# Associated production of a top-quark and a Higgs-boson in the $H \to \tau\tau$ decay channel in $pp$ collisions at $\sqrt{s} = 13\,\text{TeV}$ using the ATLAS detector

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

vorgelegt von
Florian Kirfel
aus
Mechernich

Bonn, 30.04.2025

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

| | |
|---|---|
| Gutachter und Betreuer: | Prof. Dr. Ian C. Brock |
| Gutachter: | Prof. Dr. Klaus Desch |
| Tag der Promotion: | 03.07.2025 |
| Erscheinungsjahr: | 2025 |

# Contents

CHAPTER 1

# Introduction

Understanding the fundamental structure of matter has always been a fascination of mine. In school, however, this subject was scarcely covered if not to say completely neglected. Nevertheless, from a historical point of view I was by no means unique with this desire. Since the 6th century BC with the development of atomism (in ancient Greece and India, amongst other places), the old philosophers formulated a theory concerning the composition of matter in the universe [1].

The first experimental evidence of these fundamental components however did not become apparent until the beginning of the 19th century. In 1808, English scientist John Dalton was able to show that all chemical elements consist of what he considered to be indivisible units [2]. Adopting the naming of the aforementioned ancient theories he termed these units atoms.

About 100 years later, Sir Joseph John Thomson's studies[1] demonstrated that the term atom was not quite suitable. In his experiments with cathode rays, he confirmed the existence of the electron, as a distinct component of atoms [3]. From this point on, it became evident that an atom defines a chemical element while being by no means indivisible.

Over the course of the next 120 years, many groundbreaking experiments took place. In parallel new theories arose, predicting or explaining these experiments. And like the pieces of a jigsaw puzzle, experiment and theory gradually came together, forming the picture of our current understanding of matter and its interactions: the Standard Model of particle physics.

In the analogy of a jigsaw puzzle, all fundamental particles can be seen as individual pieces of the puzzle. Each piece comes with a distinct shape that defines which interactions the associated particle can take part in.

In its current form, the Standard Model of particle physics represents one of the most successful theories in the field of physics. Its predictions on fundamental matter particles and associated interactions are impressively consistent with the increasingly more precise experimental results. Nevertheless, it remains evident that this physical jigsaw puzzle still contains serious gaps, as phenomena such as the observation of dark matter or the gravitational interactions have not yet found their way into the existing picture.

In a real jigsaw puzzle, the obvious approach to close such gaps is to analyse the shape of all adjacent edges. Similarly, the Standard Model of particle physics has shown that constant experimental progress at the edge of what is known of helps to unveil and understand new phenomena.

[1] Alongside the experiments of German physicist Emil Wiechert.

In the year 2012, one of the major remaining gaps in the Standard Model puzzle was filled when experiments at the European nuclear research centre CERN experimentally observed the Higgs-boson [4, 5]. The associated theory offers an explanation on how elementary particles acquire their mass. As often seen in the Standard Model's development, previous measurements at the frontiers (e.g. at the LEP accelerator [6]) had already partially narrowed down the rough outlines of this hitherto unknown puzzle piece.

Likewise, since this discovery, the puzzle's edges still provide the ideal place to search for the next piece. One particularly interesting approach to look for such a new puzzle piece is the investigation of the coupling between the heaviest known particles to date, the top-quark, and the Higgs-boson. Both, the strength of this coupling and its relative sign with respect to the vector boson coupling, are of outstanding interest as they sit right at the puzzle's edge. Precisely measuring this coupling will not only allow the existing theory to be tested but might possibly even open the door to new hypotheses.

In this thesis, the coupling strength is measured via the associated production of a top-quark and a Higgs-boson. The utilised experimental set of data was recorded by the ATLAS experiment at CERN. Events in which the produced Higgs-boson decays into two tau-leptons are analysed.

Initially, Chapter 2 provides the theoretical basis for understanding the formerly described Standard Model of particle physics as well as the associated production of a top-quark and a Higgs-boson. Based on this, Chapter 3 explains the experimental setup used at CERN for data acquisition. Both, the accelerator complex and the ATLAS detector, are introduced. This is followed by a brief summary of the fundamental statistical methods utilised in the analysis in Chapter 4. The subsequent Chapter 5 introduces the fundamental principles of the machine learning algorithm employed to solve multi-class classification problems.

All information given in this first part is not intended to serve as a complete introduction of each topic. Rather, an attempt is made to cover the key concepts necessary for a thorough understanding of the analysis, presented in the second part. For more detailed information, each chapter provides references to additional literature.

Building on these foundations, the second part of the thesis initially introduces the used data in combination with a general preselection in Chapter 6 as well as Chapter 7. Subsequently, the machine learning method is evaluated in Chapter 8 with the goal of optimising the separation between the process of interest and undesired background processes. The next major part, an approach to correctly estimate the different backgrounds, is presented in Chapter 9. This background estimation is crucial to precisely determine the cross-section of the associated production of a top-quark and a Higgs-boson, which is ultimately discussed in Chapter 10.

CHAPTER 2

# Theoretical concepts

This chapter is intended to serve as an introduction to the basic theoretical concepts necessary to understand this thesis. First, the Standard Model of particle physics with all its essential constituents is explained in Section 2.1. Based on this, the physics of proton-proton collisions, relevant for creating the dataset analysed in this thesis, is discussed in Section 2.2. Finally, the actual process of interest of this analysis, i.e. the associated production of a single top-quark and a Higgs-boson, is described in Section 2.3. Alongside an overview of the relevant physics, foremost the potential decay modes of all involved particles, motivation to measuring this rare process is provided.

Certain aspects of the Standard Model are only covered briefly or omitted, as a complete introduction to this topic would go beyond the scope of this thesis. The interested reader is referred to the following resources: an excellent overview of the topic with insights into the historical experimental progress can be found in [7]. For a more detailed understanding of the underlying mathematical theory, reference is made to [8].

## 2.1 Standard Model of particle physics

The field of particle physics is the centrepiece to our understanding of the universe surrounding us. The current knowledge on the fundamental particles forming our world and the forces, apart from gravity[1], that hold them together are described by the Standard Model of particle physics. Those particles and forces are embedded in a framework of quantum field theories (QFTs).

For instance, the concept of the atomic shell's structure, in which the electrically negative electrons orbit around a positive nucleus, is described by the theory of quantum electrodynamics (QED). The stable state of this positively charged atomic nucleus itself can be described by the strong interaction and the associated theory of quantum chromodynamics (QCD). Lastly, the model also includes the weak interaction, sometimes referred to as quantum flavourdynamics (QFD). Albeit not directly responsible for the structure of our atoms, weak interactions are at least partially responsible for their decay.

There is no straightforward or correct approach to explaining the Standard Model. Regardless of whether the elementary particles or the fundamental forces are discussed first, some aspects are always mentioned without a detailed introduction and covered in more depth later on.

[1] Gravity, although very relevant at macroscopic scales, can usually be ignored in particle physics.

In this thesis, the elementary particles are introduced initially in Section 2.1.1, as historically, they were already partially known before the theories concerning the fundamental interactions, presented in Section 2.1.2, were developed. Subsequently, Section 2.2 outlines the physics of proton-proton collisions, essential to create the analysed data in this thesis.

The Standard Model's predictions have impressively demonstrated their accuracy over a broad energy range in the past decades. Nonetheless, it cannot be regarded as a complete theory with infinite validity, as already mentioned in Chapter 1. In addition to gravitational effects, which currently cannot be unified with the Standard Model, there are still plenty observed phenomena that remain in conflict with the current theory, briefly discussed in Section 2.1.4.

### 2.1.1 Elementary particles

The elementary matter particles of the Standard Model are the twelve fermions[2]. Fermions are particles with a half-integer spin and are therefore described by Fermi-Dirac statistics. Spin can be understood as an intrinsic form of angular momentum, thus also being a conserved quantity. Each elementary particle is depicted in the left part of Figure 2.1, where a subdivision into three generations is clearly visible. Each generation is composed of two leptons and two quarks, whereby the particles' masses[3] increase along with the generation.

The first generation is formed by the electron ($e^-$), the electron neutrino ($\nu_e$) along with the up- ($u$) and down-quark ($d$). This primary generation is responsible for the structure of the atom and thereby that of all chemical elements: the electron orbits around the positively charged nucleus, composed of protons and neutrons, termed nucleons. The up- and down-quarks represent the constituents of these nucleons. Historically, first direct evidence of this substructure was provided in deep inelastic scattering experiments at the SLAC National Accelerator Laboratory [11].

The subsequent two generations can be understood to be, heavier but otherwise almost identical, copies of the first generation particles. To date, there is no experimental evidence for further generations of matter. On the contrary, strong arguments in favour of there being exactly three generations exist [12].

The relativistic quantum mechanical description of these particles is achieved through the Dirac equation. Negative energy solutions to this equation imply that an antiparticle with opposite charge quantum numbers must exist for each particle[4]. In other respects, i.e. in terms of mass, spin and lifetime, these antiparticles behave identically to their counterparts. In the case of charged leptons, the antiparticles carry a positive electric charge and are thus commonly labelled as $\ell^+$ (e.g. the positron $e^+$, the antiparticle of the electron $e^-$). Antiquarks as well as antineutrinos, are indicated by an overbar (e.g. the antiup-quark $\overline{u}$ and the antielectron neutrino $\overline{\nu}_e$). For the sake of simplicity, most descriptions in the following chapters of this thesis just use the particle name and symbol. Under this convention, antiparticles, if not explicitly mentioned, are always incorporated.

All particles can interact via the weak force and, as can be seen from Figure 2.1, only neutrinos do not carry electric charge, as a result being the only particles that cannot interact electromagnetically.

[2] This number depends on the way of counting. If antiparticles are included, the result is a count of 24 elementary particles. Additionally taking into account the colour charge variations of quarks results in 48 fermions.

[3] All mass values in Figure 2.1 are given in the natural unit system. See Appendix A for an explanation.

[4] In the original conception, Dirac only postulated the antiparticle of the electron, i.e. the positron, whose existence was confirmed by Occhialini and Blacket in 1933 [13, 14].

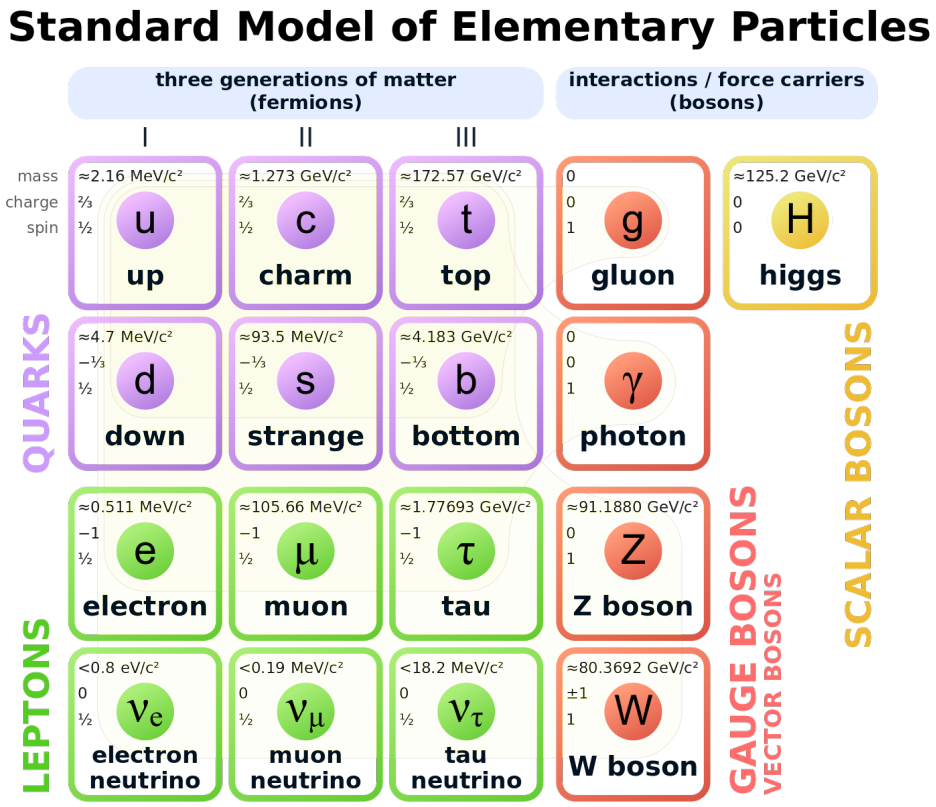



Figure 2.1: All elementary particles in the SM of particle physics. Fermions are divided into quarks (purple) and leptons (green). The gauge bosons are shown in red and the scalar Higgs-boson is shown in yellow. The sketch is adapted from [9] with all mass values being updated to the 2024 version of *Review of Particle Physics* [10].

The strong interaction, in contrast, only affects quarks via the colour charge, with its three states red, green and blue, assigned to them.

Quarks are never observed as free particles but in bound states, the hadrons. Typical hadronic signatures in experiments contain the lightest of such bound states which are categorised here: starting with the pion [15] as well as the kaon [16] in 1947, a large number of these bound states consisting of different quarks were discovered. Murray Gell-Mann first developed an approach to bring order to this chaotic particle-zoo and shortly afterwards a model to explain the structure of these particles, the quark model[5] [17, 18]. Later experimental results on the substructure of these hadrons confirmed the correctness of this quark model. Excluding exotic compounds, all known hadrons can be divided into two groups: first, there are mesons, formed by quark-antiquark pairs, such as the pion, being a constituent of secondary cosmic rays. On the other hand, a combination of three quarks or antiquarks, known as a baryons, also exists. Nucleons, for example, consist of two up-quarks and one down-quark in case of the proton and two down-quarks and one up-quark in case of neutrons. An exception to

[5] Murray Gell-Man and George Zweig independently developed a similar model. Zweig called his version of the quarks aces.

this is the heaviest quark, the top-quark, which is not found in bound states. Since its lifetime of $\tau \approx 5 \times 10^{-25}$ s [10] is shorter than the typical time scale of strong interactions the top-quark decays before forming hadrons.

The last elementary particles in the model are the uncharged leptons, i.e. the neutrinos. Originally postulated to explain the energy spectrum of $\beta$-decays by Pauli in 1930 [19], they were not discovered until 1956 [20]. Neutrinos only interact weakly and are regarded as massless[6] by the Standard Model.

Please note that technically neutrinos need a special consideration: The weak flavour eigenstates $\nu_e$, $\nu_\mu$ and $\nu_\tau$ shown in Figure 2.1 are actually quantum mechanical mixtures of the mass eigenstates $\nu_1$, $\nu_2$ as well as $\nu_3$. These mass eigenstates can oscillate between the weak flavour states. However, only the flavour eigenstate is decisive for interactions via the weak force. Therefore, neutrinos are always defined by their flavour state in the course of this thesis.

From today's perspective, the Standard Model's list of elementary particles may seem complete. However, a glance into the past shows that we already reached a similar point several times: after the neutron was discovered by Chadwick [21], an apparently satisfactory picture sufficient to explain the structure of atoms existed. Later, however, findings at the forefront of both theory and experiment repeatedly led to a modification of this picture. So, it remains to be seen whether, and if so, in what form, the list of elementary components might have to be updated and expanded.

### 2.1.2 Gauge bosons & fundamental forces

In the classical understanding, interactions are mediated by an existing field. In contrast to that, in the Standard Model view, interactions are mediated by the exchange of a particle. These are the gauge bosons with an integer spin shown on the right-hand side of Figure 2.1. A full theoretical description of the theory behind the individual forces and the bosons assigned to them is not the aim of this section. Instead, the forces' basic properties, starting with electromagnetic interactions, are explained to aid the understanding of this thesis.

At this point it should be noted, that the entire theory of the Standard Model is subject to internal symmetries from which specific properties of the interactions can be derived. These symmetry groups are mentioned in the following but not discussed in depth to avoid impairing the flow of reading. A detailed introduction can be found in [22].

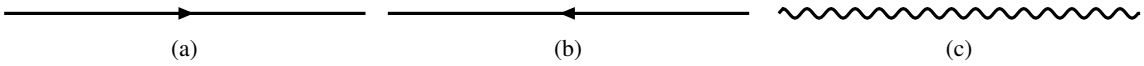
#### Electromagnetic interactions and underlying concepts

Electromagnetic interactions, described by the theory of QED, a gauge theory based on the U(1) symmetry group, are the best understood interactions in the Standard Model. The force carrier of the electromagnetic interaction is the uncharged photon ($\gamma$) which couples to all other electrically charged particles. Its experimental discovery is not easy to pin down, as several studies and experiments such as Einstein's studies on the photoelectric effect [23] and Compton's scattering experiment published in 1922 [24], gradually led to the acceptance of its existence.

Since QED is considered to be the simplest, most elegant of the interactions, it serves as a perfect starting point for giving an introduction to Feynman diagrams along with their description of scattering processes as well as decays. Feynman diagrams, named after the American physicist Richard Feynman, are an easy to understand graphical representation of the complex underlying QFT interactions. These fundamentals are translated into the diagram's surprisingly simple conventions: In all the diagrams

[6] Various experiments showed, even if only an upper mass limit could be determined, that this is not the case.

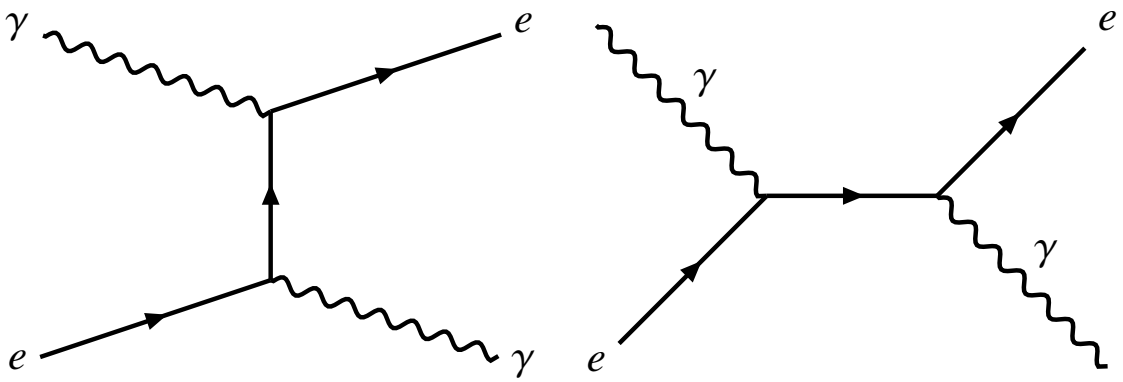
shown in this thesis, time runs horizontally from left to right. A diagram's vertical axis does not have a spatial dimension, even though the visualisation suggests otherwise, and thus does not indicate a physical distance.



(a) (b) (c)

Figure 2.2: Basic Feynman diagram elements: solid lines for (a) fermions, (b) antifermions as well as (c) a wavy line indicating photons.

Each fermion is represented by a solid line, with an associated arrow indicating whether it is a particle (arrow aligned with positive time direction, see Figure 2.2(a)) or an antiparticle (arrow pointing in the negative time direction, see Figure 2.2(b)). Bosons, i.e. the exchange particles of the forces, are represented by special lines, in the case of the photon by a wavy line (Figure 2.2(c)). The interaction of two particles is characterised by a vertex. In the simplest case, the lines of two fermions converge with the line of an exchange particle. Each vertex satisfies energy-momentum conservation and allows the generation, annihilation or scattering of particles. Lines that do not exit the diagram are called virtual particles. They are neither detectable nor do they necessarily carry the same mass as the ordinary particles listed in Figure 2.1. As an example, the Compton scattering process is shown in Figure 2.3.




(a) (b)

Figure 2.3: Two possible Feynman diagrams showing Compton scattering.

Both shown diagrams characterise the same basic process, since incoming and outgoing particles are identical. Using internal loops[7], this fact allows the creation of an infinite number of diagrams for just a single process. The mathematical framework underlying the Feynman diagrams requires the summation of this infinite number of diagrams with the same initial and final states to correctly describe the actual physics. However, this only appears to cause a problem as each vertex can be associated with a coupling constant $g$.

As a consequence, the quantum mechanical transition matrix $\mathcal{M}$ represented via each diagram is multiplied by a factor of this constant per vertex. Furthermore, the actual interaction probability of a

[7] An internal loop changes neither the initial nor the final state particles and is depicted as a cycle in a Feynman diagram. A photon for example could form a loop by decaying into a virtual electron-positron pair which annihilates back into a photon, therefore leaving the final state untouched.

process depends quadratically on this matrix element. In case of QED it is often more convenient to use the dimensionless fine-structure constant[8] $\alpha = 1/137 \propto g^2$ [7]. To provide an example of the resulting consequences based on electromagnetic interactions, one can consider the Feynman diagrams for electron-electron scattering depicted in Figure 2.4.

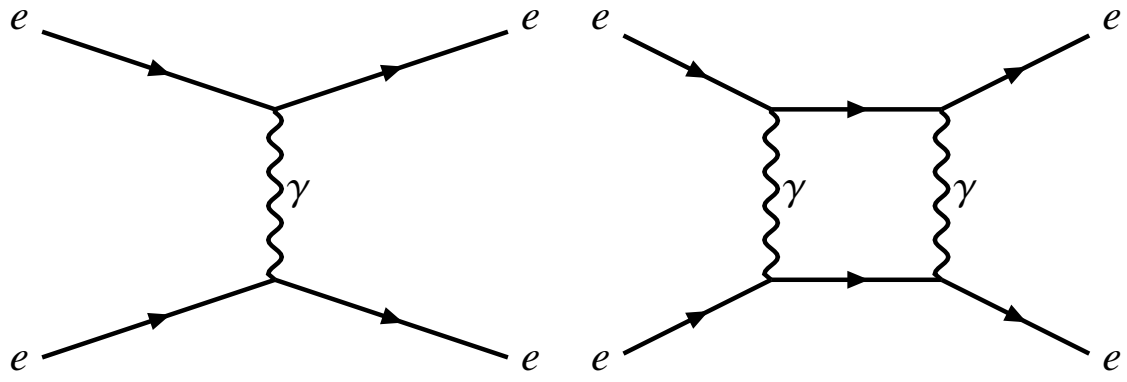



Figure 2.4: Feynman diagrams for electron-electron scattering ($e^- e^- \rightarrow e^- e^-$).

Both diagrams represent the same process and must consequently be summed. However, for left-hand diagram one obtains $|\mathcal{M}^2| \propto \alpha^2$, whereas the right-hand diagram indicates a proportionality to $\alpha^4$. As $\alpha < 1$, this quickly illustrates that only the left-hand diagram, defined as the leading order (LO), is relevant. The contribution of the next-to-leading order (NLO) is significantly smaller and as a result negligible in the majority of cases[9].

The real quantification of a process' probability is ultimately achieved by calculating the cross-section $\sigma$, which itself depends on all contributing Feynman diagrams. In a simple geometric view, it can be thought of as the area of a target particle that is visible to an incoming particle beam:

$$\sigma = \frac{\text{interactions per unit time}}{\text{beam particles per unit time} \cdot \text{scattering centres per unit area}}. \tag{2.1}$$

For any two-body to two-body process one can define it by using the magnitude of the transition matrix element $|\mathcal{M}_{fi}|$:

$$\sigma = \frac{\vec{p}_f}{64\pi^2 s \vec{p}_i} \int |\mathcal{M}_{fi}|^2 \, \mathrm{d}\Omega \,, \tag{2.2}$$

where $\mathrm{d}\Omega$ and $\vec{p}_{f,i}$ refer to the solid angle element and the final and initial state momenta in the centre-of-mass frame with energy $\sqrt{s}$.

In the case of particle decays, the general approach is quite similar and again easily visualised

[8] The term constant is technically misleading in this case, as the coupling increases with smaller distances. The reason for this effect is the generation of virtual electron-positron pairs which alter the electromagnetic field's charge distribution. This effect, known as called vacuum polarisation, causing the Lamb shift, is closely related to the concept of renormalisation. For a detailed introduction see [25].

[9] This view on Feynman diagrams is closely linked to the underlying perturbation theory.

using Feynman diagrams. The crucial aspect in all decays is that the rest mass of the combined final state has to be lower than that of the initial state. Other quantities, such as the electric charge, are obviously conserved in decays as they are in scattering processes. For most particles, several decays are theoretically possible. The transition rates of each individual mode, referred to as partial decays widths $\Gamma_j$, for a two-body decay ($A \rightarrow 1 + 2$) are given by:

$$\Gamma_j = \frac{p}{32\pi^2 m_A} \int |\mathcal{M}_{fi}|^2 \, d\Omega \,, \tag{2.3}$$

where $m_A$ is defined as the decaying particle's mass and $p$ refers to the final state particles' momenta in the centre-of-mass frame. The total decay rate per unit time $\Gamma$ can be derived by summing up all individual decay modes [7]:

$$\Gamma = \sum_i \Gamma_j \,. \tag{2.4}$$

The relative occurrence of a particular decay channel, termed the branching ratio (BR), is simply obtained by

$$\mathrm{BR} = \frac{\Gamma_j}{\Gamma} \,. \tag{2.5}$$

As particle decays are statistical processes represented through a Poisson distribution, the number of remaining particles $N(t)$ after time $t$ can be modelled by an exponential function:

$$N(t) = N(0) \exp\left(-\frac{t}{\tau}\right) \,, \tag{2.6}$$

with $\tau = 1/\Gamma$ defined as the particle's lifetime.

### Strong interactions

Strong interactions in the Standard Model are described by QCD and the associated SU(3) symmetry group. Here, the role of the force carrier is taken over by eight gluons ($g$), whose first experimental evidence was found in 1979 at the DESY research centre [26].

In contrast to the photon, the gluons themselves carry a charge, namely the colour charge with its three forms, termed red, blue and green[10]. This immediately leads to one major difference compared to the uncharged photon: due to themselves carrying colour charge, gluons not only couple to the colour-charged quarks but also take part in self-couplings. Thus, not only Feynman diagrams as in Figure 2.5(a) are allowed but also the diagram shown in Figure 2.5(b) is possible, where the gluon is indicated by a helix.

Compared to the fine-structure constant, the coupling constant $\alpha_S$ of strong interactions is large at low energy scales, $\alpha_S \sim O(1)$ [7]. As a consequence, the mere summation of the Feynman diagrams appears problematic as, unlike in QED, higher orders also have to be taken into account. Fortunately, the strong coupling constant $\alpha_S$ is not at all constant and instead changes with the energy scale,

[10] Anticolours are not listed here but also exist.

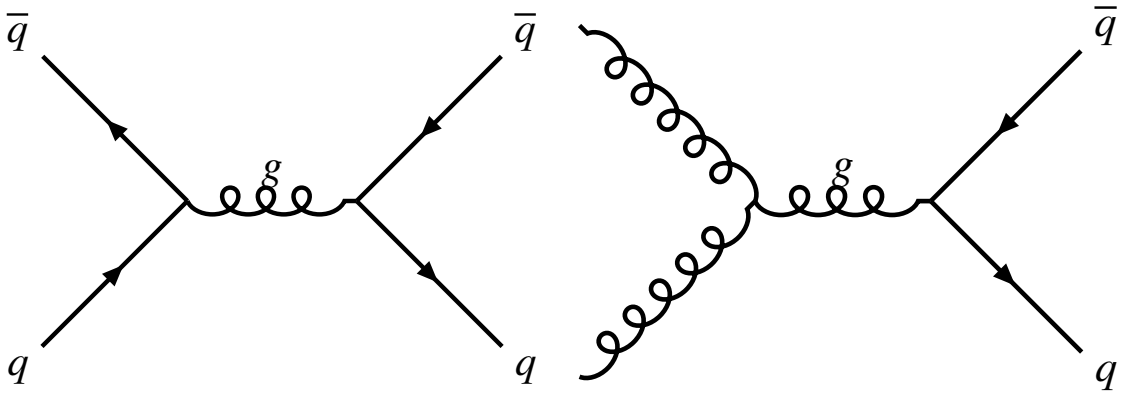



Figure 2.5: Quark-antiquark production through (a) quark-antiquark annihilation and (b) gluon-gluon fusion.

therefore often named a running coupling constant[11]. At low energies the constant is of $O(1)$, whereas at high energies the coupling's strength decreases[12] [7]. This is one of the reasons why quarks are solely observed in bound states, such as the proton, a phenomenon known as confinement. At high energies the phenomenon of asymptotic freedom arises, where quarks can be regarded as quasi-free particles.

Electric charge as well as the net colour charge are preserved in strong interactions. Furthermore, quark flavour cannot be changed, as the gluon can only couple to a vertex with a same flavour quark-antiquark pair (e.g. to a bottom- and antibottom-quark but not to a charm- and antiup-quark).

### Weak interactions

The last interaction included in the Standard Model is the weak interaction[13]. Its, in contrast to quarks and gluons, massive exchange particles are firstly the neutral $Z$-boson and secondly the charged $W^{\pm}$-bosons. In Feynman diagrams both are marked by a wavy line just as the photon. The $W$-bosons and shortly after also the $Z$-boson have been discovered at the UA1 and UA2 experiments [27–30].

The weak interaction couples to all known fermions. Any process mediated by the photon can in theory also be carried out by the neutral $Z$-boson, as in the case of electron-electron scattering shown in Figure 2.6(a) (compare to Figure 2.4). In contrast to this, neutrino scattering as shown in Figure 2.6(b) is only mediated by the $Z$-boson, but not by the photon. Those interactions, referred to as neutral current processes, do conserve flavour in the same way as strong and electromagnetic interactions.

The only way to change the quark flavour in an interaction is via coupling to the $W$-boson, termed charged current interactions. The most famous process mediated by the $W$-boson is the $\beta^-$-decay, visualised in Figure 2.7(b), in which an up-quark is converted into a down-quark while the $W^-$-boson itself decays into an electron and a neutrino. This process can be explained by linking the weak quark

[11] Similarly to QED, a renormalisation can also be applied here, this time considering the gluon-gluon self-coupling.

[12] The value still remains sufficiently large that NLO corrections cannot be entirely neglected when applying perturbation theory. Calculating QCD processes is hence still extremely challenging.

[13] In this context, the term weak reflects the small coupling constant relative to electromagnetic and strong interactions.

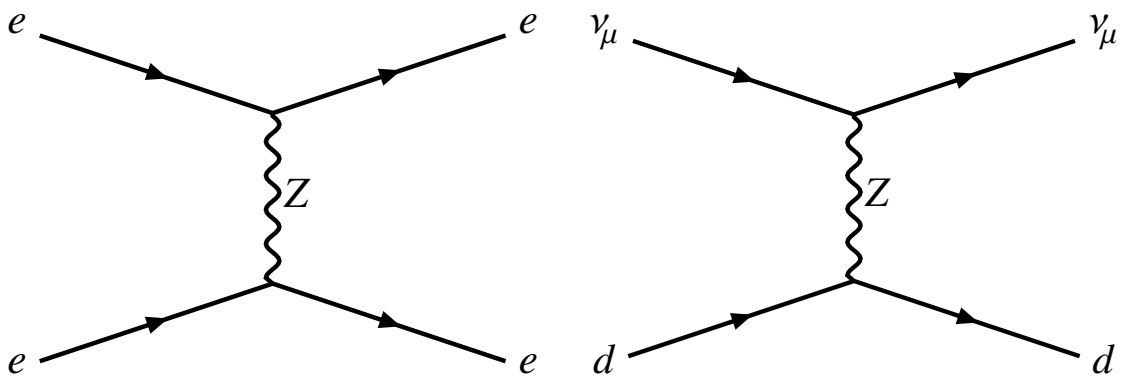



Figure 2.6: Feynman diagrams for (a) electron-electron and (b) neutrino scattering mediated by weak neutral currents.

eigenstates $q'$, to which the $W$-boson couples in such interactions, to the mass eigenstates $q$ via the Cabibbo-Kobayashi-Maskawa matrix (CKM matrix) [31]:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \cdot \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{2.7}$$

Whenever a flavour change results from a weak charged current interaction, one of the CKM matrix elements is assigned to the corresponding Feynman vertex, i.e. $V_{ud}$ in the case of the $\beta^-$-decay depicted in Figure 2.7(b). The exact values of all elements (taken from [10]) are determined in various experiments and summarised in Equation (2.8).

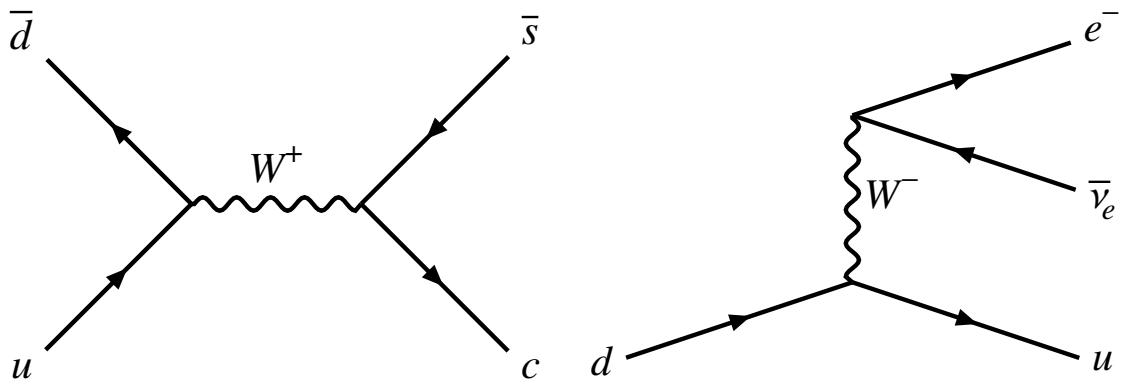
$$\begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} 0.97367 \pm 0.00032 & 0.22431 \pm 0.00085 & 0.0382 \pm 0.0020 \\ 0.221 \pm 0.004 & 0.975 \pm 0.006 & 0.0411 \pm 0.0012 \\ 0.0086 \pm 0.0002 & 0.0415 \pm 0.0009 & 1.010 \pm 0.027 \end{pmatrix}. \tag{2.8}$$

Given the CKM matrix's almost diagonal structure, a flavour change within the same quark generation is much more likely, while transitions between different generations are heavily suppressed.

In the lepton sector, a similar possibility of flavour change via the non-massless neutrinos exists. As a result, transitions between different lepton generations are possible in a similar way as for quarks. However, a violation of the lepton number conservation has not yet been observed [10].

### Electroweak unification

While trying to decipher the fundamental laws of our universe, we must inevitably ask ourselves why there are exactly four fundamental forces. One possible assumption is that all of them can be unified. In the case of electromagnetic and weak interactions, the electroweak theory, developed in the 1960s, achieves this unification. The weak charged current interactions are invariant under SU(2) local phase transformations. This condition is satisfied by three gauge fields $W_\mu^{(1,2,3)}$ manifesting as three gauge




(a) (b)

Figure 2.7: Feynman diagrams showing (a) a fully hadronic interaction and (b) the semi-leptonic $\beta^-$-decay.

bosons $W^{(1)}, W^{(2)}$ as well as $W^{(3)}$ [7]. The two physical charged bosons $W^+$ and $W^-$ arise as a linear combination of $W_\mu^{(1)}$ and $W_\mu^{(2)}$:

$$W^+{}_\mu = \frac{1}{\sqrt{2}}\left(W_\mu^{(1)} - iW_\mu^{(2)}\right), \tag{2.9}$$

$$W^-{}_\mu = \frac{1}{\sqrt{2}}\left(W_\mu^{(1)} + iW_\mu^{(2)}\right). \tag{2.10}$$

The neutral $Z$-boson, on the other hand, does not simply correspond to the leftover $W^{(3)}$-boson. Instead, the U(1) symmetry of QED is replaced by a new $\mathrm{U}(1)_Y$ symmetry, here the subscript $Y$ denotes the weak hypercharge. The gauge field $B_\mu$, arising from this new symmetry, can be linearly combined with $W_\mu^{(3)}$ to define the field $A_\mu$ associated with photons as well as $Z_\mu$ linked to the physical $Z$-boson:

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^{(3)} \end{pmatrix}, \tag{2.11}$$

where $\theta_W$ is the Weinberg or weak mixing angle, whose value has to be determined experimentally. This angle offers a link between the coupling constants of all three interactions:

$$e = g_W \sin\theta_W = g' \cos\theta_W, \tag{2.12}$$

with $g_W$ and $g'$ being associated with charged and neutral current vertices, respectively. The elementary charge is directly connected to the fine-structure constant $\alpha = e^2/4\pi$.

A unification of the electroweak and strong interactions is the subject of current research. Various models, such as the Georgi-Glashow model [32], propose solutions to this problem, but so far none has been verified experimentally.

### 2.1.3 Higgs mechanism

In the theory of electroweak interactions, $W$- and $Z$-bosons must be massless particles, otherwise local gauge invariance, one of the Standard Model's fundamental pillars, is violated. However, experiments have clearly proven that both particles do in fact have a non-zero mass (exact values are given in Figure 2.1). A solution to this problem is offered by the Higgs mechanism [33] through explaining the bosons' masses without touching the model's local gauge invariance.

Giving a short introduction to the Higgs mechanism requires introducing at least some basic concepts needed to describe continuous systems. While those are discussed in the following, reference must be made again to [7], which strongly influenced this section by providing an ingenious introduction to the topic.

In general, in addition to Newton's second law $\vec{F} = m\vec{a}$, the dynamics of a system can also be described using the Lagrange equation:

$$L = T - V\,, \tag{2.13}$$

in which $T$ and $V$ characterise kinetic and potential energy, respectively. The equations of motion can subsequently be obtained by the Euler-Lagrange equation:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{r}_i}\right) - \frac{\partial L}{\partial r_i} = 0\,. \tag{2.14}$$

For a continuous system, the generalised coordinates are replaced by fields, resulting in Lagrangian $L$ being replaced by the Lagrangian density $\mathcal{L}$:

$$L = \int \mathcal{L}\mathrm{d}^3 r\,. \tag{2.15}$$

In the QFT framework, generalised coordinates are replaced by the excitations of a quantum field, which itself satisfies the field equations. Ultimately, all interactions in the Standard Model can be represented by such a Lagrangian. According to Noether's theorem [34], a conserved quantity results from a Lagrangian that is invariant with respect to a specific transformation (such as angular momentum conservation emerges from invariance with respect to rotations).

As a simplified example, the Higgs mechanism can be illustrated via a violation of the local U(1) symmetry group. To begin with, the Lagrangian density,

$$\mathcal{L} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi\,, \tag{2.16}$$

for a spinor field $\psi$, fulfilling the Dirac equation, must be invariant under local phase transformations such as $\psi \rightarrow \psi' = \exp(iq\chi)\,\psi$. Here $\gamma^\mu$ denotes the Dirac matrices. In the given form this is not the case and one obtains

$$\mathcal{L} \rightarrow \mathcal{L}' = \mathcal{L} - q\bar{\psi}\gamma^\mu\left(\partial_\mu\chi\right)\psi \neq \mathcal{L}\,, \tag{2.17}$$

after applying the transformation.

To retain invariance under local phase transformations, $\partial_\mu$ is replaced by $\partial_\mu + iqA_\mu$. The newly introduced field behaves as $A_\mu \rightarrow A'_\mu = A_\mu - \partial_\mu\chi$ under the transformations. This addition yields

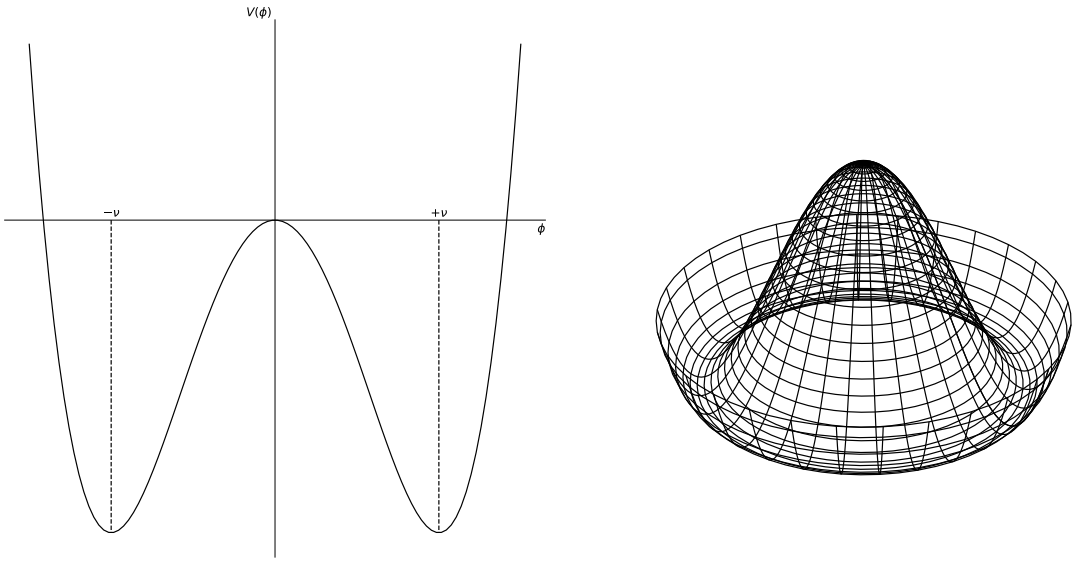
the invariant form of the Lagrangian density:

$$\mathcal{L} = \bar{\psi}\left(i\gamma^{\mu}\partial_{\mu} - m\right)\psi - q\bar{\psi}\gamma^{\mu}A_{\mu}\psi\,. \tag{2.18}$$

In case of electromagnetic interactions, $q = -e$ applies and the massless photon can be associated to the new field $A_{\mu}$. Adding a mass term $1/2 m_{\gamma}^2 A_{\mu}A^{\mu}$ to this Lagrangian immediately breaks the previously acquired invariance, as it transforms to

$$\frac{1}{2}m_{\gamma}^2 A_{\mu}A^{\mu} \rightarrow \frac{1}{2}m_{\gamma}^2\left(A_{\mu} - \partial_{\mu}\chi\right)\left(A^{\mu} - \partial^{\mu}A^{\mu}\right)\,, \tag{2.19}$$

which is obviously different from the original mass term.




(a) (b)

Figure 2.8: Potential $V(\phi)$ illustrated for (a) the one-dimensional and (b) the two-dimensional case. Both vacuum states $\pm\nu$ marked in the one-dimensional case.

For photons and gluons, albeit in the latter case for the SU(3) group, this poses no problem, as both are massless. In contrast, $W$- and $Z$-bosons need such a mass term and therefore violate the model's important invariance.

At this point, the idea of spontaneous symmetry breaking in the Higgs mechanism comes into play. Considering the general Lagrangian for a simple scalar field $\phi$ with the potential $V = 1/2\mu^2\phi^4 + 1/4\lambda\phi^4$ (shown in Figure 2.8(a)), the result is

$$\mathcal{L} = \frac{1}{2}\left(\partial_{\mu}\phi\right)\left(\partial^{\mu}\phi\right) - \left(\frac{1}{2}\mu^2\phi^2 + \frac{1}{4}\lambda\phi^4\,.\right)\,, \tag{2.20}$$

with $\lambda > 0$ but $\mu^2 < 0$. The vacuum state, defined as the field's lowest energy state, lies directly at the minimum of the potential, i.e. in this case at $\phi = \pm\nu = \pm\sqrt{-\mu^2/\lambda}$. Ending up in one of these two solutions spontaneously breaks the system's symmetry since from this point it loses its original

axial symmetry. The same process is also recognisable for the two-dimensional potential shown in Figure 2.8(b). Expanding around the positive minimum, hence using $\phi = v + \eta$, results in

$$\mathcal{L} = \frac{1}{2}\left(\partial_\mu \eta\right)\left(\partial^\mu \eta\right) - \lambda v^2 \eta^2 - \lambda v \eta^3 - \frac{1}{4}\lambda\left(\eta^4 - v^4\right), \tag{2.21}$$

where $\eta$ can be seen as the field's excitations along the $x$-axis of the potential. The second term in Equation (2.21), proportional to $\eta^2$, can be identified as a mass term $\lambda v^2 = {}^{m^2}/_2$. Should the reader seek a deeper understanding, the same derivation in a more detailed form combined with an additional example using a general complex field can be found in [7].

Inserting this mechanism in the form of a complex scalar field $\phi = {}^1/\sqrt{2}(\phi_1 + i\phi_2)$, with the same potential as in the previous example, into the U(1) local gauge symmetry first requires replacing $\partial_\mu$ by $D_\mu = \partial_\mu + igB_\mu$ analogously to Equation (2.18). The added field $B_\mu$ transforms in the same way as $A_\mu$. Combining the kinetic and potential terms in the full Lagrangian and expanding the complex field around $v$ using $\phi = {}^1/\sqrt{2}\,(v + \eta + i\delta)$ results in

$$\begin{aligned}\mathcal{L} = {} & \frac{1}{2}\left(\partial_\mu \eta\right)\left(\partial^\mu \eta\right) - \lambda v^2 \eta^2 \\ & + \frac{1}{2}\left(\partial_\mu \delta\right)\left(\partial^\mu \delta\right) - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}g^2 v^2 B_\mu B^\mu \\ & - V_{\text{int}} + gvB_\mu\left(\partial^\mu \delta\right).\end{aligned} \tag{2.22}$$

Once again, a massive scalar field $\eta$ is obtained in the form of the first two terms, additionally the third term can be interpreted as a massless boson $\delta$, also known as a Goldstone boson[14]. The kinetic term of the field $B_\mu$ is given through $F_{\mu\nu}F^{\mu\nu}$ while ${}^1/_2 g^2 v^2 B_\mu B^\mu$ corresponds to the mass term of the now massive gauge boson. The fields' self-interactions are represented by $V_{\text{int}}$. Thus it is evident that the previously massless field $B_\mu$ receives a mass term after the spontaneous symmetry breaking while the complete Lagrangian still retains the local gauge invariance under phase transformations.

In this example, only the U(1) local gauge invariance was considered, whereas in the electroweak sector of the Standard Model, the U(1) × SU(2) gauge invariance must apply. The minimal Higgs model satisfying these requirements are two complex scalar fields:

$$\phi = \frac{1}{\sqrt{2}}\begin{pmatrix}\phi_1 + i\phi_2 \\ \phi_3 + i\phi 4\end{pmatrix} \tag{2.23}$$

A completely analogous consideration, taking into account the changed symmetry requirements, results in the boson's mass terms $m_W = {}^1/_2 g_W v$ and $m_Z = {}^1/_2 v\sqrt{g_W^2 + g'^2}$ with $v = 246\,\text{GeV}$. The coupling between the Higgs and both electroweak gauge bosons $g_{\text{HVV}}$ is directly proportional to those mass terms. In a somewhat similar way, the Higgs mechanism can further explain the existence, albeit not the exact values, of fermion masses in the Standard Model. Their coupling to the Higgs field, called the Yukawa coupling, is given by $g_f = \sqrt{2}m_f/v$, obviously increasing linearly with the fermion mass $m_f$.

The actual physical scalar Higgs-boson $H$, associated with the massive scalar field and denoted by a dashed line in Feynman diagrams, could not be experimentally observed until 2012 (as described in Chapter 1) and represents, as of now, the last addition to the Standard Model. Precisely measuring the

[14] The field along with the associated Goldstone boson can be eliminated by applying a suitable gauge transformation [7].

Higgs-boson and its quantities is currently still one of the key areas of research in particle physics.

### 2.1.4 Standard Model's flaws and physics beyond

On the one hand, the Standard Model has successfully provided precise predictions for experimental observations over decades. On the other hand, this should not veil the fact that the model is far from being complete or even perfect.

Initially, the Standard Model depends on a large number of parameters that cannot be predicted but have to be determined experimentally. These include all fermion masses, the values of the CKM matrix, shown in Equation (2.8), as well as the exact values of the interactions' coupling constants. Furthermore, both inputs needed to describe the Higgs potential, i.e. $\mu$ and $\lambda$, are also neither predetermined by the theoretical foundation.

In addition, as it stands today, a number of observed phenomena cannot be explained by the model at all. For instance, it is still unclear why the known part of the universe is formed almost exclusively by particles instead of antiparticles. This asymmetry is not predicted by the Standard Model and could either have been an initial condition of the universe or evolved over the course of time. Already in the late 1960s, Russian physicist Andrei Sakharov established conditions that must be fulfilled in the creation of the universe to cause such an asymmetry [35]. The actual cause as well as the exact mechanism, however, remain a mystery to this day.

A concluding phenomenon to be mentioned is the observation of dark matter. This matter, of hitherto unknown origin, is detectable through gravitational effects as its existence is needed to explain the speed at which stars rotate around their galaxy centres. According to current assumptions, it amounts to a significantly larger fraction than the matter particles known to us. So far, none of the models put forward to explain this matter have stood up to an actual experimental test. One hypothesis concerning the existence of new gauge interactions which serve as a link between a dark sector and the Standard Model may be tested at Bonn's accelerator facility ELSA in the near future [36].

For all or at least a large number of these and other unsolved questions, various theories offering possible solutions are available. One aim of current research is to validate these theories, partially via the direct search for new, potentially predicted, particles. Secondly also, through increasingly precise measurements of the Standard Model's free parameters.

So far, no new particle has been discovered after the Higgs-boson, making the latter precision measurements ever more interesting. To take up the example used in the introduction: each of these measurements provides a better understanding of what the physical jigsaw puzzle's open areas look like. Constraining any parameter helps at deciding which puzzle piece in the form of a theory either might be able to fill a gap in the model or will never fit into the existing puzzle and can consequently be discarded along with its theory.

## 2.2 Physics of proton-proton collisions

The dataset analysed in this thesis was produced at the Large Hadron Collider (LHC), a proton-proton collider, described in Section 3.2. Therefore, sections Section 2.2.1 and Section 2.2.2 shall serve as an introduction to the relevant physical concepts of proton-proton collisions. This is followed by Section 2.2.3 which summarises properties of all key particles present in the subsequent analysis.

First of all, in contrast to an electron-positron collider, no elementary particles are brought to

collision, but instead hadrons composed of quarks and gluons. This leads to an initial state which is not well defined, as we are actually observing interactions of these constituents, described in the quark-parton model, in high energy deep inelastic scattering experiments

For this purpose, it is not quite adequate to think of the proton as a formation of just two up- and one down-quark. These three quarks, also known as valence quarks, may be sufficient to explain the proton's quantum numbers but do not fully characterise its inner kinematics. In reality, the three valence quarks permanently exchange virtual gluons, which can additionally annihilate into virtual $q\overline{q}$ pairs, referred to as sea quarks. This internal process is visualised in Figure 2.9. Experimental and theoretical consequences arising from this substructure are explained in the following.

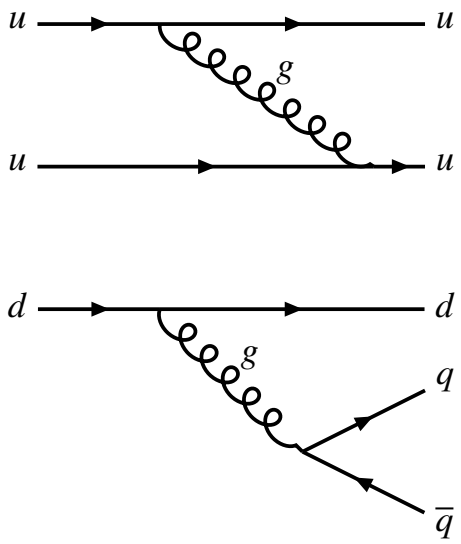



Figure 2.9: Production of virtual $q\overline{q}$ pairs and virtual gluon exchange inside the proton.

### 2.2.1 Parton distribution functions

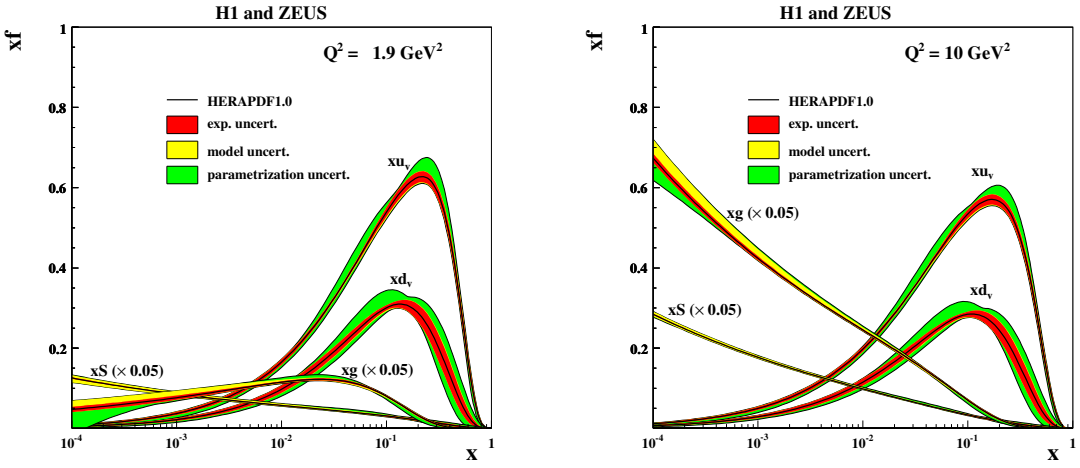
In collisions with sufficiently high momentum transfer, interactions between both protons' partons occur at length scales smaller than the characteristic confinement scale. In this case, $\alpha_S \ll 1$ applies and the partons can be considered to exist as quasi-free particles [37].

As a result, the proton's total momentum is distributed among its partons, whereby the generated sea quarks typically carry lower fractions compared to the valence quarks. A precise and correct description of these momentum distributions is crucial for any measurement involving hard hadron collisions and is obtained by parton distribution functions $f(x)$. These functions represent the probability density of finding a parton, carrying the fraction $x$ of the total proton momentum. Ultimately, they do not only depend on the momentum fraction $x$ but also on the particular energy scale, usually expressed through $Q^2$. If we consider $q^2 = -Q^2$ as the momentum squared of a virtual photon, a graphic explanation for this dependency emerges. A photon with a relatively low $q$ is not able to resolve the proton's substructure[15]. A high-energy photon, on the other hand, easily resolves the inner structure and thus the sea quarks with their small values of $x$ have a greater impact [7].

[15] Resolution increases with smaller wavelengths, i.e. greater energies.

While the actual dependency on $Q^2$ can be calculated in the QCD framework, the exact PDF itself cannot be predicted theoretically[16] but must be determined experimentally. One example of resulting PDFs for two different values of $Q^2$ from the H1 and ZEUS experiments at the HERA accelerator is shown in Figure 2.10. The discussed scaling of all PDFs with $Q^2$ is clearly visible.

An important distinction needs to be mentioned here, depending on whether the bottom-quark is described by a sea quark PDF or not. When this is the case, one refers to this as the five flavour scheme (5FS). If, on the other hand, the bottom-quark is assumed to be a real massive particle, only the four lighter quarks (4FS) are included in the distribution. In this view, the bottom-quark can only enter collisions through a gluon splitting.




(a) (b)

Figure 2.10: The PDFs for valence quarks ($xu_v$ and $xd_v$), sea quarks ($xS$) and gluons ($xg$) of HERAPDF1.0 at (a) $Q^2 = 1.9\,\mathrm{GeV}^2$ and (b) $Q^2 = 10\,\mathrm{GeV}^2$. The gluon and sea distributions are scaled down by a factor of 20. Taken from [39].

### 2.2.2 Hadronisation

If $q\bar{q}$ pairs are produced in high energy collisions, another phenomenon relevant to the comprehension of this thesis occurs: due to the running of of $\alpha_S$, quarks cannot exist freely as coloured objects. Therefore, separating a $q\bar{q}$ pair leads to an increase in the field energy stored between them. Once this energy is sufficiently high, energetically more favourable additional $q\bar{q}$ pairs are formed. This process continues until all the resulting quarks reach low energies and form colourless hadrons in a process termed hadronisation. These newly formed mesons and baryons are the particles ultimately visible in the experiment [40]. The signature of this entire process then resembles a narrow cone of hadrons, known as a particle jet or simply jet.

The bottom-quark and the associated $b$-jets play a special role in this context. Due to the fairly long lifetime of weakly decaying $B$-hadrons, a $b$-jet's point of origin is often located relatively far away

[16] Recent publications show that a calculation via lattice QCD is possible [38].

from the initial collision. With appropriate resolution, this can be utilised experimentally to identify these special jets, as described in Section 3.5.

### 2.2.3 Physics of key particles

This section serves as a introduction to properties of all particles relevant in the presented analysis. Starting with the top-quark possible production and decay channels are described.

#### Top-quark physics

With a mass of $m_t = (172.57 \pm 0.29)$ GeV [10], the top-quark is not only by far the heaviest quark but also the heaviest fundamental particle in the Standard Model. As already mentioned in Section 2.1.1, it possesses a lifetime of $\tau \approx 5 \times 10^{-25}$ s [10], shorter than the typical time scale required to form hadronic states. Consequently, no $t\bar{t}$ bound states can be observed leaving the top-quark in some sense as a free particle [37].

In proton-proton collisions, given sufficient energies, it can be produced either in $t\bar{t}$ pairs or alternatively as a single top-quark. The dominant mechanism for $t\bar{t}$ pair production is gluon-gluon fusion (Figure 2.11(a)) whereas a single top-quark is mostly produced in the $t$-channel[17] (Figure 2.11(b)) [41].

Due to the almost unitary CKM matrix in combination with the fulfilled condition $m_t > m_b + m_W$, the top-quark decays weakly into a bottom-quark as well as a real $W$-boson in $(95.7 \pm 3.4)$ % [10] of all cases. The Feynman diagrams for these decays are depicted in Figure 2.12.

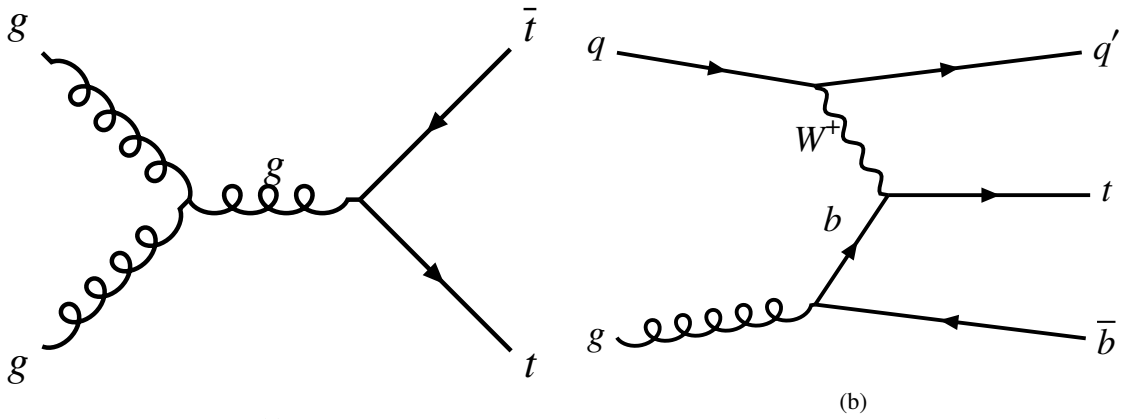
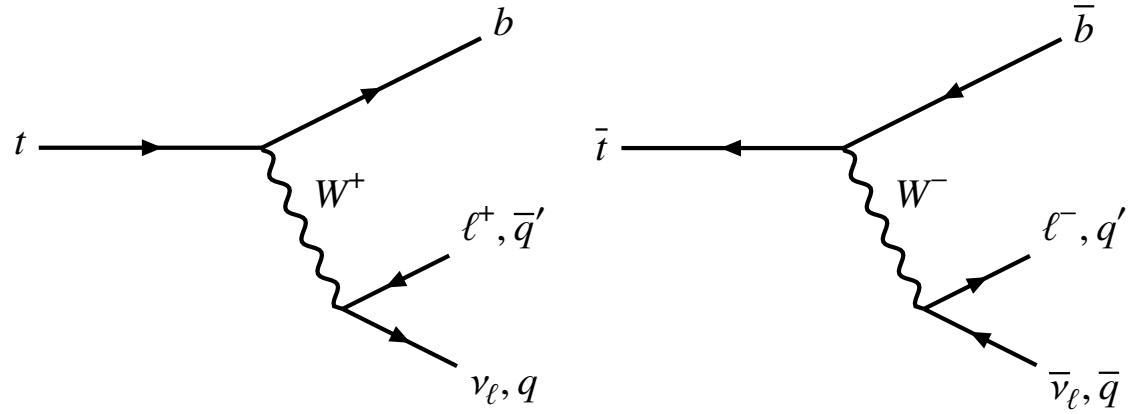



Figure 2.11: Dominant production modes for (a) $t\bar{t}$ pairs through gluon-gluon fusion and (b) single top-quark $t$-channel production.

#### Higgs physics

The Higgs-boson with its mass of $(125.20 \pm 0.11)$ GeV [10] couples to all massive elementary particles, regardless of whether they are fermions or bosons. In proton-proton collisions it is mainly produced

[17] The term $t$-channel denotes a certain representation of Feynman diagrams, closely linked to the Mandelstam variable $t$.




(a) (b)

Figure 2.12: Dominant decay mode for (a) the top-quark ($t$) and (b) the antitop-quark ($\bar{t}$). Pure leptonic $W$-boson decays are shown in both cases.

Table 2.1: All currently discovered decay modes of the Higgs-boson with a branching fraction above one percent. The decay modes $H \to WW$ and $H \to ZZ$ must contain one virtual boson. Values taken from [10].

| Decay mode | Branching fraction [%] |
|---|---|
| $H \to b\bar{b}$ | $53 \pm 8$ |
| $H \to WW$ | $25.7 \pm 2.5$ |
| $H \to \tau^-\tau^+$ | $6.0 \pm 0.8$ |
| $H \to ZZ$ | $2.8 \pm 0.3$ |

by four different processes: gluon-gluon fusion, weak-boson fusion, Higgs-Strahlung as well as the associated production with top-quarks [42]. All associated LO Feynman diagrams are displayed in Figure 2.13.

As the coupling strength between the Higgs and another particle rises proportional to the particle's mass, decays into heavier particles are generally favoured. However, since the decay into a pair of top-quarks is kinematically forbidden, the decay into a bottom-quark pair dominates, followed by the decay into a pair of $W$-bosons. In addition to those direct decays, internal loops also allow decay modes into massless particles, both options are shown in Figure 2.14. A listing of the dominant modes' respective branching ratios is given in Table 2.1.

### Tau-lepton physics

The $\tau$-lepton, or simply tau ($\tau$), is the heaviest lepton in the model and, with its mass of $(1\,776.93 \pm 0.09)$ MeV [10], approximately 3 500 times heavier than the electron. While in principle regarded as a heavy copy of the lighter two charged leptons, its greater mass opens up additional decay channels. Thus, the tau-lepton cannot only decay purely leptonically into its two lighter counterparts but in addition hadronic decay modes via the weak interaction exist, both shown in Figure 2.15. As a consequence, the signature of a tau-lepton can be significantly more complex than

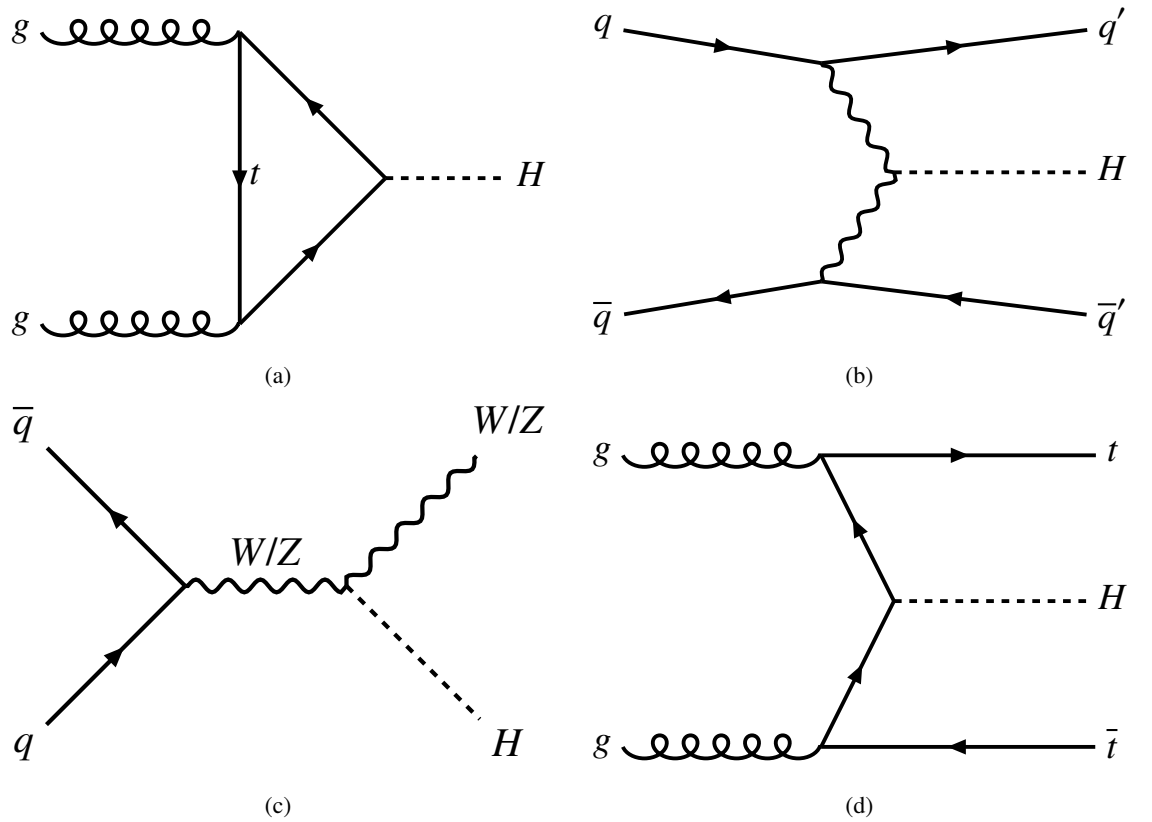



Figure 2.13: Dominant Feynman diagrams contributing to the Higgs-boson production through (a) gluon-gluon fusion, (b) weak-boson fusion, (c) Higgs-Strahlung as well as (d) the associated production with top-quarks.

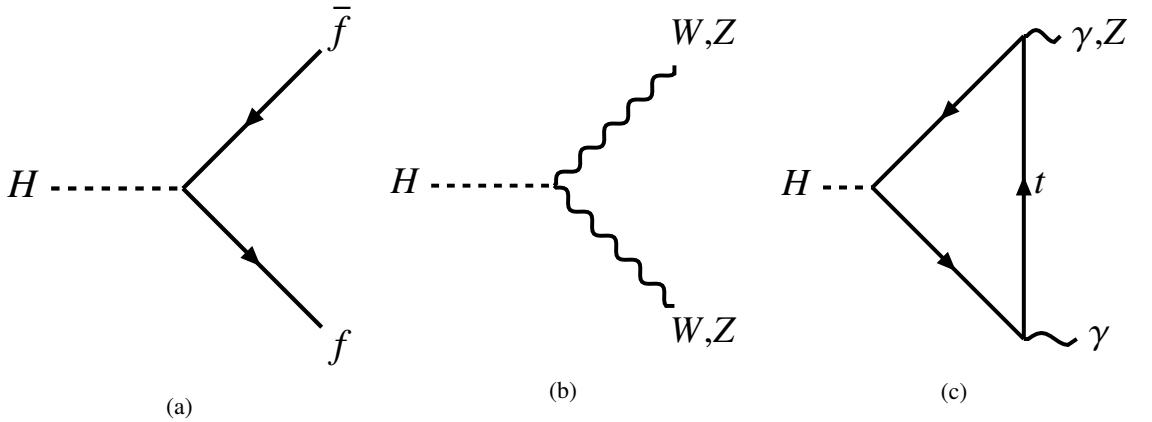
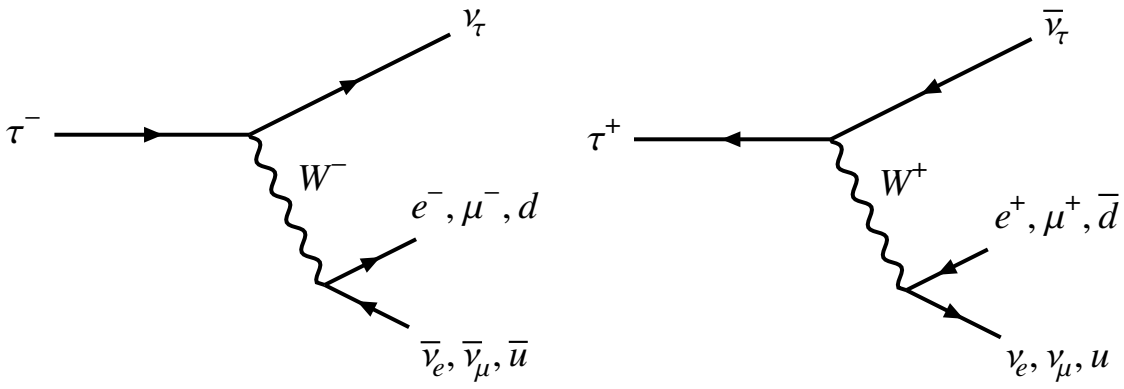



Figure 2.14: Higgs-boson decay modes into (a) a pair of fermions including bottom-quarks and all lighter particles, (b) a pair of weak bosons and (c) one example for loop induced decays.

that of an electron or muon (for more detailed information on the detection of those particles, see Section 3.5.3). In the following chapters, the latter two will be referred to as "charged" light leptons or simply light leptons[18] to avoid any confusion.




(a) (b)

Figure 2.15: Leptonic and hadronic decay modes of (a) the tau-lepton ($\tau$) and (b) the antitau-lepton ($\tau^+$).

## 2.3 Associated production of a single top-quark and a Higgs-boson

As already mentioned at the end of Section 2.1.3, precisely measuring all properties of the Higgs-boson currently still represents one of the most interesting areas of research in particle physics. In this thesis one of these parameters, namely the Yukawa coupling (see section Section 2.1.3) between the Higgs-boson and the top-quark $y_t$, is investigated. This coupling can be measured via different decay and production channels, obviously always involving the Higgs-boson as well as the top-quark.

An indirect determination is possible by analysing the Higgs-boson production through a top-quark loop, in which it is necessary to assume that no other beyond Standard Model particles contribute to the loop. A direct measurement, on the other hand, is possible via the associated production of a top-quark pair and a Higgs-boson, shown in Figure 2.18(a), a channel recently discovered [43].

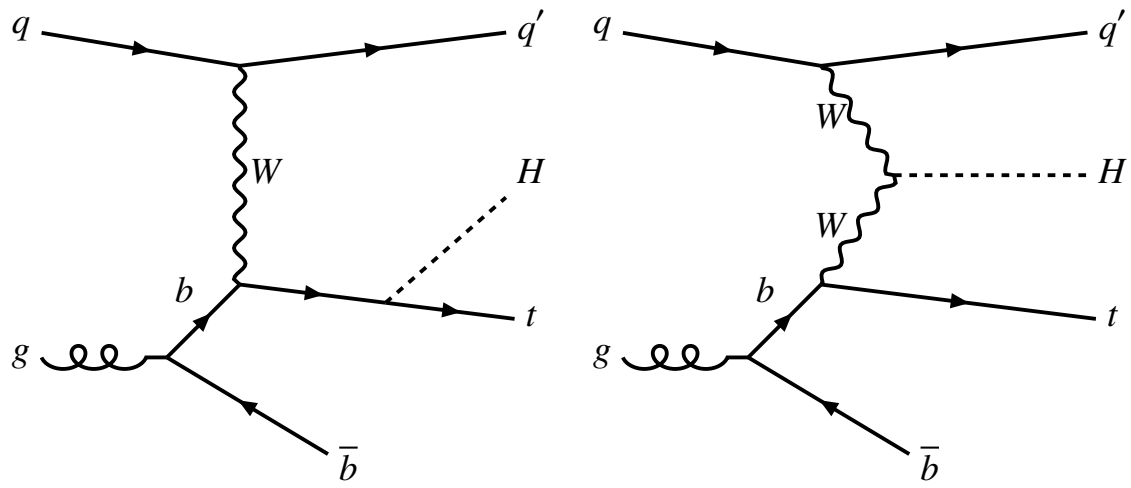
Of great interest, however, is not only the magnitude $y_t$ itself but in addition also the coupling's relative sign with respect to the gauge coupling $g_{HVV}$ between the Higgs and the massive $W$- and $Z$-bosons. Information about this relative sign can neither be obtained from the aforementioned top-quark loop nor from the associated production of a top-quark pair and a Higgs-boson. Instead, the decay of the Higgs into a pair of photons can be utilised. These measurements, although again relying on the assumption of the absence of beyond Standard Model particles, are in agreement with the Standard Model by favouring a positive sign of the Yukawa coupling [44, 45].

On the other hand, as presented in this thesis, a cross-section measurement of the associated production of a single top-quark and a Higgs-boson ($tH$) allows a direct determination of the Yukawa coupling's relative sign and magnitude. Both quantities can be extracted by exploiting interferences even in the presence of beyond Standard Model particles. The three lowest order production modes

[18] Strictly speaking, light leptons of course also include neutrinos. In this thesis, however, only electrons and muons are referred to as such.

of the $tH$ process in proton-proton collisions at $\sqrt{s} = 13$ TeV are $tHq$ (Figure 2.16) with a predicted cross-section of 74.2 fb, $tWH$ (Figure 2.17(a)) with 15.2 fb and the $s$-channel[19] (Figure 2.17(b)) with 2.9 fb [46]. In this work, the two dominant channels for proton-proton collisions, i.e. $tHq$ and $tWH$, are investigated. In both cases, an interference between a coupling of the Higgs to the top-quark and to the $W$-boson enables a determination of $y_t$'s sign. These coupling options are exemplarily shown for $tHq$ in Figure 2.16(a) and Figure 2.16(b). The much smaller contributing $s$-channel is ignored, which is why $tH$ is used synonymously for just the two dominant channels throughout this thesis.

The interference between the two diagrams shown in Figure 2.16, as predicted by the Standard Model, is maximally destructive, leading to a small cross-section. In contrast, measuring a constructive interference with its enhanced cross-section corresponds to the coupling constant's relative sign being negative [47]. The most recent publication by the CMS collaboration on investigating $tH$ in combination with top-quark pairs measured a production rate corresponding to $5.7 \pm 2.7$ (stat.) $\pm$ 3.0 (syst.) of its Standard Model prediction. The associated observed (expected) significance amounts to 1.4 (0.3) [48].




(a) (b)

Figure 2.16: Feynman diagrams for the $tHq$ production in the 4FS where the Higgs-boson is radiated off (a) the top-quark or (b) the $W$-boson.

### 2.3.1 Analysis channels

The $tH$ channel allows a split up into several different analysis channels depending on the Higgs-boson's decay (as listed in Table 2.1). This thesis focuses on Higgs-boson decays into two tau-leptons ($H \rightarrow \tau^-\tau^+$). Multi-lepton decay modes combined with decays into a pair of bottom-quarks have been analysed simultaneously. Their final results can be found in [49]. In addition, a light lepton is expected in all channels, as events are searched for, in which the $W$-boson, emitted in the weak decay of the top-quark into a bottom-quark, decays leptonically. The analysed decay channels are further

[19] The term $s$-channel describes a certain representation of Feynman diagrams similar to the already mentioned $t$-channel. In $s$-channel interactions, two particles annihilate into an exchange particle which again decays into two particles. An example would be electron-position annihilation into a photon, subsequently undergoing the process of pair production.

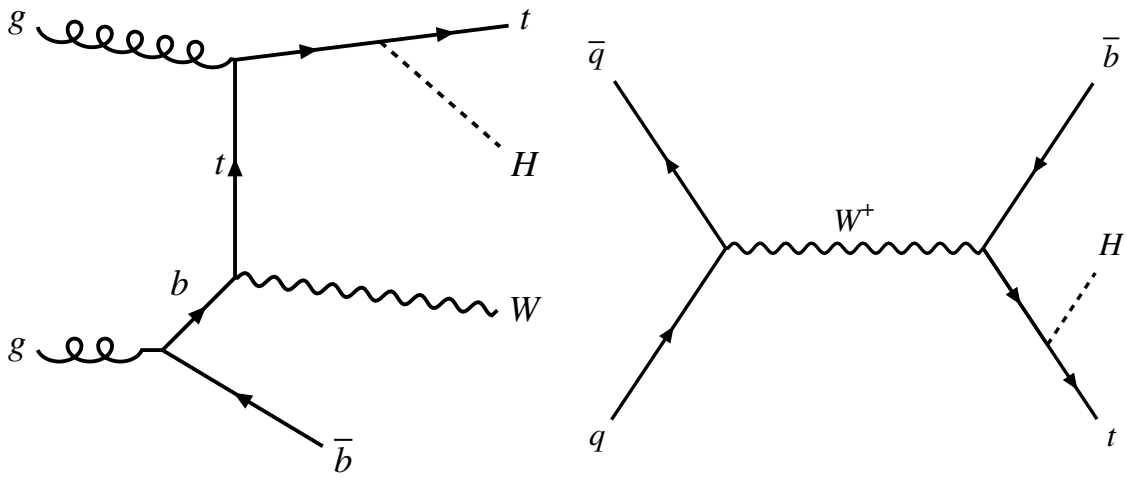



Figure 2.17: Feynman diagrams for (a) $tWH$ in the 4FS and (b) the $s$-channel process.

subdivided according to the tau-lepton decay (discussed in Section 2.2.3). If both tau-leptons decay hadronically ($\tau_{\text{had}}$), the channel is referred to as $1\ell + 2\tau_{\text{had}}$ in the following sections of this thesis. In this context $1\ell$ denotes the single involved light lepton originating from the $W$-boson decay.

If, on the other hand, one tau-lepton undergoes a leptonic decay, the channel is labelled $2\ell + 1\tau_{\text{had}}$. The two light leptons present in this case stem from the leptonic decay of the tau-lepton and, as in the $1\ell + 2\tau_{\text{had}}$ channel, from the $W$-boson.

### 2.3.2 Background processes

The entire analysis is complicated by a large number of background processes, causing the analysis to feel like searching a needle in the haystack[20], especially due to the small expected $tH$ cross-section. A precise understanding of these background processes influencing the analysis is of great importance in most particle physics analysis. Otherwise, an optimal background reduction, needed to obtain the highest significance in a measurement, cannot be achieved. In general, all processes able to mimic the $tH$ signature, consisting of a $b$-jet, two signatures similar to the tau-lepton as well as at least one light lepton, need to be considered.

The background processes arising from these criteria are first of all composed of channels mirroring $tH$ through a similar albeit not identical final state in the experiment. Their separation can be further complicated by detector uncertainties alongside incorrect identification of particles, termed fakes. Thus, although the $t\bar{t}$ process (shown in Figure 2.18(g)) is fundamentally different, it can fake the $tH$ signature through misidentified particles.

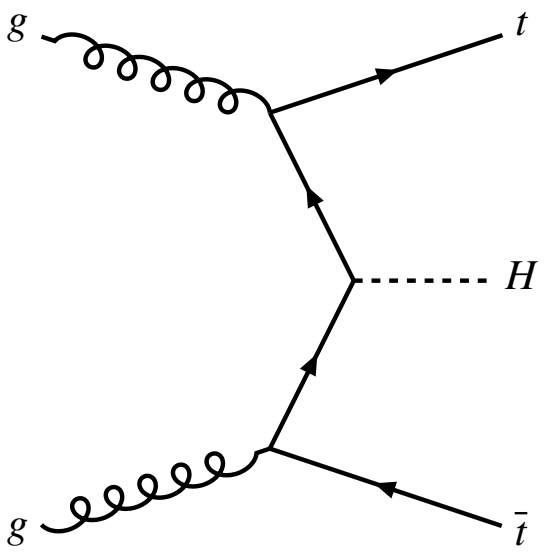
Moreover, background events can originate from processes whose final states contain the $tH$ final state as a subset. Obvious candidates for this scenario are the $tZq$ and $tWZ$ processes (shown in Figure 2.18(d) and Figure 2.18(c)) which transform into $tH$ by replacing the $Z$-boson with a Higgs-boson. Although both bosons have different masses, changing the kinematics, each can decay

[20] It is worth taking a look into [50] to see how close this frequently used metaphor is really related to the actual problem.

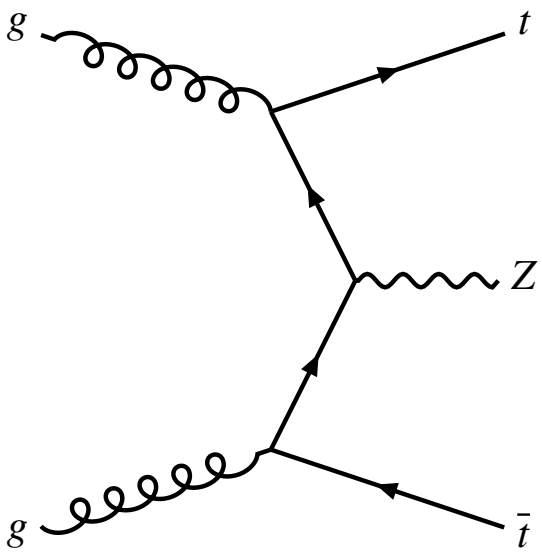
into a pair of tau-leptons and thus $tZq$ and $tH$ might produce the same final state visible in the experiment.

However, the extent to which a process obscures the signal process depends not only on the existence of similar final states but equally on the process' cross-section. As a consequence, $t\bar{t}$ with a relatively high cross-section ultimately plays a much larger role compared to $tZq$.

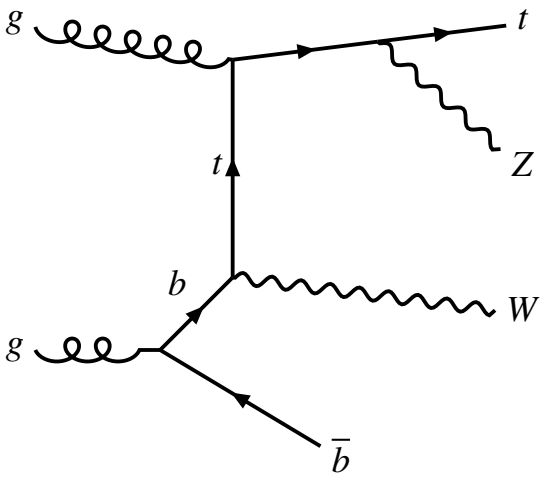
Example Feynman diagrams of the dominant and thereby relevant backgrounds for the analysis presented in this thesis are summarised in Figure 2.18. Other minor processes are not shown individually here and are instead briefly discussed in Section 6.2 and Chapter 7 along with the presentation of a pre-selection suitable to reduce the overall background contamination.



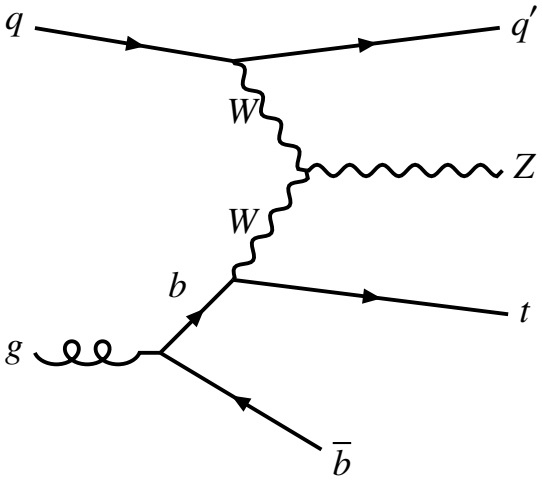

(a) $t\bar{t}H$.



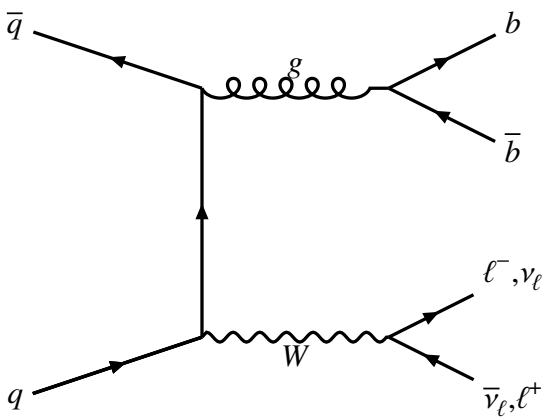

(b) $t\bar{t}Z$.



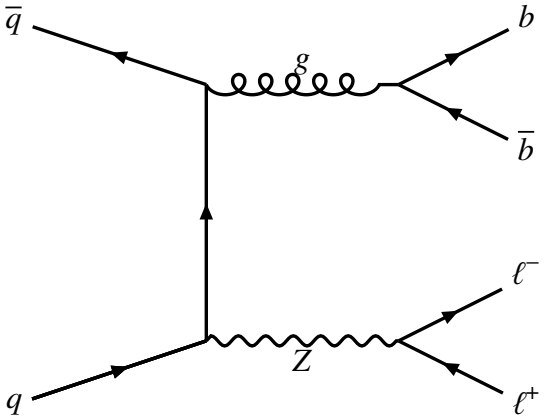

(c) $tWZ$.



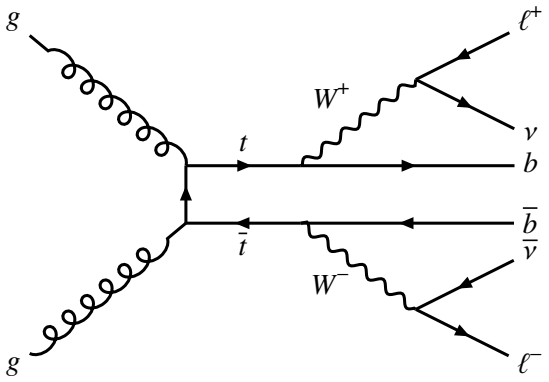

(d) $tZq$.



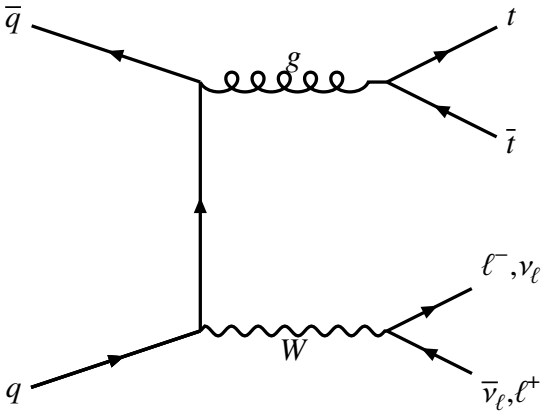

(e) $W$ + jets.



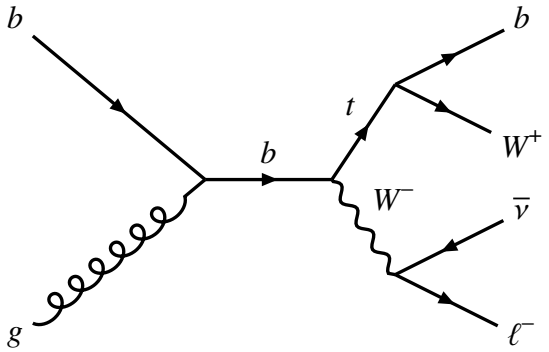

(f) $Z$ + jets.



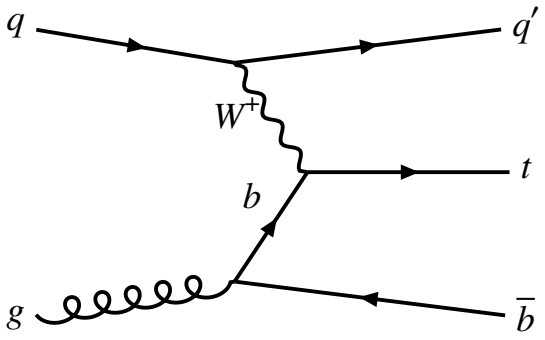

(g) $t\bar{t}$.



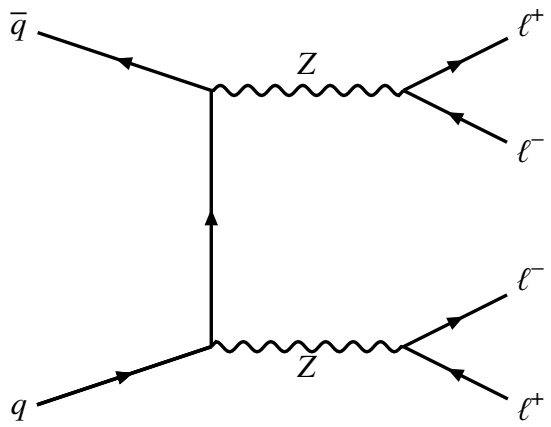

(h) $t\bar{t}W$.




(i) $tW$.




(j) Single top-quark $t$-channel.




(k) Diboson.

Figure 2.18: Lowest order Feynman diagrams of all relevant background processes in the $tH$ analysis.

CHAPTER 3

# Experimental setup

As outlined in Chapter 2, the Standard Model was developed in a direct interplay between theory and experimental work. Thus, after having already introduced the theoretical basis it is natural to continue with explaining the experimental setup, used to generate the analysed dataset. As in most high-energy particle physics experiments, an accelerator is first used to propel particles, protons in this case, to nearly the speed of light. The actual data acquisition then generally follows via a detector, located close to the accelerated particles' collision point.

Reflecting this structure, Section 3.1 covers all relevant aspects of particle acceleration. Section 3.2 capitalises on this information when introducing the accelerator used for data production in this thesis, the LHC. Afterwards, an overview of the basic mechanisms used to detect particles is presented in Section 3.3. Subsequently the ATLAS detector, responsible for the data taking, is covered in Section 3.4. Finally, identification of particles based on the detector's response alongside reconstruction methods for all key particles used in this analysis are discussed in Section 3.5.

Similar to Chapter 2, particle acceleration and detection form two thematic blocks whose detailed treatment would fill far more than just this thesis. Therefore, this chapter aims to provide sufficient information on both topics to facilitate understanding of the presented analysis. A detailed coverage is available in the following references: for accelerators, an introduction can be found in [51], the LHC in particular is explained in detail in [52]. The basics of detectors are described very comprehensively in [53], whereas for the ATLAS detector, please refer to [54].

