DEPARTMENT OF PHYSICS

SEARCH FOR CHARGED HIGGS BOSONS DECAYING INTO A TOP AND A BOTTOM QUARK WITH THE CMS DETECTOR IN pp COLLISIONS AT $\sqrt{s} = 13$ TeV WITH THE 2016 AND 2017 LHC DATA

DOCTOR OF PHILOSOPHY DISSERTATION

MARINA KOLOSOVA

2022

DEPARTMENT OF PHYSICS

SEARCH FOR CHARGED HIGGS BOSONS DECAYING INTO A TOP AND A BOTTOM QUARK WITH THE CMS DETECTOR IN pp COLLISIONS AT $\sqrt{s} = 13$ TeV WITH THE 2016 AND 2017 LHC DATA

MARINA KOLOSOVA

A Dissertation Submitted to the University of Cyprus in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

March 2022

VALIDATION PAGE

Doctoral Candidate:
Marina Kolosova

Doctoral Thesis Title:
Search for charged Higgs bosons decaying into a top and a bottom quark with the CMS detector in pp collisions at $\sqrt{s} = 13$ TeV with the 2016 and 2017 LHC data

The present Doctoral Dissertation was submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy at the Department of Physics at the University of Cyprus. It was approved on March 23, 2022 by the members of the Examination Committee.

Examination Committee:

Konstantinos Kordas
Comittee member
Associate Professor, Aristotle University of Thessaloniki

Ekaterini Tzamariudaki
Comittee member
Director of Research, National Centre for Scientific Research *Demokritos*

Nicolaos Toumbas
Committee head
Associate Professor, University of Cyprus

Halil Saka
Committee member
Lecturer, University of Cyprus

Fotios Ptochos
Research Advisor
Associate Professor, University of Cyprus

DECLARATION OF DOCTORAL CANDIDATE

The present doctoral dissertation was submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy of the University of Cyprus. It is a product of original work of my own, unless otherwise mentioned through references, notes, or any other statements.

Marina Kolosova

ΠΕΡΙΛΗΨΗ

Η παρούσα διατριβή έχει ως αντικείμενο την έρευνα για την ύπαρξη ηλεκτρικά φορτισμένων μποζονίων Higgs (H$^\pm$) μέσω της διεργασίας pp $\rightarrow$ t(b)H$^\pm$ και της ακόλουθης διάσπασής τους H$^\pm$ $\rightarrow$ tb, στην πλήρως αδρονική τελική κατάσταση. Η έρευνα βασίζεται σε γεγονότα που παρήχθησαν κατά τη διάρκεια του 2016 και 2017 στον Μεγάλο Αδρονικό Επιταχυντή (LHC) σε συγκρούσεις pp σε ενέργεια κέντρου μάζας $\sqrt{s} = 13$ TeV και συλλέχθηκαν από τον ανιχνευτή CMS. Τα γεγονότα που μελετήθηκαν αντιστοιχούν σε ολοκληρωμένη φωτεινότητα ίση με 35.9 fb^{-1} για το 2016 και 40.5 fb^{-1} για το 2017. Η έρευνα καλύπτει τις υποθέσεις σήματος του φορτισμένου μποζονίου Higgs με μάζα μεταξύ 200 GeV και 3 TeV. Ανάλογα με τη μάζα του φορτισμένου μποζονίου Higgs, υπάρχουν δύο διαφορετικές τοπολογίες συμβάντων, η προωθημένη και η διακρίσιμη. Τα φορτισμένα μποζόνια Higgs με μεγάλη μάζα της τάξης $\mathcal{O}$(TeV) δημιουργούν προϊόντα διάσπασης με μέση εγκάρσια ορμή αρκετών εκατοντάδων GeV, με αποτέλεσμα να δημιουργούνται ευθυγραμμισμένοι πίδακες σωματιδίων (jets) προερχόμενοι από τη διάσπαση του top κουάρκ. Αυτά τα ευθυγραμμισμένα προϊόντα δε μπορούν να διακριθούν με τις παραδοσιακές τεχνικές ομαδοποίησης, αλλά μπορούν να ανακατασκευαστούν ως ενιαίος πίδακας μεγάλου κώνου. Τα φορτισμένα μποζόνια Higgs με χαμηλότερες μάζες, και επομένως με λιγότερο προωθημένες τελικές καταστάσεις, μπορούν να ανακατασκευαστούν από πλήρως διακριτούς πίδακες μικρού κώνου. Η παρούσα ανάλυση στοχεύει στην διακρίσιμη τοπολογία.

Η ανάλυση με τα δεδομένα του 2016 αποτελεί την πρώτη χρονικά μελέτη σε αυτό το κανάλι και τη συγκεκριμένη τελική κατάσταση. Η ανάλυση με τα δεδομένα του 2017 αποτελεί συνέχεια της έρευνας και περιλαμβάνει πολλαπλές βελτιώσεις σε σχέση με την προηγούμενη ανάλυση, όπως τεχνικές μηχανικής μάθησης για την αναγνώριση του top κουάρκ και εξαγωγή σήματος, κατηγοριοποίηση των γεγονότων, και επικύρωση της μεθόδου εκτίμησης των διεργασιών υποβάθρου. Για την επέκταση της ευαισθησίας αναζήτησης φορτισμένων μποζονίων Higgs τα αποτελέσματα της διακρίσιμης ανάλυσης με τα δεδομένα του 2016 συνδυάστηκαν με την ανάλυση που στοχεύει την προωθημένη τοπολογία. Δεν παρατηρήθηκε πλεόνασμα γεγονότων πέρα από το αναμενόμενο υπόβαθρο διεργασιών του Καθιερωμένου Προτύπου και υπολογίστηκε, ανεξαρτήτως θεωρητικού μοντέλου, το άνω όριο της ενεργού διατομής παραγωγής ενός φορτισμένου μποζόνιου Higgs επί τον λόγο διακλάδωσης σε ζεύγος t και b κουάρκ-αντικουάρκ, $\sigma_{\text{pp}\rightarrow\text{t(b)H}^\pm} \times \mathcal{B}(\text{H}^\pm \rightarrow \text{tb})$, με 95% επίπεδο εμπιστοσύνης.

ABSTRACT

A search for charged Higgs bosons is presented in the $\mathrm{pp} \to \mathrm{t(b)H^\pm}$ process with the subsequent decay $\mathrm{H^\pm \to tb}$, in the all-jet final state. The search is based on 2016 and 2017 LHC data recorded by the CMS detector at a center-of-mass energy of $\sqrt{s} = 13$ TeV, corresponding to integrated luminosity of 35.9 and 40.5 fb^{-1}, respectively. The search covers the signal hypotheses of a charged Higgs boson with mass in the range of 200 GeV to 3 TeV. Depending on the mass of the charged Higgs boson, there exist two distinct event topologies, the boosted and the resolved. Charged Higgs bosons with large masses of $\mathcal{O}(\mathrm{TeV})$ give rise to decay products with average transverse momenta of several hundred GeV, leading to highly collimated jets originating from the decay of the top quark. These boosted products cannot be resolved with the standard clustering techniques, instead, they are reconstructed as a single large-cone jet. Charged Higgs bosons with lower masses, and thus less boosted final states, can be reconstructed from fully resolved small-cone jets. This thesis targets the resolved event topology.

