
**Search for the Standard Model Higgs boson produced in
association with $t\bar{t}$ and decaying into $b\bar{b}$ at $\sqrt{s} = 8$ TeV with
the ATLAS detector using the Matrix Element Method**

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

"Doctor rerum naturalium"

der Georg-August-Universität Göttingen

im Promotionsprogramm ProPhys

der Georg-August University School of Science (GAUSS)

vorgelegt von

Olaf Nackenhorst

aus Hamburg

Göttingen, 2015

Mitglieder des Betreuungsausschusses:

Prof. Dr. Arnulf Quadt

II. Physikalisches Institut, Georg-August-Universität Göttingen

Dr. Elizaveta Shabalina

II. Physikalisches Institut, Georg-August-Universität Göttingen

Prof. Dr. Kevin Kröninger

Experimentelle Physik IV, Technische Universität Dortmund

Mitglieder der Prüfungskommission:

Referent: Prof. Dr. Arnulf Quadt

II. Physikalisches Institut, Georg-August-Universität Göttingen

Koreferentin: Prof. Dr. Ariane Frey

II. Physikalisches Institut, Georg-August-Universität Göttingen

Weitere Mitglieder der Prüfungskommission:

PD Dr. Jörn Große-Knetter

II. Physikalisches Institut, Georg-August-Universität Göttingen

Prof. Dr. Kevin Kröninger

Experimentelle Physik IV, Technische Universität Dortmund

Prof. Dr. Wolfram Kollatschny

Institut für Astrophysik, Georg-August-Universität Göttingen

Jun.-Prof. Dr. Steffen Schumann

II. Physikalisches Institut, Georg-August-Universität Göttingen

Tag der mündlichen Prüfung: 8. Juni 2015

Referenznummer: II.Physik-UniGö-Diss-2015/01

IN MEMORY OF MY GRANDFATHER

RUDOLF JACOBS

(10.10.1913 - 01.05.2015)

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

II. Physikalisches Institut

Search for the Standard Model Higgs boson produced in association with $t\bar{t}$ and decaying into $b\bar{b}$ at $\sqrt{s} = 8$ TeV with the ATLAS detector using the Matrix Element Method

by

Olaf Nackenhorst

A search for the Standard Model Higgs boson produced in association with a pair of top quarks ($t\bar{t}H$) is presented. The analysis uses 20.3 fb^{-1} of pp collision data at $\sqrt{s} = 8$ TeV, collected with the ATLAS detector at the Large Hadron Collider during 2012. The search is designed for the $H \rightarrow b\bar{b}$ decay mode and is performed in the single lepton (electrons or muons) decay channel of the top quark pair. In order to improve the sensitivity of the search, events are categorised according to their jet and b -tagged jet multiplicities into nine different analysis regions. A matrix element method is developed and applied to regions with six jets to obtain discriminants separating $t\bar{t}H$ events from the irreducible $t\bar{t} + b\bar{b}$ background. In signal-enriched regions, a neural network is employed combining kinematic variables and variables obtained from the matrix element method to maximise the separation between signal and background events. The nine analysis regions are statistically combined using a profile likelihood fit to improve the background predictions and reduce the systematic uncertainties. The ratio of the measured $t\bar{t}H$ signal cross section to the Standard Model expectation is found to be

$$\mu = 1.2 \pm 1.3$$

assuming a Higgs boson mass of 125 GeV. No significant excess of events above the background expectation is found and an observed (expected) limit of

$$\mu < 3.6 \text{ (2.6)}$$

is set at 95% confidence level. The single lepton channel is combined with the dilepton channel of the $t\bar{t}$ decay resulting in $\mu = 1.5 \pm 1.1$ and an observed (expected) upper limit of 3.4 (2.2) times the Standard Model cross section at 95% confidence level.

Post address:
Friedrich-Hund-Platz 1
37077 Göttingen
Germany

II. Physikalisches Institut
Georg-August-Universität Göttingen
May 2015

Contents

1. Preface	1
2. The Higgs Boson and Top Quarks in the Context of the Standard Model	3
2.1. Introduction	3
2.2. A Brief Summary of the Standard Model	3
2.2.1. Mathematical Description	5
2.2.2. The Higgs Mechanism	6
2.2.3. Beyond the SM	9
2.3. The Top Quark	10
2.3.1. Top Quark Production	10
2.3.2. Top Quark Decay	11
2.4. The Higgs Boson	12
2.4.1. Higgs Boson Production	12
2.4.2. Higgs Boson Decay	14
2.4.3. Higgs Boson Measurements	16
3. The ATLAS Experiment at the LHC	19
3.1. Introduction	19
3.2. The Large Hadron Collider	19
3.3. The ATLAS Detector	21
3.3.1. Magnet System	22
3.3.2. Inner Detector	22
3.3.3. Calorimeter	24
3.3.4. Muon Spectrometer	25
3.3.5. Trigger and Data Acquisition	26
4. Experimental Data, Physics Objects and Processes	29
4.1. Introduction	29
4.2. Description of the Experimental Data	29
4.2.1. Commissioning and Performance of the LHC	30

4.3.	Object Reconstruction	32
4.3.1.	Jets	33
4.3.2.	Electrons	37
4.3.3.	Muons	38
4.3.4.	Missing Transverse Energy	39
4.4.	Modelling of the Physics Processes	39
4.4.1.	Event Simulation	40
4.4.2.	Modelling of the Signal	41
4.4.3.	Monte Carlo Simulated Background	41
4.4.4.	Tag-Rate-Function Method	47
4.4.5.	Data-Driven Estimated Background	50
5.	The Matrix Element Method	59
5.1.	Introduction	59
5.2.	Method Description	59
5.2.1.	Process Probability	59
5.2.2.	Fermi's Golden Rule	61
5.2.3.	Assignment Permutations	63
5.2.4.	Event Probability and Parameter Estimation	63
5.3.	Implementation: MEMTool	64
5.4.	Parton Distribution Functions	66
5.5.	The Matrix Elements	68
5.5.1.	Signal Hypothesis	69
5.5.2.	Background Hypothesis	71
5.5.3.	Calculation Optimisations	73
5.6.	Transfer Functions	74
5.7.	Integration over the Phase Space	80
5.7.1.	Integration Algorithm	80
5.7.2.	Reduction of Assignment Permutations	82
5.7.3.	Phase Space Alignment	82
5.7.4.	Reduction of Dimensionality	84
5.8.	Computational Challenges	86
5.8.1.	Grid Evaluation System	87
5.8.2.	MEDUSA - Grid Submission Framework	87
5.8.3.	Automated Grid Monitoring	88
5.9.	Proof of Principle	90
5.9.1.	Parton Level Events	90
5.9.2.	Single Gaussian Smeared Parton Level Events	94
5.9.3.	Double Gaussian Smeared Parton Level Events	96
6.	Search for $t\bar{t}(H \rightarrow b\bar{b})$ Production in the Single Lepton Channel	99
6.1.	Introduction	99
6.2.	Event Selection and Classification	100
6.2.1.	Event Selection	100
6.2.2.	Classification	101
6.3.	Analysis Strategy	104

6.4. Systematic Uncertainties	105
6.4.1. Luminosity	107
6.4.2. Uncertainties on Physics Objects	107
6.4.3. Uncertainties on Background Modelling	110
6.4.4. Uncertainties on Signal Modelling	113
6.5. Using the Matrix Element Method	113
6.5.1. Likelihood Functions	114
6.5.2. Likelihood Ratio	119
6.5.3. Event Reconstruction	123
6.6. The Artificial Neural Network	134
6.6.1. The NeuroBayes Package	134
6.6.2. Discriminating Variables	136
6.6.3. The Neural Network Output	144
6.7. The Profile Likelihood Fit	146
6.7.1. The Likelihood Function	146
6.7.2. Extracted Signal Strength	147
6.7.3. Improvements of the Predictions	147
6.7.4. Constraining the Systematic Uncertainties	154
6.7.5. Validation of MEM Variables	157
6.8. Combined Results	159
6.8.1. Combination with Dilepton Analysis	159
6.8.2. The Signal Strength	160
6.8.3. Setting Limits	161
6.8.4. Impact of the MEM on the Results	164
7. Conclusions and Outlook	167
7.1. Summary and Conclusion	167
7.2. Outlook	170
Bibliography	173
Appendices	189
A. Further Separation Studies using the MEM	191
A.1. Likelihood Ratio	192
A.2. Logarithm of Summed Signal Likelihood	194
A.3. Invariant Mass of Additional B-jets	196
A.4. Hadronic Top Quark Mass	200
A.5. Hadronic W Boson Mass	202
B. Further Event Reconstruction using the MEM	205
B.1. Angular Distributions of Additional B-jets	205
B.2. Reconstruction Efficiencies	209
C. Neural Networks with MEM Variables	211
C.1. Saturation of the NN	211
C.2. Alternative NN Strategies	211
C.3. Correlation among NN Variables	213

C.4. Overtraining Test of NN Output	213
D. Additional Material from the Profile Likelihood Fit	215
D.1. Nuisance Parameter	215
D.1.1. Pull and Constraints	215
D.1.2. Correlations	216
D.2. Excess in Muon Events	217
D.3. Validation of MEM Variables	220
D.4. Validation of NN Variables	223
List of Abbreviations	225
List of Figures	229
List of Tables	233
Index	235
Acknowledgements	236

1

Preface

"For me it's really an incredible thing that it's happened in my lifetime" were the words by Peter Higgs on the 4th of July 2012 after the Large Hadron Collider (LHC) experiments announced the discovery of a new particle, what we now believe is the Higgs boson predicted by Brout, Englert, Higgs and others about 50 years ago [1–4]. It is astonishing, but also proves the success of science, that a theory proposed long time ago, when the discovery seemed completely unrealistic, could finally be verified due to the great development of technology. I had the pleasure to meet Peter Higgs shortly after the discovery in August 2012 at the Scottish Universities Summer School in Physics in St. Andrews. He gave a talk for the students, in which he briefly summarised the historical development of the theory he and others had formulated. At the time they developed the theory, the picture we had of our world and the Universe was very different. Peter Higgs had also been an organiser of this summer school for a long time and thanks to him the students can enjoy a nice selection of wines during the dinners every evening. During one of these dinners, I had a chat with him and other students mostly about topics unrelated to physics. Suddenly, he was very concerned about how well he presented the topic of the talk and told us that he was nervous giving it. One year later he and Englert received the Nobel-Prize for the formulation of the mechanism, which lead to the discovery of the Higgs boson. I can only imagine how nervous he might have been during the Nobel-Prize ceremony, but the fact that he still felt insecure when giving a talk in front of students made him very human. It made us realise that we are just like him, that everyone of us can come up with a beautiful idea that might change the picture of the world. This thesis is a first attempt in that direction although admittedly a small step. However, many times in the process of working on this project a success seemed unrealistic and pioneer work had to be performed in many directions.

Although we are quite confident that we have discovered a Higgs boson, we do not know yet if the observed particle is a part of the well established Standard Model (SM) of particle physics or of a more general theory in which the SM is embedded. There are many open questions arising from observations in particle physics and astrophysics which do not have answers within the SM. Thus it is clear, that our current picture of the world is incomplete. To determine the true nature of the observed Higgs boson, it is important to study all of its properties, which include the production mechanisms and decay modes. One of the production mechanism of the Higgs boson predicted by the SM is the production in association with a top-antitop quark pair ($t\bar{t}H$), which has not yet been observed because of its small cross section. The

observation of this process would allow for a direct measurement of the top quark Yukawa coupling, which is an important parameter of the SM. Furthermore, the Higgs boson has only been discovered in bosonic decays but not in decays to fermions. For a Higgs boson with a mass around 125 GeV as discovered at the LHC, the dominant decay mode is the fermionic $H \rightarrow b\bar{b}$ decay. However, the overwhelming multijet background in the dominant production channels made it impossible to observe it so far. In the $t\bar{t}H$ production, the distinct signature of the top quark pair decay can be used to suppress the multijet background enabling a search for the $H \rightarrow b\bar{b}$ decay. Though it becomes very challenging to distinguish the signal from the irreducible and huge background arising from top quark pair production with additional b -quarks in the final state ($t\bar{t} + b\bar{b}$). In order to separate the small signal from a much larger background, sophisticated analysis techniques have to be applied. For this search the *matrix element method* has been deployed, which makes use of the theoretical description of a process in order to assign a probability weight to each observed event. This probability weight reflects how likely it is that the observed event is consistent with a certain hypothesis. This analysis technique is quite unique in particle physics, because it uses maximum available information from the theoretical model as well as from the kinematics of the observed event. A very small signal-to-background ratio and a very complex signature of the $t\bar{t}H$ process seem to be an ideal environment for employing the matrix element method. However, the calculation of the probability can be very computationally intensive and the development of the method for such a complicated process is challenging. The matrix element method has been used for precision measurements of the top quark mass [5, 6] and even for Higgs boson searches [7, 8], but never before for such a complex final state.

This thesis presents a search for the SM Higgs boson produced in association with top quark pairs and decaying into a pair of b -quarks using the matrix element method. The top quark pair decay in the single lepton final state is considered. The data used for this analysis has been recorded with ATLAS, which is one of the two multi-purpose detectors at the LHC. The content of this thesis is organised as follows. In Chapter 2, the SM of particle physics is introduced with an emphasis on top quark and Higgs boson physics. The functionality and experimental setup of the LHC and the ATLAS experiment are briefly summarised in Chapter 3. The data recorded with the ATLAS detector, the physics objects reconstructed from the data and the modelling of the physics processes relevant for this search are detailed in Chapter 4. In Chapter 5, a description of the matrix element method and its implementation for this search is given. Details about the various ingredients and the technical realisation are explained and a simple toy study is performed to prove that the method works as expected. Chapter 6 describes the actual search for the $t\bar{t}H$ process. First, the event selection and the categorisation of the events into different analysis regions is outlined. Then the general analysis strategy and the relevant systematic uncertainties are presented. Results of the matrix element method are shown and combined in an artificial neural network to maximise the separation power. Finally, the statistical analysis is performed by employing a profile likelihood fit, which is used to improve the sensitivity of the search. The $t\bar{t}H$ production rate with respect to the SM is measured and an upper limit on the production is set because no significant excess is found. The result is summarised and conclusions are presented in Chapter 7 along with a short outlook towards future searches and measurements.

Throughout this thesis, natural units implying $\hbar = c = 1$ have been used for subatomic particles, which means that their energies, momenta and masses are expressed in units of eV. However, for all other quantities not referring to subatomic scales the usual International System of Units (SI) is applied.

The Higgs Boson and Top Quarks in the Context of the Standard Model

2.1. Introduction

In this thesis a search for a Higgs boson produced in association with top quarks is presented. Both are elementary particles and part of the Standard Model (SM) of particle physics [9–11], which is briefly summarised in the following. An emphasis is placed on describing how particles acquire mass and interact with the Higgs boson as well as on the production and decay of the top quark and Higgs boson. In this context, the milestones achieved in Higgs boson physics by the two Large Hadron Collider (LHC) experiments A Toroidal LHC ApparatuS (ATLAS) [12] and Compact Muon Solenoid (CMS) [13] are briefly reviewed.

2.2. A Brief Summary of the Standard Model

The SM of particle physics represents our current understanding of elementary particles and their interactions. It is probably one of the most compelling and successful theories in science, since its predictions are being confirmed with remarkable precision in many different experiments. The constituents of matter are spin one-half particles (*fermions*), which are classified into *quarks* and *leptons*. The quarks and leptons interact with each other through the four fundamental forces, the strong, the electromagnetic, the weak and the gravitational force. The SM combines three of the four fundamental forces into one theory based on renormalisable, relativistic quantum field theory (QFT). Because a consistent QFT of gravity, with General Relativity being its low energy manifestation, has not been formulated so far, gravity is not yet included to the SM. However, the gravitational force is by far the weakest and can usually be neglected in the interactions of elementary particles. The forces of the SM are mediated by spin one particles - the gauge *bosons*. The building blocks of our present (low energy) Universe are only the electron, the electron neutrino, the up- and the down-quark, representing the first generation of particles in the SM. There are two additional generations, which basically consist of copies of the particles of the first generation differing only in their masses. The elementary particles of the SM are summarised in Figure 2.1.

An elementary particle participates in a certain interaction, if it carries the charge of the corresponding force. All quarks carry colour charge (red, green, blue) and thus interact through

2. The Higgs Boson and Top Quarks in the Context of the Standard Model

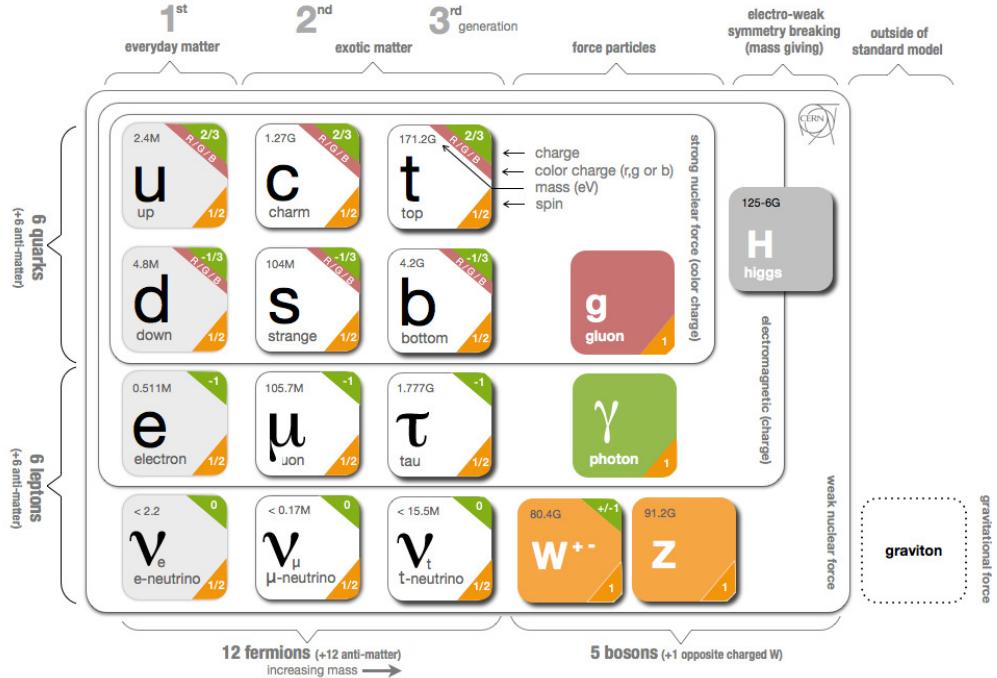


Figure 2.1.: The elementary particles of the SM: three generations of quarks and leptons, four force mediating gauge bosons and the Higgs boson.

the strong force mediated by massless gluons. The gluon exists in eight different versions, each carrying a combination of colour and anti-colour charge. The up-type quarks (up, charm, top) carry an electric charge of $+2/3e$, while the down-type quarks (down, strange, bottom) are electrically negative with a charge of $-1/3e$. The charged leptons (electron, muon, tau) have an integer charge of $-1e$. All electrically charged particles interact through the electromagnetic force mediated by the massless photon with no electric charge. Each charged lepton is complemented by a neutral lepton (electron-, muon-, tau-neutrino) with extremely low mass. All particles interact through the weak force, since they all carry an isospin, which third component is either $+1/2$ (up-type quarks, neutrinos) or $-1/2$ (down-type quarks, charged leptons). The weak force is mediated by the electrically charged W^\pm or the neutral Z vector bosons, both having considerable masses and carrying weak isospin. The W boson is the only force mediating boson carrying electrical charge. As a consequence, it couples together pairs of particles which differ by one unit in electric charge. While in case of leptons, it couples only to pairs of one generation, in case of quarks any exchange between up- and down-type quarks is allowed. The transition probabilities between physical up-type quarks and down-type quarks are given by the Cabibbo-Kobayashi-Maskawa (CKM) matrix [14, 15], which favours transitions within one generation [16]. The masses of the elementary particles are acquired through interactions with the Higgs field, which minimal excitation state is represented by the massive Higgs boson with no spin, electric charge or colour charge. Gravity is expected to be mediated by a spin two graviton, which neither has been discovered nor can be included to the SM.

2.2.1. Mathematical Description

The QFTs of the three forces, Quantum Chromodynamics (QCD), Quantum Flavourdynamics (QFD) and Quantum Electrodynamics (QED), are described in the SM by gauge theories with the unified symmetry group $SU(3) \times SU(2) \times U(1)$ [17–28]. Although not all three theories can yet be described in a combined way, the electroweak (EW) theory by Glashow, Salam and Weinberg [9, 10, 29] unifies the electromagnetic and weak interactions into the $SU(2) \times U(1)$ symmetry group. Following the Lagrangian formalism, the dynamics of a QFT is expressed in terms of a Lagrangian density. For example, the Lagrangian of a free relativistic fermion field $\psi(x)$ is given by

$$\mathcal{L}_D = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi . \quad (2.1)$$

By solving the Euler-Lagrange equation the quantum mechanical field equations are derived. In case of the non-interacting fermion the solution results in the *Dirac equation*. The interactions described by the theories are obtained by requiring *local gauge invariance* of the Lagrangian under the symmetry transformation of the group. For instance, the local phase transformation of the $U(1)$ symmetry of QED is defined by

$$\psi(x) \rightarrow \psi'(x) = e^{iq\chi(x)}\psi(x) , \quad (2.2)$$

where $\chi(x)$ is a scalar phase. The local gauge invariance can only be restored by introducing additional fields of the dimension of the symmetry, which can be identified with the force mediating bosons. This is achieved by replacing the derivative ∂_μ with the *covariant derivative* D_μ . In case of QED this is simply given by

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + iqA_\mu \quad \text{with} \quad A_\mu \rightarrow A'_\mu = A_\mu - \partial_\mu\chi , \quad (2.3)$$

where A_μ is the photon field and q is the charge of the fermion. By adding a kinetic term for the gauge bosons, the Lagrangian describing the full dynamics including interactions is defined. The Lagrangian of QED is then given by

$$\mathcal{L}_{QED} = \underbrace{\bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi}_{\text{fermion kinetic term}} - \underbrace{\frac{1}{4}F_{\mu\nu}F^{\mu\nu}}_{\text{photon kinetic term}} + \underbrace{q\bar{\psi}\gamma^\mu\psi A_\mu}_{\text{fermion-photon interaction}} , \quad (2.4)$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the field strength tensor of QED.

While the generator of the $U(1)$ symmetry of QED is a scalar phase, the generator of the $SU(2)$ and $SU(3)$ can be represented by the Pauli and Gell-Mann matrices, which do not commute. The non-Abelian property of QFD and QCD lead to self-interaction terms of the gauge bosons in the Lagrangian. Due to the number of gauge bosons in QCD, defined by the dimension of the symmetry, the self interaction leads to a special characteristic of the strong force. At low energies, the interaction becomes stronger leading to a *confinement* of the quarks, such that no free quarks can be observed. As a consequence, they form hadrons either consisting of a quark-antiquark pair $q\bar{q}$ (meson) or a triplet of quarks or antiquarks (baryons). Following the Pauli exclusion principle, the quarks in a baryon differ in colour charge and hence quarks always form colourless hadrons.

Noether's theorem applied to field theory states that a conserved current corresponds to every symmetry generated by local actions. For QED the conserved current is the probability current

$$j^\mu = \bar{\psi}\gamma^\mu\psi , \quad (2.5)$$

which satisfies the continuity equation $\partial_\mu j^\mu = 0$. As a consequence, the symmetries of the SM lead to the conservation of electric charge, weak isospin and colour charge. Further observed conservation of quantities, like lepton and baryon number, have no theoretical foundation in the SM.

2.2.2. The Higgs Mechanism

The described local gauge symmetry can only be satisfied if the introduced gauge bosons are massless. Adding mass terms for the gauge bosons breaks the local gauge invariance. While this is not a problem for QED and QCD, where the force is mediated by the massless photon and gluon, it contradicts the three observed heavy gauge bosons (W^\pm and Z) of the weak interaction. Additionally, even the mass term of the fermions, written in terms of chiral particle states, breaks the gauge invariance due to the V-A structure of the weak interaction. This is solved by introducing the *Higgs mechanism* [1–4], from which the particles of the SM acquire their masses by breaking the $SU(2) \times U(1)$ local gauge symmetry of the EW theory. In the minimal model of EW symmetry breaking, the Higgs field is represented by a weak isospin doublet of one charged and one neutral complex scalar field

$$\phi(x) = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \quad (2.6)$$

The Lagrangian for this field

$$\mathcal{L}_{Higgs} = (\partial_\mu \phi)^\dagger (\partial^\mu \phi) - V(\phi) \quad (2.7)$$

consists of a kinetic term and a Higgs potential

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2. \quad (2.8)$$

The first term can be associated with the mass of the field and the second represents the self-interaction of the field. The minima of the potential can be identified with the vacuum expectation value v of the Higgs field. While the parameter λ needs to be positive to obtain a potential with finite minima, the parameter μ can be chosen freely. Figure 2.2 visualises the influence of μ on the form of the potential for one complex scalar field.

If $\mu^2 > 0$, the potential has just one minimum at a value of zero with all fields being zero. Hence the vacuum expectation value of the Higgs field would be zero and the symmetry is preserved. However, for $\mu^2 < 0$ the potential has an infinite set of minima v given by

$$\phi^\dagger \phi = \frac{v^2}{2} = -\frac{\mu^2}{2\lambda}, \quad (2.9)$$

and the choice of the physical vacuum state spontaneously breaks the symmetry of the Lagrangian. Because the photon needs to remain massless after symmetry breaking, only the vacuum expectation value of ϕ^0 is required to be non-zero

$$\langle 0 | \phi | 0 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \quad (2.10)$$

An expansion of ϕ^0 about its vacuum state v introduces a massive scalar and three massless *Goldstone bosons*. However, the Goldstone bosons appear to be not physical and can be

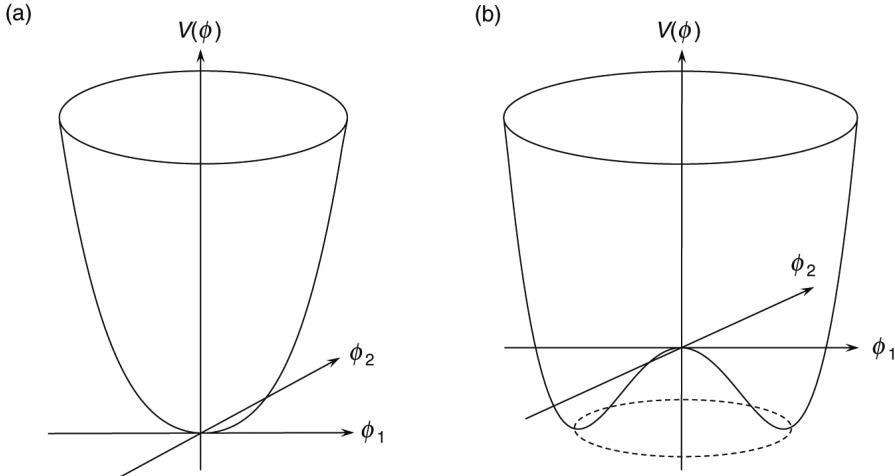


Figure 2.2.: The Higgs potential for $\mu^2 > 0$ (a) and $\mu^2 < 0$ (b).

eliminated using the *Unitary gauge*, which corresponds to choosing the Higgs doublet to be real

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}, \quad (2.11)$$

where h is a physical field which can be identified with the Higgs boson. The covariant derivative of $SU(2) \times U(1)$ acting on the Higgs field is given by

$$D_\mu \phi = \frac{1}{2} [2\partial_\mu + (ig_W \boldsymbol{\sigma} \cdot \mathbf{W}_\mu + ig' B_\mu)] \phi, \quad (2.12)$$

where $\boldsymbol{\sigma}$ are the Pauli spin matrices. The gauge fields \mathbf{W}_μ and B_μ together with the gauge couplings g_W and g' define the interactions of $SU(2) \times U(1)$. The masses of the gauge bosons are generated by the kinetic term in the Lagrangian of Equation 2.7:

$$(D_\mu \phi)^\dagger (D_\mu \phi) = \frac{1}{2} (\partial_\mu h)(\partial^\mu h) + \frac{1}{8} g_W^2 (W_\mu^{(1)} + iW_\mu^{(2)}) (W^{(1)\mu} - iW^{(2)\mu}) (v + h)^2 + \frac{1}{8} (g_W W_\mu^{(3)} - g' B_\mu) (g_W W^{(3)\mu} - g' B^\mu) (v + h)^2. \quad (2.13)$$

Identifying the quadratic terms in $W_\mu^{(1)}$ and $W_\mu^{(2)}$ with the masses of the W bosons yields

$$m_W = \frac{1}{2} g_W v. \quad (2.14)$$

The two neutral fields $W_\mu^{(3)}$ and B_μ mix with each other and two physical fields Z_μ and A_μ can be obtained by a rotation making the mass matrix diagonal

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^{(3)} \end{pmatrix}, \quad (2.15)$$

where the *Weinberg angle* θ_W is defined by $\tan \theta_W = g'/g_W$. The masses of the two neutral gauge bosons, identified with the massless photon and the massive Z boson, are then given by

$$m_A = 0 \quad \text{and} \quad m_Z = \frac{1}{2} v \sqrt{g_W^2 + g'^2}. \quad (2.16)$$

2. The Higgs Boson and Top Quarks in the Context of the Standard Model

The masses of the massive gauge bosons can be related to each other by

$$\frac{m_W}{m_Z} = \cos \theta_W , \quad (2.17)$$

which has been experimentally verified with high precision. The mass of the Higgs boson can be identified from the quadratic term in h to be

$$m_H = \sqrt{2\lambda}v . \quad (2.18)$$

Since the parameter λ is a free parameter in the SM, there is no direct prediction of the Higgs boson mass. However, if the mass is measured and the vacuum expectation value is known, statements about the stability of the Higgs potential can be made. By using the measured W boson mass and the coupling g_W the vacuum expectation value of the Higgs field is found to be

$$v = 246 \text{ GeV} . \quad (2.19)$$

The kinetic term of Equation 2.13 also contains the triple and quartic couplings of the Higgs boson and the vector bosons. Together with the self-interaction of the Higgs boson, the couplings can be summarised by the Lagrangian

$$\mathcal{L}_{hV} = \frac{g_{hhh}}{6} h^3 + \frac{g_{hhhh}}{8} h^4 + \delta_V V_\mu V^\mu \left(g_{hVV} h + \frac{g_{hhVV}}{6} h^2 \right) , \quad (2.20)$$

where V represents one of the vector bosons W, Z and $\delta_W = 1, \delta_Z = 1/2$. The couplings are purely defined by the involved masses and the vacuum expectation value

$$g_{hhh} = \frac{3m_H^2}{v}, \quad g_{hhhh} = \frac{3m_H^2}{v^2}, \quad g_{hVV} = \frac{2m_V^2}{v}, \quad g_{hhVV} = \frac{2m_V^2}{v^2} . \quad (2.21)$$

The Higgs mechanism can also be used to generate the masses of the fermions in a gauge invariant manner. In the SM fermions are described as left-handed chiral $SU(2)$ doublets L and as right-handed $SU(2)$ singlets R . Obeying the chiral structure of the electroweak interaction the gauge invariant term of the form $-y_f(\bar{L}\phi R + \bar{R}\phi^\dagger L)$ yields the Lagrangian

$$\mathcal{L}_{hf} = -\frac{y_f}{\sqrt{2}} v \bar{f} f - \frac{y_f}{\sqrt{2}} \bar{f} f h , \quad (2.22)$$

where y_f denotes the Yukawa coupling of a fermion. Hence the fermion masses m_f are proportional to the Yukawa coupling representing the strength of the coupling of the Higgs boson to the fermions

$$y_f = \sqrt{2} \frac{m_f}{v} . \quad (2.23)$$

All particle masses depend on the vacuum expectation value of the Higgs potential or in other words, the particles acquire their masses due to a non-zero vacuum expectation value of the potential after the symmetry breaking.

2.2.3. Beyond the SM

Despite its great success, the SM cannot be the final theory of particle physics, because there are several observations that are not explained by it. A very brief summary of the most prominent issues is given in the following.

- **dark matter and dark energy:** the measured rotational velocity of spiral galaxies does not decrease with $r^{-1/2}$, although the majority of the visible matter is concentrated in the central bulge [30]. Hence there must be a significant contribution of non-luminous *dark matter* to the total mass of the galaxy. From precision measurements of the fluctuations of the cosmic microwave background, it is concluded that only 5% of the energy-matter density of the Universe consists of baryonic matter, which is the massive matter described by the SM. Dark matter contributes 23% while the majority (72%) of the Universe is composed of *dark energy*, which explains the accelerated expansion of the Universe [31]. Several extensions of the SM, such as supersymmetry (SUSY) [32–40], predict the existence of weakly interacting massive particles (WIMPs), which could significantly contribute to the observed cold dark matter. There are only speculations what dark energy might consist of.
- **unification of forces:** the gauge couplings of the three forces described by the SM have similar strength and change with the energy scale (*running coupling*). Due to the gauge boson self-interactions, the couplings of the weak force α_W and the strong force α_S decrease, while the coupling of the electromagnetic force α_{EM} increases. Hence the running of the coupling constants bring their values together. If a Grand Unified Theory (GUT) combining all three forces exists, the forces described by the same mechanism should have the same coupling constant. In the simplest symmetry group $SU(5)$, which can describe all three interactions, the couplings of the SM converge but do not meet exactly. However, if additional particles, like SUSY particles, contribute to the loop corrections of the gauge boson propagators, they can converge to a single value at a scale of 1 TeV. Hence it is plausible, that the three forces are just the low energy manifestations of some larger unknown theory. Additionally, gravity could not be included in the SM, but certainly plays an important role at the Planck scale $\Lambda_P \sim 10^{19}$ GeV. Large efforts are performed to find a Theory of Everything (TOE) in which all four fundamental forces are unified.
- **hierarchy problem:** quantum loop corrections in the Higgs boson propagator, which contribute to the mass of the Higgs boson, become very large at high energy scales. If the SM is part of a GUT it must be valid to at least $\Lambda_{GUT} \sim 10^{16}$ GeV. Because the corrections are quadratic in Λ , a very precise *fine-tuning* needs to be performed to cancel these loop contributions in order to preserve the Higgs boson mass at the EW scale of roughly 100 GeV [41]. Here again SUSY would provide a natural cancellation of these corrections.
- **matter-antimatter asymmetry:** there is apparently an asymmetry of particles and anti-particles in the observed Universe, which cannot be explained by the observed charge-parity (CP) violation in the flavour sector or possible CP violation in neutrino oscillations. Hence, there should be additional CP violating effects in physics not described by the SM.

- **neutrino masses:** although the Higgs field can generate the masses of the neutrinos, the Yukawa coupling would be unnaturally small. Alternatively, neutrinos could be their own anti-particles (*Majorana particle*) and could obtain their masses through the *seesaw mechanism*, which could explain the smallness of the masses. If that is the case, neutrinoless double β -decay can occur, which is being investigated by several experiments at the moment.

Most of the existing alternative theories or extensions of the SM have been extremely constrained and many have been eliminated by the discovery of the Higgs boson and the observations of its properties. SUSY [32–40] is probably the most attractive theory. However, the symmetry must be broken since no SUSY particles could be observed so far. The minimal SUSY models predict at least five physical Higgs bosons (h, H, A, H^\pm), which need to be observed. The next run of the LHC at energies up to $\sqrt{s} = 14$ TeV will show if SUSY exists at reasonable mass scales.

2.3. The Top Quark

The top quark plays, as the heaviest known elementary particle, an important role in the SM, in particular in the electroweak symmetry breaking. The top quark was discovered in 1995 by the DØ and CDF experiments at the Tevatron [42, 43]. The first world combination of the Tevatron and LHC experiments resulted in a very precisely measured top quark mass of $173.34 \pm 0.27(\text{stat}) \pm 0.71(\text{syst})$ GeV [44]. While the top quark mass is a fundamental parameter of the SM, all of its other properties are predicted by the SM. Any deviation from the predictions would give a hint for new physics beyond the SM. Due to its very short lifetime ($\sim 10^{-25}$ s) the top quark forms no bound states, which leads to the unique opportunity to probe a bare quark. As a consequence, the properties of the top quark are propagated to the decay products and are thus directly accessible. Because it was not produced in sufficient numbers at the Tevatron, a few of the predicted properties were not measured with sufficient precision and can now be studied at the LHC. Prominent measurements of the properties include the mass, electric charge, width, $t\bar{t}$ mass difference, $t\bar{t}$ charge asymmetry, $t\bar{t}$ spin correlations, W helicity, branching fractions and the coupling to other particles. The measurements of the properties by the various experiments are summarised in Reference [16]. An important consequence of its high mass is the strong Yukawa coupling to the Higgs boson, which is approximately one according to Equation 2.23:

$$y_t = \sqrt{2} \frac{m_f}{v} \approx \sqrt{2} \frac{173 \text{ GeV}}{246 \text{ GeV}} \approx 0.99 \approx 1 . \quad (2.24)$$

This might be a coincidence but could also have a deeper reason in extended theories, which might explain the quark mass hierarchy. All these theories predict new particles, which would couple to the top quark and influence many of its properties. Hence, top quark physics is usually considered to be a unique window for searches for new physics. At the same time, top quark production is often the main background process to many direct searches for new physics due to its signatures with large numbers of jets and b -jets.

2.3.1. Top Quark Production

At the LHC, top quarks are mostly produced in pairs via the strong interaction, but also as a single top quark via the electroweak interaction. There are four leading-order (LO) Feynman

diagrams of the top quark pair production, as shown in Figure 2.3: one through $q\bar{q}$ annihilation and three through gluon fusion.

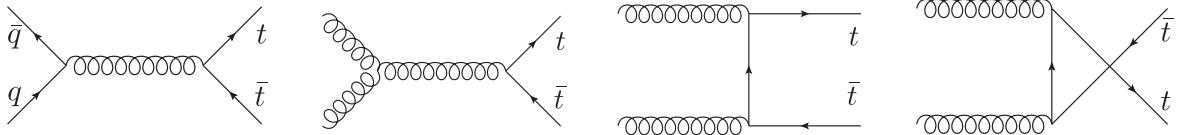


Figure 2.3.: Four LO Feynman diagrams of top quark pair production through strong interaction.

While $q\bar{q}$ annihilation was the dominant production mechanism at the Tevatron $p\bar{p}$ collider, at the LHC gluon fusion dominates the production contributing 80-90% depending on the centre-of-mass energy. The latest theoretical calculation of the top quark pair production cross section at next-to-next-to-leading-order (NNLO) including resummation of soft gluon terms with next-to-next-to-leading-logarithmic (NNLL) accuracy yields $\sigma_{t\bar{t}} = 253^{+15}_{-16} \text{ pb}$ at $\sqrt{s} = 8 \text{ TeV}$ [45–49]. The top quark pair production rate has been measured at all available centre-of-mass energies reaching recently a very high precision with uncertainties of less than 5% at the LHC [50]. The three production diagrams of the single top quark through the electroweak interaction are shown in Figure 2.4.

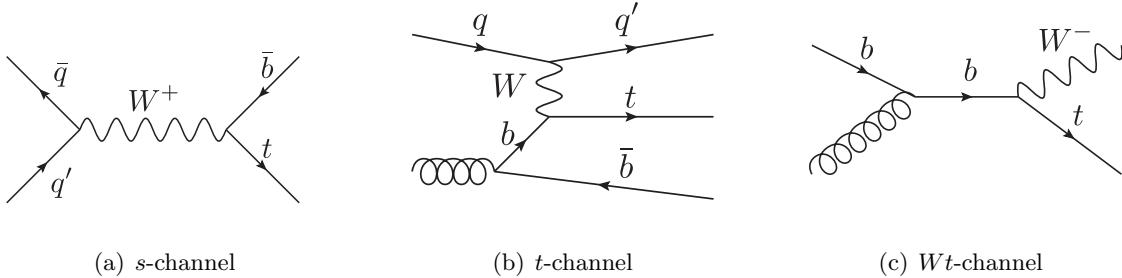


Figure 2.4.: Three LO Feynman diagrams of single top quark production through the weak interaction.

The single top quark production cross section is less than half the size of the $t\bar{t}$ production cross section at the LHC [51, 52]. The process is experimentally much more challenging, because it is accompanied by large backgrounds. While the t -channel was measured at the Tevatron and the LHC [53, 54], the Wt -channel could be only observed at the LHC [55–57]. In contrast to that, only upper limits on the virtual s -channel production cross section have been set at the LHC so far [58, 59], while this production has been observed at the Tevatron [54, 60].

2.3.2. Top Quark Decay

The top quark is the only quark heavier than the W boson and thus can decay into a real W boson. Because the CKM matrix element $|V_{tb}|$ is measured to be close to unity, the top quark decays almost exclusively into a b -quark and a W boson. The latter can either decay leptonically into a charged lepton and the corresponding neutrino or hadronically into a pair of light quarks $q\bar{q}$ ($u\bar{d}$ or $c\bar{s}$). The decays of the top quark pair are classified according to the decay of the two W bosons into a *dilepton*, *single lepton*, and *allhadronic* channel. Because the τ lepton decays leptonically or hadronically, it is usually treated separately and is not included

in the categorisation. However, the leptonic decay of τ leptons result in the same signatures as described in the following and are experimentally included into the dilepton and single lepton channels.

- **dilepton:** refers to the decay $t\bar{t} \rightarrow b\bar{b} W^+W^- \rightarrow b\bar{b} \ell\nu_\ell \ell'\nu'_\ell$ and corresponds to 5% of the branching ratio. In the LO picture two b -jets, two opposite sign charged leptons (ee , $\mu\mu$, $e\mu$) and two neutrinos resulting in large missing transverse energy are expected. Due to its very distinct signature, most of the background can be easily suppressed resulting in a very clean top quark sample. However, a full event reconstruction is difficult, since the sum of the missing transverse energy needs to be associated with the two neutrinos. The small branching ratio can result in phase space regions with low statistics.
- **single lepton:** refers to the decay $t\bar{t} \rightarrow b\bar{b} W^+W^- \rightarrow b\bar{b} q\bar{q}' \ell\nu_\ell$ and corresponds to roughly 30% of the branching ratio. In the LO picture, four jets are expected two of which are b -jets, one charged electron or muon and one neutrino resulting in missing transverse energy. This channel is the best compromise between a clear signature and sufficiently large statistics. The reconstructed missing transverse energy can be identified with the neutrino transverse momentum allowing for a full event reconstruction.
- **allhadronic** refers to the decay $t\bar{t} \rightarrow b\bar{b} W^+W^- \rightarrow b\bar{b} q\bar{q}' q''\bar{q}'''$ and corresponds to roughly 44% of the branching ratio. In the LO picture six jets are expected two of which are b -jets. The channel is contaminated with multijet background and large combinatorial background when assigning the jets to the quarks.

The given branching ratios assume lepton universality.

2.4. The Higgs Boson

On the 4th of July 2012, the two LHC experiments ATLAS [61] and CMS [62] reported the discovery of a new particle in searches for the SM Higgs boson. Until now the measurements of its couplings and its properties have strengthened the assumption that the observed particle with a mass around 125 GeV is indeed the SM Higgs boson. However, in order to verify the SM hypothesis, all possible production and decay rates need to be measured and compared to the SM prediction.

2.4.1. Higgs Boson Production

There are four important production mechanism of a Higgs boson with a mass of 125 GeV at the LHC. One representative Feynman diagram for each of these is shown in Figure 2.5 (a). The production cross section of these processes in pp collisions at a centre-of-mass energy range relevant for the LHC is presented in Figure 2.5 (b). The Higgs boson is predominantly produced via the gluon fusion process (diagram (a)), in which two merging gluons create a quark loop resulting in the creation of a Higgs boson. Since the cross section is proportional to the squared Yukawa coupling ($y_q^2 \sim m_q^2$), the production is mainly mediated by virtual top quark loops, while other quark loops are highly suppressed. The production cross section of $19.3^{+15\%}_{-15\%}$ pb at $\sqrt{s} = 8$ TeV is known with NNLO precision, with large contributions from next-to-leading-order (NLO) corrections (80%) and NNLO corrections (20%) compared to the lower order calculation [63–65].

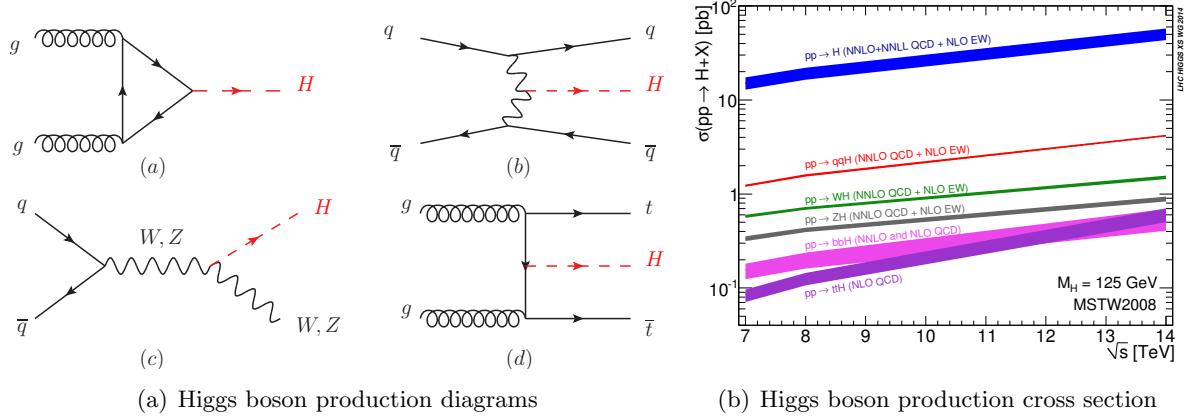


Figure 2.5.: The Higgs boson with a mass of 125 GeV is produced via four main production mechanisms at the LHC, of which representative Feynman diagrams (a) and cross sections in pp collisions (b) for centre-of-mass energies relevant at the LHC are shown [63].

The second leading production mechanism, vector boson fusion as shown in diagram (b), occurs by one order of magnitude less often with a cross section of $1.58^{+3\%}_{-2\%}$ pb at 8 TeV [63–65]. As indicated by the name, two vector bosons (W or Z) mediated from two scattering quarks merge and create a Higgs boson. The scattered quarks result in two hard jets in the forward and backward regions of the detector, whereas gluon radiation in the central region is highly suppressed. This is a very distinct characteristic, which can be used to identify this process cleanly.

As depicted in diagram (c), the Higgs boson can also be produced in association with a W or a Z boson (VH) with cross sections of $0.70^{+4\%}_{-5\%}$ pb and $0.41^{+6\%}_{-6\%}$ pb at 8 TeV, respectively [63–65]. A vector boson is produced through $q\bar{q}$ annihilation and radiates a Higgs boson. Due to a clear signature of a leptonically decaying vector boson, the Higgs boson decay into a pair of b -quarks can be studied.

Finally, the Higgs boson can be produced in association with a top quark pair ($t\bar{t}H$). This process is suppressed by two orders of magnitude compared to gluon fusion. The $t\bar{t}H$ production cross sections of $0.13^{+12\%}_{-18\%}$ pb at 8 TeV is known with NLO accuracy [63–65]. The NLO calculation significantly reduces the scale dependence and increases the LO prediction by 20%. In the energy range of Figure 2.5 (b), the $t\bar{t}H$ cross section has the strongest dependence on the centre-of-mass energy leading to an almost five times larger production at $\sqrt{s} = 14$ TeV. As indicated in diagram (d), a top quark pair is produced through the strong interaction (see Fig. 2.3) and the Higgs boson is radiated off one of the top quarks. Only this production allows for a direct measurement of the top quark Yukawa coupling to the Higgs boson, which is of particular interest. In addition, the distinct signature of the $t\bar{t}$ decay provides a relatively clean environment to identify the Higgs boson decay into a pair of b -quarks.

Similar to the $t\bar{t}H$ production, the Higgs boson can be produced in association with bottom quarks, whose cross section of $0.20^{+12\%}_{-16\%}$ pb is surprisingly higher than the one of the $t\bar{t}H$ production at 8 TeV [63–65]. Although the production is suppressed due to the small b -quark Yukawa coupling being proportional to the b -quark mass, the much smaller required mass of the $b\bar{b}H$ system allows for a production with significantly less centre-of-mass energy of the proton constituents. At energies around $\sqrt{s} = 13$ TeV, the $t\bar{t}H$ production cross section becomes larger than the $b\bar{b}H$ cross section. Because the $b\bar{b}H$ production is overwhelmed by multijet

background, it was not studied experimentally so far. Furthermore, the Higgs boson can be produced in association with a single top quark (tH) with a cross section of $0.018^{+5\%}_{-5\%}$ pb at 8 TeV [63, 66]. Similar to the $t\bar{t}H$ production, the Higgs boson is mainly radiated from the top quark produced in any of the single top quark production process, whose LO diagrams are shown in Figure 2.4. However, the Higgs boson can also be radiated from the W boson propagator. Due to interferences of these two diagram types, the tH production rate is sensitive to the sign of the top quark Yukawa coupling. While the sign is predicted to be positive in the SM resulting in destructive interference, in Beyond the Standard Model (BSM) theories the sign can be negative resulting in constructive interference, which would significantly enhance the production cross section.

2.4.2. Higgs Boson Decay

The possible SM Higgs boson decay modes are very dependent on the Higgs boson mass as shown in Figure 2.6 (a) for a Higgs boson mass range of 80 to 1000 GeV.

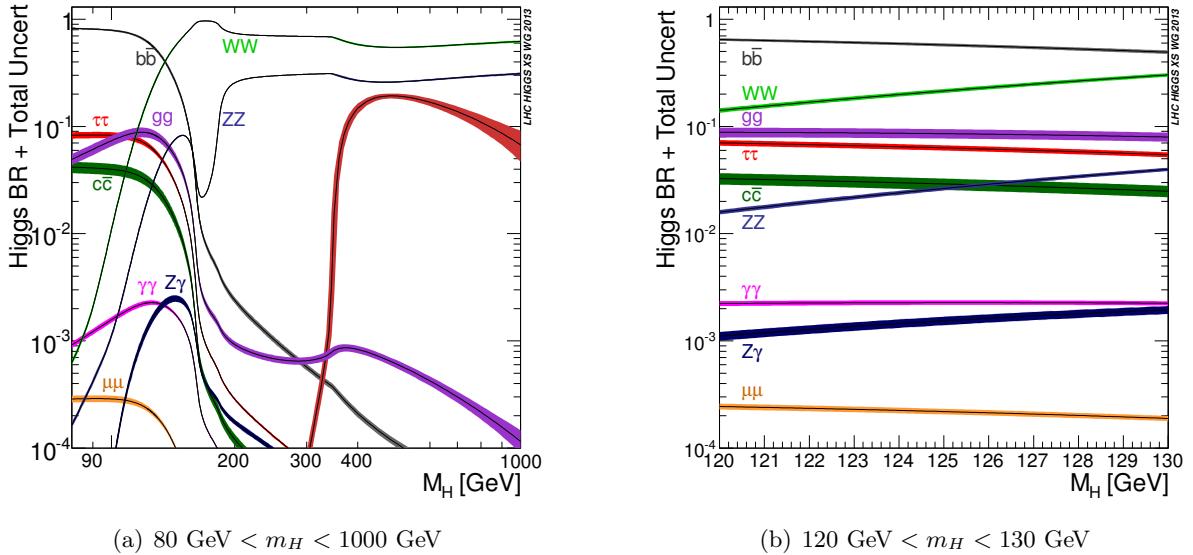


Figure 2.6.: The branching ratios and their total uncertainty of the various SM Higgs boson decay modes for two different mass ranges [63].

If the Higgs boson were heavy enough to decay into two real vector bosons, the modes $H \rightarrow WW^{(*)}$ and $H \rightarrow ZZ^{(*)}$ would have dominated the decay with small contributions from $H \rightarrow t\bar{t}$ for $m_H > 2m_{top}$. At very low masses of the Higgs boson, decays into vector bosons or $t\bar{t}$ would have played almost no role and the dominant decay mode would have been the experimentally challenging decay mode $H \rightarrow b\bar{b}$. Interestingly, the observed Higgs boson has a mass around 125 GeV, corresponding to a very special region, in which various decay modes contribute. Figure 2.6 (b) shows a closer look at the decay rates for Higgs boson masses between 120 and 130 GeV, where all but the $H \rightarrow t\bar{t}$ decay mode have a relevant and quite constant branching ratio. Hence, the Higgs boson decay leads to a large variety of signatures and a very rich phenomenology. Without sufficiently large branching ratios into the two discovery channels, $H \rightarrow ZZ^{(*)}$ and $H \rightarrow \gamma\gamma$, the early discovery of the Higgs boson at the LHC would not have been possible. The relevant branching ratios and their uncertainties for a Higgs boson mass

of $m_H = 125$ GeV are listed in Table 2.1.

Table 2.1.: The SM predictions of the Higgs boson decay branching ratios and their relative uncertainties for $m_H = 125$ GeV [64].

Decay Channel	Branching Ratio(%)	Relative uncertainty
$H \rightarrow b\bar{b}$	57.77	+ 3.2% - 3.3%
$H \rightarrow WW^{(*)}$	21.50	+ 4.3% - 4.2%
$H \rightarrow gg$	8.57	+10.2% -10.0%
$H \rightarrow \tau\tau$	6.32	+ 5.7% - 5.7%
$H \rightarrow c\bar{c}$	2.91	+12.2% -12.2%
$H \rightarrow ZZ^{(*)}$	2.64	+ 4.3% - 4.2%
$H \rightarrow \gamma\gamma$	2.28	+ 5.0% - 4.9%
$H \rightarrow Z\gamma$	1.54	+ 9.0% - 8.8%
$H \rightarrow s\bar{s}$	0.025	+ 4.9% - 4.9%
$H \rightarrow \mu\mu$	0.022	+ 6.0% - 5.9%

At that mass, the dominant decay mode of the Higgs boson is $H \rightarrow b\bar{b}$ with a branching ratio of roughly 57.8%. However, it is also one of the most challenging modes, because of the overwhelming multijet background which can easily mimic the signal. As a consequence, the $H \rightarrow b\bar{b}$ decay can only be observed if the Higgs boson is produced with an additional distinct signature, such as the leptonic decay of a vector boson (VH) or the decay of a top quark pair ($t\bar{t}H$). Despite intensive searches in VH and $t\bar{t}H$ production, this decay mode has not been observed so far [67–71]. The second most important decay mode is $H \rightarrow WW^{(*)}$ with a branching ratio of 21.5%. Since the Higgs boson mass is smaller than the combined mass of two W bosons, one W boson is produced virtually. The decay of the two W bosons provides a variety of different signatures, of which the ones involving leptons are studied at the LHC [72–74]. Although this decay mode was not one of the discovery channels, the ATLAS experiment announced a 6.5σ observation and the CMS collaboration reported a clear evidence (4.7σ) for this decay [73, 75]. Despite the fact that $H \rightarrow gg$ has a relatively large branching ratio, this decay mode is extremely difficult to detect for the same reasons as given for the $H \rightarrow b\bar{b}$ mode. While for the latter the branching ratio is much larger and the identification of b -jets helps to identify the decay, an observation of $H \rightarrow gg$ seems quite unrealistic at the moment. The same statements can be made for $H \rightarrow c\bar{c}$ and $H \rightarrow s\bar{s}$. The $H \rightarrow \tau\tau$ decay mode has a branching fraction of roughly 6.3% and is an important search channel for the study of the Higgs boson coupling to fermions. The either hadronic or leptonic τ lepton decay results in a special signature, which can be detected. First evidence for this decay has been reported by the LHC experiments [73, 76]. Despite the relatively small branching ratios of the $H \rightarrow ZZ^{(*)}$ and $H \rightarrow \gamma\gamma$ decay modes of 2.6% and 2.3%, respectively, these two channels resulted in the discovery of the Higgs boson [61, 62]. The very clear signature and excellent invariant mass resolution of the four leptons or the two photons of the $H \rightarrow ZZ^{(*)} \rightarrow 4\ell$ and the $H \rightarrow \gamma\gamma$ decays lead to the observation. The diphoton invariant mass spectrum and the invariant mass of the four leptons used to identify the Higgs boson decay at the ATLAS experiment are shown in Figure 2.7.

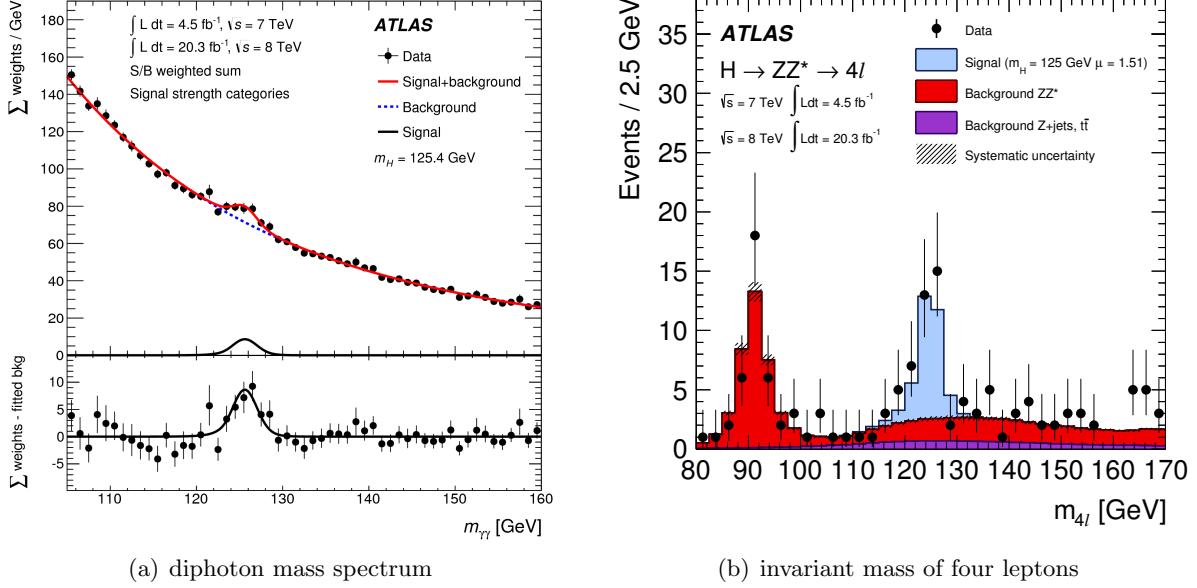


Figure 2.7.: The diphoton mass spectrum (a) and the invariant mass of the four lepton system (b) as reconstructed by the ATLAS collaboration for the discovery but with the full Run I data set [77, 78].

Although the background to the $H \rightarrow \gamma\gamma$ decay is high in the diphoton mass spectrum (a), it is expected to fall smoothly and exponentially and hence can be modelled and subtracted. In the subtracted distribution, a distinct mass peak arises around 125 GeV owing to the excellent resolution of the reconstructed photons. The peak corresponds to a 5.2σ excess over the background-only hypothesis [77]. The four leptons in the $H \rightarrow ZZ^{(*)} \rightarrow 4\ell$ decay (b) result in a very narrow mass peak around 125 GeV over a relatively small background distribution. ATLAS observes a significance of 8.1σ for this excess over the background-only predictions [78].

Finally, the $H \rightarrow Z\gamma$ and $H \rightarrow \mu\mu$ are two Higgs boson decay modes with very small branching ratios of around 0.02%. Nevertheless, they are experimentally accessible because of the clear signatures of two leptons and a photon or two opposite charged muons, respectively.

2.4.3. Higgs Boson Measurements

Since the discovery of the Higgs boson, the LHC experiments have put great effort into measuring the properties of the observed particle in order to investigate if it is a SM Higgs boson.

Mass

As shown in Figure 2.7, the discovery channels directly provide a mass measurement due to the excellent resolutions of the reconstructed leptons and photons resulting in a Higgs boson mass resolution of 1-2%. The result of the latest mass combination of the ATLAS and CMS measurements is presented in Figure 2.8.

In the combination, the SM Higgs boson mass is

$$m_H = 125.09 \pm 0.21(\text{stat}) \pm 0.11(\text{syst}) \text{ GeV} \quad (2.25)$$

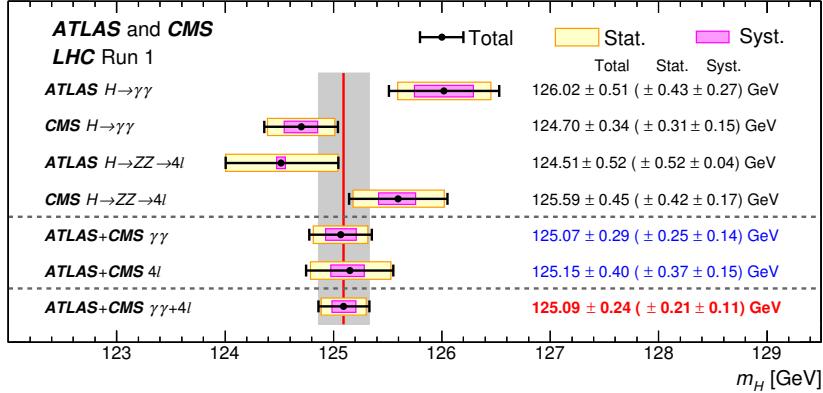


Figure 2.8.: Summary of the Higgs boson mass measurements of the ATLAS and CMS experiments in the individual channels and in combination [79].

with a remarkable precision [79]. The $H \rightarrow WW^{(*)}$ decay mode does not contribute to this measurement, because of its much poorer mass resolution of around 20% due to the neutrinos in the final state. The Higgs boson mass is a parameter in the SM. However, it can be predicted by electroweak fits using precision measurements of the top quark and W boson mass. Although the most likely value suggests a slightly smaller Higgs boson mass, the observed mass agrees within $1 - 2 \sigma$ with the expectation [80, 81].

Spin, Charge and Parity

In the SM, the Higgs boson is a CP-even scalar particle corresponding to the quantum number $J^{CP} = 0^{++}$. The spin J and CP quantum numbers were subject of studies at the ATLAS and CMS experiments [82–84]. The positive C quantum number follows from the observation of the $H \rightarrow \gamma\gamma$ decay assuming charge conservation. Similarly, the spin $J = 1$ hypothesis is excluded according to the *Landau-Yang theorem* [85, 86] due to this observation. Measurements of $H \rightarrow ZZ^{(*)} \rightarrow 4l$ and $H \rightarrow WW^{(*)} \rightarrow l\nu l\nu$ suggest a positive P quantum number [83, 87]. A wide range of models with spin $J = 2$ as well as CP-even and CP-odd BSM hypotheses with spin $J = 0$ are excluded in most cases with more than 99% confidence level (CL) favouring the SM $J^{CP} = 0^{++}$ hypothesis [82–84].

Signal Strength

Both LHC experiments performed searches for all experimentally accessible production mechanisms and decay modes of the Higgs boson [73, 88]. In these searches, the cross section relative to the SM cross section, the *signal strength* $\mu = \sigma_{\text{obs}}/\sigma_{\text{SM}}$, is measured in order to compare the production and decay rates observed in the data with the SM hypothesis. The summary of the ATLAS result of these searches is shown in Figure 2.9.

The measured signal strength of the combined $H \rightarrow \gamma\gamma$ decay mode is consistent with the SM expectation and the uncertainty is small enough to exclude the background-only hypothesis. The sensitivity is mainly driven by the analysis focussing on the gluon fusion production. In the combination with the $H \rightarrow ZZ^{(*)}$ decay mode a slightly higher signal strength is obtained and the SM expectation of $\mu = 1$ lies slightly outside the 1σ uncertainty band. With the

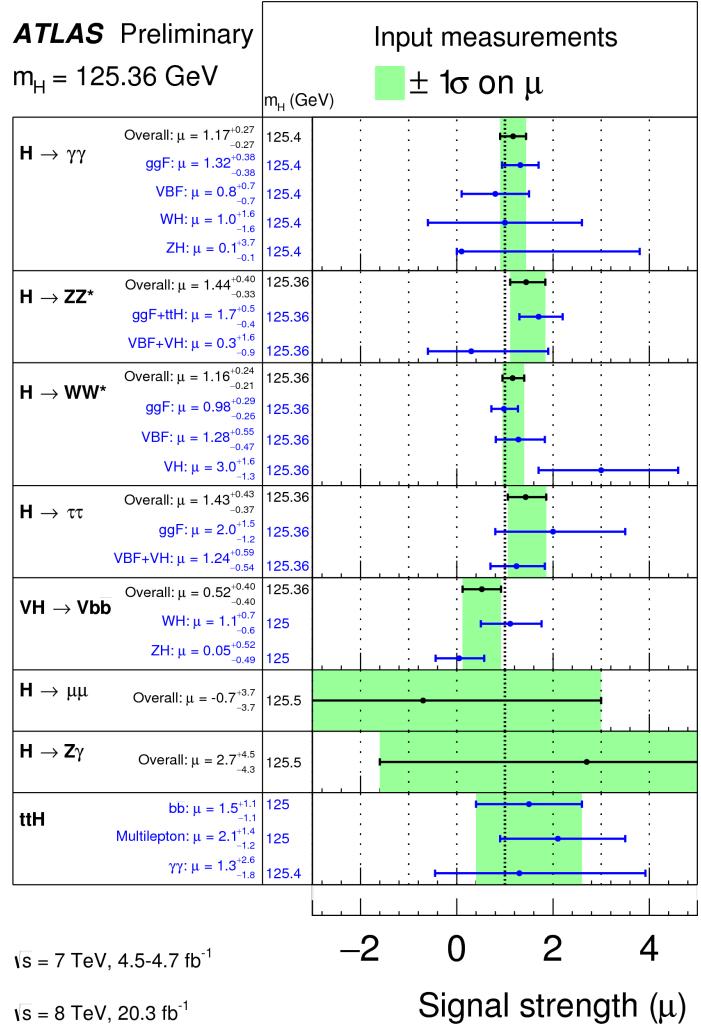


Figure 2.9: Summary of the ATLAS signal strength measurements for a Higgs boson mass around 125 GeV categorised in the different decay modes. The exact Higgs boson masses used in the analysis are quoted in the first column. The signal strength is measured in individual analyses (blue) taking the various production mechanisms into account and is combined per decay mode (black). The error bars represent the statistical and systematic uncertainties. The green shaded bands indicate the uncertainty of the combined signal strength for its decay mode. VH and $t\bar{t}H$ production is treated separately, however the combined signal strength of the $H \rightarrow \gamma\gamma$ mode includes the $t\bar{t}H$ contribution [88].

obtained uncertainty, the background-only hypothesis is excluded. The last decay mode where the Higgs boson is observed is the $H \rightarrow WW^{(*)}$ decay, whose signal strength agrees well with the SM hypothesis within remarkably small uncertainties. In all remaining measurements, the observed signal strength is consistent with the SM predictions within 1σ deviation or slightly larger deviation in case of the $H \rightarrow \tau\tau$ and $V(H \rightarrow b\bar{b})$ analyses. In particular, the measurements in the $H \rightarrow \mu\mu$ and $H \rightarrow Z\gamma$ decay channels have very large uncertainties and are not sensitive to the SM predictions yet.

All measurements suggest that the observed Higgs boson is produced and decays according to the SM expectations. The combination of all published measurements of the ATLAS experiment yields a global signal strength of $\mu = 1.18^{+0.15}_{-0.14}$, which is consistent with the SM predictions with a p-value of 18% [88]. From these measurements, only the gluon fusion process is observed with a significance of more than five sigma, while there is a strong evidence for vector boson fusion (4.3σ). The SM hypotheses of VH (2.6σ) and $t\bar{t}H$ (2.4σ) production are supported by the measurements. So far the Higgs boson has only been observed in couplings to bosons, while in the ATLAS combination there is strong evidence (4.5σ) for a coupling to down-type fermions. Similar results have been published by the CMS collaboration in Reference [73]. The latest results of searches for the $t\bar{t}H$ production are summarised in Section 7.1.

The ATLAS Experiment at the LHC

3.1. Introduction

The LHC at Conseil Européen pour la Recherche Nucléaire (CERN) near Geneva is the only collider in the world, which can produce the SM Higgs boson in association with top quark pairs in sufficient numbers. The analysis presented in this thesis is performed with the ATLAS detector - one of the experiments measuring the collisions of the LHC. The LHC and the ATLAS detector have been designed and constructed over decades by a world wide collaboration of scientists and engineers constantly pushing technological frontiers to new limits. Many of them could never profit from their work by actually analysing recorded data of the performed collisions. The regular operations of the machines started around the same time as the preparations for this thesis. Only this very privileged time allowed to perform a search for such a rare process as the $t\bar{t}H$ production. The LHC and the ATLAS detector will be very briefly introduced in the following chapter. A more detailed description can be found in References [12, 89].

3.2. The Large Hadron Collider

The LHC [89] at CERN is a circular proton or heavy ion collider and the most powerful particle accelerator in the world. It was built in the former tunnel of the Large Electron-Positron Collider (LEP) [90], which has a circumference of approximately 27 km and lies around 100 m in the underground of the French-Swiss countryside near Geneva. During its first run (Run I) from 2010 to 2013, it was operated at centre-of-mass energies of $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV bringing two proton beams to collisions inside its experiments. After Run I, the machine and its experiments went into a planned shutdown phase (LS1) of two years, which was used to upgrade and prepare them for higher energies and luminosities. The machine started its operation again in early 2015 and is expected to reach its design energy of $\sqrt{s} = 14$ TeV during the next Run II. These high energies can only be reached by a long pre-accelerator chain which is sketched in Figure 3.1.

In its main operation mode of pp collisions, ionised hydrogen atoms are accelerated in the Linear Particle Accelerator (LINAC) 2 to an energy of 50 MeV. The protons are then injected to the Proton Synchrotron BOOSTER increasing their energy to 1.5 GeV and are passed to

3. The ATLAS Experiment at the LHC

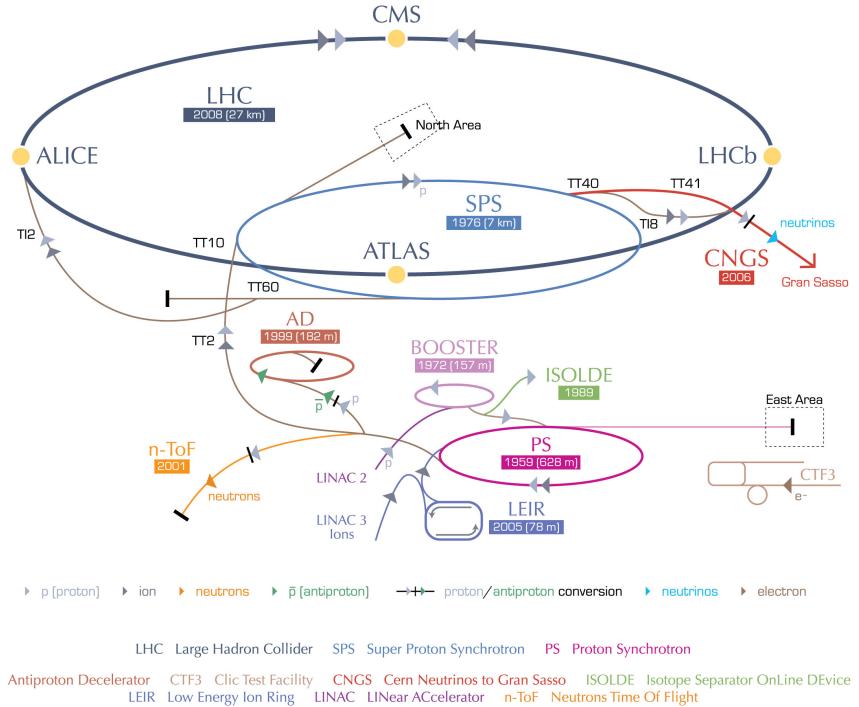


Figure 3.1.: The LHC accelerator chain and main experiments within the CERN accelerator complex.

the Proton Synchrotron (PS) in order to reach energies of 25 GeV. Finally, they are transferred to the Super Proton Synchrotron (SPS) and accelerated to the LHC injection energy of 450 GeV, before two beams of protons are filled in the LHC in opposite directions.

Inside the beam pipes of the LHC exists an ultrahigh vacuum of 10^{-10} mbar to avoid disturbances of gas particles. Radio frequency cavities along the beam pipes accelerate the beams to an energy of up to 7 TeV per beam. The proton beams are kept on their circular trajectory by 1 232 superconducting dipole magnets with field strengths of up to 8.4 T. Additionally, almost 400 superconducting quadrupole magnets focus the beams and magnetic fields of higher order correct the beam shapes from various distortions. To maintain the superconductivity of the magnets, the machine is cooled down to 1.9 K by more than 96 tonnes of superfluid helium. At full design luminosity of $10^{-34} \text{ cm}^{-2}\text{s}^{-1}$ a total of 2 808 bunches are circulating with a spacing of 25 ns inside the LHC, each consisting of around 10^{11} protons. When the bunches of the two beams are crossed inside the experiments, they can create almost a billion interactions per second.

The ATLAS [12] and CMS [13] experiments are both multi-purpose detectors designed for a large variety of physics analyses in pp collisions. However, their main goals are the discovery and measurements of the Higgs boson as well as searches for new physics. The LHC-beauty (LHCb) [91] experiment is designed asymmetric to serve the needs of heavy flavour physics investigating CP violation and searching for BSM physics. A Large Ion Collider Experiment (ALICE) [92] is specialised for heavy ion collisions and is focusing on physics of strongly interacting matter at extreme energy densities. In addition to these four main experiments, the three smaller and very specialised experiments LHC-forward (LHCf) [93], TOTal Elastic and diffractive cross section Measurement (TOTEM) [94] and Monopole and Exotics Detector At the LHC (MoEDAL) [95] are located along the beam line.

3.3. The ATLAS Detector

The ATLAS detector is housed in an underground cavern at Point 1, which is one of the interaction points of the LHC. It is designed to identify a large spectrum of traversing particles as well as to measure their tracks and energies with very high precision. At the same time, it has to handle the high interaction rates, radiation doses, particle multiplicities and energies caused by the proton collisions [12]. Following the typical onion-skin structure of particle detectors at colliders, ATLAS is almost cylindrical around the interaction point and symmetric in the forward-backward direction covering almost the full solid angle. As shown in the schematic of Figure 3.2, ATLAS is a huge technical apparatus with a length of 44 m, a height of 25 m and a weight of approximately 7000 tonnes.

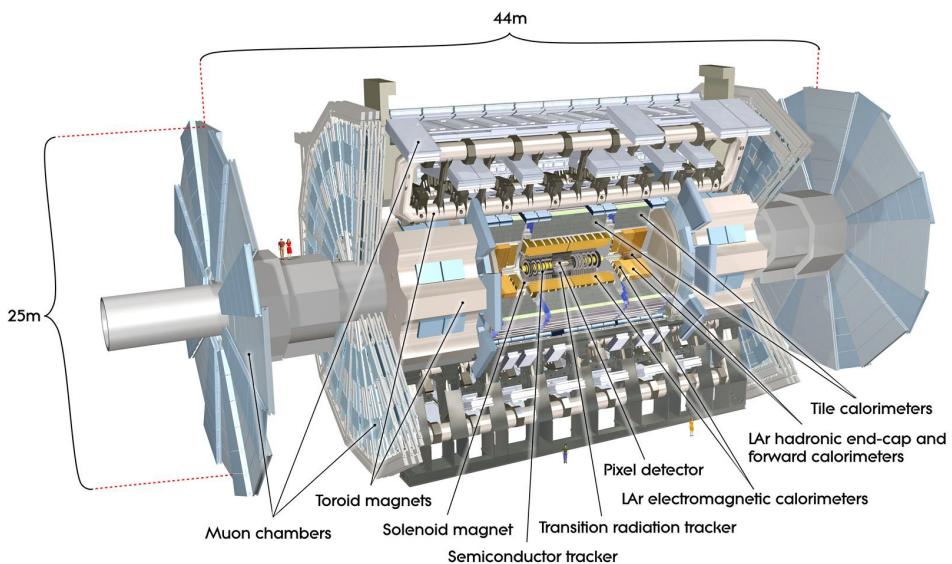


Figure 3.2.: Schematic drawings of the ATLAS detector and its components [12].

It is composed of many subcomponents, which can be grouped into the magnet system, the Inner Detector (ID), the calorimeter and the Muon Spectrometer (MS).

The ATLAS Coordinate System

In order to describe the ATLAS detector and the measured particles, a right-handed cartesian coordinate system (x, y, z) with its origin in the nominal interaction point is used. The x -axis is pointing from the interaction point to the centre of the LHC ring, the y -axis is pointing upwards and the z -axis is defined by the counter-clockwise beam direction. Transforming to a spherical coordinate system (r, θ, ϕ) the azimuthal angle ϕ lies in the x - y -plane around the beam axis z and the polar angle θ in the r - z -plane with $r = \sqrt{x^2 + y^2}$. However, when describing highly relativistic particles with $m \ll E$, it is convenient to use the *pseudorapidity*

$$\eta = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right] \quad (3.1)$$

3. The ATLAS Experiment at the LHC

instead of the angle θ . The difference in pseudorapidity $\Delta\eta$ is invariant under a Lorentz transformation. For the same reason, the distance

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \quad (3.2)$$

in η - ϕ space is often used. Finally, the transverse momentum p_T of a particle is defined as

$$p_T = \sqrt{p_x^2 + p_y^2} = |p| \sin \theta , \quad (3.3)$$

where p_x and p_y are the momentum components in the x - and y -direction, respectively, and $|p|$ is the absolute value of the momentum.

3.3.1. Magnet System

The ATLAS magnet system consists of one solenoid and three toroid superconducting magnets as sketched in Figure 3.3.

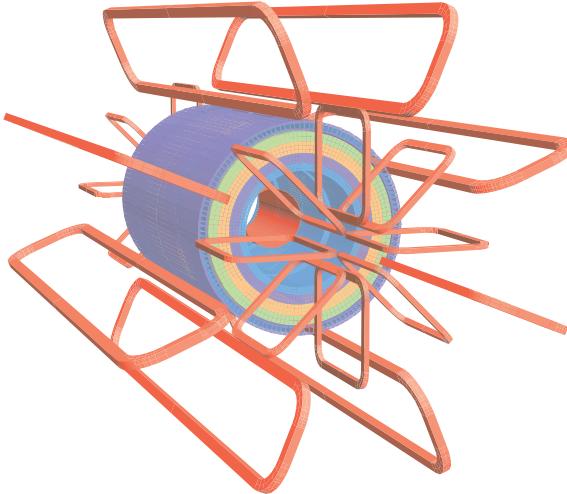


Figure 3.3.: Schematic drawing of the ATLAS magnet system [12].

The central solenoid magnet surrounds the tracking system of the ID along a distance of 5.8 m and produces a magnetic field parallel to the beam axis of 2 T. To minimise the energy loss of traversing particles, it is constructed with a minimal amount of material and shares the cooling cryostat of the calorimeter. The calorimeter and its structure also serves as return yoke for the magnetic flux.

The toroid system provides the magnetic field for the MS and is composed of one large magnet in the barrel and two smaller magnets in each end-cap region of the detector. Each toroid is constructed out of eight air-core coils, which generate strongly varying magnetic fields with field strengths of up to 4.1 T. Both magnet systems are cooled down by liquid helium to a temperature of 4.5 K to enable the superconductivity.

3.3.2. Inner Detector

The ID is the tracking system of the ATLAS detector. Electrically charged particles leave traces inside the detector by creating electron-hole pairs in semiconductors or by ionising gas.

The magnetic field of the solenoid bends the path of a charged particle to a curved trajectory. By reconstructing the tracks of the particles, their momentum and interaction vertices can be determined. After every collision more than thousand particles emerge from the interaction point causing a very high track density in the detector. A precise measurement of their tracks requires a fine granularity of the tracking system degrading with the distance to the interaction point. This is achieved by a three-component system consisting of the Pixel detector, the Semiconductor Tracker (SCT) and the Transition Radiation Tracker (TRT) as illustrated in Figure 3.4.

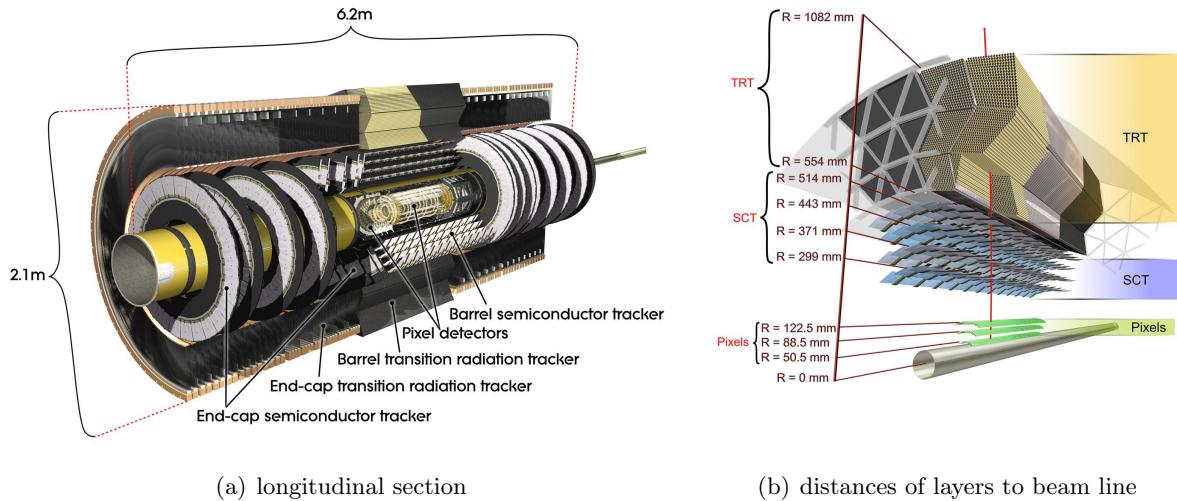


Figure 3.4.: Schematic drawing of the ATLAS Inner Detector with its subsystems Pixel, SCT and TRT in the barrel and the end-cap regions [12].

The Pixel detector and the SCT are based on doped semiconductor sensors and cover the range of $|\eta| < 2.5$. The innermost Pixel detector consists of three cylindrical barrel layers and three perpendicular oriented disc layers on each side in the end-cap regions. The three barrel layers are located in a close distance of $50.5 - 122.5$ mm to the beam line allowing for a precise reconstruction of the vertices. Aiming for a good momentum and vertex resolution, the Pixel detector has the finest granularity of the ID and is highly segmented in 1 744 sensors. Each sensor is a silicon wafer sectioned into Pixel with a minimum size of $50 \times 400 \mu\text{m}^2$ resulting in almost 50 thousand readout channels. The Pixel system is cooled down to -10°C in order to reduce thermal noise and to obtain an optimal intrinsic spatial resolution. During the LS1, a new beam pipe with smaller radius and an attached additional Pixel layer was inserted into the detector [96]. This *Insertable B-layer (IBL)* is placed at a distance of 25.7 mm to the beam line and is expected to significantly improve the reconstruction of secondary vertices caused by heavy flavour (HF) decays.

The SCT is composed of four cylindrical double strip layers parallel to the beam axis and nine end-cap disks with radially oriented strips. The barrel layers are considerably further away from the beam line at a distance of $299 - 514$ mm and half of their modules are arranged with small stereo angle to improve the spatial resolution. Each of the in total 15 912 strip sensors have a length of 6.4 cm and a strip pitch of $80 \mu\text{m}$.

The TRT uses straw tubes of 4 mm diameter filled with a $\text{Xe}/\text{CO}_2/\text{O}_2$ gaseous mixture to

3. The ATLAS Experiment at the LHC

track and identify the traversing particles. It covers a range of $|\eta| < 2.0$, but provides no track information along the beam axis in the barrel and in r -direction in the end-caps. In the barrel, the straw tubes are arranged parallel to the beam axis in 73 planes, while in the end-caps they are radially oriented in 160 planes. The barrel tubes have a distance of 554 – 1082 mm to the beam axis. The space between the tubes is filled with polyethylene with widely varying indices of refraction, which allows to identify electrons and charged pions making use of the transition radiation.

A typical track in the ID is reconstructed from three hits in the Pixel detector, eight hits in the SCT and 36 hits in the TRT. The combination of the three different systems leads to a robust track reconstruction with good transverse momentum resolution given by the *Glückstern formula* [97]

$$\frac{\sigma(p_T)}{p_T} = \frac{\sigma(x) \cdot p_T}{0.3 \cdot L^2 \cdot B} \sqrt{\frac{720}{N + 4}}, \quad (3.4)$$

where $\sigma(x)$ is the spatial uncertainty, L the length of the track, B the magnetic field of the solenoid and N the number of measured track points.

3.3.3. Calorimeter

The calorimeter system measures the energy of electromagnetically and hadronically interacting particles, such as electrons, τ leptons, photons and hadrons. The full energy can only be measured if the particles are completely contained by the calorimeter. Muons with typical energies at collider experiments act as *minimum ionizing particles (MIPs)* and escape the calorimeter leaving only small traces of ionisation. The calorimeter system of ATLAS, as presented in Figure 3.5, is divided into the electromagnetic (EM) and hadronic calorimeter both being sampling calorimeters.

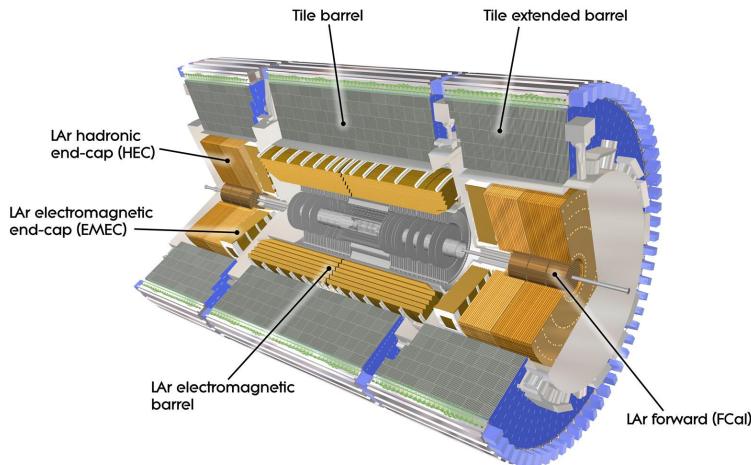


Figure 3.5.: Schematic drawing of the liquid Argon and Tile calorimeter of ATLAS [12].

In a sampling calorimeter, very dense absorber material is alternated with a highly ionisable active material. The absorber material induces particle showers, whose energy deposits are measured in the active material. Although the energy deposited in the absorber material cannot be measured, the sampling design allows to reduce the size of the calorimeter.

Electromagnetic Calorimeter

In the EM calorimeter, mostly electrons and photons create electromagnetic showers through an interplay of Bremsstrahlung and pair production. The electromagnetic barrel (EMB) and electromagnetic end-cap (EMEC) calorimeters cover a range of $|\eta| < 1.475$ and of $1.375 < |\eta| < 3.2$, respectively. They consist of accordion-shaped lead absorbers with liquid Argon (LAr) as active material in between ensuring good coverage at high granularity. The LAr is chosen as active medium due to its intrinsic linear behaviour, stable response time and radiation hardness [12]. The EMB has the finest granularity of the calorimeter and is segmented into three longitudinal layers. The forward region of $3.1 < |\eta| < 4.9$ is covered by the three layered Forward Calorimeter (FCal), which is designed to cope with high radiation and particle flux due to its vicinity to the beam pipe. The first wheel (FCal1) is part of EM calorimeter and is composed of a matrix of concentric copper rods and tubes serving as absorber with LAr as active material in the gaps. A thin LAr layer (presampler) in the range of $0 < |\eta| < 1.8$ complements the EM calorimeter estimating the energy loss due to the ID material in front of the calorimeter. With a thickness of at least 22 radiation lengths X_0 , punch-throughs to the hadronic calorimeter of electromagnetically interacting particles are mostly avoided.

Hadronic Calorimeter

Hadronically interacting particles form hadronic showers (*jets*) by interacting via various processes with the nuclei of the absorber material. The hadronic calorimeter consists of a Tile calorimeter in the barrel region as well as a LAr calorimeter in the hadronic end-cap (HEC) and forward region. In the Tile calorimeter, steel is used as absorber and plastic scintillators as active material. It is subdivided into a central and two extended barrels and covers a range of $|\eta| < 1.7$. The scintillator light is passed through wavelength shifting fibers to photomultiplier in order to measure the deposited energy. The HEC is a LAr sampling calorimeter with copper as absorber and is composed of two cylindrical wheels covering a range of $1.5 < |\eta| < 3.2$. The two hadronic FCal modules (FCal2 and FCal3) are similarly designed to FCal1, but using tungsten instead of copper as absorber to minimize the lateral spread of the hadronic showers. The relative energy resolution of the electromagnetic and hadronic calorimeter is given by

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c, \quad (3.5)$$

where a has been measured to be $a = (10.1 \pm 0.4)\%$ for electrons in the EM calorimeter and $a = (52 \pm 1.0)\%$ for pions in the hadronic calorimeter [12]. Instrumental effects proportional to $1/E$ and constant calibration effects are taken into account by the factors b and c .

3.3.4. Muon Spectrometer

Muons are the only detectable particles escaping the calorimeter. The MS is the outermost part of the ATLAS detector and is designed to track muons in addition to the ID. Since the tracks are bent by the magnetic field of the toroid system, the momentum of the muons can be measured. The various components of the MS are shown in Figure 3.6 and cover a range of $|\eta| < 2.7$.

In addition to a precise determination of the muon momentum, the MS provides information to the trigger system and hence needs to respond fast. To satisfy both requirements, the system consists of a combination of high-precision tracking and very fast trigger systems. In the

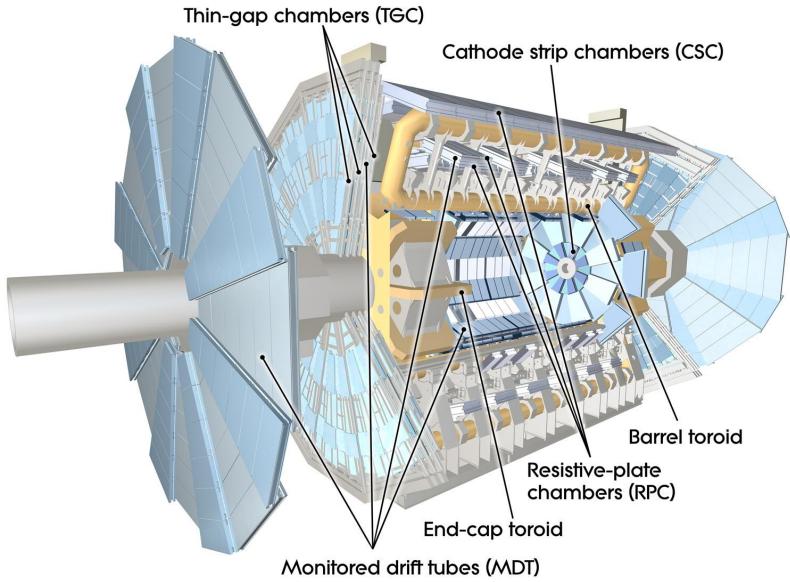


Figure 3.6.: Schematic drawing of the Muon Spectrometer consisting of the Thin Gap Chambers, Cathode Strip Chambers, Monitored Drift Tubes and Resistive Plate Chambers [12].

barrel, very reliable and robust Monitored Drift Tubes (MDTs) are used to measure the tracks precisely, while the fast Resistive Plate Chambers (RPCs) provide the trigger information. In the end-caps the precise tracking is performed by Cathode Strip Chambers (CSCs), whereas the quickly responding Thin Gap Chambers (TGCs) are used for triggering. The MDT system is composed of 1 088 chambers with 3-8 layers of proportional counters and covers a range of $|\eta| < 2.0$. The Aluminium cylinders of the counters are filled with a mixture of Ar/CO₂ gas and contain only one sense wire in the centre. Because the MDTs are limited by their counting rate, CSCs are used instead in the end-caps ($2.0 < |\eta| < 2.7$), where a higher particle flux is expected. The CSC consists of two disks each composed of 16 multi-wire proportional chambers filled with Ar/CO₂ gas. Due to a sophisticated arrangement of the wires, the drift time of the electrons is significantly reduced allowing for a much higher counting rate. The RPCs cover a range of $|\eta| < 1.05$ and consist of parallel electrode plates with a gas mixture of C₂H₂F₄, Iso-C₄H₁₀ and SF₆ in between. The TGCs consist similar to the CSCs of multi-wire proportional chambers and cover a range of $1.05 < |\eta| < 2.4$. Compared to the CSC, the wire-to-cathode distance is shorter than the wire-to-wire distance, which allows for an extremely fast charge collection time.

3.3.5. Trigger and Data Acquisition

With the design luminosity and bunch spacing of the LHC, the interaction rate inside the ATLAS detector is of the order of 1 GHz. In terms of raw data collected with all components of the detector, this corresponds to a data rate of approximately 60 Tb/s, which is technically impossible to handle. However, most of the collisions lead to rather uninteresting physics events of well known processes, which do not need to be recorded. Most of the interesting physics events have very typical signatures, such as high energetic leptons and jets or a large transverse momentum imbalance. By selecting only these events, the event rate can be significantly reduced to approximately 200 Hz resulting in a data rate below the manageable rate of

300 Mb/s. However, the selection needs to happen extremely fast, in real time and very carefully in order to not miss any events created by new physics processes. The Trigger and Data Acquisition (TDAQ) system of ATLAS uses a three level trigger system to filter interesting events. A flow chart diagram of the main components of the system is shown Figure 3.7.

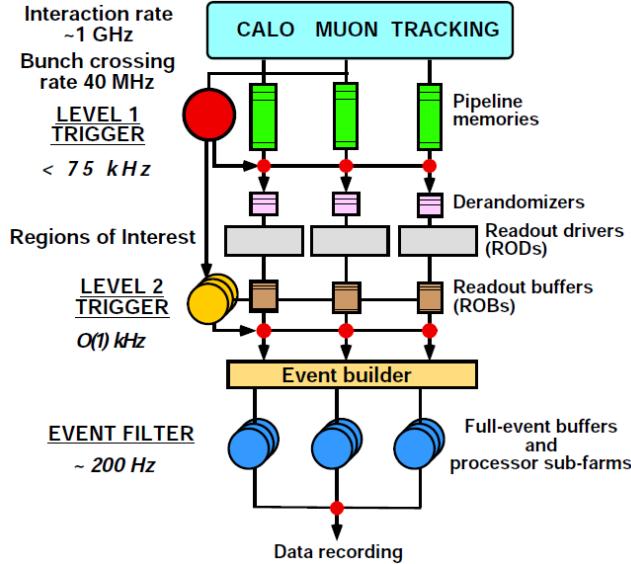


Figure 3.7.: Flow chart of the ATLAS trigger chain consisting of Level 1, Level 2 and Event Filter [12].

The first stage, the Level 1 (L1) trigger, is completely hardware-based, which means logical electronics perform the operations. By deciding if the event is of general physics interest, it reduces the event rate below 75 kHz. Because the taken data can only be stored for a very short time in pipeline memories, the decision has to be taken within $2.5 \mu\text{s}$ after the interaction occurs. Therefore the hardware is integrated into the relevant detector parts. The decision is purely based on information from the fast trigger part of the MS and the calorimeter. If a muon is identified by the RPC or TGC, its momentum is roughly estimated from the hits using look-up tables, since there is no time for a track reconstruction. The calorimeter is used with a coarse granularity looking for high p_T electrons, photons, jets, large transverse momentum imbalance or a large total amount of transverse energy. If the event is accepted by the L1 trigger, the geometry of the triggered objects is passed as Region of Interest (ROI) to the next trigger level.

The Level 2 (L2) trigger is purely software-based and analyses the ROIs with full granularity of all detector components including the tracking system. With a latency time of 10 ms, it can run more complex but still very simple and efficient reconstruction algorithms. By measuring the p_T of the objects more precisely and requiring isolation criteria on the leptons, the event rate can be reduced to the order of 1 kHz.

Finally, the global event information is passed to the software-based Event Filter (EF), which has a latency time of the order of a second. During that time, the objects of the event are fully reconstructed using complex algorithms including alignment and calibration information. Additional information, such as reconstructed vertices or globally reconstructed transverse momentum imbalance, is used to reduce the event rate to 200 Hz. Events passing the EF are stored permanently on local storage elements and are distributed around the world.

4

Experimental Data, Physics Objects and Processes

4.1. Introduction

The LHC delivers collisions of proton beams accelerated under characteristic run conditions for the experiments. The collisions inside the ATLAS detector result in a certain amount of experimental data, which can then be analysed for a specific purpose. The raw recorded data of the ATLAS detector, however, consists only of electrical currents and voltages measured in the different components of the detector. From these quantities tracks and energy deposits are derived, which are then used to reconstruct various physics objects. The characteristics of these objects can vary and are usually optimised for the individual analysis. No analysis in high energy physics can be performed without a model, which can be compared to the experimental data. This model is usually obtained from Monte Carlo (MC) simulations or is estimated using the experimental data.

In the following chapter, the commissioning and run conditions of the LHC since the start of its operations and the resulting experimental data recorded with the ATLAS detector are described. Subsequently, the physics objects such as electrons, muons and jets, used for this analysis are defined. Finally, the modelling of the relevant physics processes is specified, which is later used for comparison with the experimental data.

4.2. Description of the Experimental Data

In accelerator physics the *instantaneous luminosity* representing a particle current density is used to quantify the performance of the accelerator and thus its capability of creating collisions. In a ring collider such as the LHC, the luminosity

$$\mathcal{L} = \frac{N_1 N_2 k_b f}{A} \quad (4.1)$$

depends on the number of bunches k_b in the beams, which collide with the revolution frequency f , the number of particles per bunch (N_1, N_2) and the cross sectional area A of the bunch collision. Expressing the area $A = \frac{4\pi\sigma_x^*\sigma_y^*}{F}$ by the Gaussian width of the beam σ^* in x - and

y -direction at the interaction point, taking into account the crossing angle with a reduction factor F , Equation 4.1 takes the form

$$\mathcal{L} = \frac{N_1 N_2 k_b f}{4\pi \sigma_x^* \sigma_y^*} F . \quad (4.2)$$

The transverse beam profiles and its Gaussian width can be measured by performing van der Meer (vdM) scans [98]. The beam emittance ϵ , which is the phase space volume of the colliding particle bunches, is not conserved due to the acceleration of the beam. It is normally used in its normalised form $\epsilon_n = \beta \gamma \epsilon$ which takes the acceleration into account and hence does not change as a function of energy. It relates to the beam width at the collision point by making use of the value of the *beta function* at the interaction point denoted by β^* with $\sigma^* = \sqrt{\beta^* \epsilon}$. Using these two definitions yields the expression

$$\mathcal{L} = \frac{N_1 N_2 k_b f \gamma}{4\pi \epsilon_n \beta^*} F \quad (4.3)$$

for the luminosity. Each of these quantities, except for the crossing angle and correction factor F , are parameters of the accelerator and can be changed for each fill. Over time the collisions of the particle bunches degrade the number of particles per bunch and consequently the luminosity. The other parameters usually stay constant for each fill. The number of events of a certain process with cross section σ produced in the collisions can now be obtained by integrating over the time dependent luminosity

$$N = \sigma \int \mathcal{L} dt , \quad (4.4)$$

where $\int \mathcal{L} dt$ is referred to as *integrated luminosity*.

4.2.1. Commissioning and Performance of the LHC

The Incident

The LHC machine started its operation in 2008 and first proton beams were successfully circulated on September 10th of that year. However on September 19th a faulty electrical connection with too high resistance between two superconducting wires, driving the magnetic fields of a dipole and quadrupole magnet, triggered a safety energy discharge of the current [99]. The fast power abort induced electrical noise in the quench detectors provoking several magnet quenches of close-by magnets. The surrounding liquid helium was heated up and recovered through self actuated relief valves. An electrical arc, caused by the high voltage, damaged the helium enclosure of the cooling system. As a consequence, helium was released to the insulation vacuum of the cryostat, which degraded the insulation and beam vacuum in the neighbouring sectors. Although the helium was released into the tunnel via spring-loaded relief discs when reaching atmospheric pressure, the pressure in the vacuum enclosure continued to rise. This resulted in large pressure forces acting on the vacuum barriers between neighbouring sectors, which displaced and damaged more than 50 superconducting magnets, their support structures and the beam pipe in those sectors [100]. The helium, cooling these magnets, was rapidly released to the tunnel and triggered an emergency stop shutting down all electrical systems. Until the electrical power could be restored to close the helium valves a total of about six tonnes of helium were lost. The repair and replacement of the damaged magnets and their support structures as well as the installation of additional safety measures aiming to prevent similar incidents in the future took over a year.