## 3.1 Particle accelerators

The purpose of a particle accelerator is boosting charged particles up to extremely high energies via electromagnetic fields and ultimately colliding them in a highly focussed beam. The largest of these extremely complex machines are primarily intended to be used in fundamental research. Smaller facilities, on the other hand, are commonly employed in other areas, such as a radiation source in cancer therapy.

For research purposes, nowadays a distinction is commonly made between two types of particle accelerators: linear and circular accelerators. In both cases the particles pass through beam pipes containing a high vacuum to minimise the impact of collisions with gas molecules. As the name suggests, particles in linear accelerators travel along a linear acceleration path formed by radiofrequency

cavities. An electric field, contained inside them, reverses its polarity at a radio frequency level. When correctly timed, charged particles always feel a force in the forward direction, i.e. in the direction of their motion. To achieve this, particles do not enter a cavity in the form of a homogeneous beam but are instead split into several smaller sets, called bunches. In theory these bunches are centred around an ideally timed synchronous particle, perfectly aligned with the cavity's frequency. A slightly faster particle will reach the cavity earlier, i.e. before the field is completely reversed, therefore receiving less energy as it passes through. The opposite applies to a slower, less energetic particle. As a consequence, particles in one bunch oscillate longitudinally around this synchronous position.

As the construction of such an accelerator complex is often a trade-off between reaching the desired energy and the total cost, one disadvantage of any linear accelerator is immediately obvious: once a particle has reached the machine's end it must be brought to collision and cannot be accelerated a second time. The maximum achievable energy is thus, under the assumption of identical cavities, mostly dependent on the accelerator's length. Instead of having to extend the accelerator, it is obviously desirable to lower the cost by using the same acceleration path several times. This idea is realised by a specific type of circular accelerators known as synchrotrons. It is by no means perfectly circular but consists of one or several linear accelerators connected by arc sections.

In these arcs, strong magnetic fields deflect the beam to keep it on a closed path. The magnetic field strength must be adjusted according to the particle's kinetic energy while the resulting beam deflection causes synchrotron radiation. This electromagnetic radiation occurs whenever a charged particle is accelerated perpendicular to its original path. At relativistic speeds synchrotron radiation results in a significant energy loss inversely proportional to the fourth power of the particle's mass ($m^4$). In order to counteract corresponding losses, relatively heavy particles such as the proton are commonly accelerated in high energy circular accelerators, despite the investigation of their collisions being quite complex, as discussed in Section 2.2.

The energy ultimately available in collisions is characterised by the centre-of-mass energy

$$\sqrt{s} = \sqrt{\left(\sum_{i=1}^{2} E_i\right)^2 - \left(\sum_{i=1}^{2} \vec{p}_i\right)^2}\,, \tag{3.1}$$

with $E_i$ and $\vec{p}_i$ defined as the energies and momenta of the two colliding particles. In order to create a particle X in an event the centre-of-mass energy has to be at least as large as the particle's mass $m_\mathrm{X}$. When determining $\sqrt{s}$, a general distinction is usually made between fixed target experiments that consist of one stationary target being hit by an accelerated beam and colliding beam experiments where two accelerated particle beams are brought to collision. In the first case, the centre-of-mass energy available in a collision event, often just referred to as event, is

$$\sqrt{s} = \sqrt{\left(E_1 + m_2\right)^2 - \left(\vec{p}_1\right)^2} \approx \sqrt{2m_2E_1}\,, \tag{3.2}$$

with $E_2 = m_2$, $E_1 \gg m_{1,2}$ and $\vec{p}_2 = 0$. Assuming $E_{1,2} \gg m_{1,2}$, $\vec{p}_1 = -\vec{p}_2$ and $E_1 = E_2$ the centre-of-mass energy of a colliding beam experiment is in contrast given by

$$\sqrt{s} = \sqrt{\left(E_1 + E_1\right)^2 - \left(\vec{p}_1 - \vec{p}_1\right)^2} = 2E_1\,. \tag{3.3}$$

Comparing Equation (3.2) and Equation (3.3) clearly shows that higher energies are much simpler to

obtain in colliding beam experiments.

Although important, the highest obtainable energy is not the only decisive characteristic of an accelerator: as described in Section 2.1.2, the production and decay of particles is a statistical process. Thus, obtaining a high amount of collision events, corresponding to the data used in analyses, is crucial for any accelerator, especially in case of rare channels. This property is quantised by the luminosity L[1], which in case of a circular collider with two beams divided into $N_b$ bunches, is given by

$$\mathrm{L} = \frac{f N_\mathrm{b} n_1 n_2}{4\pi\sigma_x\sigma_y} \,. \tag{3.4}$$

Here $f$ is defined as the revolution frequency of the particles in the collider ring. Each bunch consists of a specific number of particles denoted by $n_{1,2}$. Their Gaussian distributed beam profile is described by the horizontal root-mean-square $\sigma_x$ and the vertical root-mean-square $\sigma_y$, respectively.

As the luminosity quantifies the number of interactions per area per time one often also uses the related integrated luminosity $\mathrm{L}_\mathrm{int}$ to define the interactions per area over the lifetime of an experiment through

$$\mathrm{L}_\mathrm{int} = \int \mathrm{L}\,\mathrm{d}t \,. \tag{3.5}$$

Together with the cross-section introduced in Section 2.1.2, the expected number of events $N$ associated with a certain process during the complete run time of an experiment is thus expressed by

$$N = \sigma \mathrm{L}_\mathrm{int} \,. \tag{3.6}$$

When trying to increase the luminosity, it is inevitable that at some point more than one collision per bunch crossing occurs. The average number of these nearly simultaneous collisions is referred to as pile-up. Pile-up can be subdivided into in-time pile-up caused by multiple collisions in one bunch crossing or out-of-time pile-up arising from collisions in previous or future bunch crossings. Both types complicate the detection and reconstruction of the most interesting inelastic collision. The latter is usually the collision with the highest energy, termed the hard scattering event. Its high energy enables the generation of the heaviest fundamental particles, whose properties are often targeted in analyses.

## 3.2 Large Hadron Collider

The accelerator at hand is the LHC, which, as it stands today, is the most powerful particle accelerator built by mankind. It is installed underground at the CERN research centre near Geneva at the Swiss-French border as illustrated in Figure 3.1.

CERN itself was founded in 1954 and has since become one of the largest international research centres for fundamental science. Originally intended to be a nuclear research centre, accelerator facilities to perform particle physics experiments were already installed during the starting years. This meant that in 1994, the year the LHC was approved, CERN was able to provide the ideal

[1] Luminosity is often indicated by either $L$ or $\mathcal{L}$. However, to avoid confusion with the Lagrangian and the Lagrangian density, L is used in this thesis.

infrastructure for a project of this magnitude. Not only could the existing facilities serve in the necessary pre-acceleration stages, but the tunnel of the LHC's predecessor LEP could also be reused. As an outcome the costs of civil engineering were greatly reduced.

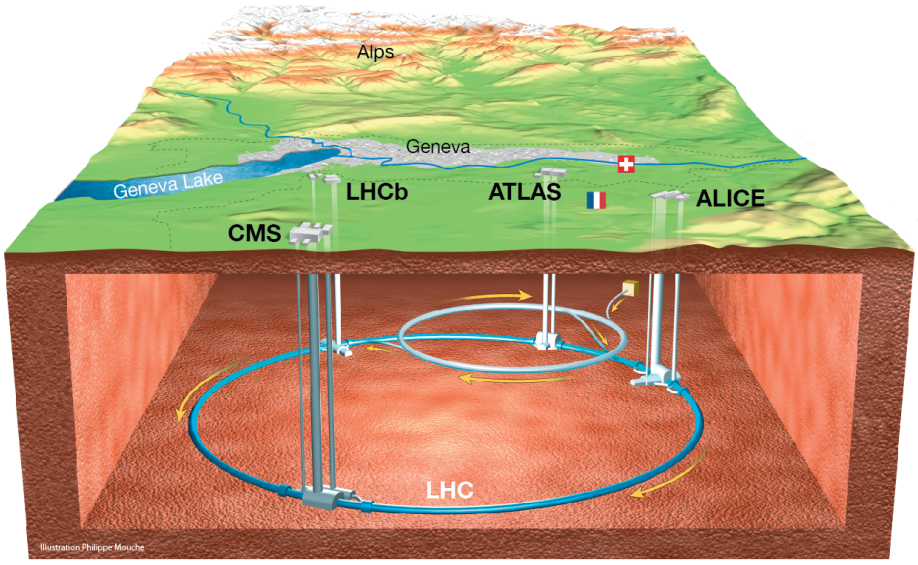



Figure 3.1: Schematic view on the subsurface installation of the LHC and its associated four large experiments near Geneva [55].

The LHC, commissioned in 2008, is designed to reach a centre-of-mass energy of $\sqrt{s} = 14\,\text{TeV}$ with a luminosity of the order of $L = 10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}$ when accelerating and colliding counter-rotating protons in high vacuum beam pipes, a combination unmatched by any other experiment [56]. A second operation mode additionally allows the acceleration of heavy ions.

In its current setup, the entire LHC can be divided into eight sections, each formed by an arc and a 500 m long straight section. While the straight sections are designed to fulfil different tasks, the arcs are almost identical. An arc represents a curved section to provide space for cells of dipole magnets, installed to bend the beam along its circular path. These magnets are operated in a superconducting mode to achieve a maximum B-field strength of 8.33 T [52]. In order to keep the beam focused and stable during this deflection, additional quadrupole magnets are included in each cell. Their force on a charged particle passing by increases linearly with the particle's spatial distance to the ideal orbit. Similar quadrupole magnets are also placed along the straight sections to maintain the beam quality.

Only one of the eight straight section provides space for the accelerator's radiofrequency cavities. These cavities, thus only taking up a relatively small part of the entire machine, are operated in a superconducting mode at a frequency[2] of 400 MHz [52]. As a consequence, particles circulate in bunches with a length of 25 ns [57]. Typically, the LHC is not completely filled with bunches leaving empty gaps in-between.

[2] The frequency of a cavity is not constant but must be adapted to the particles' speed. However, as protons already enter the LHC at a velocity close to the speed of light, the required frequency adjustments are relatively small.

The entire acceleration is not carried out by the LHC on its own, instead it is supported by a series of smaller pre-accelerators. Currently this pre-acceleration process starts at Linac4[3], which accelerates negatively charged hydrogen ions up to 160 MeV before injecting them into the Proton Synchrotron Booster. During this injection process, the hydrogen ions are stripped from their electrons before the remaining protons are accelerated to an energy of 2 GeV. This step is followed by the Proton Synchrotron, the first synchrotron built at CERN, which propels the protons until they reach 26 GeV. Finally, the last stage is represented by the Super Proton Synchrotron, further accelerating the particles to an energy of 450 GeV before being injected into the LHC [58]. This chain, embedded in the complete accelerator complex at CERN is shown in Figure 3.2, where various experiments installed at some of the pre-accelerators are marked.

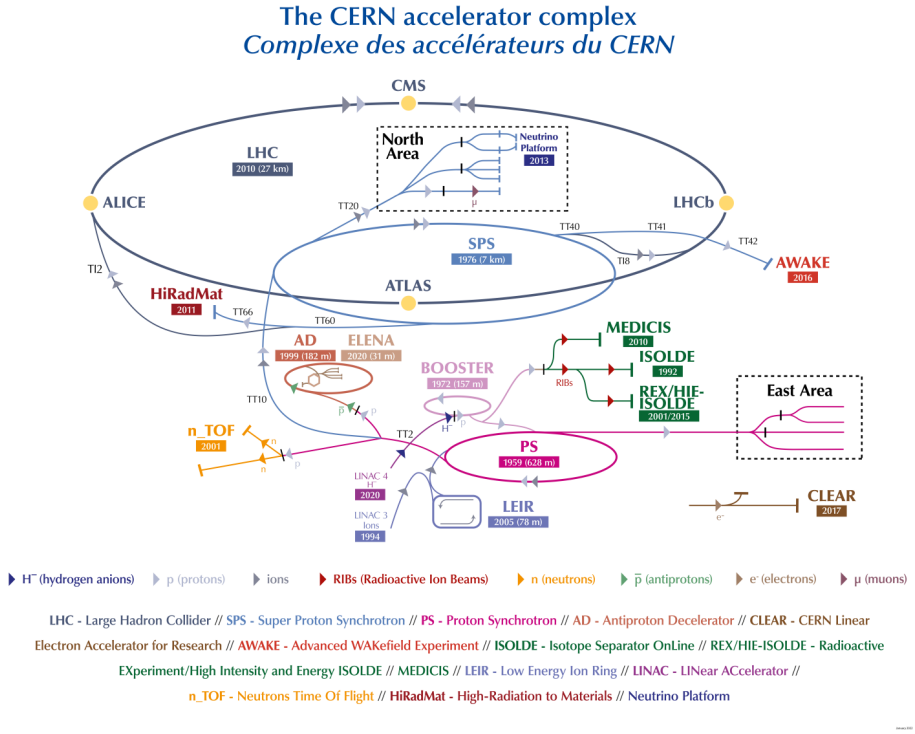



Figure 3.2: The CERN accelerator complex as of 2022, including the chain of accelerators used to feed protons into the LHC. Additionally, the acceleration path of heavy ions is illustrated, starting at Linac3 and identical to that of protons from the Proton Synchrotron onwards [59].

Injection into the LHC is realised via a combination of septum and kicker magnets. The latter are capable of rapidly building up the field required to push the protons into their intended orbit. As soon as all proton bunches are stored in the LHC, the beam energy is ramped up to the desired level before the actual data taking begins.

In the case of the LHC, this data taking is possible at four straight sections by colliding the beams in interaction points. The collision rate is increased by installing beam-focusing quadrupole magnets

[3] During the data acquisition of all events analysed in this thesis, Linac2 was still installed at this point, accelerating protons instead of $H^-$ ions. Linac4 replaced Linac2 in 2020 after the completion of Run 2 as one of the first upgrades towards the high-luminosity LHC.

close to the interaction points.

One large detector is positioned around each of these interaction points to measure the collision outcome: The two largest experiments are the general-purpose detectors ATLAS and CMS, each of them capable of measuring an extremely wide range of physics interactions. Thus, both experiments can verify and compare each other's analysis results. In addition, the ALICE detector, geared towards measuring heavy ion collisions, and LHCb which examines the asymmetry of matter and antimatter by measuring $b$-hadrons are installed in the two remaining sections.

Constantly colliding particles at each of these experiments reduces the beam's intensity, which is why after an interval of data taking the beam quality deteriorates. As a consequence, all bunches are ejected from the LHC and guided into a beam dump mostly made of graphite.

Such a data collection process has by no means been continuous since the LHC's commissioning. Instead, it is interrupted by both planned and unplanned shutdowns. A distinction is made between relatively short stops and long shutdowns. The latter are usually associated with major technical updates and maintenance work performed on both detectors and accelerators. As of autumn 2024, the current data taking period, termed Run 3, is expected to continue until end of June 2026 before the LHC enters its third long shutdown until 2030 [60]. The aim of this shutdown is to transform the facility into the high-luminosity LHC, which is expected to increase the instantaneous luminosity by a factor of ten with respect to its original design value [61].

The data analysed in this thesis were recorded during Run 2, spanning from 2015 to 2018 at a centre-of-mass energy of 13 TeV, resulting in an integrated luminosity of $140\,\mathrm{fb}^{-1}$.

## 3.3 Particle detectors

Although generating high-energy particle collisions, as described in the previous section, is a necessary prerequisite for the successful operation of a typical experiment in particle physics, it is by no means sufficient. The second essential part is the precise observation of everything happening after a collision. Obviously, the outcome of such collisions is invisible to the naked eye. Instead, it has to be recorded via sophisticated detectors, which convert particle interactions into electrical signals.

The detector's objective is usually a measurement of a particle's energy and momentum, so that statements on its mass and ultimately its identity can be made. Modern detectors used for this purpose in experiments are equipped with a wide range of technologies to perform in an optimal way in their respective field of application.

A detector combining a large selection of these technologies is termed a general-purpose detector. The ATLAS detector, described in Section 3.4, with a cylindrical structure surrounding the interaction point belongs to this class. Its structure is divided into several different layers with distinct tasks. The fundamental working principles of all techniques used in the respective layers are briefly introduced in the following sections to allow for a better understanding.

### 3.3.1 Tracking detectors

The first crucial type of detectors are tracking detectors, primarily responsible for measuring the momentum of charged particles. Detecting the ionisation caused by charged particles when passing through matter allows the reconstruction of the particles' trajectories. The associated momenta are ultimately derived by placing the tracking system in a homogeneous magnetic field to bend the

trajectory. In natural units (see Appendix A), the Lorentz force for a single electrically charged particle, the momentum $p$ and the radius $R$ of the trajectory's curvature can be combined to

$$p \left[ \frac{\text{GeV}}{\text{c}} \right] = 0.3 B \, [\text{T}] \, R \, [\text{m}] \ , \tag{3.7}$$

where $B$ is the magnetic flux density.

Nowadays, two main types of tracking detectors are used: gaseous and semiconductor detectors. Each type is split in a large number of highly specialised subtypes, which will not be discussed in this thesis, but are introduced in detail in [53].

### Gaseous detectors

The essential component of each gaseous detector is formed by a capacitor with an applied electric field. A charged particle that flies through the capacitor's gas volume will ionise the gas atoms. Electrons and positively charged ions generated in this process are accelerated towards the anode and cathode, respectively, to be detected in the form of an electric current. In general, gaseous detectors can cover a large volume at relatively low cost and moderate weight, both being of importance in detector setups.

A distinction is usually made between detectors that really enable a spatial resolution and those that only allow pure radiation detection. In all cases, the strength of the externally applied field is decisive for the exact mode of operation. At very low external field strengths, termed the recombination region, the primarily generated ionisation is not entirely measured, as electrons and ions are not separated quickly enough before recombining. A signal saturation can be achieved by increasing the electric field until secondary ionisations, resulting in charge avalanches, are created. In this region, a proportionality between primary charge and created avalanches persists. At higher field strengths space charge effects close to the anode restrict this correlation [53]. Increasing the electric field beyond this level, results in an extremely high signal that is almost independent of the primary ionisation. At this level, even a single initial ionising event is sufficient to generate an avalanche. Geiger counters operate in this mode to detect ionising radiation.

The actual spatial resolution is obtained by arranging multiple detectors next to each other, e.g. in multi-wire proportional chambers. In these chambers, several anode wires are positioned at a constant distance between two cathode plates. Adding a second level of chambers with a different anode wire orientation enables a two-dimensional position measurement.

The choice of the gas mixture varies a lot between experiments and is heavily influenced by several factors, including its cost. In most applications a mixture of noble gases and polyatomic quench gases is chosen. The latter are intended to absorb photons generated during recombination to prevent discharge effects [53].

### Semiconductor detectors

The second class of tracking detectors are semiconductor detectors which exploit the material's properties to measure ionising radiation. They are formed by a special form of p-n junction diodes mostly constructed using silicon. In this context, p and n denote the positive and negative type doping

of the semiconductor's material, i.e. the introduction of impurities. In the case of group IV[4] silicon crystal, n-type doping is achieved by adding group V elements, which provide a free valence electron. Conversely, p-type doping is accomplished through group III elements acting as electron acceptors, leaving a positive defect, known as hole.

If two differently doped layers are brought into direct contact, the different concentrations of free charge carriers create a diffusion current resulting in a recombination of negative electrons and positive holes. The stationary atomic cores remain in place and generate an electric field due to the now missing charge carriers. This field and the diffusion current have opposite directions and lead to the formation of a depletion zone, free of mobile charge carries. This depletion zone can be further enlarged by applying an external electric field.

Charged particles passing through this depletion zone, generate free electron-hole pairs in addition to exciting lattice vibrations. These pairs move towards the external field's electrodes and can ultimately be measured as a signal.

Typical semiconductor detectors consist of long strips, spaced 50 µm to 100 µm apart [53]. Each strip has an individual electrical readout to allow one-dimensional position measurements. A two-dimensional position can be resolved by using an additional strip layer with a different orientation. Improved spatial resolution can be provided by utilising pixel-shaped detectors with a typical length of 100 µm [53]. Each pixel requires its own readout unit, but their arrangement in a grid directly enables two-dimensional measurements.

Penetrating particles will typically lose more energy in a semiconductor detector due to their high density components compared to gaseous detectors, while also offering a significantly lower ionisation energy. However, the usage of semiconductors in an experiment is firstly still quite expensive and secondly requires many individual readout channels associated with a high power consumption.

In summary, both tracking systems enable momentum and position measurement for charged particles. But neither an accurate energy measurement nor the detection of neutral particles is possible inside these detectors. To overcome this issue, tracking detectors are often paired with an additional detector type, the calorimeter, as described in the next section.

### 3.3.2 Calorimeters

Calorimeters are used to measure a particle's energy by stopping it inside the detector material. In this destructive measurement method, incoming particles trigger a shower-like cascade by interacting with the detector's atoms. The measured quantities are ultimately either the electrical charge or scintillation light created in these showers. Due to their proportionality to the initial energy, the latter can be determined [53].

A clear distinction is made between electromagnetic calorimeters that mainly measure electrons as well as photons and hadronic calorimeters which determine the energy of particle jets. The shower development is fundamentally different in both cases.

#### Electromagnetic and hadronic showers

Electromagnetic showers, triggered by electrons, positrons or photons can be described using a simplified and quite illustrative model [62]. First of all, it is assumed that the energy loss due to

[4] Group refers to the periodic table of elements. Group IV corresponds to the carbon group that consists of elements with four valence electrons.

ionisation of the detector materials is itself energy-independent and the complete shower development is regarded as one-dimensional. Only the energy loss of a charged particle after scattering off an atomic nucleus, termed Bremsstrahlung, and pair production, i.e. the annihilation of a photon into an electron-positron pair, are considered to contribute to the shower. Starting with an initial electron these two processes will continue to alternate and thus lead to the formation of a shower until the critical energy $E_c$ is reached. Subsequent energy losses below $E_c$ are modelled exclusively by ionisation.

As the cross-section of Bremsstrahlung and pair production are both proportional to $Z^2$, detector materials with a high atomic number are preferred as a detector material. Those materials' stopping power is characterised in terms of the radiation length $X_0$, defined as the distance after which a particle has lost all but $1/e$ of its energy. Simultaneously, this corresponds to $7/9$ of the mean free path for a photon to produce an electron-positron pair [53]. With this simplified model the expected depths alongside contributing processes of an electromagnetic shower can be described.

Hadrons interacting with the calorimeter material cause shower formations in a similar, albeit more complex, way. A variety of processes must be considered for a correct theoretical description, some of which have significantly different signal efficiencies [53].

A hadronic shower begins with the inelastic scattering of a single hadron and an atomic nucleus. This scattering generates secondary high-energy particles, which themselves can either scatter inelastically again or radiate into photons. Each photon will cause an independent electromagnetic shower inside the hadronic shower. The remaining atomic nuclei are highly excited and lose their energy by spontaneously expelling nucleons, known as spallation. These spallation fragments are emitted continuously with an energy of around 100 MeV until the excitation energy falls below the binding energy level. Any remaining excitation energy is either released in a process called evaporation through primarily radiating nucleons or, in the case of heavy nuclei, through nuclear fission [53].

Compared to electromagnetic showers, this variety of processes typically causes a deeper penetration into the detector along with a worse energy resolution. The length scale of hadronic showers can be expressed through the nuclear absorption rate

$$\lambda_a = \frac{A}{N_A \rho \sigma_{\text{inelastic}}}\,, \tag{3.8}$$

where $\sigma_{\text{inelastic}}$ is defined as the inelastic cross-section, $A$ is the atomic mass number and $\rho$ denotes the detector material's density. Relating $X_0$ and $\lambda_a$ through

$$\frac{\lambda_a}{X_0} \approx 0.37Z\,, \tag{3.9}$$

clearly indicates that hadronic calorimeters tend to be significantly larger than their electromagnetic counterparts.

### Calorimeter setup

Calorimeters designed to stop particles are predominantly setup in two ways. The first method relies on homogeneous construction, resulting in stopping of particles and the measurement of energy deposits being carried out by the same component. The alternative comes in the form of sampling calorimeters. This sampling setup consists of alternating active layers to measure the energy and passive stopping layers. The decision in favour of a setup is usually made on the basis of the required energy resolution.

In the case of electromagnetic calorimeters, both technologies are commonly used, whereby the best energy resolution, typically around

$$\frac{\sigma_E}{E} \approx \frac{3\,\% - 5\,\%}{\sqrt{E/\mathrm{GeV}}}\,, \tag{3.10}$$

is obtained with homogeneous calorimeters [53]. Hadronic calorimeters, on the other hand, almost always use a sampling structure, as strong fluctuations present in the complex hadronic showers negate the advantages of homogeneous calorimeters [53]. These fluctuations limit the precision to which the energy deposit is measured to

$$\frac{\sigma_E}{E} \approx \frac{50\,\%}{\sqrt{E/\mathrm{GeV}}}\,. \tag{3.11}$$

In most setups, the calorimeter is not placed directly next to the interaction point but sits outside the tracking detectors. Although tracking detectors possess a relatively low stopping power, they can already initialise showers within their volume, which ultimately worsens the energy resolution. To compensate this effect, additional detectors, called presamplers, are often installed in-between tracking modules and the calorimeter system. These presamplers help at identifying showers that already started before reaching the actual calorimeter.

### 3.3.3 Missing transverse energy

Since detectors in collider experiments are incapable of directly detecting neutrinos, an indirect measurement via precise momentum determination has to be utilised. Although the initial longitudinal momentum of two colliding partons is unknown, the transverse momentum, i.e. perpendicular to the beam axis, is approximately zero. Therefore, $\sum \vec{p}_{\mathrm{T}_i} = 0$, where $\vec{p}_{\mathrm{T}_i}$ are the transverse momenta of all final state particles, must be valid by requiring momentum conversation. Thus, undetected particles, in addition to measurement uncertainties, lead to $\sum \vec{p}_{\mathrm{T}_i} \neq 0$. The missing contribution is defined as missing transverse energy[5] $E_{\mathrm{T}}^{\mathrm{miss}} = -\left|\sum \vec{p}_{\mathrm{T}_i}\right|$, often providing evidence for neutrinos in the final state.

## 3.4 ATLAS detector

The ATLAS detector (acronym for A Toroidal LHC Apparatus) is a cylindrical detector installed around one of the LHC's four interaction points near the CERN main site. With a length of 46 m and a diameter of 25 m, ATLAS is set up to function as a general-purpose detector allowing it to investigate a wide range of physics.

To fulfil this task, the structure is divided into different layers: located close to the actual interaction point is the inner detector (ID), a system of precise tracking detectors to determine the trajectory of charged particles. The ID is surrounded by the electromagnetic- and hadronic calorimeters to enable an energy measurement. Like in most general-purpose detectors, a muon detector forms the outermost

[5] The magnitude of the transverse momentum vector $p_{\mathrm{T}} = |\vec{p}_{\mathrm{T}}|$ corresponds approximately to the transverse energy $E_{\mathrm{T}}$ for massless particles or $E \gg m$.

layer, to identify this highly penetrating charged leptons[6]. As discussed in Section 3.3.1 measurements in the ID as well as the muon detector rely on magnetic fields. For this reason, a complex magnet system is integrated into the setup, forming the last essential component. Each of these components is discussed in its own section but first the coordinate system under which ATLAS operates needs introduction.

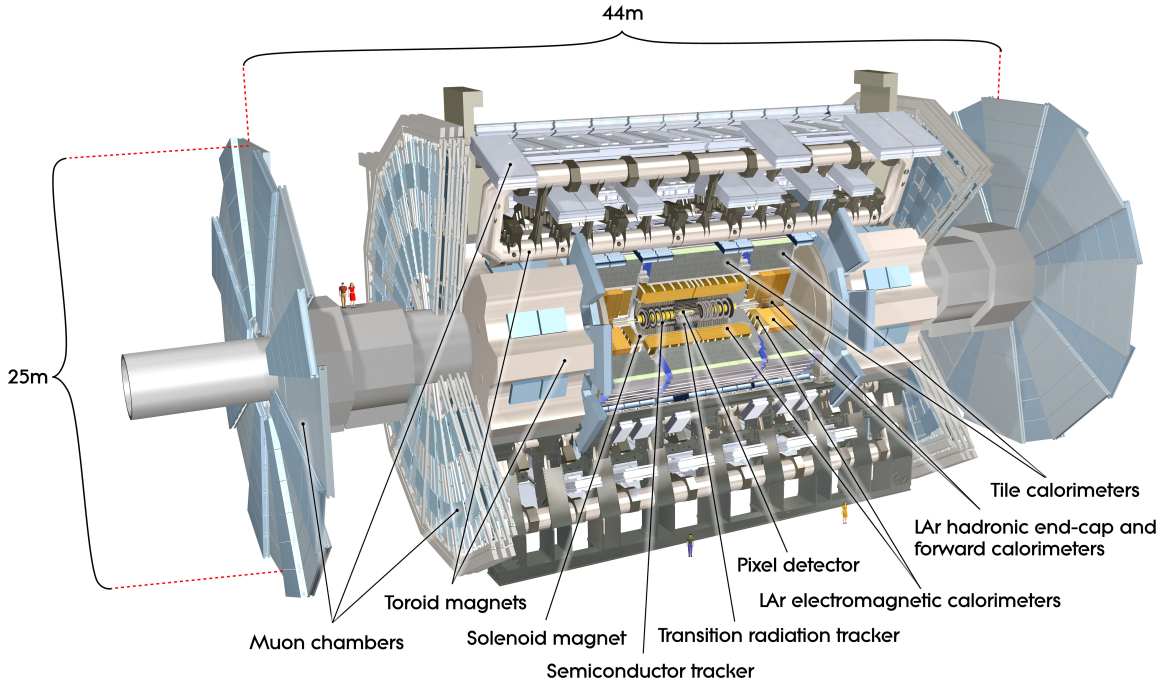



Figure 3.3: Computer generated sketch of the ATLAS detector with all its subcomponents marked [63].

### 3.4.1 Coordinate system

To simplify the description of particle collisions, a special coordinate system as shown in Figure 3.4 is defined. Its origin corresponds to the collider's interaction point with the $z$-axis running along the horizontal beam pipes. The transverse plane is formed by the $y$-axis pointing upwards and the positive $x$-axis pointing towards the centre of the LHC. While the azimuthal angle $\phi$ of classical spherical coordinates remains unchanged, the polar angle $\theta$ is often replaced by the pseudorapidity $\eta$:

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right) . \tag{3.12}$$

In a pseudorapidity interval, the flux of the particles is approximately constant, since $\Delta\eta$ represents a Lorentz invariant quantity (see Appendix B) with respect to boosts in $z$-direction. A pseudorapidity of

[6] Due to their higher mass, muons lose less energy through Bremsstrahlung compared to lighter electrons. Therefore they penetrate all previous detector layers. A signal in this outer detector is thus a clear indication of a muon being present.

$\eta = 0$ represents a position perpendicular to the beam axis while $\eta = \pm\infty$ corresponds to a position along the direction of the positive or negative $z$-axis. Distances in the plane spanned by $\phi$ and $\eta$ are determined by

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}\,. \tag{3.13}$$

By combining $\phi$, $\eta$ and the transverse momentum $p_\mathrm{T}$, one is able to derive the momentum vector in three dimensions

$$\vec{p} = \begin{pmatrix} p_x \\ p_y \\ p_z \end{pmatrix} = \begin{pmatrix} p_\mathrm{T} \cdot \cos\phi \\ p_\mathrm{T} \cdot \sin\phi \\ p_\mathrm{T} \cdot \sinh\eta \end{pmatrix} \quad \text{with} \quad p_\mathrm{T} = \sqrt{p_x^2 + p_y^2}\,. \tag{3.14}$$

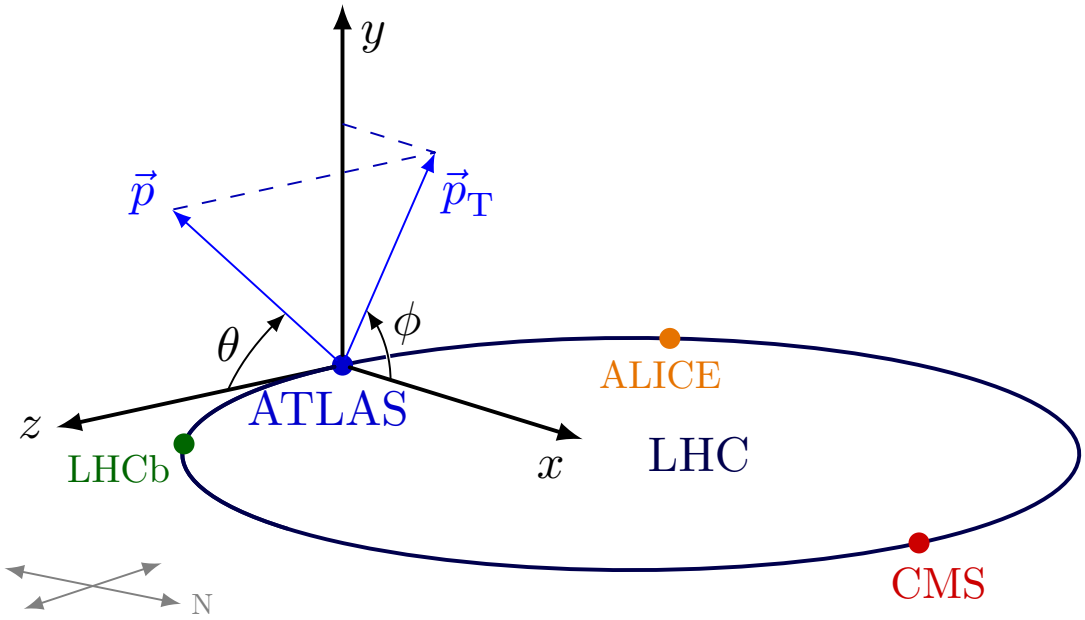



Figure 3.4: The ATLAS detector's special coordinate system alongside the LHC [64].

Beginning from this coordinate system's origin the detector's individual components are described in the following sections. The exact configuration of the detector has by no means remained unchanged since its installation back in 2008 but is subject to constant maintenance and modification. The setup outlined in this chapter corresponds to the status at the end of Run 2. The next substantial upgrade will be installed before the high-luminosity LHC [65] is commissioned to ensure precise particle detection under the new harsher conditions.

### 3.4.2 Magnet system

ATLAS' magnet system, required to determine the particle's momenta, is split into two parts. First, the entire ID is permeated by a 2 T magnetic field which is oriented along the beam axis. This field is

provided by a solenoid magnet placed between the ID and the electromagnetic calorimeter. Due to the solenoid's positioning its total material thickness of approximately 0.66 radiation lengths [66] was kept as low as possible to still allow a good energy resolution in both calorimeters.

The outer muon detector is immersed in a second magnetic field generated by toroidal magnets. The central area, perpendicular to the beam axis, is covered by a barrel toroid providing up to 3.5 T [67]. In the forward region, two smaller toroids, also known as end-cap toroids, close off the system. The combined setup, split into a central barrel region and the end-caps along the beam axis, is displayed in Figure 3.3.

### 3.4.3 Inner detector

The ID, as displayed in Figure 3.5(a), is divided into three separate systems. It enables precise track reconstruction up to $|\eta| < 2.5$ despite the harsh conditions arising from its location close to the interaction point.

The first layer is formed by a silicon pixel detector [68, 69] with sensor sizes of 50 µm × 400 µm [52]. The system originally consisted of three layers in the central region alongside three discs along each forward region. In order to maintain the detector's high efficiency throughout Run 2 and the associated increase in luminosity the insertable B-Layer (IBL)[7] [70] was installed in 2015 as a fourth layer. The IBL is placed at an average radius of 33.25 mm from the beam line and consists of an additional 280 pixel modules (with a smaller sensor size of 50 µm × 250 µm [71]) to increase precision and robustness of track reconstruction [71].

The second detection system, the semiconductor tracker (SCT) [72], a silicon strip detector, starts at a radius of 299 mm with respect to the beam axis. With a total of 4 088 strip modules, the SCT is divided into four concentric barrels in the central area and nine discs along the forward region.

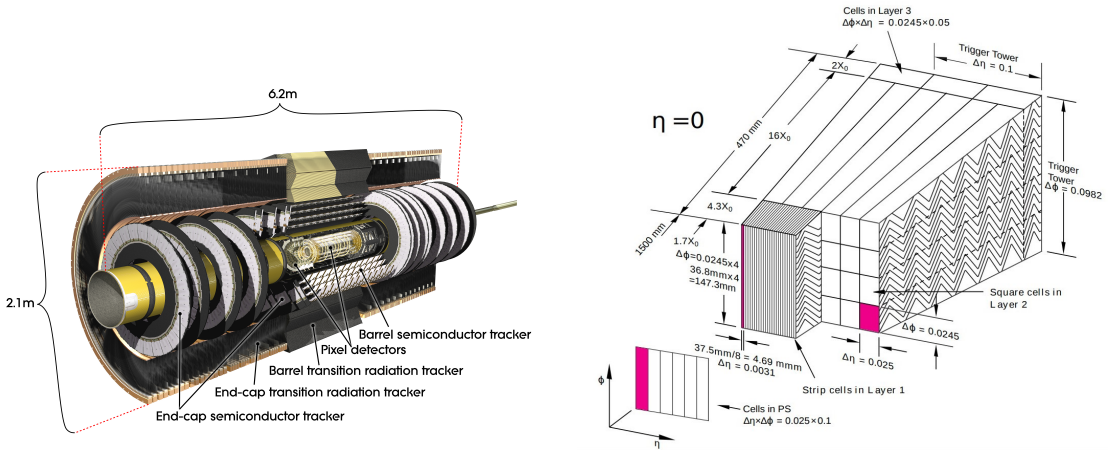
The outermost and largest component is the transition radiation tracker (TRT) [73] which allows measurements up to $|\eta| < 2.0$. This special detector type utilises transition radiation besides ionisation. Transition radiation occurs when a charged relativistic particle crosses the interface between two media with different permittivities. The radiated photon's intensity is proportional to the particle's gamma factor $\gamma = E/m$ and thus the TRT not only provides tracking information but also allows conclusions to be drawn about the particle's mass. At the TRT photons are detected by roughly 300 000 gas-filled straw tubes each housing an anode wire [73]. In the central area, all straw tubes are positioned parallel with respect to the beam axis, while the end-cap straw tubes are orientated perpendicular to it.

### 3.4.4 Calorimeters

ATLAS utilises calorimeter systems covering a combined range of up to $|\eta| = 4.9$. All individual modules rely on the sampling technique, discussed in Section 3.3.2, to enable precise energy measurement.

The first one surrounding the ID is the electromagnetic calorimeter which uses passive lead and active liquid argon layers. It consists of a central barrel region in the interval $|\eta| < 1.475$ and two end-cap units in the interval $1.375 < |\eta| < 3.2$ to enclose the interaction point as hermetically as possible. The area in the interval $1.37 \leq |\eta| < 1.52$ has a large amount of material upstream and is

[7] The name is derived from hadrons containing bottom-quarks, which decay about a millimetre away from the original interaction point as a result of their long lifetime. This distance can be measured with pixel detectors located as close as possible to the interaction point.




(a) (b)

Figure 3.5: A sketch of (a) the ATLAS ID showing the barrel layers in the central region as well as the end-caps in the forward region (IBL not included) [74] and (b) the electromagnetic calorimeter setup in the central region around $|\eta| = 0$ [75].

therefore usually not used for precision measurements. All individual layers of the central barrel region around $|\eta| = 0$, illustrated schematically in Figure 3.5(b), were constructed in a special accordion geometry to provide complete azimuthal coverage.

Up to $|\eta| < 2.5$, the calorimeter is composed of three layers. The first layer, with a thickness of about 4.4 $X_0$ around $|\eta| = 0$, is formed by high-granularity strips with a cell size of $\Delta\eta \times \Delta\phi = 0.003 \times 0.0982$ in the range $|\eta| < 1.4$ and $1.5 < |\eta| < 2.4$, while slightly larger cells ($\Delta\eta \times \Delta\phi = 0.025 \times 0.0982$) are installed in the intervals $1.4 < |\eta| < 1.5$ and $2.4 < |\eta| < 2.5$. The resulting resolution allows a distinction between prompt photons, directly created in the collision, and photons originating from the decay of neutral hadrons (such as $\pi^0 \rightarrow \gamma\gamma$) [75].

Most of a particle's energy is absorbed by the second layer with a thickness of 16 $X_0$ at $|\eta| = 0$ using a cell size of $\Delta\eta \times \Delta\phi = 0.025 \times 0.0245$. In the exceptional case of electrons or photons not being completely stopped in the electromagnetic calorimeter, their energy measurement can be corrected with the help of a thinner third layer at a thickness of 2 $X_0$ around $|\eta| = 0$ [75].

Energy loss upstream to the calorimeter is corrected through installation of a presampler in front of the first layer covering $|\eta| < 1.8$. Its thickness is constant in the range $|\eta| < 0.6$ and $0.8 < |\eta| < 1.8$ at approximately 2 $X_0$ and 3 $X_0$, respectively, increasing linearly in-between. An exception is formed by the transition region near $|\eta| = 1.7$ with a thickness of around 6 $X_0$ [75].

The complete electromagnetic calorimeter is surrounded by three detectors forming the hadronic calorimeter. The central area around $|\eta| < 1.7$ is enclosed by the tile calorimeter which is longitudinally divided into a central module and two extended-barrel modules. In this case, steel is used as a passive absorber medium while the signal detection is managed by scintillation detectors. Along the radial direction, three layers with a respective thickness of up to 3.3 $\lambda_a$ can be distinguished [76].

Close to the beam axis lies the second component, the forward calorimeter with a thickness of approximately 10 $\lambda_a$. It is split into three layers on each side of the interaction point, with copper and tungsten being used as passive elements in the first layer and outer two modules, respectively. Liquid argon is again installed as the active medium in all three layers [52].

The gap between the tile calorimeter and the forward calorimeter is bridged by the hadronic end-cap calorimeter, which covers the interval $1.5 < |\eta| < 3.2$ resulting in short overlaps with the other two systems. The choice of active and passive material is identical to the forward calorimeter's first layer, while the granularity decreases with increasing pseudorapidity [52].

### 3.4.5 Muon spectrometer

The outermost part of the ATLAS detector is the muon spectrometer (MS). This system is not only able to track charged particles up to $|\eta| < 2.7$ but also to serve as a trigger [52]. Again, an organisation into a barrel region, which cylindrically surrounds the interaction point, and two end-cap regions, as displayed in Figure 3.3, was chosen. In the central region monitored drift tubes with a spatial resolution of 80 µm are used to measure the track. The tubes are aligned perpendicular to the beam axis to enable precise determination of the trajectory in one dimension. In the innermost area of the end-caps at $2.0 < |\eta| < 2.7$, this role is taken over by cathode strip chambers, a special form of multi-wire proportional chambers. The chambers' outer cathode is divided into strips to guarantee high spatial resolution of around 60 µm [77].

The trigger system is only functioning up to $|\eta| < 2.4$ and similarly consists of two different gaseous detectors. The setup's limiting factor is the time resolution which has to be lower than the 25 ns bunch spacing in the LHC. This is secured by using resistive plate chambers in the region $|\eta| < 1.05$, while the end-caps up to $|\eta| < 2.4$ are covered by thin gap chambers. Readout strips in those systems are placed orthogonal to the monitored drift tubes. This kind of positioning allows the second coordinate of the particle track in the non-bending direction to be measured [77].

### 3.4.6 ATLAS trigger system

The LHC delivers collisions at a rate of up to 40 MHz, which corresponds to a data volume of 80 Pb/s [52]. It is neither possible nor sensible to read out all of these collisions and store them for subsequent analysis. Therefore, interesting collisions, usually the ones with high transverse momentum or special event topologies, have to be identified and selected.

In the ATLAS detector, this filtering is carried out by a two-part trigger system. Initially, the L1 trigger is used, which is able to accept or reject events in less than 2.5 µs [78]. Information of the calorimeters and the MS are combined in a central trigger processor to form a decision.

Even a trigger processing time of 2.5 µs is obviously still above the LHC's collision rate. In other words, analysing each single collision would fill up the trigger's buffer rapidly. As a countermeasure, preventive deadtimes, during which data acquisition is completely paused, are used to avoid this filling up. Each event that is analysed, i.e. enters the trigger, always leads to a fixed deadtime. In addition, the general number of events accepted by the trigger per interval of bunch crossings is limited. The resulting rate of accepted events with a chosen topology can further be limited through a prescaling. The strength of this scaling can be adjusted during data taking to counteract the decrease in luminosity towards the end of one LHC fill cycle.

Once an event is accepted, it is forwarded to the second stage, the high-level trigger. The information entering this trigger software is clustered into detector compartments, termed regions of interest. These regions are designed to be a good first choice for hinting at interesting physics. Utilising such regions the rate can be lowered further to an average of 1 kHz in roughly 200 ms [78]. The actual decision making is split into a fast initial reconstruction and a more precise but slower follow-up.

During each step, a veto can cause events to be discarded to keep the latency low. In case errors occur during this process, events are re-analysed offline with more relaxed time settings to detect problems in the trigger software.

Once an event is selected by the high-level trigger, it is processed in several event building and filtering stages before being ultimately distributed to the worldwide LHC computing grid [79]. Although the grid is not part of the detector, it is just pivotal for successful analyses. An overview is presented in [80].

## 3.5 Reconstruction and identification

While the direct output of the detector is fundamental, it is by no means sufficient to precisely analyse the collisions in the experiment. In fact, a complex combination of all subsystems' signals is required to correctly reconstruct and identify individual physics objects.

As already stated, we can neither observe the collision nor the signalling processes with our eyes. Nonetheless, the basic ideas of particle identification can be easily understood by visualising the detector response, in a way known as event displays. Initially, only the properties of long-lived particles need to be visualised and considered. All others can ultimately only be identified via their decay products as they decay before reaching the detector. A comparison of signatures that serve as a valid foundation for understanding reconstruction methods is shown in Figure 3.6 and is briefly explained in the following:

- Electrons and photons usually both deposit their entire energy in the electromagnetic calorimeter. A distinction is possible by exploiting the fact that the neutral photon, as long as it has not converted into an electron-positron pair, does not trigger a signal in the ID.
- Just like the electron, particle jets are also detected in the ID and the electromagnetic calorimeter. However, jets also penetrate into the hadronic calorimeter, where most of their energy is collected.
- As described in Section 3.4, muons lose significantly less energy in the calorimeter system than electrons and therefore pass through the entire detector. This simplifies identification as a signal in the MS represents strong evidence for a muon. Information from the ID acts as additional support.
- Neutrinos leave the detector and do not cause a signal at all as they are electrically neutral. Consequently, the only evidence of their presence is provided by detecting missing transverse energy, as mentioned in Section 3.3.3.

Using these different responses, it is possible to primitively distinguish particles based on their signature in the detector[8]. However, this visual method is neither able to keep up with the event rate of the LHC nor does it allow precise quantification of energy and momentum. The latter in particular is absolutely necessary to draw conclusions about short-lived collision products such as the Higgs-boson.

To compensate, different reconstruction algorithms are applied to measure tracks in the ID and energy deposits in the calorimeter, as explained in Section 3.5.1 and Section 3.5.2, respectively.

[8] This kind of analysis method was actually used in detector types such as cloud chambers.

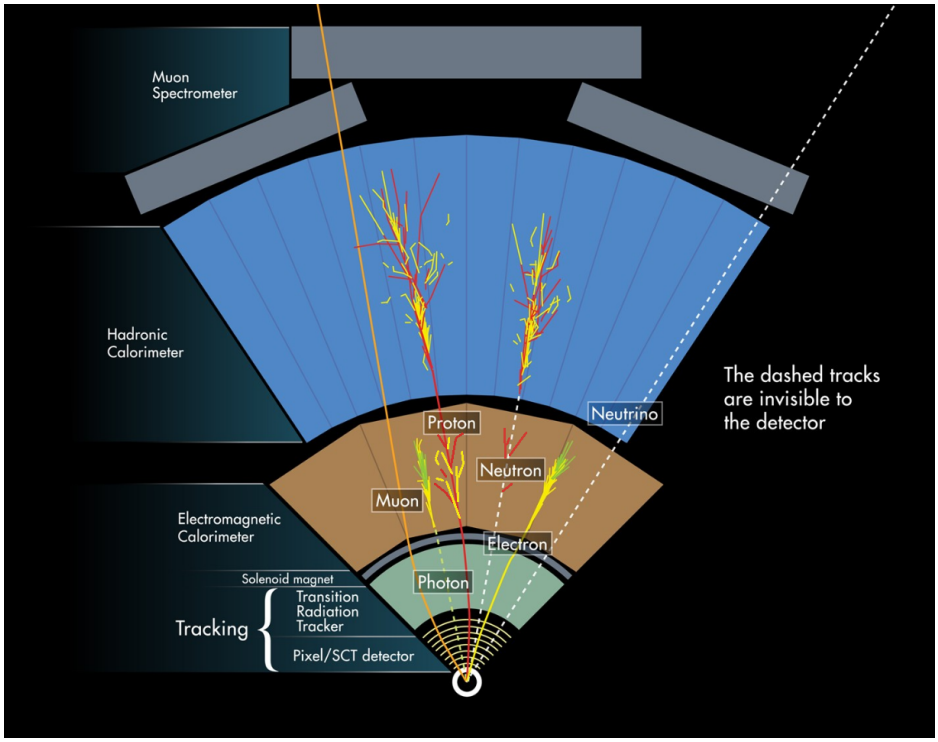



Figure 3.6: Signature of particles in the different ATLAS detector subsystems [81].

Subsequently, Section 3.5.3 focuses on the actual reconstruction of all crucial particles in the analysed $tH$ channel, i.e. charged light leptons, particle jets, hadronically decaying tau-leptons as well as neutrinos.

Optimisation and validation of these algorithms is partially based on the usage of simulated Monte Carlo (MC) events, whose generation is explained in Section 4.4. Since the underlying process of each MC event is known, a direct validation of any reconstruction algorithms' efficiency is possible.

### 3.5.1 Track reconstruction

Trajectories of charged particles are represented by five parameters with respect to a reference point which corresponds to the average position of collisions:

- The two impact parameters $d_0$ and $z_0$ that indicate the transverse and longitudinal distance between the point of closest approach and the reference point.
- The azimuthal angle $\phi$ as well as the polar angle $\theta$.
- The ratio of charge and momentum $\frac{q}{p}$ to parametrise the track curvature.

The track reconstruction is initialised by forming independent clusters in the pixel detector and the SCT through grouping signals from adjacent sensors above a threshold value. This way, three-dimensional space-points are obtained, which, when combined along different layers, create a potential track candidate, called a track seed. Every track seed that is either found to be incompatible with a fourth

space-point belonging to an adjacent layer or has large impact parameters is disregarded. This selection reduces the number of possible combinations and thus the reconstruction time [82].

In the next step a combinatorial Kalman filter [83] is applied along the expected trajectory of each seed to find adjacent clusters resulting in extended seeds. This search is performed both inwards and outwards to obtain a smooth trajectory. The resulting track extension is performed iteratively, while adding single space-points which ensures a fast reconstruction but, on the other hand, leads to ambiguities in case multiple points are compatible. Therefore, an additional algorithm assigns a score to each track candidate (based on criteria listed in [84]) to solve the ambiguities by rejecting low score candidates. All remaining track candidates are run through a high-resolution global fit before an attempt is made to extend them by adding TRT signals [84].

Based on the chosen direction, this complete method is termed the inside-out approach, optimised to reconstruct trajectories close to the primary proton-proton collision. In order to also reconstruct particles further away from the beam line, such as electrons from photon conversions, a second reconstruction with reversed direction is applied. This outside-in track reconstruction is started by seeds in the TRT, the position of which is guided by regions of interest in the electromagnetic calorimeter. The resulting track candidates are extended towards both silicon detectors and processed analogously to the inside-out method. To maintain the high reconstruction speed only candidates not already assigned by the first method are analysed.

Besides the mere reconstruction of tracks, focus is also put on their assignment to the position of the original proton-proton collision, known as vertex or primary vertex. This is achieved through adaptive multi-vertex fitting [85] by searching for potential vertex seeds via the track density along the beam axis. All tracks in the seed's immediate vicinity that pass a preselection are taken into account. The fit is performed iteratively by continuously down-weighting incompatible tracks. As soon as the position of a vertex is fixed, all incompatible tracks below a weight threshold are removed and used again to determine other vertices. This procedure is continued until no seeds are left [82].

For a visualisation of the entire procedures, please refer to [82].

### 3.5.2 Calorimeter clustering

Particles that enter the calorimeter systems spread their energy deposit longitudinally and laterally across many cells. The resulting signals in individual cells are combined into clusters and subsequently calibrated depending on the incoming particle type (see [86] for a description of this calibration).

The used clustering algorithm is split into two methods depending on the physics object: jets and missing transverse energy are reconstructed via topological clusters, which provide strong noise suppression for widely spread clusters. Similar to the track reconstruction, topological cluster creation is initialised by searching for seed cells. The natural choice are cells with a high signal significance of

$$\left|\zeta_{\text{cell}}^{\text{EM}}\right| = \left|\frac{E_{\text{cell}}^{\text{EM}}}{\sigma_{\text{noise,cell}}^{\text{EM}}}\right| > 4\,, \tag{3.15}$$

where $E_{\text{cell}}^{\text{EM}}$ corresponds to the energy deposit in a single cell, while $\sigma_{\text{noise,cell}}^{\text{EM}}$ denotes the average expected noise. The cluster formation starts around the seed cell by adding all neighbouring cells for which $\left|\zeta_{\text{cell}}^{\text{EM}}\right| > 0$ and $\left|\zeta_{\text{cell}}^{\text{EM}}\right| > 2$ applies. In this context, neighbouring cells are regarded as adjacent cells in the same calorimeter layer or as cells from other layers which overlap in the $\eta$-$\phi$ plane. The

complete growth is controlled by the following criteria:

- If $\left|\zeta_{\text{cell}}^{\text{EM}}\right| > 2$ applies to a cell candidate, not only the cell itself but also all adjacent cells are added to the cluster.
- A cell candidate with $\left|\zeta_{\text{cell}}^{\text{EM}}\right| > 4$ is a seed itself. As a consequence both growing clusters are merged.
- A cell candidate with $\left|\zeta_{\text{cell}}^{\text{EM}}\right| > 2$, that could be added to two clusters, causes both to be merged.
- A candidate with $0 < \left|\zeta_{\text{cell}}^{\text{EM}}\right| < 2$ is near the noise threshold. As a consequence, it is still added to a cluster but terminates the growth in that direction.

The magnitude is used in all cases to correctly treat negative energy entries caused by fluctuations, primarily due to out-of-time pile-up. While clusters arising from negative energy seeds are not used for the reconstruction of physics objects, they can still serve as an estimator for the general pile-up contamination [87].

A cluster formed by these rules can grow rapidly given sufficient energy and contain showers of several particles. Therefore, such clusters are subsequently split to allow a separation of individual particles. This splitting commences by finding local maxima. Around those, a cluster formation similar to the initial process is carried out. Entries in cells that could stem from two local maxima are divided according to both maxima's energy and distance [88].

The second method known as the sliding-window algorithm, is based on the merging of cells using a fixed rectangular window. This fixed size ensures a very precise calibration and is used for electrons, photons and tau-leptons [88]. For a good understanding, it is helpful to imagine the lateral surface area of an unrolled cylindrical calorimeter. The resulting surface represents the $\eta$-$\phi$ plane and can be divided into a grid with elements of size $\Delta\eta \times \Delta\phi$. The energies of each grid element are summed up along the layers of the longitudinal axis, consequently looking like towers rising on the lateral surface.

Finding a cluster seed, termed precluster, is achieved by sliding a scan window with dimensions $N_\eta \times N_\phi$ (in units of $\Delta\eta \times \Delta\phi$) across the grid until a local maximum above a threshold value is found. The exact position of this precluster is determined by calculating the energy-weighted $\eta$ and $\phi$ barycentres within this window [88]. The exact window size, the grid granularity as well as the threshold value all depend on the hypothesised particle type. If the latter is either an electron or a photon only entries in the electromagnetic calorimeter are considered, while the chosen cluster size varies in each calorimeter layer. The exact choice will always be a trade-off: on the one hand a large window contains a high fraction of the particle's total energy while on the other hand adding a high number of cells increases the total noise contribution [88]. Specific values used in the algorithms are given in [88].

### 3.5.3 Reconstructing individual physics objects

Combining the reconstructed tracks along with the energy clusters and, in some cases, the additional information from the MS allows identifying physics objects such as individual particles or jets. The methods used for all objects involved in the presented analysis are described in the following.

### Jets and flavour-tagging

As explained in Section 2.2.2, cone-shaped particle jets originate either directly from the parton collisions or arise from the decay of an intermediate product such as the Higgs-boson. The detector can only detect the final states at the end of the hadronic cascade since the initial interactions take place before its first layer is reached. A measurement of this highly fluctuating signature is far from trivial but of great importance while searching for known resonances or potential new particles.

In the ATLAS detector jet reconstruction is based on a sequential recombination algorithm starting by selecting a high energy jet seed. Objects are added iteratively to the seed based on a distance metric until a stopping criterion is reached. The metric is formed by two momentum space distances $d_{ij}$ and $d_{iB}$ [89]:

$$d_{ij} = \min\left(p_{\mathrm{T},i}^{2a}, p_{\mathrm{T},j}^{2a}\right) \cdot \frac{R_{ij}^2}{R}\,, \tag{3.16}$$

$$d_{iB} = p_{\mathrm{T},i}^{2a}\,, \tag{3.17}$$

where $p_{\mathrm{T},i}^{a}$ and $p_{\mathrm{T},j}$ are defined as the transverse momentum of two particles while $R_{ij}$ denotes the distance in the $\eta$-$\phi$ plane. The clustering algorithm works by finding the minimum in the set $\{d_{ij}, d_{iB}\}$ until all objects are assigned according to the following set of rules [89]:

- If $d_{ij}$ corresponds to the minimum, both $i$ and $j$ are combined in one jet.
- If $d_{iB}$ corresponds to the minimum, the object $i$ is considered a final jet and removed from the set.

The exact values of $a$ and the radius parameter $R$ are both dependent on the physics use case. ATLAS uses the anti-$k_{\mathrm{T}}$ algorithm [90] corresponding to $a = -1$. This choice leads to a reconstruction which primarily focuses on isolating high momentum objects. The most common choice for the radius size is $R = 0.4$, whereas $R = 1.0$ is used for strongly boosted objects, such as the hadronic decay products of a high-energy Higgs-boson. In this case both jets begin to overlap and are combined into a single large radius jet through the increased value of $R$.

Tracks from the ID and energy clusters in the calorimeters are both used as input to this algorithm. This choice is the logical consequence as both systems excel at opposite energy levels. In addition, only the calorimeters are able to resolve neutral components present in each jet.

This combination is provided through the particle flow algorithm [91] by trying to match high-quality ID tracks with calorimeter clusters. If a merge is successful, the expected energy deposit of the single particle track is subtracted from matched clusters. This subtraction is optimised on the basis to two main conditions: on the one hand, no energy entries associated with the track should remain in order to avoid double counting. On the other hand, the removal of energy entries belonging to other particles must be avoided to not impair their reconstruction.

All jets reconstructed in this algorithm require a subsequent energy calibration. First, isotropic neutral pile-up components are subtracted from the jet energy while charged pile-up contributions are directly suppressed in the particle flow algorithm. The following calibration itself consists of three main stages [92]:

- A jet energy scale (JES) is derived by comparing the true jet energy of simulated MC events with the associated reconstructed jet energy. This scale factor is determined as a function of

both energy and $\eta$.

- Fluctuations in the hadronic shower development are sequentially corrected to improve the jet energy resolution (JER).
- A complex in situ calibration using real data corrects the detector response. This adjustment results in considerable improvements, especially in the forward region.

These calibration corrections mainly solve issues arising from pile-up contributions inside of jets. In addition, pile-up can also cause formation of completely independent jets. These are suppressed or filtered by assigning them to their original vertex. ATLAS achieves this by applying a Jet Vertex Tagger based on a multivariate analysis of track properties [93].

The exact implementation and validation of all algorithms and corrections is quite technical and would rather hinder the flow of reading than contribute to a deeper understanding. Therefore, please refer to [91] and [92] for detailed information.

In many physics analyses it is crucial to identify the flavour of the quark that initiated the jet, a process known as flavour-tagging. A distinction is generally made between bottom-quarks, charm-quarks and all lighter quarks, whose associated jets are referred to as light-flavour jets.

As stated in Section 2.2.2, $B$-hadrons in particular have a relatively long lifetime, allowing them to be detected via a secondary vertex given sufficient spatial resolution. A $B$-hadron with a transverse momentum of 50 GeV, for example, will travel roughly 3 mm before decaying. This effect also applies to hadrons that contain charm-quarks, albeit to a lesser extent.

Flavour-tagging applied in this analysis is divided into two parts: firstly, various low-level algorithms are used to generate variables that provide good discrimination between different flavours. This is achieved by evaluating the track-based impact parameters and identifying the displaced secondary vertex. As bottom-quarks predominantly decay into charm-quarks, the topological structure is further exploited by also trying to find the charm-quark's tertiary vertex [94].

These pieces of information are ultimately combined in the neural network based DL1r algorithm [95]. Optimisation is realised by training on simulated MC events. The resulting multidimensional output corresponds to probabilities of a jet being a $b$-jet (see Section 2.2.2), a $c$-jet or a light-flavour jet [94].

### Hadronic tau-leptons

As a result of its relatively short lifetime of $(290.3 \pm 0.5) \times 10^{-15}$ s [10] the tau-lepton itself, in contrast to the two lighter charged leptons, is not directly visible in the detector. Therefore, the only way to identify tau-leptons is via reconstructing their decay products. As already outlined in Section 2.2.3, a distinction is made between leptonic and hadronic decay modes. This section deals exclusively with the reconstruction of the hadronically decaying tau-leptons, referred to as $\tau_{\text{had}}$. The reconstruction of leptonically decaying tau-leptons is performed in a similar way as the reconstruction of electrons and muons.

The search for $\tau_{\text{had}}$ candidates is initialised analogously to jets using the anti-$k_t$ algorithm. The radius parameter is kept at $R = 0.4$ while only calorimeter clusters are used as input. Resulting jets seeds considered as $\tau_{\text{had}}$ candidates must fulfil $|\eta| < 2.5$ along with $p_{\text{T}} > 5$ GeV [96].

Similar to $b$-jets, the lifetime of tau-leptons also allows the detection of a separate vertex determined through the designated tau vertex association algorithm [97]. Locating this vertex allows a better assignment of tracks to the $\tau_{\text{had}}$ candidate as the impact parameters are calculated relative to it [96].

Tracks which, besides fulfilling a list of quality criteria, also lie in a cone of $\Delta R = 0.25$ around the $\tau_{\text{had}}$ candidate are associated to the vertex. In addition, tracks between $0.25 < \Delta R < 0.4$ are accepted if matched to the original jet seed. The associated tracks are then divided into four categories:

- Tau tracks: all tracks that can be assigned to the $\tau_{\text{had}}$'s decay products. These are utilised to determine not only charge but also the number of charged decay products, also known as prongness.
- Conversion tracks: these tracks are associated to electrons or positrons which are created if photons undergo the process of pair production.
- Isolation tracks: tracks that belong to jets created by the remnants of hard scattering interactions.
- Fake tracks: these remaining tracks cannot be assigned to any of the other categories. They are usually either misreconstructed or belong to pile-up collisions.

All resulting $\tau_{\text{had}}$ candidates still lack a distinct separation between true and fake tau-leptons. The latter predominantly originate from quark or gluon initiated jets. This identification is realised through a recurrent neural network (RNN) [98]. Separate RNNs are optimised for 1-prong and 3-prong candidates, using the track and cluster information as input in both cases. Four working points, specified in [96], based on the actual network output, are available for physics analyses.

### Electrons

The reconstruction of electrons begins with a review of calorimeter clusters to find potential candidates. Reviewed clusters are only those that meet the listed criteria [99]:

- The cluster is required to have energy deposits of at least 400 MeV in the electromagnetic calorimeter.
- The energy deposit in the electromagnetic calorimeter must contribute more than 50 % to the total cluster energy.

Out of those, only clusters that are matched to a high-quality ID track and additionally sum up to an energy of at least 1 GeV are used as electron candidates [100]. In order to accurately reconstruct the shower development in the calorimeter, certain additional clusters are merged with the electron candidates:

- Closely neighbouring clusters are merged if their barycentres are within a range of $\Delta\eta \times \Delta\phi = 0.075 \times 0.125$.
- Clusters in a range of $\Delta\eta \times \Delta\phi = 0.125 \times 0.3$ around the candidate are merged if their associated track or the conversion vertex is identical to that of the original cluster.

A separation is carried out based on these reconstructed electron candidates to distinguish prompt electrons from background sources. Prompt electrons include not only those that are created directly in the hard scattering vertex but also those arising from the decay of heavy resonances such as the Higgs-boson [101].

A likelihood classifier provides optimised selection criteria as a function of $|\eta|$ and transverse energy by utilising highly discriminative variables. They characterise track properties, shower development and the goodness of track-cluster matching as listed in [99]. Three working points (see Section 7.1), specified in [99], provide different levels of signal acceptance and background suppression based on the classifier output.

### Muons

While optimal muon reconstruction is ensured by combining inputs of all detector subsystems, the most important ones obviously stem from the MS. Here, signals from individual detector segments are first combined into short tracks aided by Hough transformations [102]. Using a set of approximations, these short tracks are combined in a global fit to form one complete three-dimensional track candidate. Subsequently, outliers are removed and ambiguities are resolved by keeping only high-quality tracks in case of overlaps [103].

These MS tracks, in combination with information from the ID and both calorimeters, enable a global muon reconstruction. A distinction is made between five strategies and the corresponding muon types:

- Combined muons: reconstructed ID and MS tracks are matched and fitted together. Energy losses in the calorimeters are taken into account.
- Inside-out combined muons: a reconstructed track in the ID is extrapolated to the MS. Again, energy losses are taken into account, but no previously reconstructed MS track is required.
- MS extrapolated muons: in this case, the association of reconstructed MS tracks and ID tracks failed. Instead, the MS track is extrapolated back to the beam line.
- Segment-tagged muons: this method reconstructs ID tracks that cannot be matched to a global MS track, but at least to one of the short tracks formed in a single segment.
- Calorimeter-tagged muons: MS tracks are not taken into account. Instead, an attempt is made to match an ID track with a muon-like energy deposit in the calorimeter.

All reconstructed muon candidates must additionally fulfil further identification criteria in order to be used in physics analyses. These are divided into different working points depending on their strictness. The main aim is to separate prompt muons and non-prompt muons originating from jets [103]. Prompt muons refer to muons that are produced in the primary interaction, whereas non-prompt muons often stem from hadronic decays.

The separation is further enhanced by requiring an isolation condition to be met. Here, isolation refers to the hadronic activity around each muon candidate which is measured via energy deposits in a cone around muon candidate. In the case of prompt muons this activity is generally expected to be low. For a summary of efficiencies corresponding to each working point alongside detailed technical information, please refer to [103].

### Missing transverse energy

As outlined in Section 3.3.3, the missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ can indicate the presence of an undetected particle, often a neutrino. In ATLAS, $E_{\mathrm{T}}^{\mathrm{miss}} = \left|\vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right|$ is calculated by summing the

transverse momentum of all objects found in one event:

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} = -\left( \sum_{\substack{\text{selected}\\ \text{electrons}}} \vec{p}_{\mathrm{T}}^{\,e} + \sum_{\substack{\text{accepted}\\ \text{photons}}} \vec{p}_{\mathrm{T}}^{\,\gamma} + \sum_{\substack{\text{accepted}\\ \tau\text{-leptons}}} \vec{p}_{\mathrm{T}}^{\,\tau} + \sum_{\substack{\text{selected}\\ \text{muons}}} \vec{p}_{\mathrm{T}}^{\,\mu} + \sum_{\substack{\text{accepted}\\ \text{jets}}} \vec{p}_{\mathrm{T}}^{\,\mathrm{jet}} + \sum_{\substack{\text{unused}\\ \text{tracks}}} \vec{p}_{\mathrm{T}}^{\,\mathrm{track}} \right). \tag{3.18}$$

The first five terms, commonly termed the hard component, contain all selected physics objects of one event. As the object reconstructions, discussed in previous sections, are independent of one another other, double counting might occur. Such a case obviously impairs the calculation of $E_{\mathrm{T}}^{\mathrm{miss}}$ a signal ambiguity resolution. Therefore, one tries to mitigate the effect by establishing a priority before adding up the individual objects. The highest priority is assigned to reconstructed electrons while the further order corresponds to that of the terms in Equation (3.18). Lower priority objects are rejected when sharing calorimeter entries with a higher priority object. In contrast to other objects, muons are hardly affected by this overlap removal as an outcome of their clear signature [104].

After adding up these first terms, some detector signals that also originate from the hard scattering vertex remain. They are not assigned to a reconstructed object and instead summarised by the last term, known as the soft component. Besides the poor energy resolution of jets, this particular soft component is hard to pinpoint as it primarily consists of low momentum pions which are difficult to measure [105]. A detailed validation of the $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction in ATLAS is given in [104].

CHAPTER 4

# Statistical methods

In everyday life, we all encounter processes governed by their underlying stochastic nature, be it in modelling of a pandemic or simply when playing a board game. As a result, at the very latest after playing our first round of "Mensch ärgere Dich nicht" [106], we are accustomed to statements like "The dice are loaded". In most cases, such a statement is purely based on a handful of bad or good dice rolls but still sufficient to cast doubt on the hypothesis of a fair dice. In spite of your board game group accepting such statements, any educated statistician would demand a larger sample of rolls before he lets himself be drawn into such a statement.

A variety of such statistical processes also affects the field of particle physics, e.g. in the production or decay of particles following collisions. Similar to the dice roll example, any particle physics experiment usually starts with a hypothesis. This hypothesis might specify how many particles of a certain type should be produced. Unlike a statement at a board game evening with friends, a scientific evaluation of such an experiment can obviously not rely on a handful of results or on your subjective gut feeling. To ensure this, methods of statistical inference usually take a central role in every field of exact science. They form a statistical toolkit that can be utilised to estimate parameters based on data and objectively test initial hypotheses.

Having explained the experimental production of data in the previous chapter, the following sections now introduce all statistical tools required to analyse this dataset. Section 4.1 begins with a general introduction of the relevant quantities and their definitions. Subsequently, Section 4.2 outlines how a model's parameters can be adjusted to optimally describe a given dataset using the profile likelihood method. Building on this, Section 4.3 presents methods suitable to test a certain hypothesis.

Finally, and slightly detached from the previous parts, Section 4.4 briefly describes how MC simulations of particle collisions can be generated. These simulations, based on statistical assumptions, are a key part of most particle physics analyses as they allow a comparison of real data and any theoretical model.

For a better understanding of all methods presented in this chapter, a simple example is given wherever possible. Should the reader desire further information, please refer to [107]. An advanced description with direct reference to experiments at the LHC can be found in [108].

## 4.1 Fundamental concepts

Prior to the discussion of concrete statistical methods, the focus in this section is placed on introducing fundamental concepts. Over the course of both previous chapters, it has already been stated that processes such as the production and decay of particles are random in nature and each possible outcome can be assigned a probability. This directly raises the question of what exactly a probability is. In purely mathematical terms, the probability $p$ of an event $x$ can be defined using the three Kolmogorov axioms [109]:

1. $p(x) \geq 0$: the probability of event $x$ corresponds to a real non-negative number.
2. $p(X) = 1$ with $x \subset X$.
3. $p(x \cup y) = p(x) + p(y)$ must apply for all mutually exclusive[1] $x$ and $y$.

On the basis of these fundamental axioms a number of interpretations can be derived, to provide a definition of what the exact value of a probability represents.

One interpretation of probability, particularly common in particle physics, considers probability as the limit of frequency. This implies, that if an experiment is repeated $N$ times under identical conditions, the fraction $N(x)$ of cases resulting in outcome $x$ tends to a limit $p(x)$ [107]:

$$p(x) = \lim_{N \to \infty} \frac{N(x)}{N} \,. \tag{4.1}$$

The probability of obtaining a continuous random variable $x$ within a certain interval $[a, b]$ is expressed by integrating over the probability density function (pdf)[2] $f(x)$:

$$p(a \leq x \leq b) = \int_a^b f(x)\,\mathrm{d}x \,. \tag{4.2}$$

Here, $f(x)$ is a real, non-negative function which is integrable and satisfies

$$\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1 \,. \tag{4.3}$$

When trying to describe a pdf's shape more precisely, one often relies on its central moments

$$m_n = E[(x-\mu)^n] = \int_{-\infty}^{+\infty} (x-\mu)^n f(x)\,\mathrm{d}x \,, \tag{4.4}$$

all of which are centred around the arithmetic mean $\mu$. The mean itself is equivalent to the first moment $m_1$. The second moment, termed variance $V$ (whose square root defines the standard deviation $\sigma$) corresponds to the distribution's width. For a description of higher moments, please refer to [107].

In particular these first two moments play a key role throughout the next sections. Their exact values obviously depend on the associated pdf. To provide examples, two distributions of central importance are defined in the following: the Gaussian distribution and the Poisson distribution. Additionally, the

[1] Mutually exclusive random events cannot occur together.

[2] In the case of a discrete distribution, this role is taken over by a probability mass function which, in contrast to the pdf, directly represents a probability.

binomial distribution is outlined to allow a better understanding of concepts given in Section 4.2 and Section 4.3.

- The Gaussian or normal distribution's pdf

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right), \tag{4.5}$$

  is a symmetrical function defined by mean $\mu$ and variance $V = \sigma^2$.

- The Poisson distribution

$$f(x;\nu) = \frac{\nu^x}{x!} \exp(-\nu), \tag{4.6}$$

  represents the probability that $x$ events are measured in an experiment, expecting a mean number of events given by $\nu$. The mean simultaneously also represents the variance. This distribution plays an important role in particle physics, especially in the visualisation of variables in a histogram. Ultimately, the content of each bin corresponds to a Poisson distributed counting experiment.

- Finally, the discrete binomial distribution is defined by

$$f(k;n,p) = \left(\begin{array}{c} n \\ k \end{array}\right) p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}, \tag{4.7}$$

  for any random experiment with just two possible outcomes. It describes the probability of obtaining an event with intrinsic probability $p$ in $k$ out of $n$ trials. Mean and variance are given by $\mu = np$ and $V = np(1-p)$, respectively.

Building on these concepts, the following two sections introduce the key statistical concepts, i.e. the estimation of parameters as well as the validation of hypotheses.

## 4.2 Parameter estimation

Almost every analysis in particle physics involves the estimation of a model's parameter along with its uncertainty based on experimental data, colloquially referred to as fitting. Parameter estimation in this thesis is carried out by the method of maximum likelihood, whose general concept is introduced in the current section.

The fundamental principle of maximum likelihood estimation is quite easy to grasp. Provided the assumption, that a coin flip is binomially distributed, let $p$ be the probability for this coin to land on heads. To infer $p$, one can use an experiment of 100 coin flips. If, in this sample, heads were obtained in 50 out of 100 tosses, our intuition tells us to expect a value of $p = 1/2$. Estimating $p$ through the

likelihood function $L$[3], generally defined as

$$L(\mu;\, x_i) = \prod_{i=1}^{N} f(x_i;\, \mu)\,, \tag{4.8}$$

confirms this expectation as shown in the following. Here $x = (x_1, x_2, ..., x_N)$ represent the measured quantities while $\mu$ denotes the parameter of interest.

Utilising the pdf $f$, the likelihood provides a measure of the extent to which a dataset can be modelled by a certain parameter $\mu$ or a set of parameters $\vec{\mu}$. In the specific example, the likelihood is defined through the discrete binomial distribution of form

$$L(p; n = 100, k = 50) = \begin{pmatrix} 100 \\ 50 \end{pmatrix} p^{50}(1-p)^{100-50}\,. \tag{4.9}$$

The distribution is shown as a function of $p$ in Figure 4.1(a). As expected, the maximum lies at $p = 1/2$. Although the exact value of the likelihood has no direct meaning, its global maximum marks the parameter's maximum-likelihood estimator $\hat{\mu}$, which optimally describes the observed data assuming the selected model is correct.

In practice, determining the minimum of the negative log-likelihood

$$-\ln L(\mu; x_i) = -\sum_{i=1}^{N} \ln f(x_i;\, \mu)\,, \tag{4.10}$$

instead of the maximum is more convenient[4]. Nevertheless, the correct minimisation is usually highly complex and therefore carried out by numerical methods. In high energy physics the MINUIT2 [110] analysis tool is typically chosen to perform the minimisation. All likelihood fits in this thesis are carried out using the TRExFitter [111] framework based on HistFactory [112] which internally also relies on MINUIT2.

In order to determine the uncertainty on $\hat{\mu}$, one can exploit the fact that the likelihood for repeated experiments approximates a Gaussian distribution. The standard deviation $\sigma_{\hat{\mu}}$ of this Gaussian curve, i.e. the uncertainty, corresponds to the points where $-\ln L$ increases by $1/2$ with respect to its minimum [107]:

$$-\ln L(\hat{\mu} \pm \sigma_{\hat{\mu}}) = -\ln L(\hat{\mu}) + \frac{1}{2}\,. \tag{4.11}$$

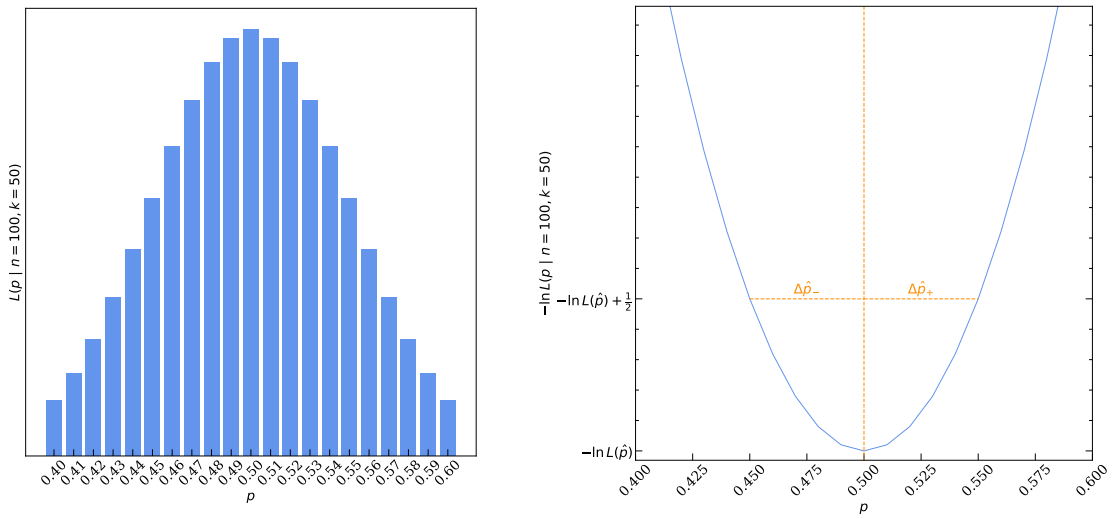
The contour of the negative log-likelihood therefore allows a direct uncertainty determination, as visualised in Figure 4.1(b) for the coin toss example.

### 4.2.1 Profiling the likelihood

A function chosen to model a dataset usually requires additional free parameters besides the parameter of interest $\mu$. A Gaussian distribution, to give an example, requires two parameters to be fully

[3] To avoid confusion at this point, the reader is reminded that in this thesis $\mathcal{L}$ denotes the Lagrangian while L denotes the luminosity.

[4] By applying the logarithm, the sum replaces the product, which is usually simpler to work with in numerical algorithms.




(a) (b)

Figure 4.1: Illustrated function showing (a) the binomial likelihood $L$ and (b) the negative log-likelihood $-\ln L(\mu \mid x_i)$ corresponding to the coin toss example. The $1\sigma$ uncertainty interval on the maximum-likelihood estimate $\hat{p} = 0.5$ is indicated through $\Delta\hat{p}_-$ as well as $\Delta\hat{p}_+$.

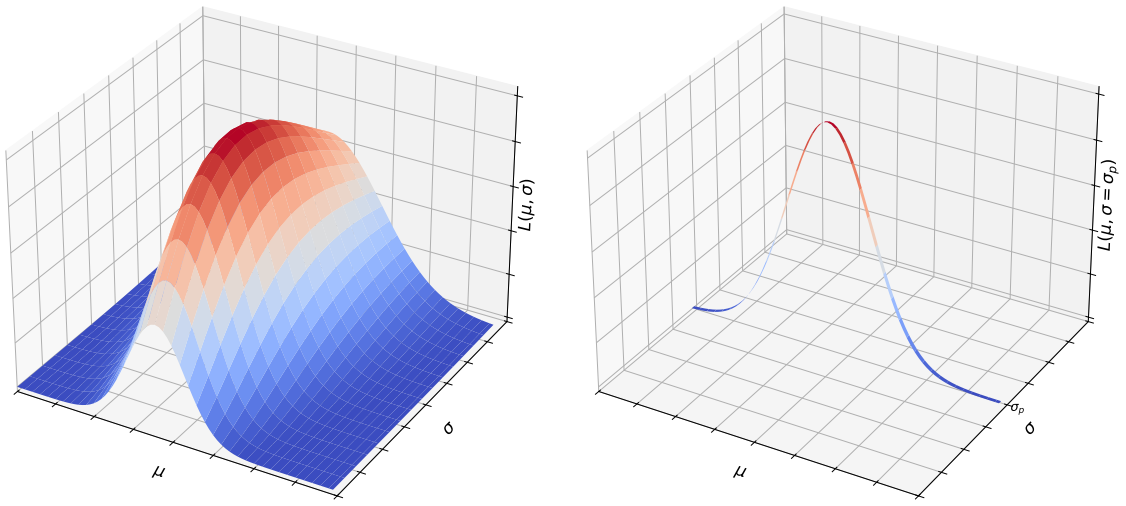
described, the mean and the standard deviation. These additional parameters $\theta$, termed nuisance parameters, obviously increase the complexity and are thus often replaced by their estimators before minimising $-\ln L$. In the case of the Gaussian function, this estimator could simply correspond to the standard deviation of a concrete dataset.

This process is termed profiling. Effectively it reduces the likelihood's dimensionality. Figure 4.2 illustrates this profiling in case of a Gaussian function. After fixing the nuisance parameter $\sigma$, one slice of the original distribution remains. A subsequent minimisation of $-\ln L$ now only depends on $\mu$.

In this thesis, the parameter of interest $\mu$ corresponds to the signal strength of the $tH$ process. A value of one is equivalent to the nominal Standard Model prediction, while $\mu = 0$ represents the complete absence of $tH$ events, termed background-only hypothesis. The nuisance parameters correspond to a large number of experimental and theoretical systematic uncertainties, a listing and explanation of which can be found in Appendix D. A nuisance parameter's value and uncertainty were determined prior to this analysis via auxiliary measurements. By expanding the likelihood function using Gaussian constraint terms, shifting a nuisance parameter away from its nominal value during minimisation penalises the fit (see Section 10.1).

### 4.2.2 Binned likelihood

As stated in Section 4.1, many measurements in particle physics can be viewed as being Poisson distributed counting experiments. Typically, the value of a kinematic variable $x$ is determined for each measured event and subsequently graphically represented in a histogram with $N$ bins [113]. The




(a) (b)

Figure 4.2: Likelihood function of a Gaussian model (a) before and (b) after profiling by setting the nuisance parameter to $\sigma = \sigma_p$.

expectation value for the event count in bin $n_i$ is given by

$$E\left[n_i\right] = \mu s_i + b_i\,, \tag{4.12}$$

where $s_i$ and $b_i$ are the mean signal and background event counts in this bin defined as

$$s_i = s_{\text{tot}} \int_{\text{bin } i} f_s(x;\vec{\theta}_s)\,\mathrm{d}x\,, \tag{4.13}$$

$$b_i = s_{\text{tot}} \int_{\text{bin } i} f_b(x;\vec{\theta}_b)\,\mathrm{d}x\,. \tag{4.14}$$

The mean number of signal and background events summed over all bins are denoted by $s_{\text{tot}}$ and $b_{\text{tot}}$ while $\vec{\theta}_{s,b}$ incorporates all nuisance parameters. By integrating the pdf within bin $n_i$, the expected fraction of $s_{\text{tot}}$ and $b_{\text{tot}}$ can be derived. Additional auxiliary measurements performed in bins $M$ can help to constrain the set of nuisance parameters. The resulting binned likelihood function is expressed via a product of Poisson probabilities [113]:

$$L(\mu,\vec{\theta}_{s,b}) = \prod_{i=1}^{N} \frac{(\mu s_i + b_i)^{n_i}}{n_i!} \exp\left(-(\mu s_i + b_i)\right) \prod_{j=1}^{M} \frac{u_j^{m_j}}{m_k!} \exp(-u_j)\,. \tag{4.15}$$

with $E\left[m_j\right] = u_j(\vec{\theta}_{s,b})$. All parameter estimations discussed in Chapter 10 are ultimately based on this special binned form of the likelihood.

## 4.3 Hypothesis tests

As indicated at the beginning of this chapter, hypotheses have to be validated with objective and unbiased tests as opposed to personal instinct. In simple examples such as verifying the fairness of a coin toss, the first steps of these tests are still quite intuitive. If the coin toss is suspected to be biased towards heads, the logical step is to repeat the experiment over and over while noting the outcome. This is followed by the more complicated part, as shown in the following example: we assume that a coin was tossed 200 times. In case heads come up 101 times, hardly anyone would deem the coin biased, whereas if it lands on heads 115 times, opinions would presumably already differ.

For any coin, there is a point at which it is not tossed anymore but returned to its purpose: serving as a currency. This obviously raises the question how this point can be determined. Hypothesis tests allow exactly this validation based on a dataset. However, it is important to note that even a hypothesis test can never actually prove any hypothesis, it can only judge whether there is enough evidence to infer that, for example, a coin is biased.

Sticking to the example of measuring heads 115 times, the null hypothesis $H_0 : p_{\text{heads}} = 1 - p_{\text{tails}} = 0.5$ would assume a fair coin with $p_{\text{heads, tails}}$ denoting the probabilities of obtaining heads and tails, respectively. The search for evidence that the coin is biased, regardless of which side is favoured, is summarised in an alternative hypothesis $H_1 : p_{\text{heads}} \neq 0.5$.

The pdf $f(t; H_0)$ under the assumption that $H_0$ is correct can be described by a binomial function of the form

$$f(t; 0.5, 200) = \left( \begin{array}{c} 200 \\ t \end{array} \right) 0.5^t (1 - 0.5)^{200-t} . \tag{4.16}$$

The corresponding expectation value lies at $\mu = np = 0.5 \cdot 200 = 100$ and the standard deviation $\sigma = \sqrt{np(1-p)} = \sqrt{200 \cdot 0.5(1 - 0.5)} > 3$ fulfils the Laplace condition[5], allowing the pdf to be approximated by a Gaussian curve.

To quantify the experimentally obtained result a test statistic $\hat{t}$ is introduced. This test statistic represents a measure of how much the actual data deviate from $H_0$ [107]. If the difference between $\mu$ and $\hat{t}$ is found to be larger than a critical threshold, $H_0$ is rejected. The exact choice of this critical value $t_c$ is ultimately arbitrary and generally defined using the significance level $\alpha$. Assuming the pdf $f(t; H_0)$ to be continuous, both quantities are related by

$$\int_{t_{c_1}}^{t_{c_2}} f(t; H_0) \, \mathrm{d}t = 1 - \alpha \, , \tag{4.17}$$

in case of a two-sided test while

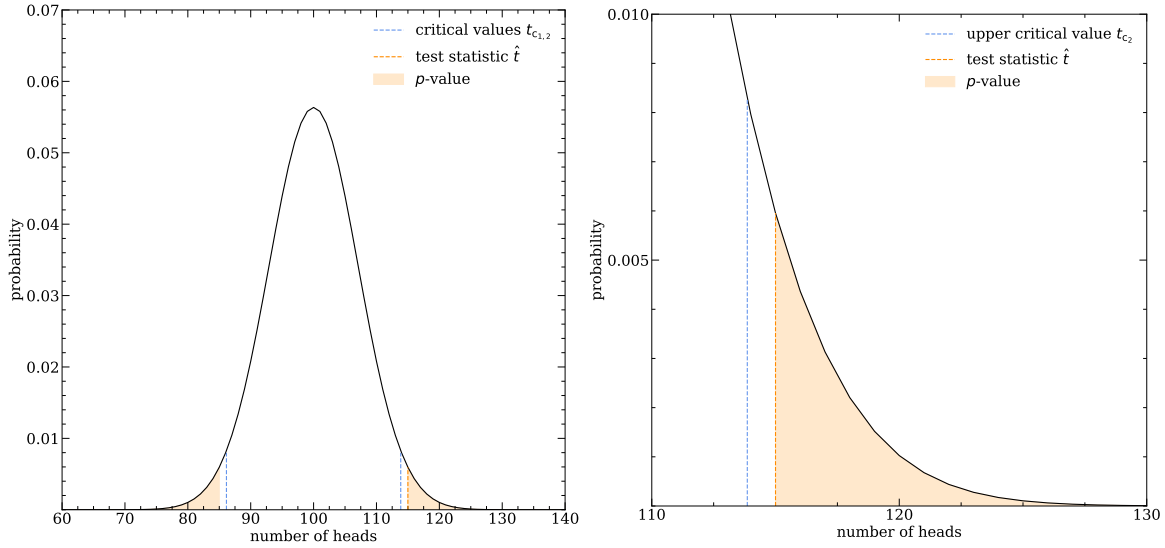
$$\int_{t_c}^{\infty} f(t; H_0) \, \mathrm{d}t = \alpha \, , \tag{4.18}$$

applies in the one-sided case.

