

# **Neural Network-based Signal Isolation and Cross Section Estimation of the $tH$ Process with the ATLAS Detector**

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

vorgelegt von  
Christian Kirfel  
aus  
Mechernich

Bonn 2023

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen  
Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Ian C. Brock  
2. Gutachter: Prof. Dr. Klaus Desch

Tag der Promotion:  
Erscheinungsjahr:

---

# Contents

---

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Particle physics</b>	<b>3</b>
2.1	The Standard Model of particle physics . . . . .	3
2.1.1	Matter particles in the Standard Model . . . . .	4
2.1.2	Feynman diagrams . . . . .	4
2.1.3	Interactions in the Standard Model . . . . .	6
2.1.4	The Higgs mechanism . . . . .	9
2.1.5	Open questions in the Standard Model . . . . .	11
2.2	Physics of proton collisions . . . . .	12
2.3	The associated production of a $t$ -quark and a Higgs boson . . . . .	13
2.3.1	Higgs boson physics . . . . .	14
2.3.2	Top physics . . . . .	14
2.3.3	Tau physics . . . . .	17
2.3.4	Predictive power of the analysis . . . . .	17
2.4	Background processes . . . . .	18
<b>3</b>	<b>Particle physics experiments</b>	<b>23</b>
3.1	Particle accelerators . . . . .	23
3.1.1	The Large Hadron Collider . . . . .	25
3.2	Particle detectors . . . . .	26
3.2.1	Tracking detectors . . . . .	28
3.2.2	Calorimeters . . . . .	28
3.2.3	Muon detectors . . . . .	31
3.2.4	Missing energy . . . . .	31
3.3	The ATLAS detector . . . . .	31
3.3.1	The ATLAS coordinate system . . . . .	32
3.3.2	The Inner Detector . . . . .	34
3.3.3	The calorimeter system . . . . .	34
3.3.4	The muon spectrometer . . . . .	35
3.3.5	The ATLAS magnet system . . . . .	35
3.3.6	The ATLAS trigger system . . . . .	36
3.4	Object reconstruction . . . . .	37
3.4.1	ID object reconstruction . . . . .	37
3.4.2	Calorimeter clusters . . . . .	38

3.4.3	Muon reconstruction	39
3.4.4	Missing transverse energy	40
3.4.5	Particle flow objects	41
3.4.6	Electron reconstruction	42
3.4.7	Jet reconstruction	43
3.4.8	Hadronic tau reconstruction	44
3.4.9	<i>b</i> -jet reconstruction	45
<b>4</b>	<b>Statistical methods</b>	<b>47</b>
4.1	Fundamentals of probability theory	48
4.2	Hypothesis tests	49
4.3	Maximum likelihood estimation	49
4.3.1	Binned likelihood estimation	51
4.3.2	Nuisance parameters	52
4.4	Monte Carlo simulation	52
4.4.1	The Asimov data set	53
<b>5</b>	<b>Machine Learning</b>	<b>55</b>
5.1	Learning algorithm	56
5.1.1	Task	56
5.1.2	Experience	56
5.1.3	Performance	57
5.2	A specific learning algorithm - the neural network	57
5.2.1	Categorical classification	59
5.2.2	An alternative approach - Boosted Decision Trees	61
5.3	Optimisation algorithms	61
5.3.1	Adaptive optimisers	63
5.3.2	Neural network metrics	64
5.4	Algorithm capacity	64
5.4.1	Dropout	67
5.4.2	Batch normalisation	67
<b>6</b>	<b>Data preprocessing</b>	<b>69</b>
6.1	Event selection	69
6.2	Object reconstruction	74
6.2.1	Lepton association	74
6.2.2	Top quark reconstruction	75
6.2.3	Higgs boson reconstruction	75
6.3	Lepton fake estimation	77
<b>7</b>	<b>Signal isolation</b>	<b>81</b>
7.1	Neural network design	81
7.2	Choice of features	82
7.3	Choice of regularisation	83
7.4	Hyperparameter optimisation	86



7.5	Model evaluation	87
7.6	K-fold cross validation	94
<b>8</b>	<b>Cross section estimation</b>	<b>97</b>
8.1	Region definition	97
8.2	Uncertainty treatment	100
8.3	Fit preparation	101
8.4	Asimov fit	101
8.5	Data fit	103
8.6	Fit combination and discussion	117
<b>9</b>	<b>Conclusion</b>	<b>119</b>
<b>A</b>	<b>Gauge symmetry</b>	<b>121</b>
<b>B</b>	<b>Symmetry breaking of a complex scalar field</b>	<b>123</b>
<b>C</b>	<b>Data and MC samples</b>	<b>125</b>
<b>D</b>	<b>Fake lepton estimation</b>	<b>127</b>
<b>E</b>	<b>Feature plots <math>2\ell</math> OS + <math>1\tau_{\text{had}}</math></b>	<b>131</b>
<b>F</b>	<b>Feature plots <math>2\ell</math> SS + <math>1\tau_{\text{had}}</math></b>	<b>137</b>
<b>G</b>	<b>Feature plots <math>1\ell</math> + <math>2\tau_{\text{had}}</math></b>	<b>143</b>
<b>H</b>	<b>Additional NN result plots</b>	<b>149</b>
<b>I</b>	<b>Performance of Lorentz invariant variables in the <math>1\ell</math> + <math>2\tau_{\text{had}}</math> channel</b>	<b>155</b>
<b>J</b>	<b>K-fold cross validation</b>	<b>159</b>
<b>K</b>	<b>Additional plots for the Asimov fits</b>	<b>161</b>
<b>L</b>	<b>Additional plots for the data fits</b>	<b>167</b>
	<b>Bibliography</b>	<b>173</b>
	<b>List of Figures</b>	<b>183</b>
	<b>List of Tables</b>	<b>187</b>
	<b>Glossary</b>	<b>189</b>
	<b>Acronyms</b>	<b>191</b>
	<b>Acknowledgements</b>	<b>193</b>



---

## Introduction

---

It was a thrilling moment when the Higgs boson, the last missing piece to a theory mostly formulated in the early 1970s, was eventually discovered in 2012 [1]. At that time, I was still in school and lucky enough to be sent to the University of Bonn for a public presentation about this discovery. I was even more excited when later on I participated in a Masterclass organised by the Netzwerk Teilchenwelt [2]. It was a course designed for pupils and I learnt hands-on what the data from a particle detector look like and how they can be analysed. I was indeed fortunate to surf this wave of excitement to a degree of my own. Now, however, more than ten years have elapsed since this last ground-breaking moment in the field. No record-breaking experiment and no discovery seems in reach to boost the public excitement, potentially giving the impression that the field had fatigued. And the Standard Model, despite its acknowledged weaknesses, looms large as an insurmountable fortress.

Reflecting this period though, it was not solely the discovery of the Higgs boson that sparked my fascination for particle physics. Initially, it was the captivating simplicity of the concepts behind particle detectors, concepts I could even grasp in school. Gaining a more complete understanding of the Standard Model and then finally exploring the prospects of machine learning methods for particle physics during my Master thesis were also massive boosts to my enthusiasm. Being a particle physicist invariably entails working at the cutting edge of technology. It also involves having exceptional colleagues who stood united during the challenges posed by a pandemic, whether in research, collaboration or supervision. Therefore, I believe that for the scientist, particle physics has not lost any of its lustre in the last decade. Nevertheless, the anticipation of the next major experiment and groundbreaking discovery remains a driving force, and it is acceptable to experience moments of discouragement from time to time. In the memory of past scientific discoveries, the steep path to the result is commonly overlooked. The Standard Model of particle physics is a fortress built over the better part of a century and the Higgs boson was the last brick to its wall. Each and every scientist working on the frontier of high energy particle physics may see himself as an engineer fortifying its walls or looking for weaknesses. And as it took the ancient Greeks ten long years to overcome the fortifications of Troy, patience and hard work are necessary to understand the weaknesses of the model.

For a fortress this weakness could be insufficient supplies, its crumbling walls and the tunnels beneath. The weakness of the Standard Model are its free parameters. In a perfect theory we could predict and understand every measurement. The Standard Model relies on measurements for some of its parameters that it is not able to predict. There is a plethora of theories out there that try to fix the

gaps in the Standard Model's predictive power and each of them makes specific predictions for its free parameters. A majority of these is connected to the Higgs boson and the underlying mechanism. An especially interesting constant is the coupling between the Higgs boson and the heaviest fermion, the top quark. Both the parameter's magnitude and its sign are a promising gateway to several theoretical predictions [3].

This thesis presents a unique avenue to measure both magnitude and sign of the coupling through the cross section of the associated production of the top quark and the Higgs boson. In particular the case in which the Higgs boson decays into a pair of  $\tau$  leptons is explored. This channel is limited by a plethora of background processes adding a layer of complexity to the analysis. Therefore, the employment of machine learning methods for signal isolation is equally suitable and enticing. An artificial neural network is chosen and meticulously optimised for the task and the results are used to extract an estimate of the process's cross section. The description of this method is embedded in an explanation of the overall framework analysis.

The first half of this thesis provides an introduction to the theoretical foundation. This introduction serves rather as direct preparation for the methods and results shown later than aiming for a complete description of the topics. Chapter 2 launches with an overview of the status quo in particle physics, the Standard Model and its weaknesses. This paves the way for the motivation of the process of interest, the associated production of a top quark and a Higgs boson. Following this up, chapter 3 summarises both the experiments used in this analysis and their technological basis. The next two chapters conclude the theoretical introductions with the mathematical tools used. Chapter 4 covers the necessities from the field of statistical methods and chapter 5 adds a comprehensive guide to machine learning methods for classification tasks. The second half of this thesis covers the individual steps of the analysis and the corresponding results. The selection and preparation of data for the analysis is presented in chapter 6. Based on this, chapter 7 explains the use of artificial neural networks to isolate the signal. The methods used and their motivation are described and the results explained. Finally, the cross section estimation is the final analysis step and the topic of chapter 8. Although this thesis is intentionally concise, the appendices aim to address particular subjects crucial for anyone attempting to replicate the results or seeking a more in-depth exploration of specific aspects. In every instance, the appendices are crafted to provide value to the engaged reader, rather than serving as a repository of plots.

# Particle physics

---

As a complete illustration of high-energy particle physics is beyond the scope of this thesis, this chapter provides a description of the scientific potential of the analysis. In a first step an introduction to the Standard Model of particle physics supplies the necessary foundation. In a short interlude the physics of proton colliders is introduced because it forms the base of the latter descriptions. Subsequently, the process of interest, the associated production of a  $t$ -quark and a Higgs boson, is introduced. The characteristics of the process as well as the resulting obstacles are explained. In addition, special attention is paid to the predictive power of the process. This predictive power embodies the possible scientific progress that is to be gained from the endeavours. Thereby, an understanding of the final results is guaranteed and motivation for each step is given. Lastly, a description of the background processes is provided.

## 2.1 The Standard Model of particle physics

The Standard Model of particle physics is a highly successful and extensively verified framework which attempts to describe a maximum of phenomena in the realm of particle physics. For contemporary physicists it is tempting to see it as the collection of particles obtained after a century of groundbreaking successes in the field. However, it is more than the sum of its parts and a multitude of particles that were predicted before their experimental discovery hint at the powerful mathematical models behind it.

Barely avoiding oversimplification, it can be stated that the Dirac equation, describing the dynamics of fermions, and the local gauge principle which is responsible for the nature of interactions form the robust theoretical foundation. Quantum field theory (QFT) and the Higgs mechanism yield the formulas that allow for the description of interactions and masses. For a comprehensive entry to these topics see [4]. Last but not least the experimental data serves two purposes. On the one hand, measurements were the guiding force in the construction of the model. On the other hand, precision measurements produce tests for its validity.

In the context of this work, a phenomenological description is adequate. The only exception applies to the Higgs mechanism which cannot be satisfactorily explained without a theoretical basis. In order to not prolong the chapter unnecessarily appendices A and B briefly showcase the necessary principles.

Generally, the model is split into matter and interactions, both of which are represented by a group of particles. This section begins with a summary of matter particles before introducing the formalism

of Feynman diagrams. This leads to a summary of the three relevant forces and the Higgs mechanism. To round things up the shortcomings of the summarised model are addressed and used as a motivation for the analysis at hand. For a more thorough introduction to modern particle physics see [5].

### 2.1.1 Matter particles in the Standard Model

The matter particles are called fermions, particles of half-integer spin defined by their quantum numbers including their charges. A charge describes the susceptibility of a particle to the respective force. The fermions of the Standard Model are grouped into three generations of ascending mass, the lightest of which yielding the building blocks of our known world. The heavier generations are short lived and with increasing mass become less common in interactions. Additionally, the fermions are split into two subcategories: quarks and leptons.

There are six quarks in the Standard Model, grouped into pairs for each generation. Every pair consists of an up-type quark with an electric charge of  $+2/3$  and a down-type quark of electric charge  $-1/3$ . In addition, quarks carry weak charge and they are the only fermions which carry the colour charge allowing them to interact with the strong force. The lightest pair, the up- and down-quark ( $u$  and  $d$ ), are the constituents of the nucleons making them essential building blocks of our universe.

The second category of fermions, the leptons, also come in a pair per generation. Each lepton generation comprises an electrically charged lepton and its electrically neutral neutrino partner  $\nu$ . All leptons carry weak charge. The lightest generation of leptons is formed by the electron  $e^-$ , thereby delivering the missing building block to the atom, and the electron neutrino  $\nu_e$ .

All particles and their quantum numbers are summarised in figure 2.1. Furthermore, each particle has its antiparticle counterpart with inverted quantum numbers.

### 2.1.2 Feynman diagrams

From the view of an experimentalist, thinking about the quantities that one wants to measure gives an interesting angle for looking at the Standard Model. There are three possible fields to study: bound states, decays and scatterings. The latter two are most important for the experiment at hand. The corresponding observables are the decay rate and the cross section<sup>1</sup>, respectively. According to Fermi's Golden Rule [7], there are two ingredients to the theoretical calculations of these quantities: (1) the amplitude  $\mathcal{M}$  and (2) the available phase space. While the phase space is determined purely from kinematics, the amplitude can be obtained through the evaluation of Feynman diagrams. Feynman diagrams are pictorial representations of the interactions of particles incorporating the mathematical expressions. More accurately, in a QFT the diagram substitutes for the sum over all possible time orderings of an interaction. Although the diagrams are a way to derive the necessary calculations, an understanding of what is depicted is sufficient for this work.

Figure 2.2 shows a diagram describing the Bhabha scattering (electron-positron scattering) as an example. The interacting fermions are embodied by the labelled arrows while the interaction is represented by the curly line labelled  $\gamma$ . Time is measured horizontally and it is important to notice that no spatial information is delivered on the vertical axis. Except for the time axis, the diagrams are purely representative of the ongoing process. Particles appear as lines and the intersections or vertices embody the interactions. The incoming and outgoing particles are grouped into initial and

---

<sup>1</sup> The cross section is a measure of probability for a certain interaction. It is expressed in units of area but it is used as a parameter for any underlying process.

## Standard Model of Elementary Particles

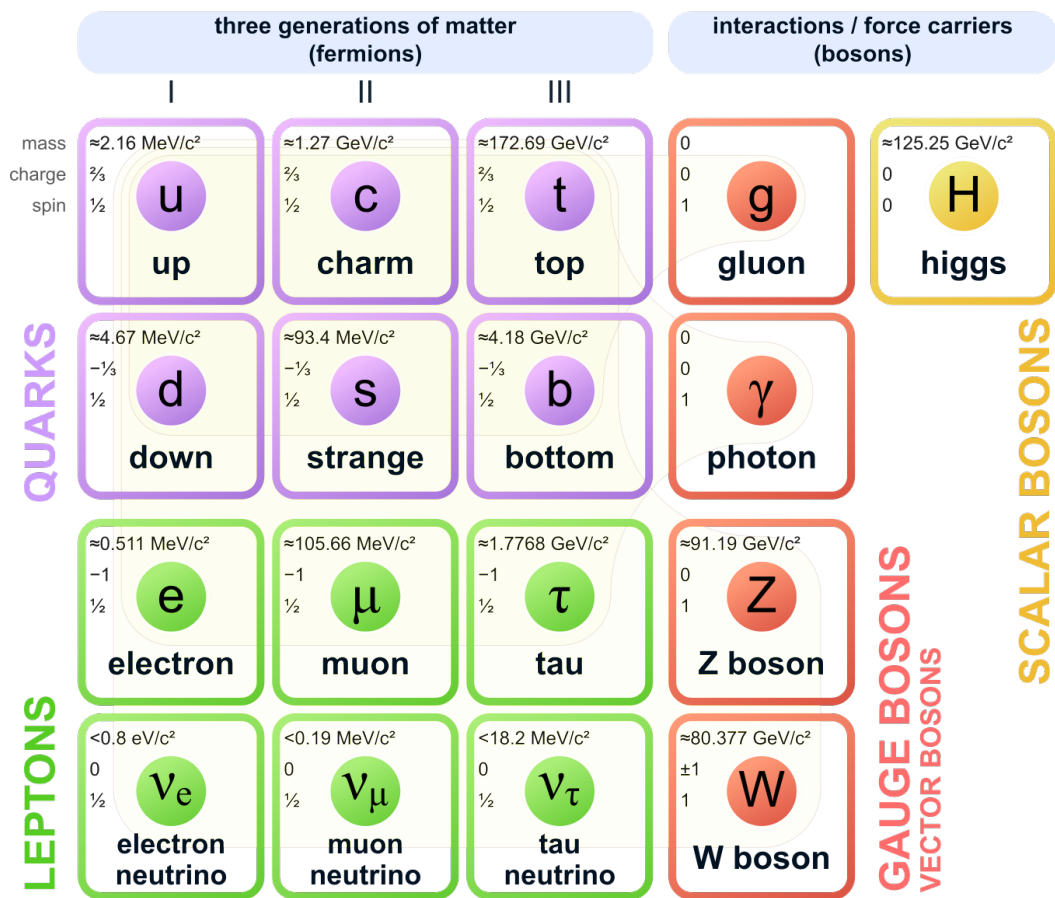


Figure 2.1: A summary of the Standard Model's constituents and their properties. Quarks are shown in purple, leptons in green, bosons in red and yellow. All properties have been updated to match the current Particle Data Group tables [3, 6].

final state respectively. Any lines in-between are virtual particles standing for the interaction. The in-between particles are not observable and simply express the manner of interaction regardless of the order. Additionally, while real particles have their fixed mass, the virtual particles can be off-shell; meaning that they can have varying masses.

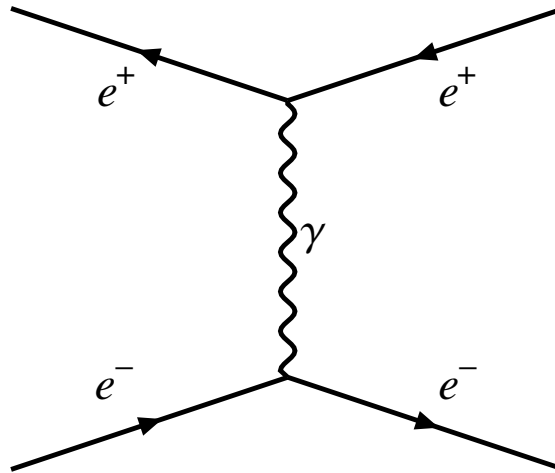


Figure 2.2: Feynman diagram of a Bhabha scattering. Electron and positron are depicted by straight lines with inverse direction for antiparticles. The exchanged  $\gamma$  is embodied by the curly line.

In general, the calculus behind the diagrams can be regarded an expansion of the process with respect to the coupling constant of the interaction. For this to be possible, the coupling constant has to be sufficiently small for the expansion to converge. Additionally, the most basic diagram usually just considers the contributions of lowest order expansion, termed leading order (LO). The next highest power of expansion is then called next to leading order (NLO). Furthermore, there is an interference between all diagrams that share the same final state. This interference can contribute a destructive or constructive term to the matrix element. For the calculus to Feynman diagrams see [8]. In this work, the diagrams are used to illustrate interactions meaningful for the analysis at hand. For each interaction the basic vertices are introduced and can then be understood as the building blocks of the advanced diagrams.

### 2.1.3 Interactions in the Standard Model

One could easily dedicate a whole chapter to the description of each of the three interactions in the Standard Model. For the sake of this work, the focus lies on the key components. Each interaction can be described by its mediator particles and its dimensionless coupling constant. Additionally, the statement of its primitive vertices in Feynman diagrams adds visual illustration. This way, a description of the necessary phenomena and guidance to understand the Feynman diagrams of the relevant processes is given.

The electromagnetic force in particle physics is described by Quantum Electrodynamics (QED) [4]. The force carrier is the photon which couples to all electrically charged particles. The most basic vertex of QED is displayed in figure 2.3. The diagram shows an incoming electron that radiates or absorbs a photon. The shown diagram uses an electron as an example but the vertex could be drawn with any other electrically charged particle instead. Due to energy and linear momentum conservation,



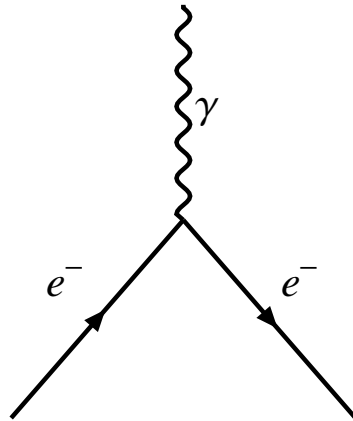


Figure 2.3: Basic QED Feynman diagram. The diagram uses electrons but any electrically charged particle could be substituted for it.

the vertex alone is not an allowed interaction. However, it can be regarded the essential ingredient of QED, easily allowing for processes like the production of an electron-positron pair or the example of Bhabha scattering earlier in figure 2.2. It is common practice to state the strength of the force via its dimensionless coupling constant. The coupling constant of the electromagnetic force is the fine-structure constant

$$\alpha = \frac{1}{137}. \quad (2.1)$$

The strong force and its underlying quantum field theory, Quantum Chromodynamics (QCD), describes interactions between particles that carry colour charge. There are three types of colour representing orthogonal states: red, blue and green. Additionally, an anticolour to each colour exists. Both a pair of colour and its anticolour, and a triplet of each colour type are regarded colour neutral. The mediator of the force is the gluon which carries a pair of colour and anti-colour, leading to a pool of eight gluons total<sup>2</sup>. The basic vertex depicted in figure 2.4(a) represents the interaction between two quarks mediated by a gluon. Since gluons, in stark contrast to the photon, also carry the charge of their interaction, they can self-interact. This results in two further primitive vertices displayed in figures 2.4(b) and 2.4(c).

It is necessary to introduce a short list of nomenclature for bound states of quarks. Generally, bound quark states are named hadrons with the special cases of mesons (quark-antiquark states) and baryons (3 quark states). The baryon number is conserved with their energetically lowest states being the nucleons of the atom. The signatures of quarks in collider physics are special in that they are not detected freely. Rather, quarks form characteristic decay chains from hadronic intermediate states and electromagnetic decays summarised as jets. The hadronisation processes for jets are essential to the reconstruction of hadronic decays in detectors and are described in a subsequent chapter. The fact that quarks are only observed in bound states leads to the postulate of colour confinement stating that quarks can only be observed in colour neutral bound states [9, 10]. At this point in time any underlying model

<sup>2</sup> Intuitively one might expect 9 gluons. The gauge group of the strong force are  $3 \times 3$  unitary matrices with trace of 0, resulting in 8 dimensions total. Alternatively, one can construct the possible gluons from the fact that they are not colour neutral.

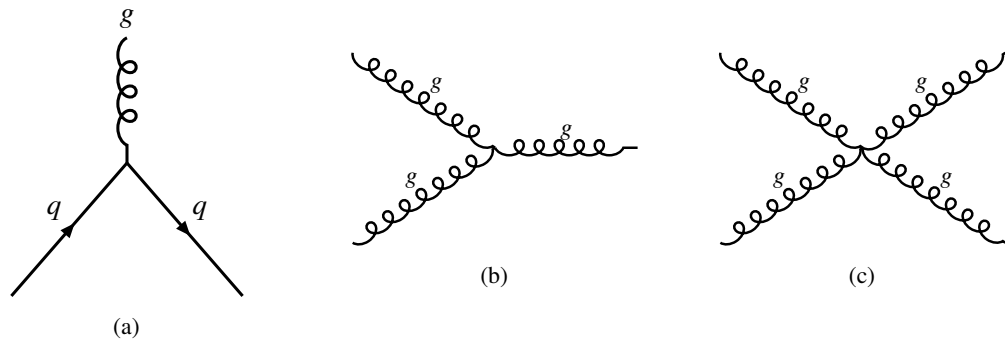


Figure 2.4: The primitive vertices of QCD. (a) shows the interaction of a gluon and a quark while (b) and (c) demonstrate the possible self-interactions of gluons.

is phenomenological only<sup>3</sup>, putting a constraint on the understanding of the underlying processes.

A frequent explanation is the nature of the coupling constant  $\alpha_S$  of QCD and its dependence on the energy scale. While for energies greater than 100 GeV  $\alpha_S$  is  $\approx 0.1$  at an energy scale of 1 GeV it becomes of order  $\mathcal{O}(1)$ . At this point perturbation theory can no longer be applied, creating a lack of theoretical modelling. A phenomenological explanation for colour confinement can then be established by considering the increasing energies for separated quarks and the gluon self-interaction. A sufficient separation of quarks provides the necessary energies for the creation of further colour neutral quark pairs. Another phenomenon that arises from the running of the strong coupling constant is asymptotic freedom. The low energy scale in bound states allows quarks to be considered as free particles.

While electric charge and colour charge are restricted to a selection of fermions, all quarks and leptons carry weak charge allowing them to interact weakly. The mediators of the weak force are the two charged  $W$  bosons and the electrically neutral  $Z$  boson. The two resulting primitive vertices are shown in figures 2.5(a) and 2.5(b). The mediator particles of the weak force are massive, putting a unique suppression on the resulting processes.

Under the electromagnetic and strong force, quark flavour is conserved. This conservation is broken for weak charged currents allowing for the transition of quark flavour. A well known example for this unique interaction are atomic  $\beta$ -decays [11]. The probability of each particular flavour transition is summarised by the CKM matrix [3]:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \quad (2.2)$$

Each of the parameters has to be determined experimentally leaving one of the biggest gaps in the model's predictive power. The Particle Data Group summarises the most recent measurements of the

<sup>3</sup> Lattice QCD offers promising explanations but is computationally extensive still.

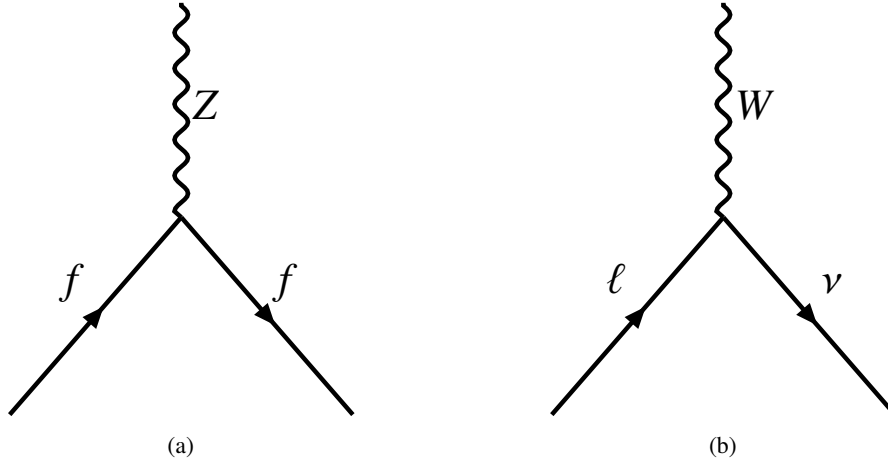


Figure 2.5: Primitive vertices of the weak interaction.

parameters as follows [3]:

$$\begin{pmatrix} 0.97401 \pm 0.00011 & 0.22650 \pm 0.00048 & 0.00361^{+0.00011}_{-0.00009} \\ 0.22636 \pm 0.00048 & 0.97320 \pm 0.00011 & 0.04053^{+0.00083}_{-0.00061} \\ 0.00854^{+0.00023}_{-0.00016} & 0.03978^{+0.00082}_{-0.00060} & 0.999172^{+0.000024}_{-0.000035} \end{pmatrix}. \quad (2.3)$$

### 2.1.4 The Higgs mechanism

The last component to the Standard Model is the Higgs Mechanism which explains how both the fermions and the heavy bosons acquire their masses. While it is true that the mathematical derivation using the principle of gauge symmetry is similar for all the components of the model, the presentation of the fermions and the three interactions has been done in an intuitive manner. It was feasible to omit the theoretical foundation since most particles can be related to more macroscopic and well known processes. This is not true, however, for the Higgs boson. Often misunderstood as the mediator of gravity, the Higgs boson is just an excitation of the associated field represented by its Proca Lagrangian. There are two mathematical concepts necessary: (1) Local gauge symmetry is used to acquire interaction terms for which appendix A gives an example. (2) Obtaining mass terms via spontaneous symmetry breaking is briefly introduced in appendix B.

It is methodical to start the introduction of the mechanism with the question it tries to answer. The  $U(1) \times SU(2)$  symmetry of the electroweak interaction is broken for both massive gauge bosons and massive fermions. The proven method in the formalism of the Standard Model to obtain terms for massive particles is the addition of a potential of the form

$$V(\phi) = \frac{1}{2}\mu^2\phi^2 + \frac{1}{4}\lambda\phi^4. \quad (2.4)$$

There are two constraints to the constants of this potential. First  $\lambda > 0$  is required to guarantee a finite minimum of  $V$ . Secondly, the potential has to allow for local symmetry breaking achieved by  $\mu^2 < 0$ . For the purpose of a descriptive representation it is expedient to examine the *Mexican hat* shape of the resulting potential provided via figure 2.6. The overall potential is symmetrical until either of the

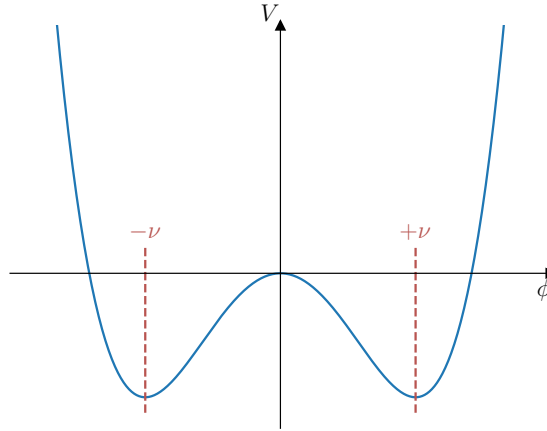


Figure 2.6: Display of a 2-dimensional mexican hat potential. The vacuum expectation values are denoted by  $\pm\nu$ .

minima  $\pm\nu = \sqrt{\frac{\mu^2}{\lambda}}$  is chosen at which point the symmetry in that minimum is broken. These minima can be interpreted as the vacuum expectation values. In the subsequent step perturbations of the field around a minimum can be used to obtain the excitations of the field representing the particle states. In the simplest case of a one-dimensional field a Lagrangian for a massive fermion is obtained. However, this is insufficient to describe the masses of the gauge bosons. Additionally, it gives no reason to expect a Higgs boson, an excitation of the field itself. Consecutively the same formalism can be used on a complex scalar field yielding not only the terms of a massive fermion but also a massless scalar, a Goldstone boson introducing an additional degree of freedom. Thereby, the formalism to obtain terms for massive particles is defined and only the associated field is missing. Requiring local gauge symmetry grants the required field term. A combination of an introduction of a new gauge field and the perturbation at a chosen minimum of the potential then gives rise to both a massive gauge field and a massless field term. Equation 2.5 portraits all resulting terms

$$\mathcal{L} = \underbrace{\frac{1}{2} (\partial_\mu \eta) (\partial^\mu \eta)}_{\text{massive } \eta} - \lambda \nu^2 \eta^2 + \underbrace{\frac{1}{2} (\partial_\mu \xi) (\partial^\mu \xi)}_{\text{massless } \xi} + \underbrace{-\frac{1}{4} F^{\mu\nu} F_{\mu\nu}}_{\text{massive gauge field}} + \underbrace{V_{\text{int}}}_{\text{interaction terms}} + \underbrace{g\nu B_\mu (\partial^\mu \xi)}_{\text{B-}\xi \text{ coupling}}. \quad (2.5)$$

Both a massive term, interpreted as the Higgs boson and a massive gauge field are present here. The massless field, called the Goldstone boson, and the coupling term can be eliminated by an appropriate gauge transformation. Thus, all necessary terms are obtained because the elimination of the Goldstone boson opens up an additional degree of freedom that can be “eaten” by the longitudinal degree of freedom of the massive gauge boson. This example explains the method to obtain all necessary terms but only works in the  $U(1)$  symmetry group. The Higgs according to the Standard Model is embedded

in the  $U(1) \times SU(2)$  symmetry of the electroweak interaction. This requirement can be fulfilled by a pair of complex scalar fields following the same methods and the potential [5, 12]:

$$V(\phi) = \mu^2 \phi^\dagger \phi + \frac{\lambda^2}{2} (\phi^\dagger \phi)^2. \quad (2.6)$$

The result is a field that permeates everything and the interaction with which lets particles acquire mass. The coupling strength to the field is likely to be proportional to the mass but is one of the parameters of the model that has to be measured experimentally. In general a plethora of parameters stems from the Higgs mechanism alone. The mechanism itself is described by its vacuum expectation value  $\nu$  and the mass of the Higgs boson  $m_H$ . Additionally, the masses of the fermions are described by their Yukawa couplings:

$$g_f = \sqrt{2} \frac{m_f}{\nu}. \quad (2.7)$$

Lastly the masses of the gauge bosons are defined as follows:

$$m_\gamma = 0, \quad (2.8)$$

$$m_W = \frac{1}{2} g_W \nu, \quad (2.9)$$

$$m_Z = \frac{1}{2} \nu \sqrt{g_W^2 + g'^2}. \quad (2.10)$$

The properties of the Standard Model Higgs boson are provided in section 2.3.1.

### 2.1.5 Open questions in the Standard Model

It is important to stress that the Standard Model neither is nor claims to be a complete representation of our universe. It is a model allowing for astonishingly good predictions of various aspects of our world and it has carried modern scientists through half a century of discoveries. The underlying theories yield a model with 19 free<sup>4</sup> parameters. There are several choices one can make regarding these depending on optimisation for theory or measurement. One possible choice (assuming massless neutrinos) goes as follows [5]:

- 3 coupling constants:  $\alpha$ ,  $\alpha_S$  and  $G_F$ .
- 4 parameters describing the CKM matrix: 3 angles and a CP-violating phase.
- 1 QCD vacuum angle:  $\theta_{\text{QCD}}$ .
- 11 parameters related to the Higgs sector of the model: 9 fermion masses or the respective Yukawa couplings, the mass of the Higgs boson and the vacuum expectation value.

These parameters of the Standard Model are commonly understood as its first imperfection. The quantities have been chosen to match experimental observations but it is generally more desirable to create a theory that allows for calculation of its parameters. However, there is a plethora of more

<sup>4</sup> Technically, the adjective free is redundant with parameter but it aims to emphasise the fact that the quantity cannot be obtained from any theoretical calculations.

profound shortcomings of the model. A renowned example is the neutrino masses. The Standard Model demands massless neutrinos but modern experiments clearly contradict this assumption [13]. Also, there are other open questions such as the origin of dark matter in the universe. From the weaknesses of the model, a broad mass of theories emerges that offer answers to the questions. Experimentally, such theories are defined by the additional predictions they make that can be verified. Commonly this is the prediction of new particles. Since the discovery of the Higgs boson, no new particles have been found and thus no clear step towards a new model could be made. However, it is important to note that precision measurements of the Standard Model and especially those performed over the last decade limit the range of possible alternative theories significantly. The progress resulting from precise measurements of the Standard Model parameters is often overlooked or even misunderstood as a scientific chore. It is therefore important to stress the gravity of parameter estimations like the one presented in this thesis because they pave the way to new experiments destined for groundbreaking discovery. In a similar fashion scientists were sure to find the nature of the mass generating mechanism with the LHC one way or another [14].

If one directs one's attention to the list of parameters the Higgs sector involves a majority of our present ignorance and is for that reason the perfect point for precision measurements. In addition, the properties of the Higgs boson have not been measured to the same standard of precision as many other particles in the model which makes for a plethora of interesting analyses. Section 2.3 describes the promises that an analysis of the associated production of a  $t$ -quark and a Higgs boson in particular makes.

## 2.2 Physics of proton collisions

For the purpose of understanding the process of proton-proton collisions, a short digression into the structure of protons and its consequence is necessary. Both the baryon number and the electric charge are conservation laws. Therefore, in the collision of two protons only final states with two baryons and an electric charge of 2 should be viable. The fact that at the LHC a multiplicity of intuitively forbidden states occurs is due to the substructure of the nucleons. As a consequence of the observed cross sections for deep inelastic electron-proton scattering it was proposed that the proton is made of point-like constituents, termed partons and later identified as quarks and gluons [11]. The observed relative freedom of the partons in nucleons is justified by the asymptotic freedom stemming from the running of the strong coupling constant. In the model each parton carries some fraction of the proton's momentum. The partons responsible for the proton's quantum numbers are named valence quarks while the sea of other quarks appears in quark-antiquark pairs only and for this reason does not contribute to the nucleon's quantum numbers. In collisions however, the hard scattering can happen between any of the protons' constituents and the conservation laws are bypassed.<sup>5</sup> Thereby, hadron scatterings give access to a wide range of final states. A significant drawback is the unknown fraction of momentum carried into the hard scattering which puts limits on the kinematic information and on the overall energy available in a single scattering. For the energy scale of nuclear processes one can assume that roughly half the momentum of the proton is carried by the sea quarks but for increasing energies the fraction increases and the sea quarks become the dominant particles in interactions [15].

---

<sup>5</sup> Of course the conservation laws are not broken for the system but a single hard scattering is not affected by the quantum numbers of the initial protons.

## 2.3 The associated production of a $t$ -quark and a Higgs boson

The target process of this analysis is the associated production of a Higgs boson and a  $t$ -quark. A  $t$ -channel Feynman diagram of this process, from now on termed  $tHq$ , in the 4-flavour<sup>6</sup> scheme is depicted in figure 2.7. For simplification, the 4-flavour scheme is omitted for the following Feynman diagrams. The final state is characterised by two heavy objects, the  $t$ -quark and the Higgs boson, that are recoiling and an additional light jet in the forward<sup>7</sup> region of the event's topology. This analysis focuses on the channel in which the Higgs boson decays to a pair of  $\tau$  leptons, the Feynman diagram for which is shown in figure 6.1. As a result, the final state contains signatures of the Higgs boson and of the heaviest quark and lepton each. These particularly massive particles take over a special role for several reasons. On the one hand, the high mass allows for a large number of decay channels. On the other hand, only a few particles can therefore decay into these heavy particles. Additionally, the high mass results in distinguished signatures in detectors. To clarify the expected signatures, the important components of the final state are briefly summarised. Finally, the foremost aim of this chapter, the predictive power of the channel, is investigated.

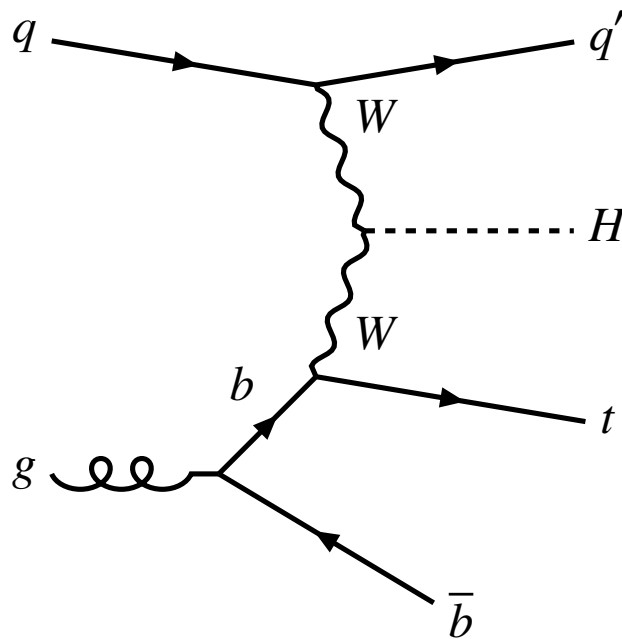


Figure 2.7:  $tHq$   $t$ -channel process in the 4-flavour scheme.

<sup>6</sup> For modelling purposes, it is assumed that the colliding protons consist only of the four lightest flavours of quarks. Consequently, the  $b$ -quark enters the collision from a gluon decay. The resulting  $\bar{b}$ -quark is assumed to be carried away in beam direction.

<sup>7</sup> Forward refers to the direction of the colliding particles.

### 2.3.1 Higgs boson physics

The Higgs boson is the central object in the  $tHq$ -process. It is a neutral scalar that couples to all fermions with coupling strength proportional to their mass and its own mass is a free parameter of the model. For this reason, the measurement of its cross sections provides an opportunity to measure the respective Yukawa couplings.

The dominant production mode of the Higgs boson at proton colliders is gluon fusion depicted in figure 2.8(a). Further production modes are the vector boson fusion (2.8(b)) and the production in association with either a vector boson (2.8(c)) or a  $t$ -quark pair (2.8(d) to 2.8(f)).

The Higgs boson couples to massive particles through the Higgs field and to massless particles via loops involving heavy particles. The possible decay modes are depicted in figure 2.9. Generally the Higgs boson tends to decay into heavier particles. The corresponding measured branching ratios are summarised in table 2.1. In a collider experiment an analysis commonly targets a specific decay

Table 2.1: List of the dominant decay modes of the Higgs boson and their respective measured branching ratios [3].

Decay mode	Branching ratio [%]
$H \rightarrow b\bar{b}$	$53.0 \pm 8.0$
$H \rightarrow WW$	$25.7 \pm 2.5$
$H \rightarrow \tau\tau$	$6.0 \pm 0.8$
$H \rightarrow ZZ$	$2.8 \pm 0.3$

channel or a related group of channels in order to efficiently isolate the events. The isolation of a signal is the first limiting factor of an analysis, the second being the magnitude of the cross section. The decay width is dominated by the  $H \rightarrow b\bar{b}$  mode with 53 %, followed by  $H \rightarrow WW$  with 25.7 %. The presented analysis targets the third highest and dominant leptonic decay  $H \rightarrow \tau\tau$  with a branching fraction of 6 %.<sup>8</sup> The high branching fraction comes with the downside of the hadronic decay modes of the  $\tau$  lepton resulting in a large pollution via reducible and irreducible backgrounds. An introduction to the dominant background processes is provided via section 2.4.

### 2.3.2 Top physics

The  $t$ -quark is the most massive fundamental particle in the Standard Model. Its mass of  $(172.69 \pm 0.30)$  GeV [3] results in a lifetime shorter than the scale of non perturbative QCD. Consequently, the  $t$ -quark forms no bound states and uniquely decays in its bare state.

The Standard Model  $t$ -quark decays via the weak force into a  $W$  and a lighter quark.<sup>9</sup> Since the CKM matrix is almost diagonal,  $|V_{tb}| \gg |V_{ts}| > |V_{td}|$ , the decay to a  $b$ -quark is dominant with a relative branching ratio measured to  $\Gamma(Wb)/\Gamma(Wq) = 0.957 \pm 0.034$  [3]. Another effect of the high mass is that the  $W$  boson in the decay can be on-shell. Figure 2.10 shows the corresponding Feynman diagram for the decay. A consequence of particular relevance for this analysis is that the Standard Model Higgs boson cannot decay directly into a pair of  $t$ -quarks due to energy conservation. As a

<sup>8</sup> While  $H \rightarrow b\bar{b}$  and  $H \rightarrow WW$  have higher cross sections they come with a downside of high missing energy and hard neutrinos respectively.

<sup>9</sup> Some theories beyond the Standard Model allow different decay modes for the  $t$ -quark



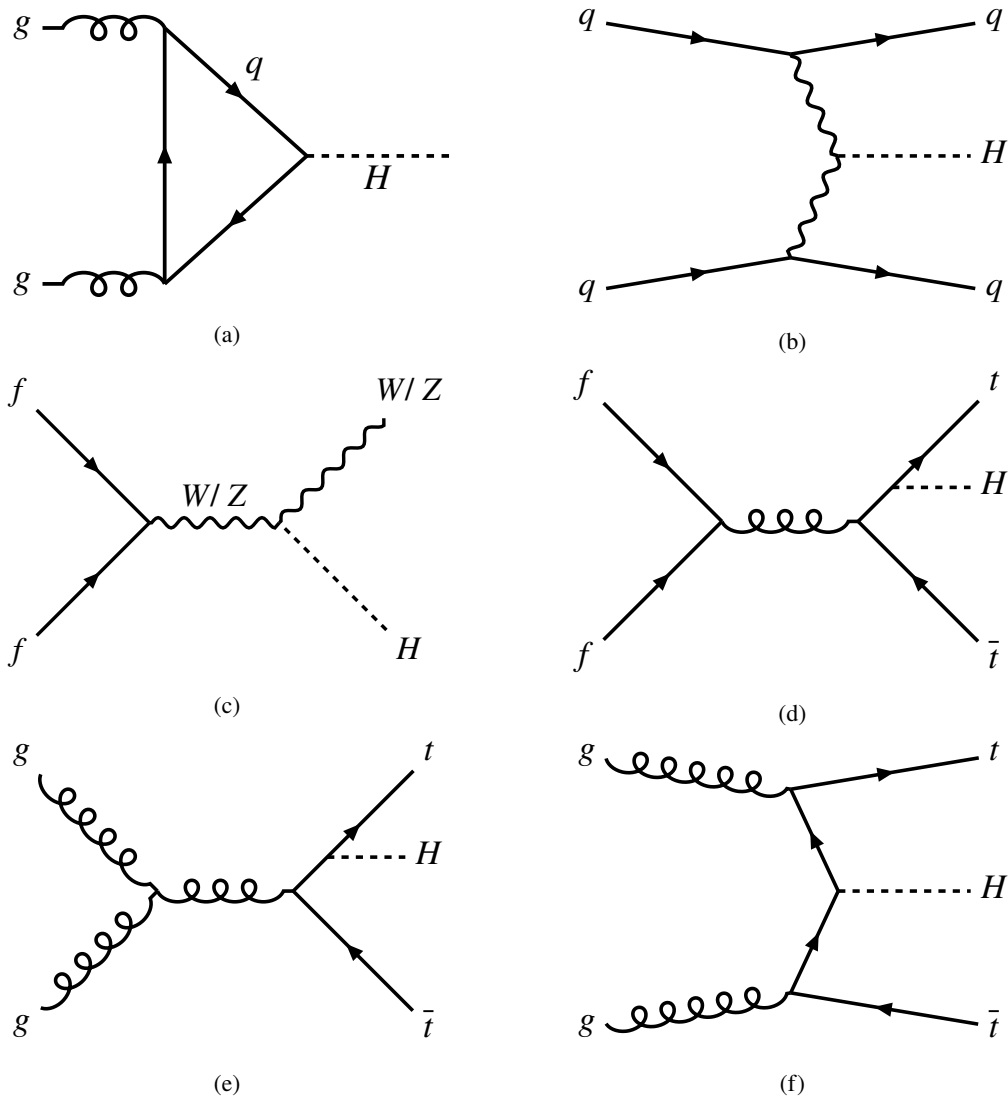


Figure 2.8: Higgs production modes at proton colliders: Gluon fusion (a), Vector boson fusion (b), associated production with a vector boson (c) and associated production with a heavy quark (d), (e), (f) [12].

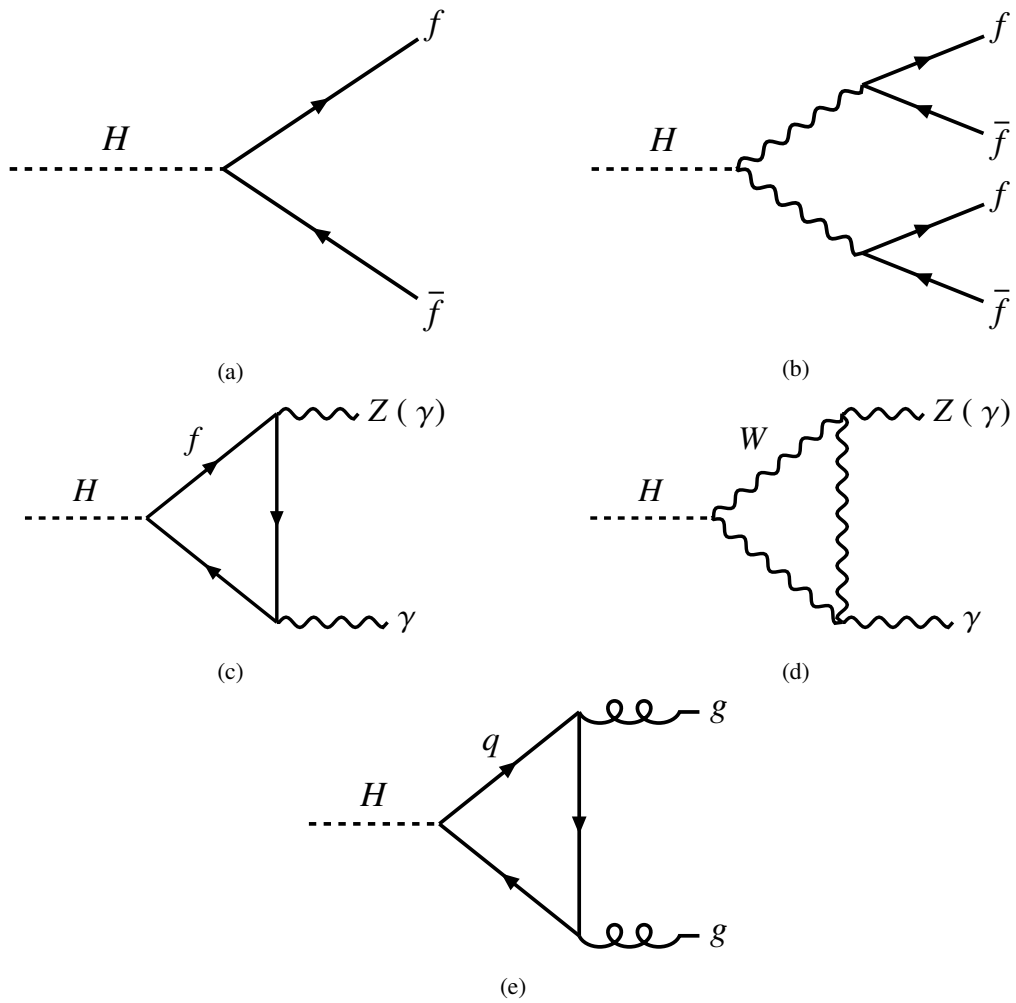


Figure 2.9: Decay modes of the Standard Model Higgs boson [12].

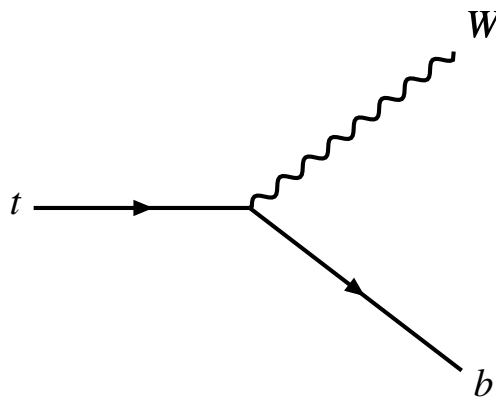


Figure 2.10: Feynman diagram of the decay of a  $t$ -quark.

consequence,  $y_t$  is the only Yukawa coupling that has to be measured indirectly. The analysis presented here is one process that allows for this measurement is explained in detail later.

### 2.3.3 Tau physics

The  $\tau$  lepton is the heaviest charged lepton with its relatively high mass of  $(1\,776.86 \pm 0.12)\text{ MeV}$ . This, as a consequence, means a strong coupling to the Higgs boson and accordingly the highest leptonic decay fraction of 6% for the  $H \rightarrow \tau\tau$  decay. Additionally, its mass makes the  $\tau$  the only lepton that can decay hadronically. The leptonic and hadronic decays are depicted in figure 2.11. The combined branching ratio of the leptonic decays makes up about one third of the decay width. The other two thirds are hadronic decay modes commonly classified by their number of charged decay particles, called ‘‘prongness’’. One-prong decays and three-prong decays account for 72% and 22%, respectively.

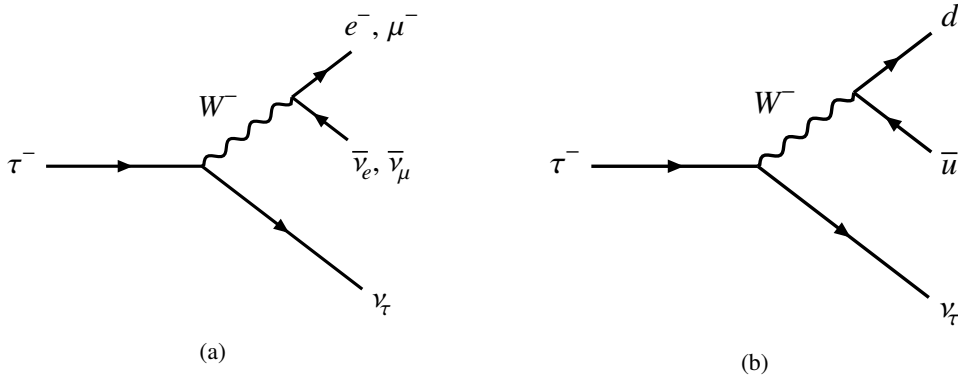


Figure 2.11: Feynman diagrams of a leptonic (a) and hadronic (b) decay of a tau lepton.

### 2.3.4 Predictive power of the analysis

With the particularities of the associated production of a  $t$ -quark and a Higgs boson,  $tH$ , in proton colliders out of the way, this section commences to derive its predictive power. It is explained what makes the process a compelling analysis and which information can be extracted from the measurement of its cross section.

The  $tH$  processes are sensitive to the  $t$ -quark Yukawa coupling  $y_t$ , the Higgs to vector boson coupling  $g_{HVV}$  and to the sign and magnitude of the relative phase of the two couplings. A sensitivity to the couplings alone makes for an interesting process because both are free parameters of the Standard Model. The  $t$ -quark Yukawa coupling as mentioned before is particularly interesting because measurements of it have to rely purely on the associated production. However, there are other processes such as  $\bar{t}tH$  that show the same sensitivity to the couplings [16, 17], and processes such as  $H \rightarrow \gamma\gamma$  that are sensitive to both sign and magnitude of the coupling [18]. Loop processes like  $H \rightarrow \gamma\gamma$  (depicted in figure 2.9(c)) rely on the knowledge of all particles contributing to the loop. For this reason, they are only sensitive to the sign of the coupling under the assumption of solely Standard Model particles contributing to the loop. This is not true for  $tH$  which is model-agnostic. To give an example of a model that suggests a negative sign of the coupling one can consider the Standard

Model Higgs mechanism. The described Higgs mechanism is the minimal model necessary to explain the current observations. There are other possible models, such as the two-Higgs doublet model, involving more than one massive excitation of the Higgs field. The modified interactions expected from these models would be visible in the sign of the coupling but also give rise to flavour changing neutral currents [19, 20]. In conclusion, the measurement of the  $tH$  cross section allows for not only  $y_t$  and  $g_{HVV}$  but also uniquely the relative sign of the two to be probed, even in the presence of [Beyond Standard Model \(BSM\)](#) physics. In the following, the origin of these sensitivities is described in detail.

There are three groups of diagrams which contribute to  $tH$ , in each of which the Higgs boson can be radiated either from the  $t$ -quark or the  $W$  boson: t-channel, s-channel and  $tWH$ . A selection of Feynman diagrams is shown in figures 2.12 to 2.14 for each channel. In all cases the two possible Higgs boson origins are depicted. Consequently, the cross section depends either on the  $t$ -quark Yukawa coupling  $y_t$  or the Higgs to vector boson coupling  $g_{HVV}$ . For clarification, figure 2.15 labels the vertex associated to the couplings for the t-channel. This alone gives rise to the sensitivity to the couplings. The unique selling point though was the advertised sensitivity to sign and magnitude of  $y_t$ .

The key to its understanding is the interference of the processes in which the Higgs couples to either of the  $t$ -quark or the  $W$  boson. The resulting interference term in the matrix element depends on sign and magnitude of the phase between the respective couplings.<sup>10</sup> Current measurements favour a positive sign of  $y_t$ . If this was true, it would lead to negative interference between the processes, and subsequently a small cross section of the  $tH$  processes. A negative sign of the coupling, on the contrary, would result in constructive interference and a significant increase in the cross section. Thereby, the result of the presented analysis can be a strong guiding factor for the acceptance of [BSM](#) theories by design.

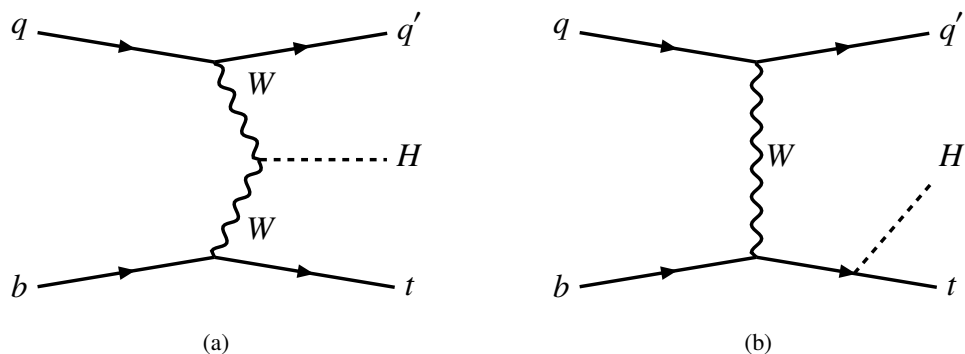


Figure 2.12: Feynman diagrams of the  $tHq$  t-channel decay with the Higgs coupling to the  $W$  (a) and the  $t$ -quark (b).

## 2.4 Background processes

A large part of this thesis deals with the separation of signal events from background events. The  $tH$  signal is described in full detail in section 2.3. For a deeper understanding of the events, it is necessary to comprehend the underlying processes and their expected signatures in the detector. This

<sup>10</sup> Since the analysis can only describe the relative sign either of the sign can be fixed to a positive value which usually is done for the Higgs to boson coupling. Thereby, the sign of  $y_t$  becomes the observable of interest.

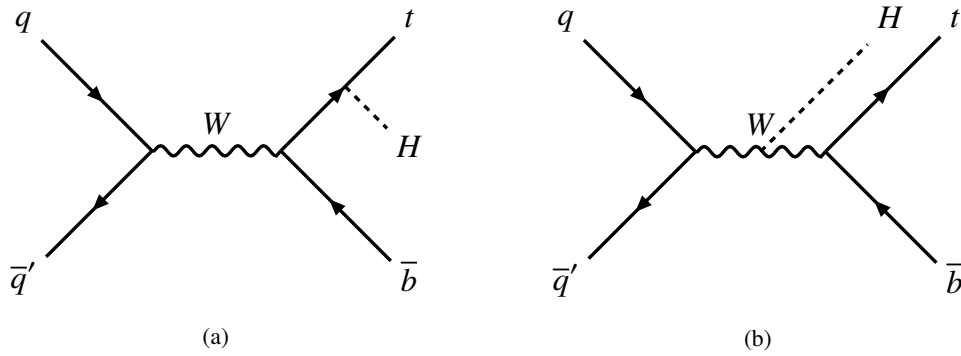


Figure 2.13: Feynman diagrams of the  $tHq$  s-channel decay with the Higgs coupling to the  $W$  (a) and the  $t$ -quark (b).

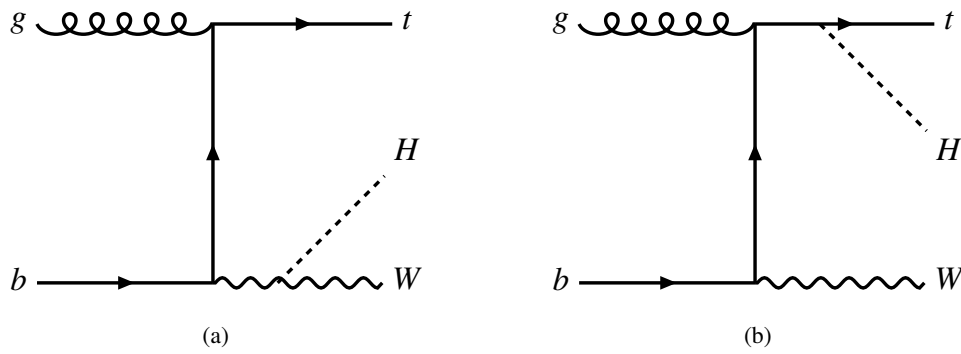


Figure 2.14: Feynman diagrams of the  $tWH$  t-channel decay with the Higgs coupling to the  $W$  (a) and the  $t$ -quark (b).

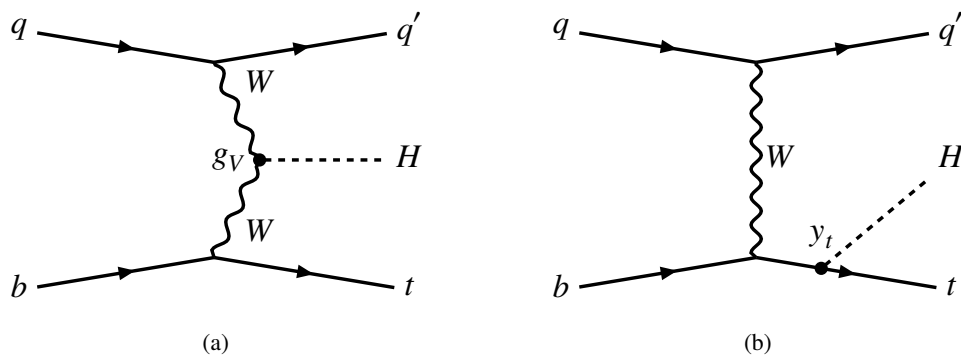


Figure 2.15: Feynman diagrams of the  $tHq$  t-channel decay with the Higgs coupling to the  $W$  (a) and the  $t$ -quark (b).

section provides the fundamental knowledge to understand the results of this thesis. In contrast to the other sections of this chapter, the description is best read in combination with chapters 6 and 7. The descriptions are limited to the practically relevant parts and provide a less theoretical and more concise introduction.

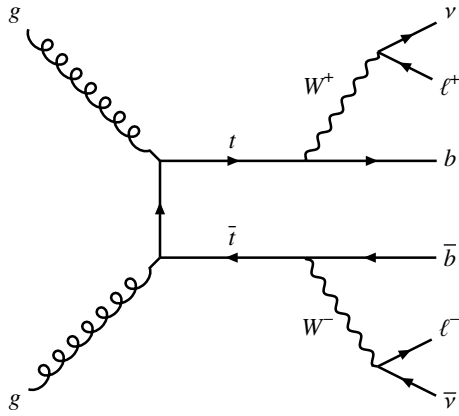
From the perspective of this analysis it is feasible to characterise processes by their final states. That means that processes can be grouped via the objects that are visible in the detector. While there is interference between processes that share the same final state, similar processes are also significantly harder to distinguish in the data. In order to understand specific contributions, it is essential to distinguish between leptonic and hadronic contributions to the final state. The following steps, which are based on the event selection in section 6.1, are then suitable for developing an understanding of a process. A good starting point is the comparison of the expected leptonic signatures' charges with the selection. Thereby, an understanding why a certain background appears at a certain rate can usually be gained. Additionally, the presented event selections require the presence of two characteristic hadronic signatures: *b*-jets and  $\tau_{\text{had}}$ . The decay of a *t* quark results in the presence of a *b*-jet.  $\tau_{\text{had}}$  can be the product of decay chains or a hadronic signature can be misidentified as a  $\tau_{\text{had}}$ . The more likely any of these criteria becomes for a given channel, the larger the contribution.

More generally, one differentiates between reducible and irreducible backgrounds. Irreducible backgrounds contain all or more objects of the signal process at LO. For this reason a separation of those backgrounds becomes severely more challenging because the signature can be fully imitated. In this case differences in the kinematics or smart decision rules can limit the contribution or create regions better for analysis. It is important to remember that a full separation is physically impossible. A good example is the *tZq* process shown in figure 2.16(f). The decay of the *Z* boson can mimic the decay of the *H* boson. An adequate feature to distinguish from this channel is the reconstruction of the boson's mass. This process is described in section 6.2. This necessity to exploit the kinematic feature of the final state in order to isolate the signal is one of the motivations for the application of the [neural network \(NN\)](#) as described in chapter 7.

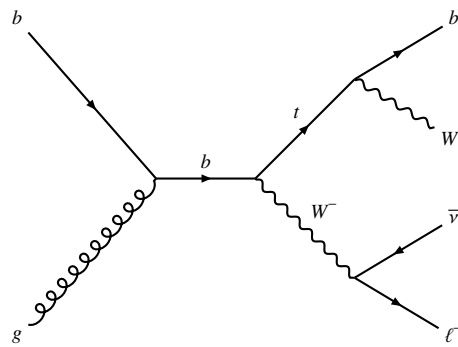
In contrast, reducible backgrounds have final states that differ from the signal and their contribution therefore should be reducible by adequate selection criteria. One of the limiting factors for the isolation from these backgrounds is the reconstruction and identification of objects in the detector. Generally, before an analysis a preselection (see section 6.1) is chosen which should reduce the contribution from reducible backgrounds significantly. The remaining contributions can largely be attributed to misidentified objects in the final state. Since the identification for leptons is superior to hadronically decaying objects, the majority of irreducible background stems from such processes. In addition, processes play a role whose leptonic part is very similar to that of the signal. In the case of this analysis, the reducible backgrounds also have a higher cross section than both the signal and most irreducible backgrounds, which further increases their contributions. The dominant reducible backgrounds are  $t\bar{t}$ , *Z*+jets and *W*+jets. Figure 2.16(a) shows the  $t\bar{t}$  process. It is a dominant or large contribution to each of the covered channels. The hadronic signatures of the *t*-quarks contain *b*-jets and can contain true or fake signatures of light leptons and  $\tau_{\text{had}}$ . *Z*+jets, depicted in figure 2.16(h) is specifically large in the  $2\ell$  OS +  $1\tau_{\text{had}}$  channel because the decay of the *Z* can result in two opposite sign leptons. This also makes it a significant contribution to the  $1\ell + 2\tau_{\text{had}}$  channel together with *W*+jets depicted in figure 2.16(g). Similarly, due to the charge of the leptons *W*+jets is more important in the  $2\ell$  SS +  $1\tau_{\text{had}}$  channel.  $2\ell$  SS +  $1\tau_{\text{had}}$  is special in so far that the dominant contributions come from irreducible backgrounds, namely the  $t\bar{t}$ +*X* processes, depicted in figures 2.16(d) and 2.16(e). Additional examples of background processes are provided for *tW* in figure 2.16(b), for single *t*-quark

processes in figure 2.16(c) and lastly for diboson processes in figures 2.16(i) and 2.16(j).

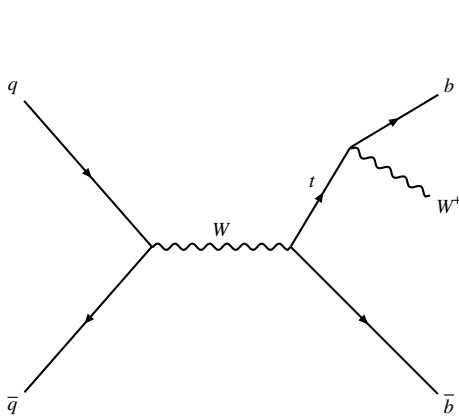
In practice, both categories of backgrounds have significant contributions. In consequence, it is essential that the amount of background events is correctly modelled in order to understand the expected separation and remaining contributions. Section 6.1 documents the contributions of the background processes to the selected region. For an in-depth explanation of backgrounds for Higgs boson final states in general purpose detectors, see [12].