Run I

After a successful restart of the machine’s operation, the first proton-proton collisions at the injection energy of 450 GeV were reported on November 23th in 2009. Only a few days later the LHC set a new world record colliding two proton beams with a centre-of-mass energy of $\sqrt{s} = 2.1$ TeV. After ensuring that the magnets manage currents that allow higher beam energies, the physics program started at a centre-of-mass energy of $\sqrt{s} = 7$ TeV in spring 2010. During that year a total integrated luminosity of 48.1 pb^{-1} was delivered by the machine [101]. At that time, the machine was filled with a maximum of 368 bunches with a spacing of 150 ns and maximal 1.2×10^{11} protons per bunch leading to a peak luminosity of $2.1 \times 10^{32} \text{ cm}^{-2}\text{s}^{-1}$ (see Tab. 4.1). The run was continued in the beginning of 2011 at the same energy but with

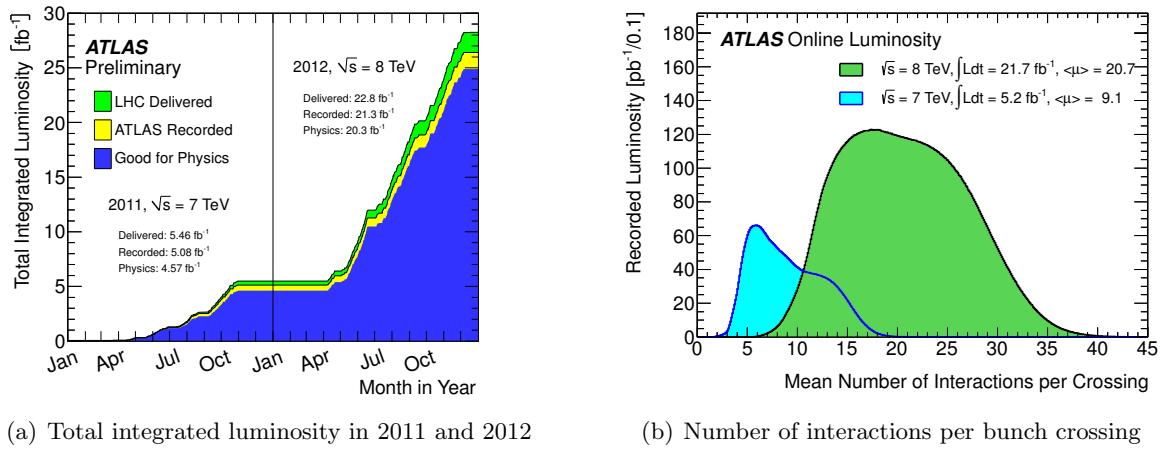
Table 4.1.: A summary of the machine parameters and peak performance of the LHC during 2010, 2011 and 2012. The design parameters and performance are shown for comparison [102, 103].

	2010	2011	2012	design
beam energy [TeV]	3.5	3.5	4	7
bunch spacing [ns]	150	50	50	25
number of bunches	368	1380	1380	2808
$\beta^*(\text{ATLAS})$ [m]	3.5	1.0	0.6	0.55
norm. emittance [mm.mrad]	2.0	2.4	2.5	3.75
protons per bunch	1.2×10^{11}	1.45×10^{11}	1.7×10^{11}	1.15×10^{11}
peak luminosity [$\text{cm}^{-2}\text{s}^{-1}$]	2.1×10^{32}	3.7×10^{33}	7.7×10^{33}	1.0×10^{34}

a reduced bunch spacing and β^* as well as increased emittance and bunch intensity, which allowed for greatly increased peak luminosity of $3.7 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ [102]. The run lasted until the end of the year and the machine was able to deliver a total integrated luminosity of 5.46 fb^{-1} [101]. The evolution of the delivered (green) and recorded (yellow) integrated luminosity is shown in Figure 4.1 (a).

In 2012 it was decided to increase the energy to 8 TeV in order to enhance the discovery potential for a SM Higgs boson. In addition, the peak luminosity could be more than doubled to $7.7 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ being already close to the design luminosity. This was possible, despite the fact that the bunches were still separated by twice the design space, due to a much higher bunch intensity than originally anticipated in the design [103]. The long run resulted in an integrated luminosity of 22.8 fb^{-1} produced by the machine of which 21.3 fb^{-1} was recorded by ATLAS (yellow). This corresponds to a data taking efficiency of roughly 93.4%. The difference is mainly due to losses of the data acquisition (DAQ) system and the *warm-start* time, which is needed to ramp up the tracking system of the detector. Of the recorded data not all can be used for physics analysis, because certain parts of the detector might malfunction or might be misconfigured and are hence not ready during stable beams. The performance of each sub-detector part is analysed and defects of different level of tolerance are set for certain run periods in case the performance is worse than expected. These defects are then translated in a so-called *Good Run List* of different quality level. In this analysis only the data with the highest quality level referred to as ALLGOOD is used and shown as a blue histogram in Figure 4.1 (a). This corresponds to an integrated luminosity of 20.3 fb^{-1} or an overall detector efficiency of roughly 95.3%. The dataset at 8 TeV taken between April and October of 2012

4. Experimental Data, Physics Objects and Processes



(a) Total integrated luminosity in 2011 and 2012

(b) Number of interactions per bunch crossing

Figure 4.1.: (a) Cumulative luminosity versus time delivered (green) to and recorded (yellow) by ATLAS and certified to be of good quality (blue) during pp collisions at 7 and 8 TeV in 2011 and 2012. (b) The number of interactions per crossing weighted by the luminosity for the 2011 and 2012 data proton-proton collisions. The average number of interactions per crossing calculated for each bunch. The number shown in the figure corresponds to the mean assuming a poisson distribution [101].

is used for this search and will be referred to as the *2012 dataset* or *8 TeV dataset* in the following.

Over the years, different beam parameters were used, which resulted in a variation of the number of interactions μ per bunch crossing, as shown in Figure 4.1 (b). At 7 TeV, an average number of interactions per bunch of $\langle\mu\rangle = 9.1$ was measured, corresponding to the mean value of the Poisson distribution. The distribution shows a rather sharp peak around 6 and a broader tail to higher numbers due to the multiple changes in the beam conditions. For the 8 TeV dataset the average number of interactions $\langle\mu\rangle$ increased to 20.7 having a rather broad and more symmetric distribution around that value. The increased number of interactions per bunch crossing also results in a higher number of energy deposits in the detector, which are not originating from the hard scattering process of interest. This *pile-up* can become an issue when defining physics objects (see Sec. 4.3), if not identified and treated accordingly. One distinguishes between *in-time pile-up*, where additional interactions in the same bunch crossing cause the energy deposits, and *out-of-time pile-up*, where the limited read-out time of the detector causes an overlay of interactions of different bunch crossings. The in-time pile-up effects are enhanced when the number of protons in a bunch is increased or the beam is more focused, while out-of-time pile-up contributions are larger when the bunch spacing is reduced. After a short period of heavy ion collisions in 2013, the machine went into a planned shut-down period until the beginning of 2015 in order to prepare the accelerator and the detectors for the design energy and instantaneous luminosity.

4.3. Object Reconstruction

Particles created in the interactions of the colliding proton beams and during the decay of other particles interact differently with the detector. Each particle leaves a characteristic signature inside the various subsystems of the detector. Special object reconstruction and identification algorithms are used to define physics objects such as jets, electrons and muons

from the measured energy deposits and tracks. Since neutrinos leave the detector unseen, their presence is inferred from an imbalance of the total transverse momentum of the entire event.

4.3.1. Jets

Jets are reconstructed from topologically connected calorimeter cells, *topo-cluster* [104, 105], using the *anti- k_t algorithm* [106, 107]. The topo-cluster formation starts from a seed cell with a significant signal-to-noise ratio, where the noise can originate from electronic or pile-up effects. Cells around that seed with at least half of the signal-to-noise ratio of the seed are iteratively included until a cluster is formed containing all significant energy deposits. In order to reduce boundary effects, all neighbouring cells to this cluster are added as well to define the final topo-cluster. One topo-cluster can be split again into several topo-clusters, if the cluster contains more than one local maximum of energy deposits with a certain threshold in order to separate overlapping showers from close-by particles. Cells inside the topo-cluster are calibrated using the *local cluster weighting (LCW) method* [108, 109], which distinguishes between electromagnetic and hadronic topo-clusters based on their energy density and longitudinal shower depth. Energy corrections obtained from simulated electromagnetic or hadronic showers are applied to take non-compensation effects, dead material and out-of-cluster leakage into account. The energy of the topo-cluster is the sum of all deposited energy in the cells, the mass is zero and the direction is calculated from the weighted averages of the directions of the single cells. The calibrated calorimeter clusters are then fed into the *anti- k_t* jet finding algorithm with a distance parameter of $R = 0.4$ utilizing the FASTJET [110] software package. This jet reconstruction is known to be independent of collinear splitting of the initial parton (*collinear safe*) and additional soft gluon radiation (*infra-red safe*). The total four-momentum of the jet is calculated summing the four-momenta of all its constituents. This calorimeter jet is then calibrated in order to take several experimental and detector effects into account. Additional pile-up interactions cause an offset of the jet energy, which is corrected by applying scale factors (SFs) derived from MC simulation as a function of the number of reconstructed primary vertices and the average of the expected number of interactions in bins of the jet η and p_T . The direction of the jet is corrected to originate from the primary event vertex. To account for energy losses inside the detector, jet energy scale (JES) corrections [104] are applied to the energy and pseudorapidity of the jet by comparing reconstructed jets to truth jets formed from stable particles in MC simulations. The response of LCW jets of different energy as a function of the detector pseudorapidity, from which the JES corrections are calculated, is shown in Figure 4.2 (a).

The JES corrections are measured using a variety of in-situ techniques exploiting the transverse momentum balance of the jet in di-jet events or to well measured reference objects like a Z boson or a photon. SFs are applied to MC events to correct for a different response in data. After the energy calibration jets are required to have $p_T > 25$ GeV and $|\eta| < 2.5$. Since semi-leptonic b -hadron decays inside jets cause losses due to the escaping muon and neutrino, corrections are applied if a jet overlaps within $\Delta R < 0.4$ with a muon by adding the energy of the measured muon to the jet. The quality of the jet is further increased by applying several cuts, which cope with energy deposits caused by non-collision background such as cosmic rays, beam-gas or beam-halo events and noise in the calorimeter. The efficiency of these cuts is close to 100% [104]. Jets from multiple pp interactions, e.g. pile-up events, tend to be rather soft and are rejected by cutting on the *jet vertex fraction (JVF)* [111, 112], which exploits the fraction of tracks within the jet coming from the primary event vertex. Any jet with $p_T < 50$ GeV and

4. Experimental Data, Physics Objects and Processes

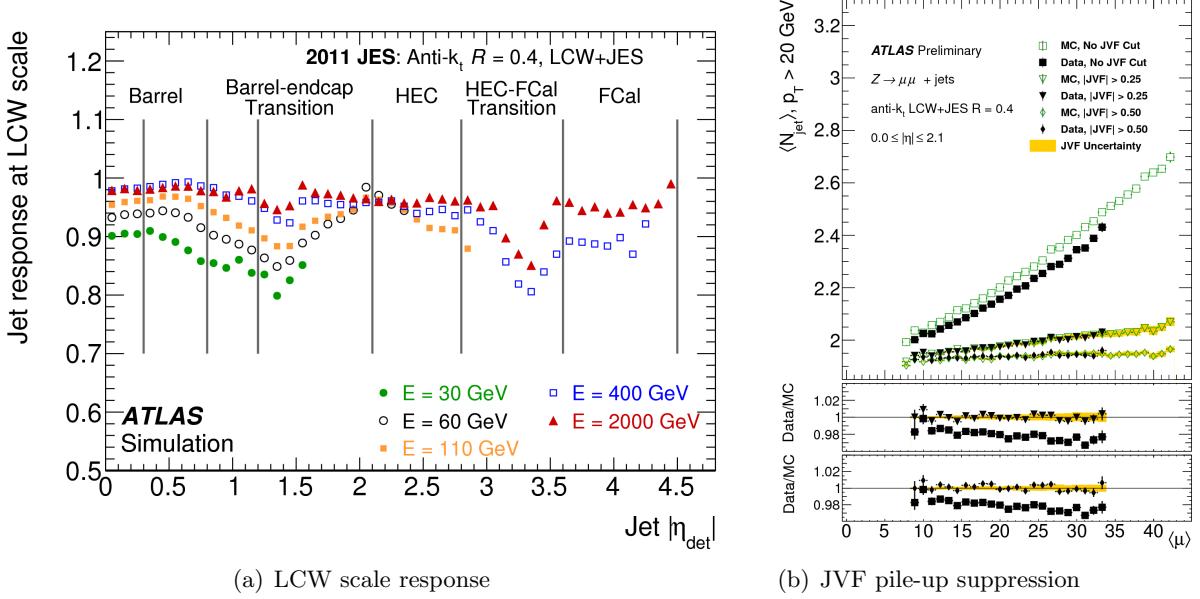


Figure 4.2.: (a) Average response of simulated (PYTHIA) jets with various truth energies for the LCW scale as a function of the detector η . The inverse corresponds to the JES correction in each η bin [105]. (b) Mean jet multiplicity ($p_T > 20 \text{ GeV}$) versus the average number of interactions per bunch crossing $\langle \mu \rangle$. Without the JVF cut (rectangle), the number of jets rises constantly with the number of interactions due to jets caused by pile-up. After applying the JVF cut of 0.5 as described above, there is almost no dependency on the number of interactions which means that almost all jets originating from pile-up are removed. Since the efficiencies of the JVF cut differ for MC simulation (green empty marker) and data (black filled marker), correction SFs are derived, which are then applied to MC simulations. No distinction is made between electrons and jets during the jet reconstruction, which is why electrons will always be reconstructed as jets. In order to avoid these jets, the single closest jet with $\Delta R < 0.2$ to an electron is removed.

$|\eta| < 2.4$ is required to have at least 50% of its summed track p_T of all tracks with $p_T > 1 \text{ GeV}$ originating from the primary vertex. The effect of the JVF cut is shown in Figure 4.2 (b), where the average multiplicity of jets with $p_T > 20 \text{ GeV}$ in $Z + \text{jets}$ events is shown dependent on the average number of interactions per bunch crossing $\langle \mu \rangle$. Without the JVF cut (rectangle), the number of jets rises constantly with the number of interactions due to jets caused by pile-up. After applying the JVF cut of 0.5 as described above, there is almost no dependency on the number of interactions which means that almost all jets originating from pile-up are removed. Since the efficiencies of the JVF cut differ for MC simulation (green empty marker) and data (black filled marker), correction SFs are derived, which are then applied to MC simulations. No distinction is made between electrons and jets during the jet reconstruction, which is why electrons will always be reconstructed as jets. In order to avoid these jets, the single closest jet with $\Delta R < 0.2$ to an electron is removed.

B-jets

Jets originating from hadronised b -quarks, also known as b -jets can be identified as such by an algorithm which utilises their unique properties. This b -jet identification is referred to as *b-tagging*. B -mesons have a relatively long lifetime of 1.5 ps, which allows them to travel several millimetres inside the detector before they decay. For example, a typical b -hadron with a transverse momentum of 70 GeV would have an average transverse decay length $\langle L_{xy} \rangle = \beta \gamma c \tau$ of 6.4 mm. Therefore, their decay vertex is considerably displaced from the events' primary interaction vertex and can be identified as a secondary vertex in a high precision tracking system such as the Pixel detector. On top of that, b -mesons are much heavier than hadrons formed from light quarks (up, down, strange) and thus their decay products tend to have a

higher momentum perpendicular to the flight direction of the b -meson which causes broader showers, higher particle multiplicities inside the jet and higher total invariant masses of the jet. Because of their high mass they can decay into muons, which can be detected in the outermost MS and be used to identify b -jets. Modern b -tagging algorithms make use of these differences by combining many variables reflecting the jet properties into one single output (*b -tag weight*) representing how likely a jet has originated from a b -quark. For this analysis, the output from three relatively simple algorithms [113], one based on secondary vertices (SV1), one using impact parameters (IP3D) and one exploiting the difference between weak b - and c -hadron decays (JETFITTER), are combined using an artificial neural network into the so called MV1 algorithm in order to separate b -, c - and light flavour (LF) jets most efficiently. This b -tagging algorithm is calibrated at several values of thresholds (*working points*) chosen to obtain specific b -tagging efficiencies. The working point corresponding to a 70% efficiency of tagging a b -quark jet as a b -jet uses a b -tag weight cut value of $w > 0.7892$ for LCW calibrated jets and was chosen in this analysis. The *rejection factor*, defined as the inverse of the mistagging rate, is 137 for light jets (u -, d -, s -quark, gluon), 13 for taus and 5 for c -quark jets. This leads to a purity of roughly 92% as determined for b -tagged jets with $p_T > 20$ GeV and $|\eta| < 2.5$ in simulated $t\bar{t}$ events [114]. In order to account for a different performance of the algorithm between recorded data and MC simulation, correction SFs depending on the jet transverse momentum are calculated selecting dileptonically decaying $t\bar{t}$ events with two oppositely charged leptons and exactly two jets in the final state [114]. For the calculation of the b -tagging efficiency a probabilistic approach is used (*PDF calibration*) defining an unbinned likelihood function containing parton distribution functions (PDFs) for the flavour of the jets, each depending on the jet momenta and covering all possible combinations of jet flavour. This method allows to significantly reduce the uncertainties on the SFs. The b -jet efficiencies and SFs obtained by combining all dilepton decays and using events with two and three jets are shown as a function of the jet momentum in Figure 4.3.

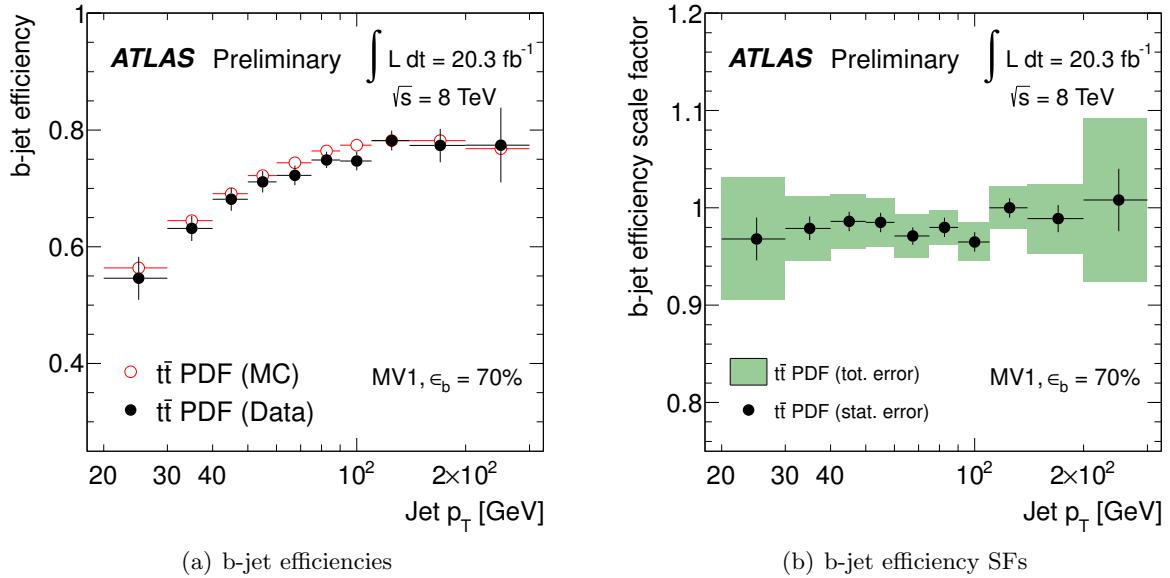


Figure 4.3.: The b -jet tagging efficiencies and SFs combining $e^\pm\mu^\mp$ and $e^+e^- + \mu^+\mu^-$ in the two and three jet bin for the 70% MV1 working point including statistical and systematic errors [114].

4. Experimental Data, Physics Objects and Processes

Since the b -tagging algorithm tags not only b -jets, but also mistags light jets and c -jets, these jets need to be calibrated as well by measuring the light jet and c -jet tagging efficiencies and corresponding SFs to account for the differences in MC simulation and data [115]. For the estimation of the c -jet tagging efficiency D^{*+} mesons containing charm quarks are reconstructed from kaons and pions matching the mass hypotheses and charge combination of the $D^{*+} \rightarrow (D^0 \rightarrow K^-\pi^+) \pi^+$ decay. The reconstructed D^{*+} meson is then associated with a jet close to its direction and the tagging efficiency for this jet is measured after subtracting background contributions and jets with beauty content. The tagging efficiency estimation for jets containing a D^{*+} meson is then extrapolated to inclusive charm jets using MC simulations. The c -tagging efficiencies in data and simulation and the resulting SFs for the 70% working point of the MV1 tagger are shown in Figure 4.4.

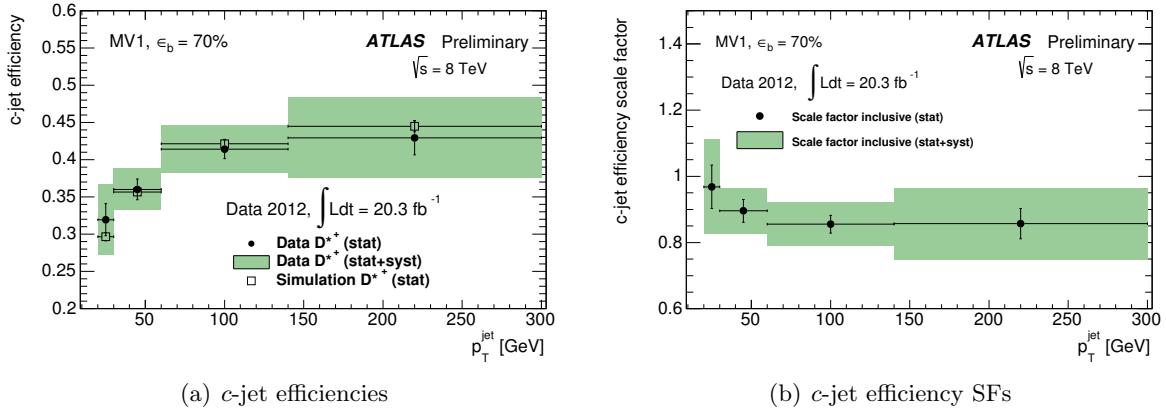


Figure 4.4.: The c -jet tagging efficiencies for jets containing D^{*+} mesons and the efficiency SFs for inclusive charm jets for the 70% MV1 working point including statistical and systematic errors [115].

The mistagging efficiency of light jets is calculated using the *negative tag* method, which defines a negative version of the tagging algorithm by internally reversing all discriminating b -tagging variables. The mistag rate can then be calculated by applying the same tag weight criteria. It needs to be corrected to take into account the finite detector resolution effects. Because of different track resolutions in the central and more forward regions of the tracking system, the light jet tag efficiencies and SFs are calculated for two η regions as shown in Figure 4.5.

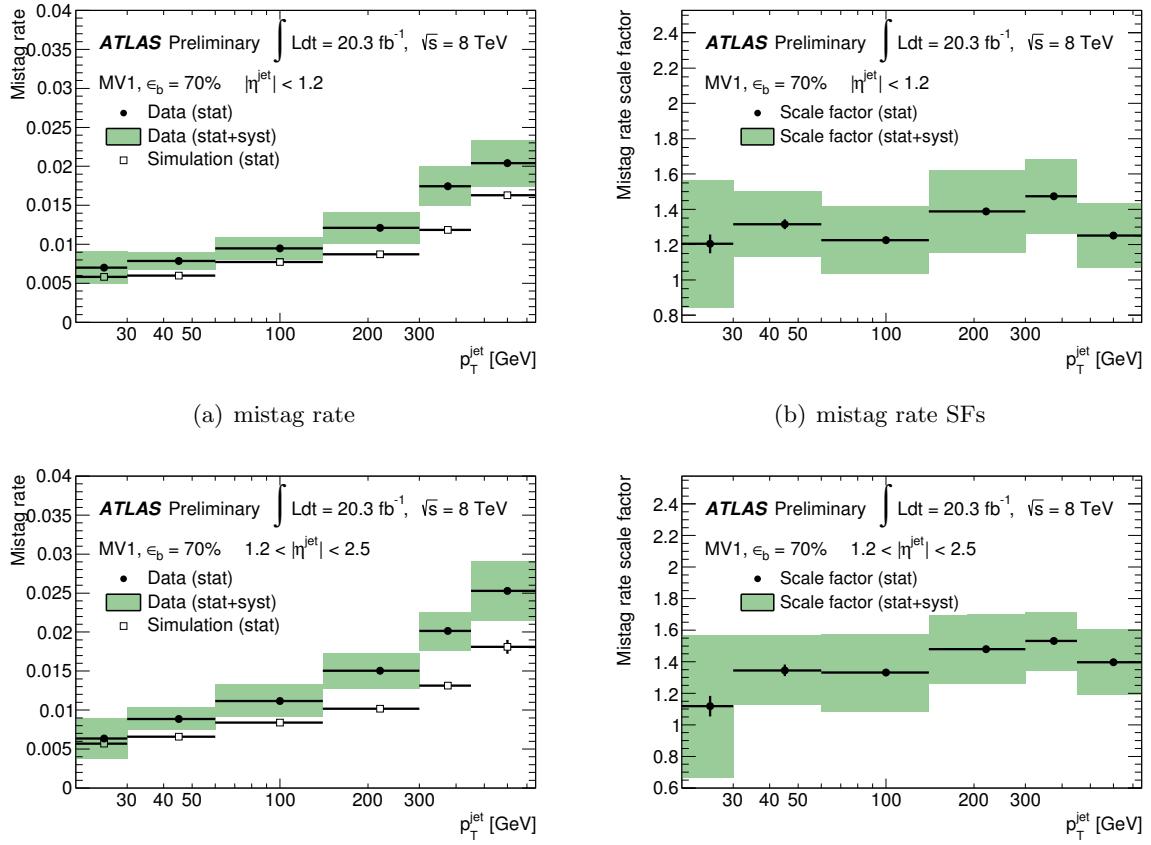


Figure 4.5.: The light jet tagging efficiencies and SFs for the 70% MV1 working point including statistical and systematic errors [115].

4.3.2. Electrons

Electron candidates [116–118] are built from energy deposits (clusters) in the central region of the electromagnetic calorimeter that are associated with reconstructed tracks in the ID. An electromagnetic cluster is reconstructed by searching for seeds above a transverse energy of 2.5 GeV with a fixed sized *sliding window* [119] spanned in the $\eta \times \phi$ plane. The direction of the window, which maximizes the covered energy, is taken as the seed direction and is required to match loosely to at least one reconstructed track of the ID. From this seed, an electron cluster is built from longitudinal towers of calorimeter cells with different sizes in the barrel and end-caps. The energy of the electron is calculated from the measured energy deposits inside the cluster taking losses from dead material in front of the calorimeter as well as lateral and longitudinal leakage into account by applying a correction factor, which is determined using a multivariate algorithm. The direction and charge of the electron is taken from the primary track in the ID (except for TRT-only tracks), which is defined as the closest track with electron track properties and is corrected for losses due to Bremsstrahlung. Only clusters with a total transverse energy of $E_T > 25$ GeV, a cluster direction of $|\eta_{\text{cluster}}| < 2.47$ and outside the transition region of the calorimeter ($1.37 < |\eta_{\text{cluster}}| < 1.52$) are considered in this analysis. The electron reconstruction efficiency is determined with the tag-and-probe (TP) method using $Z \rightarrow ee$ decays and varies between 95–99% for different η and E_T ranges [117].

In order to reject background objects such as non-prompt electrons and jets when identifying electrons, reconstructed candidates are classified into different quality levels. This is achieved by applying cuts parametrised in $|\eta|$ and E_T using calorimeter and tracking variables such as the shower shape, hadronic leakage and number of hits in the tracking system. The cut values are chosen to keep the signal efficiencies and background rejections constant under different pile-up conditions. The `tight++` selection, which has the highest purity and is chosen for this analysis, discriminates against photon conversion and makes use of the particle identification of the TRT. This selection provides the best separation from hadron decays, photons and jets misidentified as electrons, while obtaining signal efficiencies between 60-80% depending on the $|\eta|$ and E_T region. The efficiencies are calculated using the TP method in $Z \rightarrow ee$ and $J/\psi \rightarrow ee$ decays [117]. Correction factors are derived from these events to account for the different performance in MC and data depending on η and E_T , but they are generally close to one, except for low E_T and high η regions.

Backgrounds from non-prompt electrons and hadronic jets are further reduced by requiring isolation criteria of the electron candidate. First, calorimeter based isolation is required by selecting only electrons with a maximal sum of transverse energy deposited in the calorimeter cells in a cone of $\Delta R = 0.2$ around the electron. The energy is corrected for leakage from the electron to the isolation cone and for pile-up effects. Second, a track based isolation is applied by requiring a maximal sum of transverse momentum of the tracks in a cone of $\Delta R = 0.3$ around the electron track. Only tracks, which originate from the primary vertex and of good quality, are considered. Both isolation requirements have an efficiency of 90% each.

If a jet with $p_T > 25$ GeV and $|JVF| > 0.5$ is found within the range of $0.2 \leq \Delta R < 0.4$ to the electron, it is likely that the electron originates from a hadronic decay inside the jet and is thus removed. To ensure that the electron was created inside the interaction of interest, the longitudinal impact parameter of the electron track with respect to the selected event primary vertex z_0 is required to be less than 2 mm.

4.3.3. Muons

For the reconstruction of muons [120] track information from the ID and the MS is used in different ways leading to several muon types. For this analysis the combined muon (CB) type is used, for which first tracks are reconstructed independently of the ID and MS and are then combined to obtain a muon track, which defines its four-momentum vector. First, the tracks in the MS are reconstructed by searching each layer of the chambers for local track segments, which are then combined into an MS track using the track information of all layers. A track reconstructed in the ID is required to have a certain number of Pixel, SCT and TRT hits and at most two active Pixel or SCT sensors not responding to the traversing muon track candidate. If the MS track can be matched to an ID track, the muon track is obtained by a global refit of all hits assigned to the two tracks in the ID and MS. The energy and direction of the muon candidate is taken from this track. In this analysis only muons with $p_T > 25$ GeV and $|\eta| < 2.5$ are considered. The CB muons have the highest purity and yield reconstruction and identification efficiencies, obtained with the TP method in $Z \rightarrow \mu\mu$ and $J/\psi \rightarrow \mu\mu$ events, close to 99% almost independent of the $|\eta|$ or p_T region [120]. The performance in MC is calibrated to data by correcting the muon momentum scale and resolution using SFs calculated from $Z \rightarrow \mu\mu$, $J/\psi \rightarrow \mu\mu$ and $\Upsilon \rightarrow \mu\mu$ decays.

In order to suppress background from non-prompt muons, such as decays from heavy flavour hadrons, muons are required to satisfy momentum dependent and track based isolation cri-

teria [121]. If the scalar sum of the transverse momentum of the tracks in a cone of $\Delta R < 10 \text{ GeV}/p_T^\mu$ around the muon track is less than 5% of the muon p_T , the muon candidate is rejected. This isolation is designed to perform independently of the pile-up conditions and the muon momentum and yields efficiencies of roughly 97%. If the muon is found to be close to a jet within a range of $\Delta R < 0.4$, the muon is discarded to avoid muons produced by HF decays inside jets. By requiring the longitudinal impact parameter of the muon track with respect to the primary vertex z_0 to be less than 2 mm, the muon originating from the considered interaction is selected.

4.3.4. Missing Transverse Energy

The total transverse momentum of all particles escaping the detector can be indirectly measured by implying momentum conservation of the incoming and outgoing particles. In good approximation, the transverse momenta of the interacting initial partons of a hadron collider can be neglected. Therefore the missing transverse energy E_T^{miss} of an event is defined as the magnitude of the negative vector sum of the momenta of all reconstructed and calibrated objects plus additional correction terms from the calorimeter and tracking. In the reconstruction of the E_T^{miss} , energy deposits and tracks of electrons, photons, hadronically decaying τ leptons, jets and muons as well as soft calorimeter topo-cluster and tracks (SoftTerm), not associated to any high- p_T object, are taken into account [122]. The x - and y -component of the E_T^{miss} is calculated by

$$E_{x(y)}^{\text{miss}} = E_{x(y)}^{\text{miss},e} + E_{x(y)}^{\text{miss},\gamma} + E_{x(y)}^{\text{miss},\tau} + E_{x(y)}^{\text{miss,jets}} + E_{x(y)}^{\text{miss,SoftTerm}} + E_{x(y)}^{\text{miss},\mu}, \quad (4.5)$$

where each term is the negative sum of all object energies projected to the x - or y -direction. The E_T^{miss} is then given by

$$E_T^{\text{miss}} = \sqrt{(E_x^{\text{miss}})^2 + (E_y^{\text{miss}})^2}. \quad (4.6)$$

However, this definition does not directly represent the momentum imbalance of the hard scattering event, because it is compromised by an imperfect detector with limited coverage, finite resolution, electronic noise and dead material. Furthermore it is affected by tracks or energy deposits, which do not arise from the event under study, such as multiple pp interactions, cosmic rays, beam-gas or beam-halo interactions. Pile-up contributions, in particular to the jet and soft term, can be suppressed by applying cuts based on track properties or the area of the objects [122].

4.4. Modelling of the Physics Processes

The majority of analyses in high energy physics relies on a good modelling of the involved physics processes in order to make comparisons between theoretical predictions and experimental data. Due to the probabilistic nature of the physics process described by quantum field theories, MC simulations are usually used for the modelling. However, sometimes it is not easy to find a good model or the model depends on too many unknown parameters, which results in inaccurate descriptions of the data or the simulation is computationally prohibitive. In these cases the physics processes are estimated in well known phase space regions from data and are then extrapolated to the phase space region of interest. There is a large variety of these *data-driven* techniques, but each of them have their limitations. The best choice depends strongly on the physics process and there is no general recipe aiming for a good modelling.

4.4.1. Event Simulation

According to the *factorisation theorem* [123], the generation of an event can be divided into several independent sub-processes if they occur at well separated energy scales. All of these different stages are illustrated in Figure 4.6.

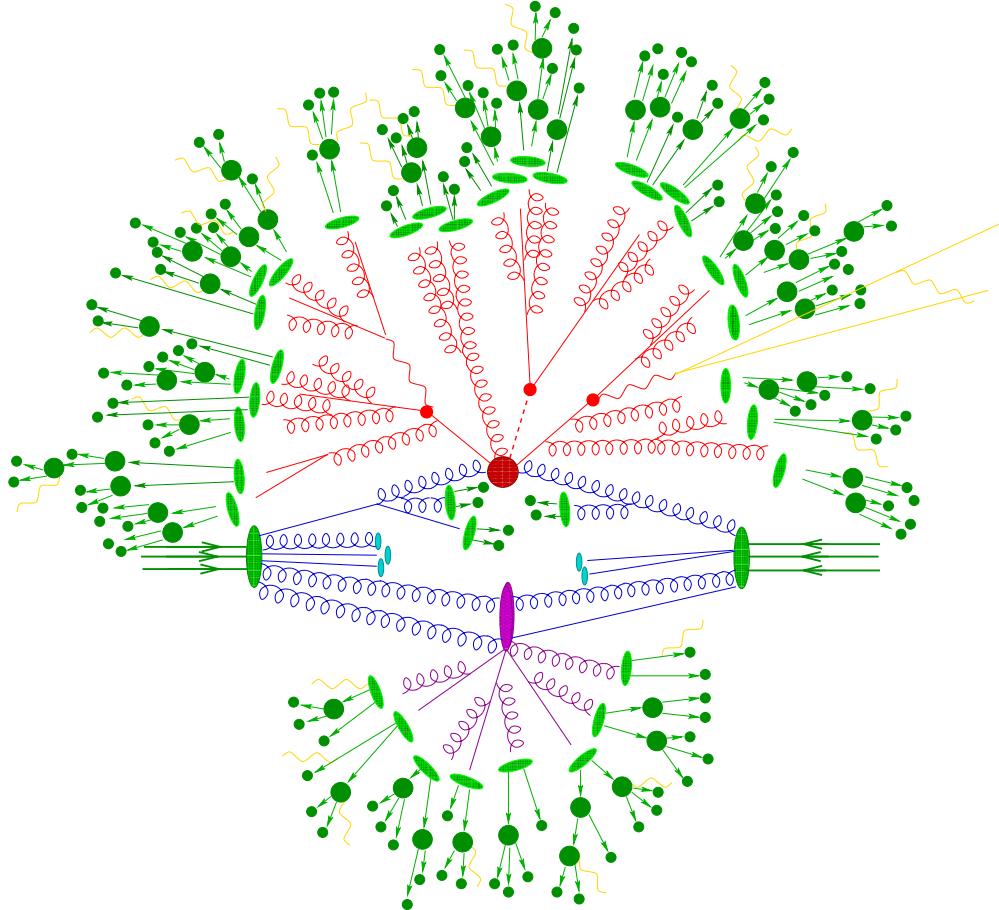


Figure 4.6.: A schematic overview of a $t\bar{t}H$ event illustrating the different steps of a MC event generation [124].

At first, the momenta of the hard scattering partons inside the colliding protons (blue) are obtained by sampling the PDFs at the energy scale of the process. The kinematic and flavour configuration of the incoming partons is then used to evaluate the matrix element of the hard scattering process (red blobs), which describes the physics process; in this case a $t\bar{t}H$ production. By integrating over the phase space of the outgoing particles the cross section and a weight at each evaluated phase space point representing the likelihood of the kinematic configuration can be obtained. The set of the momenta of the involved particles at that phase space point can then be used to represent a simulated event. Specialised matrix element generators used by the ATLAS collaboration for many hard processes are ALPGEN, MADGRAPH or POWHEG [125–127]. In a second step, all accelerated coloured partons radiate gluons (red curly lines), which can further split again into gluons or quark-antiquark pairs and iteratively create a *parton shower* until a certain energy scale is reached. This radiation corresponds to a higher order correction to the matrix element calculation and simulates initial and final

state radiation (ISR/FSR). As a next step, called *hadronisation* or *fragmentation*, the partons in the shower, which have reached non-perturbative energy scales, form colourless hadrons (green ovals). If these hadrons are unstable, they further decay into stable and observable particles (green blobs). Typically programs like PYTHIA or HERWIG [128–130] are used by the ATLAS collaboration for the simulation of the parton showering and hadronisation. PYTHIA and HERWIG differ in particular in the fragmentation model using either string or cluster fragmentation [131, 132], respectively. The last step in the event generation of a hard process is the simulation of the so-called *underlying event (UE)* (violet) usually performed by PYTHIA or alternatively JIMMY [133] in case HERWIG is used for the fragmentation modelling. It describes possible interactions between the proton remnants, contributions from initial state radiation (ISR) and final state radiation (FSR), multiple parton interaction (MPI) or in-time pile-up. Since some collisions do not produce an identifiable hard process, but still create some activity in the detector, these *minimum bias events* need to be simulated. They are overlaid with the hard process.

Finally, all stable particles on a time scale allowing them to reach the detector are processed through a full ATLAS detector simulation [134] using GEANT4 [135], which is a general framework for simulating interactions of particles with matter. In case of the $t\bar{t}H$ and $t\bar{t}$ process a fast simulation [136] of the calorimeter response is used in addition to the full simulation to increase the statistics. The simulated detector response is then used in the same way as data in order to reconstruct physics objects as described in Sections 4.3.1 – 4.3.4.

4.4.2. Modelling of the Signal

For the modelling of the $t\bar{t}H$ signal process, events are generated using matrix elements with NLO QCD accuracy obtained from the HELAC-ONELOOP package [137]. In order to interface the generated events to a parton shower program, the events are passed through POWHEG Box [127, 138, 139]. Event samples generated in this combination are also referred to as PowHEL samples. All Higgs boson decays predicted by the SM are simulated for a range of Higgs boson masses of $m_H = 115, 120, 125, 130, 135$ GeV and a fixed top quark mass of $m_{\text{top}} = 172.5$ GeV. The $t\bar{t}H$ production cross sections used for the normalisation of the samples and the Higgs boson decay branching fractions are taken from the NLO theoretical calculations [140–147], collected in Ref. [64]. The events are weighted with the CT10NLO PDF set [148] and the factorisation μ_F and renormalisation μ_R scales are set to the same static scale of $\mu_0 = \mu_F = \mu_R = m_{\text{top}} + m_H/2$. The parton shower and hadronisation is performed using PYTHIA 8.1 [129] with the CTEQ61L PDFs and the AUET2 tune to model the underlying event [149, 150].

4.4.3. Monte Carlo Simulated Background

In the phase space of interest, the main background for the $t\bar{t}H$ production in the single lepton decay channel arises from $t\bar{t}+\text{jets}$ events. Significantly smaller contributions come from single top quark production followed by vector boson production in association with jets ($W/Z+\text{jets}$), contributions from diboson ($WW/WZ/ZZ$) processes and top quark pair production in association with vector bosons ($t\bar{t}V$). The modelling of all of these processes relies on MC simulation. An overview of the matrix element generators, PDFs, parton shower programs and the precision of the cross section normalisations used in this analysis for the various physics processes is given in Table 4.2.

4. Experimental Data, Physics Objects and Processes

Table 4.2.: A summary of the various physics processes and the used matrix element generators, PDFs, parton shower programs and the normalisation precision.

Process	ME Generator	PDF	Parton Shower	Normalisation
$t\bar{t}H$	HELAC-Oneloop	CT10	Pythia 8.1	NLO
$t\bar{t} + \text{jets}$	Powheg	CT10	Pythia 6.425	NNLO+NNLL
Single top (s-chan., Wt)	Powheg	CT10	Pythia 6.426	aNNLO
Single top (t-chan.)	Powheg	CT10	Pythia 6.427	aNNLO
$t\bar{t}V$	Madgraph	CTEQ6L1	Pythia 6.425	NLO
$W + \text{jets}$	Alpgen	CTEQ6L1	Pythia 6.426	NLO
$Z + \text{jets}$	Alpgen	CTEQ6L1	Pythia 6.426	NLO
Diboson	Alpgen	CTEQ6L1	Herwig 6.520	NLO

In all simulations, photon radiation and τ lepton decays are modelled using PHOTOS 2.15 [151] and TAUOLA 1.20 [152], respectively. The minimum bias events are simulated utilizing PYTHIA 8.1 with the MSTW2008 PDF set in LO precision and the AUET2 tune. They are overlaid on the generated events to simulate pile-up and are required to match the luminosity profile of the data. All reconstructed objects of the simulated events are corrected, such that the object identification efficiencies, energy scales and energy resolutions match those determined from the data.

Top Quark Pair Production

The POWHEG NLO generator with the CT10 PDFs set is used to simulate $t\bar{t} + \text{jets}$ events assuming a top mass of $m_{\text{top}} = 172.5$ GeV. The parton shower, hadronisation and UE simulation is performed with PYTHIA 6.425 [128] using the LO CTEQ61L PDFs and the PERUGIA2011C tune [153] for the modelling of the UE. The factorisation and renormalisation scales are set to $\mu_0 = \sqrt{m_{\text{top}}^2 + p_T^2}$, where m_{top} and p_T are the top quark mass and its transverse momentum, evaluated for the underlying Born configuration (i.e. before radiation). For the normalisation of the sample a theoretical calculation with TOP++2.0 [154] at NNLO including resummation of soft gluon terms with NNLL accuracy is used, which yields a cross section of $\sigma_{t\bar{t}} = 253^{+15}_{-16}$ pb at 8 TeV [45–49].

In the matrix element calculations of the POWHEG generator $t\bar{t} + \text{HF}$ processes are not included and hence are only simulated by the parton shower. For the purpose of distinction between $t\bar{t} + \text{light}$, $t\bar{t} + c\bar{c}$ and $t\bar{t} + b\bar{b}$ events, the inclusive $t\bar{t} + \text{jets}$ events are categorised using the true flavour information of the partons. First, particle jets are reconstructed from stable particles visible to the calorimeter using the anti- k_t clustering algorithm with a radius parameter of $R = 0.4$. Particle jets with $p_T > 15$ GeV and $|\eta| < 2.5$ are then matched within a cone of $\Delta R < 0.4$ to a final state parton, not originating from the $t\bar{t}$ system. If at least one particle jet can be matched to a bottom quark, the event is classified as $t\bar{t} + b\bar{b}$. If this is not the case, but at least one particle jet can be matched to a charm quark, the event is labelled as $t\bar{t} + c\bar{c}$ instead. The remaining events, in which no particle jet can be matched to a b - or c -quark, are categorised as $t\bar{t} + \text{light}$. The $t\bar{t} + \text{HF}$ ($t\bar{t} + c\bar{c}$ or $t\bar{t} + b\bar{b}$) events are further categorised in a finer classification, in order to compare different event generator and to treat modelling systematic

uncertainties related to the flavour. For each match of a single particle jet to an extra single b - or c -quark, the $t\bar{t}$ event is labelled with a b or c extension, respectively. However, if a single particle jet is matched to a $b\bar{b}$ or $c\bar{c}$ pair, the jet is assumed to be unresolved due to gluon splitting with two close-by b - or c -quarks and the $t\bar{t}$ event gets a B or C extension.

The modelling of the irreducible $t\bar{t} + b\bar{b}$ background plays a key role in this search. To understand the effect of neglecting HF partons in the matrix element calculation of the $t\bar{t} + \text{jets}$ process, the POWHEG+PYTHIA events are compared to alternative generators which include massive b and c quarks as additional partons. The first alternative sample is obtained by generating $t\bar{t}$ events with up to three extra partons (u, d, s, c, b) with MADGRAPH 5 [126] in LO precision, using the CT10 PDF set and showering with PYTHIA 6.425. The same scales and parameters as in Reference [70] are used. For the second sample, the SHERPA [155] event generator is combined with NLO matrix element from the OPENLOOPS framework [124, 156], referred to as SHERPAOL. This calculation became only recently available for $t\bar{t} + bb$ and yields fully matched NLO predictions with massive b quarks. The renormalisation scale is set to $\mu_R = \prod_{i=t,\bar{t},b,\bar{b}} E_{T,i}^{1/4}$, where $E_{T,i}$ is the transverse energy of parton i , and the factorisation and resummation scales are set to $(E_{T,t} + E_{T,\bar{t}})/2$. In Figure 4.7 the relative fraction of events in the different $t\bar{t} + b\bar{b}$ categories to the total $t\bar{t} + b\bar{b}$ cross section are compared for the three different event generator.

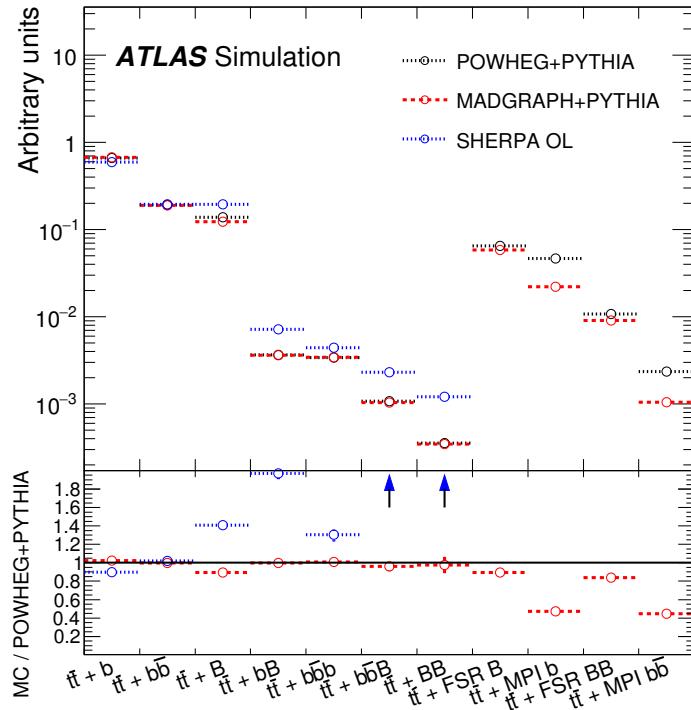


Figure 4.7.: Comparison of the relative contributions to different categories of $t\bar{t} + b\bar{b}$ events using the POWHEG+PYTHIA, MADGRAPH+PYTHIA and SHERPAOL generator. Events produced via MPI and FSR, which are not included in the SHERPAOL calculation, are labelled with $t\bar{t}+\text{MPI}$ and $t\bar{t}+\text{FSR}$, respectively. The error bars only represent the statistical uncertainties [71].

Although $t\bar{t} + \text{HF}$ events are only produced via the parton shower with the POWHEG+PYTHIA generator, the various HF contents are reasonably well reproduced compared to the LO predic-

4. Experimental Data, Physics Objects and Processes

tions of MADGRAPH+PYTHIA and NLO predictions of SHERPAOL generator. However, the SHERPAOL simulation predicts generally more events in the categories labelled with at least one B and in the $t\bar{t} + b\bar{b}$ category and does not model additional b -quarks from MPI or FSR. The yields predicted by MADGRAPH+PYTHIA in the different categories are comparable to the one from POWHEG+PYTHIA, except for categories, in which the additional $b\bar{b}$ pair is produced via MPI or FSR. The shape of the kinematic distributions across all categories is generally in reasonable agreement between POWHEG+PYTHIA and SHERPAOL. However, depending on the category, there are some small differences, for instance in the very low invariant mass and p_T region of the $b\bar{b}$ pair and in the p_T of the top quark and $t\bar{t}$ systems. As an example, the differences in a few kinematic distributions is shown for the $t\bar{t}+B$ category in Figure 4.8.

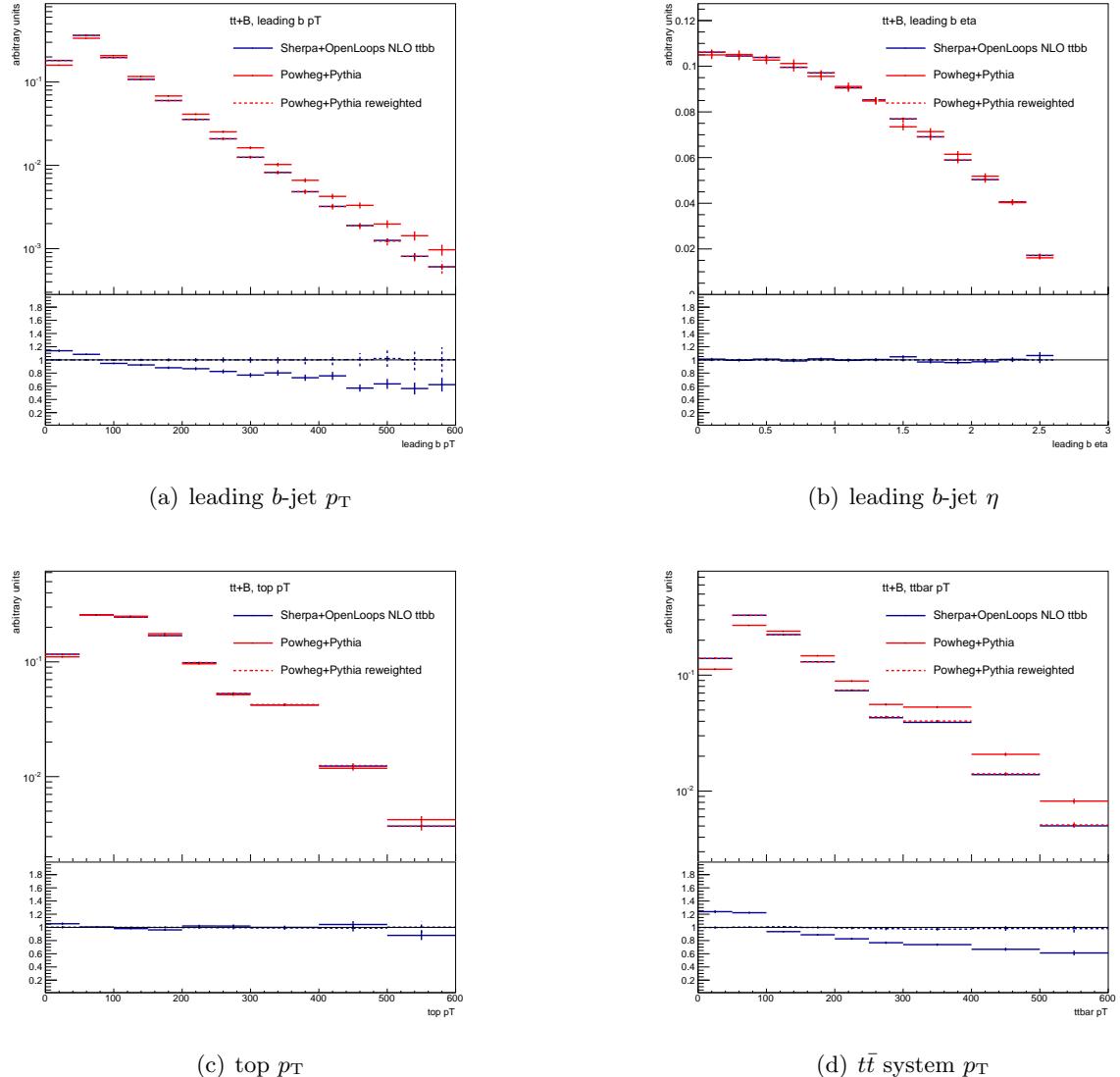


Figure 4.8.: Kinematic distributions showing differences between POWHEG+PYTHIA (solid red) and SHERPAOL (solid blue) in the $t\bar{t}+B$ category. The reweighted POWHEG distributions are also shown as a red dashed line.

Since it is expected that SHERPAOL models the HF contributions more accurately than the other two generators, the relative differences in the relative HF contributions and in the kinematic distributions in each category are used to renormalise POWHEG+PYTHIA to the NLO $t\bar{t} + b\bar{b}$ prediction and to correct the shapes of the kinematic distributions. The inclusive $t\bar{t} + b\bar{b}$ cross section for all categories is kept constant. The reweighting is done at truth level using a number of kinematic variables such as the top quark p_T , $t\bar{t}$ system p_T , the p_T and η of the leading b -jet, invariant mass, ΔR , and p_T of the dijet system not coming from the top quark decay. Comparisons of data to the POWHEG+PYTHIA prediction in the regions dominated by $t\bar{t} + \text{light}$ show that there is a visible disagreement between the predictions and data (see Fig. 4.10). The differential cross section measurement at 7 TeV [157] observed a mismodelling of the top quark and $t\bar{t}$ system p_T as shown in Figure 4.9.

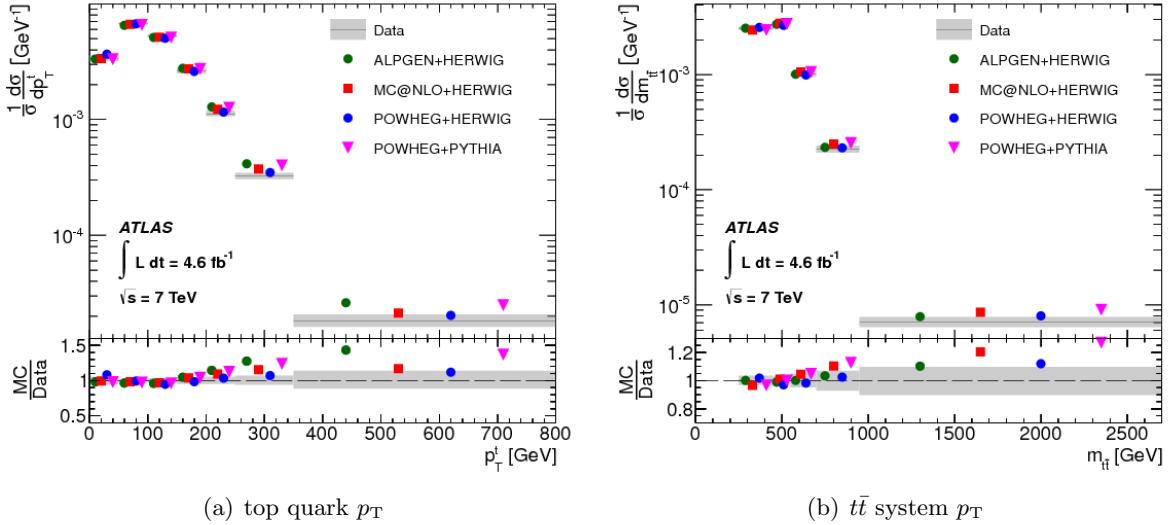


Figure 4.9.: The normalised differential cross sections of the top quark p_T (a) and the p_T of the $t\bar{t}$ system (b) for the generators ALPGEN+HERWIG (green circles), MC@NLO+HERWIG (squares), POWHEG+HERWIG (blue circles) and POWHEG+PYTHIA (triangles). The lower part of each figure shows the ratio of the generator predictions to data. The grey bands indicate the total uncertainty on the data in each bin [157].

In the high p_T regions of the top quark and $t\bar{t}$ system the POWHEG simulation tends to predict too many events, which is not entirely covered by the uncertainties. In order to achieve the best possible modelling of the very important $t\bar{t}$ background, the $t\bar{t} + \text{light}$ and $t\bar{t} + c\bar{c}$ events are reweighted based on this measurement. The following reweighting procedure was proven to be applicable to $\sqrt{s} = 8$ TeV as well. The relative differences in the top quark and $t\bar{t}$ system p_T distributions unfolded to parton level are taken as reweighting SFs. These SFs are then applied sequentially to the parton level POWHEG events taking the correlations between the two parameters fully into account. The $t\bar{t}$ system p_T reweighting helps to improve the MC modelling of the jet multiplicity distribution, while the top quark p_T reweighting recovers mismodelling of kinematic distributions in regions with high transverse momentum. The effect of the reweighting on the jet multiplicity and scalar sum of all jet p_T (H_T^{had}) is shown in Figure 4.10 requiring at least four jets and exactly two b -tags.

4. Experimental Data, Physics Objects and Processes

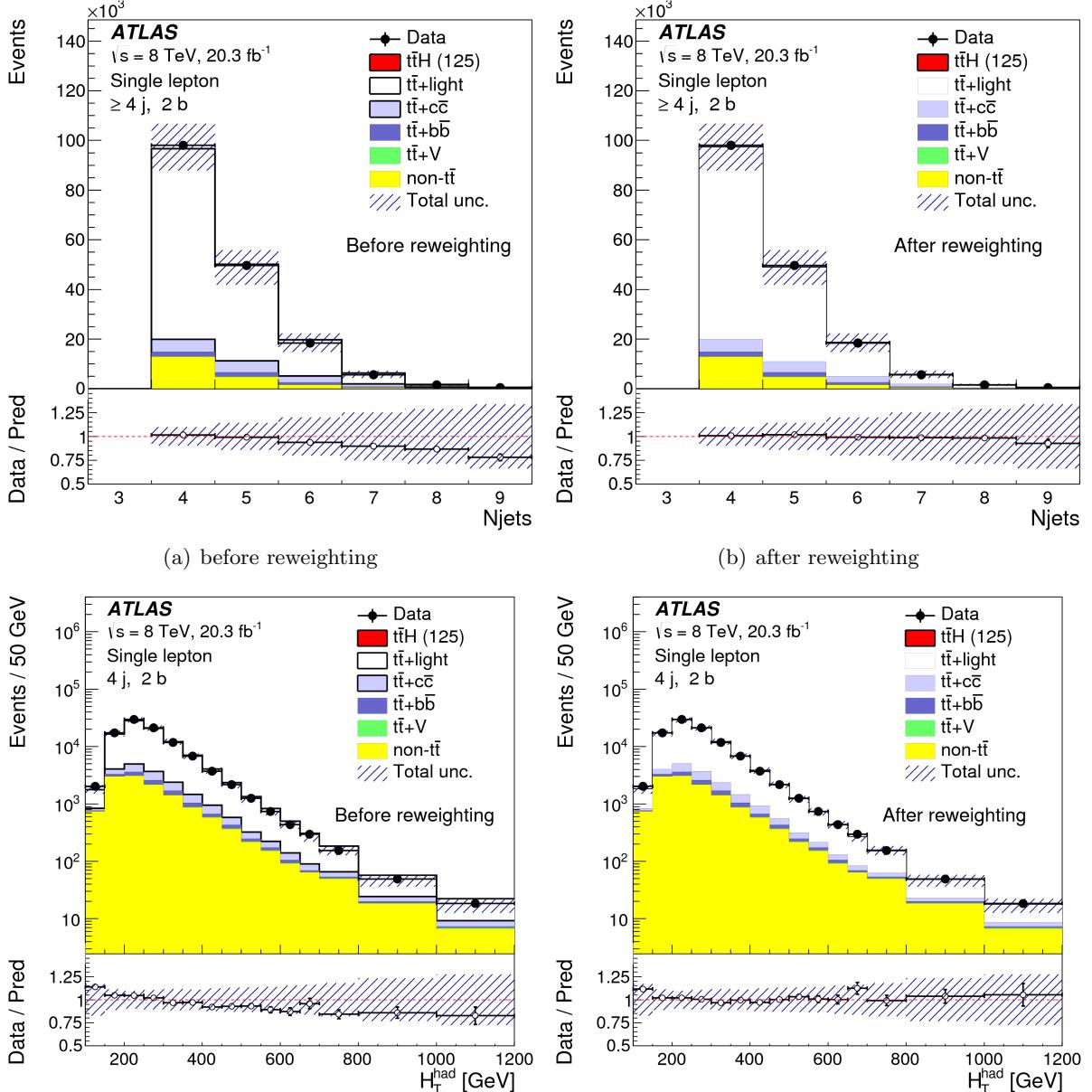


Figure 4.10.: The jet multiplicity (upper) and H_T^{had} distribution (lower) are shown before (a) and after (b) the p_T reweighting of the top quark and $t\bar{t}$ system of the POWHEG+PYTHIA events.

Single Top Quark Production

Single top quark events are produced in the s -, t - and Wt -channel with the POWHEG matrix element generator at NLO accuracy and weighted with the CT10 PDFs [148]. The interferences between diagrams in the Wt -channel at NLO are handled using the *diagram removal* method, which ensures that diagrams overlapping with the $t\bar{t}$ production are removed [158]. PYTHIA 6.425 [128] with the LO CTEQ61L PDF set [159] models the showering, hadronisation and the UE using the PERUGIA2011C tune. All three channels are normalised to approximate NNLO cross sections [51, 52] calculated with the NNLO MSTW2008 PDF.

Top Quark Pair Production in Association with Vector Bosons

Events originating from $t\bar{t}V$ ($V = W, Z$) processes are produced with the MADGRAPH 5 generator [126] and CTEQ6L1 PDFs in LO accuracy. The parton shower and fragmentation as well as the modelling of the UE is performed with PYTHIA 6.425 making use of the AUET2B tune [149, 150] for the UE. The overall normalisation is obtained by scaling to the NLO cross section predictions [160, 161].

Vector Boson Production in Association with Jets

Events with a single vector boson ($W/Z+\text{jets}$) or dibosons ($WW/WZ/ZZ+\text{jets}$) in association with jets are created with the LO matrix element generator ALPGEN v2.14 [125] with LO CTEQ6L1 PDFs [159]. The single vector boson events are then passed to PYTHIA 6.425 [128] and the diboson events to HERWIG v6.520 [130] to model the parton showering and fragmentation. The $W/Z+\text{jets}$ samples are generated with up to five additional partons, separately for $W/Z+\text{light jets}$, $W/Z+b\bar{b}+\text{jets}$, $W/Z+c\bar{c}+\text{jets}$ and $W+c+\text{jets}$. The $WW/WZ/ZZ+\text{jets}$ samples are generated with up to three additional partons. Unfortunately, WZ events are not produced with hadronic Z decays in the final state in the ALPGEN+HERWIG production. Therefore SHERPA is used to simulate this final state including massive b - and c -quarks.

Since the vector bosons are already produced with additional partons in the matrix element calculation, the parton shower might produce the same final state configuration. In order to avoid double-counting of the same partonic configurations, the *MLM matching* scheme [162] is used to remove events containing emissions from the parton shower in the phase space region, which is already covered by the matrix element calculation. Additionally, the overlap of $W/Z+q\bar{q}$ ($q = b, c$) events produced directly by the matrix element generator and by the parton shower evolution in $W+\text{light jet}$ events is prevented by the *heavy flavour overlap removal (HFOR)* algorithm based on the angular separation of the extra heavy quarks: if $\Delta R(q, \bar{q}) > 0.4$, the matrix element prediction is preferred over the parton shower prediction, which is used otherwise. The $W/Z+\text{jets}$ events are normalised to the inclusive NNLO theoretical cross sections [163], while $WW/WZ/ZZ+\text{jets}$ are scaled to their NLO prediction [164].

4.4.4. Tag-Rate-Function Method

Since the MC simulations are generated with limited statistics and the $t\bar{t}H$ signal of this search is expected in high jet and b -jet multiplicities, where the MC statistics of several processes is drastically reduced, the statistical uncertainties on the background predictions in the signal region would be large. This leads to large fluctuations in the discriminating distributions, unreliable systematic uncertainties and thus to the loss of sensitivity to the signal. Furthermore, sophisticated analysis techniques, like a multivariate analysis (MVA) or a profile likelihood fit, can not be performed, because they need smooth distributions of the input variables. Without these techniques the intended search for $t\bar{t}H$ events would be practically impossible with the available data and the current modelling of the background processes. However, the MC statistics of the phase space regions with many b -jets can be enhanced by making use of the *tag-rate-function (TRF)* method [165, 166]. This approach uses the full statistics of a sample before any b -tag is required. It assigns weights to the event representing the probability of the event to have a certain number of b -tags instead of applying b -tagging algorithms. In this way

the MC statistics is not reduced by the b -tag requirement. The probability P that an event with N_j jets has $N_b = 0, 1, 2$ b -tags is calculated as

$$P(N_b = 0) = \prod_{i=1}^{N_j} (1 - \varepsilon_i) \quad (4.7)$$

$$P(N_b = 1) = \sum_{i=1}^{N_j} \left(\varepsilon_i \prod_{j \neq i} (1 - \varepsilon_j) \right) \quad (4.8)$$

$$P(N_b = 2) = \sum_{i=1}^{N_j} \left[\varepsilon_i \sum_{j=i+1}^{N_j} \left(\varepsilon_j \prod_{k \neq i, k \neq j} (1 - \varepsilon_k) \right) \right], \quad (4.9)$$

where ε_j is the b -tagging efficiencies $\varepsilon^{MC}(f, |\eta|, p_T)$ depending on the true jet flavour f and its η and p_T . Probabilities for higher b -tag multiplicities can be defined accordingly. Because the official measurement of these efficiencies (see Sec. 4.3.1) is obtained from inclusive $t\bar{t}$ events, the usage of these efficiencies leads to discrepancies between direct tagging and the TRF approach. Hence more accurate efficiencies are calculated separately for $t\bar{t} + \text{light}$, $t\bar{t} + c\bar{c}$ and $t\bar{t} + b\bar{b}$ events using the same selection as applied for this analysis (see Sec. 6.2.1). It is further distinguished if the probed light, c - or b -jet originates from a top quark, W boson or from other sources. Since the b -tagging efficiencies ε^{MC} are obtained in MC simulation and differences between data and predictions are observed, b -tagging SFs are applied to the efficiencies $\varepsilon = SF \times \varepsilon^{MC}$ when calculating the probabilities. The event probability is derived from all possible permutations of N_b jets being b -tagged with probability ε and of $(N_j - N_b)$ jets being untagged with probability $(1 - \varepsilon)$. Inclusive probabilities

$$P(N_b \geq M) = 1 - \sum_{i=0}^{M-1} P(N_b = i) \quad (4.10)$$

are calculated via the complementary probabilities. In this analysis, it is not enough to know the event probability. To build the discriminating variables one has to know which jets are b -tagged. In order to decide this, the sum $S = \sum_{p=1}^{N_p} w_i$ of all TRF weights w_i for each possible permutation p of having N_b b -jets out of N jets is calculated. Then a random number X uniformly distributed between 0 and S is drawn. One permutation p_s is chosen, for which X lies in the range of its partial sums, which means if $\sum_{p=1}^{p_s-1} w_i < X \leq \sum_{p=1}^{p_s} w_i$ then p_s is the selected permutation. In this way it is ensured, that the probability to pick a certain permutation is proportional to its tagging probability. To validate the TRF method, kinematic distributions obtained by applying the TRF weights to MC and the direct b -tagging are compared. In Figure 4.11 the scalar sum of the p_T of all hadronic objects H_T^{had} , reflecting well the kinematics of the jets, is shown for $t\bar{t} + \text{light}$ and $t\bar{t} + b\bar{b}$ events in different phase space regions for the purpose of validation.

When requiring exactly four jets and two b -tags (upper row), the distributions obtained with the TRF method (red dashed line) and the cut based distributions (blue dots) agree very well for both $t\bar{t}$ sub-samples up to high p_T regions, where the statistical uncertainties start to increase. When requiring at least six jets and four b -tags (lower row), the MC statistics of the $t\bar{t} + b\bar{b}$ and $t\bar{t} + \text{light}$ events is significantly reduced and large statistical fluctuations are visible, in particular for the $t\bar{t} + \text{light}$ events in the cut based distribution. The distributions using the

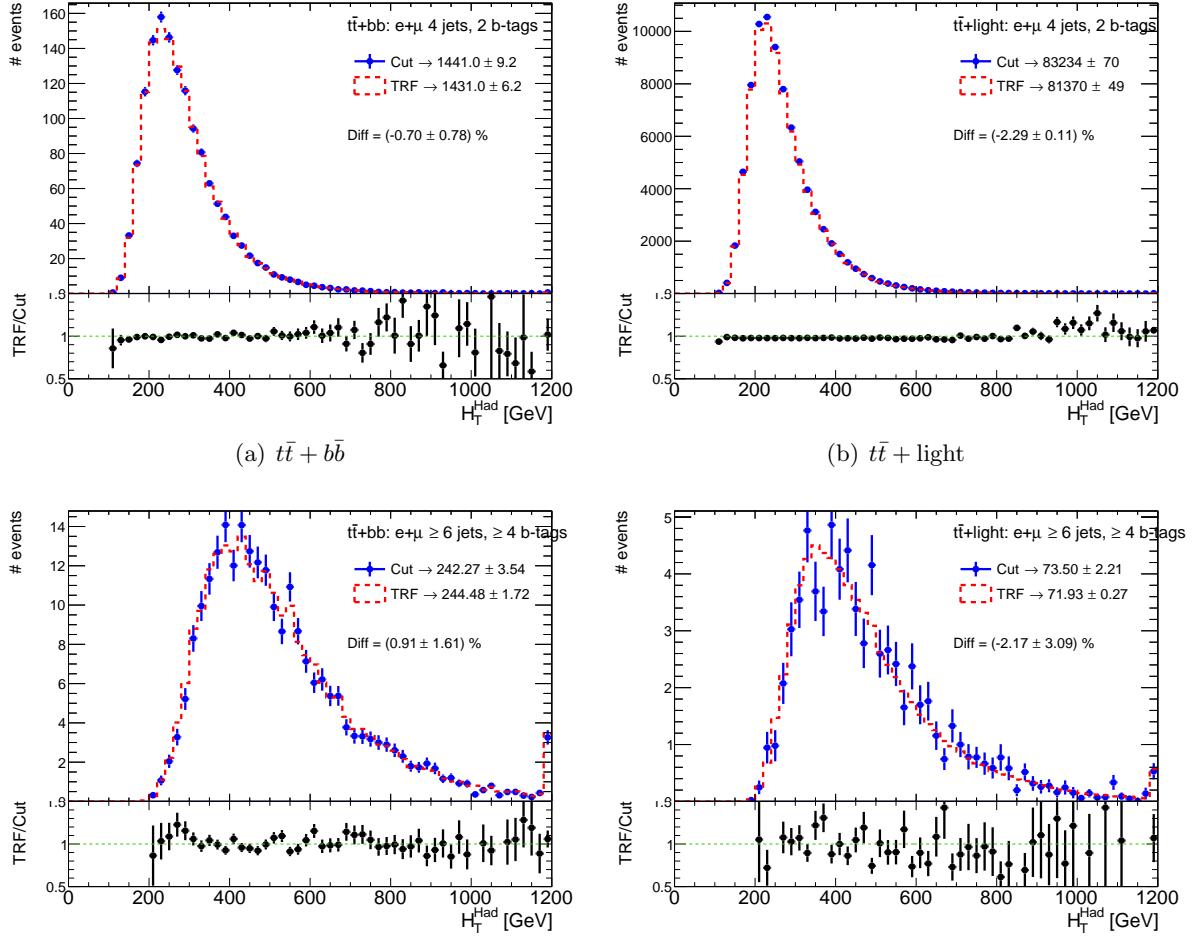


Figure 4.11.: The H_T^{had} distributions are shown exemplary in the (4j, 2b) and (≥ 6 j, ≥ 4 b) channel for the $t\bar{t} + b\bar{b}$ and $t\bar{t} + \text{light}$ MC simulations. The distributions obtained with the TRF method (red dashed line) are compared to the cut based distributions (blue dots).

TRF weights instead of the b -tag cuts are much smoother and agree well with the cut based distributions within the statistical uncertainties.

4.4.5. Data-Driven Estimated Background

Although multijet events with a misidentified isolated lepton (*fake lepton*) can be generally well suppressed, they still contribute significantly in certain regions of the phase space. Since an accurate modelling via MC simulation is extremely difficult in all phase space regions, a data-driven approach is used for the modelling and is described in more detail in the following.

Misidentified Lepton Background

Although the efficiencies to select a real lepton with the definitions described in Sections 4.3.2 – 4.3.3 are relatively high, it may happen that an object of the origin other than a prompt lepton, satisfies the selection criteria. These misidentified leptons, produced together with additional jets, contribute to the physics background. They mainly originate from non-prompt leptons from semileptonic decays of HF hadrons or in case of the electron also from converted photons or jets with large fraction of energy deposited in the EM calorimeter. The most common technique to estimate this background from data is the *matrix method*.

Matrix Method In order to apply the matrix method in the single lepton channel, two different selection criteria for the lepton need to be defined - a loose and a tight criterion. When applying them to the data sample, two sets are produced. In general, the tight set with N^{tight} events is a subset of the loose set with N^{loose} events. Both quantities can be expressed as the linear combination of the number of events with a real lepton N_{real} and with a misidentified lepton N_{fake} . This leads to the set of equations

$$\begin{pmatrix} N^{\text{loose}} \\ N^{\text{tight}} \end{pmatrix} = \begin{pmatrix} N_{\text{real}}^{\text{loose}} + N_{\text{fake}}^{\text{loose}} \\ N_{\text{real}}^{\text{tight}} + N_{\text{fake}}^{\text{tight}} \end{pmatrix}. \quad (4.11)$$

It can be converted using the transition efficiencies ϵ that a loose event passes also the tight criteria,

$$\epsilon_{\text{real}} = \frac{N_{\text{real}}^{\text{tight}}}{N_{\text{real}}^{\text{loose}}}, \quad \epsilon_{\text{fake}} = \frac{N_{\text{fake}}^{\text{tight}}}{N_{\text{fake}}^{\text{loose}}} \quad (4.12)$$

into the set of equations

$$\begin{pmatrix} N^{\text{loose}} \\ N^{\text{tight}} \end{pmatrix} = \begin{pmatrix} N_{\text{real}}^{\text{loose}} + N_{\text{fake}}^{\text{loose}} \\ \epsilon_{\text{real}} N_{\text{real}}^{\text{loose}} + \epsilon_{\text{fake}} N_{\text{fake}}^{\text{loose}} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ \epsilon_{\text{real}} & \epsilon_{\text{fake}} \end{pmatrix} \begin{pmatrix} N_{\text{real}}^{\text{loose}} \\ N_{\text{fake}}^{\text{loose}} \end{pmatrix}, \quad (4.13)$$

which gives the name to the method. This system of equations can be solved for the number of tight fake events

$$N_{\text{fake}}^{\text{tight}} = \frac{\epsilon_{\text{fake}}}{\epsilon_{\text{fake}} + \epsilon_{\text{real}}} (\epsilon_{\text{real}} N_{\text{real}}^{\text{loose}} - N^{\text{tight}}), \quad (4.14)$$

corresponding to the fake background contribution in the selected sample. This means, if the efficiencies ϵ_{real} and ϵ_{fake} are known, the number of fake lepton events of the tight sample can simply be estimated by counting the number of tight and loose events in the data sample. In order to predict the kinematics of the fake lepton events depending on the kinematic properties of the event, Equation 4.14 can be translated into a weight ω_i

$$\omega_i = \frac{\epsilon_{\text{fake}}}{\epsilon_{\text{fake}} + \epsilon_{\text{real}}} (\epsilon_{\text{real}} - \delta_i) \quad \text{with} \quad \delta_i = \begin{cases} 1 & \text{if event is tight} \\ 0 & \text{if event is only loose} \end{cases}, \quad (4.15)$$

which is applied to each loose event in a certain phase space point of the data set.

Measurement of the Efficiencies The real and fake efficiencies ϵ are determined in control regions which are dominated by real or fake leptons, respectively, and represent similar kinematics and sources of the leptons as in the signal region, to which the estimations are extrapolated. In the optimal case, they are also required to be orthogonal to the signal region. Naturally, the efficiencies depend on the detector properties and on the kinematics of the event. The fake efficiency ϵ_{fake} is also sensitive to the properties of the possible sources of the fake lepton. Thus it makes sense to measure the efficiencies dependent on kinematic properties of the event. Unfortunately, such measurement is usually limited by data statistics prohibiting a multidimensional parametrisation of the efficiencies and only one-dimensional projections to the parameter spaces can be used. The crucial part of the fake lepton background estimation is hence finding an optimal set of preferably uncorrelated parameters for the measurement of the ϵ_{real} and ϵ_{fake} .

For the selection of the single lepton events, two lepton triggers with different p_T threshold and isolation criteria are used. The one with the lower p_T threshold has an isolation requirement, which is tighter than the definition used for the loose lepton. The efficiencies are thus expected to differ depending on which trigger has fired to select the event, and are accordingly derived separately for the two triggers. They are then applied exclusively above the p_T threshold of the particular trigger favouring the high p_T trigger with the looser isolation criteria and avoiding turn-on effects. For the measurement [167] of the real efficiencies ϵ_{real} in the single electron and single muon channel, the TP method is used by selecting $Z \rightarrow ee$ or $Z \rightarrow \mu\mu$ events from the data. The events are required to contain one tight lepton (tag) and one loose lepton (probe) with opposite charges and with an invariant dilepton mass close to the Z boson mass as shown in Figure 4.12 (a) exemplary for the muon channel.

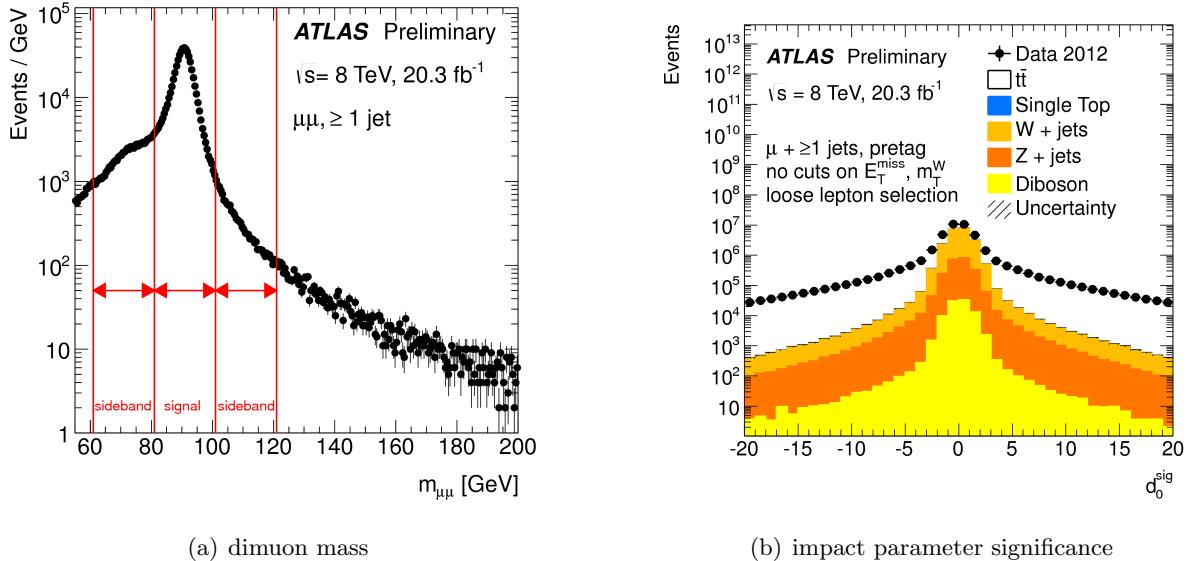


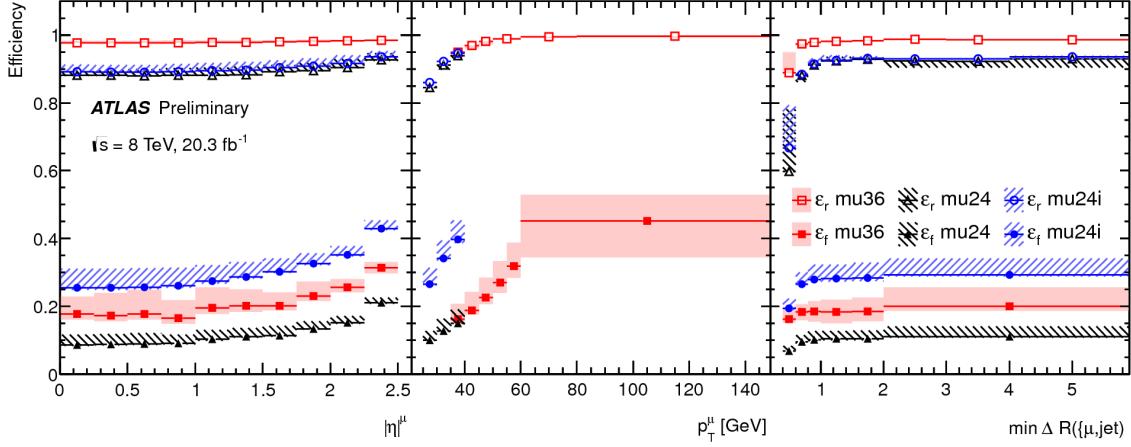
Figure 4.12.: (a) The real efficiencies are obtained in the real lepton dominated region (signal region) of the invariant mass spectrum of the opposite charged muon pair with one loose and one tight muon using $Z \rightarrow \mu\mu$ events. The fake lepton contamination is determined in the side bands and subtracted. (b) The fake efficiencies are obtained in the fake lepton dominated region defined by the high impact parameter significance ($|d_0^{\text{sig}}| > 5.0$). The real lepton contamination is determined from MC and subtracted [167].

The remaining background contribution coming from fake leptons, which is found to be at the percent level [167], is estimated in the side bands and then subtracted. The real efficiencies ϵ_{real} are now determined by calculating the rate of the loose probe lepton to also satisfy the tight selection criteria. In the electron channel, the efficiencies obtained in the $Z \rightarrow ee$ events are expected to differ slightly from those in a $t\bar{t}$ dominated sample, due to the higher jet activity. This is corrected by applying SFs obtained from comparing $t\bar{t}$ and Z boson MC simulations. The fake efficiencies ϵ_{fake} are obtained in a fake lepton enriched data control region of the data set, which is defined by requiring a high lepton impact parameter significance ($|d_0^{\text{sig}}| > 5.0$) in case of the muon (see Fig. 4.12 (b)) and by requiring low E_T^{miss} and a low transverse W boson mass ($m_T^W < 20 \text{ GeV} \& m_T^W + E_T^{\text{miss}} < 60 \text{ GeV}$) in case of the electron. Additionally, both regions have to contain one loose lepton and at least one jet with $p_T > 25 \text{ GeV}$. The real lepton contamination of these regions from processes with one prompt lepton, like $Z+\text{jets}$, $W+\text{jets}$, $t\bar{t}$, single top and diboson events, is estimated and subtracted by using MC simulation. It is found to be of the order of 15% and 50% in events with one tight muon or electron, respectively. The fake efficiencies are then determined by calculating the ratio of events with one loose lepton and one tight lepton.

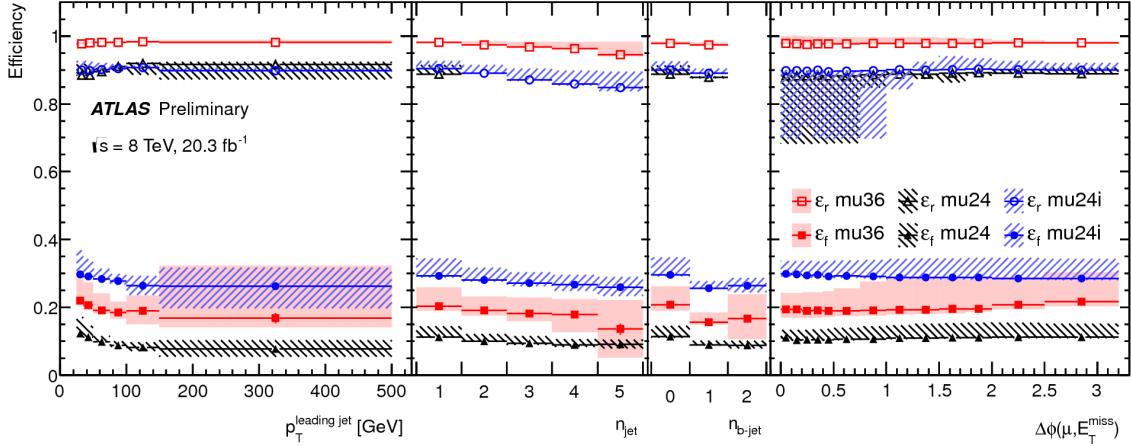
As motivated previously, both efficiencies are measured dependent on several different variables suggested by the physics processes and the detector geometry. The most prominent dependencies are shown in Figure 4.13, where the efficiencies in the muon channel are calculated separately for the three different triggers (mu24, mu24i, mu36) and plotted versus the muon $|\eta|$, p_T and its distance to the closest jet, the leading jet p_T , the jet and b -jet multiplicity as well as the angle $\Delta\phi$ between the muon and E_T^{miss} in the transverse plane.

In all plots the different responses due to the different trigger requirements are visible. While the efficiencies for the pre-scaled trigger with no isolation and a p_T threshold of 24 GeV (mu24) are shown only for comparison, the efficiencies of the isolated low p_T trigger mu24i ($p_T > 24 \text{ GeV}$) are used for muon $p_T < 37 \text{ GeV}$ and the ones of the non-isolated high p_T trigger mu36 ($p_T > 36 \text{ GeV}$) are used for muon $p_T \geq 37 \text{ GeV}$. Although higher muon momenta yield higher efficiencies, the isolation criteria profoundly effect the fake efficiencies, which causes the ϵ_{fake} of the mu24i to be higher than the one of the mu36, while the ϵ_{real} of the mu36 are generally higher than the one of the mu24i trigger. There is some clear dependency visible for the real and fake efficiencies on the muon η , p_T and its distance to the closest jet. This is well motivated by the detector geometry and the fact, that mostly non-prompt muons, originating from the decay of HF hadrons, cause the misidentification of an isolated muon and the overlap removal as described in Section 4.3.3 might accidentally remove a prompt muon if it is close to a jet. Both efficiencies are thus parametrised dependent on these three quantities. In addition, the fake efficiency is slightly dependent on the number of b -jets, since higher HF activity causes more fake muons, and hence this variable is also considered in the parametrisation. Both efficiencies are also slightly dependent on the number of jets, due to increased hadronic activity. However, this variable is only included in the parametrisation of the real efficiency due to high correlations to the b -jet multiplicity. The final choice of parameters for both lepton channels are summarised in Table 4.3 and is driven by the strength of the dependencies, correlations among the variables and the stability of the estimate.

In the parametrisation of the efficiencies, correlations between discrete and continuous variables are taken into account, but not within one category, since the correlations are proven to be negligible. Several sources of systematic uncertainties on the efficiency calculations are considered and are shown in combination with the statistical uncertainty as hashed area in Figure 4.13. Although the choice of the control regions, in which the real and fake efficiencies are calculated,



(a) ϵ_{real} and ϵ_{fake} depending on muon $|\eta|$, p_T and its distance to the closest jet



(b) ϵ_{real} and ϵ_{fake} depending on the leading jet p_T , jet & b -jet multiplicity and $\Delta\phi$ between μ & E_T^{miss}

Figure 4.13.: The real and fake efficiencies in the single muon channel for three different trigger requirements (mu24, mu24i, mu36) and dependent on various quantities. The hashed area represents the combination of the statistical and systematic uncertainties on the efficiency measurements [167].

is motivated by physics, other control regions are suitable as well and should yield compatible results. Alternative control regions for the calculation are chosen and the differences in the results are interpreted as a systematic uncertainty resulting from the choice of the control region. For the real and fake efficiencies a high and a low missing transverse energy region ($E_T^{\text{miss}} > 150$ GeV and $E_T^{\text{miss}} < 20$ GeV) is used in the electron channel, while a high and a low transverse W boson mass region ($m_T^W > 100$ GeV and $m_T^W < 20$ GeV & $E_T^{\text{miss}} + m_T^W < 60$ GeV) is selected in the muon channel. Fake lepton contamination in the regions, where the real efficiencies are calculated, are negligibly small in these cases. The obtained uncertainties cover effects stemming from different fake lepton compositions and the modelling of the background shape. They are below 3% for the real efficiency and between 2-5% for the fake efficiency. Since the real lepton contributions are subtracted using MC simulations for the fake efficiency calculation, the shape and normalisation uncertainties on the various predicted processes are

Table 4.3.: A summary of the variables used for the parametrisation of the real and fake efficiencies in the two lepton channels. Correlations between discrete and continuous variables are taken into account.

	discrete variables			continuous variables				
	Trigger	n_{jet}	$n_{\text{b-jet}}$	$ \eta^\ell $	p_T^ℓ	$p_T^{\text{lead. jet}}$	$\Delta R(\ell, \text{jet})$	$\Delta\phi(\ell, E_T^{\text{miss}})$
$\epsilon_{\text{real}}(e)$	✓	✓		✓	✓			✓
$\epsilon_{\text{real}}(\mu)$	✓	✓		✓	✓			✓
$\epsilon_{\text{fake}}(e)$	✓		✓	✓		✓		✓
$\epsilon_{\text{fake}}(\mu)$	✓		✓	✓	✓			✓

also taken into account. This turns out to be the dominating uncertainty, mainly due to uncertainties on the $W+\text{jets}$ and $Z+\text{jets}$ simulation, with a relative size of 3-13% on the fake efficiency. Finally, the efficiencies depend on the choice of the parametrisation. Thus alternative dependencies are used in order to estimate the uncertainty arising from not taken into account choices of variables and from correlations among the variables.

Validation of the Estimate The efficiencies and their uncertainties are used to predict the shape and rate of the fake lepton background and its uncertainty in validation regions and the phase space of physics interest. The total background predictions are compared to data to validate the fake lepton estimate. Figure 4.14 shows the distributions of the missing transverse energy (a) and the transverse W boson mass (b) in the single muon channel requiring exactly two jets with $p_T > 25$ GeV and before (top) and after (bottom) requiring at least one b -jet. The uncertainty on the prediction (hashed) includes only the statistical uncertainties and the uncertainty on the fake muon background estimation. The ratio of data over total background prediction is plotted in the inlay below the distribution. In this phase space region, the background is dominated by $W+\text{jets}$ and $Z+\text{jets}$ events (light and dark orange), but contains also a large fraction of fake muons (purple) dominating the low E_T^{miss} and m_T^W region. The prediction agrees well with the observed data within the uncertainties, wherever there are sizeable contributions from non-prompt muons. The deviations observed in the real lepton dominated peak region of the transverse W boson mass, in particular after requiring a b -tag, is caused by inaccurate simulation of $W+\text{jets}$ events. This and the disagreement between the overall normalisation of the MC simulation and the data after the b -tag requirement is well covered by the uncertainties on the MC simulation not shown in this plot. In Figure 4.13 the same distributions are shown in a higher jet multiplicity by requiring at least four jets. In this phase space region $t\bar{t}$ events contribute a larger fraction to the total background. The kinematics of this region is closer to the one of $t\bar{t}H$ events. In the fake lepton dominated region, the predictions agree very well within the uncertainties on the fake estimation with the observed data.

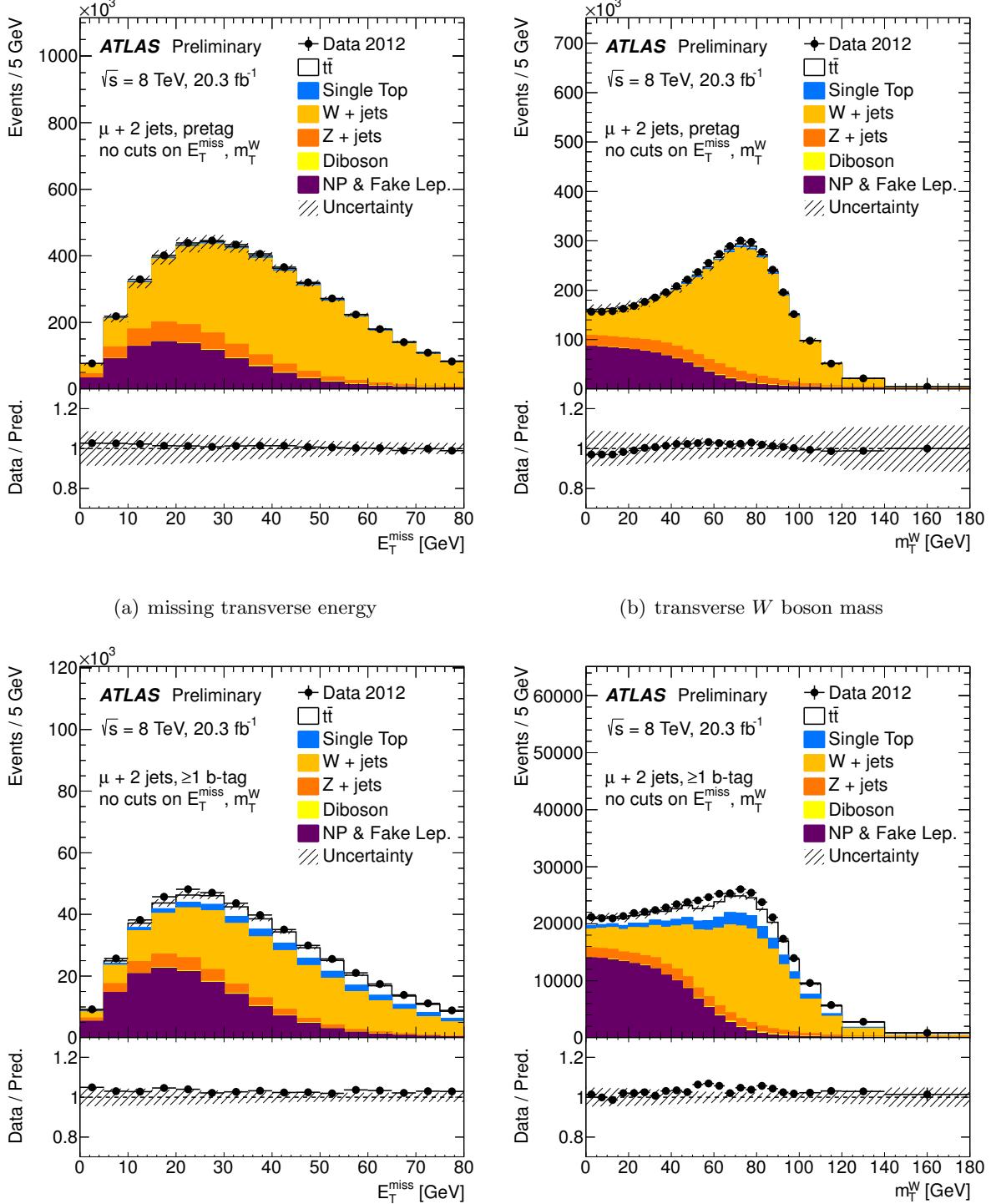


Figure 4.14.: The missing transverse energy (a) and transverse W boson mass (b) in the single muon channel with exactly two jets and before (top) and after (bottom) requiring at least one b -jet. The data is compared to the expectations from MC plus the fake lepton contribution (purple). The hashed area represents the combined statistical and systematic uncertainty on the fake lepton estimation [167].

4. Experimental Data, Physics Objects and Processes

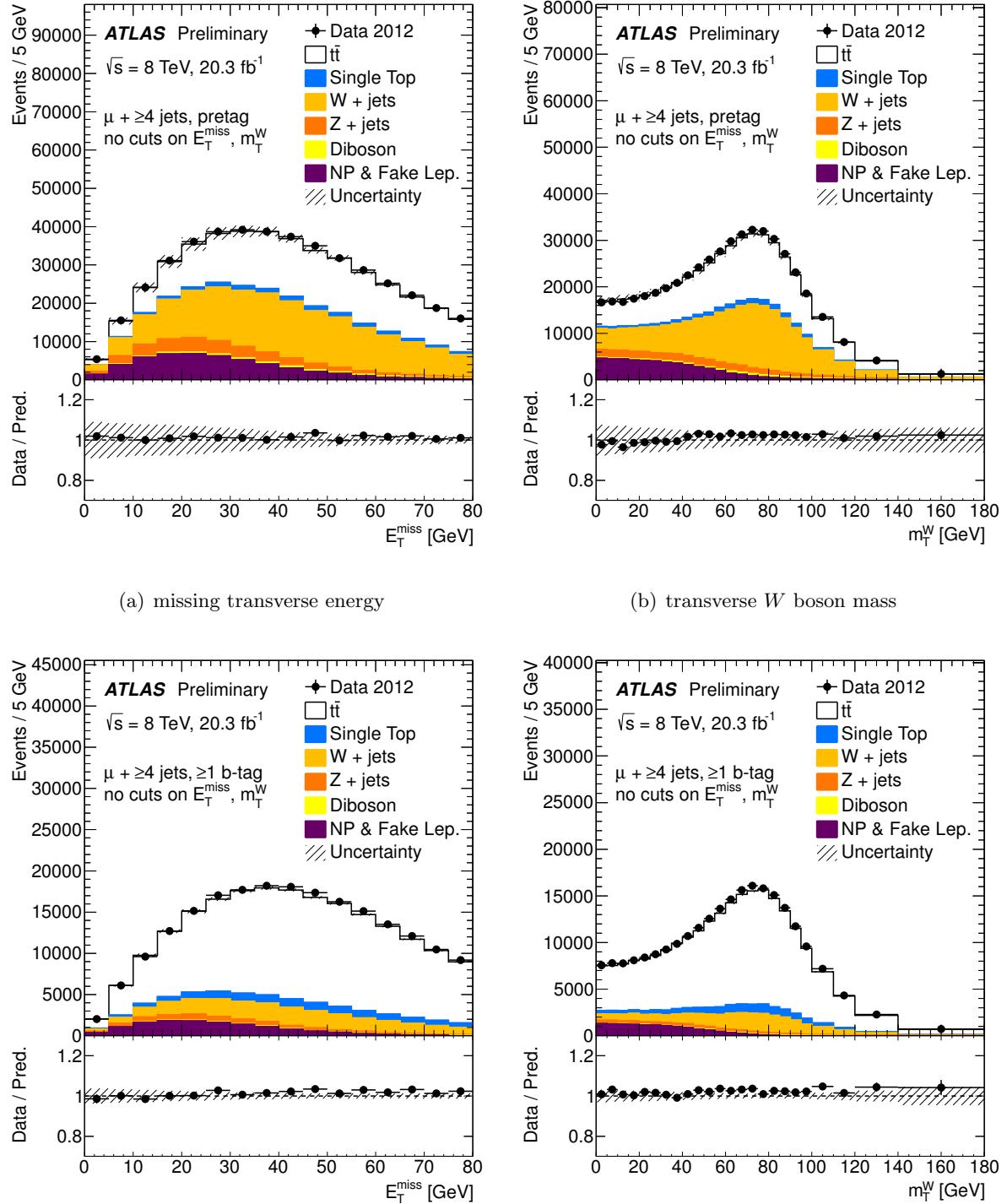


Figure 4.15.: The missing transverse energy (a) and the transverse W boson mass (b) in the single muon channel with at least four jets and before (top) and after (bottom) requiring at least one b -jet. The data is compared to the expectations from MC plus the estimated fake lepton contribution (purple). The hashed area represents the statistical and systematic uncertainty on the fake lepton estimation [167].

V+jets Corrections

Transverse Momentum The ALPGEN generator predicts a V boson with a too-hard transverse momentum spectrum. The discrepancies with data are similar to missing higher-order electroweak corrections [168]. Correction SFs for V +jets MC events are calculated in a phase space dominated by Z bosons decaying into two leptons by requiring at least two jets, a dilepton invariant mass close to the Z boson mass and either zero or one b -tag, respectively. After subtracting other contributing processes from data based on MC simulations, the observed and predicted Z boson p_T spectra are compared and the ratio is taken as an event weight to correct the MC simulation of V +jets events, keeping the total predicted cross section constant.

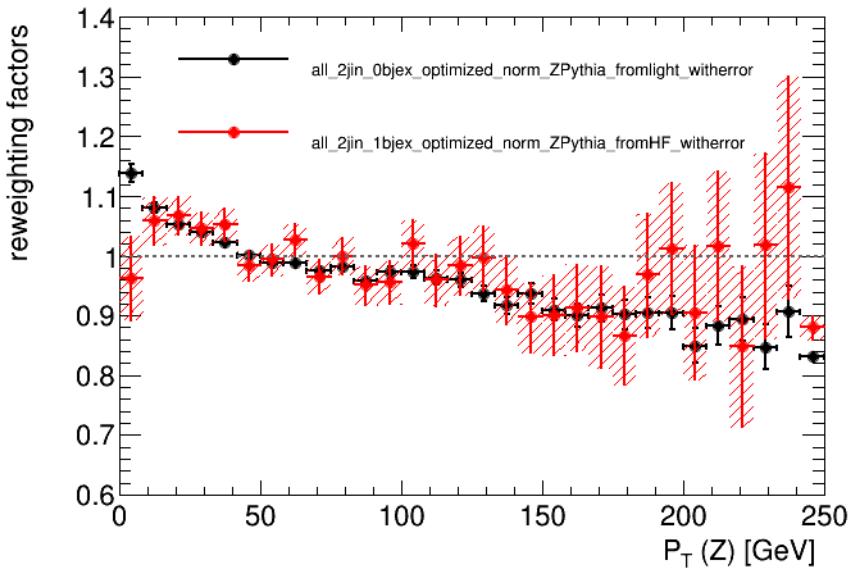


Figure 4.16.: Reweighting factors calculated from differences between data and MC simulation in the Z boson transverse momentum requiring at least two jets and zero (black) or one (red) b -tag [169].

As shown in Figure 4.16, the corrections are consistent within statistical uncertainties in both b -tag bins and thus independent of the HF content. The corrections can therefore be extrapolated to higher jet and b -tag multiplicities in order to improve the modelling the V bosons transverse momentum.

