Bergische Universität Wuppertal
Fakultät für
Mathematik und Naturwissenschaften
Fachgruppe Physik

Search for flavor-changing neutral currents via anomalous ctH and utH couplings

Thesis zur Erlangung des akademischen Grades
Master of Science (M.Sc.)
im Studiengang Physik
der Bergischen Universität Wuppertal

CERN-THESIS-2019-327
20/12/2019

vorgelegt von Oliver Thielmann

Erstgutachter: Herr Prof. Dr. Wolfgang Wagner
Zweitgutachter: Herr PD Dr. Frank Ellinghaus

Abgabedatum: December 2, 2019
Überarbeitete Version vom: February 11, 2020

Abstract

In this master thesis, flavor-changing neutral currents are investigated at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector. It is searched for two different flavor-changing neutral current processes. These processes arise either from an anomalous ctH or utH coupling. The search for them is complicated due to the large amount of background events from processes like $t\bar{t}$ and single top-quark production. In order to optimize the search for the flavor-changing neutral current processes, neural networks are used to produce a quantity, which strongly separates simulated signal events from background events. To train the neural network, kinematic properties of the signal and background processes are used.

The main part of this thesis focuses on the optimization of the neural networks. The optimized neural networks are used to determine an expected upper exclusion limit on the branching ratios of the flavor changing neutral current processes. For this, a binned profile likelihood fit is performed with the help of a statistical evaluation framework and including statistical and systematic uncertainties. The expected upper exclusion limit on the branching ratio for the left-handed processes are $\mathcal{B}(t \to uH) = 1.59 \times 10^{-3}$ and $\mathcal{B}(t \to cH) = 3.33 \times 10^{-3}$.

Contents

1 Introduction

The Standard Model of particle physics is a theory, which describes all known fundamental particles and the interactions between them up to a high precision. The most recent large success of the Standard Model was most likely the prediction of the Higgs boson by Peter Higgs in the 1960s and the discovery of the Higgs boson in 2012, which has a mass of approx. 125 GeV. Cross sections, a measure of the probability of a process, are also correctly predicted over many orders of magnitude. Nevertheless, there is the theory of general relativity, which cannot yet be unified with the Standard Model. Also, the CP violation predicted by the Standard Model is not sufficient to explain the asymmetry between matter and antimatter in the known universe. Testing the limits of the Standard Model and discovering clues to an underlying theory is one of the main tasks of particle physics today.

This thesis investigates flavor changing neutral currents (FCNC), which describe the process of an up-(down) quark being converted into a different up-(down) quark by emitting a neutral boson. Those FCNC processes are highly suppressed in the Standard Model but can be significantly enhanced by many extensions of the Standard Model (BSM theories). The top-quark-Higgs-boson FCNC sector is of special interest because the enormous mass of the top quark leads to some unique properties. On the one hand, the large mass of the top quark results in a short average lifetime, so that the top quark decays before it can hadronize. On the other hand, the coupling between the Higgs boson and other fermions is proportional to the mass of the fermions, so that the top-quark-Higgs-boson FCNC sector could be particularly pronounced.

For a better overview, this analysis is divided into several key steps:

Event selection and background estimates
: The first important step in separating the signal from the background is to find an event selection that retains as much signal as possible but suppresses the background as much as possible. The event selection used is described and explained in Section 6.

Event reconstruction
: In the signal considered, particles are formed which do not occur in most backgrounds. This leads to kinematic properties of the signal which can be well separated from the backgrounds. However, since these particles decay further, a reconstruction is necessary, which is discussed in Chapter 7.

Multivariate analysis
: To separate the signal as well as possible from the background, a neural network is used, which uses properties of the signal processes and the major background processes to generate a single variable, which separates the signal

and the background as well as possible. Further information can be found in Section 8.

Statistical analysis and interpretation

The neural network output variable is then used in different regions to perform a binned profile likelihood fit. If no excess of data over expectation is observed, exclusion limits are derived for the cross section and the branching ratio for the FCNC process. This is done in Section 11.

In addition to the key steps, the normalization of the cross section of a signal process is performed (Chapter 2.4), the advantage of a two-dimensional neural network output is investigated (Section 8.6) and the considered systematic uncertainties are discussed (Chapter 9).

2 Theoretical context

This chapter discusses the theoretical context, including more detailed information about the Standard Model. Furthermore, it is explained why FCNC are highly suppressed in the Standard Model and the way of including them in a beyond-Standard Model theory.

2.1 Standard Model

The Standard Model (SM) of elementary particle physics describes all known elementary particles and their interactions with each other. The elementary particles are divided into fermions and bosons. The fermions have spin 1/2 and are further divided into leptons and quarks with three generations each. In addition, each fermion has an anti-fermion partner with opposite charges but the same mass. Ordinary everyday matter is mostly composed of fermions of the first generation, i.e. up- and down quarks, and electrons. The fermions of the higher generations are generated mainly in particle accelerators or other high-energy collisions. Bosons have integer spin and are the exchanged particles of the three fundamental forces: The photon mediates electromagnetism, Z^0 and $W^\pm$ the weak force and gluons the strong force. The Higgs boson is a quantum excitation of the Higgs field.

Mathematically, the SM is a renormalizable gauge quantum field theory. The Lagrange density of the SM depends on fermion fields ψ, electroweak boson fields W_1, W_2, W_3 and B, the gluon fields G_a and the Higgs field ϕ. Within the Lagrange density, the allowed interactions between the elementary particles are realized by interactions of the fields. The gauge group of the SM is SU(3) $\bigotimes$ SU(2) $\bigotimes$ U(1), where SU(3) acts on G_a, SU(2) acts on W_1, W_2, W_3 and ϕ, and U(1) acts on B and ϕ. After the spontaneous symmetry breaking, mass terms for the W- and Z boson and the fermions arise. Figure 2.1 provides an overview of the particles of the SM and their interactions.

Interactions between the fermions and the bosons are indicated by grayed out backgrounds.

2.1.1 Quark mixing in the Standard Model

In the SM, up-type quarks (positively charged quarks) can be converted to down-type quarks (negatively charged quarks) and vice versa while emitting a W boson such that charge conservation is given. Conversion of an up-type quark (down-type

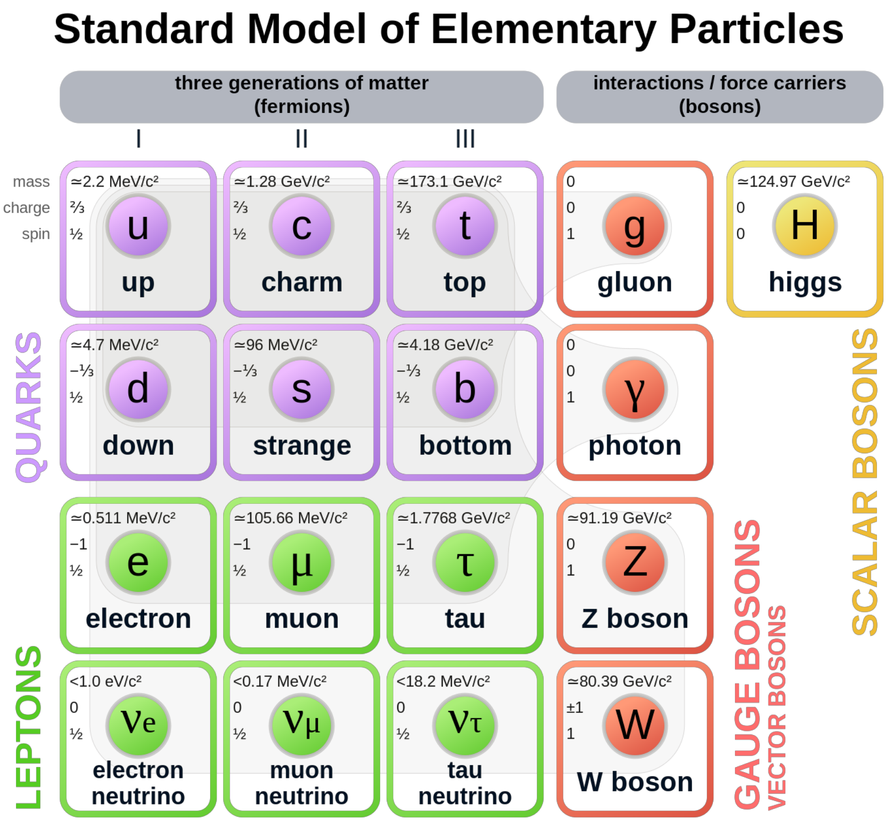


Figure 2.1: *Summary of all known elemental particles in the SM and their interactions with each other. [1]*

quark) into an up-type quark (down-type quark) of a different generation by emitting an electrically neutral boson is prohibited in the SM on tree level[1].

Since the mass eigenstates of the quarks do not correspond to the eigenstates of the weak interaction, the mass eigenstates of the quarks are mixed by the so-called CKM matrix (short for Cabibbo–Kobayashi–Maskawa) after an interaction with a $W^{\pm}$ boson:

$$\begin{bmatrix} d' \\ s' \\ b' \end{bmatrix} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} \cdot \begin{bmatrix} d \\ s \\ b \end{bmatrix} \tag{2.1}$$

where d', s' and b' are eigenstates of the weak interaction and d, s and b are mass eigenstates of the down-type quarks. $V_{i,j}$ are the entries of the (complex unitary)

[1]The tree level of a process denotes the leading order of the process in perturbation theory.

CKM matrix. The absolute values of the CKM matrix can be seen in Equation (2.2) [2]:

$$(|V_{i,j}|) = \begin{bmatrix} 0.97427 & 0.22534 & 0.00351 \\ 0.22520 & 0.97344 & 0.0412 \\ 0.00867 & 0.0404 & 0.999146 \end{bmatrix} \tag{2.2}$$

As one can see, the CKM matrix has entries close to 1 on the diagonal. This means that an up-type quark decays primarily into the down-type quark of the same generation and vice versa. This behavior is especially pronounced for the top quark, which almost exclusively decays into a W boson and a bottom quark.

2.1.2 Special aspects of the top quark and the Higgs boson

Since the top quark has by far the largest rest mass in the SM, it has some characteristics that distinguish it from the other elementary particles. On the one hand, a large rest mass is associated with a short average lifetime, which in the case of the top quark is about $5 \times 10^{-25}\,\mathrm{s}$ [3]. On the other hand, the mass of the fermions is proportional to the coupling strength to the Higgs boson.

The short average lifetime of the top quark means that the top quark does not form hadrons because it most likely decays before. This makes it possible to study the top quark itself, and not just hadrons containing a top quark. In addition, the fact that the top quark does not hadronize simplifies identification in the detector and also simplifies the theoretical calculation. The large rest mass of the top quark also means that it plays an important role in BSM theories. The effect of BSM theories will be stronger with higher energy under consideration. Since the top quark and the Higgs boson are the particles with the highest masses in the SM, they have a special role in BSM searches.

2.2 Flavor-changing neutral currents

FCNC describe the process that an up-type quark (down-type quark) converts into an up-type quark (down-type quark) of a different generation, emitting a neutral boson. These processes do not exist at tree level in the SM. In addition, FCNC are highly suppressed in the SM beyond the tree-level by the Glashow-Iliopolus-Maiani (GIM) mechanism [4], which states that the contributions of the possible Feynman diagrams are almost completely canceled by each other due to the properties of the unitary CKM matrix. Nevertheless, FCNC can be strongly enhanced in BSM theories (see Figure 2.2), making the search for them an interesting aspect of high-energy physics. BSM theories can increase the probability of FCNC processes by either abandoning flavor conservation (FC) (for example as in the 2-Higgs-Doublet model (2HDM) (**F**lavor **V**iolation) theory) or introducing FCNC vertices with new or already known particles that would allow the FCNC process. Figure 2.2 provides an overview of the upper limits on top-quark FCNC processes in the SM, expressed in terms of the branching ratios of the FCNC decays. Furthermore, previous searches and possible extensions of the SM are shown.

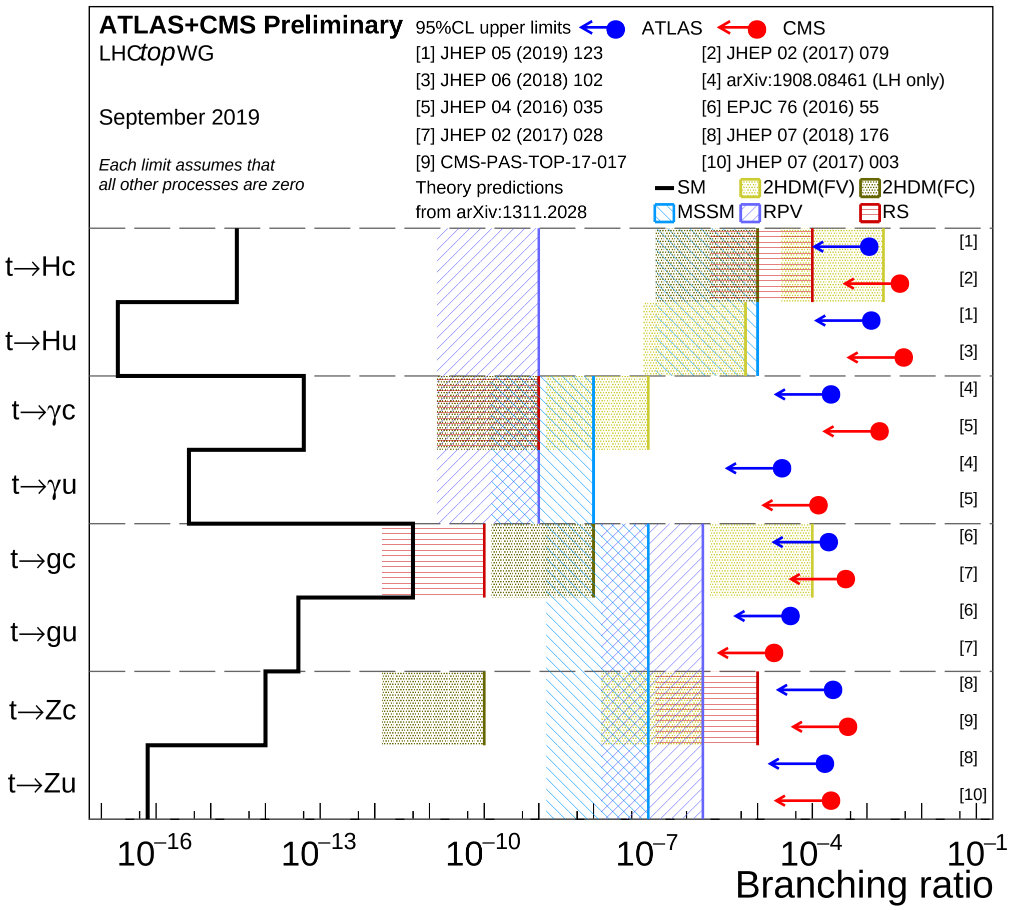


Figure 2.2: *Summary of the current 95% confidence level observed limits on the branching ratios of the top quark decays via flavour changing neutral currents to a quark and a neutral boson t→Xq (X=g, Z, γ or H; q=u or c) by the ATLAS and CMS Collaborations compared to several new physics models [5].*

As can be seen, according to some extensions of the SM, an FCNC signal could already be measurable. However, no FCNC process has been discovered so far.

2.3 Effective field theory

An effective field theory (EFT) is used to describe and simulate the FCNC processes. The advantage of an EFT is that knowledge of the fundamental theory is not needed, and potential new physics can be described by operators of the EFT. Equation (2.3) shows the general Lagrangian of an EFT whereas the used framework only uses the first order beyond the SM.

$$\mathcal{L}_{EFT} = \mathcal{L}_{SM} + \sum_i \frac{C_i}{\lambda^2}\mathcal{O}_i \tag{2.3}$$

where $\mathcal{L}_{EFT}$ is the Lagrangian of the EFT model, $\mathcal{L}_{SM}$ is the SM Lagrangian, Λ is a mass scale beyond where new physics is expected, C_i are dimensionless coefficients and $\mathcal{O}_i$ are dimension-i operators. The dimension-i operators contain the information about the possible BSM vertices, the dimensionless coefficients, also called Wilson coefficients, influence the strength of the BSM vertices.

This EFT is implemented in the `TopFCNC` FeynRules [6,7] model using dimension-six operators. In combination with the `MadGraph5_aMC@NLO` [8] Monte Carlo (MC) generator, it is possible to make next-to leading order (NLO) calculations in QCD. The coefficients, which correspond to vertices which are considered in this analysis, are listed in Table 2.1. The corresponding Feynman diagram of the FCNC-vertex is shown in Figure 2.3.

Table 2.1: *Summary of Wilson coefficients which correspond to vertices, which are considered in this analysis. The 6D block parameter is used for internal naming inside the configuration files for* `MadGraph5_aMC@NLO`.

6D block parameter	6D coefficient	Comment
RCtphi	$C_{u\phi}^{13}$	Left-handed utH coupling
RCuphi	$C_{u\phi}^{31}$	Right-handed utH coupling
RCtcphi	$C_{c\phi}^{23}$	Left-handed ctH coupling
RCctphi	$C_{c\phi}^{32}$	Right-handed ctH coupling

The FCNC vertex, which is shown in Figure 2.3, is considered in this analysis in two different processes. These processes have overlapping end states, so both processes are taken into account. The two processes can be seen in Figure 2.4. From now on, the left process in Figure 2.4 is called “production-FCNC process” and the right process is called “decay-FCNC process”.

The further decay of the two processes and the resulting selections are discussed in detail in Chapter 6.

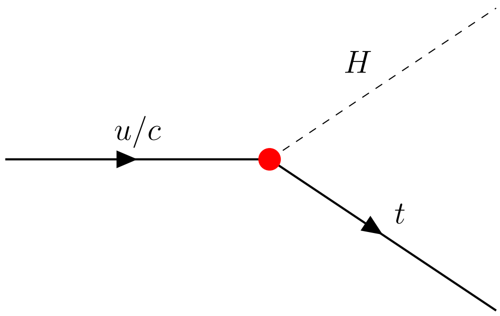


Figure 2.3: Feynman diagram for a possible process, involving the FCNC vertex, colored in red.

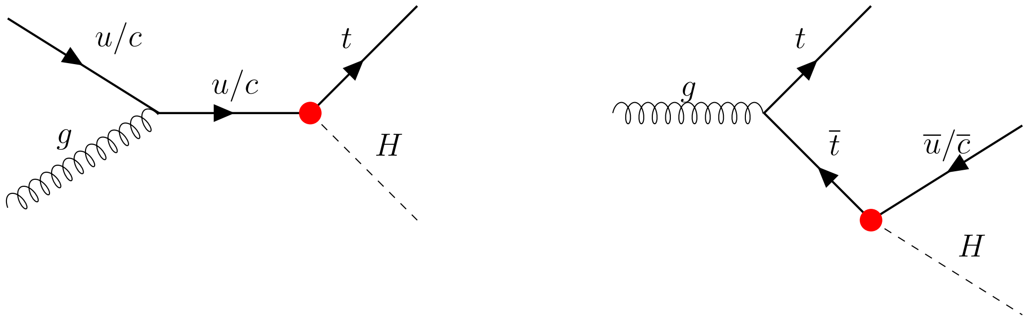


Figure 2.4: Feynman diagram of the considered production-FCNC process on the left and the considered decay-FCNC process on the right. The FCNC vertex is colored in red.

2.4 Signal sample normalization

When using `MadGraph5_aMC@NLO` and `MadSpin` to simulate the signal-FCNC processes, problems occurred with the decay-FCNC process. While the simulation of the production-FCNC process is possible without problems with a Wilson coefficient of $|C_{u\phi}| = |C_{c\phi}| = 1$, this was not the case for the decay-FCNC process. However, in order to consider both processes simultaneously in the analysis, both processes must be normalized to the same Wilson coefficient. The normalization of the decay-FCNC process to a Wilson coefficient of $|C_{u\phi}| = |C_{c\phi}| = 1$ is discussed in the following sections.

2.4.1 Limitations of the decay-FCNC sample generation

The computation of the $t\bar{t}, t \to qH, \bar{t} \to W^-\bar{b}$ cross section (the corresponding process $t\bar{t}, \bar{t} \to \bar{q}H, t \to W^+b$ will be discussed later) is performed using `MadGraph5_aMC@NLO` and depends on the Wilson coefficients $|C_{u\phi}|$ and $|C_{c\phi}|$. Limits on these coefficients can be interpreted as limits on the branching ratio $\mathcal{B}(t \to qH)$.

With the production cross section of $t\bar{t}$, the cross section for the decay-FCNC process can be calculated:

$$\sigma_{decay-FCNC} = \sigma_{t\bar{t}} \cdot \mathcal{B}(t \to qH) \cdot \mathcal{B}(t \to Wb), \tag{2.4}$$
$$= \sigma_{t\bar{t}} \cdot \mathcal{B}(t \to qH) \cdot (1 - \mathcal{B}(t \to qH))\,, \tag{2.5}$$

where $\mathcal{B}(t \to qH)$ is defined as

$$\mathcal{B}(t \to qH) = \frac{\Gamma_{t\to qH}}{\Gamma_{t\to qH} + \Gamma_{t\to Wb}}, \tag{2.6}$$

with $\Gamma_{t\to qH}$ and $\Gamma_{t\to Wb}$ being the partial decay width for the $t \to qH$ and $t \to Wb$ process respectively.

Since the partial decay width $\Gamma_{t\to qH}$, if calculated with Wilson coefficients smaller than 12, is smaller than the QCD-scale, the decay width is automatically set to zero by `MadSpin`. This is done by `MadSpin` to prevent unphysical results. Following Equation (2.6) and (2.5), a partial decay width of zero results in a branching ratio of zero and subsequently in a cross section for the decay-FCNC process of zero. But since it is expected to have a cross section for arbitrary values for the coefficients, the normalization of the cross section of the decay-FCNC sample is required.

Changing the coefficients $|C_{u\phi}|$ and $|C_{c\phi}|$ only changes the cross section of the processes, not the kinematic behaviour. It is thus possible to produce the decay-FCNC samples with $|C_{u\phi}|= |C_{c\phi}|= 12$ and afterwards normalize the cross section to a value, which corresponds to the coefficients $|C_{u\phi}|= |C_{c\phi}|= 1$. The procedure for normalizing the decay-FCNC sample is discussed in the following sections.

2.4.2 Cross section of the production-FCNC process

The theoretical dependence of the cross section for the production-FCNC process as function of the coefficient C can be derived as follows:

$$\mathcal{L} = a' \cdot C + b', \qquad \mathcal{M} \propto \mathcal{L}, \qquad \sigma \propto |\mathcal{M}|^2, \tag{2.7}$$

with $\mathcal{L}$ being the Lagrangian of the effective field theory, $\mathcal{M}$ being the matrix element for the process and a,b,d being arbitrary constants. It follows that

$$\sigma_{prod-FCNC}(C) = a \cdot C^2 + b \cdot C + d. \tag{2.8}$$

Requiring $\sigma(C = 0) = 0$ and $\frac{d\sigma}{dC}\big|_{C=0} = 0$, it can be concluded that

$$\sigma_{prod-FCNC}(C) = a \cdot C^2. \tag{2.9}$$

Figure 2.5 shows the distribution of the cross section, calculated by `MadGraph5_aMC@NLO` with varying Wilson coefficients and which is fitted by the theory expectation in Equation (2.9). There occurs no technical problem while generating this process with small Wilson coefficients because `MadGraph5_aMC@NLO` is used to calculate the matrix element.

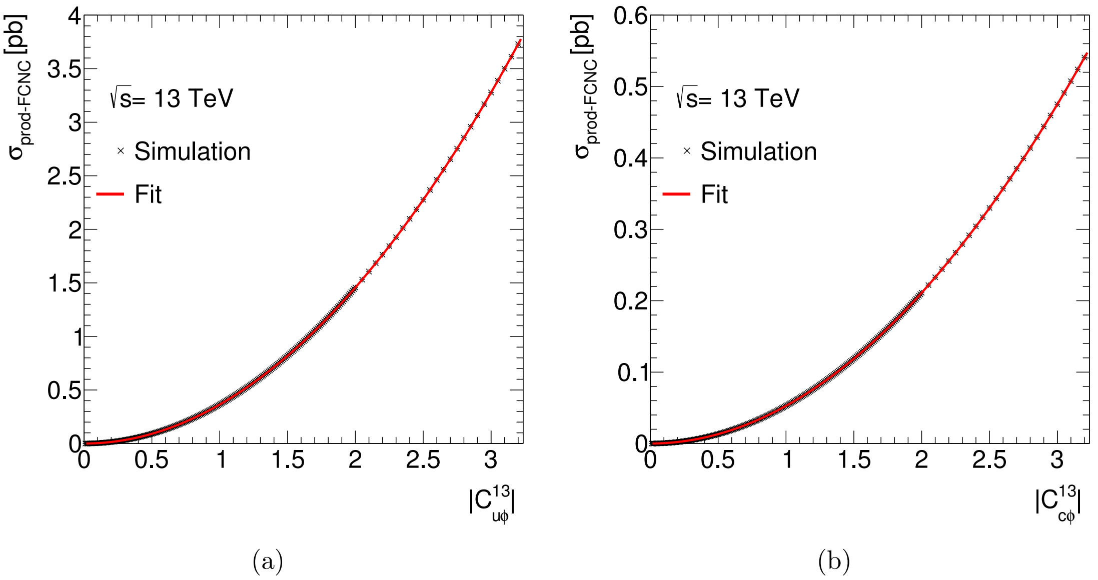


Figure 2.5: *Cross section of the production-FCNC process, depending on the Wilson coefficient used. The Wilson coefficients used correspond to the left-handed utH coupling (a) or to the left-handed ctH coupling (b) respectively.*

Within the uncertainties of the generator and the fit, no difference is found between the cross section of the left-handed and right-handed coupling.

2.4.3 Cross section of the decay-FCNC process

The cross section of the decay-FCNC process is determined by Equation (2.5), thus the cross section is limited by the $t\bar{t}$ cross section.

Starting from Equation (2.5) and inserting Equation (2.6), the following can be derived

$$\sigma_{decay-FCNC} = \sigma_{t\bar{t}} \cdot \frac{\Gamma_{t\to qH}}{\Gamma_{t\to qH} + \Gamma_{t\to Wb}} \cdot \left(1 - \frac{\Gamma_{t\to qH}}{\Gamma_{t\to qH} + \Gamma_{t\to Wb}}\right) \tag{2.10}$$

$$= \sigma_{t\bar{t}} \cdot \frac{\Gamma_{t\to qH} \cdot \Gamma_{t\to Wb}}{(\Gamma_{t\to qH} + \Gamma_{t\to Wb})^2}. \tag{2.11}$$

With $\Gamma_{t\to qH} \propto C^2$, it follows that

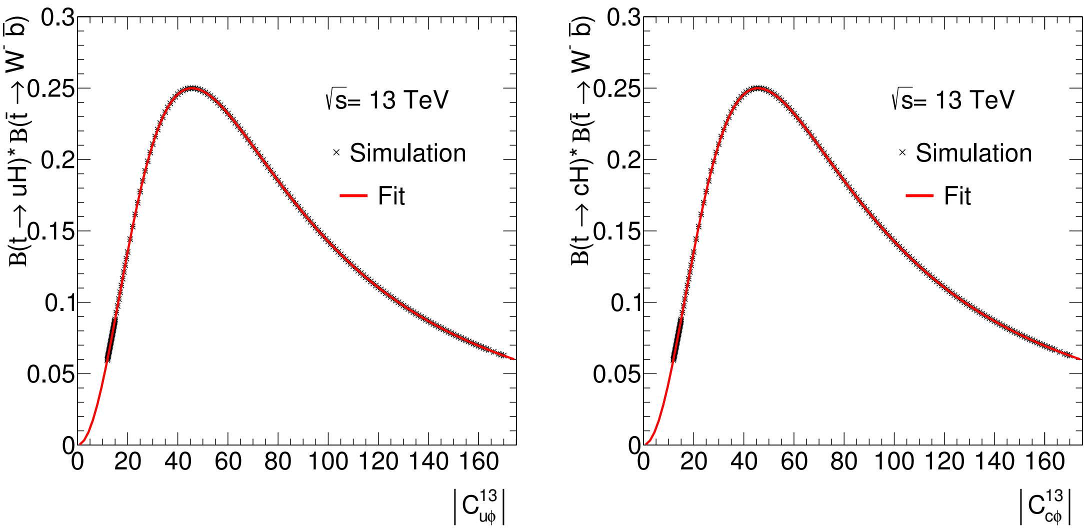
$$\sigma_{decay-FCNC}(C) = \sigma_{t\bar{t}} \cdot \frac{aC^2 \cdot b}{a^2C^4 + 2abC^2 + b^2}. \tag{2.12}$$

with a and b being constants. In contrast to the production-FCNC case, the `MadSpin` package is used to handle the top-quark decay. Since the FCNC decay is a rare process, the analysis considers the case where one top quark decays using the FCNC vertex, but the other top quark according to the SM. Using `MadSpin` with varying coupling constant generates a branching ratio for the $t \to qH, \bar{t} \to W^-\bar{b}$ process. These branching ratios are plotted in Figure 2.6 for the left-handed $q = c, u$ coupling.

Again, within the uncertainty, there is no difference between left-handed and right-handed coupling. In contrast to the cross section of the production-FCNC process, there is a maximum for the cross section of the decay-FCNC process. In addition, the cross section goes towards zero for large Wilson coefficients. This is because one top quark is required to decay using the FCNC coupling, the other, however, according to the SM. For very large Wilson coefficients, it is likely that the second top quark will also decay by FCNC, such that the product of branching ratios $\mathcal{B}(t \to qH) \times \mathcal{B}(\bar{t} \to W^-\bar{b})$ will decrease. The simulated Wilson coefficients in Figure 2.6 are orders of magnitude above the current upper exclusion limit for the FCNC process considered. The simulation of the large FCNC values should confirm the derived formula for the branching ratio, which was successful. In addition, many data points reduce the error of the fit so that a better extrapolation can be performed.

2.4.4 FCNC cross section ratio

This section covers the ratio of the production-FCNC process and the decay-FCNC process. Using Equation (2.9) and (2.12), the following proportionality can be derived:



(a) (b)

Figure 2.6: *Branching ratio for the process $t \to qH, \bar{t} \to W^-\bar{b}$, depending on the Wilson coefficients used. The Wilson coefficients used correspond to the left-handed utH coupling (a) or to the left-handed ctH coupling (b) respectively.*

$$\frac{\sigma_{prod-FCNC}}{\sigma_{decay-FCNC}}(C) \propto \frac{C^2 \cdot (C^2 + a')^2}{b' \cdot C^2 \cdot d'} = a \cdot C^4 + b \cdot C^2 + d. \tag{2.13}$$

For small C, the ratio becomes approximately constant:

$$\left.\frac{\sigma_{prod-FCNC}}{\sigma_{decay-FCNC}}(C)\right|_{C\approx 0} \approx d. \tag{2.14}$$

The ratio of cross sections for varying Wilson coefficients can be seen in Figure 2.7.

The approximately constant ratio between both processes for small values of the Wilson coefficient is used in the evaluation in Chapter 11.4.

