DISS. ETH No. 24964

Search for the production of the Higgs boson in association with a top quark pair with CMS at $\sqrt{s} = 13$ TeV

A thesis submitted to attain the degree of

Doctor of Sciences of ETH Zurich
(Dr. sc. ETH Zurich)

presented by

Joosep Pata
M.Sc. (Physics), University of Tartu
born on 24.05.1990
citizen of Estonia

accepted on the recommendation of

Prof. Dr. Günther Dissertori, examiner
Prof. Dr. Nigel Glover, co-examiner
Prof. Dr. Stefano Pozzorini, co-examiner

2018

Abstract

This thesis summarises the search for a rare process involving the Higgs boson, where the Higgs boson is produced in association with a top quark pair. This analysis was carried out using data from proton-proton collisions in the Large Hadron Collider (LHC), collected by the Compact Muon Solenoid (CMS) detector during the 2016 data taking period. In 2015, the LHC was restarted at an unprecedented centre-of-mass energy of $\sqrt{s} = 13$ TeV, significantly increasing the discovery potential of the experiments. The discovery of the top quark pair associated Higgs ($\mathrm{t\bar{t}H}$) production process would be a direct confirmation of the standard model (SM) mechanism of mass generation for the most massive known quark. Deviations from the SM expectation would signal the presence of beyond the standard model (BSM) physics affecting the recently-discovered Higgs boson. The $\mathrm{t\bar{t}H}$ process has a production cross-section that is about two orders of magnitude below the Higgs production via gluon fusion, which was the channel where this particle was discovered in 2012. Due to the extremely low production rate, all detectable decay modes of the Higgs need to be considered. We concentrated on the dominant decay mode for the Higgs boson, where it decays to bottom quarks which further hadronise to jets. In order to reduce the contribution of multi-jet processes not involving top quarks, we additionally require the presence of charged leptons, which can arise from the leptonic decays of the W bosons from top quark decays. This search, which is made challenging by the presence of an overwhelming background arising from QCD production of $\mathrm{t\bar{t}}$+jets, necessitated the use of sophisticated particle reconstruction and identification algorithms in the detector. We contributed to the development of a new method for identifying and tagging jets arising from the hadronisation of bottom quarks, which are a feature of both the signal and background processes, by combining information across different detector subsystems using machine learning. This combined b tagging algorithm improved over the state of the art by reducing the rate of erroneously tagged jets by approximately 10-20% and was employed in the CMS analysis that saw first evidence for the $\mathrm{H} \to \mathrm{b\bar{b}}$ decay. For the identification of the $\mathrm{t\bar{t}H}$ process, we implemented an algorithm that directly uses predictions from quantum field theory via the evaluation of matrix elements to disentangle the signal process from the $\mathrm{t\bar{t}}$+jets background. We employed this matrix element method in an analysis of CMS data to determine an upper limit on the $\mathrm{t\bar{t}H}$ production cross-section at the level of $\mu = \sigma/\sigma_{\mathrm{SM}} < 1.52$ at a 95% confidence level, with an upper limit of 1.57 expected in the case of no signal. This method is now available to the CMS collaboration and plays a central role in the Run 2 analysis of $\mathrm{t\bar{t}H}(\to \mathrm{b\bar{b}})$ in the Higgs group.

Advanced data mining techniques based on artificial intelligence have recently made significant progress on problems in various fields such as machine vision and translation that were previously thought to be intractable using computers. During an internship at the private company Lingvist Technologies, we investigated the use of these adaptive algorithms based on deep learning for a problem where accurate predictions based on theory are not available. We developed a data-driven method based on neural networks that significantly improved the accuracy of the estimated vocabulary size for people learning a second language. The use of such adaptive techniques based on machine learning also shows promise in the field of natural sciences, particularly in experimental high-energy physics, where the discovery potential can be enhanced by making maximal use of the information recorded by the detector, but also in effectively reducing, filtering and modelling the large amounts of data that are foreseen in the coming decades.

Zusammenfassung

Die vorliegende Dissertationsarbeit beschreibt die Suche nach der Produktion eines Higgs-Bosons in Assoziation mit einem Top-Quark-Paar. Durchgeführt wurde die Analyse mit Daten aus Proton-Proton-Kollisionen am Large Hadron Collider (LHC), die während des Jahres 2016 mit dem Compact Muon Solenoid (CMS) Detektor aufgenommen wurden. Seit 2015 operiert LHC mit einer bislang unerreichten Schwerpunktsenergie von $\sqrt{s} = 13$ TeV, welche das Potential neuer Entdeckungen durch die Experimente massgeblich erhöht. Die Entdeckung der assoziierten Produktion eines Higgs-Bosons mit einem Top-Quark-Antiquark-Paar ($\mathrm{t\bar{t}H}$) wäre eine direkte Bestätigung für den Mechanismus des Standardmodells (SM), der den Ursprung der Masse des Top-Quarks erklärt, welches das schwerste Teilchen im SM darstellt. Abweichungen von Vorhersagungen des SM würden die Existenz bislang unbekannter physikalischer Prozesse beweisen, die auch das Verständnis über das vor kurzem entdeckte Higgs-Boson verändern könnten. Der $\mathrm{t\bar{t}H}$-Prozess hat einen Wirkungsquerschnitt, der um etwa zwei Grössenordnungen kleiner ist als der des Higgs-Produktionsmechanismus durch die Fusion von Gluonen, durch welchen das Higgs-Boson 2012 entdeckt wurde. Wegen dieser äußerst kleinen Produktionsrate des $\mathrm{t\bar{t}H}$-Prozesses ist es notwendig, alle möglichen Zerfallsprozesse des Higgs-Bosons zu betrachten. Diese Arbeit beschäftigt sich mit dem wahrscheinlichsten Zerfallsprozess des Higgs-Bosons, im Rahmen dessen es in Bottom-Quarks, die desweiteren in Jets hadronisieren, zerfällt. Um den Beitrag von Ereignissen mit mehreren Jets ohne Top-Quarks zu verringern, werden nur solche Ereignisse betrachtet, in denen geladene Leptonen vorhanden sind. Leptonen entstehen beim Zerfall von W-Bosonen, die wiederum aus dem Zerfall von Top-Quarks stammen. Ausgeklügelte Techniken zur Teilchenrekonstruktion und Identifizierungsalgorithmen sind für diese Suche notwendig, um den Signalprozess von dem dominierenden Untergrund, der hauptsächlich aus QCD-Produktion von $\mathrm{t\bar{t}}$+jets besteht, unterscheiden zu können. Im Rahmen dieser Arbeit wurde zur Entwicklung einer neuen Methode beigetragen, die Jets, die aus der Hadronisierung von Bottom-Quarks sowohl im Signal- als auch im Untergrundprozess entstehen, identifiziert. Die Methode stützt sich auf das Kombinieren von Informationen aus verschiedenen Detektorkomponenten mithilfe von maschinellem Lernen. Dieser kombinierte Algorithmus für die Identifizierung von Bottom-Quarks führt zu einer 10-20% Verringerung der Anzahl an Jets, die fälschlicherweise als b-Jets identifiziert werden und wurde bereits in der CMS Analyse verwendet, die als Erste den Zerfall von Higgs-Bosonen zu zwei Bottom-Quarks nachgewiesen hat. Zusätzlich wurde für die Identifizierung des $\mathrm{t\bar{t}H}$Prozesses ein weiterer Algorithmus verwendet, der Vorhersagungen der Quantenfeldtheorie nutzt, um Matrixelemente zu berechnen. Diese ermöglichen es, Signal- von Untergrundprozessen zu unterscheiden. Die Matrix-Element-Methode wurde in einer Analyse von CMS-Daten verwendet, um eine obere Grenze für den $\mathrm{t\bar{t}H}$-Produktionswirkungsquerschnitt zu berechnen. Die Messdaten lieferten eine Grenze von $\mu = \sigma/\sigma_{\mathrm{SM}} < 1.52$ bei einem 95% Konfidenzniveau, wobei eine Grenze von 1.57 ohne Signal erwartet worden wäre. Die Matrix-Element-Methode ist nun für die gesammte CMS-Kollaboration zugänglich und spielt eine zentrale Rolle in der Run 2 Analyse des $\mathrm{t\bar{t}H}(\rightarrow \mathrm{b\bar{b}})$-Prozesses in der Higgs-Gruppe.

Fortgeschrittene, auf künstlicher Intelligenz basierende Data-Mining-Techniken, haben in letzter Zeit bemerkenswerte Fortschritte bei Problemen erreicht, die lange als unlösbar galten. Dies gilt zum Beispiel für Aufgaben wie der Bearbeitung von Bildern und für die Übersetzung. Während eines Praktikums im Unternehmen Lingvist wurde die Anwendung solcher adaptiven, auf tiefneuronalen Netzwerken basierenden Algorithmen auf Aufgaben, von denen es keine präzisen theoretischen Vorhersagungen gibt, untersucht. Ein adaptives neuronales Netzwerk wurde entwickelt, welches mit höherer Genauigkeit den Wortschatzes einer Person, die eine Fremdsprache erlernt, schätzt. Die Anwendung solcher auf maschinellem Lernen basierender Techniken ist ebenfalls vielversprechend für die Naturwissenschaften, wie zum Beispiel für die experimentelle Hochenergiephysik. Das Entdeckungspotential kann durch eine optimale Verwendung der vom Detektor gemessenen Informationen verbessert werden, aber auch durch eine effiziente Selektierung, Filterung und Modellierung der grossen Datenmenge, die in den nächsten Jahrzehnten erwartet wird.

Contents

Introduction 1

Acknowledgements 3

1 Theoretical background 5

1.1 The Standard Model . . . 5
1.1.1 The Lorentz group and particle states . . . 5
1.1.2 Relativistic quantum mechanics . . . 5
1.1.3 Interactions via gauge theory . . . 6
1.1.4 Quantum Electrodynamics . . . 7
1.1.5 Quantum Chromodynamics . . . 7
1.1.6 Hadrons . . . 9
1.1.7 Weak interaction . . . 9
1.1.8 Electroweak unification . . . 9
1.1.9 Higgs mechanism . . . 10
1.1.10 Top quark physics . . . 10
1.2 The Standard Model at colliders . . . 11
1.2.1 The parton model . . . 11
1.2.2 Jet physics . . . 13
1.2.3 Monte Carlo event generators . . . 13
1.3 Higgs phenomenology . . . 14
1.3.1 Higgs at the LHC . . . 14
1.3.2 Production modes . . . 14
1.3.3 Decay channels . . . 16
1.3.4 Experimental characterisation . . . 16
1.3.5 Top-Higgs coupling . . . 17
1.3.6 Current results on $t\bar{t}H$ . . . 18
1.4 Summary . . . 18

2 Experimental setup 21

2.1 The Large Hadron Collider . . . 21
2.1.1 The accelerator complex . . . 21
2.1.2 Experiments at the LHC . . . 23
2.2 The CMS detector . . . 24
2.2.1 The superconducting magnet . . . 24
2.2.2 Inner tracking system . . . 25
2.2.3 Electromagnetic calorimeter . . . 26
2.2.4 Hadronic calorimeter . . . 27
2.2.5 The muon systems . . . 29
2.2.6 Trigger, data acquisition and computing . . . 30
2.2.7 Particle flow reconstruction . . . 30
2.3 Summary . . . 31

3 Identification of jets from bottom quarks 33

3.1 Discriminating variables . . . 33
3.1.1 Track selection . . . 33
3.1.2 Vertex reconstruction . . . 34
3.2 Multivariate b taggers . . . 34
3.3 The combined multivariate b tagger . . . 36
3.3.1 Implementation . . . 36

3.3.2 Optimisation and validation . . . 38
3.4 Summary and outlook . . . 43

4 Matrix Element Method **45**
4.1 Hypothesis testing . . . 45
4.2 Description of MEM . . . 46
4.3 Implementation . . . 47
4.3.1 Transfer functions . . . 48
4.3.2 Lost quarks . . . 50
4.3.3 Treatment of QCD radiation . . . 52
4.3.4 Event interpretations . . . 53
4.3.5 Integration . . . 54
4.3.6 Profiling and optimisation . . . 56
4.3.7 Computational budget . . . 57
4.3.8 Propagating uncertainties . . . 57
4.3.9 MEM on the WLCG . . . 59
4.4 Expected performance . . . 59
4.4.1 Semileptonic categories . . . 61
4.4.2 Dileptonic categories . . . 63
4.5 Summary . . . 63

5 Search for top-quark pair associated Higgs production **71**
5.1 Data and simulation . . . 71
5.1.1 Modelling of $\mathrm{t\bar{t}}$+jets . . . 73
5.2 Event reconstruction and object identification . . . 74
5.2.1 Charged leptons . . . 74
5.2.2 Jets . . . 75
5.2.3 B tagging . . . 77
5.3 Analysis . . . 79
5.3.1 Event selection and categorisation . . . 79
5.3.2 Signal extraction . . . 81
5.3.3 Systematic uncertainties . . . 82
5.3.4 Jet energy correction uncertainties . . . 82
5.3.5 b tagging systematic uncertainties . . . 84
5.3.6 Other experimental systematic uncertainties . . . 86
5.3.7 Theoretical uncertainties . . . 86
5.3.8 Control regions . . . 88
5.3.9 Statistical method . . . 93
5.3.10 Analysis of the statistical model . . . 94
5.4 Results . . . 95
5.4.1 Systematic uncertainties . . . 107
5.5 Discussion . . . 109
5.6 Summary . . . 110

6 Modelling memory in language acquisition **113**
6.1 Introduction . . . 113
6.2 Problem statement and data description . . . 114
6.2.1 Knowledge estimation . . . 114
6.2.2 Data . . . 116
6.3 Model description . . . 116
6.3.1 Reference models . . . 116
6.3.2 Machine learning model . . . 117
6.3.3 Model architecture . . . 119
6.3.4 Optimisation . . . 121
6.4 Results . . . 121
6.4.1 Studies of the RNN model . . . 121

6.4.2 Discussion 122
6.4.3 Future work 124
6.5 Summary 125

7 Conclusions and outlook **129**

Bibliography **135**

Introduction

The aim of particle physics is to describe and understand the fundamental laws of nature. The Universe has been found to consist of particles, the quarks and leptons, and the interactions between these particles, mediated by weak, strong and electromagnetic forces. The standard model (SM) of particle physics is the achievement of more than a century of experimental and theoretical work by the physics community, starting from the discovery of the electron by J. J. Thomson in the end of the 19th century and the discovery of the atomic nucleus by E. Rutherford in the early 20th century. Crucially, it is a mathematical theory that can be used to describe high-energy processes in physics with a high degree of accuracy, as has been verified by collider experiments. The discovery of the Higgs boson in 2012 confirmed the basic mechanism of mass generation through electroweak symmetry breaking, but many open questions remain.

Several observed phenomena do not find an explanation, or at least a sufficient one, in present-day particle physics. In particular, the existence of non-luminous gravitating matter or dark matter (DM) is necessary to explain features on the astrophysical and cosmological scales, such as the rotation curves of galaxies and the energy spectrum of the cosmic microwave background radiation. It is plausible that this gravitational effect results from weakly interacting particles that are distinct from quarks or leptons. It is possible these DM particles could be detected (I) in collider searches, (II) by directly observing the nuclear recoil from the DM halo in our galaxy due to weak interactions, or (III) indirectly through the detection of cosmic rays resulting from DM annihilations.

Furthermore, the non-zero neutrino masses and mixings necessary to explain the observed neutrino flavour oscillations currently do not have an explanation in the SM. It is possible that neutrinos, which have masses between six to ten orders of magnitude below other fermions, acquire mass by a different mechanism than Yukawa interactions with the Higgs field. Thus, research in neutrino physics is expected to shed light on the nature, the mass generation mechanism and hierarchy and the mixing properties of the neutrinos. There is considerable asymmetry between matter and anti-matter in the Universe, but the sources of charge-parity (CP) violation that would allow this to happen are not sufficient in the SM. The experimental study of the flavour sector of high-energy physics complements direct searches for new physics at the energy frontier.

The Higgs field is the first observed fundamental scalar field, therefore, it is interesting to study the properties of the observed Higgs boson more deeply. It is possible that the Higgs boson is a composite particle, a hypothesis which can be tested by precisely measuring the couplings of the Higgs boson to SM particles.

There are other aspects of the SM that cannot yet be explained satisfactorily. In particular, fermions are found to fall into three generations, but there seems to be no underlying reason for this. Furthermore, many of the parameters of the SM, such as the masses and mixing properties of the fermions, cannot be deduced from theory alone and have to be determined experimentally. All of the aforementioned questions motivate a further study of the physics of high-energy processes with the hope of arriving at a more complete mathematical understanding of the Universe. The Large Hadron Collider (LHC) experiments, which in the last years have collected a few percent of the data foreseen over their lifetimes and have experimentally confirmed the foundations of the SM, are essential for progress on these questions.

In this thesis, we focus on improving our understanding of the Higgs sector by developing a measurement for the Higgs production cross-section in a rare production mode, where the Higgs boson is produced in association with a top quark pair. The experimental confirmation of the $\mathrm{t\bar{t}H}$ process would be a direct verification of the mechanism for mass generation for the most massive known quark. It is expected that using Run 2 data, the presence of this production mode can be determined with a significance exceeding five Gaussian standard deviations (σ) and thus the Yukawa coupling of the top quark can be measured with a precision of $\simeq 10\%$. This requires effort in reducing the effect of experimental uncertainties and of the background contribution arising from

QCD production of top quark pairs, which is the overall focus of this thesis. We have developed a data analysis method based on the direct computation of matrix elements on observed events that does not rely on the simulation and subsequent fitting of millions of complex multi-jet events using Monte Carlo. This method is applied in the $\mathrm{t\bar{t}H}(\rightarrow \mathrm{b\bar{b}})$ search using 35.9 fb^{-1} of data collected during 2016 by the CMS experiment to extract upper limits on the $\mathrm{t\bar{t}H}$ production cross-section.

The accurate reconstruction and identification of jets is essential for this search. We have improved upon the algorithms used for the identification of jets from the hadronisation of bottom quarks at CMS by developing a classifier based on machine learning that combines information from various subsystems of the detector. This improved algorithm was used during data-taking at CMS during the 2016 data taking period.

The LHC experiments will face unprecedented data rates in the coming decades during the high-luminosity LHC (HL-LHC) project, where the total amount of collected data will increase by two orders of magnitude. This presents an opportunity for physics as well as a challenge in terms of the reduction, analysis and storage of these data. Recent advances in the field of machine learning have made it possible to use algorithms that directly learn from and adapt to data instead of being constructed by human experts. Using these techniques based on mathematical optimisation, human-level performance has been achieved or exceeded in many areas where an algorithmic solution was previously thought to be many decades away, such as computer vision or the game of Go. During an internship at the private company Lingvist Technologies we applied these machine learning methods to model language learning behaviour in humans, where theoretical models are less predictive than in physics. The use of such data-driven techniques shows great promise in fields where the underlying model is not known, but can also benefit physics in cases where theoretical models are not yet predictive enough, such as the reconstruction and identification of complex signals spanning the detector.

This thesis is structured as follows. In chapter 1, we introduce the Standard Model of particle physics and the theoretical background for Higgs physics. We discuss the experimental setup of the LHC machine and the CMS experiment in chapter 2. The method used at CMS for identifying jets from bottom quarks is introduced in chapter 3, where we also discuss our contribution to the CMS state of the art. We introduce the matrix element method used for the $\mathrm{t\bar{t}H}(\rightarrow \mathrm{b\bar{b}})$ search in chapter 4 and discuss the implementation, improvements and validation studies that we carried out. This method is applied in a search for $\mathrm{t\bar{t}H}(\rightarrow \mathrm{b\bar{b}})$ using CMS data, which is described in chapter 5. We describe the data and simulation samples, the reconstruction of the signal, the systematic uncertainties affecting the measurement and the procedure used to extract the signal strength parameter and the upper limit on $\mathrm{t\bar{t}H}$ production. We also compare our results to the latest analysis from the ATLAS collaboration. Finally, in chapter 6 we discuss our work on modelling language learning using data-driven machine learning techniques. We conclude with a summary and outlook in chapter 7.

Acknowledgements

I would like to acknowledge some of the people from whom I've benefited the most during my PhD studies. First and foremost, I'm deeply indebted to Günther Dissertori for accepting me as a student and for his continuous advice and support. Secondly, I would like to thank Lorenzo Bianchini and Gregor Kasieczka, who were very generous with their time and from whom I was able to learn a lot during my studies. I would also like to thank Nigel Glover and the rest of the people who created the HiggsTools Initial Training Network, which made it possible for me to undertake PhD studies at ETH Zürich. I'm very grateful to Joe Incandela, whose generosity via the CMS Fundamental Physics Scholarship supported my first year at CERN. I learned a lot from all my colleagues at the Institute of Particle Physics and Astrophysics at ETH Zürich, I would like to thank in particular Pasquale Musella and Malte Backhaus, whose additional comments were essential for this text. Furthermore, I would like to thank my colleagues at the $\mathrm{t\bar{t}H}$ group at ETH, Maren Meinhard and Christina Reissel, for a very engaging work environment, and for helping me with the German translation of the abstract of this thesis. I'm thankful to my Estonian HEP colleagues at NICPB, Tallinn, thanks to whom I was able to learn about particle physics at CERN during my undergraduate studies. It was very interesting to work with the Lingvist team in Tallinn during my internship and I'm thankful for their kind hospitality. I'm deeply grateful to Jackie for all the support and patience, and for helping me with English. Finally, I would like to thank my parents and the rest of my family for nurturing and encouraging my interest in science.

1 Theoretical background

In this chapter, we give an overview of the theoretical background necessary for Higgs boson searches at the LHC. We start with an overview of the standard model (SM) of particle physics, followed by a description of the phenomenology of proton-proton collisions and Higgs boson production. We conclude with a discussion on the relevance of the search for the associated production of the Higgs boson with top quarks.

1.1 The Standard Model

The fundamental building blocks of the SM are complex quantum fields $\phi(x)$ that depend on the space-time coordinates x. The interaction of these fields is governed by Quantum Field Theory (QFT). These fields can be classified according to their transformation properties under various symmetry groups, in particular the Lorentz group, and thus associated to particle states. The principle of gauge invariance allows us to use these transformation properties to describe fundamental interactions via particle exchange. Much of particle physics is concerned with the measurement of decay rates and scattering cross sections, which are computed using perturbation series in QFT.

The particle content of the SM is summarised in fig. 1.1, where we see that the fundamental particle states can be grouped into bosons with integer spin and fermions with spin 1/2. The fermions are divided into quarks that carry colour charge, fractional electric charge and weak isospin, and leptons which are divided into charged leptons and neutrinos and carry no colour charge. In the following, we describe the types of fields in the QFT description of particle physics.

1.1.1 The Lorentz group and particle states

The Lorentz group consists of coordinate transformations $x^\mu = \Lambda^\mu{}_\nu x^\nu$ that preserves the space-time interval $\mathrm{d}s^2 = \mathrm{d}x^\mu \eta_{\mu\nu} \mathrm{d}x^\nu$ defined through the metric tensor $\eta_{\mu\nu}$. The Lorentz transformation satisfies $\Lambda^\mu{}_\rho \eta^{\rho\sigma} \Lambda^\nu{}_\sigma = \eta^{\mu\nu}$, such that $\det \Lambda = +1$ and sign $\Lambda^0{}_0 = +1$. The transformations form a group $\mathrm{SO}^+(3,1)$, which can be decomposed $\mathrm{SO}^+(3,1) \simeq \mathrm{SU}(2)_L \times \mathrm{SU}(2)_R$. This allows the angular momenta (j_1, j_2) of the decomposition to be used to group the fields as scalars $(j_1 = 0, j_2 = 0)$, left and right handed spinors $(\frac{1}{2}, 0), (0, \frac{1}{2})$ and vectors $(1, 1)$ based on their transformation properties under the Lorentz group. For example, a scalar field $\phi(x)$ transforms as $\phi(x) \to \phi(\Lambda^{-1}x)$, a vector field as $A^\mu(x) \to \Lambda^\mu{}_\nu A^\nu(\Lambda^{-1}x)$ and a spinor field as $\phi^\alpha(x) = S[\Lambda]^\alpha{}_\beta \phi^\beta(x)$, where $S[\Lambda]$ is a spinor built from 4×4 Dirac γ matrices in the chiral representation.

These fields can be identified with particle states, which have a definite mass m and spin s [2]. According to the spin-statistics theorem, particles with integer spin (bosons) follow Bose-Einstein statistics, whereas particles with half-integer spin (fermions) follow Fermi-Dirac statistics [3].

1.1.2 Relativistic quantum mechanics

After having identified quantum fields with definite Lorentz transformation properties as the central objects in QFT, the next step is to derive dynamical relations for the free fields in order to describe the propagation of free particles.

The Klein-Gordon wave equation,

$$(\partial^\mu \partial_\mu + m^2)\psi = 0 \tag{1.1}$$

which can be derived from Einstein's energy-momentum relation $E^2 = \boldsymbol{p}^2 + m^2$ by replacing energy and momentum with operators acting on the wavefunction ψ, is a manifestly Lorentz-invariant relation between energy and momentum for the quantum mechanical wavefunction. However, it admits solutions with negative energy and negative probability densities, which are unphysical.

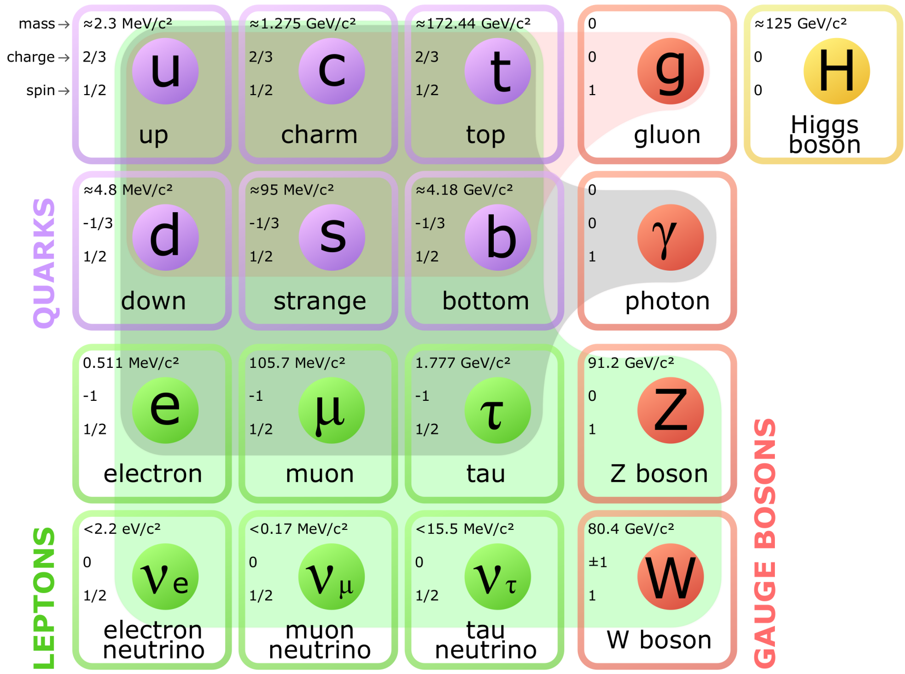


Figure 1.1: The particle content of the Standard Model. Figure adapted from [1].

These negative-probability states led Dirac to search for a relation linear in $\hat{\mathbf{p}}$ and E, which resulted in the Dirac equation:

$$(i\gamma^\mu \partial_\mu - m)\psi = 0 \tag{1.2}$$

where γ^μ are the 4×4 Dirac γ-matrices which satisfy the Clifford algebra anti-commutation relation $\{\gamma^\mu, \gamma^\nu\} = \gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu}\mathbb{I}$ and ψ is a four-component spinor. Using eq. (1.2), we can describe the dynamics, spin and magnetic moment of free spin-half fermions. The probability densities predicted by Dirac's equation are positive, but it still admits solutions with negative energy. In the Feynman-Stückelberg interpretation, these $E < 0$ solutions can be interpreted as negative-energy particles moving backwards in time or equivalently as anti-particles moving forwards in time. The predictions by Dirac's equation were spectacularly confirmed by the discovery of the positron in 1932 [4].

Both of these equations of motion can be derived from a Lagrangian using the principle of least action:

$$S = \int \mathrm{d}^4x \, \mathcal{L}_{\text{free}}, \quad \delta S = 0, \tag{1.3}$$

where $\mathcal{L}_{\text{free}}$ is the Lagrangian density for non-interacting fields for the Klein-Gordon or Dirac case respectively and the integral is taken over all possible path variations of the fields. The Dirac action for a single free spinor field can be written as

$$\mathcal{L}_{\text{Dirac}} = \bar{\psi} i\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi, \quad \bar{\psi} = \psi^\dagger \gamma^0 \tag{1.4}$$

and varied with respect to ψ to derive the Dirac equation.

1.1.3 Interactions via gauge theory

The concept of interactions mediated by fields is central to QFT. In order to encode the observed particle interactions in the theory, we require the Lagrangian to have additional symmetries. In particular, if we require the laws of physics to be invariant under local transformations $\psi(x) \to U(x)\psi(x)$, where the continuous and differentiable transformations $U(x)$ form Lie groups, then the principle of gauge symmetry allows additional degrees of freedom corresponding to the mediator fields or gauge bosons to be naturally incorporated in the Lagrangian. First, we show how quantum electrodynamics, the relativistic quantum theory of electron-photon interaction and thus electromagnetism, arises from the gauge principle.

1.1.4 Quantum Electrodynamics

For the U(1) symmetry, which corresponds to the transformation $\psi(x) \to e^{iq\varphi(x)}\psi(x)$, in order for the Lagrangian to be invariant under this symmetry, the derivative in the momentum operator in eq. (1.4) needs to be changed to

$$\partial_\mu \to \mathcal{D}_\mu = \partial_\mu + iqA_\mu(x). \tag{1.5}$$

This ensures that the covariant derivative $\mathcal{D}_\mu\psi$ transforms as the field ψ itself under the gauge transformation U. The field A_μ corresponds to a massless gauge boson with spin 1 that is coupled to the fermion field ψ via a coupling constant q in the term $q\bar{\psi}\gamma^\mu A_\mu\psi$. The formulation of quantum electrodynamics (QED) as a gauge theory generated by the Abelian group $\mathrm{U}(1)_{\mathrm{EM}}$ was first done by Tomonaga, Feynman and Schwinger [5, 6, 7]. In QED, the spinor ψ is associated to the electron, the vector A_μ to the photon and q is the electric charge of the fermion.

In order to compute observable decay rates or scattering cross sections under an interaction Hamiltonian V, we use Fermi's golden rule, which relates the scattering rate Γ_{fi} to the transition matrix element between the initial and final state derived using perturbation theory:

$$\Gamma_{fi} \propto |\mathcal{M}_{fi}|^2\rho, \quad \mathcal{M}_{fi} = \langle\psi_f|V|\psi_i\rangle \tag{1.6}$$

where ρ is the density of final states which generally depends on the final state energy E_f.

The matrix element in eq. (1.6) is a Lorentz scalar and can be explicitly computed from the Lagrangian using Feynman rules [8], where we represent each term in the perturbation expansion of eq. (1.6) as a graphical diagram with the ingoing and outgoing lines. The vertices and the propagators are associated with quantities that are specified by the form of the interaction Lagrangian. The basic QED interaction vertex thus allows us to construct diagrams corresponding to any QED interaction and calculate scattering cross sections for processes such as electron-electron scattering, as depicted in fig. 1.2.

It is remarkable that the gauge theory formulation of QED allows us to both recover Maxwell's equations of electromagnetism and predict the anomalous electric dipole moment of the electron, which has been confirmed to be accurate to about one part per billion [10]. This underscores the predictive power of symmetries in the SM.

1.1.5 Quantum Chromodynamics

The interaction of quarks and gluons can be described by the theory of quantum chromodynamics (QCD), which arises from the invariance of the Lagrangian under a local $\mathrm{SU}(3)_C$ symmetry, where C stands for a colour charge that is N_c-valent, with $N_c = 3$ in the SM. The spinor field transforms under this group as

$$\psi(x) \to \psi'(x) = \exp\left[ig_s\alpha^a(x)T^a\right]\psi(x) \tag{1.7}$$

where T^a are the generators of the group represented by the 3×3 Gell-Mann matrices λ^a as $T^a = \frac{1}{2}\lambda^a$, g_s is the gauge coupling and $\alpha^a(x)$ are the local gauge transformations corresponding to the eight generators. The covariant derivative can then be written as

$$\partial_\mu \to \mathcal{D}_\mu = \partial_\mu + ig_sG^a_\mu T^a \tag{1.8}$$

where G^a_μ must transform as $G^a_\mu \to G^a_\mu - \partial_\mu\alpha^a - g_sf_{bca}\alpha^bG^c_\mu$. The last term arises from the non-Abelian nature of QCD, which means that the generators T^a do not commute, but are instead related through the structure constants: $[T^a, T^b] = T^aT^b - T^bT^a = if_{abc}T^c$. The Lagrangian for a single quark field can then be written as

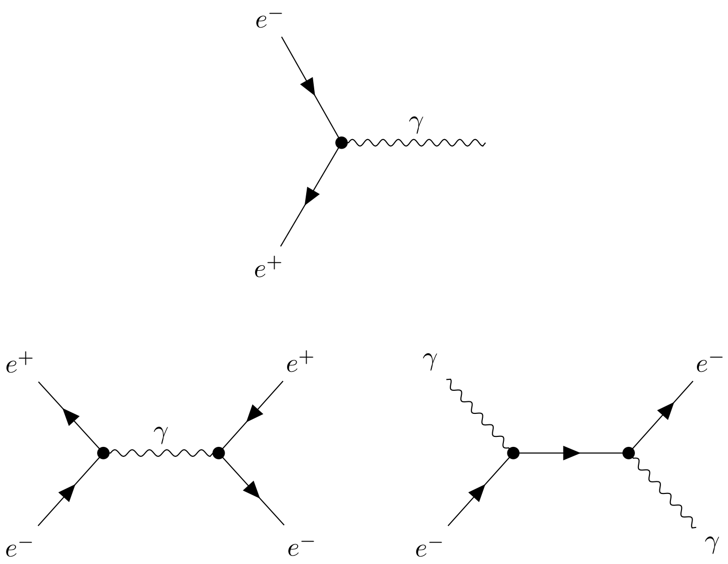


Figure 1.2: The basic QED vertex (top), a representative leading order diagram for e^+e^- annihilation and subsequent pair production (left) and for Compton scattering (right). Time flows left to right. These figures were created using the `tikz-feynman` package [9].

$$\mathcal{L}_{\text{quark}} = \bar{\Psi}(i\gamma^\mu\partial_\mu - m - g_s\gamma^\mu G^a_\mu T^a)\Psi \tag{1.9}$$

and we can associate G^a_μ to the gluons. We note that $\Psi = (\psi_i)$ has three components corresponding to colour states, each of which is a 4-component Dirac spinor. The QCD quark-gluon interaction vertex is then

$$-g_s\bar{\Psi}\gamma^\mu G^a_\mu T^a\Psi. \tag{1.10}$$

At this stage, G^a_μ is simply a field associated with external sources, which can be made dynamical by adding a gauge-invariant term

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{2}\text{Tr}\;[F^{\mu\nu}F_{\mu\nu}] \tag{1.11}$$

to the QCD Lagrangian, where $F_{\mu\nu} = \partial_\mu G_\nu - \partial_\nu G_\mu + ig_s[G_\mu, G_\nu]$, with $G_\mu = G^a_\mu T^a$, is the QCD gluon field strength tensor. The non-vanishing structure constants f_{abc} imply that gluons carry colour charge and thus interact with each other, giving rise to a triple-gluon vertex (ggg) proportional to g_s and a 4-gluon vertex (gggg) proportional to g_s^2.

The $\text{SU}(3)_\text{C}$ symmetry of QCD thus implies the existence of a conserved trivalent colour charge, which is exchanged between quarks by eight gluons carrying colour-anticolour in QCD vertices. As free quarks have not been experimentally detected, the colour charge is hypothesized to be confined, such that quarks are only observed bound to colourless hadrons - mesons ($\text{q}\bar{\text{q}}$) and baryons (qqq). Furthermore, this means that hadrons have to be colour singlets, which for baryons, composed of three quarks, implies that the colour wavefunction must be totally antisymmetric under the exchange of any two quarks, since the colour singlet $\psi_c = \frac{1}{\sqrt{6}}(rgb - rbg + gbr - grb + brg - bgr)$ from the decomposition $\mathbf{3}\otimes\mathbf{3}\otimes\mathbf{3} = \mathbf{10}\oplus\mathbf{8}\oplus\mathbf{8}\oplus\mathbf{1}$ is totally antisymmetric.

The success of QCD in describing the phenomenology of high-energy interactions of protons relies on a model for building hadrons out of the elementary constituents of QCD - the quarks and gluons.

1.1.6 Hadrons

The known quarks of the SM come in six different flavours, grouped into three generations: up and down (I), charm and strange (II), top and bottom (III). The underlying reason for the SM having exactly three generations is unknown, but an approximate symmetry between flavours allows us to predict allowable hadronic states that can be formed from quarks. Strong interaction is approximately invariant under u ↔ d exchange, which implies an SU(2) flavour symmetry. This group has three generators $\hat{T}_i$ that can be represented as 2×2 Pauli spin matrices. This means that for a given flavour state ψ, the quantised isospin $I_3 \leftrightarrow \hat{T}_3$ and the total isospin $I \leftrightarrow \hat{T}^2$ are conserved in analogy to spin, and can therefore be used to label composite states of quarks as $\phi(I, I_3)$.

We can construct the flavour wavefunction proton, which contains three valence quarks (uud), by combining three isospin doublets $\mathbf{2} \otimes \mathbf{2} \otimes \mathbf{2} = \mathbf{4} \oplus \mathbf{2}$, which results in a spin $I = 3/2$ quadruplet and two spin $I = 1/2$ doublets $\phi_A = \frac{1}{\sqrt{2}}(\mathrm{udu} - \mathrm{duu})$ and $\phi_S = \frac{1}{\sqrt{6}}(2\mathrm{uud} - \mathrm{duu} - \mathrm{udu})$, that are (anti)symmetric under the exchange of the first two quarks. The proton wavefunction is then a superposition of the two flavour doublets, multiplied by the corresponding (anti)symmetric wavefunctions for the spin states, such that the flavour-spin wavefunction is completely symmetric under the exchange of any two quarks. This, combined with the completely antisymmetric colour wavefunction described in section 1.1.5, guarantees that the proton wavefunction is completely antisymmetric under the exchange of any two quarks.

The SU(2) isospin symmetry of flavour allows us to write down the wavefunction of the proton in the ground state and predict the existence and approximate masses of the excited states such as the Δ-baryons. It is not an exact symmetry, as illustrated by the difference in the proton and neutron masses, which should vanish under precise SU(2) isospin symmetry, but nevertheless underscores the role of symmetries in describing hadronic states. Further hadronic states can be described by the SU(3) model, where the u, d and c quarks are grouped into a triplet

1.1.7 Weak interaction

The weak force is responsible for radioactive β-decay and the leptonic decay of mesons. It is the only fundamental interaction known to violate parity [11]. The weak interaction couples neutrinos to charged leptons and up-type quarks to down-type quarks via the exchange of massive vector bosons, the charged W-boson and the neutral Z-boson. At low energy ($q \ll m_W$) and before the observation of parity violations, the weak force was modelled as a Fermi theory with a matrix element $\mathcal{M} = G_F g_{\mu\nu}[\bar{\psi}_3 \gamma^\mu \psi_1][\bar{\psi}_4 \gamma^\mu \psi_2]$ corresponding to a point interaction of four fermions proportional to the Fermi constant G_F. After the observation of parity violation, the weak interaction vertex needed to be modified to include axial vector coupling in addition to vector coupling, such that it is proportional to $g\gamma^\mu(1 - \gamma^5)$. The γ^5 operator serves to project out the left (right) handed (anti)-particle chiral states, such that only these states interact via weak boson exchange.

The weak interaction can be described in the SM as a $\mathrm{SU}(2)_L$ gauge symmetry. The left-handed fermions form doublets (ν_L, ℓ_L) for leptons and (q_L, q'_L) for quarks, whereas right-handed fermions ℓ_R, q_R are singlets under weak isospin. The generators T_i of $\mathrm{SU}(2)_L$ are related to the three Pauli matrices $T_i = \frac{1}{2}\sigma_i$, which can then be associated to two charged vector boson fields that mediate the weak force: $W^\pm_\mu = \frac{1}{\sqrt{2}}(W^1_\mu \mp iW^2_\mu)$. The gauge structure predicts the existence of a massive electrically neutral vector boson W^3_μ, which can be associated to the SM Z boson through electroweak unification, described in the following section.

1.1.8 Electroweak unification

The weak and electromagnetic forces both suggest an electrically neutral boson, so the physically observed states of the photon and the Z-boson must be superpositions of the two, implying a connection between the weak and electromagnetic forces. The mixing between the neutral gauge boson states can be expressed as $A = B\cos\theta_W + W^3 \sin\theta_W$ and is characterised by the weak mixing angle $\theta_W = \tan^{-1} g'/g$ between a new $\mathrm{U}(1)_Y$ gauge field B with a coupling g' and the neutral component of the weak force W^3. In the Glashow-Weinberg-Salam theory of electroweak unification [12, 13, 14], the Lagrangian is symmetric under the group $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$, whereas the vacuum state is symmetric only under the QED gauge symmetry $\mathrm{U}(1)_{\mathrm{EM}}$. In the unified theory, it is necessary to introduce a new quantum number Y, the hypercharge, which is equal for both components of the $\mathrm{SU}(2)_L$ doublets, so that the left-handed doublets would be invariant under both $\mathrm{SU}(2)_L$ and $\mathrm{U}(1)_Y$ symmetries. The observed electric charges of the fermions are then related to the weak hypercharge Y and weak isospin I_3 by $Q = I_3 + Y/2$.

The theory of electroweak unification was confirmed by the observation of weak neutral currents in the Gargamelle detector [15] and the following discovery of the W and Z-boson by the UA1 and UA2 experiments at the SPS collider [16], which couples to both left-handed and right-handed fermions. This has allowed the weak mixing angle θ_W between the neutral vector boson states to be measured. The principle of local gauge invariance has great predictive power for the overall structure of the observed interactions, but it does not account for the mechanism by which electroweak symmetry breaking (EWSB) is realised nor the mass of the $\mathrm{W}^{\pm}$ and Z bosons, which would violate gauge invariance. In order to incorporate these phenomena, we turn to the Higgs mechanism. The decay $\mathrm{Z} \rightarrow \nu\bar{\nu}$ can be used to determine the number of light neutrino generations by measuring the total width of the Z-boson Γ_Z and the decay widths to visible fermions - the charged leptons $\mathrm{e}^{\pm}, \mu^{\pm}, \tau^{\pm}$ and quarks [17].

1.1.9 Higgs mechanism

We can introduce mass terms for the heavy gauge bosons by coupling them to two complex scalar fields arranged in a weak isospin doublet $\phi = \frac{1}{\sqrt{2}}(\phi^{+},\ \phi_0)^T$ [18, 19, 20, 21]. The Lagrangian density for this scalar field is $\mathcal{L}_{\phi} = (\partial_{\mu}\phi)^{\dagger}(\partial^{\mu}\phi) - V(\phi)$, where the Higgs potential $V(\phi) = \mu^2(\phi^{\dagger}\phi) + \lambda(\phi^{\dagger}\phi)^2$ has degenerate minima $\phi^{\dagger}\phi = v^2/2 = -\mu^2/2\lambda$ for $\mu^2 < 0$. If the physical vacuum state does not have the same symmetries as the Lagrangian, then it is possible to introduce gauge-invariant mass terms for gauge bosons and fermions.

The Higgs field couples to the gauge fields through the kinetic term $(\partial_{\mu}\phi)^{\dagger}(\partial^{\mu}\phi)$, which is made gauge invariant by $\partial_{\mu} \rightarrow D_{\mu} = \partial_{\mu} + igT^aW^a_{\mu} + ig'YB_{\mu}/2$. The vacuum state is then chosen to be $\langle 0|\phi|0\rangle = \frac{1}{\sqrt{2}}(0,\ v)^T$, so the Higgs field can be expanded around the vacuum, resulting in a massive scalar field η and three massless Goldstone fields ϕ_1, ϕ_2, ϕ_3. The Goldstone bosons can be removed from the Lagrangian by fixing the gauge, such that they correspond to the degrees of freedom of longitudinal polarisation states of the Z and $\mathrm{W}^{\pm}$ bosons. The masses of the gauge bosons can then be written as $M_W = \frac{1}{2}gv$ and $m_Z = \frac{1}{2}gv/\cos\theta_W$. The theory of EWSB predicts $m_W/m_Z = \cos\theta_W$, which has been confirmed experimentally [22].

The explicit mass terms for fermions in the form of $m\bar{\psi}\psi$ are not gauge invariant. By arranging the left-handed fermions in SU(2) doublets L and the right-handed fermions in singlets R, the mass terms for Dirac fermions can be introduced through spontaneous symmetry breaking through the terms $y_f\bar{L}\phi R + y_f(\bar{L}\phi R)^{\dagger}$ for down-type leptons $e^{\pm}$, $\mu^{\pm}$, $\tau^{\pm}$ and quarks d, s and b. The constant $y_f = \sqrt{2}m_f/v$ is the Yukawa coupling of the fermions, which has to be fixed from the experimental determination of the masses. A conjugate Higgs doublet $\phi_c = -i\sigma_2\phi^*$ is used to give the mass terms for the up-type quarks. From this mechanism, the Yukawa coupling of the top quark is found to be compatible with unity, a scale very different from the rest of the quarks.

To summarise, the Higgs mechanism can accommodate the observed masses of heavy gauge bosons and fermions in a unified electroweak theory. It predicts the existence of a massive scalar boson, which can decay to fermions through $\mathrm{H} \rightarrow \mathrm{f\bar{f}}$ or bosons through $\mathrm{H} \rightarrow \mathrm{W^+W^-}$ and $\mathrm{H} \rightarrow \mathrm{ZZ}$, and through higher-order processes to $\mathrm{H} \rightarrow \gamma\gamma$.

1.1.10 Top quark physics

The top quark is the most massive particle in the SM, with $m_t = 173.34 \pm 0.27(\mathrm{stat}) \pm 0.71(\mathrm{syst})$ GeV [23], such that it is kinematically allowed to decay through weak interaction $\mathrm{t} \rightarrow \mathrm{bW^+}$. This is the primary decay mode, since $|V_{tb}| \gg |V_{ts}|, |V_{td}|$ from the Cabibbo-Kobayashi-Maskawa mass matrix, which parametrises the strength of the Yukawa interaction between the quarks and the Higgs field [24, 25], and the hadronisation timescale $\Lambda^{-1}_{QCD} \simeq 3 \times 10^{-24}$ s is much larger than the top quark lifetime $\tau = 1/\Gamma_t \simeq 5 \times 10^{-25}$ s [26]. This means that the top quark mass can be measured accurately from a relatively small number of decay products, among which the mesons containing a b quark have a distinctive experimental signature due to their large lifetime.

The top quark was discovered in 1995 at the CDF experiment at Tevatron [27] and is now ubiquitous at the LHC, where it is produced primarily through gluon-gluon fusion in the form of top quark pairs [28].

1.2 The Standard Model at colliders

1.2.1 The parton model

Through the study of deep inelastic scattering (DIS) experiments, where an electron transfers sufficient four-momentum Q^2 to a proton for it to break up in the process $\mathrm{e}^-\mathrm{p} \to \mathrm{e}^-\mathrm{X}$, with X being a proton remnant, it was possible to establish that the electrons scatter elastically off point-like spin-half constituents of the proton, the partons [29]. This can be seen as an analogy to the Rutherford experiment, where α-particles were scattered off the nucleus of an atom to reveal the point-like structure of the nucleus within. By identifying the partons as the quarks from QCD, electron-proton interactions can thus be described in terms of the more fundamental electron-quark interactions.

Furthermore, by measuring the Q^2-dependence of the QCD coupling constant $\alpha_s(Q^2) = g_s^2/4\pi$ experimentally and from theoretical considerations of QCD [30, 31], it has been possible to establish that at high Q^2, the coupling constant $\alpha_S(Q^2)$ becomes small ($\alpha_s \simeq 0.1$), such that quarks can be treated as free particles in the asymptotic limit. The Q^2-dependence of α_s can be seen in fig. 1.3, confirming the QCD prediction of asymptotic freedom. Only when the coupling $\alpha_S(Q^2)$ is sufficiently small can perturbative QCD (pQCD) be used to compute cross-sections of the underlying processes. However, it is important to note that the magnitude of the strong coupling constant in the perturbative, high Q^2 regime is still significantly larger than the electromagnetic coupling constant, which means that higher-order loop processes are important in QCD.

The other feature of the running of α_s is that at a momentum scale comparable to typical hadron sizes $\Lambda_{\mathrm{QCD}} = 0.1, \ldots 0.3$ GeV, the coupling constant becomes very large, implying the breakdown of perturbative QCD at $Q < \Lambda_{\mathrm{QCD}}$. This means that in processes with low energy scales such as the production or interaction of hadrons, non-perturbative dynamics of QCD become important.

In order to model a proton-proton interaction with a hard scattering, such as the Drell-Yan process $\mathrm{q\bar{q}} \to \ell^+\ell^-$ at high Q^2, we describe the protons in terms of parton distribution functions (PDFs) $f_a^{\mathrm{p}}(x)$, which follows from the factorisation theorem in QCD [34]. These specify the fraction of proton momentum x carried by a quasi-free constituent quark or gluon a, b and factorise the proton-proton interaction to a hard interaction between quarks and gluons, as depicted in fig. 1.4. This allows us to write the factorised cross-section for the process $\mathrm{pp} \to cd$ as

$$\mathrm{d}\sigma(\mathrm{pp} \to cd) = \int_0^1 \mathrm{d}x_1 \mathrm{d}x_2 \sum_{a,b} f_a^{\mathrm{p}}(x_1, \mu_F^2) f_b^{\mathrm{p}}(x_2, \mu_F^2)\, \mathrm{d}\hat{\sigma}^{ab\to cd}(Q^2, \mu_f^2), \tag{1.12}$$

which is evaluated at the factorisation scale μ_F^2. In a qualitative sense, emissions with a transverse momentum above μ_F are accounted in the cross-section, otherwise, they are contained in the PDF. As it is not possible to describe the structure of hadrons and the PDFs perturbatively, they have to be determined from experimental data in terms of x and μ_F^2 [35, 36]. The proton is a dynamical system of bound quarks that interact strongly via virtual gluon exchange, so the protons are found to contain a *sea* of gluons and quarks and anti-quarks from the vacuum fluctuations of $\mathrm{g} \to \mathrm{q\bar{q}}$ in addition to the up and down *valence* quarks expected from flavour symmetry. The PDFs at different scales μ_F are related through the DGLAP [37, 38, 39][*] evolution equations, which result from QCD corrections and involve splitting kernels that arise from the basic QCD vertices. This allows us to deduce the PDFs at a certain scale from measurements at a different scale, making it possible to use data from various collider experiments in combined PDF fits and for making predictions at the LHC. The PDFs for protons at different factorisation scales are shown in fig. 1.5.

[*] Dokshitzer, Gribov, Lipatov, Altarelli and Parisi

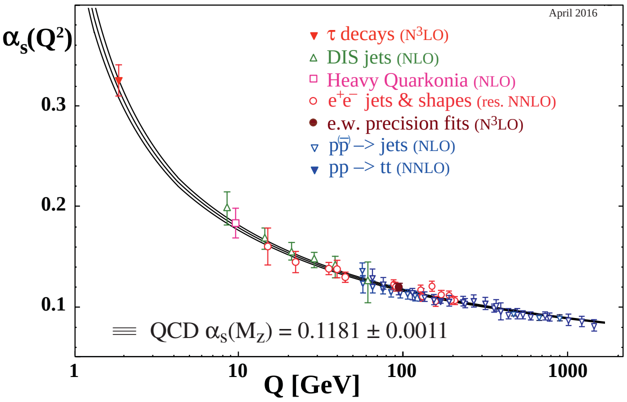


Figure 1.3: The measured values of the strong coupling $\alpha_S(Q^2)$, compared to the prediction from perturbative QCD with five-loop corrections [32]. We see that α_S decreases as the momentum scale Q^2 increases, reflecting the asymptotically free nature of QCD. Figure from [33].

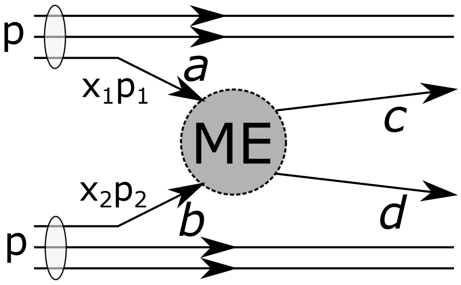


Figure 1.4: An illustration of the factorisation of a proton-proton interaction to a hard parton-parton interaction.

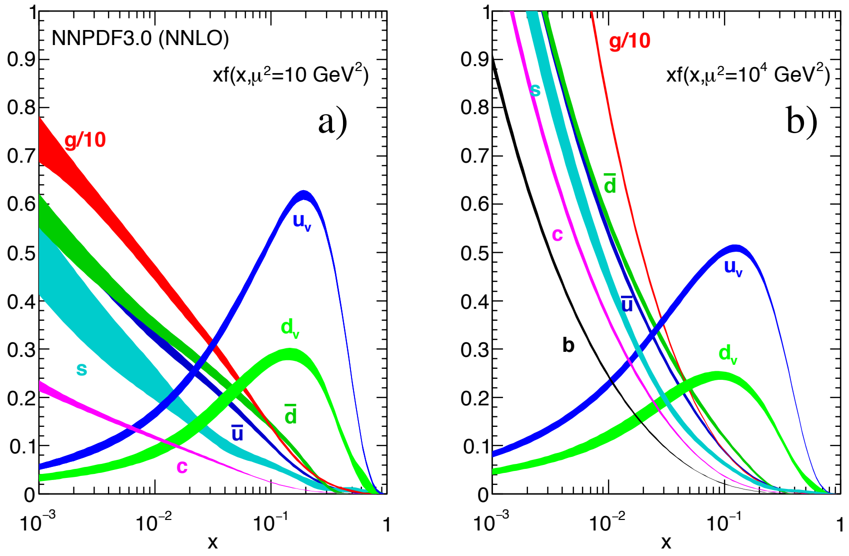


Figure 1.5: The parton distribution functions for protons at $\mu_F^2 = 10\ \mathrm{GeV}^2$ (left) and $\mu_F^2 = 10^4\ \mathrm{GeV}^2$. We see that at higher momenta μ_F and low x, the proton consists almost entirely out of gluons. Figure from [33].

1.2.2 Jet physics

In a process where QCD partons (quarks or gluons) are created, we cannot observe them directly, since colour confinement forbids the existence of isolated coloured states. Events with quarks and gluons in the final state can be observed experimentally as collimated jets of hadrons. This feature, which emerges from the short to long distance evolution of QCD, makes it possible to make measurements on macroscopic distance scales and connect them to otherwise inaccessible microscopic quantities such as the parton momenta or quantum numbers [40].

Jets are formed through the creation of a parton with high virtuality Q^2, which undergoes showering until it reaches a soft hadronisation scale, where non-perturbative processes create stable particles, mainly charged pions and protons ($\simeq 60\%$), neutral pions decaying to $\gamma\gamma$ ($\simeq 30\%$) and other neutral particles ($\simeq 10\%$). Due to the soft and collinear divergences in the QCD splitting functions, the jets are collimated along the direction of the original parton momentum.

In order to compare QCD predictions to experimental data, jet-related observables must be defined in such a way that they exhibit infrared safety, i.e. stability with respect to the addition of low-p_T particles, and collinear safety, such that the momentum of a single particle can be split up between collinear daughters. Furthermore, they must also be practical to compute for events with very high jet multiplicities that are expected in hadron colliders. This can be done using sequential recombination, where energy clusters with momenta $k_{t,i}$ and corresponding distances $d_{ij} = \min(k_{t,i}^{2p}, k_{t,}^{2p})\Delta_{ij}/R^2$ are recombined until stable jets can be formed. These jet algorithms are characterised by the jet radius parameter R and the parameter p, where $p = -1$ corresponds to the *anti-k_T* algorithm [41].

1.2.3 Monte Carlo event generators

In order to compute experimental observables such as differential distributions from the underlying theory of the SM, we typically use Monte Carlo (MC) simulators. Although inclusive observables and in some cases differential distributions can be calculated analytically, it is in general not possible to apply experimental cuts, therefore the observable phase space is sampled probabilistically to generate simulated events, which can be compared to the measurement [42].

The MC simulators generally have several stages, starting with importance sampling the phase space of the hard scattering process, such that events occur with probability proportional to $|\mathcal{M}|^2$. This is simple at leading order, but in higher orders, the cancellations between real and virtual contributions need to be accounted for and the procedure becomes more complicated [43]. The hard process defines the overall energy flow in the event.

The hard process is followed by the simulation of the parton shower, which evolves down from the hard scale by gluon emission or g → $q\bar{q}$ splittings. Showering is generally process-independent, but may depend on the inherent momentum scales in the process. At a scale of around 1 GeV, hadronisation takes over, which is usually treated using phenomenological models [44] for the non-perturbative dynamics. The overall multiplicity of final state particles is defined by the hadronisation models.

Additional radiation from gluon emission and splitting can be generated either at the level of the hard matrix element or in the parton shower in initial (ISR) or final state radiation (FSR). The matching between these two stages is therefore crucially important in accurately predicting the differential distributions, especially for events with high jet multiplicities. Furthermore, all MC implementations contain a number of free parameters, which need to be tuned to experimental data [45].

1.3 Higgs phenomenology

1.3.1 Higgs at the LHC

The discovery of the Higgs boson with $m_H \simeq 125$ GeV in 2012 at the LHC by the CMS [46] and ATLAS [47] collaborations confirmed the basic mechanism of EWSB and mass generation and completed the particle spectrum of the SM. Following this, a new experimental and theoretical program has opened in experimentally verifying the properties of the Higgs boson. In particular, it should be established whether the Higgs boson couples to SM gauge bosons and fermions as expected by observing these processes directly. In Run 1 of the LHC, the coupling of the Higgs boson to gauge bosons has been established at a relative precision of $\simeq 10\%$, however, to be able to determine the couplings to fermions, Run 2 data of the LHC are necessary [48].

Beyond establishing the existence of the predicted production and decay channels, a crucial test of the Higgs mechanism is determining the coupling strengths of the new scalar field to SM fields and comparing them to the SM predictions. The top quark Yukawa coupling y_t , which is $\mathcal{O}(10^5)$ times larger than that of the first-generation quarks, determines the evolution of the Higgs self coupling λ under renormalisation and is currently only known indirectly, as will be discussed further in section 1.3.5. Next, we discuss how the Higgs boson can be produced at the LHC.

1.3.2 Production modes

The main production modes of the Higgs boson at the LHC, shown in fig. 1.7, are through gluon-gluon fusion (ggF), with a cross-section at $\sqrt{s} = 13$ TeV of $\sigma_{\mathrm{ggF}} = 48.6 \pm 5\%$ pb, weak or vector boson fusion (VBF) with a cross-section $\sigma_{\mathrm{VBF}} = 3.78 \pm 2\%$ pb, associated production with vector bosons (VH) with $\sigma_{\mathrm{WH}} = 1.37 \pm 2\%$ pb, $\sigma_{\mathrm{ZH}} = 0.88 \pm 5\%$ pb and through the associated production with top quark pairs ($\mathrm{t\bar{t}H}$) with $\sigma_{\mathrm{t\bar{t}H}} = 0.50^{+9\%}_{-13\%}$ pb or with a single top quark (tH) [33]. We note that the difference between the cross sections for the ggF and $\mathrm{t\bar{t}H}$ production modes is roughly two orders of magnitude.

The ggF production proceeds predominantly via the top quark loop and is known to N3LO [49]. This was the main production mode for the initial observation with the $\mathrm{H} \to \gamma\gamma$, $\mathrm{H} \to \mathrm{W^+W^-} \to \mathrm{e}\nu\mu\nu$ and $\mathrm{H} \to \mathrm{ZZ} \to 4\ell$ signatures, which are important for the measurement of the production cross-section, J^P and mass of the Higgs boson. It also serves as an indirect constraint on the top-Higgs Yukawa coupling, due to the presence of the top quark loop in ggF. For the direct decay of the Higgs boson to fermions, which has a higher branching ratio than $\mathrm{H} \to \gamma\gamma$, additional production modes with a lower cross-section can be used. The VBF mode, where two (anti)quarks scatter by exchanging a weak boson in $\mathrm{qq} \to \mathrm{qqH}$ is important both for discovery and determination of the couplings, as it can be distinguished from QCD background through the presence of two jets in the opposite forward regions of the detector originating from the scattered quarks. The VH mode allows the use of leptonic decays of the W/Z bosons to reduce the multijet background, such that this mode has recently been used to establish the $\mathrm{H} \to \mathrm{b\bar{b}}$ decay [50]. While inclusive Higgs cross section measurements in the ggF channel are indirectly sensitive to the top quark Yukawa

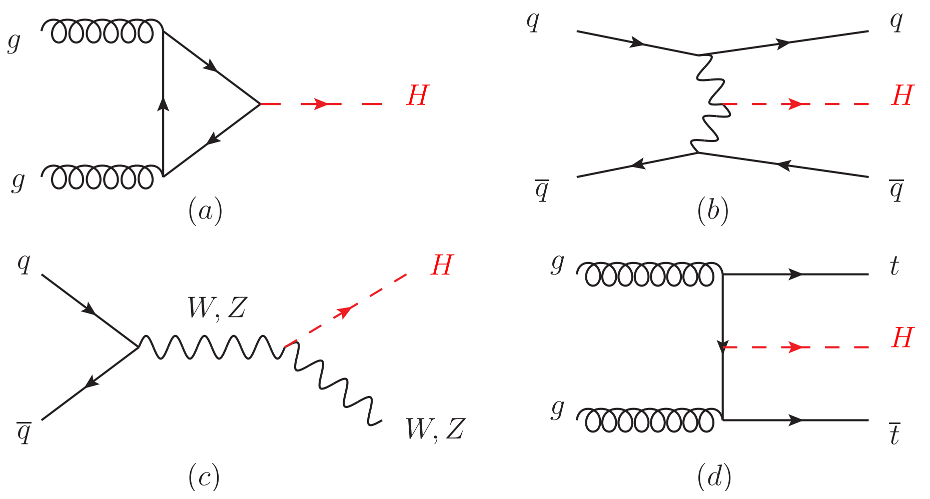


Figure 1.6: The generic tree-level Feynman diagrams for Higgs boson production: the gluon fusion process (ggF) (a), vector boson fusion (b), associated production with vector bosons (VH) (c) and associated production with top quarks (t$\bar{t}$H) (d). Figure from the PDG [33].

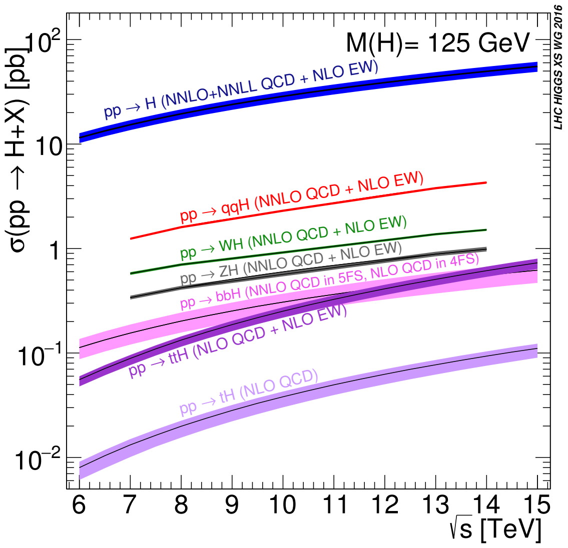


Figure 1.7: The production cross section of the SM Higgs boson with $m_H = 125$ GeV as a function of $\sqrt{s}$ along with theoretical uncertainties. Figure from the PDG [33].

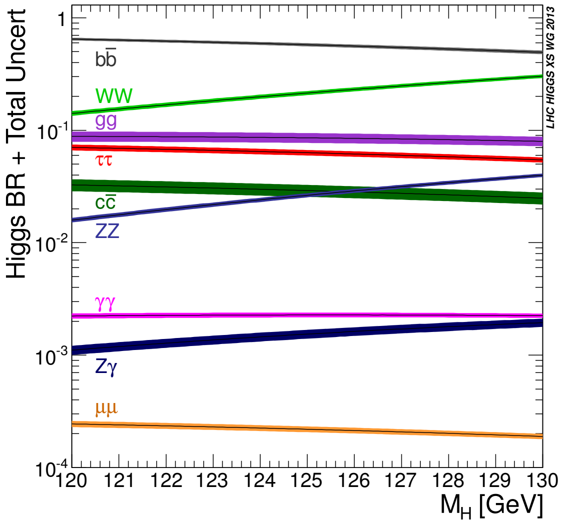


Figure 1.8: The branching ratios of the SM Higgs boson around $m_H = 125$ GeV along with the theoretical uncertainties. Figure from the PDG [33].

coupling y_t, the $\mathrm{t\bar{t}H}$ and tH channels can be used to probe y_t directly, such that they provide an important independent verification of the mass generation mechanism for the top quark.

1.3.3 Decay channels

The Higgs boson decays to SM particles with decay widths Γ that are proportional to the mass of the daughter particles. For $m_H \simeq 125$ GeV, the dominant decay channels are the decay to bottom quarks $\mathrm{H} \to \mathrm{b\bar{b}}$ with a branching ratio $\mathrm{BR} = \Gamma/\Gamma_H = 0.584 \pm 0.019$, the decay to an on-shell and an off-shell W-boson $\mathrm{H} \to \mathrm{WW}^*$ ($\mathrm{BR} = 0.214 \pm 0.009$), the decay to Z bosons $\mathrm{H} \to \mathrm{ZZ}$ ($\mathrm{BR} = 0.026 \pm 0.001$) and the decay to tau leptons $\mathrm{H} \to \tau^+\tau^-$ ($\mathrm{BR} = 0.063 \pm 0.004$) [33]. Through loop-induced processes, which are enhanced by the large top quark mass, the Higgs boson can also decay to massless particles in the channels $\mathrm{H} \to \mathrm{gg}, \gamma\gamma$ with non-negligible branching fractions. The branching ratios for a SM Higgs boson as a function of m_H can be seen in fig. 1.8.