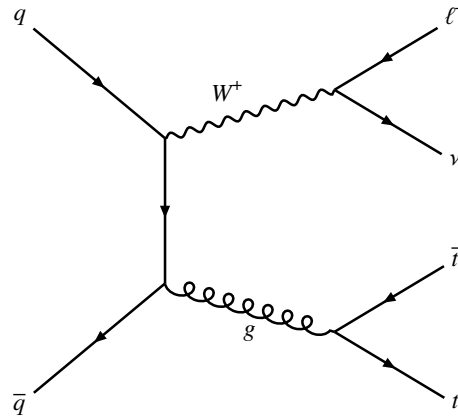
(a) Feynman diagram of the  $t\bar{t}$  process.



(b) Feynman diagram of the  $tW$  process.



(c) Feynman diagram of the production of a single  $t$ -quark in the s-channel.



(d) Feynman diagram of the  $t\bar{t}W$  process.

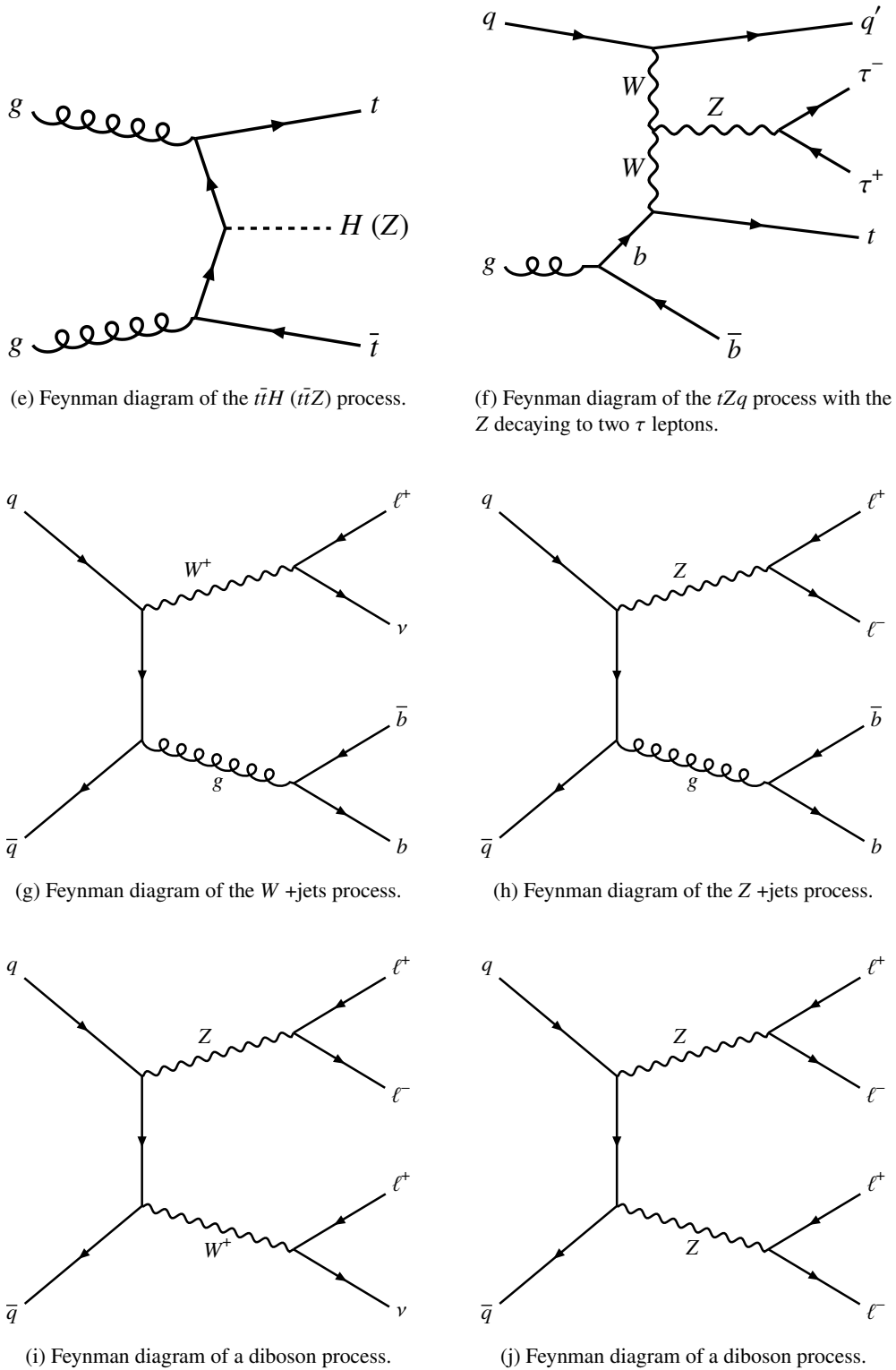


Figure 2.16: Feynman diagrams of important background processes.



---

## Particle physics experiments

---

For most research in modern particle physics there are two predominant constraints. The first one arises from the statistical nature of decay and creation processes in particle physics. Many of the most interesting events occur extremely rarely and require large amounts of data, or more precisely high integrated luminosity, to achieve statistically significant results. Secondly, the energy supplied has to be sufficiently high to provide the necessary creation energies. An essential component to understanding analyses in particle physics are the choices made in the experimental approach. There are different strategies regarding particle production and detection, resulting in critically different experimental setups. Each method comes with its own focus and specific limitations. Understanding the aim of the experiment and its components is mandatory to comprehend its results. This chapter introduces the underlying concepts of particle acceleration and detection. Firstly, the production of particles via acceleration and collision is described leading to a description of the [Large Hadron Collider \(LHC\)](#). Following this up, the primary methods of particle detection and their underlying interactions are explained. The [ATLAS](#) detector and its components are introduced. Lastly, the details most important to the analysis presented in this thesis are summarised.

A very comprehensible guide to the technologies utilised in the field of high energy particle physics is provided in [15]. For a more in-depth source, see [21].

### 3.1 Particle accelerators

It is a justifiable simplification to describe the experimental setup as consisting of two components, the particle accelerator and the detector with an appropriate choice of either being crucial to the whole setup. Before the detector is explained in sections 3.2 and 3.3, this section introduces the concepts of particle acceleration and discusses the technical details of the [LHC](#).

In order to measure the properties of particles or even discover new ones, the energy for their production needs to be provided at a sufficiently high rate. This becomes a limiting factor for particles that either have a very high mass or a minuscule production cross section respectively. Consequently, the result is a steadily increasing demand in both luminosity and centre-of-mass energy for particle physics experiments. Naturally, the technical requirements grow with both quantities. These requirements are specified in this chapter and used to motivate the design decisions for particle accelerators.

One distinguishes collider experiments utilising a fixed target or colliding beams. While fixed target experiments offer some unique advantages like the limitation of the kinematic region and a free choice of target material, colliding beam experiments are better suited for achieving high centre-of-mass energies. The LHC is a colliding beam experiment and thus this section focusses on this kind of setup. Colliding beam experiments are usually set up as storage rings allowing for efficient collision. However, it also poses the challenges of synchrotron radiation and bending the particles' tracks.

Synchrotron radiation is the emission of electromagnetic radiation by charged particles moving in a curved path under the influence of magnetic fields, typically produced in synchrotron accelerators. The energy lost by a particle of charge  $q$  and Lorentz factor  $\gamma$  in a storage ring of radius  $R$  is its synchrotron radiation and given by:

$$\Delta E \approx \frac{1}{3\epsilon_0} \frac{q^2}{R} \gamma^4 \propto \frac{1}{R} \frac{E^4}{m^4}. \quad (3.1)$$

Here  $\epsilon_0$  denotes the vacuum permittivity. The formula makes apparent that this energy loss becomes more costly for higher energetic particles and due to the higher power of the energy cannot be mitigated effectively by the ring's radius. Having said that, an increase in the accelerated particle's mass makes the effect less dominant. Thus, for heavier particles the magnetic bending field becomes the limiting factor.

Apart from the particle's mass the type of particle has a defining influence on the experiment's potential. Lepton colliders are more limited in final states but also need a factor 6 less energy for the same discovery potential<sup>1</sup>. Nonetheless, for general analysis, the reduced demand in energy and the access to strongly interacting particles makes hadron colliders more suitable for discovery. It is worth noting, that the interaction of the partons generally leads to a wider range in outcomes. Lepton colliders on the other hand are basically considered precision experiments.

Before talking any further about collider physics, it is instructive to outline the quantities that classify the named discovery potential and its limitations. While the maximum available energy in an experiment is described by the centre-of-mass energy, the second key performance indicator of a collider is its instantaneous luminosity  $\mathcal{L}$ . It has the dimension of events per time and area, thereby quantifying the rate of particle collisions. For a colliding beam experiment it is defined by

$$\mathcal{L} = \frac{f n_1 n_2}{4\pi \sigma_x \sigma_y}, \quad (3.2)$$

where  $f$  is the beam crossing frequency,  $n_i$  the number of particles per beam and  $\sigma_i$  the Gaussian beam sizes in both directions. The luminosity is linked to the event rate of a process in the experiment via another quantity, the cross section  $\sigma$ , following:

$$\frac{dN}{dt} = \mathcal{L} \cdot \sigma. \quad (3.3)$$

The total number of events for a process is then defined by:

$$N = \sigma \cdot \int \mathcal{L} dt = \sigma \mathcal{L}_{int}. \quad (3.4)$$

---

<sup>1</sup> The factor of six is a rule of thumb and uses the fact that leptons are elementary particle. In contrast, a collision of protons includes three valence quarks and a sea of gluons instead.

In practice, it is more effective to use the integrated luminosity,  $\mathcal{L}_{int}$ , since it can be measured more precisely via channels with well known cross sections.

To complete the picture, a particle's decay rate needs to be defined. Knowing the frequency and probability of a certain process to occur does not suffice. Instead, most particles in the initial process are short-lived and decay into particles that then form the measurable object in the event. The associated decay rate is defined as the inverse of the lifetime:

$$\tau = \frac{1}{\Gamma_{tot}}. \quad (3.5)$$

The subscript “tot” denotes the total decay rate. It summarises the decay rates for all possible decay modes of a particle, each denoted by their individual decay rate  $\Gamma_i$ . This means that  $\Gamma_{tot}$  is given by

$$\Gamma_{tot} = \sum_{i=1}^n \Gamma_i, \quad (3.6)$$

while the branching fraction denoting the share of a certain decay mode is defined as

$$\mathcal{B}_i = \frac{\Gamma_i}{\Gamma_{tot}}. \quad (3.7)$$

### 3.1.1 The Large Hadron Collider

The Large Hadron Collider, located at the facilities of the European Organization of Nuclear Research (CERN) close to Geneva, was built to extend the frontiers of modern particle physics by delivering unrivalled luminosities and reaching unprecedented high energies, thereby providing data benefiting multiple particle physics experiments.

The **LHC** is a circular particle collider with a circumference of 26.7 km designed to accelerate and collide two counter-rotating proton beams. The protons are accelerated in bunches of up to  $1.2 \times 10^{11}$  protons, at energies up to 6.5 TeV in run 2. This way, the outstanding luminosity of  $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  and centre-of-mass energy of up to 13 TeV is achieved. The bunches are obtained by stripping electrons from hydrogen atoms before injecting them into an accelerator complex. In the accelerator complex the bunches are successively accelerated before being inserted in the final storage ring. An overview of the acceleration system is given in figure 3.2. In the storage ring dipole magnets keep the protons on the circular trajectory, while quadrupole magnets bring the beams to collision. For more detailed information one can refer to the **LHC** design report [22]. The four main interaction points at which the beams are brought to collision contain the main experiments of the **LHC**. Two of them are general purpose detectors, namely **ATLAS** [23] and **CMS** [24]. That means they are designed to cover a wide range of final states rather than focusing on a single analysis. The third is the **LHCb** [25] which focuses on bottom quark physics. Lastly, **ATLAS** [26] is designed for the investigation of heavy ion collisions. Figure 3.1 shows a sketch of the **LHC**'s location and the positions of the four main experiments.

A far-reaching consequence of the high luminosity is the occurrence of multiple interactions per bunch crossing. The resulting events are referred to as **pile-up (PU)** events. There are two types of **PU**, in-time and out-of-time. In-time **PU** describes additional events arising from the same bunch crossing but not the hard scatter vertex. Out-of-time **PU** summarises remnant signatures from preceding or subsequent bunch crossings. **PU** is a polluting contribution to the analysis complicating reconstruction

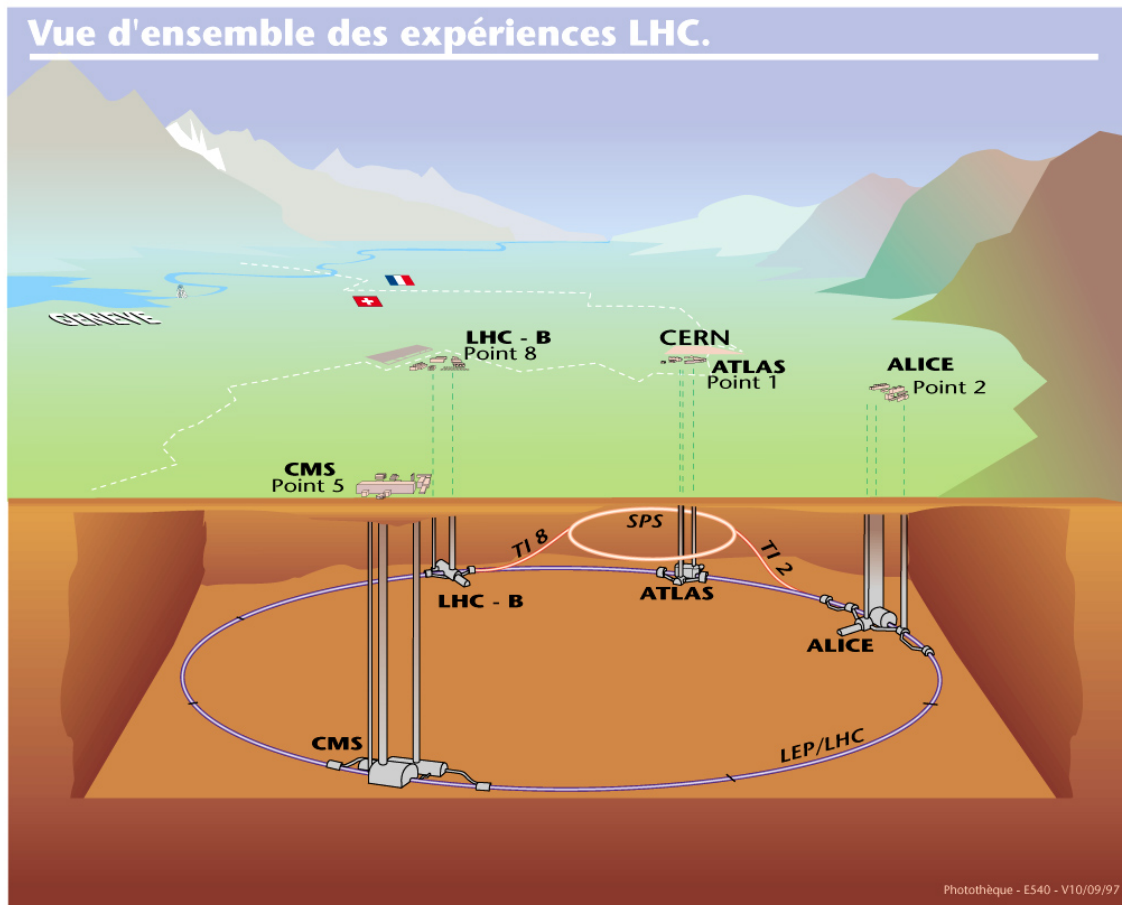


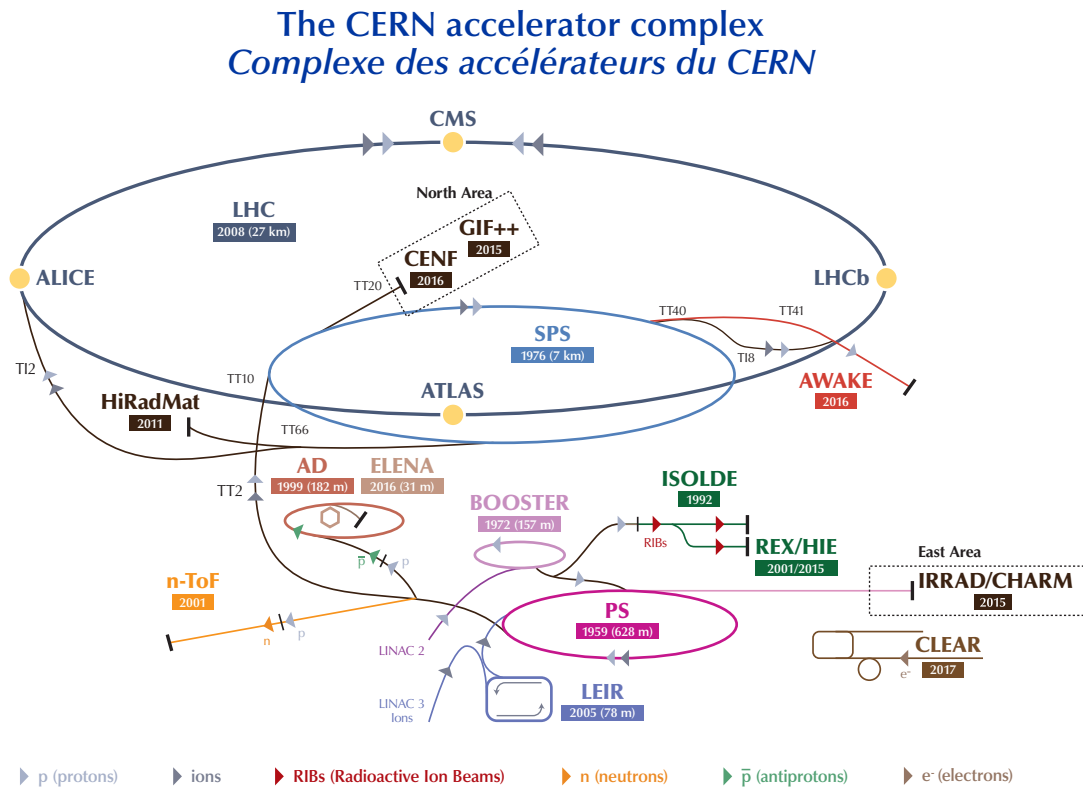
Figure 3.1: Sketch of the LHC ring, the position of the experiments, and the surrounding countryside. The four big LHC experiments are indicated (ATLAS, CMS, LHC-B and ALICE) along with their injection lines [27].

of objects particularly. A common way to mitigate it is to isolate it by taking the comparatively low energies of its constituents into account. Additionally, an excellent reconstruction of the point of collision can deliver criteria to extract PU contributions.

### 3.2 Particle detectors

The development of particle detectors over the last century represents a mesmerising field. It gave birth to a gigantic range of different concepts, each making measurable what we cannot see or feel. Ranging from specialised detectors, probing specific particles or interactions to the concept of general purpose detectors, the noble goal of measuring a wide range of final states. While the exact shape depends on the accelerator, a general purpose detector typically consists of three layers increasing in thickness. The first layer contains tracking detectors and is kept thin<sup>2</sup> to reduce energy loss. The second layer is formed by calorimeters, thick enough to contain the full shower of a particle decay.

<sup>2</sup> While this can result in spatially thin components, thin refers to a low stopping power of the detector.



LHC - Large Hadron Collider // SPS - Super Proton Synchrotron // PS - Proton Synchrotron // AD - Antiproton Decelerator // CLEAR - CERN Linear Electron Accelerator for Research // AWAKE - Advanced WAKEfield Experiment // ISOLDE - Isotope Separator OnLine // REX/HIE - Radioactive EXperiment/High Intensity and Energy ISOLDE // LEIR - Low Energy Ion Ring // LINAC - LINear ACcelerator // n-ToF - Neutrons Time Of Flight // HiRadMat - High-Radiation to Materials // CHARM - Cern High energy AcceleraTOR Mixed field facility // IRRAD - proton IRRADiation facility // GIF++ - Gamma Irradiation Facility // CENF - CERN Neutrino platForm

Figure 3.2: The LHC is the last link in a chain of particle accelerators. The figure provides a summary of the pre-acceleration chain. Both the path for protons and ions is depicted. The image shown here reflects the status used for the presented analysis. As of 2020 LINAC2 is replaced by LINAC4 as the initial accelerator [28].

Muon spectrometers commonly form the last layer. In this section the technological concept of each part of a general purpose detector is outlined summarising their basic concepts and limitations. Based on this the [ATLAS](#) detector is explained specifically.

A crucial decision in the construction of a detector system is the choice of its specific technology. There is a wide range of options for each type of detector each coming with its own advantage. In the same manner as the choice between fixed target or colliding particles is made with regard to the aim of the experiment, the choice of detector technology must be made.

Before introducing and motivating the specific components of the [ATLAS](#) detector, this section outlines the notable, crucial approaches and technologies. Thereby, the reader is offered the pivotal talking points to comprehend the ideas and hurdles of the analysis presented.

### 3.2.1 Tracking detectors

The term tracking detector summarises a group of detectors allowing the measurement of trajectory, momentum and (electrical) charge of electrically charged particles. The general principle is the determination of the movement of charged particles, sometimes even allowing for impressive visualisation. There are two main types of tracking detectors: gaseous detectors and semiconductor detectors. The principle is as simple as beautiful and has not changed much since the invention of the Geiger Müller Zählrohr [29]. Ionising radiation traverses a medium creating pairs of charged particles. In an electric field the resulting current becomes a measurable quantity allowing for signal detection or precise measurement in space and time.

In gaseous detectors the signal is twofold. On the one hand the emission of photons from excited atoms can be detected by photomultipliers, on the other hand the resulting charged particles, i.e. electrons and ions, can be guided to electrodes via an electrostatic field. In the latter case the electronic noise becomes the limiting factor for a signal extraction. A sufficiently high electric field can be utilised to increase the charged particle's energy resulting in an ionisation of the gas and raising the signal above noise level. The resulting amplification is the gas gain. It allows the mode of operation to be chosen between a highly quantified or a pure detection measurement. The gas mixture in a gaseous tracker typically consists of a noble gas and a quench gas with the latter stabilising the cascades while the former showing the best properties for ionisation.

Basically, semiconductor trackers operate similarly to gaseous detectors. Traversing ionising radiation creates electrons and electron holes, creating a measurable current between two electrodes. The faster timing and higher granularity as well as the larger number of electron/hole pairs that can be reached with these detectors is worth emphasising.

The momentum measurement in a tracking detector is achieved by bending the traversing particle's trajectory in a magnetic field. A precise extraction of the curvature allows for the calculation utilising the Lorentz force.

### 3.2.2 Calorimeters

A calorimeter in the context of particle physics measures the energy of particles. The customary process is the full absorption of a particle in the detector's material and the subsequent measurement of the energy lost in the absorption. A full absorption of the particle makes the measurement a destructive process which is represented by a particle shower. A particle shower is a cascade of particles originating from inelastic scatterings. These inelastic scatterings are the conceptual fundament to calorimetry in

particle physics. In the sense of expediency one needs to distinguish between electromagnetic showers, for electrons, positrons and photons, and hadron showers, for energetic hadrons. The technical details of the showers allow an understanding of functionality and limitations of the respective calorimeter technology. Additionally, the kinematics of the particles play a role, but for the sake of this thesis a description at the energy of high energy particle colliders will suffice. For this reason, a conceptual introduction is offered, outlining the defining quantities for the calorimeter design.

The characteristic quantity of an electromagnetic shower is the radiation length,  $X_0$ , the length after which an electron's energy has decreased to  $\frac{E}{e}$ . It is defined via

$$\left(\frac{dE}{dx}\right)_{\text{rad}} = -\frac{E}{X_0} \quad (3.8)$$

and is used to quantify the size of the expected shower. It grows inversely proportional to the square of the nuclear charge  $Z$ ,

$$X_0 \propto \frac{1}{Z^2}. \quad (3.9)$$

Furthermore, the absorption for photons is exponential and can be related to the radiation length with the definition of an absorption length

$$\lambda \approx \frac{9}{7}X_0. \quad (3.10)$$

Measuring the shower shape in units of  $X_0$  results in an approximate material independence [21]. For a simplified but effective description of electromagnetic showers three assumptions can be made [30]:

- Only Bremsstrahlung and pair production are taken into account.
- Energy loss due to ionisation is independent of the scale and equal to the critical energy per radiation length.
- Showers can be described in one dimension ignoring multiple scattering.

Utilizing these assumptions the dominant processes become Bremsstrahlung and pair production for  $e^-/e^+$  and photons respectively. Until reaching the critical energy  $E_k$ , after which ionisation becomes dominant, we obtain for the energy loss per length

$$\left.\frac{dE}{dx}\right|_{\text{ion}}(E_k) \approx -\frac{E_k}{X_0}. \quad (3.11)$$

The total number of shower particles,  $N_{\text{tot}}$ , and the length of the shower,  $s_{\text{tot}}$ , become simply

$$N_{\text{tot}} \approx \frac{E_0}{E_k}, \quad (3.12)$$

$$s_{\text{tot}} \approx \frac{E_0}{E_k}X_0. \quad (3.13)$$

If each particle always loses half its energy in every step until  $E_k$  is reached, then for the maximum

number of steps  $t_{\max}$  holds

$$E_k = \frac{E_0}{2^{t_{\max}}}, \quad (3.14)$$

$$t_{\max} = \frac{\ln \frac{E_0}{E_k}}{\ln 2}. \quad (3.15)$$

This results in a linear dependence of particle count on energy and a logarithmic growth of the shower length. It should be noted that this description is based on simplifications and is sufficient for an understanding of the characteristics but not for a simulation in practice.

For hadrons the description of the particle shower cannot be made with corresponding elegance because both the contributions to the shower are manifold and their share is highly statistical. Generally, highly energetic hadrons collide inelastically with a nucleus or a single nucleon. In this collision new particles are created, resulting in a cascade. In the collision the nucleus enters an excited state and via spallation radiates nucleons. These cascades in the nucleus go on until de-excitation or until the particles leave the nucleus. Furthermore, for some of the secondary particles electromagnetic and weak decays can also occur. The whole energy deposition in a hadron calorimeter needs to be separated into the underlying processes:

$$E_{\text{dep}} = (f_{\text{em}} + f_{\text{ion}} + f_n + f_\gamma + f_B) E. \quad (3.16)$$

$f_{\text{em}}$  is the electromagnetic component, while the others are summarized as the hadronic components.  $f_{\text{ion}}$  is the ionization component,  $f_n$  deals with the energy loss of neutrons,  $f_\gamma$  are photons from nuclear reactions and  $f_B$  describes the binding energy. An in-depth analysis of the hadron shower components exceeds the scope of this thesis [21]. It is important though, to establish a measure of length for the hadron shower as well. This is the nuclear absorption length,

$$\lambda_a = \frac{A}{N_a \rho \sigma_{\text{inel}}} \approx 35 \text{ g cm}^{-2} \frac{A^{\frac{1}{3}}}{\rho} \quad (3.17)$$

It is related to the radiation length via

$$\frac{\lambda_a}{X_0} \approx 0.37 Z. \quad (3.18)$$

Due to this relation of the interaction length, hadron calorimeters are commonly much larger than electromagnetic calorimeters. With this in mind, the general concept of shower development, absorption and shape has been established. However, the second important quantity is the resolution of the calorimeter. It depends on the type of detector. Nevertheless, it can be described generally by three components

$$\frac{\sigma_E}{E} = \sqrt{\frac{a^2}{E} + \frac{b^2}{E^2} + c^2} \quad (3.19)$$

The coefficients represent uncertainties as follows:

- $a$  represents statistical fluctuations due to the fluctuations of the particles contributing to the shower. The uncertainty depends on the energy via  $\propto \frac{1}{\sqrt{E}}$ .
- $b$  quantifies the electronic noise.



- $c$  is a constant term, therefore dominating at large energies. It consists of all irregularities originating from calibration, electronics, geometry of the detector or mechanical components. It summarizes the major systematic uncertainties.

The design of a calorimeter always incorporates a part inducing the particle shower, the passive medium, and an active medium allowing a proportionality between signal and energy depositions. These can be realised in two ways. (1) In a homogeneous calorimeter a single medium fulfils both functions. (2) An inhomogeneous, sampling calorimeter, alternates between layers of active and passive media.

The medium needs to allow for a proportionality between signal and energy loss. The active medium usually works via charge collection or light measurement.

It might seem intuitive, that muons or even taus should behave similar to electrons in matter. However, due to the high mass and the short lifetime and hadronic decay channels respectively, this is not the case. Instead taus need to be identified by their signature. The detection of muons is explained next.

### 3.2.3 Muon detectors

Muon detectors can be regarded as a subcategory of tracking detectors. In the context of combined detector systems, like general purpose detectors however, they serve a distinct purpose. The muon is the heavier cousin of the electron resulting in a similar behaviour in tracking detectors. Its heavier mass though, makes it a minimum ionising particle, MIP [21], in the calorimeters. For that reason a muon traverses most detector components, losing only a small fraction of its energy. As a result of this, muons are the only remaining electrically charged particles reaching the outermost detector layer. A secondary outer tracking detector can therefore be understood as a muon detector and allows for precision measurements, track matching and triggering. This becomes especially useful due to the fact that leptons are an interesting signature of many non-QCD processes. Allowing for a combination of tracking detectors in the inner- and outermost layers of the detector, yields a unique opportunity to trigger on highly energetic muons.

### 3.2.4 Missing energy

Missing energy is an observable defined for certain detector geometries representing the particles that are *invisible* in the detector. These are mostly neutrinos or yet undetected particles from beyond the SM physics. These particles interact rarely and, while not leaving a detectable signature, can carry a significant momentum fraction. This fact allows to find indirect evidence by requiring momentum conservation.

## 3.3 The ATLAS detector

The [ATLAS](#), “A Toroidal LHC Apparatus”, detector is a general purpose detector designed for a maximum coverage of final states. This allows for many topics of research within the realm of particle physics. It has a cylindrical shape measuring 46 m in length and 25 m in diameter. Its dimensions result in a weight of 7 000 t, making it the largest detector ever built for a particle accelerator experiment.

**ATLAS** has the distinguishing structure of a general purpose detector. Its innermost part is formed by tracking detectors directly surrounding the interaction point, followed by calorimeters, and a final layer for muon tracking. All the components and its dimensions are visualized in figure 3.3 including two humans to give an impression of the detector's size.

The detector's components in combination with an advantageous choice of coordinates allows for an efficient detection and reconstruction of a plethora of processes. This section begins with an introduction to the choice of coordinate system, followed by a description of the sub-detectors. On the basis of this, the ability to trigger on and distinguish different processes in **ATLAS** is highlighted. Finally, the signatures and reconstruction algorithms of the objects most relevant to this thesis are presented.

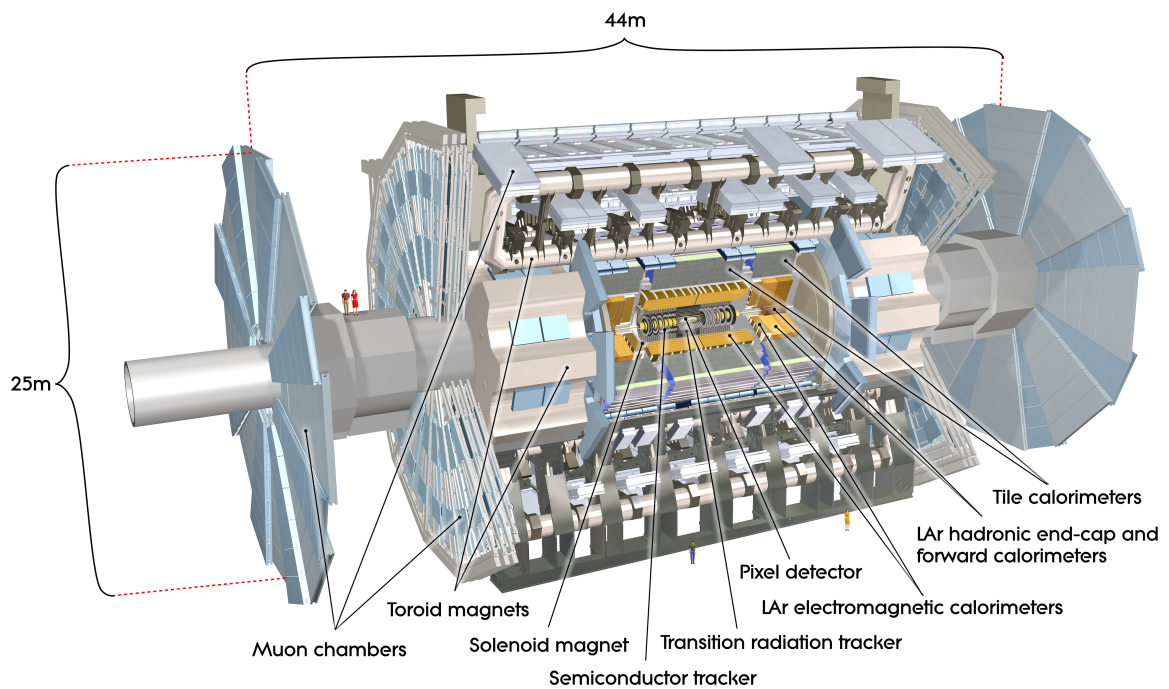


Figure 3.3: The computer generated graphic depicts the **ATLAS** detector, all its subcomponents and its scale. Close to the left hand side end-caps and at the bottom a human couple is added for scale [31].

### 3.3.1 The ATLAS coordinate system

It is advantageous to define coordinates that are well suited for the experiment's key data and the detector's geometry. The **ATLAS** detector utilizes coordinates well adapted to both the kinematics of particles at the high collision energy and the exploitation of its cylindrical shape. The coordinate system is right-handed and right-angled with the  $z$ -axis pointing along the **LHC**'s beam pipe. The corresponding transverse plane is defined by the  $x$ -axis pointing towards the ring's centre while the  $y$ -axis points upwards. The origin of the system is defined by the nominal point of interaction. The polar angle,  $\theta$ , is the angle towards the  $z$ -axis and  $\phi$  is the polar angle in the  $x$ - $y$ -plane. Figure 3.4

provides a visualisation of the coordinate system. Since the momentum of the colliding particles in the

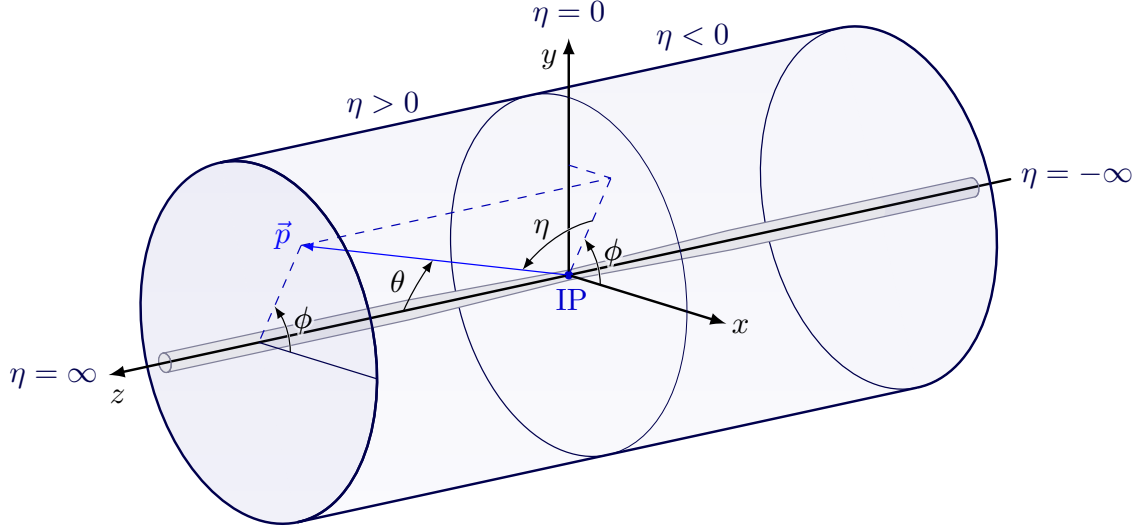


Figure 3.4: The vector graphic visualises the orientation and coordinates of the [ATLAS](#) detector. Based on a vector graphic provided in [32].

transverse plane is assumed 0, the definition of a variable quantifying the direction of emission after the collision with respect to the beam axis is of benefit. The rapidity yields the demanded properties and is defined as

$$y = \frac{1}{2} \ln \left( \frac{E + p_z c}{E - p_z c} \right). \quad (3.20)$$

The rapidity ranges from 0 for particles perpendicular to the beam axis to  $\pm\infty$  for particles down the beam axis. Additionally, one can prove that differences in rapidity are invariant under Lorentz boosts. However, the rapidity has the drawback of its reliance on precise measurements of a particle's energy and momentum. Especially, in the direction of the beam axis, the precision in accelerator experiments is insufficient. Utilising the fact that for highly relativistic particles, the mass becomes negligible, another auxiliary quantity, named pseudo-rapidity is established as

$$\eta = -\ln \tan \frac{\theta}{2}. \quad (3.21)$$

For the high collision energies at the [LHC](#)  $y \simeq \eta$  applies, while providing a quicker and more precise estimation of the quantity. The angular variables of the system are defined within

$$\eta \in (-\infty, \infty), \phi \in [-\pi, \pi). \quad (3.22)$$

Additionally, the transverse momentum is defined as

$$p_T = \sqrt{p_x^2 + p_y^2}, \quad (3.23)$$

where  $p_x$  and  $p_y$  are the momenta along the corresponding axes. The angular separation of two objects is measured using their  $R$ -value. The angular difference,  $\Delta R$ , is defined as:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}. \quad (3.24)$$

Lastly, two further quantities must be mentioned representing the displacement of an object from a point in space, in this case the point of hard-scatter interaction. Here,  $d_0$  refers to the closest distance in the  $r - \phi$ -plane and  $z_0$  in the longitudinal plane.

The shape of the [ATLAS](#) detector makes the spherical system the obvious choice. This comes with the additional benefit of a well defined transverse plane where the sum of all vectors in the final state is approximately zero because the initial state has no transverse momentum. This is the essential prerequisite to the reconstruction of invisible particles.

### 3.3.2 The Inner Detector

The innermost part of the detector is aptly named the [Inner Detector \(ID\)](#). It consists of several tracking detectors layered around the beam pipe close to the interaction point [33, 34]. The [ID](#) consists of two silicon detectors, namely the Pixel Detector and a Silicon Microstrip Tracker (SCT), as well as a straw detector named Transition Radiation Tracker (TRT). Each sub-detector consists of a barrel and two end-caps, covering a total range of  $|\eta| < 2.5$ . The Pixel detector consists of four barrel layers, the innermost of which is the Insertable-B-Layer (IBL) [35, 36]. It has been added in Run 2 and its high granularity and close proximity to the beam pipe allow for an improved reconstruction of secondary vertices, particularly of  $b$ -quark decays. Additionally, it has three end-cap disks associated with it. The SCT comprises four cylindrical layers in the barrel and nine planar disks in each end-cap. Lastly the TRT is a straw-tube detector made of three rings and 300 000 straws in total. It enables radially extended track reconstruction up to  $|\eta| = 2.0$ . A sketch of the layers of the [ID](#) is given in figure 3.5. The [ID](#) allows for precise measurement of the trajectory, momentum and electrical charge of electrically charged particles. Additionally, its close proximity to the interaction point enables reconstruction of the primary and secondary vertices.

### 3.3.3 The calorimeter system

The [ATLAS](#) calorimeter system comprises three parts, the [electromagnetic calorimeter \(ECal\)](#), the [hadronic calorimeter \(HCal\)](#) and the [forward calorimeter \(FCal\)](#), covering a range of  $|\eta| < 4.9$  in combination [38, 39].

The [ECal](#) follows up on the [ID](#). It is a liquid-argon sampling detector with lead absorbers, utilising an accordion geometry for faster readout. As the name implies, it is an electromagnetic calorimeter spanning 24 radiation lengths in the barrel and 26 in the end-caps. A pre-sampler covers a range of  $|\eta| < 1.8$  permitting a measurement of showers starting before the calorimeter and thereby improving the energy resolution.

Surrounding the [ECal](#) is the [HCal](#). It uses plastic-scintillator-tiles in the barrel and liquid-argon sampling with copper absorbers in the end-caps.

Lastly the [FCal](#) represents three layers in the [HCal](#) end-caps focussing on the measurement of particles at high  $\eta$ .

A summary of coverage and interaction length of the sub-calorimeters is depicted in figure 3.6.

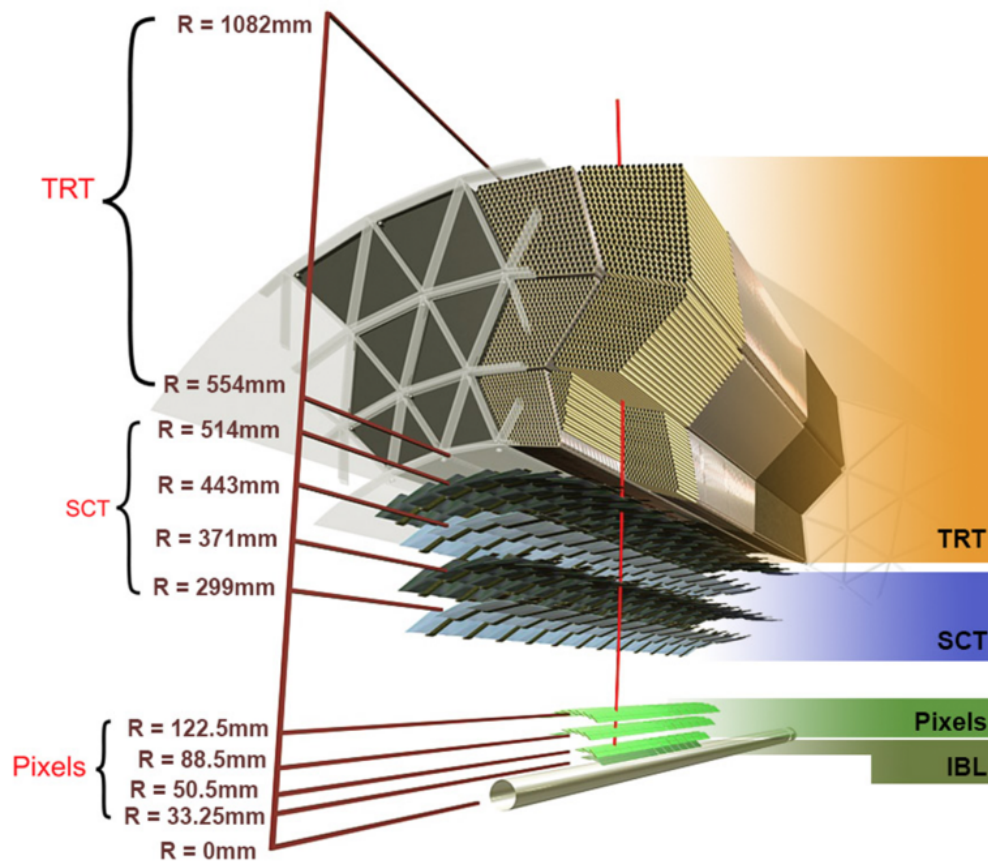


Figure 3.5: A sketch of the layers of the ID surrounding the beam pipe in the barrel [37].

### 3.3.4 The muon spectrometer

The **Muon Spectrometer (MS)** forms the outermost layer of the **ATLAS** detector [41]. Following the structure of the other sub-detectors, it is split into a barrel and two end-caps covering regions of  $|\eta| < 1.2$  and  $1 < |\eta| < 2.7$  respectively. Its design enables not only its utilisation as a trigger for interesting events but also the high precision measurement of muon kinematics. To fulfil these two tasks four different tracking technologies are applied. Thin Gap Chambers (TGC) and Resistive Plate Chambers (RPC) are used for trigger purposes. The functionality allows for a decision making time of less than  $2.5 \mu\text{s}$ . Additionally, Monitored Drift Tubes (MDT) and Cathode Strip Chambers (CSC) in combination with a magnetic field orthogonal to the muon trajectory provide a precise measurement of the muon kinematics.

### 3.3.5 The ATLAS magnet system

The superconducting magnet system is an essential component of **ATLAS** as it allows an extraction of momentum information from the **ID** and **MS** via the Lorentz force [42–45]. It consists of a central solenoid magnet surrounding the **ID** and three toroid magnets, one in the barrel (BT) and two in the

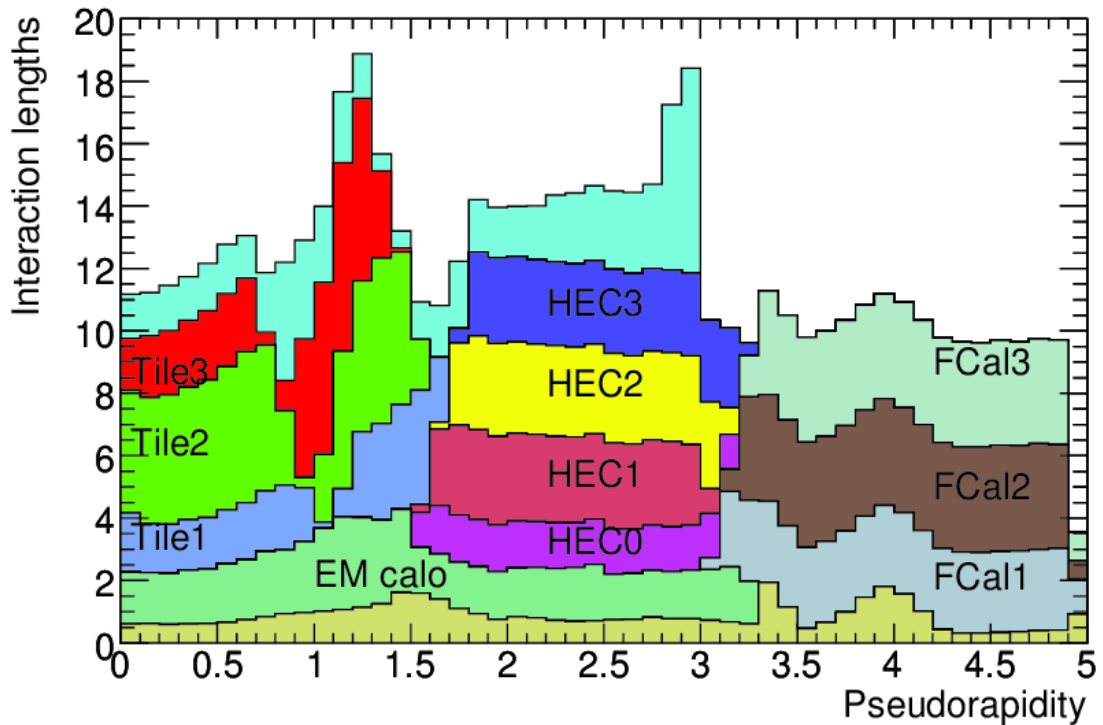


Figure 3.6: The graph shows the cumulative material of the calorimeter parts in units of the interaction length as a function of  $\eta$ . The plot allows for a better understanding of coverage of each sub-calorimeter and layer [40].

end-caps (ECT). The central solenoid is 5.8 m long, 2.56 m in diameter and provides a field of 2 T in the ID. The BT and ECT provide peak magnetic fields of 3.9 T and 4.1 T respectively.

### 3.3.6 The ATLAS trigger system

The LHC's record luminosity provides the data volume necessary for many physics analyses. But the statistical nature of proton collisions together with the high event rate lead to a significant amount of uninteresting and PU events that exceeds the magnitude of event by event analysis and the storage space. Therefore, the ATLAS detector is equipped with a Trigger and Data Acquisition system (TDAQ) enabling the selection of interesting physics events for study [46]. The system consists of a hardware-based first level trigger (L1) and a software-based high level trigger (HLT) overall reducing the rate of data recording to 1 kHz from a bunch crossing rate of 40 MHz [47, 48]. The L1 uses information from the calorimeter (L1Calo) and the MS (L1Muon) in reduced granularity to define regions of interest. The remaining event rate of 100 kHz is the input to the HLT. In the HLT a multiple feature algorithm based on a trigger hypothesis performs a reconstruction of the regions of interest. If necessary information from the full detector can be added for object or full event reconstruction. This results in a further selection and a decrease to 1 kHz in recording.



### 3.4 Object reconstruction

The description of the sub-detectors already permits an elegant understanding of object distinction. Figure 3.7 shows the signatures of some representative particles in the schematic of the detector and illustrates their intuitive distinction. For proper analysis usage however, more sophisticated object reconstruction is needed. Here, object reconstruction refers to the process of identifying and quantifying a particle's signature in the detector. In this section the fundamental methods for both the initial interpretation of a single sub-detector's information and the combination of multiple sub-detector signals are described. To limit the extent of the explanation, a focus is placed on the relevant objects for this analysis and a detailed description is provided especially where it is conducive to later understanding.

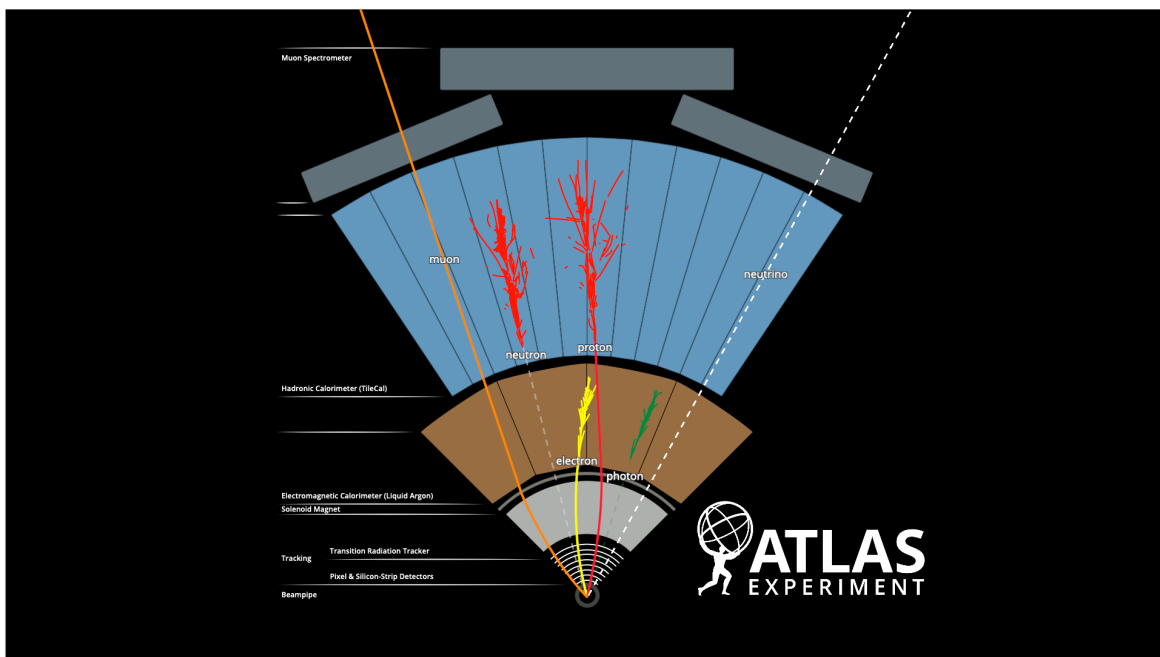


Figure 3.7: The schematic shows the signatures of an electron, a photon, a muon, a neutrino and two hadrons, a proton and a neutron in the different sub-detectors. A dotted line represents no detection. Curved lines are shown for electrically charged particles in the magnetic fields [49].

#### 3.4.1 ID object reconstruction

The objects directly reconstructed from **ID** information are vertices and tracks associated to charged particles [50, 51]. The reconstruction of the later begins with the inside-out sequence. The sequence reconstructs a track step-wise as follows:

1. **Global track seeding** is the generation of three-dimensional representations from the silicon detectors. From these seed objects, track candidates are built.
2. **Kalman filtering** [52] and **smoothing** is used to match further hits to the seeded track candidates.

3. **Ambiguity solving** aims at the removal of fake or overlapping track candidates. It applies a scoring algorithm focusing on fully reconstructed tracks and scaling hits with the sub-detector's precision.
4. In the **TRT track extension** seeded tracks that survived the ambiguity solving are matched to TRT hits.

The inside-out algorithm, however, relies on a good seed in the silicon trackers. This makes it weak for tracks with ambiguous seeds or even to those originating from secondary vertices behind the first **ID** layers. For that reason a secondary algorithm, that goes *outside-in*, is applied:

1. Initially, applicable track segments in the TRT are defined via a Hough transformation mechanism. It is important to note that tracks already used in the first sequence are not reused, which also significantly reduces CPU time.
2. Backtracking these segments to previously unused segments in the silicon layers finalises the tracks in this sequence.
3. Lastly, ambiguity solving is performed.

Beyond the reconstruction of tracks, the **ID**'s information is also responsible for the definition of primary and secondary vertices. The definition of vertices is important to find the hard-scatter vertex and to expose its distinction from the **PU** vertices. A vertex candidate has to be associated with at least two tracks. If several candidates exist, the highest sum of squared transverse momenta of the associated tracks  $\sum p_{\perp}^2$  defines the hard scatter vertex.

### 3.4.2 Calorimeter clusters

From the calorimeters, groups of energy depositions in cells are reconstructed as clusters. Ideally, a cluster corresponds to the energy deposition of a single particle. There are two approaches to cluster extraction.

It is phenomenologically sensible to expect the majority of energy depositions surrounding the particle's original trajectory. Therefore, the topological cell clustering algorithm builds clusters around cells with a large energy deposition. For this reason, cell significance, the signal to noise ratio, is chosen as a base metric. It is defined as:

$$\zeta_{\text{cell}}^{\text{EM}} = \frac{E_{\text{cell}}^{\text{EM}}}{\sigma_{\text{noise,cell}}^{\text{EM}}}. \quad (3.25)$$

$E_{\text{cell}}^{\text{EM}}$  is the uncalibrated cell energy and  $\sigma_{\text{noise,cell}}^{\text{EM}}$  is the expected noise in the cell. Thresholds in  $\zeta_{\text{cell}}^{\text{EM}}$  are used to form clusters following a simple 4-2-0 scheme. First, a primary seed is defined as a cell exceeding the noise by at least a factor of 4,  $|\zeta_{\text{cell}}^{\text{EM}}| > 4$ . In the second step all neighbouring cells satisfying the growth requirement  $|\zeta_{\text{cell}}^{\text{EM}}| > 2$  are added to the cluster. Lastly all neighbouring cells are added. The resulting clusters can be too large for a good measurement of depositions. In particular a cluster can contain several local maxima. In that case a cluster is split in all three spatial dimensions around the local maxima [53].



The resulting topological clusters are best suited for jet and missing energy reconstruction. For the indicative clusters of electrons, photons and  $\tau$ -leptons a sliding-window algorithm is the method of choice. This approach differentiates between electromagnetic clusters for electrons and photons and combined clusters (combining information from both calorimeter parts) for  $\tau$ -lepton identification. The reconstruction follows three consecutive steps:

1. In **tower building** the  $\eta \times \phi$ -space is divided into a grid of rectangle-size  $N_\eta \times N_\phi$  and element size  $\Delta_\eta \times \Delta_\phi$ . The energy per rectangle is then longitudinally summed to create towers with specific energy.
2. The **sliding-window precluster seed finding** uses a window of fixed size  $N_\eta^{\text{window}} \times N_\phi^{\text{window}}$  in units of tower size to define preclusters based on an energy threshold  $E_T^{\text{thres}}$ . Both size and threshold are optimised to allow for a trade-off between high efficiency and noise reduction. The position of the precluster is then defined as the energy-weighted barycentre of all cells and duplicate clusters are removed. For electromagnetic clusters a third step follows. For combined clusters, the encompassed cells form the final cluster.
3. **EM cluster formation** accounts for the different shapes of electron and photon depositions. In this step final clusters are built from the preclusters in different sizes. The starting point is defined as the barycentre of the precluster in the ECal middle layer. The size is defined based on the expected signature in different regions of magnetic field.

Thereby, the algorithm provides clusters appropriate for electrons, photons and  $\tau$ -leptons respectively. A more detailed description of the algorithm exceeds the scope of the thesis and is given in [54].

### 3.4.3 Muon reconstruction

The **MS** aims at the identification and reconstruction of muons [55]. For this purpose, there are two types of muon reconstruction. The stand-alone reconstruction uses exclusively **MS** information:

1. The **MS** stand-alone reconstruction starts with short straight-line segments in individual **MS** stations via a Hough transformation.
2. The segments are loosely combined to form candidates.
3. 3-dimensional models of the candidates are defined and a fit is performed.
4. Based on the fit result, outliers are removed and better hits are added to the model.
5. Ambiguous tracks are removed based on a quality ranking.

Alternatively, knowing the unique signatures of muons in the detector, complete detector information can be used to establish muon candidates. There are five methods and correspondingly five types of muons.

- Combined muons (CB) stem from the matching of tracks in **ID** and **MS** associated with energy loss in the calorimeters. This algorithm follows the intuitive signature of a muon traversing the detector as a minimum-ionising-particle (MIP).

- Inside-out combined muons (IO) are **ID** tracks extrapolated to the **MS**. Calorimeter deposition are again taken into account. This method is suited for precision-weak regions of the **MS**, utilising high quality **ID** information.
- Muon spectrometer extrapolated muons (ME) are based on **MS** tracks that, due to the larger acceptance in pseudo-rapidity of the **MS**, could not be matched to **ID** tracks.
- Segment-tagged muons (ST) require a tight angle agreement between an **ID** track and at least one **MS** segment.
- Calorimeter-tagged muons (CT) are **ID** tracks extrapolated to calorimeter deposition corresponding to a potential MIP. It allows to tag tracks in the **ID** as muons specifically.

The selected muon candidates are refined further depending on the needs of an analyses via:

- An identification working point orders muons in terms of prompt-muon identification efficiency, resolution of the momentum measurement and non-prompt muon rejection. Prompt refers to muons originating from the primary vertex. Non-prompt muons can either stem from the decay of light or heavy flavour hadrons.
- Vertex association criteria enable a further distinction of hard-scatter muons from **PU** or cosmic rays.
- A muon isolation requirement is established to quantify the likelihood of a muon to come from the decay of a boson or a hadronic source. Isolation is measured as the transverse momentum in a cone surrounding the muon candidate relative to the muon's momentum.

### 3.4.4 Missing transverse energy

The **ATLAS** sub-detectors allow for the detection of a wide range of particles and their subsequent reconstruction. But neutrinos and potential unknown particles leave no measurable signatures in the detector. Therefore, a proxy quantity for the transverse momentum carried by undetected particles can be introduced. This quantity is denoted as missing transverse energy  $E_T^{\text{miss}}$  for it uses the detector's geometry. Given the fact that from momentum conservation no momentum is expected in the transverse plane any deviation can be understood as the inverse of an undetected signature [56].

$E_T^{\text{miss}}$  is calculated from signals in the final state. There are two contributions to it. The first contribution comes from fully reconstructed and calibrated objects, electrons, photons, taus, muons and jets. The second component is soft-event signals associated with the primary vertex but no hard objects. The starting point of observable production are the components  $p_{x(y)}$  of the transverse momentum vector. The missing transverse momentum is then calculated to:

$$E_{x(y)}^{\text{miss}} = - \sum_{i \in \{\text{hard objects}\}} p_{x(y),i} - \sum_{j \in \{\text{soft objects}\}} p_{x(y),j}. \quad (3.26)$$

From this definition three observables are constructed, namely the vector  $\mathbf{E}_T^{\text{miss}}$ , its magnitude  $E_T^{\text{miss}}$  and its direction in the transverse plane  $\phi^{\text{miss}}$ :

$$\mathbf{E}_T^{\text{miss}} = \left( E_x^{\text{miss}}, E_y^{\text{miss}} \right), \quad (3.27)$$

$$E_T^{\text{miss}} = |\mathbf{E}_T^{\text{miss}}| = \sqrt{(E_x^{\text{miss}})^2 + (E_y^{\text{miss}})^2}, \quad (3.28)$$

$$\phi^{\text{miss}} = \arctan(E_y^{\text{miss}}/E_x^{\text{miss}}). \quad (3.29)$$

$\mathbf{E}_T^{\text{miss}}$  can then be broken down into its components as follows:

$$\mathbf{E}_T^{\text{miss}} = - \underbrace{\sum_{\text{selected electrons}} \mathbf{p}_T^e - \sum_{\text{accepted photons}} \mathbf{p}_T^\gamma - \sum_{\text{accepted } \tau\text{-leptons}} \mathbf{p}_T^{\tau_{\text{had}}} - \sum_{\text{selected muons}} \mathbf{p}_T^\mu - \sum_{\text{accepted jets}} \mathbf{p}_T^{\text{jet}}}_{\text{hard term}} - \underbrace{\sum_{\text{unused tracks}} \mathbf{p}_T^{\text{track}}}_{\text{soft term}}. \quad (3.30)$$

### 3.4.5 Particle flow objects

Tracks in the **ID**, clusters in the calorimeters, muon candidates and the missing transverse energy can be considered the most direct reconstructions from detector signals. They each represent the purpose of a certain sub-detector or of the chosen geometry. Muon candidates, however, already combine information from all detector parts. A similar approach can be chosen in combining **ID** and calorimeter signatures exploiting the advantages of both sub-detectors. The **ID** convinces with superior resolution in angle and vertex, especially at low  $p_T$ . The calorimeter system provides the ability to measure electrically neutral particles and improves the energy resolution for high  $p_T$  particles.

While there are several approaches to employ the combination of **ID** and calorimeter information, the analysis presented uses a particle flow algorithm [57]. Particle flow is a cell-based energy subtraction algorithm, matching tracks to calorimeter clusters. The biggest advantages are improved suppression of **PU** events and an improvement in missing energy reconstruction due to better reconstruction of low  $p_T$  objects.

A summary of the particle flow algorithm is provided by figure 3.8. The algorithm's starting points are well reconstructed tracks from the **ID** which are, in descending  $p_T$  order, matched to clusters in the calorimeter. Track matching is based on the distance between the track and the barycentre of the energy deposition following the metric  $\Delta R'$  defined as:

$$\Delta R' = \sqrt{\left(\frac{\Delta\phi}{\sigma_\phi}\right)^2 + \left(\frac{\Delta\eta}{\sigma_\eta}\right)^2}. \quad (3.31)$$

$\sigma_\phi$  and  $\sigma_\eta$  denote the respective uncertainties. The matched cells are subsequently subtracted by matching the expected energy from the track to the calorimeter deposition. The expected energy of a track with measured momentum  $p^{\text{trk}}$  can be estimated using a reference track via

$$\langle E_{\text{dep}} \rangle = p^{\text{trk}} \langle E_{\text{ref}}^{\text{clus}} / p_{\text{ref}}^{\text{clus}} \rangle. \quad (3.32)$$

A simple discriminant using the difference between expected and measured energy is then established as

$$S(E^{\text{clus}}) = \frac{E^{\text{clus}} - \langle E_{\text{dep}} \rangle}{\sigma(E_{\text{dep}})} \quad (3.33)$$

and is used to recover showers split over multiple clusters. Lastly remnants can be removed and the

remaining clusters together with the particle flow objects, representing a track matched to clusters, should represent a set of objects with no energy overlap.

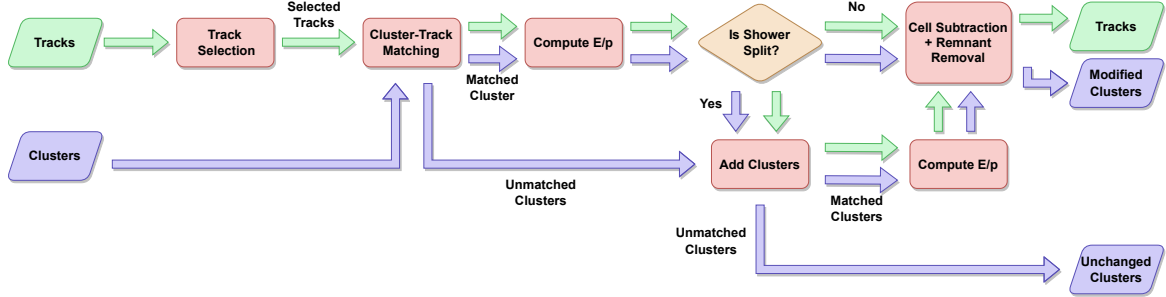


Figure 3.8: The flowchart depicts the formation of particle flow objects starting with the initial tracks and clusters (based on [58]).

### 3.4.6 Electron reconstruction

The expected signal of an electron is a track in the ID and a matching energy deposition in the ECal. In terms of the reconstruction, therefore, there are three components to an electron candidate [59]:

- a localised cluster of energy deposits in the ECal
- charged particle tracks in the ID (if the shower starts early, several tracks can belong to a single electron)
- close  $\eta \times \phi$ -matching of tracks and cluster

The construction of an electron candidate starts with seeds from the EMC using the electromagnetic sliding-window algorithm. Overlapping seeds are ordered according to their transverse energy. Tracks and cluster are then matched via a fitting algorithm. In order to quantify the quality of an electron candidate a total efficiency  $\epsilon_{\text{total}}$  is defined as:

$$\epsilon_{\text{total}} = \epsilon_{\text{EMClus}} \times \epsilon_{\text{reco}} \times \epsilon_{\text{id}} \times \epsilon_{\text{iso}} \times \epsilon_{\text{trig}} = \left( \frac{N_{\text{cluster}}}{N_{\text{all}}} \right) \times \left( \frac{N_{\text{reco}}}{N_{\text{cluster}}} \right) \times \left( \frac{N_{\text{id}}}{N_{\text{reco}}} \right) \times \left( \frac{N_{\text{iso}}}{N_{\text{id}}} \right) \times \left( \frac{N_{\text{trig}}}{N_{\text{iso}}} \right). \quad (3.34)$$

The contributions to  $\epsilon_{\text{total}}$  are summarised as follows:

- The efficiency of cluster reconstruction in the ECal,  $\epsilon_{\text{EMClus}}$ , is defined as the ratio of clusters and true electrons in simulation.
- The reconstruction efficiency,  $\epsilon_{\text{reco}}$ , represents the number of clusters reconstructed as electron candidates.
- The identification efficiency,  $\epsilon_{\text{id}}$ , further limits the candidates through a likelihood function.
- The isolation efficiency,  $\epsilon_{\text{iso}}$ , represents the decrease in candidates due to insufficient separation from photons and hadrons.

- Lastly, the trigger efficiency,  $\epsilon_{\text{trig}}$ , is the number of candidates that also pass the trigger requirements.

An additional BDT-derived criterion is applied to the passing electrons to suppress charge misidentification.

### 3.4.7 Jet reconstruction

The energy loss or short life time of heavy particles leads to the emergence of particle showers. This collimated bunch of hadrons flying roughly in the same direction is referred to as a jet. Jets are the characteristic signatures of many interesting particles in the detector, such as heavy quarks or  $\tau$ -leptons. This makes the distinct appearance of jets in the detector useful for the trigger and essential for many analyses. A phenomenological understanding and well calibrated reconstruction of jets is consequently crucial. For lepton colliders the number of jets in an interaction can be known or at least be limited. In this case one can make some assumptions to easily assign detected particles to a jet cone. Unfortunately, in the presented case of a hadron collider, this task becomes ambiguous because **PU** and high energy events make the correct association of signals to a jet non-trivial. The presented analysis utilises jets reconstructed from Particle Flow objects via the anti- $k_t$ -algorithm [60, 61].