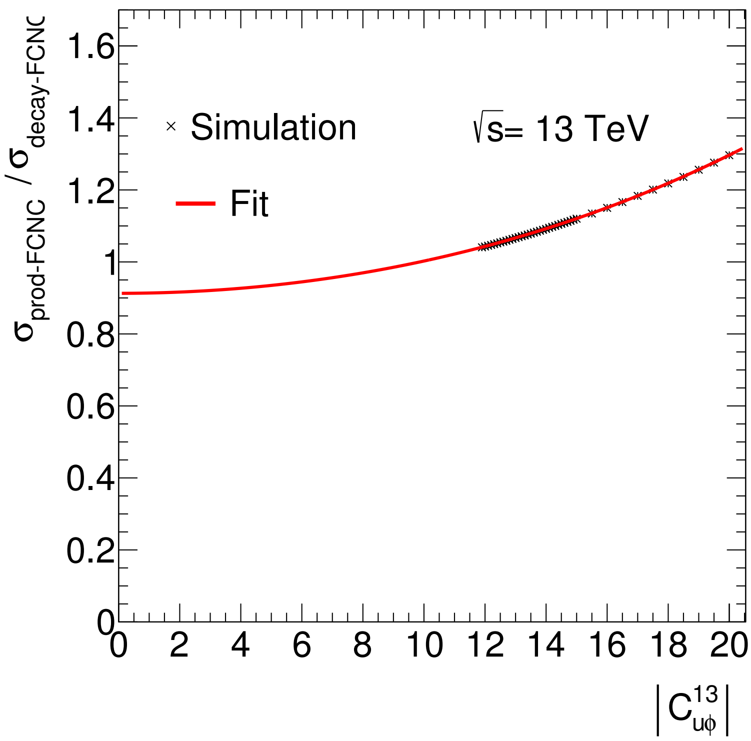


Figure 2.7: *Ratio of the production-FCNC cross section, divided by the decay-FCNC cross section, with left-handed utH-coupling enabled.*

2.4.5 Theoretical calculation of the $t \to qH$ partial decay width

Besides computing the cross section of the decay-FCNC process with the numerical tool `MadSpin`, it is also possible to calculate the partial decay width at leading order [9]:

$$\Gamma^{(0)}_{t\to qH} = \frac{|C_{q\phi}|^2}{\Lambda^4}\frac{\sqrt{2}G_F m_t^7}{8\pi}\left(1-\frac{m_H^2}{m_t^2}\right)^2 . \tag{2.15}$$

With

$$\begin{aligned}\Lambda &= 1000\,\text{GeV},\\ G_F &= 1.166\,37 \times 10^{-5}\,/\text{GeV}^2,\\ m_t &= 172.5\,\text{GeV},\\ m_h &= 125\,\text{GeV},\\ \Gamma_{t\to Wb} &= 1.4526\,\text{GeV},\end{aligned}$$

the branching ratio for $t \to qH$, using Equation (2.6) can be calculated. Note that the authors of the papers mentioned next-to leading order corrections in the order of magnitude of 10% and that the equation assumes, that the up-quark and the charm-quark masses are negligible compared to the top-quark mass.

2.4.6 Cross section extrapolation at $|C_{q\phi}|^2 = 1$

If the results of the simulations of the previous sections are used, a cross section of

$$\sigma_{decay-FCNC} = 0.3987\,\text{pb} \tag{2.16}$$

is obtained for the left/right-handed process $pp \to t\bar{t}, t \to qH, \bar{t} \to W^-\bar{b}$ for a Wilson coefficient of $|C_{q\phi}|^2 = 1$.

This result is consistent with that of the theoretical approach from equation 2.15, which is $\sigma_{decay-FCNC,theo} = 0.3846\,\text{pb}$.

In addition to the cross sections, the selection efficiency is also needed to calculate the expected events. Since it can be assumed that the selection efficiency does not depend on the Wilson coefficient, the selection efficiency is calculated with a coefficient of $|C_{q\phi}| = 12$ and then applied to the coefficient $|C_{q\phi}| = 1$.

The resulting event yield can be seen in Table 6.1. It must be noted, that for the event yield the expected events for the decay-FCNC process have been doubled, compared to the expectation for the cross section in Equation (2.16). This is due to the fact that in the cross section calculation above, only the process $pp \to t\bar{t}, t \to qH, \bar{t} \to W^-\bar{b}$ has been considered. However, there is also the process where the top antiquark decays via FCNC and the top quark decays according to the SM. These two processes have the same probability in the used EFT model so that one had to double the expected events for the decay-FCNC process.

3 LHC and the ATLAS experiment

The following chapter describes the basic functionality of the Large Hadron Collider and the ATLAS experiment.

3.1 LHC

The Large Hadron Collider (LHC) is a particle collider with a designed center-of-mass energy of up to 14 TeV. After the end of the Large Electron-Positron (LEP) experiment, the accelerator ring with a circumference of about 27 km was rebuilt to allow proton-proton (and heavy-ion) collisions. With the help of several pre-accelerators and the LHC itself (see Figure 3.1), the protons (heavy ions) are accelerated in opposite directions to almost the speed of light and brought to collision at four different positions in the ring.

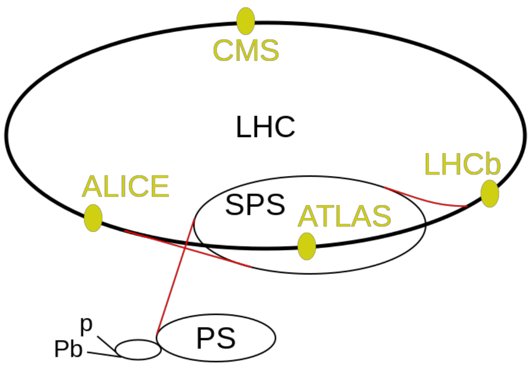


Figure 3.1: The LHC along with its pre-accelerators. [10]

The charged particles are accelerated in two separated vacuum pipes in opposite directions. The separated pipes are necessary to have a magnetic field with opposite direction at the location of the proton beams. Furthermore, the separated pipes ensure that collisions only occur in the experiments. The charged particles themselves are accelerated in bunches within radio-frequency (RF) cavities. If a particle (with already nearby speed of light) in a bunch has more energy than the nominal particle, the higher momentum will result in an increased orbit radius. Due to the increased orbit radius and, however, still approximately the same speed, a particle with higher energy needs longer for one revolution so that it will arrive at the RF cavity shortly after the particle with the nominal energy. This can be compensated if the bunches are accelerated on the falling flank of the electromagnetic wave inside the RF-cavity, where delayed particles receive less energy than the nominal particle. The opposite is true for particles with less energy than the nominal particle: These particles have a smaller orbit radius so that they need less time for one revolution and receive more energy in the RF-cavity. Thus the bunches are held together in the longitudinal direction. Dipole magnets are used to force the bunches onto the circular path. To hold the bunches together in transversal direction alternating quadrupole magnets are used. In a single quadrupole magnet, one transverse direction is focusing but the other transverse direction is defocusing at the same time. The alternating configuration of the quadrupole magnets focuses the previously defocused plane in the following quadrupole and defocuses the previously focused plane. However, as the field strength increases to the poles of the magnets, charged particles that are further away from the nominal particle are focused stronger. Thus, multiple alternating quadrupole magnets create an overall focusing effect on the entire bunch in the transverse direction. In addition to dipole and quadrupole magnets, magnets of even higher order are also used to correct for small imperfections.

The Luminosity L is a measure of the number of interactions per time and area unit at an accelerator and is defined by Equation (3.1).

$$L = \frac{n \cdot N_B^2 \cdot f}{4\pi\sigma_x\sigma_y} \tag{3.1}$$

with N_B being the number of protons per bunch, n the number of bunches which collide with the frequency f and σ_x and σ_y being the transverse profiles of the bunches. In general, this instantaneous luminosity must be corrected because the bunches collide at a certain angle in the detectors. If the instantaneous luminosity is integrated over a certain period of time, the integrated luminosity $\int L\ dt$ is obtained. This is a measure of how many interactions have taken place in the period under consideration. Using the cross section σ of a process the expected number of produced events in an accelerator N for a given integrated luminosity can be calculated according to Equation (3.2).

$$N = \sigma \cdot \int L\ dt \tag{3.2}$$

This analysis uses the ATLAS data set from 2015-2018 with a center-of-mass energy of 13 TeV, which corresponds to an integrated luminosity of $\int L\ dt = 139\,\mathrm{fb}^{-1}$.

Following Equation (3.2), the large amount of recorded data makes it possible to search for rare processes with a small cross section.

3.2 The ATLAS experiment

The ATLAS experiment [11] is, alongside CMS, Alice and LHCb, one of the four major experiments at the LHC. While Alice is focused on lead ion collisions and LHCb on the measurement of hadrons containing b quarks, CMS and ATLAS are general-purpose particle detectors designed for the measurement of high-energy proton-proton collisions.

The ATLAS detector consists of an inner detector (ID) for tracking, surrounded by a superconducting solenoid, followed by an electromagnetic and hadronic calorimeter and finally a muon spectrometer including superconducting toroid magnets. The ID consists of several layers of silicon pixel detectors, a silicon microstrip tracker and a transition radiation tracker. Each layer of detector should ideally detect electrically charged particles in order to reconstruct their trajectory. The innermost layer, and thus the layer closest to the beam pipe, is the Insertable B-layer [12] (IBL) and was subsequently installed in 2014 to increase the tracking performance. The silicon microstrip tracker (SCT) and the transition radiation tracker (TRT) are used to further improve the particle identification and trajectory reconstruction. The ID is surrounded by a 2 T axial magnetic field which bends the trajectories of electrically charged particles depending on their momentum.

The electromagnetic and hadronic calorimeter measure the energy of electrons/photons and hadrons respectively in a range of $|\eta| < 4.9$ [1]. The electromagnetic calorimeter consists of alternating layers of lead/stainless steel and liquid argon. The layers of lead/stainless steel have a short radiation length so that the showers are better developed. The secondary electrons ionize the liquid argon so that an electrical signal can be measured at copper electrodes. The measured signal is proportional to the energy of the showers, so after calibrating the electrical signals the energy of the showers can be determined. The hadronic calorimeter works analogously, but the barrel consists of alternating layers of steel and scintillators and the endcaps of layers of copper and liquid argon.

Further outside the calorimeters is the muon spectrometer (MS) together with the toroid magnets. The MS is used to identify muons which are not absorbed by the

[1] In particle physics, the pseudorapidity $\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]$ is preferred over the polar angle θ because the flow of particles in hadron-hadron collisions is approximately constant per pseudorapidity.

electromagnetic calorimeter and thus deflected by the toroid magnets. Figure 3.2 shows a section through the ATLAS detector in the transverse plane.

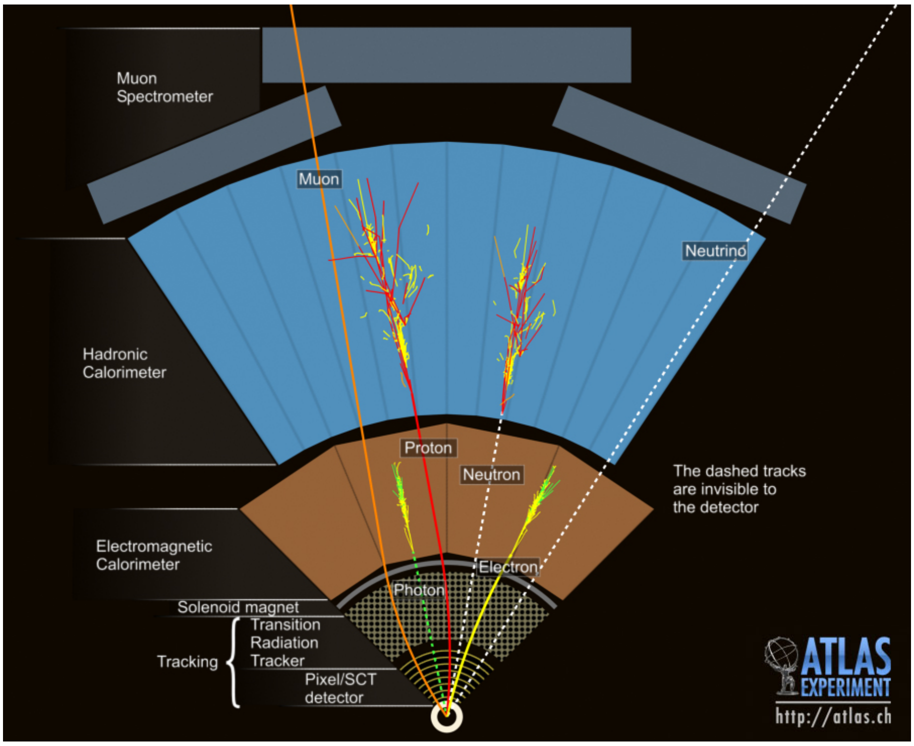


Figure 3.2: *Section through the ATLAS detector in the transverse plane. [13]*

It can be well seen, how different particles influence the response from the detector, depending e.g. on the electric charge or whether its a hadron or not.

In order to reduce the number of events to be recorded, a 2-level trigger system [14, 15] is used, which sorts out supposedly uninteresting events. The first trigger stage uses only a part of the total recorded information to make a decision as fast as possible and to reduce the frequency of the recorded events from 40 MHz to 100 kHz. The second trigger stage uses all recorded information and lowers the frequency of the events to be recorded to about 1 kHz. The remaining events are stored offline for further analysis.

The ATLAS detector uses a right-handed coordinate system, where the coordinate origin is in the interaction point. The z-axis is tangential to the beam pipe, the x-axis is directed to the center of the accelerator ring and the y-axis is directed upwards. The transverse plane (i.e. the x-y-plane) is of special interest (see Section 5.5).

4 Monte Carlo event generation

In order to compare data recorded by the ATLAS detector with the theoretical prediction, events are generated using MC event generators. The complicated process of simulating events is divided into different steps.

First the matrix element of the respective process has to be calculated to a certain order in perturbation theory. For this, either `Madgraph5_aMC@NLO` [8], `Powheg-Box` [16–19] or `Sherpa` [20] is used. After the matrix element has been determined, a shower simulation is used to simulate further radiation of particles and the hadronization of quarks, using the parton-shower generators `Pythia` [21] or `Sherpa` [20]. To simulate the particles' interaction with the detector material and the response of the detector, `GEANT4` [22] is used.

The following table summarizes the matrix element generators and the parton-shower generators, which were used for the different processes. The matrix element generators calculate the matrix element in next-to leading order (NLO) precision in QCD. For the FCNC processes, the $t\bar{t}$ process and the single top-quark production, the decay of the top quarks is simulated at leading order using `MadSpin` [23, 24] to preserve all spin correlations.

Table 4.1: *Summary of the generators , used for the matrix element generation and the parton shower + hadronization simulation of the considered processes.*

process	Matrix element	Parton shower + hadronization
prod-FCNC process	`Madgraph5_aMC@NLO`	`Pythia`v8.230
decay-FCNC process	`Powheg-Box` v2	`Pythia`v8.230
$t\bar{t}$	`Powheg-Box` v2	`Pythia`v8.230
single top-quark production	`Powheg-Box` v2	`Pythia`v8.230
W+jets	`Sherpa` v2.2	`Sherpa` v2.2
Z+jets	`Sherpa` v2.2	`Sherpa` v2.2
Diboson	`Sherpa` v2.2	`Sherpa` v2.2

The inclusive $t\bar{t}$ NLO cross-section is normalized to the next-to-next-to leading order (NNLO) prediction in QCD, including the resummation of next-to-next-to leading log (NNLL) soft-gluon terms [25].

5 Object definitions and overlap removal

This section describes the reconstruction and selection procedures of electrons, muons, jets and missing transverse momentum. These objects will be used in this analysis.

5.1 Electrons

Electrons created by a process in the final state are perceived as electron candidates in the detector. If an object leaves tracks in the ID and additionally generates isolated energy clusters in the electromagnetic calorimeter without leaving a signature in the MS and without exceeding a threshold of energy deposits in the hadronic calorimeter, this object is treated as an electron candidate [26,27]. These candidates must have a transverse momentum $p_T > 10\,\text{GeV}$ and a pseudorapidity of $|\eta|< 2.47$, excluding the region of $1.37 < |\eta| < 1.52$ due to the design of the electromagnetic calorimeter. After the electron candidates have been defined, multivariate algorithms and a likelihood-based quality criterion are used to achieve a high electron selection efficiency and at the same time a high rejection rate for fake electrons. Fake electrons are objects, which are wrongly classified as electrons. During this selection process, a distinction is made between electrons being more or less likely fake electrons. These two classes of electrons are called "loose" and "tight" electrons. This analysis uses tight electrons, which uses a TightLH quality criteria and the gradient isolation. Furthermore, a veto against loose electrons is used, which uses LooseAndBLayerLH quality criteria and no isolation. To ensure that the electron comes from the primary vertex and is not generated by decays of other particles, further cuts to $|d_0^{BL}|< 5\,\text{mm}$ and $|z_0^{BL}\sin(\theta)|< 0.5\,\text{mm}$ [1] are applied. The energy measured in the electromagnetic calorimeter is calibrated by applying scales to the measured data. These scales are validated by MC to data comparisons in $Z \to ee$ resonances [28].

5.2 Muons

Similarly to the electrons, muons are perceived as muon candidates in the detector. If an object leaves tracks in the ID and additionally leaves a signature in the MS,

[1] $|d_0^{BL}|$ ($|z_0^{BL}|$) is the shortest distance between the extrapolated trajectory and the z-axis in the transversal (longitudinal) direction.

this object is treated as a muon candidate [29]. These candidates must have a transverse momentum $p_T > 10\,\text{GeV}$ and a pseudorapidity of $|\eta| < 2.5$. Similarly to the electrons, quality and isolation criteria are used to achieve a high muon selection efficiency and simultaneously a high fake muon rejection rate. During this selection process, a distinction is made between muons being more or less likely fake muons, which are then called “loose” or “tight” muons. The tight condition are used for this analysis for selecting muons, which use the medium quality and the gradient isolation. Loose muons are defined with the loose quality criteria and no isolation. To ensure that the muon comes from the primary vertex, further cuts to $|d_0^{BL}| < 3\,\text{mm}$ and $|z_0^{BL}\sin(\theta)| < 0.5\,\text{mm}$ are applied. The energy scale calibration of the muons is performed analogously to that of the electrons.

5.3 Jets

Unlike leptons, quarks are not measured as such in the detector. Due to the nature of the strong interaction, quarks (and also gluons) cannot exist as free particles but form colorless hadrons. This principle is also known as confinement. Quarks are perceived in the detector as a “jet” of energy deposits, which must then be reconstructed. Jets are reconstructed from topological clusters [30]. These topological clusters are built from deposits in the calorimeters using the anti-k_T algorithm [31] with the radius parameter $R = 0.4$. This radius parameter has an influence on the shape of the jets. Jets must have a transverse momentum of $p_T > 30\,\text{GeV}$ and a pseudorapidity of $|\eta| < 2.5$ in this analysis. To reduce the contribution of jets, arising from pileup, the Jet Vertex Tagger (JVT) discriminant is used. The JVT discriminant must be greater than 0.59 for jets with $p_T < 120\,\text{GeV}$ and $|\eta| < 2.4$ [32]. For the energy calibration of the jets, several corrections are made, which are determined by MC-to-data comparisons [33]. These corrections include a calibration of the four-momentum of the reconstructed jet to the simulated particles inside the jet.

5.4 b-Jets

Hadrons, which consist of a b quark or a c quark, behave differently compared to hadrons consisting only of light quarks. This allows to state with a certain probability, which flavor the quark had, which initiated the jet. Among other things, b hadrons have a longer lifetime before they decay, compared to c hadrons or light hadrons, allowing them to fly on average a few millimeters before decaying. This leads to a secondary vertex that has a measurable distance from the primary interaction point. In addition to the secondary vertex, other properties of the jets, such as the topology inside the jets or the impact parameter of charged-particle tracks, are also investigated. To classify jets as b jets, the used b-tagging algorithm "MV2c10" [34] uses a neural network, which combines the different properties of the jets into a single score. This score is then cut into four bins (2,3,4,5). These bins correspond to the selection efficiencies (85%,77%,70%,60%) of the b jets in $t\bar{t}$ events. If a jet is classified into bin i by the algorithm, the properties of this jet correspond to a score of the b-tagging algorithm, which places the jet at the upper X_i percent of all b jets. In this analysis, the b-jet efficiency operating point, also called "working point" (WP), for selecting b jets is at 85%. In addition to the information whether a jet was classified as b jet by the b-tagging algorithm or not, the b-tagging score of the jet (in form of the efficiency bin) is also further analyzed. This method is known as "pseudo-continuous b-tagging". This method has the advantage over only a fixed working point that more information is retained so that a suitable analysis strategy can ideally benefit from this.

5.5 Missing transverse momentum

In the reference frame of the ATLAS detector, protons have nearly all their momentum in the longitudinal direction before the collision. The momentum of the protons in the transverse plane is negligibly small. Given the momentum conservation principle, the sum of all momenta of final state particles in the transverse plane is expected to be zero. Since neutrinos leave no signal in the ATLAS detector but carry momentum, this generally leads to missing transverse momentum. The missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}}$ is the vectorial sum of all transverse momenta from all reconstructed objects plus any contributions from tracks or energy deposits in the calorimeter that were not associated to any of the particles mentioned above [35, 36].

5.6 Overlap removal

After all objects have been reconstructed, objects must be removed that have double-counted tracks or energy deposits. This prevents counting tracks or energy deposits more than once, which would distort the total energy deposited. The overlap removal procedure is done in the following order, following the ATLAS collaboration recommendations [37]. If an object is discarded, it will not be considered in the following steps.

- Any electron found with a track in the ID, overlapping with any other electron, is removed.
- Any calorimeter muon[2] found to share a track in the ID with an electron is removed.
- Any electron found to share a track in the ID with a muon is removed.
- Any jet found within a $\Delta R \leq 0.2$[3] of an electron is removed.
- Any electron subsequently found within $\Delta R \leq 0.4$ of a jet is removed.
- Any jet with less than 3 tracks in the ID associated to it found within $\Delta R \leq 0.2$ of a muon is removed.
- Any jet with less than 3 tracks in the ID associated to it which has a muon ID track ghost-associated[4] to it, is removed.
- Any muon subsequently found within $\Delta R \leq 0.4$ of a jet is removed.

[2] A muon, which leaves a track in the ID and the calorimeter, but not in the MS because of a detector gap, is called a calorimeter muon.

[3] $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ is a measure for the spatial angle between two objects in the detector.

[4] Ghost-association describe the procedure of implementing an object into a jet. The four-vector of the object to be added, is added to the list of objects, which are reconstructed to a jet and the jet clustering algorithm is repeated. Ghost-association allows to account for jets with an unusual shape.

6 Event selection and background estimates

This chapter describes the motivation to use a signal region and a control region. In addition, the selections are explained and the expected numbers of events are presented.

6.1 Signal regions

As already briefly discussed in Section 2.3, this analysis considers the same FCNC vertex in two different processes, each with a similar signature. In the first process, an u/c quark is converted into top quark by emitting a Higgs boson, with the top quark and Higgs boson decaying according to the SM. This process is called "production-FCNC process". In the other process, a $t\bar{t}$ pair is generated according to the SM and one of the two top quarks then decays into an u/c quark and a Higgs boson mediated by the FCNC vertex. This process is called "decay-FCNC process". Figure 6.1 shows the Feynman diagram for the production-FCNC process, Figure 6.2 shows the Feynman diagram for the decay-FCNC process, including the further SM decays.

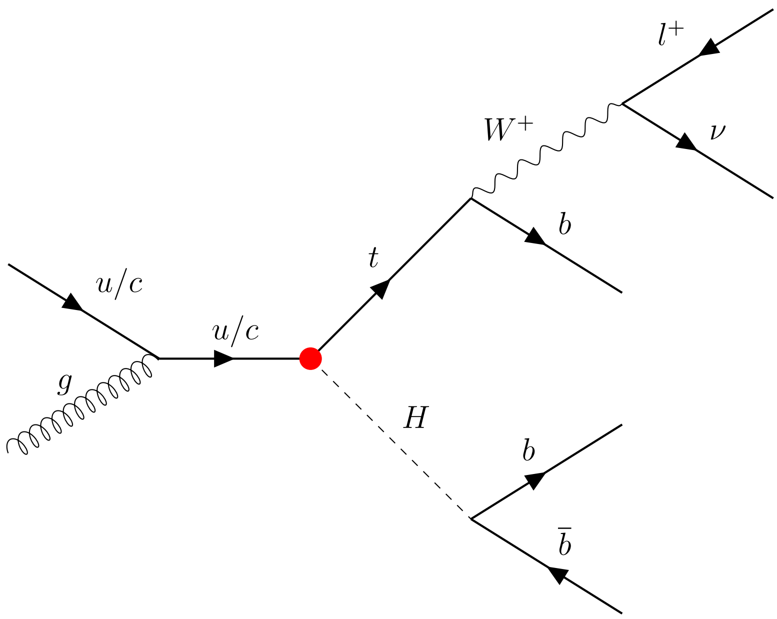


Figure 6.1: Feynman diagram of the production-FCNC process involving the FCNC vertex with further SM decays. The qtH vertex is colored in red.

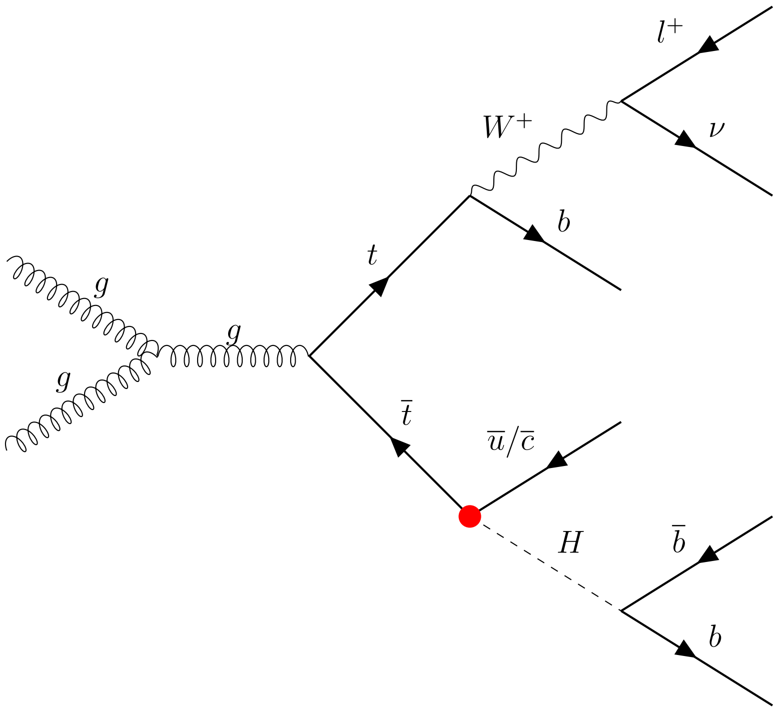


Figure 6.2: *Feynman diagram of the decay-FCNC process involving the FCNC vertex with further SM decays. The qtH vertex is colored in red.*

In order to maximize the signal over background ratio as much as possible, the definitions of the signal regions are adapted to the expected number of leptons, jets, b jets and $E_{\mathrm{T}}^{\mathrm{miss}}$. Since additional radiations of gluons from initial and final state particles can occur, causing further light- or c jets, the selection is not limited to exactly 3 jets. Furthermore, those additional radiations would influence the kinematic variables, thus the definition of the signal regions distinguishes between three different jet channels. The different requirements for the **three** signal regions are listed below:

- **exactly one lepton** (electron or muon, $p_T \geq 28\,\mathrm{GeV}$)
  - veto against a second loose lepton
- $\boldsymbol{E_{\mathrm{T}}^{\mathrm{miss}}}\geq 30\,\mathrm{GeV}$
- either **3 jets** or **4 jets** or **$\geq$5 jets** ($p_T \geq 30\,\mathrm{GeV}$)
  - of which **exactly 3** are ***b* tagged** (85% WP)

Since the resolution of jets is worse than the resolution of leptons, the decay of the Higgs boson into a $b\bar{b}$ pair is not as clear in the detector as for example into two W- or Z-bosons, which then (partially) decay leptonically. However, the $b\bar{b}$

channel has the largest branching ratio (approx 58% [38]) so that a better significance is expected by the increased statistics [39] and the usage of continuous b-tagging techniques.

The main background is $t\bar{t}$, due to the large cross section. Furthermore, the $t\bar{t}$ signature is already very similar to the selections for the signal region. Only one top quark must decay leptonically and the other top quark hadronically, so that with another b jet (for example as mis-tag or from additional gluon radiation) the signature of the signal is already imitated. A mis-tagged b jet is a light- or c jet, which was wrongly classified as a b jet. Single top-quark production includes the s-channel, t-channel and the tW-production processes. The single top-quark production can have events in the signal region, if the top quark decays leptonically. The remaining b jets are either generated/mis-tagged by the particles involved in the hard scattering process or caused by additional radiation. W+jets production with a leptonically decaying W boson is also an important background. The backgrounds Z+jets, Diboson, ttH, ttV $(V = W, Z)$ and the SM process tHq are also considered, but their contributions are very small.

The event yield for the three signal regions can be seen in Table 6.1. Included are all considered backgrounds and the FCNC signals, divided into utH/ctH and production/decay. Furthermore, the $t\bar{t}$ background is split into three different subprocesses. The definition and the motivation for this is described in Section 6.2.

It can be seen, that the amount of production-FCNC events differ between the utH and ctH channel. This is due to the parton distribution functions[1]. Since the up quark exists as valence quark in the proton and the c quark only as a sea quark, it is more likely to have the utH vertex than the ctH vertex. The expected number of decay-FCNC events, on the other hand, does not differ between the utH and the ctH channel. This is because in the calculation of the decay width for the processes $t \to uH$ and $t \to cH$, the same (negligible) mass of the up- and charm quark, compared to the top quark, was assumed. This leads to the same decay width for $t \to uH$ and $t \to cH$, respectively, and consequently to the same number of expected events. This assumption is justified because of the small mass difference between the up quark and the charm quark relative to the top-quark mass. A graphical illustration of the fraction of events of the backgrounds can be seen in Figure 6.4.

[1]The parton distribution function $f_i(x, Q^2)$ gives the probability of finding a parton i with the momentum x in the proton with Q being the momentum scale of the hard interaction.

Table 6.1: Event yields in the signal regions before correcting the $t\bar{t}$+HF miss modeling for an integrated luminosity of $\int Ldt = 139\,\mathrm{fb}^{-1}$.

	3 jet 3 b-tag SR	4 jet 3 b-tag SR	≥5 jet 3 b-tag SR
utH prod-FCNC	213 ± 5	145 ± 4	78 ± 3
utH decay-FCNC	210 ± 5	316 ± 7	314 ± 7
ctH prod-FCNC	41 ± 3	26 ± 2	12 ± 1
ctH decay-FCNC	210 ± 5	316 ± 7	314 ± 7
$t\bar{t}$(W→cq)	146 000 ± 33 000	240 000 ± 50 000	240 000 ± 60 000
$t\bar{t}$+j(bbb/c)	10 400 ± 2600	23 000 ± 6000	43 000 ± 11 000
$t\bar{t}$+j(bbl)	57 000 ± 14 000	86 000 ± 22 000	111 000 ± 31 000
t-channel	7100 ± 1600	4400 ± 1500	3200 ± 1700
s-channel	540 ± 120	350 ± 90	190 ± 50
Wt-channel	7500 ± 1600	11 700 ± 2900	11 000 ± 4000
W+jets	19 000 ± 6000	16 200 ± 1800	17 000 ± 5000
Z+jets,VV	2600 ± 500	2700 ± 500	3100 ± 600
SM tHq	33.9 ± 1.5	31.6 ± 1.5	24.3 ± 1.3
ttH	85.3 ± 1.5	391 ± 7	1396 ± 24
ttV	250 ± 50	790 ± 160	2600 ± 500
Total background	250 000 ± 40 000	390 000 ± 60 000	430 000 ± 80 000

6.2 Control regions

The rate of the $t\bar{t}$+jets process is not well predicted by the NLO + parton shower generators. In this analysis, the rate of additional gluon radiations, which results in either a $b\bar{b}$-quark pair or a $c\bar{c}$-quark pair, is not well predicted and has a potentially high impact on the signal region due to the chance of (mis-)tagging one of the quarks from the radiation as third b jet. A reason for the bad prediction is, that the parton shower is only calculated at leading log precision. To better account for the misprediction, the $t\bar{t}$ process is divided into three subprocesses using truth information in the MC samples. The subprocesses are denoted as follows:

$t\bar{t}$ ($W \to cq$)
: This process is defined by having one hadronically decaying W boson and one leptonically decaying W boson. This subprocess has only a negligible amount of events in the control region but is the major background in the signal region.

$t\bar{t}$+j (bbb/c)
: In this process, both W bosons decay leptonically and there are additional radiations of b or c quarks. The additional radiations are then responsible for the third b tagged jet. This process is also called "$t\bar{t}$ + Heavy Flavor (HF)" and has to be normalized.