In the presented example, a significance level of $\alpha = 0.05$ is chosen, which corresponds to critical

[5] The Moivre-Laplace theorem states that a binomial distribution for $t \to \infty$ and $0 < p < 1$ converges towards the normal distribution. The approximation is considered sufficient if the Laplace condition $\sigma > 3$ is fulfilled.

values of $t_{c_{2,1}} = \mu \pm 1.96\sigma \approx 100 \pm 7$[6]. Given the obtained test statistic $\hat{t} = 115 > t_{c_2}$, the null hypothesis is rejected. The percentage of repeated measurements that would yield test statistics at least as extreme as $\hat{t}$ is characterised through the $p$-value. As an overview, the concrete values corresponding to the example are visualised in Figure 4.3.




(a) (b)

Figure 4.3: Test statistic $\hat{t}$, critical values as well as the $p$-value corresponding to the coin flip example shown for (a) a number of heads between 60 and 140 and (b) zoomed into the range around the upper critical value. The pdf is assumed to be Gaussian.

Even though $H_0$ was rejected, a quick look at the pdf in Figure 4.3 indicates that $\hat{t} = 115$ could also be obtained with an unbiased coin. Therefore, a hypothesis test is never a proof. Generally, a distinction is made between two different types of false conclusions

- Type I error: $H_0$ is falsely rejected. The probability of this error is represented by the significance level $\alpha$.
- Type II error: $H_0$ is failed to be rejected even though it is incorrect.

In particle physics, very strict limits are applied when performing hypothesis tests to avoid false statements caused by a type I error. To declare a discovery of new physics, a significance level of $\alpha = 2.87 \times 10^{-7}$ corresponding to a $5\sigma$ window is required [108].

In contrast to the coin toss example, the test statistic used in this thesis is formed by the likelihood ratio

$$t_\mu = -2 \ln \lambda(\mu) = -2 \ln \frac{L(\mu, \hat{\hat{\theta}})}{L(\hat{\mu}, \hat{\theta})}, \tag{4.19}$$

[6] The integral of a Gaussian distribution in the interval $[\mu - 1.96\sigma, \mu + 1.96\sigma]$ corresponds to roughly 95 % of the total integral.

where $\hat{\hat{\theta}}$ denotes the value of all nuisance parameters that maximise $L$ at the null hypothesis value $\mu$. In contrast, the denominator represents the maximised unconditional likelihood by using the maximum likelihood estimators $\hat{\mu}$ and $\hat{\theta}$ [113]. Despite the fact that the exact value of each likelihood is not directly interpretable, as discussed in Section 4.2, the ratio bound to $0 \leq \lambda(\mu) \leq 1$ still permits identifying which parameter receives more support given a specific dataset. A value of $\lambda(\mu) = 1$ obviously implies the best agreement.

In analogy to Equation (4.18) the $p$-value associated with the observed test statistic $t_{\mu,\,\text{observed}}$ is defined as

$$p = \int_{t_{\mu,\,\text{observed}}}^{\infty} f(t_\mu; \mu)\,\mathrm{d}t_\mu\,. \tag{4.20}$$

Depending on the exact type of hypothesis, the applied test statistic is slightly adjusted. The resulting cases are briefly explained below, closely following the detailed description provided in [113]:

- Test statistic $\tilde{t}_\mu$ for $\mu \geq 0$: a hypothesised signal can be modelled by a purely positive signal strength $\mu$ if an increase in the event rate is expected in any case. Should $\hat{\mu} < 0$ correspond to the maximum likelihood estimator, $\mu = 0$ receives the most support from a given a dataset. Therefore, one defines

$$\tilde{t}_\mu = -2\ln\tilde{\lambda}(\mu) = \begin{cases} -2\ln\frac{L(\mu,\hat{\hat{\theta}})}{L(0,\hat{\theta})}\,, & \hat{\mu} < 0\,, \\ -2\ln\frac{L(\mu,\hat{\hat{\theta}})}{L(\hat{\mu},\hat{\theta})}\,, & \hat{\mu} \geq 0\,. \end{cases} \tag{4.21}$$

- Test statistic $q_0$ for rejection of $\mu = 0$: still assuming $\mu \geq 0$, the special case in which the hypothesis $\mu = 0$ is rejected directly leads to the discovery of a new signal. In this case, only $\hat{\mu} > 0$ is considered to indicate a lack of agreement between data and the background-only hypothesis. Consequently, one obtains

$$q_0 = \begin{cases} -2\ln\frac{L(0,\hat{\hat{\theta}})}{L(\hat{\mu},\hat{\theta})}\,, & \hat{\mu} \geq 0\,, \\ 0\,, & \hat{\mu} < 0\,. \end{cases} \tag{4.22}$$

- Test statistic $q_\mu$ for upper limits: in certain scenarios, a signal cannot yet be discovered, but one at least aims to set an upper limit on $\mu$. In these cases, $\hat{\mu} > \mu$ is not viewed as an indication of less compatibility between the hypothesised signal strength and data, so the following applies

$$q_\mu = \begin{cases} -2\ln\frac{L(\mu,\hat{\hat{\theta}})}{L(0,\hat{\theta})}\,, & \hat{\mu} \leq \mu\,, \\ 0\,, & \hat{\mu} > \mu\,. \end{cases} \tag{4.23}$$

## 4.4 Monte Carlo simulation

The previous two sections made use of very simple examples to introduce the concepts of parameter estimation and hypothesis tests. Aided by the binomial distribution a descriptive model along with expectation values was quickly assigned to the coin toss examples. Defining a corresponding model in

particle physics experiments relies on Standard Model predictions and is far from being trivial. In practice, this process partially relies on the generation of simulated data using Monte Carlo (MC) methods, therefore termed MC events.

In general, MC methods refer to a broad group of numerical algorithms that exploit (pseudo-)random numbers to provide an approximate solution or simulate processes [114]. They are frequently applied to tackle problems for which other simulation methods cannot be carried out in a reasonable time. Nevertheless, the accuracy of MC methods likewise increases when allowing more runtime. A simple example provided in Appendix C illustrates this effect by obtaining the solution of an integral through MC methods. The simulation of MC events at ATLAS can essentially be broken down into four parts, briefly discussed in the following:

1. Event generation: starting from a proton-proton collision, this initial step simulates the prompt decays of short-lived particles together with all stable particles, actually visible to the detector. First, the total cross-section of the core hard scattering process is obtained by relying on MC methods to perform the numerical integration [115]. Building on this, the complex simulation of QCD effects is added in the form of the parton showers. Radiation off the parton's initial (ISR) and final state (FSR) are taken into account. In addition, other processes such as hadronisation, introduced in Section 2.2.2, or the possibility of multiple parton interactions must be accounted for. Already in this first step, a filtering of events by applying certain kinematic cuts is possible. In this way, a simulation can, for example, be restricted to certain decay channels. A detailed description of these phenomena is beyond the scope of this thesis, so please refer to [115] which serves as an excellent introduction to the topic.

   Within ATLAS, event generation takes place inside the Athena [116] framework into which the externally developed generators are integrated. The latter can be split into general-purpose generators designed to produce complete events, and specialised generators, which usually take over small parts in the production chain to increase the accuracy of a particular process [117]. The generators used to simulate all processes relevant to this analysis are listed in Chapter 6.

2. Detector simulation: during this step, the interaction of all generated stable particles with the ATLAS detector's active and passive components is simulated. For this purpose, the external toolkit Geant4 [118] is commonly utilised as it offers a detailed modelling of the entire detector. Alternatively, a fast, less accurate simulation of the detector response is available via Atlfast3 [119] or the previous version Atlfast2.

3. Digitisation: this step emulates the output of the detector readout alongside the L1 trigger decision. Noise arising from different sources and pile-up effects are taken into account.

4. Reconstruction: finally, the reconstruction of physical objects is conducted as described in Section 3.5. In contrast to real data, simulated events are usually saved with additional information, termed truth objects. These objects represent particles with their true kinematics prior to the addition of distorting detector effects.

In contrast to real data, with unit weights, each raw MC event is assigned an individual weight. Usually, more events of a given process are simulated than expected in the real experiment. The simulation's statistical uncertainty is therefore often not a limiting factor. Each event is consequently weighted down to obtain the correct normalisation. The opposite case, i.e. weighting up, is also conceivable,

especially if the MC simulation is computationally intensive. During generation, however, events with negative weights can also occur, thereby reducing the statistical power [120].

The ATLAS experiment relies on MC simulations in several areas. One important application in the statistical interpretation of this analysis is briefly discussed in the following section.

### 4.4.1 Asimov dataset

While performing a statistical analysis, the expected sensitivity is often inferred. In particle physics, predominantly two quantities are of interest: the median significance of a measurement assuming a certain signal strength $\mu'$ or the median exclusion significance while setting upper limits. To determine this sensitivity $f$, as specified in Equation (4.20), must be known. This pdf can be determined by using computationally intensive sampling techniques. However, as deduced in [113], its parameters can also be obtained from just a representative single dataset, termed the Asimov[7] dataset.

When performing a binned profile likelihood fit while trying to estimate a signal strength $\mu'$, this Asimov dataset must fulfil

$$n_{i,\mathrm{A}} = E[n_i] = \mu' s_i + b_i \,, \tag{4.24}$$

in each bin $n_i$. This implies that the number of Asimov events per bin $n_{i,\mathrm{A}}$ has to be equal to the expectation values $E[n_i]$.

In an analysis that searches for physics predicted by the Standard Model, such as $tH$, the selected signal strength corresponds to the nominal value $\mu' = 1$. In this case, the Asimov dataset can be generated by simply using the simulated MC events. A profile likelihood fit based on an Asimov dataset is carried out at the beginning of Chapter 10. Its results allow the measurement's expected sensitivity to be estimated before fits to real data are performed.

[7] The name is inspired by the science fiction short story *Franchise* written in 1955 by Isaac Asimov. The plot centres around a computer which conducts the U.S. presidential elections after surveying a single person.

CHAPTER 5

# Machine learning concepts

Just as the industrial revolution led to the automation of simple physical work, machine learning (ML) methods are developed on the premise of automating simple mental tasks. Recent developments, particularly in the area of large language models[1], indicate that these methods are still in their early stages of development, as their full potential is far from being discovered. Nevertheless, they already have an enormous impact on everyday life. Whether it is Spotify's algorithm for recommending songs and audiobooks [122] or pattern recognition techniques used in clinical oncology to diagnose cancer [123].

What sets ML apart from classical algorithms can be aptly described by quoting the French software engineer François Chollet, "Traditionally, software engineering combined human created rules with data to create answers to a problem. Instead, ML uses data and answers to discover the rules behind a problem". ML algorithms allow the computer to learn autonomously, aided by data, while automatically improving its decision criteria.

At this point, the reader will rightfully ask where the link between ML methods and a particle physics analysis lies: first of all, ML algorithms usually benefit from being trained on large datasets in order to tune their predictions. Large scale particle physics experiments, as carried out at the LHC, are easily able to generate such datasets. Along with the corresponding MC simulations, they provide an ideal playground for testing and developing ML techniques. In addition, certain steps in the physics analysis pose problems whose solution and underlying laws are not simple to grasp. One common example where ML algorithms excel is the separation of the targeted signal process from backgrounds. As a result, significance is increased which is otherwise hard to obtain in the search for rare processes. One such process is the presented $tH$ channel.

The specific branch of ML techniques chosen to tackle this problem is supervised learning. The term supervised refers to methods in which the algorithm receives data assigned with correct labels (called ground truth or target), in this case represented by MC simulations. Based on these labelled datasets, the algorithm is able to improve its prediction and finally make more accurate decisions for new, unseen datasets associated with the same underlying problem.

The specific algorithms employed in this thesis are artificial neural networks (NNs). Their structure and learning process is first introduced in Section 5.1 on the basis of a simple binary classifier. Building on this, Section 5.2 discusses parameters and techniques crucial when setting up a NN to

[1] Large language models are special ML models capable of understanding and generating human language [121].

optimally solve a specific problem. In addition, an overview of validation methods and their metrics is given. Finally, Section 5.3 demonstrates how a binary classifier can be extended to carry out multidimensional classification as utilised in Chapter 8.

As mentioned at the start of this chapter, the field of ML is quite broad and constantly evolving. Since the following sections only cover one specific algorithm, please refer to [124] for a general overview. In addition, an ingenious visual representation of a NN's working principle can be found in [125].

## 5.1 Neural network based binary classification

In the following section, the structure of a NN is described using the example of a binary classifier, which provides discrimination between two classes based on a set of input variables. With regard to particle physics, this often corresponds to the separation of signal and background events.

Generally, NNs can be viewed as a biologically motivated method to analyse the underlying patterns of a dataset [126]. Human neurons are responsible for the transmission of stimuli in the human body and transmit signals in the form of electrical impulses. Each neuron is linked to a multitude of other neurons via synapses to enable highly complex processing. In order to trigger transmission, an external stimulus must first exceed a certain threshold value, which can be reached by summing up various signals.

Similarly, the NN's artificial neurons, are linked to each other to mirror this biological processing and summation of inputs. The actual setup of all NNs trained in this thesis is based on Google's TensorFlow [127], an open source software library for ML, using the Keras [128] Python interface.

### 5.1.1 Neural network structure - nodes and layers

The aforementioned neurons, commonly named nodes, represent the key element of a NN and are arranged in a user defined number of layers. The first layer, referred to as input layer, consists of a number of nodes that exactly correspond to the number of input variables. These variables are subsequently propagated through all layers before ultimately reaching the output layer. The interconnecting layers between input and output, termed hidden layers, are formed by an, again, user defined number of nodes. The actual connection between layers is obtained by linking each node in one layer to all nodes of both neighbouring layers. The final output layer completes the NN and is formed by just a single node in case of a binary classifier. This node typically provides an output value between zero (background event in the physics use case) and one (signal event in the physics use case) to ensure classification. In general, however, the output layer can also be composed of several nodes, as discussed in Section 5.3. As an example, the structure of a binary classifier with three inputs and two hidden layers is shown in Figure 5.1.

Internally, the $m$th node in the $l$th layer represents a scalar value $a_m^l$. In case of the first layer these scalars correspond to the provided input variables. Each link between the $m$th node of the $l$th layer with the $n$th node of the $(l+1)$th layer corresponds to a weight parameter $w_{mn}^{l+1}$. The sum of all these weights can additionally be shifted by a constant value, the bias $b_n^{l+1}$, resulting in a new single scalar

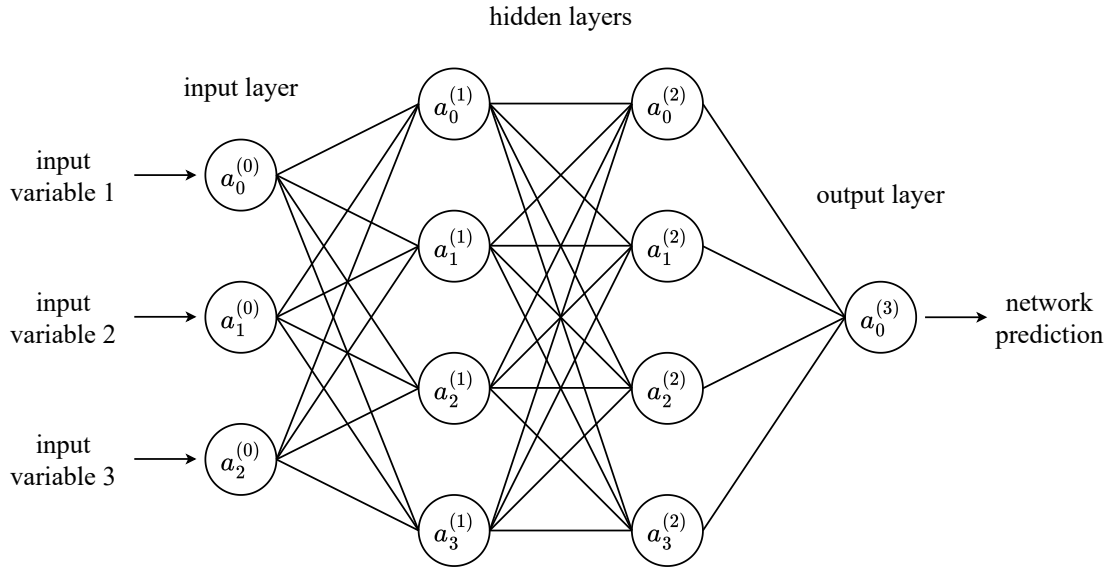



Figure 5.1: Sketch of a binary classifier NN with three input variables, two hidden layers with four nodes and a single node output layer.

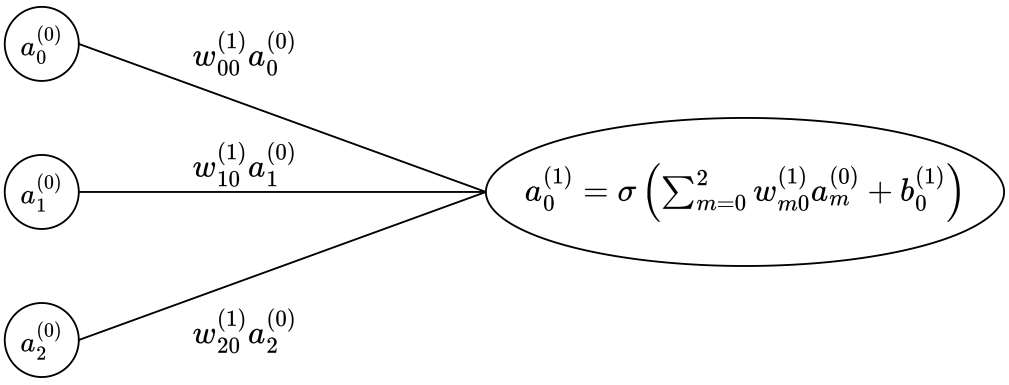



Figure 5.2: Exemplary node calculation using three nodes of the 0th layer $a_m^{(0)}$ to determine the value a first layer node $a_0^{(1)}$ via their weights $w_{m0}^{(1)}$ and bias $b_0^{(1)}$ as the result of an activation function $\sigma$.

value

$$c_n^{l+1} = \sum_{m=0}^{M-1} w_{mn}^{l+1} a_m^l + b_n^{l+1}\,, \tag{5.1}$$

serving as an input to the $(l+1)$th layers' $n$th node. Here $M$ denotes the total number of nodes in the $l$th layer. The final scalar value $a_n^{l+1}$, represented by the $n$th node in the $(l+1)$th layer, is ultimately obtained by using $c_n^{l+1}$ as input to an activation function $\sigma$:

$$a_n^{l+1} = \sigma\left(c_n^{l+1}\right) = \sigma\left(\sum_{m=0}^{M-1} w_{mn}^{l+1} a_m^l + b_n^{l+1}\right)\,. \tag{5.2}$$

An activation function usually not only restricts a node's value to a certain interval but can also introduce a non-linearity. The latter greatly helps when trying to adequately model complex non-linear problems. A summary of this calculation for four nodes is depicted in Figure 5.2.

All activation functions relevant in this thesis are listed below and additionally depicted in Figure 5.3:

- The exponential linear unit (ELU) given by

$$\sigma(x) = \begin{cases} x\,, & x > 0\,, \\ \alpha(\exp(x) - 1)\,, & x \leq 0\,, \end{cases} \tag{5.3}$$

  with $\alpha > 0$, represents the activation function applied to all hidden layers in this thesis.

- Binary classifiers usually employ the sigmoid function

$$\sigma(x) = \frac{1}{1 + \exp(-x)}\,, \tag{5.4}$$

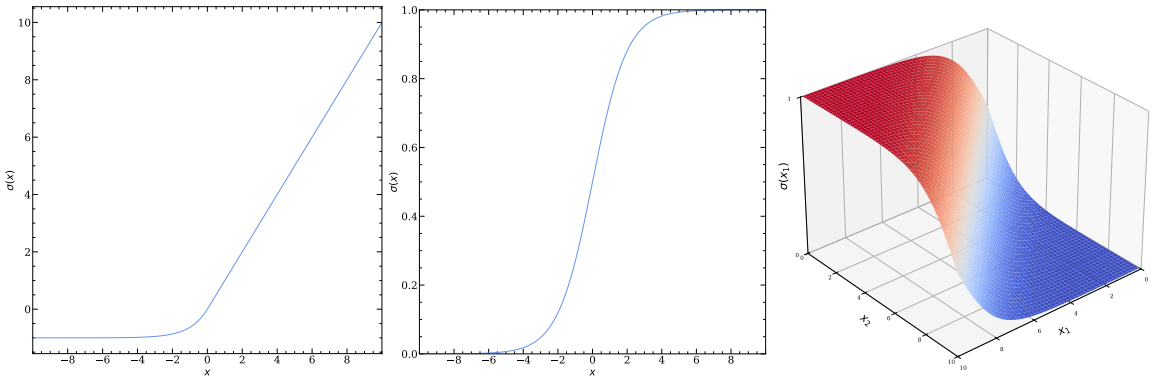
  as their last layer's activation function, since it restricts its own output value to the interval $[0, 1]$.

- If the output of a layer is supposed to represent probabilities, the softmax function

$$\sigma(\vec{x})_i = \frac{\exp(x_i)}{\sum_{j=1}^{n} \exp(x_j)}\,, \tag{5.5}$$

  is a common choice. It transforms an $n$-dimensional vector $\vec{x}$ into the likewise $n$-dimensional vector $\vec{\sigma}$, whose components sum to one. This characteristic is particularly helpful when encountering a multidimensional classification problem as outlined in Section 5.3.

Based on this definition of a NN's architecture along with the underlying mathematical representation, the next section presents the actual learning algorithm's working principle.



(a) (b) (c)

Figure 5.3: Activation functions (a) ELU, (b) sigmoid and (c) softmax plotted for input variables in the range $[-10, 10]$. The softmax function is displayed via its first output component $\sigma(\vec{x})_1$ for a two-dimensional input vector $\vec{x}$.

### 5.1.2 Neural network training - backpropagation and optimisers

In the case of supervised learning, a labelled dataset is required to optimise the network predictions. This implies, that the correct class $y \in Z \cap [0, 1]$ of each individual sample in the dataset is known. Before the optimisation is initialised, all weights and biases in the network are set to random values, which obviously results in a suboptimal prediction. A quantification of this prediction is achieved by comparing the true labels $y$ and predictions $p \in [0, 1]$ through a loss function. For binary classifiers, the binary cross-entropy

$$L_{\text{BCE}} = -(y \log(p) + (1 - y) \log(1 - p)) \,, \tag{5.6}$$

is the most common choice. Its shape, shown in Figure 5.4, causes an exponential increase for large discrepancies between the NN prediction and the true label. Extremely confident but incorrect predictions therefore result in large values. Such a behaviour is desirable as the training's central objective is a minimisation of the loss function by adjusting the weights and biases.

These adjustments are carried out by a class of algorithms, termed optimisers, which generally rely on a determination of the loss function's negative gradient $-\nabla L_{\text{BCE}}$. Updating the network's parameters according to the negative gradient's direction will reduce the loss function's value as rapidly as possible by definition.

Internally, this gradient is calculated via the backpropagation algorithm (first introduced in [129] in its modern form), which computes $-\nabla L_{\text{BCE}}$ starting at the output layer before iterating backwards through the NN's structure. Using the chain rule to obtain the derivative of a composite function, the gradient can be determined for each free parameter, i.e. the scalar node value $a_m^{l-1}$, the weight $w_{mn}^{l}$

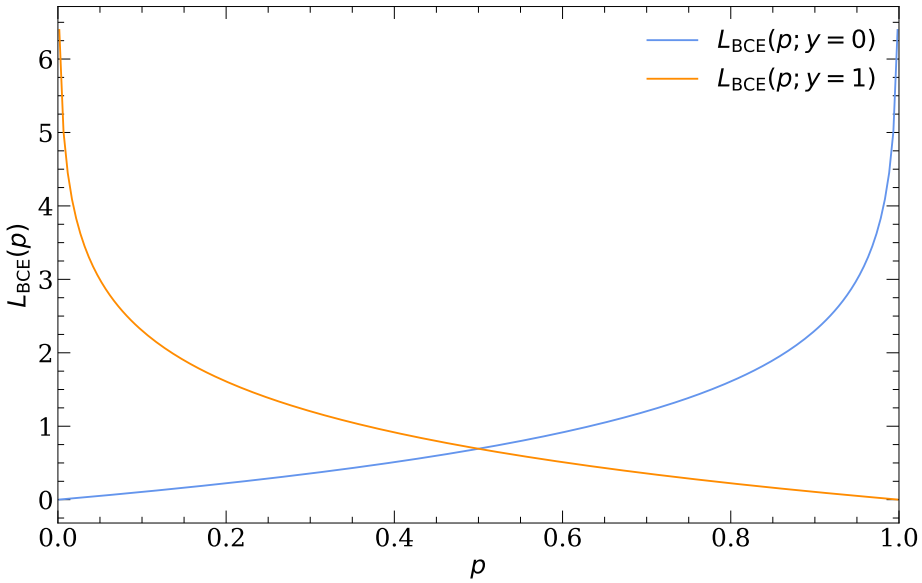



Figure 5.4: Binary cross entropy $L_{\mathrm{BCE}}$ as a function of the prediction $p$ given for true labels $y = 0$ and $y = 1$.

and bias $b_n^l$. For a single sample this results in

$$\frac{\partial L_i}{\partial a_m^{l-1}} = \sum_{n=0}^{N_l - 1} \frac{\partial c_n^l}{\partial a_m^{l-1}} \frac{\partial a_n^l}{\partial c_n^l} \frac{\partial L_i}{\partial a_n^l} , \tag{5.7}$$

$$\frac{\partial L_i}{\partial w_{mn}^l} = \frac{\partial c_n^l}{\partial w_{mn}^l} \frac{\partial a_n^l}{\partial c_n^l} \frac{\partial L_i}{\partial a_n^l} , \tag{5.8}$$

$$\frac{\partial L_i}{\partial b_n^l} = \frac{\partial c_n^l}{\partial b_n^l} \frac{\partial a_n^l}{\partial c_n^l} \frac{\partial L_i}{\partial a_n^l} , \tag{5.9}$$

where $N_l$ denotes the number of nodes in the $l$th layer.

Using these derivations, the chosen optimiser can finally carry out the actual parameter adjustment. A simple example of such an optimiser is the stochastic gradient descent (SGD) [130] which updates the initial parameters $\beta_i$ to $\beta_f$ after validating a single training sample with input $x$ and target $y$ by applying

$$\beta_f = \beta_i - \alpha \nabla L_{\mathrm{BCE}}(x, y, \beta_i) . \tag{5.10}$$

Since SGD updates all parameters after each individual sample, the resulting steps are subject to strong fluctuations. As a countermeasure, more sophisticated optimisers rely on averaging over a certain number of training samples, called batch size, before applying an adjustment. Although determining this average increases the computation time, it leads to a more stable convergence towards the minimum compared to SGD.

The additional factor $\alpha$, usually referred to as learning rate, scales the step size along the gradient's direction and thus influences the training process enormously. Its exact value is always a trade-off:

choosing a large learning rate increases the risk of overshooting the minimum while small learning rates tend to get stuck in local minima or drastically increase training time by slowing down the convergence towards a minimum.

To mitigate this problem, most modern optimisers constantly adjust $\alpha$ during training. A simple option is to decrease the learning rate after each iteration by applying

$$\alpha' = \frac{\alpha}{1+d}\,, \tag{5.11}$$

with a fixed decay rate $d$. As a consequence, the optimisation begins with large steps while slowing down when approaching the minimum, similar to zooming in on a map when searching for an unknown location.

Apart from such a decay rate, one often replaces the gradient with an exponential moving average of current and previous gradients $m_{i,i-1}$ through applying

$$\beta_f = \beta_i + m_i\,, \tag{5.12}$$

with

$$m_i = \gamma \cdot m_{i-1} - \alpha \nabla L(x_i, y_i, \beta_i)\,, \tag{5.13}$$

where $\gamma$ denotes the freely selectable momentum parameter [131]. This way, the step size increases, i.e. builds up momentum, when the gradient points towards the same direction over several iterations.

The optimiser employed in this thesis, known as Adam (acronym for adaptive moment estimation) [132], also makes use of such a momentum parameter. Adam benefits from being able to determine the parameter-dependent learning rate autonomously during training. Its calculation is aided by the gradient's first and second moments through defining

$$\hat{m}_i = \frac{m_i}{1-\delta_1^i} \quad \text{with} \quad m_i = \delta_1 m_{i-1} + (1-\delta_1)\nabla L(x_i, y_i, \beta_i)\,, \tag{5.14}$$

and

$$\hat{v}_i = \frac{v_i}{1-\delta_2^i} \quad \text{with} \quad v_i = \delta_2 v_{i-1} + (1-\delta_2)\left(\nabla L(x_i, y_i, \beta_i)\right)^2\,, \tag{5.15}$$

$$v_t = \delta_2 v_p + (1-\delta_2)\left(\frac{1}{N_b}\sum_{i=0}^{N_b-1}\nabla L(x_i, y_i, \beta_i)\right)^2\,. \tag{5.16}$$

In this case $\delta_{1,2}$ denote user-defined parameters to control the decay rates and correct for potential biases before obtaining $\hat{m}_i$ as well as $\hat{v}_i$. Based on this, the final adjustment of any parameter is specified via

$$\beta_f = \beta_i - \frac{\alpha}{\sqrt{\hat{v}_i}+\epsilon}\cdot \hat{m}_i\,, \tag{5.17}$$

where $\epsilon$ is a small positive constant to prevent dividing by zero.

As discussed, all these optimisers make iterative adjustments to the model's parameters. In most cases, the minimum is not found after iterating just once over each individual sample or batch of a dataset. Therefore, the dataset is scanned several times during training, with each complete iteration being referred to as an epoch.

At this point. the attentive reader will already have noticed that the presented optimisation process enables automatic adjustment of a NN's weights and biases but does not constrain certain characteristics such as the number of nodes and layers used. Similarly, neither the batch size nor the number of iterated epochs are specified. Correctly choosing these parameters, known as hyperparameters, is crucial for any NN and covered in the next section.

## 5.2 Neural network hyperparameter optimisation

In theory NNs are able to autonomously learn patterns of a dataset. Nonetheless, certain values, called hyperparameters, must be set prior to the training process in order to achieve satisfactory results. These hyperparameters cannot be changed during the learning process and are thus distinct in their definition from the constantly updated weights and biases. The most common hyperparameters were previously described: the number of layers and the number of nodes per layer, the activation functions used, the number of epochs and the batch size. Their choice is highly problem-dependent and neither trivial nor generalisable but crucial to obtain a well functioning model. As a consequence, a substantial effort should be invested in optimising the hyperparameters.

Obviously, a metric is needed to determine a chosen set of set of hyperparameter's performance. As discussed in the previous section, an optimal choice of hyperparameters should enable the NN to find a global minimum of the chosen loss function. When using the same dataset, a lower final loss value is therefore an indicator for a more accurate prediction. In addition, its shape should show a decrease over several epochs while remaining stable towards the end of the training. Nevertheless, the loss function is only of limited use for comparing two trained NNs, as its concrete values do not allow an easy and direct comparison. Consequently, another relevant metric to evaluate the predictive power of a classifier is introduced in the following section.

### 5.2.1 Receiver operating characteristic

A binary classifier usually returns a continuous prediction in the interval $[0, 1]$. In the common particle physics context, high values should be more likely to be signal events and vice versa.

Any arbitrary choice of threshold value based on this prediction output allows a sample to be labelled as signal or background. For each specific choice of threshold, the false positive rate (FPR)

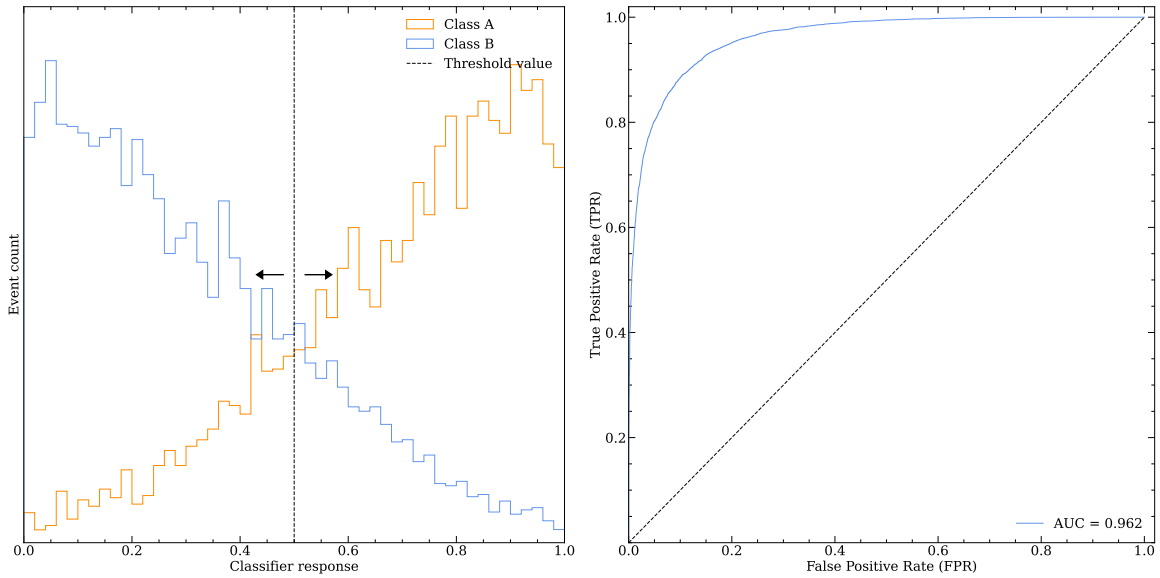
$$\text{FPR} = \frac{\text{False Positives}}{(\text{False Positives} + \text{True Negatives})}, \tag{5.18}$$

and the true positive rate (TPR)

$$\text{TPR} = \frac{\text{True Positives}}{(\text{True Positives} + \text{False Negatives})}, \tag{5.19}$$

can be derived. Varying this threshold within the response interval and calculating FPR and TPR for each point results in an evaluation metric, known as the receiver operating characteristic (ROC) curve,

as depicted in Figure 5.5. Evaluating and comparing different sets of hyperparameters is possible through computing the area under the ROC curve (AUC). Perfect classification results in AUC = 1.0 while random decisions yield AUC = 0.5.




(a) (b)

Figure 5.5: Sketch of (a) the response curve of a binary classifier separating classes A and B and (b) the corresponding ROC curve obtained by varying the depicted threshold value.

Thus, achieving a high AUC value based on training data, represents evidence for a good choice of hyperparameters. However, it cannot provide any information about the network's performance on different datasets belonging to the same problem. In order to clarify this important generalisation capability, a network therefore needs testing on an independent dataset, not used in the training. Two common methods to realise such a test are covered in the next section.

### 5.2.2 Evaluating and optimising generalisation capability

Once a network is optimised, the final model should ideally be able to make predictions for unknown data with similar characteristics as the training samples. To ensure this, the network is evaluated on an independent dataset which was not used for optimisation. Primarily two approaches are used for this task:

- In the first option, an existing dataset is divided into three subsets. The largest dataset serves as a training dataset, when adjusting the model's parameter. Secondly, a validation dataset is formed to enable an unbiased comparison of different hyperparameter configurations. Finally, a model with fixed hyperparameters is evaluated on the remaining test dataset to assess its generalisation. The exact split between all three sets depends on the concrete use case.

- The second approach, named $k$-fold cross-validation, is often chosen when training on small datasets. All available samples are randomly categorised into $k$ groups or folds. Now $k$ models

are trained, whereby model number $k$ uses the $k$th fold as a test set, the $(k+1)$th fold as validation set and the remaining folds for training. This permutation ensures that each individual sample is available for training $k-2$ models while representing a test or validation sample for the remaining two models. The final results are often specified by the mean of all models. A slightly adopted version of this method is utilised in Section 8.1.2.

Correct application of these methods allows clear identification of problems in a model. Arguably the most common problem, known as overtraining, occurs especially in networks with a complexity not justified by the dataset (e.g. small sample size). It is recognised by obtaining significantly better results on training data when compared to test and validation data. An overtrained NN simply memorises the whole training data while not learning the underlying patterns. Thus, it cannot make any reliable predictions for independent datasets.

To prevent such overtraining, various regularisation options are available that can intervene in the training process. One option, used in this thesis, relies on deactivating random nodes with a user-defined probability during each training step. Deactivation corresponds to ignoring a node during weight adjustments. This method is termed dropout [133]. The actual nodes, affected by this deactivation, change after each iteration. As an outcome, reducing the overall co-adaptation between certain nodes and decreasing the network's focus on certain connections. Alternative countermeasures are summarised in [134].

Nevertheless, there may be cases where overtraining cannot be completely avoided. An indicator is the difference between training and validation loss, which should be monitored and compared after each epoch. In case the validation loss increases, the training can be terminated using the early stopping method before overtraining poses a problem. In order to eliminate the influence of statistical fluctuations, the optimisation is usually only stopped when the loss deteriorates over a certain number of epochs, referred to as patience.

### 5.2.3 Hyperparameter scans

The previous section discussed how the choice of hyperparameters affects a NN's performance and how the latter can be quantified. Finally, it remains to be clarified in which way the best hyperparameters can be found.

While several different methods are available, this thesis utilises a slightly modified version of a grid search to determine the optimal hyperparameters. This corresponds to firstly defining a subspace of values in advance and then finding the optimum out of every possible combination. Obviously, such a grid search is not only a resource-intensive but also a time-consuming process. But since the individual combinations are independent, a parallelisation is feasible given sufficient computing power. In this thesis, such a parallelisation is realised by making use of the BAF2 cluster [135]. A more detailed description of the grid search along with a list of scanned parameters is provided in Section 8.2.2.

## 5.3 Neural network based categorical classification

All previous sections made use of a simple binary classification example to introduce important ML and NN concepts. However, the typical separation of signal and background in any particle physics analysis does not quite match this example. The individual background processes are frequently very

different allowing them to be not only separated from a signal but also from each other. Overall, considering these additional classes can lead to an improved signal isolation.

In order to enable such a separation, a binary NN is extended to a categorical neural network (CNN), which can theoretically separate any given number of classes. To set the CNN up, the number of nodes in its output layer is chosen to match the number of classes. Additionally employing a softmax activation function in this layer allows the prediction of a single node to be interpreted as the probability of an event belonging a certain class.

The training and optimisation of a CNN is carried out by minimising the categorical cross-entropy

$$L_{\text{CCE}} = -\sum_{i=1}^{N} y_i \log(p_i)\,, \tag{5.20}$$

where $y_i$ and $p_i$ denote true label and prediction for each of the $N$ classes. Once a model is trained, the response of each individual output node can be used to generate an individual ROC curve. Such a curve enables the separation between a single class and all others to be evaluated. The exact choice of classes for the concrete application of CNNs in this thesis is discussed in Section 8.1.1.

CHAPTER 6

# Data and simulated Monte Carlo samples

All previous chapters of this thesis laid the foundation to understand the *tH* channel's analysis. Not only was the field of particle physics introduced, but in addition all crucial data analysis methods were covered. Building on this, all subsequent chapters cover each individual analysis step in detail.

To begin with, this chapter is intended to provide a brief overview of the analysed dataset. First, Section 6.1 describes the real data taken by the ATLAS detector, before Section 6.2 outlines the simulation of all individual MC processes.

## 6.1 Data samples

The analysed dataset was recorded by ATLAS during LHC's Run 2 at a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$. Selected events had to pass single-lepton triggers [136], i.e. contain at least one light lepton candidate, as well as a variety of quality criteria. The latter ensured that collisions originated from stable particle beams and were only collected if all individual detector modules performed correctly. Data taking sequences that met these conditions are included in so called good-run lists, summarised in [137].

All collision events available after applying these selections are presented in Table 6.1 for each individual year and in total. Overall, these samples amount to an integrated luminosity of $L_{\text{int}} = (140.1 \pm 1.2)\,\text{fb}^{-1}$ [138].

Table 6.1: Available data samples of each individual year and for the complete LHC Run 2. Integrated luminosity $L_{\text{int}}$ is given alongside the corresponding event count.

| Year | Number of events | $L_{\text{int}}$ [pb$^{-1}$] |
|---|---|---|
| 2015 | $220.58 \times 10^6$ | $3\,244 \pm 37$ |
| 2016 | $1\,057.84 \times 10^6$ | $33\,400 \pm 300$ |
| 2017 | $1\,340.80 \times 10^6$ | $44\,630 \pm 500$ |
| 2018 | $1\,716.77 \times 10^6$ | $58\,790 \pm 650$ |
| Combined | $4\,335.99 \times 10^6$ | $140\,100 \pm 1\,200$ |

## 6.2 Simulated Monte Carlo samples

Besides real data, many steps of the analysis rely on the use of simulated MC events, introduced in Section 4.4. Therefore, all relevant processes alongside technical aspects of their generation are outlined in this section. Guided by the two $tH$ signal processes, $tHq$ and $tWH$, background processes with a similar signature were simulated. This comprises, in addition to the major backgrounds listed in Section 2.3, the following processes:

- Triboson production.
- Single top-quark $s$-channel production.
- The production of three ($t\bar{t}t/t\bar{t}\bar{t}$) as well as four ($t\bar{t}t\bar{t}$) top-quarks.
- Higgs-boson production via gluon-gluon fusion ($ggH$) and weak-boson fusion ($qqH$).
- Higgs-Strahlung, i.e. Higgs-boson production in association with $W$- ($WH$) and $Z$-bosons ($ZH$).

As discussed in Section 7.1.1, these additional backgrounds do not contribute significantly to the analysis region but are still considered to ensure a good agreement between data and simulation. Other processes are negligible as they are not expected to pass the preselection of the $tH$ analysis. The specifics of this preselection are discussed in Chapter 7.

The actual generation of the respective simulations was not part of this work. Instead, all MC simulations were provided centrally by the ATLAS collaboration, which facilitates the analysis. If possible, simulations that utilise the complete GEANT4 detector response are chosen. However, for processes where this full simulation is not available, as in the case of $tHq$ and $tWH$ events, the fast simulation via ATLFAST2 [117] is used instead.

In similar fashion, chosen event generators along with the tools to model parton showers are also process-dependent. The resulting selection is summarised in Table 6.2 along with a specification of each utilised PDF. A deep knowledge of how the individual generators work is not essential at all to understand the presented analysis. Their listing simply serves the purpose of providing a complete documentation and thereby eases potential reproducibility. Interested readers are referred to [139] for more details.

Generating all these simulations is a very resource intensive procedure. Nevertheless, each analysis benefits greatly when increasing the number of simulated events as statistical uncertainties are reduced.

In contrast, the second source of uncertainties, i.e. systematic uncertainties, is not reduced by simply enhancing the MC event production as it introduces a one-sided trend or offset to a measurement. These systematics can have a plethora of origins, be it uncertainties of certain theory parameters or imperfections when modelling the detector. A correct measurement is only feasible if all of them are properly taken into account.

One way to accomplish this is by assessing their effect through simply reweighting the MC events. The uncertainty on $L_{int}$, for example, can be modelled by globally reweighting the nominal sample. As an alternative, parameters that influence a process' modelling are often varied within their uncertainties. This way, additional samples which reflect the resulting effect are generated.

The exact definitions of systematic uncertainties are often highly technical and not directly transferable to other experiments. Therefore, to not hinder the flow of reading, Appendix D provides a summary of all individual systematic effects that are considered in the final parameter estimation.

Table 6.2: Summary of all simulated signal and background event samples used in the $tH$ analysis.

| Process | Generator | Order (scheme) | PDF set | Parton shower | PDF set (tune) |
|---|---|---|---|---|---|
| | | Signal | | | |
| $tHq$ | MadGraph5_aMC@NLO 2.6.2 | NLO (4FS) | NNPDF3.0nlo nf4 | Pythia 8.230 | NNPDF2.3lo (A14 tune) |
| $tWH$ | MadGraph5_aMC@NLO 2.8.1 | NLO (5FS, DR) | NNPDF3.0nlo | Pythia 8.245p3 | NNPDF2.3lo (A14 tune) |
| | | Backgrounds | | | |
| $t\bar{t}$ | Powheg Box v2 | NLO (5FS) | NNPDF3.0nlo | Pythia 8.230 | NNPDF2.3lo (A14 tune) |
| $V$+jets | Sherpa 2.2.1 | NLO+LO | NNPDF3.0nnlo | - | - |
| Diboson | Sherpa 2.2.1-2 | NLO+LO | NNPDF3.0nnlo | - | - |
| Triboson | Sherpa 2.2.2 | NLO+LO | NNPDF3.0nnlo | - | - |
| $t\bar{t}Z$ | MadGraph5_aMC@NLO 2.3.3 | NLO | NNPDF3.0nlo | Pythia 8.210 | NNPDF2.3lo (A14 tune) |
| $t\bar{t}W$ | Sherpa 2.2.10 | NLO | NNPDF3.0nnlo | - | - |
| $t\bar{t}H$ | Powheg Box v2 | NLO (5FS) | NNPDF3.0nlo | Pythia 8.230 | NNPDF2.3lo (A14 tune) |
| Single top-quark $t$-channel | Powheg Box v2 | NLO (4FS) | NNPDF3.0nlo nf4 | Pythia 8.230 | NNPDF2.3lo (A14 tune) |
| $tW$ | Powheg Box v2 | NLO (5FS, DR) | NNPDF3.0nlo | Pythia 8.230 | NNPDF2.3lo (A14 tune) |
| Single top-quark $s$-channel | Powheg Box v2 | NLO | NNPDF3.0nlo | Pythia 8.230 | NNPDF2.3lo (A14 tune) |
| $tZq$ | MadGraph5_aMC@NLO 2.3.3 | NLO | NNPDF3.0nlo | Pythia 8.230 | NNPDF2.3lo (A14 tune) |
| $tWZ$ | MadGraph5_aMC@NLO 2.3.3 | NLO | NNPDF3.0nlo | Pythia 8.212 | NNPDF2.3lo (A14 tune) |
| $t\bar{t}t/t\bar{t}\bar{t}$ | MadGraph5_aMC@NLO 2.2.2 | NLO | NNPDF3.1nlo | Pythia 8.186 | NNPDF2.3lo (A14 tune) |
| $t\bar{t}t\bar{t}$ | MadGraph5_aMC@NLO 2.3.3 | NLO | NNPDF3.1nlo | Pythia 8.230 | NNPDF2.3lo (A14 tune) |
| $ggH$ | Powheg Box v2 | NLO | CT 10 | Pythia 8.210 | CTEQ 6L1 (AZNLO tune) |
| $qqH$ | Powheg Box v1 | NLO | CT 10 | Pythia 8.186 | CTEQ 6L1 (AZNLO tune) |
| $WH$ | Pythia 8.186 | LO | NNPDF2.3lo | - | - |
| $ZH$ | Pythia 8.186 | LO | NNPDF2.3lo | - | - |

CHAPTER 7

# Object definition and event selection

The previous chapter described the analysed dataset alongside the associated MC simulation. Directly performing a parameter estimation with this dataset would not yield any statistically significant results. The haystack that is searched for the needle, i.e. $tH$ events, is simply too large. In order to reduce its size, the initial step of this analysis aims to establish a general event preselection. Any preselection's foundation is a significant reduction of backgrounds while simultaneously requiring little to no change to the signal acceptance. All chosen selection criteria along with the resulting background composition are discussed in Section 7.1.

The final separation of signal and background events on top of this preselection, discussed in Chapter 8, is ensured by NNs. In order to improve their discrimination power, it is advantageous to reconstruct each event signature as accurately as possible. This implies that besides physics objects measured in the detector, all important short-lived particles also need to be reconstructed. In case of the $tH$ channel focus was laid on reconstructing the Higgs-boson and the top-quark. The corresponding methods are explained in the Section 7.2.

## 7.1 Event selection

The intention of this section is to introduce the general preselection for all subsequent analysis steps. Each individual selection should be aimed at reducing the overall background contamination, i.e. achieving a high background rejection. At the same time, the signal efficiency, i.e. the fraction of $tH$ events passing a certain selection, should remain close to one.

The most fundamental of these selections are motivated by considering the final state particles present in $tH(H \to \tau\tau)$ events (cf. Figure 2.16). As discussed in Section 2.3, this analysis focuses on two orthogonal channels depending on the tau-lepton's decay mode[1]. In each channel the following requirements must be met:

- $2\ell + 1\tau_{\text{had}}$: one hadronic tau-lepton ($\tau_{\text{had}}$) must always be detected. In addition, exactly two light leptons ($2\ell$) are required. In $tH$ events, one of them originates from the leptonic tau decay ($\tau_{\text{lep}}$) while the second light lepton is expected to stem from the leptonic $W$-boson decay.

[1] The decay mode involving two leptonic tau-leptons is not part of this work but analysed within the ATLAS collaboration, as noted in Section 2.3.1.

- $1\ell + 2\tau_{\text{had}}$: only events with exactly two $\tau_{\text{had}}$ are accepted. Just as in case of $tH$ events in the $2\ell + 1\tau_{\text{had}}$ channel, an additional light lepton arising from the $W$-boson decay is required.

Adding up tau-leptons and light leptons, each channel yields a total of three leptons. Out of those, the one with the highest transverse momentum, known as the leading lepton, must fulfil $p_{\text{T}} > 27\,\text{GeV}$. In case this lepton is a tau-lepton, the leading light lepton must also fulfil $p_{\text{T}} > 27\,\text{GeV}$. The subleading lepton as well as the softest lepton must pass a slightly relaxed selection of $p_{\text{T}} > 20\,\text{GeV}$ and $p_{\text{T}} > 10\,\text{GeV}$, respectively.

In addition, between two and six jets are required to be reconstructed in each event. Out of those jets, at least one but a maximum of two have to be $b$-jets, which pass the second strictest working point[2] of the DL1r-based jet flavour-tagging [94]. This working point corresponds to correctly tagging 77 % of all $b$-jets in simulated $t\bar{t}$ events. All jets must fulfil $p_{\text{T}} > 20\,\text{GeV}$ and $|\eta| < 4.5$ while $b$-jets are chosen based on stricter selections, specified by $p_{\text{T}} > 20\,\text{GeV}$ and $|\eta| < 2.5$. Furthermore, missing energy is accepted in a range of $5\,\text{GeV} < E_{\text{T}}^{\text{miss}} < 800\,\text{GeV}$, since neutrinos are expected to be produced in the leptonic decays of $W$-bosons and $\tau_{\text{lep}}$.

Besides those requirements, certain quality criteria are imposed on the individual physics objects in the final state. These criteria are designed to reduce the fraction of so-called "fake events", i.e. events with misidentified objects as discussed in Section 2.3.2. In the $tH$ analysis, this includes both fake light lepton and fake-$\tau$ events. The most important associated conditions are listed in the following:

- Selected electron candidates must fulfil $p_{\text{T}} > 10\,\text{GeV}$ and $|\eta| < 2.47$, excluding the interval $1.37 \leq |\eta| < 1.52$ (see Section 3.4.4). Furthermore, an electron not only has to pass the tight identification working point [101] but also the tight isolation criteria [140]. In addition, requirements to reduce electrons with misidentified charge and electrons originating from $\gamma$-conversions are imposed [141, 142].

- Muons, that enter the analysis, must fulfil $p_{\text{T}} > 10\,\text{GeV}$ and $|\eta| < 2.5$. Similar to electrons, the tight isolation working point is chosen as well. However, the muon identification selection is slightly more relaxed by requiring the respective medium working point [103].

- Hadronic tau-leptons have to fulfil $p_{\text{T}} > 20\,\text{GeV}$ and $|\eta| < 2.47$. As for electrons the interval $1.37 \leq |\eta| < 1.52$ is excluded. The amount of jet-faking $\tau_{\text{had}}$ is reduced by requiring each candidate to pass the medium RNN based working point [96].

The combination of all mentioned selections is referred to as the tight preselection. However, for the purpose of estimating fake events, as discussed in Chapter 9, a second slightly less strict selection is introduced. In this loose preselection region, light leptons and $\tau_{\text{had}}$ are only required to pass their respective loose working points. Since all other selections remain unchanged, events passing the tight preselection form a subset of the loose region. As a result, the loose region contains significantly more events than the tight preselection region. This way, statistical uncertainties can be reduced when performing a fake estimation, covered in Chapter 9. To provide a clear overview, the differences between both selections are summarised in Table 7.1.

The definition of these selections is a prerequisite to this analysis and not the main topic of this thesis. A detailed discussion on concrete threshold values can be found in [143]. However, the specific

[2] A working point defines a certain threshold based on a decision criterion such as the response of the RNNs in case of $\tau_{\text{had}}$ candidates. Usually, several working points are provided, commonly labelled as tight, medium and loose in order of decreasing strictness.

adjustments such as discarding events with more than three leptons are exclusive to the presented analysis.

Applying the tight selection results in an expected signal-to-background ratio ($S/B$) of roughly $1.6 \times 10^{-3}$ in case of the $1\ell + 2\tau_{\text{had}}$ channel. In contrast, the $2\ell + 1\tau_{\text{had}}$ channel yields $S/B \approx 4.2 \times 10^{-4}$, which is an order of magnitude lower. To still perform a statistically significant parameter estimation, the $2\ell + 1\tau_{\text{had}}$ channel is further split up based on the light leptons' electric charge as follows:

- $2\ell$ OS + $1\tau_{\text{had}}$: both light leptons are required to have an opposite sign (OS) electric charge. The light lepton with opposite sign charge compared to the $\tau_{\text{had}}$ is associated with the Higgs-boson decay ($\ell_{\text{Higgs}}$). The remaining light lepton is expected to arise from the top-quark decay ($\ell_{\text{top}}$).

- $2\ell$ SS + $1\tau_{\text{had}}$: both light leptons carry a same sign (SS) electric charge. In this case, the assignment is not trivial and therefore ensured by a BDT-based[3] classifier, optimised on MC simulations. Initially, MC truth information is exploited to correctly assign the reconstructed leptons to the Higgs-boson or the top-quark. Using the resulting labelled dataset the BDT classifier can be trained on a set of highly discriminative variables. Cutting on the final classifier's response allows correct assignment in approximately 88 % of all cases [145].

This charge-based division results in a drastic reduction of most dominant backgrounds in case of the $2\ell$ SS + $1\tau_{\text{had}}$ channel and consequently enables an improved measurement.

As an outcome, three distinct channels are available for analysis at this point. Initially, their background compositions are estimated via MC simulation as discussed in the next section.

Table 7.1: Selection requirements imposed on light leptons and $\tau_{\text{had}}$ in order to enter the loose or tight preselection region. The region $1.37 \leq |\eta| < 1.52$ is excluded for electrons and $\tau_{\text{had}}$ in both selections.

| Physics object | Electron ($e$) | | Muon ($\mu$) | | Hadronic tau ($\tau_{\text{had}}$) | |
|---|---|---|---|---|---|---|
| Preselection | Loose | Tight | Loose | Tight | Loose | Tight |
| $p_{\text{T}}$ | $> 10$ GeV | | $> 10$ GeV | | $> 20$ GeV | |
| $\lvert\eta\rvert$ | $< 2.5$ | | $< 2.47$ | | $< 2.47$ | |
| Identification working point | Loose | Tight | Loose | Medium | Loose | Medium |
| Isolation working point | Loose | Tight | Loose | Tight | - | - |

### 7.1.1 Preselection composition

This section is intended to provide an overview of the expected channel compositions after applying the loose or tight selection. All results based on MC simulations are displayed in Tables 7.2 to 7.4. The raw number of available MC events is shown in addition to the important weighted contribution of each process. All background processes that contribute less than one weighted event in the respective channel's tight preselection region are merged in a single "minor backgrounds" category. Depending on the channel, this results in the following merging:

[3] Boosted decision trees (BDTs), just like NNs, represent a branch of ML algorithms which can be utilised to tackle classification problems. See [144] for an introduction.

- $1\ell + 2\tau_{\text{had}}$ minor backgrounds: triboson, single top $s$-channel, $t\bar{t}/t\bar{t}\bar{t}$, $t\bar{t}t\bar{t}$ as well as the Higgs production modes $ggH$, $qqH$, $WH$ and $ZH$.
- $2\ell$ SS + $1\tau_{\text{had}}$ minor backgrounds: triboson, single top $s$- and $t$-channel, $t\bar{t}/t\bar{t}\bar{t}$, $t\bar{t}t\bar{t}$ as well as the Higgs production modes $ggH$, $qqH$, $WH$ and $ZH$.
- $2\ell$ OS + $1\tau_{\text{had}}$ minor backgrounds: triboson, single top $s$- and $t$-channel, $t\bar{t}/t\bar{t}\bar{t}$, $t\bar{t}t\bar{t}$ and $W$ + jets events.

For better visualisation, the composition of all channels is also shown in Figures 7.1 to 7.3, with each individual process ranked according to its relative contribution.

Generally, all three channels show a distinct composition. The $1\ell + 2\tau_{\text{had}}$ channel consists almost exclusively of $t\bar{t}$ events. In the case of the $2\ell$ OS + $1\tau_{\text{had}}$ channel, the most dominant backgrounds are likewise $t\bar{t}$ events with the addition of $Z$ + jets events. In contrast, the $2\ell$ SS + $1\tau_{\text{had}}$ preselection region is formed by a variety of processes, while $Z$ + jets events are almost not present at all. Since the latter are expected to produce two opposite sign leptons, this reduction is a direct effect of the applied charge requirement. This reduction is also reflected when assessing the signal-to-background ratio as well as the signal significance ($S/\sqrt{B}$), summarised in Table 7.5. Consequently, the $2\ell$ SS + $1\tau_{\text{had}}$ channel's tight region with $S/\sqrt{B} \approx 1.6 \times 10^{-1}$ offers the highest signal significance of all listed options. The lowest significance is obtained for both regions of the $2\ell$ OS + $1\tau_{\text{had}}$ channel. However, at the same time this channel provides by far the most events, which greatly reduces statistical uncertainty when performing any parameter estimation.

Before reaching the end of this section, it is important to note that the presented lists of individual processes do not differentiate between prompt events and fake events. The latter category includes all events with at least one fake lepton, be it a light lepton or a $\tau_{\text{had}}$. As an example, one can consider the $t\bar{t}$ process, as displayed in Figure 2.18(g). In its final state only two prompt leptons from the leptonic $W$-boson decays should be present. Therefore, no $t\bar{t}$ event should meet the required preselection criteria. Nevertheless, additional fake leptons cause $t\bar{t}$ events to represent the dominant background across the entire analysis. Consequently, a correct estimation of fake contributions is of great importance when studying the $tH$ process. The actual estimation methods chosen in this thesis are thoroughly explained in Chapter 9.

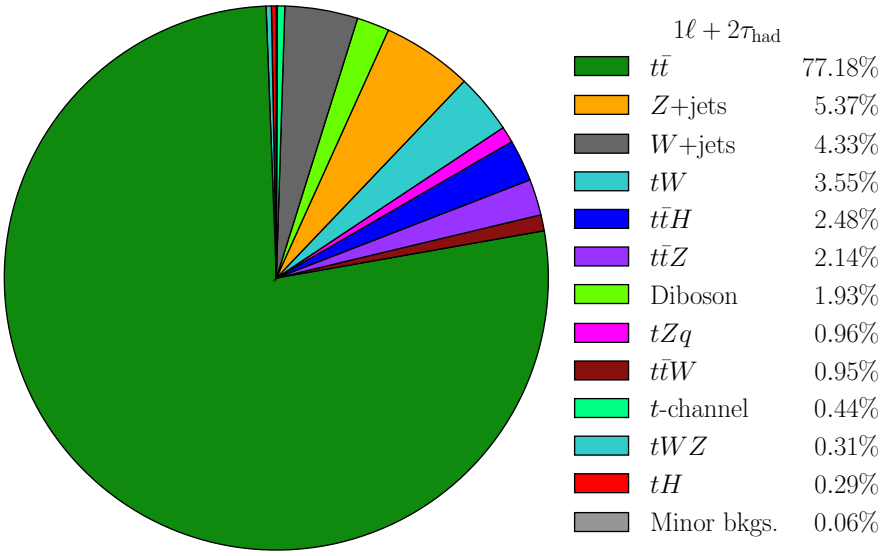



Figure 7.1: Expected tight preselection region composition in the $1\ell + 2\tau_{\text{had}}$ channel. All processes are ordered by relative contribution.

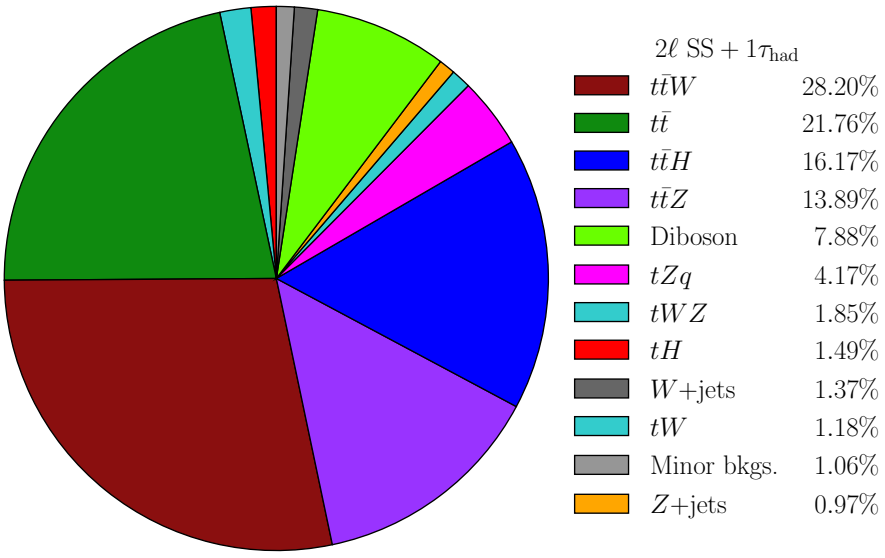



Figure 7.2: Expected tight preselection region composition in the $2\ell$ SS + $1\tau_{\text{had}}$ channel. All processes are ordered by relative contribution.

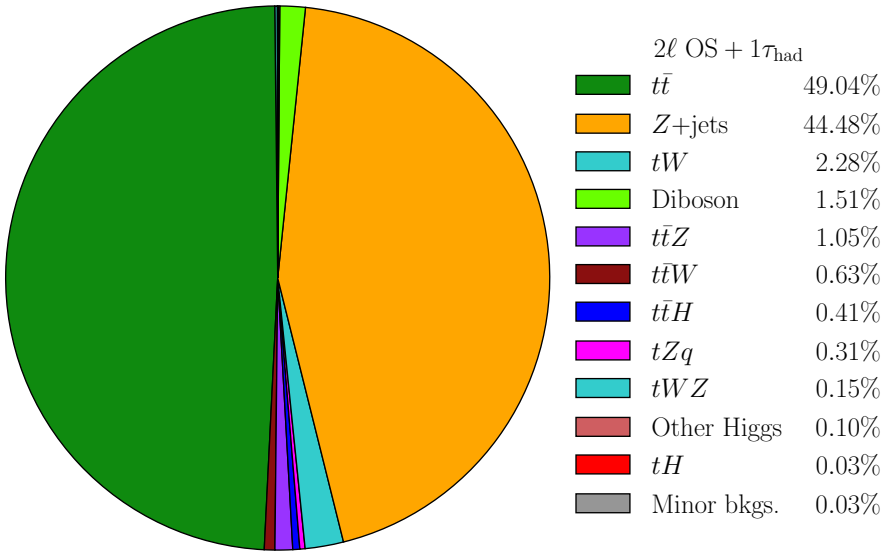



Figure 7.3: Expected tight preselection region composition in the $2\ell$ OS + $1\tau_{\text{had}}$ channel. All processes are ordered by relative contribution.

Table 7.2: Expected raw and weighted yields in the $1\ell + 2\tau_{\text{had}}$ channel's loose and tight preselection regions.

| $1\ell + 2\tau_{\text{had}}$ | Loose preselection | | Tight preselection | |
|---|---|---|---|---|
| Process | Raw yields | Weighted yields | Raw yields | Weighted yields |
| $tHq$ | 55 760 ± 240 | 3.022 ± 0.025 | 38 500 ± 200 | 2.078 ± 0.026 |
| $tWH$ | 1 683 ± 41 | 1.298 ± 0.011 | 1 186 ± 34 | 0.907 ± 0.011 |
| $tWZ$ | 51 170 ± 230 | 4.53 ± 0.04 | 36 180 ± 190 | 3.20 ± 0.04 |
| $t\bar{t}$ | 16 700 ± 130 | 2 047 ± 17 | 6 477 ± 80 | 791 ± 10 |
| $t\bar{t}W$ | 6 162 ± 78 | 17.42 ± 0.15 | 3 436 ± 59 | 9.76 ± 0.12 |
| $t\bar{t}Z$ | 14 770 ± 120 | 32.12 ± 0.27 | 10 400 ± 100 | 21.97 ± 0.28 |
| $t\bar{t}H$ | 44 710 ± 210 | 36.56 ± 0.31 | 30 880 ± 180 | 25.39 ± 0.32 |
| $tZq$ | 87 930 ± 300 | 14.14 ± 0.12 | 61 120 ± 250 | 9.87 ± 0.12 |
| $tW$ | 925 ± 30 | 106.7 ± 0.9 | 312 ± 18 | 36.4 ± 0.5 |
| $Z$ + jets | 20 410 ± 140 | 159.6 ± 1.3 | 7 030 ± 84 | 55.1 ± 0.7 |
| $t$-channel | 423 ± 21 | 18.03 ± 0.15 | 103 ± 10 | 4.49 ± 0.06 |
| $W$ + jets | 3 873 ± 62 | 172.4 ± 1.4 | 1 099 ± 33 | 44.4 ± 0.6 |
| Diboson | 8 678 ± 93 | 32.73 ± 0.27 | 5 880 ± 77 | 19.78 ± 0.25 |
| Minor bkg. | 1 617 ± 40 | 2.847 ± 0.024 | 1 123 ± 34 | 0.612 ± 0.008 |
| Total | 314 800 ± 561 | 2 648 ± 22 | 203 700 ± 451 | 1 025 ± 13 |
| Data | 2 710 | | 1 006 | |

Table 7.3: Expected raw and weighted yields in the $2\ell$ SS + $1\tau_{\text{had}}$ channel's loose and tight preselection regions.

| $2\ell$ SS + $1\tau_{\text{had}}$ | Loose preselection | | Tight preselection | |
|---|---|---|---|---|
| Process | Raw yields | Weighted yields | Raw yields | Weighted yields |
| $tHq$ | 40 220 ± 200 | 2.125 ± 0.013 | 20 160 ± 140 | 1.081 ± 0.022 |
| $tWH$ | 1 457 ± 38 | 1.157 ± 0.007 | 672 ± 26 | 0.531 ± 0.011 |
| $tWZ$ | 41 740 ± 200 | 3.640 ± 0.023 | 23 110 ± 150 | 1.99 ± 0.04 |
| $t\bar{t}$ | 24 610 ± 160 | 3 042 ± 19 | 204 ± 14 | 23.5 ± 0.5 |
| $t\bar{t}W$ | 20 770 ± 140 | 57.2 ± 0.4 | 10 970 ± 100 | 30.4 ± 0.6 |
| $t\bar{t}Z$ | 18 830 ± 140 | 32.84 ± 0.21 | 9 424 ± 97 | 15.00 ± 0.31 |
| $t\bar{t}H$ | 55 080 ± 240 | 37.20 ± 0.23 | 25 600 ± 160 | 17.5 ± 0.4 |
| $tZq$ | 57 540 ± 240 | 8.73 ± 0.05 | 29 690 ± 170 | 4.50 ± 0.09 |
| $tW$ | 836 ± 29 | 100.0 ± 0.6 | 11 ± 3 | 1.270 ± 0.026 |
| $Z$ + jets | 9 328 ± 97 | 48.69 ± 0.31 | 253 ± 16 | 1.050 ± 0.022 |
| $W$ + jets | 1 658 ± 41 | 44.70 ± 0.28 | 10 ± 3 | 1.483 ± 0.003 |
| Diboson | 6 168 ± 79 | 20.39 ± 0.13 | 3 096 ± 56 | 8.51 ± 0.17 |
| Minor bkg. | 6 963 ± 83 | 24.17 ± 0.15 | 3 496 ± 59 | 1.143 ± 0.023 |
| Total | 285 200 ± 534 | 3 423 ± 22 | 126 700 ± 356 | 107.9 ± 2.2 |
| Data | 3 025 | | 92 | |

Table 7.4: Expected raw and weighted yields in the $2\ell$ OS + $1\tau_{\text{had}}$ channel's loose and tight preselection regions.

| $2\ell$ OS + $1\tau_{\text{had}}$ | Loose preselection | | Tight preselection | |
|---|---|---|---|---|
| Process | Raw yields | Weighted yields | Raw yields | Weighted yields |
| $tHq$ | 61 430 ± 250 | 3.342 ± 0.011 | 29 680 ± 170 | 1.668 ± 0.008 |
| $tWH$ | 3 482 ± 59 | 2.741 ± 0.009 | 1 714 ± 41 | 1.370 ± 0.007 |
| $tWZ$ | 294 620 ± 540 | 26.15 ± 0.08 | 176 890 ± 420 | 15.67 ± 0.07 |
| $t\bar{t}$ | 136 960 ± 370 | 16 800 ± 50 | 4 350 ± 66 | 5 291 ± 25 |
| $t\bar{t}W$ | 46 188 ± 220 | 127.0 ± 0.4 | 24 550 ± 160 | 67.63 ± 0.32 |
| $t\bar{t}Z$ | 134 570 ± 370 | 205.2 ± 0.7 | 77 980 ± 280 | 113.0 ± 0.5 |
| $t\bar{t}H$ | 176 970 ± 420 | 88.41 ± 0.29 | 92 980 ± 300 | 44.64 ± 0.21 |
| $tZq$ | 365 490 ± 600 | 54.86 ± 0.18 | 222 070 ± 470 | 33.67 ± 0.16 |
| $tW$ | 6 493 ± 81 | 748.1 ± 2.4 | 2 118 ± 46 | 246.1 ± 1.2 |
| $Z$ + jets | 1 291 400 ± 1 100 | 13 930 ± 50 | 495 300 ± 700 | 4 799 ± 23 |
| Diboson | 84 360 ± 290 | 357.5 ± 1.2 | 47 650 ± 220 | 162.8 ± 0.8 |
| Other Higgs | 45 ± 7 | 18.48 ± 0.06 | 24 ± 5 | 10.37 ± 0.05 |
| Minor bkg. | 19 300 ± 140 | 121.1 ± 0.4 | 8 421 ± 92 | 2.724 ± 0.013 |
| Total | 2 621 300 ± 1 600 | 32 490 ± 110 | 1 222 400 ± 1 100 | 10 790 ± 52 |
| Data | 30 459 | | 10 060 | |

Table 7.5: Expected number of signal events $S$, signal-to-background ratio $S/B$ and signal significance $S/\sqrt{B}$ for all three channels in the loose and tight preselection regions.

| Channel | $1\ell + 2\tau_{\text{had}}$ | | $2\ell$ SS $+ 1\tau_{\text{had}}$ | | $2\ell$ OS $+ 1\tau_{\text{had}}$ | |
|---|---|---|---|---|---|---|
| Preselection | Loose | Tight | Loose | Tight | Loose | Tight |
| $S$ | 4.3 | 3.0 | 3.3 | 1.6 | 6.1 | 3.0 |
| $S/B$ | $1.6 \times 10^{-3}$ | $2.9 \times 10^{-3}$ | $1.0 \times 10^{-3}$ | $1.5 \times 10^{-2}$ | $0.2 \times 10^{-3}$ | $0.3 \times 10^{-3}$ |
| $S/\sqrt{B}$ | $8.4 \times 10^{-2}$ | $9.3 \times 10^{-2}$ | $5.6 \times 10^{-2}$ | $1.6 \times 10^{-1}$ | $3.4 \times 10^{-2}$ | $2.9 \times 10^{-2}$ |

## 7.2 Object reconstruction

As shown in the previous section, $tH$ events only represent a minor contribution to all preselection regions. Given the low cross-section predicted by the Standard Model, this is to be expected. However, the high share of backgrounds ultimately impedes a significant measurement of the $tH$ signal strength. The subsequent Chapter 8 addresses this issue by discussing the optimisation of CNNs to further enhance separation between signal and background processes.

To enable and support this separation, usage of input variables that properly characterise a given process is essential. First of all, this encompasses all final states in the detector. In addition, information about short-lived objects can offer valuable insights when trying to tell apart signal and background events. However, such objects decay before reaching the first layers of the ATLAS detector. This results in their reconstruction being highly non-trivial and only indirectly possible via the measurement and correct assignment of their decay products. In case of $tH$ events, the Higgs-boson and the top-quark represent two such important short-lived objects. Sophisticated methods that allow and improve their reconstruction are subject of the following sections.

### 7.2.1 Reconstructing the top-quark

The most straightforward approach to reconstruct the top-quark is based on simply adding the Lorentz vectors[4] of all its visible decay products. Presuming that the top-quark decays into a bottom-quark and the radiated $W$-boson decays leptonically, one ends up with just two visible decay products in the detector. These correspond to a $b$-jet and the light lepton associated with the $W$-boson decay $\ell_{\text{top}}$. Assuming that the leading $b$-jet originates from the top-quark decay, this yields

$$p_{\text{top}} = p_{\text{lead. } b\text{-jet}} + p_{\ell_{\text{top}}}, \tag{7.1}$$

where $p$ denotes a particle's Lorentz vector.

In this simple reconstruction, the neutrino $\nu_{\text{top}}$ arising from the leptonic $W$-boson decay is completely ignored. This obviously affects the reconstruction's precision. To address this shortcoming, a second reconstruction method, based on MC truth studies presented in [146], is carried out. This method allows an approximate reconstruction of the neutrino's kinematics by exploiting correlations with other decay products. The complete procedure relies on the assumption, that the total missing transverse

[4] Lorentz vectors are four-vectors that play an important role when studying particle collisions and decays. A brief explanation is provided in Appendix B.

momentum in $tH$ events is exclusively caused by neutrinos, present in the Higgs-boson and top-quark decays.

On this basis, correlations between all involved particles are investigated on MC truth level. These studies indicate a correlation between the kinematics of the light lepton and the neutrino, as visualised in Figure 7.4. Utilising these correlations

$$p_{\mathrm{T},\,\nu_{\mathrm{top}}} = \frac{1\,615.98\,\mathrm{GeV}^2}{p_{\mathrm{T},\,\ell_{\mathrm{top}}}}\,, \tag{7.2}$$

and

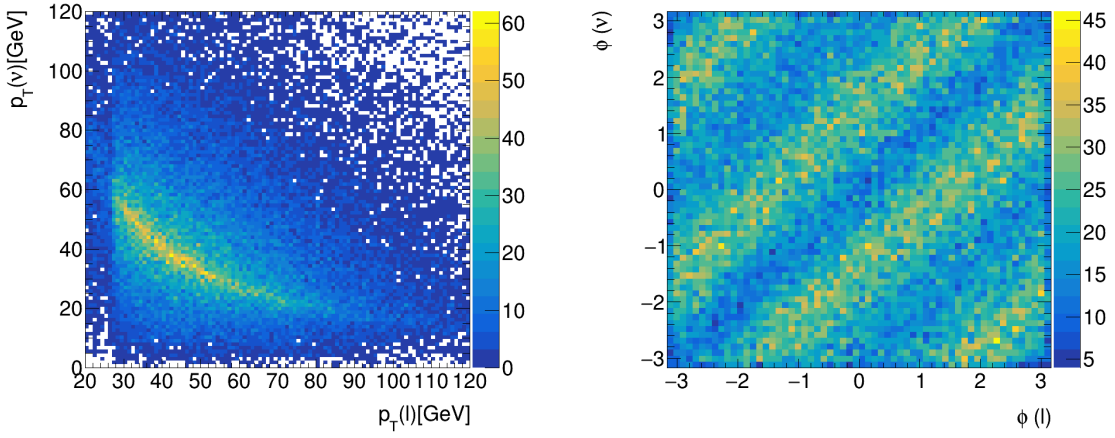
$$\phi_{\nu_{\mathrm{top}}} = \phi_{\ell_{\mathrm{top}}} \pm \frac{\pi}{2}\,, \tag{7.3}$$

can be derived. The ambiguity in Equation (7.3) can be resolved by additionally considering correlations between the neutrino and the leading $b$-jet via

$$\phi_{\nu_{\mathrm{top}}} = \begin{cases} \phi_{\ell_{\mathrm{top}}} + \frac{\pi}{2}\,, & \phi_{\text{lead. } b\text{-jet}} - \phi_{\ell_{\mathrm{top}}} < 0\,, \\ \phi_{\ell_{\mathrm{top}}} - \frac{\pi}{2}\,, & \phi_{\text{lead. } b\text{-jet}} - \phi_{\ell_{\mathrm{top}}} \geq 0\,. \end{cases} \tag{7.4}$$

These approximations allow the $W$-boson to be reconstructed by adding the neutrino and the light lepton. Lastly, the top-quark is obtained by combining the $W$-boson along with the leading $b$-jet.

As a comparison, the resulting top-quark mass for $1\ell + 2\tau_{\mathrm{had}}$ events is shown in Figure 7.5 for both reconstruction methods. The more sophisticated second method clearly provides results closer to the expected top-quark mass of $m_t = 172.57\,\mathrm{GeV}$. At the same time, modelling of the distribution's tails along with the width is improved.




(a) (b)

Figure 7.4: Correlations between (a) $p_{\mathrm{T},\,\nu_{\mathrm{top}}}$ and $p_{\mathrm{T},\,\ell_{\mathrm{top}}}$ as well as (b) $\phi_{\nu_{\mathrm{top}}}$ and $\phi_{\ell_{\mathrm{top}}}$ observed using MC truth information in the $2\ell$ OS + $1\tau_{\mathrm{had}}$ channel. In [146] fits were performed on both distributions which result in Equation (7.2) and Equation (7.3) [143].

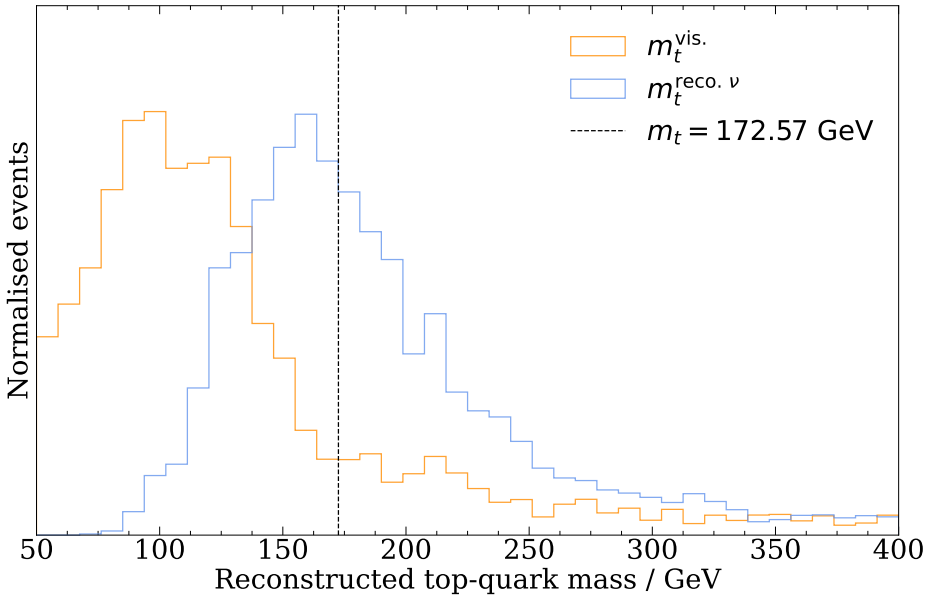



Figure 7.5: Results of both top-quark reconstruction methods for $tH$ events in the $1\ell + 2\tau_{\text{had}}$ channel. The mass reconstructed using the visible decay products is denoted by $m_t^{\text{vis.}}$ while $m_t^{\text{reco. }\nu}$ represents the more sophisticated reconstruction based on truth studies. The correct top-quark mass $m_t$ is indicated.

### 7.2.2 Reconstructing the Higgs-boson

The second short-lived particle to be reconstructed is the Higgs-boson. Depending on the channel a decay into two $\tau_{\text{had}}$ or one $\tau_{\text{had}}$ and one $\tau_{\text{lep}}$ has to be considered. In both cases at least two neutrinos are present, making the reconstruction non-trivial. A first approach can be carried out analogously to the top-quark reconstruction. Again, adding the visible decay products while completely ignoring neutrinos yields

$$p_{\text{Higgs, }1\ell+2\tau_{\text{had}}} = p_{\text{lead. }\tau_{\text{had}}} + p_{\text{sublead. }\tau_{\text{had}}}\,, \tag{7.5}$$

$$p_{\text{Higgs, }2\ell+1\tau_{\text{had}}} = p_{\tau_{\text{had}}} + p_{\ell_{\text{Higgs}}}\,. \tag{7.6}$$

The result only provides a rough estimate, which is why a second improved method based on the missing mass calculator (MMC) [147] is utilised. The MMC was developed to reconstruct the invariant mass of resonances decaying into two tau-leptons, such as the $Z$- and Higgs-boson.

It relies on the assumption that the resonance mass is significantly greater than that of its decay products. As a result, each tau-lepton and its own decay products, are collimated. The MMC takes this condition into account but compared to other reconstruction methods also allows for similar albeit not identical flight directions of all decay products [148]. These prerequisites enable a system of underdetermined equations to be set up. Even if no exact analytical solution can be found, additional kinematic information, such as the distance $\Delta R$ between all decay products, help to distinguish between likely and unlikely solutions [147]. Internally, the most probable solution, yielding a mass estimate, is found by likelihood maximisation.

In addition to the Higgs-boson's visible decay products, i.e. leading and subleading $\tau_{\text{had}}$ or $\tau_{\text{had}}$ and

$\ell_{\text{Higgs}}$, the MMC also requires $E_{\text{T}}^{\text{miss}}$ as an input. Since the method expects that $E_{\text{T}}^{\text{miss}}$ is entirely allocated to the neutrinos produced during the Higgs-boson's decay, the previously derived contribution from $\nu_{\text{top}}$ must be subtracted. The resulting Higgs-boson mass is shown in Figure 7.6 for both reconstruction methods. As expected, a determination via the MMC provides a more accurate estimate.

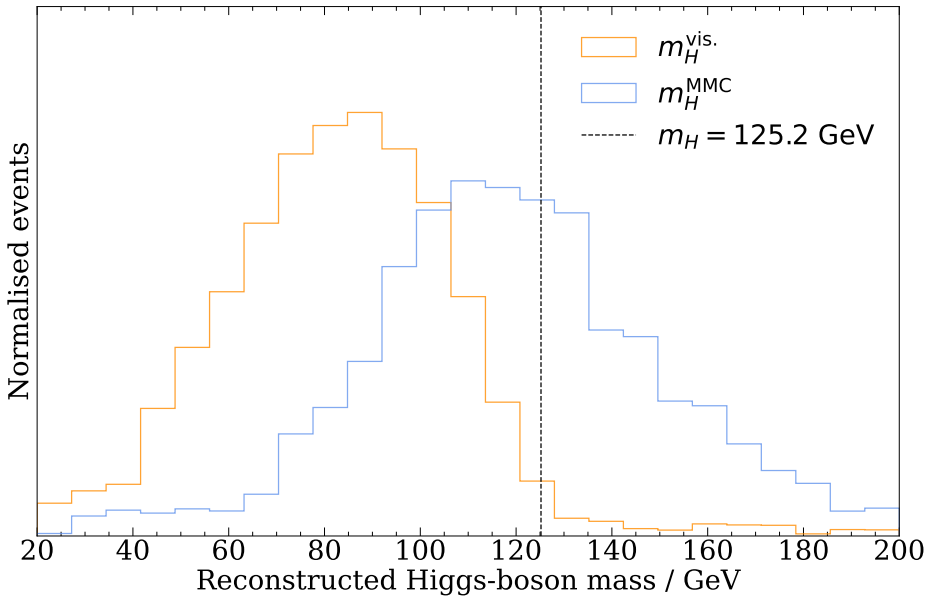



Figure 7.6: Results of both Higgs-boson reconstruction methods for $tH$ events in the $1\ell + 2\tau_{\text{had}}$ channel. The mass reconstructed using the visible decay products is denoted by $m_H^{\text{vis.}}$ while $m_H^{\text{MMC}}$ represents the more sophisticated reconstruction based on the MMC. The correct Higgs-boson mass $m_H$ is indicated.

CHAPTER 8

# Multivariate signal and background separation

The *tH* cross-section, as predicted by the Standard Model, is orders of magnitude lower than that of most individual background processes. As a consequence, the tight preselection, presented in Chapter 7, forms an essential foundation to perform a significant measurement. However, since the resulting signal-to-background ratio of around 2 % still remains quite low, this selection is not considered to be sufficient. Therefore, this chapter presents an approach to improve separation between signal and background events by relying on CNNs, introduced in Chapter 5.

The final cross-section estimation, discussed in Chapter 10, is based on a binned profile likelihood fit. Thus, the objective of an improved separation is not necessarily to define a completely new and improved analysis region. Rather, the networks' response should correspond to a variable which provides an optimal discrimination between signal and background events in the tight preselection region. An adequately binned fit, performed on this variable, can then exploit resulting bins with a greatly increased signal significance. This enables a more sensitive evaluation, less influenced by background processes.

To begin with, Section 8.1 outlines the general strategy as well as the chosen CNN design. Subsequently, Section 8.2 discusses the optimisation of each individual network. This concerns not only the choice of input variables but also the optimisation of all crucial network hyperparameters. Finally, Section 8.3 concludes with an evaluation of the final networks' performance.

The described method builds on and modifies the studies documented in [149]. Throughout the chapter, reference is made to this *previous optimisation* in order to emphasise differences and similarities.

## 8.1 Neural network design and training strategy

Generally, the selection described in Chapter 7 results in three distinct channels. Given the different background compositions, an individual optimisation of the signal isolation in each channel is justified. However, as the $2\ell$ OS + $1\tau_{\text{had}}$ channel's expected signal significance is substantially lower compared to the other two channels, an independent cross-section estimation is not performed in this region. As a consequence, the analysis is not expected to benefit from an enhanced signal and background separation in this channel. Thus, in contrast to [149], only two NNs need to be trained, one in the $1\ell + 2\tau_{\text{had}}$ channel and one in the $2\ell$ SS + $1\tau_{\text{had}}$ channel. Their general design and setup are explained

in the following sections.

### 8.1.1 Choice of categories

Both remaining channels, $1\ell + 2\tau_{\text{had}}$ and $2\ell\,\text{SS} + 1\tau_{\text{had}}$, do not exclusively consist of just a single background or a set of similar backgrounds. To reflect this complex composition CNNs are employed instead of binary NNs. By using a CNN's response, every individual event is assigned a probability of belonging to a certain category.

The exact definition of categories, i.e. the number of output nodes, is motivated as follows: the obvious choice is that of a signal, i.e. $tH$ category in both channels. The CNN's response regarding this category, termed $\text{NN}_{tH}$, thus represents the networks confidence for an event to be a signal event.

What remains is the non-trivial split of background processes into different categories. Firstly, it is not to be expected that each individual background process can be easily distinguished by any trained network. Otherwise, suitable selections would have already been introduced in Chapter 7. In addition, it has to be borne in mind that the entire optimisation relies on a labelled dataset. This is realised by carrying out the training on the basis of MC simulations. Consequently, a sufficiently large simulation sample size is required for each process to obtain a generalised classification and minimise the risk of overtraining. Validating the raw event yields listed in Table 7.2 and Table 7.3 indicates that this is not a given for every process. For instance, only eleven unweighted $tW$ events are available after applying the tight preselection in the $2\ell\,\text{SS} + 1\tau_{\text{had}}$ channel. Therefore, an overly fine splitting is not justified and might even lead to a decrease in performance. This is also supported by previous studies performed in the $2\ell\,\text{OS} + 1\tau_{\text{had}}$ channel, as summarised in [150].

Taking these restrictions into account, various sets of background categories (see Figures 7.1 to 7.3) are tested. The best results are achieved with the categorisation given below:

- $1\ell + 2\tau_{\text{had}}$ channel: besides the signal category, the $t\bar{t}$ process, being the dominant background, is assigned its own category. All remaining backgrounds are grouped into a third category. The corresponding CNN responses are termed $\text{NN}_{tH}$, $\text{NN}_{t\bar{t}}$ and $\text{NN}_{\text{background}}$.
- $2\ell\,\text{SS} + 1\tau_{\text{had}}$ channel: in contrast to the $1\ell + 2\tau_{\text{had}}$ channel, the $2\ell\,\text{SS} + 1\tau_{\text{had}}$ events are not dominated by a single background. Nonetheless, the best results are obtained by again using three categories. The first two are a signal category ($\text{NN}_{tH}$) and a combination of $t\bar{t}$, $t\bar{t}H$, $t\bar{t}Z$ and $t\bar{t}W$ events ($\text{NN}_{t\bar{t}\,+\,t\bar{t}X}$). The third category and the corresponding CNN response $\text{NN}_{\text{background}}$, are formed by all remaining processes in analogy to the $1\ell + 2\tau_{\text{had}}$ channel.

To summarise, two CNNs with three categories each are used for the following analysis. In the remainder of this section, the division of all datasets utilised for their training and evaluation is discussed.

### 8.1.2 Splitting the input data

The division of datasets into training and evaluation sets prior to optimising the CNNs represents a major difference compared to the previous optimisation. In order to understand the chosen split, it is necessary to first consider the conventional use case of NNs, employed to solve classification problems: both training and evaluation of a model are performed on an existing labelled dataset. Once the optimisation is carried out, a NN is supposed to provide predictions for new, previously unseen data. Consequently, the initially used labelled dataset is no longer relevant.

Conversely, the discussed particle physics use case represents a different scenario: MC simulations, i.e. the labelled data, used during training, also play an important role in the signal strength estimation. To avoid introducing bias, simulations that have already been used to optimise a network should not be predicted on by the same network. If, under this condition, the standard three-way split (see Section 5.2.2) into a training, validation and test set is chosen, only the test dataset can be predicted on as it is held back during the optimisation. Therefore, only this dataset's events will be available in a subsequent parameter estimation. Since the number of simulated MC events is limited, this approach poses a problem by leading to an increased statistical uncertainty.