Normalisation and Heavy Flavour Contribution The total normalisation and the HF composition is generally poorly predicted by leading order generators like ALPGEN and needs to be estimated with a data driven approach. Since the zero and one b -tag bin contain considerably different flavour contributions, they can be used to estimate correction SFs for the different flavours of Z +jets background by simultaneously fitting the MC simulation to data with the LF and HF fractions being free parameters. The Z +light fraction is reduced by a factor of 0.94 and the Z +HF contribution is increased by a factor of 1.50, which also leads to a renormalisation of the total predicted cross section. The correction factors measured in Z +jets events are also applied to W +jets events. It is further assumed that the same SFs apply to higher jet and b -tag multiplicities [169].

4. Experimental Data, Physics Objects and Processes

Validation In order to show the effect of the V boson momentum reweighting and HF correction and to validate the procedure, the distributions of the scalar sum of the p_T of all final state objects (H_T) are shown before and after reweighting the zero b -tag and extrapolating to the two b -tag bins in Figure 4.17.

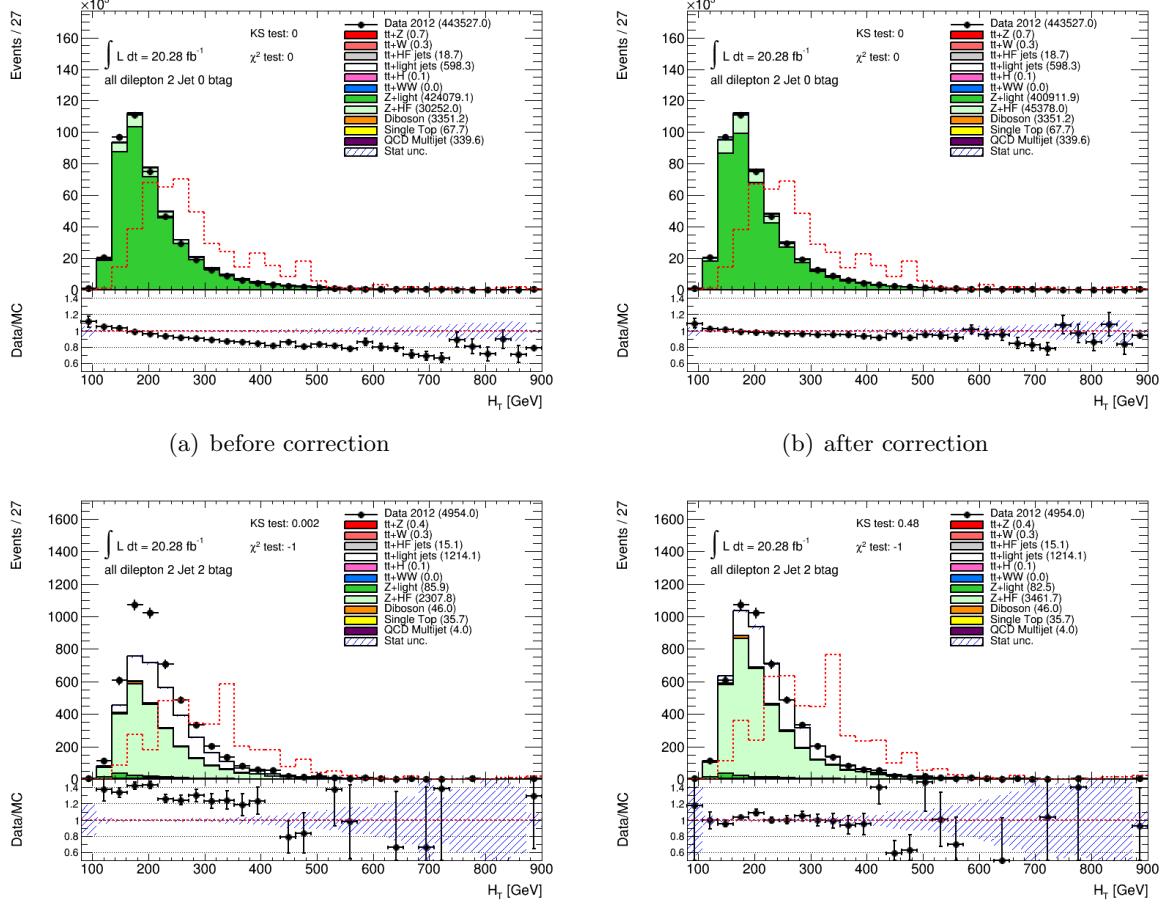


Figure 4.17.: The H_T distribution in the two jet inclusive channel before (left) and after (right) V boson transverse momentum and flavour fraction corrections. In the upper row the zero b -tag bin is shown, in which the corrections have been calculated, while in the lower row the corrections are applied to the two b -tag bin. The red dashed line corresponds to the $t\bar{t}V$ signal normalised to the integral since the correction was derived within the $t\bar{t}V$ analysis [169].

In the zero b -tag bin (upper row), in which the SFs have been calculated, the slope visible in the data over MC ratio of the H_T distribution almost disappears after the corrections and the MC simulation agrees well with the data within the uncertainties. The LF contribution is slightly decreased, while significantly more events with HF jets contribute after the correction. In the two b -tag bin (lower row), which is dominated by $Z+HF$, the MC prediction clearly underestimates the data before the reweighting, while the rescaled MC events agree well within their uncertainties with the data events.

5

The Matrix Element Method

5.1. Introduction

The *matrix element method (MEM)*, which was originally proposed as *Dynamical Likelihood Method* [170–172], has been invented by the DØ [5] and used by the CDF experiments [6] for precision measurements of the top quark mass [173], for the single top observations [174, 175], for evidence of spin correlations in top quark pair decays [176] and for simple Higgs boson searches [7, 8]. Several authors [177, 178] have suggested to use the MEM for searches of the Higgs boson and measurement of its properties at the LHC. In particular in searches for the $t\bar{t}H$ process where the dominant and irreducible $t\bar{t}$ background has a very similar kinematic and flavour signature, one can gain from making the most complete use of experimental and theoretical information by using the matrix element method [179]. Recently, this technique has been used for the $t\bar{t}H$ search by the CMS experiment [180].

5.2. Method Description

The MEM is universal and can be applied to any theoretically described process. Amongst all data analysis techniques in particle physics, this method is unique due to its direct link between theoretical calculations and reconstructed events.

5.2.1. Process Probability

The method is based on a probabilistic approach, in which it is required to calculate the probability density function (p.d.f.) that an event observed in the detector is consistent with a certain theoretical hypothesis. The probability \mathcal{P}_i , that an event with perfectly measured kinematic quantities Y is originating from a possible process i occurring with a cross section σ_i , is given by integrating over the normalised differential cross section

$$\mathcal{P}_i(Y \in \Phi) = \int_{Y \in \Phi} P_i(\mathbf{y}) d\mathbf{y} = \frac{1}{\sigma_i} \int_{\sigma_i(Y)} d\sigma_i(\mathbf{y}). \quad (5.1)$$

5. The Matrix Element Method

Here it is assumed that Y is contained in the parton level phase space of the initial and final states \mathbf{y} with volume $\Phi = \int d\mathbf{y}$. The integrals over the entire phase space are normalised to unity fulfilling $\int_{\Phi} P_i(\mathbf{y}) d\mathbf{y} = \frac{1}{\sigma_i} \int_{\sigma_i(\Phi)} d\sigma_i = 1$. Thus the p.d.f. $P_i(\mathbf{y})$ of a process is defined by

$$P_i(\mathbf{y}) = \frac{1}{\sigma_i} \frac{d\sigma_i(\mathbf{y})}{d\mathbf{y}}. \quad (5.2)$$

However, in reality every quantity is measured with a finite resolution and the most likely measured values might not agree with the true parton level quantity. In order to take the detector response into account, the differential cross section needs to be convoluted with the *transfer functions* (TFs) $W(\mathbf{y}|\mathbf{x})$, which map the detector level objects with quantities \mathbf{x} to the parton level description \mathbf{y} , leading to

$$P_i(\mathbf{x}) = \int P_i(\mathbf{y}) W(\mathbf{y}|\mathbf{x}) d\mathbf{y} = \frac{1}{\sigma_i} \int d\sigma_i(\mathbf{y}) W(\mathbf{y}|\mathbf{x}). \quad (5.3)$$

Details about the transfer functions can be found in Section 5.6.

Let us consider now a $2 \rightarrow N$ process, where two initial state particles A and B collide and create N final state particles $1, 2, \dots, N$

$$A + B \rightarrow 1 + 2 + \dots + N \quad (5.4)$$

The parton level phase space \mathbf{y} can be expressed using the four-momentum vector p of the initial and final states

$$\mathbf{y} = (p_A, p_B; p_1, p_2, \dots, p_N) \quad (5.5)$$

At hadron colliders like the LHC, where two protons consisting of many partons collide, the momentum fraction x of the initial state particles is unknown, and can only be described probabilistically by a convolution of the differential cross section with the parton distribution functions (PDFs) $f(x, Q^2)$ at a certain scale Q^2 and summing over all possible flavours

$$P_i(\mathbf{x}) = \frac{1}{\sigma_i} \sum_{\text{flavours}} \int dp_A dp_B f(p_A, Q^2) f(p_B, Q^2) d\sigma_i(\mathbf{y}) W(\mathbf{y}|\mathbf{x}). \quad (5.6)$$

The differential cross section might depend on several parameters α of the theoretical model of interest, as well as the TFs can depend on model parameters β of instrumental nature. Taking this into account and hiding the sum over the flavours $\mathbf{f}(x, Q^2) = \sum f(x, Q^2)$, the equation above can be written as

$$\underbrace{P_i(\mathbf{x}|\alpha, \beta)}_{\substack{\text{probability} \\ \text{normalisation}}} = \underbrace{\frac{1}{\sigma_i(\alpha)}}_{\substack{\text{normalisation}}} \int \underbrace{dp_A dp_B \mathbf{f}(p_A) \mathbf{f}(p_B)}_{\substack{\text{parton} \\ \text{distribution} \\ \text{function}}} \underbrace{d\sigma_i(\mathbf{y}|\alpha)}_{\substack{\text{differential} \\ \text{cross} \\ \text{section}}} \underbrace{W(\mathbf{y}|\mathbf{x}, \beta)}_{\substack{\text{transfer} \\ \text{functions}}} \quad (5.7)$$

This equation is schematically visualised in Figure 5.1 for the $t\bar{t}H$ signal process.

The cross section normalizes the probability, usually including experimental acceptance and efficiency effects. The PDFs describe the production mechanism via the incoming partons (green). The differential cross section defines the process hypothesis by describing the hard scattering process (blue). The TFs map the detector quantities to the parton level quantities \mathbf{y} of the final state particles of the hard scattering process (red). The integration is performed over the phase space volume of the initial and final state particles.

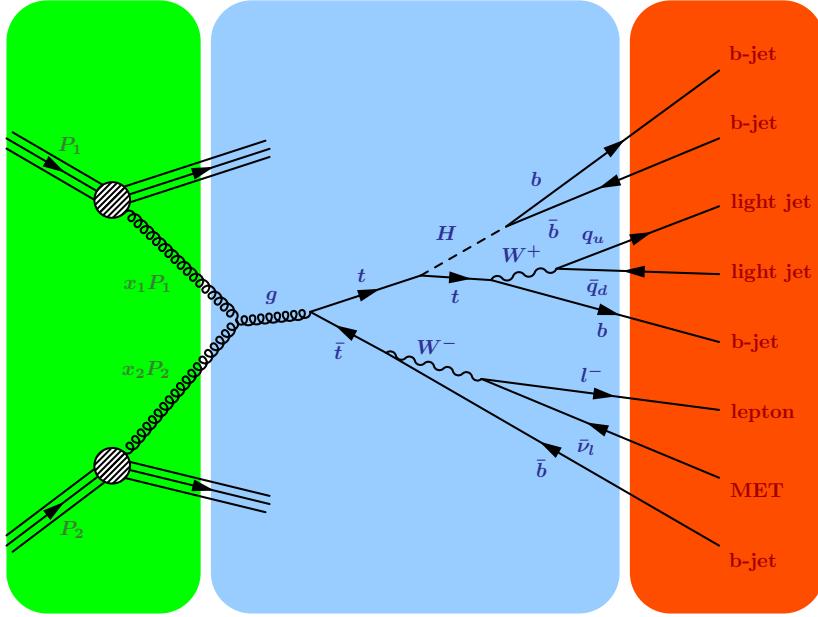


Figure 5.1.: A schematic visualisation of the content of the $t\bar{t}H$ process probability using the MEM.

5.2.2. Fermi's Golden Rule

The differential cross section can be related to the matrix element of the hard scattering process according to *Fermi's Golden Rule* [181]. The rule was originally derived for quantum mechanical transitions of initial to final states from perturbation theory using the time-dependent Schrödinger equation and assuming that the time of the transition is sufficiently small compared to the measurement. In the non-relativistic case, the rule states that the total transition rate is given by

$$\Gamma_{fi} = 2\pi|T_{fi}|^2\rho(E_i) , \quad (5.8)$$

where T_{fi} is the *transition matrix element* and

$$\rho(E_i) = \left| \frac{dn}{dE_f} \right|_{E_i} = \int \delta(E_i - E) dn \quad (5.9)$$

is the *density of states* with the number of accessible states dn in the energy range $[E, E+dE]$. In Equation 5.8 the transition matrix element describes the fundamental physics, while the density of states contains the kinematics of the considered process. For the $2 \rightarrow N$ process the number of accessible states can be transformed into the momentum space. Making at the same time use of momentum conservation leads to the expression

$$dn = (2\pi)^3 \prod_{i=1}^N \frac{d^3 p_i}{(2\pi)^3} \delta^3 \left(\mathbf{p}_A + \mathbf{p}_B - \sum_{i=1}^N \mathbf{p}_i \right) \quad (5.10)$$

where A and B denote the initial state particles and i are the indices for the final state particles. The factor $(2\pi)^{-3}$ is obtained from the boundary conditions of each wave function

5. The Matrix Element Method

in momentum space. The non-relativistic transition matrix element T_{fi} can be substituted by a Lorentz-invariant form

$$|\mathcal{M}_{fi}|^2 = |T_{fi}|^2 \prod_j 2E_j , \quad (5.11)$$

where the product over j includes all initial and final state particles. With these equations, Fermi's Golden Rule of Equation 5.8 for the $2 \rightarrow N$ process can be written in a Lorentz-invariant and relativistic way as

$$\Gamma_{fi} = \frac{(2\pi)^4}{2E_A 2E_B} \int |\mathcal{M}_{fi}|^2 \delta \left(E_A + E_B - \sum_{i=1}^N E_i \right) \delta^3 \left(\mathbf{p}_A + \mathbf{p}_B - \sum_{i=1}^N \mathbf{p}_i \right) \prod_{i=1}^N \frac{d^3 \mathbf{p}_i}{(2\pi)^3 2E_i} \quad (5.12)$$

or using four-momentum vectors and the relation $\frac{1}{2E} = \int \delta(E^2 - \mathbf{p}^2 - m^2) dE$ as

$$\Gamma_{fi} = \frac{(2\pi)^4}{2E_A 2E_B} \int |\mathcal{M}_{fi}|^2 \delta^4 \left(p_A + p_B - \sum_{i=1}^N p_i \right) \prod_{i=1}^N (2\pi)^{-3} \delta(p_i^2 - m_i^2) d^4 p_i , \quad (5.13)$$

where now the δ -functions intuitively ensure overall energy and momentum conservation and the Einstein relation $p^2 = m^2$. Simplifying the equation by using the definition of the Lorentz-invariant N -body phase space element

$$\begin{aligned} d\Phi_N(p_A + p_B; p_1, p_2, \dots, p_N) &= \delta^4 \left(p_A + p_B - \sum_{i=1}^N p_i \right) \prod_{i=1}^N \frac{d^3 \mathbf{p}_i}{(2\pi)^3 2E_i} \\ &= \delta^4 \left(p_A + p_B - \sum_{i=1}^N p_i \right) \prod_{i=1}^N \delta(p_i^2 - m_i^2) \frac{d^4 p_i}{(2\pi)^3} , \end{aligned} \quad (5.14)$$

the transition rate can be expressed as

$$\Gamma_{fi} = \frac{(2\pi)^4}{4E_A E_B} \int |\mathcal{M}_{fi}|^2 d\Phi_N . \quad (5.15)$$

Given that cross section σ of the process is related to the transition rate by

$$\sigma = \frac{\Gamma_{fi}}{v_A + v_B} , \quad (5.16)$$

where v_A and v_B being the four-velocity vectors of the colliding particle A and B, Fermi's Golden Rule can be rewritten as

$$\sigma = \frac{(2\pi)^4}{4E_A E_B (v_A + v_B)} \int |\mathcal{M}_{fi}|^2 d\Phi_N . \quad (5.17)$$

The denominator

$$\mathcal{F} = 4E_A E_B (v_A + v_B) = 4(E_B p_A + E_A p_B) \quad (5.18)$$

is known as the Lorentz-invariant flux factor and can be written as

$$\mathcal{F} = 4 \left[(p_A \cdot p_B)^2 - m_A^2 m_B^2 \right]^{\frac{1}{2}} , \quad (5.19)$$

assuming that the velocities of the incident particles A and B are collinear. Consequently, the differential cross section of the process probability of Equation 5.7 is directly proportional to the squared Lorentz-invariant transition matrix element of the process and can be calculated by using

$$d\sigma_i(\mathbf{y}|\boldsymbol{\alpha}) = \frac{(2\pi)^4}{\mathcal{F}} \int |\mathcal{M}_i(\mathbf{y}|\boldsymbol{\alpha})|^2 d\Phi_N(\mathbf{y}) . \quad (5.20)$$

The flux factor \mathcal{F} describes the kinematics of the initial state particle collision, the transition matrix element \mathcal{M} the pure physics defined by the Feynman diagrams of the hard scattering process and the phase space element $d\Phi_N$ the kinematic of the process. Finally, the process probability can be obtained by integrating numerically over the entire phase space of the initial and final state particles

$$P_i(\mathbf{x}|\boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{(2\pi)^4}{\sigma_i(\boldsymbol{\alpha})} \int d\Phi_N(\mathbf{y}) \mathbf{f}(p_A) \mathbf{f}(p_B) \frac{|\mathcal{M}_i(\mathbf{y}|\boldsymbol{\alpha})|^2}{\mathcal{F}} W(\mathbf{y}|\mathbf{x}, \boldsymbol{\beta}) . \quad (5.21)$$

This equation gives the name to the method, indicating that the full physics information is used for calculating the probability that an observed event is originating from a certain hard scattering process described by the transition matrix element.

5.2.3. Assignment Permutations

The assignment of the reconstructed objects, such as charged leptons and jets, to the final state partons in the hard process described by the matrix element contains multiple ambiguities. Hence the process probability density needs to be calculated for each allowed assignment permutation of the reconstructed objects to the final state quark or lepton of the hard process. Assignment permutations are only allowed within one object type. They can be further constrained by making additional use of detector information such as b -jet tagging or by exploiting the topology of the event. For instance, decays of intermediate states can result in invariant assignment permutations, which result in the same probability value. Two choices of building a process likelihood function are obtained by either summing the process probabilities for the N_p allowed assignment permutations

$$\mathcal{L}_i^{\text{sum}}(\mathbf{x}|\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{p=1}^{N_p} P_i^p(\mathbf{x}|\boldsymbol{\alpha}, \boldsymbol{\beta}) \quad (5.22)$$

or by using the permutation with the maximal process probability

$$\mathcal{L}_i^{\text{max}}(\mathbf{x}|\boldsymbol{\alpha}, \boldsymbol{\beta}) = \max \left\{ P_i^1(\mathbf{x}|\boldsymbol{\alpha}, \boldsymbol{\beta}), P_i^2(\mathbf{x}|\boldsymbol{\alpha}, \boldsymbol{\beta}), \dots, P_i^{N_p}(\mathbf{x}|\boldsymbol{\alpha}, \boldsymbol{\beta}) \right\} . \quad (5.23)$$

5.2.4. Event Probability and Parameter Estimation

If several processes can generate the same final state, the event probability is given by

$$P_{\text{evt}}^{\text{sum/max}}(\mathbf{x}|\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_i f_i \mathcal{L}_i^{\text{sum/max}}(\mathbf{x}|\boldsymbol{\alpha}, \boldsymbol{\beta}) , \quad (5.24)$$

5. The Matrix Element Method

where the f_i are the contribution fractions of the non-interfering processes. For a sample of n events, one global likelihood function can be obtained by multiplying all event probabilities

$$\mathcal{L}(\mathbf{x}|\boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{j=1}^n P_{\text{evt}}^j(\mathbf{x}_j|\boldsymbol{\alpha}, \boldsymbol{\beta}) . \quad (5.25)$$

Maximizing the likelihood function will return the best estimators for the model parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ of a given hypothesis under study. In this way, theory parameters $\boldsymbol{\alpha}$ can be estimated by measuring at the same time experimental parameters $\boldsymbol{\beta}$, which would degrade the precision and bias the result of the best estimator of $\boldsymbol{\alpha}$ due to systematic uncertainties.

The different process probabilities of a single event can also be used to distinguish signal from background events by calculating a parameter dependent likelihood ratio of the signal and background processes

$$r_{\text{sig}}(\mathbf{x}|\boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{\mathcal{L}_{\text{sig}}(\mathbf{x}|\boldsymbol{\alpha}, \boldsymbol{\beta})}{\sum_{\text{bkg}} f_{\text{bkg}} \mathcal{L}_{\text{bkg}}(\mathbf{x}|\boldsymbol{\alpha}, \boldsymbol{\beta})} . \quad (5.26)$$

According to the theorem of Neyman-Pearson [182], this is the most powerful discriminant between background and signal. In reality, the probabilities are not known with perfect precision and simplifications and approximations are needed during the integration process due to experimental and computational constraints. This leads to a degradation of the discrimination power, which can be partially recovered by using the likelihood ratio along with other kinematical and topological variables in a MVA.

The MEM can not only be used for precision measurements of the model parameter $\boldsymbol{\alpha}$ or to obtain great separation power coming from the process probability information, but also for a full event reconstruction of an arbitrary final state. The latter can be performed by choosing the assignment permutation, which gives the highest likelihood value \mathcal{L}_i^{\max} for a certain process i after the integration. Since the integration is performed over the full kinematic phase space and for all possible assignments of detector objects to parton level objects, one is not constrained to have enough kinematic information available as in kinematic likelihood fits, in which under-constrained systems are usually an issue.

5.3. Implementation: MEMTool

Following initial tests of the stand-alone MADWEIGHT software [183], it was determined that numerous improvements could be made by implementing a more tailored integration framework. A software package, called MEMTOOL, which was originally developed by the group of the University of Göttingen for a top mass measurement in the single lepton channel of the $t\bar{t}$ decay [184], provided a flexible approach and a good basis. For the purpose of this analysis, it has been completely generalised being able to run on any kind of event topology and testing any possible model. MEMTOOL is a very modular framework, written in C++, and designed to interface with multiple external libraries for various elements of the calculation. Some of the classes make use of the ROOT library [185], several PDF sets are made available using the LHAPDF 5.9.1 package [186] and the GSL 1.16 library [187] provides all kind of integration techniques. The most important structures of MEMTOOL and their relations are shown in Figure 5.2.

The EVENTREADER contains several interfaces to different input files, i.e. LHCO or ROOT

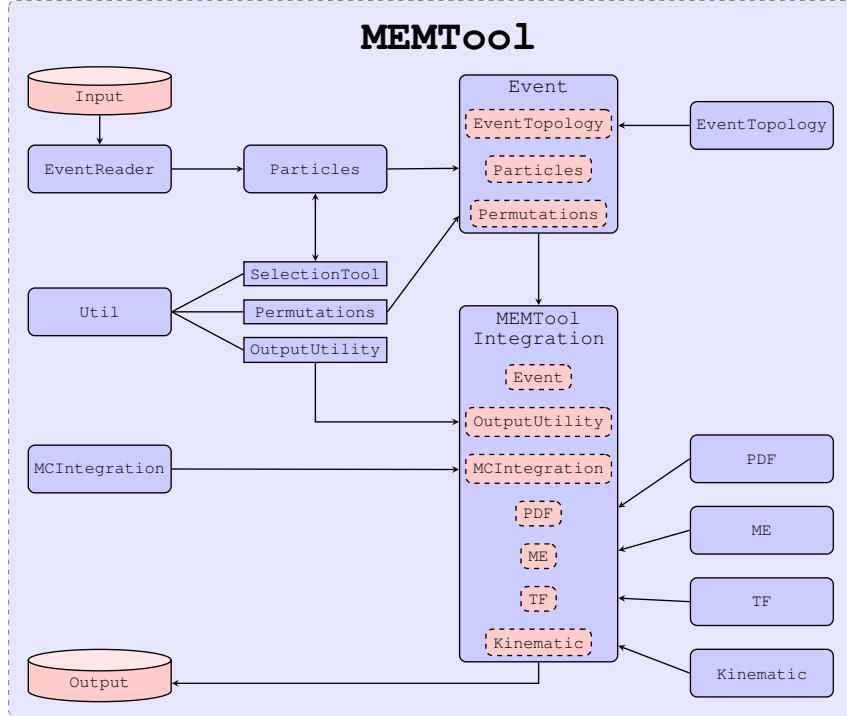


Figure 5.2.: A schematic drawing of the principal objects and functions of MEMTOOL.

files of various formats. It fills containers of PARTICLES objects of different type with its four-momentum vector plus some additional information, depending on the particle type, e.g. in case of jets, a b -tag flag, its b -tag weight, efficiency and rejection and its η in the detector coordinate system. With the help of a SELECTIONTOOL being part of a UTIL namespace, a selection on the particles can be applied depending on their properties and numbers. The particle containers are then passed to an EVENT object containing all the necessary event information. In addition to the PARTICLES, that are the EVENTTOPOLOGY defining the final states of the hard process and the possible detector level to parton level PERMUTATIONS defined by the particle containers and the topology of the event. The entire event is then passed to the MEMTOOLINTEGRATION structure, where all the information necessary for the matrix element integration is collected. Inside the PDF structure, all PDFs provided by the LHAPDF libraries are made available and are cached if desired for faster access. Several possible matrix element calculation methods, i.e. analytical calculations as well as a wrapper for generically generated matrix elements, are interfaced via the ME structure. The TF structure contains several sets of transfer functions, for instance simple δ -distributions, one-dimensional Gaussian functions and parametrised double-Gaussian functions from the KLFITTER package [188] described in Section 5.6. The KINEMATIC structure provides for the given process several kinematic transformations, which can be utilised to speed up the integration convergence and are described in Section 5.7.3. Finally, a MCINTEGRATION algorithm provided for instance by the GSL library is the last ingredient for the process probability calculation passed to the MEMTOOLINTEGRATION structure. An OUTPUTUTILITY, which is also part of UTIL, writes out the results of the matrix element integration into a standalone output file or a copy of the input file. The entire framework is designed in such a way, that any object of a certain structure can be easily replaced without changing anything else inside the framework.

5.4. Parton Distribution Functions

At hadron colliders, the kinematic configuration of the initial state particles in the hard scattering process can not be deterministically described by the machine set-up. Since protons and hence not elementary particles collide at the LHC, the momentum of the interacting particles is only reflected by a probability density function. The *parton distribution functions (PDFs)* $f(x, Q^2)$ are defined as the probability to find a certain constituent of the proton with a longitudinal momentum fraction x (*Bjorken x*) of the total proton's momentum probed at the energy scale Q^2 . The proton does not consist of three static valence quarks (uud), each carrying a momentum fraction of $\frac{1}{3}$, but of valence quarks interacting with each other by mediating gluons. These can split into virtual $q\bar{q}$ pairs of arbitrary flavour (*sea-quarks*) or further into additional gluons. Due to the interactions, the momentum fraction is smeared out and even at low momentum transfer of 1 GeV sea-quarks enhance the PDFs of the valence type quarks at low Bjorken x as shown in Figure 5.3 (a).

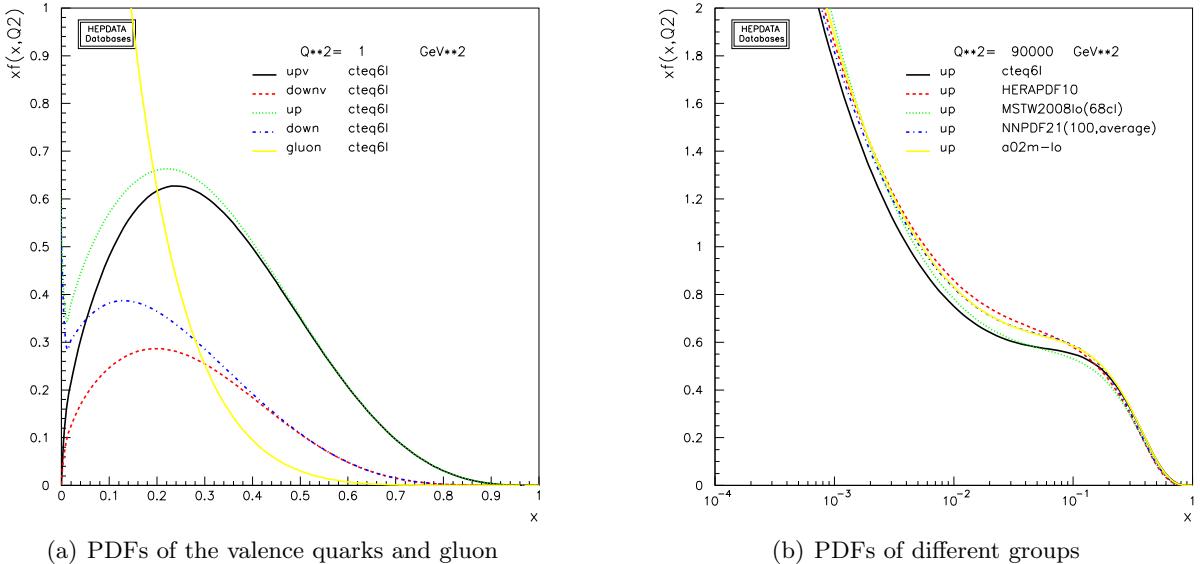


Figure 5.3.: The LO PDFs of the valence quark types and gluons (a) from the CTEQ group at $Q^2 = 1 \text{ GeV}^2$ and of the up quark (b) obtained from the CTEQ (black), HERAPDF (red), MSTW (green), NNPDF (blue) and ABM (yellow) at $Q^2 = 300^2 \text{ GeV}^2$.

The momentum fraction of the proton carried by an individual quark flavour or gluon can be obtained by integrating $xf(x, Q^2)$ over the entire Bjorken x range. The momentum fraction carried by the valence quarks is just above 50% (36% by the two up quarks and 18% by the down quark), while a large fraction of the momentum is carried by the gluons interacting between the quarks. Since the gluon propagator is anti-proportional to the squared momentum transfer, gluons with large momenta are suppressed and the sea-quarks tend to have low values of x . The structure of the proton cannot be calculated from first principles using perturbation theory, due to the large coupling constant α_S of the QCD. Hence the PDFs need to be determined experimentally using data from deep inelastic scattering or hadron-hadron collisions. There are several groups worldwide providing PDF sets by extracting them from fits to various experimental data. The LO PDF of the up quark, provided by the most prominent groups

CTEQ [148, 159], HERAPDF [189], MSTW [190, 191], NNPDF [192] and ABM [193], are shown for an energy scale of $Q^2 = (m_{\text{top}} + m_H/2)^2 \approx 300^2 \text{ GeV}^2$ in Figure 5.3 (b). This energy scale is a typical static scale choice for $t\bar{t}H$ production simulations. The groups agree within 20% variation to each other depending on the Bjorken x and the chosen energy scale. The largest discrepancies are visible between the PDF sets of the CTEQ and HERAPDF groups. For the MEM calculation the PDFs provided by the CTEQ group are utilised, because they are most commonly used in the MC simulations. Since only matrix elements with LO precisions are employed, the set with LO accuracy CTEQ6L is chosen. The full set of these PDFs for all proton constituents is shown in Figure 5.4 (a) again for the same $Q^2 = 300^2 \text{ GeV}^2$ on a double-logarithmic scale for Bjorken $x > 10^{-4}$.

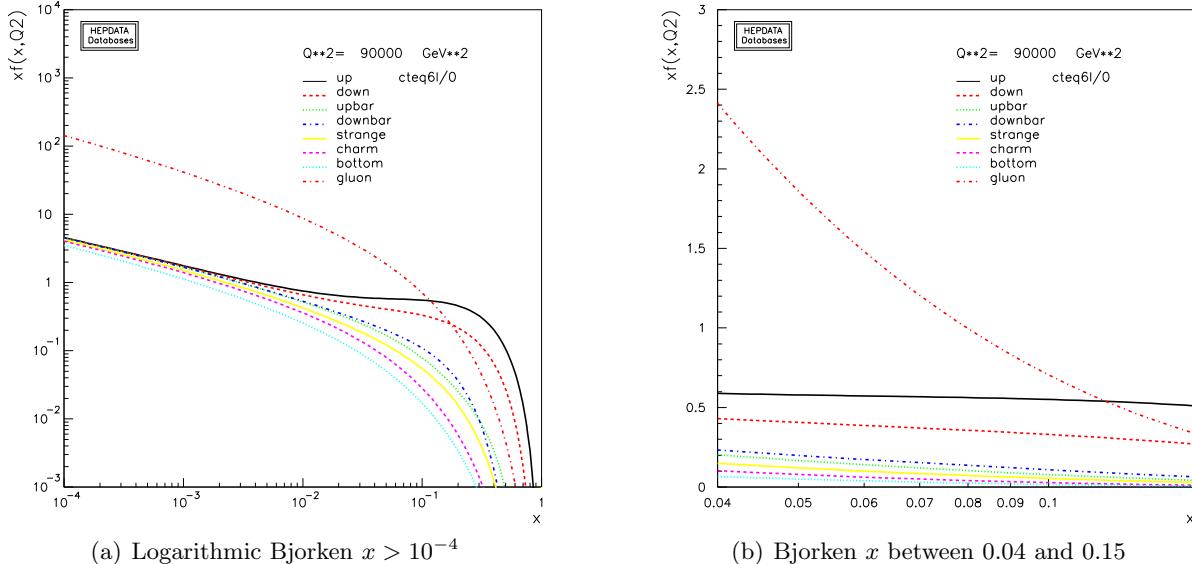


Figure 5.4.: The parton distribution functions of the CTEQ6L set for all proton constituents

The heavy quarks charm and bottom contribute only at very low x to the total momentum of the proton and thus are generally negligible in the production of high mass systems such as the $t\bar{t}H$ system. Nevertheless the contributions of all proton constituents are used in the calculation of the process probabilities. Assuming that both initial particles collide with the same momentum fraction $x_1 = x_2 = x$ at a centre-of-mass energy of $\sqrt{s} = 8 \text{ TeV}$, $t\bar{t}$ events are produced with a minimal Bjorken $x_{\min} = \frac{2m_{\text{top}}}{8 \text{ TeV}} \simeq 0.044$, while for $t\bar{t}H$ events at least a $x_{\min} = \frac{2m_{\text{top}}+m_H}{8 \text{ TeV}} \simeq 0.059$ is needed due to the additional Higgs mass in the system. The same PDFs are shown in Figure 5.4 (b) on a non-logarithmic scale in a range of $x \in [0.04, 0.15]$, which is the most relevant scale of the production of the two processes. Since $t\bar{t}$ events can be produced at lower x , where the gluon PDF dominates, the ratio of $gg/q\bar{q}$ -production is roughly 80/20 at 8 TeV. However, the gluon PDF drops rapidly, while the quark PDFs change only marginally with increasing x . As a consequence, $t\bar{t}H$ events, which need a higher x , are only produced in roughly 60% of the cases via gluon fusion and in 40% via $q\bar{q}$ -annihilation. Furthermore, at moderate $x \in [0.01, 0.5]$ there is only a very weak dependence on the energy scale Q^2 , since scaling violations appear only at very low or very high Bjorken x . As an example the LO CTEQ PDFs are shown for a Bjorken $x = 0.05$ varied for energy scales $Q^2 \in [10, 10^9] \text{ GeV}^2$ in Figure 5.5 (a).

5. The Matrix Element Method

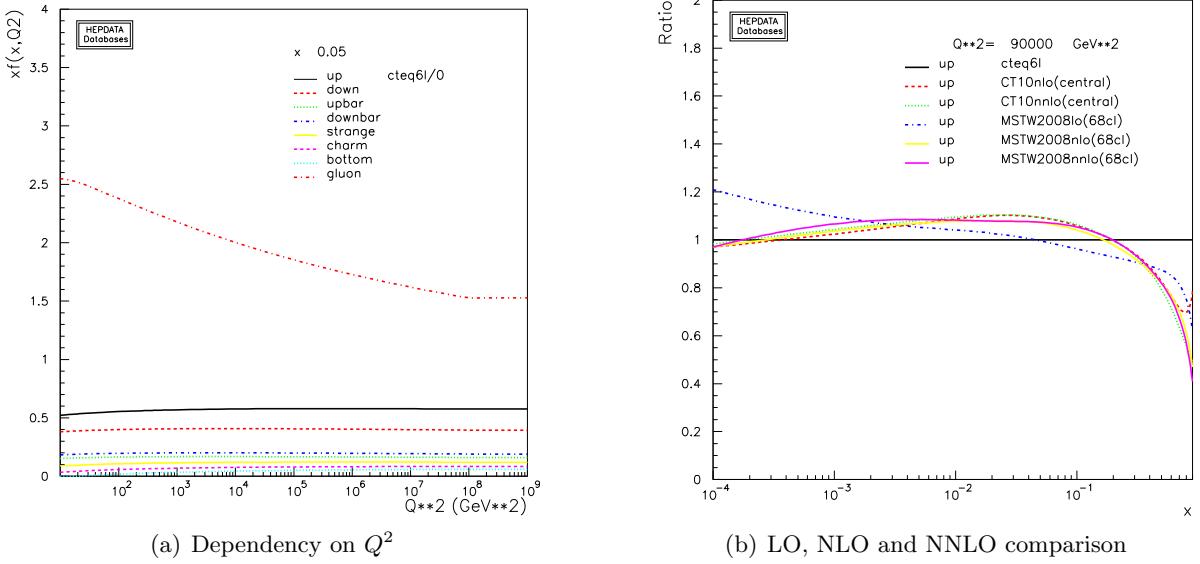


Figure 5.5.: The parton distribution functions of the CTEQ6L set dependent on Q^2 (a) and the up quark PDF compared to higher-order precision and the one from the MSTW group (b).

As expected, there is almost no dependency visible for the quark PDFs and also the gluon PDF decreases only slightly over the large range of Q^2 . Finally, up quark PDFs using different fitting models with LO, NLO and NNLO accuracy are compared for the sets from the CTEQ and MSTW group in Figure 5.5 (b). The Figure shows the ratio of the PDFs with higher-order precision or from the alternative group and the default CTEQ6L PDF. While NLO and NNLO agree very well within one group and also quite well between the groups for $x > 0.01$, there are deviations of up to 15% between the PDFs with LO and higher-order accuracy. The two LO PDFs have been already compared in Figure 5.3 (b). At $x > 0.6$ the values of the PDFs are so small that tiny deviations start to result in large variations in the ratio.

5.5. The Matrix Elements

The transition matrix element defines the hypothesis of the process probability as given in Equation 5.21. The matrix element can be formulated from the Feynman diagram of the hard scattering process by applying the Feynman rules. However, the calculation of the squared matrix element of complex processes such as $t\bar{t}H$ or its dominant background $t\bar{t} + b\bar{b}$ is already at tree-level very challenging because of many contributing diagrams. In the definition of the signal and background hypothesis of this analysis are thus only LO diagrams considered. For simplicity, all diagrams are produced in the four-flavour scheme. In the diagrams of the signal as well as of the background hypothesis it is required to have a top quark pair as intermediate state resulting in exactly four b -quarks, two light quarks, one charged lepton (electron or muon) and one neutrino in the final state. By assuming lepton universality and invariance under charge conjugation, diagrams of only one lepton flavour and of only negative charge (electron) need to be considered.²

5.5.1. Signal Hypothesis

In the signal hypothesis it is required that a Higgs boson is radiated off one of the top quarks, which is also known as the Higgs boson production in association with a top quark pair. Hence no coupling of the Higgs boson to the W boson is accepted in the diagrams to allow for a consistent treatment when performing the kinematic transformation. The Higgs boson is then further required to decay into a pair of b -quarks, while the top quark pair decays into the single lepton final state. Neglecting the decays of the $t\bar{t}H$ intermediate state, there are four types of production diagrams, one via $q\bar{q}$ annihilation (a) and three via gluon fusions (b, c, d) as shown in Figure 5.6.

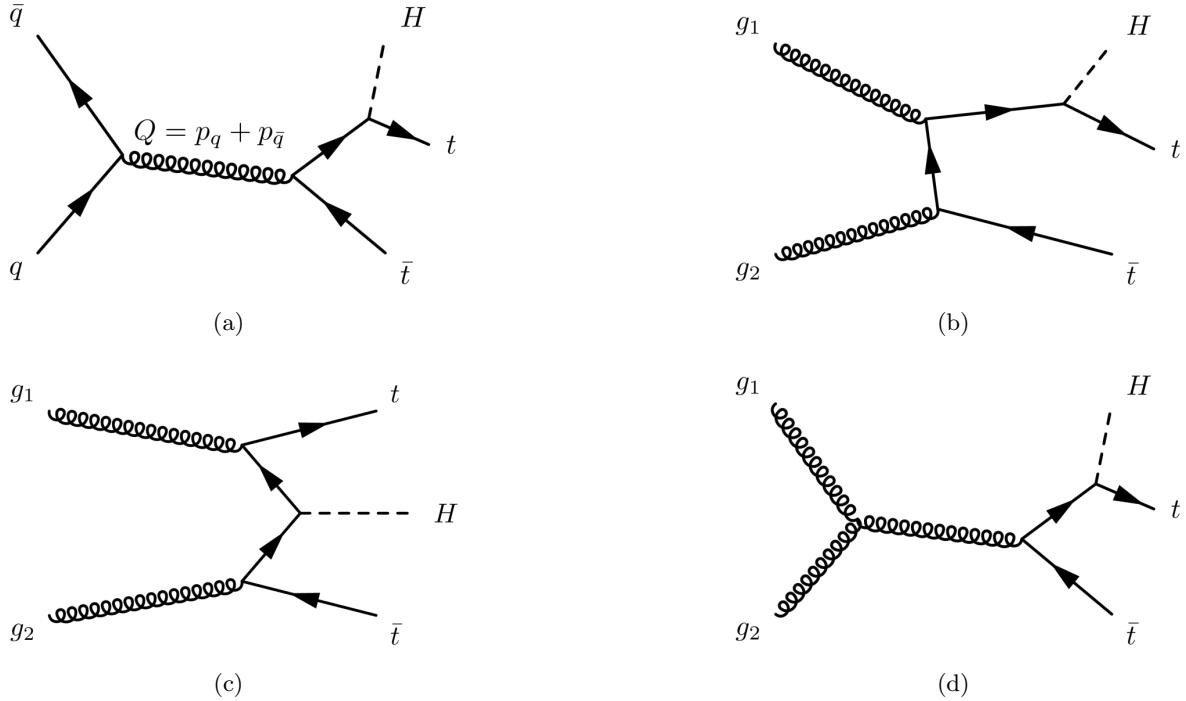


Figure 5.6.: The four basic types of production diagrams of the $t\bar{t}H$ process. By interchanging fermions and antifermions new diagrams of the same type are obtained.

The analytical expression of the squared matrix element of the simplest diagram via $q\bar{q}$ annihilation (see Fig. 5.6 (a)) is given by [194]

$$\begin{aligned}
 |\mathcal{M}_{q\bar{q} \rightarrow t\bar{t}H}|^2 &= \frac{32}{(2p_H \cdot p_t + m_H^2)(2p_H \cdot p_{\bar{t}} + m_H^2)} \\
 &\left\{ Q^2(Q \cdot p_H)^2 \left[1 + \frac{(4m_t^2 - m_H^2)Q^2}{(2p_H \cdot p_t + m_H^2)(2p_H \cdot p_{\bar{t}} + m_H^2)} \right] \right. \\
 &+ \left(\frac{Q^2}{2}m_t^2 - 2p_t \cdot p_q p_t \cdot p_{\bar{q}} \right) \left[(Q^2 + m_H^2 - 4m_t^2) + \frac{2Q \cdot p_H (4m_t^2 - m_H^2)}{(2p_H \cdot p_t + m_H^2)} \right] \quad (5.27) \\
 &+ \left(\frac{Q^2}{2}m_t^2 - 2p_{\bar{t}} \cdot p_{\bar{q}} p_{\bar{t}} \cdot p_{\bar{q}} \right) \left[(Q^2 + m_H^2 - 4m_t^2) + \frac{2Q \cdot p_H (4m_t^2 - m_H^2)}{(2p_H \cdot p_{\bar{t}} + m_H^2)} \right] \\
 &- \left. (Q^2 + m_H^2 - 4m_t^2) (2p_t \cdot p_q p_{\bar{t}} \cdot p_{\bar{q}} + 2p_t \cdot p_{\bar{q}} p_{\bar{t}} \cdot p_q - Q^2 p_t \cdot p_{\bar{t}}) \right\}
 \end{aligned}$$

5. The Matrix Element Method

which is already a complex and complicated expression and the decay of the $t\bar{t}H$ system is not yet contained. Even if the following decay is factorised, which means the intermediate particles contribute each with a propagator $\propto [(q^2 - M^2)^2 - M^2\Gamma^2]^{-1}$ and with a decay factor $\propto \Gamma^{-1} \frac{d\Gamma}{d\Omega}$, an analytical expression of the resulting squared matrix element would be very complicated and it would neglect colour flow, spin-correlations and polarisations. Furthermore the calculation of the squared matrix element of the three gluon fusion diagrams of Figure 5.6 (b-d) are even more complicated since the three production diagrams interfere with each other [194]:

$$|\mathcal{M}_{gg \rightarrow t\bar{t}H}|^2 = \left| \mathcal{M}_{gg \rightarrow t\bar{t}H}^b + \mathcal{M}_{gg \rightarrow t\bar{t}H}^c + \mathcal{M}_{gg \rightarrow t\bar{t}H}^d \right|^2 \quad (5.28)$$

In order to avoid the approximations implied by factorisation and to easily obtain any kind of hypothesis for the MEM, the generically generated matrix element calculations by MADGRAPH 5 in full LO precision are used for this analysis in C++ output. The output is automatically incorporated into the MEMTOOL framework to be called during the integration phase. The ten contributing diagrams generated by MADGRAPH are presented in Figure 5.7.

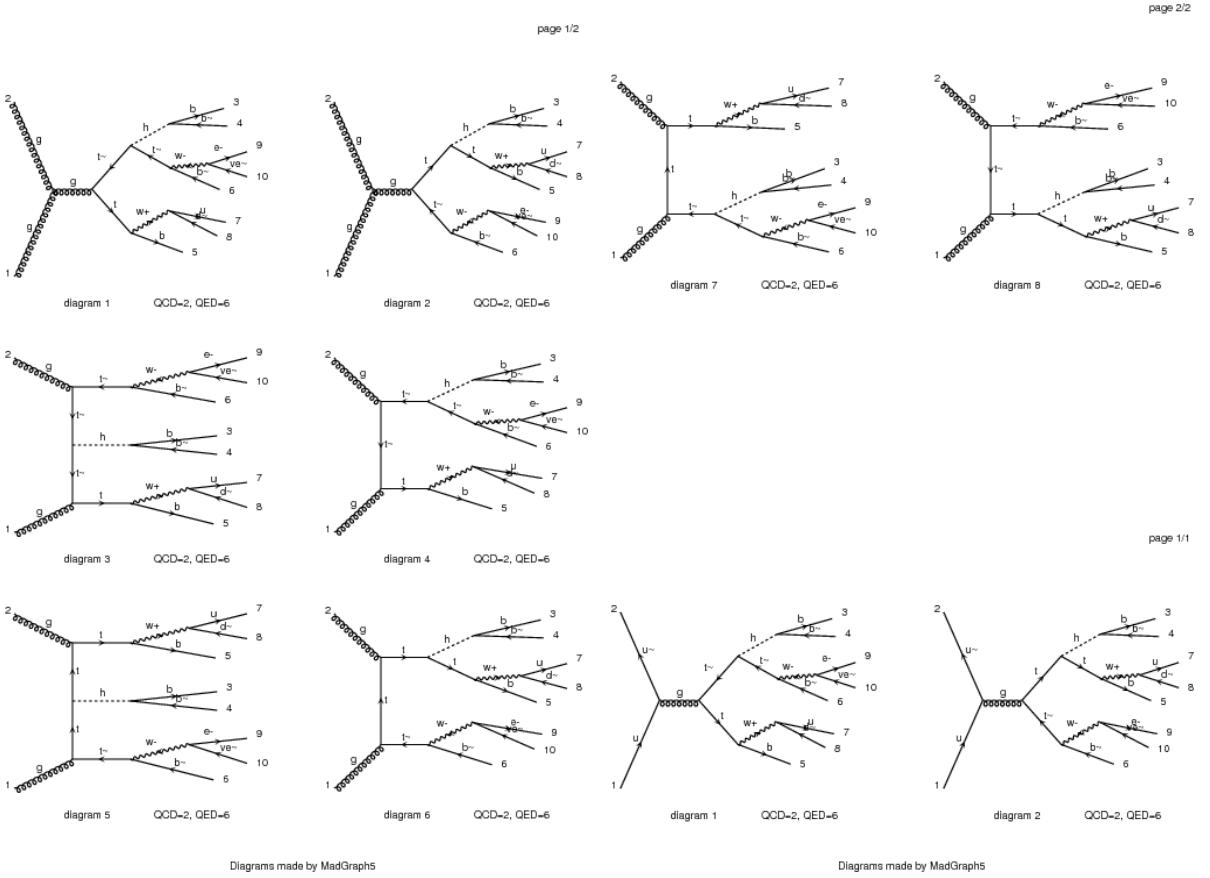


Figure 5.7.: The ten LO diagrams of the signal process $q\bar{q}/gg \rightarrow t\bar{t}H \rightarrow b\bar{b} + b\bar{b} + l\nu + qq'$. The first eight diagrams show gluon fusion production, while the last two diagrams show $q\bar{q}$ annihilation production. The diagrams and matrix element calculations are generated with MADGRAPH 5 [126].

When generating events with MADGRAPH 5, the three different types of gluon fusion production as discussed in Figure 5.6 can be distinguished in a simplified non-interfering manner

for a qualitative consideration. At $\sqrt{s} = 8$ TeV the diagrams of type (b) as realised in the diagrams 4, 6, 7, 8 of Figure 5.7 contribute with the largest fraction of around 70%. The most prominent type (c) as used in the diagrams 3 and 5 generate only approximately 24% of the events, while type (d), manifested in the diagrams 1 and 2, can be almost neglected, since it is only responsible for 6% of the events due to the three-gluon vertex.

5.5.2. Background Hypothesis

As already stated, for this analysis only the diagrams of the irreducible background process $t\bar{t} + b\bar{b}$ define the background hypothesis, since this is the most challenging background to distinguish from $t\bar{t}H$ events and is also dominant in the most signal sensitive analysis regions. However, it is expected that $t\bar{t} + jj$ events respond similarly to the hypothesis, since in most of the cases they differ only in the quark masses and can thus also be distinguished from signal events with the obtained process probability. The full LO picture with a few restrictions is used in the generation of the diagrams, obtained in the same way as for the signal hypothesis with MADGRAPH 5. Diagrams with internal b -quark lines were not allowed in the first round of studies, since they mostly contribute only with a negligible fraction to the total cross section. However, they are added in the presented final analysis for completeness and because the inclusion slightly improves the separation without increasing too much the computing time. No gluon radiation from the final state quarks is allowed, since this is kinematically suppressed and difficult to treat in any kinematic transformation aiming for phase space alignment during the integration process. The seven diagrams produced via $q\bar{q}$ annihilation are presented in Figure 5.8.

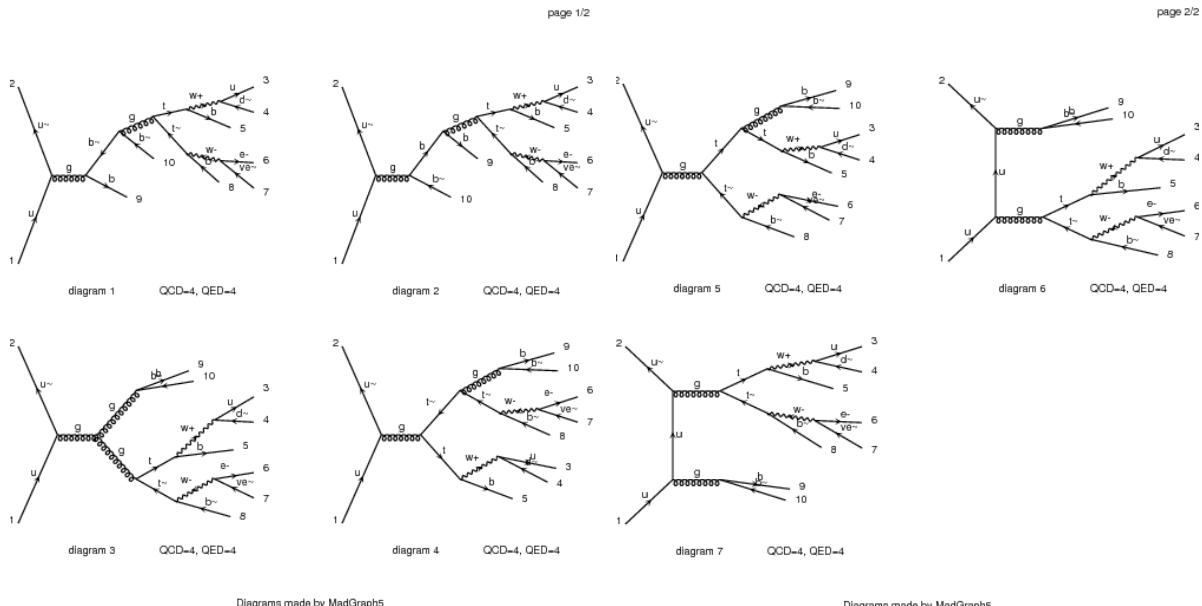


Figure 5.8.: Seven LO diagrams via $q\bar{q}$ annihilation of the background process $q\bar{q} \rightarrow t\bar{t} + b\bar{b} \rightarrow b\bar{b} + b\bar{b} + \ell\nu + qq'$. The diagrams and matrix element calculations are generated with MADGRAPH 5.

The $q\bar{q}$ annihilation diagrams can be grouped into four independent types. Type (a) (diagrams 1, 2), in which a gluon radiates off a $b\bar{b}$ production and splits into a top quark pair, is highly

5. The Matrix Element Method

suppressed ($< 10^{-4}$). Diagram 3 shows type (b), where a gluon splits into two gluons decaying into a pair of top and b -quarks. The contribution to the total production rate is almost negligible with a fraction of around 5%. Type (c) is presented in diagram 4 and 5, where the gluon radiates off one top quark similar as in the $t\bar{t}H$ production of diagram type (a). This diagram type contributes only with around 20% to the production via $q\bar{q}$ annihilation. The largest contribution of more than 70% comes from the t -channel diagrams of type (d) (diagram 6, 7), where the b -quark pair is produced from an initial state radiating gluon. The 36 $t\bar{t} + b\bar{b}$ diagrams produced via gluon fusions are shown in Figure 5.9.

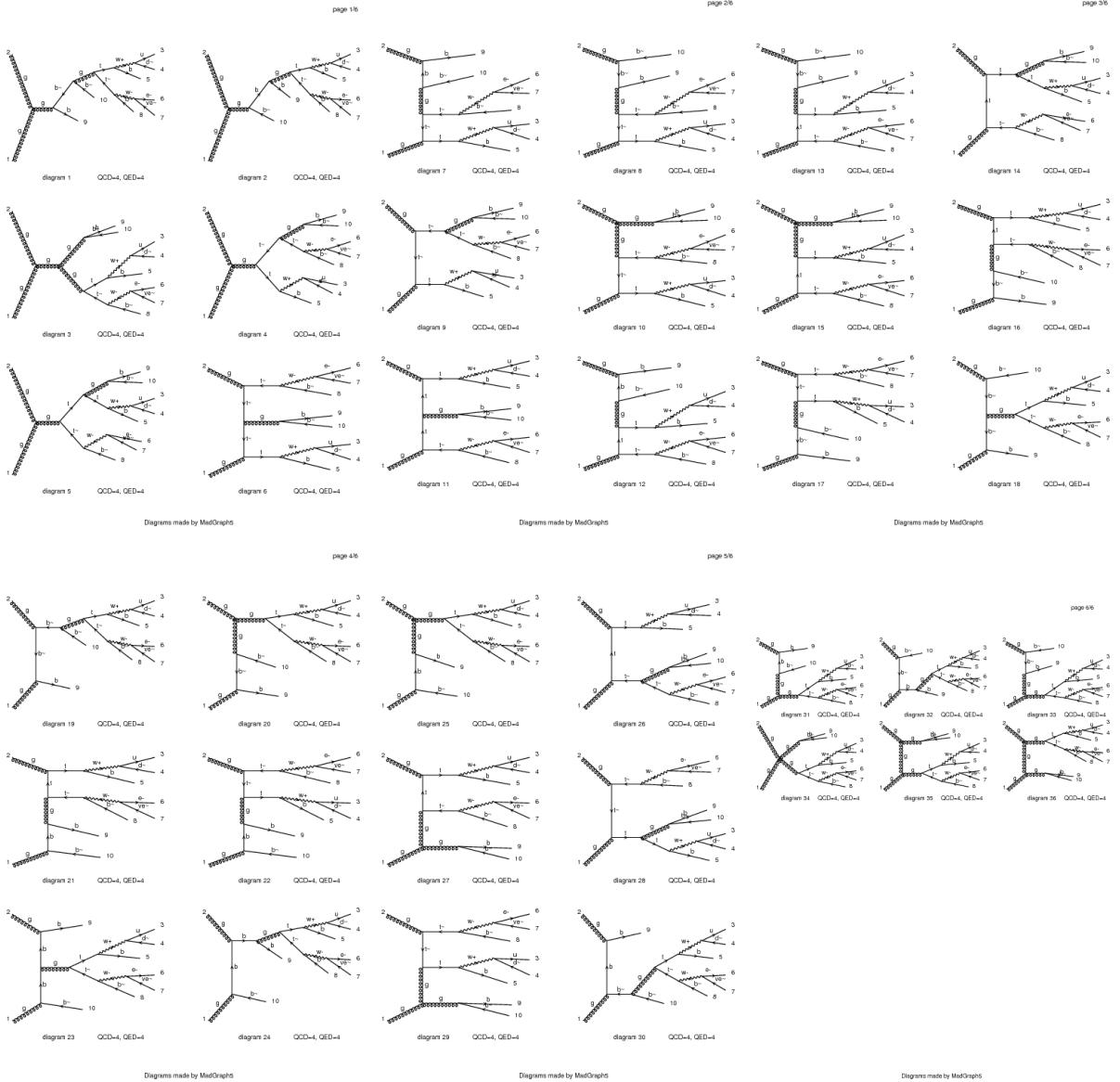


Figure 5.9.: The 36 LO diagrams via gluon fusion of the background process $gg \rightarrow t\bar{t} + b\bar{b} \rightarrow b\bar{b} + b\bar{b} + \ell\nu + q\bar{q}$. The diagrams and matrix element calculations are generated with MADGRAPH 5 [126].

Following the previous procedure ,the diagrams can be grouped into twelve different produc-

tion mechanisms. Some of them are kinematically suppressed due to internal b -quark lines radiating a gluon, which splits into a top quark pair (diagrams 1, 2, 18, 19, 23, 24, 30, 32) or contribute only negligibly due to suppressed vertex sequences (diagrams 3, 4, 5, 6). As in the $q\bar{q}$ annihilation, ISR-like diagrams (10, 15, 20, 25, 27, 29, 31, 33) contribute the most with approximately 45% to the gluon fusion production. The second largest contribution arises from the four-gluon vertex production (diagram 34) and the di-gluon production diagrams (diagrams 35, 36) contributing with 10% each.

Adding all other background processes simulated by MC (see Sec. 4.4.3), is ideally desirable and from the implementational point-of-view straight-forward, but unfortunately computationally prohibited. Studies were performed with the aim to group most of the other background process types together with a single hypothesis. This hypothesis is defined by requiring exactly one leptonically decaying W boson plus four additional b -quarks and two light quarks in the final state. A minimal set of diagrams was chosen in order to cover the production of single top, W +jets and most of the diboson events. Unfortunately, the computing time was very long and the impact on the discrimination was not satisfying enough to pursue further studies. However, $t\bar{t}V$ events, which are expected to be more signal- than background-like, might be worth a dedicated consideration in the future.

5.5.3. Calculation Optimisations

Over the course of development, many approximations were tested with the goal of optimising the signal to background discrimination while achieving reasonable computing times. In general, the approximations either decrease the time required to sample the integration phase space or serve to improve the convergence rate of the VEGAS adaptive MC integration algorithm. Since the function which calculates the squared matrix element for a certain phase space point is called thousands of times during the phase space sampling, any optimisation which can be made here is highly beneficial. When investigating possible speed improvements it turned out that the generic MADGRAPH code was not optimal. In particular the fact that the matrix element was evaluated for all possible helicity state configurations was identified as subject with the largest potential for improvements. In processes with many final state objects, such as $t\bar{t}H$ or $t\bar{t}+b\bar{b}$ events, there are many possible helicity state configurations in the matrix element calculation. To be precise, there are ten states (two initial and eight final states), each of them can have a helicity eigenvalue of +1 or -1, which leads to $2^{10} = 1024$ possible helicity permutations. Some configurations are not allowed by the underlying physics laws and result in a vanishing matrix element. These helicity state permutations are automatically evaluated and removed by MADGRAPH after the first call. For the $t\bar{t}H$ and $t\bar{t}+b\bar{b}$ process only 64 helicity permutations with non-zero contribution remain. However, in certain kinematic phase space regions some of the remaining helicity permutations are highly suppressed by physics and their contribution is almost negligible. Furthermore, in certain processes some of the configurations result in the exact same matrix element response and are thus invariant under permutation and there is no need to calculate them twice. In order to evaluate dynamically only leading helicity states of a given kinematic configuration and to avoid spending computing time on invariant helicity state permutations the following optimisation has been applied. After the first two initial calls to the matrix element calculation method, all invariant helicity states with exactly the same matrix element contribution are combined and non-physical states are excluded from the calculation. Following the next 100 calls, the remaining helicity states are ranked according to their relative contribution in the evaluated phase space. Those contributing less

than 10% of the maximum contributing helicity state are subsequently excluded from further calculation. Finally, after every further 2000 calls, states contributing less than 1% of the maximum contributing helicity state are excluded from further calculation. Studies have shown that the relative change in the likelihood response is smaller than 3%. This approximation does not perceptibly change the final separation power of the final discriminant, but reduces the per-sample calculation time by approximately a factor of 20. The relevant MADGRAPH routines are modified accordingly in a completely automated way during the implementation process of the MADGRAPH code to MEMTOOL.

5.6. Transfer Functions

The transfer functions describe the probability to observe the four-momenta of all partonic final states \mathbf{y} with the reconstructed four-momenta \mathbf{x} of the observed objects in the detector. It is assumed that the TFs can be factorised into the contributions of the k final state particles

$$W(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^k W(p_i^y|p_i^x) \quad (5.29)$$

and follow their normalisation condition

$$\int W(p_i^y|p_i^x) dp_i^x = 1 \quad (5.30)$$

to conserve the probability. Assuming that each four-momentum quantity can be factorised

$$W(p^y|p^x) = W(E^y|E^x) W(p_x^y|p_x^x) W(p_y^y|p_y^x) W(p_z^y|p_z^x) \quad (5.31)$$

and transformed to the angular variables η and ϕ , the TFs of particles with negligible mass can be written as

$$W(p^y|p^x) = W(E^y|E^x) W(\eta^y|\eta^x) W(\phi^y|\phi^x) . \quad (5.32)$$

The detector responds differently to the various particle types, i.e. light quarks, b -quarks, electrons, muons and neutrinos, and different physics objects are identified from these signatures as defined in Section 4.3. Consequently, the TFs are distinguished according to this classification and need to be extracted for each particle type. For light quarks, b -quarks, electrons and muons it is assumed that the angular variables are measured with negligible uncertainties, thus the TFs are modelled simply by a Dirac delta function

$$W(p^y|p^x) = W(E^y|E^x) \delta(\eta^y - \eta^x) \delta(\phi^y - \phi^x) . \quad (5.33)$$

Accordingly the only remaining TF to be extracted for these objects are the energy TFs, which are taken from the KLFITTER package [188]. The detector response caused by a traversing particle is in general not a symmetric distribution around the parton level true value. In particular the calorimeter shows a tendency to measure less energy due to losses, e.g. in dead material, transition regions, escaping particles or out-of-cone contributions. The energy TFs are therefore modelled by a double Gaussian function

$$W(E^y|E^x) = \frac{1}{\sqrt{2\pi}(p_2 + p_3 p_5)} \left[\exp\left(-\frac{(\Delta E - p_1)^2}{2p_2^2}\right) + p_3 \exp\left(-\frac{(\Delta E - p_4)^2}{2p_5^2}\right) \right] \quad (5.34)$$

with $\Delta E = \frac{E^y - E^x}{E^y}$ and the parameters of the double Gaussian $p_1 - p_5$. The energy losses and the energy resolution depends on the detector geometry, due to the different material and technologies of the sub detector parts, and on the energy of the partonic particle. Although these effects are partially taken into account during the calibration of the objects (see Sec. 4.3), the parameters p_i of the energy TF are derived for different $|\eta|$ regions and are parametrised depending on the true energies E^y . In the case of the muon, the transverse momentum p_T instead of the energy E is measured and is therefore used in the transfer functions.

The TFs are extracted from $t\bar{t}$ events generated with POWHEG and showered with PYTHIA as described in Section 4.4.3. Every partonic final state is required to match bi-uniquely to the corresponding reconstructed object within a distance of $\Delta R < 0.3$. A jet matched to a b -quark is treated as a b -jet independently on the b -tagging identification. The parametrisations of the parameters p_i are partially motivated by physics but are optimised for the best fit result and summarised in Table 5.1.

Table 5.1.: The parametrisation of the TF parameters p_i as a function of the partonic energy E^y in case of light quarks, b -quarks and electrons and as a function of the partonic p_T^y in case of muons.

p_i	light quarks	b -quarks	electrons	muons
p_1	$a_1 + b_1/E^y$	$a_1 + b_1/E^y$	$a_1 + b_1 E^y$	$a_1 + b_1 p_T^y$
p_2	$a_2 + b_2/\sqrt{E^y}$	$a_2 + b_2/\sqrt{E^y}$	$a_2 + b_2/\sqrt{E^y}$	$a_2 + b_2 p_T^y$
p_3	$a_3 + b_3/E^y$	$a_3 + b_3/E^y$	$a_3 + b_3 E^y$	$a_3 + b_3 p_T^y$
p_4	$a_4 + b_4/\sqrt{E^y}$	$a_4 + b_4/\sqrt{E^y}$	$a_4 + b_4/\sqrt{E^y}$	$a_4 + b_4 p_T^y$
p_5	$a_5 + b_5 E^y$	$a_5 + b_5 E^y$	$a_5 + b_5 E^y$	$a_5 + b_5 p_T^y$

In the case of quarks and electrons, the parameter p_2 , corresponding to the width of the main Gaussian, is parametrised according to the calorimeter resolution. In the case of the muon, where the momentum resolution is directly proportional to the transverse momentum p_T according to the *Glückstern formula* [97], a linear parametrisation is used. The $|\eta|$ binning is chosen motivated by the detector geometry resulting in the interval borders [0, 0.8, 1.37, 1.52, 2.5] for quarks and electrons and [0, 1.11, 1.25, 2.5] for muons. The binning for the true partonic energy (transverse momentum) is chosen depending on the statistics within a single bin in order to minimize statistical fluctuations. The minimum and maximum energy (transverse momentum) depend on the objects and on the $|\eta|$ -bin. First, the ΔE (Δp_T) distributions are fitted with a double Gaussian function locally for each object and bin in $|\eta|$ and partonic energy (transverse momentum). Secondly, the obtained parameters p_i are plotted depending on the partonic energy (transverse momentum) and are fitted with the described parametrisations to obtain the ten parameters a_i, b_i . Last, in each bin of $|\eta|$ a two-dimensional fit in ΔE (Δp_T) and E^y (p_T^y) is performed to obtain the globally defined TFs using the parameters a_i, b_i as starting value. As an example, the resulting parameters for the light and b -quarks within $0.8 < |\eta| < 1.37$ are summarised in Table 5.2. Figure 5.10 shows the MC distributions of ΔE (Δp_T) for light quarks, b -quarks, electrons and muons with the local (brown) and global (red) double Gaussian fit for the central $|\eta|$ region and for energies (transverse momenta) around 100 GeV. The separate components of the two single Gaussian (green and blue) of the local fit are also shown.

The distributions for the quarks show clearly an asymmetric tail on the right side and a large contribution of the second Gaussian (blue) due to losses in the calorimeter and during the jet

5. The Matrix Element Method

Table 5.2.: Globally fitted parameters for the transfer functions of light quarks and b -quarks in the range of $0.8 < |\eta| < 1.37$. The units for a_2 and b_2 can be identified from Table 5.1.

light quarks			b-quarks	
p_i	a_i	b_i	a_i	b_i
p_1	-0.009 ± 0.001	1.1 ± 0.1	-0.013 ± 0.001	5.3 ± 0.2
p_2	0.047 ± 0.002	0.70 ± 0.02	0.055 ± 0.002	0.55 ± 0.03
p_3	-0.088 ± 0.003	46.1 ± 1.1	-0.141 ± 0.008	102.5 ± 2.2
p_4	0.50 ± 0.01	-4.48 ± 0.07	0.39 ± 0.01	-2.89 ± 0.04
p_5	0.201 ± 0.002	$4.2 \pm 0.2 \times 10^{-4}$	0.223 ± 0.001	$0.7 \pm 0.1 \times 10^{-4}$

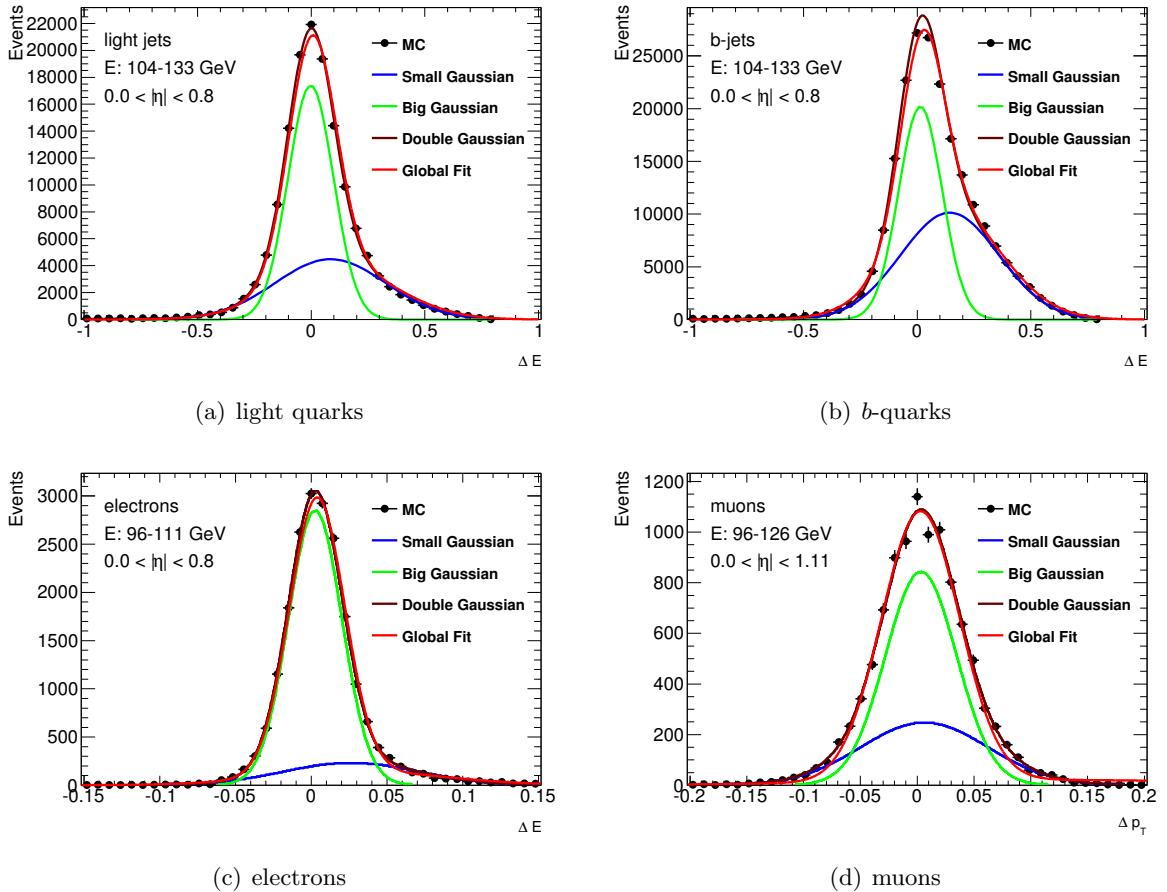


Figure 5.10.: The ΔE (Δp_T) distributions for quarks and leptons and the corresponding local (brown) and global (red) double Gaussian fit and the subcomponents (green, blue) to extract the TF are shown for central $|\eta|$ regions and energies (transverse momenta) around 100 GeV.

reconstruction. As expected, the b -quark TF is significantly broader than the light quark TF due to the higher mass of the decaying b -hadron resulting in a wider shower profile. The global fit (red) differs slightly from the best local fit (brown), but still agrees well with the MC simu-

lation. The electron energy TF has only a small contribution from the second Gaussian term on the right side and is much narrower, since the losses in the EM calorimeter are significantly smaller. The muon transverse momentum TF is more symmetric, since the energy losses of the muon passing the detector are negligible. However, the second Gaussian term contributes to broaden the side-bands. At this energy the muon TF is also slightly broader than the electron TF. The global fits of both lepton TF are very similar to the local fits and agree well with the MC predictions. In Figure 5.11 the evolution of the light quark (a) and b -quark (b) energy TFs for partonic energies E^y in a range of [50, 650] GeV in steps of 50 GeV are shown in the central region of $0.0 < |\eta| < 0.8$ (top) and in a more forward region of $1.52 < |\eta| < 2.50$ (bottom).

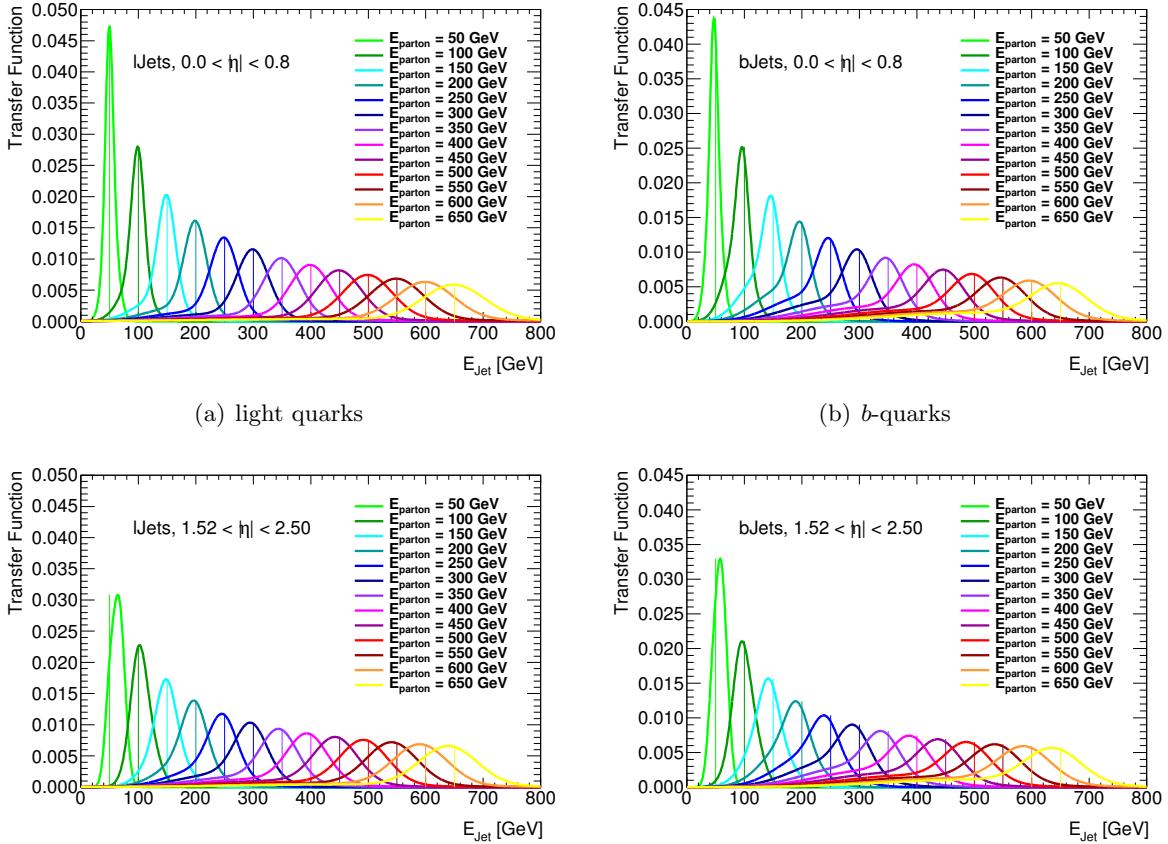


Figure 5.11.: The evolution of the light quark (a) and b -quark (b) energy TFs for central (top) and more forward (bottom) $|\eta|$ regions.

At very low energies the b -quark energy TFs are comparable to the light quark energy TFs. However, they become much broader at very high energies and develop a very long tail to lower energies due to losses from escaping muons and neutrinos from semileptonic b -hadron decays. In the forward region, the dead material in the upstream and the coarser granularity of the hadronic calorimeter results in significantly broader TFs of the quarks. The light quark TFs are in particular in the central detector region quite symmetric and the most likely measured value tends to correspond to the partonic energy. On the contrary, the b -quark TFs start to become very asymmetric at energies around 200 GeV and the central value of the double

5. The Matrix Element Method

Gaussian can be significantly lower than the true energy at parton level. Figure 5.12 shows the evolution of the electron (a) and muon (b) energy and transverse momentum TFs for the same partonic energies (transverse momentum) and similar $|\eta|$ regions.

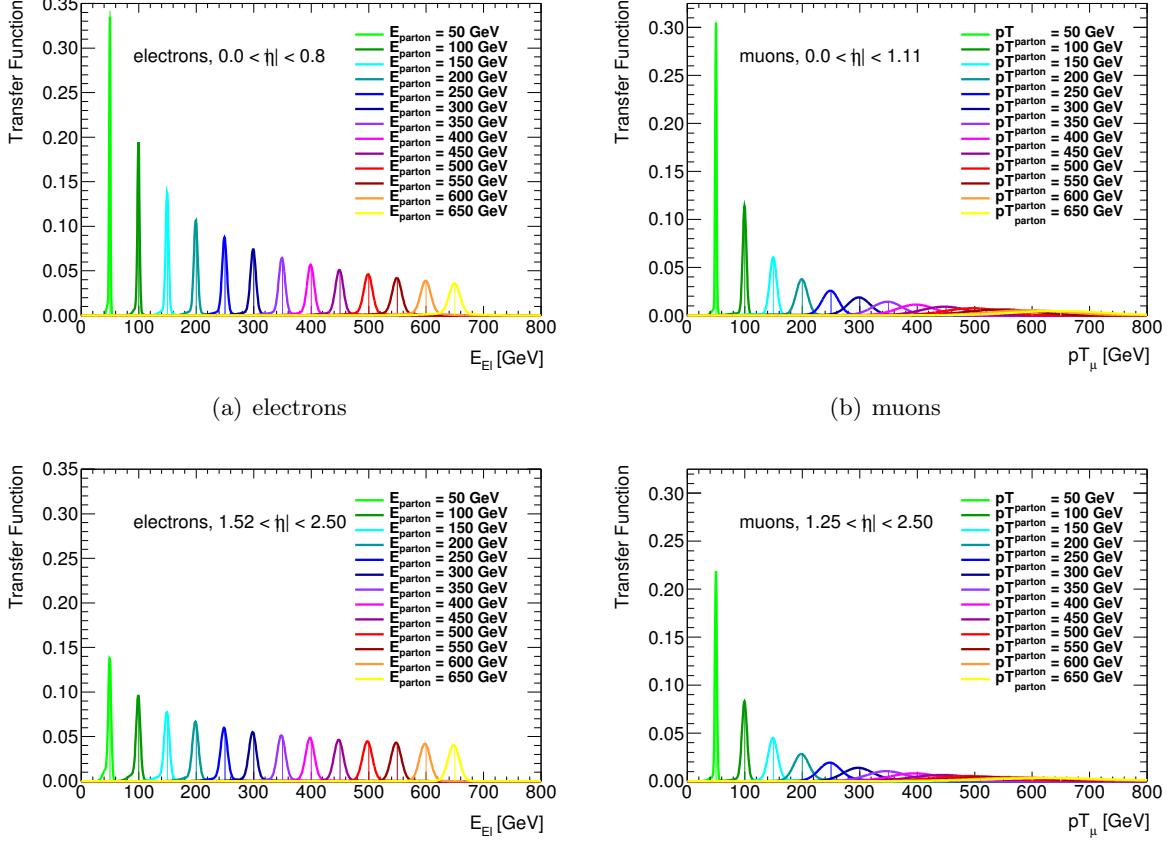


Figure 5.12.: The evolution of the electron energy TF (a) and muon transverse momentum TF (b) for central (top) and more forward (bottom) $|\eta|$ regions.

The resolutions of the electron energy TF are very sharp even for extremely high energies, but get slightly worse in the more forward region of the EM calorimeter due to dead material and coarser granularity. At low energies they tend to be slightly asymmetric with a tail to lower energies due to more significant energy losses in the ID, which are not relevant at higher energies. Although at lower energies the muon transverse momentum TF are comparable to the electron energy TF, they are significantly broader for $p_T > 250$ GeV since the resolution is directly proportional to the transverse momentum [97]. Because the momentum measurement is only marginally influenced by the material in front of the MS and the components of the MS perform similar in higher $|\eta|$ regions, the resolution degrades not as much as for electrons. In the case of the escaping neutrino, no energy or momentum can be measured directly in the detector. However, in events where the neutrino is expected to be the only invisible particle to the detector, the missing transverse momentum observed in the detector can be related to the neutrino transverse momenta. The E_T^{miss} definition as defined in Section 4.3.4 corrects for detector effects and objects not originating in the hard scattering process. Hence the x - and

y -components of the E_T^{miss} describe in good approximation the neutrino momentum in x - and y -direction. The four-momentum TF of the neutrino of Equation 5.31 can be written as

$$W(p^y|p^x) = W(p_x^y|E_x^{\text{miss}}) W(p_y^y|E_y^{\text{miss}}) \quad (5.35)$$

and the x - and y -components are expected to follow a single Gaussian function dependent on $\Delta p_{x/y} = E_{x/y}^{\text{miss}} - p_{x/y}^y$

$$W(p_{x/y}^y|E_{x/y}^{\text{miss}}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\Delta p_{x/y}^2}{\sigma^2}\right). \quad (5.36)$$

The Gaussian width σ depends on the total energy deposited in the detector and is hence parametrised with a generic Sigmoid function of the scalar sum of all deposited energy $\sum E_T$

$$\sigma(\sum E_T) = p_0 + \frac{p_1}{1 + \exp[-p_2(\sum E_T - p_3)]}. \quad (5.37)$$

However, it does not depend on the detector geometry because the objects used for the E_T^{miss} reconstruction are distributed over the entire detector. Similar to the parameter extraction for the double Gaussian functions, the width σ is obtained in different $\sum E_T$ regions by fitting the MC distributions of $\Delta p_{x/y}$ with a single Gaussian function. Then the width is plotted depending on $\sum E_T$ in order to obtain the parameters of the Sigmoid function. No differences between the x - and y -direction are observed and accordingly the same parametrisation is used for both directions. The evolution of the neutrino TF for values of $\sum E_T = 200, 400, 600, 800, 1000$ and 1200 GeV is shown in Figure 5.13.

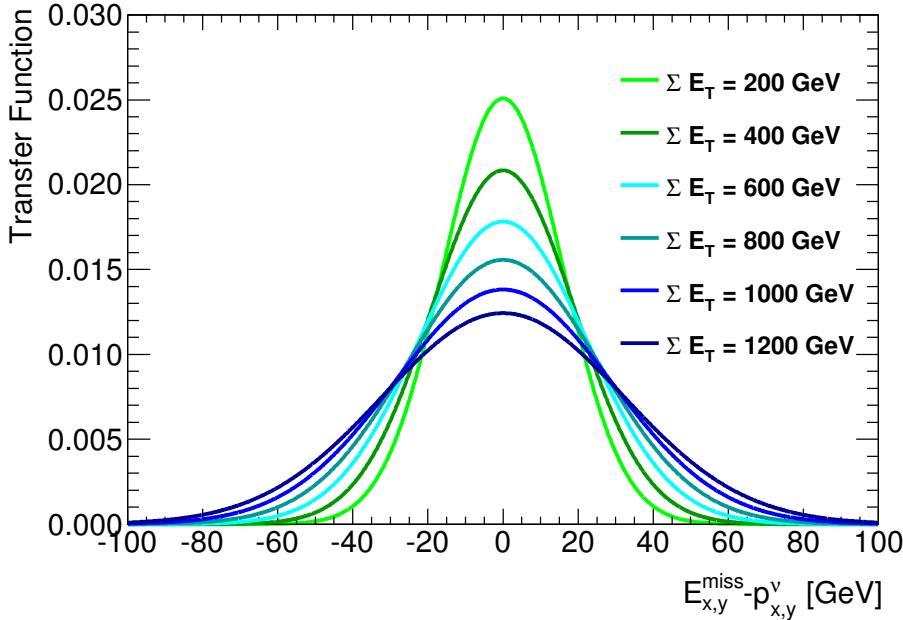


Figure 5.13.: The evolution of neutrino momentum TF in x - and y -directions.

Apart from very high and very low $\sum E_T$ the Sigmoid function models a linear function, hence the width of the resolution behaves almost linear with increasing $\sum E_T$. Already at low $\sum E_T$ the momentum resolution is relatively broad with a width of around 15 GeV.

5.7. Integration over the Phase Space

5.7.1. Integration Algorithm

MC integration techniques are usually preferred for higher dimensional integrals, because of their faster convergence rate compared to classical numerical integration techniques. They use random numbers instead of fixed grids to define the sample points at which the integrand is evaluated. For a sufficiently large number N of sample points, any d -dimensional integration over the volume V of a function $f(\mathbf{x})$ can be approximated by

$$I = \int d^d \mathbf{x} f(\mathbf{x}) \approx \frac{V}{N} \sum_{n=1}^N f(\mathbf{x}_n). \quad (5.38)$$

The uncertainty $\sigma_I = \frac{V\sigma_f}{\sqrt{N}}$ on the integral I depends on the standard deviation σ_f of the function and scales like $\frac{1}{\sqrt{N}}$ independent on the dimension d of the integral. Although this is a huge advantage compared to classical methods, in which the convergence rate usually scales exponentially with the dimensions, in practice the integration is still relatively slow. However, there are several techniques aiming to reduce the uncertainty estimate. Instead of sampling the integration volume uniformly, certain algorithms divide the integration space into subspaces (*stratified sampling*) or change the integration variables (*importance sampling*) to probe the important integration volume faster.

For this analysis the VEGAS algorithm [195] provided by the GSL 1.16 library [187] is chosen for the phase space integration. Its realisation in the C programming language simplified the development phase and also provides a modest performance boost compared to the FORTAN version. VEGAS is an adaptive MC integration algorithm combining the basic concepts of importance and stratified sampling. The latter is manifested in the approach that VEGAS splits the integration space into hypercubes, in which the integration is performed. The idea of importance sampling is to identify the important regions of the integrand and sample in these regions with a higher density. By introducing a probability density function $p(\mathbf{x}) = \frac{dP(\mathbf{x})}{d\mathbf{x}}$, the integrand $f(\mathbf{x})$ of Equation 5.38 can be rewritten to

$$I = \int d\mathbf{x} f(\mathbf{x}) = \int dP(\mathbf{x}) \frac{f(\mathbf{x})}{p(\mathbf{x})} \approx \frac{V}{N} \sum_{n=1}^N \frac{f(\mathbf{x}_n)}{p(\mathbf{x}_n)}. \quad (5.39)$$

As a consequence, the random numbers are now generated according to the prior distribution of $p(\mathbf{x})$. An estimator for the statistical uncertainty on the integral can be obtained from the variance

$$\sigma_I^2 = \frac{V^2}{N} \sum_{n=1}^N \left(\frac{f(\mathbf{x}_n)}{p(\mathbf{x}_n)} \right)^2 - I^2. \quad (5.40)$$

For fast convergence, it is desired to choose $p(\mathbf{x})$ as close in shape as possible to the function $f(\mathbf{x})$. The ideal choice would be $p(\mathbf{x}) = \frac{f(\mathbf{x})}{I}$, however the knowledge of I is the goal of the integration and thus unknown. VEGAS starts hence with a uniformly distributed $p(\mathbf{x})$, evaluates the integrand at a certain number of random points and maintains contribution histograms along each of the integration dimensions. After each iteration, the histograms from the previous iteration are used to model $p(\mathbf{x})$ in order to concentrate the sampling

in regions with the highest contributions. This is achieved by adjusting the edge length of the hypercubes according to their contribution, but keeping the number of sample points in each hypercube constant. Consequently, the phase space with the highest contribution is contained in the smallest hypercube and thus probed with the finest granularity. Each subsequent iteration refines $p(\mathbf{x})$ by adjusting the hypercube grid and hence improves the estimate I and reduces its uncertainty. However, the prior probability and thus the grid of the hypercubes is usually stiffened after a warm-up phase. This allows for evaluating the integral faster and with higher precision afterwards by using more sample points and ignoring the estimates from the warm-up phase. In order to avoid d -dimensional histograms with exponentially growing bin numbers, it is assumed that the prior probability factorizes in each dimension corresponding to $p(\mathbf{x}) = p(x_1)p(x_2)\dots p(x_d)$. This ensures linearly growing number of bins, but also means that the peak structures of the integrand need to be well separated to be projected onto the coordinate axes of the integration variables. This can be achieved by a transformation of the integration variables to quantities which are well aligned with the peak structures of the integrand (see Sec. 5.7.3).

VEGAS utilizes simultaneously performed and statistically independent iterations in order to calculate a χ^2 measure of convergence, which gives a stronger constraint on the estimated uncertainty. A combined estimate \hat{I} is calculated from the weighted average of the m iterations

$$\hat{I} = \left(\sum_{j=1}^m \frac{N_j I_j}{\sigma_{I_j}^2} \right) \left(\sum_{j=1}^m \frac{N_j}{\sigma_{I_j}^2} \right)^{-1} \quad (5.41)$$

and the χ^2 per degree of freedom (dof) represents the consistency between the estimates

$$\chi^2/\text{dof} = \frac{1}{m-1} \sum_{j=1}^m \frac{(I_j - \hat{I})^2}{\sigma_{I_j}^2}. \quad (5.42)$$

Finally, the modularity of this VEGAS implementation allows saving and restoring the state of an interrupted integration. This property is primarily utilised when ordering and reducing the number of assignment permutations according to their likelihood contribution (see Sec. 5.7.1 and Sec. 5.7.2).

Integration Optimisations

Adaptive MC integrations converge at rates between \sqrt{N}^{-1} and N^{-1} depending on the ability of the adaptive method to approximate the inverse integrand distribution $p(\mathbf{x})$. In the worst case scenario, the number of sample points N needs to be increased by a factor of four in order to reduce the uncertainty on the estimate by a factor of two. Thus defining the allowed upper bound of the estimate uncertainty significantly impacts the run time of the integration.

For this analysis, the integration procedure has been optimised as follows. Before any actual evaluation of the integral, a plain MC integration algorithm is used as a pre-run to sample the phase space uniformly at 100 points in order to evaluate the contributing helicity states of the matrix element calculation as described in Section 5.5.3. Afterwards the warm-up phase of the VEGAS integration is initialised, by using three parallel iterations with 2048 sample points each to determine $p(\mathbf{x})$. During that phase the integrals of all possible assignment permutations are evaluated and the permutations are ordered according to their relative contribution to the summed likelihood. After that the integration state of each permutation in descending order is

5. The Matrix Element Method

restored and the grid of the hypercubes is stiffened. Using this grid but neglecting the result of the warm-up phase, the full integration is performed starting again with three iterations and 2048 sample points. The integration is considered to have converged if the relative uncertainty on the estimate is less than 10% and the estimates of the different iterations are consistent (i.e. satisfy $|\chi^2 - 1| < 0.5$). If both of the conditions are not achieved with the current integral estimate, the number of iterations is increased by one, the number of sample points is doubled and the procedure is repeated until the integration converges or a maximum number of iterations is reached. If the uncertainty is acceptable, but the estimates of the different iterations are not consistent, only the number of iterations is increased but the sample points are kept constant. In addition, to avoid wasting time on the convergence of integrals which contribute only negligibly ($< 10^{-4}$) to the total likelihood or by far less than the highest ranked permutation, the convergence conditions are loosened with descending order of permutations.

5.7.2. Reduction of Assignment Permutations

Due to the large number of final state quarks in the described matrix element of the $t\bar{t}H$ or $t\bar{t} + b\bar{b}$ processes, there are many possible ways to assign the reconstructed jets to the six quarks of the matrix element calculation. Even if only six jets are considered in the calculation, without any further information there are $6! = 720$ possible assignment permutations for which the integral needs to be evaluated for every single event. This blows up the integration time, so it is desired to reduce the number of assignment permutations. For this analysis, only six out of the n reconstructed jets are chosen: the four jets with the highest b -tagging weight and the two jets among the remaining with an invariant mass closest to the W boson mass of 80.4 GeV. If a jet is b -tagged, it can not be assigned to a light quark in the matrix element description. In the case of more than four b -tagged jets, only the four with the highest b -tag weight are treated as b -tagged. If two jets have the same b -tag weight, priority is given to the jet with larger p_T . Assignment permutations between the two light quarks of the hadronically decaying W boson and between the two b quarks originating from the Higgs boson or gluon result in the same likelihood value and thus are not considered. Following these descriptions, there are only 12 or 36 assignment permutations for events with four or three b -tags, respectively, which need to be evaluated in the warm-up phase. In order to avoid running on more than 12 permutations, a MISER MC algorithm [196] is used with 400 sample points to evaluate the integral for each permutation and to select only the 12 with the highest contributions, which are then evaluated as described in Section 5.7.1.

5.7.3. Phase Space Alignment

Since VEGAS projects the $p(\mathbf{x})$ onto the axes of the integration dimensions, the choice of the coordinate system is crucial for the convergence speed when calculating the process probability by numerical integration. Furthermore, aligning the integration phase space with the peak structures of the integrand caused by propagator divergence allows to constrain the integration ranges, due to the knowledge of the location and decay width of the intermediate states. However, too complicated transformations result in time consuming calculations with multiple solutions, which might even increase the integration time. Finding the optimal kinematic transformation is not easy and a lot of development effort has been devoted to test different kinematic transformations and find the perfect balance between phase space alignment and simplicity in the transformations.

According to the *transformation theorem*, an integral behaves in the following way under a coordinate transformation represented by the diffeomorphism $\Phi : \Omega \rightarrow \Phi(\Omega)$

$$\int_{\Phi(\Omega)} f(\mathbf{x}) d\mathbf{x} = \int_{\Omega} |\det(J(\Phi(\mathbf{x}')))| f(\Phi(\mathbf{x}')) d\mathbf{x}' , \quad (5.43)$$

where $J(\Phi(\mathbf{x}'))$ is the *Jacobian matrix* and $\det(J(\Phi(\mathbf{x}')))$ the *Jacobian determinant* of Φ . If one of the integration variables is transformed to the invariant mass m (or squared mass) of an unstable particle with pole mass M and negligible decay width Γ , the *narrow width approximation (NWA)* can be applied by approximating the Breit-Wigner (BW) resonance

$$\frac{1}{(m^2 - M^2)^2 + \Gamma^2 M^2} \approx \frac{\pi}{\Gamma M} \delta(m^2 - M^2) , \quad (5.44)$$

which reduces the integration variable m . Thus a transformation to the masses of the intermediate states might not only serve for alignment of the propagator peaks to the integration dimensions, but can also be used to reduce the integration dimensions by one.

The invariant mass m of any particle decaying into two products with four-momenta p_1 and p_2 and negligible masses is given by

$$m^2 = (p_1 + p_2)^2 = m_1^2 + m_2^2 + 2E_1 E_2 - 2\vec{p}_1 \cdot \vec{p}_2 \approx 2E_1 E_2 (1 - \cos \theta_{1,2}) . \quad (5.45)$$

Hence the Jacobian component when transforming the energy E_1 to the mass m or squared mass m^2 is

$$J_{m,E_1} = \frac{\partial m}{\partial E_1} = \frac{\partial}{\partial E_1} (2E_1 E_2 (1 - \cos \theta_{1,2}))^{\frac{1}{2}} = \frac{m}{2E_1} \quad (5.46)$$

$$J_{m^2,E_1} = \frac{\partial m^2}{\partial E_1} = \frac{\partial}{\partial E_1} (2E_1 E_2 (1 - \cos \theta_{1,2})) = \frac{m^2}{E_1} . \quad (5.47)$$

Several kinematic transformations have been tested, in which one of the energies of the decay products of the intermediate states is transformed to the mass of its parent particle. The following change of integration variables turned out to be the optimal choice in terms of integration time and minimal bias on the resulting integral reflected in the separation of the Neyman-Pearson likelihood ratio. First, a kinematic transformation of the energy of the light quark with lower energy to the hadronically decaying W boson mass is performed. Similarly, the energy of the b -quark with lower energy is transformed to the invariant mass of the Higgs boson or gluon. Although b -quark masses are generally not negligible, compared to the Higgs boson mass ($2m_b^2 \ll m_H^2$) and considering the high jet energies at the LHC, the assumption of Equation 5.45 is still valid when aligning to the Higgs boson or gluon peak. In case of the leptonically decaying W boson, the p_z -component of the neutrino momenta instead of the charged lepton energy is used for a transformation to the squared mass m_W^2 , because that integration variable is the least constrained. Unfortunately, there are two solutions for the neutrino p_z -momentum due to the quadratic equation

$$m_W^2 = (p_\ell + p_\nu)^2 \approx 2E_\ell ((p_\nu^x)^2 + (p_\nu^y)^2 + (p_\nu^z)^2)^{\frac{1}{2}} - 2p_\ell^x p_\nu^x - 2p_\ell^y p_\nu^y - 2p_\ell^z p_\nu^z \quad (5.48)$$

5. The Matrix Element Method

and thus the integrand needs to be evaluated twice. The Jacobian component of transforming p_ν^z to m_W or m_W^2 is given by

$$J_{m_W, p_\nu^z} = \frac{\partial m_W}{\partial p_\nu^z} = \frac{E_\ell p_\nu^z - E_\nu p_\ell^z}{E_\nu m_W} \quad (5.49)$$

$$J_{m_W^2, p_\nu^z} = \frac{\partial m_W^2}{\partial p_\nu^z} = \frac{2E_\ell p_\nu^z - 2E_\nu p_\ell^z}{E_\nu} . \quad (5.50)$$

Kinematic transformations to the top quark propagator are also possible, but result in more non-vanishing entries in the Jacobian matrix and thus make the computations much more complicated. Furthermore, in case of the leptonically decaying top quark, the number of possible solutions for the neutrino p_z -momentum increases to four, each of them needs to be evaluated during the integration. Since the phase space can also be constrained making use of the TF width and no speed up in integration time could be gained, no transformations to the top quark masses are made. Although the integration phase space is not perfectly aligned to the integrand, which is not optimal for the VEGAS algorithm, this resulted in no loss in separation power in the likelihood ratio distribution.

After these transformations several combinations of NWA have been tested for the W boson and Higgs boson propagator. Since the propagator peak and the peak due to the energy TFs do not always overlap, the resulting probability is biased when using the NWA for the Higgs boson mass. It was thus decided not to use a NWA for the hadronically decaying W boson and the Higgs boson, but only apply it to the leptonically decaying W boson.

5.7.4. Reduction of Dimensionality

Having two initial and eight final state particles involved in the hard scattering process results in a high dimensional phase space which needs to be integrated. Each particle is defined by a four-momentum vector, thus there are in principle $10 \cdot 4 = 40$ integration dimensions. However, evaluating the δ -distributions of the phase space element of Equation 5.14, which ensure energy and momentum conservation of all particles and make use of the Einstein relation, reduces the number of integration dimensions significantly. The dimensionality can be further reduced by making several approximations on the particles or event properties as summarised in Table 5.3.

At hadron colliders like the LHC, it is fair to assume that the interacting initial state partons A and B have negligible masses and transverse momenta p_x and p_y compared to the longitudinal component p_z . The p_z -components of these partons can be constrained by assuming longitudinal momentum balance of all involved particles and energy conservation between the initial and final state particles of the matrix element. These assumptions follow the leading-order picture of the matrix element and neglect any additional radiation. However, studies were performed, in which any additional observed objects, not used in the matrix element calculation, boost the $t\bar{t}H$ system and are thus treated as ISR. The boost appeared slightly random, since no distinction between ISR, FSR, UE or pile-up jets could be made and acceptance cuts bias the momentum imbalance as well. Since no improvement in separation of the likelihood ratio could be obtained, additional objects are not considered in this analysis.

As already mentioned in Section 5.6, all angular quantities of the observed final state particles are assumed to be measured perfectly and are thus modelled by Dirac δ -functions. If the masses are known, only the energies of the quarks and the charged lepton remain as integration dimensions. After some studies with the chosen kinematic transformation described in

Table 5.3.: Assumptions and approximations made in order to reduce the number of integration dimensions. Masses m are given in units of GeV.

