beam.00149146.physics_RNDM.merge.NTUP_TRKVALID.f214_m370_c0.sctc...	349554	sctcalibhv	FINISHED	1	0	0	0	0	0
149146	data10_1beam.00149146.physics_RNDM.merge.NTUP_TRKVALID.f214_m370_c0.sctc...	349553	sctcalibns	FINISHED	1 (1)	0	0	0	0	0
149135	data10_1beam.00149135.physics_RNDM.merge.NTUP_TRKVALID.f214_m370_c0.sctc...	349549	sctcalibhv	FINISHED	1	0	0	0	0	0
149135	data10_1beam.00149135.physics_RNDM.merge.NTUP_TRKVALID.f214_m370_c0.sctc...	349548	sctcalibns	FINISHED	1 (1)	0	0	0	0	0
149133	data10_1beam.00149133.physics_RNDM.merge.NTUP_TRKVALID.f214_m370_c0.sctc...	349544	sctcalibhv	FINISHED	1	0	0	0	0	0
149133	data10_1beam.00149133.physics_RNDM.merge.NTUP_TRKVALID.f214_m370_c0.sctc...	349543	sctcalibns	FINISHED	1 (1)	0	0	0	0	0
149130	data10_1beam.00149130.physics_RNDM.merge.NTUP_TRKVALID.f214_m370_c0.sctc...	349539	sctcalibhv	FINISHED	1	0	0	0	0	0
149130	data10_1beam.00149130.physics_RNDM.merge.NTUP_TRKVALID.f214_m370_c0.sctc...	349538	sctcalibns	FINISHED	1	0	0	0	0	0
149129	data10_1beam.00149129.physics_RNDM.merge.NTUP_TRKVALID.f214_m370_c0.sctc...	349534	sctcalibhv	FINISHED	1	0	0	0	0	0
149129	data10_1beam.00149129.physics_RNDM.merge.NTUP_TRKVALID.f214_m370_c0.sctc...	349533	sctcalibns	FINISHED	1	0	0	0	0	0
149061	data10_1beam.00149061.physics_RNDM.merge.NTUP_TRKVALID.f214_m370_c0.sctc...	349526	sctcalibhv	FINISHED	1	0	0	0	0	0
149061	data10_1beam.00149061.physics_RNDM.merge.NTUP_TRKVALID.f214_m370_c0.sctc...	349525	sctcalibns	FINISHED	1 (1)	0	0	0	0	0
148991	data10_1beam.00148991.physics_RNDM.merge.NTUP_TRKVALID.f214_m370_c0.sctc...	349518	sctcalibhv	FINISHED	1	0	0	0	0	0
148991	data10_1beam.00148991.physics_RNDM.merge.NTUP_TRKVALID.f214_m370_c0.sctc...	349517	sctcalibns	FINISHED	1	0	0	0	0	0
148909	data10_1beam.00148909.physics_RNDM.merge.NTUP_TRKVALID.f214_m370_c0.sctc...	349515	sctcalibhv	FINISHED	1	0	0	0	0	0
148909	data10_1beam.00148909.physics_RNDM.merge.NTUP_TRKVALID.f214_m370_c0.sctc...	349512	sctcalibns	FINISHED	1 (1)	0	0	0	0	0
148900	data10_1beam.00148900.physics_RNDM.merge.NTUP_TRKVALID.f214_m370_c0.sctc...	349513	sctcalibhv	FINISHED	1	0	0	0	0	0
148900	data10_1beam.00148900.physics_RNDM.merge.NTUP_TRKVALID.f214_m370_c0.sctc...	349510	sctcalibns	FINISHED	1	0	0	0	0	0
148884	data10_1beam.00148884.physics_RNDM.merge.NTUP_TRKVALID.f214_m370_c0.sctc...	349514	sctcalibhv	FINISHED	1	0	0	0	0	0
148884	data10_1beam.00148884.physics_RNDM.merge.NTUP_TRKVALID.f214_m370_c0.sctc...	349511	sctcalibns	FINISHED	1	0	0	0	0	0

Figure 3.13: Screenshot of the TaskLister webpage for monitoring of tasks. For each task the numbers of successful, running and failed jobs are shown. For failed jobs a link is provided to the log file of the job.

3.10.2 SCT Prompt Calibration

For the SCT Calibration, a special instance of the TOM deamon is running, controlling SCT calibration task, while the centrally managed Tier-0 Eowyn deamon controls all jobs running at the Tier-0 including the SCT calibration jobs. Upon initialisation, the TOM deamon reads in a configuration file with all the tasks description [7]. Having a separate instance of the TOM deamon, enables easy change task parameters (both criteria and description) without affecting the rest of the Tier-0 workflow. The task criteria includes a search string for which datasets to process and a requirement on the minimal number of events in the dataset, while the task parameters contain more SCT specific settings.

Since TMS was initially developed to handle the bulk reconstruction workflow, several features needed to be implemented in order to facilitate the calibration tasks. In the reconstruction workflow the processing of datasets is split up into several identical sub-jobs running in parallel to allow faster processing. After all sub-jobs are done the output from all sub-jobs are merged into a new dataset which can be passed on to the next step in the reconstruction chain. For the SCT calibration, each dataset needs to be processed as a whole. This quickly generated problems as the size of datasets rose and the data became to large to be shipped to a single node on the Tier-0. In order to cope with these large datasets, a mechanism was developed to continuously stream data from the storage to the Tier-0 nodes. In addition to the space issues, the SCT calibration also needed to run several different tasks for each run – one for each calibration routine since these produce different output and in some cases take different input types.

[7] An example of a SCT TOM config file is given in appendix A

Part III

Analysing Pairs of τ-leptons

CHAPTER 4

Reconstructing the $\tau^+\tau^-$ rest frame

4.1 Reconstructing the $\tau^+\tau^-$ rest frame

At LHC energies, due to the short life time of the τ-lepton, it is only possible to study τ-leptons indirectly through their decay products. This introduces great challenges because of the neutrinos always present in the decays, which leads to an inconsistency between the visible 4-momentum of the decay products, the τ-jet, and the 4-momentum of the true τ-lepton.

This deviation in the 4-momentum makes it impossible to exactly calculate the rest frame, RF, of a resonance, decaying into a pair of τ-leptons as well as accurately reconstruct the mass of the decaying resonance. Although, an approximation of the RF, is given by the centre-of-mass frame of the two visible τ-jets, this frame will, by construction, always reconstruct the energy of the two τ-jets to be equal. Hence this will give a poor frame to study helicity correlation of the two τ-leptons.

Furthermore, the invariant mass of the two τ-jets, M_{vis}, will greatly underestimate the invariant mass of the resonance as shown in figure 4.1 for Z^0-bosons.

In the RF of a heavy resonance, i.e. Z^0-bosons, decaying into a pair of τ-leptons, the direction of the two τ-jets will only deviate slightly from that of the τ-leptons. Hence the two τ-jets will appear nearly back-to-back in this frame. This can be exploited to find an approximation to the RF. For any given frame of reference the *acollinearity* – defined as angular difference for the two τ-jets to be completely back-to-back, $\alpha = \pi - \angle(\tau_1, \tau_2)$ – is a well-defined number between zero and π. Hereafter, α denotes the acollinearity between the two τ-jets and α_{RF} denotes the acollinearity in the resonance rest frame.

At LHC energies, Z^0-bosons (or possible light Higgs bosons) are produced with a large boost with respect to the detector frame predominately in the direction along the beam axis[1] as shown in figure 4.2. Therefore, the measured 4-momenta of the τ-jets in the detector frame will deviate significantly from those in the RF.

As shown in figure 4.3, α_{RF} is expected to be small, with an average value of 0.28 rad. for Z^0-bosons and 0.50 rad. for 125 GeV ggH Higgs bosons. This fact has been used

[1] In the case of a Higgs boson, both the size and direction of the boost will depend on the production mechanism.

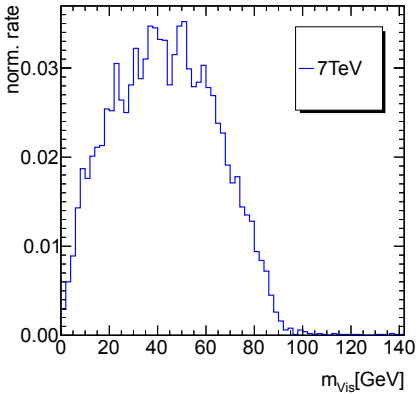
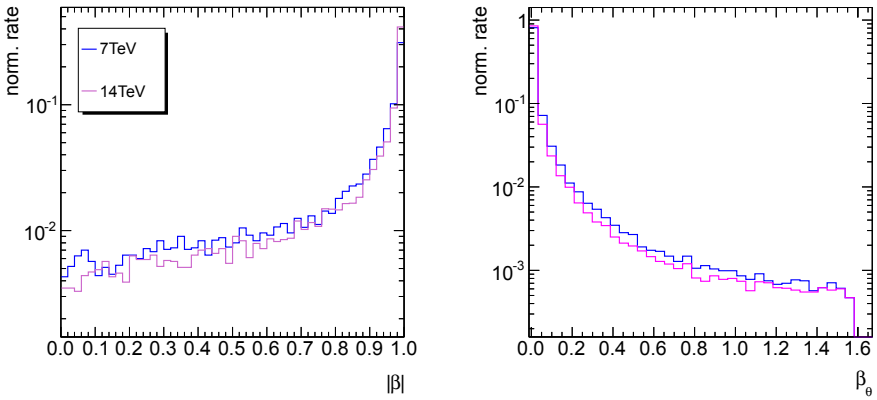


Figure 4.1: Visible mass distribution for $Z \to \tau\tau$+jets in PYTHIA8+TAUOLA simulations pp-collisions at $\sqrt{s} = 7$ TeV.

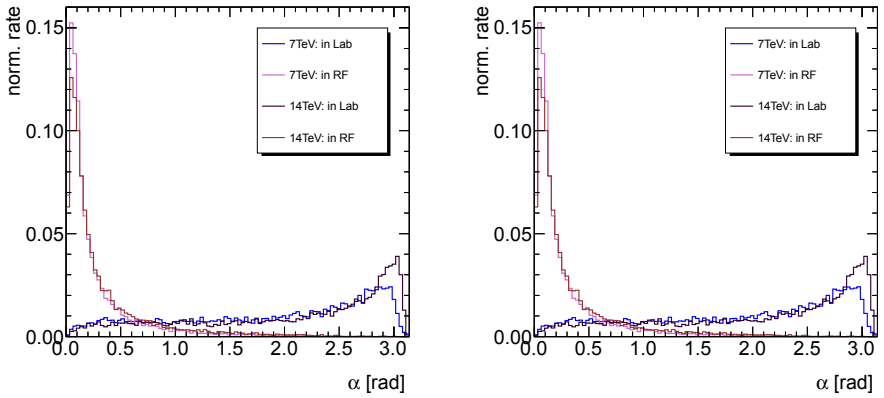


(a) Boost magnitudes, $|\beta|$.

(b) Polar angle of the boost,β_θ

Figure 4.2: Boost magnitudes and angles for Z^0 and Higgs bosons at different energies simulation using PYTHIA8. At both energies both Z^0 and Higgs bosons are created with a high boost predominantly along the beam axis.

to develop a technique for reconstructing the RF [29]. The technique is divided into two categories, depending on whether the transverse part of the boost should be estimated.



(a) $Z \to \tau\tau$ (b) ggH $\to \tau\tau$

Figure 4.3: Acollinearity in the laboratory frame and the true rest frame for Z^0-bosons and Higgs at different energies.

4.1.1 The `BoostZ`-method

In the simplest method, the p_{T} of the resonance is assumed to be negligible and therefore only the longitudinal boost, β_{z} is reconstructed. The method, hereafter referred to as `BoostZ`-method [29], consists of searching for the β_{z} that minimises α. Since α is a well-defined positive number for any applied boost and α is a one-dimensional function of β_{z} with a single minimum, this minimisation is highly reliable and the minimum can be found using a simple and fast binary search[2].

In figure 4.4, α is shown as a function of applied β_{z} for different $Z \to \tau\tau$ events in 7 TeV pp-collisions simulated using PYTHIA8 [30]. In all the events the minimum of the curves is close to the true β_{z} of the Z^0-boson.

4.1.2 The `BoostXYZ`-method

The `BoostZ`-method works well for events with relative small β_{T} which covers most part of both Z^0 and Higgs boson events produced at the LHC. On the other hand since events where the resonance is produced with a large β_{T} are easier to trigger on at the LHC, it can be vital to extend this method to also be able to cope with such events.

In general, there always exists an infinitely set of frames of reference with $\alpha = 0$. Therefore it is not possible to do a sensible global minimisation of α. Instead one has to restrict the allowed search directions – preferably to a subset of possible boosts which only contains a single minimum. This is exactly the strategy in searching for the boost along the beam axis in the `BoostZ`-method.

[2]The minimisation procedure has also been cross checked using more sophisticated minimisation routines

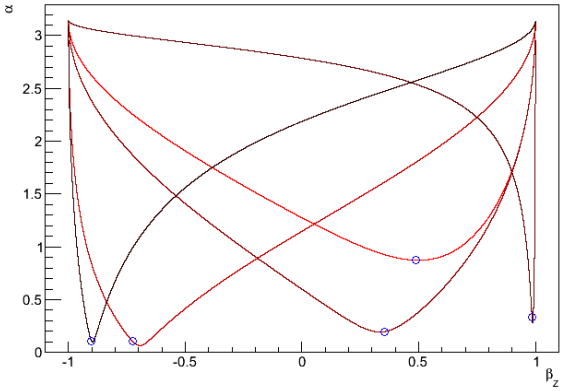


Figure 4.4: Acollinearity as a function of applied longitudinal boost for various $Z \to \tau\tau$ events for PYTHIA8 simulations of 7 TeV pp-collisions. The blue circles mark the true longitudinal β_z and true α_{RF} of the Z^0-bosons. The functions all have a single minimum situated close to the true β_z.

A straightforward extension of the `BoostZ`-method is also proposed in [29]. This method, referred to as the `BoostXYZ`-method, consists of first finding the β_z using the `BoostZ`-method and hereafter finding the transverse component by minimising α by searching for a transverse boost component while keeping the longitudinal component fixed. The transverse search direction is derived by summing up the missing transverse momentum and the two visible τ-momenta. This will of course be very dependent on the reliability of the E_T^{miss} reconstruction which can be very difficult at hadron colliders due to pileup.

4.1.3 Performance using ATLAS simulations

In [29], the performance of both the `BoostZ`-method and the `BoostXYZ`-method is documented using generator level quantities. However, in order to fully test the performance of this technique, it is vital to do more complete studies using realistic detector reconstructed variables.

Since the `BoostZ`-method only relies on measured particles directions, its performance is not expected to be significantly degraded using reconstructed quantities. For the `BoostXYZ`-method the performance will depend heavily on the ability to estimate the transverse direction.

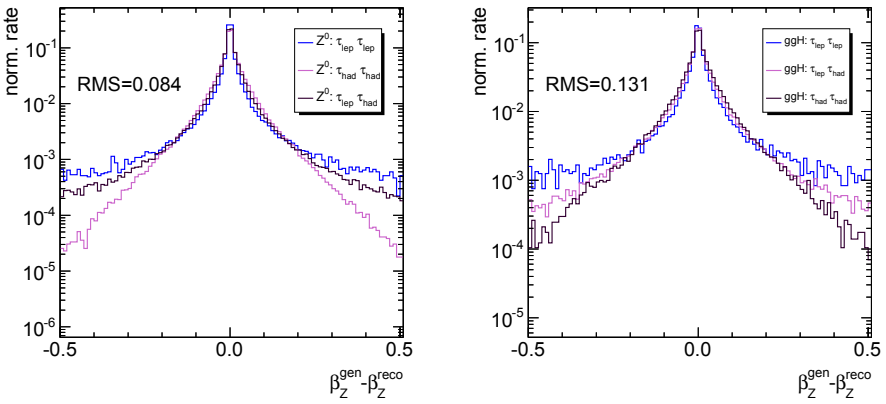
In the following the performance of the `BoostZ`-method and `BoostXYZ`-method is tested for reconstructed variables using full ATLAS detector Monte Carlo simulations listed in table 4.1. To get a realistic performance a loose preselection is applied on the events. The selection requires exactly two reconstructed τ-leptons with a $p_T > 15$ GeV and a $|\eta| < 2.5$ to be within the detector resolution. Only τ-leptons with 1 or 3 associated tracks are considered. Furthermore to be able to compare the reconstructed RF with the true boson RF both reconstructed τ-leptons are required to be matched to a true lepton (e, μ or τ) in the generator level event record.

Figure 4.5 shows the difference in reconstructed and generated β_z for the different

Process	Generator	Events
ggH $\rightarrow \tau\tau$	POWHEG+PYTHIA	250k
$Z \rightarrow \tau\tau$	PYTHIA8	3M

Table 4.1: Dataset used to evaluate the rest frame reconstruction performance. For Higgs samples a Higgs mass of 120 GeV has been used.

samples divided into the decay mode of the $\tau^+\tau^-$-system. In both Higgs and Z^0 events, the performance of reconstructing the β_z is good with an overall RMS[3] of 0.08 for Z^0 events and a RMS of 0.13 for Higgs events. Despite of presence of more neutrinos in the leptonic τ-decays than in the hadronic decays, the distributions are slightly narrower close to zero but have much larger tails away from zero. The large spread around zero for hadronic decays, could be due to the large fraction of decays containing neutral hadrons.



(a) $Z \rightarrow \tau\tau$ (b) ggH $\rightarrow \tau\tau$

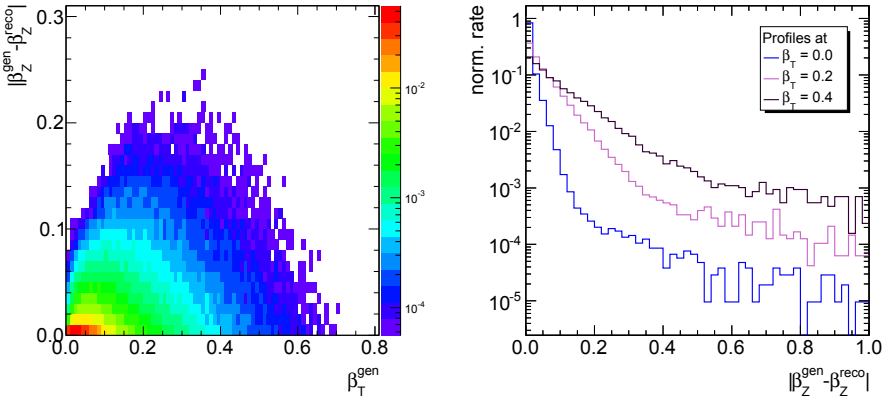
Figure 4.5: Reconstruction performance of the longitudinal boost using the `BoostZ`-method. The longitudinal boost reconstruction is identical to that in the `BoostXYZ`-method.

The worsening in performance for Higgs events is mainly due to the larger transverse boost in these events compared to the Z^0 events – as illustrated in figure 4.6(a) where the performance is shown as function of β_T for $Z \rightarrow \tau\tau$ events. This fact clearly signal that an extension of the `BoostZ`-method is needed in case of large β_T.

To estimate the goodness of the full reconstruction, the vectorial difference of true and reconstruction boost vectors, $|\vec{\beta}_{gen} - \vec{\beta}_{reco}|$ is computed for both reconstruction methods. In [29] performance numbers based on generator level information have been derived for events with either leptonic or 1-prong τ-decays. In table 4.2 the numbers from [29] are compared to the numbers derived here. Here the numbers are derived including all hadronic τ-decays, but it is checked that the performance on 1-prong is very similar to that of multi-prongs.

As expected there is no significant difference between the generator level studies and full detector simulation Monte Carlo in the performance of `BoostZ`-method. However,

[3]Here, the overall RMS is defined as the spread of the $\beta_z^{gen} - \beta_z^{recon}$ distribution



(a) β_z performance as function of β_T.

(b) Profiles of β_z performance for different values of β_T.

Figure 4.6: Dependence on transverse boost in the `BoostZ` reconstruction performance in Z^0 events. Neglecting the transverse component clearly has an effect for high transverse boost.

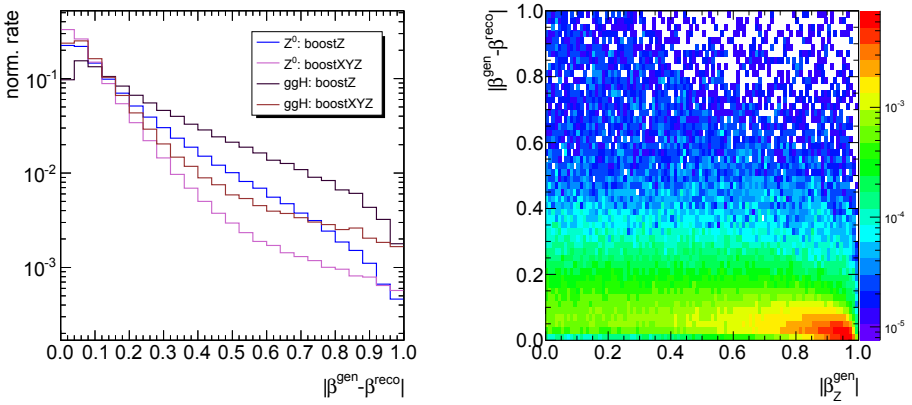
	Average $\lvert\vec{\beta}_{gen} - \vec{\beta}_{reco}\rvert$					
	`BoostZ`			`BoostXYZ`		
	Gen. Level	Det. Level		Gen. Level	Det. Level	
Decay mode	Z^0	Z^0	ggH	Z^0	Z^0	ggH
$\tau_\ell\tau_\ell$	–	0.160	0.259	–	0.125	0.195
$\tau_\ell\tau_{had}$	–	0.157	0.237	–	0.111	0.137
$\tau_{had}\tau_{had}$	–	0.142	0.221	–	0.089	0.110
Overall	0.149	0.148	0.234	0.076	0.098	0.132

Table 4.2: Comparison of rest frame reconstruction methods for different decay modes in Z^0 and Higgs events. For `BoostZ`, the performance is equally good at generator level and fully detector simulated objects while the performance is reduced for the `BoostXYZ`-method in full detector simulation.

for the `BoostXYZ`-method a reduction in the performance is seen in the detector simulation. For both methods, the best result is found in the fully hadronic final states as expected since only a single neutrino is generated in each decay. Although the bosons are predominantly longitudinally boosted, clear improvements are found when the search for the transverse boost is included in both Z^0 and Higgs events as also displayed in figure 4.7(a).

Since quantities such as particle momenta and energy transform highly non-linear as function of β, it is important to check how well the reconstruction works not just on an overall basis, but also in differential distributions. In figure 4.7(b), the performance of the `BoostXYZ`-method is shown as function of β_z and although the most part of events is concentrated near high β_z, the spread is more or less constant over the whole range – except for a small subset near 1 where effects from the requirement $|\beta| < 1$ are resulting in a small spread.

Since the `BoostZ`-method does not try to estimate the transverse component, $|\vec{\beta}_{gen} -$



(a) Vectorial difference between reconstructed and generated boosts.