The analysis performed with the 2016 LHC data constitutes the first report on this channel and final state. A continuation of the search is performed with the 2017 data and contains multiple improvements with respect to the 2016 analysis, such as machine learning techniques for top quark identification and signal extraction, event categorization, and validation of the background estimation methods. To extend the search sensitivity, the 2016 results are combined with the analysis targeting the boosted event topology. No significant excess is observed above the expected background and model-independent upper limits at 95% confidence level are set on the product of the charged Higgs boson production cross-section and branching fraction to the top and bottom quark-antiquark pair decay, $\sigma_{\mathrm{pp \to t(b)H^\pm}} \times \mathcal{B}(\mathrm{H^\pm \to tb})$. The results are interpreted using different minimal supersymmetric extensions of the Standard Model.

Hoping for peace.

ACKNOWLEDGEMENTS

for inspiring me with her determination and positive attitude. To Mikela, for always being next to me.

I would also like to thank my closest friends Michaella, Dimitra, Tatiana, Anastasia and Andriani for their love and understanding. Special thanks to my parents Yuri and Chariklia, and my sister Alexia. I could not have completed this dissertation without their love, support and encouragement. Last but certainly not least, I would like to thank Sergio for bringing happiness to my life.

Author's contribution

Publications

- CMS Collaboration, "Search for charged Higgs bosons decaying into a top and a bottom quark in the all-jet final state of pp collisions at $\sqrt{s} = 13$ TeV", *JHEP*, vol. 07, p. 126, 2020. DOI: 10.1007/JHEP07(2020) 126. arXiv: 2001.07763 [hep-ex].

Detector Perfomance Notes

- CMS Collaboration. Performance of top triggers in Run 2. DP-2019/026
- CMS Collaboration. Level-1 E/Gamma trigger performance in Run II. DP-2019-020

CMS Internal

- CMS Note. "Search for charged Higgs bosons with the $H^{\pm} \rightarrow tb$ decay in the fully hadronic final state at $\sqrt{s} = 13$ TeV with full Run 2 data" (CMS AN-2021/060)

- CMS Note. "Resolved top quark identification using deep neural network" (CMS AN-2021/019)

- CMS Note. "Search for charged Higgs bosons in the neutral Higgs boson and W boson decay channel at $\sqrt{s} = 13$ TeV with full Run 2 data" (CMS AN-2020/107)

- Approval talk. "Search for charged Higgs bosons with the $H^{\pm} \rightarrow tb$ decay in the fully hadronic final states at 13 TeV" at "Approvals for LHCP 2019, 10 May 2019"

- Talk. "$H^{\pm} \rightarrow tb$ fully hadronic (resolved and boosted analysis)" at "CMS Exotic Higgs Workshop, 29-30 Nov. 2018"

- Talk. "First look at 2017 data" at "CMS Exotic Higgs Workshop, 19-20 Oct. 2017"

Conference-related activities

- Plenary talk. "CMS Overview on Searches for charged Higgs" at "Charged H: Charged Higgs at HL-LHC, Charged Higgs Online workshop, 30-31 Aug 2021, Mohammed V University in Rabat, Morocco and BNL, online (Morocco)"

- Plenary talk. "Searches for exotic visible Higgs boson decays at CMS and ATLAS" at "HC2019: Higgs Couplings 2019, 30 Sep-5 Oct 2019, Oxford (United Kingdom)" (plenary talk)

- Poster. "The CMS electron and photon trigger for the LHC Run 2 and plans for Run 3" at "EPS-HEP2019: European Physical Society Conference on High Energy Physics, 10-17 Jul 2019, Ghent (Belgium)"

- Poster. "Search for charged Higgs bosons decaying into a top and a bottom quark in the fully hadronic final states at 13 TeV" at "CMSWeek2018: CMS Week in Budapest, 1-5 Oct 2018, Budapest (Hungary)"

Activities within the CMS Collaboration

The author contributed to the efforts of online data-taking during the Run 2 of the LHC with the development and monitoring of new Level-1 seeds and High Level Triggers for the Top Physics Analysis Group and B-tagging Physics Object Group. During the 2018 data-taking period, the author was a contact of the online Level-1 e/γ trigger algorithm. Since 2020, the author is responsible for the maintenance of the offline Pileup Jet identification (PUID) algorithm.

Contents

List of Tables

List of Figures

1 Introduction

High Energy Physics is the field of physics aiming to understand the constituents of the universe as well as the fundamental forces governing the interactions between them. To do so, physicists use accelerators to boost particles at speeds close to that of light and bring them to collisions. These collisions might produce massive particles that are typically unstable and rapidly transform into a cascade of lighter and more stable particles. By studying the remnants of these collisions, physicists can infer the presence of a new particle. Since 1930's, the field enjoyed many new discoveries that led to theoretical developments. These findings were encapsulated into a well-established mathematical framework that is currently known as the best theory describing nature, the Standard Model (SM) of particle physics [1, 2, 3, 4, 5, 6, 7]. To date, all predictions of the SM have been verified by experimental data with remarkable precision [8].

Despite being an outstandingly successful theory, the SM fails to answer some major questions. For instance, it cannot explain why everything around us, from the smallest form of life to the biggest astronomical objects, consists almost entirely of matter rather than antimatter [9]. During the Big Bang, equal amounts of matter and antimatter were created. An unknown mechanism could have caused the extinction of antimatter. Another unexplained phenomenon is that the galaxies in our universe are rotating faster than they should, such that the gravitational pull of their visible matter cannot hold them together. Physicists gave a possible explanation for this with the introduction of dark matter which could give additional mass to these galaxies. Dark matter, an entity that cannot absorb, reflect or emit light and therefore cannot be directly observed, is not described by the SM but makes up about 27% of our universe [9]. These, together with other unresolved questions indicate that the SM is yet incomplete and not appropriate for scales beyond the electroweak (EW) scale. It is rather viewed as an effective low-energy approximation of a more fundamental theory that is valid up to a scale smaller than the Planck scale ($\mathrm{M_P} \sim 10^{19}$ GeV), where gravitational effects become relevant. Nevertheless, there exist several extensions of the SM proposing solutions to its shortcomings. In some of them, the existence of additional Higgs bosons [10], either electrically neutral, charged, or both, is favored.

The discovery of the Higgs boson in 2012 by the ATLAS [11] and CMS [12] collaborations at the Large Hadron Collider (LHC) was one of the most important scientific breakthroughs. It marked a leap towards the understanding of the electroweak symmetry breaking mechanism, based on which any elementary particle interacting with the Higgs field can ac-

quire mass. During its second run in 2015-2018, the LHC operated with record-breaking energy of 13 TeV and produced large statistics datasets allowing to perform precise measurements of SM processes, including measurements on the properties of the Higgs boson [13, 14]. In addition, the LHC's physics program involves a rich variety of searches beyond the SM, such as Supersymmetry and exotic physics, as well as flavor-physics. The work presented here focuses on the search for charged Higgs bosons ($H^{\pm}$) decaying into a top and bottom quark-antiquark pair in the all-jet final state. The search is based on LHC proton-proton collision data recorded with the CMS detector in 2016 [15] and 2017 at a center-of-mass energy of 13 TeV, corresponding to 35.9 and 40.5 fb^{-1}, respectively. The search targets heavy charged Higgs bosons and exploits machine-learning and data-driven techniques for signal-to-background discrimination and background estimation. No excess above the expected SM background is observed and upper limits at 95% confidence level (C.L.) are set on the charged Higgs boson production cross section and branching fraction of its subsequent decay into a top and a bottom quark-antiquark pair, $\sigma_{pp \to t(b)H^{\pm}} \times \mathcal{B}(H^{\pm} \to tb)$.