1.3.4 Experimental characterisation

Accurate predictions for the Higgs boson production cross sections and branching ratios for decays allow the interpretation of experimental data and measuring the properties of the Higgs boson in various channels. The mass m_H has been measured most accurately in the $\mathrm{H} \to \gamma\gamma$ and $\mathrm{H} \to \mathrm{ZZ} \to 4\ell$ channels, with a combined value of $m_H = 125.09 \pm 0.21$ (stat.) ± 0.11 (syst) GeV, dominated by statistical uncertainties, with the photon momentum scale uncertainties being the most significant systematic uncertainty [51]. This has been confirmed in a recent measurement of the Higgs boson properties in the $\mathrm{H} \to \mathrm{ZZ} \to 4\ell$ channel, where the Higgs boson mass has been determined to be $m_H = 125.26 \pm 0.2$ (stat.) ± 0.08 (syst.), with the systematic uncertainties dominated by the lepton momentum scale [52]. The spin and parity J^P of the Higgs boson are probed independently of the mass and total cross section and are found to be compatible with the SM 0^+ hypothesis, excluding the pseudo-scalar hypothesis at a $98-99\%$ confidence level [53, 54]. The width of the Higgs boson cannot be measured directly at the LHC, since the mass resolution in the diphoton and 4ℓ channels is $1-2$ GeV, three orders of magnitude larger than the expected SM line width $\Gamma_H = 4.2$ MeV [33]. However, it can be constrained by comparing the on-shell and off-shell $\mathrm{H} \to \mathrm{VV}$ cross sections, thus setting upper limits on Γ_H that are around 5-6 times the SM value [55].

It is important to experimentally confirm that the coupling strengths of the Higgs boson to SM particles correspond to those predicted by the SM. Any deviation from the values predicted by the SM Higgs mechanism could thus signal physics beyond the SM (BSM). The simplest way to experimentally characterise discrepancies in the couplings is the so-called κ-framework [56], where

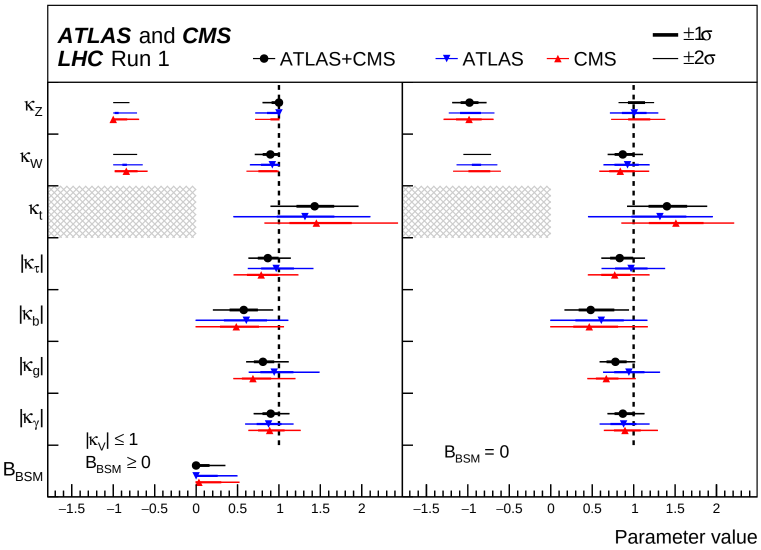


Figure 1.9: The combined signal strength modified factors κ from the CMS and ATLAS collaborations with Run 1 data. Figure from [48].

the couplings to SM particles are rescaled by factors κ_i for $i \in \{\mathrm{Z}, \mathrm{W}, \mathrm{f}, \mathrm{g}, \gamma, \mathrm{Z}\gamma\}$ which can be determined from signal strength ($\mu = \sigma/\sigma_{\mathrm{SM}}$) measurements, without modifying the SM structure of the theory. In this formalism, the t$\bar{\mathrm{t}}$H signal strength modifier is given by $\mu_{\mathrm{t\bar{t}H}} = \kappa_{\mathrm{t}}^2$. The CMS and ATLAS collaborations have extracted these signal modifier values from Run 1 data in a combined fit, with the results shown in fig. 1.9. The couplings, while compatible with the SM, have significant uncertainties, in particular for the Higgs-top quark coupling κ_t, thus paving the way for Run 2 measurements with improved sensitivity.

While the κ-framework is relatively simple to apply for small deviations in signal strength, the clear shortcoming of this approach is that any BSM physics would necessarily change the structure of the theory, rendering the results potentially invalid. In particular, the above approach assumes that the loop contributions in ggF and H → $\gamma\gamma$ are not modified by new physics. However, contributions with a different Lorentz structure from the SM cannot be captured in the κ framework. Therefore, it has been suggested to use either Higgs pseudo-observables [57] or effective field theories (EFT) to further parametrise any possible deviations from the SM [58, 59].

1.3.5 Top-Higgs coupling

In the SM, the coupling between the top quark and the Higgs boson is predicted to be $y_t = \sqrt{2}m_t/v$, which can be verified by measuring the cross sections of top quark pair associated Higgs production. This implies that the direct determination of the top-Higgs coupling in t$\bar{\mathrm{t}}$H is a test of the EWSB model at a natural scale where $y \simeq 1$.

Vacuum stability

The top-Higgs coupling plays an important role in vacuum stability, since it controls the evolution of the self-coupling λ of the Higgs potential through renormalisation evolution $\frac{\mathrm{d}\lambda}{\mathrm{d}\ln\mu}$, where it gives a quartic negative contribution. In particular, if y_t is sufficiently large but well within the bounds set by experimental uncertainties, the self-coupling λ becomes negative at a renormalisation scale μ below the Planck scale $M_{\mathrm{Pl}} = \sqrt{\hbar c/G} \simeq 10^{19}$ GeV, where gravitation becomes important. This means that the Higgs potential develops an additional minimum, which would possibly make the SM vacuum state unstable or metastable, depending on the exact values of m_H and m_t [60]. Given that we have not observed a transition between the vacuum states, this would imply the existence of new physics that would prevent this. Since the scale where the scalar self coupling becomes negative depends strongly on y_t, an accurate determination of y_t can help to pinpoint the scale of new physics in the absence of clear BSM signals [61] and establish whether the vacuum is stable on cosmological timescales.

Anomalous couplings

Furthermore, the top-Higgs coupling, which is purely scalar in the SM, can be extended quite generally to contain scalar and pseudoscalar interactions, making it possible to use results from $\mathrm{t\bar{t}H}$ for setting direct constraints on anomalous top-Higgs couplings [62]. Such anomalous couplings can arise from a two-Higgs doublet model [63, 64] that appears in several BSM scenarios, such as supersymmetry [65], axions [66] or from models with a composite Higgs [67]. Measurements of the $\mathrm{t\bar{t}H}$ cross-section can help to constrain such anomalous couplings.

1.3.6 Current results on $\mathrm{t\bar{t}H}$

Both the CMS and ATLAS collaborations have searched for $\mathrm{t\bar{t}H}$ in Run 1 and Run 2 of the LHC by measuring $\mu = \sigma_{\mathrm{t\bar{t}H}}/\sigma_{\mathrm{SM}}$ in a $\mathrm{t\bar{t}H}$-enriched region using different Higgs decay channels, with the $\mathrm{H} \to \mathrm{WW}^*, \mathrm{ZZ}^* \to$ multilepton, $\mathrm{H} \to \mathrm{b\bar{b}}$ and $\mathrm{H} \to \tau\tau$ decay channels providing the highest sensitivity in Run 2, followed by $\mathrm{t\bar{t}H}$-tagged $\mathrm{H} \to \gamma\gamma$ decay in the diphoton analysis. The results are summarised in table 1.1. The analyses are systematically dominated and quite involved in terms of statistical methods. A recent combination by ATLAS while this work was in preparation has been able to measure a cross-section of $\sigma_{\mathrm{t\bar{t}H}} = 590^{+160}_{-150}$ fb, compared to the SM value of $\sigma^{\mathrm{SM}}_{\mathrm{t\bar{t}H}} = 507^{+35}_{-50}$ fb, with the equivalent measurement from CMS in the $\mathrm{H} \to \mathrm{b\bar{b}}$ channel being a major part of this thesis.

decay channel	CMS	ATLAS
$\mathrm{H} \to \mathrm{b\bar{b}}$	$\mu = -0.19 \pm 0.45$ (stat) $\pm\, 0.68$ (syst) $\mu < 1.5$ (1.7) obs (exp) @ 95% CL 12.9 fb^{-1} [68]	$\mu = 0.84 \pm 0.29$ (stat) $\pm\, 0.57$ (syst) $\mu < 2.0$ (1.2) obs (exp) @ 95% CL 36.1 fb^{-1} [69]
multi-lepton	$\mu = 1.5 \pm 0.5$ $\mu < 2.0$ (2.2) @ 95% CL 35.9 fb^{-1} $(\mathrm{WW}^*, \mathrm{ZZ}^*, \tau_\ell\tau_\ell)$ [70]	$\mu = 1.6^{+0.5}_{-0.4}$ 4.1σ obs (2.8σ exp.) 35.9 fb^{-1} [71]
$\mathrm{H} \to \tau\tau$	$\mu = 0.72^{+0.62}_{-0.53}$ $\mu < 1.3$ (1.4) @ 95% CL 35.9 fb^{-1} $(\tau_h\tau_l)$ [72]	included in multi-lepton
$\mathrm{H} \to \gamma\gamma$	$\mu = 2.2^{+0.9}_{-0.8}$ 3.3σ obs (1.5σ exp.) 35.9 fb^{-1}, $\mathrm{t\bar{t}H}$ tag [73]	$\mu = 0.5 \pm 0.6$ $\mu < 1.7$ (2.3) @ 95% CL 36.1 fb^{-1}, $\mathrm{t\bar{t}H}$ tag [74]

Table 1.1: Current results on $\mathrm{t\bar{t}H}$ production from Run 2 of the LHC.

1.4 Summary

The discovery of the Higgs boson in 2012 has opened up a new field of research in experimentally probing the properties of this fundamental scalar field. Measurements of the Higgs boson are entering the precision era with an increase of attention on rare production and decay modes, which allow the nature of the EWSB mechanism to be verified experimentally. In Run 2 of the

LHC, several rare production modes have confirmed the coupling of the Higgs boson to quarks and leptons. Many BSM theories predict changes to the EWSB sector, with the determination of the top quark Yukawa coupling being an important test of the mechanism of mass generation. The LHC experiments are already probing this coupling, but work remains to be done on both the experimental side in reducing measurement uncertainties and on the side of theoretical predictions. The current dataset collected by the CMS and ATLAS experiments is approximately 1% of the total foreseen over the lifetime of the LHC project; therefore considerable progress can be made in precision measurements of the Higgs boson and thus the SM.

2 Experimental setup

In this section, we give an overview of the experimental setup of the Large Hadron Collider (LHC) and the Compact Muon Solenoid (CMS) experiment as it pertains to the $\mathrm{t\bar{t}H}$ search during Run 2 of the LHC. First, we summarise the essential details of the LHC machine and the main experiments on the LHC ring. At the time of writing, the LHC, which is located at CERN nearby Geneva, is the highest-energy hadron collider in the world and the only experimental apparatus where Higgs bosons can be produced and studied directly. Next, we discuss the CMS experiment, which is one of the two general-purpose detectors on the LHC ring. In particular, we describe the subsystems and reconstruction algorithms of the detector that are crucial to the $\mathrm{t\bar{t}H}(\rightarrow \mathrm{b\bar{b}})$ search.

2.1 The Large Hadron Collider

The LHC is a hadron collider situated in the 26.7-km tunnel of the former Large Electron Positron (LEP) machine. The hadrons, either protons or heavy ions, are accelerated in several stages by a pre-accelerator complex, shown in fig. 2.1. In the description that follows, we consider only proton-proton collisions for concreteness. The protons are accelerated by the linear accelerator LINAC to 50 MeV, the proton synchrotron booster (PSB) to 1.4 GeV, the proton synchrotron (PS) to 26 GeV and by the super proton synchrotron to 450 GeV. The protons are injected in bunches of $N_b \simeq 10^{11}$ protons per bunch into the main rings with a frequency of 40 MHz, such that the nominal bunch spacing is 25 ns and there are approximately $n_b \simeq 2500$ bunches per beam.

2.1.1 The accelerator complex

The protons are accelerated to the centre-of-mass energy $\sqrt{s} = 13$ TeV in the main accelerator system consisting of two concentric counter-rotating rings, where superconducting magnets with a nominal B-field $B = 8.33$ T are used to bend and focus the beams. Dipole magnets, depicted in fig. 2.2, are used to bend the beam and multipole magnets to focus it. Both beams are located in beam pipes with an inner diameter of 48 mm within the same vacuum chamber in a so-called twin-bore design, dictated by the size of the tunnel. The magnets are kept at an operating temperature below 2K using superfluid He-4. These proton bunches are collided at four interaction points and the beams can be sustained for up to 24 hours. The machine is characterised by the nominal instantaneous luminosity of $L = 10^{34}$ $\mathrm{cm^{-2}s^{-1}}$, which has been exceeded in 2017 by a factor of 2.06 [76].

The proton-proton collisions at the LHC interaction points result in a number of events per second given by $N_i = L\sigma_i$ for a process that has a cross-section σ_i at a given instantaneous luminosity L. For the LHC, the machine luminosity is given by

$$L = \frac{N_b^2 n_b f_{\mathrm{rev}} \gamma_r}{4\pi \epsilon_n \beta^*} F \tag{2.1}$$

where f_{rev} is the number of revolutions per second and γ_r is the relativistic Lorentz factor. The beam is further described by the normalised transverse emittance ϵ_n and the beta function β^* of the beam at the interaction point, which are related to the transverse beam size at a location s along the beam through $\sigma(s) = \sqrt{\epsilon_n \beta(s)}$. The transverse emittance characterises the spread of particles in the position-momentum phase space throughout their orbits. The beams cross at the interaction points at an angle θ_c, which results in luminosity reduction by a factor

$$F = \left[1 + (\frac{\theta_c \sigma_z}{2\sigma^*})^2\right]^{-1/2} \tag{2.2}$$

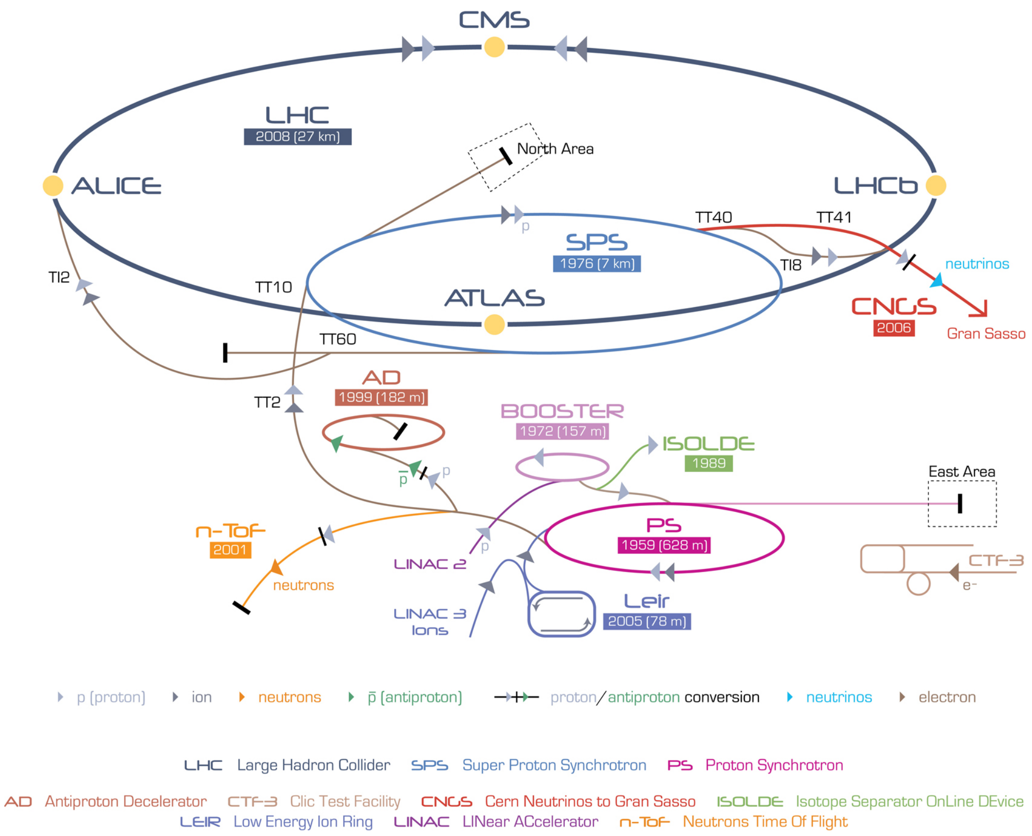


Figure 2.1: The LHC accelerator complex, where the protons are accelerated by LINAC 2, the PS and the SPS before being injected into the main LHC ring, where they are collided in the four interaction points. Figure from [75].

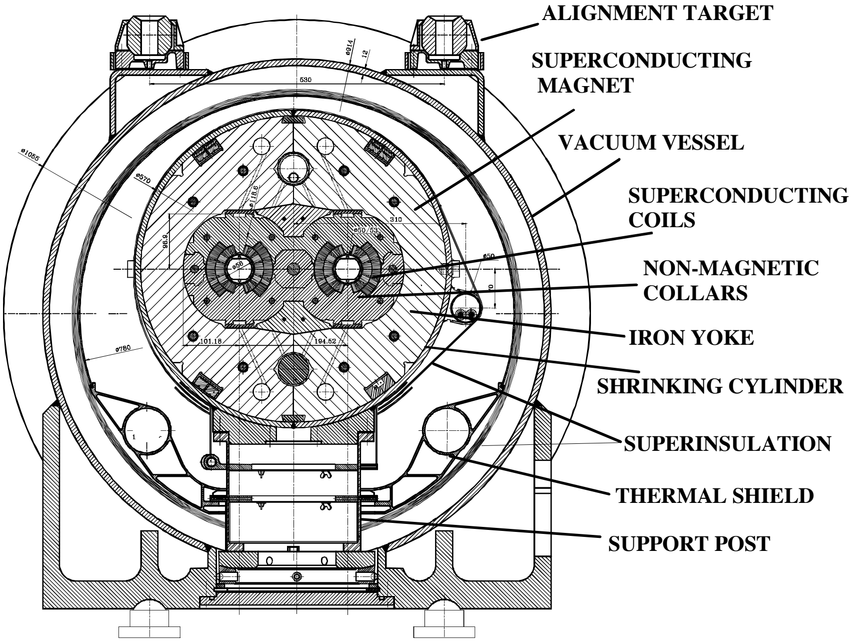


Figure 2.2: Cross-section of the LHC dipole magnet. Figure from [77].

where σ_z is the bunch length and σ^* the transverse bunch size. The β^* parameter dictates the size of the beam at the interaction point, which is tuned to the luminosity requirements of the experiment and in turn limited by the aperture of the focusing magnets. The maximum beam size in the transverse direction is $\sigma = 1.2$ mm, limited by the dimensions of the beam screen. The instantaneous luminosity decays over time during a physics run due to beam loss from the collisions, such that the luminosity decreases by $1/e$ during approximately 15 hours.

The total number of collisions in a given unit of time is characterised by the integrated luminosity $\mathcal{L} = \int L \, \mathrm{d}t$ and is limited to around $80 - 120$ fb^{-1} per year under perfect conditions, assuming around 200 days of operation and on average seven hours of turn-around time between runs for filling the beams and accelerating to data-taking energies [77].

During the 2016 data taking period considered in this thesis, the LHC operated at a centre-of-mass energy of $\sqrt{s} = 13$ TeV, $\beta^* = 40$ cm and delivered around $\mathcal{L} = 40$ fb^{-1} of proton-proton data to the ATLAS and CMS experiments. With an inelastic pp cross-section of $\sigma_{pp} \simeq 77$ mb [78], this corresponds to about 10^{15} inelastic pp interactions per year.

2.1.2 Experiments at the LHC

The collision data from the LHC are recorded by two general purpose detectors, CMS and ATLAS (A Toroidal LHC ApparatuS), and two experiments with a more specialised physics program, LHCb and ALICE (A Large Ion Collider Experiment), located at the four interaction points. The general properties and physics goals of the main experiments are described in the following section.

CMS

The main characteristic of CMS is the superconducting solenoid, which provides a magnetic field of 3.8T that enables the momentum of charged particles to be measured with high accuracy. Inside the solenoid volume are silicon pixel and strip trackers, an electromagnetic calorimeter (ECAL) comprised of $PbWO_4$ crystals and a brass-scintillator hadronic calorimeter (HCAL). Outside the solenoid volume is the steel return yoke for the magnetic field, which contains gas-ionisation chambers used to measure muons. The CMS detector was designed to meet a dimuon, diphoton and dielectron mass resolution of $\simeq 1\%$ at 100 GeV [79].

ATLAS

The overall layout of the ATLAS detector differs from CMS mainly with respect to the configuration of the magnetic fields, where a central superconducting solenoid with B=2 T houses the semiconductor trackers, with the lead-liquid argon (LAr) electromagnetic calorimeter and the hadronic calorimeters outside the solenoid volume. The muon systems are embedded in an outer air-core toroidal system that minimises multiple scattering [80].

LHCb

The primary goal of the LHCb experiment is to study heavy flavour physics, in particular rare decays of beauty and charm hadrons and searching for indirect evidence for new physics in CP-violation, exploiting the large rate of B-meson production at the LHC. The LHCb detector is a single-arm spectrometer with a forward angular coverage of 10 to 250-300 mrad, featuring a beryllium beampipe that is highly transparent to particle fluxes and an accurate vertexing system. LHCb operates at an instantaneous luminosity that is two orders of magnitude lower than CMS and ATLAS in order to minimise multiple pp interactions per bunch crossing [81].

ALICE

The ALICE detector is designed to study heavy ion collisions, focusing on QCD measurements, in particular the study of strongly interacting matter at the high temperatures and densities achievable in nucleon-nucleon collisions. It features a barrel region embedded in a solenoid which measures hadrons, electrons and photons, and muon spectrometers in the forward direction. The ALICE detector is specifically optimised to study global event observables such as particle multiplicity and energy flow [82].

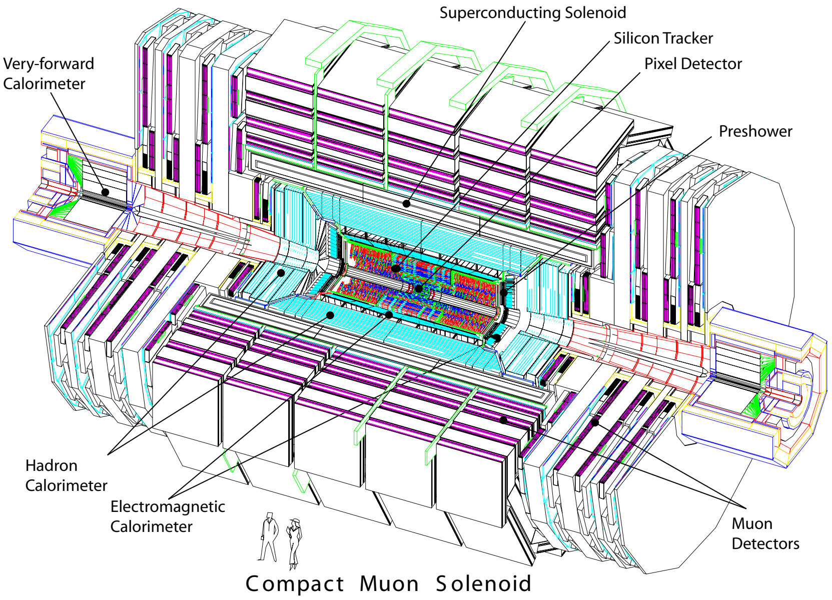


Figure 2.3: The cross-section of the CMS experiment. Figure from [79].

Purpose

Both the CMS and ATLAS experiments have the broad physics goal of discovering the Higgs boson, studying its properties and searching for any new resonances or other new phenomena at high energies. The discovery of the Higgs boson was realised during Run I of the LHC, with both experiments reporting a significant excess in 2012 that is compatible with a SM Higgs boson [47, 46]. In the following section, we discuss the essential aspects of the CMS experiment in more detail.

2.2 The CMS detector

The coordinate system adopted by CMS is centred at the collision point, with the x-axis pointing inward towards the LHC ring, the y-axis vertically upward and the z-axis along the beamline towards east in the direction of the Jura mountains. The azimuthal angle ϕ is measured from the x-axis in the plane transverse to the beam. The polar angle is measured from the z-axis and it defines the pseudorapidity $\eta = -\ln \tan \theta/2$. The CMS detector follows a layered design that encapsulates the interaction region completely in the azimuthal direction and provides good coverage in the polar direction. In order to achieve this, the detector is divided into a barrel component and two endcaps, as can be seen in fig. 2.3. A transverse slice of the experiment can be seen in fig. 2.4, where the overall layout of the sub-detectors is depicted along with example particle trajectories for muons, electrons, charged and neutral hadrons and photons.

2.2.1 The superconducting magnet

The 3.8T magnetic field, which is essential for measuring the momenta and charge of charged particles, is created by the 220 ton superconducting solenoid, which has a diameter of 6 m and a length of 12.5 m. The energy stored in the relatively thin NbTi conductor reaches up to 2.6 GJ and mechanical deformations in the magnet during energising can be significant ($\simeq 0.15\%$). The iron return yoke of the magnetic field consists of sections which house the muon chambers and thus guarantees a sufficient field strength in the muon spectrometer region [79].

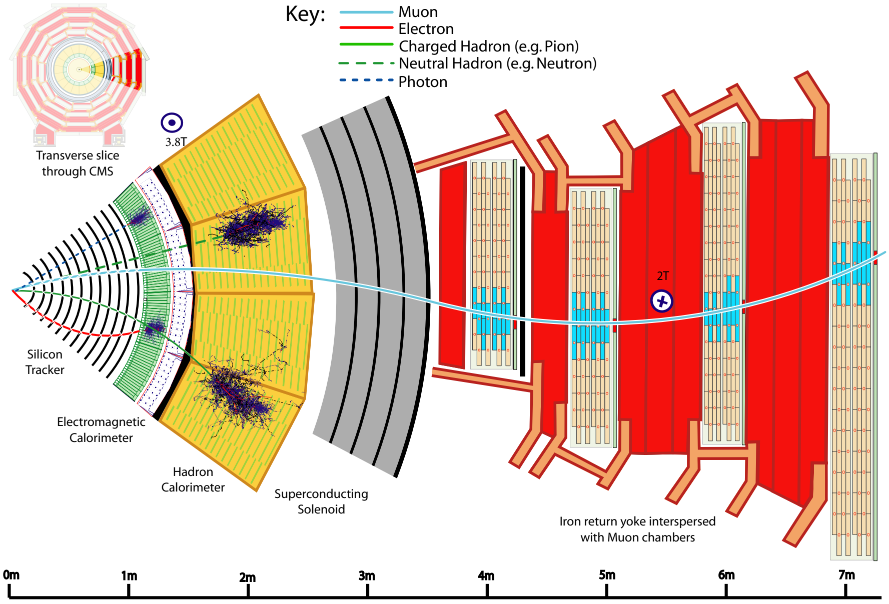

Key:
Muon
Electron
Charged Hadron (e.g. Pion)
Neutral Hadron (e.g. Neutron)
Photon
Transverse slice through CMS
3.8T
2T
Silicon Tracker
Electromagnetic Calorimeter
Hadron Calorimeter
Superconducting Solenoid
Iron return yoke interspersed with Muon chambers
0m
1m
2m
3m
4m
5m
6m
7m

Figure 2.4: A view of a transverse slice of the CMS detector with the subsystems.

2.2.2 Inner tracking system

The inner tracking system measures the momenta, trajectories and charge of charged particles from the primary vertex in the interaction point and any secondary vertices associated to the decay of long-lived particles such as b hadrons. The magnetic field is homogeneous within the tracker volume, which has a length of 5.8 m and a diameter of 2.5 m. The number of simultaneous inelastic collision (pileup) events per bunch crossing in pp collisions can be significant at the LHC, with nominal values around $N_{PV} = 20 - 50$. Therefore, the tracking system has to cope with high particle fluxes of up to 10^3 particles per bunch crossing every 25 ns and be able to associate signals to the correct bunch crossing.

Pixel detector

In order to keep the hit occupancy around 1%, pixel detectors have to be used in the inner region of the tracker. The inner tracker consists of a three-layer silicon pixel detector with layers at 4.4 cm, 7.3 cm and 10.2 cm and a silicon strip detector with ten layers extending out to radii of 1.1 m. The pixels in the inner layer measure $100 \times 150\ \mu\text{m}^2$ in the $r - \phi$ and z directions, which is driven by the secondary vertex and impact parameter resolution necessary for the detection of heavy flavour states. The pixel detector covers the range $|\eta| < 2.5$ and consists of the barrel layers (BPix) and two endcap discs (FPix), located in such a way as to guarantee at least three pixel hits over almost the full range, as seen in fig. 2.5. By reading out the analog pulse height using charge sharing which results from the B field, a hit resolution of 15-20 μm can be achieved. During Run 1 of the LHC and the 2015-2016 data taking period, the Phase-0 version pixel detector, covering an area of 1 m^2 and consisting of around 66 million pixels was in operation. The pixel detector was upgraded during the 2016-2017 winter shut-down procedure as part of the Phase-1 upgrade, moving the inner layer from $r = 4.4$ cm to $r = 2.9$ cm, adding a new outer layer at $r = 16$ cm and new disks in the forward directions. The capabilities of the readout chip (ROC) have been improved to cope with an instantaneous luminosity of $L = 2 \times 10^{34}\ \text{cm}^{-2}\text{s}^{-1}$ [83].

Silicon strip tracker

At larger radii, the occupancy decreases, such that silicon strips with a typical size of 10cm × 80μm are used in the silicon strip tracker. The tracker consists of several layered subsystems as

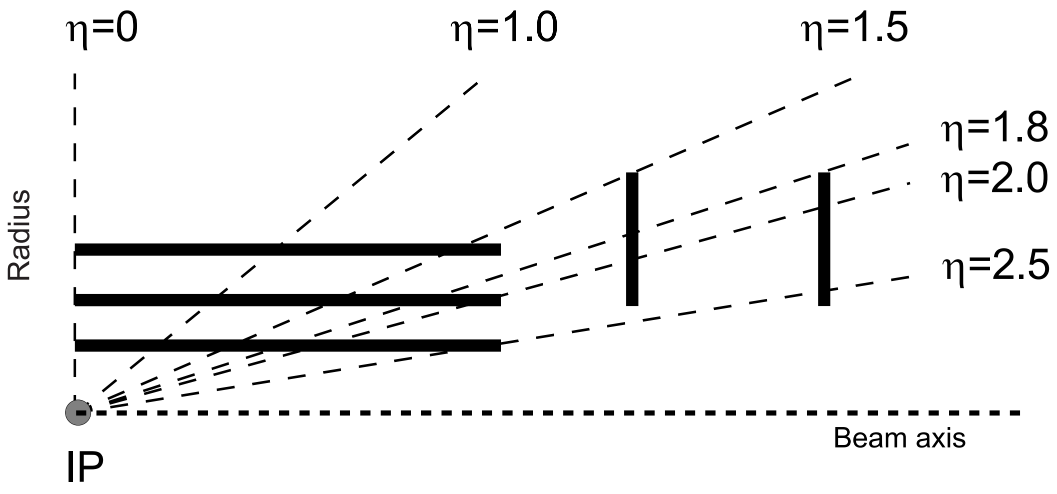


Figure 2.5: A schematic overview of the geometrical layout of the Phase-0 CMS pixel tracker, showing the three BPix layers at mean radii of 4.4, 7.3 and 10.2 cm and the two FPix disks at $z = \pm 34.5$ cm and ± 46.5 cm. Figure from [79].

shown in fig. 2.6, in particular, the tracker inner barrel and discs (TIB/TID), the outer barrel (TOB) and the tracker endcaps (TEC). The TIB/TID delivers four measurements of a track in the $r - \phi$ direction, the TOB six measurements, and the TEC nine measurements per particle trajectory. The strip pitch increases at larger radii to compensate for the reduction in particle flux.

Overall, the tracking system has to maximise the number of measurement points for each particle trajectory while keeping the material budget at a minimum. The amount of interacting material can be measured in terms of a radiation length X_0, which corresponds to the distance over which the energy of a high-energy charged particle is reduced by a factor e, mainly due to Bremsstrahlung. Over the whole functional η range, the tracking system contributes between 0.4 and 1.8 radiation lengths, with the largest radiation losses in the region around $|\eta| \simeq 1.5$ due to the TIB/TID transition. The transverse impact parameter resolution is around 10 μm for high momentum tracks. The tracking system contributes significantly to muon reconstruction, resulting in a transverse momentum resolution of around $1-2\%$ up to $|\eta| \leq 1.6$ and a reconstruction efficiency of around 99% over most of the η range for muons with $p_T \simeq 100$ GeV [79, 84].

2.2.3 Electromagnetic calorimeter

The primary function of the electromagnetic calorimeter is to measure the energy of electrons and photons through the production of scintillation light from electromagnetic cascades produced by high-energy electrons or photons. The ECAL is situated within the solenoid volume in order to minimise energy losses from radiative processes, pair creation and hadronic interactions and consists of around 76 000 lead-tungstate ($PbWO_4$) crystals arranged in the barrel and endcaps, as seen in fig. 2.7. This material is characterised by a high density ($\rho = 8.28$ g/cm^3), a short radiation length ($X_0 = 0.89$ cm) and a small Molière radius ($R_M = 2.19$ cm) [33], which determines the transverse size of the electromagnetic shower. Furthermore, the scintillation decay time is short, such that 80% of the light is emitted during the 25 ns bunch spacing. The light is emitted with a broad maximum in the 420-430 nm range [79].

In the barrel (endcaps), the crystals are coupled to avalanche photodiodes (vacuum phototriodes) for light collection, with a 1 MeV particle producing a yield of about 4.5 photoelectrons. Due to the high radiation damage expected throughout the lifetime of the ECAL, the light transmission properties of the crystals are monitored using injected laser light at $\lambda = 440$ nm. A two-layer lead

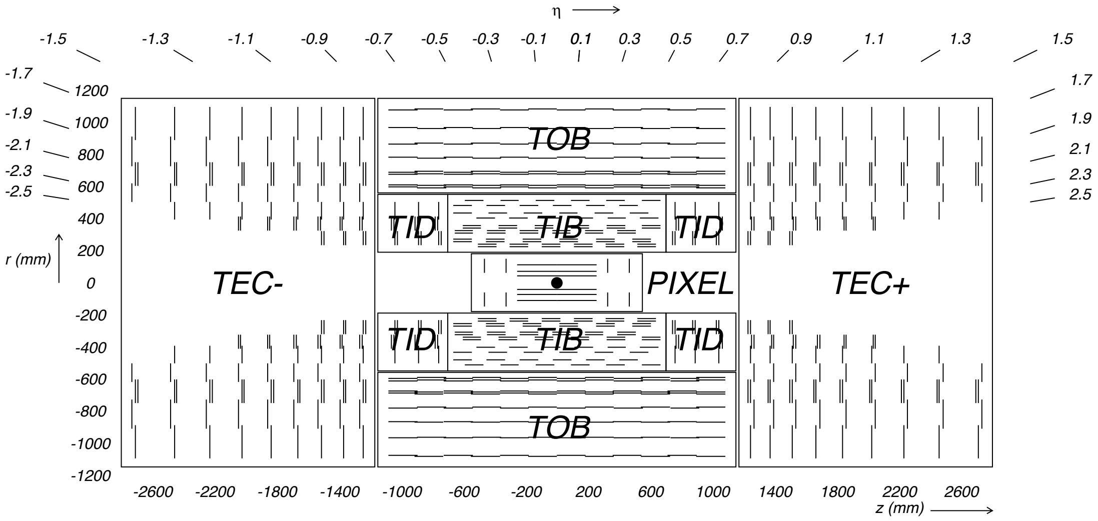


Figure 2.6: The schematic overview of the CMS strip tracking system. Figure from [79].

absorber and silicon strip sensor preshower detector is located between the endcaps and the interaction point, covering $1.65 < |\eta| < 2.6$, in order to improve the discrimination between photons and neutral pions.

The ECAL barrel region extends to $|\eta| < 1.479$, with a 360-fold (190-fold) granularity in the azimuthal (polar) direction. The endcaps cover the range between $1.479 < |\eta| < 3.0$. Since the photon emission in scintillation and subsequent amplification are temperature-dependent, the ECAL has to be maintained at a constant temperature within 0.05° C. In order to reconstruct the signal pulse from the photodetectors, a new technique is used in Run 2, where the signal amplitude templates from up to nine bunch crossings around the in-time signal are fitted to the observed ten-sample signal, in order to determine the signal amplitude in the presence of both in-time and out-of-time pileup [85]. The energy resolution of the ECAL has been measured in electron test beams, arranging the crystals in a 3×3 matrix to minimise energy leakage, and found to be described by

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c \tag{2.3}$$

with $a = 2.8\%$ being the stochastic term, $b = 12\%$ the noise term and $c = 0.3\%$ the irreducible term from non-uniformities for the barrel [86].

Since there is about $1-2$ radiation lengths of material in front of the ECAL and the crystals are about one Moliere radius in the lateral dimension, the energy from a single electromagnetic shower is spread over multiple crystals. In order to reconstruct the energy of incident particles, the energy that is spread over multiple ECAL crystals is clustered by merging crystals into superclusters. The ECAL has an excellent energy resolution, with photons from the decay of a 125-GeV Higgs boson being reconstructed in the barrel with an energy resolution between 1-3% [87] and a similar resolution for electrons [88].

2.2.4 Hadronic calorimeter

The purpose of the HCAL is to measure the energies of hadron jets and to have a hermetic energy coverage of the detector, such that the missing transverse energy resulting from neutrinos or hypothetical weakly interacting massive particles which may arise from BSM theories [89] could be determined. The HCAL consists of a barrel and endcaps extending to $|\eta| < 3.0$, with a forward calorimeter covering the range up to $|\eta| < 5.0$, as seen in fig. 2.8.

The HCAL barrel is situated between the ECAL and the superconducting coil in the region $1.77 \text{ m} < R < 2.95 \text{ m}$. As this constrains the volume and thus the amount of material,

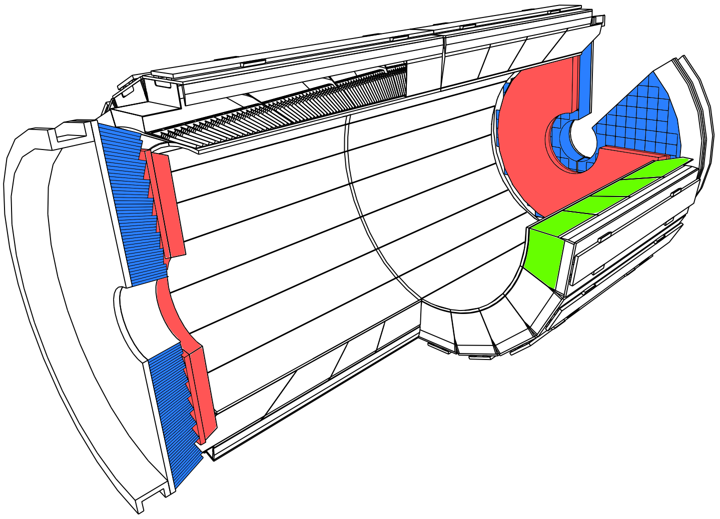


Figure 2.7: The CMS electromagnetic calorimeter, with the barrel crystals in green, the endcaps in blue and the pre-shower in red. Figure from [79].

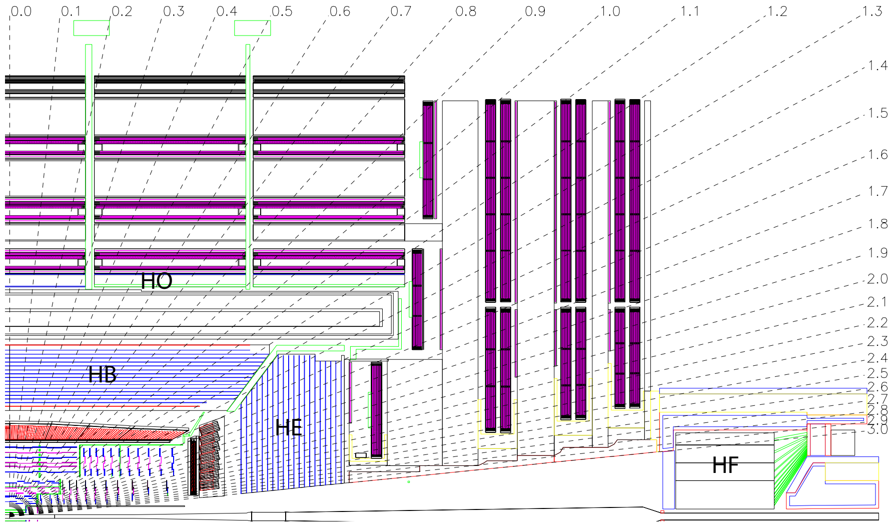


Figure 2.8: The cross-section of the CMS hadronic calorimeter, showing the barrel region (HB), the endcap (HB), the outer calorimeter (HO) and the forward calorimeter (HF). Figure from [79].

an outer hadron calorimeter is installed outside the solenoid in the barrel region, such that the material amounts to around 11 hadronic radiation lengths. The HCAL barrel regions extends to $|\eta| < 1.3$ and consists of an absorber made from brass sandwiched between steel plates with embedded scintillator tiles. The scintillation light is read out by bringing the light to photodiodes in readout towers using wavelength shifting fibres. Hybrid photodiodes are used due to their low sensitivity to the magnetic field and their large dynamic range.

The HCAL endcaps cover the pseudorapidity range $1.3 < |\eta| < 3$, thus they need to handle high counting rates and to be radiation hard. They follow a similar construction as the barrel, with an absorber/scintillator design with a segmentation granularity of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ for $|\eta| < 1.6$ and $\Delta\eta \times \Delta\phi \simeq 0.17 \times 0.17$ for the rest of the endcap [79].

The forward calorimeter (HF) located at ±11 m from the interaction point, covering the range $3.0 < |\eta| < 5.2$, is situated in an extreme radiation environment of up to 100 Mrad/year, thus radiation hardness has been the primary design criterion. It is based on the collection of Cherenkov light collected in quartz fibres embedded in steel absorber plates, read out by photomultiplier tubes. The HF is mostly sensitive to the electromagnetic component of showers [90]. The HF can also be used for luminosity monitoring at CMS to infer the mean number of interactions per bunch crossing and thus an accurate determination of the normalisation for physics analyses.

The energy resolution of the HCAL has been determined in test beams using single pions and found to be approximately $\sigma/E = 110\%/\sqrt{E} \oplus 9\%$ [91] with a typical readout noise of 200 MeV per tower. The particle flow algorithm is further used to build a global representation of the event based on the detector signals from other subsystems [92], as described in more detail in section 2.2.7.

2.2.5 The muon systems

Muon detection is of central importance to CMS, as muons can be detected relatively easily and are produced in several interesting decays, such as $\mathrm{H} \to \mathrm{ZZ}^* \to 4\mu$. The muon systems at CMS are used to identify muons over a wide angular range up to $|\eta| < 2.4$, to measure their charge and momentum and for triggering purposes. The punch-through of hadronic particles to the muon systems is negligible ($\simeq 0.2\%$) due to the large amount of material preceding the muon systems ($\simeq 16$ radiation lengths). The muon reconstruction efficiency can be factorised to the tracking efficiency $\epsilon_{\mathrm{track}} \simeq 0.99$ [93], reconstruction and identification efficiency $\epsilon_{\mathrm{rec+id}} \simeq 0.96 - 0.98$, isolation efficiency ϵ_{iso} and trigger efficiency ϵ_{trig} [94]. The overall muon efficiency is around 95-99% and the p_T-resolution between 15% in the barrel and 25% in the endcap at the trigger level. Since the muon systems cover a large area around the detector in the form of a barrel and endcaps, they have to be inexpensive and robust.

In the barrel region, drift tubes are organised into four stations arranged in concentric cylinders within the return yoke. The drift tubes use a mixture of $\mathrm{Ar/CO_2}$ gas and each chamber consists of four layers arranged into a superlayer. This provides a timing resolution of a few nanoseconds and allows the muon system readout to be assigned to a bunch crossing [79]. The spatial resolution of the drift chambers has been measured in test beams to be around 300 μm, determined by the dispersion of the drift time and distortions of the drift caused by magnetic field [95]. The bunch crossing identification efficiency, which is important for triggering, is better than 90%, driven by the timing resolution and muons producing electromagnetic showers [79].

In the endcaps, the muon system consists of cathode strip chambers (CSC), which have the advantage that they can operate at the high rates and at the non-uniform magnetic field present in the forward region. The spatial resolution of a hit is around ~ 80 μm in the combined six-plane CSC chamber and the bunch tagging efficiency is around 98-99% [79].

In order to complement the time resolution of the drift tubes and the CSC, a trigger system based on resistive plate chambers (RPC) exists in the barrel and endcaps. The RPC operates on the principle of an avalanche generated in a gas gap between two resin plates, such that the bunch crossing assignment is possible with rates up to 1 kHz/cm^2. The timing resolution of a few nanoseconds provided by the RPC improves the trigger efficiency [79].

Due to manufacturing tolerances, the intense magnetic field and thermal stress can cause the geometry of the muon system to change at the level of up to a few centimetres, which is monitored using an optical alignment system [96].

2.2.6 Trigger, data acquisition and computing

The data from the 40 MHz LHC collisions needs to be reduced by a factor of 10^6 for storage and analysis. This means that highly efficient trigger logic needs to be applied in order to select the most interesting physics events containing high-energy particles. The CMS experiment employs a two-level trigger system, where the Level-1 (L1) trigger is implemented in custom programmable electronics and operates on the level of the calorimeters and the muon systems, whereas the high-level trigger (HLT) has access to the complete event readout and is implemented on a conventional CPU farm.

The L1 trigger is composed of the trigger primitive generators, which operate on the level of calorimeter trigger towers, track segments and hit patterns on muon chambers. This information is combined using regional triggers to determine trigger objects in limited spatial regions of the detector. These objects are then compared by the global calorimeter and muon triggers, which determine if sufficient good-quality muon or calorimeter objects are present to accept the event. The processing is pipelined such that the deadtime is minimised [79].

The global calorimeter trigger works on the basis of calorimeter jets, total transverse energy, missing transverse energy and H_T, the scalar transverse energy sum of calorimeter jets. Furthermore, it provides isolated and non-isolated e/γ candidates. The muon system trigger works on the basis of reconstructing track candidates from hits in the drift tubes, the CSCs and the RPCs. In the global muon trigger, the muon candidates are identified by p_T, charge, η, ϕ and quality parameters, as well as isolation information from the global calorimeter trigger primitive [79].

In case the event passes the L1 trigger, the full detector is read out. This produces data with a maximal output rate of 100 kHz, further fed into the HLT, which reduces the recorded events to a rate of about 1000 Hz. The data are divided to luminosity sections of 2^{20} LHC orbits (93 s), during which trigger thresholds and trigger prescale factors, which sample the trigger acceptance, are not changed. The total amount of zero-suppressed data recorded for a bunch crossing is on the order of 1 MB [79].

If an event is accepted by the HLT, it is transferred to the CMS offline computing infrastructure, which consists of computer farms linked with high-speed networks, with the bulk of the computing resource located in computing centres around the globe. The computing resources are divided into Tiers according to function and expected reliability. The Tier 0 centre at CERN performs the immediate reconstruction of the data and transfers it to several Tier 1 centres for storage, where late-stage reconstruction with improved calibrations can take place. Data analysis and MC simulation happens primarily at Tier 2 centres, which are associated to Tier 1 sites and divide the resources between the CMS collaboration and the local physics community. A typical Tier 2 site hosts 1-2 PB of data and O(5000) CPUs [79].

2.2.7 Particle flow reconstruction

In case a collision event passes the HLT, it is recorded as signals from different subsystems of the detector, such as energies deposited in the calorimeter cells and particle hits in the muon system. At CMS, the reconstruction algorithm creates physics objects such as jets, leptons and missing transverse energy by linking the signals across different sub-detectors using the particle flow (PF) algorithm to arrive at a global event description. This algorithm relies on high-granularity sub-detectors, which allow signals from various subsystems to be correlated.

The PF algorithm associates charged tracks from the silicon tracker to calorimeter clusters based on geometrical proximity and proceeds by "subtracting" objects from the event in order of decreasing reconstruction accuracy, starting from muons, followed by electrons and isolated photons, such that neutral hadrons and non-isolated photons are built from calorimeter clusters that are not associated to any tracks. The PF algorithm improves the response and resolution of the detector, parametrised by the jet response $R = p_T / p_{T,\mathrm{ref}}$, the ratio between the transverse momentum of a reconstructed jet and the transverse momentum of the generator-level jet, and the jet resolution, defined as the Gaussian spread of the response distribution. The mean jet response is corrected from around 60% at $p_T \simeq 100$ GeV for calorimeter jets to around $R = 95\%$ for particle flow jets, additionally reducing the momentum dependence. The jet energy resolution is similarly improved, corrected from 15% for calorimeter jets to about 10% for particle-flow jets with $p_T \simeq 100$ GeV [92]. This allows to achieve an uncertainty in the jet energy scale of $< 3\%$ over the region covering most analyses ($p_T < 30$ GeV, $|\eta| < 5.0$) [97].

2.3 Summary

Overall, the LHC is uniquely suited to studying the Higgs boson and thus the mechanism of electroweak symmetry breaking in high-energy processes. The accelerator has been operating close to or above design capabilities with around $35 - 40$ fb^{-1} of integrated luminosity in proton-proton collisions delivered per year at the centre-of-mass energy of 13 TeV. The CMS experiment has excellent characteristics for detecting and studying the SM particles produced in the $\mathrm{t\bar{t}H}$ process and in the subsequent decay, thus making it possible to experimentally clarify the nature of the recently-discovered Higgs boson by determining the coupling of the scalar field to fermions. In the next chapter, we discuss how the tracking and vertexing capabilities of the CMS detector can be used to distinguish between jets arising from the hadronisation of bottom quarks and light quarks.

3 Identification of jets from bottom quarks

Accurately identifying jets from bottom quarks is a crucial part of the physics program involving top quarks, which decay before hadronisation to bottom quarks and W-bosons, and Higgs bosons, which decay primarily to bottom quarks. Since we cannot directly access the quantum numbers of a parton that gave rise to a jet, we instead measure a number of jet-related observables such as hadron lifetime and multiplicity of charged tracks that are correlated to the flavour of the underlying parton and thus allow us to obtain a statistical indication of the flavour.

The hadronisation properties of the bottom quark and the relatively large lifetime and possible semileptonic decays of b hadrons allow us to experimentally identify the jets that arise from bottom quarks through the technique of b tagging, or more generally identify the flavour of the jet through flavour tagging. A b hadron has a relatively long lifetime of 10^{-12} seconds and a large Lorentz boost factor, resulting in jet production with a secondary vertex that is displaced by several millimetres with respect to the primary interaction point.

The idea of b tagging is to use a combination of discriminating variables to distinguish between jets from bottom, charm and light quarks on a statistical basis by assigning a discriminator value for jets that is on average higher for jets arising from bottom quarks than for the ones arising from light quarks. Jets that are associated to discriminator values above certain pre-determined thresholds are taken to be tagged as a certain flavour. This technique based on the presence of a secondary vertex was used extensively at Tevatron in the analyses that led to the discovery of the top quark [98, 27].

In this chapter, we give an overview of b tagging at CMS, with a specific focus on the development and re-optimisation of the Combined Multivariate (cMVAv2) b tagging algorithm for Run 2 at the LHC. We start with an overview of the discriminating variables important for b tagging, followed by a description of the existing b tagging algorithms at CMS. We then introduce the cMVAv2 algorithm and describe the optimisation and performance of the retrained version for Run 2 analyses. We conclude with an outlook on b tagging.

3.1 Discriminating variables

The most important discriminating variables for flavour tagging are the kinematic properties of jets and the associated leptons, as well as the existence and properties of tracks from charged particles and vertices associated to either the primary hard interaction in the proton-proton collision or the secondary vertices from the decay of b or c hadrons. B hadrons in the ground state, where the bottom quark is accompanied by a light quark (down, up, strange), decay through weak interaction and the decay modes are well described by the decay of the bottom quark in $\mathrm{b} \to \mathrm{cW}^*$, followed by a leptonic or hadronic decay of the off-shell W-boson. The secondary vertex associated with the b hadron decay, which is displaced with respect to the primary vertex due to the long life time of b hadrons, is a salient feature of b hadron decays that the b tagging algorithms seek to exploit through accurate track and vertex reconstruction. Therefore, the selection of good tracks and the reconstruction of secondary vertices is necessary for efficient b tagging.

3.1.1 Track selection

The tracks reconstructed by the inner tracking detector and associated to jets are used for b tagging only in case they are of sufficient quality, i.e. they pass the track selection, detailed below. They must have a transverse momentum of at least 1 GeV, a high-quality track fit* and at least eight hits

*Normalised χ^2 quality parameter for the track fit to hits below 5.

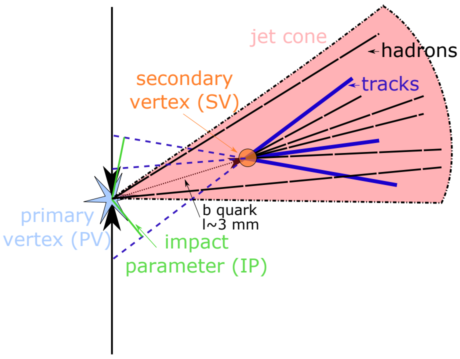


Figure 3.1: An illustration of the decay of a b quark, along with the definition of the secondary vertex and the impact parameter.

in the tracker with at least two in the inner pixel detector. Furthermore, the impact parameter (IP), defined as the distance of closest approach between the primary vertex and the track trajectory, shown on fig. 3.1, is required to be less than 0.2 (17) cm in the transverse (longitudinal) direction. This ensures that the tracks are sufficiently close to the primary vertex and thus reduces the contribution from pileup. The distance of closest approach between the jet axis and the track must be smaller than 0.07 cm, and the decay length, defined as the distance between the primary vertex and the point of closest approach between the track and the jet must be less than 5 cm. Jets with $p_T \simeq 100$ GeV are estimated to have around seven associated tracks, with the track multiplicity rising to about ten at $p_T \simeq 200$ GeV. [99].

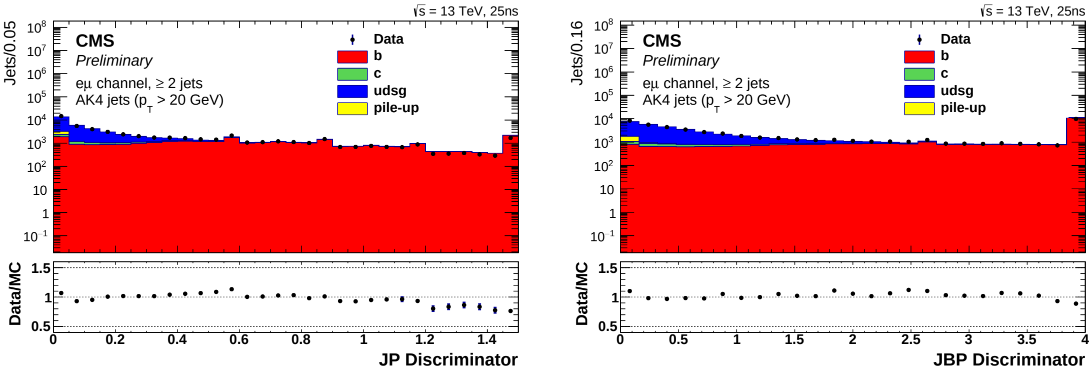
3.1.2 Vertex reconstruction

Tracks passing the pileup-suppressing selection are used for vertex reconstruction in the adaptive vertex reconstruction (AVR) algorithm [100]. Vertices reconstructed by AVR must pass further selection criteria designed to suppress vertices that are unlikely to originate from b hadron decay. These selection criteria require a vertex to have a sufficient number of tracks, a high-significance flight distance, a mass of less than 6.5 GeV and incompatible with the K^0_S hadron mass. Upon fulfilling these criteria, a jet is assigned to contain an AVR vertex [99].

An alternative to AVR, where tracks are required to be associated to jets and therefore vertex reconstruction is seeded by jets, is the inclusive vertex finder (IVF) algorithm [101]. As the name implies, the IVF algorithm starts with the set of all tracks in the event that pass looser selection criteria than those used for AVR. Vertices are reconstructed by fitting all tracks simultaneously using an adaptive fitting algorithm looking for clusters of tracks. A further arbitration step assigns tracks to either the primary or secondary vertex based on compatibility and pixel hits, or removes secondary vertices of low quality. The IVF algorithm reconstructs about 10% (15%) more often the vertices from bottom (charm) hadrons, but also increases the fraction of vertices reconstructed for light jets by about 8%. The overlap between the two algorithms is about 60%, meaning that the two fitters provide some independent information on the event [99].

3.2 Multivariate b taggers

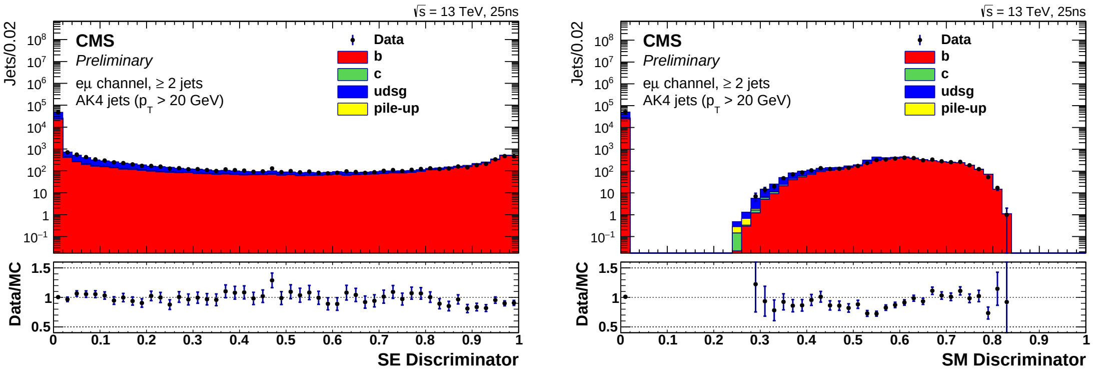
Using machine learning, signals from various regions of the detector can be combined effectively to develop a discriminator between jet that arise from the hadronisation of bottom quarks (b jets) and from uds quarks or gluons (light jets). Such techniques are especially suited to exploit sources of information that are partially correlated, such as the AVR and IVF vertices. The cMVAv2 b tagging algorithm combines the output of different low and high level b tagging algorithms, developed independently and using partially correlated variables, into a single high level discriminator. Before we can describe the cMVAv2 in detail, we must therefore discuss the different b tagging algorithms used at CMS.



(a) The JP b tagger

(b) The JBP b tagger

Figure 3.2: The jet probability b taggers in dileptonic (eμ) $\mathrm{t\bar{t}}$+jets events. Figures from [99].



(a) The soft electron b tagger

(b) The soft muon b tagger

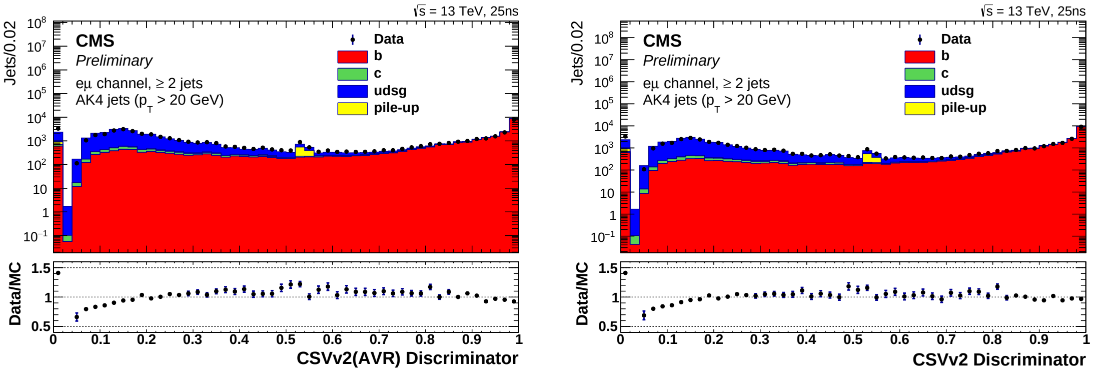
Figure 3.3: The discriminator distributions of the soft lepton b taggers in dileptonic (eμ) $\mathrm{t\bar{t}}$+jets events. The large fraction of jets with a soft muon discriminator value around zero arises from cases where no muon was associated to the jet. Figures from [99].

Jet probability taggers

The CMS Jet Probability (JP) tagger was developed during Run 1 of the LHC and is a simple multivariate likelihood discriminator based on track properties. In JP, the likelihood for a jet to originate from the primary vertex, as opposed to a secondary vertex, is computed by multiplying the per-track likelihoods, based on the track impact parameter and detector resolution. In a version of the JP tagger, the four tracks that have the highest IP significance ($\mathrm{IP}/\sigma_{\mathrm{IP}}$) are given a higher weight such that the new tagger, denoted Jet B-Probability (JBP) is more efficient in selecting b jets from light jets [102]. The discriminator distributions for the JP and JBP b taggers are shown in fig. 3.2.

Soft lepton taggers

The semileptonic decays of the b hadron to muons through $\mathrm{b} \to \mu^- \bar{\nu}_\mu \mathrm{c}$, which happen with a branching fraction of about $\simeq 11\%$ [103], make it possible to exploit the presence of a reconstructed muon in a jet for b tagging. The Soft Muon (SM) algorithm at CMS relies on the presence of a muon in the jet constituents, but not on the presence of a secondary vertex. Similarly, the decay of a b hadron to an electron is exploited through the Soft Electron (SE) tagger [99]. The discriminator distributions are shown in fig. 3.3.



(a) The CSVv2 AVR b tagger

(b) The CSVv2 IVF b tagger

Figure 3.4: The discriminator distributions of the CSVv2 b taggers in dileptonic (eμ) $t\bar{t}$+jets events. Figures from [99].

The CSVv2 b tagger

The Combined Secondary Vertex algorithm V2 (CSVv2) is based on the original CSV implementation introduced in Run 1 [102] and uses machine learning to combine track and secondary vertex information such as vertex mass or flight distance. Based on the presence and quality of the secondary vertex, the CSVv2 is optimised independently in several categories:

- presence of a good secondary vertex, in which case the flight distance and other vertex related variables are defined
- the pseudo-vertex category with two good tracks but no vertex fit, in which case the track parameters are used
- a no vertex category that uses information only from displaced tracks.

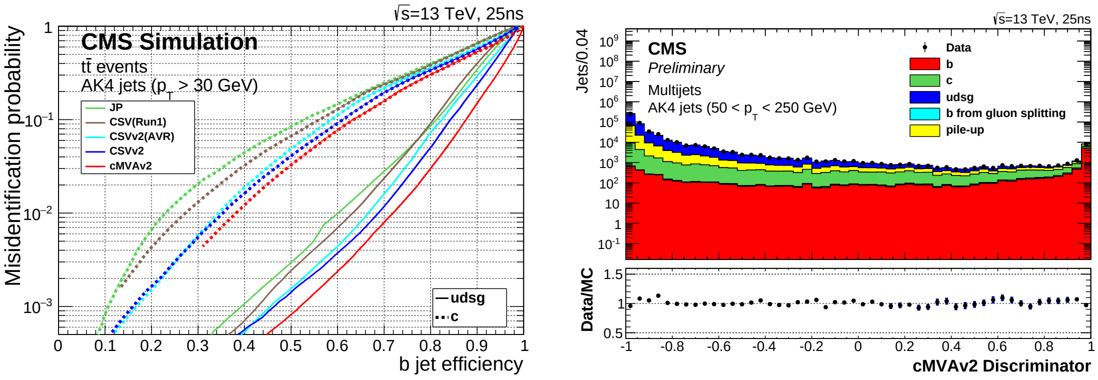
The final CSVv2 discriminant is a likelihood combination of binary classifiers based on independent artificial neural networks with a single hidden layer in each of the three categories. The CSVv2 algorithm has been deployed on both the AVR vertices as CSVv2 (AVR) and the IVF vertices, denoted CSVv2 (IVF). The CSVv2 (IVF) has an efficiency of about 66% for b jets at a mistag rate for light jets (udsg-associated) of about 1% based on $t\bar{t}$ simulation at the CSVv2 medium working point [99]. The CSVv2 b discriminator is the primary b tagging algorithm used in the end of Run 1 and the beginning of Run 2 at CMS. The discriminator distributions are shown in fig. 3.4.

3.3 The combined multivariate b tagger

For Run 2, we have developed an improved b tagger algorithm for CMS that combines the aforementioned individual b taggers, relying on various sources of information, into a single discriminator using boosted decision trees (BDTs) via the `scikit-learn` package [104]. This combined discriminator, denoted cMVAv2, can make use of both the AVR and IVF vertex reconstruction algorithms, the track-based jet probability taggers and the presence or lack of soft leptons and secondary vertex information simultaneously. The tagger is optimised on $t\bar{t}$+jets simulation, with cross-validation on a multi-jet simulation sample. At a similar b-jet efficiency to the CSVv2 medium working point ($\epsilon_b \simeq 70\%$), the cMVAv2 b tagger algorithm reduces the mistag rate for light jets from 1% to about 0.5%, as seen in fig. 3.5.