(b) `BoostXYZ` reconstruction performance as function of longitudinal boost for Higgs events.

Figure 4.7: Full rest frame reconstruction performance. For both Higgs and Z^0-events the performance of `BoostXYZ` is the best and the performance of `BoostXYZ` is shown to be constant as function of longitudinal boost for Higgs events.

$\vec{\beta}_{\text{reco}}|$ will always be larger than the true β_{T} as illustrated in figure 4.8(a). Because the `BoostXYZ`-method estimates the β_{T}, one would expect to see this effect disappear, but – as shown in figure 4.8(b) – the effect of a transverse boost is still seen.

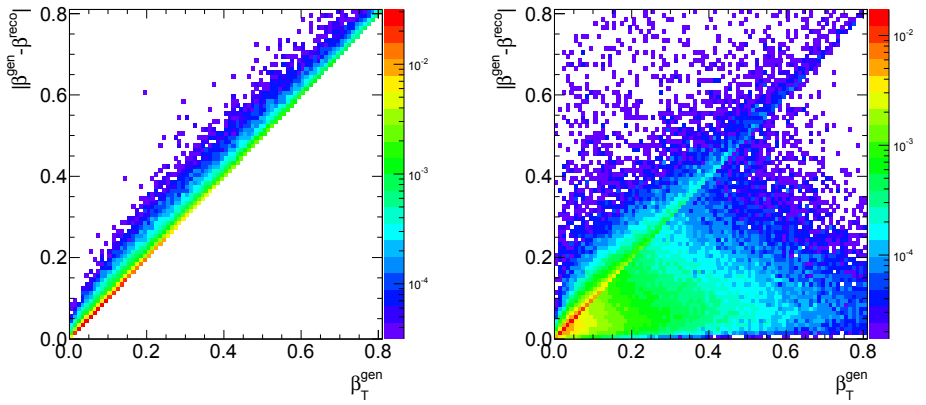
As mentioned above the `BoostXYZ`-method is highly sensitive to the reliability of the reconstruction of the transverse direction. A quantity that can be used to estimate the reliability of the reconstruction of the transverse direction is the magnitude of the reconstructed $E_{\text{T}}^{\text{miss}}$. As shown in figure 4.9 the reliability of the transverse direction decreases with decreasing $E_{\text{T}}^{\text{miss}}$. Therefore a good requirement in an analysis situation would be to introduce a lower cut on the $E_{\text{T}}^{\text{miss}}$ before extending the RF-reconstruction from `BoostZ` to `BoostXYZ`.

The method presented here searches for the longitudinal component of the boost first before estimating the transverse component. Alternating this sequence gives a slightly different performance, however given that the longitudinal component of β is the dominant part in mosts cases, estimating this component first gives the best performance.

4.2 Mass Estimation

One of the applications for reconstructing the RF of resonances is a new technique for estimating the resonance mass. In $\tau^+\tau^-$ pair systems, a trivial reconstruction of the invariant mass of the decaying mother boson is in principle not possible due the presence of neutrinos in the τ-decays.

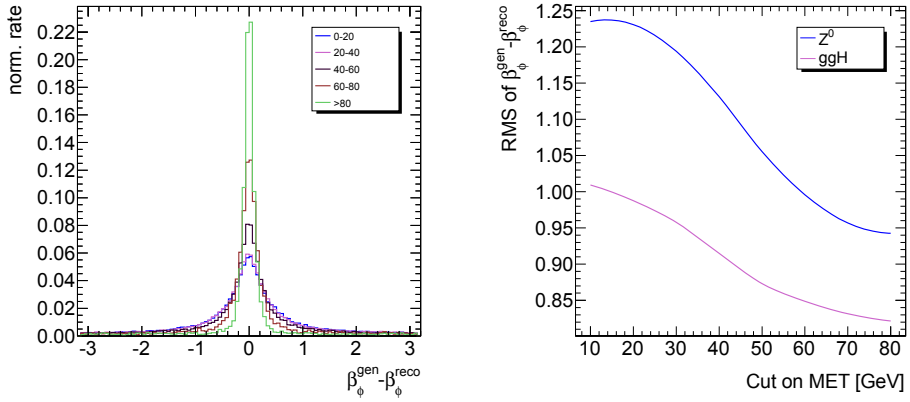
The visible invariant mass, M_{vis}, distribution peaks far below the real mass as demonstrated in figure 4.1. Although M_{vis} can often be used as a powerful discriminator between bosons with different mass, it is insufficient when trying to estimate the mass of a new resonance.



(a) `BoostZ`-method

(b) `BoostXYZ`-method

Figure 4.8: Full rest frame reconstruction as function of transverse boost for $H \to \tau\tau$ events. The exclusion of the transverse component is clear visible in the `BoostZ`-method. Although the transverse boost is taken into account in the `BoostXYZ`-method it only managed to correct for this in a fraction of the events.



(a) Difference between true and estimated transverse direction for $H \to \tau\tau$ for different values of E_T^{miss}.

(b) RMS of the difference between true and estimated transverse direction as function of E_T^{miss} magnitude.

Figure 4.9: Estimation of the transverse direction heavily depend on the magnitude of the E_T^{miss}, but for most part of events summing the τ-momenta and missing transverse momenta gives a reasonable estimate.

A method exists – called *Collinear Approximation* [31] – which corrects for the missing neutrinos by assuming that the neutrinos are collinear with the mother τ-lepton. By doing this it can include $E_{\mathrm{T}}^{\mathrm{miss}}$ by projecting it onto the τ-jet directions. However, this projection is only possible and reliable in a fraction of the events with a large transverse boost and well aligned $E_{\mathrm{T}}^{\mathrm{miss}}$ [32].

Due to the large fraction of events where the collinear mass approximation fails, most analyses involving pairs of τ-leptons have switch to likelihood based techniques. In ATLAS, the most frequently used is called the missing mass calculator, MMC [33]. It uses probability density functions based on Monte Carlo simulations to assign a probability for the boson mass given the event topology of the visible particles – not just the two τ-lepton– and $E_{\mathrm{T}}^{\mathrm{miss}}$. In the process, it exploits a lot of assumptions from τ-polarisations to object reconstruction and pileup activity in the detector. For most cases, these assumptions can be evaluated by studying control regions or pre-identified objects, but in searches for new particles, these assumptions rely heavily on Monte Carlo predictions. Furthermore, the constant re-tuning to new detector environment and the event-on-event calculations can be very computational intensive tasks.

A more detailed description of the present mass techniques is found in section 6.1. The technique presented here functions as an alternative way of estimating a boson mass.

4.2.1 Finding the kinematic edge

In the RF of a decaying boson the energy of either of τ-jets cannot exceed half of the boson mass. Therefore the mass could be found by finding the end-point of twice the energy of the leading τ-jet in the RF. A quantity which henceforth will be denoted, M_{Boost}.

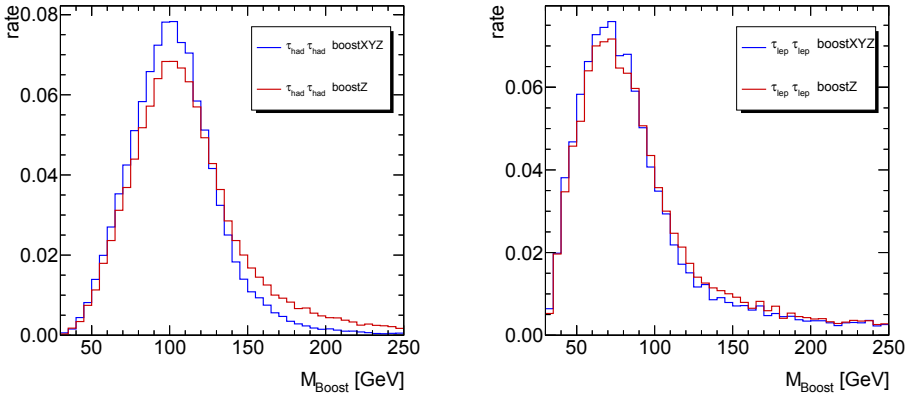
When both τ-leptons decays into hadrons the distribution of M_{Boost} – shown in figure 4.10(a) – will accumulate towards the end-point due to the presence of a single neutrino in each τ-decay, generating an almost triangular shape. Finding the kinematic edge of such a distribution is preferable to finding the end-point of a broad distribution, e.g. $\mathrm{M_{vis}}$, when including backgrounds distributions. The triangular shape is distorted by several effects, τ-polarisation, the detector resolution and the width of the resonance all smear the distribution into a distribution which here is approximated with a triangle convoluted with gaussian.

For the final states involving leptonic τ-decays, the peak of the M_{Boost}-distributions are shifted towards lower values, resulting in a less profound edge. However, estimating the edge in these events is still possible. The M_{Boost}-distributions for the leptonic final states are displayed in figure 4.10(b)

As opposed to the mass obtained from the collinear approximation, $\mathrm{M_{CA}}$, the M_{Boost} quantity can be calculated for all events without cpu intensive calculations since the both `BoostZ` and `BoostXYZ` methods are computational fast.

Fitting a triangle convoluted with a gaussian function to the distribution in figure 4.10(a) raises some numerical difficulties since such a fit is highly unstable due to the correlations between the gaussian width and the triangular end-point.

Numerical calculations in MATHEMATICA shows that taking the position of the steepest point in the M_{Boost} distribution as an estimator of the boson mass, $\hat{m}_{\mathrm{boost}}$, gives a reasonable agreement. Only a small offset towards a lower estimated mass is expected. The $\hat{m}_{\mathrm{boost}}$ is trivial to calculate, but will show disadvantages in low statistics scenarios.



(a) Fully hadronic final states

(b) Fully leptonic final states

Figure 4.10: Distributions of M_{Boost} in ggH $\rightarrow \tau\tau$ event with either the fully hadronic or fully leptonic final states. The M_{Boost} distributions have been made both using both the `BoostZ` and `BoostXYZ`-method.

To evaluate the precision of $\hat{m}_{\text{boost}}$, a series of pseudo-experiments are performed. Using the shape from the distributions in figure 4.10, 1.000 pseudo-distributions are generated with a fixed number of events and $\hat{m}_{\text{boost}}$ is calculated using `BoostXYZ`. This is repeated with different numbers of events in each pseudo-distribution and the resulting mean and spread of $\hat{m}_{\text{boost}}$ as a function of the event statistic is shown in figure 4.11.

Instead of estimating the boson mass from the steepest point in the distribution, a more precise result might be achieved by fitting the M_{Boost} distribution using templates. This approach is proposed for measuring the mass of the W-boson [34], but requires a huge amount of simulations in order to generate the templates. Therefore, an approach using of $\hat{m}_{\text{boost}}$ is more feasible.

4.3 Helicity correlations

The helicity of a particle is given as the projection of its spin onto its momentum. Since a τ-lepton is a spin-$\frac{1}{2}$ fermion, it can have either -1 (left-handed) or +1 (right-handed) helicity. Information on the helicity is in principle achievable, since the decay of the τ-lepton is affected by its helicity and the τ-lepton decays inside the detector, but the information is hidden by the fact that the visible decay product cannot be compared to the originating τ-lepton. Therefore, one cannot hope to reconstruct the polarisation for a single τ-lepton. Instead the τ-polarisation is defined on a statistical basis as the relative difference in cross sections for left-handed, σ_L, and right-handed,σ_R, τ-leptons

$$P_\tau = \frac{\sigma_L - \sigma_R}{\sigma_L + \sigma_R} \tag{4.1}$$

At the LEP experiments, several powerful observables [35] were constructed for studying the polarisation of τ-leptons in Z^0-decays. Since Z^0-bosons were produced at rest at

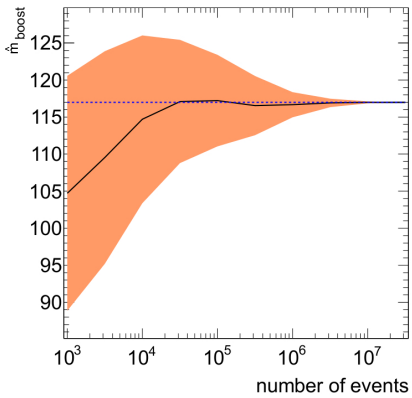


Figure 4.11: Performance of $\hat{m}_{\text{boost}}$ as a function events in the distribution in simulated ggH events with $m_H = 120$ GeV. The performance is evaluated using 1000 pseudo-experiments for each point in the input statistics. The solid lines indicate the means and the filled areas indicate the spread of the distribution.

LEP, all these variables are defined in the RF of the Z^0-boson. As previously demonstrated, bosons producing pairs of τ-leptons at the LHC will be highly boosted. Therefore a reconstruction of the RF is vital in order to use these observables at LHC.

The most dominant effect of τ-polarisation is present in the single charge pion decay channel, $\tau^\pm \to \pi^\pm \nu_\tau$, hereafter referred to as the *pion mode*. Due to the single neutrino produced in the pion mode and that neutrinos have to be left-handed, a right-handed τ-lepton will prefer to emit a neutrino in the direction of flight as pictured in figure 4.12.

In the Z^0 rest frame, the decay angle, i.e. direction of the visible τ-jet, translates directly to the fraction of the τ-lepton energy carried by the visible τ-jet, e.g. $\frac{E^-_\pi}{E^-_\tau}$, as illustrated in figure 4.13(a) for the pion mode. In the Monte Carlo simulation format used in ATLAS, the helicity information is not stored. Therefore this plots are made using generator level quantities in PYTHIA8+TAUOLA [36].

Since one does not have access to the true energy of the τ-lepton, the π-meson energy is normalised by the mass of the boson which in the correct RF should be twice E_τ. Henceforth, E_{vis} will denote energy of the visible τ-jet.

The correlation between helicity and decay angle is also present in the other decay modes, although it is much weakened – in the leptonic modes due to the second neutrino and the $\tau^\pm \to \rho^\pm \nu \to h^\pm h^0 \nu$ decay, referred to as the rho mode, due to the non-zero spin of the ρ-meson. However in the rho mode, which constitutes $\sim 25\%$ of all τ-decays, the decay of the ρ-meson can be used to estimate the polarisation. The decay directions for the neutral and charged pion correlates to the energy sharing between them in the Z^0 frame. As shown in figure 4.13(b), the charged energy fraction defined as

$$\Upsilon = \frac{E^{\text{ch}} - E^{\text{neu}}}{E^{\text{ch}} + E^{\text{neu}}} \tag{4.2}$$

in the τ-jet correlates well to the helicity of the τ-lepton. Since this quantity only compares an energy sharing of two very close objects, it only depends weakly on the frame of reference and it gives information on the τ-helicity based on a single τ-lepton without

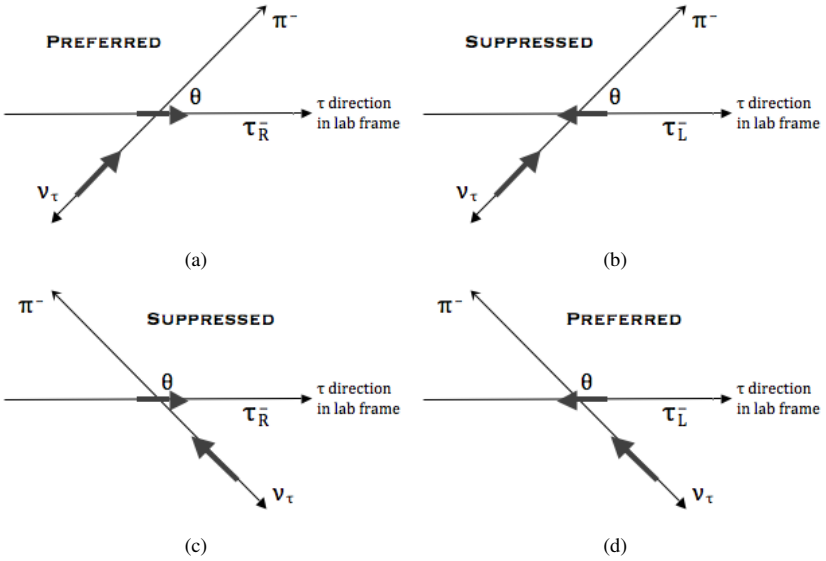
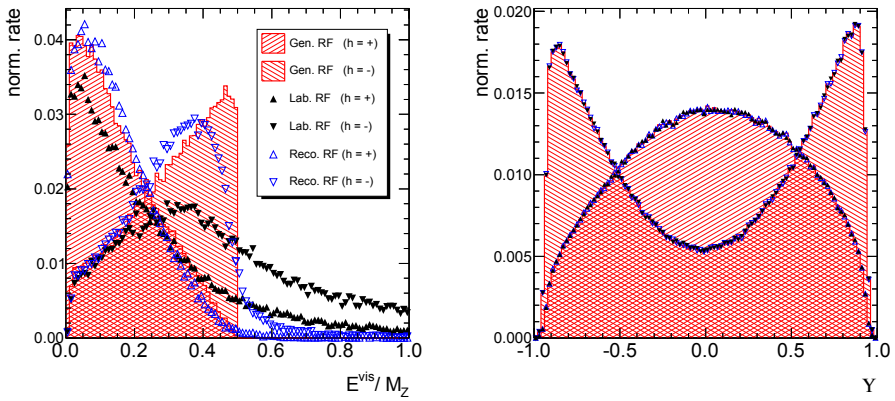


Figure 4.12: Illustration of the preferred and suppressed decay distributions for $\tau \to h^{\pm}\,\nu_{\tau}$.



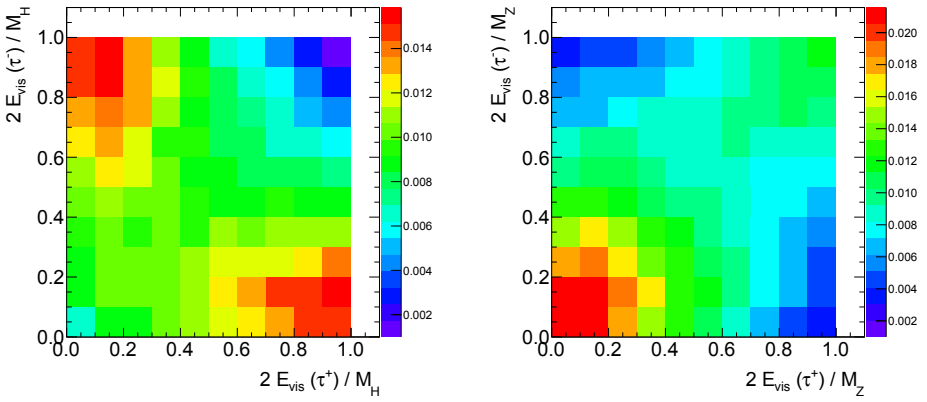
(a) E_{vis}/M_Z for τ-leptons decaying to $\pi^{\pm}\nu_{\tau}$.

(b) Υ for τ-leptons decaying to $\rho^{\pm}\nu_{\tau}$.

Figure 4.13: Polarisation variables for τ-leptons with positive and negative helicities shown in the generated and reconstructed RF using the `BoostXYZ`-method as well as the laboratory frame. Improvements when going from laboratory to the reconstructed RF are only seen for the Υ.

using the mass of the boson nor information about the mother τ-lepton. Therefore this variable is very useful experimentally and was recently the primary variable used in the τ-polarisation measurement $W \to \tau\nu$ in ATLAS [37].

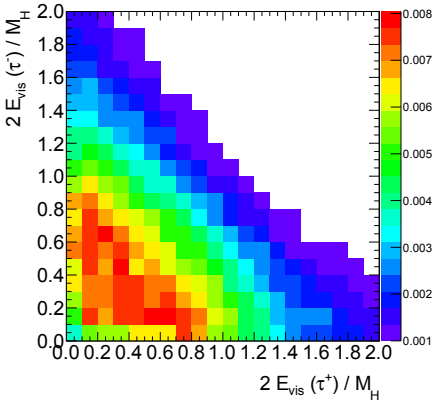
For a boson decaying to a pair of τ-leptons, the helicity configuration of the τ-leptons is determined by the spin of the boson. A spin 1 boson, e.g. Z^0, will produce a set of τ-leptons with same helicity while a spin 0 boson, e.g. a Higgs boson, will produce the opposite. Thus, if reconstructed well enough, the helicity correlations of the τ-leptons can determine the spin of the boson. Figure 4.14 shows the normalised E_{vis} distributions for pairs of τ-leptons coming from Higgs and Z^0 bosons in the true RF of the boson. For Higgs bosons, the distributions are concentrated around (1,0) and (0,1) while for Z^0 events the distribution are predominantly located at (0,0) and (1,1). The larger peak around (0,0) than at (1,1) is due to the axial-vector coupling of the Z^0.



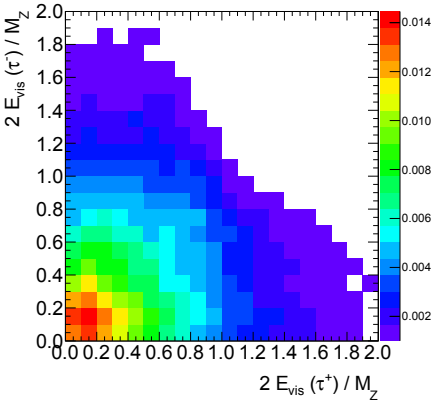
(a) Spin 0 (H) (b) Spin 1 (Z^0).

Figure 4.14: Energy correlations of the visible τ-jet observed in the true RF of the boson. Only events where both τ-leptons decaying to $h^{\pm}\,\nu_{\tau}$are considered. To avoid effects coming from the mass differences all energies are scaled with the mass of the boson.

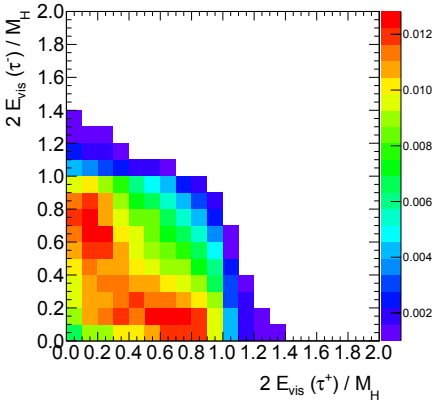
These correlations are mostly washed out when observing the τ pair from the laboratory frame as seen in figures 4.15(a)-4.15(b), however some of the correlations are regained when reconstructing the RF as seen in figures 4.15(c)-4.15(d) for the `BoostXYZ`-method.



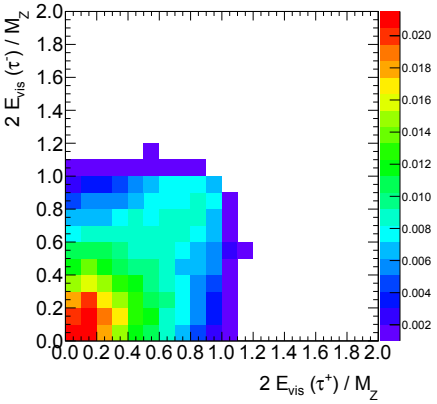
(a) Spin 0 (H sample).



(b) Spin 1 (Z^0 sample).