The thesis is structured as follows. The theoretical background that motivated the search for charged Higgs bosons is reviewed in chapter 2. This includes an introduction to the SM of particle physics and its shortcomings, to theories beyond the SM that predict the presence of charged Higgs bosons, as well as the phenomenology of charged Higgs bosons at the LHC. Chapter 3 describes the LHC accelerator complex and the Compact Muon Solenoid (CMS) detector used to deliver and collect the data. The reconstruction of the collision events and the physics objects used to conduct this search is described in chapter 4. Chapter 5 gives an overview of the common analysis strategy and statistical methods used to analyze the 2016 and 2017 LHC data. Chapter 6 and chapter 7 describe separately the experimental methods and developments used in each year to extract the signal, together with the corresponding results. A summary of the searches and prospects are presented in chapter 8.

2 Theoretical Background

Particle physicists aim to provide answers to fundamental questions, such as the origin of the building blocks of nature and their interactions. The ultimate goal is to reduce all natural phenomena at all energy scales into a set of basic laws, which can quantitatively reproduce experimental observations and infer the evolution of our universe. Section 2.1 gives an overview of these efforts through the Standard Model (SM) of the strong [1, 2, 3, 4] and electroweak interactions [5, 6, 7] of elementary particles. Following, section 2.2 briefly describes some of the yet-unresolved problems of the SM, and section 2.3 gives a brief overview on two minimal extensions of the SM. Finally, section 2.4 introduces the reader to the search for charged Higgs bosons at the LHC from a phenomenological point of view.

2.1 The Standard Model of Particle Physics

Our current picture of the universe is embodied in the Standard Model of particle physics; a renormalizable quantum field theory based on gauge symmetries. The SM describes the basic building blocks of matter (*fermions*) and the interaction particles mediating the fundamental forces (*bosons*). Its underlying local gauge symmetry is the $SU(3)_C \times SU(2)_L \times U(1)_Y$. Each factor in the SM gauge symmetry group corresponds to a fundamental force that is represented by one or more dynamical fields. The excitations of these fields correspond to the gauge bosons that act as force carriers. The color gauge group $SU(3)_C$ describes Quantum Chromodynamics (QCD) and is associated with eight in total fields G_μ, corresponding to *gluons*, the mediators of the strong force. The $SU(2)_L$ describes the weak isospin interactions experienced only by left-handed fermions and is associated with three gauge bosons W_μ. The $U(1)_Y$ corresponds to the weak hypercharge interactions and is associated with the gauge field B_μ. The combined symmetry $SU(2)_L \times U(1)_Y$ describes the *electroweak* interactions and the gauge fields W_μ and B_μ mix to form the physical $W^\pm$, Z^0, and γ bosons.

2.1.1 Fundamental particles and interactions

The fundamental particles of the SM are listed in Table 2.1. Fermions carry half-integer spin and follow the Fermi-Dirac statistics [16]. They comprise of *quarks* and *leptons* and come in threefold family replication. Besides their masses, the particles of each family share identical quantum numbers. For every fermion, there exists a corresponding antiparticle of the same

Table 2.1: The fundamental particles of the SM. Their masses and quantum numbers are listed.

		Particles	Mass	T_{L3}	$U(1)_Y$	Q
Fermions	Leptons	$\begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix}, \begin{pmatrix} \nu_{\mu L} \\ \mu_L \end{pmatrix}, \begin{pmatrix} \nu_{\tau L} \\ \tau_L \end{pmatrix}$	$m_{\nu_e} < 1.1\,\text{eV}, m_{\nu_\mu} < 190\,\text{keV}, m_{\nu_\tau} < 18.2\,\text{MeV}$	$1/2$	$-1/2$	0
			$m_e = 0.51\,\text{MeV}, m_\mu = 0.11\,\text{GeV}, m_\tau = 1.78\,\text{GeV}$	$-1/2$	$-1/2$	-1
		e_R, μ_R, τ_R		0	-1	-1
	Quarks	$\begin{pmatrix} u_L \\ d_L \end{pmatrix}, \begin{pmatrix} c_L \\ s_L \end{pmatrix}, \begin{pmatrix} t_L \\ b_L \end{pmatrix}$	$m_u = 2.16\,\text{MeV}, m_c = 1.27\,\text{GeV}, m_t = 172.76\,\text{GeV}$	$1/2$	$1/6$	$2/3$
			$m_d = 4.67\,\text{MeV}, m_s = 93\,\text{MeV}, m_b = 4.18\,\text{GeV}$	$-1/2$	$1/6$	$-1/3$
		u_R, c_R, t_R		0	$2/3$	$2/3$
		d_R, s_R, b_R		0	$-1/3$	$-1/3$
Bosons	Vectors	$W^\pm$	$m_{W^\pm} = 80.4\,\text{GeV}$	± 1	0	± 1
		Z^0	$m_{Z^0} = 91.2\,\text{GeV}$	0	0	0
		γ	$m_\gamma = 0$	0	0	0
		g	$m_g = 0$	0	0	0
	Scalars	H^0	$m_{H^0} = 125.1\,\text{GeV}$	$-1/2$	$1/2$	0

mass and opposite internal quantum numbers. Quarks come in six distinct *flavors*, three up-type (up, charm, top) and three down-type (down, strange, bottom) quarks and carry a color charge and fractional electric charge. They are subject to all interactions described by the SM and are never observed as free states but only confined in colorless bound states, known as *hadrons*. Hadrons can be further classified into *mesons* and *baryons*, with the former being composed of a quark-antiquark pair, and the latter composed of three quarks or three antiquarks. Protons and neutrons are two examples of baryons, consisting of uud and udd states, respectively. The idea of hadrons with more than the minimal quark content of mesons or baryons was first proposed by Gell-Mann in 1964 [1] and was experimentally observed in the form of tetraquarks in $\bar{B}^0 \to \psi' K^- \pi^+$ decays [17, 18, 19] and more recently in the form of pentaquarks in $\Lambda_b^0 \to J/\psi K^- p$ decays [20, 21]. Leptons are colorless states and comprise of the electrically charged electrons (e), muons (μ), and taus (τ), and the electrically neutral neutrinos ν_e, ν_μ, and ν_τ. Electrons, muons, and taus are subject to electroweak interactions, while neutrinos interact only via the weak force.

The strong and electroweak interactions arise from the exchange of fermions and bosons. Bosons have integer spin and follow the Bose-Einstein statistics. Eight massless and electrically neutral vector gauge bosons, called *gluons* (g) mediate the strong force. Gluons themselves carry color charge and are thus, self-coupled. The photon (γ), is one of the four gauge bosons of the electroweak interactions. It is massless and interacts with all electrically charged particles, but is itself neutral. The remaining three force-carriers of the electroweak interactions are the massive gauge bosons W^-, W^+, and Z^0. The last elementary particle and the only spin-0 scalar boson of the SM is the *Higgs* boson (H^0), discovered in 2012 by the ATLAS [11] and CMS [12] experiments.