To overcome this issue, all CNNs trained in this thesis rely on the $k$-fold cross-validation method (introduced in Section 5.2.2), as opposed to the previous optimisation. The available dataset is split into five folds, i.e. five independent models are formed and optimised per channel. Each individual model receives three folds' events in training. The remaining two folds along with their associated events are available for validation and testing/prediction, respectively. When assigning a set of folds to each model different permutations are exploited. This way it is ensured that no model predicts on test samples used in its own training. The division is schematically depicted in Figure 8.1 and allows the complete dataset to be used in training and validation while still obtaining an unbiased prediction through the test fold.

Individual events are allocated to each fold by performing Euclidean division based on their event number. This number is generated directly during the simulation and is unique for each event of a certain process. Rerunning the simulation with different theory parameters (see Section 6.2) will generate events with different kinematics and yet identical event numbers. Accordingly, each fold fulfils the following conditions:

- Each fold is formed by approximately the same number of events.
- Any fold's relative composition regarding different processes mirrors the tight preselection's composition, i.e. the entirety of all folds.
- Any nominal MC event and its counterpart found in an alternative sample, generated to estimate systematic uncertainties, end up in the same fold. This ensures that both are predicted by the identical model.

Ultimately, relying on this method guarantees that statistical uncertainties are kept low without introducing any bias in the prediction.

## 8.2 Neural network optimisation

Having determined the number of categories, this section outlines the general optimisation of both channels' CNNs. Initially, the selection of an ideal input variable set is discussed. Based on these inputs, tuning and evaluation of different hyperparameter settings is carried out.

### 8.2.1 Input variables

In theory, a NN is able to represent any linear or non-linear function that maps a selection of inputs to a target. However, such a mapping can only be successful if the input variables contain enough information to correctly map to the target. As an example, an attempt to predict the development or

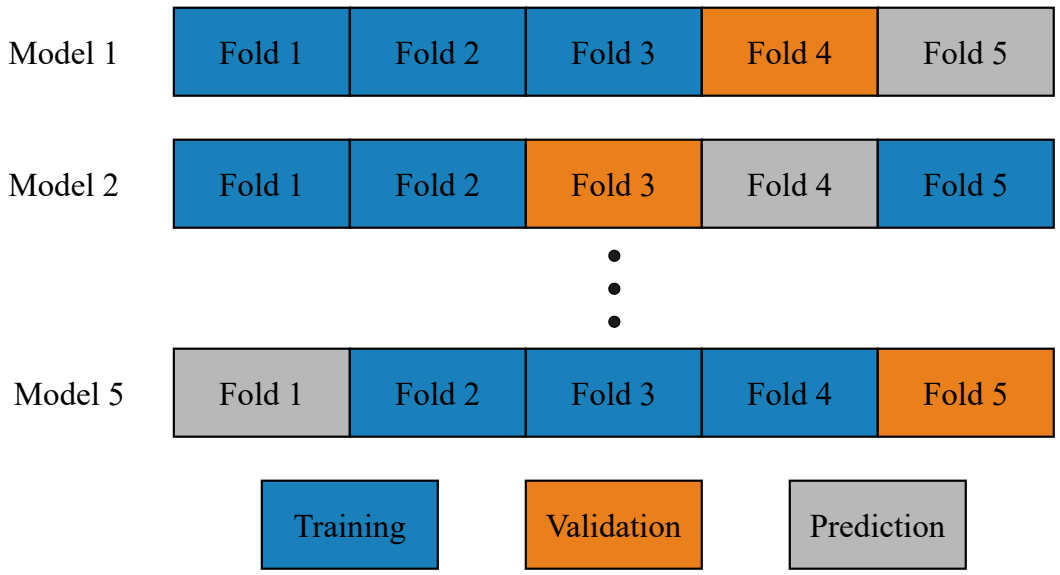



Figure 8.1: Schematic illustration showing the mapping of all folds to the five independent CNNs models. Each model uses different sets of folds in training, validation and testing. Here, the test folds are referred to as prediction folds. Only their predictions are unbiased and thus available in a parameter estimation.

value of any given share price on the basis of its history is doomed to fail. The history simply does not contain sufficient information to forecast future developments.

Therefore, the concrete choice of input variables is crucial in order to obtain a well performing network. The naive approach in the actual use case corresponds to simply utilising all reconstructed variables to provide as much information as possible. However, just like a lack of information, this abundance of inputs might also limit a model's predictive capability. This phenomenon, often termed the "curse of dimensionality", can be explained by considering the increasingly sparse distribution of data points in a high-dimensional space. This situation hinders the identification of relevant correlations and meaningful patterns. Thus, optimising a NN on too many inputs, partially with redundant information, often results in longer training times and poor generalisation.

Respecting these conditions, the input variable set should be as small as possible, while still accurately describing the signature of signal and background processes. The corresponding selection procedure was carried out in analogy to the previous optimisation. First, kinematic information of the final states is used. This includes the $\tau_{\text{had}}$, the light leptons' signature, as well as the leading jet and the leading $b$-jet. In addition, the quantification of missing transverse energy $E_{\text{T}}^{\text{miss}}$ is added to the list of inputs. Furthermore, knowledge on the Higgs-boson and the top-quark, such as the MMC's mass estimate derived in Section 7.2, is fed to the CNNs.

The resulting set of variables accompanied by their concrete definition is summarised in Table 8.1 for the $1\ell + 2\tau_{\text{had}}$ channel and in Table 8.2 for the $2\ell$ SS $+ 1\tau_{\text{had}}$ channel. Their respective distributions are displayed in Appendix E. As these variable sets match those of the previous optimisation, a direct comparison of both approaches is facilitated.

An additional supporting study, testing a reduced number of inputs, was conducted in [149]. For this purpose, a set of Lorentz invariant variables was used. In theory, these variables are sufficient to fully describe the entire process. To test the contrary case, results with larger input variable sets are also investigated as part of this work. Particular focus is placed on tests with an increased amount of jet kinematics to characterise each event's final state more precisely. These tests, as well as the studies in [149] are unable to show improvements with respect to the input sets given in Tables 8.1 and 8.2, and consequently not pursued further.

Table 8.1: Input variables used during training of a CNN in the $1\ell + 2\tau_{\text{had}}$ channel.

| Variable name | Description |
|---|---|
| $\eta_{\text{forward jet}}$ | Pseudorapidity of the forward jet |
| $p_{\text{T, forward jet}}$ | Transverse momentum of the forward jet |
| $\phi_{\text{forward jet}}$ | Azimuthal angle of the forward jet |
| $\eta_{\text{lead. } b\text{-jet}}$ | Pseudorapidity of the leading $b$-jet |
| $p_{\text{T, lead. } b\text{-jet}}$ | Transverse momentum of the leading $b$-jet |
| $\phi_{\text{lead. } b\text{-jet}}$ | Azimuthal angle of the leading $b$-jet |
| $\eta_{\text{lead. jet}}$ | Pseudorapidity of the leading jet |
| $p_{\text{T, lead. jet}}$ | Transverse momentum of the leading jet |
| $\phi_{\text{lead. jet}}$ | Azimuthal angle of the leading jet |
| $E_{\text{lead. } \tau_{\text{had}}}$ | Energy of the leading $\tau_{\text{had}}$ |
| $\eta_{\text{lead. } \tau_{\text{had}}}$ | Pseudorapidity of the leading $\tau_{\text{had}}$ |
| $p_{\text{T, lead. } \tau_{\text{had}}}$ | Transverse momentum of the leading $\tau_{\text{had}}$ |
| $\phi_{\text{lead. } \tau_{\text{had}}}$ | Azimuthal angle of the leading $\tau_{\text{had}}$ |
| $E_{\text{sublead. } \tau_{\text{had}}}$ | Energy of the subleading $\tau_{\text{had}}$ |
| $\eta_{\text{sublead. } \tau_{\text{had}}}$ | Pseudorapidity of the subleading $\tau_{\text{had}}$ |
| $p_{\text{T, sublead. } \tau_{\text{had}}}$ | Transverse momentum of the subleading $\tau_{\text{had}}$ |
| $\phi_{\text{sublead. } \tau_{\text{had}}}$ | Azimuthal angle of the subleading $\tau_{\text{had}}$ |
| $E_{\ell_{\text{top}}}$ | Energy of the light lepton associated to the top-quark $\ell_{\text{top}}$ |
| $\eta_{\ell_{\text{top}}}$ | Pseudorapidity of the light lepton associated to the top-quark $\ell_{\text{top}}$ |
| $p_{\text{T, } \ell_{\text{top}}}$ | Transverse momentum of the light lepton associated to the top-quark $\ell_{\text{top}}$ |
| $\phi_{\ell_{\text{top}}}$ | Azimuthal angle of the light lepton associated to the top-quark $\ell_{\text{top}}$ |
| $E_{\text{T}}^{\text{miss}}$ | Missing transverse energy |
| $\text{MMC}_{\text{Out}}$ | MMC Higgs mass estimation |
| $\eta_{\text{visible. Higgs}}$ | Pseudorapidity of the Lorentz vector sum of both $\tau_{\text{had}}$ |
| $p_{\text{T, visible Higgs}}$ | Transverse momentum of the Lorentz vector sum of both $\tau_{\text{had}}$ |
| $\eta_{\text{visible. top}}$ | Pseudorapidity of the visible top-quark decay products' Lorentz vector sum |
| $p_{\text{T, visible top}}$ | Transverse momentum of the visible top-quark decay products' Lorentz vector sum |
| $M_{\text{lead. } b\text{-jet + forward jet}}$ | Mass of the Lorentz vector sum of the leading $b$-jet and the forward jet |
| $H_{\text{T}}$ | Total transverse energy |
| $\Delta R_{\text{lead. } \tau_{\text{had}}, \text{ sublead. } \tau_{\text{had}}}$ | $\Delta R$ of both $\tau_{\text{had}}$ |
| $\Delta\phi_{\text{lead. } \tau_{\text{had}}, \text{ sublead. } \tau_{\text{had}}}$ | $\Delta\phi$ of both $\tau_{\text{had}}$ |

Table 8.2: Input variables used during training of a CNN in the $2\ell$ SS + $1\tau_{\text{had}}$ channel.

| Variable name | Description |
|---|---|
| $\eta_{\text{forward jet}}$ | Pseudorapidity of the forward jet |
| $p_{\text{T, forward jet}}$ | Transverse momentum of the forward jet |
| $\phi_{\text{forward jet}}$ | Azimuthal angle of the forward jet |
| $\eta_{\text{lead. } b\text{-jet}}$ | Pseudorapidity of the leading $b$-jet |
| $p_{\text{T, lead. } b\text{-jet}}$ | Transverse momentum of the leading $b$-jet |
| $\phi_{\text{lead. } b\text{-jet}}$ | Azimuthal angle of the leading $b$-jet |
| $\eta_{\text{lead. jet}}$ | Pseudorapidity of the leading jet |
| $p_{\text{T, lead. jet}}$ | Transverse momentum of the leading jet |
| $\phi_{\text{lead. jet}}$ | Azimuthal angle of the leading jet |
| $\eta_{\tau_{\text{had}}}$ | Pseudorapidity of the leading $\tau_{\text{had}}$ |
| $p_{\text{T, }\tau_{\text{had}}}$ | Transverse momentum of the leading $\tau_{\text{had}}$ |
| $\phi_{\tau_{\text{had}}}$ | Azimuthal angle of the leading $\tau_{\text{had}}$ |
| $\eta_{\ell_{\text{Higgs}}}$ | Pseudorapidity of the lepton associated to the Higgs-boson $\ell_{\text{Higgs}}$ |
| $p_{\text{T, }\ell_{\text{Higgs}}}$ | Transverse momentum of the lepton associated to the Higgs-boson $\ell_{\text{Higgs}}$ |
| $\phi_{\ell_{\text{Higgs}}}$ | Azimuthal angle of the lepton associated to the Higgs-boson $\ell_{\text{Higgs}}$ |
| $\eta_{\ell_{\text{top}}}$ | Pseudorapidity of the lepton associated to the top-quark $\ell_{\text{top}}$ |
| $p_{\text{T, }\ell_{\text{top}}}$ | Transverse momentum of the lepton associated to the top-quark $\ell_{\text{top}}$ |
| $\phi_{\ell_{\text{top}}}$ | Azimuthal angle of the lepton associated to the top-quark $\ell_{\text{top}}$ |
| $E_{\text{T}}^{\text{miss}}$ | Missing transverse energy |
| $\text{MMC}_{\text{Out}}$ | MMC Higgs mass estimation |
| $\eta_{\text{visible. Higgs}}$ | Pseudorapidity of the Lorentz vector sum of $\ell_{\text{Higgs}}$ and $\tau_{\text{had}}$ |
| $p_{\text{T, visible Higgs}}$ | Transverse momentum of the Lorentz vector sum of $\ell_{\text{Higgs}}$ and $\tau_{\text{had}}$ |
| $\eta_{\text{visible. top}}$ | Pseudorapidity of the visible top-quark decay products' Lorentz vector sum |
| $p_{\text{T, visible top}}$ | Transverse momentum of the visible top-quark decay products' Lorentz vector sum |
| $M_{\text{visible top}}$ | Mass of the visible top-quark decay products' Lorentz vector sum |
| $M_{\text{lead. } b\text{-jet + forward jet}}$ | Mass of the Lorentz vector sum of the leading $b$-jet and the forward jet |
| $H_{\text{T}}$ | Total transverse energy |
| $\Delta\phi_{\tau_{\text{had}},\ \ell_{\text{Higgs}}}$ | $\Delta R$ of $\ell_{\text{Higgs}}$ and $\tau_{\text{had}}$ |
| $\Delta R_{\tau_{\text{had}},\ \ell_{\text{Higgs}}}$ | $\Delta\phi$ of $\ell_{\text{Higgs}}$ and $\tau_{\text{had}}$ |

### 8.2.2 Hyperparameter optimisation

The previous results were obtained by CNNs with an optimised set of hyperparameters. Since the models utilised in this thesis rely on identical input variables and share the same output categories, the hyperparameters of the previous optimisation are expected to already form a solid starting point. Their original optimisation was conducted via an iterative "evolutionary approach" [149]: first, CNNs were initialised with hyperparameters randomly selected from a predefined range of values. Once trained, all models were ranked according to the AUC of their signal response's ROC curve. This ranking was then used to define the next set of hyperparameters. Besides the best model which always enters the next iteration, all parameter sets ranking above average were shuffled to form new test models. This complete evolution step was repeated five times in total. Afterwards, all hyperparameters except for the batch size were fixed to their best value based on the highest AUC. Subsequently, the batch size was optimised via a one-dimensional grid search [149]. All final values for both channels are displayed in Table 8.3 and Table 8.4.

Compared to the previous optimisation, the presented analysis is carried out with two noteworthy changes. The first difference concerns the previously mentioned use of the $k$-folding approach. In addition, the $Z+\text{jets}$ and $W+\text{jets}$ backgrounds are modelled by an updated MC simulation which offers an increased number of unweighted events.

Both modifications are not expected to fundamentally change the classification problem, as event signatures and kinematics remain unchanged. Thus, the previous hyperparameters are likely close to being optimal and rerunning the time-consuming optimisation from the beginning is not considered reasonable. Instead, a grid search, focused on hyperparameters near the previous results, is carried out. The tested range of each individual hyperparameter is summarised in Table 8.5 for both channels. The differences in batch size between the two channels can be explained by the relatively low number of raw events in the $2\ell\,\text{SS} + 1\tau_{\text{had}}$ channel. In particular, highly weighted processes with low raw event counts such as $t\bar{t}$ (see Table 7.3) can lead to unstable models when choosing a small batch size. Evaluated models are set up with the ELU activation function for each hidden layer[1], whereas the softmax function is applied to the output layer. The latter ensures a correct normalisation of all three output categories per channel.

In order to stabilise a model's training, the inputs of individual layers are normalised per batch. This adjustment, simply termed batch normalisation [151], prevents outliers from causing overtaining and allows higher learning rates to be used. This speeds up the training, i.e. the number of epochs can be reduced. The actual number of epochs itself is not directly determined as part of the grid search. Instead, the training and validation loss curves are monitored for each fold. The training is simply continued until a discrepancy between both loss curves starts to develop. At this point, the model's training is terminated to avoid overtraining via early stopping, as discussed in Section 5.2. Robustness against fluctuations is ensured by requiring a patience value of ten epochs. Fluctuations in the loss curves usually only persist over one or two epochs, making a patience of ten epochs a conservative estimate.

Ultimately, all individual hyperparameter sets are ranked according to the best signal response by studying the ROC curve's AUC, similar to the previous optimisation. Due to the presence of five folds, i.e. five models, the AUC was averaged across all models. Hyperparameters that cause large performance discrepancies between individual models are discarded.

[1] In principle, the activation function of a hidden layer is also a hyperparameter. However, initial results showed that the use of ELU led to the best results, which is why this hyperparameter is not varied.

All AUC values are computed using the weighted event yields per process, as listed in Table 7.2 and Table 7.3. In contrast, a reweighting of the form

$$w_i' = \frac{w_i}{\sum_{i=1}^{N} w_i}, \tag{8.1}$$

is applied to all $N$ event weights $w_i$ of each output category before training individual models. This ensures that the total event weight in each output category equals one ($\sum_{i=1}^{N} w_i' = 1$), providing equal importance to all three categories. Otherwise, the loss function's minimisation would be completely governed by correctly classifying the dominant backgrounds. Properly predicting a signal event, on the other hand, would have a barely noticeable influence on the minimisation. This would obviously hinder the intended discrimination between signal and background events.

As a summary, the best hyperparameters extracted from this grid search are listed in Table 8.3 and Table 8.4 for the respective channels. A direct comparison with the previous optimisation reveals that the initial assumption is correct: their hyperparameters provide a stable model and only require minor adjustments to obtain optimal results.

Table 8.3: Final hyperparameters used in the $1\ell + 2\tau_{\text{had}}$ channel's CNN compared to the previous optimisation.

| Hyperparameter | Previous optimisation | Final optimisation |
|---|---|---|
| Number of layers | 6 | 5 |
| Nodes per layer | 120 | 100 |
| Batch size | 25 000 | 25 000 |
| Batch normalisation | True | True |
| Optimiser | Adam | Adam |
| Learning rate | 0.001 | 0.001 |
| Dropout fraction | 0.2 | 0.2 |

Table 8.4: Final hyperparameters used in the $2\ell\,\text{SS} + 1\tau_{\text{had}}$ channel's CNN compared to the previous optimisation.

| Hyperparameter | Previous optimisation | Final optimisation |
|---|---|---|
| Number of layers | 7 | 6 |
| Nodes per layer | 70 | 70 |
| Batch size | 100 000 | 80 000 |
| Batch normalisation | True | True |
| Optimiser | Adam | Adam |
| Learning rate | 0.001 | 0.01 |
| Dropout fraction | 0.2 | 0.2 |

## 8.3 Neural network evaluation

Both previous sections focused on the CNNs' general structure and their hyperparameter optimisation. Consequently, this section concludes the discussion by evaluating the final model's performance at

Table 8.5: Scanned hyperparameters in both channels' grid search. If an interval is given, a parameter is varied according to the specified step size.

| Hyperparameter | $1\ell + 2\tau_{\text{had}}$ | $2\ell\,\text{SS} + 1\tau_{\text{had}}$ |
|---|---|---|
| Number of layers | [4, 8], step 1 | [5, 9], step 1 |
| Nodes per layer | [80, 160], step 10 | [30, 110], step 10 |
| Batch size | [10 000, 40 000], step 5 000 | [60 000, 140 000], step 20 000 |
| Learning rate | 0.0001, 0.001, 0.01 | 0.0001, 0.001, 0.01 |
| Dropout fraction | [0.0, 0.4], step 0.1 | [0.0, 0.4], step 0.1 |

separating signal and background events. Following the $k$-folding approach, a total of five identical models are trained per channel. They all use the corresponding hyperparameters listed in Table 8.3 and Table 8.4.

The first evaluation relies on studying the loss curves obtained during training and validation, displayed in Figure 8.2 and Figure 8.3. The training of all models in the $2\ell\,\text{SS} + 1\tau_{\text{had}}$ channel is terminated by early stopping after approximately 150 epochs during hyperparameter optimisation. In contrast, the $1\ell + 2\tau_{\text{had}}$ channel training is carried out over around 260 epochs. During the final training, however, all folds of one channel are trained for an identical number of epochs to obtain comparable models. This difference between both channels can be explained by comparing the available number of events (see Table 7.2 and Table 7.3). Roughly 1.6 times more events are available in the $1\ell + 2\tau_{\text{had}}$ channel, which increases its training process' robustness against overtraining and allows for a longer optimisation.

Apart from the mentioned difference in the number of epochs both channel's loss curves show the expected shape. Training and validation loss within one fold lie close to each other and converge to a stable minimum. This is a first indication that worrisome overtraining is avoided in all models. Furthermore, no major discrepancies or outliers are observed when comparing the loss curve across all folds.

Nonetheless, two additional aspects are striking and need discussion: the training loss curve is subject to significantly more fluctuations than the validation loss. Simultaneously, the latter dataset yields lower loss values over the first epochs. The fluctuations are a direct consequence of using Adam as an optimiser. Internally, Adam relies on batches to determine the gradient, whose reduced sample size introduces this behaviour. The validation loss, on the other hand, is determined once per epoch using all events in the respective fold and thus less prone to fluctuations. The second aspect, i.e. the noticeable discrepancy between training and validation across the first epochs, can be attributed to the application of dropout. This regularisation method only affects the training loss, whereas no node connections are dropped during validation. Therefore, unoptimised models initially tend to fluctuate and provide worse training results. In case of the $2\ell\,\text{SS} + 1\tau_{\text{had}}$ channel, only this second aspect is relevant, since the large batch size (see Table 8.4) effectively corresponds to computing the gradient descent based on the entire training dataset.

The second step of the evaluation centres around analysing each channels' signal node response ($\text{NN}_{tH}$), as shown in Figure 8.4 and Figure 8.5, respectively. Both distributions are obtained by combining the prediction datasets across all five folds. The remaining response categories, i.e. $\text{NN}_{\text{background}}$ and $\text{NN}_{t\bar{t}}$ or $\text{NN}_{t\bar{t}\,+\,t\bar{t}X}$, are merged and termed "orthogonal categories". Generally, the pattern required to enable an improved separation can be obtained, as $tH$ events tend to yield

a prediction closer to one. Directly comparing the two channels clearly shows that a more distinct separation is achieved in the $1\ell + 2\tau_{\text{had}}$ channel. This result can be partially explained by the significantly smaller training dataset, available in the $2\ell\,\text{SS} + 1\tau_{\text{had}}$ channel. In addition, the charge requirement already prevents some large backgrounds such as $Z$ + jets from entering the $2\ell\,\text{SS} + 1\tau_{\text{had}}$ channel. As an outcome, the $2\ell\,\text{SS} + 1\tau_{\text{had}}$ channel consists of backgrounds which more similar to $tH$ events and thus more difficult to separate. The same studies are also carried out for the orthogonal outputs and displayed in Appendix F.

Subsequently, a final validation is carried out by studying the ROC curves derived on the basis of the $\text{NN}_{tH}$ response. The corresponding training and validation results are shown in Figure 8.6 and Figure 8.7 for both channels. Besides all five individual folds, their average along with its standard deviation is additionally displayed. Investigating these ROC curves also contributes to the overall image of a successful training. The good agreement between individual folds implies the absence of critical overtraining. Visible differences between training and test results are still regarded as tolerable. Higher fluctuations in the validation dataset's ROC curves are expected due to the reduced sample size. Especially in the $2\ell\,\text{SS} + 1\tau_{\text{had}}$ channel, highly weighted processes such as $t\bar{t}$, with just 204 raw events, pose a problem and cause these fluctuations. Results in the $1\ell + 2\tau_{\text{had}}$ channel suggest that such effects are mitigated by increasing the number of simulated events. The residual responses provided by both remaining output nodes, i.e. $\text{NN}_{\text{background}}$ and $\text{NN}_{t\bar{t}}$ or $\text{NN}_{t\bar{t}\,+\,t\bar{t}X}$, are not included in this chapter. Instead, their corresponding distributions are shown in Appendix F along with the associated ROC curves. In summary, all investigated metrics are able to verify that a well optimised model is available in both the $1\ell + 2\tau_{\text{had}}$ channel and the $2\ell\,\text{SS} + 1\tau_{\text{had}}$ channel. Overtraining is not observed in any model. Furthermore, studying the validation dataset confirms that both models are successful at generalisation.

On this basis, a direct comparison between these final results and the previous optimisation is possible. The AUC, obtained from all output nodes' ROC curves, are summarised in Table 8.6 and Table 8.7. As the previous optimisation did not hold back an unbiased dataset, no comparison is possible in case of the prediction dataset. This comparison demonstrates conclusively that re-optimising the CNNs is justified. Here, the focus lies clearly on the signal response AUC. This result is ultimately the important one, as the associated network response $\text{NN}_{tH}$ should provide a variable which allows optimal discrimination between signal and background events. With regard to the validation datasets an improvement of 0.16 in the $1\ell + 2\tau_{\text{had}}$ channel and 0.01 in the $2\ell\,\text{SS} + 1\tau_{\text{had}}$ channel is obtained. The results belonging to the two background outputs fail to show any clear trend. In some cases, the previous optimisation is outperformed and vice versa. However, their outputs are not used in the further course of the analysis, which is why the result associated to the signal response is deemed decisive.

Subsequently, it is possible to examine whether the expected objective has been realised. As stated, the signal response $\text{NN}_{tH}$ of both networks is supposed to represent a variable that ensures the best possible separation of $tH$ and background events. To verify this assumption, all reconstructed variables are compared to $\text{NN}_{tH}$ in terms of their discriminative power. A quantification is obtained by the Jenson-Shannon divergence (JSD) [152]. This metric measures the difference between two probability distributions $Q$ and $P$ via

$$\text{JSD}(P \,||\, Q) = \frac{1}{2}\left(\text{KLD}(P \,||\, M) + \text{KLD}(Q \,||\, M)\right) , \tag{8.2}$$

with $M = (P+Q)/2$. In this equation, KLD is defined as the Kullback-Leibler divergence [153] given by

$$\mathrm{KLD}(P \,||\, Q) = \sum_i P_i \log\left(\frac{P_i}{Q_i}\right) , \tag{8.3}$$

where the sum runs over all $i$ discrete intervals, i.e. bins, of a distribution. The JSD is not only symmetrical but also limited to the interval $[0, 1]$, with higher values indicating larger differences.

The five variables yielding the highest JSD per channel are listed in Table 8.8. The calculation is performed by comparing the distributions of $tH$ events and the four dominant background processes per channel. Minor backgrounds do not enter this calculation to avoid biases caused by statistical fluctuations. This comparison reveals that the signal response $\mathrm{NN}_{tH}$ provides by far the largest discrepancy and is thus best suited to perform a $tH$ signal strength estimation on.

To conclude this evaluation, the exact distributions of $\mathrm{NN}_{tH}$ for all individual background processes are displayed in Figure 8.8 and Figure 8.9. Both distributions take only statistical uncertainties into account while $tH$ events are scaled up for better visibility.

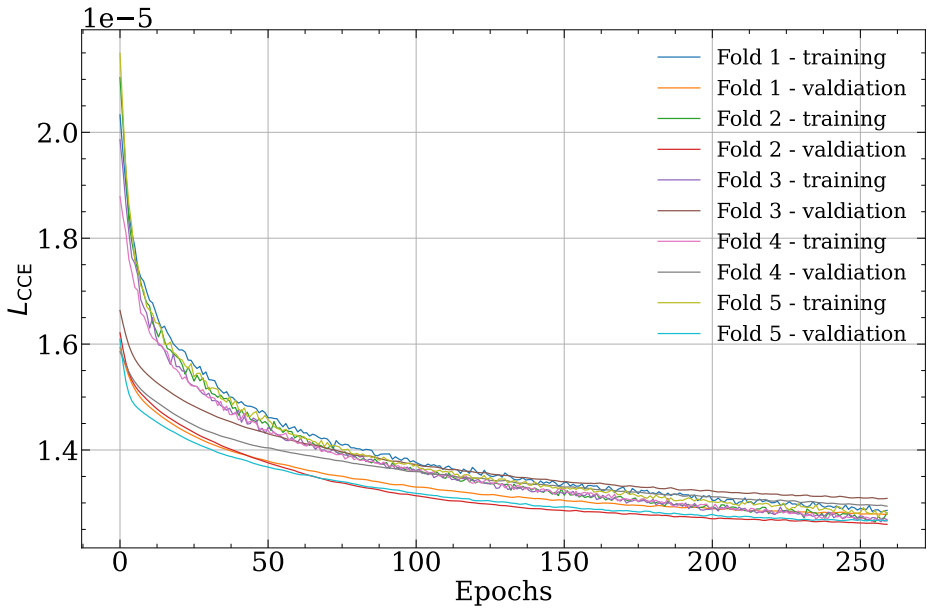



Figure 8.2: Categorical cross-entropy loss plotted for the training and validation dataset of all five folds in the $1\ell + 2\tau_{\mathrm{had}}$ channel. Training loss shows more fluctuations (see text for an explanation).

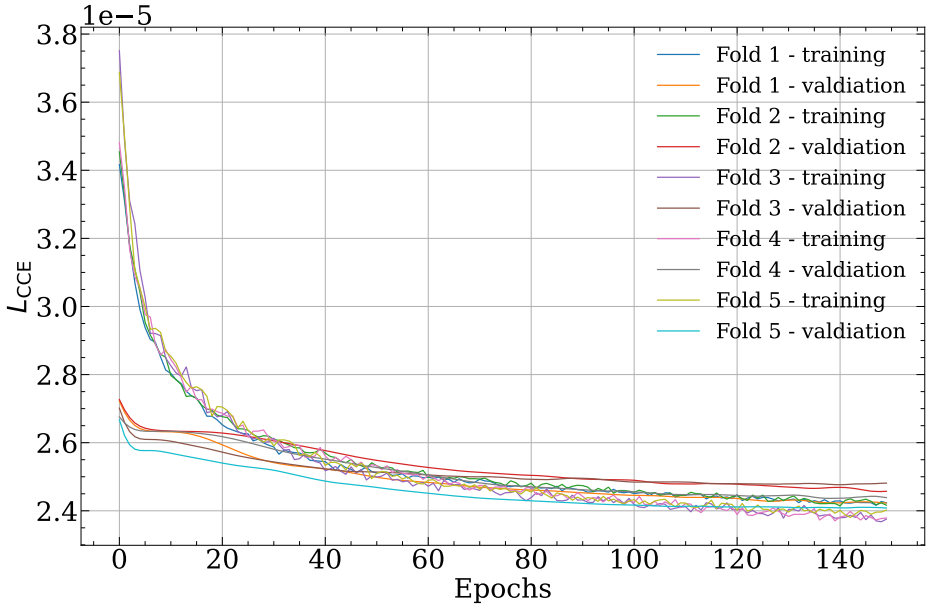



Figure 8.3: Categorical cross-entropy loss plotted for the training and validation dataset of all five folds in the $2\ell\,\mathrm{SS} + 1\tau_{\mathrm{had}}$ channel. Training loss shows more fluctuations (see text for an explanation).

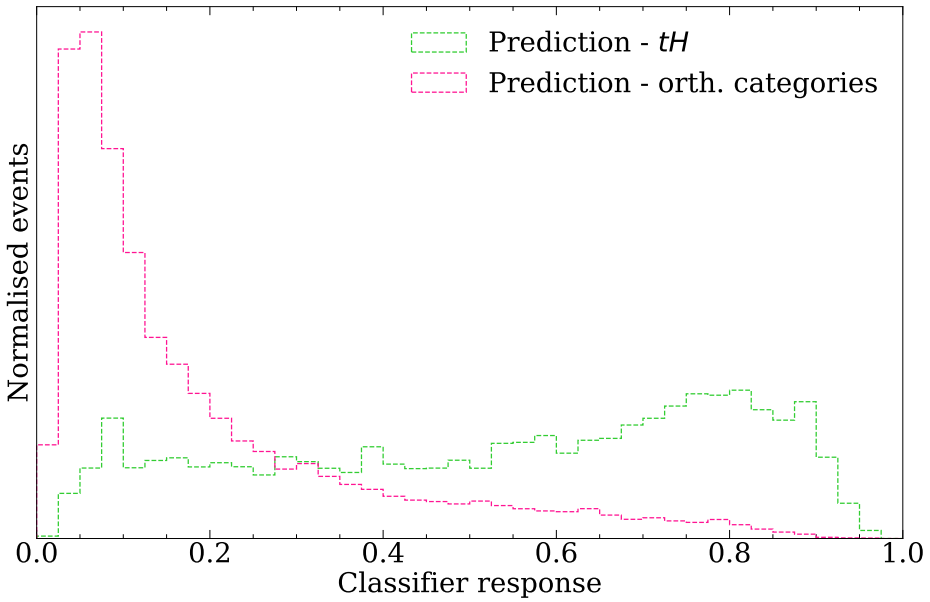



Figure 8.4: CNN response of the $tH$ output node based on the prediction dataset for each fold in the $1\ell + 2\tau_{\mathrm{had}}$ channel. All $t\bar{t}$ events are combined with the remaining background events and termed orthogonal categories.

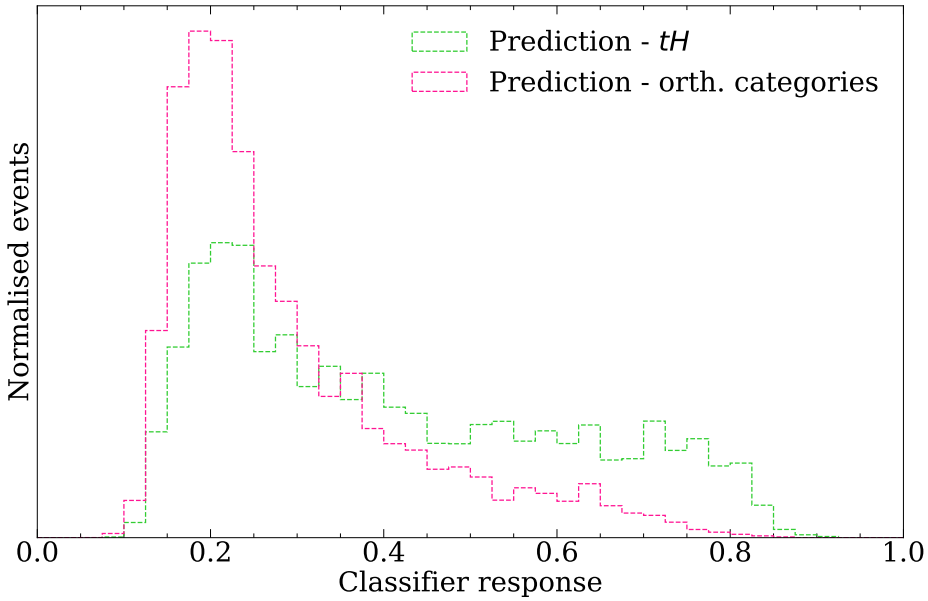
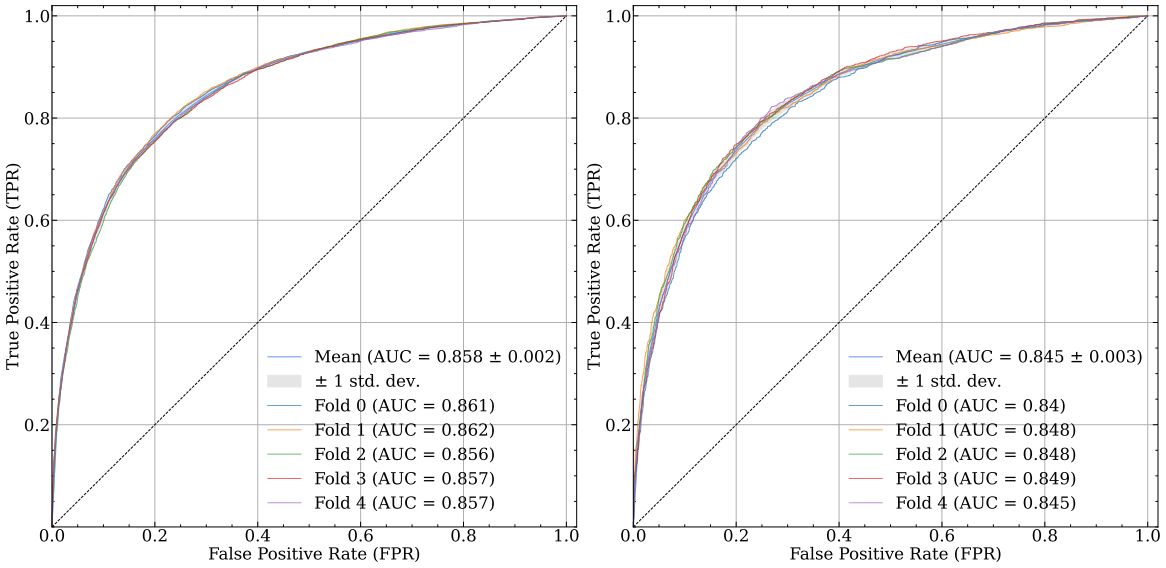
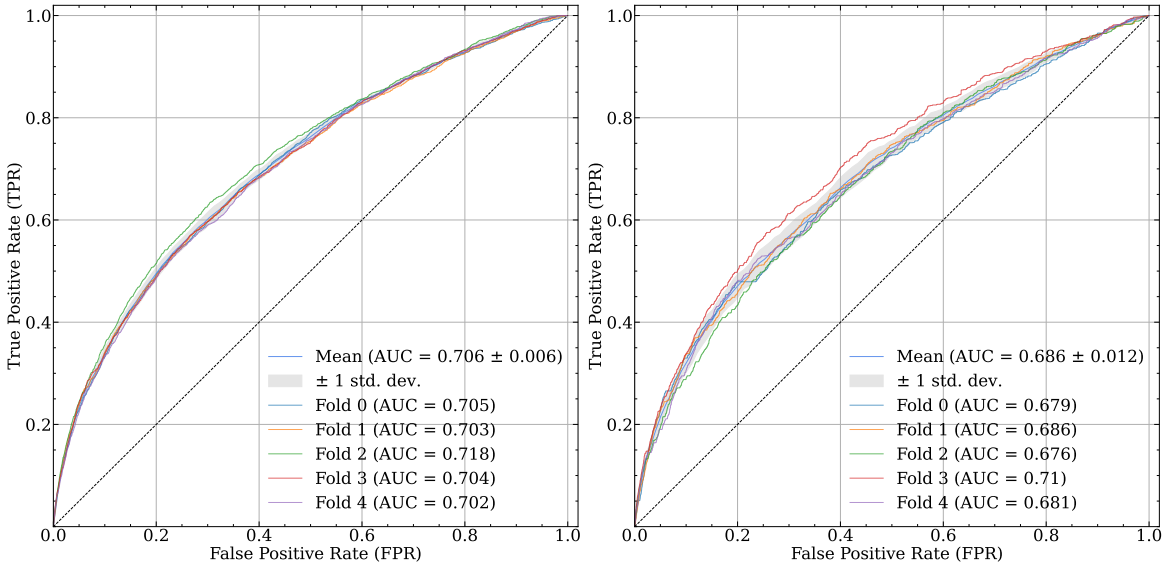



Figure 8.5: CNN response of the $tH$ output node based on the prediction dataset for each fold in the $2\ell\,\mathrm{SS}+1\tau_{\mathrm{had}}$ channel. All $t\bar{t}$ + $t\bar{t}X$ events are combined with the remaining background events and termed orthogonal categories.




(a) (b)

Figure 8.6: ROC curve corresponding to each model's $\mathrm{NN}_{tH}$ output based on (a) training data and (b) validation data in the $1\ell + 2\tau_{\mathrm{had}}$ channel. The average across all folds is shown along with its standard deviation.




(a) (b)

Figure 8.7: ROC curve corresponding to each model's $\mathrm{NN}_{tH}$ output based on (a) training data and (b) validation data in the $2\ell$ SS + $1\tau_{\mathrm{had}}$ channel. The average across all folds is shown along with its standard deviation.

Table 8.6: Comparison of the AUC values obtained from all output nodes' ROC curves between the previous optimisation and the final optimisation in the $1\ell + 2\tau_{\mathrm{had}}$ channel. The listed value in case of the final results corresponds to the mean of all five folds. As the previous optimisation did not provide a prediction dataset, no comparison is possible.

| Dataset | AUC - Previous optimisation | AUC - Final optimisation |
|---|---|---|
| Signal response training | 0.70 | 0.86 |
| Signal response validation | 0.69 | 0.85 |
| Signal response prediction | - | 0.85 |
| Background response training | 0.68 | 0.68 |
| Background response validation | 0.67 | 0.65 |
| Background response prediction | - | 0.65 |
| $t\bar{t}$ response training | 0.82 | 0.75 |
| $t\bar{t}$ response validation | 0.80 | 0.72 |
| $t\bar{t}$ response prediction | - | 0.72 |

Table 8.7: Comparison of the AUC values obtained from all output nodes' ROC curves between the previous optimisation and the final optimisation in the $2\ell$ SS + $1\tau_{\text{had}}$ channel. The listed value in case of the final results corresponds to the mean of all five folds. As the previous optimisation did not provide a prediction dataset, no comparison is possible.

| Dataset | AUC - Previous optimisation | AUC - Final optimisation |
|---|---|---|
| Signal response training | 0.67 | 0.71 |
| Signal response validation | 0.68 | 0.69 |
| Signal response prediction | - | 0.69 |
| Background response training | 0.61 | 0.75 |
| Background response validation | 0.61 | 0.71 |
| Background response prediction | - | 0.70 |
| $t\bar{t}$ + $t\bar{t}X$ response training | 0.82 | 0.78 |
| $t\bar{t}$ + $t\bar{t}X$ response validation | 0.81 | 0.75 |
| $t\bar{t}$ + $t\bar{t}X$ response prediction | - | 0.75 |

Table 8.8: Listing of the five variables providing the highest JSD in the $1\ell + 2\tau_{\text{had}}$ and $2\ell$ SS + $1\tau_{\text{had}}$ channel. All $tH$ events and the four dominant backgrounds per channel are used in the calculation.

| $1\ell + 2\tau_{\text{had}}$ | | $2\ell$ SS + $1\tau_{\text{had}}$ | |
|---|---|---|---|
| Variable | JSD | Variable | JSD |
| $\text{NN}_{tH}$ | 0.40 | $\text{NN}_{tH}$ | 0.24 |
| $p_{\text{T, visible Higgs}}$ | 0.30 | $M_{\text{lead. } b\text{-jet + forward jet}}$ | 0.14 |
| $\text{NN}_{\text{background}}$ | 0.27 | $\eta_{\text{forward jet}}$ | 0.12 |
| $p_{\text{T, lead. } \tau_{\text{had}}}$ | 0.26 | $p_{\text{T, } \ell_{\text{top}}}$ | 0.11 |
| $\text{MMC}_{\text{Out}}$ | 0.21 | $p_{\text{T, visible Higgs}}$ | 0.10 |

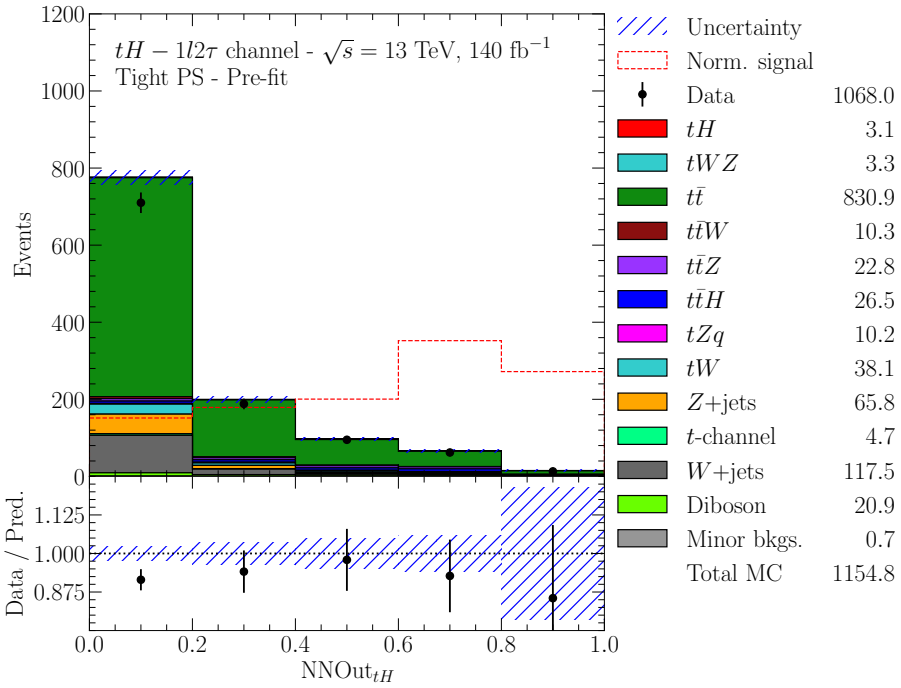



Figure 8.8: Final prediction of the signal output node $\mathrm{NN}_{tH}$ for all individual processes in the $1\ell + 2\tau_{\mathrm{had}}$ channel. Only statistical uncertainties are displayed. The signal contribution is normalised to match the total background yield and is shown as a red dashed line (consistent across all subsequent histograms).

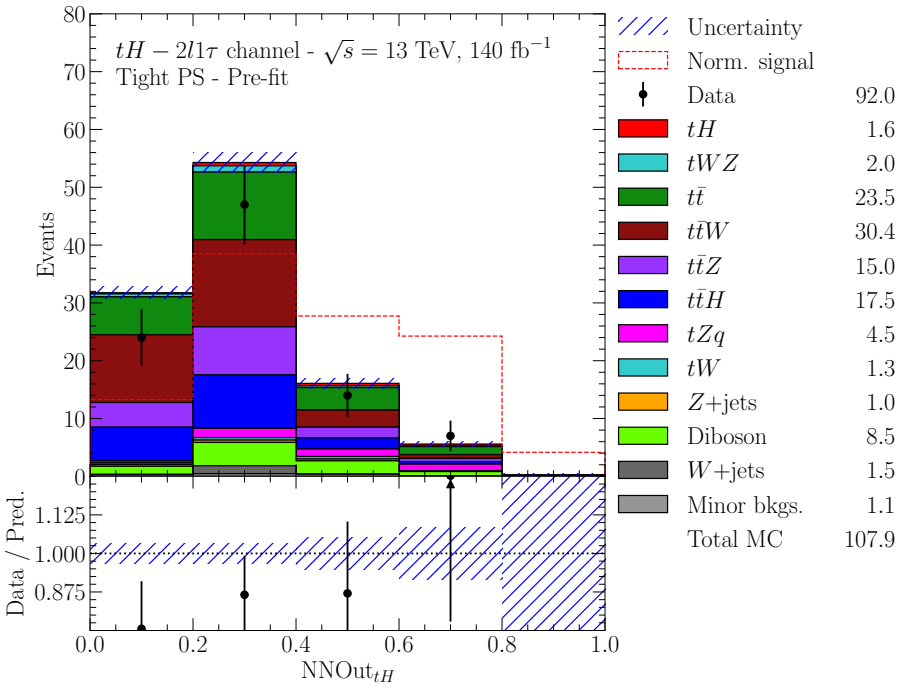



Figure 8.9: Final prediction of the signal output node NN$_{tH}$ for all individual processes in the $2\ell$ SS + $1\tau_{\text{had}}$ channel. Only statistical uncertainties are displayed.

CHAPTER 9

# Background estimation

Estimating the $tH$ signal strength via a profile likelihood fit in Chapter 10 represents the final and most important step of the analysis. Put simply, this fit scales the simulated $tH$ events by modifying the signal strength until the best agreement between data and MC simulation events is obtained. Obviously, such an estimation is only valid under the assumption that all relevant background processes are accurately modelled by their corresponding simulations. State-of-the-art MC generators are capable of providing such precise simulations for most event topologies. However, specific types of events, in particular those containing fake objects, mentioned in Section 2.3.2, are not expected to be modelled as accurately. Since the analysed $tH$ events feature both tau-leptons and light leptons in their final states, a distinction must be made. Events might feature fake tau-leptons, fake light leptons or even a combination of both.

The former are mainly caused by the misidentification of quark- or gluon-initiated jets in the detector. MC generators fail to correctly model those signatures as a detailed simulation of soft jet properties and the detector's response to hadrons is difficult [154].

In the case of light leptons, a distinction is made between actual fake light leptons caused by misidentified objects and non-prompt[1] leptons. Fake light leptons typically arise from light jets, which are falsely reconstructed and identified in the detector. In contrast, non-prompt leptons are real leptons that often originate from semi-leptonic hadron decays or, in the case of electrons, from photon conversions. The combined presence of these poorly modelled objects in selected events hinders an accurate determination of the $tH$ signal strength.

Initially, the tight preselection was introduced in Chapter 7 to suppress the amount of these fake events in the analysis. The obtained data to MC simulation agreement was displayed in Tables 7.2 to 7.4. Those results already indicate that the applied selection was only partially successful in mitigating disagreement caused by fake objects. As a consequence, a precise estimation of the remaining fake events is still a necessity and is discussed over the course of this chapter.

First of all, Section 9.1 outlines all relevant categories of fake events and how exploiting MC truth information enables their differentiation. Utilising this differentiation, the expected fake event rate in all channels is analysed for both the loose and tight preselection. This basis allows a decision to be made as to which corrections are required in each channel. Subsequently, Section 9.2 and Section 9.3

[1] A definition of prompt leptons was given in Chapter 3: prompt leptons include those that are created directly in the hard scattering vertex as well as those arising from the decay of heavy resonances such as the Higgs- and $W/Z$-bosons.

explain how these corrections are incorporated into the $1\ell + 2\tau_{\text{had}}$ channel, as well as the $2\ell + 1\tau_{\text{had}}$ channel. As part of this explanation, all regions that enter the final profile likelihood fit, presented in Chapter 10 are introduced.

## 9.1 Fake composition

Prior to discussing any channel-specific studies, this section is supposed to introduce the general methodology for fake background estimation. Firstly, an introduction on how misidentified events are detected and categorised in MC simulations is given. Subsequently, the concrete approach, chosen to correct individual contributions, is explained.

### 9.1.1 Event categorisation

The first step towards estimating all misidentified backgrounds entails an analysis of both the loose and tight preselection regions' compositions. This study, performed for all channels, tries to identify the dominant processes leading to incorrectly modelled tau-leptons and light leptons. The task is accomplished by exploiting MC truth information, which allows a mapping of reconstructed objects to the corresponding primordial particles.

In the case of hadronic tau-leptons $\tau_{\text{had}}$, this mapping is performed by the TauTruthMatchingTool [155]. This tool attempts to match the reconstructed jet, associated to the $\tau_{\text{had}}$, with a truth object within $\Delta R \leq 0.2$. The following categorisation is used:

- Reconstructed $\tau_{\text{had}}$ is matched to a truth tau-lepton but not to a truth jet.
- Reconstructed $\tau_{\text{had}}$ is matched to a truth electron but not to a truth jet.
- Reconstructed $\tau_{\text{had}}$ is matched to a truth jet initiated by a quark.
- Reconstructed $\tau_{\text{had}}$ is matched to a truth jet initiated by a gluon.

Only the first category is considered to be a prompt $\tau_{\text{had}}$, while all remaining categories are grouped under the label fake-$\tau$ to avoid statistical fluctuations caused by small sample sizes. Their contribution is simply listed as "$\tau$-fakes" in all graphics.

In contrast, the charged light leptons, i.e. electrons and muons, are classified according to the TruthIFFClassification [156] scheme. Similar to the hadronic tau-leptons, this scheme allows leptons to be split into the listed classes:

- Prompt electrons or prompt muons: electrons or muons created in the hard scattering vertex or in the decay of heavy resonances.
- Electrons with misidentified charge: electrons whose electric charge is mistakenly reconstructed with the opposite sign. This is caused by an incorrectly measured track curvature. Electrons with high momentum are more probable to fall into this category, as their less curved track has a higher chance of being incorrectly reconstructed. In addition, oppositely charged electrons or positrons from a Bremsstrahlung photon also need to be considered. They can contribute to this category by depositing energy reconstructed in the calorimeters as part of prompt electrons.

- Muons with misidentified charge: similar to electrons, a muon can also be falsely reconstructed with the opposite sign electric charge. However, this case is strongly suppressed by the additional measurement performed in the detector's MS.
- Prompt $\gamma$-conversions: this category comprises all electrons that are produced when a prompt photon undergoes the process of pair production.
- Hadronic tau decays: electrons or muons that are generated in the cascade of hadronically decaying tau-leptons.
- Hadron decays: electrons or muons that arise from the decay of individual hadrons within a particle jet.

Events which contain charge-flip electrons or muons are merged into one category. The last two categories, i.e. events with tau or quark decays which are misidentified as light leptons or events that generate non-prompt leptons, are termed fake-$e$ and fake-$\mu$ events, or fake light leptons when combined. In analogy to fake-$\tau$ events, the fake light lepton contributions are termed "$e/\mu$-fakes" in all graphics.

As all analysed channels feature three leptons, including $\tau_{\text{had}}$, a single event might contain more than one misidentified object. Therefore, events might simultaneously fall into several of the categories mentioned above. Taking this into account, different compositions are possible depending on the concrete channel. For example, an event entering $1\ell + 2\tau_{\text{had}}$ channel can yield up to two fake-$\tau$ objects, whereas $2\ell + 1\tau_{\text{had}}$ events can have a maximum of one.

Unfortunately, neither the classification tools nor the MC simulations are without flaws. As a consequence, cases where the classification fails due to a lack of information also need to be considered. Corresponding events are summarised in an additional category termed "other fake" events. This group of events is not expected to make a significant contribution to the analysis and is therefore not scaled in the final corrections.

Applying the complete categorisation to the loose and tight preselection regions of all channels results in the compositions shown in Tables 9.1 to 9.3. Although the breakdown presented in each table displays the complete composition of one region, it does not allow any conclusions to be drawn about the composition of any individual background processes. Therefore, to provide a complete picture, the categorisation for each process can be found in Appendix G.

Comparing all results indicates a quite inhomogeneous composition across the three channels. The $1\ell + 2\tau_{\text{had}}$ channel, for example, is clearly dominated by fake-$\tau$ events, contributions from fake-$e$ and fake-$\mu$ events are negligible. The $2\ell\,\text{SS} + 1\tau_{\text{had}}$ channel, on the other hand, is populated by a significant amount of both fake-$\tau$ and fake-$e/\mu$ events. As a consequence, a separate background estimation is carried out per channel as discussed in Section 9.2 and Section 9.3. Prior to that, the general estimation strategy, shared by both channels, is explained.

### 9.1.2 Fake background estimation strategy

The general strategy for estimating the misidentified background processes is identical in both the $1\ell + 2\tau_{\text{had}}$ and $2\ell\,\text{SS} + 1\tau_{\text{had}}$ channel. However, their relative fractions of fake-$\tau$ and fake-$e/\mu$ are quite different. Respecting this inhomogeneous composition, the effect of misidentified backgrounds is individually estimated in each channel.

Table 9.1: Weighted yields in the $1\ell + 2\tau_{\text{had}}$ channel's loose and tight preselection regions after applying the fake event classification.

| $1\ell + 2\tau_{\text{had}}$ | Weighted yields | |
|---|---|---|
| Process | Loose preselection | Tight preselection |
| $tH$ | 3.403 ± 0.063 | 2.485 ± 0.054 |
| $tWZ$ | 3.558 ± 0.034 | 2.625 ± 0.029 |
| $t\bar{t}$ | – | – |
| $t\bar{t}W$ | 6.45 ± 0.19 | 4.73 ± 0.16 |
| $t\bar{t}Z$ | 22.12 ± 0.44 | 16.57 ± 0.38 |
| $t\bar{t}H$ | 26.73 ± 0.19 | 19.81 ± 0.16 |
| $tZq$ | 12.04 ± 0.19 | 8.62 ± 0.16 |
| $tW$ | – | – |
| $Z$ + jets | – | – |
| $t$-channel | – | – |
| $W$ + jets | – | – |
| Diboson | 19.66 ± 0.59 | 14.23 ± 0.44 |
| Minor bkg. | 0.74 ± 0.59 | 0.104 ± 0.011 |
| Fake-$e$ | 10.0 ± 1.1 | 0.60 ± 0.21 |
| Fake-$\mu$ | 16.4 ± 1.3 | 0.27 ± 0.15 |
| Charge-misidentification | 0.0259 ± 0.0008 | 0.0150 ± 0.0062 |
| $\gamma$-conversion | 0.280 ± 0.098 | 0.131 ± 0.032 |
| Fake-$\tau$ | 1 229 ± 13 | 559.4 ± 8.6 |
| Fake-$\tau$ + fake-$e$ | 13.4 ± 1.3 | 0.60 ± 0.25 |
| Fake-$\tau$ + fake-$\mu$ | 23.9 ± 1.8 | 0.30 ± 0.17 |
| Fake-$\tau$ + charge-misidentification | 0.40 ± 0.22 | 0.16 ± 0.14 |
| Fake-$\tau$ + $\gamma$-conversion | 2.75 ± 0.87 | 1.07 ± 0.72 |
| Fake-$\tau$ + fake-$\tau$ | 1 514 ± 29 | 522 ± 18 |
| Fake-$\tau$ + fake-$\tau$ + fake-$e$ | 5.7 ± 1.1 | 0.12 ± 0.10 |
| Fake-$\tau$ + fake-$\tau$ + fake-$\mu$ | 8.8 ± 1.5 | 0.11 ± 0.10 |
| Fake-$\tau$ + fake-$\tau$ + charge-misidentification | – | – |
| Fake-$\tau$ + fake-$\tau$ + $\gamma$-conversion | 1.98 ± 0.56 | 0.74 ± 0.42 |
| Other fakes | 1.39 ± 0.53 | 0.58 ± 0.45 |
| Total | 2 922 ± 32 | 1 155 ± 20 |

Table 9.2: Weighted yields in the $2\ell$ SS + $1\tau_{\text{had}}$ channel's loose and tight preselection regions after applying the fake event classification. Rare cases containing a fake-$e/\mu$ alongside a second light lepton which is assigned to the "charge-misidentification" or "$\gamma$-conversion" category are included in the fake-$e/\mu$ contribution.

| $2\ell$ SS + $1\tau_{\text{had}}$ | Weighted yields | |
|---|---|---|
| Process | Loose preselection | Tight preselection |
| $tH$ | 2.438 ± 0.055 | 1.375 ± 0.041 |
| $tWZ$ | 2.530 ± 0.028 | 1.464 ± 0.022 |
| $t\bar{t}$ | 2.63 ± 0.60 | 0.25 ± 0.18 |
| $t\bar{t}W$ | 31.11 ± 0.44 | 20.56 ± 0.36 |
| $t\bar{t}Z$ | 18.87 ± 0.39 | 10.75 ± 0.29 |
| $t\bar{t}H$ | 24.66 ± 0.17 | 13.91 ± 0.13 |
| $tZq$ | 5.86 ± 0.13 | 3.169 ± 0.098 |
| $tW$ | – | – |
| $Z$ + jets | – | – |
| Diboson | 10.55 ± 0.38 | 5.73 ± 0.25 |
| $W$ + jets | – | – |
| Minor bkg. | 0.93 ± 0.33 | 0.38 ± 0.02 |
| Fake-$e$ | 882 ± 11 | 5.03 ± 0.72 |
| Fake-$\mu$ | 851 ± 11 | 7.47 ± 0.98 |
| Charge-misidentification | 0.133 ± 0.026 | 0.062 ± 0.021 |
| $\gamma$-conversion | 17.6 ± 1.5 | 3.56 ± 0.65 |
| Fake-$e$ + fake-$e$ | 0.94 ± 0.33 | 0.006 ± 0.004 |
| Fake-$e$ + fake-$\mu$ | 2.25 ± 0.53 | 0.032 ± 0.012 |
| Fake-$\mu$ + fake-$\mu$ | 1.79 ± 0.47 | 0.048 ± 0.022 |
| Fake-$\tau$ | 48.4 ± 1.2 | 21.08 ± 0.46 |
| Fake-$\tau$ + fake-$e$ | 761 ± 11 | 4.36 ± 1.21 |
| Fake-$\tau$ + fake-$\mu$ | 705.6 ± 9.9 | 3.97 ± 0.71 |
| Fake-$\tau$ + charge-misidentification | 4.83 ± 0.71 | 1.01 ± 0.28 |
| Fake-$\tau$ + $\gamma$-conversion | 13.4 ± 1.3 | 1.64 ± 0.42 |
| Fake-$\tau$ + fake-$e$ + fake-$e$ | 1.45 ± 0.59 | – |
| Fake-$\tau$ + fake-$e$ + fake-$\mu$ | 1.12 ± 0.47 | 0.003 ± 0.003 |
| Fake-$\tau$ + fake-$\mu$ + fake-$\mu$ | 0.40 ± 0.25 | – |
| Other fakes | 32.1 ± 2.2 | 2.08 ± 0.48 |
| Total | 3 423 ± 22 | 107.9 ± 2.2 |

Table 9.3: Weighted yields in the $2\ell$ OS + $1\tau_{\text{had}}$ channel's loose and tight preselection regions after applying the fake event classification. Rare cases containing a fake-$e/\mu$ alongside a second light lepton which is assigned to the "charge-misidentification" or "$\gamma$-conversion" category are included in the fake-$e/\mu$ contribution.

| $2\ell$ OS + $1\tau_{\text{had}}$ | Weighted yields | |
|---|---|---|
| Process | Loose preselection | Tight preselection |
| $tH$ | 3.817 ± 0.072 | 2.26 ± 0.05 |
| $tWZ$ | 17.515 ± 0.073 | 11.52 ± 0.06 |
| $t\bar{t}$ | 2.75 ± 0.60 | 0.26 ± 0.19 |
| $t\bar{t}W$ | 67.01 ± 0.68 | 44.39 ± 0.52 |
| $t\bar{t}Z$ | 115.52 ± 0.84 | 73.83 ± 0.66 |
| $t\bar{t}H$ | 55.16 ± 0.22 | 32.65 ± 0.16 |
| $tZq$ | 43.29 ± 0.37 | 27.79 ± 0.30 |
| $tW$ | 0.25 ± 0.18 | – |
| $Z$ + jets | 0.16 ± 0.13 | 0.04 ± 0.03 |
| Diboson | 113.3 ± 1.3 | 72.03 ± 0.95 |
| Other Higgs | 4.3 ± 1.4 | 2.7 ± 1.2 |
| Minor bkg. | 1.473 ± 0.038 | 0.92 ± 0.03 |
| Fake-$e$ | 873 ± 11 | 7.24 ± 0.87 |
| Fake-$\mu$ | 846 ± 11 | 6.87 ± 0.88 |
| Charge-misidentification | 0.46 ± 0.23 | 0.05 ± 0.02 |
| $\gamma$-conversion | 17.8 ± 1.4 | 4.76 ± 0.72 |
| Fake-$e$ + fake-$e$ | 1.94 ± 0.52 | 0.05 ± 0.01 |
| Fake-$e$ + fake-$\mu$ | 7.35 ± 0.99 | 0.05 ± 0.01 |
| Fake-$\mu$ + fake-$\mu$ | 3.86 ± 0.70 | 0.02 ± 0.02 |
| Fake-$\tau$ | 27 780 ± 100 | 10 435 ± 51 |
| Fake-$\tau$ + fake-$e$ | 1 161 ± 13 | 10.2 ± 1.2 |
| Fake-$\tau$ + fake-$\mu$ | 1 130 ± 13 | 7.27 ± 0.98 |
| Fake-$\tau$ + charge-misidentification | 0.80 ± 0.33 | 0.02 ± 0.01 |
| Fake-$\tau$ + $\gamma$-conversion | 114.2 ± 4.8 | 27.9 ± 2.0 |
| Fake-$\tau$ + fake-$e$ + fake-$e$ | 1.75 ± 0.60 | 0.01 ± 0.01 |
| Fake-$\tau$ + fake-$e$ + fake-$\mu$ | 5.6 ± 1.1 | 0.01 ± 0.01 |
| Fake-$\tau$ + fake-$\mu$ + fake-$\mu$ | 4.3 ± 1.1 | 0.05 ± 0.04 |
| Other fakes | 121.3 ± 4.2 | 21.6 ± 1.2 |
| Total | 32 490 ± 100 | 10 790 ± 52 |

The fundamental idea envisages the creation of background control regions (CRs) that are completely or at least predominately composed of a single poorly modelled process, such as fake-$\tau$ events. In such a region, it is assumed that any disagreement between data and simulation can be attributed to this single process. Subsequently, the disagreement is corrected in a profile likelihood fit by rescaling the process' weighted contribution to match real data as described in Chapter 10.

To ensure that all corrections are successful and justified, additional validation regions (VRs) need to be defined. These regions do not enter the profile likelihood fit. Instead, all corrections obtained from a fit are applied to the VRs afterwards. This allows the effectiveness of all corrections to be validated.

Once such corrections are applied and validated, the background processes are expected to be properly modelled within the limits of their uncertainties. Broadly speaking, any remaining disagreement between data and simulation can be attributed to the $tH$ process, which will be adjusted by fitting the signal strength.

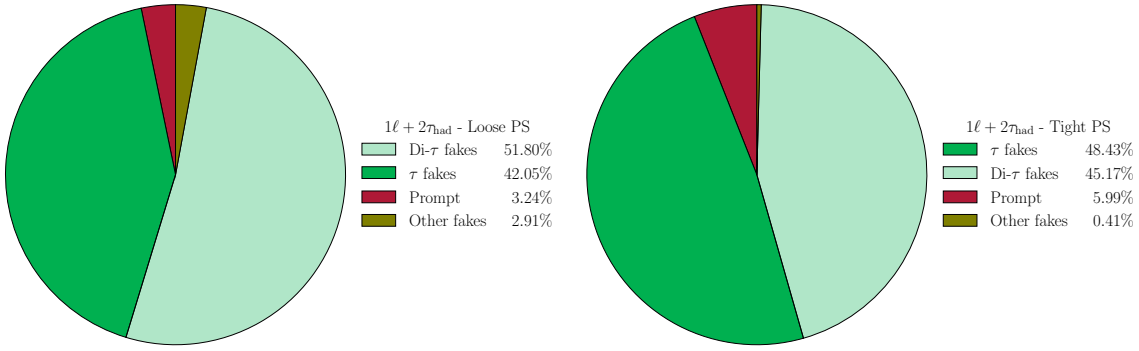
A correctly performed profile likelihood fit requires that each individual event appears in only one statistically independent region. Respecting this condition, the obvious choice for a region to perform the signal strength extraction in is each channel's tight preselection region, given its relatively high signal significance. As a consequence, orthogonality with respect to this tight preselection represents a mandatory requirement for each region of background estimation. Therefore, CRs and VRs are defined within the loose preselection region. These regions offer an increased fraction of fake events as an outcome of their more relaxed selection criteria, shown in Tables 9.1 to 9.3. This way, any obtained correction will be less susceptible to statistical fluctuations.

The two following sections discuss the concrete region definition for each channel. In addition, correlations between the strength of data to simulation disagreement and kinematic variables are investigated and taken into account where necessary. To simplify this procedure, all fake processes that contribute less than 1 % to a channel's tight preselection region are included in the "other fakes" category. The residual prompt background events are merged in a similar way. All contributions lower than the $tH$ contribution supplement each channel's minor backgrounds introduced in Section 7.1.

## 9.2 $1\ell + 2\tau_{\text{had}}$ - background estimation

In the following, the strategy presented in Section 9.1.2 is implemented in the $1\ell + 2\tau_{\text{had}}$ channel. Initially, all dominant poorly modelled fake processes, that contribute more than 1 % to the tight preselection, are identified based on Table 9.1. Only events with one or two fake-$\tau$ objects need to be considered. The latter are referred to as fake-di-$\tau$ events. In contrast, fake light leptons account for less than one percent of the tight preselection region and are consequently included but not modified in the estimation. Following the outlined strategy, they are instead assigned to the "other fakes" category along with further insignificant processes. The resulting breakdown of the loose and tight preselection regions is shown in Figure 9.1, where each individual process is ranked according to its relative contribution. The displayed prompt category includes both signal and background processes.

Firstly, reconstructed variables are examined to identify correlations between event signatures or kinematics and the level of disagreement between real data and simulation. A significant disagreement is immediately evident in variables where $W$ + jets events are centred in a narrow range, as depicted in Figure 8.8. A more detailed investigation reveals that this pronounced disagreement is caused by a few, relatively highly weighted $W$ + jets events. Inspecting their signatures paints an almost identical




(a) (b)

Figure 9.1: Composition in (a) the loose and (b) the tight preselection regions of the $1\ell + 2\tau_{\text{had}}$ channel. Signal and background events are both merged in the prompt contribution while small fake processes are grouped together as "other fakes".

picture for all events, visualised in Figure 9.2. Both reconstructed $\tau_{\text{had}}$ objects pass through the detector almost back-to-back with respect to the azimuthal angle ($\Delta\phi_{\text{lead. } \tau_{\text{had}}, \text{ sublead. } \tau_{\text{had}}} \approx \pi$), while located in a nearly identical pseudorapidity region ($\Delta\eta_{\text{lead. } \tau_{\text{had}}, \text{ sublead. } \tau_{\text{had}}} \approx 0$). Without exception, all events in question possess at least one fake-$\tau$ as according to Table 9.1 no prompt $W$ + jets events enter the $1\ell + 2\tau_{\text{had}}$ tight preselection region.

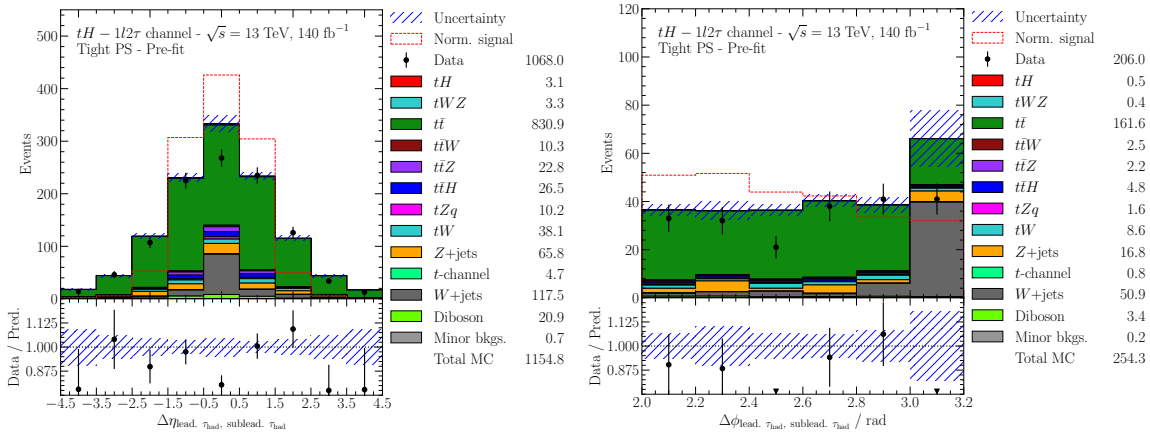
Given the level of disagreement, these $W$ + jets events would strongly affect the correction of all other fake-$\tau$ events. To avoid such bias, the selection $\Delta\phi_{\text{lead. } \tau_{\text{had}}, \text{ sublead. } \tau_{\text{had}}} < 3.05$ is applied in all regions of the $1\ell + 2\tau_{\text{had}}$ channel, on top of the general preselection.

Subsequently, further potential correlations between the disagreement level and kinematic information or special event signatures are studied. To begin with, dependencies on the $p_{\text{T}}$ and $\eta$ of the leading and subleading[2] $\tau_{\text{had}}$ are investigated. The corresponding distributions are given in Figure 9.3 and Figure 9.4. While no constant level of disagreement is observed in the leading $\tau_{\text{had}}$'s $\eta$, the $p_{\text{T}}$ distribution shows a clear trend. In case of the subleading $\tau_{\text{had}}$, the $\eta$ distribution shows a fluctuating level of disagreement but no clear trend. The $p_{\text{T}}$ distribution, on the other hand, mirrors the trend observed for the leading $\tau_{\text{had}}$. The disagreement generally increases along with the $\tau_{\text{had}}$'s transverse momentum. Although the trend is not exactly identical for the leading and subleading $\tau_{\text{had}}$, the differences are covered by the uncertainties.

This correlation is investigated in more detail by additionally examining differences between 1-prong and 3-prong $\tau_{\text{had}}$ decays (see Section 3.5.3). Their respective distributions are summarised in Figure 9.5 as well as Figure 9.6. Besides a mere visual comparison of all distributions, additional preliminary profile likelihood fits are conducted. These allow for a direct verification of whether the required corrections for 1-prong and 3-prong decays differ. Again, no statistically significant differences are observed. Therefore, it is assumed that only the fake-$\tau$'s transverse momentum

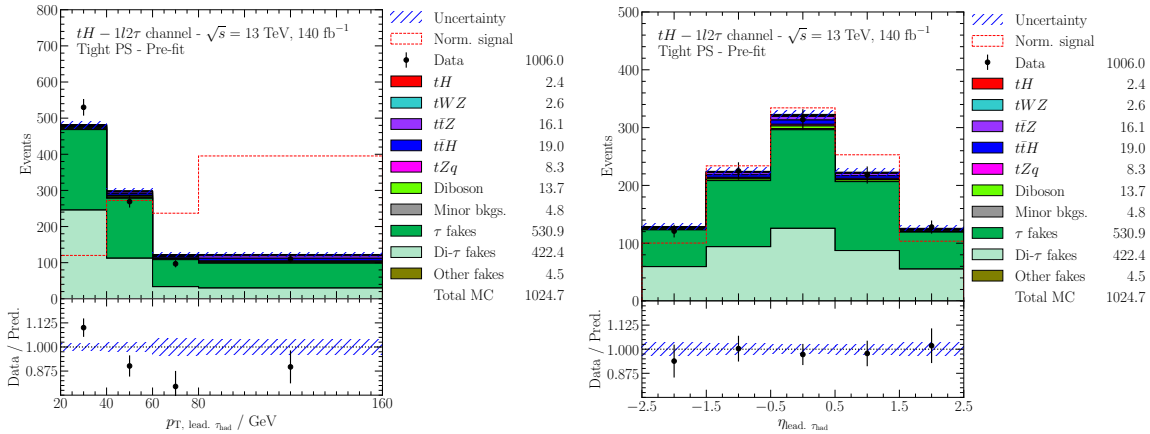
[2] As outlined in Section 7.1, identical physics objects, in this case $\tau_{\text{had}}$, are commonly ordered according to their $p_{\text{T}}$. The $\tau_{\text{had}}$ with the highest $p_{\text{T}}$ is referred to as the leading $\tau_{\text{had}}$, while the one with the second highest $p_{\text{T}}$ is termed the subleading $\tau_{\text{had}}$. The same convention applies to light leptons.

influences the level of data to simulation disagreement.



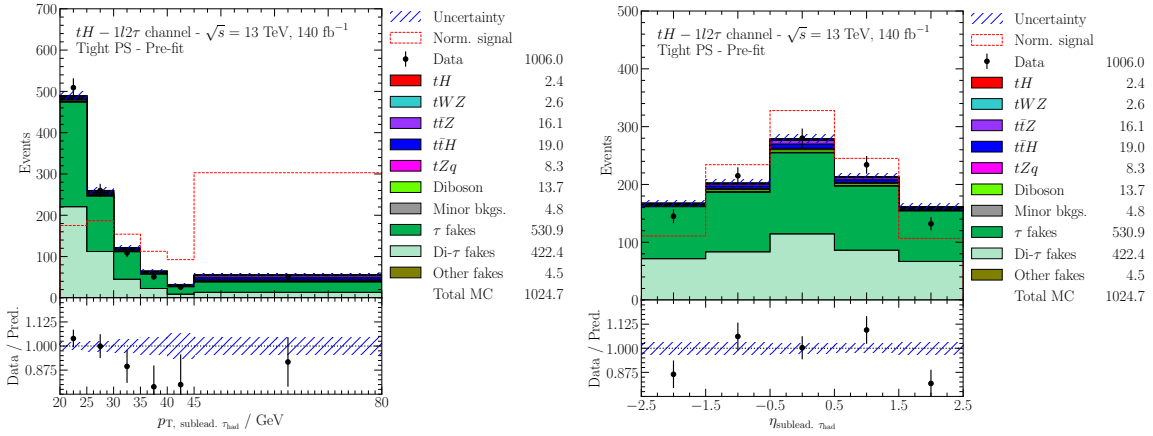

(a) (b)

Figure 9.2: Distributions of both $\tau_{\text{had}}$'s difference in (a) pseudorapidity $\Delta\eta_{\text{lead. }\tau_{\text{had}}, \text{ sublead. }\tau_{\text{had}}}$ and (b) azimuthal angle $\Delta\phi_{\text{lead. }\tau_{\text{had}}, \text{ sublead. }\tau_{\text{had}}}$ in the $1\ell + 2\tau_{\text{had}}$ channel's tight preselection region. In both variables, $W + \text{jets}$ events are clearly peaked in bins with large disagreement between data and simulation. Only statistical uncertainties are displayed.



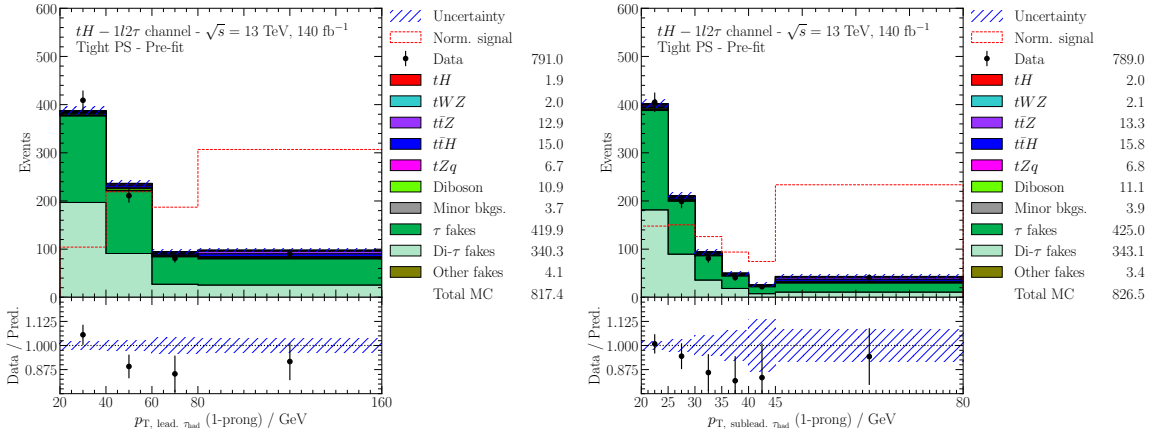

(a) (b)

Figure 9.3: Correlation studies showing (a) $p_{\text{T}}$ and (b) $\eta$ of the leading $\tau_{\text{had}}$ in the $1\ell + 2\tau_{\text{had}}$ channel's tight preselection region. Only statistical uncertainties are displayed.



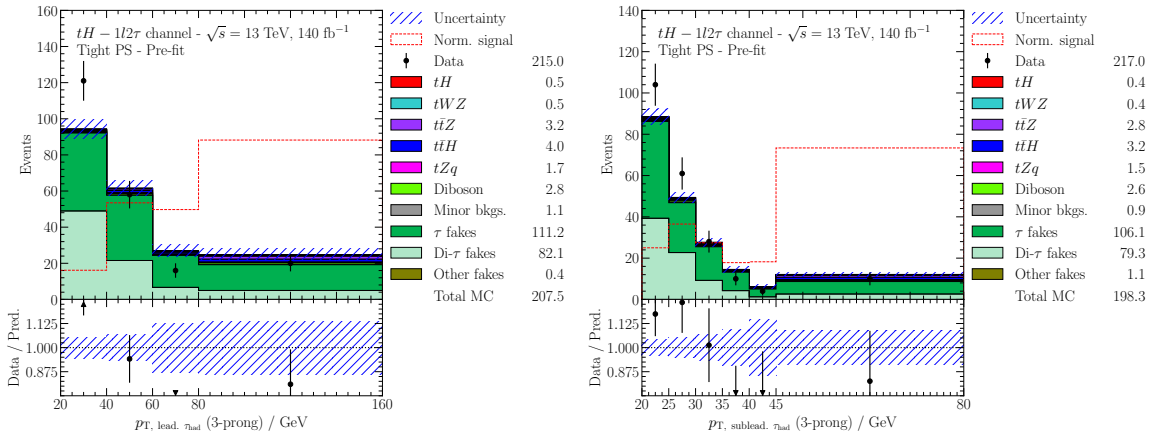

(a) (b)

Figure 9.4: Correlation studies showing (a) $p_T$ and (b) $\eta$ of the subleading $\tau_{had}$ in the $1\ell + 2\tau_{had}$ channel's tight preselection region. Only statistical uncertainties are displayed.




(a) (b)

Figure 9.5: Correlation studies showing the transverse momentum for (a) leading and (b) subleading 1-prong $\tau_{had}$ decays in the $1\ell + 2\tau_{had}$ channel's tight preselection region. Only statistical uncertainties are displayed.




(a) (b)

Figure 9.6: Correlation studies showing the transverse momentum for (a) leading and (b) subleading 3-prong $\tau_{\text{had}}$ decays in the $1\ell + 2\tau_{\text{had}}$ channel's tight preselection region. Only statistical uncertainties are displayed.

### 9.2.1 Region definitions

In the following, all regions which enter the profile likelihood fit are defined. The exact definitions take the outcome of the previously presented studies into account.

At the beginning, the tight preselection region is divided into two regions. The first region, called the signal region (SR), represents the area in which the highest signal significance is to be expected. It is defined via the CNN's signal response by selecting all events that fulfil $\text{NN}_{tH} > 0.45$. In addition, an orthogonal VR is defined via $\text{NN}_{tH} < 0.45$. As a result, the VR's composition is, apart from a lower relative fraction of $tH$ events, as similar as possible to the SR. This proves to be useful later on, when analysing whether the correction of poorly modelled fake-$\tau$ events was successful[3].

Since the tight preselection is a subset of the loose preselection region, only a fraction of the events are left to form the CRs, needed to correct all fake-$\tau$ contributions. On the one hand, these regions are required to be completely dominated by these particular fake processes. On the other hand, their phase space should not differ too much from both the SR and the VR, while still being orthogonal. Choosing a similar phase space region ensures that the corrected fakes have comparable signatures. Two regions matching these criteria are obtained via the following conditions:

- Events in which the leading $\tau_{\text{had}}$ fulfils the medium RNN working point while the subleading $\tau_{\text{had}}$ passes the loose RNN working point selection but fails the medium one (termed anti-medium subleading $\tau_{\text{had}}$).
- Events in which the subleading $\tau_{\text{had}}$ fulfils the medium RNN working point while the leading $\tau_{\text{had}}$ passes the loose RNN working point selection but fails the medium one (termed anti-medium leading $\tau_{\text{had}}$).

The tight preselection criteria of the single light lepton are retained. Given the fact that no significant

[3] The relevance of this region is more evident when taking into account that the analysis was initially blinded. This principle, thoroughly introduced in Chapter 10, is supposed to prevents biasing an analysis in a similar way as blinding in clinical trials.

differences in the strength of disagreement with respect to the leading and subleading $\tau_{\text{had}}$ were found, both regions are merged. As an outcome, the increased event count facilitates a statistically robust estimation.

As a next step, this entire region is further partitioned to account for the correlation between poorly modelled fake-$\tau$ events and the transverse momentum. The exact split into several $p_{\text{T}}$-bins is a trade-off: choosing a fine binning allows to model the trend more accurately at the risk of any correction being distorted by statistical fluctuations. In this concrete analysis, it is decided to use a total of three $p_{\text{T}}$ bins. Their corresponding CRs are defined as follows:

- $\text{CR}_{\text{fake-}\tau,p_{\text{T},1}}$: $20\,\text{GeV} \leq$ anti-medium $\tau_{\text{had}}$ $p_{\text{T}} < 30\,\text{GeV}$.
- $\text{CR}_{\text{fake-}\tau,p_{\text{T},2}}$: $30\,\text{GeV} \leq$ anti-medium $\tau_{\text{had}}$ $p_{\text{T}} < 60\,\text{GeV}$.
- $\text{CR}_{\text{fake-}\tau,p_{\text{T},3}}$: $60\,\text{GeV} \leq$ anti-medium $\tau_{\text{had}}$ $p_{\text{T}}$.

As noted, associating events to a region is carried out based on the anti-medium $\tau_{\text{had}}$'s momentum. Given their looser selection criteria, these $\tau_{\text{had}}$ have a higher probability of being a fake-$\tau$.

The composition of each region is shown in Figure 9.7 and displays the expected results. The CRs and the VR are clearly dominated by fake-$\tau$ or fake-di-$\tau$ contributions. The SR is likewise dominated by both kinds of fake events. However, this region simultaneously features the largest relative share of prompt events. Other fake sources are negligible throughout all regions. A more detailed listing, broken down into all individual processes in analogy to Table 9.1, can be found in Appendix G. Ultimately, all regions except the VR serve as input to a single profile likelihood fit as discussed in Chapter 10. By adding free parameters to this fit, both the signal strength and the $p_{\text{T}}$-dependent fake-$\tau$ corrections can be estimated in parallel.

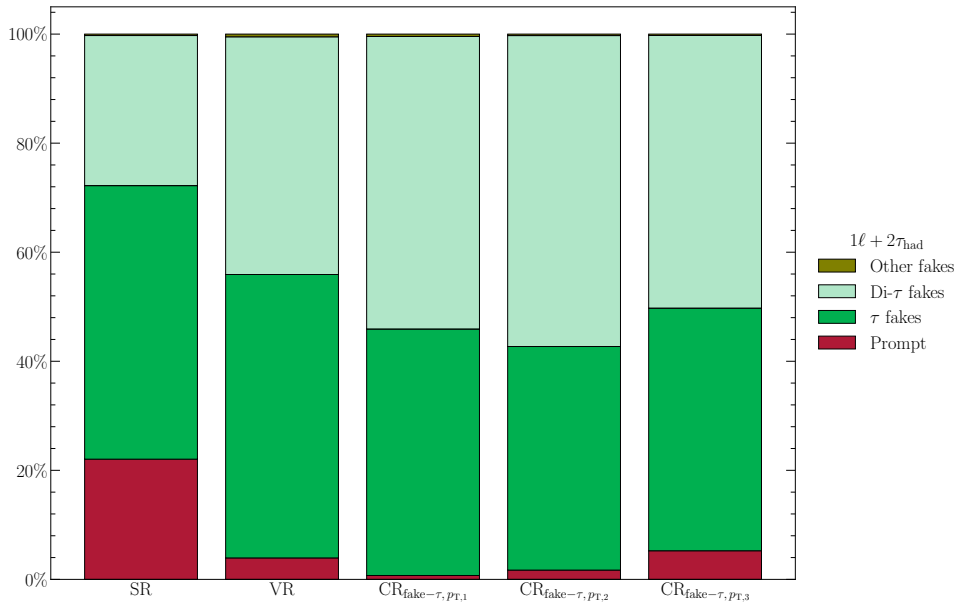



Figure 9.7: Composition of all regions entering the $1\ell + 2\tau_{\text{had}}$ channel's profile likelihood fit.

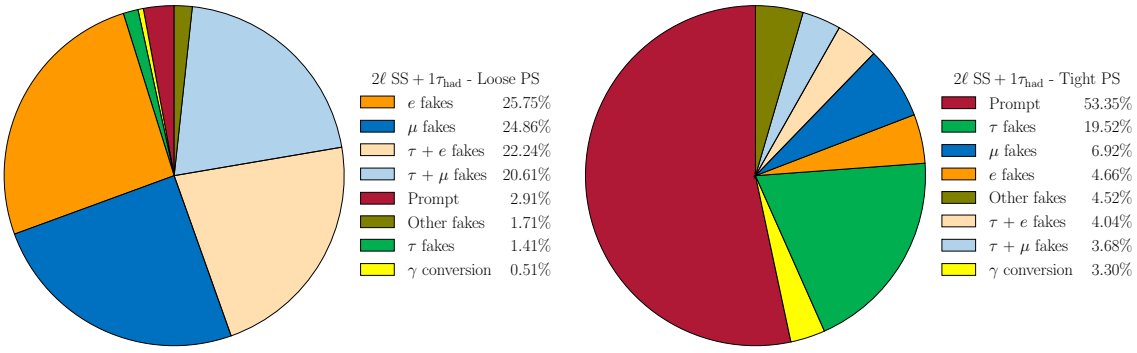
## 9.3 $2\ell + 1\tau_{\text{had}}$ - background estimation

In the following, the fake estimation in the $2\ell + 1\tau_{\text{had}}$ channel is discussed analogously to Section 9.2. Already in Chapter 8, a decision was made that $2\ell$ OS$+1\tau_{\text{had}}$ events would only serve in the background estimation given the channel's low signal significance. Hence, the emphasis of this section lies on the $2\ell$ SS $+ 1\tau_{\text{had}}$ channel.

Similar to the $1\ell + 2\tau_{\text{had}}$ channel, the results listed in Table 9.2 represent a good starting point to identify all dominant background processes. These results indicate that the situation in the $2\ell$ SS$+1\tau_{\text{had}}$ channel is more complex. Once again, the predominant contribution corresponds to fake-$\tau$ events. This time, however, the fake light leptons, i.e. fake-$e$ and fake-$\mu$ events, also need to be taken into account. A non-negligible number of events contain either a single fake light lepton or one in the presence of an additional fake-$\tau$ object. In contrast, events with two fake light leptons are heavily suppressed and can be ignored. The last relevant contribution, exceeding 1 % in the tight preselection region, is formed by electrons arising from the conversions of prompt photons. Conducting identical studies in the $2\ell$ OS $+ 1\tau_{\text{had}}$ channel demonstrates that fake-$\tau$ events represent the only significant contribution in in this analysis region.