(c) Spin 0 (H sample).



(d) Spin 1 (Z^0 sample).

Figure 4.15: Energy correlations between the visible τ-jets, observed in the detector frame (a,b) and in the reconstructed RF using the `BoostXYZ`-method (c,d). Only events where both τ-leptons decaying to $h^{\pm}\,\nu_{\tau}$ are considered. To avoid effects coming from the mass differences all energies are scaled with the mass of the boson.

CHAPTER 5

Searching for the Higgs boson in the τ channel

As described in section 2.1, the $H \to \tau\tau$ decay mode becomes an important search channel in the low Higgs mass range. In ATLAS the $\tau^+\tau^-$ channel is divided into three subchannels, the fully leptonic final states, $\tau_\ell\tau_\ell$, where both τ-leptons decay into leptons, the semi-leptonic final states, $\tau_\ell\tau_{\text{had}}$, where one τ-lepton decays into leptons and the other into hadrons, and finally the fully hadronic final state, $\tau_{\text{had}}\tau_{\text{had}}$, where both τ-leptons decay into hadrons.

In this chapter, a search for the Higgs boson in the $\tau_{\text{had}}\tau_{\text{had}}$ final state is presented. This final state suffers from a large background of QCD-jets which have a very similar signature in the detector. Furthermore, the QCD backgrounds are known not to be well described by simulations and therefore a urgent need for data-driven background estimations is present. However, as will be shown the different topology of QCD jet event compared to Z^0 and Higgs events, can be used quite effectively to suppress the QCD background, and in fact the irreducible background from Z^0 bosons will turn out to be the main obstacle.

On the other hand, the $\tau_{\text{had}}\tau_{\text{had}}$ final state gains by having a large branching ratio, BR, compared to the $\tau_\ell\tau_\ell$ and $\tau_\ell\tau_{\text{had}}$ final states. Table 5.1 summarises the cross sections for all production modes of the Higgs boson for different mass hypotheses in proton-proton collisions with a centre-of-mass energy of 7 TeV. The last column reports the branching ratio of the Higgs boson decaying to a pair of τ-leptons. The $\tau_{\text{had}}\tau_{\text{had}}$ final state accounts for 42% of all τ-pairs: $\text{BR}(\tau \to h) \times \text{BR}(\tau \to h) = 0.648^2 = 0.4199$.

The selection of Higgs candidates events is divided into two steps: a preselection step and a signal region selection. The preselection performs a minimal set of cuts to select any pair of real τ-leptons. This stage in the selection also works as a control region where agreement between data and simulation is validated. Ideally, a guaranteed signal-free region would be the best control region, however as long as the mass of the Higgs boson is low or unknown, such a region is hard to obtain. Instead the preselection stage is used. At this stage, the *contamination* from Higgs events compared to the background processes is very low and therefore this region can serve as a control region.

After the preselection an optimised selection of Higgs events is performed. The optimised selection is split into a ggH and a VBF selection. The resulting events of each

m_H [GeV]	ggH [pb]	VBF [pb]	WH [pb]	ZH [pb]	BR ($H \to \tau\tau$)	$H \to \tau_{\text{had}}\tau_{\text{had}}$ [pb]
100	24.00	1.546	1.186	0.631	0.0836	0.961
105	21.77	1.472	1.018	0.545	0.0825	0.859
110	19.84	1.398	0.875	0.472	0.0802	0.761
115	18.14	1.332	0.755	0.411	0.0765	0.663
120	16.65	1.269	0.656	0.360	0.0710	0.564
125	15.32	1.211	0.573	0.316	0.0637	0.466
130	14.16	1.154	0.501	0.278	0.0548	0.370
135	13.11	1.100	0.439	0.245	0.0452	0.283
140	12.18	1.052	0.386	0.217	0.0354	0.206
145	11.33	1.004	0.341	0.193	0.0261	0.141
150	10.58	0.962	0.300	0.171	0.0178	0.090

Table 5.1: NNLO cross sections for the SM Higgs production in proton-proton collision at 7 TeV and branching ratios for Higgs decaying in a pair of τ-lepton [38, 39].

of these categories are passed on to the limit setting described in chapter 9.

5.1 Background processes

Since the final state in this analysis consist of two hadronic final states, a large background arises from QCD-events, and the sensitivity of the analysis depends strongly on the ability to separate the huge amount of fake τ-pairs from QCD events from real τ-pairs.

Unfortunately, simulations does not describe QCD events very well. Therefore, a *data-driven* estimation of the QCD events is used. This method, described in section 7.2, consists in taking pairs where the τ-candidates do not have opposite reconstructed electrical charge, referred to as not-opposite-sign events. Since the rate of not-opposite-sign and opposite-sign pairs coming from QCD events will not be identical, the number of not-opposite-sign events is normalised by the $\Delta\eta(\tau_1, \tau_2)$ fit described in section 7.4.

The second large background is the irreducible background from $Z \to \tau\tau$ events. This background process is estimated using the partly data-driven method of *embedded* $Z \to \mu\mu$ events as described in section 7.1.

The background processes and their cross sections are listed in table 5.2

Process	Cross section [pb] ($\times$BR)
$W \to l\nu$ ($l = e, \mu, \tau$) + jets	10.46×10^3
$Z^0 \to ll$+ jets ($l = e, \mu$)	0.71×10^3
$t\bar{t}$	166.7
Single top $t-$, $s-$ and Wt-channel	58.7, 3.9, 13.1
WW,WZ^0 and Z^0Z^0	46.2, 18.0, 5.6

Table 5.2: Cross sections of Monte Carlo background processes. The cross sections for W and Z^0 includes a sum over the branching ratios indicated final states.

5.2 Input Samples

The analysis presented here uses data from proton-proton collisions with the centre-of-mass energy of 7 TeVrecorded by the ATLAS experiment. All data taken in 2011 with all the subsystems of the ATLAS detector operating is used. This amounts to 4.7 fb^{-1} of data.

The event generations of the Higgs samples are done with POWHEG [40, 41, 42] which is based on next-to-leading order (NLO) perturbative calculations. The parton shower, hadronisation and underlying event simulations are provided through an interface with PYTHIA8 [30]. The associated production samples are based on the leading order simulations using PYTHIA8. Several mass points have been produced and used in this analysis.

In the background processes, the W/Z^0 + jets event generation is done with ALPGEN [43]. This generator employs the MLM matching scheme [44] between the hard process, calculated with leading-order matrix elements for up to five jets, and the parton shower. The $t\bar{t}$ and di-boson (WW,WZ^0,Z^0Z^0) samples are produced with MC@NLO [45] with NLO precision. Single-top events are simulated with ACERMC+PYTHIA [46]. The loop-induced $gg \to WW$ processes are generated using GG2WW [47].

In all the background Monte Carlo samples, the parton shower and the hadronisation are simulated using HERWIG [48] and the underlying event using JIMMY [49]. The τ-decay is simulated using TAUOLA [36], and PHOTOS [50] is used to simulate additional photon radiation from the charged leptons.

The detector simulation in all the Monte Carlo samples is performed with GEANT4 [27] and the reconstruction is made with the same software as is used for data.

In order to get a correct modelling of the different pile-up conditions in all of the data taking periods, the Monte Carlo samples are reweighted according to the average number of interactions per bunch crossing. This reweighting is done by comparing distributions of the average number of interactions per bunch crossing in the recorded data and in the simulated events, and therefore the reweighting is dependent on the data periods used in each analysis.

Additional weights are applied on the simulations in order to correct for a mismodelling of the trigger efficiency [51]. These weights are derived as a function of p_{T} and η of the τ-candidates, and are smaller than 5% over the full p_{T} and η range.

For the ggH signal samples a reweighting is also done as a function of the p_{T} of the Higgs boson. This reweighting is done to incorporate NNLO effects of on the kinematics of the Higgs events.

5.3 Selection of τ-pairs

In this section, the preselection of hadronic decaying τ-pairs is described. The preselection aims to select a wide range of well reconstructed τ-pairs, while filtering out regions in which simulations are known not to describe data accurately.

The preselection is divided into three step: The first step selects *good* events, i.e. events which has been selected by the triggering system and which satisfy some Data Quality cuts, the second step selects the individual objects, i.e. leptons, jets, by a series of quality cuts on the individual objects and the final step selects events with a suitable

τ-object selection criteria
Cluster seeded
Electrical charge = ± 1
`BDT-Medium` Identification
BDT-based Electron Veto
Cut-based Muon Veto
Matched to the EF Trigger within $\Delta R < 0.2$
$\lvert\eta\rvert < 2.5$
$p_{\mathrm{T}} > 20$ GeV
Leading track not within $1.37 < \lvert\eta\rvert < 1.52$

Table 5.3: Selection criteria for the tau candidates.

pair of hadronically decaying τ-leptons.

5.3.1 Identification of Hadronic τ-decays

Since τ-leptons decaying into leptons will be reconstructed as either a electron or a muon in the detector, reconstructed τ-candidates only refers to the τ-leptons decaying into hadrons. The τ-reconstruction is performed using clusters in the electromagnetic and hadronic calorimeters [52] and is seeded by a jets with $p_{\mathrm{T}} > 10$ GeV, reconstructed using the anti-k_{t} jet-algorithm [53]. Reconstructed τ-candidates are required to are have $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.5$ to be within the reach of the tracking detectors. Furthermore, to avoid the transition regions from the barrel to end-cap, the leading[1] track of the candidates are required not to be within $1.37 < |\eta| < 1.51$.

For the τ-identification a Boosted Decision Tree (BDT) algorithm is used [52] to separate true τ-leptons from fakes from QCD-jets. The BDT algorithm is based on input from shower profile in the calorimeter and topology of the associated tracks. Three standard cuts on the BDT score have been defined by the τ-identification group, `Loose`, `Medium` and `Tight`, depending on the purity of the selection. Here both τ-candidates are required to be `Medium` which corresponds to an efficiency of approximately 45%. Furthermore, a specific veto again electron and muon with is done each candidate using a cut-based approach for the muon veto and a BDT-based one for the electron rejection [52]. Finally, each candidates are required to be matched to the region of interest (RoI) where the trigger was fired. The selection criteria for τ-leptons are summarised in table 5.3.

5.3.2 Trigger

The first step of selecting τ-pair is choosing the trigger from the ATLAS trigger menu described in section 3.7. In this analysis, different double hadronic triggers are chosen depending on the run period due to the increasing luminosity of the LHC. Usually the unprescaled trigger with the lowest p_{T} requirement is chosen. All the used triggers are designed with an asymmetric p_{T} requirement for the two candidates, which is lower at Level 1 than at Event Filter Level. The `medium` in the trigger name, refers to the tightness

[1]The highest p_{T} track is referred to as the leading track.

of the identification at the trigger stage. Since less information is available during the trigger processing than in the offline reconstruction, this identification value will not match completely the BDT value in the offline identification. Some of the triggers also have a requirement on the number of associated tracks to be either 1, 2 or 3 which is marked with an `1` in the trigger name. The used triggers for the different periods are listed in table 5.4.

EF Trigger	Periods	$\mathcal{L}$ [pb^{-1}]
EF_tau29_medium1_tau20_medium1 (loose)	B-E	222
EF_tau29_medium1_tau20_medium1 (default)	F-K	2047
EF_tau29T_medium1_tau20T_medium1	L-M	2400

Table 5.4: Summary of the triggers used through the 2011 data taking.

5.3.3 Event Cleaning

In order to only consider good quality data a set of event cleaning cuts is applied. The first of these cuts is is *GoodRunsList*(GRL) filtering provided by the ATLAS data quality group which guarantees that all parts of the detector have reported a green flag at the time of this event. To further remove non-collision events, e.g. beam halo and cosmic rays events, at least one reconstructed primary vertex is required with more than four associated tracks. These cuts are only applied to data. On top for this a cleaning cut is applied to the jets in both simulated events and in data. This cut provided by the JetEtMiss Performance group [54] rejects event poorly reconstructed or out-of-time jets.

5.3.4 Muons and Electrons

Since this analysis only focuses on events with two hadronic decaying τ-leptons, a veto again events containing electrons or muon is done. Here the leptons are required to better reconstructed than in the lepton veto used on the individual τ objects.

The rejection of events containing reconstructed leptons also makes the combination with the semi-leptonic and fully-leptonic final states easier since an event cannot enter in both analysis and double counting is avoided.

The muons are chosen from the STACO algorithm [55] and required at be identify as `Loose` tag. The muon are required to have $p_{\mathrm{T}} > 10$ GeV and be within $|\eta| < 2.5$. The muon are required to have at least certain number of hits in each subdetector in the ID, and no more the three holes-on-track[2]. Additionally, a set of standard quality cuts is applied on the associated track. For the electrons is reconstruction is seeded by clusters in the electromagnetic calorimeter. Electrons are required to have an associated track and they are required to be tagged as `MediumPP` by the identification algorithm. They are required to have $p_{\mathrm{T}} > 10$ GeV and $|\eta| < 2.47$, and not be in the transition region at $1.37 < |\eta| < 1.52$. As for the muons additional quality criteria made by the EGamma performance group is imposed.

The lepton selection criteria are listed the table 5.5.

[2]The number of holes-on-track are defined as the expected number of hits from the reconstruction track trajectory over the measured number of hits

Electrons	Muons
$p_{\mathrm{T}} > 15$ GeV	$p_{\mathrm{T}} > 10$ GeV
Author = 1 or 3	Staco reconstructed
Isolated and `medium`	`Loose`
$\lvert\eta\rvert < 2.47$	$\lvert\eta\rvert < 2.5$
Not in $1.37 < \lvert\eta\rvert < 1.52$	Hits in all ID subdetectors
Charge, $\lvert Q_e\rvert = 1$	Less than 3 holes-on-track
Tagged as well-reconstructed	Tagged as well reconstructed

Table 5.5: Criteria for lepton definitions used in the lepton veto.

5.3.5 Jets

In ATLAS jets are reconstructed jets using a set of different clustering algorithms. The default clustering algorithm in ATLAS – also used in here – is the anti-k_{t} -algorithm [53] with an cone size of ΔR 0.4. The input for the clustering is topological clusters [56] calibrated with the Local Hadronic Calibration scheme [57] which calibrates the energy deposit in the electromagnetic and hadronic calorimeters using weighting scheme based on different response for electrons and pions obtained by simulation and test-beam data. This calibration also takes into account the dead material and out-of-cluster cells to get the optimal response and noise suppression.

Only jets with $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 4.5$ are selected. Furthermore, jets which are within a cone of $\Delta R < 0.2$ of the τ-candidates are removed.

5.3.6 Missing Transverse Energy

The missing transverse energy is reconstructed based on the energy deposited in calorimeters and on the muons reconstructed in the muon spectrometer [58]. The energy deposited in the cells is calibrated according to the reconstructed physics object to which they belong. The cells are associated to reconstructed objects in the following order: electrons, photons, hadronically decaying τ-leptons, jets and muons. Also cells not associated to any of these objects are taken into account. All the calibrated energies are summed up to form a vector in the transverse plan. From all selected muons the summed p_{T} is added to this vector.

5.3.7 Pair selection

Besides the trigger and event cleaning cuts, a series of cuts are used to define selected events of possible Higgs candidates. First of all, a veto against leptons selected by the criteria described in section 5.3.4 are performed. Exactly two τ candidates are required using the selection documented in section 5.3.1 with at least one of the two candidates have `Tight` identification. The two candidates are required to have p_{T} above 35 GeV and 25 GeV. These p_{T} requirements are mainly imposed to get rid of effects from the trigger turn-on which is known to be poorly modelled in the simulations. Furthermore, the two candidates are required to have opposite charges and no outlying tracks counted as described in section 7.3.1.

A cut on the visible mass of the selected τ-pairs is imposed by removing events with $M_{vis} < 60$ GeV since the background modelling of QCD and $Z \to \tau\tau$ events is known to be inaccurate at low masses. For $Z \to \tau\tau$ events the inaccuracy comes from the fact that only $Z \to \mu\mu$ events above a certain threshold is used as input for the $Z \to \tau\tau$ modelling as described in section 7.1. Furthermore, since the direction of the E_T^{miss}-vector is used further on in the selection, a cut requiring $|E_T^{miss}| > 10$ GeV is imposed. This analysis tries to make minimal use of the E_T^{miss}, because large uncertainties are associated with the E_T^{miss} reconstruction.

Finally, in order to reject QCD events only τ-pairs with $\Delta\eta(\tau_1, \tau_2) < 2$ are selected. Figure 5.1 shows the $\Delta\eta(\tau_1, \tau_2)$ distribution before making this cut. Clearly, true τ-jets are produced much closer in η than fake τ-jets from QCD.

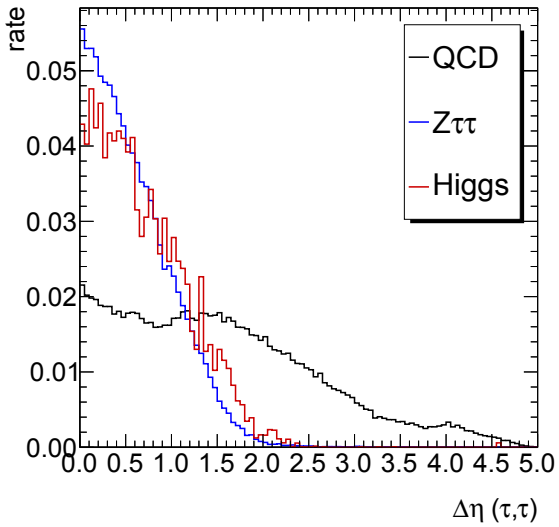


Figure 5.1: Normalised distributions of $\Delta\eta(\tau_1, \tau_2)$ for during preselection before cutting at $\Delta\eta(\tau_1, \tau_2) = 2$. The Higgs processes includes both ggH, VBF and VH events which are weighted by their cross sections before the normalisation.

Table 5.6 summarised the expected and observed number of events surviving the steps in the preselection. Here both background and signal samples are kept split up to reflect the signal and background composition. Only the statistical uncertainties on the numbers are presented in table 5.6.

After applying the preselection cuts, the dominant background sources are $Z \to \tau\tau$ and QCD jets. Both of these background processes are normalised using the fit described in section 7.4. The normalisation is done before the $\Delta\eta(\tau_1, \tau_2)$-cut, and this normalisation are kept throughout the selection steps. For all other background and signal processes the cross sections listed in table 5.2 and table 5.1 are used.

Events

Cut	Data	Background							Signal ($m_H = 125$ GeV)		
		Total	QCD-jets	$Z \to \tau\tau$	W	top	$Z \to \ell\ell$	Di-Bosons	ggH	VBF	VH
Trigger	9867680	-	-	111646.0	201130.0	59752.8	1200831.8	17013.4	196.7	30.7	18.2
Exactly 2 τ-jets	51901 ± 227.8	50289.7 ± 83.8	41174.9 ± 64.4	8050.6 ± 27.6	966.3 ± 43.9	53.4 ± 2.0	23.6 ± 13.1	20.9 ± 3.9	37.2 ± 0.8	5.4 ± 0.1	2.0 ± 0.1
$M_{vis} > 60$ GeV	49376 ± 222.2	48266.5 ± 82.2	39427.6 ± 63.0	7846.0 ± 27.1	905.5 ± 43.0	45.5 ± 1.8	23.4 ± 13.1	18.6 ± 3.9	36.6 ± 0.8	5.1 ± 0.1	1.9 ± 0.1
$\lvert E_T^{miss}\rvert > 10$	35472 ± 188.3	35477.4 ± 73.5	28368.1 ± 53.6	6169.2 ± 24.1	857.1 ± 41.8	43.9 ± 1.7	21.6 ± 13.0	17.4 ± 3.9	30.7 ± 0.8	4.7 ± 0.1	1.7 ± 0.1
$\Delta\eta(\tau_1, \tau_2) < 2$	26741 ± 163.5	26619.8 ± 65.7	19667.2 ± 44.9	6142.0 ± 24.1	740.4 ± 39.5	36.9 ± 1.5	21.0 ± 13.0	12.3 ± 0.7	30.4 ± 0.7	4.6 ± 0.1	1.7 ± 0.1

Cut Efficiencies

Cut	Data	Background							Signal ($m_H = 125$ GeV)		
		Total	QCD-jets	$Z \to \tau\tau$	W	top	$Z \to \ell\ell$	Di-Bosons	ggH	VBF	VH
Exactly 2 τ-jets	0.005	–	–	–	0.005	0.001	<0.001	0.001	0.189	0.176	0.110
$M_{vis} > 60$ GeV	0.951	0.960	0.958	0.975	0.937	0.852	0.992	0.890	0.984	0.944	0.950
$E_T^{miss} > 10$ GeV	0.718	0.735	0.719	0.786	0.947	0.965	0.923	0.935	0.839	0.922	0.895
$\Delta\eta(\tau_1, \tau_2) < 2$	0.754	0.750	0.693	0.996	0.864	0.841	0.972	0.707	0.990	0.979	1.000

Table 5.6: Cutflow table for preselection of τ-pairs. At each step of the selection the number accepted events including statistical errors is shown for both simulations and data. In the bottom table the efficiencies of each cut wrt. the previous step in shown.

5.3.8 Validation

The purpose of validation procedure is to check if the variables used in the analysis behave according to expectation, and evaluate the level of disagreement in terms of systematic errors of the final results. In the procedure it is desirable to define signal free control regions where particular backgrounds can be studied and correction factors can be extrapolated to the signal region. However, no region has been found which is both guaranteed signal-free and in which all variables of interest can be validated. Therefore, the preselection stage is used as validation region. A validation of all variables used in the analysis is presented in appendix B.

Figure 5.2 shows the predicted and observed distributions of the mass variable $M_{\text{Boost}-\text{XYZ}}$, along with a ratio plot of the observed data over the expected background. Only statistical errors and errors from the normalisation procedure are included in the ratio plot, but even then good agreement is found between data and expectation. The excepted signal in case of a 125 GeV Higgs boson is also included. For visual purposes the Higgs signal is enhanced by a factor 30.

Even though good agreement is found in the $M_{\text{Boost}-\text{XYZ}}$ presented here, discrepancies are found in some variables such as the $E_{\text{T}}^{\text{miss}}$.

5.4 Optimised Higgs Selections

After having preselection τ-candidate pairs, an optimised selection of Higgs events are done. Since ggH and VBF events have slightly different signatures, two different selection strategies are applied to the event, resulting in two signal regions or categories. The two strategies are optimised to select predominantly VBF or ggH events. Since the expected sensitivity is highest for VBF due to the small theoretical uncertainty on VBF production, the VBF-selection is executed first. This means that only events which are not accepted in the VBF category, is passed on to the ggH selection.

5.4.1 VBF category

Since VBF Higgs bosons will be more boosted in the transverse plan than Z^0 events as discussed in chapter 4, a cut is made on the reconstructed transverse boost of the τ-pair system using the boost reconstruction method `BoostXYZ`. The reconstructed boost is required to have $\beta_{\text{T}} > 0.3$ and $\beta_{\text{z}} < 0.9$. Figure 5.3 shows the reconstructed β_{z} and β_{T} distributions for signal and two main background processes.