3.3.1 Implementation

The cMVAv2 b discriminator is optimised as a binary classifier in a supervised training mode on jets derived from $t\bar{t}$+jets simulation. Simulated jets are assigned a flavour using *ghost clustering*



(a) The b tagging and misclassification probability.

(b) The distribution of the cMVAv2 discriminator.

Figure 3.5: The performance of the CMS b taggers in Run 2. On (**a**), the efficiency to correctly tag b jets is shown with respect to the misclassification probability for light jets (udsg-associated, solid line) and charm jets (dashed line), based on $t\bar{t}$+jets simulation. The medium working point corresponds to a light jet efficiency of $\epsilon_l = 1\%$ and we see that the cMVAv2 b discriminator has $\epsilon_b \simeq 72\%$, compared to the CSVv2 b-efficiency $\epsilon_b \simeq 66\%$. On (**b**), we show the comparison of the cMVAv2 discriminator between data and simulation, where data-driven scale factors are applied. Figures from [99].

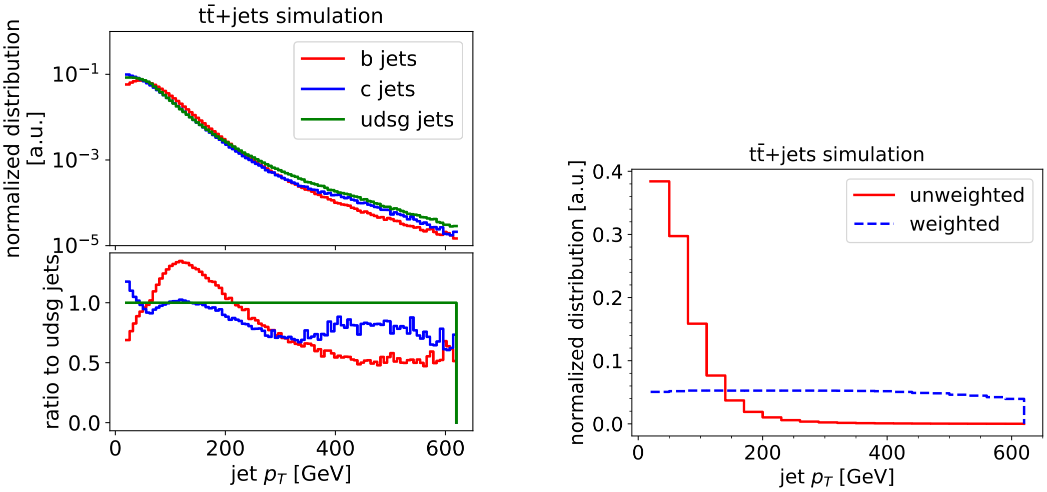
by re-clustering the stable jet constituents alongside generator-level hadrons. In order to prevent the generator-level hadrons from changing the momentum of the jet, the modulus of the hadron four momentum is set to a small number. Based on the clustered hadrons, a jet is assigned to be a b jet, a c jet or a light jet in decreasing priority [105]. We use particle flow jets with $p_T >$ 20 GeV, $|\eta| < 2.5$ and enhance the selected jet sample at high p_T and $|\eta|$ by limiting the amount of jets in the binned 2D distribution of $p_T \in [20, 620], |\eta| \in [0, 2.5]$ with 100 bins per dimension to $N_{\max} = 10000$ per bin. This sub-sampling technique effectively gives higher weight to the tails of the momentum and pseudorapidity distribution and reduces the simulated data to an amount that fits in the random access memory of a conventional personal computer. We use an inclusive $t\bar{t}$+jets sample generated with `POWHEG` and `Pythia 8` and an independent multi-jet sample generated with `Pythia 8` for cross-validation. We list the details of the simulated samples in table 3.1.

MC sample	events	b jets	c jets	light jets
$t\bar{t}$+jets	28M	9.5M	3.7M	11M
multi-jet (QCD)	29M	1.2M	2.3M	17M

Table 3.1: The simulated samples used for optimising and validating the cMVAv2 b tagger. The jets are filtered with a $p_T, |\eta|$-dependent selection that restricts the amount of jets in the high-statistics, low p_T and $|\eta|$ part of the distribution.

Kinematic reweighting

Since the kinematic distributions in p_T and $|\eta|$ of bottom, charm and light jets are different due to their different production modes, we need to ensure that the b tagger algorithm distinguishes between jets of different flavour based on quantities that are mostly independent of kinematics, such that features beyond jet kinematics are used to discriminate between jets of different flavour. In addition to the subsampling technique mentioned above, we reweight the jet sample used in the training of the binary classifier such that the distributions are roughly uniform in p_T and $|\eta|$ using a weight $w(p_T, |\eta|, \text{flavor})$ derived on simulation, as shown in fig. 3.6. This reweighting does not have a significant effect on the overall final b tagging performance of the discriminator, however, it helps to make the optimisation less sample-dependent.



(a) The jet p_T distributions by flavour.

(b) The reweighted jet p_T distribution.

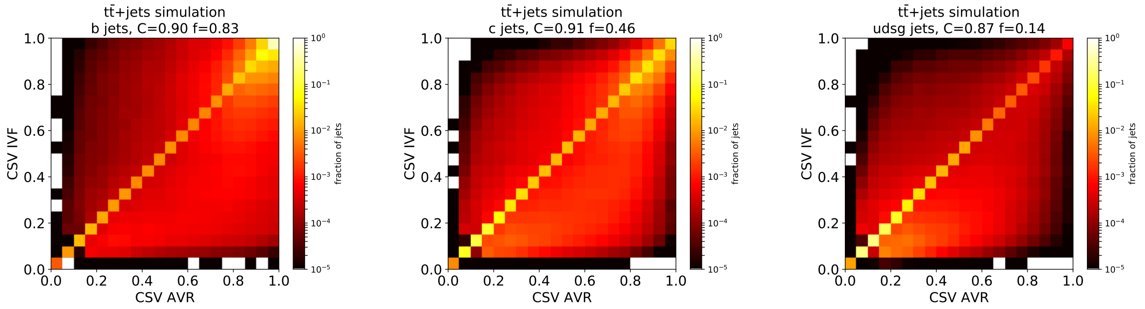
Figure 3.6: Comparison of jet kinematics by associated MC flavour before and after the kinematic reweighting. On (**a**), we compare the jet p_T distributions for bottom, charm and light jets. We see that b jets have a momentum distribution that is narrower and peaks around 100 GeV. On (**b**), we show the jet p_T distribution before and after applying the kinematic reweighting $w(p_T, |\eta|, \text{flavor})$. We see that the reweighted distribution is roughly uniform, with small differences caused by simulation statistics in the derivation of the weight. These distributions are derived using $\mathrm{t\bar{t}}$+jets simulation.

Input variables

We have used the b tagger algorithms described earlier as inputs to the cMVAv2 algorithm. In particular, we use the following six b discriminators: CSV (AVR), CSV (IVF), JP, JBP, SoftMuon and SoftElectron. We have studied the correlation between CSV (AVR) and CSV (IVF), as shown in fig. 3.7. Although the linear correlation coefficient is around $C \simeq 0.9$ for all flavours, we see a non-negligible contribution from jets which have different vertices from the two vertexing algorithms and thus different CSV discriminator values. Therefore, we expect that a combined discriminator will outperform either of the two. Furthermore, the information provided by the JP and JBP discriminators can complement the CSV in cases no vertices could be found. To verify whether adding the additional input b discriminators provides useful information for b tagging, we optimise a set of BDT classifiers iteratively by including the six input discriminators one by one.

3.3.2 Optimisation and validation

The BDT is optimised using gradient boosting [106] to distinguish between two classes: b flavoured and non-b (charm, udsg) flavoured jets, minimising the deviance between the true value $y_i = 1$ for b jets ($y_i = 0$ for non-b jets) and the predicted value $f(x_i)$, defined as $\sum_i \ln\left[1 + e^{-y_i f(x_i)}\right]$, based on the inputs x_i. We use 200 decision trees with a maximum depth of three, such that the contribution of each successive tree is reduced by a factor of 0.1, requiring at least 100 entries per split and per leaf in order not to be sensitive to statistical fluctuations in simulation. This specific choice of model parameters (hyper-parameters) guarantees a convergence in approximately 1.5 hours with a single CPU on the given dataset. We have carried out an optimisation of the parameters using a grid scan, however, we did not find statistically significant improvements. The optimisation and validation of the BDTs is carried out on statistically independent subsets of the $\mathrm{t\bar{t}}$ +jets simulation sample and the statistical over-training is negligible after 200 boosting iterations, as can be seen in fig. 3.8. This is due to the large amount ($\mathcal{O}(10^7)$ jets) of available simulation statistics, combined with the input variable space being relatively low-dimensional ($N = 6$). We have further validated the training using k-fold cross-validation, where the training is repeated k times on sub-samples of



(a) bottom jets **(b)** charm jets **(c)** light (udsg) jets

Figure 3.7: The correlation between the CSV b discriminator value using AVR and IVF vertices. Despite the high linear correlation coefficient C, we see a significant fraction f of jets with non-equal values of the two discriminators, visible as the non-diagonal elements in the 2D distribution. These distributions are derived using the tt̄+jets simulation.

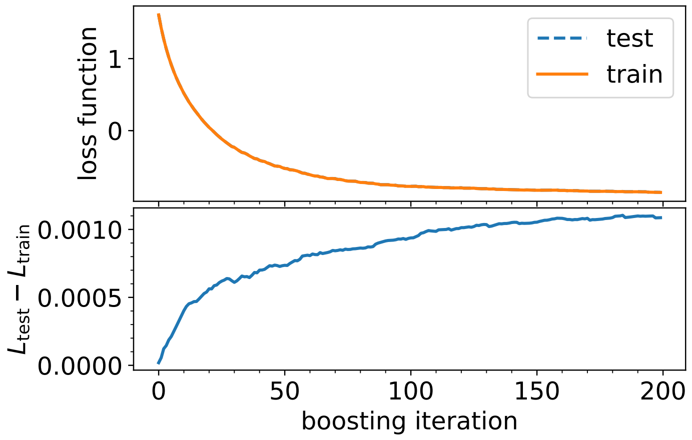
the data, and found the statistical uncertainty between sub-samples to be negligible.

We analyse the performance of the optimised b discriminator in terms of the efficiency for selecting b jets (signal) with respect to charm and light jets (background). On fig. 3.9, we show the receiver operating characteristic (ROC) area-under-curve (AUC) performance metric, which is lower for better discrimination, of the various stages of cMVAv2 optimisation. In general, we see that the BDTs are able to take advantage of all the input taggers, with the inclusion of each input improving the overall discrimination power. The optimisation is carried out on the tt̄ +jets simulation sample and we cross-checked the results on the multi-jet simulation, where we see an equivalent improvement.

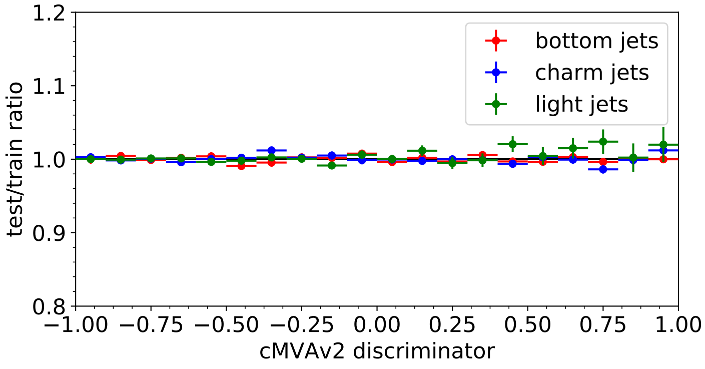
The performance of the optimised cMVAv2 b discriminator is validated by comparing the efficiency of selecting light jets or charm jets as b tagged at a fixed b tagging efficiency for b jets as a function of jet kinematics. We see that the cMVAv2 b discriminator has somewhat more stable performance with respect to jet kinematics than the CSV b discriminators and an improved discrimination power in the forward region $|\eta| > 1.5$. The cMVAv2 b discriminator uniformly outperforms the CSV algorithm over the studied jet momentum range $20 < p_T < 320$, $0 < |\eta| < 2.5$. Overall, the cMVAv2 algorithm has a mistagging efficiency lower by at least 20% (10%) for light jets (charm jets) over the full range, as shown in fig. 3.10.

The final validation and calibration of the cMVAv2 algorithm was carried out using the tag counting method, where the b tagging efficiency and the corresponding scale factor between data and simulation is determined from tt̄+jets dilepton events relying on the t → bW decay and thus the presence of two true b jets. The correction for the light jet mistagging efficiency is extracted from multi-jet events based on the discriminator distribution using jets with only negative impact parameters and thus enriched in light jets. The cMVAv2 discriminator corrections have independently been determined with the differential reweighting method described further in section 5.2.3 and the results are compatible with scale factors derived from tag counting [99].

The evaluation of the cMVAv2 b tagger has been implemented in the CMS software environment (`CMSSW`) independently of the BDT optimisation. As such, the cMVAv2 was one of the two default b discriminators of CMS during the 2016 data-taking period. In order to interface the decision trees optimised using `scikit-learn` to the CMS online software, we have developed a stand-alone library [107] to convert between BDT implementations from several commonly used open source machine learning toolkits, namely `scikit-learn` [104], `xgboost` [108] and `TMVA` [109]. The set of decision trees was then exported to the CMS software, such that the cMVAv2 b discriminator could be deployed through standard reconstruction software by simply evaluating the decision trees.

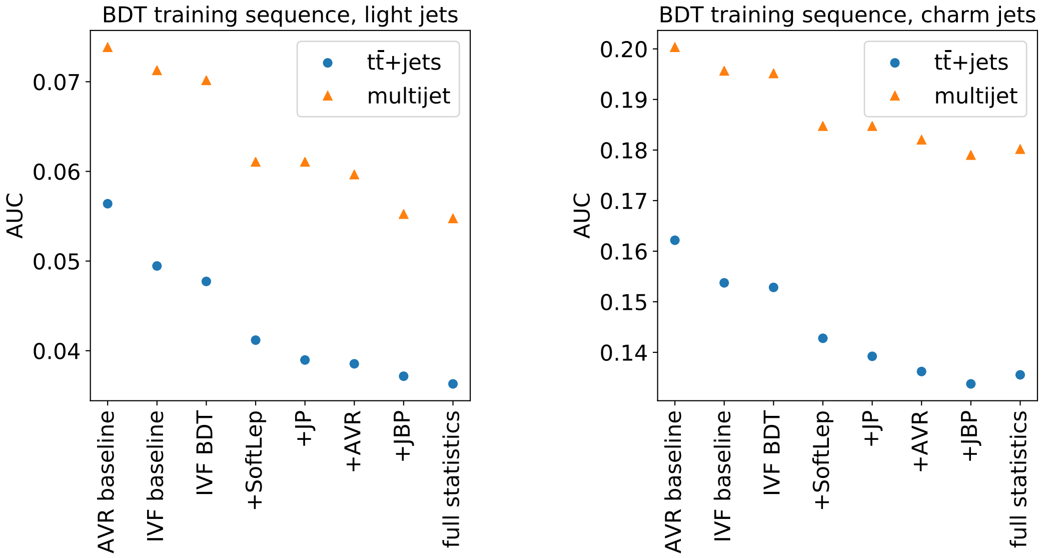


(a) The loss function (top) with respect to boosting iteration and the difference between training and validation loss (bottom).



(b) Ratio of discriminator distributions.

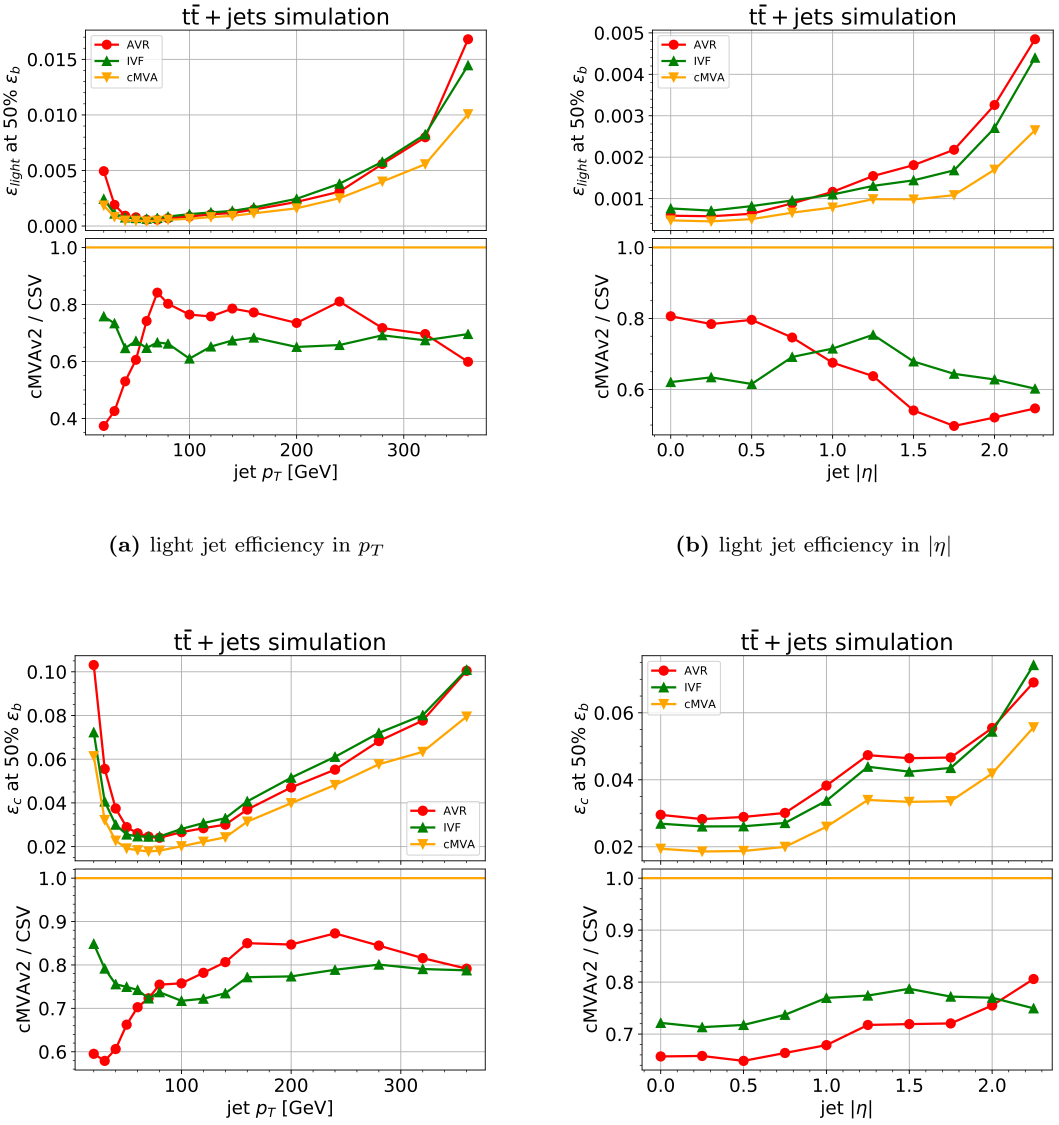
Figure 3.8: Validation of the cMVAv2 against over-training by comparing the loss function between the training set and the statistically independent validation set as a function of the boosting iteration on (**a**). We see good convergence after $N = 200$ boosting steps, with a negligible difference between the loss functions. Furthermore, on (**b**) we compare the discriminator distributions on the training and test sets, with the ratio being compatible with unity within statistical uncertainties.



(a) b vs. light jets

(b) b vs. charm jets

Figure 3.9: The performance characteristic ROC AUC as a function of the number of variables included in the BDT training, where lower values correspond to better discrimination. We use the ROC AUC as opposed to an efficiency measure at a specific b discriminator working point value, as this metric describes the performance over the full efficiency range inclusively. The baseline CSV b discriminators with IVF and AVR vertices are compared to the cMVAv2, which is optimised by successively including the IVF, SoftLepton, JP, AVR, JBP taggers with partial training statistics. We see that the inclusion of the Soft Muon and Soft Electron b taggers provides a significant improvement to the b jet vs. light jet discrimination (**a**), as well as b jet vs. charm jet (**b**). The best discrimination against light jets is reached by using the full simulation sample. We have performed the optimisation on the $t\bar{t}$ +jets simulation sample, and cross-checked the results on the multi-jet sample. Statistical uncertainties on the performance are verified using cross-validation and are negligible. We observe a slightly reduced discrimination for the full training sample compared to the limited sample in terms of charm jet rejection. This can be understood as an absolute shift in the performance characteristic caused by selecting an equal number (50 000) of light, charm and b jets for the limited-statistics training, whereas in the case of full statistics, the b : c : l ratios are determined from simulation. Thus, in the limited-statistics case, the cMVAv2 is slightly more optimal against charm jets than in the full statistics case. This is not expected to change the overall conclusions of the optimisation.



(a) light jet efficiency in p_T

(b) light jet efficiency in $|\eta|$

(c) charm jet efficiency in p_T

(d) charm jet efficiency in $|\eta|$

Figure 3.10: The p_T and $|\eta|$-dependent mistagging efficiency ϵ_{light}, ϵ_{c} at a fixed b-jet efficiency ϵ_b in tt̄ +jets simulation. We compare the cMVAv2 discriminator to the CSV IVF and AVR discriminators in terms of the light jet and charm jet efficiency (mistagging) at the working point with 50% b jet efficiency. We see that the cMVAv2 discriminator performs well over a wide p_T and $|\eta|$ range, with ϵ_{light} in the forward region being around 40% lower than for the CSV algorithm. While the efficiency of the discriminator overall still depends on jet kinematics due to the inherent correlation between the kinematics and the underlying variables sensitive to b tagging, the dependence is somewhat lower for cMVAv2 than for CSV.

3.4 Summary and outlook

Identifying jets that arise from the hadronisation of bottom quarks is crucial for analyses involving top quarks or Higgs bosons in the final state, such as the search for $\mathrm{t\bar{t}H}(\rightarrow \mathrm{b\bar{b}})$. We have developed an improved b tagging algorithm, the cMVAv2, for Run 2 which combines the discrimination power of several independently optimised b discriminators that use different vertexing algorithms and information from different subsystems of the detector. The cMVAv2 b discriminator outperforms the present state of the art at CMS by around 10-20% in terms of reduced rate of erroneously tagging jets from light or charm quarks. This method was deployed during the 2016 data-taking period as one of the standard b tagging algorithms at CMS. This algorithm has been used in several CMS analyses, such as in the analysis that reported evidence for the decay of the Higgs boson to bottom quark-antiquark pairs [110]. We demonstrated that using state of the art machine learning methods developed outside the experimental physics community is a viable way to make progress on object classification and identification tasks in high energy physics. However, extensive reliance on simulation in these b tagging algorithms means that a careful evaluation of the systematic uncertainties is necessary. Such multivariate b discriminators need to be calibrated with respect to data. With the increasingly large data samples available in Run 2, these calibrations can be carried out differentially with respect to jet kinematics, thereby increasing their accuracy. Advances in algorithm development have allowed the techniques of b tagging to be extended, such that it has recently become possible to separate charm jets from b jets and carry out b tagging in the regime with boosted jets [111].

4 Matrix Element Method

In the search for $t\bar{t}H(\rightarrow b\bar{b})$, the irreducible $t\bar{t}+b\bar{b}$ background presents an experimental challenge, as no single observable clearly distinguishes between the signal and background processes. Therefore, the information in multiple observables needs to be combined using multivariate analysis techniques (MVA), where the joint distribution of the observables is used to construct a classifier between the signal and background process hypotheses.

In the following chapter, we describe the implementation of a multivariate technique that is based on the direct computation of theory-motivated *ab initio* likelihoods for the observations using the underlying matrix element for the hard scattering process. The matrix element method (MEM) thus allows to connect the observable quantities at the detector level, such as the momenta of jets and charged leptons, to the dynamics of the scattering process. We show that the MEM provides a suitable framework for interpreting data from multi-parton final states and can be used to construct a discriminator between irreducible processes that is both theoretically motivated and practical for analysis.

We motivate the MEM approach from a statistical point of view and describe the implementation of the MEM likelihood. Furthermore, we discuss in detail the improvements made to the MEM as applied to the $t\bar{t}H(\rightarrow b\bar{b})$ analysis in Run 2, where we have extended the method to incorporate effects from mis-reconstructed jets and additional QCD radiation. This considerably extended the phase space which can be analysed using the MEM. Finally, we study the expected performance of the MEM in simulation and propose an analysis strategy for the $t\bar{t}H(\rightarrow b\bar{b})$ search based on the method.

4.1 Hypothesis testing

We can formulate the problem of deciding whether an event arose from a signal or background process as performing a hypothesis test. We distinguish between the background hypothesis (H_0) and the signal hypothesis (H_1) on an event-by-event basis based on the event-level observables $\boldsymbol{y}$, the momenta of the final state particles. We classify an event as signal by rejecting the hypothesis H_0 in favour of H_1. In the framework of statistical hypothesis testing, we define the test statistic $\lambda(\boldsymbol{y})$ as a scalar function of the observables, such that the sampling distributions under the hypotheses can be estimated using MC simulation. Based on the observed value of $\lambda(\boldsymbol{y})$ for an event, we decide if it should be interpreted as signal or background. In practice, we use the continuous value of the test statistic in a template fit, as described further in section 5.3.9. The choice of the test statistic is determined by the desired size of the test α, i.e. the probability of falsely rejecting H_0 if it is true (Type I error), and the power of the test $1-\beta$, where β is the probability of accepting the null hypothesis in case it is false (Type II error).

In the case of two simple hypotheses that do not depend on additional parameters, the Neyman-Pearson lemma [112] states that the likelihood ratio

$$\lambda(\boldsymbol{y}) = \frac{L(\boldsymbol{y}|\mathrm{H}_1)}{L(\boldsymbol{y}|\mathrm{H}_0)} \tag{4.1}$$

is the most powerful test statistic, meaning that the probability of Type II errors is minimised. Therefore, the task of computing the likelihood $L(\boldsymbol{y}|\mathrm{H}_i)$ of an observation characterised by $\boldsymbol{y}$ under a given hypothesis is central to a simple hypothesis test. Naively, this could be achieved using MC simulation to estimate the multidimensional probability density $f(\boldsymbol{y}|\mathrm{H}_i)$ and using it as the likelihood, however, this quickly becomes intractable as the dimensionality of $\boldsymbol{y}$ increases. Fortunately, the underlying theory of particle physics provides a way to compute probability densities in the form of differential cross sections for processes involving scattering and interactions. In the next section, we discuss how we can construct a useful likelihood function $L(\boldsymbol{y}|\mathrm{H})$ using the theoretical framework provided by QFT.

4.2 Description of MEM

The use of MVA techniques has a long history in HEP data analysis, starting with multivariate discriminators in the analysis of bubble chamber data [113] and investigations of artificial neural networks (ANN) for tracking [114]. Classical likelihood methods are based on likelihood ratios constructed from 1-dimensional distributions estimated from simulation, which do not properly account for the correlations between variables, or machine learning (ML) techniques such as BDTs or ANNs, as introduced for b tagging in chapter 3. An advantage of the ML based methods is that they can faithfully approximate complex functions and take into account correlations between the input variables, however, they suffer from several important drawbacks:

- in HEP, such ML methods are typically optimised completely on simulation, making the prediction susceptible to over-fitting[*] in case the simulation does not represent data accurately,
- they require extensive MC simulation to be optimised effectively, such that $\mathcal{O}(10^8)$ MC events are routinely simulated with full detector simulation to have sufficient statistics in high jet multiplicity categories,
- the predictions are often opaque, combining dozens of variables in complex ways, such that it is not generally possible to understand why an event was classified as signal or background.

Therefore, we seek an alternative that would be less susceptible to the aforementioned issues. The MEM belongs to a class of MVA methods in which we directly compute the per-event probability density that depends on model parameters $\boldsymbol{\theta}$ from first principles via

$$P_{\boldsymbol{\theta}}(\boldsymbol{y}) = \frac{1}{\sigma}\frac{\mathrm{d}\sigma_{\boldsymbol{\theta}}(\boldsymbol{y})}{\mathrm{d}\boldsymbol{y}} \tag{4.2}$$

and use it in hypothesis testing as component of the test statistic. Through the differential cross section $\mathrm{d}\sigma_{\boldsymbol{\theta}}$, the MEM explicitly depends on $\boldsymbol{y} = (\tilde{p}_{i\in\mathrm{jets}}, \tilde{p}_{i\in\mathrm{leptons}}, \dots)$, the vector containing the detector-level 4-momenta of the reconstructed particles, in particular, jets and leptons. The differential cross-section of eq. (4.3) depends on the squared matrix element $|\mathcal{M}_{\boldsymbol{\theta}}|^2$ that we use to take into account the dynamics of the relevant underlying processes. The kinematics of the $2 \to n$ scattering process are encoded in the n-body phase space element (eq. (4.4)), where the delta function enforces the conservation of momentum between the initial and final state particles.

$$\mathrm{d}\sigma_{\boldsymbol{\theta}} = \frac{|\mathcal{M}_{\boldsymbol{\theta}}(q_1, q_2; p_1, \dots, p_n)|^2}{4\sqrt{q_1 \cdot q_2 - m_1^2 m_2^2}} \times \mathrm{d}\Phi_n(q_1 + q_2; p_1, \dots, p_n) \tag{4.3}$$

$$\mathrm{d}\Phi_n = (2\pi)^4 \delta^4(q_1 + q_2 - \sum_{i=1}^{n} p_i) \prod_{i=1}^{n} \frac{\mathrm{d}^3 p_i}{(2\pi)^3 2E_i} \tag{4.4}$$

We have to consider several additional effects to be able to compute the MEM probability $P_{\boldsymbol{\theta}}(\boldsymbol{y})$:

- we cannot directly observe the momenta of the initial state partons q_1 and q_2 in a proton-proton collision,
- we do not measure the momenta of the neutrinos or jets that do not pass a transverse momentum threshold,
- the detector has a finite energy resolution,
- we do not know which of the observed jets are matched to which final state partons,

[*] to be distinguished from *over-training* in the sense of being sensitive to statistical fluctuations on a simulation sample

- the presence of non-zero final transverse momentum caused by large-angle initial state radiation, spoiling the momentum balance in eq. (4.4).

To address the first issue, we need to use parton density functions $g(x_{1,2})$ to weight the differential cross-section, integrating over the momentum fractions $x_{1,2} = E_{q_{1,2}}/E_{\text{beam}}$. In order to take into account detector effects, we integrate over unmeasured or poorly measured quantities using a transfer function $W(\boldsymbol{y}, \boldsymbol{p})$. The transfer function relates final state parton-level quantities $\boldsymbol{p}$ to measurable quantities on the detector level $\boldsymbol{y}$ and encodes our knowledge about detector resolution and reconstruction efficiencies.

Jet-to-parton matching is addressed by summing over the N_a possible combinations of jet-to-parton matching, which depends on assumptions about the process and the number of observed final state particles and is encoded in the exact form of the transfer function $W(\boldsymbol{y}, \boldsymbol{p})$, further described in section 4.3.1.

The modelling of the non-zero transverse recoil $\boldsymbol{\rho}_T = -\sum_{i=1}^{n} \boldsymbol{p}_{i,T}$ is treated empirically using a transfer function $\mathcal{R}(\tilde{\boldsymbol{\rho}}_T, \boldsymbol{\rho}_T)$ determined on simulation that relates the parton-level transverse momentum of the system $\boldsymbol{\rho}_T$ to the observed recoil $\tilde{\boldsymbol{\rho}}_T$.

The evaluation of the scattering amplitude $|\mathcal{M}_{\boldsymbol{\theta}}(\boldsymbol{p})|^2$ requires full knowledge of the initial and final state momenta $\boldsymbol{p}$, as well as the parameters of the model, summarised in $\boldsymbol{\theta}$. In particular, the parameters of the model consist of the hypothesis $\mathcal{H} \in \{\mathrm{t\bar{t}H}, \mathrm{t\bar{t}} + \mathrm{b\bar{b}}\}$ about the underlying process and the assumptions about which of the partons formed reconstructed jets $\mathcal{C}$, such that $\boldsymbol{\theta} = (\ldots, \mathcal{H}, \mathcal{C})$. In the case of $\mathrm{t\bar{t}H}$ with the top quark pair decaying semileptonically (SL) as $\mathrm{t\bar{t}H} \to (\ell^- \bar{\nu}_\ell \mathrm{b})\ (\mathrm{qq'\bar{b}})\ (\mathrm{b\bar{b}})$, we may consider the fully reconstructed category where all six of the final state partons are reconstructed as jets, denoted as $2_W 2_H 2_t$, the case where one of the quarks from the hadronic decay $\mathrm{W} \to \mathrm{qq'}$ is out of acceptance, denoted as $1_W 2_H 2_t$ and so forth, such that $\mathcal{C}_{\text{SL}} \in \{2_W 2_H 2_t, 1_W 2_H 2_t, \ldots\}$. The number of unknown quantities to be integrated over depends on the reconstruction category $\mathcal{C}$, as described in detail in the next section. Thus, the per-event probability density takes the form

$$P(\boldsymbol{y}, \boldsymbol{\theta}) = \sum_{k=1}^{N_a} \int \frac{\mathrm{d}x_1 \mathrm{d}x_2}{2 x_1 x_2 s} \int \prod_{i=1}^{n} \frac{\mathrm{d}^3 p_i}{(2\pi)^3 2 E_i} \tag{4.5}$$

$$\times\, \delta^4(q_1 + q_2 - \sum_{i=1}^{n} p_i) \tag{4.6}$$

$$\times\, g(x_1) g(x_2) \tag{4.7}$$

$$\times\, \mathcal{R}(\tilde{\boldsymbol{\rho}}_T, \boldsymbol{\rho}_T) \tag{4.8}$$

$$\times\, |\mathcal{M}_{\boldsymbol{\theta}}(q_1, q_2, p_1, \ldots, p_n)|^2 \tag{4.9}$$

$$\times\, W(\boldsymbol{y}, \boldsymbol{p}). \tag{4.10}$$

After having been first proposed for reconstructing events with missing momentum [115], the MEM has been used in Tevatron for Higgs boson searches [116, 117], in a precise measurement of the top quark mass [118] and in the analysis that saw evidence for single top quarks [119] at DØ. After first phenomenological studies showed that MEM could be used effectively for $\mathrm{t\bar{t}H}$ in a multi-parton final state [120], it has been used in searches for $\mathrm{t\bar{t}H}(\to \mathrm{b\bar{b}})$ by the CMS and ATLAS experiments at the LHC [121, 122]. The MEM approach is closely related to the matrix element likelihood approach (MELA) [123] that has been extensively used in $\mathrm{H} \to \mathrm{ZZ} \to 4\ell$ searches and J^P measurements, however, in MELA, unreconstructed particles and transfer functions are not considered and the matrix elements generally have simple analytical forms. In the following sections, we discuss in detail the implementation and improvements that were made to the MEM in the search for $\mathrm{t\bar{t}H}(\to \mathrm{b\bar{b}})$ during Run 2.

4.3 Implementation

In the case of the semileptonic or dileptonic $\mathrm{t\bar{t}H}$ final state, the observables $\boldsymbol{y}$ consist of the energies (or equivalently transverse momenta) and the directions of the jets, the momenta of the charged leptons and the measured recoil of the system $\tilde{\boldsymbol{\rho}}_T$. As eq. (4.5) needs to be integrated numerically, we first need to define the phase space volume element explicitly in terms of variables that are convenient and suitable for integration, namely energies, solid angles and combinations of invariant masses. The Jacobian transformations can be carried out analytically for all particles and are of the form shown in Equations (4.11) to (4.15) for the $\mathrm{H} \to \mathrm{b\bar{b}}$ decay, the unassociated $\mathrm{b\bar{b}}$ from $\mathrm{t\bar{t}+b\bar{b}}$ and the top decay with $\mathrm{qq'}$ corresponding to the quarks or leptons from the W boson decay:

$$\mathrm{H} \to \mathrm{b\bar{b}}: \prod_{i=1}^{2} \frac{\mathrm{d}^3 p_i}{(2\pi)^3 2E_i} = \left(\frac{1}{16\pi^3}\right)^2 \times \frac{p_b p_{\bar{b}}}{8|E_b - \boldsymbol{p}_b \cdot \frac{\boldsymbol{e}_{\bar{b}}}{\beta_{\bar{b}}}|} \times \tag{4.11}$$

$$\times\, \mathrm{d}E_b\ \mathrm{d}m^2_{b\bar{b}}\ \mathrm{d}\Omega_b\ \mathrm{d}\Omega_{\bar{b}} \tag{4.12}$$

$$\mathrm{b\bar{b}}: \prod_{i=1}^{2} \frac{\mathrm{d}^3 p_i}{(2\pi)^3 2E_i} = \left(\frac{1}{16\pi^3}\right)^2 \times \frac{p_b p_{\bar{b}}}{4}\ \mathrm{d}E_b\ \mathrm{d}E_{\bar{b}}\ \mathrm{d}\Omega_b\ \mathrm{d}\Omega_{\bar{b}} \tag{4.13}$$

$$\mathrm{t} \to \mathrm{bqq'}: \prod_{i=1}^{3} \frac{\mathrm{d}^3 p_i}{(2\pi)^3 2E_i} = \left(\frac{1}{16\pi^3}\right)^3 \times \frac{p_q p_{q'} p_b E_{q'}}{16 m^2_{qq'} |(E_q + E_{q'}) - (\boldsymbol{p}_q + \boldsymbol{p}_{q'}) \cdot \frac{\boldsymbol{e}_b}{\beta_b}|} \times \tag{4.14}$$

$$\times\, \mathrm{d}E_q\ \mathrm{d}m^2_{qq'}\ \mathrm{d}m^2_{qq'b}\ \mathrm{d}\Omega_q\ \mathrm{d}\Omega_{q'}\ \mathrm{d}\Omega_b, \tag{4.15}$$

where we have expressed the Lorentz-invariant phase space in terms of particle energies, solid angles ($\mathrm{d}\Omega = \sin\theta\ \mathrm{d}\theta \mathrm{d}\phi$) and invariant masses $m_{qq'} = (q+q')^2$, $m_{qq'b} = (q+q'+b)^2$. The scattering amplitude written as $|\mathcal{M}_{\boldsymbol{\theta}}(\boldsymbol{p})|^2$ depends only on particle momenta, hence it is implied to be summed over spin and colour states. Furthermore, we treat the production and decay of the top quarks, W and Higgs bosons in the narrow-width approximation (NWA), meaning we factorise the production and decay of these particles, as seen in eq. (4.16).

$$|\mathcal{M}_{\mathrm{t\bar{t}H} \to (qq'b)(qq'\bar{b})(b\bar{b})}|^2 = |\mathcal{M}_{gg \to\ \mathrm{t\bar{t}H}}(g_1, g_2, p_t, p_{\bar{t}}, p_h)|^2 \tag{4.16}$$

$$\times \prod_{r=t,\bar{t}} [\frac{\pi}{m_t \Gamma_t} \delta(p_t^2 - m_t^2) |\mathcal{M}_t(p_q, p_{\bar{q}}, p_b)|^2]_r \tag{4.17}$$

$$\times \frac{\pi}{m_H \Gamma_H} \delta(p_h^2 - m_H^2) |\mathcal{M}_H(p_b, p_{\bar{b}})|^2 \tag{4.18}$$

The decay amplitude of the top quark assuming the NWA for the W-boson is

$$|\mathcal{M}_t(p_q, p_{q'}, p_b)|^2 = \frac{64\pi m_t^2 g^4}{m_W \Gamma_W} \frac{(p_q \cdot p_t)}{1} (1 - \frac{m_b^2}{m_t^2} - \frac{2 p_q \cdot p_t}{m_t^2}) \times \delta((p_q + p_{q'})^2 - m_W^2) \tag{4.19}$$

and for the $\mathrm{H} \to \mathrm{b\bar{b}}$ decay

$$|\mathcal{M}_H(p_b, p_{\bar{b}})|^2 = 2\sqrt{2} m_b^2 m_H^2 (1 - \frac{4 m_b^2}{m_H^2}). \tag{4.20}$$

4.3.1 Transfer functions

The transfer function $W(\boldsymbol{y}|\boldsymbol{p})$ maps a point $\boldsymbol{p} \in \Omega$ in the phase space of the hard scattering process to a point $\boldsymbol{y} \in \mathcal{A}$ in the space of detector-level reconstructed variables and it is ensured to be normalised to unity using eq. (4.21),

$$\int_{\mathcal{A}} \mathrm{d}\boldsymbol{y}\ W(\boldsymbol{y}|\boldsymbol{p}) = 1, \tag{4.21}$$

which means that in an observable fiducial region $\mathcal{A}^* \subset \mathcal{A}$, a phase space point has an efficiency $\epsilon(\boldsymbol{p}) \leq 1$ to be reconstructed. We note that since the same transfer functions are applied for all hypotheses, the choice of transfer functions ultimately affects the model sensitivity, but not the correctness, that is, we are not introducing any bias into the analysis with assumptions on the transfer functions. This would not be the case if we were determining a parameter such as the Higgs boson mass by maximising the MEM probability of eq. (4.5). Furthermore, the MEM is validated on MC simulation in order to make sure that the assumptions do not degrade the sensitivity significantly.

We can make further progress by factorising the transfer function in terms of individual reconstructed objects, the jets and leptons:

$$W(\boldsymbol{y}|\boldsymbol{p}) = \prod_{i\in \text{jets}} W(E_i, \boldsymbol{e}_i|E_{q_i}, \boldsymbol{e}_{q_i}) \prod_{i\in \ell^\pm} W(E_i, \boldsymbol{e}_i|E_{\ell_i}, \boldsymbol{e}_{\ell_i}) = \prod_{i\in \text{jets}} W_{i,j} \prod_{i\in \ell^\pm} W_{i,\ell}, \tag{4.22}$$

where q_i (ℓ_i) is shorthand for the quark (charged lepton) assumed to give rise to the i-th measured jet (charged lepton). If we are dealing with indistinguishable objects such as jets, for which it is generally not possible to determine the charge or flavour of the originating parton, we have to sum over all possible ways of matching the detector-level and parton-level objects, such that

$$W_{i,j} \propto \sum_{q_i\in \text{quarks}} W(E_i, \boldsymbol{e}_i|E_{q_i}, \boldsymbol{e}_{q_i}). \tag{4.23}$$

Although the techniques of b tagging can be extended to determine if a jet arose from a b or anti-b quark based on a statistical combination of the charge properties of the secondary vertex, the purity of the charge determination is still relatively low, at a level of b quark efficiency $\epsilon_{\text{b}} \simeq 20\%$ for anti-b efficiency of $\epsilon_{\bar{\text{b}}} \simeq 50\%$, such that in an event with multiple jets arising from b quarks, the probability of mis-reconstructing the charge of a single b jet is significant [124].

Therefore we make no assumption on jet charge to assign out of four reconstructed b-jets two to the quarks from H → b$\bar{\text{b}}$, we have $4!/2!2! = 6$ combinations. Without distinguishing between b and $\bar{\text{b}}$, we have a further $2!/1!1! = 2$ ways of assigning the b tagged jets to bottom quarks from the top or antitop quarks, such that assigning the fully reconstructed $\text{t}\bar{\text{t}}\text{H}(\rightarrow \text{b}\bar{\text{b}})$ hypothesis with four b quarks and two light quarks among six jets amounts to twelve combinations. We describe the approach taken to reducing the number of required combinations further in section 4.3.4.

Lepton energies and directions are measured with an order of magnitude higher resolution than jet energies [92], therefore we assume those to be perfectly reconstructed such that their transfer functions are Dirac delta functions. Furthermore, we assume that the jet energy and angular transfer functions can be factorised such that

$$W(E_i, \boldsymbol{e}_i|E_{q_i}, \boldsymbol{e}_{q_i}) = W(E_i|E_{q_i}) \times W(\boldsymbol{e}_i|\boldsymbol{e}_{q_i}). \tag{4.24}$$

With these assumptions, the transfer function for observed particles reduces to a product of $W(E_i|E_{q_i})$ over the jets.

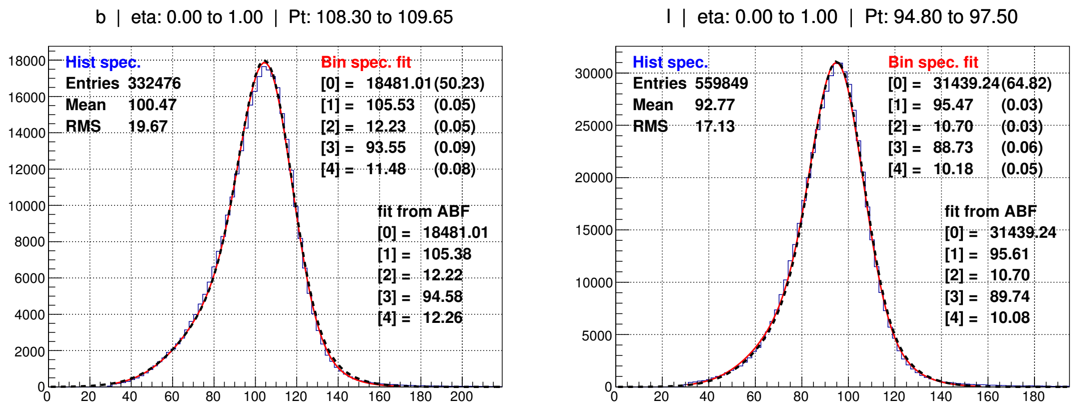
We use a double Gaussian function, shown in eq. (4.25),

$$W(p_{T,j}|p_{T,\text{gen}}) = N\left[0.7\exp\left(\frac{p_{T,\text{gen}} - p_{T,j} - \alpha_1}{\alpha_2}\right)^2 + 0.3\exp\left(\frac{p_{T,\text{gen}} - p_{T,j} - \alpha_3}{\alpha_2 + \alpha_4}\right)^2\right], \tag{4.25}$$

to model the parton-to-jet transfer function for both light jets and b jets. This functional form arises from a Gaussian detector response, which forms the core, and the modelling of parton fragmentation outside the jet cone, described by a non-Gaussian tail.

The parameters of these transfer functions are derived from MC simulation, assuming that they depend on quark energy, direction and flavour. We extract the transfer functions from a $\text{t}\bar{\text{t}}$ MC sample by matching the generator-level partons geometrically to jets using $\Delta R(q, j) < 0.3$ and fitting the functional form $W(p_{T,j}|p_{T,q})$ of the distributions in eq. (4.25). We fit the parameters $\alpha_1 \ldots \alpha_4$ of eq. (4.25) in bins of $p_{T,q}$, $|\eta_q|$ and flavour $\text{f} \in \text{l}, \text{b}$, shown in fig. 4.1. Additionally, in order to have a smooth dependence of the transfer functions on the generated transverse momentum $p_{T,\text{gen}}$, we fit polynominals to the per-bin parameters $\alpha_n(p_{T,\text{gen}}|\eta_q, \text{f})$, shown in fig. 4.2 for b jets in the central detector region $0 < |\eta| \leq 1.0$. Thus, we are able to evaluate the probability density for a quark with $p_{T,\text{gen}}, \eta$ to hadronise into a jet continuously over the full range of $p_{T,\text{gen}}$.

As we are considering both semileptonic and dileptonic $\text{t}\bar{\text{t}}\text{H}$ decays that contain neutrinos in the final state, the reconstruction resolution of MET has to be taken into account via a transfer



(a) Fit for b jets.

(b) Fit for light jets.

Figure 4.1: The double Gaussian (eq. (4.25)) fit (red) to the simulated transfer function (light blue) for b jets (**a**) and light jets (**b**) in the central detector region ($0 < |\eta| \leq 1$) in a $p_{T,\mathrm{gen}} \simeq 100$ GeV region. In general, we see that the double Gaussian function is sufficiently flexible to describe the transfer functions for both light jets and b jets. Furthermore, the interpolated double Gaussian using the fitted parameters $\alpha_1 \ldots \alpha_4$ (black) reproduces the exact double Gaussian fit reasonably well. The plots are derived using $\mathrm{t\bar{t}}$+jets simulation.

function. We model it via a Gaussian with a resolution $\sigma^2_{MET} = 30$ GeV, which is of similar magnitude to detector resolution and does not affect the results significantly. It is possible to evaluate the covariance matrix of MET on an event-by-event basis and thus take into account the correlation between the MET_x and MET_y components in the MEM [125]. This leads to the full MET transfer function in the form of a multivariate Gaussian

$$W_{\mathrm{MET}}(\boldsymbol{\rho}_T | \sum_k \boldsymbol{p}_k) = \frac{1}{2\pi |\Sigma|^{1/2}} \exp\left[-\frac{1}{2} (\boldsymbol{\rho}_T - \sum_k \boldsymbol{p}_k)^T \Sigma^{-1} (\boldsymbol{\rho}_T - \sum_k \boldsymbol{p}_k) \right], \tag{4.26}$$

where we currently have assumed $\Sigma = \sigma_{\mathrm{MET}} \mathbf{I}$ as a simplification.

4.3.2 Lost quarks

We have extended the MEM to deal with scenarios where one (or more) of the quarks labelled by q from the underlying $\mathrm{pp} \to \mathrm{t\bar{t}H}(\to \mathrm{b\bar{b}})$, $\mathrm{t\bar{t}} + \mathrm{b\bar{b}}$ processes are not reconstructed due to either being out of the geometrical detector acceptance $|\eta_j| > \eta_{\mathrm{cut}}$ or due to hadronising into jets that are below an experimental energy threshold $E_j < E_{\mathrm{cut}}$. To achieve this, we formally extend the space of observables $\boldsymbol{y} \to \boldsymbol{y}' = (\boldsymbol{y}, (E_q, \boldsymbol{e}_q)_{q \in \mathrm{lost}})$ and the per-event MEM probability

$$P(\boldsymbol{y}) \to P'(\boldsymbol{y}') = \frac{1}{\sigma'} \frac{\mathrm{d}\sigma_i}{\mathrm{d}\boldsymbol{y} \prod_{q \in \mathrm{lost}} \mathrm{d}E_q \mathrm{d}\boldsymbol{e}_q} \tag{4.27}$$

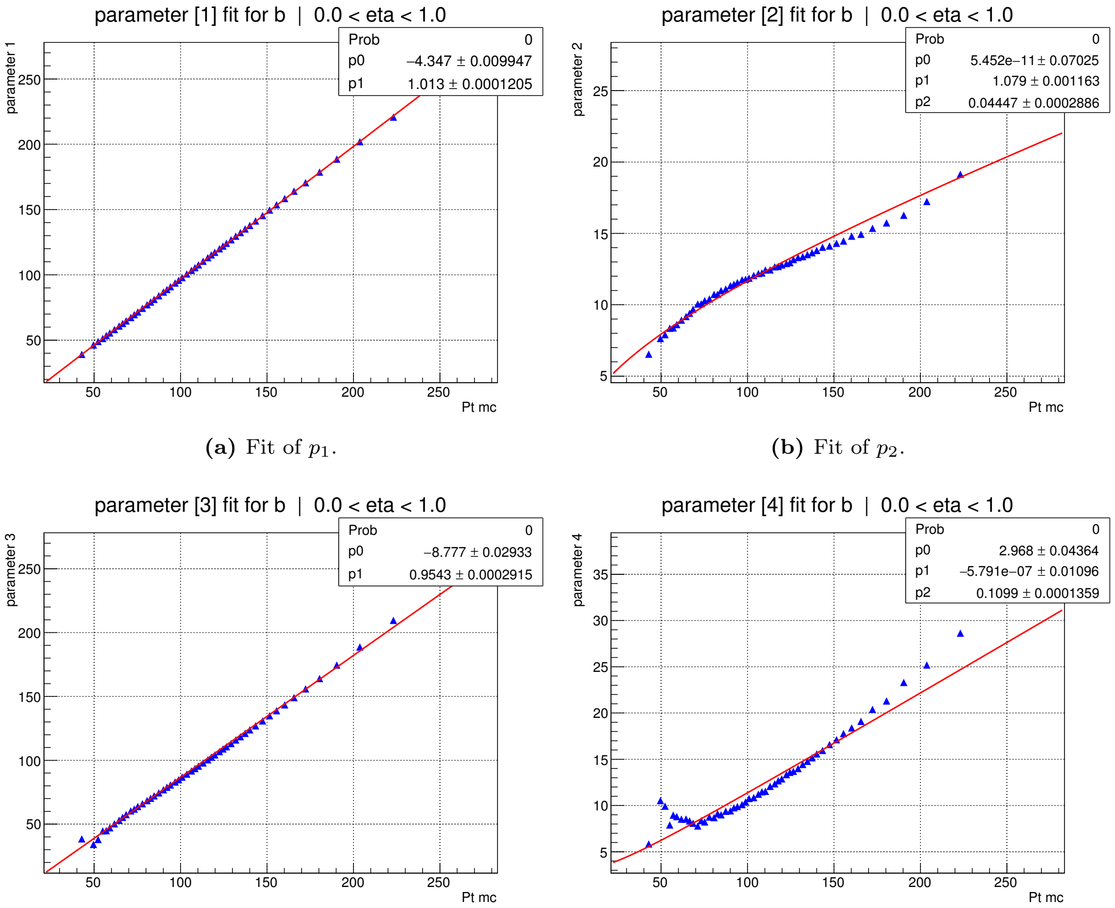
and integrate out the unobserved quantities:

$$P(\boldsymbol{y}) = \int_{\mathcal{A}'} \Big[\prod_{q \in \mathrm{lost}} \mathrm{d}E_q \mathrm{d}\boldsymbol{e}_q \Big] P'(\boldsymbol{y}'). \tag{4.28}$$

As the quark can be out of acceptance either due to failing the geometrical acceptance in η or falling below the energy threshold, the integral over the lost quarks simplifies to

$$P(\boldsymbol{y}) = \int \cdots \times \prod_{q \in \mathrm{lost}} \left[\int_{|\eta_q| \leq \eta_{\mathrm{cut}}} \mathrm{d}\Omega_q \epsilon(E_q, \eta_q) \cdots + \int_{|\eta_q| > \eta_{\mathrm{cut}}} \mathrm{d}\Omega_q \ldots \right] \times \ldots \tag{4.29}$$

where $\epsilon(E_q, \eta_q)$ is the probability for a quark of energy E_q at a pseudo-rapidity η_q to hadronise to a jet below the energy threshold $E_{\mathrm{cut}}(\eta_q)$ and thus fails to be reconstructed:



(a) Fit of p_1.

(b) Fit of p_2.

(c) Fit of p_3.

(d) Fit of p_4.

Figure 4.2: The fitted parameters of the function eq. (4.25) in the form of $\alpha_n(p_{T,\text{gen}}|\eta_q, \text{flavour})$. Overall, we see a sufficiently flexible description of the parameters derived in the individual fits (blue) by the polynomial fit (red). The bias in α_4 near $p_{T,\text{gen}} \simeq 50$ GeV is due to the underlying double Gaussian fit not being able to accommodate the jets that fall below the p_T threshold when reconstructed with the experimental resolution and does not affect the MEM significantly. The fits are derived using $\mathrm{t\bar{t}}$+jets simulation.

$$\epsilon(E_q,\eta_q) = \int_0^{E_{\mathrm{cut}}(\eta_q)} \mathrm{d}E_j W(E_j|E_q). \tag{4.30}$$

This means that for every lost quark, we add two integration variables through $\mathrm{d}\Omega_q$, as well as an extra combination for choosing which of the quarks did not produce a jet. The flexibility afforded by this technique, which makes the MEM applicable for cases where we may not always reconstruct the exact multi-particle final state, thus comes at a computational cost which is evaluated in section 4.3.7.

4.3.3 Treatment of QCD radiation

The MEM as formulated above does not account for QCD radiation, which at the LHC can be substantial [126]. In particular, we estimate using simulation that the ratio between seven and six reconstructed jets is $N_{7\mathrm{jet}}/N_{6\mathrm{jet}} \simeq 0.5$ for $\mathrm{t\bar{t}H}(\to \mathrm{b\bar{b}})$, whereas the hard interaction produces only six partons at LO. Therefore there is a substantial fraction of events with more jets than is naively required for the MEM and we need to either define a selection among those jets or describe them via a model in the MEM probability of eq. (4.5). Additionally, ISR that is not reconstructed as jets also affects the kinematics of the event with an unknown component in the final momentum balance. We have implemented and compared several alternative techniques that extend the MEM to final states with additional jets arising from radiation.

First, to deal with unreconstructed ISR, we note that momentum balance can be restored by performing a Lorentz boost with $\beta(\boldsymbol{p}_k) = (\sum_k p_{k,x}, \sum_k p_{k,y}, \beta_z)$ in the transverse plane, such that a Born-like configuration with a null transverse momentum is achieved. The longitudinal component β_z of the boost is in principle unknown and should be integrated out. However, we find that it can be ignored by setting $\beta_z = 0$, since it corresponds to different values for gluon fractions $x_{1,2}$ which are found not to affect the performance of the matrix element significantly. This simple treatment of ISR kinematics is necessary to evaluate the LO matrix element with proper momentum balance, but it does not take into account the dynamics of ISR production nor the properties of QCD radiation in general.

A possible step forward would be to use additional matrix elements with more final state partons to take into account extra jets. In particular, for events with one extra jet, we can use the $\mathrm{pp} \to \ \mathrm{t\bar{t}H}(\to \mathrm{b\bar{b}})\ + \mathrm{g}$ matrix element with an additional gluon [127]. For a semileptonic top decay, we would thus have seven final state partons that need to be matched to seven jets. This approach has the advantage of not making the assumption that the extra jet arises necessarily from ISR, but is instead a full treatment of the $2 \to 7$ scattering with perturbation theory. However, the additional computational complexity is significant, especially if (anti)quark diagrams are included in addition to the gluon. We study this approach further in section 4.4.

Sudakov reweighting allows us to approximate the effects of ISR in the scattering amplitude in terms of splitting functions derived from QCD [126]. At detector level, the Sudakov factor is approximated by a log-normal transfer function of the form

$$W(p_T) = \frac{1}{\sqrt{2\pi}\sigma p_T} \exp\left[\frac{-(\ln{[p_T/1\ \mathrm{GeV}]} - \mu)^2}{2\sigma^2}\right] \tag{4.31}$$

that approximates the probability of the parton-level transverse momentum p_T resulting in an observed recoil ρ_T below an experimental threshold $\rho_{T,0} < 30$ GeV, taking into account the detector resolution from simulations. Here the values $\mu = 4.1$ and $\sigma = 1.35$ are estimated from MC simulation and the unknown ISR momentum p_T is integrated out. We have implemented this empirical factor and compared it with the nominal LO MEM, however, we have seen that the changes are very small and compatible with MC statistical uncertainties.

In case a significant missing transverse momentum is observed in the event, a separate double-Gaussian transfer function would be appropriate in the Sudakov factor. However, since we independently have to consider a transfer function for the MET due to the presence of neutrinos, the modelling of ISR recoil can be absorbed in the MET recoil transfer function, as explained in section 4.3.1.

4.3.4 Event interpretations

The scattering amplitude $|\mathcal{M}_{\boldsymbol{\theta}}|$ depends on the assumed process $\mathcal{H}$ and the interpretation of the event $\mathcal{C}$. Various interpretations are possible, depending on the observed multiplicity of charged leptons and jets. First, the number of charged leptons fixes the choice of the top decay amplitudes between the semileptonic, dileptonic or all-hadronic. This in turn determines the number of quarks in the final state. Depending on the nature and number of the observed jets, we consider four classes of event interpretations:

- **Fully reconstructed events**: In this case $N_{\mathrm{jets}} = N_{\mathrm{quarks}}$, such that each quark is associated with a jet. This is the standard case.
- **Over-reconstructed events**: If $N_{\mathrm{jets}} > N_{\mathrm{quarks}}$, we may choose up to N_{quarks} jets out of the set of reconstructed jets, such that each quark is associated with a jet and the left-over jets are ignored. In doing this, we sum over the possible choice of jets using the combinatorial treatment of the MEM.
- **Over-reconstructed events with an ISR interpretation**: If $N_{\mathrm{jets}} = N_{\mathrm{quarks}} + 1$, we may act as above, but choose to interpret the extra jet as arising from gluon radiation in the initial state using the LO diagrams $\mathrm{t\bar{t}H + g}$ and $\mathrm{t\bar{t} + b\bar{b} + g}$ in a full MEM approach by extending the reconstructed phase space $\boldsymbol{y}$. Due to the larger complexity of the involved diagrams, this approach is computationally more costly than the above, but includes information on the dynamics and kinematics of the additional jet.
- **Partially-reconstructed events**: In case $N_{\mathrm{jets}} < N_{\mathrm{quarks}}$, we treat a number of quarks as lost and integrate over their directions as described in section 4.3.2. This allows the MEM to be used as a discriminator on a larger set of events, but we may also apply this hypothesis in case we suspect some jets may be mismeasured and not correspond to the underlying hard interaction, thus integrating them out despite a measurement.

In the most general case, we could evaluate all possible event interpretations and use them as a combined event discriminator that makes maximal use of kinematics and the prior probabilities for each hypothesis. However, this is computationally prohibitive, requiring many numerical integrals per event, with the evaluation of each hypothesis taking $\mathcal{O}(10^1)\ldots\mathcal{O}(10^3)$ seconds on a modern CPU. Therefore, it is necessary to restrict the interpretation space using assumptions that are tested on MC simulation, comparing the expected performance of various interpretation strategies and choosing one providing an acceptable trade-off between discriminator performance and computational cost. We list the interpretations that were considered in table 4.1.

interpretation	bottom quarks	light quarks	description
SL $2_{\mathrm{W}}2_{\mathrm{h}}2_{\mathrm{t}}$	4	2	fully-reconstructed semileptonic
SL $1_{\mathrm{W}}2_{\mathrm{h}}2_{\mathrm{t}}$	4	1	$(l\nu_l'\mathrm{b})_{\mathrm{t}}(\not{\mathrm{q}}\mathrm{q}'\bar{\mathrm{b}})_{\bar{\mathrm{t}}}(\mathrm{b\bar{b}})_{\mathrm{h}}$
SL $0_{\mathrm{W}}2_{\mathrm{h}}2_{\mathrm{t}}$	4	0	$(l\nu_l'\mathrm{b})_{\mathrm{t}}(\not{\mathrm{q}}\not{\mathrm{q}}'\bar{\mathrm{b}})_{\bar{\mathrm{t}}}(\mathrm{b\bar{b}})_{\mathrm{h}}$
SL $2_{\mathrm{W}}2_{\mathrm{h}}2_{\mathrm{t}} + 1\mathrm{g}$	4	3	fully-reconstructed, additional ISR gluon
DL $2_{\mathrm{h}}2_{\mathrm{t}}$	4	0	fully-reconstructed dileptonic

Table 4.1: The detailed event interpretations for semileptonic and dileptonic $\mathrm{t\bar{t}H}$ (signal) and $\mathrm{t\bar{t}} + \mathrm{b\bar{b}}$ (background) events. In the semileptonic channel, we consider cases where up to two light quarks may be lost. The minimum number of jets required for a hypothesis is the sum of the number of quarks. For the SL $1_{\mathrm{W}}2_{\mathrm{h}}2_{\mathrm{t}}$ hypothesis the direction and energy of one of the light quarks is integrated out, denoted by $\not{\mathrm{q}}$. For the fully-reconstructed semileptonic case with seven jets, we also consider the ISR-modified interpretation.

We use a number of strategies to restrict the number of combinations that need to be considered in the transfer functions of eq. (4.23).

- As we are dealing with up to two oppositely-charged leptons, we can neglect the small effect of charge confusion and assume the leptons are identified perfectly in case of a dileptonic event.

- We assume that the efficiency of correctly b tagging jets arising from bottom quarks (mistagging light jets) is sufficiently high (low) that we find bottom quark (light quark) candidates only among jets that are likely to arise from bottom quarks (light quarks) according to their b discriminator properties.
- If we have more than three candidates for the light quarks, we select the three that are most compatible with a $\mathrm{W} \to \mathrm{qq}'$ decay according to invariant mass.
- The scattering amplitude $|\mathcal{M}|^2$ is symmetric under charge exchange for $\mathrm{W} \to \mathrm{qq}'$, therefore, we compute the scattering amplitude with only one combination corresponding to a particular charge assignment of quarks.

We choose the set of jets to be interpreted as b quarks in the MEM that is most compatible with arising from four b quarks by using a likelihood ratio based on jet b discriminators, as will be explained further in section 5.2.3. Overall, the number of transfer function combinations that are required for each hypothesis in the categories we use is shown in table 4.2. These assumptions are found not to reduce the performance of the MEM significantly, but they decrease the number of combinations (and thus the computational burden) by an order of magnitude, as seen on figure fig. 4.3. In fact, we find that being more strict with the assumptions can increase the discriminating power of the MEM, since the likelihood is only an approximation, therefore discarding some unlikely permutations in eq. (4.23) is an optimisation in both CPU time and performance.

interpretation	8+ jets	7 jets	6 jets	5 jets	4 jets
SL $2_\mathrm{W}2_\mathrm{h}2_\mathrm{t}$	72/36	36	12	-	-
SL $1_\mathrm{W}2_\mathrm{h}2_\mathrm{t}$	-	-	-	12	-
SL $0_\mathrm{W}2_\mathrm{h}2_\mathrm{t}$	-	-	-	-	12
SL $2_\mathrm{W}2_\mathrm{h}2_\mathrm{t} + 1\mathrm{g}$	72/36	36	-	-	-
DL $2_\mathrm{h}2_\mathrm{t}$	12	12	12	12	12

Table 4.2: The number of MEM combinations in associating quarks to jet for various MEM categories. In SL events with seven or more jets, we choose up to four candidates based on invariant mass among the light jets for the W-boson reconstruction, in order to prevent a combinatorial explosion for events with a very high jet multiplicity. In DL events, we always choose exactly four candidates for the b quarks.