$t\bar{t}$+j (bbl)
: In this process, again both W bosons decay leptonically and there are additional radiations of light quarks which are responsible for the third b tagged jet.

Since the rate prediction does not change the kinematic behavior of the process, the amount of expected events can be corrected using data. To normalize the $t\bar{t}$ + HF process, control regions, where the amount of signal events is small, are used to identify this process by truth-matching and then rescale this process to observed data. This scale factor is then propagated into the signal region to correct the rate prediction.

An example of a Feynman diagram of the $t\bar{t}$ + HF process, which has to be corrected, can be seen in Figure 6.3.

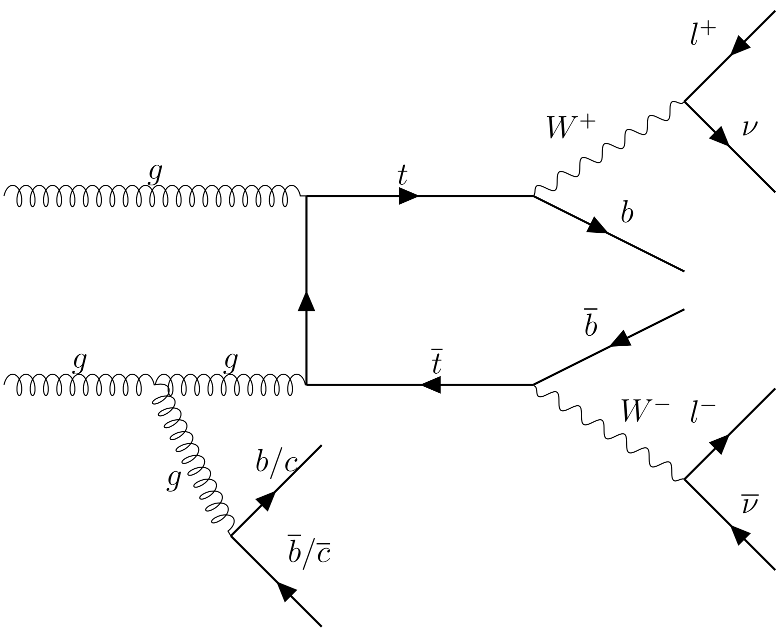


Figure 6.3: *Feynman diagram of the $t\bar{t}$ + HF process.*

To define the control regions orthogonal to the signal regions and in addition have a $t\bar{t}$+HF enriched region, the following selection is made for the **three** control regions:

- **exactly two leptons** (ee, $e\mu$, $\mu\mu$; $p_T \geq 28\,\text{GeV}$)
  - veto against a third tight lepton
- $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$$\geq$30 GeV
- either **3 jets** or **4 jets** or **≥5 jets** ($p_T \geq 30\,\text{GeV}$)

– of which **exactly 3** are ***b* tagged** (85% WP)

As can be seen, the selection, except the number of leptons, is equal to the selection for the signal region. The requirement of the two leptons on the one hand almost ensures that both W bosons from the top quark decays decay leptonically, so that the three b jets result from the b quarks from the top quark decay and from the additional radiation. This allows to rescale exactly the process whose rate is not well predicted. On the other hand, the requirement of the two leptons ensures that only a negligible part of the signal is in the control region. Furthermore, it can be assumed that the scale factor, determined in the control region, is similar enough to be applied to the signal region because the only difference between the $t\bar{t}$+HF process in the signal and the control region is the W boson decay, which has no impact on the production rate. Each control region is used to determine a correction factor which scales the $t\bar{t}$+HF process. The scale factors from each control region are combined into a single scale factor which is afterwards applied in all signal regions.

The event yield for the three control regions can be seen in Table 6.2. All considered backgrounds are included.

Table 6.2: *Event yields in the control regions before correcting the $t\bar{t}$+HF miss modeling for an integrated luminosity of $\int Ldt = 139\,\mathrm{fb}^{-1}$.*

	3 jet 3 b-tag CR	4 jet 3 b-tag CR	≥5 jet 3 b-tag CR
$t\bar{t} + j(bbb/c)$	4600 ± 1200	3800 ± 900	2800 ± 800
$t\bar{t} + j(bbl)$	7600 ± 2000	6100 ± 1700	4700 ± 1400
t-channel	0.28 ± 0.28	1.1 ± 0.9	1.1 ± 2.0
s-channel	0.09 ± 0.09	0.03 ± 0.07	0.05 ± 0.10
Wt-channel	370 ± 190	270 ± 140	170 ± 100
W+jets	0.15 ± 0.08	0.8 ± 0.5	1.04 ± 0.35
Z+jets,VV	1300 ± 500	1030 ± 320	1040 ± 340
SM tHq	0.44 ± 0.10	0.20 ± 0.11	0.24 ± 0.12
ttH	32.6 ± 0.6	40.4 ± 0.7	41.4 ± 0.7
ttV	73 ± 15	137 ± 28	260 ± 50
Total background	14 000 ± 2500	11 400 ± 2100	9000 ± 1900

A graphical illustration of the relative fraction of events of the backgrounds can be seen in Figure 6.4.

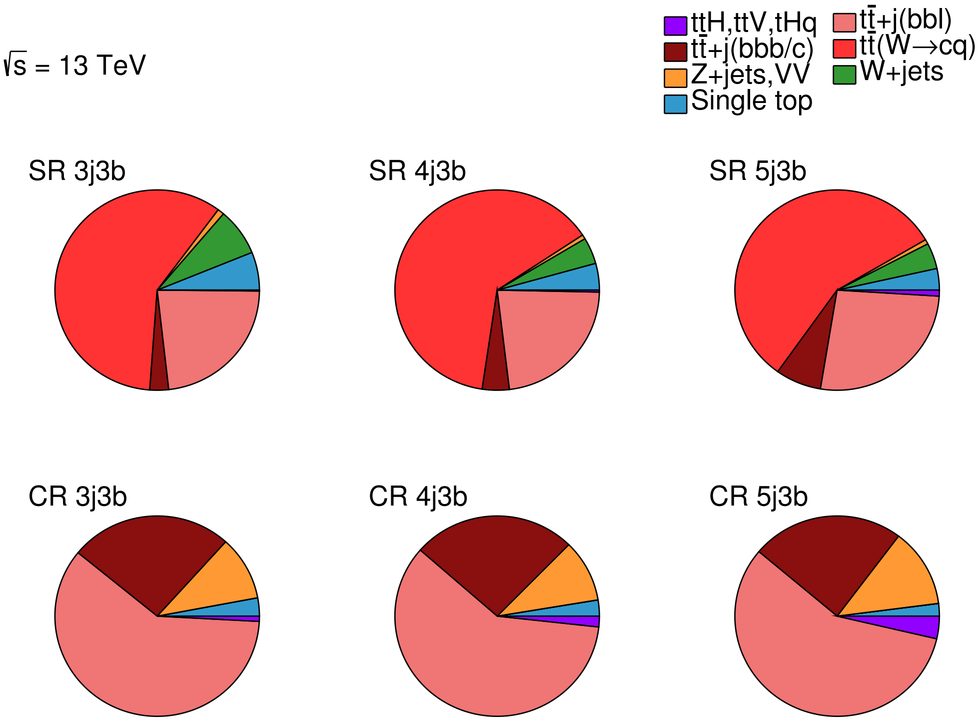


Figure 6.4: *Illustration of the relative fraction of background events in the signal region and in the control region.*

Even if the signal regions have already been designed to obtain a good signal-to-background-ratio, the number of background events is by orders of magnitude larger than the number of expected signal events. In order to still be able to discover a potential signal, the reconstruction is tried to be optimized and multivariate analysis techniques are used to discriminate signal from background. This is done in Chapter 7 and 8.

7 Event Reconstruction

An important step in differentiating between signal and background is a good reconstruction of the process. Since a top quark nearly always decays into a W boson and a b quark, the W boson must first be reconstructed in order to reconstruct the top quark. This step and also two different approaches to reconstruct the top quark and the Higgs boson are discussed in this chapter.

7.1 Reconstruction of the neutrino

The leptonic decay of the W boson has the advantage that the resulting charged lepton is well detectable and suppresses some backgrounds. The disadvantage of leptonic decay is the associated production of a neutrino, which carries part of the momentum of the W boson, but is never detected. However, since there is practically no momentum in the transverse plane before the collision of the protons, it can be assumed that there is no momentum in the transverse plane after the collision. Thus in the case of the processes considered in this analysis, it is assumed that one single neutrino is responsible for the entire $E_{\mathrm{T}}^{\mathrm{miss}}$, about which one only knows the x- and y-components, but not the z-component. Thus, it holds that $E_{\mathrm{T}}^{\mathrm{miss}} = p_{T,\nu}$. Using the fact that the W boson is on-shell[1], a quadratic equation can be obtained using the pole-mass of $m_W = 80.4\,\mathrm{GeV}$, which leads to two possible solutions for the z-component of the neutrino. This equation can be seen in the following [40]

$$p_{z,\nu}^2 - 2 \cdot \frac{\mu \cdot p_{z,l}}{E_l^2 - p_{z,l}^2} \cdot p_{z,\nu} + \frac{E_l^2 \cdot p_{T,\nu}^2 - \mu^2}{E_l^2 - p_{z,l}^2} = 0 \tag{7.1}$$

with

$$\mu = \frac{m_W^2}{2} + \cos(\Delta\Phi) \cdot p_{T,l} \cdot p_{T,\nu} \quad . \tag{7.2}$$

E_l, $p_{T,l}$ and $p_{z,l}$ stand for the energy, the transverse momentum and the momentum in z-direction of the charged lepton, respectively. The properties of the neutrino ν are named analogously. $\Delta\Phi$ is the azimuthal angle between the charged lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$. Resolving Equation (7.1) after $p_{z,\nu}$ gives Equation (7.3)

[1] Being on-shell means, that the mass of the W boson is equal to its pole mass

$$p_{z,\nu}^{\pm} = \frac{\mu \cdot p_{z,l}}{p_{T,l}^2} \pm \sqrt{\frac{\mu^2 \cdot p_{z,l}^2}{p_{T,l}^4} - \frac{E_l^2 \cdot p_{T,\nu}^2 - \mu^2}{p_{T,l}^4}} \ . \tag{7.3}$$

In the case of two real solutions, both are retained and in a later step of the reconstruction, it is decided which one will be further used. If, due to detector effects, $m_{T,W}$ is larger than m_W, $E_{\mathrm{T}}^{\mathrm{miss}}$ is adjusted in such a way that the square root in Equation (7.3) is zero, so that $m_{T,W} = m_W$ applies. This prevents a complex, non-physical solution in Equation (7.3). Further information about the W boson reconstruction can be found in Ref. [40].

7.2 Reconstruction of the top quark and the Higgs boson

After selecting the final state objects, described in Chapter 6, it is now necessary to find out which of the b jets comes from the top-quark decay and which two b jets come from the decay of the Higgs boson. A good reconstruction should assign the b jets correctly as often as possible so that kinematic distributions result for the FCNC processes, which differ significantly from the backgrounds which mostly do not involve a Higgs boson. Thus, in a later part of the thesis, the signal can be better separated from the background with the help of multivariate analysis techniques.

7.2.1 The χ^2 algorithm

The χ^2 algorithm loops over all possible permutations of b jets and solutions for the neutrino. In each permutation, a top quark and a Higgs boson are reconstructed from the b jets and the W boson. Their invariant masses are then used in Equation (7.4) to obtain a χ^2 value. The permutation with the smallest χ^2 value is then used as the correct reconstruction.

$$\chi^2 = \frac{\left(m_H^{reco} - m_H^{ref}\right)^2}{{\sigma_H^{ref}}^2} + \frac{\left(m_{top}^{reco} - m_{top}^{ref}\right)^2}{{\sigma_{top}^{ref}}^2} \tag{7.4}$$

Within this formula, m_H^{reco} corresponds to the invariant mass of the reconstructed Higgs boson, m_H^{ref} is a reference value for the Higgs mass and σ_H^{ref} is a reference value for the reconstructed decay width of the Higgs boson including detector effects. The corresponding values for the top quark are named analogously. To have a consistent set of MC samples, those are not always generated with the most up-to-date literature values for masses and decay widths for all particles involved. These reference values are extracted from truth information of the samples and then later used for reconstruction. The used values of the reference values are $m_H^{ref} = 123.33\,\text{GeV}$, $\sigma_H^{ref} = 3.83\,\text{GeV}$, $m_{top}^{ref} = 172.23\,\text{GeV}$ and $\sigma_{top}^{ref} = 2.76\,\text{GeV}$. A typical distribution of the lowest χ^2 value of all permutations in an event (denoted as $\min.\chi^2$) can be seen in Figure 7.1.

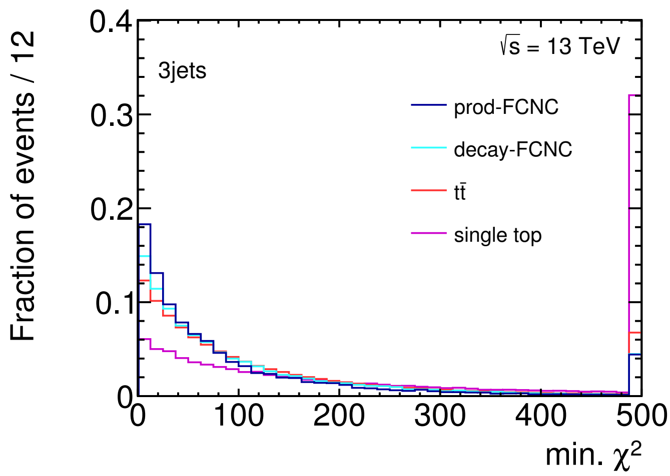


Figure 7.1: *Distribution of the* $\min.\chi^2$ *variable in the 3 jet* utH *channel.* $t\bar{t}$ *and single top-quark production are the two main backgrounds.*

As can be seen, the production-FCNC process has a higher probability for a smaller $\min.\chi^2$ value than the backgrounds $t\bar{t}$ and single top-quark production. A smaller $\min.\chi^2$ value means a higher probability of correctly assigning all measured objects to their actual mother particles. The FCNC processes having a smaller $\min.\chi^2$ value can be explained by the fact, that in the signal there should always be 2 b jets coming from the Higgs-boson decay and one b jet coming from the top-quark decay while for the two backgrounds, the Higgs boson is reconstructed from the left-over b jets, which were not assigned to the top quark. This causes the invariant mass of the reconstructed Higgs boson to be further away from the reference mass in the background processes than in the signal processes. According to Equation (7.4), this leads to a higher $\min.\chi^2$ value. Also, the similarity of the decay-FCNC process to the $t\bar{t}$ process can be seen, since the distributions of the decay-FCNC process are not as distinguishable from the background $t\bar{t}$ as the production-FCNC process is. This kinematic similarity between the decay-FCNC process and $t\bar{t}$ can be observed in many different variables as Section 8.1 will show further. This is because the only difference between the two processes is that in the decay-FCNC process a top quark decays via the FCNC coupling.

7.2.2 The KLFitter algorithm

Within the **K**inematic log-**L**ikelihood **Fitter** (KLFitter) framework [41], a likelihood value for all possible permutations of input objects is calculated, following Equation (7.5) The permutation with the highest value is then used as the correct reconstruction.

$$\begin{aligned}\mathcal{L} = {} & \mathcal{B}(m_{q_1q_2}|m_H, \Gamma_H) \cdot \mathcal{B}(m_{q_3l\nu}|m_{top}, \Gamma_{top}) \cdot \mathcal{B}(m_{l\nu}|m_W, \Gamma_W) \cdot \\ & \prod_{i=1}^{n} W_{jet}(E_{jet,i}^{meas}|E_{jet,i}) \cdot W_l(E_l^{meas}|E_l) \cdot W_{miss}(E_x^{miss}|p_x^{\nu}) \cdot W_{miss}(E_y^{miss}|p_y^{\nu})\end{aligned} \tag{7.5}$$

with $\mathcal{B}$ being the Breit-Wigner-function which quantifies the agreement of the fitted four-vectors of the reconstructed particles with their reference values and $W\,(X|Y)$ being predefined transfer-functions which quantify the agreement of the fitted energies and missing transverse momenta with their measured values.

KLFitter stores the value of the log-likelihood function for each permutation, so that the log-likelihood value of a permutation, divided by the sum of the log-likelihood values of all permutations, can be interpreted as a probability for the correct permutation. The highest probability of all permutations of an event is stored for a later step in the analysis (analogous to the χ^2 value). This variable is called "prob_KLFitter". The distribution of the "prob_KLFitter" variable for the 3-jets utH channel can be seen in Figure 7.2.

It can be seen that the FCNC signal tends to slightly higher values than the backgrounds. Figure 7.3 shows the resulting distribution of the "prob_KLFitter" variable in the 3-jets ctH channel.

If one compares the distribution of the "prob_KLFitter" variable in the ctH-channel with that in the utH-channel, it is noticeable that the ctH signal is no longer different from the background.

For the reconstruction of the top quark and the Higgs boson, the KLFitter framework offers the possibility to "fix" the masses. Fixing the masses causes the detector effects in the likelihood function to be adjusted in such a way that the resulting masses of the reconstructed particles correspond more likely to the reference value. The effect of this functionality on the reconstruction, as well as the use of the likelihood value itself, will be investigated. In order to compare the quality of the

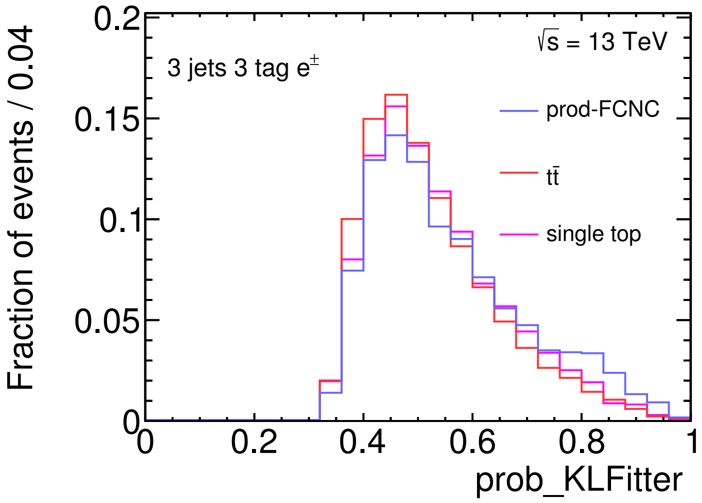


Figure 7.2: Distribution of the "prob_ KLFitter" variable in the 3-jets utH channel.

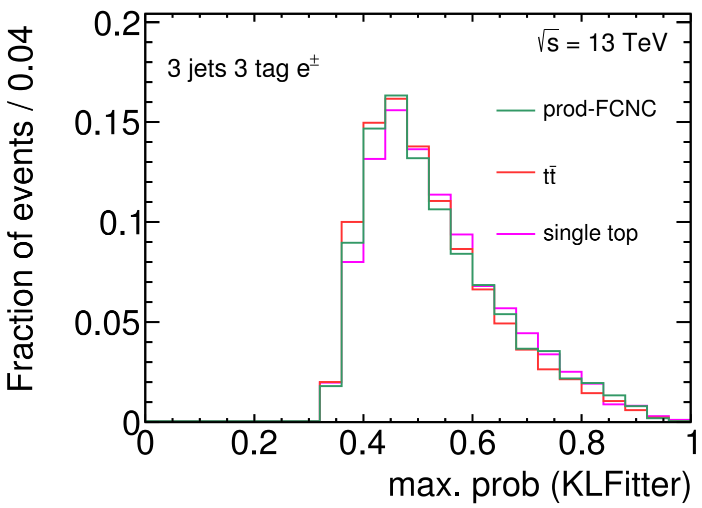
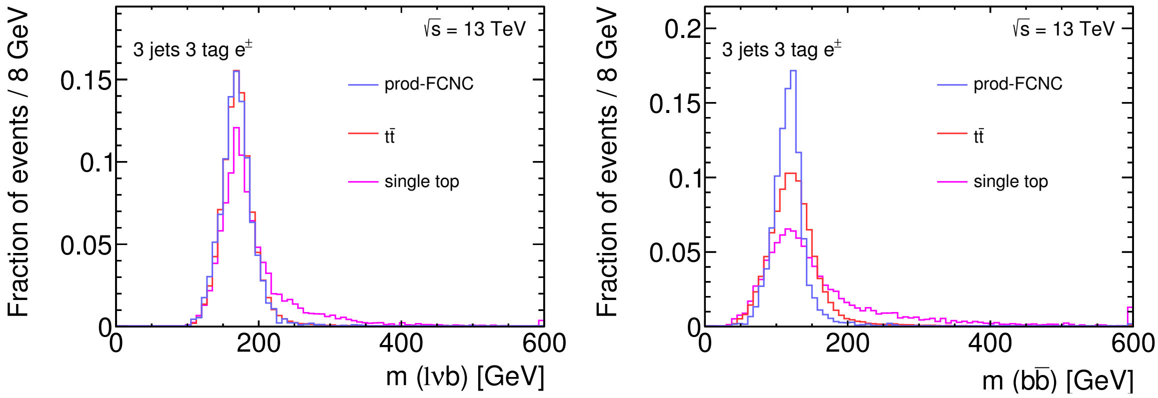


Figure 7.3: Distribution of the "prob_ KLFitter" variable in the 3 jet ctH channel.

different reconstructions, trainings of a neural network are performed with the help of the reconstructed variables. The resulting quality of the training (in terms of Gini index) is used as a measure of the quality of the reconstruction. More on the functionality of neural networks and on the one used in this analysis can be found in Chapter 8. Furthermore, it should be mentioned that the reconstruction is only studied for the production-FCNC process and only for the 3-jets and 4-jets channel.

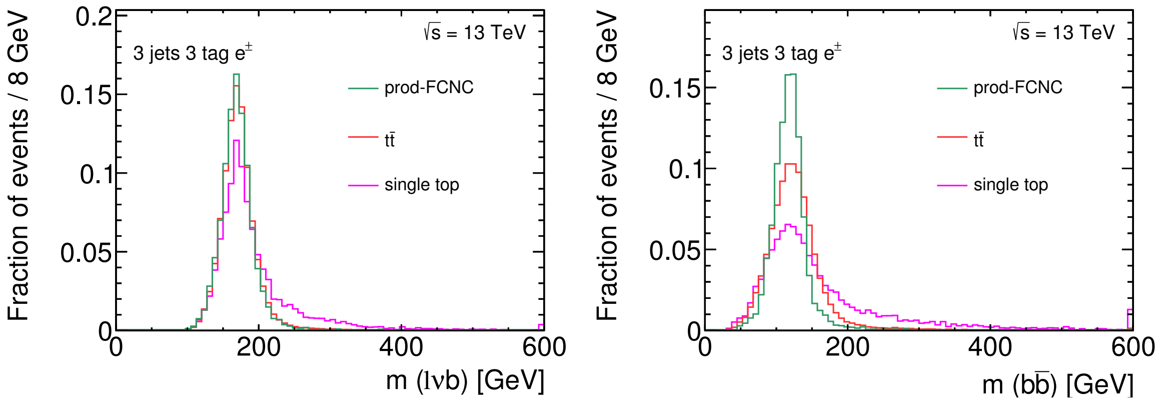
Reconstruction algorithm comparison

This section summarizes the results with the χ^2 algorithm and the KLFitter algorithm for the utH and ctH channel. The distributions of the invariant masses are shown in Figure 7.4 and 7.5 for the case, that the top quark and the Higgs boson are reconstructed with the χ^2 algorithm. The production-FCNC process is only compared to the main backgrounds $t\bar{t}$ and single top-quark production.



(a) (b)

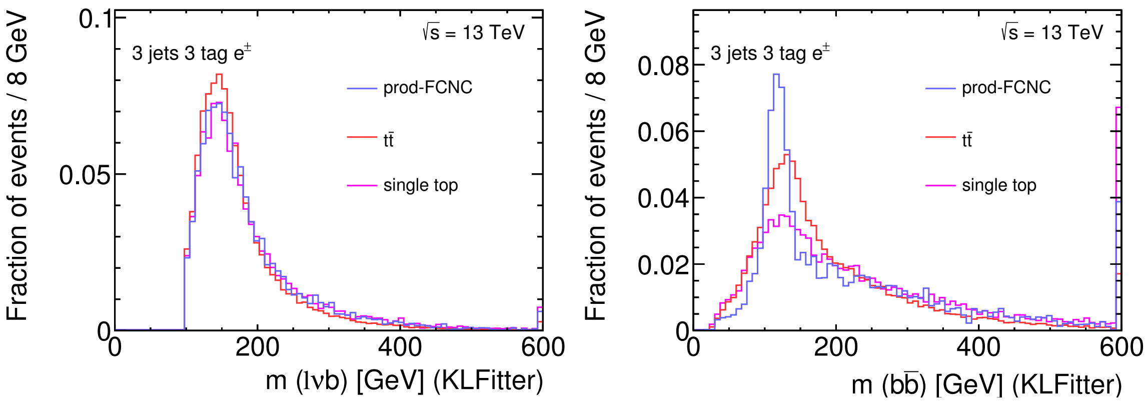
Figure 7.4: *Invariant mass distributions of the reconstructed top quark (a) and the Higgs boson (b). Reconstructed with the χ^2 algorithm in the 3-jets utH channel.*



(a) (b)

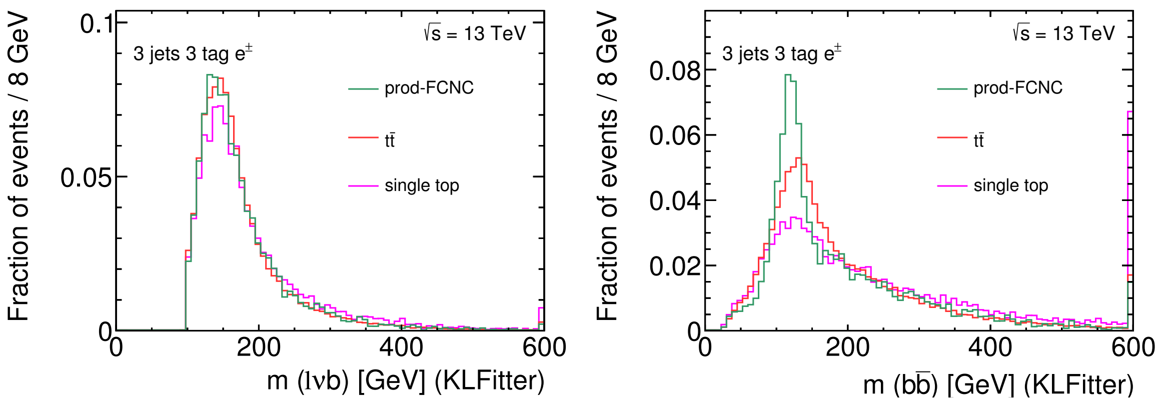
Figure 7.5: *Invariant mass distributions of the reconstructed top quark (a) and the Higgs boson (b). Reconstructed with the χ^2 algorithm in the 3-jets ctH channel.*

It can be seen that the distributions for the invariant mass of the reconstructed top quark do not differ significantly between signal and background. This is because a top quark actually occurs in all processes under consideration. The distributions of the invariant mass of the Higgs boson, on the other hand, differ significantly, since a Higgs boson only occurs in the FCNC processes. In the case of the backgrounds the remaining b jets, which were not reconstructed to the top quark, are reconstructed to the Higgs boson, which results in a distribution that does not have such a pronounced peak at the Higgs boson mass. A similar behavior can be seen for the KLFitter reconstruction in Figure 7.6 and 7.7.



(a) (b)

Figure 7.6: Invariant mass distributions of the reconstructed top quark (a) and the Higgs boson (b). Reconstructed with the KLFitter algorithm in the 3-jets utH channel.

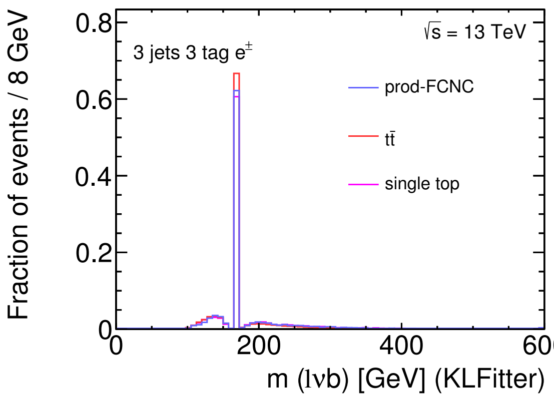


(a) (b)

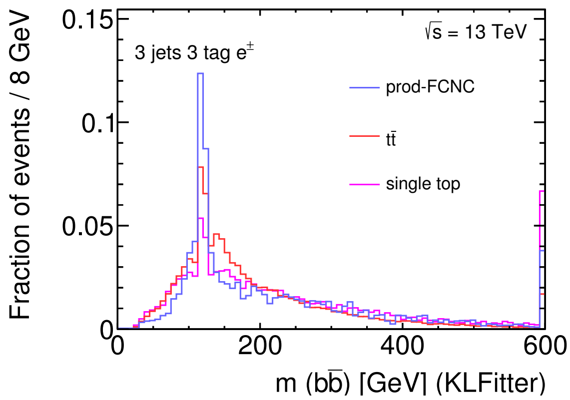
Figure 7.7: Invariant mass distributions of the reconstructed top quark (a) and the Higgs boson (b). Reconstructed with the KLFitter algorithm in the 3-jets ctH channel.

First, it can be seen that the shapes of the distributions of the invariant masses differ from those of the χ^2 algorithm. This is because of the formula used in the algorithms: In the χ^2 algorithm, deviations from the reference value are considered quadratic, so that large deviations become unlikely. This also results in a (nearly) symmetrical distribution around the reference value. In the likelihood function of KLFitter, on the other hand, the decay widths of the particles are described with the help of Breit-Wigner distributions, whereby the more pronounced tails can form automatically. However, the same behavior between top mass and Higgs mass can be observed: The distributions of the invariant top mass are similar for signal and background processes, but significant differences can be observed for the Higgs mass.

If the “fixed-mass” option is used for the reconstruction with KLFitter, distributions result which are shown in Figure 7.8.



(a)



(b)

Figure 7.8: *Invariant mass distributions of the reconstructed top quark (a) and the Higgs boson (b). Reconstructed with the KLFitter algorithm in the 3-jets utH channel with the fixed-mass option being enabled.*

It is noteworthy, that the distribution of the fixed Higgs-boson mass is not influenced by fixing the top-quark mass and vice versa. The mass distributions of the top quark and the Higgs boson in the case of the ctH signal with the “fixed mass” option look similar to the ones of the utH signal. Adjusting the masses near the reference mass results in an even more pronounced peak at the reference mass. On the other hand, there are less events in the immediate vicinity of the reference mass, since these events have all been adjusted into the reference mass. This effect is more pronounced for the top-quark mass than for the Higgs-boson mass. Even though the masses often correspond exactly to the reference mass, shape differences between signal and background are almost completely removed. However, since these shape differences in the kinematic distributions are the important property required for the further analysis, it is questionable whether this option helps to improve the separation between signal and background.

In addition to the trainings of a neural network, where all kinematic variables are reconstructed with either the χ^2 algorithm or the KLFitter algorithm, trainings are also done where only the “prob_KLFitter” variable is added to a χ^2-variables training. The resulting Gini indices for the utH signal are shown in Table 7.1. The functional principle of a neural network and the definition of the Gini index can be seen in Section 8.2.

Table 7.1: Results of the Gini-index, which compares the quality of different neural networks in the utH channel.