A cut is also put on the alignment of the $E_{\text{T}}^{\text{miss}}$-vector. In order to expressed the position of the $E_{\text{T}}^{\text{miss}}$-vector in the transverse plan relative to the τ-jets, a continuous variable $E_{\text{T}}^{\text{miss}}$-centrality is defined by

$$E_{\text{T}}^{\text{miss}} - \text{centrality} = \frac{A + B}{\sqrt{A^2 + B^2}},$$

where

$$A = \frac{\sin(\phi_{E_{\text{T}}^{\text{miss}}} - \phi_{\tau_1})}{\sin(\phi_{\tau_2} - \phi_{E_{\text{T}}^{\text{miss}}})}, \quad B = \frac{\sin(\phi_{\tau_2} - \phi_{E_{\text{T}}^{\text{miss}}})}{\sin(\phi_{\tau_2} - \phi_{E_{\text{T}}^{\text{miss}}})}$$

which is above 1 when the $E_{\text{T}}^{\text{miss}}$-vector is in between the two τ-jets in the transverse plan. The $E_{\text{T}}^{\text{miss}}$-centrality is defined in the range $[-\sqrt{2}, \sqrt{2}]$ with $\sqrt{2}$ being right between the

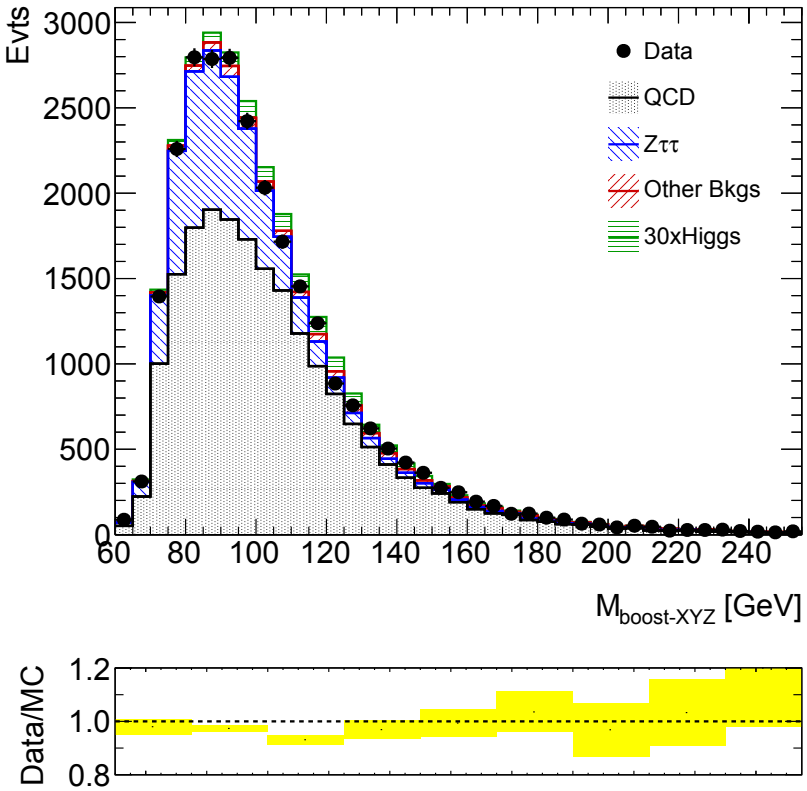
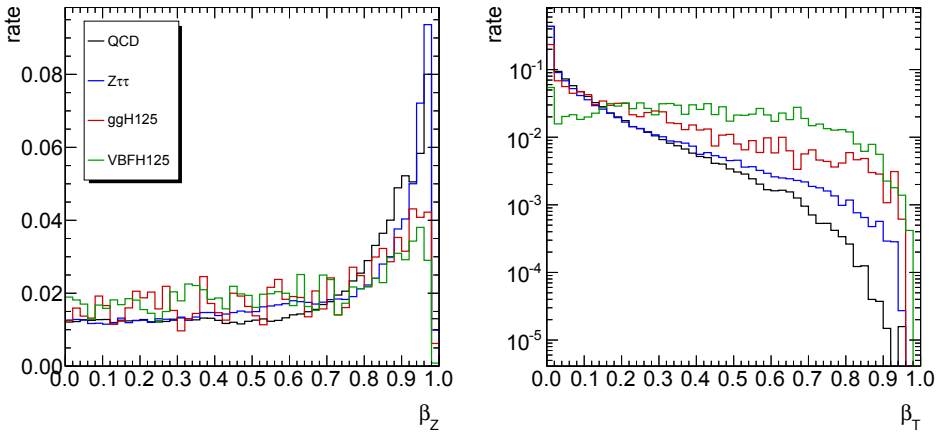


Figure 5.2: Validation plot of the $M_{\text{Boost}-\text{XYZ}}$ variable at preselection. Higgs signal which includes both ggH, VBF and VH events has been enhanced by a factor 30 in order to be visible. A ratio plot is shown between observed data and background predictions (labeled MC). Only statistical errors and uncertainties from the normalisation procedure are included in the ratio plot.

two τ-jets. The normalised distributions at the preselection stage of the $E_{\text{T}}^{\text{miss}}$-centrality is shown in figure 5.4. Only events with $E_{\text{T}}^{\text{miss}}$-centrality above zero is selected.

The VBF event topology is characterised by having two forward jets in the detector with a large rapidity gap as mentioned in section 2.1.2. Therefore, events are required to have at least two jets with $p_{\text{T}} > 30$ GeV and with these two jets are satisfying $\Delta\eta(\text{jets}) > 2.5$ and $\eta_1 \times \eta_2 < 0$.

The event cutflow from the preselection to the VBF category is presented in table 5.7 for the background processes and 125 GeV Higgs samples. Table 5.7 also lists the efficiencies for each cut in the selection for the different samples.



(a) Longitudinal boost (b) Transverse boost

Figure 5.3: Normalised distributions for the reconstructed boost variables at the preselection stage for the two main background and signal processes.

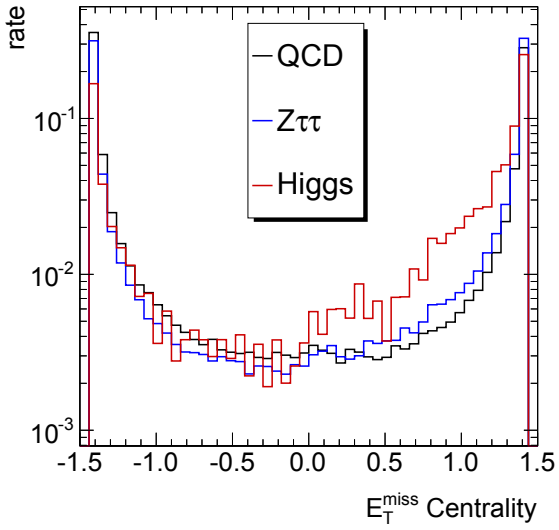


Figure 5.4: Normalised distributions (at logarithmic scale) of $E_{\mathrm{T}}^{\mathrm{miss}}$-centrality defined in equation 5.1 at the preselection stage. The Higgs processes includes both ggH, VBF and VH events which are weighted by their cross sections before the normalisation.

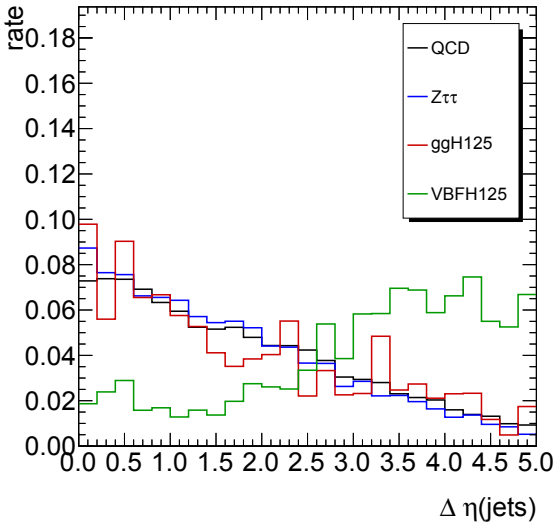


Figure 5.5: Normalised distribution of $\Delta\eta$(jets) at preselection stage. The ggH and VBF Higgs processes are displayed separately to show the distinctive signature of VBF events.

Events

Cut	Background							Signal ($m_H = 125$ GeV)		
	Total	QCD-jets	$Z \to \tau\tau$	W	top	$Z \to \ell\ell$	Di-Bosons	ggH	VBF	VH
$\beta_T > 0.3$, $\beta_z < 0.9$	2376.6 ± 19.1	1589.8 ± 12.8	674.2 ± 8.6	92.2 ± 11.2	16.0 ± 1.1	0.0 ± 0.0	4.3 ± 0.4	7.8 ± 0.4	2.7 ± 0.1	0.8 ± 0.0
E_T^{miss}-centrality> 0.5	991.2 ± 11.2	555.3 ± 7.5	398.4 ± 6.7	27.7 ± 4.7	7.3 ± 0.7	0.0 ± 0.0	2.4 ± 0.4	5.7 ± 0.3	2.3 ± 0.1	0.6 ± 0.0
≥ 2 jets with $p_T > 30$ GeV	225.7 ± 5.6	99.5 ± 3.3	107.5 ± 3.7	10.4 ± 2.5	6.5 ± 0.7	0.0 ± 0.0	1.9 ± 0.3	2.2 ± 0.2	1.5 ± 0.1	0.4 ± 0.0
$\eta_1 \times \eta_2$(jets)	99.2 ± 3.7	44.8 ± 2.2	45.1 ± 2.4	5.3 ± 1.8	3.2 ± 0.5	0.0 ± 0.0	0.7 ± 0.2	1.1 ± 0.1	1.3 ± 0.1	0.1 ± 0.0
$\Delta\eta$(jets) > 2.5	49.7 ± 2.6	26.3 ± 1.6	18.8 ± 1.5	3.4 ± 1.4	0.9 ± 0.3	0.0 ± 0.0	0.3 ± 0.1	0.6 ± 0.1	1.2 ± 0.1	0.0 ± 0.0

Cut Efficiencies

Cut	Background							Signal ($m_H = 125$ GeV)		
	Total	QCD-jets	$Z \to \tau\tau$	W	top	$Z \to \ell\ell$	Di-Bosons	ggH	VBF	VH
$\beta_T > 0.3$, $\beta_z < 0.9$	0.089	0.081	0.110	0.125	0.434	0.000	0.350	0.257	0.587	0.471
E_T^{miss}-centrality > 0.5	0.417	0.349	0.591	0.300	0.456	0.000	0.558	0.731	0.852	0.750
≥ 2 jets with $p_T > 30$ GeV	0.228	0.179	0.270	0.375	0.890	0.000	0.792	0.386	0.652	0.667
$\eta_1 \times \eta_2$(jets)	0.440	0.450	0.420	0.510	0.492	0.000	0.368	0.500	0.867	0.250
$\Delta\eta$(jets) > 2.5	0.501	0.587	0.417	0.642	0.281	0.000	0.429	0.545	0.923	0.000

Table 5.7: Cutflow table for the VBF category. At each step of the selection the number accepted events including statistical errors is shown for both simulations and data. In the bottom table the efficiencies of each cut wrt. the previous step in shown. The β_T and β_z are reconstructed using the `BoostXYZ`-method described in chapter 4. The cut efficiencies on in the first row are calculated wrt. the last of the preselection. table 5.6.

5.4.2 ggH category

The ggH category is optimised to mainly selected ggH events. The main discriminant in this selection is the $E_{\mathrm{T}}^{\mathrm{miss}}$ and the transverse boost of the τ-pair system, which is not as dominant as for VBF production, but still higher than the background processes.

First, only events with a high $E_{\mathrm{T}}^{\mathrm{miss}}$ value is selected. This cut mainly suppressed QCD since neutrinos are also present in the Z^0 and W events. Here a cut on $|E_{\mathrm{T}}^{\mathrm{miss}}|$ is chosen at 30 GeV.

In order to select boosted τ-pairs, a cut is put on the reconstructed boost using the `BoostXYZ`-method is required to have $\beta_{\mathrm{T}} > 0.2$ and $\beta_{\mathrm{z}} < 0.9$.

Furthermore, the $E_{\mathrm{T}}^{\mathrm{miss}}$-vector is required to satisfy $E_{\mathrm{T}}^{\mathrm{miss}}$-centrality$> 0$, i.e. the $E_{\mathrm{T}}^{\mathrm{miss}}$-vector must point in the same ϕ-hemisphere as the two τ-jets. The two τ-jets are required to be close in η by requiring $\Delta\eta(\tau_1, \tau_2) < 1.5$. And finally, the two τ-jets must be produced in associated with at least one hard jet (p_{T} of the jet must be greater than 50 GeV).

With this selection the background have been greatly reduce ($\sim 5\%$) while keeping a much higher acceptance for the signal ($\sim 25\%$ for ggHat $m_H = 125$ GeV).

Events

Cut	Background							Signal ($m_H = 125$ GeV)		
	Total	QCD-jets	$Z \to \tau\tau$	W	top	$Z \to \ell\ell$	Di-Bosons	ggH	VBF	VH
$\beta_T > 0.2$, $\beta_z < 0.85$	3708.7 ± 24.2	2594.6 ± 16.3	950.3 ± 10.0	139.8 ± 14.8	18.1 ± 1.1	1.2 ± 0.8	4.7 ± 0.4	9.7 ± 0.4	1.9 ± 0.1	0.9 ± 0.0
E_T^{miss}-centrality > 0	1907.0 ± 15.5	1212.5 ± 11.1	629.6 ± 8.2	50.7 ± 7.1	10.9 ± 0.8	0.6 ± 0.6	2.8 ± 0.4	7.5 ± 0.4	1.6 ± 0.1	0.7 ± 0.0
$\Delta\eta(\tau_1, \tau_2) < 1.5$	1601.7 ± 14.2	946.7 ± 9.8	601.9 ± 8.1	41.1 ± 6.2	8.7 ± 0.7	0.6 ± 0.6	2.7 ± 0.4	7.1 ± 0.4	1.5 ± 0.1	0.7 ± 0.0
≥ 1 jet with $p_T > 50$ GeV	510.7 ± 8.3	216.9 ± 4.8	259.4 ± 5.6	24.2 ± 3.8	8.1 ± 0.7	0.0 ± 0.0	2.2 ± 0.3	4.4 ± 0.3	1.2 ± 0.1	0.6 ± 0.0

Cut Efficiencies

Cut	Background							Signal ($m_H = 125$ GeV)		
	Total	QCD-jets	$Z \to \tau\tau$	W	top	$Z \to \ell\ell$	Di-Bosons	ggH	VBF	VH
$\beta_T > 0.2$, $\beta_z < 0.85$	0.139	0.132	0.155	0.189	0.491	0.057	0.382	0.319	0.413	0.529
E_T^{miss}-centrality > 0	0.514	0.467	0.663	0.363	0.602	0.500	0.596	0.773	0.842	0.778
$\Delta\eta(\tau_1, \tau_2) < 1.5$	0.840	0.781	0.956	0.811	0.798	1.000	0.964	0.947	0.938	1.000
≥ 1 jet with $p_T > 50$ GeV	0.319	0.229	0.431	0.589	0.931	0.000	0.815	0.620	0.800	0.857

Table 5.8: Cutflow table for the ggH category. At each step of the selection the number accepted events including statistical errors is shown for both simulations and data. In the bottom table the efficiencies of each cut wrt. the previous step in shown. The β_T and β_z are reconstructed using the `BoostXYZ`-method described in chapter 4. The cut efficiencies on in the first row are calculated wrt. the last of the preselection.

CHAPTER 6

Mass Reconstruction of Tau Pair Systems

As mentioned in chapter 4, reconstructing the mass of bosons decaying into τ-pairs is highly non-trivial due to the escaping neutrinos in the τ decays. In this chapter, the performance of the mass reconstruction technique presented in section 4.2 is compared to some of the existing techniques given the selection of τ-pairs presented in section 5.3. The performance of these methods all depends on the topologies of the events, and the performance is therefore specific to this selection.

6.1 Mass Reconstruction Techniques

In the following, a brief introduction to three of the most commonly used mass variables is given.

6.1.1 Visible Mass

The *visible mass* is simply the invariant mass of the sum of the two visible τ-jets, $\mathrm{M_{vis}} = \sqrt{(E_{\tau_1^{\mathrm{vis}}} + E_{\tau_2^{\mathrm{vis}}})^2 - (\vec{p}_{\tau_1^{\mathrm{vis}}} + \vec{p}_{\tau_2^{\mathrm{vis}}})^2}$, where $p_{\tau_i^{\mathrm{vis}}}$ denotes the four momenta of the i^{th} τ-jet.

6.1.2 Collinear Approximation Mass

The *Collinear Approximation*, CA, [31] is based on the assumption that the particles generated in the τ-decay are emitted in a very narrow cone. This is true in most cases due to the large boost of the τ-leptons. Using the four momenta of the two τ-jets and the projection of $E_{\mathrm{T}}^{\mathrm{miss}}$ vector along the τ-jet directions, it is possible to write down two equations which gives an estimation of the energy of the two τ-leptons. The boson mass can then be easily evaluated.

The projection of the $E_{\mathrm{T}}^{\mathrm{miss}}$ onto the two τ-jets can be written as

$$\begin{aligned} E_{\mathrm{x}}^{\mathrm{miss}} &= E_{\mathrm{x},\nu_1} + E_{\mathrm{x},\nu_2} = E_{\nu_1}\frac{p_{\mathrm{x},\tau_1^{\mathrm{vis}}}}{|\vec{p}_{\tau_1^{\mathrm{vis}}}|} + E_{\nu_2}\frac{p_{\mathrm{x},\tau_2^{\mathrm{vis}}}}{|\vec{p_{\tau_2^{\mathrm{vis}}}}|} \\ E_{\mathrm{y}}^{\mathrm{miss}} &= E_{\mathrm{y},\nu_1} + E_{\mathrm{y},\nu_2} = E_{\nu_1}\frac{p_{\mathrm{y},\tau_1^{\mathrm{vis}}}}{|\vec{p}_{\tau_1^{\mathrm{vis}}}|} + E_{\nu_2}\frac{p_{\mathrm{y},\tau_2^{\mathrm{vis}}}}{|\vec{p_{\tau_2^{\mathrm{vis}}}}|} \end{aligned}$$

These equations can be inverted into

$$\begin{aligned} E_{\nu,1} &= |\vec{p}_{\tau_1^{\mathrm{vis}}}|\frac{E_{\mathrm{x}}^{\mathrm{miss}}\, p_{\mathrm{y},\tau_2^{\mathrm{vis}}} - E_{\mathrm{y}}^{\mathrm{miss}}\, p_{\mathrm{x},\tau_2^{\mathrm{vis}}}}{p_{\mathrm{y},\tau_2^{\mathrm{vis}}}\, p_{\mathrm{x},\tau_1^{\mathrm{vis}}} - p_{\mathrm{x},\tau_2^{\mathrm{vis}}}\, p_{\mathrm{y},\tau_1^{\mathrm{vis}}}} \\ E_{\nu,2} &= |\vec{p}_{\tau_2^{\mathrm{vis}}}|\frac{E_{\mathrm{x}}^{\mathrm{miss}}\, p_{\mathrm{y},\tau_1^{\mathrm{vis}}} - E_{\mathrm{y}}^{\mathrm{miss}}\, p_{\mathrm{x},\tau_1^{\mathrm{vis}}}}{p_{\mathrm{y},\tau_1^{\mathrm{vis}}}\, p_{\mathrm{x},\tau_2^{\mathrm{vis}}} - p_{\mathrm{x},\tau_1^{\mathrm{vis}}}\, p_{\mathrm{y},\tau_2^{\mathrm{vis}}}} \end{aligned} \tag{6.1}$$

The two τ-leptons four momenta, p_{τ_i}, can then be expressed by summing up the τ-jet four momenta with the reconstructed neutrinos as follows

$$p_{\tau_i} = \left(E_{\nu_i} + E_{\tau_i^{\mathrm{vis}}}\ ,\ (E_{\nu_i} + |\vec{p}_{\tau_i^{\mathrm{vis}}}|)\,\frac{\vec{p}_{\tau_i^{\mathrm{vis}}}}{|\vec{p}_{\tau_i^{\mathrm{vis}}}|} \right)$$

From here the collinear approximation mass, $\mathrm{M_{CA}}$, can be constructed as the invariant mass of the two reconstructed τ-leptons.

The drawback of this method is that it requires certain criteria on the event topology to be fulfilled. First of all, the two τ-jets must not be completely back-to-back in the transverse plan, in order to keep the projection of the $E_{\mathrm{T}}^{\mathrm{miss}}$ onto the two τ-jets unique. This is typically controlled by putting a cut on the ϕ-angle between the two τ-jets, $\cos\Delta\phi(\tau_1^{\mathrm{vis}}, \tau_2^{\mathrm{vis}}) >$ cut value where a typical cut value used in the ATLAS [59] is -0.9. This is a rather harsh cut on both Z^0 and Higgs events.

Another limitation with this method is that sometimes renders unphysical solutions to the neutrino energies given in equation 6.1. Therefore, it is convenient to define the fraction, x, of τ-lepton's energy carried by the visible τ-jet

$$x = \frac{E_{\tau_i^{\mathrm{vis}}}}{E_{\tau_i}} = \frac{E_{\tau_i^{\mathrm{vis}}}}{E_{\tau_i^{\mathrm{vis}}} + E_{\nu_i}} \tag{6.2}$$

which has to be in the range $[0, 1]$ for both τ-jets in order to get a physical solution to the neutrino energies. This requirement is identical to requiring that in the transverse plan the ϕ of the $E_{\mathrm{T}}^{\mathrm{miss}}$ is in between the ϕ of the two τ-jets.

6.1.3 Missing Mass Calculator

Another method to for mass estimation is the *Missing Mass Calculator*, MMC [33]. This method does not have the same constrains on the phase space as CA. The MMC is a likelihood-based method, which takes the neutrino directions from probability distributions rather than assuming neutrinos are collinear with the visible decay products. In the τ-pair system, the following four equations relate the known and unknown kinematic

quantities

$$\begin{aligned} E_{\mathrm{x}}^{\mathrm{miss}} &= p_{\nu_1}\sin\theta_{\nu_1}\cos\phi_{\nu_1} + p_{\nu_2}\sin\theta_{\nu_2}\cos\phi_{\nu_2} \\ E_{\mathrm{y}}^{\mathrm{miss}} &= p_{\nu_1}\sin\theta_{\nu_1}\sin\phi_{\nu_1} + p_{\nu_2}\sin\theta_{\nu_2}\sin\phi_{\nu_2} \\ m_{\tau_1}^2 &= m_{\tau_1^{\mathrm{vis}}}^2 + 2\sqrt{p_{\tau_1^{\mathrm{vis}}}^2 + m_{\tau_1^{\mathrm{vis}}}^2}\cdot p_{\nu_1} - 2p_{\tau_1^{\mathrm{vis}}}p_{\nu_1}\cos\Delta\theta_{\nu\tau 1} \\ m_{\tau_2}^2 &= m_{\tau_2^{\mathrm{vis}}}^2 + 2\sqrt{p_{\tau_2^{\mathrm{vis}}}^2 + m_{\tau_2^{\mathrm{vis}}}^2}\cdot p_{\nu_2} - 2p_{\tau_2^{\mathrm{vis}}}p_{\nu_2}\cos\Delta\theta_{\nu\tau 2} \end{aligned} \tag{6.3}$$

where $\theta_{\nu\tau i}$ denotes the angle between the τ-jet and the neutrino. The mass of the τ-lepton is known ($m_\tau = 1.777$ GeV) and the mass of the τ-jets are set to either the π-meson mass ($m_\pi = 139$ MeV) or ρ-meson mass ($m_\rho = 770$ MeV) depending on the number of tracks associated with the τ-jet.