Generally, approaches to reconstruct jets evolve around regions of high energy depositions. Assuming a concentration of energy around the original trajectory of the initial particle makes this a sensible assumption for its central axis. In a sequential recombination objects can be aggregated around the resulting seed based on two distance metrics:

$$d_{ij} = \min(p_{Ti}^2, p_{Tj}^2) \frac{\Delta R_{ij}^2}{R^2}, \quad (3.35)$$

$$d_{iB} = p_{Ti}^2. \quad (3.36)$$

$d_{ij}$  quantifies the distance between particles and  $d_{iB}$  between a particle in momentum space and the beam axis. Objects are then iteratively added to the jet based on both distance measures until only few objects are left.  $R$  is a free parameter determining the radius of the jet. For narrow jets  $R$  is chosen to be 0.4 and 1.0 for jets arising from boosted heavy objects. The resulting object is cone-shaped and for that reason phenomenologically sensible. The anti- $k_t$ -algorithm additionally provides infrared safety, an insensitivity to soft and collinear emissions.

In order to refine the resulting jet objects for analysis usage, a number of adjustments are necessary. The foremost of those calibrations are the jet-vertex-tagger (JVT) [62–64] and the jet-energy-scale (JES) [65]. The purpose of the JVT is the suppression of **PU** jets in the event. To achieve this, a multivariate combination of track-based variables is established that yields a better association of jets with the hard-scatter vertex. The two important variables are the jet-vertex-fraction (JVF) and  $R_{p_T}$ . JVF denotes the fraction of jet track  $p_T$  originating from the hard-scatter vertex via:

$$\text{JVF} = \frac{\sum_k p_T^{\text{trk}_k}(\text{PV}_0)}{\sum_l p_T^{\text{trk}_l}(\text{PV}_0) + \sum_{n \geq 1} \sum_l p_T^{\text{trk}_l}(\text{PV}_n)}. \quad (3.37)$$

Here,  $\text{PV}_0$  denotes the hard-scatter and  $\text{PV}_n$  with  $n \geq 1$  corresponds to **PU** primary vertices. A cut on JVF allows for an effective reduction of **PU** jets. It is worth noting that JVF depends on the number

of vertices in the event and is in practice replaced by the corrected JVF (corrJVF):

$$JVF = \frac{\sum_k p_T^{\text{trk}_k}(\text{PV}_0)}{\sum_l p_T^{\text{trk}_l} \text{PV}_0 + \frac{\sum_{n \geq 1} \sum_l p_T^{\text{trk}_l}(\text{PV}_n)}{k \cdot n_{\text{trk}}^{\text{PU}}}}. \quad (3.38)$$

The correction term  $k \cdot n_{\text{trk}}^{\text{PU}}$  contains the factor  $k$  accounting for the slope of increase in momentum stemming from **PU** jets with respect to **PU** tracks in the event. The aforementioned second variable,  $R_{p_T}$ , quantifies the fraction of transverse momentum originating from the hard-scatter vertex with respect to the jet's  $p_T$  after **PU**-correction:

$$R_{p_T} = \frac{\sum_k p_T^{\text{trk}_k}(\text{PV}_0)}{p_T^{\text{jet}}}. \quad (3.39)$$

This variable peaks at 0 and falls steeply for **PU**-jets, thereby allowing for good decision power.

The second calibration step is the scaling of the jet energy to establish a desired agreement between simulated and measured jet objects. The scaling consists of a sequence of simulation-based techniques followed by *in situ* corrections between data and Monte Carlo directly. Firstly the excess energy resulting from **PU** is removed. Consecutively, the *absolute JES* calibration determines correction factors for both direction and energy of the jet via matching to truth jets in simulation. The third step, formed by a global sequential calibration (GSC), considers flavour dependence and energy leakage. Finally, an *in situ* rescaling of the data itself gauges remaining differences stemming from imperfections in simulation of both detector materials and the physics processes [65].

### 3.4.8 Hadronic tau reconstruction

$\tau$  leptons can decay hadronically or leptonically. Consequently, the reconstruction has to be separate for both cases. In a leptonic decay the  $\tau$  lepton can decay to an electron or a muon creating the corresponding signature in the detector. This section describes the reconstruction of the remaining 65 % of  $\tau$  leptons that decay into hadrons [66].

The visible constituents of the decay are measured as a narrow jet-shaped object. For this reason, the starting point of a  $\tau_{\text{had}}$  is a jet reconstructed from topological clusters via the anti- $k_t$ -algorithm ( $R = 0.4$ ), demanding  $p_T > 10 \text{ GeV}$  and  $|\eta| < 2.5$ . Subsequently, requirements are set according to the expectation of a  $\tau_{\text{had}}$  jet signature. Given the  $\tau$  decay length of  $87 \mu\text{m}$  the decay vertex can be before the IBL but does not have to coincide with the primary vertex. Therefore, a vertex association is performed. A primary vertex candidate is matched to the sum of all track  $p_T$  in the region around the jet seed ( $R < 0.2$ ). The matched vertex is defined as a tau vertex and the impact parameters are calculated accordingly. A second rule of classification in  $\tau$  decays is the prongness, referring to the distinct number of charged particles in the  $\tau$  decay and resulting tracks. 72 % of  $\tau$  are one-prong and 22 % are three-prong. While higher numbers of prongness are possible the two majority cases are used in reconstruction. For tracks to be associated to the  $\tau$  they are required to be in the core region of  $\Delta R < 0.2$ . Requirements of transverse and longitudinal distance are then imposed on these core jets to maximise the number of  $\tau_{\text{had}}$  reconstructed with correct prongness. The reconstruction efficiency is calculated for one-prong and three-prong jets separately.

Until now, the described procedure is trying to find a good method for  $\tau_{\text{had}}$  signature description.

It does however not provide a way to effectively separate the  $\tau_{\text{had}}$  candidates from similar jet-like signatures. For this reason an RNN is applied to yield a better identification [67]. The method uses low-level and high-level variables from both ID and calorimeters. The result is a selection of working points according to signal efficiency and background rejection pairs for either prongness.

### 3.4.9 *b*-jet reconstruction

Flavour-tagging describes the identification of jets originating from the hadronisation of a certain flavour of particle. Especially important to this and many other analyses is the separation of *b*-jets, jets originating from a *b*-quark, from *c*-jets and light flavour jets. An efficient detection of *b*-jets is essential to many high  $p_T$  analyses, for example top physics or generally channels expecting *t*-quarks in the final state.

The method is based on the characteristic properties of the decay of B hadrons. These are distinguished amongst others by high mass, frequent semi-leptonic decays and above all by a detectable secondary vertex. The process of tagging is split into low level and high level decision making. The low level taggers extract quantities sensitive to the aforementioned features of the jet. This includes but is not limited to a soft muon tagger algorithm, targeting muons from semi-leptonic particle decays, and secondary vertex finding algorithms. The information of the low level taggers in combination with a set of suitable variables is fed to the high level tagger. The high level tagger in this work is the DL1r-tagger. DL1 is a family of deep neural network based algorithms yielding a multidimensional output representing the probability of a light-, *b*- or *c*-flavoured jet. As for the other objects an in-depth exploration of the algorithms exceeds the scope of this document and can be found in references [68–71].





---

## Statistical methods

---

It is not necessary to go to the texts of the Bible, already in fables and fairy tales we find powerful tools to convey a lesson. Anecdotes make their message much more durable than the pure information would be. This truth takes on particular force when one remembers the millennia of human existence before writing was invented. Early laws and morals were stories. We use anecdotes to rally others to our cause, we use them when we teach and we use them every day to help us remember. It is little wonder then that even neuroscience states that our brains are hardwired for stories [72]. A story in its purest nature is a direct connection of cause and effect. Unfortunately anecdotes have severe drawbacks. Their perspective is limited and they do not exclude confirmation bias. It is seldom generalisable and makes us susceptible to confuse correlation with causation. That is why scientists must painstakingly rewire their brains. Scientific decision-making demands a different procedure and a rigorous control of it. The mathematical tool-set employed to meet these requirements is statistical methods. While different types of analysis may demand a different tool-set, each one requires a scientific statement from experimentation, a number with an associated uncertainty. In this work, frequentist methods are applied and therefore only an explanation of those is given. An excellent example of the alternative Bayesian approach is provided in reference [12].

Statistical methods describe our instruments of data collection and interpretation. They allow for the provision of an analysis framework which enables the results of an experiment to be quantified in a convincing and supportable way. The mathematical groundwork is given by the field of probability theory and must be discussed first. The first section provides this foundation, together with a description of probability density functions, formulating the probability of specific outcomes. Following this, the method of hypothesis tests forms the link between mathematical probability and experiment. A hypothesis, a statement about the probability of the data, provides a probability density function for a measurement and defines the parameters of interest. Given a hypothesis and its parameters, a measurement can be assigned a likelihood regarding a hypothesis. The estimation of the parameters via maximisation of the likelihood is the method of choice to determine the parameters, and is described in the third section. Following this, the consideration of systematic uncertainties in the form of nuisance parameters is discussed. This last section covers the demands of the experiment at hand a bit more specifically. For an elegant introduction into the statistical methods of particle physics see [73]. Additionally, for a very thorough description of the methods applied, consult the review article on statistics from the Particle Data Group [3].

## 4.1 Fundamentals of probability theory

Probability theory defines the mathematical quantification of probability and the calculation of the same, together with the associated uncertainty. While a comprehensive introduction to the field of probability theory goes beyond the scope of what is feasible here, a brief description of the basic concepts is useful for a complete description of the latter methods. Probability itself can be understood as the likelihood of an event occurring. We have an intuitive understanding of it when betting on a coin flip or the roll of a die. For more abstract processes however a more formal description of the term is required. In the case of radioactive decays or the development of traffic flow an intuitive understanding is quickly lost or hardly sustainable. A detailed mathematical representation is therefore necessary. This, however, does not mean that we have to leave the intuitive relationship to probabilities behind completely. On the contrary, let us start with the description of an  $n$ -sided die. Rolling the die allows for  $n$  different possible outcomes forming a sample space  $S = x_1, x_2, \dots, x_n$ . The likelihood to obtain a certain result  $x_i$  is the probability  $P(x_i)$ . A renowned way of defining this property are the Kolmogorov axioms [74]. Let  $S$  be the set of possible outcomes. To each subset  $A \subset S$  assign a real number  $P(A)$  defined by:

1.  $P(A) \geq 0$ , the probability of each possible event cannot be smaller than zero.
2. Let  $A, B$  be disjoint subsets of  $S$ ;  $A \cap B \implies P(A \cup B) = P(A) + P(B)$ .
3.  $P(S) = 1$ , consequently the possible outcomes are disjunct.

In the die example each outcome is assigned the same probability and the number of outcomes is, for most dice we can imagine, limited. It is however necessary to expand the description of outcome and probability to a continuous set of outcomes and in-homogeneous probabilities. An intermediate step for this can be the histogram as a display of a counting experiment sampling from observable  $S$  (especially as it is common tool in particle physics). In a histogram the observations are plotted in bins of width  $\Delta x_i$  containing measurements in the interval  $[x_i, x_i + \Delta x_i]$ . For an infinite number of observations a normalised histogram can be understood as a quantification of probability for the sample space. Given an infinitesimal bin width  $[x_i, x_i + dx]$ , a continuous display of probability, a probability density function (pdf), is derived. The probability for the observation to lie in a certain interval  $[x_a, x_b]$  for a pdf  $f(x)$  is then given by the integral over the interval,

$$P(a < X < b) = \int_a^b f(x)dx, \quad (4.1)$$

and the integral over the whole set equals 1:

$$\int_S f(x)dx = 1. \quad (4.2)$$

Additionally, it is worthwhile to define a cumulative probability function, the probability sum of outcomes below a certain threshold  $x$ ,

$$C(x) = \int_{-\infty}^x f(x')dx'. \quad (4.3)$$

This definition of probability directly allows the determination of the expectation value of  $x$  as an estimator  $\hat{x}$  of the true value  $x'$ :

$$\hat{x} = E[x] = \int_S x f(x) dx. \quad (4.4)$$

An estimator  $\hat{x}$  is a function of the data used to estimate  $x$ . It should be consistent, unbiased, efficient and robust. [3] The method of estimation is described in section 4.3. Before that, however, the parametrisation of the pdf needs to be described. This connection between probability theory and statistical data evaluation is satisfied by the introduction of an hypothesis in the following section.

## 4.2 Hypothesis tests

A hypothesis H should be understood as statement about the probability of the data  $x$ , denoted as  $P(x|H)$ . It represents a set of parameters  $\vec{\theta}$  defining a pdf for a measurement. A probability that is a function of a hypothesis H,  $P(x|H) = P(x|\vec{\theta})$ , can be interpreted as a likelihood of given hypothesis and its parameters  $\theta$ :

$$L(\vec{\theta}) = P(\vec{x}|\vec{\theta}). \quad (4.5)$$

For frequentist methods we require a full specification of  $P$  as a function of  $\vec{x}$  and H. Thereby the comparison between hypothesis and observation becomes quantifiable.

The  $n$ -sided die on which every integer number from 1 to  $n$  is mapped exactly once forms a suitable example. A simple null-hypothesis would be that the die is not loaded. The resulting assumption is that the average result after  $k$  rolls of the die is a normal distribution of mean  $(n + 1)/2$ . The average can be used as a test statistic because a divergence from the expected mean decreases the degree of belief in the null-hypothesis. However, the amount of divergence to reject the hypothesis is chosen by the user as the critical value  $\alpha$ . The significance of the chosen value is quantified by the cumulative probability of all results leading to rejection. Additionally, for a given observation, the p-value is defined as the proportion of repeated samples as extreme or more extreme than the generated result. It quantifies the significance of an observation.

Figure 4.1 shows an example of a normal distribution together with the rejection threshold and the p-value resulting from a measurement. In the example of the die the number of rolls of a certain side can be regarded a test statistic for the hypothesis. The test statistic used in this work specifically is derived in the next section together with the specific estimator. It is intuitive that the statistical nature of the test allows for both the rejection of a true hypothesis or the acceptance of a false one. A visualisation of these two errors is provided via figure 4.2. Consequently, the result of a measurement is incomplete without a significance and uncertainty because a hypothesis can never be proven from a statistical test. Instead, just inference is possible.

## 4.3 Maximum likelihood estimation

Until now, the parameters  $\vec{\theta}$  which define a hypothesis were assumed to be known. If this is not the case, the parameters need to be estimated from the measurement. A proven method for parameter estimation is the [maximum likelihood estimation \(MLE\)](#). MLE provides the parameters that makes the observed sample the most likely sample amongst all possible samples. This is a simple minimisation problem. The underlying likelihood equations are the derivatives of the likelihood with respect to

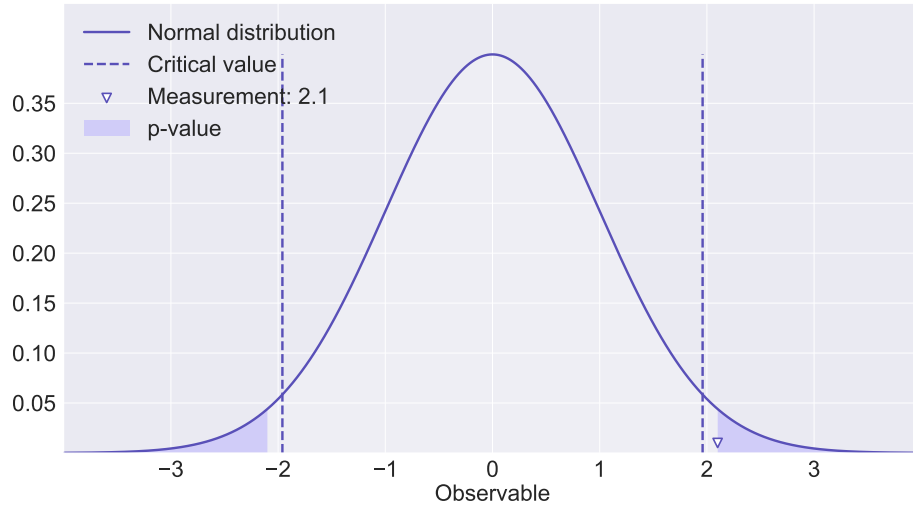
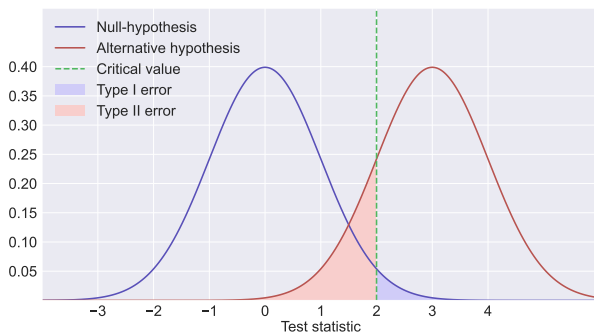


Figure 4.1: The plot shows a normal distribution with dashed lines at the critical values for a two-sided test with a significance of 5%. The triangle marks a possible observation with the corresponding p-value as a shadowed area.



(a) The graphic displays the pdf of a null-hypothesis (blue) and an alternative hypothesis (red). A dashed line marks a possible critical value for rejection and the shaded areas depict the probability of a corresponding type I and type II error.

		True value	
		True	False
Measurement	Accept	Correct Conclusion	Type I Error
	Reject	Type II Error	Correct Conclusion

(b) The table shows the different cases for the two types of error limiting inference from a hypothesis test.

Figure 4.2: Visualisation of the error types in hypothesis tests.

each parameter:

$$\frac{dL}{d\theta_i} = 0, \quad i = 1 \dots m. \quad (4.6)$$

To explain this for the analysis at hand, the respective likelihood function needs to be addressed. The likelihood for  $n$  statistically independent measurements, in this analysis referred to as events, is given by the product of their pdfs:

$$\mathcal{L}(\vec{\theta}) = \prod_{i=1}^n f(x_i; \vec{\theta}). \quad (4.7)$$

If, however, the probability of observing  $n$  events itself depends on  $\vec{\theta}$  the extended likelihood is obtained as follows:

$$\mathcal{L}(\vec{\theta}) = \frac{\mu^n}{n!} e^{-\mu} \prod_{i=1}^n f(x_i; \vec{\theta}). \quad (4.8)$$

While it is true that a higher likelihood communicates a higher agreement with a hypothesis, its exact value is meaningless without context. Therefore, a test statistic should take an alternative hypothesis into account to make the critical region unique. The Neyman Pearson lemma states that the optimal test statistic is given by the ratio of likelihoods for null and alternative hypothesis [75]:

$$\lambda = \frac{\mathcal{L}(H_0)}{\mathcal{L}(H_1)} = \frac{P(x|\vec{\theta}_0)}{P(x|\vec{\theta}_1)}. \quad (4.9)$$

In the next subsection the details for the hypothesis at hand is outlined.

### 4.3.1 Binned likelihood estimation

The analysis at hand is performed on binned data and using a signal strength  $\mu$  as a **parameter of interest (POI)**, to be estimated in the **MLE**. The following formulae are derived for that specific case [76]. A measurement  $\vec{x}$  populates  $N$  bins,  $n_1 \dots n_N$ , resulting in a bin-wise expectation value of:

$$E[n_i] = \mu s_i + b_i. \quad (4.10)$$

Here,  $s_i$  and  $b_i$  denote the mean number of entries per bin for signal and background events respectively. Both values are defined via their pdfs  $f_s$  and  $f_b$ :

$$s_i = s_{\text{tot}} \int_{\text{bin } i} f_s(x_i; \vec{\theta}_s) dx, \quad (4.11)$$

$$b_i = b_{\text{tot}} \int_{\text{bin } i} f_b(x_i; \vec{\theta}_b) dx. \quad (4.12)$$

To further constrain the parameters of the model, subsidiary measurement are performed. These are denoted by bins  $m_i$  and their expectation values  $u_i$ . Assuming Poisson distributions for all bins, the resulting likelihood function becomes:

$$\mathcal{L}(\mu, \vec{\theta}) = \prod_{j=1}^N \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \prod_{k=1}^M \frac{u_k^{m_k}}{m_k!} e^{-u_k}. \quad (4.13)$$

Any hypothesis is then characterised by its assumed value for the **POI**  $\mu$ . A test statistic  $\lambda(\mu)$  is given by the profile likelihood ratio:

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\hat{\theta}})}{\mathcal{L}(\hat{\mu}, \hat{\hat{\theta}})}. \quad (4.14)$$

The denominator states the maximised likelihood for a specific value of  $\mu$  and is divided by the unconditional maximised likelihood.  $\lambda(\mu)$  displays values between 0 and 1, where a larger value means a good agreement between hypothesis and measurement. Equivalently but computationally more efficient, the test statistic  $t_\mu$  is defined,

$$t_\mu = -2 \ln \lambda(\mu) \quad (4.15)$$

for which an increase communicates a larger incompatibility. For a given observation of test statistic  $t_{\mu,\text{obs}}$ , the disagreement can then be quantified via:

$$p_\mu = \int_{t_{\mu,\text{obs}}}^{\infty} f(t_\mu | \mu) dt_\mu. \quad (4.16)$$

The description of the test statistic is concluded here. For different assumptions about  $\mu$ , different versions of  $\lambda$  and  $t$  might arise and a full description is provided in [76]. A more detailed explanation of the technical implementation of the methods is given at the place of their application with the associated results in chapter 8. The following subsection briefly covers the impact of nuisance parameters on the fit before finally the representation of the pdf of a hypothesis is discussed.

### 4.3.2 Nuisance parameters

From a theoretical point of view, the number of parameters that an **MLE** can estimate is not limited. In practice, obviously, the accuracy of such a fit suffers massively from an excess of parameters. For this reason, most of the parameters are usually fixed or limited to a small interval of freedom. Free parameters are then only the **POIs**. In this case one speaks of a profile likelihood fit. The remaining parameters are nuisance parameters (NP). In high energy particle physics there is a long list of initial parameters that are unknown or have high uncertainties. These parameters are estimated from dedicated experiments and calculations. Hence, heavy profiling is employed in most analyses. However, the influence of these unknowns on the analysis and the associated degrees of freedom are not directly known. Rather, the uncertainties can be embodied by independent data sets whose influence on the **MLE** can then be determined indirectly instead of directly. Thereby, the impact in terms of uncertainty can still be accessed for partially profiled parameters in the fit. In general, the representation of hypotheses by simulated samples is explained in the following section.

## 4.4 Monte Carlo simulation

In this chapter the theory is represented by a hypothesis. The methods described explain how a measurement can be compared to a hypothesis to extract a scientific statement. So far, this chapter has left the question of how such a hypothesis is mathematically represented unanswered. For the purpose of introducing the statistical tools, it was sufficient to assume that a fully defined function underlies it.

In reality, it is much more likely that the function contains one or more statistical components. In that case, samples have to be generated from the function to represent the hypothesis.

In the analysis documented here, the theory is based on numerous and multiple statistical or insufficiently defined components. The foundation is provided by the [Standard Model \(SM\)](#) of Particle Physics. In an event, the collision of two protons needs to be understood and finally the expected signature in a detector needs to be explained. The sheer number of poorly understood parts of the processes in addition to their statistical nature makes deterministic methods impossible to use. Instead, [Monte Carlo \(MC\)](#) simulations are applied. These are a class of computational techniques that use randomness and probability to approximate complex mathematical problems. Generally processes and the deviations in nuisance parameters are represented by designated [MC](#) samples. An in depth description of the methods goes beyond the scope of this thesis. However, [appendix C](#) provides a summary of the generators used for specific samples.

#### 4.4.1 The Asimov data set

The samples used in the statistical tests in this work are [MC](#) simulations as described in [section 4.4](#). To calculate the statistics (here a statistic refers to any quantity calculated from a sample independently of the model parameters), sampling from the simulations becomes necessary but computationally expensive. Instead, it can be shown that sampling from an artificial data set, the Asimov data set, is sufficient [\[76\]](#). It gets its inspiration from a short story by Isaac Asimov where a whole democratic election is sampled from a single vote [\[77\]](#). While the sample at hand does not go just as far, it is defined in a way such that any parameter estimation yields the true parameters. This can be used to sample the median value of the test statistic  $t_\mu$ . The statistical analysis in this thesis is executed using the [TRexFitter](#) framework based on [Histfitter](#) [\[78\]](#) and [Histfactory](#) [\[79\]](#). A more hands-on approach to the methods is given in [chapter 8](#) with the respective results. For a more thorough explanation of the methods see the respective documentation.





---

## Machine Learning

---

Over the last decades, computers have become indispensable tools of science; handling large amounts of data, completing tedious calculations, and controlling sophisticated experiments. For the most part, these machines were assigned discrete tasks and they followed step-by-step commands, designed beforehand by human users, and had expected outcomes. For particle physics in particular, computers have been used to select and process data from extremely large samples, allowing the processing of these data at a speed far beyond human capabilities. However, the selection rules always had to be generated by the user, therefore requiring an in depth understanding of the underlying system.

In contrast, the field of machine learning enables a program to establish its own decision rules. A machine learning algorithm can understand a system by itself and opens up various fields of application. It comes as no surprise that this family of algorithms has found its way into almost every aspect of scientific work and its novelty has been a driving force of many a young and old scientist's projects. In the field of particle physics, machine learning influences the entirety of an analysis. Whether it is the preparation or correction of simulations, the reconstruction of signatures, or the isolation of a signal, in all of these areas machine learning has been used with success. As a consequence, such algorithms are also applied in several places in this thesis. For this reason and, because the isolation of the signal via a machine learning algorithm is the central aspect of this work, the topic deserves an insightful introduction.

Machine learning is commonly exemplified by drawing an analogy to the human learning process. While this is an elegant way to provide an initial understanding, it also commonly overlooks a few necessary abstractions. This chapter takes a different route and begins with an abstract definition of a learning algorithm. Then, step by step the components of the learning algorithm are established and with every step more of the abstractions are omitted. This ensures that the basics are covered while the algorithm of particular relevance can be used as a direct example. This is done in the second section which establishes the artificial NN as an example of machine learning algorithms explaining its potential for establishing decision rules. Additionally, a brief description of an alternative algorithm, the Boosted Decision Tree, is provided. Following this up the self-learning capabilities of the algorithm are introduced before in the fourth section the challenge of generalisation for NNs is explained. For a very thorough introduction to the field of machine learning refer to Goodfellow [80].

## 5.1 Learning algorithm

Before explaining a specific algorithm, a starting point has to be an abstract definition of learning applicable to computer programs. An excellent definition is provided by Mitchell [81]: “A computer program is said to learn from experience  $E$  with respect to some class of tasks  $T$  and performance measure  $P$ , if its performance at tasks in  $T$ , as measured by  $P$ , improves with experience  $E$ .” [81, p. 2] Given this definition, there are three components at the base of any learning algorithm; task, performance and experience. This section takes a closer look at each of these and provides examples wherever possible.

### 5.1.1 Task

The starting point of the learning algorithm is the task assigned to it. In this case, there is no need to generalise any further than to a classification task. There is a wide variety of further applications but the task of classification is central to this thesis and sufficient for an introduction of all necessary ideas. Classification means that the algorithm tries to match a set of objects with a set of classes. A possible classification task is the isolation of a signal. In particle physics, that means the categorisation of signatures, called events, into interesting and uninteresting physics. In this thesis these categories are called signal and background.

Given a certain task, the underlying system needs to be understood. Humans have their senses to easily break down their observations into useful features and concepts that can then be processed for decision-making. A machine has no or only limited senses. This means that the step of filtering information for a relevant subset of features still has to be done by humans or a good pre-processing algorithm. Often it means that a list of variables describing an event is submitted to the algorithm. In particle physics this could be kinematic variables of the particles in an event. In simple terms, the task is then defined by a set of information  $x$  and a desired prediction of class  $\hat{y}$  for true class  $y$ . The algorithm  $\phi$  that makes the classification is defined by a set of parameters  $\Theta$  which is translated to the prediction via some function  $w$ :

$$\hat{y} = \phi(x; \Theta) w^\top. \quad (5.1)$$

Later on this general definition of the algorithm is related to specific techniques but for now it is sufficient.

### 5.1.2 Experience

Learning always relies on experience as its fundament. The human learning process can be fuelled by a multitude of experiences. We learn from good teachers, from well-crafted exercises and from repetition. Even the most stubborn of us will learn from their mistakes if they are painful enough. So, most children have to touch a hot stove at least once.<sup>1</sup> This intuitive learning pervades our whole life and cannot be fully imitated in its many facets. Instead, an algorithm needs to learn from a data set specifically provided for this purpose. There are several approaches to this provided learning data, called training samples. The one that is used in this thesis is supervised learning. In supervised learning, the algorithm is provided a training sample that contains the true class  $y$ , called a label.

---

<sup>1</sup> And additionally the cigarette lighter in their parents' car if they are like me.

Thereby, a comparison between predicted class  $\hat{y}$  and true class  $y$  can be established for a decision-rule  $\phi$ .

### 5.1.3 Performance

To make a meaningful improvement based on a training sample, the performance needs to be quantifiable. For supervised learning this means that a comparison between the prediction  $\hat{y}$  and the provided true value  $y$  needs to be translated to a number. For each algorithm  $\phi(x; \Theta)$  the deviation from the truth labels can then be used as a performance metric. This has to be done in two steps. First,  $\phi$  is translated into a prediction  $\hat{y}$  via some function  $w$ . Then for  $\hat{y}$  a performance metric can be calculated. For a binary classification task the output can be as simple as 1 or 0. However, it is usually more practical to give a probability for one of the classes. A common way to achieve this is to use a sigmoid function for the representation of the output  $w$  [82]:

$$\frac{1}{1 + e^{-\phi}}. \quad (5.2)$$

The sigmoid function can represent a probability as it saturates for high positive and negative values. It is also depicted in table 5.1. In a subsequent step the output of  $w$  can be used as the probability  $p$  for class  $\hat{y}$  and compared to the true class  $y$ . This gives a measure for the deviation from the true label for an algorithm  $\phi(x; \Theta)$ . Such metrics are termed cost or loss. In this work both terms are used interchangeably. A common cost function is the *binary crossentropy* for a binary output result:

$$C = -(y \log p + (1 - y) \log(1 - p)). \quad (5.3)$$

The cost is the primary indicator for the training quality. This training quality must not be mixed up with the overall quality of the generated model. Only after a test on a different sample or real data can the model be fully evaluated which is discussed in section 5.4.

With these definitions a task characterised by features  $x$  together with decision-algorithm  $\phi(x; \Theta)$  to be learnt from a sample of labelled data is described. The following section introduces an example of such an algorithm.

## 5.2 A specific learning algorithm - the neural network

The previous section introduced a set of features  $x$  corresponding to a class  $y$  that together define a classification task. A decision-rule for this task is to be found via an algorithm  $\phi(x; \Theta)$ . The algorithm is defined by its internal parameters  $\Theta$  and its performance is measured through some cost function  $C$ . This section presents artificial NNs as an example for a learning algorithm. The artificial NN, or just NN, is a widely used approach to machine learning. Its structure is inspired by the neurons forming the human brain which is also where its name is derived. Instead of neurons a NN consists of numerous very simple processors, called nodes. These nodes are usually structured into several layers as presented in figure 5.1.

The human nervous system relies on some uncertainty when processing information through its net of neural cells. In this net, the output of each neuron is taken as input for the surrounding neurons. The resulting uncertainty cannot be directly implemented for computer algorithms and gives rise to the challenge of representing this fuzziness via many, somewhat more discrete calculations. In a NN,

the neurons and their fuzzy interactions are represented by the nodes. Like neurons, each node can use input from many other nodes to create a new output signal. Thereby the sets of input information can be linked to each other in numerous ways. For this section, it is fully sufficient to consider the most general case. Here, each node of a layer receives input from each node in the previous layer and transmits its information to each node of the following layer. This connection between the nodes is now to be described in detail.

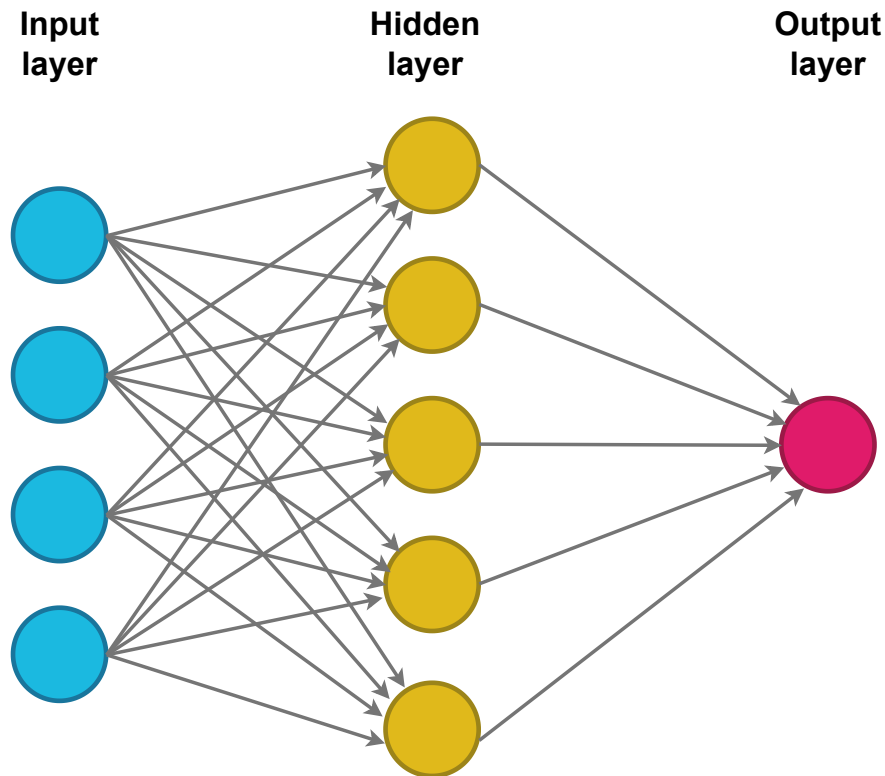


Figure 5.1: Sketch of the typical NN structure. In this example only one hidden layer is included and the output is a single node.

Generally, the first layer of nodes represents the input features  $x$ . Accordingly, the last layer is the output layer and encapsulates the decision-rule. The remaining layers are the moving parts of the algorithm, connecting input and output. They are described by the parameters  $\Theta$  that are to be learnt. Usually, the parameters of the layers are invisible to the user and thus are termed the hidden layers.

The input of every node is the weighted output of all previous nodes, as follows:

$$z_j^L = \sum_{k=0}^N \omega_{jk}^L a_k^{L-1} + b_k. \quad (5.4)$$

$z_j^L$  is the input to the  $j$ -th node in the  $L$ -th layer.  $\omega_{jk}^L$  is the weight from the  $k$ -th node in the previous layer to this node, weighting the output  $a_k^{L-1}$ . Additionally,  $b_k$  is the relevant bias, the offset of the linear connection, representing a possible intercept of the function. Bias can be implemented directly or it can be a consequence of the average input of all nodes in larger networks. The sum indicates that

all  $k$  previous nodes contribute to the input to the  $j$ -th node. In simple terms, the nodes are connected with weights. The weight allows a network to predict which variables are correlated or allow for better decision rules when combined. In addition, the weights can simply mark the strongest variables and features. However, all arithmetic operations so far are linear. To allow the NN to portray non-linear operations an activation function  $\sigma$  is added. This activation function is applied to the output of each node and an example for this was already given with the sigmoid function in the previous section:

$$a_j^L = \sigma(z_j^L) = \frac{1}{1 + e^{-z_j^L}}. \quad (5.5)$$

The choice of activation function is non-trivial and often ambiguous because the choice generally has to be found through trial and error. Table 5.1 summarises a selection of common activation functions and gives a visual representation of their shapes. The full propagation formulae including the activation from one layer to the next is summarised in figure 5.2.

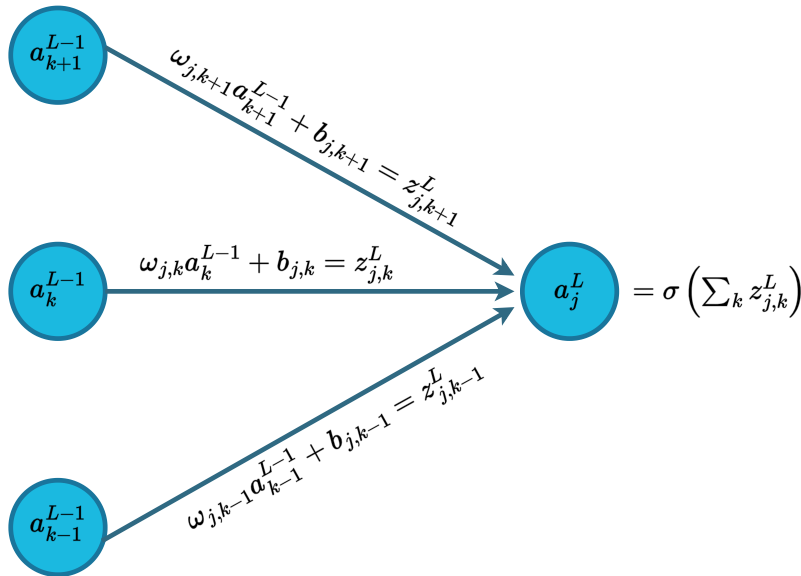


Figure 5.2: Network propagation from layer  $(L - 1)$  to layer  $L$ . The linear connections and their nomenclature is indicated. The non-linear activation of the node is written as  $\sigma$ .

### 5.2.1 Categorical classification

Up to this point a binary classification task has been described. The transition to a classification of multiple classes is trivial. Just the output layer and the loss function need to be adapted. In a multi-classification the output becomes a vector in which each component corresponds to one of the classes. The components  $s_i$  of the vector  $s$  can be calculated by the softmax function:

$$s_i = \sigma(\phi) = \frac{e^{a_i}}{\sum_j^c e^{a_j}}. \quad (5.6)$$

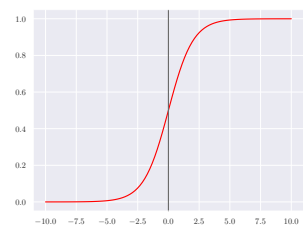
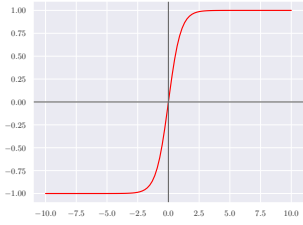
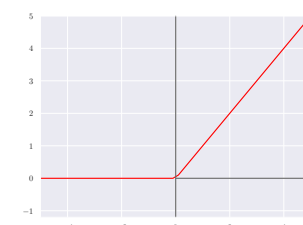
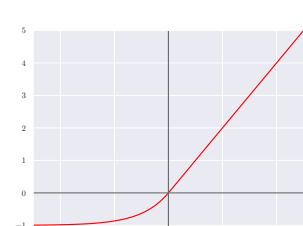
Name	Function	Plot
Sigmoid	$f(x) = \frac{1}{1+e^{-x}}$	
Hyperbolic tangent	$f(x) = \frac{2}{1+e^{-2x}} - 1$	
Rectified Linear Unit, ReLU	$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases}$	
Exponential Linear Unit, ELU	$f(x) = \begin{cases} \alpha(e^x - 1) & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases}$	

Table 5.1: Selection of activation functions as defined in the Keras documentation [82].

Softmax can be considered an expansion of the sigmoid function to multiple dimensions. It guarantees that the sum of the vector components equals 1 and normalises the output to be interpreted as class-wise probabilities. The loss can be expanded from the binary cross-entropy to the categorical cross-entropy:

$$C = - \sum_{i=1} y_i \log \hat{y}_i.$$

With these simple differences understood, the training process is no different to the previously described binary classification task.

### 5.2.2 An alternative approach - Boosted Decision Trees

The biggest competing algorithm to the [NN](#) in particle physics is the [Boosted Decision Tree \(BDT\)](#). A decision tree [83] is a growing set of binary cut-decisions that divides a provided feature space into cuboid regions. In a classification task each region is assigned a class label or respective probability. Starting from an initial set of training events, represented by a list of features, each node of the tree subdivides the set into two subsets. These subsets are then subdivided further until a stop criterion is met.

Boosting is a method of combining multiple weaker classifiers to improve the performance. A common choice of Boosting algorithm that provides an excellent example is AdaBoost [84]. In simplified terms, the weak classifiers are trained iteratively. In each iteration the events that performed poorly are assigned higher weights for the next iteration. Thereby, the classification power is well generalised for the whole set. Finally, the resulting set of weak classifiers is reweighed before its combination to the final classifier. Applying the boosting method to a set of decision trees completed the [BDT](#) algorithm.

## 5.3 Optimisation algorithms

Given the [NN](#) architecture and the cost function, both a way for the creation of a decision-rule  $\Theta$  and its evaluation are presented. However, so far no direction from a poor parametrisation  $\Theta$  to an improved  $\Theta'$  is provided. This step of directional learning is with respect to the human pendant the hardest to compare. For a complex problem we can come up with even more complex ideas to solve it, and the ways there are far from linear.

Again, heat and the painful sensation of getting burned presents a good example. If we sit too close to a fire we might get burned or at least uncomfortably warm. A human would rapidly identify the fire as the source of discomfort and increase his distance to it. A computer algorithm, provided a useful description of the situation, might well be able to identify the discomfort, too. However, recognising the fire as the source of the same and subsequently increasing the distance to it, is not an obvious action. The connection between heat sensation and distance to the fire is intuitive to a human but not to a machine. It might as well move into the fire and step by step try out every spot in the room until the connection between distance and cost is made. For an n-dimensional problem this becomes the main challenge to overcome in machine learning. Given a cost  $C$  of a decision-rule  $\Theta$ , an efficient direction to decrease the cost needs to be defined. The algorithm that takes this role is termed an optimisation algorithm or just an optimiser. This optimisation has two sub-tasks. In a first step, the direction of learning is determined. This is done via backpropagation. Subsequently, a step in terms

of the parameters  $\Theta$  of the NN is calculated.

In the presented NN the information flows forward from the input  $x$  to the output  $\hat{y}$ . This is the forward-propagation. The output  $\hat{y}$  is then transformed into a cost  $C$ . To reverse this propagation the gradient of the loss with respect to all of the NN's parameters  $\Theta$  needs to be calculated:

$$g = \nabla_{\Theta} C(\Theta). \quad (5.7)$$

The algorithm may be computationally expensive but mathematically it can be easily derived via application of the chain rule of differentiation.

With the gradient calculated a step in the opposite direction of the gradient can be taken. This second part of the optimisation is generally called a gradient descent algorithm. Typically, a stochastic gradient descent (SGD) is implemented which only calculates the gradient for a fraction  $m$  of the entire training sample before updating  $\Theta$ . Lastly, the scale of the update is determined by a parameter  $\eta$  termed learning rate:

$$\Theta' = \Theta - \eta g, \quad (5.8)$$

with  $g$  being the SGD defined as:

$$g = \frac{1}{m} \nabla_{\Theta} \sum_j C(\phi(\hat{y}^j; \Theta), y^i). \quad (5.9)$$

In summary, the parameters  $\Theta$  are updated for a fraction of the training sample  $m$  called a batch until the full sample is processed. A full step of processing is named an epoch and can be condensed as follows:

1. The input features  $x$  are propagated forward through all layers of a network resulting in an estimator  $\hat{y}$  at each output node.
2. Based on the truth labels  $y$  a cost  $C$  is calculated.
3. The gradient of the cost is calculated. Given the nomenclature used in this chapter, the resulting equation is:

$$\frac{\partial C}{\partial a_k^{L-1}} = \sum_{j=1}^N \frac{\partial z_j^L}{\partial a_k^{L-1}} \frac{\partial a_j^L}{\partial z_j^L} \frac{\partial C}{\partial a_j^L}.$$

4. Lastly  $\Theta$  is updated based on the negative gradient.
5. The process is repeated for each batch until the full sample has been processed.

The here described SGD is the most basic kind of gradient descent algorithm and comes with a single user-controlled parameter, the learning rate  $\eta$ . A good learning rate should be small enough to avoid oscillations around minima but high enough to approach a minimum efficiently. There are various works on how to effectively decide on such a step-size in algorithms. A common estimate is given by the Robbins Monro condition [85]. Nevertheless, no choice of learning rate is perfect for each part of the problem's topology. *Momentum*,  $\nu$ , can be introduced as a second parameter to the optimiser [82]. The desired effect is twofold; momentum should increase the learning rate in the direction of the



minimum and lower it when approaching said minimum. Momentum scales each step by how aligned previous steps were, meaning it will allow avoiding local minima or moving slowly along a slope. This is accomplished by enlarging steps at the beginning of the training but diminishing them at the end close to the minimum where the slope decreases. It promises to speed up the training with less risk of large oscillations thereby avoiding the negative effects of higher learning rates. Momentum also takes a single scaling user-defined parameter  $\alpha$  and is updated each step in the following way:

$$v' = \alpha v - \eta \frac{1}{m} \nabla_{\Theta} \sum_j L(f(\hat{y}^j; \Theta), y^j), \quad (5.10)$$

$$\Theta' = \Theta + v'. \quad (5.11)$$

Alternatively one can use *Nesterov momentum* [82], which is a more advanced adoption of momentum as it updates the step a further time after applying the gradient:

$$v' = \alpha v - \eta \frac{1}{m} \nabla_{\Theta} \sum_j L(f(\hat{y}^j; \Theta + \alpha \times v), y^j), \quad (5.12)$$

$$\Theta' = \Theta + v'. \quad (5.13)$$

Finally, it can be helpful to decrease the learning rate of the network step-wise while approaching a minimum to avoid oscillations or even missing the minimum completely. This can be accomplished via learning rate *decay*. Decay simply decreases the learning rate in each iteration,  $t$ , by a small fraction,  $d$ , following the assumption that smaller steps are sufficiently close to the minimum [82]. The learning rate is then defined as:

$$\eta' = \frac{\eta}{1 + dt}. \quad (5.14)$$

This sums up the optimisation process and summarises its underlying algorithms. The essential user-controlled parameters have been introduced. The next subsection concludes the description by providing an attempt to automatise these parameters.

### 5.3.1 Adaptive optimisers

In addition to the merely gradient based optimisers, there are *adaptive optimisers*. Learning rate and momentum as previously described are difficult to tune to every part of the training process as the topology of the problem might rapidly change. Therefore adaptive optimisers update their parameters during the training process. From all of the adaptive optimisers [82] *Adam* is probably the most popular [86]. *Adam* updates both its learning rate and momentum over the course of the training based on an exponentially decaying average of past gradients and squared gradients. The average makes sure that the parameters keep getting updated based on past steps. They should be decaying, as otherwise, the parameters would rapidly shrink. The decay of the averages is defined by a two decay parameters  $\beta_1$  and  $\beta_2$  resulting in these gradient definitions:

$$\hat{g}^2 = \frac{\sum g^2}{1 - \beta_1}, \quad (5.15)$$

$$\hat{g} = \frac{\sum g}{1 - \beta_2}. \quad (5.16)$$

The model's parameters are then updated according to:

$$\Theta' = \Theta + \frac{\eta}{\sqrt{\hat{g}^2 + \epsilon}} \hat{g}, \quad (5.17)$$

where  $\epsilon$  is a small control parameter. *Adam* is often considered an excellent algorithm as it enables many corrections during the training, and consequently allows for easier optimisation. However, it also needs significantly more computational power.

### 5.3.2 Neural network metrics

The final part of the introduction to the algorithm in form of the **NN** are the output metrics. The metrics allow for the user to gauge the performance of an algorithm  $\phi$  beyond the pure cost value. This section provides an introduction and an example graph of the important metrics. The plots contain a test sample that is introduced in section 5.4 completing the description of the plots.

The first of these metrics is a plot of the cost (or loss) with respect to the epochs. This plot allows for estimation of progress in the training. During a training the loss is expected to decrease steadily. Once the decrease becomes negligible the training should be stopped and the model saved. Strong fluctuation in the loss are an indicator for an unstable algorithm and demand trouble-shooting. An example of a loss curve is depicted in figure 5.3.

The **receiver operating characteristic (ROC)** curve is a graphical plot for the performance of a binary classifier. The y-axis shows the true positive rate (TPR), i.e. the percentage of correctly assigned labels of value  $\hat{y}$ . The x-axis shows the false positive rate (FPR) and represents the percentage of wrongly assigned  $\hat{y}$  labels. The **ROC** curve can be used to evaluate the quality of a model generated by a classifier. A diagonal curve means the model is as good as a random guess. Curves below the diagonal are worse than a random guess and can indicate that something is inherently wrong with the model. **ROC** curves are commonly summarised via their *Area Under the Curve (AUC)*. The **AUC** gives a robust first indicator to compare model quality because a high value corresponds to a more robust classifier. For an example of a **ROC** curve see figure 5.4.

Lastly, for classification tasks a visual representation of the output per event is useful. This way the assignment of likelihoods over the sample can be observed. The response curve is a histogram of the events in each truth class  $y$  with respect to their estimated class  $\hat{y}$ . The response plot corresponding to the **ROC** curve in figure 5.4 is shown in figure 5.5. The visualisation of the response becomes particularly important if special weights need to be used for events in the training. A common example for this are the negative weights that events can have for subtracting contributions from histograms. These additive weights cannot be used directly in a **NN**. For this reason, the response should be visualised with both types of weights to observe possible instabilities.

## 5.4 Algorithm capacity

Unfortunately humans can be limited by too much and too little *intelligence*. On the one hand, there are tasks that are too complicated for us to solve. On the other hand, under-challenging often leads

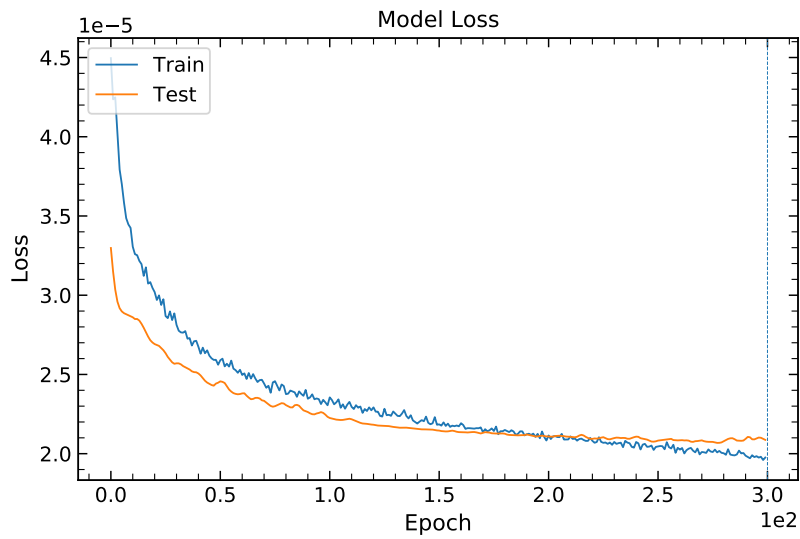


Figure 5.3: The graph shows an example of a loss curve for both the training and test sample. In the final epochs a small divergence between training and test hints at the beginning of overtraining.

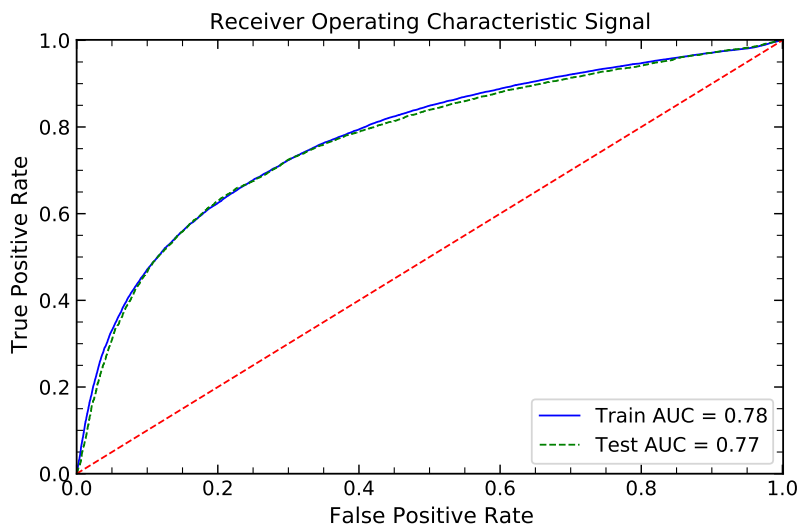


Figure 5.4: The graph shows an example of a ROC curve for both the training and test sample including the AUC value for both distributions. The agreement between both curves is satisfactory.

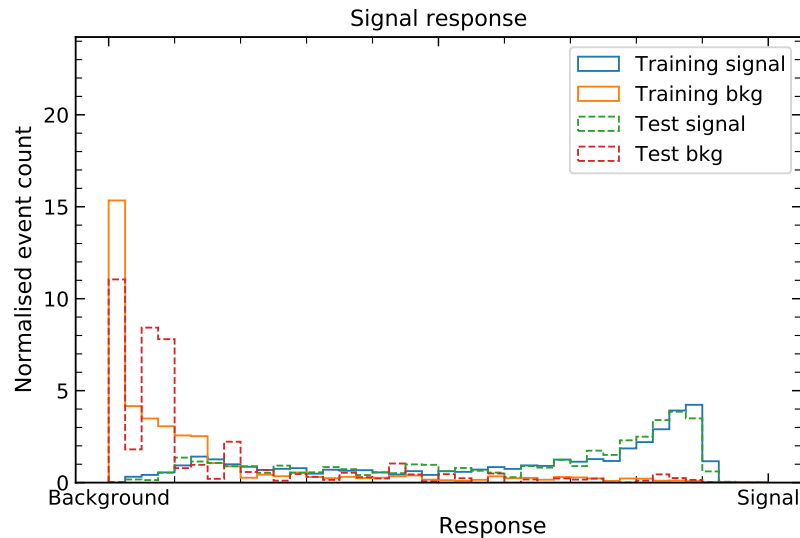


Figure 5.5: The graph shows a response plot corresponding to the ROC curve in figure 5.4. While the ROC curve shows good agreement between training and test, fluctuations in the respective responses are clearly visible. A possible reason can be the weights of the events.

to inefficiency or an unnecessary complication of work. We need the right brain for the right task. The same is true for NNs. For a NN we define this as its capacity and two limitations arise from it: an insufficient capacity results in a lack of cost minimisation and an excessive capacity causes bad generalisation. Generalisation is a core property of learning algorithms that this chapter has omitted until now. A decision-rule in supervised learning is always dependent on the provided sample of labelled training data. If the sample is flawed or at least not fully representative, then the algorithm can never learn the perfect decision-rule. Furthermore, given enough epochs the algorithm can mimic features present in the training sample but not applicable to a general sample.

Clearly, part of the challenge of generalisation lies within the choice of a good training sample. However, tuning the capacity according to the task can minimise the problem of generalisation significantly. Any parameters of the algorithm that the user beforehand defines contribute to the capacity and are termed hyperparameters. These differ from the parameters  $\Theta$  in so far that the user has full control and that they are not adapted during the learning process. The essential hyperparameters have already been mentioned and include the number of nodes and layers, the learning rate and the number of epochs. The number of epochs, defining the length of the learning endeavour, makes for an excellent example to understand its impact on the generalisation. Previously, it was mentioned that a NN might pick up features only present in the specific training sample. This is termed overfitting and in the most extreme scenario the NN is large and deep enough to pick up every single feature in the training sample, thus mirroring it. If the training sample has been chosen well, overfitting should only happen once the training has advanced significantly. For this reason, a smaller number of epochs can mitigate the impact of overfitting.

With the concept understood, the generalisation of a decision-rule needs to be quantified. This is done using a test<sup>2</sup> sample. This can be a fraction of the provided labelled training sample that is

<sup>2</sup> It is sometimes differentiated between test and validation samples. Should this be necessary, it is pointed out.

not used in the training process. Applying a decision-rule to this sample yields a measure for how well it performs on unknown events. For this to work, the test sample needs to be large enough to be representative beyond the statistical fluctuations. Using a proper test sample, all distributions for training and test can be compared and strong disagreements hint at bad generalisation. If for example the loss curves for training and test sample diverge, the NN probably starts to overtrain. For an example of test distributions for each metric see figures 5.3 to 5.5.

Controlling the capacity is one way to minimise these issues. There are more dedicated approaches to solving the problem called *regularisations* of a NN. The most commonly used solution is a so called *dropout* layer described in the following subsection. Additionally, a *batch normalisation* can have an effect of regularisation and is therefore introduced in succession. There is a lot more to a full understanding of regularisation and learning algorithms in general. This chapter only provides a basis of understanding for the rest of this thesis. The relevant chapters also point out more details if it helps to understand the applied methods.

### 5.4.1 Dropout

One computationally inexpensive but widely applicable method of regularisation is dropout [87]. Dropout is based on the fact that an ensemble of networks often times outperforms a single NN. Unfortunately, the cost of using multiple NNs is high. Dropout circumvents this challenge by creating the artificial ensemble of all networks formed by removing a finite number of non-output nodes from a base network. This can be applied to any layer of a NN via a mask  $\mu$  that temporarily sets a random fraction of nodes to a weight of zero. The probability of choosing a node to be dropped is a hyperparameter of the method.

Dropout can be described as an additional layer which attempts to hinder the network from relying on less dominant features. It forces the network to build models that are not based on strong correlations between nodes, making the weights less interdependent. In short, it means training several NNs depending on which nodes are turned on during a training epoch which keeps the training in motion for a high number of epochs. Figure 5.6 sketches the process.

Dropout is added to each layer of a network and can also be restricted to a subset of layers. It slows down the training as the additional motion decelerates the process of finding a minimum. However, it also accelerates each epoch slightly as it simplifies the network architecture.

### 5.4.2 Batch normalisation

Batch normalisation is technically a method for adaptive reparametrisation which comes with regularising capabilities. Its necessity arises from the shortcomings of backpropagation and gradient-descent based algorithms for particularly deep NNs. The optimisation algorithm updates all layers simultaneously. Since only the first derivative of the parameters is considered, there can be unexpected second order effects between the layers. These second order effects are impossible to take into account. Instead, a renormalisation of the output of each layer mitigates the impact. This is done by normalising each output to the batch mean  $\mu_B$  and the batch standard deviation  $\sigma_B^2$ :

$$\mu_B = \frac{1}{m} \sum_i x_i, \quad (5.18)$$

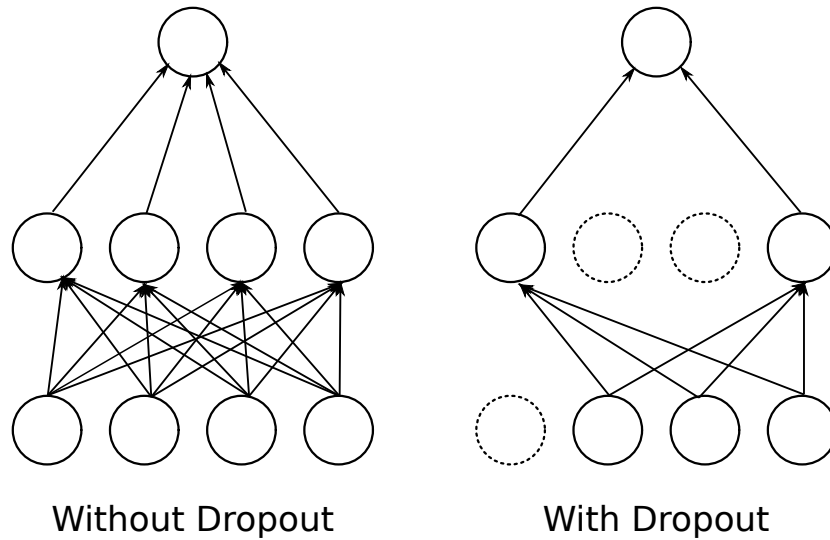


Figure 5.6: Sketch of a NN before and after the inclusion of dropout. On the left hand side dropout is not applied and all nodes are connected. On the right hand side the dashed circles are nodes excluded by dropout and therefore not effectively connected to the other nodes.

$$\sigma_B^2 = \delta \frac{1}{m} \sum_i (x_i - \mu_B)^2, \quad (5.19)$$

$$x_{i,\text{norm}} = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2}}. \quad (5.20)$$

Here,  $\delta$  is a small control parameter to avoid unwanted behaviour around critical points. By doing this, the learning process of an algorithm can be supported.

The additional regularisation power comes from a reduction of covariance shifts. A covariance shift is a drastic change in features over batches and best explained via an example. Imagine a classifier distinguishing between pictures that show cars and pictures that do not show cars. If the training set contains predominantly green cars the colour green might end up as a strong indicator for the classification car. In general the colour green will not be as dominant and the network will perform slightly worse when trying to classify cars of a different colour. This problem becomes especially severe in very deep NNs. In this case small shifts in the sample can have large and unforeseeable consequences in deeper layers. The renormalisation of a batch reduces the impact that such a poorly generalisable batch can have and stabilises the learning process.

---

## Data preprocessing

---

The remaining chapters of this thesis document the isolation of the signal and the subsequent cross section measurement of the  $tH$  process including  $\tau_{\text{had}}$  in its final state. Before any more involved analysis is undergone, a preselection region is defined and the expected background contributions are estimated. This selection of events is split into several phases, each of which has its own justification. The first section describes the selection of events in combination with the rationale. The following section describes the reconstruction of the objects in the final state until, in a final section, the estimation of misidentified leptons is documented.

The provision of the data sets used and a complete description of the information presented is a complex and multifaceted topic. A summary of the used data and MC samples is provided in appendix C and an explanation of additional details is given wherever helpful. Additionally, some important properties of the samples are listed here:

- MC generation is computationally expensive. This means that even for simulated samples statistical limitations can arise.
- Events are assigned weights to match the expected event count in data taking. Some MC generators allow for negative weights in the contribution. These weights are meant to be used in histograms in a counting experiment and various challenges can arise from usage in other methods.
- The simulations contain several layers of information ranging from parton level to reconstruction level. The lowest level of simulation is referred to as truth level and can be used to find the true origin of reconstructed particles.

### 6.1 Event selection

The process of choosing the region for an analysis has multiple components and differs significantly depending on the overall goal of the effort. To understand the steps, it is helpful to be aware of three main objectives: (1) The foremost goal is always the reduction of background processes and an associated improvement of the signal to background ratio. (2) Additionally, the selection has to be done in way that avoids overlap with other analyses covering different final states. Thereby, a later combination of results becomes possible. (3) Lastly, not only a reduction of total background

count but also a limitation of the types of background is useful because it gives more control over the background estimation.

The main ingredient to the selection are the objects in the signal's final state, depicted for  $tHq$  in figure 6.1, and their signatures in the detector. The final state consists of two  $\tau$  leptons originating from the decay of the Higgs boson and a  $t$ -quark decaying into a  $b$ -quark and a  $W$  boson. Lastly, as a consequence of the  $t$ -channel production, an additional light jet is expected, termed the spectator jet. The  $\tau$  leptons can result in hadronic or leptonic signatures and an additional leptonic signature is expected from the  $W$  decay. Since hadronic signatures yield less clean detector responses leptonic signatures are an excellent basis for a first rough selection.

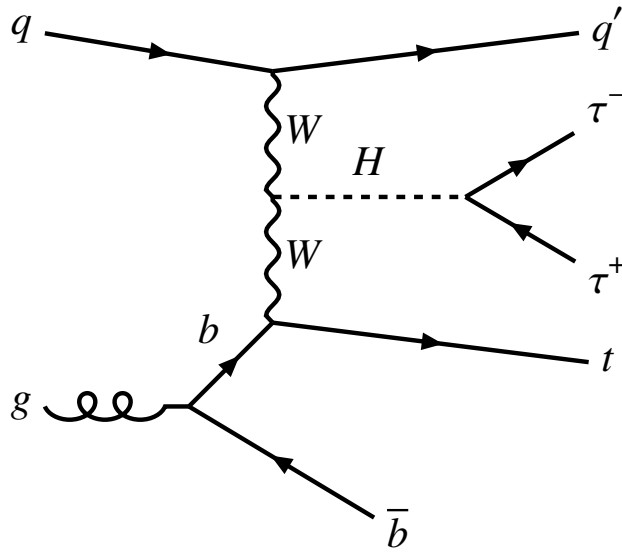


Figure 6.1:  $tHq$   $t$ -channel process with decay into  $\tau\tau$  in the 4-flavour scheme.

The data at the LHC is recorded over long stretches of time and the quality is not always stable. For this reason an initial selection is made via the “good run list”, ensuring that all selected events belong to high quality data taking periods. There are additional requirements applied to a variety of objects following the recommendations of the reconstruction groups, summarised here as object quality requirements. An introduction to the reconstruction process and a basic understanding of possible quality controls is given in chapter 3.4. In general, tighter regulations mean a lower signal efficiency in exchange for a lower misidentification rate.

The first layer of analysis-specific selection cuts is formed by the requirement of at least one reconstructed primary vertex and at least one single lepton trigger (either electron or muon). This rough selection is common for a wider range of  $tH$  analyses, a distinction to which is created through the requirement of  $\tau_{\text{had}}$  in the final state.

After this latter requirement ensures the orthogonality to the other channels of the  $tH$  analysis, the exact composition of the leptons in the final state is further specified in order to characterise channels with less and specific backgrounds. The first one is the  $1\ell + 2\tau_{\text{had}}$  channel in which two of the three leptons are required to be  $\tau_{\text{had}}$  with opposite charge. The second one is the  $2\ell + 1\tau_{\text{had}}$  channel in which only one of the three leptons is required to be a  $\tau_{\text{had}}$ . The  $2\ell + 1\tau_{\text{had}}$  channel is subsequently split into two sub-channels depending on the charge of the light leptons:  $2\ell \text{ OS} + 1\tau_{\text{had}}$  for opposite



sign (OS) light leptons and  $2\ell$  SS +  $1\tau_{\text{had}}$  for same sign (SS) light leptons. This creates three channels with distinct background compositions. The number of events after a selection is termed yields. A pie chart and a table of yields for each channel are provided in figures 6.2 to 6.4 and tables 6.1 to 6.3. The pie charts give an understanding of the signal to background ratio and the dominant background contributions to each channel. It becomes apparent that the dominant background in the  $1\ell + 2\tau_{\text{had}}$  selection is  $t\bar{t}$  with additional significant contributions from Z+jets and W+jets. The  $2\ell$  OS +  $1\tau_{\text{had}}$  channel is dominated by Z+jets and  $t\bar{t}$  background events. Lastly, the  $2\ell$  SS +  $1\tau_{\text{had}}$  channel has the most diverse background composition with significant contributions from all  $t\bar{t} + X$  processes. For a later MLE, it is beneficial to understand the dominant contributions and to find regions in which each contribution can be controlled specifically. The yield tables reflect the same composition for each channel. Additionally, the raw yields are listed. Raw yields represent the number of events before any weights are applied to represent the expected event count. It is worth noting, that  $t\bar{t}$  is proportionally stronger represented in the weighted than in the raw yields. This results in statistical limitations for further analysis of the raw events.