3 jets results	χ^2	$\chi^2 + prob\_KLFitter$	KLFitter
no masses fixed	32.2	32.2	29.2
Higgs fixed	/	32.2	29.9
top fixed	/	32.2	29.5
both fixed	/	32.2	30.0
4 jets results	χ^2	$\chi^2 + prob\_KLFitter$	KLFitter
no masses fixed	32.8	32.8	30.0
Higgs fixed	/	32.8	30.6
top fixed	/	32.8	30.0
both fixed	/	32.8	29.6

As can be seen, the training of a neural network with variables, which are reconstructed with the χ^2 algorithm, has an up to 10% better Gini index than all trainings with variables, which are reconstructed with the KLFitter algorithm. Moreover, adding the “prob_KLFitter” variable to the χ^2 training does not improve the training. This is due to the fact that the neural network does not use variables if they have too little significance, i.e. they do not have additional separation power after taking into account correlated variables. That this is the case with the “prob_KLFitter” variable is partly due to the fact that the variable has a high correlation to other variables, but also to the fact that the separation strength is worse than that of the “minχ^2” variable. Table 7.2 summarizes the results of the Gini indices for the ctH signal process.

Again, the variables reconstructed by the KLFitter algorithm result in a training that yields an up to 10% worse Gini index. However, in the 4-jets ctH case for the first time, it can be observed that adding the prob_KLFitter variable to the training is helpful. Since the impact is small, it is decided that the reconstruction will be done with the χ^2 algorithm for the rest of the analysis to have a consistent analysis strategy. A possible explanation that the KLFitter algorithm is not needed is that the reconstruction is not too complex for the χ^2 algorithm. In the 3-jets channel,

Table 7.2: Results of the Gini-index, which compares the quality of different neural networks in the ctH channel.

3 jets results	χ^2	$\chi^2 + prob\_KLFitter$	KLFitter
no masses fixed	29.4	29.4	26.8
Higgs fixed	/	29.4	27.4
top fixed	/	29.4	26.7
both fixed	/	29.4	27.6
4 jets results	χ^2	$\chi^2 + prob\_KLFitter$	KLFitter
no masses fixed	29.1	29.2	25.9
Higgs fixed	/	29.2	26.6
top fixed	/	29.1	25.7
both fixed	/	29.1	26.1

there are only six permutations of b jets which have to be considered. Note also that half of the permutations are omitted since the permutations for the Higgs boson were unnecessarily counted twice, only three permutations remain to be considered. This reconstruction seems to be sufficiently simple so that the χ^2 algorithm already performs sufficiently well. However, no reason can be given with certainty why the KLFitter algorithm yields kinematic distributions which perform significantly worse in the training of the neural network than the χ^2 algorithm.

8 Separation of signal and background processes

This chapter discusses the differences between the two FCNC processes and the backgrounds. Furthermore, the functional principle of artificial neural networks (NN) is covered, specific details about the used NN are given and the undertaken steps to optimize the NN used are explained.

8.1 Discriminating signal from background

To separate the FCNC processes from the background processes, kinematic variables must be found in which the respective processes are well separable. First, the pseudo-continuous distribution of the b-tagging algorithm in Figure 8.1 should be considered.

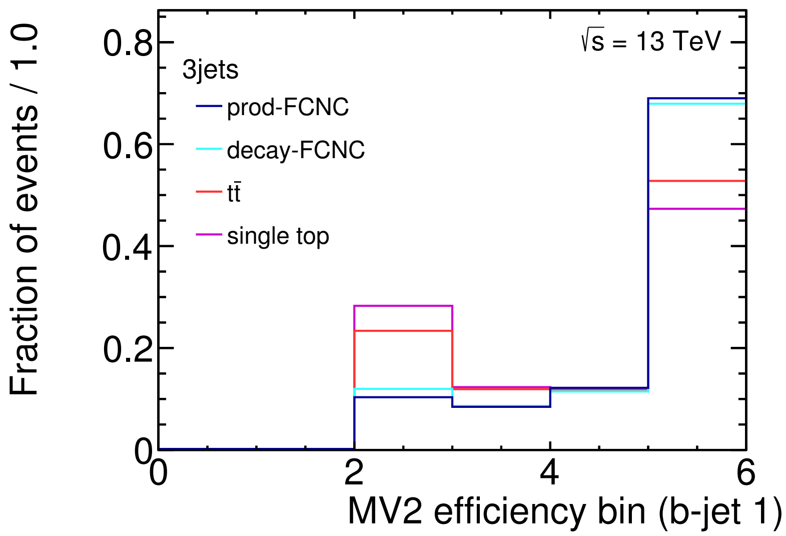


Figure 8.1: *MV2 efficiency bin distribution of leading b jet of the two FCNC processes and the two main backgrounds $t\bar{t}$ and single top-quark production in the 3-jets utH channel.*

It can be seen that the jet with the highest transverse momentum (b-jet 1) from the FCNC processes is with a significantly higher probability in a high b-tagging algorithm bin than in the backgrounds. This is also true for the other two tagged b jets. The reason for this are the final states of the processes. The signal processes actually generate three b jets, which can be tagged afterwards. In the case of the backgrounds $t\bar{t}$ and singletop there are (mostly) no real three b jets, but a c or light quark is mis-tagged. These mis-tagged jets have properties that lead to a lower b-tagging-algorithm score and thus to a lower bin.

Since a Higgs boson only occurs in the FCNC processes, it is reasonable to assume that variables, concerning the reconstruction of the Higgs boson, have an overall good separation power. As one example of variables, concerning the Higgs-boson reconstruction, the reconstructed Higgs-boson mass can be seen in Figure 8.2.

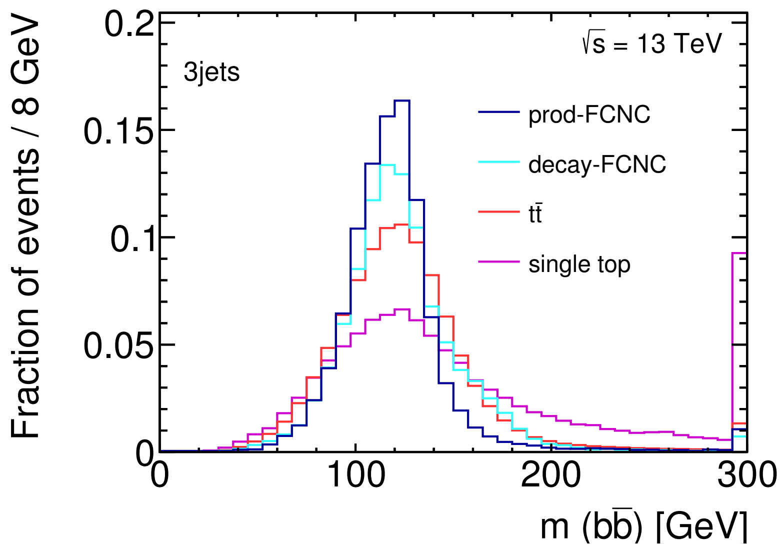


Figure 8.2: *Distribution of the invariant mass of the reconstructed Higgs boson of the two FCNC processes and the two main backgrounds $t\bar{t}$ and single top-quark production in the 3-jets utH channel.*

It can be seen that the FCNC processes have a clearer peak in the distribution of the invariant mass of the reconstructed Higgs boson than the backgrounds. The fact that the backgrounds, although there is no Higgs boson at all in the process, still have a peak at the expected Higgs mass is due to the χ^2 reconstruction algorithm. This algorithm uses the permutation of jets, which has the smallest square distance to the expected masses. The top quark is not preferred or disadvantaged to the Higgs boson in the reconstruction. If, for example, it turns out that one of the permutations randomly yields the mass of the Higgs boson, a higher deviation from the top quark can be accepted. The fact that the top-quark mass does not always result in the reconstruction, although a top quark was always present in the major background processes, is due to the finite detector resolution, which smears the energies and thus the reconstructed masses. Figure 8.2 again shows that the decay-

FCNC process has a larger similarity in the shape of the distribution of the Higgs mass to the $t\bar{t}$ background than the production-FCNC process.

In Figure 8.3, the ΔR distribution between the two b jets can be seen, which according to the reconstruction come from the Higgs boson. In this variable, the similarity between the decay-FCNC process and the $t\bar{t}$ is especially pronounced.

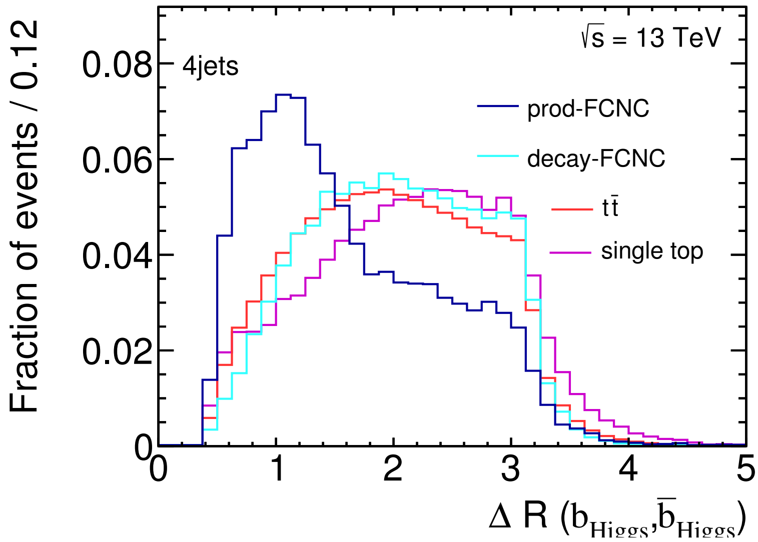


Figure 8.3: *ΔR distribution of the two b jets coming from the Higgs boson according to the reconstruction. The two FCNC processes are compared to the main backgrounds $t\bar{t}$ and single top-quark production in the 4-jets utH channel.*

By construction, it is expected that the spatial angle between the two b jets, coming from the Higgs boson, is smaller than any other combination of b jets. This behavior can be well seen for the production-FCNC process. For the decay-FCNC process, on the other hand, almost no difference can be seen to the $t\bar{t}$ background.

In order to combine all information from all distributions to separate the FCNC processes from the backgrounds in the best possible way, an artificial NN is used.

8.2 Functional principle of feedforward neural networks

An NN combines the information of several input distributions into a smaller amount of variables (often a single one), which then have a stronger separation power than the input distributions themselves. A feedforward NN consists of a predefined number of input nodes, at least one hidden layer and one or more output nodes. Each input node corresponds to one input variable. Each input node is connected to every hidden node in the first hidden layer. Each hidden node in the first hidden layer is again connected to every hidden node in the second hidden layer and so on. The hidden nodes in the last hidden layer are finally connected to the output node(s). Figure 8.4 illustrates the structure of a simple artificial NN.

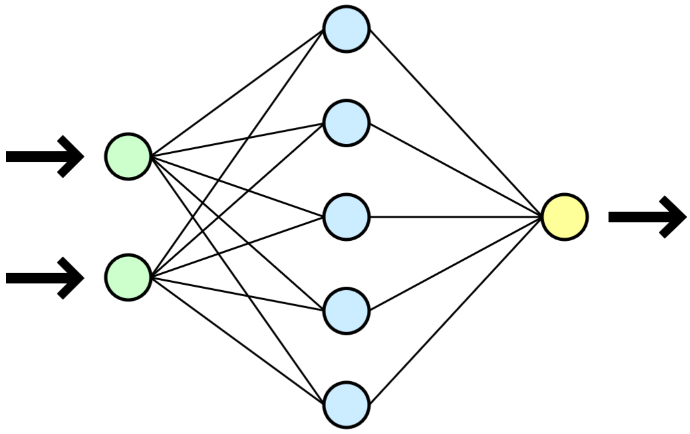


Figure 8.4: Simplified scheme of an artificial neural network with two inputs, one hidden layer with five hidden nodes and one output [42].

A line between two nodes indicates a connection. This connection is multiplied by a weight so that the relative importance of the node for the next node can be adjusted. The input value of a node is calculated according to Equation (8.1)

$$y = \sum_{i=1}^{N} a_i \cdot w_i \tag{8.1}$$

where y is the input for a given node, N is the number of nodes, which have a connection into the given node, a_i is the output values of the connected nodes and w_i is their corresponding weight. The output of a node is then calculated from the input using an activation function. The best activation function depends on the topology of the NN, but activation functions are generally always continuous, monotonously rising functions. The benefit of such activation functions, such as a sigmoid function, is that the effect of outliers is limited, making the NN more robust. An nn is trained by varying the weights of the connections between the nodes, using events where it is known if it is a signal or background event, so that a given loss-function is minimized.

8.3 NeuroBayes

NeuroBayes [43] is a multivariate analysis tool, which uses NN techniques and Bayesian statistics. Kinematic distributions, such as the energy of the lepton or the invariant mass of the reconstructed Higgs boson, serve as input variables for the NN. In addition to the kinematic distributions, another input is used, which is a constant offset to remove a potential bias. The NeuroBayes framework does a so-called pre-processing, where the distributions are transformed for a better and more robust training. These pre-processed variables are then passed into a feed-forward NN with a single hidden layer and a single output node. The number of hidden nodes in the hidden layer, among other parameters, can be set and optimized as needed. The output of the pre-processing procedure for a specific variable can be seen in Figure 8.5.

This example shows the steps of the pre-processing for the invariant mass of the reconstructed Higgs boson. The signal is the red distribution and the background the black distribution. As a first step, the bin widths are varied so that the sum of signal and background events is equal in each bin. The second graph shows the purity, i.e. the ratio of signal and signal plus background events per bin. A clear shape in the purity curve is desired since a flat line without any shape would mean, that the signal and the background are indistinguishable. The third graph shows the transformed distribution, which has a mean value of 0 and a variance of 1. This variable is finally passed into the NN. The last graph shows the signal purity, plotted against the signal efficiency. The signal efficiency is calculated by cutting at each bin and keeping either everything above or below the cut. The fraction of the retained signal and the total signal is the signal efficiency.

After pre-processing, the input variables are sorted according to their importance. This is done by first computing the $N \times N$ correlation matrix. Then each variable is temporarily removed from the matrix and the loss of correlation to the target[1] is calculated. The variable that causes the smallest loss of correlation to the target (so the least important variable) is removed. The procedure is then repeated with the $(N-1) \times (N-1)$ correlation matrix to get the second least important variable and so on until one has the full ranking of the variables. In addition to the ranking of the variables, one gets information that can be seen in Figure 8.5 at the top. The added significance is equal to the loss of correlation to the target, multiplied by $\sqrt{n}$, where n is the sample size. "only this" refers to the significance of the variable to the target, multiplied by $\sqrt{n}$. When calculating the significance to the target, correlations to all other variables are not taken into account. "signi. loss" is the loss of correlation to the target, multiplied by $\sqrt{n}$, when removing only this variable from the $N \times N$ correlation matrix and recalculating the correlation to the target with the

[1]The target variable is defined as 1 for a signal event and as 0 for a background event

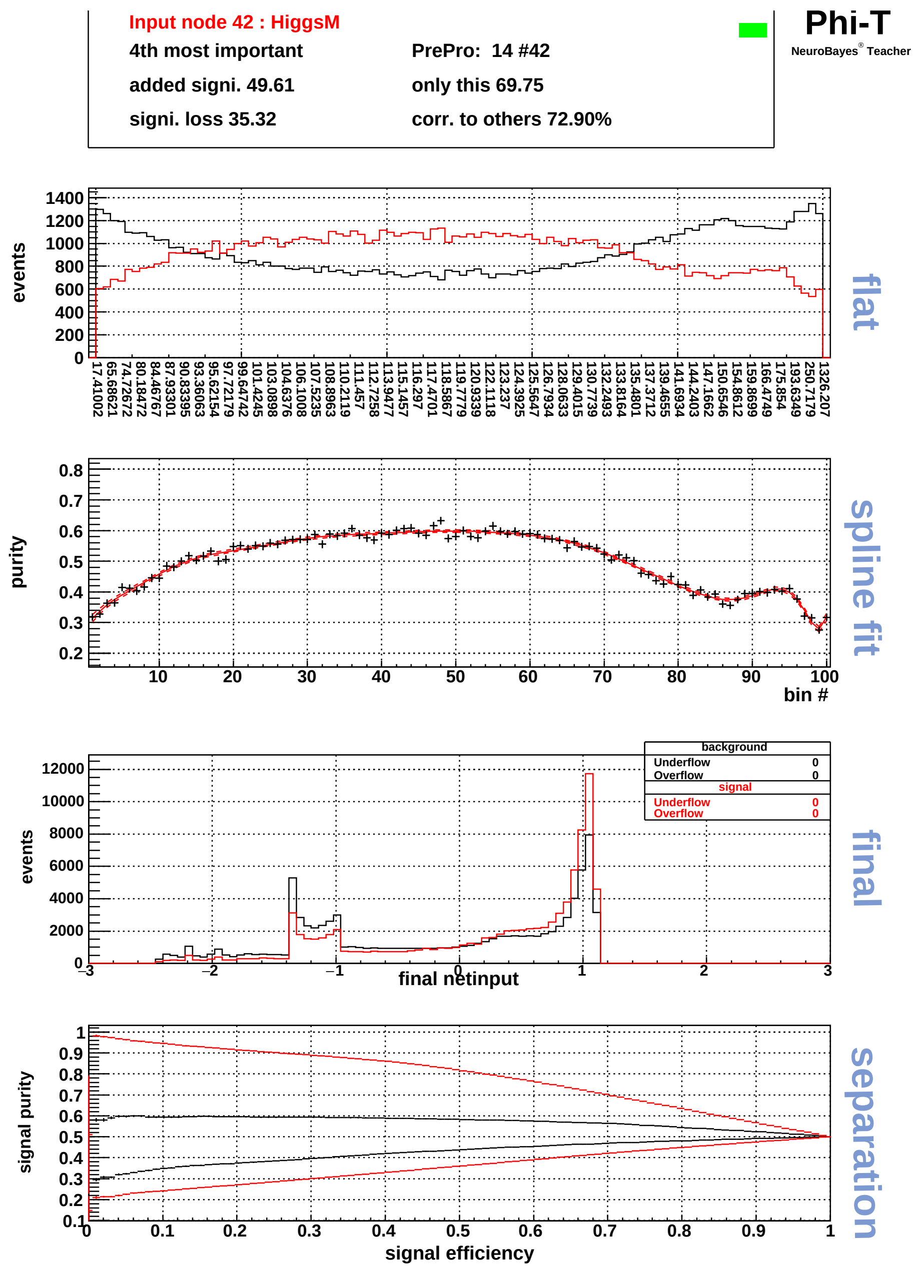
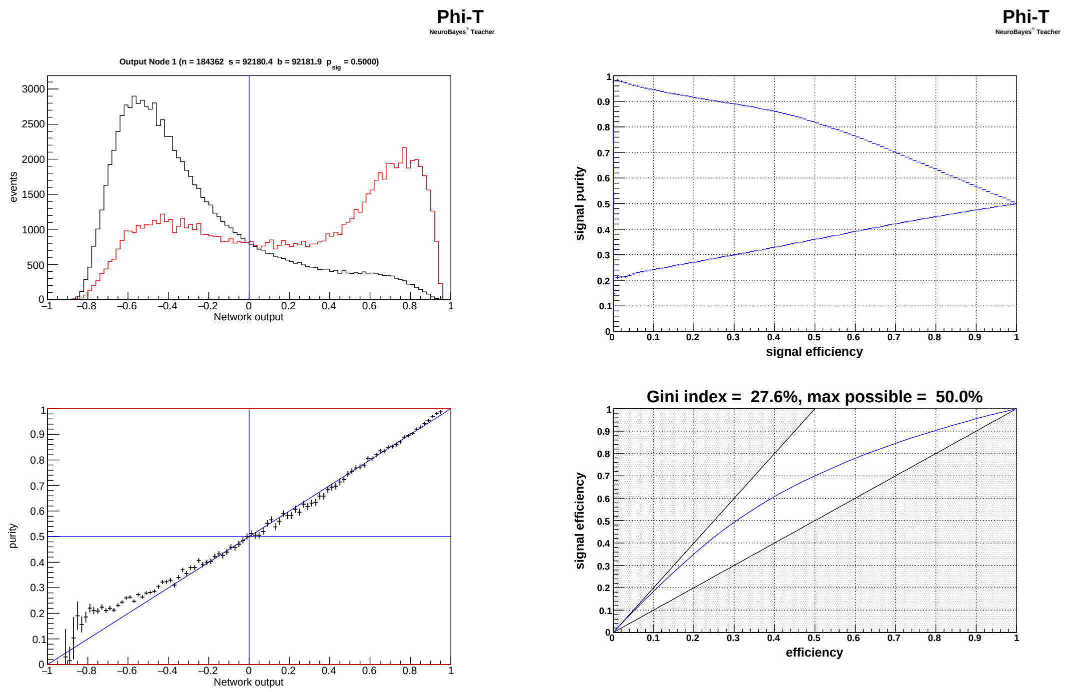


Figure 8.5: NeuroBayes output for the various steps of the pre-processing in the case of the reconstructed Higgs-boson mass. The first plot shows the signal (red) and background (black) distribution of the Higgs mass, the second plot the purity. The third plot shows the purity, transformed to a distribution with a mean of 0 and a variance of 1. The fourth plot shows the signal purity against the signal efficiency.

resulting $(N-1) \times (N-1)$ correlation matrix. The "correlation to others" indicates how strongly a variable is correlated to other variables in the $N \times N$ correlation matrix.

After a successful training additional plots are created, which are shown in Figure 8.6.



(a) (b)

Figure 8.6: *(a) shows the NN output with the signal in red and the background in black and the purity in each bin. (b) shows the signal purity against the signal efficiency and the signal efficiency against the total efficiency including the Gini index.*

In the top left corner the distribution of the output node, also called NNout, can be seen for the signal in red and for the background in black. The NN tries to assign high NNout values to signal events while at the same time tries to assign low NNout values to background events as good as possible. A high separation of the two curves is desired. The separation of the processes is also influenced by other parameters of the NN. In the bottom left corner the purity against the NNout for each bin individually can be seen. If the data points are on the diagonal, then the probability that a randomly selected event with a given NNout is actually a signal event is equal to $p = \frac{\text{NNout}+1}{2}$. In the top right corner the signal purity is plotted against the signal efficiency. This happens for the two possibilities to keep either all data above or below the cut. In the lower right corner the signal efficiency is plotted against the total efficiency of selected events. Areas that cannot be reached with the data set used are grayed out in the plot. The area between the resulting curve and the diagonal corresponds to the Gini index. This is a measure of how well a signal can be separated from the background. A higher Gini index means in principle a better separable signal. For the training the signal and background processes are reweighted so that half of the data used are signal events and the other half of the data used are background events, thus the best possible Gini-index would theoretically be 50%. However, caution is advised if one wants to quantify the quality of a training solely through the Gini index. The final result of the search for FCNC processes is influenced by many other factors, such as the binning of the NN discriminant in the statistical analysis and the influence of systematic uncertainties. As a result, especially small changes in the Gini index do not necessarily result in the better end result.

8.4 Investigation of the training strategy

Since two different processes are considered when searching for the FCNC coupling (production-FCNC and decay-FCNC process), the question arises as to whether both processes should also be considered during training. A possible counter-argument would be that the NN may not be complex enough to simultaneously classify two kinematically different processes as signals. First, a training is generated, which must simultaneously learn the characteristics of the production-FCNC process and the decay-FCNC process at the same time. The NNout distribution can be seen in Figure 8.7.

It can be seen that a large part of the decay-FCNC process has a low NNout value. This means that this part of the FCNC signal can practically not be used in statistical analysis since the small values of the NNout are dominated by the background.

Figure 8.8 shows on the left side the NNout variable of an NN, which was only trained on the production-FCNC process, and on the right side the NNout variable of an NN, which was only trained on the decay-FCNC process.

It can be seen, that each FCNC process is better separated from the background by the NN, which was specifically trained on the FCNC process respectively, compared to the combined training. Furthermore, it can be seen that the production-FCNC process is always better separable from the backgrounds than the decay-FCNC process, regardless of the training used.

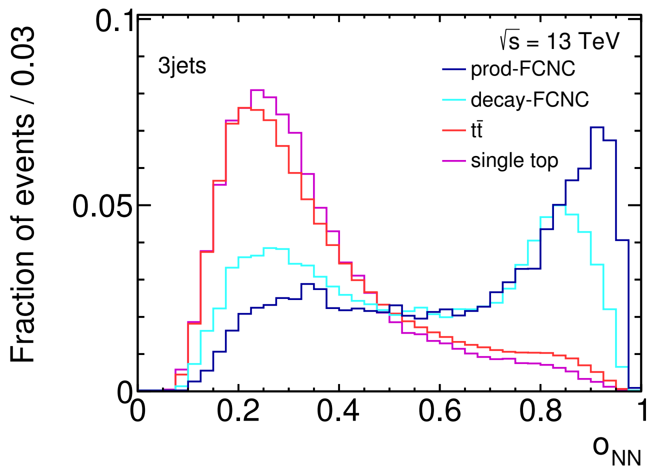


Figure 8.7: *NNout variable from an NN which was trained simultaneously on both FCNC processes in the 3-jets utH channel.*

The NNout distributions of these trainings are then evaluated using receiver operating characteristic (ROC) curves. In a ROC curve, the background efficiency is plotted against the signal efficiency for each bin. The efficiencies are calculated by dividing all events above/below each bin by all events. Since the NN classifies signal events to high and background events to low NNout values, for the ROC curves used, all events above a selection are used in the numerator for the calculation of the efficiency. The ROC curves for the different jet channels for the utH-FCNC processes can be seen in Figure 8.9. A is the area that the curve encloses with the diagonal. A larger A means that the distribution between signal and background is better differentiable. $A = 0$ means that signal and background cannot be distinguished, $A = 0.5$ means that signal and background have no overlapping events and are therefore perfectly separable.

It can be seen that the area A of the production-FCNC based training is always smaller than that of the other two trainings. Furthermore, the NN, which was trained on both FCNC processes at the same time, is better in the 3-jets channel and in the $\geq$ 5-jets channel and about as good in the 4-jets channel as the decay-FCNC based training. The fact that the production-FCNC based training does not perform so well is due to the fact that the events of the production-FCNC process are well classified, whereas those of the decay-FCNC process are not since they have more similarities with $t\bar{t}$. If the decay-FCNC based training is used, the decay-FCNC process will be better separated, and the generally easier to separate production-FCNC process will be separated automatically. This is not the case vice versa. To have a consistent analysis strategy, it is decided to use the training that was trained on both FCNC processes at the same time.

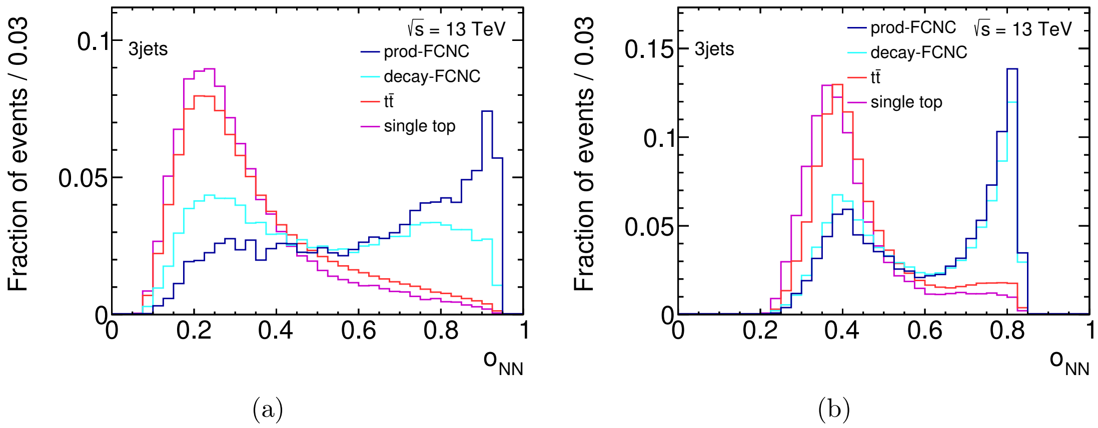


Figure 8.8: *NNout of an NN, which is trained only on the production-FCNC process as signal process (a) or only on the decay-FCNC process as signal process (b) in the 3-jets* utH *channel.*

The list of variables used in training, ordered by their importance in the training, is shown in Table 8.1.

Most of the important variables in training have something to do with the reconstructed Higgs boson. Since the Higgs boson only occurs in the FCNC processes, this is naturally one of the most separate features.

8.5 Optimization of the hyperparameters

NeuroBayes offers the possibility to adjust various hyperparameters to optimize the training. The choice of hyperparameters depends on the available statistics - low statistics in the signal or background data sets require other hyperparameters than high statistics. The adjustable hyperparameters are explained below. In addition, the procedure for determining the optimal hyperparameters is described and the hyperparameters used are shown.

- iterations: Defines the number of times, how often the training patterns are presented to the network.
- maxlearn: Used to set an upper limit to the learning rate

[2]The sphericity S is defined as $S = \frac{3}{2}(\lambda_2 + \lambda_3)$ with λ_i being the eigenvalues of the sphericity tensor $S^{\alpha,\beta} = \frac{\sum_i p_i^\alpha p_i^\beta}{\sum_i |p_i|^2}$ where $\alpha, \beta = x, y, z$ and i runs over the objects in the event. Furthermore, $\lambda_1 \geq \lambda_2 \geq \lambda_3$ and $\lambda_1 + \lambda_2 + \lambda_3 = 1$ must apply [44].

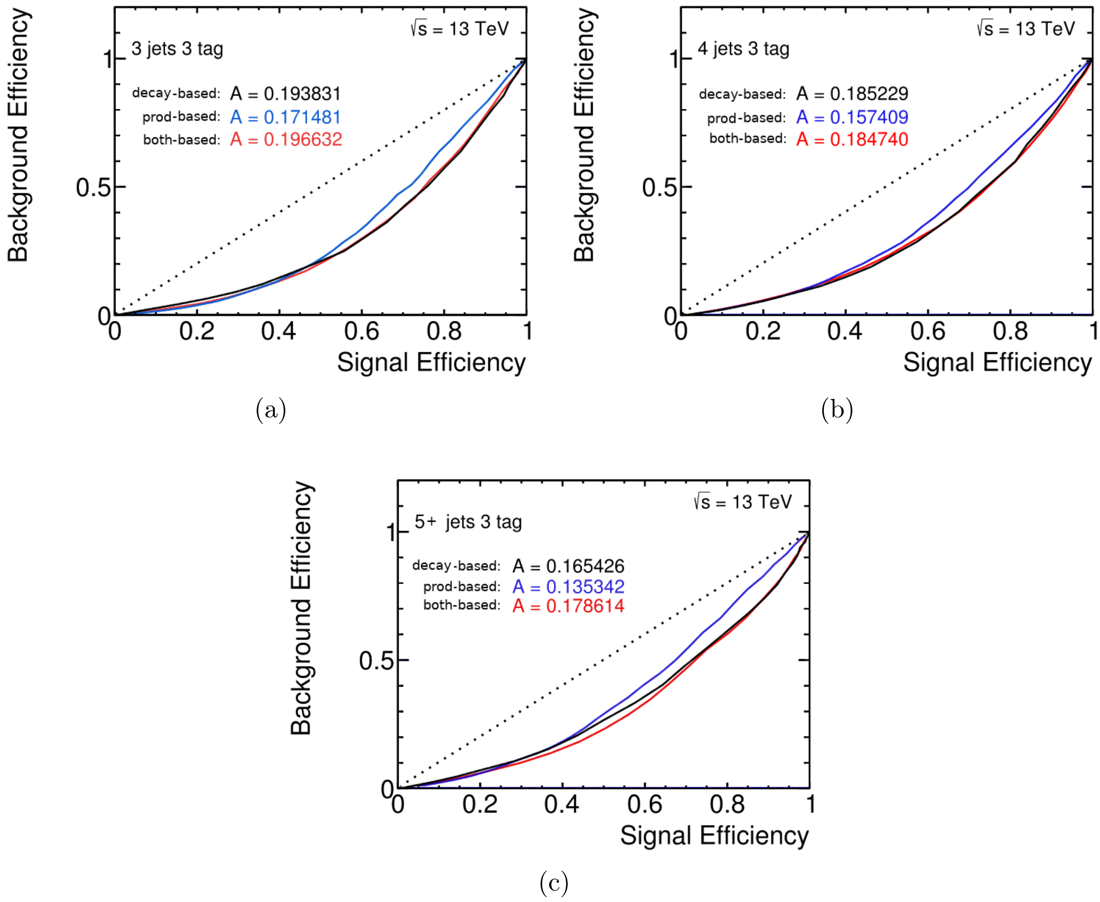


Figure 8.9: *ROC curves of differently based trainings, split into the 3-jets channel (a), 4-jets channel (b) and the ≥ 5-jets channel (c).*

- speed: The speed is a factor, which is multiplied to the learning rate. A speed larger than 1 results in a faster training but eventually the NN does not learn as good.
- mom: Adding a fraction of the previous weight update to the current one, the momentum term can help the NN to get out of local minima and to reach a global minimum faster
- weightupdate: The weightupdate defines the number of events, which are processed before a weightupdate is done.
- prunemin and prunemax: During the learning iterations, all connections are multiplied by small weights. If the network does not “revive” the connection, the contribution of a connection will become too small to contribute and can be pruned away. Pruning helps to reduce the computing power needed.