From these equations the most likely mass of the τ-pair system can be estimated based on simulations. These simulations can also accounts for reconstruction performance of the τ-jets and of the $E_{\mathrm{T}}^{\mathrm{miss}}$.

6.1.4 Boost Mass

In section 4.2 an alternative way to estimating the mass of τ-pair system are described. Two different versions of the mass estimator are presented based on whether the transverse component of the boost is neglected ($M_{\mathrm{Boost-Z}}$) or if it is estimated using the direction of the $E_{\mathrm{T}}^{\mathrm{miss}}$-vector($M_{\mathrm{Boost-XYZ}}$).

As mentioned in section 4.2, which of these two mass estimators is best will depend on the size of the transverse boost and on the resolution of the $E_{\mathrm{T}}^{\mathrm{miss}}$ direction. An easy way to combine the two mass estimator would be to choose a threshold based on the magnitude of $E_{\mathrm{T}}^{\mathrm{miss}}$-vector, and use the $M_{\mathrm{Boost-Z}}$ below this threshold and the $M_{\mathrm{Boost-XYZ}}$ above. In figure 6.1 the mixed M_{Boost} is shown on simulated $Z \to \tau\tau$ events at preselection level for different threshold values of $|E_{\mathrm{T}}^{\mathrm{miss}}|$.

Since no signification change in the mass distribution is observed in figure 6.1 and both the signal regions in the analysis target events with a high transverse boost of the Higgs, $M_{\mathrm{Boost-XYZ}}$ is chosen as the default mass estimator, and henceforth will be denoted M_{Boost}.

6.2 Performance comparison

In order to compare the performance of the different mass estimation techniques a *separation power*, S, is defined as

$$S = \frac{|\mu_{\mathrm{a}} - \mu_{\mathrm{b}}|}{\sqrt{\frac{1}{2}(\sigma_{\mathrm{a}}^2 + \sigma_{\mathrm{b}}^2)}} \tag{6.4}$$

where σ denotes the spread and μ the mean of the distributions, and a,b denotes the two samples under test. Here, the performance is evaluated at the preselection level and the two samples will be the total background vs. the Higgs samples. These distributions along with the derived S are shown in figure 6.2 for the M_{Boost} mass estimator, in figure 6.3 for the visible mass and in figure 6.4 for the likelihood-based $\mathrm{M_{MMC}}$ variable.

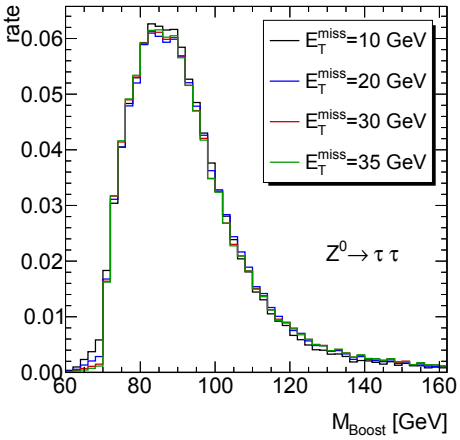
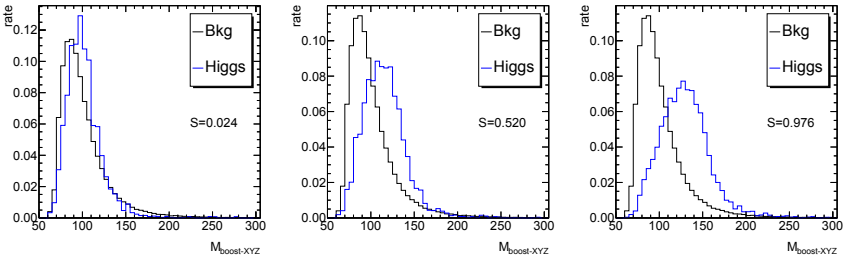


Figure 6.1: Combined $M_{\mathrm{Boost-Z}}$ and $M_{\mathrm{Boost-XYZ}}$ distribution for different threshold values of $E_{\mathrm{T}}^{\mathrm{miss}}$. The combined M_{Boost} variable exploits the $M_{\mathrm{Boost-XYZ}}$ when thee $E_{\mathrm{T}}^{\mathrm{miss}}$ value is above the threshold and otherwise uses the $M_{\mathrm{Boost-Z}}$variable.



(a) $m_H = 110$ GeV (b) $m_H = 130$ GeV (c) $m_H = 150$ GeV

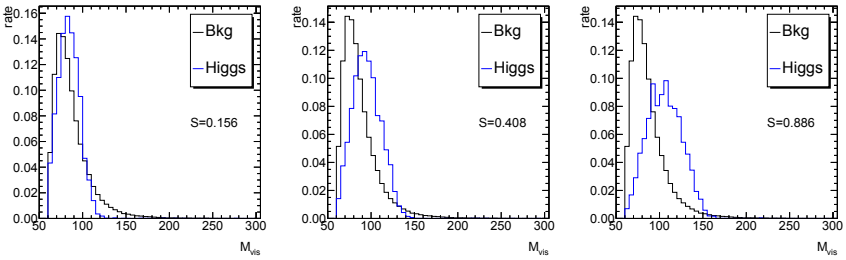
Figure 6.2: Normalised distributions of M_{Boost} at preselection for expected background and signal events. The separation power, S, is derived using equation 6.4.

It should be noted that the cut on $\mathrm{M_{vis}}$ and the p_{T} cuts on the τ-objects truncates the $\mathrm{M_{vis}}$ distribution and make the width unnatural small. Therefore the obtained values for the separation power will be biased, especially for the low Higgs masses. For two highest mass point, best separation is found using the $\mathrm{M_{MMC}}$ variable, while the $\mathrm{M_{vis}}$ variable has the worst separation.

6.3 Principal Component Analysis

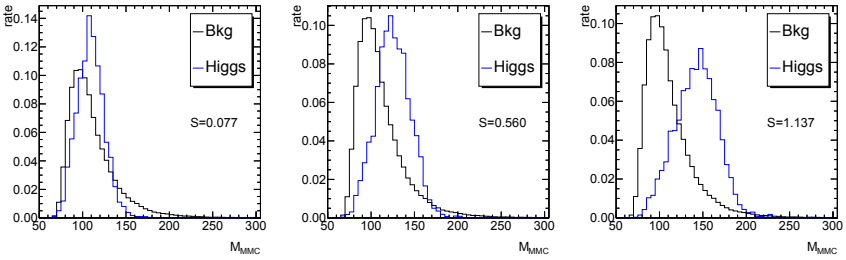
As shown in [29], the $\mathrm{M_{vis}}$ and M_{Boost} are fairly correlated with a correlation factor of 0.67[1], therefore using both variables should add information. However, it is possible to transform the correlated $\mathrm{M_{vis}}$ and M_{Boost} variables into a set of linearly uncorrelated variables. The Principal Component Analysis transforms input variables into linearly

[1]This correlation factor were derived at generator level using $Z \rightarrow \tau\tau$ events.



(a) $m_H = 110$ GeV (b) $m_H = 130$ GeV (c) $m_H = 150$ GeV

Figure 6.3: Normalised distributions of M_{vis} at preselection for expected background and signal events. The separation power, S, is derived using equation 6.4.



(a) $m_H = 110$ GeV (b) $m_H = 130$ GeV (c) $m_H = 150$ GeV

Figure 6.4: Normalised distributions of M_{MMC} at preselection for expected background and signal events. The separation power, S, is derived using equation 6.4.

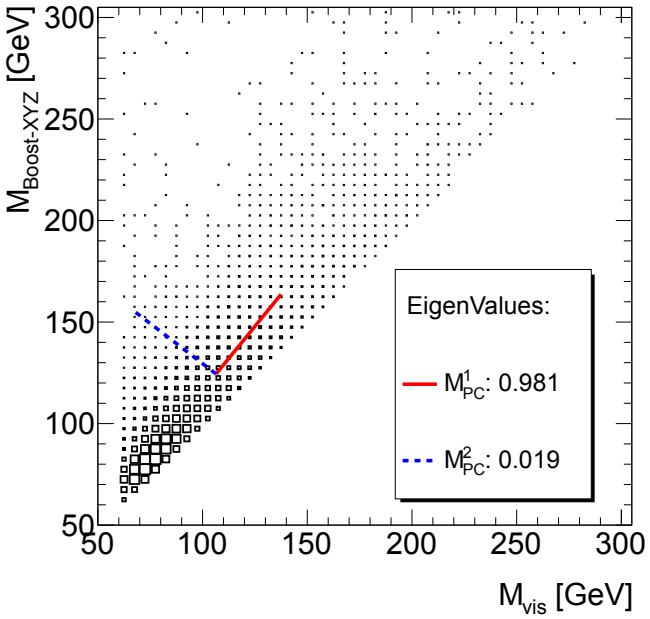


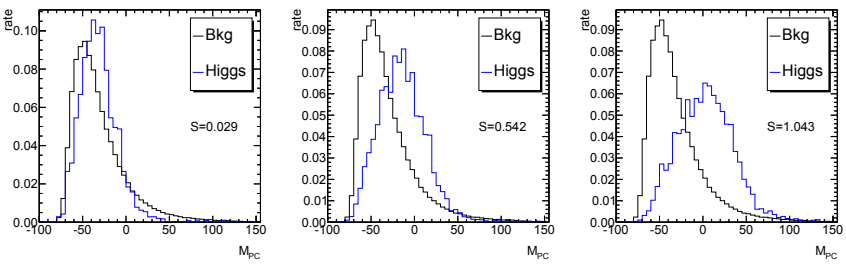
Figure 6.5: Correlation between M_{vis} and M_{Boost} at preselection for data, along with the principal components found by the Principal Component Analysis and their eigenvalues.

uncorrelated variables, referred to as principal components, by an orthogonal transformation. The new principal components are ranked by the variance of data in this variable. Hence a ranking of the expected importance of each principal component variable is achieved. Figure 6.5 shows the correlation between M_{vis} and M_{Boost} at preselection for the observed data, as well as the found principal components and their eigenvalues. Since the eigenvalue of the first component is much larger than the second eigenvalue, most for the information is store in this variable and this component will hereafter be referred to as M_{PC}. In figure 6.6, the separation powers are calculated for the M_{PC} variable as well. Only slightly better separation powers are found using the M_{PC} over the M_{Boost} variable.

The separation powers are summarised in table 6.1, although these numbers should not be used as proof of which variable is best in a search or exclusion scenario, they do contain some hint of the ranking of these variables. Therefore the M_{MMC} is used as the default mass variable when setting a limit on the Higgs cross section in section 9.3.

Variable	$m_H = 110$	$m_H = 130$	$m_H = 150$
M_{vis}	0.156	0.408	0.886
M_{Boost}	0.024	0.520	0.976
M_{PC}	0.029	0.540	1.043
M_{MMC}	0.077	0.560	1.137

Table 6.1: Separation power for different mass variables.



(a) $m_H = 110$ GeV (b) $m_H = 130$ GeV (c) $m_H = 150$ GeV

Figure 6.6: Normalised distributions of M_{PC} at preselection for expected background and signal events. The separation power, S, is derived using equation 6.4.

CHAPTER 7

Background Estimation

In this chapter, the modelling of the two main backgrounds, i.e. $Z \to \tau\tau$ and QCD jets, are described. In order to model the differential distributions of these processes, data-driven methods are used, as described in section 7.1 and section 7.2, while the normalisation of both samples is done using the track multiplicity fit described in section 7.3.

7.1 Embedded $Z \to \tau\tau$ events

Since $Z \to \tau\tau$ is a dominant background process for this analysis, getting the best possible prediction on its distribution is very important. The idea behind the embedding technique is to use Z^0 events from data to model the Z^0 background. By doing so one could get a low statistical uncertainty on the distributions (due to the large cross section of Z^0 events) and avoid systematics uncertainties related to the underlying events and jet kinematics in these events.

However, getting a clean and unbiased $Z^0 \to \tau_{\text{had}}\tau_{\text{had}}$ samples from data is not possible due to the large QCD background and possible contamination from Higgs events. Therefore, $Z \to \mu\mu$ events are used instead, where the muons from data are substituted with τ-decays from Monte Carlo simulations. Due to the clean signatures and high energy resolution of muons in the detector, it is possible to select a almost un-contaminated $Z \to \mu\mu$ sample from data with an extremely good prediction on the Z^0-mass shape from data. By removing the muons from the events and replacing them with simulated τ-decays, hybrid events can be constructed. Since τ-decays are well understood and well modelled in simulations this procedure should be safe. However, global quantities, i.e. $E_{\text{T}}^{\text{miss}}$, have to be recomputed after inserting the τ-decays.

In the following, the construction of the hybrid or *embedded* events is described

- **Selection of $Z \to \mu\mu$ events:**
  $Z \to \mu\mu$ are selected from data by requiring exactly two well reconstructed and isolated muons with $p_{\text{T}} > 20$ GeV. The muons are required to have an invariant mass $\text{m}_{\mu\mu} > 55$ GeV and to come from the same primary vertex.

- **Extracting $Z \to \tau\tau$ information:**
  The Z^0 boson is reconstructed from the selected muons. In the rest frame of the Z^0 boson, the reconstructed muons are replace with τ-leptons at generator level.

Here the four-momenta of the τ-leptons are scaled in order to account for the mass difference between the muon and τ-lepton.

- **Simulation of τ-decays**
  The generator level τ-leptons and their decay are generated with TAUOLA and PHOTOS, in order to correctly simulate the polarisation and final state radiation of the τ-leptons. The generator level event (containing only the two τ-decays and their final state particles) is then processed by the full ATLAS detector simulation.

- **Merging data and simulation:**
  First, all tracks and calorimeter energy associated with the muons are subtracted from the events. The energy subtraction in the calorimeter cells is based on simulated $Z \to \mu\mu$ events where the kinematics have be fixed to match the $Z \to \mu\mu$ kinematics in this event. Subsequently, the tracks and energy deposited in the calorimeters from the $Z \to \tau\tau$ events are added to the original event.

- **Re-reconstruction of the hybrid event:**
  The resulting $Z \to \tau\tau$ hybrid event is then re-processed by the full event reconstruction to ensure that all global objects,e.g. $E_{\mathrm{T}}^{\mathrm{miss}}$, are correctly recomputed.

The *embedded* events have been throughly tested in ATLAS and have been shown to describe data well. However, no trigger information is available in the *embedded* samples. Therefore, additional weights provided by the ALTAS Tau Working Group are assigned to the events to incorporated the effects of the trigger as function of the η and p_{T} of the τ-jets. The overall normalisation of the *embedded events* are done with the track fit described in section 7.3.

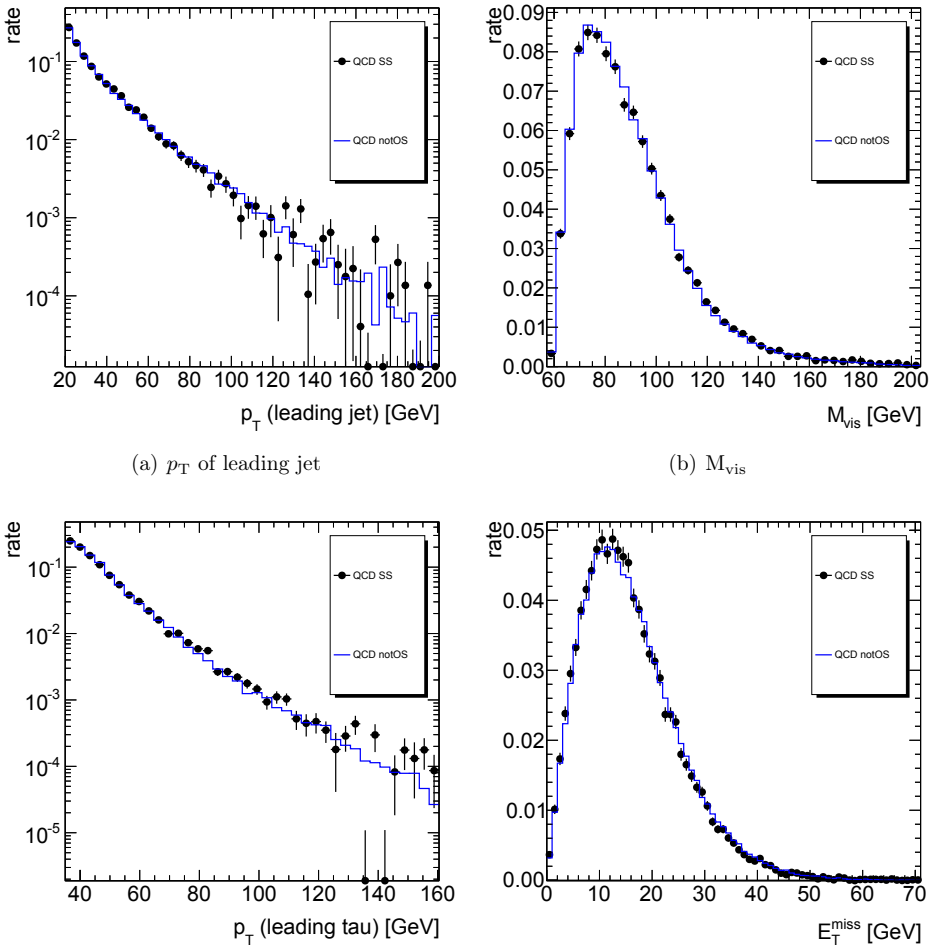
7.2 QCD modelling

Unfortunately, Monte Carlo simulations does not describe events only containing QCD-jets (QCD events) very well. Therefore, a *data-driven* estimation of the QCD events is used. This method consists of taking pairs of τ-candidates with same reconstructed charge from data, referred to as same-sign events. A data-driven method of estimating QCD events, can therefore be achieved by taking same-sign events in data, and subtracting the small contamination of other background processes. However, the rate of same-sign and opposite-sign QCD events, is not guaranteed to be equal. Therefore a normalisation of the same-sign events is needed. This is done by the track-multiplicity fit described in section 7.3.

Although, the same-sign QCD events is assumed to describe well in most of the phase space, the statistical uncertainties on the predicted shapes become large when going to the tight signal regions. Therefore, the method is extended to use all events which does not satisfy the charge production $= -1$ requirement of the selection, i.e. the *not-opposite-sign* events. These events will have a higher contamination from other background processes, however these contaminations can still be estimated from simulations.

Hence, the QCD events are estimated from *not-opposite-sign* events from data with the *not-opposite-sign* events from the Monte Carlo samples subtracted. Figure 7.1 shows a comparison at preselection level of the shape of QCD events obtained from *same-sign*

and *not-opposite-sign* events. Using the *not-opposite-sign* events results increases the statistics for the QCD description by a factor of ~ 12.



(a) p_{T} of leading jet

(b) $\mathrm{M}_{\mathrm{vis}}$

(c) p_{T} of leading τ

(d) $E_{\mathrm{T}}^{\mathrm{miss}}$

Figure 7.1: Comparison of QCD models using same-sign, SS, or not-opposite-sign, notOS, events. The comparison is done at the preselection stage.

7.3 Normalisation from Track Multiplicity Fit

A τ-lepton decaying into hadrons typically creates a more collimated jet in the detector than jets originating from quarks or gluons, QCD-jets, and typically τ-jets generate less tracks than QCD-jets, which is why these two features are central in the τ-identification [52]. In most cases, a hadronic τ-decay will generate one or three charged pions contained in a narrow cone around the direction of the τ-lepton. Therefore, the

τ-reconstruction searches for tracks in a cone of $\Delta R \leq 0.2$ which can be associated to the τ-candidate. This cone size has been proven optimal for the τ-reconstruction. However, counting the number of tracks outside this cone, can help separating real τ-lepton from QCD-jets as well as give a way to estimate the fraction of selected real and fake τ-jet.

7.3.1 Track counting method

In addition to the standard τ-identification cuts further rejection against QCD-jets can be achieved by counting the tracks close to, but outside the cone used by the τ-reconstruction. Tracks with a distance in the range of $0.2 < \Delta R \leq 0.6$ from the centre of the τ-jet is counted in an anti-k_t -like way, first used in [59], in order to be as independent of the underlying event and pileup-conditions as possible.

A real τ-jet will typically only have 1 or 3 associated tracks in the cone of $\Delta R \leq 0.2$, *core* tracks, and tracks in the outer cone $\Delta R \leq 0.6$, *outer* tracks, will most likely come from the underlying event or from pileup, and therefore be uncorrelated. In contrast, a QCD-jet faking a τ-jet will be likely to produce additional *outer* tracks which are correlated with the existing *core* tracks by the splitting kernels [60, 61].

Tracks are only selected as *outer* tracks if they satisfy certain quality criteria[1] and have $p_T > 500$ MeV. In addition, outer tracks are required to have $D_{max} < 4$ where

$$D_{\max} = \max \sum_{\text{core}} \left(\frac{p_T^{\text{core}} \Delta R(\text{core}, \text{outer})}{p_T^{\text{outer}}} \right) \tag{7.1}$$

The D_{max} distance measure mimics the distance measure used in the anti-k_t -clustering algorithm [53] and the requirement $D_{max} < 4$ has been found as that best threshold to filter out uncorrelated tracks [62, 63].

Figure 7.2 shows the core+outer track distribution for same-sign, SS, events in data and for real τ-jets obtained from $Z \to \tau\tau$ simulations at the preselection level described in section 5.3, but before making the $\Delta\eta(\tau_1, \tau_2)$ cut and without imposing the track multiplicity cut on the τ-candidates. In the following this region will be referred to as the *normalisation stage*. Both distributions are peaking at 1 and 3 due to the requirement that reconstructed electrical charge should be ± 1 for both candidates. However, a clearly shift towards higher tracks multiplicities is seen for QCD-jet events compared to real τ-candidates.

7.3.2 2-dimensional Fit of Track Multiplicities

In addition to rejecting QCD events, the track multiplicity distributions can also be used to estimate the fraction of real and fake τ-candidates in the selected ensemble. The fractions are found by creating 2-dimensional histograms of the track multiplicity of leading and sub-leading τ-candidates, before cutting on the number of tracks. These templates are generated for the following type of events: two real τ-candidates,T_{real}, two fake τ-candidates, T_{fake}, and one fake and one real τ-candidate, T_{mixed}.

The T_{real} and T_{mixed} are generated from simulations of $Z \to \tau\tau$ events and $W \to \tau\nu$ and top events, while the T_{fake} template is mainly taken from *same-sign* events in data.