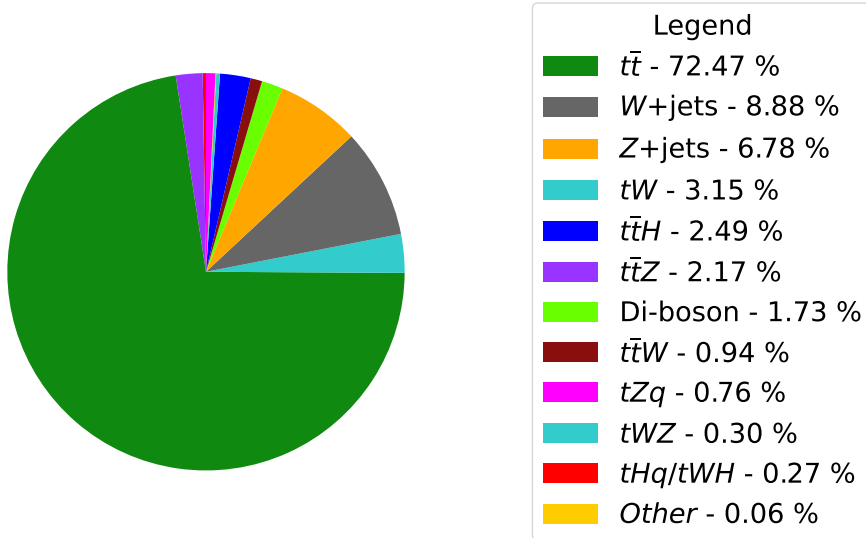


Figure 6.2: Pie chart of the contributions to the channel after the  $1\ell + 2\tau_{\text{had}}$  selection is applied.

All three channels share another layer of selection criteria. These are chosen to get a cleaner signal region but are of lesser importance to the understanding of the remaining chapters. One light lepton needs to match a trigger object. All leptons need to have at least 20 GeV of  $p_T$  and the leading lepton is required to have a  $p_T$  of at least 27 GeV. Events are allowed to contain 2 to 6 EMPFlow jets including one or two b-tagged jets. All jets need to fulfil  $|\eta| \leq 4.5$  and  $p_T \geq 20$  GeV. The *b*-jets need to fulfil  $|\eta| \leq 2.5$  and  $p_T \geq 20$  GeV at the working point DL1r\_70 [69–71]. Each event may include 5 GeV to 800 GeV of missing energy. Electrons are restricted to  $|\eta| \leq 2.47$  under exclusion of  $|\eta| \in [1.37, 1.52]$ . Muons are allowed in  $0.01 \geq |\eta| \leq 2.5$ .  $\tau_{\text{had}}$  leptons need to fulfil  $|\eta| \leq 2.5$  under exclusion of  $|\eta| \in [1.37, 1.52]$  and pass the RNNMedium [67] requirement.

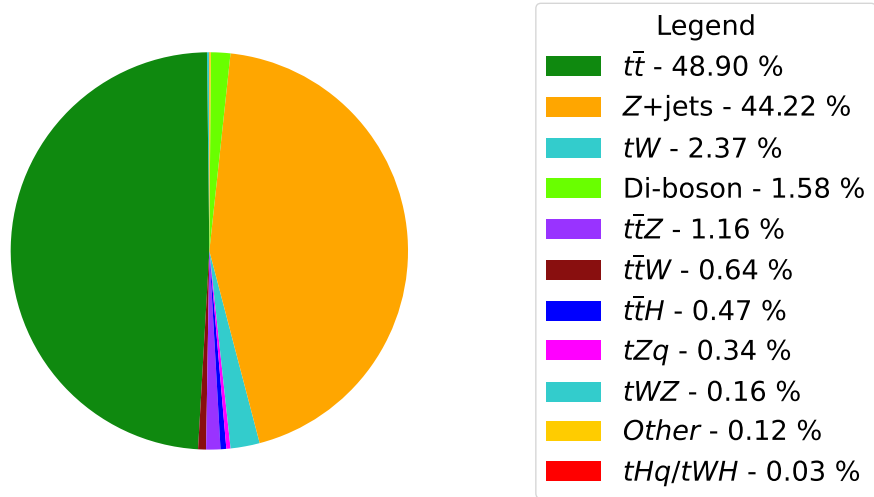


Figure 6.3: Pie chart of the contributions to the channel after the  $2\ell$  OS +  $1\tau_{\text{had}}$  selection is applied.

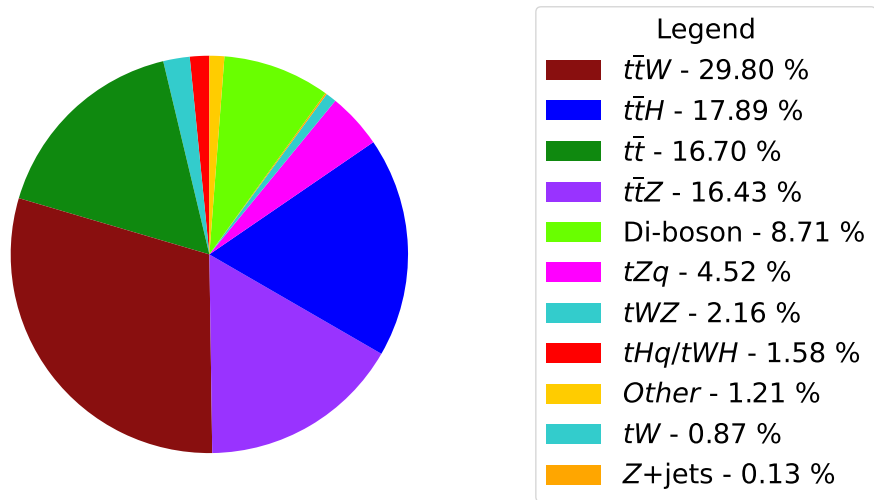


Figure 6.4: Pie chart of the contributions to the channel after the  $2\ell$  SS +  $1\tau_{\text{had}}$  selection is applied.

Table 6.1: Raw and weighted yields in the  $1\ell + 2\tau_{\text{had}}$  channel. The raw yields are calculated before the inclusion of luminosity scaling. The displayed uncertainties are as calculated by the MLE software tool. The uncertainty of the raw yields is Poisson distributed and then propagated to the weighted yields. The uncertainty of the total yields is the sum of the separate Poisson uncertainties. If processes are combined in a fit, this can be treated differently.

Channel	Raw yields	Weighted yields
$tHq$	34 020 $\pm$ 180	1.94 $\pm$ 0.01
$tWH$	1 832 $\pm$ 43	1.54 $\pm$ 0.04
$tWZ$	42 430 $\pm$ 210	3.96 $\pm$ 0.02
$t\bar{t}$	7 176 $\pm$ 80	947 $\pm$ 11
$t\bar{t}W$	3 951 $\pm$ 60	12.29 $\pm$ 0.19
$t\bar{t}Z$	12 726 $\pm$ 110	28.4 $\pm$ 0.2
$t\bar{t}H$	39 890 $\pm$ 200	32.5 $\pm$ 0.2
$tZq$	59 490 $\pm$ 240	9.97 $\pm$ 0.04
$tW$	342 $\pm$ 18	41.1 $\pm$ 2.1
Z+jets	2 814 $\pm$ 50	84.7 $\pm$ 1.5
Z+jets (low mass)	8 $\pm$ 3	3.9 $\pm$ 1.5
$t$ -channel	112 $\pm$ 11	4.9 $\pm$ 0.5
W+jets	528 $\pm$ 23	116 $\pm$ 5
Diboson	7 410 $\pm$ 86	22.6 $\pm$ 0.3
Minor bkg.	1 556 $\pm$ 40	0.76 $\pm$ 0.02
Total background	214 300 $\pm$ 460	1 312 $\pm$ 3

Table 6.2: Raw and weighted yields in the  $2\ell$  OS +  $1\tau_{\text{had}}$  channel. The raw yields are calculated before the inclusion of luminosity scaling. The displayed uncertainties are as calculated by the MLE software tool. The total uncertainty of the raw yields is Poisson distributed and then propagated to the weighted yields. The uncertainty of the total yields is the sum of the separate Poisson uncertainties. If processes are combined in a fit, this can be treated differently.

Channel	Raw yields	Weighted yields
$tHq$	24 518 $\pm$ 160	1.377 $\pm$ 0.009
$tWH$	2 447 $\pm$ 49	1.94 $\pm$ 0.04
$tWZ$	189 210 $\pm$ 400	16.74 $\pm$ 0.04
$t\bar{t}$	43 780 $\pm$ 210	5 382 $\pm$ 26
$t\bar{t}W$	25 240 $\pm$ 160	69.6 $\pm$ 0.4
$t\bar{t}Z$	84 640 $\pm$ 290	121.9 $\pm$ 0.4
$t\bar{t}H$	103 020 $\pm$ 320	48.77 $\pm$ 0.15
$tZq$	228 830 $\pm$ 500	34.6 $\pm$ 0.1
$tW$	2 143 $\pm$ 50	249 $\pm$ 6
Z+jets	208 770 $\pm$ 500	4 689 $\pm$ 11
Z+jets (low mass)	820 $\pm$ 29	135.5 $\pm$ 0.5
Diboson	49 560 $\pm$ 220	167 $\pm$ 1
Other Higgs	24 $\pm$ 5	10.4 $\pm$ 2.2
Minor bkg.	9 814 $\pm$ 100	2.64 $\pm$ 0.03
Total background	972 800 $\pm$ 1 000	10 930 $\pm$ 11

Table 6.3: Raw and weighted yields in the  $2\ell$  SS +  $1\tau_{\text{had}}$  channel. The raw yields are calculated before the inclusion of luminosity scaling. The displayed uncertainties are as calculated by the MLE software tool. The total uncertainty of the raw yields is Poisson distributed and then propagated to the weighted yields. The uncertainty of the total yields is the sum of the separate Poisson uncertainties. If processes are combined in a fit, this can be treated differently.

Channel	Raw yields	Weighted yields
$tHq$	16 560 $\pm$ 130	1.297 $\pm$ 0.010
$tWH$	952 $\pm$ 31	1.078 $\pm$ 0.035
$tWZ$	26 710 $\pm$ 163	3.24 $\pm$ 0.02
$t\bar{t}$	208 $\pm$ 14	25.1 $\pm$ 1.7
$t\bar{t}W$	11 285 $\pm$ 110	44.8 $\pm$ 0.4
$t\bar{t}Z$	11 506 $\pm$ 110	24.7 $\pm$ 0.2
$t\bar{t}H$	28 800 $\pm$ 170	26.9 $\pm$ 0.2
$tZq$	30 770 $\pm$ 180	6.79 $\pm$ 0.04
$tW$	11 $\pm$ 3	1.31 $\pm$ 0.36
Z+ jets	90 $\pm$ 10	0.474 $\pm$ 0.053
Z+ jets (low mass)	0 $\pm$ 0	0 $\pm$ 0
Diboson	3 419 $\pm$ 60	13.1 $\pm$ 0.2
W+ jets	2 $\pm$ 1	0.134 $\pm$ 0.067
Minor bkg.	4 081 $\pm$ 64	1.75 $\pm$ 0.03
Total background	134 400 $\pm$ 400	150.6 $\pm$ 0.5

## 6.2 Object reconstruction

The selection relies on the object reconstruction described in chapter 3. Given an expected signature, it can be helpful to reconstruct additional objects in the event. In this case, these are the Higgs boson and the  $t$ -quark. Reconstruction means an estimation of the particle's kinematic variables, first and foremost the mass, and provides powerful information for further analysis. A good motivation is the expected deviation in the reconstructed Higgs boson mass for backgrounds that do not contain a Higgs boson. The key component to the process is a matching of final state objects to their parent particles, which is complicated by the presence of neutrinos. The reconstruction algorithms described below focus on the  $tHq$  signature. For a more detailed documentation of the procedures see [88].

### 6.2.1 Lepton association

In order to reconstruct the  $t$ -quark and the Higgs boson the leptons in the final state need to be matched to their origin particles. Depending on sign and number of light leptons the procedure is different for each channel. In the following, fundamental details are presented for each case.

In the  $1\ell + 2\tau_{\text{had}}$  channel the two  $\tau_{\text{had}}$  are assumed to stem from the decay of the Higgs boson and should have opposite charge. Based on this assumption the  $\tau_{\text{had}}$  are assigned correctly in 86.1 % of all simulated  $tHq$  events. This means that the signal sample contains events with  $\tau_{\text{had}}$  that do not originate from the decay of the Higgs boson, e.g. from a  $t$ -quark. The remaining light lepton is subsequently associated to the  $t$ -quark. An additional study for  $tWH$  samples is not performed and the same assignment is applied there.

In the  $2\ell$  OS +  $1\tau_{\text{had}}$  channel only one  $\tau_{\text{had}}$  is present, which in the majority of events is associated to the Higgs boson. Knowing the charge of the  $\tau_{\text{had}}$ , the light lepton with opposite charge is associated

to the Higgs boson. The other light lepton is associated to the  $t$ -quark. Similarly, no additional study for the  $tWH$  signal is performed.

In the  $2\ell SS + 1\tau_{\text{had}}$  channel the light leptons have the same charge and an association is more complicated. A **BDT** based on the TMVA package [89] is employed to associate the leptons to their origin [90]. Initially, using truth information the reconstructed leptons are matched to their truth origin in a cone of  $\Delta R$ . Both the Higgs boson and the  $t$ -quark are possible origins. Secondly, a set of variables with high discrimination power is established followed by an optimisation of the **BDT**'s hyperparameters. It is important to note that the **BDT** only uses positively weighted events due to the algorithm's limitations. Each lepton is assigned a probability for both possible origins. Subsequently, a threshold in **BDT** output is then chosen for the association resulting in a correct label in 88.4 % of events.

### 6.2.2 Top quark reconstruction

The expected decay products of the  $t$ -quark are a  $b$ -jet, a light lepton from the  $W$  decay and a neutrino. The  $b$ -jet with the largest  $p_{\text{T}}$  is associated to the  $t$ -quark and the light lepton is the remaining one after the association to the Higgs boson. The neutrino results in an unknown contribution to  $E_{\text{T}}^{\text{miss}}$ . While the majority of missing energy can be attributed to the decay of the Higgs boson and the  $t$ -quark, no clear separation between the two contributions can be drawn. However, if an estimation of one's contribution can be calculated, the other can be derived using:

$$p_{\text{T}}^{\text{miss, total}} \approx p_{\text{T}}^{\text{miss, H}} + p_{\text{T}}^{\text{miss, t}}. \quad (6.1)$$

The reconstruction of the  $t$  quark is chosen as a starting point and two constraints for its missing transverse momentum are found using truth information [91]:

$$p_{\text{T}}^{\text{miss, t}} = \frac{1\,615.98\text{ GeV}^2}{p_{\text{T}}^{\ell, t}}, \quad (6.2)$$

$$\phi^{\text{miss, t}} = \phi^{\ell, t} \pm \frac{\pi}{2}. \quad (6.3)$$

Here,  $\ell$  denotes the lepton from the decay of the  $t$ -quark. The kinematic information of the associated  $b$ -jet is utilised to solve the ambiguity in  $\phi$ . Combined with the associated  $b$ -jet and light lepton the mass of the  $t$ -quark can be reconstructed. Additionally, the  $b$ -jet associated to the  $t$ -quark indirectly defines the spectator jet as the jet that has the highest combined mass with it.

### 6.2.3 Higgs boson reconstruction

A reconstruction of the Higgs boson and its mass promises a feature with high discriminating power with respect to many background events. There are two ingredients to the reconstruction. Firstly, the associated leptons present the visible part of the signature. The second, invisible contribution is given by the neutrinos. The association of the leptons is described in section 6.2.1 while section 6.2.2 gives an estimate of the neutrino contribution via equation 6.1. Both ingredients come with large uncertainties, the impact of which can be reduced by a more sophisticated mass reconstruction method, namely the Missing Mass Calculator (MMC) [92, 93]. An essential assumption of the MMC is that the  $\tau$  leptons due to the mass of their parent particle are highly boosted. As a consequence the decay

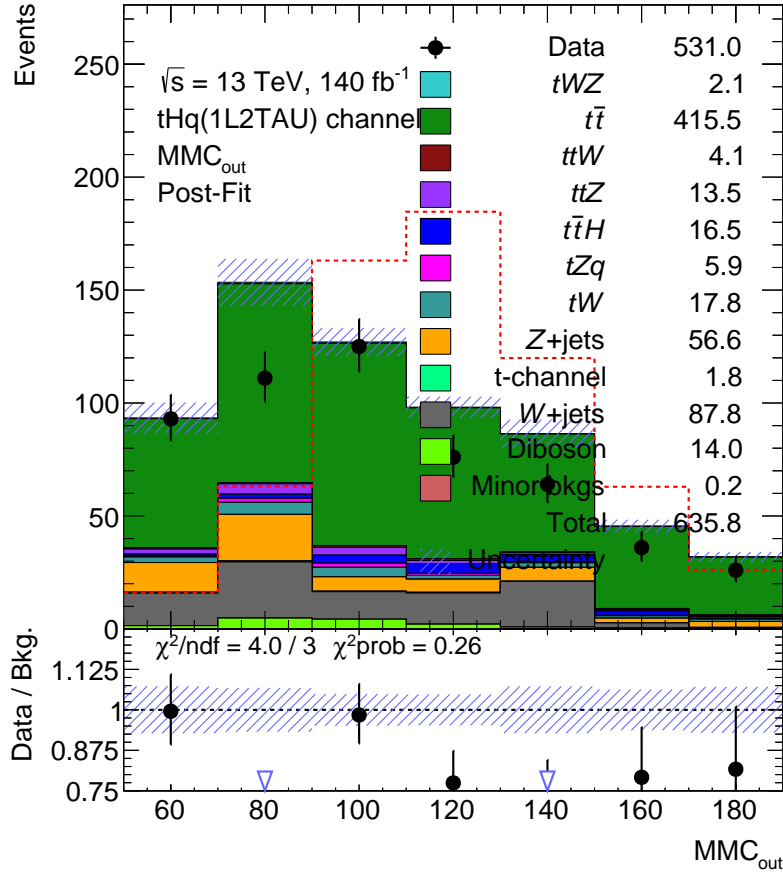


Figure 6.5: Example of the MMC output measured in GeV in the  $1\ell + 2\tau_{\text{had}}$  channel with full systematic uncertainties. The signal contribution is represented by the dashed red line. It is normalised to the total yields of the background and visibly peaks around the Higgs boson mass. Notably, the  $W+jets$  and  $Z+jets$  distributions peak around the masses of the contributing bosons.

products are highly collimated although not completely collinear leading to a small divergence  $d\theta$ . The small divergence is assumed to only depend on the decay mode and the kinematics of the  $\tau$  leptons. The MMC takes all combinations of leptonic and hadronic decay modes as well as the prongness into account. Combined with the assumption that all missing energy stems from the neutrinos in the decay, constraints on the decay products can be created. A weighted scan of the unknown parameters based on simulation samples yields a maximum likelihood estimation of the Higgs boson mass. An example of the output with full systematic uncertainties is displayed in figure 6.5 for the  $1\ell + 2\tau_{\text{had}}$  channel. The contribution of the signal sample is scaled up and peaks around the mass of the Higgs boson. Notably, the distributions of  $Z+jets$  and  $W+jets$  peak around the respective masses of the bosons.

## 6.3 Lepton fake estimation

The primary cause of the contributions of the reducible backgrounds are misidentified leptons. A lepton can either be falsely associated to an object or another object can be falsely identified as a lepton. The latter will be referred to as lepton fakes. In the event selection light leptons are required to be `tight` and  $\tau_{\text{had}}$  have to pass the `RNNMedium` criteria. These quality requirements reduce the contribution of fake  $\tau$  leptons significantly. However, the exact rates of fake leptons are hard to simulate. For this reason, data-driven techniques are applied to correct the rates and estimate the associated uncertainties. A template fit method is applied to each analysis channel to estimate the rates of  $\tau_{\text{had}}$  and light lepton fakes in combination with the respective uncertainties. This estimation is based on control regions and an improved agreement between data and `MC` can be the consequence. Generally, the agreement is expected to be within the resulting uncertainties and a net increase in disagreement means that either the simulation is flawed in some other way or the pure factor calculated from the control region is not completely applicable. This is not generally worrisome as the provided uncertainty can still be correctly used in the subsequent steps of the analysis. In every channel a template fit method or the simpler counting method is applied. The idea is always the same and the complexity is defined by the available event count and the associated statistical limitations. In each case essentially the same four steps are followed [94]:

1. In a first step, template fit regions need to be chosen. These regions need to be orthogonal, meaning statistically independent, to the selection region and contain a high fraction of lepton fakes. This is achieved by inverting the quality selections on the respective leptons.
2. These regions are then subdivided in bins of observables that are expected to have different fake rates. This step is limited by statistical uncertainties because each region is required to have a sufficient event count.
3. To allow a scaling of the fake contribution the events containing fake leptons need to be labelled accordingly. The matching is performed through the association of truth level objects to reconstruction level objects via a cone in  $\Delta R$ . For light leptons the `IFFTruthClassifier` and for  $\tau$  leptons the `TauTruthMatchingTool` is used. Both tools are analysis internal methods designed to find the origin of the respective leptons.
4. Finally, the simulated fake rates of falsely reconstructed objects are scaled to match data in all control regions. If the number of regions was previously limited by statistical uncertainties, the fit results can be controlled in more regions to verify the scaling.

It is important to emphasise that any significant disagreement between data and `MC` in the control regions is expected to come from lepton fakes. Otherwise, the background estimation is flawed. This will show in an `MLE` via the impact of other statistical uncertainties. In the following the details per channel are listed with the resulting factors summarised in appendix `D`. In all cases an example of the impact of the scale factors is presented for the  $\eta$  distribution of the leading  $\tau_{\text{had}}$ .

The  $2\ell \text{ OS} + 1\tau_{\text{had}}$  channel has the highest yields and is therefore least limited by statistical uncertainties. This means that the most in-depth fake rate estimation is possible here. In a first step, the normalisation of  $\tau$  fakes is estimated by scaling the fake events in `MC` to match data. This is done in bins of prongness, number of `b-jets` and  $p_T$ . Secondly, the shape of contributions for gluon- and quark-initiated  $\tau_{\text{had}}$  fakes are estimated via separate templates. The distinction of the templates is

achieved by means of a **BDT** which employs observables that are different for quark and gluon jets. For the fit the same bins as in step 1 are used. To conclude, the light lepton fake rates are corrected for electrons and muons. The fit is performed in regions of each possible combination of leading or subleading lepton fakes, separated in number of  $b$ -jets. The final impact of the factors is exemplified in figure 6.6. After the scaling the agreement between data and **MC** improves.

In the  $2\ell$  SS +  $1\tau_{\text{had}}$  channel the  $\tau$  fakes are estimated in regions of prongness, number of  $b$ -jets and  $p_T$ . No additional study of the quark- and gluon-initiated jets is performed due to statistical limitations. The light lepton fake rates are estimated in regions of fake lepton number only. The final impact of the factors is exemplified in figure 6.7.

In the  $1\ell + 2\tau_{\text{had}}$  channel the hadronic  $\tau$  fake rates are estimated in regions of prongness, number of  $b$ -jets and number of fake  $\tau$  leptons. No distinction between gluon- and quark-initiated fakes is made due to statistical limitations. Subsequently, the light lepton fake rates are estimated in one inclusive region to be  $1.40 \pm 0.31$  and verified in bins of  $p_T$ . The final impact of the factors is exemplified in figure 6.8. The corrected **MC** gets a larger tension with the data in the preselection region. But this tension is within the uncertainty bands. The dominant impact is statistical.

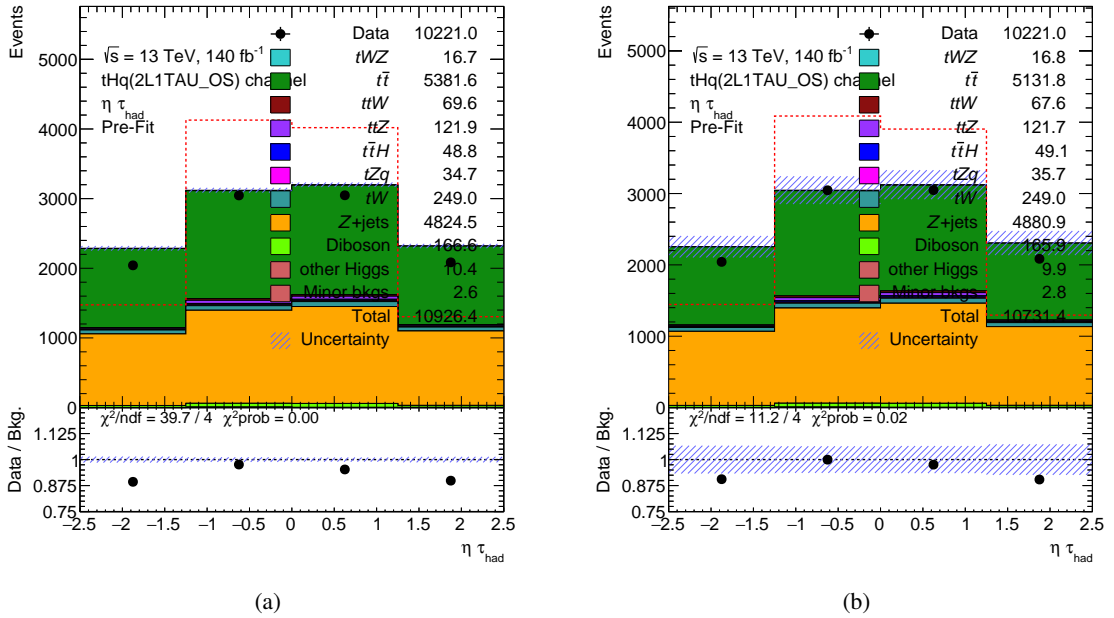


Figure 6.6: Pre- (a) and post-lepton-fake-fit (b)  $\eta$  distributions of the  $\tau_{\text{had}}$  in  $2\ell$  OS +  $1\tau_{\text{had}}$  with statistical and full (statistical and systematic) uncertainties respectively. The **MC** is represented by the stacked histograms and the data is represented by the black dots. The Pre-Fit label in both plots corresponds to the final **MLE**. After the scaling the agreement improves.



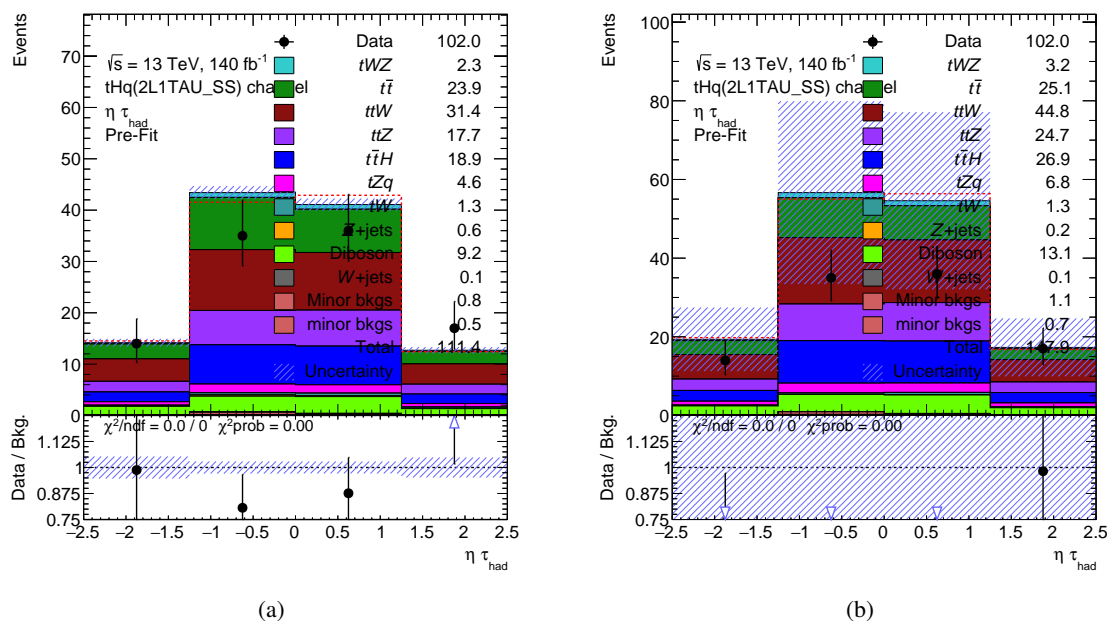


Figure 6.7: Pre- (a) and post-lepton-fake-fit (b)  $\eta$  distributions of the  $\tau_{\text{had}}$  in  $2\ell \text{ SS} + 1\tau_{\text{had}}$  with statistical and full (statistical and systematic) uncertainties respectively. The MC is represented by the stacked histograms and the data is represented by the black dots. The Pre-Fit label in both plots corresponds to the final MLE. After the scaling the agreement worsens, implying a poor agreement between the fake rates in the signal region and the template fit region.

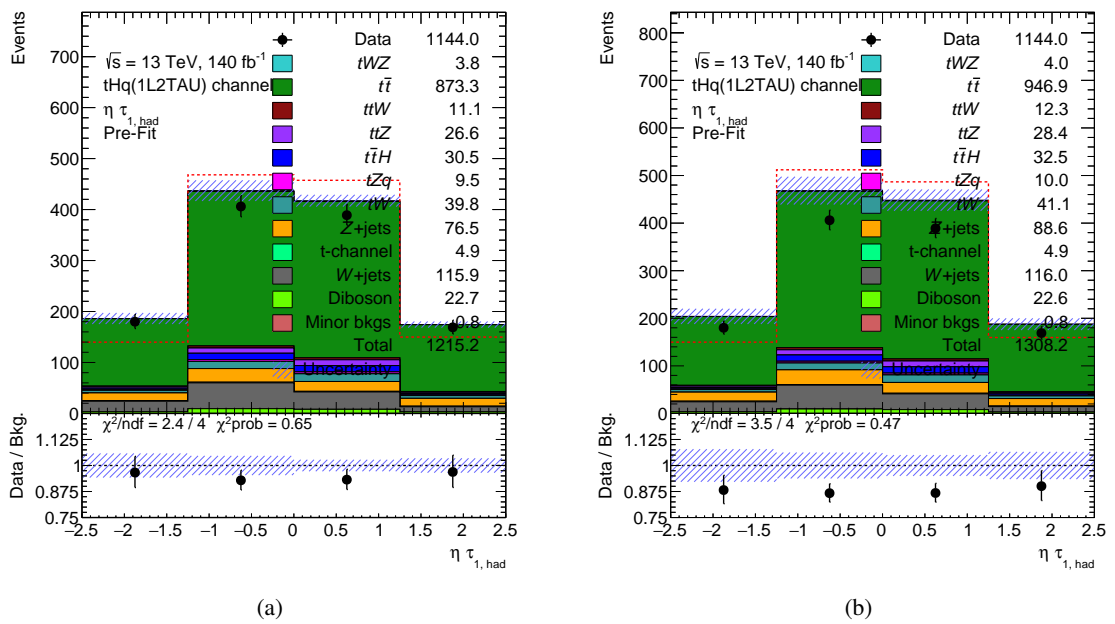


Figure 6.8: Pre- (a) and post-lepton-fake-fit (b)  $\eta$  distributions of the leading  $\tau_{\text{had}}$  in  $1\ell + 2\tau_{\text{had}}$  with statistical and full (statistical and systematic) uncertainties respectively. The MC is represented by the stacked histograms and the data is represented by the black dots. The Pre-Fit label in both plots corresponds to the final MLE. After the scaling the agreement worsens, implying a poor agreement between the fake rates in the signal region and the template fit region.

---

## Signal isolation

---

Applying the preselection to the events achieves a limitation of background processes and an overall improvement of signal to background ratio. Especially for smaller signal processes more involved methods of signal isolation become a necessity. The available phase space is therefore further exploited for variables with signal isolating properties generating two main benefits. The first and obvious benefit is an additional increase in signal to background ratio. Secondly, the output variables of the NN commonly yield distributions that are less correlated for signal and background processes. This disentanglement makes them perfect regions for a later MLE.

An advanced way of designing variables with high separation power is the usage of an artificial NN for process classification. This chapter describes the employed methods for the NN beginning with the initial design decisions. The documentation of input features and regularisation methods leads to an introduction to the hyperparameter optimisation. The final NNs for each channel are then reported on. All NNs in this work were designed using Keras with TensorFlow backend [82, 95]. The next chapter demonstrates the advantages of the resulting distributions for the concluding cross section estimation.

### 7.1 Neural network design

There is a broad range of design decisions available for the setup of a NN. Many of these options exceed the introduction provided in this thesis, but even within the described options manifold decisions need to be made. The key design choices are summarised below.

Firstly, the presented NNs are fully connected feed-forward networks. That means that information is only propagated once and the impact is calculated backwards. Kinematic features of the final state directly present in the samples are chosen as input features and normalised via Lecun normalisation<sup>1</sup> [96]. The hidden layers are chosen to contain the same number of nodes each. There are arguments for different structures of nodes. However, since initial tests did not yield a strong dependence on the distribution of the nodes over the layers an equal distribution is chosen. The largest benefit is an easier automation of the hyperparameters' optimisation in the later process. The output is chosen to be three

---

<sup>1</sup> Lecun normalisation works best in combination with a matching version of activation function and dropout. The details are omitted for being purely technical. The interested reader is strongly encouraged to read the detailed description in the reference.

categories, signal, generic background and specific background. More options were tested but did not show a significant improvement in the results [97]. It is true, that a small dependence on the number of categories is recorded. Nevertheless, the resulting networks are less stable and stability is a core feature of a method that needs to be rerun during an ongoing analysis. Each category is then weighted according to the event weights and subsequently rescaled to correspond to a third of the input each. This reweighting is a necessity in order to sufficiently weight underrepresented categories in the classification. The usage of event weights in the training is debatable because even one raw event always corresponds to correct physics. The drawback is that strongly weighted events can dominate the training. The training uses the absolute magnitude of the event weights to cancel out the impact of the negative weights without losing the statistical power of the simulated events completely. Additionally a large and static test sample of 30 % is chosen for validation purposes. This sample is never used for training. An alternative would be a k-fold cross validation [80]. However, due to the divergence between raw and weighted yields for some simulations, the decision to keep the training sample as large as possible was made. An additional cross validation was run, just for stability report purposes.

## 7.2 Choice of features

The big incentive to use a **NN** is the confidence that the algorithm will independently find composite variables that optimise the isolation of a signal. This statement is based on the fact that an ideal **NN** can model any function. Another consequence of this statement is that a sufficient **NN** can perform a principle axis transformation on its own. This means that all meaningful correlations can be maximally exploited while on the other hand correlations in the input are no hindrance to the algorithm.

The choice made for this analysis is a set of features mainly describing the kinematic properties of the final state objects. The choice is very similar for all three analysed channels and summarised in tables 7.1 to 7.3. Part of the features are chosen on reconstruction level but the reconstructed properties of the Higgs boson and the  $t$ -quark are also taken advantage of. A noteworthy addition is the combined mass of the b-jet and the spectator jet which proved too powerful in separation power to leave out.

A perfect ranking of features based on their separation power in a **NN** is not achievable given the strongly entangled inputs. This is because the network, much less than for example a BDT, is based on combining the provided knowledge and exploiting smaller bits of information. Typically, a ranking would be done by removing one feature from the model iteratively. The drop in performance should be anti-proportional to the importance of the variable. Doing this, a batch of features can be excluded. However, for the remaining sample no definite ranking is possible. Excluding numerous less influential variables also contradicts the principle of the method. After all, a neural network is used to maximise the usage of smaller bits of information. In this way, information that would be difficult to analyse manually can be included.

In addition, a set of Lorentz-invariant variables is tested, that describes the event almost completely. Based on the assumption that a **NN** can mimic any function, a set of Lorentz-invariant variables is a popular approach. Only limited by the uncertainties of the reconstruction, the variables allow a complete description of the event. Optimally, a neural network should infer all further variables from the given information. The tests in this work did not show a superior or even comparable performance for Lorentz-invariant variables. An example is shown for the  $1\ell + 2\tau_{\text{had}}$  channel in appendix I. For an in-depth description of Lorentz-invariant variables and their possible exploitation in the  $tZq$  analysis,

see [98].

Table 7.1: Input features of the  $2\ell$  SS +  $1\tau_{\text{had}}$  channel NN.

Raw feature name	Description
eta_jf	$\eta$ of the spectator jet
pt_jf	$p_T$ of the spectator jet
phi_jf	$\phi$ of the spectator jet
eta_jet1	$\eta$ of the leading jet
pt_jet1	$p_T$ of the leading jet
phi_jet1	$\phi$ of the leading jet
eta_b	$\eta$ of the leading b-jet
pt_b	$p_T$ of the leading b-jet
phi_b	$\phi$ of the leading b-jet
MMC_out_1	MMC estimation of the Higgs mass
m_met	Total $E_T^{\text{miss}}$
had_tau_pt	$p_T$ of the $\tau_{\text{had}}$
had_tau_eta	$\eta$ of the $\tau_{\text{had}}$
had_tau_phi	$\phi$ of the $\tau_{\text{had}}$
deltaRTau	$\Delta R$ of the associated $\tau$ leptons
deltaPhiTau	$\Delta\phi$ of the associated $\tau$ leptons
HvisPt	$p_T$ of Lorentz-vector sum of the associated $\tau$ leptons
HvisEta	$\eta$ of Lorentz-vector sum of the associated $\tau$ leptons
TvisPt	Visible $p_T$ of the reconstructed $t$ -quark
TvisEta	Visible $\eta$ of the reconstructed $t$ -quark
M_b_jf	Mass of the Lorentz-vector sum of the leading b-jet and spectator jet
HT	Total energy in the transverse plane
lep_Top_pt	$p_T$ of the $t$ associated light lepton
lep_Top_eta	$\eta$ of the $t$ associated light lepton
lep_Top_phi	$\phi$ of the $t$ associated light lepton
lep_Higgs_pt	$p_T$ of the Higgs associated light lepton
lep_Higgs_eta	$\eta$ of the Higgs associated light lepton
lep_Higgs_phi	$\phi$ of the Higgs associated light lepton

### 7.3 Choice of regularisation

The NN models described here were trained in parallel to the ongoing analysis efforts. In consequence, any optimised model has to be adapted if changes to the analysis process are made. The reasons can be manifold, including a change in samples or weights. Any resulting model should be stable and robust with respect to changes in the input information. Should this not be the case, the re-optimisation process takes a significant amount of time and brings the analysis to a standstill. The full algorithm for optimisation is described in section 7.4. In general a decision is made to use a combination of dropout and batch normalisation. The result is a set of NN models that perform similarly on a range of samples and require little to no re-optimisation upon small changes in the input.

The parameter stabilised in this way is the generalisation of the NN model commonly controlled via the loss curves. While the interpretation of the loss is already discussed in section 5.3.2, it is worthwhile to put a few things into context. The starting point of the discussion needs to be the generalisability of the training and test sample. A perfect training sample flawlessly reflects the truth.

Table 7.2: Input features of the  $2\ell$  OS +  $1\tau_{\text{had}}$  channel NN.

Raw feature name	Description
eta_jf	$\eta$ of the spectator jet
pt_jf	$p_{\text{T}}$ of the spectator jet
phi_jf	$\phi$ of the spectator jet
eta_jet1	$\eta$ of the leading jet
pt_jet1	$p_{\text{T}}$ of the leading jet
phi_jet1	$\phi$ of the leading jet
eta_b	$\eta$ of the leading b-jet
pt_b	$p_{\text{T}}$ of the leading b-jet
phi_b	$\phi$ of the leading b-jet
MMC_out_1	MMC estimation of the Higgs mass
m_met	Total $E_{\text{T}}^{\text{miss}}$
had_tau_pt	$p_{\text{T}}$ of the $\tau_{\text{had}}$
had_tau_eta	$\eta$ of the $\tau_{\text{had}}$
had_tau_phi	$\phi$ of the $\tau_{\text{had}}$
deltaRTau	$\Delta R$ of the associated $\tau$ leptons
deltaPhiTau	$\Delta\phi$ of the associated $\tau$ leptons
HvisPt	$p_{\text{T}}$ of Lorentz-vector sum of the associated $\tau$ leptons
HvisEta	$\eta$ of Lorentz-vector sum of the associated $\tau$ leptons
TvisPt	Visible $p_{\text{T}}$ of the reconstructed $t$ -quark
TvisEta	Visible $\eta$ of the reconstructed $t$ -quark
M_b_jf	Mass of the Lorentz-vector sum of the leading b-jet and spectator jet
HT	Total energy in the transverse plane
lep_Top_pt	$p_{\text{T}}$ of the $t$ associated light lepton
lep_Top_eta	$\eta$ of the $t$ associated light lepton
lep_Top_phi	$\phi$ of the $t$ associated light lepton
lep_Higgs_pt	$p_{\text{T}}$ of the Higgs associated light lepton
lep_Higgs_eta	$\eta$ of the Higgs associated light lepton
lep_Higgs_phi	$\phi$ of the Higgs associated light lepton

Accordingly, a perfect test sample mirrors anything truthfully represented by the training sample. In this case any divergence between training and test loss curves implies overtraining. Usually, given the extreme statistic uncertainties, this assumption can neither be made for the training nor for the test sample. Instead, a decrease in training loss and a stable test loss can imply that the NN has learnt features only present in the training sample. This is not a typical example for overtraining. Nevertheless, it signals a poor understanding of the model and should be avoided. The actual worst case is an increase of test loss. This means that the NN learns features that are false in the test sample. This behaviour is always and strictly to be avoided. All things considered, it is a common mistake to call any amount of divergence in the loss curves overtraining. The divergence is only then most problematic when the loss of the test sample increases significantly. This is combined with a secondary misconception being that any amount of overtraining in a NN model is a mistake in an analysis. Generally this is not true in the presented type of analysis because the model is not used for direct inference of a statement.

Instead, the NN model is used to redistribute the simulated events based on the provided input variables. If no overtraining is recorded and all tests point at a well generalised model, this redistribution of events is similar in the data sample. It is still possible that the agreement between data and MC becomes worse for the response curves of the NN. The reason is then likely to be the

Table 7.3: Input features of the  $1\ell + 2\tau_{\text{had}}$  channel NN.

Raw feature name	Description
eta_jf	$\eta$ of the spectator jet
pt_jf	$p_T$ of the spectator jet
phi_jf	$\phi$ of the spectator jet
eta_jet1	$\eta$ of the leading jet
pt_jet1	$p_T$ of the leading jet
phi_jet1	$\phi$ of the leading jet
eta_b	$\eta$ of the leading b-jet
pt_b	$p_T$ of the leading b-jet
phi_b	$\phi$ of the leading b-jet
MMC_out_1	MMC estimation of the Higgs mass
m_met	Total $E_T^{\text{miss}}$
had_tau_1_pt	$p_T$ of the leading $\tau_{\text{had}}$
had_tau_1_px	$p_x$ of the leading $\tau_{\text{had}}$
had_tau_1_py	$p_y$ of the leading $\tau_{\text{had}}$
had_tau_1_E	Energy of the leading $\tau_{\text{had}}$
had_tau_1_eta	$\eta$ of the leading $\tau_{\text{had}}$
had_tau_1_phi	$\phi$ of the leading $\tau_{\text{had}}$
had_tau_2_pt	$p_T$ of the subleading $\tau_{\text{had}}$
had_tau_2_px	$p_x$ of the subleading $\tau_{\text{had}}$
had_tau_2_py	$p_y$ of the subleading $\tau_{\text{had}}$
had_tau_2_E	Energy of the subleading $\tau_{\text{had}}$
had_tau_2_eta	$\eta$ of the subleading $\tau_{\text{had}}$
had_tau_2_phi	$\phi$ of the subleading $\tau_{\text{had}}$
deltaRTau	$\Delta R$ of the associated $\tau$ leptons
deltaPhiTau	$\Delta\phi$ of the associated $\tau$ leptons
HvisPt	$p_T$ of Lorentz-vector sum of the associated $\tau$ leptons
HvisEta	$\eta$ of Lorentz-vector sum of the associated $\tau$ leptons
TvisPt	Visible $p_T$ of the reconstructed $t$ -quark
TvisEta	Visible $\eta$ of the reconstructed $t$ -quark
M_b_jf	Mass of the Lorentz-vector sum of the leading b-jet and spectator jet
HT	Total energy in the transverse plane
lep_Top_pt	$p_T$ of the $t$ associated light lepton
lep_Top_eta	$\eta$ of the $t$ associated light lepton
lep_Top_phi	$\phi$ of the $t$ associated light lepton
lep_Higgs_pt	$p_T$ of the Higgs associated light lepton
lep_Higgs_eta	$\eta$ of the Higgs associated light lepton
lep_Higgs_phi	$\phi$ of the Higgs associated light lepton

exploitation of previously poorly modelled information in the **MC**. In a perfect scenario this is not the case and the risk of it is reduced by controlling the input feature for good behaviour (in shape and data to **MC** agreement). However, during the final fit and cross section estimation, the truth labelling of the simulations can be used. This means the disagreement is not hidden from the fit. Accordingly, even a faulty model does not result in a false statement in the final **MLE**. Rather, the quality of the statement is worsened. This does not mean that overtraining is desirable. The consequence is merely a more careful handling of the interpretation.

During the presented optimisation, the training is stopped based on divergence between training and test loss. Thereby any unwanted behaviour of the model can be excluded. To avoid ambiguities by saving the model at a previous stage in training, the optimal model is retrained and re-controlled for the optimised number of epochs. Afterwards the response curves and the **ROC** curves are additionally controlled for their training and test agreement as documented in section 7.5.

## 7.4 Hyperparameter optimisation

The associated production of a  $Z$  boson and a  $t$ -quark, in short  $tZq$ , can be considered the predecessor analysis to  $tHq$  [99]. It has a larger cross section but a similar final state. Many of the insights gained in the  $tZq$  analysis were reused for the presented work. The basis of the presented **NN** framework is an example for the said exploitation of previous experience. Initially, a shallow neural network from the NeuroBayes package was used for signal isolation [100]. Afterwards a Keras-based **NN** was optimised to isolate a  $tZq$  signal from its backgrounds. In that case, an extensive grid search of the hyperparameters was performed to maximise the model's performance. Assuming that the tasks are comparable in magnitude, the knowledge obtained was thus used as a basis for optimising the  $tH$  signal isolation. This means that the optimised hyperparameters are used as a starting point for further optimisation and it is assumed that a slightly increased capacity with the same order of magnitude will yield the optimal results.

Of course, the obtained hyperparameters have to be confirmed. However, another pure grid search is a blunt and therefore computationally expensive method. A valid alternative is a class of algorithms that finds its inspiration in genetic evolution. This is comical yet appropriate given the analogy drawn to the human brain for the **NN**. This evolutionary algorithm starts from a pool of models seeded with random hyperparameters. The models are then evaluated, ranked and recombined based on their performance. The specific algorithm was written to harmonise with the computing cluster in Bonn [101] to minimise computing times. The four phases are:

1. An initial group of **NNs**, termed a generation, is generated with random hyperparameters. The range of initial hyperparameters is based on the former optimisation for the  $tZq$  signal.
2. The generation of **NNs** is trained and ranked based on their **AUC**.
3. Based on the ranking a new generation is created. Firstly, the best **NN** is always kept. Secondly, the hyperparameters of the top-performing 50 % of **NNs** are mixed to form new **NNs**. In a last step, the resulting hyperparameters are randomly changed by up to 20 %. This is referred to as *mutation* and avoids local minima in the evolution
4. The new generation is trained and its ranking forms the basis of the next step of the process.



The number of generations needs to be chosen by the user. Since the process requires the training of many NNs it takes a significant amount of time.<sup>2</sup> Therefore, the evolution is stopped after five generations. Tracking the highest ranking NN shows that the AUC becomes almost stable after two generations as demonstrated in figure 7.1 for an extensive 8 generations. This has two consequences: the duration of five generations can be assumed sufficient and the acquired hyperparameters should correspond to a stable model. The full set of final NNs is analysed and used as input ranges for a fine grid search. The initial and resulting ranges of hyperparameters are summarised in table 7.4.

Table 7.4: Tested and final ranges of hyperparameters in the evolutionary optimisation. Due to the mutation step the initial range can be exceeded further in every generation

Hyperparameter	Range	Resulting range
Layers	[1, 10]	[5, 8]
Nodes	[1, 200]	[60, 170]
Dropout fraction	[0, 1]	[0.2, 0.8]
Learning rate	[0.0001, 1]	0.001
Optimiser	Adam, SGD	Adam

During the final optimisation, an increased focus is placed on the stability and generalisability of the models. One important ingredient to the stability was the inclusion of both batchnormalisation and a significant dropout fraction.<sup>3</sup> The resulting hyperparameters for each channel are summarised in tables 7.5 to 7.7. Lastly, it is worth noting that the batchsize was determined after the optimisation by hand. Especially for channels with lower event yields, a larger batchsize can reduce fluctuations. A suitable example is given by a comparison of raw and weighted yields for  $t\bar{t}$  and  $tHq$  in the  $2\ell$  SS +  $1\tau_{\text{had}}$  channel summarised in table 6.3.  $tHq$  contributes about 15 000 raw events that are translated into roughly 1 weighted event. However, for  $t\bar{t}$  only 200 raw events result in 20 weighted events. Therefore, the event weights for the  $t\bar{t}$  events must be larger than for  $tHq$  by three orders of magnitude. Even for a reweighting this effect will be visible in smaller batches and as a result a training can become unstable. To reduce this effect, the batchsize was increased step-wise until either no further growth in stability or a decrease in performance is observed.

## 7.5 Model evaluation

The previous sections provide documentation on the design decisions for both setup and subsequent optimisation. This section concludes the description with the full performance report for all three channels. The final hyperparameters are listed in tables 7.5 to 7.7. While the hyperparameters were generally a pure result of the evolutionary optimisation, the choice of specific background category is different for each channel. This is due to the specific background composition in each case. In  $1\ell + 2\tau_{\text{had}}$ ,  $t\bar{t}$  is the dominant background by far and the obvious choice for a category. For  $2\ell$  OS +  $1\tau_{\text{had}}$ , Z+jets is chosen because  $t\bar{t}$  showed less stable training results, likely due to the low number of raw events. Lastly, due to the generally low event yields in  $2\ell$  SS +  $1\tau_{\text{had}}$  the  $t\bar{t}$ +X processes are combined as a specific category. In all cases the remaining backgrounds are given a common

<sup>2</sup> The main effort was done before the wider availability of GPUs in the Bonn cluster. If GPUs should be available a longer optimisation of a more advanced algorithm can become feasible.

<sup>3</sup> See section 5.4.2 for the benefits of batchnormalisation beyond its regularising effect.

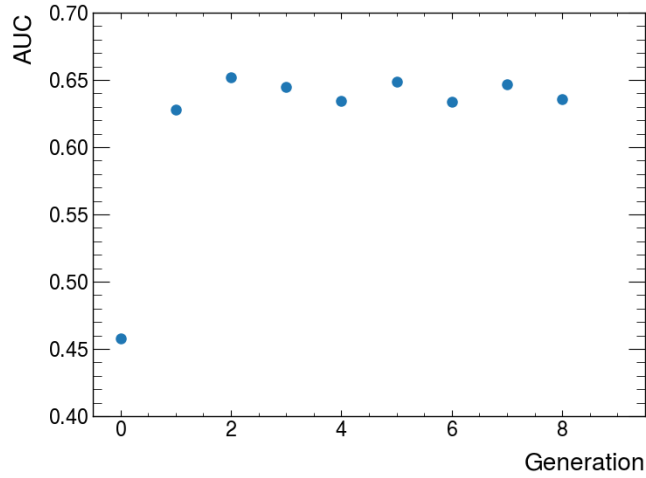


Figure 7.1: Development of the  $AUC$  of the  $1\ell + 2\tau_{\text{had}}$  channel over 8 generations of evolutionary optimisation. It becomes apparent that the  $AUC$  stagnates after one or two generations.

Table 7.5: Optimised hyperparameters of the NN for the  $1\ell + 2\tau_{\text{had}}$  channel.

Hyperparameter	Optimised value
Number of layers	6
Nodes per layer	120
Dropout fraction	20 %
Batchnormalisation	True
Hidden activations	elu
Output activation	softmax
Optimiser	Adam
Learning rate	0.001
Decay	False
Batchsize	25 000
Epochs	260
Category	$i\bar{i}$

Table 7.6: Optimised hyperparameters of the NN for the  $2\ell \text{ OS} + 1\tau_{\text{had}}$  channel.

Hyperparameter	Optimised value
Number of layers	6
Nodes per layer	120
Dropout fraction	65 %
Batchnormalisation	True
Hidden activations	elu
Output activation	softmax
Optimiser	Adam
Learning rate	0.001
Decay	False
Batchsize	1 000
Epochs	200
Category	Z+ jets

Table 7.7: Optimised hyperparameters of the NN for the  $2\ell$  SS +  $1\tau_{\text{had}}$  channel.

Hyperparameter	Optimised value
Number of layers	7
Nodes per layer	70
Dropout fraction	20 %
Batchnormalisation	True
Hidden activations	elu
Output activation	softmax
Optimiser	Adam
Learning rate	0.001
Decay	False
Batchsize	100 000
Epochs	300
Category	$t\bar{t}+X$

target in the training. The first part of the report is given by the loss curves and the associated stability of the training. Provided a good stability, the ROC curve and the response distributions are evaluated for their classification performance. In addition, the agreement between the train and test curves allows a further examination of the stability. The response and ROC curve of the signal category are directly shown in this section while the distributions for the other two categories are exported to appendix H. In general, the response for each category is plotted against the cumulative response of the remaining two categories. For the signal category, this cumulative response corresponds to the overall background response (including general and specific background). In the other cases, the cumulative distribution is denoted as the *orthogonal* sample. Figure 7.8 shows the process-wise signal responses with full uncertainties, as introduced in section 8.2, for all three channels. The respective distributions for the categorical and background responses are shown in appendix H in figures H.7 and H.8. This style of plot while slightly more convoluted increases the understanding of the behaviour of distinct processes in the model. While the training is performed on absolute weights response and ROC curve are plotted using the true weights. A final report of the impact of the event weights on the training thus concludes the summary.

$2\ell$  OS +  $1\tau_{\text{had}}$  has the highest raw yields of the three channels by far. This results in a very stable training depicted in figure 7.2. Any divergence between the train and test loss is due to the dropout percentage only being applied to the training model. The response and AUC for the  $tH$  signal is depicted in figure 7.5. In both cases the training and test agreement is excellent. The responses are very well separated promising to be an efficient cut variable for later region definition. The same is true for the background and  $Z+$  jets results depicted in figure H.1 and figure H.4.

$2\ell$  SS +  $1\tau_{\text{had}}$  has the lowest raw yields and the lowest weighted yields by far. Additionally, the divergence between raw and weighted yields is the largest in this channel. This puts some constraints on the performance of the NN. The NN in this channel required intensive re-optimisation after each change in the analysis. The achieved loss curve and its stability shown in figure 7.3 is an outstanding result because, given the low raw yields in the channel, large fluctuations in the loss curve are common for other models. Generally, the responses and ROC curves (figures 7.6, H.2 and H.5) for all three categories slightly underperform with respect to the other two channels. However, especially the results for the  $t\bar{t}+X$  processes show a strong separation power with an AUC of 0.82 and used as a cut selection allows for a further exploitation of the preselection region. The visible fluctuations in

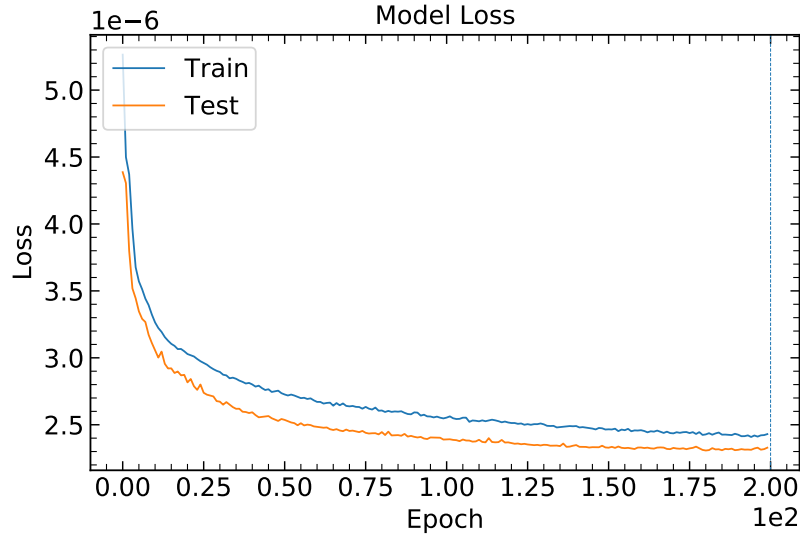


Figure 7.2: Loss curve of the  $2\ell$  OS +  $1\tau_{\text{had}}$  NN training. The loss shows neither worrisome signs of overtraining nor strong fluctuations. An initial superior performance of the test sample is due to the application of the dropout percentage to the training evaluation only.

the response curve are a consequence of the low yields and the negative event weights and cannot be controlled perfectly provided the strong statistical limitations.

The raw and weighted yields in the  $1\ell + 2\tau_{\text{had}}$  channel are smaller than in the  $2\ell$  OS +  $1\tau_{\text{had}}$  channel by a factor of 5 and 10 respectively (refer to tables 6.1 to 6.3). This is still sufficient for a stable loss result as shown in figure 7.4. Response and ROC curve for the  $tH$  signal are depicted in figure 7.7. The additional two categories for background and  $t\bar{t}$  are summarised in figures H.3 and H.6. Proper classifications can be achieved for all three categories. Only the stability of the response curves shows the statistical uncertainty, which is increased by the weights. In general, all three categories promise to be excellent for a region definition or an MLE. The loss curves generally document satisfying stability for all three models. Nevertheless, fluctuations occur in the response distributions, especially for the channels with low raw yields. The reason for this is, on the one hand, the differences in the ratios between pure and weighted events and on the other hand, the negative weights play a role. These are not included in the training, but in the responses. With an uneven distribution across the bins of the histograms, fluctuations are unavoidable. For verification, the response distributions were plotted with absolute and correct weights. The impact is demonstrated in figure 7.9. The fluctuations decrease visibly and can therefore be clearly assigned to the negative weights. Since all events correspond to correct physics, this does not pose a problem for generalisability. It is important to remember that no direct decision is made based on the NN classification. It becomes clear that a majority of fluctuations are due to the inconsistent distribution of negatively weighted events between the bins of the distribution.

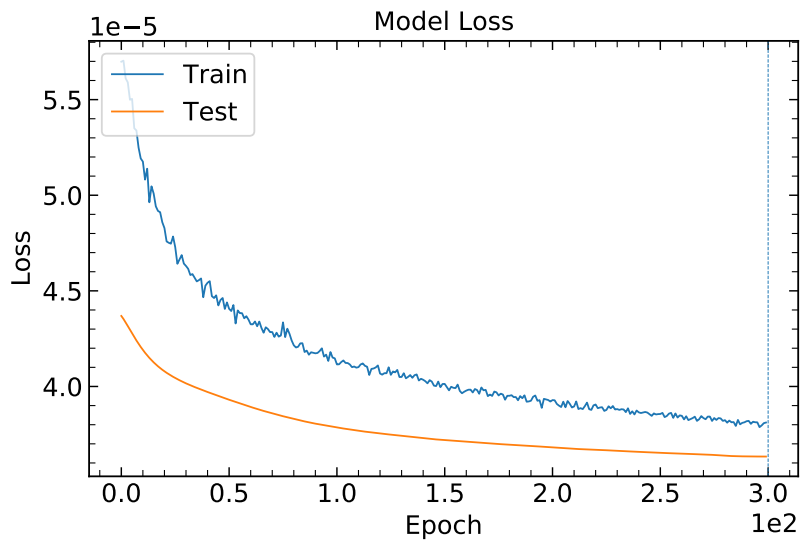


Figure 7.3: Loss curve of the  $2\ell$  SS +  $1\tau_{\text{had}}$  NN training. The loss shows neither worrisome signs of overtraining nor strong fluctuations. An initial superior performance of the test sample is due to the application of the dropout percentage to the training evaluation only.

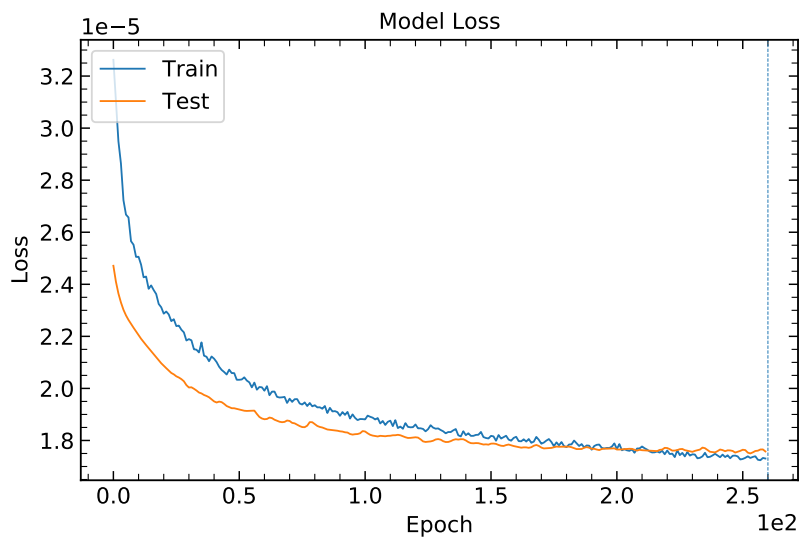


Figure 7.4: Loss curve of the  $1\ell + 2\tau_{\text{had}}$  NN training. The loss shows neither worrisome signs of overtraining nor strong fluctuations. An initial superior performance of the test sample is due to the application of the dropout percentage to the training evaluation only.

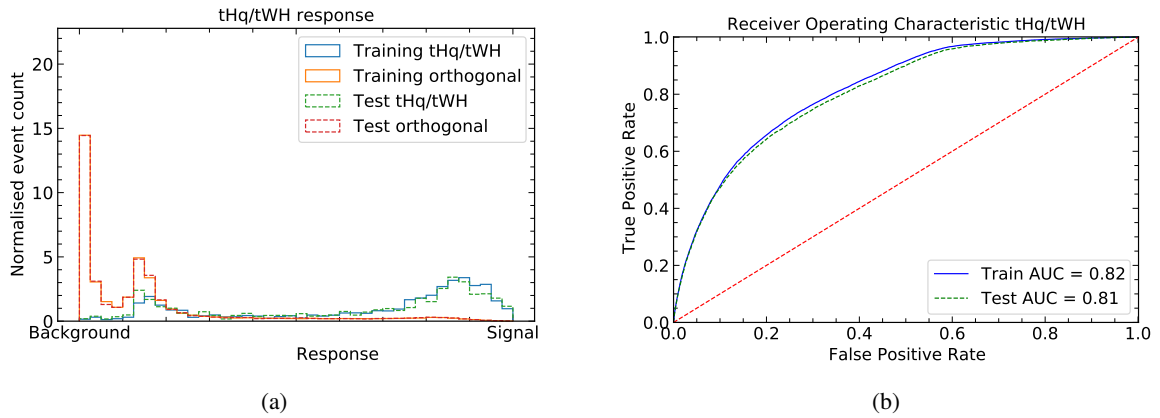


Figure 7.5: Display of the signal response (a) and the ROC (b) curve of the final model for the  $2\ell$  OS +  $1\tau_{\text{had}}$  channel.

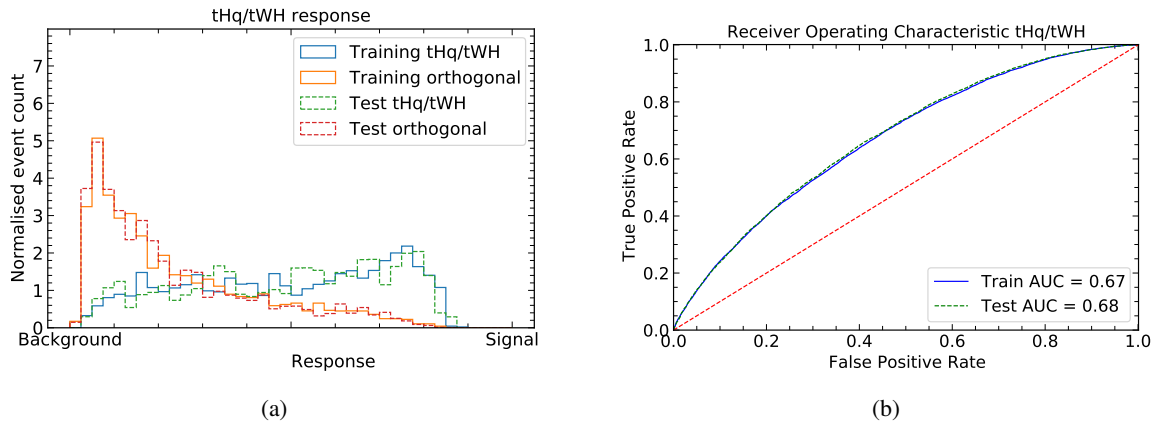


Figure 7.6: Display of the signal response (a) and the ROC (b) curve of the final model for the  $2\ell$  SS +  $1\tau_{\text{had}}$  channel.

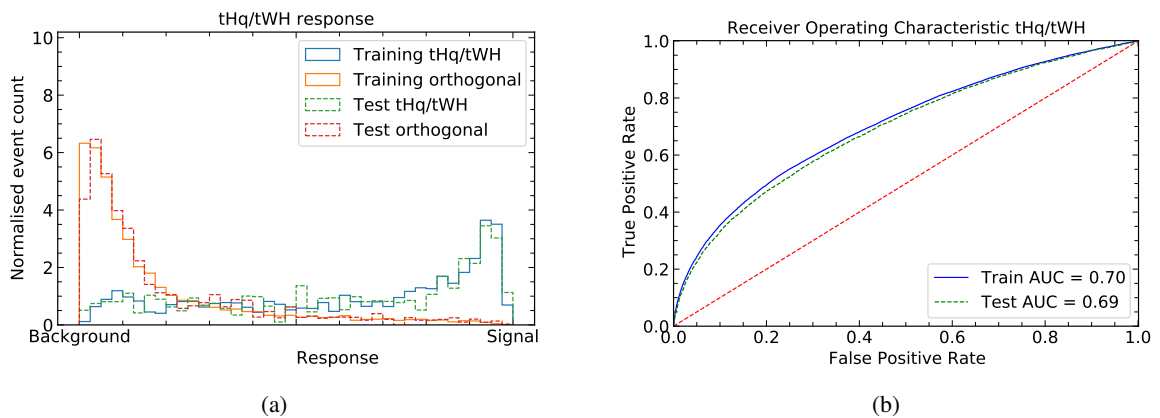


Figure 7.7: Display of the signal response (a) and the ROC (b) curve of the final model for the  $1\ell + 2\tau_{\text{had}}$  channel.