4.3.5 Integration

The MEM is implemented as a dedicated code in C++, relying on the `OpenLoops` library [127] interfaced via C++ for the evaluation of the hard scattering amplitude. The `ROOT` package [128] is used for numerical Lorentz algebra and `CLING` [129] for interfacing the code to Python and the rest of the $\mathrm{t\bar{t}H}(\to \mathrm{b\bar{b}})$ analysis. The PDFs are evaluated using the CTEQ6.6 [130] set via the `LHAPDF` package [131]. The numerical integration routines rely on the `VEGAS` algorithm [132] that uses multiple passes to refine the integration grid, with the maximal number of evaluations tuned for approximately $\Delta I/I < 2.5 \ldots 5\%$ relative numerical accuracy on the integral, suitable for use in a discriminator. We use the `CUBA` package [133] for numerical integration, as it supports vector-valued integrands. The distribution of expected numerical accuracy is shown in fig. 4.5 and illustrates the convergence of the numerical integration. We see that the computational cost is around 1-2 CPU minutes per event including both the signal and background hypotheses and that the numerical integration is accurate to within 5% on average. Transfer functions can be provided in a flexible parametrisation using `ROOT`, however, as described in section 4.3.6, we have also provided faster, optimised versions of the Gaussian transfer functions.

In section 4.3 we have expressed the integral of eq. (4.5) in terms of variables which are aligned with the peaks of the integrand. In general, we could integrate over the full range of these variables, however, to improve convergence, we restrict the integration over energies to a symmetric confidence region around the reconstructed energy based on the transfer functions. In particular, we derive the lower ($E_L(E_j, \alpha)$) and upper ($E_H(E_j, \alpha)$) integration boundaries for a jet with energy E_j from

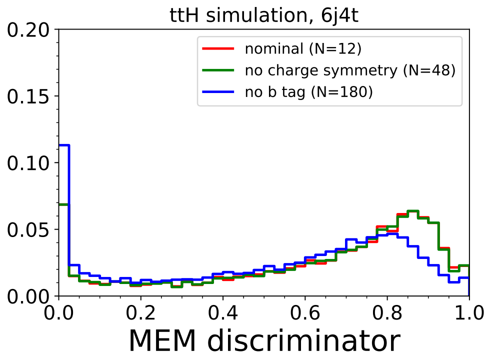


Figure 4.3: The effect of combination assumptions in the MEM discriminator based on $t\bar{t}H(\rightarrow b\bar{b})$ simulation. We compare the nominal case (red), where the combinations are restricted according to b tagging and assuming a single charge assignment, with the case where the charge assignment restriction is removed (green, 48 combinations) and the case where the b tagging restriction is removed (blue, 180 combinations). We see that the charge assignment has negligible effect on the overall discriminator shape, whereas the lack of b tagging significantly increases the number of permutations and thus reduces the discriminating power of the MEM.

$$\int_{E_L(E_j,\alpha)}^{E_H(E_j,\alpha)} \mathrm{d}E_q W(E_j|E_q) = \alpha \tag{4.32}$$

with $\alpha = 0.95$ and choose the integration variable x_q for the quark energy such that $E_q(x_q) = E_L + x_q(E_H - E_L)$, to restrict the integration into the range $x \in [0,1]$. The integration with respect to the angular variables is performed over the ranges $\phi \in [-\pi, +\pi]$ and $\cos\theta \in [-1,+1]$.

VEGAS algorithm

In computing the full ME scattering amplitude and transfer function convolution of eq. (4.5), we rely on numerical Monte Carlo (MC) integration in the VEGAS approach [132]. The integral is approximated by a weighted sum over points $\mathbf{x}_i$ sampled from $p(\mathbf{x})$ such that

$$I = \int_V f(\mathbf{x})\ \mathrm{d}\mathbf{x} \simeq \tilde{I} = \frac{V}{N}\sum_i^N \frac{f(\mathbf{x}_i)}{p(\mathbf{x}_i)}. \tag{4.33}$$

The variance in the estimation $\tilde{I}$ is given by

$$\sigma^2_{\tilde{I}} \simeq \frac{1}{N-1}\left(\frac{V^2}{N}\sum_i \left[\frac{f^2(\mathbf{x}_i)}{p(\mathbf{x}_i)}\right]^2 - \tilde{I}^2\right).$$

Choosing the sampling distribution $p(\mathbf{x})$ appropriately allows the variance to be reduced, such that the variance is optimal for $p(\mathbf{x}) = \frac{|f(\mathbf{x}_i)|}{\int |f(\mathbf{x})|\mathrm{d}\mathbf{x}}$, however, this is not possible without knowing the integral I that we seek. The VEGAS algorithm relies on importance sampling, which increases the amount of points where $|f|$ is large, by constructing the sampling distribution $p(\mathbf{x})$ as a piecewise constant function in the integration hypercube, refined by iteratively increasing the number of integration samples. We use between $\mathcal{O}(10^3)$ and $\mathcal{O}(5 \times 10^4)$ integration points in five successive stages. An example of the final VEGAS integration grid is shown on figure fig. 4.4, where we see that the integration points are clustered around $\alpha = 0.5$, which corresponds to an on-peak quark reconstruction $E_q = E_j$ and somewhat more spread out in the neutrino polar angle. We can see that the peak dimensions are aligned with the integration axes by the construction of the Jacobian, which is necessary for efficient integral evaluation via VEGAS.

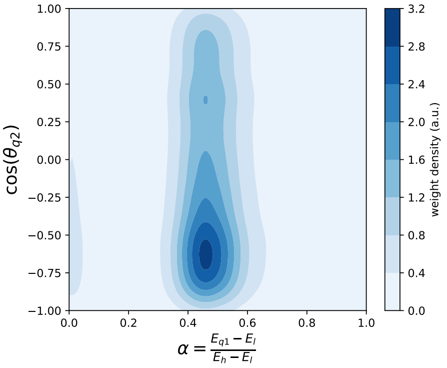
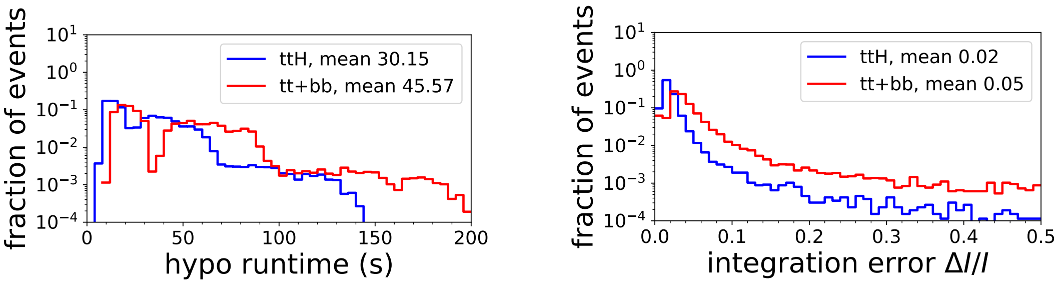


Figure 4.4: An example of the VEGAS integration weights in the plane defined by the light quark energy fraction and neutrino polar angle.



(a) Integration time of the MEM.

(b) Numerical uncertainty of the MEM

Figure 4.5: The integration time (**a**) and uncertainty resulting from numerical integration (**b**) of the MEM discriminator for the fully reconstructed hypothesis. In general, we see that the average time to compute the MEM discriminator is below a minute for both the signal and background hypotheses, with a significant spread due to the varying number of jets in events and thus the number of combinations that need to be considered. The background hypothesis is significantly more computationally expensive, due to the complexity of the $\mathrm{t\bar{t}+b\bar{b}}$ diagrams compared to $\mathrm{t\bar{t}H(\rightarrow b\bar{b})}$. The uncertainty from numerical integration is below 5% on average for both the signal and background hypotheses, with a larger spread for the background hypothesis.

For the SL $2_\mathrm{W}2_\mathrm{h}2_\mathrm{t}$ signal hypothesis, the integration is carried out over four variables: the top-associated light quark energy E_{q_t}, the neutrino directions $\cos\theta_\nu$ and $\cos\phi_\nu$ and the energy of the Higgs-associated b quark E_{b_H}. An additional integration axis is introduced for the $\mathrm{t\bar{t}+b\bar{b}}$ hypothesis, integrating over the energy of the second b quark not associated to top quark decay ($E_{\bar{b}_H}$). The addition of each lost quark adds a further energy integration, such that for the SL $0_\mathrm{W}2_\mathrm{h}2_\mathrm{t}$ hypothesis, we have six integration dimensions for the $\mathrm{t\bar{t}H(\rightarrow b\bar{b})}$ hypothesis and seven for the $\mathrm{t\bar{t}+b\bar{b}}$ hypothesis. In the dileptonic channel, the DL $2_\mathrm{h}2_\mathrm{t}$ signal hypothesis has five integration variables: two angles per neutrino, plus the energy of one of the Higgs-associated b-quarks E_{b_H}.

4.3.6 Profiling and optimisation

In order to deploy the MEM on the large data samples and the corresponding simulation expected in Run 2 of the LHC, we have optimised the code by studying the computational cost of various parts in a *profiling* study. We have used the `igprof` sampling profiler [134, 135] to analyse the computational budget spent in various subroutines of the program, which uses a random sampling technique to estimate the computational cost of various subroutines. In general, we find that the overwhelming majority of time is spent within the integrand, out of which about 40% is spent computing the transfer functions, largely in the tails of the transfer functions, 35% is spent evaluating the scattering amplitude of the hard process, 10% on computing the PDFs and about 10% on manipulating the phase space volume. The evaluation of the transfer functions at a single phase space point is about an order of magnitude faster than the scattering amplitude. In order to achieve this ratio, we implemented the transfer functions explicitly as optimised C++ functions, instead of relying on a more generic approach using symbolic functions supported in ROOT. Additionally, as a large fraction of time in optimising the VEGAS integration grid is spent in evaluating the exponential tails of the transfer function, we have used a piecewise exponential function that is suppressed far in the tails.

Currently, the MEM algorithm as implemented here can only be run on standard x86 CPU architectures. Although it has been shown that GPUs may offer strong parallelisation benefits in evaluating the integral, it would be necessary to completely port and optimise the `OpenLoops` toolset, or another matrix element library, on the GPU in a significant engineering effort [136], furthermore, GPU clusters are currently not commonplace in the Worldwide LHC Grid (WLCG), limiting the applicability of the code. However, in the future, as massively parallelised resources and automatic code translation tools become more widely available, the phase space integration could benefit significantly from parallel resources.

4.3.7 Computational budget

In this section, we present a feasibility estimation on using the MEM in a Run 2 analysis. This is necessary in order to predict the amount of computing resources that will be required. The computing time depends strongly on the number of combinations and integration variables needed for a given interpretation and event topology, as well as the total number of simulated events that are needed for the analysis. In section 4.3.7, we show the required CPU budget for evaluating the MEM on various event topologies. Based on this, we identify the MEM interpretations to apply on a given event topology. In particular, we see that the treatment of the additional gluon radiation in the scattering amplitude increases the computational cost by about a factor of five, mainly due to the $\mathrm{t\bar{t}+b\bar{b}}$ amplitude. Furthermore, we find that computing the MEM with a hypothesis that considers all the reconstructed jets provides the best trade-off between performance and computing cost.

method	time $\mathrm{t\bar{t}H}$ (s)	time $\mathrm{t\bar{t}+b\bar{b}}$ (s)	ROC AUC	ϵ_bkg	total (h) / 1k
SL, ≥ 7jet, $2_\mathrm{W}2_\mathrm{h}2_\mathrm{t}$	45.8 ± 18.9	69.4 ± 26.1	0.315	0.232	32.00
SL, ≥ 7jet, $2_\mathrm{W}2_\mathrm{h}2_\mathrm{t}1_\mathrm{g}$	71.7 ± 27.1	471.7 ± 50.6	0.317	0.233	150.94
SL, ≥ 6jet, $2_\mathrm{W}2_\mathrm{h}2_\mathrm{t}$	30.2 ± 21.0	45.4 ± 30.9	0.321	0.233	21.00
SL, ≥ 6jet, $1_\mathrm{W}2_\mathrm{h}2_\mathrm{t}$	64.8 ± 22.7	101.1 ± 33.0	0.307	0.210	46.07
SL, ≥ 6jet, $0_\mathrm{W}2_\mathrm{h}2_\mathrm{t}$	83.9 ± 20.4	136.3 ± 28.9	0.294	0.218	61.16
SL, 5jet, $1_\mathrm{W}2_\mathrm{h}2_\mathrm{t}$	25.4 ± 7.1	39.9 ± 9.6	0.293	0.198	18.13
SL, 5jet, $0_\mathrm{W}2_\mathrm{h}2_\mathrm{t}$	84.7 ± 20.3	136.9 ± 28.9	0.291	0.217	61.56
SL, 4jet, $0_\mathrm{W}2_\mathrm{h}2_\mathrm{t}$	84.3 ± 20.7	136.0 ± 29.2	0.333	0.275	61.21
DL, ≥ 4jet, $0_\mathrm{W}2_\mathrm{h}2_\mathrm{t}$	55.7 ± 13.7	90.4 ± 19.3	0.223	0.124	40.58

Table 4.3: The CPU budget and separation power of the MEM in the SL channel using various event interpretations. We show the time required to evaluate the signal and background hypotheses, the receiver operating characteristic (ROC) area under curve (AUC), the efficiency of selecting background events at a signal selection efficiency of 50% (ϵ_bkg) and the total time required to compute the MEM discriminator for 1000 events.

4.3.8 Propagating uncertainties

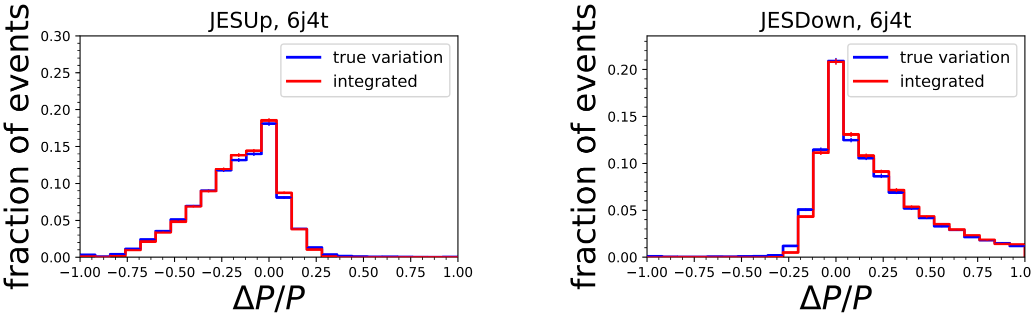
When using the MEM in a realistic experimental analysis, we need to evaluate the effect of systematic uncertainties on the MEM discriminant. In general, uncertainties modify the observables $\boldsymbol{y} \rightarrow \boldsymbol{y}^*$, for example the jet energies may be modified by uncertainties in the jet energy scale calibration. The naive approach to estimate the sensitivity of the discriminator would be to recompute the MEM discriminator probabilities $P_{\boldsymbol{\theta}}(\boldsymbol{y}) \to P(\boldsymbol{y}^*)$. However, this turns out to be impractical, since the number of individual variations that need to be considered can easily reach $\mathcal{O}(10^2)$ and it is not realistic or practical to expend two orders of magnitude more computational resources.

In order to reduce the computing time, we have developed an approximation for the effect of jet energy scale uncertainties using the analytical form of the MEM. We first note that the variations are generally small, such that $\boldsymbol{y}^* \simeq \boldsymbol{y}+\delta\boldsymbol{y}$. Therefore, the numerical integration described in section 4.3 would be performed on almost the same phase space, with an equivalent VEGAS integration grid.

Furthermore, we see from eq. (4.5) that the observables enter the definition of the MEM probability primarily through the transfer functions $W(\boldsymbol{y}|\boldsymbol{p})$ and affect the integration volume only secondarily. The most computationally costly part in the integrand is the evaluation of the LO scattering amplitudes for the hard process. Therefore, if we can promote the integrand to a vector-valued quantity, such that

$$|\mathcal{M}(\boldsymbol{p})|^2 W(\boldsymbol{y}|\boldsymbol{p}) \to |\mathcal{M}(\boldsymbol{p})|^2 \begin{pmatrix} W(\boldsymbol{y}|\boldsymbol{p}) \\ W(\boldsymbol{y}+\delta\boldsymbol{y}_1|\boldsymbol{p}) \\ \dots \\ W(\boldsymbol{y}+\delta\boldsymbol{y}_n|\boldsymbol{p}) \end{pmatrix}, \tag{4.34}$$

then the integration of the nominal and varied weight can be performed in a single pass using a grid optimised for the whole integration. We have tested this approach by comparing the variation evaluated using vector integration to the full computation with shifted inputs, showing the comparison of the full variation and the approximation in fig. 4.6. As we only wish to estimate the sensitivity of the analysis to such sources of uncertainty, it is sufficient if the approximated variation has the same magnitude and direction as the true variation.

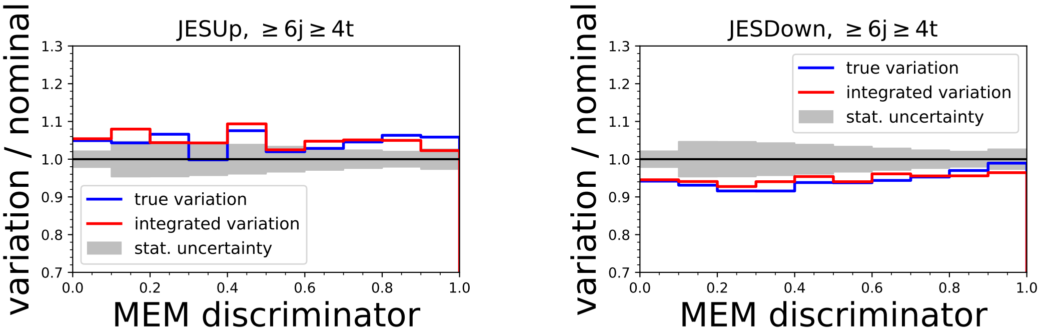


(a) Up-variation

(b) Down-variation

Figure 4.6: The effect of JES up and down variations on the MEM signal probability $P_{\mathrm{t\bar{t}H(\to b\bar{b})}}$ for the signal hypothesis with the full computation (blue) and with the vector integration approximation method. We see that both the up and down shifts have the correct direction and magnitude. This estimation is done on $\mathrm{t\bar{t}H(\to b\bar{b})}$ simulation with exactly six jets and exactly four b tags, requiring that the jet variations do not change the jet multiplicity in the final state.

Additional complexity is introduced due to variations in the uncertainties possibly changing the topology of the reconstructed final state, as scaling jet energies down (up) may cause jets to migrate under (above) the experimental threshold $p_{T\mathrm{cut}}$. In order to account for this, in case a particular variation $\boldsymbol{y}+\delta\boldsymbol{y}_n$ changes the reconstructed final state, the MEM is still recomputed using the new topology. On fig. 4.7, we compare the fully varied MEM discriminator to the approximation, taking into account these jet multiplicity migrations.



(a) Up-variation

(b) Down-variation

Figure 4.7: The effect of JES up and down variations on the MEM discriminator (likelihood ratio) with the full computation (blue) and with the vector integration approximation method (red). As before, we see that both the up and down shifts have the correct direction and magnitude. The differences are not significant compared to the MC statistical uncertainty. This estimation is done on $\mathrm{t\bar{t}H}(\to \mathrm{b\bar{b}})$ simulation with at least six jets and at least four b tags, taking into account possible bin-to-bin migrations due to JEC uncertainties.

4.3.9 MEM on the WLCG

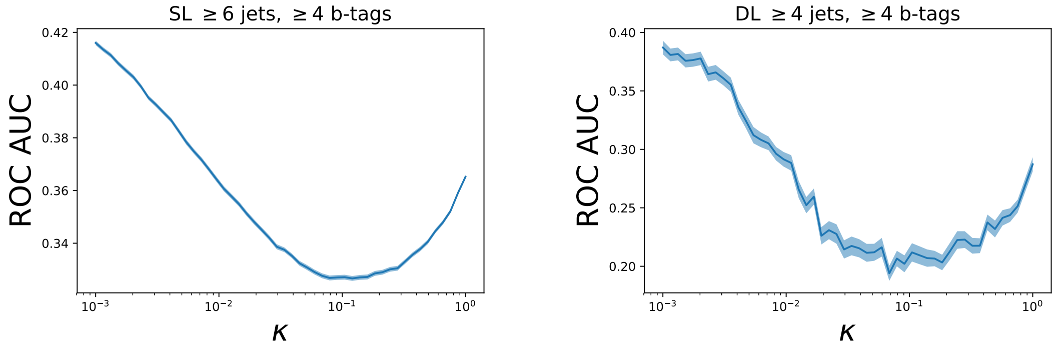
From the computational cost of the MEM shown in section 4.3.7 it is apparent that it is necessary to use distributed computing systems in order to have a reasonable turn-around time for the analysis. Therefore, we have parallelised the workflow both on the level of a computing cluster using `grid-control` and the WLCG using `CRAB`. On the WLCG, we have thus been able to take advantage of CMS computing resources opportunistically and have demonstrated that the MEM as implemented here is able to run on a wide range of data centres on a planetary scale. For this, we relied on `CMSSW` to provide a consistent environment along with user-provided external dependencies such as `OpenLOOPS`. We integrated the MEM into a multi-step workflow that produced the final analysis data sets directly from the CMS MiniAOD data stage in a single pass on the WLCG. This way, we were able to benefit from load balancing using data locality at CMS and reduced the number of manual intermediate steps and data management, which can be error prone. Overall, we were able to iterate with the full analysis from MiniAOD to the final limits in a matter of a few days under optimal conditions.

4.4 Expected performance

We study the expected performance of the MEM on a MC simulation sample of $\mathrm{t\bar{t}H}(\to \mathrm{b\bar{b}})$ and $\mathrm{t\bar{t}}$+jets. First, in fig. 4.9, we verify that the signal and background probabilities indeed behave as expected on their respective MC simulations. In particular, we see that the signal probability $P_{\mathrm{t\bar{t}H}(\to \mathrm{b\bar{b}})}$ is on average higher on the $\mathrm{t\bar{t}H}(\to \mathrm{b\bar{b}})$ sample and vice versa, as we would expect. This allows us to construct an efficient likelihood ratio discriminator in the form

$$P_{\mathrm{s/b}}(\mathbf{y}) = \frac{P_{\mathrm{t\bar{t}H}}(\boldsymbol{y})}{P_{\mathrm{t\bar{t}H}}(\boldsymbol{y}) + \kappa \cdot P_{\mathrm{t\bar{t}+b\bar{b}}}(\boldsymbol{y})} \tag{4.35}$$

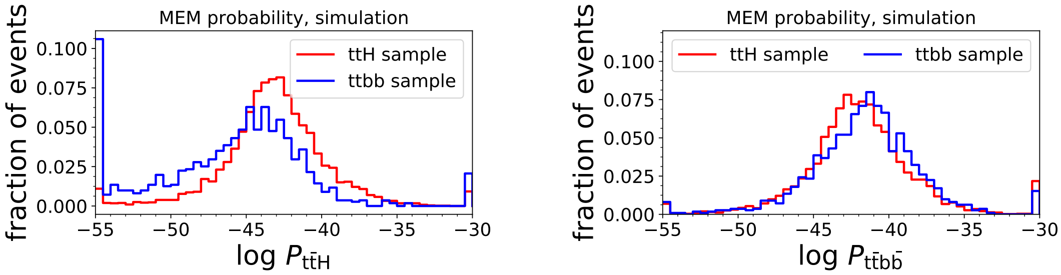
The scale factor κ is optimised to adjust the relative normalisation of the signal and background probabilities and is introduced to allow the dynamic range of $P_{\mathrm{s/b}}$ to be uniform in the range $P_{\mathrm{s/b}} \in [0,1]$. Adjusting this coefficient does not change the signal-to-background discrimination power of the discriminant as it is a monotonic rescaling, but allows us to discretise the distribution into a small number of bins without losing sensitivity. We find that the value $\kappa = 0.1$ results in the optimal signal-to-background discrimination in all categories, as can be verified in fig. 4.8.



(a) Semileptonic category

(b) Dileptonic category

Figure 4.8: The MEM discriminator performance as a function of the signal-to-background scale factor κ. We compute the ROC AUC from histograms of the MEM discriminant with six bins uniformly in the range $[0, 1]$ and find that within statistical uncertainties, the optimal discrimination is achieved for $\kappa = 0.1$.



(a) MEM probability for the $\mathrm{t\bar{t}H}$ hypothesis.

(b) MEM probability for the $\mathrm{t\bar{t}+b\bar{b}}$ hypothesis

Figure 4.9: The expected distribution of the signal probability $P_{\mathrm{t\bar{t}H(\to b\bar{b})}}$ and the background probability $P_{\mathrm{t\bar{t}+b\bar{b}}}$ on MC simulation. We see that for the signal sample, the signal probability is on average higher than the background probability, and vice versa for the background. Here, we have selected events with exactly one isolated lepton, at least six jets, out of which four must be b tagged. Furthermore, the jets are required to be matched to quarks from the corresponding hard interaction on generator level.

The performance of the MEM depends crucially on whether the jets in the observed final state can be correctly associated to the particles from the hard interaction. For the following comparisons, we define the full $\mathrm{t\bar{t}H}(\to \mathrm{b\bar{b}})$ selection, corresponding to all the events that pass the detector-level selection criteria; and the matched selection, where the full selection is augmented by further requiring that all the jets in the event can be matched to the generator-level partons of the corresponding hypothesis. This matching is done geometrically using a cone size of $\Delta R = 0.3$.

For example, for the signal hypothesis in the $2_{\mathrm{W}}2_{\mathrm{h}}2_{\mathrm{t}}$ interpretation, we require that the two light jets can be matched to the quarks from the hadronic W decay, two b jets can be matched to bottom quarks from the Higgs and two b jets to bottom quarks from the top quark decay. This is done using generator-level information with the full decay chain. For the $1_{\mathrm{W}}2_{\mathrm{h}}2_{\mathrm{t}}$ hypothesis, we require only one of the light jets to be matched to a quark from the W boson, and for the $0_{\mathrm{W}}2_{\mathrm{h}}2_{\mathrm{t}}$ hypothesis, the only the b jets are required to be matched to b quarks. We show the estimated matching fractions in section 4.4. We find that in the semileptonic $\geq 6j \geq 4b$ category, the jets can be fully matched to the quarks in approximately 20% of the cases, whereas in the dileptonic $\geq 4j \geq 4b$ category, the fraction is 59%, allowing the MEM computation to be more effective.

category	hypothesis	full final state	Higgs and top	Higgs only
SL $\geq$6j$\geq$4b	SL $2_\mathrm{W}2_\mathrm{h}2_\mathrm{t}$	0.20	0.51	0.69
SL $\geq$6j3b	SL $2_\mathrm{W}2_\mathrm{h}2_\mathrm{t}$	0.09	0.24	0.52
SL 5j$\geq$4b	SL $1_\mathrm{W}2_\mathrm{h}2_\mathrm{t}$	0.36	0.59	0.74
SL 5j3b	SL $1_\mathrm{W}2_\mathrm{h}2_\mathrm{t}$	0.14	0.23	0.51
SL 4j4b	SL $0_\mathrm{W}2_\mathrm{h}2_\mathrm{t}$	0.61	0.61	0.75
SL 4j3b	SL $0_\mathrm{W}2_\mathrm{h}2_\mathrm{t}$	0.18	0.18	0.47
DL $\geq$4j$\geq$4b	DL $2_\mathrm{h}2_\mathrm{t}$	0.59	0.59	0.74
DL $\geq$4j3b	DL $2_\mathrm{h}2_\mathrm{t}$	0.30	0.30	0.55

Table 4.4: The fraction of events where the jets could be matched to partons from the hard interaction under the full hypothesis consisting of all the quarks from the W-boson, top and Higgs decay; matching only the b quarks from the Higgs and top decay and finally only the quarks from the Higgs decay. We see that for the SL $\geq 6j \geq 4b$ category, 20% of events could be matched to the full final state, whereas in the DL $\geq 4j \geq 4b$ category, the equivalent fraction is 59%. This stark difference arises due to the difficulty of reconstructing light quarks from the hadronic W-boson decay. Furthermore, we see that in the categories with three b tags, the fraction of events with full matching is significantly lower, especially in terms of matching for the top quark decay products. The estimation is made using $\mathrm{t\bar{t}H}(\rightarrow \mathrm{b\bar{b}})$ simulation.

4.4.1 Semileptonic categories

We compare the MEM distributions and expected performance on events with four jets, five jets and ≥ 6 jets in the single-lepton channel. Using these three broad categories, we can see the effect of adding additional information about jet kinematics to the MEM discriminator. We further split the events into two subcategories requiring at least four b tagged jets, corresponding to a high-purity selection, and exactly three b tagged jets, corresponding to a lower-purity MEM control region. This allows us to evaluate the effect of incorrect hypotheses on the MEM performance, as the lower-purity category will have a large fraction of events which cannot be matched according to the above prescription.

4-jet final state

On the SL categories with only four jets using only the kinematics of the b jets, we are able to achieve a background efficiency of $\simeq 30\%$ at a signal efficiency of 50% by selecting a sample where the four jets are likely to arise from bottom quarks using b tagging, as seen in fig. 4.10. Extending the MEM to the three b tag category shows the importance of having selected the right candidates for the b quarks, as can be seen in fig. 4.11, where the lower matching fraction in signal results in a lower discriminator performance. Therefore, we conclude that it is feasible to use the newly-introduced $0_\mathrm{W}2_\mathrm{h}2_\mathrm{t}$-hypothesis in the 4-jet category, in case the four jets are likely to arise from the hadronisation of b quarks from the decay of the Higgs boson and the top quark.

5-jet final state

By adding the information of an additional jet in the four jet categories, the background efficiency decreases to 20% at the benchmark point of 50% signal efficiency in the four tag category in fig. 4.13. Thus we see that the additional information provided by the kinematics of the extra jet helps to constrain the MEM integration significantly. The category with three b tags performs slightly worse, with the main effect coming from the additional light jet not being associated to the W-boson, as can be seen from fig. 4.12.

6-jet final state

In the categories with six or more jets, we verify that the MEM is able to exploit the information in the fully-reconstructed category, with a $\simeq 10\%$ background efficiency in the case where the jets could be fully matched to the partons of the underlying hypothesis, as shown in fig. 4.15 for events with at least four b tagged jets. On the extended events with three b tagged jets, the MEM still

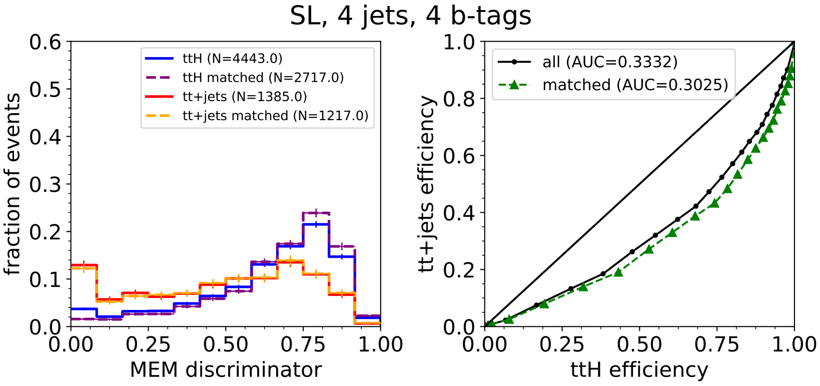


Figure 4.10: The MEM discriminator with the interpretation $0_W 2_h 2_t$, with the momenta of the two light quarks from the W-boson integrated over, in the single-leptonic category with four jets, four b tags. Compared to fig. 4.11, we see an improvement in performance, as the requirement of four b tags enhances the fraction of events where the jets can be matched to quarks, thereby increasing the fraction of events where the correct MEM hypothesis is computed.

shows relatively good discrimination (fig. 4.14), with the most significant reduction coming from the fraction of events where one of the light jets assumed to arise from $W \to qq'$ was instead spurious. We also verify the performance of the MEM discriminator against different $t\bar{t}$+jets subprocesses in fig. 4.16, where we see the best classification efficiency against $t\bar{t} + b\bar{b}$, as would be expected.

Mis-reconstructed events in the 6-jet categories

Motivated by the significant effect on the MEM from the unmatched jets, we investigate a possibility to mitigate the loss in discriminator performance by integrating over the jet momenta in case we are unsure if they correspond to the correct quarks on the level of the hard scattering. In particular, since we have seen that a significant fraction of the light jets in the fully-reconstructed category are not matched to the quarks from the W-boson decay, we consider relaxing the assumption that all the light quarks are reconstructed as jets.

This amounts to applying an interpretation such as $1_W 2_h 2_t$ that only uses the kinematics of five jets on an event with six or more jets. In general, we see that for realistic detector-level events without generator-level matching, this strategy results in a slightly improved performance, as it is less sensitive to the mis-reconstruction of the light jets, as can be seen comparing fig. 4.17 and fig. 4.15. In particular, the ROC AUC performance characteristic for the unmatched events decreases from AUC = 0.32 to AUC = 0.31 and the difference between the $t\bar{t}H(\to b\bar{b})$ MEM distribution between the matched and unmatched event samples decreases.

However, the computational cost , elaborated in section 4.3.7, would increase by about a factor of two with this choice, therefore, we do not use this approach in the $t\bar{t}H(\to b\bar{b})$ analysis at this stage. In the future, it may be interesting to explore complex hypothesis tests, which effectively combine the probabilities under various jet reconstruction hypotheses in a uniform way, smoothly interpolating between a fully-reconstructed and a partially reconstructed hypothesis based on some event-level observables.

7-jet categories

Furthermore, we have studied the effect of including diagrams containing additional QCD radiation, as outlined in section 4.3.3. These additional diagrams are only valid on events with at least seven jets. We have compared the standard fully reconstructed hypothesis in fig. 4.18 to the one with additional gluon radiation in fig. 4.19 in this category. We see that using the diagram for additional radiation has a very small effect on the final discrimination, since the effect of spurious

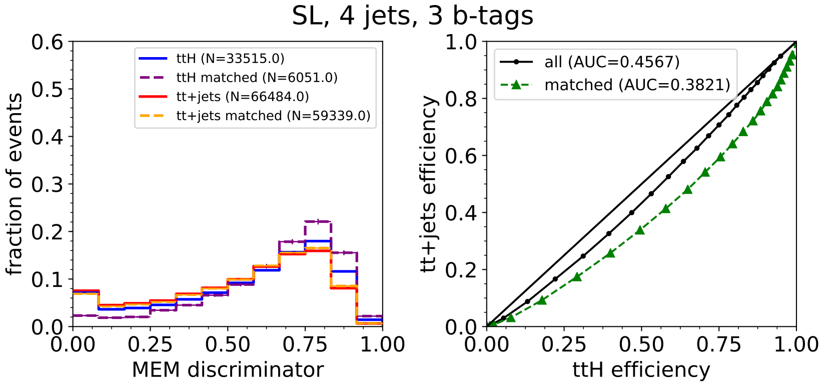


Figure 4.11: The MEM discriminator with the interpretation $0_W 2_h 2_t$, with the momenta of the two light quarks from the W-boson integrated over, in the single-leptonic category with four jets, three b tags. On the left, we show the discriminator distributions, on the right, the expected performance as characterised by the ROC for all events (black) and the events for which the jets could be matched to the partons from the hard process (green). Based on the kinematics of the fourth b jet, the MEM is already able to achieve a degree of separation (AUC = 0.41) in this category. The distributions are derived using the full $t\bar{t}H(\to b\bar{b})$ and $t\bar{t}$+jets simulations, including detector effects.

jets incorrectly assigned to the hadronic W-boson decay is significantly larger than that of the better description provided by the accurate treatment of the additional radiation. From this, we conclude that the most straightforward way to improve the performance of the MEM discriminator is to ensure that the hypothesis corresponds to the particles reconstructed in the final state. Therefore, we consider this as a cross-check of the MEM and do not investigate it further in the $t\bar{t}H(\to b\bar{b})$ analysis, concluding that improved object reconstruction and identification is more important than an accurate matrix-element level description at this stage.

4.4.2 Dileptonic categories

We have also evaluated the MEM discriminator in the dileptonic categories, where the W-bosons from both top quarks decay leptonically. At the level of the hard interaction, we expect four b quarks: two from the $H \to b\bar{b}$ decay and two from the top quark decay. In order to associate the jets to the quarks, we select the set of four jets that are most compatible with b quarks. Thus, we have twelve combinations to assign the four quarks to the four b tagged jets. In general, we see excellent performance of the MEM in the dilepton categories with both ≥ 4 and three b tags, since in around 60% of the cases, the jets in the event can be matched to the b quarks from the underlying hard process, as seen in fig. 4.20 and fig. 4.21. This means that the MEM can be used as a powerful signal-to-background discriminator in the dilepton final states, despite the complexity of the integration that is required to account for the momenta of the two neutrinos.

4.5 Summary

To summarise, we have implemented the MEM discriminator for the $t\bar{t}H(\to b\bar{b})$ analysis in Run 2 with significant improvements over the approach used in Run I. The MEM can now be computed in all relevant final states in the $t\bar{t}H(\to b\bar{b})$ analysis, and we have studied the expected performance on MC simulation. The performance studies are carried out in several categories with a varying number of jets, corresponding to the gradual increase of information available to the MEM in a multi-particle final state. We have seen that in case all the partons from the hard scattering process are present in the final state, the MEM is able to distinguish between the signal and background hypotheses with a background efficiency of $\simeq 10\%$ for a signal efficiency of 50% based only on high-level event kinematic properties. The effect of the quark jets being outside the acceptance region

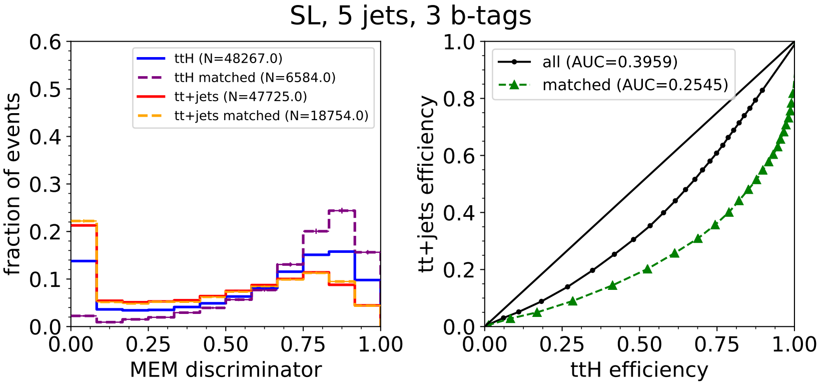


Figure 4.12: The MEM discriminator with the interpretation $1_W 2_h 2_t$, with the momentum of one of the light quarks from the W-boson integrated over, in the single-leptonic category with five jets, three b tags. Compared to fig. 4.11, we have added the information about one of the quarks from the hadronic W-boson decay, thus constraining the integration. From the ROC on the right plot, we see that for the matched case (green), the MEM performs significantly better than the corresponding discriminator in fig. 4.11, thanks to the additional information available. However, the low fraction of matched events ($<$ 10% for signal) significantly degrades the observed performance (black). Most of the mismatched events come from the additional light quark not corresponding to the hadronic W-boson decay. Thus, we see the importance of deploying the MEM on final states where the relevant particles are reconstructed.

degrades the performance of the discriminator, as the spurious jets that are instead associated to the quarks do not provide any constraining power in the integrand. Therefore, an accurate reconstruction of the final state is very important for the MEM discriminator, especially in the signal sample.

Nevertheless, the expected realistic performance of the MEM, based on full detector simulation, is excellent in the high jet and b tag multiplicity categories. In the best semileptonic category, the $t\bar{t}$+jets efficiency is around 20% at a reference signal efficiency of 50%, whereas in dilepton, the background rejection is even higher. We have further studied the resilience of the MEM to QCD radiation and found that the performance does not change markedly by introducing effective Sudakov reweighting or diagrams with additional jets. In conclusion, we find that the MEM is a suitable discriminator in the $t\bar{t}H(\rightarrow b\bar{b})$ search, given the performance and the solid theoretical foundation on which it is based.

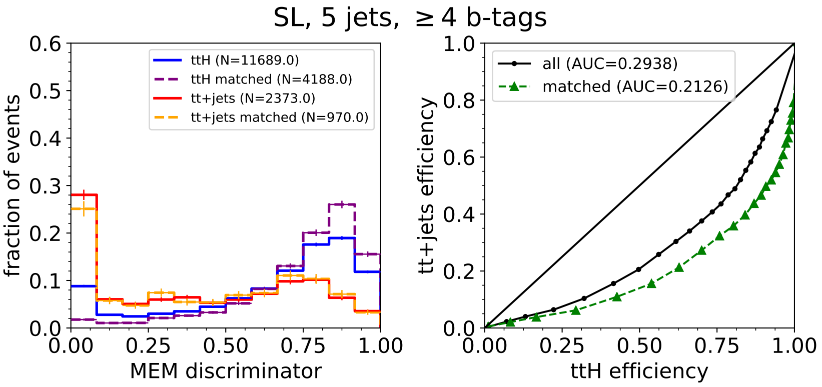


Figure 4.13: The MEM discriminator with the interpretation $1_W 2_h 2_t$, with the momentum of one of the light quarks from the W-boson integrated over, in the single-leptonic category with five jets, four b tags. By requiring four b tagged jets, the jets in the final state correspond to the partons from the underlying hypothesis in about 35% of cases for signal. We also see that the performance of the MEM discriminator is thus increased compared to fig. 4.12.

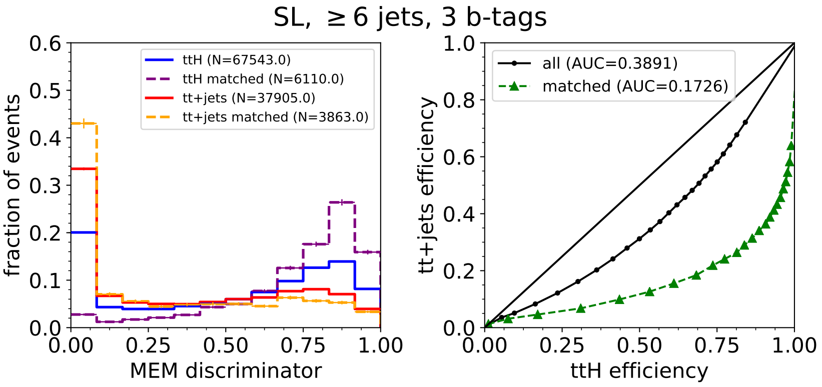


Figure 4.14: The MEM discriminator with the interpretation $2_W 2_h 2_t$, i.e. the fully reconstructed hypothesis, in the single-leptonic category with at least six jets, three b tags. We see that while the performance on the events that can be matched to the quarks from the underlying hard process is excellent, the fraction of matched events is below 10%, degrading the performance on realistic events.

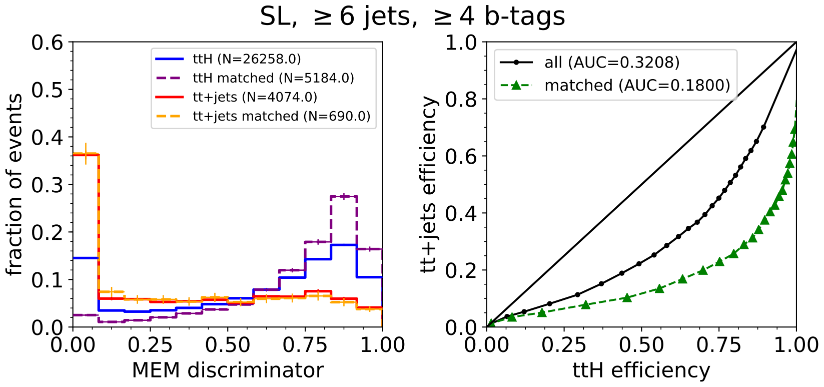


Figure 4.15: The MEM discriminator with the interpretation $2_W 2_h 2_t$, i.e. the fully reconstructed hypothesis, in the single-leptonic category with at least six jets, at least four of which must be b tagged. This is the category with the highest signal to background ratio and with a full reconstruction of the two light quark, four bottom quark final state corresponding to $t\bar{t}H(\rightarrow b\bar{b})$. We see that in case all the jets were matched to the quarks, the signal to background separation provided by the MEM is significant, with a background efficiency of around 10% at a median signal efficiency. After relaxing the matching criterion, the separation decreases by about a factor of two at the expense of the signal distribution becoming more background-like.

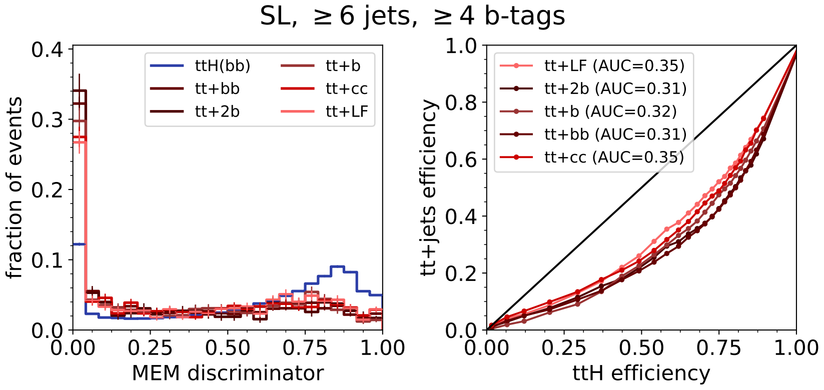


Figure 4.16: The performance of the MEM for different $t\bar{t}$+jets subprocesses, introduced in section 5.1.1, in the semileptonic ≥6-jet, ≥4-b tag category. In general, we see that the discrimination is best against $t\bar{t}+b\bar{b}$ and $t\bar{t}+2b$, followed by $t\bar{t}+b$. This arises from the use of the $t\bar{t}+b\bar{b}$ diagram as the representative diagram for the computation of the background amplitude.

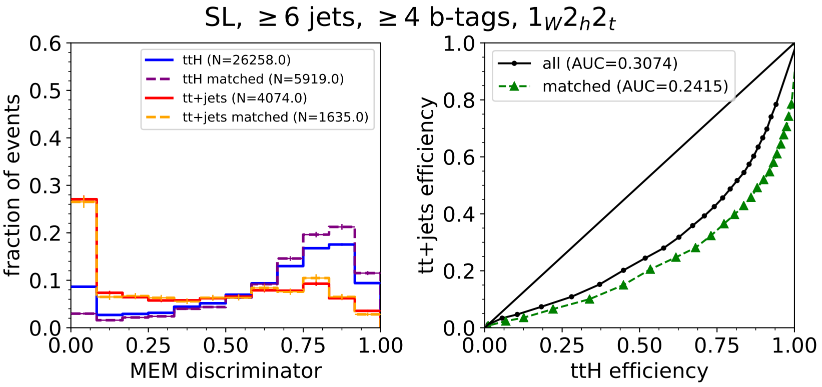


Figure 4.17: The MEM discriminator with the interpretation $1_{\mathrm{W}}2_{\mathrm{h}}2_{\mathrm{t}}$ in the single-leptonic category with at least six jets, at least four of which must be b tagged. In comparison to fig. 4.15 , we have integrated over the momentum of one of the light jets, assuming that only one of the two light jets corresponds to a quark from the W-boson decay. We see a slightly improved performance compared to the fully reconstructed hypothesis in the case of all events (black), and a slight decrease for events that are fully matched. This is consistent with the performance of the MEM being strongly dependent on whether the objects in the final state correspond to the hard scattering process.

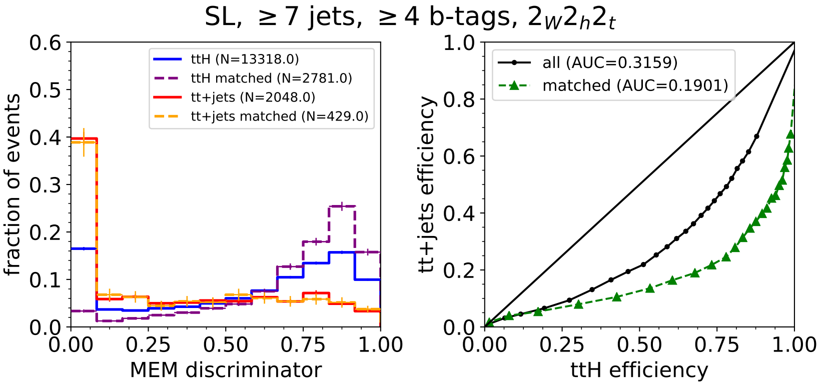


Figure 4.18: The MEM discriminator with the interpretation $2_{\mathrm{W}}2_{\mathrm{h}}2_{\mathrm{t}}$, i.e. the fully reconstructed interpretation, in the single-leptonic category with at least seven jets, at least four of which must be b tagged. In order to interpret the light jets as quarks, we sum over all possibilities of choosing two light quarks from the three light jets most compatible with a W-boson decay according to invariant mass. We see that in the presence of an additional jet, the fully-reconstructed MEM interpretation is still able to distinguish between signal and background at an acceptable level.

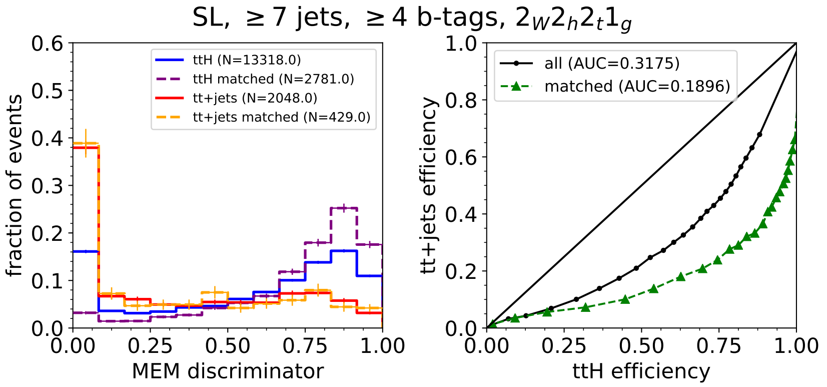


Figure 4.19: The MEM discriminator with the interpretation $2_W 2_h 2_t 1_g$, where we include the gluon radiation in the signal and background hypotheses, in the single-leptonic category with at least seven jets, at least four of which must be b tagged. Compared to the hypothesis without gluon radiation in fig. 4.18 , we do not see an improved discrimination from including this more complex hypothesis.

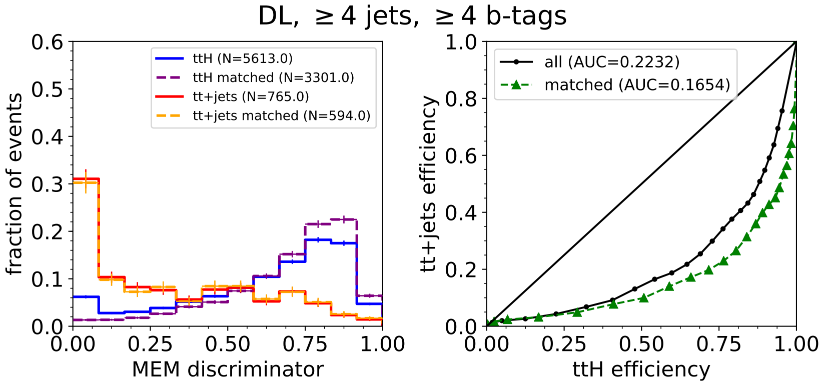


Figure 4.20: The MEM discriminator with the interpretation $2_h 2_t$ in the dileptonic category with at least four jets, all of which must be b tagged. We see that at 50% signal efficiency, the inclusive $t\bar{t}$ efficiency is around 17% as estimated on simulation.

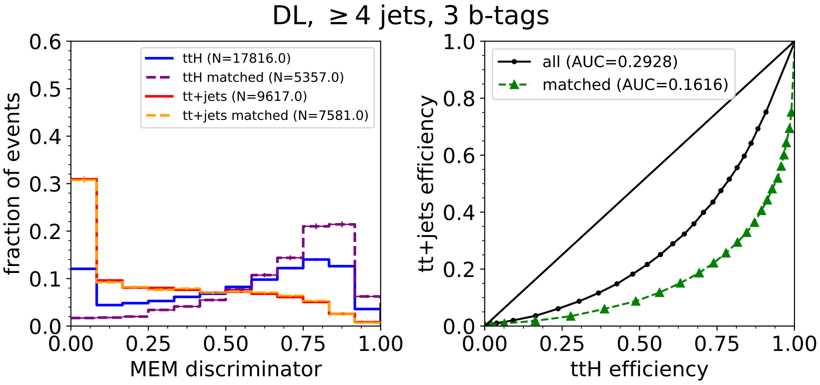


Figure 4.21: The MEM discriminator with the interpretation $2_\mathrm{h}2_\mathrm{t}$ in the dileptonic category with at least four jets, out of which three are b tagged. Compared to fig. 4.20, the expected performance decreases somewhat, with a background efficiency of 20% at the reference signal efficiency. Nevertheless, it is possible to use the DL MEM in the three b tag category as an efficient tt+bb discriminator.

5 Search for top-quark pair associated Higgs production

In this chapter, we describe the search for $\mathrm{t\bar{t}H}(\rightarrow \mathrm{b\bar{b}})$ using the matrix element method at the CMS experiment (MEM) in Run 2 of the LHC. This is based on preliminary results from CMS [137, 68] and ongoing work. We concentrate on the semileptonic (SL) and dileptonic (DL) decay channels of the top quark pair and the application of the MEM in this search, where the primary background arises from QCD production of $\mathrm{t\bar{t}}$+jets. We are showing results with the 2016 dataset of CMS data, the use of which for this thesis has been endorsed by the CMS Higgs group.

Briefly, the analysis proceeds as follows. Throughout the analysis, we use the MC simulation and data samples described in section 5.1. The detailed identification criteria for the physics objects (jets and leptons) are motivated by established top quark analyses and are described in section 5.2. The selection is motivated by the semileptonic and dileptonic decays of the top quark pair, which results in a final state containing several jets, charged leptons and MET. We select events with at least one (two) charged lepton(s) in the semileptonic (dileptonic) channel and at least four jets, out of which three must be b tagged.

The statistical analysis is described in section 5.3. We further divide the events into independent categories based on the jet and b tag multiplicity, as described in section 5.3.1. These categories are introduced in order to constrain the various sub-processes of $\mathrm{t\bar{t}}$+jets, described further in section 5.1.1, together with the $\mathrm{t\bar{t}H}$ signal cross-section. In order to extract the signal strength modifier $\mu = \sigma/\sigma_{\mathrm{SM}}$ for the $\mathrm{t\bar{t}H}$ process, we perform a combined template fit in all the categories, relying on the discriminating power provided by the MEM in categories with a high signal-to-background ratio and on other multivariate techniques in background-enhanced control regions. The fit is described in section 5.3.9.

An important component in the fit is the handling of systematic uncertainties, which drives the determination of the confidence interval of the estimated signal strength and is described in section 5.3.3. We study the fit model with respect to the systematic uncertainties in section 5.3.10 before determining the compatibility of the model with data. Finally, we present the results of the analysis in section 5.4, followed by a discussion in section 5.5 and a summary and outlook in section 5.6.

5.1 Data and simulation

We use proton-proton collision data collected by the CMS experiment at a center-of-mass energy of $\sqrt{s} = 13$ TeV, corresponding to a total integrated luminosity of 35.9 fb^{-1} in the semileptonic

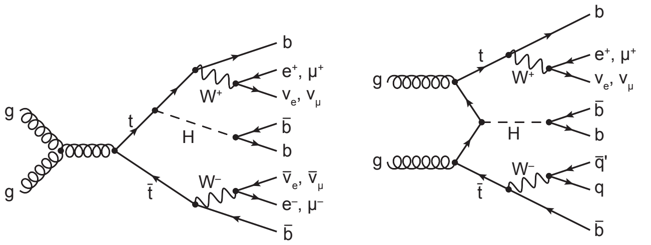


Figure 5.1: Representative leading-order Feynman diagrams for the $\mathrm{t\bar{t}H}(\rightarrow \mathrm{b\bar{b}})$ process in the gluon-fusion production channel. On the left, we show the dileptonic (DL) decay of the top quark pair, whereas on the left, we show the semileptonic (SL) decay channel.

(SL) and dileptonic (DL) channels. We have used the complete dataset of 2016 as an update over the analysis finalised in the middle of 2016, which used roughly 12.9 fb^{-1} of data [68].

We use MC simulation to model the signal and background processes, interfaced to a parton shower and hadronisation as appropriate. In order to model the detector effects, we use a detailed simulation of the detector based on `GEANT4` [138]. For the $\mathrm{t\bar{t}H}$ signal, $\mathrm{t\bar{t}}$+jets and single-top backgrounds, we use the NLO generator `POWHEG` [139, 140]. The use of an NLO model for the signal and the primary background processes is a significant advancement over Run I $\mathrm{t\bar{t}H}$ analyses at the LHC, where only a LO model was available. Besides $\mathrm{t\bar{t}}$+jets, we need to model the production of W or Z/γ^{*} bosons with additional jets (denoted W/Z+jets or commonly V+jets), which is simulated using `MG5_aMCNLO`(v. 2.2.2) [141] and diboson production (WW, WZ, ZZ), simulated using `Pythia 8` [142]. Throughout, we assume the value of the top quark mass to be $m_t \simeq 172.5\ \mathrm{GeV}/c^2$ and of the Higgs boson to be $m_H \simeq 125\ \mathrm{GeV}/c^2$. In order to describe the substructure of the protons via the parton density functions, we use the PDF parametrisation provided by NNPDF3.0 and `Pythia 8` for showering and hadronisation.

In order to model the production of hadrons, the parameters in the `Pythia 8` model have been tuned to Tevatron, LEP and LHC data [143, 144]. It has been observed that the default tune in `Pythia 8` does not reproduce the observed number of jets in data in the $\mathrm{t\bar{t}}$+jets dominated region [145]. Therefore, CMS has created a custom tune where the α_{ISR} parameter, which controls the amount of initial-state radiation, and h_{damp} parameter, which suppresses real emissions in `POWHEG`, have been adapted to reproduce the spectrum of the number of jets observed at CMS [145]. This MC tuning has been carried out on $\sqrt{s} = 8$ TeV data and significantly improves the modelling of the jet multiplicity.

sample	generator	events	cross-section
$\mathrm{t\bar{t}H}(\rightarrow \mathrm{bb})$	`POWHEG`	3 845 992	0.293 pb [146]
$\mathrm{t\bar{t}H}(\rightarrow \mathrm{non\ b\bar{b}})$	`POWHEG`	3 981 250	0.215 pb [146]
$\mathrm{t\bar{t}}$+jets (SL)	`POWHEG`	152 720 952	365.45 pb [147]
$\mathrm{t\bar{t}}$+jets (DL)	`POWHEG`	79 092 400	88.34 pb [147]
$\mathrm{t\bar{t}}+\mathrm{W}(\rightarrow \ell\nu)$	`MG5_aMCNLO`	5 280 565	0.2043 pb [148]
$\mathrm{t\bar{t}}+\mathrm{W}(\rightarrow \mathrm{qq}')$	`MG5_aMCNLO`	833 298	0.4062 pb [148]
$\mathrm{t\bar{t}}+\mathrm{Z}(\rightarrow \ell\ell, \nu\nu)$	`MG5_aMCNLO`	13 908 701	0.2529 pb [148]
$\mathrm{t\bar{t}}+\mathrm{Z}(\rightarrow \mathrm{qq}')$	`MG5_aMCNLO`	749 400	0.5297 pb [148]
WW	`Pythia 8`	7 981 136	118.7 pb [149]
WZ	`Pythia 8`	1 988 098	47.13 pb [150]
ZZ	`Pythia 8`	3 995 828	16.523 pb [150]
single top (tW)	`POWHEG`, 4FS	992 024	35.85 pb [151]
single (anti)top (tW)	`POWHEG`, 4FS	998 276	35.85 pb [152]
single top (t)	`POWHEG`, 5FS	5 993 676	136.02 pb [152]
single (anti)top (t)	`POWHEG`, 5FS	3 928 063	80.95 pb [152]
single top (s), ($\mathrm{W} \rightarrow \ell\nu$)	`MG5_aMCNLO`, 4FS	3 928 063	3.70 pb [152]
W+jets ($\mathrm{W} \rightarrow \ell\nu$)	`MG5_aMCNLO`	29 705 748	61526.7 pb [153]
$\mathrm{Z} \rightarrow \ell\ell + \mathrm{jets}, m_{\ell\ell} < 50$ GeV	`MG5_aMCNLO`	35 291 566	18610 pb [154]
$\mathrm{Z} \rightarrow \ell\ell + \mathrm{jets}, m_{\ell\ell} \geq 50$ GeV	`MG5_aMCNLO`	145 803 217	5765.4 pb [154]

Table 5.1: The MC samples used in the $\mathrm{t\bar{t}H}(\rightarrow \mathrm{b\bar{b}})$ analysis. For the simulation of the parton shower and subsequent hadronisation, `Pythia 8` is used for all samples, whereas `POWHEG` or `MG5_aMCNLO` are used for the hard process. The single top tW-channel process is simulated in the four flavour scheme (4FS), whereas other processes use the five flavour scheme (5FS) with a massless b quark being present in the proton.

We list the details of the MC samples in table 5.1 and of data in table 5.2. In order to compare simulation to data, the simulated events are weighted according to the integrated luminosity and the predicted cross sections, which are taken from the best available inclusive calculations. In particular, the $\mathrm{t\bar{t}H}$ cross section is known at NLO accuracy [155, 156, 157, 158, 159]. The Higgs boson branching fraction for $\mathrm{H} \rightarrow \mathrm{b\bar{b}}$ is affected by radiative corrections that are known up to N4LO (QCD) and NLO (electroweak), resulting in an uncertainty of about 1-2% [160, 161, 146].

dataset	trigger threshold	data taking period	integrated luminosity
single muon	24 GeV	Run B-H, 2016	35.9 fb^{-1}
single electron	27 GeV	Run B-H, 2016	35.9 fb^{-1}
dimuon	17 (8) GeV	Run B-H, 2016	35.9 fb^{-1}
dielectron	23 (12) GeV	Run B-H, 2016	35.9 fb^{-1}
muon, electron	23_μ (12_e), 23_e (8_μ) GeV	Run B-H, 2016	35.9 fb^{-1}

Table 5.2: The data samples used in this analysis. We list the transverse momentum thresholds in the trigger for the leading (subleading) charged leptons.

The inclusive cross section for $\mathrm{t\bar{t}}$+jets production is known at NNLO accuracy and includes soft gluon resummation to next-to-next leading log (NNLL) [147]. The cross-sections of the minor backgrounds are known to at least NLO, as summarised in table 5.1.

In addition to the hard interaction and the subsequent showering and hadronisation, events from additional pp interactions within the same bunch crossing (pileup) are superimposed on the simulated event for all processes. The multiplicity distribution of these additional pileup events is reweighted to match the observed number of interactions in data. Furthermore, we correct the MC simulation with additional data-driven correction factors for b tagging and lepton efficiencies as described in section 5.3.3.

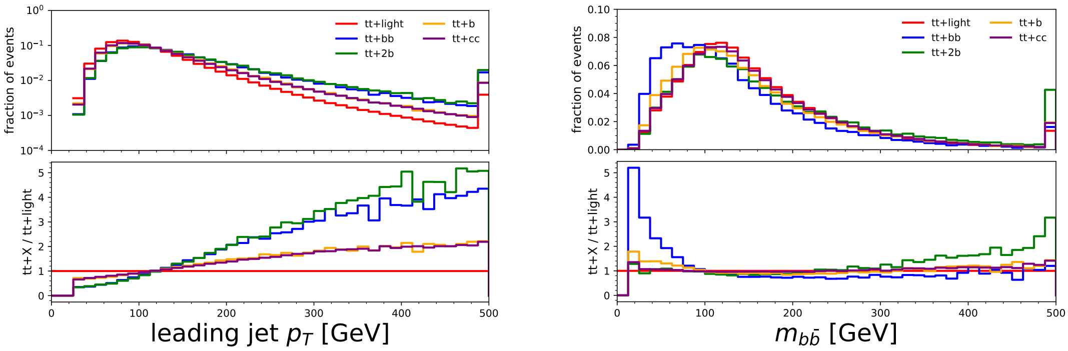
5.1.1 Modelling of $\mathrm{t\bar{t}}$+jets

The main background in the $\mathrm{t\bar{t}H}(\rightarrow \mathrm{b\bar{b}})$ search is the QCD production of $\mathrm{t\bar{t}}$+jets, for which we are using an NLO model based on `POWHEG`. This process is affected by considerable scale uncertainties at the level of 70-80% at LO and reduced to about 20-30% at NLO [162]. Furthermore, there are multiple scales present in the $\mathrm{t\bar{t}+b\bar{b}}$ process, with a large hierarchy between the scale set by the top quark mass and the mass of the bottom quark pair. Despite the available NLO QCD corrections, at high jet multiplicities, the differential distributions are modelled only at LO or parton shower accuracy, depending on the momenta and multiplicity of jets. Therefore, the residual uncertainties in $\mathrm{t\bar{t}}$+heavy flavour production are significant in terms of the differential distributions, with the differences of up to 100% between the `POWHEG` and `MG5_aMCNLO` MC models in variables sensitive to parton shower and quark mass effects [146]. In general, to have an effective treatment of high-multiplicity events in a precise $\mathrm{t\bar{t}H}(\rightarrow \mathrm{b\bar{b}})$ measurement, an improved theoretical understanding of the underlying QCD process for $\mathrm{pp} \rightarrow \mathrm{t\bar{t}b\bar{b}}$, along with possible interference effects with the signal process is necessary [163].