[1] The standard track quality criteria consist of a minimum number of hits in each sub-detector in the ID and cuts on impact parameters

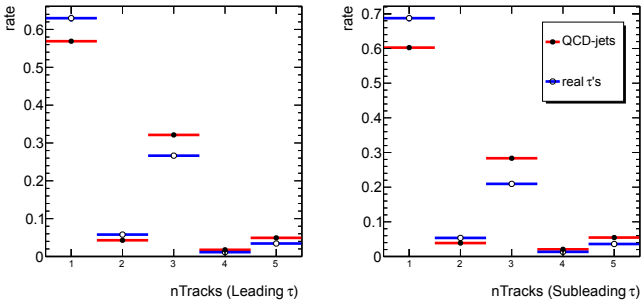


Figure 7.2: Number of associated tracks –both core and outer – at normalisation stage for QCD and real taus. QCD event taken from from same-sign events in data, while real τ-pairs are taken from $Z \to \tau\tau$ simulations.

Since, the $Z \to \tau\tau$ contamination of the T_{fake} template due to the probability of charge misreconstruction is low, this effect is taken into account by subtracting these events from both data and the T_{fake} template.

Each template is generated at the same stage of the selection, i.e. at the normalisation stage, to ensure the same kinematics in each templates. At this stage of the selection, the number of real and fake τ-pairs can then be estimated, by fitting the ratio of the three templates to the observed data distributions. Since number of events from the third template is expected to be very small, the normalisation of this template is kept constant. In practice, the T_{mixed} template is subtracted from data before fitting the two other templates to the remaining data.

The fraction of QCD events, f_{QCD} is found using a binned maximum likelihood fit, with the log-likelihood function, $\log \mathcal{L}$, defined as

$$\log \mathcal{L} \quad = \quad \sum_i \log \text{ Poisson}(n_i|\mu_i) \tag{7.2}$$

where for each bin i in the 2-dimensional track multiplicity space, n_i denotes the number of observed events and μ_i mean value given by the normalised T_{real} and T_{fake} using

$$\mu_i = f_{\text{QCD}}\, T_{\text{fake}}(i) + (1 - f_{\text{QCD}})\, T_{\text{real}}(i) \tag{7.3}$$

Figure 7.3 shows the log-likelihood ratio, q, at the normalisation stage, defined as

$$q = -\log \frac{\mathcal{L}(f_{\text{QCD}})}{\mathcal{L}(\hat{f}_{\text{QCD}})} \tag{7.4}$$

where $\hat{f}_{\text{QCD}}$ is the value of f_{QCD} that maximises $\mathcal{L}(f_{\text{QCD}})$. The q normalises $\log \mathcal{L}$ to its maximal value. The q shown in figure 7.3 clearly has a single minimum, and is well defined over a large range around it.

Figure 7.4 shows the fitted combination of two templates, T_{real} and T_{fake}, to the observed data at the normalisation stage. Since the fit only evaluates the fraction of T_{fake} vs. T_{real} events, the overall normalisation of events are fixed by the number of events in data. The fit can easily be extended to also fitting the overall normalisation by

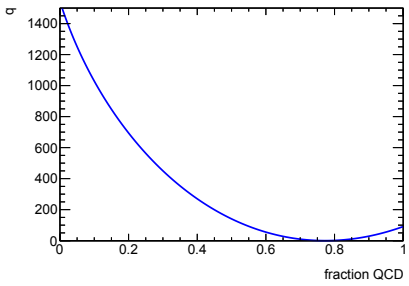


Figure 7.3: Log Likelihood Ratio, q, for the track multiplicity likelihood function in equation 7.2.

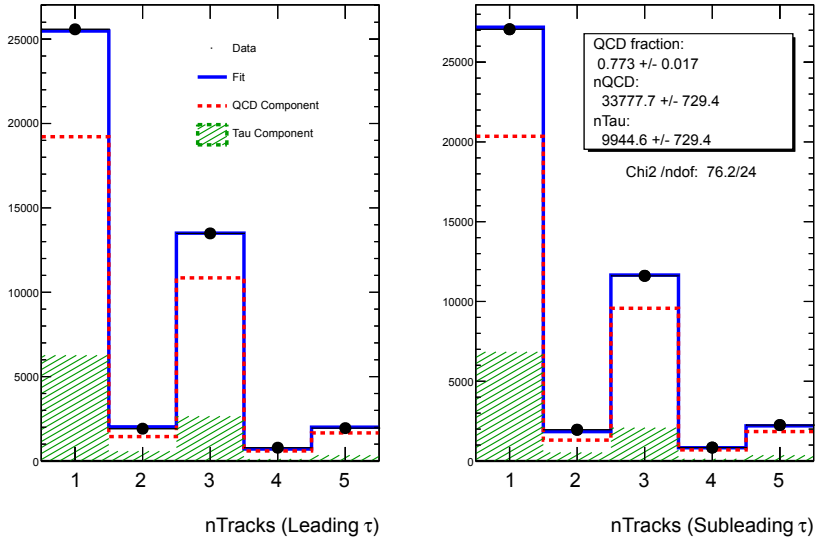


Figure 7.4: Projections on track-multiplicity of leading and subleading τ-candidates. Full line show the fitted combination of the real and fake τ-pair templates, while the dashed lines shows the individual templates used in the combination. The points are the measured data points.

substituting the f_{QCD} parameter with 2 parameters N_{fake} and N_{real} denoting the total number of QCD or real τ-pair events. In this case the extended likelihood function is

$$\begin{aligned} \log \mathcal{L} &= \sum_i \log \text{ Poisson}(n_i|\mu_i) \\ &\quad - \log \text{ Poisson}(N_{\text{fake}} + N_{\text{real}}, N_{\text{observed}}) \end{aligned} \tag{7.5}$$

where

$$\mu_i = \frac{N_{\text{fake}}}{N_{\text{fake}} + N_{\text{real}}} T_{\text{fake}}(i) + \frac{N_{\text{real}}}{N_{\text{fake}} + N_{\text{real}}} T_{\text{real}}(i)$$

Fitting the 2-parameter likelihood given in equation 7.5, has been shown to give a consistent result with the one parameter fit in figure 7.4. And therefore the 1-parameter fit will be used here.

The ratio between QCD events and real τ-pairs could in principle be determined at any stage of the selection. However, the number of events containing two real τ-leptons might not be purely from Z^0, but can contain possible signal events. This would introduce a small bias of the on the $Z \to \tau\tau$ normalisation and possible hide a signal. Therefore the $Z \to \tau\tau$ samples has to be normalised at an early stage of the selection, i.e. the normalisation stage, where the bias is very small[2] and therefore can be neglected. The normalisation of $Z \to \tau\tau$ events can hereafter be extrapolated from the normalisation stage to the signal regions.

7.4 Normalisation from $\Delta\eta(\tau_1, \tau_2)$ Fit

An alternative way of determining the amount of QCD and Z^0 events is to use the difference in the $\Delta\eta(\tau_1, \tau_2)$ distributions. The $\Delta\eta(\tau_1, \tau_2)$ distribution for τ-pairs coming from $Z \to \tau\tau$ event is much more centred around 0 and has an upper limit at 2, while distribution for QCD-jets is much broader, as illustrated in Figure 5.1. This difference is already exploited in the selection of good τ-pairs, however this difference also provides good way for estimating the ratio between QCD and $Z \to \tau\tau$ events.

By using the template distributions for QCD and $Z \to \tau\tau$ shown in figure 5.1, a maximum likelihood fit of the two distributions are performed to match the observed data at the *normalisation stage*. Just like for the track-multiplicity fit, the normalisation stage is defined as the preselection with the last cut on $\Delta\eta(\tau_1, \tau_2) < 2$. Before performing the fit, the other background processes are subtracted from the data points using the predicted distributions from simulations, so that the data only contains the two backgrounds of interest (and possibly a small contamination of signal events).

Figure 7.5 shows the observed and the fitted $\Delta\eta(\tau_1, \tau_2)$ distributions at normalisation stage, along with the QCD and $Z \to \tau\tau$ components of the fit. The obtained number of QCD events from fitting the $\Delta\eta(\tau_1, \tau_2)$ distributions is ~ 100 events more than from track multiplicity fit which is within the error of the fit. Hence, the two fit methods are consistent, but since the $\Delta\eta(\tau_1, \tau_2)$ fit give the best χ^2, the normalisations further on are taken from this fit.

[2]The signal contamination wrt. the number of $Z \to \tau\tau$ at the normalisation stage is $\approx 0.38\%$ (for $m_H = 125$ GeV

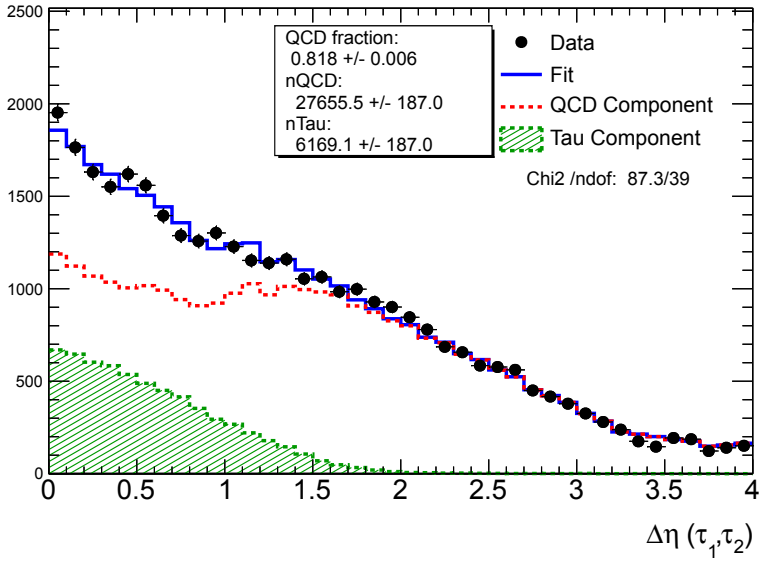


Figure 7.5: Fit of $\Delta\eta(\tau_1, \tau_2)$ distribution to the observed data at the normalisation stage.

CHAPTER 8

Theoretical Uncertainties

The theoretical uncertainties (THU) related the Higgs production are often among the largest uncertainties in the Higgs searches at the LHC. The ability to claim either discovery or exclusion of the Higgs, is very dependent on getting the best prediction from theory on the expected signal yield in the analysis.

Three types of THU usually enter in the analysis in order to get the expected signal yield.

- The uncertainty on the inclusive cross section for a given production mode.
- The uncertainty on the branching ratios, BR, into different final states.
- The uncertainty on the differential cross section or on the kinematics in the Higgs event.

In order to get a consistent treatment of these uncertainties between ATLAS and CMS, the THU are evaluated by the LHC Higgs Cross Section Working Group, XSWG [38, 39] without taking experimental cuts or acceptance of the detector into account. This ensures a consistent use of THU across the different final states searches and across experiments, which is important when comparing the different results and crucial when trying to combine them.

The procedure of using values given by the Higgs XSWG, guarantees that the analyses use not only common, but also the most up-to-date numbers from theory both on the central values and on the associated uncertainties.

Since the THU's listed above are derived without considering analysis cuts or detector acceptance, these still have to be evaluated by the individual analyses. Some recommendations on how to do this consistently is given in Handbook of LHC Higgs Cross Sections [39]. In the latter part of this chapter, a description of how these uncertainties are derived for the $H \to \tau\tau \to \tau_{\text{had}}\tau_{\text{had}}$ analysis is given.

8.1 Inclusive Cross Section Uncertainties

As described in section 2.1 the main Higgs production mechanism at the LHC is ggH where the Higgs boson is produced by two gluons through a heavy quark loop. Hence

this production mechanism is mainly controlled by the strong interaction and therefore one would expect sizeable QCD corrections.

The leading order, LO, cross section for ggH [64] is proportional to α_S^2 which can be seen from the diagram in figure 2.3(a). The next-to-leading order, NLO, QCD corrections [65, 66, 67] enhances the cross section with $\sim 80 - 100\%$. These corrections are computed both with and without using the large-m_t limit ($m_t \to \infty$), while the next-to-next-to-leading-order, NNLO [68, 69, 70], corrections are only computed in the large-m_t limit. These corrections further increase the cross section with $\sim 25\%$. The NNLO cross section is improved by resumming the soft-gluon contributions [71] up to next-to-next-leading-log, NNLL which leads to an additional increase in the cross section of $7 - 9\%$.

In addition to the QCD corrections, also EW corrections has to be considered. Since the ggH is initiated by strong dynamics, the EW corrections are much smaller. Currently, two-loops EW corrections are computed [72, 73], and leads to a $\sim 5\%$ increase in the cross section for a low mass Higgs.

The uncertainties on the inclusive cross section comes from the following sources

- Higher-order radiative corrections from both QCD and EW corrections
- Limited knowledge of the parton distribution functions, PDFs, and uncertainty on α_S
- Effects of the large-m_t approximation
- Uncertainties on the input quark masses, in particular the m_t and m_b.
- Choice of the renormalisation scale, μ_R, and factorisation scale, μ_F

Of these uncertainties the two dominating sources are the uncertainties on the PDF+α_S[1] and uncertainties from QCD scales (μ_R,μ_F) which are comparable in size ($\sim 10\%$) for most of the m_H range.

For the VBF and VH production modes also described in section 2.1, the QCD corrections are typically small since this production modes is controlled by weak interactions. The cross section have been calculated with both NLO QCD and EW corrections [9, 10]. These corrections also takes into account real photon emissions.

The main sources still comes from QCD scales and PDF+α_S, however since the kinematics of this production mode is controlled by weak interaction, a large part of the uncertainties on the QCD scales are shifted to the PDF uncertainties.

The main THUs on the inclusive Higgs cross section taken from [38] are summarised in table 8.1.

8.2 Uncertainties on Jet Categories

Since the background composition often varies with the number of associated jets, the Higgs searches are often split up into categories based on the jet kinematics of the event in order to enhance sensitivity. Also these categories can be used to target ggH and

[1] Since the uncertainties on α_S and the PDF's are highly correlated it is customary to list these as a combined uncertainty.

Source	ggH [%]	VBF [%]	VH [%]
PDF+α_S	$^{+7.9}_{-7.7}$	$^{+2.7}_{-2.1}$	$^{+3.9}_{-3.9}$
QCD Scale	$^{+13.3}_{-8.6}$	$^{+0.4}_{-0.4}$	$^{+0.7}_{-0.8}$

Table 8.1: Theoretical uncertainties on the inclusive Higgs cross section due to PDF+α_S and QCD scale taken from [38].

VBF events separately. When splitting up the events into different jet categories, it becomes important to estimate the THU on the signal yield in each category. Since the uncertainty from PDF+α_S does not modify the jet kinematics of the events, only the THU from QCD scales are considered important here.

Even for analyses which only have a single signal region, the uncertainty on the signal yield in this signal region still needs to be evaluated – especially if the jet kinematics is used to define the signal region. This was the case in the $H \to \tau\tau \to \tau_{\text{had}}\tau_{\text{had}}$ analysis for Moriond 2012 [51], where HNNLO [74, 75] was used to estimate the uncertainty on the Higgs+ $\geq$ 1jet cross section.

For analyses with multiple exclusive jet categories, a natural approach for estimating the uncertainties on the signal yield in each category would be to compute the *exclusive* cross section for the each jet category while varying the QCD scales, and treat the uncertainty for each category as uncorrelated. However, this approach has been shown to severely underestimate the uncertainty [39].

Instead, an improved prescription for estimating the uncertainties is given by Stewart and Tackman [76]. In the ATLAS $H \to \tau_{\text{had}}\tau_{\text{had}}$ working group, the following three categories are proposed, *VBF*, *Boosted* and *Rest* with the cuts used to define the categories listed in table 8.2. These categories do not entirely match the categories defined in section 5.4, but the VBF and Boosted category presented here have very similar definitions as the VBF and ggH categories, and therefore the THU are assumed to be the same.

VBF	Boosted	Rest
≥ 2 jets with:	Not in the VBF	Not in the VBF
jet$_1$ $p_T \geq 50$ GeV	≥ 1 jet with:	Not in Boosted
jet$_2$ $p_T \geq 30$ GeV	Jet$_1$ $p_T \geq 50$ GeV	
$\eta_{\text{jet1}} \cdot \eta_{\text{jet2}} < 0$		
$\Delta\eta_{\text{jet}_1,\text{jet}_2} > 2.6$		
$M_{jj} < 350$ GeV		
Only jets within $\lvert\eta\rvert < 4.5$ are considered		

Table 8.2: Proposed cuts to the define the jet categories in the $H \to \tau_{\text{had}}\tau_{\text{had}}$ analysis.

For ggH production, the *inclusive* cross sections $\sigma_{\geq 2\text{jets}}$, $\sigma_{\geq 1\text{jets}}$ and σ_{Total} have been computed using the Parton level Monte Carlo program, HNNLO, which is able to calculate the Higgs production at NNLO in QCD using a set of user-specific cuts on the associated jets. The $\sigma_{\geq 2\text{jets}}$ cross section is computed using the cuts for the VBF category, while the $\sigma_{\geq 1\text{jets}}$ cuts are defined as either the VBF or Boosted category. The uncertainties on the $\sigma_{\geq 2\text{jets}}$, $\sigma_{\geq 1\text{jets}}$ and σ_{Total} are found by computing the cross sections

while varying the QCD scales, μ_F and μ_R, around the nominal value at $\mu_R = \mu_F = m_H$. The QCD scales are varied among the following values[2]

$$\mu_R, \mu_F \in \{{}^{m_H}\!/_2, m_H, 2mH\} \tag{8.1}$$

For any combination of μ_R and μ_F, the HNNLO program is run for a sufficiently long time to make the statistical errors on the cross sections negligible[3] ($< 0.1\%$). The uncertainties, $\Delta\sigma_{\geq 2}$, $\Delta\sigma_{\geq 1}$ and $\Delta\sigma_{\text{Total}}$ on the inclusive cross sections are taken as the maximal variation from the nominal value.

From the inclusive cross sections, the *exclusive* cross section in each category is found by assuming

$$\sigma_{\text{VBF}} = \sigma_{\geq 2\text{jets}}, \quad \sigma_{\text{Boosted}} = \sigma_{\geq 1\text{jets}} - \sigma_{\geq 2\text{jets}}, \quad \sigma_{\text{Rest}} = \sigma_{\text{Total}} - \sigma_{\geq 1\text{jets}} \tag{8.2}$$

while the signal yield fractions in each category are defined as

$$f^{\text{VBF}} = \frac{\sigma_{\text{VBF}}}{\sigma_{\text{Total}}}, \quad f^{\text{Boosted}} = \frac{\sigma_{\text{Boosted}}}{\sigma_{\text{Total}}}, \quad f^{\text{Rest}} = \frac{\sigma_{\text{Rest}}}{\sigma_{\text{Total}}} \tag{8.3}$$

				HNNLO			MCFM
m_H [GeV]	f^{Rest}	f^{Boosted}	f^{VBF}	$\delta\sigma_{\text{Total}}$	$\delta\sigma_{\geq 1}$	$\delta\sigma_{\geq 2}$	$\delta\sigma_{\geq 2}$
100	0.615	0.377	0.008	23.2 %	22.5 %	75.1 %	25.7 %
110	0.577	0.414	0.009	22.6 %	22.4 %	74.0 %	27.0 %
120	0.545	0.446	0.009	24.3 %	22.2 %	73.0 %	27.3 %
130	0.514	0.476	0.010	22.3 %	22.5 %	72.2 %	27.7 %
140	0.485	0.505	0.010	22.6 %	22.5 %	71.5 %	27.2 %
150	0.461	0.528	0.011	23.1 %	23.0 %	70.8 %	27.4 %

Table 8.3: Fraction of signal yield in each jet category, f, for ggH Higgs production, along with the relative uncertainties on the cross sections in the inclusive jet bins, $\delta\sigma_{\text{Total}}$, $\delta\sigma_{\geq 1}$ and $\Delta\sigma_{\geq 2}$. Calculated using HNNLO and MCFM. The large uncertainties $\delta\sigma_{\geq 2}$ found by HNNLO are replaced by the uncertainties found by MCFM.

Table 8.3 shows the signal yield in each category and the associated relative uncertainties, $\delta\sigma_i = {}^{\Delta\sigma_i}\!/_{\sigma_i}$ on the *inclusive* cross sections obtained by HNNLO in the mass range of m_H from 100 GeV to 150 GeV. Since HNNLO calculates ggH Higgs at NNLO, this means that the cross sections σ_{Total}, $\sigma_{\geq 1\text{jets}}$ and $\sigma_{\geq 2\text{jets}}$ are all calculated to same order in α_S. However, the prediction for $H+ \geq 2$ jets is a LO prediction, and hence this cross section and its uncertainty are very sensitive to virtual corrections. Therefore, the uncertainty for the $\sigma_{\geq 2\text{jets}}$ is also computed using Monte Carlo event generator MCFM [77] which is able to calculate $H+ \geq 2$ jets at NLO. The results are also listed in table 8.3. As expected, the uncertainty on the $\sigma_{\geq 2\text{jets}}$ drops significantly when including the higher order corrections, and the high uncertainties found by HNNLO for $\sigma_{\geq 2\text{jets}}$ are replaced by the uncertainties found by MCFM.

Using the Stewart-Tackman (S.T.) approach, the uncertainties on the *inclusive* cross sections should be treated as correlated, and the uncertainties on the *exclusive* cross

[2]Both the central values and the variation range of the QCD scales μ_R and μ_F are purely chosen by convention.

[3]The cpu time for each cross section computation depends on the specific cuts, but a typical computation takes ~ 10 hours.

sections can be found from error propagation using the covariance matrix, C, for the cross sections $\{\sigma_{\text{Total}}, \sigma_{\text{Rest}}, \sigma_{\text{Boosted}}, \sigma_{\text{VBF}}\}$[76]

$$\begin{bmatrix} \Delta\sigma^2_{\text{Total}} & \Delta\sigma^2_{\text{Total}} & 0 & 0 \\ \Delta\sigma^2_{\text{Total}} & \Delta\sigma^2_{\text{Total}} + \Delta\sigma^2_{\geq 1} & -\Delta\sigma^2_{\geq 1} & 0 \\ 0 & -\Delta\sigma^2_{\geq 1} & \Delta\sigma^2_{\geq 1} + \Delta\sigma^2_{\geq 2} & -\Delta\sigma^2_{\geq 2} \\ 0 & 0 & -\Delta\sigma^2_{\geq 2} & \Delta\sigma^2_{\geq 2} \end{bmatrix} \tag{8.4}$$

When using THU in the likelihood fit described in section 9.1, the uncertainties are introduced as nuisance parameters, which are usually assumed to be gaussian distributed. However, due to the large values of the QCD scales uncertainties, these uncertainties are assumed to be distributed according to a log-normal distribution of width κ. In practice, the uncertainties on the exclusive cross section are not used as nuisance parameters, because of there correlation. Instead, one or two nuisance parameters for the inclusive cross sections are given for each category. These nuisance parameters are derived as follows [78]

Name	VBF	Boosted	Rest
QCDscale_ggH	$\kappa_{\text{tot}}^{1/f_0}$	-	-
QCDscale_ggH1in	$\kappa_{\geq 1}^{-f_1 + f_2/f_0}$	$\kappa_{\geq 1}^{-f_1 + f_2/f_1}$	-
QCDscale_ggH2in	-	$\kappa_{\geq 2}^{-f_2/f_1}$	$\kappa_{\geq 2}$

where the f's are the fraction listed in table 8.3 and $\kappa_i = 1 + \delta\sigma_i$. The value of κ_{tot} is taken from the Higgs XSWG [38].