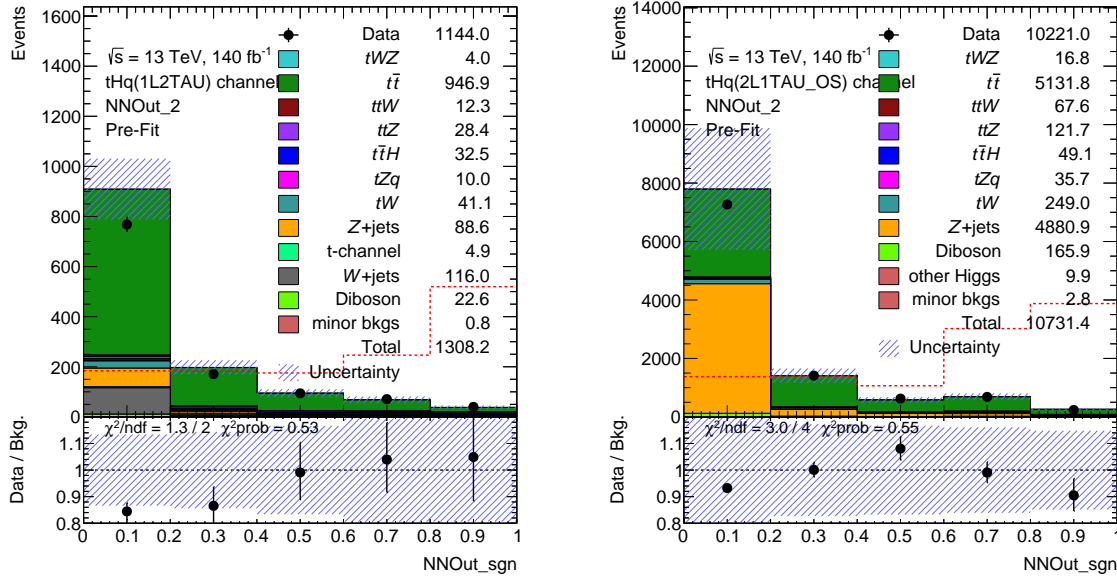
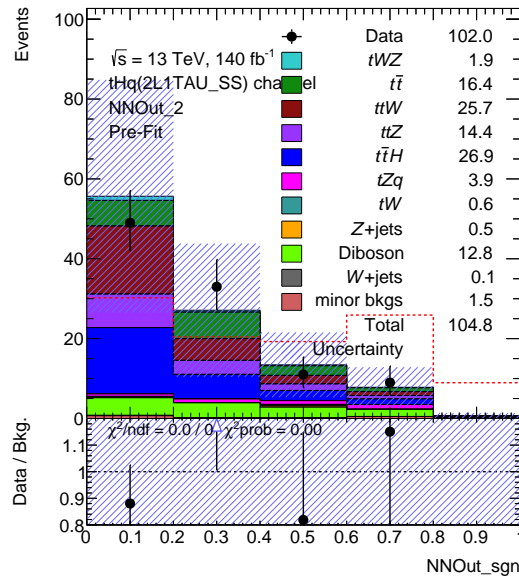
(a)  $1\ell + 2\tau_{\text{had}}$  signal response(b)  $2\ell \text{ OS} + 1\tau_{\text{had}}$  signal response(c)  $2\ell \text{ SS} + 1\tau_{\text{had}}$  signal response

Figure 7.8: Display of the process-wise response distributions of the signal responses using full uncertainties. The signal is scaled up and displayed as a red-dashed line.

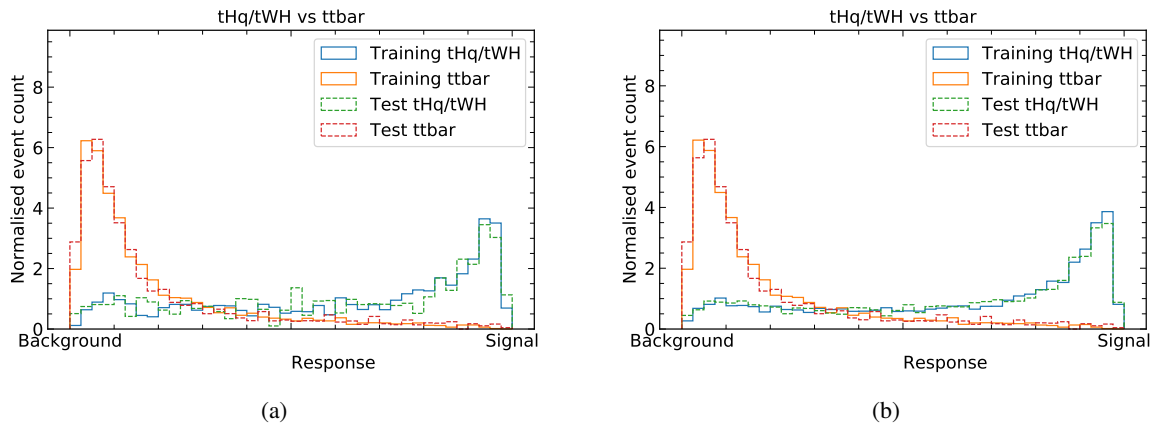


Figure 7.9: Demonstration of the impact of the negative weights on the stability of the response distributions in the  $1\ell + 2\tau_{\text{had}}$  channel. (a) shows the distributions using the correct weights while absolute weights are displayed in (b). In this example the response is displayed for signal versus  $t\bar{t}$ . It becomes apparent that the majority of fluctuations stems from the binning and thus the distributions of the negative weights.

## 7.6 K-fold cross validation

K-fold cross validation is an alternative to a fixed test sample. It makes it so that every event can be used for training and validation in turns. For this the training sample is divided into  $k$  subsets. Given the low raw yields and the large discrepancies between raw and weighted yields in all three channels, a smaller training sample causes more foreseeable instabilities than the exclusion of a fixed test sample would. Nevertheless, an additional investigation of stability over at least 5 batches was made for each channel using the final hyperparameters. Only the number of epochs has been reduced, as overtraining on small batches would otherwise be unavoidable. To quantify the generalisability of a training over different batches a ROC curve for each batch is calculated. Subsequently the mean and standard deviation is calculated. An example is displayed in figure 7.10 for the  $1\ell + 2\tau_{\text{had}}$  channel. For all three channels the fluctuation between the batches are minimal which indicates that the training sample does not contain dangerous inconsistencies. A full documentation of the studies for each channel is provided via appendix J.



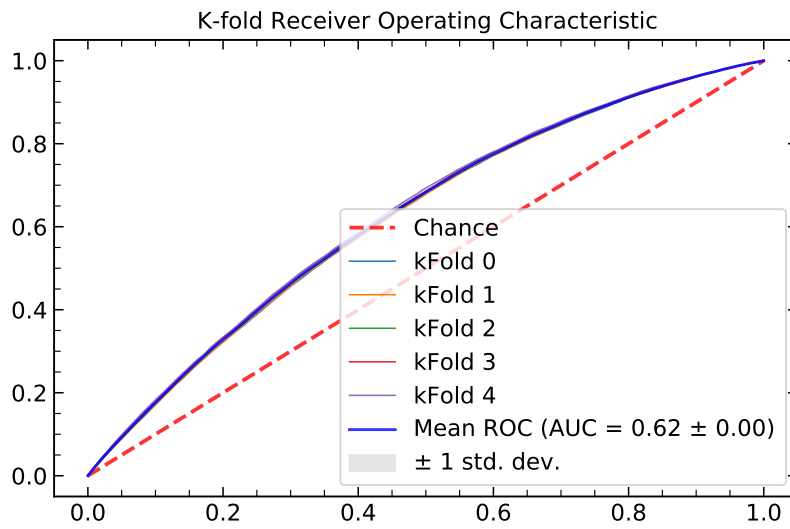


Figure 7.10: Display of the ROC curves of the k-fold cross validation in  $1\ell + 2\tau_{\text{had}}$ . The mean ROC and its standard deviation are plotted to observe significant deviations in any fold.



---

## Cross section estimation

---

The aim of this work is to obtain an estimator and the associated uncertainty for the cross section of the  $tH$ <sup>1</sup> process. For this purpose, the MC simulation based on the SM prediction is scaled to achieve an optimal match to the measured data. In the corresponding MLE fit the rate of expected signal events as well as the expected rates of certain background processes become free parameters of the fit. The values by which the fit scales them are termed **normalisation (abbreviated to norm) factors (NFs)** symbolised by the Greek letter  $\mu$ . The resulting factor can be applied to the SM prediction of the cross section to obtain an estimate.

The basis for the measurement is the signal-rich preselection region as defined in section 6, which is further split into regions and successively into bins using the NN models. The first section of this chapter documents this region definition. The second section contains the documentation of sources and treatment of uncertainties. With these prerequisites covered the fit to Asimov data is described which serves as a stability test for the final fit. Lastly, the fit to data and the final results are covered.

### 8.1 Region definition

To maximise the significance of the measurement and to minimise the influence of background processes the significance in the signal region needs to be maximised. Additionally, the expected distributions for background and signal events should be as uncorrelated as possible. Thereby, the available information can be exploited best and the uncertainty on the POI is minimised. Both are achieved by using the NN response distributions described in section 7. On the one hand, the response variables are suitable for narrowing down regions that are purer in signal or purer in other processes. On the other hand, the  $tH$  distribution in the signal response is maximally uncorrelated from the background distributions. Other features were also tested for their correlation between the processes, but the signal response of the NN offered the best separation in all cases.

In addition to the definition of the signal region, commonly one or more control regions are defined. Control regions are used to estimate behaviour of significant background contributions in the fit. To make this possible, such regions should meet two criteria. First, a control region is chosen in a phase space that is orthogonal yet similar to the signal region. Thus, it is expected that the behaviour of the

---

<sup>1</sup> In this analysis  $tHq$  t-channel production and  $tWH$  are summarised as  $tH$  signal. The s-channel production due to its small cross section is omitted.

events is comparable. Secondly, a control region should be dominated by one or more background processes and contain few to no signal events.<sup>2</sup> A control region that by this definition contains no or only very few signal events results in the correct background proportions for both an excess and an absence of signal events.

In the following the definitions of the regions are described for each channel. The criteria are based on the response vector of the respective **NN**. A display of the signal responses can be found in figure 7.8. The background responses are shown in figures H.7 and H.8.

In the  $1\ell + 2\tau_{\text{had}}$  channel the signal region is required to have a **NN**  $tH$  response greater than 0.45. This requirement is inverted for the control region events. The  $\bar{t}\bar{t}$  control region is contains all remaining events for which the background response is smaller than 0.5. The opposite selection is chosen for the  $V$ +jets control region. In the  $2\ell$  OS +  $1\tau_{\text{had}}$  channel the signal region is defined via the requirement of **NN** signal response higher than 0.4 and  $Z$ +jets response smaller than 0.1. Thereby, the region purest in  $tH$  events is defined and additionally freed from a large fraction of the remaining  $Z$ +jets events. Successively, the control regions are separated by an initial inversion of either of the two requirements. The resulting events are then further divided into a designated  $\bar{t}\bar{t}$  and a  $Z$ +jets control region. The  $\bar{t}\bar{t}$  control region corresponds to the remaining events with a background response higher than 0.6 and the inverted selection is chosen for the  $Z$ +jets control region. The  $2\ell$  SS +  $1\tau_{\text{had}}$  signal region is similarly defined by a combination of signal and category response cuts, which are signal response higher than 0.35 and  $\bar{t}\bar{t}+X$  response smaller than 0.4. The  $\bar{t}\bar{t}+X$  control region is selected by inverting either of the two selection criteria. In all cases the region definitions include all events from the respective preselection region for the final fit. The yields in the regions together with the significance of the targeted process are summarised in tables 8.1 to 8.3.

Table 8.1: Weighted pre-fit yields in the signal and control regions of the  $1\ell + 2\tau_{\text{had}}$  channel. The significance is listed for the process that the region focusses on with their uncertainty propagated from the Poisson uncertainty of the unweighted yields.

Process	Signal region	$\bar{t}\bar{t}$ control region	$V$ +jets control region
Signal	$2.41 \pm 0.27$	$0.91 \pm 0.09$	$0.159 \pm 0.026$
$\bar{t}\bar{t}$	$110 \pm 28$	$760 \pm 130$	$81 \pm 16$
$V$ +jets	$7.7 \pm 20.6$	$56.85 \pm 9.74$	$140 \pm 27$
Background	$54.46 \pm 5.56$	$84.88 \pm 11.11$	$16.89 \pm 2.73$
Significance	0.18	63.63	14.14

Once the regions are defined, the binning within the regions is chosen. The aim is to map the shapes of the distributions as accurately as possible without unnecessarily increasing the influence of the statistical uncertainties per bin. The baseline of the binning is set through a binning algorithm based on a merging threshold  $Z$  [102]:

$$Z = z_b \frac{n_b}{N_b} + z_s \frac{n_s}{N_s}. \quad (8.1)$$

The sum of the user defined parameters  $z_b$  and  $z_s$  defines the total number of bins in the region. A bin is formed when the threshold  $Z = 1$  is met. Finally, it is checked whether the shapes of the distributions are well represented by the bins without increasing the statistical uncertainties. The final binning is displayed in the distributions for the Asimov fit in figures 8.1 to 8.3. The plots show the

<sup>2</sup> In the presented analysis, the expected cross section is so low that this condition is met in all cases.

Table 8.2: Weighted pre-fit yields in the signal and control regions of the  $2\ell$  OS +  $1\tau_{\text{had}}$  channel. The significance is listed for the process that the region focusses on with their uncertainty propagated from the Poisson uncertainty of the unweighted yields.

Process	Signal region	$t\bar{t}$ control region	Z+jets control region
Signal	$2.35 \pm 0.17$	$0.61 \pm 0.14$	$0.394 \pm 0.057$
$t\bar{t}$	$1\ 010 \pm 220$	$3\ 400 \pm 800$	$680 \pm 150$
Z+jets	$99 \pm 36$	$206 \pm 90$	$4\ 576 \pm 2\ 000$
Background	$212.7 \pm 18.9$	$263.6 \pm 33.6$	$240.02 \pm 25.23$
Significance	0.06	157	150

Table 8.3: Weighted pre-fit yields in the signal and control regions of the  $2\ell$  SS +  $1\tau_{\text{had}}$  channel. The significance is listed for the process that the region focusses on with their uncertainty propagated from the Poisson uncertainty of the unweighted yields.

Process	Signal region	$t\bar{t}+X$ control region
Signal	$1.17 \pm 0.52$	$1.17 \pm 0.55$
$t\bar{t}+X$	$17.1 \pm 4.4$	$80 \pm 19$
Background	$20.8 \pm 10.0$	$31.76 \pm 13.27$
Significance	0.2	14

data as black dots. However, in an Asimov fit the data is not fitted to and consequently no change in the agreement is expected. The following discussion covers the pre-fit plots while the post-fit plots are described in the designated section.

For  $1\ell + 2\tau_{\text{had}}$  all regions contain a significant contribution of  $t\bar{t}$  events. The  $t\bar{t}$  control region displayed in figure 8.1(c) additionally shows a peak around a response of 0.1. This is expected as it reflects the shape of the NN output. Since it is a result of the softmax output it contains a different process contribution than the other bins and consequently should be represented as such in the fit. The second point of note is the disagreement between data and MC in figure 8.1(c). Studies show that the fluctuation in MC is caused by a single digit number of  $W$ +jets events that are highly weighted. As such the events cannot be simply excluded from the analysis. However, the disagreement only becomes visible after the contribution of  $t\bar{t}$  has been reduced in the region definition. In the fit a decision is made to use the regions to estimate the respective background contributions via NFs. Both contributions should be dominated by fake lepton objects. For this reason, the corresponding systematic uncertainty terms (as derived in section 6.3) is excluded in order to avoid double counting the effects.

In the  $2\ell$  OS +  $1\tau_{\text{had}}$  channel the high yields lead to stable regions and a large purity of processes for all regions. Again, a contribution of  $t\bar{t}$  cannot be avoided in all regions. In figure 8.2(e) a small flaw in modelling can be observed for the left-hand side bins.

The  $2\ell$  SS +  $1\tau_{\text{had}}$  selection has the highest expected significance of all channels. Another consequence of the selection, however, is the relatively low event count, which leads to a correspondingly large statistical uncertainty. It must be pointed out that the influences of the fake lepton estimation, described in section 6.3, are omitted in this channel. The justification for this is a technical difficulty at the time of this thesis, as reported on in section 8.3. As a consequence data and MC do not agree within their uncertainty bands. A NF is applied to the  $t\bar{t}+X$  and  $t\bar{t}$  processes combined which utilises

the control region to incorporate possible effects in the fit. In general, both regions show significant contributions from various processes which limits the power of the control region.

## 8.2 Uncertainty treatment

Uncertainties enter the fit in the form of **nuisance parameters (NPs)**. In the likelihood function the uncertainties are represented by two categories of contributions:

$$\mathcal{L}(\mu_{\text{sig}}, \theta) = \underbrace{\prod_{i=1}^{\text{stat}} e^{-\nu_i} \frac{\nu_i^{n_i}}{n_i!}}_{\text{stat term}} \underbrace{\prod_j^{\text{syst}} \mathcal{G}(\theta_j^0 - \theta_j)}_{\text{syst term}}. \quad (8.2)$$

The first contribution is the statistical Poisson uncertainty of the **MC** samples. For each bin and sample a **NP** is created using a Poisson constraint. In that way, statistical uncertainties can be treated similarly to the systematic uncertainties which are described in the following. These **NPs** are denoted as  $\gamma$  factors.

The second term denotes the Gaussian systematic uncertainties. Systematic uncertainties denote any missing knowledge in the analysis that is not of statistical nature. To understand the full impact of these systematic uncertainties on the fit all associated degrees of freedom need to be known. However, this approach is not feasible here due to the complexity of the system. Instead, a more pragmatic yet elegant treatment is chosen in which each uncertainty needs to be represented by a known model parameter and an associated uncertainty. The consequences of each uncertainty are then represented by designated samples<sup>3</sup> in which the parameter is varied by  $\pm\sigma$ . An example could be a parameter stemming from a subsidiary calibration measurement or the estimation of a background process's cross section. In the fit each **NP** is given the freedom of one standard deviation in that uncertainty.

Given this representation, two effects of the **NPs** can be examined. Firstly, the direct influence that the **NPs** have on the yields and subsequently the influence that the uncertainties have directly on the result of the **MLE** are considered.

This influence of a **NP** on the yields can be twofold. If a **NP** merely changes the overall number of events by the same rate in all bins, it is considered a normalisation component. If on the other hand the total yields stay the same while their distribution in the histogram changes the shape is affected. A **NP** can possess both a normalisation and a shape component. In an additional step, termed *pruning*, the shape, the normalisation or the whole contribution of a **NP** can be dropped if its impact is too low. This is done to not pollute the final fit with small uncertainties, likely stemming from statistical fluctuations, while making sure that possible impacts have been considered. Lastly, the impact of each **NP** on each region can be checked for its behaviour. A smoothing algorithm is applied to each region to mitigate artificial effects from statistical fluctuations.

After pruning and smoothing have been applied the influence of the resulting **NPs** on the fit can be evaluated. For each **NP** four subsidiary fits are performed in which the parameter is fixed to its pre-fit value or post-fit value and  $\pm$  the respective uncertainties. The difference in **POI**  $\mu$  is then used as a quantification of the **NP**'s impact  $\Delta\mu$  on the fit. Since the subsidiary fits are run after the full **MLE**

<sup>3</sup> Some uncertainties can be represented by a set of event weights instead of a fully designated sample but the consequences are the same.

for each NP  $\theta$  the maximum likelihood estimator  $\hat{\theta}$  is known. Using this, the pre-fit impact can be calculated using the initial NP  $\theta$  with  $\Delta\theta = 1$  and the post-fit impact using  $\hat{\theta}$  and  $\Delta\hat{\theta} \leq 1$ .

In addition to the impact, the information that is gained about the NPs is studied. A *constraint* is put on a NP if its uncertainties can be reduced with respect to the original  $\Delta\theta$  during the fit. A *pull* is recorded if the fit changes the nominal value of the NP. In both cases this means that the fit can extract information about the NP from the preselection region. In Asimov fits this is usually unwanted behaviour since the nominal values with its provided uncertainties should be sufficient to describe the uncertainties in the region. The exception are significant contributions from processes that contribute through misidentified objects. Such contributions can be poorly studied in the preselection region which implies a constraint in the fit.

### 8.3 Fit preparation

This section describes a few important details that are slightly more practical than those described previously yet relevant for completeness. In the MLE the NFs obtain uncertainties from the fit. This means that any normalisation components of uncertainties would be double counting. Therefore, the necessary samples are re-scaled to remove the normalisation differences while keeping possible shape information.

Section 6.3 describes the estimation of fake rates and their associated uncertainties in each channel. The results enter the fit as a NP. Since dominant backgrounds like  $t\bar{t}$  and  $Z$ +jets are fake lepton dominated, the lepton fake rate correction needs to be dropped additionally if the respective cross sections are left free floating in the fit. While for  $1\ell + 2\tau_{\text{had}}$  and  $2\ell \text{ OS} + 1\tau_{\text{had}}$  the remaining terms of the fake lepton uncertainties are taken into account, the uncertainties (but not the weights) are dropped entirely in the  $2\ell \text{ SS} + 1\tau_{\text{had}}$  channel. The reason for this are, at the time of this thesis, poorly understood effects in the impact of the NPs on the fit. Instead, the effects are represented by a combined NF for  $t\bar{t}$  and  $t\bar{t} + X$  processes based on the control region.

Lastly, in any fit any bin and region that is expected to have a signal contribution higher than 0.5% is excluded from the fit to the real data. This is called blinding the signal and its purpose is to avoid a possible bias from looking at the data before investigating the fit sufficiently. In addition to this, before using data in the fit a fit of the control regions alone can be used to gauge its stability.

### 8.4 Asimov fit

A first fit is performed using the Asimov sample, in which the contributions for all processes are set to their SM predictions. While this fit will not alter any rates, it can estimate the impact of the systematic uncertainties. Additionally, the sensitivity of the analysis is gauged in this initial fit.<sup>4</sup> The obvious first result of an Asimov fit is then the NF of 1 and its expected uncertainty. This uncertainty is expected to lie in a range that agrees with the physical expectation. Both a very high and very low uncertainty are a cause for investigation. The second plot to observe is the ranking of the impact parameters before and after the fit. The leading NPs should not exceed the expected impact. Additionally, the impact is expected to be symmetrical for most parameters in the Asimov fit because the uncertainties are usually

<sup>4</sup> Extracting sensitivity from an Asimov sample means calculating the expected significance for the case in which the data looks exactly like the simulation.

symmetric. Deviations might result from technical reasons in the MLE algorithm. Any worrisome behaviour in ranking and normalisation can often be tracked to constraints of NPs. In the Asimov fit no NP can be pulled. Every occurring constraint should be either understood or eliminated. To do this it is possible to list impact and constraints by process and by NP. For any NP additionally the pre- and post-fit impact is controlled. To accomplish this, the deviation in distributions in each region is observed for each NP after pruning and smoothing. A complete description of the constraints and correlations in the fit are provided via appendix K. Most noteworthy are the effects on the theoretical systematic uncertainties of the  $t\bar{t}$  process in all channels.

Figure 8.1 shows the pre- and post-Asimov-fit distributions for the  $1\ell + 2\tau_{\text{had}}$  channel. The presented data points correspond to the actual and not the Asimov data in all unblinded regions but were not used in the fit. The initial noteworthy observation is the decrease in uncertainties, particularly in the signal region, following the Asimov fit. The resulting uncertainties on the NFs are:

$$\mu_{\text{signal, dihad}} = (1 \begin{smallmatrix} +4 \\ -4 \end{smallmatrix} \text{ (stat.)} \begin{smallmatrix} +3.5 \\ -3.1 \end{smallmatrix} \text{ (sys.)}) = (1 \begin{smallmatrix} +6 \\ -5 \end{smallmatrix}) \quad (8.3)$$

$$\mu_{t\bar{t}, \text{ dihad}} = (1.00 \begin{smallmatrix} +0.04 \\ -0.04 \end{smallmatrix} \text{ (stat.)} \begin{smallmatrix} +0.06 \\ -0.06 \end{smallmatrix} \text{ (sys.)}) = (1.00 \begin{smallmatrix} +0.07 \\ -0.07 \end{smallmatrix}) \quad (8.4)$$

$$\mu_{V+\text{jets, dihad}} = (1.00 \begin{smallmatrix} +0.11 \\ -0.11 \end{smallmatrix} \text{ (stat.)} \begin{smallmatrix} +0.25 \\ -0.25 \end{smallmatrix} \text{ (sys.)}) = (1.00 \begin{smallmatrix} +0.27 \\ -0.27 \end{smallmatrix}). \quad (8.5)$$

The contributions of statistical and systematic uncertainty are of roughly the same size. The ranking of the impact is displayed in figure 8.4. The highest ranking systematic is the NLO generator variation for the  $t\bar{t}$  sample. A high impact and a possible constraint for this NP has been previously observed for this channel [88]. A possible reason is relatively poor modelling of  $t\bar{t}$  systematic samples in the signal which can stem from a large contribution of fake objects. Studies with alternative samples have been conducted but the samples contained insufficient events in the selected region. Given the large fraction of  $t\bar{t}$  events in all regions, the resulting impact can be large. In addition, both NFs for the control region have a significant impact. Furthermore, besides minor jet related constraints, a large group of constraints related to the  $t\bar{t}$  process remains for the Asimov fit.

Figure 8.2 shows the pre- and post-Asimov-fit distributions for the  $2\ell \text{ OS} + 1\tau_{\text{had}}$  channel. The presented data points correspond to the actual and not the Asimov data in all unblinded regions. After the Asimov fit a decrease in uncertainties is also visible for this channel. The resulting uncertainties on the NFs are:

$$\mu_{\text{signal, OS}} = (1 \begin{smallmatrix} +12 \\ -11 \end{smallmatrix} \text{ (stat.)} \begin{smallmatrix} +11 \\ -10 \end{smallmatrix} \text{ (sys.)}) = (1 \begin{smallmatrix} +16 \\ -15 \end{smallmatrix}) \quad (8.6)$$

$$\mu_{t\bar{t}, \text{ OS}} = (1.000 \begin{smallmatrix} +0.020 \\ -0.020 \end{smallmatrix} \text{ (stat.)} \begin{smallmatrix} +0.030 \\ -0.030 \end{smallmatrix} \text{ (sys.)}) = (1.00 \begin{smallmatrix} +0.04 \\ -0.04 \end{smallmatrix}) \quad (8.7)$$

$$\mu_{Z+\text{jets, OS}} = (1.00 \begin{smallmatrix} +0.02 \\ -0.02 \end{smallmatrix} \text{ (stat.)} \begin{smallmatrix} +0.10 \\ -0.10 \end{smallmatrix} \text{ (sys.)}) = (1.00 \begin{smallmatrix} +0.10 \\ -0.10 \end{smallmatrix}). \quad (8.8)$$

As before the contributions of statistical and systematic uncertainty are of roughly the same size. The ranking of the impact is displayed in figure 8.5. In this channel, the  $t\bar{t}$  related NPs have by far the highest impact. This is not entirely unexpected since  $t\bar{t}$  is the dominant process in its own control region and in the signal region. Additionally the statistical uncertainties of the signal region play an important role in the fit. This is not entirely unexpected because the yields decrease significantly for the higher bins of the signal region.

Figure 8.3 shows the pre- and post-Asimov-fit distributions for the  $2\ell \text{ SS} + 1\tau_{\text{had}}$  channel. The presented data points correspond to the actual and not the Asimov data in all unblinded regions. There is an absence of any reduction in uncertainties evident in the fit. The reason for this is, that due to the



complete exclusion of  $\tau_{\text{had}}$  fake uncertainties, the pre-fit uncertainties are presumably underestimated. Studies have shown that for an inclusion of these uncertainties a reduction of uncertainty bands in the fit is observed. As a consequence, the **NF** reflecting this uncertainty,  $\mu(\bar{t}\bar{t} + \bar{t}\bar{t} + X)$  has the leading impact on the fit. The other impacts are comparatively low with important contributions from the signal region statistical uncertainties and the  $\bar{t}\bar{t}$  related systematic uncertainties. The resulting uncertainties on the **NFs** are:

$$\mu_{\text{signal, SS}} = (1 \begin{smallmatrix} +6 \\ -5 \end{smallmatrix} \text{ (stat.)} \begin{smallmatrix} +3 \\ -5 \end{smallmatrix} \text{ (sys.)}) = (1 \begin{smallmatrix} +7 \\ -5 \end{smallmatrix}) \quad (8.9)$$

$$\mu_{\bar{t}\bar{t} + \bar{t}\bar{t} + X, \text{ SS}} = (1.00 \begin{smallmatrix} +0.14 \\ -0.14 \end{smallmatrix} \text{ (stat.)} \begin{smallmatrix} +0.11 \\ -0.11 \end{smallmatrix} \text{ (sys.)}) = (1.00 \begin{smallmatrix} +0.18 \\ -0.18 \end{smallmatrix}). \quad (8.10)$$

As predicted, the  $2\ell$  SS +  $1\tau_{\text{had}}$  channel despite its low yields shows a good overall uncertainty and the lowest impact of systematic uncertainties.

Overall, the Asimov fits are well understood in all cases. For future studies and improved estimation of the  $\tau_{\text{had}}$  fakes especially in the  $2\ell$  SS +  $1\tau_{\text{had}}$  channel and a better estimation of the  $\bar{t}\bar{t}$  events in all regions could yield better results for the fit.

## 8.5 Data fit

The final step is the **MLE** using data. This is done in two steps: a background-only fit and a full data fit. In the background-only fit, only the data in control regions is used while the signal region is treated as its Asimov sample. Using the background hypothesis in this fit gives an estimate of the behaviour of the region for the case that no signal can be detected. This gives strong implications on both the behaviour of the dominant backgrounds and the stability of the fit provided the available data. The final results are controlled for unexpected correlations and especially pulls that exceed the expected range of one standard deviation.

The background-only fits are very similar to the Asimov fits in all three channels. No unexpected changes in correlation occur and no **NP** is pulled out of the allowed boundaries. The largest pulls are visible for the previously constrained parameters, which is expected. For this reason, the full documentation of the background-only fits is omitted in this thesis. After performing these checks the signal region is unblinded and the complete available data is used in the fit. For this final fit the same steps of investigation as for the Asimov fit are followed and the presented results correspond to the best possible fit at the time of this thesis. The constraints and correlation matrices for all three channels in the data fit are summarised in appendix L. No **NP** is pulled out its expectation range. The largest pulls are visible for the  $\bar{t}\bar{t}$  theory systematic uncertainties which were previously constrained in the Asimov fits. However, especially for the  $2\ell$  OS +  $1\tau_{\text{had}}$  channel additional pulls appear for jet-related **NPs**.

Figure 8.7 shows the pre- and post-fit distributions for the  $1\ell + 2\tau_{\text{had}}$  channel. From the figures it becomes evident that the fit finds good information to achieve agreement in the  $V + \text{jets}$  control region. In the signal and  $\bar{t}\bar{t}$  control region, the **MC** rates are only adjusted slightly by the fit. The resulting **NFs** are:

$$\mu_{\text{signal, dihad}} = (3.4 \begin{smallmatrix} +4.0 \\ -4.0 \end{smallmatrix} \text{ (stat.)} \begin{smallmatrix} +4.0 \\ -3.2 \end{smallmatrix} \text{ (sys.)}) = (3 \begin{smallmatrix} +6 \\ -5 \end{smallmatrix}) \quad (8.11)$$

$$\mu_{\bar{t}\bar{t}, \text{ dihad}} = (0.94 \begin{smallmatrix} +0.04 \\ -0.04 \end{smallmatrix} \text{ (stat.)} \begin{smallmatrix} +0.04 \\ -0.04 \end{smallmatrix} \text{ (sys.)}) = (0.94 \begin{smallmatrix} +0.06 \\ -0.06 \end{smallmatrix}) \quad (8.12)$$

$$\mu_{V + \text{jets, dihad}} = (0.42 \begin{smallmatrix} +0.08 \\ -0.08 \end{smallmatrix} \text{ (stat.)} \begin{smallmatrix} +0.14 \\ -0.14 \end{smallmatrix} \text{ (sys.)}) = (0.42 \begin{smallmatrix} +0.16 \\ -0.16 \end{smallmatrix}). \quad (8.13)$$

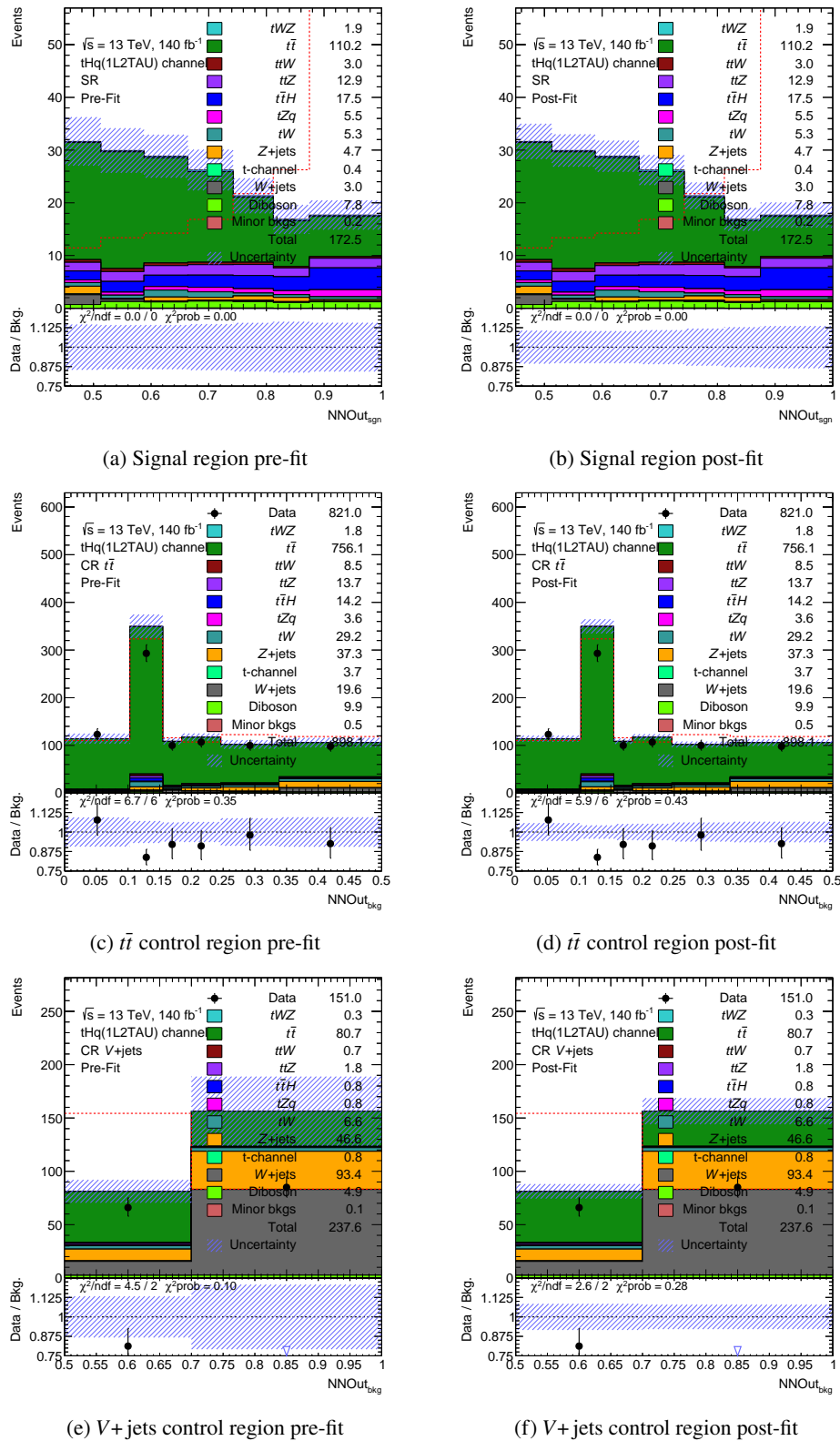


Figure 8.1: Displayed are the pre- and post-fit distributions of both the signal (a), (b) region and the  $t\bar{t}$  (c), (d) and V+jets (e), (f) control regions for the Asimov fit in the  $1\ell + 2\tau_{\text{had}}$  channel.

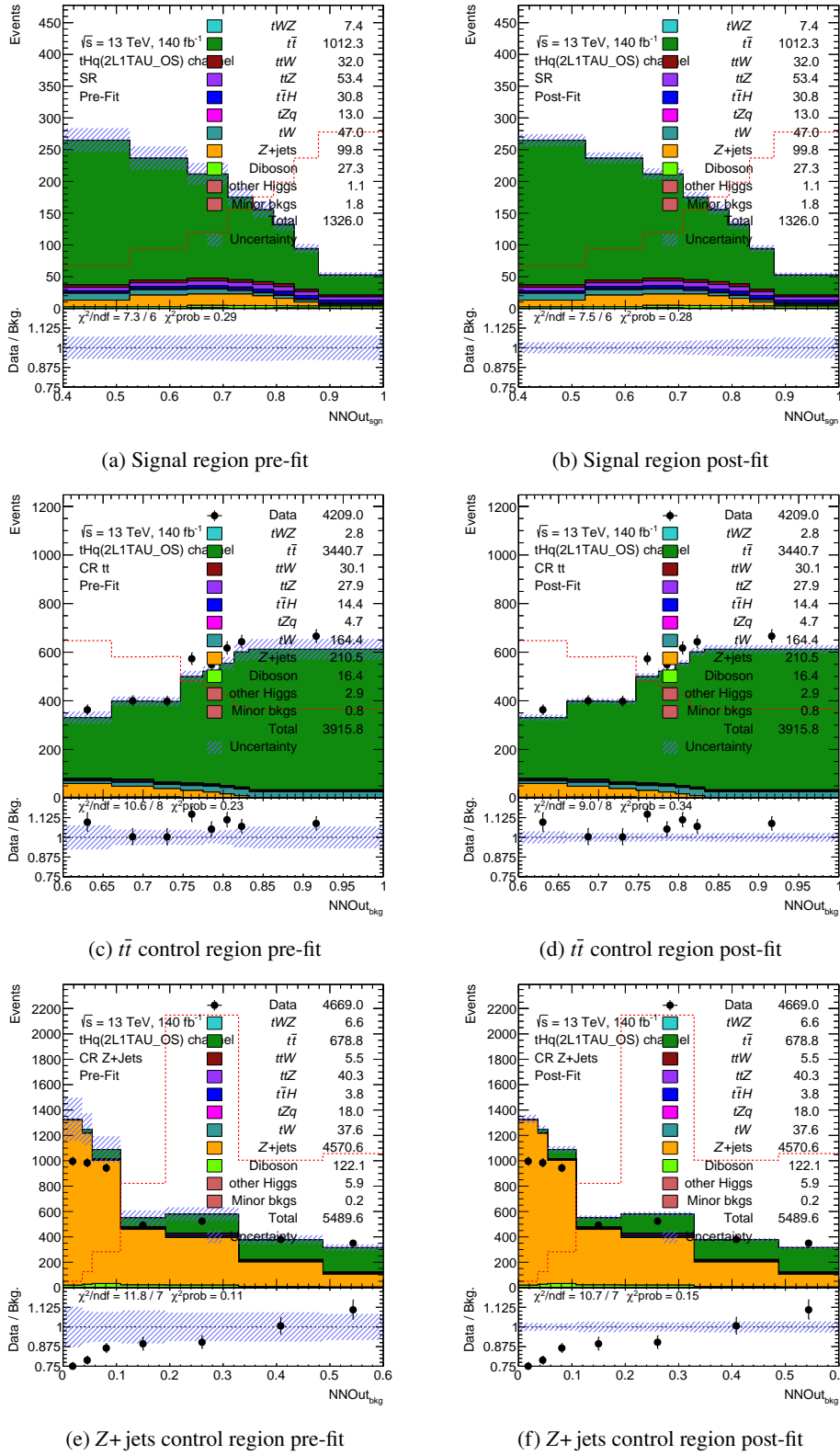


Figure 8.2: Displayed are the pre- and post-fit distributions of both the signal (a), (b) region and the  $t\bar{t}$  (c), (d) and Z+jets (e), (f) control regions for the Asimov fit in the  $2\ell$  OS +  $1\tau_{\text{had}}$  channel.

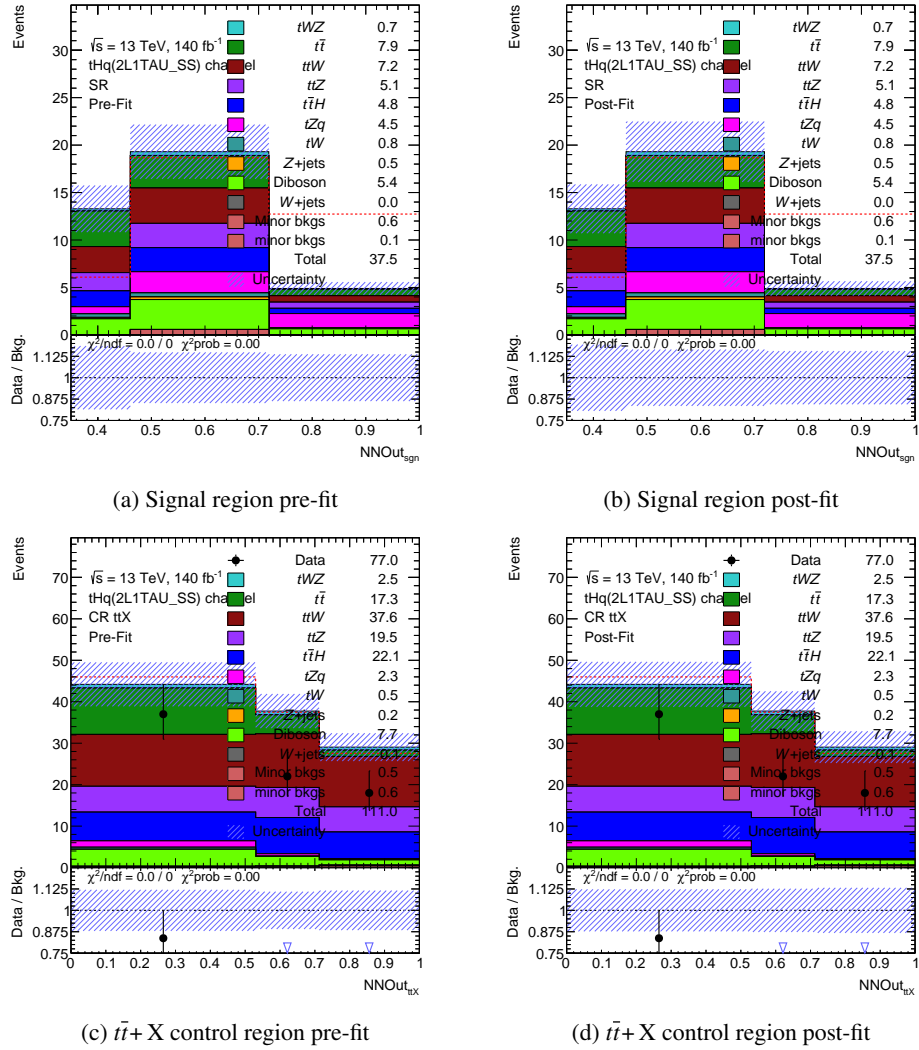


Figure 8.3: Displayed are the pre- and post-fit distributions of both the signal (a), (b) region and the  $t\bar{t} + X$  (c), (d) control region for the Asimov fit in the  $2\ell \text{ SS} + 1\tau_{\text{had}}$  channel.

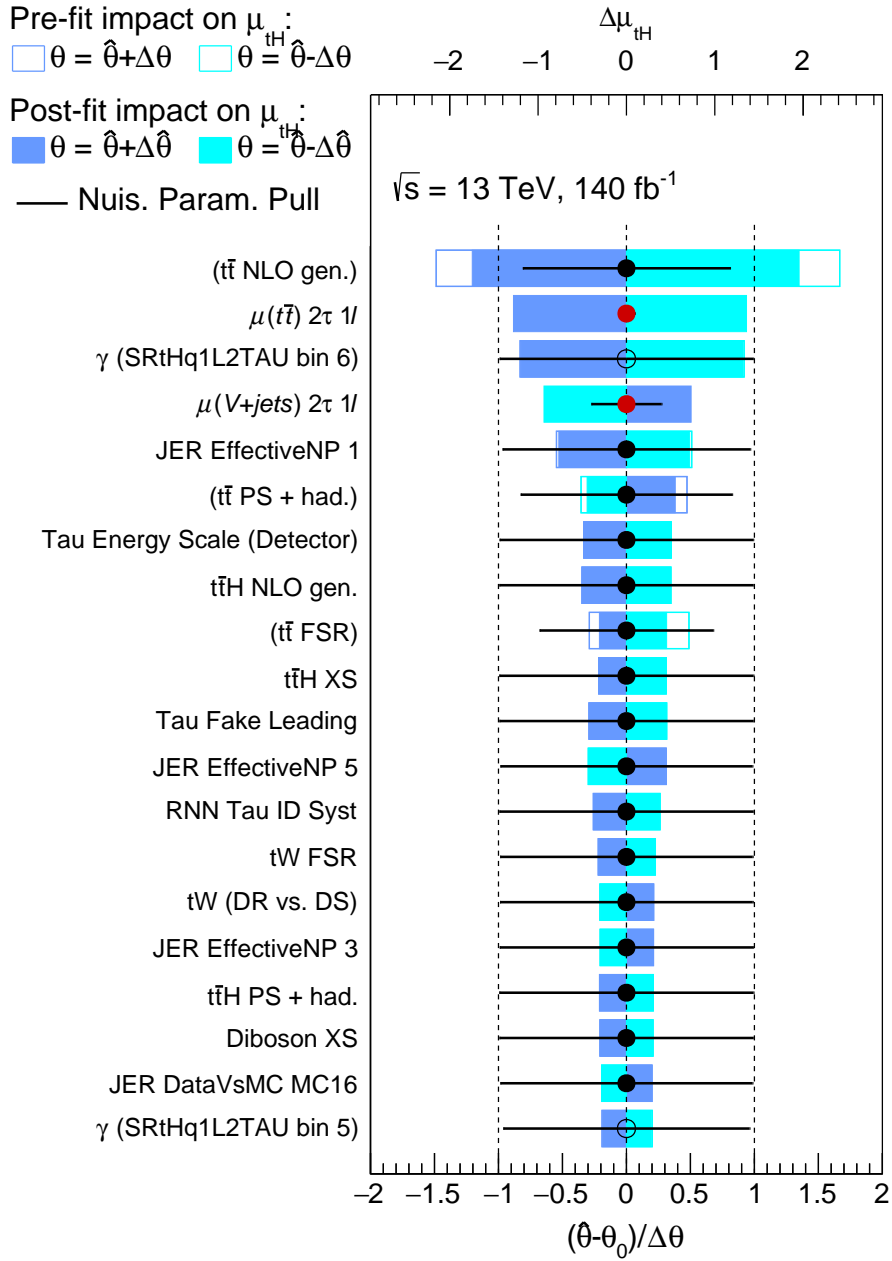


Figure 8.4: Asimov ranking plot of the  $1\ell + 2\tau_{\text{had}}$  channel. The upper axis shows the impact on the POI which is displayed by the empty and filled rectangles for pre- and post-fit impact respectively. The bottom axis denotes the pull of the parameter in units of pre-fit uncertainty.

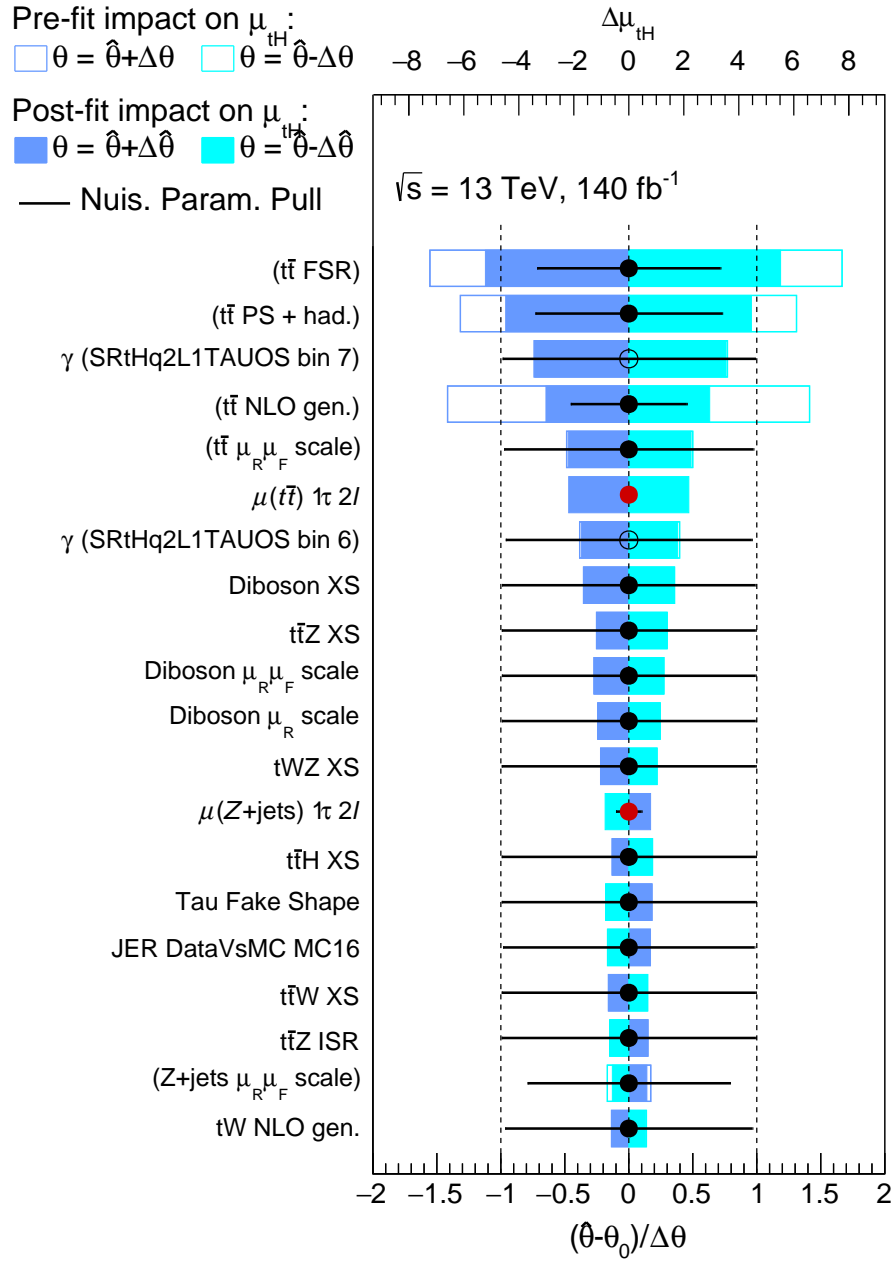


Figure 8.5: Asimov ranking plot of the  $2\ell$  OS +  $1\tau_{\text{had}}$  channel. The upper axis shows the impact on the POI which is displayed by the empty and filled rectangles for pre- and post-fit impact respectively. The bottom axis denotes the pull of the parameter in units of pre-fit uncertainty.

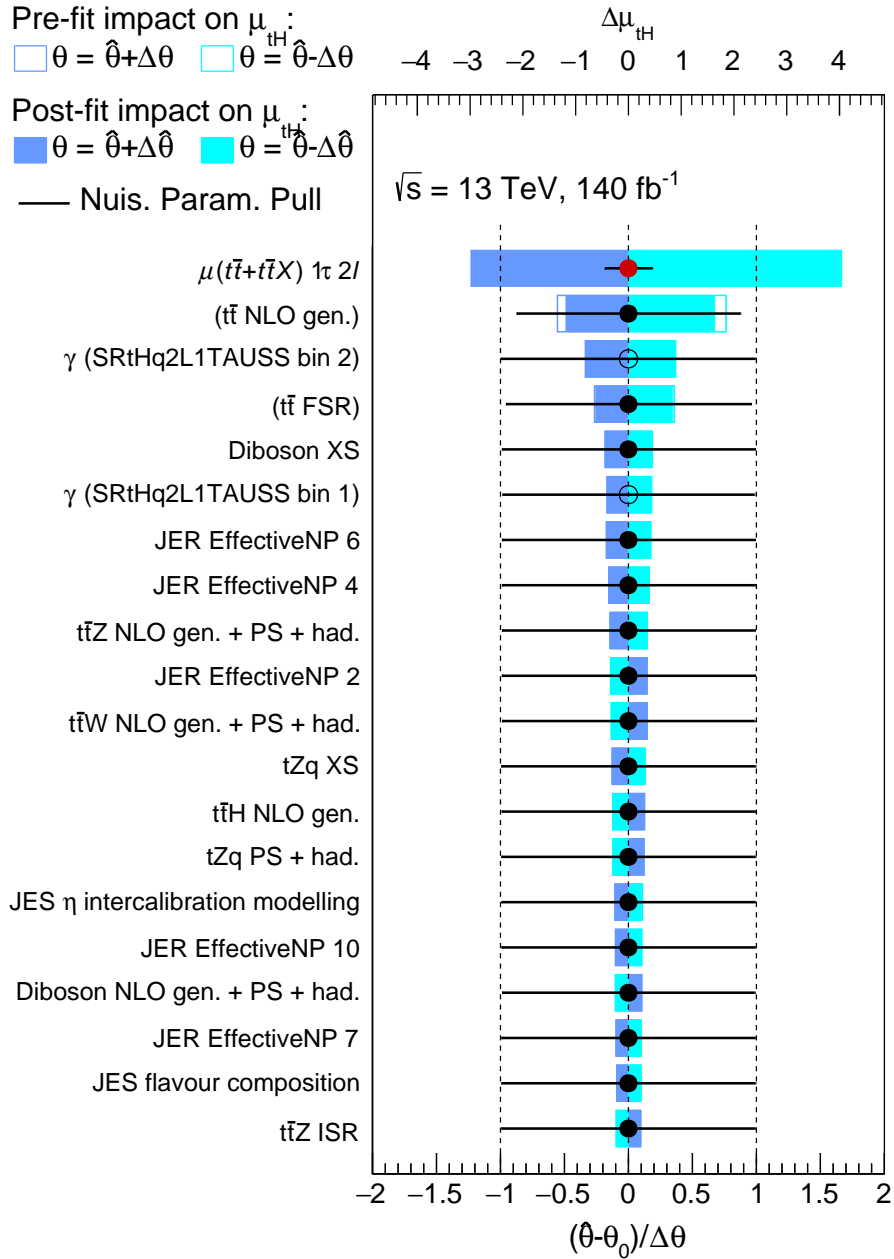


Figure 8.6: Asimov ranking plot of the  $2\ell$  SS +  $1\tau_{\text{had}}$  channel. The upper axis shows the impact on the POI which is displayed by the empty and filled rectangles for pre- and post-fit impact respectively. The bottom axis denotes the pull of the parameter in units of pre-fit uncertainty.

While slightly statistically dominated, the contributions are roughly of the same size. The slight asymmetry of the uncertainties stems mostly from the statistical uncertainties. Figure 8.10 shows the ranking for the data fit. In addition to the  $t\bar{t}$  theory systematics and the statistical limitations in the signal region, the background NFs appear as leading impacts. Especially the  $V$ + jets NF is pulled significantly. Another impact associated with a strong pull is the  $t\bar{t}$  FSR NP.

The pre- and post-fit distributions of the  $2\ell$  OS +  $1\tau_{\text{had}}$  channel are displayed in figure 8.8. This channel had the best pre-fit agreement in all regions which the fit can consistently exploit. However, a large negative value for the POI is extracted:

$$\mu_{\text{signal, OS}} = (-27^{+11}_{-10} \text{ (stat.) } ^{+9}_{-9} \text{ (sys.)}) = (-27^{+14}_{-14}) \quad (8.14)$$

$$\mu_{t\bar{t}, \text{OS}} = (1.100^{+0.020}_{-0.020} \text{ (stat.) } ^{+0.030}_{-0.030} \text{ (sys.)}) = (1.10^{+0.04}_{-0.04}) \quad (8.15)$$

$$\mu_{Z+\text{jets, OS}} = (0.78^{+0.01}_{-0.01} \text{ (stat.) } ^{+0.08}_{-0.08} \text{ (sys.)}) = (0.78^{+0.08}_{-0.08}) . \quad (8.16)$$

The highest ranking impacts are again  $t\bar{t}$  theory NPs followed by the  $t\bar{t}$  NFs. Additionally, the statistical uncertainty in one signal bin is ranked highly.

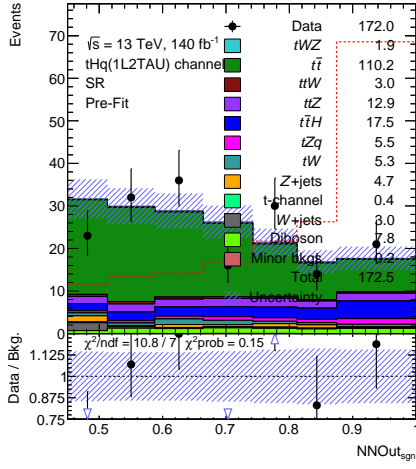
In the  $2\ell$  SS +  $1\tau_{\text{had}}$  fit, as displayed in figure 8.9, the fit achieves an overall improvement of the agreement in both the signal and the control region. The resulting NFs are:

$$\mu_{\text{signal, SS}} = (-1.8^{+5.0}_{-3.5} \text{ (stat.) } ^{+1.7}_{-1.6} \text{ (sys.)}) = (-2^{+5}_{-4}) \quad (8.17)$$

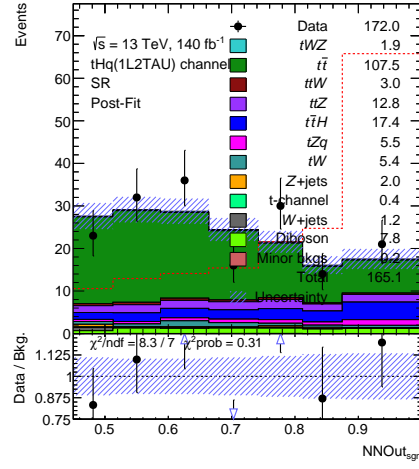
$$\mu_{t\bar{t} + t\bar{t}+X, \text{SS}} = (0.65^{+0.11}_{-0.11} \text{ (stat.) } ^{+0.07}_{-0.07} \text{ (sys.)}) = (0.65^{+0.13}_{-0.13}) . \quad (8.18)$$

The  $2\ell$  SS +  $1\tau_{\text{had}}$  channel has the smallest impact from systematic uncertainties and is, as expected, dominated by the statistical uncertainty. The ranking is shown in figure 8.12. The highest impact by far is the combined NF of the  $t\bar{t}$  and  $t\bar{t}+X$  processes. The impact of this NF is assumed to be an estimator for the uncertainty stemming from the fake lepton contributions. It can be expected that a corresponding set of designated NPs would also impact the result significantly.

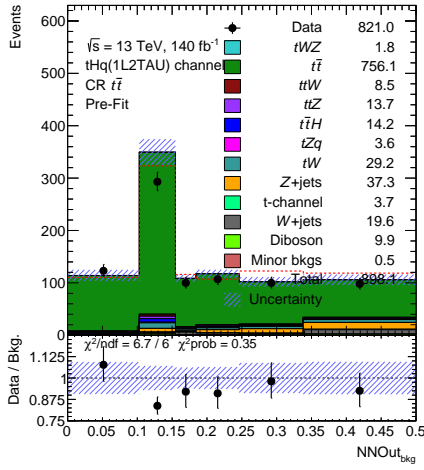
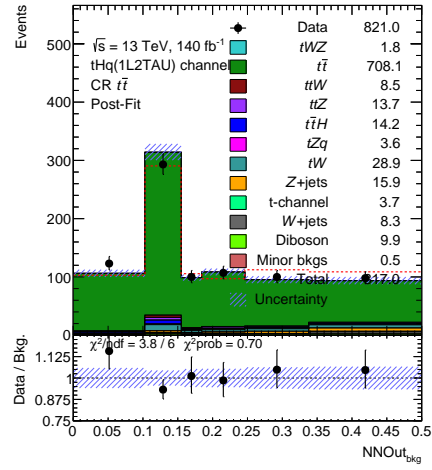
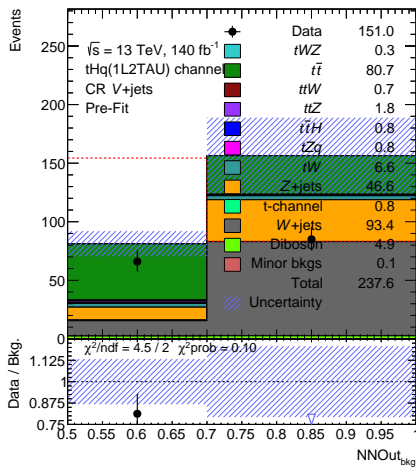




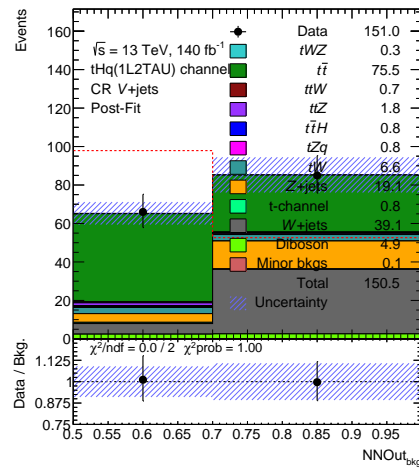
(a) Signal region pre-fit



(b) Signal region post-fit

(c)  $t\bar{t}$  control region pre-fit(d)  $t\bar{t}$  control region post-fit

(e) V+ jets control region pre-fit



(f) V+ jets control region post-fit

Figure 8.7: Displayed are the pre- and post-fit distributions of both the signal (a), (b) region and the  $t\bar{t}$  (c), (d) and V+ jets (e), (f) control regions for the fit to data in the  $1\ell + 2\tau_{\text{had}}$  channel.

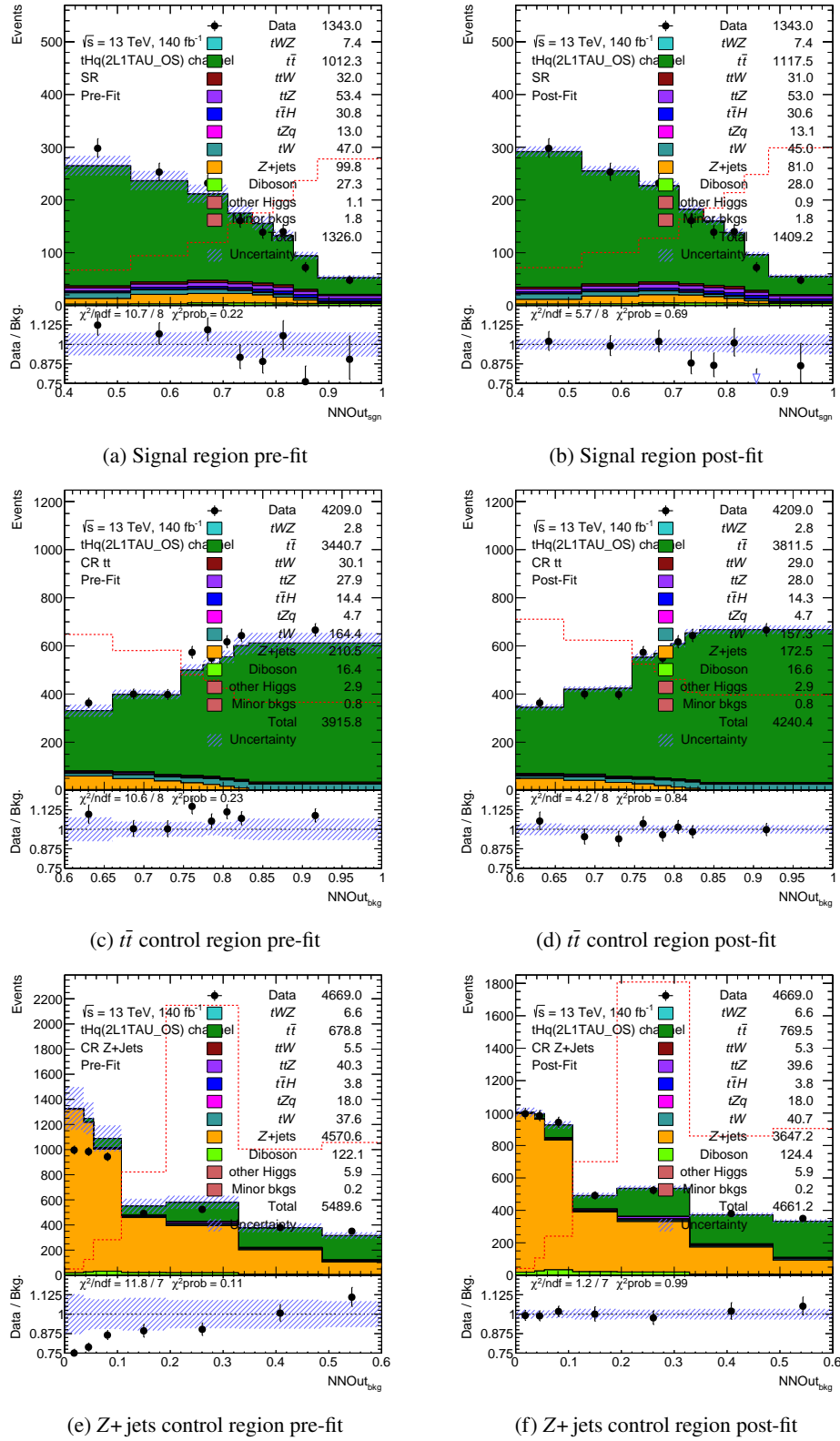


Figure 8.8: Displayed are the pre- and post-fit distributions of both the signal (a), (b) region and the  $t\bar{t}$  (c), (d) and Z+jets (e), (f) control regions for the fit to data in the  $2\ell$  OS +  $1\tau_{\text{had}}$  channel.

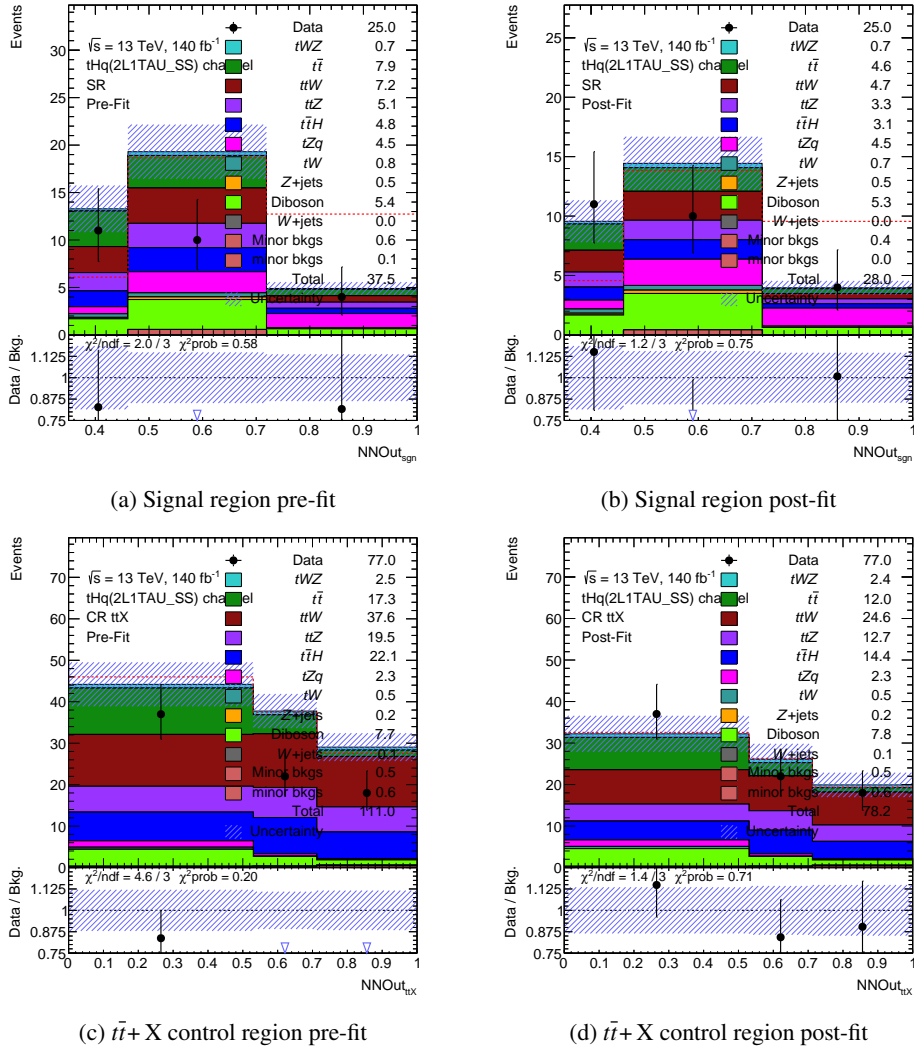


Figure 8.9: Displayed are the pre- and post-fit distributions of both the signal (a), (b) region and the  $t\bar{t}+X$  (c), (d) control region for the fit to data in the  $2\ell \text{ SS} + 1\tau_{\text{had}}$  channel.