Description		Integration Dimensions			
Object	Parent	E	p_x	p_y	p_z
parton A	-	$m = 0.0$	$p_x^A = 0$	$p_y^A = 0$	$\sum p_z = 0$
parton B	-	$m = 0.0$	$p_x^B = 0$	$p_y^B = 0$	$\sum E_i = \sum E_f$
b-quark 1	t_{had}	\mathbf{d}_1	$\delta(\Delta\eta)$	$\delta(\Delta\phi)$	$m = 4.7$
b-quark 2	t_{lep}	\mathbf{d}_2	$\delta(\Delta\eta)$	$\delta(\Delta\phi)$	$m = 4.7$
b-quark 3	H/g	\mathbf{d}_3	$\delta(\Delta\eta)$	$\delta(\Delta\phi)$	$m = 4.7$
b-quark 4	H/g	$\delta(\mathbf{d}_4^2 - (p^{\text{bq}3} + p^{\text{bq}4})^2)$	$\delta(\Delta\eta)$	$\delta(\Delta\phi)$	$m = 4.7$
light quark 1	W_{had}	\mathbf{d}_5	$\delta(\Delta\eta)$	$\delta(\Delta\phi)$	$m = 0.0$
light quark 2	W_{had}	$\delta(\mathbf{d}_6^2 - (p^{\text{lq}1} + p^{\text{lq}2})^2)$	$\delta(\Delta\eta)$	$\delta(\Delta\phi)$	$m = 0.0$
charged lepton	W_{lep}	$\delta(\Delta E)$	$\delta(\Delta\eta)$	$\delta(\Delta\phi)$	$m = 0.0$
neutrino	W_{lep}	$m = 0.0$	$\sum p_x = 0$	$\sum p_y = 0$	$\delta(m_W^2 - (p^\ell + p^\nu)^2)$

Section 5.7.3, for the final integration, the energies of the charged leptons are assumed to be measured with negligible uncertainty. As shown in Figure 5.12, this assumption holds only for low energies or momenta. However, the bias introduced by this assumption affects both the signal and background hypothesis in the same way. When building the likelihood ratio, this bias is cancelled out such that the separation power is not reduced.

The dimensions of the neutrino transverse momenta p_x and p_y can be reduced by imposing momentum conservation of all initial and final state particles. In consistency with the LO picture, the transverse momentum of the neutrino is given by the negative sum of all other parton level objects involved in the processes. The deviation to the measured p_x and p_y component of the E_T^{miss} results in a weight from the neutrino momentum TFs. Assuming negligible neutrino mass and applying a NWA to the leptonically decaying W boson reduce the other two dimensions of the neutrino four-momentum. The following two alternative treatments of the neutrino transverse momenta have been studied. They lead to either less separation or to slower integration convergence and are therefore rejected. Additional objects influence the transverse momentum balance, so the neutrino transverse momenta could be treated as free integration dimensions constrained by the neutrino TFs. Although the integration results in a slightly better separation, this approach is rejected because of much slower integration convergence with the increase of integration dimensions. Similar to the charged lepton energy approach, it could be assumed that the measured E_T^{miss} directly reflects the neutrino momenta without any uncertainty and the introduced bias cancels out in the likelihood ratio. Using δ -functions for the neutrino momenta does not degrade the separation of the likelihood ratio but of other distributions and does not converge significantly faster than the momentum balance option.

With all the described approximations and simplifications, the dimensions of the integral can be reduced from initially 40 to six. After considering the described kinematic transformations, the remaining integration parameters are three b -quark energies $\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3$, one light-quark energy \mathbf{d}_5 , the hadronically decaying W boson mass \mathbf{d}_6 and the invariant mass of the two

5. The Matrix Element Method

b -quarks \mathbf{d}_4 originating either from the Higgs boson or a gluon. Using the Jacobian factors due to the kinematic transformation and the NWA as discussed in Section 5.7.3, the 8-body phase space element (without the δ -functions) reduces to

$$d\Phi_8 = \frac{dE^{bq1}}{(2\pi)^3} \cdot \frac{dE^{bq2}}{(2\pi)^3} \cdot \frac{dE^{bq3}}{(2\pi)^3} \cdot \frac{dE^{lq1}}{(2\pi)^3} \cdot \frac{2E^{bq4} dm_{H/g}}{(2\pi)^3 m_{H/g}} \cdot \frac{2E^{lq1} dm_W^{\text{had}}}{(2\pi)^3 m_W^{\text{had}}} \cdot \frac{E_\nu \Gamma_W M_W}{2\pi(2\pi)^3 (E_\ell p_\nu^z - E_\nu p_\ell^z)} . \quad (5.51)$$

The total integration volume is significantly restricted by setting the limits of the integration dimensions to optimum values based upon the observed values and the width of the transfer functions or the propagator peaks in the matrix elements. The integration ranges for the various integration parameters are shown in Table 5.4.

Table 5.4.: The upper and lower limits of the integration dimensions \mathbf{d}_1 - \mathbf{d}_6 based on the TF widths and masses of the intermediate states. The upper and lower energies of the decay products $E_{u/l}^q = E^j \pm 3 \cdot \sigma(\text{TF})$ are used to calculate the limits on the invariant masses corresponding to the intermediate states.

Parameter	Dimension	Integration Ranges	
		lower limit	upper limit
d_1	E^{bq1}	$E^{bj1} - 3 \cdot \sigma(\text{b-quark TF})$	$E^{bj1} + 3 \cdot \sigma(\text{b-quark TF})$
d_2	E^{bq2}	$E^{bj2} - 3 \cdot \sigma(\text{b-quark TF})$	$E^{bj2} + 3 \cdot \sigma(\text{b-quark TF})$
d_3	E^{bq3}	$E^{bj3} - 3 \cdot \sigma(\text{b-quark TF})$	$E^{bj3} + 3 \cdot \sigma(\text{b-quark TF})$
d_4	$m_{H/g}$	$\text{Min}\left(m_H - 50 \text{ GeV}, \sqrt{2E_l^{bj3} E_l^{bj4} - \cos(\theta^{bj3,bj4})}\right)$	$\text{Max}\left(m_H + 100 \text{ GeV}, \sqrt{2E_u^{bj3} E_u^{bj4} - \cos(\theta^{bj3,bj4})}\right)$
d_5	E^{lq1}	$E^{lj1} - 3 \cdot \sigma(\text{l-quark TF})$	$E^{bj1} + 3 \cdot \sigma(\text{l-quark TF})$
d_6	m_W	$\text{Min}\left(m_W - 50 \text{ GeV}, \sqrt{2E_l^{lj1} E_l^{lj2} - \cos(\theta^{lj1,lj2})}\right)$	$\text{Max}\left(m_W + 100 \text{ GeV}, \sqrt{2E_u^{lj1} E_u^{lj2} - \cos(\theta^{lj1,lj2})}\right)$

For the b -quark and light quark energies (E^{bq} , E^{lq}), the ranges are set to $\pm 3 \sigma$ around the observed jet energies (E^{bj} , E^{lj}), where σ is the Gaussian width of the dominant Gaussian term of the corresponding double-Gaussian TF. In case of the W boson or the Higgs boson (gluon) masses, the minimal ranges are set to $[m_{W/H} - 50 \text{ GeV}, m_{W/H} + 100 \text{ GeV}]$ to account for the asymmetries in the energy resolutions. However, lower and upper integration limits are also calculated from the energy TF, by varying the energies of the measured decay products by $\pm 3 \sigma$ of the main Gaussian and then calculating the corresponding invariant di-jet masses while neglecting the quark masses. If the latter limits extend the range, they are taken instead of the minimal ranges. This procedure ensures in particular that the integration range for the invariant mass of the two b -quarks originating from the gluon is large enough.

All the optimisations described in this chapter could reduce the computing time needed to calculate the signal and background process probability for a single event, while considering all possible assignment permutations, from initially 24 hours to less than 2 minutes.

5.8. Computational Challenges

Although the integration time could be significantly reduced by several orders of magnitude, the resulting 2 min per event for the process probability calculation still requires to highly parallelize the computation. The worldwide LHC computing grid (WLCG) provides large computing resources around the world with more than 170 computing centres (*grid sites*) in around 40 countries. These resources can be utilised by the collaborations of the LHC exper-

iments. However, in order to allow for fair share, restrictions per user are made constraining the usage with a personal priority quota. As a consequence, the workload needs to be shared among several users in order to run a large scale MEM analysis. To realise the presented analysis, a large scale computing framework has been developed of which the main components are briefly introduced in the following.

5.8.1. Grid Evaluation System

The MC simulation files need to be highly segmented in order to calculate the process probabilities of large MC statistics in a reasonable time scale. To simplify the merging procedure afterwards and to avoid technical difficulties it was decided to run each MC file on only one grid site. Because not all grid sites perform equally well at all times, the current performance of the grid sites are evaluated before each run. With the goal to find recent, current and possibly future reliable grid sites, all available grid sites with a decent number of analysis nodes and usable storage elements are analysed. By making use of the ATLAS Grid Information System (AGIS) and the ATLAS grid monitoring system, the performance of the sites over the last week and 24 h is assessed. The evaluation system checks for scheduled downtimes, the available number of nodes, the current load, the relative queue length and failure rate. Depending on criteria evaluating these quantities, it provides a list of good grid sites to which the input data is automatically distributed. During the analysis run, typically around 20 grid sites were identified as good.

5.8.2. MEDUSA - Grid Submission Framework

In order to run the presented MEM analysis in a relatively short time, the user friendly submission framework Management of Exigent, Distributed User Submission at ATLAS (MEDUSA) was developed. It aims to simplify and automatise all user related processes as much as possible. MEDUSA is a general software framework simplifying the share of any kind of large scale grid-based analysis. The main components and work flow of MEDUSA is presented in Figure 5.14.

Without going into too much detail, the general structure and idea of MEDUSA is outlined in the following. A template file defines all details of a certain project, such as information about the input files on which the analysis should run, the grid sites to which the input has been distributed and the workload per job. From this template, a project is organised according to the defined specifications and making use of the information provided by the grid evaluation system. Job tables are created specifying how many events of which file are required to run on which site. Finally, an online assignment daemon is initialised waiting for users to request jobs. A user willing to help to run the analysis can comfortably request a certain number of jobs for a defined project via a web interface. The assignment daemon assigns the requested jobs to the user based on the availability, creates a grid submission script and sends the user an email with submission instructions. The user can simply follow the one-line instruction sent by email, ideally submitting a batch job, which remotely submits the actual analysis jobs to the grid. The admin of the project receives reports of the submission containing all relevant information and status reports about the completion of the assignment. Requesting and submitting the jobs takes less than one minute for an experienced user.

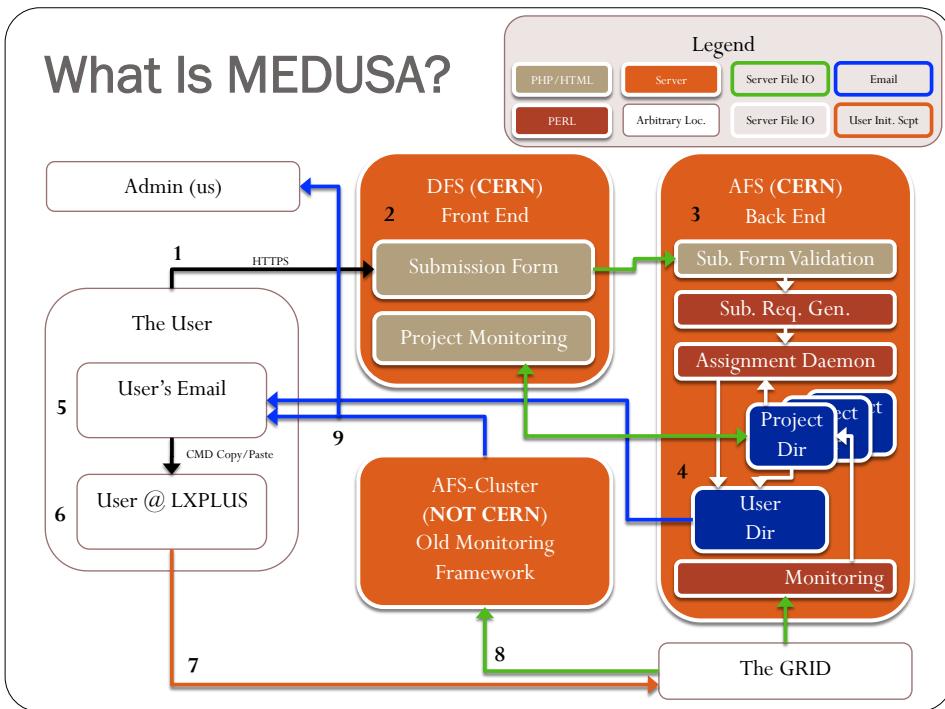


Figure 5.14.: Work flow diagram of the main components of MEDUSA.

5.8.3. Automated Grid Monitoring

After a project has been completely submitted to the grid, many thousand jobs of 20-30 user need to be monitored, merged and downloaded. To do that most efficiently, a fully automated monitoring system has been developed, whose work flow is schematically illustrated in Figure 5.15

The monitoring is started with a master script, which initialises several programs executed in certain time periods. The recovery tool restarts the assignment daemon once a day and recovers possible data loss in the submission log containing all relevant information of all user submissions. Using this information, a monitor job checks every 6 h all analysis jobs running currently on the grid by multiple users. It writes a summary report for each input MC sample summarising the current status of all related grid jobs by counting how many jobs are waiting, running, done or failed. It recognises automatically if the run of one MC file is fully completed, initialises then a merging job on the grid and writes the information to a log file. In case jobs are failed, it generates one retry script per user and notifies the user per email once a day, that actions are needed to retry the failed jobs. In order to retry the jobs, the user executes again a one-line command, which will at the same time resubmit jobs in case something with the initial submission went wrong. Another program monitors the merging jobs and writes similarly as described previously a summary of the current status of the jobs. If the merging of one MC sample is fully completed, it automatically initialises a download job running on a batch system, which retrieves the data set from the grid. Because of file sizes restrictions on the

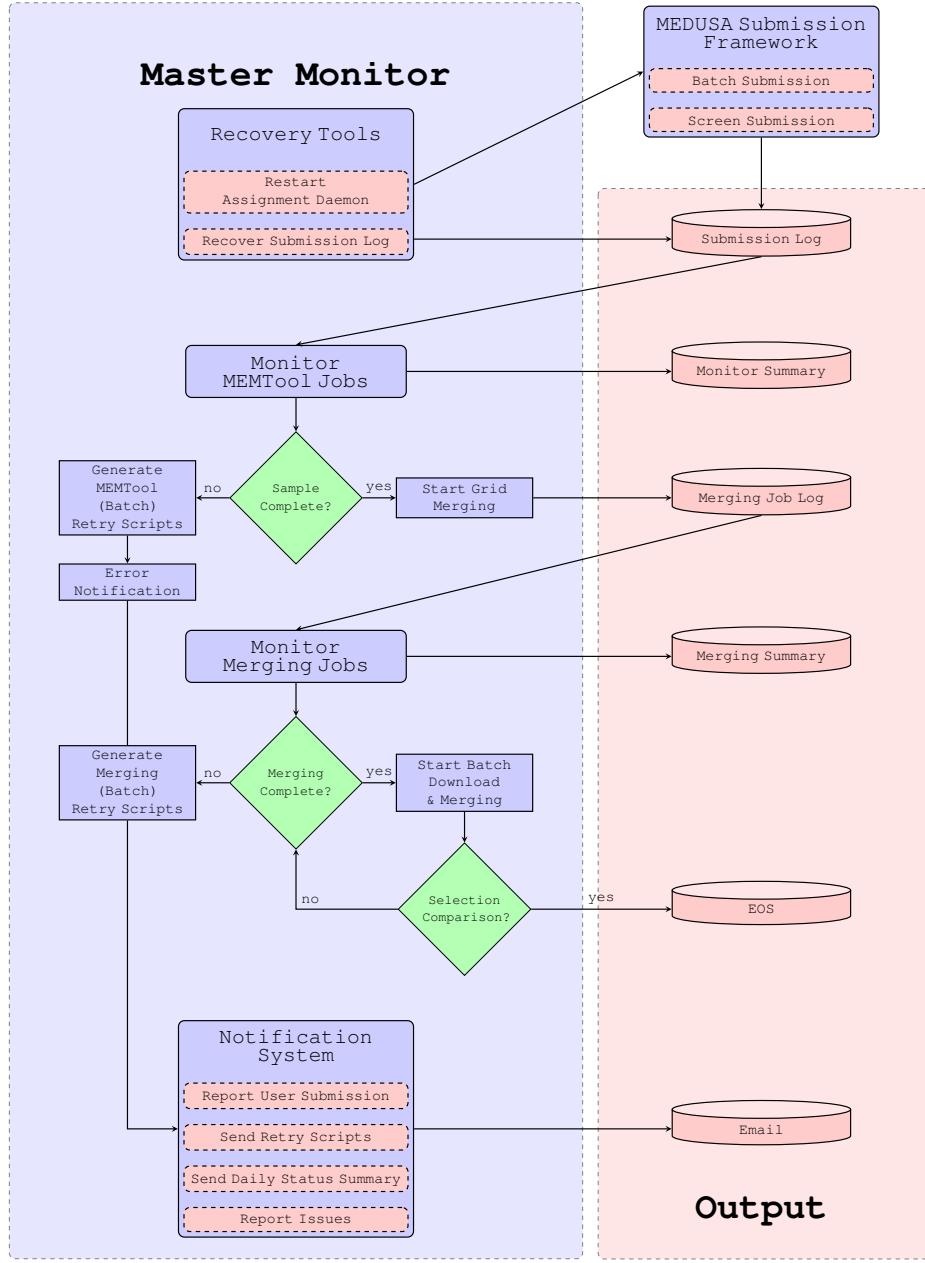


Figure 5.15.: Work flow diagram of the main components of the automated monitoring.

grid, the output files need to be further merged, if the MC file size is large. The job then runs an event selection on the output file ensuring that the number of selected events agrees with the one obtained with the input file. If that is the case, the file is uploaded and permanently stored on an EOS storage element. However, if the grid merging is not complete and jobs are failed, it generates a retry script and notifies the admin to retry some of the merging jobs once a day. All the notifications are executed by the notification system, which sends also daily summaries of the grid run status and merging status to the admin. If at any point of the chain something goes wrong, it sends reports to the admin with detailed information about the issue.

Without the MEDUSA and monitoring frameworks simplifying and automating multiple user submissions to the grid, a computationally intensive MEM is impossible. Many projects were run using this framework and at peak times around 11 k computing nodes located on 40 different grid sites around the world were used simultaneously. The user action could be minimised and simplified requiring a time of less than 5 min a day from them.

5.9. Proof of Principle

In order to validate the implementation of the MEM, several studies based on parton level events are performed. Two samples containing each 1000 $t\bar{t}H$ and $t\bar{t} + b\bar{b}$ events with LO precision are generated with MADGRAPH 5 [126]. These events are referred to as MADEVENTS and are produced at a centre-of-mass energy of $\sqrt{s} = 8$ TeV using the CTEQ6L1 PDF set. In both samples, the top quark pair decays into the single lepton (electron) final state and the top quark mass is set to 174.3 GeV. The Higgs boson decays exclusively into a pair of b -quarks and its mass is set to $m_H = 125$ GeV. Very loose acceptance cuts requiring for instance a jet p_T of 15 GeV are applied to the events. The process probabilities for the described $t\bar{t}H$ and the $t\bar{t} + b\bar{b}$ hypotheses are calculated on both samples. Under the $t\bar{t}H$ hypothesis the Higgs boson mass is varied as a model parameter between 115–135 GeV in steps of 1 or 2 GeV. The best estimator of the Higgs boson mass is extracted by maximizing the global likelihood function as described in Equation 5.25. Using the $t\bar{t}H$ process as the signal hypothesis and the $t\bar{t} + b\bar{b}$ process as the background hypothesis, a slightly modified version of Equation 5.26 is used to define the likelihood ratio

$$D1 = \frac{\mathcal{L}_{t\bar{t}H}^{\text{sum}}}{\mathcal{L}_{t\bar{t}H}^{\text{sum}} + \alpha \cdot \mathcal{L}_{t\bar{t}+b\bar{b}}^{\text{sum}}} , \quad (5.52)$$

where α is a relative normalisation factor automatically chosen to optimise the performance of the discriminant given the finite bin sizes of the $D1$ distribution. In this definition, signal-like and background-like events have $D1$ values close to one and zero, respectively.

5.9.1. Parton Level Events

First studies are performed using the four-momentum vectors of the parton level objects. Since no detector resolution is implied, the TFs of the six quarks, the electron and the neutrino transverse momentum are modelled as δ -distributions. The same reductions of dimensionality based on energy-momentum conservation, Einstein relation and negligible initial state transverse momentum as described in Section 5.7.4 are applied, but no kinematic transformation is used. With all these, the only integration variable left is the neutrino p_z momentum component. The true flavour of the quarks is used as a perfect b -tagging allowing to reduce the number of assignment permutations to twelve. A typical negative $\ln \mathcal{L}$ distribution using the signal hypothesis and varying the Higgs boson mass in 1 GeV steps is shown in Figure 5.16 for a single $t\bar{t}H$ event (a) and a single $t\bar{t} + b\bar{b}$ event (b).

In the upper distributions the signal $-\ln \mathcal{L}$ of each assignment permutation are shown separately for the two type of events. In case of the $t\bar{t}H$ event (a), the permutation 0 (black circle) and 1 (red circle) have significantly lower $-\ln \mathcal{L}$ values over the Higgs boson mass range and show a distinct minimum at a Higgs boson mass of 125 GeV. Permutation 0 is the correct assignment of the parton level objects to the quarks of the matrix element and thus results in

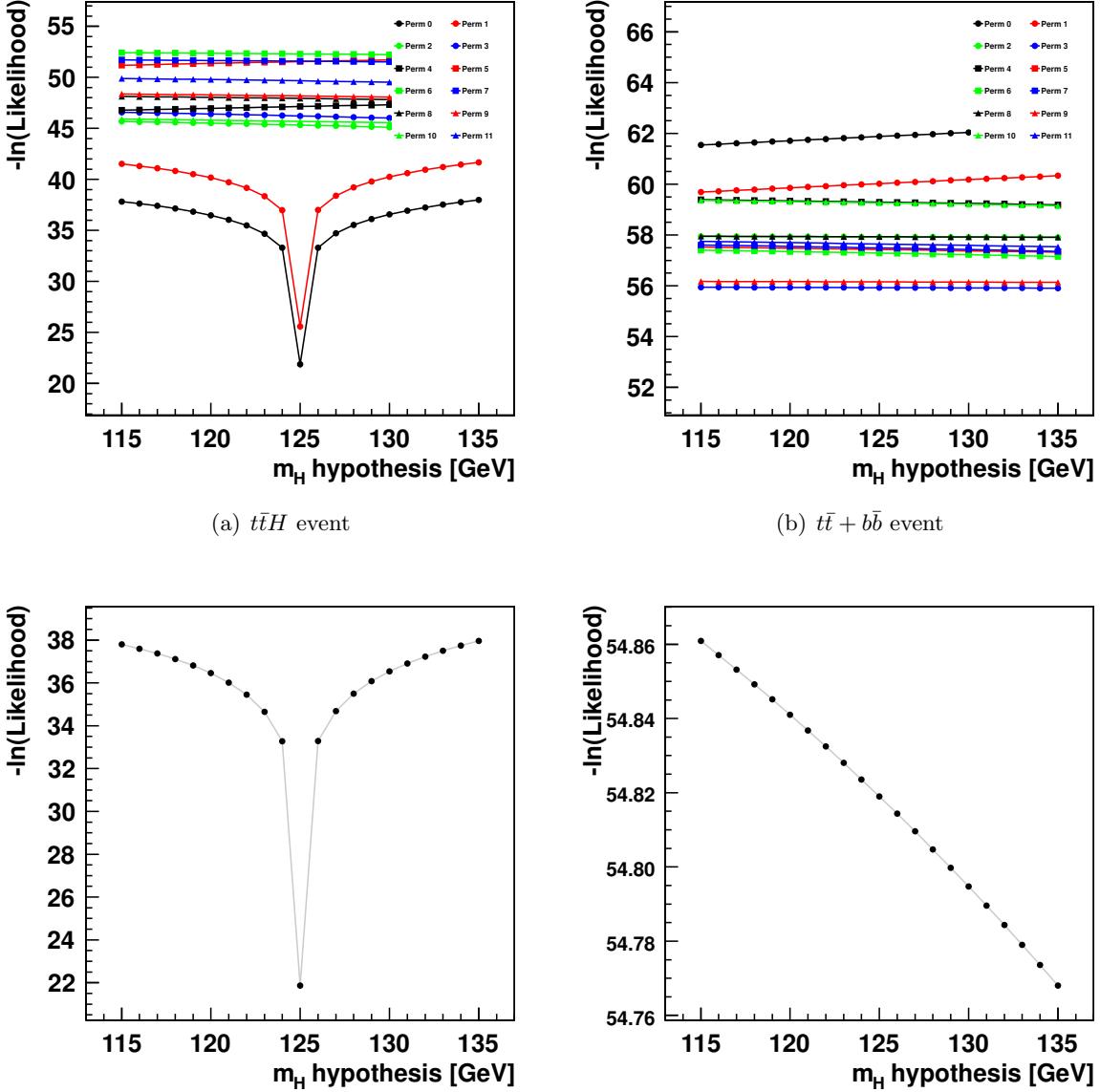


Figure 5.16.: The negative logarithm of the signal likelihood per permutation (upper) and sum of all permutations (lower) for a single $t\bar{t}H$ (a) and $t\bar{t} + b\bar{b}$ (b) event at parton level for different Higgs boson mass hypotheses.

the lowest $-\ln \mathcal{L}$ values. In the assignment of permutation 1, the two b -quarks of the decaying top quarks are interchanged, but the two b -quarks of the Higgs boson are assigned correctly. Thus the resulting $-\ln \mathcal{L}$ distribution results in slightly higher values than for permutation 0. The contribution to the likelihood from the top quark divergence is smaller, but the likelihood distribution still reaches a minimum at the correct Higgs boson mass, since the correct b -quarks are used to reconstruct the Higgs boson divergence. In all other assignments, one of the b -quarks originating from the Higgs boson is not correctly assigned and thus the response is almost independent of the Higgs boson mass parameter. In case of testing a $t\bar{t} + b\bar{b}$ event

5. The Matrix Element Method

with the $t\bar{t}H$ hypothesis (b), there is no visible dependence on the Higgs boson mass parameter and all permutations show a similar flat behaviour. Since there is no large contribution from the Higgs boson propagator, the overall $-\ln \mathcal{L}$ value is also significantly higher than for the $t\bar{t}H$ event. The correct assignment of permutation 0 results in the most unlikely value: even if the top quark pair is correctly reconstructed, the invariant mass of the two b -quarks originating from the gluon do not match the Higgs boson mass. Thus using b -quarks from the top quark decay for the Higgs boson divergence results in higher likelihood values, but none of the b -quark combinations have an invariant mass in the range of 115-135 GeV, which would result in a local minimum. In certain events this might happen by accident resulting in single or multiple peak structures, however the total $-\ln \mathcal{L}$ value would be by far higher than the one for a $t\bar{t}H$ event.

In the lower distributions, the $t\bar{t}H$ process probabilities are combined to a single event likelihood by summing over all the different assignment permutations. For the $t\bar{t}H$ event the likelihood contributions of the two permutations with the correct assignment of the b -quarks originating from the Higgs boson are really dominant. The remaining permutations have a negligible contribution and result in an overall shift in the $-\ln \mathcal{L}$ value. As a consequence, the combined $-\ln \mathcal{L}$ distribution still shows a pronounced and sharp minimum at a Higgs boson mass of 125 GeV. This likelihood distribution could be used to extract the Higgs boson mass of the event. The combined $-\ln \mathcal{L}$ distribution for the $t\bar{t} + b\bar{b}$ event is dominated by the likelihood of some wrong permutations which results in an almost flat distribution. The average of the $-\ln \mathcal{L}$ values is significantly higher than the ones of the $t\bar{t}H$ event, thus combining one signal and one background event would result in a very pronounced minimum at the Higgs boson mass of the signal event. Even at parton level, the situation is not always that clear and in a small fraction of the events the $-\ln \mathcal{L}$ distribution has multiple minima as shown for two example events in Figure 5.17.

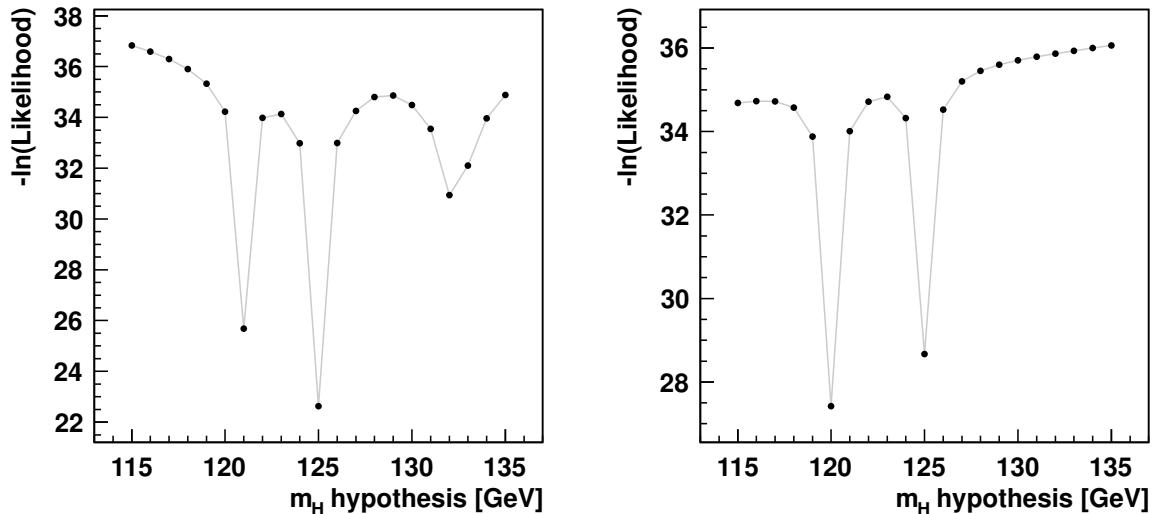


Figure 5.17.: The negative logarithm of the $t\bar{t}H$ likelihood with varied Higgs boson mass parameter for two different signal events with no clear kinematic properties resulting in ambiguous minimal Higgs boson mass hypotheses.

In the left plot, there are three local minima for Higgs boson masses of 121, 125 and 132 GeV. The additional minima are caused by assignment permutations in which the invariant mass of the two b -quarks assigned to the Higgs boson are coincidentally close to the tested Higgs boson pole masses even though at least one of the b -quarks originates from a top quark. In the right plot there is a second minimum at 120 GeV, which is even stronger pronounced than the one of the correct assignment of all quarks. However, the $-\ln \mathcal{L}$ values of both minima are higher than the values obtained in the other events. Thus either the integration of the neutrino p_z did not cover all the relevant phase space or the invariant mass of the two b -quarks originating from the Higgs boson is further away from the Higgs boson pole mass of 125 GeV. All the likelihoods from the single events can now be combined to a global likelihood function (see Eq. 5.25). The global $-\ln \mathcal{L}$ distributions of the $t\bar{t}H$ hypothesis combined for the 1000 signal and background events are shown in Figure 5.18.

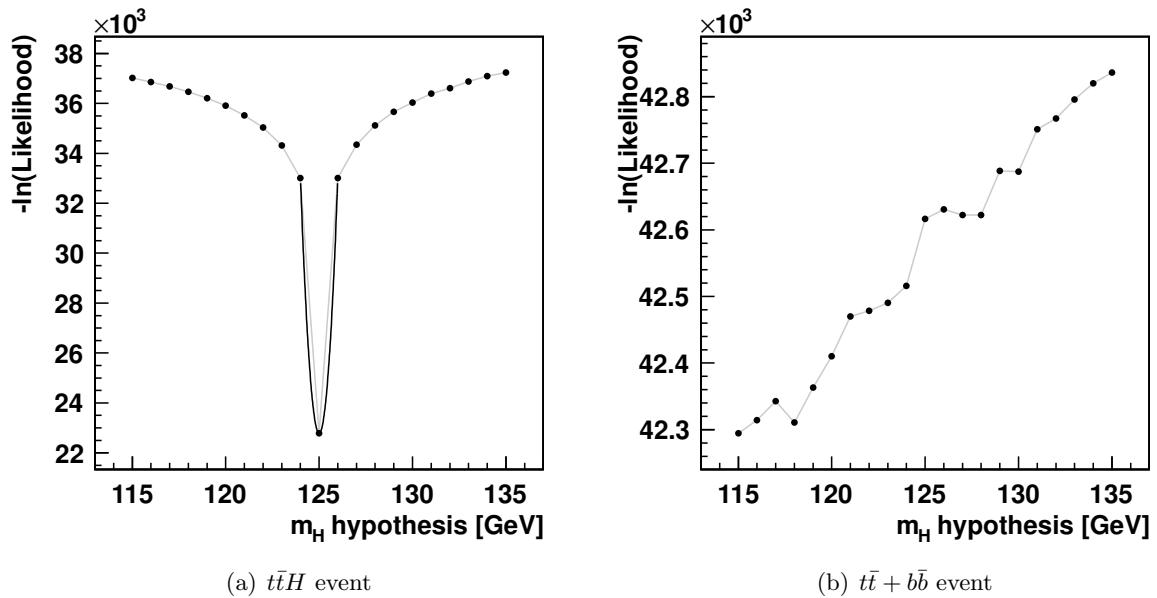


Figure 5.18.: The signal likelihood is shown combined for 1000 $t\bar{t}H$ (a) and $t\bar{t} + b\bar{b}$ events at parton level for different Higgs boson mass hypotheses.

In case of the $t\bar{t}H$ events (a), the distribution looks very similar to the one of an ideal single event multiplied by a factor of 1000. Multiple peak structures arising from accidental matching invariant masses of two b -quarks as seen in Figure 5.17 are smoothed out. The global minimum at 125 GeV is extremely sharp, such that neighbouring mass points have 10^4 higher $-\ln \mathcal{L}$ values. Assuming a Gaussian distribution, the $-\ln \mathcal{L}$ distribution is fitted with a parabola in order to extract the Higgs boson mass. The minimum of the parabola is exactly located at the generated Higgs boson mass of 125 GeV. The response of the 1000 $t\bar{t} + b\bar{b}$ events (b) to the signal hypothesis is as expected independent of the Higgs boson mass parameter. Thus the combined $-\ln \mathcal{L}$ distribution is rather flat with some fluctuations and a small rising slope. The difference between the smallest and the highest $-\ln \mathcal{L}$ value in this parameter range is approximately 500, which corresponds to less than 5% of the differences for $t\bar{t}H$ events. Furthermore, the average $-\ln \mathcal{L}$ is significantly higher than the $-\ln \mathcal{L}$ values of the $t\bar{t}H$ events. Thus, combining even a small fraction of $t\bar{t}H$ events with a large sample of $t\bar{t} + b\bar{b}$ events will result in a very nice $-\ln \mathcal{L}$.

5. The Matrix Element Method

distribution with a very pronounced minimum. The likelihood of the $t\bar{t} + b\bar{b}$ hypothesis is also calculated and the $D1$ likelihood ratio of Equation 5.52 is shown in Figure 5.19 for three Higgs boson mass values for the entire $t\bar{t}H$ (red) and $t\bar{t} + b\bar{b}$ (blue) events.

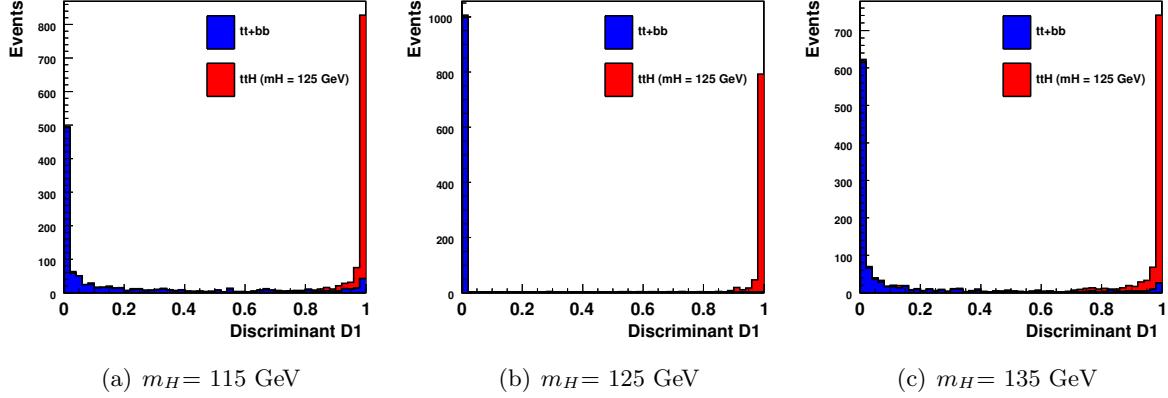


Figure 5.19.: The discriminant $D1$ for a Higgs boson mass parameter of $m_H = 115, 125$ and 135 GeV for $1000 t\bar{t}H$ (red) and $t\bar{t} + b\bar{b}$ (blue) events at parton level.

The separation between $t\bar{t}H$ and $t\bar{t} + b\bar{b}$ events is perfect (no overlap) for the correct Higgs boson mass parameter of 125 GeV . While all $t\bar{t} + b\bar{b}$ events have a background-like value very close to zero, most of the $t\bar{t}H$ events have a signal-like value close to one and only a few events spread to values below one. However, the discrimination is slightly degraded when $D1$ is calculated for a Higgs boson mass of 115 GeV and 135 GeV , since the signal likelihood response from the $t\bar{t}H$ sample is not at its minimum. Although the distributions of the $t\bar{t} + b\bar{b}$ and $t\bar{t}H$ sample overlap and some of the $t\bar{t} + b\bar{b}$ events even have very signal-like values of close to one, there is great separation.

5.9.2. Single Gaussian Smeared Parton Level Events

For the next validation step, more realistically studies are performed by smearing the energies of the final state quarks and electron and the transverse momentum of the neutrino using simple Gaussian functions that take the detector resolution into account. The width of the Gaussian distributions are chosen to approximately match the resolutions of the ATLAS detector and are summarised in Table 5.5.

Table 5.5.: Width of the Gaussian functions used to smear the energies of the light quarks, b -quarks and electrons, as well as the p_x and p_y momentum components of the neutrino.

Object	Gaussian width σ [GeV]
light quark energy	$0.5 \cdot \sqrt{E}$
b -quark energy	$0.5 \cdot \sqrt{E}$
electron energy	$0.1 \cdot \sqrt{E}$
neutrino momentum p_x p_y	20

Accordingly, all quark and electron energy TF and the neutrino momentum p_x and p_y TF are modelled with the same single Gaussian function. The directions of the particle are assumed to be measured with negligible uncertainty and are thus represented by δ -distributions. The integration dimensions are reduced in the same way as for the parton level events. In addition, kinematic transformations to the Higgs boson and leptonically decaying W boson mass are performed in order to apply a NWA. As a consequence, the dimensions are reduced to six: the three b -quark energies, two light quark energies and the electron energy. Because of the increased dimensionality and thus integration time, the process probabilities are calculated for only 500 events and the Higgs boson mass hypothesis is varied in steps of 2 GeV. As before, the number of assignment permutations is reduced by exploiting the true quark flavour. The negative logarithm of the global signal likelihood function combining all $t\bar{t}H$ (a) and $t\bar{t} + b\bar{b}$ (b) events are shown in Figure 5.20.

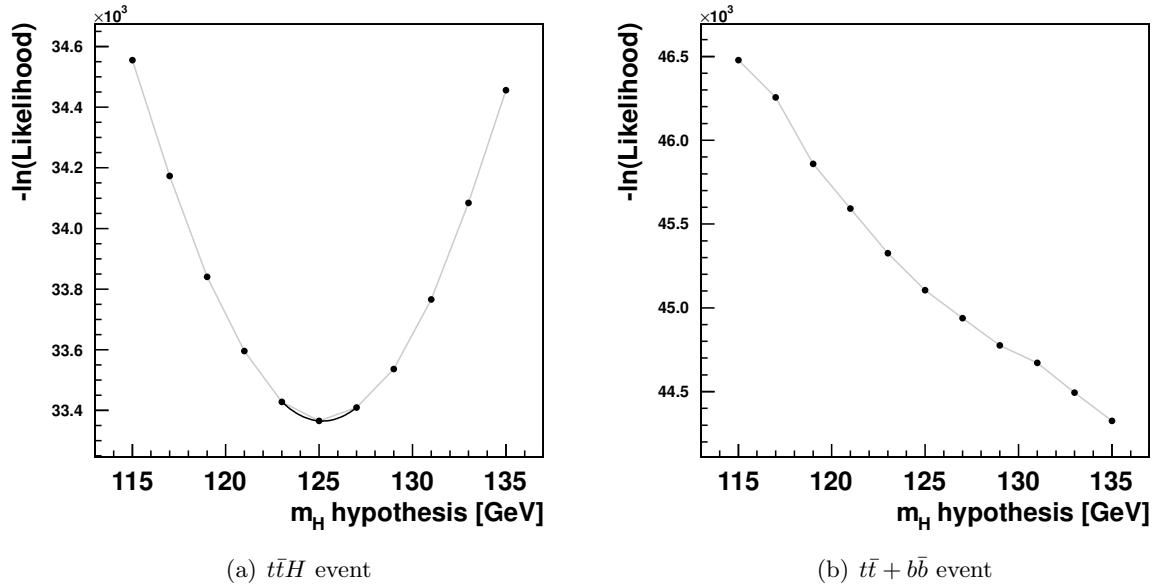


Figure 5.20.: The signal likelihood is shown combined for 500 $t\bar{t}H$ (a) and $t\bar{t} + b\bar{b}$ Gaussian smeared parton level events for different Higgs boson mass hypotheses.

Although the width of the $-\ln \mathcal{L}$ distribution of the $t\bar{t}H$ events is significantly increased due to the detector resolutions and increased number of integration dimensions, the distribution is still very symmetric and has a minimum at the true Higgs boson mass of the sample. The most likely Higgs boson mass of the sample can be again extracted by fitting the minimum with a parabola and a value very close to 125 GeV is obtained. Unlike for the parton level events, the minimum does not span over many orders of magnitude in likelihood value, thus its error is much larger. The response of the background sample $t\bar{t} + b\bar{b}$ is not as flat as before, but drops with a constant slope. However, the likelihood value is still roughly ten orders of magnitude higher. Thus when combining background and signal events, the Higgs boson mass can still be extracted with negligible bias. In Figure 5.21, the discriminant $D1$ is shown again for three different $t\bar{t}H$ likelihood hypotheses with Higgs boson masses of 115, 125 and 135 GeV. For the input mass of 125 GeV the separation between $t\bar{t}H$ (red) and $t\bar{t} + b\bar{b}$ (blue) is again almost perfect, although a large fraction of the $t\bar{t}H$ events have now significantly smaller val-

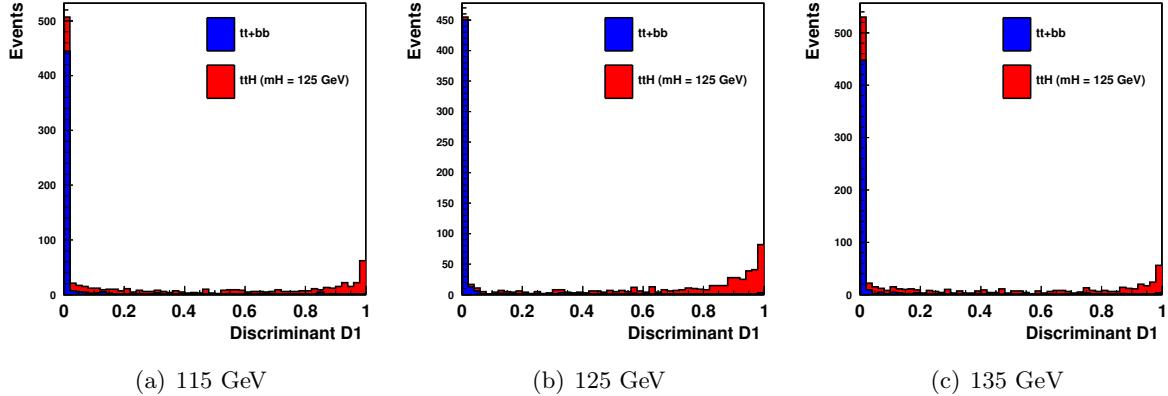


Figure 5.21.: The discriminant $D1$ for a Higgs boson mass parameter of 115, 125 and 135 GeV for 500 $t\bar{t}H$ (red) and $t\bar{t} + b\bar{b}$ (blue) Gaussian smeared parton level events.

ues than one and some events have even values around zero. Most of the $t\bar{t} + b\bar{b}$ events have background-like values close to zero allowing the $D1$ variable to be still a very powerful discriminant. For Higgs boson mass hypotheses 10 GeV away from the true value, the discrimination is clearly degraded. Despite the overlap, the separation is still good enough to distinguish the two processes.

5.9.3. Double Gaussian Smeared Parton Level Events

For the last validation step, the measurable quantities of the final state particles are smeared using the TF obtained from a simulation of the ATLAS detector as described in Section 5.6. This means that the energies of the quarks and the electron are smeared with an asymmetric double Gaussian function and the central value is shifted in order to account for possible energy losses. The neutrino transverse momentum is modelled with a single Gaussian function dependent on the total summed E_T . Although showering and hadronisation as well as multiple parton interactions and radiations are not considered, this is the most realistic picture matching the LO model of the process probabilities. Naturally, the same TF are used for the measured quantities. Since the same approximations and phase space transformation as for the single Gaussian smeared events are applied, the same integration variables define the six dimensional space. Again only 500 events are integrated, the Higgs boson mass parameter is varied in 2 GeV steps and truth flavour tagging is used to reduce the assignment permutations. The $-\ln \mathcal{L}$ distributions of the signal hypothesis is shown in Figure 5.22 for the $t\bar{t}H$ and $t\bar{t} + b\bar{b}$ events.

The distribution of the signal events is now asymmetric and broader than before with a minimum slightly shifted from the true value to a higher value of 127 GeV. The minimum of the parabolic fit is biased due to the asymmetric shape of the distribution. Measuring the Higgs boson mass with this method is still feasible since the observed bias can be calibrated using several simulated samples with different Higgs boson masses. The minimum is also less pronounced compared to the previous tests and the likelihood values of the $t\bar{t} + b\bar{b}$ background events are closer to those for $t\bar{t}H$, but still around three to four orders of magnitude higher. Thus, if the signal fraction is large enough, the Higgs boson mass can still be extracted in a combined sample, though the uncertainty will be significantly larger and an additional bias

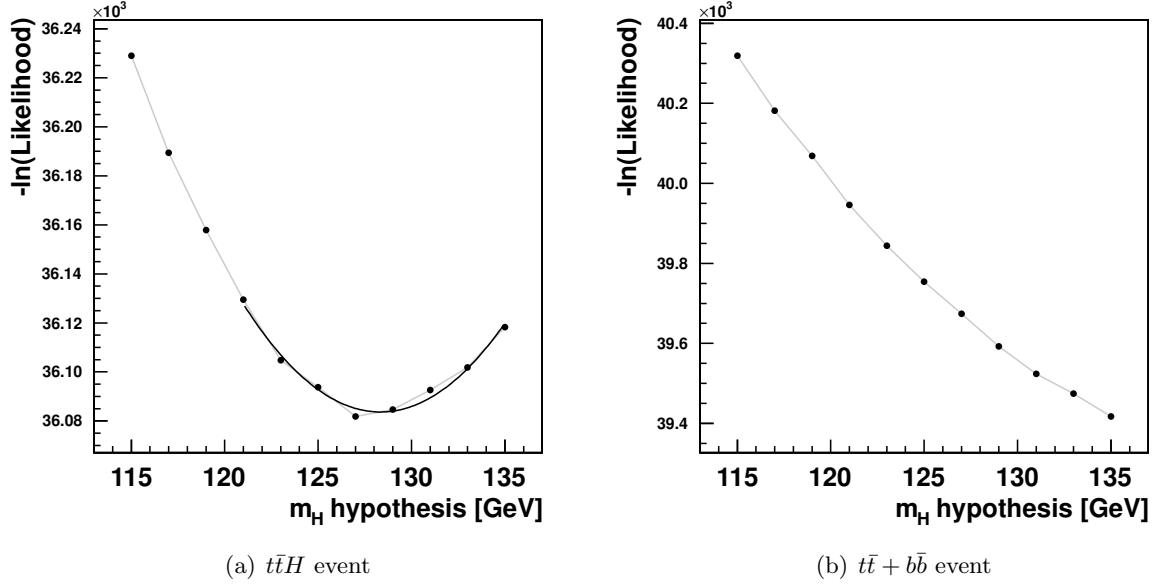


Figure 5.22.: The signal likelihood combined for 500 $t\bar{t}H$ (a) and $t\bar{t} + b\bar{b}$ double Gaussian smeared parton level events for different Higgs boson mass hypotheses.

might be introduced. The $t\bar{t} + b\bar{b}$ events respond similar as before to the Higgs boson mass variation with a constantly falling slope to higher mass values. As expected from the signal likelihood distributions, the discriminant $D1$, again shown for the different Higgs boson mass in Figure 5.23, is significantly less dependent on the Higgs boson mass hypothesis.

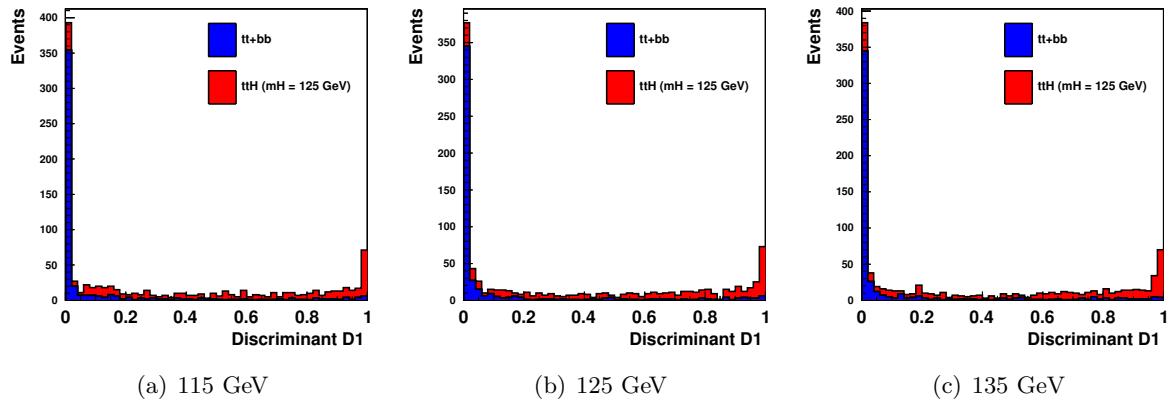


Figure 5.23.: The discriminant $D1$ for a Higgs boson mass parameter of 115, 125 and 135 GeV for 500 $t\bar{t}H$ (red) and $t\bar{t} + b\bar{b}$ (blue) double Gaussian smeared parton level events.

The signal events overlap now more with the background events, even for the true mass value of $m_H = 125$ GeV. The distribution of the $t\bar{t}H$ events is flatter over the range of $D1$ values with peaks at zero and one. The $t\bar{t} + b\bar{b}$ events populate mainly the region zero, with a spread to higher values. Overall, the separation power of the likelihood ratio is still very good allowing to distinguish $t\bar{t}H$ and $t\bar{t} + b\bar{b}$ events.

6

Search for $t\bar{t}(H \rightarrow b\bar{b})$ Production in the Single Lepton Channel

6.1. Introduction

After the discovery of a Higgs boson by the two collaborations ATLAS and CMS [61, 62], there is by now clear evidence of this particle in the $H \rightarrow \gamma\gamma$, $H \rightarrow ZZ^{(*)} \rightarrow 4\ell$, $H \rightarrow WW^{(*)} \rightarrow \ell\nu\ell\nu$ decay channels [72, 82]. By measuring its properties as well as the couplings to other particles [73, 82, 83, 197], more confidence is gained in the SM Higgs boson hypothesis of the observation. However, it is not yet certain if the observed Higgs boson is part of the well established SM or of a more general theory in which the SM is embedded. Observing the Higgs boson in all possible production and decay processes predicted by the SM and measuring its properties will show the true nature of the observed particle. The production of the Higgs boson in association with a pair of top quarks ($t\bar{t}H$) [143, 198–200], is one of the four main production mechanisms of the Higgs boson predicted by the SM. Despite the fact that many searches for this production mechanism are being performed at the LHC [70, 71, 74, 180, 201–206], it has not been observed yet. In addition, the Higgs boson has been only discovered in bosonic decays, though recently first evidence for a fermionic decay $H \rightarrow \tau\tau$ has been reported [76]. Although for a low mass Higgs boson as found at the LHC, the dominant decay mode is $H \rightarrow bb$, this decay has not been observed yet, because it is impossible to distinguish it from the overwhelming multijet background in the main production channels. Observing the $t\bar{t}H$ production with a Higgs boson decaying into a pair of b -quarks would allow for verifying the coupling of the Higgs boson to fermions in the production as well as in the subsequent decay. Due to the clear signature of the $t\bar{t}$ decay, the multijet background can be suppressed, which is alternatively only possible in Higgs boson production in association with a vector boson (VH) [67–69]. However, a new challenge is to distinguish the $t\bar{t}H$ signal from the irreducible $t\bar{t} + b\bar{b}$ background. The main source of background to this search comes from $t\bar{t}$ pairs produced in association with additional jets. The dominant source is the $t\bar{t} + b\bar{b}$ production, whose decay results in the same final state signature with kinematic properties similar to the signal. A second contribution arises from $t\bar{t}$ production in association with light-quark (u, d, s) or gluon jets, referred to as $t\bar{t} +$ light background, and from $t\bar{t}$ production in association with c quarks, referred to as $t\bar{t} + c\bar{c}$. The size of the second contribution depends on the mistag rate of the b -tagging algorithm. Obtaining a good separation between $t\bar{t}H$ and $t\bar{t} + b\bar{b}$ events and

6. Search for $t\bar{t}(H \rightarrow b\bar{b})$ Production in the Single Lepton Channel

understanding the $t\bar{t}$ background in all regions of the phase space are the key aspects of this search.

Since in LO the cross section of the $t\bar{t}H$ process is proportional to the squared matrix element and thus to the squared top quark Yukawa coupling, the observation of such a process enables the most direct measurement of the magnitude of the top quark Yukawa coupling. Up to now only indirect statements can be made regarding this coupling by measuring loop-induced processes like $gg \rightarrow H$ or $H \rightarrow \gamma\gamma$. While in the SM top quark and W boson loops dominate such processes, particles which are not contained in the SM may contribute as well. Any deviation in a direct measurement of the top quark Yukawa coupling from these indirect measurements would be a very robust and model independent test for such new particles. Due to the large measured top quark mass, the top quark Yukawa coupling is close to unity and might give insight to the scale of new physics [207].

The following chapter describes a search for the SM Higgs boson in the $t\bar{t}H$ production mode designed to be primarily sensitive to the $H \rightarrow b\bar{b}$ decay, although other Higgs boson decay modes are also treated as signal. It focuses on the decay of the $t\bar{t}$ system with a single lepton in the final state. For this search the full 8 TeV dataset as described in Section 4.2 with a total integrated luminosity of 20.3 fb^{-1} is used. The data is modelled by the relevant processes obtained from MC simulations and data-driven estimates as summarised in Section 4.4, which are then normalised to the same integrated luminosity. Given the measurements of the Higgs boson mass, this analysis primarily tests a Higgs boson mass of $m_H = 125 \text{ GeV}$.

6.2. Event Selection and Classification

6.2.1. Event Selection

The event selection of this analysis aims to reduce the non- $t\bar{t}$ background by requiring a final state consistent with $t\bar{t} + \text{jets}$ production. It considers the single lepton (electron or muon) decay channel of the top quark pair and is optimised for the $H \rightarrow b\bar{b}$ decay of the Higgs boson. In order to maximise the signal acceptance, relatively loose cuts are applied and many phase space regions are examined. The events are required to be recorded, because at least one of two single lepton triggers has fired. In order to maximise the trigger efficiency, two single lepton triggers with different p_T thresholds and isolation criteria are combined. In case of electrons, the p_T thresholds of the triggers are 24 and 60 GeV, while in case of muons they are 24 and 36 GeV. The triggers with the lower p_T threshold require in addition a loose isolation criterion on the lepton candidate, whose inefficiency at large transverse momenta is recovered by the trigger with the higher p_T threshold. The triggered events are further required to have at least one vertex, which is reconstructed with at least five associated tracks and close to the beam collision region in the x - y plane. The vertex with tracks resulting in the largest sum of squared transverse momentum is taken as the primary vertex of the hard scattering process. The single lepton channel is defined by requiring exactly one identified electron or muon with $p_T > 25 \text{ GeV}$, which matches the lepton found by the trigger within a cone radius of $\Delta R < 0.15$. If an additional lepton with opposite charge and $p_T > 15 \text{ GeV}$ is found, the event is removed to avoid statistical overlap with the dilepton analysis. In case of a second electron, the identification criterion is slightly looser and a p_T dependent isolation criterion is applied. Events are further discarded if any jet with $p_T > 20 \text{ GeV}$ is identified as out-of-time activity from a previous pp collision or as calorimeter noise [208]. Finally, the event is required to contain at least four jets with $p_T > 25 \text{ GeV}$ of which at least two are b -tagged,

which significantly reduces the non- $t\bar{t}$ background. No requirements on E_T^{miss} or transverse W boson mass typically used to suppress multijet background are applied in this analysis. The contribution from multijet events with a misidentified lepton is small due to the requirements of large number of jets and in particular b -tags.

6.2.2. Classification

The events are then further categorised into different analysis regions dependent on the jet and b -tag multiplicity. A region with m jets, of which n are b -tagged is referred to as "($m j, n b$)" in the following. A total of nine independent regions are considered. Event yields in these regions for the signal and background model compared to data are summarised in Table 6.1.

The given uncertainties include all statistical and systematic uncertainties (see Sec. 6.4), which are added in quadrature. In order to visualize the comparison of the predictions with the data, the expected and observed number of events for each analysis region are shown on a logarithmic scale in Figure 6.1 (a).

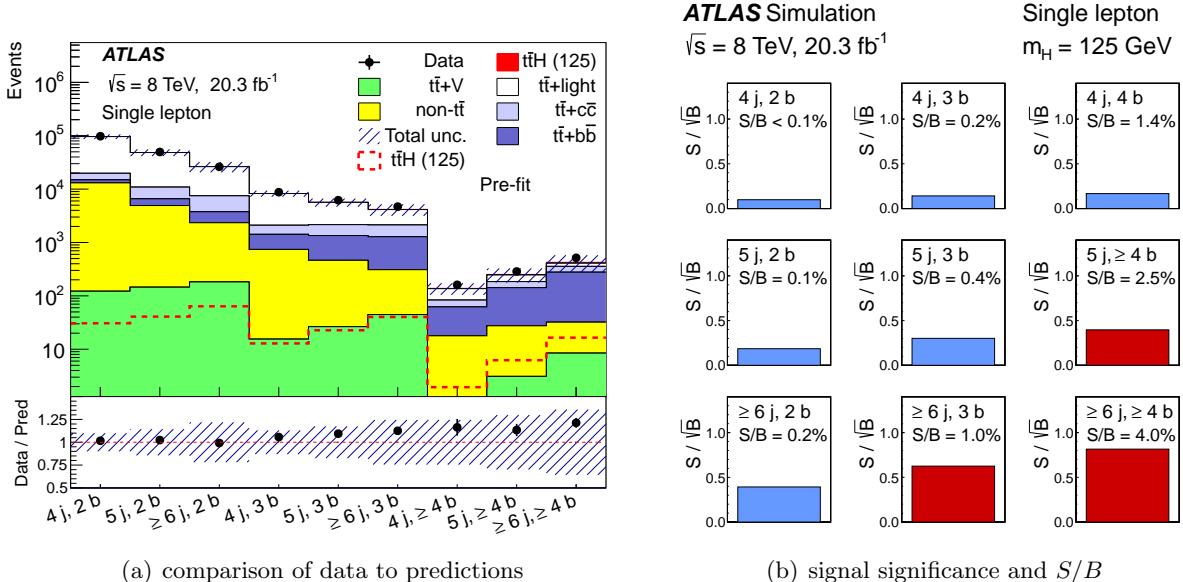


Figure 6.1.: (a) Comparison of prediction to data in all analysis regions. The signal, normalised to the SM prediction, is shown both as a filled red area stacked on the backgrounds and separately as a dashed red line. The hashed area corresponds to the total uncertainty on the yields [71].
 (b) The signal significance and signal-to-background ratio is shown for the nine analysis regions. The jet and b -tag multiplicities are ordered by row and column, respectively. The signal-enriched regions are highlighted in dark red, while background-dominated regions are shaded in light blue [71].

The total predicted event yield obtained from the model agrees in all categories well within the uncertainties (hashed area) with the observed data yield. However, in the regions with high jet and b -tag multiplicities the model tends to underestimate the event in data and the relative uncertainties become very large. Any $t\bar{t}H$ signal, as shown stacked on the background (filled red) and in addition separately (dashed red), would be completely hidden in the total uncertainties. Without improving the predictions and constraining the systematic uncertainties, no statement about the presence or absence of a signal can be made.

6. Search for $t\bar{t}(H \rightarrow b\bar{b})$ Production in the Single Lepton Channel

Table 6.1.: Pre-fit event yields for signal, backgrounds and data in each of the analysis regions. The quoted uncertainties are the sum in quadrature of the statistical and systematic uncertainties on the yields.

	4 j, 2 b	4 j, 3 b	4 j, 4 b	5 j, 2 b
$t\bar{t}H$ (125)	30.7 ± 2.8	12.9 ± 1.4	2.0 ± 0.3	40.9 ± 2.1
$t\bar{t}$ + light	$76\,700 \pm 7500$	6170 ± 750	53 ± 12	$37\,600 \pm 5500$
$t\bar{t} + c\bar{c}$	4870 ± 3000	682 ± 390	21 ± 12	4300 ± 2400
$t\bar{t} + b\bar{b}$	1840 ± 1100	680 ± 380	44 ± 25	1670 ± 880
W +jets	5120 ± 3000	225 ± 130	5.5 ± 3.3	1940 ± 1200
Z +jets	1130 ± 600	50 ± 27	0.9 ± 0.6	405 ± 240
Single top	4930 ± 640	337 ± 60	6.8 ± 1.6	1880 ± 360
Diboson	217 ± 71	11.5 ± 4.1	0.2 ± 0.1	97 ± 39
$t\bar{t}V$	122 ± 40	15.5 ± 5.1	0.9 ± 0.3	145 ± 48
Lepton misID	1560 ± 620	102 ± 37	3.5 ± 1.3	461 ± 170
Total	$96\,500 \pm 9500$	8280 ± 1100	138 ± 34	$48\,500 \pm 7000$
Data	98 049	8752	161	49 699
S/B	< 0.001	0.002	0.014	0.001
S/\sqrt{B}	0.099	0.141	0.167	0.186

	5 j, 3 b	5 j, ≥ 4 b	≥ 6 j, 2 b	≥ 6 j, 3 b	≥ 6 j, ≥ 4 b
$t\bar{t}H$ (125)	22.7 ± 1.8	6.2 ± 0.8	63.7 ± 5.0	40.2 ± 3.5	16.5 ± 2.0
$t\bar{t}$ + light	3480 ± 520	61 ± 15	$18\,800 \pm 4400$	2010 ± 460	52 ± 17
$t\bar{t} + c\bar{c}$	810 ± 460	43 ± 25	3730 ± 2100	846 ± 480	79 ± 46
$t\bar{t} + b\bar{b}$	886 ± 480	115 ± 63	1420 ± 770	974 ± 530	245 ± 130
W +jets	135 ± 87	5.9 ± 3.9	912 ± 620	97 ± 66	8.6 ± 6.2
Z +jets	29 ± 17	1.5 ± 0.9	183 ± 120	19 ± 12	1.5 ± 1.0
Single top	195 ± 41	8.3 ± 1.3	836 ± 220	122 ± 35	11.9 ± 3.7
Diboson	8.0 ± 3.4	0.4 ± 0.2	50 ± 24	6.0 ± 3.0	0.5 ± 0.3
$t\bar{t}V$	26.5 ± 8.6	3.1 ± 1.0	182 ± 59	44.6 ± 14	8.5 ± 2.8
Lepton misID	70 ± 28	8.3 ± 3.7	181 ± 66	21.3 ± 7.6	1.1 ± 0.5
Total	5670 ± 980	252 ± 75	$26\,400 \pm 5800$	4180 ± 1000	426 ± 150
Data	6199	286	26 185	4701	516
S/B	0.004	0.03	0.002	0.01	0.04
S/\sqrt{B}	0.301	0.40	0.393	0.63	0.82

From the number of predicted signal S and background B events, the signal-to-background ratio S/B and the signal significance S/\sqrt{B} are calculated for each region. To indicate which regions are dominated by background and which regions are sensitive to the signal, these two

numbers are visualised in Figure 6.1 (b). While the signal-to-background ratio gives an idea of the relative signal contribution in the corresponding category, the signal significance takes also the absolute numbers into account. The most sensitive regions with a signal significance of 0.82 and 0.63 and a signal-to-background ratio of 4% and 1% are the ($\geq 6j, \geq 4b$) and ($\geq 6j, 3b$) categories, respectively. It is clear, that with such low signal-to-background ratios no cut-and-count analysis can be performed, since the systematic uncertainties on the normalisations are much larger. The regions with a signal-to-background ratio $S/B > 1\%$ and $S/\sqrt{B} > 0.3$ are referred to as *signal-enriched regions*, as they provide most of the sensitivity to the signal. The remaining regions are referred to as *background-dominated regions*, since they are almost purely background-only regions and are used to improve the background predictions and constrain their systematic uncertainties in the signal-enriched regions. There are six background-dominated regions, (4j, 2b), (4j, 3b), (4j, 4b), (5j, 2b), (5j, 3b), ($\geq 6j, 2b$), and three signal-enriched regions, (5j, $\geq 4b$), ($\geq 6j, 3b$) and ($\geq 6j, \geq 4b$). Each region is analysed separately and later combined statistically to maximise the overall sensitivity.

In order to visualize the background composition in the different analysis regions, the fractional contributions of the various background processes to the total prediction are summarised in a series of pie charts shown in Figure 6.2 (a).

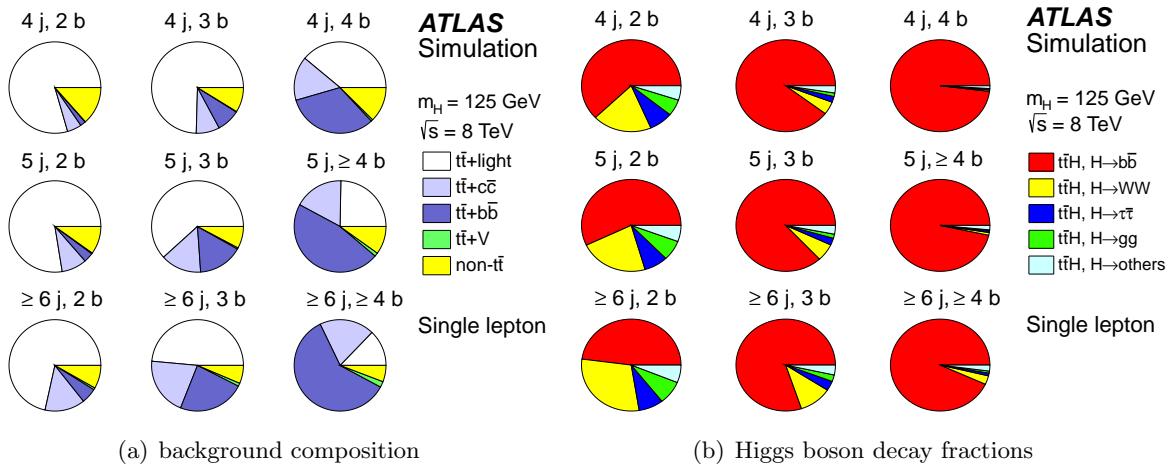


Figure 6.2.: The background compositions (a) and the fractions of the Higgs boson decay modes (b) are shown for the nine analysis regions. The jet and b -tag multiplicities are ordered by row and column, respectively [71].

In all regions, the background is clearly dominated by the $t\bar{t} +$ jets process with more than 90% contribution. While in the low jet and b -tag multiplicities it consists mainly of $t\bar{t} +$ light (white), the most signal sensitive regions with a high number of b -tags are dominated by $t\bar{t} + c\bar{c}$ (light blue) and $t\bar{t} + b\bar{b}$ (dark blue) events. Thus it is very important to have a good model of these backgrounds. Since the fractional contributions of the $t\bar{t} +$ HF ($t\bar{t} + b\bar{b}$, $t\bar{t} + c\bar{c}$) processes vary significantly in the different categories, the separate treatment of the regions gives a good handle to constrain the uncertainties on their normalisation. In the signal-enriched regions there are also small contributions from the not very well known $t\bar{t}V$ processes (green). The non- $t\bar{t}$ processes (yellow) are generally very small and thus combined in one slice. The main contribution to this slice comes from single top quark production. In the low jet and b -tag multiplicities there are also visible contributions from $W+$ jets background. Diboson and $Z+$ jets

6. Search for $t\bar{t}(H \rightarrow b\bar{b})$ Production in the Single Lepton Channel

production is almost completely negligible in all regions, while multijet production plays in the four jet bin and the ($5j, \geq 4b$) region a small role and is negligible elsewhere.

Although all Higgs boson decays are included in the $t\bar{t}H$ simulation and are treated as signal, this analysis is optimised for the $H \rightarrow b\bar{b}$ decay. The fractions of the different decay modes after the selection are shown for the different analysis regions in Figure 6.2 (b). In almost all regions and in particular in the signal-enriched regions the $H \rightarrow b\bar{b}$ decay (red) is clearly the dominant mode with more than 90% contribution in the most sensitive region ($\geq 6j, \geq 4b$). Only in categories with two b -tags, where the signal fractions are negligibly small, there are sizeable contributions from the $H \rightarrow WW^{(*)}$ decay (yellow). Hence it is justified to neglect the other decay modes when optimising the separation of signal and background in the signal-enriched regions.

6.3. Analysis Strategy

In order to maximize the search sensitivity and to learn as much as possible from the data about the predictions and their systematic uncertainties, each region is analysed separately and is later combined in a binned profile likelihood fit. Aiming for maximal sensitivity, an artificial neural network (NN) is employed in the signal-enriched regions. The NN uses in all regions kinematical variables and in the six jet regions additionally variables obtained with the MEM. It is trained to maximise the separation between background and $t\bar{t}H$ signal. In addition to the signal-enriched regions, a special NN is trained in the ($5j, 3b$) region to separate $t\bar{t} +$ light from $t\bar{t} +$ HF. Instead of aiming for signal sensitivity, this NN is employed to learn more from the data about the systematic uncertainties unique to the $t\bar{t} +$ HF processes. In all the remaining background-dominated regions, the scalar sum of the transverse energy of all jets H_T^{had} is used as a simple variable representing topological properties of the event. This variable is expected to be different in shape for the various processes and regions, but is at the same time easy to understand and to control. Furthermore, H_T^{had} is sensitive to the JES uncertainties, which strongly depend on the number of jets in the event. By using this variable in categories with different jet multiplicities in a profile likelihood fit, the JES uncertainties are expected to be constrained by the data. In Table 6.2 the variables used as discriminants in the different analysis regions are summarised.

Table 6.2.: Summary of the variables used in the profile likelihood fit in the different analysis regions.

Region	2 b -tags	3 b -tags	≥ 4 b -tags
4 jets	H_T^{had}	H_T^{had}	H_T^{had}
5 jets	H_T^{had}	NNHF	NN
≥ 6 jets	H_T^{had}	NN	NN

Due to the different contributing processes and kinematic properties, the impact of the various systematic uncertainties is very different in the individual analysis regions. Region ($4j, 2b$) has by far the largest statistics and thus provides powerful constraints on the overall normalisation of the background dominated by the $t\bar{t}$ process. The categories with two b -tags consist mainly of $t\bar{t} +$ light events. Its modelling uncertainties affecting shape and normalisation can be constrained due to the shape variation of H_T^{had} resulting from the different number of jets in the regions. By making use of regions with two, three and four b -tags, the tagging uncer-

tainties can be restricted. There are large contributions from $t\bar{t}$ +light events in the three b -tag categories, which can be used to regulate the uncertainties on mistagging. In particular the (4j, 3b) region, in which the third b -tag of $t\bar{t}$ +light events is mainly caused by a c -quark from the $W \rightarrow cs$ decay of the $t\bar{t}$ system, helps to decrease the uncertainties on c -tagging. Finally, the different fractions of $t\bar{t} + b\bar{b}$ and $t\bar{t} + c\bar{c}$ in the three and four b -tag regions provides very useful information to reduce the uncertainty on the normalisation of these processes.

Using the H_T^{had} and the NN output distributions, a binned profile likelihood fit is performed to combine all the different analysis regions. The likelihood is a function of the signal strength μ (parameter of interest) and the systematic uncertainties (*nuisance parameters (NPs)*). This allows for measuring the signal strength primarily from the signal-enriched regions and simultaneously constraining the systematic uncertainties using the statistical power of data in the background-dominated regions.

To avoid biasing the result of this analysis by looking at the outcome or the shape of the data in the signal sensitive regions, this analysis is at first performed blindly until the background and its improved prediction after the fit are fully understood. All the signal-enriched regions are blinded by removing the phase space with the highest signal significance using an appropriate cut on the NN output. As a consequence, the constraints from the data on the irreducible $t\bar{t} + b\bar{b}$ background dominating the signal-enriched regions can be studied, without being influenced by the signal. Although the profile likelihood fit is performed in the full phase space, the signal strength is not quantified, but merely the behaviour of the NPs and the improvements of the background predictions in the signal-depleted phase space is analysed. This procedure allows for developing a robust analysis strategy and verifying the improved modelling of the background without being biased by the presence of a signal. After the analysis is proven to be robust and unbiased, the cut on the NN variables is removed. At this stage no changes of the analysis strategy and procedure is allowed. All of the following studies show the results after the data has been unblinded.

6.4. Systematic Uncertainties

Systematic uncertainties might arise from an imperfect calibration of the experiment leading to a systematic bias of the measurement, from not considered experimental effects or from an inaccurate model to which the experimental data is compared. Unlike statistical uncertainties, they can not be reduced by simply increasing the experimental data, but by improving the knowledge about the experiment or the physics model. In high energy physics experiments, systematic uncertainties arise mainly from the imperfect knowledge or inaccuracies of the detector, which are propagated to the physics object definitions, and from certain assumptions made in the MC simulations to predict the experimental data. Some of the uncertainties might change the object properties, such that the selection acceptance is affected resulting in a different set of objects, on which the analysis is performed. In this case, the full analysis needs to be carried out on the modified set. This can be time consuming depending on the analysis methods. Other uncertainties only affect the number of events of a certain process or the shape of a certain kinematic distributions. In that case, the nominal analysis can be simply modified by applying a specific weight to the individual events without repeating the analysis. The smaller the experimental effect of interest or the higher the precision of the measurement, the more sources of systematic uncertainties need to be taken into account. In the case of this search, the expected signal fraction is tiny compared to the overwhelming background, as

6. Search for $t\bar{t}(H \rightarrow b\bar{b})$ Production in the Single Lepton Channel

discussed in Section 6.2. Thus many sources of systematic uncertainties need to be considered. Table 6.3 gives a summarised list of the evaluated systematic uncertainties consisting of more than 100 components.

Table 6.3: List of the considered systematic uncertainties. The type "N" and "S" means that the uncertainty for all processes and regions affects only normalisation or shape, respectively, whereas "SN" means that the uncertainty affects both. Some of the systematic uncertainties are split into several components for a more accurate treatment.

Systematic uncertainty	Type	Components
Luminosity	N	1
Physics Objects		
Electron	SN	5
Muon	SN	6
Jet reconstruction	SN	1
Jet energy scale	SN	22
Jet vertex fraction	SN	1
Jet energy resolution	SN	1
b -tagging efficiency	SN	6
c -tagging efficiency	SN	4
Light-jet tagging efficiency	SN	12
High- p_T tagging efficiency	SN	1
Background Model		
$t\bar{t}$ cross section	N	1
$t\bar{t}$ modelling: parton shower	SN	3
$t\bar{t}$ modelling: p_T reweighting	SN	9
$t\bar{t}+c\bar{c}$: p_T reweighting	SN	2
$t\bar{t}+h$ -flavour: normalisation	N	2
$t\bar{t}+b\bar{b}$: NLO Shape	SN	8
$t\bar{t}+c\bar{c}$: generator	SN	4
$W+jets$ normalisation	N	3
W p_T reweighting	SN	1
$Z+jets$ normalisation	N	3
Z p_T reweighting	SN	1
Diboson+jets normalisation	N	3
Lepton misID normalisation	N	2
Lepton misID shape	S	2
Single top cross section	N	1
Single top model	SN	1
$t\bar{t}V$ cross section	N	1
$t\bar{t}V$ model	SN	1
Signal Model		
$t\bar{t}H$ scale	SN	2
$t\bar{t}H$ generator	SN	1
$t\bar{t}H$ hadronisation	SN	1
$t\bar{t}H$ PDF	SN	1

The type indicates if the systematic uncertainty affects only the normalisation (N), only the shape (S) of the discriminating variables or both (SN). For a more accurate treatment, some of the sources of systematic uncertainties are divided into components. The number of components is also specified in the table. All systematic uncertainties are generally applied to all signal and background processes and in all analysis regions, if applicable and not specified differently. Independent sources of systematic uncertainties are treated as uncorrelated. However, if correlations exist between sources, components or among processes, they are maintained. In general, the uncertainties are correlated across the analysis regions. This leads to a reduction of the systematic uncertainties in the signal-enriched regions, when using them as NPs in the profile likelihood fit, by exploiting the constraining power of the data in the high statistics background-dominated regions. The following section describes individual sources of systematic uncertainties of this analysis in more detail.

6.4.1. Luminosity

The integrated luminosity of the 8 TeV dataset is measured with an uncertainty of 2.8% [101]. Since the MC simulations are normalised to the same integrated luminosity, this uncertainty is applied to all simulated processes, affecting only the normalisation.

6.4.2. Uncertainties on Physics Objects

Several systematic uncertainties arise from the object reconstruction as described in Section 4.3 due to the determination of correction SF applied to MC to compensate the differences between data and predictions.

Lepton

In case of the lepton selection, several sources of systematic uncertainties originate from the measurement of the efficiencies of the trigger, reconstruction, identification and isolation as well as from the lepton momentum scale and resolution corrections. There are in total five uncertainty components for electrons and six for muons. Since the correction SFs applied to MC are derived in bins of lepton η and p_T (E_T), the uncertainties on it affect the shape as well as the normalisation. Due to the independent measurements of the various SF, the uncertainties of each component are treated uncorrelated. Combining all components of the lepton uncertainties, the total yield changes by roughly 1.5% across all processes and regions, which is a minor effect.

Jet

In case of the jet selection, uncertainties arise from the jet reconstruction, the JES calibration, the JVF requirement and corrections to the jet energy resolution (JER). Among these the JES uncertainty has the largest influence on the analysis. Since all of the quantities are measured depending on the jet kinematics, they all affect the shape and the normalisation of the discriminant.

Jet Reconstruction Although the reconstruction efficiency is consistent between simulation and data for high p_T jets, for jets with a momentum below 30 GeV the efficiency obtained in simulations is by 0.2% lower than in data [104]. Since jets with at least 25 GeV are required

6. Search for $t\bar{t}(H \rightarrow b\bar{b})$ Production in the Single Lepton Channel

for this analysis, the uncertainty on the discriminating variables is evaluated by removing randomly 0.2% of the jets with $25 \text{ GeV} \leq p_T \leq 30 \text{ GeV}$. Since this procedure affects the acceptance and thus events entering the different analysis regions, the selection is repeated for the modified set of jets. The total prediction in the most sensitive analysis regions changes by around 3% due to this uncertainty.

Jet Energy Scale The JES and its uncertainties are determined by combining information from test-beam data, LHC collision data and simulation using partially complementary in-situ calibration techniques, in which the momentum balance of multijets, $Z+\text{jets}$ and $\gamma+\text{jets}$ events is exploited [104]. Although more than 50 sources of systematic uncertainties from the independent measurements are partially combined, in the end there are still 22 mostly uncorrelated components to be taken into account. They arise from the following sources:

- **in-situ-techniques (12):** uncertainties from different in-situ techniques are grouped into **statistical (3)**, **detector (3)**, **modelling (4)** and **mixed (2)** categories. Correlations among the sources and across p_T bins are preserved.
- **pile-up (4):** uncertainties on the pile-up corrections due to mismodelling, on the number of primary vertices and average interactions per bunch crossing.
- **η -intercalibration (2):** the calibration in the forward regions is improved by comparing to more precisely measured central jets making use of the p_T -balance of jets in different η regions. A statistical component and a MC modelling uncertainty are considered.
- **flavour response and composition (2):** jets originating from quarks or gluons respond differently to the calorimeter, while mainly quark-initiated jets are calibrated using the in-situ techniques. The differences to gluon jets in MC simulations are taken as uncertainties and are weighted with the relative composition in the various processes.
- **b-JES (1):** jets originating from b -quarks respond differently to the calorimeter than the quark jets used for the calibration.
- **high p_T jets (1):** due to limited statistics JES uncertainties are derived from single-particle response measurements for jets with $p_T > 1 \text{ TeV}$.

The components depend on the jet p_T and η and thus affect the shape and normalisation. In Figure 6.3 the relative JES uncertainty split into categories is shown as a function of transverse momentum (a) and pseudorapidity (b) of the jet.

The uncertainties are below 4% in the full p_T range in central η regions. While they are relatively constant for jet $|\eta| < 2.0$, they drop from 4% for low p_T jets to below 2% for jets with p_T above 100 GeV. Only for very high momentum the uncertainties increase again due to the low data statistic and usage of alternative calibration methods. Although the uncertainties seem to be small, they add up for multiple jets leading to uncertainties of larger than 10% in the most sensitive regions of this analysis. Since the JES uncertainties affect the jet p_T and consequently the acceptance, the full analysis needs to be run for each component, which is in particular challenging for computationally intensive steps like the calculation of the process probabilities using the MEM. However, a separate treatment of the components is highly beneficial in the profile likelihood fit, since they can be reduced individually taking the correlations among analysis regions into account. In this analysis, the largest JES uncertainty

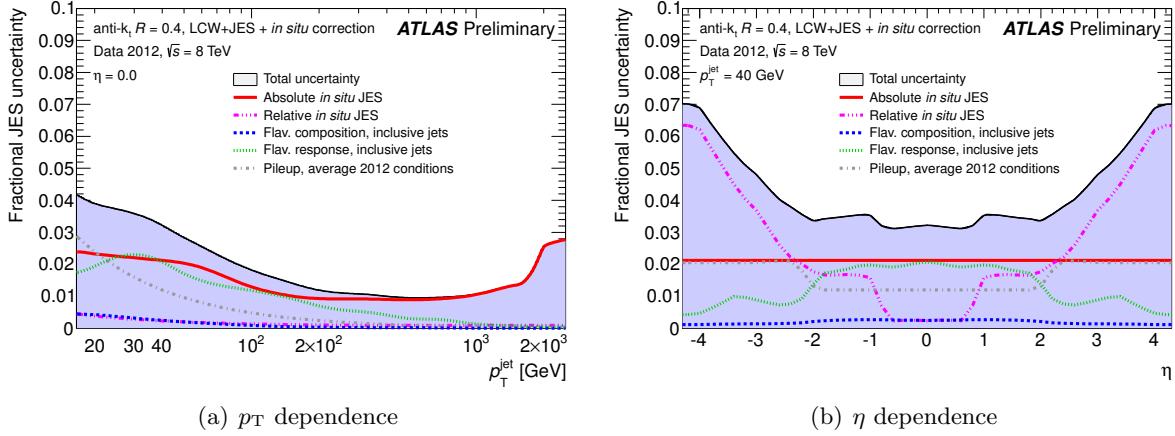


Figure 6.3.: Fractional in-situ and sample dependent JES systematic uncertainty as a function of jet p_T (a) and η (b). The total uncertainty (all components summed in quadrature) is shown as a filled blue region topped by a solid black line. Average 2012 pile-up conditions are used and topology dependent components are taken from inclusive di-jet samples [104].

arises from the modelling uncertainties in the η -intercalibration. Because in the high jet multiplicity regions the shape effect on the $t\bar{t}$ distributions is correlated to the signal distribution, it is the second leading uncertainty.

Jet Vertex Fraction As shown in Figure 4.2 (b) the JVF efficiencies for a single jet are measured in $Z(\rightarrow \ell^+\ell^-)+1$ -jet events enriched with pile-up and hard scattering processes. They differ in data and simulation and uncertainties for the corresponding correction SFs are determined. The corresponding uncertainty in the analysis is evaluated by changing the nominal JVF cut value by its uncertainty, leading to a modified set of jets on which the analysis is performed.

Jet Energy Resolution The JER is measured using the di-jet balance and bisector technique depending on the jet p_T and η in data and simulation [209]. Although the resolution agrees generally well between data and simulation, differences in particular in high- p_T and high- η regions are observed. They give rise to an additional source of systematic uncertainty. The uncertainty is estimated by smearing the p_T of the simulated jet by this residual difference, which affects again the selection of the jets. Since this smearing will only broaden the resolution, this uncertainty by definition produces a one-sided effect on the normalisation and shape of the discriminant. It is symmetrised for the profile likelihood fit. In the ($\geq 6j, \geq 4b$) region, the yield of the $t\bar{t} + b\bar{b}$ and $t\bar{t} +$ light events is affected by 1.6% and 4.4%, respectively.

Heavy- and Light-Flavour Tagging

As described in Section 4.3.1, the measured b -, c - and light-tagging efficiencies differ in data and simulations. Systematic uncertainties arise from the derived correction SF. In case of the b -tagging SF uncertainties, the ten bins in jet p_T as shown in Figure 4.3 (b) are combined into six bins. In case of the c -tagging and light-tagging SF uncertainties, the same jet p_T binning as shown in Figure 4.4 (b) and Figure 4.5 (b) is used leading to four and twelve uncertainties,

respectively. The light-tagging SFs are calculated for two different η regions, each composed of six p_T bins. In order to take the bin-to-bin correlations into account, the uncertainties are obtained by diagonalising the covariance matrix containing the uncertainties per jet p_T bin and their corresponding correlations, which is also referred to as *eigenvector method*. An additional uncertainty arises from the extrapolation of the b -tagging efficiency measurement to high p_T regions. The systematic uncertainties are assumed to be uncorrelated between b -, c - and light-tagging. All of the tagging uncertainties affect the shape and the normalisation of the discriminant due to the dependence on the jet p_T and η in case of light-tagging. Since the differences in the discriminating variables using direct tagging and the TRF predictions (see Sec. 4.4.4) are negligible compared to the tagging uncertainties, no additional uncertainty is assigned due to the use of the TRF method. In the ($\geq 6j, \geq 4b$) region, the uncertainties on the b -tagging efficiencies affect the $t\bar{t} + b\bar{b}$ event yield by roughly 9%, the uncertainties on the c -tagging efficiencies affect the $t\bar{t} + c\bar{c}$ event yield by 12% and the uncertainties on the light-tagging efficiencies the $t\bar{t} +$ light event yield by 19%. The effect of the high p_T tagging uncertainty on the yield is below 1% for all processes.

6.4.3. Uncertainties on Background Modelling

$t\bar{t}+\text{jets}$ Modelling

The inclusive $t\bar{t}$ production cross section is calculated at NNLO with an uncertainty of roughly 6%, which includes uncertainties on the PDF, the choice of α_S , the choices of the various scales and the top quark mass [45–49, 154]. The uncertainties arising from the PDF and α_S choice are calculated by using the envelope of the uncertainties of the PDF sets from CT10 NNLO [210], MSTW2008 NNLO (68% CL) [190, 191] and NNPDF2.3 5F FFN [192] obtained with the PDF4LHC prescription [211]. The uncertainty on the $t\bar{t}$ cross section affects naturally only the normalisation of the sample.

In order to evaluate the uncertainty on the parton shower and hadronisation model, the POWHEG+PYTHIA events are compared to POWHEG+HERWIG events and the symmetrised differences in shapes of the discriminants are taken as uncertainty. Because the change of the shower and hadronisation model affects the number of jets as well as the flavour composition, the uncertainty is split into three uncorrelated components acting separately on the $t\bar{t} +$ light, $t\bar{t} + c\bar{c}$ and $t\bar{t} + b\bar{b}$ contribution. These uncertainties result in variations of the $t\bar{t}$ yield in the most sensitive region by 11–16%.

As described in Section 4.4.3, $t\bar{t} +$ light and $t\bar{t} + c\bar{c}$ events are reweighted to correct for the disagreements observed in the top quark and $t\bar{t}$ p_T distributions based on the $\sqrt{s} = 7$ TeV differential cross section measurement [157]. The nine largest uncertainties affecting the differential top quark and $t\bar{t}$ p_T measurements represent approximately 95% of the total experimental uncertainty on this measurement and are taken as systematic uncertainties on the reweighting procedure. These shape and normalisation affecting uncertainties include the uncertainty on the radiation modelling in $t\bar{t}$ events, the generator choice, the jet energy scale and resolution as well as the flavour tagging. The measurement is performed on inclusive $t\bar{t}$ events, which consist mainly of $t\bar{t} +$ light events, and hence the size of the uncertainties applicable to the $t\bar{t} + c\bar{c}$ component is unknown. In order to account for that, the full effect of the reweighting is taken as two additional uncorrelated uncertainties applied to $t\bar{t} + c\bar{c}$ events. In the ($\geq 6j, \geq 4b$) region, the yield of $t\bar{t} +$ light and $t\bar{t} + c\bar{c}$ events is changed by roughly 5% and 6%, respectively. Although more precise NLO predictions of the $t\bar{t} + b\bar{b}$ cross section became recently available

reducing the uncertainties to 30% [155], these predictions are sensitive to the considered phase space and HF definitions. In case of $t\bar{t} + c\bar{c}$ events, an NLO prediction is not available. Because this analysis is very sensitive to the $t\bar{t} + b\bar{b}$ modelling and the constraining power of the data allows to determine the $t\bar{t} + \text{HF}$ normalisation with better precision, a conservative uncertainty of 50% is assigned to the $t\bar{t} + b\bar{b}$ and $t\bar{t} + c\bar{c}$ production. Detailed generator studies comparing the production of $t\bar{t} + b\bar{b}$ events using LO predictions from POWHEG+PYTHIA and NLO predictions from SHERPAOL show that the cross sections agree within this uncertainty. The uncertainties on the $t\bar{t} + b\bar{b}$ and $t\bar{t} + c\bar{c}$ normalisation are considered uncorrelated between the two processes, but each is correlated across all analysis regions. They are the leading and third leading uncertainty in the analysis, respectively.

From the reweighting of the $t\bar{t} + b\bar{b}$ background, as described in Section 4.4.3, additional eight uncertainties affecting shape and normalisation arise. Three scale uncertainties are estimated changing the SHERPAOL prediction by varying the scale by a factor of two up and down, by using an alternative functional form of the renormalisation scale $\mu_R = (m_t m_{b\bar{b}})^{1/2}$ and by changing the factorisation μ_F and resummation μ_Q scales to $\mu_F = \mu_Q = \prod_{i=t,\bar{t},b,\bar{b}} E_{T,i}^{1/4}$. These variations affect the distribution of events between the sub categories of the $t\bar{t} + b\bar{b}$ sample as shown in Figure 6.4 (a).

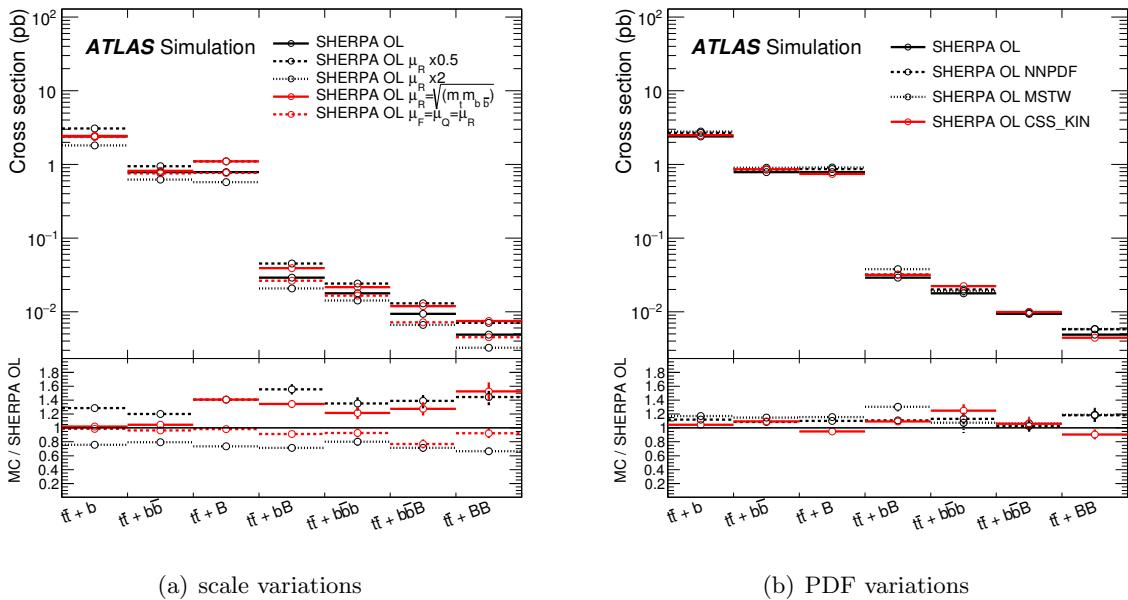


Figure 6.4.: Systematic uncertainties on the $t\bar{t} + b\bar{b}$ contribution based on (a) scale variations and (b) PDF choice and shower recoil model of the SHERPAOL simulation. The effect of a given systematic uncertainty is shown across the different $t\bar{t} + b\bar{b}$ sub categories. The effect of migration between categories is covered by variations of these systematic uncertainties [71].

Additionally, the SHERPAOL predictions are varied by choosing a different PDF set (NNPDF, MSTW) and by changing the shower recoil model (CSS_KIN). The differences to the default SHERPAOL prediction in the different $t\bar{t} + b\bar{b}$ sub categories are shown in Figure 6.4 (b) and are taken as three uncertainty components. As visible in Figure 4.7 $t\bar{t} + b\bar{b}$ production via MPI and FSR is not modelled with SHERPAOL. Thus two further uncertainties are estimated by comparing the POWHEG+PYTHIA predictions with and without these two effects. The uncer-

6. Search for $t\bar{t}(H \rightarrow b\bar{b})$ Production in the Single Lepton Channel

tainty due to the scale choice $\mu_R = (m_t m_{b\bar{b}})^{1/2}$ and the shower recoil scheme result in a large shape variation of the discriminants in the most sensitive analysis region and are the fourth and sixth leading uncertainties in this analysis.

Four systematic uncertainties on the $t\bar{t} + c\bar{c}$ prediction are determined by comparing the POWHEG+PYTHIA predictions with an alternative MADGRAPH+PYTHIA simulation, since the latter includes matrix element calculations of the $t\bar{t} + c\bar{c}$ process. The MADGRAPH+PYTHIA prediction is then varied by simultaneously changing the factorisation and renormalisation scales by a factor of two, by doubling the matching threshold of 20 GeV and varying the c -quark mass by 100 MeV. In the ($\geq 6j, \geq 4b$) region the $t\bar{t} + b\bar{b}$ and $t\bar{t} + c\bar{c}$ event yields are varied by roughly 8% and 16% due to these HF modelling uncertainties.

$W/Z+\text{jets}$ and Diboson Modelling

Three normalisation uncertainties on each of the $W/Z+\text{jets}$ and diboson+jets background, as described in Section 4.4.3, arise from the normalisation to the inclusive cross section calculations and the extrapolation to the five jet exclusive and the six jet inclusive jet multiplicities [164, 212]. The extrapolation uncertainties are large in order to cover both normalisation and HF composition uncertainties in high jet multiplicity bins and are considered uncorrelated between the regions and the processes. One additional uncertainty is assigned to the $W/Z+\text{jets}$ backgrounds due to the vector boson p_T reweighting as discussed in Section 4.4.5, which is estimated by taking the symmetrised full difference between applying and not applying the reweighting as systematic uncertainty. Although the same p_T reweighting is applied to both V bosons, the uncertainty is assumed to be uncorrelated between the two processes.

Misidentified Lepton Background Modelling

The general sources of systematic uncertainties on the misidentified lepton background are already described in Section 4.4.5. Since this analysis exploits a phase space with high jet and b -tag multiplicities, in which the estimate suffers from limited data statistics of loose and tight events, a conservative normalisation uncertainty of 50% is used. The uncertainties are considered to be uncorrelated between both lepton channels, but correlated across the analysis regions. Two shape uncertainties for both lepton channel are estimated from the uncertainty on the prompt lepton background subtraction and the alternative efficiency measurements as explained in Section 4.4.5.

Single Top Quark Modelling

Uncertainties of +5%/-4% are assigned to the normalisation of the single top quark production, which corresponds to the weighted average of the uncertainties on the theoretical cross section calculation of the s -, t - and Wt -channel [51, 52]. One shape uncertainty is assessed by comparing two sets of MC simulations of the Wt -channel using different schemes to account for the interference with $t\bar{t}$ production diagrams.

$t\bar{t}V$ Modelling

The uncertainty on the normalisation of the $t\bar{t}V$ production is assumed to be 30% due to the uncertainties on the theoretical cross section calculation [160, 161]. One more shape and

normalisation affecting uncertainty arises from the variation of the amount of ISR. Because the $t\bar{t}Z$ background with the Z boson decaying into a $b\bar{b}$ pair is an irreducible background to the $t\bar{t}(H \rightarrow b\bar{b})$ signal with very similar kinematic properties, the shape of the discriminant is similar to the one of the signal. As a consequence, the signal extraction is very sensitive to variations of the $t\bar{t}V$ contribution. Therefore the $t\bar{t}V$ background normalisation is the fifth leading uncertainty in the analysis.

6.4.4. Uncertainties on Signal Modelling

In order to evaluate the scale dependency of the $t\bar{t}H$ production, alternative sets of NLO PowHEL events are produced at parton level, in which the default static factorisation and renormalisation scales $\mu_F = \mu_R = m_t + m_H/2$ are varied by a factor of two up and down and are replaced by the dynamic scale $\mu_F = \mu_R = (m_T^t m_T^{\bar{t}} m_T^H)^{1/3}$. The effect on the nominal PowHEL sample as described in Section 4.4.2 is then assessed by reweighting the events at parton level to reproduce the variations. The two uncertainties related to the scale variations, which affect both shape and normalisation, are then obtained by comparing the discriminants at reconstruction level.

Another uncertainty, which accounts for the choice of the generator, is estimated by comparing the PowHEL+PYTHIA events to MADGRAPH5_AMC@NLO+HERWIG++ events [66, 213, 214]. The uncertainty due to the choice of the shower and hadronisation model is evaluated by comparing the discriminants obtained with PowHEL+PYTHIA and PowHEL+HERWIG samples. Finally, an uncertainty arising from the choice of the PDF is estimated following the PDF4LHC prescription [211]. All $t\bar{t}H$ modelling uncertainties affect the yield of the process in the ($\geq 6j, \geq 4b$) region by 2.7%.

6.5. Using the Matrix Element Method

Following the descriptions of Chapter 5, the process probabilities for the $t\bar{t}H$ and $t\bar{t} + b\bar{b}$ hypotheses are calculated for all predicted and data events in the ($\geq 6j, \geq 4b$) and ($\geq 6j, 3b$) regions. The top quark mass is set to $m_{top} = 172.5$ GeV for both hypotheses, since this value is used in the event generation of the MC simulations. In case of the $t\bar{t}H$ hypothesis a Higgs boson mass of 125 GeV is used. The full statistics of the nominal samples with the full detector simulation as well as with the fast simulation in case of $t\bar{t}H$ and $t\bar{t} +$ jets events is evaluated. In addition, almost all systematic variations, which affect the acceptance of the events, corresponding to 35 variations of the nominal sample, are evaluated as well. Since the statistics of the samples are kept high for the NN training by making use of the TRF method, the process probabilities for a total of around 62 Million events (6 M nominal, 56 M systematic variations) are calculated. With an average of two minutes per event to calculate the likelihood of both hypotheses for the twelve most contributing assignment permutations, this corresponds to a total runtime of 2.1 Million CPUh or 236 CPU years, which has been completed within two month of real time. It is clear, that without a huge computational effort (see Sec. 5.8) and without making heavy use of parallel computing provided by the WLCG, this analysis would not have been possible. In the following section the results from this run are discussed, by showing a selection of the distributions obtained with the MEM. Since the main goal of the application of this method in this analysis is to improve the separation between $t\bar{t}H$ and background events, the distributions of the signal predictions are often compared to

6. Search for $t\bar{t}(H \rightarrow b\bar{b})$ Production in the Single Lepton Channel

the total background distributions. In order to quantify the separation between distributions the definition for separation

$$\langle s^2 \rangle = \frac{1}{2} \int_{-\infty}^{\infty} \frac{(S(x) - B(x))^2}{S(x) + B(x)} dx \quad (6.1)$$

between two histograms $S(x), B(x)$ is used, which is related to the non-overlapping area of the distributions.