For the VBF production mode, the cross sections are computed using the Monte Carlo generator HAWK [79]. which is able to compute VBF production at NLO. Since the THU from QCD scales are much smaller in VBF than for ggH, the uncertainties on the exclusive cross section are not derived using the S.T. procedure, following the recommendations from the Higgs XSWG [39]. Instead, the exclusive cross sections are computed for different values of μ_R and μ_F, and the uncertainties are taken as the maximal variation from the nominal value. Here, the nominal value is chosen for $\mu_R = \mu_F = m_W$, and the QCD scales are varied independently between the following values $\{1/2 m_W, m_W, 2m_W\}$. The resulting relative uncertainties from HAWK are listed in table 8.4.

m_H [GeV]	f^{Rest}	f^{Boosted}	f^{VBF}	$\delta\sigma_{\text{Rest}}$	$\delta\sigma_{\text{Boosted}}$	$\delta\sigma_{\text{VBF}}$
110	0.840	0.086	0.073	1.6%	3.2 %	1.7 %
120	0.822	0.091	0.086	5.6 %	5.8 %	2.8 %
130	0.810	0.095	0.095	5.9 %	3.0 %	1.1 %
140	0.813	0.089	0.097	5.7 %	2.9 %	1.8 %
150	0.803	0.089	0.107	6.4 %	4.2 %	1.7 %

Table 8.4: Fraction of signal yield in the different jet categories, f, for VBF Higgs production and the associated relative uncertainties, $\delta\sigma_{\text{Rest}}$, $\delta\sigma_{\text{Boosted}}$ and $\delta\sigma_{\text{VBF}}$ calculated using HAWK.

8.3 Background cross sections

The production cross section uncertainties for the Z and W background as well as diboson backgrounds are taken from [80] and listed in table 8.5.

For the cross-section uncertainty on top pair production, calculations have been used that approximates NNLO in QCD with Hathor 1.2 [81] using the MSTW2008 90% NNLO PDF sets [82] incorporating PDF+α_S uncertainties according to the MSTW prescription [83].

Background	Total uncertainty [%]
Z^0+jets	5.0
W+jets	5.0
WW	$^{+5.5}_{-4.5}$
WZ^0	$^{+7.1}_{-5.9}$
Z^0Z^0	$^{+5.0}_{-4.1}$
$t\bar{t}$	$^{+9.9}_{-10.7}$
Single top	$^{+6.8}_{-2.2}$

Table 8.5: Theoretical uncertainties for relevant background production cross sections.

CHAPTER 9

RESULTS

In this chapter, the results obtained from the $H \to \tau_{\text{had}}\tau_{\text{had}}$ analysis in ATLAS from January 2012 is presented and compared to the sensitivity of the analysis presented in in this thesis. The expected and observed limits on the Higgs cross section are presented for Higgs masses in the range $m_H \in [100, 150]$ GeV.

9.1 Limits Setting Technique

The exclusion limits on the Higgs production cross section is set using a binned profile likelihood fit based on methods described in [78]. It is customary to express the limits in terms of the signal strength parameter μ which normalises the limit on the cross section by the SM Higgs cross section, i.e. a limit on μ below 1 corresponds to an exclusion of the SM Higgs cross section at 95% Confidence Level, CL. It should be noted that μ only expresses a change in the cross section on the Higgs production while the Higgs BR are assumed fixed to the SM values.

The overall likelihood is given as the product of the likelihoods in each category. For a given category a likelihood is defined as

$$\mathcal{L}(x_i|\mu,\theta) = \prod_i \text{Poisson}(x_i|\mu \cdot s_i(\theta) + b_i(\theta)) \cdot \mathcal{L}(\theta) \tag{9.1}$$

where

- x_i denotes the number of events in each bin[1].
- s_i and b_i is expected signal and background yields in each bin.
- θ is the array of systematic uncertainties expressed as nuisance parameters.

In the following, a procedure for quantifying the compatibility of data, x_i with the two hypotheses of *background-only* and *signal+background* for an assumed signal strength μ is described. In the limit setting procedure, this routine is performed as an iterative process in order to find an upper limit on μ.

[1]Here, these events can refers to both events from pseudo-data or from observed data

First, a test statistic, $\tilde{q}_\mu$ [84], is defined as

$$\tilde{q}_\mu = \begin{cases} -2\ln \mathcal{L}(x_i|\mu,\hat{\theta}_0)/\mathcal{L}(x_i|0,\hat{\theta}) & \text{for} \quad \hat{\mu} < 0 \\ -2\ln \mathcal{L}(x_i|\mu,\hat{\theta}_\mu)/\mathcal{L}(x_i|\hat{\mu},\hat{\theta}) & \text{for} \quad 0 \leq \hat{\mu} \leq \mu \end{cases} \tag{9.2}$$

where

- $\hat{\theta}_\mu$ are the conditional Maximum Likelihood Estimators (MLE) of θ given μ.
- $\hat{\theta}$ and $\hat{\mu}$ represent the MLE's that globally maximises $\mathcal{L}$.

From the *observed* data, the MLE's for the nuisance parameters that best describe data under the *background-only* hypothesis, $\hat{\theta}_0^{\text{data}}$, and under the *signal+background* hypothesis, $\hat{\theta}_\mu^{\text{data}}$, are found by maximising the likelihood in equation 9.1. From these MLE's the probability density functions, p.d.f.'s, of the test statistics $\tilde{q}_\mu$ are generated.

The p.d.f of the test statistic, $f(\tilde{q}_\mu|\mu', \hat{\theta}_{\mu'}^{\text{data}})$, for the true value of $\mu = \mu'$, is derived by generating a large ensemble of Monte Carlo pseudo-data with an input value of $\mu = \mu'$ and evaluating $\tilde{q}_\mu$. The corresponding p.d.f. of the test statistic under the *background-only* hypothesis, $f(\tilde{q}_\mu|0, \hat{\theta}_0^{\text{data}})$ is also derived by pseudo-data. It should be noted that in generating the pseudo-data for both hypothesis the nuisance parameters are fixed to MLE's derived from the observed data, $\hat{\theta}_\mu^{\text{data}}$ and $\hat{\theta}_0^{\text{data}}$.

Given the value of the test statistic derived from data $\tilde{q}_\mu^{\text{data}}$, the two probabilities of getting an equal or large value of the test statistic is calculated: under the *signal+background* hypothesis, p_μ, and under the *background-only* hypothesis, p_b.

$$\begin{aligned} p_\mu &= P(\tilde{q}_\mu \geq \tilde{q}_\mu^{\text{data}}|\,\text{signal+background}) \\ &= \int_{\tilde{q}_\mu^{\text{data}}}^{\infty} f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{\text{obs}}) \\ 1 - p_0 &= P(\tilde{q}_\mu \geq \tilde{q}_\mu^{\text{data}}|\,\text{background-only}) \\ &= \int_{\tilde{q}_\mu^{\text{data}}}^{\infty} f(\tilde{q}_\mu|0, \hat{\theta}_0^{\text{obs}}) \end{aligned} \tag{9.3}$$

With these definitions the probabilities, p_μ and $1 - p_0$ corresponds to the confidence levels, $CL_{\text{s+b}}$ and CL_b. From these, the confidence level ratio CL_s [85] is computed as

$$CL_s = \frac{p_\mu}{1 - p_0} \tag{9.4}$$

and the Higgs is said to be excluded at 95% CL if $CL_s \leq 0.05$ for $\mu = 1$.

Now, to set an upper limit on μ, this whole routine of pseudo-data generation and calculations is executed while modifying the input value of μ until a μ is found which correspond to a 95% CL exclusion.

9.2 Current ATLAS Limit

The current limit published by ATLAS in the $H \to \tau\tau \to \tau_{\text{had}}\tau_{\text{had}}$ search channel [51](see attached papers) is based on a H + 1jet category similar to the ggH category presented

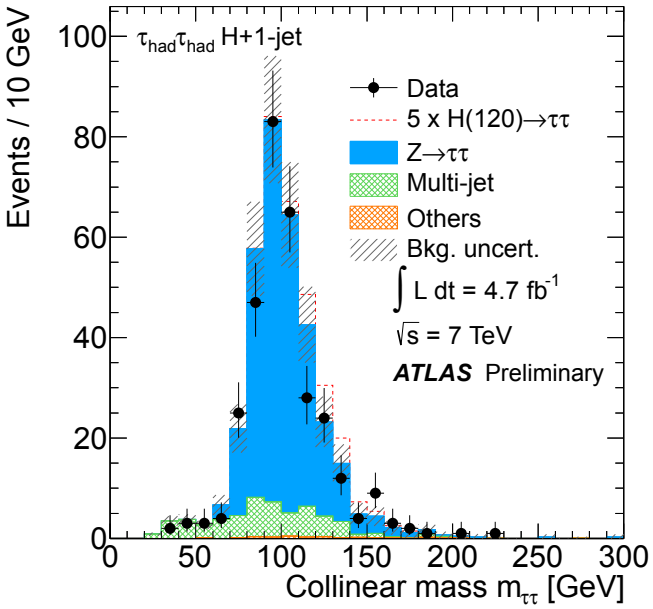


Figure 9.1: M_{CA} distribution in the $H+ \geq 1$jet signal region [51]

in section 5.4.2 with the collinear approximation mass as the discriminating variable. Figure 9.1 shows the M_{CA} distribution in the $H + 1$jet category for both data and simulations.

Since no clear excess is observed an upper limit on the Higgs production cross section is set using the machinery described in section 9.1. Both the expected upper limit in case of the *background-only* hypothesis (no Higgs) and the observed upper limit are derived, and both limits are normalised to the SM Higgs cross section as described above.

The limits are derived using a limit setting tool, `HistFactory` [86], that ensures all limits in ATLAS are derived consistently across the Higgs search channels. Statistical and systematics uncertainties are taken into account as bin-by-bin fluctuations in the signal and background shapes. Four different shape are considered as input to the limit setting: signal, $Z \to \tau\tau$, QCD-jet event and other background (including $Z \to \ell\ell$, W+jets, top and di-boson events).

The QCD and $Z \to \tau\tau$ events are normalised using the fitted $\Delta\eta(\tau_1, \tau_2)$ distributions as described in section 7.4, while the normalisation of both signal and other backgrounds are taken from theoretical predictions. Uncertainties on the shape of templates are derived by varying the tau and jet energy scale and the $E_{\mathrm{T}}^{\mathrm{miss}}$ scale and resolution. Since the $Z \to \tau\tau$ are derived partly from data, the jet energy does not affect its shape. The tau and jet energy scale uncertainties are taken to be completely correlated. All uncertainties are considered uncorrelated between background and signal samples which will lead to a more conservative limit.

Figure 9.2 shows the expected and observed upper limits on the Higgs cross section in the mass range $m_H = [100, 150]$ GeV. The expected sensitivity is still far away from $\mu = 1$ at all mass points, hence the presence of a SM Higgs boson can not be excluded

yet in the τ-decay channel for any mass point.

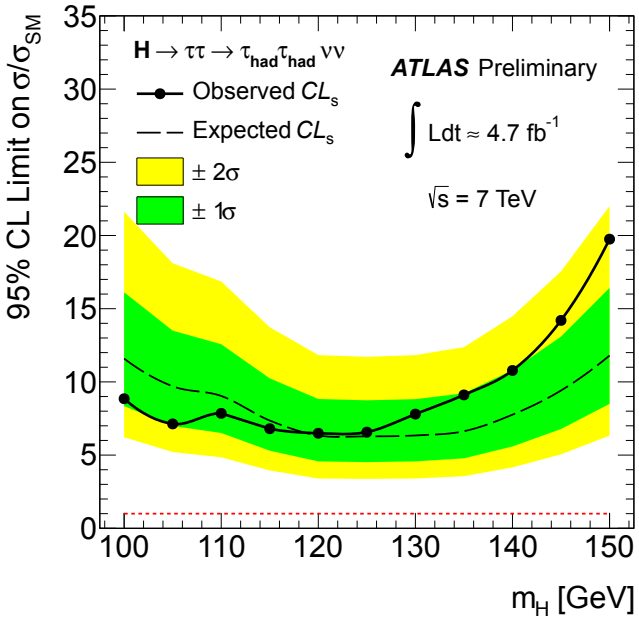


Figure 9.2: Current ATLAS upper limit on the 95% CL exclusion of the Higgs production cross section from the fully hadronic τ final states. The limit on the cross section is normalised to the SM Higgs cross section. Plot taken from [51].

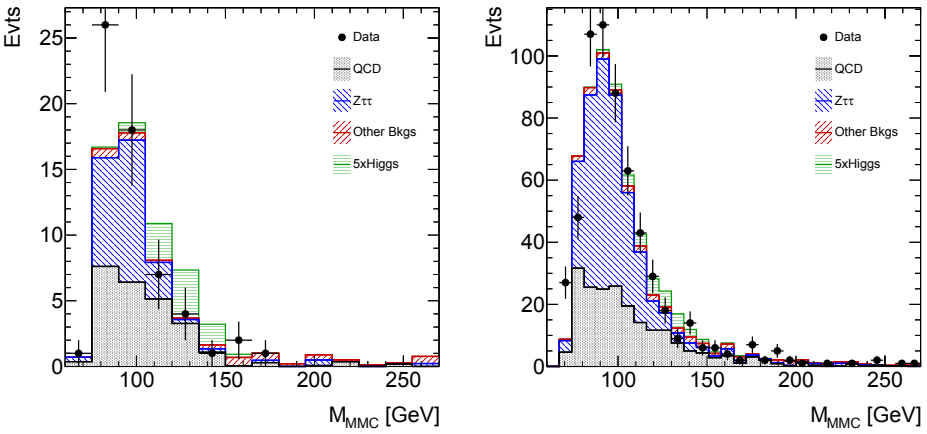
9.3 Exclusion Limits Comparison

In chapter 5, an alternative selection for $H \to \tau\tau$ events is presented. There are two main differences between this selection and the selection used in the previous section: the use of specific ggH and VBF categories and the selection of events based on the reconstructed boost of the Higgs candidates using the techniques developed in chapter 4.

Up until this point a *blinded* analysis has been carried out. Hence, the optimised selection has been derived using only the expected distributions from simulations and the data-driven background techniques. The *true* data has only been seen at the preselection level, and none of the cuts in the optimised part of the selection have been tuned by looking at the data. This is crucial for maintaining an unbiased selection.

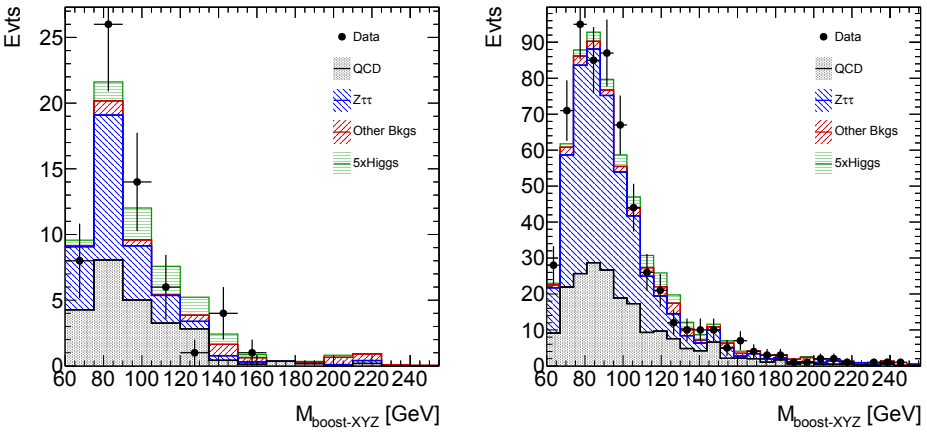
Figure 9.3 shows the expected and observed $\mathrm{M_{MMC}}$ distributions in the two signal signal regions, while figure 9.4 shows the equivalent distributions for the M_{Boost} variable. In order to avoid empty bins due to fluctuations in the background samples, a bin width of 7 GeV has been chosen for the ggH category and 15 GeV for the VBF category.

A full evaluation of all systematic uncertainties is not done here, however the following uncertainties on the number of accepted events in each signal regions are considered for the processes based on Monte Carlo predictions: the uncertainty on the integrated luminosity delivered in ATLAS during the data taking period, uncertainties on the energy



(a) VBF category (b) ggH category

Figure 9.3: The expected and observed M_{MMC} distributions in the signal regions for both background and signal processes.



(a) VBF category (b) ggH category

Figure 9.4: The expected and observed M_{Boost} distributions in the signal regions for both background and signal processes.

	Relative Uncertainties [%]			
	ggH	VBF	VH	MC Backgrounds
Luminosity	1.8	1.8	1.8	1.8
Energy Scale	7.9	12.0	0.6	$^{+7.9}_{-6.8}$
Energy Resolution	0.7	0.5	10.0	4.5
$E_{\mathrm{T}}^{\mathrm{miss}}$	0.7	0.7	0.7	0.8
τ-ID	8.0	8.0	8.0	5.0
Trigger Efficiency	4.0	4.0	4.0	2.3
PDF	$^{+7.6}_{-7.1}$	2.3	$^{+3.6}_{-3.4}$	$^{+4.7}_{-4.1}$
QCD Scales	$^{+29.8}_{-23.0}$	1.6	1.5	0.5
Underlying Event	5.7	5.7	5.7	-

Table 9.1: Systematic uncertainties on the acceptance of events from simulated samples in the VBF signal region.

scale for both τ-candidates and jets, the uncertainties on the energy resolution for both τ-leptons and jets, the resolutions of the reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$-vector, the efficiencies on the used triggers and the efficiency of the τ-identification methods. All these uncertainties are evaluated by tools or numbers provided by the performs groups. The uncertainties on the energy scales for τ-leptons and jets have been taken to be 100% correlated.

Additionally, uncertainties from theory are evaluated as described in chapter 8. These uncertainties are among the largest for the signal samples. In addition to the theoretical uncertainties described in chapter 8, an uncertainties on the underlying event model is included. This uncertainty is determined by running the selection on samples generated with different models of the underlying event. The systematic uncertainties are summarised in table 9.1 for the VBF category and in Table 9.2 for the ggH category. In addition to the uncertainties on the simulated events, uncertainties on the normalisation of the data-driven samples for QCD-jets (17%) and $Z \to \tau\tau$ events are also used. Besides the uncertainties on the normalisation, the embedded $Z \to \tau\tau$ events also have an uncertainty from the energy scale of the reconstructed τ-objects and an uncertainty from the embedding technique coming from the selection of $Z \to \mu\mu$ events. The systematic uncertainties related to the $Z \to \tau\tau$ sample are listed in table 9.3.

The obtained values for all systematic uncertainties are derived using a very similar re-optimised selection of events proposed in the $H \to \tau_{\mathrm{had}}\tau_{\mathrm{had}}$ working group. The proposed re-optimised selection is described in appendix C. Since none of the deviations seen in the validation plots in appendix B exceeds the deviations found by using the re-optimised selection, the systematic uncertainties listed here are assumed to cover the uncertainties for this selection as well. However, no uncertainties on the mass shapes for either signal or background processes are included here.

Using the $\mathrm{M_{MMC}}$ distributions from figure 9.3 as the discriminating variable, an updated expected exclusion limit on the Higgs production cross section is shown in figure 9.5. This expected limit is achieved by combining the two signal regions. In addition to the combined expected limit, the updated expected limits using only events from the ggH or VBF categories are also shown. Despite the very different number of events in the two categories, similar sensitivity is achieve from the the two categories. The combined expected limit shows clear improvement compared to the current expected limit.

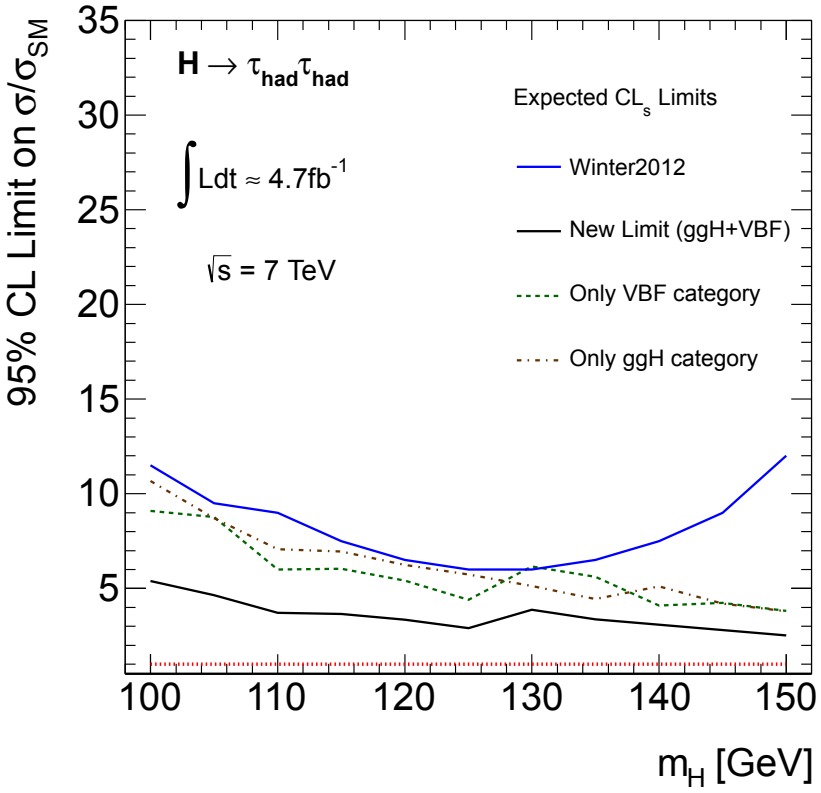


Figure 9.5: Current and updated expected upper limits on the 95% CL exclusion of the Higgs production cross section. The limits on the cross section are normalised to the SM Higgs cross section. For the updated limit the individual expected limits using only the ggH or VBF categories are also shown.

	Relative Uncertainties [%]			
	ggH	VBF	VH	MC Backgrounds
Luminosity	1.8	1.8	1.8	1.8
Energy Scale	1.5	0.4	1.9	6.8
Energy Resolution	3.4	0.7	1.4	1.4
$E_{\mathrm{T}}^{\mathrm{miss}}$	0.7	0.7	0.7	0.8
τ-ID	8.0	8.0	8.0	6.0
Trigger Efficiency	4.0	4.0	4.0	3.0
PDF	$^{+7.6}_{-7.1}$	2.3	$^{+3.6}_{-3.4}$	$^{+4.7}_{-4.1}$
QCD Scales	19.0	$^{+3.8}_{-3.7}$	1.5	0.5
Underlying Event	6.9	6.9	6.9	-

Table 9.2: Systematic uncertainties on the acceptance of events from simulated samples in the ggH signal region.