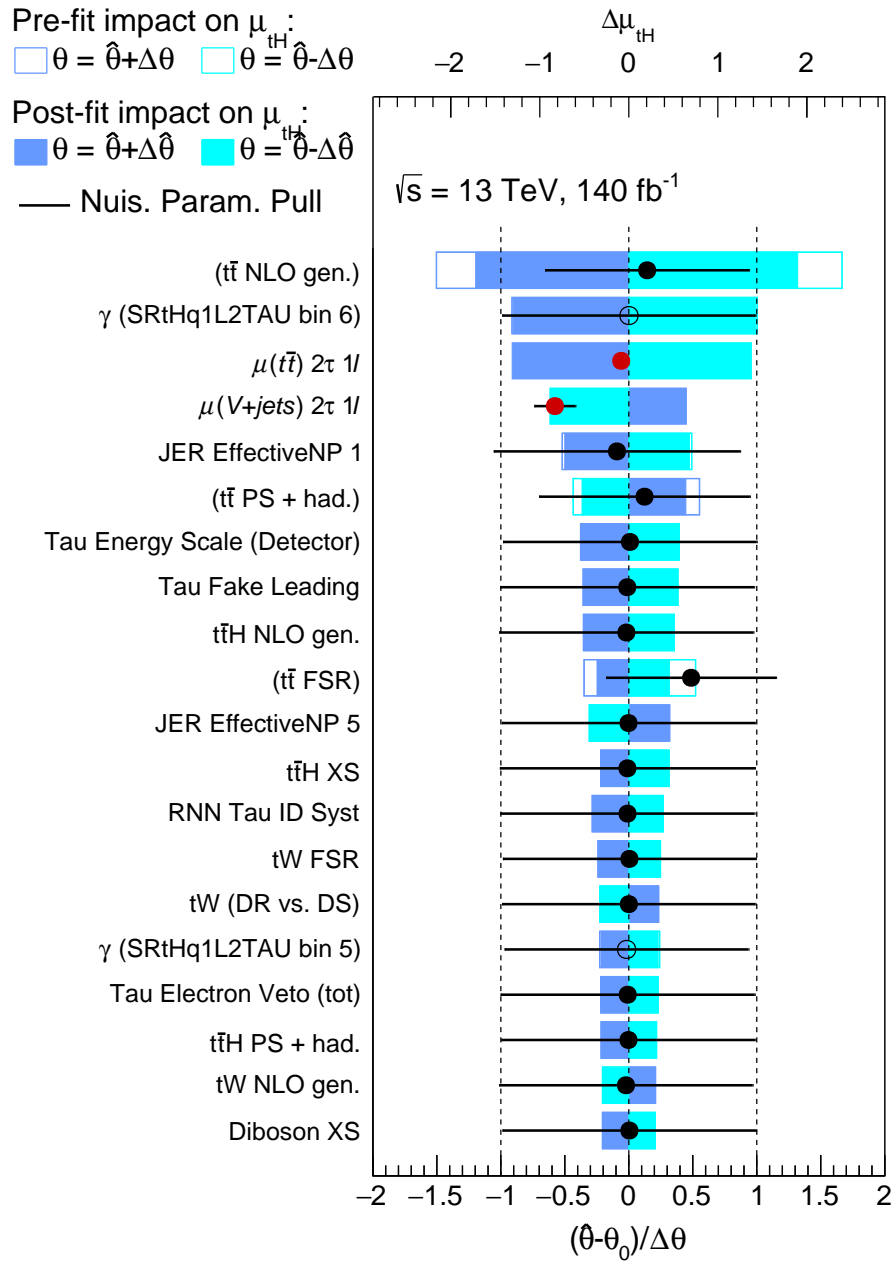


Figure 8.10: Ranking of the NPs' impact on the NF of the fit to data in the  $1\ell + 2\tau_{\text{had}}$  channel. The upper axis shows the impact on the POI which is displayed by the empty and filled rectangles for pre- and post-fit impact respectively. The bottom axis denotes the pull of the parameter in units of pre-fit uncertainty.

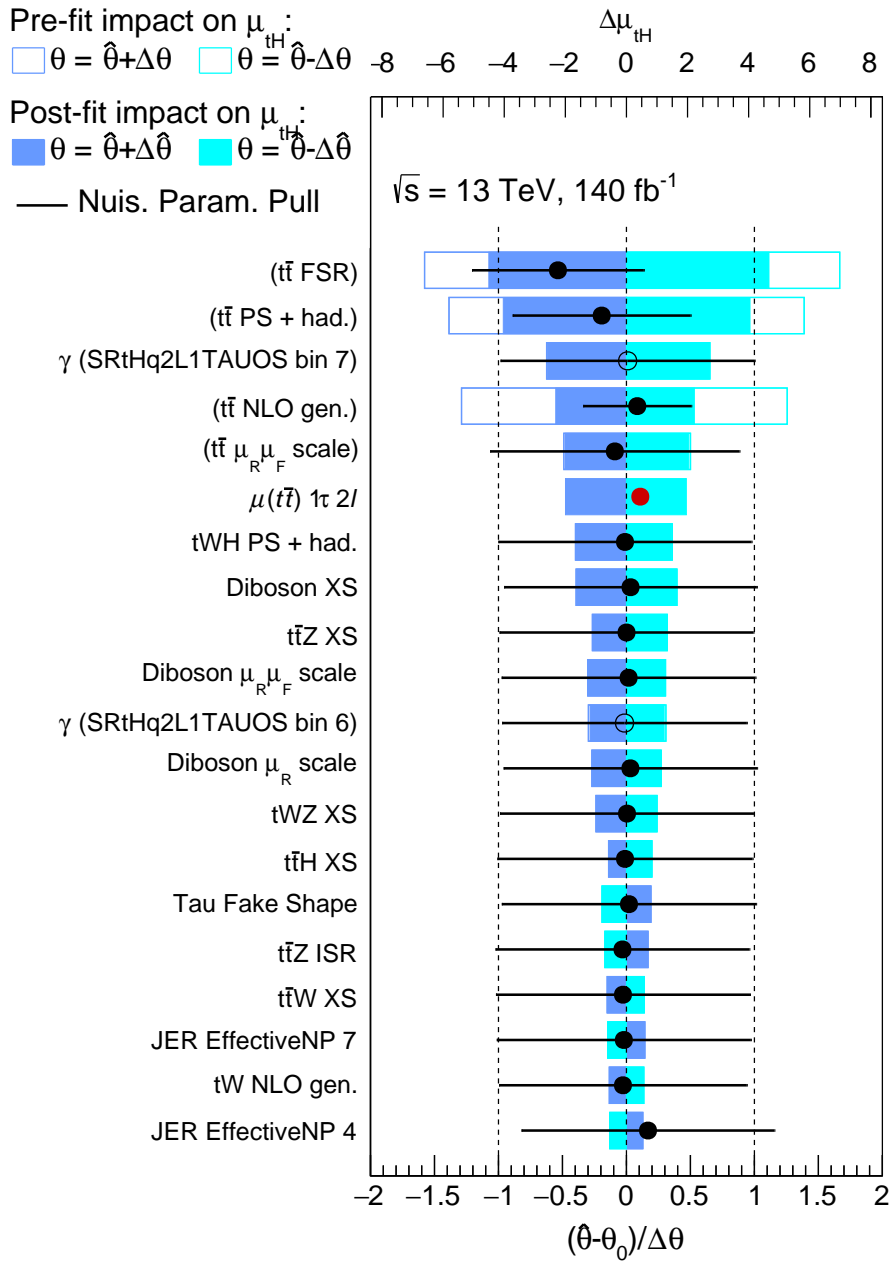


Figure 8.11: Ranking of the NPs' impact on the NF of the fit to data in the  $2\ell$  OS +  $1\tau_{\text{had}}$  channel. The upper axis shows the impact on the POI which is displayed by the empty and filled rectangles for pre- and post-fit impact respectively. The bottom axis denotes the pull of the parameter in units of pre-fit uncertainty.

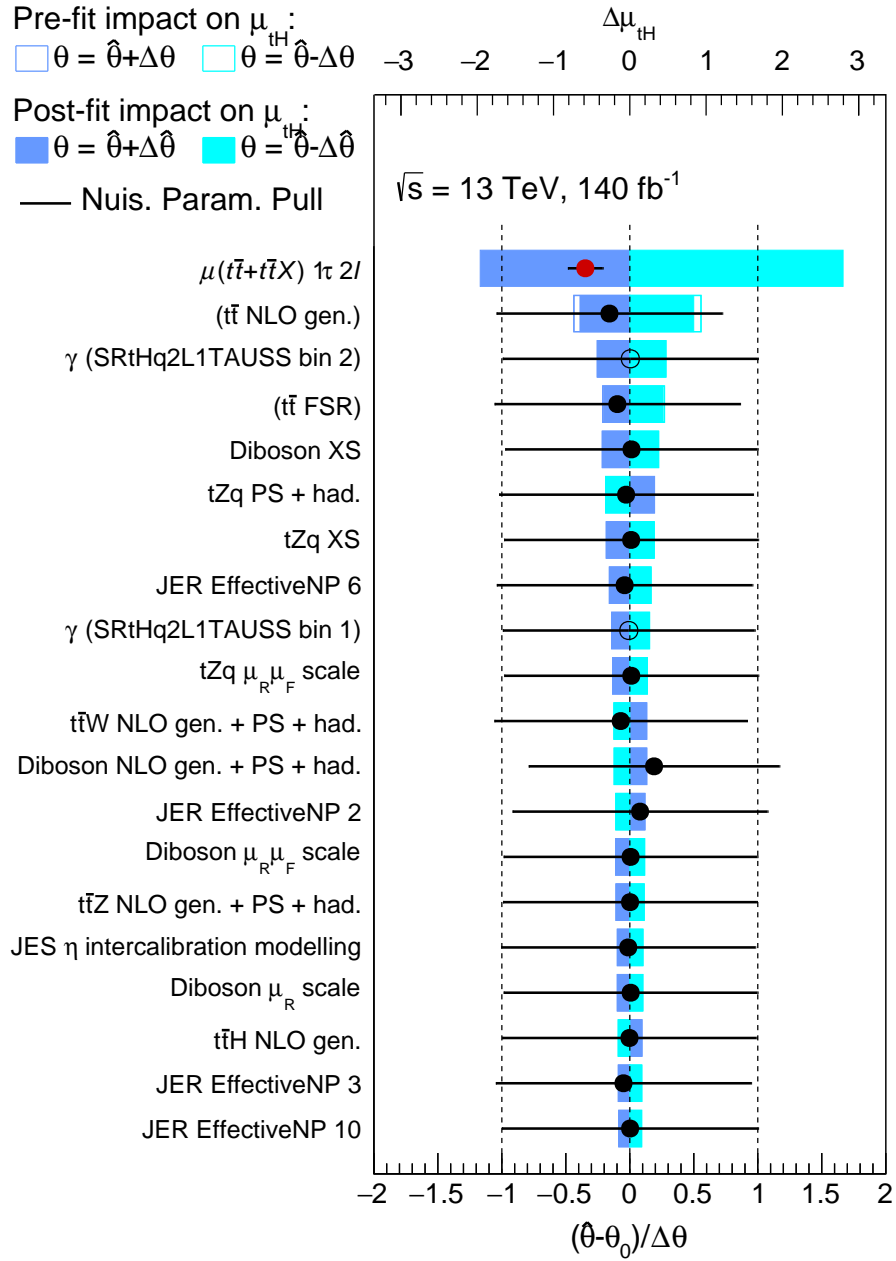


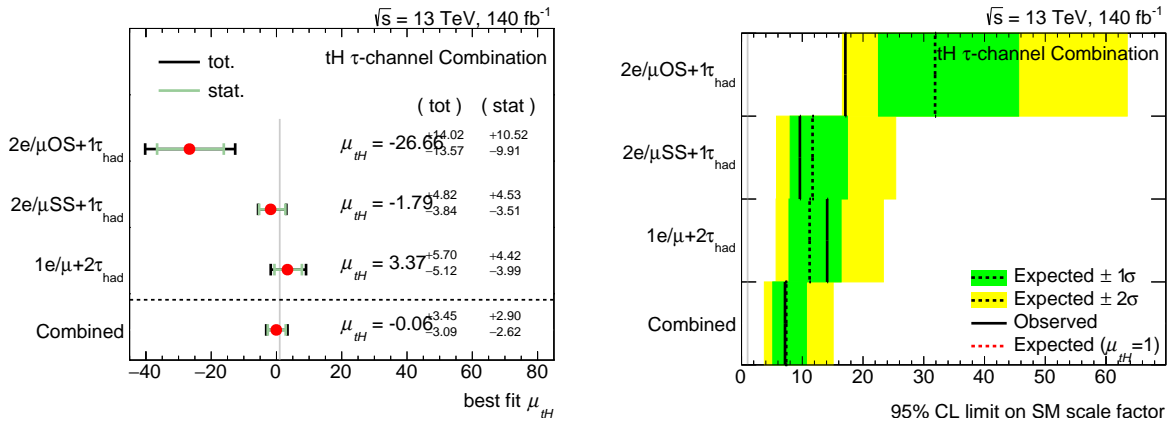
Figure 8.12: Ranking of the NPs' impact on the NF of the fit to data in the  $2\ell \text{ SS} + 1\tau_{\text{had}}$  channel. The upper axis shows the impact on the POI which is displayed by the empty and filled rectangles for pre- and post-fit impact respectively. The bottom axis denotes the pull of the parameter in units of pre-fit uncertainty.

## 8.6 Fit combination and discussion

In a final step the results of all channels are combined. This is possible because the chosen regions are orthogonal. All NFs are summarised in figure 8.13(a). The  $2\ell$  SS +  $1\tau_{\text{had}}$  channel has the lowest overall uncertainties. Both  $2\ell$  SS +  $1\tau_{\text{had}}$  and  $1\ell + 2\tau_{\text{had}}$  are in good agreement with the SM.  $2\ell$  OS +  $1\tau_{\text{had}}$  is only within two standard deviations from the SM which can happen for any channel with low significance. However, the other two channels as are in good agreement and the combination is consistent with the SM prediction given its NF of:

$$\mu_{\text{combined}} = (-0.1^{+2.9}_{-2.6} \text{ (stat.) } +1.9^{+1.9}_{-1.6} \text{ (sys.)}) = (-0.1^{+3.5}_{-3.1}). \quad (8.19)$$

In addition, the test statistic can be used to extract a statement about the cross section estimation. The significance in the combination is insufficient for a discovery and instead upper limits on the cross section can be determined. The estimation of a limit is the calculation of a value that is greater than the true value of the observable with 95 % confidence.<sup>5</sup> To obtain the interval, the maximum likelihood is compared to the likelihood for different fixed values of signal strength. In this scan the hypothesis for a signal strength is found for which the obtained results are outside of the desired confidence interval. The expected limit is extracted from MC under the assumption that the data perfectly match the simulation. Figure 8.13(b) summarises the limits for all channels and their combination excluding a cross section of about seven times the SM prediction.



(a) Displayed are all NFs and their combination. The vertical, grey line represents the SM expectation value.

(b) Expected and observed limits on the combined NF for a 95 % confidence level (CL).

Figure 8.13: Shown are the results of the fit combination. (a) shows the NFs and their combination while (b) shows the expected and observed limits on the SM scale for each channel and the combination.

<sup>5</sup> The value chosen for the confidence is a convention. In general any chosen confidence is arbitrary. In a hypothesis test a statement can be inferred but never proven.





---

## Conclusion

---

With the discovery of the Higgs boson completing the Standard Model, precision measurements have become the core focus of high energy collider physics. The Standard Model of particle physics contains numerous open questions and incorporates various free parameters. Each theory that goes beyond our current understanding makes its own predictions about these parameters and a robust theory with promising predictions is an essential drive towards new experiments. Therefore, a precise measurement of the parameters may pave the way towards the next groundbreaking discovery in the field. A majority of these parameters, including some of the most intriguing ones, are associated to the Higgs sector of the model.

One parameter of particular interest is the coupling between the Higgs boson and the heaviest of the quarks, the top quark. Both, sign and magnitude of the coupling are compelling observables and the associated production of a top quark and a Higgs boson,  $tH$ , offers a unique opportunity to measure both, even in the presence of new physics. This gives the process a special role compared to analyses like the decay of a Higgs boson to two photons which only allows the same conclusions under exclusion of new physics. This thesis in particular covers the channel in which the Higgs boson decays into a pair of  $\tau$  leptons of which one or both decay hadronically using the Run 2 data from the ATLAS detector at the LHC. The big downside of  $tH$  is its small cross section combined with a huge number of similar background processes. This is a problem that an increasingly large number of analyses have in common while the luminosity and energy of the LHC are increased and rarer processes are targeted.

A compelling tool to mitigate this problem is the usage of machine learning methods. While machine learning algorithms have found their place in almost every aspect of modern particle physics, this thesis presents the employment of an artificial neural network for signal isolation. In this context, a neural network is expected to be able to extract the optimal selections from the kinematic information. A Keras-based deep neural network with three target categories is documented. For the purpose of optimisation an evolutionary algorithm is implemented based on the computational cluster at the Physikalisches Institut Bonn. Of particular importance is the demonstrated stability of the model, which makes the tool robust towards changes in the analysis workflow. Accordingly, the result of neural networks for three different analysis channels is reported on. In each case a stable separation is presented from both a general selection of backgrounds and a specific background.

Subsequently, a maximum likelihood fit is used to extract upper limits on the cross section of  $tH$ . For this, the neural network's application for region definition and as a fit distribution is demonstrated. In a combination of all channels a scale factor with respect to the Standard Model prediction is

extracted as:

$$\mu_{\text{combined}} = (-0.1^{+2.9}_{-2.6} \text{ (stat.) } ^{+1.9}_{-1.6} \text{ (sys.)}) = (-0.1^{+3.5}_{-3.1}). \quad (9.1)$$

The result implies good agreement with the Standard Model prediction for sign and magnitude of the probed coupling. Since no evidence for the production is found a limit is set to seven times the Standard Model prediction which matches the expected limit for the analysis.

In conclusion, a full exploitation of neural network-based signal isolation and a subsequent cross section estimation for the  $tH$  process are shown using Run 2 data from the ATLAS detector. Of great relevance is the neural network model's stability, which makes it a reliable link in the analysis chain.

---

## Gauge symmetry

---

The usage of gauge symmetries to establish a Lagrangian and to understand conservation laws is an essential part of the groundwork of particle physics. This appendix provides a simplified example of the establishment of a field interaction term via the requirement of an invariance. The Lagrangian of a free Dirac field representing a particle of mass  $m$  is given by:

$$\mathcal{L} = \bar{\Psi} \left( i\gamma^\mu \partial_\mu - m \right) \Psi. \quad (\text{A.1})$$

A possible local phase transformation for a field  $\Psi$  could be as follows

$$\Psi(x) \rightarrow \Psi' = e^{iq\chi(x)} \Psi(x). \quad (\text{A.2})$$

Transforming the associated Lagrangian according to this rule yields an additional term  $q\bar{\Psi}\gamma^\mu \left( \partial_\mu \chi(x) \right) \Psi$  making it not invariant under the proposed transformation:

$$\mathcal{L} \rightarrow \mathcal{L}' = e^{iq\chi(x)} \bar{\Psi} \gamma^\mu \left[ e^{iq\chi(x)} \partial_\mu \Psi + iq \left( \partial_\mu \chi(x) \right) e^{iq\chi(x)} \Psi \right] - m e^{iq\chi(x)} \bar{\Psi} e^{iq\chi(x)} \Psi, \quad (\text{A.3})$$

$$= \mathcal{L} - q\bar{\Psi}\gamma^\mu \left( \partial_\mu \chi(x) \right) \Psi. \quad (\text{A.4})$$

To restore the invariance one requires a covariant derivative  $D_\mu$  that introduces an additional field  $A_\mu$

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + iqA_\mu. \quad (\text{A.5})$$

The newly introduced field then has to be defined to transform as follows to cancel the initially undesired term:

$$A_\mu \rightarrow A'_\mu = A_\mu - \partial_\mu \chi. \quad (\text{A.6})$$

Following these basic steps, the original Lagrangian receives an additional interaction term and becomes invariant under the required transformation:

$$\mathcal{L} = \bar{\Psi} \left( i\gamma^\mu \partial_\mu - m \right) \Psi - q\bar{\Psi}\gamma^\mu A_\mu \Psi. \quad (\text{A.7})$$

Using the example of the electromagnetic interaction this interaction term with the new field could be interpreted as the interaction with the photon. The same prescription can be used to obtain the field terms of QCD. For the Higgs mechanism a similar method with a small twist is applied. This is described in appendix B. Please be aware that this appendix neither aims to be a complete introduction to the construction of an interaction term via Gauge Symmetry nor does it provide a complete construction of the Lagrangian of the electromagnetic interaction. The concept is merely introduced to fill one of the main gaps leading to the introduction of the Higgs mechanism. For a thorough understanding of the underlying formalism, see [4].

---

## Symmetry breaking of a complex scalar field

---

This appendix briefly covers the mathematical foundation of the creation of a mass term via the concept of local symmetry breaking. It does in no way attempt to give a complete introduction to the formalism but rather fills the gap in the description of the Higgs mechanism. For a full introduction see [12]. In the following a brief description of the perturbation at the minimum of a Mexican hat potential for a complex scalar field is given. This does not match the exact calculations for the two complex fields underlying the Standard Model Higgs mechanism. Instead, it is the minimal example that shows both the terms for the massive field and the massless field [5]. The starting point is a simple complex field with components  $\phi_1$  and  $\phi_2$

$$\phi = \frac{1}{\sqrt{2}} (\phi_1 + i\phi_2), \quad (\text{B.1})$$

and a corresponding Lagrangian

$$\mathcal{L} = (\partial_\mu \phi)^* (\partial^\mu \phi) - V(\phi), \quad (\text{B.2})$$

with a potential

$$V(\phi) = \mu^2 (\phi^* \phi) + \lambda (\phi^* \phi)^2. \quad (\text{B.3})$$

Writing out the Lagrangian with the complex components yields

$$\mathcal{L} = \frac{1}{2} (\partial_\mu \phi_1) (\partial^\mu \phi_1) + \frac{1}{2} (\partial_\mu \phi_2) (\partial^\mu \phi_2) - \frac{1}{2} \mu^2 (\phi_1^2 + \phi_2^2) - \frac{1}{4} \lambda (\phi_1^2 + \phi_2^2)^2. \quad (\text{B.4})$$

Assuming a negative value for the parameter  $\mu^2$  gives the potential the desired Mexican hat shape with a valley of minima defined by

$$\phi_1^2 + \phi_2^2 = \frac{-\mu^2}{\lambda} = \nu^2 \quad (\text{B.5})$$

and interpreted as the vacuum expectation value. The only following step is to perform a perturbation around one particular solution of the minimal valley,  $\phi_1(x) = \eta(x) + \nu$  and  $\phi_2(x) = \xi(x)$  [7]:

$$\phi = \frac{1}{\sqrt{2}} (\eta + \nu + i\xi). \quad (\text{B.6})$$

This results in a Lagrangian of the form

$$\mathcal{L} = \frac{1}{2} (\partial_\mu \eta) (\partial^\mu \eta) + \frac{1}{2} (\partial_\mu \xi) (\partial^\mu \xi) - V(\eta, \xi), \quad (\text{B.7})$$

where  $V(\eta, \xi)$  summarises all possible interaction terms

$$V(\eta, \xi) = -\frac{1}{4} \lambda v^4 + \lambda v^2 \eta^2 + \lambda v \eta^3 + \frac{1}{4} \lambda \eta^4 + \frac{1}{4} \lambda \xi^4 + \lambda v \eta \xi^2 + \frac{1}{2} \lambda \eta^2 \xi^2. \quad (\text{B.8})$$

The term that is quadratic in the perturbed field is interpreted as a mass term. This way both a mass term for the initial field and a new massless field is obtained yielding the following Lagrangian:

$$\mathcal{L} = \frac{1}{2} (\partial_\mu \eta) (\partial^\mu \eta) - \frac{1}{2} m_\mu^2 \eta^2 + \frac{1}{2} (\partial_\mu \xi) (\partial^\mu \xi) - V_{\text{int}}(\eta, \xi). \quad (\text{B.9})$$

The massless field is identified as a Goldstone boson and adds an additional degree of freedom. However, as gauge symmetry for the initial field was postulated the term can be removed by an additional gauge transformation which later can be identified as the unitary gauge [5]. This removal and the interpretation of the additional degree of freedom as the longitudinal freedom of a massive field is sometimes called eating the Goldstone boson.

---

## Data and MC samples

---

The analysed data event samples were taken from 25 ns pp collisions delivered by the LHC from 2015 to 2018 at  $\sqrt{s} = 13$  TeV and collected by the ATLAS detector. Events were selected from a common data stream using single-lepton triggers as described in references [103–105]. Events that fired single-electron triggers in Muon data stream were not selected in order to avoid double counting of events.

The registered data were filtered using good-runs lists<sup>1</sup>. The amount of data used by each analysis channel corresponds to an integrated luminosity of  $140 \text{ fb}^{-1}$ . The uncertainty in the combined 2015–2018 integrated luminosity is 0.83 % [106], obtained using the LUCID-2 detector [107] for the primary luminosity measurements, complemented by measurements using the inner detector and calorimeters. The partial and total integrated luminosities together with their uncertainties and some additional details are given in table C.1. Samples of events generated using MC simulations

Table C.1: Integrated luminosity per year with their relative uncertainties. Additionally, run numbers per year are shown.

Year	Periods	Run numbers	Number of events	Integrated luminosity [ $\text{pb}^{-1}$ ]
2015	D–H,J	276262–284484	220.58M	$3244.54 \pm 1.13\%$
2016	A–G,I,K,L	297730–311481	1057.84M	$33402.2 \pm 0.89\%$
2017	B–F,H,I,K	325713–340453	1340.80M	$44630.6 \pm 1.13\%$
2018	B–D,F,I,K,L,M,O,Q	348885–364292	1716.77M	$58791.6 \pm 1.10\%$
2015–2018	All	276262–364292	4335.99M	$140068.94 \pm 0.83\%$

are produced using different event generators interfaced to various shower/hadronisation generators within the MC16a/d/e production campaigns. After the event generation step, the trigger and detector simulation is performed with the dedicated ATLAS software infrastructure (named Athena) [108] either making fully use of the GEANT4 [109] framework for a detailed physics description simulation or At1fast2 [108] framework (AFII) for fast simulation. In these analyses, full-simulated (FS) event samples are always used as baseline samples unless not available. Thus, fast-simulated (AFII) event samples are used for some baseline samples (in particular for the  $tHq$  and  $tWH$  signal processes. The

<sup>1</sup> Good run lists document the periods of stable data taking. This excludes periods of poor read-out or detector response for a variety of reasons.

effect of multiple interactions in the same and neighbouring bunch crossings (PU) was modelled by overlaying the simulated hard-scattering event with inelastic pp events generated with Pythia 8.186 [110] using the NNPDF[2.3lo] set of parton distribution functions (PDFs) [111] and the third ATLAS set of tuned parameters for minimum-bias events (A3 tune) [112] over the original hard-scattering event. The MC events were weighted to reproduce the distribution of the average number of interactions per bunch crossing ( $\langle\mu\rangle$ ) observed in the data. The  $\langle\mu\rangle$  value in data was rescaled by a factor of  $1.03 \pm 0.04$  to improve agreement between data and simulation in the visible inelastic pp cross section [113]. In these analyses, samples of events generated using MC simulations are produced for  $tHq$  signal and most of the background processes, and are used to evaluate models of efficiency and resolution, and to estimate systematic uncertainties. Table C.2 summarises the simulated signal and background event samples used in the  $tHq$  multi-lepton analyses.

Table C.2: Summary of the baseline simulated signal and background event samples used in the  $tHq$  multi-lepton analyses. A renormalisation is applied to  $t\bar{t}W$  as documented in [114].

Process	Generator	Order (scheme)	PDF set	Parton shower	PDF set (tune)
Signal					
$tHq$	MADGRAPH5_AMC@NLO 2.6.2	NLO (4FS)	NNPDF3.0 <sub>NLO</sub> nf4	PYTHIA 8.230	NNPDF2.3 <sub>LO</sub> (A14 tune)
Backgrounds					
$t\bar{t}$	POWHEG BOX v2	NLO (5FS)	NNPDF3.0 <sub>NLO</sub>	PYTHIA 8.230	NNPDF2.3 <sub>LO</sub> (A14 tune)
$V$ +jets	SHERPA 2.2.1	NLO+LO	NNPDF3.0 <sub>NNLO</sub>	-	-
Diboson	SHERPA 2.2.1-2	NLO+LO	NNPDF3.0 <sub>NNLO</sub>	-	-
Triboson	SHERPA 2.2.2	NLO+LO	NNPDF3.0 <sub>NNLO</sub>	-	-
$t\bar{t}Z$	MADGRAPH5_AMC@NLO 2.3.3	NLO	NNPDF3.0 <sub>NLO</sub>	PYTHIA 8.210	NNPDF2.3 <sub>LO</sub> (A14 tune)
$t\bar{t}W$	SHERPA 2.2.10	NLO	NNPDF3.0 <sub>NNLO</sub>	-	-
$t\bar{t}H$	POWHEG BOX v2	NLO (5FS)	NNPDF3.0 <sub>NLO</sub>	PYTHIA 8.230	NNPDF2.3 <sub>LO</sub> (A14 tune)
$t$ -channel	POWHEG BOX v2	NLO (4FS)	NNPDF3.0 <sub>NLO</sub> nf4	PYTHIA 8.230	NNPDF2.3 <sub>LO</sub> (A14 tune)
$tW$	POWHEG BOX v2	NLO (5FS, DR)	NNPDF3.0 <sub>NLO</sub>	PYTHIA 8.230	NNPDF2.3 <sub>LO</sub> (A14 tune)
$s$ -channel	POWHEG BOX v2	NLO	NNPDF3.0 <sub>NLO</sub>	PYTHIA 8.230	NNPDF2.3 <sub>LO</sub> (A14 tune)
$tZq$	MADGRAPH5_AMC@NLO 2.3.3	NLO	NNPDF3.0 <sub>NLO</sub>	PYTHIA 8.230	NNPDF2.3 <sub>LO</sub> (A14 tune)
$tWH$	MADGRAPH5_AMC@NLO 2.8.1	NLO (5FS, DR)	NNPDF3.0 <sub>NLO</sub>	PYTHIA 8.245p3	NNPDF2.3 <sub>LO</sub> (A14 tune)
$tWZ$	MADGRAPH5_AMC@NLO 2.3.3	NLO	NNPDF3.0 <sub>NLO</sub>	PYTHIA 8.212	NNPDF2.3 <sub>LO</sub> (A14 tune)
$ttt$	MADGRAPH5_AMC@NLO 2.2.2	NLO	NNPDF3.1 <sub>NLO</sub>	PYTHIA 8.186	NNPDF2.3 <sub>LO</sub> (A14 tune)
$t\bar{t}t$	MADGRAPH5_AMC@NLO 2.3.3	NLO	NNPDF3.1 <sub>NLO</sub>	PYTHIA 8.230	NNPDF2.3 <sub>LO</sub> (A14 tune)
$ggH$	POWHEG BOX v2	NLO	CT10	PYTHIA 8.210	CTEQ6L1 (AZNLO tune)
$qqH$	POWHEG BOX v1	NLO	CT10	PYTHIA 8.186	CTEQ6L1 (AZNLO tune)
$WH$	PYTHIA 8.186	LO	NNPDF2.3 <sub>LO</sub>	-	-
$ZH$	PYTHIA 8.186	LO	NNPDF2.3 <sub>LO</sub>	-	-



## Fake lepton estimation

This appendix summarises the fake lepton correction factors calculated for the channels and regions. Depending on the method yields and scale factors are documented per region.

Table D.1: Yields and estimated  $\tau_{\text{had}}$  fake factors for the  $1\ell + 2\tau_{\text{had}}$  region.

$p_T(\tau_{\text{had}})$ [GeV]	$N_{\text{data}}$	$N_{\text{MC}}^{\tau+e+\mu}$	$N_{\text{MC}}^{\text{unknown}}$	$N_{\text{MC}}^{q+g}$	$N^{q+g}$ est.	SF( $q+g$ )
1-prong, 1 b-jet						
20–30	3 753	17.1	967.4	2 842.8	$2\,768.6 \pm 67.1$	$0.974 \pm 0.028$
30–40	1 110	6.0	148.9	1 020.4	$955.0 \pm 34.9$	$0.936 \pm 0.039$
>40	954	12.4	52.6	1 058.4	$888.9 \pm 31.3$	$0.840 \pm 0.032$
3-prong, 1 b-jet						
20–30	2 564	11.1	294.2	2 323.9	$2\,258.7 \pm 52.9$	$0.972 \pm 0.028$
30–40	902	3.7	47.1	937.9	$851.1 \pm 30.2$	$0.907 \pm 0.037$
>40	623	9.0	22.2	773.4	$591.8 \pm 26.3$	$0.765 \pm 0.036$
1-prong, 2 b-jet						
20–30	1 046	4.1	315.5	706.4	$726.5 \pm 33.0$	$1.028 \pm 0.049$
30–40	314	1.8	50.5	257.8	$261.7 \pm 17.9$	$1.015 \pm 0.073$
>40	270	3.7	16.3	300.3	$250.0 \pm 16.6$	$0.833 \pm 0.058$
3-prong, 2 b-jet						
20–30	768	2.1	81.3	620.7	$684.6 \pm 27.9$	$1.103 \pm 0.050$
30–40	240	0.9	14.5	269.2	$224.6 \pm 15.6$	$0.834 \pm 0.061$
>40	212	2.5	2.2	226.5	$207.3 \pm 14.6$	$0.916 \pm 0.068$

## Appendix D Fake lepton estimation

Table D.2: Summary of the jet  $\tau_{\text{had}}$  fakes obtained using the template fit method in the  $2\ell$  OS +  $1\tau_{\text{had}}$  channel. The uncertainties on the scale factors for quark- and gluon-initiated jets are anti-correlated due to the constraint  $f_q + f_g = 1$ .

$p_T(\tau_{\text{had}})$ [GeV]	$f_S$ (fixed)	$f_q$ (derived)	$q$ -jet SF	$g$ -jet SF
1-prong, 1 b-jet				
20–30	0.1	$0.41 \pm 0.04$	$0.63 \pm 0.07$	$2.15 \mp 0.18$
30–40	0.06	$0.59 \pm 0.08$	$0.72 \pm 0.10$	$2.02 \mp 0.46$
>40	0.04	$0.69 \pm 0.06$	$0.67 \pm 0.06$	$2.30 \mp 0.48$
3-prong, 1 b-jet				
20–30	0.05	$0.68 \pm 0.05$	$0.85 \pm 0.06$	$1.59 \mp 0.27$
30–40	0.02	$0.80 \pm 0.08$	$0.85 \pm 0.08$	$1.24 \mp 0.51$
>40	0.03	$0.81 \pm 0.07$	$0.70 \pm 0.06$	$1.57 \mp 0.67$
1-prong, 2 b-jets				
20–30	0.11	$0.43 \pm 0.05$	$0.75 \pm 0.09$	$1.90 \mp 0.22$
30–40	0.06	$0.59 \pm 0.09$	$0.83 \pm 0.12$	$1.93 \mp 0.49$
>40	0.04	$0.79 \pm 0.07$	$0.79 \pm 0.07$	$1.45 \mp 0.58$
3-prong, 2 b-jets				
20–30	0.03	$0.75 \pm 0.05$	$0.99 \pm 0.06$	$1.66 \mp 0.34$
30–40	0.02	$0.88 \pm 0.07$	$0.82 \pm 0.08$	$1.04 \mp 0.58$
>40	0.02	$0.97 \pm 0.07$	$0.88 \pm 0.08$	$1.33 \mp 1.22$

Table D.3: Event yields and derived scale factors for the light lepton fakes in the  $2\ell$  OS +  $1\tau_{\text{had}}$  channel. The templates with large scale factor uncertainties are not corrected in the signal region, and a 100% uncertainty is assigned instead.

CR name	$N_{\text{data}}$	$N_{\text{sim,sim}}$	$N_{\text{sim,jet}}$	$N_{\text{jet,sim}}$	$N_{\text{jet,jet}}$
$T\bar{T} (n_b = 1)$	4975	2536.4	2224.3	1.2	0.9
$\bar{T}T (n_b = 1)$	1057	687.6	0.9	282.3	0.1
$\bar{T}\bar{T} (n_b = 1)$	797	257.6	191.5	71.9	17.1
Measured SF			$1.09 \pm 0.02$	$1.30 \pm 0.11$	$13.87 \pm 3.08$
$T\bar{T} (n_b = 2)$	921	649.0	167.1	0.1	0.0
$\bar{T}T (n_b = 2)$	197	189.2	0.2	26.9	0.1
$\bar{T}\bar{T} (n_b = 2)$	120	72.1	17.6	4.3	1.0
Measured SF			$1.62 \pm 0.09$	$0.19 \pm 0.72$	$18.75 \pm 24.31$

Table D.4: Scale factors for the  $\tau_{\text{had}}$  fakes in the  $2\ell$  SS +  $1\tau_{\text{had}}$  channel.

$p_T(\tau_{\text{had}})$ bin [GeV]	1-prong	3-prong
20-30	$3.6 \pm 1.2$	$1.1 \pm 0.7$
30-40	$1.7 \pm 1.0$	$3.2 \pm 2.3$
>40	$1.6 \pm 0.8$	$0.1 \pm 1.0$

Table D.5: Scale factors for the light lepton fakes in the  $2\ell$  SS +  $1\tau_{\text{had}}$  channel. In the 2 b-jet region for (fake, fake) the factor is taken from MC with a 100% uncertainty.

$p_T(\tau_{\text{had}})$ bin [GeV]	1 b-jet	2 b-jet
fake, true	$0.63 \pm 0.01$	$0.63 \pm 0.04$
true, fake	$0.52 \pm 0.04$	$0.94 \pm 0.23$
fake, fake	$6.42 \pm 3.35$	$1.0 \pm 1.0$

Table D.6: Scale factors for  $\tau_{\text{had}}$  fakes in the  $1\ell + 2\tau_{\text{had}}$  channel in the subregions. tau indicates a truth  $\tau$  and jet indicates a fake.

template	1 b-jet	2 b-jets
1-prong, 1-prong		
tau+jet	$1.27 \pm 0.21$	$1.18 \pm 0.22$
jet+tau	$1.54 \pm 0.58$	$1.05 \pm 0.49$
jet+jet	$0.65 \pm 0.12$	$0.71 \pm 0.15$
1-prong, 3-prong		
tau+jet	$1.58 \pm 0.63$	$0.92 \pm 0.24$
jet+tau	$4.36 \pm 2.77$	$0.12 \pm 0.49$
jet+jet	$0.12 \pm 0.44$	$1.11 \pm 0.17$
3-prong, 1-prong		
tau+jet	$1.03 \pm 0.29$	$1.58 \pm 0.43$
jet+tau	$1.73 \pm 0.45$	$1.93 \pm 0.69$
jet+jet	$0.74 \pm 0.11$	$0.58 \pm 0.25$
3-prong, 3-prong		
tau+jet	$0.64 \pm 0.38$	$0.76 \pm 0.37$
jet+tau	$2.92 \pm 1.52$	$1.66 \pm 1.28$
jet+jet	$0.85 \pm 0.19$	$1.23 \pm 0.26$



# Feature plots $2\ell$ OS + $1\tau_{\text{had}}$

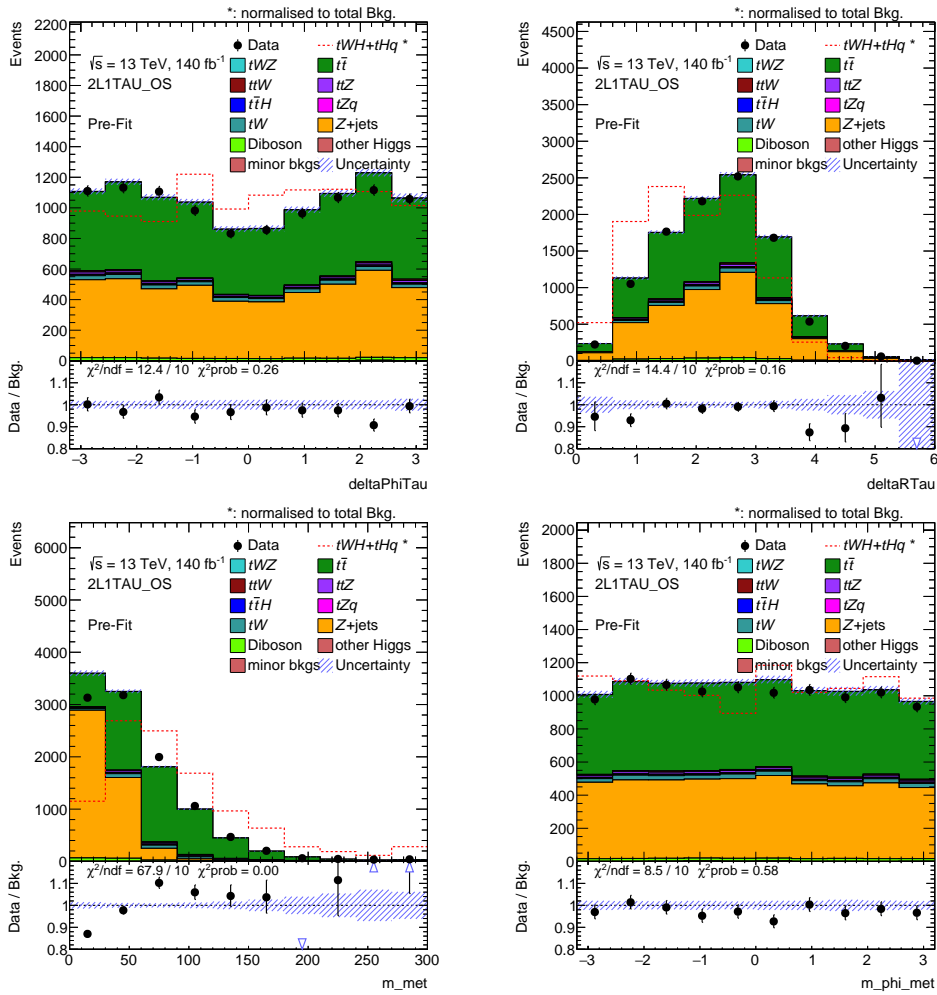
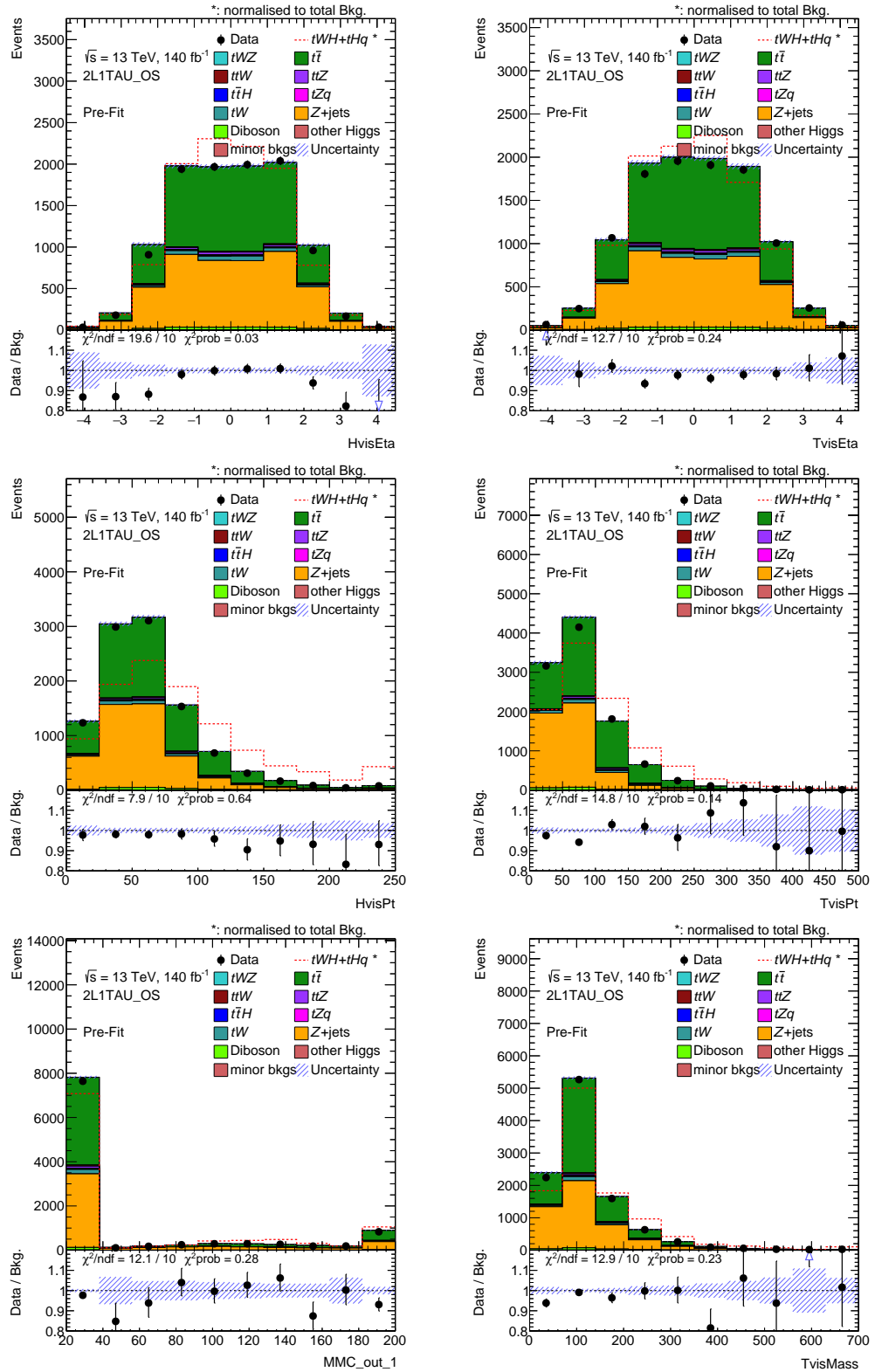


Figure E.1: Data to MC agreement for the features used in the neural network training for the  $2\ell$  OS +  $1\tau_{\text{had}}$  channel. For a definition of the features see section 7.2


 Figure E.1: Data to MC agreement for the features used in the neural network training for the  $2\ell$  OS +  $1\tau_{\text{had}}$  channel. For a definition of the features see section 7.2

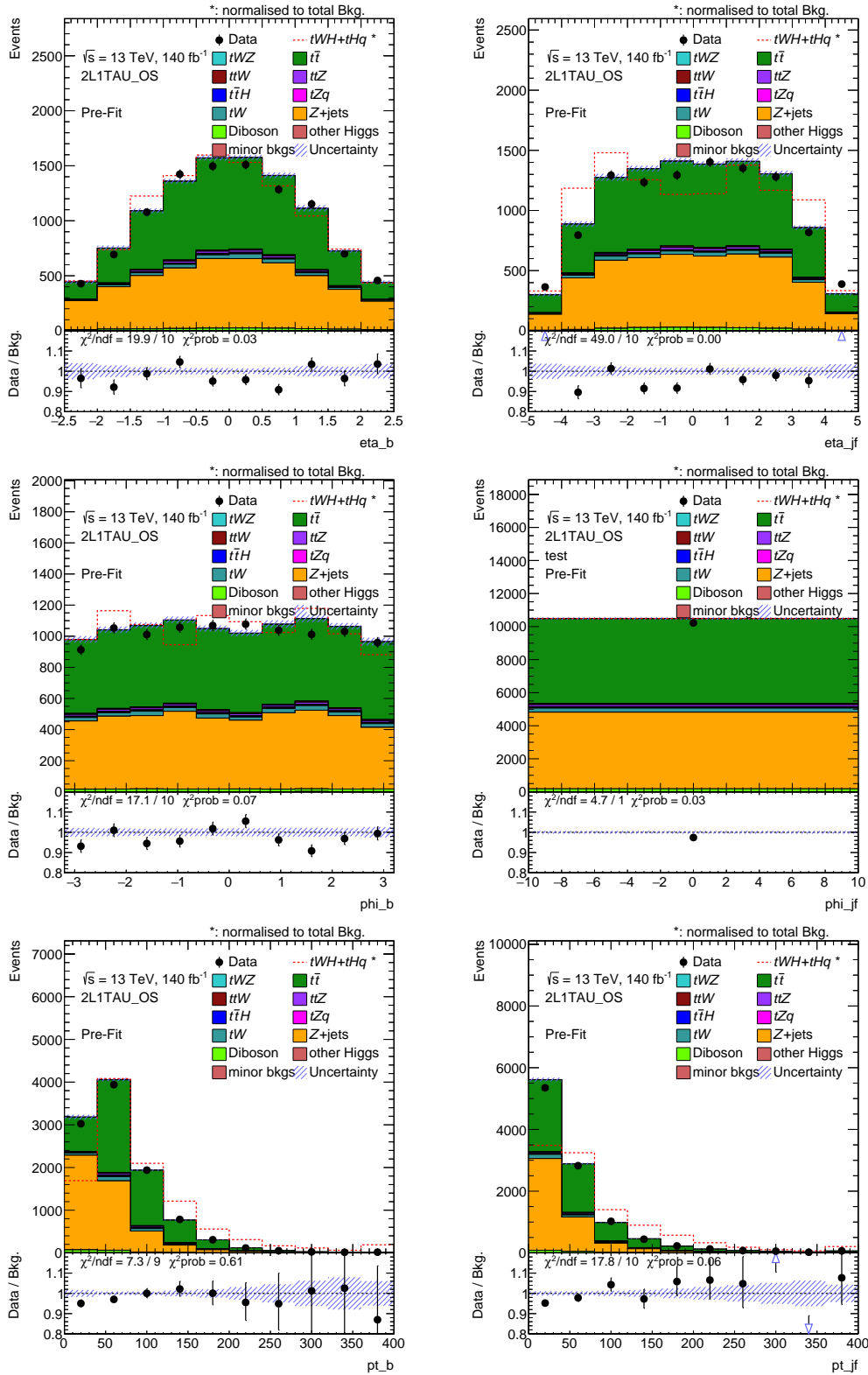
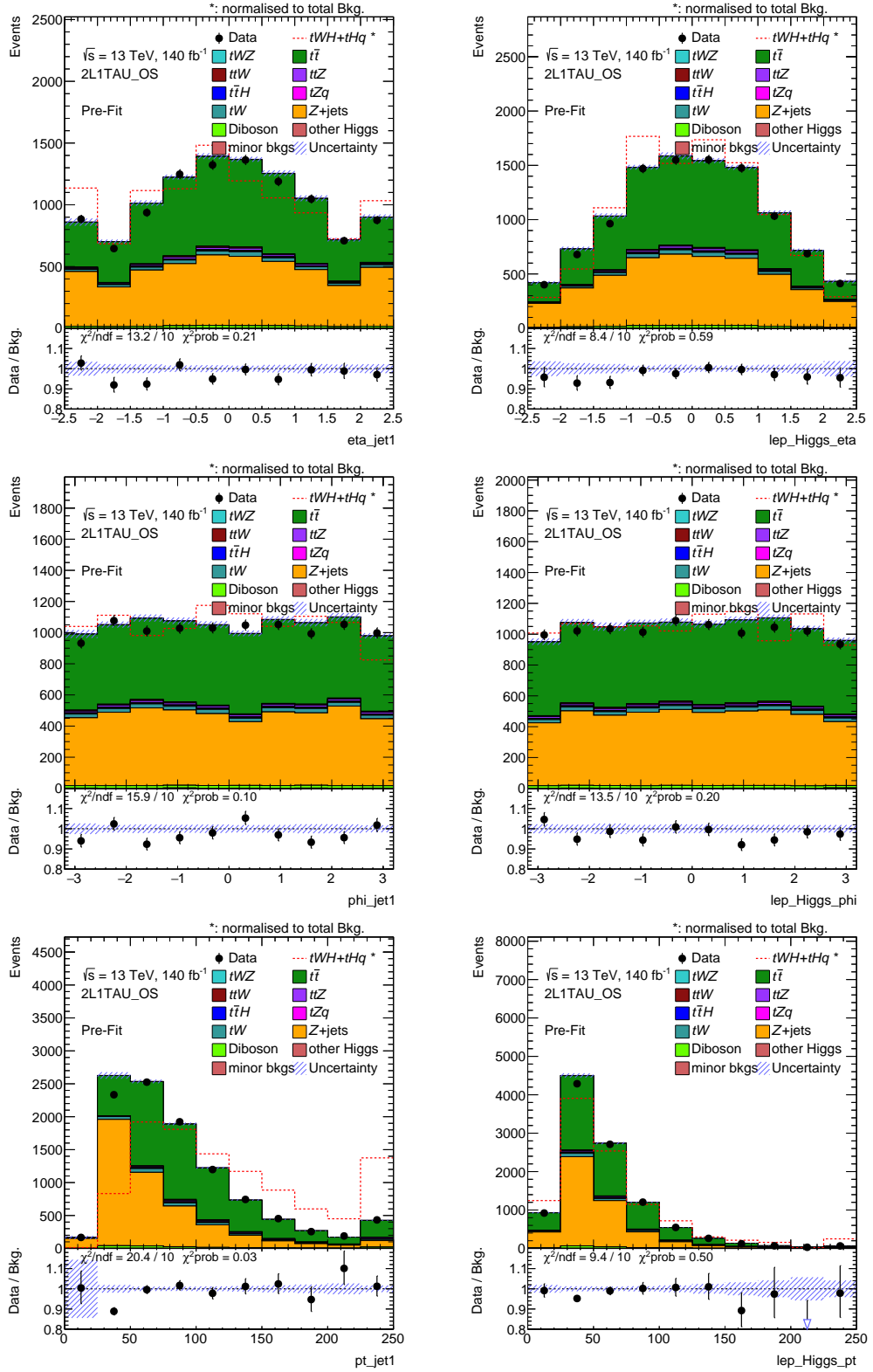


Figure E.1: Data to MC agreement for the features used in the neural network training for the  $2\ell$  OS +  $1\tau_{\text{had}}$  channel. For a definition of the features see section 7.2


 Figure E.1: Data to MC agreement for the features used in the neural network training for the  $2\ell$  OS +  $1\tau_{\text{had}}$  channel. For a definition of the features see section 7.2



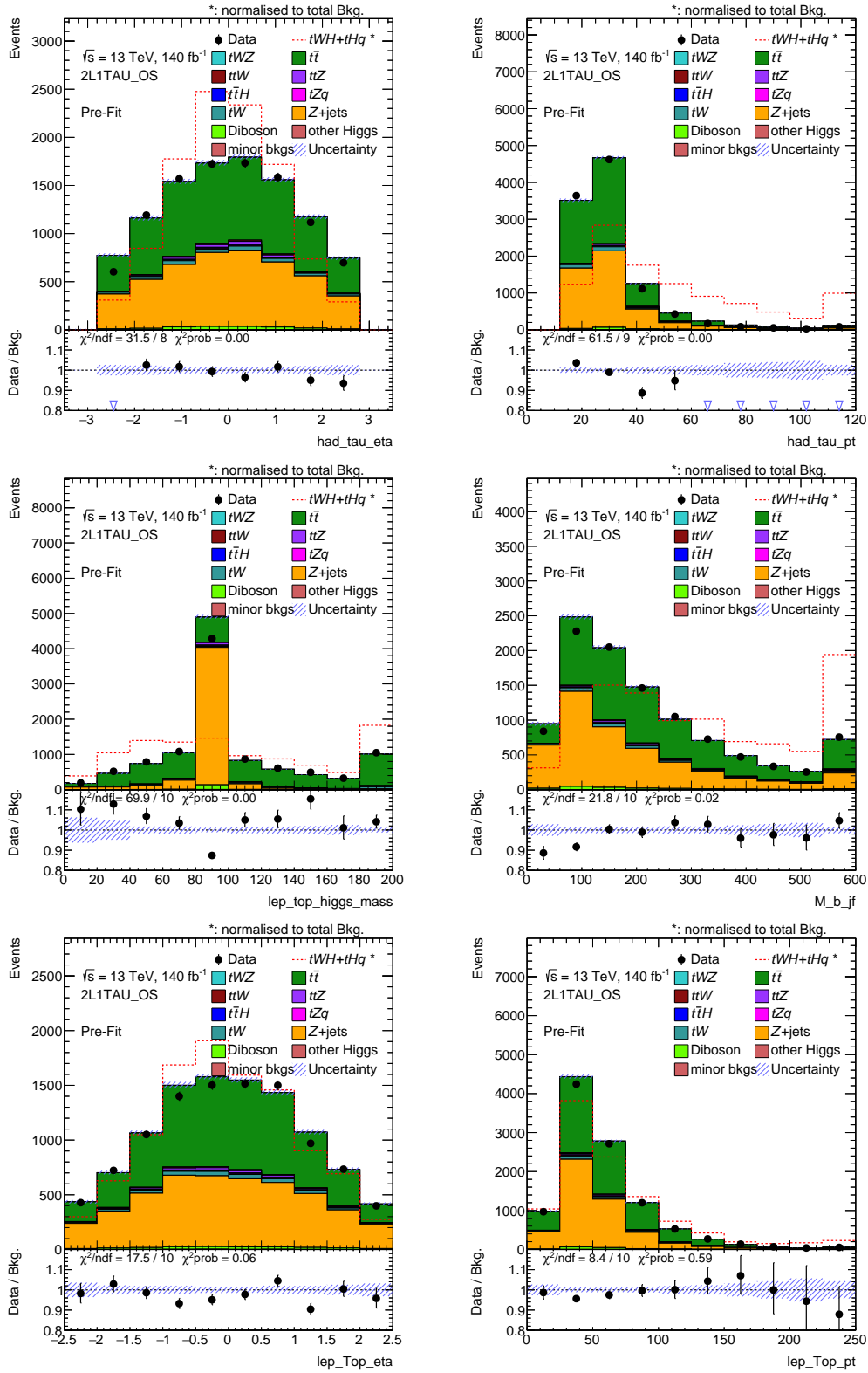


Figure E.1: Data to MC agreement for the features used in the neural network training for the  $2\ell$  OS +  $1\tau_{\text{had}}$  channel. For a definition of the features see section 7.2

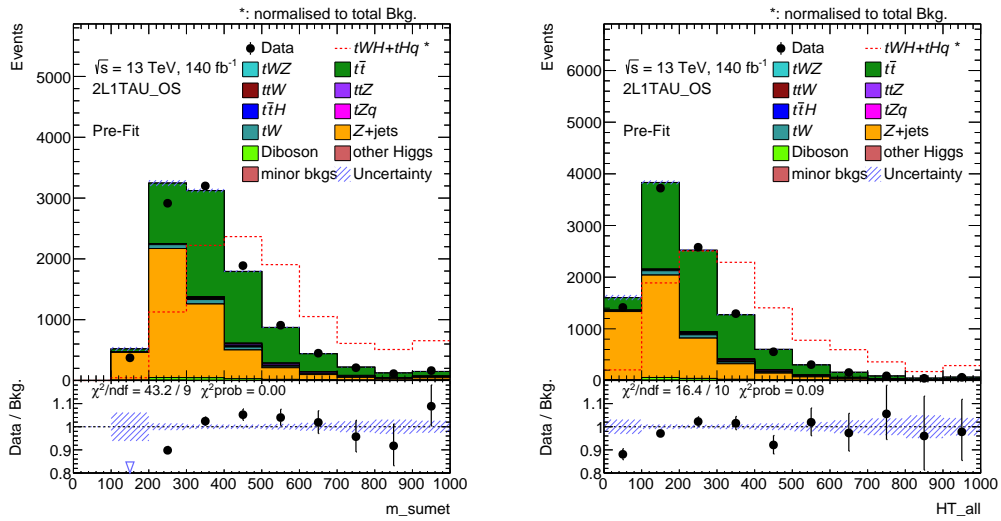


Figure E.1: Data to MC agreement for the features used in the neural network training for the  $2\ell$  OS +  $1\tau_{\text{had}}$  channel. For a definition of the features see section 7.2

## Feature plots $2\ell$ SS + $1\tau_{\text{had}}$

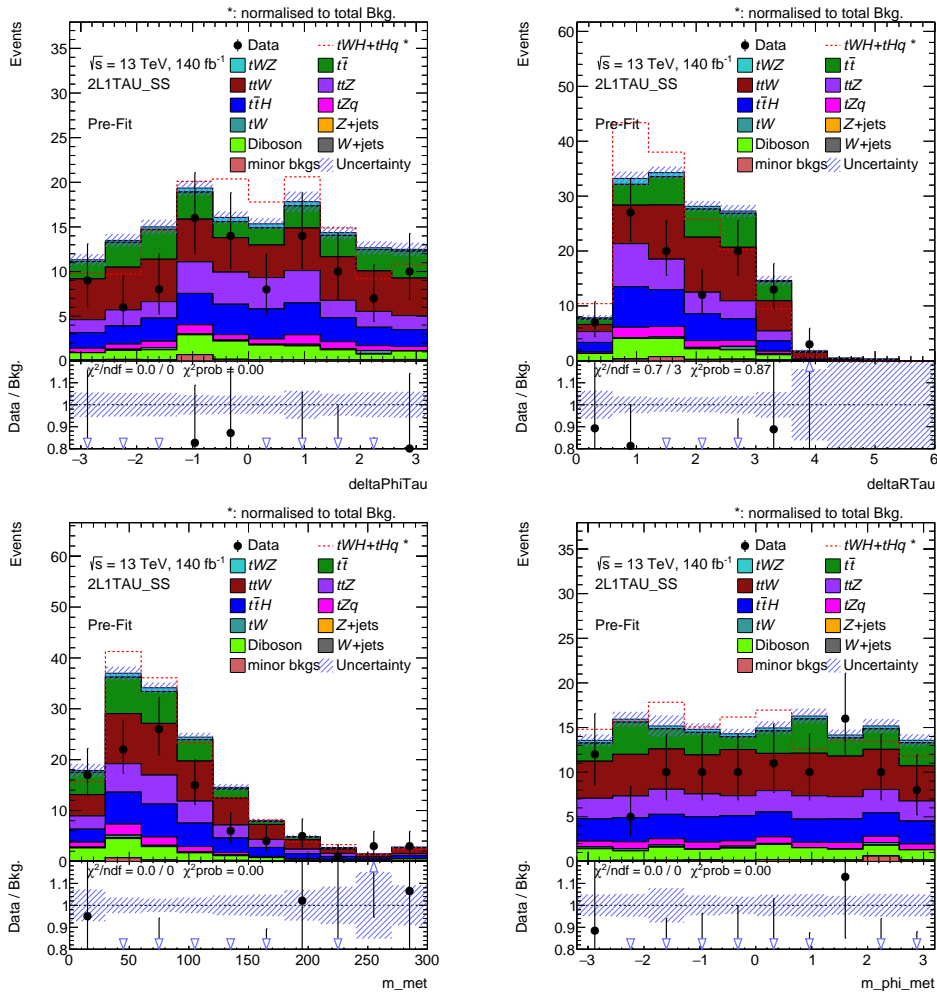
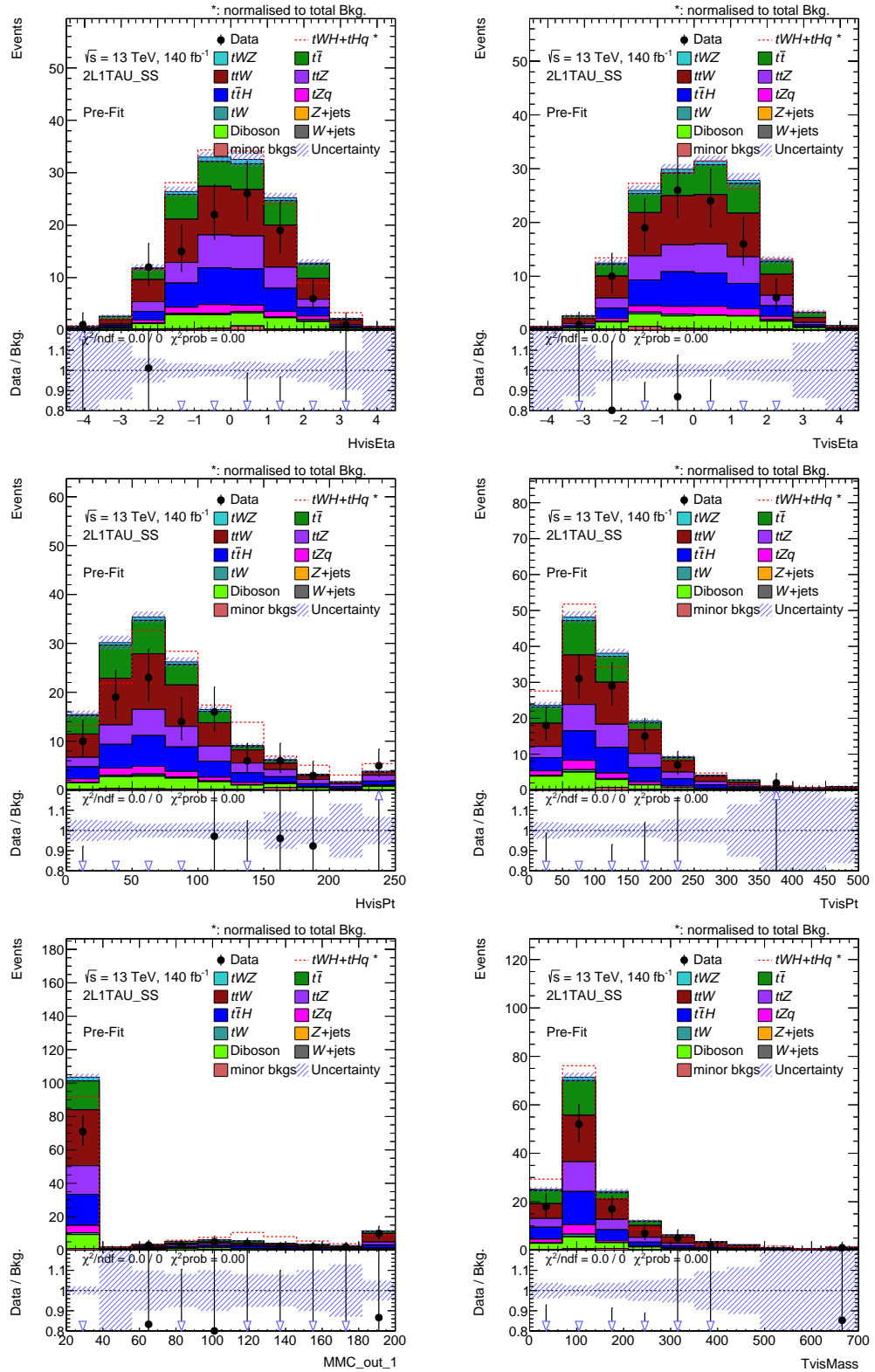


Figure F.1: Data to MC agreement for the features used in the neural network training for the  $2\ell$  SS +  $1\tau_{\text{had}}$  channel. For a definition of the features see section 7.2


 Figure F.1: Data to MC agreement for the features used in the neural network training for the  $2\ell$  SS +  $1\tau_{\text{had}}$  channel. For a definition of the features see section 7.2