In the absence of an experimentally verified model that accurately describes these high-multiplicity and heavy flavour processes, we account for these modelling uncertainties by assigning conservative uncertainties on the theoretical modelling. We subdivide the $\mathrm{t\bar{t}}$+jets sample based on the generator-level flavour of additional jets, as the underlying processes that give rise to these different flavour categories are different. In particular, it has been shown that $\mathrm{t\bar{t}+b\bar{b}}$ can be significantly affected by the modelling of gluon splitting to collinear bottom quarks [164]. We distinguish between the following $\mathrm{t\bar{t}}$+jets sub-processes:

- $\mathrm{t\bar{t}+b\bar{b}}$, where two additional bottom jets are created from one or more b hadrons,
- $\mathrm{t\bar{t}+b}$, with only one additional bottom jet, which can arise from a process with two b hadrons, with one b hadron being out of acceptance,
- $\mathrm{t\bar{t}+2b}$, where jets from two b hadrons merge to produce one resolved bottom jet, which can arise from collinear gluon splitting,
- $\mathrm{t\bar{t}+c\bar{c}}$, if there are no additional bottom jets but at least one charm flavoured jet,
- $\mathrm{t\bar{t}+LF}$(light flavour), in case there are no bottom or charm flavoured jets.

The jet flavour is assigned using so-called *ghost clustering*, where simple geometrical matching between partons and jets is superseded by clustering the generator-level partons and hadrons



(a) Leading jet p_T

(b) The invariant mass of the closest b jet pair

Figure 5.2: The modelling of the leading jet transverse momentum (**a**) and the invariant mass of the closest b jet pair (**b**) for the various $\mathrm{t\bar{t}}$+jets sub-processes based on the $\mathrm{t\bar{t}}$+jets POWHEG simulation from events with at least two jets with $p_T > 30$ GeV, $|\eta| < 2.5$.

along with the detector-level jet constituents using standard jet algorithms. Information from the generator-level decay chain is used to assign the flavour of the jet according to the parton that gave rise to that jet [165], such that only hadrons that did not arise from the decay chain of a top quark are used in this categorisation. The aim of this splitting is to have experimental constraints on the uncertainties of the various $\mathrm{t\bar{t}}$ sub-processes separately, effectively relaxing some of the modelling assumptions. We illustrate the differences in the modelling of these processes in fig. 5.2 by comparing the distributions of leading jet transverse momentum (p_T) and the mass of the geometrically closest b jet pair $m_{\mathrm{b\bar{b}}}$, with significant differences between distributions visible for these processes.

5.2 Event reconstruction and object identification

We use particle flow to reconstruct events from particle candidates based on signals from all subdetectors, as described in section 2.2.7. This allows us to perform the analysis at the level of physics objects, namely, jets, charged leptons and MET from neutrinos, which arise from the decay of the top quark pair. A top quark decays almost exclusively through $\mathrm{t} \rightarrow \mathrm{W^+b}$, such that the leptonic or hadronic decay of the W-boson, which happen with a branching ratio of 33% or 67% respectively, determines the final state. In the case of the semileptonic decay, which happens with a total branching ratio of 44%, we expect one charged lepton ($\mathrm{e}^\pm, \mu^\pm$), MET from the corresponding neutrino, at least six jets, out of which four arise from bottom quarks and two from light quarks from the W-boson decay. In the dileptonic decay of the top quark pair, which happens with a branching ratio of $\simeq 11\%$, we expect two charged leptons of opposite sign, MET from neutrinos and at least four jets, all of which arise from bottom quarks. The fully hadronic decay of the top quark pair is treated in a separate analysis at CMS, as the background processes and required triggers are significantly different between the leptonic and hadronic channels. The hadronic analysis [166] makes use of a discriminator based on matrix elements between the $\mathrm{t\bar{t}H}$ signal and the multi-jet background that is similar to the method developed for the leptonic channel. The possible final state configurations are shown in fig. 5.3.

5.2.1 Charged leptons

The leptonic decay of at least one of the W-bosons from top decay is required in order to trigger the event at the HLT. In order to suppress leptons from the multi-jet QCD background, the charged leptons are required to be sufficiently isolated from hadronic activity using an isolation variable, which is computed within a cone of radius ΔR around the lepton direction (defined by

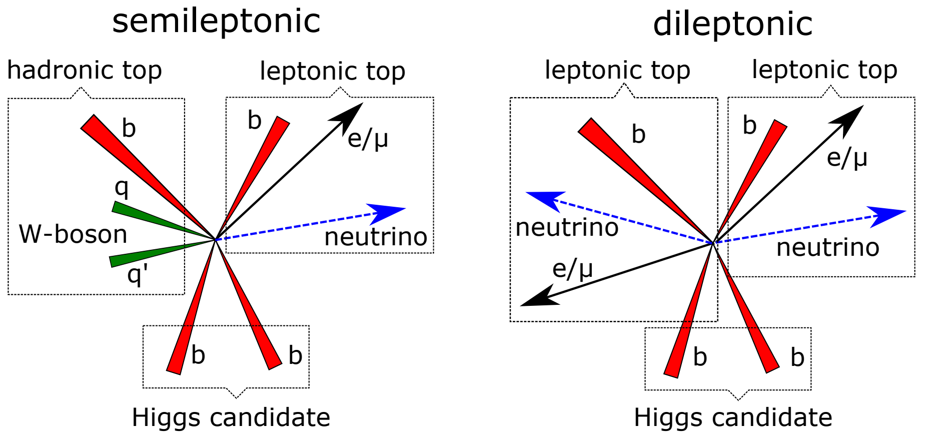


Figure 5.3: The semileptonic (left) and dileptonic (right) final states of $\mathrm{t\bar{t}H}(\to \mathrm{b\bar{b}})$.

the track) from the primary vertex as shown in eq. (5.1) for muons and eq. (5.2) for electrons. In order to compute the isolation, we sum over the transverse momenta of all particle candidates ($p_T^{c.h.}$ for charged hadrons, $E_T^{n.h.}$ for neutral hadrons, E_T^{γ} for photons) excluding the lepton itself and subtracting the neutral component from pileup events based on either the average pile-up energy (ρ) and effective area (A) for electrons or pile-up associated charged hadrons for muons [105]. The pre-factor 1/2 for the pile-up component for muons is used to account for the approximate charged-to-neutral fraction in the hadronisation of pile-up interactions [167].

Furthermore, in order to suppress leptons from non-prompt decays, we apply identification (ID) criteria based on various reconstruction parameters for the leptons. For muons, we apply the *tight ID*, which is a cut-based selection that suppresses decays in flight and is based on properties of the global track fit, number of hits in the pixel detector, tracker and muon chambers and sufficient proximity to the primary vertex [94, 168]. For electrons, the ID is based on a multivariate discriminator combining track-to-cluster matching observables, super cluster structure and cluster shapes [88, 169].

We summarise the lepton selection criteria in all the considered channels in section 5.2.1 and describe the event selection in terms of leptons further in section 5.3.1.

$$\mathrm{Iso}_{\mu} = \sum_{\Delta R<0.4} p_T^{c.h.} + \max\left(0, \sum_{\Delta R<0.4} [E_T^{n.h.} + E_T^{\gamma} - \frac{1}{2} p_T^{\mathrm{PU}}]\right) \tag{5.1}$$

$$\mathrm{Iso}_{\mathrm{e}} = \sum_{\Delta R<0.3} p_T^{c.h.} + \max\left(0, \sum_{\Delta R<0.3} [E_T^{n.h.} + E_T^{\gamma} - \rho A(\eta)]\right) \tag{5.2}$$

channel	trigger	offline p_T	$\|\eta\|$	isolation
$\mu^{\pm}$	$p_T > 24$ GeV	$p_T > 25$ GeV	$\|\eta\| < 2.1$	$\mathrm{Iso}/p_T < 0.15$
$\mathrm{e}^{\pm}$	$p_T > 27$ GeV	$p_T > 30$ GeV	$\|\eta\| < 2.1$	$\mathrm{Iso}/p_T < 0.15$
$\mathrm{e}^{\pm}\mathrm{e}^{\mp}$	$p_T > 23(12)$ GeV	$p_T > 25(15)$ GeV	$\|\eta\| < 2.1$	$\mathrm{Iso}/p_T < 0.15$
$\mathrm{e}^{\pm}\mu^{\mp}$	$p_T > 23_{\mathrm{e}}(8_{\mu})$ GeV	$p_T > 25(15)$ GeV	$\|\eta\| < 2.4$	$\mathrm{Iso}/p_T < 0.25_{\mu}(0.15_{\mathrm{e}})$
$\mu^{\pm}\mathrm{e}^{\mp}$	$p_{T,} > 23_{\mu}(12_{\mathrm{e}})$ GeV	$p_T > 25(15)$ GeV	$\|\eta\| < 2.4$	$\mathrm{Iso}/p_T < 0.25_{\mu}(0.15_{\mathrm{e}})$
$\mu^{\pm}\mu^{\mp}$	$p_T > 17(8)$ GeV	$p_T > 25(15)$ GeV	$\|\eta\| < 2.4$	$\mathrm{Iso}/p_T < 0.25$

Table 5.3: The selection and ID criteria for the charged leptons.

5.2.2 Jets

As the signal process is expected to produce between four to six jets in the leading order description and additional jets due to QCD radiation and pileup, an accurate reconstruction of jets is critical for this analysis. We use the anti-k_T clustering algorithm [41] in the FASTJET implementation [170] with a distance parameter $\Delta R = 0.4$ to reconstruct jets from particle flow candidates [171, 172, 173] and use the CMS particle flow (PF) jet ID algorithm, described in section 2.2.7, to reject reconstruction failures and noise. The noise rejection works on the basis of cuts on jet energy fractions from various types of PF candidates, namely muons, electrons, photons, charged hadron and neutral hadron candidates and has a noise rejection of around 99% [174].

Charged hadron subtraction

Charged particles from pileup interactions are removed from clustering via the process of charged hadron subtraction (CHS). As CHS relies on the reconstruction of tracks and the association of charged hadrons to tracks, the procedure is applied within the tracker volume ($|\eta| < 2.5$). We choose the leading primary vertex (PV) as the one that has the highest magnitude of total track transverse momentum squared ($\sum |p_T^{\text{track}}|^2$), with the rest of the PVs passing certain quality criteria as subleading PVs. Charged hadrons that are associated to tracks that are compatible with subleading PVs are removed. The subtraction procedure reduces the amount of jets arising from pileup from about 20% to 5% in the tracker region and also improves the momentum resolution and angular resolution ($\Delta R \simeq 0.01$) of jets [175].

Jet energy scale calibration

The experimentally measured energies of the jets have to be calibrated in terms of jet energy scale (JES) and resolution (JER) in both data and simulation. This is done using jet energy corrections (JEC), which correct for offset energy from pileup, the detector response based on simulation, the residual differences between data and simulation based on well-understood channels such as di-jet production, and the detector response to jet flavour.

The presence of pileup interactions generates a diffuse energy component that results in an energy offset in the jets. This offset correction is estimated using simulation by comparing jets in a MC sample without pileup events to the same simulation overlayed with pileup, geometrically matching them to the same underlying jet on the generator level [97]. An additional scale factor between data and simulation is extracted from zero-bias data using the random cone method [176].

The detector response is defined as the ratio between the reconstructed jet and a geometrically matched particle-level jet, averaged over a sample of jets: $R = \langle p_T \rangle / \langle p_{T,\text{particle}} \rangle$. It is estimated using a detailed model of the detector geometry, alignment and calibration, implemented in GEANT4 and evaluated in bins of particle-level jet momentum and reconstructed jet η. The corrections bring the response to a deviation of approximately 1% from unity based on simulations. A residual data to simulation correction scale factor is applied on data based on transverse momentum balance from dijet, Z/γ^*+jets and multi-jet events, with the momentum projection fraction (MPF) with respect to the missing transverse momentum being used as an alternative derivation for the corrections. These relative corrections rely on comparing the jet under calibration (probe) to a reference object (tag) and are of the order of a few percent in the central region considered in this analysis [176, 97].

The jet flavour response is estimated using simulation by taking the difference between the jet response for quark and gluon jets as predicted by Pythia 8 and HERWIG++. The magnitude of the flavour response is generally within a few percent, with differences between the flavours arising from fragmentation, where gluons fragment the most into soft particles that may remain unreconstructed and thus have the lowest response, and particle composition, with the neutral hadronic component having the largest effect. The flavour corrections are validated in Z+b jet data and the residual correction between data and simulation is found to be consistent with unity [176].

Jet energy resolution

In contrast to jet energy scale, which is known with a total uncertainty better than 3% over the relevant phase space, the jet energy resolution is known to around $10 - 20\%$. The resolution can be determined using p_T balance as for JES, but measuring the width instead of the mean of the response distribution. Both Z/γ^*+jet and dijet events are used to determine the JER response [176]. In the implementation of the MEM, we also use approximate jet resolution functions derived from simulation in order to account for detector effects in the phase space integral as explained in section 4.3.1.

5.2.3 B tagging

Since the $\mathrm{t\bar{t}H}(\rightarrow \mathrm{b\bar{b}})$ signal is characterised by the presence of four bottom quarks in the final state, with two arising from the top quarks and two from the Higgs boson decay, an accurate identification of b jets arising is important in this analysis. We rely on the combined secondary vertex algorithm (CSVv2) [102] to identify b jets. The CSVv2 algorithm uses secondary vertex properties such as the impact parameter along with track-based lifetime information to create a robust combined discriminator ξ optimised to distinguish between jets arising from bottom quarks and light quarks using supervised learning. In Run 2 of the LHC, the CSVv2 algorithm has been improved with a new vertex reconstruction algorithm, as well as using artificial neural networks instead of a likelihood method to combine the input variables, such that correlations between the inputs are taken into account, as described in chapter 3.

The threshold value of the b tagging discriminant, above which a jet is considered to be b tagged is chosen such that the efficiency of misidentifying jets arising from light quarks (u,d,s) or gluons as b jets would be around $\simeq 1\%$. This corresponds to an efficiency of $\simeq 70\%$ of correctly identifying bottom quarks and of around $\simeq 20\%$ for mis-identifying charm quarks. This threshold value is denoted the CSVv2 medium working point (CSVM) and an event where N jets pass this threshold contains N b tags.

B tagging likelihood

We further use the values of the per-jet b tagging discriminants $\boldsymbol{\xi} = [\xi_1, \xi_2, \ldots, \xi_{N_j}]$ with N_j being the number of jets to construct a per-event likelihood discriminator $\mathcal{BLR}(\boldsymbol{\xi})$ between the hypotheses that the event contained four bottom quarks (“4b”) or two bottom quarks (“2b”).

First, we use the probability density that the k-th jet has a discriminator value ξ_k assuming that it originated from a bottom quark (light quark), $f(\xi_k|\mathrm{b})$ $[f(\xi_k|\mathrm{l})]$, to define a likelihood for the given observed discriminator values $\boldsymbol{\xi}$ to result from M bottom quarks $\mathcal{BL}(\boldsymbol{\xi}|M\mathrm{b})$, as shown in eq. (5.3).

$$\mathcal{BL}(\boldsymbol{\xi}|M\mathrm{b}) = \sum_{i \in C_{M,N}} \left[\prod_{k \in \mathrm{b}_i} f(\xi_k|\mathrm{b}) \prod_{k \in \mathrm{l}_i} f(\xi_k|\mathrm{l}) \right] \tag{5.3}$$

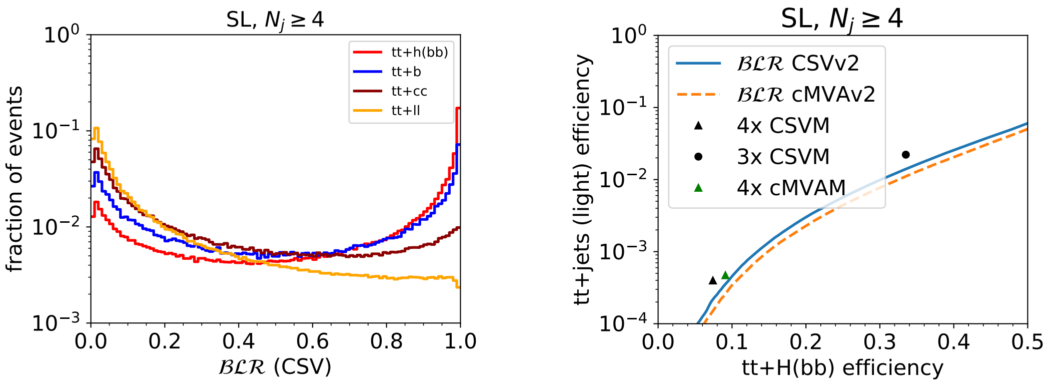
The sum in eq. (5.3) is performed over all the combinations $C_{M,N}$ of associating M jets out of N to bottom quarks and the rest to light quarks, such that b_i (l_i) refers to the M ($N-M$) jets associated to bottom quarks (light quarks) in the i-th combination. There are $\frac{N!}{(N-M)!M!}$ unique combinations to choose M jets to be b tagged out of N jets. For example, for $M=4$ and $N=6$, the combinations $C_{M,N}$ are formed by dividing up the jet indices $k = 1 \ldots 6$ between the b quark and light quark hypotheses, with one combination out of 15 being $b_i = \{1,2,3,4\}$, $l_i = \{5,6\}$.

The likelihoods for hypotheses with a particular number of bottom quarks are then used to construct a likelihood ratio discriminator $\mathcal{BLR}(\boldsymbol{\xi})$ (eq. (5.4)) that is optimised to suppress the $\mathrm{t\bar{t}}+$ LF background with two bottom quarks in favour of the $\mathrm{t\bar{t}H}(\rightarrow \mathrm{b\bar{b}})$ signal with four bottom quarks.

$$\mathcal{BLR}(\boldsymbol{\xi}) = \frac{\mathcal{BL}(\boldsymbol{\xi}|4\mathrm{b})}{\mathcal{BL}(\boldsymbol{\xi}|4\mathrm{b}) + \mathcal{BL}(\boldsymbol{\xi}|2\mathrm{b})} \tag{5.4}$$

We assess the performance of this discriminator in terms of $\mathrm{t\bar{t}H}(\rightarrow \mathrm{b\bar{b}})$ vs. $\mathrm{t\bar{t}}$+LF discrimination using simulation. In fig. 5.4, we see that the $\mathcal{BLR}$ discriminant improves over a fixed cut of ≥ 4 jets passing the CSVv2 medium working point ($N_{\mathrm{CSVM}} \geq 4$) by about 50% ($\epsilon_{\mathrm{t\bar{t}+LF}} = 0.04\% \rightarrow 0.022\%$) in terms of background rejection at the same signal efficiency ($\epsilon_{\mathrm{t\bar{t}H}(\rightarrow \mathrm{b\bar{b}})} \simeq 7\%$).

Furthermore, we have studied the efficiency of the $\mathcal{BLR}$ in correctly reducing the number of permutations in the MEM by assigning jets to be b tagged or untagged. For this, we evaluated the fraction of events where the final jets can be correctly matched to quarks from the hard interaction as a function of $\mathcal{BLR}$. We see in fig. 5.5 that the likelihood discriminator is positively correlated with the probability that all the quarks have been matched to jets, where around 50% of bottom



(a) Simulated shape of the discriminant.

(b) Expected performance of the discriminant.

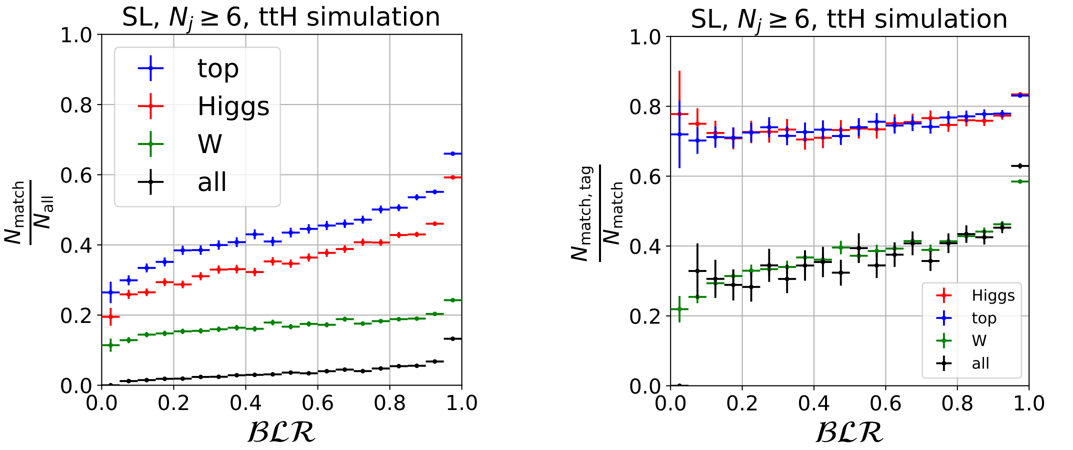
Figure 5.4: Simulated distribution and expected performance of $\mathcal{BLR}$ discriminant in the SL channel, requiring at least four good jets. On (**a**), we show the simulated shapes of the discriminant for signal ($\mathrm{t\bar{t}H}(\to \mathrm{b\bar{b}})$) and the various $\mathrm{t\bar{t}}$+jets backgrounds. On (**b**), we compare the efficiency to select $\mathrm{t\bar{t}H}(\to \mathrm{b\bar{b}})$ and $\mathrm{t\bar{t}}+$ LF events. We see that the $\mathcal{BLR}$ discriminant compares favourably to a fixed cut on number of b tags. The $\mathcal{BLR}$ discriminator defined with the cMVAv2 b tagger algorithm further improves the performance over the full range.

quarks from top decay, 40% of bottom quarks from Higgs decay and around 20% of the light quarks from the W boson decay have been reconstructed as jets at $\mathcal{BLR} \simeq 0.8$. Furthermore, we see a positive correlation between the likelihood discriminator and the probability that the highest-probability permutation in the sum in eq. (5.3) with the 4 bottom quark hypothesis corresponds to the bottom quarks from top or Higgs decay. In other words, the likelihood discriminator successfully tags the bottom quarks on an event-by-event basis.

The likelihood ratio as defined here ignores the differences in $f(\xi_k|\mathrm{b})$ and $f(\xi_k|\mathrm{c})$, meaning that in the case of $\mathrm{W} \to \mathrm{c\bar{s}}(\mathrm{\bar{d}})$ decays, the discriminator is suboptimal. We have investigated extending this likelihood to also account for the possibility of such decays by a straightforward extension to $\mathcal{BL}(\boldsymbol{\xi}|M\mathrm{b}\ 1\mathrm{c})$. However, we have found that the additional combinatorial complexity suppresses any increased discrimination power and further progress would likely require methods that are better able to deal with the combinatorial problem using additional information, for example jet kinematics. We use this b tagging likelihood ratio as a discriminator between the various $\mathrm{t\bar{t}}$+jets sub-processes, as well as to select the bottom quark candidates in the application of the MEM, as described in section 5.3.2. We have additionally studied whether switching to the new combined multivariate b tagging algorithm (cMVAv2) introduced in chapter 3 would improve the analysis. Since the irreducible $\mathrm{t\bar{t}} + \mathrm{b\bar{b}}$ background contains the same number of b jets as the $\mathrm{t\bar{t}H}(\to \mathrm{b\bar{b}})$ signal, improved b tagging can only reduce contribution from the background components with light or charm quarks, which are inherently less problematic than $\mathrm{t\bar{t}} + \mathrm{b\bar{b}}$. We find that although there is a small increase in the signal rate by switching the b discriminator, it is not sufficient to motivate a re-optimisation of the analysis at this stage.

B discriminator shape calibration

As we have used the detailed b discriminator shape information in constructing the b tagging likelihood ratio, we must experimentally calibrate the full range of this observable using data. This is accomplished by deriving a reweighting factor between data and simulation that depends on the jet b discriminator value, jet kinematics and flavour using a tag-and-probe method. In this approach, the tag jet is required to pass the medium operating point that has been described earlier and the discriminator distribution of the probe jet is corrected by reweighting the MC simulation. In order to extract the weight for the b jets, the procedure relies on dilepton $\mathrm{t\bar{t}}$+jets events with the contribution from light jets and backgrounds subtracted using simulation, whereas for the scale



(a) Fraction of events with correct matching.

(b) Fraction of matched events with correct tagging.

Figure 5.5: Estimation of the fraction of events where the bottom quarks from the top quark, the Higgs boson and the light quarks from the W boson are reconstructed as jets in the final state on (**a**). We study the event-level b tagging efficiency on (**b**), where we plot the fraction of events where the highest-likelihood permutation correctly assigned the bottom quarks to jets with respect to all events where the quarks were reconstructed as jets without considering tagging.

factor for light jets, Z+jets events are used. The procedure is iterative, as the scale factor for light jets is required for the extraction of the b jet scale factor and vice versa [177, 99]. The systematic uncertainties from this reweighting method are described in section 5.3.3.

Missing transverse energy

The leptonic decays of the W boson produce neutrinos, which are only partially reconstructed by the detector as MET, defined as the negative sum of all the momenta of the reconstructed particles in the transverse plane. In the SL channel, we can directly associate the MET with the transverse momentum of the neutrino through the modelling of the recoil as described in section 4.3.1 whereas in the DL channel, only the total momentum of both neutrinos is constrained by the MET.

5.3 Analysis

5.3.1 Event selection and categorisation

First, the large multi-jet QCD background is reduced to negligible levels by requiring that at least one of the top quarks in the $\mathrm{t\bar{t}H}(\to \mathrm{b\bar{b}})$ process decays leptonically. We divide events into two exclusive lepton categories: SL and DL, based on the multiplicity of the reconstructed charged leptons passing the quality cuts described in section 5.2.1. This is achieved by vetoing events with any additional leptons passing loosened quality criteria. We further suppress the γ^* background by requiring $m_{\ell\ell} > 20$ GeV and Z+jets in the DL categories and reject events around the resonant Z boson peak with 76 GeV $< m_{\ell\ell} <$ 106 GeV, where $m_{\ell\ell}$ is the invariant mass of the dilepton system. Furthermore, the leptons are required to have opposite charge. In the DL same-flavour categories, we require MET $>$ 40 GeV. We do not explicitly distinguish between cases where the top quark decays to τ leptons, although these events can still pass the selection in case the τ lepton decays leptonically and are considered as signal.

Events from $\mathrm{t\bar{t}H}(\to \mathrm{b\bar{b}})$ have a large number of jets and b tags compared to the V+jets backgrounds. Therefore, we require the presence of at least four jets passing the quality criteria (section 5.2.2), out of which at least three must be b tagged according to the medium working point

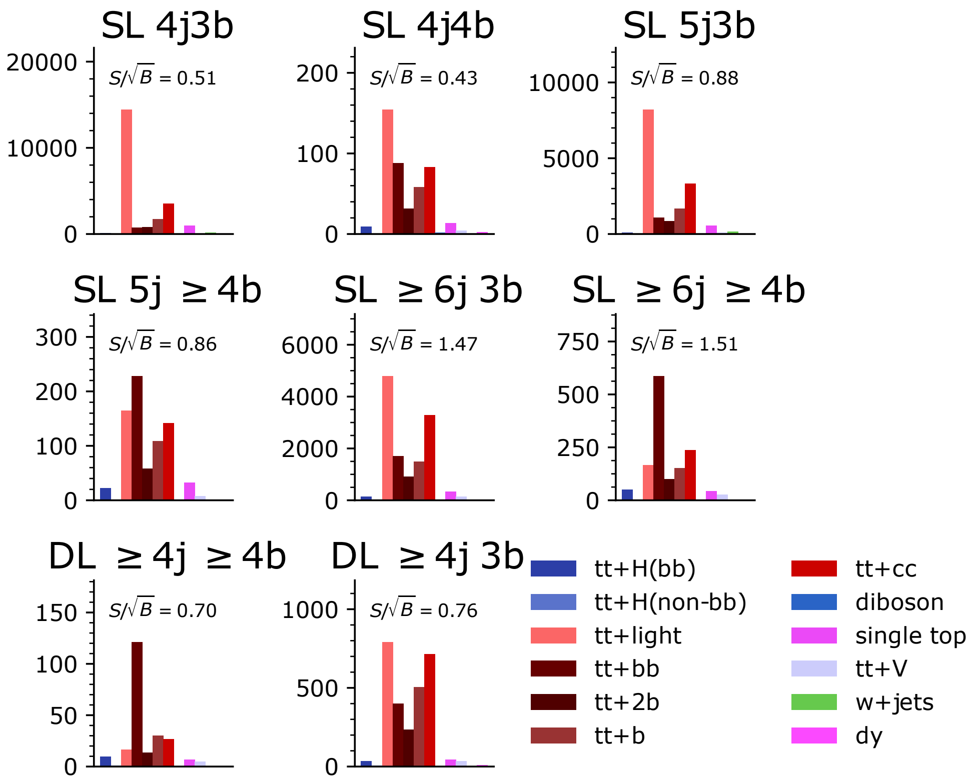


Figure 5.6: The expected signal and background yields in the semileptonic (SL) and dileptonic (DL) analysis categories. We also show the expected signal over background ratio $S/\sqrt{B}$. As can be seen, the expected signal yield is quite low compared to the predicted background, even in the most high-purity categories of SL ≥ 6 jets (j), ≥ 4 b tags (t) and DL ≥ 4 jets, ≥ 4 b tags.

(section 5.2.3). This brings us to the $\mathrm{t\bar{t}}$+jets dominated region, where we further distinguish between six categories in the SL channel

- $\geq$6 jets, $\geq$4 b tags; 5 jets, $\geq$4 b tags and 4 jets, 4 b tags, that are the most signal-enriched categories,

- $\geq$6 jets, 3 b tags; 5 jets, 3 b tags and 4 jets, 3 b tags, that contain a significant amount of $\mathrm{t\bar{t} + c\bar{c}}$ and $\mathrm{t\bar{t} + b\bar{b}}$,

and two categories in the DL channel

- $\geq$ 4 jets, $\geq$ 4 b tags,

- $\geq$ 4 jets, 3 b tags,

resulting in a total of eight exclusive categories, as shown in fig. 5.6.

We use the MEM as a $\mathrm{t\bar{t}H(\rightarrow b\bar{b})}$ to $\mathrm{t\bar{t} + b\bar{b}}$ discriminator in the categories with $\geq$ 4 b tags as these categories are enhanced in the signal fraction and have been shown to have an excellent discriminator performance for the MEM. The categories with three b tags are retained as control regions where we determine the $\mathrm{t\bar{t}}$+jets background rates by fitting the b tagging likelihood discriminator. We now turn to the description of the signal extraction using a template fit.

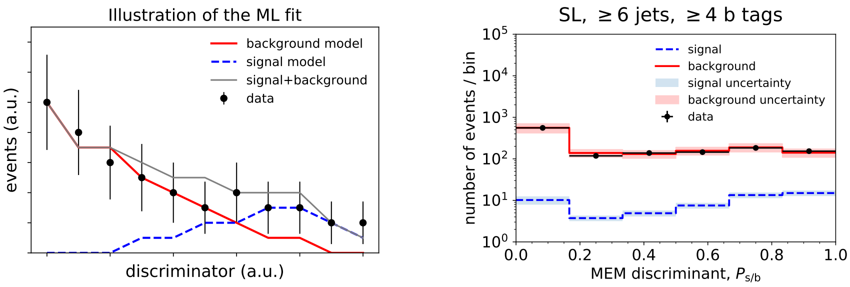


Figure 5.7: An illustration of the maximum likelihood template fit. On the left, we show a fictitious case of a signal and background model of a sensitive discriminator compared with data. The normalisations of the signal and background components are adjusted separately so as to maximise the likelihood of the combined signal plus background model giving rise to the data. On the right, we show a realistic but simplified case, where the MEM discriminator distribution is compared between the total signal and background processes, along with the data, in the most signal-rich semileptonic ≥ 6 jets, ≥ 4 b tags category. In the latter case, the overall signal and background yields are as predicted by MC simulation, hence the signal to background ratio is realistic and the discrimination between the background only model and the signal plus background model is much more difficult. We also show the approximate total systematic uncertainty on the signal and background distributions, neglecting the sources of uncertainty that do not change the discriminator shape. In practice, the fit is carried out simultaneously over all the analysis categories.

5.3.2 Signal extraction

The likelihood discriminant based on b tagging enhances the $\mathrm{t\bar{t}}$+heavy flavour component, but the cross-section of $\mathrm{t\bar{t}+b\bar{b}}$ is still an order of magnitude larger than that of $\mathrm{t\bar{t}H}(\to \mathrm{b\bar{b}})$. Thus, a simple determination of the signal cross-section or the upper limit on the signal strength modifier μ from the comparison of predicted and observed event counts does not result in significant sensitivity. Furthermore, we cannot directly reconstruct the resonant peak of the $\mathrm{H} \to \mathrm{b\bar{b}}$ decay as a natural discriminant between the signal and non-resonant background. Even though the width of the SM Higgs boson is relatively small compared to detector resolution ($\Gamma_{\mathrm{SM}} = 4.07 \times 10^{-3}$ GeV), the presence of multiple additional bottom quarks due to top decay in the final state creates a combinatorial self-background in the form of an ambiguity in choosing the candidate jets for the $\mathrm{H} \to \mathrm{b\bar{b}}$ decay. An estimator for the Higgs candidate invariant mass built from randomly selected jet pairs results in a much broader distribution compared to experimental resolution, whereas choosing the pair of jets that would give a mass closest to m_H would cause also the background to exhibit a signal-like peak.

In order to determine the signal strength modifier, we thus carry out a maximum likelihood template fit, where we rely on differences between the signal and background processes in sensitive differential cross-sections which are predicted using MC simulation in the form of one-dimensional histogram templates. As illustrated on fig. 5.7, we determine the normalisation scale factors for the signal and background processes such that the likelihood of the signal plus background model to give rise to the observed data distribution is maximised. Therefore, the sensitivity of the analysis is directly driven by the expected signal to background separation in these sensitive variables, as well as the susceptibility of these predictions to systematic uncertainties. The details of the fit are further described in section 5.3.9.

We use the MEM discriminant, introduced in chapter 4, to compute theory-motivated weights $P_{\mathrm{t\bar{t}H}(\to\mathrm{b\bar{b}})}$ and $P_{\mathrm{t\bar{t}+b\bar{b}}}$ for each candidate event. We construct a signal-to-background discriminant $P_{\mathrm{s/b}}$ based on the likelihood ratio of these weights, as described in section 4.4, which based on the Neyman-Pearson lemma, described in section 4.1, is the optimal test statistic between the signal and background hypotheses. As an improvement over the search for $\mathrm{t\bar{t}H}(\to \mathrm{b\bar{b}})$ performed by the CMS experiment in Run I [122], we use the MEM discriminant also in categories which are not fully reconstructed, but still contain a large amount of signal, namely five-jet and four-jet categories in the SL channel. The details of the additional MEM hypotheses are described in section 4.3.4. We list the discriminants that have been used in the different categories in table 5.4.

Table 5.4: The analysis categories and the discriminators used in those categories.

category	discriminant
SL ≥ 6 jets, ≥ 4 tags	MEM SL $2_{\mathrm{W}}2_{\mathrm{h}}2_{\mathrm{t}}$
SL ≥ 6 jets, 3 tags	The b tagging likelihood ratio
SL 5 jets, ≥ 4 tags	MEM SL $1_{\mathrm{W}}2_{\mathrm{h}}2_{\mathrm{t}}$
SL 5 jets, 3 tags	The b tagging likelihood ratio
SL 4 jets, ≥ 4 tags	MEM SL $0_{\mathrm{W}}2_{\mathrm{h}}2_{\mathrm{t}}$
SL 4 jets, 3 tags	The b tagging likelihood ratio
DL 4 jets, ≥ 4 tags	MEM DL $2_{\mathrm{h}}2_{\mathrm{t}}$
DL 4 jets, 3 tags	The b tagging likelihood ratio

We use the $\mathrm{t\bar{t}}+\mathrm{b\bar{b}}$ matrix element as a representative background diagram in all categories. This gives the best separation in the most signal-rich categories against $\mathrm{t\bar{t}}+\mathrm{b\bar{b}}$ and is further motivated by simulation, where we see that using this process as background still achieves a high rate of separation in categories enriched in other $\mathrm{t\bar{t}}$+jets sub-processes. As an optimisation, considering additional background hypotheses in different categories in the future is expected to improve the signal-to-background discrimination at the cost of computational complexity.

As the b tagging likelihood ratio method is optimised to identify the set of jets that are most compatible with arising from four bottom quarks, we further augment the MEM by assuming that the bottom quarks need to be considered only among those four jets, as explained in section 4.3.4. This means that we have exactly four candidates for the bottom quarks from $\mathrm{H} \to \mathrm{b\bar{b}}$ and $\mathrm{t} \to \mathrm{Wb}$ decay, whereas the remaining jets are assumed to arise from $\mathrm{W} \to \mathrm{qq'}$ or from unspecified sources.

Both of these discriminators are complex multivariate functions based on quantities which have significant modelling uncertainties affecting the shapes of the distributions. Therefore, a realistic description and propagation of the systematic uncertainties is crucial in the interpretation of data.

5.3.3 Systematic uncertainties

Among the experimental uncertainties, the dominant ones are uncertainties on the jet energy scale and resolution corrections (section 5.3.4) and the reweighting of the b tagging discriminant (section 5.3.5). Both of these can affect the predicted yields of all the processes, since they change the selection efficiency, as well as the shapes of the final discriminants. In case the source of an uncertainty is the same across several categories, the nuisance parameters associated with the uncertainties are treated as fully correlated, otherwise, the uncertainties are treated as uncorrelated in the fit.

Concerning theory systematics, the uncertainties on the modelling of the $\mathrm{t\bar{t}}$+jets background are the ones with the largest impact on the measurement. As introduced in section 5.1.1, the $\mathrm{t\bar{t}}$+jets `POWHEG` model we currently use in the analysis treats the $\mathrm{t\bar{t}} + \mathrm{b\bar{b}}$ process only at leading order accuracy, where the $\mathrm{b\bar{b}}$ pair is generated from gluon splittings using a parton shower, such that we have considerable additional theoretical uncertainties on the modelling of $\mathrm{t\bar{t}} + \mathrm{b\bar{b}}$, as will be described in section 5.3.7.

We give a detailed overview of the most important experimental and theoretical uncertainties along with their estimation in the next sections. The full list of systematic uncertainties along with the assumed priors is shown in table 5.5.

5.3.4 Jet energy correction uncertainties

We apply jet energy scale (JES) and resolution (JER) corrections between data and simulation, as described in section 5.2.2. Thus, we need to understand the effect of the uncertainties on these corrections. In Run 2, we consider various uncorrelated sources of jet energy scale correction uncertainties with their corresponding correlations, as opposed to a single bulk JES uncertainty as

uncertainty	normalisation	shape	processes	prior
JES (26 sources)	yes	yes	all	Gaussian, $< 5\%$
JER	yes	yes	all	Gaussian, $< 1\%$
b tagging (9 sources)	yes	yes	all	Gaussian, $0 - 20\%$
pileup	yes	yes	all	Gaussian, $0 - 5\%$
lepton ID	yes	yes	all	Gaussian, $\simeq 1\%$
lepton isolation	yes	yes	all	Gaussian, $\simeq 1\%$
luminosity	yes	no	all	log-normal, 2.4%
limited MC statistics	no	yes	all	bin-by-bin Poisson
$\mathrm{t\bar{t}}$+jets ISR, FSR	yes	partly	$\mathrm{t\bar{t}}$+jets	log-normal, 0-15%
$\mathrm{t\bar{t}}$+jets tune, h_{damp}	yes	partly	$\mathrm{t\bar{t}}$+jets	log-normal, 0-15%
$\mathrm{t\bar{t}+b\bar{b}}$ norm.	yes	no	$\mathrm{t\bar{t}}$+jets	log-normal, 50%
$\mathrm{t\bar{t}+b}$ norm.	yes	no	$\mathrm{t\bar{t}}$+jets	log-normal, 50%
$\mathrm{t\bar{t}+2b}$ norm.	yes	no	$\mathrm{t\bar{t}}$+jets	log-normal, 50%
$\mathrm{t\bar{t}+c\bar{c}}$ norm.	yes	no	$\mathrm{t\bar{t}}$+jets	log-normal, 50%
PDF (gg)	yes	no	$\mathrm{t\bar{t}H}$, $\mathrm{t\bar{t}}$+jets	log-normal, 4%
PDF (qq')	yes	no	W+jets	log-normal, 2%
PDF (qg)	yes	no	single top	log-normal, 3%
μ_R, μ_F scale	yes	yes	$\mathrm{t\bar{t}}$+jets	Gaussian
scale uncertainties in norm.	yes	no	primarily $\mathrm{t\bar{t}H}$	log-normal

Table 5.5: Systematic uncertainties in the $\mathrm{t\bar{t}H}(\rightarrow \mathrm{b\bar{b}})$ analysis. We list whether a particular source of uncertainty changes the predicted normalisation of processes and whether it has an effect on the discriminator distributions, i.e. the shapes. For each uncertainty, we also show for illustrative purposes the size of the prior uncertainty in a relevant variable, e.g. jet momentum spectrum for JES, and the distributions used to model the nuisance parameters associated to each uncertainty. We further distinguish between experimental uncertainties, which affect all processes, and theoretical uncertainties that are specific to particular processes. Although the $\mathrm{t\bar{t}}$+jets ISR, FSR, MC tune and h_{damp} modelling uncertainties affect also the discriminator distributions, due to limited MC simulation statistics, we only account for the normalisation effect as a function of jet multiplicity, as further described in section 5.3.7, therefore these uncertainties are listed as having a partial effect on the shape.

was done for this analysis in Run I. This significant advancement has resulted from an improved modelling of the detector performance and better calibration techniques developed with more data. By treating the various sources of JES uncertainties independently, the assumptions that allow the profile likelihood fit to constrain the combined JES uncertainty significantly in Run I are thus relaxed and the final uncertainty is a more realistic estimate of the true uncertainty.

The magnitude and correlation of the uncertainties on JES and JER are determined in a dedicated CMS analysis and are provided as a vector of per-jet corrections with p_T and η dependent correlations [97]. The most important groups of correction uncertainties are the following:

- Pileup offset, which results from extra energy deposited in jets from additional pp interactions within the same bunch crossing (in-time pileup) or due to the finite signal decay time in the calorimeters (out-of-time pileup). The uncertainty for this source results from the η-dependent scale factor used to correct the offset distribution measured in simulation.

- Relative η-dependent corrections, which calibrate the forward regions of the detector with respect to the central region. Uncertainties of this type arise from jet energy resolution and from the modelling of initial and final state radiation (ISR+FSR).

- Uncertainties on the absolute energy scale, which are derived using Z/γ+jet and multijet data. The energy scale uncertainties are driven by the muon momentum scale and the single pion response in the HCAL. Furthermore, the uncertainties in fragmentation are assessed in a comparison of the `PYTHIA` and `HERWIG++` MC models.

- Uncertainties in the modelling of the detector response for jet flavour, which are assessed using simulation and are largest for gluon jets.

- Finally, due to radiation damage, there is a residual time-dependent uncertainty in the scale corrections, which is estimated using dijet events in different run periods.

These five broad groups factorise into approximately 26 independent sources. In order to account for the JES scale uncertainties in the analysis, we propagate the uncertainties in jet energy scale and resolution corrections to the jet momenta and all the event-level observables that are derived from them, such as the jet and b tag multiplicities, by changing the jet energy scales and resolutions by one standard deviation up and down from the nominal values. This is done separately for all the sources so that we can fully account for the correlations between the various sources. Thus, we are able to account for both the changes in efficiency (normalisation) and discriminator shape in the final analysis categories. We find that the normalisation effects are of the order of 0-4% for all JES and JER sources for the SL channel and around 0-1% in the DL channel, with the largest variations resulting from the jet flavour response modelling.

In order to propagate the uncertainty to high-level multivariate observables such as the MEM discriminator, they need to be recomputed using the varied jets. We use the approximate MEM vector integration technique for this as described in section 4.3.8. In addition to uncertainties in the jet energy scale corrections, we also consider uncertainties on the jet energy resolution corrections. In the tracker acceptance region, the relative JER uncertainty is around 2-4%, depending on the jet pseudorapidity $|\eta|$. The JER uncertainty is propagated by shifting the JER scale factor up and down by one standard deviation corresponding to the uncertainty, thus it is fully deterministic. The overall effect of the JES and JER uncertainties on the leading jet transverse momentum distribution is shown in fig. 5.8. These uncertainties will mostly affect the predicted yields in our final analysis categories.

5.3.5 b tagging systematic uncertainties

Due to the high expected number (≥ 4 in the signal regions) of b tagged jets in the $\mathrm{t\bar{t}H}(\rightarrow \mathrm{b\bar{b}})$ final state, this analysis is sensitive to uncertainties in b tagging, which arise from the MC modelling of the input variables used in the construction of the multivariate b tagging discriminator. As we have described in section 5.2.3, we correct for mis-modelling in the b tagging discriminator shape using a tag and probe method, such that the detailed b discriminator shape can be used in further template fitting. The effect of this shape correction is shown in fig. 5.9, where we see that the correction improves the description by the MC, with the corrected distribution agreeing with data within the systematic uncertainties.

The uncertainties of this correction include the propagation of jet energy scale uncertainty, which affects the determination of the correction through changes in efficiency and the discriminator value. Furthermore, simulation is used to subtract the non-relevant jet flavour components in determining the scale factor for bottom (light) jets. For the scale factor for light jets, the fraction of bottom (charm) jets is varied within 20% of the MC prediction in the Z+jets simulation used to determine the scale factors. Similarly, for the extraction of the b jet scale factor, the light flavour component in the $\mathrm{t\bar{t}}$+jets dileptonic sample arises from additional radiation and is estimated to be $\simeq 20\%$ [99].

As the b discriminator scale factor is determined in bins of the discriminator value, statistical fluctuations in bins with a low number of data and simulated events introduce an uncertainty on the final scale factor. This uncertainty is only significant in case the size of the fluctuations varies systematically over the b discriminant range. Since the discriminator has a roughly monotonous increasing (decreasing) shape for b jets (light jets), this condition is fulfilled. The statistical uncertainties are accounted for by a sum of polynomials of first and second order, where the nuisance parameter is the overall scale of the distortion.

There is currently no dedicated b discriminator scale factor for charm jets, therefore, the uncertainties on the charm flavour scale factor are assumed to be twice as large as for the b jet scale factor, with the scale factor being unity. We propagate the uncertainties from b tagging in the form of a set of per-event weights, which are determined from the individual per-jet weights that are used to correct the jet b discriminator distributions. The uncertainties on the b discriminator scale factor result in both normalisation effects due to acceptance changes, which can be up to 20% in some cases, and shape effects on the templates. An example of the effect of b discriminator uncertainties is shown in fig. 5.10.

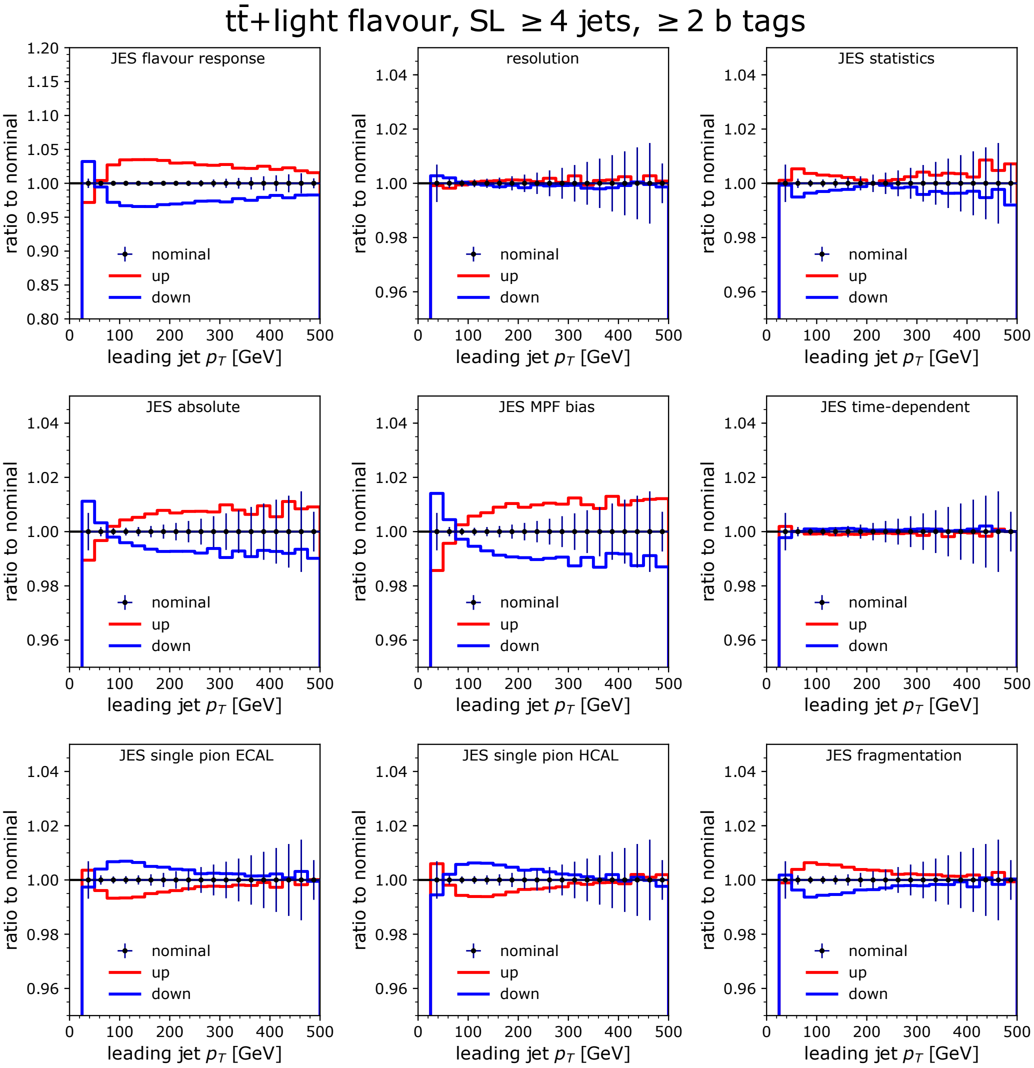


Figure 5.8: The effect of jet energy corrections on the leading jet transverse momentum distribution. In the top row, we show the distribution under variations of the flavour composition (left), jet energy resolution (middle) and the statistical uncertainties in the absolute scale determination (right). In the middle row, we show the overall uncertainty from the absolute scale variation (left), the uncertainty arising from the corrections derived using the missing momentum projection fraction (MPF, middle) and the time-dependent momentum scale variation (right). On the bottom row, we show the uncertainties in the single pion response in the ECAL (left), HCAL (middle) and the fragmentation model (right). In general, we see that the variations are within a few percent of the nominal, with the largest effect from the uncertainties on the flavour response (top left).

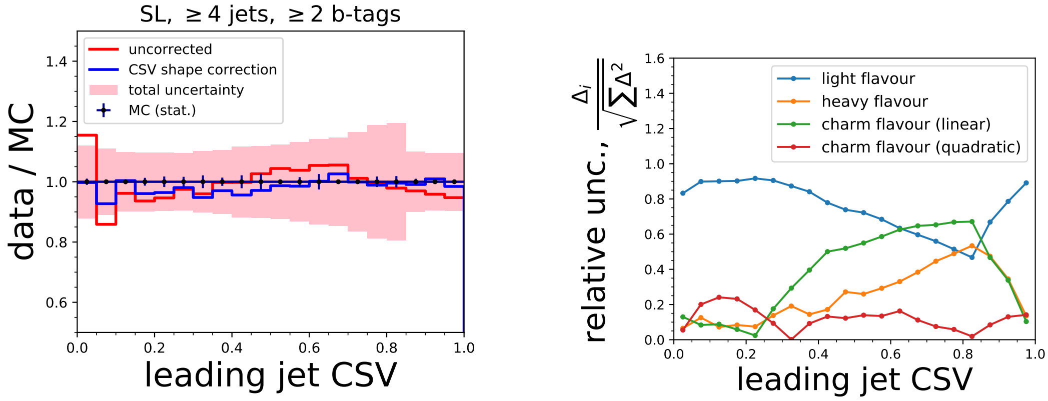


Figure 5.9: The effect of the b tag reweighting on the MC modelling (left). We see that the b tag reweighting technique improves the modelling from the nominal case (red) to the corrected case (blue), with the total uncertainties (pink) over-covering the difference. On the right plot, we see that the overall uncertainty band consists mainly of the light flavour (blue), charm flavour (green) and heavy flavour (orange) uncertainty components.

5.3.6 Other experimental systematic uncertainties

We also assess the effect of uncertainties in the lepton identification, isolation and trigger selection, which may have different efficiencies in data and simulation and are thus corrected using scale factors. For muons, we assign a 1% normalisation uncertainty for the lepton ID, 1% for isolation and 0.5% for the effect of highly-ionising particles, on top of the statistical uncertainties on the muon scale factor [178]. For electrons, we use p_T and ECAL supercluster η-dependent scale factor uncertainties derived using a tag-and-probe method, which are generally below 1% [179].

As the pileup profile in simulation is corrected to data using a pileup-dependent scale factor, we estimate the uncertainty in the pileup correction by varying the minimum bias cross section from $\sigma = 69.2$ mb by 4.6%, corresponding to the uncertainty in the number of interactions in minimum bias events from luminosity and cross section[180]. This results in both a normalisation and shape effect in the predicted templates.

Furthermore, the uncertainty on the total integrated luminosity is estimated to be 2.4% using cluster counting in the pixel detector and affects the normalisation of all processes in a correlated way [181, 182].

5.3.7 Theoretical uncertainties

The most important theoretical uncertainties arise from the modelling of the $\mathrm{t\bar{t}}$+heavy flavour processes, namely $\mathrm{t\bar{t}+b\bar{b}}$, $\mathrm{t\bar{t}+2b}$, $\mathrm{t\bar{t}+b}$ and $\mathrm{t\bar{t}+c\bar{c}}$. Currently, it is not possible to isolate a pure $\mathrm{t\bar{t}+b\bar{b}}$ control region which would not contain a significant amount of $\mathrm{t\bar{t}H(\to b\bar{b})}$ and thus this background cannot be determined directly from data. Although inclusive measurements of the $\mathrm{t\bar{t}+b\bar{b}}$ cross-section have been carried out in CMS [183] with $\sigma_{\mathrm{t\bar{t}+b\bar{b}}}$ determined with a $\simeq 35\%$ relative accuracy, these analyses treat $\mathrm{t\bar{t}H(\to b\bar{b})}$ as an irreducible background and thus cannot directly be used to set the prior uncertainties in our analysis. On the other hand, NLO estimates for the inclusive cross-section of $\sigma_{\mathrm{t\bar{t}+b\bar{b}}}$ have residual theoretical uncertainties at the level of $\simeq 20\%$ [184], with important differences between alternative models in the differential predictions. Therefore, we heuristically assign a conservative 50% normalisation uncertainty on all the $\mathrm{t\bar{t}}$+heavy flavour processes, which is assumed to be uncorrelated across the aforementioned sub-processes. Correlating these uncertainties would reduce their overall impact. Effectively, this allows us to use the data in the control regions to determine the best fit values of the cross-sections for the $\mathrm{t\bar{t}}$+heavy flavour processes in a consistent way with the extraction of the signal strength modifier. We do not use additional extrapolation uncertainties for the $\mathrm{t\bar{t}}$+jets processes between

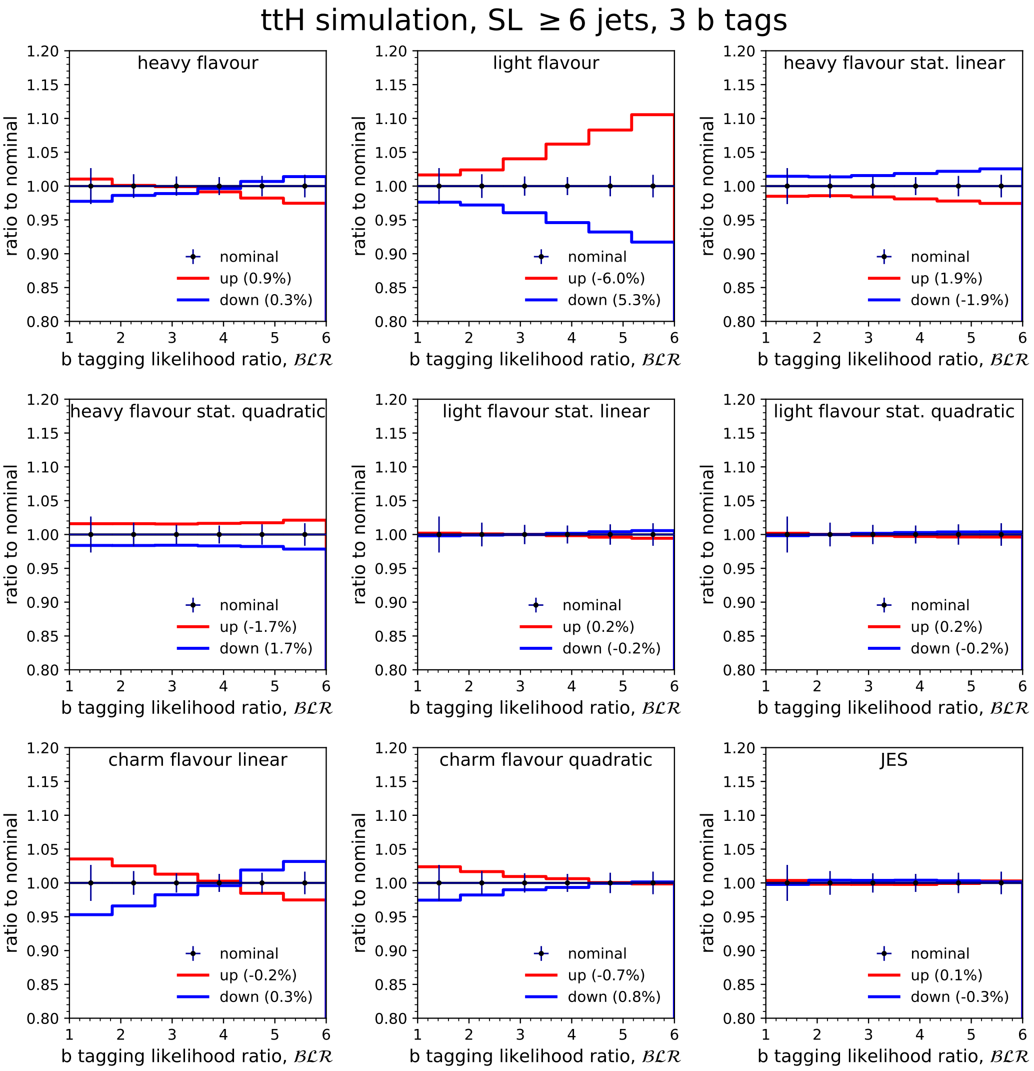


Figure 5.10: The effect of b tagging uncertainties on the b tagging likelihood ratio distribution for the $\mathrm{t\bar{t}H}(\rightarrow \mathrm{b\bar{b}})$ sample. In the top row, we show the effect of the heavy flavour (left), light flavour (center) and the linear heavy flavour distortion from statistical uncertainties (right). In the middle row, we show the effect of the quadratic distortion on the heavy flavour scale factor from statistical uncertainties (left) and the linear and quadratic uncertainties on the light flavour scale factor (middle, right). In the bottom row, we show the uncertainties for the charm flavour jets (left, middle) and the uncertainty on the scale factor arising from the propagation of JES uncertainties. These templates are derived in the SL ≥ 6 jet, three b tag category on $\mathrm{t\bar{t}H}$ simulation.

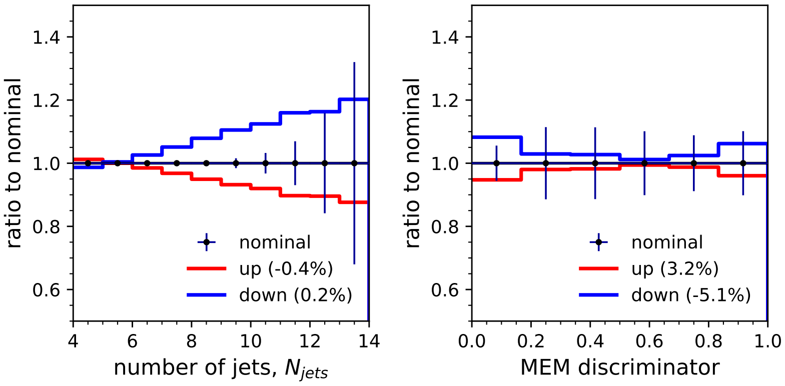


Figure 5.11: The effect of the renormalisation and factorisation scale variations (μ_r and μ_f) on the modelling of the jet multiplicity (left) in the SL $\geq$ 4 jet, $\geq$ 2 b tag inclusive region and the MEM discriminator (right) in the SL $\geq$ 6 jet, $\geq$ 4 b tag signal region. While the effect of the scale changes is normalised to be shape-changing in the inclusive region, it can introduce migrations between jet-tag bins among the final categories. These distributions are derived using $t\bar{t}$ + LF simulation. We also show the relative change in normalisation in the particular category resulting from the up and down variations.

the control and signal regions, as the combined fit provides a consistent framework for determining the best-fit estimates given the data.

The inclusive cross sections of all involved signal and background processes are known to at least NLO accuracy, with a 4% renormalisation and factorisation scale uncertainty and a 4% PDF uncertainty on the gluon-gluon dominated production of $t\bar{t}$ +jets. Shape uncertainties from PDF variations are found to be negligible and thus not considered further in the analysis. We use MC simulation to estimate the shape effect of the renormalisation and factorisation scale (μ_R and μ_F) on the final discriminant shape by changing the nominal values of μ_R and μ_F by 0.5 (2.0) for the down (up) variation. This is achieved using the embedded matrix-element dependent weights in the MC simulation. The effect of these variations is illustrated in fig. 5.11 and is generally small on the final observables, but has a significant effect on the jet multiplicity and transverse momentum distributions.

For the parton shower uncertainties, in particular the effects of ISR and FSR, we have limited MC simulation samples that can only be used to determine the overall effect on normalisation, whereas the shape distortions on the distributions are generally consistent with no change. These background modelling uncertainties primarily affect jet kinematics and thus the number of reconstructed jets in the final state. Therefore, we model these uncertainties through per-subprocess normalisation factors that depends on the jet multiplicity, with the magnitude of the uncertainties generally between 5-15%, as seen in fig. 5.12. The overall normalisation and shape effect of the most important shape changing uncertainties is shown in fig. 5.13.

5.3.8 Control regions

We validate the MC simulation in the inclusive semileptonic and dileptonic control regions with at least four jets, out of which at least two must be b tagged by comparing the simulated distributions of jet and lepton kinematic variables to data. The distributions in the semileptonic control region can be seen in fig. 5.14 and for the dileptonic in fig. 5.15. In general, we see that both the inclusive yields and differential distributions for the kinematic variables are well-described within the systematic uncertainties. We observe a residual mismodelling in the jet multiplicity distribution, which we attribute to uncertainties in the modelling of the parton shower and the $t\bar{t}$+jets MC tuning.

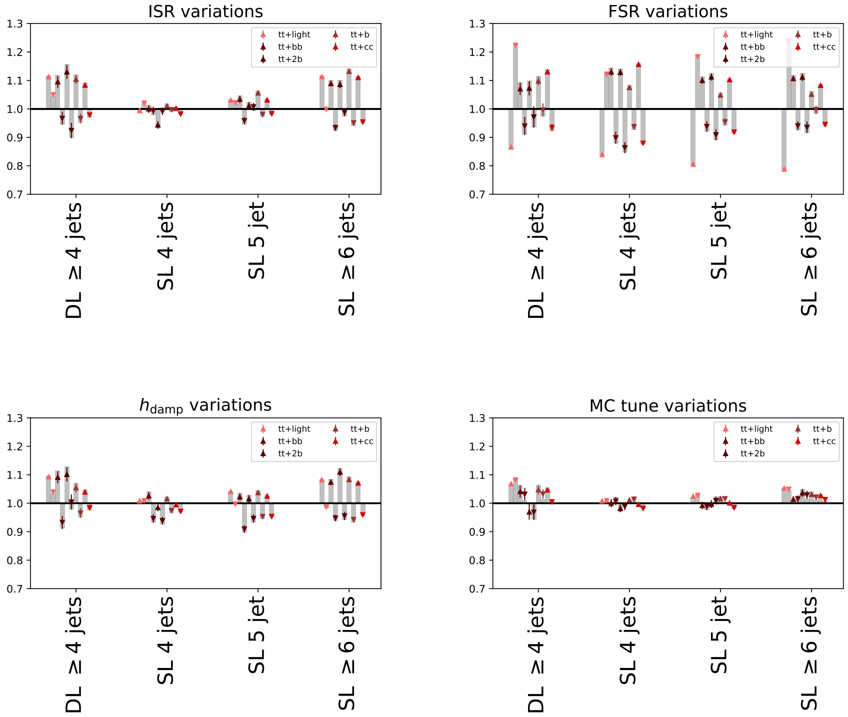


Figure 5.12: The $\mathrm{t\bar{t}}$+jets ISR, FSR, h_{damp} and MC tune uncertainties in terms of a scale factor that depends on jet multiplicity. We extract this scale factor by comparing the yield predicted by the MC simulation with varied parameter values to the nominal, adding the statistical uncertainty on this prediction. Generally, these scale factors are symmetric around the nominal and change the yield by $< 15\%$, with the most significant effects on the $\mathrm{t\bar{t}} + \mathrm{LF}$ process.