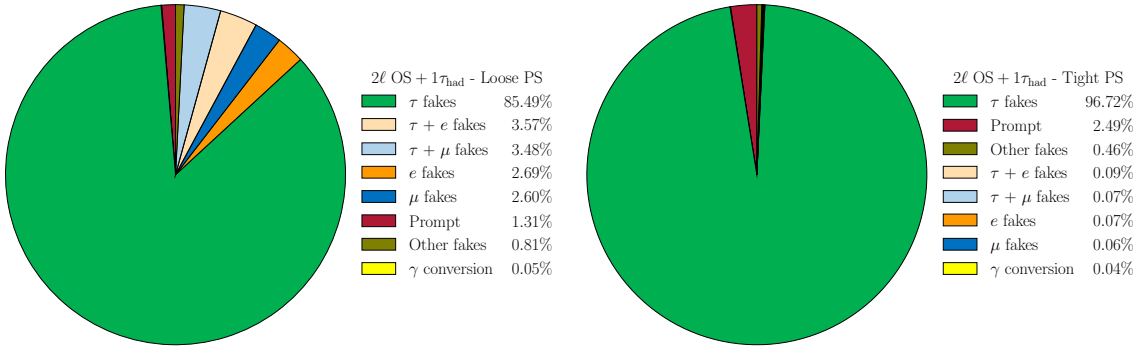
The remaining incorrectly modelled processes, as well as the residual small prompt backgrounds, are merged with the the "other fakes" category and the minor backgrounds according to the criteria listed in Section 9.1.2. The obtained compositions in the loose and tight preselection region are visualised in Figure 9.8 and Figure 9.9 for both channels. Based on these results, potential correlations between the strength of the data to simulation disagreement and certain event kinematics are investigated. These studies are conducted separately for fake-$\tau$ events and fake light lepton events, and are discussed in the following two sections.

In the case of the remaining improperly modelled process, i.e. electrons originating from prompt photon conversion, no individual study is carried out. This contribution is not corrected in a predetermined region as the available number of events is found to be too small. Instead, these events enter the final profile likelihood with a conservatively estimated uncertainty, motivated in Chapter 10.




(a) (b)

Figure 9.8: Composition in (a) the loose and (b) the tight preselection regions of the $2\ell$ SS + $1\tau_{\text{had}}$ channel. Signal and background events are both merged in the prompt contribution while small fake processes are grouped together as "other fakes".




(a) (b)

Figure 9.9: Composition in (a) the loose and (b) the tight preselection regions of the $2\ell\,\mathrm{OS} + 1\tau_{\mathrm{had}}$ channel. Signal and background events are both merged in the prompt contribution while small fake processes are grouped together as "other fakes".

### 9.3.1 Fake-$\tau$ - correlation studies

Particularly the results in the $2\ell\,\mathrm{OS} + 1\tau_{\mathrm{had}}$ regions (see Figure 9.9) suggest that the associated events are well suited for correcting the fake-$\tau$ background in the $2\ell\,\mathrm{SS} + 1\tau_{\mathrm{had}}$ regions.

Given that both channels share the same selection criteria apart from the charge requirement, it is initially assumed that there are no fundamental differences between the fake-$\tau$ events. As a consequence, all fake-$\tau$ related studies are conducted in the $2\ell\,\mathrm{OS} + 1\tau_{\mathrm{had}}$ channel. Similar to the $1\ell + 2\tau_{\mathrm{had}}$ channel, correlations between event kinematics and the agreement of data and simulation, are searched for. With this channel's increased number of events, a statistically more reliable result can be obtained.

Initially, studies in the $2\ell\,\mathrm{OS} + 1\tau_{\mathrm{had}}$ channel reveal that the data to simulation agreement is highly correlated to $Z+$jets events as shown in Figure 9.10. Regions with a high fraction of $Z+$jets events have a substantially overestimated yield compared with the inclusive $2\ell\,\mathrm{OS} + 1\tau_{\mathrm{had}}$ region. As a consequence, $Z+$jets events strongly influence any fake-$\tau$ correction, while also biasing a potential extrapolation to the $2\ell\,\mathrm{SS} + 1\tau_{\mathrm{had}}$ region. To mitigate this effect, the fraction of $Z+$jets events is reduced by introducing an additional selection. The corresponding criterion is obtained by analysing the invariant mass of both light leptons $M_{\ell\ell}$, as depicted in Figure 9.10. The majority of $Z+$jets events will generate the two light leptons through the leptonic $Z$-boson decay, causing $M_{\ell\ell}$ to peak around the $Z$-boson mass of $m_Z \approx 91\,\mathrm{GeV}$. Excluding the range $66\,\mathrm{GeV} < M_{\ell\ell} < 116\,\mathrm{GeV}$ proves to be successful at substantially lowering the amount of $Z+$jets events in the $2\ell\,\mathrm{OS} + 1\tau_{\mathrm{had}}$ channel. This selection is applied for all subsequent studies in the $2\ell\,\mathrm{OS} + 1\tau_{\mathrm{had}}$ channel.

In the remaining $2\ell\,\mathrm{OS} + 1\tau_{\mathrm{had}}$ region, a dependency with respect to the $\tau_{\mathrm{had}}$'s $p_{\mathrm{T}}$ and $\eta$ is investigated in analogy to the $1\ell + 2\tau_{\mathrm{had}}$ channel. This study's outcome is displayed in Figure 9.11 and leads to the same conclusion as in Section 9.2: with regard to $\eta$, no dependency was found, whereas the strength of the disagreement increases along with the transverse momentum of the $\tau_{\mathrm{had}}$.

Once again, potential differences between 1-prong and 3-prong $\tau_{\mathrm{had}}$ decays are probed. Neither the

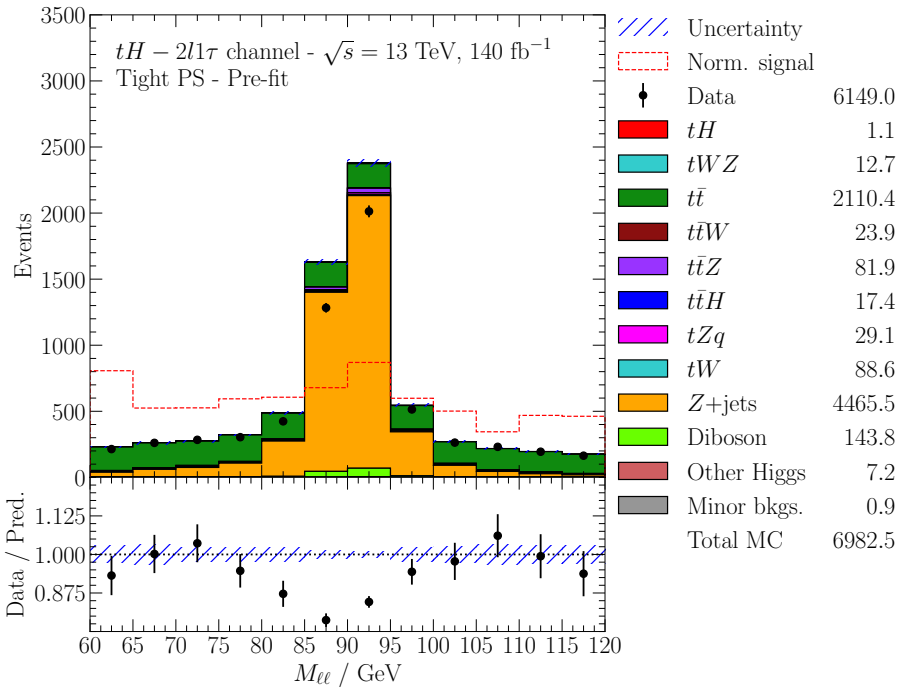
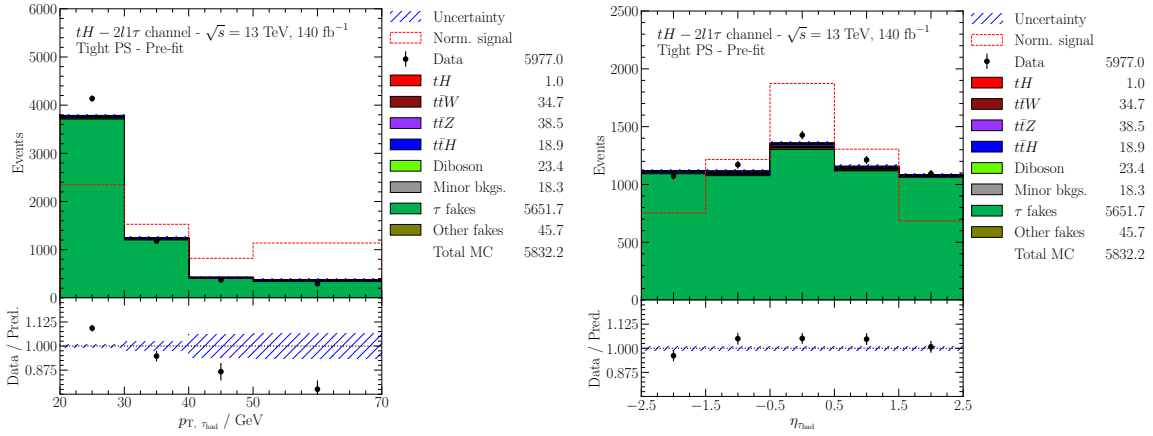



Figure 9.10: Invariant mass distribution of both light leptons $M_{\ell\ell}$ in the $2\ell$ OS + $1\tau_{\text{had}}$ channel. The data to simulation agreement decreases around the $Z$-boson peak.

corresponding distributions, visualised in Figure 9.12, nor preliminary profile likelihood fits show significant deviations in the required corrections for 1-prong and 3-prong decays. As a conclusion, the fake-$\tau$ transverse momentum is considered to be the sole dependency of the disagreement.

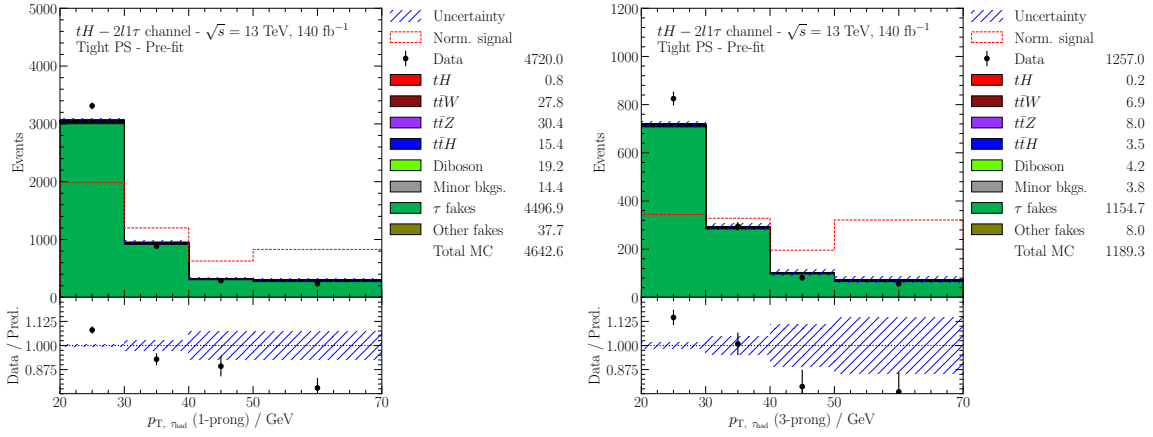
### 9.3.2 Fake light leptons - correlation studies

Subsequently, the modelling of the fake-$e$ and fake-$\mu$ events is likewise analysed for possible dependencies. Both the leading and subleading light leptons kinematics are examined in the $2\ell$ SS + $1\tau_{\text{had}}$ tight preselection region. The results shown in Figure 9.13 and Figure 9.14 give no reason to suspect a dependency with regard to $p_{\text{T}}$ and $\eta$. A second, more detailed study, shown in Figure 9.15 as well as Figure 9.16, which considers fake-$e$ and fake-$\mu$ events separately, also fails to show any consistent dependency. The agreement between these findings demonstrates that both fake light lepton contributions do not require any further splitting. Each contribution can be estimated and controlled in its own inclusive region.



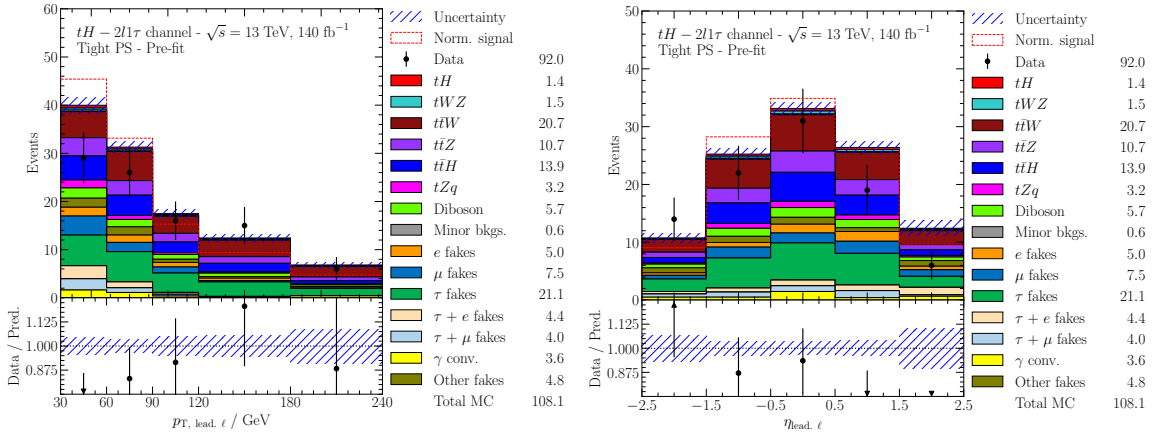

(a) (b)

Figure 9.11: Correlation studies showing (a) $p_{\mathrm{T}}$ and (b) $\eta$ of the $\tau_{\mathrm{had}}$ in the $2\ell$ OS + $1\tau_{\mathrm{had}}$ channel's tight preselection region. Only statistical uncertainties are displayed.



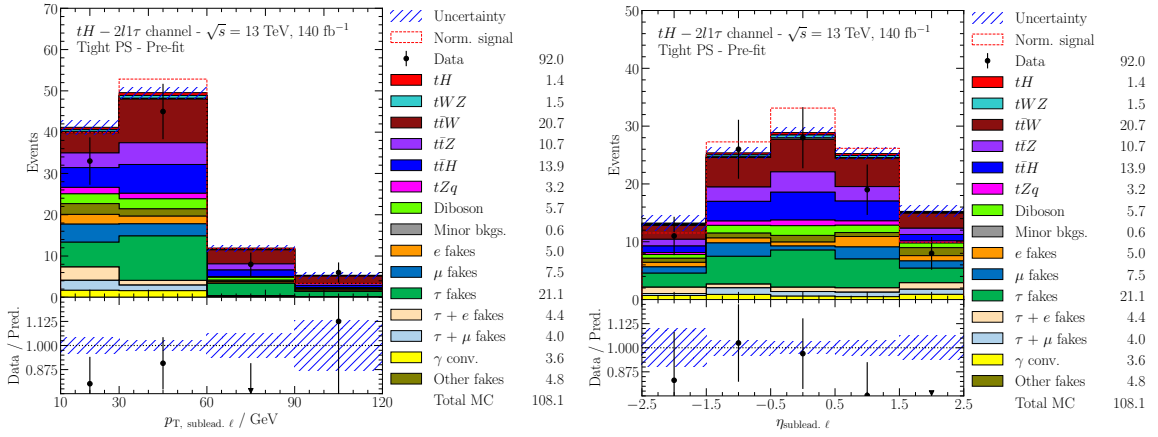

(a) (b)

Figure 9.12: Correlation studies showing the transverse momentum for (a) 1-prong (b) 3-prong $\tau_{\mathrm{had}}$ decays in the $2\ell$ OS + $1\tau_{\mathrm{had}}$ channel's tight preselection region. Only statistical uncertainties are displayed.



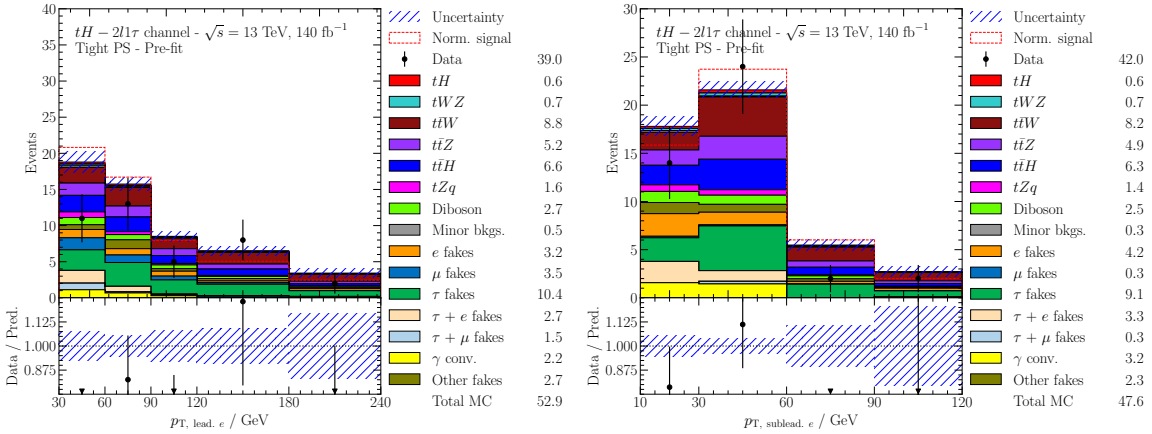

(a) (b)

Figure 9.13: Correlation studies showing (a) $p_{\text{T}}$ and (b) $\eta$ of the leading light lepton in the $2\ell$ SS + $1\tau_{\text{had}}$ channel's tight preselection region. Only statistical uncertainties are displayed.



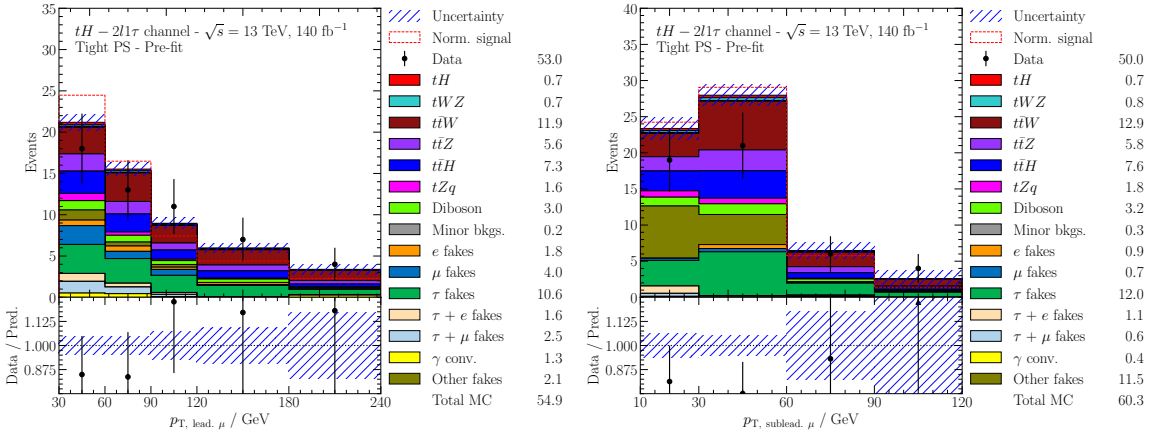

(a) (b)

Figure 9.14: Correlation studies showing (a) $p_{\text{T}}$ and (b) $\eta$ of the subleading light lepton in the $2\ell$ SS + $1\tau_{\text{had}}$ channel's tight preselection region. Only statistical uncertainties are displayed.




(a) (b)

Figure 9.15: Correlation studies showing the transverse momentum for events with (a) leading and (b) subleading electrons in the $2\ell$ SS + $1\tau_{\text{had}}$ channel's tight preselection region. Only statistical uncertainties are displayed.




(a) (b)

Figure 9.16: Correlation studies showing the transverse momentum for events with (a) leading and (b) subleading muons in the $2\ell$ SS + $1\tau_{\text{had}}$ channel's tight preselection region. Only statistical uncertainties are displayed.

### 9.3.3 Region definitions

The last section of this chapter covers the definition of all regions used in the fit, in analogy to Section 9.2.1. Again, the choice of individual regions is strongly guided by the results of the correlation studies summarised in Section 9.3.1 as well as Section 9.3.2.

On account of the small number of events in the $2\ell$ SS + $1\tau_{\text{had}}$ channel, it is decided to use the entire tight preselection as a SR. As a second step, a VR is defined to confirm the fake estimation's accuracy. To map a phase space similar to the SR, this VR is also formed by $2\ell$ SS + $1\tau_{\text{had}}$ events. Orthogonality is ensured by requiring an anti-medium $\tau_{\text{had}}$ while the light leptons are required to pass their looser working point.

This means, that the selection criteria for light leptons in the SR and the VR can be kept identical at the strictest level. Conversely, all areas of the loose preselection region in the $2\ell$ SS + $1\tau_{\text{had}}$ channel with a more relaxed light lepton requirement are still available. Given their increased fraction of fake light leptons, they represent a suitable choice for the formation of CRs. Generally, the physics processes that lead to fake-$e$ or fake-$\mu$ events are not expected to be identical. Therefore, an independent correction of both components is required, generating the demand of two distinct regions: one of which is supposed to be dominated by fake-$e$ events and the other by fake-$\mu$ events. This is ensured by applying the following selection:

- $\text{CR}_e$: the subleading light lepton has to be an electron. In addition, at least one of the two light leptons is required to pass its looser working point and fail the stricter one (termed anti-tight light lepton).

- $\text{CR}_\mu$: the subleading light lepton has to be a muon. In addition, at least one of the two light leptons is required to pass its looser working point and fail the stricter one (termed anti-tight light lepton).

The $\tau_{\text{had}}$ must pass the medium RNN working point in both CRs to minimise the contribution of fake-$\tau$ events. As a consequence, the only difference between $\text{CR}_e$ and $\text{CR}_\mu$ lies in the type of the subleading light lepton. Since this light lepton has a higher probability of being a fake-$e/\mu$ than the leading one, each obtained region is only dominated by exactly one source of fake light leptons. The composition of all regions, displayed in Figure 9.21, confirms this statement. As visible, fake-$\mu$ events represent a small fraction of $\text{CR}_e$ and vice versa.

As no significant dependence on $p_{\text{T}}$ and $\eta$ is found, an inclusive correction is obtained in the CRs, which is ultimately extrapolated to the SR. Therefore, potential kinematic differences between the fake light leptons of both regions can introduce a bias. To rule out such a bias, the kinematics of all three regions are compared. The corresponding results for $\text{CR}_e$ and $\text{CR}_\mu$ are displayed in Figures 9.17 to 9.20. Only events in which the leading or subleading light lepton corresponds to a fake light lepton are considered in those studies. The visible deviations can be attributed to the relatively high statistical uncertainties in the SR. Apart from those deviations, no fundamentally different trend in the kinematics' distributions is evident. In conclusion, it is assumed that a correction, obtained in the CRs, can be safely extrapolated into the SR.

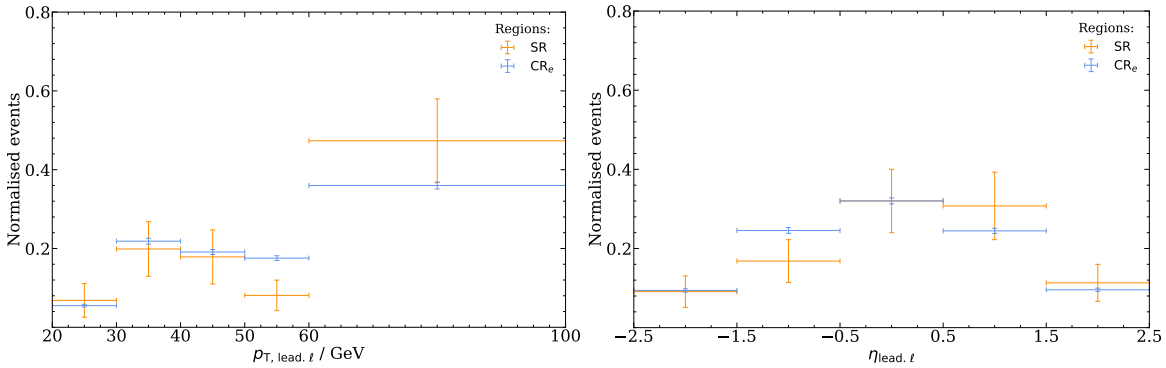
What remains to be discussed is a definition of all CRs used to correct the fake-$\tau$ contributions. Taking into account the results presented in Table 9.3, the $2\ell$ OS + $1\tau_{\text{had}}$ channel represents the obvious choice to define these regions in. This channel is completely dominated by fake-$\tau$ events while orthogonality to all other regions is already ensured by the charge requirement. In order to mirror

fake-$\tau$ events in the SR as closely as possible, the tight preselection region forms a prerequisite of all fake-$\tau$ CRs. The concrete definition of these regions incorporates the $p_{\mathrm{T}}$ dependency, found in Section 9.3.1. Analogous to the $1\ell + 2\tau_{\mathrm{had}}$ channel, three CRs are formed:

- $\mathrm{CR}_{\text{fake-}\tau,p_{\mathrm{T},1}}$: $20\,\mathrm{GeV} \leq$ anti-medium $\tau_{\mathrm{had}}$ $p_{\mathrm{T}} < 30\,\mathrm{GeV}$.
- $\mathrm{CR}_{\text{fake-}\tau,p_{\mathrm{T},2}}$: $30\,\mathrm{GeV} \leq$ anti-medium $\tau_{\mathrm{had}}$ $p_{\mathrm{T}} < 60\,\mathrm{GeV}$.
- $\mathrm{CR}_{\text{fake-}\tau,p_{\mathrm{T},3}}$: $60\,\mathrm{GeV} \leq$ anti-medium $\tau_{\mathrm{had}}$ $p_{\mathrm{T}}$.

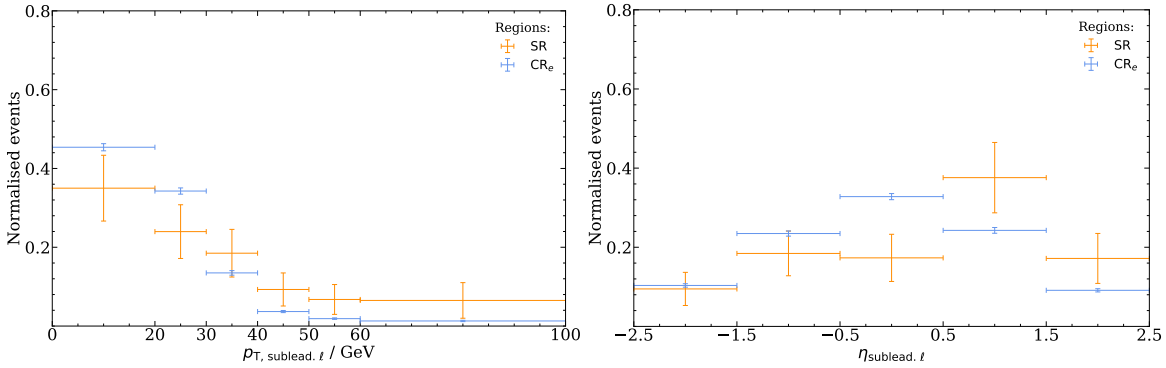
The relative region composition is visualised in Figure 9.22. It is evident that all regions are heavily dominated by fake-$\tau$ events.

A profile likelihood fit that uses the combination of all defined regions therefore should be capable of estimating the $tH$ signal strength under simultaneous correction of fake-$\tau$ and fake light lepton contributions.



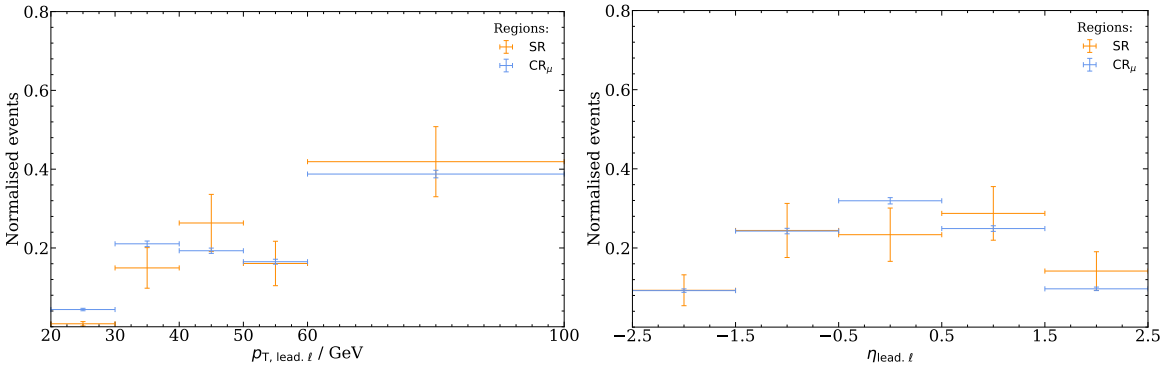

(a) (b)

Figure 9.17: Kinematic distributions of leading light leptons compared between the $\mathrm{CR}_e$ and the SR region in the $2\ell\,\mathrm{SS} + 1\tau_{\mathrm{had}}$ channel. Displayed are (a) $p_{\mathrm{T}}$ and (b) $\eta$ for all events with a fake leading light lepton.



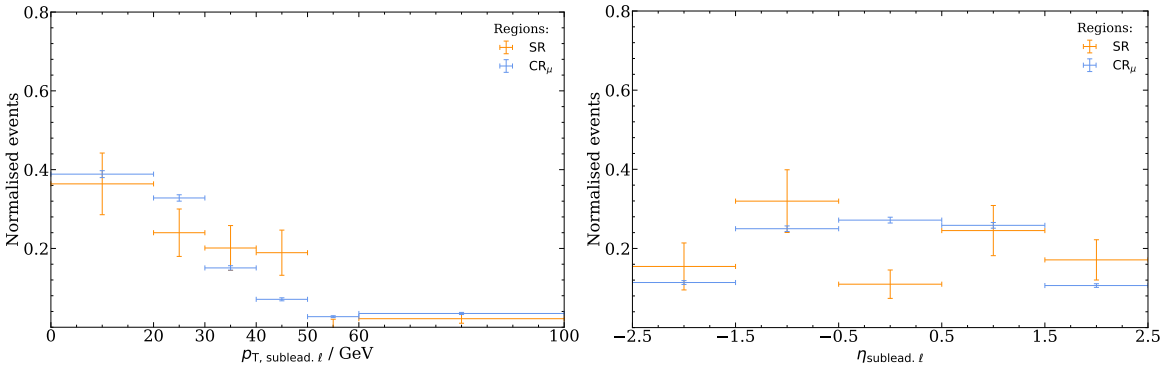

(a) (b)

Figure 9.18: Kinematic distributions of subleading light leptons compared between the $\mathrm{CR}_e$ and the SR region in the $2\ell\,\mathrm{SS} + 1\tau_{\mathrm{had}}$ channel. Displayed are (a) $p_{\mathrm{T}}$ and (b) $\eta$ for all events with a fake subleading light lepton.




(a) (b)

Figure 9.19: Kinematic distributions of leading light leptons compared between the CR$_\mu$ and the SR region in the $2\ell$ SS + $1\tau_{\text{had}}$ channel. Displayed are (a) $p_{\text{T}}$ and (b) $\eta$ for all events with a fake leading light lepton.




(a) (b)

Figure 9.20: Kinematic distributions of subleading light leptons compared between the CR$_\mu$ and the SR region in the $2\ell$ SS + $1\tau_{\text{had}}$ channel. Displayed are (a) $p_{\text{T}}$ and (b) $\eta$ for all events with a fake subleading light lepton.

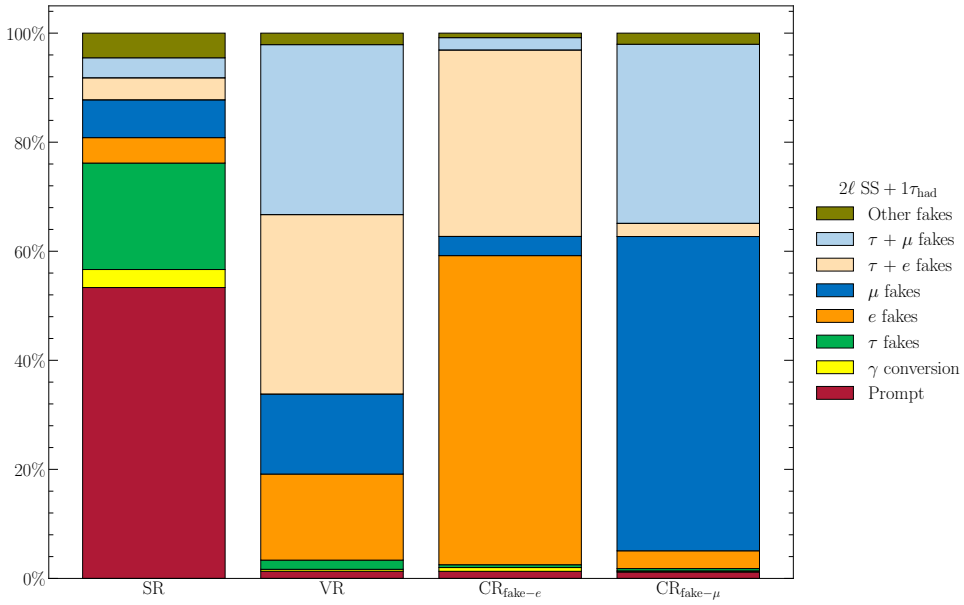



Figure 9.21: Composition of all $2\ell$ SS + $1\tau_{had}$ regions entering the profile likelihood fit.

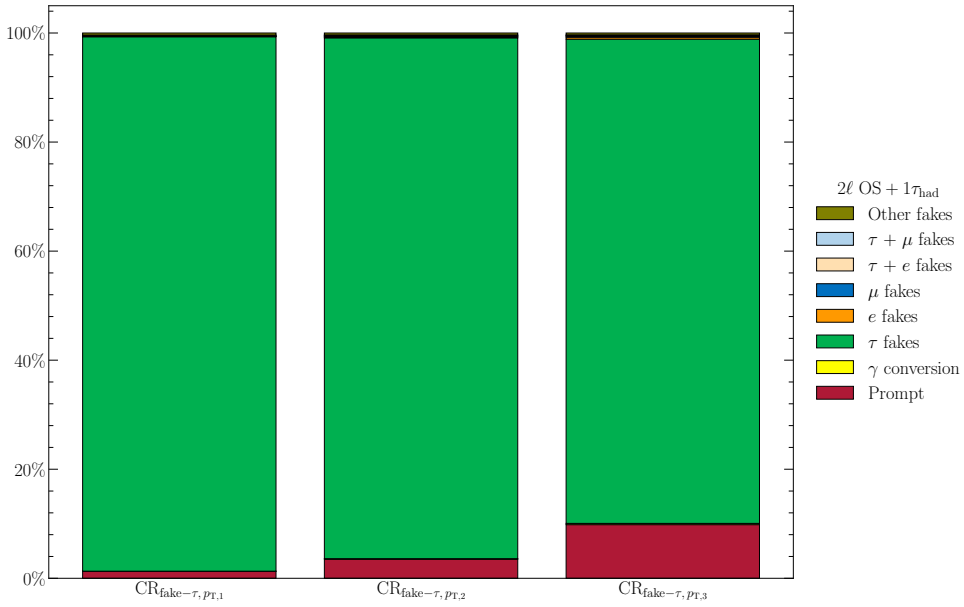



Figure 9.22: Composition of all $2\ell$ OS + $1\tau_{had}$ regions entering the profile likelihood fit.

CHAPTER 10

# Cross-section estimation

The goal of this analysis is to estimate the $tH$ cross-section. This is accomplished by performing a binned profile likelihood fit in the regions defined in Chapter 9. As an outcome, the fit provides a scaled $tH$ signal strength, defined in Section 4.2.1.

Setting up and optimising a binned profile likelihood fit is a complex process. Therefore, all required steps and validation methods are discussed in the beginning of this chapter. First of all, the general fit approach is covered in Section 10.1, before Section 10.2 presents fits performed on the Asimov dataset (see Section 4.4.1) in both channels. This section also aims to familiarise the reader with the common diagnostic tools used in the analysis of fit results. The following two sections, Section 10.3 and Section 10.4, focus on fit results obtained with real data.

All these fits are performed under the assumption of the Standard Model hypothesis, which implies a destructive interference in the $tH$ channel (see Section 2.3). In contrast, an additional fit is performed in Section 10.5 to investigate the inverse scenario, with an increased cross-section due to constructive interference. Finally, Section 10.6 discusses the combined fit results of the two channels for both hypotheses.

## 10.1 General strategy

Before discussing the fit results in detail, this section outlines the general approach. Similar to the fake background estimation in Chapter 9, both channels share the same fit strategy, which is adapted to take into account the individual characteristics. Initially, the binned profile likelihood function introduced in Section 4.2 needs slight modification to account for systematic uncertainties. Building on this, the iterative fit procedure for cross-section estimation is explained and motivated.

### 10.1.1 Uncertainty treatment

Performing a binned profile likelihood fit not only provides an estimate for the parameter of interest but also assesses its uncertainties. Statistical uncertainties are already accounted for through each bin's Poisson term, as given in Equation (4.15). However, to also incorporate systematic uncertainties,

Equation (4.15) has to be extended to

$$L(\vec{\mu}, \vec{\theta};\ \vec{n}, \vec{a}) = \prod_i \text{Poisson}\left(\nu_i(\vec{\mu}, \vec{\theta});\ n_i\right) \prod_j c_j\left(\theta_j;\ a_j\right) , \tag{10.1}$$

where the second term represents $j$ Gaussian constraint terms $c_j$, with mean $\theta_{j,0}$ and standard deviation $\sigma_j = \Delta\theta_j$. Each such term characterises a nuisance parameter, obtained in an auxiliary measurement[1] performed on dataset $a_j$. This includes all systematic uncertainties listed in Appendix D. Prominent examples are detector calibration measurements, background rate uncertainties, or theoretical uncertainties affecting the event simulation. In addition, the statistical uncertainties of the MC simulations in each bin are accounted for through an additional constraint term, referred to as a $\gamma$-factor.

To investigate the impact that a $\pm 1\sigma$ deviation of a variable would have on the nominal sample, two methods, introduced in Section 4.4, are available. This involves either reweighting the nominal MC events or generating new samples under modified simulation assumptions. Compared to the nominal MC, such variations can exhibit one or both of the following effects:

- Normalisation uncertainty: the overall normalisation of one or all processes is uniformly affected across all bins and regions, as depicted in Figure 10.1(a). For instance, varying the integrated luminosity $L_{\text{int}}$ within its uncertainties increases ($+1\sigma$) or decreases ($-1\sigma$) the expected number of events in all bins to the same extent.

- Shape uncertainty: the systematic variation alters the shape of a binned distribution as shown in Figure 10.1(b). Such an uncertainty could, for example, lead to a higher average transverse momentum of fake-$\tau$ events in the preselection region without affecting the overall event yield.

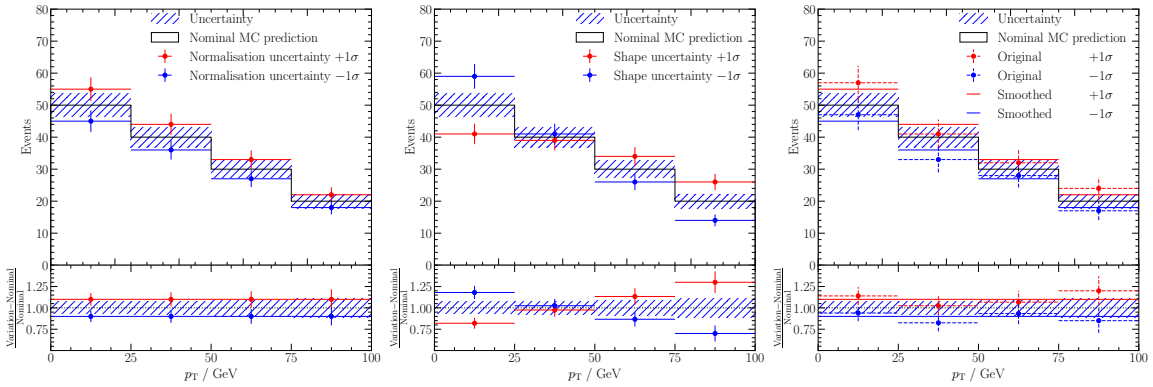
The up- and down-variations of all nuisance parameters are evaluated individually for each bin in each region. This procedure can introduce statistical fluctuations, particularly when alternative samples with limited statistics are used to model the variations. Such statistical fluctuations are problematic as they might bias the fit results. To mitigate this effect, a smoothing algorithm described in [111] is applied. This algorithm smooths the up- and down-variations across bins within a region, as displayed in Figure 10.1(c).

Not all of these nuisance parameters are expected to have a significant impact on the parameter estimation. Thus, following the smoothing procedure, negligible systematics are "pruned", i.e. excluded from the likelihood function, to prevent unnecessary complexity in the fit model. This pruning is performed individually per process and per region based on the following criteria:

- Pruning the normalisation component: if the systematic variation within the $\pm 1\sigma$ interval affects the global normalisation by less than 0.5 %, it will be removed.

- Pruning the shape component: when trying to prune the shape component, any normalisation effect of a nuisance parameter is eliminated first. This is achieved by rescaling both the nominal sample and the systematic variation to have the same total event yield per region. If the remaining deviation in any bin is found to be below 1 %, the shape component will be pruned.

[1] Most of these auxiliary measurements, particularly those addressing detector uncertainties, are carried out by different groups within the ATLAS collaboration.




(a) (b) (c)

Figure 10.1: Comparison of systematic variation showing (a) a pure normalisation effect, (b) a pure shape effect, as well as (c) a flat smoothing applied to a fluctuating variation.

All remaining nuisance parameters are kept and can be shifted when performing a fit. However, as already stated in Section 4.2, the Gaussian constraint terms penalise deviations of a nuisance parameter from its nominal value $\theta_{j,0}$, referred to as a "pull", during the likelihood minimisation. Additionally, the observed dataset may not allow as much variation in a nuisance parameter as initially assumed by its uncertainty. The nuisance parameter, or more precisely its uncertainty, is constrained. Since the initial uncertainties of all nuisance parameters are determined beforehand in dedicated measurements, constraints are generally considered unlikely. Any analysis that is not focussed on measuring a specific nuisance parameter should generally not yield a higher sensitivity, meaning it should not impose a constraint. Thus, constraints need to be carefully reviewed and explained in order to assess whether the initial uncertainties may have been over-conservative or whether the fit is affected by instabilities. Methods for validating the influence of certain nuisance parameters and to investigate pulls and constraints are introduced in Section 10.2 when discussing the first fit results.

### 10.1.2 Blinding procedure

As explained above, a nuisance parameter might get pulled when the fit scales MC events to correctly model the real data. However, fitting to real data represents the final step of the analysis. Prior to this, any fit setup needs to be evaluated in a blinded analysis. In this context, blinding refers to conducting the analysis without directly examining the real experimental data. This ensures objectivity and minimises the risk of introducing confirmation bias, i.e. the analysis is not unintentionally adjusted to affirm an expected result.

To enable evaluation of the fit model while maintaining blinding, the Asimov dataset, introduced in Section 4.4.1, is utilised, in which all processes correspond to their Standard Model expectations. Consequently, a measurement's expected sensitivity can be assessed without influencing or biasing the analysis. Any nuisance parameter that gets pulled in this scenario clearly indicates fundamental issues in the fit setup. Starting from this fully blinded fit, the analysis is progressively unblinded through the following iterations:

1. Asimov fit: this fit exploits the Asimov dataset in all regions. Such a fit provides the expected sensitivity and allows to rule out technical problems.

2. Hybrid fit: real data are used in all CRs under the assumption of a background-only hypothesis. All bins with an expected signal-to-background ratio below 0.3 % remain blinded. With this setup, potential nuisance parameter pulls and constraints can be studied in the real data. To still utilise the full constraining power and determine the expected sensitivity, the SR enters this fit on the basis of a modified pseudo dataset. This pseudo dataset corresponds to a modified Asimov dataset, obtained by applying the fitted nuisance parameters from all CRs.

3. Fit to data: this fit is completely unblinded and yields the final result. It is not performed until all other analysis steps have been fully validated. Once conducted, only well-justified changes should be implemented.

Over the course of this analysis, a variety of fit setups are tested before converging towards a final configuration. The different unblinding stages, presented in the following sections, all reflect this final configuration. As a result, the validation methods, as well as their findings are often highly similar and always grouped at the end of each section. Initially, these methods are thoroughly explained using the Asimov fit as a case study.

In the absence of significant deviations or conspicuous results, some standardised validation methods are omitted in subsequent sections. This is done to avoid artificially lengthening the sections at the cost of clarity. However, in order to still provide a comprehensive overview, all these results are summarised in Appendix H, to which the interested reader is referred at this point.

## 10.2 Asimov fit

This section covers the Asimov fit results, independently performed in both channels. In addition to validating the technical setup, this fit also allows for the extraction of the expected sensitivity. Each channel's discussion begins with a discussion of channel specific setup characteristics, followed by an introduction to the diagnostic tools and their respective results.

### 10.2.1 $1\ell + 2\tau_{\text{had}}$ - Asimov fit

The Asimov fit results for the $1\ell + 2\tau_{\text{had}}$ channel are presented in the following. The goal is to estimate not only the $tH$ signal strength, but also correct all dominant fake backgrounds in the fit[2]. To achieve this simultaneous estimation, the single parameter of interest $\mu$ is replaced by a vector $\vec{\mu}$ as indicated in Equation (10.1). Unlike the nuisance parameters with their Gaussian constraints, each vector component, known as a normalisation factor, can be adjusted in the fit without any restrictions. The selection of these components is guided by the $p_{\text{T}}$ dependence of fake-$\tau$ events, observed in Section 9.2. As a result, the vector $\vec{\mu}$ is formed by the following four normalisation factors:

- $\mu_{tH}$: this normalisation factor represents the $tH$ signal strength and scales all $tH$ events in the fit. It expresses the ratio of the observed value and the Standard Model's prediction and is the actual parameter of interest in this fit.

- $\mu(\text{fake-}\tau,\ p_{\text{T},1})$: this normalisation factor scales all events with a fake-$\tau$ object whose transverse momentum fulfils $20\,\text{GeV} < p_{\text{T}} < 30\,\text{GeV}$.

[2] This objective can only be achieved when fitting to real data. However, the Asimov fit already utilises the same setup to estimate the expected sensitivity.

- $\mu(\text{fake-}\tau,\ p_{\text{T},2})$: this normalisation factor scales all events with a fake-$\tau$ object whose transverse momentum fulfils $30\,\text{GeV} < p_{\text{T}} < 60\,\text{GeV}$.
- $\mu(\text{fake-}\tau,\ p_{\text{T},3})$: this normalisation factor scales all events with a fake-$\tau$ object whose transverse momentum fulfils $60\,\text{GeV} < p_{\text{T}}$.

Events containing fake-di-$\tau$ objects are consequently scaled using the product of two normalisation factors $\mu(\text{fake-}\tau,\ p_{\text{T},i})$ and $\mu(\text{fake-}\tau,\ p_{\text{T},j})$ for $i, j \in \{1, 2, 3\}$. Furthermore, $tH$ events can also contain fake-$\tau$ objects (see Appendix G). To ensure correct handling, these are scaled by $\mu_{tH} \cdot \mu(\text{fake-}\tau,\ p_{\text{T},i})$ for single fake-$\tau$ objects and by $\mu_{tH} \cdot \mu(\text{fake-}\tau,\ p_{\text{T},i}) \cdot \mu(\text{fake-}\tau,\ p_{\text{T},j})$ in case of fake-di-$\tau$ objects.

The fit utilises both the SR and the CRs, defined in Section 9.2.1. The normalisation factors are simultaneously fitted in all regions. Since the respective CRs are dominated by fake-$\tau$ events, they exert a large influence on the final determination of $\mu(\text{fake-}\tau,\ p_{\text{T},1,2,3})$. In contrast, $\mu_{tH}$ is primarily influenced by events in the SR, which contains the highest relative fraction of $tH$ events. Consequently, shifting $\mu_{tH}$ has a substantial impact in this region.

To achieve the highest possible sensitivity, $\text{NN}_{tH}$ (defined in Section 8.1.1) is chosen as the fit variable for the SR. The baseline binning is derived by scanning the complete distribution and forming a bin every time a specified fraction of signal and background events is reached. The associated threshold value $Z$ is defined through

$$Z = z_b \frac{n_b}{N_b} + z_s \frac{n_s}{N_s} \geq 1\,, \tag{10.2}$$

where $n_{s,b}$ defines the number of signal and background events in a bin while $N_{s,b}$ corresponds to the total signal and background event count in a region [157]. The tunable parameters $z_s$ and $z_b$ are both set to 2. This approach exploits a variable's discriminatory power, while simultaneously the creation of bins with high statistical uncertainties is avoided. Nonetheless, initial tests indicate that the most sensitive bin's statistical uncertainty highly impacts the fit. As a countermeasure, the two uppermost bins are merged, which stabilises the fit at the cost of slightly reducing the sensitivity.

In contrast, only a single bin is used in each CR. This choice is based on the fact that the relevant $p_{\text{T}}$ binning for fake-$\tau$ events is already incorporated into the definition of all three CRs. An additional division does not only lead to higher statistical uncertainty but additionally increases the risk of generating artificial pulls due to statistical fluctuations in resulting bins. By not requiring any binning in the CRs, the choice of the variable, in this case $\text{NN}_{\text{bkg.}}$, has no impact on the fit.

The pre- and post-fit distributions are displayed in Figure 10.3. Fitting to the Asimov dataset means that the yields in all regions are unchanged. This is also reflected in the resulting normalisation factors listed in the following:

$$\begin{aligned}
\mu_{tH} &= 1.0^{+5.6}_{-5.2}\,(\text{stat.})^{+4.1}_{-3.9}\,(\text{syst.}) &&= 1.0^{+7.0}_{-6.5}\,,\\
\mu(\text{fake-}\tau,\ p_{\text{T},1}) &= 1.00^{+0.04}_{-0.04}\,(\text{stat.})^{+0.03}_{-0.03}\,(\text{syst.}) &&= 1.00^{+0.05}_{-0.05}\,,\\
\mu(\text{fake-}\tau,\ p_{\text{T},2}) &= 1.00^{+0.05}_{-0.05}\,(\text{stat.})^{+0.05}_{-0.05}\,(\text{syst.}) &&= 1.00^{+0.07}_{-0.07}\,,\\
\mu(\text{fake-}\tau,\ p_{\text{T},3}) &= 1.00^{+0.13}_{-0.13}\,(\text{stat.})^{+0.25}_{-0.25}\,(\text{syst.}) &&= 1.00^{+0.28}_{-0.28}\,.
\end{aligned}$$

The expected value of one is obtained for all four normalisation factors. Generally, statistical uncertainties (stat.) slightly dominate, but are still of the same order of magnitude as the systematic uncertainties (syst.). The increase in uncertainty for higher $p_{\text{T}}$ normalisation factors can be explained

by the smaller event count in the corresponding CRs. Validating the fake background estimation, i.e. studying the post-fit data to MC agreement in the VR, is obviously meaningless in case of the Asimov fit. Nevertheless, all variables examined in this VR become relevant when presenting the unblinded fit results and are listed in the following:

- The leading and subleading $\tau_{\text{had}}$'s $p_{\text{T}}$, $\eta$ and RNN score.
- The light lepton's $p_{\text{T}}$ and $\eta$.
- The leading jet's and $b$-jet's $p_{\text{T}}$ and $\eta$.
- The two non-$tH$ CNN output distributions.
- The total transverse energy as well as the mass of the Lorentz vector sum of the leading $b$-jet and the forward jet.

The next step of this Asimov fit involves further validation through different diagnostic tools, starting with an investigation of potential pulls and constraints. For each nuisance parameter, a study as illustrated in Figure 10.2 is performed. This study reveals deviation between the post-fit value of a nuisance parameter $\hat{\theta}$ and its initially expected value $\theta_0$. To facilitate comparability, this deviation is normalised to be expressed in units of a nuisance parameter's standard deviation $\Delta\theta$. Both the $\pm1\sigma$ (green) and $\pm2\sigma$ (yellow) intervals are indicated. Additionally, the normalised post-fit uncertainty is shown. Whenever this uncertainty is smaller than the $\pm1\sigma$ interval, the corresponding nuisance parameter is considered constrained. The results of this test for all nuisance parameters in the $1\ell+2\tau_{\text{had}}$ channel are presented in Figure H.5. No concerning constraints are found, indicating a stable fit setup.

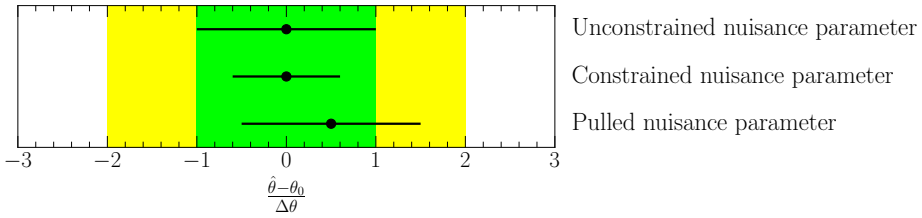



Figure 10.2: Exemplary investigation of three nuisance parameter's post-fit pulls and constraints. The plot includes an unconstrained nuisance parameter that is not pulled, a constrained parameter and a pulled parameter. The $x$-axis is given in units of a nuisance parameter's standard deviation, with the $\pm1\sigma$ ($\pm2\sigma$) interval marked in green (yellow).

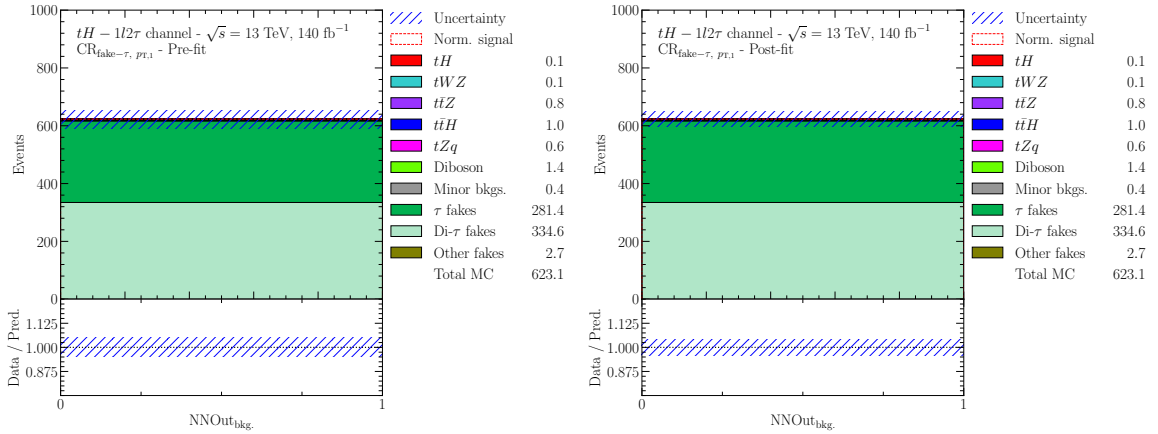
Besides studying pulls and constraints, it is essential to analyse the correlations between individual nuisance parameters. In general, all nuisance parameters, as well as normalisation factors, should be uncorrelated. Strong correlations between parameters need to be properly understood to prevent uncertainties in the fit from being incorrectly estimated. The correlation matrix associated to the $1\ell+2\tau_{\text{had}}$ Asimov fit is shown in Figure 10.4. Only nuisance parameters and normalisation factors with correlations exceeding 30 % are displayed. Notably, negative correlations between the normalisation factors $\mu(\text{fake-}\tau,\ p_{\text{T},1})$ and $\mu(\text{fake-}\tau,\ p_{\text{T},2})$ are observed. However, this correlation can be explained by considering fake-di-$\tau$ events. These cause both normalisation factors to scale an identical event, which can introduce correlations. Furthermore, both factors are positively correlated with additional

nuisance parameters such as the parton showering uncertainty of $t\bar{t}$ events ($t\bar{t}$ PS). This common dependency can likewise lead to increased correlations.

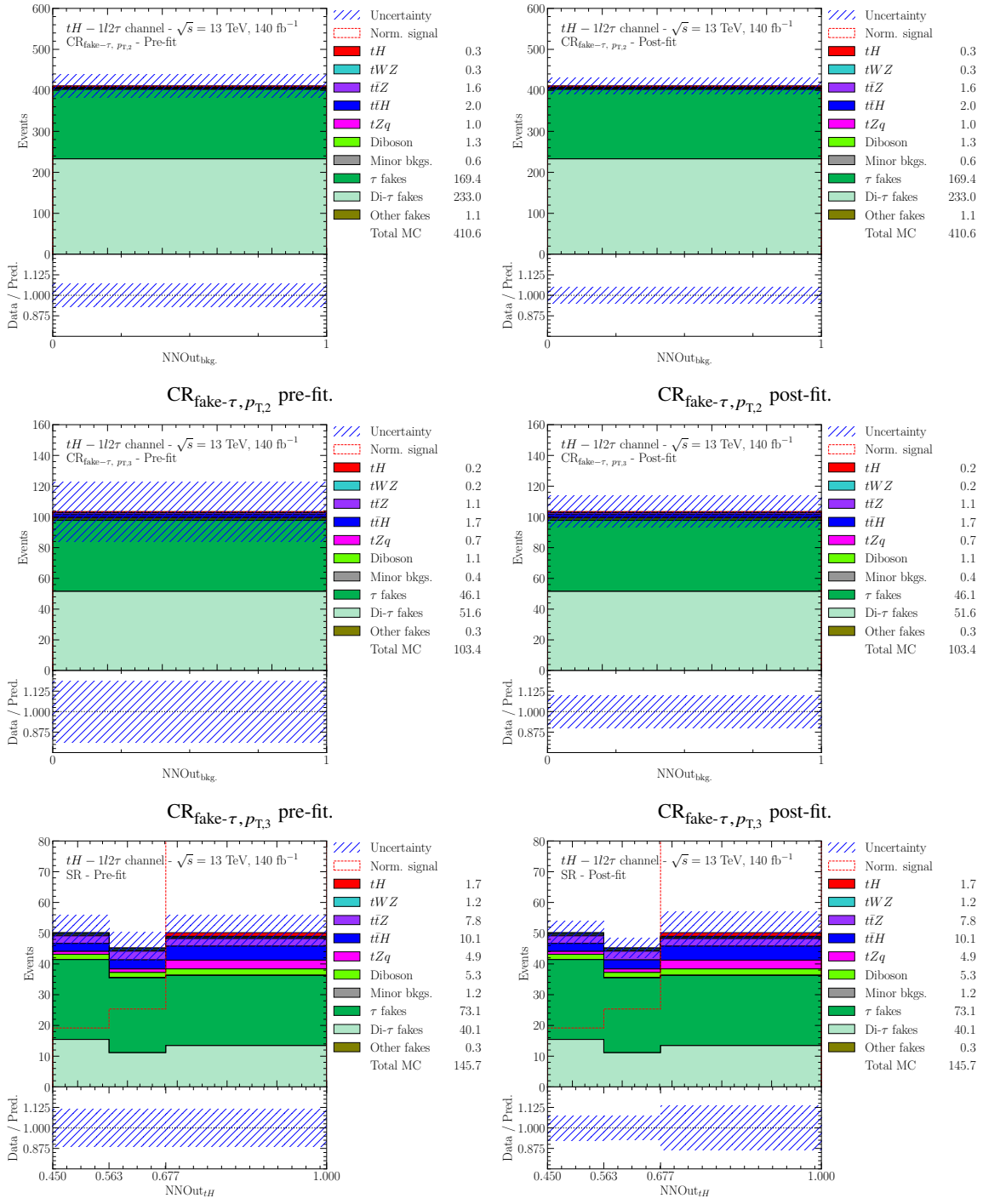
Finally, the nuisance parameters with the greatest influence on the parameter of interest, $\mu_{tH}$, are identified. The twenty most impactful parameters in the $1\ell + 2\tau_{\text{had}}$ channel are ranked in Figure 10.5. Following the definition in Figure 10.4, the pull of a nuisance parameter is indicated by $(\hat{\theta}-\theta_0)/\Delta\theta$. Additionally, the influence on the signal strength $\Delta\mu_{tH}$ can be observed for four different fit scenarios: the two pre-fit variations $\hat{\theta} \pm \Delta\theta$ and the two the post-fit variations $\hat{\theta} \pm \Delta\hat{\theta}$. Since nuisance parameters can be constrained, the post-fit impact may be reduced. Each of the four studies is performed by fixing just a single nuisance parameter to its lower or upper variation and rerunning the fit.

No unexpected results in the form of significant constraints are observed. In general, $\mu_{tH}$ is most strongly influenced by the fake-$\tau$ estimation, which is expected as the SR is primarily populated by fake-$\tau$ events (see Figure 9.7). These factors are followed in the ranking by the statistical uncertainty of the SR's highest bin. In principle, this impact could be reduced by merging bins. However, as this reduces the expected sensitivity, the binning of the SR is retained.

In conclusion, this fit yields the expected results with no indications of technical setup issues. The next iteration, a partially unblinded fit, is discussed in Section 10.3.




$\text{CR}_{\text{fake-}\tau,p_{\text{T},1}}$ pre-fit. $\text{CR}_{\text{fake-}\tau,p_{\text{T},1}}$ post-fit.




$CR_{fake\text{-}\tau, p_{T,2}}$ pre-fit. $CR_{fake\text{-}\tau, p_{T,2}}$ post-fit.

$CR_{fake\text{-}\tau, p_{T,3}}$ pre-fit. $CR_{fake\text{-}\tau, p_{T,3}}$ post-fit.

SR pre-fit. SR post-fit.

Figure 10.3: Pre- and post-fit distributions of the SR and all CRs used in the $1\ell + 2\tau_{had}$ channel's Asimov fit.

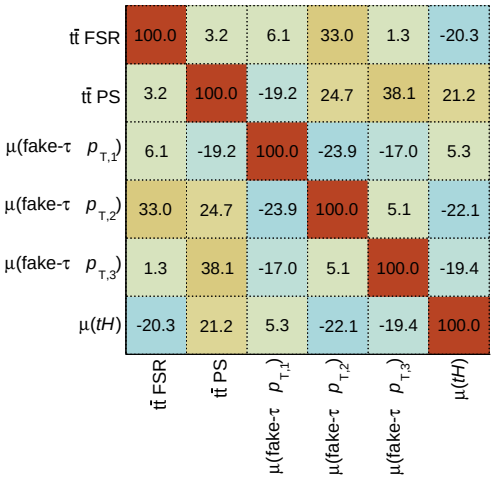



Figure 10.4: Correlation matrix for the $1\ell + 2\tau_{\text{had}}$ channel's Asimov fit.

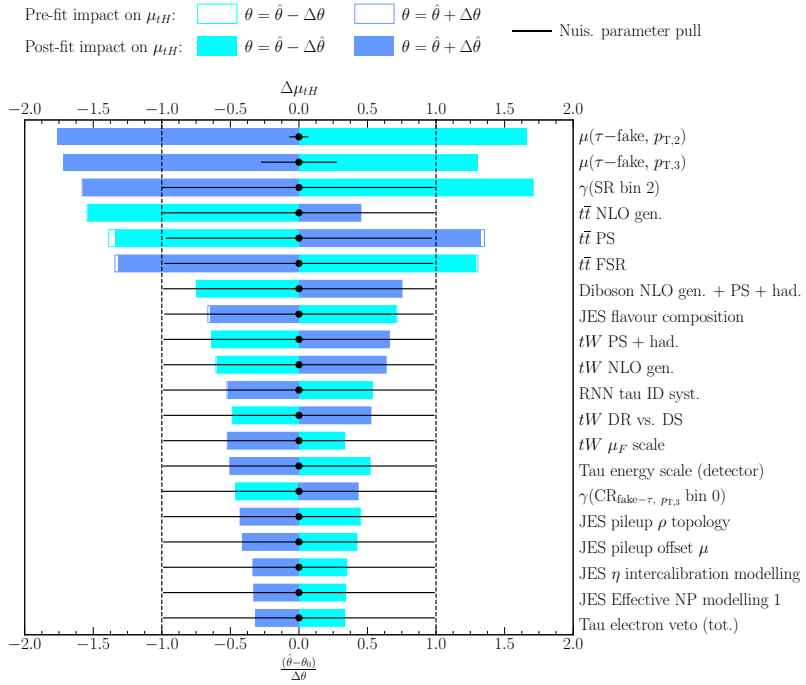



Figure 10.5: Nuisance parameter impact ranking for the $1\ell + 2\tau_{\text{had}}$ channel's Asimov fit.

### 10.2.2 $2\ell + 1\tau_{\text{had}}$ - Asimov fit

This section presents the $2\ell + 1\tau_{\text{had}}$ channel's Asimov fit, using the same approach as in the $1\ell + 2\tau_{\text{had}}$ channel. As a starting point, all normalisation factors are introduced:

- $\mu_{tH}$: this normalisation factor represents the $tH$ signal strength and scales all $tH$ events in the fit. It expresses the ratio of the observed value and the Standard Model's prediction and is the actual parameter of interest in this fit.

- $\mu(\text{fake-}\tau,\ p_{\text{T},1})$: this normalisation factor scales all events with a fake-$\tau$ object whose transverse momentum fulfils $20\,\text{GeV} < p_{\text{T}} < 30\,\text{GeV}$.

- $\mu(\text{fake-}\tau,\ p_{\text{T},2})$: this normalisation factor scales all events with a fake-$\tau$ object whose transverse momentum fulfils $30\,\text{GeV} < p_{\text{T}} < 60\,\text{GeV}$.

- $\mu(\text{fake-}\tau,\ p_{\text{T},3})$: this normalisation factor scales all events with a fake-$\tau$ object whose transverse momentum fulfils $60\,\text{GeV} < p_{\text{T}}$.

- $\mu(\text{fake-}e)$: this normalisation factor scales all events with a fake-$e$ object.

- $\mu(\text{fake-}\mu)$: this normalisation factor scales all events with a fake-$\mu$ object.

In this case, as in the $1\ell+2\tau_{\text{had}}$ channel, events with more than one fake object are scaled by the product of the corresponding normalisation factors. The same applies to $tH$ events containing fake objects. As mentioned in Section 9.3, an additional nuisance parameter is added to account for poorly modelled electrons from prompt photon conversions. Given this contribution's low event yield, estimation in a dedicated CR is not feasible. Instead, such events are corrected via a conservative normalisation uncertainty of 30 %. This value is derived by comparing results of the multi-lepton analysis channels referenced in Section 2.3. In these $tH$ decay channels, electrons from photon conversions play a central role. Therefore, such events are corrected through an independent normalisation factor, which, depending on the studied channel, leads to a correction of +20 % or −16 %, respectively [49].

All CRs and the SR established in Section 9.3.3 enter the Asimov fit. In case of the SR, the fit is performed on the $\text{NN}_{tH}$ distribution, which is again binned according to Equation (10.2). The fake-$\tau$ CRs as well as the two light lepton CRs are fit inclusively without any additional binning. The reasoning for the fake-$\tau$ CRs is identical to the $1\ell + 2\tau_{\text{had}}$ channel; the $p_{\text{T}}$ dependence is already incorporated into the region definition. In the case of the fake light lepton CRs, no significant shape dependence is expected according to the studies in Section 9.3.2. Therefore, forming multiple bins would simply increase the risk of introducing artificial pulls. As a fit variable, $\text{NN}_{\text{bkg.}}$ is used for the three fake-$\tau$ CRs, while the leading light lepton's $p_{\text{T}}$ is used for $\text{CR}_e$ and $\text{CR}_\mu$. Due to the single bin setup, this choice does not affect the fit. The pre-fit and post-fit distributions of these regions are displayed in Figure 10.6. As a consequence of using the Asimov dataset, all pre- and post-fit yields

remain identical. The expected normalisation factor uncertainties amount to

$$\begin{aligned}
\mu_{tH} &= 1.0^{+5.4}_{-4.8}\,(\text{stat.})^{+3.8}_{-3.5}\,(\text{syst.}) &&= 1.0^{+6.6}_{-6.0}\,,\\
\mu(\text{fake-}\tau,\ p_{\text{T},1}) &= 1.00^{+0.02}_{-0.02}\,(\text{stat.})^{+0.05}_{-0.05}\,(\text{syst.}) &&= 1.00^{+0.05}_{-0.05}\,,\\
\mu(\text{fake-}\tau,\ p_{\text{T},2}) &= 1.00^{+0.02}_{-0.02}\,(\text{stat.})^{+0.05}_{-0.05}\,(\text{syst.}) &&= 1.00^{+0.05}_{-0.05}\,,\\
\mu(\text{fake-}\tau,\ p_{\text{T},3}) &= 1.00^{+0.08}_{-0.08}\,(\text{stat.})^{+0.08}_{-0.08}\,(\text{syst.}) &&= 1.00^{+0.11}_{-0.11}\,,\\
\mu(\text{fake-}e) &= 1.00^{+0.03}_{-0.03}\,(\text{stat.})^{+0.06}_{-0.06}\,(\text{syst.}) &&= 1.00^{+0.07}_{-0.07}\,,\\
\mu(\text{fake-}\mu) &= 1.00^{+0.03}_{-0.03}\,(\text{stat.})^{+0.05}_{-0.05}\,(\text{syst.}) &&= 1.0^{+0.06}_{-0.06}\,.
\end{aligned}$$

Compared to the $1\ell + 2\tau_{\text{had}}$ channel, the $2\ell + 1\tau_{\text{had}}$ fit results are slightly more sensitive but likewise more affected by statistical uncertainties.

The remaining VR is used to investigate the agreement between data and MC simulation for the following variables:

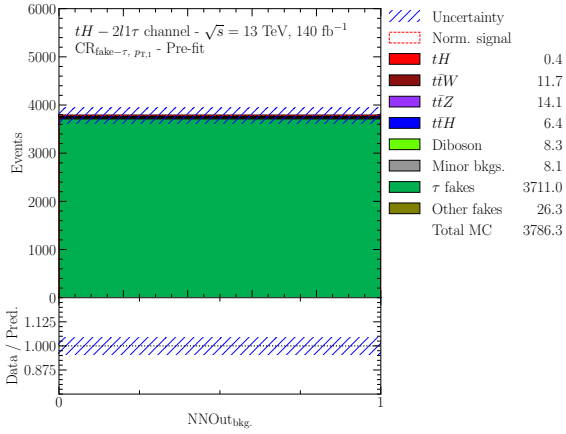
- The $p_{\text{T}}$ and $\eta$ of the light lepton associated to the top-quark ($\ell_{\text{top}}$) and to the Higgs-boson ($\ell_{\text{Higgs}}$).
- The $\tau_{\text{had}}$'s $p_{\text{T}}$, $\eta$ and RNN score.
- The leading jet's and $b$-jet's $p_{\text{T}}$ and $\eta$.
- The two non-$tH$ CNN output distributions.
- The total transverse energy as well as the mass of the Lorentz vector sum of the leading $b$-jet and the forward jet.

All VR studies become relevant only in the case of partially unblinded fits and are therefore omitted for the Asimov fit.

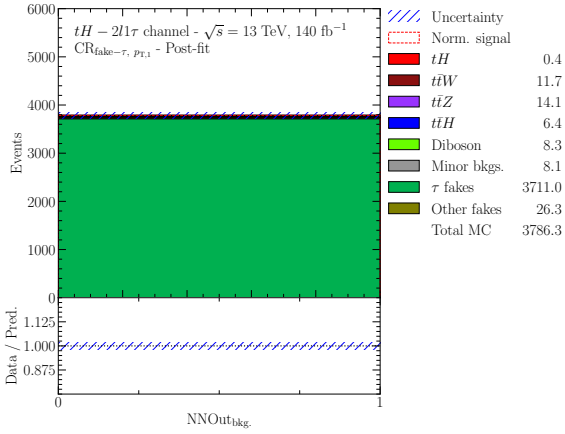
Subsequently, the fit is validated by first reviewing potential nuisance parameter pulls and constraints. The results are summarised in Figure H.9 and do not indicate instabilities in the setup. The associated correlations matrix is shown in Figure 10.7. The highest correlation is observed between the $\mu(\text{fake-}\tau,\ p_{\text{T},1})$ and $\mu(\text{fake-}\tau,\ p_{\text{T},2})$. This correlation arises from both normalisation factors' nearly identical anti-correlation with respect to the electron identification efficiency (Electron ID eff.). The latter also shows the largest anti-correlation $\mu(\text{fake-}e)$. Since both parameters affect the fake-$e$ normalisation, this dependency is to be expected. The same argument applies to $\mu(\text{fake-}\tau,\ p_{\text{T},3})$ and the $t\bar{t}$ parton showering uncertainty.

The last validation concerns the impact of all nuisance parameters, ranked in Figure 10.8. As clearly visible, various parton showering uncertainties have the highest impact on $\mu_{tH}$. The impact of the most sensitive SR bin's statistical uncertainty, ranked fourth highest, is expected given the low event count in the $2\ell\,\text{SS} + 1\tau_{\text{had}}$ region. In comparison to the $1\ell + 2\tau_{\text{had}}$ fit, fake background normalisation factors are less influential.

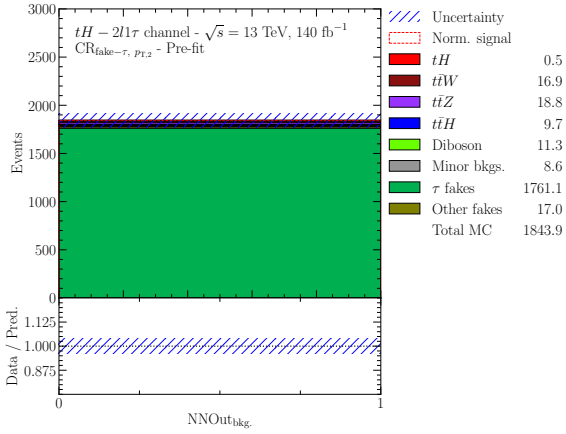
The partially unblinded hybrid fit, conducted with the same setup, is presented in Section 10.3.2.



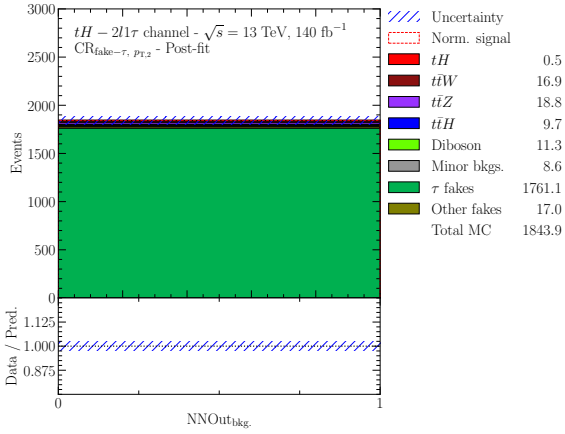

CR$_{\text{fake-}\tau, p_{\text{T},1}}$ pre-fit.



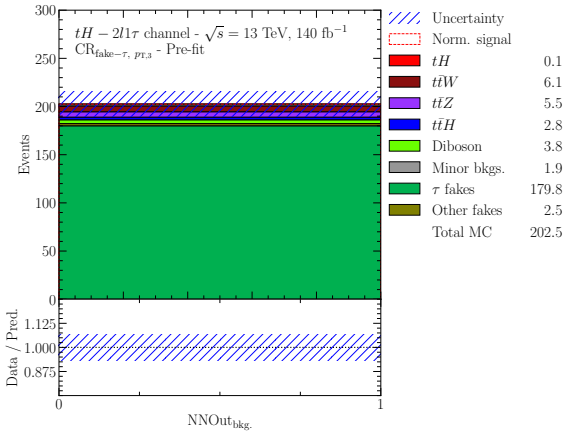

CR$_{\text{fake-}\tau, p_{\text{T},1}}$ post-fit.



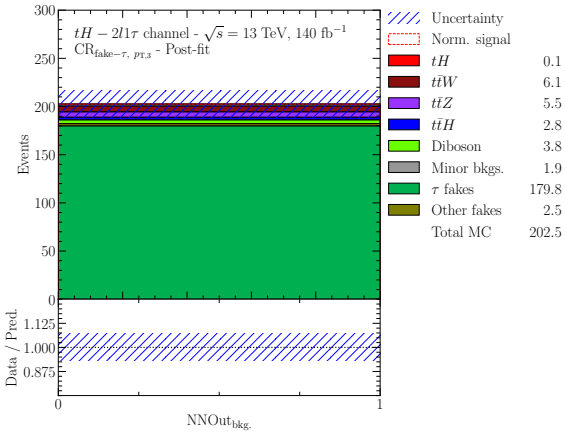

CR$_{\text{fake-}\tau, p_{\text{T},2}}$ pre-fit.



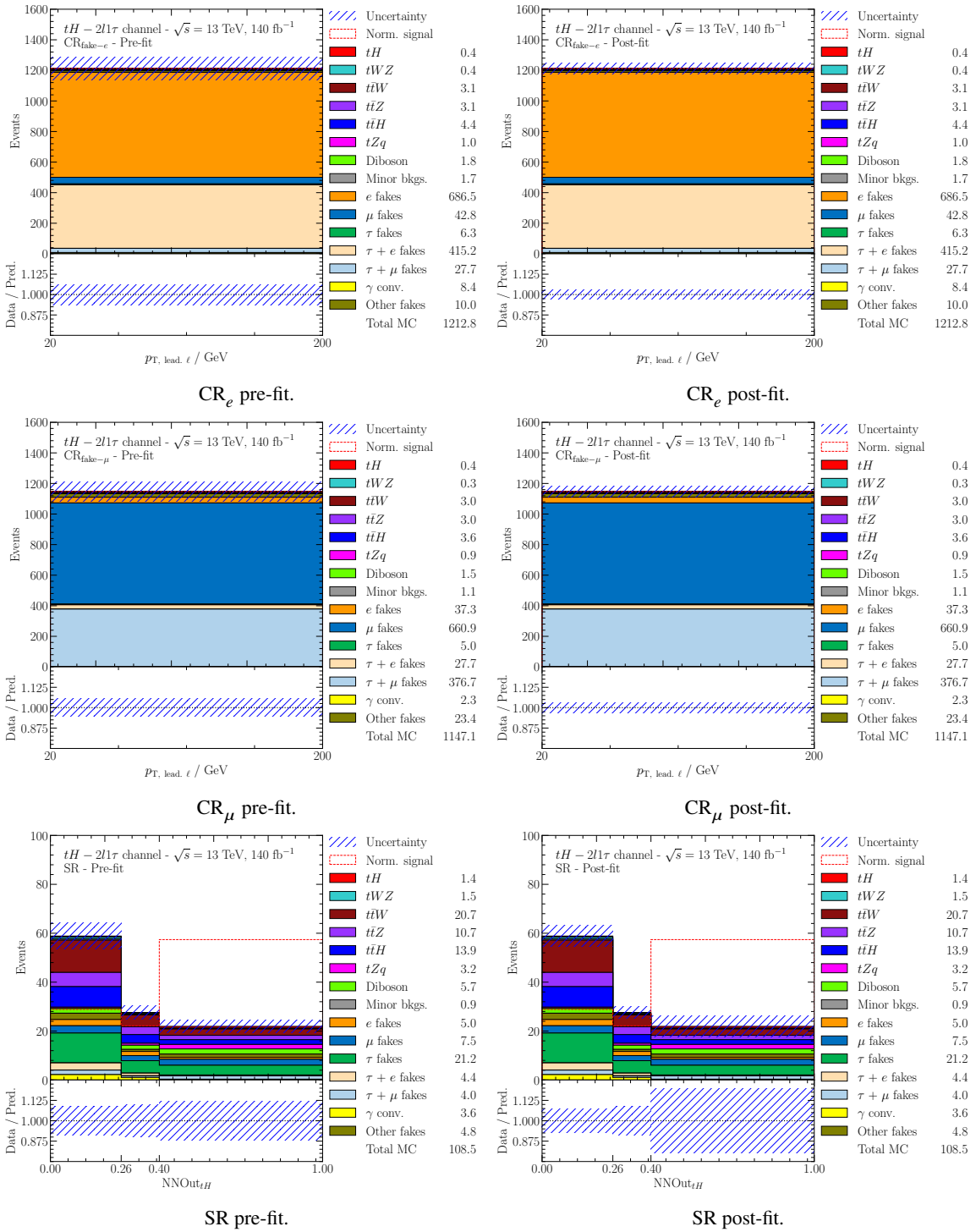

CR$_{\text{fake-}\tau, p_{\text{T},2}}$ post-fit.




CR$_{\text{fake-}\tau, p_{\text{T},3}}$ pre-fit.




CR$_{\text{fake-}\tau, p_{\text{T},3}}$ post-fit.




CR$_e$ pre-fit. CR$_e$ post-fit.

CR$_\mu$ pre-fit. CR$_\mu$ post-fit.

SR pre-fit. SR post-fit.

Figure 10.6: Pre- and post-fit distributions of the SR and all CRs used in the $2\ell + 1\tau_{\text{had}}$ channel's Asimov fit.

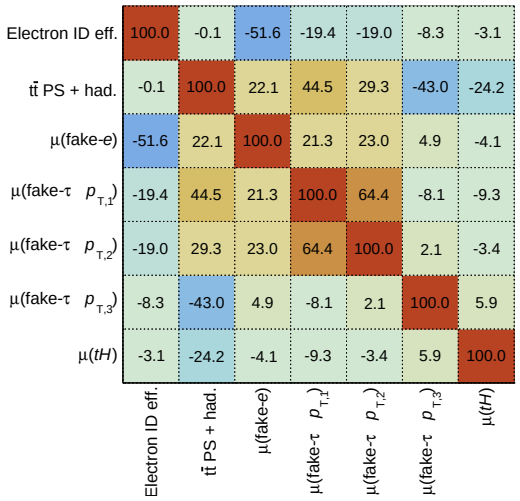



Figure 10.7: Correlation matrix for the $2\ell + 1\tau_{\text{had}}$ channel's Asimov fit.

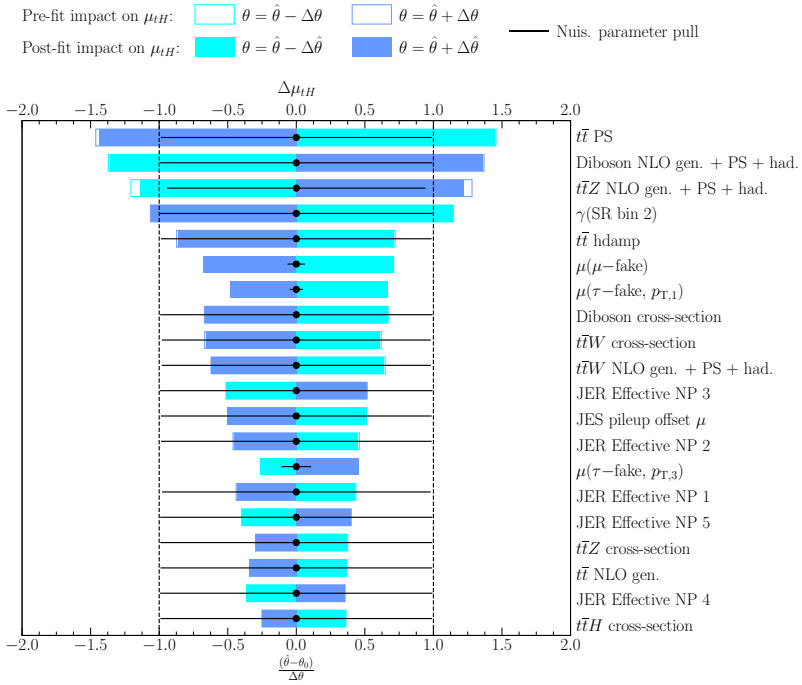



Figure 10.8: Nuisance parameter impact ranking for the $2\ell + 1\tau_{\text{had}}$ channel's Asimov fit.

## 10.3 Hybrid fit

Subsequent to the Asimov fit, the hybrid fit approach is presented in this section. All CRs bins are fitted to real data, provided they fall below the defined blinding threshold of $S/B < 0.3\,\%$. The primary objective of this fit is to verify that all background processes are well understood. A good post-fit agreement between data and the MC simulation indicates that fake-$\tau$ and fake light lepton contributions are scaled correctly. As before, the two analysis channels are discussed in dedicated subsections, beginning with the $1\ell + 2\tau_{\text{had}}$ channel.

### 10.3.1 $1\ell + 2\tau_{\text{had}}$ - hybrid fit

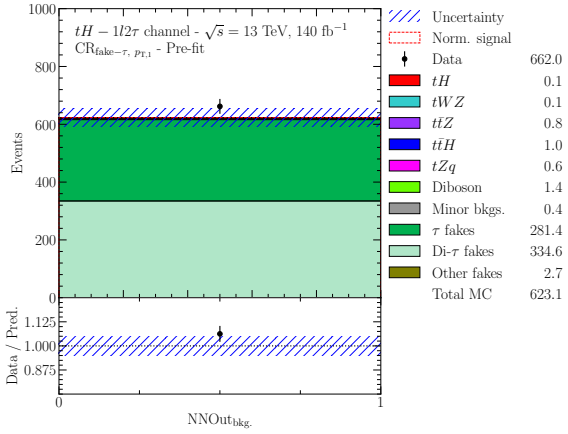
The fit setup as described in Section 10.2.1 remains unchanged, with the distinction that all three CRs lie below the blinding threshold, allowing real data to be used in the fit. As a result, the fit yields:

$$\begin{aligned}
\mu_{tH} &= 1.0^{+5.4}_{-5.0}\,(\text{stat.})^{+3.9}_{-3.6}\,(\text{syst.}) &&= 1.0^{+6.7}_{-6.1}\,,\\
\mu(\text{fake-}\tau,\ p_{\text{T},1}) &= 1.08^{+0.04}_{-0.04}\,(\text{stat.})^{+0.03}_{-0.03}\,(\text{syst.}) &&= 1.08^{+0.05}_{-0.05}\,,\\
\mu(\text{fake-}\tau,\ p_{\text{T},2}) &= 0.93^{+0.05}_{-0.05}\,(\text{stat.})^{+0.05}_{-0.05}\,(\text{syst.}) &&= 0.93^{+0.07}_{-0.07}\,,\\
\mu(\text{fake-}\tau,\ p_{\text{T},3}) &= 0.70^{+0.12}_{-0.12}\,(\text{stat.})^{+0.16}_{-0.16}\,(\text{syst.}) &&= 0.70^{+0.20}_{-0.20}\,.
\end{aligned}$$

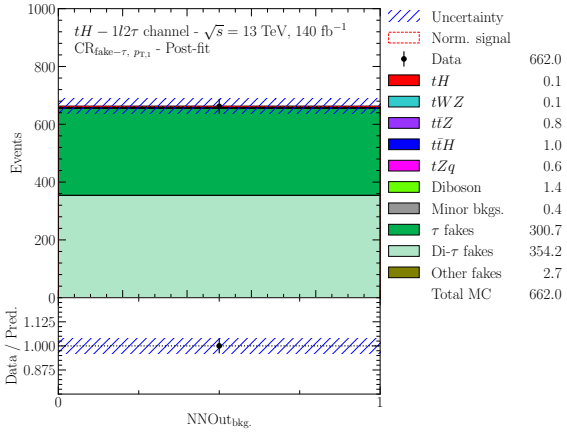
The listed normalisation factors precisely reflect the $p_{\text{T}}$ dependence of fake-$\tau$ events. At high $p_{\text{T}}$, the MC simulation tends to increasingly overestimate the data, correspondingly requiring a larger downscaling. In contrast, the low $p_{\text{T}}$ fake-$\tau$ events are upscaled by 8 %, which is consistent with the findings discussed in Section 9.3.1. The fit results obtained in the CRs are used to adjust the blinded SR's Asimov dataset, creating a pseudo dataset. The pre-fit and post-fit distributions for the SR and all three CRs are displayed in Figure 10.9 and show the expected behaviour. Given the single bin setup, the normalisation factor scaling is able to achieve a perfect post-fit agreement. The visible change in the SR's total event yield arises from the generated pseudo dataset.

As the fit is now performed on real data, the VR can additionally be used to assess the effect of the fake background estimation for all variables listed in Section 10.2.1. Each individual bin still has to fulfil the blinding requirement. As an example, the subleading $\tau_{\text{had}}$'s $p_{\text{T}}$ is shown in Figure 10.10. All investigated distributions can be found in Figure H.13. Compared to the CRs, all VRs show a better pre-fit agreement, as data and simulation agree within their $1\sigma$ uncertainties for most variables. This improved agreement can be attributed to the stricter selection criteria applied in the VR, which effectively suppresses fake background contributions (see Section 9.2.1). As this level of agreement is retained post-fit, it is assumed that the fit successfully corrected fake-$\tau$ contributions. Therefore, derived normalisation factors can be safely applied in the SR.

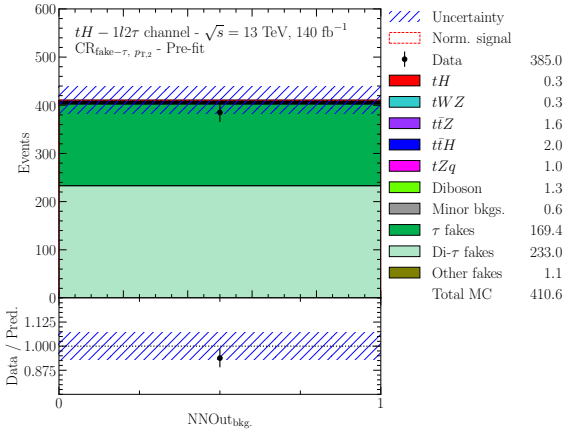
In analogy to the Asimov fit study, the pulls and constraints, correlations, and the impact on $\mu_{tH}$ are analysed for all nuisance parameters and displayed in Figures H.1(a), H.3(a) and H.6. As no inconsistencies are observed, the final unblinded fit is conducted in Section 10.4.1, following the same fit strategy.