Table 8.1: List of variables used in the trainings. Sorted by importance in the training in descending order.

Variable	added Significance
$mv2c10(HiggsJet2)$	104.88
$mv2c10(HiggsJet1)$	95.58
$\Delta\Theta(b,\bar{b})$	58.37
$m(Higgs)$	49.61
$mv2c10(TopJet)$	37.05
$p_T(Jet1)$	25.53
$sphericity$[2]	15.84
$\Delta\phi(Top, Higgs)$	13.95
$m(Top)$	11.24
$\Delta\phi(b,\bar{b})$	9.09
$\Delta\eta(b,\bar{b})$	9.94
$\Delta\eta(Higgs,l)$	6.17
$\Delta\phi(Higgs,l)$	5.61
$\eta(Higgs)$	4.74
$p_T(Higgs)$	3.54
$\eta(l)$	3.10

- hiddenNodes: Defines the number of hidden nodes in the hidden layer.

For a given data set, it is difficult to predict which combination of hyperparameters is the best possible. To scan the entire phase space of 8 different hyperparameters, a script is used which at first sets all hyperparameters to only three different values. After determining which hyperparameters have a large influence on the NNout distribution and thus on the quality of the training, these hyperparameters are varied more precisely. The criteria to decide which training is the best is the Gini index and the shape of the NNout distribution itself. It is expected that in Figure 8.6 in the purity plot, the data points are statistically distributed around the diagonal for the best possible training [45]. Existing shapes around the diagonal indicate a non-optimal training as will be shown below:

Equation (8.2) describes the loss function, except for a regularization term and two regularization factors. However, the regularization term does not influence the following proof, since it would be dropped in the derivation.

$$E_D = \sum_{i=1}^{N} \ln\left(\frac{1+T_i\cdot o_i}{2}\right) \tag{8.2}$$

with N being the number of training events, T_i being the target for event i (thus -1 for a background event and +1 for a signal event) and o_i being the actual NN output for the event i. This loss function is also called the entropy loss function and it's absolute value is minimized in the learning process. The advantage of this loss-function is that the NN learns early to avoid a completely wrong assignment of signal and background because otherwise the absolute of the loss-function runs against infinity. For a fixed NN output o, the mean entropy loss function $\bar{E_D^o} = \frac{E_D}{N}$ for all events with the output o can be expressed as follows

$$\begin{aligned}
\bar{E_D^o} &= \frac{1}{N}\sum_{i=1}^{S} \ln\left(\frac{1+o}{2}\right) + \frac{1}{N}\sum_{i=S+1}^{N} \ln\left(\frac{1-o}{2}\right) && (8.3)\\
&= \frac{S}{N}\ln\left(\frac{1+o}{2}\right) + \frac{B}{N}\ln\left(\frac{1-o}{2}\right) && (8.4)\\
&= P\ln\left(\frac{1+o}{2}\right) + (1-P)\ln\left(\frac{1-o}{2}\right) && (8.5)
\end{aligned}$$

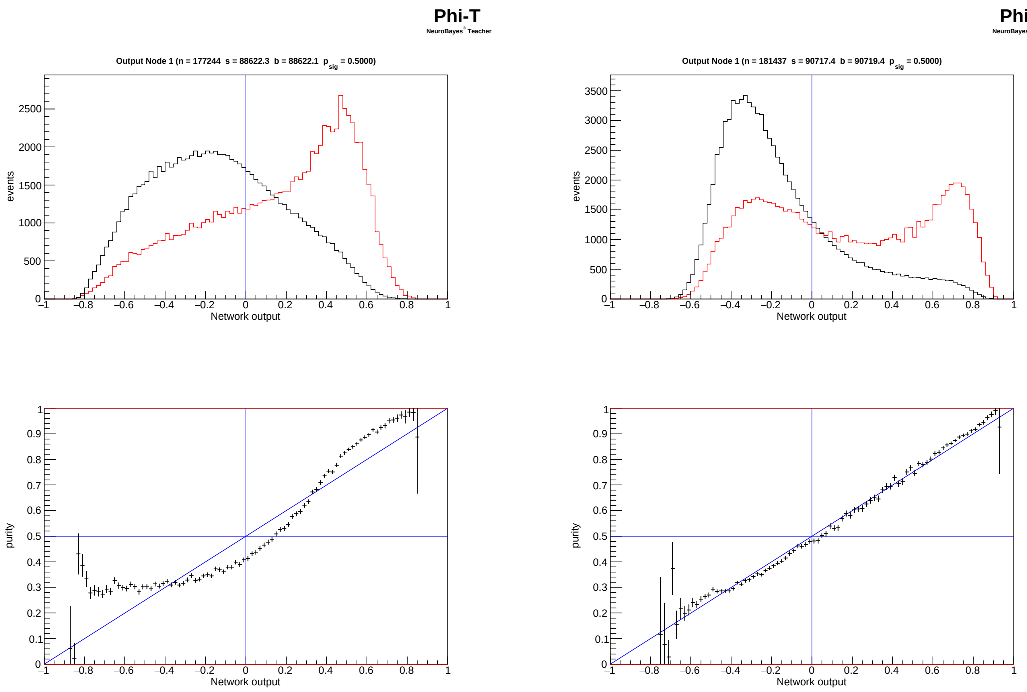
with S being the number of signal events, B being the number of background events and P being the purity. An optimal training minimizes the loss function, so the following applies $\frac{\partial \bar{E_D^o}}{\partial o} = 0$. Using Equation (8.5), it follows:

$$\begin{aligned}
0 &= P\frac{\partial}{\partial o}\ln\left(\frac{1+o}{2}\right) + (1-P)\frac{\partial}{\partial o}\ln\left(\frac{1-o}{2}\right) && (8.6)\\
&= P\cdot\frac{1}{1+o} + (1-P)\cdot\frac{1}{o-1} && (8.7)\\
\Rightarrow P &= \frac{o+1}{2} && (8.8)
\end{aligned}$$

thus one expects that for an optimal training, the purity P is distributed around the diagonal.

Figure 8.10 shows an NNout distribution on the left, which was generated with the default hyperparameters of NeuroBayes. On the right, the NNout distribution can be seen, which results from the optimization of the hyperparameters.

It can clearly be seen that the S-shape around the diagonal has disappeared due to the optimization of the hyperparameters. Also, the Gini index has improved from 23.1 to 23.7. The hyperparameters that have a significant influence on the training are “speed”, “maxlearn”, “momentum” and “hiddenNodes”. “hiddenNodes” is an important hyperparameter because it allows the NN to be sufficiently complex.



(a) (b)

Figure 8.10: *2 NNout distributions for the $utH \geq 5$-jets channel with the default hyperparameters from NeuroBayes (a) and an optimized set of hyperparameters (b).*

The other three important hyperparameters all have an influence on the NN not getting stuck in a local minimum. This seemed to be the case when using the default hyperparameters.

All optimized parameters, separated into jet multiplicity and utH- and ctH-channel, are listed in Table 8.2.

It is interesting to note that the signal in Figure 8.10(b) in the optimized NNout distribution is still divided into two classes. A naive consideration suggests that these two classes are caused by the two different FCNC processes, where the easy to separate production-FCNC process causes the class with the higher NNout and the harder to separate decay-FCNC process causes the class with the lower NNout. However, this is not the case, as Figure 8.7 shows. The decay-FCNC process is already divided into two classes: Some events are easy to separate, other events seem to be too similar to $t\bar{t}$. The production-FCNC process is also divided into two classes, although not as pronounced as the decay-FCNC process. A possible explanation for the two classes in the NNout distribution of the decay-FCNC process is that this process may be mis-reconstructed more frequently than the production-

Table 8.2: Summary of the default hyperparameters of Neurobayes and the optimized hyperparameters for each channel.

		utH			ctH		
	Default (range)	3jets	4jets	≥5jets	3jets	4jets	≥5jets
Iterations	100	100	100	100	100	100	100
maxlearn	1.0	10.0	3.0	3.0	10.0	4.0	3.0
speed	1.0	8.0	50.0	50.0	8.0	50.0	50.0
momentum	0.0 ([0,1[ )	0.85	0.8	0.75	0.85	0.75	0.8
weightupdate	200	200	200	200	200	200	200
hiddenNodes	15 ([1,99])	30	15	15	30	20	30

FCNC process. This would lead to less pronounced characteristics in the kinematic variables, which can be observed.

To test for overtraining, only 80% of the available data is used to train the NN. The remaining 20% serve as a test sample to compare the error, i.e. the value of the entropy loss function, of the NN between learn sample and test sample. It is expected that the error of the NN decreases with an increasing number of iterations and runs into a plateau. In the case of an overtrained NN, the error between the learning and the test sample will most likely differ in the area, where the plateau is expected. The distribution of the error from the NN for the learning and for the test sample in the ≥5-jets channel is shown in Figure 8.11.

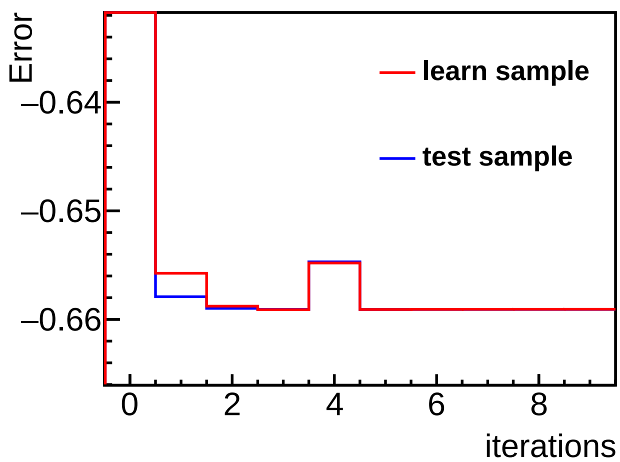


Figure 8.11: The error as a function of the training iterations for the learning and the test sample in the ≥5-jets utH channel.

The error distributions in the other channels are similar, so it is assumed that there is no overtraining.

8.6 Two-dimensional discriminant study

Section 8.4 already examined the processes on which the NN should be trained. The NN, which was only trained on one FCNC process, could also best separate this process from the background. However, the NN lost sensitivity to the other FCNC process, so it is best to have trained the NN on both processes. The motivation of the following study is to investigate, whether a more complex usage of the NN can achieve a better separation between signal and background. By creating a training that has been optimized for the production-FCNC process and one that has been optimized for the decay-FCNC process, correlations between both specialized trainings may help to achieve a better separation power.

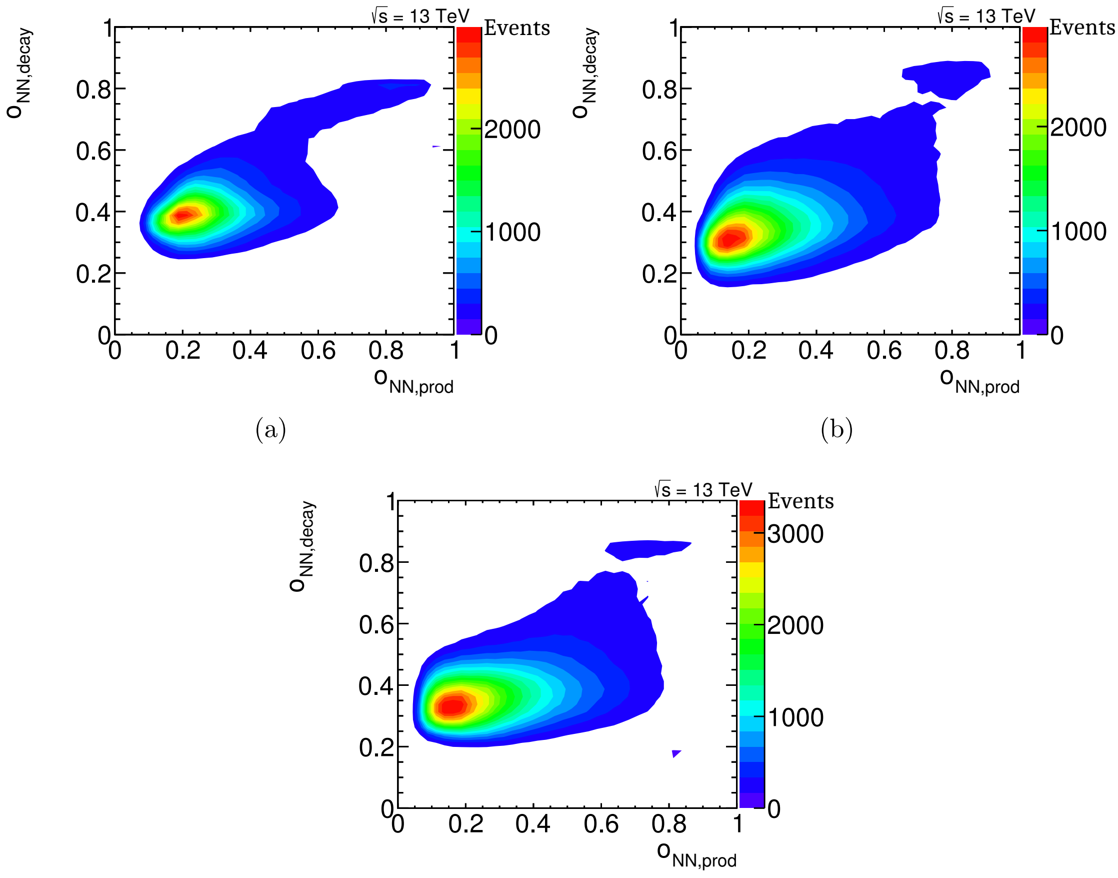
8.6.1 Two dimensional NNout

To combine both trainings, two-dimensional NNout plots are generated. The NNout of an event resulting from the production-based training is on the X-axis and the NNout resulting from the decay-based training is on the Y-axis. The two-dimensional NNout distribution for the $t\bar{t}$ background in the different jet multiplicities can be seen in Figure 8.12.

It can be seen that as the number of jets increases, it becomes more difficult for the NN, which has only been trained on the production-FCNC process, to assign $t\bar{t}$ events to low NNout values. In addition, $t\bar{t}$ has certain events in all jet multiplicities which could not be successfully assigned to low NNout values by neither of the two NN. These events have kinematic properties that are too similar to those of the FCNC signals. Thus those events cannot be assigned to low NNout values without unreasonable losses in the signal selection efficiency. The two-dimensional NNout distribution for the FCNC processes can be seen in Figure 8.13.

For the production-FCNC process, the effects of low statistics of the MC sample are noticeable because many outliers without a smooth transition can be seen. However, one can still see that the majority of all events are correctly assigned to high NNout values by both NN. With increasing jet multiplicity a bend in the direction of the lower right corner of the distribution can be seen more clearly. This means that events exist in the production-FCNC sample which are assigned to high NNout values by the production FCNC-based training, but to low NNout values by the decay-FCNC-based training. If the backgrounds do not show such behaviour, those correlations could be exploited by the two-dimensional NNout analysis to improve the final separation.

For the decay-FCNC process, it is interesting to see that the two-dimensional NNout yields two main classes of events, one that achieves high NNout values in both



(a) (b) (c)

Figure 8.12: *Two-dimensional NNout of the process $t\bar{t}$ in the 3-jets channel (a), 4-jets channel (b) and the $\geq$ 5-jets channel (c).*

NN and one that achieves low NNout values in both NN. In addition, a horizontal accumulation of events with high NNout values can be observed in all jet multiplicities, which means that these events are always reliably classified by the decay-based training to high NNout values, but not by the production-based training.

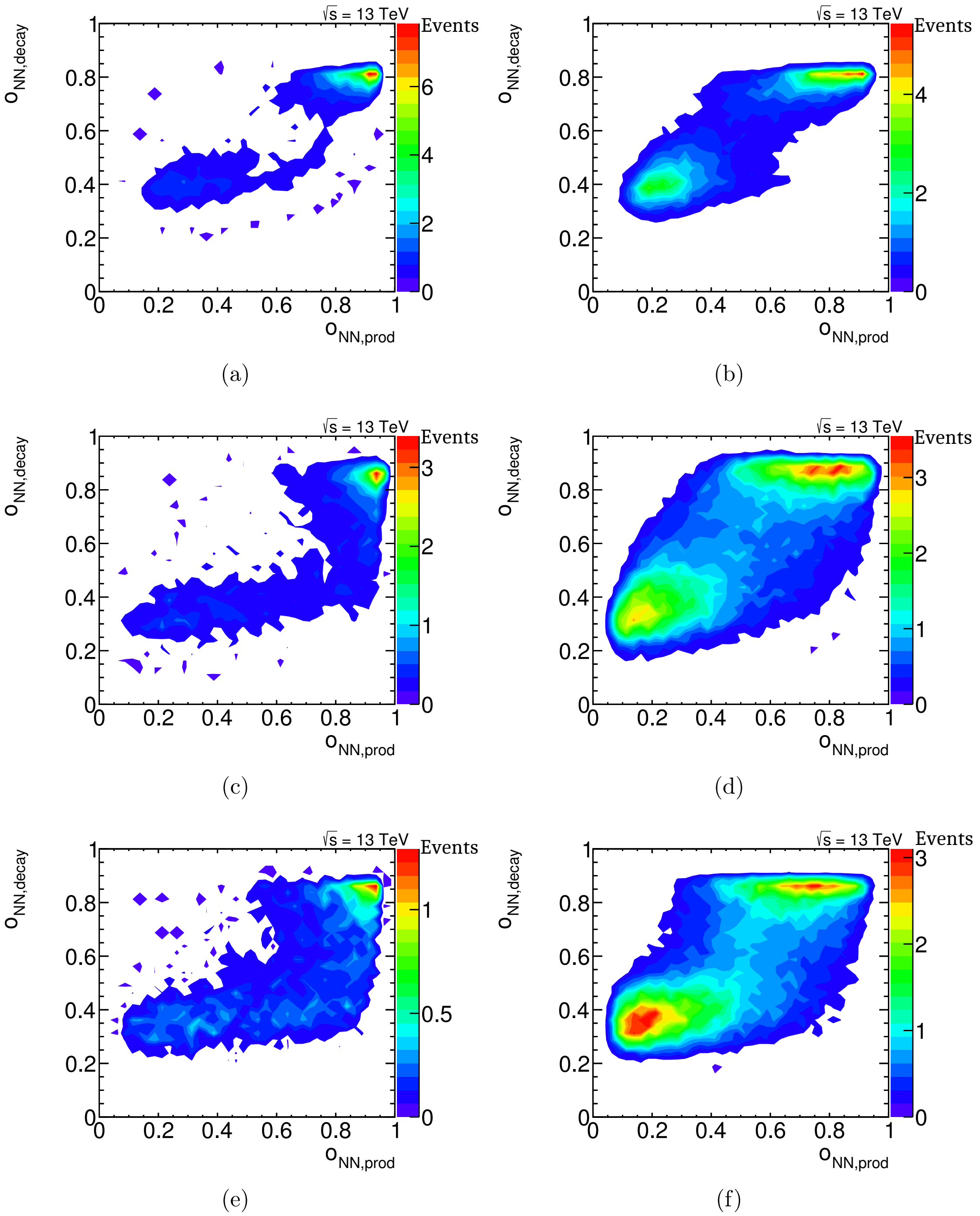


Figure 8.13: *Two-dimensional NNout of the production-FCNC process in the 3-jets channel (a), 4-jets channel (c) and $\geq$ 5-jets channel (e) and the decay-FCNC process in the 3-jets channel (b), 4-jets channel (d) and the $\geq$ 5-jets channel (f).*

8.6.2 Unrolled discriminant

Since a one-dimensional distribution is required for the statistical evaluation tool used, the two-dimensional NNout distribution must be transformed back to a one-dimensional distribution without discarding the correlations obtained. Since the two-dimensional NNout distribution of the decay-FCNC process has a horizontal accumulation of events, the following procedure is used: In steps of 0.2, the two-dimensional NNout distribution is split on the Y-axis, and the distribution along the X-axis in each subdivision is plotted in a histogram. The resulting "unrolled NNout" distribution retains the characteristics of the horizontally accumulated events. Figure 8.14 shows the unrolled NNout distribution in the 3-jets channel, Figure 8.15 in the 4-jets channel and Figure 8.16 in the $\geq$ 5-jets channel.

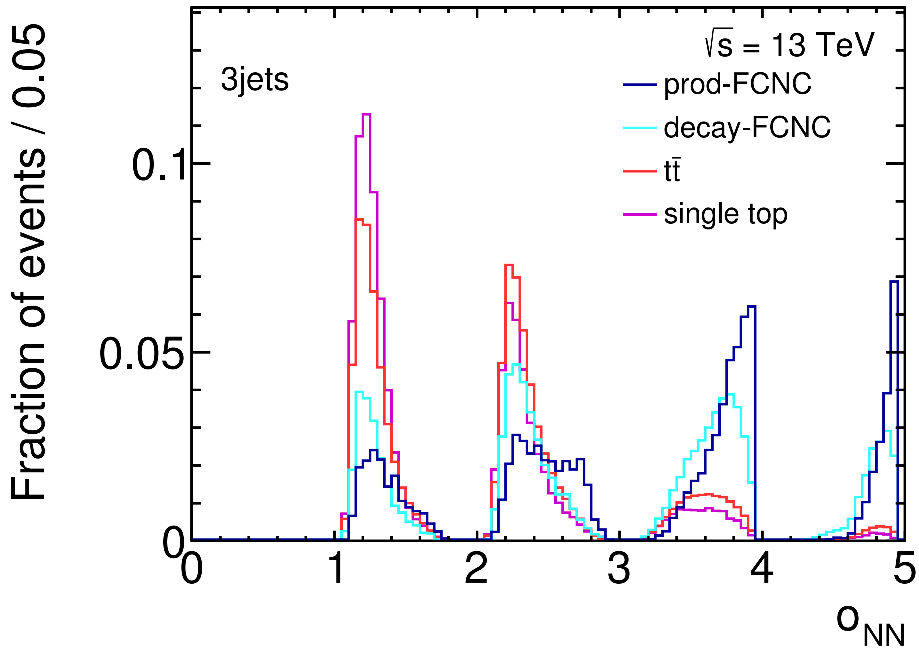


Figure 8.14: *The two-dimensional NNout distribution, unrolled to a one-dimensional in the 3-jets utH channel.*

The statistical analysis comparing the unrolled NNout with the NNout from the training, based on both FCNC processes, is done in Section 11.1.

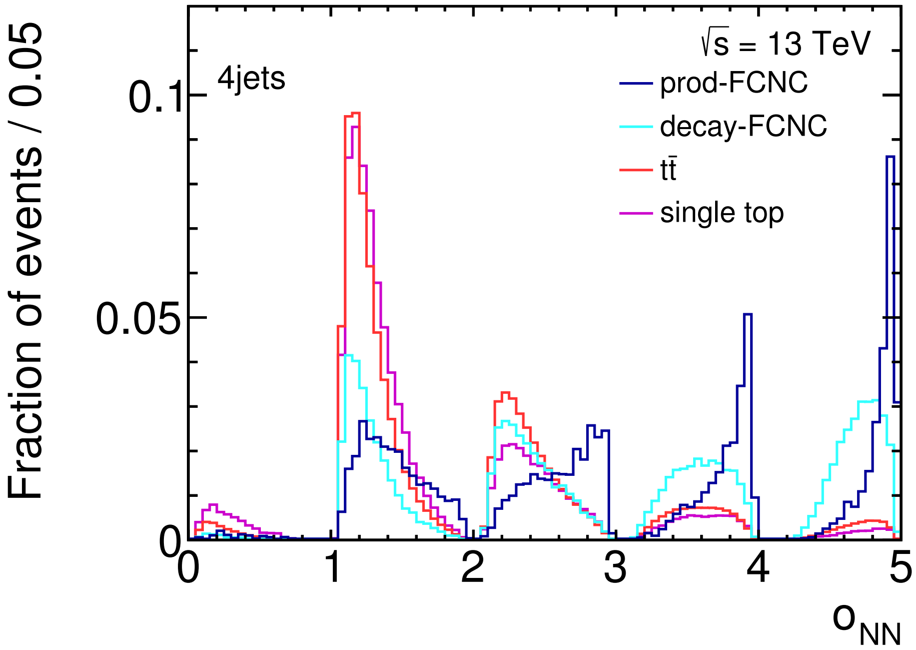


Figure 8.15: *The two-dimensional NNout distribution, unrolled to a one-dimensional in the 4-jets utH channel.*

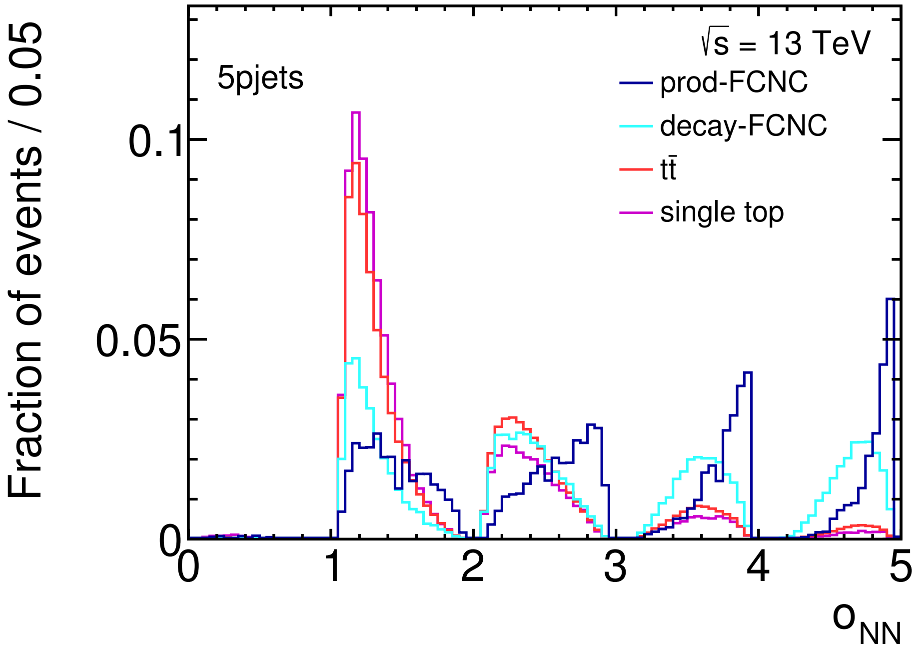


Figure 8.16: *The two-dimensional NNout distribution, unrolled to a one-dimensional in the ≥ 5-jets utH channel.*

9 Systematic uncertainties

When measuring an arbitrary quantity, two types of uncertainty may occur. On the one hand, there is the statistical uncertainty, which is caused by statistical fluctuations. This statistical uncertainty can be reduced by repeated measurements. On the other hand, systematic uncertainties may occur, for example, due to assumptions of the experimentalist or the nature of the measuring apparatus. Repeated measurements cannot reduce this systematic uncertainty since the measuring apparatus will always yield the same biased value.

This chapter describes the sources of systematic uncertainties that may have an influence on the signal strength and are thus considered in the later analysis. In general, a systematic uncertainty is a variation of a quantity by one standard deviation (1σ) upwards and downwards. In case there is no up- or down variation, but just one of them, this systematic uncertainty is called one-sided uncertainty. Those uncertainties are subsequently symmetrized, which means that the one-sided uncertainty is also applied to the other direction.

9.1 Luminosity

The generated MC events are scaled to the total integrated luminosity recorded by the ATLAS detector. To obtain the uncertainty on the luminosity, different luminosity detectors and algorithms are compared [46]. Subsequently, the absolute calibration of luminosity is performed using the van der Meer method, where x-y beam-separation scans of the luminosity scale from 2015-2018 are evaluated. The resulting systematic uncertainty on the luminosity is $\pm 1.7\%$ [47].

9.2 Pile-up

At each bunch crossing in the LHC, several inelastic collisions occur simultaneously, which is called pile-up. The pile-up distribution must generally be corrected by a scale factor to rescale the MC samples to correspond to the data [48]. An uncertainty on this scale factor is considered in the later analysis. This uncertainty is based on the disagreement between the instantaneous luminosity in MC events and in data.

9.3 Lepton uncertainties

Differences between the simulation of the MC events and the data in the identification, reconstruction, isolation and trigger efficiencies of leptons have an influence on the energy and the resolution of the leptons and are considered by scale factors. These scale factors are determined by applying tag-and-probe techniques to the resonance $Z \to l^+l^-$. The uncertainties on the scale factors are determined for electrons and muons in different ways, which is described below.

For electrons the tool in Ref. [49] is used, varying the energy calibration and the resolution of the electrons up and down by 1σ each. The systematic uncertainty on the scale factor of the energy calibration is denoted as `EG_SCALE_ALL`, the systematic uncertainty on the scale factor of the resolution as `EG_RESOLUTION_ALL`.

For muons the tool in Ref. [50] is used, providing uncertainties for the transverse momentum and resolution of the muons. Since the muons, in general, do not only leave a track in the ID, but also in MS, there are three uncertainties for the muons: `MUON_ID`, `MUON_MS` and `MUON_SCALE`. `MUON_ID` smears the track in the inner detector, `MUON_MS` smears the track in the Muon spectrometer and `MUON_SCALE` varies the transverse momentum.

9.4 Jet uncertainties

The response of jets in the calorimeters must be calibrated using MC simulations and data to determine the energy of jets in later measurements [51]. This calibration is provided with systematic uncertainties called Jet Energy Scale (JES). Up to 75 nuisance parameters, derived in p_T bins and η bins, are combined into a smaller set of two-sided systematic uncertainties. In addition to the JES uncertainties, the Jet Energy Resolution (JER) uncertainties are also considered. These are caused by the limited resolution of the calorimeters and are determined by comparing Di-jet events in the 2015 data set with MC simulations. The obtained one-sided systematic is subsequently symmetrized.

9.5 Heavy-flavor jet tagging uncertainties

The used pseudo-continuous b-tagging algorithm [52] comes along with a set of 74 independent and almost uncorrelated systematic uncertainties. Those uncertainties are determined by calibrating the b-tagging algorithm to measured b-jet efficiency and miss-tagging rates in each pseudo-continuous bin. All 74 uncertainties are considered in the later analysis. Because of the construction of the tagging uncertainties, the influence of the uncertainties is in a (depending on the considered phase space) approximate descending order for each flavor component, starting with the zeroth component.

9.6 Missing transverse momentum uncertainties

Since the $E_{\mathrm{T}}^{\mathrm{miss}}$ is composed of the sum of the reconstructed objects among others, all uncertainties of the reconstructed leptons and jets are propagated into the calculation of the uncertainties of the $E_{\mathrm{T}}^{\mathrm{miss}}$. In addition, the uncertainties of the soft tracks (i.e. the non-reconstructed objects) are considered as follows: A systematic uncertainty is implemented for the p_T of the soft tracks (`MET_SoftTrk_Scale`), one for the p_T resolution of the soft tracks parallel to the mean p_T of the hard tracks (`MET_SoftTrk_ResoPara`) and one for the p_T resolution of the soft tracks perpendicular to the mean p_T of the hard tracks (`MET_SoftTrk_ResoPerp`) [53]. Hard tracks denote tracks, which were associated to leptons or jets.