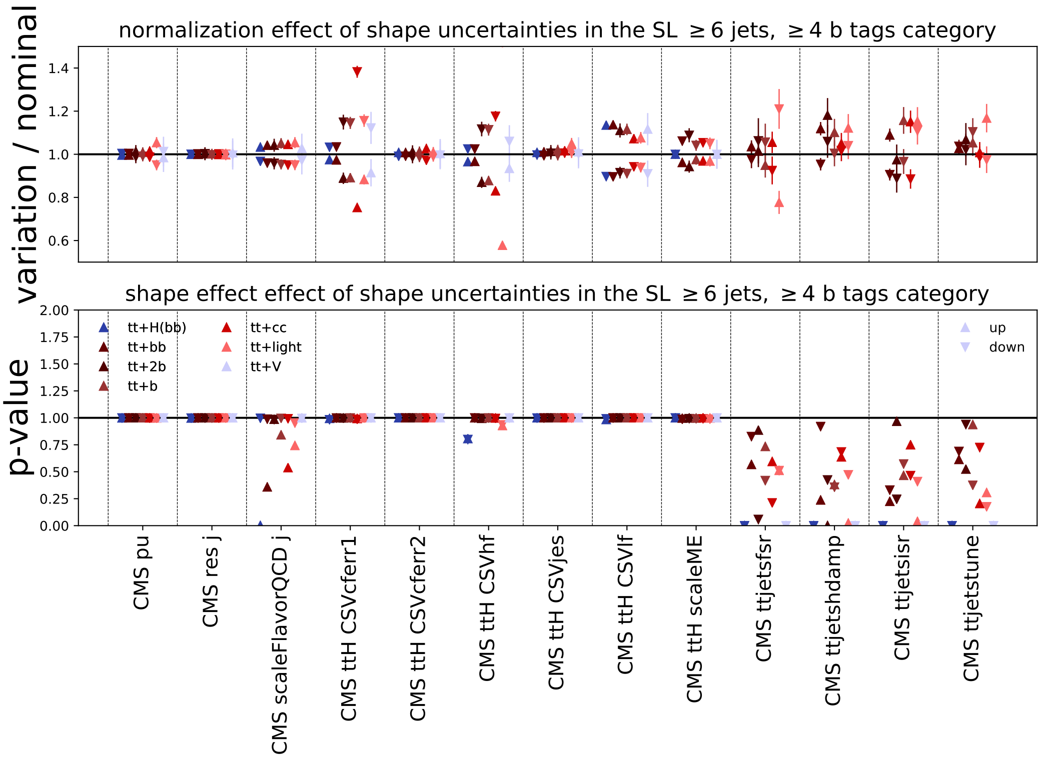


Figure 5.13: The normalisation and shape effect of the most important shape-changing uncertainties. In the top figure, we show the effect on the normalisation in terms of the ratio between the varied and nominal predictions. In the figure in the bottom, we show the estimated effect on the shape of the template by computing the p-value from the χ^2 test between the two histograms. We see that generally the effect of shape variations is symmetric, with a yield change that is under 20%. Most shape changes are consistent with no change, apart from the JES flavour uncertainty and the b tagging heavy flavour uncertainty. The ISR, FSR, h_{damp} and MC tune uncertainties do not have enough simulation statistics to determine the presence of shape changes. These uncertainties are verified using MC simulation in the semileptonic category with at least six jets, out of which four must be b tagged.

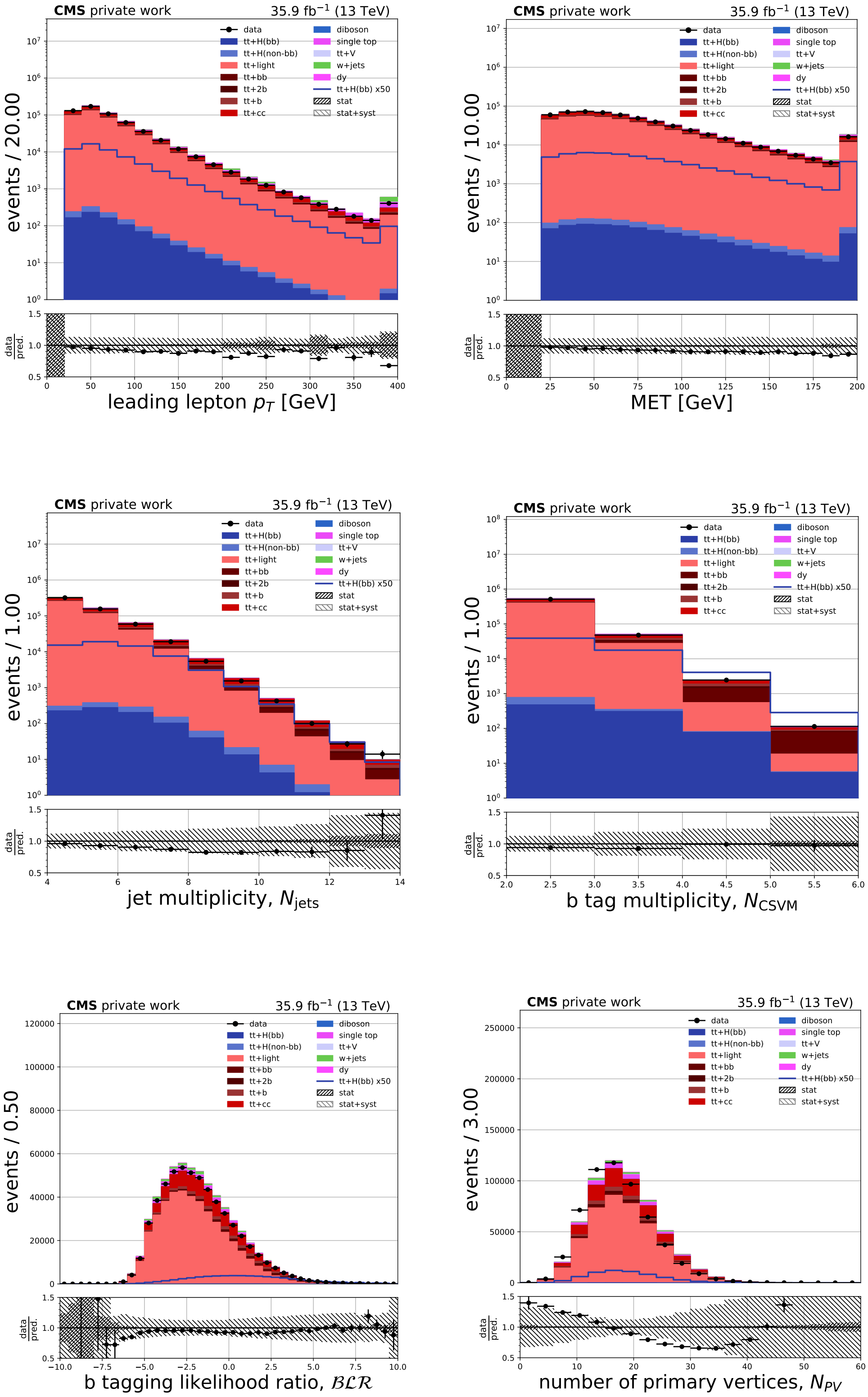


Figure 5.14: The modelling of the most important analysis variables in the semileptonic control region with at least four jets, out of which two must be b tagged.

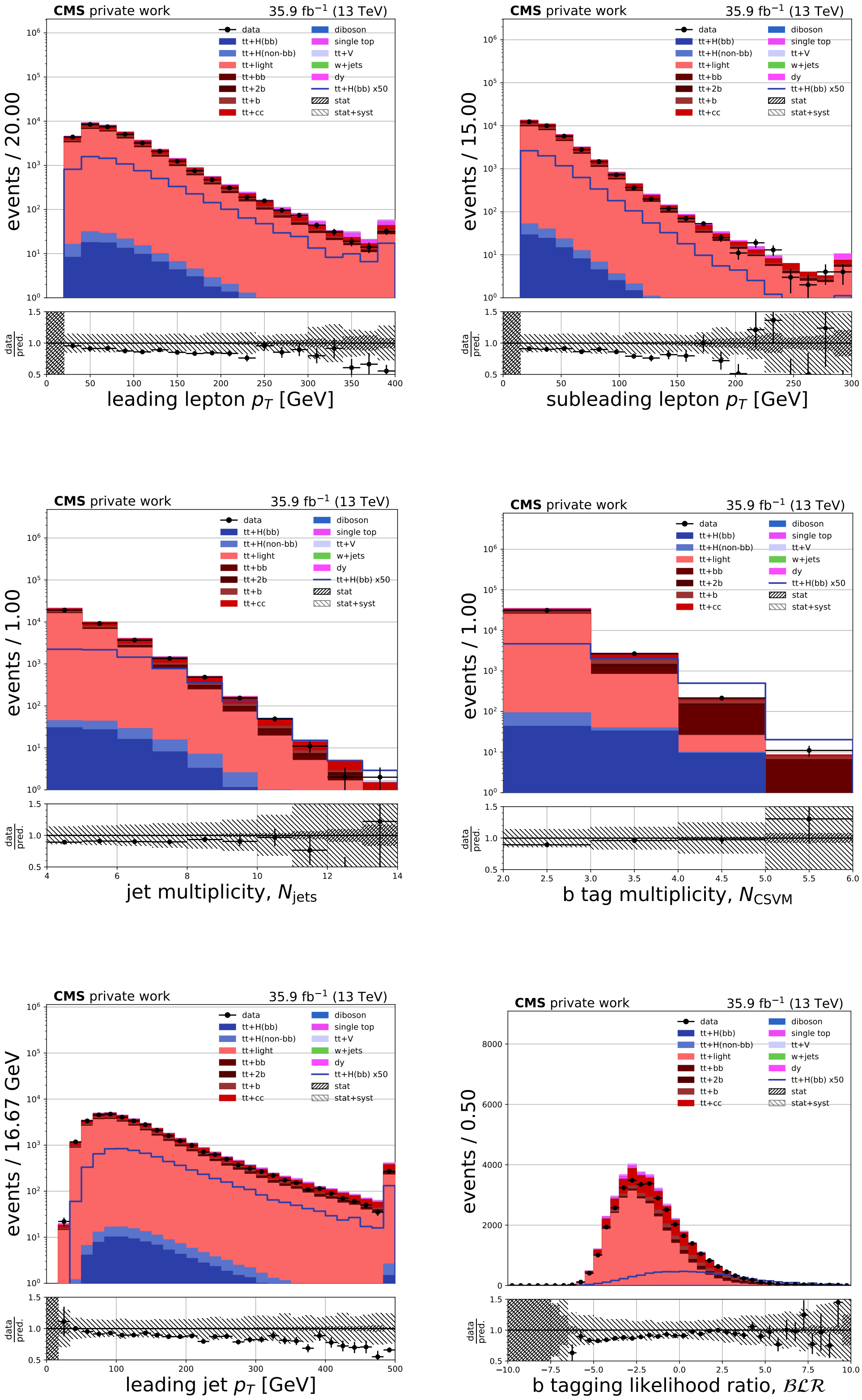


Figure 5.15: The modelling of the most important analysis variables in the dileptonic control region with at least four jets, out of which two must be b tagged.

5.3.9 Statistical method

In order to interpret the data, we use the same statistical framework as has been used for other Higgs boson searches in the CMS collaboration [46, 185, 186]. We wish to measure the signal strength modifier $\mu = \sigma_{\mathrm{t\bar{t}H}}/\sigma^{\mathrm{SM}}_{\mathrm{t\bar{t}H}}$ and in the absence of an observed signal, exclude $\mu \geq \mu^{CL}$ at a certain confidence level. The null hypothesis (H_0) is therefore the presence of a signal with a given μ, whereas the alternative hypothesis is no signal ($H_1, \mu = 0$). Based on the data, we seek to exclude the null hypothesis above a certain μ.

The predicted distributions for both signal (denoted as s) and background (denoted as b) are subject to uncertainties introduced in section 5.3.3 such that the expectations are functions of the nuisance parameters θ through $s(\theta)$ and $b(\theta)$. The uncertainties are assumed to be either fully correlated or uncorrelated, as is more appropriate and conservative, which allows the likelihood function to be written in a factorised form.

To determine confidence intervals on the Higgs boson production cross section and thus quantify the presence or absence of a signal, we use the CL_s method [187, 188], which defines the likelihood function $\mathcal{L}(\mathrm{data}|\mu,\theta)$ as

$$\begin{aligned}\mathcal{L}(\mathrm{data}|\mu,\theta) =& \mathrm{Poisson}[\mathrm{data}|\mu\cdot s(\theta)+b(\theta)]\cdot p(\tilde{\theta}|\theta) \qquad (5.5)\\ =& \prod_{i\in\mathrm{bins}} \frac{(\mu s_i+b_i)^{n_i}}{n_i!}\exp\left[-(\mu s_i+b_i)\right]\cdot p(\tilde{\theta}|\theta). \qquad (5.6)\end{aligned}$$

We have used Poisson probabilities to model the observation of n_i events in the bin i of a discretised distribution, given an expectation $\mu s_i + b_i$. The distribution $p(\tilde{\theta}|\theta)$ encodes the prior knowledge on the nuisance parameters, which have default values $\tilde{\theta}$. This likelihood function can be computed both with observed data and with "pseudo-data", which is constructed from simulation under a specific hypothesis.

We use the test statistic $\tilde{q}_\mu$, based on the profile likelihood ratio [189], to assess the compatibility of the data with either the *background-only* or *signal+background* hypotheses:

$$\tilde{q}_\mu = -2\ln\frac{\mathcal{L}(\mathrm{data}|\mu,\hat{\theta}_\mu)}{\mathcal{L}(\mathrm{data}|\hat{\mu},\hat{\theta})} = -2\ln\lambda(\mu),\ \ 0\leq\hat{\mu}\leq\mu. \qquad (5.7)$$

This test statistic is constructed such that it considers only models with $\mu \geq 0$, furthermore it is constrained to be one-sided by $\hat{\mu} \leq \mu$ such that data with $\hat{\mu} > \mu$ are not used as part of the rejection region for the test on the upper limit of μ.

Here $\hat{\theta}_\mu$ is the conditional maximum likelihood estimator of θ given a fixed value μ, whereas $\hat{\mu}$ and $\hat{\theta}$ refer to the overall maximum likelihood estimators of both quantities. For a given signal strength modifier μ that we test, we first find the observed value of $\tilde{q}^{\mathrm{obs}}_\mu$ and the nuisance parameters $\hat{\theta}_0$ (background hypothesis) and $\hat{\theta}_\mu$ (signal hypothesis). Then, in order to compute the $\mathrm{CL}_s(\mu)$, we compute the p-values of the signal and background hypotheses using

$$p_\mu = \int_{\tilde{q}_\mu \mathrm{obs}}^{\infty} f(\tilde{q}_\mu|\mu,\hat{\theta}_\mu)\ \mathrm{d}\tilde{q}_\mu \qquad (5.8)$$

and

$$1-p_b = \int_{\tilde{q}^{\mathrm{obs}}_\mu}^{\infty} f(\tilde{q}_\mu|0,\hat{\theta}_0)\ \mathrm{d}\tilde{q}_\mu. \qquad (5.9)$$

The p-values are the probabilities of observing results as extreme or more given the underlying hypothesis and are derived from the probability densities of $\tilde{q}_\mu$ under a given hypothesis: $f(\tilde{q}_\mu|\mu,\hat{\theta}^{\mathrm{obs}}_\mu)$. We find the 95% confidence level on the upper limit of μ by adjusting μ until

$$\mathrm{CL}_s(\mu) = \frac{p_\mu}{1-p_b} < 0.05. \qquad (5.10)$$

Equivalently, if $\mathrm{CL}_s < \alpha$ at a given μ, then the Higgs boson is excluded at a production rate of μ or higher with a confidence level $1-\alpha$.

In order to compute the upper limit on μ given the observed data, we need the PDFs $f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{\text{obs}})$, which can be derived using a Monte Carlo method by generating pseudo-data assuming the given signal strength μ and fitting the observed data to evaluate the test statistic. As the MC procedure for generating the PDFs can be very time consuming, we use an approximate asymptotic distribution [189] for the PDF $\tilde{q}_\mu$, which results from the Wald approximation for the profile likelihood [190]:

$$-2\ln\lambda(\mu) = \frac{(\mu - \hat{\mu})^2}{\sigma^2} + \mathcal{O}(1/\sqrt{N}) \tag{5.11}$$

where σ is the standard deviation of $\hat{\mu}$ derived from the full covariance matrix of the likelihood function.

Using the asymptotic distribution for $f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{\text{obs}})$, we find the upper limit for μ at a confidence level of $1-\alpha$ to be

$$\mu = \hat{\mu} + \sigma\Phi^{-1}(1-\alpha) \tag{5.12}$$

where Φ^{-1} is the inverse of the cumulative distribution of the Gaussian PDF. The standard deviation of μ can be computed from the likelihood function eq. (5.5) using the so-called Asimov data set, where the SM prediction is used for s_i and b_i.

We will also quote the expected sensitivity of the measurement, which is derived from simulation, assuming $\mu = 1$ and computing the median expected upper limit on μ using the asymptotic formulae.

Uncertainties in the statistical model

Our prior knowledge of the systematic uncertainties is encoded in $p(\tilde{\theta}|\theta)$, where the values of the nuisance parameters θ are determined by minimising the likelihood function in a frequentist sense. We use $\tilde{\theta}$ to represent our best pre-fit estimate of the nuisance parameters, which can be

- Gaussian, used for shape variations,
- log-normal used for nuisance parameters for which negative values are unphysical,
- flat, in case we cannot assign a prior uncertainty.

In order to account for limited MC statistics, we create nuisance parameters for each process and bin in the template distributions, corresponding to the Poisson uncertainties from the limited number of simulation events.

5.3.10 Analysis of the statistical model

In this section, we will study the expected sensitivity as predicted by the statistical model. We have used pseudo-data constructed from MC simulation with the SM expectation ($\mu = 1$) and a no-signal model ($\mu = 0$) as representative datasets. The nuisance parameters are set to the priors, with central values at 0 and relative variances at 1, compared to the pre-fit expectation. This is the Asimov principle, where an ensemble of datasets is replaced by a single representative dataset [189].

First, in order to validate the fit model, we study the effect of systematic uncertainties in the form of pulls and constraints on the nuisance parameters after the fit to pseudo-data derived from MC. We compute the pulls and constraints, defined as the central value and the width of the distribution $(\hat{\theta} - \theta_0)/\Delta\theta$ with respect to the pre-fit values θ_0, with the uncertainty $\Delta\theta$ determined around the minimum of the likelihood function. By fitting the signal+background and background-only models on the background-only dataset, we verify that both models result in equivalent constraints and no pulls, as can be seen in fig. 5.17.

We further determine that in the fit to the $\mu = 1$ Asimov dataset, some of the nuisances in the background-only model will be shifted with respect to their pre-fit values to compensate the mismatch between the background-only model and the signal+background dataset. In particular, we observe a negative pull in the heavy flavour modelling of the b discriminator and positive pulls

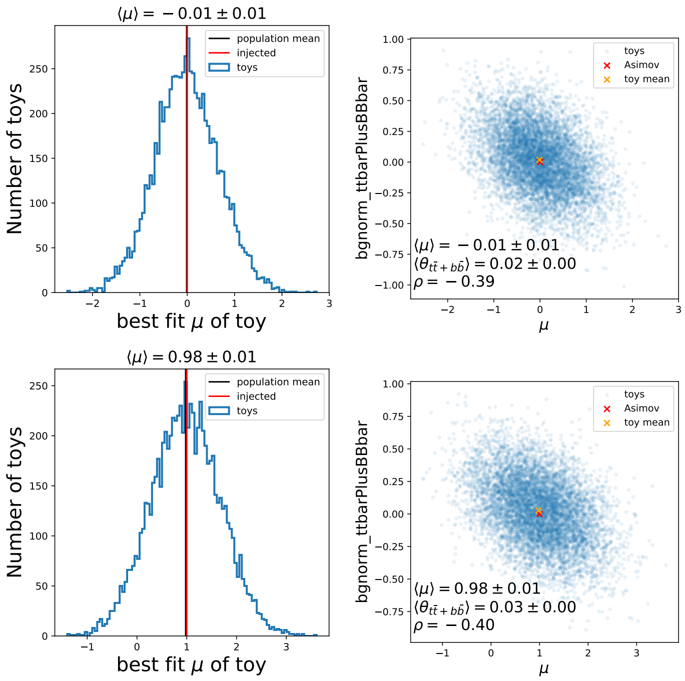


Figure 5.16: On the left, the distribution of the best-fit values $\hat{\mu}$ for toy experiments sampled from the post-fit $\mu = 0$ dataset (top row) and from the $\mu = 1$ dataset (bottom row). On the right, we show $\hat{\mu}$ with respect to $\hat{\theta}_{\mathrm{t\bar{t}+b\bar{b}}}$. We see that the expected anti-correlation between these parameters is reproduced by the model.

for the $\mathrm{t\bar{t}+b\bar{b}}$ normalisation and $\mathrm{t\bar{t}}$+jets ISR modelling, showing that these nuisance parameters are (anti)-correlated to the signal strength parameter μ. When fitting the signal+background model on the $\mu = 1$ dataset, we observe constraints at the level of $\mathrm{Var}[(\hat{\theta} - \theta_0)/\Delta\theta] \simeq 0.5$ for these nuisances, from which we surmise that the model has sensitivity for these nuisance parameters with respect to the prior uncertainties. This is expected, as the prior uncertainties on some of the nuisance parameters are assumed to be quite large in the absence of a detailed description by MC. We can see that the correlation between the best-fit values of the signal strength μ and the nuisance parameters is significant, by determining the best-fit values as a function of the $\mathrm{t\bar{t}+b\bar{b}}$ background normalisation nuisance parameter $\theta_{\mathrm{t\bar{t}+b\bar{b}}}$, as shown in fig. 5.18.

Furthermore, we study the post-fit distribution of the signal strength parameter $\hat{\mu}$ and the nuisance parameters $\hat{\theta}$ by MC sampling. We see in fig. 5.16 that the mean of the best-fit signal strength parameter $\langle\hat{\mu}\rangle$ reproduces the prior value to within $\Delta\mu \simeq 0.01$.

5.4 Results

After having validated the statistical model on simulation using the Asimov dataset, we carry out the analysis on the observed data by using the real dataset instead of the Asimov dataset from the MC expectation. A combined fit across all categories is used to extract the signal strength parameter μ. We show the pre-fit and post-fit distributions in the final categories in figs. 5.21 to 5.24. The uncertainty is reduced by the fit and the post-fit description of the data is generally good, with p-values for the goodness of fit test at the level of $p = 0.8 - 0.99$ for the individual categories. We observe a downward fluctuation in the MEM distribution in the DL ≥ 4 jets, ≥ 4 tags category, which is compatible with the MC expectation within one standard deviation. The post-fit p-value in this category is $p \simeq 0.87$.

We compare the post-fit distributions of the nuisance parameters in fig. 5.19 to the expected distributions in fig. 5.17. In general, we see that the data admit strongest constraints, at the level of $\mathrm{Var}[(\hat{\theta} - \theta_0)/\Delta\theta] \simeq 25\%$ for the nuisance parameters of the $\mathrm{t\bar{t}}$+jets parton shower modelling. The most significant pull is for the CSV heavy flavour modelling nuisance parameter, at $\langle(\hat{\theta}-\theta_0)/\Delta\theta\rangle \simeq$

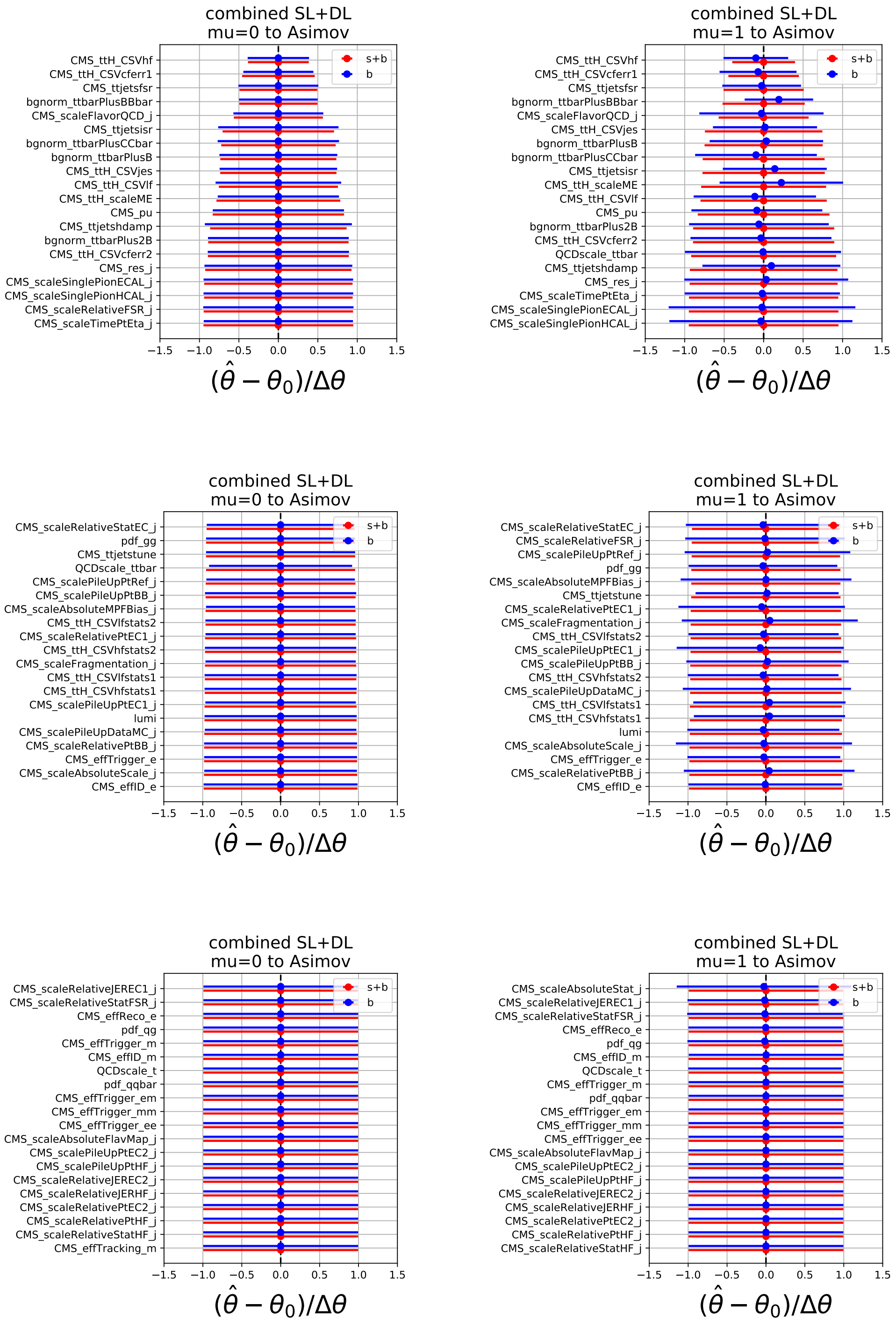


Figure 5.17: The pulls and constraints on the nuisance parameters of the background-only model (blue) and the signal+background model (red) on the $\mu = 0$ Asimov dataset (left) and the $\mu = 1$ Asimov dataset (right). The nuisances are ordered by ascending constraint size (width of the pull distribution), shown as the error bar around the pull $\hat{\theta}/\theta$. We only show the first 60 nuisance parameters that experience the most significant constraints, out of the full list of $\mathcal{O}(600)$. For the $\mu = 1$ Asimov dataset, we observe the strongest constraints at around $\hat{\sigma_\theta} \simeq 0.5$ for the CSV b discriminator heavy flavour modelling (`CMS_ttH_CSVhf`), the $\mathrm{t\bar{t}} + \mathrm{b\bar{b}}$ normalisation (`bgnorm_ttbarPlusBBbar`), the $\mathrm{t\bar{t}}$+jets FSR modelling (`CMS_ttjetsfsr`) and the b discriminator charm flavour modelling (`CMS_ttH_CSVcferr1`).

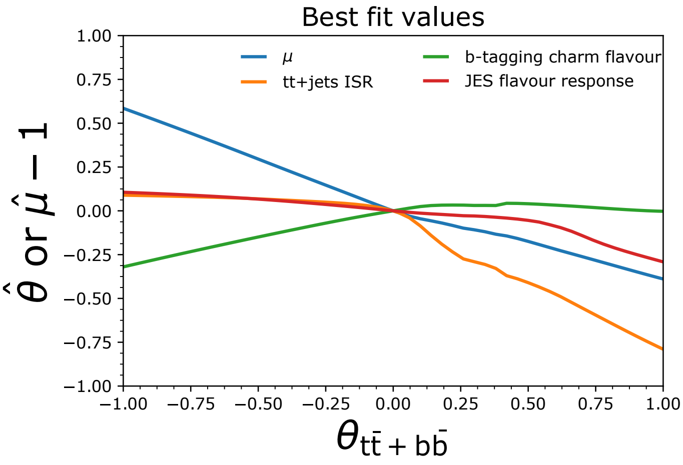


Figure 5.18: The best-fit estimations of the signal strength parameter μ and other significant nuisance parameters as a function of the $\mathrm{t\bar{t}+b\bar{b}}$ normalisation uncertainty ($\theta_{\mathrm{t\bar{t}+b\bar{b}}}$) nuisance parameter. We see that the best-fit value of the signal strength parameter is anti-correlated to $\theta_{\mathrm{t\bar{t}+b\bar{b}}}$, and other nuisance parameters have a non-trivial dependence on the chosen value of $\theta_{\mathrm{t\bar{t}+b\bar{b}}}$.

-0.8. This can be understood as the model compensating for the shape mismodelling of the b tagging likelihood discriminator.

In order to test the compatibility of the results between the semileptonic and dileptonic categories, we carry out additional fits in the semileptonic and dileptonic categories separately. In the semileptonic channel, we find $\mu = 0.71 \pm 0.88$, whereas in the dileptonic channel, the best-fit value for the signal strength parameter is $\mu = -1.61^{+1.25}_{-1.20}$. This is understood to be the effect of the downward fluctuation in data in the dileptonic discriminator distribution. We also find the upper limit on the signal strength at a 95% confidence level in the combined case and for the SL and DL categories. The results of these fits are shown in fig. 5.26. Furthermore, the post-fit agreement between the model and the data can be visualised by sorting all the bins of the fitted templates according to the expected signal over background ratio, as shown in fig. 5.27. Overall, we find the best fit value and the upper limit for the signal strength parameter to be

$$\hat{\mu} = -0.07^{+0.28}_{-0.27}\ \mathrm{stat}^{+0.73}_{-0.74}\ \mathrm{syst} \tag{5.13}$$

$$= -0.07^{+0.78}_{-0.79} \tag{5.14}$$

$$\mu^{95\%CL} = 1.52 \text{ obs. } (1.57 \text{ exp.}). \tag{5.15}$$

The compatibility of the data with the SM hypothesis is at the level of $Z = 1.4\sigma$, as shown in fig. 5.20. The data are also compatible with the background only hypothesis. In order to study the compatibility of the combined fit with the individual categories and to understand the contributions of the various categories, we have carried out individual fits in all the categories separately, eight in total. The results are shown in fig. 5.28 and in table 5.6. In general, we see that the fit in the semileptonic categories prefers a positive signal strength, whereas in the dileptonic categories, a negative signal strength is preferred. This is consistent with the pre-fit distributions and the observed downward fluctuation in the dilepton channel. For these individual fits, the nuisance parameters obtain different and possibly inconsistent values for each fit. Therefore, we have also carried out a multidimensional fit with an individual μ for each category, but the same nuisance parameters across the categories. The results from this multidimensional fit are compatible with the combined fit with a p-value of $p \simeq 0.56$, estimated using toy experiments drawn from MC distributions.

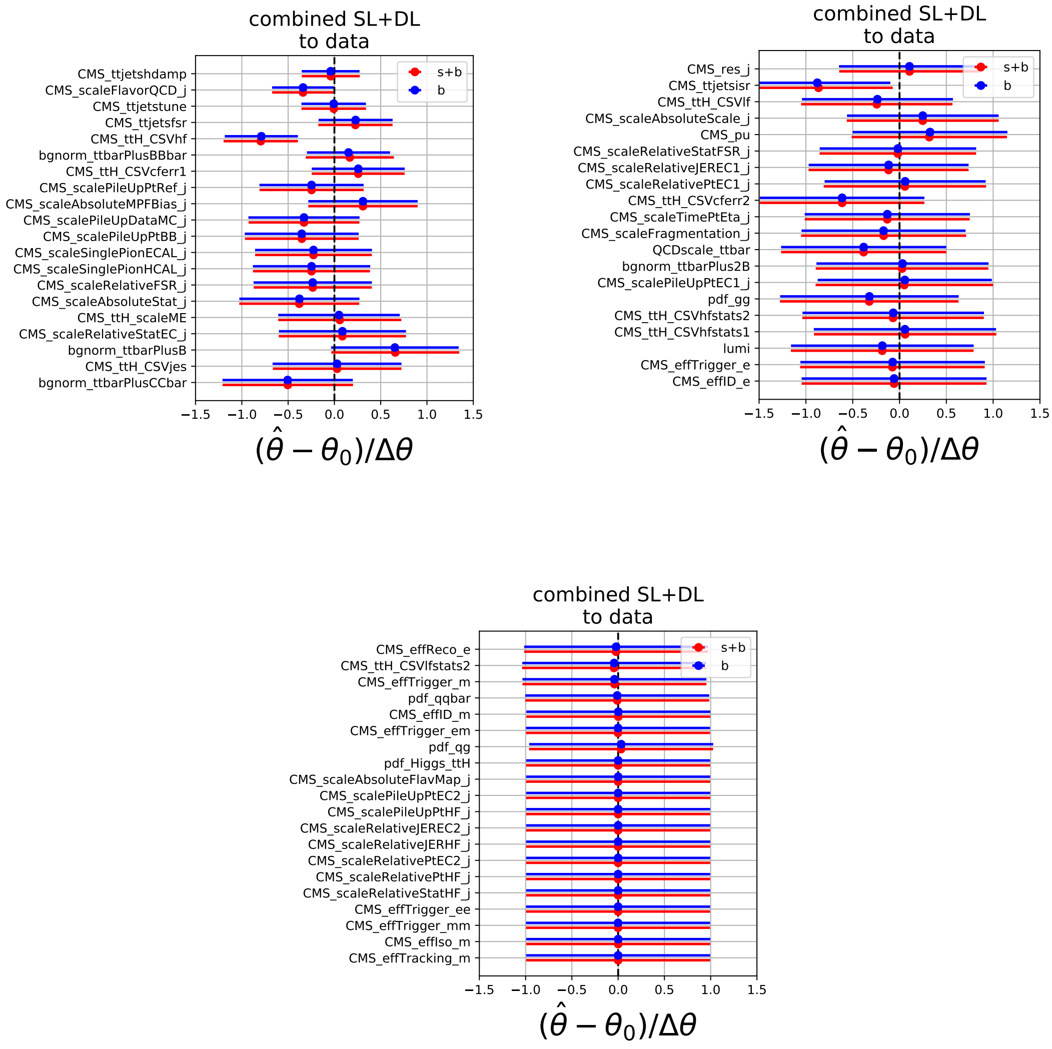


Figure 5.19: The constraints and pulls on the nuisance parameters, fitting the model to data. We see the strongest constraints for the $t\bar{t}$+jets modelling, the JES flavour modelling and the CSV b discriminator heavy flavour modelling.

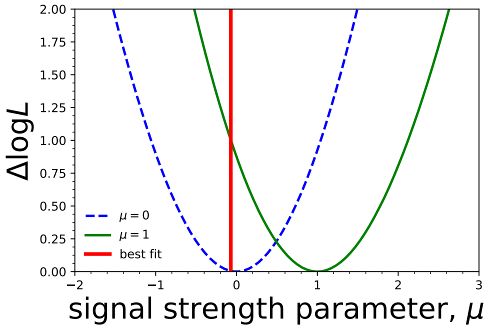


Figure 5.20: The likelihood as a function of μ on the background only ($\mu = 0$) Asimov dataset (blue) and the signal+background ($\mu = 1$) Asimov dataset. We also see that the uncertainty in μ is reduced when removing the systematic uncertainties from the model.

category	best fit μ	95% upper limit on μ		
		expected	observed	injected
DL $\geq$4j 3t	$-1.93^{+5.87}_{-7.24}$	$10.91^{15.21}_{7.90}$	9.91	12.28
DL $\geq$4j $\geq$4t	$-1.51^{+1.26}_{-1.23}$	$3.01^{4.53}_{2.06}$	2.03	3.97
DL	$-1.61^{+1.25}_{-1.20}$	$3.04^{4.55}_{2.09}$	1.96	3.94
SL 4j3t	$9.21^{+10.79}_{-15.86}$	$31.12^{43.41}_{22.57}$	39.04	28.52
SL 4j4t	$0.34^{+6.05}_{-5.89}$	$12.09^{17.35}_{8.56}$	12.43	12.51
SL 5j3t	$7.94^{+8.32}_{-9.83}$	$17.81^{24.84}_{12.90}$	22.79	16.14
SL 5j $\geq$ 4t	$1.73^{+2.56}_{-2.47}$	$5.11^{7.41}_{3.59}$	6.53	5.64
SL $\geq$6j 3t	$0.53^{+5.78}_{-7.32}$	$11.00^{15.12}_{8.12}$	11.32	11.07
SL $\geq$6j $\geq$4t	$1.25^{+1.34}_{-1.36}$	$2.63^{3.80}_{1.85}$	3.66	3.46
SL	$0.71^{+0.88}_{-0.88}$	$1.80^{2.57}_{1.29}$	2.33	2.61
comb.	$-0.07^{+0.78}_{-0.79}$	$1.57^{2.24}_{1.11}$	1.52	2.37

Table 5.6: The best-fit values on the signal strength parameter μ, along with the median expected limit, the 68% confidence interval on the limit and the observed upper limits on μ at a 95% CL. For comparison, we also list the expected limit when injecting a signal with $\mu = 1$.

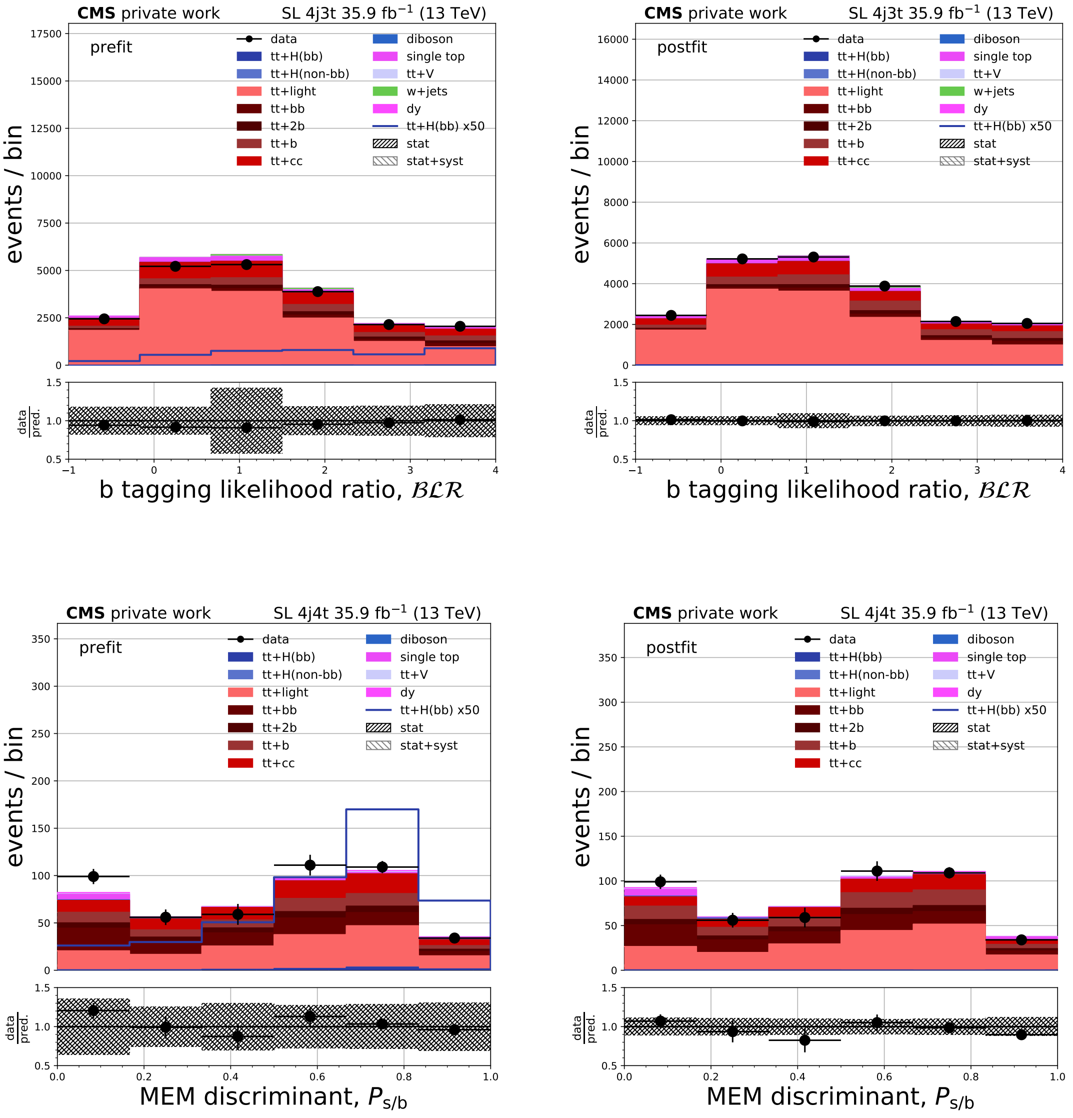


Figure 5.21: The pre-fit (left column) and post-fit (right column) distributions in the semileptonic 4 jet, 3 b tag category (top row) and the 4 jet, 4 b tag category (bottom row).

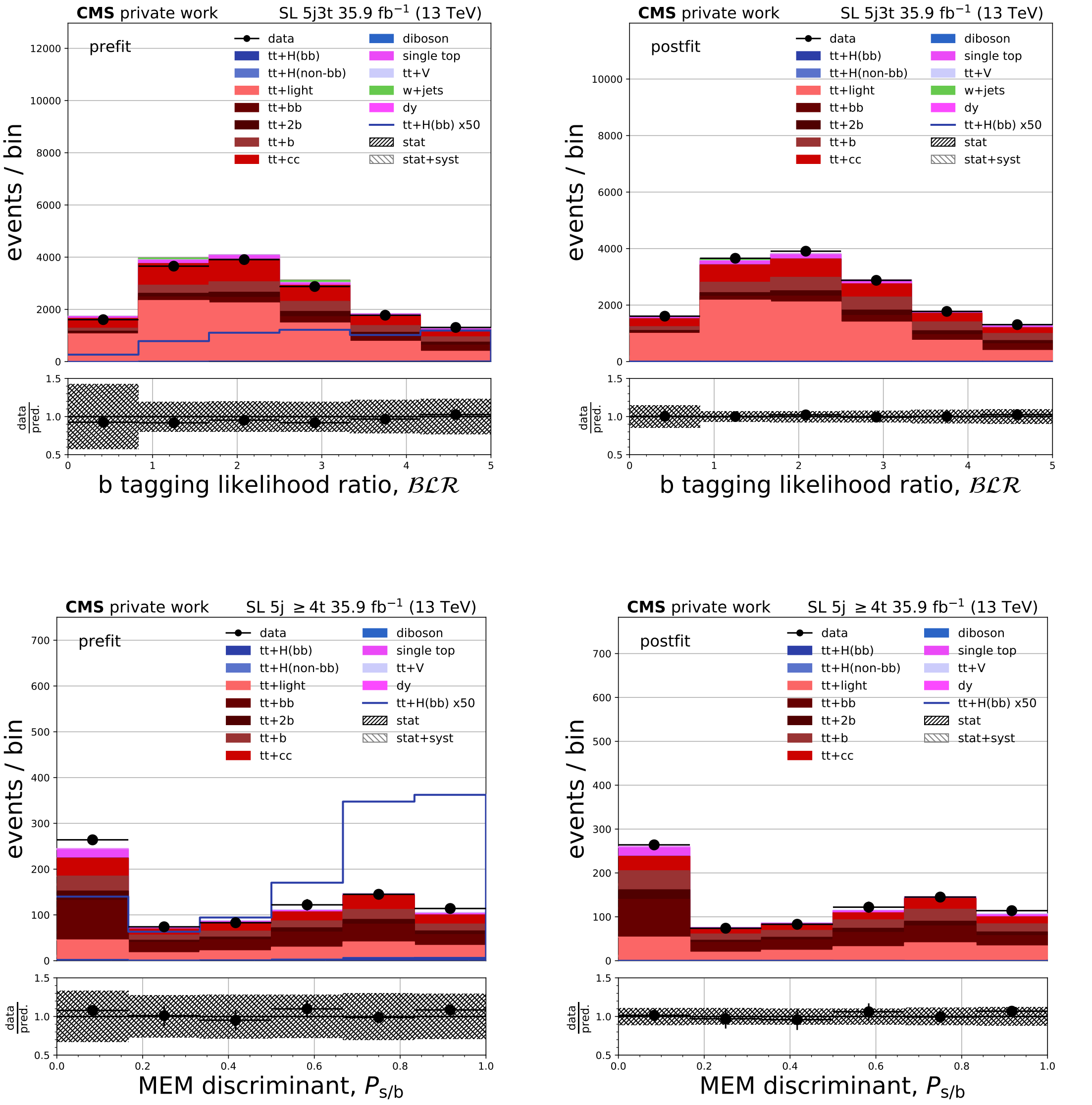


Figure 5.22: The pre-fit (left column) and post-fit (right column) distributions in the semileptonic 5 jet 3 b tag category (top row) and the 5 jet, ≥ 4 b tag category (bottom row).

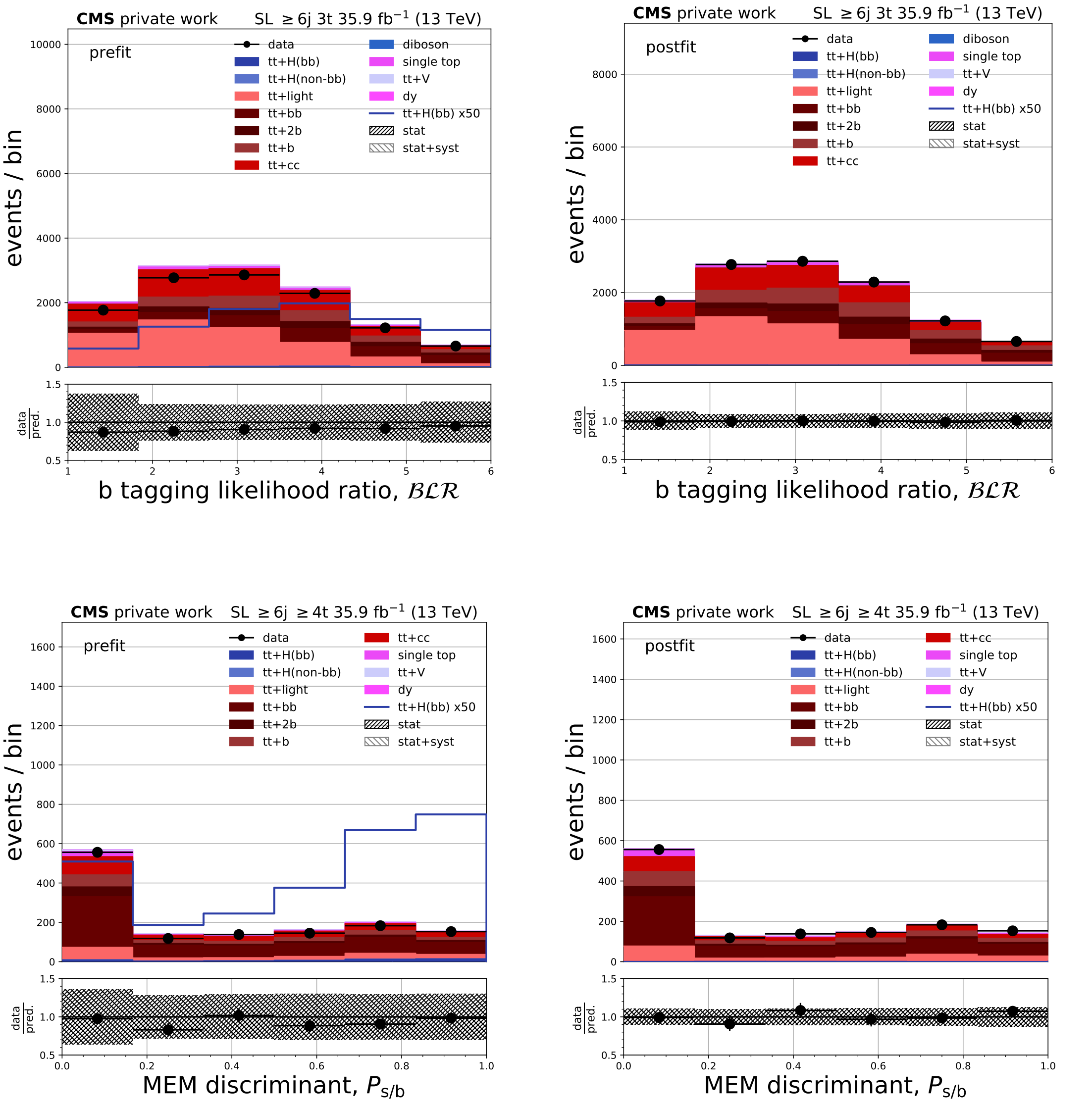


Figure 5.23: The pre-fit (left column) and post-fit (right column) distributions in the semileptonic ≥ 6 jet, 3 b tag category (top row) and the ≥ 6 jet, ≥ 4 b tag category (bottom row).

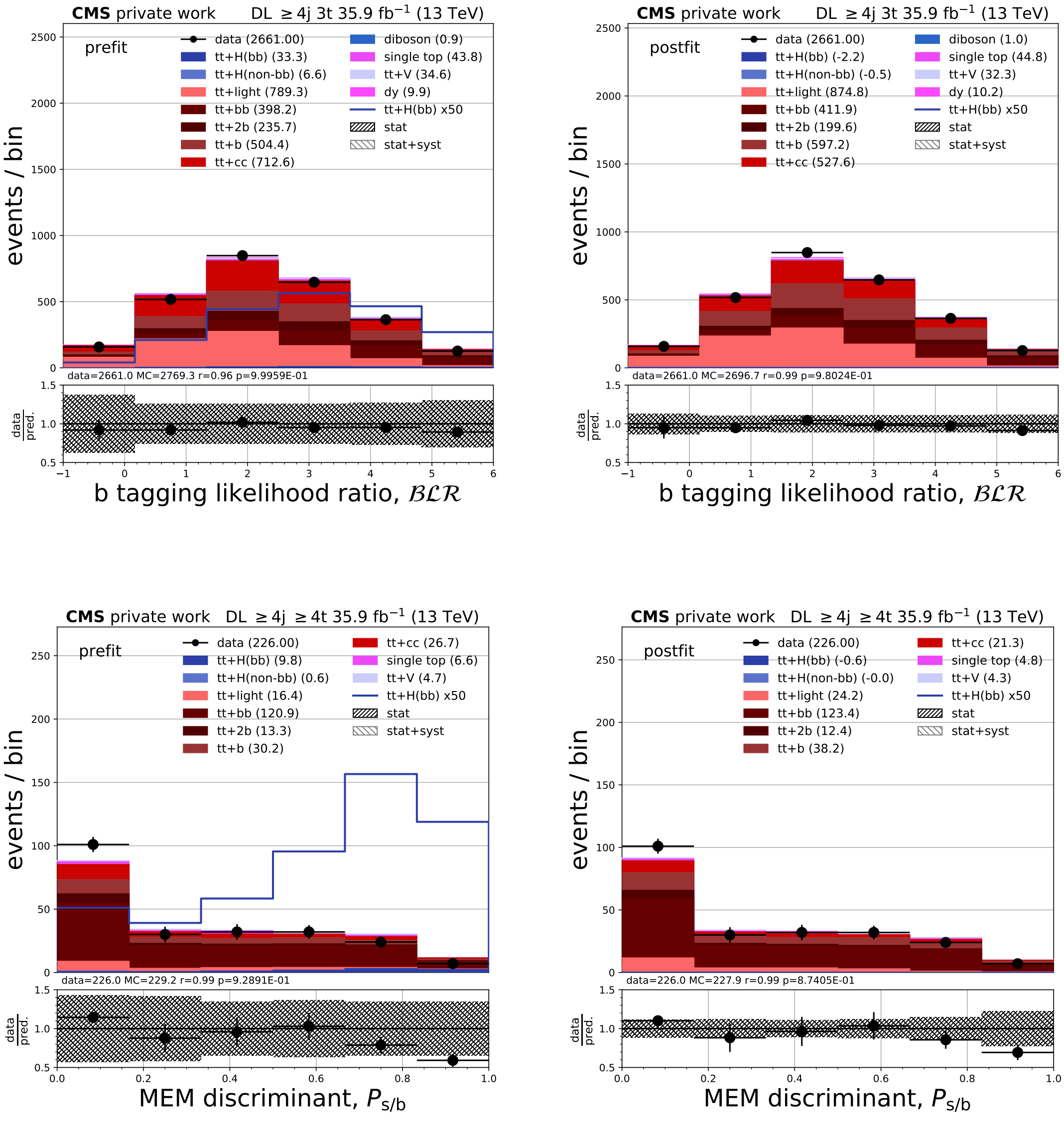


Figure 5.24: The pre-fit (left column) and post-fit (right column) discriminator distributions in the dileptonic categories, with the ≥ 4 jet, 3 b tag category in the top row and the ≥ 4 jet, ≥ 4 b tag category in the bottom row.

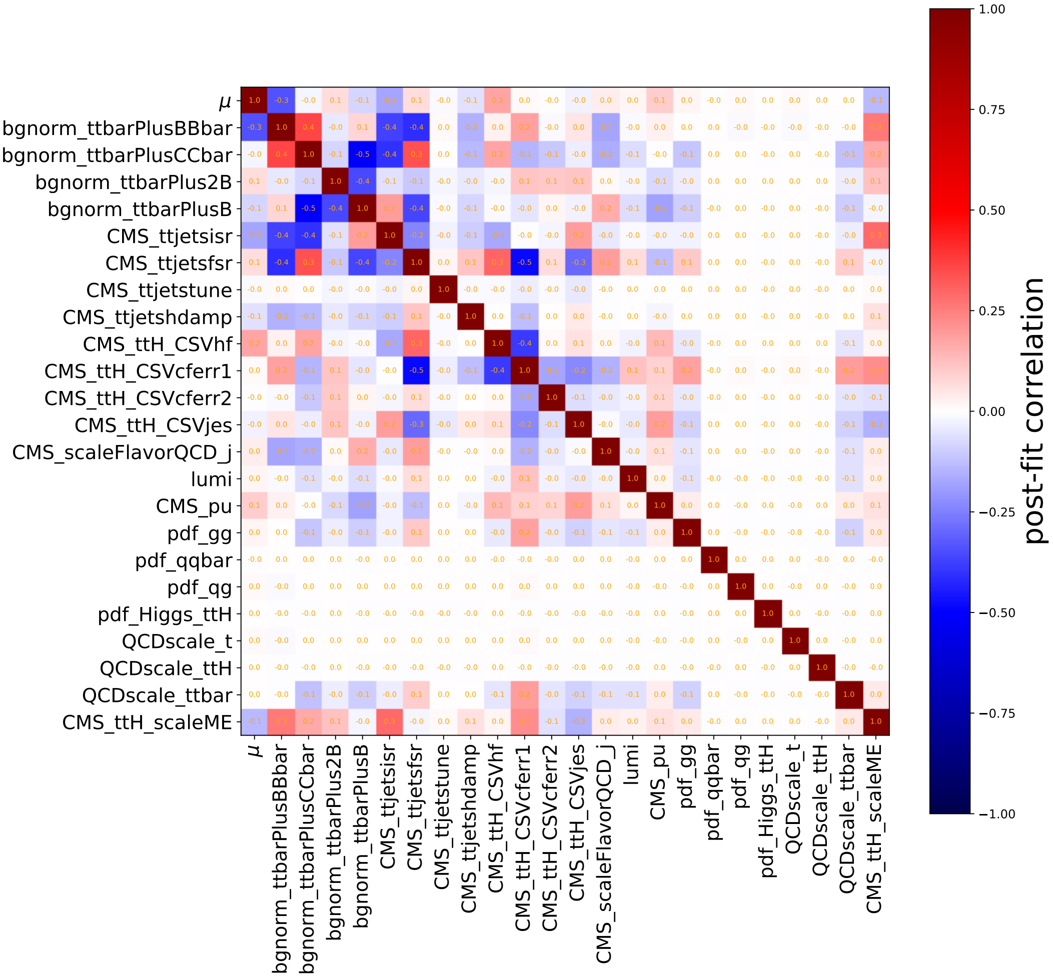


Figure 5.25: The post-fit correlation coefficient between the signal strength parameter μ and the nuisance parameters. We see an anti-correlation at the level of 30% between the signal strength modifier and the $t\bar{t} + b\bar{b}$ normalisation. Furthermore, we find that there is non-negligible residual correlation between the uncertainties.

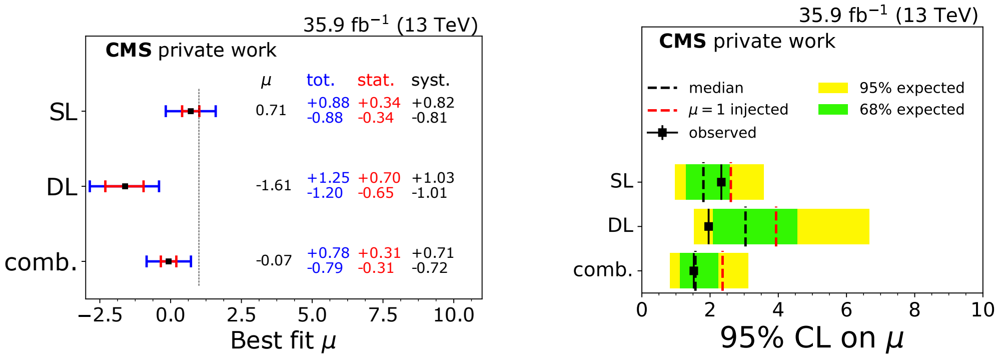


Figure 5.26: The best-fit values (left) and the 95% upper limits (right) on the signal strength parameter in the semileptonic and dileptonic categories and the combined fit.

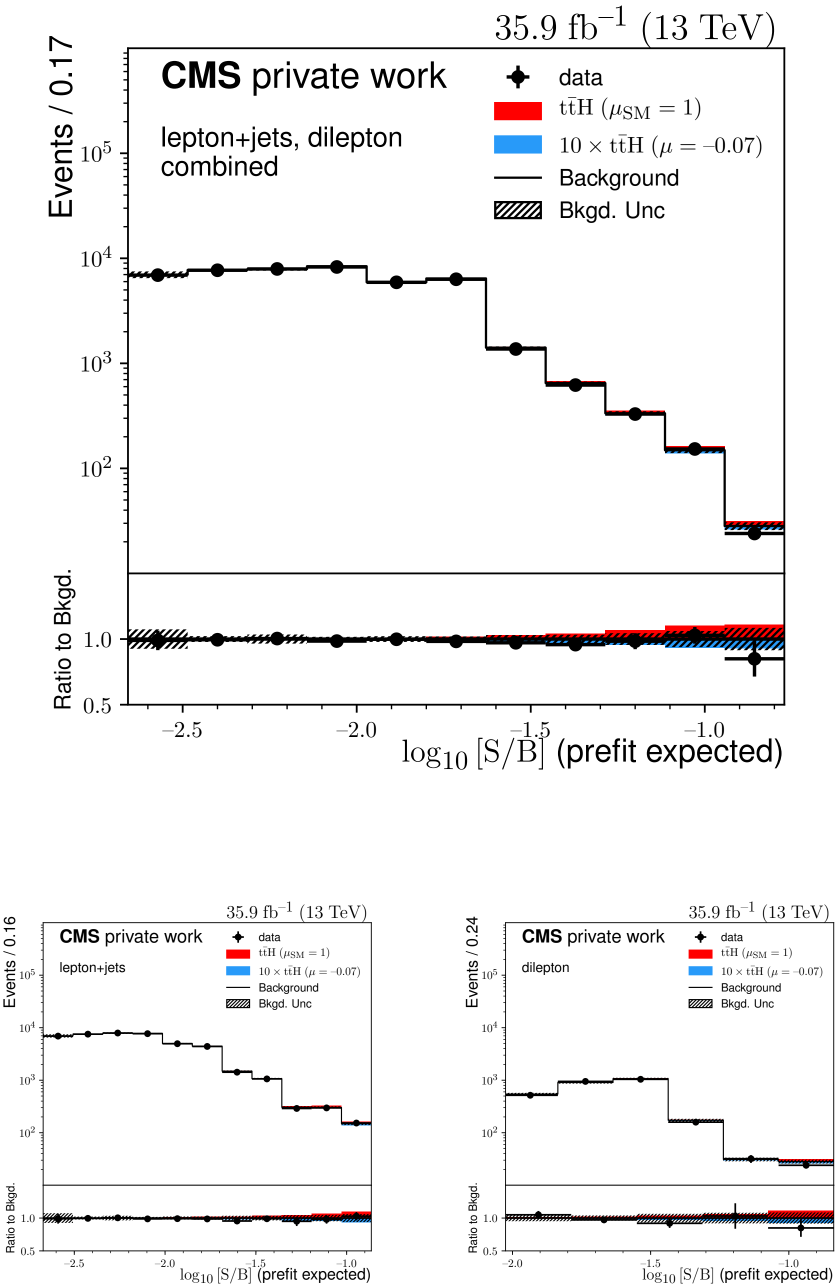


Figure 5.27: The best-fit signal and background distributions in the bins of the final discriminant, arranged by S/B, along with the measured data. We show the combination of the semileptonic and dileptonic channels on the top, with the individual channels shown on the bottom. In determining the uncertainty of the background prediction, we have neglected correlations between the uncertainties of the bins. We show the expected SM $t\bar{t}H$ process in red and the measured value, scaled up by a factor of 10, in blue.

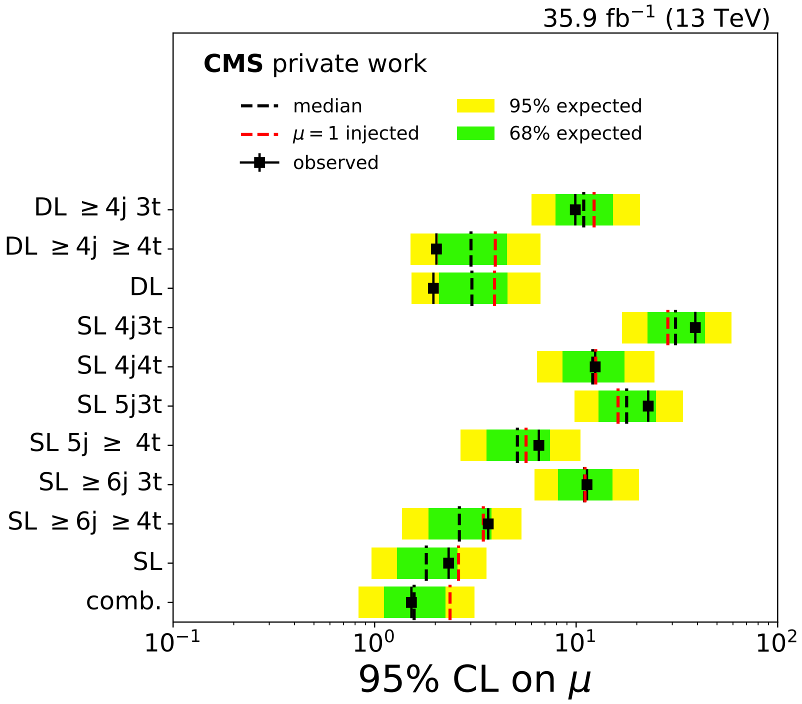


Figure 5.28: The expected and observed upper limits on the signal strength parameter at a 95% confidence limit in the analysis categories. We show the 95% and 68% range of expected limit under no signal, along with the median of the distribution. This can be compared to the expected limit when a nominal SM signal is injected. We also show the observed limit. The numerical values can be found in table 5.6.

uncertainty source	$-\Delta\mu$	$+\Delta\mu$
bkg. norm.	-0.32	$+0.25$
ISR, FSR	-0.14	$+0.15$
pdf norm.	-0.03	$+0.04$
Q^2 scale	-0.14	$+0.12$
MC tune, h_{damp}	-0.05	$+0.05$
theory	-0.53	$+0.49$
MC stats	-0.24	$+0.24$
JES, JER	-0.11	$+0.09$
b tagging	-0.24	$+0.16$
PU, lumi, lepton	-0.09	$+0.08$
experimental	-0.35	$+0.30$
systematic	-0.72	$+0.71$
statistical	-0.31	$+0.31$
total	-0.79	$+0.78$

Table 5.7: The post-fit uncertainties on the signal strength parameter μ in the $\mathrm{t\bar{t}H}(\rightarrow \mathrm{b\bar{b}})$ analysis. The uncertainties are estimated by freezing sets of nuisance parameters in the fit and evaluating the change in post-fit uncertainty with respect to the nominal case. Due to correlations, the uncertainties of the experimental and theory groups cannot be added in quadrature.

5.4.1 Systematic uncertainties

After fitting the model to data, we determine the impacts of the major uncertainties on the signal strength parameter, with the results shown in table 5.7. These are found by freezing a particular set of nuisance parameters and determining the change in the post-fit uncertainty on μ, under the assumption that the uncertainties are uncorrelated. This assumption holds only approximately, as can be seen from the correlation matrix in fig. 5.25, thus the individual uncertainty components cannot be treated as independent and Gaussian-distributed, i.e. they cannot be added in quadrature.

Overall, we see from table 5.7 that this analysis is dominated by systematic uncertainties ($\Delta\mu \simeq \pm 0.73$) in the theoretical modelling ($\Delta\mu \simeq \pm 0.53$) of the background processes ($\Delta\mu \simeq \pm 0.32$). The considerable normalisation uncertainties assigned to the various $\mathrm{t\bar{t}}$+heavy flavour processes, specifically the $\mathrm{t\bar{t}} + \mathrm{b\bar{b}}$ process, have a significant effect on the analysis. The second largest source of theoretical uncertainty is the parton shower modelling of the initial state radiation with $\Delta\mu \simeq \pm 0.15$, which affects the predicted yield of the $\mathrm{t\bar{t}}$+jets background. This is followed by uncertainties on the ME scale with $\Delta\mu \simeq \pm 0.13$, resulting in distortions to the final discriminant shapes.