2.1.2 Strong interactions

Quantum Chromodynamics (QCD) is the non-abelian gauge field theory that describes the strong interactions of colored quarks and gluons. It was proposed in 1974, by Politzer [4], Wilczek [2], and Gross [3], and is based on the gauge group $SU(3)_C$, where the subscript C stands for color. Quarks come in three colors (*red*, *green* and *blue*) and are described by field spinors in the fundamental represention of $SU(3)_C$:

$$\Psi = \begin{pmatrix} \psi_{q\ red} \\ \psi_{q\ green} \\ \psi_{q\ blue} \end{pmatrix}, \quad q = u, d, c, s, t, b. \tag{2.1}$$

Gluons fields, eight in total, transform under the adjoint representation of the $SU(3)_C$ color group. The free Lagrangian of a single quark field with mass m_q is given by

$$\mathcal{L}_{q\ free} = \bar{\Psi}_q(i\gamma^\mu \partial_\mu - m_q)\Psi_q, \tag{2.2}$$

where γ^μ are the Dirac matrices and the symbol ∂_μ stands for the differentiation operator with respect to the space-time coordinates, $\partial_\mu \equiv \frac{\partial}{\partial x^\mu}$. The interactions between quarks and gluons are introduced through the covariant derivative $D_\mu = \partial_\mu + ig_s\frac{\lambda_a}{2}G_\mu^a$, where g_s is the coupling constant for strong interactions (usually redefined as $\alpha_s = g_s^2/4\pi$) and G_μ^a corresponds to the eight independent gluon fields where each of them is associated with a generator of $SU(3)_C$. The generators of $SU(3)_C$ are the 3×3, hermitian, and traceless Gell-Mann matrices:

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

$$\lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix} \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \lambda_8 = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}, \tag{2.3}$$

and satisfy the commutation relation:

$$\left[\frac{\lambda_a}{2}, \frac{\lambda_b}{2}\right] = if_{abc}\frac{\lambda_c}{2}, \qquad a, b, c = 1, ..., 8, \tag{2.4}$$

where there is an implicit sum over c. The f_{abc} are the structure constants of the $SU(3)_C$ and are antisymmetric under the interchange of any pair of color indices. The gauge field kinetic energy is expressed in terms of the strength tensor $G_{\mu\nu}^a$ as $-\frac{1}{4}G^{a\mu\nu}G_{\mu\nu}^a$, where $G_{\mu\nu}^a$ is defined as

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a - g_s f_{abc} G_\mu^b G_\nu^c. \tag{2.5}$$
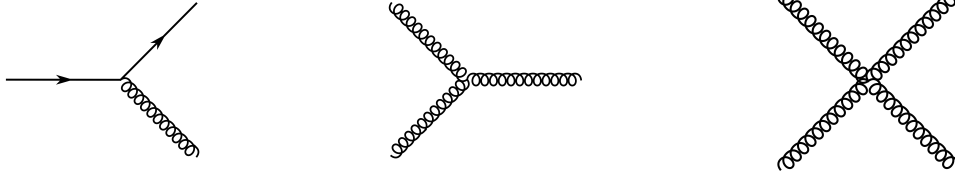
Figure 2.1: The quark-gluon (*left*), 3-gluon (*middle*), and 4-gluon (*right*) QCD interaction vertices.

The overall QCD Lagrangian is given by

$$\mathcal{L}_{\text{QCD}} = -\frac{1}{4}G^{a\mu\nu}G^a_{\mu\nu} + \sum_{q=1\ldots6} \bar{\Psi}_q(i\gamma^\mu D_\mu - m_q)\Psi_q, \tag{2.6}$$

where all terms are invariant under the local $\text{SU}(3)_{\text{C}}$ transformations:

$$\Psi_q \to \Psi'_q = e^{-ig_s\frac{\lambda}{2}\theta_a(x)}\Psi_q$$
$$G^a_\mu \to G'^a_\mu = G^a_\mu + \partial_\mu\theta^a + g_s f_{abc}\theta^b G^c_\mu, \tag{2.7}$$

where $\theta_a(x)$ are arbitrary parameters. Owing to the non-abelian character of the color group, the non-linear term in $G^a_{\mu\nu}$ in equation 2.6 generates cubic and quartic gluon self-interactions. The strength of these interactions, shown in Fig. 2.1, is given by the same coupling constant and is proportional to the g_s and g_s^2, respectively.

The dependence of the strong coupling constant $\alpha_s(Q^2)$ on the momentum transfer Q between quarks and gluons encodes the two prominent features of QCD, *color confinement* and *asymptotic freedom*. In the lowest order of perturbation theory, the strong coupling behaves as follows:

$$\alpha_s(Q^2) \propto \frac{1}{\ln\left(\frac{Q^2}{\Lambda_{\text{QCD}}}\right)}, \tag{2.8}$$

where Λ_{QCD} is an arbitrary scale at which non-perturbative effects take over. The α_s decreases as a function of Q^2, meaning that for very high values of Q^2, the coupling tends towards zero and quarks behave as free particles (asymptotic freedom). On the other hand, for very low values of Q^2, below Λ_{QCD}, the coupling constant is strong enough to confine the quarks into composite and colorless states (color confinement). The two aforementioned phenomena provide a rich, yet-complex phenomenology in hadron colliders. Quarks cannot be observed as free states. Instead, they become part of hadrons following a process known as *hadronization*. The top quark is an exception of this statement as it decays before it has time to hadronize due to its small lifetime (5×10^{-25} s).

2.1.3 Electroweak interactions

The SM of the electroweak interactions was introduced by Glashow [5], Weinberg [6], and Salam [7]. It describes the electromagnetic [22, 23, 24, 25] and weak [26] interactions between quarks and leptons and is based on the gauge symmetry group $SU(2)_L \times U(1)_Y$ of weak left-handed isospin and weak hypercharge.

Electromagnetic interactions

Quantum Electrodynamics (QED) is an abelian gauge theory, which describes the interactions of electrically charged particles with photons. It is based on the symmetry group $U(1)_{QED}$. A free fermion field ψ (four-component spinor) with mass m_f has a free Lagrangian:

$$\mathcal{L}_{f\ free} = \bar{\psi}(x)(i\gamma^\mu \partial_\mu - m_f)\psi(x), \tag{2.9}$$

where the adjoint spinor $\bar{\psi}$ is defined as $\bar{\psi} = \psi^\dagger \gamma^0$. When considering the interaction of a fermion of electric charge q with a photon, the electromagnetic field A_μ is introduced and the Lagrangian is modified as follows:

$$\mathcal{L}_{f\ QED} = \bar{\psi}(x)(i\gamma^\mu D_\mu - m_f)\psi(x), \tag{2.10}$$

where D_μ is the covariant derivative defined as $D_\mu = \partial_\mu + iqA_\mu$. The replacement of the differentiation operator ∂_μ with the covariant derivative ensures the gauge invariance of the Lagrangian under local $U(1)_{QED}$ transformations:

$$\psi(x) \to e^{-iq\chi(x)}\psi(x), \quad \bar{\psi}(x) \to \bar{\psi}e^{iq\chi(x)}, \quad A^\mu \to A^\mu + \partial^\mu \chi. \tag{2.11}$$

The kinetic Lagrangian of the photon field is also gauge-invariant and in the absence of fermions is given by

$$\mathcal{L}_{gauge\ kin} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{2.12}$$

where $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ is the electromagnetic or Maxwell strength tensor. Combining 2.10 and 2.12 leads to the QED Lagrangian:

$$\mathcal{L}_{QED} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \bar{\psi}(x)(i\gamma^\mu D_\mu - m)\psi(x) \tag{2.13}$$

A mass term for the photon of the form $\frac{1}{2}m_\gamma^2 A_\mu A^\mu$ is forbidden as it spoils the gauge invariance of the Lagrangian. Hence, the mass of the gauge field A_μ is set to zero. Experimentally, we know that $m_\gamma < 6 \times 10^{-17}$ eV [8]. A basic QED vertex is shown in Fig. 2.2, where the strength of the interaction is quantified by the coupling constant g, which is proportional to the fermion electric charge. The coupling strength is usually expressed in terms of the dimensionless QED *fine-structure constant* α, which, close to the Thomson limit, has a value of
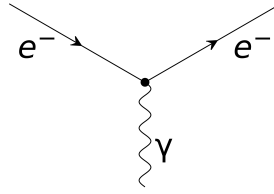
Figure 2.2: A basic QED vertex

$\alpha^{-1} \sim 137.036$ [8]. At higher energy scales above a few hundred MeV, low energy hadronic effects introduce a theoretical uncertainty on α. The QED fine-structure constant is then predicted and observed to increase with the momentum transfer due to the effect of *screening* of the bare electric charge caused by the polarized cloud of virtual particles.