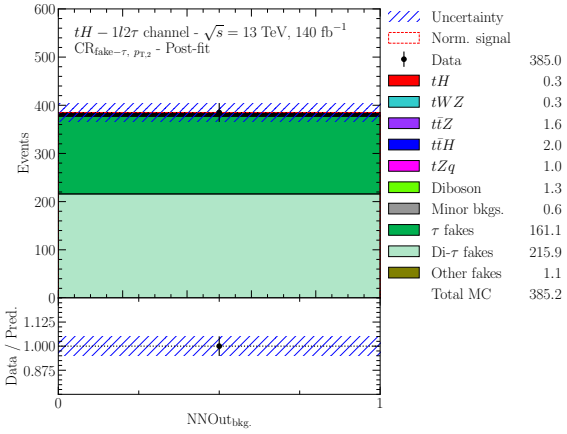

$CR_{\text{fake-}\tau, p_{T,1}}$ pre-fit.



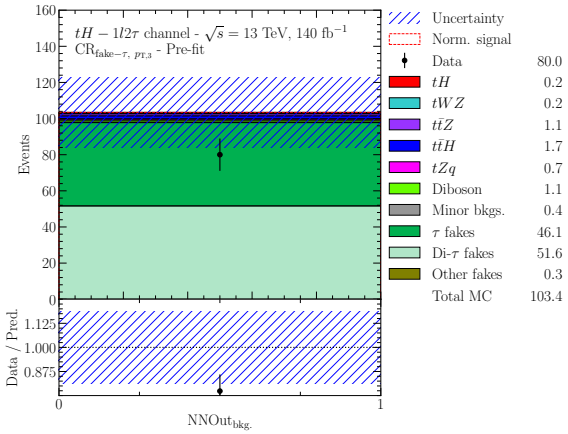

$CR_{\text{fake-}\tau, p_{T,1}}$ post-fit.



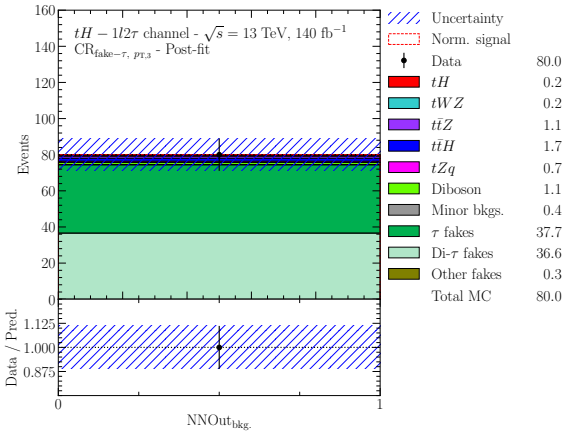

$CR_{\text{fake-}\tau, p_{T,2}}$ pre-fit.



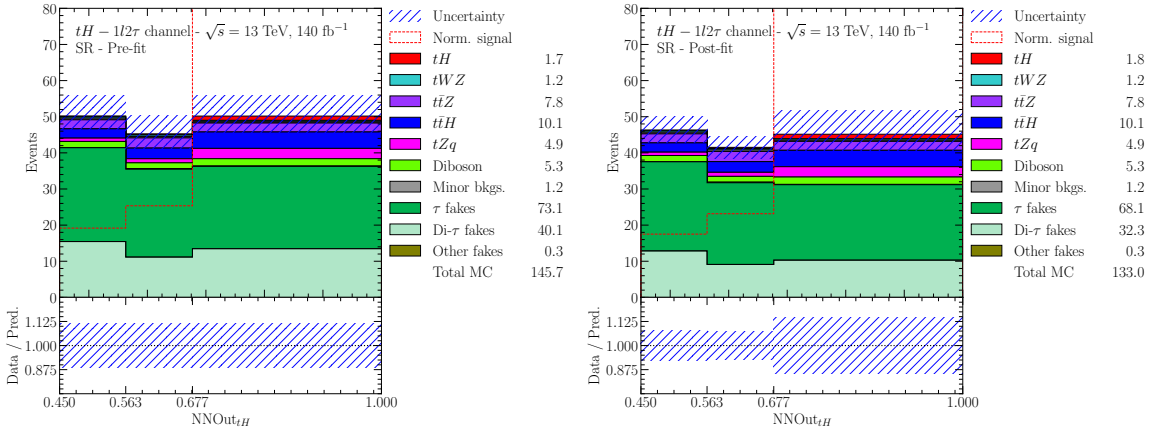

$CR_{\text{fake-}\tau, p_{T,2}}$ post-fit.




$CR_{\text{fake-}\tau, p_{T,3}}$ pre-fit.

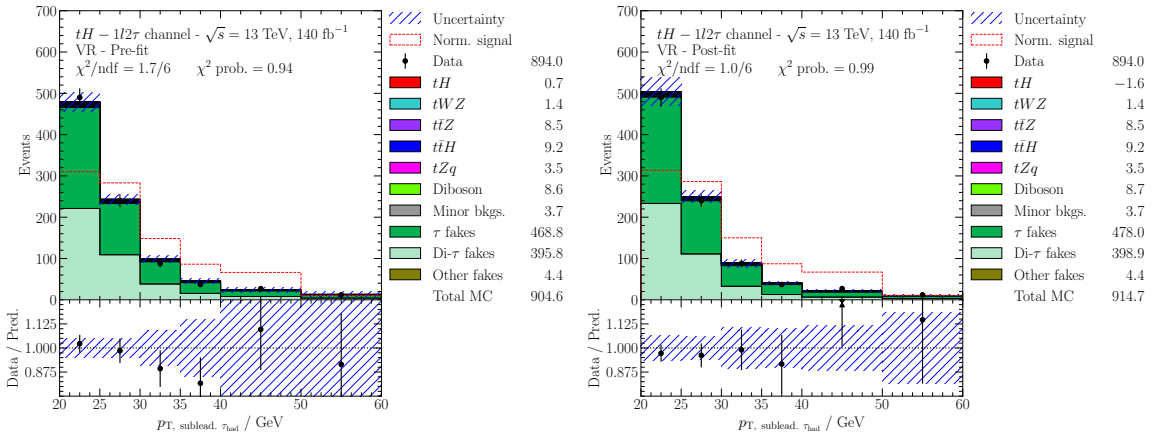



$CR_{\text{fake-}\tau, p_{T,3}}$ post-fit.




SR pre-fit. SR post-fit.

Figure 10.9: Pre- and post-fit distributions of the SR and all CRs used in the $1\ell + 2\tau_{\text{had}}$ channel's hybrid fit.




VR pre-fit. VR post-fit.

Figure 10.10: Pre- and post-fit distributions of the subleading $\tau_{\text{had}}$'s $p_{\text{T}}$ in the VR used in the $1\ell + 2\tau_{\text{had}}$ channel's hybrid fit.

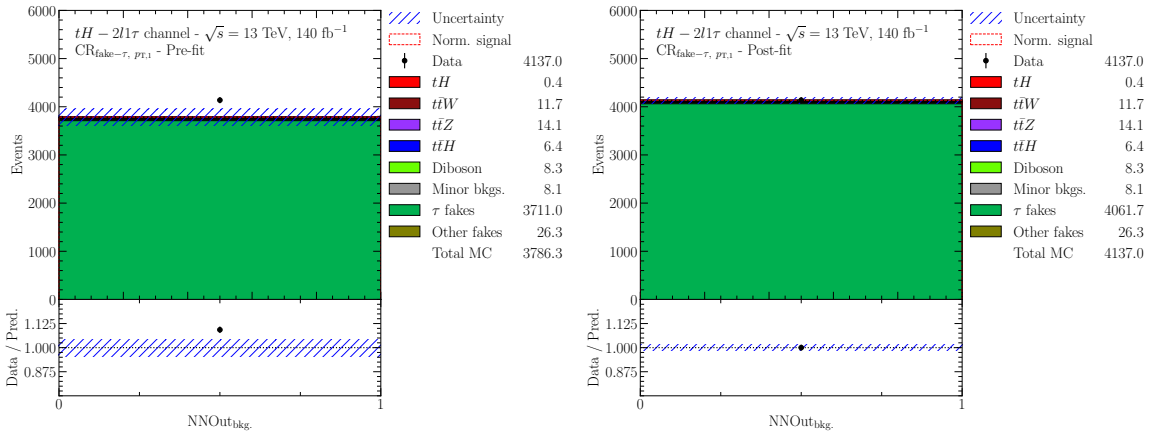
### 10.3.2 $2\ell + 1\tau_{\text{had}}$ - hybrid fit

Analogous to the previous section, the hybrid fit is carried out in the $2\ell + 1\tau_{\text{had}}$ channel. Again, all CRs pass the blinding threshold, allowing real data to be used. The resulting normalisation factors are as follows:

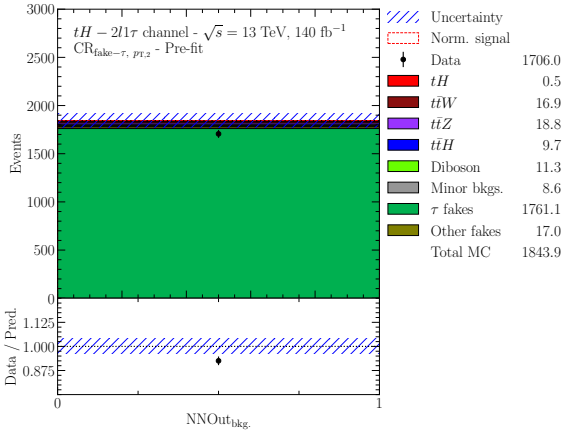
$$\begin{aligned}
\mu_{tH} &= 1.0^{+5.3}_{-4.7}\,(\text{stat.})^{+3.8}_{-3.4}\,(\text{syst.}) &&= 1.0^{+6.5}_{-5.8}\,,\\
\mu(\text{fake-}\tau,\ p_{\text{T},1}) &= 1.09^{+0.02}_{-0.02}\,(\text{stat.})^{+0.05}_{-0.05}\,(\text{syst.}) &&= 1.09^{+0.05}_{-0.05}\,,\\
\mu(\text{fake-}\tau,\ p_{\text{T},2}) &= 0.92^{+0.02}_{-0.02}\,(\text{stat.})^{+0.03}_{-0.03}\,(\text{syst.}) &&= 0.92^{+0.04}_{-0.04}\,,\\
\mu(\text{fake-}\tau,\ p_{\text{T},3}) &= 0.62^{+0.06}_{-0.06}\,(\text{stat.})^{+0.05}_{-0.05}\,(\text{syst.}) &&= 0.62^{+0.07}_{-0.07}\,,\\
\mu(\text{fake-}e) &= 0.91^{+0.03}_{-0.03}\,(\text{stat.})^{+0.05}_{-0.05}\,(\text{syst.}) &&= 0.91^{+0.06}_{-0.06}\,,\\
\mu(\text{fake-}\mu) &= 0.90^{+0.03}_{-0.03}\,(\text{stat.})^{+0.05}_{-0.05}\,(\text{syst.}) &&= 0.90^{+0.06}_{-0.06}\,.
\end{aligned}$$

The expected $p_{\text{T}}$ dependence of fake-$\tau$ events is properly accounted for by the normalisation factors. Fake light leptons contributions are inclusively downscaled by approximately 10 %. The pre- and post-fit distributions are displayed in Figure 10.11.

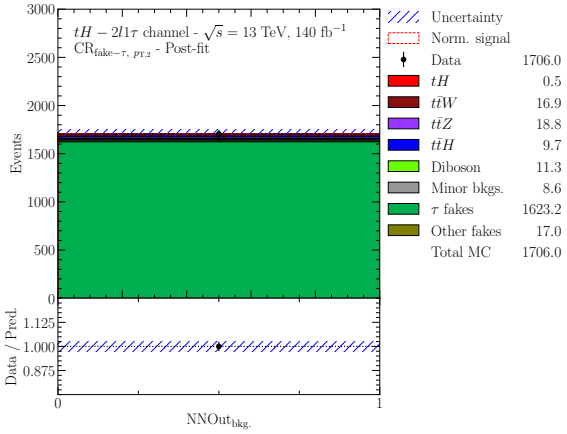
In addition, all fake background corrections can be studied in the VR. As an example, the $\tau_{\text{had}}$'s pre- and post-fit $p_{\text{T}}$ distributions are shown in Figure 10.12. The distributions of all remaining variables are summarised in Figure H.14. Compared to the $1\ell + 2\tau_{\text{had}}$ channel, a worse pre-fit agreement is observed. This is directly connected to the VR's object definition, described in Section 9.3.3: requiring events with an anti-medium $\tau_{\text{had}}$ increases the chance of selecting fake-$\tau$ events. A comparison of the pre- and post-fit distributions, quantified via the shown $\chi^2$ test results, demonstrates that the initial disagreement is resolved through rescaling the fake background contributions. Further nuisance parameter validations reveal no issues and are summarised in Figures H.2(a), H.4(a) and H.10. The final fully unblinded fit based on the same setup is covered in Section 10.4.2.



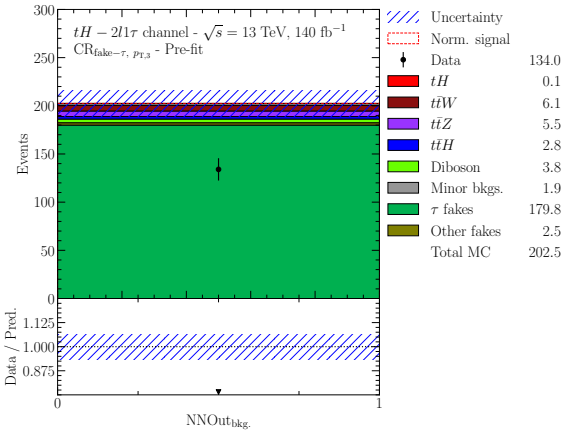

$\text{CR}_{\text{fake-}\tau, p_{\text{T},1}}$ pre-fit. $\text{CR}_{\text{fake-}\tau, p_{\text{T},1}}$ post-fit.



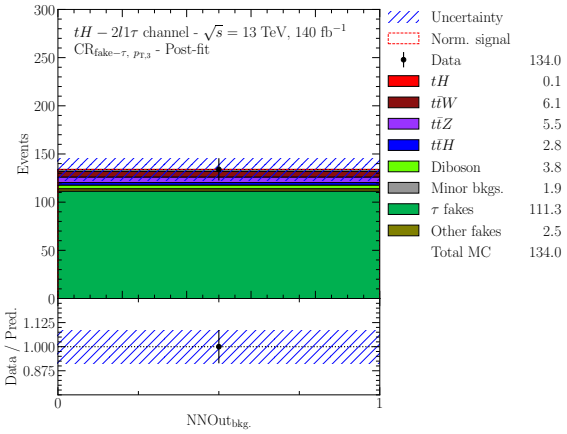

CR$_{\text{fake-}\tau, p_{T,2}}$ pre-fit.



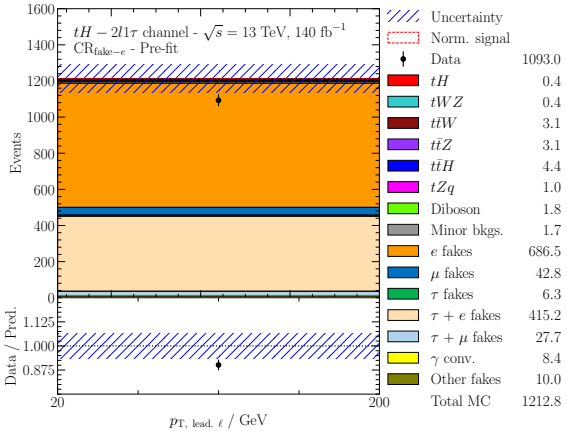

CR$_{\text{fake-}\tau, p_{T,2}}$ post-fit.



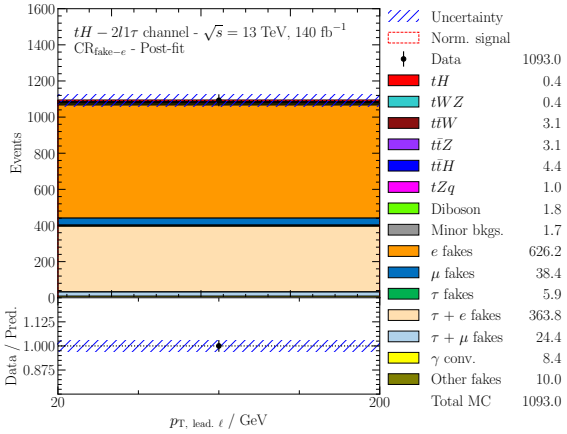

CR$_{\text{fake-}\tau, p_{T,3}}$ pre-fit.



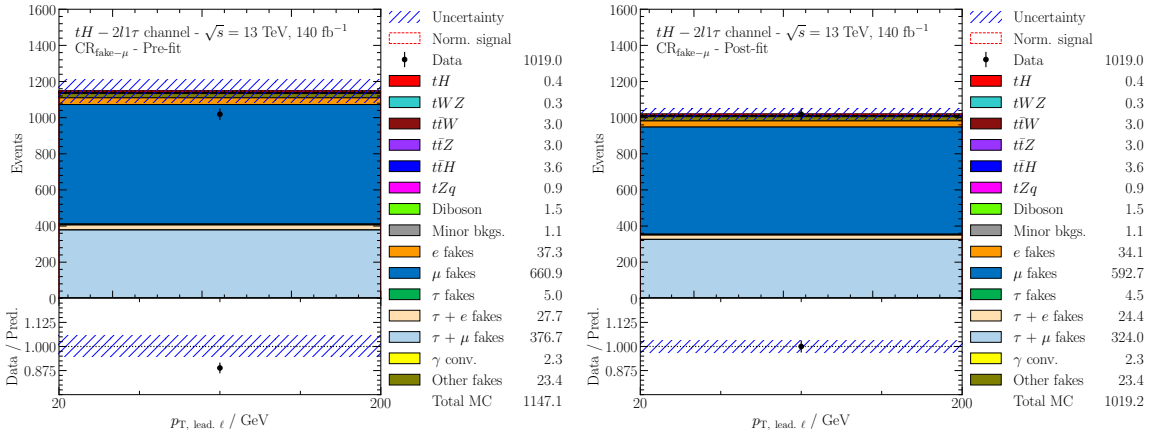

CR$_{\text{fake-}\tau, p_{T,3}}$ post-fit.



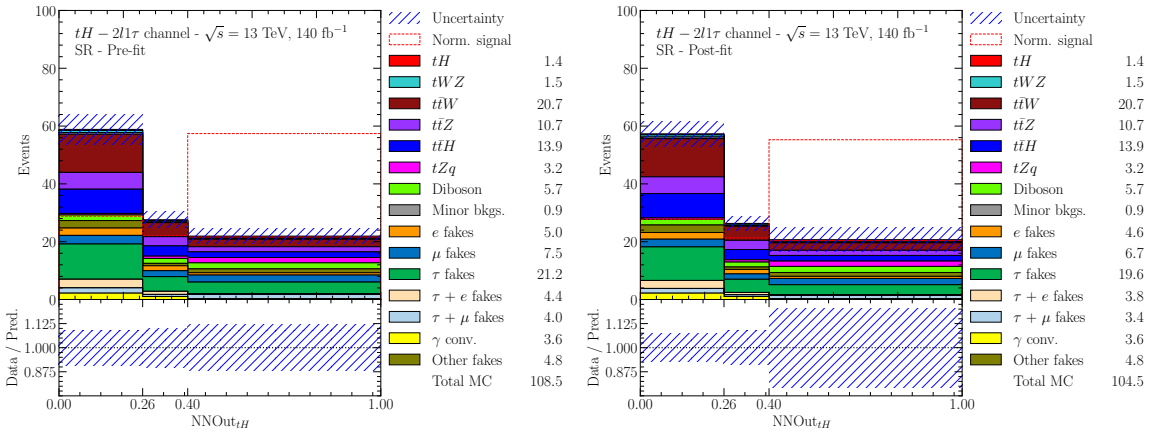

CR$_e$ pre-fit.




CR$_e$ post-fit.

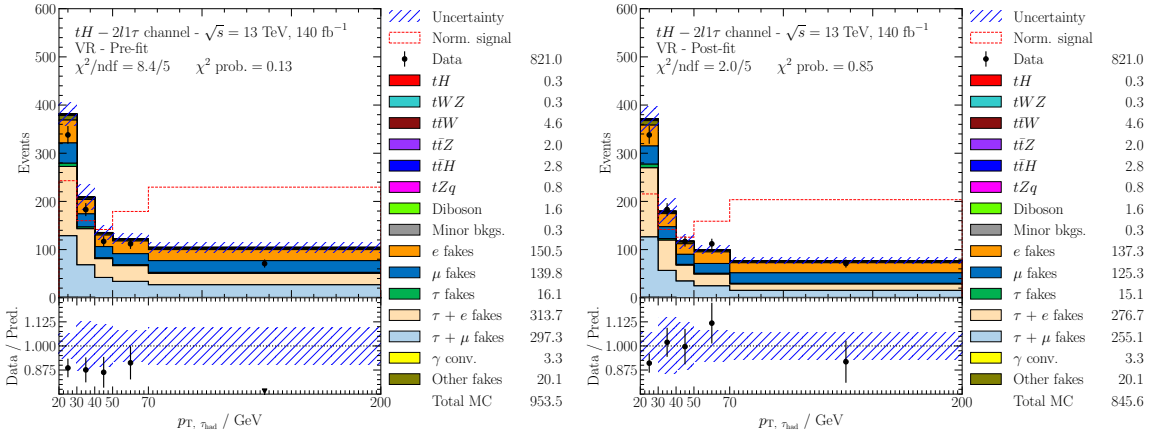



CR$_\mu$ pre-fit. CR$_\mu$ post-fit.




SR pre-fit. SR post-fit.

Figure 10.11: Pre- and post-fit distributions of the SR and all CRs used in the $2\ell + 1\tau_{\text{had}}$ channel's hybrid fit.




VR pre-fit. VR post-fit.

Figure 10.12: Pre- and post-fit distributions of the $\tau_{\text{had}}$'s $p_{\text{T}}$ in the VR used in the $2\ell + 1\tau_{\text{had}}$ channel's hybrid fit.

## 10.4 Fit to data

As a final step, the fully unblinded fit is performed in both channels by using real data in the SRs. The two preceding fits have undergone extensive optimisation and multiple adjustments to achieve a stable and reliable setup. In contrast, the unblinded fit is carried out only once with a finalised setup to avoid introducing any potential bias into the analysis.

### 10.4.1 $1\ell + 2\tau_{\text{had}}$ - data fit

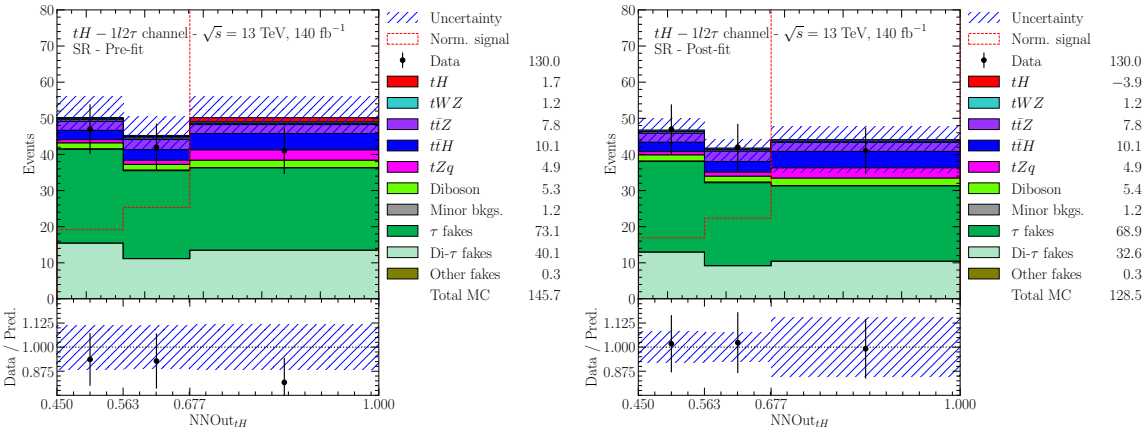
Initially, all normalisation factors in the $1\ell + 2\tau_{\text{had}}$ channel are examined and listed as follows:

$$\begin{aligned}
\mu_{tH} &= -2.3^{+5.2}_{-4.8}\,(\text{stat.})^{+3.7}_{-3.6}\,(\text{syst.}) \quad = -2.3^{+6.4}_{-6.0}\,, \\
\mu(\text{fake-}\tau,\ p_{\text{T},1}) &= 1.08^{+0.04}_{-0.04}\,(\text{stat.})^{+0.03}_{-0.03}\,(\text{syst.}) = 1.08^{+0.05}_{-0.05}\,, \\
\mu(\text{fake-}\tau,\ p_{\text{T},2}) &= 0.94^{+0.05}_{-0.05}\,(\text{stat.})^{+0.05}_{-0.05}\,(\text{syst.}) = 0.94^{+0.07}_{-0.07}\,, \\
\mu(\text{fake-}\tau,\ p_{\text{T},3}) &= 0.71^{+0.12}_{-0.12}\,(\text{stat.})^{+0.17}_{-0.17}\,(\text{syst.}) = 0.71^{+0.21}_{-0.21}\,.
\end{aligned}$$

Compared with the hybrid fit approach, only minor deviations of 1 % are observed in case of the fake-$\tau$ normalisation factors. Each factor's value is predominantly determined in the dedicated fake-$\tau$ CR. Therefore, including real data in the SR only leads to the observed minor corrections. Any significant deviation would hint at a potential issue, suggesting that the modelling of fake-$\tau$ events in the CRs does not reliably extrapolate to the SR.

For $\mu_{tH}$, the fit yields a negative value. In the absence of interference effects with other processes, such a value is, in principle, unphysical. However, a negative result can occur when the observed data are lower than expected by the background-only hypothesis. Allowing $\mu_{tH}$ to take on negative values ensures that the likelihood function remains approximately symmetrical near zero. This is particularly important for the proper estimation of confidence intervals and when combining multiple analysis channels. Nonetheless, the Standard Model prediction, corresponding to $\mu_{tH} = 1$, lies well within the measurement's $\pm 1\sigma$ interval.

The small changes in the fake background normalisation factors, compared to the hybrid fit in Section 10.3.1, indicate that unblinding the SR does not affect distributions in the CRs and the VR. Consequently, only the pre- and post-fit SR results are visualised in Figure 10.13. Studies on the nuisance parameters' correlations, pulls and constraints can be found in Figures H.1(b) and H.7, while the ranked impact is shown in Figure 10.14. Unblinding the SR causes minor pulls, well within the $\pm 1\sigma$ range.




SR pre-fit. SR post-fit.

Figure 10.13: Pre- and post-fit distributions of the SR used in the $1\ell + 2\tau_{\text{had}}$ channel's fully unblinded fit.

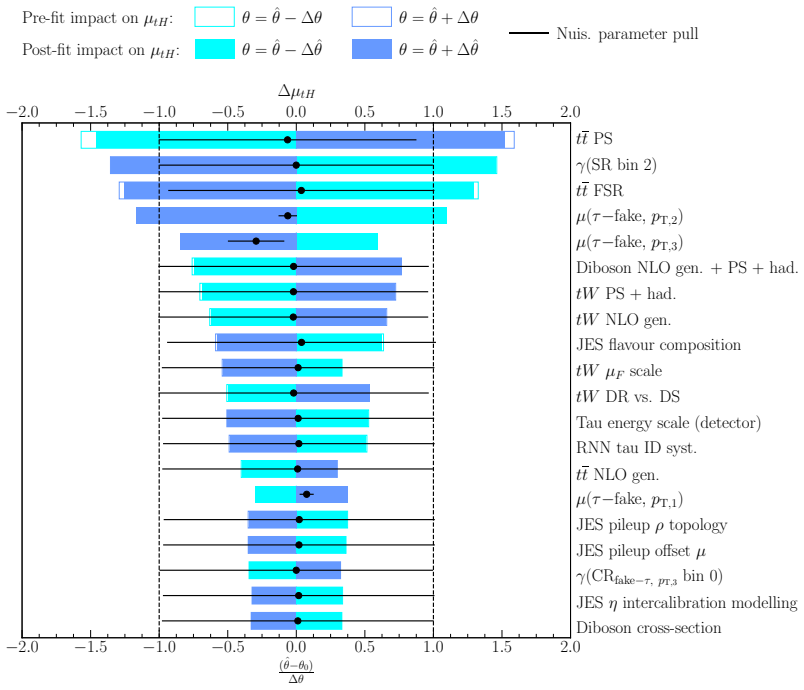



Figure 10.14: Nuisance parameter impact ranking for the $1\ell + 2\tau_{\text{had}}$ channel's fully unblinded fit.

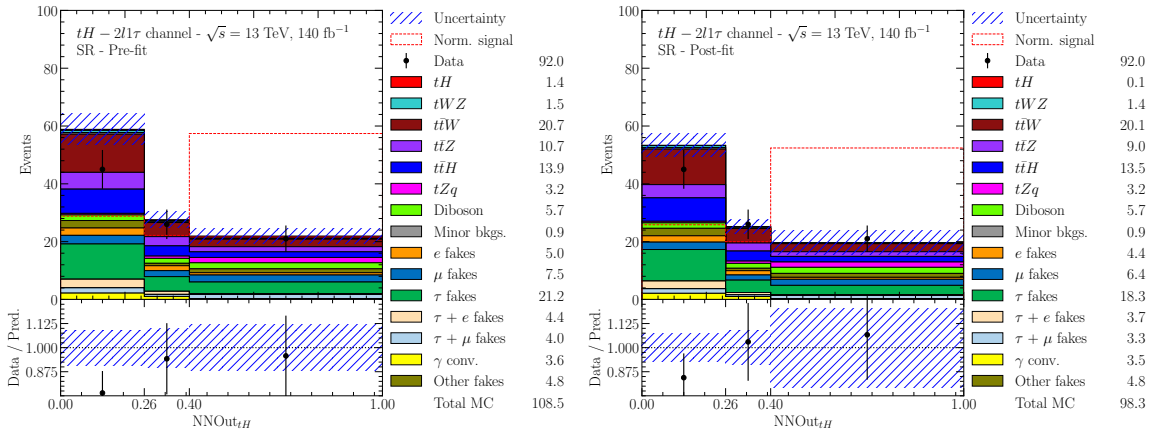
### 10.4.2 $2\ell + 1\tau_{\text{had}}$ - data fit

Completely unblinding the $2\ell + 1\tau_{\text{had}}$ channel results in the following normalisation factors:

$$\begin{aligned}
\mu_{tH} &= 0.1^{+5.0}_{-4.5}\,(\text{stat.})^{+3.5}_{-3.4}\,(\text{syst.}) &&= 0.1^{+6.1}_{-5.6}\,,\\
\mu(\text{fake-}\tau,\ p_{\text{T},1}) &= 1.09^{+0.02}_{-0.02}\,(\text{stat.})^{+0.05}_{-0.05}\,(\text{syst.}) &&= 1.09^{+0.05}_{-0.05}\,,\\
\mu(\text{fake-}\tau,\ p_{\text{T},2}) &= 0.92^{+0.02}_{-0.02}\,(\text{stat.})^{+0.03}_{-0.03}\,(\text{syst.}) &&= 0.92^{+0.04}_{-0.04}\,,\\
\mu(\text{fake-}\tau,\ p_{\text{T},3}) &= 0.63^{+0.06}_{-0.06}\,(\text{stat.})^{+0.05}_{-0.05}\,(\text{syst.}) &&= 0.63^{+0.08}_{-0.08}\,,\\
\mu(\text{fake-}e) &= 0.89^{+0.03}_{-0.03}\,(\text{stat.})^{+0.05}_{-0.05}\,(\text{syst.}) &&= 0.89^{+0.06}_{-0.06}\,,\\
\mu(\text{fake-}\mu) &= 0.91^{+0.03}_{-0.03}\,(\text{stat.})^{+0.05}_{-0.05}\,(\text{syst.}) &&= 0.90^{+0.06}_{-0.06}\,.
\end{aligned}$$

As already observed in the unblinded $1\ell + 2\tau_{\text{had}}$ channel, the normalisation factors for the fake background processes deviate by at most 1 % from their corresponding hybrid fit results. This confirms that unblinding the SR does not impact the corrections, thereby justifying the extrapolation from the CRs. The result for $\mu_{tH}$ agrees with the Standard Model expectation within $\pm 1\sigma$. The obtained post-fit SR distribution is shown in Figure 10.15. CRs, as well as VR results are omitted, as the associated distributions are almost unchanged with respect to Figure 10.11.

The impact of each nuisance parameter on $\mu_{tH}$ is illustrated in Figure 10.16. Again, only small pulls, well within the $\pm 1\sigma$ interval, are observed. These pulls arise from the SR's first bin with its more pronounced deviation between data and simulation. Assuming that all fake backgrounds are roughly uniformly distributed, a perfect agreement cannot be achieved merely by a flat renormalisation. Instead, the fit corrects this first bin by pulling nuisance parameters which impact the distribution's shape. Further nuisance parameter studies can be found in Figures H.2(b) and H.11.




SR pre-fit. SR post-fit.

Figure 10.15: Pre- and post-fit distributions of the SR used in the $2\ell + 1\tau_{\text{had}}$ channel's fully unblinded fit.

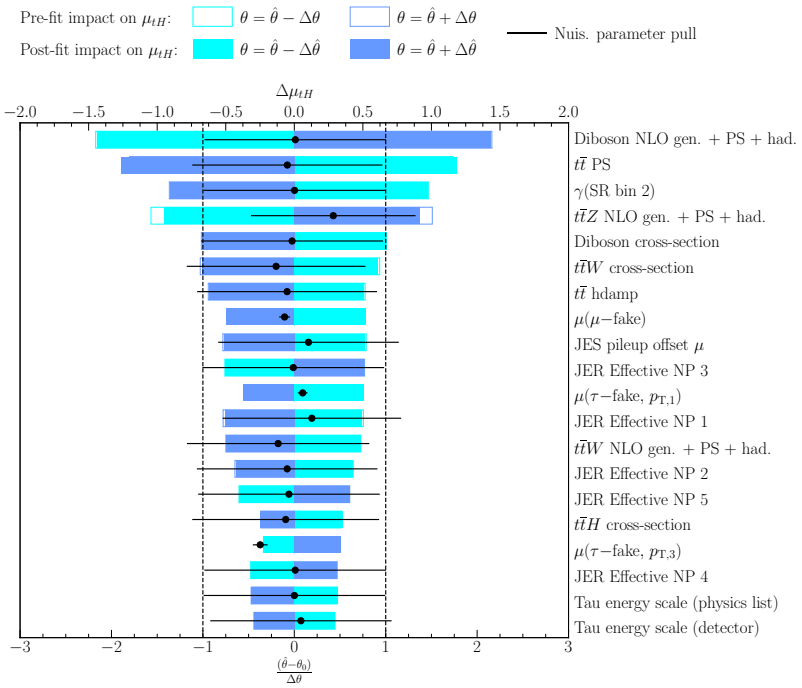



Figure 10.16: Nuisance parameter impact ranking for the $2\ell + 1\tau_{\text{had}}$ channel's fully unblinded fit.

## 10.5 Fit to data - inverted coupling ($y_t = -1$)

The entire analysis is designed towards measuring the $tH$ process as predicted by the Standard Model. The simulated $tHq$ and $tWH$ events, used in the CNN training and the background estimation, are generated under the assumption of a destructive interference, i.e. a positive sign of the Yukawa coupling between the Higgs-boson and the top-quark ($y_t = 1$). Nevertheless, the final fit setup also allows testing of the alternative inverted coupling hypothesis ($y_t = -1$). As discussed in Section 2.3, this hypothesis predicts a constructive interference in the $tH$ channel, thereby enhancing the cross-section. The corresponding fit is performed in both channels by relying on the fully unblinded setup discussed in Section 10.4. The original MC simulation of $tH$ events is replaced by a modified sample that assumes $y_t = -1$.

In the case of the $1\ell + 2\tau_{\text{had}}$ channel, the resulting normalisation factors are given as follows:

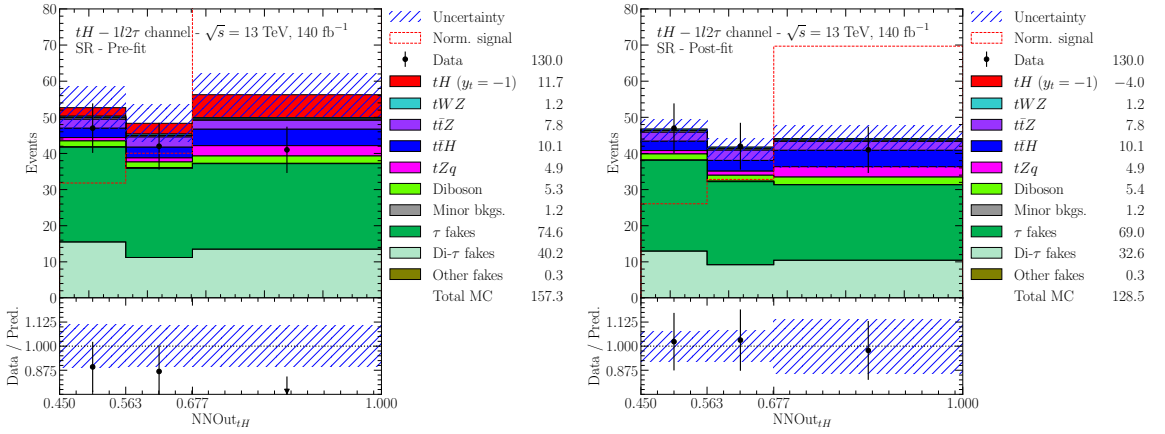
$$\begin{aligned}
\mu_{tH}(y_t = -1) &= -0.3^{+0.9}_{-0.9}\,(\text{stat.})^{+0.8}_{-0.8}\,(\text{syst.}) \quad = -0.3^{+1.2}_{-1.2}\,, \\
\mu(\text{fake-}\tau,\ p_{\text{T},1}) &= 1.08^{+0.04}_{-0.04}\,(\text{stat.})^{+0.03}_{-0.03}\,(\text{syst.}) = 1.08^{+0.05}_{-0.05}\,, \\
\mu(\text{fake-}\tau,\ p_{\text{T},2}) &= 0.94^{+0.05}_{-0.05}\,(\text{stat.})^{+0.05}_{-0.05}\,(\text{syst.}) = 0.94^{+0.07}_{-0.07}\,, \\
\mu(\text{fake-}\tau,\ p_{\text{T},3}) &= 0.71^{+0.12}_{-0.12}\,(\text{stat.})^{+0.17}_{-0.17}\,(\text{syst.}) = 0.71^{+0.21}_{-0.21}\,.
\end{aligned}$$

For the $2\ell + 1\tau_{\text{had}}$ channel one obtains:

$$\begin{aligned}
\mu_{tH}(y_t = -1) &= -0.2^{+0.8}_{-0.7}\,(\text{stat.})^{+0.6}_{-0.7}\,(\text{syst.}) \quad = -0.2^{+1.0}_{-1.0}\,, \\
\mu(\text{fake-}\tau,\ p_{\text{T},1}) &= 1.09^{+0.02}_{-0.02}\,(\text{stat.})^{+0.05}_{-0.05}\,(\text{syst.}) = 1.09^{+0.05}_{-0.05}\,, \\
\mu(\text{fake-}\tau,\ p_{\text{T},2}) &= 0.92^{+0.02}_{-0.02}\,(\text{stat.})^{+0.03}_{-0.03}\,(\text{syst.}) = 0.92^{+0.04}_{-0.04}\,, \\
\mu(\text{fake-}\tau,\ p_{\text{T},3}) &= 0.63^{+0.06}_{-0.06}\,(\text{stat.})^{+0.05}_{-0.05}\,(\text{syst.}) = 0.63^{+0.08}_{-0.08}\,, \\
\mu(\text{fake-}e) &= 0.89^{+0.03}_{-0.03}\,(\text{stat.})^{+0.05}_{-0.05}\,(\text{syst.}) = 0.89^{+0.06}_{-0.06}\,, \\
\mu(\text{fake-}\mu) &= 0.91^{+0.03}_{-0.03}\,(\text{stat.})^{+0.05}_{-0.05}\,(\text{syst.}) = 0.90^{+0.06}_{-0.06}\,.
\end{aligned}$$

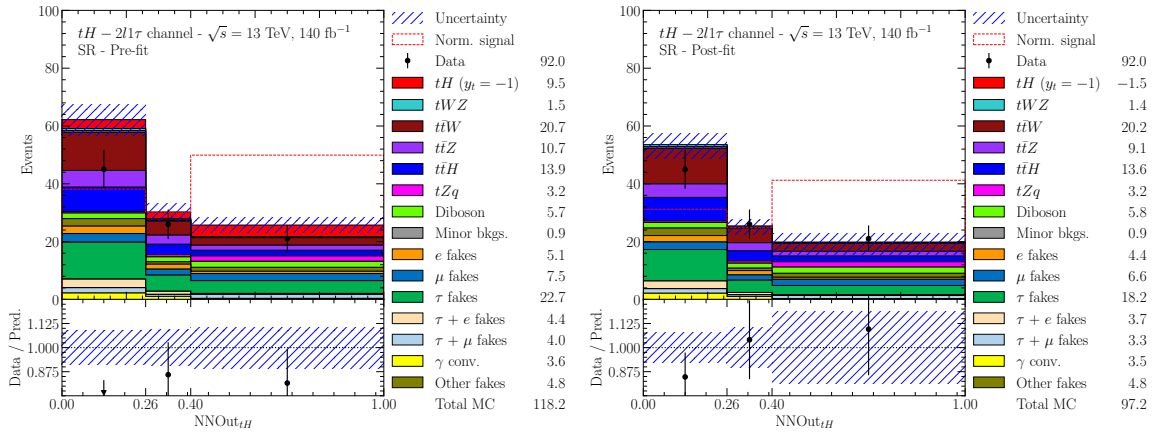
In both channels, the modified signal model leads to a shift for $\mu_{tH}$ with respect to the results in Section 10.4, while the background normalisation factors remain unaffected. Thus, only the SRs for the respective channels are shown in Figures 10.17 and 10.18. A negative best-fit value for the $tH$ signal strength is obtained in both channels. A comparison with the results in Section 10.4 indicates that this is in good agreement with the expectations. Even for the Standard Model hypothesis, the observed data are best described by a downscaled $tH$ contribution. In case of the alternative hypothesis, given its associated constructive interference, more $tH$ events are expected in all regions, especially in the SR. Therefore, a stronger downscaling is required.

Taking this argument into account, the $2\ell + 1\tau_{\text{had}}$ channel's $\mu_{tH}$ result may appear somewhat counterintuitive, as a lower normalisation factor is obtained when testing the Standard Model hypothesis (see Section 10.4.2). However, under the alternative hypothesis the SR's pre-fit $tH$ contribution is approximately seven times larger. Since this contribution is directly multiplied by $\mu_{tH}$, a larger negative normalisation factor value would cause the observed data to be underestimated. A comparison of Figure 10.15 and Figure 10.18 demonstrates that both $\mu_{tH}$ results lead to an identical post-fit contribution, indicating a stable fit setup. Further studies are performed in analogy to the Standard Model hypothesis fits. Their results are summarised in Figures H.1(c), H.3(b) and H.8 as well as Figures H.2(c), H.4(b) and H.12 for the $1\ell + 2\tau_{\text{had}}$ channel and the $2\ell + 1\tau_{\text{had}}$ channel.




SR pre-fit. SR post-fit.

Figure 10.17: Pre- and post-fit distributions of the SR used in the $1\ell + 2\tau_{\text{had}}$ channel's fully unblinded inverted coupling fit.




SR pre-fit. SR post-fit.

Figure 10.18: Pre- and post-fit distributions of the SR used in the $2\ell + 1\tau_{\text{had}}$ channel's fully unblinded inverted coupling fit.

## 10.6 Combined fit

To enhance the overall sensitivity and statistical significance, the results from both channels are combined in the final step of the analysis. Since the two channels are statistically independent, their combination can be expressed as the product of the individual likelihoods:

$$L(\mu, \vec{\theta}) = \prod_{k=1}^{2} L_k(\mu, \vec{\theta}_k) \,. \tag{10.3}$$

Consequently, the whole dataset is exploited when testing the Standard Model hypothesis as well as the $y_t = -1$ scenario. For the Standard Model hypothesis, the combined results are displayed in Figure 10.19, leading to

$$\mu_{tH} = -1.1^{+3.6}_{-3.3}\,(\text{stat.})^{+2.5}_{-2.6}\,(\text{syst.}) = -1.1^{+4.4}_{-4.2}\,.$$

The Standard Model hypothesis is found to be within one standard deviation for all individual results, as well as for the combined measurement. Compared with systematic effects, statistical uncertainties have a greater impact on the final result. Therefore, repeating this analysis with a larger dataset would significantly increase the sensitivity.

As shown in Figure 10.19(b), at a 95 % confidence level, conventionally used in particle physics, a cross-section exceeding 8.5 times the Standard Model prediction can be excluded. This sets an upper limit of $\sigma(tH) = 760\,\text{fb}$ on the $tH$ cross-section.

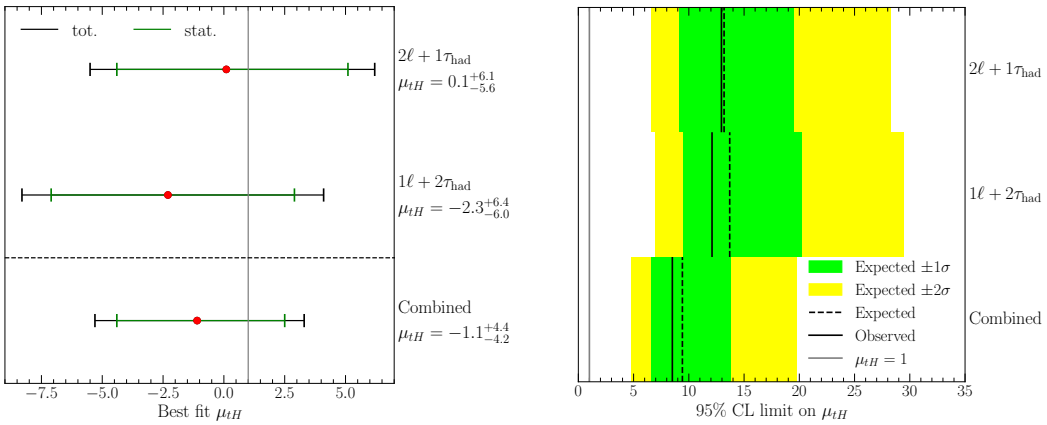
Combining the inverted coupling results yields

$$\mu_{tH}(y_t = -1) = -0.3^{+0.6}_{-0.6}\,(\text{stat.})^{+0.5}_{-0.5}\,(\text{syst.}) = -0.3^{+0.8}_{-0.8}\,,$$

as shown in Figure 10.20. The nominal signal strength in the case of $y_t = -1$ lies outside the $\pm 2\sigma$

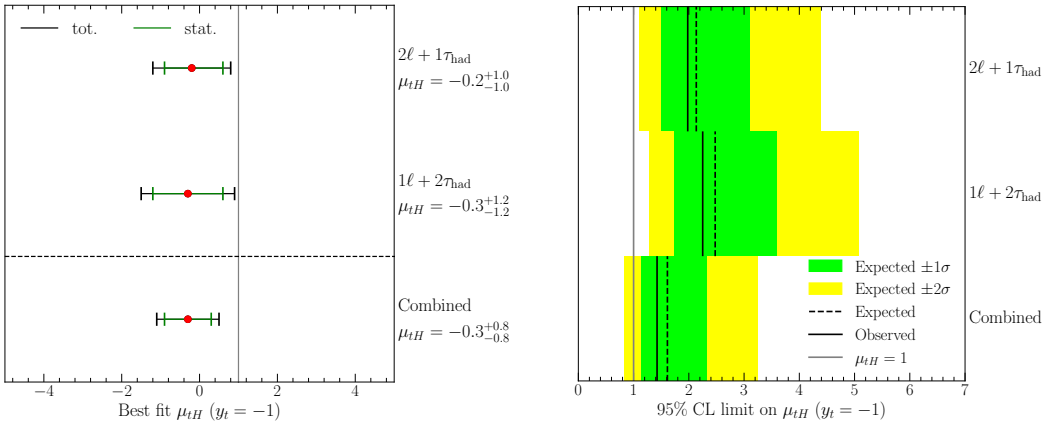
interval of the measurement. At a 95 % confidence level, any cross-section exceeding 1.4 times the inverted coupling hypothesis can be excluded.

To ultimately evaluate the stability of the fit results obtained for both hypotheses, a goodness-of-fit test is performed. This test is carried out by comparing each model with a saturated model [158]. Such a model is set up with enough degrees of freedom to perfectly describe the observed data without pulling any nuisance parameters, thus avoiding penalising the likelihood. The ratio of both models' likelihoods follows a $\chi^2$-distribution, allowing for a standard goodness-of-fit test. This test yields a probability of 83 % (84 %) for the $y_t = 1$ ($y_t = -1$) case, supporting the validity of each performed combination.




(a) (b)

Figure 10.19: Results of the combined fit under the Standard Model hypothesis, showing (a) the best-fit signal strength value $\mu_{tH}$ as well as (b) the exclusion limits at a 95 % confidence level. The expected limit is shown along with the one- and two-standard deviation uncertainty bands, in addition to the observed limit.




(a) (b)

Figure 10.20: Results of the combined fit under the inverted coupling hypothesis, showing (a) the best-fit signal strength value $\mu_{tH}$ as well as (b) the exclusion limits at a 95 % confidence level. The expected limit is shown along with the one- and two-standard deviation uncertainty bands, in addition to the observed limit.

CHAPTER 11

# Conclusion

In 2012, the discovery of the Higgs-boson marked the most recent addition of a pivotal piece to the particle physics jigsaw puzzle known as the Standard Model. Nevertheless, experimental evidence keeps revealing major gaps in this model. To close such gaps, finding puzzle pieces that explain and incorporate phenomena like dark matter or gravity is crucial. So, it comes as no surprise that the search for these missing puzzle pieces is central to today's efforts in particle physics.

This work presents an analysis aimed at refining and constraining the known puzzle pieces' contours by investigating properties of the heaviest elementary particles, the top-quark and the Higgs-boson. Through measuring the cross-section of the associated production of a top-quark and a Higgs-boson ($tH$), both the magnitude and importantly also the relative sign of the Yukawa coupling $y_t$ can be determined. The Standard Model predicts a positive sign ($y_t = 1$), causing a destructive interference in the $tH$ channel and thus complicating the search by reducing the expected cross-section.

The analysis uses data from $pp$ collisions recorded at $\sqrt{s} = 13\,\text{TeV}$ by the ATLAS detector during Run 2 of the LHC, corresponding to an integrated luminosity of $L_{\text{int}} = 140\,\text{fb}^{-1}$. The $H \to \tau\tau$ channels are targeted and split into two distinct regions according to the number of hadronically decaying tau-leptons ($\tau_{\text{had}}$). To enhance the overall sensitivity, a categorical neural network is trained and optimised in each region, providing an improved separation between signal and background processes. Given their high discriminatory power, the resulting predictions are chosen to serve as a fit variable in a binned profile likelihood fit to measure the signal strength ($\mu_{tH}$), defined as the ratio between the measured and the predicted $tH$ cross-section. A reliable result is obtained by simultaneously validating and correcting the modelling of all dominant background processes in dedicated regions.

Combining both channels in a final fit yields a signal strength of

$$\mu_{tH} = -1.1\,^{+3.6}_{-3.3}\,(\text{stat.})\,^{+2.5}_{-2.6}\,(\text{syst.}) = -1.1\,^{+4.4}_{-4.2}\,,$$

which is in agreement with the Standard Model prediction and consistent with the results reported by the CMS collaboration [48] at the $2\sigma$ level. At a 95 % confidence level, cross-sections exceeding 8.5 times the Standard Model prediction can be excluded, allowing an upper limit of $\sigma(tH) = 760\,\text{fb}$ to be set. In addition, the alternative hypothesis of a modified Yukawa coupling with inverted sign, leading to the expectation of an enhanced cross-section, is evaluated. Using the same profile likelihood fit

setup a signal strength of

$$\mu_{tH}(y_t = -1) = -0.3^{+0.6}_{-0.6}\,(\text{stat.})^{+0.5}_{-0.5}\,(\text{syst.}) = -0.3^{+0.8}_{-0.8}\,,$$

is obtained. Corresponding 95 % confidence level limits lie at 1.4 times the predicted cross-section.

For both hypotheses, the measurement is dominated by the statistical uncertainty of the data available. Therefore, the presented analysis provides an ideal foundation for a future combination of this dataset with the full ATLAS Run 3 dataset expected by summer 2026. This enlarged dataset would enable stricter limits to be set.

APPENDIX A

# Natural units

All calculations in this thesis are performed in the natural unit system. This system is commonly used in particle physics with the purpose of simplifying expressions. Using the reduced Planck constant $\hbar = 1.0546 \times 10^{-34}$ Js and the speed of light $c = 2.9979 \times 10^{8}$ m/s, one can convert units of distance, mass, and time to powers of units of energy.

In this system, energies $E$ are not expressed in units of joule but in eV, where 1 eV is defined as the kinetic energy that an electron receives when passing through an electric potential difference of 1 V. For simplicity one additionally defines:

$$\hbar = c = 1\,. \tag{A.1}$$

This simplifies the energy mass relation to

$$E^2 = p^2 + m^2\,. \tag{A.2}$$

Obviously, mass and momentum can also be expressed in units of eV in this system. Further considering the quantised photon energy

$$E = h\nu = 2\pi\hbar\nu = 2\pi\nu\,, \tag{A.3}$$

where $\nu = 1/T$ denotes the frequency, it quickly becomes clear that time is specified in units of $\mathrm{eV}^{-1}$. Similarly, using the de Broglie wavelength

$$p = \frac{h}{\lambda} = \frac{2\pi\hbar}{\lambda} = \frac{2\pi}{\lambda}\,, \tag{A.4}$$

it can be shown that lengths are also given in units of $\mathrm{eV}^{-1}$, areas consequently in units of $\mathrm{eV}^{-2}$.

APPENDIX B

# Lorentz vectors and transformations

In particle physics, it is common to express space-time coordinates using a four-vector $\vec{r}$, given by

$$\vec{r} = \begin{pmatrix} t \\ x \\ y \\ z \end{pmatrix}. \tag{B.1}$$

These vectors, also called Lorentz vectors, share their linearity properties with standard Euclidean vectors, i.e.

$$\vec{r}_1 + \vec{r}_2 = \begin{pmatrix} t_1 \\ x_1 \\ y_1 \\ z_1 \end{pmatrix} + \begin{pmatrix} t_2 \\ x_2 \\ y_2 \\ z_2 \end{pmatrix} = \begin{pmatrix} t_1 + t_2 \\ x_1 + x_2 \\ y_1 + y_2 \\ z_1 + z_2 \end{pmatrix}, \tag{B.2}$$

applies. The scalar multiplication also matches that of Euclidean vectors. However, a vector's magnitude is defined as

$$|\vec{r}| = \sqrt{t^2 - x^2 - y^2 - z^2}, \tag{B.3}$$

and represents a quantity, which is invariant under Lorentz transformations. Lorentz transformations form the basis of special relativity. They allow transformations between two inertial systems, that are moving with a constant velocity $v$ relative to each other [7].

The simplest transformation considers a particle to be located at $\vec{r}_1$ in one inertial system and at $\vec{r}_2$ in a second inertial system. The second system moves with velocity $v$ along the $x$-direction relative to the first system. Both systems' coordinates can be linked by a Lorentz transformation in natural units via

$$t_2 = \gamma(t_1 - \beta x_1), \qquad x_2 = \gamma(x_1 - \beta t_1), \qquad y_2 = y_1 \qquad \text{and} \qquad z_2 = z_1\,. \tag{B.4}$$

In this case $\gamma = 1/\sqrt{1-\beta^2}$ denotes the Lorentz factor with $\beta = v/c$. For $v \ll c$, this transformation corresponds to Galilei transformations used in Newtonian mechanics. Using the relativistic expressions

for energy $\gamma m$ and the three-dimensional momentum's magnitude $\gamma\beta m$, the four-momentum Lorentz vector

$$\vec{p} = \begin{pmatrix} E \\ p_x \\ p_y \\ p_z \end{pmatrix}, \tag{B.5}$$

is defined similar to the space-time four-vector $\vec{r}$. Its magnitude

$$|\vec{p}| = \sqrt{E^2 - p_x^2 - p_y^2 - p_z^2} = m \ , \tag{B.6}$$

yields the energy-momentum relationship. As a direct consequence, a particle's mass $m$ represents a Lorentz invariant quantity. Four-vectors reconstructed in the ATLAS coordinates system can be transformed back into Cartesian momenta by

$$\vec{p} = \begin{pmatrix} E \\ p_x \\ p_y \\ p_z \end{pmatrix} = \begin{pmatrix} E \\ p_\mathrm{T} \cos\phi \\ p_\mathrm{T} \sin\phi \\ p_\mathrm{T} \sinh\eta \end{pmatrix}. \tag{B.7}$$

APPENDIX C

# Integration with Monte Carlo methods

As discussed in Section 4.4, the field of application for MC methods is broad but always utilises repeated random sampling. In general, the selected sampling numbers must be uniformly distributed, and each method's accuracy increases with the sample size. This appendix provides a short explanation to this idea by discussing the best-known example: determining the number $\pi$ by performing integration.

First imagine a quadrant with radius $r = 1$ plotted within a unit square as shown in Figure C.1. Now the random numbers $x$ as well as $y$ are sampled from a uniform distribution in the interval $[0, 1)$ and subsequently plotted in the unit square. The probability of one scatter point lying within the quadrant obviously corresponds to the ratio of the two areas

$$p = \lim_{N\to\infty} \frac{N(\text{quadrant})}{N} = \frac{A_{\text{quadrant}}}{A_{\text{square}}} = \frac{\frac{1}{4}\pi r^2}{r^2} = \frac{1}{4}\pi\,, \tag{C.1}$$

where $N(\text{quadrant})$ denotes the number of all scatter points $N$ that lie inside the quadrant's area, i.e. fulfil $\sqrt{x^2 + y^2} < 1$.

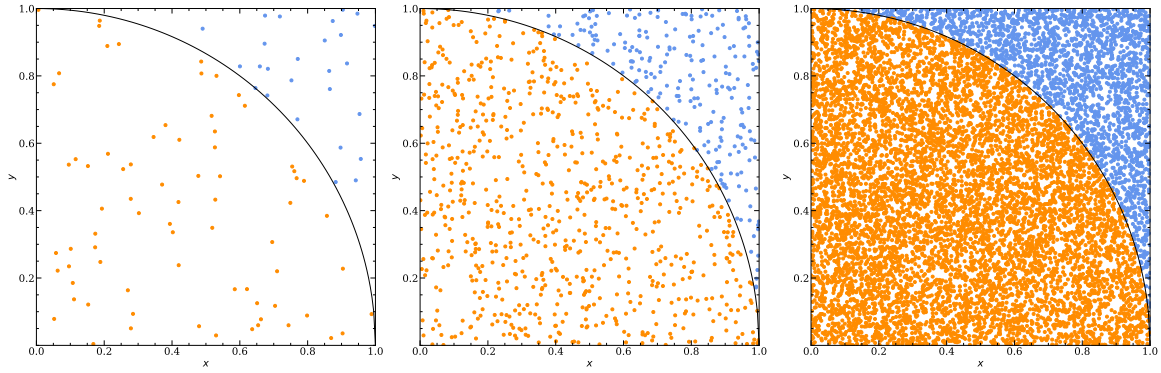
The integration in this example was carried out with $N_1 = 100$, $N_2 = 1000$ and $N_3 = 10000$. The results, visualised in Figures C.1(a) to C.1(c), are:

$$\pi = 4\frac{N(\text{quadrant})}{N} = 4\frac{76}{100} \approx 3.04\,, \tag{C.2}$$

$$\pi = 4\frac{N(\text{quadrant})}{N} = 4\frac{795}{1000} \approx 3.18\,, \tag{C.3}$$

$$\pi = 4\frac{N(\text{quadrant})}{N} = 4\frac{7868}{10000} \approx 3.15\,. \tag{C.4}$$

Evidently, the value of $\pi$ can be determined using the MC integration while the increase in accuracy for a larger sample is apparent. A generalisation of this method combined with a detailed introduction to the topic can be found in [114].




(a) (b) (c)

Figure C.1: Approximating $\pi$ with MC methods by using (a) 100, (b) 1 000 and (c) 10 000 uniformly distributed scattering points.

APPENDIX D

# Systematic uncertainties

This appendix provides a summary of all systematic uncertainties considered in this analysis. All uncertainties can be divided into two main categories:

- Instrumental uncertainties, which affect the reconstruction of physics objects in the detector.
- Modelling uncertainties arising when generating MC samples.

Instrumental uncertainties are evaluated by dedicated groups within the ATLAS collaboration and are simulated by rescaling the nominal MC samples. The following uncertainties are taken into account:

- The uncertainty in the integrated luminosity measurement is primarily determined by the LUCID-2 detector [159] and complemented by additional measurements within the ATLAS detector itself.
- The uncertainty associated with pile-up reweighting, i.e. rescaling the MC pile-up distribution to match the data, is estimated by varying the derived reweighting factors.
- Lepton selection uncertainties include variations in trigger, reconstruction, identification, and isolation efficiencies. Additionally, uncertainties in the lepton momentum scale and resolution are included, addressing potential systematic offsets and statistical fluctuations.
- The uncertainty on $E_{\mathrm{T}}^{\mathrm{miss}}$ is modelled through variations in the scale and resolution of the soft component in Equation (3.18).
- Uncertainties in jet reconstruction primarily arise from the determination of the JER and the JES. These effects are parametrised using a set of nuisance parameters that account for dependencies on $\eta$ and $p_{\mathrm{T}}$. Additionally, variations in the Jet Vertex Tagger's efficiency are considered.
- Flavour-tagging uncertainties are treated independently for $b$-jets, $c$-jets, and light-flavour jets. In all cases, the uncertainties are modelled as functions of $p_{\mathrm{T}}$.

In contrast, modelling uncertainties – apart from cross-section uncertainties – are generally not evaluated through simple rescaling. Instead, they are assessed either by varying parameters within the nominal simulation or by comparing the nominal simulation to alternative simulations produced using different MC generators. In the following, each modelling uncertainty is briefly introduced, along with a list of the processes for which it is included in the analysis.

- Perturbative QCD calculations depend on the renormalisation ($\mu_R$) and factorisation ($\mu_F$) scales, which influence the running of the strong coupling constant and the PDFs. Uncertainties arising from missing higher-order corrections in this parametrisation are estimated by varying $\mu_R$ and $\mu_F$ by factors of 2 and 0.5 in the calculation. These uncertainties are included for diboson, $tW$, $t\bar{t}$, $t\bar{t}H$, $t\bar{t}W$, $t\bar{t}Z$, $tZq$, and $Z$ + jets events.

- Uncertainties in the modelling of FSR and ISR are covered using a modified nominal sample, generated by using adjusted parameters in the simulation. For FSR variations, an altered value of $\alpha_S$ is used in the simulation, while inital state radiation (ISR) variations are estimated by adjusting the VAR3C eigentune of the A14 tune within the nominal POWHEG + PYTHIA 8 simulation [160]. These uncertainties are included for $tW$, $t\bar{t}$, $t\bar{t}H$, $t\bar{t}W$, $tZq$, and $t$-channel events.

- Uncertainties associated with the matrix element calculation and the parton shower modelling are estimated by comparing the nominal sample to an additional sample simulated using an alternative MC generator. These uncertainties are included for diboson, $tHq$, $tWH$, $tW$, $t\bar{t}$, $t\bar{t}H$, $t\bar{t}W$, $t\bar{t}Z$ and $t$-channel events.

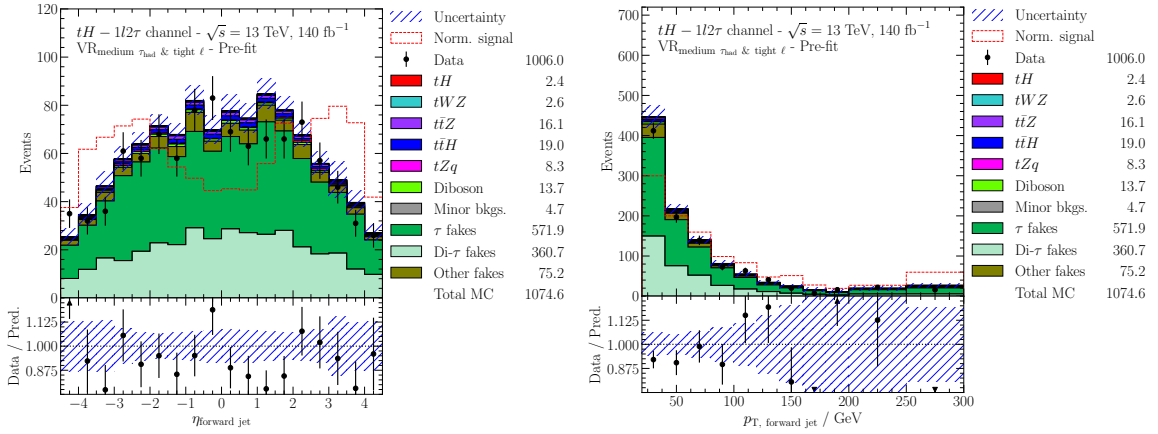
- Uncertainties connected to the used PDFs sets are assessed by comparing to results obtained with alternative sets. These uncertainties are included for diboson, $tHq$, $tWH$, $tW$, $t\bar{t}$, $t\bar{t}H$, $t\bar{t}W$, $t\bar{t}Z$ and $tZq$ events.

- In the case of $t\bar{t}$ events, an additional uncertainty arising from the selection of the $h_{damp}$ parameter must be taken into account. Within POWHEG, this parameter regulates the high-$p_T$ radiation against which the $t\bar{t}$ system recoils [161]. To evaluate this uncertainty, an alternative simulation with a doubled $h_{damp}$ parameter is used.

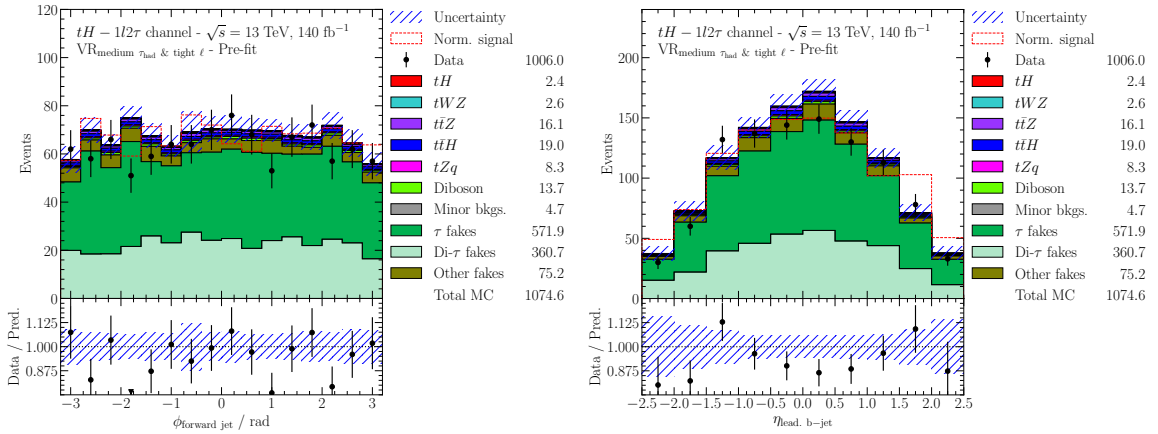
# APPENDIX E

# Neural network input variables

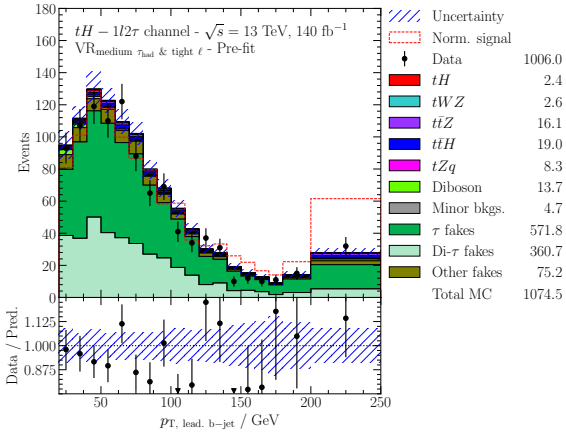
Figures E.1 and E.2 display all CNN input variables based on the tight preselection in the $1\ell + 2\tau_{\text{had}}$ channel and the $2\ell\,\text{SS} + 1\tau_{\text{had}}$ channel. In addition the responses of all three output nodes are shown. A comprehensible definition of each variable is given in Section 8.2.1.



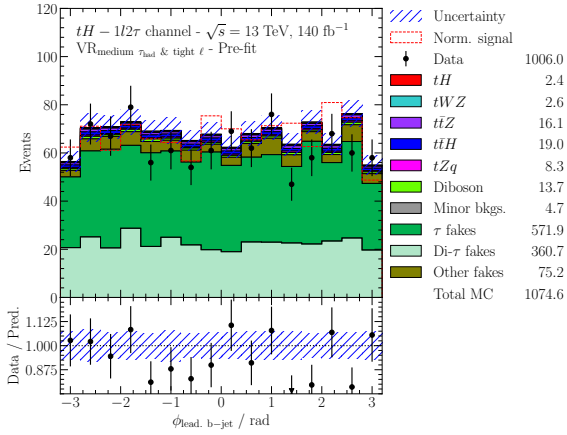

$\eta_{\text{forward jet}}$ $p_{\text{T, forward jet}}$



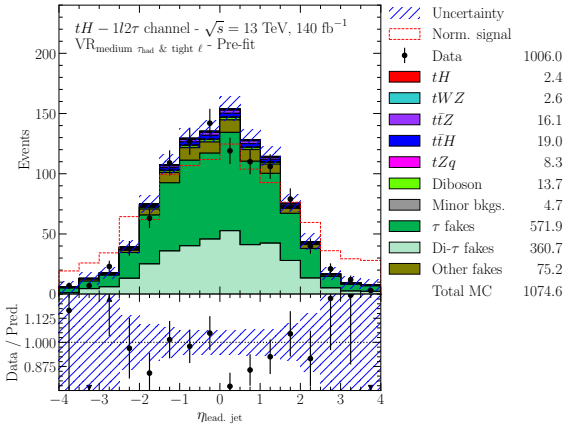

$\phi_{\text{forward jet}}$ $\eta_{\text{lead. } b\text{-jet}}$



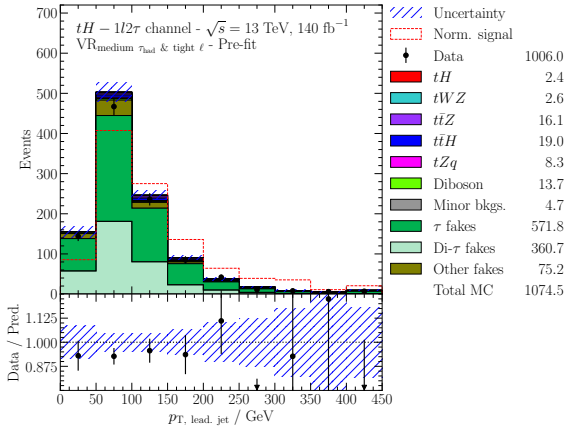

$p_{\text{T, lead. } b\text{-jet}}$.



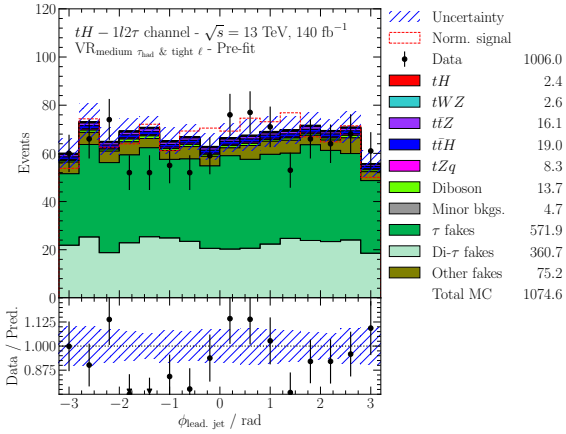

$\phi_{\text{lead. } b\text{-jet}}$.



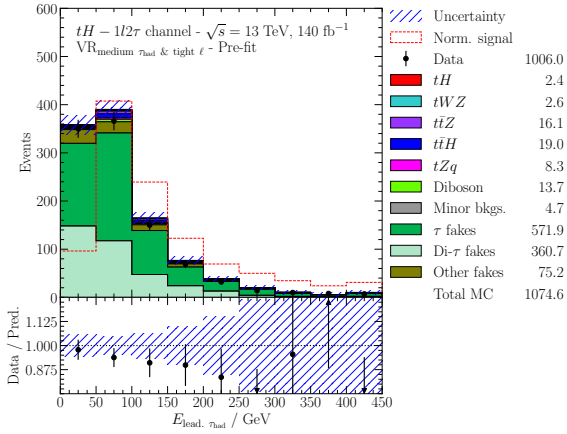

$\eta_{\text{lead. jet}}$.



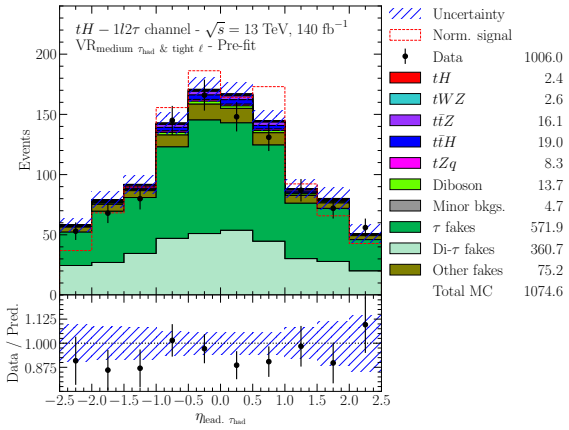

$p_{\text{T, lead. jet}}$.



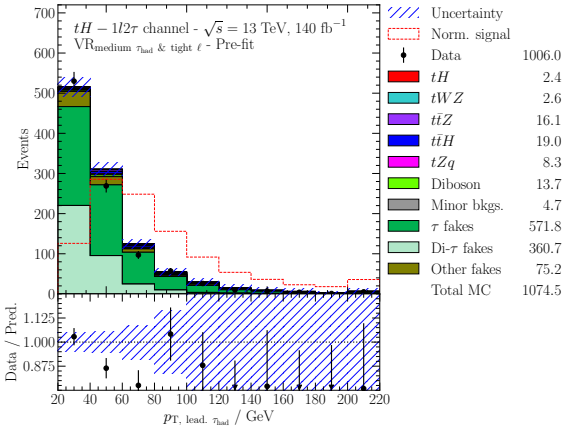

$\phi_{\text{lead. jet}}$.



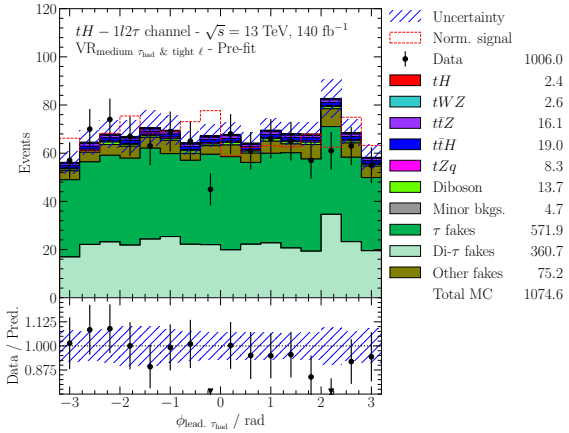

$E_{\text{lead. } \tau_{\text{had}}}$.



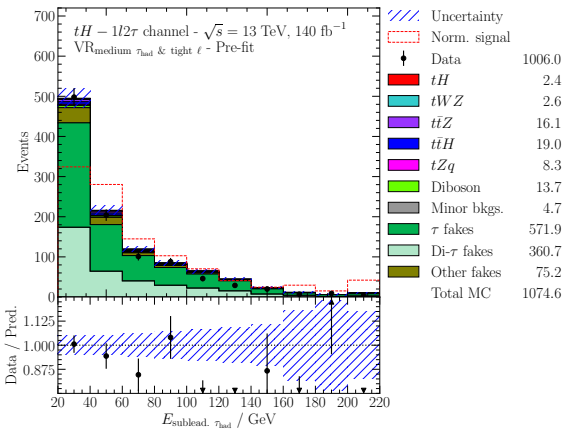

$\eta_{\text{lead. } \tau_{\text{had}}}$.



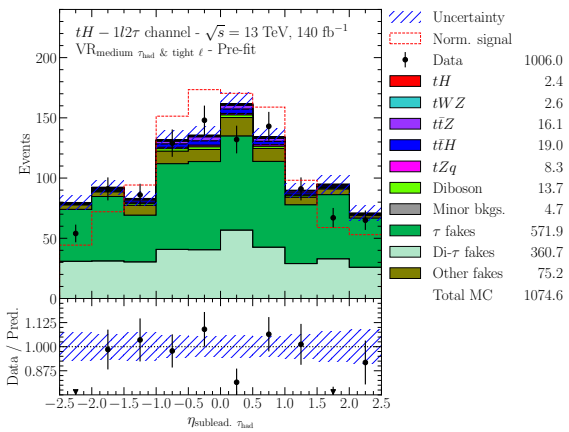

$p_{\text{T, lead. } \tau_{\text{had}}}$.



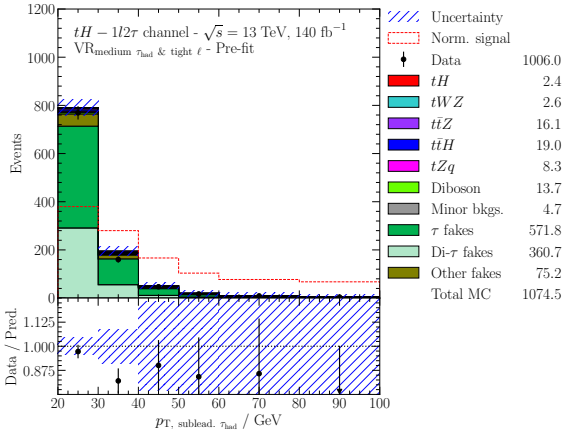

$\phi_{\text{lead. } \tau_{\text{had}}}$.



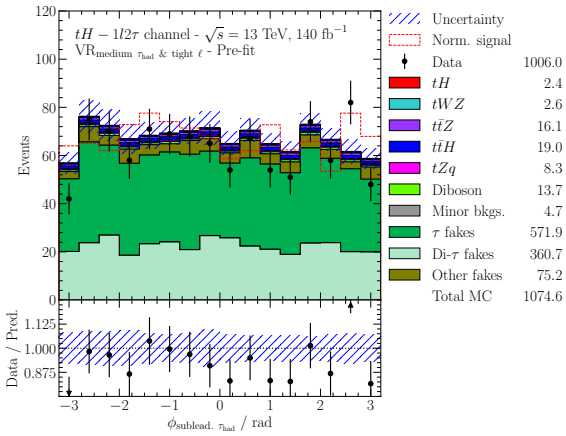

$E_{\text{sublead. } \tau_{\text{had}}}$.



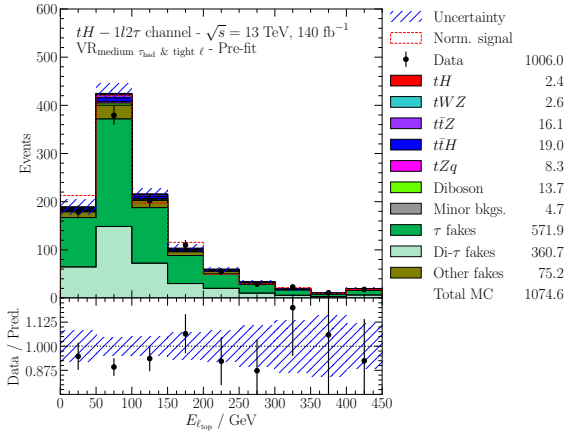

$\eta_{\text{sublead. } \tau_{\text{had}}}$.



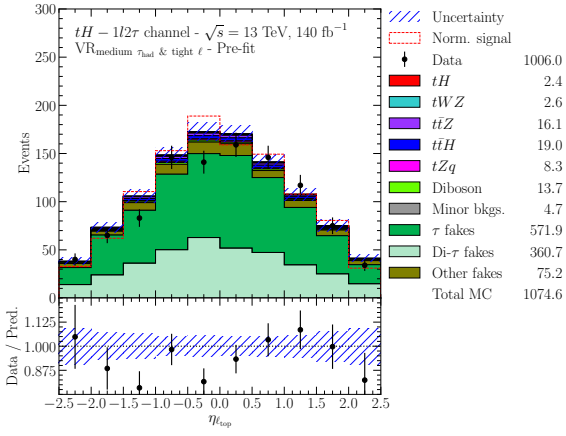

$p_{\text{T, sublead. } \tau_{\text{had}}}$.



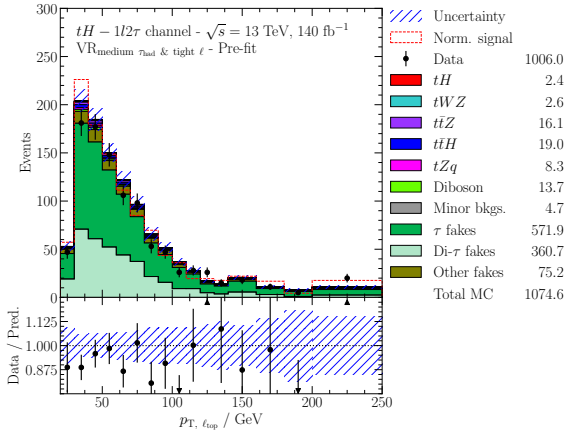

$\phi_{\text{sublead. } \tau_{\text{had}}}$.



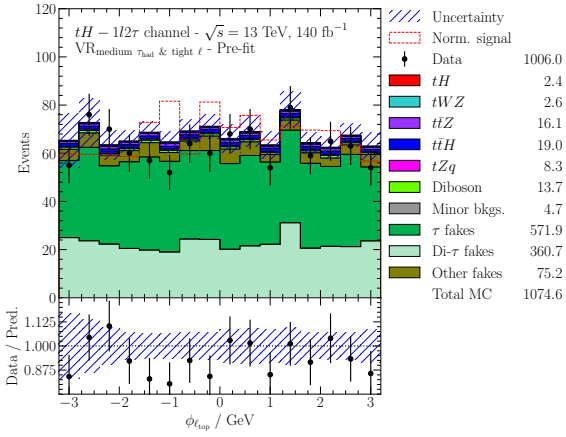

$E_{\ell_{\text{top}}}$.



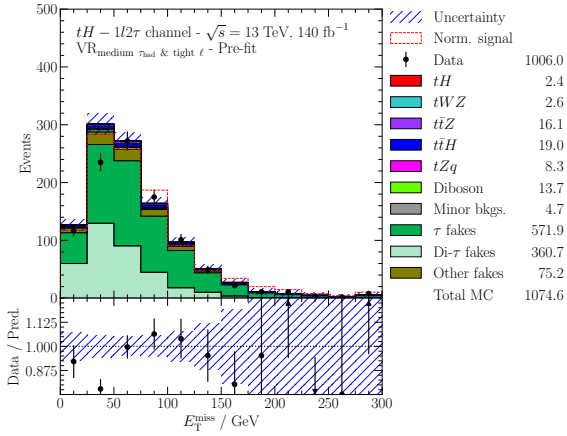

$\eta_{\ell_{\text{top}}}$.



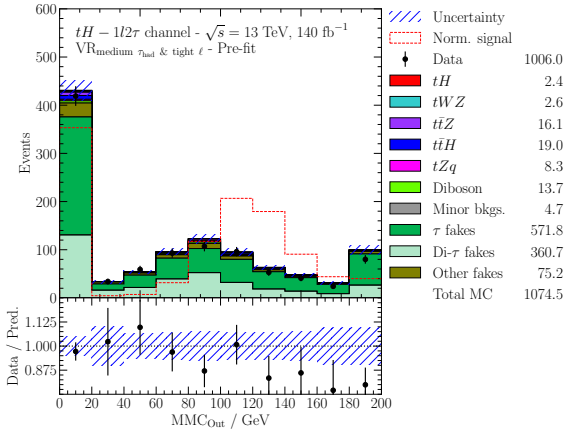

$p_{\text{T, } \ell_{\text{top}}}$.



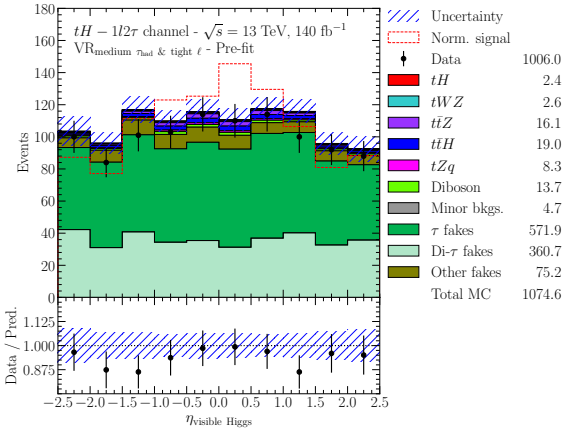

$\phi_{\ell_{\text{top}}}$.



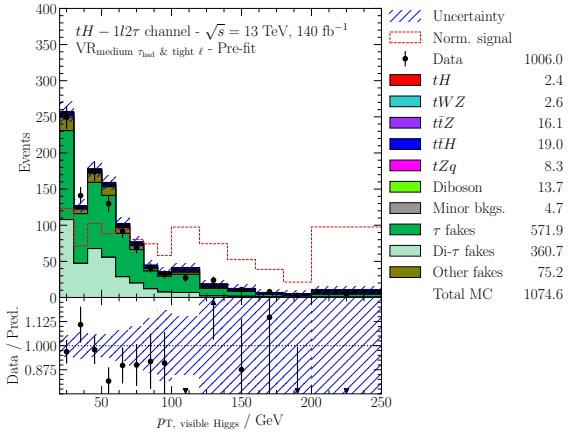

$E_{\text{T}}^{\text{miss}}$.



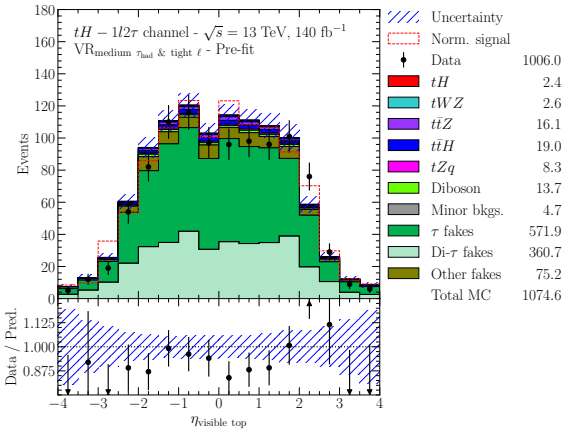

MMC$_{\text{Out}}$.



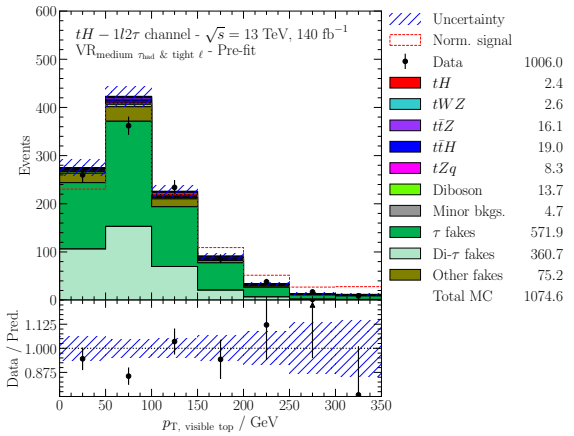

$\eta_{\text{visible. Higgs}}$.



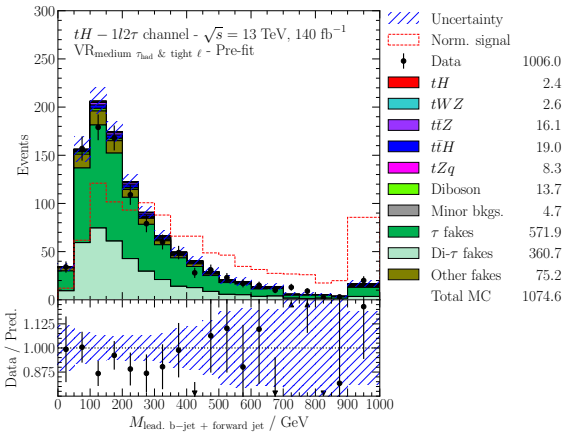

$p_{\text{T, visible Higgs}}$.




$\eta_{\text{visible. top}}$.




$p_{\text{T, visible top}}$.




$M_{\text{lead. } b\text{-jet + forward jet}}$.