6.5.1. Likelihood Functions

The logarithms of the $t\bar{t}H$ likelihood function in the best assignment permutation $\ln \mathcal{L}_{t\bar{t}H}^{\max}$ (see Eq. 5.23) and summed over all possible permutations $\ln \mathcal{L}_{t\bar{t}H}^{\text{sum}}$ (see Eq. 5.22) are shown in Figure 6.5 (a) and (b), respectively, in the ($\geq 6j, \geq 4b$) region for the total background (solid blue) and the $t\bar{t}H$ signal (dashed red) normalised to unity.

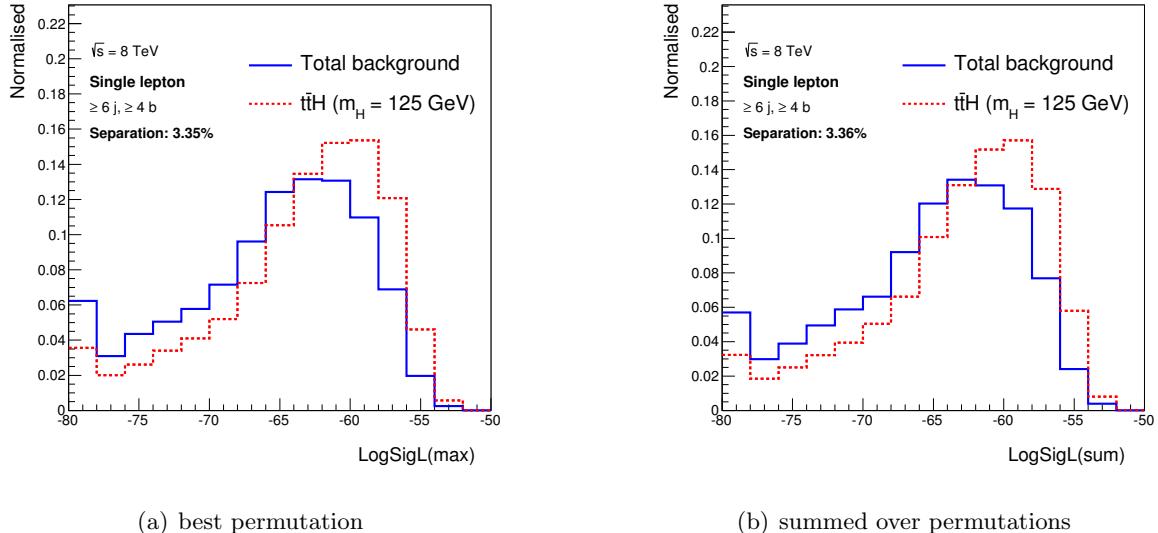


Figure 6.5.: The logarithm of the $t\bar{t}H$ likelihood functions (a) in the best and (b) summed over all possible assignment permutations in the ($\geq 6j, \geq 4b$) region for the total background (solid blue) and the $t\bar{t}H$ signal (dashed red) predictions.

Since the largest contribution to the $\ln \mathcal{L}_{t\bar{t}H}^{\text{sum}}$ distribution (b) comes from the best permutation, the two distributions look very similar in Figure 6.5 (a) and (b) and yield both a good separation of approximately 3.4%. The $t\bar{t}H$ MC events give as expected a higher likelihood response to the signal hypothesis with a peak around $\ln \mathcal{L}$ values of -57. The higher values are mainly caused by a large contribution from the Higgs boson mass divergence in the matrix element calculation coming from events, in which the two b -jets originating from a Higgs boson are present and correctly assigned. The background events, consisting dominantly of the $t\bar{t} + b\bar{b}$ process, respond on average with significantly lower $\ln \mathcal{L}$ values. The $\ln \mathcal{L}$ distributions have a less pronounced peak around -63 and a large spread to very low values, since most of them do not match the $t\bar{t}H$ hypothesis very well. However, few of the background events yield $\ln \mathcal{L}$ values close to the best values obtained for $t\bar{t}H$ events, because two jets might match accidentally.

tally the Higgs boson mass hypothesis. The likelihood distributions obtained with the $t\bar{t} + b\bar{b}$ background hypothesis are presented in Figure 6.6

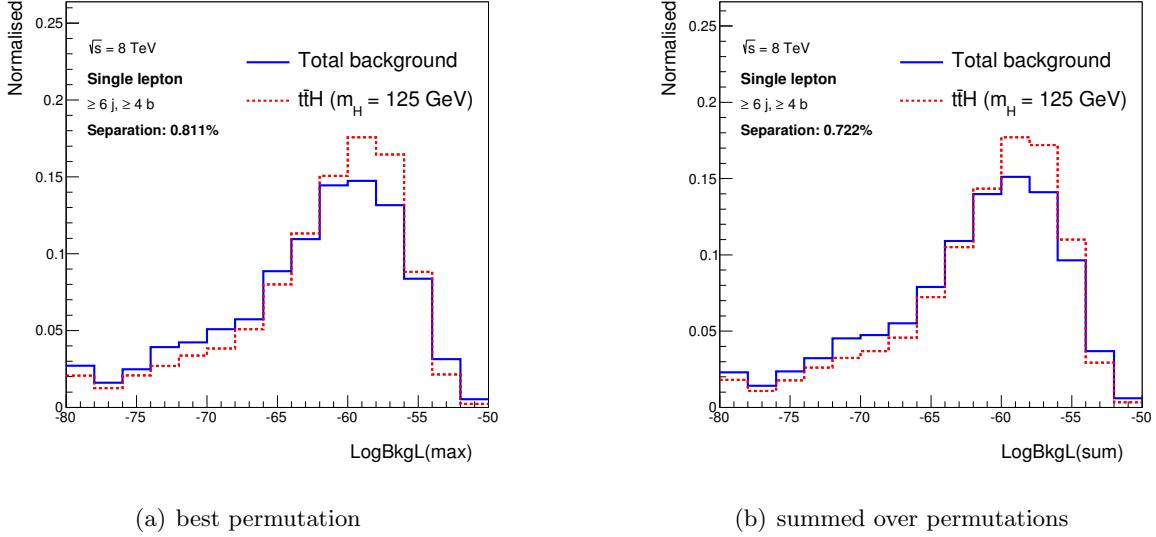


Figure 6.6.: The logarithm of the $t\bar{t} + b\bar{b}$ likelihood functions (a) in the best and (b) summed over all possible assignment permutations in the ($\geq 6j, \geq 4b$) region for the total background (solid blue) and the $t\bar{t}H$ signal (dashed red) predictions.

The $\ln \mathcal{L}_{t\bar{t}+b\bar{b}}^{\max}$ distributions are again very similar to the $\ln \mathcal{L}_{t\bar{t}+b\bar{b}}^{\text{sum}}$ distributions, though the first yields a slightly better separation of 0.81% compared to 0.72% in the latter. Although the highest likelihood values are calculated using background events, the response from $t\bar{t}H$ events is very similar compared to the one from the background. Both event types peak around $\ln \mathcal{L}$ values of -58. Both peaks are slightly sharper than for the $t\bar{t}H$ hypothesis. The peak from the $t\bar{t}H$ events is even more pronounced than the one from the background. It is much easier for jets originating from the Higgs boson decay to mimic a gluon radiating from the $t\bar{t}$ system in the $t\bar{t} + b\bar{b}$ hypothesis than the other way around. Due to the higher required mass of the $t\bar{t}H$ system, it is less boosted by ISR and thus fits in general better the LO picture of the matrix element calculation. Apart from the Higgs boson divergence, the differences in the kinematics of the reconstructed final states between $t\bar{t}H$ and $t\bar{t} + b\bar{b}$ events are minor. Hence it is difficult to distinguish these two processes by testing the $t\bar{t} + b\bar{b}$ hypothesis. The $\ln \mathcal{L}_{t\bar{t}H}^{\max}$ and $\ln \mathcal{L}_{t\bar{t}H}^{\text{sum}}$ distributions in the ($\geq 6j, 3b$) region are shown for the total background and $t\bar{t}H$ events in Figure 6.7.

In this region, the $t\bar{t}$ +light events contribute considerably more to the total background. Typically one b -tag arises from a mistag of a c -quark originating from the hadronically decaying W boson. Since b -tagged jets are not assigned to the quarks from the W boson decay this leads to a wrong assignment of the jets to the partons from the start. As before, both distributions in Figure 6.7 look similar. Compared to the ($\geq 6j, \geq 4b$) region, the separation between the two processes is significantly reduced to approximately 1.6% in both cases. However, the $t\bar{t}H$ events still occupy the higher $\ln \mathcal{L}$ values on average and peak around a slightly higher $\ln \mathcal{L}$ value of -59. With only three b -tags, there are now 36 possible jet to parton assignments instead of twelve, for which the process probability is calculated. With such increase of free-

6. Search for $t\bar{t}(H \rightarrow b\bar{b})$ Production in the Single Lepton Channel

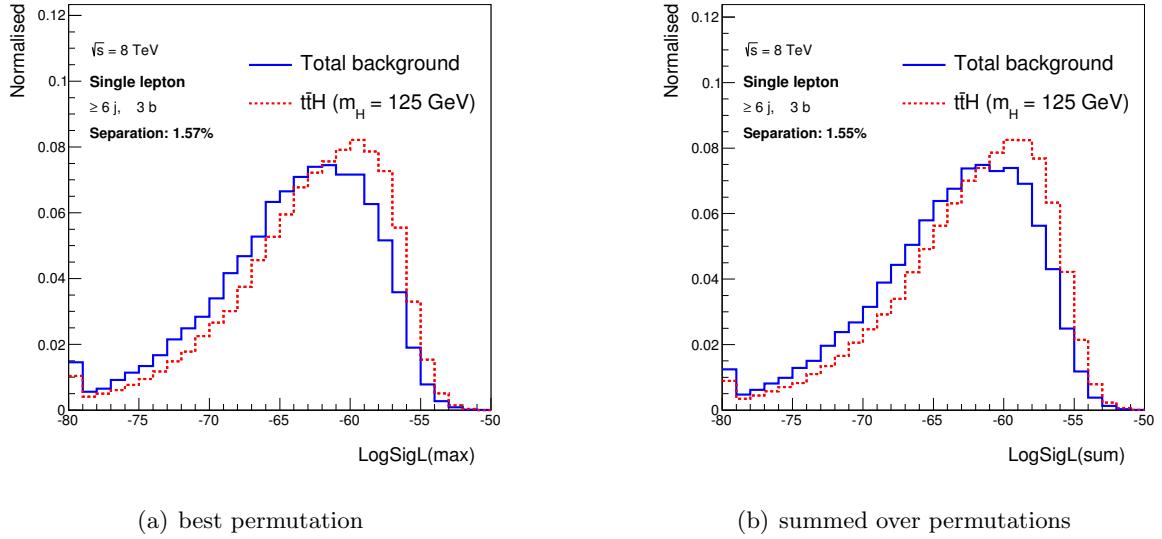


Figure 6.7.: The logarithm of the $t\bar{t}H$ likelihood functions (a) in the best and (b) summed over all possible assignment permutations in the ($\geq 6j, 3b$) region for the total background (solid blue) and the $t\bar{t}H$ signal (dashed red) predictions.

dom, it is more difficult to highly rank the correct assignment in the permutation evaluation phase of the integration for $t\bar{t}H$ events. This results in less precise integrations of the correct permutation and hence in lower likelihood values. On the other hand, it is much easier to find two jets in $t\bar{t} + bb$ events, which are close to the evaluated Higgs boson mass. Accordingly, the background events yield higher likelihood values leading to less separation. However, there is still valuable information inside these distributions, in particular when aiming for separating the non- $t\bar{t}$ background from the signal (see Fig. A.4). Figure 6.8 shows the $\ln \mathcal{L}_{t\bar{t}+bb}^{\max}$ and $\ln \mathcal{L}_{t\bar{t}+bb}^{\sum}$ distributions in the ($\geq 6j, 3b$) region.

For both choices of calculating the likelihood the distributions look similar. Due to the increased number of possible assignment permutations the responses from signal and background events result in almost the same shape. The peak for $t\bar{t}H$ events is more pronounced due to the smaller $t\bar{t}H$ system p_T . The resulting separation of around 0.26% is very small, indicating that $t\bar{t}H$ and $t\bar{t} + bb$ events are almost indistinguishable under the $t\bar{t} + bb$ hypothesis. Since no major differences between the two choices of the likelihood distribution \mathcal{L}^{\max} and \mathcal{L}^{\sum} are visible and the latter is more robust against approximations during the integration, only the summed likelihood \mathcal{L}^{\sum} will be discussed in the following.

In order to compare the differences in response to the $t\bar{t}H$ and $t\bar{t} + bb$ hypothesis from $t\bar{t} +$ light and $t\bar{t} +$ HF events, the corresponding $\ln \mathcal{L}_{t\bar{t}H}^{\sum}$ and $\ln \mathcal{L}_{t\bar{t}+bb}^{\sum}$ distributions are presented in Figure 6.9 for the ($\geq 6j, \geq 4b$) region.

For both hypotheses the $t\bar{t} +$ HF events (shaded pink) yield slightly higher likelihood values than the $t\bar{t} +$ light events (filled blue), since the objects as well as the kinematics of the final state match better the tested hypothesis. However, the separation between these two event types of 0.26% and 0.74% for the signal and background hypothesis, respectively, is relatively small. In the ($\geq 6j, 3b$) region shown in Figure 6.10, these difference are even smaller due to the increase of freedom in the assignments discussed above.

In case of the signal hypothesis (a) the separation is only 0.16% and in case of the background

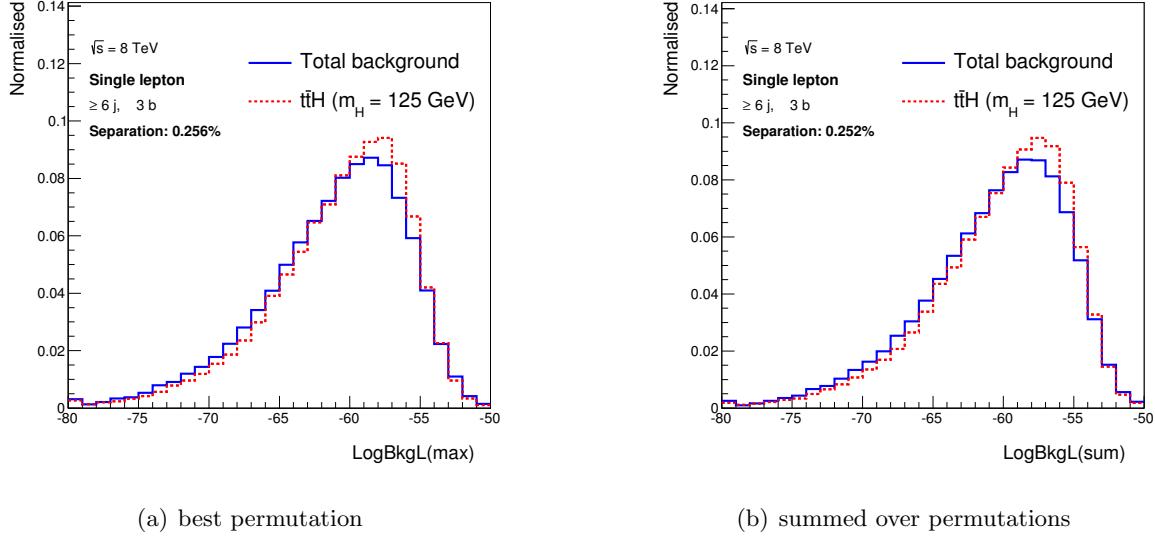


Figure 6.8.: The logarithm of the $t\bar{t} + b\bar{b}$ likelihood functions (a) in the best and (b) summed over all possible assignment permutations in the ($\geq 6j, 3b$) region for the total background (solid blue) and the $t\bar{t}H$ signal (dashed red) predictions.

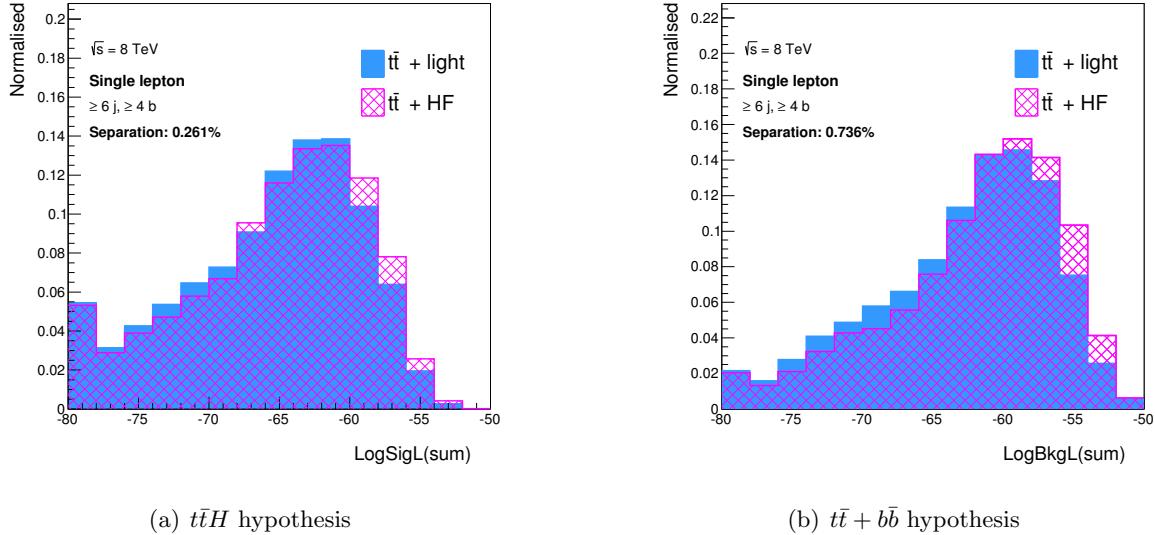


Figure 6.9.: The logarithm of (a) the $t\bar{t}H$ and (b) $t\bar{t} + b\bar{b}$ likelihood functions summed over all possible assignment permutations in the ($\geq 6j, \geq 4b$) region for $t\bar{t} + \text{light}$ (filled blue) and $t\bar{t} + \text{HF}$ (shaded pink) events.

hypothesis (b) the discrimination is consistent with zero. Finally, the $\ln \mathcal{L}_{t\bar{t}H}^{\text{sum}}$ distribution is presented in the ($\geq 6j, 3b$) and ($\geq 6j, \geq 4b$) regions for the model predictions compared to data in Figure 6.11.

As expected from the observed disagreement in the total yield between data and predictions, as shown in Figure 6.1, the predictions significantly underestimate the data in both regions.

6. Search for $t\bar{t}(H \rightarrow b\bar{b})$ Production in the Single Lepton Channel

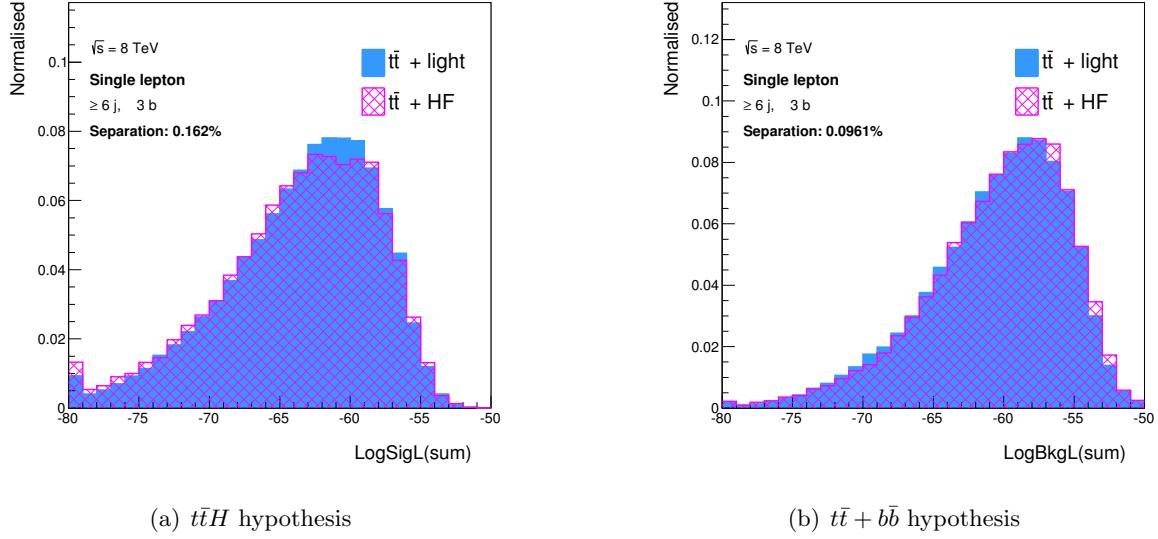


Figure 6.10.: The logarithm of (a) the $t\bar{t}H$ and (b) $t\bar{t} + b\bar{b}$ likelihood functions summed over all possible assignment permutations in the ($\geq 6j, 3b$) region for $t\bar{t}$ + light (filled blue) and $t\bar{t}$ + HF (shaded pink) events.

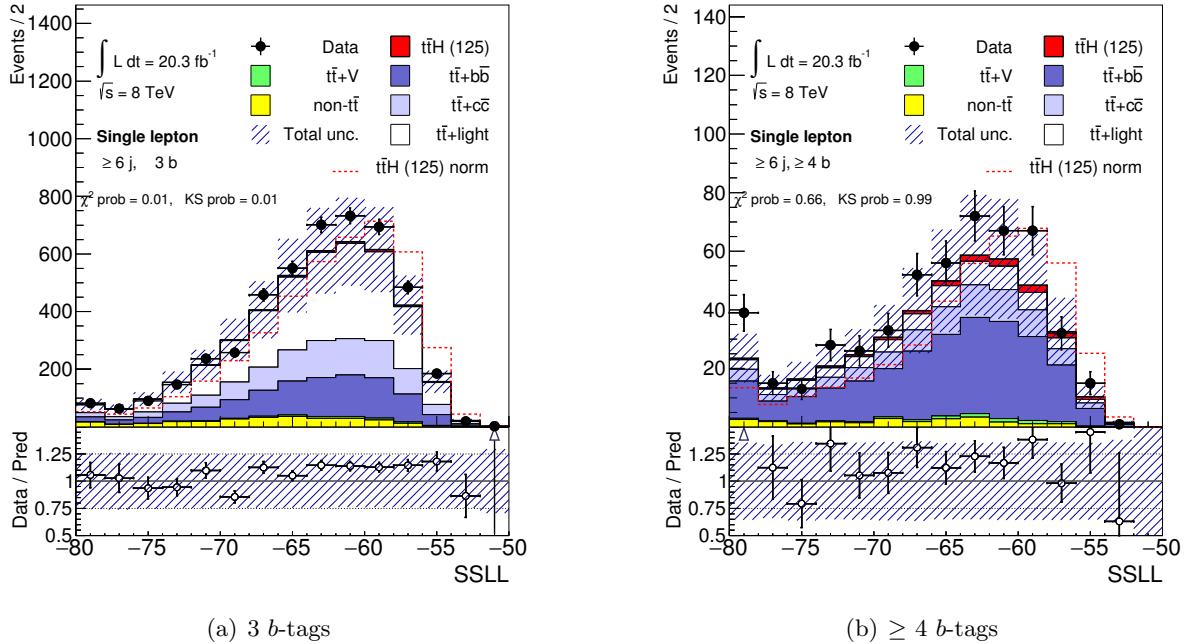


Figure 6.11.: The logarithm of the $t\bar{t}H$ likelihood functions summed over all possible assignment permutations in the (a) ($\geq 6j, 3b$) and (b) ($\geq 6j, \geq 4b$) regions for the predicted background and signal compared to data. The dashed red line shows the $t\bar{t}H$ prediction normalised to the total background prediction. The hashed area represents the total uncertainty on the prediction. The ratio of data to the total prediction is shown below.

However, the differences are well covered by the total uncertainties (hashed area) in every bin of the distribution. The $t\bar{t}H$ signal (filled red) contributes only marginally in the ($\geq 6j, 3b$) region but is clearly visible on top of the background in the ($\geq 6j, \geq 4b$) region. Two statistical tests, the χ^2 [215] and Kolmogorov-Smirnov (KS) [216, 217] test, are performed to compare the agreement between data and predictions. The probabilities of the tests, representing the goodness of the agreement of the distributions with each other, are shown in the plots. In the ($\geq 6j, 3b$) region, the probabilities are very low, corresponding to a poor agreement, since the prediction systematically underestimates the data in the regions with the highest statistics. In the ($\geq 6j, \geq 4b$) region, the high probabilities confirm a good agreement. It is obtained due to lower statistics and thus larger statistical uncertainties used for the KS probability calculation. Because of the large systematic uncertainties on the shape and the normalisation of the predictions, no strong statement about the agreement can be made and the expected signal is completely hidden. Accordingly, the predictions need to be improved and the uncertainties decreased before a statement about the modelling of the $t\bar{t}H$ likelihood response can be made.

6.5.2. Likelihood Ratio

From the just discussed likelihood functions the likelihood ratio $D1$ (see Eq. 5.52), introduced in Section 5.9, can be built. If the likelihood is calculated for several background processes, the likelihood of each process needs to be properly normalised according to its fractional contribution. However, when testing only one background contribution, the choice of the relative normalisation factor α between the signal and background likelihood is arbitrary. Assuming infinite statistics and infinitesimally small bin sizes, the separation between the signal and background distributions should be the same for any choice of α . In a real analysis this is not the case and the normalisation factor can be chosen to optimise the visual separation. The information contained in the distributions, however, does not depend on α . First, the likelihood ratio $D1$ in the ($\geq 6j, \geq 4b$) region is shown in Figure 6.12 with $\alpha = \sigma_{t\bar{t}H}^{\text{obs}} / \sigma_{t\bar{t} + b\bar{b}}^{\text{obs}}$, where σ^{obs} are the observed cross sections taking selection efficiencies and acceptance into account.

The distributions are shown for the best assignment permutation $D1^{\text{max}}$ (a), using the highest likelihood value in the signal as well as in the background hypothesis, and summed over all permutations $D1^{\text{sum}}$ (b). Both distributions are normalised to unity and show remarkable separation between the total background (solid blue) and the $t\bar{t}H$ signal (dashed red). The separation is slightly better for the summed likelihood yielding around 4.4% compared to 4.1% for the best permutation. This is due to the fact, that additional topological information of the event is evaluated by testing the other assignments. Although the most background-like bin close to zero is dominated by background events and the most signal-like bin close to one is dominated by $t\bar{t}H$ events, the background distributions looks on average more signal-like. Approximately 44% of the background events populate the most signal-like bin, in which also more than 65% of the $t\bar{t}H$ events are present, which causes a significant loss in separation. However, if $\alpha = 0.23$ is chosen, a value optimised for the best separation given the bin size, the signal and the background events migrate to more background-like values. The $D1^{\text{max}}$ and $D1^{\text{sum}}$ distributions for this choice of α are shown in Figure 6.13.

The separation of 5.4% in the $D1^{\text{max}}$ (a) and of 6.6% in the $D1^{\text{sum}}$ (b) distributions has improved significantly compared to the cross section normalisation presented in Figure 6.12. By changing the normalisation, the distribution in (a) and (b) look now significantly different, revealing the structure, which was previously hidden in the most signal-like bin. As a consequence of the additional structure, the $D1^{\text{sum}}$ distribution discriminates much better signal

6. Search for $t\bar{t}(H \rightarrow b\bar{b})$ Production in the Single Lepton Channel

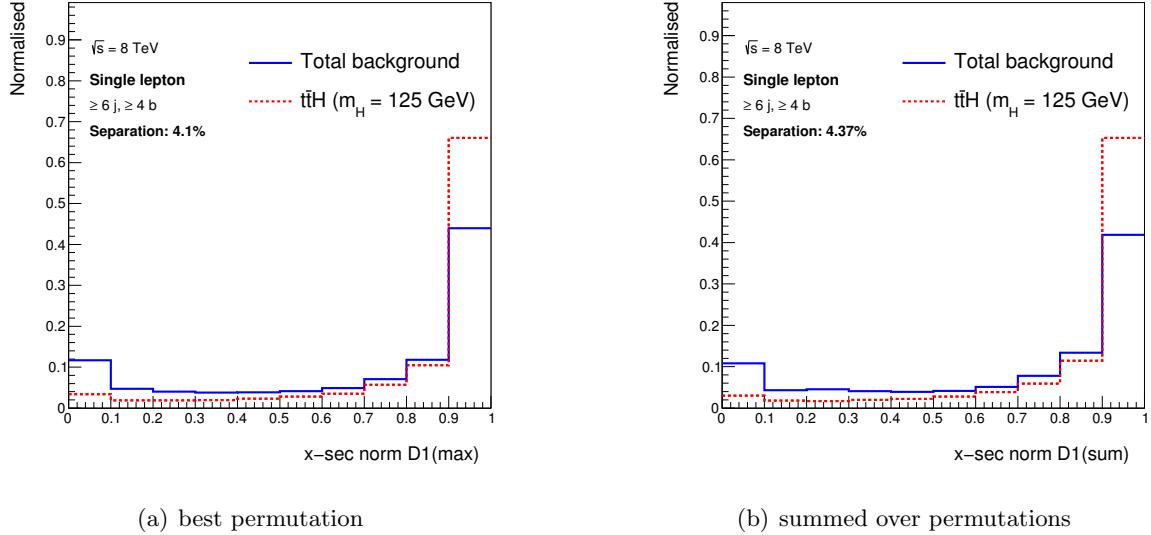


Figure 6.12.: The likelihood ratio $D1$ with cross section normalisation (a) in the best and (b) summed over all possible assignment permutations in the ($\geq 6j, \geq 4b$) region for the total background (solid blue) and the $t\bar{t}H$ signal (dashed red) predictions.

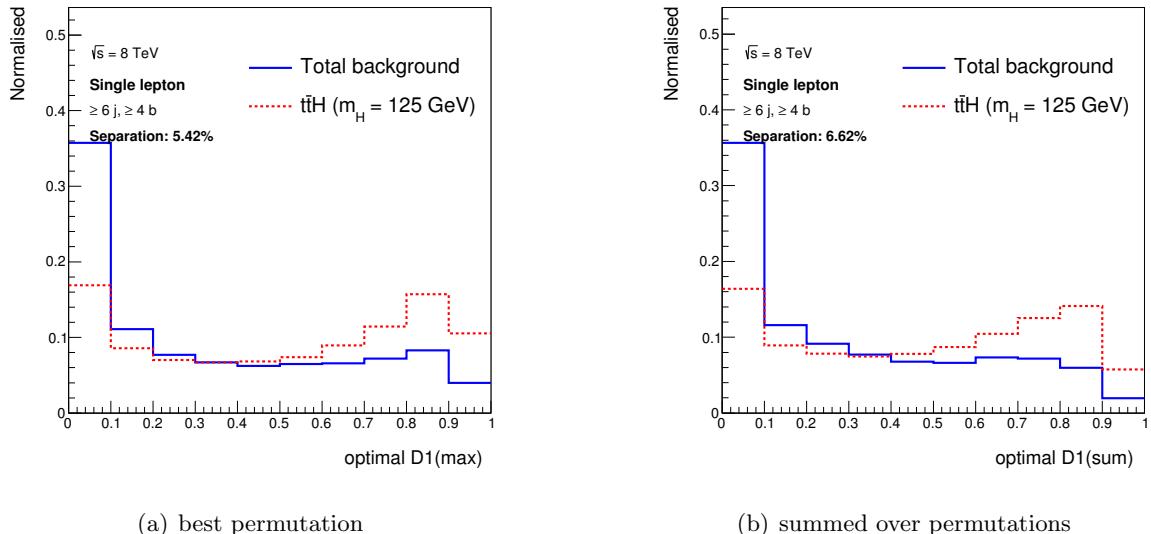


Figure 6.13.: The likelihood ratio $D1$ with optimal normalisation (a) in the best and (b) summed over all possible assignment permutations in the ($\geq 6j, \geq 4b$) region for the total background (solid blue) and the $t\bar{t}H$ signal (dashed red) predictions.

from background. In both distributions the events migrate from the signal-like values to more background-like values. As a result, more than 35% of the background and approximately 18% of the signal events populate the bin closest to zero. Instead of having most of the signal events in the bin closest to one, they distribute now over several signal-like bins, which substructure allows to distinguish significantly better these events. This choice of α will be used

in the following $D1$ distributions. In Figure 6.14 the normalised $D1$ distributions for the total background and signal are shown in the ($\geq 6j, 3b$) region.

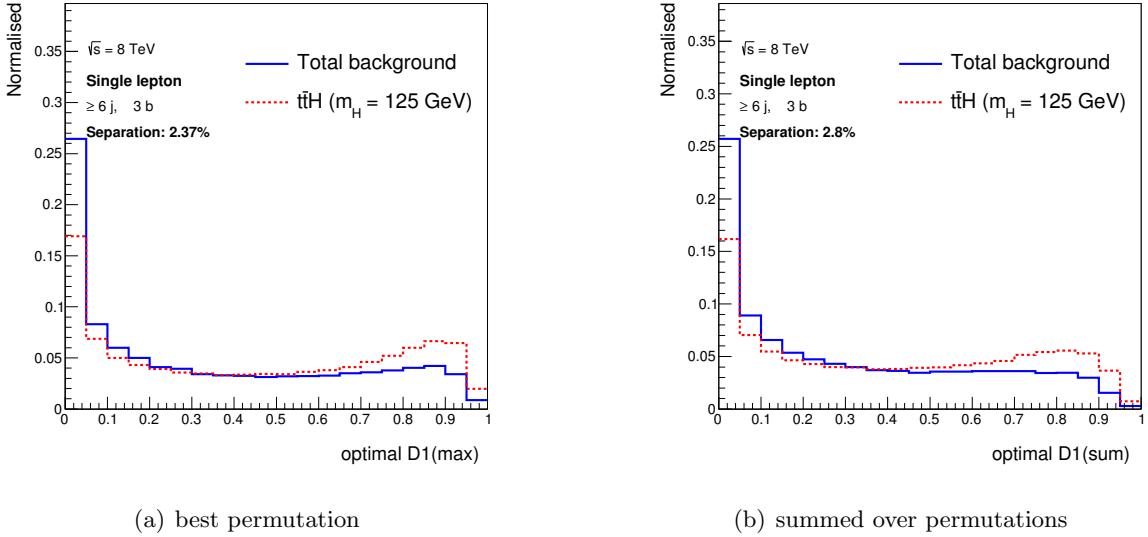


Figure 6.14.: The likelihood ratio $D1$ (a) in the best and (b) summed over all possible assignment permutations in the ($\geq 6j, 3b$) region for the total background (solid blue) and the $t\bar{t}H$ signal (dashed red) predictions.

Due to the larger statistics in this region, a finer binning is chosen. Nevertheless, the increase of assignment permutations and partially also the larger fraction of $t\bar{t} + \text{light}$ events significantly reduce the discrimination power. The separation is only approximately 2.4% for the $D1^{\max}$ (a) and 2.8% for the $D1^{\text{sum}}$ (b) distribution. Because of the generally better separation, the $D1^{\text{sum}}$ variable is preferred over the $D1^{\max}$ variable in the following, but will be denoted as $D1$ for simplicity. Comparisons of the $D1$ distributions between $t\bar{t} + \text{light}$ (filled blue) and $t\bar{t} + \text{HF}$ (shaded pink) events in the ($\geq 6j, 3b$) and ($\geq 6j, \geq 4b$) regions in Figure 6.15, reveals that $t\bar{t} + \text{light}$ events tends to have much more signal-like values.

With the largely increased fraction of $t\bar{t} + \text{light}$ events in the ($\geq 6j, 3b$) region, this difference in response degrades the separation between signal and the total background. Although the likelihood distributions of Figure 6.10 show little differences in shape between $t\bar{t} + \text{light}$ and $t\bar{t} + \text{HF}$, kinematic differences between the two processes are enhanced by building the likelihood ratio. From these studies and the comparisons with other processes presented in Appendix A.1, it is clear that the $D1$ variable is optimised for discrimination of $t\bar{t}H$ and $t\bar{t} + b\bar{b}$ events, which are the only hypotheses tested in the MEM calculation.

In order to avoid populations near zero and one, the likelihood ratio can also be constructed by dividing the signal likelihood by the background likelihood. The logarithm of this likelihood ratio $D2 = \log(\mathcal{L}_{t\bar{t}H}^{\text{sum}} / \mathcal{L}_{t\bar{t} + b\bar{b}}^{\text{sum}})$ is presented in Figure 6.16 for the ($\geq 6j, 3b$) (a) and ($\geq 6j, \geq 4b$) (b) regions comparing the total background and signal prediction normalised to unity.

The logarithm function reveals more substructure, providing a better separation between signal and background in both regions. Although the visible separation reflected in the calculated separation is significantly improved, no additional information compared to the $D1$ variable is contained in these distributions. As the $D1$ variable is easier to understand and the used NN

6. Search for $t\bar{t}(H \rightarrow b\bar{b})$ Production in the Single Lepton Channel

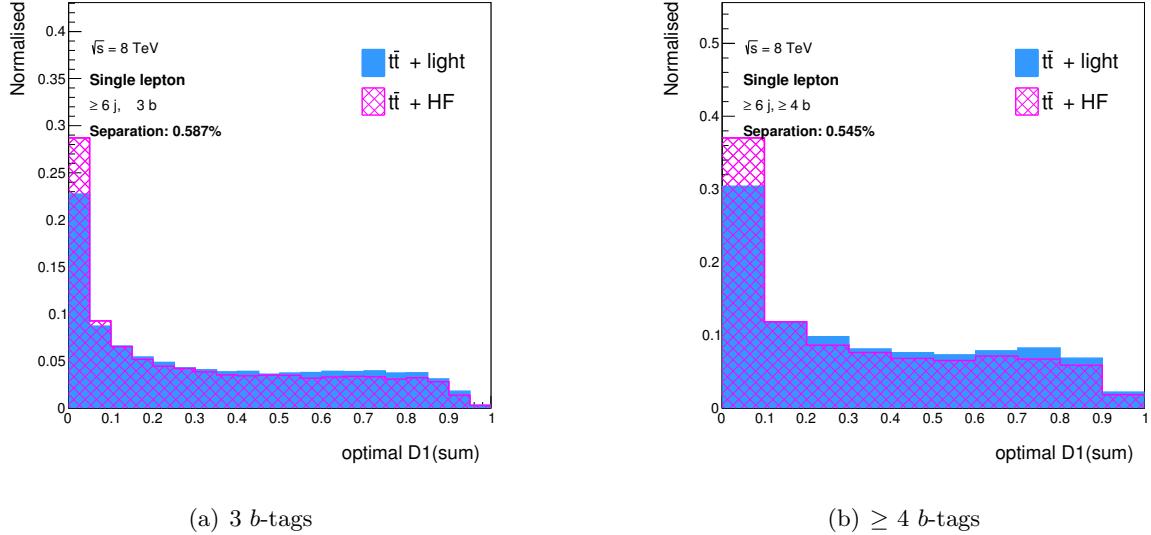


Figure 6.15.: The likelihood ratio $D1$ using the summed likelihoods in the (a) ($\geq 6j, 3b$) and (b) ($\geq 6j, \geq 4b$) regions for $t\bar{t} + \text{light}$ (filled blue) and $t\bar{t} + \text{HF}$ (shaded pink) events.

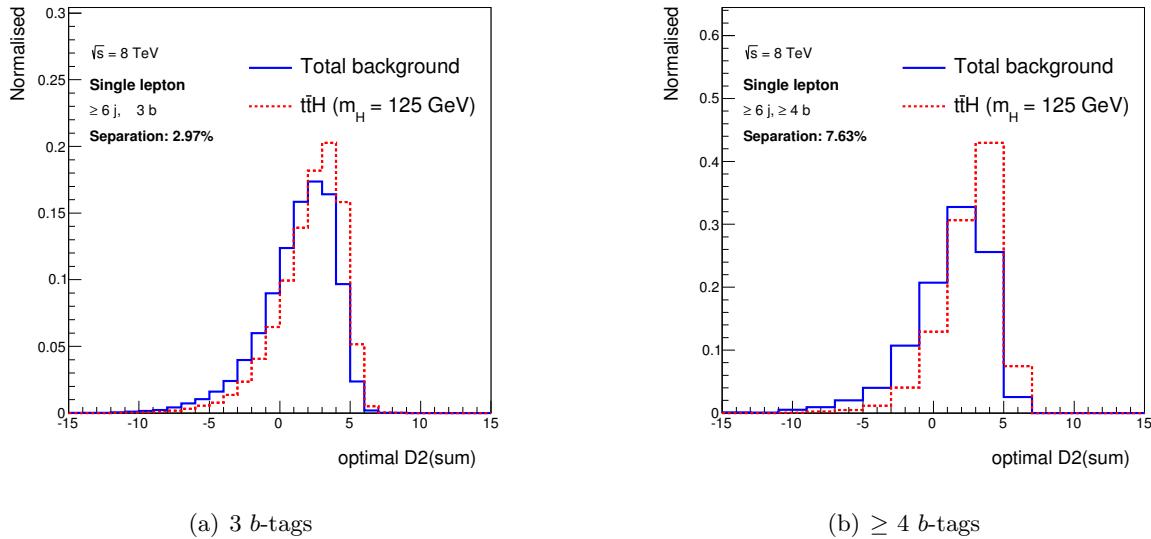


Figure 6.16.: The likelihood ratio $D2$ using the summed likelihoods is shown in the (a) ($\geq 6j, 3b$) and (b) ($\geq 6j, \geq 4b$) regions for the total background (solid blue) and the $t\bar{t}H$ signal (dashed red) predictions.

program transforms the input variables to a similar shape, it was decided to use $D1$ in the NN. The $D1$ distribution for background and signal compared to data in the ($\geq 6j, 3b$) (a) and ($\geq 6j, \geq 4b$) regions (b) are shown in Figure 6.17.

As for the likelihood distributions, the data is underestimated by the predictions in both regions, but both distributions agree well within the total uncertainty. No unexpected excess in

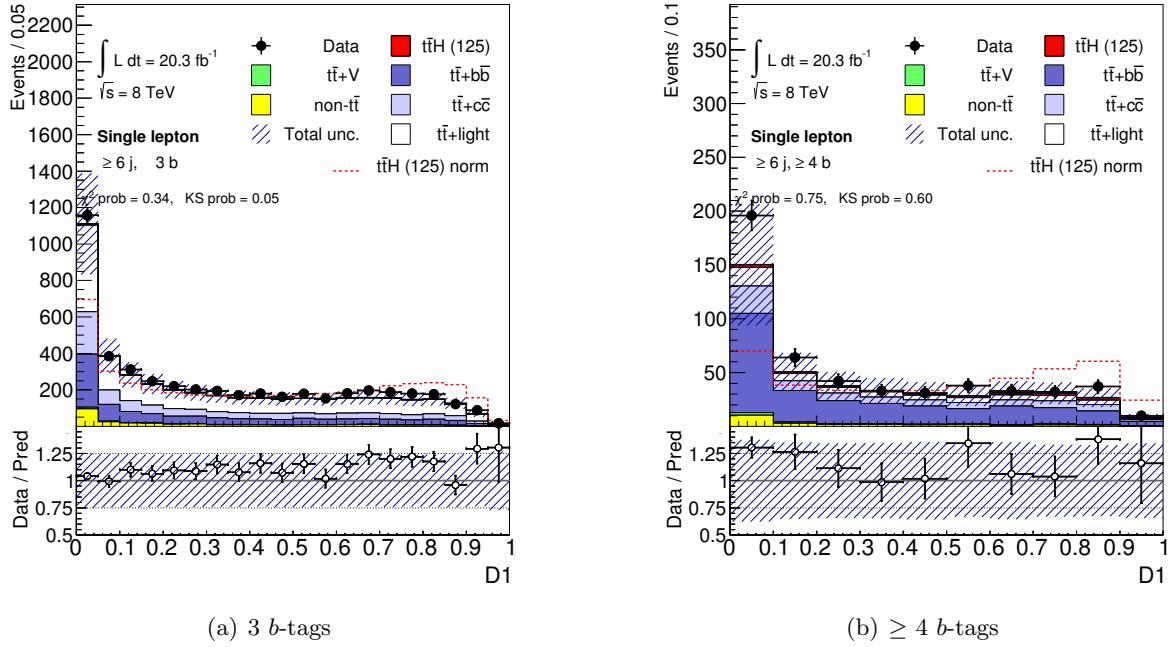


Figure 6.17.: The likelihood ratio $D1$ using the summed likelihoods is shown in the (a) ($\geq 6j, 3b$) and (b) ($\geq 6j, \geq 4b$) regions for the predicted background and signal compared to data. The dashed red line shows the $t\bar{t}H$ prediction normalised to the total background prediction. The hashed area represents the total uncertainty on the prediction. The ratio of data to the total prediction is shown below.

data is visible in the most signal-like regions of the distribution. In the ($\geq 6j, 3b$) region the statistical tests return relatively low probabilities, since almost all data points are above the predictions. In the ($\geq 6j, \geq 4b$) region, where significantly smaller statistics leads to larger statistical uncertainties, the tests confirm reasonably good agreement.

6.5.3. Event Reconstruction

A full event reconstruction can be performed using the MEM output by choosing the jet to parton assignment with the maximal likelihood value. Since two different hypotheses are tested, there are two choices for the reconstruction, defined by the assignment permutations maximising either the $t\bar{t} + b\bar{b}$ or the $t\bar{t}H$ likelihood of a given event. In the following, the permutation maximizing the signal likelihood is called *best $t\bar{t}H$ permutation* and the one maximizing the background likelihood *best $t\bar{t} + b\bar{b}$ permutation*.

W-boson mass

The W boson mass can be reconstructed by calculating the invariant mass of the two light jets, which are assigned to the light quarks of the hadronically decaying W boson. However, assignments of b -tagged jets to light quarks are not allowed as described in Section 5.7.2. Hence the reconstruction of the W boson in the ($\geq 6j, \geq 4b$) region is defined by the described selection of two non-tagged jets. In the ($\geq 6j, 3b$) region on the other hand, three jets are permuted to reconstruct the W boson mass, thus the best choice is defined by the likelihood value. The reconstructed W boson mass for the total background (solid blue) and signal

6. Search for $t\bar{t}(H \rightarrow b\bar{b})$ Production in the Single Lepton Channel

(dashed red) predictions for both regions is presented in Figure 6.18.

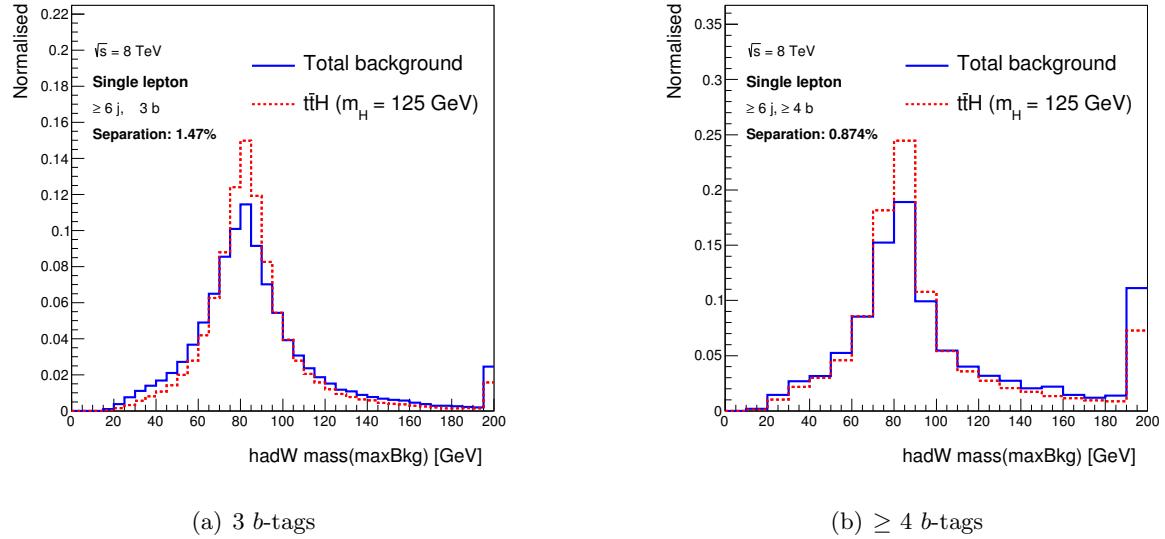


Figure 6.18.: The reconstructed W boson mass using the $t\bar{t} + b\bar{b}$ likelihoods is shown in the (a) ($\geq 6j, 3b$) and (b) ($\geq 6j, \geq 4b$) regions for the total background (solid blue) and the $t\bar{t}H$ signal (dashed red) predictions.

The mass reconstruction is very similar for both hypotheses, since the assignment of the jets to the light quarks of the W boson depends little on the additional radiation of a gluon or a Higgs boson. Thus only the reconstruction in the best $t\bar{t} + b\bar{b}$ permutation is shown. In both regions a very nice mass spectrum is obtained, which peaks at the W boson mass of 80.4 GeV for both processes. In the ($\geq 6j, 3b$) region, the peak is slightly more pronounced, since in total three different di-jet combinations are tested to fit the W boson mass. In addition, the distribution obtained with $t\bar{t}H$ events is slightly sharper. This is partially explained by the higher energy required for the creation of the $t\bar{t}H$ system and subsequently less additional radiation and initial momentum. The distributions show a separation between signal and background of 1.48% in the ($\geq 6j, 3b$) (a) region and of 0.87% in the ($\geq 6j, \geq 4b$) region. This is one of the few distributions, where the separation in the ($\geq 6j, 3b$) region is actually better than in the ($\geq 6j, \geq 4b$) region. The cause for that can be understood from Figure 6.19, where the same distributions are shown for the $t\bar{t} + \text{light}$ and $t\bar{t} + \text{HF}$ events normalised to unity.

The $t\bar{t} + \text{HF}$ events, similar to the $t\bar{t}H$ events, show a significantly better mass reconstruction than $t\bar{t} + \text{light}$ events. When requiring three or four b -tags, at least one of the jets in $t\bar{t} + \text{light}$ events has to be mistagged and it is typically the jet originating from a c -quark from the hadronically decaying W boson. If this jet is b -tagged, it will not be assigned to a light quark, and the W boson mass will not be correctly reconstructed. Although this happens more often in the ($\geq 6j, \geq 4b$) region (b), which is reflected in the larger separation between $t\bar{t} + \text{HF}$ and $t\bar{t} + \text{light}$ events, this affects substantially more the ($\geq 6j, 3b$) region due to larger fraction of $t\bar{t} + \text{light}$ events. The W boson mass distribution for the total prediction is compared to data in the ($\geq 6j, 3b$) (a) and ($\geq 6j, \geq 4b$) (b) regions in Figure 6.20

In both regions the data agrees well with the predictions within the uncertainties, which is confirmed by the statistical tests.

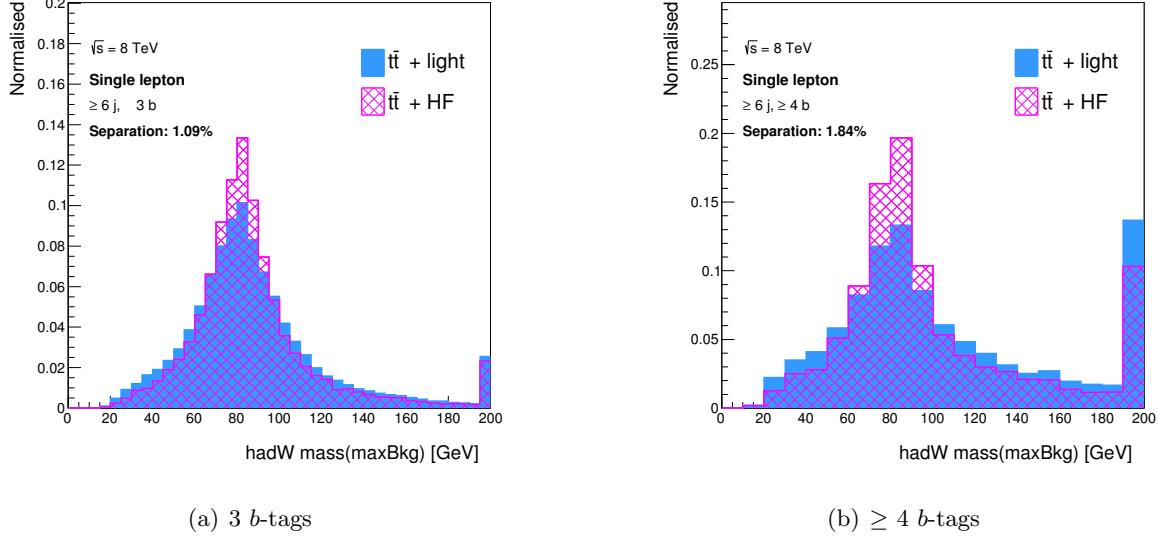


Figure 6.19.: The reconstructed W boson mass using the $t\bar{t} + b\bar{b}$ likelihoods is shown in the (a) ($\geq 6j, 3b$) and (b) ($\geq 6j, \geq 4b$) regions for $t\bar{t} + \text{light}$ (filled blue) and $t\bar{t} + \text{HF}$ (shaded pink) events.

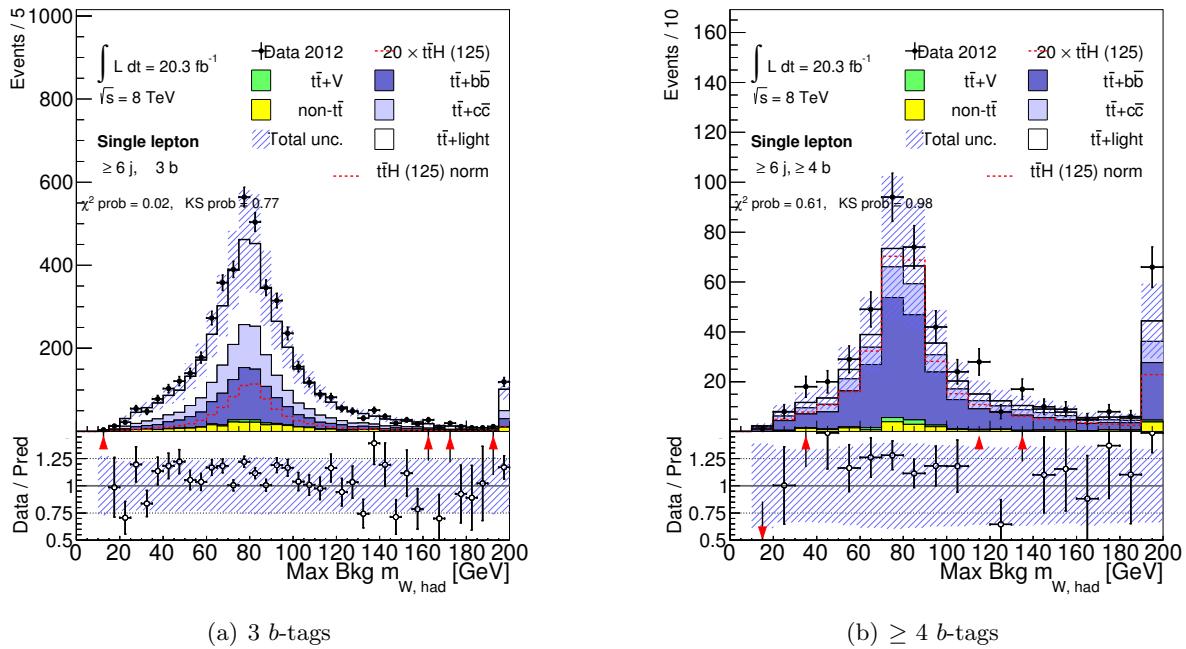


Figure 6.20.: The reconstructed W boson mass using the $t\bar{t} + b\bar{b}$ likelihoods in the (a) ($\geq 6j, 3b$) and (b) ($\geq 6j, \geq 4b$) regions for the predicted background and signal compared to data. The dashed red line shows the $t\bar{t}H$ prediction normalised to the total background prediction. The hashed area represents the total uncertainty on the prediction. The ratio of data to the total prediction is shown below.

6. Search for $t\bar{t}(H \rightarrow b\bar{b})$ Production in the Single Lepton Channel

Top quark mass

By combining the two jets assigned to the W boson with one additional b -jet the hadronically decaying top quark can be reconstructed. Because b -tagged jets can be either assigned to one of the top quarks or the gluon or Higgs boson, the reconstructed top quark mass is sensitive to the chosen hypothesis. The hadronic top quark mass is shown for the best $t\bar{t}H$ permutation (a) and best $t\bar{t} + b\bar{b}$ permutation (b) in Figure 6.21 in the ($\geq 6j, \geq 4b$) region for the total background (solid blue) and $t\bar{t}H$ signal (dashed red) normalised to unity.

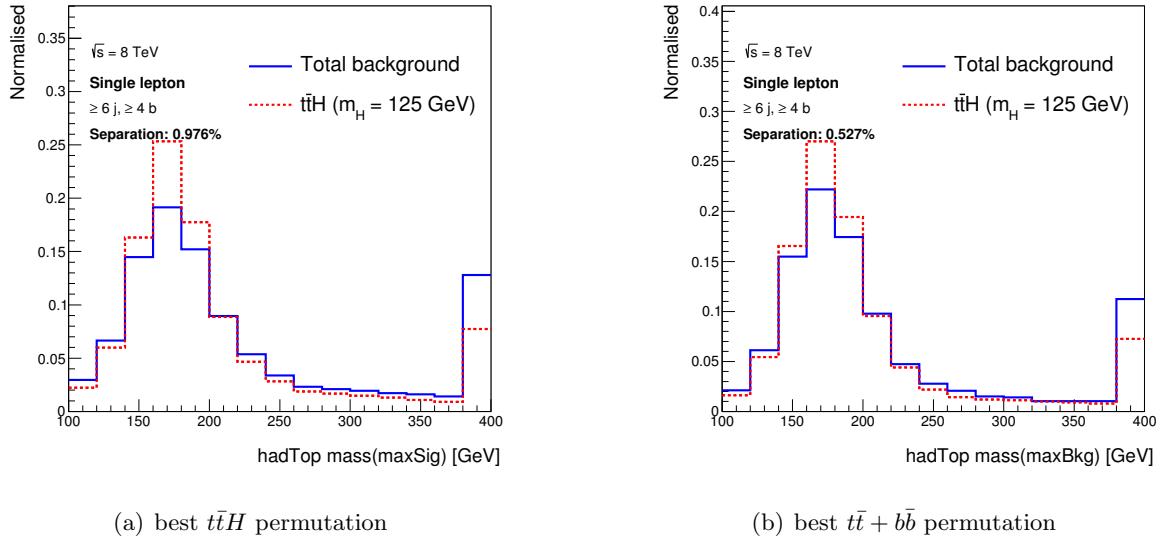


Figure 6.21.: The reconstructed hadronically decaying top quark mass in (a) best $t\bar{t}H$ permutation and (b) best $t\bar{t} + b\bar{b}$ permutation in the ($\geq 6j, \geq 4b$) region for the total background (solid blue) and the $t\bar{t}H$ signal (dashed red) predictions.

In both hypotheses, a nice top quark mass distribution peaking at values close to the used pole mass of 172.5 GeV is obtained, despite the large number of possible assignments and thus large combinatorial background. Similar to the W boson mass reconstruction, the $t\bar{t}H$ events yield a better mass reconstruction, partially for the same reasons as discussed before. However, differences are arising from the additional gluon or Higgs boson radiation. If background events are tested with the $t\bar{t}H$ hypothesis, the large contribution from the very sharp Higgs boson mass divergence to the likelihood favours a di-jet system with an invariant mass close to the Higgs boson mass as product of the decaying Higgs boson. As a consequence, less importance is given to the correct top quark reconstruction. This is reflected in a worse top quark mass reconstruction in background events using the best $t\bar{t}H$ permutation (a) instead of the best $t\bar{t} + b\bar{b}$ permutation (b), which is confirmed in the reconstruction efficiency studies shown in Appendix B.2. The separation between signal and background of 0.98% is better for the signal hypothesis compared to 0.53% for the background hypothesis. Although the differences in the top quark reconstruction efficiencies in $t\bar{t}H$ events are marginal whether the best $t\bar{t}H$ permutation or the best $t\bar{t} + b\bar{b}$ permutation is used, the mass peak is slightly more pronounced with the best $t\bar{t} + b\bar{b}$ permutation. The reason for that is the large fraction of events, in which the actual two b -jets originating from the Higgs boson decay are not present among the selected jets. The reconstructed top quark masses obtained in the ($\geq 6j, 3b$) region

are shown in Figure 6.22 for the total background prediction and the signal normalised to unity.

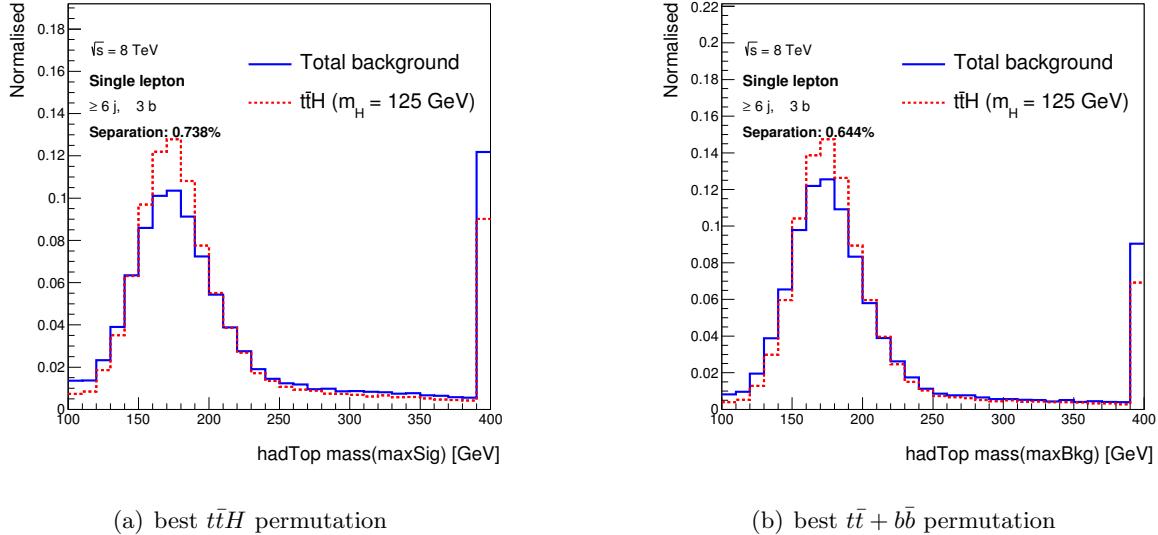


Figure 6.22.: The reconstructed hadronically decaying top quark mass in (a) best $t\bar{t}H$ permutation and (b) best $t\bar{t} + b\bar{b}$ permutation in the ($\geq 6j, 3b$) region for the total background (solid blue) and the $t\bar{t}H$ signal (dashed red) predictions.

Compared to the ($\geq 6j, \geq 4b$) region, a very similar top quark mass spectrum is obtained, despite the increased combinatorial freedom. However, the latter leads to a slightly better reconstruction of the background events and hence less difference between the background and signal when using the best $t\bar{t}H$ permutation. For the same reasons as discussed above, the mass is better reconstructed using the best $t\bar{t} + b\bar{b}$ permutation for both processes. In $t\bar{t} +$ light events, two light jets originate from a gluon, though the assignment of light jets to the gluon is only allowed in the ($\geq 6j, 3b$) region. Since the $t\bar{t} +$ light fraction is quite high in this region, this leads additionally to an improved reconstruction of background events using the best $t\bar{t} + b\bar{b}$ permutation. The differences in the top quark mass reconstruction between $t\bar{t} +$ light and $t\bar{t} +$ HF events shown in Figure 6.23 for both analysis regions are much smaller than for the W boson mass reconstruction.

The correct three-jet system corresponding to the hadronically decaying top quark can still be built, even though the wrong jets are assigned to the W boson. This leads to a generally better reconstruction of the top quark and less differences between $t\bar{t} +$ light and $t\bar{t} +$ HF events. In Figure 6.24 the data in the ($\geq 6j, 3b$) (a) and ($\geq 6j, \geq 4b$) regions (b) is compared to the background and signal predictions.

Within the large systematic uncertainties, the data agrees well with the predictions and the statistical tests result in high probabilities for the ($\geq 6j, \geq 4b$) region. However, due a systematic underestimation of the data in the high statistics bins of the top quark mass distribution, the probabilities of the tests are low in the ($\geq 6j, 3b$) region.

6. Search for $t\bar{t}(H \rightarrow b\bar{b})$ Production in the Single Lepton Channel

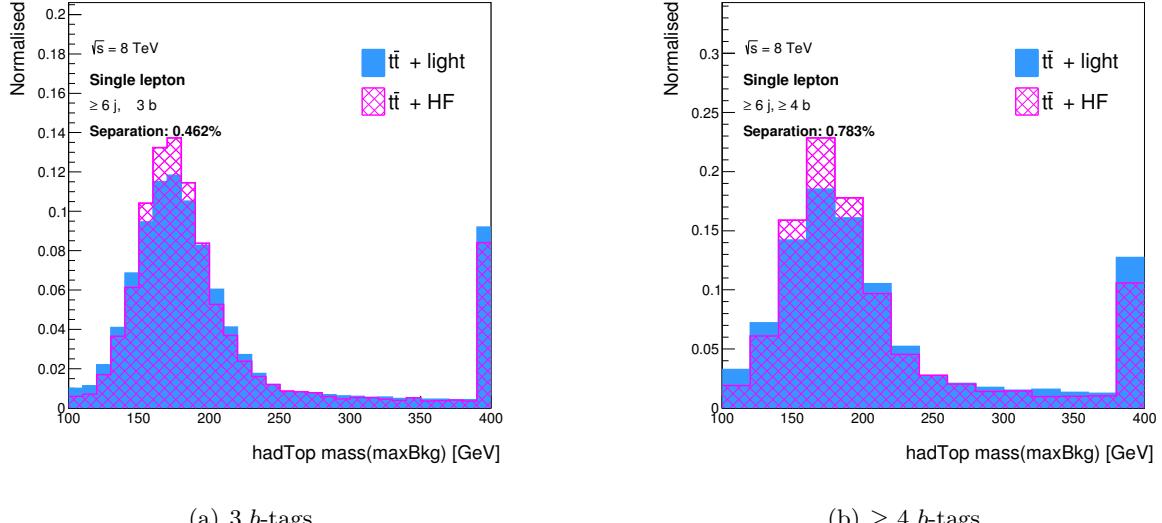


Figure 6.23.: The reconstructed hadronically decaying top quark mass in (a) best $t\bar{t}H$ permutation and (b) best $t\bar{t} + b\bar{b}$ permutation in the ($\geq 6j, 3b$) region for $t\bar{t} + \text{light}$ (filled blue) and $t\bar{t} + \text{HF}$ (shaded pink) events.

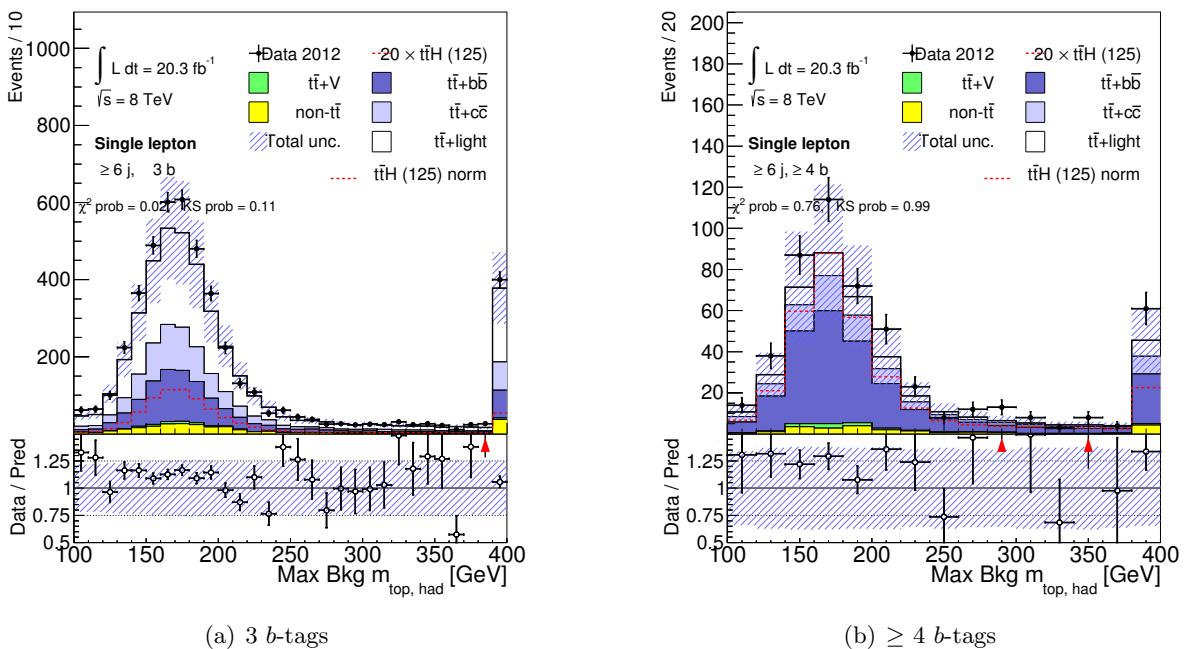


Figure 6.24.: The reconstructed hadronically decaying top quark mass using the $t\bar{t} + b\bar{b}$ likelihoods in the (a) ($\geq 6j, 3b$) and (b) ($\geq 6j, \geq 4b$) regions for the predicted background and signal compared to data. The dashed red line shows the $t\bar{t}H$ prediction normalised to the total background prediction. The hashed area represents the total uncertainty on the prediction. The ratio of data to the total prediction is shown below.

Properties of additional b-jets

The di-jet mass of the two jets, not used in the $t\bar{t}$ reconstruction is of particular interest for the $t\bar{t}H$ event reconstruction. In the background model it corresponds to the invariant mass of the gluon decaying into two b -jets, while in the signal model it is the mass of the Higgs boson radiated off one of the top quarks and decaying into two b -jets. This variable is called $m_{b\bar{b}}$ in the following. The separation power of the $m_{b\bar{b}}$ variable between the total background (solid blue) and $t\bar{t}H$ predictions normalised to unity and reconstructed in the best $t\bar{t}H$ permutation (a) and best $t\bar{t} + b\bar{b}$ permutation (b) is shown in Figure 6.25 for the ($\geq 6j, \geq 4b$) region.

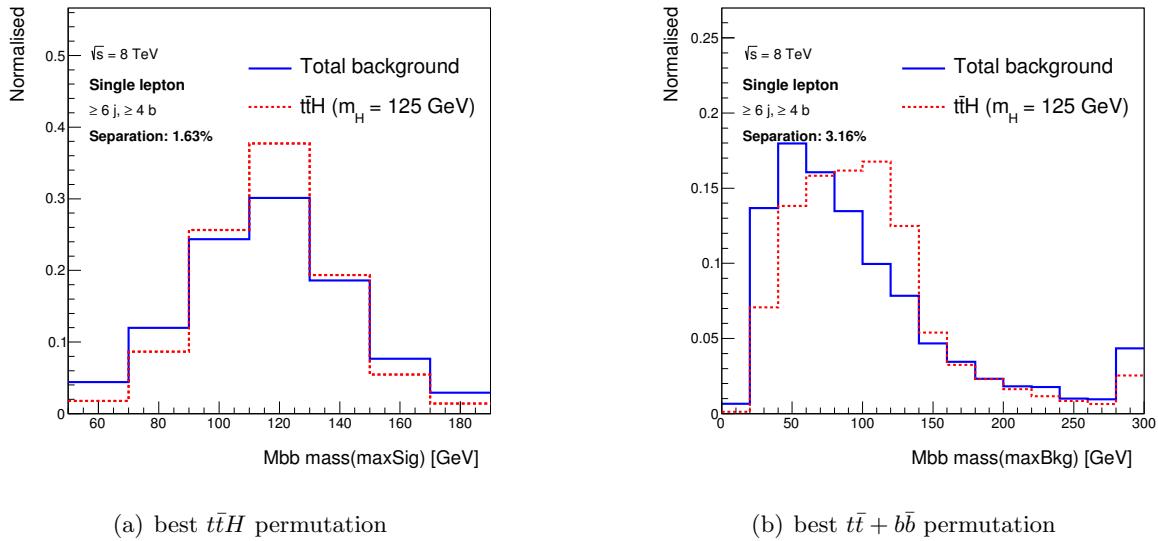


Figure 6.25.: The reconstructed invariant mass $m_{b\bar{b}}$ in (a) best $t\bar{t}H$ permutation and (b) best $t\bar{t} + b\bar{b}$ permutation in the ($\geq 6j, \geq 4b$) region for the total background (solid blue) and the $t\bar{t}H$ signal (dashed red) predictions.

When using the best $t\bar{t}H$ permutation (a), the $m_{b\bar{b}}$ distributions peaks for $t\bar{t}H$ as well as for background events at a mass around the evaluated mass of 125 GeV. The largest likelihood value is always obtained when a jet pair with an invariant mass close to the Higgs boson mass parameter is assigned to the Higgs boson, because of the large contribution from the very narrow Higgs boson divergence in the matrix element calculation. Owing to the large combinatorial freedom and the jet p_T cut, there is always a di-jet system with such property in the event regardless of the actual physics process. In $t\bar{t}H$ events, this is of course more likely, since the jets originating from the Higgs boson are expected to be present. Hence the mass peak is more pronounced, but the separation between background and signal of 1.6% is quite low. However, when maximising to the $t\bar{t} + b\bar{b}$ hypothesis (b), the $m_{b\bar{b}}$ distribution of the total background has a broad combinatorial-like shoulder, expected from pure gluon radiation and combinatorial background with the applied jet p_T cuts. In case of $t\bar{t}H$ events though, a peak around the Higgs boson mass used in the MC simulation arises on top of this combinatorial distribution. This peak is now purely caused by events in which the two jets originating from the Higgs boson decay are present and are correctly assigned to the Higgs boson in the matrix element calculation. With the much looser assumption of just having a gluon radiated, the Higgs boson mass peak can be more accurately reconstructed and measured. Thus for the

6. Search for $t\bar{t}(H \rightarrow b\bar{b})$ Production in the Single Lepton Channel

studies of the Higgs boson properties, it is advisable to use the assignment, which maximises the likelihood of the $t\bar{t} + b\bar{b}$ hypothesis, where nothing is assumed about the Higgs boson. This choice leads also to a good separation between signal and background events of around 3.2%, despite the similarities in the kinematics of the $t\bar{t} + b\bar{b}$ and $t\bar{t}H$ events. The $m_{b\bar{b}}$ distributions in the ($\geq 6j, 3b$) region are shown in Figure 6.26.

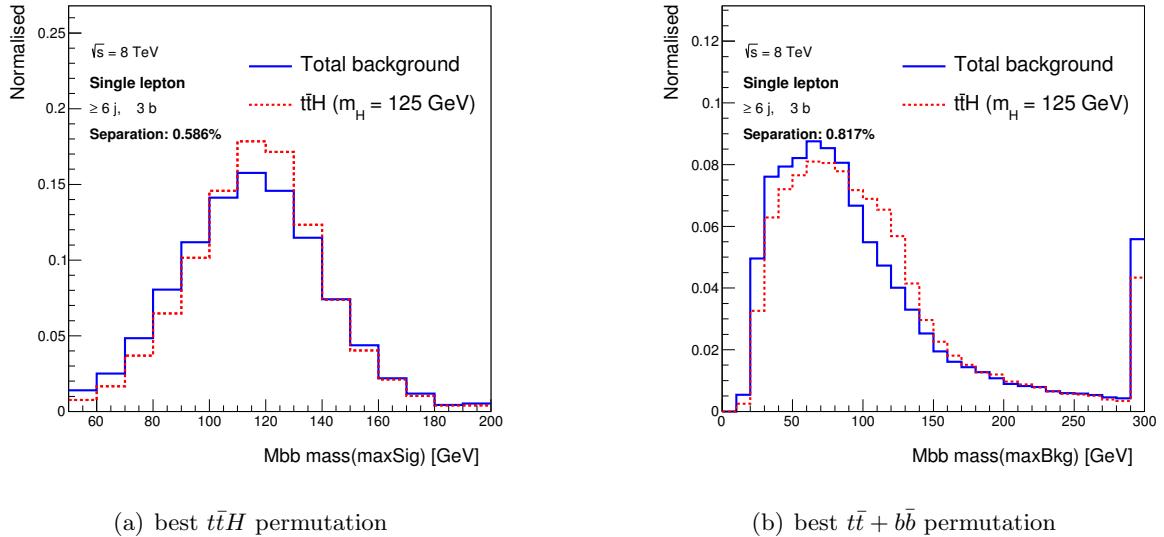


Figure 6.26.: The reconstructed invariant mass $m_{b\bar{b}}$ in (a) best $t\bar{t}H$ permutation and (b) best $t\bar{t} + b\bar{b}$ permutation in the ($\geq 6j, 3b$) region for the total background (solid blue) and the $t\bar{t}H$ signal (dashed red) predictions.

In this region the statements made for the ($\geq 6j, \geq 4b$) region still hold. However, the separation is diluted for both hypotheses by the increase of possible jet to parton assignments. With the increased statistics and hence finer binning, it is further visible, that the peak reconstructed with the best $t\bar{t}H$ permutation is slightly biased to lower Higgs boson masses and not perfectly symmetric. This is caused by the fact that on average more jets with lower energy are available and are combined arbitrarily. When using the best $t\bar{t} + b\bar{b}$ permutation, the peak arising from the Higgs boson decay in the $m_{b\bar{b}}$ distribution of the $t\bar{t}H$ events is significantly less pronounced due to the increase of combinatorics. The $m_{b\bar{b}}$ distribution of the background events in the best $t\bar{t} + b\bar{b}$ permutation looks similar to the one in the four b -tag region, but the peak is slightly shifted to higher masses. This is caused by the increased fraction of $t\bar{t} +$ light events, whose additional jets have on average slightly higher masses. This can be observed in the $m_{b\bar{b}}$ distributions of Figure 6.27, where $t\bar{t} +$ light and $t\bar{t} +$ HF events are compared in the ($\geq 6j, 3b$) (a) and ($\geq 6j, \geq 4b$) regions (b) using the best $t\bar{t} + b\bar{b}$ permutation.

As already discussed, it is quite likely, that at least one of the additional b -tagged jets is originating from the W boson and is consequently expected to have higher energy. As a result, the peak in the $m_{b\bar{b}}$ distributions of the $t\bar{t} +$ light events is closer to the W boson mass. This leads to less distinction from the Higgs boson mass peak when comparing to the $m_{b\bar{b}}$ distribution of $t\bar{t}H$ events. In Figure 6.28 the $m_{b\bar{b}}$ distribution of the total predictions is compared to the events in the ($\geq 6j, 3b$) (a) and ($\geq 6j, \geq 4b$) regions (b) using the best $t\bar{t}H$ permutation and best $t\bar{t} + b\bar{b}$ permutation, respectively.

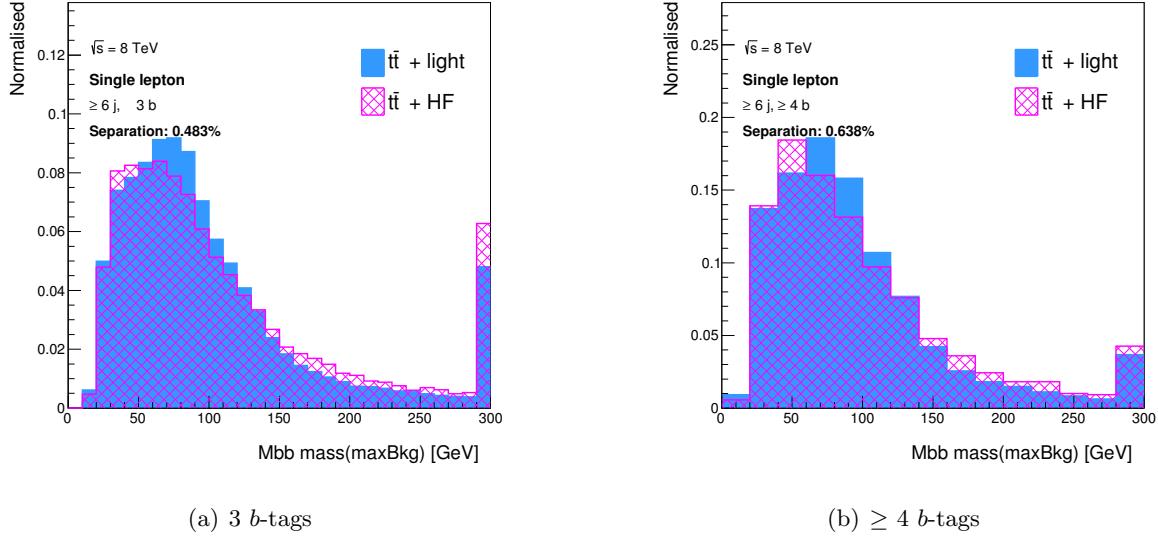


Figure 6.27.: The reconstructed invariant mass $m_{b\bar{b}}$ using the $t\bar{t} + b\bar{b}$ likelihoods is shown in the (a) ($\geq 6j, 3b$) and (b) ($\geq 6j, \geq 4b$) regions for $t\bar{t} + \text{light}$ (filled blue) and $t\bar{t} + \text{HF}$ (shaded pink) events.

Although the data is underestimated, the predictions agree well within the uncertainties with the observed data. In case of the best $t\bar{t}H$ permutation, the statistical tests confirm a good agreement in the ($\geq 6j, \geq 4b$) region, but not in the ($\geq 6j, 3b$) region. In case of the best $t\bar{t} + b\bar{b}$ permutation, the predictions agree much better with data in the ($\geq 6j, 3b$) region based on the statistical tests.

Similar to the $m_{b\bar{b}}$ reconstruction, the reconstructed angle between the two additional jets depends very strongly on the chosen hypothesis. The decaying particles of a massive Higgs boson at rest would escape in opposite directions, but at the LHC the Higgs boson can have significant momentum. Still, they are expected to have a larger opening angle than the one of the $q\bar{q}$ pair originating from the massless gluon, which is small due to the lower invariant mass and the properties of QCD. The angle in the ϕ plane of the two jets assigned to the gluon or the Higgs boson for the $t\bar{t}H$ (a) and $t\bar{t} + b\bar{b}$ (b) hypotheses is shown in Figure 6.29 for the ($\geq 6j, 3b$) region.

Since the best $t\bar{t}H$ permutation (a) picks two jets with high invariant mass, they tend to be back-to-back in the ϕ plane regardless of the process, because they likely originate from the two different top quark hemispheres. In case of a $t\bar{t}H$ event, the Higgs boson is radiated from one of the top quarks. It unbalances the $t\bar{t}$ system resulting in smaller angle. Furthermore, in some cases the correct jets are assigned to the Higgs boson. Subsequently, the $\Delta\phi$ distribution is expected to peak at medium values ($|1| \geq \Delta\phi \geq |2|$) due to the smaller mass. If the best $t\bar{t} + b\bar{b}$ permutation is used (b), the opening angle is small, which is the expected behaviour for background events. However, the distribution of the $t\bar{t}H$ events differs only marginally with a trend to have higher values. The resulting separation is negligible. In Appendix B.1 the $\Delta\phi$, $\Delta\eta$, and ΔR distributions using the best $t\bar{t}H$ permutation and best $t\bar{t} + b\bar{b}$ permutation in both b -tag regions are shown.

Since a full event reconstruction is performed, the angle between the hadronically decaying top quark and the $m_{b\bar{b}}$ system, which is supposed to be the Higgs boson in case of $t\bar{t}H$ events, can be calculated. This variable is expected to be very different for signal and background events.

6. Search for $t\bar{t}(H \rightarrow b\bar{b})$ Production in the Single Lepton Channel

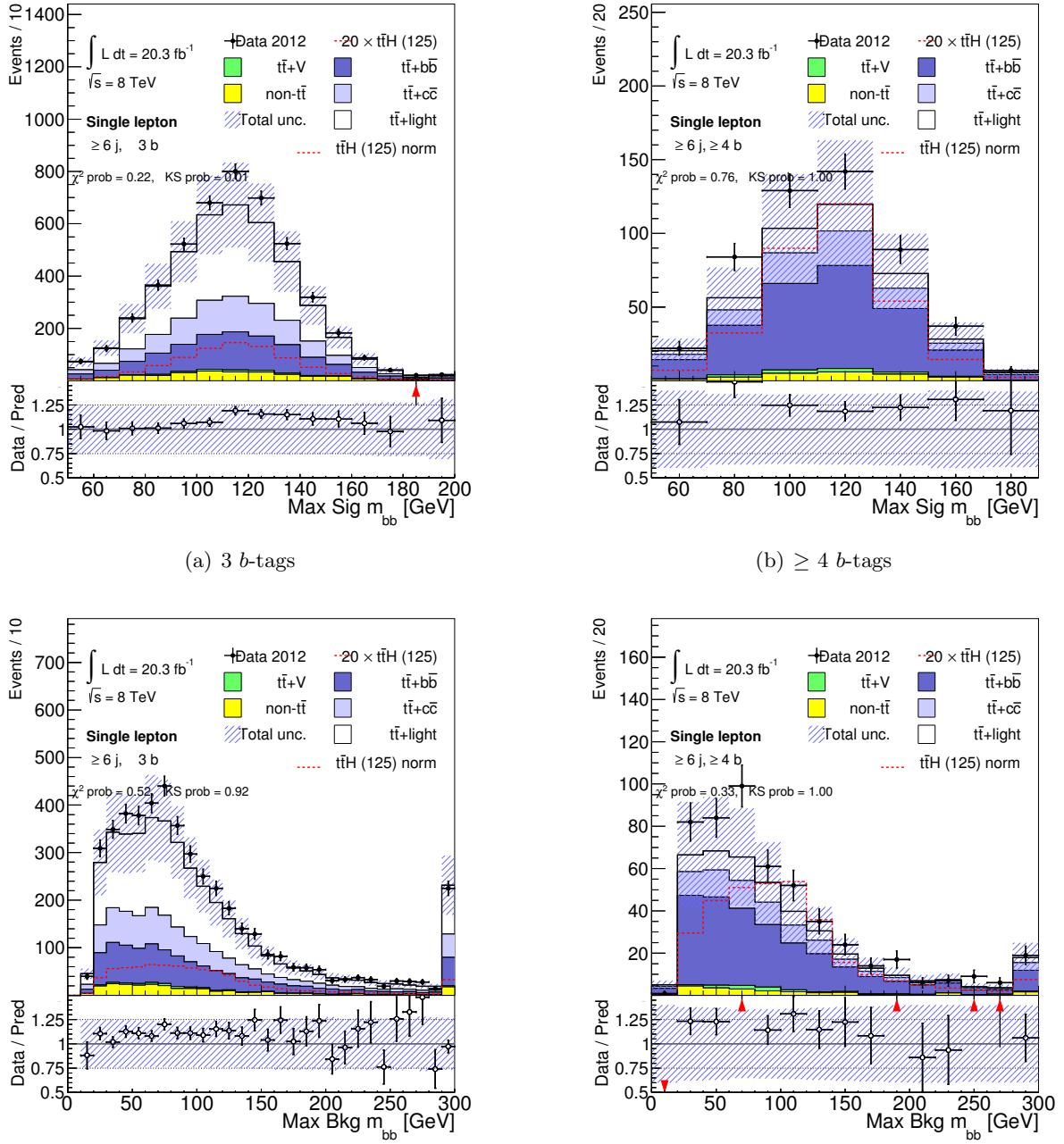


Figure 6.28.: The reconstructed invariant mass $m_{b\bar{b}}$ using the $t\bar{t}H$ likelihood (upper) and $t\bar{t} + b\bar{b}$ likelihood (lower) is shown in the (a) ($\geq 6j, 3b$) and (b) ($\geq 6j, \geq 4b$) regions for the predicted background and signal compared to data. The dashed red line shows the $t\bar{t}H$ prediction normalised to the total background prediction. The hashed area represents the total uncertainty on the prediction. The ratio of data to the total prediction is shown below.

Because the best $t\bar{t} + b\bar{b}$ permutation is more suitable for the reconstruction since it does not force the background events to look like signal, this assignment is used for the reconstruction. The $\Delta\eta$ between the hadronic top quark and the $m_{b\bar{b}}$ system is shown in Figure 6.30 for the

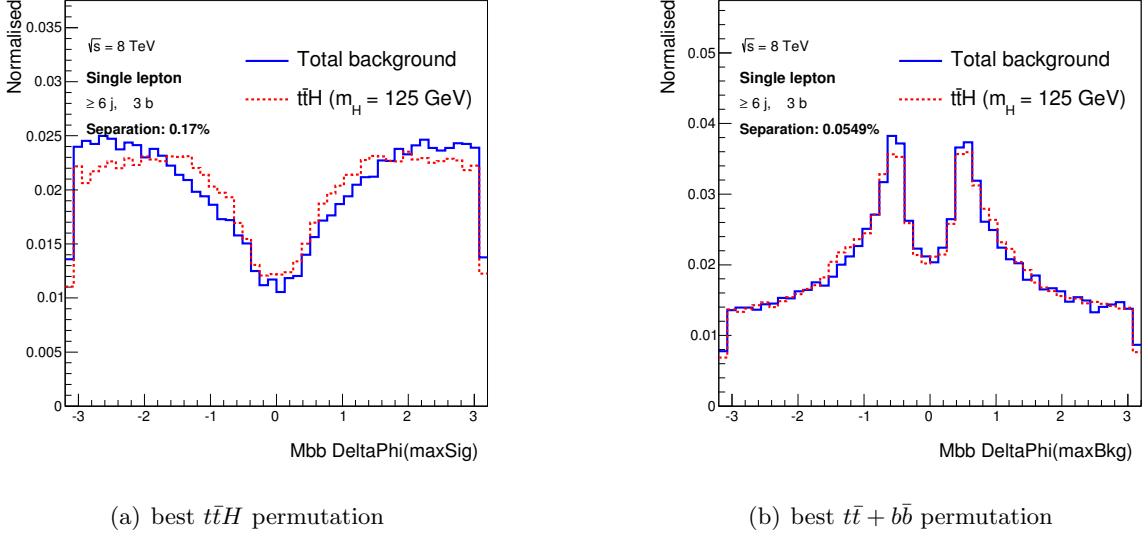


Figure 6.29.: The reconstructed $\Delta\phi$ between the two additional b -jets in (a) best $t\bar{t}H$ permutation and (b) best $t\bar{t} + b\bar{b}$ permutation in the ($\geq 6j, 3b$) region for the total background (solid blue) and the $t\bar{t}H$ signal (dashed red) predictions.

($\geq 6j, 3b$) (a) and ($\geq 6j, \geq 4b$) (b) region.

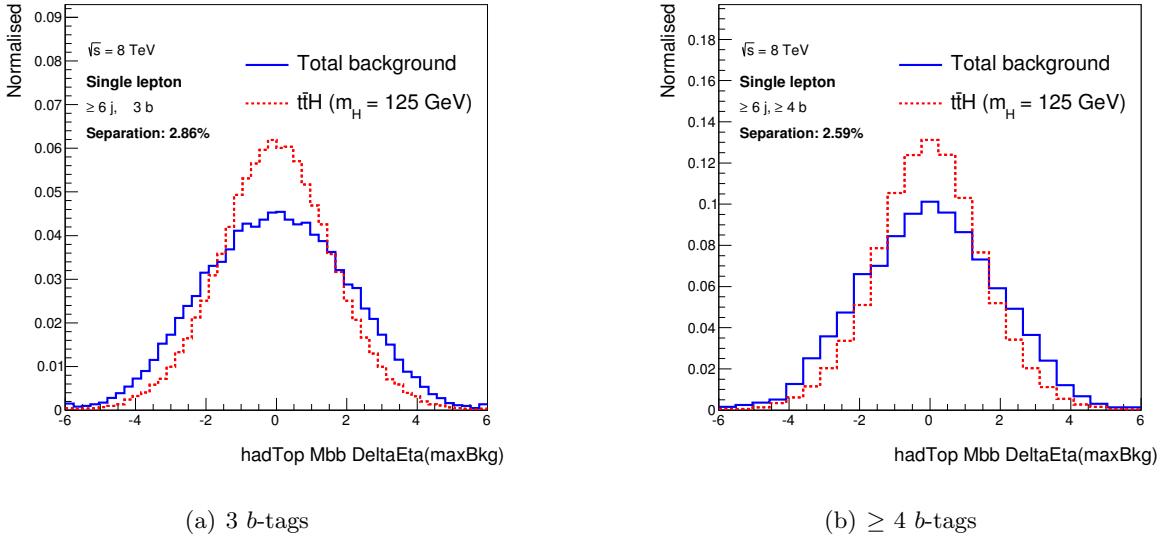


Figure 6.30.: The reconstructed $\Delta\eta$ between the hadronic top quark and the $m_{b\bar{b}}$ system in the (a) ($\geq 6j, 3b$) and (b) ($\geq 6j, \geq 4b$) regions for the total background (solid blue) and the $t\bar{t}H$ signal (dashed red) predictions.

In both regions, the distribution has a much more pronounced peak around zero for $t\bar{t}H$ events. The distributions of the background is significantly broader. This results in a very good separation of around 2.9% and 2.6% in the region with three and four b -tags, respectively. In the ($\geq 6j, 3b$) region, this separation is even slightly better than the one obtained with the

likelihood ratio $D1$.

More than sixty different variables reconstructed using the output of the MEM have been studied with the goal of obtaining well separating variables. Among them are very complex event properties, particle properties or angular distributions in the rest frame of certain intermediate states making use of the full event reconstruction. However, any further detailed discussion of these variables is beyond the scope of this thesis. The accurate reconstruction of the W boson and the top quark mass gives confidence that the method works well and provides a good event reconstruction. A small selection of further event reconstruction studies is shown in Appendix B.

6.6. The Artificial Neural Network

A *multivariate analysis (MVA)* refers to a class of statistical analysis methods, which cope simultaneously with multiple observed variables. They examine the relations between the variables and usually study their relevance for a certain problem with the aim to reduce the number of variables without losing important information. In high energy physics, MVAs are mainly used for classification of objects, events or assignment permutations by recognising patterns in the observed data. One of these techniques is the family of *artificial neural networks*, which are machine learning algorithms inspired by the central nervous system of animals. Similar to the biological neural networks, they build connections between *nodes* modelling the biological neurons, which are usually structured into a certain number of layers. During a supervised learning phase, weights are adaptively assigned to the connections between the nodes representing the importance of the connection for the problem. Finally, unknown input can be evaluated by using the weights and the structure of the network to obtain a response for the problem. Depending on the realisation of the nodes by its activation function, the design of the network and the learning algorithm, the NN is capable of learning non-linear relations between the input variables. If the output layer consist only of a single node, the information of the multiple input variables can be combined into a single variable containing ideally all of the information of the input variables.

6.6.1. The NeuroBayes Package

The NEUROBAYES package [218, 219] is a realisation of an artificial NN following a Bayesian statistics approach, such that the output can be directly interpreted as a posteriori probability. One essential advantage of NEUROBAYES is that overtraining is basically prevented by advanced regularisation and pruning techniques. By using physics knowledge as Bayesian prior probabilities, unphysical results can be excluded. Another very useful feature of NEUROBAYES is the ranking of the input variables according to their significance for a certain target, taking all correlations among the input variables into account. A target could be maximal separation between two classes of input types. The relatively simple NN is designed of only three layers, of which one is the input layer, one is a hidden layer and one is the output layer. The input layer consists of n nodes, representing the n input variables, plus one additional node providing internal information used to adjust the output of the hidden nodes. The complexity of the network can be varied by changing the number of nodes in the hidden layer. The information is then combined into one single output node. As an example of such a designed network, the actual trained NN in the ($\geq 6j, \geq 4b$) region is shown in Figure 6.31.

In this architecture of the network, nodes of one layer are only connected to the nodes of the

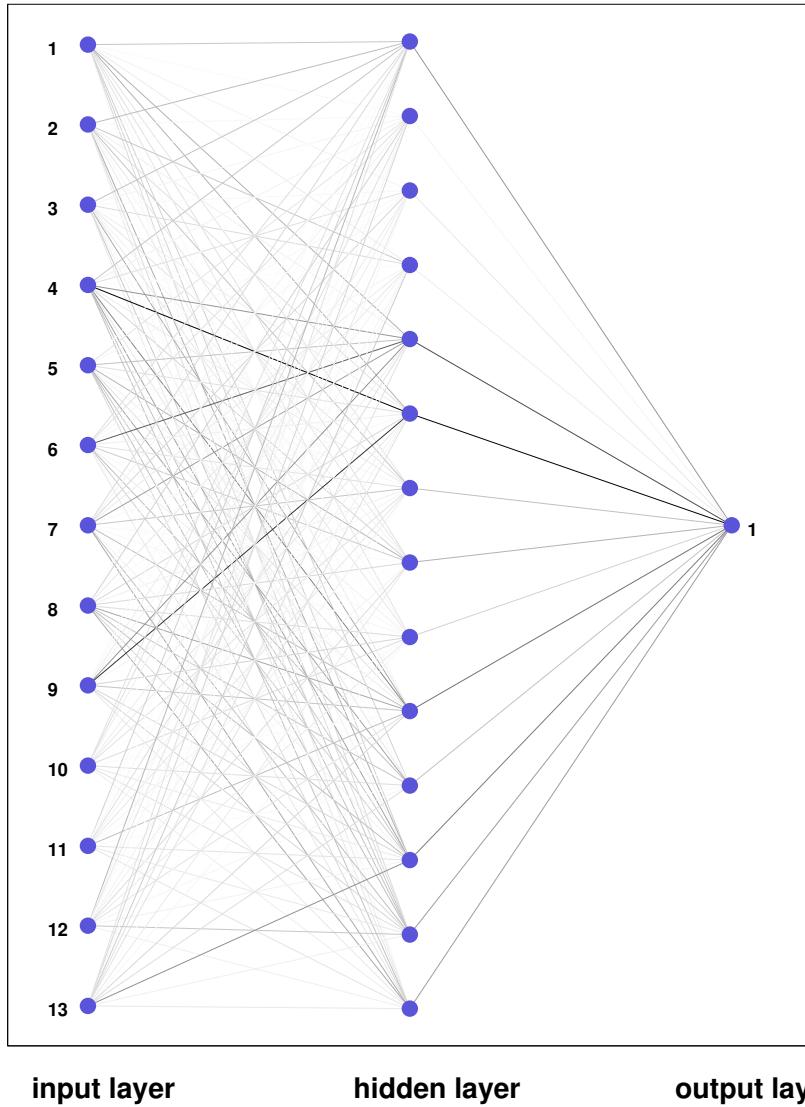


Figure 6.31: The trained artificial neural network of NEUROBAYES in the ($\geq 6j, \geq 4b$) region is shown. It consists of a three layered node system with one input layer, one hidden layer and one output layer. The input layer consists of twelve nodes representing the input variables and one additional node. They are then combined via the 13 nodes of the hidden layer to one single output node. The weight of the connection is represented by the strength (darkness of grey tone) of the drawing.

next layer, which is referred to as a *feed-forward NN*. The weight assigned to the connections during the learning phase is represented by the strength (darkness of grey tone) of the drawing. During the sophisticated preprocessing phase of NEUROBAYES, the input variables are equalised by transformation to a normalised distribution with a mean of zero. This is achieved by first transforming the distribution of arbitrary range and shape to a flat distribution by a non-linear transformation. Subsequently, the flat distribution is transformed by a Gaussian function with mean zero and a width of one. From this basis, the correlation matrix of the input variables is calculated and their total importance for the separation is determined taking the correlations between them into account. The variables are then ranked by iteratively removing one variable at a time and recalculating the significance of the separation to estimate the loss. The least important variable is removed and the procedure is repeated until only the most important variable remains. The ranking of the variables can be used to remove insignificant variables in order to keep the analysis as simple as possible. Finally, the variables are linearly decorrelated by performing a *Jacobi rotation* of the covariance matrix.

Each node in the network is modelled by an activation function $g(t)$, which is in this case the sigmoid function. The output of node j in the layer l is given by iteratively calculating

$$x_j^l = g \left(\sum_{k=0}^{N^{l-1}} w_{jk}^l \cdot x_k^{l-1} + \mu_j^l \right), \quad (6.2)$$

where w_{jk}^l are the weights of the connections to the node from the N^{l-1} nodes of the previous layer each having an output x_k^{l-1} . The μ_j^l define the activation threshold of the node given by the additional bias node of the previous layer, whose connections are not modified by a weight. The output is thus given by evaluating the activation function with the weighted sum of all nodes in the previous layers modified by the threshold. During the learning phase the weights assigned to each connection are adjusted by minimising a loss function

$$E = \sum_i^{N_{\text{evts}}} \log \left(\frac{1+T_i}{2} x_i^o \right), \quad (6.3)$$

which depends on the output of the final node x^o for the event i and the target T , which is zero for background and one for signal events. The minimisation is performed using a memory optimised *backpropagation* algorithm [220]. In order to treat outliers efficiently and avoid overtraining, a Bayesian regularisation scheme is applied [219]. After a certain number of iterations, in which each sub-sample of the total number of events is evaluated, the minimisation converges and the final weights are assigned to the connections. Events of unknown origin can now be evaluated by calculating the output of the final node using the weights of the training.