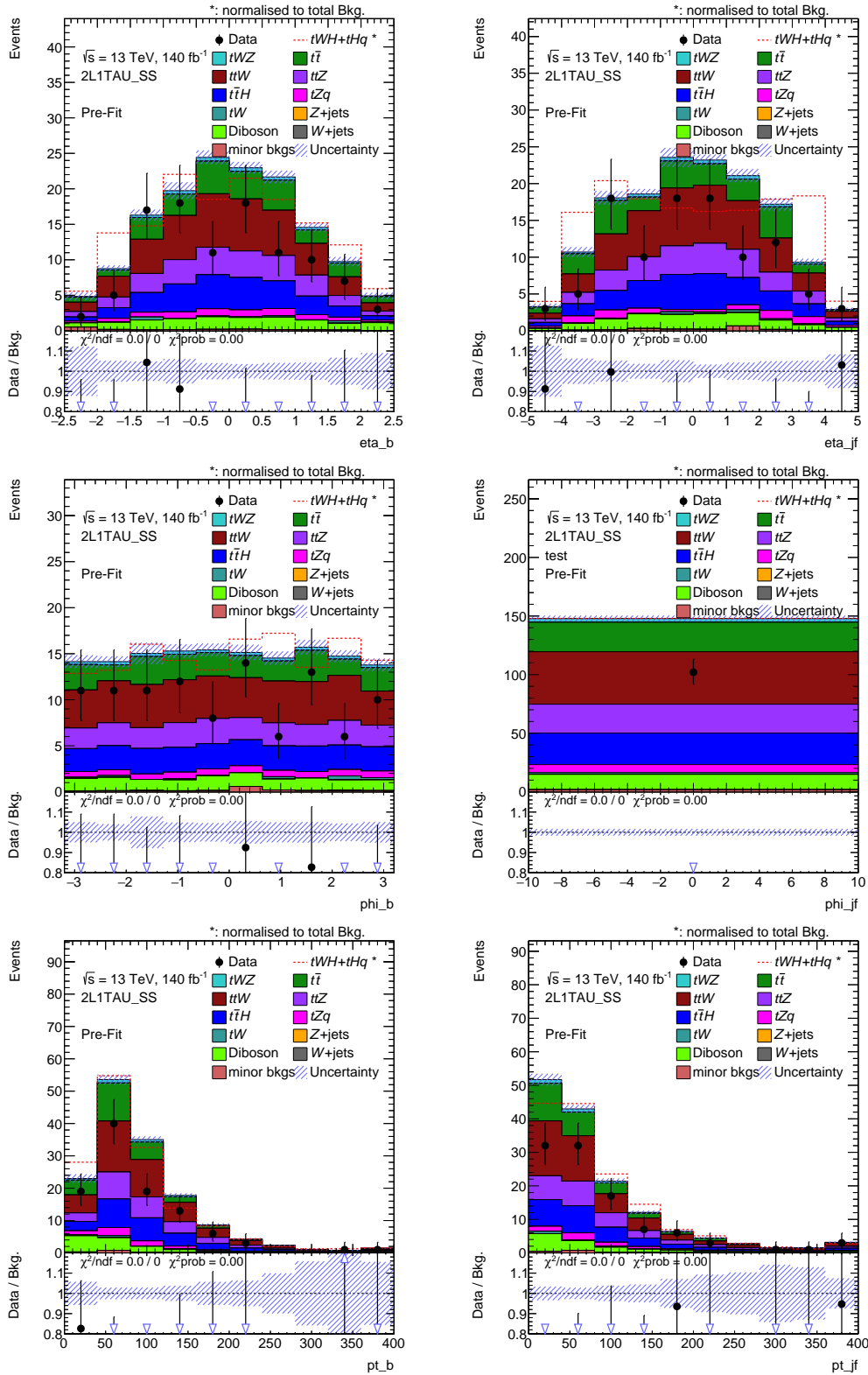
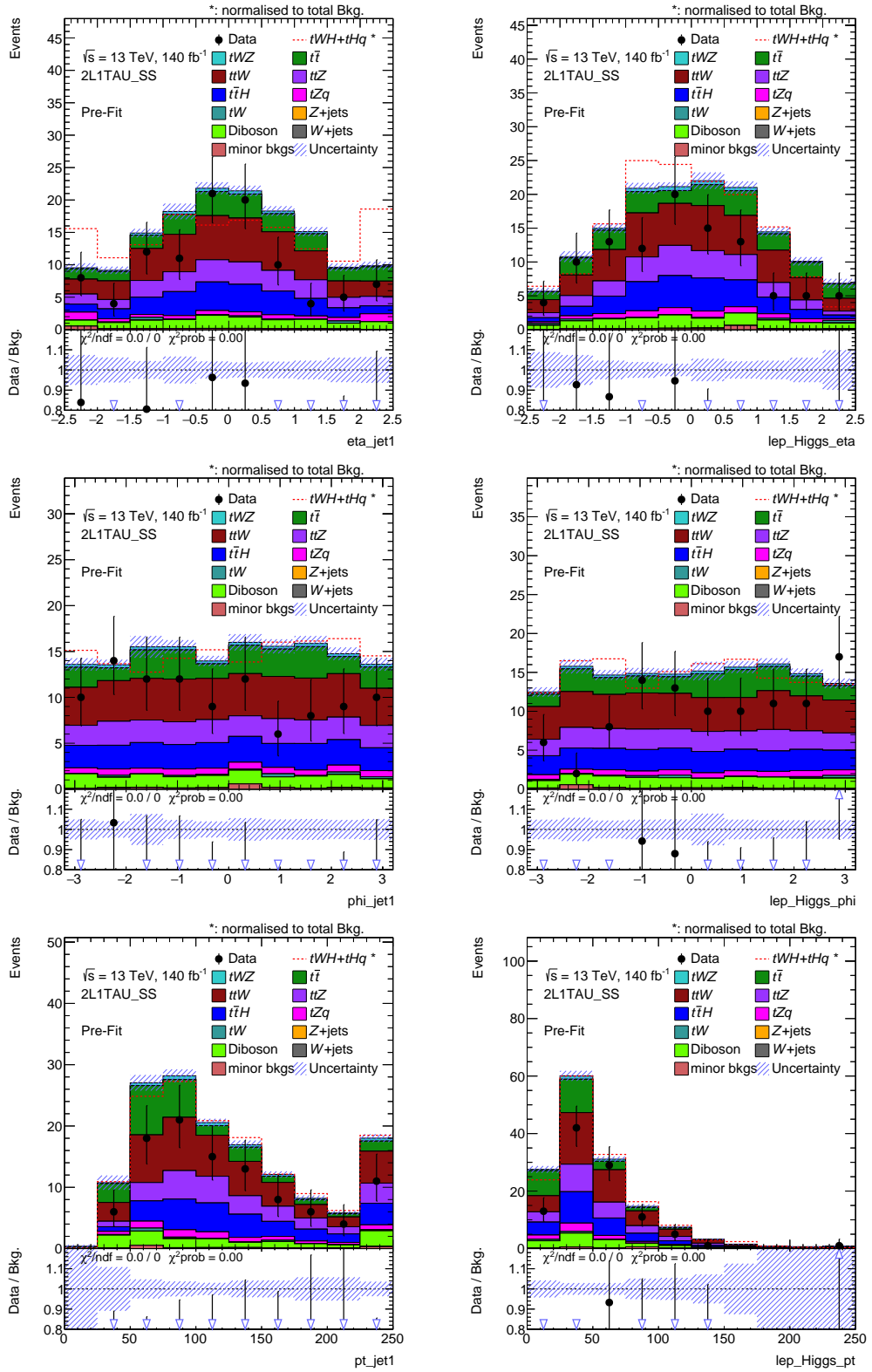


Figure F.1: Data to MC agreement for the features used in the neural network training for the  $2\ell$  SS +  $1\tau_{\text{had}}$  channel. For a definition of the features see section 7.2


 Figure F.1: Data to MC agreement for the features used in the neural network training for the  $2\ell$  SS +  $1\tau_{\text{had}}$  channel. For a definition of the features see section 7.2

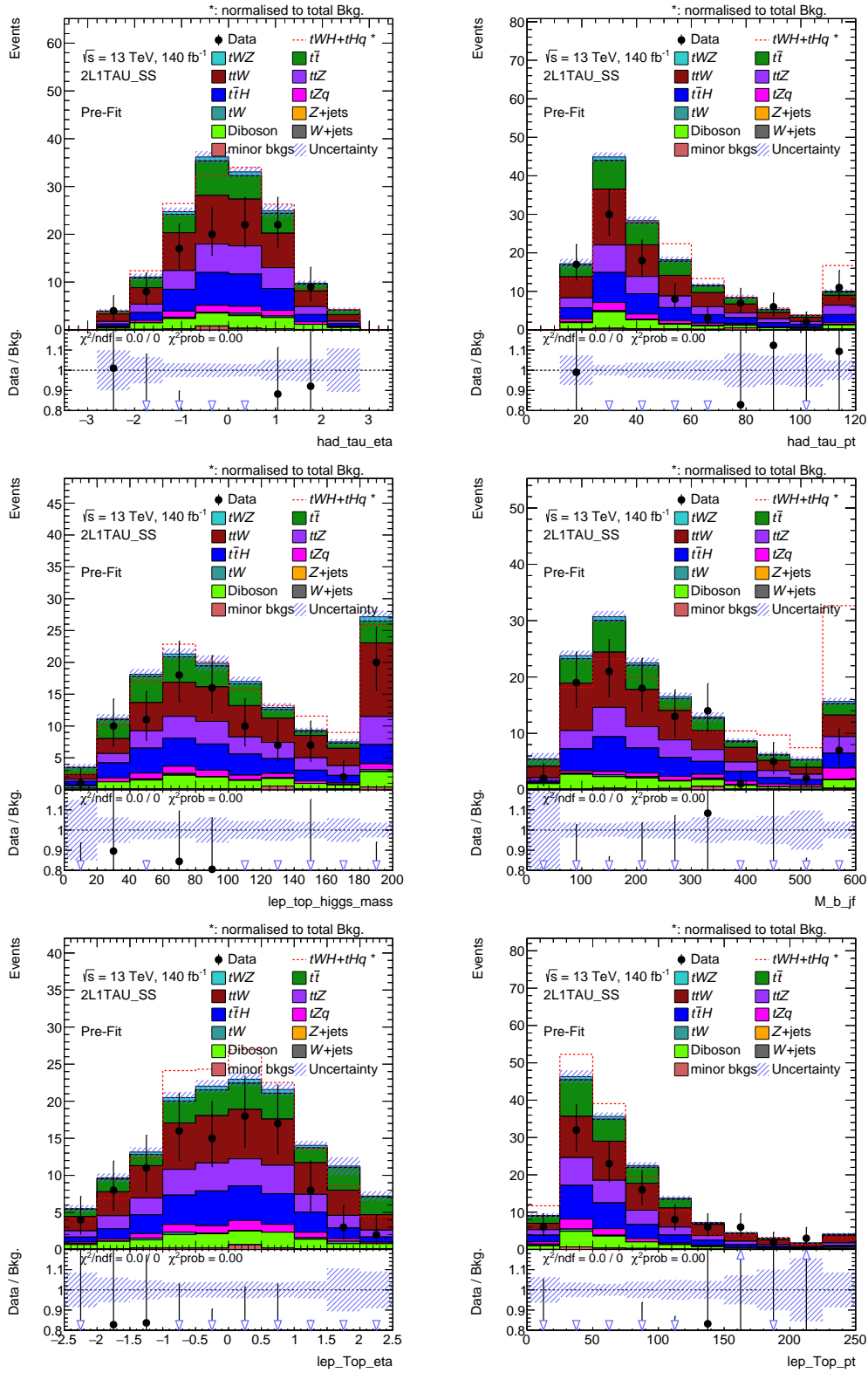


Figure F.1: Data to MC agreement for the features used in the neural network training for the  $2\ell$  SS +  $1\tau_{\text{had}}$  channel. For a definition of the features see section 7.2

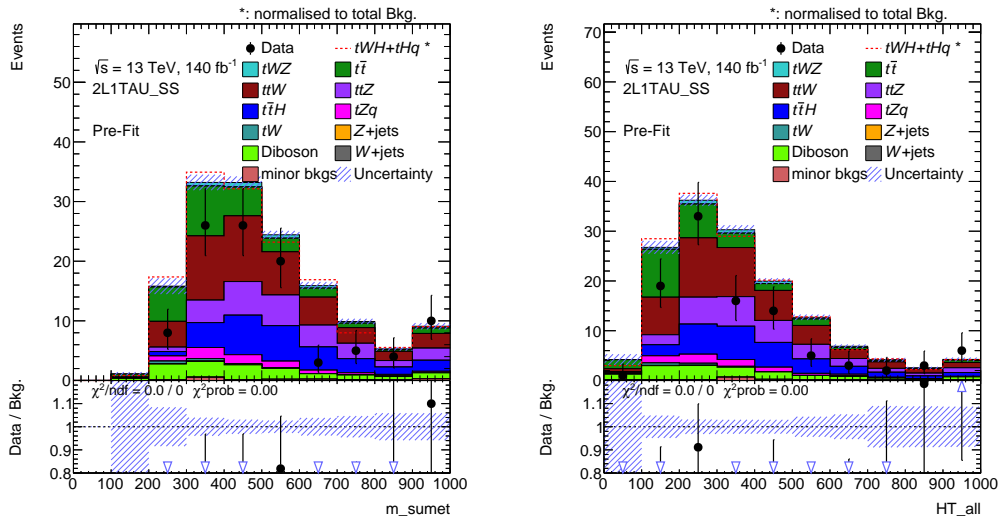


Figure F.1: Data to MC agreement for the features used in the neural network training for the  $2\ell$  SS +  $1\tau_{\text{had}}$  channel. For a definition of the features see section 7.2



# Feature plots $1\ell + 2\tau_{\text{had}}$

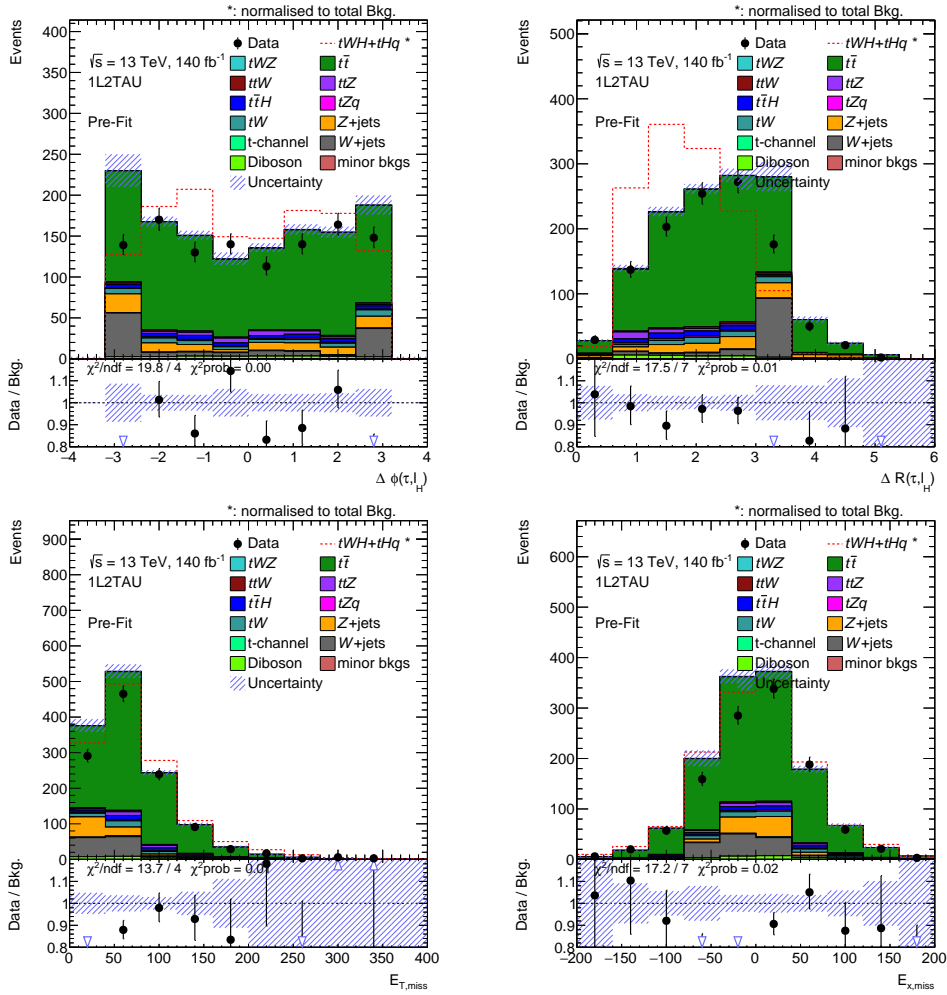


Figure G.1: Data to MC agreement for the features used in the neural network training for the  $1\ell + 2\tau_{\text{had}}$  channel. For a definition of the features see section 7.2

Appendix G Feature plots  $1\ell + 2\tau_{\text{had}}$

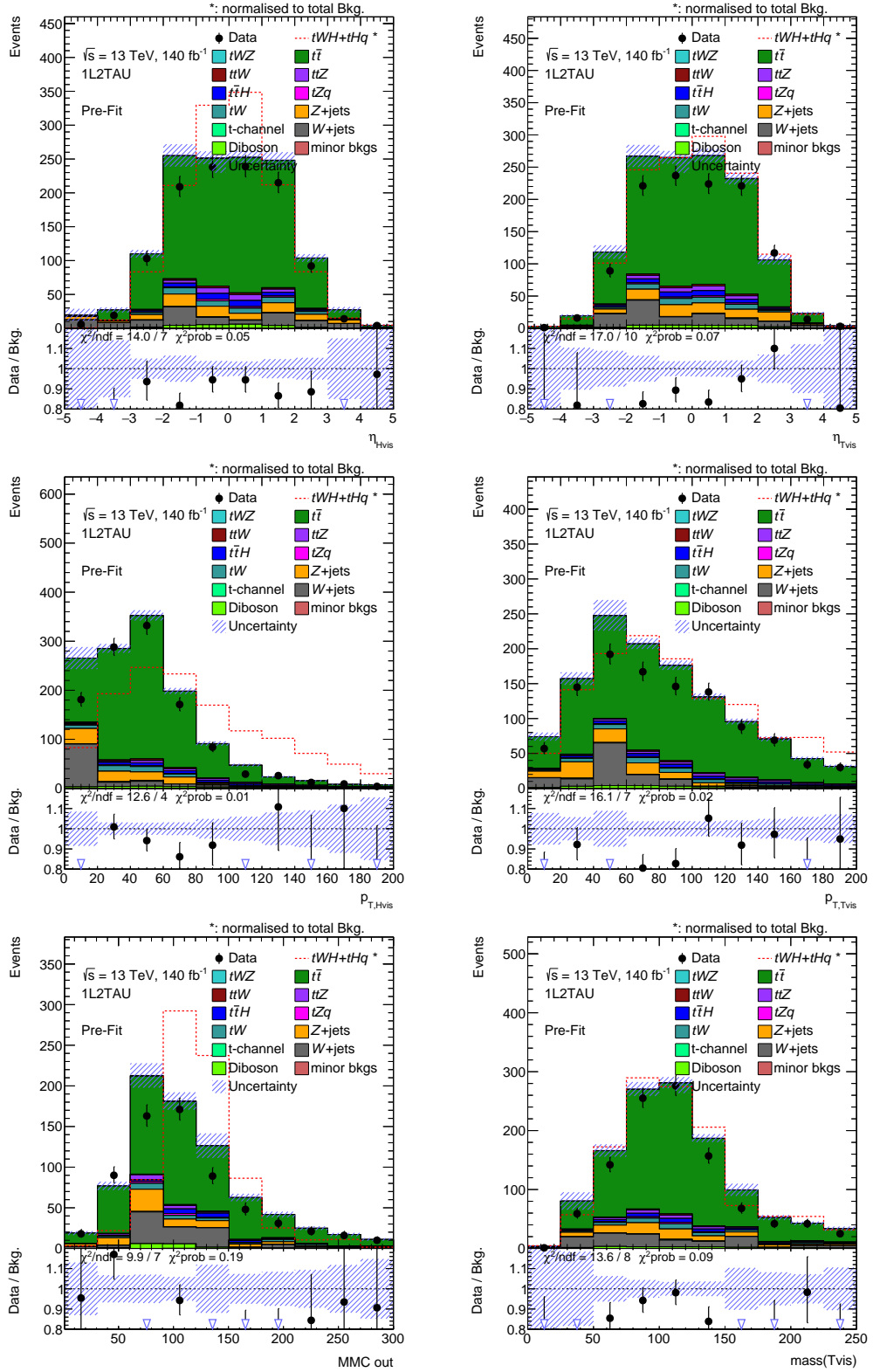


Figure G.1: Data to MC agreement for the features used in the neural network training for the  $1\ell + 2\tau_{\text{had}}$  channel. For a definition of the features see section 7.2

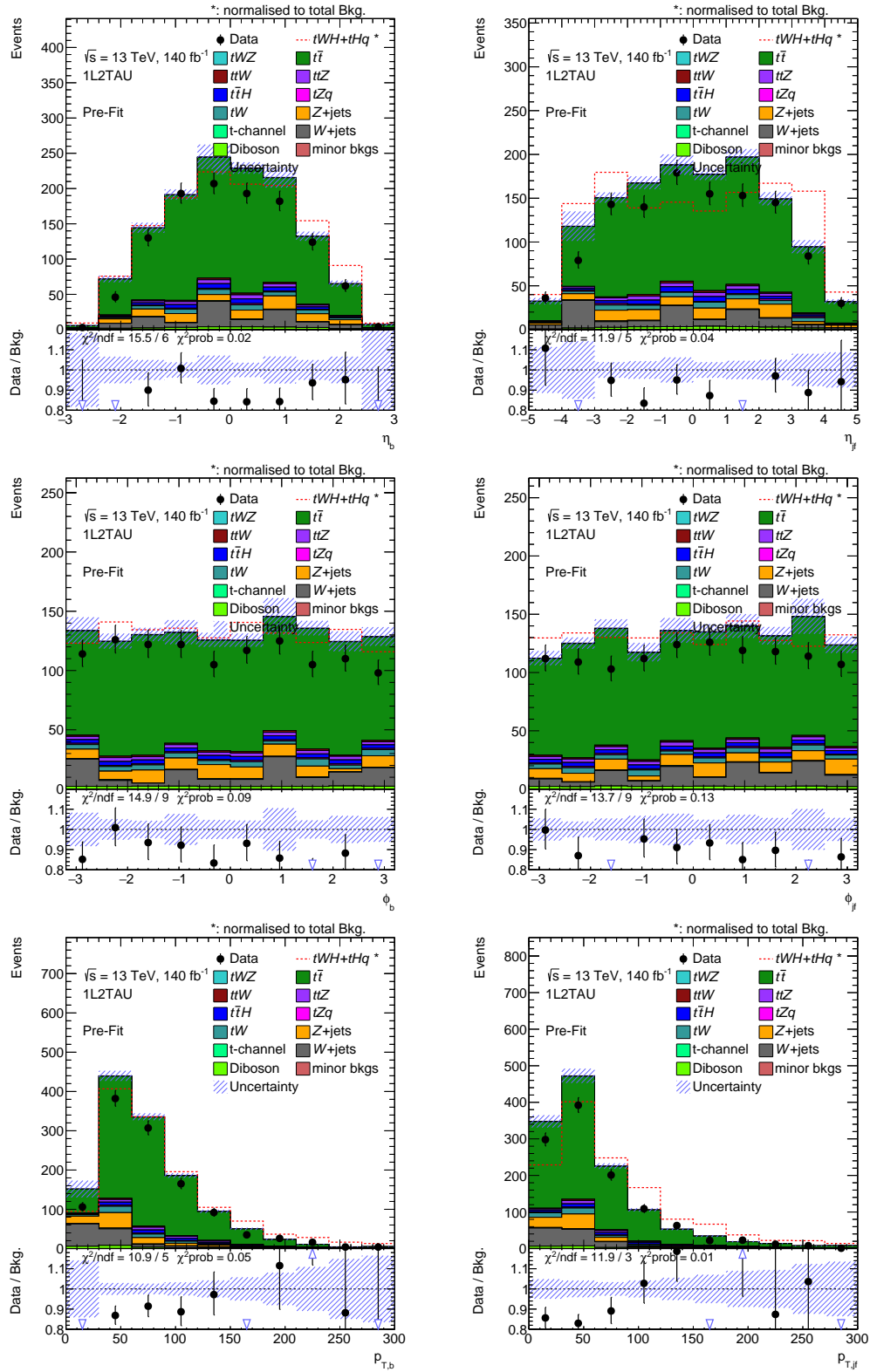


Figure G.1: Data to MC agreement for the features used in the neural network training for the  $1\ell + 2\tau_{\text{had}}$  channel. For a definition of the features see section 7.2

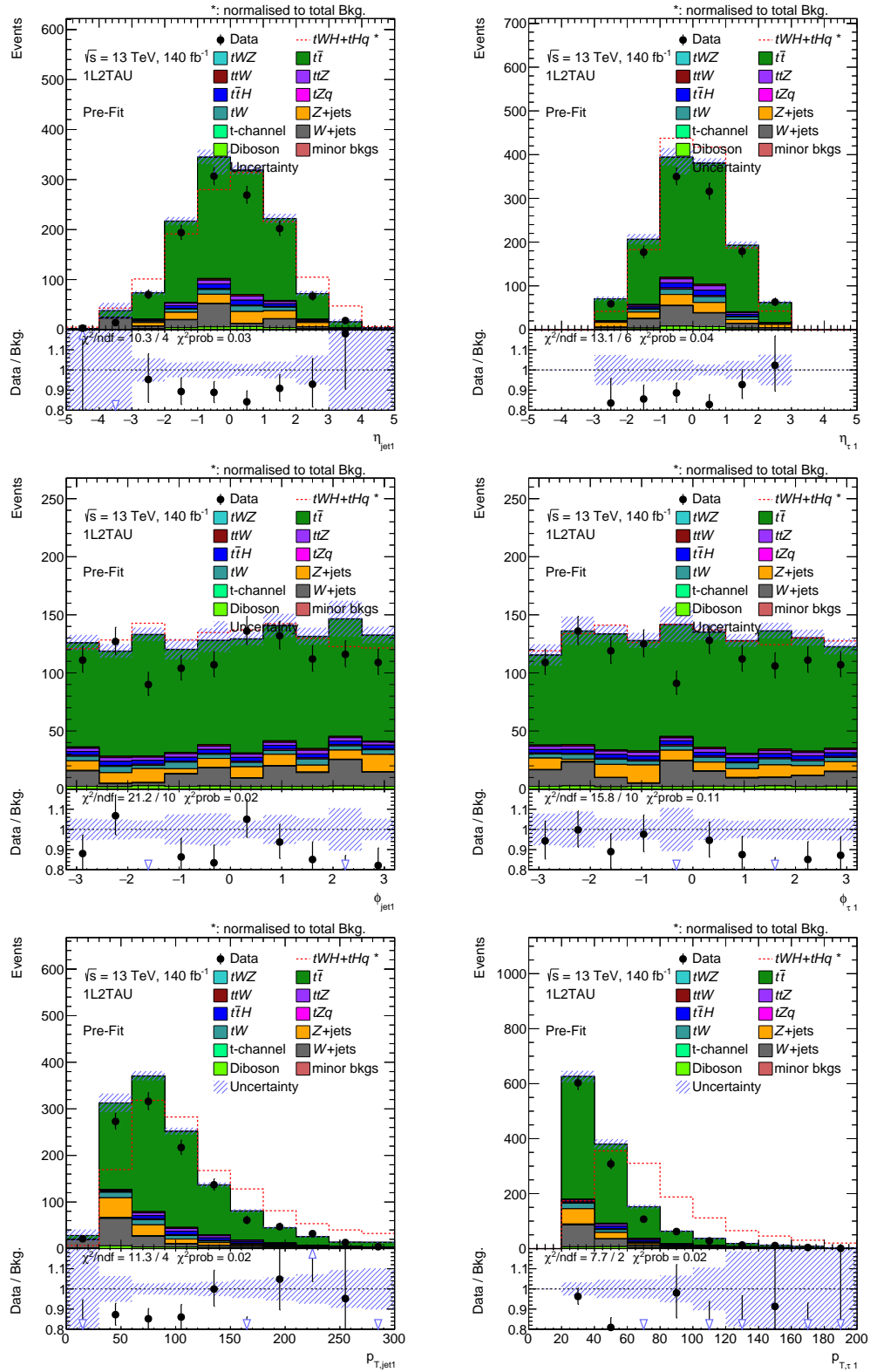


Figure G.1: Data to MC agreement for the features used in the neural network training for the  $1\ell + 2\tau_{\text{had}}$  channel. For a definition of the features see section 7.2

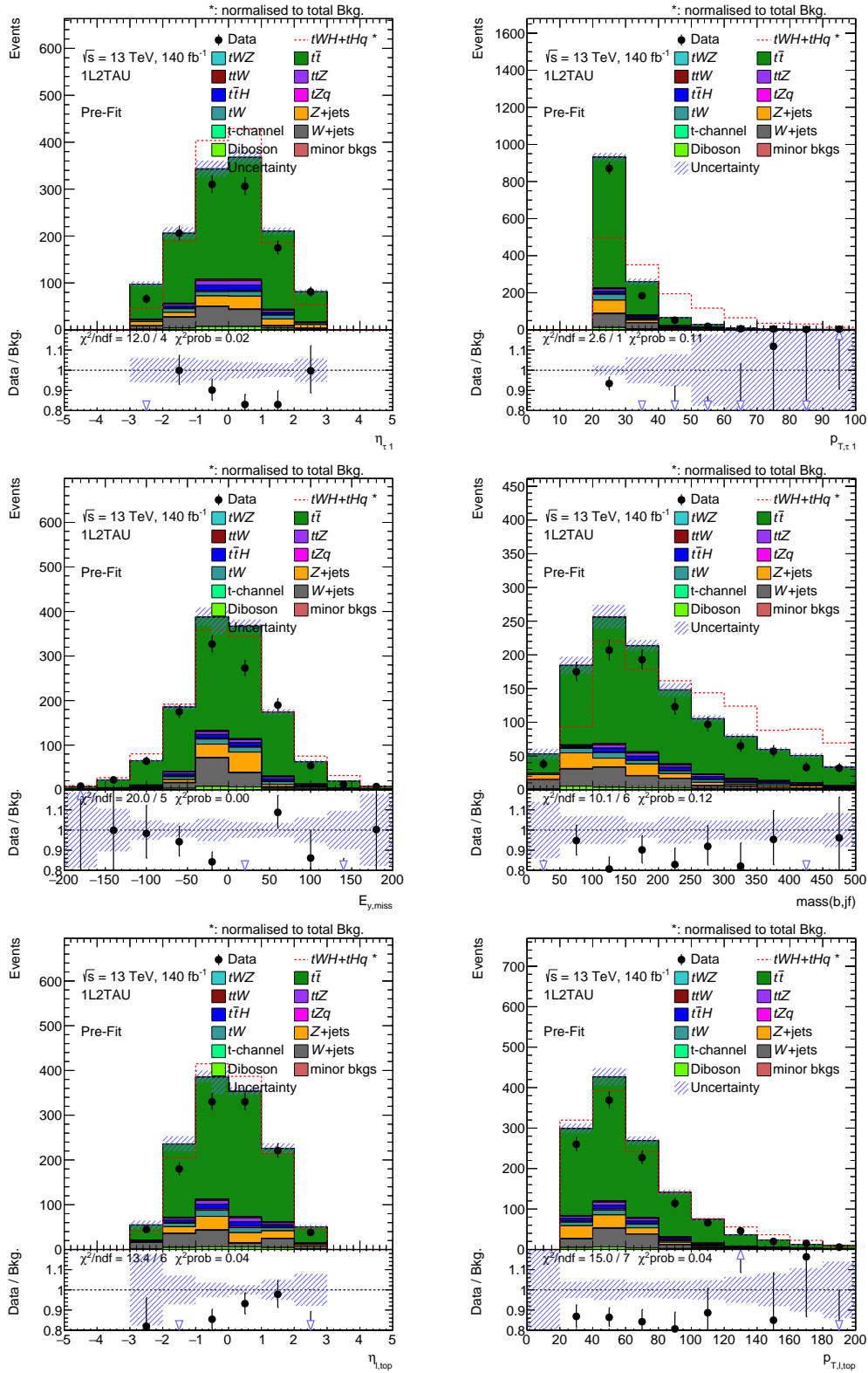


Figure G.1: Data to MC agreement for the features used in the neural network training for the  $1\ell + 2\tau_{\text{had}}$  channel. For a definition of the features see section 7.2

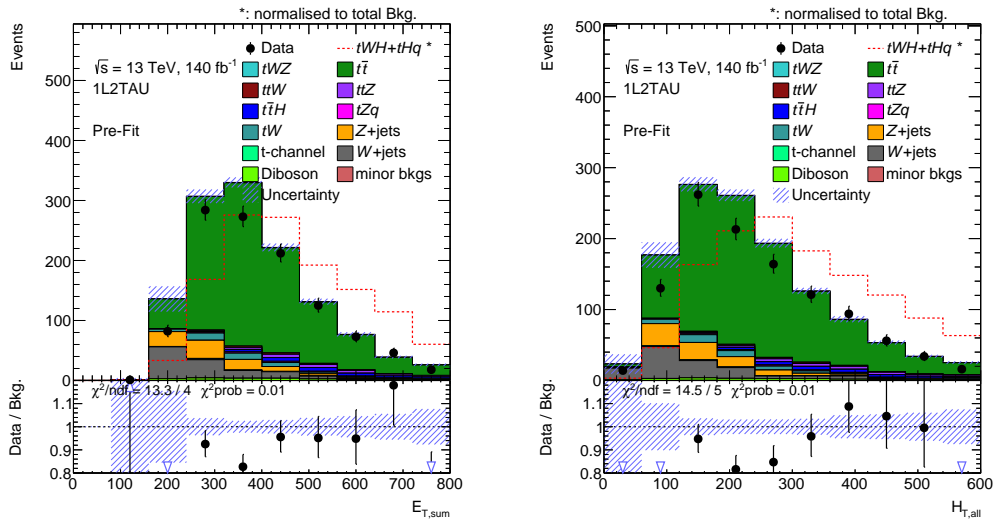


Figure G.1: Data to MC agreement for the features used in the neural network training for the  $1\ell + 2\tau_{\text{had}}$  channel. For a definition of the features see section 7.2

## Additional NN result plots

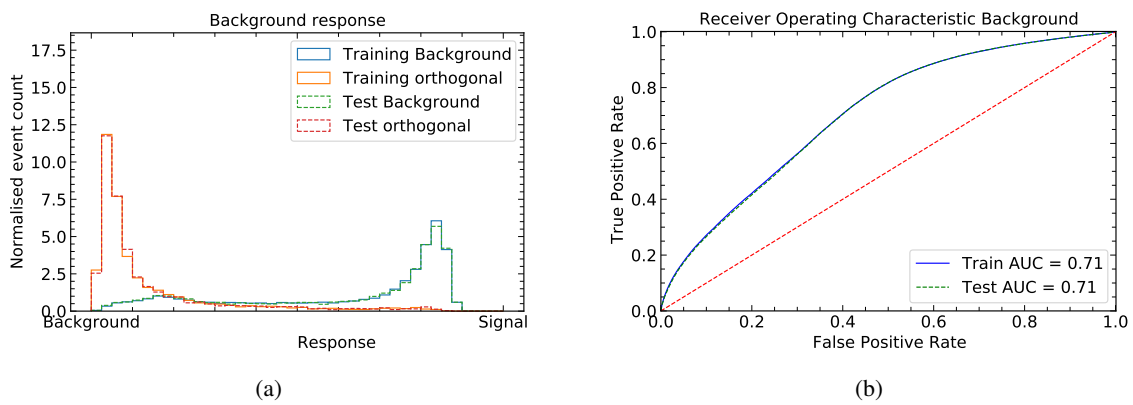


Figure H.1: Display of the background response (a) and the ROC (b) curve of the final model for the  $2\ell$  OS +  $1\tau_{\text{had}}$  channel.

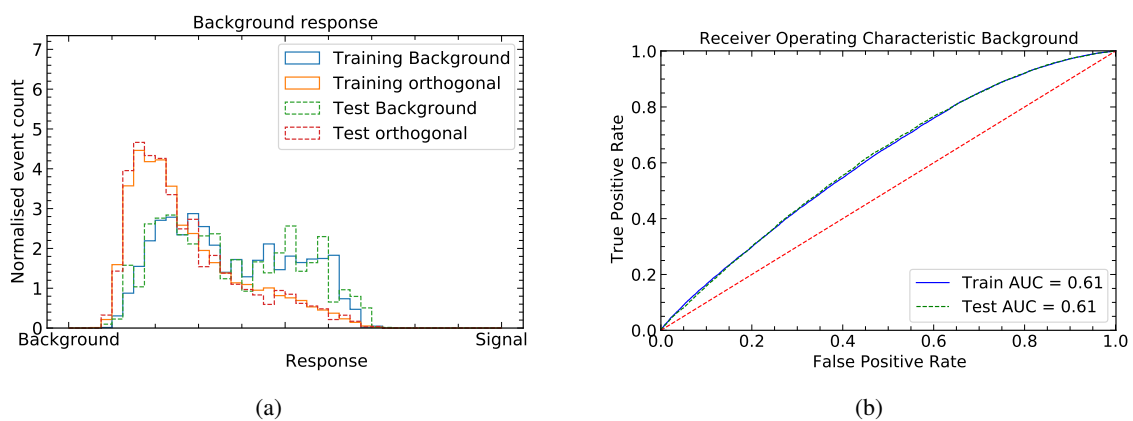


Figure H.2: Display of the background response (a) and the ROC (b) curve of the final model for the  $2\ell$  SS +  $1\tau_{\text{had}}$  channel.

## Appendix H Additional NN result plots

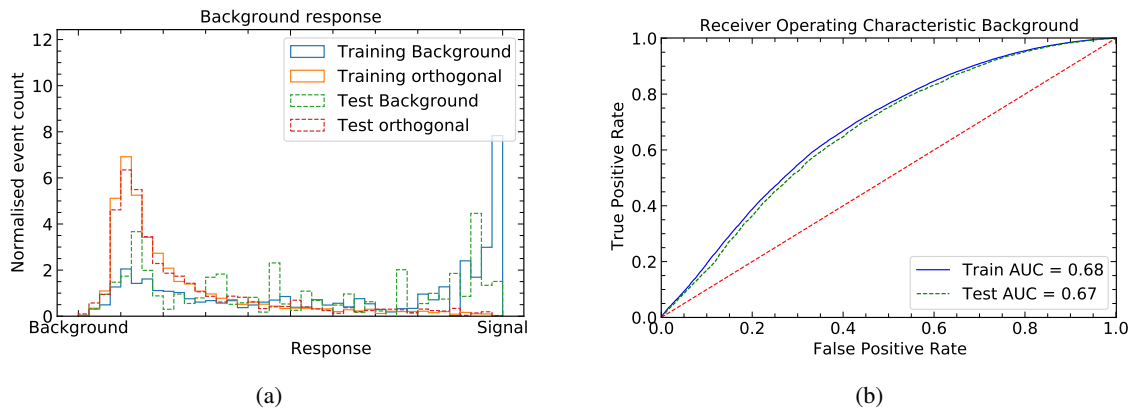


Figure H.3: Display of the background response (a) and the ROC (b) curve of the final model for the  $1\ell + 2\tau_{\text{had}}$  channel.

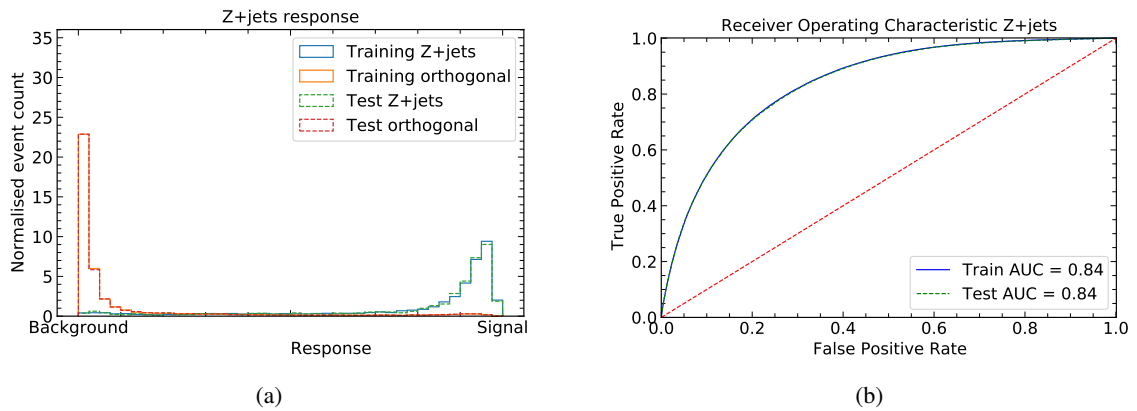


Figure H.4: Display of the Z+jets response (a) and the ROC (b) curve of the final model for the  $2\ell \text{ OS} + 1\tau_{\text{had}}$  channel.

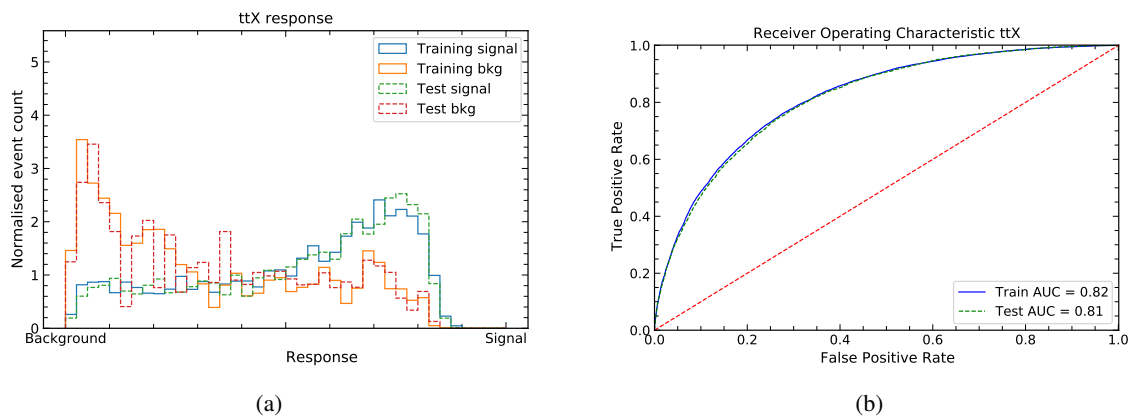


Figure H.5: Display of the  $t\bar{t} + X$  response (a) and the ROC (b) curve of the final model for the  $2\ell \text{ SS} + 1\tau_{\text{had}}$  channel.



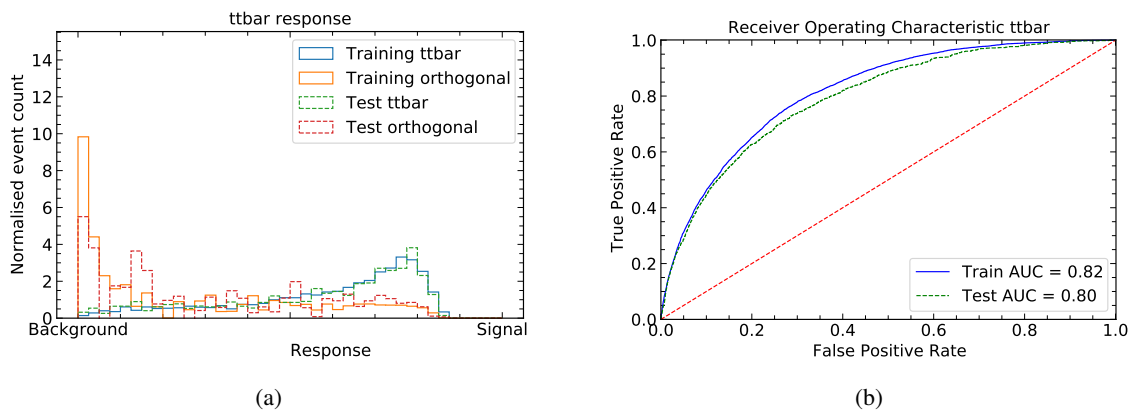
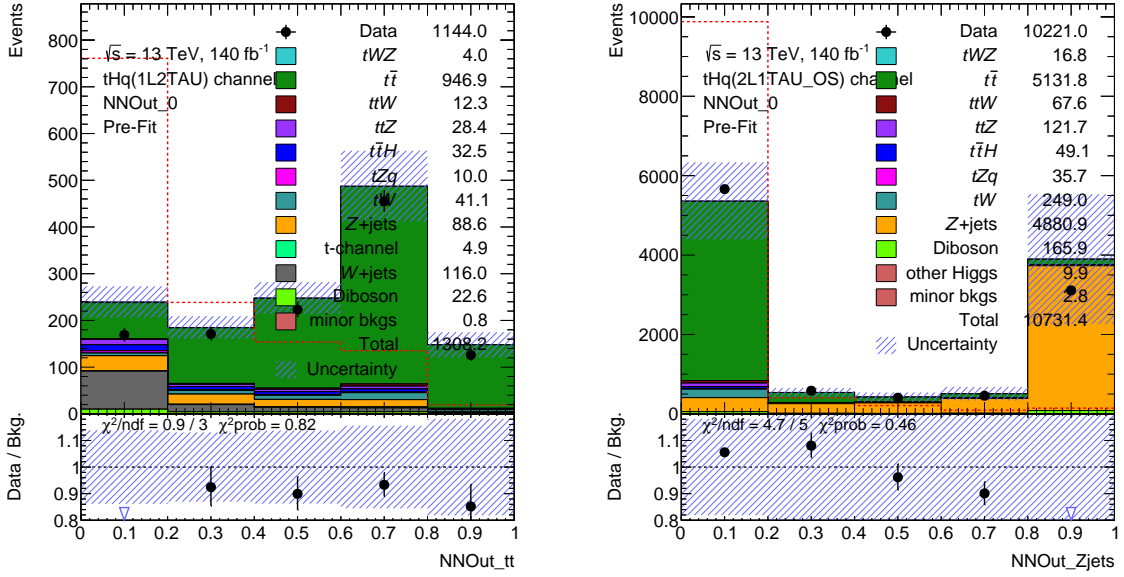
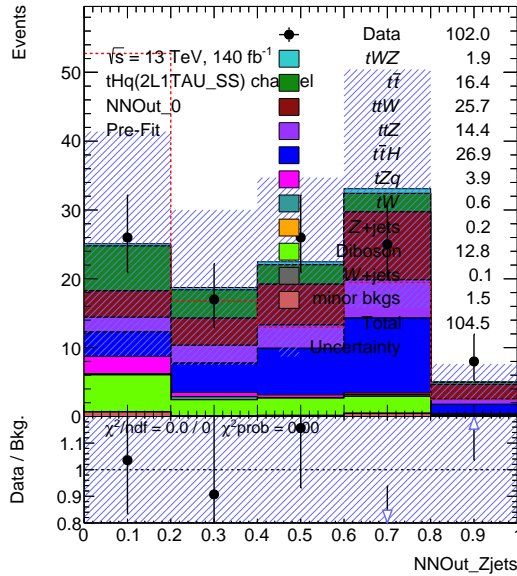


Figure H.6: Display of the  $t\bar{t}$  response (a) and the ROC (b) curve of the final model for the  $1\ell + 2\tau_{\text{had}}$  channel.



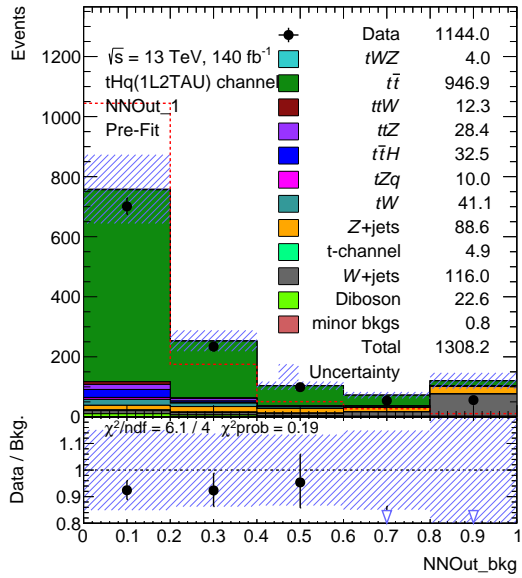
(a)  $1\ell + 2\tau_{\text{had}} \bar{t}\bar{t}$  response

(b)  $2\ell \text{ OS} + 1\tau_{\text{had}} \text{ Z+jets}$  response

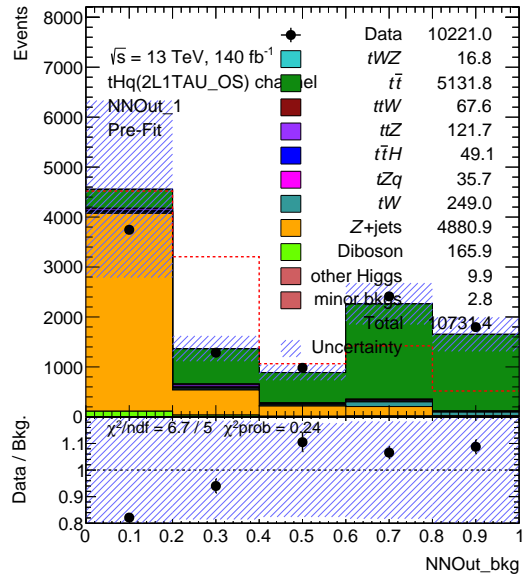


(c)  $2\ell \text{ SS} + 1\tau_{\text{had}} \bar{t}\bar{t} + \text{X}$  response

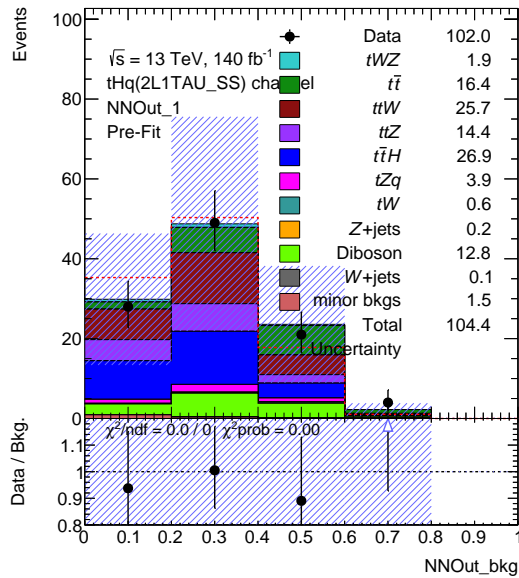
Figure H.7: Display of the process-wise response distributions of the categorical responses using full uncertainties. The signal is scaled up and displayed as a red-dashed line.



(a)  $1\ell + 2\tau_{\text{had}}$  background response



(b)  $2\ell \text{ OS} + 1\tau_{\text{had}}$  background response



(c)  $2\ell \text{ SS} + 1\tau_{\text{had}}$  background response

Figure H.8: Display of the process-wise response distributions of the background responses using full uncertainties. The signal is scaled up and displayed as a red-dashed line.



---

## Performance of Lorentz invariant variables in the $1\ell + 2\tau_{\text{had}}$ channel

---

As described in section 7.2, the usage of Lorentz-invariant variables as input features to a NN is a tempting choice due to their full and minimalistic coverage of the information in the final state. This appendix shows an optimised and relatively stable model for  $1\ell + 2\tau_{\text{had}}$  using a set of Lorentz-invariant variables in combination with the missing energy and reconstructed Higgs mass. The loss is shown in figure I.1. It shows no worrisome instabilities. The ROC curves and responses for all three categories are displayed in figures I.2 to I.4. The overall performance is clearly worse than for the usage of advanced kinematic variables presented in section 7.5. In addition, the instabilities exceed what is expected from the impact of negatively weighted events. For the  $tZq$  signal, an analysis comparable to  $tH$ , even a more in-depth optimisation of the model did not yield comparable results [98].

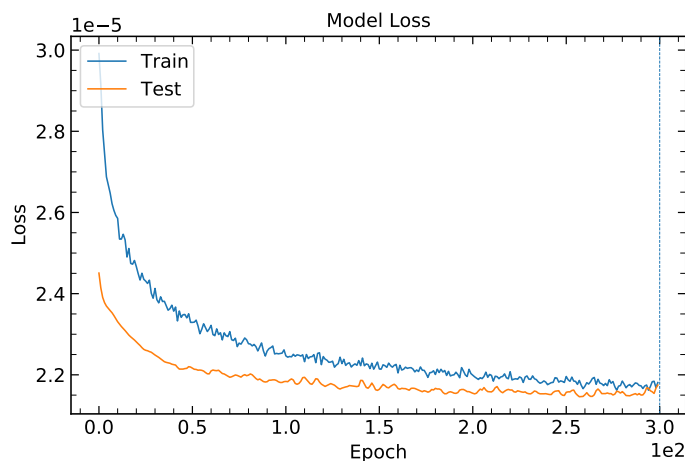


Figure I.1: Display of the training and test loss for the  $1\ell + 2\tau_{\text{had}}$  channel using Lorentz-invariant variables.

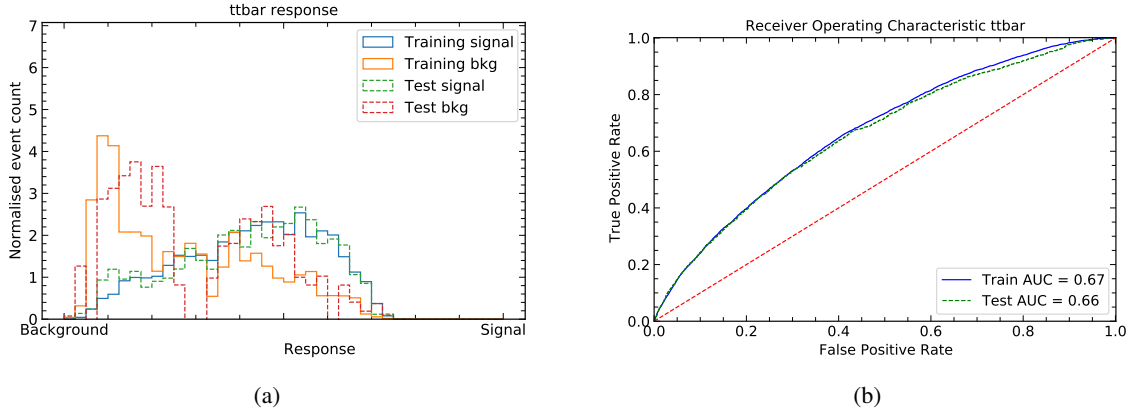


Figure I.2: Display of the categorical response (a) and the ROC (b) curve of the final model for the  $1\ell + 2\tau_{\text{had}}$  channel for Lorentz-invariant variables. Both response and ROC curve communicate a poor classification with respect to the kinematic features. In addition the fluctuations between training and test sample are too large to be acceptable.

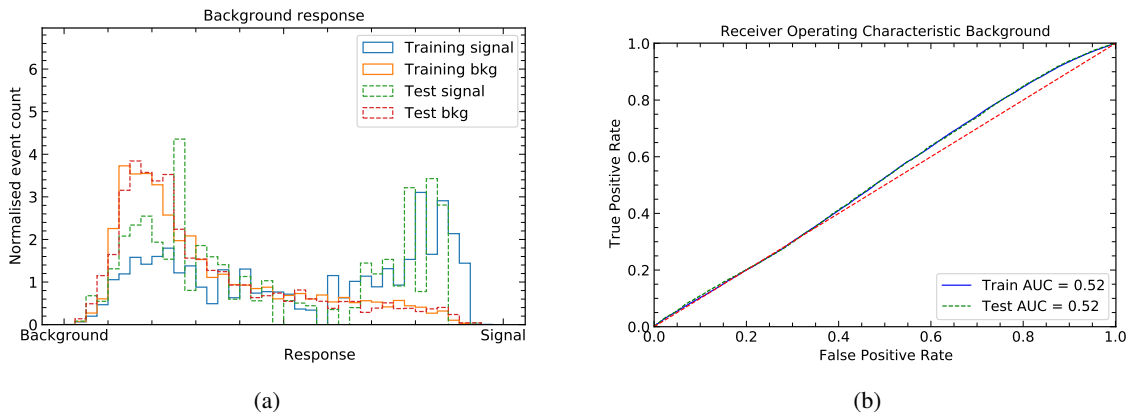


Figure I.3: Display of the background response (a) and the ROC (b) curve of the final model for the  $1\ell + 2\tau_{\text{had}}$  channel for Lorentz-invariant variables. Both response and ROC communicate a poor classification with respect to the kinematic features. In addition the fluctuations between training and test sample are too large to be acceptable.

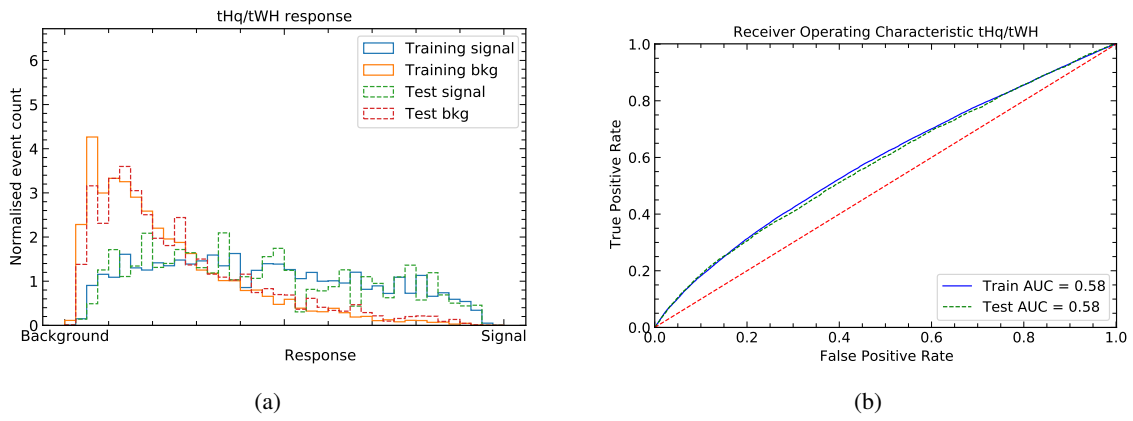


Figure I.4: Display of the signal response (a) and the ROC (b) curve of the final model for the  $1\ell + 2\tau_{\text{had}}$  channel for Lorentz-invariant variables. Both response and ROC communicate a poor classification with respect to the kinematic features. In addition the fluctuations between training and test sample are too large to be acceptable.





## K-fold cross validation

Section 7.6 covers the additional 5-fold cross validation performed for the optimised models. In the additional test both loss and ROC curve were compared and checked for similar behaviour for each channel and subset. It is reported that in each case the behaviour is very comparable and therefore little generalisation problems are to be expected from the training sample beyond the statistical uncertainties. This appendix presents the ROC curves and losses for each batch and all channels in figures J.1 to J.3. In all cases the final hyperparameters were used for decreased number of epochs to accommodate for the higher uncertainties. The ROC curve or the loss behaviour shows no problematic deviation in any of the cases. For  $2\ell$  OS +  $1\tau_{\text{had}}$  even eight folds were stable for 200 epochs each. This is due to the higher raw yields in the channel.

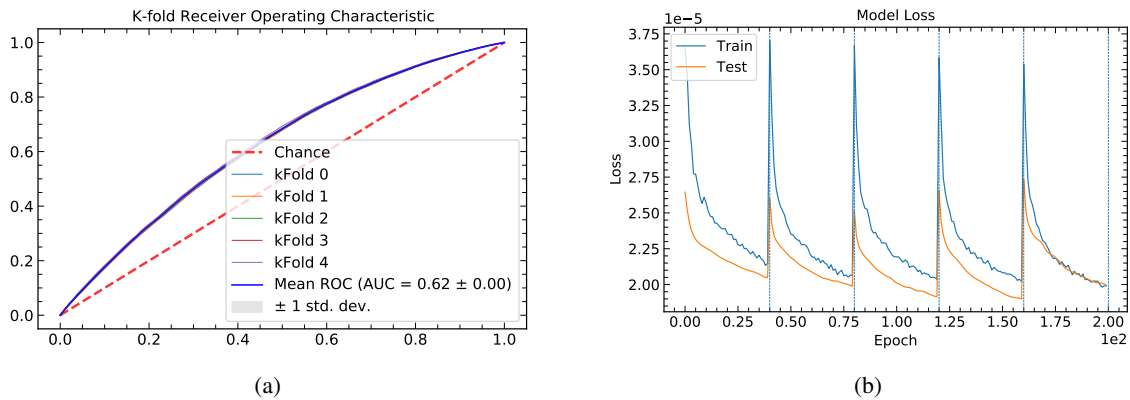


Figure J.1: ROC (a) and loss curve (b) for the 5-fold cross validation in the  $1\ell + 2\tau_{\text{had}}$  channel.

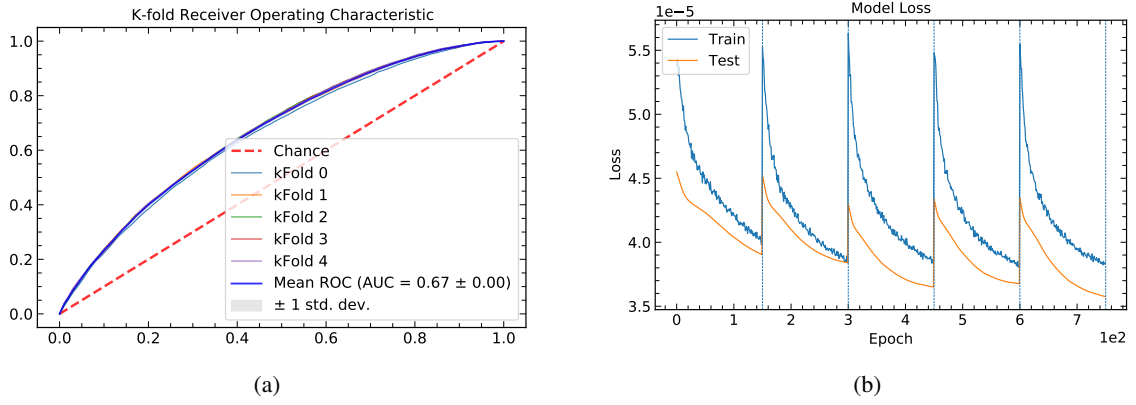


Figure J.2: ROC (a) and loss curve (b) for the 5-fold cross validation in the  $2\ell$  SS +  $1\tau_{\text{had}}$  channel.

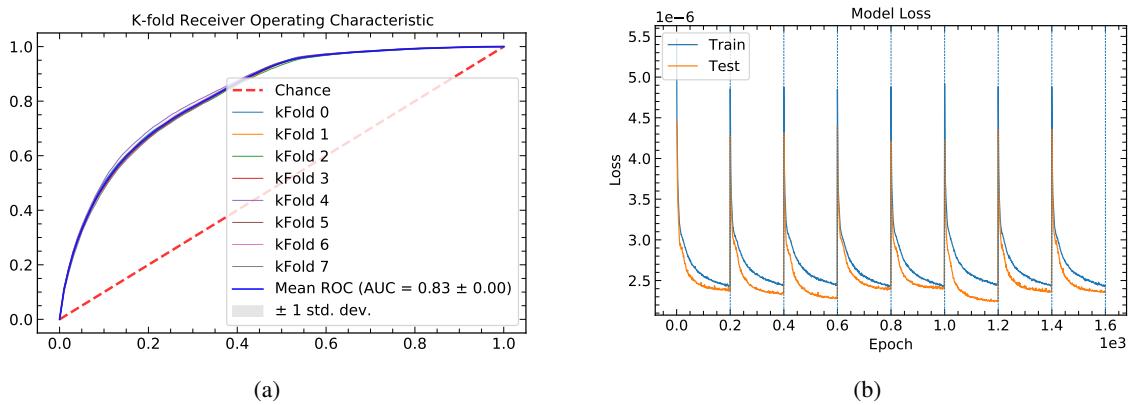


Figure J.3: ROC (a) and loss curve (b) for the 8-fold cross validation in the  $2\ell$  OS +  $1\tau_{\text{had}}$  channel.

---

## Additional plots for the Asimov fits

---

This appendix summarises a list of stability control plots for the Asimov fits. These plots are useful to gain a more in-depth understanding of the performance. The constraints are displayed in figures [K.1](#) to [K.3](#). All three collections of plots have the constraints for the  $t\bar{t}$  systematics in common. While future studies are expected to give further insights into these effects, the best explanation is the poor modelling of  $t\bar{t}$  in the preselection region. In the preselection region  $t\bar{t}$  is fake dominated which is not a situation for which the modelling is optimised. Alternative modelling samples that promise a better performance were studied but do not contain sufficient yields after the selection is applied. In general the constraints originate from the  $t\bar{t}$  dominated regions which is expected.

The important correlations between the samples are collected in figures [K.4](#) to [K.6](#). For  $1\ell + 2\tau_{\text{had}}$  no correlations are outstanding. The most noteworthy are the correlation of the background NFs for which mainly  $t\bar{t}$  and  $\tau$  related systematics are correlated. This is within reason because a certain connection between these contributions is physically implied. The situation is comparable for  $2\ell \text{ OS} + 1\tau_{\text{had}}$  with the exception of the positive correlation between the NFs of  $t\bar{t}$  and  $Z+$  jets. The correlation shows the shape component of the jets originating from either quarks or gluons. The yields in the region are large enough that even a NF can reflect the difference in fake origin for the two contributions in the fit. In  $2\ell \text{ SS} + 1\tau_{\text{had}}$  a negative correlation between the NF for  $t\bar{t}$  and  $t\bar{t}+X$  with respect to a  $t\bar{t}W$  systematic is recorded. Studies have shown that the  $t\bar{t}W$  modelling in this region is imperfect. If more events were available the  $t\bar{t}W$  sample should be treated independently from the other  $t\bar{t}+X$  samples. The strong negative correlation between the two NFs is likely an effect of the poor separation of processes in the bins and regions.