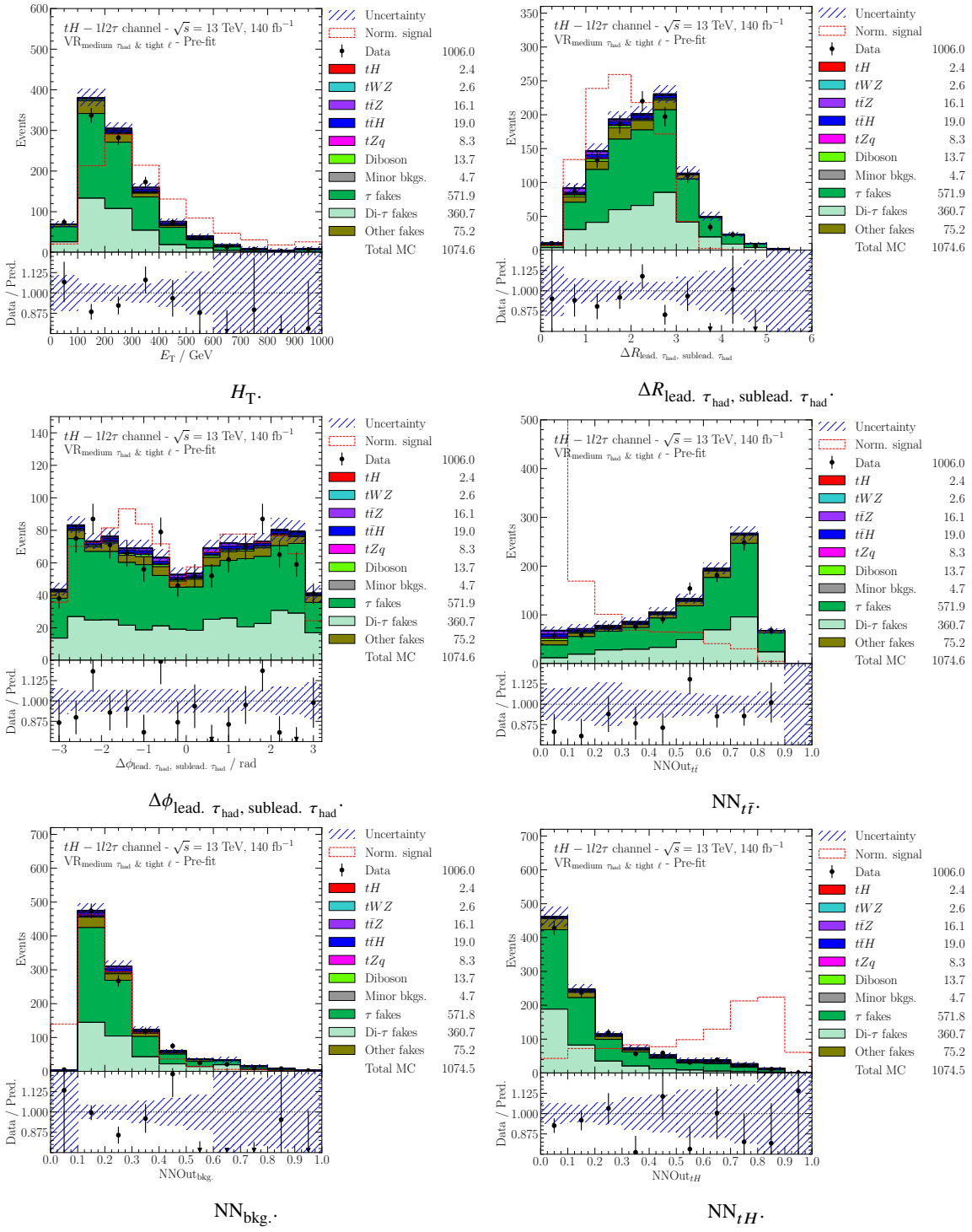
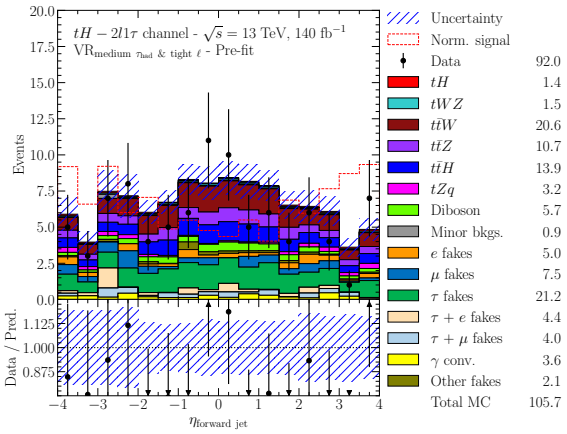
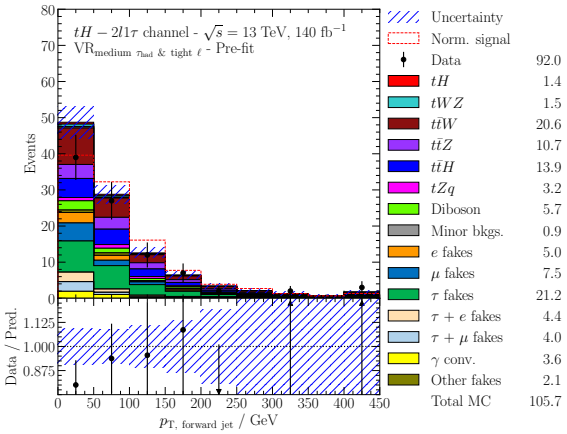


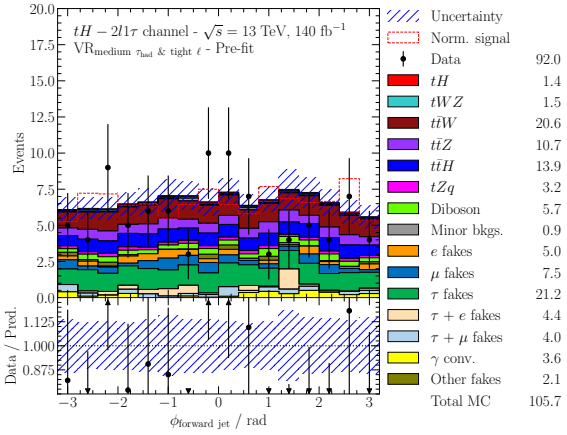

Figure E.1: All variables used as input to the CNN along with the resulting network response in the $1\ell + 2\tau_{\text{had}}$ channel. The lower ratio plot indicates the agreement between real data and simulated MC events.



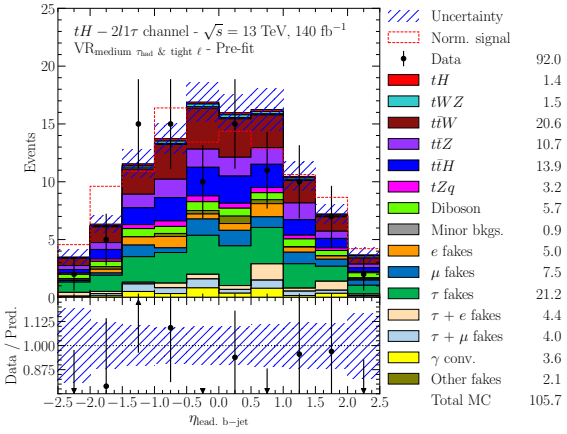

$\eta_{\text{forward jet}}$·



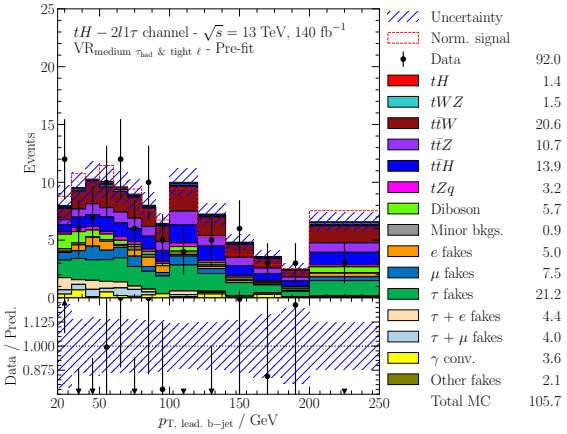

$p_{\text{T, forward jet}}$·



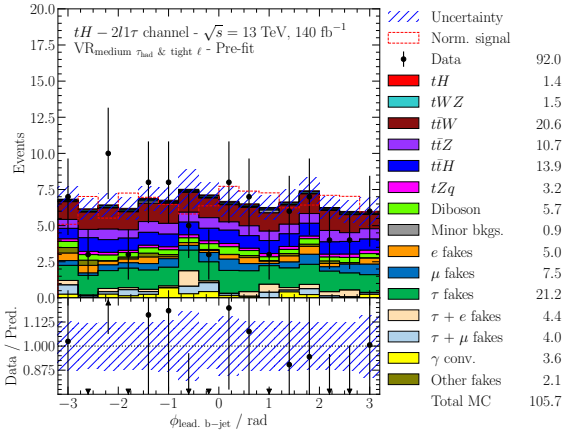

$\phi_{\text{forward jet}}$·



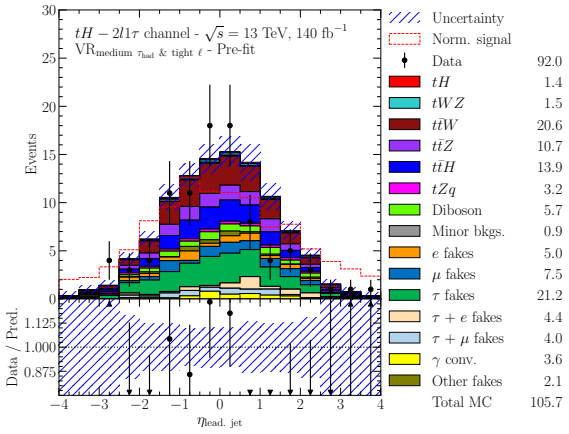

$\eta_{\text{lead. } b\text{-jet}}$·



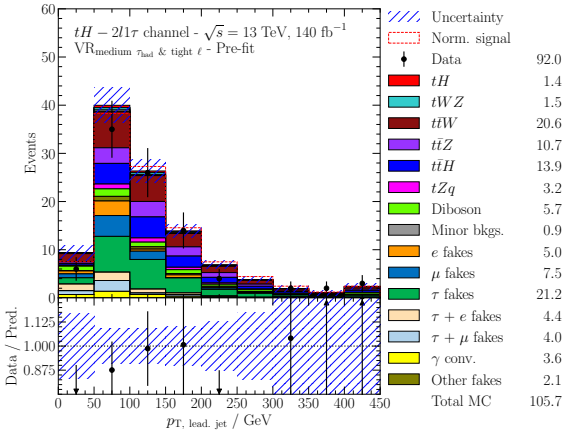

$p_{\text{T, lead. } b\text{-jet}}$·



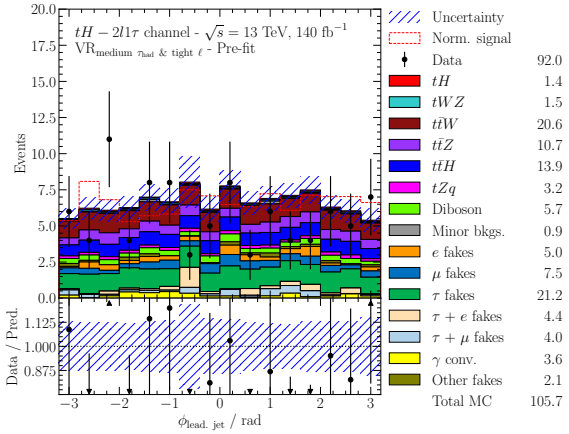

$\phi_{\text{lead. } b\text{-jet}}$·



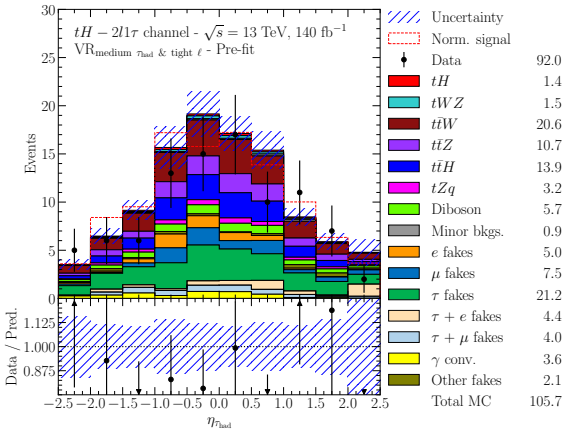

$\eta_{\text{lead. jet}}$.



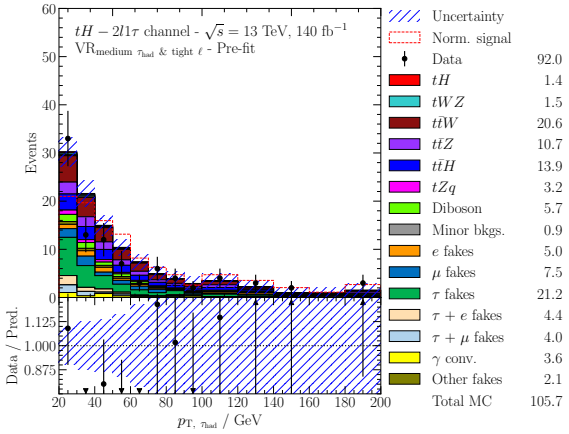

$p_{\text{T, lead. jet}}$.



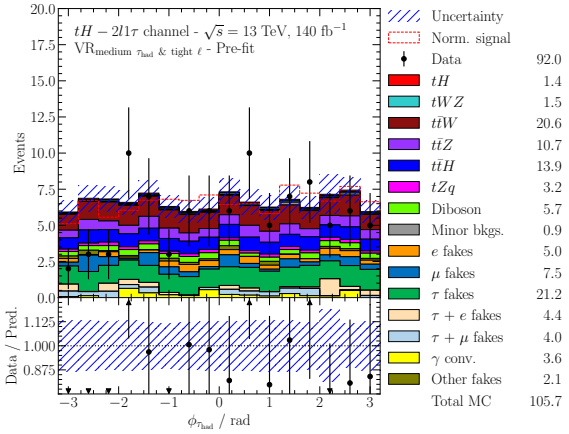

$\phi_{\text{lead. jet}}$.



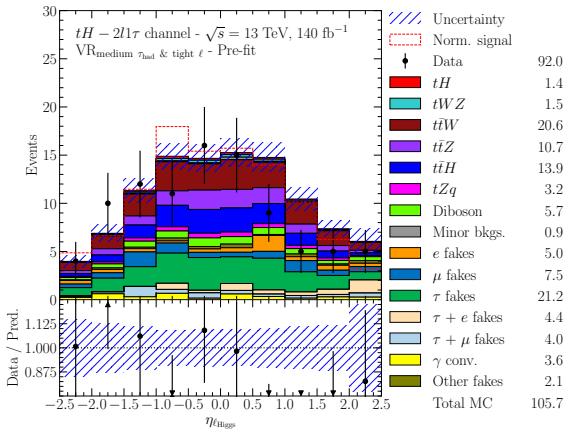

$\eta_{\tau_{\text{had}}}$.



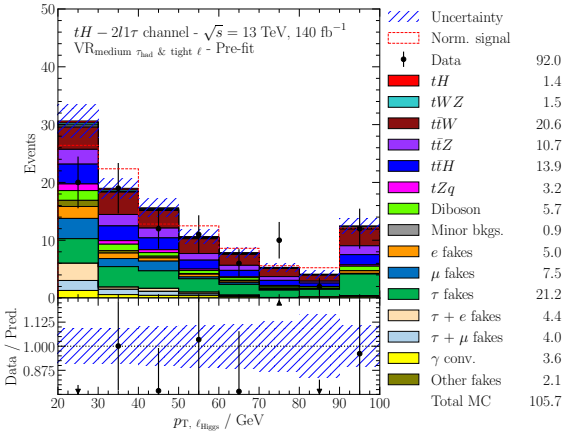

$p_{\text{T, } \tau_{\text{had}}}$.



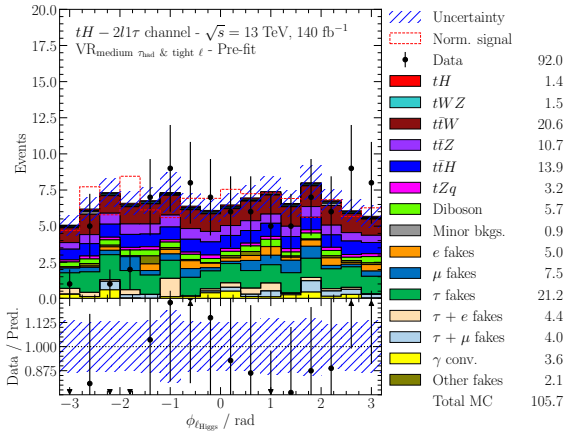

$\phi_{\tau_{\text{had}}}$.



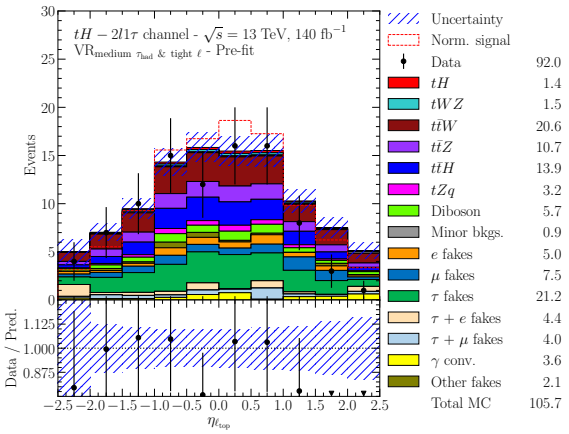

$\eta_{\ell_{\text{Higgs}}}$.



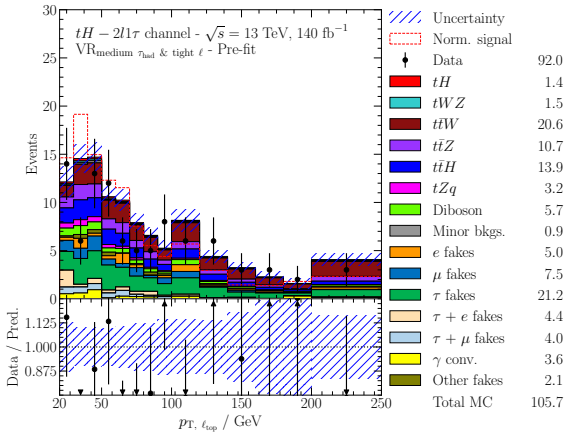

$p_{\text{T, }\ell_{\text{Higgs}}}$.



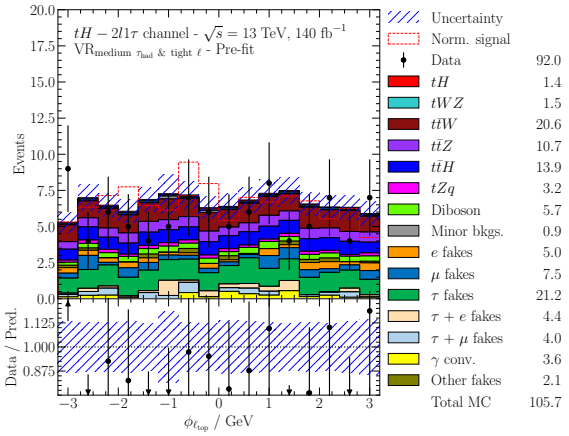

$\phi_{\ell_{\text{Higgs}}}$.



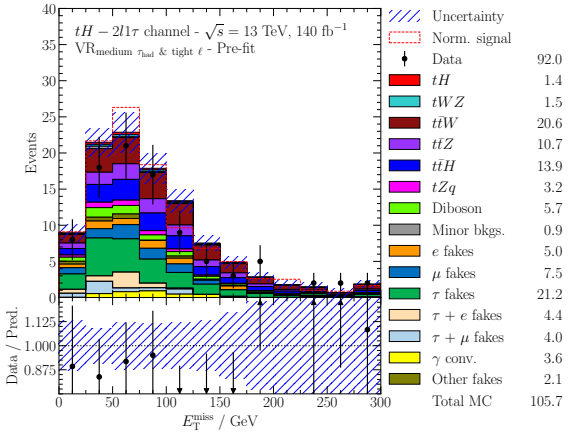

$\eta_{\ell_{\text{top}}}$.



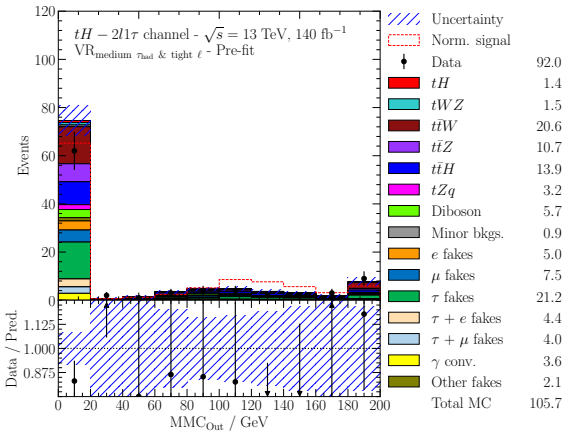

$p_{\text{T, }\ell_{\text{top}}}$.



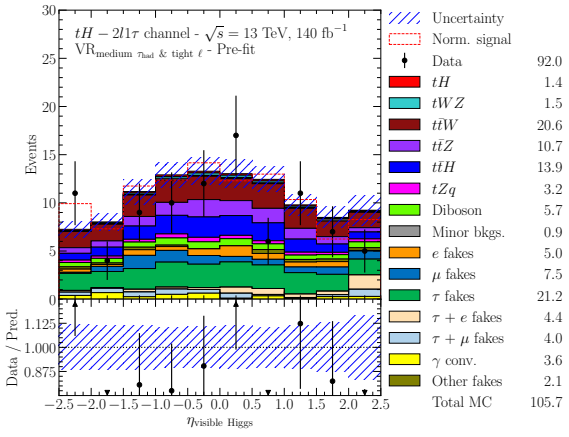

$\phi_{\ell_{\text{top}}}$.



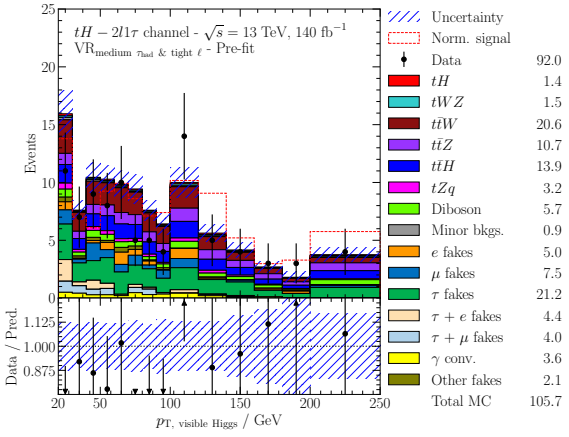

$E_{\text{T}}^{\text{miss}}$.



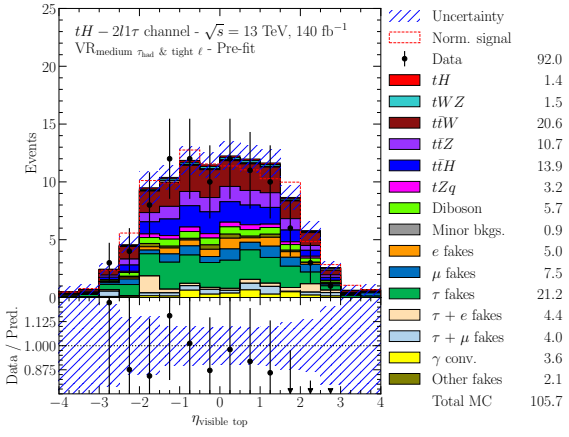

MMC$_{\text{Out}}$.



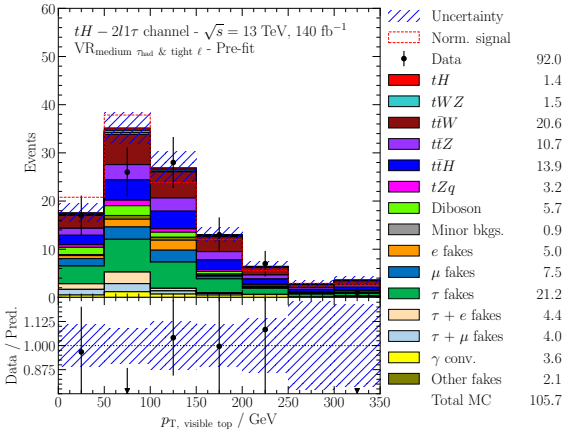

$\eta_{\text{visible. Higgs}}$.



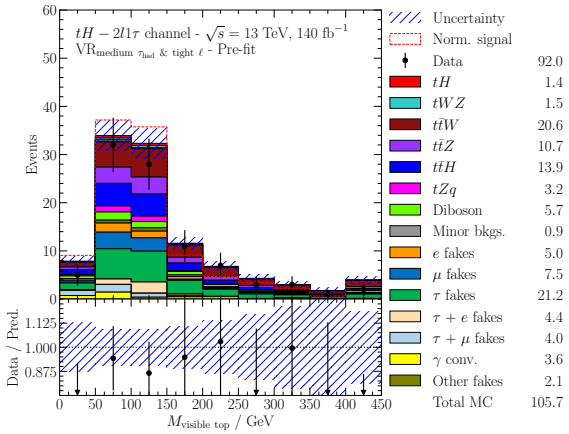

$p_{\text{T, visible Higgs}}$.



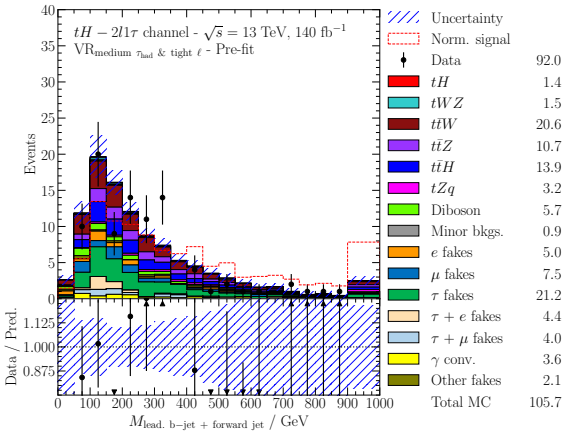

$\eta_{\text{visible. top}}$.



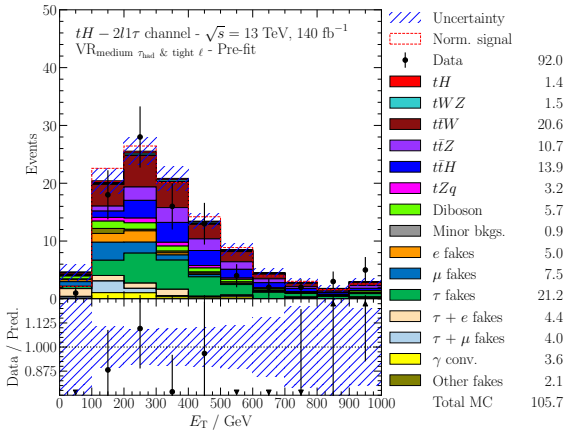

$p_{\text{T, visible top}}$.



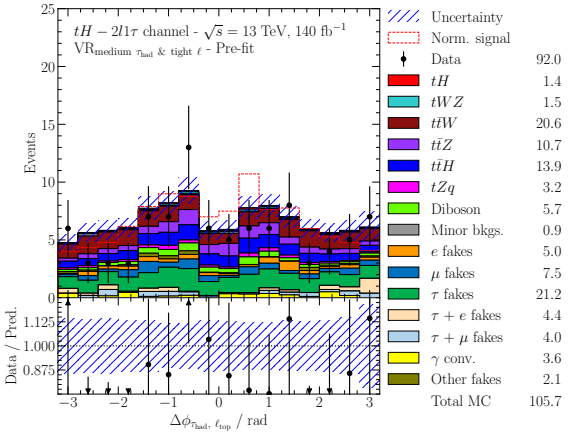

$M_{\text{visible top}}$.



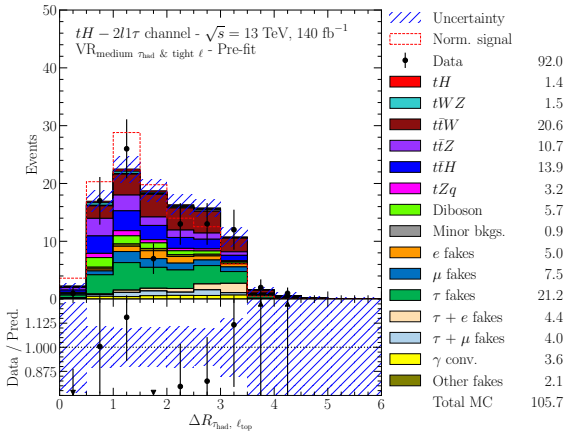

$M_{\text{lead. } b\text{-jet + forward jet}}$.



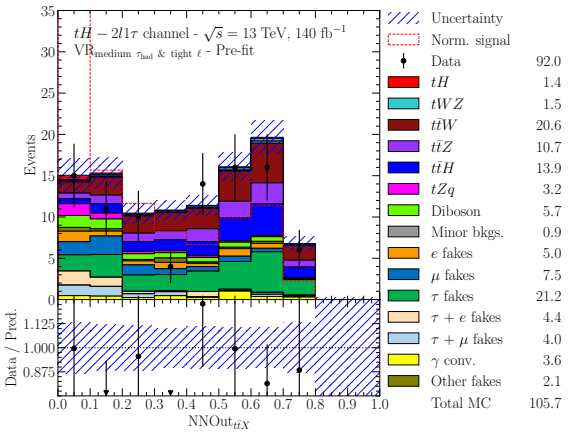

$H_{\text{T}}$.



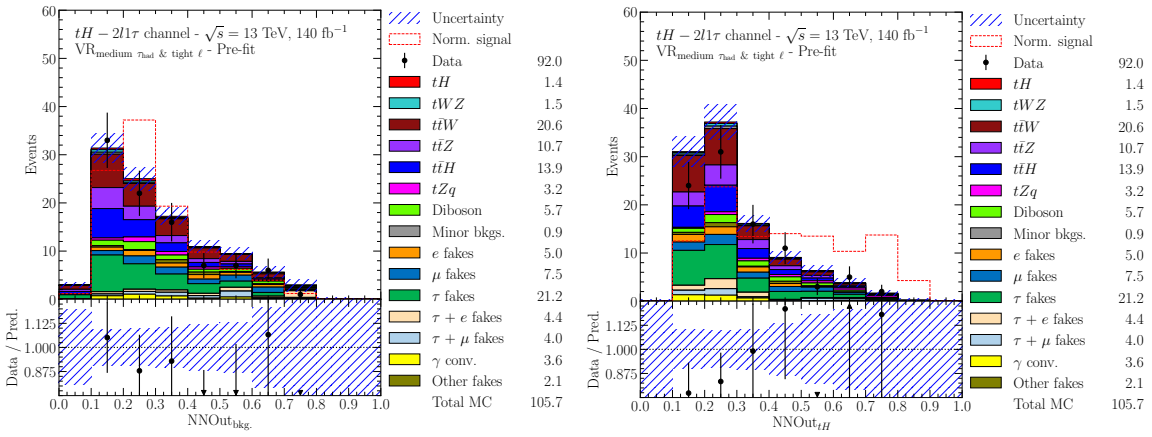

$\Delta\phi_{\tau_{\text{had}}, \ell_{\text{Higgs}}}$.




$\Delta R_{\tau_{\text{had}}, \ell_{\text{Higgs}}}$.




NN$_{t\bar{t}X}$.



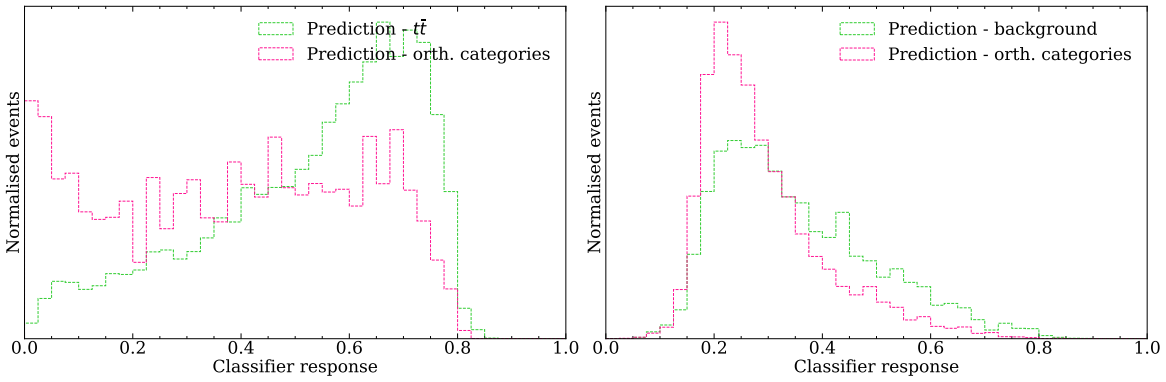

$\text{NN}_{\text{bkg.}}$ $\text{NN}_{tH}$

Figure E.2: All variables used as input to the CNN along with the resulting network response in the $2\ell\,\text{SS} + 1\tau_{\text{had}}$ channel. The lower ratio plot indicates the agreement between real data and simulated MC events.

APPENDIX F

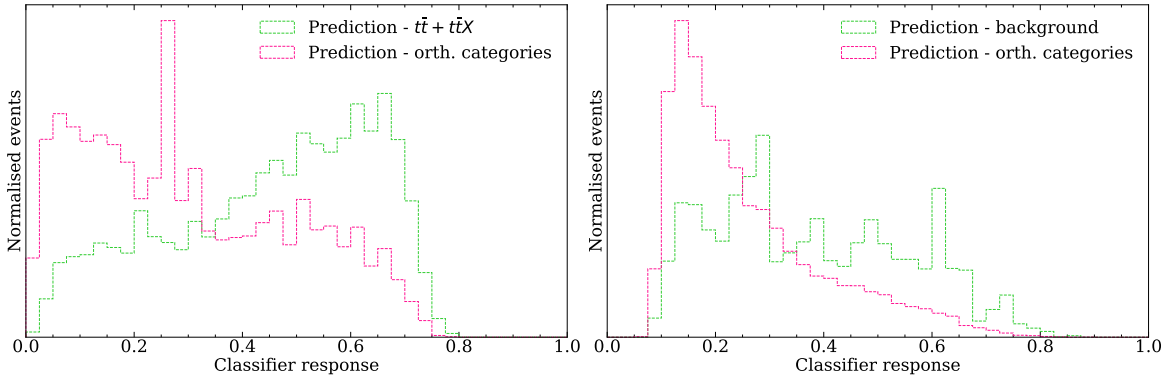
# Supplementary neural network results

This appendix summarises additional results of the signal and background separation in the $1\ell + 2\tau_{\text{had}}$ channel and the $2\ell\,\text{SS} + 1\tau_{\text{had}}$ channel. The response distributions corresponding to the non-$tH$ output nodes are shown in Figures F.1 and F.2. To obtain this response each model's outputs based on the respective prediction folds are merged. Furthermore, Figures F.3 to F.6 illustrate the training and validation ROC curve corresponding to these two output nodes for all five models. Lastly, Figures F.7 and F.8 display the remaining ROC curve for all three output nodes determined via each fold's prediction sample.



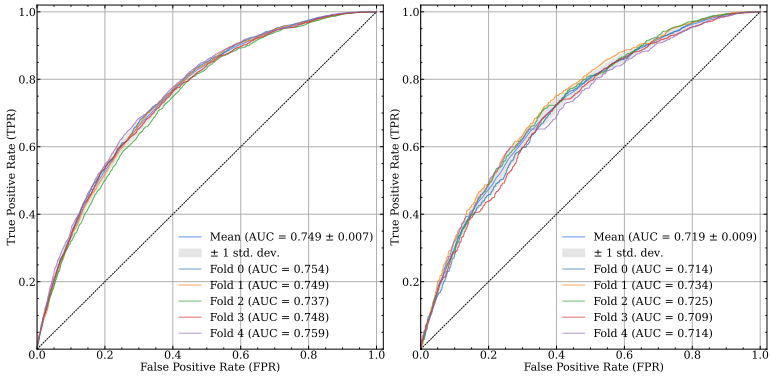

(a) (b)

Figure F.1: CNN response of (a) the $t\bar{t}$ output node and (b) the background output node based on the prediction dataset in the $1\ell + 2\tau_{\text{had}}$ channel. All remaining processes are combined and termed orthogonal categories.



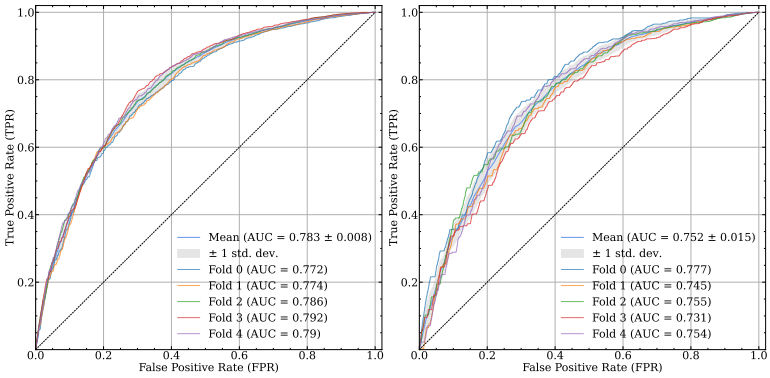

(a) (b)

Figure F.2: CNN response of (a) the $t\bar{t}$ + $t\bar{t}X$ output node and (b) the background output node based on the prediction dataset in the $2\ell$ SS + $1\tau_{\text{had}}$ channel. All remaining processes are combined and termed orthogonal categories.



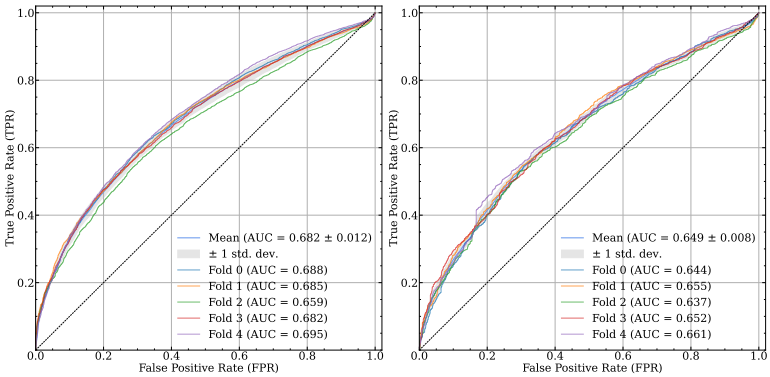

(a) (b)

Figure F.3: ROC curve corresponding to each model's $t\bar{t}$ output node response based on (a) training data and (b) validation data in the $1\ell + 2\tau_{\text{had}}$ channel. The average across all folds is shown along with its standard deviation.



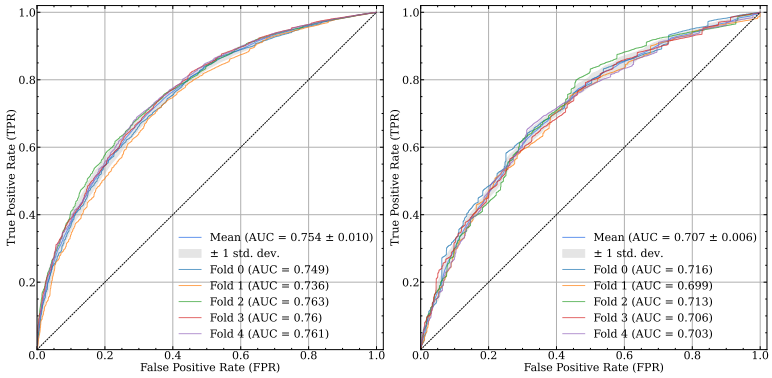

(a) (b)

Figure F.4: ROC curve corresponding to each model's $t\bar{t}$ + $t\bar{t}X$ output node response based on (a) training data and (b) validation data in the $2\ell$ SS + $1\tau_{\text{had}}$ channel. The average across all folds is shown along with its standard deviation.



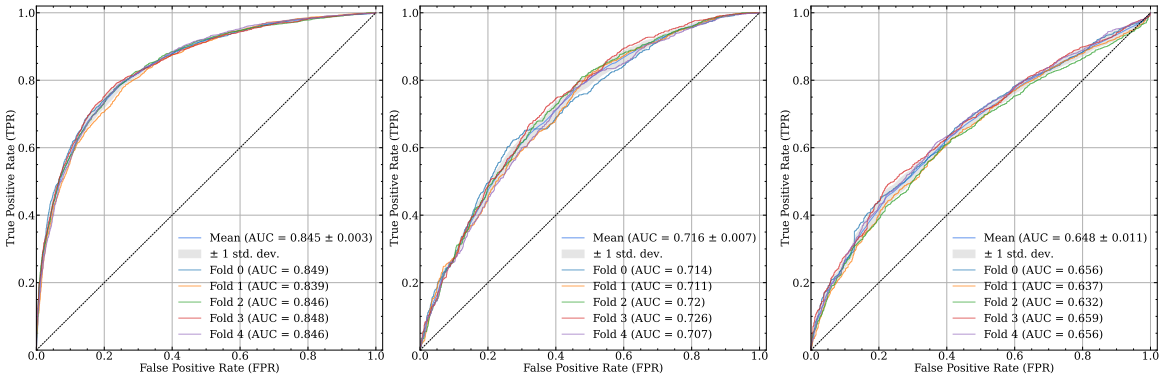

(a) (b)

Figure F.5: ROC curve corresponding to each model's background output node response based on (a) training data and (b) validation data in the $1\ell + 2\tau_{\text{had}}$ channel. The average across all folds is shown along with its standard deviation.



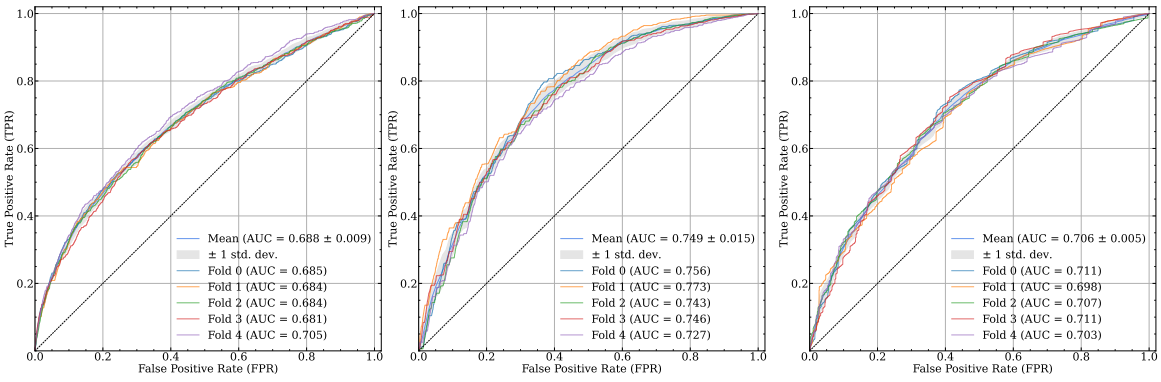

(a) (b)

Figure F.6: ROC curve corresponding to each model's background output node response based on (a) training data and (b) validation data in the $2\ell\,\mathrm{SS} + 1\tau_{\mathrm{had}}$ channel. The average across all folds is shown along with its standard deviation.




(a) (b) (c)

Figure F.7: ROC curve corresponding to each model's (a) $tH$, (b) $t\bar{t}$ and (c) background output node response based on the prediction data in the $1\ell + 2\tau_{\mathrm{had}}$ channel. The average across all folds is shown along with its standard deviation.




(a) (b) (c)

Figure F.8: ROC curve corresponding to each model's (a) $tH$, (b) $t\bar{t} + t\bar{t}X$ and (c) background output node response based on the prediction data in the $2\ell$ SS + $1\tau_{\text{had}}$ channel. The average across all folds is shown along with its standard deviation.

APPENDIX G

# Supporting background estimation studies

This appendix provides a detailed breakdown of all regions used in the profile likelihood fit into their individual processes. Small processes are not merged and instead listed individually. Table G.1 shows the results corresponding to the $1\ell + 2\tau_{\text{had}}$ channel while Table G.2 and Table G.3 summarise those of the $2\ell + 1\tau_{\text{had}}$ channel. In addition, Figures G.1 to G.3 show the composition of all processes that enter the tight preselection region in the different channels.

Table G.1: Weighted event yields for all regions used in the $1\ell + 2\tau_{\text{had}}$ channel's profile likelihood fit.

| $1\ell + 2\tau_{\text{had}}$ | Weighted yields | | | | |
|---|---|---|---|---|---|
| Process | SR | VR | $\text{CR}_{\text{fake-}\tau,p_{\text{T},1}}$ | $\text{CR}_{\text{fake-}\tau,p_{\text{T},3}}$ | $\text{CR}_{\text{fake-}\tau,p_{\text{T},3}}$ |
| $tH$ | 1.719 ± 0.044 | 0.6956 ± 0.0003 | 0.13 ± 0.01 | 0.254 ± 0.017 | 0.226 ± 0.016 |
| $tWZ$ | 1.1969 ± 0.0002 | 1.3865 ± 0.0021 | 0.1472 ± 0.0069 | 0.2770 ± 0.0094 | 0.2304 ± 0.0009 |
| $t\bar{t}$ | – | – | – | – | – |
| $t\bar{t}W$ | 1.078 ± 0.086 | 3.50 ± 0.13 | 0.366 ± 0.042 | 0.552 ± 0.057 | 0.403 ± 0.043 |
| $t\bar{t}Z$ | 7.78 ± 0.27 | 8.53 ± 0.27 | 0.789 ± 0.081 | 1.58 ± 0.11 | 1.12 ± 0.11 |
| $t\bar{t}H$ | 10.08 ± 0.12 | 9.20 ± 0.11 | 0.992 ± 0.037 | 1.9575 ± 0.0051 | 1.662 ± 0.047 |
| $tZq$ | 4.90 ± 0.12 | 3.51 ± 0.10 | 0.646 ± 0.046 | 1.00 ± 0.06 | 0.679 ± 0.042 |
| $tW$ | – | – | – | – | – |
| $Z$ + jets | – | – | – | – | – |
| $t$-channel | – | – | – | – | – |
| $W$ + jets | – | – | – | – | – |
| Diboson | 5.32 ± 0.24 | 8.56 ± 0.36 | 1.38 ± 0.31 | 1.34 ± 0.17 | 1.09 ± 0.13 |
| Minor bkg. | 0.0293 ± 0.0054 | 0.0704 ± 0.0001 | 0.0031 ± 0.0013 | 0.0083 ± 0.0022 | 0.0121 ± 0.0068 |
| Fake-$e$ | 0.088 ± 0.027 | 0.52 ± 0.21 | 0.17 ± 0.14 | 0.055 ± 0.018 | 0.029 ± 0.014 |
| Fake-$\mu$ | 0.042 ± 0.015 | 0.23 ± 0.15 | 0.022 ± 0.017 | 0.24 ± 0.18 | 0.0176 ± 0.0082 |
| Charge-misident. | 0.0046 ± 0.0002 | 0.0103 ± 0.0059 | 0.0007 ± 0.0005 | 0.0002 ± 0.0002 | 0.0004 ± 0.0003 |
| $\gamma$-conversion | 0.042 ± 0.015 | 0.09 ± 0.03 | 0.099 ± 0.089 | 0.0077 ± 0.0021 | 0.0006 ± 0.0022 |
| Fake-$\tau$ | 73.1 ± 3.0 | 468.0 ± 7.9 | 281.4 ± 6.4 | 167.55 ± 5.05 | 45.9 ± 2.4 |
| Fake-$\tau$ + fake-$e$ | 0.045 ± 0.020 | 0.54 ± 0.25 | 0.37 ± 0.18 | −0.060 ± 0.097 | 0.036 ± 0.016 |
| Fake-$\tau$ + fake-$\mu$ | 0.0136 ± 0.0074 | 0.28 ± 0.17 | 0.095 ± 0.042 | 0.035 ± 0.030 | 0.14 ± 0.13 |
| Fake-$\tau$ + charge-misident. | 0.0026 ± 0.0018 | 0.16 ± 0.14 | 0.096 ± 0.095 | 0.0044 ± 0.0038 | – |
| Fake-$\tau$ + $\gamma$-conversion | 0.0183 ± 0.0094 | 1.05 ± 0.72 | 0.89 ± 0.37 | 0.0121 ± 0.0056 | 0.030 ± 0.026 |
| fake-di-$\tau$ | 40.1 ± 2.5 | 392.5 ± 9.6 | 334.2 ± 8.9 | 233.0 ± 7.6 | 51.6 ± 2.8 |
| fake-di-$\tau$ + fake-$e$ | 0.0001 ± 0.0036 | 0.12 ± 0.10 | 0.15 ± 0.15 | 0.56 ± 0.35 | – |
| fake-di-$\tau$ + fake-$\mu$ | 0.10 ± 0.10 | 0.0081 ± 0.0058 | 0.36 ± 0.36 | 0.10 ± 0.10 | – |
| fake-di-$\tau$ + charge-misident. | 0.0059 ± 0.0059 | 0.021 ± 0.018 | 0.0050 ± 0.0036 | −0.12 ± 0.12 | – |
| fake-di-$\tau$ + $\gamma$-conversion | −0.010 ± 0.013 | 0.93 ± 0.37 | 0.41 ± 0.23 | 0.15 ± 0.15 | 0.0046 ± 0.0046 |
| Other fakes | 0.029 ± 0.011 | 0.52 ± 0.45 | 0.0575 ± 0.0034 | 0.11 ± 0.15 | 0.0072 ± 0.0038 |
| Total | 145.6 ± 4.0 | 901 ± 13 | 623 ± 11 | 408.6 ± 9.1 | 103.2 ± 3.7 |

Table G.2: Weighted event yields for all regions used in the $2\ell$ SS + $1\tau_{\text{had}}$ channel's profile likelihood fit.

| $2\ell$ SS + $1\tau_{\text{had}}$ | Weighted yields | | | |
|---|---|---|---|---|
| Process | SR | VR | $CR_{\text{fake-}e}$ | $CR_{\text{fake-}\mu}$ |
| $tH$ | 1.375 ± 0.041 | 0.277 ± 0.019 | 0.428 ± 0.023 | 0.358 ± 0.021 |
| $tWZ$ | 1.464 ± 0.022 | 0.3130 ± 0.0097 | 0.434 ± 0.012 | 0.320 ± 0.010 |
| $t\bar{t}$ | 0.25 ± 0.18 | 0.25 ± 0.18 | 1.52 ± 0.46 | 0.61 ± 0.28 |
| $t\bar{t}W$ | 20.56 ± 0.36 | 4.55 ± 0.15 | 3.06 ± 0.13 | 2.93 ± 0.14 |
| $t\bar{t}Z$ | 10.75 ± 0.29 | 2.04 ± 0.13 | 3.09 ± 0.17 | 2.99 ± 0.15 |
| $t\bar{t}H$ | 13.909 ± 0.127 | 2.754 ± 0.055 | 4.440 ± 0.072 | 3.557 ± 0.065 |
| $tZq$ | 3.169 ± 0.098 | 0.794 ± 0.047 | 0.996 ± 0.057 | 0.898 ± 0.053 |
| $tW$ | – | – | – | – |
| $Z$ + jets | – | – | – | – |
| Diboson | 5.73 ± 0.25 | 1.56 ± 0.20 | 1.76 ± 0.14 | 1.50 ± 0.15 |
| $W$ + jets | – | – | – | – |
| Minor bkg. | 0.379 ± 0.020 | 0.0417 ± 0.0055 | 0.114 ± 0.014 | 0.40 ± 0.32 |
| Fake-$e$ | 5.03 ± 0.72 | 150.5 ± 4.5 | 688.7 ± 9.8 | 37.34 ± 2.27 |
| Fake-$\mu$ | 7.47 ± 0.98 | 139.8 ± 4.3 | 42.8 ± 2.4 | 660.9 ± 9.5 |
| Charge-misident. | 0.0618 ± 0.0021 | 0.0230 ± 0.0092 | 0.039 ± 0.011 | 0.0092 ± 0.0039 |
| $\gamma$-conversion | 3.56 ± 0.65 | 3.34 ± 0.69 | 8.4 ± 1.1 | 2.28 ± 0.58 |
| Fake-$e$ + fake-$e$ | 0.0056 ± 0.0004 | 0.022 ± 0.025 | 0.91 ± 0.32 | – |
| Fake-$e$ + fake-$\mu$ | 0.032 ± 0.012 | 0.47 ± 0.25 | 0.55 ± 0.24 | 1.20 ± 0.39 |
| Fake-$\mu$ + fake-$\mu$ | 0.048 ± 0.022 | 0.54 ± 0.25 | – | 1.20 ± 0.40 |
| Fake-$\tau$ | 21.07 ± 0.46 | 16.10 ± 0.38 | 6.29 ± 0.94 | 4.92 ± 0.34 |
| Fake-$\tau$ + fake-$e$ | 4.4 ± 1.2 | 313.9 ± 7.2 | 415.2 ± 7.8 | 27.7 ± 1.9 |
| Fake-$\tau$ + fake-$\mu$ | 3.97 ± 0.71 | 297.3 ± 6.4 | 27.7 ± 1.9 | 376.7 ± 7.2 |
| Fake-$\tau$ + charge-misident. | 1.01 ± 0.28 | 2.45 ± 0.50 | 0.89 ± 0.35 | 0.48 ± 0.24 |
| Fake-$\tau$ + $\gamma$-conversion | 1.64 ± 0.42 | 6.10 ± 0.88 | 4.42 ± 0.75 | 1.26 ± 0.39 |
| Fake-$\tau$ + fake-$e$ + fake-$e$ | – | 0.102 ± 0.094 | 1.34 ± 0.58 | – |
| Fake-$\tau$ + fake-$e$ + fake-$\mu$ | 0.0031 ± 0.0031 | 0.47 ± 0.31 | 0.25 ± 0.16 | 0.40 ± 0.30 |
| Fake-$\tau$ + fake-$\mu$ + fake-$\mu$ | – | 0.042 ± 0.040 | – | 0.35 ± 0.24 |
| Other fakes | 2.08 ± 0.48 | 9.9 ± 1.2 | 1.683 ± 0.442 | 18.4 ± 1.7 |
| Total | 108.0 ± 2.2 | 954 ± 12 | 1 215 ± 13 | 1 147 ± 12 |

Table G.3: Weighted event yields for all regions used in the $2\ell$ OS + $1\tau_{\text{had}}$ channel's profile likelihood fit.

| $2\ell$ OS + $1\tau_{\text{had}}$ | Weighted yields | | |
|---|---|---|---|
| Process | $\text{CR}_{\text{fake-}\tau, p_{\text{T},1}}$ | $\text{CR}_{\text{fake-}\tau, p_{\text{T},3}}$ | $\text{CR}_{\text{fake-}\tau, p_{\text{T},3}}$ |
| $tH$ | 0.384 ± 0.022 | 0.451 ± 0.026 | 0.119 ± 0.014 |
| $tWZ$ | 1.788 ± 0.023 | 2.621 ± 0.029 | 0.945 ± 0.017 |
| $t\bar{t}$ | 0.11 ± 0.11 | 0.15 ± 0.15 | – |
| $t\bar{t}W$ | 11.03 ± 0.26 | 16.10 ± 0.31 | 5.80 ± 0.19 |
| $t\bar{t}Z$ | 14.12 ± 0.29 | 18.79 ± 0.34 | 5.55 ± 0.18 |
| $t\bar{t}H$ | 6.382 ± 0.071 | 9.708 ± 0.085 | 2.806 ± 0.038 |
| $tZq$ | 5.18 ± 0.13 | 5.40 ± 0.13 | 0.831 ± 0.051 |
| $tW$ | – | – | – |
| $Z$ + jets | 0.0046 ± 0.0037 | – | – |
| Diboson | 8.30 ± 0.32 | 11.33 ± 0.46 | 3.77 ± 0.21 |
| Other Higgs | 0.74 ± 0.53 | – | – |
| Minor bkg. | 0.165 ± 0.013 | 0.258 ± 0.017 | 0.0974 ± 0.0094 |
| Fake-$e$ | 1.20 ± 0.36 | 2.83 ± 0.55 | 0.72 ± 0.28 |
| Fake-$\mu$ | 1.46 ± 0.41 | 2.13 ± 0.48 | 0.37 ± 0.18 |
| Charge-misident. | 0.009 ± 0.011 | 0.0083 ± 0.0004 | 0.0025 ± 0.0019 |
| $\gamma$-conversion | 1.15 ± 0.39 | 1.46 ± 0.40 | 0.47 ± 0.21 |
| Fake-$e$ + fake-$e$ | 0.0070 ± 0.0044 | 0.0127 ± 0.0059 | 0.0024 ± 0.0024 |
| Fake-$e$ + fake-$\mu$ | 0.0179 ± 0.0007 | 0.0024 ± 0.0056 | 0.0051 ± 0.0038 |
| Fake-$\mu$ + fake-$\mu$ | 0.0068 ± 0.0049 | 0.0132 ± 0.0068 | – |
| Fake-$\tau$ | 3 711 ± 24 | 1 761 ± 16 | 179.8 ± 5.2 |
| Fake-$\tau$ + fake-$e$ | 5.45 ± 0.85 | 2.82 ± 0.60 | 0.22 ± 0.15 |
| Fake-$\tau$ + fake-$\mu$ | 3.19 ± 0.66 | 2.02 ± 0.53 | 0.26 ± 0.18 |
| Fake-$\tau$ + charge-misident. | 0.0154 ± 0.0068 | −0.0001 ± 0.0018 | −0.0007 ± 0.0007 |
| Fake-$\tau$ + $\gamma$-conversion | 11.3 ± 1.2 | 4.79 ± 0.75 | 0.27 ± 0.16 |
| Fake-$\tau$ + fake-$e$ + fake-$e$ | – | 0.0021 ± 0.0021 | – |
| Fake-$\tau$ + fake-$e$ + fake-$\mu$ | – | 0.0001 ± 0.0001 | – |
| Fake-$\tau$ + fake-$\mu$ + fake-$\mu$ | 0.031 ± 0.031 | 0.019 ± 0.019 | – |
| Other fakes | 3.09 ± 0.53 | 1.71 ± 0.36 | 0.42 ± 0.17 |
| Total | 3 786 ± 24 | 1 844 ± 16 | 202.4 ± 5.3 |


Figure G.1: Relative prompt and fake event composition of all processes entering the tight preselection in the $1\ell + 2\tau_{\text{had}}$ channel. Minor backgrounds are omitted. All different categories are ranked according to their contribution.

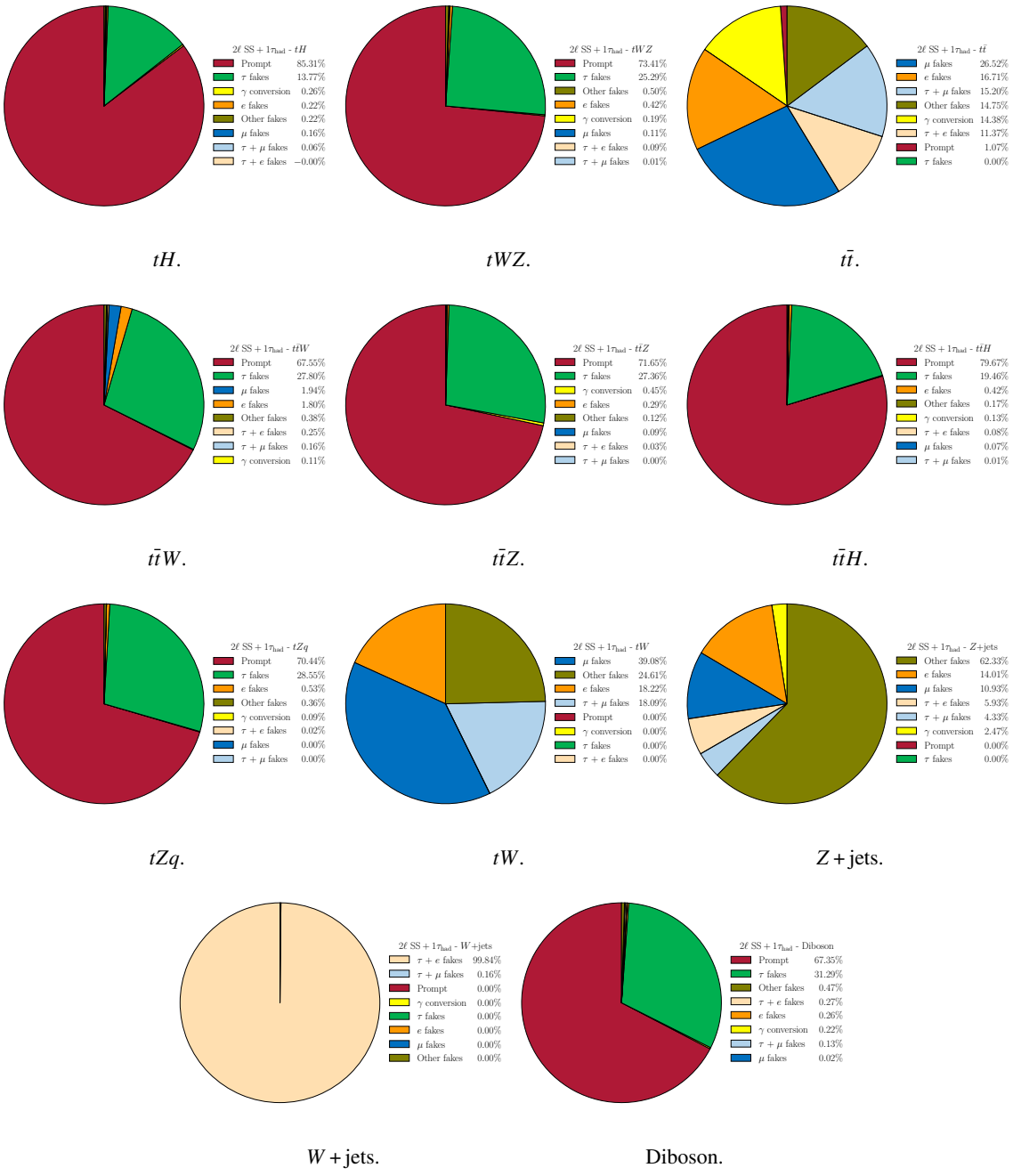


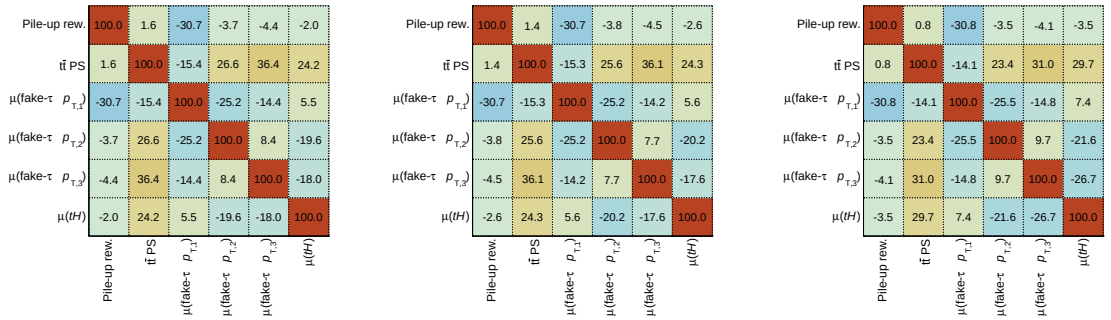

Figure G.2: Relative prompt and fake event composition of all processes entering the tight preselection in the $2\ell$ SS + $1\tau_{\text{had}}$ channel. Minor backgrounds are omitted. All different categories are ranked according to their contribution.


Figure G.3: Relative prompt and fake event composition of all processes entering the tight preselection in the $2\ell\,\mathrm{OS} + 1\tau_{\mathrm{had}}$ channel. Minor backgrounds are omitted. All different categories are ranked according to their contribution.

APPENDIX H

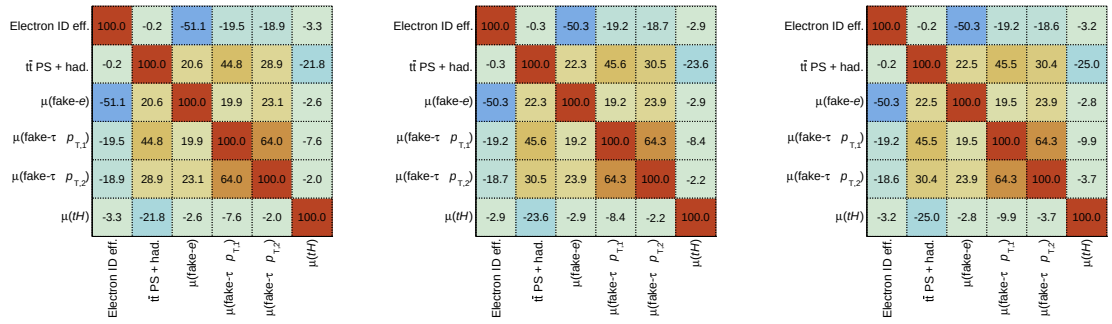
# Supplementary fit results

This appendix presents supplementary results for all fits performed in the $1\ell + 2\tau_{\text{had}}$ channel and the $2\ell + 1\tau_{\text{had}}$ channel. Summarising all studies in this appendix is supposed to avoid unnecessarily extending Chapter 10, while still providing a complete overview of the analysis. Any noteworthy or unexpected results are always discussed separately within Chapter 10.



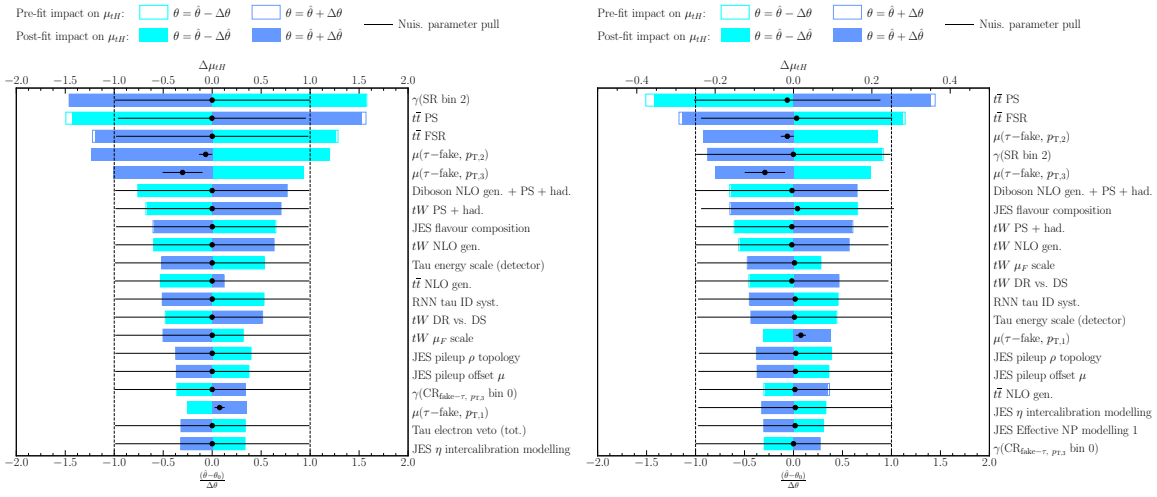

(a) (b) (c)

Figure H.1: Correlation matrix for the $1\ell + 2\tau_{\text{had}}$ channel's (a) hybrid, (b) fully unblinded and (c) fully unblinded inverted coupling fit.



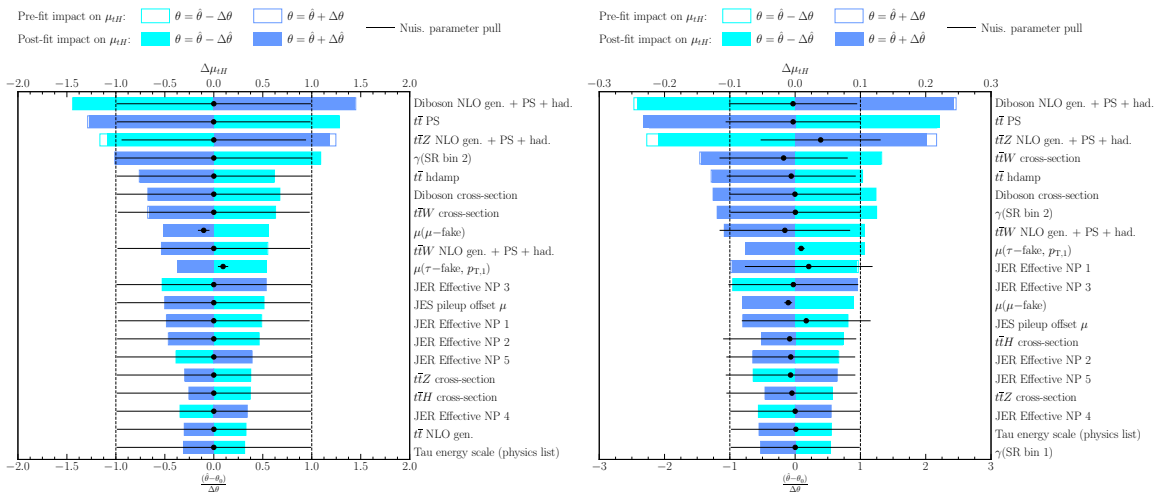

(a) (b) (c)

Figure H.2: Correlation matrix for the $2\ell + 1\tau_{\text{had}}$ channel's (a) hybrid, (b) fully unblinded and (c) fully unblinded inverted coupling fit.



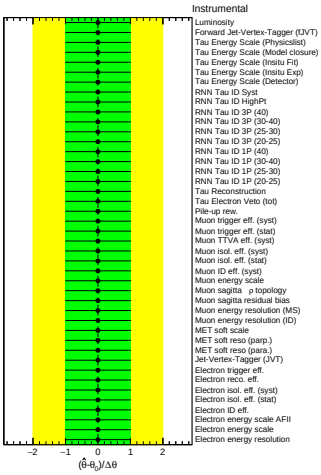

(a) (b)

Figure H.3: Nuisance parameter impact ranking for the $1\ell + 2\tau_{\text{had}}$ channel's (a) hybrid and (b) fully unblinded inverted coupling fit.

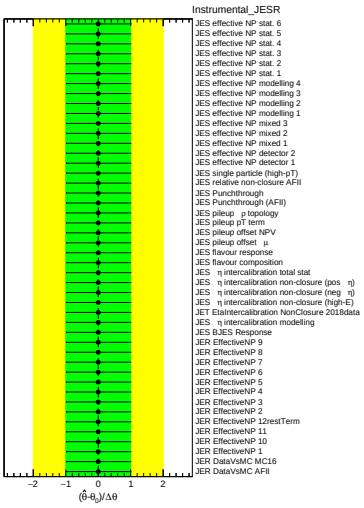



(a) (b)

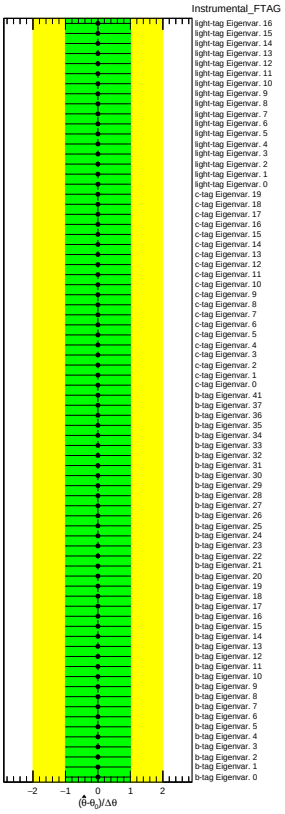
Figure H.4: Nuisance parameter impact ranking for the $2\ell + 1\tau_{\text{had}}$ channel's (a) hybrid and (b) fully unblinded inverted coupling fit.



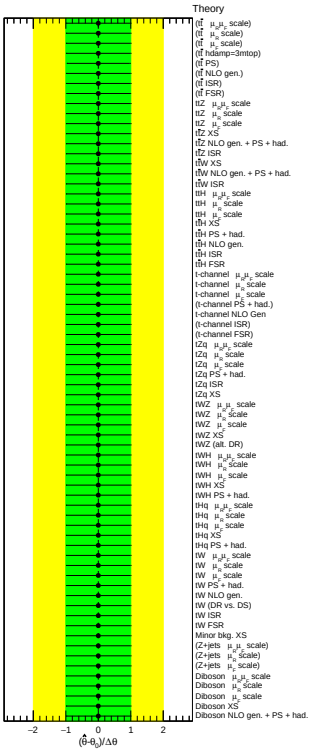

(a) Instrumental nuisance parameters.




(b) Instrumental JES/JER nuisance parameters.

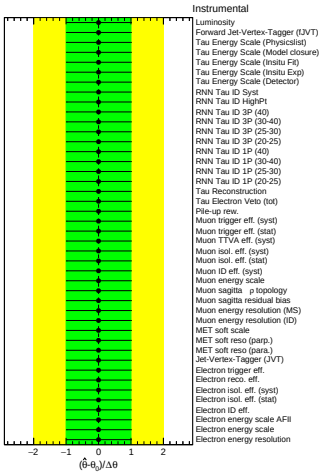



(c) Instrumental flavour tagging nuisance parameters.

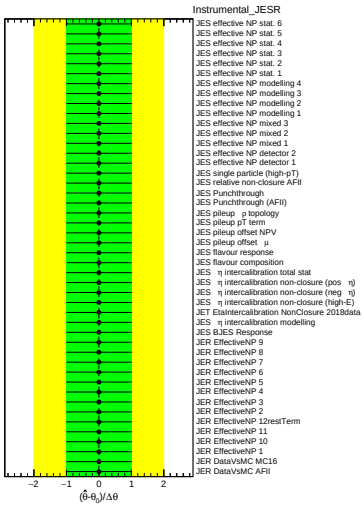



(d) Theory nuisance parameters.

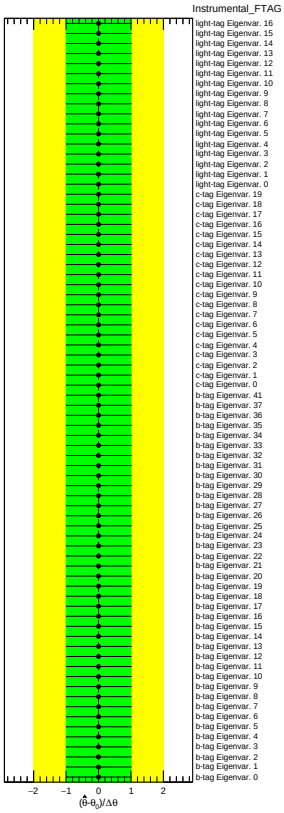
Figure H.5: Nuisance parameter constraint and pull studies in the $1\ell + 2\tau_{\text{had}}$ channel's Asimov fit.



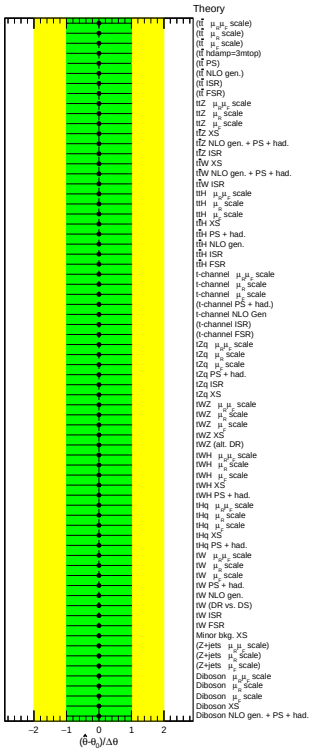

(a) Instrumental nuisance parameters.




(b) Instrumental JES/JER nuisance parameters.

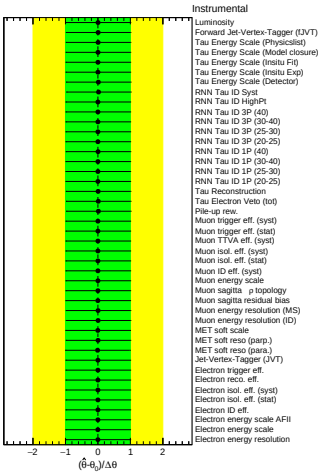



(c) Instrumental flavour tagging nuisance parameters.

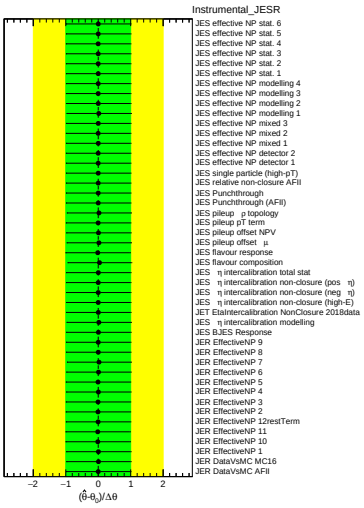



(d) Theory nuisance parameters.

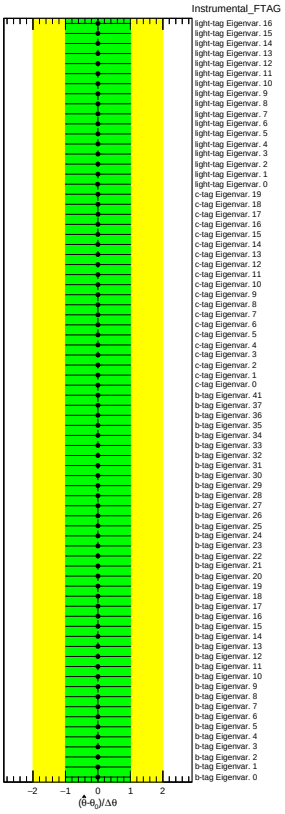
Figure H.6: Nuisance parameter constraint and pull studies in the $1\ell + 2\tau_{\text{had}}$ channel's hybrid fit.



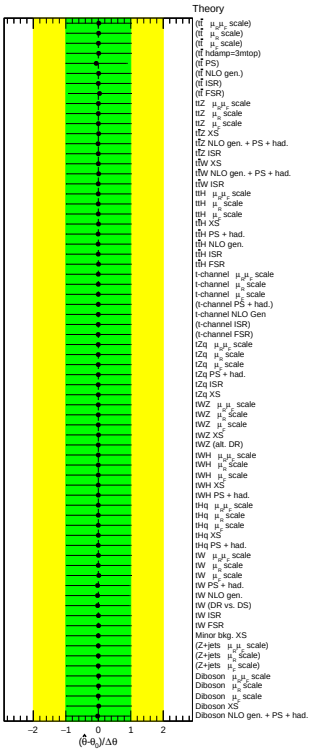

(a) Instrumental nuisance parameters.




(b) Instrumental JES/JER nuisance parameters.

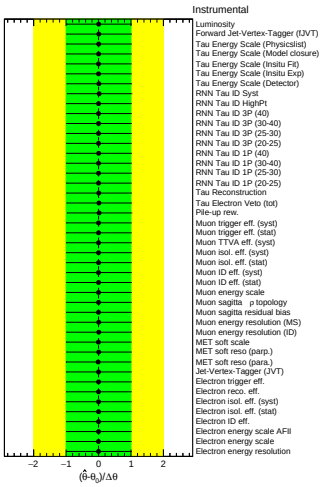



(c) Instrumental flavour tagging nuisance parameters.

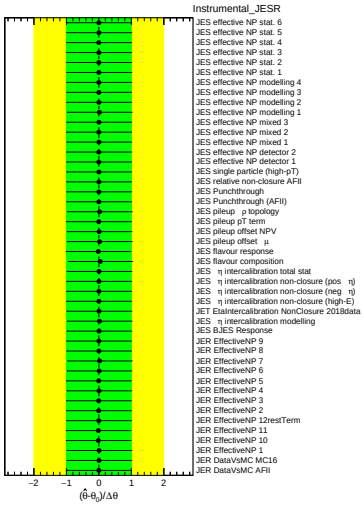



(d) Theory nuisance parameters.

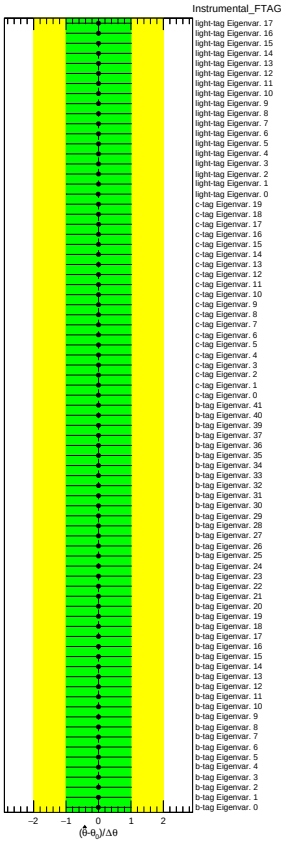
Figure H.7: Nuisance parameter constraint and pull studies in the $1\ell + 2\tau_{\text{had}}$ channel's fully unblinded fit.



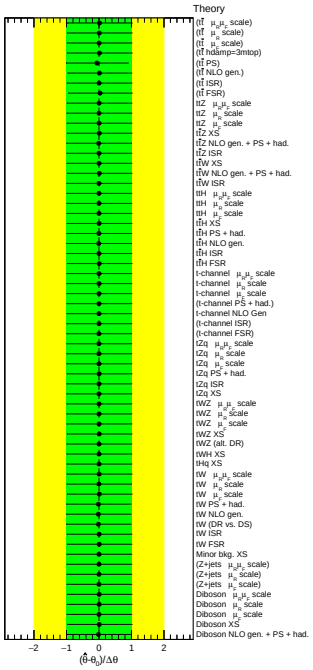

(a) Instrumental nuisance parameters.




(b) Instrumental JES/JER nuisance parameters.

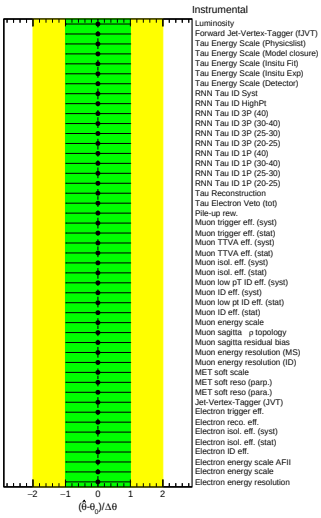



(c) Instrumental flavour tagging nuisance parameters.

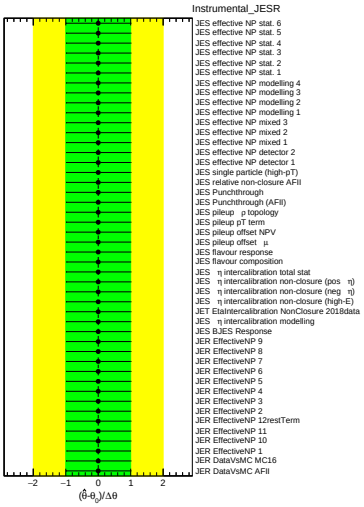



(d) Theory nuisance parameters.

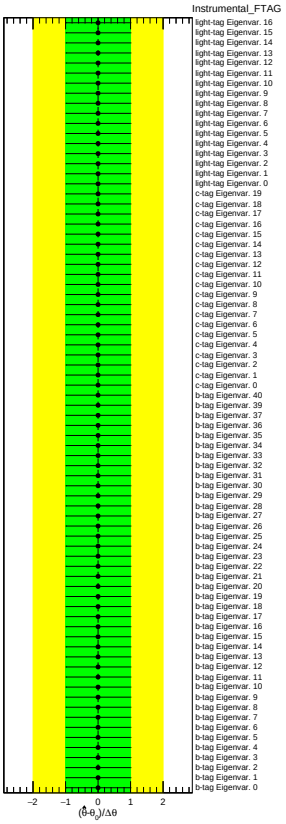
Figure H.8: Nuisance parameter constraint and pull studies in the $1\ell + 2\tau_{\text{had}}$ channel's fully unblinded inverted coupling fit.



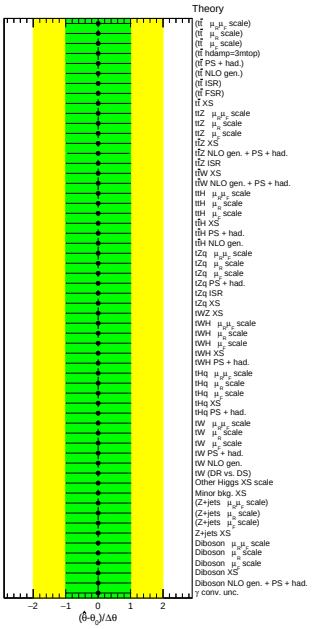

(a) Instrumental nuisance parameters.




(b) Instrumental JES/JER nuisance parameters.

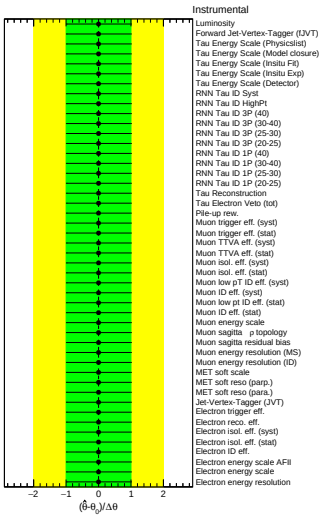



(c) Instrumental flavour tagging nuisance parameters.

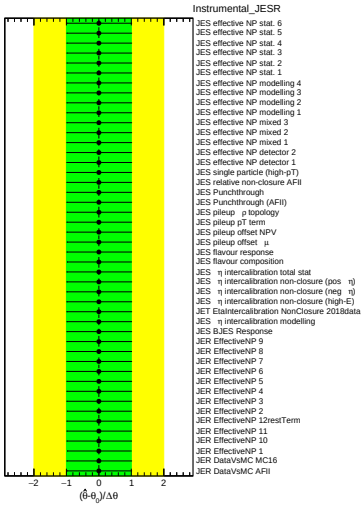



(d) Theory nuisance parameters.

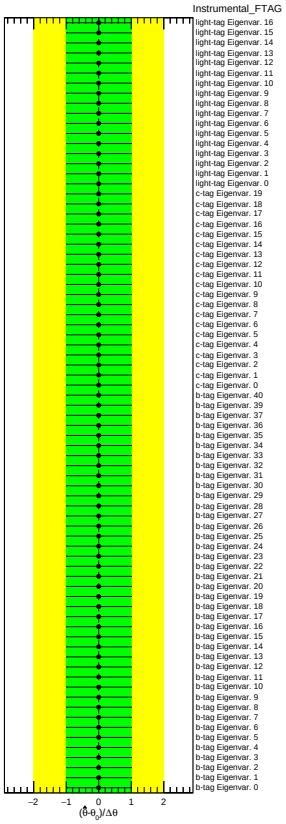
Figure H.9: Nuisance parameter constraint and pull studies in the $2\ell + 1\tau_{\text{had}}$ channel's Asimov fit.



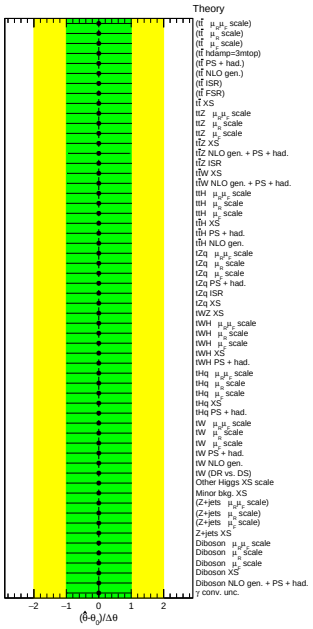

(a) Instrumental nuisance parameters.




(b) Instrumental JES/JER nuisance parameters.

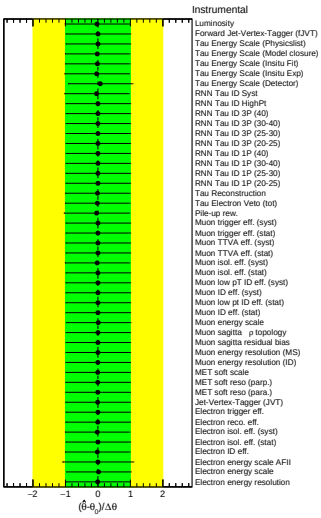



(c) Instrumental flavour tagging nuisance parameters.

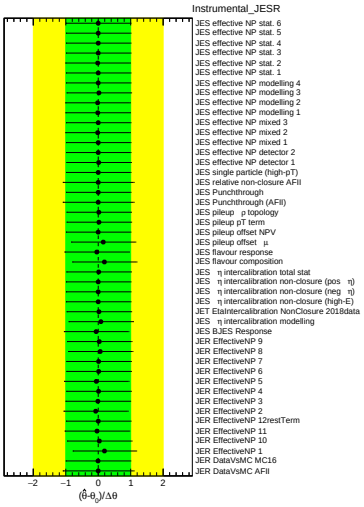



(d) Theory nuisance parameters.

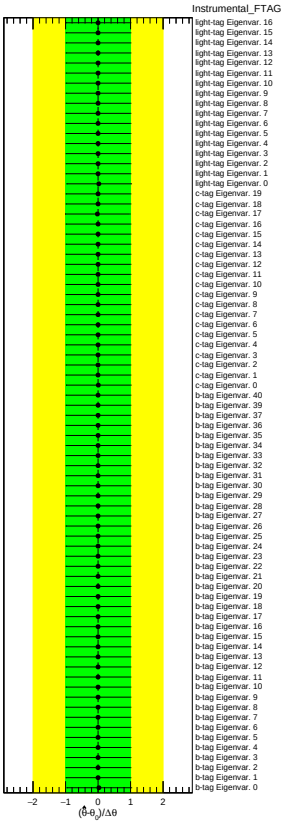
Figure H.10: Nuisance parameter constraint and pull studies in the $2\ell + 1\tau_{\text{had}}$ channel's hybrid fit.