We have studied the sources of uncertainty in more detail by evaluating the shift in the best-fit value of the signal strength parameter when the nuisance parameters are shifted around their post-fit values by one standard deviation. These results are shown in fig. 5.29, where we see that the $\mathrm{t\bar{t}} + \mathrm{b\bar{b}}$ normalisation uncertainty and the modelling of the ISR have the largest impacts on the signal strength.

Out of the experimental factors, the major sources of uncertainty are MC simulation statistics with $\Delta\mu \simeq \pm 0.25$, which remains a challenge in the high-multiplicity final state, followed by the detailed modelling of the experimental observables related to b tagging with $\Delta\mu \simeq \pm 0.23$ and jet energy corrections with $\Delta\mu \simeq \pm 0.11$. In general, all these sources of uncertainty have considerable effects on the predicted number of events and the distributions of the final discriminators in the analysis categories, such that the final nuisance parameters have non-negligible correlations, for example the nuisance parameters for the CSV heavy flavour and charm flavour modelling are anti-correlated with $\rho \simeq -0.4$, as seen in fig. 5.25.

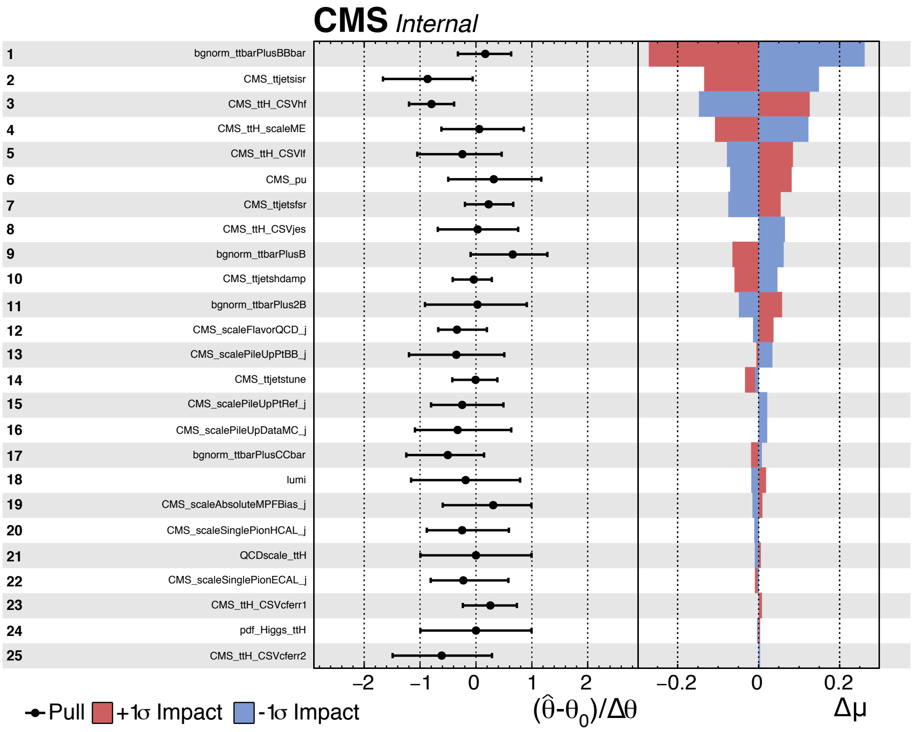


Figure 5.29: The impact of nuisance parameters on the signal strength. The impacts are derived by shifting the nuisance parameter under question by one standard deviation up or down around the post-fit value $\hat{\theta}$ and computing the change in the best-fit signal strength value. We see the most significant impacts from the $\mathrm{t\bar{t}+b\bar{b}}$ uncertainty, followed by the heavy flavour b tagging uncertainty and the parton shower modelling uncertainties. The impact of some nuisance parameters, such as the JES modelling in the CSV b discriminator, may be one-sided, i.e. always positive or negative. This happens when the other nuisance parameters are able to compensate for the shift in this particular nuisance parameter. In comparison to the data in table 5.7, we list the impacts of individual nuisance parameters, instead of logical groups of nuisances. We omit the uncertainties from the limited MC statistics from this list.

5.5 Discussion

This result has been derived using the 2016 dataset from CMS at $\sqrt{s} = 13$ TeV corresponding to an integrated luminosity of 35.9 fb^{-1}. We can compare it to the equivalent CMS analysis at $\sqrt{s} = 8$ TeV [122] with 19.5 fb^{-1} of data, where a similar analysis strategy with the MEM was employed, and to the latest result on $\mathrm{t\bar{t}H}(\to \mathrm{b\bar{b}})$ from the ATLAS collaboration [69] with 36.1 fb^{-1} of data.

We see from the comparison in table 5.8 that this work improves by about a factor of 2 in sensitivity over the equivalent search of CMS at 8 TeV. This can be attributed partly to the increased dataset, but also to the favourable centre-of-mass scaling of the cross-section ratio for signal $\sigma^{13\ \mathrm{TeV}}_{\mathrm{t\bar{t}H}}/\sigma^{8\ \mathrm{TeV}}_{\mathrm{t\bar{t}H}} \simeq 3.9$ and background $\sigma^{13\ \mathrm{TeV}}_{\mathrm{t\bar{t}+jets}}/\sigma^{8\ \mathrm{TeV}}_{\mathrm{t\bar{t}+jets}} \simeq 3.3$ and the improved reconstruction and background rejection, with the signal-over-background fraction improving from 3.4% in the highest purity category at 8 TeV to 4.2% at 13 TeV. The main sources of systematic uncertainties affecting this analysis have not changed with respect to the 8 TeV analysis. In particular, the dominant source of uncertainty in both cases is the 50% normalisation uncertainty on the $\mathrm{t\bar{t}+b\bar{b}}$ background. Although recent analyses from CMS have determined the $\mathrm{t\bar{t}+b\bar{b}}$ cross-section to a relative precision of about $\simeq 30\%$, it is compatible with the post-fit constraints that we observe in this work and thus a more precise measurement would be needed to significantly reduce the effect of this uncertainty. Furthermore, in the aforementioned analysis, the $\mathrm{t\bar{t}H}(\to \mathrm{b\bar{b}})$ process is considered as a background for the $\mathrm{t\bar{t}+b\bar{b}}$ measurement, thus, any $\mathrm{t\bar{t}H}(\to \mathrm{b\bar{b}})$ signal would be absorbed into the $\mathrm{t\bar{t}+b\bar{b}}$ normalisation. A possible direction of future research would be the consistent and simultaneous measurement of $\mathrm{t\bar{t}H}(\to \mathrm{b\bar{b}})$ and $\mathrm{t\bar{t}+b\bar{b}}$ in a multidimensional fit.

It is interesting to compare the work presented in this thesis to the latest results from ATLAS, where an equivalent dataset is used. First, we see that the sensitivity of the ATLAS analysis is about $\simeq 20\%$ higher than that reported in this work. This is likely due to the significantly more complex statistical analysis techniques employed in the ATLAS analysis, in particular a staged approach with a BDT that is responsible for choosing the optimal reconstruction hypothesis for the observed jets, followed by another classification BDT optimised to distinguish between signal and background. These algorithms are optimised on simulation on an event-by-event basis and rely heavily on the detailed modelling of b tagging discriminators. For comparison, only the MEM and the b tagging likelihood ratio classifiers are used in this work, neither of which require significant MC simulation to be evaluated on a new dataset. Furthermore, ATLAS is already making use of improved vertexing and b-tagging in the detector through the Insertable B-Layer [191]. A similarly-upgraded pixel detector is available at CMS since the beginning of 2017 [192].

In terms of uncertainties, the ATLAS analysis has a slightly different treatment of the $\mathrm{t\bar{t}}$+jets modelling and the corresponding uncertainty, where the subcomponents of the default `POWHEG` MC sample are scaled to an NLO $\mathrm{t\bar{t}+b\bar{b}}$ sample generated using `SHERPA`+`OpenLoops` in the 4-flavour scheme. This scaling is most significant for events with more than one jet containing multiple b hadrons, which make up between 1-2% of the full $\mathrm{t\bar{t}+b\bar{b}}$ cross-section [69]. In our analysis, all events with at least one jet containing multiple b hadrons are grouped under the $\mathrm{t\bar{t}+2b}$ process. In the ATLAS analysis, the systematic uncertainties on the NLO modelling are derived by comparing the nominal `POWHEG` model to a 5-flavour `SHERPA`+`OpenLoops` model and independently a 4-flavour model. The normalisation of the various $\mathrm{t\bar{t}}$+jets subprocesses is left freely floating in the fit, described by k-factors. Overall, the CMS and ATLAS procedures for the $\mathrm{t\bar{t}}$+jets modelling uncertainties are different, but in both cases, the background modelling uncertainties have the largest impacts. The ATLAS analysis observes a post-fit scale factor of $k(\mathrm{t\bar{t}}+ \geq 1\mathrm{b}) = 1.24 \pm 0.1$ for the $\mathrm{t\bar{t}+b\bar{b}}$ process, whereas in the analysis presented in this thesis, the post-fit k-factor for this process is consistent with unity. The parton shower modelling uncertainties are estimated by comparing `HERWIG++` and `Pythia 8`, whereas for this analysis, it is obtained by varying the parameters in the nominal `POWHEG`+`Pythia 8` model, as explained in section 5.3.7. Both the ATLAS analysis and the analysis presented in this thesis are mainly impacted by similar systematic uncertainties: the $\mathrm{t\bar{t}}$+jets modelling, limited MC statistics, b-tagging and JES. It is also clear that it is necessary for the experiments to adopt an improved theoretical model of the $\mathrm{t\bar{t}+b\bar{b}}$ process in a way that is consistent and comparable between the experiments.

During 2017, the CMS experiment has recorded approximately 45 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV. Furthermore, the upgrade of the pixel detector and the commissioning of

category	quantity	CMS (8 TeV)	ATLAS (13 TeV)	this work
SL	best-fit μ	$1.7^{+2.0}_{-1.8}$	$0.95^{+0.65}_{-0.62}$	$0.71^{+0.88}_{-0.88}$
	exp. UL	$4.2^{6.2}_{2.9}$	$1.4^{1.99}_{1.01}$	$1.80^{2.57}_{1.29}$
	obs. UL	5.5	1.95	2.33
DL	best-fit μ	$1.0^{+3.3}_{-3.0}$	$-0.24^{+1.02}_{-1.05}$	$-1.61^{+1.25}_{-1.20}$
	exp. UL	$6.9^{15.8}_{3.4}$	$2.74^{3.86}_{1.98}$	$3.04^{4.55}_{2.09}$
	obs. UL	7.7	2.64	1.96
combined	best-fit μ	$1.2^{+1.6}_{-1.5}$	$0.84^{+0.64}_{-0.61}$	$-0.07^{+0.78}_{-0.79}$
	exp. UL	$3.3^{4.9}_{2.3}$	$1.24^{1.77}_{0.89}$	$1.57^{2.24}_{1.11}$
	obs. UL	4.2	1.96	1.52

Table 5.8: A comparison of the CMS [122] and ATLAS [69] results on $t\bar{t}H(\to b\bar{b})$ and the results obtained in this work. We compare the best-fit μ, the expected 95% upper limit (UL) under the no signal hypothesis with $\pm1\sigma$ uncertainties and the observed upper limit.

improved b tagging algorithms promises to improve the sensitivity of this analysis. On the other hand, progress will need to be made in incorporating the latest improvements in the theoretical modelling, in particular, the NLO 4-flavour models for $t\bar{t} + b\bar{b}$ properly matched to an accurate parton shower.

5.6 Summary

We have presented a search for the $t\bar{t}H$ process in the data collected by the CMS experiment during the 2016 run period, in the channels where the Higgs boson decays to b quarks and at least one of the top quarks decays leptonically. The observation of this process, where the Higgs boson is produced in association with top quarks, would make it possible to directly study the process of mass generation for up-type quarks and to confirm the mechanism of electroweak symmetry breaking for the heaviest known quark. The analysis is challenging due to the presence of a significant background arising from the QCD production of $t\bar{t}$+jets. Specifically, the $t\bar{t} + b\bar{b}$ process, where two additional b quarks are produced in the final state in association with the top quark pair, is irreducible with respect to the particles in the final state, as we expect between four to six jets, out of which four arise from b quarks from both the $t\bar{t} + b\bar{b}$ background and the $t\bar{t}H(\to b\bar{b})$ signal. Furthermore, the analysis is complicated by the presence of a combinatorial self-background, as we cannot directly reconstruct the Higgs boson candidate invariant mass peak due to the presence of several additional b quark candidates arising from the top decay.

We have shown that it is possible to construct a discriminator based on the direct computation of matrix elements from the kinematic properties of the observed jets and leptons in the event, without relying on large amounts of MC simulation for the $t\bar{t}+b\bar{b}$ process, which is affected by considerable theoretical uncertainties. The observed data are compatible with both the SM signal hypothesis and the background-only hypothesis. We have been able to establish an upper limit on the signal strength modifier at the level of $\mu < 1.52$ at a confidence level of 95%, with $\mu < 1.57$ expected under the SM hypothesis. The best fit signal strength value is $\mu = -0.07^{+0.28}_{-0.27}$ (stat.)$^{+0.73}_{-0.74}$ (syst.), with the total combined uncertainty being $\Delta\mu \simeq \pm0.79$. The largest components in the final uncertainty are of systematic origin, arising from the modelling of the $t\bar{t}$+heavy flavour background and the overall theoretical uncertainties in the modelling of the $t\bar{t}$+jets background.

Additionally, we find simulation statistics and experimental uncertainties in the detailed calibration of the b discriminator shape to have a sub-leading but significant effect on the final measurement. This means that in addition to an improved theoretical treatment of the $t\bar{t}$+jets background, which is crucial for this analysis, improving the MC simulation in terms of a more efficient use of the generated events and an improved modelling of the quantities related to b tagging can improve this analysis. Both of these problems are well-suited to be solved with the matrix element method discriminator, which does not rely on the precise modelling of various low-level experimental observables such as b discriminator distributions or extensive simulation statistics.

It is expected that the dataset from the 2017 run period will allow us to further improve our understanding of the $\mathrm{t\bar{t}H}$ process, where the $\mathrm{H} \to \mathrm{b\bar{b}}$ decay channel benefits from a high branching ratio. The discovery of the $\mathrm{t\bar{t}H}$ production mode at a 5σ significance level is feasible within 2018 through a combination of multiple channels and datasets. However, a focused experimental and theoretical effort is needed to fully make use of these additional data, as this analysis is affected by significant systematic uncertainties.

6 Modelling memory in language acquisition

In this chapter, we describe the work that we[*] carried out in the context of an internship at the private company Lingvist Technologies[†] during May-July 2017. The main product of the company is a computerised language learning tool that allows users to practice vocabulary in a foreign language using spaced repetition [193]. Spaced repetition is a technique to improve retention of new information by spreading out the amount of time allocated to studying, first described quantitatively by H. Ebbinghaus in the late 19th century [194].

The product has over 10^6 users learning several languages, thereby presenting a rich dataset where techniques of computer-assisted education can be tested and improved. Taking into account the population of users, statistical learning can be used to optimise the contents of the course by modelling the learning process and prior knowledge of users.

Our main contribution to this project was developing a data-driven model based on machine learning (ML) that estimates the prior vocabulary knowledge of users based on a small set of probe questions and the overall statistical properties of the user population.

In this chapter, we will describe the problem of vocabulary prediction, give an overview of the model and validate it with respect to data. We discuss also the reference model that we formalized and the metrics we developed to compare the improved model to the reference. We found that the ML-based model improved over the existing prediction algorithm by about 40% in terms of prediction efficiency, such that this algorithm was put into use in the user-facing product in a test phase.

6.1 Introduction

It has been shown that individual tutoring can improve the results of average students by up to two standard deviations [195]. However, access to high-quality individual tutoring is limited, therefore using computerised assistants presents an opportunity to make learning more effective and accessible.

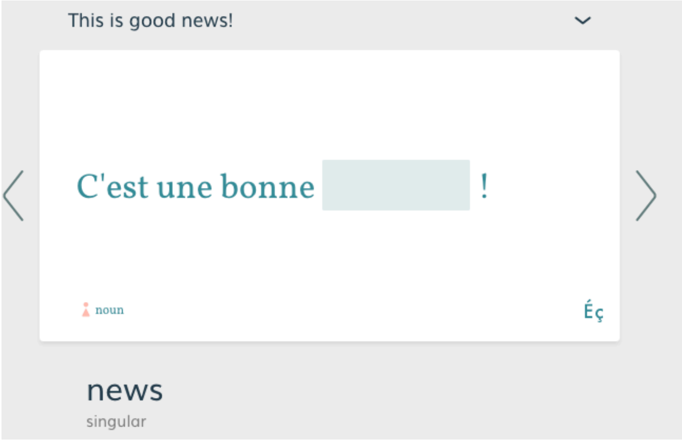
In a computerised learning course, tailoring the course contents to the specific pre-existing knowledge of an individual learner allows the learning material to be covered more efficiently by focussing on subjects that are not yet well understood, dedicating less resources to subjects that the learner (user) is likely to know well or that are too advanced for the learner. Knowledge estimation deals with the question of parametrising and modelling the knowledge of various topics or items on a per-user basis.

Furthermore, if the individual knowledge of users can be modelled as a function of time, then these predictions can be used to test the learner on topics that are at the threshold of their existing knowledge, such that learning speed would be maximised. This is one of the main goals of the learning system at Lingvist.

We describe a computational method to estimate the overall second-language vocabulary of users, based on answers to a small set of trial words on which the users are tested. Effectively, users entering the learning environment are presented with words in their target language, for which they provide their best answer as free-form text. After this attempt, the user is presented with the correct answer. These exercises are repeated after some time in order to improve vocabulary acquisition and test the users' knowledge. An example is shown in fig. 6.1. Such user-word-answer tuples form the basic unit of data in the learning environment and a basis for the modelling.

[*] In this chapter, "we" refers to J. Pata

[†] http://lingvist.io



This is good news!

C'est une bonne ______ !

news

singular

Figure 6.1: An example of an exercise in the Lingvist environment, where the user is learning French based on English.

By implementing a simplified version of Deep Knowledge Tracing (DKT) [196], we show that by predicting the knowledge of individual items on a per-user basis using a sequence of previous guesses, we can develop an accurate estimation procedure for the overall vocabulary knowledge of the user. We achieve this by compressing the sparse and variable-length guess sequences into a fixed-size representation using a recurrent neural network (RNN) and furthermore mapping this small fixed-size representation into a set of numbers that can be interpreted as distinct per-item knowledge likelihoods[‡].

This chapter is organised as follows: first, in section 6.2, we review the mathematical background of the problem of knowledge estimation and present an overview of the data. In section 6.3, we discuss the details of the RNN model that was used to estimate the user knowledge and compare it with other models. In section 6.4, we analyse the results and finally in section 6.5, we summarise the studies that were made to optimise the RNN model and discuss possible improvements that could be made to the model.

6.2 Problem statement and data description

In the following section, we describe the problem of estimating the users' vocabulary knowledge in terms of guess likelihoods for basic language items and describe the dataset based on which we will construct the prediction model.

6.2.1 Knowledge estimation

The problem of knowledge estimation can formally be stated as follows. We have a set of items, indexed by $n = 1, 2, \ldots, N$, for which users, indexed by $m = 1, 2, \ldots, M$ provide guesses that can be correct or incorrect, denoted by $g_{nm} \in \{1, 0\}$, with $g_{nm} = 1$ ($g_{nm} = 0$) corresponding to the n-th item being answered correctly (incorrectly) by the m-th user. Therefore for each user, the guess data form a sequence

$$S_m = [(n_1, g_1), \ldots, (n_i, g_i), \ldots, (n_T, g_T)]$$

where T is the number of guesses a user has made. The individual knowledge of a specific user m can then be written as $\boldsymbol{g}_m = (g_1, g_2, \ldots, g_N)$. Note that we have no control over the order in which items are presented to the user nor the set of items which are presented, such that the vector $\boldsymbol{g}_m$ is sparsely filled. This arises from the nature of the function that generates the sequence of questions S_m, which is randomized on a per-user basis. For example, user $m = 1$ can answer items in the

[‡] To be distinguished from a probabilistic interpretation, where the predicted knowledge values are required to sum to unity over a population.

order $(n = 1, n = 17, n = 3, \dots)$, whereas another user $m = 2$ may instead have answers for $(n = 2, n = 1, n = 10, \dots)$.

Lexical unit The items in the guess sequence themselves may be arbitrary, but in our case, they always represent word pairs, asking the user to guess a word in their target language based on an explanation in their source language. We use the term *lexical unit* (LU) to denote such word pairs. An example LU would be the pair *news* - *la nouvelle* in the English-French language pair, embedded as a completion in a sentence such as: *Cette une bonne . . . (news)*, as illustrated in fig. 6.1.

These data can be represented as a $(N \times M)$ matrix $\mathbb{G} = (g_{nm})$, with N being the total number of users and M the total number of items. In general, users may answer questions from a randomized LU sequence, thus, this matrix is sparsely filled, as can be seen in fig. 6.2, where we have shown the guess data for a small subset of users and LUs. We explicitly distinguish between the case where a user doesn't know an item $g_{nm} = 0$ and where no answer has been provided by storing the matrix $\mathbb{G}' = (g'_{nm})$, $g'_{nm} \in \{\text{available}, \text{not available}\}$, that records whether a given user has provided an answer to a given question. We use this matrix to mask out the unknown questions in the optimization phase, as explained further in section 6.3.2.

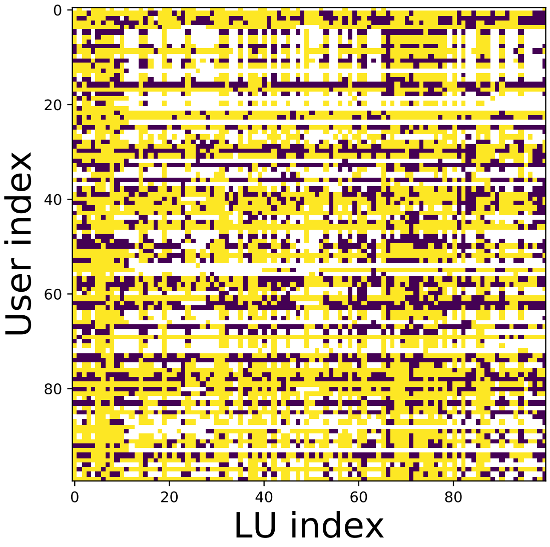


Figure 6.2: Guess data for the first 100 users and the first 100 LUs, a small subset of the full user population and LU space. Correct guess attempts are shown as a yellow pixel, incorrect attempts where the users guess did not correspond to the correct answer with a dark purple pixel. Cases where data were not available are shown as white pixels. The relative sparsity of data results from the question algorithm sampling the LU space in a pseudo-random fashion.

Prediction task The task of knowledge estimation is then to predict the guess probabilities $\boldsymbol{g}_m$ for all items for an individual user m, given a short ($\sim$ 5% of the total or about 10-50 guesses) subsequence of that user's guesses. This can further be seen in the context of knowledge tracing, where given a sequence of observations about a user $\mathbf{x}_0 \dots \mathbf{x}_t$ over time steps t, the task is to predict properties of the interaction $\mathbf{x}_{t+1}$. Here, the interaction is formally a pair $\mathbf{x}_t = \{n_t, g_t\}$ of the knowledge item and the guess. Time-dependent prediction is important for modelling the memory and learning capacity of users. In the studies that follow, we decoupled the study of time-dependent properties of learning from the initial knowledge, focusing on optimising the method for predicting the pre-existing vocabulary.

The guess probabilities of different items may be related to each other, such that knowing a difficult word would be a good predictor for knowing other words of similar difficulty. Therefore, the task is to model the full probability distribution $p(\boldsymbol{g}_m|\boldsymbol{g}_m^{\text{obs}})$, given the observed data $\boldsymbol{g}_m^{\text{obs}}$ for a user. However, we do not have an explicit labelling of which items are related and we hope to capture the essential properties of the knowledge probability by directly using the data.

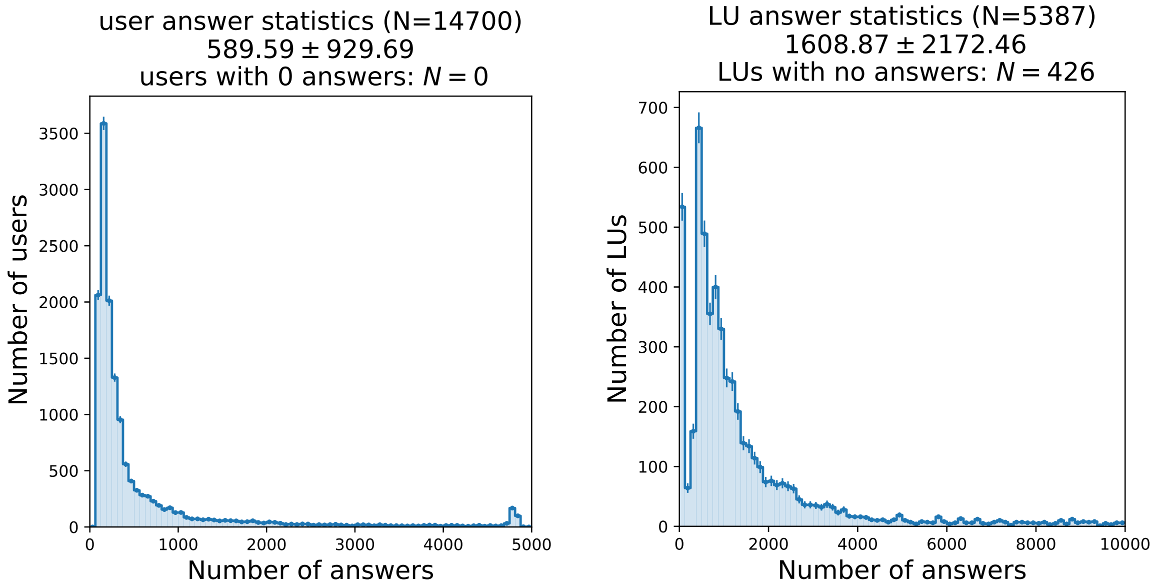


Figure 6.3: On the left, we show the distribution of users by number of answered LUs. We have required that a user has answered at least 100 LUs. On the right, we show the corresponding distribution of LUs by number of answers from users. At this stage, LUs have not been filtered, hence the whole course is considered. The slight peak in the tail of the user answer distribution represents very enthusiastic language learners who have progressed through the complete course.

6.2.2 Data

We use guess data for about six months between January - June 2017 from roughly 15 000 users from the English to French course at Lingvist, requiring that users have answered at least 100 LUs. This corresponds to about 10 million user - guess pairs. The course consists of roughly 5400 individual LUs. Overall, an average (mean) user has answered about 590 LUs and an average (mean) LU has about 1600 answers, with the distributions following a rough power law form, as can be seen in fig. 6.3. Although we used the English-French language pair for the primary studies, we also validated the model on several other language pairs with less data, namely Russian-French, English-Standard Chinese (Mandarin) and English-German, and confirmed that the results are broadly replicable.

The problem of predicting the properties of the guess matrix $\mathbb{G}$ as stated above is well known in literature as matrix completion [197], where the task is to discover hidden or latent variables that would allow the unspecified elements of a matrix to be estimated based on assumptions about rows or columns being sampled from a common distribution. Such problems are typically solved with some combination of collaborative filtering and matrix factorization. However, as in the future we may also want to model time-dependent behaviour of the estimated knowledge, we seek a more generic solution to the problem that would make it possible to add arbitrary information about users and guesses and derive the appropriate model without much expert knowledge. Furthermore, in the future, the definition of a lexical unit (LU) may be expanded to encompass more general exercises beyond word pair guesses, for which a simple matrix representation may not be sufficiently flexible.

6.3 Model description

6.3.1 Reference models

In this section we describe the baseline models that will be used as reference implementations to estimate the improved performance from the proposed approach. The simplest model for predicting the prior knowledge rate of words would simply be to assume that any user is an average user and each word with index $n = 1 \dots N$ has a probability p_n of being known by any user, where the probability is simply estimated by computing the ratio of correct to all guesses of this word by all users. Clearly this does not allow the learning experience to be tailored, but it is a reasonable first guess about the user in the absence of other data.

As the user proceeds through the course, their guesses will allow us to update our estimation of their knowledge in a Bayesian framework. In general, Bayesian Knowledge Tracing [198], where the skill-specific performance is modelled using transition probabilities from an unlearned to a learned state based on evidence, has been shown to perform at least as well as the best machine-learning based models, however, such models require considerable domain knowledge and manual item labelling to be used successfully [199], thus we investigate supervised neural networks as a possible solution.

In order to measure how well a model is able to predict user knowledge, we need to establish a set of metrics. In particular we use the number of correctly predicted items that the user actually knows (true positives) and the number of items incorrectly predicted to be known (false positives). Thus, we effectively treat the problem as binary classification with each guess attempt being an independent trial of equal weight.

We use these true and false positive rates to construct the receiver operating characteristic area-under curve (ROC AUC) metric as an overall average performance indicator. Such a metric considers all trials as equal, meaning it is tuned to measure the performance of the model on the same distribution of LUs that is used in the optimisation. This is referred to as the trial-weighted case. An alternative would be to treat each LU, regardless of how many guesses have been made for it, as equal, effectively reducing the weight of the LUs for which very many users have provided guess data. We refer to this later as the LU-weighted case.

6.3.2 Machine learning model

Since we may want to continuously make predictions as the user progresses through the course, the model should be able to deal with sequence data of variable length. This leads us to consider recurrent neural networks as a possible model architecture. RNNs are a type of models that are suited for processing sequential data through the use of parameter sharing between successive recurrence steps [200]. In particular, RNNs can be thought of as dynamical systems in time t characterized by a function $f(\mathbf{h}_t, \mathbf{x}_t, \theta)$ and a set of parameters θ, where an external input $\mathbf{x}_t$ modifies the hidden state $\mathbf{h}_t$ of the model through the recurrence relation

$$\mathbf{h}_t = f(\mathbf{h}_{t-1}, \mathbf{x}_t, \theta) \tag{6.1}$$

such that the hidden state vector at time t is a compressed representation of the sequence $\mathbf{x}_1 \dots \mathbf{x}_t$. The model parameters are optimized using the back-propagation through time (BPTT) approach, where the recurrence relation is unrolled in time and the weights adjusted using gradient descent and error propagation.

In general, RNN models are problematic to optimise on sequences with long-range correlations, as they suffer from issues with vanishing gradients over long sequences. This is known as the "vanishing gradient" problem, where the weight updates in the back-propagation step either decay to very small values or increase exponentially ("exploding gradient") over long sequences [201]. Inspired by the Deep Knowledge Tracing framework [196], we use a Long Short-Term Memory (LSTM) [202] architecture to summarise a guess history of arbitrary length as a vector of fixed length. The advantage of the LSTM-type model over standard RNN-s is that the network is augmented with a memory cell and gates for storing and forgetting the item in memory, as shown in fig. 6.4. This has been shown to help solve the vanishing gradient problem typical in RNN applications.

In practice, on a per-user basis, the questions that are used to estimate the knowledge of a particular user are fed one-by-one into the LSTM cell (eq. (6.1)), which creates a fixed-size representation of the user's knowledge, that is then mapped into the output space $\boldsymbol{g}_{\text{pred}}$. Between the LSTM layer and the output layer, we use a number of densely connected layers interleaved with dropout layers [203] in order to approximate the knowledge decoding transformations between the knowledge representation and the output item space.

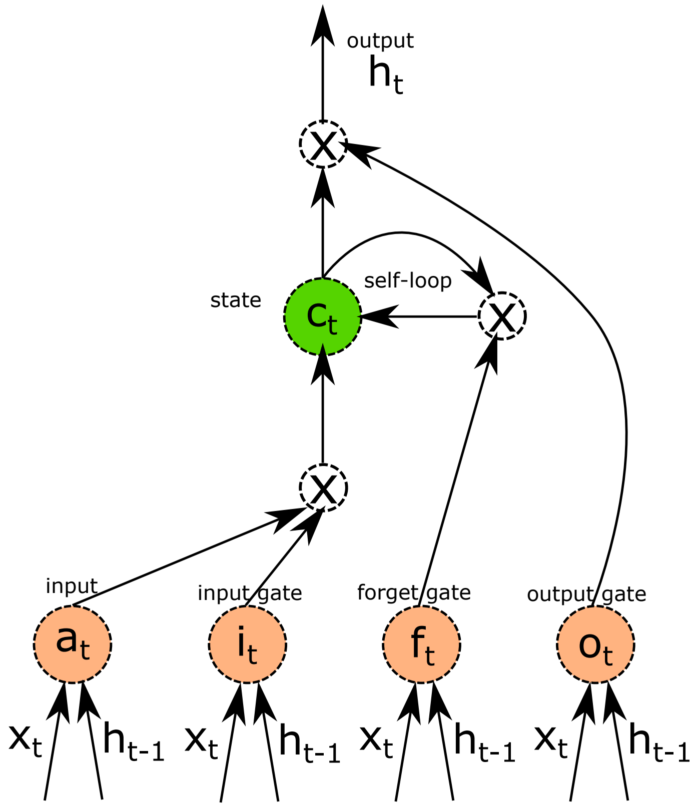


Figure 6.4: A functional overview of the LSTM cell. The input x_t and the output of the previous timestep h_{t-1} are modulated by the input activation function $f(x_t, x_{t-1}) = a_t$. The cell state c_t is updated by $c_t = i_t \cdot a_t + f_t \cdot c_{t-1}$, such that the previous state of the cell c_{t-1} is modulated by the forget gate f_t. The cell state is converted to the output h_t by the output modulation gate. All the gates have a form of non-linearity and weights and biases for the inputs.

We use a RNN as opposed to a dense network working on a fixed-size input in order to be able to give accurate estimations of the user's knowledge before the full sequence is seen. Nevertheless, the RNN needs to be optimised with a fixed input size N_{sequence}, so we pad the input matrix with zeroes for the missing inputs. The parameters to optimise in the model are the input, output and forget gate weights and biases and the weights of the dense layers connecting the LSTM output to the final output vector. As hyper-parameters, we optimise the size of the LSTM layer, the number and size of the intermediate dense layers and the amount of dropout that is applied. We use a sigmoid activation function in the output layer to guarantee an output $g_{\text{pred}} \in [0, 1]$, rectified linear units (ReLu) in the intermediate densely connected layers and a hyperbolic tangent function in the LSTM layer. The specific choice of activation functions in the intermediate layer did not have a significant effect on the performance of the model.

Care must be taken when computing the loss function that is minimised to optimise the weights. Since for any user, the known guess vector $\boldsymbol{g}_{\text{true}}$ (the prediction target) is sparse, we must compare only those prediction values for which the true value is known. Technically, we implement this by propagating the boolean mask of available answers as an auxiliary input along with the guess data and compute the binary cross-entropy loss only over available answers.

6.3.3 Model architecture

In order to feed the $(\text{LU}, \text{answer})$ pairs into the LSTM, we need to define an appropriate numerical representation of the LUs in the form of a fixed-length vector. As the set of all possible questions is known at training time, we could simply consider each LU as a unique symbol that can be embedded into a real vector space of dimension k by some appropriate function $f(n) \Rightarrow \mathbb{R}^k$. The simplest approach would be to represent each LU in the "one-hot" encoding with $k = n$, such that $f_k(n) = \delta_{nk}$. This suffers from a dimensionality problem, as the number of LUs can in principle be arbitrarily large.

In our case, each LU is also a word, or more precisely a word pair in the source-target languages, thus we could use an existing word embedding such as `word2vec` [204] or `GLOVE` [205]. These methods specify the embedding function based on contextual similarity that is extracted by data-mining a large corpus of text. However, a predefined word embedding would limit the model to only be useful for word-based question items, making it difficult to use data from other types of learning activities, such as grammar exercises or speech.

Another option would be to derive the embedding directly as part of the neural network optimisation procedure by using an embedding layer that reduces the dimensionality of the one-hot encoding. The advantage of this approach is that the embedding function $f(n)$ can be optimised as part of the model, thus, the theoretical performance is the highest. This was the first approach we adopted and overall, it showed promising results. However, since over the lifetime of the model the items in LU set can change due to modifications to the course structure, we have investigated an encoding which is resilient to such changes and does not need immediate re-optimisation upon adding new items.

Instead of an optimised embedding described above, we may use random vectors to represent the questions. Using compressed sensing, we can reconstruct any d-dimensional k-sparse signal using random vectors that have at most $r \simeq k \log d/k$ dimensions [206]. In our case, the dimensionality is the number of individual LU items ($N \simeq 5000$) or symbols that we want to represent and the sparsity is $k = 1$ for a one-hot encoding, thus, a uniform random vector of length 4-5 would be sufficient to represent all the word pairs in the full course. We illustrate the sparse coding method in fig. 6.5. In this approach, any new items would still be representable and the model would ideally return an average knowledge estimate for these unknown items.

As mentioned above, the contents of the course may change over the lifetime of the model. Therefore, we wish to ensure that any new LU could be represented by the chosen embedding. We choose the random vector approach, as we can easily use the unique `UUID4`-based identifiers of the LUs used in the internal database at Lingvist. The `UUID4` protocol prescribes universally unique identifiers for data in computer systems with negligible duplication probabilities based on pseudo-random bits. We do this by mapping the 128 pseudo-random bits of `UUID4` to 16 random 8-bit floats in the range $[-1, +1]$, as illustrated in fig. 6.6. With this method, it is straightforward to keep track of item (LU) to representation associations throughout the lifetime of the model, as

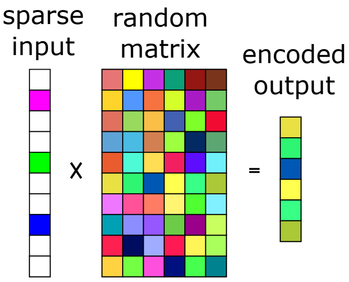


Figure 6.5: An illustration of the sparse coding technique using random projections. In this example, we encode a 3-sparse 10-dimensional signal to a 6-dimensional space using a matrix of Gaussian distributed weights.

they are always uniquely specified just by the item's identity in the item database. This practical necessity arose when implementing the model in the production system at Lingvist which evaluates the model predictions for the users in real time.

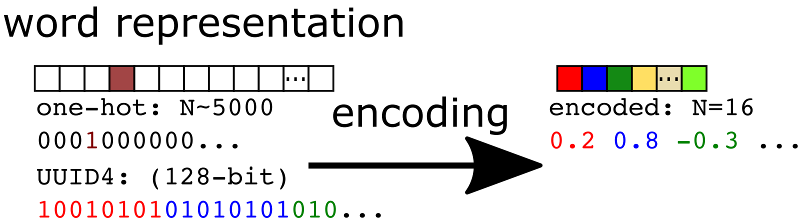


Figure 6.6: The word embedding scheme, where we use the 128-bit unique `UUID4` representation to construct an encoded representation of 16 8-bit floats in the range $[-1,+1]$. The actual bit strings and floating point values are illustrative.

For each user, we take up to a fixed number $N_{\mathrm{seq}} = 10\ldots100$ of guesses as the input. These guesses are then represented as a $N_{\mathrm{seq}} \times (r+1)$ matrix, where $r = 16$ based on the compressed sensing arguments discussed above. Each row of the matrix corresponds to a tuple of the LU representation and the guess value $g \in \{0,1\}$. The output of the model is the n-dimensional knowledge vector $\boldsymbol{g}$, such that we predict the static knowledge of all LUs, based on a short guess sequence. The overall structure of the neural network based model is shown in fig. 6.7, where the sequence size is typically $50\ldots100$, the fixed-size representation size is 32 and the target vector has a typical size of $\mathcal{O}(5000)$. For the output layer, the activation function has to be sigmoidal such that the outputs can be interpreted as probabilities.

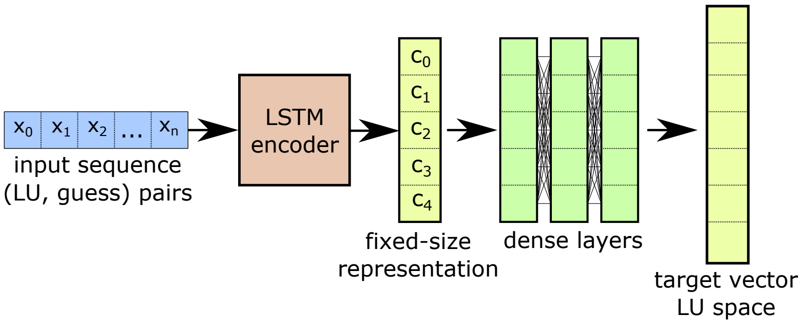


Figure 6.7: The overall architecture of the knowledge prediction neural network model. The arbitrary-length input sequence of (LU, answer) tuples is fed into a recurrent LSTM unit, which encodes the sequence into a fixed-size latent representation. This representation is decoded by a multi-layer densely connected feedforward unit into the final output space.

6.3.4 Optimisation

The loss function of a particular user m for the model optimisation is based on a binary cross-entropy between the predicted knowledge and the true guess for all the items for which this user has provided answers. It has the form

$$L_m = \sum_{n \in \mathrm{ans}_m} \mathrm{H}(y_n, g_n) \tag{6.2}$$

where $\mathbf{y} = (y_n)$ is the predicted knowledge vector, g_n is the guess (prediction target) for a given item n in the set of answered items ans_m for user m. The binary cross-entropy between two vectors $y_n \in [0,1]$ and $g_n \in [0,1]$ is defined as

$$\mathrm{H}(y_n, g_n) = -[g_n \ln y_n + (1 - g_n) \ln (1 - y_n)].$$

We note that in the loss function, we sum over all the answers, including the ones used in the input sequence. This does not pose a problem and serves as a cross-check on the model, where we expect the output for an item that was seen in the input to be predicted with high confidence.

We use the binary cross-entropy, since it can be shown that this leads to a maximum likelihood formulation of the problem. The binary cross-entropy is related to the Kullback-Leibler (KL) divergence, defined over the true probability distribution $g(x)$ and the assumed distribution $h(x)$ as

$$\mathcal{D}_{\mathrm{KL}}(g||h) = \mathbb{E}_g \ln \frac{g(x)}{h(x)} = \int g(x) \ln g(x) \ \mathrm{d}x - \int g(x) \ln h(x) \ \mathrm{d}x \geq 0 \tag{6.3}$$

through $\mathrm{H}(g,h) = \mathrm{H}(g) - \mathcal{D}_{\mathrm{KL}}(g||h)$ with $\mathrm{H}(g) = -\mathbb{E}_g \log g$ being the Shannon entropy. In the case of binary discrimination, the distributions are defined by a single probability value $g(1) = p$ and $g(0) = 1 - p$, such that the Shannon entropy is maximal for $p = 0.5$. The KL divergence is a discrepancy measure between the true and assumed distributions, which satisfies $\mathcal{D}_{\mathrm{KL}}(g||g) = 0$, but is not symmetric nor does it satisfy the triangle inequality, hence it is not a true distance measure.

We optimise the model with respect to the loss function over the whole dataset, using random subsamples (minibatches) of 100 users for about 10-100 epochs with the Adam stochastic gradient descent optimiser[§]. The Adam method estimates first and second moments of the gradient of the loss function and uses it to adaptively update the parameter estimates [207]. The model was implemented using the tensorflow package [208] and the optimisation was performed on a single AWS `g2.2xlarge` instance, with an epoch time around $\mathcal{O}(10\ \mathrm{s})$.

6.4 Results

In this section, we discuss the optimisation studies of the model and the overall performance as estimated on held-out data.

6.4.1 Studies of the RNN model

In general, the model performance has only a mild dependence on the choice of hyperparameters, introduced in section 6.3.2. Nevertheless, we scan the hyperparameter space by computing a cross-validated mean AUC at each point in a 5-dimensional space, consisting of the size of the LSTM layer, the size of the intermediate dense layers, the number of intermediate dense layers, the amount of dropout and the learning rate. We find that the hyper-parameter scan prefers the encoding by the LSTM to be rather small, around 32 to 64 units, and the intermediate dense layers are preferred to be small (128-256) and deep (3-4 layers). The final optimised loss function is shown in fig. 6.8, where we see from the loss on the training set that the model has sufficient generality to be able to fit features of the data, with the performance on the held-out test set being still acceptable after 10-15 epochs. We stopped the model training when the loss function on the test set had not decreased for five epochs, using the optimisation only up to the point when the loss decreased.

[§] We used a learning rate of $\alpha = 0.001$, exponential decay rates for the first and second moment $\beta_1 = 0.9$, $\beta_2 = 0.999$ and a learning rate decay of $d = 0.0$.

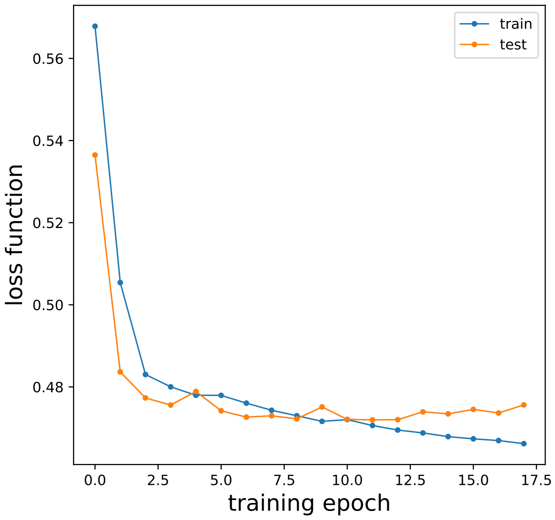


Figure 6.8: The loss function of the final optimisation of the model. The loss on the training set is shown in blue, the validation set in orange. In general, we see a relatively good convergence of the model after approximately 10-15 optimisation epochs. There is a small amount of over-training present after ten epochs, as visible from the increasing loss function on the test set.

One of the more important free parameters of the model is the length of the input guess sequence N_{seq}, introduced in section 6.3.3. While RNN-type networks can work with an input of arbitrary length, there is a clear trade-off between the amount of information provided to the model vs. the amount it needs to predict. We studied how this affects the performance of the model by retraining the model on the same dataset, selecting up to N_{seq} guesses for each user. In order not to bias the analysis, as users with a different number of guesses may have different knowledge characteristics, we optimise on the same underlying data where users have at least $N_{\mathrm{seq}} = 100$ guesses, masking the additional sequence elements. As expected, the model performs better with longer sequences, as more information is provided about the user. We observe a roughly linear relationship between the AUC and the guess sequence length, as can be seen in fig. 6.9. The performance measurement was carried out on an independent sub-sample that was held out from the model optimisation. We note that since the model predicts all items, including the ones that were seen in the question phase, we expect the overall true correct rate to increase with an increasing input sequence length.

Additionally, we have studied how the amount of training statistics affects the performance of the model. Interestingly, we see that with even as few as 1000 users, the model is able to approximate the knowledge characteristics of an average user. In general, adding more data improves the performance more significantly for words for which there are very few answers provided, as can be seen in fig. 6.10. Adding more data improves the model performance, especially for words with very low answer rates.

6.4.2 Discussion

As a first step, we compare the average performance of the model over all guesses to the reference toy model. We show the average performance of the RNN model in fig. 6.11 on a held-out dataset that was not seen during optimisation. In general, we see that the RNN model is more accurate than the average model, with the average true positive rate improving from 0.38 to 0.54 at an average false positive rate of 0.1 for the trial-weighted case, an improvement of about 40% in efficiency. This conclusion also holds if all LUs are given equal weight. This is not surprising, as the reference model has no discriminating information about the individual users or LUs. This increase in efficiency directly translates into reduced learning time, as the LUs correctly predicted can be defocused in the study period without losing the coverage of the course.

Furthermore, we studied some first-order statistics of the predictions, namely the per-user and

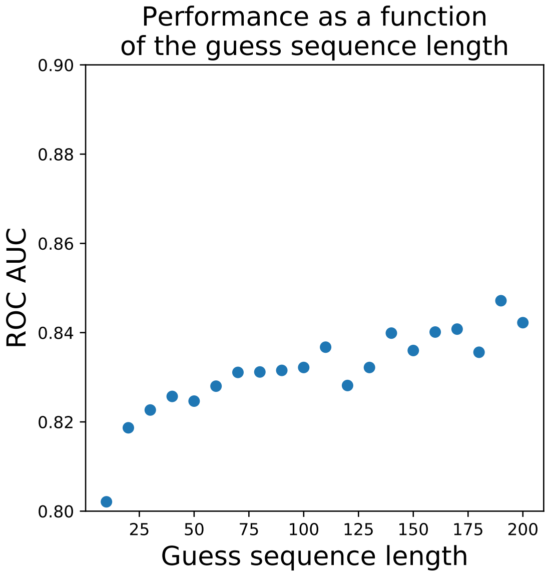


Figure 6.9: Performance of the model as a function of the guess sequence length. For each point, we use only up to N_{sequence} guesses from all users as the input to the model. The model is trained from scratch on the full dataset and evaluated in the same way for all cases.

per-LU averages. In a very simple model, each user (LU) can be characterized by their average correct rate, treating all LUs (users) as equal. More sophisticated treatments based on item response theory would also be possible, where the skills assigned to users and difficulties assigned to items are determined using a maximum likelihood procedure [209]. In fig. 6.12 we see that the average per-LU and per-user correct rates are also reproduced to a good degree by the model, with linear correlation coefficients above 90%. We can see that the correct rate for users and LUs with a low (high) average correct rate are slightly under (over) estimated. This could result from the model not having enough counterexamples to sufficiently learn the probability distribution $p(\boldsymbol{g})$ for very simple or difficult LUs and users with very low or high correct rate. We take this as an example that the model is able to summarise the user and LU population and approximate the mean correct rates.

We can further study the average performance of the model on a per-LU, per-user basis as a function of number of guesses per LU. In fig. 6.13 we see that the overall performance of the model is strongly dependent on how many guess attempts a given LU has, with the average ROC AUC dropping significantly for rarer LUs. This performance degradation is expected, as we can see from fig. 6.3 that the number of answers per word drops steeply going further in the course, reflecting that this model was trained treating each trial equivalently. With increasing guess statistics, it may be worthwhile to investigate a re-weighting or re-sampling of the trials such that words that have fewer trials have a higher weight and would thus still be optimised for in the training. This trend also introduces a natural cut-off in terms of when the predictions from the model are no longer considered reliable at a priority index of around 4000 for LUs. The priority index arises from an ordering of the LUs by when they appear in the course on average and can simply be considered a specific way to order the LUs.

Additionally, we studied the performance as a function of the LU average guess rate, which can be interpreted as the LU difficulty. We see in fig. 6.14 that the performance of the model is broadly similar over a wide range of difficulties. For words that are very easy or difficult, the variance in the trained model is considerable, since the model has less counterexamples for optimisation. Furthermore, we see that the LUs with vanishing performance all have 0 correct guesses in data, meaning they can be excluded from the predictions.

Prediction uncertainty We have also investigated the prediction uncertainty arising from the model in terms of sensitivity to the inputs. In particular, for a word that is predicted to be known

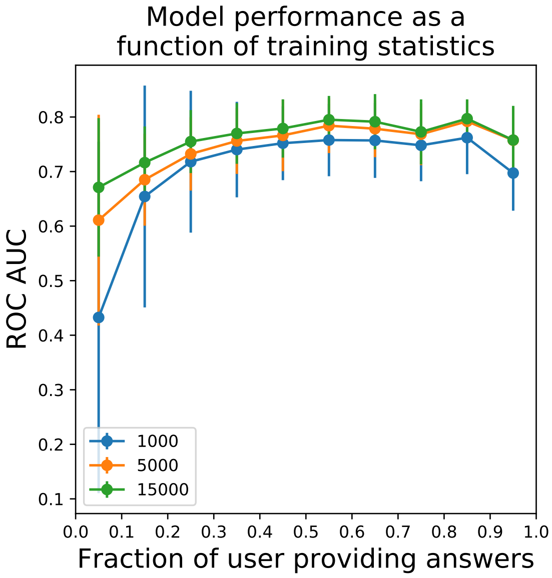


Figure 6.10: The effect of the amount of training data on the model performance We investigate three cases: 1000 users, 5000 users, 15000 users (full dataset) and plot the performance characteristics of the model (ROC AUC), averaged over words that have a certain fraction of users providing answers. Overall, we see that the model performance increases significantly for words with few answers when adding training data.

at a knowledge likelihood of $y_m = 0.9$, as an example, how certain is the model of this prediction and how much would it vary if the input sequence changed slightly? For neural networks, it has been shown that using dropout in the evaluation phase is a good approximation for the inherent sensitivity [210]. In fig. 6.15, we can see the evaluated uncertainty of the guess probability using ten dropout rounds, compared to the data, in terms of the running mean of the average correct rate as a function of the LU priority index. It is interesting to see that the 3σ confidence interval around the predicted mean covers the data, however, the predicted mean is systematically lower than the true guess probability as a function of the priority index. This systematic underestimation clearly shows that while sensitivity to inputs accounts for a large fraction of the variability, there is room for improvement in terms of the bias of the prediction.

6.4.3 Future work

As possible future improvements, extending the information in the LU feature vector beyond a random representation may prove to be a straightforward way to improve the performance of the model. In particular, the random representation could be augmented with features such as difficulty or context based on expert labelling. This is particularly important, as the LUs are embedded in sentences, so context priming effects could be significant [211].

Additionally, the Deep Knowledge Tracing approach was originally conceived for time-dependent knowledge modelling and the current framework is naturally amenable to predicting the behaviour of the guess sequences in time, such that effects of acquisition and forgetting could naturally be incorporated.

Furthermore, the current model works only on a single language pair, but it could easily be extended to multiple simultaneous languages by adding language data to the source representation vector and decoding the internal knowledge representation into different target language spaces. This would allow training a single model on all language pairs simultaneously, which could help bootstrapping new courses that don't yet have enough user interaction data.

Interpretable models Having optimised a model that is capable of estimating the knowledge vector $\boldsymbol{g}$ based on a subset of guesses, it is interesting to study the explanatory factors of that

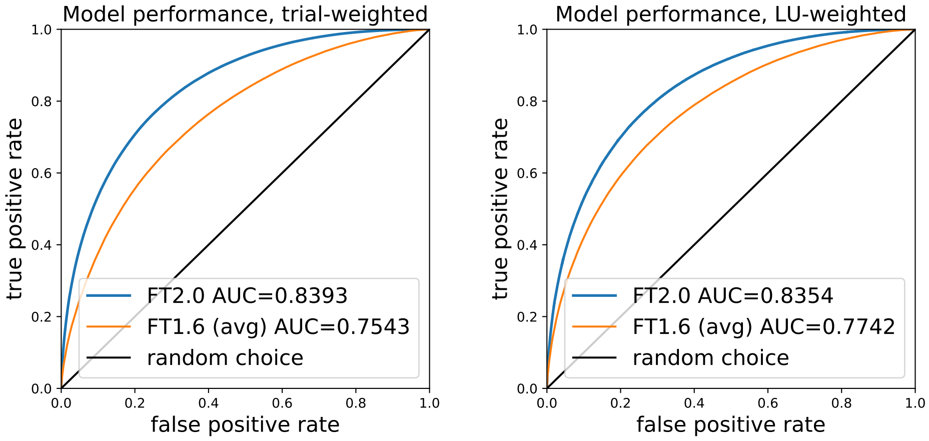


Figure 6.11: Overall average performance of the model, as characterised by the receiver operating characteristic (false positive to true positive rate curve) area under curve (ROC AUC). On the left, we show the average treating all trials equally, on the right, the average by weighting trials such that each LU is treated equally. We see that the RNN model (blue) outperforms the average model (orange).

model. In general, explaining the decisions or predictions of complex "black box" models is a topic of intense research activity [212]. Recently, locally interpretable linear models (LIME) have been proposed as a way of explaining the features a model makes use of for particular prediction outcomes [213]. In the LIME approach, any classifier or regressor predictions can be explained in a locally faithful way by fitting local linear approximations to the decision. We have used the method to validate our optimised vocabulary model on a few examples and see that it uses guess information in an intuitively correct way, as can be seen in section 6.4.3. We mention this as a possible direction for future research, as understanding the factors in a prediction by a complex model is an essential step in validation.

6.5 Summary

We showed that by formulating the problem of predicting existing knowledge as a sequence-to-vector binary classification problem, RNNs present a simple and flexible solution with a wide range of applicability. This method outperforms the existing reference implementation at Lingvist based on average correct rates by about 45% in terms of true positive rate at a fixed false positive rate. In order to achieve this result, we formalized the reference model based on average guess accuracies and formulated the vocabulary prediction problem as binary classification. We found that sparse coding from random bits in database identifiers can be used to represent categorical items in a practical and resilient fashion.

Moreover, we saw that a single flexible model architecture based on recursive neural networks is able to predict knowledge probabilities of thousands of words based on just a small amount of randomly chosen guesses. The RNN model is easily extensible should more information about LUs or users become available. In particular, we see extension to predicting knowledge in multiple languages simultaneously or incorporating time-dependent information as possible future improvements.

Furthermore, we studied the reliability of the predictions by evaluating the sensitivity of the model to the inputs using a sampling technique, which allows to attach uncertainties to predictions, and by deriving locally-interpretable models around predictions. Using this sensitivity analysis in HEP, where multivariate predictions are commonplace, would allow the reliability of these results to be studied from a new perspective.

Overall, in cases where the amount of available data is significant but the underlying theory is not well known from first principles, such as language acquisition, the use of ML for modelling and prediction tasks can provide significant advantages in terms of flexibility and speed over approaches that require expert annotations, such as Bayesian Knowledge Tracing.

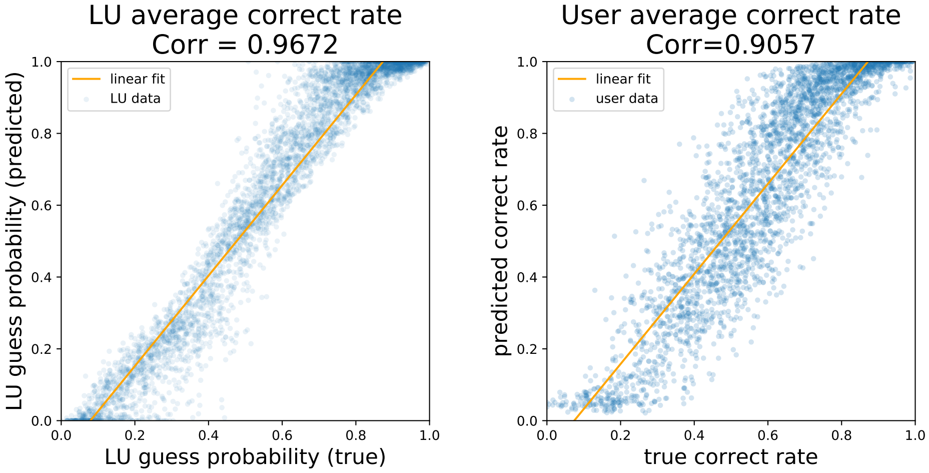


Figure 6.12: Predicted vs. true average guess rate by LU (left) and user (right). We see that the model performs best for LUs which have a true correct rate around 50%, with slight underestimation at low correct rates and overestimation at high true correct rates.

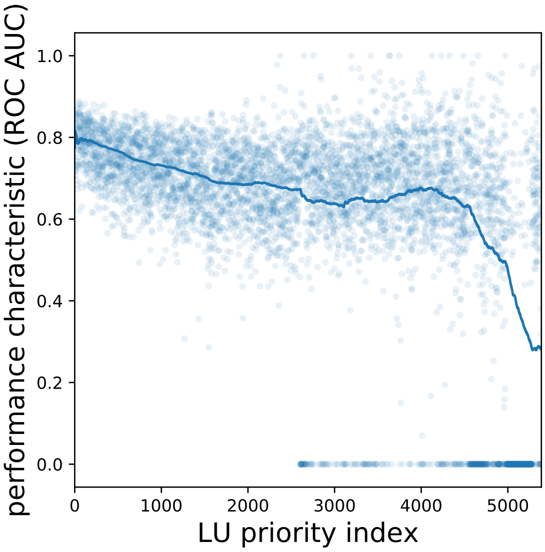


Figure 6.13: The performance characteristic (ROC AUC) as a function of lexical unit priority index. The priority index is a monotonously growing index that roughly corresponds to the order in which words are shown to users, with lower indices being shown earlier and more often. The words with a ROC AUC of 0 or 1 correspond to outliers where all the answers are either correct or incorrect. We show the individual lexical units (blue dots), as well as the running mean of the distribution with a window size of 100 (blue line).

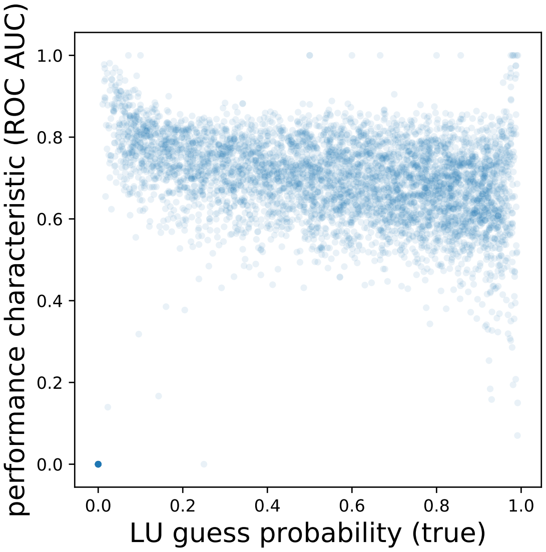


Figure 6.14: The performance of the model on a per-word basis as a function of the guess probability (prior difficulty) of the word. We see that for words with very high or low guess probability, the variance of the model is significant, resulting from few training examples.

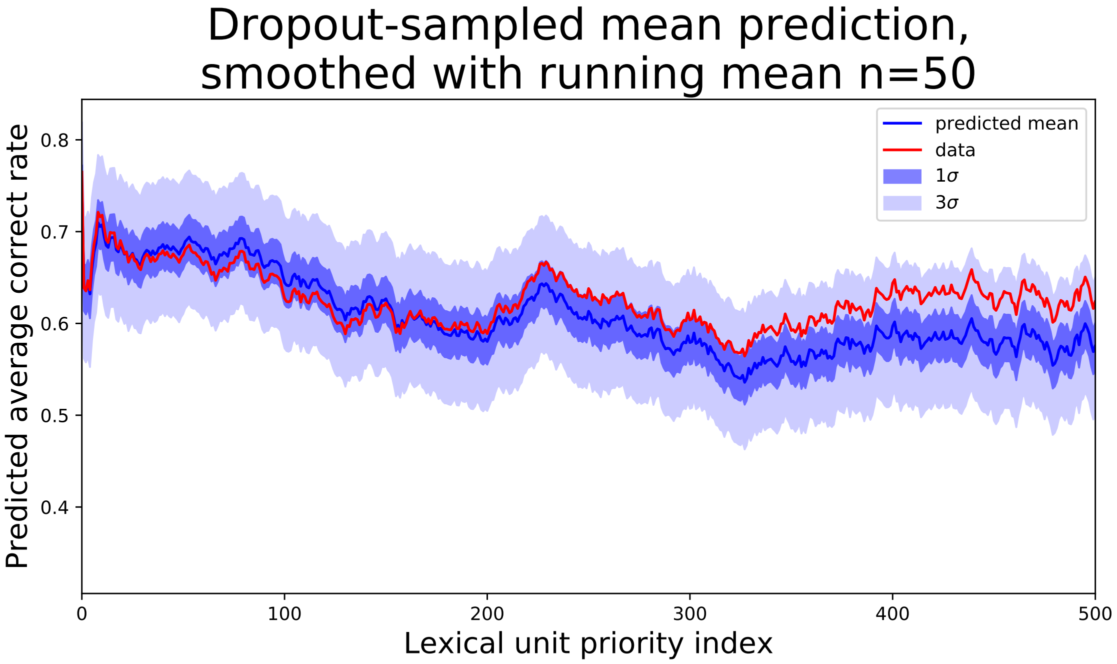


Figure 6.15: The predicted (blue) and true (red) guess probability as a function of LU priority index for the first 500 LUs. We see that the prediction is generally sufficiently close to the true mean correct rate, but trends somewhat lower than the data.

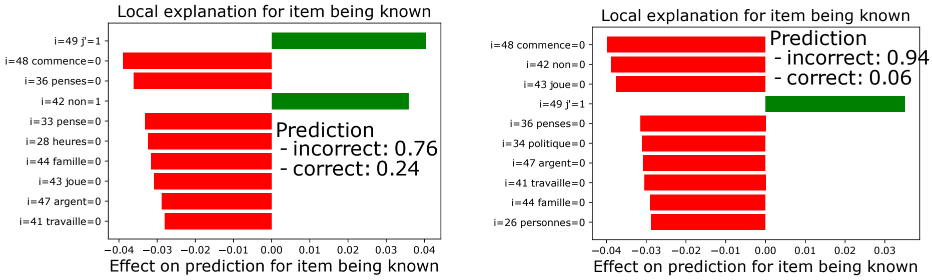


Figure 6.16: Local explanations for a word that was known (left) and not known (right). In the case where the word was known, the algorithm predicts the knowledge likelihood to be $g = 0.24$, based on the correct guesses of the LUs *j'* and *non*, but incorrect guesses for most other words in the trial set. In the case where the word was not known on the right, the model predicts the guess likelihood to be $g = 0.06$ based on incorrect guesses for most words. We see that incorrect guesses ($= 0$, red) tend to affect the prediction in a negative way, whereas correct guesses ($= 1$, green) tend to reinforce the knowledge, as would be expected.

7 Conclusions and outlook

The SM is a successful and predictive theory of high-energy processes. Nevertheless, it is clearly not a complete theory of the physical Universe. Among the open questions, the nature of the electroweak symmetry breaking mechanism needs to be clarified experimentally. Deviations between the couplings of the recently-discovered Higgs boson and SM particles can arise in BSM scenarios and thus must be tested experimentally. The Large Hadron Collider and the corresponding general purpose detector experiments, CMS and ATLAS, are in a unique position to clarify the nature of this fundamental scalar. The couplings of the Higgs boson to vector bosons have been determined in Run 1 of the LHC to a relative precision of $\simeq 10\%$. In Run 2, thanks to the increased centre-of-mass energy and luminosity, the couplings to the fermion sector can be explored in more detail.