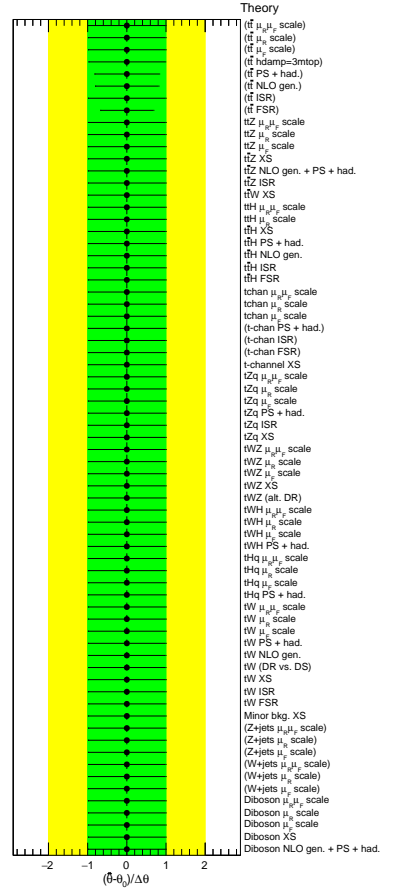
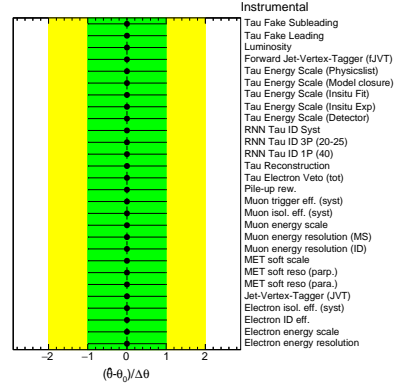
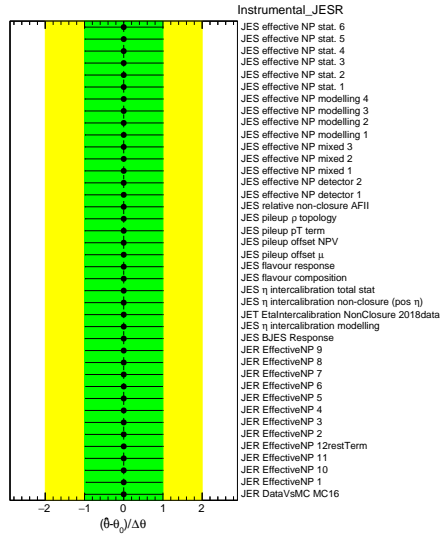
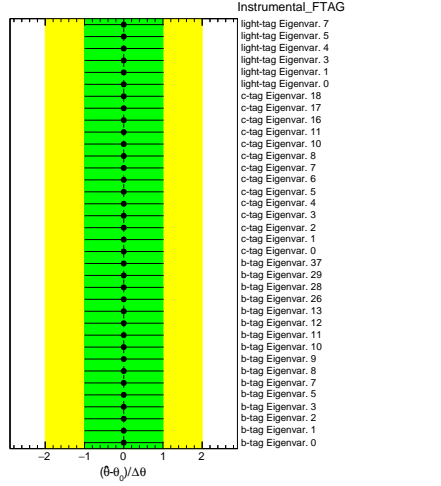
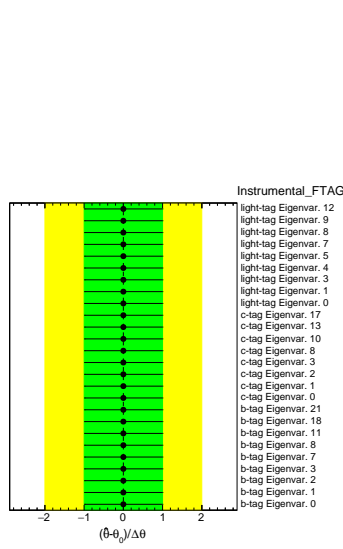
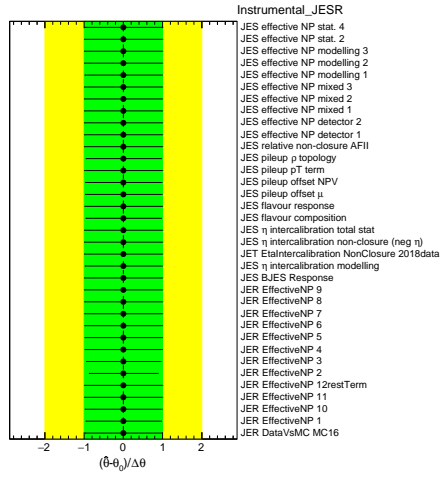


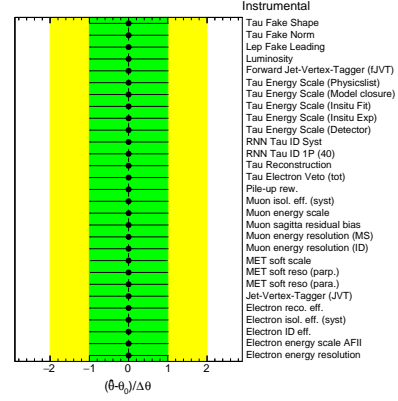
Figure K.1: Constraints on the NPs in the  $1\ell + 2\tau_{\text{had}}$  Asimov fit.



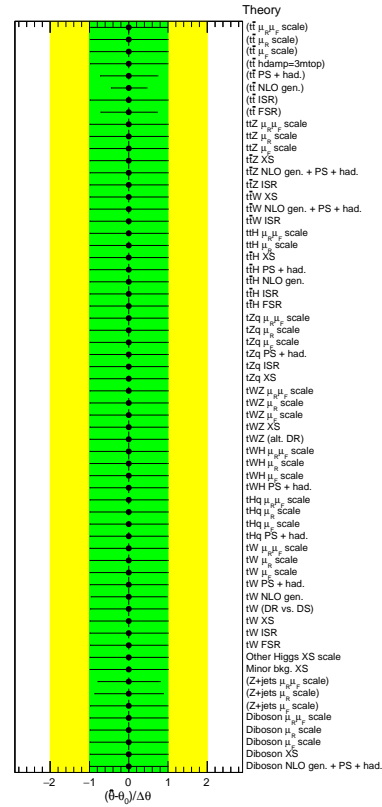
(a) Constraints of instrumental FTAG NPs.



(b) Constraints of instrumental JSER NPs.



(c) Constraints of instrumental NPs.



(d) Constraints of theory NPs.

Figure K.2: Constraints on the NPs in the  $2\ell$  OS +  $1\tau_{\text{had}}$  Asimov fit.

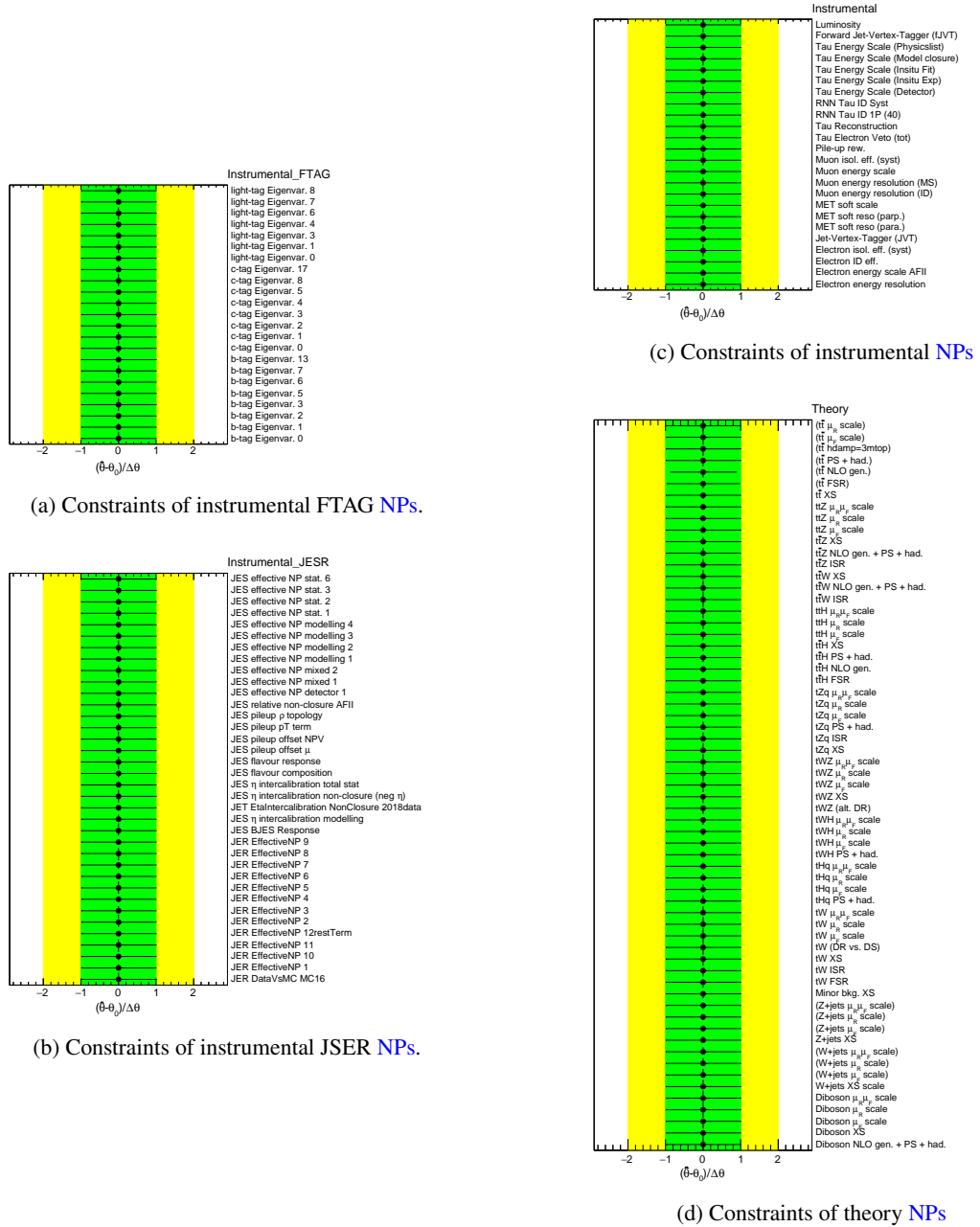


Figure K.3: Constraints on the NPs in the  $2\ell$  SS +  $1\tau_{\text{had}}$  Asimov fit.

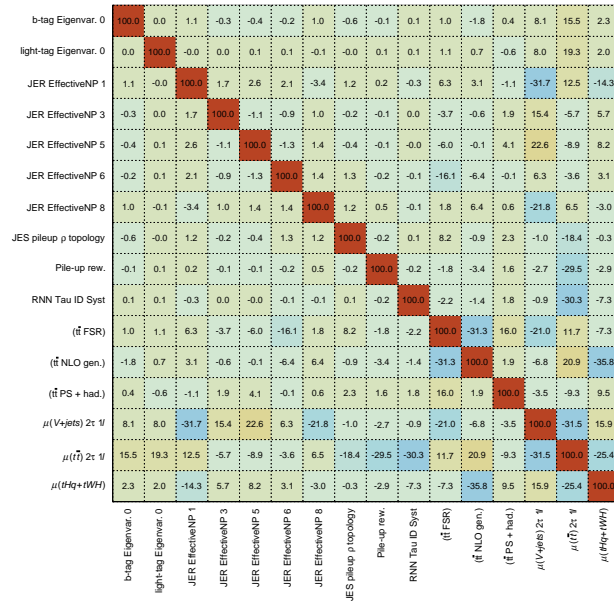


Figure K.4: Correlation matrix of the  $1\ell + 2\tau_{\text{had}}$  Asimov fit.

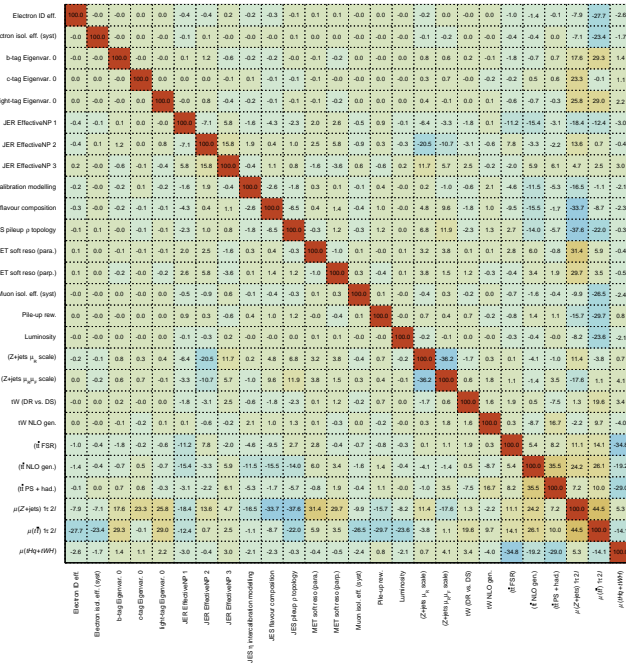


Figure K.5: Correlation matrix of the  $2\ell \text{ SS} + 1\tau_{\text{had}}$  Asimov fit.

JER EffectiveNP 3	100.0	-0.5	-0.4	-0.0	-0.5	-1.0	15.8	-3.2
JER EffectiveNP 4	-0.5	100.0	-0.8	0.1	-2.0	-4.0	18.4	-6.4
JER EffectiveNP 6	-0.4	-0.8	100.0	0.4	-1.8	-3.7	16.1	-7.1
$t\bar{t}W$ XS	-0.0	0.1	0.4	100.0	0.4	0.8	-26.5	2.3
$(t\bar{t}$ FSR)	-0.5	-2.0	-1.8	0.4	100.0	-15.1	7.8	-12.1
$(t\bar{t}$ NLO gen.)	-1.0	-4.0	-3.7	0.8	-15.1	100.0	14.1	-23.8
$\mu(t\bar{t}+t\bar{t}X)$ $1\tau$ $2l$	15.8	18.4	16.1	-26.5	7.8	14.1	100.0	-57.3
$\mu(tHq+tWH)$	-3.2	-6.4	-7.1	2.3	-12.1	-23.8	-57.3	100.0
JER EffectiveNP 3								
JER EffectiveNP 4								
JER EffectiveNP 6								
$t\bar{t}W$ XS								
$(t\bar{t}$ FSR)								
$(t\bar{t}$ NLO gen.)								
$\mu(t\bar{t}+t\bar{t}X)$ $1\tau$ $2l$								
$\mu(tHq+tWH)$								

Figure K.6: Correlation matrix of the  $2l$  SS +  $1\tau_{\text{had}}$  Asimov fit.



---

## Additional plots for the data fits

---

This appendix summarises a list of stability control plots for the data fits. These plots are useful to gain a more in-depth understanding of the performance. The constraints are displayed in figures L.1 to L.3. All three collections of plots have the pull effects for the  $t\bar{t}$  systematics in common. While future studies are expected to give further insights into these effects, the best explanation is the poor modelling of  $t\bar{t}$  systematics in the given fake object enriched region. In the preselection region  $t\bar{t}$  is fake dominated which is not a situation for which the modelling is optimised. Alternative modelling samples that promise a better performance were studied but do not contain sufficient yields after the selection is applied. In general the constraints originate from the  $t\bar{t}$  dominated regions which is expected. Additional pulls appear for jet related variables. The by far largest number of pulls is recorded in the  $2\ell$  OS +  $1\tau_{\text{had}}$  channel. These effects can be attributed to the higher yields in the selected region. No NP is pulled out of its expected band.

The important correlations between the samples are collected in figures L.4 to L.6. For  $1\ell + 2\tau_{\text{had}}$  no correlations are outstanding. The most noteworthy are the correlation of the background NFs for which mainly  $t\bar{t}$  and  $\tau$  related systematics are correlated. This is within reason because a certain connection between these contributions is physically implied. The situation is comparable for  $2\ell$  OS +  $1\tau_{\text{had}}$  with the exception of the positive correlation between the NFs of  $t\bar{t}$  and  $Z$ +jets. The correlation shows the shape component of the jets originating from either quarks or gluons. The yields in the region are large enough that even a NF can reflect the difference in fake origin for the two contributions in the fit. In  $2\ell$  SS +  $1\tau_{\text{had}}$  a negative correlation between the NF for  $t\bar{t}$  and  $t\bar{t}+X$  with respect to a  $t\bar{t}W$  systematic is recorded. Studies have shown that the  $t\bar{t}W$  modelling in this region is imperfect. If more events were available the  $t\bar{t}W$  sample should be treated independently from the other  $t\bar{t}+X$  samples. The strong negative correlation between the two NFs is likely an effect of the poor separation of processes in the bins and regions.

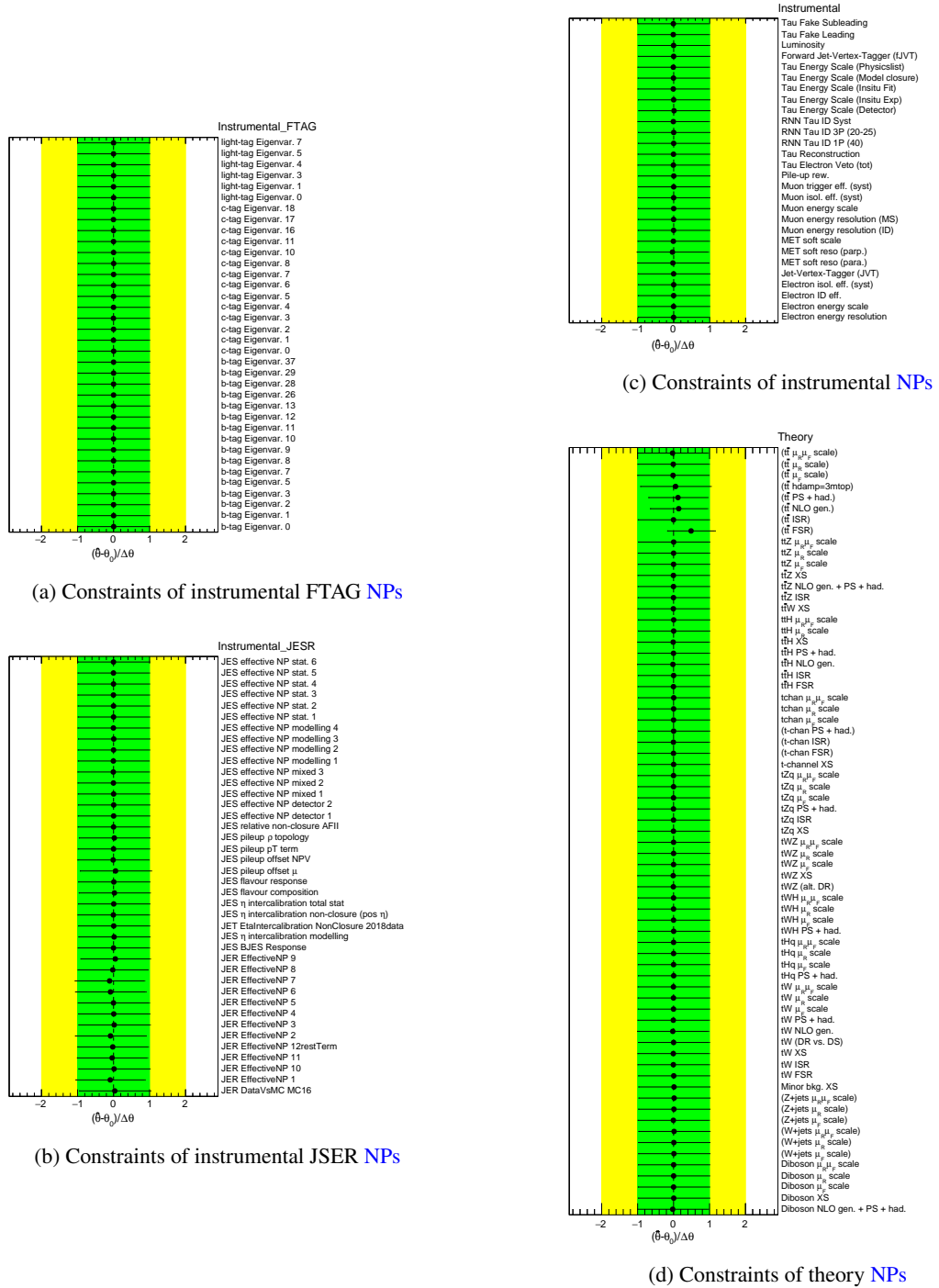


Figure L.1: Constraints on the NPs in the  $1\ell + 2\tau_{\text{had}}$  data fit.

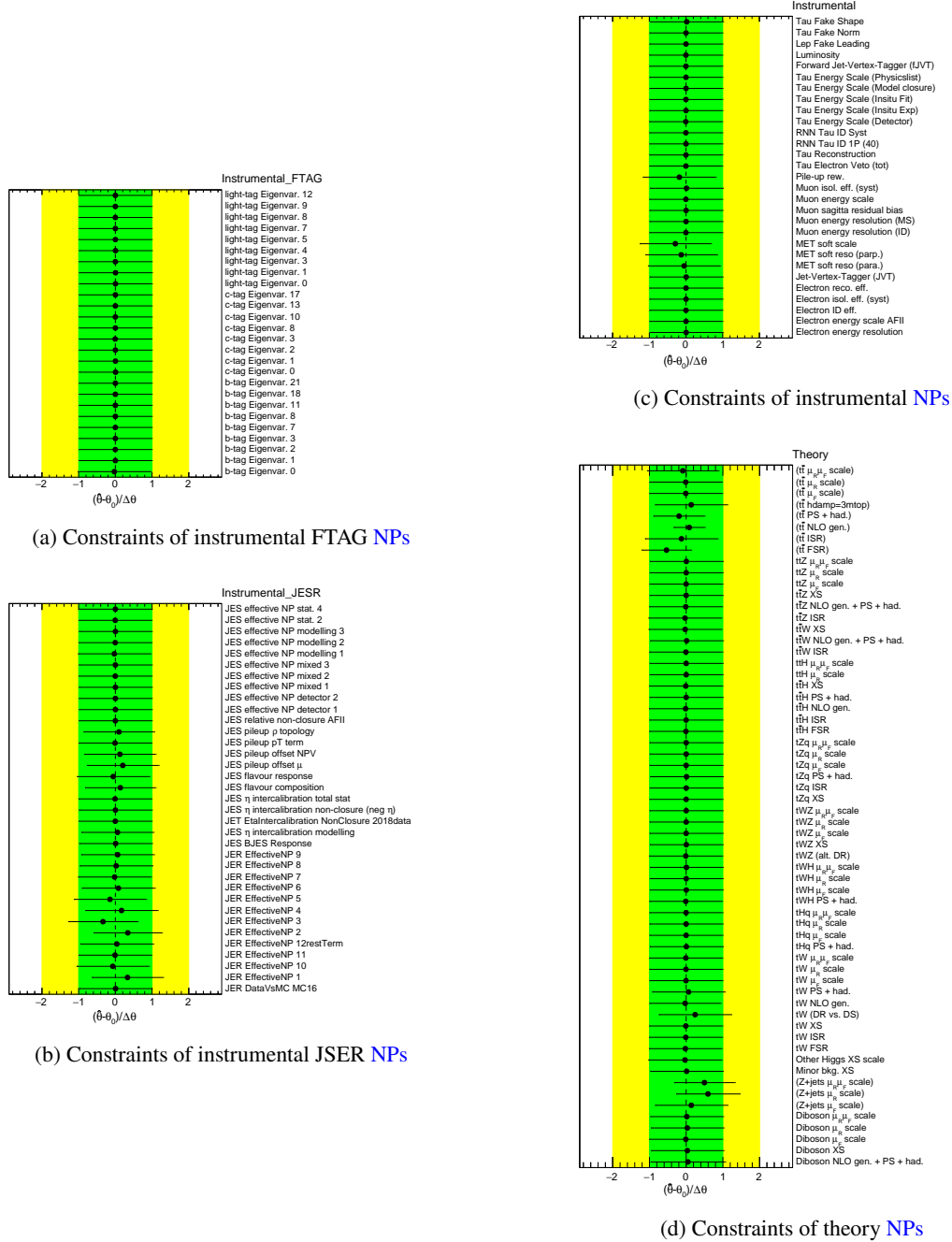
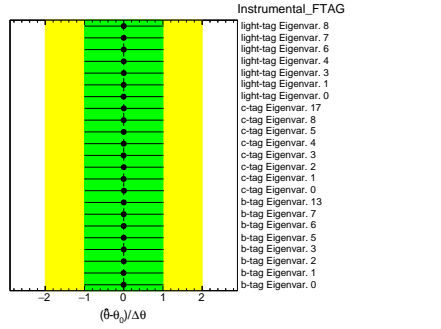
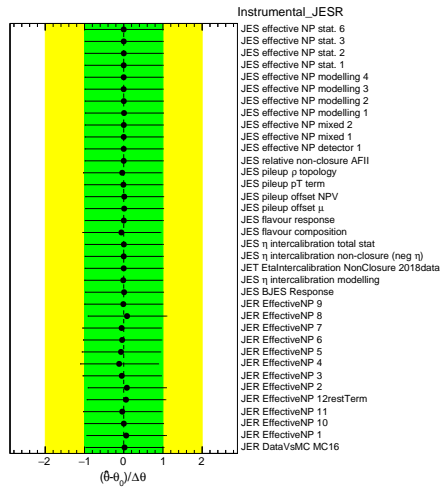


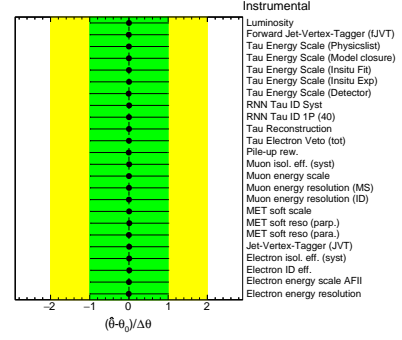
Figure L.2: Constraints on the NPs in the  $2\ell$  OS +  $1\tau_{\text{had}}$  data fit.



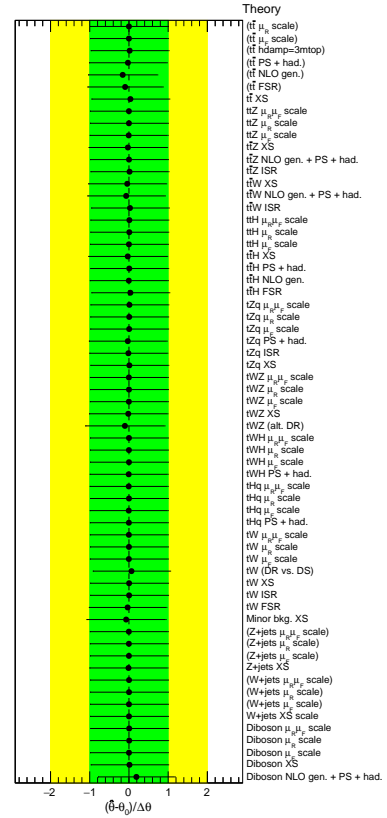
(a) Constraints of instrumental FTAG NPs



(b) Constraints of instrumental JSER NPs



(c) Constraints of instrumental NPs



(d) Constraints of theory NPs

Figure L.3: Constraints on the NPs in the  $2\ell$  SS +  $1\tau_{\text{had}}$  data fit.



JER EffectiveNP 4	100.0	0.1	-3.3	15.4	-3.4
$t\bar{t}W$ XS	0.1	100.0	1.0	-23.9	1.9
( $t\bar{t}$ NLO gen.)	-3.3	1.0	100.0	10.4	-18.8
$\mu(t\bar{t}+t\bar{t}X)$ $1\tau$ $2l$	15.4	-23.9	10.4	100.0	-53.1
$\mu(tHq+tWH)$	-3.4	1.9	-18.8	-53.1	100.0
	JER EffectiveNP 4	$t\bar{t}W$ XS	( $t\bar{t}$ NLO gen.)	$\mu(t\bar{t}+t\bar{t}X)$ $1\tau$ $2l$	$\mu(tHq+tWH)$

Figure L.6: Correlation matrix of the  $2\ell$  SS +  $1\tau_{\text{had}}$  data fit.

---

## Bibliography

---

- [1] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012) 1, arXiv: [1207.7214](https://arxiv.org/abs/1207.7214) [[hep-ex](#)] (cit. on p. 1).
- [2] *Netzwerk Teilchenwelt*, URL: <https://www.teilchenwelt.de/> (visited on 28/12/2023) (cit. on p. 1).
- [3] Particle Data Group, *Review of Particle Physics*, *PTEP* **2022** (2022) 083C01 (cit. on pp. [2](#), [5](#), [8](#), [9](#), [14](#), [47](#), [49](#)).
- [4] M. E. Peskin and D. V. Schroeder, *An introduction to quantum field theory*, 1st ed., Westview Press, 2007 (cit. on pp. [3](#), [6](#), [122](#)).
- [5] M. Thomson, *Modern Particle Physics*, 2nd ed., Wiley-VCH, 2008 (cit. on pp. [4](#), [11](#), [123](#), [124](#)).
- [6] Wikipedia contributors, *Elementary particles included in the Standard Model*, [Online; accessed 29-August-2023], 2006, URL: [https://commons.wikimedia.org/wiki/File:Standard\\_Model\\_of\\_Elementary\\_Particles.svg](https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg) (cit. on p. [5](#)).
- [7] R. Shankar, *Principles of Quantum Mechanics*, 2nd ed., Springer, 1995 (cit. on pp. [4](#), [123](#)).
- [8] D. Griffith, *Introduction to Elementary Particles*, Cambridge University Press, 2013 (cit. on p. [6](#)).
- [9] M. Gell-Mann, *A Schematic Model of Baryons and Mesons*, *Phys. Lett.* **8** (1964) (cit. on p. [7](#)).
- [10] K. Nishijima, *BRS invariance, asymptotic freedom and color confinement (a review)*, *Czech J Phys* **46** (1996) (cit. on p. [7](#)).
- [11] B. Povh, *Teilchen und Kerne eine Einführung in die physikalischen Konzepte*, 9th ed., Springer, 2014 (cit. on pp. [8](#), [12](#)).
- [12] A. Nisati and V. Sharma, *Discovery of the Higgs Boson*, World Scientific, 2017 (cit. on pp. [11](#), [15](#), [16](#), [21](#), [47](#), [123](#)).
- [13] The Super-Kamiokande Collaboration, “The Super-Kamiokande Experiment”, *Neutrino Oscillations*, WORLD SCIENTIFIC, 2008, URL: [https://doi.org/10.1142/2F9789812771971\\_0002](https://doi.org/10.1142/2F9789812771971_0002) (cit. on p. [12](#)).
- [14] M. Chanowitz, *Strong WW scattering at the end of the 90’s: theory and experimental prospects*, 1998, arXiv: [hep-ph/9812215](https://arxiv.org/abs/hep-ph/9812215) [[hep-ph](#)] (cit. on p. [12](#)).

- [15] I. C. Brock and T. Schorner-Sadenius, *Physics at the Terascale*, Weinheim: Wiley, 2011, URL: <https://cds.cern.ch/record/1354959> (cit. on pp. 12, 23).
- [16] D. Madaffari, *Higgs boson production in association with a top quark pair at the LHC*, tech. rep., CERN, 2018, URL: <https://cds.cern.ch/record/2626861> (cit. on p. 17).
- [17] ATLAS Collaboration, *Evidence for the associated production of the Higgs boson and a top quark pair with the ATLAS detector*, *Phys. Rev. D* **97** (2018) 072003, arXiv: 1712.08891 [hep-ex] (cit. on p. 17).
- [18] ATLAS Collaboration, *Measurements of Higgs boson properties in the diphoton decay channel with  $36\text{fb}^{-1}$  of  $pp$  collision data at  $\sqrt{s} = 13\text{TeV}$  with the ATLAS detector*, *Phys. Rev. D* **98** (2018) 052005, arXiv: 1802.04146 [hep-ex] (cit. on p. 17).
- [19] A. Celis, V. Ilisie and A. Pich, *LHC constraints on two-Higgs doublet models*, *Journal of High Energy Physics* **2013** (2013), URL: <https://doi.org/10.1007%2Fjhep07%282013%29053> (cit. on p. 18).
- [20] D. Chowdhury and O. Eberhardt, *Update of global Two-Higgs-Doublet model fits*, *Journal of High Energy Physics* **2018** (2018), URL: <https://doi.org/10.1007%2Fjhep05%282018%29161> (cit. on p. 18).
- [21] N. Wermes and H. Kolanoski, *Teilchendetektoren Grundlagen und Anwendungen*, 1st ed., Springer Verlag, 2016, ISBN: 9783662453490 (cit. on pp. 23, 29–31).
- [22] L. Evans and P. Bryant, *LHC Machine*, *JINST* **3** (2008) S08001 (cit. on p. 25).
- [23] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08003 (cit. on p. 25).
- [24] CMS Collaboration, *The CMS Experiment at the CERN LHC*, *JINST* **3** (2008) S08004 (cit. on p. 25).
- [25] A. A. Alves Jr. et al., *The LHCb Detector at the LHC*, *JINST* **3** (2008) S08005 (cit. on p. 25).
- [26] K. Aamodt et al., *The ALICE experiment at the CERN LHC*, *JINST* **3** (2008) S08002 (cit. on p. 25).
- [27] J.-L. Caron, “Overall view of LHC experiments.. Vue d’ensemble des experiences du LHC.”, AC Collection. Legacy of AC. Pictures from 1992 to 2002., 1998, URL: <https://cds.cern.ch/record/841555> (cit. on p. 26).
- [28] E. Mobs, *The CERN accelerator complex - August 2018. Complexe des accélérateurs du CERN - Août 2018*, (2018), General Photo, URL: <https://cds.cern.ch/record/2636343> (cit. on p. 27).
- [29] H. Geiger and W. Müller, *Elektronenzählrohr zur Messung schwächster Aktivitäten*, *Die Naturwissenschaften* (1928) (cit. on p. 28).
- [30] B. Rossi, *High Energy Particles*, Prentice Hall, 1952 (cit. on p. 29).
- [31] J. Pequeno, “Computer generated image of the whole ATLAS detector”, 2008, URL: <https://cds.cern.ch/record/1095924> (cit. on p. 32).
- [32] Neutelings, Izaak, *CMS coordinate system*, URL: [https://tikz.net/axis3d\\_cms/](https://tikz.net/axis3d_cms/) (cit. on p. 33).



- 
- [33] ATLAS Collaboration, *ATLAS Inner Detector: Technical Design Report, Volume 1*, ATLAS-TDR-4; CERN-LHCC-97-016, 1997, URL: <https://cds.cern.ch/record/331063> (cit. on p. 34).
- [34] ATLAS Collaboration, *ATLAS Inner Detector: Technical Design Report, Volume 2*, ATLAS-TDR-5; CERN-LHCC-97-017, 1997, URL: <https://cds.cern.ch/record/331064> (cit. on p. 34).
- [35] ATLAS Collaboration, *ATLAS Insertable B-Layer: Technical Design Report*, ATLAS-TDR-19; CERN-LHCC-2010-013, 2010, URL: <https://cds.cern.ch/record/1291633> (cit. on p. 34), Addendum: ATLAS-TDR-19-ADD-1; CERN-LHCC-2012-009, 2012, URL: <https://cds.cern.ch/record/1451888>.
- [36] A. L. Rosa, *The ATLAS Insertable B-Layer: from construction to operation*, 2016, arXiv: [1610.01994](https://arxiv.org/abs/1610.01994) [[physics.ins-det](https://arxiv.org/abs/1610.01994)] (cit. on p. 34).
- [37] ATLAS Collaboration, “Experiment Briefing: Keeping the ATLAS Inner Detector in perfect alignment”, General Photo, 2020, URL: <https://cds.cern.ch/record/2723878> (cit. on p. 35).
- [38] ATLAS Collaboration, *ATLAS Calorimeter Performance: Technical Design Report*, ATLAS-TDR-1; CERN-LHCC-96-040, 1996, URL: <https://cds.cern.ch/record/331059> (cit. on p. 34).
- [39] ATLAS Collaboration, *Measurement of the energy response of the ATLAS calorimeter to charged pions from  $W^\pm \rightarrow \tau^\pm (\rightarrow \pi^\pm \nu_\tau) \nu_\tau$  events in Run 2 data*, *Eur. Phys. J. C* **82** (2021) 223, arXiv: [2108.09043](https://arxiv.org/abs/2108.09043) [[hep-ex](https://arxiv.org/abs/2108.09043)] (cit. on p. 34).
- [40] C. Solans Sanchez, J. Ferrer and E. Higon-Rodriguez, *Implementation of the ROD Crate DAQ Software for the ATLAS Tile Calorimeter and a Search for a MSSM Higgs Boson decaying into Tau pairs*, (2010) (cit. on p. 36).
- [41] ATLAS Collaboration, *ATLAS Muon Spectrometer: Technical Design Report*, ATLAS-TDR-10; CERN-LHCC-97-022, CERN, 1997, URL: <https://cds.cern.ch/record/331068> (cit. on p. 35).
- [42] ATLAS Collaboration, *ATLAS Magnet System: Magnet Project Technical Design Report, Volume 1*, ATLAS-TDR-6; CERN-LHCC-97-018, 1997, URL: <https://cds.cern.ch/record/338080> (cit. on p. 35).
- [43] ATLAS Collaboration, *ATLAS Barrel Toroid: Magnet Project Technical Design Report, Volume 2*, ATLAS-TDR-7; CERN-LHCC-97-019, 1997, URL: <https://cds.cern.ch/record/331065> (cit. on p. 35).
- [44] ATLAS Collaboration, *ATLAS End-Cap Toroids: Magnet Project Technical Design Report, Volume 3*, ATLAS-TDR-8; CERN-LHCC-97-020, CERN, 1997, URL: <https://cds.cern.ch/record/331066> (cit. on p. 35).

- [45] ATLAS Collaboration, *ATLAS Central Solenoid: Magnet Project Technical Design Report, Volume 4*, ATLAS-TDR-9; CERN-LHCC-97-021, CERN, 1997, URL: <https://cds.cern.ch/record/331067> (cit. on p. 35).
- [46] ATLAS Collaboration, *Operation of the ATLAS trigger system in Run 2*, *Journal of Instrumentation* **15** () P10004, URL: <https://dx.doi.org/10.1088/1748-0221/15/10/P10004> (cit. on p. 36).
- [47] ATLAS Collaboration, *ATLAS Level-1 Trigger: Technical Design Report*, ATLAS-TDR-12; CERN-LHCC-98-014, 1998, URL: <https://cds.cern.ch/record/381429> (cit. on p. 36).
- [48] ATLAS Collaboration, *ATLAS High-Level Trigger, Data Acquisition and Controls: Technical Design Report*, ATLAS-TDR-16; CERN-LHCC-2003-022, 2003, URL: <https://cds.cern.ch/record/616089> (cit. on p. 36).
- [49] S. Mehlhase, *ATLAS detector slice (and particle visualisations)*, (2021), URL: <https://cds.cern.ch/record/2770815> (cit. on p. 37).
- [50] ATLAS Collaboration, *Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC run 2. Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC run 2*, *Eur. Phys. J. C* **77** (2017) 673, URL: <https://cds.cern.ch/record/2261156> (cit. on p. 37).
- [51] F. Meloni, *Primary vertex reconstruction with the ATLAS detector*, tech. rep. 12, CERN, 2016, URL: <https://cds.cern.ch/record/2222390> (cit. on p. 37).
- [52] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **262** (1987) 444, ISSN: 0168-9002, URL: <https://www.sciencedirect.com/science/article/pii/0168900287908874> (cit. on p. 37).
- [53] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1.*, *Eur. Phys. J. C* **77** (2017) 490, arXiv: 1603.02934, URL: <https://cds.cern.ch/record/2138166> (cit. on p. 38).
- [54] W. Lampl et al., *Calorimeter Clustering Algorithms: Description and Performance*, tech. rep., CERN, 2008, URL: <https://cds.cern.ch/record/1099735> (cit. on p. 39).
- [55] ATLAS Collaboration, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at  $\sqrt{s} = 13$  TeV*, *Eur. Phys. J., C* **81** (2021) 578, arXiv: 2012.00578, URL: <http://cds.cern.ch/record/2746302> (cit. on p. 39).
- [56] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton proton collisions at  $\sqrt{s} = 13$  TeV*, *The European Physical Journal C* **78** (), URL: <https://doi.org/10.1140/EPJc/2Fs10052-018-6288-9> (cit. on p. 40).

- 
- [57] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, *Eur. Phys. J. C* **77** (2017) 466, arXiv: 1703.10485 [hep-ex] (cit. on p. 41).
- [58] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, *The European Physical Journal C* **77** (), URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-017-5031-2> (cit. on p. 42).
- [59] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at  $\sqrt{s} = 13$  TeV*, *Eur. Phys. J. C* **79** (2019) 639, arXiv: 1902.04655, URL: <https://cds.cern.ch/record/2657964> (cit. on p. 42).
- [60] M. Cacciari, G. P. Salam and G. Soyez, *The anti-kt jet clustering algorithm*, *Journal of High Energy Physics* **2008** () 063, URL: <https://doi.org/10.1088%2F1126-6708%2F2008%2F04%2F063> (cit. on p. 43).
- [61] M. Cacciari, G. P. Salam and G. Soyez, *FastJet user manual*, *The European Physical Journal C* (), URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-012-1896-2> (cit. on p. 43).
- [62] *Tagging and suppression of pileup jets with the ATLAS detector*, tech. rep., CERN, 2014, URL: <https://cds.cern.ch/record/1700870> (cit. on p. 43).
- [63] ATLAS Collaboration, *Jet Vertex Tagger for Run 2 in reco and analysis*, 2021, URL: <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/JetVertexTaggerTool> (cit. on p. 43).
- [64] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at  $\sqrt{s} = 8$  TeV using the ATLAS detector*, *The European Physical Journal C* (), URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-016-4395-z> (cit. on p. 43).
- [65] ATLAS Collaboration, *Jet energy scale and resolution measured in proton-proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, *Eur. Phys. J. C* (2021), arXiv: 2007.02645, URL: <https://cds.cern.ch/record/2722869> (cit. on pp. 43, 44).
- [66] ATLAS Collaboration, *Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at  $\sqrt{s} = 13$  TeV*, tech. rep., CERN, 2017, URL: <https://cds.cern.ch/record/2261772> (cit. on p. 44).
- [67] ATLAS Collaboration, *Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment*, tech. rep., CERN, 2019, URL: <https://cds.cern.ch/record/2688062> (cit. on pp. 45, 71).
- [68] ATLAS Collaboration, *Measurements of b-jet tagging efficiency with the ATLAS detector using  $t\bar{t}$  events at  $\sqrt{s} = 13$  TeV*, *JHEP* **08** (2018) 089, arXiv: 1805.01845 [hep-ex] (cit. on p. 45).

- [69] ATLAS Collaboration, *Commissioning of the ATLAS  $b$ -tagging algorithms using  $t\bar{t}$  events in early Run-2 data*, tech. rep., CERN, 2015, URL: <https://cds.cern.ch/record/2047871> (cit. on pp. 45, 71).
- [70] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS  $b$ -tagging algorithms for the 2017-18 LHC run*, (2017), URL: <https://cds.cern.ch/record/2273281> (cit. on pp. 45, 71).
- [71] ATLAS Collaboration, *Identification of Jets Containing  $b$ -Hadrons with Recurrent Neural Networks at the ATLAS Experiment*, (2017), URL: <https://cds.cern.ch/record/2255226> (cit. on pp. 45, 71).
- [72] P. Zak, *Why Inspiring Stories Make Us React: The Neuroscience of Narrative*, Cerebrum: the Dana forum on brain science **2015** (2015) 2 (cit. on p. 47).
- [73] O. Behnke, K. Kröniger, G. Schott and T. Schörner-Sadenius, *Data Analysis in High Energy Physics*, Wiley-VCH, 2013 (cit. on p. 47).
- [74] G. Cowan, *Statistical Data Analysis*, Oxford Science, 1998 (cit. on p. 48).
- [75] J. Neyman and E. S. Pearson, *On the problem of the most efficient tests of statistical hypotheses*, Philosophical Transactions of the Royal Society of London. Series A (1933), URL: <https://doi.org/10.1098/rsta.1933.0009> (cit. on p. 51).
- [76] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *The European Physical Journal C* **71** (2011), URL: <https://doi.org/10.1140/epjc%2Fs10052-011-1554-0> (cit. on pp. 51–53).
- [77] I. Asimov, *Franchise*, Quinn Publishing Company, 1955 (cit. on p. 53).
- [78] M. Baak et al., *HistFitter software framework for statistical data analysis*, *The European Physical Journal C* **75** (2015), URL: <https://doi.org/10.1140/epjc%2Fs10052-015-3327-7> (cit. on p. 53).
- [79] K. Cranmer, G. Lewis, L. Moneta, A. Shibata and W. Verkerke, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, tech. rep., New York U., 2012, URL: <https://cds.cern.ch/record/1456844> (cit. on p. 53).
- [80] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*, 1st ed., The MIT Press, 2016 (cit. on pp. 55, 82).
- [81] T. Mitchell, *Machine Learning*, 1st ed., McGraw-Hill Science, 1997 (cit. on p. 56).
- [82] F. Chollet et al., *Keras*, 2015, URL: <https://keras.io> (cit. on pp. 57, 60, 62, 63, 81).
- [83] C. M. Bishop, *Pattern Recognition and Machine Learning*, 1st ed., Springer Science+Business Media, 2006 (cit. on p. 61).
- [84] Y. Freund and R. E. Schapire, *A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting*, *Journal of Computer and System Sciences* (1997), URL: <https://www.sciencedirect.com/science/article/pii/S002200009791504X> (cit. on p. 61).

- 
- [85] H. Robbins and S. Munro, *A Stochastic Approximation Method*, (1951) (cit. on p. 62).
- [86] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, arXiv e-prints, arXiv:1412.6980 (2014) arXiv:1412.6980, arXiv: [1412.6980](https://arxiv.org/abs/1412.6980) [[cs.LG](#)] (cit. on p. 63).
- [87] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*, *Journal of Machine Learning Research* (2014) (cit. on p. 67).
- [88] T. Holm, *Towards a measurement of the  $tHq$  process in channels with hadronic tau lepton decays*, Bonn, 2023, URL: <https://web.physik.uni-bonn.de/group/view.php?&group=3&lang=en&c=t&id=60> (cit. on pp. 74, 102).
- [89] A. Hoecker et al., *TMVA - Toolkit for Multivariate Data Analysis*, 2007, arXiv: [physics/0703039](https://arxiv.org/abs/physics/0703039) [[physics.data-an](#)] (cit. on p. 75).
- [90] P. Martinez Agullo, personal communication, 2023 (cit. on p. 75).
- [91] M. Weiß, *Improvements on the mass reconstruction of single top quark associated Higgs production at ATLAS analyzing the decay channel Higgs to tau-tau*, MA thesis: Bonn, 2023, URL: <https://web.physik.uni-bonn.de/group/view.php?&group=3&lang=en&c=t&id=63> (cit. on p. 75).
- [92] A. Elagin, P. Murat, A. Pranko and A. Safonov, *A new mass reconstruction technique for resonances decaying to di-tau*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **654** (2011) 481, ISSN: 0168-9002, URL: <http://www.sciencedirect.com/science/article/pii/S0168900211014112> (cit. on p. 75).
- [93] M. Hübner, *Effects of tau decay product reconstruction in a Higgs CP analysis with the ATLAS experiment*, MA thesis: Bonn, 2016, URL: <https://web.physik.uni-bonn.de/group/view.php?&group=1&lang=en&c=t&id=81> (cit. on p. 75).
- [94] O. Kivernyk, personal communication, 2023 (cit. on p. 77).
- [95] G. B. Team, *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, Software available from [tensorflow.org](https://tensorflow.org/), 2015, URL: <http://tensorflow.org/> (cit. on p. 81).
- [96] G. Klambauer, T. Unterthiner, A. Mayr and S. Hochreiter, *Self-Normalizing Neural Networks*, (2017) (cit. on p. 81).
- [97] F. A. J. Obando Molina, *Implementation and evaluation of a categorical classifier for signal and background isolation in the  $tHq$  channel at ATLAS*, Bonn, 2023, URL: <https://web.physik.uni-bonn.de/group/view.php?&group=3&lang=en&c=t&id=62> (cit. on p. 82).
- [98] R. Baumann, *Hyperparameter Optimization of a Neural Network using Lorentz invariant Variables in the  $tZq$  Channel at 13 TeV with the ATLAS Experiment*, Bonn, 2020, URL: <https://web.physik.uni-bonn.de/group/view.php?&group=3&lang=en&c=t&id=39> (cit. on pp. 83, 155).

- [99] Irina Antonela Cioară, *Associated Production of a Top Quark and a Z Boson in pp Collisions at  $\sqrt{s} = 13$  TeV using the ATLAS Detector*, PhD thesis: Bonn, 2018 (cit. on p. 86).
- [100] M. Feindt and U. Kerzel, *The NeuroBayes neural network package*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **559** (2006) 190, URL: <https://www.sciencedirect.com/science/article/pii/S0168900205022679> (cit. on p. 86).
- [101] O. Freyermuth, P. Wienemann, P. Bechtle and K. Desch, *Operating an HPC/HTC Cluster with Fully Containerized Jobs Using HTCondor, Singularity, CephFS and CVMFS*, *Computing and Software for Big Science* **5** (2021) (cit. on p. 86).
- [102] T. P. Calvet, *Search for the production of a Higgs boson in association with top quarks and decaying into a b-quark pair and b-jet identification with the ATLAS experiment at LHC*, Aix-Marseille University, 2017, URL: <https://cds.cern.ch/record/2296985> (cit. on p. 98).
- [103] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, *Eur. Phys. J. C* **77** (2017) 317, arXiv: 1611.09661 [hep-ex] (cit. on p. 125).
- [104] ATLAS Collaboration, *2015 start-up trigger menu and initial performance assessment of the ATLAS trigger using Run-2 data*, ATL-DAQ-PUB-2016-001, 2016, URL: <https://cds.cern.ch/record/2136007> (cit. on p. 125).
- [105] ATLAS Collaboration, *Trigger Menu in 2016*, ATL-DAQ-PUB-2017-001, 2017, URL: <https://cds.cern.ch/record/2242069> (cit. on p. 125).
- [106] ATLAS Collaboration, *Luminosity determination in pp collisions at  $\sqrt{s} = 13$  TeV using the ATLAS detector at the LHC*, (2022), arXiv: 2212.09379 [hep-ex] (cit. on p. 125).
- [107] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, *JINST* **13** (2018) P07017 (cit. on p. 125).
- [108] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, *Eur. Phys. J. C* **70** (2010) 823, arXiv: 1005.4568 [physics.ins-det] (cit. on p. 125).
- [109] GEANT4 Collaboration, *GEANT4 – a simulation toolkit*, *Nucl. Instrum. Meth. A* **506** (2003) 250 (cit. on p. 125).
- [110] T. Sjöstrand, S. Mrenna and P. Skands, *A brief introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852, arXiv: 0710.3820 [hep-ph] (cit. on p. 126).
- [111] R. D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys. B* **867** (2013) 244, arXiv: 1207.1303 [hep-ph] (cit. on p. 126).
- [112] ATLAS Collaboration, *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model*, ATL-PHYS-PUB-2016-017, 2016, URL: <https://cds.cern.ch/record/2206965> (cit. on p. 126).
- [113] ATLAS Collaboration, *Measurement of the Inelastic Proton–Proton Cross Section at  $\sqrt{s} = 13$  TeV with the ATLAS Detector at the LHC*, *Phys. Rev. Lett.* **117** (2016) 182002, arXiv: 1606.02625 [hep-ex] (cit. on p. 126).



- 
- [114] R. Frederix and I. Tsirikos, *On improving NLO merging for  $t\bar{t}W$  production*, *Journal of High Energy Physics* **2021** (2021),  
URL: <https://doi.org/10.1007%2Fjhep11%282021%29029> (cit. on p. 126).





---

## List of Figures

---

2.1	The Standard Model of Particle Physics	5
2.2	Bhabha scattering Feynman diagram	6
2.3	Basic QED vertex	7
2.4	Primitive vertices of QCD	8
2.5	Primitive vertices of the weak interaction	9
2.6	Mexican hat potential	10
2.7	$tHq$ t-channel process in the 4-flavour scheme	13
2.8	Higgs production modes at proton colliders	15
2.9	Decay modes of the Higgs boson	16
2.10	Decay of a top quark	16
2.11	Tau decay modes	17
2.12	$tHq$ t-channel coupling diagrams	18
2.13	$tHq$ s-channel coupling diagrams	19
2.14	$tWH$ t-channel coupling diagrams	19
2.15	$tHq$ t-channel coupling diagrams displaying the couplings	19
2.16	Feynman diagrams of important background processes.	22
3.1	Sketch of the LHC	26
3.2	The LHC accelerator complex	27
3.3	Computer generated image of the ATLAS detector	32
3.4	ATLAS coordinate system	33
3.5	Sketch of the ID layers	35
3.6	Calorimeter interaction length as a function of $\eta$	36
3.7	ATLAS schematic visualising particle signatures	37
3.8	Particle Flow flowchart	42
4.1	A normal distribution, p-value and critical values	50
4.2	Visualisation of the error types in hypothesis tests	50
5.1	Sketch of a typical neural network structure	58
5.2	Nomenclature of the hidden parameters in a NN	59
5.3	Example of a loss curve	65
5.4	Example of a ROC curve	65
5.5	Example of a response plot	66
5.6	Dropout sketch	68

6.1	$tHq$ t-channel process with decay into $\tau\tau$ in the 4-flavour scheme . . . . .	70
6.2	Contributions to the $1\ell + 2\tau_{\text{had}}$ channel . . . . .	71
6.3	Contributions to the $2\ell$ OS + $1\tau_{\text{had}}$ channel . . . . .	72
6.4	Contributions to the $2\ell$ SS + $1\tau_{\text{had}}$ channel . . . . .	72
6.5	Example of the MMC output for $1\ell + 2\tau_{\text{had}}$ . . . . .	76
6.6	Pre- and post-lepton-fake-fit $\eta$ distributions of the $\tau_{\text{had}}$ in $2\ell$ OS + $1\tau_{\text{had}}$ . . . . .	78
6.7	Pre- and post-lepton-fake-fit $\eta$ distributions of the $\tau_{\text{had}}$ in $2\ell$ SS + $1\tau_{\text{had}}$ . . . . .	79
6.8	Pre- and post-lepton-fake-fit $\eta$ distributions of the leading $\tau_{\text{had}}$ in $1\ell + 2\tau_{\text{had}}$ . . . . .	80
7.1	Development of the AUC in an evolutionary optimisation . . . . .	88
7.2	Loss curve of the $2\ell$ OS + $1\tau_{\text{had}}$ NN training . . . . .	90
7.3	Loss curve of the $2\ell$ SS + $1\tau_{\text{had}}$ NN training . . . . .	91
7.4	Loss curve of the $1\ell + 2\tau_{\text{had}}$ NN training . . . . .	91
7.5	$tHq/tWH$ response and ROC curve for $2\ell$ OS + $1\tau_{\text{had}}$ . . . . .	92
7.6	$tHq/tWH$ response and ROC curve for $2\ell$ SS + $1\tau_{\text{had}}$ . . . . .	92
7.7	$tHq/tWH$ response and ROC curve for $1\ell + 2\tau_{\text{had}}$ . . . . .	92
7.8	Process-wise response distributions of the signal responses . . . . .	93
7.9	Impact of the negative weights in the $1\ell + 2\tau_{\text{had}}$ channel . . . . .	94
7.10	ROC curves of the k-fold cross validation in $1\ell + 2\tau_{\text{had}}$ . . . . .	95
8.1	Asimov pre- and post-fit regions in $1\ell + 2\tau_{\text{had}}$ . . . . .	104
8.2	Asimov pre- and post-fit regions in $2\ell$ OS + $1\tau_{\text{had}}$ . . . . .	105
8.3	Asimov pre- and post-fit regions in $2\ell$ SS + $1\tau_{\text{had}}$ . . . . .	106
8.4	Asimov fit impact parameter ranking plot of the $1\ell + 2\tau_{\text{had}}$ channel . . . . .	107
8.5	Asimov fit impact parameter ranking plot of the $2\ell$ OS + $1\tau_{\text{had}}$ channel . . . . .	108
8.6	Asimov fit impact parameter ranking plot of the $2\ell$ SS + $1\tau_{\text{had}}$ channel . . . . .	109
8.7	Pre- and post-fit regions in $1\ell + 2\tau_{\text{had}}$ . . . . .	111
8.8	Pre- and post-fit regions in $2\ell$ OS + $1\tau_{\text{had}}$ . . . . .	112
8.9	Pre- and post-fit regions in $2\ell$ SS + $1\tau_{\text{had}}$ . . . . .	113
8.10	Fit impact ranking plot of the $1\ell + 2\tau_{\text{had}}$ channel . . . . .	114
8.11	Fit impact ranking plot of the $2\ell$ OS + $1\tau_{\text{had}}$ channel . . . . .	115
8.12	Fit impact ranking plot of the $2\ell$ SS + $1\tau_{\text{had}}$ channel . . . . .	116
8.13	Combination and exclusion limits of the fits . . . . .	117
E.1	Feature plots $2\ell$ OS + $1\tau_{\text{had}}$ . . . . .	136
F.1	Feature plots $2\ell$ SS + $1\tau_{\text{had}}$ . . . . .	142
G.1	Feature plots $1\ell + 2\tau_{\text{had}}$ . . . . .	148
H.1	Background response and ROC curve for $2\ell$ OS + $1\tau_{\text{had}}$ . . . . .	149
H.2	Background response and ROC curve for $2\ell$ SS + $1\tau_{\text{had}}$ . . . . .	149
H.3	Background response and ROC curve for $1\ell + 2\tau_{\text{had}}$ . . . . .	150
H.4	$Z$ + jets response and ROC curve for $2\ell$ OS + $1\tau_{\text{had}}$ . . . . .	150
H.5	$t\bar{t}$ + X response and ROC curve for $2\ell$ SS + $1\tau_{\text{had}}$ . . . . .	150
H.6	$t\bar{t}$ response and ROC curve for $1\ell + 2\tau_{\text{had}}$ . . . . .	151

---

H.7	Process-wise response distributions of the categorical responses . . . . .	152
H.8	Process-wise response distributions of the background responses . . . . .	153
I.1	Loss for the $1\ell + 2\tau_{\text{had}}$ channel for Lorentz-invariant input variables . . . . .	155
I.2	Categorical response and ROC curve for $1\ell + 2\tau_{\text{had}}$ for Lorentz-invariant variables . .	156
I.3	Background response and ROC curve for $1\ell + 2\tau_{\text{had}}$ for Lorentz-invariant variables .	156
I.4	Signal response and ROC curve for $1\ell + 2\tau_{\text{had}}$ for Lorentz-invariant variables . . . .	157
J.1	K-fold cross validation results for $1\ell + 2\tau_{\text{had}}$ . . . . .	159
J.2	K-fold cross validation results for $2\ell \text{ SS} + 1\tau_{\text{had}}$ . . . . .	160
J.3	K-fold cross validation results for $2\ell \text{ OS} + 1\tau_{\text{had}}$ . . . . .	160
K.1	Constraints on the NPs in the $1\ell + 2\tau_{\text{had}}$ Asimov fit . . . . .	162
K.2	Constraints on the NPs in the $2\ell \text{ OS} + 1\tau_{\text{had}}$ Asimov fit . . . . .	163
K.3	Constraints on the NPs in the $2\ell \text{ SS} + 1\tau_{\text{had}}$ Asimov fit . . . . .	164
K.4	Correlation matrix of the $1\ell + 2\tau_{\text{had}}$ Asimov fit . . . . .	165
K.5	Correlation matrix of the $2\ell \text{ OS} + 1\tau_{\text{had}}$ Asimov fit . . . . .	165
K.6	Correlation matrix of the $2\ell \text{ SS} + 1\tau_{\text{had}}$ Asimov fit . . . . .	166
L.1	Constraints on the NPs in the $1\ell + 2\tau_{\text{had}}$ data fit . . . . .	168
L.2	Constraints on the NPs in the $2\ell \text{ OS} + 1\tau_{\text{had}}$ data fit . . . . .	169
L.3	Constraints on the NPs in the $2\ell \text{ SS} + 1\tau_{\text{had}}$ data fit . . . . .	170
L.4	Correlation matrix of the $1\ell + 2\tau_{\text{had}}$ data fit . . . . .	171
L.5	Correlation matrix of the $2\ell \text{ OS} + 1\tau_{\text{had}}$ data fit . . . . .	171
L.6	Correlation matrix of the $2\ell \text{ SS} + 1\tau_{\text{had}}$ data fit . . . . .	172



---

## List of Tables

---

2.1	Branching ratios of the Higgs boson	14
5.1	Selection of activation functions	60
6.1	Yields in the $1\ell + 2\tau_{\text{had}}$ channel	73
6.2	Yields in the $2\ell$ OS + $1\tau_{\text{had}}$ channel	73
6.3	Yields in the $2\ell$ SS + $1\tau_{\text{had}}$ channel	74
7.1	Input features of the $2\ell$ SS + $1\tau_{\text{had}}$ NN	83
7.2	Input features of the $2\ell$ OS + $1\tau_{\text{had}}$ NN	84
7.3	Input features of the $1\ell + 2\tau_{\text{had}}$ NN	85
7.4	Tested ranges of hyperparameters in the evolutionary optimisation	87
7.5	Final hyperparameters of the $1\ell + 2\tau_{\text{had}}$ NN	88
7.6	Final hyperparameters of the $2\ell$ OS + $1\tau_{\text{had}}$ NN	88
7.7	Final hyperparameters of the $2\ell$ SS + $1\tau_{\text{had}}$ NN	89
8.1	Yields in the $1\ell + 2\tau_{\text{had}}$ fit regions	98
8.2	Yields in the $2\ell$ OS + $1\tau_{\text{had}}$ fit regions	99
8.3	Yields in the $2\ell$ SS + $1\tau_{\text{had}}$ fit regions	99
C.1	Integrated luminosity per year	125
C.2	Summary of the baseline simulated signal and background event samples	126
D.1	Yields and $\tau_{\text{had}}$ fake factors for the $1\ell + 2\tau_{\text{had}}$ region	127
D.2	Summary of the jet $\tau_{\text{had}}$ fakes obtained in the $2\ell$ OS + $1\tau_{\text{had}}$ channel	128
D.3	Event yields and scale factors for the light lepton fakes in the $2\ell$ OS + $1\tau_{\text{had}}$ channel	128
D.4	Scale factors for the $\tau_{\text{had}}$ fakes in the $2\ell$ SS + $1\tau_{\text{had}}$ channel	128
D.5	Scale factors for the light lepton fakes in the $2\ell$ SS + $1\tau_{\text{had}}$ channel	129
D.6	Scale factors for $\tau_{\text{had}}$ fakes in the $1\ell + 2\tau_{\text{had}}$ channel	129



---

## Glossary

---

***b*-jet** A jet that originates from the decay of a *b* quark. [20](#), [45](#), [71](#), [75](#), [77](#), [78](#)

**ATLAS** A Toroidal LHC Apparatus - the largest general-purpose particle detector experiment at the Large Hadron Collider. [23](#), [25](#), [26](#), [28](#), [31–36](#), [40](#), [183](#)

**CMS** The Compact Muon Solenoid experiment - one of the two large general-purpose particle physics detectors built on the Large Hadron Collider. [25](#)

**ECal** Electromagnetic calorimeter - electromagnetic calorimeter in the ATLAS calorimeter system. [34](#)

**FCal** Forward calorimeter - Forward calorimeter in the ATLAS calorimeter system. [34](#)

**HCal** Hadronic calorimeter - hadronic calorimeter in the ATLAS calorimeter system. [34](#)

**ID** Inner Detector - Tracking detectors forming the innermost part of the ATLAS detector. [34](#)

**LHC** Large Hadron Collider - the world's largest and highest-energy particle collider. [23](#)

**LHCb** Large Hadron Collider beauty - an experiment at the Large Hadron Collider optimised for the study of *b*-physics. [25](#)

**MS** Muon spectrometer - Muon spectrometer of the ATLAS detector. [35](#)

**PU** Pile-up - additional events in a single interaction. [25](#)





---

## Acronyms

---

**AUC** Area Under the Curve. [64](#), [65](#), [86–89](#), [184](#)

**BDT** Boosted Decision Tree. [61](#), [75](#), [78](#)

**BSM** beyond Standard Model. [18](#)

**ECal** Electromagnetic calorimeter. [34](#), *Glossary*: [ECal](#)

**FCal** Forward calorimeter. [34](#), *Glossary*: [FCal](#)

**HCal** Hadronic calorimeter. [34](#), *Glossary*: [HCal](#)

**ID** Inner Detector. [34–42](#), [45](#), *Glossary*: [ID](#)

**LHC** Large Hadron Collider. [23–27](#), [32](#), [33](#), [36](#), [183](#), *Glossary*: [LHC](#)

**MC** Monte Carlo. [53](#), [69](#), [77–80](#), [84](#), [86](#), [97](#), [99](#), [100](#), [103](#), [117](#)

**MLE** maximum likelihood estimation. [49](#), [51](#), [52](#), [71](#), [73](#), [74](#), [77–81](#), [86](#), [90](#), [97](#), [100–103](#)

**MS** Muon spectrometer. [35](#), [36](#), [39](#), [40](#), *Glossary*: [MS](#)

**NF** normalisation (abbreviated to norm) factor. [97](#), [99](#), [101–103](#), [110](#), [114–117](#), [161](#), [167](#)

**NN** neural network. [20](#), [55](#), [57–59](#), [61](#), [62](#), [64](#), [66–68](#), [81–87](#), [89–91](#), [97–99](#), [155](#), [183](#), [184](#), [187](#)

**NP** nuisance parameter. [100–103](#), [110](#), [114–116](#), [162–164](#), [167–170](#)

**POI** parameter of interest. [51](#), [52](#), [97](#), [100](#), [107–110](#), [114–116](#)

**PU** Pile-up. [25](#), [26](#), [36](#), [38](#), [40](#), [41](#), [43](#), [44](#), *Glossary*: [PU](#)

**ROC** receiver operating characteristic. [64–66](#), [86](#), [89](#), [90](#), [92](#), [94](#), [95](#), [149–151](#), [155–157](#), [159](#), [160](#), [183–185](#)

**SM** Standard Model. [53](#), [97](#), [101](#), [117](#)



---

## Acknowledgements

---

First and foremost, I would like to thank my supervisor, Ian C. Brock, who despite all challenges of the last few years, always managed to be there for his group. I thank him for his meticulousness, not only in dealing with typesetting systems, which forced me to surpass myself many times. I would like to thank Klaus Desch for his readiness to review my thesis but also for, together with Ian, enabling many a Jamboree, a tradition that was even more important for team-building in the last years. I would also like to thank Prof. Carsten Urbach and Prof. Raoul-Martin Memmesheimer for their uncomplicated willingness to become part of my commission.

I am indebted to the generation of scientists who made up this group when I first became part of it in 2016. They were all role models for me in their own way and gave me the courage to get involved myself: Rui for countless discussions and his infectious laugh, Ozan for his unwavering warmth and Anji for his quick-witted humour and above all for his straightforward help even to the final days of my thesis.

When I started my PhD in 2020, I was eagerly awaiting my opportunity to leave my own mark on academia. A global pandemic ensured that little turned out as I had imagined. I must take this opportunity to thank all those who have been my friend and support during a phase in which I had to come to terms with this reality. I would like to thank all those who got involved in their own way and, despite the circumstances, worked with me to keep communication and cohesion high in our group. And I certainly couldn't have done it without Rico, who was always a good friend to me, even during the bleak and discouraging months. A big thank you is also due to those who, with a zest for action and a smile on their faces, breathed new life into the group when we returned to the office. Oliver, Richard and Nilima seized the opportunity and turned us into a group that I will look back on fondly. I would like to thank Tanja and Oleh for their support during this work and for engaged discussions far beyond the topic. My gratitude also goes to Oliver Freyermuth, who is accomplishing a Herculean task and whose answers at any time of day or night are more valuable than he himself perhaps knows. There are many others that have shaped this path and made it a little more memorable. I can't name them all here because I would inevitably forget someone.

Meine Freundin, meine Freunde und Familie waren mir stets eine so starke Stütze, dass ich manchmal selbst staune, wieviel Balast sie ungerührt ertragen haben. Ich danke meinem Vater, der es bis heute versteht eine kindliche Faszination für die Fragen der Welt mit einem Pragmatismus zu verbinden, der ihn zum besten Lehrer macht, den ich je kennen durfte. Meiner Mutter danke ich, dass sie es stets versteht, mich auf den Boden der Tatsachen zurückzuholen und dafür, dass sie immer ein offenes Ohr für mich hat, auch wenn es sie manche Stunde Schlaf kostet. Und mein letzter Dank gebührt Florian, dessen Tatendrang scheinbar endlos war und dessen Verlässlichkeit mir Hoffnung gab, wo ich sie schon verloren glaubte. Einen besseren Bruder hätte ich mir nie wünschen können.