Weak interactions

Weak interactions [7] are described by a non-abelian theory based on the $\mathrm{SU}(2)_{\mathrm{L}}$ group. Compared to QED and QCD, the weak interaction has several unique features. It is the only interaction that allows for flavor change. For example, in beta decays a proton decays to a neutron, a positron and an electron neutrino: $p \to n + e^+ + \nu_e$, which at the quark level is $u \to d + e^+ + \nu_e$, as shown in Fig. 2.3. Another important feature is that weak interaction
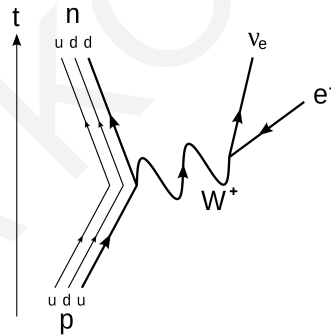
Figure 2.3: Leading-order Feynman diagram for β^+ decay of a proton into a neutron, positron and electron neutrino through an intermediate W^+ boson.

violates parity symmetry. Parity ($\hat{P}$) is the operation of spatial inversion; $\hat{P}\psi(\vec{r}) \to \psi(-\vec{r})$, and its violation was experimentally confirmed by C.S. Wu in 1957 [27]. To account for the parity violation, weak interactions are described by a chiral gauge theory, which introduces an *axial vector* coupling in the Lagrangian. Any fermionic field can be expressed in terms of its left and right-handed chiral components via the projection operators P_R and P_L:

$$\psi = \frac{1}{2}(1+\gamma^5)\psi + \frac{1}{2}(1-\gamma^5)\psi = P_R\psi + P_L\psi = \psi_R + \psi_L, \qquad (2.14)$$

where the operator $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ has eigenvalues of ± 1. A field with eigenvalue of $+1$ (-1) is referred to as left-chirality (right-chirality) field. Left-chirality fields are represented as $\mathrm{SU}(2)_{\mathrm{L}}$ doublets (Ψ_L), while right-chirality fields as $\mathrm{SU}(2)_{\mathrm{L}}$ singlets (ψ_R), leading to the

following representations for leptons:

$$\Psi_L = \begin{pmatrix} \nu_{\ell_L} \\ \ell_L \end{pmatrix}, \ \ \ell_R \quad (\ell = e, \mu, \tau) \tag{2.15}$$

and for quarks:

$$\Psi_L = \begin{pmatrix} q_L \\ q'_L \end{pmatrix}, \ \ q_R, \ q'_R \quad (q = u, c, t \text{ and } q' = d, s, b). \tag{2.16}$$

The quarks that are primed are *weak eigenstates* related to *mass eigenstates* by the Cabibbo-Kobayashi-Maskawa (CKM) matrix [28, 29]:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \hat{V}_{\text{CKM}} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{2.17}$$

Given the above representations, the free Lagrangian for a single fermion, omitting the mass term, is written as:

$$\mathcal{L}_{f \ free} = i\bar{\Psi}_L \gamma^\mu \partial_\mu \Psi_L + i\bar{\psi}_R \gamma^\mu \partial_\mu \psi_R + i\bar{\psi}'_R \gamma^\mu \partial_\mu \psi'_R. \tag{2.18}$$

The weak interaction arises from $\text{SU}(2)_L$ local phase transformations acting only on the left-handed components of the fermion fields:

$$\Psi_L \to e^{\frac{i}{2}\vec{\theta}\cdot\vec{T}} \Psi_L, \qquad \psi_R \to \psi_R, \tag{2.19}$$

where $\vec{\theta}(x)$ is a vector with three real parameters and $T^i = \frac{1}{2}\sigma^i$ are the generators of $\text{SU}(2)_L$. The matrices σ^i are the Pauli matrices. The conserved charge of the $\text{SU}(2)_L$ group is called *weak isospin* and has three components. Since the right-chirality fermions are singlets under $\text{SU}(2)_L$, they do not carry weak isospin charge. The left-chirality fermions have isospin equal to $1/2$ and a third component of the weak isospin equal to $T_3 = +1/2$ for neutrinos and up-type quarks, and $T_3 = -1/2$ for charged leptons and down-type quarks. The theory becomes $\text{SU}(2)_L$ gauge-invariant through the introduction of a multiplet of gauge vector fields $W^i_\mu(x)$ in the covariant derivative:

$$D_\mu = \partial_\mu + ig_w \frac{\sigma^i}{2} W^i_\mu, \quad i = 1, 2, 3 \tag{2.20}$$

where g_w is the coupling constant of the weak interactions. The weak strength tensor is defined as:

$$W^i_{\mu\nu} = \partial_\mu W^i_\nu - \partial_\nu W^i_\mu - g_w \varepsilon_{ijk} W^j_\mu W^k_\nu, \tag{2.21}$$

where ε_{ijk} is the Levi-Civita symbol, corresponding to the structure constants of $\mathrm{SU(2)_L}$ and satisfying the commutation relation $[\frac{\sigma_i}{2}, \frac{\sigma_j}{2}] = i\varepsilon_{ijk}\frac{\sigma_k}{2}$. Since the field strengths $W_{\mu\nu}^i$ contain a quadratic term, the Lagrangian of the weak force gives rise to cubic and quartic self-interactions among the gauge fields, the strength of which is given by the same $\mathrm{SU(2)_L}$ coupling constant, g_w. Some of the basic vertices of the weak interactions are shown in Fig. 2.4. The gauge bosons of the weak theory are massless as an explicit mass term would break the $\mathrm{SU(2)_L}$ gauge invariance. The same holds for the fermion masses. Since left- and right-handed fields transform differently under $\mathrm{SU(2)_L}$, a fermion mass term would lead to mixing of the two chirality states: $m\bar{\psi}\psi = m(\bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R)$.

In nature, there exist three physical mediators of the weak interactions: the charged $W^{\pm}$ bosons responsible for the *charged-current* interactions, and the Z^0 boson which mediates the *neutral-current* interactions. The physical $W^{\pm}$ bosons are associated with the fields W_{μ}^1 and W_{μ}^2 as follows:

$$W_{\mu}^{\pm} = \frac{1}{\sqrt{2}}(W_{\mu}^1 \mp iW_{\mu}^2). \tag{2.22}$$

The third weak gauge boson W_{μ}^3 is a neutral massless gauge boson but cannot be identified as the massive Z^0 boson, nor as the photon field A_{μ} as the electromagnetic current does not involve any ν's. Remarkably, the unification of the electromagnetic and weak forces can relate the W_{μ}^3 with the Z_{μ} and A_{μ} boson fields after the spontaneous breaking of $\mathrm{SU(2)_L} \times \mathrm{U(1)_Y}$ to $\mathrm{U(1)_{QED}}$, described in section 2.1.4. The term $\mathrm{U(1)_Y}$ corresponds to an abelian gauge group, where Y stands for its conserved charge, the *weak hypercharge*. The weak hypercharge of a field is related to its electric charge (Q) and the third component of weak isospin (T_3) through the Gell-Mann-Nishijima formula:

$$\frac{Y}{2} = Q - T_3. \tag{2.23}$$

Imposing local $\mathrm{U(1)_Y}$ gauge invariance on both left- and right-handed fermion components introduces the associated gauge field B_{μ} in the covariant derivative $D_{\mu} = \partial_{\mu} + ig_Y\frac{Y}{2}B_{\mu}$, where g_Y is the coupling constant of $\mathrm{U(1)_Y}$. The $\mathrm{U(1)_Y}$ strength tensor $B_{\mu\nu}$ is defined as $\partial_{\mu}B_{\nu} - \partial_{\nu}B_{\mu}$. The overall electroweak Lagrangian before the spontaneous breaking of $\mathrm{SU(2)_L} \times \mathrm{U(1)_Y}$ and for a single family of quarks and leptons, omitting the mass term can be given by:

$$\mathcal{L}_{\mathrm{SU(2)_L \times U(1)_Y}} = -\frac{1}{4}W_{\mu\nu}^i W^{i\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} + i\bar{\Psi}_L\gamma^{\mu}D_{\mu}\Psi_L + i\bar{\psi}_R\gamma^{\mu}D_{\mu}\psi_R + i\bar{\psi}'_R\gamma^{\mu}D_{\mu}\psi'_R \tag{2.24}$$

where the covariant derivatives acting on left- and right-chirality fermions differ as below:

$$D_{\mu}\psi_R = (\partial_{\mu} - ig_Y\frac{Y}{2}B_{\mu})\psi_R \tag{2.25}$$

$$D_{\mu}\Psi_L = (\partial_{\mu} - ig_w\frac{\sigma^i}{2}W_{\mu}^i - ig_Y\frac{Y}{2}B_{\mu})\Psi_L. \tag{2.26}$$
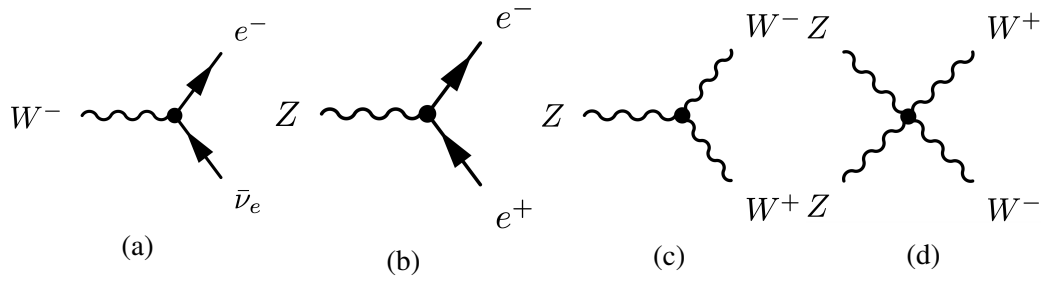
Figure 2.4: Basic vertices in weak interactions (non-exhaustive).

2.1.4 The Brout-Englert-Higgs mechanism

So far, no mass terms for any of the fields were introduced as they would spoil the gauge invariance of the SM Lagrangian. To generate masses for the gauge bosons, the local gauge symmetry must break. The local gauge symmetry of the SM breaks in the vacuum, causing the spontaneous symmetry breaking (SSB) of the electroweak symmetry group into the electromagnetic subgroup:

$$\mathrm{SU(3)_C \times SU(2)_L \times U(1)_Y \to SU(3)_C \times U(1)_{QED}}. \tag{2.27}$$

This mechanism, known as the Brout-Englert-Higgs (BEH) mechanism, was proposed in 1964 by Englert and Brout [30], Higgs [10] and Guralnik, Hagen and Kibble [31], and is realized through the introduction of two complex scalar fields in an $\mathrm{SU(2)_L}$ doublet, ϕ and a weak hypercharge of $1/2$:

$$\phi = \begin{pmatrix} \phi^\dagger \\ \phi^0 \end{pmatrix} \tag{2.28}$$

The Lagrangian corresponding to this new scalar field is:

$$\mathcal{L}_\phi = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi), \tag{2.29}$$

where the covariant derivative is given by equation 2.26 and $V(\phi)$ is the scalar field potential. Renormalizability and $\mathrm{SU(2)_L \times U(1)_Y}$ invariance require that the scalar field potential is of the form:

$$V(\phi) = \mu^2 (\phi^\dagger \phi) + \lambda (\phi^\dagger \phi)^2, \tag{2.30}$$

where μ^2 has dimensions of mass squared and λ is a dimensionless parameter that is required to be positive such that the energy is bounded from below. The minima of the scalar field potential can be found by imposing:

$$0 = \frac{\partial V}{\partial \phi^\dagger} = \mu^2 \phi + 2\lambda (\phi^\dagger \phi)\phi \ \Rightarrow \ \phi = \begin{cases} \phi = 0 \text{ for } \mu^2 > 0 \\ \phi^\dagger \phi = -\frac{\mu^2}{2\lambda} \text{ for } \mu^2 < 0 \end{cases} \tag{2.31}$$

Since the electric charge is a conserved quantity, only the neutral scalar field ϕ^0 can acquire a vacuum expectation value (VEV) for $\mu^2 < 0$. Due to the symmetry of $V(\phi)$ there is an infinite set of degenerate states with minimum energy satisfying $\phi^\dagger \phi = -\frac{\mu^2}{2\lambda} \equiv \frac{v^2}{2}$, as shown in Fig. 2.5. Once a particular ground state is chosen, $\phi = \frac{1}{\sqrt{2}} \left(\begin{smallmatrix} 0 \\ v \end{smallmatrix} \right)$, the $\mathrm{SU(2)_L \times U(1)_Y}$
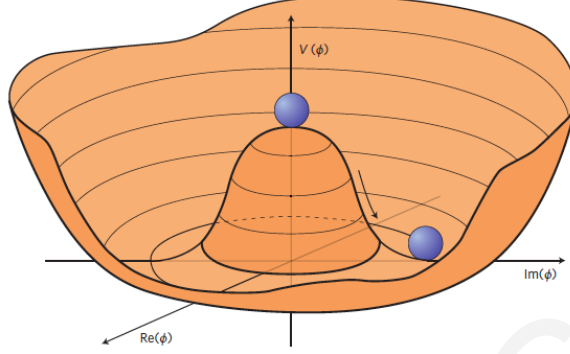
Figure 2.5: The shape of the scalar potential. For $\mu^2 < 0$, there is an infinite number of degenerate vacua, connected through a massless excitation field.

symmetry gets spontaneously broken into $\mathrm{U(1)_{QED}}$, which still remains a true symmetry of the vacuum. Based on the Goldstone theorem [32], there must exist as many massless spin-0 particles (*Goldstone bosons*) as the broken generators, in this case, three massless states should appear. Without loss of generality, the scalar doublet can be written in terms of four real fields $\theta_{1,2,3}(x)$ and $H(x)$:

$$\phi(x) = \begin{pmatrix} \theta_2 + i\theta_1 \\ \frac{1}{\sqrt{2}}(v + H) - i\theta_3 \end{pmatrix} = \frac{1}{\sqrt{2}} e^{i\frac{\sigma_i}{2}\theta^i(x)} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}. \tag{2.32}$$