9.7 Theoretical cross section uncertainties

The prediction for the cross sections of the considered background processes is provided with systematic uncertainties. For the $t\bar{t}$ process an uncertainty of $^{+5.58\%}_{-6.11\%}$ [54] is used, for the single top-quark t-channel production processes an uncertainty of $^{+4.11\%}_{-5.51\%}$, for the s-channel an uncertainty of $^{+3.38\%}_{-3.49\%}$ and for the Wt-channel an uncertainty of $^{+5.37\%}_{-5.37\%}$ [55]. For the remaining considered processes W+jets, Z+jets and "diboson" comparatively large uncertainties of $\pm 20\%$ are assumed.

9.8 Modelling uncertainties

Since $t\bar{t}$ and single top-quark production are the largest backgrounds, the systematic uncertainty of the modeling is considered for these. The same process is simulated with the help of other generators / other parameters in the simulation. The difference between the resulting samples is then assumed as an uncertainty. In the case of another generator, where no uncertainty can be determined upwards and downwards, since there is only one more "data point", this uncertainty is symmetrized. For varying the parton shower and the hadronization, the nominal generator "Pythia" is compared with "Herwig". For the ISR and FSR the parameters

μ_f, μ_r and h_{damp} are varied. The μ_f parameter provides the factorization scale, μ_r the re-normalization scale and h_{damp} is a parameter that effectively regulates the high-p_T radiations.

9.9 Parton Distribution Function uncertainties

To consider the uncertainty of the Parton Distribution Function (PDF), events are reweighted using the combined PDF set "PDF4LHC15" [56]. This PDF set has 30 eigenvectors. Each eigenvector leads to systematic variation, which is later taken into account.

10 Statistical analysis

This chapter describes the working principle of a binned profile likelihood fit and further important aspects like binning and smoothing of systematic uncertainties. Furthermore, the computation of exclusion limits is discussed.

The motivation to use a profile likelihood approach for statistical evaluation is, among other things, the possibility of constraining systematic uncertainties using control regions. Those constrained systematic uncertainties can subsequently be propagated into the signal region and thus reduce the uncertainty on the signal.

10.1 Binned profile likelihood fit

With the NNout distribution of the expectation (MC) and of the data, the agreement of both distributions can be evaluated with the help of a binned profile likelihood fit. The evaluation of the agreement is done by maximizing the following function:

$$\mathcal{L}(\mu, \vec{\alpha}) = \prod_{i=1}^{N_{bins}} \mathcal{P}\left(n_i|\mu \cdot \left(s_i^{prod}(\vec{\alpha}) + s_i^{decay}(\vec{\alpha})\right) + b_i(\vec{\alpha})\right) \quad \cdot \quad \prod_{j=1}^{N_{NP}} \mathcal{G}\left(\alpha_j|0,1\right) \tag{10.1}$$

with $\mathcal{P}(x|y)$ being the Poisson probability distribution at x with the parameter y and $\mathcal{G}(x|y_1, y_2)$ being the Gauß function at x with mean y_1 and standard deviation y_2. μ is a so-called "parameter of interest" (POI), a free-floating parameter which characterizes the strength of the signal processes and is determined by maximizing the likelihood function. Furthermore, n_i is the number of measured events in bin i, $s_i^{prod,decay}$ is the number of MC events in bin i from the production-FCNC process and the decay-FCNC process respectively, b_i is the number of MC events from the backgrounds in bin i and $\vec{\alpha}$ is a vector of all nuisance parameters (NP). Each systematic uncertainty belongs to one NP. The NP α_j is a measure of how much the shape or normalization of a process has been adapted to those of the corresponding systematic uncertainty j. If the nominal template has been replaced by the upward (downward) systematic variation, the corresponding NP is at 1 (-1). In between, interpolation is performed. For the shape extrapolation, a piecewise-linear function is used. For the normalization extrapolation, a polynomial/exponential function is used [57]. An NP $\neq 0$ is also called a "pull" on the corresponding systematic uncertainty. In addition to the NP for the systematic uncertainties, each bin in each region also has its own NP, which are called gamma factors. These gamma factors account for the MC statistical uncertainties in the respective bins.

The magnitude of systematic uncertainties can be decreased by the fit. To determine the post-fit uncertainty, the NP of an uncertainty is varied and the likelihood function is maximized again. The value of the NP, where the likelihood function deviates by 1σ, is the post-fit uncertainty of the respective systematic uncertainty. If the post-fit uncertainty is smaller than the pre-fit uncertainty, the systematic uncertainty is called "constrained".

10.2 Binning, smoothing and pruning

The NNout distribution of the NN is in general rebinned. A higher number of bins means a higher sensitivity to the signal because more bins better preserve the shape of the signal. Furthermore, systematic uncertainties might be constrained in the control region which then can be transferred as information into the signal region. These influences improve the uncertainty of the signal strength. However, too many bins are accompanied by statistical fluctuations in individual bins, so that in the end a compromise must always be found. The statistical analysis framework, used in this analysis, offers an automatic binning. With the automatic binning the bin boundaries are calculated based on 2 parameters. The ratio of both parameters gives a measure of how flat the signal or background should be. If the ratio is large, the boundaries are designed such that the signal is as flat as possible. A ratio of zero results in a flat background. The sum of both parameters gives the total number of bins used.

Smoothing is used for systematic uncertainties, as these are often not generated with the same number of MC events as the nominal templates. As a result, systematic uncertainties are more likely to be affected by statistical fluctuations. The statistical uncertainties of the systematic uncertainties are, however, not taken into account in the fit. This may lead to nonphysical constraints or pulls of systematic uncertainties in the likelihood fit. For this reason, smoothing is used, which determines an approximated function of the systematic uncertainty that (ideally) preserves the basic characteristics but smoothes out statistical fluctuations.

Systematic uncertainties can be "pruned", which means that these systematic uncertainties are not considered in the likelihood fit. In order to prune the shape of a systematic uncertainty, it is checked whether at least one bin (after normalizing the events of the systematic uncertainty to those of the nominal sample) has a deviation of $\geq 1\%$. If this is not the case, the shape of the systematic uncertainty is pruned. The normalization of the systematic uncertainty is pruned if the amount of events differs less than 1% from the amount of events in the nominal sample. Pruning the systematic uncertainties helps to significantly reduce the required computing power in the likelihood fit.

10.3 Asimov template

An Asimov template contains all background MC processes considered. This can be treated as if it were data. A likelihood fit, using the Asimov template as data, is also called an Asimov fit. An Asimov fit is used to obtain an expected upper exclusion limit that takes into account the considered integrated luminosity of the data set and already indicates constraints on systematic uncertainties without using actual data. The only difference between the Asimov template and the MC template is the statistical uncertainty ΔN_i in the individual bins i, which is calculated according to $\Delta N_i = \sqrt{N_i}$ with N_i being the number of expected events in bin i. This means that no pulls in systematic uncertainties can be observed or BSM processes discovered. However, the expected upper exclusion limit on the signal strength μ is a measure for the sensitivity of the physics analysis, since the amount and distribution of bins also have an influence.

10.4 Exclusion limits and the CLs method

If no significant excess of data over the MC expectation is found, an upper exclusion limit on the signal strength is determined. This exclusion limit describes a signal strength where all higher values can be excluded with a probability of at least 95%. The null hypothesis is defined as the hypothesis that only includes all backgrounds, i.e. all SM processes ($\mu = 0$). In contrast, there is the alternative hypothesis, including all SM processes plus the FCNC signals ($\mu > 0$). The test statistic $\lambda(\mu)$ is defined in Equation (10.2)

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\hat{\vec{\alpha}}})}{\mathcal{L}(\hat{\mu}, \hat{\vec{\alpha}})} \tag{10.2}$$

with $\hat{\hat{\vec{\alpha}}}$ being the set of nuisance parameters, which maximize the likelihood function for a given μ. $\hat{\mu}$ and $\hat{\vec{\alpha}}$ is the signal strength and the set of nuisance parameters, which overall maximize the likelihood function. Thus for the test statistics $\lambda(\mu) \in]0, 1] \; \forall \; \lambda$

applies, where the value μ, which best describes the data, generates a value of 1 for the test statistics. With

$$q_\mu = -2 \cdot \ln\left(\lambda(\mu)\right) \tag{10.3}$$

and the probability density function $f\left(q_\mu|\mu\right)$, which describes the distribution of q_μ for a given μ, one can define the p-value as follows:

$$p_\mu = \int_{q_{\mu,obs}}^{\infty} f\left(q_\mu|\mu\right) dq_\mu \tag{10.4}$$

The p-value for a given μ is the probability to obtain test results at least as extreme as the observed results assuming that the null hypothesis is correct. To define exclusion limits that are not affected by statistical fluctuations of the background, the CLs method [58, 59] is used:

$$CLs = \frac{p_\mu}{1 - p_0} \tag{10.5}$$

with p_μ being the p-value for a μ, such that $p_\mu = 0.05$ applies. p_0 is the p-value for $\mu = 0$. Given Equation (10.5), signals with a signal strength μ can be excluded at a confidence level of 95%, if $CLs < 0.05$ holds.

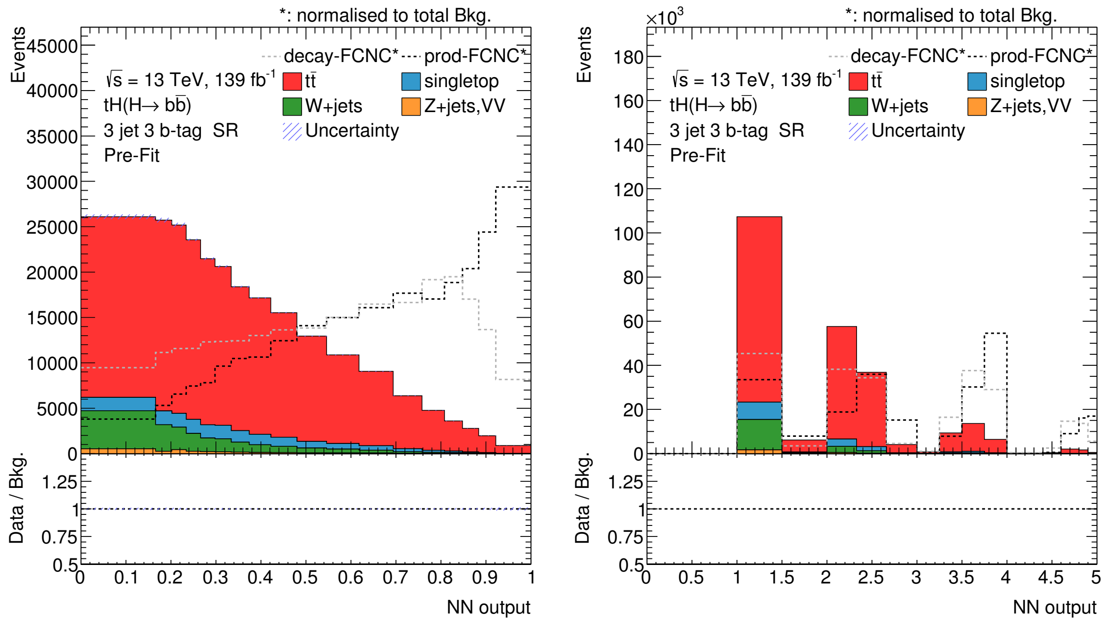
11 Results and interpretation

This chapter first compares the unrolled NNout distribution with the NNout from the training, based on both FCNC processes. Subsequently, the result of the statistical analysis of the NNout distribution including the systematic uncertainties and the different interpretations of the final result are presented.

11.1 Statistical analysis of the unrolled NNout

The statistical evaluation framework "TRExFitter" [60] is used for the statistical analysis of the NNout distribution. This section compares the sensitivity of the unrolled- with the baseline NNout distribution. The baseline NNout distribution denotes the NNout distribution from the NN, which was trained simultaneously on both FCNC processes. This comparison considers only the main backgrounds without the splitting of the $t\bar{t}$ background. Furthermore, this comparison does not include the effect of systematic uncertainties. When comparing the expected upper exclusion limits of the FCNC signals with TRExFitter, it should be considered that rebinning and especially the number of bins used have a large effect on the expected upper exclusion limits. Since no systematic uncertainties are considered in this section, it is important not to use too many bins, which is normally prevented by the systematic uncertainties. In addition, the same binning cannot be used for both NNout distributions because, for example, bins in the unrolled NNout distribution that would combine two of the five subsections are not physically meaningful. This means that the automated binning of TRExFitter cannot be used for the unrolled NNout distribution. The baseline- and the unrolled NNout distribution, processed by TRExFitter, can be seen in Figure 11.1. For the baseline NNout, the automatic binning of TRExFitter with the parameters (12,6) was used. For the unrolled NNout distribution the following binning is used: The first subdivision in combined into a single bin, the second subdivision into two bins and so on. In total, 15 bins are used for the unrolled NNout distribution. A similar amount of bins in each distribution ensures that the determined upper exclusion limits can be compared.

For determining the exclusion limits, a simultaneous fit in all three signal regions is performed for each NNout distribution, respectively. The extracted upper exclusion limits on the signal strength are shown in Table 11.1.



(a)

(b)

Figure 11.1: *Rebinned baseline NNout (a) and the rebinned unrolled NNout distribution (b) in the 3-jets channel for the left-handed FCNC processes. The error band includes statistical uncertainties only.*

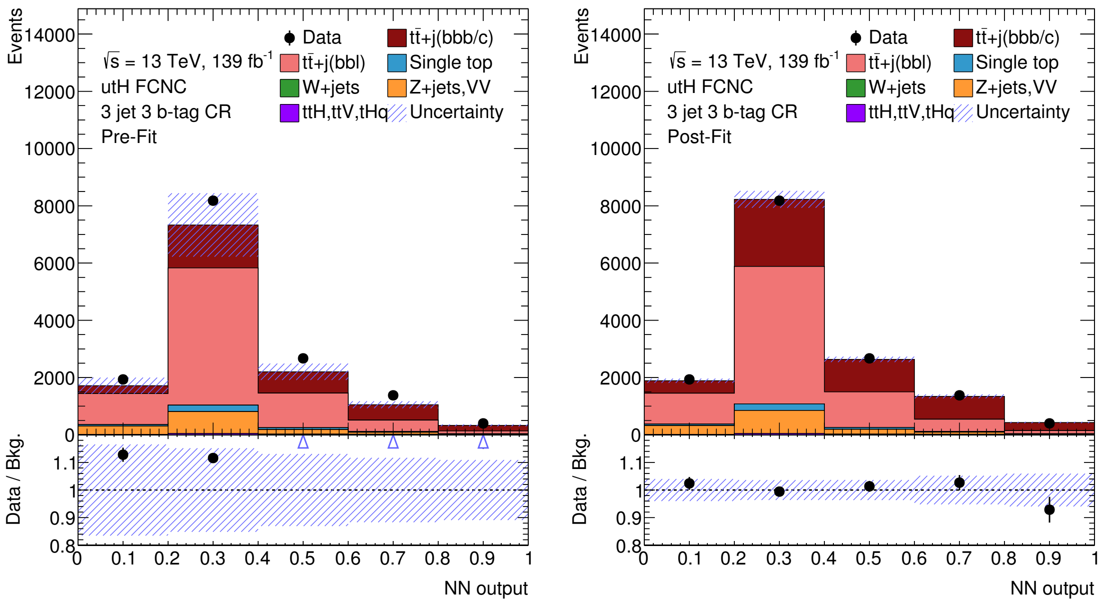
Table 11.1: *Expected upper exclusion limits on the signal strength for both NNout distributions with the corresponding 1σ and 2σ bands. Only the statistical uncertainty is included in the fit.*

	-2σ	-1σ	exp. upper limit	$+1\sigma$	$+2\sigma$
baseline NNout	0.353	0.474	0.658	0.916	1.23
unrolled NNout	0.372	0.499	0.692	0.964	1.29

It turns out that the upper exclusion limit, which is determined using the unrolled NNout distribution, is about 5% worse than the upper exclusion limit of the baseline NNout distribution. One possible reason for this is that the binning has not been optimized for the unrolled NNout distribution. Although the baseline NNout distribution has not been optimized either, it may be that an optimized binning is more important for the unrolled NNout distribution than for the baseline NNout distribution. A further reason could be that the determination of the one-dimensional distribution from the two-dimensional NNout distribution was not optimally chosen in order to preserve the characteristics as well as possible. Another approach could produce a distribution from the two-dimensional NNout distribution, which better preserves the characteristics and thus achieves a stronger separation strength.

11.2 Scale factor for the $t\bar{t}$ + HF estimate

As already discussed in Chapter 6.2, one scale factor for the $t\bar{t}$ + HF normalization is determined from a combined fit to data in the control regions. The pre- and postfit plots for the 3-jets, 4-jets and ≥5-jets channel in the control region are shown in Figure 11.2 - 11.4. The uncertainty bands contain statistical and systematic uncertainties.



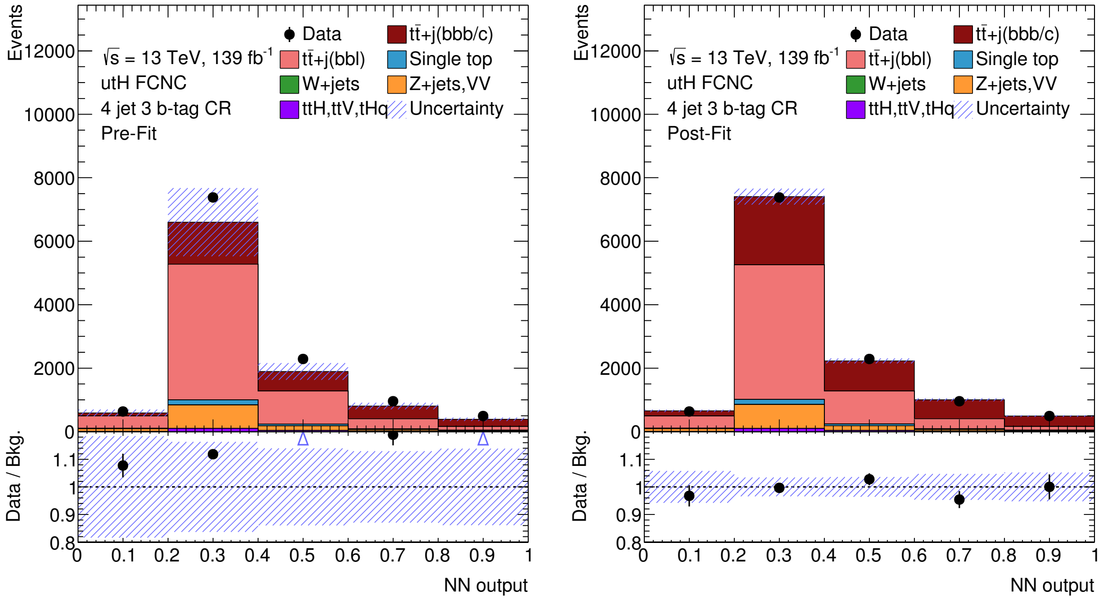
(a) (b)

Figure 11.2: *Prefit (a) and postfit (b) NNout distributions in the 3-jets control region.*

The resulting scale factor for the $t\bar{t}$ + HF process, which is denoted as $t\bar{t}$ + j(bbb/c) in the plots, is

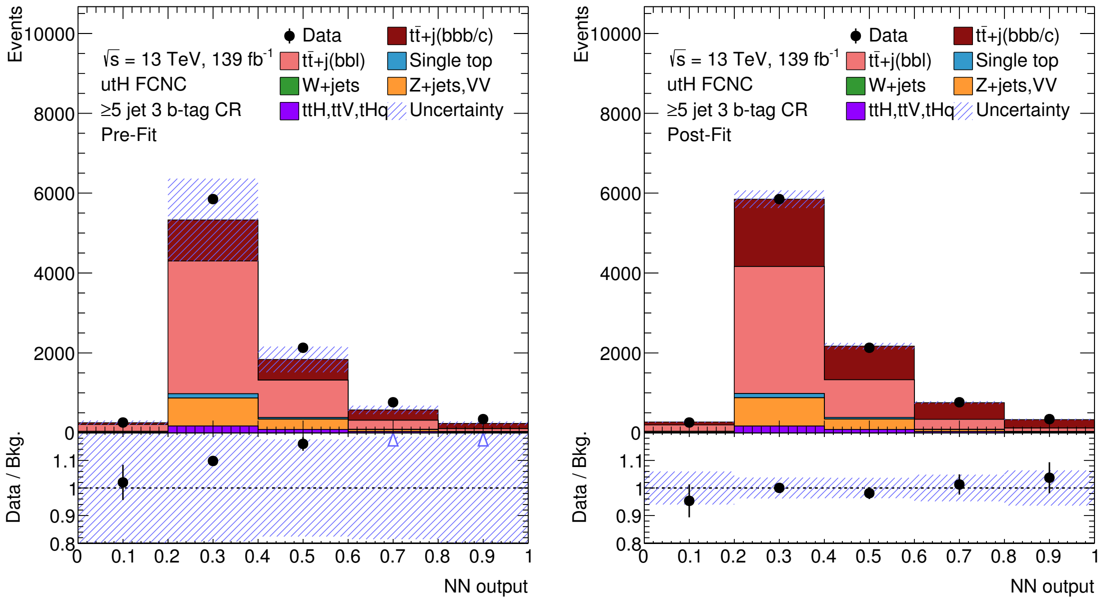
$$\beta_{t\bar{t}+HF} = 1.41^{+0.17}_{-0.15} \,. \tag{11.1}$$

The considered systematic uncertainties are shown in Figure 13.1 - 13.10. It can be seen that in the prefit plot the data does not agree with the MC prediction within the



(a) (b)

Figure 11.3: *Prefit (a) and postfit (b) NNout distributions in the 4-jets control region.*



(a) (b)

Figure 11.4: *Prefit (a) and postfit (b) NNout distributions in the ≥ 5-jets control region.*

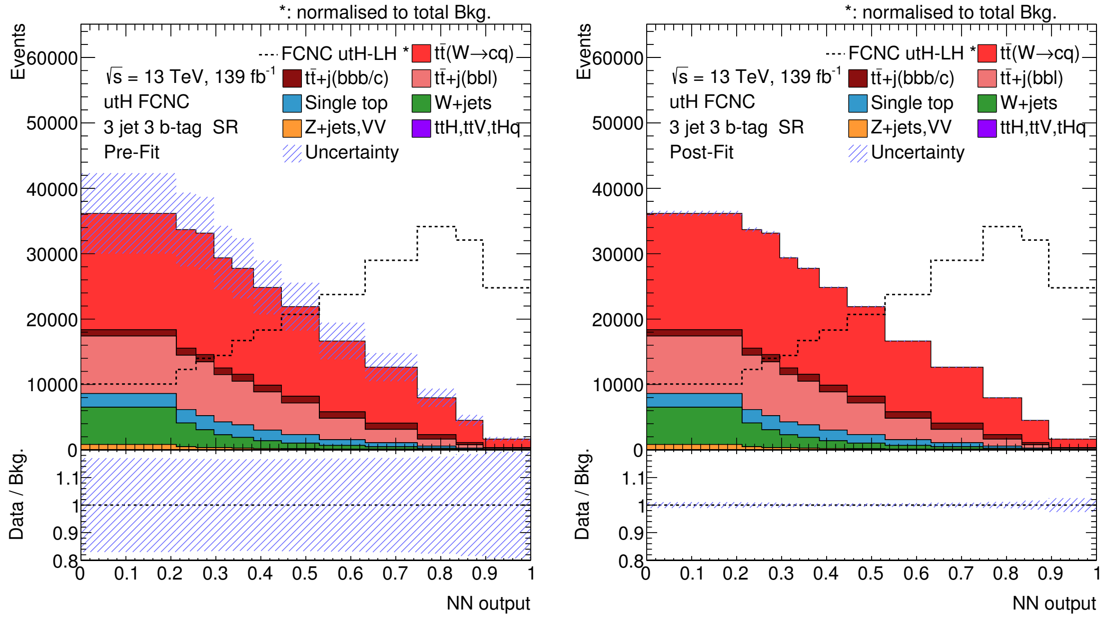
uncertainties. This is due to the fact that the uncertainty on the MC processes does not include any specific uncertainty for the rate prediction of the $t\bar{t}$+HF process, since exactly this rate is fitted as a free floating parameter. The data points in the post-fit plots coincide to the MC within the uncertainties. The uncertainty band in the post-fit plots becomes smaller due to constraints and correlations of systematic uncertainties.

11.3 Asimov fit result in the signal region with systematic uncertainties

The scale factor determined, together with its uncertainty, is applied in the signal region, which influences the number of MC events and thus the uncertainty. A binned profile likelihood fit on an Asimov data set is then performed with the considered systematic uncertainties (Figure 13.1 - 13.10). Plots which summarize the constraints on the systematic uncertainties can be seen in Figure 13.11 - 13.14. Constraints exist in the b-tagging and the shower systematic uncertainties. In case of the tagging uncertainties, the zeroth components of the c component and the light component of the tagging uncertainties are constrained (called `bTagSF_PC_C_0` and `bTagSF_PC_Light_0` in Figure 13.11). Because the main background $t\bar{t}$ requires to have one mis-tagged jet, the signal strength μ can be sensitive to the c- and light components of the tagging algorithm.

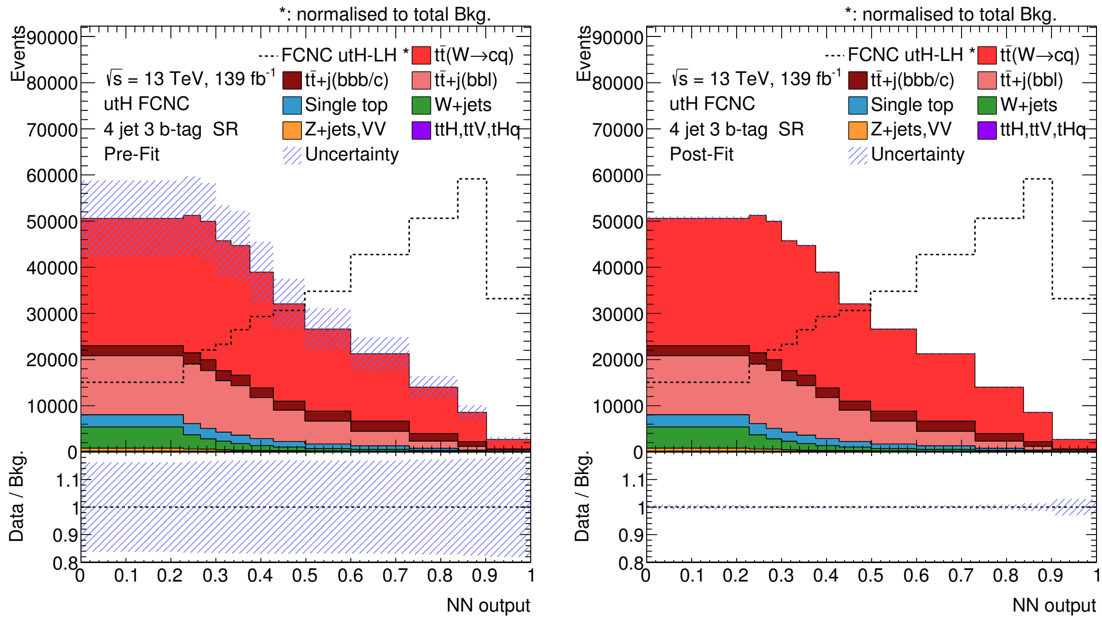
Normally, the $t\bar{t}$ shower systematic uncertainty should be determined by another generator and taken into account in the fit. However, this could not be achieved due to technical problems. In order to at least partially compensate for the effect of this systematic, an overall uncertainty of 20% on $t\bar{t}$ is assumed. However, this does not take into account the shape of the uncertainty. Since there is already a systematic uncertainty concerning the overall normalization of $t\bar{t}$, it is very likely that the uncertainty on the overall normalization is overestimated, which explains the constraint on the $t\bar{t}$ shower uncertainty (called $t\bar{t}$ `shower` in Figure 13.13).

For each FCNC vertex (left/right-handed utH/ctH) considered, a combined fit to Asimov data in all three signal regions is performed to determine an expected upper exclusion limit with a CL of 95% on the signal strength μ. The pre- and postfit plots for the left-handed utH process are shown in Figure 11.5 - 11.7. "FCNC utH-LH" denotes both, the left-handed production- and decay-FCNC process. The expected upper exclusion limits of the signal strengths are listed in Table 11.2.



(a) (b)

Figure 11.5: *Prefit (a) and postfit (b) NNout distributions in the 3-jets signal region.*



(a) (b)

Figure 11.6: *Prefit (a) and postfit (b) NNout distributions in the 4-jets signal region.*

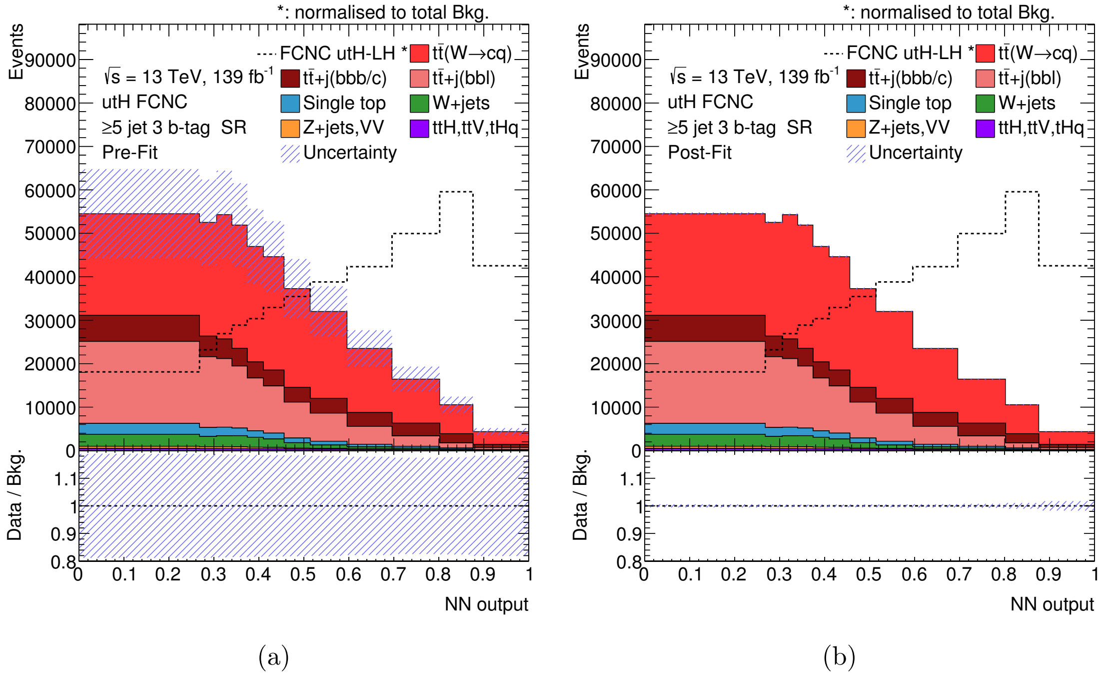


Figure 11.7: *Prefit (a) and postfit (b) NNout distributions in the* ≥ 5*-jets signal region.*

Table 11.2: *Results of the expected upper exclusion limits on the signal strength for the four different couplings considered.*

	-2σ	-1σ	exp. upper limit	$+1\sigma$	$+2\sigma$
$\mu_{exp}^{utH,lh}$	1.25	1.68	2.33	3.19	4.24
$\mu_{exp}^{utH,rh}$	1.37	1.84	2.56	3.49	4.63
$\mu_{exp}^{ctH,lh}$	2.62	3.52	4.89	6.54	8.56
$\mu_{exp}^{ctH,rh}$	2.59	3.47	4.82	6.43	8.41

The uncertainty band in Figure 11.5 - 11.7 contains all statistical and systematic uncertainties. It can be seen, that the uncertainty is significantly reduced by the fit, which can be explained by the constraints on the systematic uncertainties and by exploiting the correlations between the systematic uncertainties. It is interesting to see, that the inclusion of the systematic uncertainties results into expected upper exclusion limits, which are nearly one order of magnitude worse than the expected upper exclusion limits without systematic uncertainties.