	Energy Scale	Embedding Selection	Normalisation
ggH category	9.5	$^{+6.0}_{-4.0}$	4.7
VBF category	$^{+10.0}_{-8.0}$	$^{+6.1}_{-2.1}$	5.1

Table 9.3: Systematic uncertainties on the acceptance of events from the $Z \to \tau\tau$ events in the ggH and VBF signal regions.

In figure 9.5, the $\mathrm{M_{MMC}}$ is used as discriminating variable, because most sensitivity is expected in this variable based on the separation power comparison done in section 6.2. Figure 9.6 shows the updated expected limits derived using $\mathrm{M_{MMC}}$, M_{Boost} and $\mathrm{M_{MMC}}$ as the discriminating variable, along with the observed limit found using $\mathrm{M_{MMC}}$. For all updated limits, the same selection of events is used, and the same set of systematic uncertainties are assign. As expeced the $\mathrm{M_{MMC}}$ variable gives the best limit, while the M_{Boost} and $\mathrm{M_{PC}}$ are right on top of each other and only deviating significantly at the last mass point. By checking the numerical results the $\mathrm{M_{PC}}$ and M_{Boost} do differ at the other mass points as well, however only very slightly. Hence, the ranking of the mass variables given in section 6.2 holds for the limits on the Higgs cross section. The observed limit shows only minor deviations from the expected limit. The small excess at the low mass points are within the 2σ errorband, and fluctuations of this size are to be expected.

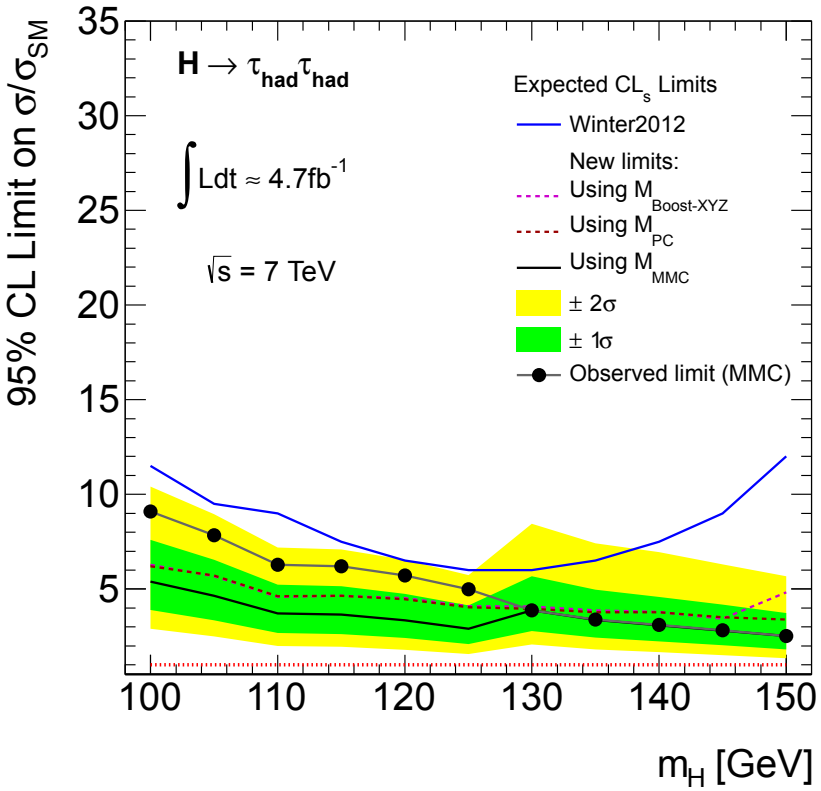


Figure 9.6: Expected upper limits on the 95% CL exclusion of the Higgs production cross section for different discriminating variables, along with the observed limit when using the M_{MMC}. The 1σ and 2σ errorbands on the μ are shown for the expected limit using M_{MMC}.

CHAPTER 10

Conclusion

In this thesis, a new technique for reconstructing the rest frame of τ-pair systems at hadron colliders is described. This technique is based on minimising the acollinearity of the τ-pair by boosting the system along a given direction. No assumptions on the parton distribution functions in the hard collision or the spin of the resonance are used to derived the reconstructed rest frame, which should make the technique applicable for other final states as well.

Two sub-methods of the rest frame reconstruction are presented depending on whether the transverse boost of the system is neglected or not. Both sub-methods can be applied to all event topologies and have no or little dependence on the $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction. The methods were first presented in [29] where the performance was evaluated using generator level studies. In this thesis, an extensive study of the performance on Z^0 and Higgs events at the LHC is presented using full ATLAS detector simulation. The study demonstrates that the rest frame is well reconstructed and the performance from the generator level studies is only slightly worsened by including detector reconstruction.

The latter part of this thesis focuses on the search for the Higgs boson in the fully hadronic τ-decay channel. This decay mode is very difficult due to the huge background from QCD-jet events, and the irreducible background from $Z \to \tau\tau$. It is shown, that the reconstructed boost of the τ-pair rest frame can be used to efficiently select τ-pairs from both VBF and ggH Higgs bosons. Here, an alternative selection of Higgs candidate events is given, which deviates from both the latest ATLAS selection published last winter and the proposed re-optimised selection to be published by ATLAS. The biggest difference between the selection presented here and the other selections, is the use of the reconstructed boost of the Higgs candidate.

Using the reconstructed rest frame, a way of estimating mass of the τ-pair system is also developed. Although, the mass estimation was originally intended as an estimation technique based on an ensemble distribution, the event-by-event mass estimator, M_{Boost}, is shown to have comparable performance to the existing mass estimation methods. Two key features of this mass reconstruction techniques are that it works for all event topologies and only makes minimal use of the missing transverse energy, $E_{\mathrm{T}}^{\mathrm{miss}}$. Using a principal component analysis, the new mass variable is combined with the visible invariant mass of the τ-pair system, and a principal component mass variable is constructed. However, little or no sensitivity is gained by combining these variables in

the current search for Higgs boson.

An important issue in all searches in the τ-pair channel, is the data-driven normalisation of the two main backgrounds. An alternative way of normalising the QCD and $Z \to \tau\tau$ backgrounds using the $\Delta\eta$ distribution between the two τ-jets is presented. This method is shown to give a consistent result with the normalisation used in the current ATLAS analysis.

Using the new selection, an updated limit on the 95% exclusion on the Higgs cross section is computed using three different mass estimator as the discriminating variable. The some preliminary systematic uncertainties have been taking from the very similar re-optimised selection. The largest uncertainties are coming from the theoretical predictions on the signal yield in each signal region. A part of this thesis work has gone into deriving these uncertainties both for analysis published last winter and for the proposed re-optimised selection. The uncertainties are computed through numerous calculations using the programs HNNLO and HAWK.

For all three mass variables, the updated limit on the Higgs cross section shows significant improvement over the current ATLAS limit. The best limit is found using the $\mathrm{M_{MMC}}$ with an upper limit on 2.9 times the Standard Model Higgs cross section for a Higgs mass of 125 GeV. However, no systematic uncertainties on the shapes of these mass variables have been included in the limit setting, and therefore a worsening the these limits is to be expected when including a full set of systematic uncertainties. Here it should be noted that since the M_{Boost} variable makes minimal use of the $E_{\mathrm{T}}^{\mathrm{miss}}$-vector and the production mechanism in the events, a larger uncertainties should be expected on the $\mathrm{M_{MMC}}$ than on the M_{Boost} distributions.

Also, an updated observed upper limit on the Higgs production cross section using the $\mathrm{M_{MMC}}$ variable is presented. For all mass points, the observed limit is consistent within 1.5σ deviations from the expected limit, and hence no significant excess is seen.

For a boson decaying into a pair of τ-leptons, the spin of the boson will determine the polarisation configuration of the two τ-leptons, which leads to distinct energy correlations between the visible τ-objects. These correlations are hidden in the detector frame by large boost of the system. It is shown, how the rest frame reconstruction, regains part of the energy correlations in the $\tau^{\pm} \to h^{\pm}\,\nu_{\tau}$ decay mode. Unfortunately, the harsh requirement on the p_{T} of the τ-candidates enforced by the trigger, is biasing these energy correlations and therefore these correlations has yet to be used in the full analysis.

Finally, this thesis documents the structure of the prompt calibration model for the Semi Conductor Tracker. The automation of these processes has guaranteed a consistent and robust processing of the calibration data, and helped save a huge amount of working hours.

Bibliography

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

[2] Particle Data Group Collaboration, *Review of particle physics*, *J. Phys. G.* **33** (2006) 1.

[3] F. Mandl and G. Shaw, *Quantum Field Theory*. Wiley, 2005.

[4] G. Karagiorgi, A. Aguilar-Arevalo, J. M. Conrad, M. H. Shaevitz, K. Whisnant, M. Sorel, and V. Barger, *Leptonic cp violation studies at miniboone in the $(3+2)$ sterile neutrino oscillation hypothesis*, *Phys. Rev. D* **75** (Jan, 2007) 013011.

[5] The ATLAS Collaboration, *ATLAS detector and physics performance: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1999.

[6] ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, The LEP Working Group for Higgs Boson Searches, *Search for the standard model higgs boson at lep*, *Physics Letters B* **565** (2003), no. 0 61 – 75.

[7] The ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys.Lett.B* (2012) [arXiv:1207.7214].

[8] The CMS Collaboration, *Observation of a new boson with a mass near 125 gev*, Tech. Rep. CMS-PAS-HIG-12-020, 2012.

[9] M. Ciccolini, A. Denner, and S. Dittmaier, *Strong and electroweak corrections to the production of Higgs + 2-jets via weak interactions at the LHC*, *Phys. Rev. Lett.* **99** (2007) 161803, [arXiv:0707.0381].

[10] M. Ciccolini, A. Denner, and S. Dittmaier, *Electroweak and QCD corrections to Higgs production via vector-boson fusion at the LHC*, *Phys. Rev.* **D77** (2008) 013002, [arXiv:0710.4749].

[11] S. Asai et al., *Prospects for the search for a standard model higgs boson in atlas using vector boson fusion*, *Eur. Phys. J.* **C32S2** (2004) 19.

[12] LHC Higgs XSEC WG. https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CrossSectionsFigures.

[13] The LEP Electroweak Working Group, *A combination of preliminary electroweak measurements and constraints on the standard model*, *Phys. Lett.* (2005) [hep-ex/0511027].

[14] D. L. Rainwater et al., *Searching for $h \to \tau\tau$ in weak boson fusion at the lhc*, *Phys. Rev.* **D59** (1999) 014037.

[15] M. Duhrssen et al., *Extracting higgs boson couplings from lhc data*, *Phys. Rev.* **D70** (2004) 113009.

[16] The CMS Collaboration, *Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc*, *Phys. Lett. B* **716** (Jul, 2012) 30–61. 59 p.

[17] O. S. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock, *LHC Design Report*. CERN, Geneva, 2004.

[18] D. Green, *High p_T physics*. Cambridge University Press, 2000.

[19] The ATLAS Collaboration, *The atlas experiment at the cern large hadron collider*, *Journal of Instrumentation* **3** (2008), no. 08 S08003.

[20] The ATLAS Collaboration Tech. Rep. ATL-PHYS-PUB-2009-002, CERN, Geneva.

[21] J. Carter et al., *The silicon microstrip sensors of the atlas semiconductor tracker*, *Nucl. Instrum. Meth.* **A578** (2007) 98–118.

[22] C. Gabaldon, *Performance of the atlas trigger system*, *Journal of Instrumentation* **7** (2012), no. 01 C01092.

[23] ATLAS Tau Trigger Public Results. https://twiki.cern.ch/twiki/bin/view/AtlasPublic/TauTriggerPublicResults.

[24] The ATLAS Collaboration, *The atlas computing technical design report*, Tech. Rep. ATLAS-TDR-017 ; CERN-LHCC-2005-022, June, 2005.

[25] R. W. L. Jones and D. Barberis, *The evolution of the atlas computing model*, *Journal of Physics: Conference Series* **219** (2010), no. 7 072037.

[26] "Wlcg." http://wlcg.web.cern.ch.

[27] S. A. et al., *Geant4—a simulation toolkit*, *Nucl.Inst.Methods in Physics Research Section A* **506** (2003), no. 3 250 – 303.

[28] D. Adams, D. Barberis, C. P. Bee, R. Hawkings, S. Jarp, R. Jones, D. Malon, L. Poggioli, G. Poulard, D. Quarrie, and T. Wenaus, *The atlas computing model*, Tech. Rep. ATL-SOFT-2004-007. ATL-COM-SOFT-2004-009. CERN-ATL-COM-SOFT-2004-009. CERN-LHCC-2004-037-G-085, CERN, Geneva, Dec, 2004.

[29] P. L. Rosendahl, T. Burgess and B. Stugu, *A method to estimate the boson mass and to optimise sensitivity to helicity correlations of tau+tau- final states*, *Journal of High Energy Physics* **2012** (2012) 043, [arXiv:1105.6003].

[30] T. Sjostrand, S. Mrenna, and P. Skands, *A Brief Introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852–867, [arXiv:0710.3820].

[31] R. K. Ellis et al., *Higgs decay to $\tau^+\tau^-$: A possible signature of intermediate mass Higgs bosons at high energy hadron colliders*, *Nucl. Phys. B* **297** (1988) 221–243.

[32] B. Mellado, W. Quayle and S. L. Wu, *Prospects for the Observation of a Higgs Boson with $H \to \tau\tau$ Associated with One Jet at the LHC*, *Phys. Lett. B* **611** (2005) 60.

[33] A. Elagin, P. Murat, A. Pranko and A. Safonov, *A New Mass Reconstruction Technique for Resonances Decaying to di-tau*, arXiv:1012.4686.

[34] E. Klinkby, *W mass measurement and simulation of the Transition Radiation Tracker at the ATLAS experiment.* PhD thesis, Niels Bohr Institute, 2008.

[35] M. Davier, L. Duflot, F. Le Dibirder, and A. Rougé, *The optimal method for the measurement of tau polarization*, *Phys. Lett. B* **306** (1993) 411.

[36] N. Davidson, G. Nanava, T. Przedzinski, E. Richter-Was, and Z. Was, *Universal Interface of TAUOLA Technical and Physics Documentation*, arXiv:1002.0543.

[37] The ATLAS Collaboration, *Measurement of τ polarization in $w \to \tau\nu$ decays with the atlas detector in pp collisions at $\sqrt{s}$ = 7 tev. oai:cds.cern.ch:1445243*, *Eur. Phys. J. C* **72** (May, 2012) 2062. 25 p.

[38] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.), *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*, *CERN-2011-002* (CERN, Geneva, 2011) [arXiv:1101.0593].

[39] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.), *Handbook of LHC Higgs Cross Sections: 2. Differential Distributions*, *CERN-2012-002* (CERN, Geneva, 2012) [arXiv:1201.3084].

[40] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, *JHEP* **0411** (2004) 040, [hep-ph/0409146].

[41] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, *JHEP* **0711** (2007) 070, [arXiv:0709.2092].

[42] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **1006** (2010) 043, [arXiv:1002.2581].

[43] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, *JHEP* **07** (2003) 001, [hep-ph/0206293].

[44] M. L. Mangano, M. Moretti, and R. Pittau, *Multijet matrix elements and shower evolution in hadronic collisions: $Wb\bar{b}$ + n jets as a case study*, *Nucl. Phys.* **B632** (2002) 343–362, [hep-ph/0108069].

[45] S. Frixione and B. R. Webber, *Matching nlo qcd computations and parton shower simulations*, *JHEP* **0206** (2002) 029.

[46] B. P. Kersevan and E. Richter-Was, *The Monte Carlo event generator AcerMC version 2.0 with interfaces to PYTHIA 6.2 and HERWIG 6.5*, hep-ph/0405247.

[47] T. Binoth, M. Ciccolini, N. Kauer, and M. Kramer, *Gluon-induced W-boson pair production at the LHC*, *JHEP* **0612** (2006) 046, [hep-ph/0611170].

[48] G. Corcella *et. al.*, *HERWIG 6.5 release note*, hep-ph/0210213.

[49] J. M. Butterworth, J. R. Forshaw, and M. H. Seymour, *Multiparton interactions in photoproduction at hera*, *Z. Phys.* **C72** (1996) 637–646, [hep-ph/9601371].

[50] N. Davidson, T. Przedzinski, and Z. Was, *PHOTOS Interface in C++: Technical and Physics Documentation*, arXiv:1011.0937.

[51] The ATLAS Collaboration, *Search for the Standard Model Higgs boson in the H to tau+ tau- decay mode in sqrt(s) = 7 TeV pp collisions with ATLAS*, arXiv:1206.5971.

[52] *Reconstruction, energy calibration, and identification of hadronically decaying tau leptons*, Tech. Rep. ATLAS-CONF-2011-077, CERN, Geneva, May, 2011.

[53] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-kt jet clustering algorithm*, *Journal of High Energy Physics* **2008** (2008), no. 04 063.

[54] *Data-quality requirements and event cleaning for jets and missing transverse energy reconstruction with the atlas detector in proton-proton collisions at a center-of-mass energy of sqrts = 7 tev*, Tech. Rep. ATLAS-CONF-2010-038, CERN, Geneva, Jul, 2010.

[55] B. Resende, *Muon identification algorithms in atlas*, Tech. Rep. ATL-PHYS-PROC-2009-113, CERN, Geneva, Sep, 2009.

[56] W. Lampl et Al., *Calorimeter clustering algorithms: Description and performance*, Tech. Rep. ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, CERN, Geneva, Apr, 2008.

[57] T. Barillari et al Tech. Rep. ATL-LARG-PUB-2009-001-2. ATL-COM-LARG-2008-006. ATL-LARG-PUB-2009-001, CERN, Geneva, Jun, 2008.

[58] T. A. Collaboration, *Performance of missing transverse momentum reconstruction in proton-proton collisions at 7 tev with atlas*, 1108.5602.

[59] The ATLAS Collaboration, *Expected performance of the atlas experiment – detector, trigger and physics*, arXiv:0901.0512.

[60] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language*, *Nucl.Phys.* **B126** (1977) 298.

[61] V. Gribov and L. Lipatov, *Deep inelastic e p scattering in perturbation theory*, *Sov.J.Nucl.Phys.* **15** (1972) 438–450.

[62] *Performance of the reconstruction and identification of hadronic tau decays with atlas*, Tech. Rep. ATLAS-CONF-2011-152, CERN, Geneva, Nov, 2011.

[63] *Measurement of hadronic tau decay identification efficiency using w->taunu events*, Tech. Rep. ATLAS-CONF-2011-093, CERN, Geneva, Jul, 2011.

[64] H. Georgi, S. Glashow, M. Machacek, and D. V. Nanopoulos, *Higgs bosons from two gluon annihilation in proton proton collisions*, *Phys.Rev.Lett.* **40** (1978) 692Ð694.

[65] A. Djouadi, M. Spira, and P. Zerwas, *Production of Higgs bosons in proton colliders: QCD corrections*, *Phys.Lett.* **B264** (1991) 440–446.

[66] S. Dawson, *Radiative corrections to Higgs boson production*, *Nucl.Phys.* **B359** (1991) 283–300.

[67] M. Spira, A. Djouadi, D. Graudenz, and P. Zerwas, *Higgs boson production at the LHC*, *Nucl.Phys.* **B453** (1995) 17–82, [hep-ph/9504378].

[68] R. V. Harlander and W. B. Kilgore, *Next-to-next-to-leading order Higgs production at hadron colliders*, *Phys.Rev.Lett.* **88** (2002) 201801, [hep-ph/0201206].

[69] C. Anastasiou and K. Melnikov, *Higgs boson production at hadron colliders in NNLO QCD*, *Nucl.Phys.* **B646** (2002) 220–256, [hep-ph/0207004].

[70] V. Ravindran, J. Smith, and W. L. van Neerven, *NNLO corrections to the total cross-section for Higgs boson production in hadron hadron collisions*, *Nucl.Phys.* **B665** (2003) 325–366, [hep-ph/0302135].

[71] S. Catani, D. de Florian, M. Grazzini, and P. Nason, *Soft gluon resummation for Higgs boson production at hadron colliders*, *JHEP* **0307** (2003) 028, [hep-ph/0306211].

[72] U. Aglietti, R. Bonciani, G. Degrassi, and A. Vicini, *Two loop light fermion contribution to Higgs production and decays*, *Phys.Lett.* **B595** (2004) 432–441, [hep-ph/0404071].

[73] S. Actis, G. Passarino, C. Sturm, and S. Uccirati, *NLO electroweak corrections to Higgs boson production at hadron colliders*, *Phys.Lett.* **B670** (2008) 12–17, [arXiv:0809.1301].

[74] S. Catani and M. Grazzini *Phys.Rev.Lett.* **98** (2007) 222002.

[75] M. Grazzini arXiv:0801.3232.

[76] I. W.Stewart and F. J. Tackmann, *Theory Uncertainties for Higgs and Other Searches Using Jet Bins*, *Phys.Rev.* **D85** (2012) 034011, [arXiv:1107.2117].

[77] J. M. Campbell, R. K. Ellis and G. Zanderighi, *Next-to-leading order Higgs + 2 jet production via gluon fusion*, *JHEP* **0610** (2006) 028, [0608194].

[78] *Procedure for the lhc higgs boson search combination in summer 2011*, Tech. Rep. ATL-PHYS-PUB-2011-011, CERN, Geneva, Aug, 2011.

[79] A. Denner, S. Dittmaier, S. Kallweit, and A. Mück, *Electroweak corrections to Higgs-strahlung off W/Z bosons at the Tevatron and the LHC with HAWK*, arXiv:1112.5142.

[80] J. Butterworth, E. Dobson, U. Klein, B. Mellado Garcia, T. Nunnemann, J. Qian, D. Rebuzzi, and R. Tanaka, *Single Boson and Diboson Production Cross Sections in pp Collisions at sqrts = 7 TeV*, .

[81] M. Aliev, H. Lacker, U. Langenfeld, S. Moch, P. Uwer, and M. Wiedermann, *Hathor – hadronic top and heavy quarks cross section calculator*, *Computer Physics Communications* **182** (2011), no. 4 1034 – 1046.

[82] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, *Parton distributions for the LHC*, *Eur. Phys. J.* **C63** (2009) 189–285, [arXiv:0901.0002].

[83] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, *Uncertainties on alpha(S) in global PDF analyses and implications for predicted hadronic cross sections*, *Eur.Phys.J.* **C64** (2009) 653–680, [arXiv:0905.3531].

[84] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *European Physical Journal C* **71** (Feb., 2011) 1554, [arXiv:1007.1727].

[85] A. L. Read, *Presentation of search results: the cl s technique*, *Journal of Physics G: Nuclear and Particle Physics* **28** (2002), no. 10 2693.

[86] K. Cranmer, G. Lewis, L. Moneta, A. Shibata, and W. Verkerke, *Histfactory: A tool for creating statistical models for use with roofit and roostats. oai:cds.cern.ch:1456844*, Tech. Rep. CERN-OPEN-2012-016, SUNY, Stony Brook, New York, Jan, 2012.