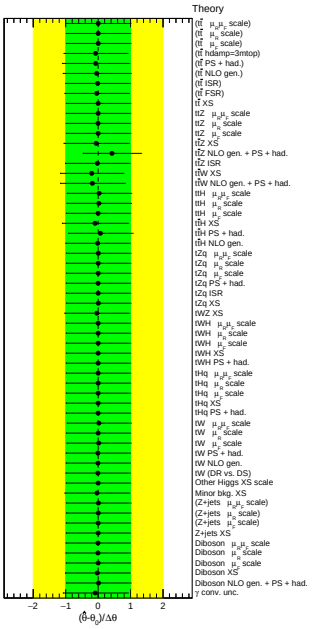

(a) Instrumental nuisance parameters.




(b) Instrumental JES/JER nuisance parameters.

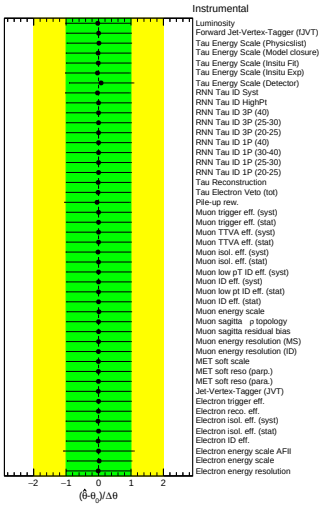



(c) Instrumental flavour tagging nuisance parameters.

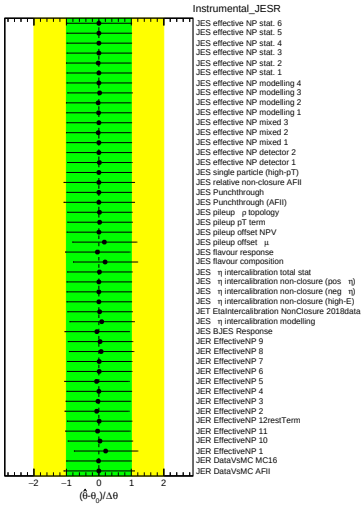



(d) Theory nuisance parameters.

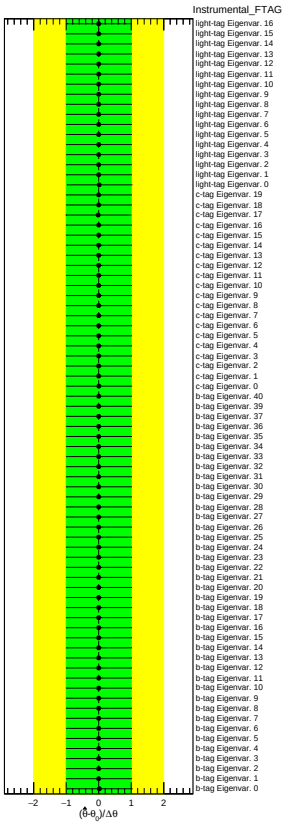
Figure H.11: Nuisance parameter constraint and pull studies in the $2\ell + 1\tau_{\text{had}}$ channel's fully unblinded fit.



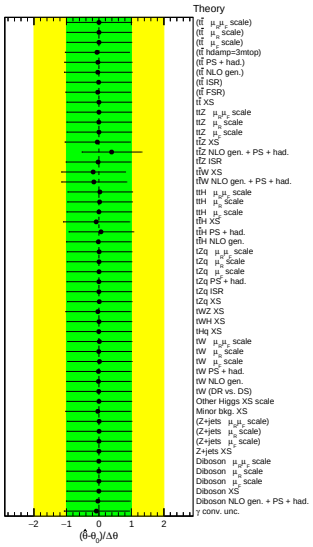

(a) Instrumental nuisance parameters.




(b) Instrumental JES/JER nuisance parameters.

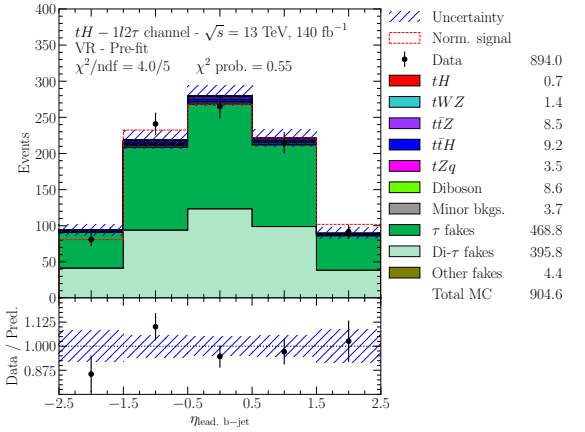



(c) Instrumental flavour tagging nuisance parameters.

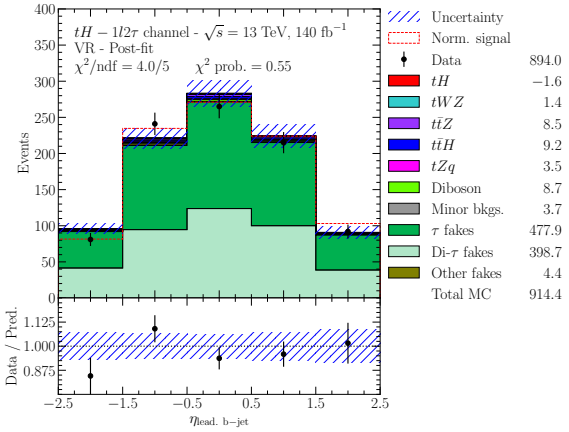



(d) Theory nuisance parameters.

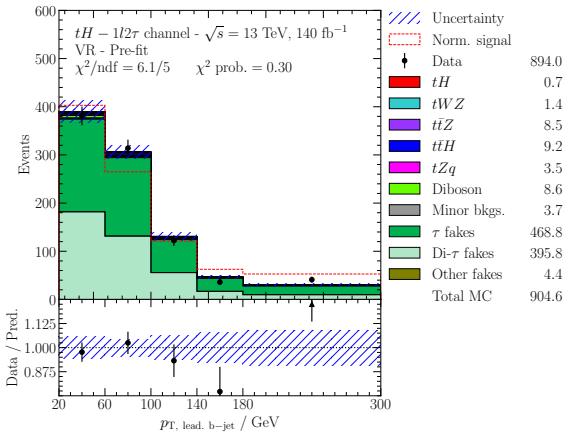
Figure H.12: Nuisance parameter constraint and pull studies in the $2\ell + 1\tau_{\text{had}}$ channel's fully unblinded inverted coupling fit.



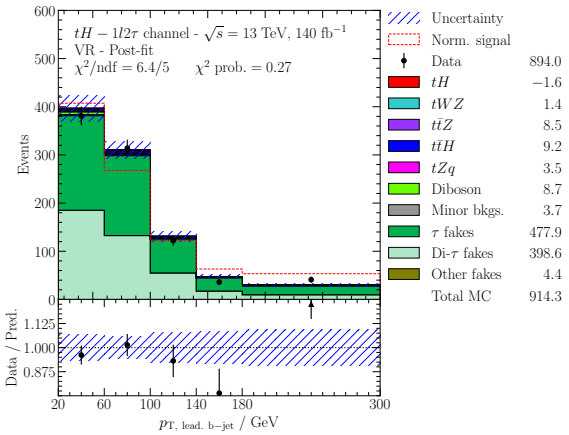

VR pre-fit.



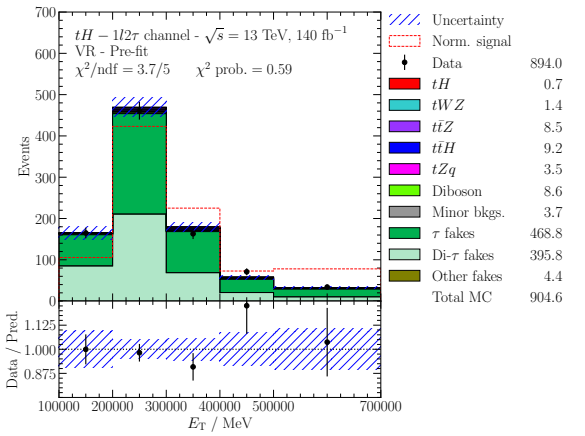

VR post-fit.



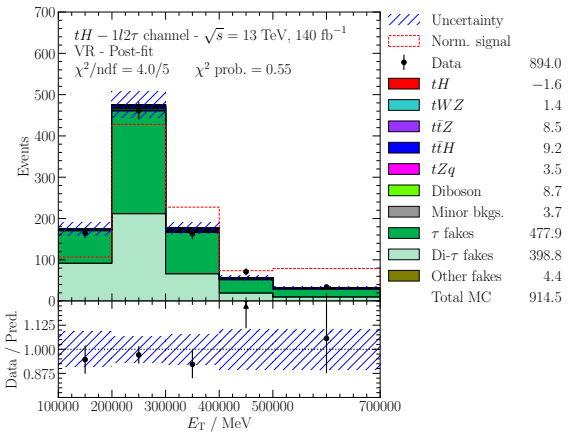

VR pre-fit.



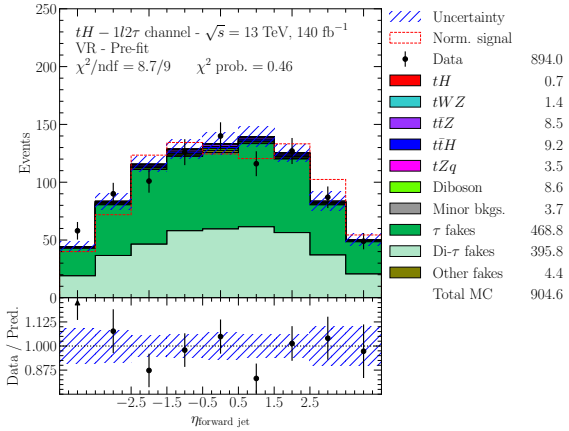

VR post-fit.



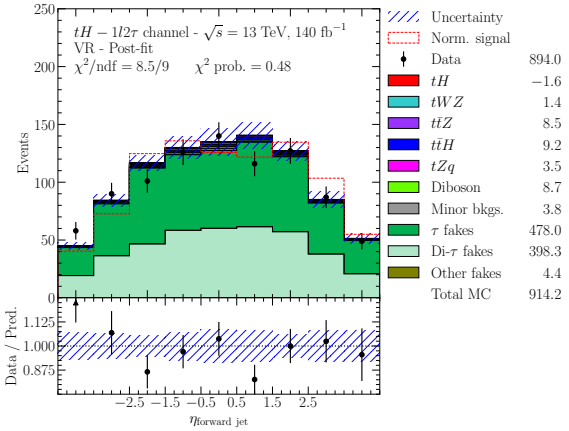

VR pre-fit.



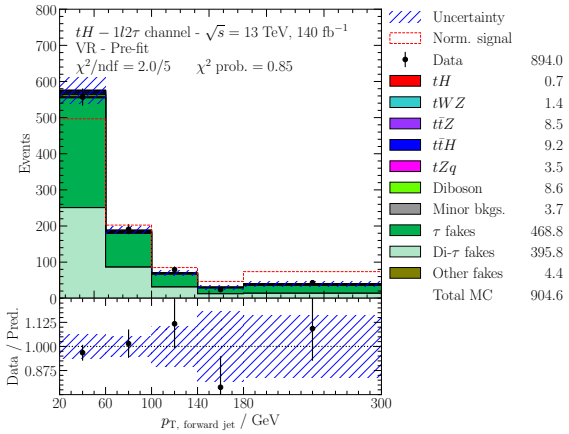

VR post-fit.



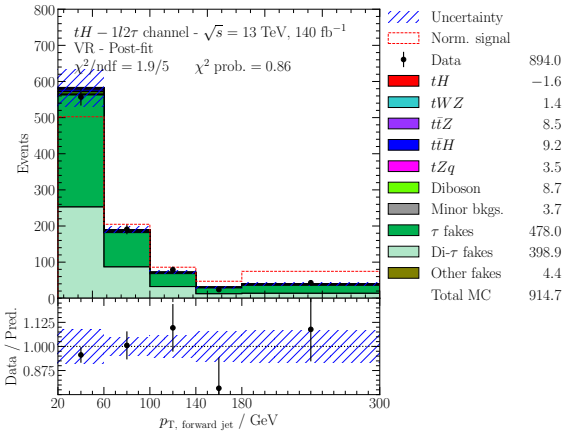

VR pre-fit.



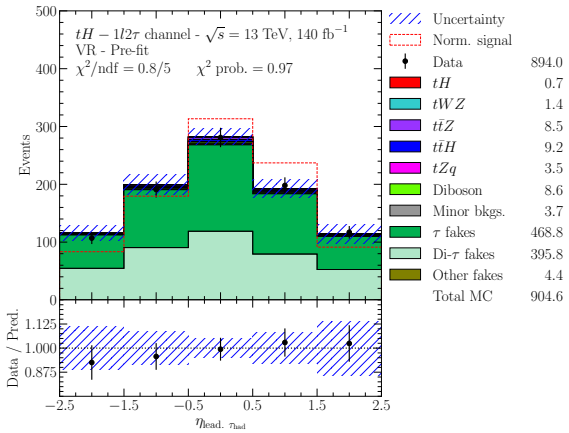

VR post-fit.



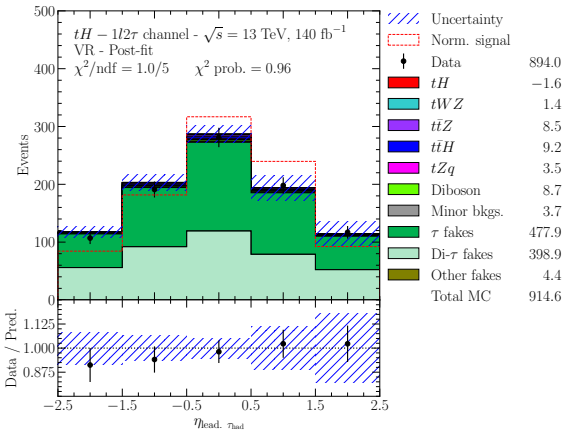

VR pre-fit.



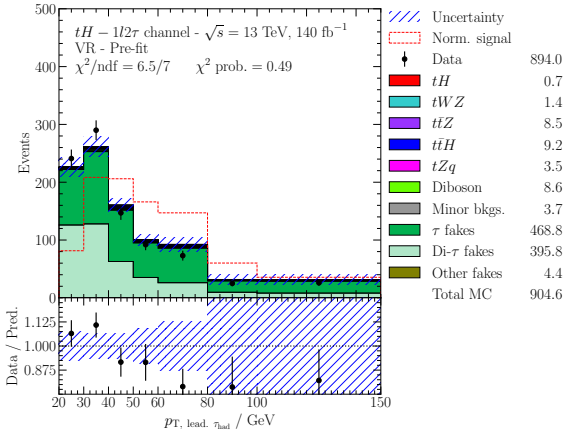

VR post-fit.



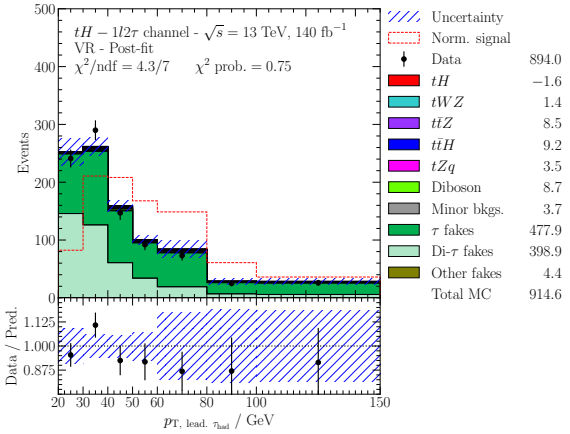

VR pre-fit.



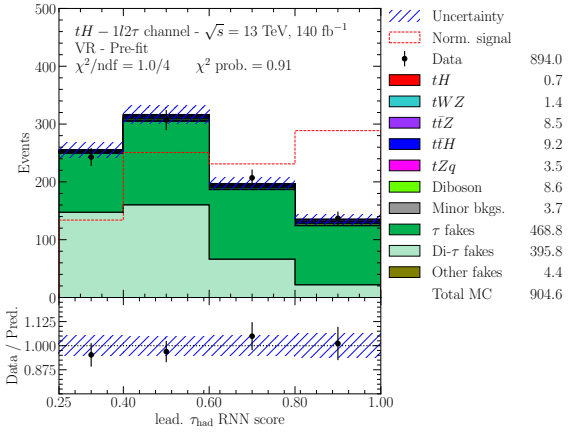

VR post-fit.



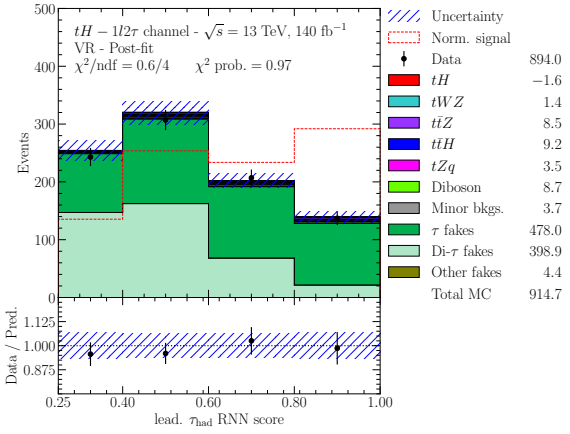

VR pre-fit.



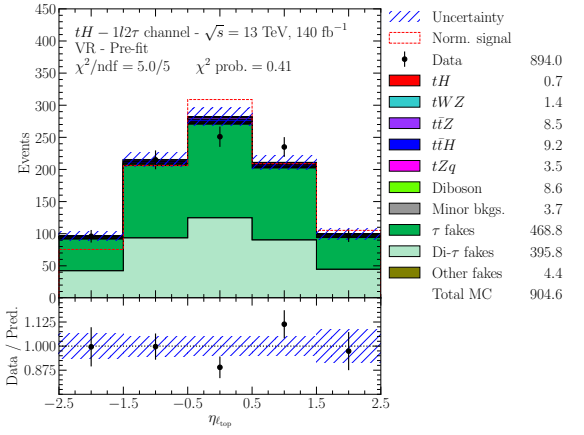

VR post-fit.



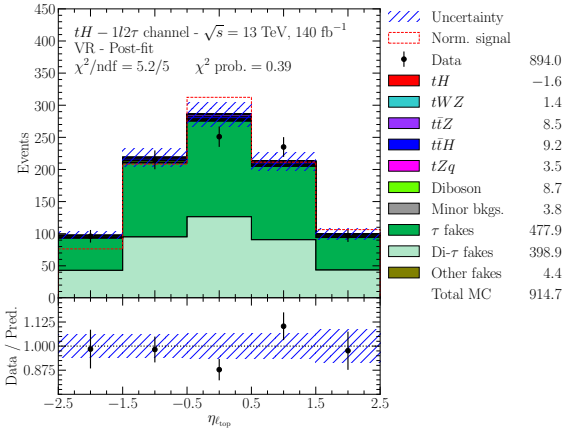

VR pre-fit.



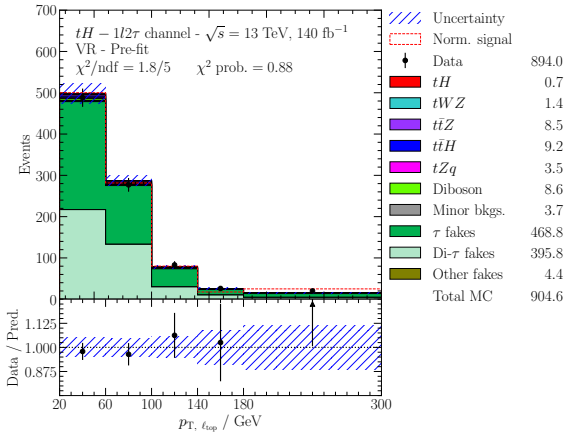

VR post-fit.



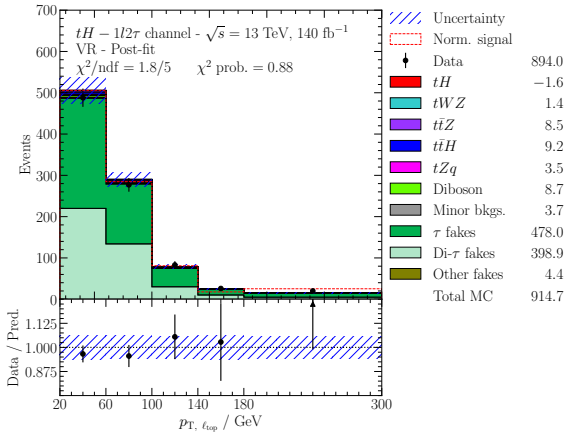

VR pre-fit.



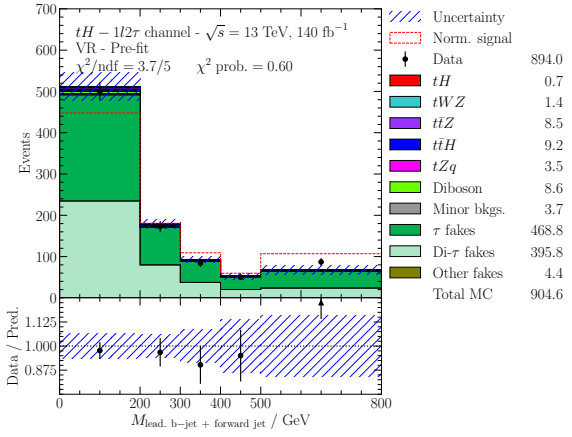

VR post-fit.



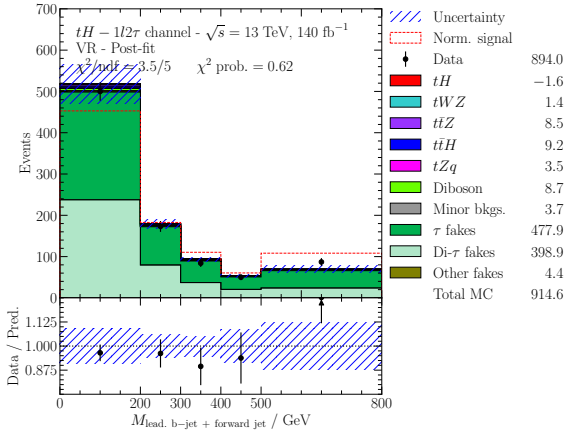

VR pre-fit.



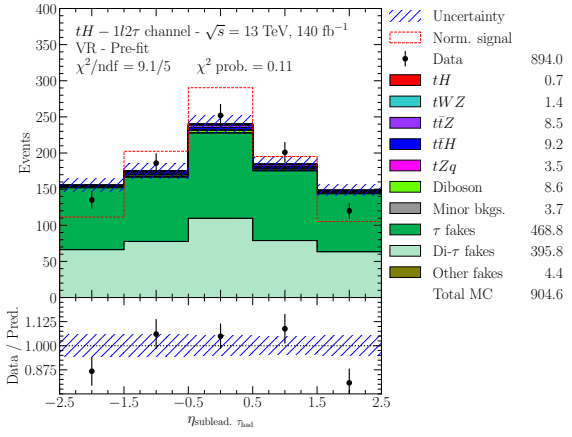

VR post-fit.



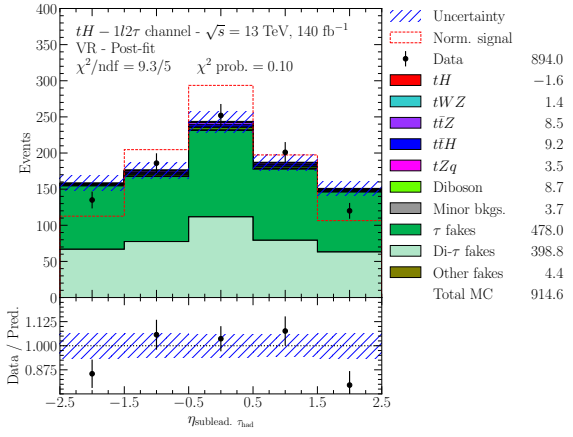

VR pre-fit.



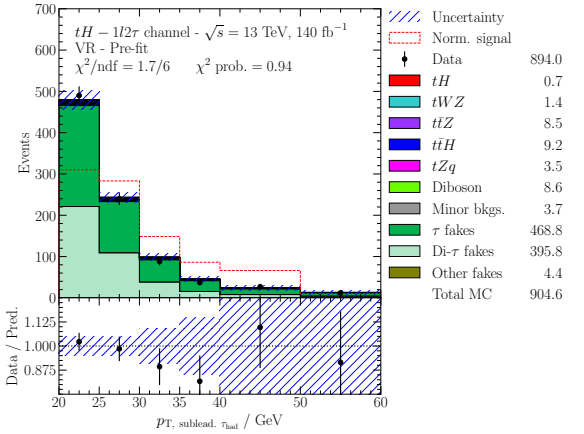

VR post-fit.



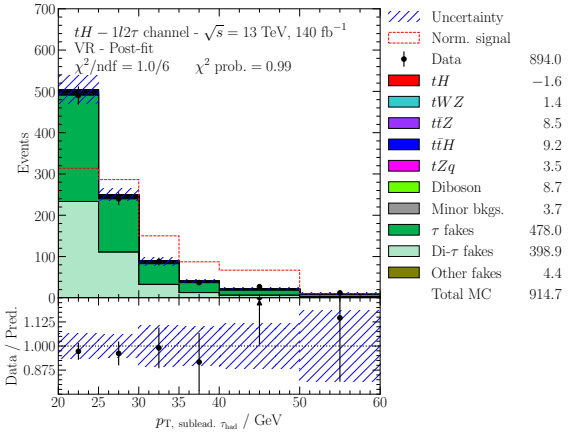

VR pre-fit.



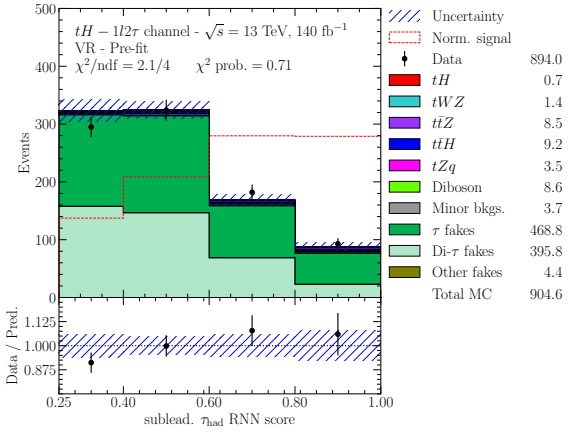

VR post-fit.



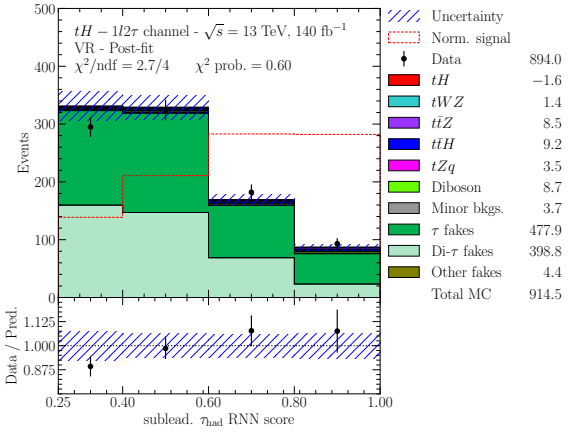

VR pre-fit.



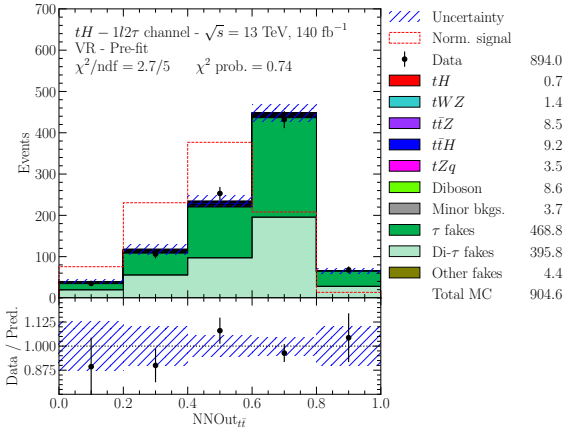

VR post-fit.



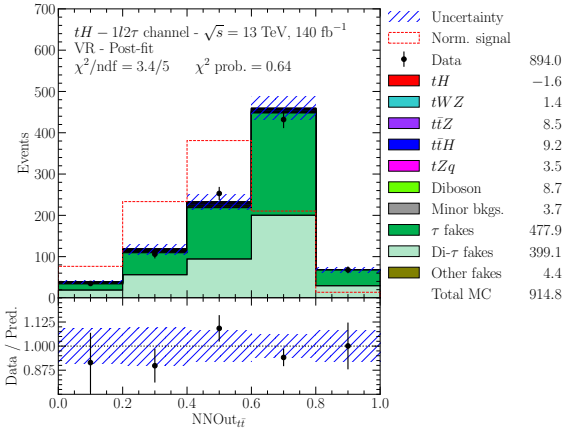

VR pre-fit.



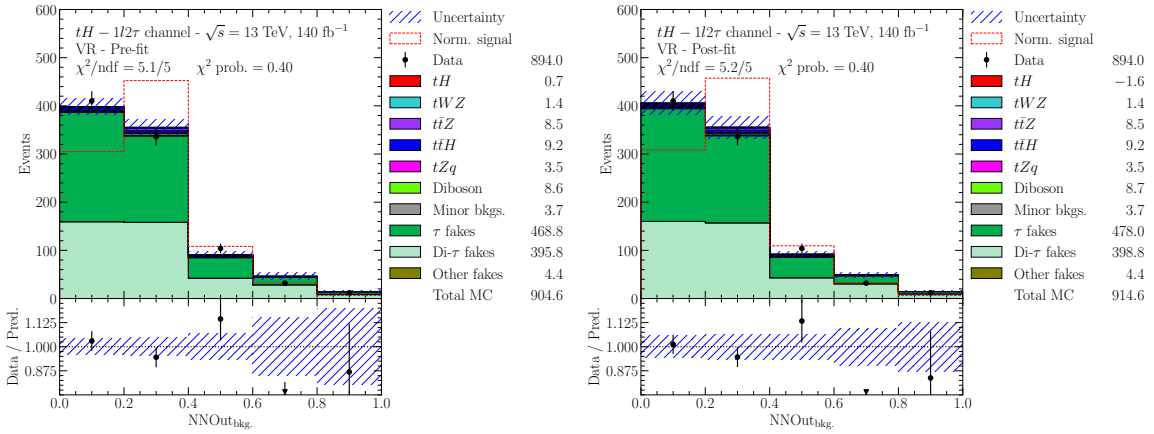

VR post-fit.




VR pre-fit.

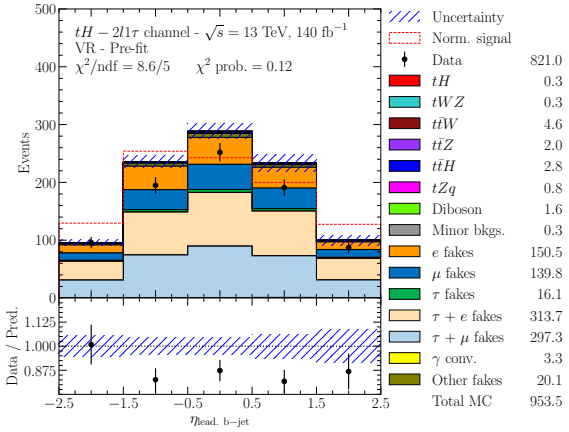



VR post-fit.

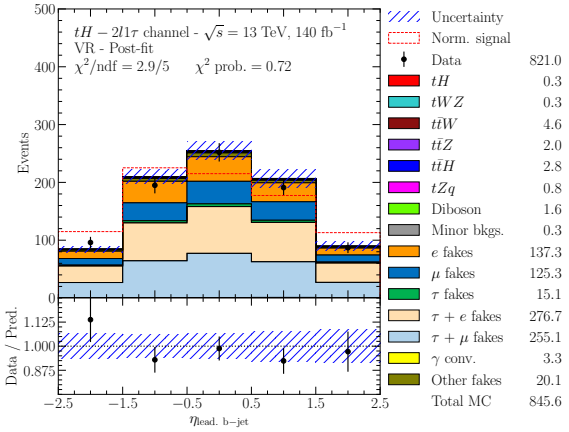



VR pre-fit. VR post-fit.

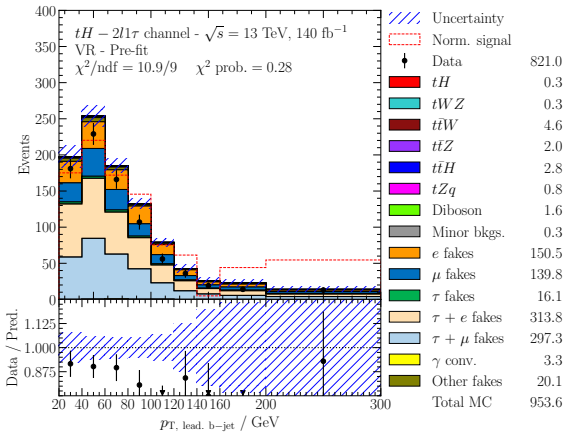
Figure H.13: All VR variables validated in the $1\ell + 2\tau_{\text{had}}$ channels hybrid fit.



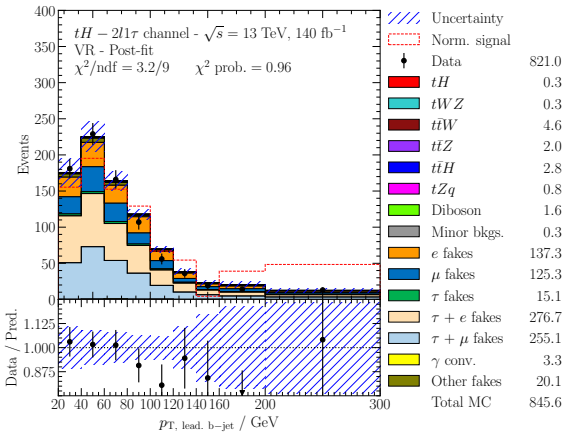

VR pre-fit.



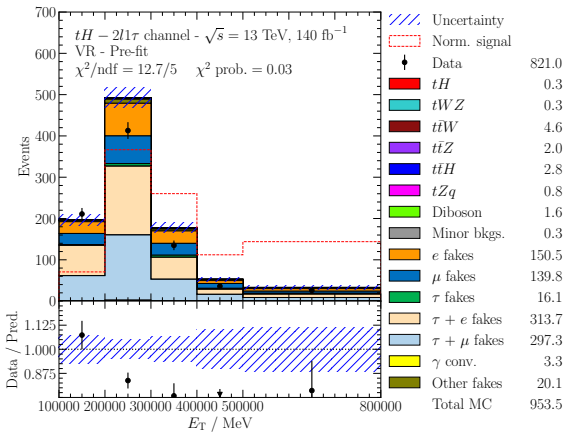

VR post-fit.



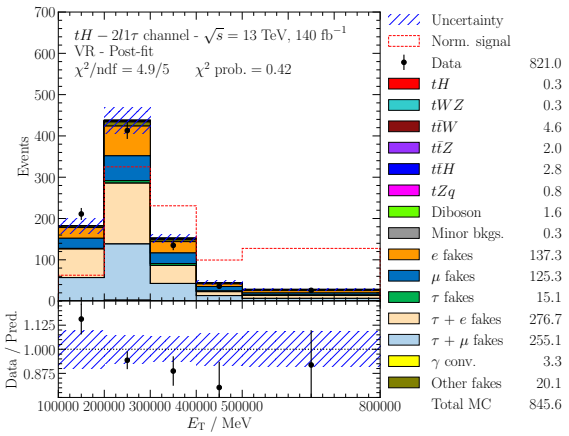

VR pre-fit.



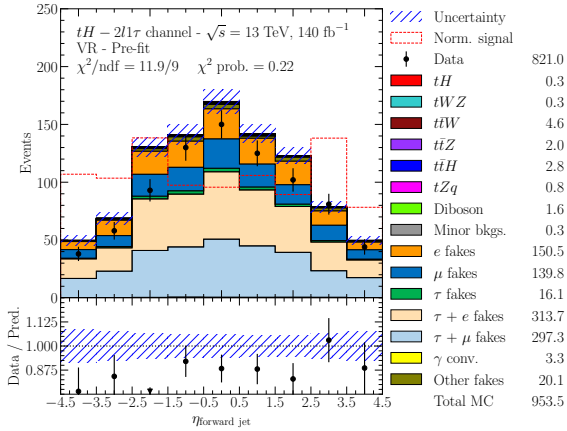

VR post-fit.



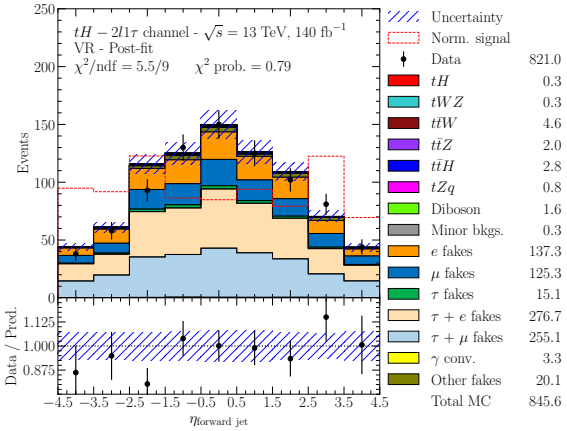

VR pre-fit.



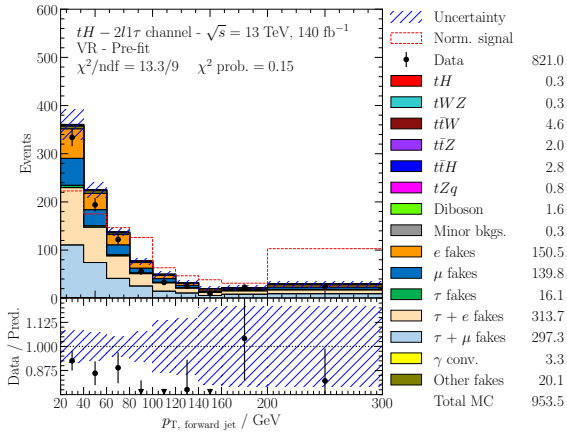

VR post-fit.



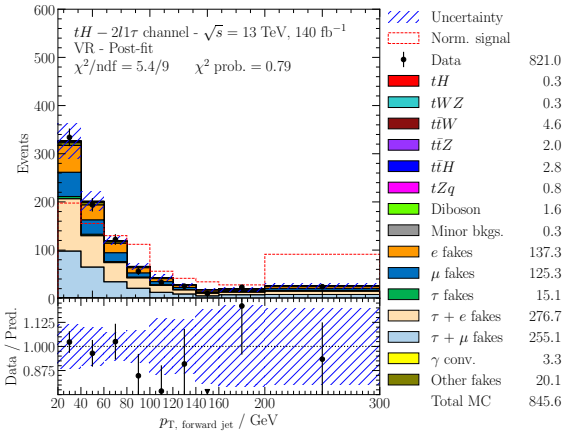

VR pre-fit.



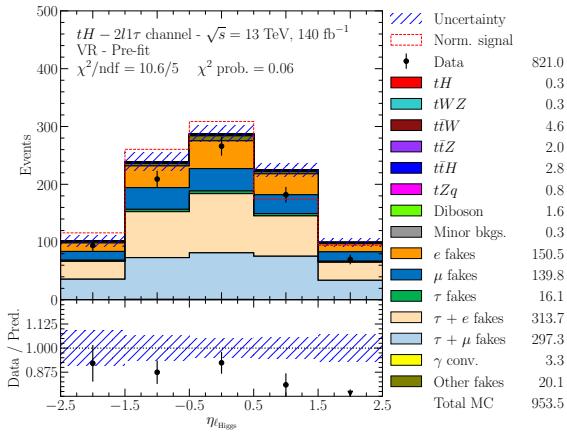

VR post-fit.



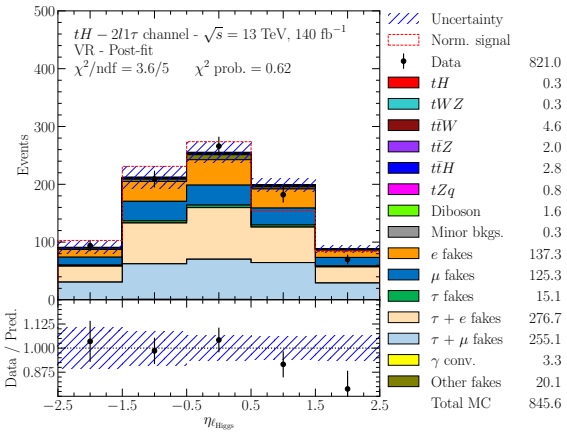

VR pre-fit.



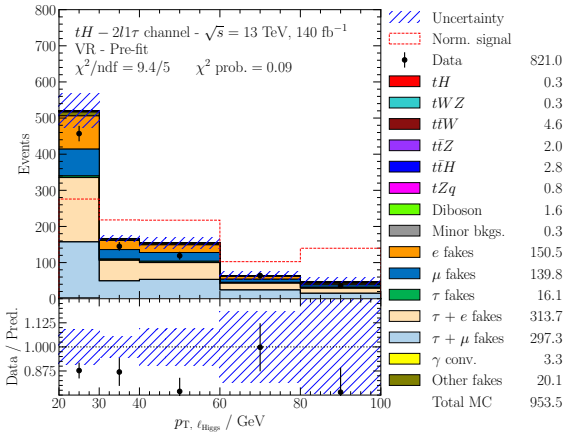

VR post-fit.



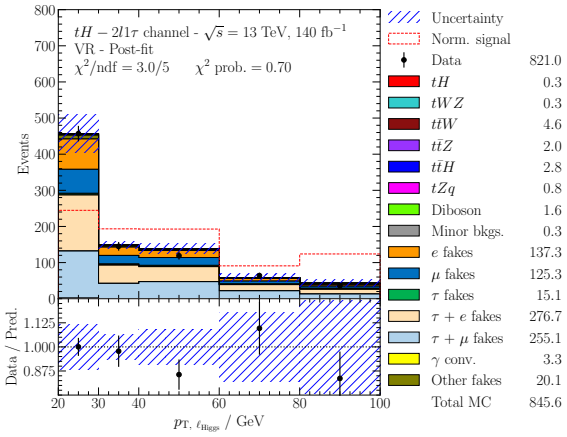

VR pre-fit.



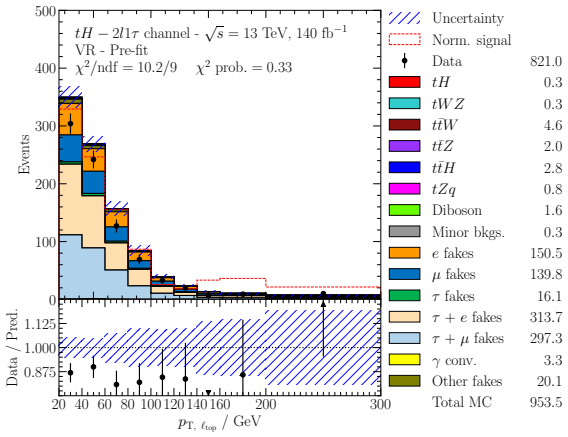

VR post-fit.



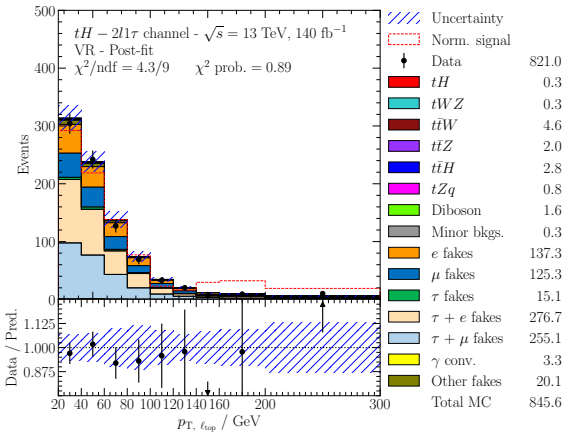

VR pre-fit.



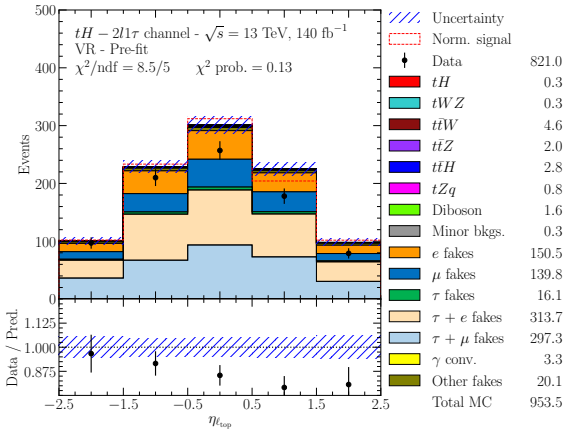

VR post-fit.



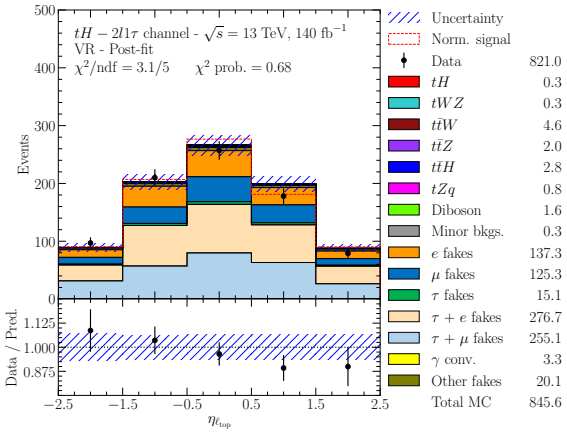

VR pre-fit.



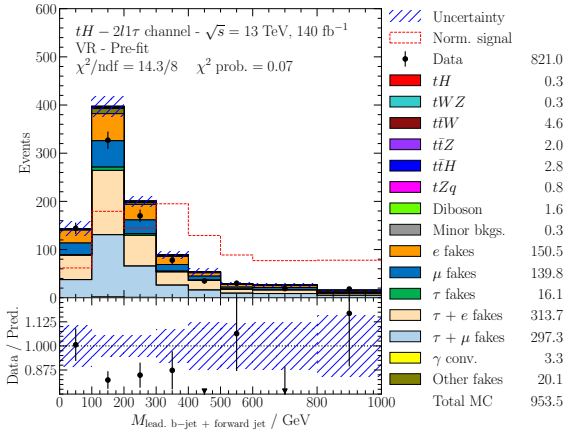

VR post-fit.



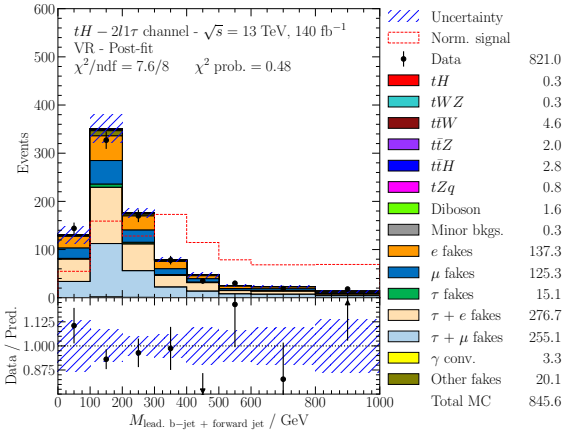

VR pre-fit.



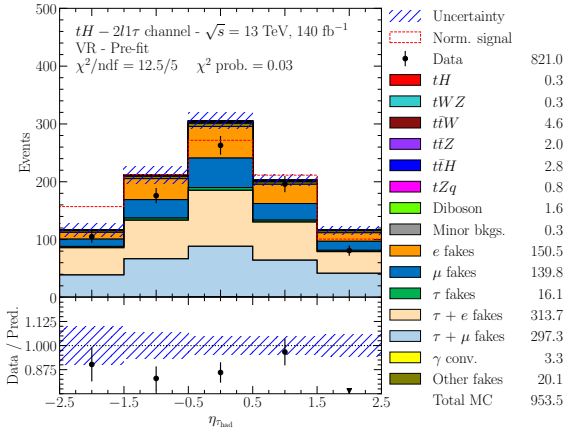

VR post-fit.



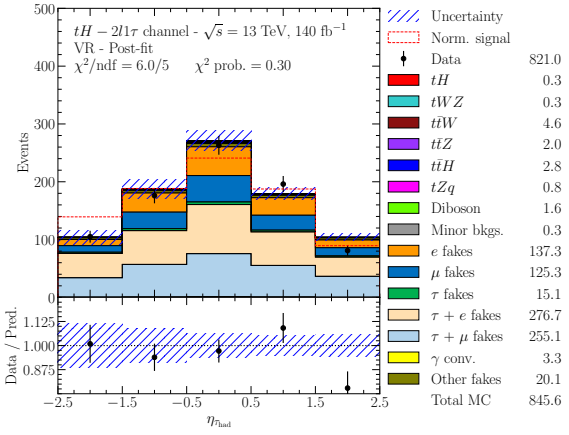

VR pre-fit.



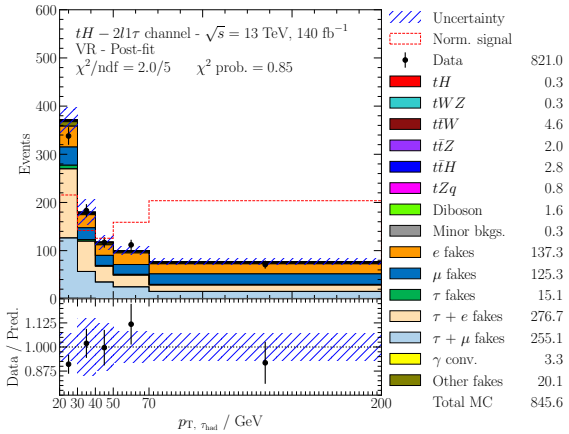

VR post-fit.



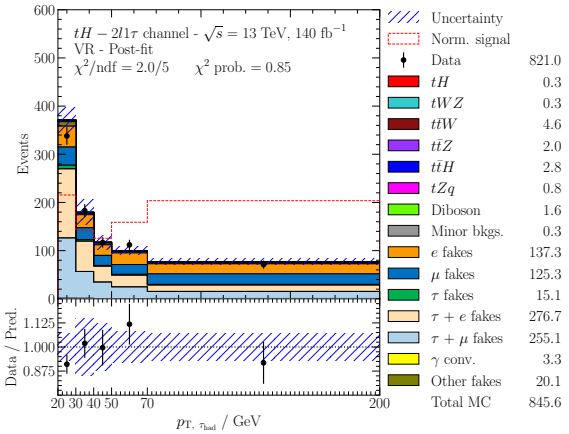

VR pre-fit.



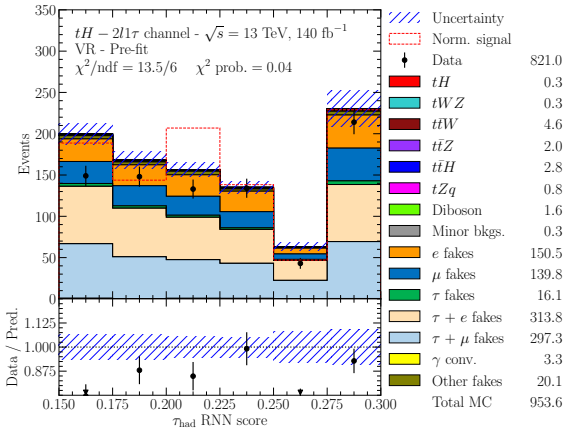

VR post-fit.



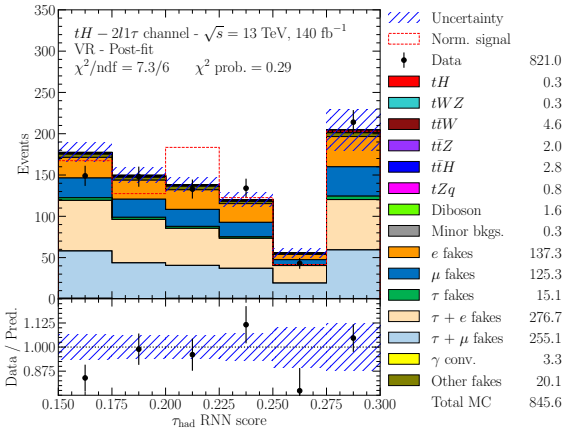

VR pre-fit.



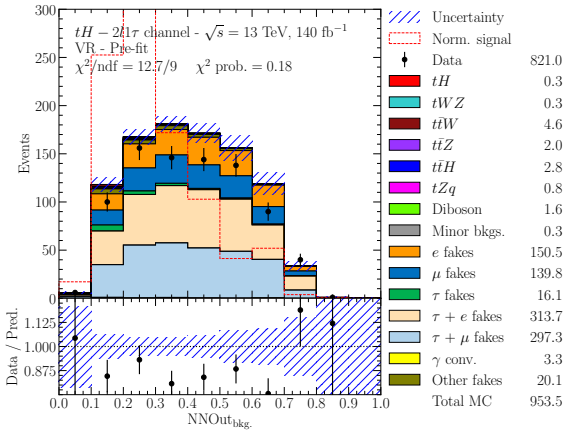

VR post-fit.



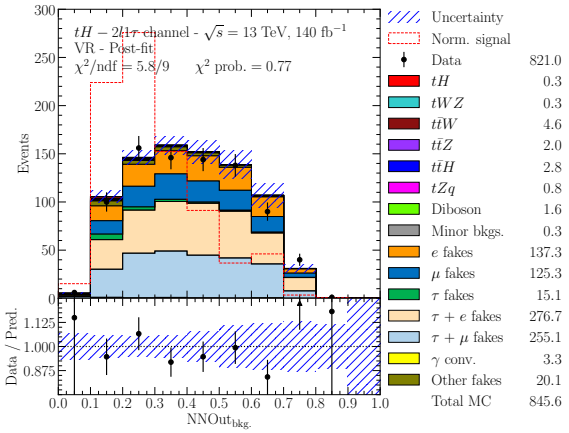

VR pre-fit.



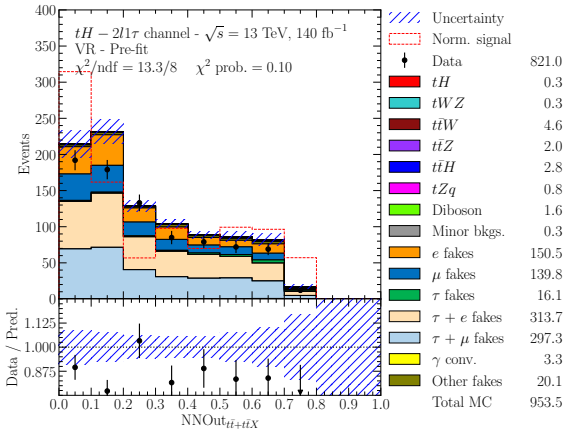

VR post-fit.



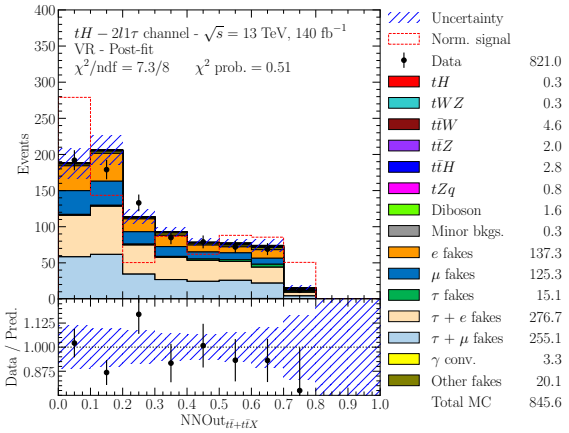

VR pre-fit.




VR post-fit.




VR pre-fit.




VR post-fit.

Figure H.14: All VR variables validated in the $2\ell + 1\tau_{\text{had}}$ channels hybrid fit.

# Bibliography


[1] S. Berryman, "Ancient Atomism", *Stanford Encyclopedia of Philosophy*, 2008 (cit. on p. 1).

[2] J. Dalton, *A new system of chemical philosophy*, vol. 1, Cambridge University Press, 2010 (cit. on p. 1).

[3] J. J. Thomson, *XL. Cathode rays*, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **44** (1897) 293 (cit. on p. 1).

[4] 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 [hep-ex] (cit. on p. 2).

[5] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B **716** (2012) 30, arXiv: 1207.7235 [hep-ex] (cit. on p. 2).

[6] R. Barate et al., *Search for the standard model Higgs boson at LEP*, Phys. Lett. B **565** (2003) 61, arXiv: hep-ex/0306033 (cit. on p. 2).

[7] M. Thomson, *Modern Particle Physics*, Cambridge University Press, 2013 (cit. on pp. 3, 8–10, 12, 13, 15, 17, 165).

[8] M. D. Schwartz, *Quantum Field Theory and the Standard Model*, Cambridge University Press, 2013 (cit. on p. 3).

[9] Cush, *Standard Model of Elementary Particles*, 2019, URL: https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg#file (cit. on p. 5).

[10] S. Navas et al., *Review of particle physics*, Phys. Rev. D **110** (2024) 030001 (cit. on pp. 5, 6, 11, 19, 20, 49).

[11] J. I. Friedman and H. W. Kendall, *Deep inelastic electron scattering*, Annual Review of Nuclear Science **22** (1972) 203 (cit. on p. 4).

[12] S. Mele, *The Measurement of the Number of Light Neutrino Species at LEP*, Adv. Ser. Dir. High Energy Phys. **23** (2015) 89, URL: https://cds.cern.ch/record/2103251 (cit. on p. 4).

[13] P. A. M. Dirac, *Quantised singularities in the electromagnetic field*, Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character **133** (1931) 60 (cit. on p. 4).

[14] P. M. S. Blackett and G. P. S. Occhialini,
*Some photographs of the tracks of penetrating radiation*,
Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character **139** (1933) 699 (cit. on p. 4).

[15] C. M. G. Lattes, H. Muirhead, G. P. S. Occhialini and C. F. Powell,
*Processes involving charged mesons*, Nature **159** (1947) 694 (cit. on p. 5).

[16] G. D. Rochester and C. Butler,
*Evidence for the existence of new unstable elementary particles*, Nature **160** (1947) 855 (cit. on p. 5).

[17] M. Gell-Mann, *A Schematic Model of Baryons and Mesons*, Phys. Lett. **8** (1964) 214 (cit. on p. 5).

[18] G. Zweig, *An $SU_3$ model for strong interaction symmetry and its breaking; Version 2*, (1964), URL: https://cds.cern.ch/record/570209 (cit. on p. 5).

[19] R. L. Mößbauer, "History of neutrino physics: Pauli's letters",
*Proceedings of the Fourth SFB-375 Ringberg Workshop Neutrino Astrophysics*, 1998 3 (cit. on p. 6).

[20] C. L. Cowan Jr, F. Reines, F. Harrison, H. Kruse and A. McGuire,
*Detection of the free neutrino: a confirmation*, Science **124** (1956) 103 (cit. on p. 6).

[21] J. Chadwick, *The existence of a neutron*, Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character **136** (1932) 692 (cit. on p. 6).

[22] A. Zee, *Group theory in a nutshell for physicists*, Princeton University Press, 2016 (cit. on p. 6).

[23] A. Einstein,
*On a heuristic point of view concerning the production and transformation of light*,
Annalen der Physik **17** (1905) 1 (cit. on p. 6).

[24] A. H. Compton, *A quantum theory of the scattering of X-rays by light elements*,
Physical review **21** (1923) 483 (cit. on p. 6).

[25] J. C. Collins, *Renormalization: an introduction to renormalization, the renormalization group and the operator-product expansion*, Cambridge university press, 1984 (cit. on p. 8).

[26] R. Brandelik et al., *Evidence for planar events in e+ e- annihilation at high energies*,
Physics Letters B **86** (1979) 243 (cit. on p. 9).

[27] G. Arnison et al., *Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s}$ = 540 GeV*, Physics Letters B **122** (1983) 103,
ISSN: 0370-2693,
URL: https://www.sciencedirect.com/science/article/pii/0370269383911772
(cit. on p. 10).

[28] M. Banner et al., *Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider*,
Physics Letters B **122** (1983) 476, ISSN: 0370-2693,
URL: https://www.sciencedirect.com/science/article/pii/0370269383916052
(cit. on p. 10).

[29] UA2 Collaboration, *Evidence for $Z \to e^+e^-$ - at the CERN $\overline{p}p$ collider*, Phys. Lett. B **129** (1983) 130, URL: https://cds.cern.ch/record/146503 (cit. on p. 10).

[30] UA1 Collaboration, *Experimental observation of lepton pairs of invariant mass around $95 GeV/c^2$ at the CERN SPS collider*, Phys. Lett. B **126** (1983) 398, URL: https://cds.cern.ch/record/163857 (cit. on p. 10).

[31] M. Kobayashi and T. Maskawa, *CP-Violation in the Renormalizable Theory of Weak Interaction*, Progress of Theoretical Physics **49** (1973) 652, ISSN: 0033-068X, eprint: https://academic.oup.com/ptp/article-pdf/49/2/652/5257692/49-2-652.pdf, URL: https://doi.org/10.1143/PTP.49.652 (cit. on p. 11).

[32] H. Georgi and S. L. Glashow, *Unity of All Elementary-Particle Forces*, **32** (1974) 438 (cit. on p. 12).

[33] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508, ed. by J. C. Taylor (cit. on p. 13).

[34] E. Noether, "Invariante variationsprobleme", *Gesammelte Abhandlungen-Collected Papers*, Springer, 1983 231 (cit. on p. 13).

[35] A. D. Sakharov, "Violation of CP-invariance, C-asymmetry, and baryon asymmetry of the Universe", *In The Intermissions... Collected Works on Research into the Essentials of Theoretical Physics in Russian Federal Nuclear Center, Arzamas-16*, World Scientific, 1998 84 (cit. on p. 16).

[36] P. Bechtle et al., *A Proposal for the Lohengrin Experiment to Search for Dark Sector Particles at the ELSA Accelerator*, 2024, arXiv: 2410.10956 [hep-ex], URL: https://arxiv.org/abs/2410.10956 (cit. on p. 16).

[37] I. C. Brock and T. Schörner-Sadenius, *Physics at the Terascale*, Wiley Online Library, 2011 (cit. on pp. 17, 19).

[38] X. Ji, *Parton Physics on a Euclidean Lattice*, Physical Review Letters **110** (2013), ISSN: 1079-7114, URL: http://dx.doi.org/10.1103/PhysRevLett.110.262002 (cit. on p. 18).

[39] H1, ZEUS Collaborations, *Combined measurement and QCD analysis of the inclusive $e \pm p$ scattering cross sections at HERA*, Journal of High Energy Physics **2010** (2010), ISSN: 1029-8479, URL: http://dx.doi.org/10.1007/JHEP01(2010)109 (cit. on p. 18).

[40] D. Griffiths, *Introduction to elementary particles*, John Wiley & Sons, 2020 (cit. on p. 18).

[41] P. de Jong, *Top Physics at the LHC*, 2009, arXiv: 0902.4798 [hep-ex], URL: https://arxiv.org/abs/0902.4798 (cit. on p. 19).

[42] C. Grojean, *Higgs Physics*, en, CERN Yellow Reports (2016), URL: https://e-publishing.cern.ch/index.php/CYR/article/view/435 (cit. on p. 20).

[43] ATLAS Collaboration, *Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector*, Phys. Lett. B **784** (2018) 173, arXiv: 1806.00425 [hep-ex] (cit. on p. 22).

[44] ATLAS Collaboration, *CP Properties of Higgs Boson Interactions with Top Quarks in the $t\bar{t}H$ and $tH$ Processes Using $H \to \gamma\gamma$ with the ATLAS Detector*, Phys. Rev. Lett. **125** (2020) 061802, arXiv: 2004.04545 [hep-ex] (cit. on p. 22).

[45] CMS Collaboration, *Measurements of Higgs boson production cross sections and couplings in the diphoton decay channel at $\sqrt{s} = 13$ TeV*, JHEP **07** (2021) 027, arXiv: 2103.06956 [hep-ex] (cit. on p. 22).

[46] *SM Higgs production cross sections at $\sqrt{s} = 13$ TeV*, 2025, URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt13TeV (visited on 20/01/2025) (cit. on p. 23).

[47] F. Demartin, F. Maltoni, K. Mawatari and M. Zaro, *Higgs production in association with a single top quark at the LHC*, 2015, arXiv: 1504.00611 [hep-ph], URL: https://arxiv.org/abs/1504.00611 (cit. on p. 23).

[48] CMS Collaboration, *Measurement of the Higgs boson production rate in association with top quarks in final states with electrons, muons, and hadronically decaying tau leptons at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **81** (2021) 378, arXiv: 2011.03652 [hep-ex] (cit. on pp. 23, 161).

[49] V. Vecchio et al., *Search for associated production of a Higgs boson and a single top quark in multi-lepton final states using $pp$ collisions at 13 TeV with the ATLAS detector*, tech. rep., CERN, 2022, URL: https://cds.cern.ch/record/2802431 (cit. on pp. 23, 142).

[50] P. Bechtle et al., *Faszinierende Teilchenphysik*, 2023 (cit. on p. 24).

[51] H. Wiedemann, *Particle accelerator physics*, Springer Nature, 2015 (cit. on p. 29).

[52] L. Evans, *The Large Hadron Collider: a marvel of technology; 2nd ed.* Physics (EPFL Press), On the cover : Including the discovery of the higgs boson, Lausanne: EPFL Press, 2018, URL: https://cds.cern.ch/record/2645935 (cit. on pp. 29, 32, 41–43).

[53] H. Kolanoski and N. Wermes, *Teilchendetektoren*, Springer, 2016 (cit. on pp. 29, 35–38).

[54] *Atlas Technical Design Reports*, 2025, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/AtlasTechnicalDesignReports (visited on 29/01/2025) (cit. on p. 29).

[55] P. Mouche, *Overall view of the LHC. Vue d'ensemble du LHC*, (2014), General Photo, URL: https://cds.cern.ch/record/1708847 (cit. on p. 32).

[56] O. S. Brüning et al., *LHC Design Report*, CERN Yellow Reports: Monographs, Geneva: CERN, 2004, URL: https://cds.cern.ch/record/782076 (cit. on p. 32).

[57] J. Gareyte, *LHC main parameters*, Part. Accel. **50** (1995) 61 (cit. on p. 32).

[58] *Accelerators*, 2025, URL: https://home.cern/science/accelerators (visited on 24/01/2025) (cit. on p. 33).

[59] E. Lopienska, *The CERN accelerator complex, layout in 2022. Complexe des accélérateurs du CERN en janvier 2022*, (2022), General Photo,
URL: https://cds.cern.ch/record/2800984 (cit. on p. 33).

[60] *Longer term LHC schedule*, 2025,
URL: https://lhc-commissioning.web.cern.ch/schedule/LHC-long-term.htm
(visited on 24/01/2025) (cit. on p. 34).

[61] I. Zurbano Fernandez et al.,
*High-Luminosity Large Hadron Collider (HL-LHC): Technical design report*,
**10/2020** (2020), ed. by I. Béjar Alonso et al. (cit. on p. 34).

[62] B. B. Rossi, *High-energy particles*, (No Title) (1952) (cit. on p. 36).

[63] J. Pequenao, "Computer generated image of the whole ATLAS detector", 2008,
URL: https://cds.cern.ch/record/1095924 (cit. on p. 39).

[64] Izaak Neutelings, *ATLAS coordinate system*, 2025,
URL: https://tikz.net/axis3d_cms/ (visited on 28/01/2025) (cit. on p. 40).

[65] O. Aberle et al., *High-Luminosity Large Hadron Collider (HL-LHC): Technical design report*,
CERN Yellow Reports: Monographs, Geneva: CERN, 2020,
URL: https://cds.cern.ch/record/2749422 (cit. on p. 40).

[66] A. Yamamoto et al., *Progress in ATLAS central solenoid magnet*,
IEEE Trans. Appl. Supercond. **10** (2000) 353,
URL: https://cds.cern.ch/record/438924 (cit. on p. 41).

[67] *ATLAS Magnet System*, 2025,
URL: https://atlas-public.web.cern.ch/Discover/Detector/Magnet-System
(visited on 27/01/2025) (cit. on p. 41).

[68] 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. 41).

[69] 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. 41).

[70] B. Abbott et al., *Production and integration of the ATLAS Insertable B-Layer*,
JINST **13** (2018) T05008, arXiv: 1803.00844 [physics.ins-det] (cit. on p. 41).

[71] ATLAS Collaboration, *Performance of the reconstruction of large impact parameter tracks in the inner detector of ATLAS*, Eur. Phys. J. C **83** (2023) 1081, arXiv: 2304.12867 [hep-ex]
(cit. on p. 41).

[72] ATLAS Collaboration,
*Operation and performance of the ATLAS semiconductor tracker in LHC Run 2*,
JINST **17** (2022) P01013, arXiv: 2109.02591 [physics.ins-det] (cit. on p. 41).

[73] ATLAS Collaboration, *Performance of the ATLAS Transition Radiation Tracker in Run 1 of the LHC: tracker properties*, JINST **12** (2017) P05002, arXiv: 1702.06473 [hep-ex]
(cit. on p. 41).

[74] J. Pequenao, "Computer generated image of the ATLAS inner detector", 2008, URL: https://cds.cern.ch/record/1095926 (cit. on p. 42).

[75] ATLAS Collaboration, *Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run 2 data collected in 2015 and 2016*, Eur. Phys. J. C **79** (2019) 205, arXiv: 1810.05087 [hep-ex] (cit. on p. 42).

[76] ATLAS Collaboration, *Operation and performance of the ATLAS Tile Calorimeter in Run 1*, Eur. Phys. J. C **78** (2018) 987, arXiv: 1806.02129 [hep-ex] (cit. on p. 42).

[77] 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. 43).

[78] A. Martínez, *The Run-2 ATLAS Trigger System*, Journal of Physics: Conference Series **762** (2016) 012003 (cit. on p. 43).

[79] *The Worldwide LHC Computing Grid*, 2025, URL: https://wlcg-public.web.cern.ch (visited on 30/01/2025) (cit. on p. 44).

[80] J. Shiers, *The Worldwide LHC Computing Grid (worldwide LCG)*, Comput. Phys. Commun. **177** (2007) 219, URL: https://cds.cern.ch/record/1063847 (cit. on p. 44).

[81] S. Mehlhase, *ATLAS detector slice (and particle visualisations)*, (2021), URL: https://cds.cern.ch/record/2770815 (cit. on p. 45).

[82] *ATLAS Tracking Software Tutorial*, 2025, URL: https://atlassoftwaredocs.web.cern.ch/internal-links/tracking-tutorial/index.html (visited on 03/02/2025) (cit. on p. 46).

[83] 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 (cit. on p. 46).

[84] ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2*, Eur. Phys. J. C **77** (2017) 673, arXiv: 1704.07983 [hep-ex] (cit. on p. 46).

[85] R. Frühwirth and W. Waltenberger, *Adaptive Multi-vertex fitting*, tech. rep., CERN, 2005, URL: https://cds.cern.ch/record/803519 (cit. on p. 46).

[86] *Calibration and Performance of the Electromagnetic Calorimeter*, tech. rep., CERN, 2009, URL: https://cds.cern.ch/record/1171938 (cit. on p. 46).

[87] 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 [hep-ex] (cit. on p. 47).

[88] W. Lampl et al., *Calorimeter Clustering Algorithms: Description and Performance*, tech. rep., CERN, 2008, URL: https://cds.cern.ch/record/1099735 (cit. on p. 47).

[89] R. Atkin, *Review of jet reconstruction algorithms*, Journal of Physics: Conference Series **645** (2015) 012008, URL: https://dx.doi.org/10.1088/1742-6596/645/1/012008 (cit. on p. 48).

[90] M. Cacciari, G. P. Salam and G. Soyez, *The anti-$k_t$ jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: 0802.1189 [hep-ph] (cit. on p. 48).

[91] 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 pp. 48, 49).

[92] 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 **81** (2021) 689, arXiv: 2007.02645 [hep-ex] (cit. on pp. 48, 49).

[93] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*, ATLAS-CONF-2014-018, 2014, URL: https://cds.cern.ch/record/1700870 (cit. on p. 49).

[94] ATLAS Collaboration, *ATLAS flavour-tagging algorithms for the LHC Run 2 $pp$ collision dataset*, Eur. Phys. J. C **83** (2023) 681, arXiv: 2211.16345 [physics.data-an] (cit. on pp. 49, 82).

[95] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*, ATL-PHYS-PUB-2017-013, 2017, URL: https://cds.cern.ch/record/2273281 (cit. on p. 49).

[96] ATLAS Collaboration, *Reconstruction, Identification, and Calibration of hadronically decaying tau leptons with the ATLAS detector for the LHC Run 3 and reprocessed Run 2 data*, ATL-PHYS-PUB-2022-044, 2022, URL: https://cds.cern.ch/record/2827111 (cit. on pp. 49, 50, 82).

[97] ATLAS Collaboration, *Identification and energy calibration of hadronically decaying tau leptons with the ATLAS experiment in $pp$ collisions at $\sqrt{s}$ = 8 TeV*, Eur. Phys. J. C **75** (2015) 303, arXiv: 1412.7086 [hep-ex] (cit. on p. 49).

[98] ATLAS Collaboration, *Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment*, ATL-PHYS-PUB-2019-033, 2019, URL: https://cds.cern.ch/record/2688062 (cit. on p. 50).

[99] 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 [physics.ins-det] (cit. on pp. 50, 51).

[100] *Electrons and photons*, 2025, URL: https://atlassoftwaredocs.web.cern.ch/internal-links/lectures/2024/egamma/ (visited on 10/02/2025) (cit. on p. 50).

[101] ATLAS Collaboration, *Electron and photon efficiencies in LHC Run 2 with the ATLAS experiment*, JHEP **05** (2024) 162, arXiv: 2308.13362 [hep-ex] (cit. on pp. 50, 82).

[102] J. Illingworth and J. Kittler, *A survey of the hough transform*, Computer Vision, Graphics, and Image Processing **44** (1988) 87, ISSN: 0734-189X, URL: https://www.sciencedirect.com/science/article/pii/S0734189X88800331 (cit. on p. 51).

[103] 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 [hep-ex] (cit. on pp. 51, 82).

[104] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at* $\sqrt{s}$ = 13 *TeV*, Eur. Phys. J. C **78** (2018) 903, arXiv: 1802.08168 [hep-ex] (cit. on p. 52).

[105] *Jets and Missing Transverse Energy*, 2025, URL: https://atlassoftwaredocs.web.cern.ch/internal-links/lectures/2024/jetetmiss/ (visited on 04/02/2025) (cit. on p. 52).

[106] *Mensch ärgere dich nicht*, 2025, URL: https://en.wikipedia.org/wiki/Mensch_%C3%A4rgere_Dich_nicht (cit. on p. 53).

[107] O. Behnke, K. Kröninger, G. Schott and T. Schörner-Sadenius, *Data analysis in high energy physics: a practical guide to statistical methods*, John Wiley & Sons, 2013 (cit. on pp. 53, 54, 56, 59).

[108] K. Cranmer, *Practical Statistics for the LHC*, 2015, arXiv: 1503.07622 [physics.data-an], URL: https://arxiv.org/abs/1503.07622 (cit. on pp. 53, 60).

[109] A. N. Kolmogorov, *Foundations of the theory of probability: Second English Edition*, Courier Dover Publications, 2018 (cit. on p. 54).

[110] *Minuit2*, 2025, URL: https://root.cern.ch/doc/master/Minuit2Page.html (cit. on p. 56).

[111] *TRExFitter documentation*, 2025, URL: https://trexfitter-docs.web.cern.ch/trexfitter-docs/ (cit. on pp. 56, 134).

[112] 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. 56).

[113] 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), ISSN: 1434-6052, URL: http://dx.doi.org/10.1140/epjc/s10052-011-1554-0 (cit. on pp. 57, 58, 61, 63).

[114] C. Theis and W. Kernbichler, *Grundlagen der Monte Carlo Methoden*, TU Graz (2002) (cit. on pp. 62, 167).

[115] *Jets and Missing Transverse Energy*, 2025, URL: https://atlassoftwaredocs.web.cern.ch/internal-links/lectures/2023/mc_generators/ (visited on 24/02/2025) (cit. on p. 62).

[116] *Athena basics*, 2025, URL: https://atlassoftwaredocs.web.cern.ch/athena/basics/ (visited on 24/02/2025) (cit. on p. 62).

[117] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823, arXiv: 1005.4568 [physics.ins-det] (cit. on pp. 62, 78).

[118] *Geant4*, 2025, URL: https://geant4.web.cern.ch/ (visited on 24/02/2025) (cit. on p. 62).

[119] ATLAS Collaboration, *AtlFast3: The Next Generation of Fast Simulation in ATLAS*, Comput. Softw. Big Sci. **6** (2022) 7, arXiv: 2109.02551 [hep-ex] (cit. on p. 62).

[120] K. Danziger, S. Höche and F. Siegert, *Reducing negative weights in Monte Carlo event generation with Sherpa*, arXiv preprint arXiv:2110.15211 (2021) (cit. on p. 63).

[121] IBM, *What are large language models (LLMs)*, 2025, URL: https://www.ibm.com/think/topics/large-language-models (visited on 06/03/2025) (cit. on p. 65).

[122] M. D. Nadai et al., *Personalized Audiobook Recommendations at Spotify Through Graph Neural Networks*, 2024, arXiv: 2403.05185 [cs.IR], URL: https://arxiv.org/abs/2403.05185 (cit. on p. 65).

[123] A. Erle et al., *Evaluating a Machine Learning Tool for the Classification of Pathological Uptake in Whole-Body PSMA-PET-CT Scans*, Tomography **7** (2021) 301 (cit. on p. 65).

[124] F. Chollet, *Deep Learning mit Python und Keras: Das Praxis-Handbuch vom Entwickler der Keras-Bibliothek*, MITP-Verlags GmbH & Co. KG, 2018 (cit. on p. 66).

[125] *Neural Networks*, 2025, URL: https://www.3blue1brown.com/topics/neural-networks (visited on 01/03/2025) (cit. on p. 66).

[126] T. Rashid, *Neuronale Netze selbst programmieren: ein verständlicher Einstieg mit Python*, O'Reilly, 2017 (cit. on p. 66).

[127] Google Brain Team, *TensorFlow*, 2025, URL: https://www.tensorflow.org/ (visited on 06/03/2025) (cit. on p. 66).

[128] F. Chollet, *Keras*, 2025, URL: https://keras.io/ (visited on 06/03/2020) (cit. on p. 66).

[129] S. Linnainmaa, *The representation of the cumulative rounding error of an algorithm as a Taylor expansion of the local rounding errors*, PhD thesis: Master's Thesis (in Finnish), Univ. Helsinki, 1970 (cit. on p. 69).

[130] S. Ruder, *An overview of gradient descent optimization algorithms*, 2017, arXiv: 1609.04747 [cs.LG], URL: https://arxiv.org/abs/1609.04747 (cit. on p. 70).

[131] G. Nakerst, J. Brennan and M. Haque, *Gradient descent with momentum — to accelerate or to super-accelerate?*, 2020, arXiv: 2001.06472 [cs.LG], URL: https://arxiv.org/abs/2001.06472 (cit. on p. 71).

[132] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, arXiv preprint arXiv:1412.6980 (2014) (cit. on p. 71).

[133] 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 **15** (2014) 1929, URL: http://jmlr.org/papers/v15/srivastava14a.html (cit. on p. 74).

[134] X. Ying, *An Overview of Overfitting and its Solutions*, Journal of Physics: Conference Series **1168** (2019) 022022, URL: https://dx.doi.org/10.1088/1742-6596/1168/2/022022 (cit. on p. 74).

[135] 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. 74).

[136] 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. 77).

[137] ATLAS Collaboration, *GoodRunLists for analysis (Run 2)*, 2025, URL: https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/GoodRunListsForAnalysisRun2 (cit. on p. 77).

[138] ATLAS Collaboration, *Luminosity determination in $pp$ collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*, Eur. Phys. J. C **83** (2023) 982, arXiv: 2212.09379 [hep-ex] (cit. on p. 77).

[139] J. Campbell et al., *Event Generators for High-Energy Physics Experiments*, (2025), arXiv: 2203.11110 [hep-ph], URL: https://arxiv.org/abs/2203.11110 (cit. on p. 78).

[140] *Recommended isolation working points (Rel. 21)*, 2025, URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/RecommendedIsolationWPs (visited on 12/03/2025) (cit. on p. 82).

[141] *Electron Charge ID Selector Tool*, 2025, URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/ElectronChargeFlipTaggerTool (visited on 12/03/2025) (cit. on p. 82).

[142] *Electron and Photon Selection and Identification for Run2 and Run3*, 2025, URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/EGammaIdentificationRun2#Using_the_Ambiguity_tool_in_anal (visited on 12/03/2025) (cit. on p. 82).

[143] T. Holm, *Towards a measurement of the tHq process in channels with hadronic tau lepton decays*, (2023) (cit. on pp. 82, 89).

[144] Y. Coadou, *Boosted Decision Trees and Applications*, EPJ Web of Conferences **55** (2013) 02004 (cit. on p. 83).

[145] P. Martinez Agullo, *Search for the Higgs boson produced in association with a top quark using $\tau$ leptons with ATLAS*, Presented 11 Mar 2024, Valencia U., 2024, URL: https://cds.cern.ch/record/2892621 (cit. on p. 83).

[146] M. Weiß, *Improvements on the mass reconstruction of single top quark associated Higgs production at ATLAS analyzing the decay channel $H \to \tau\tau$*, (2023) (cit. on pp. 88, 89).

[147] 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://dx.doi.org/10.1016/j.nima.2011.07.009 (cit. on p. 90).

[148] M. Hübner, *Effects of tau decay product reconstruction in a Higgs CP analysis with the ATLAS experiment*, PhD thesis: unpublished, MA thesis:
Mathematisch-Naturwissenschaftlichen Fakultät der . . ., 2016 (cit. on p. 90).

[149] C. Kirfel, *Neural Network-based Signal Isolation and Cross Section Estimation of the tH Process with the ATLAS Detector*,
PhD thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2024,
URL: https://hdl.handle.net/20.500.11811/11375 (cit. on pp. 93, 97, 99).

[150] F. A. J. O. Molina, *Implementation and evaluation of a categorical classifier for signal and background isolation in the tHq channel at ATLAS*, 2023 (cit. on p. 94).

[151] S. Ioffe and C. Szegedy, *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*, 2015, arXiv: 1502.03167 [cs.LG],
URL: https://arxiv.org/abs/1502.03167 (cit. on p. 99).

[152] I. Grosse et al., *Analysis of symbolic sequences using the Jensen-Shannon divergence*,
Physical Review E **65** (2002) 041905 (cit. on p. 102).

[153] H. Kantz and T. Schreiber, *Nonlinear time series analysis*, vol. 7,
Cambridge university press, 2004 (cit. on p. 103).

[154] ATLAS Collaboration, *Estimation of backgrounds from jets misidentified as $\tau$-leptons using the Universal Fake Factor method with the ATLAS detector*, 2025,
arXiv: 2502.04156 [hep-ex], URL: https://arxiv.org/abs/2502.04156
(cit. on p. 111).

[155] *TauTruthMatchingTool*, 2025, URL: https://gitlab.cern.ch/atlas/athena/-/blob/cbd64ce535b0a84a7487b9c66177a62cd53d4c2d/PhysicsAnalysis/TauID/TauAnalysisTools/doc/README-TauTruthMatchingTool.rst (cit. on p. 112).

[156] *TruthClassification*, 2025, URL: https://gitlab.cern.ch/atlas/athena/-/tree/21.2/PhysicsAnalysis/AnalysisCommon/TruthClassification
(cit. on p. 112).

[157] 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*,
Presented 08 Nov 2017, Aix-Marseille U., 2017,
URL: https://cds.cern.ch/record/2296985 (cit. on p. 137).

[158] R. D. Cousins, *Generalization of chisquare goodness-of-fit test for binned data using saturated models, with application to histograms*,
URL http://www. physics. ucla. edu/cousins/stats/cousins_saturated. pdf (2013)
(cit. on p. 159).

[159] G. Avoni et al.,
*The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*,
JINST **13** (2018) P07017, URL: https://cds.cern.ch/record/2633501 (cit. on p. 169).

[160] ATLAS Collaboration,
*Studies on top-quark Monte Carlo modelling with Sherpa and MG5_aMC@NLO*,
ATL-PHYS-PUB-2017-007, 2017, URL: https://cds.cern.ch/record/2261938
(cit. on p. 170).

[161] ATLAS Collaboration, *Studies on the improvement of the matching uncertainty definition in top-quark processes simulated with Powheg+Pythia8*, ATL-PHYS-PUB-2023-029, 2013,
URL: https://cds.cern.ch/record/2872787 (cit. on p. 170).

# List of Figures

# List of Tables

# Acronyms

**Adam** adaptive moment estimation. 71, 101

**ATLAS** A Toroidal LHC Apparatus. 2, 29, 34, 38–43, 45, 48, 49, 51, 52, 62, 63, 77, 78, 81, 88, 134, 161, 162, 166, 169, 229

**AUC** area under the curve. 73, 99, 100, 102, 106, 107

**CERN** conseil européen pour la recherche nucléaire. 2, 31, 33, 38, 229

**CKM matrix** Cabibbo-Kobayashi-Maskawa matrix. 11, 16, 19

**CNN** categorical neural network. 75, 88, 93–100, 102, 104, 105, 121, 138, 143, 156, 161, 171, 176, 182–184, 230, 231, 233

**CR** control region. 117, 121, 122, 129, 130, 136–138, 140, 142, 145, 147, 149, 150, 152, 153, 155, 230, 231

**ELU** exponential linear unit. 68, 69, 99, 229

**FPR** false positive rate. 72

**FSR** final state radiation. 62, 170

**IBL** insertable B-Layer. 41, 42

**ID** inner detector. 38–42, 44, 48, 50, 51, 229

**ISR** inital state radiation. 62, 170

**JER** jet energy resolution. 49, 169, 198–205

**JES** jet energy scale. 48, 169, 198–205

**JSD** Jenson-Shannon divergence. 102, 103, 107, 233

**KLD** Kullback-Leibler divergence. 103

**LHC** Large Hadron Collider. 16, 29, 31–34, 38–40, 43, 44, 53, 65, 77, 161, 229

**LO** leading order. 8, 20

**MC** Monte Carlo. 45, 48, 49, 53, 62, 63, 65, 77, 78, 81, 83, 88, 89, 94, 95, 99, 111–113, 134, 135, 138, 143, 147, 156, 167–170, 176, 182, 230, 231

**ML** machine learning. 2, 65, 66, 74, 83

**MMC** missing mass calculator. 90, 91, 96

**MS** muon spectrometer. 43, 44, 47, 51, 113

**NLO** next-to-leading order. 8, 10

**NN** neural network. 65–69, 72, 74, 75, 81, 83, 93–96, 229

**PDF** parton distribution function. 17, 18, 78, 170, 229

**pdf** probability density function. 54–56, 58–60, 63

**QCD** quantum chromodynamics. 3, 9, 10, 18, 62, 170

**QED** quantum electrodynamics. 3, 6, 8–10, 12

**QFD** quantum flavourdynamics. 3

**QFT** quantum field theory. 3, 6, 13

**RNN** recurrent neural network. 50, 82, 121, 129, 138, 143

**ROC** receiver operating characteristic. 72, 73, 75, 99, 102, 105–107, 183–187, 229–231

**SCT** semiconductor tracker. 41, 45

**SGD** stochastic gradient descent. 70

**SR** signal region. 121, 122, 129–131, 136, 137, 139, 140, 142, 143, 145, 147, 149, 152–155, 157, 158, 230, 231

**TPR** true positive rate. 72

**TRT** transition radiation tracker. 41, 46

**VR** validation region. 117, 121, 122, 129, 138, 143, 147, 149, 150, 152, 153, 155, 211, 216, 231, 232

# Acknowledgements

First and foremost, I would like to thank Ian C. Brock for the opportunity to write this thesis in his research group, for his continued support and guidance, even after his retirement, and for creating and refining his thesis guide, which undoubtedly saved me countless hours during the writing process. I am grateful to Klaus Desch for his immediate willingness to review my thesis, despite his undoubtedly very busy schedule. My sincere thanks also go to PD Bastian Kubis and Prof. Maja Köhn for kindly agreeing to be part of my doctoral committee.

Although I am Ian's last PhD student, I never felt like being part of a diminishing group or left on my own. This is largely thanks to the alumni of the Brock group, who always offered their help and company whenever needed, be it constructing a "Doktorwagen" or proofreading parts of my thesis. At the same time, I'm also thankful to the members of Klaus Desch's group and the many great memories we shared, whether at Jamborees or simply during short breaks. Simon and Luka, you can decide for yourselves which group you belong to. Thanks to all of you!

I especially want to thank Tanja and Oleh for introducing me to the $tH$ analysis and all the engaging discussions along the way. A big thanks to Rico for your dark humour, which helped lighten the mood during the Corona days back in my Master's. Thanks to my office mate Oliver (I suspect deep down you still consider yourself as part of the Brock group) for always being ready to turn a short question into an insightful, hour-long deep-dive discussion. I hope we will get around to planning a road trip to Brandenburg someday. And thanks, Nilima, for the time we shared in the group. Our debrief on the prancing horse's latest misery always made me look forward to Monday mornings.

It goes without saying that this thesis would still be unfinished without the help of Oliver Freyermuth and Michael Hübner. Therefore, I want to express my gratitude for their countless detailed and almost instantaneous answers to all my questions and problems, no matter how naive they may have seemed.

Thanks to all my friends – especially to Quentin, whose company during lunch breaks throughout the writing of this thesis often provided much-needed relief.

I am grateful to Annette, who not only always listened patiently when I needed to vent but also tried to navigate the intricacies of English punctuation for me.

Special thanks go to my parents for their unwavering support throughout my entire studies and life in general. Each of you, in your own unique but equally important ways, made it possible for me to complete this work.

Christian, you know full well that you're far better at writing things like this than I am – so I'll just keep it simple: *Danke für alles*.