6.6.2. Discriminating Variables

Several scenarios have been studied, in which purely kinematic variables and variables obtained with the MEM are used in order to find an optimal set of input variables for the NN making use of the ranking procedure of NEUROBAYES. More than 300 kinematic variables are created, classified into object kinematics, object pair properties, global event variables and event shape variables. In the construction of jet related variables, b -jets are generally preferred before the remaining jets ranked by the transverse momentum. In the regions with ≥ 6 jets, no more than seven jets are considered in the construction in order to avoid mismodelling from higher jet multiplicities. The MEM provides more than 60 MEM variables, created either directly from the likelihood information or by performing an event reconstruction. Among these complicated variables, obtained from boosting to the rest frame of the $t\bar{t}$ system or the intermediate states making full use of the event reconstruction. The ideal scenario is to use all of the roughly 400 variables in the preprocessing phase of NEUROBAYES and let the ranking system decide which and how many variables should be used. These studies have been performed together with other possible scenarios and are shown in Appendix C.2. The discriminating variables are not only required to separate the signal from background but also to describe the data well in simulations. In particular in the ($\geq 6j, \geq 4b$) region, where good separation can be obtained using the MEM variables, this is hard to confirm in a blinded analysis with a poor modelling before the predictions are corrected by the profile likelihood fit. Using the ($\geq 6j, 3b$) region as validation region might work for variables constructed only from the likelihood information and ideally making no distinction between the assignment permutations. However,

due to the slightly different realisation of the MEM, the approximations made in the assignment permutation ranking and the different composition of the background, it is not clear without extensive studies, that all reconstructed MEM variables behave the same way in both regions. Because the kinematic NN was already well established, while there had been little experience with the reconstructed MEM variables, it was decided to use only the likelihood ratio $D1$ and the logarithm of the summed signal likelihood $\ln \mathcal{L}_{t\bar{t}H}^{\text{sum}}$ as input to the final NN. These two variables contain only event information and sum up the information of the single assignment permutations. When adding them as input to the established and highly optimised kinematic NN, the difference of the performance compared to the ideal scenario is small (see Appendix C.2). The approach described in the following is a compromise between minimizing the change of the analysis and gaining sensitivity to the signal by adding MEM variables.

In the three signal-enriched regions, more than 300 kinematic variables are used in the pre-processing of NEUROBAYES to determine the best variables for separating $t\bar{t}H$ events from the total background simulated by MC. A saturation effect in terms of separation is observed for more than ten variables and no significant improvement of discrimination is achieved by adding more (see Fig. C.1). Therefore, only the best ten variables are selected for the NN. The variables and their pairwise correlations (see Fig. C.3) are required to model the data well in all analysis regions. In the signal-enriched regions with at least six jets, the two MEM variables are added as input to the NN, the preprocessing is repeated and the training is performed. In the (5j, 3b) region, $t\bar{t} + \text{HF}$ events are defined as signal and $t\bar{t} + \text{light}$ events as background in order to find the best seven variables discriminating $t\bar{t} + \text{HF}$ from $t\bar{t} + \text{light}$ events. This is mainly achieved by exploiting the different sources of the third b -tagged jet, which is in case of the $t\bar{t} + \text{light}$ events a mistagged jet and often originates from the hadronically decaying W boson. In case of the $t\bar{t}$ and $t\bar{t}H$ events, the fast ATLAS simulations are used for the training in order to increase the available statistics, while the full simulation is used for the validation of the variables. The best variables selected by NEUROBAYES, their definition and ranking in the different analysis region, where an NN is employed, are summarised in Table 6.4.

In the most sensitive region ($\geq 6j, \geq 4b$) the MEM variable $D1$ is ranked as the most important variable of the NN with large significance advantage, since it is optimised to separate signal from the dominant $t\bar{t} + b\bar{b}$ background. The kinematic variables centrality, $p_T^{\text{jet}5}$, the second Fox-Wolfram moment $H1$ [221, 222] and the average ΔR for all b -tagged jet pairs $\Delta R_{bb}^{\text{avg}}$ follow in the ranking. Despite the relatively high correlation to $D1$ (see Fig. C.3), the $\ln \mathcal{L}_{t\bar{t}H}^{\text{sum}}$ is the sixth most important variable, because it helps to separate against the non- $t\bar{t}$ background in particular (see Fig. A.3). The separation of all variables used in the NN in the ($\geq 6j, \geq 4b$) region are shown in descending order in Figure 6.34 comparing the normalised total background (solid blue) to the signal (dashed red).

Apart from the centrality, the visible separation between signal and background for the four best discriminating variables obtained from pure kinematics is clearly worse than the one of the summed likelihood ratio $D1$. However in combination, they all help to improve the performance of the NN. The centrality, which is a measure of the fraction of the total energy of the event deposited in the transverse plane of the detector, is expected to be different for signal events due to the higher total mass of the $t\bar{t}H$ system. Since it uses a lot of information about the event topology, it is the best separating kinematic variable. For the same reason it has also the highest correlation to the $D1$ variable (see Fig. C.3). Two other event shape variables, the Fox-Wolfram moment $H1$ and the aplanarity of the b -jets, are also ranked among the best variables but show significantly worse separation. The jet p_T of the fifth jet is the only single object property, which enters the ranking due to the slightly harder p_T spectrum of $t\bar{t}H$ events.

6. Search for $t\bar{t}(H \rightarrow b\bar{b})$ Production in the Single Lepton Channel

Table 6.4.: The variable names, definitions and rankings of the variables considered in each of the regions where a NN is used [71].

Variable	Definition	Neural Network Rank				
		$\geq 6j, \geq 4b$	$\geq 6j, 3b$	$5j, \geq 4b$	$5j, 3b$	
$D1$	Neyman–Pearson MEM discriminant	1	10	-	-	
Centrality	Scalar sum of the p_T divided by sum of the E for all jets and the lepton	2	2	1	-	
p_T^{jet5}	p_T of the fifth leading jet	3	7	-	-	
$H1$	Second Fox–Wolfram moment [221, 222] computed using all jets and the lepton	4	3	2	-	
$\Delta R_{bb}^{\text{avg}}$	Average ΔR for all b -tagged jet pairs	5	6	5	-	
$\ln \mathcal{L}_{t\bar{t}H}^{\text{sum}}$	Logarithm of the summed signal likelihoods	6	4	-	-	
$m_{bb}^{\min \Delta R}$	Mass of the combination of the two b -tagged jets with the smallest ΔR	7	12	4	4	
$m_{bj}^{\max p_T}$	Mass of the combination of a b -tagged jet and any jet with the largest vector sum p_T	8	8	-	-	
$\Delta R_{bb}^{\max p_T}$	ΔR between the two b -tagged jets with the largest vector sum p_T	9	-	-	-	
$\Delta R_{\text{lep}-bb}^{\min \Delta R}$	ΔR between the lepton and the combination of the two b -tagged jets with the smallest ΔR	10	11	10	-	
$m_{uu}^{\min \Delta R}$	Mass of the combination of the two untagged jets with the smallest ΔR	11	9	-	2	
$A_{\text{plan}_{b-\text{jet}}}$	$1.5\lambda_2$, where λ_2 is the second eigenvalue of the momentum tensor [223] built with only b -tagged jets	12	-	8	-	
N_{40}^{jet}	Number of jets with $p_T \geq 40$ GeV	-	1	3	-	
$m_{bj}^{\min \Delta R}$	Mass of the combination of a b -tagged jet and any jet with the smallest ΔR	-	5	-	-	
$m_{jj}^{\max p_T}$	Mass of the combination of any two jets with the largest vector sum p_T	-	-	6	-	
H_T^{had}	Scalar sum of jet p_T	-	-	7	-	
$m_{jj}^{\min \Delta R}$	Mass of the combination of any two jets with the smallest ΔR	-	-	9	-	
$m_{bb}^{\max p_T}$	Mass of the combination of the two b -tagged jets with the largest vector sum p_T	-	-	-	1	
$p_{T,uu}^{\min \Delta R}$	Scalar sum of the p_T of the pair of untagged jets with the smallest ΔR	-	-	-	3	
$m_{bb}^{\max m}$	Mass of the combination of the two b -tagged jets with the largest invariant mass	-	-	-	5	
$\Delta R_{uu}^{\min \Delta R}$	Minimum ΔR between the two untagged jets	-	-	-	6	
m_{jjj}	Mass of the jet triplet with the largest vector sum p_T	-	-	-	7	

The fact that predominantly complicated variables enter the NN indicates how similar $t\bar{t}H$ and $t\bar{t} + b\bar{b}$ events are kinematically. Pair properties, like $\Delta R_{bb}^{\text{avg}}$, $m_{bb}^{\min \Delta R}$, $m_{bj}^{\max p_T}$, $\Delta R_{bb}^{\max p_T}$, which aim to reconstruct the angular distance or the mass of the two additional non- $t\bar{t}$ jets are also natural choices for distinguishing $t\bar{t}H$ from $t\bar{t} + b\bar{b}$ events. The $\Delta R_{\text{lep}-bb}^{\min \Delta R}$ exploits the angular distance of a b -jet pair to the lepton and $m_{uu}^{\min \Delta R}$ aims to reconstruct the hadronically

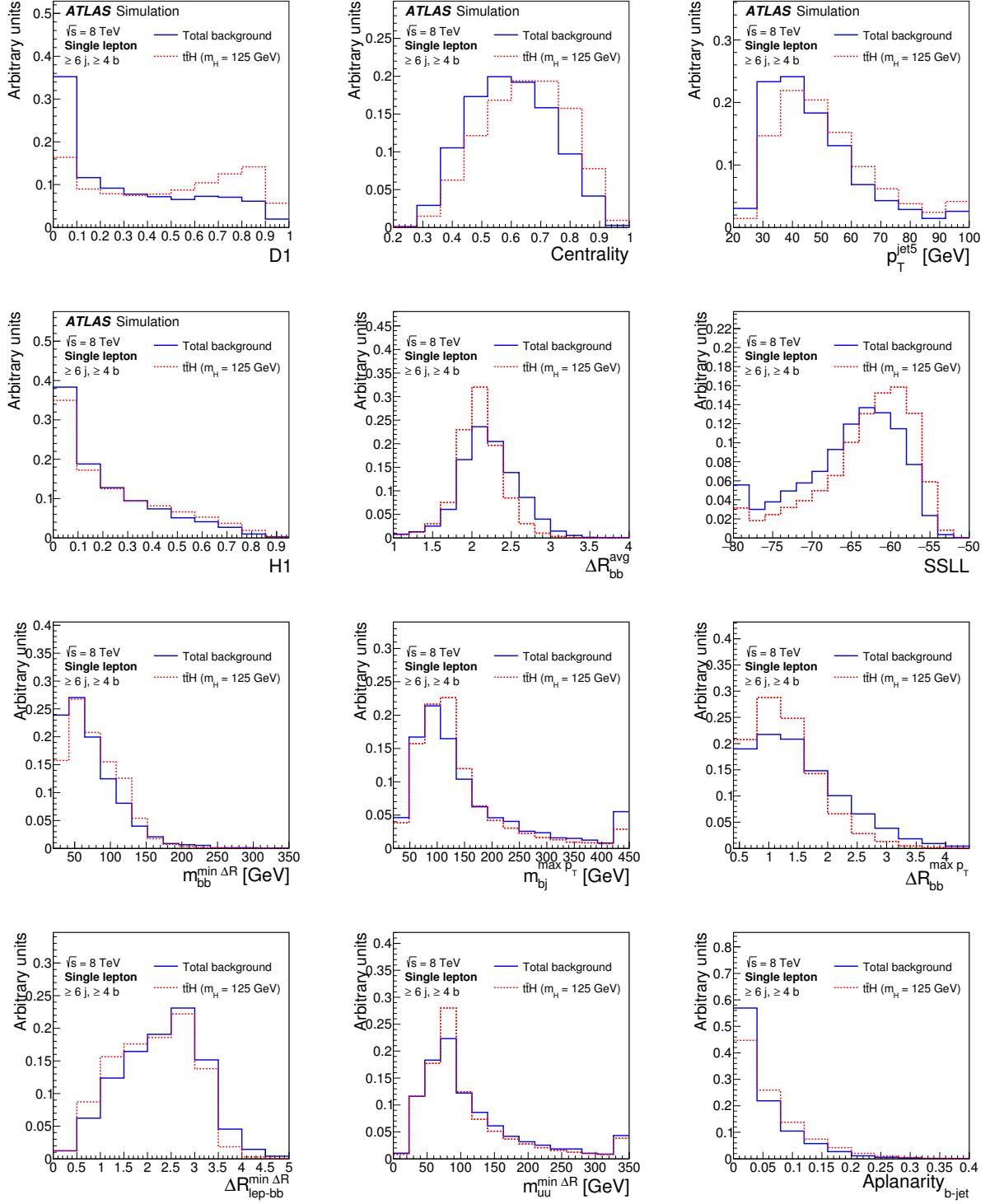


Figure 6.32.: The twelve discriminating variables in descending order used in the NN in the ($\geq 6j, \geq 4b$) region for the normalised distributions of the total background (solid blue) and the $t\bar{t}H$ signal (dashed red) predictions [71].

6. Search for $t\bar{t}(H \rightarrow b\bar{b})$ Production in the Single Lepton Channel

decaying W boson mass. The separation of the twelve highest ranked variables in descending order are shown in Figure 6.33 for the ($\geq 6j, 3b$) region.

In this region, the $D1$ variable is ranked in position ten, however the $\ln \mathcal{L}_{t\bar{t}H}^{\text{sum}}$ is the fourth most important variable and helps to improve the separation of the NN output quite significantly. The number of jets with $p_T \geq 40$ GeV, N_{40}^{jet} , is the most important variable, since $t\bar{t}H$ events tend to have more jets with high p_T due to the decaying Higgs boson. The event shape variables, centrality and $H1$, are again very important discriminating variables, as well as kinematic variables exploiting pair properties, which are mostly the same variables as chosen in the ($\geq 6j, \geq 4b$) region. All of the variables show generally less separation compared to the ($\geq 6j, \geq 4b$) region due to the less stringent b -tag requirement. The ten most important variables in the last signal-enriched region ($5j, \geq 4b$), in which no MEM variables are used, are shown in Figure 6.34.

In this region the three most important variables, centrality, $H1$ and N_{40}^{jet} , as well as H_T^{had} and the aplanarity exploit the event topology. The remaining variables are pair properties comparable to the previously used variables, but with a few exchanges of variables with similar properties. For instance, instead of the $m_{uu}^{\min \Delta R}$ the $m_{jj}^{\min \Delta R}$ and instead of $m_{bj}^{\max p_T}$ the $m_{jj}^{\max p_T}$ variable is used. But in each case both are exploiting the same properties of the event. The discrimination of the signal from the background is slightly better than in the ($\geq 6j, 3b$) region, however this signal-enriched region has the lowest statistics. In the ($5j, 3b$) region, where the NN is trained to separate $t\bar{t} + \text{HF}$ from $t\bar{t} + \text{light}$, the highest ranked seven variables naturally differ considerably from the ones discussed above, as can be seen in Figure 6.35.

The invariant mass of the two b -tagged jets with the highest vectorial transverse momentum, $m_{bb}^{\max p_T}$, yields the best separation, since it is expected to be higher for $t\bar{t} + b\bar{b}$ events. Very similar variables exploiting the same properties are $m_{bb}^{\min \Delta R}$ and $m_{bb}^{\max m}$ but they are ranked at fourth and fifth position due to less discrimination and correlations among the three variables. The second and third best variable are combinations using untagged jets, $m_{uu}^{\min \Delta R}$ and $p_{T,uu}^{\min \Delta R}$, which are expected to be the W boson mass and p_T . In case of $t\bar{t} + \text{light}$ events, the W boson properties are poorly reconstructed, because of the large fraction of events, in which a c -quark originating from the W boson is mistagged as a b -jet. Similarly, the expected angular distance between the two jets from the W boson $\Delta R_{uu}^{\min \Delta R}$ shows good separation, but is only ranked sixth, because of its correlation to $m_{uu}^{\min \Delta R}$. Finally, the mass of the jet triplet with the largest vector sum p_T , m_{jjj} which is often used for a simple reconstruction of the hadronically decaying top quark mass, yields a better reconstruction in case of $t\bar{t} + \text{light}$ events and enters the list last.

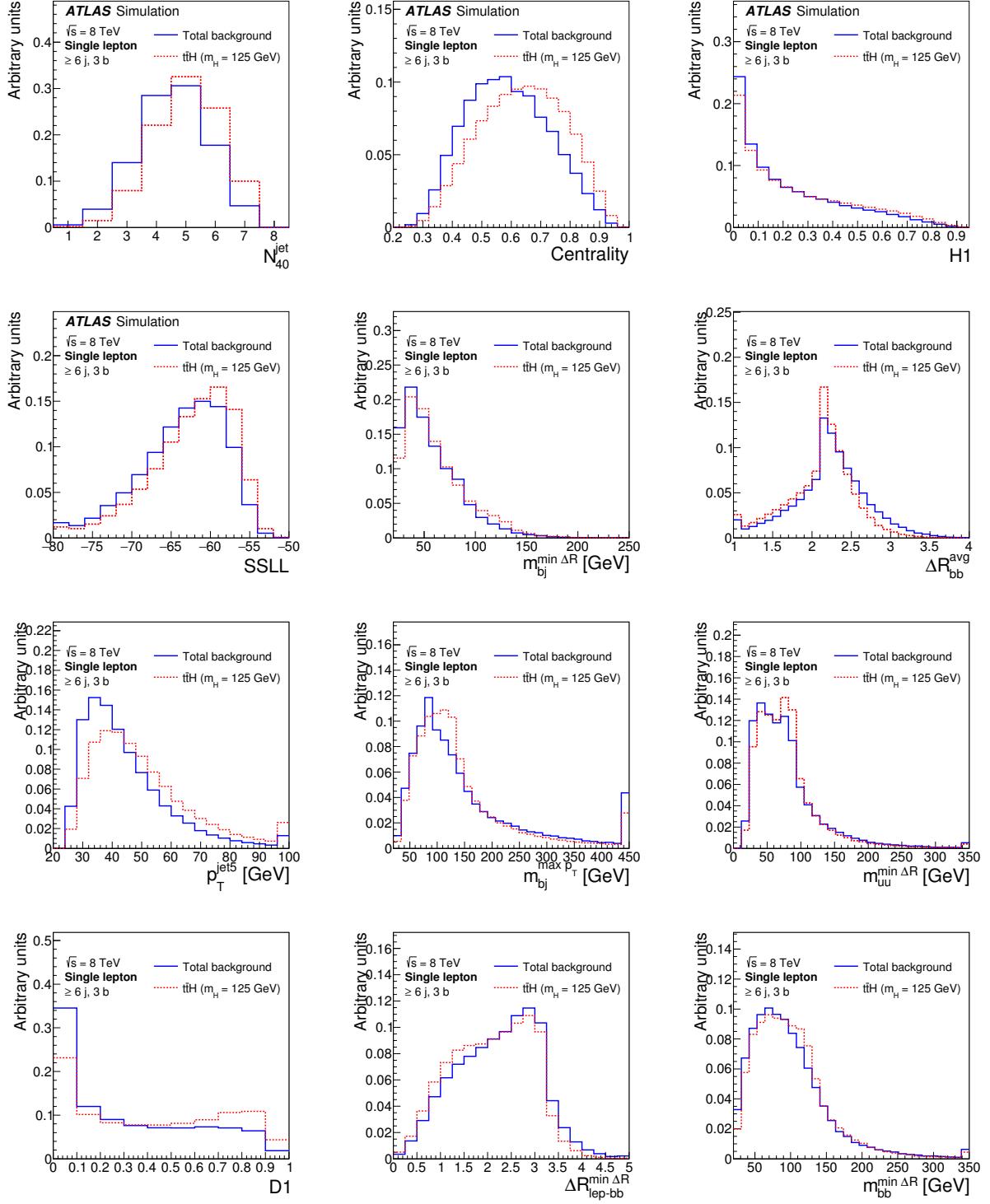


Figure 6.33.: The twelve discriminating variables in descending order, used in the NN in the ($\geq 6j, 3b$) region for the normalised distributions of the total background (solid blue) and the $t\bar{t}H$ signal (dashed red) predictions [71].

6. Search for $t\bar{t}(H \rightarrow b\bar{b})$ Production in the Single Lepton Channel

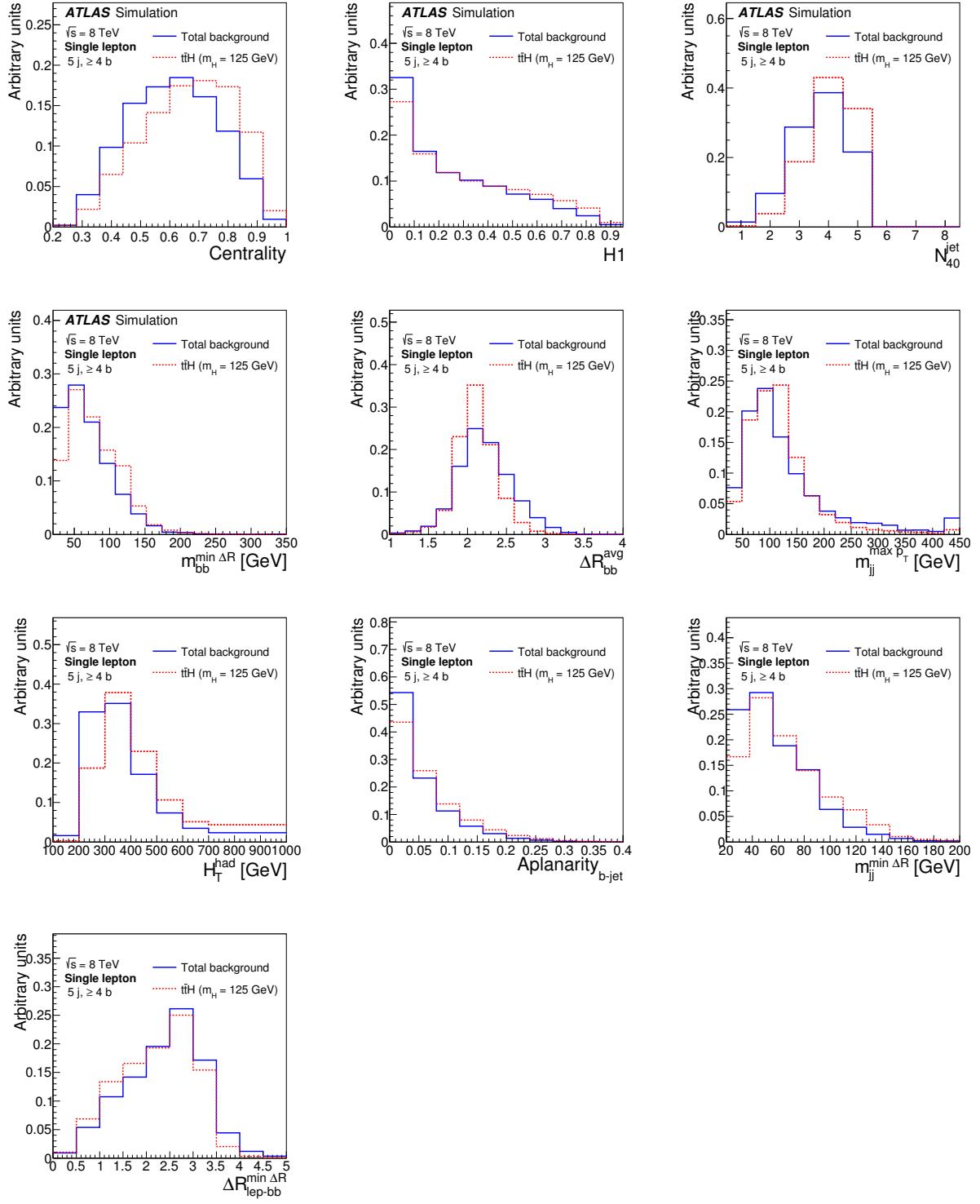


Figure 6.34.: The ten discriminating variables in descending order used in the NN in the $(5j, \geq 4b)$ region for the normalised distributions of the total background (solid blue) and the $t\bar{t}H$ signal (dashed red) predictions [71].

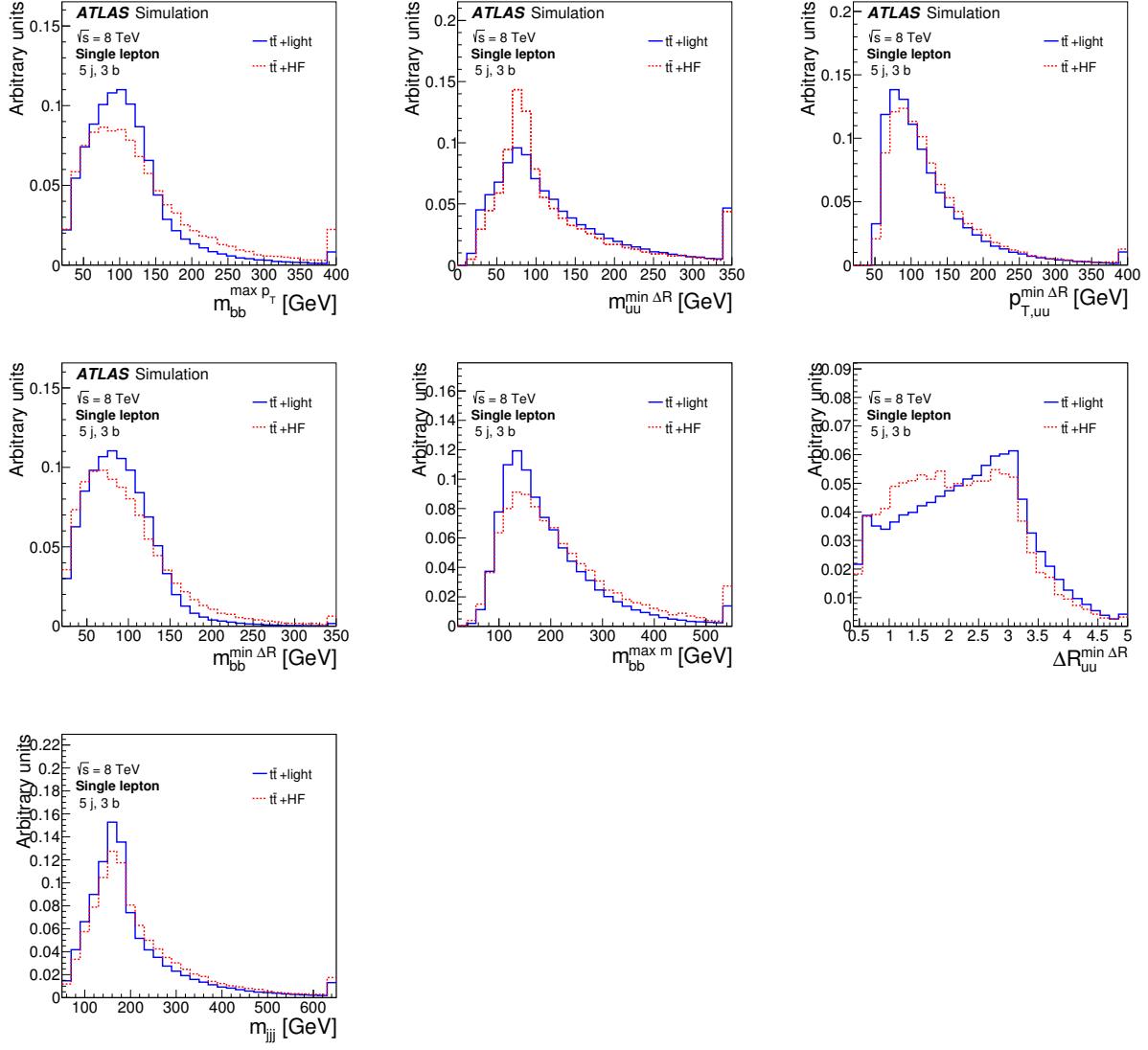


Figure 6.35.: The seven discriminating variables in descending order used in the NN in the (5j, 3b) region for the normalised distributions of the $t\bar{t} + \text{light}$ (solid blue) and the $t\bar{t} + \text{HF}$ (dashed red) predictions [71].

6. Search for $t\bar{t}(H \rightarrow b\bar{b})$ Production in the Single Lepton Channel

6.6.3. The Neural Network Output

After the NN training the NN output is obtained for all predictions using the full simulations and data. In case of the six jet inclusive regions, the output is once evaluated using the purely kinematic NN and the extended one including the two MEM variables to study the improvements due to the MEM. In Figure 6.36 the normalised distributions of the NN outputs for the total background (solid blue) and the $t\bar{t}H$ signal (dashed red) are shown in the ($\geq 6j, \geq 4b$) region.

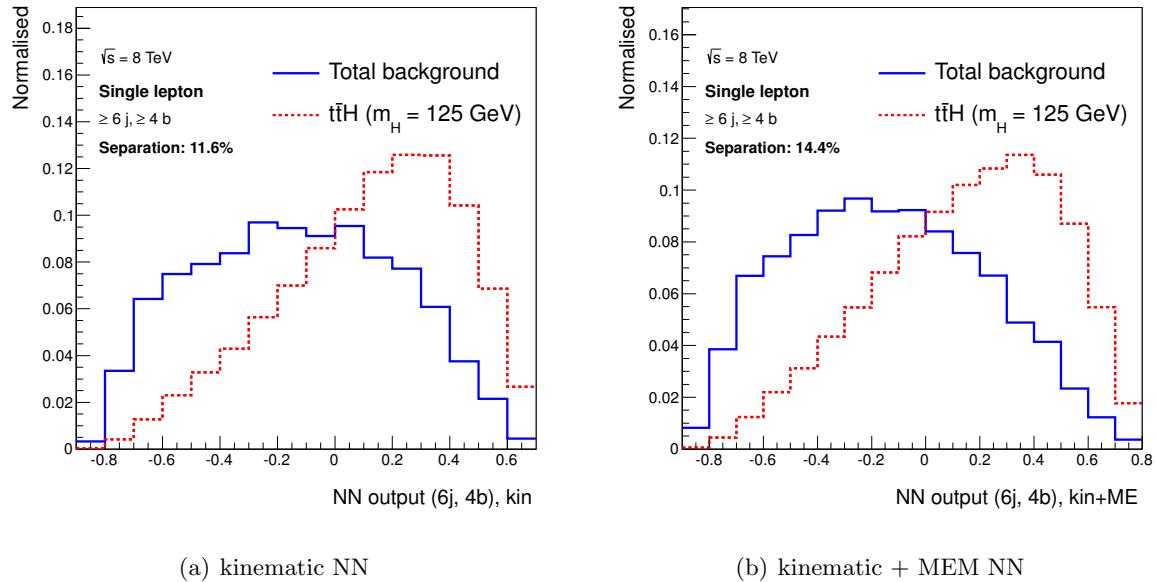


Figure 6.36.: The NN output using purely kinematic (a) and in addition MEM (b) variables in the ($\geq 6j, \geq 4b$) region for the total background (solid blue) and the $t\bar{t}H$ signal (dashed red) predictions.

The highly optimised kinematic NN output with a great separation of 11.6% is further improved by the addition of the MEM variables leading to a separation of 14.4%, corresponding to a relative improvement of roughly 24%. Considering that more than 300 kinematic variables are used to select the best ten variables for the kinematic NN and no significant improvement could be gained by the addition of more variables (see Fig. C.1), this is a remarkable achievement. The output of the two NNs in the ($\geq 6j, 3b$) region is shown in Figure 6.37 for the total background and signal normalised to unity.

In this region, the separation of the NN improves from 10.4% to 11.6% by using the extra two MEM variables. As expected from the ranking of the variables, the relative improvement of 12% is smaller than in the ($\geq 6j, \geq 4b$) region, however, the MEM variables still help considerably. Finally, the response of the NN in the ($5j, \geq 4b$) region and the output of the dedicated NN separating $t\bar{t} + \text{HF}$ from $t\bar{t} + \text{light}$ in the ($\geq 6j, 3b$) region are shown in Figure 6.38.

Given that only kinematic variables are used, the ($5j, \geq 4b$) region (a) yields a great separation of 12.4%, which is even better than the one from the purely kinematic NN in the ($\geq 6j, \geq 4b$) region, but not as good as the final one with the addition of the MEM. Although the discrimination is better than in the ($\geq 6j, 3b$) region, the ($5j, \geq 4b$) region is still less sensitive due to the significantly smaller statistics. The NN output of the ($\geq 6j, 3b$) region (b) shows a

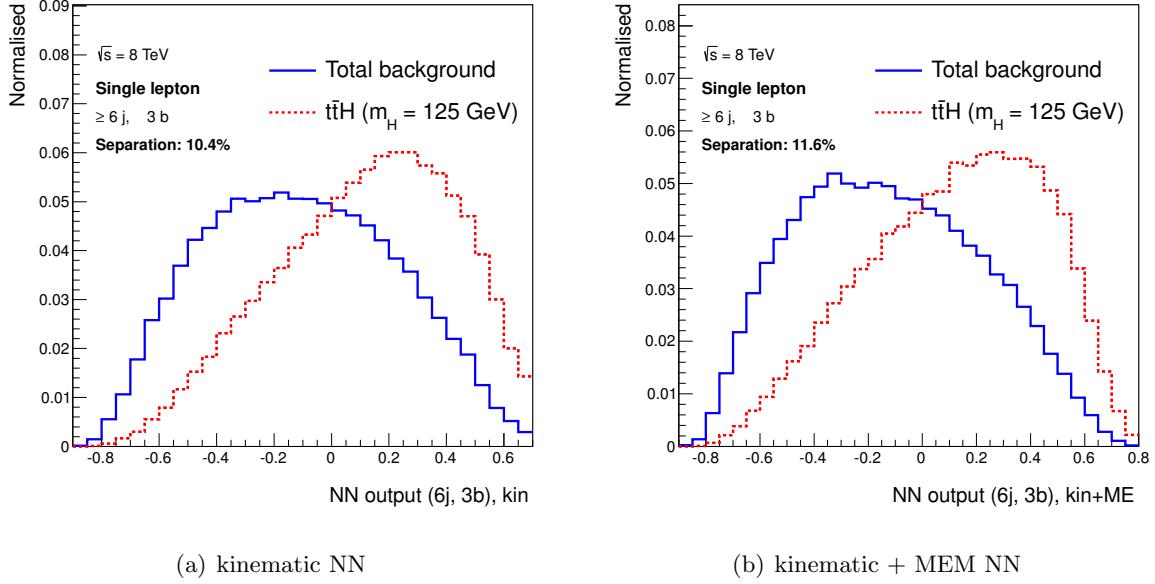


Figure 6.37.: The NN output using purely kinematic (a) and in addition MEM (b) variables in the ($\geq 6j, 3b$) region for the total background (solid blue) and the $t\bar{t}H$ signal (dashed red) predictions.

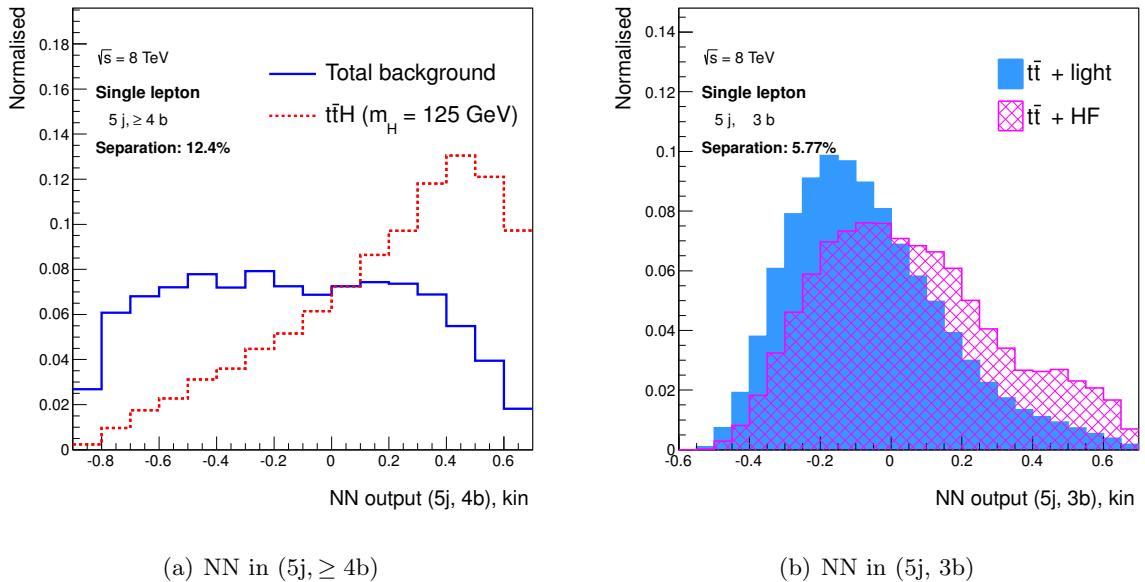


Figure 6.38.: The NN output discriminating signal from background in the ($5j, \geq 4b$) region (a) and discriminating $t\bar{t} + \text{light}$ from $t\bar{t} + \text{HF}$ events in the ($5j, 3b$) region (b).

separation of 5.8% between $t\bar{t} + \text{HF}$ and $t\bar{t} + \text{light}$ events. Since $t\bar{t} + \text{HF}$ and $t\bar{t} + \text{light}$ events are kinematically very similar it is difficult to find good discriminating variables. However employing the NN in this region helps considerably to reduce the systematic uncertainties related to these processes.

6.7. The Profile Likelihood Fit

6.7.1. The Likelihood Function

A binned profile likelihood fit, using the H_T^{had} and the NN distributions as discriminants, is performed simultaneously in all nine analysis regions in order to test for the presence of a signal, assuming a Higgs boson mass of $m_H = 125$ GeV. Without taking the systematic uncertainties into account the purely statistical likelihood function $\mathcal{L}_{\text{stat}}(\mu)$ depends only on the observed signal strength $\mu = \sigma_{t\bar{t}H}^{\text{obs}}/\sigma_{t\bar{t}H}^{\text{SM}}$ and is defined as the product of the Poisson probabilities of the N_{bins} histogram bins of each distribution and in each region comparing the total predictions to data

$$\mathcal{L}_{\text{stat}}(\mu) = \prod_{i=0}^{N_{\text{bins}}} \frac{(\mu s_i + b_i)^{n_i}}{n_i!} e^{-(\mu s_i + b_i)}, \quad (6.4)$$

where n_i are the observed data events, s_i and b_i are the predicted signal and backgrounds events, respectively, in the i -th bin. However the predictions s_i and b_i are not perfect and vary within systematic uncertainties, of which some only influence certain processes. Thus the total background prediction $b_i(\alpha) = \sum_j \alpha_j p_{ij}$ is split into the number of events of the different contributing processes $p_{k,i}$, where α_j are normalisation parameters of the processes, which can be varied according to the uncertainties and are equal to unity for the nominal predictions. Assuming Gaussian uncertainties σ_j , the parameter α_j are constrained by a Gaussian prior probability leading to

$$\mathcal{L}_{\text{norm}}(\mu, \alpha) = \prod_{i=0}^{N_{\text{bins}}} \frac{(\mu s_i + b_i(\alpha))^{n_i}}{n_i!} e^{-(\mu s_i + b_i(\alpha))} \prod_{j=0}^{N_{\text{proc}}} \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{(\alpha_j - 1)^2}{2\sigma_j^2}}. \quad (6.5)$$

Technically, any Gaussian function with observable a , best estimate \hat{a} and uncertainty σ_a can be transformed to a normal distribution with $\hat{\theta} = 0$ and $\sigma_\theta = 1$ by the relation $a = \hat{a}(1 + \sigma_a\theta)$. Not only a single uncertainty on the cross section of the process allows the predictions to vary in each bin, but many other systematic uncertainties as discussed in Section 6.4. Each of these uncertainties can be represented by such a NP θ influencing the predicted signal or backgrounds events. Assuming that all of them are represented by Gaussian functions, the likelihood is given by

$$\mathcal{L}(\mu, \theta) = \prod_{i=0}^{N_{\text{bins}}} \frac{(\mu s_i(\theta) + b_i(\theta))^{n_i}}{n_i!} e^{-(\mu s_i(\theta) + b_i(\theta))} \prod_{k=0}^{N_{\text{syst}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{\theta_k^2}{2}}. \quad (6.6)$$

However, the prior distributions $\rho(\theta)$ for the uncertainties are not necessarily Gaussian. Positively defined variables can be represented by log-normal functions, statistical uncertainties associated with the number of MC events are modelled by gamma distributions and a flat prior can be assigned to unconstrained parameters. Such an unconstrained parameter is the signal strength μ , which can vary freely, but is required to be the same in all analysis regions. Finally, correlations among the parameters and among the bins (analysis regions) are taken into account. The variations of the bin content are only known for 1σ variations of the uncertainties corresponding to $\theta_k = \pm 1$, for which templates are created. The values in between are interpolated and values larger than $\pm 1\sigma$ are extrapolated by using various *vertical template*

morphing procedures [224]. In order to take the finite MC statistics into account, one additional NP for each bin is introduced, which allows the template to vary within the statistical uncertainty of the MC [225]. Statistical fluctuations in the templates might cause an artificial shape variation, which can bias the fit results. In order to avoid that, each template undergoes an automatic *smoothing* procedure, in which bins are merged until a shape change is significant with respect to the statistical fluctuations. Systematic uncertainties affecting the total normalisation or bin of the distribution by less than 0.5% are neglected. Such uncertainties play no role in the fit, but increase the convergence time. This procedure is referred to as *pruning*. This pruning does not affect by any means the final result of the analysis. The profile likelihood fit in this analysis is performed using the ROOFIT framework [224, 226].

The described procedure takes advantage of the highly populated background-dominated regions to reduce the systematic uncertainties in the signal-enriched regions, which requires a good understanding of the normalisation and shape effects of the various uncertainties on the discriminants. The effects before (*pre-fit*) and after (*post-fit*) the profile likelihood fit have been studied extensively and are well understood.

6.7.2. Extracted Signal Strength

By maximising the likelihood or minimising the $-\ln \mathcal{L}$ the best estimators $\hat{\mu}$ and $\hat{\theta}$ are obtained. In order to obtain the uncertainties on the estimators, the profile likelihood function is derived by minimising the likelihood function for each value of the estimator. For sufficiently large statistics, the profile likelihood function is dominated by Gaussian probabilities. In that case, the 1σ uncertainties are defined by a variation of ± 0.5 of the log-likelihood value with respect to the maximum. The impact of the various systematic uncertainties on the uncertainty on $\hat{\mu}$ can be determined by scanning the likelihood profile and keeping the particular NP fixed. The statistical uncertainty on the estimator $\hat{\mu}$ is found by scanning the profile likelihood and allowing no variations of the NPs.

The best fit value of the signal strength for a Higgs boson mass of $m_H = 125$ GeV is

$$\hat{\mu}(m_H = 125 \text{ GeV}) = 1.2 \pm 1.3 (0.8) , \quad (6.7)$$

where 1.3 is the total uncertainty and 0.8 is the statistical component. The signal strength agrees with the SM expectation of $\mu = 1$, however the large uncertainty shows that the search is not sensitive to the SM value yet. The observed excess corresponds to a 0.9σ deviation from the background-only hypothesis. The obtained uncertainty agrees with the expected uncertainty obtained from *Asimov pseudo data* [227]. The following results show the outcome of the fit under the signal-plus-background hypothesis with the extracted signal strength $\hat{\mu}$. However, when setting the signal strength $\mu = 0$, very similar results on the background corrections and the constraints on the systematic uncertainties are obtained.

6.7.3. Improvements of the Predictions

The pre-fit and post-fit distributions of the discriminating distributions in the nine analysis regions serving as input to the profile likelihood fit are discussed in the following. In Figure 6.39 the H_T^{had} pre-fit (a) and post-fit (b) distributions in the analysis regions with exactly four jets are shown.

While in the (4j, 2b) region the predictions agree quite well with the data before the fit and the uncertainties seem to be overestimated, the predictions underestimate the data considerably in

6. Search for $t\bar{t}(H \rightarrow b\bar{b})$ Production in the Single Lepton Channel

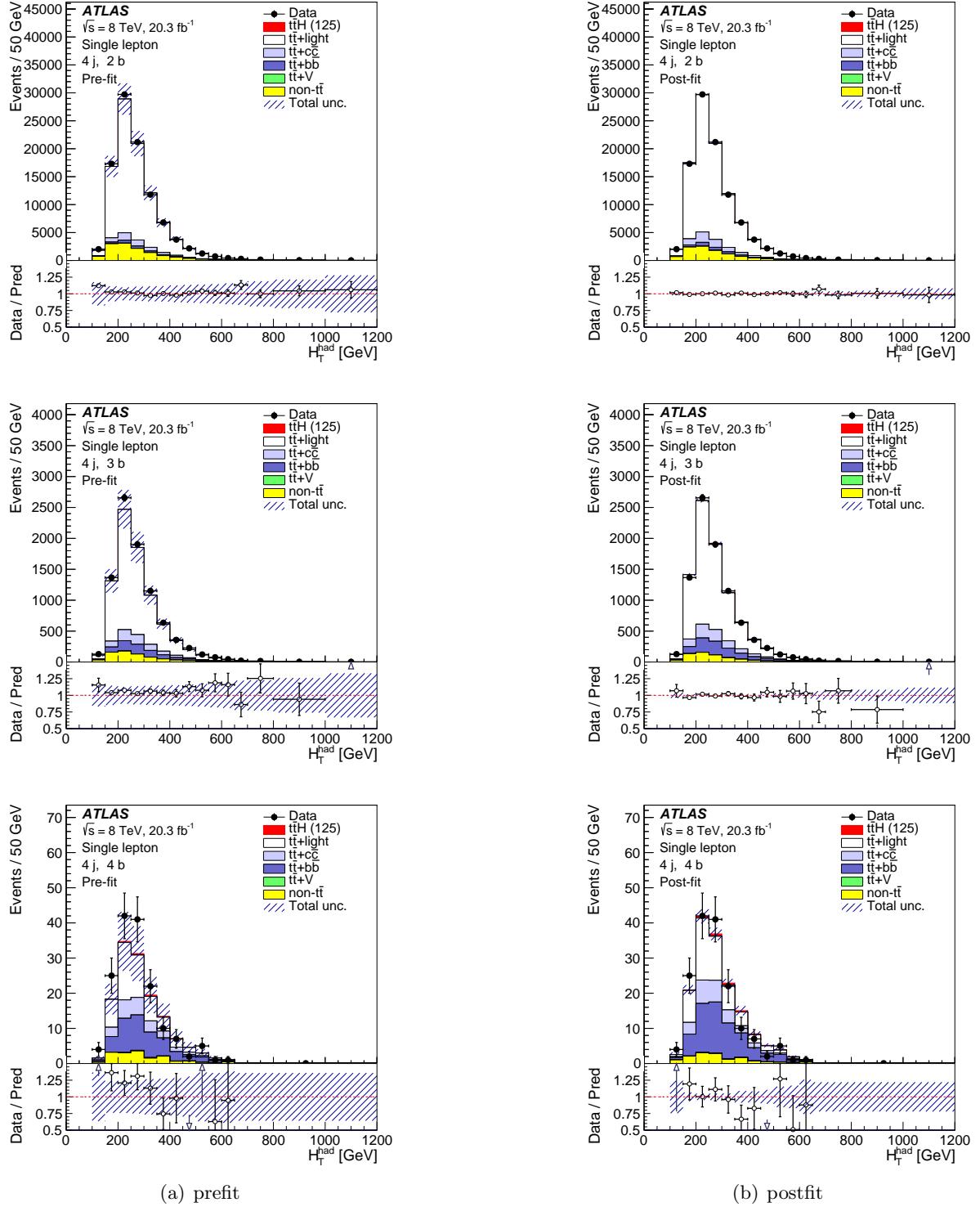


Figure 6.39.: The H_T^{had} distributions in the analysis regions with exactly four jets before (a) and after (b) the fit for the total predictions and data. The $t\bar{t}H$ signal yield (solid red) is normalised to $\mu = 1$ and the best fitted $\hat{\mu}$ before and after the fit, respectively [71].

the (4j, 3b) and slightly in the low statistics (4j, 4b) region. The deviations are still within the large systematic uncertainties covering up to 25% of the predictions. However in the post-fit distributions, the uncertainties are significantly reduced. In particular in the (4j, 3b) region the predictions are notably improved after the fit such that they agree very well with the data. The pre-fit and post-fit distributions in the analysis regions with exactly five jets, using either H_T^{had} ($= 2$ b -tag) or the NN output (≥ 3 b -tag) are presented in Figure 6.40.

In the region with two b -tags the agreement of the H_T^{had} distribution between data and prediction is quite good before the fit, such that the uncertainties seem conservative. In the three b -tag region, where the dedicated NN output separating $t\bar{t} + \text{HF}$ from $t\bar{t} + \text{light}$ is used, the model systematically predicts too few events, though the shape of the output seems to agree quite well in bins with high statistics. In the four b -tag region, insufficient data statistics is available leading to large fluctuations. After the fit, the systematic uncertainties are considerably reduced in all regions and the normalisation discrepancy in the (5j, 3b) region is corrected leading to a nice agreement between expectations and data. In the (5j, ≥ 4 b) region large statistical fluctuations do not allow to make conclusions about the data agreement with the model. Finally, the input distributions for the profile likelihood fit in the six jet region are shown in Figure 6.41.

Similar observations as in the lower jet multiplicities can be made about the pre-fit distributions. In the two b -tag region, the predictions agree very well with data, while some normalisation discrepancy seem to occur in the three b -tag region. In the four b -tag region, which is the most sensitive region, the agreement with data is reasonable before the fit. In the (≥ 6 j, 3b) and (≥ 6 j, ≥ 4 b) regions the predictions are further improved and its uncertainties are reduced significantly after the fit. In summary, the uncertainties decrease significantly in all regions due to constraints provided by the data and correlations between different sources of uncertainty introduced by the fit to the data. This can be also seen in the visualisation of the pre-fit (a) and post-fit (b) event yields as presented in Figure 6.42 for the nine analysis regions.

While in the high b -tag regions the predictions significantly underestimate the data before the fit, the agreement in the yields improves greatly after the profile likelihood fit, such that they almost perfectly match the data. By using the large amount of data in the high statistics background-dominated regions better estimates on the uncertainties are obtained, which remarkably reduces the pre-fit uncertainties. Only with these reduced uncertainties a meaningful search for the very small signal can be performed. The post-fit yields for the various backgrounds are summarised in Table 6.5.

Compared to the pre-fit yields as presented in Table 6.1, the agreement between total prediction and data is significantly improved within much smaller uncertainties on the predictions across the analysis regions. In almost all regions, the signal+background predictions agree within 1σ deviation with the data. The expectations of the (5j, ≥ 4 b), (≥ 6 j, 2b) and (≥ 6 j, 3b) regions agree within 2σ uncertainty. The single uncertainties of a certain process are often larger than the uncertainty of the total prediction. The uncertainty of the total prediction is considerably smaller due to anti-correlations of uncertainties among processes. The expectations of the $t\bar{t}H$ event yield generally increase slightly due to an excess of data events in the signal sensitive regions.

6. Search for $t\bar{t}(H \rightarrow b\bar{b})$ Production in the Single Lepton Channel

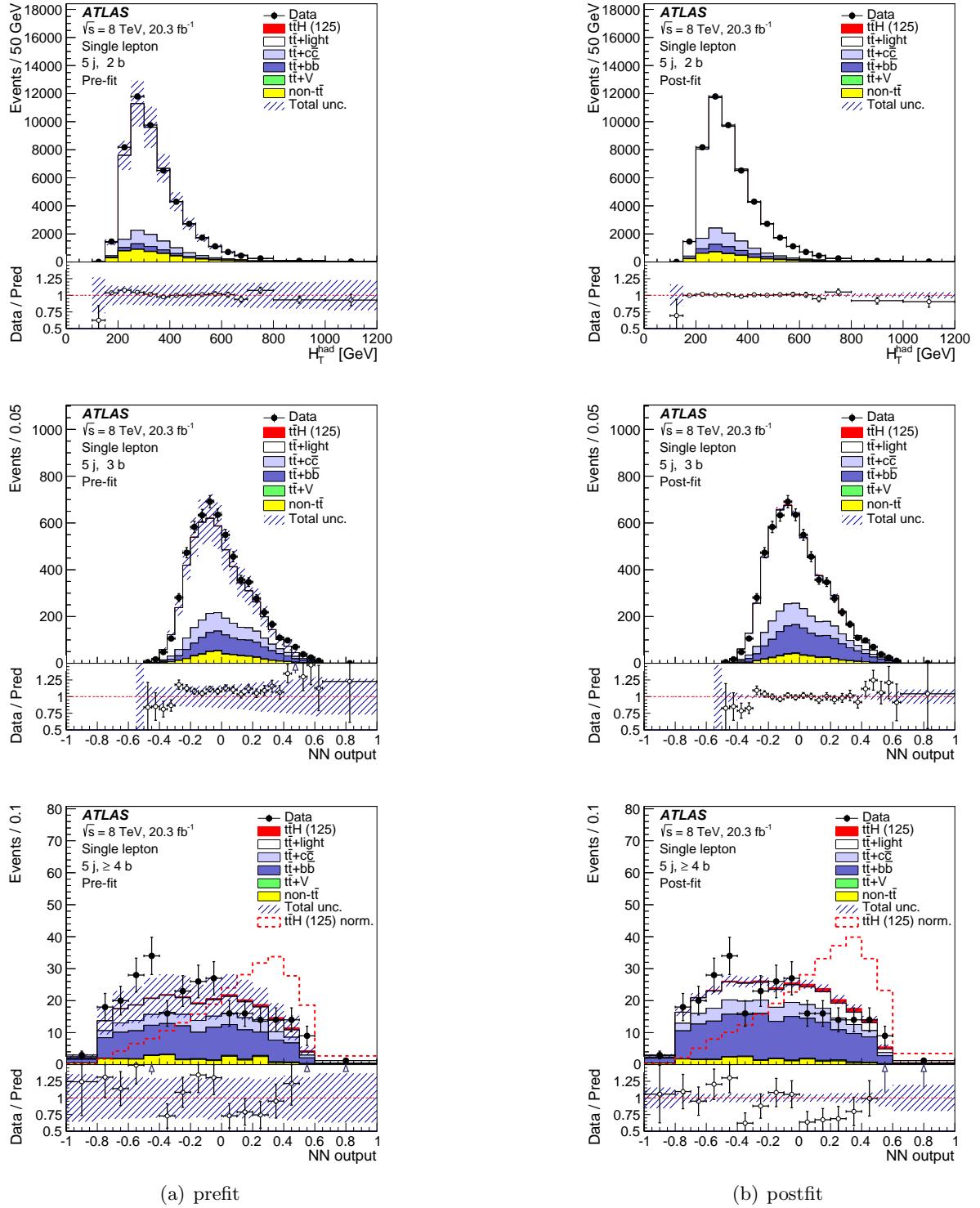


Figure 6.40.: The H_T^{had} and NN distributions in the analysis regions with exactly five jets before (a) and after (b) the fit for the total predictions and data. The $t\bar{t}H$ signal yield (solid red) is normalised to $\mu = 1$ and the best fitted $\hat{\mu}$ before and after the fit, respectively [71].

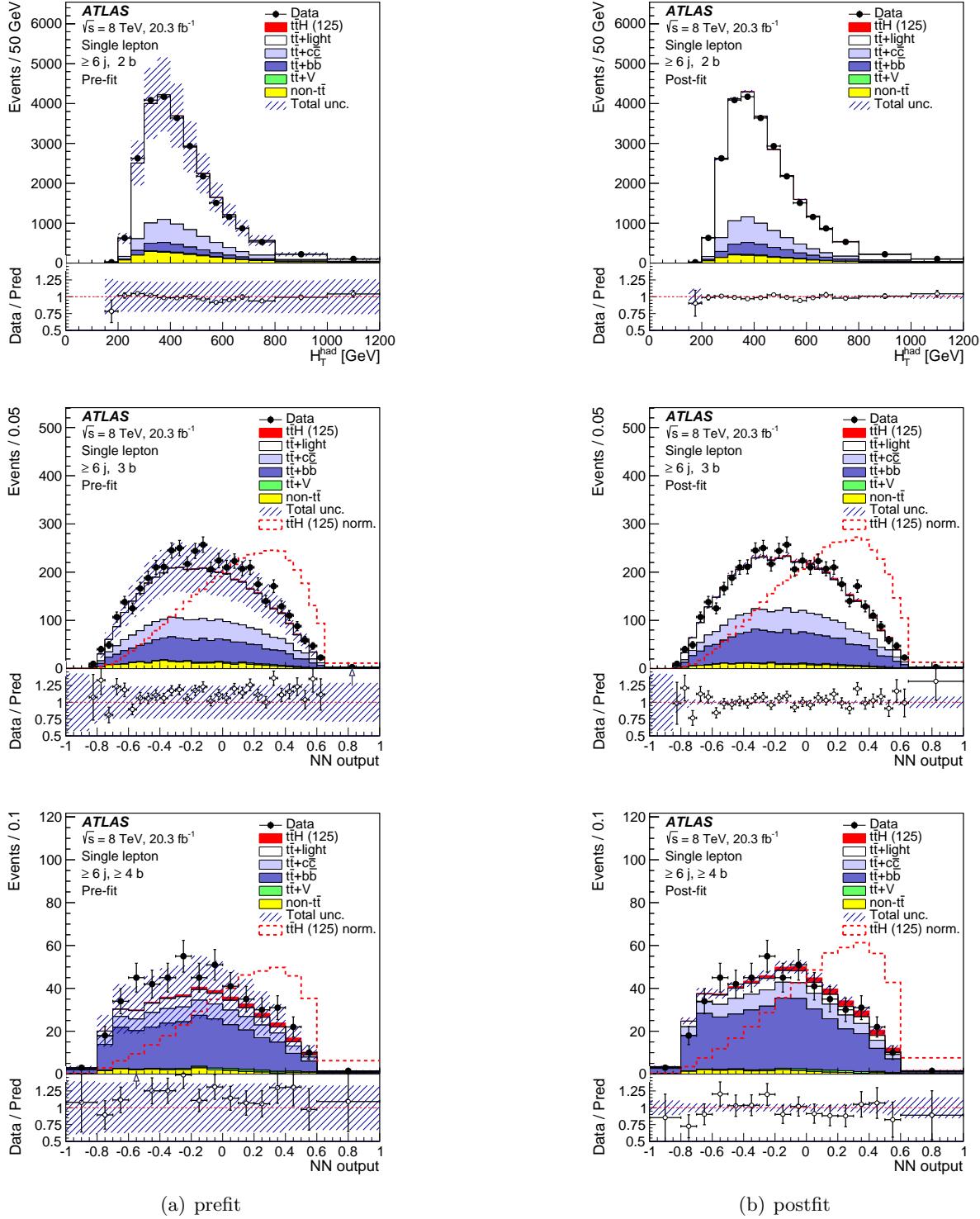


Figure 6.41.: The H_T^{had} and NN distributions in the analysis regions with at least six jets before (a) and after (b) the fit for the total predictions and data. The $t\bar{t}H$ signal yield (solid red) is normalised to $\mu = 1$ and the best fitted $\hat{\mu}$ before and after the fit, respectively [71].

6. Search for $t\bar{t}(H \rightarrow b\bar{b})$ Production in the Single Lepton Channel

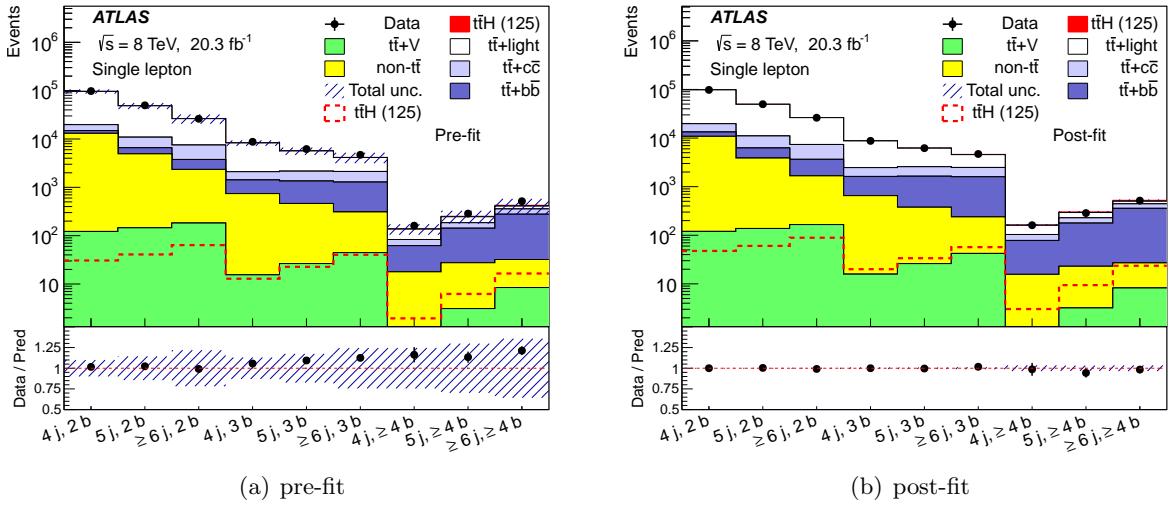


Figure 6.42.: Comparison of the total predictions and data in all nine analysis regions before (a) and after the fit (b). The signal, normalised to the SM prediction (a) and to the best fitted $\hat{\mu}$, is shown both as a filled red area stacked on the backgrounds and separately as a dashed red line. The hashed area corresponds to the total uncertainty on the yields [71].

Table 6.5.: Post-fit event yields for signal, backgrounds and data in each of the analysis regions. The quoted uncertainties are the sum in quadrature of the statistical and systematic uncertainties on the yields, computed taking into account correlations among nuisance parameters and among processes. Due to these correlations the uncertainties on a single process can be larger than the uncertainty on the total prediction.

	4 j, 2 b	4 j, 3 b	4 j, 4 b	5 j, 2 b
$t\bar{t}H$ (125)	48 ± 35	20 ± 15	3.0 ± 2.2	60 ± 44
$t\bar{t}$ + light	$78\,200 \pm 1600$	6260 ± 160	56.5 ± 4.7	$38\,400 \pm 1000$
$t\bar{t} + c\bar{c}$	6430 ± 1800	845 ± 220	25.5 ± 6.5	4800 ± 1200
$t\bar{t} + b\bar{b}$	2480 ± 490	969 ± 150	62.5 ± 8.5	2380 ± 360
W +jets	3650 ± 1100	166 ± 51	4.0 ± 1.2	1210 ± 420
Z +jets	1060 ± 540	49 ± 25	1.1 ± 0.6	368 ± 200
Single top	4710 ± 320	333 ± 28	6.8 ± 0.7	1730 ± 150
Diboson	216 ± 65	11.3 ± 3.7	0.3 ± 0.1	94 ± 35
$t\bar{t}V$	120 ± 38	15.8 ± 4.9	0.9 ± 0.3	138 ± 43
Lepton misID	1080 ± 370	78 ± 26	2.6 ± 1.0	343 ± 110
Total	$98\,000 \pm 340$	8750 ± 82	163 ± 6	$49\,500 \pm 220$
Data	98 049	8752	161	49 699

	5 j, 3 b	5 j, ≥ 4 b	≥ 6 j, 2 b	≥ 6 j, 3 b	≥ 6 j, ≥ 4 b
$t\bar{t}H$ (125)	34 ± 25	9.4 ± 6.9	89 ± 65	57 ± 42	24 ± 17
$t\bar{t}$ + light	3610 ± 120	65.3 ± 5.6	$18\,900 \pm 700$	2080 ± 87	57.9 ± 5.3
$t\bar{t} + c\bar{c}$	935 ± 230	51 ± 12	3730 ± 890	888 ± 210	85 ± 21
$t\bar{t} + b\bar{b}$	1260 ± 180	155 ± 20	1980 ± 310	1360 ± 190	331 ± 37
W +jets	87 ± 31	4.0 ± 1.5	455 ± 170	51 ± 19	4.4 ± 1.9
Z +jets	28 ± 16	1.4 ± 0.8	152 ± 86	15.6 ± 8.9	1.2 ± 0.7
Single top	185 ± 18	8.2 ± 0.7	734 ± 83	111 ± 14	11.4 ± 1.6
Diboson	8.0 ± 3.1	0.5 ± 0.2	45 ± 20	5.6 ± 2.6	0.5 ± 0.2
$t\bar{t}V$	26.1 ± 8.1	3.2 ± 1.0	166 ± 52	42 ± 13	8.2 ± 2.5
Lepton misID	44 ± 16	5.7 ± 2.2	117 ± 41	13.8 ± 5.3	1.1 ± 0.5
Total	6220 ± 54	303 ± 10	$26\,400 \pm 160$	4620 ± 55	525 ± 18
Data	6199	286	26 185	4701	516

6.7.4. Constraining the Systematic Uncertainties

As observed in the various post-fit distributions, the systematic uncertainties are significantly reduced by the profile likelihood fit. The effect of the fit on the most important systematic uncertainties and their impact on the signal strength is illustrated in Figure 6.43.

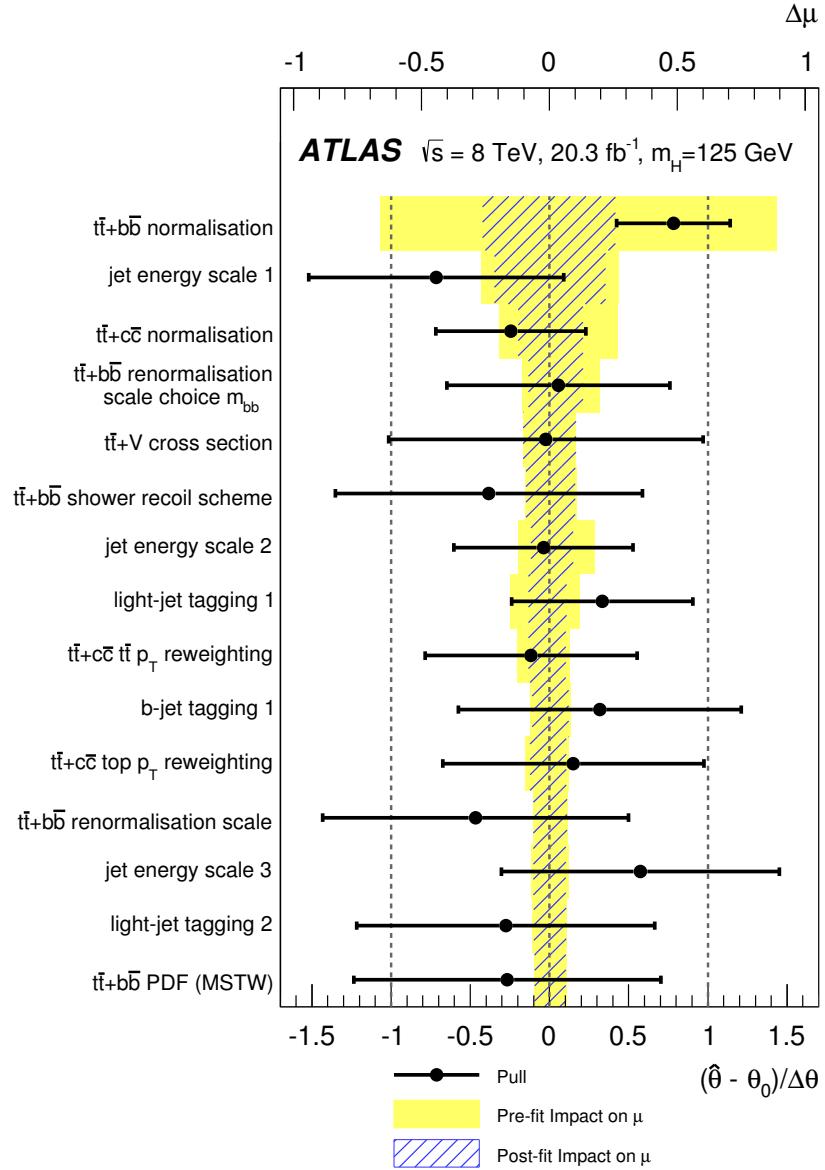


Figure 6.43: The pull and constraints of the most important NPs having the largest impact on the uncertainty of $\hat{\mu}$ ranked in descending order. The points (black) show the deviation of the fitted values $\hat{\theta}$ from the pre-fit values θ_0 in units of their pre-fit uncertainty $\Delta\theta$ (bottom axis). The error bars correspond to the post-fit uncertainties $\sigma_\theta/\Delta\theta$. The error bars smaller than one indicate a reduced uncertainty with respect to the initial uncertainty. The shaded yellow area corresponds to the pre-fit uncertainty contribution to $\Delta\mu$, while the hashed blue area represents the post-fit impact on $\Delta\mu$ (top axis) [71].

The uncertainties are shown in descending order dependent on their post-fit contribution to the total uncertainty of the extracted signal strength. The pull of the systematic uncertainties, which is defined as difference of the post-fit value $\hat{\theta}$ to the pre-fit value θ_0 in units of the pre-fit uncertainty $\Delta\theta$, is drawn as black points (bottom axis). The error bars correspond to the relative post-fit uncertainties σ_θ and the vertical dashed lines indicate a 1σ deviation with respect to the pre-fit uncertainties. The shaded yellow area corresponds to the pre-fit uncertainty contribution to $\Delta\mu$, while the hashed blue area represents the post-fit impact on $\Delta\mu$ (top axis). The post-fit impact on μ is estimated by fixing the corresponding NP at

$\hat{\theta} \pm \sigma_\theta$ and performing the fit again. The most important systematic uncertainty is the one on the $t\bar{t} + b\bar{b}$ normalisation, whose central value is significantly pulled to higher values ($\approx 0.8\sigma$) resulting in an increase in the observed $t\bar{t} + b\bar{b}$ yield. This is not unexpected given the differences in yields between data and simulation in the high b -tag regions before the fit. The 50% pre-fit uncertainty is larger than the difference between latest theoretical calculation. This leaves the fit enough freedom to correct the predictions to the data in the signal-depleted regions of the discriminants in the ($\geq 6j, 3b$) and ($\geq 6j, \geq 4b$) analysis region. At the same time, the large amount of data allows the fit to reduce the uncertainty significantly. This also reduces the impact on the uncertainty of $\hat{\mu}$, which is visualised in the comparison of the shaded yellow and hashed blue area. Since the information about of the $t\bar{t} + b\bar{b}$ normalisation can be obtained only in the signal-enriched regions, an artificial increase or an over-constrain could bias the result of the signal strength extraction. Thus intensive studies have been performed to ensure that this is not the case. For example, if no Gaussian prior is applied to the $t\bar{t} + b\bar{b}$ normalisation, corresponding to a freely floating parameter, the fit still prefers an increase in the amount of $t\bar{t} + b\bar{b}$ background by about 40%. In Table 6.6 various normalisation uncertainties for the main background processes are compared before and after the fit in the ($\geq 6j, \geq 4b$) region. In that region, the uncertainty on the $t\bar{t} + b\bar{b}$ normalisation is reduced from 50% to 14%. Other

Table 6.6.: Normalisation uncertainties (expressed in %) on signal and the main background processes for the systematic uncertainties considered, before and after the fit to data in the ($\geq 6j, \geq 4b$) region. The total uncertainty can be different from the sum in quadrature of individual sources due to the anti-correlations between them [71].

	Pre-fit				Post-fit			
	$t\bar{t}H$ (125)	$t\bar{t} + \text{light}$	$t\bar{t} + c\bar{c}$	$t\bar{t} + b\bar{b}$	$t\bar{t}H$ (125)	$t\bar{t} + \text{light}$	$t\bar{t} + c\bar{c}$	$t\bar{t} + b\bar{b}$
Luminosity	± 2.8	± 2.8	± 2.8	± 2.8	± 2.6	± 2.6	± 2.6	± 2.6
Lepton efficiencies	± 1.4	± 1.4	± 1.4	± 1.5	± 1.3	± 1.3	± 1.3	± 1.3
Jet energy scale	± 6.4	± 13	± 11	± 9.2	± 2.3	± 5.3	± 4.7	± 3.6
Jet efficiencies	± 1.7	± 5.2	± 2.7	± 2.5	± 0.7	± 2.3	± 1.2	± 1.1
Jet energy resolution	± 0.1	± 4.4	± 2.5	± 1.6	± 0.1	± 2.3	± 1.3	± 0.8
b -tagging efficiency	± 9.2	± 5.6	± 5.1	± 9.3	± 5.0	± 3.1	± 2.9	± 5.0
c -tagging efficiency	± 1.7	± 6.0	± 12	± 2.4	± 1.4	± 5.1	± 10	± 2.1
l -tagging efficiency	± 1.0	± 19	± 5.2	± 2.1	± 0.6	± 11	± 3.0	± 1.1
High p_T tagging efficiency	± 0.6	—	± 0.7	± 0.6	± 0.3	—	± 0.4	± 0.3
$t\bar{t}$: p_T reweighting	—	± 12	± 13	—	—	± 5.1	± 5.8	—
$t\bar{t}$: parton shower	—	± 13	± 16	± 11	—	± 3.6	± 10	± 6.0
$t\bar{t}$ +HF: normalisation	—	—	± 50	± 50	—	—	± 28	± 14
$t\bar{t}$ +HF: modelling	—	—	± 11	± 8.3	—	—	± 8.1	± 7.1
Theoretical cross sections	—	± 6.3	± 6.3	± 6.3	—	± 4.1	± 4.1	± 4.1
$t\bar{t}H$ modelling	± 2.7	—	—	—	± 2.6	—	—	—
Total	± 12	± 32	± 59	± 54	± 6.9	± 9.2	± 23	± 12

important uncertainties affecting mainly the shape of the $t\bar{t} + b\bar{b}$ predictions are the $t\bar{t} + b\bar{b}$ modelling uncertainties due to the scale variations, choice of the recoil scheme in SHERPAOL or the choice of the PDF, which are ranked fourth, sixth, 12th and 15th. The most important one among those is the variation obtained by changing the $t\bar{t} + b\bar{b}$ renormalisation scale to

6. Search for $t\bar{t}(H \rightarrow b\bar{b})$ Production in the Single Lepton Channel

$\mu_R = (m_t m_{b\bar{b}})^{1/2}$, which makes the NN output shape in the signal-enriched regions appear more signal-like. It is only marginally pulled to higher values, but notably constrained by the fit. All constraints on the $t\bar{t} + b\bar{b}$ modelling uncertainties result only in a small reduction on the yield from 8.3% to 7.1% in the most sensitive region, since they mostly affect the shape. The second most important uncertainty is the modelling component originating from the η -intercalibration of the JES (see Sec. 6.4). It is in particular important in the signal-enriched regions at high jet η values and is correlated to the signal strength. Its initial value is pulled negatively and the uncertainty can only be marginally constrained leading to just a small reduction of the impact on $\hat{\mu}$. As a consequence, the post-fit impact is almost as large as the one from the $t\bar{t} + b\bar{b}$ normalisation. Another important JES uncertainty, ranked in 7th position, is the first component of the modelling uncertainties originating from the in-situ calibration. This is the largest JES uncertainty and can be significantly reduced by the fit of the H_T^{had} distribution in the background-dominated regions. The pull of the central value is negligible. Ranked on the 13th position is the JES uncertainty originating from the different response of the calorimeter due to the jet flavour. It is pulled to higher values but not much constrained. In the ($\geq 6j, \geq 4b$) region, the total JES uncertainty can be reduced from values between 9-13% to values between 3-6% on the main background.

The uncertainty on the $t\bar{t} + c\bar{c}$ normalisation is ranked third and the central value is pulled to slightly lower values. Nevertheless the post-fit yield of $t\bar{t} + c\bar{c}$ events is increased in all analysis regions compared to the pre-fit yields. This is only possible due to the interplay with other systematic uncertainties affecting the $t\bar{t} + c\bar{c}$ predictions positively. The component of the parton shower modelling uncertainty acting on the relative $t\bar{t} + c\bar{c}$ contribution is considerably pulled to higher values to correct the predictions to data in the four and five jet regions (see Fig. D.1). This leads to an increase of the total $t\bar{t} + c\bar{c}$ yield in all analysis regions. Because this gain in yield is not consistent with data in the six jet regions with large contributions from $t\bar{t} + c\bar{c}$ events, the $t\bar{t} + c\bar{c}$ normalisation is pulled to negative values to compensate for that. At the same time the uncertainty on the $t\bar{t} + c\bar{c}$ normalisation is considerably constrained. In the most sensitive region it is reduced from 50% to 28% after the fit.

The fifth most important uncertainty is the one on the $t\bar{t}V$ cross section. It is correlated to the signal strength, since $t\bar{t}V$ events produce a similar NN response as signal. The fit has no ability to pull or constrain this uncertainty. In particular the irreducible $t\bar{t}Z$ background with $Z \rightarrow b\bar{b}$ decay has a large impact due to the similar event kinematics. The effect of the $t\bar{t}V$ normalisation uncertainty on $\hat{\mu}$ is notable with $d\mu/d\sigma(t\bar{t}V) = 0.3$.

Further important uncertainties arise from the flavour tagging, due to mistags of light jets (light-tagging 1 & 2) ranked as 8th and 14th as well as the b -tagging (10th position). The large impact of the mistags is explained by the relatively large fraction of $t\bar{t} +$ light events in the four b -tag regions of the analysis. In the three b -tag regions, $t\bar{t} +$ light events enter mainly due to a mistag of a c -jet originating from the hadronically decaying W boson. Since the effect of the mistag uncertainty on the uncertainty of $\hat{\mu}$ arises exclusively from the four b -tag regions with low statistics, the uncertainties cannot be constrained significantly. The effect of the light tagging uncertainties on the yield of the $t\bar{t} +$ light events in the ($\geq 6j, \geq 4b$) region are reduced from 19% to 11%. The b -tag uncertainty is even less constrained and slightly pulled to higher values. The constraints on all b -tagging uncertainties lead to a reduction of the uncertainty on the yield of $t\bar{t} + b\bar{b}$ events in the ($\geq 6j, \geq 4b$) region from 9% to 5%.

Finally, the uncertainties arising from the top quark $t\bar{t}$ and p_T reweighting of the $t\bar{t} + c\bar{c}$ events are ranked 9th and 11th. They have a considerable effect on the shape of the NN and H_T^{had} distributions. The central values are only marginally pulled, but the uncertainties are

slightly reduced. The normalisation uncertainties on the $t\bar{t} + c\bar{c}$ events in the most sensitive region due to the p_T reweighting are reduced from 13% to 6%. The pulls and constraints of all uncertainties considered in the fit can be found in Appendix D.1. All constraints are consistent with the expectations obtained from Asimov pseudo data [227].

6.7.5. Validation of MEM Variables

In order to validate the variables obtained with the MEM, the total predictions after the fit are compared to data. Figure 6.44 presents such comparison for the $\ln \mathcal{L}_{t\bar{t}H}^{\text{sum}}$ distributions in the ($\geq 6j, 3b$) (a) and ($\geq 6j, \geq 4b$) (b) regions.

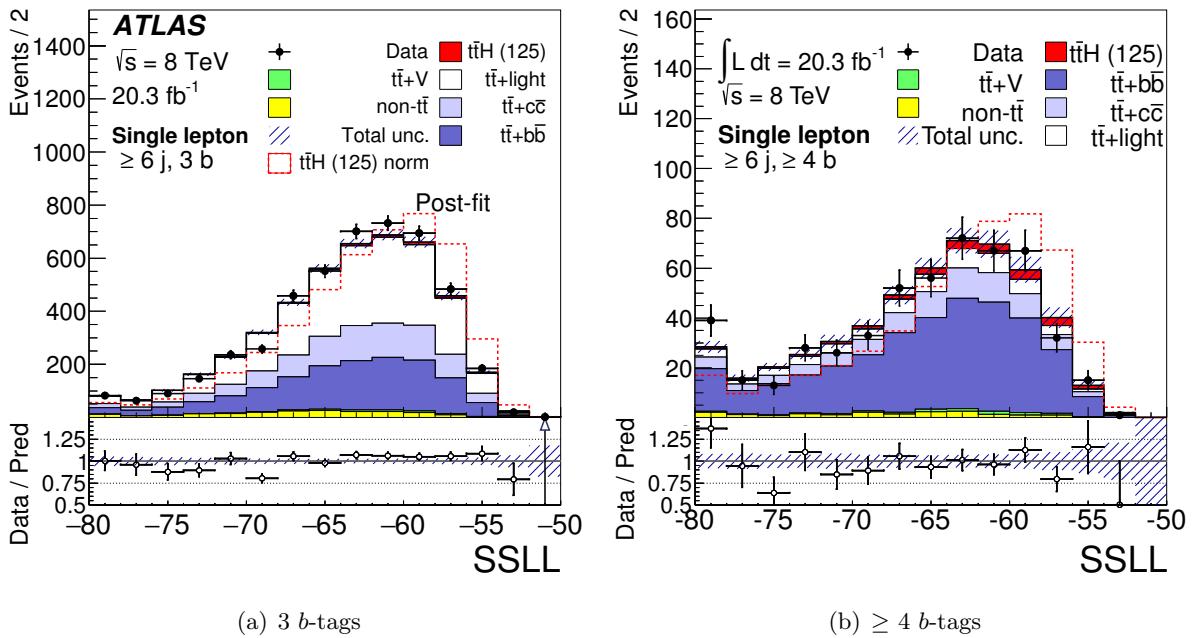


Figure 6.44.: The logarithm of the summed signal likelihoods is shown in the ($\geq 6j, 3b$) (a) and ($\geq 6j, \geq 4b$) (b) region, comparing the total predictions after the fit to data. The $t\bar{t}H$ prediction is shown normalised to the fitted $\hat{\mu}$ (filled red) and to the total background prediction (dashed red). The hashed area represents the post-fit uncertainty on the prediction [71].

The signal prediction is shown twice, once stacked on top of the total background (filled red) normalised to the best fit value $\hat{\mu}$ and once normalised to the total background (dashed red). The agreement with data after the fit is significantly improved in both regions compared to the pre-fit distributions as shown in Figure 6.11. In the three b -tag region, where enough statistics is available, the expectations agree very well with the data except for one outlier in one bin with values around $\ln \mathcal{L} = -69$. Since the uncertainties are small, this deficit in data is now notable. However it is still below 3σ considering statistical and systematic uncertainties. Furthermore, the events in this bin are background-like and thus have a negligible impact on the search for a $t\bar{t}H$ signal. In the region with at least four b -tags, the statistical uncertainty is much larger and the predictions agree very well everywhere with the data within the uncertainties. No visible excess of data is observed in the signal-like regions. The likelihood ratio $D1$ using the summed likelihoods is shown in Figure 6.45 for the total predictions and data.

6. Search for $t\bar{t}(H \rightarrow b\bar{b})$ Production in the Single Lepton Channel

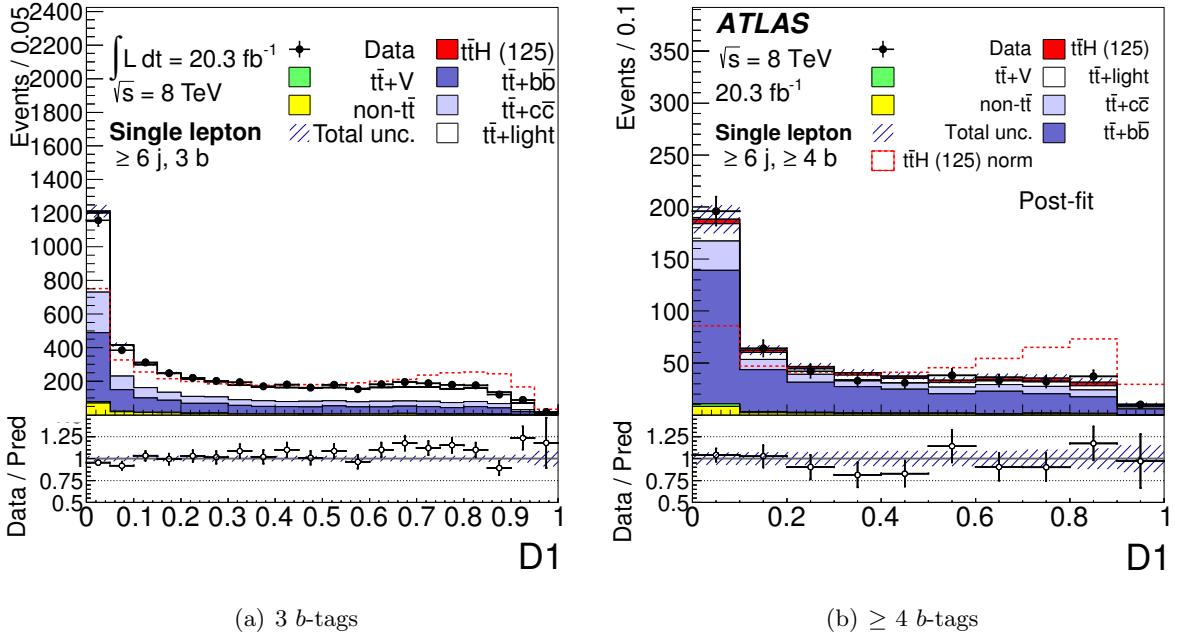


Figure 6.45.: The likelihood ratio $D1$ using the summed likelihoods is shown in the ($\geq 6j, 3b$) (a) and ($\geq 6j, \geq 4b$) (b) region, comparing the total predictions after the fit to data. The $t\bar{t}H$ prediction is shown normalised to the fitted $\hat{\mu}$ (filled red) and to the total background prediction (dashed red). The hashed area represents the post-fit uncertainty on the prediction [71].

In both regions, the predictions agree generally well with the observed data within the uncertainties and no unexpected excess of data events is observed in the signal-like regions at high discriminant values. However, in the three b -tag region a small excess of data events is present, in particular between values of 0.60 and 0.85. This excess is visible in events with exactly one muon and is discussed in more detail in the Appendix D.2. Based on these studies it is concluded, that the excess is a fluctuation in data with a negligible impact on the extracted signal strength. This feature is not present in the $\geq 4 b$ -tag region and very good agreement with data is achieved, especially in the signal sensitive region close to one. Finally, the reconstructed invariant $m_{b\bar{b}}$ distribution of the two jets not originating from the $t\bar{t}$ decay using the best $t\bar{t} + b\bar{b}$ permutation is shown in Figure 6.46.

Although only one assignment permutation is used for the reconstruction of this distribution, the predictions agree very well with data with high probabilities in the statistical tests. In the regions around $m_H = 125$ GeV, where a Higgs boson mass peak is expected to be visible if an excess is observed, the data agrees very well with the expectations. No bias is observed from the MEM used for event reconstruction. More post-fit distributions of the MEM variables and of the most important NN input variables can be found in Appendix D.3 and Appendix D.4, respectively.

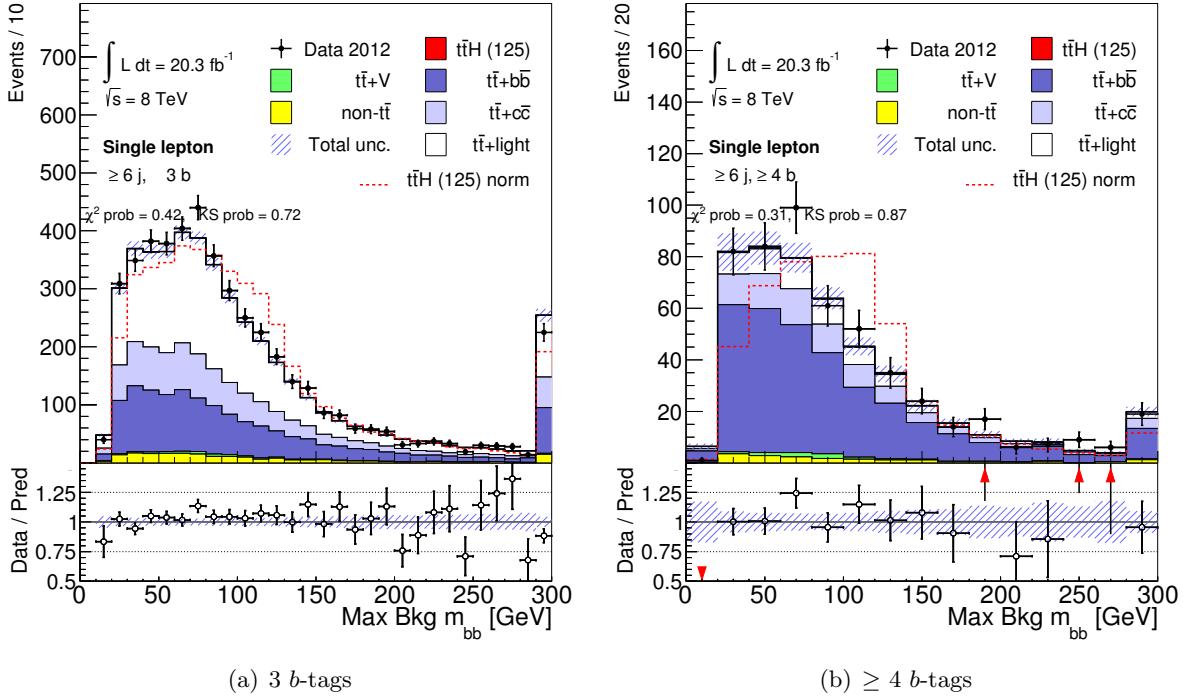


Figure 6.46.: The reconstructed invariant mass $m_{b\bar{b}}$ using the $t\bar{t} + b\bar{b}$ likelihood is shown in the ($\geq 6j, 3b$) (a) and ($\geq 6j, \geq 4b$) (b) region, comparing the total predictions after the fit to data. The ttH prediction is shown normalised to the fitted $\hat{\mu}$ (filled red) and to the total background prediction (dashed red). The hashed area represents the post-fit uncertainty on the prediction.

6.8. Combined Results

6.8.1. Combination with Dilepton Analysis

The single lepton analysis is combined with the dilepton analysis, which pursues a very similar strategy [71]. Instead of one lepton, it requires exactly two opposite sign leptons (electron or muon) with slightly different object identification in case of the electron to increase the acceptance. Additional requirements on the scalar sum of the transverse energy of leptons and jets, H_T , and on the invariant mass of the two leptons are applied to avoid a contamination from vector boson, J/ψ and Υ events. The selections of the two analysis are designed to be orthogonal. A similar classification into analysis regions according to the jet and b -tag multiplicities is applied leading to six analysis regions. In the three most signal sensitive regions a NN using NEUROBAYES is employed, but no MEM has been exploited so far. In the background-dominated regions the scalar sum of the transverse energy of all jets and leptons H_T is used as discriminating variable for the profile likelihood fit. Apart from a few exceptions, the same systematic uncertainties are evaluated and the same fit model is applied. Most of the systematic uncertainties are fully correlated between the two analysis. This allows for a straight-forward combination which includes 15 analysis regions of the two analyses and allows to learn the most from the two orthogonal datasets about the background predictions and the systematic uncertainties.

6.8.2. The Signal Strength

The profile likelihood fit is performed separately and in combination of the two analysis channels. The best estimators for the signal strength are summarised in Figure 6.47.

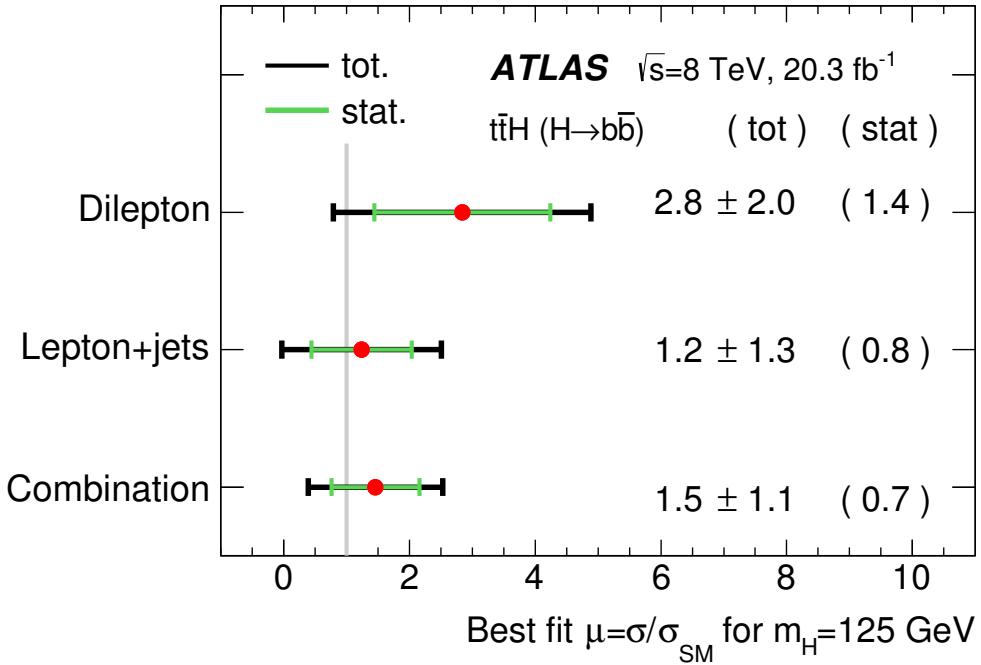


Figure 6.47.: The fitted values of the signal strength and their uncertainties for the individual analyses as well as their combination, assuming $m_H = 125$ GeV are shown in comparison with the SM of $\mu = 1$. The statistical uncertainty (green) and total uncertainty (black) on the signal strength are separately shown as error bars [71].

The individual measurements are consistent with each other and the measured signal strengths are compatible with the SM expectation. However, in the dilepton analysis more events than expected from the SM predictions are observed, leading to a signal strength of $\hat{\mu} = 2.8 \pm 2.0$. This excess is insignificant and agrees within 1σ with the SM due to the large total uncertainty. The combination yields a signal strength of $\hat{\mu} = 1.5 \pm 1.1$. Its central value and uncertainty is clearly driven by the single lepton analysis, but the combination with the dilepton analysis helps to reduce the uncertainty by 0.2 and shifts the minimum of the likelihood to a higher signal strength. The measurement of the signal strength is dominated by the statistical component of the uncertainty. However, a large part of this component is arising from the measurement of the NPs representing the systematic uncertainties. Although the effect of the systematic uncertainties is largely reduced by the profile likelihood fit, the statistical uncertainty on $\hat{\mu}$ is increased due to correlations between μ and the NPs resulting in an enlarged contour of the likelihood profile. The expected uncertainty of the combination of ± 1.1 obtained with pseudo data agrees with the observed one. The pull and constraints on the NPs compared for both analyses separately and the combination are shown in Figure D.1 of the Appendix. They agree generally well between the two analysis channels. The fit values of the combination are dominated by the single lepton channel.

6.8.3. Setting Limits

Because the measured signal strength shows no significant excess above the background, hypothesis tests are performed based on a frequentist approach in order to determine the significance of the excess and to set upper limits on the observed cross section relative to the SM cross section. The *test statistic* q_μ for this hypothesis tests is defined based on the profile likelihood ratio [227]

$$q_\mu(\mathbf{x}) = -2 \ln \left(\frac{\mathcal{L}(\mathbf{x} | \mu, \hat{\theta}_\mu)}{\mathcal{L}(\mathbf{x} | \hat{\mu}, \hat{\theta})} \right) \quad \text{with } 0 \leq \hat{\mu} \leq \mu , \quad (6.8)$$

where \mathcal{L} is the likelihood function of the profile likelihood fit and \mathbf{x} represents the data or pseudo data. The $\hat{\mu}$ and $\hat{\theta}$ are the parameter values, which maximise the global likelihood function and the $\hat{\theta}_\mu$ are the profiled values of the NPs θ , which maximise \mathcal{L} at a given value μ . Since physics dictates a positive signal strength and the exclusion limit should not be smaller than the best estimator, the signal strength μ is constrained to be $0 \leq \hat{\mu} \leq \mu$. The test statistic is used to asses the validity of the background-only hypothesis with $\mu = 0$ given the distributions of the observed data as well as to make statistical inferences about μ , such as upper limits using the CL_s method [227–229] as implemented in the `Roofit` package [226, 230]. In order to perform a hypothesis test the p.d.f.s of the two alternative hypotheses $f(q_\mu | 0, \hat{\theta}_0^{\text{obs}})$ and $f(q_\mu | \mu, \hat{\theta}_\mu^{\text{obs}})$ need to be constructed [231]. For the construction, the NPs describing best the experimental data under the background-only hypothesis $\hat{\theta}_0^{\text{obs}}$ and under the signal+background hypothesis $\hat{\theta}_\mu^{\text{obs}}$ need to be determined from the profile likelihood fit. While with alternative test statistics, the p.d.f.s are determined using computing intensive toy MC pseudo experiments, this test statistic has the important advantage that its p.d.f.s can be approximated by asymptotic formulae based on Wilks and Wald theorems [232, 233] using the Asimov data set [227]. The Asimov data set is a single representative of the ensemble reflecting the p.d.f., in which all parameters perfectly match the expected background, leading to zero pulls of the NPs. With the obtained p.d.f.s of the test statistic the p-values for the signal+background hypothesis μ

$$p_\mu = P(q_\mu \geq q_\mu^{\text{obs}} | \mu s + b) = \int_{q_\mu^{\text{obs}}}^{\infty} f(q_\mu | \mu, \hat{\theta}_\mu^{\text{obs}}) dq_\mu \quad (6.9)$$

and for the background-only hypothesis

$$1 - p_b = P(q_\mu \geq q_\mu^{\text{obs}} | b) = \int_{q_0^{\text{obs}}}^{\infty} f(q_\mu | 0, \hat{\theta}_0^{\text{obs}}) dq_\mu \quad (6.10)$$

can be calculated for a given observed test statistic q_μ^{obs} defined by the signal strength μ under test. The CL of the signal hypothesis is then given by

$$CL_s(\mu) = \frac{CL_{s+b}}{CL_b} = \frac{p_\mu}{1 - p_b} . \quad (6.11)$$

Accordingly, an upper limit on μ with a CL of α can be determined by adjusting μ until $CL_s = 1 - \alpha$. Alternatively, the CL for a given value of μ can be derived, for instance, the SM

6. Search for $t\bar{t}(H \rightarrow b\bar{b})$ Production in the Single Lepton Channel

expectation $\mu = 1$ can be excluded with a CL of 95%, if $CL_s \leq 0.05$. Finally, the p-value can be converted into a significance Z in units of standard deviations by

$$Z = \Phi^{-1}(1 - p) , \quad (6.12)$$

where Φ^{-1} is the inverse cumulative distribution function of the Gaussian distribution. The observed and the expected median upper limits on μ with a CL of 95% for the background-only hypothesis with $\mu = 0$ and the SM hypothesis $\mu = 1$ of a SM Higgs boson with $m_H = 125$ GeV are summarised in Table 6.7 for the two analyses channels and the combination.

Table 6.7.: Observed and expected (median, for the background-only hypothesis) upper limits on $\sigma(t\bar{t}H)$ relative to the SM prediction with a 95% CL for the individual analysis channels as well as their combination assuming $m_H = 125$ GeV. The 68% and 95% confidence intervals around the expected limits under the background-only hypothesis are also provided, denoted by $\pm 1\sigma$ and $\pm 2\sigma$, respectively. The expected (median) 95% CL upper limits assuming the SM prediction for $\sigma(t\bar{t}H)$ are shown in the last column.

Channel	Observed	-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$	Median ($\mu = 1$)
Single Lepton	3.6	1.4	1.9	2.6	3.7	4.9	3.6
Dilepton	6.7	2.2	3.0	4.1	5.8	7.7	4.7
Combination	3.4	1.2	1.6	2.2	3.0	4.1	3.1

The $\pm 1\sigma$ and $\pm 2\sigma$ uncertainty bands on the median for the background-only hypothesis are determined from the crossing of the cumulative probability distribution with the corresponding quantiles without recalculation of the p.d.f.. The obtained results are additionally visualised in Figure 6.48.

In the single lepton channel a cross section of 3.6 times the SM cross section is excluded by the measurement, while an exclusion of 2.6 and 3.6 times $\sigma(t\bar{t}H)$ assuming only background and the SM cross section is expected, respectively. Thus the observed limit agrees perfectly with the SM expectation assuming that the Higgs boson is produced in association with top quarks. However, the uncertainty bands of the median using the background-only hypothesis are large and the observed value lies within the 1σ band. The dilepton channel is much less sensitive to the SM expectations leading to an observed upper limit on μ of 6.7, while 4.1 and 4.7 is expected for the background-only and the SM signal hypothesis, respectively. The upward fluctuation of the observed value agrees within 2σ with the background-only hypothesis. In the combination a signal 3.4 times larger than predicted by the SM is excluded at 95% CL, while 2.2 and 3.1 times is expected without and with a SM Higgs boson, respectively. The observed value lies in the 2σ uncertainty band of the background-only hypothesis. The upper limits obtained in the combination are clearly driven by the single lepton analysis channel. Owing to the combination with the dilepton analysis, the expected median of the background-only hypothesis is reduced from 2.6 to 2.2, which corresponds to a relative improvement of 15%. Similarly, the uncertainty bands on the median are reduced. The observed (expected) p-value of the excess given the background-only hypothesis is 8% (15%), which corresponds to an observed (expected) significance of the signal of 1.4 (1.1) standard deviations. The measurement is thus not sensitive enough to make any statement about the SM expectations.

Figure 6.49 shows the event yield after the combined fit as a function of $\log_{10}(S/B)$.