Furthermore it should be noted, that the NN was not optimized for the ctH process, but instead the same NN was used as in the utH case. This may lead to worse upper exclusion limits for the ctH process, since the NN is no longer optimized for the respective process. This was done due to time and computing power reasons. As already discussed, the shower systematic uncertainty of $t\bar{t}$ was not treated optimally in the fit, which could also have a negative effect on the result. The overall difference between the expected upper exclusion limits on the utH and the ctH processes can be explained by the different production rate and kinematic differences between both processes.

The ranking plot can be seen in Figure 11.8. This plot sorts the systematic uncertainties according to their influence on the signal strength. All systematic uncertainties are successively set to a standard deviation upwards and downwards and the fit is repeated. The systematic uncertainties that most influence the signal strength are recorded in descending order in the ranking plot. The bluish bars indicate the influence of the respective NP on the signal strength, which must be read at the upper axis. Furthermore the constraints of the ranked NP can be seen at the lower axis.

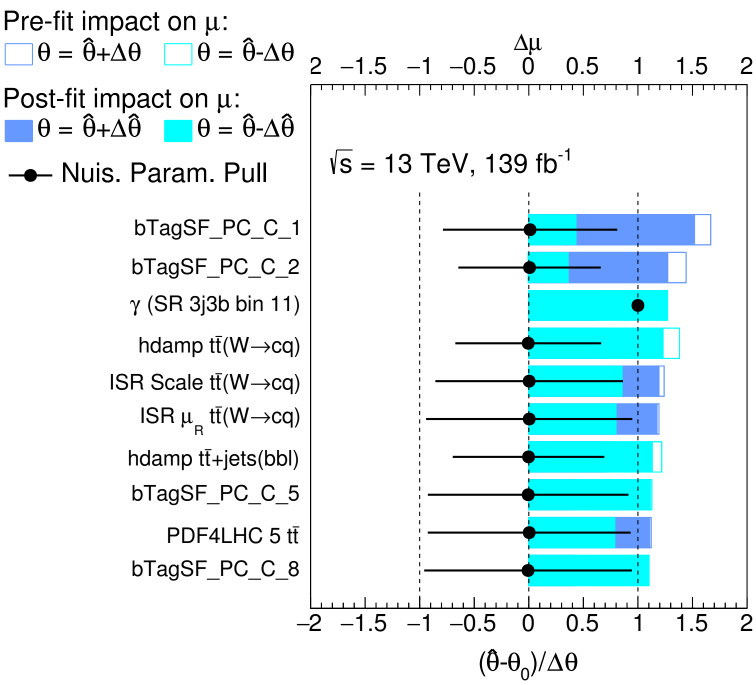


Figure 11.8: *Ranking plot, which sorts the systematic uncertainties by means of their influence on the signal strength in descending order.*

First, it is noticeable that the ranking plot only indicates positive changes in μ. This is due to the used Asimov data set, which was created without the FCNC processes ($\mu = 0$). Since μ cannot become smaller than zero by construction, the ranking plot shows no influences of systematic uncertainties on the signal strength in the negative direction. Besides that, it can be seen that the systematic uncertainties of the c component of the flavor tagging are the two uncertainties that have the largest influence on the signal strength with their variations. The 11th bin in the signal region is the bin with the best signal-background-ratio (the numbering starts with the zeroth bin and the signal regions have 12 bins each), so that also the MC statistics has an influence on the final result. Most of the remaining uncertainties have to do with the $t\bar{t}$ modeling. Since $t\bar{t}$ is the major background, which also provides most events from all backgrounds in the high-NNout-value regime, it is comprehensible that the $t\bar{t}$ modeling has an influence on the signal strength.

11.4 Considered interpretations

This section summarizes the derivations of the interpretations of the upper exclusion limits determined.

11.4.1 Interpretation in terms of a cross section

The determined upper exclusion limit on the signal strength μ can be used to calculate an upper exclusion limit on the cross section of the FCNC processes. The cross section of the FCNC processes is related to the signal strength as follows:

$$\sigma_{tot}^{lim} = \sigma_{tot}^{MC} \cdot \mu^{lim} \Rightarrow \mu^{lim} = \frac{\sigma_{tot}^{lim}}{\sigma_{tot}^{MC}} = \frac{\sigma_{prod}^{lim} + \sigma_{decay}^{lim}}{\sigma_{prod}^{MC} + \sigma_{decay}^{MC}} \tag{11.2}$$

As discussed in Chapter 2.4.4, for a $C_{u\phi} = 1$ as used in the MC production, one can use the following approximation

$$\frac{\sigma_{prod}}{\sigma_{decay}} = const \quad . \tag{11.3}$$

Since the fraction of cross sections for the MC processes and the cross section limits is the same constant, one can conclude that

$$\frac{\sigma_{prod}^{lim}}{\sigma_{decay}^{lim}} = \frac{\sigma_{prod}^{MC}}{\sigma_{decay}^{MC}} \Rightarrow \sigma_{decay}^{lim} = \sigma_{prod}^{lim} \cdot \frac{\sigma_{decay}^{MC}}{\sigma_{prod}^{MC}} \quad . \tag{11.4}$$

Injecting Equation (11.4) into (11.2) yields an equation for the limit of the signal strength, depending on the production-FCNC process cross section:

$$\mu^{lim} = \frac{\sigma_{prod}^{lim}}{\sigma_{prod}^{MC}} \tag{11.5}$$

thus the following holds

$$\boxed{\sigma_{prod}^{lim} = \mu^{lim} \cdot \sigma_{prod}^{MC}}. \tag{11.6}$$

The results for the production-FCNC process cross section can be seen in Section 11.5.

11.4.2 Interpretation in terms of a Wilson coefficient

The determined upper exclusion limit of the signal strength can also be used to calculate an upper exclusion limit on the Wilson coefficient in the EFT context.

Inserting Equation (2.9) into (11.5), one obtains

$$\mu^{lim} = \frac{\sigma_{decay}^{lim}}{\sigma_{prod}^{MC}} = \left(\frac{C_{u\phi}^{lim}}{C_{u\phi}^{MC}} \right)^2 \tag{11.7}$$

since the constant a in Equation (2.9) does not depend on the Wilson coefficient $C_{u\phi}$.

By transforming Equation (11.7), one obtains an equation for the upper exclusion limit on the Wilson coefficient of the EFT:

$$\boxed{C_{u\phi}^{lim} = \sqrt{\mu^{lim}} \cdot C_{u\phi}^{MC}}. \tag{11.8}$$

The results for the upper exclusion limits of the Wilson coefficients can be seen in Section 11.5.

11.4.3 Interpretation in terms of a branching ratio

In order to be able to compare the results of this analysis with other analyses without having to rely on the same EFT, the signal strength is also interpreted as an upper exclusion limit on the branching ratio $\mathcal{B}(t \to qH)$.

Starting from Equation (11.7) and inserting Equation (2.15), one obtains

$$\mu^{lim} = \frac{\Gamma^{lim}_{t\to qH}}{\Gamma^{MC}_{t\to qH}}. \tag{11.9}$$

Inserting the approximation $\mathcal{B}(t \to qH) \approx \frac{\Gamma_{t\to qH}}{\Gamma_{t\to Wb}}$ yields

$$\mu^{lim} = \frac{\mathcal{B}(t\to qH)^{lim}}{\mathcal{B}(t\to qH)^{MC}} \tag{11.10}$$

$$\Rightarrow \boxed{\mathcal{B}(t\to qH)^{lim} = \mu^{lim}\cdot\mathcal{B}(t\to qH)^{MC}}. \tag{11.11}$$

The approximation used is justified since $\Gamma_{t\to qH}$ is negligible small compared to $\Gamma_{t\to Wb}$ at a Wilson coefficient of 1, which was used for producing the MC samples.

The results for the upper exclusion limits on the branching ratios for the process $t \to qH$ can be seen in Section 11.5.

11.5 Results of the various interpretations

The following tables contain the expected upper exclusion limits including the uncertainty bands for the various interpretations for the four FCNC couplings considered.

Table 11.3: *Results of the various interpretations for the expected upper limit μ^{lim} for the left-handed utH coupling.*

	-2σ	-1σ	exp. upper limit	$+1\sigma$	$+2\sigma$
μ^{lim}	1.25	1.68	2.33	3.19	4.24
σ^{lim}_{prod} [pb]	0.457	0.614	0.852	1.16	1.55
$C^{lim}_{u\phi}$	1.12	1.30	1.53	1.79	2.06
$\mathcal{B}(t\to uH)[\cdot 10^{-3}]$	0.854	1.15	1.59	2.17	2.89

Table 11.4: *Results of the various interpretations for the expected upper limit* μ^{lim} *for the right-handed* utH *coupling.*

	-2σ	-1σ	exp. upper limit	$+1\sigma$	$+2\sigma$
μ^{lim}	1.37	1.84	2.56	3.49	4.63
σ^{lim}_{prod} [pb]	0.502	0.673	0.935	1.27	1.69
$C^{lim}_{u\phi}$	1.17	1.36	1.60	1.87	2.15
$\mathcal{B}(t \to uH)[\cdot 10^{-3}]$	0.936	1.26	1.74	2.38	3.16

Table 11.5: *Results of the various interpretations for the expected upper limit* μ^{lim} *for the left-handed* ctH *coupling.*

	-2σ	-1σ	exp. upper limit	$+1\sigma$	$+2\sigma$
μ^{lim}	2.62	3.52	4.89	6.54	8.555
σ^{lim}_{prod} [pb]	0.139	0.186	0.258	0.346	0.452
$C^{lim}_{c\phi}$	1.62	1.88	2.21	2.56	2.93
$\mathcal{B}(t \to cH)[\cdot 10^{-3}]$	1.79	2.40	3.33	4.45	5.82

Table 11.6: *Results of the various interpretations for the expected upper limit* μ^{lim} *for the right-handed* ctH *coupling.*

	-2σ	-1σ	exp. upper limit	$+1\sigma$	$+2\sigma$
μ^{lim}	2.59	3.47	4.82	6.43	8.41
σ^{lim}_{prod} [pb]	0.137	0.184	0.255	0.340	0.445
$C^{lim}_{c\phi}$	1.61	1.86	2.20	2.54	2.90
$\mathcal{B}(t \to cH)$ $[\cdot 10^{-3}]$	1.76	2.36	3.28	4.38	5.72

The most recent published 95% CL expected upper exclusion limit on the branching ratios are $\mathcal{B}(t \to uH) = 4.9 \times 10^{-3}$ and $\mathcal{B}(t \to cH) = 4.0 \times 10^{-3}$ [61]. This analysis was carried out with $\int L\ dt = 36\,\text{fb}^{-1}$ with the decay-FCNC process only, using the $H \to b\bar{b}$ channel. As can be seen, the obtained expected upper exclusion limits for the branching ratio are better than the most recent results, especially for the utH case. One reason for the improvement is the improved considered statistics of $\int L\ dt = 139\,\text{fb}^{-1}$. In order to further improve the expected upper exclusion limit obtained, the binning in TRExFitter could be optimized individually for each process. In addition, the NN could also be trained for the ctH process when searching for the ctH process. It is also interesting that the result is so different between the left-handed and the right-handed utH coupling. This is also due to the binning of the NNout distribution in TRExFitter. Since the automatic binning function was used, the bins could differ slightly, even with similar shapes of the two signal distributions. The boundary of the last bin was exactly on the falling edge of the NNout distribution due to the automatic binning function from TRExFitter, so small differences in the bin boundary caused a big difference in the number of events in the last bin. Since the highest bin has the best signal-to-background ratio, this bin has a high sensitivity to the signal strength, so the difference between the left-handed and right-handed results can be explained by this.

12 Conclusion and outlook

This thesis is about the search for FCNC processes in the top-quark-Higgs-boson sector. Signal and control regions are defined based on the number on leptons, b-jets and missing transverse momentum. The definition of the signal region is adapted to the signature of the FCNC processes. The control region is used to extract a scale factor for the normalization of the $t\bar{t}$+heavy-flavor jets, whose production rate is not well predicted.

Afterwards, the reconstruction of the objects is discussed, which is an important step towards the discrimination between signal and background. Two reconstruction techniques, using either a χ^2 algorithm or a likelihood approach called “KLFitter”, are tested. It turns out that the χ^2 algorithm is sufficient for the reconstruction since not many permutations have to be considered.

In order to take into account a further FCNC process, which has other kinematic properties but nevertheless has a significant part in the signal region, the cross section of this process has to be normalized. The Wilson coefficient in the Monte Carlo production of the “decay-FCNC process” could not be set equal to the “production-FCNC process” due to technical problems. This problem is solved by deriving the dependence of the cross section on the Wilson coefficient theoretically, confirming it using simulations and then extrapolating the cross section at the desired value for the Wilson coefficient.

A major part of the analysis is the use of artificial neural networks, which combine kinematic distributions of the FCNC processes and the backgrounds into a single variable that best separates signal and background. The kinematic differences in the processes, as well as the optimization of the neural networks, is discussed. Since the neural network has to simultaneously differentiate two kinematically different FCNC processes from the background, it is investigated to what extent a two-dimensional distribution of the neural network output variable can help. Trainings of the neural network are used which were only trained on the production-FCNC process or the decay-FCNC process. Correlations between the two trainings could improve the separation strength. However, it turns out that the NNout, which is based on both FCNC processes simultaneously, leads to a better expected upper exclusion limit, than the processed two-dimensional NNout distribution.

This search for FCNC signals uses a binned profile likelihood fit with systematic uncertainties for the statistical analysis. In the control region, data is used to normalize the $t\bar{t}$+heavy-flavor jets, whose production rate is not well predicted. In the signal region, an Asimov data set is used to obtain expected upper exclusion limits on the branching ratio of the FCNC processes. In addition, constraints on the systematic uncertainties are discussed and the systemic uncertainties with the highest influences are identified. The expected upper exclusion limits on the branching ratio for the left-handed processes are $\mathcal{B}(t \to uH) = 1.59 \times 10^{-3}$ and $\mathcal{B}(t \to cH) = 3.33 \times 10^{-3}$.

Possible improvements of the analysis strategy are to optimize the neural network for the ctH process. This could result in a better expected upper exclusion limit than the currently determined one. Also, the shower systematic uncertainty of $t\bar{t}$, which for technical reasons could not be implemented correctly but was assumed to be without shape, could be re-implemented. Finally, it could be investigated how the influence of the systematic uncertainties on the final result could be reduced since the expected upper exclusion limit without systematic uncertainties is about one order of magnitude smaller than the expected upper exclusion limit with systematic uncertainties.

13 Appendix

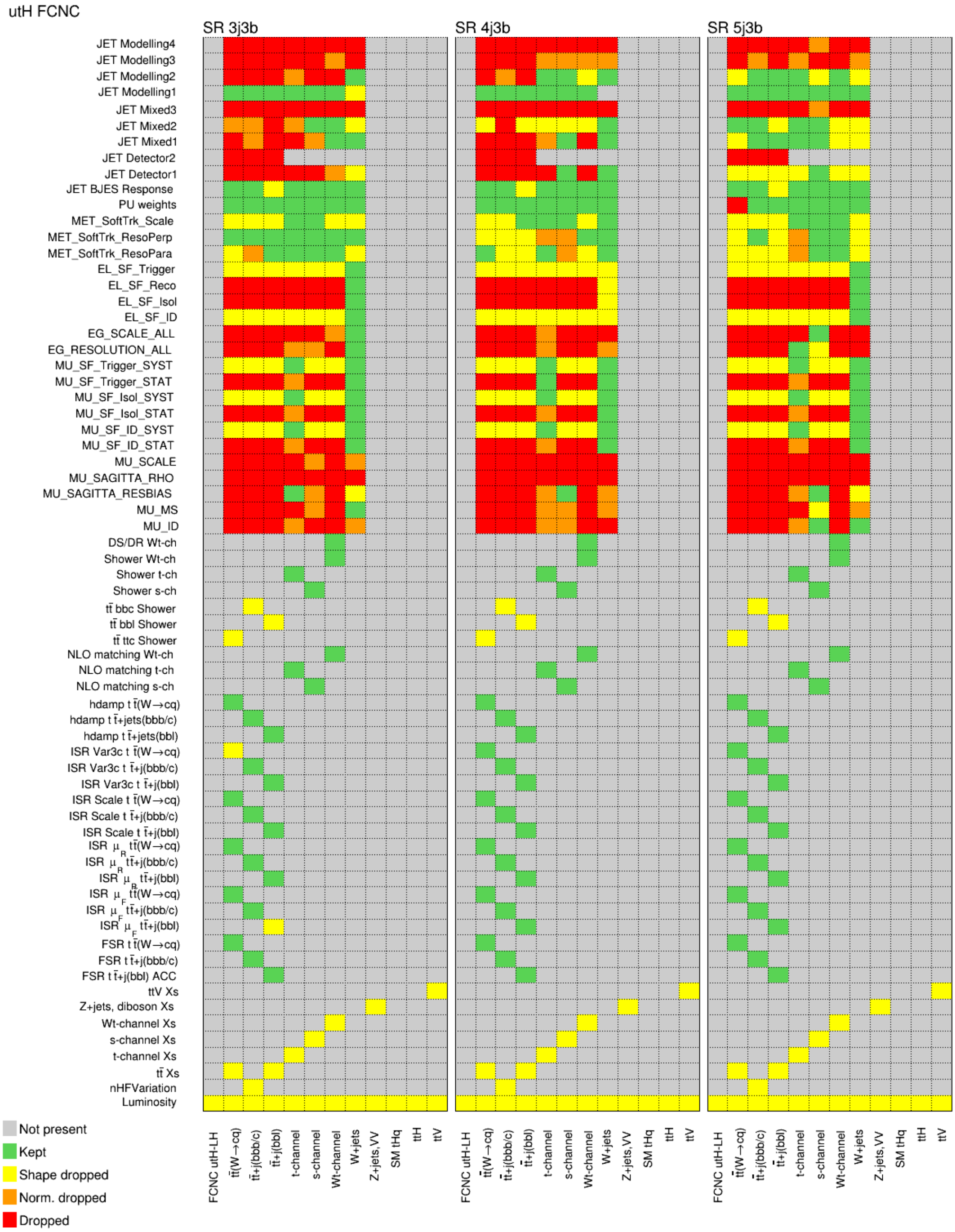


Figure 13.1: *Considered systematic jet uncertainties in the signal region with the left-handed utH-process as signal.*

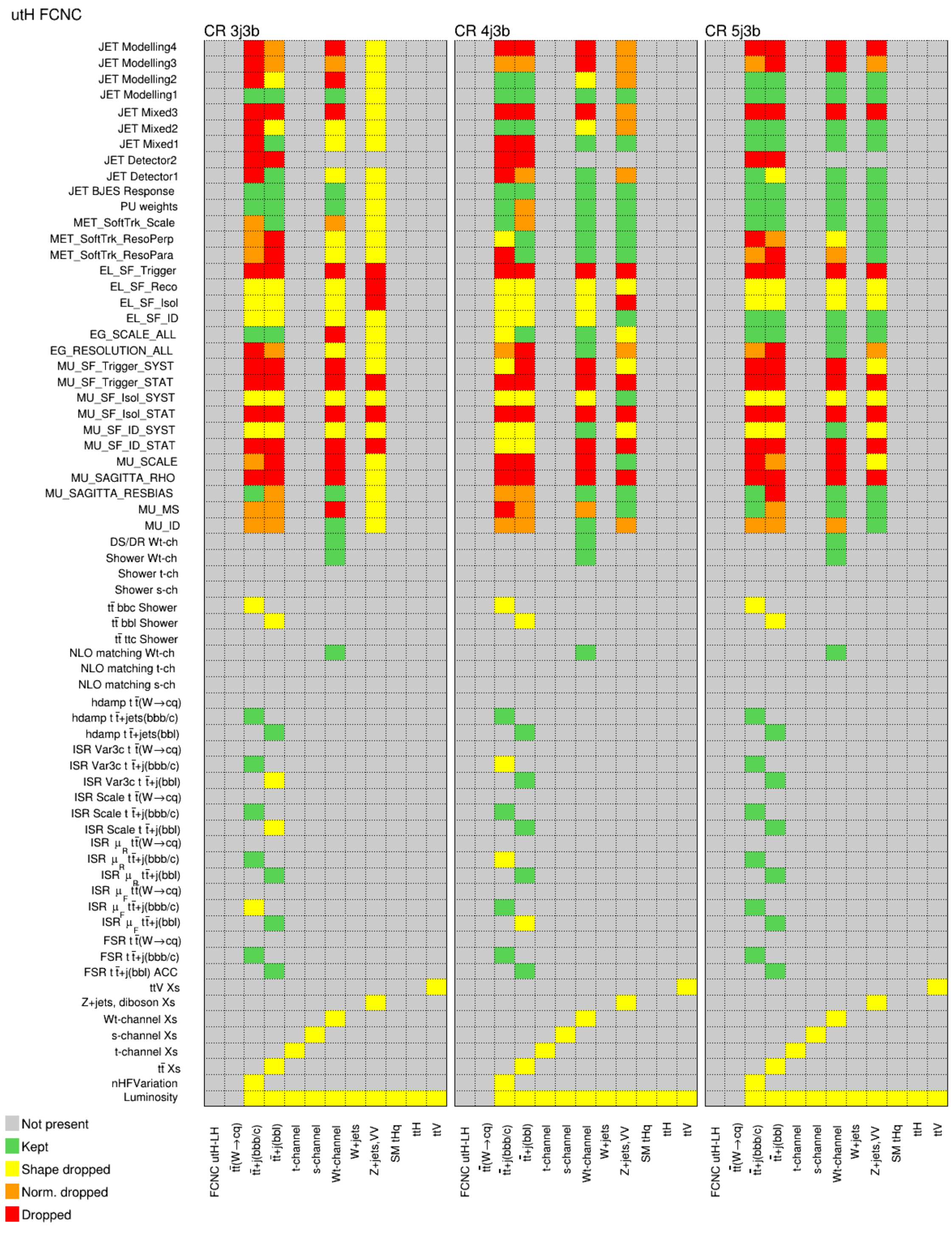


Figure 13.2: *Considered systematic jet uncertainties in the control region.*

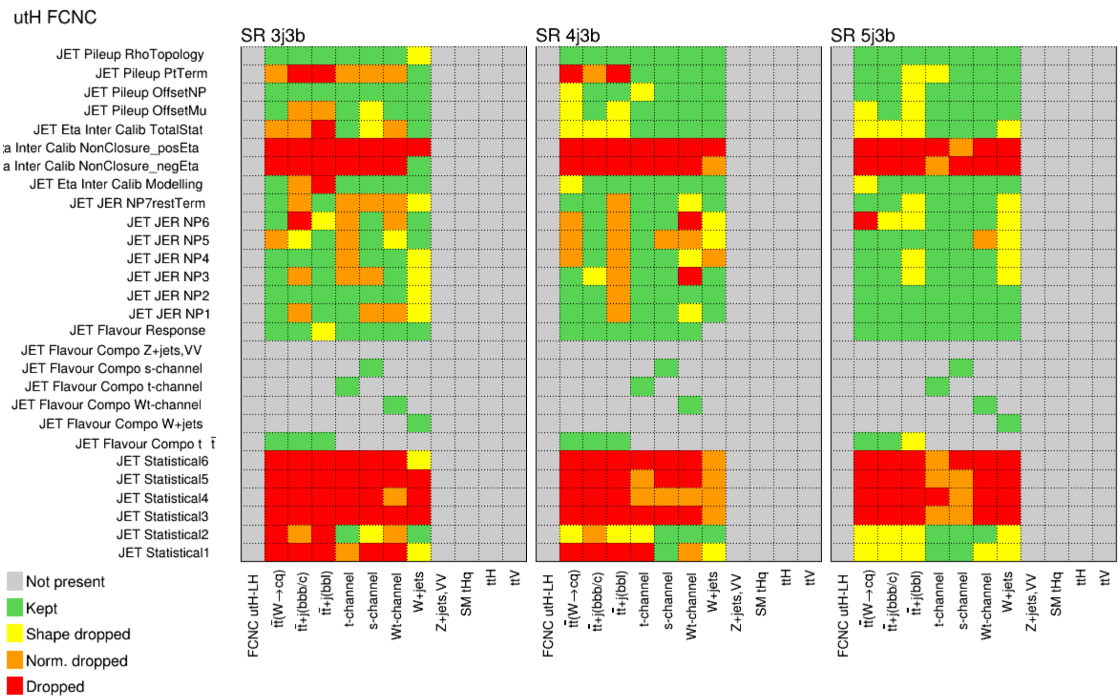


Figure 13.3: *Further considered systematic jet uncertainties in the signal region with the left-handed utH-process as signal.*

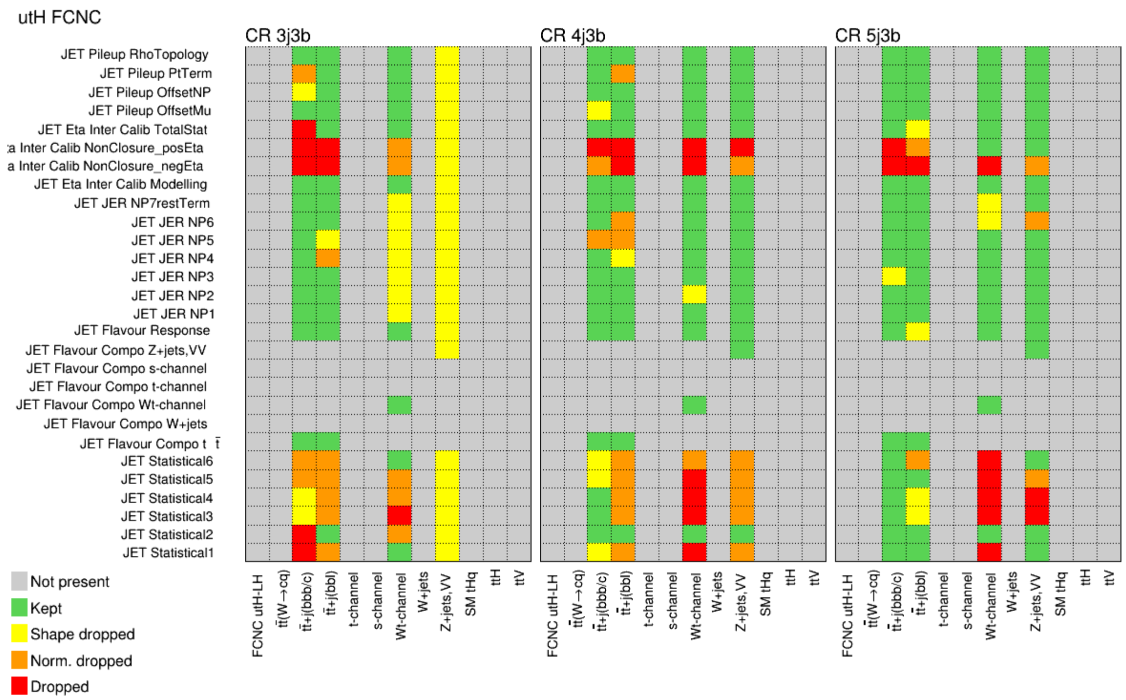


Figure 13.4: *Further considered systematic jet uncertainties in the control region.*

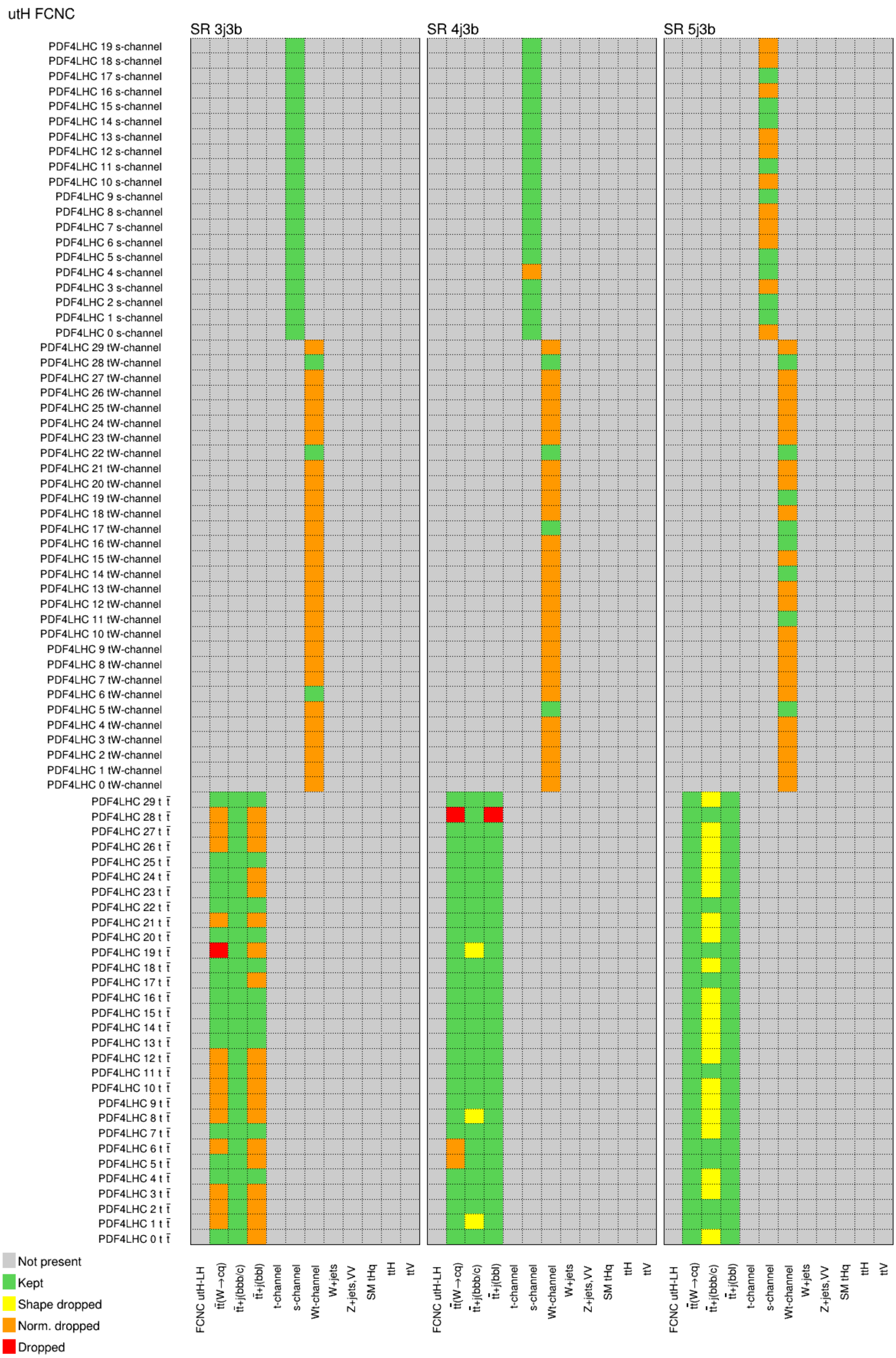


Figure 13.5: *Considered systematic modeling uncertainties in the signal region with the left-handed utH-process as signal.*

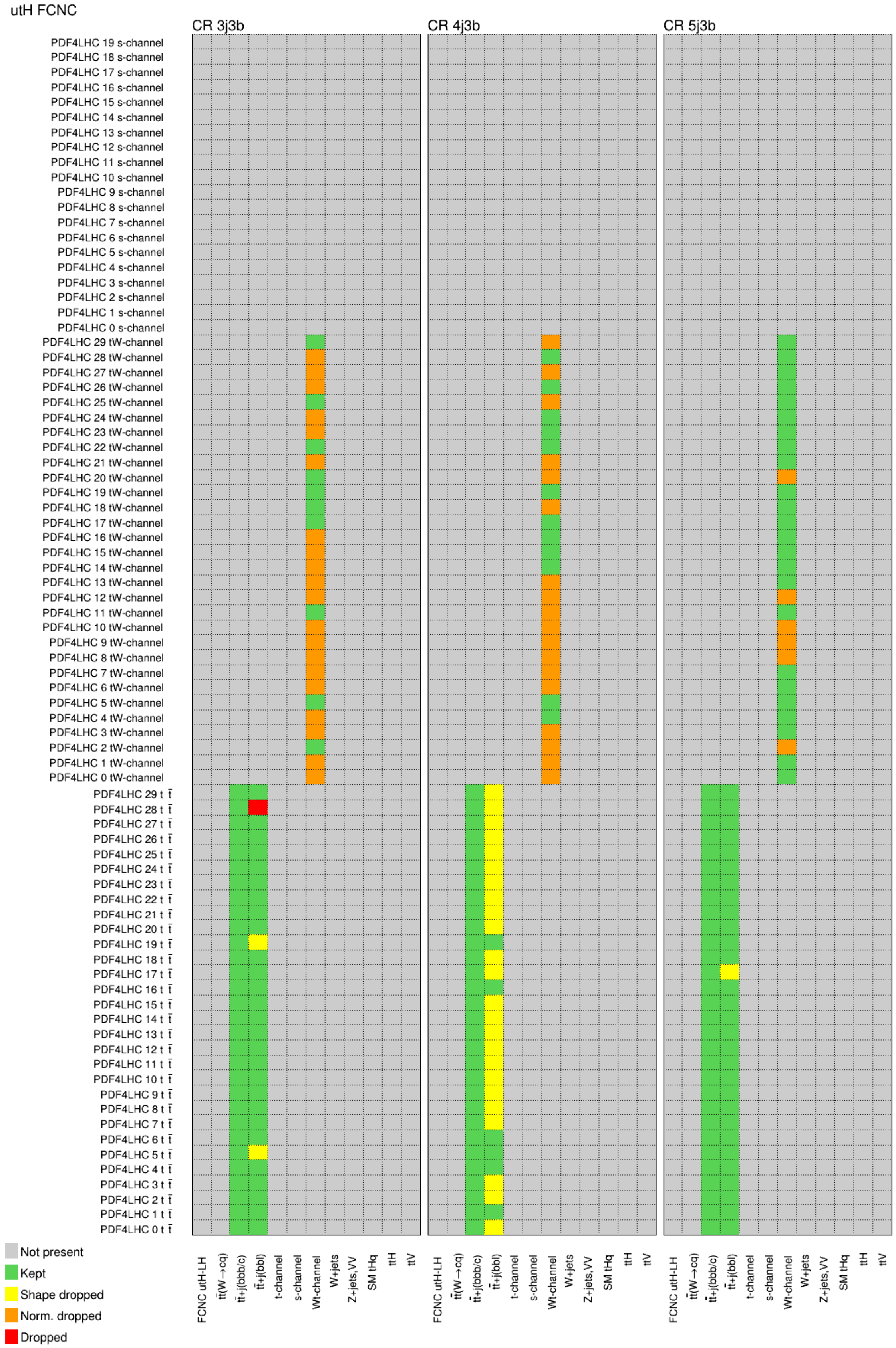


Figure 13.6: Considered systematic modeling uncertainties in the control region.