The focus of this thesis was the search for the top quark pair associated Higgs production mode ($\mathrm{t\bar{t}H}$) in the channels where at least one of the top quarks decays leptonically and the Higgs boson decays to bottom quarks. This is a sensitive channel which contributes to the overall $\mathrm{t\bar{t}H}$ cross-section measurement and thus the determination of the top quark Yukawa coupling. However, the analysis is challenging due to the presence of an irreducible background arising from the QCD production of top quark pairs and the complex multi-jet final state with a combinatorial ambiguity between the jets. Multivariate analysis techniques that exploit the differences in the dynamics of the signal and background processes are necessary to achieve sufficient sensitivity. Traditionally, LHC analyses have employed machine learning techniques, where large amounts of MC simulation are used to construct a signal-to-background classifier. This approach works well and has been essential for many of the results of the LHC experiments. However, for final states with a large number of jets, it is challenging to generate sufficient MC simulation events that pass the experimental cuts in the most sensitive regions of the phase space, making it difficult to optimise the machine learning algorithms. Such predictions also have significant theoretical uncertainties arising from the challenges of matching the hard process to the parton shower and the multiple momentum scales present in the hard process.

We have investigated an alternative approach for analysis, where the signal and background processes are distinguished by making use of the matrix elements directly at the event observable level. We made significant contributions to the matrix element method as applied to $\mathrm{t\bar{t}H}(\rightarrow \mathrm{b\bar{b}})$ in CMS and this approach is now standard within the collaboration for this analysis. We have also carried out a complete search for $\mathrm{t\bar{t}H}(\rightarrow \mathrm{b\bar{b}})$ in the leptonic channels in the 2016 data collected by CMS using the matrix element method. We have not observed any significant excess and have thus established an upper limit on the signal strength factor $\mu = \sigma_{\mathrm{t\bar{t}H}}/\sigma^{\mathrm{SM}}_{\mathrm{t\bar{t}H}}$ at 1.52 (1.57 expected) at a confidence level of 95%. The upcoming years and the data collected by the High-Luminosity LHC project promise to be particularly interesting for $\mathrm{t\bar{t}H}$, with a discovery of this process being within reach in the next years. On the other hand, theoretical and experimental uncertainties are significant in this analysis. The former can be reduced by using more accurate theoretical models, which are increasingly becoming available. There, the matrix element method can be a useful way to incorporate latest predictions into the analysis. For the latter, the analysis techniques, in particular the reconstruction of jets and the identification of jets from bottom quarks, need to be refined.

For object reconstruction and identification, machine learning presents a useful way for combining signals across various channels. We have investigated a method for b jet identification where several independently optimised b discriminators based on different vertexing algorithms, track properties and presence of leptonic decays of hadrons are combined into a multivariate super-discriminator between jets arising from bottom quarks, light quarks and charm flavoured quarks. This approach presented an improvement over the state-of-the art at CMS and was deployed during the 2016 data taking and used in several physics analyses. The success of this discriminator relied partly on benefiting from the developments made in the field of data science outside of high-energy physics. Currently, such discriminators rely fully on simulation of the hard scattering, the subsequent showering and hadronisation and the detector simulation and thus are assigned significant uncertainties based on data-to-simulation corrections. An interesting direction for research would be to reduce the sensitivity of such discriminators by making use of the theoretical and experimental uncertainties during the optimisation phase. The direct use of collider data together with the appropriate simulation in a semi-supervised learning environment may be possible to reduce model dependence.

Although the SM of particle physics predicts many observable quantities accurately from a relatively small set of principles and parameters, there are features that are less understood due to the presence of non-perturbative physics, such as hadronisation or the precise parton content of protons in high-energy processes. In such cases, modelling the observed data in a phenomenological way and performing global fits presents a way to make testable predictions, as has been evidenced by the usefulness of effective hadronisation models and PDF fits for LHC experiments. During an internship at the private company Lingvist Technologies, we investigated and developed such methods in the field of human cognition and language learning. We proposed a data-driven model based on deep learning that predicted the performance of learners in a vocabulary test. Using such predictions, it is possible to optimise the learning process and make it personal on a large scale. Our proposed model out-performed the existing approach employed by the learning environment offered by the company. Such data-driven methods may also be useful in physics for problems where our understanding is not yet sufficient to have accurate predictions based on well-understood theory. However, it remains an open question as to how to best develop these phenomenological models such that they take into account the physical principles that have been well established by experiments.

In conclusion, the successful operation of the LHC and the detectors has opened up a new program in fundamental physics, making it possible to experimentally study the properties of the recently discovered Higgs boson, the only known fundamental scalar.

List of Figures

1.1 The particle content of the Standard Model 6
1.2 The basic QED vertex and representative processes 8
1.3 The measured values of $\alpha_S(Q^2)$ compared the perturbative QCD prediction 12
1.4 An illustration of the factorisation of a proton-proton interaction 12
1.5 The parton distribution functions for protons. 13
1.6 Tree-level Feynman diagrams for Higgs production 15
1.7 The production cross-section of the Higgs boson 15
1.8 The branching ratios of the Higgs boson 16
1.9 The Higgs signal strength modifiers as measured by CMS and ATLAS 17

2.1 The LHC accelerator complex 22
2.2 Cross-section of the LHC dipole magnet 22
2.3 The cross-section of the CMS experiment. 24
2.4 A view of a transverse slice of the CMS detector with the subsystems. 25
2.5 The CMS pixel tracker 26
2.6 The CMS strip tracker 27
2.7 The structure of the CMS electromagnetic calorimeter 28
2.8 The cross-section of the CMS hadronic calorimeter 28

3.1 Illustration of b quark decay 34
3.2 The jet probability b discriminator distributions 35
3.3 The soft lepton b discriminator distributions 35
3.4 The CSVv2 b tagger discriminator distributions 36
3.5 The performance of the CMS b taggers in Run 2 37
3.6 Comparison of jet kinematics in MC simulation by jet flavour 38
3.7 The correlation between the CSV b discriminators using AVR or IVF vertices . . . 39
3.8 The cMVAv2 BDT loss function as a function of boosting iteration 40
3.9 The cMVAv2 performance as a function of training variables. 41
3.10 The cMVAv2 mistagging probability as a function of jet kinematics. 42

4.1 The double-Gaussian transfer function fit in $\mathrm{t\bar{t}}$ simulation. 50
4.2 The dependence of the transfer function fit parameters on jet kinematics 51
4.3 The effect of combinations assumptions in the MEM 55
4.4 VEGAS integration grid example 56
4.5 MEM integration time. 56
4.6 The distributions of the MEM signal and background probabilities 58
4.7 The full variation in the MEM discriminator 59
4.8 The MEM discriminator performance as a function of κ 60
4.9 The distributions of the MEM signal and background probabilities 60
4.10 MEM with the $0_\mathrm{W}2_\mathrm{h}2_\mathrm{t}$ interpretation in the 4 jet, 4 b tag category 62
4.11 MEM with the $0_\mathrm{W}2_\mathrm{h}2_\mathrm{t}$ interpretation in the 4 jet, 3 b tag category 63
4.12 MEM with the $1_\mathrm{W}2_\mathrm{h}2_\mathrm{t}$ interpretation in the 5 jet, 3 b tag category 64
4.13 MEM with the $1_\mathrm{W}2_\mathrm{h}2_\mathrm{t}$ interpretation in the 5 jet, $\geq$ 4 b tags category 65
4.14 MEM with the $2_\mathrm{W}2_\mathrm{h}2_\mathrm{t}$ interpretation in the $\geq$ 6 jet, 3 b tag category 65
4.15 MEM with the $2_\mathrm{W}2_\mathrm{h}2_\mathrm{t}$ interpretation in the $\geq$ 6 jet, $\geq$ 4 b tag category 66
4.16 Performance of the MEM for different $\mathrm{t\bar{t}}$+jets subprocesses 66
4.17 MEM with the $1_\mathrm{W}2_\mathrm{h}2_\mathrm{t}$ interpretation in the $\geq$ 6 jet, $\geq$ 4 b tag category 67
4.18 MEM with the $2_\mathrm{W}2_\mathrm{h}2_\mathrm{t}$ interpretation in the $\geq$ 7-jet, $\geq$ 4 b tag category 67
4.19 MEM with the $2_\mathrm{W}2_\mathrm{h}2_\mathrm{t}1_\mathrm{g}$ interpretation in the $\geq$ 7 jet, $\geq$ 4 b tag category 68

4.20 MEM with the $2_\mathrm{h}2_\mathrm{t}$ interpretation in the dileptonic ≥ 4 jet, ≥ 4 b tag category . . 68
4.21 MEM with the $2_\mathrm{h}2_\mathrm{t}$ interpretation in the dileptonic ≥ 4 jet, 3 b tag category . . . 69

5.1 Representative leading-order Feynman diagrams for the $\mathrm{t\bar{t}H}(\rightarrow \mathrm{b\bar{b}})$ process 71
5.2 Modelling of the $\mathrm{t\bar{t}}$+X sub-processes 74
5.3 $\mathrm{t\bar{t}H}(\rightarrow \mathrm{b\bar{b}})$ leptonic final states 75
5.4 Expected performance of the b tag likelihood ratio discriminant. 78
5.5 Fraction of events with correct matching to the hard process 79
5.6 The expected signal and background yields in the analysis categories 80
5.7 An illustration of the maximum likelihood template fit 81
5.8 Effect of jet energy corrections. 85
5.9 b tagging reweighting and uncertainty 86
5.10 The effect of b tagging uncertainties on the b tagging likelihood ratio. 87
5.11 The effect of μ_r and μ_f variations. 88
5.12 The $\mathrm{t\bar{t}}$+jets modelling uncertainties 89
5.13 The normalisation and shape effect of uncertainties 90
5.14 The modelling of the kinematic distributions in the semileptonic control region . . 91
5.15 The modelling of the kinematic distributions in the dileptonic control region 92
5.16 Fit model validation on toy experiments 95
5.17 The pulls and constraints of the combined fit model with the Asimov datasets . . . 96
5.18 The best-fit estimations of nuisance parameters as a function of the $\theta_{\mathrm{t\bar{t}+b\bar{b}}}$ nuisance parameter 97
5.19 The post-fit pulls 98
5.20 The likelihood with respect to the minimum as a function of μ 99
5.21 The pre-fit and post-fit distributions in the semileptonic 4-jet categories 100
5.22 The pre-fit and post-fit distributions in the semileptonic 5-jet categories 101
5.23 The pre-fit and post-fit distributions in the semileptonic $\geq$ 6-jet categories 102
5.24 The pre-fit and post-fit distributions in the dileptonic categories 103
5.25 The post-fit correlation between the signal strength parameter and the nuisance parameters 104
5.26 The best-fit value and the 95% upper limits on the signal strength parameter . . . 104
5.27 The final analysis bins, arranged by the signal over background ratio S/B 105
5.28 The expected and observed upper limits on the signal strength parameter. 106
5.29 The impact of nuisance parameters on the signal strength. 108

6.1 An example exercise in the Lingvist environment 114
6.2 Word pair (LU) guess data for a subset of users and words 115
6.3 Distributions of guesses by LU and user 116
6.4 The long short-term memory cell 118
6.5 An illustration of sparse coding 120
6.6 The word embedding scheme 120
6.7 Overall network architecture 120
6.8 Loss function for the knowledge estimation model 122
6.9 Performance as a function of sequence length 123
6.10 The effect of training statistics on the model performance 124
6.11 Overall knowledge estimation model performance 125
6.12 Predicted vs. true average guess rate 126
6.13 Performance of knowledge estimation as a function of LU priority 126
6.14 Model performance on a per-LU basis as a function of average LU guess probability 127
6.15 The sensitivity estimation of the model using dropout sampling 127
6.16 Local explanations of the model prediction 128

List of Tables

1.1 Current results on $t\bar{t}H$ . . . 18

3.1 The training and validation samples for cMVAv2 . . . 37

4.1 The MEM event interpretations considered for different final state topologies . . . 53
4.2 The number of transfer function combinations in different categories . . . 54
4.3 The CPU budget and separation of the MEM in different categories . . . 57
4.4 The fraction of matched events in the analysis categories . . . 61

5.1 The MC samples used in the $t\bar{t}H(\rightarrow b\bar{b})$ analysis . . . 72
5.2 The data samples used in the $t\bar{t}H(\rightarrow b\bar{b})$ analysis . . . 73
5.3 The selection and ID criteria for the charged leptons. . . . 75
5.4 The analysis categories for the $t\bar{t}H(\rightarrow b\bar{b})$ analysis. . . . 82
5.5 Systematic uncertainties in the $t\bar{t}H(\rightarrow b\bar{b})$ analysis. . . . 83
5.6 The best-fit values and upper limits per category . . . 99
5.7 The post-fit uncertainties in the $t\bar{t}H(\rightarrow b\bar{b})$ analysis . . . 107
5.8 Comparison of the CMS and ATLAS results on $t\bar{t}H(\rightarrow b\bar{b})$ . . . 110

Bibliography

[1] Wikimedia Commons. *Standard Model of Elementary Particle Physics*. https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles_modified_version.svg. [Online; accessed 05-Oct-2017]. 2017 (cit. on p. 6).

[2] Eugene Wigner. “On unitary representations of the inhomogeneous Lorentz group”. In: *Annals of mathematics* (1939), pp. 149–204 (cit. on p. 5).

[3] Wolfgang Pauli. “The connection between spin and statistics”. In: *Physical Review* 58.8 (1940), p. 716 (cit. on p. 5).

[4] Carl D Anderson. “The positive electron”. In: *Physical Review* 43.6 (1933), p. 491 (cit. on p. 6).

[5] Julian Schwinger. “On Quantum-Electrodynamics and the Magnetic Moment of the Electron”. In: *Phys. Rev.* 73 (4 Feb. 1948), pp. 416–417. DOI: 10.1103/PhysRev.73.416. URL: http://link.aps.org/doi/10.1103/PhysRev.73.416 (cit. on p. 7).

[6] Julian Schwinger. “Quantum Electrodynamics. I. A Covariant Formulation”. In: *Phys. Rev.* 74 (10 Nov. 1948), pp. 1439–1461. DOI: 10.1103/PhysRev.74.1439. URL: http://link.aps.org/doi/10.1103/PhysRev.74.1439 (cit. on p. 7).

[7] R. P. Feynman. “Space-Time Approach to Quantum Electrodynamics”. In: *Phys. Rev.* 76 (6 Sept. 1949), pp. 769–789. DOI: 10.1103/PhysRev.76.769. URL: http://link.aps.org/doi/10.1103/PhysRev.76.769 (cit. on p. 7).

[8] R. P. Feynman. “Space-time approach to nonrelativistic quantum mechanics”. In: *Rev. Mod. Phys.* 20 (1948), pp. 367–387. DOI: 10.1103/RevModPhys.20.367 (cit. on p. 7).

[9] Joshua Ellis. “TikZ-Feynman: Feynman diagrams with TikZ”. In: *Comput. Phys. Commun.* 210 (2017), pp. 103–123. DOI: 10.1016/j.cpc.2016.08.019. arXiv: 1601.05437 [hep-ph] (cit. on p. 8).

[10] T. Aoyama et al. “Revised value of the eighth-order QED contribution to the anomalous magnetic moment of the electron”. In: *Phys.Rev.* D77 (2008), p. 053012. DOI: 10.1103/PhysRevD.77.053012. arXiv: 0712.2607 [hep-ph] (cit. on p. 7).

[11] C. S. Wu et al. “Experimental Test of Parity Conservation in Beta Decay”. In: *Phys. Rev.* 105 (1957), pp. 1413–1414. DOI: 10.1103/PhysRev.105.1413 (cit. on p. 9).

[12] S.L. Glashow. “Partial Symmetries of Weak Interactions”. In: *Nucl.Phys.* 22 (1961), pp. 579–588. DOI: 10.1016/0029-5582(61)90469-2 (cit. on p. 9).

[13] Steven Weinberg. “A Model of Leptons”. In: *Phys. Rev. Lett.* 19 (21 Nov. 1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264. URL: http://link.aps.org/doi/10.1103/PhysRevLett.19.1264 (cit. on p. 9).

[14] Abdus Salam. “Weak and Electromagnetic Interactions”. In: *Conf.Proc.* C680519 (1968), pp. 367–377 (cit. on p. 9).

[15] F. J. Hasert et al. “Observation of Neutrino Like Interactions Without Muon Or Electron in the Gargamelle Neutrino Experiment”. In: *Phys. Lett.* 46B (1973), pp. 138–140. DOI: 10.1016/0370-2693(73)90499-1 (cit. on p. 10).

[16] G. Arnison et al. “Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c**2 at the CERN SPS Collider”. In: *Phys. Lett.* 126B (1983), pp. 398–410. DOI: 10.1016/0370-2693(83)90188-0 (cit. on p. 10).

[17] M. Z. Akrawy et al. “Measurement of the Z^0 Mass and Width with the OPAL Detector at LEP”. In: *Phys. Lett.* B231 (1989), pp. 530–538. DOI: 10.1016/0370-2693(89)90705-3 (cit. on p. 10).

[18] Peter W. Higgs. “Broken symmetries, massless particles and gauge fields”. In: *Phys. Lett.* 12 (1964), pp. 132–133. DOI: 10.1016/0031-9163(64)91136-9 (cit. on p. 10).

[19] Peter W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. In: *Phys.Rev.Lett.* 13 (1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508 (cit. on p. 10).

[20] F. Englert and R. Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons”. In: *Phys. Rev. Lett.* 13 (1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321 (cit. on p. 10).

[21] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. “Global Conservation Laws and Massless Particles”. In: *Phys. Rev. Lett.* 13 (1964), pp. 585–587. DOI: 10.1103/PhysRevLett.13.585 (cit. on p. 10).

[22] S. Schael et al. “Precision electroweak measurements on the Z resonance”. In: *Phys. Rept.* 427 (2006), pp. 257–454. DOI: 10.1016/j.physrep.2005.12.006. arXiv: hep-ex/0509008 [hep-ex] (cit. on p. 10).

[23] *First combination of Tevatron and LHC measurements of the top-quark mass.* Tech. rep. arXiv:1403.4427. ATLAS-CONF-2014-008. CDF-NOTE-11071. CMS-PAS-TOP-13-014. D0-NOTE-6416. Comments: 34 pages, 7 figures. Geneva: CERN, Mar. 2014. URL: https://cds.cern.ch/record/1669819 (cit. on p. 10).

[24] Nicola Cabibbo. “Unitary Symmetry and Leptonic Decays”. In: *Phys. Rev. Lett.* 10 (1963). [648(1963)], pp. 531–533. DOI: 10.1103/PhysRevLett.10.531 (cit. on p. 10).

[25] Makoto Kobayashi and Toshihide Maskawa. “CP Violation in the Renormalizable Theory of Weak Interaction”. In: *Prog. Theor. Phys.* 49 (1973), pp. 652–657. DOI: 10.1143/PTP.49.652 (cit. on p. 10).

[26] Werner Bernreuther. “Top quark physics at the LHC”. In: *J. Phys.* G35 (2008), p. 083001. DOI: 10.1088/0954-3899/35/8/083001. arXiv: 0805.1333 [hep-ph] (cit. on p. 10).

[27] The CDF Collaboration. “Observation of top quark production in $\bar{p}p$ collisions”. In: *Phys. Rev. Lett.* 74 (1995), pp. 2626–2631. DOI: 10.1103/PhysRevLett.74.2626. arXiv: hep-ex/9503002 [hep-ex] (cit. on pp. 11, 33).

[28] Michal Czakon, Paul Fiedler, and Alexander Mitov. “Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha\frac{4}{S})$”. In: *Phys. Rev. Lett.* 110 (2013), p. 252004. DOI: 10.1103/PhysRevLett.110.252004. arXiv: 1303.6254 [hep-ph] (cit. on p. 11).

[29] J. D. Bjorken and Emmanuel A. Paschos. “Inelastic Electron Proton and gamma Proton Scattering, and the Structure of the Nucleon”. In: *Phys. Rev.* 185 (1969), pp. 1975–1982. DOI: 10.1103/PhysRev.185.1975 (cit. on p. 11).

[30] David J. Gross and Frank Wilczek. “Ultraviolet Behavior of Non-Abelian Gauge Theories”. In: *Phys. Rev. Lett.* 30 (26 June 1973), pp. 1343–1346. DOI: 10.1103/PhysRevLett.30.1343. URL: https://link.aps.org/doi/10.1103/PhysRevLett.30.1343 (cit. on p. 11).

[31] H. David Politzer. “Reliable Perturbative Results for Strong Interactions?” In: *Phys. Rev. Lett.* 30 (26 June 1973), pp. 1346–1349. DOI: 10.1103/PhysRevLett.30.1346. URL: https://link.aps.org/doi/10.1103/PhysRevLett.30.1346 (cit. on p. 11).

[32] F. Herren and M. Steinhauser. “Version 3 of {\tt RunDec} and {\tt CRunDec}”. In: *ArXiv e-prints* (Mar. 2017). arXiv: 1703.03751 [hep-ph] (cit. on p. 12).

[33] C. Patrignani et al. “Review of Particle Physics”. In: *Chin. Phys.* C40.10 (2016), p. 100001. DOI: 10.1088/1674-1137/40/10/100001 (cit. on pp. 12–16, 26).

[34] John C Collins et al. *Perturbative quantum chromodynamics.* 1989 (cit. on p. 11).

[35] M. Diemoz et al. “Parton Densities from Deep Inelastic Scattering to Hadronic Processes at Super Collider Energies”. In: *Z. Phys.* C39 (1988), p. 21. DOI: 10.1007/BF01560387 (cit. on p. 11).

[36] Richard D. Ball et al. “Parton distributions for the LHC Run II”. In: *JHEP* 04 (2015), p. 040. DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849 [hep-ph] (cit. on p. 11).

[37] Guido Altarelli and G. Parisi. “Asymptotic Freedom in Parton Language”. In: *Nucl. Phys.* B126 (1977), pp. 298–318. DOI: 10.1016/0550-3213(77)90384-4 (cit. on p. 11).

[38] Yuri L. Dokshitzer. “Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics.” In: *Sov. Phys. JETP* 46 (1977). [Zh. Eksp. Teor. Fiz.73,1216(1977)], pp. 641–653 (cit. on p. 11).

[39] V. N. Gribov and L. N. Lipatov. “Deep inelastic e p scattering in perturbation theory”. In: *Sov. J. Nucl. Phys.* 15 (1972). [Yad. Fiz.15,781(1972)], pp. 438–450 (cit. on p. 11).

[40] George F. Sterman and Steven Weinberg. “Jets from Quantum Chromodynamics”. In: *Phys. Rev. Lett.* 39 (1977), p. 1436. DOI: 10.1103/PhysRevLett.39.1436 (cit. on p. 13).

[41] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. “The Anti-k(t) jet clustering algorithm”. In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph] (cit. on pp. 13, 76).

[42] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. “PYTHIA 6.4 Physics and Manual”. In: *JHEP* 05 (2006), p. 026. DOI: 10.1088/1126-6708/2006/05/026. arXiv: hep-ph/0603175 [hep-ph] (cit. on p. 13).

[43] Stefano Frixione and Bryan R. Webber. “Matching NLO QCD computations and parton shower simulations”. In: *JHEP* 06 (2002), p. 029. DOI: 10.1088/1126-6708/2002/06/029. arXiv: hep-ph/0204244 [hep-ph] (cit. on p. 14).

[44] Bo Andersson et al. “Parton Fragmentation and String Dynamics”. In: *Phys. Rept.* 97 (1983), pp. 31–145. DOI: 10.1016/0370-1573(83)90080-7 (cit. on p. 14).

[45] Peter Zeiler Skands. “Tuning Monte Carlo Generators: The Perugia Tunes”. In: *Phys. Rev.* D82 (2010), p. 074018. DOI: 10.1103/PhysRevD.82.074018. arXiv: 1005.3457 [hep-ph] (cit. on p. 14).

[46] The CMS Collaboration. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. In: *Phys. Lett.* B716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex] (cit. on pp. 14, 24, 93).

[47] The ATLAS Collaboration. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Phys. Lett.* B716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex] (cit. on pp. 14, 24).

[48] The ATLAS and CMS Collaborations. “Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s}$ = 7 and 8 TeV”. In: *JHEP* 08 (2016), p. 045. DOI: 10.1007/JHEP08(2016)045. arXiv: 1606.02266 [hep-ex] (cit. on pp. 14, 17).

[49] Charalampos Anastasiou et al. “High precision determination of the gluon fusion Higgs boson cross-section at the LHC”. In: *JHEP* 05 (2016), p. 058. DOI: 10.1007/JHEP05(2016)058. arXiv: 1602.00695 [hep-ph] (cit. on p. 14).

[50] The ATLAS Collaboration. “Evidence for the $H \to b\bar{b}$ decay with the ATLAS detector”. In: *JHEP* 12 (2017), p. 024. DOI: 10.1007/JHEP12(2017)024. arXiv: 1708.03299 [hep-ex] (cit. on p. 14).

[51] The ATLAS and CMS Collaborations. “Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments”. In: *Phys. Rev. Lett.* 114 (2015), p. 191803. DOI: 10.1103/PhysRevLett.114.191803. arXiv: 1503.07589 [hep-ex] (cit. on p. 16).

[52] “Measurements of properties of the Higgs boson decaying into the four-lepton final state in pp collisions at $\sqrt{s} = 13$ TeV”. In: *JHEP* 11 (2017), p. 047. DOI: 10.1007/JHEP11(2017)047. arXiv: 1706.09936 [hep-ex] (cit. on p. 16).

[53] The CMS Collaboration. “Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV”. In: *Phys. Rev.* D92.1 (2015), p. 012004. DOI: 10.1103/PhysRevD.92.012004. arXiv: 1411.3441 [hep-ex] (cit. on p. 16).

[54] The ATLAS Collaboration. “Evidence for the spin-0 nature of the Higgs boson using ATLAS data”. In: *Phys. Lett.* B726 (2013), pp. 120–144. DOI: 10.1016/j.physletb.2013.08.026. arXiv: 1307.1432 [hep-ex] (cit. on p. 16).

[55] The CMS Collaboration. “Constraints on the Higgs boson width from off-shell production and decay to Z-boson pairs”. In: *Phys. Lett.* B736 (2014), pp. 64–85. DOI: 10.1016/j.physletb.2014.06.077. arXiv: 1405.3455 [hep-ex] (cit. on p. 16).

[56] S Heinemeyer et al. *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group.* Tech. rep. CERN-2013-004. CERN-2013-004. Comments: 404 pages, 139 figures, to be submitted to CERN Report. Working Group web page: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CrossSections. Geneva, 2013. URL: https://cds.cern.ch/record/1559921 (cit. on p. 16).

[57] Martin Gonzalez-Alonso et al. “Pseudo-observables in Higgs decays”. In: *Eur. Phys. J.* C75 (2015), p. 128. DOI: 10.1140/epjc/s10052-015-3345-5. arXiv: 1412.6038 [hep-ph] (cit. on p. 17).

[58] W. Buchmuller and D. Wyler. “Effective Lagrangian Analysis of New Interactions and Flavor Conservation”. In: *Nucl. Phys.* B268 (1986), pp. 621–653. DOI: 10.1016/0550-3213(86)90262-2 (cit. on p. 17).

[59] B. Grzadkowski et al. “Dimension-Six Terms in the Standard Model Lagrangian”. In: *JHEP* 10 (2010), p. 085. DOI: 10.1007/JHEP10(2010)085. arXiv: 1008.4884 [hep-ph] (cit. on p. 17).

[60] Giuseppe Degrassi et al. “Higgs mass and vacuum stability in the Standard Model at NNLO”. In: *JHEP* 08 (2012), p. 098. DOI: 10.1007/JHEP08(2012)098. arXiv: 1205.6497 [hep-ph] (cit. on p. 17).

[61] Fedor Bezrukov and Mikhail Shaposhnikov. “Why should we care about the top quark Yukawa coupling?” In: *J. Exp. Theor. Phys.* 120 (2015). [Zh. Eksp. Teor. Fiz.147,389(2015)], pp. 335–343. DOI: 10.1134/S1063776115030152. arXiv: 1411.1923 [hep-ph] (cit. on p. 18).

[62] Archil Kobakhidze et al. “Implications of CP-violating Top-Higgs Couplings at LHC and Higgs Factories”. In: *Phys. Rev.* D95.1 (2017), p. 015016. DOI: 10.1103/PhysRevD.95.015016. arXiv: 1610.06676 [hep-ph] (cit. on p. 18).

[63] T. D. Lee. “A Theory of Spontaneous T Violation”. In: *Phys. Rev.* D8 (1973). [,516(1973)], pp. 1226–1239. DOI: 10.1103/PhysRevD.8.1226 (cit. on p. 18).

[64] G. C. Branco et al. “Theory and phenomenology of two-Higgs-doublet models”. In: *Phys. Rept.* 516 (2012), pp. 1–102. DOI: 10.1016/j.physrep.2012.02.002. arXiv: 1106.0034 [hep-ph] (cit. on p. 18).

[65] Howard E. Haber and Gordon L. Kane. “The Search for Supersymmetry: Probing Physics Beyond the Standard Model”. In: *Phys. Rept.* 117 (1985), pp. 75–263. DOI: 10.1016/0370-1573(85)90051-1 (cit. on p. 18).

[66] Jihn E. Kim. “Light Pseudoscalars, Particle Physics and Cosmology”. In: *Phys. Rept.* 150 (1987), pp. 1–177. DOI: 10.1016/0370-1573(87)90017-2 (cit. on p. 18).

[67] Da Liu, Ian Low, and Carlos E. M. Wagner. “Modification of Higgs Couplings in Minimal Composite Models”. In: *Phys. Rev.* D96.3 (2017), p. 035013. DOI: 10.1103/PhysRevD.96.035013. arXiv: 1703.07791 [hep-ph] (cit. on p. 18).

[68] *Search for* $\mathrm{t\bar{t}H}$ *production in the* $\mathrm{H} \to \mathrm{b\bar{b}}$ *decay channel with 2016 pp collision data at* $\sqrt{s} = 13$ TeV. Tech. rep. CMS-PAS-HIG-16-038. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2231510 (cit. on pp. 18, 71, 72).

[69] *Search for the Standard Model Higgs boson produced in association with top quarks and decaying into a* $\boldsymbol{b\bar{b}}$ *pair in* $\boldsymbol{pp}$ *collisions at* $\sqrt{\boldsymbol{s}}$ *=13 TeV with the ATLAS detector.* Tech. rep. ATLAS-CONF-2017-076. Geneva: CERN, Nov. 2017. URL: https://cds.cern.ch/record/2291393 (cit. on pp. 18, 109, 110).

[70] *Search for Higgs boson production in association with top quarks in multilepton final states at* $\sqrt{s} = 13$ TeV. Tech. rep. CMS-PAS-HIG-17-004. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2256103 (cit. on p. 18).

[71] *Evidence for the associated production of the Higgs boson and a top quark pair with the ATLAS detector.* Tech. rep. ATLAS-CONF-2017-077. Geneva: CERN, Nov. 2017. URL: https://cds.cern.ch/record/2291405 (cit. on p. 18).

[72] *Search for the associated production of a Higgs boson with a top quark pair in final states with a τ lepton at $\sqrt{s} = 13$ TeV*. Tech. rep. CMS-PAS-HIG-17-003. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2257067 (cit. on p. 18).

[73] *Measurements of properties of the Higgs boson in the diphoton decay channel with the full 2016 data set*. Tech. rep. CMS-PAS-HIG-16-040. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2264515 (cit. on p. 18).

[74] *Measurements of Higgs boson properties in the diphoton decay channel with 36.1 fb^{-1} pp collision data at the center-of-mass energy of 13 TeV with the ATLAS detector*. Tech. rep. ATLAS-CONF-2017-045. Geneva: CERN, July 2017. URL: https://cds.cern.ch/record/2273852 (cit. on p. 18).

[75] R. Garoby. "Plans for upgrading the CERN proton accelerator complex". In: *J. Phys. Conf. Ser.* 110 (2008), p. 112003. DOI: 10.1088/1742-6596/110/11/112003 (cit. on p. 22).

[76] Corinne Pralavorio. *Record luminosity: well done LHC*. Nov. 2017. URL: http://cds.cern.ch/record/2295027 (cit. on p. 21).

[77] Lyndon Evans and Philip Bryant. "LHC Machine". In: *JINST* 3 (2008), S08001. DOI: 10.1088/1748-0221/3/08/S08001 (cit. on pp. 22, 23).

[78] H. Van Haevermaet. "Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV". In: *PoS* DIS2016 (2016), p. 198. arXiv: 1607.02033 [hep-ex] (cit. on p. 23).

[79] The CMS Collaboration. "The CMS Experiment at the CERN LHC". In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004 (cit. on pp. 23, 24, 26–30).

[80] The ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *JINST* 3 (2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003 (cit. on p. 23).

[81] A. Augusto Alves Jr. et al. "The LHCb Detector at the LHC". In: *JINST* 3 (2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005 (cit. on p. 23).

[82] K. Aamodt et al. "The ALICE experiment at the CERN LHC". In: *JINST* 3 (2008), S08002. DOI: 10.1088/1748-0221/3/08/S08002 (cit. on p. 23).

[83] Vittorio R. Tavolaro. "The Phase1 CMS Pixel detector upgrade". In: *JINST* 11.12 (2016), p. C12010. DOI: 10.1088/1748-0221/11/12/C12010 (cit. on p. 25).

[84] *Muon Identification and Isolation efficiency on full 2016 dataset*. Tech. rep. Mar. 2017. URL: https://cds.cern.ch/record/2257968 (cit. on p. 26).

[85] Luca Brianza. "Precision crystal calorimetry in LHC Run II with the CMS ECAL". In: *JINST* 12.01 (2017), p. C01069. DOI: 10.1088/1748-0221/12/01/C01069 (cit. on p. 27).

[86] P. Adzic et al. "Energy resolution of the barrel of the CMS electromagnetic calorimeter". In: *JINST* 2 (2007), P04004. DOI: 10.1088/1748-0221/2/04/P04004 (cit. on p. 27).

[87] The CMS Collaboration. "Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in *pp* Collisions at $\sqrt{s} = 7$ TeV". In: *JINST* 8 (2013). [JINST8,9009(2013)], P09009. DOI: 10.1088/1748-0221/8/09/P09009. arXiv: 1306.2016 [hep-ex] (cit. on p. 27).

[88] The CMS Collaboration. "Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV". In: *JINST* 10.06 (2015), P06005. DOI: 10.1088/1748-0221/10/06/P06005. arXiv: 1502.02701 [physics.ins-det]. URL: https://arxiv.org/abs/1502.02701 (cit. on pp. 27, 75).

[89] Gerard Jungman, Marc Kamionkowski, and Kim Griest. "Supersymmetric dark matter". In: *Phys. Rept.* 267 (1996), pp. 195–373. DOI: 10.1016/0370-1573(95)00058-5. arXiv: hep-ph/9506380 [hep-ph] (cit. on p. 27).

[90] N. Akchurin and R. Wigmans. "Quartz fibers as active elements in detectors for particle physics". In: *Rev. Sci. Instrum.* 74 (2003), pp. 2955–2972. DOI: 10.1063/1.1570945 (cit. on p. 29).

[91] Victor Daniel Elvira. *Measurement of the Pion Energy Response and Resolution in the CMS HCAL Test Beam 2002 Experiment*. Tech. rep. CMS-NOTE-2004-020. Geneva: CERN, Sept. 2004. URL: https://cds.cern.ch/record/800406 (cit. on p. 29).

[92] The CMS Collaboration. "Particle-flow reconstruction and global event description with the CMS detector". In: (2017). arXiv: 1706.04965 [physics.ins-det] (cit. on pp. 29, 31, 49).

[93] The CMS Collaboration. "Measurements of Inclusive W and Z Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV". In: *JHEP* 01 (2011), p. 080. DOI: 10.1007/JHEP01(2011)080. arXiv: 1012.2466 [hep-ex] (cit. on p. 29).

[94] The CMS Collaboration. "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV". In: *JINST* 7 (2012), P10002. DOI: 10.1088/1748-0221/7/10/P10002. arXiv: 1206.4071 [physics.ins-det]. URL: https://arxiv.org/abs/1206.4071 (cit. on pp. 29, 75).

[95] The CMS Collaboration. "Performance of the CMS Drift Tube Chambers with Cosmic Rays". In: *JINST* 5 (2010), T03015. DOI: 10.1088/1748-0221/5/03/T03015. arXiv: 0911.4855 [physics.ins-det] (cit. on p. 29).

[96] The CMS Collaboration. "Alignment of the CMS Silicon Tracker during Commissioning with Cosmic Rays". In: *JINST* 5 (2010), T03009. DOI: 10.1088/1748-0221/5/03/T03009. arXiv: 0910.2505 [physics.ins-det] (cit. on p. 29).

[97] The CMS Collaboration. "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV". In: *JINST* 12.02 (2017), P02014. DOI: 10.1088/1748-0221/12/02/P02014. arXiv: 1607.03663 [hep-ex] (cit. on pp. 31, 76, 83).

[98] The CDF Collaboration. "Evidence for top quark production in $\bar{p}p$ collisions at $\sqrt{s} = 1.8$ TeV". In: *Phys. Rev.* D50 (1994), pp. 2966–3026. DOI: 10.1103/PhysRevD.50.2966 (cit. on p. 33).

[99] *Identification of b quark jets at the CMS Experiment in the LHC Run 2*. Tech. rep. CMS-PAS-BTV-15-001. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2138504 (cit. on pp. 34–37, 39, 79, 84).

[100] Wolfgang Waltenberger. *Adaptive Vertex Reconstruction*. Tech. rep. CMS-NOTE-2008-033. Geneva: CERN, July 2008. URL: https://cds.cern.ch/record/1166320 (cit. on p. 34).

[101] The CMS Collaboration. "Measurement of $B\bar{B}$ Angular Correlations based on Secondary Vertex Reconstruction at $\sqrt{s} = 7$ TeV". In: *JHEP* 03 (2011), p. 136. DOI: 10.1007/JHEP03(2011)136. arXiv: 1102.3194 [hep-ex] (cit. on p. 34).

[102] The CMS Collaboration. "Identification of b-quark jets with the CMS experiment". In: *JINST* 8 (2013), P04013. DOI: 10.1088/1748-0221/8/04/P04013. arXiv: 1211.4462 [hep-ex] (cit. on pp. 35, 36, 77).

[103] M. Acciarri et al. "Measurement of the branching ratios b –> e neutrino X, mu neutrino X, tau-neutrino X and neutrino X". In: *Z. Phys.* C71 (1996), pp. 379–390. DOI: 10.1007/s002880050184 (cit. on p. 35).

[104] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830 (cit. on pp. 36, 39).

[105] Matteo Cacciari and Gavin P. Salam. "Pileup subtraction using jet areas". In: *Phys. Lett.* B659 (2008), pp. 119–126. DOI: 10.1016/j.physletb.2007.09.077. arXiv: 0707.1378 [hep-ph] (cit. on pp. 37, 75).

[106] Jerome H Friedman. "Greedy function approximation: a gradient boosting machine". In: *Annals of statistics* (2001), pp. 1189–1232 (cit. on p. 38).

[107] J. Pata. *The mlglue package*. https://github.com/jpata/mlglue. [Online; accessed 19-Sep-2017]. 2017 (cit. on p. 39).

[108] T. He T. Chen M. Benesty et al. *The xgboost package*. https://github.com/dmlc/xgboost/. [Online; accessed 22-Nov-2017]. 2017 (cit. on p. 39).

[109] Andreas Hocker et al. "TMVA - Toolkit for Multivariate Data Analysis". In: *PoS* ACAT (2007), p. 040. arXiv: physics/0703039 [PHYSICS] (cit. on p. 39).

[110] The CMS Collaboration. "Evidence for the Higgs boson decay to a bottom quark-antiquark pair". In: (2017). arXiv: 1709.07497 [hep-ex] (cit. on p. 43).

[111] The CMS Collaboration. “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”. In: (2017). arXiv: 1712.07158 [physics.ins-det] (cit. on p. 43).

[112] Jerzy Neyman and Egon S Pearson. “On the problem of the most efficient tests of statistical hypotheses”. In: *Breakthroughs in statistics*. Springer, 1992, pp. 73–108 (cit. on p. 45).

[113] W. Van Doninck. “Application of a Multivariate Discriminant Analysis to High-Energy Physics in Bubble Chambers”. In: *Beyond the Standard Model. Proceedings, 18th Rencontres de Moriond, La Plagne, France, March 13-19, 1983. vol. 2.* 1983, pp. 265–278. URL: http://inspirehep.net/record/209453/files/Pages_from_C83-03-13--1_265-278.pdf (cit. on p. 46).

[114] Bruce H. Denby. “Neural Networks and Cellular Automata in Experimental High-energy Physics”. In: *Comput. Phys. Commun.* 49 (1988), pp. 429–448. DOI: 10.1016/0010-4655(88)90004-5 (cit. on p. 46).

[115] K. Kondo. “Dynamical Likelihood Method for Reconstruction of Events With Missing Momentum. 1: Method and Toy Models”. In: *J. Phys. Soc. Jap.* 57 (1988), pp. 4126–4140. DOI: 10.1143/JPSJ.57.4126 (cit. on p. 47).

[116] The CDF Collaboration. “Search for a Higgs Boson in $WH \to \ell\nu b\bar{b}$ in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV”. In: *Phys. Rev. Lett.* 103 (2009), p. 101802. DOI: 10.1103/PhysRevLett.103.101802. arXiv: 0906.5613 [hep-ex] (cit. on p. 47).

[117] The 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”. In: *Phys. Rev.* D85 (2012), p. 072001. DOI: 10.1103/PhysRevD.85.072001. arXiv: 1112.4358 [hep-ex] (cit. on p. 47).

[118] D0 Collaboration. “A precision measurement of the mass of the top quark”. In: *Nature* 429.June (June 2004), pp. 10–14. DOI: 10.1038/nature02614.1.. URL: http://dx.doi.org/10.1038/nature02589 (cit. on p. 47).

[119] The D0 Collaboration. “Evidence for production of single top quarks”. In: *Phys. Rev.* D78 (2008), p. 012005. DOI: 10.1103/PhysRevD.78.012005. arXiv: 0803.0739 [hep-ex] (cit. on p. 47).

[120] Pierre Artoisenet et al. “Unravelling $t\bar{t}h$ via the Matrix Element Method”. In: *Phys. Rev. Lett.* 111.9 (2013), p. 091802. DOI: 10.1103/PhysRevLett.111.091802. arXiv: 1304.6414 [hep-ph] (cit. on p. 47).

[121] The ATLAS Collaboration. “Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector”. In: *Eur. Phys. J.* C75.7 (2015), p. 349. DOI: 10.1140/epjc/s10052-015-3543-1. arXiv: 1503.05066 [hep-ex] (cit. on p. 47).

[122] The 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”. In: *Eur. Phys. J.* C75.6 (2015), p. 251. DOI: 10.1140/epjc/s10052-015-3454-1. arXiv: 1502.02485 [hep-ex] (cit. on pp. 47, 81, 109, 110).

[123] Yanyan Gao et al. “Spin determination of single-produced resonances at hadron colliders”. In: *Phys. Rev.* D81 (2010), p. 075022. DOI: 10.1103/PhysRevD.81.075022. arXiv: 1001.3396 [hep-ph] (cit. on p. 47).

[124] *A new tagger for the charge identification of b-jets*. Tech. rep. ATL-PHYS-PUB-2015-040. Geneva: CERN, Sept. 2015. URL: https://cds.cern.ch/record/2048132 (cit. on p. 49).

[125] CMS Collaboration. “Missing transverse energy performance of the CMS detector”. In: *Journal of Instrumentation* 6 (Sept. 2011), p. 9001. DOI: 10.1088/1748-0221/6/09/P09001. arXiv: 1106.5048 [physics.ins-det] (cit. on p. 50).

[126] J. Alwall, A. Freitas, and O. Mattelaer. “The Matrix Element Method and QCD Radiation”. In: *Phys. Rev.* D83 (2011), p. 074010. DOI: 10.1103/PhysRevD.83.074010. arXiv: 1010.2263 [hep-ph] (cit. on p. 52).

[127] Fabio Cascioli, Philipp Maierhofer, and Stefano Pozzorini. “Scattering Amplitudes with Open Loops”. In: *Phys. Rev. Lett.* 108 (2012), p. 111601. DOI: 10.1103/PhysRevLett.108.111601. arXiv: 1111.5206 [hep-ph] (cit. on pp. 52, 54).

[128] R. Brun and F. Rademakers. “ROOT: An object oriented data analysis framework”. In: *Nucl. Instrum. Meth.* A389 (1997), pp. 81–86. DOI: 10.1016/S0168-9002(97)00048-X (cit. on p. 54).

[129] V. Vasilev et al. “Cling: The new interactive interpreter for ROOT 6”. In: *J. Phys. Conf. Ser.* 396 (2012), p. 052071. DOI: 10.1088/1742-6596/396/5/052071 (cit. on p. 54).

[130] Pavel M. Nadolsky et al. “Implications of CTEQ global analysis for collider observables”. In: *Phys. Rev.* D78 (2008), p. 013004. DOI: 10.1103/PhysRevD.78.013004. arXiv: 0802.0007 [hep-ph] (cit. on p. 54).

[131] Andy Buckley et al. “LHAPDF6: parton density access in the LHC precision era”. In: *Eur. Phys. J.* C75 (2015), p. 132. DOI: 10.1140/epjc/s10052-015-3318-8. arXiv: 1412.7420 [hep-ph] (cit. on p. 54).

[132] G. Peter Lepage. “A New Algorithm for Adaptive Multidimensional Integration”. In: *J. Comput. Phys.* 27 (1978), p. 192. DOI: 10.1016/0021-9991(78)90004-9 (cit. on pp. 54, 55).

[133] T. Hahn. “CUBA: A Library for multidimensional numerical integration”. In: *Comput. Phys. Commun.* 168 (2005), pp. 78–95. DOI: 10.1016/j.cpc.2005.01.010. arXiv: hep-ph/0404043 [hep-ph] (cit. on p. 54).

[134] G. Eulisse and Lassi A. Tuura. “IgProf profiling tool”. In: *Computing in high energy physics and nuclear physics. Proceedings, Conference, CHEP’04, Interlaken, Switzerland, September 27-October 1, 2004.* 2005, pp. 655–658. URL: http://doc.cern.ch/yellowrep/2005/2005-002/p655.pdf (cit. on p. 56).

[135] Lassi A. Tuura, V. Innocente, and G. Eulisse. “Analysing CMS software performance using IgProf, OProfile and callgrind”. In: *J. Phys. Conf. Ser.* 119 (2008), p. 042030. DOI: 10.1088/1742-6596/119/4/042030 (cit. on p. 56).

[136] Doug Schouten, Adam DeAbreu, and Bernd Stelzer. “Accelerated Matrix Element Method with Parallel Computing”. In: *Comput. Phys. Commun.* 192 (2015), pp. 54–59. DOI: 10.1016/j.cpc.2015.02.020. arXiv: 1407.7595 [physics.comp-ph] (cit. on p. 57).

[137] *Search for* $\mathrm{t\bar{t}H}$ *production in the* $\mathrm{H} \to \mathrm{b\bar{b}}$ *decay channel with* $\sqrt{s} = 13$ TeV *pp collisions at the CMS experiment.* Tech. rep. CMS-PAS-HIG-16-004. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2139578 (cit. on p. 71).

[138] S. Agostinelli et al. “GEANT4: A Simulation toolkit”. In: *Nucl. Instrum. Meth.* A506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8 (cit. on p. 72).

[139] Stefano Frixione, Paolo Nason, and Carlo Oleari. “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method”. In: *JHEP* 11 (2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph] (cit. on p. 72).

[140] Emanuele Re. “Single-top Wt-channel production matched with parton showers using the POWHEG method”. In: *Eur. Phys. J.* C71 (2011), p. 1547. DOI: 10.1140/epjc/s10052-011-1547-z. arXiv: 1009.2450 [hep-ph] (cit. on p. 72).

[141] Valentin Hirschi et al. “Automation of one-loop QCD corrections”. In: *JHEP* 05 (2011), p. 044. DOI: 10.1007/JHEP05(2011)044. arXiv: 1103.0621 [hep-ph] (cit. on p. 72).

[142] Torbjorn Sjostrand, Stephen Mrenna, and Peter Skands. “A Brief Introduction to PYTHIA 8.1”. In: *Comput. Phys. Commun.* 178 (2008), pp. 852–867. DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820 [hep-ph] (cit. on p. 72).

[143] *Underlying Event Tunes and Double Parton Scattering.* Tech. rep. CMS-PAS-GEN-14-001. Geneva: CERN, 2014. URL: https://cds.cern.ch/record/1697700 (cit. on p. 72).

[144] Peter Skands, Stefano Carrazza, and Juan Rojo. “Tuning PYTHIA 8.1: the Monash 2013 Tune”. In: *Eur. Phys. J.* C74.8 (2014), p. 3024. DOI: 10.1140/epjc/s10052-014-3024-y. arXiv: 1404.5630 [hep-ph] (cit. on p. 72).

[145] *Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of* $t\bar{t}$ *at* $\sqrt{s} = 8$ *and 13 TeV*. Tech. rep. CMS-PAS-TOP-16-021. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2235192 (cit. on p. 72).

[146] D. de Florian et al. “Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector”. In: (2016). DOI: 10.23731/CYRM-2017-002. arXiv: 1610.07922 [hep-ph] (cit. on pp. 72, 73).

[147] Michal Czakon and Alexander Mitov. “Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders”. In: *Comput. Phys. Commun.* 185 (2014), p. 2930. DOI: 10.1016/j.cpc.2014.06.021. arXiv: 1112.5675 [hep-ph] (cit. on pp. 72, 73).

[148] 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”. In: *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph] (cit. on p. 72).

[149] T. Gehrmann et al. “W^+W^- Production at Hadron Colliders in Next to Next to Leading Order QCD”. In: *Phys. Rev. Lett.* 113.21 (2014), p. 212001. DOI: 10.1103/PhysRevLett.113.212001. arXiv: 1408.5243 [hep-ph] (cit. on p. 72).

[150] John M. Campbell, R. Keith Ellis, and Ciaran Williams. “Vector boson pair production at the LHC”. In: *JHEP* 07 (2011), p. 018. DOI: 10.1007/JHEP07(2011)018. arXiv: 1105.0020 [hep-ph] (cit. on p. 72).

[151] Nikolaos Kidonakis. “Two-loop soft anomalous dimensions for single top quark associated production with a W- or H-”. In: *Phys. Rev.* D82 (2010), p. 054018. DOI: 10.1103/PhysRevD.82.054018. arXiv: 1005.4451 [hep-ph] (cit. on p. 72).

[152] P. Kant et al. “HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions”. In: *Comput. Phys. Commun.* 191 (2015), pp. 74–89. DOI: 10.1016/j.cpc.2015.02.001. arXiv: 1406.4403 [hep-ph] (cit. on p. 72).

[153] Ryan Gavin et al. “W Physics at the LHC with FEWZ 2.1”. In: *Comput. Phys. Commun.* 184 (2013), pp. 208–214. DOI: 10.1016/j.cpc.2012.09.005. arXiv: 1201.5896 [hep-ph] (cit. on p. 72).

[154] Kirill Melnikov and Frank Petriello. “Electroweak gauge boson production at hadron colliders through O(alpha(s)**2)”. In: *Phys. Rev.* D74 (2006), p. 114017. DOI: 10.1103/PhysRevD.74.114017. arXiv: hep-ph/0609070 [hep-ph] (cit. on p. 72).

[155] S Dittmaier et al. *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*. CERN Yellow Reports: Monographs. Geneva: CERN, 2011. URL: https://cds.cern.ch/record/1318996 (cit. on p. 72).

[156] W. Beenakker et al. “Higgs radiation off top quarks at the Tevatron and the LHC”. In: *Phys. Rev. Lett.* 87 (2001), p. 201805. DOI: 10.1103/PhysRevLett.87.201805. arXiv: hep-ph/0107081 [hep-ph] (cit. on p. 72).

[157] W. Beenakker et al. “NLO QCD corrections to t anti-t H production in hadron collisions”. In: *Nucl. Phys.* B653 (2003), pp. 151–203. DOI: 10.1016/S0550-3213(03)00044-0. arXiv: hep-ph/0211352 [hep-ph] (cit. on p. 72).

[158] S. Dawson et al. “Associated top quark Higgs boson production at the LHC”. In: *Phys. Rev.* D67 (2003), p. 071503. DOI: 10.1103/PhysRevD.67.071503. arXiv: hep-ph/0211438 [hep-ph] (cit. on p. 72).

[159] S. Dawson et al. “Associated Higgs production with top quarks at the large hadron collider: NLO QCD corrections”. In: *Phys. Rev.* D68 (2003), p. 034022. DOI: 10.1103/PhysRevD.68.034022. arXiv: hep-ph/0305087 [hep-ph] (cit. on p. 72).

[160] A. Djouadi, J. Kalinowski, and M. Spira. “HDECAY: A Program for Higgs boson decays in the standard model and its supersymmetric extension”. In: *Comput. Phys. Commun.* 108 (1998), pp. 56–74. DOI: 10.1016/S0010-4655(97)00123-9. arXiv: hep-ph/9704448 [hep-ph] (cit. on p. 72).

[161] J. M. Butterworth et al. "THE TOOLS AND MONTE CARLO WORKING GROUP Summary Report from the Les Houches 2009 Workshop on TeV Colliders". In: *Physics at TeV colliders. Proceedings, 6th Workshop, dedicated to Thomas Binoth, Les Houches, France, June 8-26, 2009*. 2010. arXiv: 1003.1643 [hep-ph]. URL: https://inspirehep.net/record/848006/files/arXiv:1003.1643.pdf (cit. on p. 72).

[162] A. Bredenstein et al. "NLO QCD corrections to pp —> t anti-t b anti-b + X at the LHC". In: *Phys. Rev. Lett.* 103 (2009), p. 012002. DOI: 10.1103/PhysRevLett.103.012002. arXiv: 0905.0110 [hep-ph] (cit. on p. 73).

[163] Ansgar Denner, Robert Feger, and Andreas Scharf. "Irreducible background and interference effects for Higgs-boson production in association with a top-quark pair". In: *JHEP* 04 (2015), p. 008. DOI: 10.1007/JHEP04(2015)008. arXiv: 1412.5290 [hep-ph] (cit. on p. 73).

[164] Fabio Cascioli et al. "NLO matching for $t\bar{t}b\bar{b}$ production with massive b-quarks". In: *Phys. Lett.* B734 (2014), pp. 210–214. DOI: 10.1016/j.physletb.2014.05.040. arXiv: 1309.5912 [hep-ph] (cit. on p. 73).

[165] Nazar Bartosik and Achim Geiser. "Associated Top-Quark-Pair and b-Jet Production in the Dilepton Channel at $\sqrt{s} = 8$ TeV as Test of QCD and Background to tt+Higgs Production". Presented 03 Jul 2015. Aug. 2015. URL: https://cds.cern.ch/record/2047049 (cit. on p. 74).

[166] CMS Collaboration. "Search for ttH in all-jet final states in pp collisions at sqrt(s) = 13 TeV". In: *JHEP* (2018). Forthcoming (cit. on p. 74).

[167] The CMS Collaboration. "Search for neutral Higgs bosons decaying to tau pairs in pp collisions at $\sqrt{s} = 7$ TeV". In: *Physics Letters B* 713.2 (2012), pp. 68–90. DOI: https://doi.org/10.1016/j.physletb.2012.05.028. URL: http://www.sciencedirect.com/science/article/pii/S0370269312005564 (cit. on p. 75).

[168] The CMS Muon POG. *Baseline muon selections for Run-II*. https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonIdRun2?rev=28#Tight_Muon. [Online; accessed 19-Sep-2017]. 2017 (cit. on p. 75).

[169] The CMS EGamma POG. *Multivariate Electron Identification for Run2*. https://twiki.cern.ch/twiki/bin/view/CMS/MultivariateElectronIdentificationRun2?rev=31. [Online; accessed 19-Sep-2017]. 2017 (cit. on p. 75).

[170] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "FastJet User Manual". In: *Eur. Phys. J.* C72 (2012), p. 1896. DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph] (cit. on p. 76).

[171] *Jet Performance in pp Collisions at 7 TeV*. Tech. rep. CMS-PAS-JME-10-003. Geneva: CERN, 2010. URL: https://cds.cern.ch/record/1279362 (cit. on p. 76).

[172] *Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET*. Tech. rep. CMS-PAS-PFT-09-001. Geneva: CERN, Apr. 2009. URL: https://cds.cern.ch/record/1194487 (cit. on p. 76).

[173] *Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector*. Tech. rep. CMS-PAS-PFT-10-001. 2010. URL: https://cds.cern.ch/record/1247373 (cit. on p. 76).

[174] *Jet algorithms performance in 13 TeV data*. Tech. rep. CMS-PAS-JME-16-003. Geneva: CERN, 2017. URL: http://cds.cern.ch/record/2256875 (cit. on p. 76).

[175] *Pileup Removal Algorithms*. Tech. rep. CMS-PAS-JME-14-001. Geneva: CERN, 2014. URL: https://cds.cern.ch/record/1751454 (cit. on p. 76).

[176] The CMS Collaboration. "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS". In: *JINST* 6 (2011), P11002. DOI: 10.1088/1748-0221/6/11/P11002. arXiv: 1107.4277 [physics.ins-det] (cit. on p. 76).

[177] *Search for Higgs Boson Production in Association with a Top-Quark Pair and Decaying to Bottom Quarks or Tau Leptons*. Tech. rep. CMS-PAS-HIG-13-019. Geneva: CERN, 2013. URL: https://cds.cern.ch/record/1564682 (cit. on p. 79).

[178] The CMS Muon POG. *Baseline muon selections for Run-II*. https://twiki.cern.ch/twiki/bin/viewauth/CMS/MuonWorkInProgressAndPagResults?rev=29#Results_on_the_full_2016_data. [Online; accessed 19-Sep-2017]. 2017 (cit. on p. 86).

[179] The CMS EGamma Physics Object Group. *EGamma ID recipes for Run-II*. https://twiki.cern.ch/twiki/bin/view/CMS/EgammaIDRecipesRun2?rev=43#Electron_efficiencies_and_scale. [Online; accessed 25-Sep-2017]. 2017 (cit. on p. 86).

[180] Michael Hildreth. *Estimating Systematic Errors Due to Pileup Modeling*. https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupSystematicErrors. [Online; accessed 25-Sep-2017]. 2017 (cit. on p. 86).

[181] *CMS Luminosity Measurements for the 2016 Data Taking Period*. Tech. rep. CMS-PAS-LUM-17-001. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2257069 (cit. on p. 86).

[182] The CMS Luminosity Physics Object Group. *Current recommendations for luminosity estimations in Run-II*. https://twiki.cern.ch/twiki/bin/view/CMS/TWikiLUM?rev=79#CurRec. [Online; accessed 25-Sep-2017]. 2017 (cit. on p. 86).

[183] The CMS Collaboration. "Measurements of $t\bar{t}$ cross sections in association with b jets and inclusive jets and their ratio using dilepton final states in pp collisions at $\sqrt{s} = 13$ TeV". In: *Phys. Lett.* B776 (2018), pp. 355–378. DOI: 10.1016/j.physletb.2017.11.043. arXiv: 1705.10141 [hep-ex] (cit. on p. 86).

[184] Axel Bredenstein et al. "NLO QCD Corrections to Top Anti-Top Bottom Anti-Bottom Production at the LHC: 2. full hadronic results". In: *JHEP* 03 (2010), p. 021. DOI: 10.1007/JHEP03(2010)021. arXiv: 1001.4006 [hep-ph] (cit. on p. 86).

[185] The CMS Collaboration. "Combined results of searches for the standard model Higgs boson in pp collisions at $\sqrt{s} = 7$ TeV". In: *Phys. Lett.* B710 (2012), pp. 26–48. DOI: 10.1016/j.physletb.2012.02.064. arXiv: 1202.1488 [hep-ex] (cit. on p. 93).

[186] *Procedure for the LHC Higgs boson search combination in Summer 2011*. Tech. rep. CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11. Geneva: CERN, Aug. 2011. URL: https://cds.cern.ch/record/1379837 (cit. on p. 93).

[187] Thomas Junk. "Confidence level computation for combining searches with small statistics". In: *Nucl. Instrum. Meth.* A434 (1999), pp. 435–443. DOI: 10.1016/S0168-9002(99)00498-2. arXiv: hep-ex/9902006 [hep-ex] (cit. on p. 93).

[188] A L Read. "Presentation of search results: the CL s technique". In: *Journal of Physics G: Nuclear and Particle Physics* 28.10 (2002), p. 2693. URL: http://stacks.iop.org/0954-3899/28/i=10/a=313 (cit. on p. 93).

[189] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *Eur. Phys. J.* C71 (2011). [Erratum: Eur. Phys. J.C73,2501(2013)], p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0, 10.1140/epjc/s10052-013-2501-z. arXiv: 1007.1727 [physics.data-an] (cit. on pp. 93, 94).

[190] Abraham Wald. "Tests of statistical hypotheses concerning several parameters when the number of observations is large". In: *Transactions of the American Mathematical society* 54.3 (1943), pp. 426–482 (cit. on p. 94).

[191] M Capeans et al. *ATLAS Insertable B-Layer Technical Design Report*. Tech. rep. CERN-LHCC-2010-013. ATLAS-TDR-19. Sept. 2010. URL: https://cds.cern.ch/record/1291633 (cit. on p. 109).

[192] The CMS Collaboration. *CMS Technical Design Report for the Pixel Detector Upgrade*. Tech. rep. CERN-LHCC-2012-016. CMS-TDR-11. Additional contacts: Jeffrey Spalding, Fermilab, Jeffrey.Spalding@cern.ch Didier Contardo, Universite Claude Bernard-Lyon I, didier.claude.contardo@cern.ch. Sept. 2012. URL: https://cds.cern.ch/record/1481838 (cit. on p. 109).

[193] Lingvist Technologies. *The Lingvist learning environment*. https://learn.lingvist.com/. [Online; accessed 18-Oct-2017]. 2017 (cit. on p. 113).

[194] H. Ebbinghaus. *Memory: A Contribution to Experimental Psychology*. https://web.archive.org/web/20050504104838/http://psy.ed.asu.edu/~classics/Ebbinghaus/index.htm. [Online; accessed 23-Sep-2017]. 1885 (cit. on p. 113).

[195] Albert Corbett. "Cognitive computer tutors: Solving the two-sigma problem". In: *User Modeling 2001* (2001), pp. 137–147 (cit. on p. 113).

[196] Chris Piech et al. "Deep Knowledge Tracing". In: *CoRR* abs/1506.05908 (2015). URL: http://arxiv.org/abs/1506.05908 (cit. on pp. 114, 117).

[197] Emmanuel J Candès and Benjamin Recht. "Exact matrix completion via convex optimization". In: *Foundations of Computational mathematics* 9.6 (2009), p. 717 (cit. on p. 116).

[198] A. T. Corbett and J. R. Anderson. "Knowledge tracing: Modeling the acquisition of procedural knowledge". In: *User modelling and user-adapted interaction* 4.4 (1994), pp. 253–278 (cit. on p. 117).

[199] Mohammad Khajah, Robert V Lindsey, and Michael C Mozer. "How deep is knowledge tracing?" In: *arXiv preprint arXiv:1604.02416* (2016) (cit. on p. 117).

[200] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. "Learning Representations by Back Propagating Errors". In: 323 (Oct. 1986), pp. 533–536 (cit. on p. 117).

[201] Yoshua Bengio, Patrice Simard, and Paolo Frasconi. *Learning Long-Term Dependencies with Gradient Descent is Difficult*. 1994 (cit. on p. 117).

[202] Felix A. Gers, Jürgen A. Schmidhuber, and Fred A. Cummins. "Learning to Forget: Continual Prediction with LSTM". In: *Neural Comput.* 12.10 (Oct. 2000), pp. 2451–2471. DOI: 10.1162/089976600300015015. URL: http://dx.doi.org/10.1162/089976600300015015 (cit. on p. 117).

[203] Nitish Srivastava et al. "Dropout: a simple way to prevent neural networks from overfitting." In: *Journal of machine learning research* 15.1 (2014), pp. 1929–1958 (cit. on p. 117).

[204] Tomas Mikolov et al. "Efficient estimation of word representations in vector space". In: *arXiv preprint arXiv:1301.3781* (2013) (cit. on p. 119).

[205] Jeffrey Pennington, Richard Socher, and Christopher D Manning. "Glove: Global vectors for word representation." In: *EMNLP*. Vol. 14. 2014, pp. 1532–1543 (cit. on p. 119).

[206] Richard G Baraniuk. "Compressive sensing". In: *IEEE signal processing magazine* 24.4 (2007), pp. 118–121 (cit. on p. 119).

[207] Diederik P. Kingma and Jimmy Ba. "Adam: A Method for Stochastic Optimization". In: *CoRR* abs/1412.6980 (2014). URL: http://arxiv.org/abs/1412.6980 (cit. on p. 121).

[208] M. Abadi et al. "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems". In: *ArXiv e-prints* (Mar. 2016). arXiv: 1603.04467 [cs.DC] (cit. on p. 121).

[209] Susan E Embretson and Steven P Reise. *Item response theory*. Psychology Press, 2013 (cit. on p. 123).

[210] Yarin Gal and Zoubin Ghahramani. "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning". In: *international conference on machine learning*. 2016, pp. 1050–1059 (cit. on p. 124).

[211] Irina Elgort. "Deliberate learning and vocabulary acquisition in a second language". In: *Language Learning* 61.2 (2011), pp. 367–413 (cit. on p. 124).

[212] W. Samek, T. Wiegand, and K.-R. Müller. "Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models". In: *ArXiv e-prints* (Aug. 2017). arXiv: 1708.08296 [cs.AI] (cit. on p. 125).

[213] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "Why should I trust you?: Explaining the predictions of any classifier". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM. 2016, pp. 1135–1144 (cit. on p. 125).