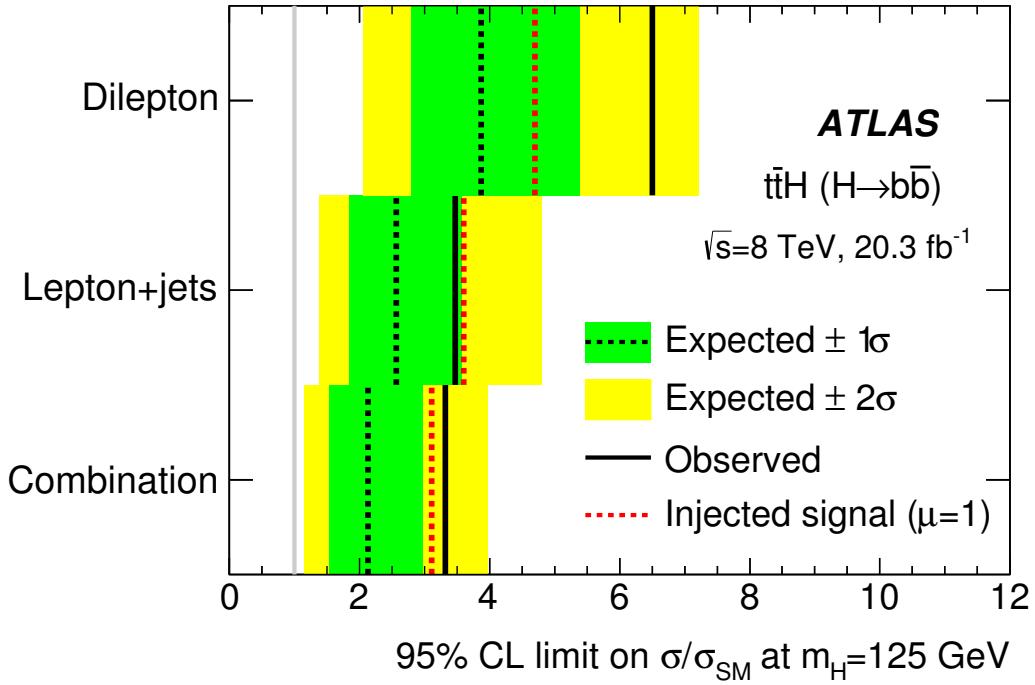


Figure 6.48.: The 95% CL upper limits on $\sigma(t\bar{t}H)$ relative to the SM prediction, $\sigma/\sigma_{\text{SM}}$, for the individual channels as well as their combination. The observed limits (solid lines) are compared to the expected (median) limits under the background-only hypothesis and under the signal-plus-background hypothesis assuming the SM prediction for $\sigma(t\bar{t}H)$ and pre-fit prediction for the background. The surrounding shaded bands correspond to the 68% and 95% confidence intervals around the expected limits under the background-only hypothesis, denoted by $\pm 1\sigma$ and $\pm 2\sigma$, respectively [71].

The signal-over-background ratios of each event are calculated from the bins of the post-fit distributions, H_T^{had} , H_T , and the NN output of the profile likelihood fit. The total background prediction (black line) is compared to the observed data events. In all background dominated bins, the post-fit background predictions agree very well with the data. This demonstrates that the profile likelihood fit corrects perfectly the predictions according to the observed data in the background-dominated phase space. Only in bins with very high S/B ratio, the data disagrees with the background-only prediction and thus gives room for a possible signal. The bin with the highest S/B ratio contains the last two bins of the NN output in the most signal-rich regions of the single lepton ($\geq 6j, \geq 4b$) and dilepton ($\geq 4j, \geq 4b$) analyses. The $t\bar{t}H$ signal normalised to the best estimator of $\hat{\mu} = 1.5$ is shown stacked on the histogram bins (filled red) of the background prediction and in the ratio below added to the total background prediction (red line). As expected, the data agrees very well with this signal-plus-background hypothesis. In addition the $t\bar{t}H$ signal hypothesis with a signal strength of $\mu = 3.4$ (dashed orange) is displayed, which can be excluded with a 95% CL from the observed data.

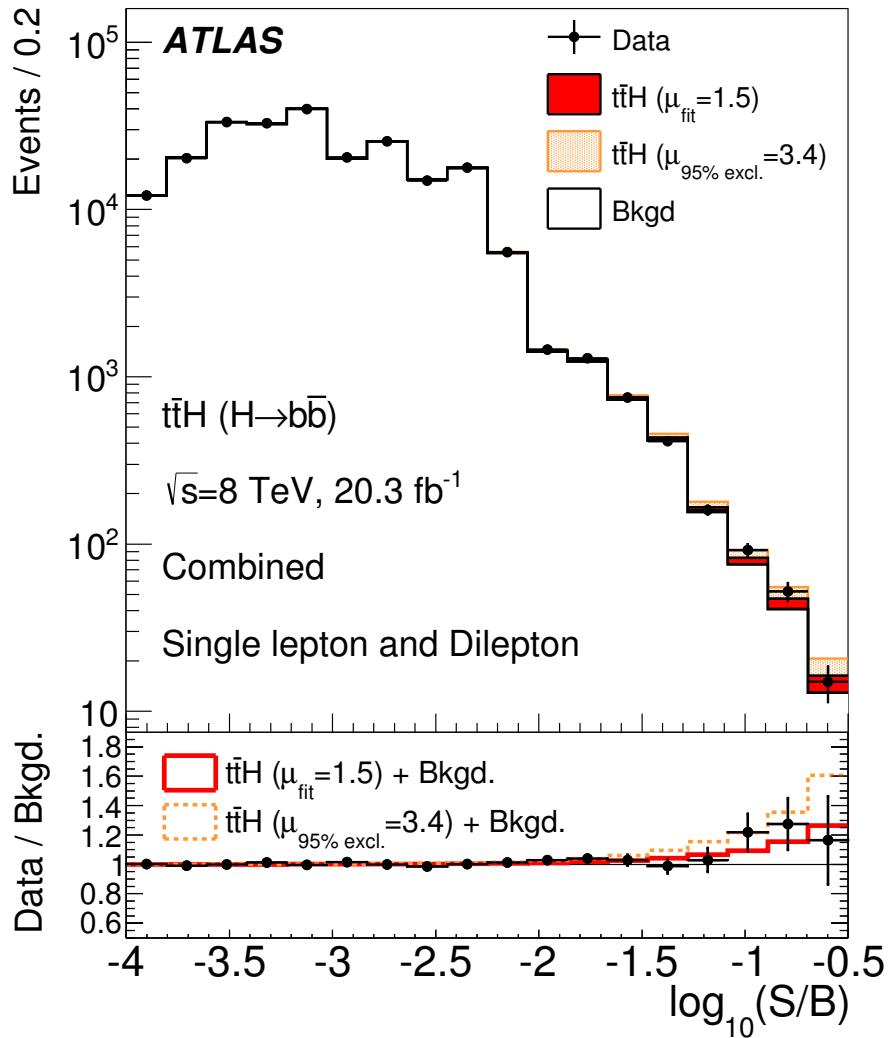


Figure 6.49.: Event yields as a function of $\log_{10}(S/B)$, where S (signal yield) and B (background yield) are taken from the H_T^{had} , H_T , and NN output bin of each event. The predicted background is obtained from the global signal-plus-background fit. The $t\bar{t}H$ signal is shown both for the best fit value ($\mu = 1.5$) and for the upper limit at 95% CL ($\mu = 3.4$) [71].

6.8.4. Impact of the MEM on the Results

In order to quantify the impact of the inclusion of the MEM on the analysis, the results are compared to the previous results obtained without the MEM [205]. Although other small changes were introduced in addition, the improvements in the results are mainly driven by the improvement in separation due to the MEM. In Table 6.8 the observed signal strength and the expected 95% CL upper limit for the background-only hypothesis are shown for the analysis channels separately and their combination.

With the inclusion of the MEM in the single lepton analysis, the uncertainty on the signal strength is reduced from 1.6 to 1.3, which corresponds to a relative improvement of 19%. The observed central value is shifted to a slightly lower value matching better the SM prediction.

Table 6.8.: The observed signal strength and the median of the expected upper limit on the signal strength with a CL of 95% for the background-only hypothesis are compared to the results obtained without the MEM [205].

Channel	Signal Strength		Expected Limit $\mu = 0$	
	no MEM	with MEM	no MEM	with MEM
Single Lepton	1.3 ± 1.6	1.2 ± 1.3	3.1	2.6
Dilepton	2.9 ± 2.3	2.8 ± 2.0	4.3	4.1
Combination	1.7 ± 1.4	1.5 ± 1.1	2.6	2.2

This effect, however, does not indicate the improvement. Due to minor changes in the dilepton channel, mainly related to the background modelling, the uncertainty on the signal strength is reduced by 13%, although no MEM was used. In the combination, both improvements result in a relative reduction of the uncertainty of 21%.

Similarly, the expected upper limit on the signal strength of the background-only hypothesis is reduced from 3.1 to 2.6 in the single lepton corresponding to a relative improvement of 16%. This coincides with the improvement obtained from the combination with the dilepton channel when no MEM was used. With the additional reduction of the expected limit in the dilepton channel by 5%, the combined limit is reduced from 2.6 to 2.2. This corresponds to a relative improvement of 18%. Given that the analysis was already using powerful methods like a highly optimised NN and a profile likelihood fit making use of the information from a large phase space region, the improvement accomplished at that level due to the inclusion of the MEM is a remarkable achievement.

Conclusions and Outlook

7.1. Summary and Conclusion

The SM of particle physics is currently the most complete and elegant picture of the structure of the Universe. It describes all constituents of matter and combines three of the four fundamental forces in a coherent way. In the context of the SM, elementary particles acquire their mass by the mechanism of the electroweak symmetry breaking predicting a massive scalar particle - the Higgs boson. However, several observations in astrophysics and particle physics cannot be explained by the SM, accordingly it cannot be the final picture. The first run of the LHC being the world's most powerful particle accelerator turned out to be a great success. It allowed for probing the SM of particle physics at the new energy frontier. Many rare processes were observed for the first time and particle properties were measured with unprecedented precision. No measurements contradict the predictions of the SM so far. The greatest success of the LHC was the discovery of the Higgs boson in July 2012 by the ATLAS and CMS collaborations. The discovery opened up a completely new field of particle physics and many measurements of the Higgs boson properties have been performed already with a remarkable precision. Precision measurements of the Higgs boson properties are expected to show the direction of the search for new physics necessary to extend the SM. Hence it is of large importance to measure all of them, but in particular the couplings to other particles, which are expressed in the production and decay rates of the Higgs boson. The measurement of the $t\bar{t}H$ production cross section with a subsequent Higgs boson decay into bottom quarks addresses the measurement of the coupling to heavy fermions in the production as well as in the decay. A direct and precise measurement of the top quark Yukawa coupling is of particular interest. In the SM it is expected to be close to unity and a comparison to indirect measurements would be a very robust and model independent test for new particles.

A search for this process has been presented in this thesis using the $\sqrt{s} = 8$ TeV dataset recorded with the ATLAS detector. The search is focussed on the single lepton decay channel of the top quark pair and is optimised for the $H \rightarrow b\bar{b}$ decay. Because large systematic uncertainties on the background predictions prevent a simple and direct search for the Higgs boson, sophisticated analysis techniques have to be applied. Events are categorised into nine analysis regions according to the number of observed jets and b -jets, of which six are dominated by background processes and three are sensitive to the signal. The resulting analysis regions yield very different compositions of the processes, which can be exploited when comparing the

7. Conclusions and Outlook

predictions to data in a profile likelihood fit. In order to improve the separation between signal and the main $t\bar{t} + b\bar{b}$ background the MEM has been developed and applied in the most signal sensitive regions with at least six jets. The MEM makes use of the theoretical calculation of the transition matrix element of processes in order to assign a probabilistic weight to an observed event based on the kinematic information. The method is very computationally intensive and many approximations and optimisations are needed to make the calculation of the probabilities feasible for a large number of events. Probabilities are calculated using the full LO theoretical description of the $t\bar{t}H$ process as signal hypothesis and the $t\bar{t} + b\bar{b}$ process as a prototype of the background hypothesis. The two output probabilities are used to calculate powerful discriminating variables separating signal from background as well as to perform a full reconstruction of the observed event. The most powerful discriminating variable is the Neyman-Pearson likelihood ratio $D1$, which uses the probabilities of both processes simultaneously. Another robust variable is the logarithm of the summed signal likelihood $\ln \mathcal{L}_{t\bar{t}H}^{\text{sum}}$ representing how signal-like a probed event is. In most of the background-dominated regions H_T^{had} - a simple kinematic variable sensitive to the main systematic uncertainties - is used as input distribution to the profile likelihood fit. In the signal-enriched regions a NN is employed combining several kinematic variables and the MEM discriminants $D1$ and $\ln \mathcal{L}_{t\bar{t}H}^{\text{sum}}$ in regions with six jets. The use of the MEM in the NN leads to a relative improvement of 24% in separation of the NN output in the most signal sensitive regions with $D1$ being the most sensitive variable. This is a great achievement considering that the kinematic NN has been highly optimised using more than 300 kinematic variables. By performing the profile likelihood fit simultaneously in all analysis regions, the systematic uncertainties on the background are significantly reduced and the nominal predictions are improved. The signal strength μ for a Higgs boson mass of $m_H = 125$ GeV is extracted from the profile likelihood fit to be

$$\mu = \sigma_{t\bar{t}H}^{\text{obs}} / \sigma_{t\bar{t}H}^{\text{SM}} = 1.2 \pm 1.3 . \quad (7.1)$$

Despite the great constraining power of data in the background-dominated regions, the measurement is still dominated by systematic uncertainties. The largest impact on the uncertainty on μ arises from the uncertainty on the $t\bar{t} + b\bar{b}$ normalisation, although it was reduced by more than half from the initial value of 50%. Because the measurement is still consistent with the background-only hypothesis, an observed (expected) upper limit on the signal strength with a 95% CL of

$$\mu < 3.6 \quad (2.6) \quad (7.2)$$

is set. The inclusion of the MEM in the analysis reduces the uncertainty on the signal strength by 19% and improves the upper limit by 16%. The single lepton channel has been combined with the dilepton channel of the $t\bar{t}$ decay resulting in a signal strength of $\mu = 1.5 \pm 1.1$ and an observed upper limit on the signal strength of 3.4, while 2.2 was expected for the background-only hypothesis at a 95% CL.

Comparison with other Searches for $t\bar{t}H$ Production

Both LHC experiments have performed searches for the $t\bar{t}H$ production in various decay modes of the Higgs boson and have combined the results [70, 71, 74, 88, 180, 206]. Because no significant excess above the background-only expectation has been observed so far, the analyses set upper limits on the signal strength with a 95% CL. The expected (observed) upper limits assuming

the background-only hypothesis for a Higgs boson mass around 125 GeV are compared between the two experiments in Table 7.1.

Table 7.1.: The expected and observed upper limits on the signal strength of $t\bar{t}H$ production assuming the background-only hypothesis for a Higgs boson mass around 125 GeV set by the ATLAS and CMS experiments with the Run I data [70, 71, 74, 88, 206]. *In the single lepton channel of the $H \rightarrow b\bar{b}$ search, CMS has published an analysis using the MEM, which has not been combined with the remaining analyses but yields a better limit [180].

Higgs Decay	Channel	Expected (Observed) Limit	
		ATLAS	CMS
$H \rightarrow b\bar{b}$	single lepton	2.6 (3.6)	4.2 (5.5)*
	dilepton	4.1 (6.7)	6.7 (7.0)
	combined	2.2 (3.4)	3.5 (4.1)
$H \rightarrow \gamma\gamma$	leptonic	6.6 (10.7)	6.8 (8.2)
	hadronic	10.1 (9.0)	10.7 (8.0)
	combined	4.9 (6.7)	4.7 (7.4)
$H \rightarrow WW/ZZ/\tau\tau$	2ℓ	3.9 (6.7)	3.4 (9.0)
	3ℓ	3.8 (6.8)	4.1 (7.5)
	4ℓ	15 (18)	8.8 (6.8)
	$2\tau_{\text{had}}$	18 (13)	14.2 (13.0)
	$2\ell 1\tau_{\text{had}}$	8.4 (7.5)	-
	combined	2.4 (4.7)	2.4 (6.6)
Combination		1.4 (3.2)	1.7 (4.5)

The expected limits using the background-only hypothesis represent the sensitivity of the search. The searches are categorised in three decay channels $H \rightarrow b\bar{b}$, $H \rightarrow \gamma\gamma$ and $H \rightarrow$ multilepton. The latter includes all decays with at least two charged leptons in the final state and investigates mainly the $H \rightarrow WW^{(*)}$, $H \rightarrow ZZ^{(*)}$ and $H \rightarrow \tau\tau$ decay modes with different fractional contributions. Recently, the CMS collaboration has published a search for $H \rightarrow b\bar{b}$ in the single lepton channel which also exploits the discrimination power of the MEM [180]. This search is tailored for the application of the MEM and calculates the probabilities only for the most signal-like events selected using b -tagging information. This leads to a reduction of MC statistics by three orders of magnitude compared to the presented analysis. As a consequence, the MEM variables could not be utilised in a NN and the analysis could not be combined with the other CMS searches. The expected upper limit of 3.3 is significantly higher than the one obtained in the presented analysis. The combined search in the single and dilepton channel by CMS employs a MVA. The resulting upper limit of 3.5 is considerably higher than the limit of 2.2 obtained by the presented search of the ATLAS collaboration.

The search of the $t\bar{t}H$ production in the $H \rightarrow \gamma\gamma$ mode divides the analysis in leptonic and hadronic channel according to the top quark decay. Similar to the discovery in the $H \rightarrow \gamma\gamma$ channel, a search for a resonance in the invariant diphoton mass distribution is performed. The requirement of two photons results in a clean environment. However due to the tiny branching ratio, the search is statistically limited. Both ATLAS and CMS obtain very similar

7. Conclusions and Outlook

expected combined limits of 4.9 and 4.7, respectively. The search for $H \rightarrow$ multilepton is split into five sub-channels according to the observed number of leptons (electrons or muons) and τ leptons. Although the rate is relatively low, the high signal purity allows for a cut and count analysis. Both collaborations obtain an identical expected upper limit of 2.4. In the final combination of all decay channels searching for the $t\bar{t}H$ process, ATLAS obtains an expected (observed) upper limit of 1.4 (3.2), which corresponds to a significance of 2.4σ for an excess above the background-only hypothesis. The $t\bar{t}H$ combination of CMS collaboration yields an expected (observed) limit of 1.7 (4.5), which represents an excess above the background-only expectation of 3.4σ . With the obtained sensitivities both experiments reach a regime where they become sensitive to the SM production cross section of the $t\bar{t}H$ process. The two LHC collaborations are currently working on combining their results.

In conclusion, the analysis presented in this thesis has performed the single most sensitive search for $t\bar{t}H$ production. The obtained result is highly competitive and considerably more sensitive than the comparable CMS analysis. The $H \rightarrow b\bar{b}$ analysis of ATLAS is the most sensitive decay channel in searches for $t\bar{t}H$ production. The presented search significantly contributes to the global combination of $t\bar{t}H$ searches in ATLAS resulting in an upper limit with the background-only hypothesis close to the SM expectation. However, no significant excess above the background-only hypothesis has been observed so far.

7.2. Outlook

With the LHC Run I combinations being already close to the SM sensitivity, a discovery of $t\bar{t}H$ production is expected in Run II, if the process is produced according to the SM expectations. At the higher energies of $\sqrt{s} = 13 - 14$ TeV, the $t\bar{t}H$ cross section increases by a factor of five, while the production rates of the main background processes increase less. Due to the much higher energy, boosted topologies of the $t\bar{t}H$ process will become interesting, because physics and combinatorial backgrounds can be much easier suppressed. Preliminary studies of $t\bar{t}H$ predictions in the boosted regime applied to the $H \rightarrow b\bar{b}$ decay show that a 5σ discovery might be possible with the first 100 fb^{-1} of data [234]. The measurements of the $t\bar{t}H$ production rate with decays to $H \rightarrow WW^{(*)}$ and $H \rightarrow \gamma\gamma$ are expected to be still statistically limited with up to 300 fb^{-1} [235, 236]. However, these channels will eventually enable a measurement of the top quark Yukawa coupling with an uncertainty of 5-10% at the LHC [235, 236]. All these studies are rather conservative estimates and the results of Run I have shown, that analysers will likely outperform the expectations.

Concerning the presented $t\bar{t}H$ analysis, the most crucial point for any increase in sensitivity is to improve the knowledge of the $t\bar{t} + b\bar{b}$ background. The Run II data is expected to allow for a fiducial or differential cross section measurements of the $t\bar{t} + b\bar{b}$ process. Although efforts are ongoing in this direction, only a relatively simple measurement of the cross section ratio of $t\bar{t} + b\bar{b}$ to $t\bar{t} + jj$ events has been performed by CMS [237]. These measurements are very important input for improving the MC generators and theoretical calculations in the future. There are also studies ongoing to measure the $t\bar{t}H$ production in the allhadronic and single lepton + τ lepton final state of the $t\bar{t}$ decay. A combination with these results will potentially improve the expected limits. Recently, the MEM was implemented in the dilepton channel in the presented framework of MEMTOOL. Preliminary studies at $\sqrt{s} = 8$ TeV show that the separation provided by the likelihood ratio $D1$ is significantly better than in the single lepton channel. Pioneer studies conclude that the dilepton decay of $t\bar{t}$ will eventually outperform the

sensitivity of the single lepton channel [179]. Although a MEM in the dilepton channel could not be realised in time for the Run I publication, everything is being prepared for Run II. However, to perform MEM analyses in the single lepton and dilepton channels, it is critical to further reduce the runtime of the calculations. With increasing MC statistics and less time to perform the analysis in Run II, the calculation of the probabilities might not be feasible any more. Although the implemented permutation ranking could be more exploited, there is limited potential for further optimisations. The best solution for the huge computing time would be to create look-up tables by sampling the relevant phase space and calculate the process probabilities in advance. Because the Run I results provided a phase space sampling of almost a billion 6-dimensional phase space points, this approach has already been tried. Since the kinematics of the MC events have been used instead of a systematic sampling of the phase space, the density distribution of the phase space points is not optimal. In order to obtain a smooth transitions between phase space points, the results need to be interpolated, which is not trivial in the 6-dimensional space. Furthermore, storage issues arise if the results are translated into 6-dimensional histograms with one billion entries. Hence an optimal parametrisation with reduced dimensions is required to create the look-up tables. Initial tests using the masses of the intermediate states and variations of the jet energies as parameters did not lead to closure of the likelihood distributions obtained with the look-up tables and the full integrations. This might not be a serious issue if the MC simulation and data behaves the same way, however it degrades the separation power obtained with the MEM. Nevertheless, this seems to be the most promising approach for future MEM analysis in $t\bar{t}H$, but it certainly requires more studies. Once the computing time issues are overcome, additional matrix element of background processes such as the $t\bar{t}V$ production could be added to further improve the separation. So far only analysis regions with six jets are exploited, where the number of measured objects matches the number of final states. With the MEM, the five jet regions can also be investigated as carried out by the CMS collaboration [180]. Because missing quantities can simply be marginalised by the integration, missing objects are not a serious issue but result in larger integration times. A transformation to the intermediate states might even make the integration of the missing quantities unnecessary. Interestingly, an event reconstruction can still be performed for this under-constrained systems which is not possible for likelihood-based reconstruction algorithms.

Once a signal in $t\bar{t}H$ is established, the MEM could be used for a Higgs boson mass measurement as suggested in Section 5.9. For this analysis, the probabilities of the $t\bar{t}H$ hypothesis have actually been calculated with a Higgs boson mass parameter variation between 110-140 GeV. Furthermore, MC simulations of the $t\bar{t}H$ process with five different Higgs boson masses have been probed in order to calibrate the measurement. These studies go beyond the scope of this thesis but yield quite interesting results. A Higgs boson mass measurement in the $H \rightarrow b\bar{b}$ decay, cannot compete with the precisions obtained in the $H \rightarrow \gamma\gamma$ or $H \rightarrow ZZ^{(*)} \rightarrow 4\ell$ decays. Nevertheless, a measurement of the Higgs boson mass is still interesting and a stronger Higgs boson mass dependence of this analysis might be desirable.

Finally, there is a large variety of applications of the MEM. Since MEMTOOL has a modular and mostly process independent design, its computing framework can be used for any kind of MEM analysis, whether it be precision measurements or searches for new physics.

Bibliography

- [1] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321.
- [2] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508.
- [3] P. W. Higgs, *Broken Symmetries, Massless Particles and Gauge Fields*, Phys. Lett. **12** (1964) 132.
- [4] G. Guralnik, C. Hagen, and T. Kibble, *Global Conservation Laws and Mass-less Particles*, Phys. Rev. Lett. **13** (1964) 585.
- [5] D0 Collaboration, *A Precision Measurement of the Mass of the Top Quark*, Nature **429** (2004) 638–642, [arXiv:0406031 \[hep-ph\]](https://arxiv.org/abs/hep-ph/0406031).
- [6] CDF Collaboration, *Precision measurement of the top-quark mass from dilepton events at CDF II*, Phys. Rev. D **75** (2007) 031105.
- [7] CDF Collaboration, *Search for a Standard Model Higgs Boson in $WH \rightarrow \ell\nu b\bar{b}$ in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV*, Phys. Rev. Lett. **103** (2009) 101802.
- [8] CDF Collaboration, *Search for Standard Model Higgs Boson Production in Association with a W Boson Using a Matrix Element Technique at CDF in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV*, Phys. Rev. D **85** (2012) 072001, [arXiv:1112.4358 \[hep-ex\]](https://arxiv.org/abs/hep-ex/1112.4358).
- [9] S. L. Glashow, *Partial symmetries of weak interactions*, Nucl. Phys. **22** (1961) 579.
- [10] S. Weinberg, *A model of leptons*, Phys. Rev. Lett. **19** (1967) 1264–1266.
- [11] A. Salam, *Weak and electromagnetic interactions*, Proc. of the 8th Nobel Symposium (1969) 367.
- [12] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

Bibliography

- [13] CMS Collaboration, *The CMS experiment at the CERN LHC*, JINST **3** (2008) S08004.
- [14] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10** (1963) 531.
- [15] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49** (1973) 652.
- [16] Particle Data Group, *Review of Particle Physics*, Chin. Phys. **C38** (2014) 090001.
- [17] H. Weyl, *The Classical Groups: Their Invariants and Representations*, Westview Press (1995).
- [18] S. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579.
- [19] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264.
- [20] S. Glashow, J. Iliopoulos, and L. Maiani, *Weak Interactions with Lepton-Hadron Symmetry*, Phys. Rev. **D2** (1970) 1285.
- [21] H. Georgi and S. L. Glashow, *Unified Weak and Electromagnetic Interactions without Neutral Currents*, Phys. Rev. Lett. **28** (1972) 1494.
- [22] H. D. Politzer, *Reliable Perturbative Results for Strong Interactions?*, Phys. Rev. Lett. **30** (1973) 1346.
- [23] H. D. Politzer, *Asymptotic freedom: An approach to strong interactions*, Phys. Rep. **14** (1974) 129.
- [24] D. J. Gross and F. Wilczek, *Ultraviolet Behavior of Nonabelian Gauge Theories*, Phys. Rev. Lett. **30** (1973) 1343.
- [25] S. Weinberg, *The Making of the standard model*, Eur. Phys. J. **C34** (2004) 5, [arXiv:0401010 \[hep-ph\]](https://arxiv.org/abs/hep-ph/0401010).
- [26] G. 't Hooft, *Renormalizable Lagrangians for Massive Yang-Mills Fields*, Nucl. Phys. **B35** (1971) 167.
- [27] G. 't Hooft and M. Veltman, *Regularization and Renormalization of Gauge Fields*, Nucl. Phys. **B44** (1972) 189.
- [28] G. 't Hooft and M. Veltman, *Combinatorics of gauge fields*, Nucl. Phys. **B50** (1972) 318.
- [29] A. Salam and J. C. Ward, *Electromagnetic and weak interactions*, Phys. Lett. **13** (1964) 168–171.
- [30] F. Zwicky, *Die Rotverschiebung von extragalaktischen Nebeln*, Helv. Phys. Acta **6** (1933) 110–127.
- [31] M. S. Turner and D. Huterer, *Cosmic Acceleration, Dark Energy, and Fundamental Physics*, J. Phys. Soc. Jap. **76** (2007) 111015.
- [32] H. Miyazawa, *Baryon number Changing Currents*, Prog. Theor. Phys. **36** (1966) 1266.

- [33] P. Ramond, *Dual Theory for Free Fermions*, Phys. Rev. **D3** (1971) 2415.
- [34] Y. Golfand and E. Likhtman, *Extension of the Algebra of Poincare Group Generators and Violation of p Invariance*, JETP Lett. **13** (1971) 323.
- [35] A. Neveu and J. Schwarz, *Factorizable dual model of pions*, Nucl. Phys. **B31** (1971) 86.
- [36] A. Neveu and J. Schwarz, *Quark Model of Dual Pions*, Phys. Rev. **D4** (1971) 1109.
- [37] J.-L. Gervais and B. Sakita, *Field Theory Interpretation of Supergauges in Dual Models*, Nucl. Phys. **B34** (1971) 632.
- [38] D. Volkov and V. Akulov, *Is the Neutrino a Goldstone Particle?*, Phys. Lett. **B46** (1973) 109.
- [39] J. Wess and B. Zumino, *A Lagrangian Model Invariant Under Supergauge Transformations*, Phys. Lett. **B49** (1974) 52.
- [40] J. Wess and B. Zumino, *Supergauge Transformations in Four-Dimensions*, Nucl. Phys. **B70** (1974) 39.
- [41] P. H. Chankowski et al., *Cosmological fine tuning, supersymmetry and the gauge hierarchy problem*, Phys. Lett. B **452** (1998) 28–38.
- [42] D0 Collaboration, *Observation of the Top Quark*, Phys. Rev. Lett. **74** (1995) 2632, [arXiv:9503003 \[hep-ex\]](https://arxiv.org/abs/hep-ex/9503003).
- [43] CDF Collaboration, *Observation of Top Quark Production in $p\bar{p}$ Collision with Collider Detector at Fermilab*, Phys. Rev. Lett. **74** (1995) 2626, [arXiv:9503002 \[hep-ex\]](https://arxiv.org/abs/hep-ex/9503002).
- [44] ATLAS, CDF, CMS and D0 Collaboration, *First combination of Tevatron and LHC measurements of the top-quark mass*, [arXiv:1403.4427 \[hep-ex\]](https://arxiv.org/abs/1403.4427).
- [45] M. Cacciari et al., *Top-Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*, Phys. Lett. **B710** (2012) 612, [arXiv:1111.5869 \[hep-ph\]](https://arxiv.org/abs/1111.5869).
- [46] P. Barnreuther et al., *Percent Level Precision Physics at the Tevatron: First Genuine NNLO QCD Corrections to $q\bar{q} \rightarrow t\bar{t}$* , Phys. Rev. Lett. **109** (2012) 132001, [arXiv:1204.5201 \[hep-ph\]](https://arxiv.org/abs/1204.5201).
- [47] M. Czakon and A. Mitov, *NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels*, JHEP **1212** (2012) 054, [arXiv:1207.0236 \[hep-ph\]](https://arxiv.org/abs/1207.0236).
- [48] M. Czakon and A. Mitov, *NNLO corrections to top-pair production at hadron colliders: the quark-gluon reaction*, JHEP **1301** (2013) 080, [arXiv:1210.6832 \[hep-ph\]](https://arxiv.org/abs/1210.6832).
- [49] P. F. M. Czakon and A. Mitov, *The total top quark pair production cross-section at hadron colliders through $\mathcal{O}(\alpha_S^4)$* , Phys. Rev. Lett. **110** (2013) 252004, [arXiv:1303.6254 \[hep-ph\]](https://arxiv.org/abs/1303.6254).

Bibliography

- [50] ATLAS and CMS Collaboration, *Combination of ATLAS and CMS top quark pair cross section measurements in the $e\mu$ final state using proton-proton collisions at $\sqrt{s} = 8$ TeV*, ATLAS-CONF-2014-054, CMS-PAS-TOP-14-016 (2014).
- [51] N. Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production*, Phys. Rev. D **83** (2011) 091503, [arXiv:1103.2792 \[hep-ph\]](https://arxiv.org/abs/1103.2792).
- [52] N. Kidonakis, *Next-to-next-to-leading logarithm resummation for s-channel single top quark production*, Phys. Rev. D **81** (2010) 054028.
- [53] ATLAS and CMS Collaboration, *Combination of single top-quark cross-sections measurements in the t-channel at $\sqrt{s} = 8$ TeV with the ATLAS and CMS experiments*, CMS-PAS-TOP-12-002, ATLAS-CONF-2013-098 (2013).
- [54] CDF and D0 Collaboration, *Tevatron Measurement of Single Top Quark Cross Sections and the CKM Matrix Element V_{tb}* , FERMILAB-CONF-14-370-E.
- [55] ATLAS Collaboration, *Evidence for the associated production of a W boson and a top quark in ATLAS at $\sqrt{s} = 7$ TeV*, Phys. Lett. **B716** (2012) 142, [arXiv:1205.5764 \[hep-ex\]](https://arxiv.org/abs/1205.5764).
- [56] CMS Collaboration, *Evidence for associated production of a single top quark and W boson in pp collisions at $\sqrt{s} = 7$ TeV*, Phys. Rev. Lett. **110** (2013) 022003, [arXiv:1209.3489 \[hep-ex\]](https://arxiv.org/abs/1209.3489).
- [57] ATLAS and CMS Collaboration, *Combination of cross-section measurements for associated production of a single top-quark and a W boson at $\sqrt{s} = 8$ TeV with the ATLAS and CMS experiments*, ATLAS-CONF-2014-052, CMS-PAS-TOP-14-009 (2014).
- [58] ATLAS Collaboration, *Search for s-channel single top-quark production in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, Phys. Lett. **B740** (2015) 118–136, [arXiv:1410.0647 \[hep-ex\]](https://arxiv.org/abs/1410.0647).
- [59] CMS Collaboration, *Search for s-channel single top-quark production in pp collisions at $\sqrt{s} = 8$ TeV*, CMS-PAS-TOP-13-009 (2013).
- [60] CDF and D0 Collaboration, *Observation of s-channel production of single top quarks at the Tevatron*, Phys. Rev. Lett. (2014), [arXiv:1402.5126 \[hep-ex\]](https://arxiv.org/abs/1402.5126).
- [61] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. **B716** (2012) 1, [arXiv:1207.7214 \[hep-ex\]](https://arxiv.org/abs/1207.7214).
- [62] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. **B716** (2012) 30, [arXiv:1207.7235v1 \[hep-ex\]](https://arxiv.org/abs/1207.7235v1).
- [63] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, [arXiv:1307.1347 \[hep-ph\]](https://arxiv.org/abs/1307.1347).

- [64] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*, arXiv:1101.0593 [hep-ph].
- [65] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 2. Differential Distributions*, arXiv:1201.3084 [hep-ph].
- [66] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **1407** (2014) 079, arXiv:1405.0301 [hep-ph].
- [67] ATLAS Collaboration, *Search for the bb decay of the Standard Model Higgs boson in associated W/ZH production with the ATLAS detector*, JHEP **01** (2015) 069, arXiv:1409.6212 [hep-ex].
- [68] CMS Collaboration, *Search for the standard model Higgs boson produced in association with a W or a Z boson and decaying to bottom quarks*, Phys. Rev. D **89** (2014) 012003, arXiv:1310.3687 [hep-ex].
- [69] CDF and D0 Collaboration, *Evidence for a Particle Produced in Association with Weak Bosons and Decaying to a Bottom-Antibottom Quark Pair in Higgs Boson Searches at the Tevatron*, Phys. Rev. Lett. **109** (2012) 071804, arXiv:1207.6436 [hep-ex].
- [70] CMS Collaboration, *Search for the associated production of the Higgs boson with a top-quark pair*, JHEP **09** (2014) 087, arXiv:1408.1682 [hep-ex].
- [71] ATLAS Collaboration, *Search for the Standard Model Higgs boson produced in association with top quarks and decaying into b̄b in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, arXiv:1503.05066 [hep-ex].
- [72] ATLAS Collaboration, *Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC*, Phys. Lett. **B726** (2013) 88, arXiv:1307.1427 [hep-ex].
- [73] CMS Collaboration, *Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV*, arXiv:1412.8662 [hep-ex].
- [74] ATLAS Collaboration, *Search for the associated production of the Higgs boson with a top quark pair in multi-lepton final states with the ATLAS detector*, ATLAS-CONF-2015-006 (2015).
- [75] ATLAS Collaboration, *Observation and measurement of Higgs boson decays to WW* with the ATLAS detector*, arXiv:1412.2641 [hep-ex].
- [76] ATLAS Collaboration, *Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS Detector*, arXiv:1501.04943 [hep-ex].
- [77] ATLAS Collaboration, *Measurement of Higgs boson production in the diphoton decay channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector*, Phys. Rev. **D90** (2014) 112015, arXiv:1408.7084 [hep-ex].

Bibliography

- [78] ATLAS Collaboration, *Measurements of Higgs boson production and couplings in the four-lepton channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector*, Phys. Rev. **D91** (2015) 012006, [arXiv:1408.5191 \[hep-ex\]](https://arxiv.org/abs/1408.5191).
- [79] ATLAS and CMS Collaboration, *Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments*, [arXiv:1503.07589 \[hep-ex\]](https://arxiv.org/abs/1503.07589).
- [80] M. Baak et al., *The Electroweak Fit of the Standard Model after the Discovery of a New Boson at the LHC*, Eur. Phys. J. **C72** (2012) 2205, [arXiv:1209.2716 \[hep-ph\]](https://arxiv.org/abs/1209.2716).
- [81] Gfitter Group, *The global electroweak fit at NNLO and prospects for the LHC and ILC*, Eur. Phys. J. **C74** (2014) 3046, [arXiv:1407.3792 \[hep-ph\]](https://arxiv.org/abs/1407.3792).
- [82] ATLAS Collaboration, *Evidence for the spin-0 nature of the Higgs boson using ATLAS data*, Phys. Lett. **B726** (2013) 120, [arXiv:1307.1432 \[hep-ex\]](https://arxiv.org/abs/1307.1432).
- [83] CMS Collaboration, *Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV*, [arXiv:1411.3441 \[hep-ex\]](https://arxiv.org/abs/1411.3441).
- [84] ATLAS Collaboration, *Study of the spin and parity of the Higgs boson in HVV decays with the ATLAS detector*, ATLAS-CONF-2015-008 (2015).
- [85] L. Landau, *On the angular momentum of a two-photon system*, Dokl. Akad. Nauk Ser. Fiz. **60** (1948) 207–209.
- [86] C. N. Yang, *Selection Rules for the Dematerialization of a Particle into Two Photons*, Phys. Rev. **77** (1950) 242–245.
- [87] ATLAS Collaboration, *Determination of spin and parity of the Higgs boson in the $WW^* \rightarrow e\nu\mu\nu$ decay channel with the ATLAS detector*, [arXiv:1503.03643 \[hep-ex\]](https://arxiv.org/abs/1503.03643).
- [88] ATLAS Collaboration, *Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s} = 7$ and 8 TeV in the ATLAS experiment*, ATLAS-CONF-2015-007 (2015).
- [89] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001.
- [90] LEP Collaboration, *LEP design report*, CERN-LEP-84-01 (1984).
- [91] LHCb Collaboration, *The LHCb Detector at the LHC*, JINST **3** (2008) S08005.
- [92] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002.
- [93] LHCf Collaboration, *The LHCf detector at the CERN Large Hadron Collider*, JINST **3** (2008) S08006.
- [94] TOTEM Collaboration, *TOTEM: Technical design report. Total cross section, elastic scattering and diffraction dissociation at the Large Hadron Collider at CERN*, CERN-LHCC-2004-002 (2004).

- [95] MoEDAL Collaboration, *Technical Design Report of the MoEDAL Experiment*, CERN-LHCC-2009-006, MoEDAL-TDR-001 (2009).
- [96] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*, CERN-LHCC-2010-013, ATLAS-TDR-19 (2010).
- [97] R. L. Glückstern, *Uncertainties in Track Momentum and Direction due to Multiple Scattering and Measurement Errors*, Nucl. Instr. Meth. A **24** (1963) 381.
- [98] S. van der Meer, *Calibration of the effective beam height in the ISR*, CERN-ISR-PO-68-31, ISR-PO-68-31 (1968).
- [99] *Summary of the analysis of the 19 September 2008 incident at the LHC*, Document EDMS 973073 (2008).
- [100] *Follow up of the incident of 19 September 2008 at the LHC*, CERN PR17.08 (2008).
- [101] ATLAS Collaboration, *Improved luminosity determination in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector at the LHC*, Eur. Phys. J. C **73** (2013) 2518, arXiv:1302.4393 [hep-ex].
- [102] D. Fournier, *Performance of the LHC, ATLAS and CMS in 2011*, EPJ Web Conf. **28** (2012) 01003, arXiv:1201.4681 [hep-ex].
- [103] M. Lamont, *Status of the LHC*, J. Phys. Conf. Ser. **455** (2013) 012001.
- [104] ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 7$ TeV*, Eur. Phys. J. C**73** (2013) 2304, arXiv:1112.6426 [hep-ex].
- [105] ATLAS Collaboration, *Jet energy measurement and its systematic uncertainty in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, arXiv:1406.0076 [hep-ex].
- [106] M. Cacciari et al., *The anti- k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv:0802.1189 [hep-ph].
- [107] M. Cacciari and G. P. Salam, *Dispelling the N^3 myth for the k_t jet-finder*, Phys. Lett. B**641** (2006) 57, arXiv:0512210v2 [hep-ph].
- [108] C. Cojocaru et al., *Hadronic calibration of the ATLAS liquid argon end-cap calorimeter in the pseudorapidity region $1.6 < |\eta| < 1.8$ in beam tests*, Nucl. Instr. Meth. A **531** (2004) 481, arXiv:0407009 [physics].
- [109] T. Barillari et al., *Local hadronic calibration*, ATL-LARG-PUB-2009-001 (2009).
- [110] M. Cacciari et al., *FastJet User Manual*, Eur. Phys. J. C**72** (2012) 1896, arXiv:1111.6097 [hep-ph].
- [111] ATLAS Collaboration, *Pile-up subtraction and suppression for jets in ATLAS*, ATLAS-CONF-2013-083 (2013).

Bibliography

- [112] D0 Collaboration, *Measurement of the $p\bar{p} \rightarrow t\bar{t}$ production cross section at $\sqrt{s} = 1.96\text{-TeV}$ in the fully hadronic decay channel.*, Phys. Rev. **D76** (2007) 072007, [arXiv:0612040 \[hep-ex\]](#).
- [113] ATLAS Collaboration, *Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data*, ATLAS-CONF-2011-102 (2011).
- [114] ATLAS Collaboration, *Calibration of b-tagging using dileptonic top pair events in a combinatorial likelihood approach with the ATLAS experiment*, ATLAS-CONF-2014-004 (2014).
- [115] ATLAS Collaboration, *Calibration of the performance of b-tagging for c and light-flavour jets in the 2012 ATLAS data*, ATLAS-CONF-2014-046 (2014).
- [116] ATLAS Collaboration, *Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data*, Eur. Phys. J. C **72** (2012) 1909, [arXiv:1110.3174 \[hep-ex\]](#).
- [117] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton-proton collision data*, ATLAS-CONF-2014-032 (2014).
- [118] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data*, [arXiv:1407.5063 \[hep-ex\]](#).
- [119] W. Lampl et al., *Calorimeter clustering algorithms: Description and performance*, ATL-LARG-PUB-2008-002 (2008).
- [120] ATLAS Collaboration, *Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton-proton collision data*, [arXiv:1407.3935 \[hep-ex\]](#).
- [121] K. Rehermann and B. Tweedie, *Efficient Identification of Boosted Semileptonic Top Quarks at the LHC*, JHEP **1103** (2011) 059, [arXiv:1007.2221 \[hep-ph\]](#).
- [122] ATLAS Collaboration, *Performance of Missing Transverse Momentum Reconstruction in ATLAS studied in Proton-Proton Collisions recorded in 2012 at 8 TeV*, ATLAS-CONF-2013-082 (2013).
- [123] J. C. Collins et al., *Factorization of Hard Processes in QCD*, Adv. Ser. Direct. HEP **5** (1988) 1–91, [arXiv:0409313 \[hep-ph\]](#).
- [124] T. Gleisberg et al., *Event generation with SHERPA 1.1*, JHEP **0902** (2009) 007, [arXiv:0811.4622 \[hep-ph\]](#).
- [125] M. L. Mangano et al., *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, JHEP **07** (2003) 001, [arXiv:0206293 \[hep-ph\]](#).
- [126] J. Alwall et al., *MadGraph 5 : Going Beyond*, [arXiv:1106.0522](#) (2011).
- [127] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040.

- [128] T. Sjöstrand, S. Mrenna, and P. Skands, *Pythia 6.4 Physics and Manual*, JHEP **05** (2006) 026, [arXiv:0603175v2 \[hep-ph\]](https://arxiv.org/abs/0603175v2).
- [129] T. Sjöstrand, S. Mrenna, and P. Skands, *A Brief Introduction to Pythia 8.1*, [arXiv:0710.3820 \[hep-ph\]](https://arxiv.org/abs/0710.3820).
- [130] G. Corcella et al., *HERWIG 6: an event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)*, JHEP **01** (2001) 010.
- [131] B. Andersson et al., *Parton fragmentation and string dynamics*, Phys. Rep. **97** (1983) 31–145.
- [132] G. Marchesini and B. R. Webber, *Simulation of QCD jets including soft gluon interference*, Nucl. Phys. B **238** (1984) 1–29.
- [133] J. Butterworth, J. Forshaw, and M. Seymour, *Multiparton interactions in photoproduction at HERA*, Z. Phys. C **72** (1996) 637, [arXiv:9601371 \[hep-ph\]](https://arxiv.org/abs/9601371).
- [134] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823, [arXiv:1005.4568 \[physics.ins-det\]](https://arxiv.org/abs/1005.4568).
- [135] S. Agostinelli et al., *Geant4: a simulation toolkit*, Nucl. Instr. Meth. A **506** (2003) 250.
- [136] ATLAS Collaboration, *The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim*, ATL-PHYS-PUB-2010-013 (2010).
- [137] G. Bevilacqua et al., *HELAC-NLO*, Comput. Phys. Commun. **184** (2013) 986, [arXiv:1110.1499v2 \[hep-ph\]](https://arxiv.org/abs/1110.1499v2).
- [138] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, [arXiv:0709.2092 \[hep-ph\]](https://arxiv.org/abs/0709.2092).
- [139] S. Alioli et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 040, [arXiv:1002.2581 \[hep-ph\]](https://arxiv.org/abs/1002.2581).
- [140] S. Dawson et al., *Associated Higgs production with top quarks at the Large Hadron Collider: NLO QCD corrections*, Phys. Rev. D**68** (2003) 034022, [arXiv:0305087 \[hep-ph\]](https://arxiv.org/abs/0305087).
- [141] L. Reina and S. Dawson, *Next-to-leading order results for $t\bar{t}h$ production at the Tevatron*, Phys. Rev. Lett. **87** (2001) 201804, [arXiv:0107101 \[hep-ph\]](https://arxiv.org/abs/0107101).
- [142] W. Beenakker et al., *NLO QCD corrections to $t\bar{t}H$ production in hadron collisions*, Nucl. Phys. B**653** (2003) 151–203, [arXiv:0211352 \[hep-ph\]](https://arxiv.org/abs/0211352).
- [143] W. Beenakker et al., *Higgs radiation off top quarks at the Tevatron and the LHC*, Phys. Rev. Lett. **87** (2001) 201805, [arXiv:0107081 \[hep-ph\]](https://arxiv.org/abs/0107081).
- [144] A. Djouadi et al., *HDECAY: A Program for Higgs boson decays in the standard model and its supersymmetric extension*, Comput. Phys. Commun. **108** (1998) 56–74, [arXiv:9704448 \[hep-ph\]](https://arxiv.org/abs/9704448).

Bibliography

- [145] A. Bredenstein et al., *Precise predictions for the Higgs-boson decay $H \rightarrow WW/ZZ \rightarrow 4$ leptons*, Phys. Rev. **D74** (2006) 013004, [arXiv:0604011 \[hep-ph\]](#).
- [146] S. Actis et al., *NNLO Computational Techniques: The Cases $H \rightarrow \gamma\gamma$ and $H \rightarrow gg$* , Nucl. Phys. **B811** (2009) 182–273, [arXiv:0809.3667 \[hep-ph\]](#).
- [147] A. Denner et al., *Standard Model Higgs-Boson Branching Ratios with Uncertainties*, Eur. Phys. J. **C71** (2011) 1753, [arXiv:1107.5909 \[hep-ph\]](#).
- [148] H.-L. Lai et al., *New parton distributions for collider physics*, Phys. Rev. D **82** (2010) 074024, [arXiv:1007.2241 \[hep-ph\]](#).
- [149] ATLAS Collaboration, *ATLAS tunes of PYTHIA6 and PYTHIA8 for MC11*, ATL-PHYS-PUB-2011-008 (2011).
- [150] ATLAS Collaboration, *New ATLAS event generator tunes to 2010 data*, ATL-PHYS-PUB-2011-009 (2011).
- [151] P. Gonka and Z. Wąs, *PHOTOS Monte Carlo: a precision tool for QED corrections in Z and W decays*, Eur. Phys. J. C **45** (2006) 97, [arXiv:0506026 \[hep-ph\]](#).
- [152] S. Jadach, *TAUOLA - a library of Monte Carlo programs to simulate decays of polarized τ leptons*, Comput. Phys. Commun. **64** (1991) 275.
- [153] P. Z. Skands, *Tuning Monte Carlo Generators: The Perugia Tunes*, Phys. Rev. **D82** (2010) 074018, [arXiv:1005.3457 \[hep-ph\]](#).
- [154] M. Czakon and A. Mitov, *Top++: a program for the calculation of the top-pair cross-section at hadron colliders*, [arXiv:1112.5675 \[hep-ph\]](#).
- [155] F. Cascioli et al., *NLO matching for $t\bar{t}bb$ production with massive b -quarks*, [arXiv:1309.5912 \[hep-ph\]](#).
- [156] F. Cascioli et al., *Scattering Amplitudes with Open Loops*, Phys. Rev. Lett. **108** (2012), [arXiv:1111.5206 \[hep-ph\]](#).
- [157] ATLAS Collaboration, *Measurements of normalized differential cross sections for $t\bar{t}$ production in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, Phys. Rev. D **90** (2014) 072004.
- [158] S. Frixione et al., *Single-top production in MC@NLO*, JHEP **03** (2006) 092, [arXiv:0512250 \[hep-ph\]](#).
- [159] P. M. Nadolsky et al., *Implications of CTEQ global analysis for collider observables*, Phys. Rev. D **78** (2008) 013004, [arXiv:0802.0007 \[hep-ph\]](#).
- [160] J. M. Campbell and R. K. Ellis, *$t\bar{t}W$ production and decay at NLO*, [arXiv:1204.5678 \[hep-ph\]](#).
- [161] M. V. Garzelli et al., *$t\bar{t}W$ and $t\bar{t}Z$ Hadroproduction at NLO accuracy in QCD with Parton Shower and Hadronization effects*, JHEP **1211** (2012) 056, [arXiv:1208.2665 \[hep-ph\]](#).

- [162] M. L. Mangano et al., *Multijet matrix elements and shower evolution in hadronic collisions: $Wb\bar{b} + n$ jets as a case study*, Nucl. Phys. **B** **632** (2002) 343, arXiv:0108069 [hep-ph].
- [163] K. Melnikov and F. Petriello, *Electroweak gauge boson production at hadron colliders through $\mathcal{O}(\alpha_s^2)$* , Phys. Rev. D **74** (2006) 114017, arXiv:0609070 [hep-ph].
- [164] J. Campbell and R. Ellis, *An update on vector boson pair production at hadron colliders*, Phys. Rev. D **60** (1999) 113006, arXiv:9905386 [hep-ph].
- [165] D0 Collaboration, *Measurement of the $t\bar{t}$ production cross section in pp collisions at $\sqrt{s} = 1.96$ TeV using secondary vertex b-tagging*, Phys. Rev. D **74** (2006) 112004, arXiv:0611002 [hep-ex].
- [166] A. Shibata and B. Clement, *Tagging Rate Function B-Tagging*, ATL-PHYS-PUB-2007-011 (2007).
- [167] ATLAS Collaboration, *Estimation of non-prompt and fake lepton backgrounds in final states with top quarks produced in proton-proton collisions at 8 TeV with the ATLAS detector*, ATLAS-CONF-2014-058 (2014).
- [168] ATLAS Collaboration, *Measurement of the Production Cross section of Jets in Association with a Z Boson in pp Collisions at 7 TeV Using the ATLAS Detector*, JHEP **07** (2013) 32.
- [169] ATLAS Collaboration, *Evidence for the associated production of a vector boson (W, Z) and top quark pair in the dilepton and trilepton channels in pp collision data at $\sqrt{s} = 8$ TeV collected by the ATLAS detector at the LHC*, ATLAS-CONF-2014-038 (2014).
- [170] K. Kondo, *Dynamical Likelihood Method for Reconstruction of Events With Missing Momentum. 1: Method and Toy Models*, J. Phys. Soc. Jap. **57** (1988) 4126–4140.
- [171] K. Kondo, *Dynamical likelihood method for reconstruction of events with missing momentum. 2: Mass spectra for $2 \rightarrow 2$ processes*, J. Phys. Soc. Jap. **60** (1991) 836–844.
- [172] K. Kondo et al., *Dynamical Likelihood Method for Reconstruction of Events with Missing Momentum. III. Analysis of a CDF High p_T $e\mu$ Event as $t\bar{t}$ Production*, J. Phys. Soc. Jap. **62** (1993) 1177–1182.
- [173] D0 Collaboration, *Precision Measurement of the Top Quark Mass in Lepton + Jets Final States*, Phys. Rev. Lett. **113** (2014) 032002.
- [174] D0 Collaboration, *Observation of Single Top-Quark Production*, Phys. Rev. Lett. **103** (2009) 092001.
- [175] CDF Collaboration, *Observation of Electroweak Single Top-Quark Production*, Phys. Rev. Lett. **103** (2009) 092002.
- [176] D0 Collaboration, *Measurement of spin correlation in $t\bar{t}$ production using a matrix element approach*, Phys. Rev. Lett. **107** (2011) 032001, arXiv:1104.5194 [hep-ex].
- [177] J. S. Gainer et al., *Improving the sensitivity of Higgs boson searches in the golden channel*, JHEP **1111** (2011) 027, arXiv:1108.2274 [hep-ph].

Bibliography

- [178] J. M. Campbell, W. T. Giele, and C. Williams, *The Matrix Element Method at Next-to-Leading Order*, JHEP **1211** (2012) 043, arXiv:1204.4424 [hep-ph].
- [179] F. M. Pierre Artoisenet, Priscila de Aquino and O. Mattelaer, *Unravelling $t\bar{t}h$ via the matrix element method*, arXiv:1304.6414 [hep-ph].
- [180] CMS Collaboration, *Search for a standard model Higgs boson produced in association with a top-quark pair and decaying to bottom quarks using a matrix element method*, arXiv:1502.02485 [hep-ex].
- [181] E. Fermi, *Nuclear Physics*. University of Chicago Press, 1950.
- [182] J. Neyman and E. Pearson, *On the Problem of the Most Efficient Tests of Statistical Hypotheses*, Phil. Trans. R. Soc. Lond. A **231** (1933) 694–706, 289–337.
- [183] J. Alwall et al., *New Developments in MadGraph/MadEvent*, arXiv:0809.2410 (2008).
- [184] Carsten Brachem, *Studies for a top quark mass measurement using the matrix element method in the semileptonic channel*, II.Physik-UniG-MSc-2012/01 (2012).
- [185] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, Nucl. Instr. Meth. A **A 389** (1997) 81–86.
- [186] M. R. Whalley, D. Bourilkov, and R. C. Group, *The Les Houches accord PDFs (LHAPDF) and LHAGLUE*, arXiv:0508110 [hep-ph].
- [187] M. Galassi et al., *GNU Scientific Library Reference Manual (3rd Ed.)*, ISBN 0954612078 (2009).
- [188] J. Erdmann et al., *A likelihood-based reconstruction algorithm for top-quark pairs and the KLFitter framework*, Nucl. Instrum. Meth. A **748** (2014) 18 – 25.
- [189] ZEUS and H1 Collaboration, *PDF Fits at HERA*, arXiv:1112.2107 [hep-ph].
- [190] A. D. Martin et al., *Parton distributions for the LHC*, Eur. Phys. J. C **63** (2009) 189, arXiv:0901.0002 [hep-ph].
- [191] A. D. Martin et al., *Uncertainties on α_S in global PDF analyses and implications for predicted hadronic cross sections*, Eur. Phys. J. C **64** (2009) 653, arXiv:0905.3531 [hep-ph].
- [192] R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. **B867** (2013) 244, arXiv:1207.1303 [hep-ph].
- [193] S. Alekhin, *Parton distributions from deep inelastic scattering data*, Phys. Rev. **D68** (2003) 014002, arXiv:0211096 [hep-ph].
- [194] J. N. Ng and P. Zakarauskas, *QCD-parton calculation of conjoined production of Higgs bosons and heavy flavors in $p\bar{p}$ collisions*, Phys. Rev. D **29** (1984) 876–886.
- [195] G. P. Lepage, *A New Algorithm for Adaptive Multidimensional Integration*, J. Comput. Phys. **27** (1978) 192.

- [196] W. H. Press and G. R. Farrar, *Recursive Stratified Sampling for Multidimensional Monte Carlo Integration*, Comput. Phys. **4** (1990) 190–195.
- [197] CMS Collaboration, *Evidence for the direct decay of the 125 GeV Higgs boson to fermions*, Nature Physics **10** (2014) 557, arXiv:1401.6527 [hep-ex].
- [198] J. N. Ng and P. Zakarauskas, *QCD-parton calculation of conjoined production of Higgs bosons and heavy flavors in p anti-p collisions*, Phys. Rev. D **29** (1984) 876.
- [199] Z. Kunszt, *Associated production of heavy Higgs boson with top quarks*, Nucl. Phys. B **29** (1984) 876.
- [200] S. Dawson et al., *Associated top quark Higgs boson production the LHC*, Phys. Rev. D **67** (2003) 071503, arXiv:0211438 [hep-ph].
- [201] CDF Collaboration, *Search for the standard model Higgs boson produced in association with top quarks using the full CDF data set*, Phys. Rev. Lett. **109** (2012) 181802, arXiv:1208.2662 [hep-ex].
- [202] CMS Collaboration, *Search for Higgs boson production in association with top quark pairs in pp collisions*, CMS-PAS-HIG-2012-025 (2012).
- [203] CMS Collaboration, *Search for the standard model Higgs boson produced in association with a top-quark pair in pp collisions at the LHC*, JHEP **05** (2013) 145, arXiv:1303.0763 [hep-ex].
- [204] ATLAS Collaboration, *Search for the Standard Model Higgs boson produced in association with top quarks in proton-proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, ATLAS-CONF-2012-135 (2012).
- [205] ATLAS Collaboration, *Search for the Standard Model Higgs boson produced in association with top quarks and decaying to $b\bar{b}$ in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector at the LHC*, ATLAS-CONF-2014-011 (2014).
- [206] ATLAS Collaboration, *Search for $H \rightarrow \gamma\gamma$ produced in association with top quarks and constraints on the Yukawa coupling between the top quark and the Higgs boson using data taken at 7 TeV and 8 TeV with the ATLAS detector*, Phys. Lett. **B740** (2015) 222–242, arXiv:1409.3122 [hep-ex].
- [207] F. Bezrukov and M. Shaposhnikov, *Why should we care about the top quark Yukawa coupling?*, arXiv:1411.1923 [hep-ph].
- [208] ATLAS Collaboration, *Selection of jets produced in proton-proton collisions with the ATLAS detector using 2011 data*, ATLAS-CONF-2012-020 (2012).
- [209] ATLAS Collaboration, *Jet energy resolution in proton-proton collisions at $\sqrt{s} = 7$ TeV recorded in 2010 with the ATLAS detector*, Eur. Phys. J. **C73** (2013) 2306, arXiv:1210.6210 [hep-ex].
- [210] J. Gao et al., *The CT10 NNLO Global Analysis of QCD*, arXiv:1302.6246 [hep-ph].
- [211] M. Botje et al., *The PDF4LHC Working Group Interim Recommendations*, arXiv:1101.0538 [hep-ph].

Bibliography

- [212] ATLAS Collaboration, *Estimation of the W+Jets Background for Top Quark Re-Discovery in the Single Lepton+Jets Channel*, ATL-COM-PHYS-2010-834 (2010).
- [213] M. Bähr et al., *Herwig++ Physics and Manual*, Eur. Phys. J. **C58** (2008) 639, [arXiv:0803.0883 \[hep-ph\]](https://arxiv.org/abs/0803.0883).
- [214] J. Bellm et al., *Herwig++ 2.7 Release Note*, [arXiv:1310.6877 \[hep-ph\]](https://arxiv.org/abs/1310.6877).
- [215] K. Pearson, *On the criterion that a given system of derivations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling*, LED **50** (1900) 157–175.
- [216] A. N. Kolmogorov, *Sulla Determinazione Empirica di una Legge di Distribuzione*, Giornale dell’Istituto Italiano degli Attuari **4** (1933) 83–91.
- [217] N. Smirnov, *Table for Estimating the Goodness of Fit of Empirical Distributions*, Ann. Math. Statist. **19** (1948) 279–281.
- [218] Phi-T GmbH, *NeuroBayes package*, <http://neurobayes.phi-t.de/>.
- [219] M. Feindt and U. Kerzel, *The NeuroBayes neural network package*, Nucl. Instrum. Meth. **A559** (2006) 190–194.
- [220] R. H. Byrd et al., *A Limited Memory Algorithm for Bound Constrained Optimization*, SIAM J. Sci. Comput. **16** (1995) 1190–1208.
- [221] G. Fox and S. Wolfram, *Observables for the Analysis of Event Shapes in e^+e^- Annihilation and Other Processes*, Phys. Rev. Lett. **41** (1978) 1581.
- [222] C. Bernaciak et al., *Fox-Wolfram Moments in Higgs Physics*, Phys. Rev. **D87** (2013) 073014, [arXiv:1212.4436 \[hep-ph\]](https://arxiv.org/abs/1212.4436).
- [223] V. Barger, J. Ohnemus, and R. Phillips, *Event shape criteria for single lepton top signals*, Phys. Rev. D **48** (1993) 3953, [arXiv:9308216 \[hep-ph\]](https://arxiv.org/abs/9308216).
- [224] ROOT Collaboration, K. Cranmer et al., *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, CERN-OPEN-2012-016 (2012).
- [225] R. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*, Comput. Phys. Commun. **77** (1993) 219 – 228.
- [226] W. Verkerke and D. Kirkby, *The RooFit toolkit for data modeling*, [arXiv:0306116 \[physics\]](https://arxiv.org/abs/0306116).
- [227] G. Cowan et al., *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. C **71** (2011) 1554, [arXiv:1007.1727 \[physics.data-an\]](https://arxiv.org/abs/1007.1727).
- [228] T. Junk, *Confidence level computation for combining searches with small statistics*, Nucl. Instr. Meth. A **434** (1999) 435, [arXiv:9902006 \[hep-ex\]](https://arxiv.org/abs/9902006).
- [229] A. Read, *Presentation of search results: the CL_s technique*, J. Phys. G **28** (2002) 2693.
- [230] W. Verkerke and D. Kirkby, *RooFit Users Manual*, <http://roofit.sourceforge.net/>.

- [231] ATLAS Collaboration, *Procedure for the LHC Higgs boson search combination in summer 2011*, ATL-PHYS-PUB-2011-011 (2011).
- [232] S. S. Wilks, *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*, Ann. Math. Statist. **9** (1938) 60–62.
- [233] A. Wald, *Tests of statistical hypotheses concerning several parameters when the number of observations is large*, Trans. Amer. Math. Soc. **54** (1943) 426–482.
- [234] T. Plehn et al., *Fat Jets for a Light Higgs Boson*, Phys. Rev. Lett. **104** (2010) 111801.
- [235] P. Onyisi et al., *Analysis of $t\bar{t}H$ Events at $\sqrt{s}=14$ TeV with $H \rightarrow WW^{(*)}$* , arXiv:1307.7280 [hep-ex].
- [236] ATLAS Collaboration, *Physics at a High-Luminosity LHC with ATLAS*, ATL-PHYS-PUB-2013-007 (2013).
- [237] CMS Collaboration, *Measurement of the cross section ratio $ttbb/ttjj$ in pp Collisions at 8 TeV*, CMS-PAS-TOP-13-010 (2013).

Appendices

A

Further Separation Studies using the MEM

The Figures A.1 – A.1 show distributions normalised to unity of the most important MEM variables comparing the $t\bar{t}H$ process with the total background, the $t\bar{t} + b\bar{b}$ process, the non- $t\bar{t}$ background and the $t\bar{t}V$ process. Additionally, a comparison of $t\bar{t}H$ and the single top quark Higgs boson production with SM and BSM couplings together with the total background is shown. Single top quark Higgs boson production is sensitive to the sign of the top quark Yukawa coupling. For completeness, a comparison of $t\bar{t} + \text{light}$ and $t\bar{t} + \text{HF}$ events is also presented.

A. Further Separation Studies using the MEM

A.1. Likelihood Ratio

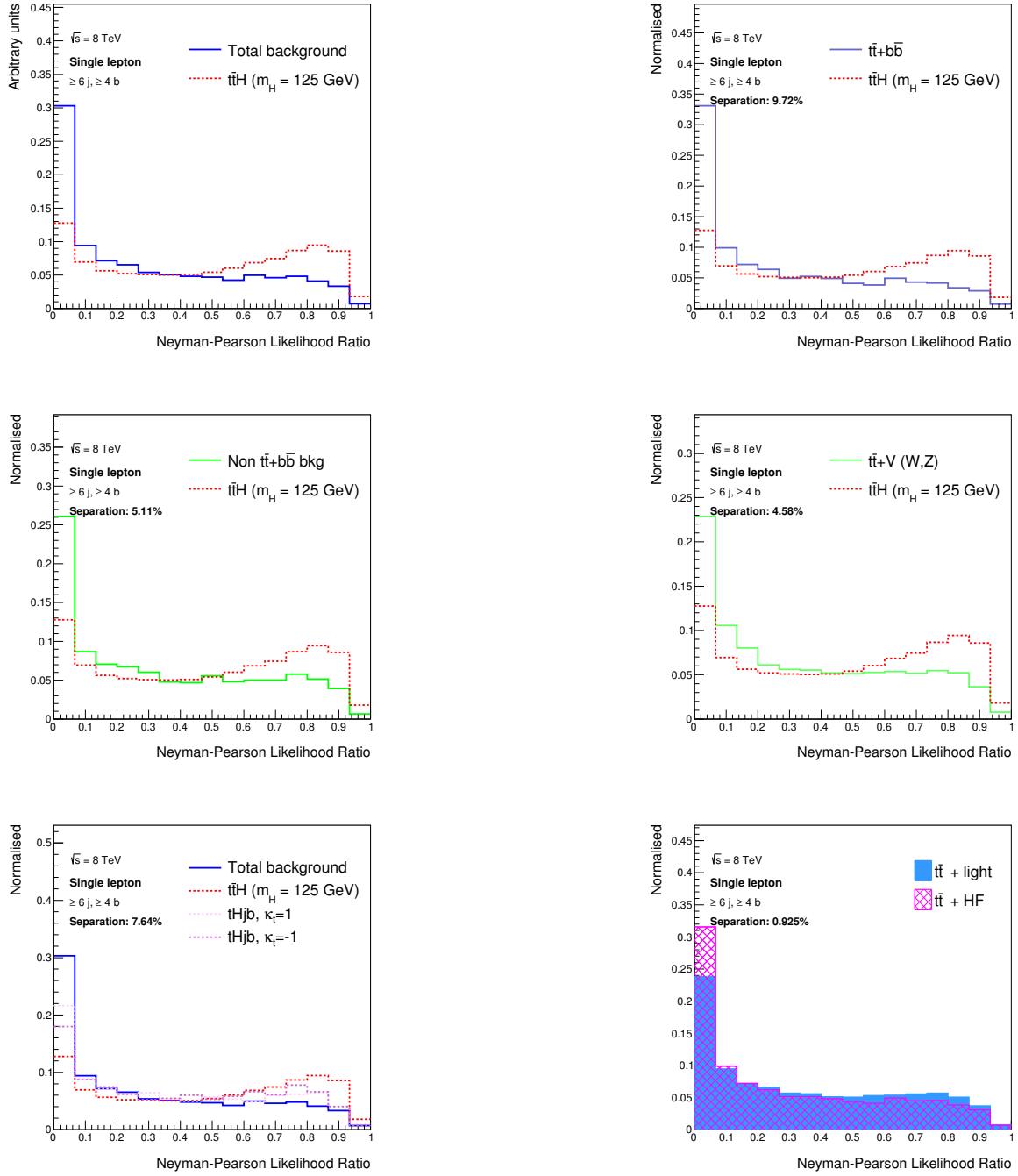


Figure A.1.: The likelihood ratio $D1$ with optimal normalisation summed over all possible assignment permutations in the ($\geq 6j, \geq 4b$) region for the $t\bar{t}H$ signal prediction (dashed red) to various background processes, the single top Higgs boson production as well as $t\bar{t} + \text{HF}$ to $t\bar{t} + \text{LF}$ events.

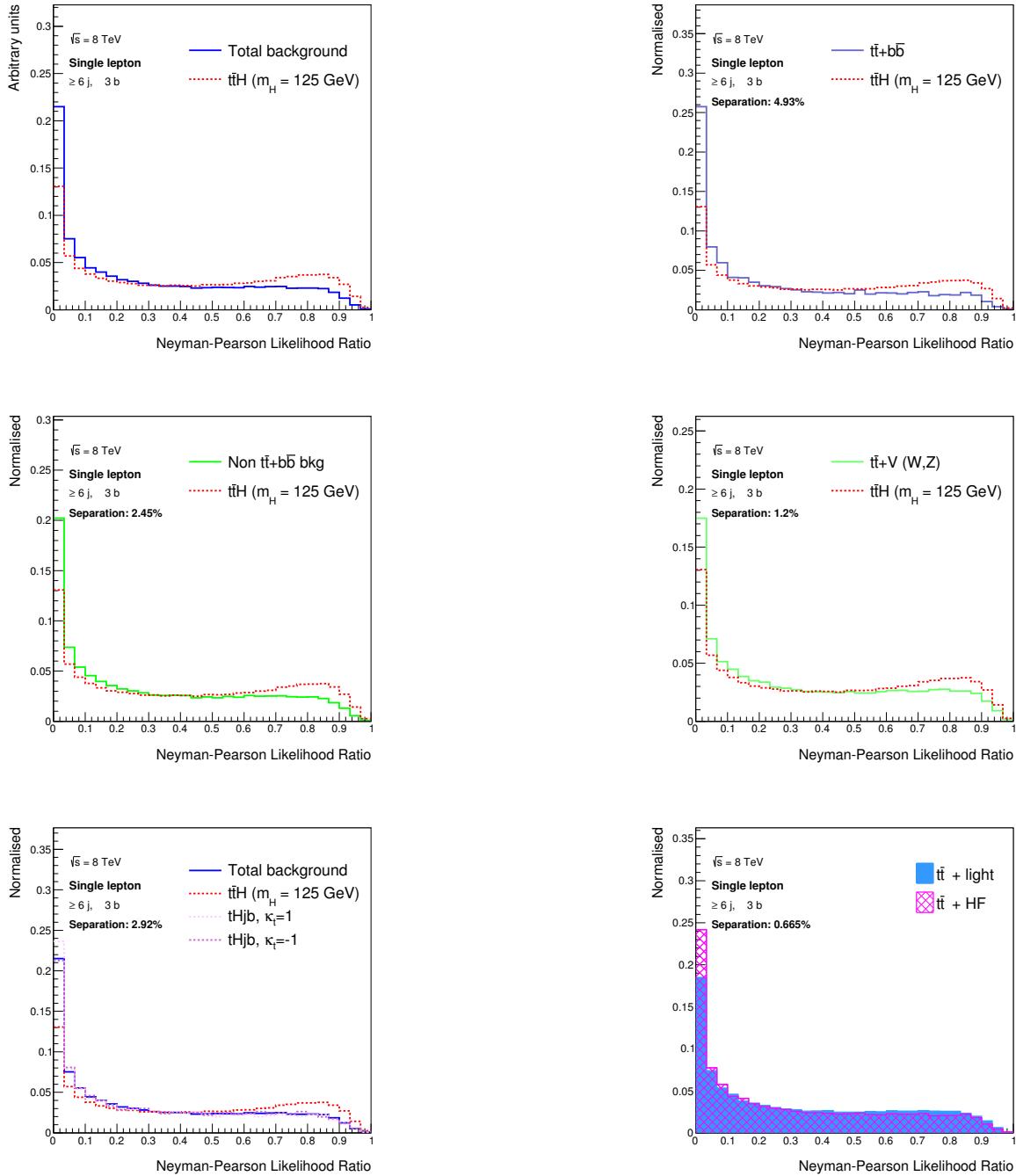


Figure A.2.: The likelihood ratio D_1 with optimal normalisation summed over all possible assignment permutations in the ($\geq 6j, 3b$) region for the $t\bar{t}H$ signal prediction (dashed red) to various background processes, the single top Higgs boson production as well as $t\bar{t} + \text{HF}$ to $t\bar{t} + \text{LF}$ events.

A. Further Separation Studies using the MEM

A.2. Logarithm of Summed Signal Likelihood

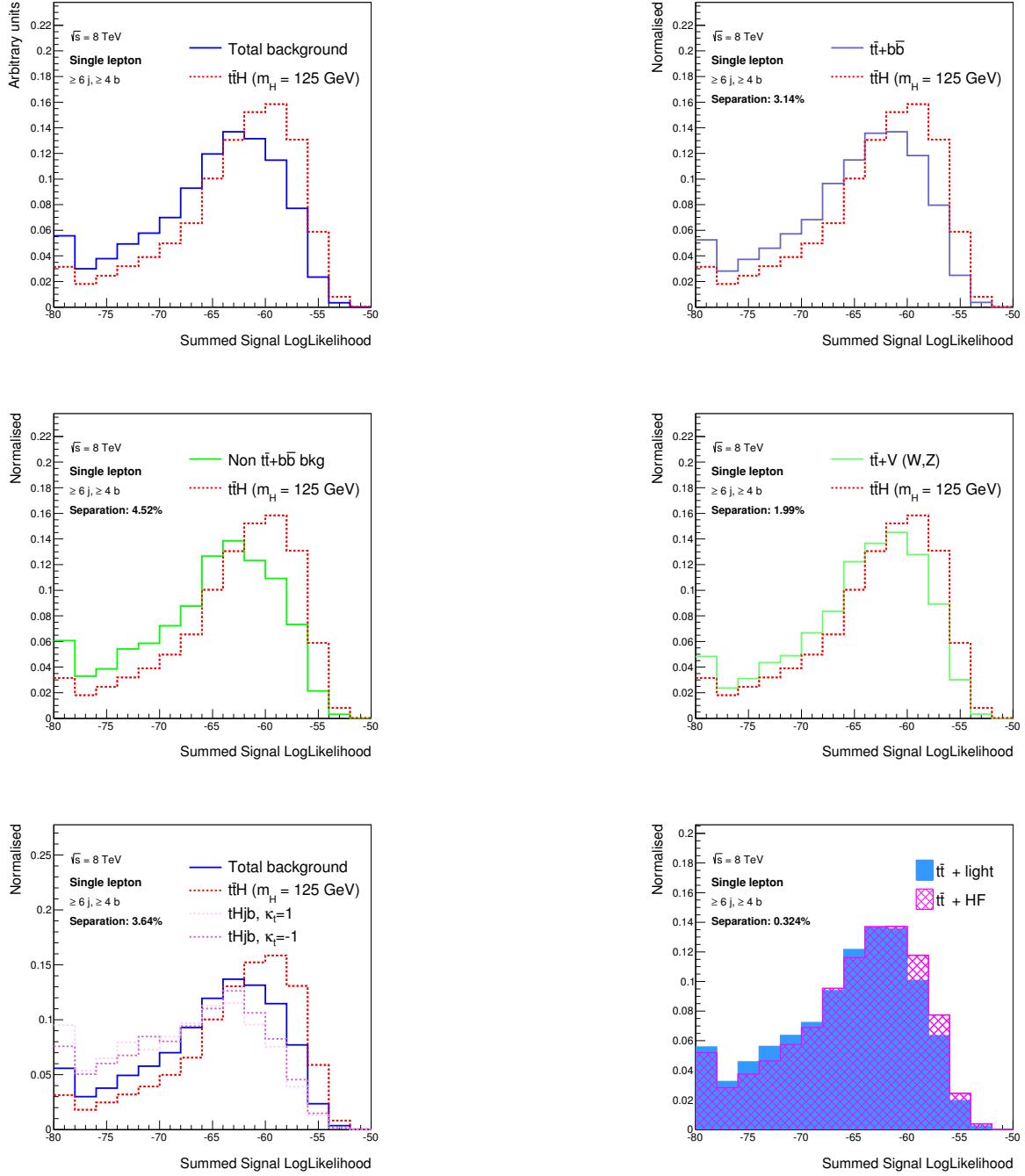


Figure A.3.: The logarithm of summed signal likelihood in the ($\geq 6j, \geq 4b$) region for the $t\bar{t}H$ signal prediction (dashed red) to various background processes, the single top Higgs boson production as well as $t\bar{t} + \text{HF}$ to $t\bar{t} + \text{LF}$ events.

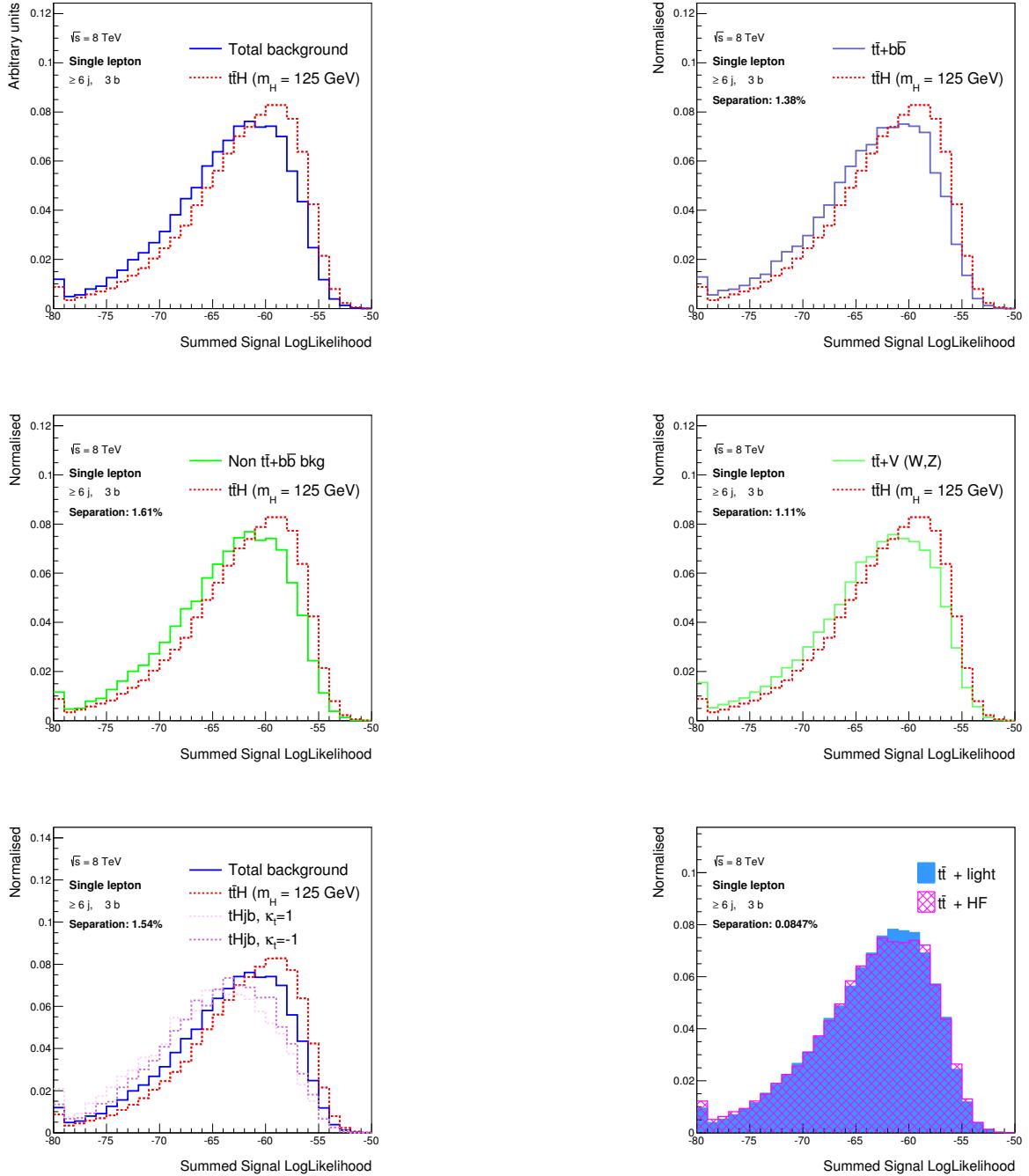


Figure A.4.: The logarithm of summed signal likelihood in the ($\geq 6j, 3b$) region for the $t\bar{t}H$ signal prediction (dashed red) to various background processes, the single top Higgs boson production as well as $t\bar{t} + \text{HF}$ to $t\bar{t} + \text{LF}$ events.

A.3. Invariant Mass of Additional B-jets

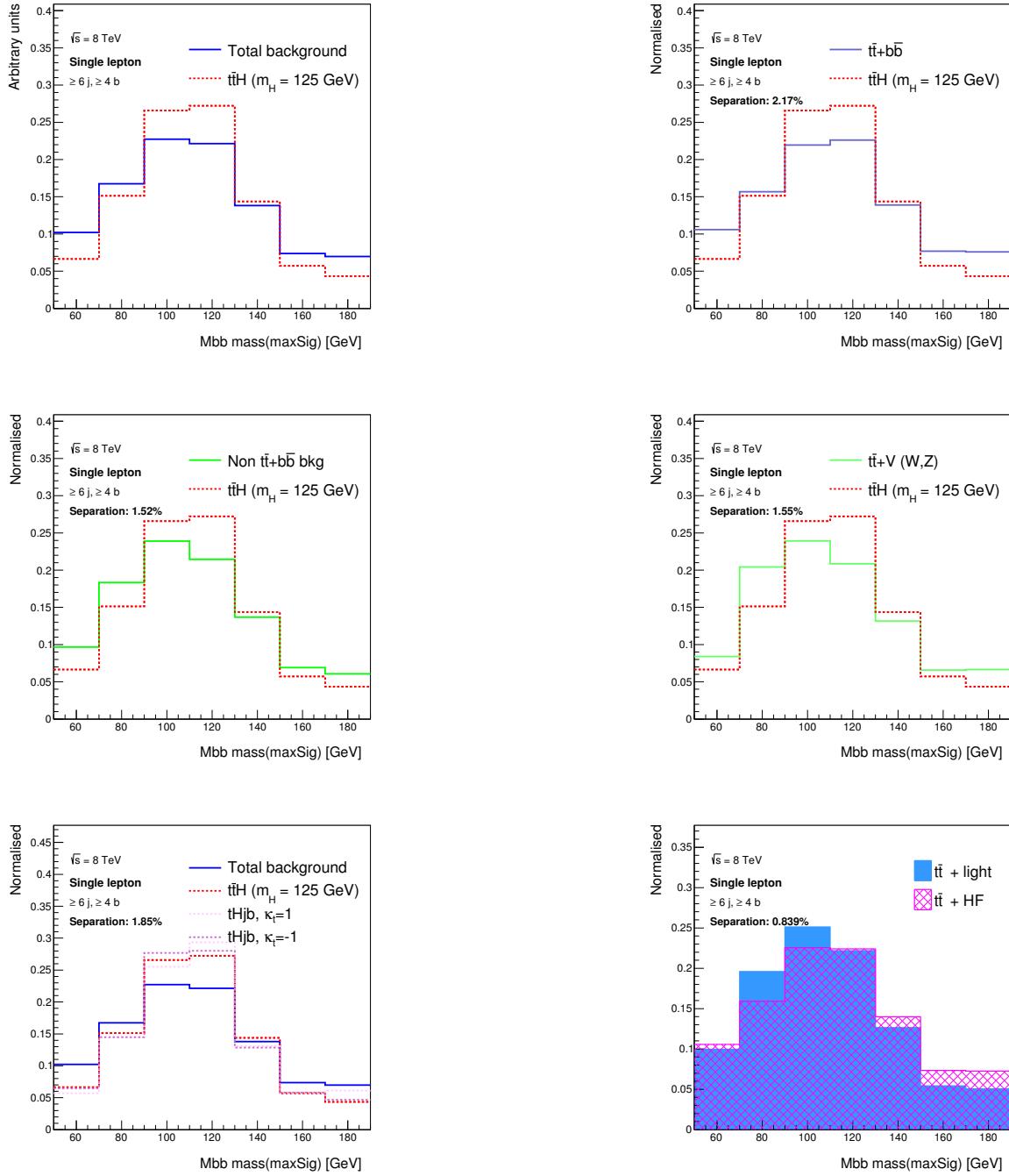


Figure A.5.: The invariant mass of additional b -jets using the $t\bar{t}H$ likelihood in the ($\geq 6j, \geq 4b$) region for the $t\bar{t}H$ signal prediction (dashed red) to various background processes, the single top Higgs boson production as well as $t\bar{t} + \text{HF}$ to $t\bar{t} + \text{LF}$ events.

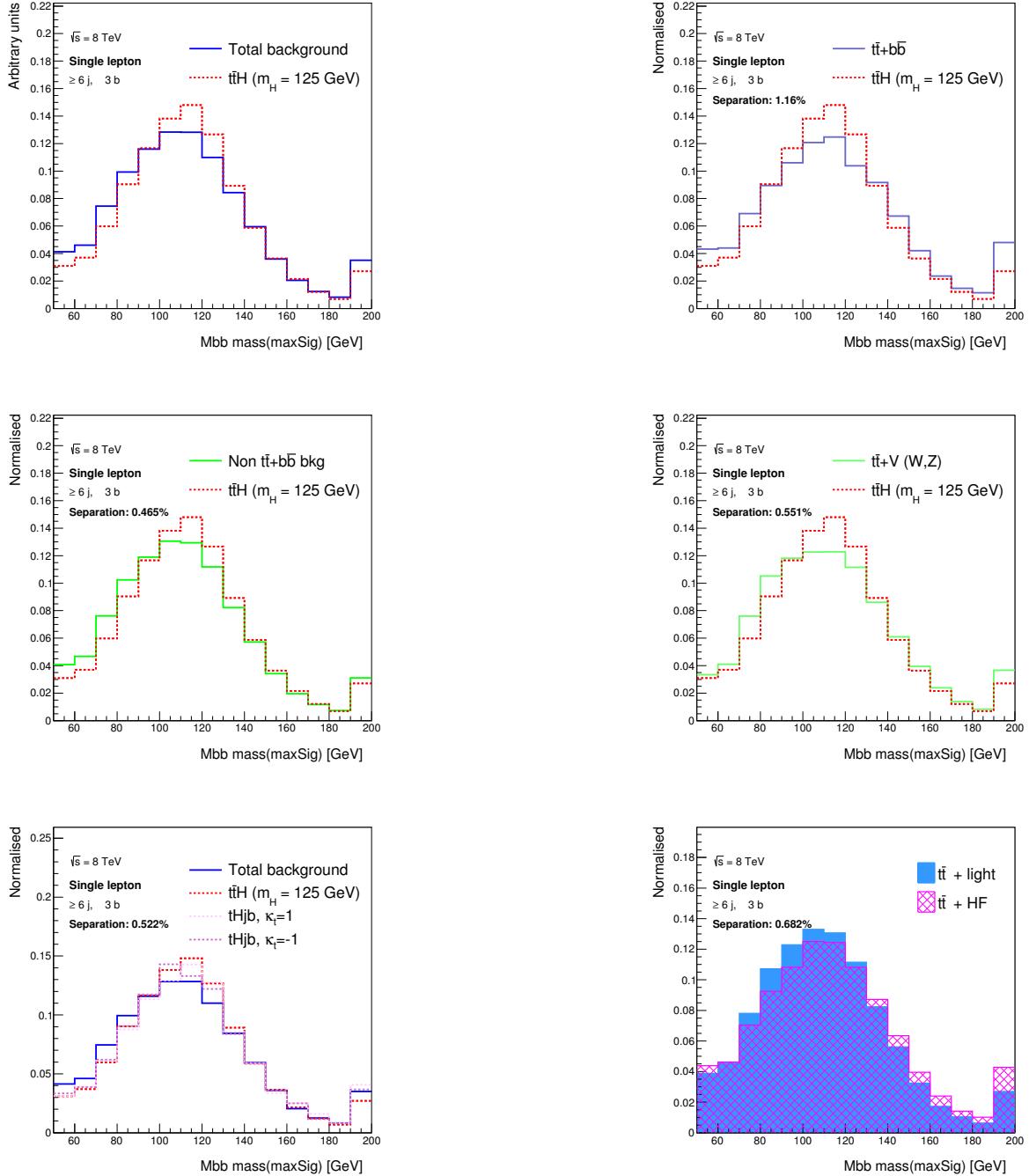


Figure A.6.: The invariant mass of additional b -jets using the $t\bar{t}H$ likelihood in the ($\geq 6 \text{j}, 3 \text{b}$) region for the $t\bar{t}H$ signal prediction (dashed red) to various background processes, the single top Higgs boson production as well as $t\bar{t} + \text{HF}$ to $t\bar{t} + \text{LF}$ events.

A. Further Separation Studies using the MEM

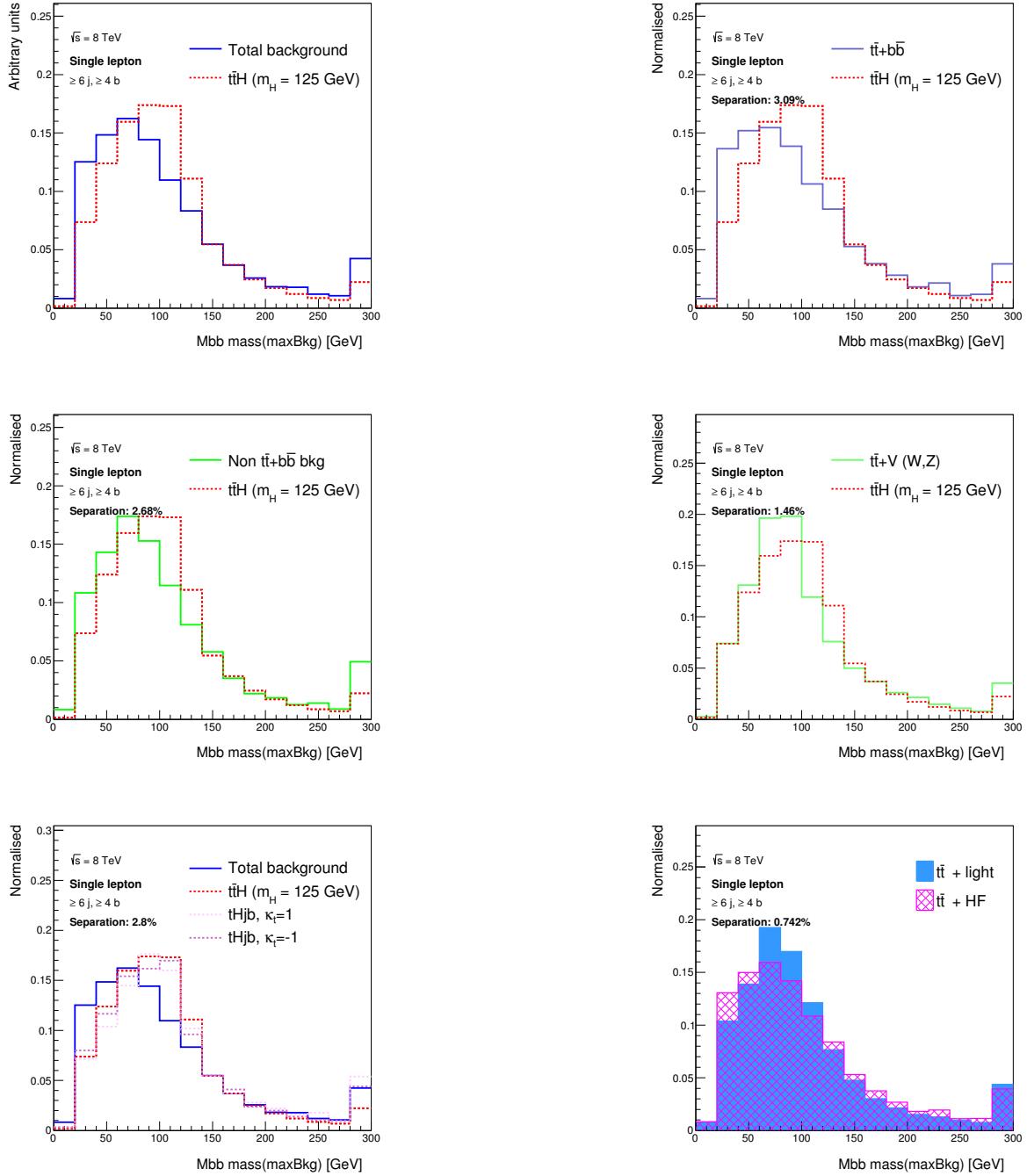


Figure A.7.: The invariant mass of additional b -jets using the $t\bar{t} + b\bar{b}$ likelihood in the ($\geq 6j, \geq 4b$) region for the $t\bar{t}H$ signal prediction (dashed red) to various background processes, the single top Higgs boson production as well as $t\bar{t} + \text{HF}$ to $t\bar{t} + \text{LF}$ events.

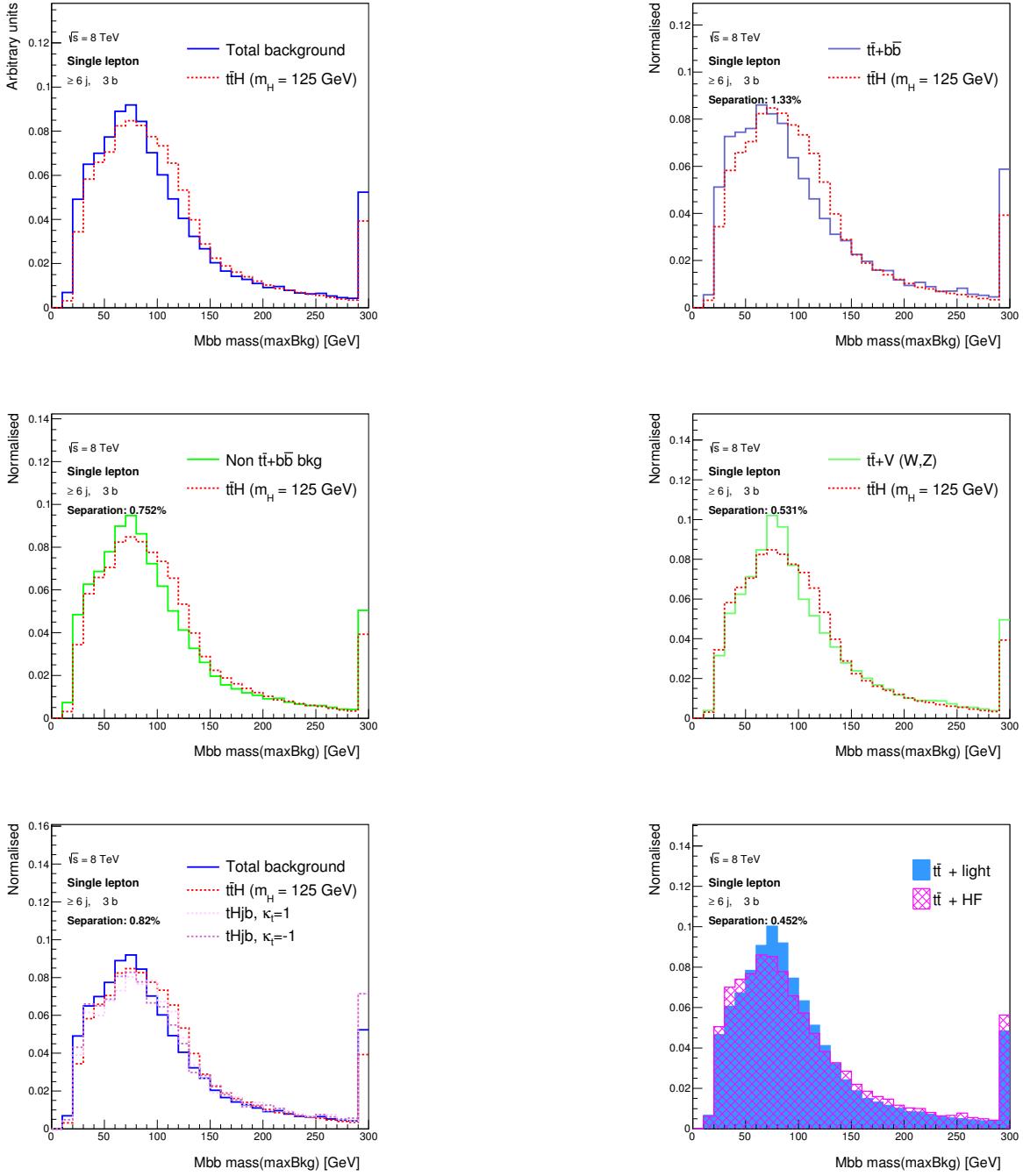


Figure A.8.: The invariant mass of additional b -jets using the $t\bar{t} + b\bar{b}$ likelihood in the ($\geq 6j, 3b$) region for the $t\bar{t}H$ signal prediction (dashed red) to various background processes, the single top Higgs boson production as well as $t\bar{t} + \text{HF}$ to $t\bar{t} + \text{LF}$ events.

A. Further Separation Studies using the MEM

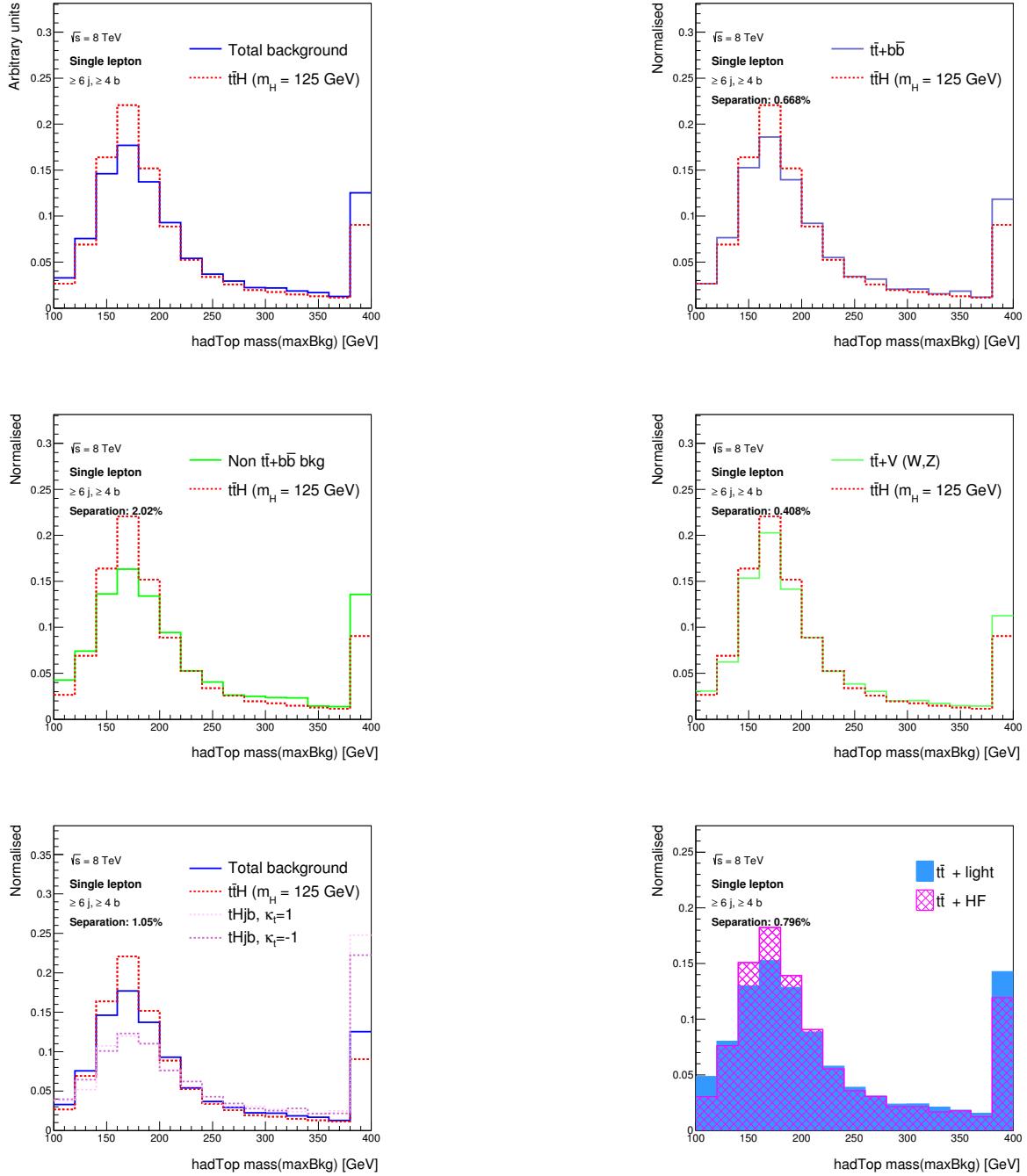


Figure A.9.: The hadronic top quark mass using the $t\bar{t} + b\bar{b}$ likelihood in the ($\geq 6 j, \geq 4 b$) region for the $t\bar{t}H$ signal prediction (dashed red) to various background processes, the single top Higgs boson production as well as $t\bar{t} + \text{HF}$ to $t\bar{t} + \text{LF}$ events.