The $H(x)$ is the massive field and θ^i are the Goldstone bosons associated with the broken symmetry. Under a gauge transformation, one can move to the *unitary gauge*, where these θ^i get absorbed. To generate the masses of the gauge bosons, we need to fully expand the kinetic term with the covariant derivative in the Lagrangian:

$$|D_\mu \phi|^2 \Longrightarrow \left| \left(-i g_w \frac{\sigma^i}{2} W_\mu^i - i \frac{g_Y}{2} B_\mu \right) \phi \right|^2 = \frac{1}{8} \left| \begin{pmatrix} g_w W_\mu^3 + g_Y B_\mu & \sqrt{2} g_w W_\mu^- \\ \sqrt{2} g_w W_\mu^+ & -g_w W_\mu^3 + g_Y B_\mu \end{pmatrix} \begin{pmatrix} 0 \\ v \end{pmatrix} \right|^2$$

$$= \left(\frac{g_w v}{2} \right)^2 W_\mu^+ W_\mu^- + \frac{v^2}{8} \begin{pmatrix} W_\mu^3 & B_\mu \end{pmatrix} \underbrace{\begin{pmatrix} g_w^2 & -g_w g_Y \\ -g_w g_Y & g_Y^2 \end{pmatrix}}_{det\mathcal{M}=0} \begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix}. \tag{2.33}$$

The mass of the charged $\mathrm{W^\pm}$ bosons is given by the first term and is equal to $\mathrm{m_W} = \mathrm{g_w} v/2$. The second term contains a mass matrix with zero determinant, which is not unexpected since one of the eigenstates needs to be the massless photon (A_μ). Thus, we get the two

neutral states:

$$A_\mu = \frac{g_w B_\mu + g_Y W_\mu^3}{\sqrt{g_w^2 + g_Y^2}} \quad \text{with} \quad m_A = 0, \quad \text{and}$$

$$Z_\mu = \frac{g_w W_\mu^3 - g_Y B_\mu}{\sqrt{g_w^2 + g_Y^2}} \quad \text{with} \quad m_Z = \frac{v}{2}\sqrt{g_w^2 + g_Y^2}. \tag{2.34}$$

Introducing the *Weinberg* (or *weak*) angle θ_W [5]:

$$\cos\theta_W \equiv \frac{g_w}{\sqrt{g_w^2 + g_Y^2}}, \quad \sin\theta_w \equiv \frac{g_Y}{\sqrt{g_w^2 + g_Y^2}}, \tag{2.35}$$

the gauge bosons A_μ and Z_μ can be expressed as:

$$A_\mu = \cos\theta_W B_\mu + \sin\theta_W W_\mu^3$$

$$Z_\mu = \cos\theta_W W_\mu^3 - \sin\theta_W B_\mu. \tag{2.36}$$

The mass of the Z boson is also related to the $W^\pm$ boson mass: $m_Z \cos\theta_W = m_W$. The overall scalar Lagrangian $\mathcal{L}_\phi$ of equation 2.29 can then be written in terms of three parts:

$$\mathcal{L}_\phi = \frac{1}{4}\lambda v^4 + \mathcal{L}_H + \mathcal{L}_{Hgauge}, \tag{2.37}$$

where the first term is the scalar potential, the second term is the kinetic term of a new massive and scalar boson, the Higgs boson, and the third is the kinetic term correlating this new scalar with the gauge bosons. The $\mathcal{L}_H$ is equal to:

$$\mathcal{L}_H = \frac{1}{2}\partial_\mu H \partial^\mu H - \frac{1}{2}m_H^2 H^2 - \frac{m_H^2}{2v}H^3 - \frac{m_H^2}{8v^2}H^4, \tag{2.38}$$

where the Higgs boson mass is given by: $m_H = \sqrt{-2\mu^2} = \sqrt{2\lambda}v$, and is a free parameter of the SM. The cubic and quartic terms reveal that it is possible for three or four Higgs bosons to interact via a single vertex. The term $\mathcal{L}_{Hgauge}$ is given by:

$$\mathcal{L}_{Hgauge} = m_W^2 W_\mu^+ W^\mu \left(1 + \frac{2}{v}H + \frac{H^2}{v^2}\right) + \frac{1}{2}m_Z^2 Z_\mu Z^\mu \left(1 + \frac{2}{v}H + \frac{H^2}{v^2}\right), \tag{2.39}$$

The first terms in each parenthesis correspond to the mass terms, while the remaining terms give rise to interactions between the scalar and the gauge bosons, the strengths of which are always proportional to the squared mass of the coupled gauge boson.

Fermion masses

Fermions can acquire mass through the *Yukawa interactions* that couple a right-handed fermion with its left-handed doublet and the scalar field, after SSB. To generate masses for the first family of fermions, the following gauge-invariant fermion-scalar coupling must

be included in the SM Lagrangian:

$$\mathcal{L}_{\text{Yukawa}} = f_e \bar{\ell}_L \phi e_R + f_u \bar{q}_L \tilde{\phi} u_R + f_d \bar{q}_L \phi d_R + (\text{h.c}). \tag{2.40}$$

where $f_{e,u,d}$ are the Yukawa couplings for electrons, up-, and down-type quarks, and the scalar doublet $\tilde{\phi}$ corresponds to the charge conjugate of the Higgs field ($\tilde{\phi} = i\sigma_2\phi^*$) and has a hypercharge of $-1/2$. After SSB, the Yukawa Lagrangian takes the simpler form:

$$\mathcal{L}_{\text{Yukawa}} = \frac{1}{\sqrt{2}}(v + H)\left\{ f_e \underbrace{(\bar{e}_L e_R + \bar{e}_R e_L)}_{\bar{e}e} + f_u(\bar{u}_L u_R + \bar{u}_R u_L) + f_d(\bar{d}_L d_R + \bar{d}_R d_L) \right\} \tag{2.41}$$

from which the masses for the fermions can be read off:

$$m_i = -\frac{f_i v}{\sqrt{2}}, \quad i = e, u, d. \tag{2.42}$$

The Yukawa coupling constants f_i are arbitrary and are inferred from measurements of the fermion masses. Since neutrinos do not have right-handed partners in the SM, they cannot acquire mass terms through the Yukawa coupling. Besides the mass terms, the Yukawa Lagrangian contains interaction terms with the Higgs boson, the strength of which is proportional to the mass of the fermion.

To summarize, the Lagrangian of the SM is constructed based on the strong and electroweak interactions. The BEH mechanism gives masses to the gauge bosons and fermions through the SSB of the $\text{SU}(3)_\text{C} \times \text{SU}(2)_\text{L} \times \text{U}(1)_\text{Y}$ to $\text{SU}(3)_\text{C} \times \text{U}(1)_\text{QED}$. The degrees of freedom before and after SSB remain the same. Before SSB, the complex doublet ϕ holds four degrees of freedom, the massless spin-1 gauge bosons B_μ and the three W_μ have two possible polarizations that double their degrees of freedom to eight, making a total of twelve degrees of freedom. After SSB, only one real massive scalar boson exists with one degree of freedom. The photon remains massless and has two degrees of freedom, while the three massive weak gauge bosons $\text{W}^\pm$ and Z^0 have six and three degrees of freedom, respectively.

2.2   Shortcomings of the SM

The SM constitutes one of the most brilliant achievements in modern physics. However, there are still many open questions that the SM fails to answer. Gravity, one of the fundamental forces in nature, is not included in the SM. There is no explanation for the origin of the three generations of fermions and their mixing, nor for the origin of Dark Matter (DM) and Dark Energy (DE). The mechanism responsible for the observed matter-antimatter asymmetry is still unknown. The implication of all these shortcomings is profound: the SM is incomplete and further work is needed to fully understand nature. In the following, some of the most pressing problems of the SM are discussed.