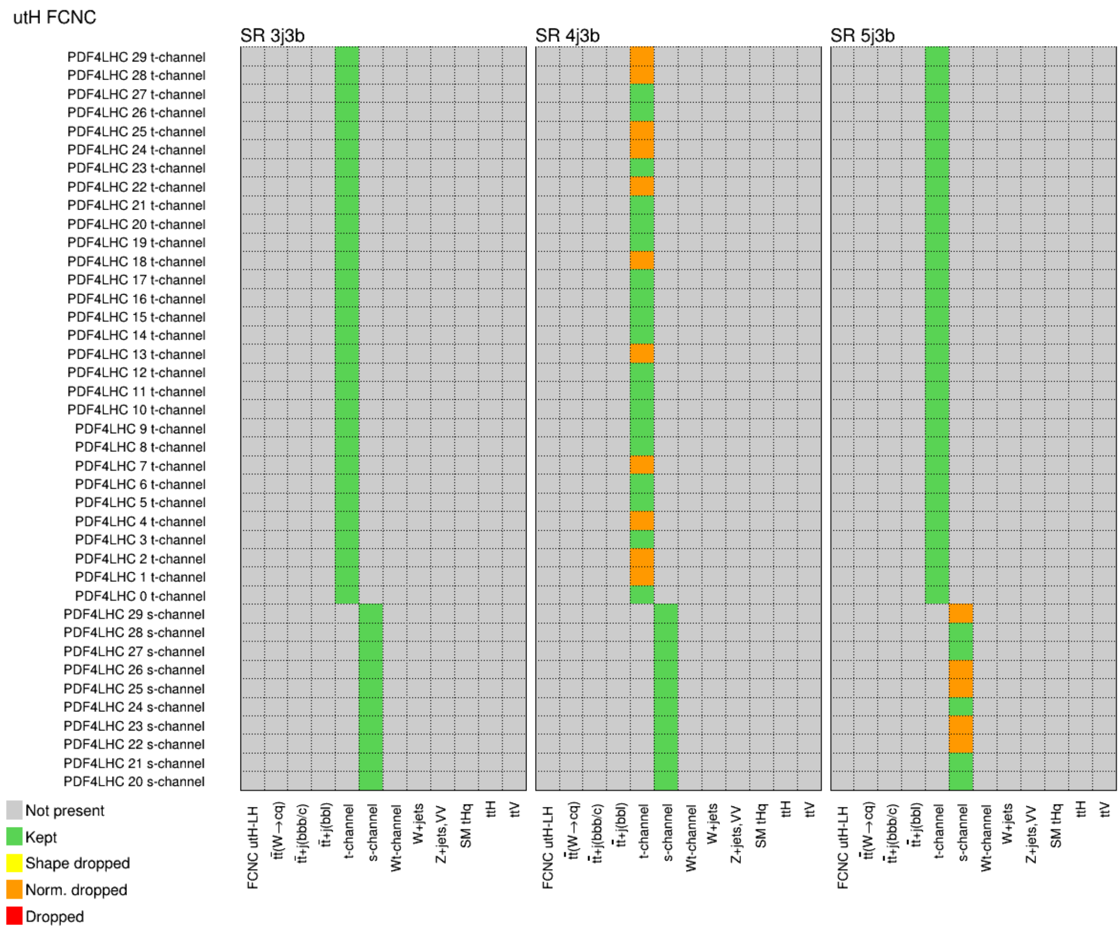


Figure 13.7: Further considered systematic modeling uncertainties in the signal region with the left-handed utH-process as signal.

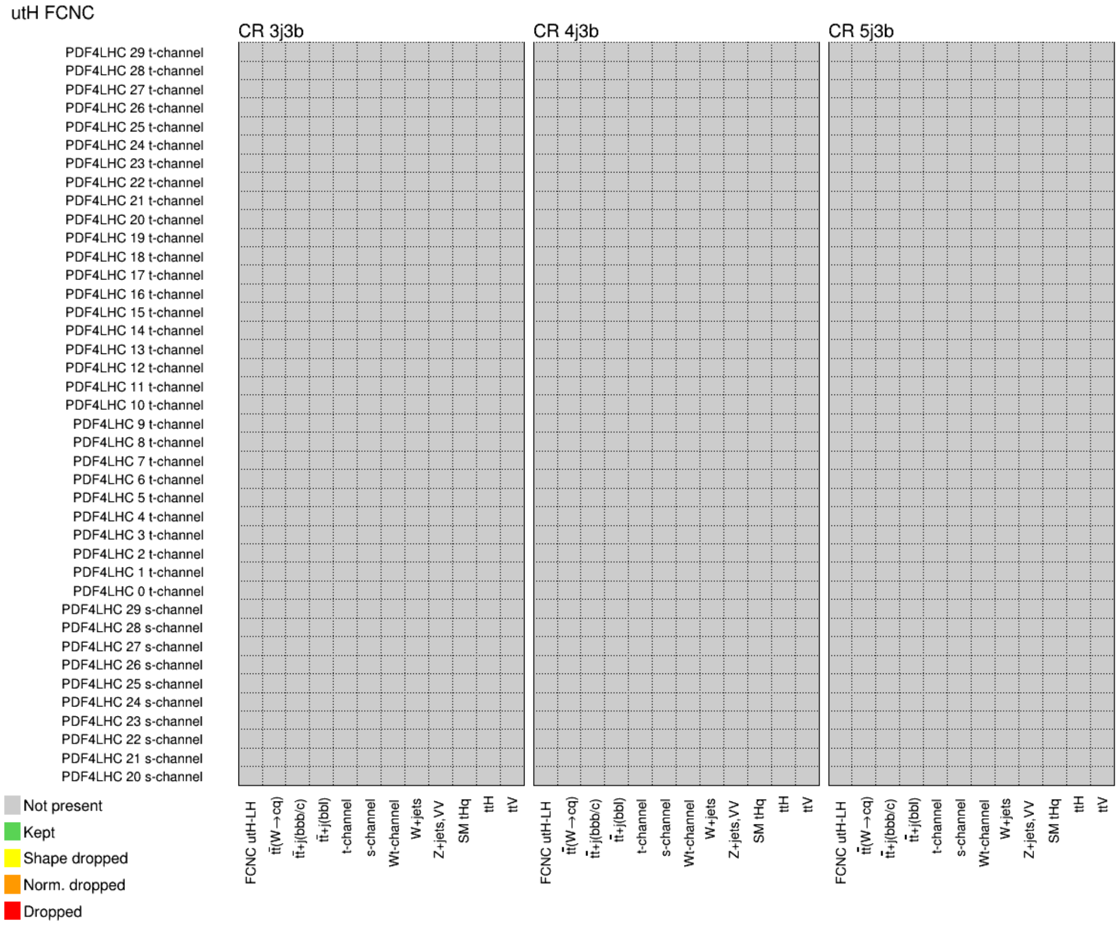


Figure 13.8: *Further considered systematic modeling uncertainties in the control region.*

Figure 13.9: *Considered systematic tagging uncertainties in the signal region with the left-handed utH-process as signal.*

Figure 13.10: *Considered systematic tagging uncertainties in the control region.*

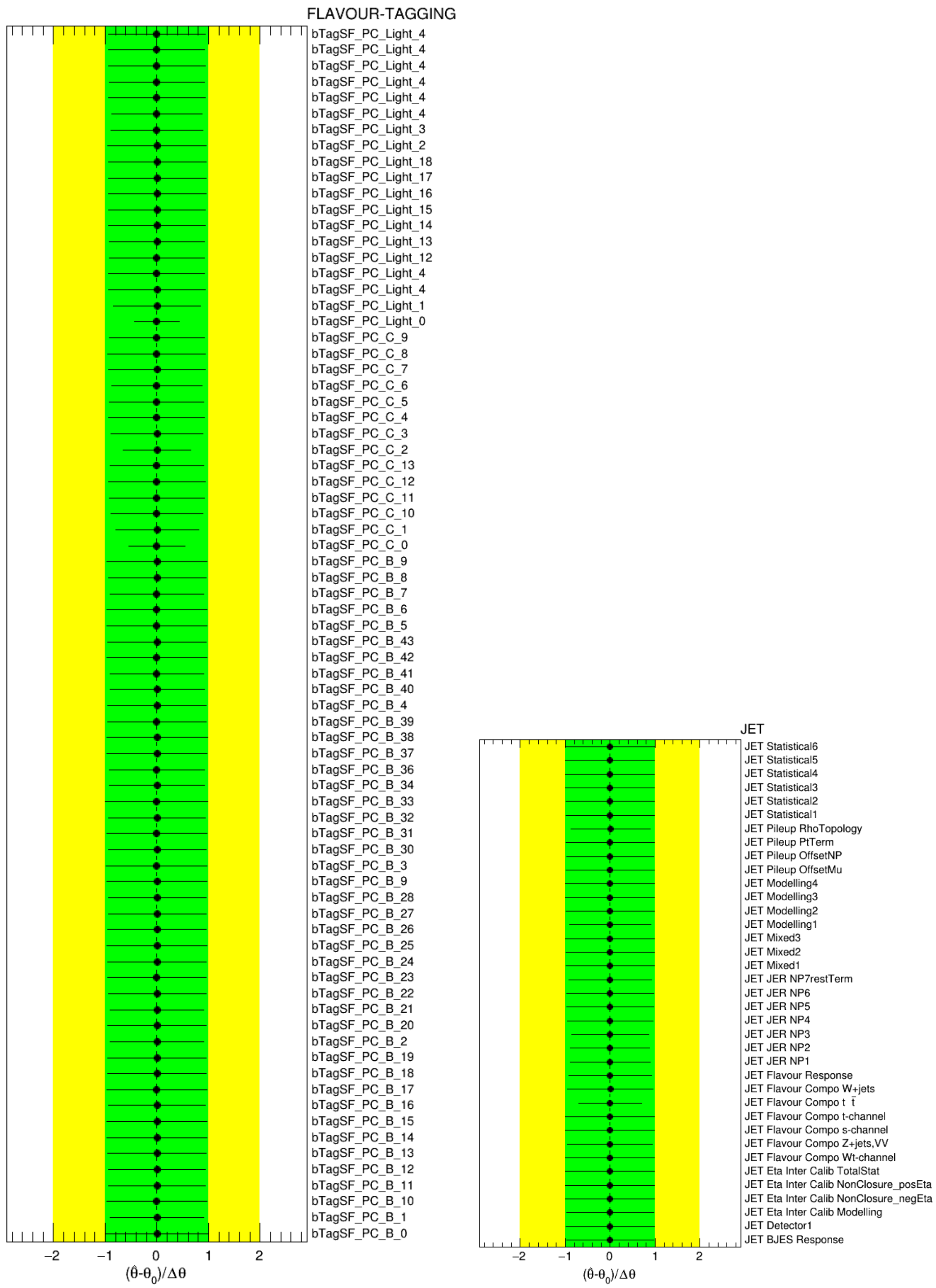


Figure 13.11: *Constraints of the nuisance parameters for the tagging- and JET uncertainties in the case of the left-handed utH process.*

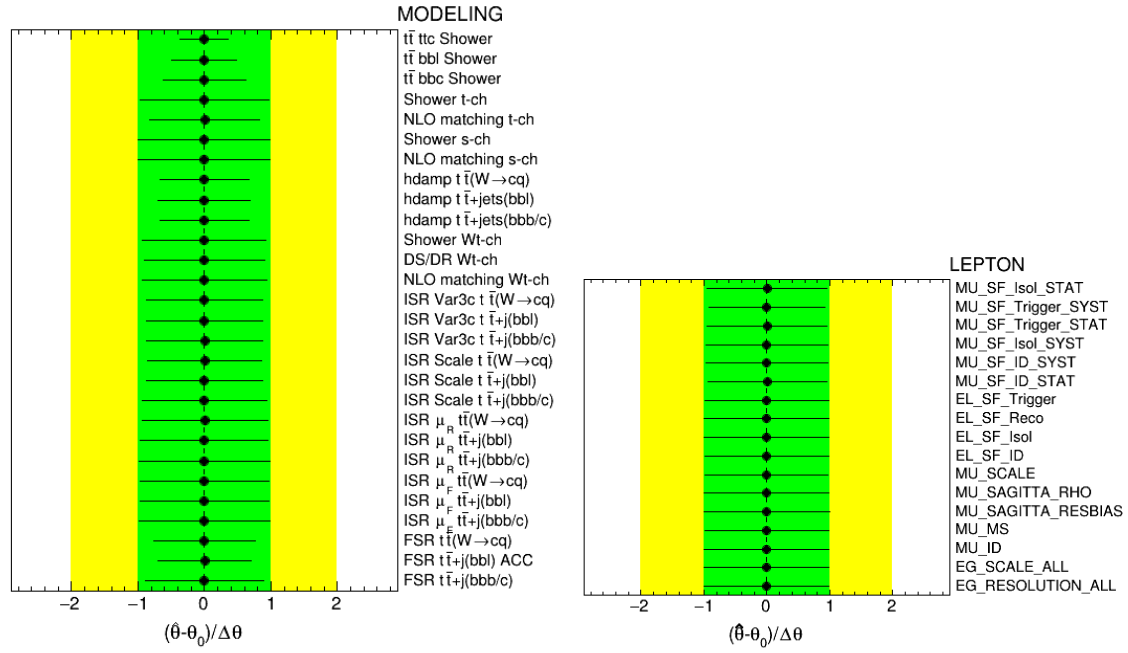


Figure 13.12: *Constraints of the nuisance parameters for the modeling- and lepton uncertainties in the case of the left-handed utH process.*

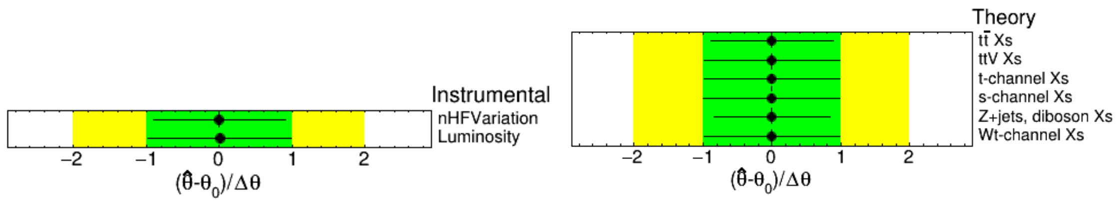


Figure 13.13: *Constraints of the nuisance parameters for the instrumental- and theory uncertainties in the case of the left-handed utH process.*

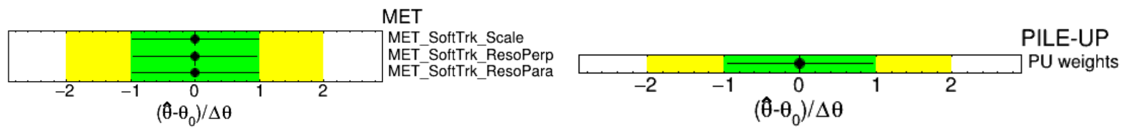


Figure 13.14: *Constraints of the nuisance parameters for the MET- and pileup uncertainties in the case of the left-handed utH process.*

Bibliography

[1] Summary of all known elemental particles in the Standard Model and their interactions with each other. . URL: https://en.wikipedia.org/wiki/Standard_Model#/media/File:Standard_Model_of_Elementary_Particles.svg.

[2] J. Beringer et al. Review of particle physics. *Phys. Rev. D*, 86:010001, Jul 2012. URL: https://link.aps.org/doi/10.1103/PhysRevD.86.010001, doi:10.1103/PhysRevD.86.010001.

[3] PDG review of the top quark 2018. . URL: http://pdg.lbl.gov/2019/reviews/rpp2018-rev-top-quark.pdf.

[4] S. L. Glashow, J. Iliopoulos, and L. Maiani. Weak interactions with lepton-hadron symmetry. *Phys. Rev. D*, 2:1285–1292, Oct 1970. URL: https://link.aps.org/doi/10.1103/PhysRevD.2.1285, doi:10.1103/PhysRevD.2.1285.

[5] Summary of the current 95ratios of top quark FCNC decays. . URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2019-038/fig_01.png.

[6] Adam Alloul, Neil D. Christensen, Céline Degrande, Claude Duhr, and Benjamin Fuks. Feynrules 2.0 - a complete toolbox for tree-level phenomenology. *Computer Physics Communications*, 185(8):2250 – 2300, 2014. URL: http://www.sciencedirect.com/science/article/pii/S0010465514001350, doi:https://doi.org/10.1016/j.cpc.2014.04.012.

[7] Céline Degrande, Claude Duhr, Benjamin Fuks, David Grellscheid, Olivier Mattelaer, and Thomas Reiter. Ufo – the universal feynrules output. *Computer Physics Communications*, 183(6):1201 – 1214, 2012. URL: http://www.sciencedirect.com/science/article/pii/S0010465512000379, doi:https://doi.org/10.1016/j.cpc.2012.01.022.

[8] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. MadGraph 5 : Going Beyond. *JHEP*, 06:128, 2011. arXiv:1106.0522, doi:10.1007/JHEP06(2011)128.

[9] Cen Zhang and Fabio Maltoni. Top-quark decay into Higgs boson and a light quark at next-to-leading order in QCD. *Phys. Rev.*, D88:054005, 2013. arXiv:1305.7386, doi:10.1103/PhysRevD.88.054005.

[10] The LHC along with its pre-accelerators. URL: https://en.wikipedia.org/wiki/Large_Hadron_Collider#/media/File:LHC.svg.

[11] The ATLAS Collaboration. The ATLAS experiment at the CERN large hadron collider. *Journal of Instrumentation*, 3(08):S08003–S08003, aug 2008. URL: https://doi.org/10.1088%2F1748-0221%2F3%2F08%2Fs08003, doi:10.1088/1748-0221/3/08/s08003.

[12] M Capeans, G Darbo, K Einsweiller, M Elsing, T Flick, M Garcia-Sciveres, C Gemme, H Pernegger, O Rohne, and R Vuillermet. ATLAS Insertable B-Layer Technical Design Report. Technical Report CERN-LHCC-2010-013. ATLAS-TDR-19, Sep 2010. URL: https://cds.cern.ch/record/1291633.

[13] Joao Pequenao. Event Cross Section in a computer generated image of the ATLAS detector. Mar 2008. URL: https://cds.cern.ch/record/1096081.

[14] M. Aaboud, G. Aad, B. Abbott, J. Abdallah, O. Abdinov, B. Abeloos, R. Aben, O. S. AbouZeid, N. L. Abraham, and et al. Performance of the atlas trigger system in 2015. *The European Physical Journal C*, 77(5), May 2017. URL: http://dx.doi.org/10.1140/epjc/s10052-017-4852-3, doi:10.1140/epjc/s10052-017-4852-3.

[15] S. Artz et al. Upgrade of the ATLAS central trigger for LHC run-2. *Journal of Instrumentation*, 10(02):C02030–C02030, feb 2015. URL: https://doi.org/10.1088%2F1748-0221%2F10%2F02%2Fc02030, doi:10.1088/1748-0221/10/02/c02030.

[16] Stefano Frixione, Giovanni Ridolfi, and Paolo Nason. A positive-weight next-to-leading-order monte carlo for heavy flavour hadroproduction. *Journal of High Energy Physics*, 2007(09):126–126, sep 2007. URL: https://doi.org/10.1088%2F1126-6708%2F2007%2F09%2F126, doi:10.1088/1126-6708/2007/09/126.

[17] Paolo Nason. A new method for combining NLO QCD with shower monte carlo algorithms. *Journal of High Energy Physics*, 2004(11):040–040, nov 2004. URL: https://doi.org/10.1088%2F1126-6708%2F2004%2F11%2F040, doi:10.1088/1126-6708/2004/11/040.

[18] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with parton shower simulations: the POWHEG method. *Journal of High Energy Physics*, 2007(11):070–070, nov 2007. URL: https://doi.org/10.1088%2F1126-6708%2F2007%2F11%2F070, doi:10.1088/1126-6708/2007/11/070.

[19] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing nlo calculations in shower monte carlo programs: the powheg box. *Journal of High Energy Physics*, 2010(6):43, Jun 2010. doi:10.1007/JHEP06(2010)043.

[20] T Gleisberg, S Höche, F Krauss, M Schönherr, S Schumann, F Siegert, and J Winter. Event generation with SHERPA 1.1. *Journal of High Energy Physics*, 2009(02):007–007, feb 2009. URL: https://doi.org/10.1088%2F1126-6708%2F2009%2F02%2F007, doi:10.1088/1126-6708/2009/02/007.

[21] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to pythia 8.2. *Computer Physics Communications*, 191:159–177, Jun 2015. URL: http://dx.doi.org/10.1016/j.cpc.2015.01.024, doi:10.1016/j.cpc.2015.01.024.

[22] S. Agostinelli et al. Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250 – 303, 2003. URL: http://www.sciencedirect.com/science/article/pii/S0168900203013688, doi:https://doi.org/10.1016/S0168-9002(03)01368-8.

[23] Stefano Frixione, Eric Laenen, Patrick Motylinski, and Bryan R Webber. Angular correlations of lepton pairs from vector boson and top quark decays in monte carlo simulations. *Journal of High Energy Physics*, 2007(04):081–081, apr 2007. URL: https://doi.org/10.1088%2F1126-6708%2F2007%2F04%2F081, doi:10.1088/1126-6708/2007/04/081.

[24] Pierre Artoisenet, Rikkert Frederix, Olivier Mattelaer, and Robbert Rietkerk. Automatic spin-entangled decays of heavy resonances in monte carlo simulations. *Journal of High Energy Physics*, 2013(3):15, Mar 2013. doi:10.1007/JHEP03(2013)015.

[25] Inclusive ttbar cross-section, normalized to the NNLO prediction in QCD including NNLL soft-gluon terms. . URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO.

[26] Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton-proton collision data. Technical Report ATLAS-CONF-2014-032, CERN, Geneva, Jun 2014. URL: https://cds.cern.ch/record/1706245.

[27] Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data. Technical Report ATLAS-CONF-2016-024, CERN, Geneva, Jun 2016. URL: https://cds.cern.ch/record/2157687.

[28] Electron and photon energy calibration with the ATLAS detector using data collected in 2015 at $\sqrt{s} = 13$ TeV. Technical Report ATL-PHYS-PUB-2016-015, CERN, Geneva, Aug 2016. URL: https://cds.cern.ch/record/2203514.

[29] Georges Aad et al. Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$ =13 TeV. *Eur. Phys. J.*, C76(5):292, 2016. arXiv:1603.05598, doi:10.1140/epjc/s10052-016-4120-y.

[30] Georges Aad et al. Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1. *Eur. Phys. J.*, C77:490, 2017. arXiv:1603.02934, doi:10.1140/epjc/s10052-017-5004-5.

[31] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008. arXiv:0802.1189, doi:10.1088/1126-6708/2008/04/063.

[32] Recommendations for the Pileup jet treatment by the ATLAS collaboration. . URL: https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/PileupJetRecommendations.

[33] M. Aaboud et al. Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the atlas detector. *Phys. Rev. D*, 96:072002, Oct 2017. URL: https://link.aps.org/doi/10.1103/PhysRevD.96.072002, doi:10.1103/PhysRevD.96.072002.

[34] Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run. Technical Report ATL-PHYS-PUB-2017-013, CERN, Geneva, Jul 2017. URL: https://cds.cern.ch/record/2273281.

[35] Morad Aaboud et al. Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J.*, C78(11):903, 2018. arXiv:1802.08168, doi:10.1140/epjc/s10052-018-6288-9.

[36] $E_{\mathrm{T}}^{\mathrm{miss}}$ performance in the ATLAS detector using 2015-2016 LHC p-p collisions. Technical Report ATLAS-CONF-2018-023, CERN, Geneva, Jun 2018. URL: https://cds.cern.ch/record/2625233.

[37] Recommendations from the ATLAS collaboration for the overlap removal. URL: https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/TopCommonObjects13TeV#Object_overlap_removal.

[38] Probabilities for the decay of the Higgs boson into the specific channels. . URL: http://opendata.atlas.cern/books/current/get-started/_book/the-higgs-boson.html.

[39] Oliver Thielmann. Sensitivitätsstudien zur Suche nach Top-Quark-Higgs-Boson-Produktion durch flavour-ändernde neutrale Ströme am Large Hadron Collider, Aug 2017. Presented 27 Sep 2017. URL: https://cds.cern.ch/record/2299371.

[40] Phillip Sturm. W Boson Reconstruction. URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/NeutrinoPz?rev=3.

[41] Johannes Erdmann, Stefan Guindon, Kevin Kroeninger, Boris Lemmer, Olaf Nackenhorst, Arnulf Quadt, and Philipp Stolte. A likelihood-based reconstruction algorithm for top-quark pairs and the KLFitter framework. *Nucl. Instrum. Meth.*, A748:18–25, 2014. arXiv:1312.5595, doi:10.1016/j.nima.2014.02.029.

[42] Simplified scheme of a artificial neural network. . URL: https://de.wikipedia.org/wiki/K%C3%BCnstliches_neuronales_Netz.

[43] M. Feindt and U. Kerzel. The neurobayes neural network package. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 559(1):190 – 194, 2006. Proceedings of the X International Workshop on Advanced Computing and Analysis Techniques in Physics Research. URL: http://www.sciencedirect.com/science/article/pii/S0168900205022679, doi:https://doi.org/10.1016/j.nima.2005.11.166.

[44] James D. Bjorken and Stanley J. Brodsky. Statistical model for electron-positron annihilation into hadrons. *Phys. Rev. D*, 1:1416–1420, Mar 1970. URL: https://link.aps.org/doi/10.1103/PhysRevD.1.1416, doi:10.1103/PhysRevD.1.1416.

[45] Michael Feindt. A Neural Bayesian Estimator for Conditional Probability Densities. *arXiv e-prints*, page physics/0402093, Feb 2004. arXiv:physics/0402093.

[46] Morad Aaboud et al. Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC. *Eur. Phys. J.*, C76(12):653, 2016. arXiv:1608.03953, doi:10.1140/epjc/s10052-016-4466-1.

[47] Uncertainty of the luminosity, indicated by the ATLAS collaboration. URL: https://twiki.cern.ch/twiki/bin/viewauth/Atlas/LuminosityForPhysics#Proton_proton_data.

[48] Extended pile-up reweighting guide by the ATLAS collaboration. . URL: https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/ExtendedPileupReweighting.

[49] Louis Fayard, Christophe Raymond Goudet, Thibault Guillemin, Lydia Iconomidou-Fayard, Stefano Manzoni, Kerstin Tackmann, Ruggero Turra, Paolo Mastrandrea, and Guillaume Unal. Energy calibration prerecommendation for Run 2. Technical Report ATL-COM-PHYS-2015-1300, CERN, Geneva, Oct 2015. URL: https://cds.cern.ch/record/2060328.

[50] Guideline for the muon systematic uncertainties by the ATLAS collaboration. . URL: https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/MCPAnalysisGuidelinesMC15.

[51] Uncertainty of the jets, indicated by the ATLAS collaboration. URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/ApplyJetCalibrationR21.

[52] ATLAS Collaboration. Atlas b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ tev, 2019. arXiv:1907.05120.

[53] Calculation of the systematic uncertainties of the missing transverse momentum. URL: https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/MissingETSystematics.

[54] NNLO+NNLL top-quark-pair cross sections. URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO.

[55] Recommended uncertainties for single top cross sections. . URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SingleTopRefXsec.

[56] Jon Butterworth et al. PDF4LHC recommendations for LHC Run II. *J. Phys.*, G43:023001, 2016. arXiv:1510.03865, doi:10.1088/0954-3899/43/2/023001.

[57] Default extrapolation functions between systematic uncertainty templates and the nominal template in the TRExFitter framework. URL: https://indico.cern.ch/event/700646/contributions/2936548/attachments/1627872/2595678/TRExFitterTutorial.pdf.

[58] A L Read. Presentation of search results: the CL_s technique. *J. Phys. G*, 28(10):2693–704, 2002. URL: https://cds.cern.ch/record/722145, doi:10.1088/0954-3899/28/10/313.

[59] Thomas Junk. Confidence level computation for combining searches with small statistics. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 434(2):435 – 443, 1999. URL: http://www.sciencedirect.com/science/article/pii/S0168900299004982, doi:https://doi.org/10.1016/S0168-9002(99)00498-2.

[60] Twiki page of the TRExFitter binned profiled likelihood fit framework. . URL: https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/TtHFitter.

[61] Morad Aaboud et al. Search for top-quark decays $t \to Hq$ with 36 fb^{-1} of pp collision data at $\sqrt{s} = 13$ TeV with the ATLAS detector. *JHEP*, 05:123, 2019. arXiv:1812.11568, doi:10.1007/JHEP05(2019)123.

Name, Vorname: ....

Erklärung

gem. § 14 Abs. 7 Prüfungsordnung vom 15.04.2013

Hiermit erkläre ich, dass ich die von mir eingereichte Abschlussarbeit (Master-Thesis) selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie Stellen der Abschlussarbeit, die anderen Werken dem Wortlaut oder Sinn nach entnommen wurden, in jedem Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

....

Datum Unterschrift

Erklärung

Hiermit erkläre ich mich damit einverstanden, dass meine Abschlussarbeit (Master-Thesis) wissenschaftlich interessierten Personen oder Institutionen und im Rahmen von externen Qualitätssicherungsmaßnahmen des Studienganges zur Einsichtnahme zur Verfügung gestellt werden kann.
Korrektur- oder Bewertungshinweise in meiner Arbeit dürfen nicht zitiert werden.

....

Datum Unterschrift