A.4. Hadronic Top Quark Mass

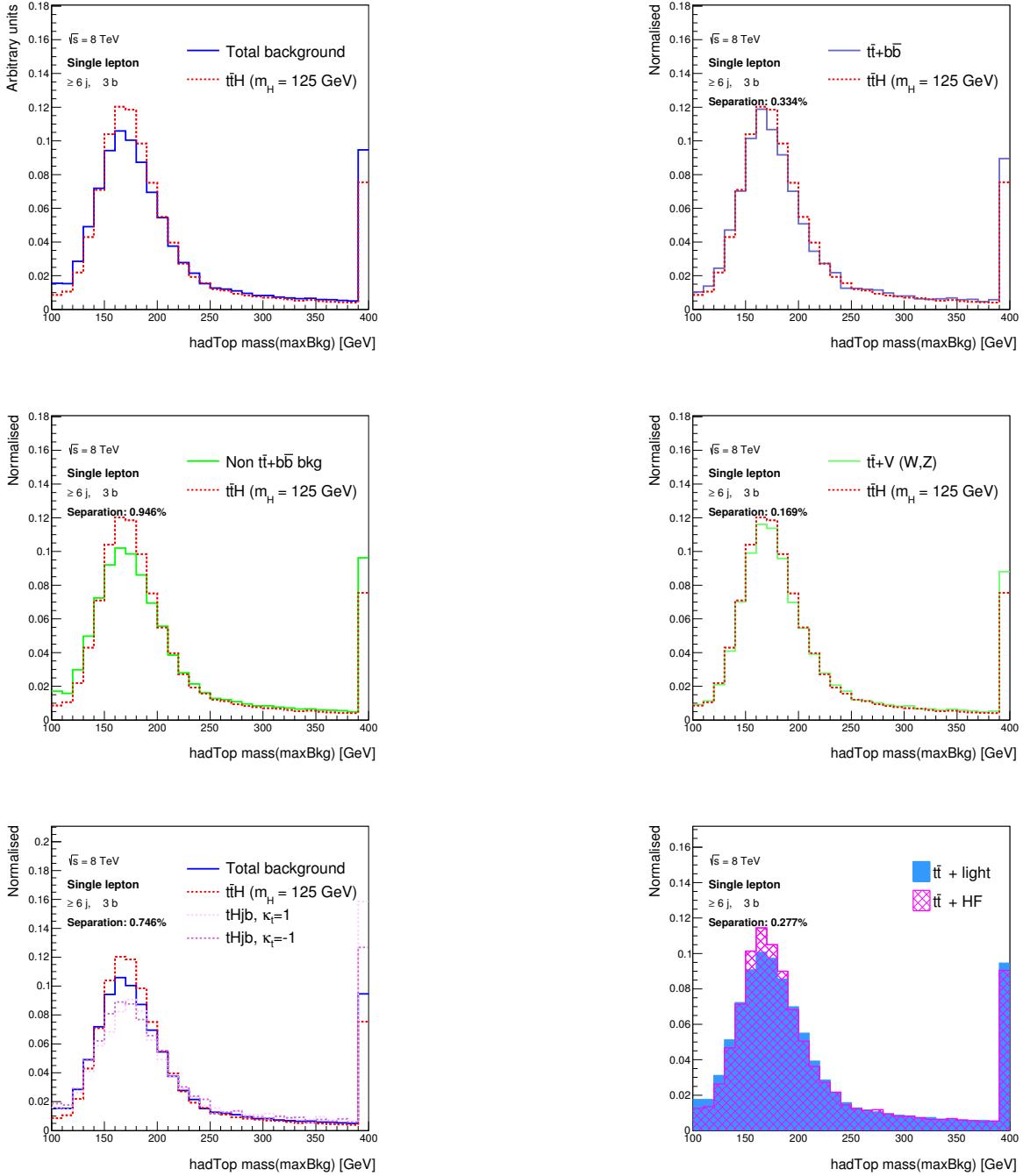


Figure A.10.: The hadronic top quark mass using the $t\bar{t} + b\bar{b}$ likelihood in the ($\geq 6j, 3b$) region for the $t\bar{t}H$ signal prediction (dashed red) to various background processes, the single top Higgs boson production as well as $t\bar{t} + \text{HF}$ to $t\bar{t} + \text{LF}$ events.

A.5. Hadronic W Boson Mass

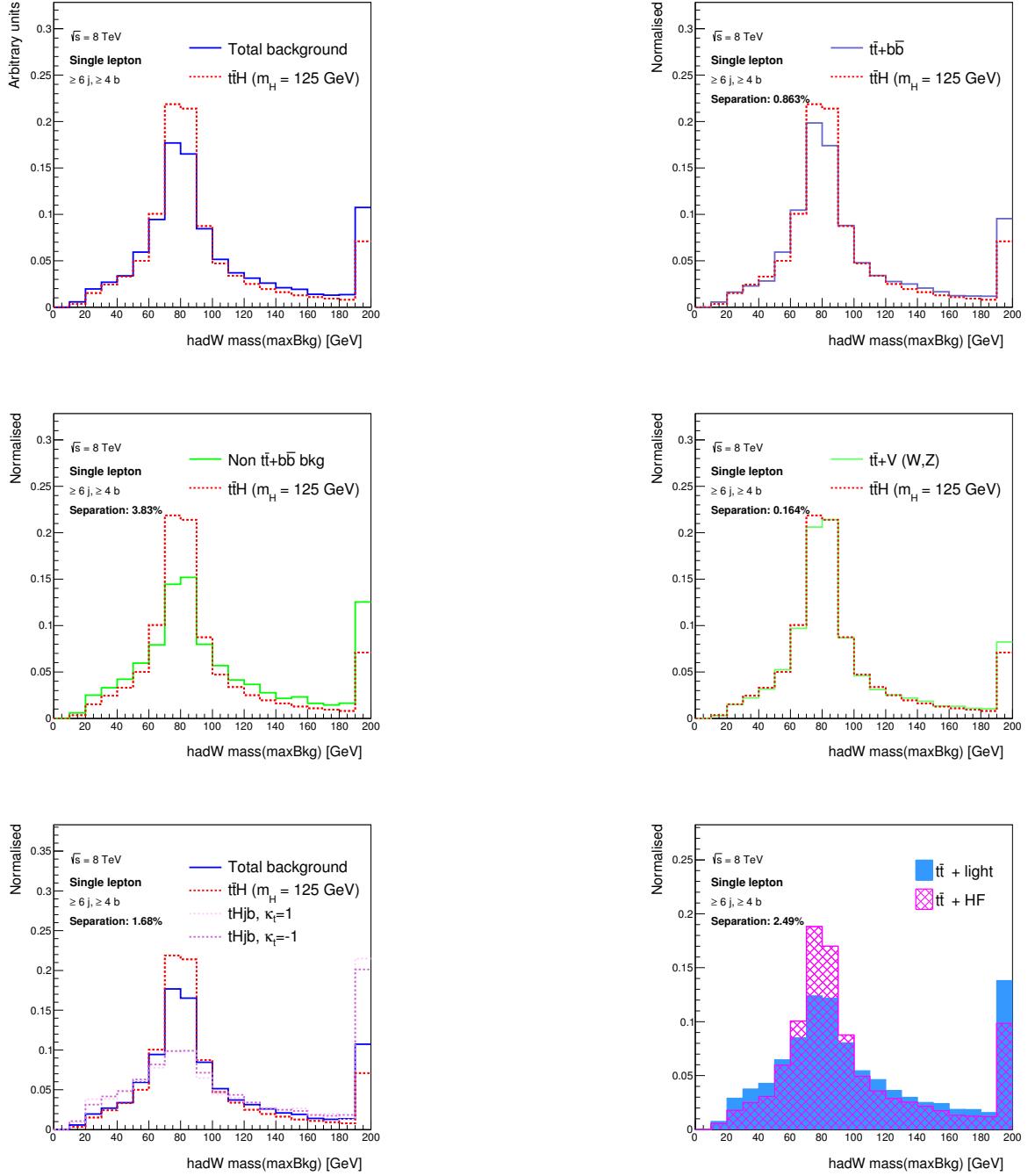


Figure A.11.: The hadronic W boson mass using the $t\bar{t} + b\bar{b}$ likelihood in the ($\geq 6j, \geq 4b$) region for the $t\bar{t}H$ signal prediction (dashed red) to various background processes, the single top Higgs boson production as well as $t\bar{t} + \text{HF}$ to $t\bar{t} + \text{LF}$ events.

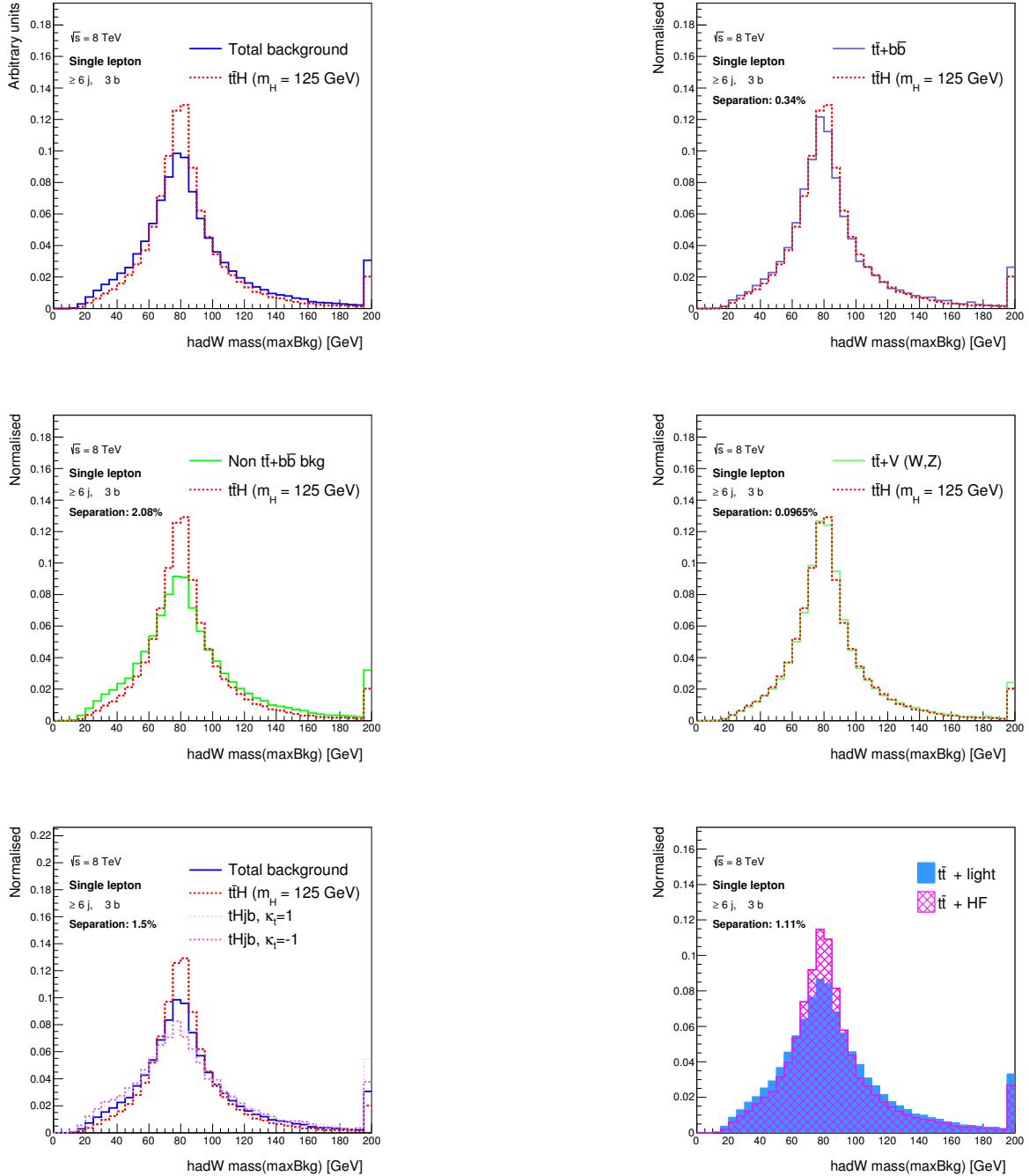


Figure A.12.: The hadronic W boson mass using the $t\bar{t} + b\bar{b}$ likelihood in the ($\geq 6j, 3b$) region for the $t\bar{t}H$ signal prediction (dashed red) to various background processes, the single top Higgs boson production as well as $t\bar{t} + \text{HF}$ to $t\bar{t} + \text{LF}$ events.

Further Event Reconstruction using the MEM

In Appendix B.1 the remaining angular distributions between the two additional jets not originating from the top quark system are shown in the ($\geq 6j, 3b$) and the ($\geq 6j, \geq 4b$) region. In Appendix B.2 reconstruction efficiencies studies at $\sqrt{s} = 7$ TeV are presented.

B.1. Angular Distributions of Additional B-jets

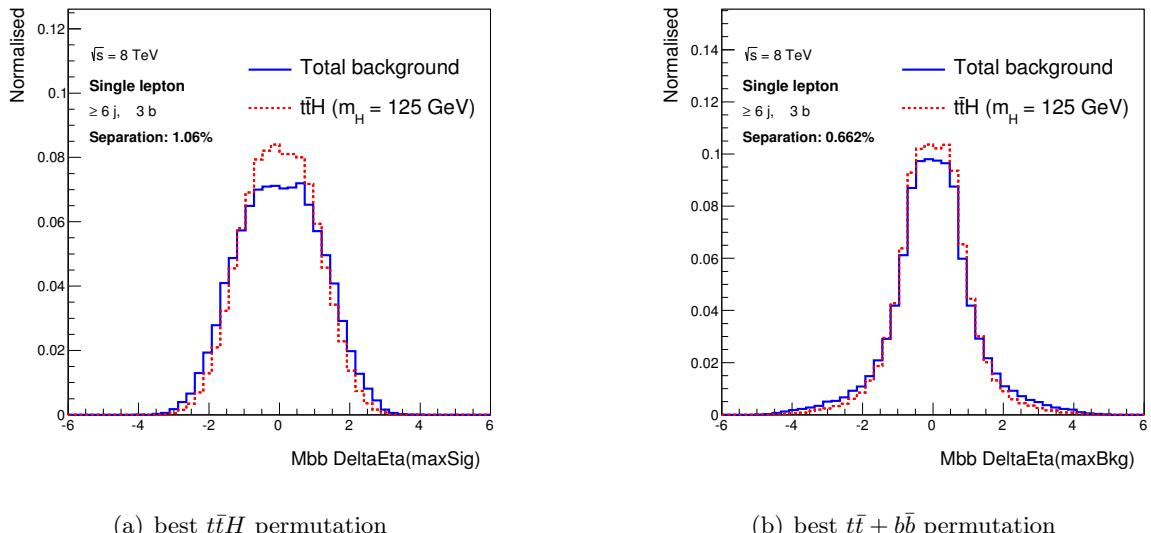


Figure B.1.: The reconstructed $\Delta\eta$ between the two additional b -jets in (a) best $t\bar{t}H$ permutation and (b) best $t\bar{t} + b\bar{b}$ permutation in the ($\geq 6j, 3b$) region for the total background (solid blue) and the $t\bar{t}H$ signal (dashed red) predictions.

B. Further Event Reconstruction using the MEM

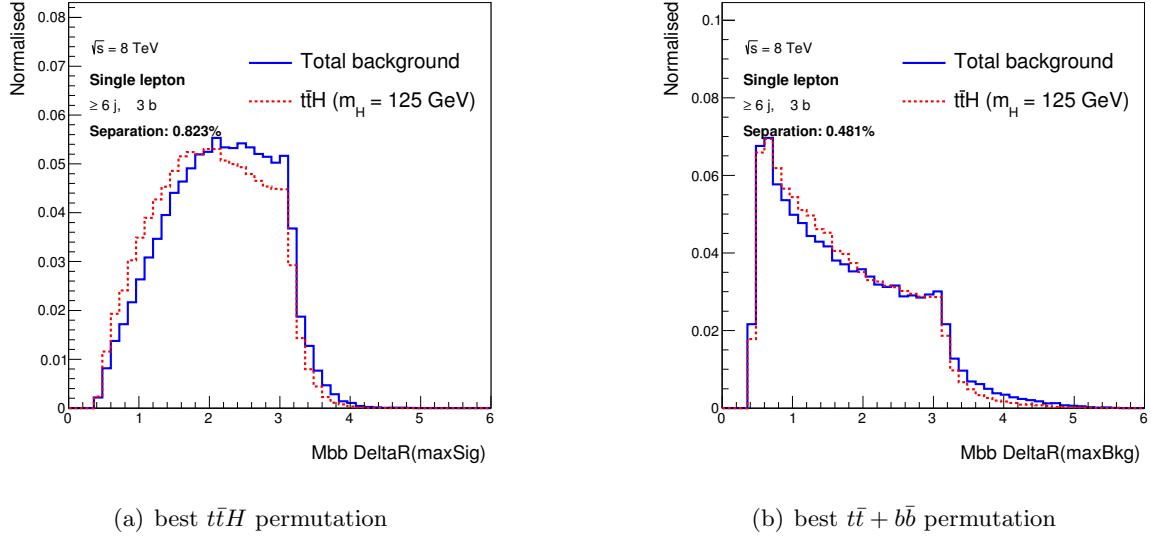


Figure B.2.: The reconstructed ΔR between the two additional b -jets in (a) best $t\bar{t}H$ permutation and (b) best $t\bar{t} + b\bar{b}$ permutation in the ($\geq 6j, 3b$) region for the total background (solid blue) and the $t\bar{t}H$ signal (dashed red) predictions.

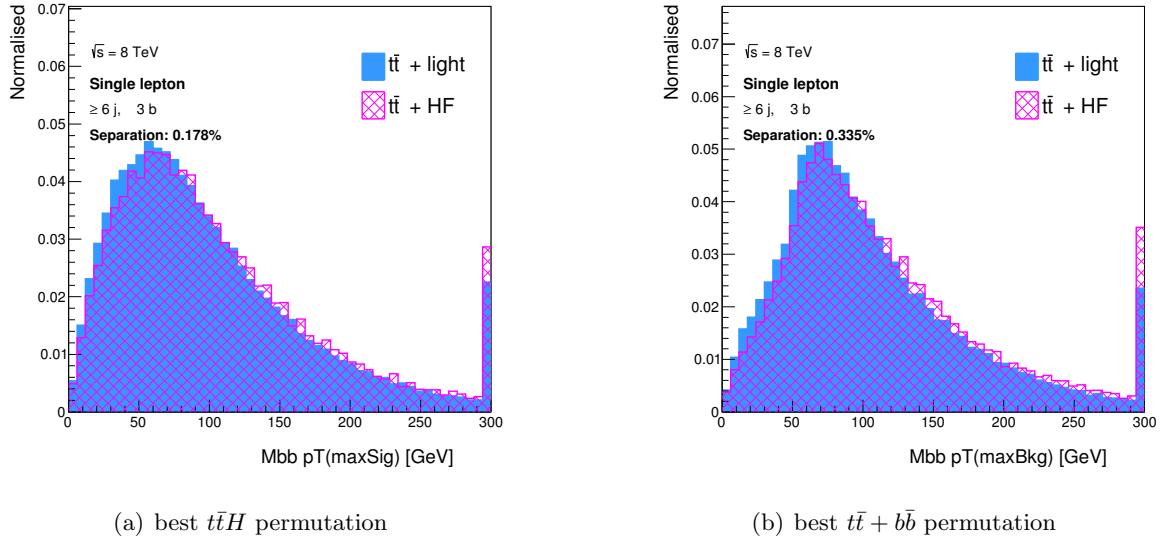


Figure B.3.: The reconstructed p_T between the two additional b -jets in (a) best $t\bar{t}H$ permutation and (b) best $t\bar{t} + b\bar{b}$ permutation in the ($\geq 6j, 3b$) region for the total background (solid blue) and the $t\bar{t}H$ signal (dashed red) predictions.

B.1. Angular Distributions of Additional B-jets

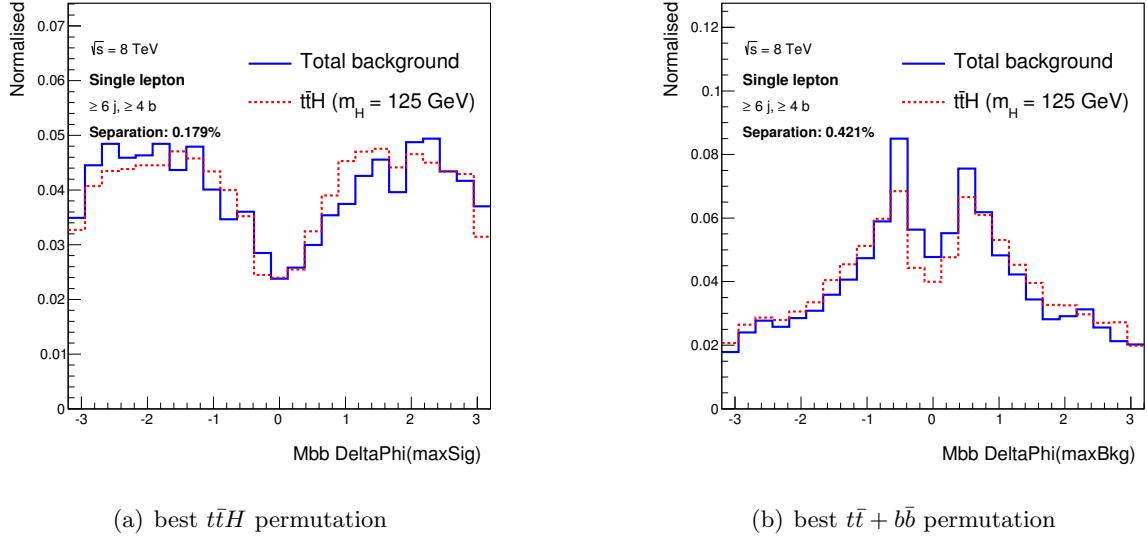


Figure B.4.: The reconstructed $\Delta\phi$ between the two additional b -jets in (a) best $t\bar{t}H$ permutation and (b) best $t\bar{t} + b\bar{b}$ permutation in the ($\geq 6j, \geq 4b$) region for the total background (solid blue) and the $t\bar{t}H$ signal (dashed red) predictions.

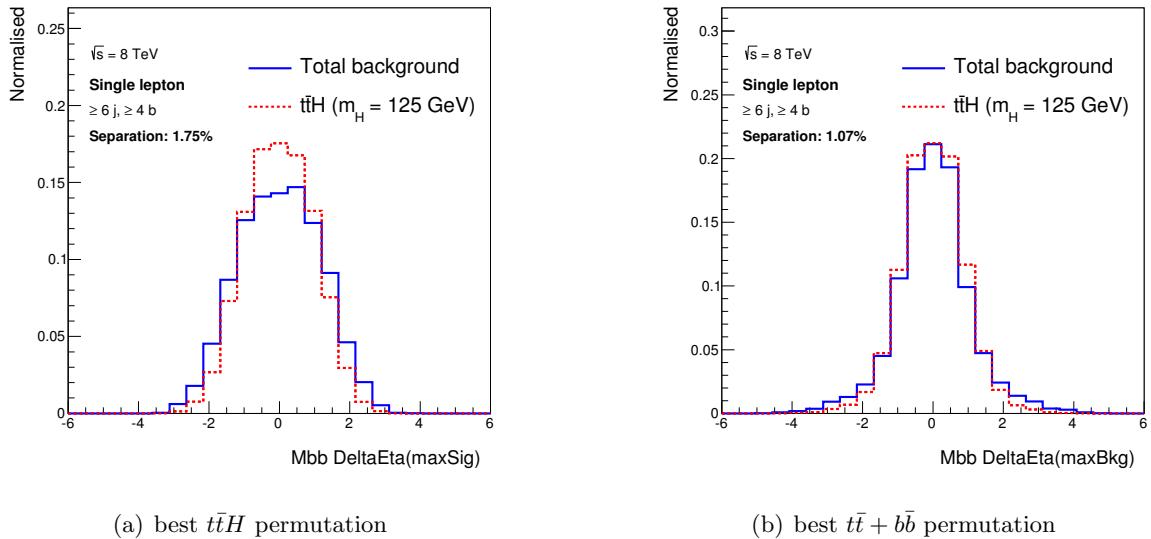


Figure B.5.: The reconstructed $\Delta\eta$ between the two additional b -jets in (a) best $t\bar{t}H$ permutation and (b) best $t\bar{t} + b\bar{b}$ permutation in the ($\geq 6j, \geq 4b$) region for the total background (solid blue) and the $t\bar{t}H$ signal (dashed red) predictions.

B. Further Event Reconstruction using the MEM

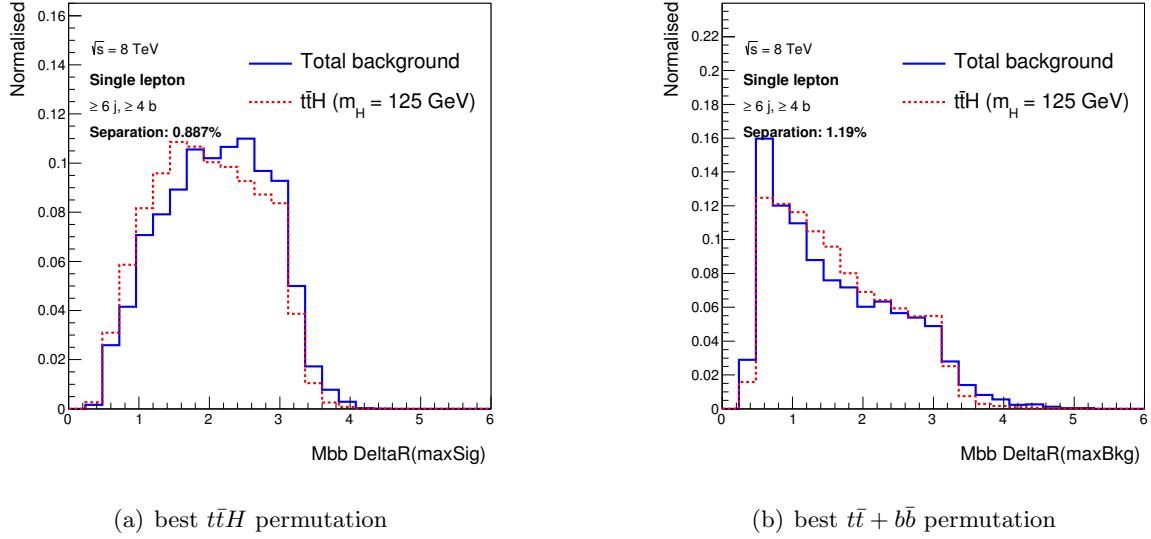


Figure B.6.: The reconstructed ΔR between the two additional b -jets in (a) best $t\bar{t}H$ permutation and (b) best $t\bar{t} + b\bar{b}$ permutation in the ($\geq 6j, \geq 4b$) region for the total background (solid blue) and the $t\bar{t}H$ signal (dashed red) predictions.

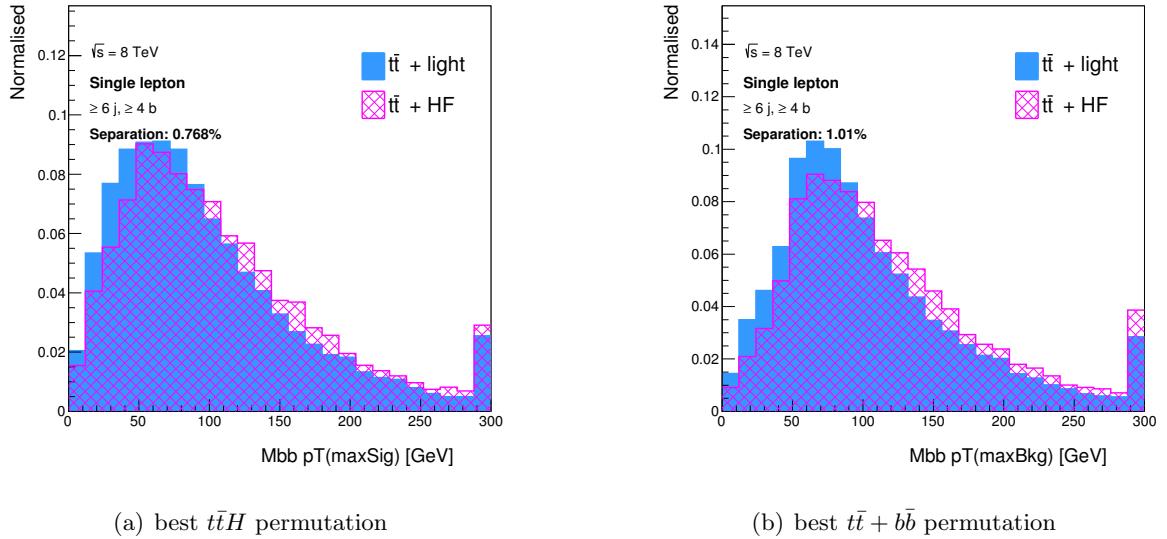


Figure B.7.: The reconstructed p_T between the two additional b -jets in (a) best $t\bar{t}H$ permutation and (b) best $t\bar{t} + b\bar{b}$ permutation in the ($\geq 6j, \geq 4b$) region for the total background (solid blue) and the $t\bar{t}H$ signal (dashed red) predictions.

B.2. Reconstruction Efficiencies

The reconstruction efficiencies are calculated based on events in which each parton level quark could be matched bi-uniquely with a reconstructed jet within $\Delta R < 0.3$. If the jet is assigned to the matched quark by the reconstruction algorithm the reconstruction is considered to be correct. In this way, the fractions of events in which the jets are correctly assigned to the Higgs boson H_{bb} , the g_{bb} , the hadronic W boson W_{jj} , the hadronic top quark t_{bjj} , the leptonic $t_{ble\nu}$ and all final states are calculated. The calculation is performed on $t\bar{t}H$ (Tab. B.1) and $t\bar{t} + b\bar{b}$ (Tab. B.2) events in the ($\geq 6j, \geq 4b$) region at $\sqrt{s} = 7$ TeV using the $t\bar{t}H$ and $t\bar{t} + b\bar{b}$ hypotheses of the MEM and the KLFITTER [188] reconstruction algorithm assuming only a $t\bar{t}$ system without additional radiations.

In the ($\geq 6j, \geq 4b$) region, the assignment to the W_{jj} is defined by the b -tagging algorithm and the choice of the additional light jets. However, for the studies with KLFITTER a looser b -tagging is applied resulting in less mistags of the light jets and hence more events with correctly assigned jets to W_{jj} . As a consequence, the comparisons is not entirely fair, because the MEM start with poorer initial conditions of only 70.3% and 77.2% of the $t\bar{t}H$ and $t\bar{t} + b\bar{b}$ events, respectively, which have a chance to be correctly reconstructed. Nevertheless, the MEM outperforms KLFITTER in the reconstruction of $t\bar{t}H$ events using the $t\bar{t}H$ hypothesis and of $t\bar{t} + b\bar{b}$ events using the $t\bar{t} + b\bar{b}$ hypothesis. The best reconstruction of the Higgs boson in $t\bar{t}H$ events can be obtained using the $t\bar{t}H$ hypothesis. The reconstruction of $t\bar{t} + b\bar{b}$ events using the $t\bar{t}H$ hypothesis is poor, since the Higgs boson divergence forces two jets close to the Higgs boson mass to be assigned to H_{bb} . As a result, the in only 12.3% of the events the correct jets are assigned to the gluon and the two top quarks are poorly reconstructed. The reconstruction efficiencies for $t\bar{t} + b\bar{b}$ events are significantly higher using the $t\bar{t} + b\bar{b}$ hypothesis resulting in 47.0% of the events which are completely correct reconstructed. Because much more $t\bar{t} + b\bar{b}$ events are present in the ($\geq 6j, \geq 4b$) region, this study suggest to use the $t\bar{t} + b\bar{b}$ hypothesis for an event reconstruction.

Method	Reconstruction Efficiency [%]				
	H_{bb}	W_{jj}	t_{bjj}	$t_{ble\nu}$	all
MEM ($t\bar{t}H$)	41.5	70.3	53.7	55.5	33.5
MEM ($t\bar{t} + b\bar{b}$)	34.7	70.3	52.2	52.9	27.0
KLFitter ($t\bar{t}$)	34.3	75.2	52.9	50.2	28.5

Table B.1: The reconstruction efficiencies of the intermediate states and all final states of $t\bar{t}H$ events in the ($\geq 6j, \geq 4b$) region at $\sqrt{s} = 7$ TeV comparing the MEM reconstruction efficiency using the $t\bar{t}H$ hypothesis, the $t\bar{t} + b\bar{b}$ hypothesis and a reconstruction efficiency obtained with KLFITTER [188]

Method	Reconstruction Efficiency [%]				
	g_{bb}	W_{jj}	t_{bjj}	$t_{ble\nu}$	all
MEM ($t\bar{t}H$)	12.3	77.2	38.3	40.4	10.4
MEM ($t\bar{t} + b\bar{b}$)	58.3	77.2	63.1	62.8	47.0
KLFitter ($t\bar{t}$)	38.8	80.5	55.0	53.7	33.0

Table B.2: The reconstruction efficiencies of the intermediate states and all final states of $t\bar{t} + b\bar{b}$ events in the ($\geq 6j, \geq 4b$) region at $\sqrt{s} = 7$ TeV comparing the MEM reconstruction efficiency using the $t\bar{t}H$ hypothesis, the $t\bar{t} + b\bar{b}$ hypothesis and a reconstruction efficiency obtained with KLFitter [188]

C

Neural Networks with MEM Variables

C.1. Saturation of the NN

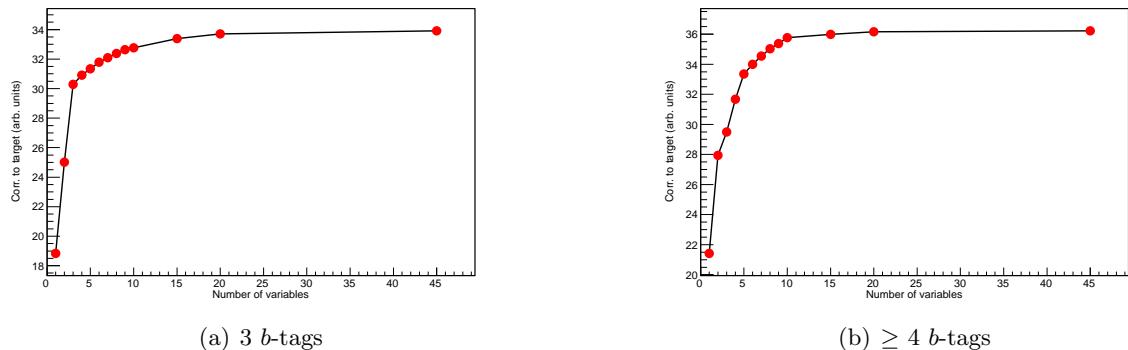


Figure C.1.: The evolution of NN discrimination with the addition of kinematic variables in the ($\geq 6j$, 3b) (a) and ($\geq 6j$, $\geq 4b$) (b) regions.

C.2. Alternative NN Strategies

Figure C.2 shows the NN output for six different scenarios of using kinematic and MEM in the neural network in the ($\geq 6j$, $\geq 4b$) region. Various sets of input variables are used in NEUROBAYES to chose the best variables for maximal separation of the NN output. In Figure C.2 (a) around 60 MEM variables are used without any kinematic variables and in Figure C.2 (b) only the 300 kinematic variables are used without MEM variables. In Figure C.2 (c) the $D1$ variable is added to the best ten kinematic variables and in Figure C.2 (d) the $D1$ and $\ln \mathcal{L}_{t\bar{t}H}^{\text{sum}}$ variables are added to the best ten kinematic variables. In Figure C.2 (e) the best ten are chosen out of the best ten kinematic variables plus all MEM variables and Figure C.2 (f) shows the ideal scenario of using all kinematic and MEM variables to let NEUROBAYES find

C. Neural Networks with MEM Variables

the best ten. The separation of 15.4% in the ideal scenario (f) is only slightly better than the separation of 14.8% in scenario (d) which is the chosen scenario for this analysis. Compared to the separation of 11.7% of the purely kinematic NN (b) this is a large improvement. The NN in which only MEM are used yields also a better separation of 13.3% than the purely kinematic NN. The addition of the $D1$ variable yields the largest improvement.

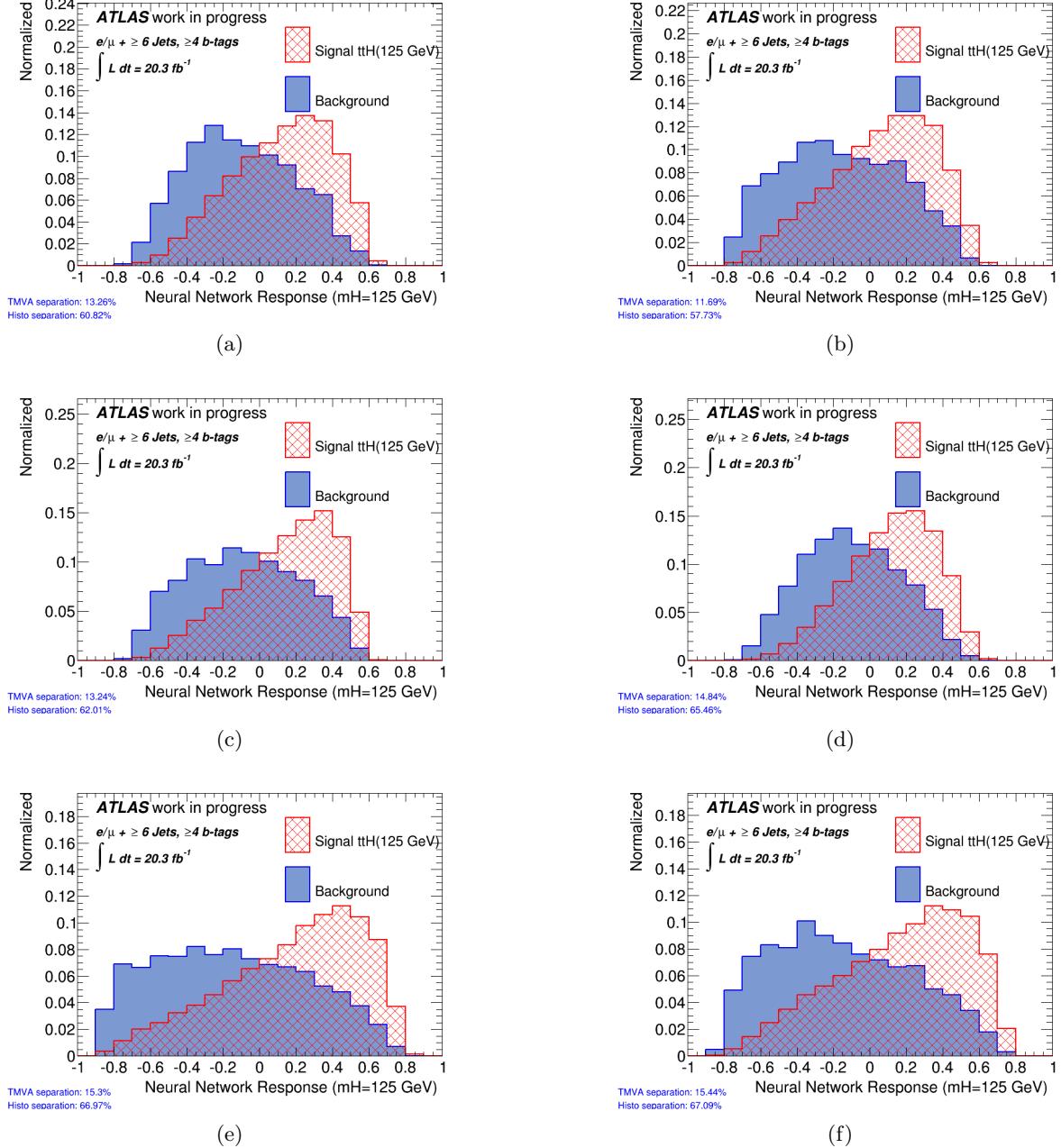


Figure C.2.: The NN output using the best ten variables from only MEM (a), purely kinematic (b), best kinematic + D1 (c), best kinematic + D1 + SSLL (d), best kinematic + all MEM (e) and all kinematic + all MEM (f) variables in the ($\geq 6j, \geq 4b$) region, comparing the total background (solid blue) and the $t\bar{t}H$ signal (dashed red) predictions.

C.3. Correlation among NN Variables

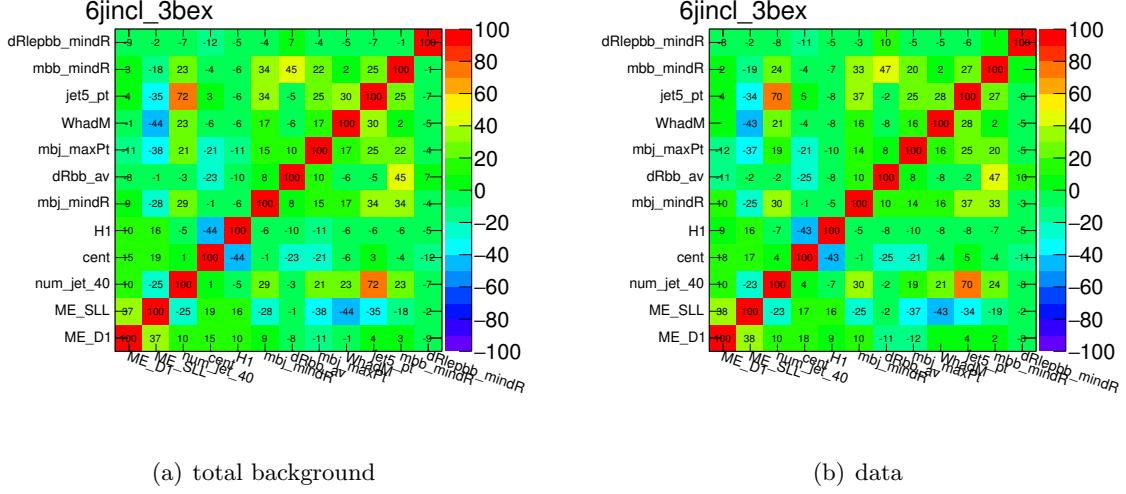


Figure C.3.: The two dimensional correlation between the NN used in the ($\geq 6j, 3b$) region

C.4. Overtraining Test of NN Output

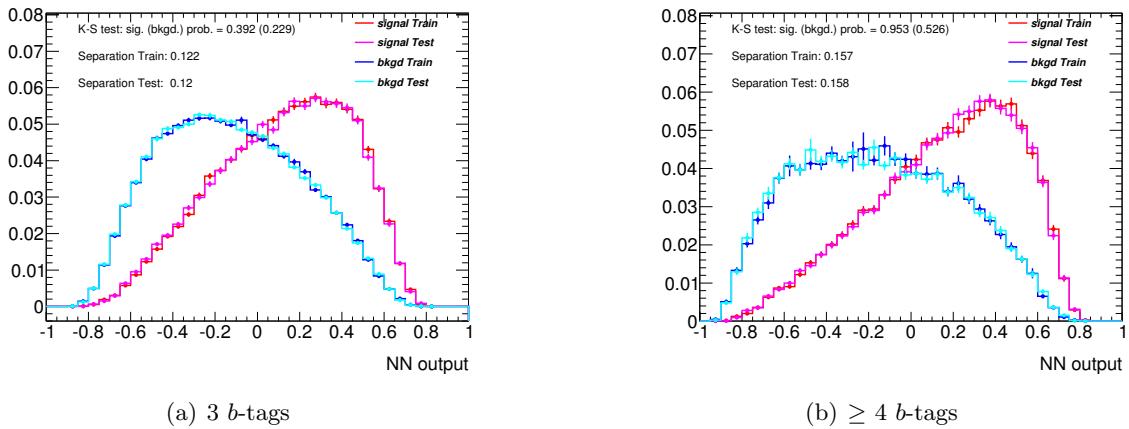


Figure C.4.: Overtraining test of NN output comparing the training and the evaluation in the ($\geq 6j, 3b$) (a) and ($\geq 6j, \geq 4b$) (b) regions

\mathcal{D}

Additional Material from the Profile Likelihood Fit

D.1. Nuisance Parameter

D.1.1. Pull and Constraints

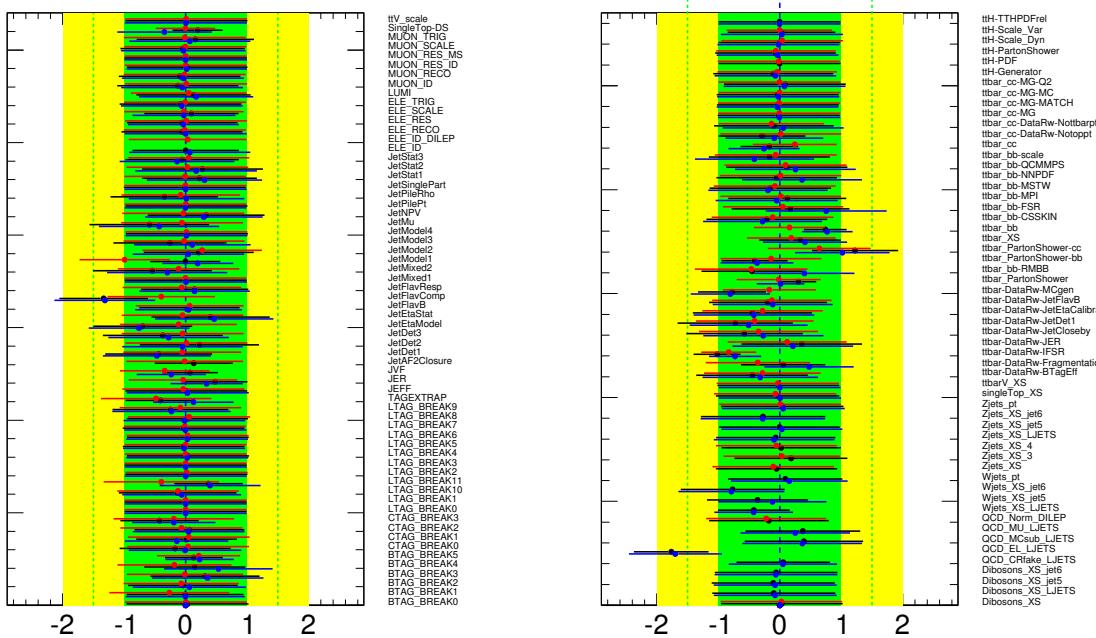


Figure D.1.: The pull and constraints of all NP considered in the fit showing the result of the single lepton (blue), dilepton (red) channel and the combination (black)

D.1.2. Correlations

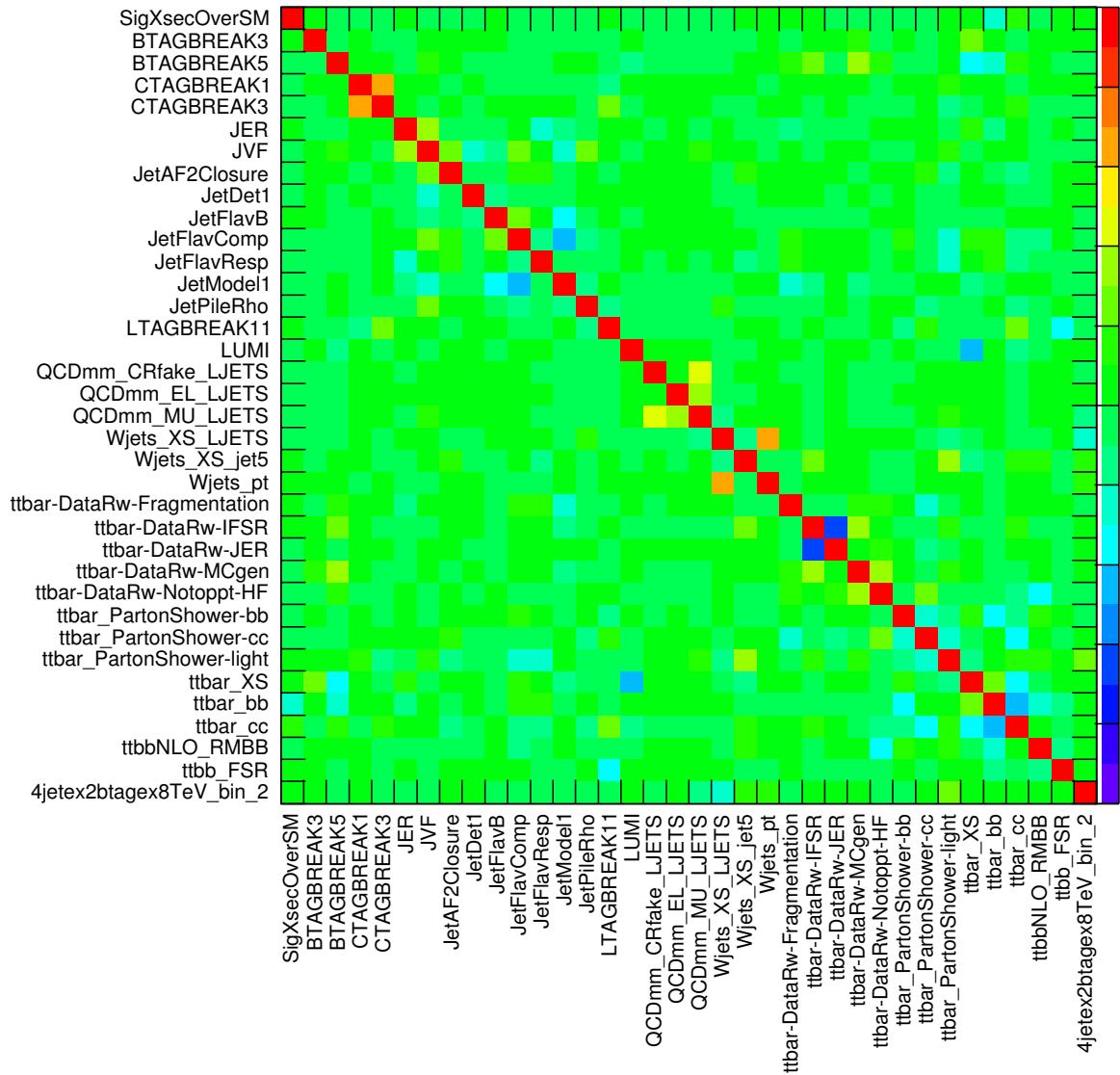


Figure D.2.: The correlations between the most important NP, where red means a correlation of 100%, green means no correlation (0%) and violet means an anti-correlation of 100%.

D.2. Excess in Muon Events

An excess of data events has been observed exclusively in events with one muon and only in the ($\geq 6j, 3b$) analysis region. This excess is not visible in events with electrons or in any region with other jet or b -tag multiplicities. In particular the high statistics ($\geq 6j, 2b$) region, which is kinematically most comparable to the ($\geq 6j, 3b$) region shows very good and consistent agreement between the electron and the muon channel. Extensive studies have been performed to investigate this excess, however no hint of the origin could be found. The checks included studies of muon isolation and quality, second lepton veto, cut variations, using alternative MC generator for $t\bar{t}$, dependence on data taking periods, semileptonic jet correction. The studies proved that the excess is not related to $t\bar{t}$ mis-modelling or is originating from non-prompt muons. Furthermore, the excess could not be identified as being particular signal- or background-like since it shows up in different regions depending on the variable distributions. The excess is flat in the NN output distribution and it is proven to be not related to the MEM variables. As a consequence, the excess is concluded to be a statistical fluctuation in data. The excess has been quantified by assuming lepton universality and subtracting the NN output observed in data in the electron channel from the distribution in the muon channel taking the different efficiencies in the event selection into account. This estimate yields an excess of 232 ± 76 events, which corresponds to a significance with respect to the MC prediction of about 2.5σ . The total bias on the extracted signal strength is -0.2 , which is negligible compared to the current total uncertainty on the measurement.

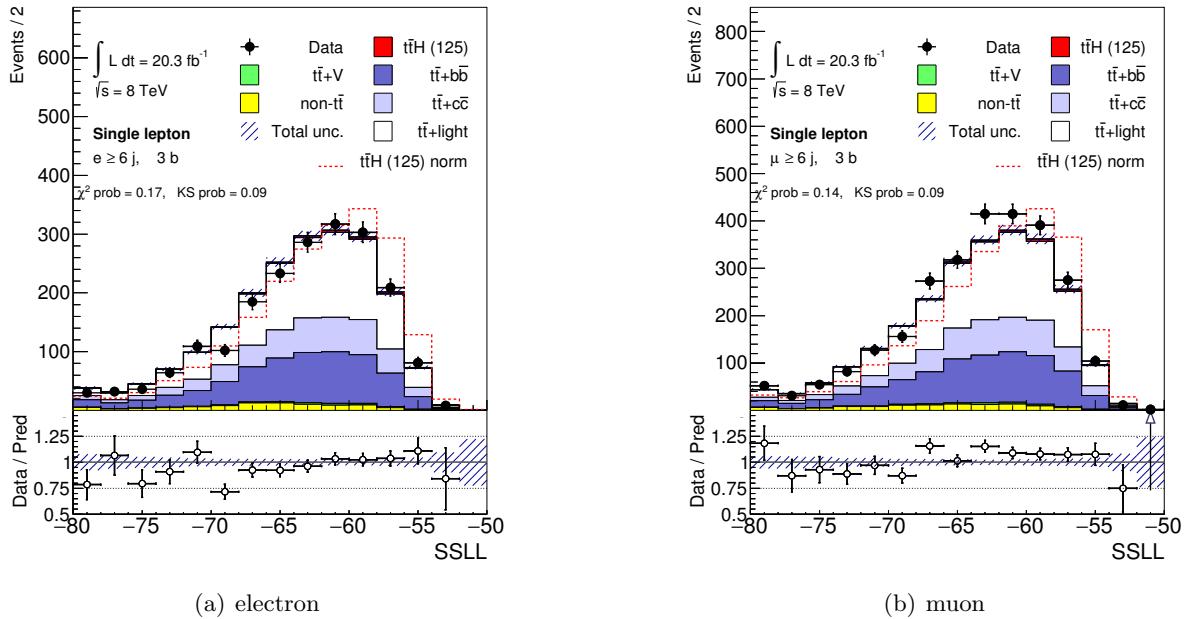


Figure D.3.: The logarithm of the summed signal likelihoods is shown in the electron (a) and muon (b) channel in the ($\geq 6j, 3b$) region, comparing the total predictions after the fit with data. The hashed area represents the post-fit uncertainty on the prediction.

D. Additional Material from the Profile Likelihood Fit

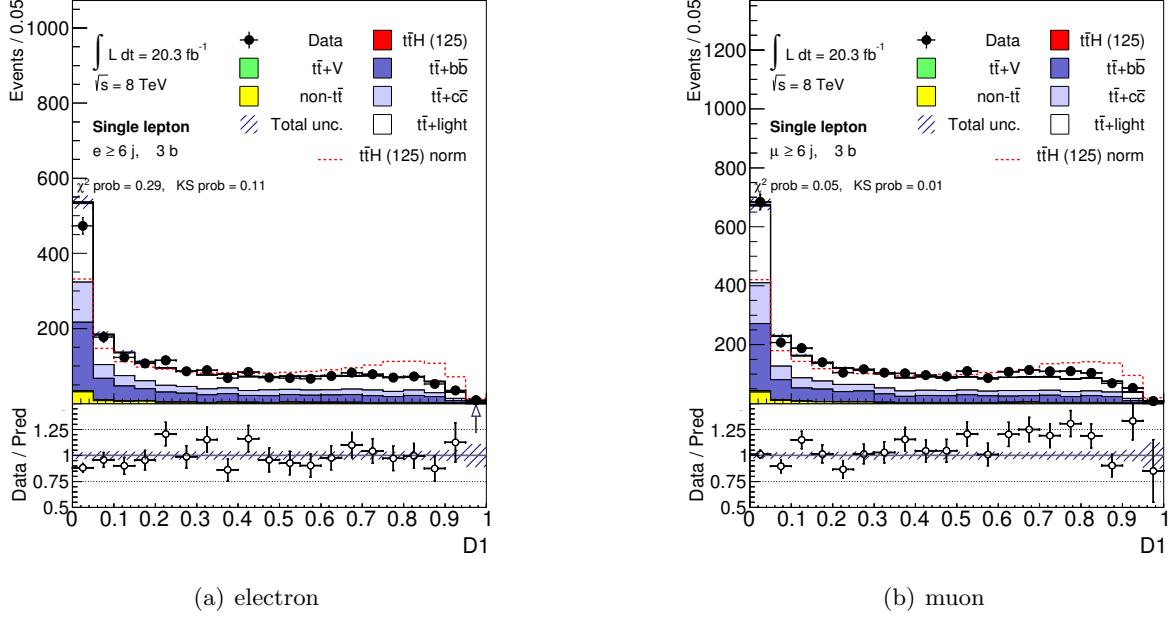


Figure D.4.: The likelihood ratio $D1$ using the summed likelihoods is shown in the electron (a) and muon (b) channel in the ($\geq 6j, 3b$) region, comparing the total predictions after the fit with data. The hashed area represents the post-fit uncertainty on the prediction.

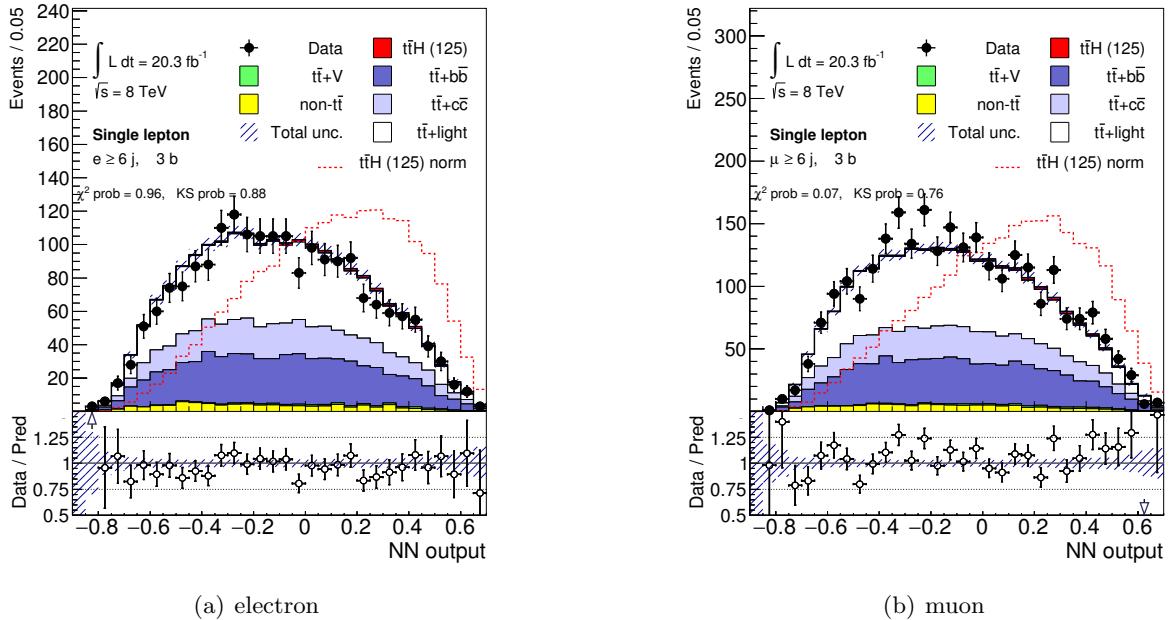


Figure D.5.: The NN output using only kinematic variables s is shown in the electron (a) and muon (b) channel in the ($\geq 6j, 3b$) region, comparing the total predictions after the fit with data. The hashed area represents the post-fit uncertainty on the prediction.

D.3. Validation of MEM Variables

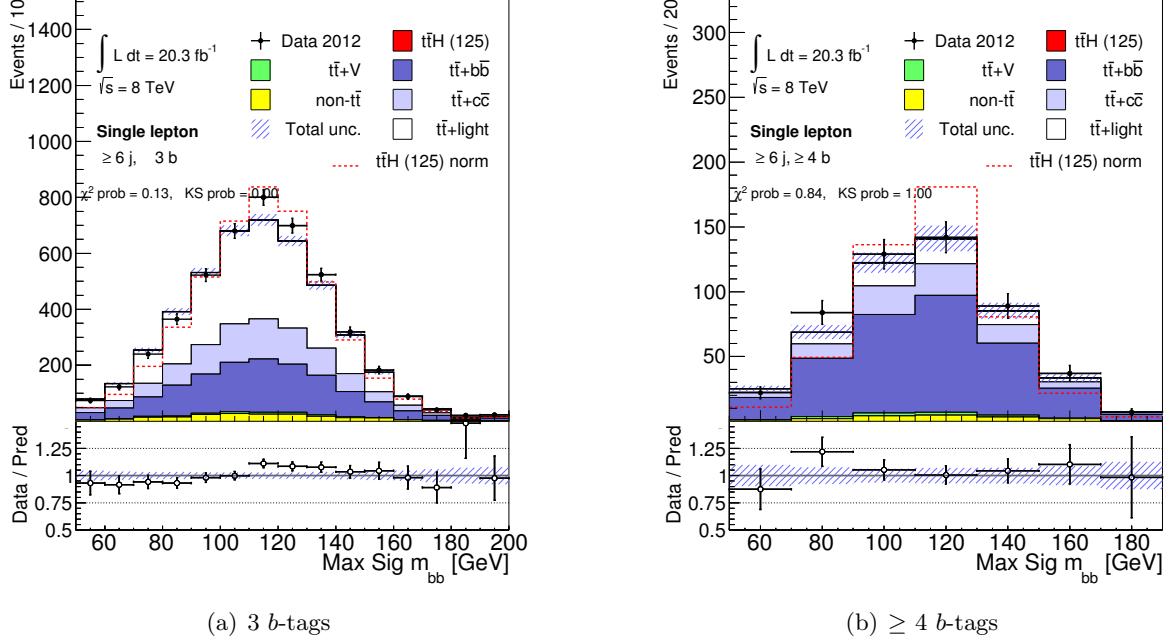


Figure D.6.: The reconstructed invariant mass m_{bb} using the $t\bar{t}H$ likelihood is shown in the ($\geq 6j, 3b$) (a) and ($\geq 6j, \geq 4b$) (b) region, comparing the total predictions after the fit with data. The $t\bar{t}H$ prediction is shown normalised to the fitted $\hat{\mu}$ (filled red) and to the total background prediction (dashed red). The hashed area represents the post-fit uncertainty on the prediction.

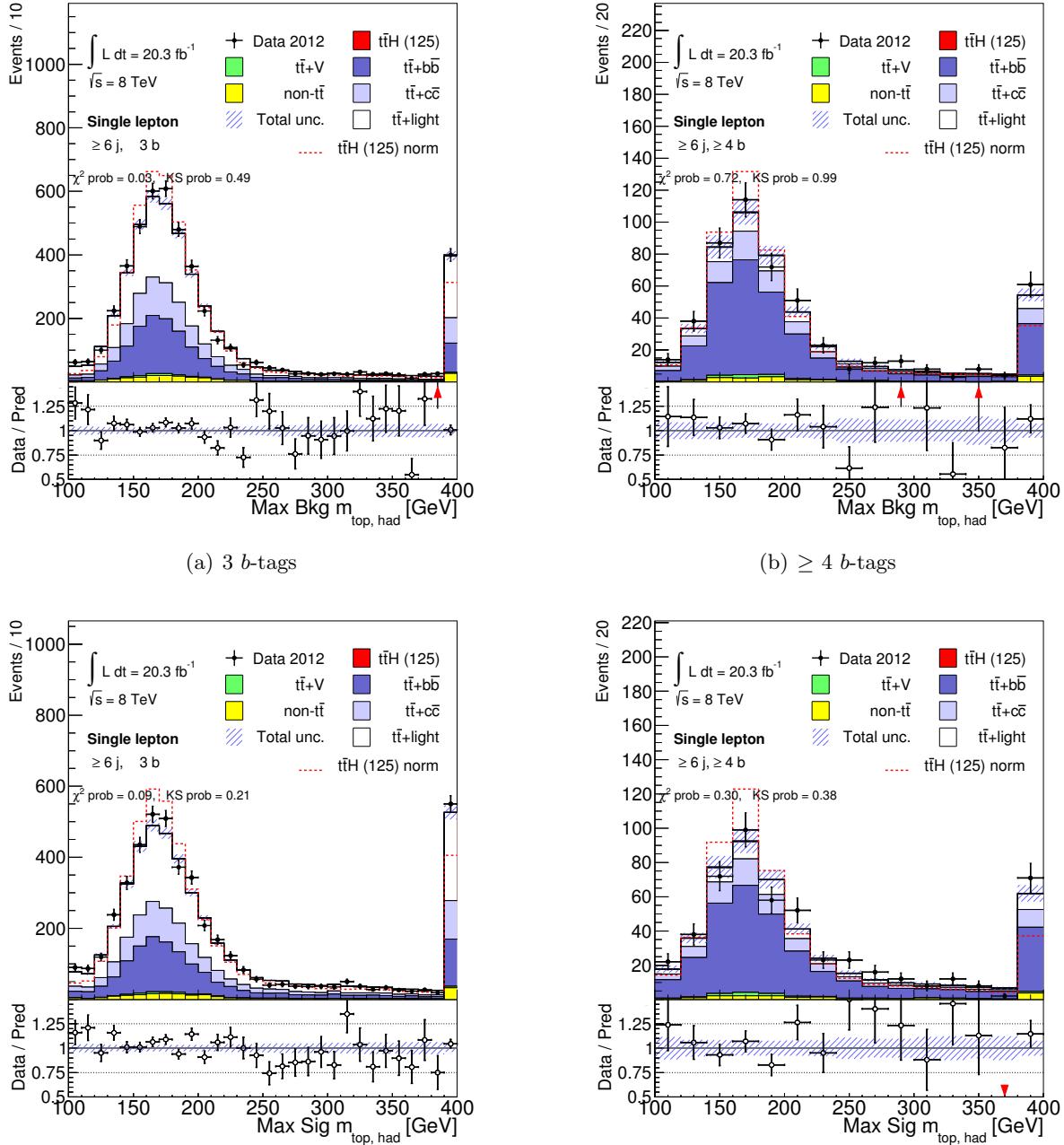


Figure D.7.: The reconstructed hadronic top quark mass using the $t\bar{t} + b\bar{b}$ likelihood (upper) and $t\bar{t}H$ likelihood (lower) is shown in the ($\geq 6j, 3b$) (a) and ($\geq 6j, \geq 4b$) (b) region, comparing the total predictions after the fit with data. The $t\bar{t}H$ prediction is shown normalised to the fitted $\hat{\mu}$ (filled red) and to the total background prediction (dashed red). The hashed area represents the post-fit uncertainty on the prediction.

D. Additional Material from the Profile Likelihood Fit

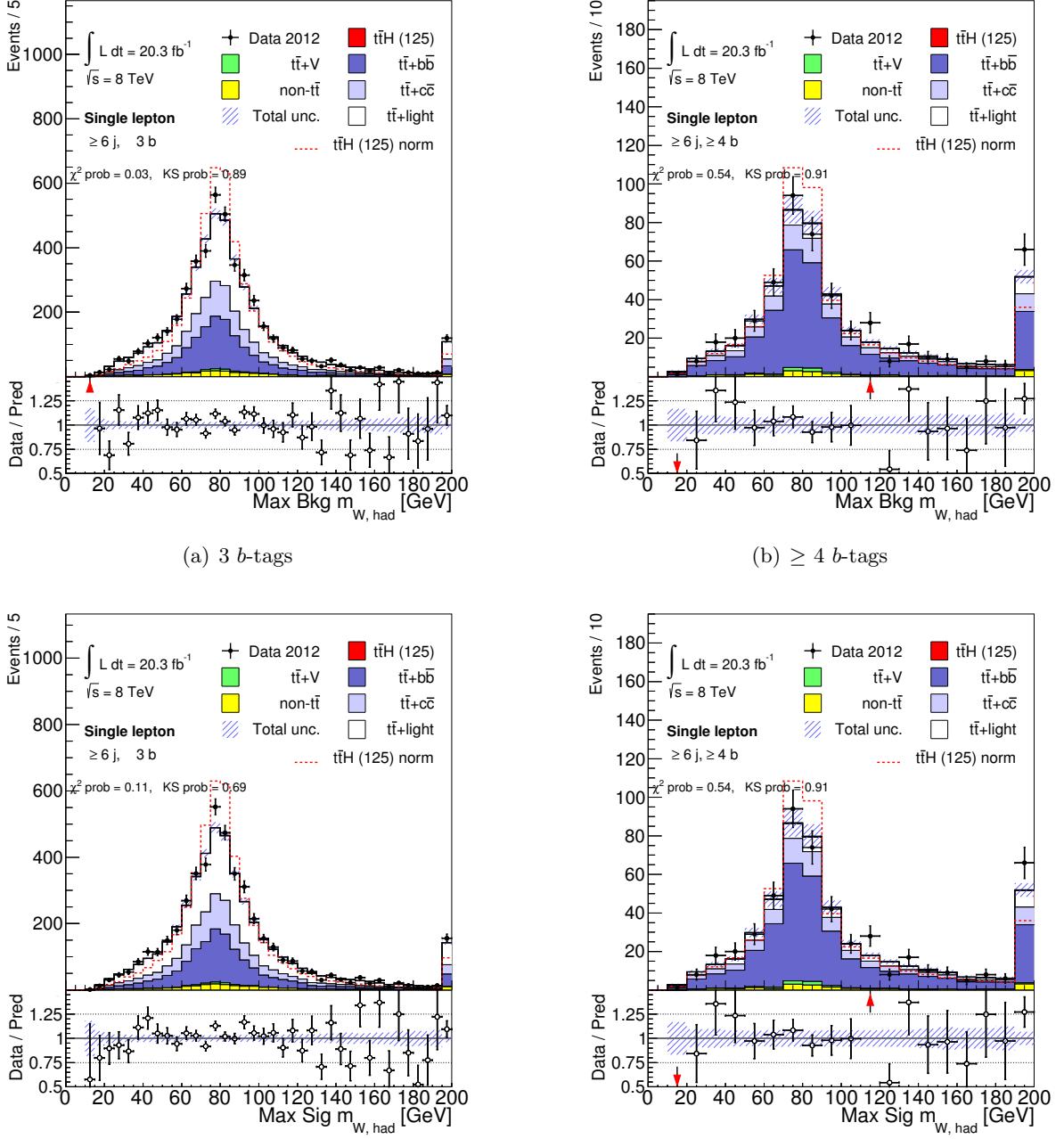


Figure D.8.: The reconstructed hadronic W boson mass using the $t\bar{t} + b\bar{b}$ likelihood (upper) and $t\bar{t}H$ likelihood (lower) is shown in the ($\geq 6j, 3b$) (a) and ($\geq 6j, \geq 4b$) (b) region, comparing the total predictions after the fit with data. The $t\bar{t}H$ prediction is shown normalised to the fitted $\hat{\mu}$ (filled red) and to the total background prediction (dashed red). The hashed area represents the post-fit uncertainty on the prediction.

D.4. Validation of NN Variables

Figures D.9 – D.9 show a comparison of data and prediction for the four most important input variables to the NN in the region where it is employed. All of the plots are made using post-fit predictions.

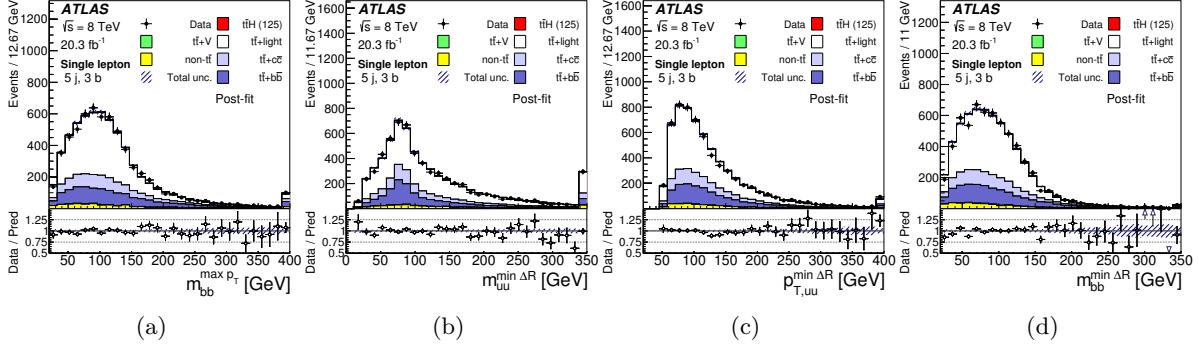


Figure D.9.: Post-fit comparison of data and prediction for the four top-ranked input variables in the (5j, 3b) region. The plots include (a) $m_{bb}^{\max PT}$, (b) $m_{uu}^{\min \Delta R}$, (c) $p_{T,uu}^{\min \Delta R}$ and (d) $m_{bb}^{\min \Delta R}$. The hashed area represents the post-fit uncertainty on the background.

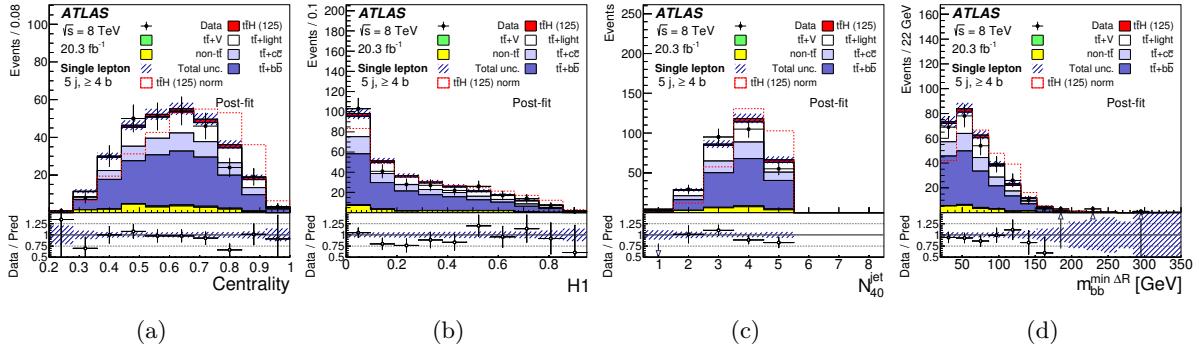


Figure D.10.: Post-fit comparison of data and prediction for the four top-ranked input variables in the (5j, $\geq 4b$) region. The plots include (a) Centrality, (b) H_1 , (c) N_{40}^{jet} and (d) $m_{bb}^{\min \Delta R}$. The hashed area represents the post-fit uncertainty on the background. The $t\bar{t}H$ signal distribution is shown normalised to background prediction (dashed red) and to the fitted $\hat{\mu}$ (filled solid).

D. Additional Material from the Profile Likelihood Fit

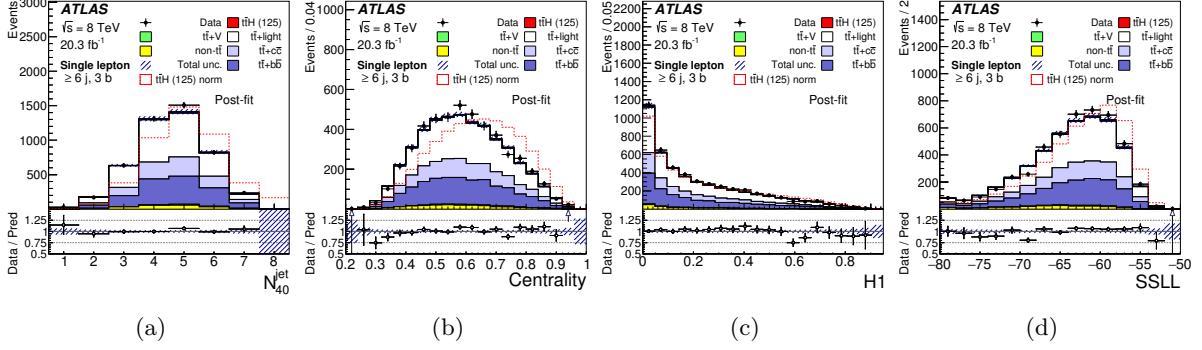


Figure D.11.: Post-fit comparison of data and prediction for the four top-ranked input variables in ($\geq 6j, 3b$) region. The plots include (a) N_{40}^{jet} , (b) Centrality, (c) H_1 , and (d) SSLL. The hashed area represents the post-fit uncertainty on the background. The $t\bar{t}H$ signal distribution is shown normalised to background prediction (dashed red) and to the fitted $\hat{\mu}$ (filled solid).

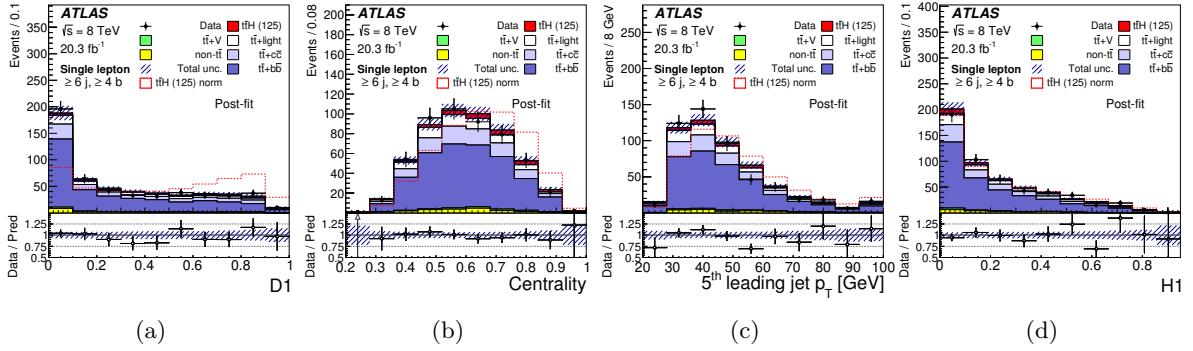


Figure D.12.: Post-fit comparison of data and prediction for the four top-ranked input variables in ($\geq 6j, \geq 4b$) region. The plots include (a) $D1$, (b) Centrality, (c) $p_T^{\text{jet}5}$, and (d) H_1 . The hashed area represents the post-fit uncertainty on the background. The dashed line shows $t\bar{t}H$ signal distribution normalised to background yield. The $t\bar{t}H$ signal distribution is shown normalised to background prediction (dashed red) and to the fitted $\hat{\mu}$ (filled solid).

List of Abbreviations

AGIS	ATLAS Grid Information System	87
ALICE	A Large Ion Collider Experiment.....	20
ATLAS	A Toroidal LHC ApparatuS.....	3
BSM	Beyond the Standard Model	14
BW	Breit-Wigner.....	83
CB	combined muon	38
CERN	Conseil Européen pour la Recherche Nucléarie.....	19
CKM	Cabibbo-Kobayashi-Maskawa.....	4
CL	confidence level	17
CMS	Compact Muon Solenoid	3
CP	charge-parity.....	9
CPU	central processing unit	
CSC	Cathode Strip Chamber.....	26
DAQ	data acquisition.....	31
dof	degree of freedom	81
EF	Event Filter.....	27
EMB	electromagnetic barrel	25
EMEC	electromagnetic end-cap.....	25
EM	electromagnetic	24
EW	electroweak	5
FCal	Forward Calorimeter	25
FS	final state	
FSR	final state radiation	41
GR	General Relativity	

D. Additional Material from the Profile Likelihood Fit

GUT	Grand Unified Theory	9
HEC	hadronic end-cap	25
HEP	high energy physics	
HF	heavy flavour	23
HFOR	heavy flavour overlap removal	47
IBL	Insertable B-layer	23
ID	Inner Detector	21
IS	initial state	
ISR	initial state radiation	41
JER	jet energy resolution	107
JES	jet energy scale	33
JVF	jet vertex fraction	
KS	Kolmogorov-Smirnov	119
L1	Level 1	27
L2	Level 2	27
LAr	liquid Argon	25
LCW	local cluster weighting	33
LEP	Large Electron-Positron Collider	19
LF	light flavour	35
LHCb	LHC-beauty	20
LHCf	LHC-forward	20
LHC	Large Hadron Collider	3
LINAC	Linear Particle Accelerator	19
LO	leading-order	10
MC	Monte Carlo	29
MDT	Monitored Drift Tube	26
MEDUSA	Management of Exigent, Distributed User Submission at ATLAS	87
ME	matrix element	
MEM	matrix element method	59
MIP	minimum ionizing particle	24
MoEDAL	Monopole and Exotics Detector At the LHC	20
MPI	multiple parton interaction	41
MS	Muon Spectrometer	21
MVA	multivariate analysis	47
NLO	next-to-leading-order	12
NNLL	next-to-next-to-leading-logarithmic	11

NNLO	next-to-next-to-leading-order	11
NN	neural network	104
NP	nuisance parameter	105
NWA	narrow width approximation	83
PDF	parton distribution function	35
PL	parton level	
p.d.f.	probability density function	59
PSB	Proton Synchrotron BOOSTER	
PS	Proton Synchrotron	20
QCD	Quantum Chromodynamics	5
QED	Quantum Electrodynamics	5
QFD	Quantum Flavourdynamics	5
QFT	quantum field theory	3
ROI	Region of Interest	27
RPC	Resistive Plate Chamber	26
SCT	Semiconductor Tracker	23
SF	scale factor	33
SI	International System of Units	2
SM	Standard Model	3
SPS	Super Proton Synchrotron	20
SUSY	supersymmetry	9
TDAQ	Trigger and Data Acquisition	27
TF	transfer function	60
TGC	Thin Gap Chamber	26
TOE	Theory of Everything	9
TOTEM	TOTal Elastic and diffractive cross section Measurement	20
TP	tag-and-probe	37
TRF	tag-rate-function	47
TRT	Transition Radiation Tracker	23
UE	underlying event	41
vdM	van der Meer	30
VEV	vacuum expectation value	
WIMP	weakly interacting massive particle	9
WLCG	wordwide LHC computing grid	86

List of Figures

2.1.	The Standard Model of Particle Physics	4
2.2.	The Higgs potential	7
2.3.	Leading order top quark pair production	11
2.4.	Leading order single top quark production	11
2.5.	Higgs boson production	13
2.6.	Higgs boson decay	14
2.7.	Mass spectrum of Higgs boson discovery channels	16
2.8.	Higgs boson mass combination	17
2.9.	ATLAS Higgs boson signal strength combination	18
3.1.	CERN accelerator complex	20
3.2.	Overview of the ATLAS detector	21
3.3.	ATLAS magnet system	22
3.4.	ATLAS Inner Detector	23
3.5.	ATLAS Calorimeter	24
3.6.	ATLAS Muon Spectrometer	26
3.7.	ATLAS trigger system	27
4.1.	Integrated luminosity and average interactions for 2011 and 2012	32
4.2.	LCW scale response and JVF pile-up suppression	34
4.3.	b -jet tagging efficiencies and SFs	35
4.4.	c -jet tagging efficiencies and SFs	36
4.5.	Light jet tagging efficiencies and SFs	37
4.6.	Schematic of $t\bar{t}H$ event generation	40
4.7.	Comparison of generators for $t\bar{t} + b\bar{b}$ event categories	43
4.8.	Reweighting $t\bar{t} + b\bar{b}$ POWHEG events to SHERPA	44
4.9.	Top and $t\bar{t}$ system p_T reweighting	45
4.10.	Effect of top quark and $t\bar{t}$ system p_T reweighting	46
4.11.	Validation of tag-rate-function method	49
4.12.	Control regions for measurement of real and fake efficiencies	51
4.13.	Measurement of real and fake efficiencies for fake estimation	53

List of Figures

4.14. Validation of fake estimation in two jet exclusive bin	55
4.15. Validation of fake estimation in four jet inclusive bin	56
4.16. Vector boson transverse momentum reweighting	57
4.17. Validation of $V+jets$ corrections	58
5.1. MEM process probability calculation of $t\bar{t}H$	61
5.2. Schematic of MEMTool package	65
5.3. Valence quarks and parton distribution functions groups comparison	66
5.4. CTEQ6L parton distribution functions	67
5.5. Q^2 dependency and higher-order comparison of PDFs	68
5.6. Types of the $t\bar{t}H$ production.	69
5.7. LO $t\bar{t}H$ diagrams used for MEM	70
5.8. LO $q\bar{q} \rightarrow t\bar{t} + b\bar{b}$ diagrams used for MEM	71
5.9. LO $gg \rightarrow t\bar{t} + b\bar{b}$ diagrams used for MEM	72
5.10. Double Gaussian fit of the quark and lepton TFs	76
5.11. Evolution of quark energy TF	77
5.12. Evolution of lepton energy and momentum TF	78
5.13. Evolution of the neutrino momentum TF.	79
5.14. MEDUSA work flow	88
5.15. Automated monitoring work flow	89
5.16. Signal likelihood of one parton level event	91
5.17. Signal likelihood of parton level event with multiple peaks	92
5.18. Signal likelihood of 1000 parton level events	93
5.19. $D1$ for $t\bar{t}H$ and $t\bar{t} + b\bar{b}$ parton level events	94
5.20. Signal likelihood of Gaussian smeared parton level events	95
5.21. $D1$ for $t\bar{t}H$ and $t\bar{t} + b\bar{b}$ Gaussian smeared parton level events	96
5.22. Signal likelihood of double Gaussian smeared parton level events	97
5.23. $D1$ for $t\bar{t}H$ and $t\bar{t} + b\bar{b}$ double Gaussian parton level events	97
6.1. Pre-fit event yield and signal significance	101
6.2. Background compositions and Higgs boson decay fractions	103
6.3. JES uncertainty components depending on p_T and η	109
6.4. $t\bar{t} + b\bar{b}$ systematic uncertainty variations	111
6.5. Best and summed $t\bar{t}H$ likelihood in ($\geq 6j, \geq 4b$)	114
6.6. Best and summed $t\bar{t} + b\bar{b}$ likelihood in ($\geq 6j, \geq 4b$)	115
6.7. Best and summed $t\bar{t}H$ likelihood in ($\geq 6j, 3b$)	116
6.8. Best and summed $t\bar{t} + b\bar{b}$ likelihood in ($\geq 6j, 3b$)	117
6.9. Signal and BG likelihood HF separation in ($\geq 6j, \geq 4b$)	117
6.10. Signal and BG likelihood HF separation in ($\geq 6j, 3b$)	118
6.11. Pre-fit distribution of $t\bar{t}H$ likelihood in ($\geq 6j, 3b$) and ($\geq 6j, \geq 4b$)	118
6.12. Best and summed likelihood ratio $D1$ cross section in ($\geq 6j, \geq 4b$)	120
6.13. Best and summed likelihood ratio $D1$ optimal in ($\geq 6j, \geq 4b$)	120
6.14. Best and summed likelihood ratio $D1$ in ($\geq 6j, 3b$)	121
6.15. Summed likelihood ratio $D1$ HF separation in ($\geq 6j, 3b$) and ($\geq 6j, \geq 4b$)	122
6.16. Summed likelihood ratio $D2$ in ($\geq 6j, 3b$) and ($\geq 6j, \geq 4b$)	122
6.17. Pre-fit distribution of $D1$ in ($\geq 6j, 3b$) and ($\geq 6j, \geq 4b$)	123
6.18. Reconstructed W boson mass separation in ($\geq 6j, 3b$)	124

6.19. Reconstructed W boson mass HF separation in ($\geq 6j, 3b$)	125
6.20. Pre-fit W boson mass with $t\bar{t} + b\bar{b}$ likelihood in ($\geq 6j, 3b$) and ($\geq 6j, \geq 4b$)	125
6.21. Reconstructed hadronic top quark mass separation in ($\geq 6j, \geq 4b$)	126
6.22. Reconstructed hadronic top quark mass separation in ($\geq 6j, 3b$)	127
6.23. Reconstructed hadronic top quark mass HF separation in ($\geq 6j, 3b$)	128
6.24. Pre-fit top quark mass with $t\bar{t} + b\bar{b}$ likelihood in ($\geq 6j, 3b$) and ($\geq 6j, \geq 4b$)	128
6.25. Reconstructed $m_{b\bar{b}}$ mass separation in ($\geq 6j, \geq 4b$)	129
6.26. Reconstructed $m_{b\bar{b}}$ mass separation in ($\geq 6j, 3b$)	130
6.27. Reconstructed $m_{b\bar{b}}$ mass HF separation in ($\geq 6j, 3b$) and ($\geq 6j, \geq 4b$)	131
6.28. Pre-fit $m_{b\bar{b}}$ mass with $t\bar{t}H$ likelihood in ($\geq 6j, 3b$) and ($\geq 6j, \geq 4b$)	132
6.29. Reconstructed $\Delta\phi$ between two additional b -jets separation in ($\geq 6j, 3b$)	133
6.30. Reconstructed $\Delta\phi$ between two additional b -jets separation in ($\geq 6j, \geq 4b$)	133
6.31. Design of the neural network	135
6.32. Separartion of all NN variables in ($\geq 6j, \geq 4b$)	139
6.33. Separartion of all NN variables in ($\geq 6j, 3b$)	141
6.34. Separartion of all NN variables in ($5j, \geq 4b$)	142
6.35. Separartion of all NN variables in ($5j, 3b$)	143
6.36. The NN output separation in ($\geq 6j, \geq 4b$)	144
6.37. The NN output separation in ($\geq 6j, 3b$)	145
6.38. The NN output separation in ($5j, 3b$) and ($5j, \geq 4b$)	145
6.39. Pre-fit and post-fit distributions of discriminants in four jet regions	148
6.40. Pre-fit and post-fit distributions of discriminants in five jet regions	150
6.41. Pre-fit and post-fit distributions of discriminants in six jet regions	151
6.42. Prefit and postfit event yields in all regions	152
6.43. Pull and constraints of most important NPs	154
6.44. Post-fit distribution of $t\bar{t}H$ likelihood in ($\geq 6j, 3b$) and ($\geq 6j, \geq 4b$)	157
6.45. Post-fit distribution of $D1$ in ($\geq 6j, 3b$) and ($\geq 6j, \geq 4b$)	158
6.46. Post-fit $m_{b\bar{b}}$ mass with $t\bar{t} + b\bar{b}$ likelihood in ($\geq 6j, 3b$) and ($\geq 6j, \geq 4b$)	159
6.47. Signal strength of combined analysis	160
6.48. Limits of combined analysis	163
6.49. Event yield as function of S/B	164
A.1. Likelihood ratio $D1$ separation for various processes in ($\geq 6j, \geq 4b$)	192
A.2. Likelihood ratio $D1$ separation for various processes in ($\geq 6j, 3b$)	193
A.3. $\ln \mathcal{L}_{t\bar{t}H}^{\text{sum}}$ separation for various processes in ($\geq 6j, \geq 4b$)	194
A.4. $\ln \mathcal{L}_{t\bar{t}H}^{\text{sum}}$ separation for various processes in ($\geq 6j, 3b$)	195
A.5. $m_{b\bar{b}}$ separation using $t\bar{t}H$ likelihood for various processes in ($\geq 6j, \geq 4b$)	196
A.6. $m_{b\bar{b}}$ separation using $t\bar{t} + b\bar{b}$ likelihood for various processes in ($\geq 6j, 3b$)	197
A.7. $m_{b\bar{b}}$ separation for various processes in ($\geq 6j, \geq 4b$)	198
A.8. $m_{b\bar{b}}$ separation for various processes in ($\geq 6j, 3b$)	199
A.9. m_{top} separation for various processes in ($\geq 6j, \geq 4b$)	200
A.10. m_{top} separation for various processes in ($\geq 6j, 3b$)	201
A.11. m_W^{had} separation for various processes in ($\geq 6j, \geq 4b$)	202
A.12. m_W^{had} separation for various processes in ($\geq 6j, 3b$)	203
B.1. Reconstructed $\Delta\eta$ between two additional b -jets separation in ($\geq 6j, 3b$)	205
B.2. Reconstructed ΔR between two additional b -jets separation in ($\geq 6j, 3b$)	206

List of Figures

B.3. Reconstructed p_T between two additional b -jets flavoration in ($\geq 6j, 3b$)	206
B.4. Reconstructed $\Delta\phi$ between two additional b -jets separation in ($\geq 6j, \geq 4b$)	207
B.5. Reconstructed $\Delta\eta$ between two additional b -jets separation in ($\geq 6j, \geq 4b$)	207
B.6. Reconstructed ΔR between two additional b -jets separation in ($\geq 6j, \geq 4b$)	208
B.7. Reconstructed p_T between two additional b -jets flavoration in ($\geq 6j, \geq 4b$)	208
C.1. The NN evolution with addition of variables	211
C.2. The NN output scenarios in ($\geq 6j, \geq 4b$)	212
C.3. Correlation of NN variables in ($\geq 6j, 3b$)	213
C.4. Overtraining test of the NN	213
D.1. The pull and constraints of all NP	215
D.2. The correlations between the most important NP	216
D.3. Post-fit distribution of $t\bar{t}H$ likelihood in electron and muon events	217
D.4. Post-fit distribution of $D1$ in electron and muon events	218
D.5. Post-fit distribution of kinematic NN in electron and muon events	218
D.6. Post-fit $m_{b\bar{b}}$ mass with $t\bar{t}H$ likelihood in ($\geq 6j, 3b$) and ($\geq 6j, \geq 4b$)	220
D.7. Post-fit top quark mass with $t\bar{t} + b\bar{b}$ likelihood in ($\geq 6j, 3b$) and ($\geq 6j, \geq 4b$)	221
D.8. Post-fit W boson mass with $t\bar{t} + b\bar{b}$ likelihood in ($\geq 6j, 3b$) and ($\geq 6j, \geq 4b$)	222
D.9. Post-fit distributions of NN input in ($5j, 3b$)	223
D.10. Post-fit distributions of NN input in ($5j, \geq 4b$)	223
D.11. Post-fit distributions of NN input in ($\geq 6j, 3b$)	224
D.12. Post-fit distributions of NN input in ($5j, 3b$)	224

List of Tables

2.1.	Higgs boson decay branching ratio	15
4.1.	LHC performance and parameter during Run I	31
4.2.	Summary of MC simulation used for physics processes	42
4.3.	Parametrisation of real and fake efficiencies used for fake estimation	54
5.1.	The parametrisation of the TF parameters p_i	75
5.2.	Parameter of quark TF for $0.8 < \eta < 1.37$	76
5.3.	Summary of Reduction of Dimensions	85
5.4.	Integration Ranges of Dimensions	86
5.5.	Gaussian width of Smearing	94
6.1.	Pre-fit event yields	102
6.2.	Summary of discriminants per region	104
6.3.	Summary of systematic uncertainties	106
6.4.	Variable ranking in NN	138
6.5.	Post-fit event yields	153
6.6.	Pre-fit / Post-fit uncertainties in ($\geq 6j, \geq 4b$)	155
6.7.	Limits of combined analysis	162
6.8.	Impact of MEM on results	165
7.1.	Upper limits on $t\bar{t}H$ production by ATLAS and CMS	169
B.1.	Reconstruction efficiencies	209
B.2.	Reconstruction efficiencies	209

Index

- b*-tag weight, 35
- heavy flavour overlap removal (HFOR), 47
- Insertable B-layer (IBL), 23
- jet vertex fraction (JVF), 33
- local cluster weighting (LCW) method, 33
- matrix element method (MEM), 59
- multivariate analysis (MVA), 134
- narrow width approximation (NWA), 83
- tag-rate-function (TRF), 47
- underlying event (UE), 41
- nuisance parameters (NPs), 105
- parton distribution functions (PDFs), 66
- transfer functions (TFs), 60
- matrix element method, 2
- 2012 dataset, 32
- 8 TeV dataset, 32
- allhadronic, 11
- anti- k_t algorithm, 33
- artificial neural networks, 134
- Asimov pseudo data, 147
- b-tagging, 34
- background-dominated regions, 103
- backpropagation, 136
- best $t\bar{t} + b\bar{b}$ permutation, 123
- best $t\bar{t}H$ permutation, 123
- beta function, 30
- Bjorken x, 66
- bosons, 3
- collinear safe, 33
- confinement, 5
- covariant derivative, 5
- dark energy, 9
- dark matter, 9
- data-driven, 39
- density of states, 61
- diagram removal, 46
- dilepton, 11
- Dirac equation, 5
- Dynamical Likelihood Method, 59
- eigenvector method, 110
- factorisation theorem, 40
- fake lepton, 50
- feed-forward NN, 135
- Fermi's Golden Rule, 61
- fermions, 3
- fine-tuning, 9
- fragmentation, 41
- Glückstern formula, 75
- Goldstone bosons, 6
- Good Run List, 31
- grid sites, 86
- hadronisation, 41
- Higgs mechanism, 6
- importance sampling, 80
- in-time pile-up, 32
- infra-red safe, 33
- instantaneous luminosity, 29

List of Tables

integrated luminosity, 30	quarks, 3
Jacobi rotation, 135	rejection factor, 35
Jacobian determinant, 83	running coupling, 9
Jacobian matrix, 83	sea-quarks, 66
jets, 25	seesaw mechanism, 10
Landau-Yang theorem, 17	signal strength, 17
leptons, 3	signal-enriched regions, 103
local gauge invariance, 5	single lepton, 11
Majorana particle, 10	sliding window, 37
matrix method, 50	smoothing, 147
minimum bias events, 41	stratified sampling, 80
MLM matching, 47	test statistic, 161
negative tag, 36	topo-cluster, 33
nodes, 134	transformation theorem, 83
out-of-time pile-up, 32	transition matrix element, 61
parton shower, 40	Unitary gauge, 7
PDF calibration, 35	vertical template morphing, 147
pile-up, 32	warm-start, 31
post-fit, 147	Weinberg angle, 7
pre-fit, 147	working points, 35

Acknowledgements

This thesis would not have been possible without the support and guidance of many people to whom I owe my gratitude. I have met so many people over the last years, that it is impossible to thank all of them. My apologies to whomever I have forgotten or could not consider.

First and foremost, I would like to express my sincere gratitude to my thesis advisor Prof. Arnulf Quadt. He not only gave me the opportunity to write a Diploma and PhD thesis at the II. Institute of Physics but also supported me in many ways over the last years. Thanks to him I could do research at CERN for a year, collaborated with interesting physicist in Amsterdam for a couple of month and gained experience during conferences, workshops or summer schools at nice places. His good advise, suggestions and expertise certainly contributed to this work. I would also like to thank Prof. Ariane Frey for kindly agreeing to be the second referee of this thesis.

During the time at CERN I had the pleasure to meet many people whom I would like to thank. The most important one is Lisa Shabalina, who suggested me to change my thesis topic after almost two years and work on the fantastic project resulting in this thesis. I have no regrets, it was one of the best decisions I have made. Probably the most challenging project I could choose but I have learned so much and it has awoken my enthusiasm for physics again. Thank you for the countless physics discussions during coffee breaks, taking care of me in so many ways, for giving me so much freedom and for trusting in my qualities. You really made me a better physicist.

During the review process of our paper many experienced particle physicist told us that this was the most sophisticated analysis they have ever seen. I will certainly not disagree with them, but it was the work of many highly skilled physicist, who worked on the leptonic $t\bar{t}$ ($H \rightarrow b\bar{b}$) analysis and to whom I owe my gratitude. Thank you all and in particular Jahred, Austin, Julian, Ian, Valerio, Stefan, Steffen, Aurelio, Andrea, Eve, Javier, Maria, Mark, Michele, Leonid and Lisa, for creating such a very healthy research environment and realising probably the closest to an ideal scientific collaboration I will ever encounter. Carrying out such a sophisticated analysis was a big achievement and it was a great pleasure to meet so many very talented physicist.

A special thanks goes to Austin, with whom I have accomplished the most astonishing collaboration I have ever experienced. Only the very close cooperation of similar motivated and talented physicist could make such an unrealisable seeming project possible. By sharing the

work according to our skills, I have certainly learned a lot about physics and programming from him - as well that I should not mess with Texas! Thank you for all the fun and late hour work making last minute fixes before and during the big grid runs. And sorry to your wife about the latter...

I would especially like to thank the Russian-Mexican, Leonid, who was my "suffering comrade" in the institute working on the same analysis. Thanks for being sometimes so Mexican preferring *fiestas y cervezas* over work and for being so Russian sharing everything you have. Along the road to this thesis, I have met so many extraordinary people. I would like to give a special mention to Kevin for always having an open ear despite supporting Herne-West, to Johannes for being the most correct person (in a positive way) I have met, to Boris for never saying no, to Anna for enjoying beer as much as football statistics, to Stefan for being so little Bavarian, to Andrea for always helping despite having no time and to Jason for always being so incredibly confused.

In Amsterdam I have met many friendly and welcoming Dutch, but I would like to thank in particular Duc and Priscilla for the great fun and the many cooking sessions. Priscilla, I miss you and your spaghetti alla carbonara. I would not have imagined that such a simple dish can be that tasty. Duc, I miss skiing with you through the deepest snow and the great evenings in Geneva or the French countryside.

Thanks to all my friends and colleagues at the II. Institute. I have been there for such a long time that I would most certainly forget someone if I started naming anyone. But everyone of you has managed to create such a relaxed working atmosphere that I really enjoyed my time there. I will miss the "test screen" espresso breaks, the nice chats about physics and non-physics, the times I shared with you outside the institute and beating you in tabletop football.

A huge thanks to Lisa, Maria, Oleg and Boris for reading the drafts of this thesis so thoroughly and attentively. Sorry for the many long sentences I tend to produce. I can improve. Really. Many friends have helped me stay sane through these years. They helped me to diverse my mind and supported me during the entire time of my studies. I would especially like to thank Bahne, Bela, Heiner, Maik, Matze and Philipp for the great DOKO evenings, for all the parties - for making my time in Göttingen such a great one. I greatly value your friendship and hope we are going to stay in touch.

Most importantly, I would like to express my heart-felt gratitude to my family not only for any kind of support during my time as a student but especially for their guidance through my life. My parents Rainer and Margrit and my sister Tanja have always been supportive of my enthusiasm for physics and are always there for me if I need help, advise or motivation. Thank you for your love, patience and strength.

I would also like to thank top and bottom who gave me comfort in the absence of a Higgs boson. They were always there for me no matter if I needed them or not.

Last but not least, the deepest gratitude to my love, Lara. She helped me through the hardest phase of my thesis, cheered me up, supported me in my everyday life and always believed in me. Thank you, for really understanding me, for accepting all my weaknesses and imperfections, for the happiest time of my life - for loving me like I am. Nothing can say it better than the words by Carl Sagan: "In the vastness of space and the immensity of time, it is my joy to share a planet and an epoch with you."