Neutrino oscillations

The SM successfully describes all existing particle physics data with the exception of neutrino masses. Neutrinos in the SM are part of $\mathrm{SU}(2)_\mathrm{L}$ doublets. There exist three neutrinos, one for each charged lepton and they do not have strong nor electromagnetic interactions. They are only involved in weak charged current (CC) interactions with their corresponding charged leptons and in neutral current (NC) interactions among themselves:

$$
\begin{aligned}
-\mathcal{L}_\mathrm{CC} &= \frac{g_w}{2} \sum_\ell \bar{\nu}_{L,\ell} \gamma^\mu \ell_L^- W_\mu^+ + (h.c.), \\
-\mathcal{L}_\mathrm{NC} &= \frac{g_w}{2\cos\theta_\mathrm{W}} \sum_\ell \bar{\nu}_{L,\ell} \gamma^\mu \nu_{L,\ell} Z_\mu^0.
\end{aligned}
\tag{2.43}
$$

The $\mathcal{L}_\mathrm{NC}$ determines the width of invisible decays of Z^0, the measurement of which implies that only three ($N_\nu = 2.984 \pm 0.008$ [8]) light neutrinos exist in the SM or any extension of it. To date, there is no explanation why only three generations of neutrinos, or fermions in general, must exist. Moreover, a series of experimental observations [33] involving solar, atmospheric, accelerator, and reactor neutrinos gave strong evidence that neutrinos do have mass [34]. Neutrinos of one flavor were observed to *oscillate* into another flavor while traveling, implying that neutrinos have non-zero mass. In order to describe these flavor transitions in the SM, a 3×3 mixing matrix analogous to the CKM matrix for the quarks needs to be introduced in the SM to connect these flavor eigenstates with the mass eigenstates.

Dark matter

The nature and properties of this mysterious substance dominating the dynamics of the universe called DM, is among the most crucial problems that physicists try to solve. The observation of approximately flat rotational curves of gas in galaxies in the 1930s and high-velocity dispersions of galaxies in galaxy clusters in the 1970s [35, 36] gave strong evidence for the hypothesis of missing mass that escapes detection. Since then, many direct and indirect attempts have been made to identify its nature, including measurements of its gravitational lensing effects, and measurements of the Cosmic Microwave Background (CMB) anisotropies. Combined measurements of the CMB anisotropies and Large Scale Structure (LSS) of the Planck collaboration [9] show that only 5% of the energy content of the universe originates from the SM baryonic matter, while 27% appears as DM and 68% as DE. Several models beyond the SM, such as Supersymmetry (SUSY), and models with rich dark sectors, predict the existence of different DM candidates. One of the most popular hypotheses is that non-baryonic DM consists of Weakly Interacting Massive Particles (WIMPs). WIMPs must be electrically neutral, weakly interacting, stable (or at least long-lived) to survive till today, and massive enough or non-relativistic at the time of structure formation. At the LHC, direct searches for DM particles have been performed under the assumption that they escape the detectors without interacting with ordinary matter, and therefore, result in significant amounts

of missing energy and momentum. Unfortunately, no strong evidence for DM particles has been observed so far.

The hierarchy problem

Heavy quarks, such as the top quark, give rise to large radiative corrections to the Higgs boson mass. The one-loop Feynman diagram illustrated in Fig. 2.6 yields a quantum correction to the Higgs boson squared mass of:

$$\Delta m_H^2 = -\frac{3|f_t|^2}{8\pi^2}\Lambda^2 \tag{2.44}$$

where f_t is the top quark Yukawa coupling and Λ is a momentum cutoff introduced to regularize the loop integral. The Λ corresponds to the scale in which new physics is expected to exist. Knowing that quantum gravitational effects become important at the Planck scale it is natural to choose the Λ near that scale. However, the quadratic divergences on the Λ drag the mass of the Higgs boson to extremely high values. This disturbing sensitivity implies an enormous and rather unnatural fine-tuning to cancel these corrections and bring the Higgs boson mass back to the electroweak scale. One of the theories beyond the SM suggesting a solution to the hierarchy problem without applying excessive fine-tuning is SUSY.
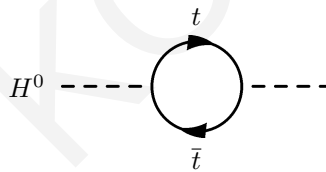
Figure 2.6: One-loop radiative corrections to the Higgs boson mass from the top quark.

Matter-antimatter asymmetry

The mechanism to generate the observed matter-antimatter asymmetry in the universe is still one of the key questions that are still unanswered. At the beginning of time, equal amounts of matter and antimatter were created. One would expect matter and anti-matter to annihilate into pure energy, producing photons and nothing else. However, the universe as we know it today is dominated by matter. To explain this asymmetry, Andrei Sakharov (1967) [37] proposed the idea of the violation of the combined symmetry of charge conjugation (C) and parity (P). The former conjugates all internal quantum numbers, meaning that it transforms particles into antiparticles and vice versa, while the latter inverts all spacial coordinates. A violation of the combined $\mathcal{CP}$ symmetry allows for a total distinction between matter and antimatter to be made. $\mathcal{CP}$ violation is embedded in the weak interactions of the SM via the complex phase of the CKM matrix. Measurements on $\mathcal{CP}$ violation in the quark sector have been performed for over 50 years in K, D, and B meson systems. However, the

predictions on the number of baryons derived by the SM still fall several orders of magnitude short compared to what is observed by astronomers. Additional sources of $\mathcal{CP}$ violation are required to explain the matter-antimatter asymmetry and several extensions of the SM naturally include them.

2.3 Theories beyond the Standard Model

Theoretical developments beyond the Standard Model (BSM) aim to solve the deficiencies of the SM. Several of these theories argue for the presence of extra scalar particles. The most straightforward minimal approach is to extend the SM with an additional singlet field or a Higgs doublet. When doing so, one has to be careful not to violate experimental and theoretical constraints. The most significant constraints are given by the ρ parameter and the non-existence of Flavor Changing Neutral Currents (FCNC). The ρ parameter is defined as

$$\rho = \frac{M_W^2}{M_Z^2 \cos^2\theta_W} \tag{2.45}$$

has been measured experimentally and found to be very close to unity, whereas in the SM it takes automatically the tree-level value of 1. For n scalar multiplets ϕ_i of weak isospin I_i, weak hypercharge Y_i and VEV υ_i of the neutral components, the tree-level ρ parameter can be expressed as:

$$\rho = \frac{\sum_{i=1}^{n} \left[I_i(I_i + 1) - \frac{1}{4}Y_i^2 \right] \upsilon_i}{\sum_{i=1}^{n} \frac{1}{2}Y_i^2 \upsilon_i}, \tag{2.46}$$

and both, $\mathrm{SU}(2)$ singlets ($Y = 0$) and doublets ($Y = \pm 1$) satisfy $I(I + 1) = \frac{3}{4}Y^2$, and thus, $\rho = 1$. FCNC are not allowed in the SM and this absence is also confirmed by experimental data. Models with extended Higgs sectors, however, can include FCNC terms. Charged Higgs bosons are predicted in such theories, including two-Higgs-doublet Models (2HDMs) and Supersymmetry (SUSY).

2.3.1 Two-Higgs Doublet Models

The simplest possible extension of the SM is called the 2HDM and includes two complex $\mathrm{SU}(2)_\mathrm{L}$ Higgs doublets of hypercharge $Y = +1$:

$$\Phi_i = \begin{pmatrix} \phi_i^\dagger \\ \phi_i^0 \end{pmatrix}, \tag{2.47}$$

where $i = 1, 2$ is the Higgs flavor index. The most general and renormalizable scalar potential of the two doublets Φ_1 and Φ_2 is:

$$
V_{\text{Higgs}} = m_{11}^2 \Phi_1^\dagger \Phi_1 + m_{22}^2 \Phi_2^\dagger \Phi_2 - \left[m_{12}^2 \Phi_1^\dagger \Phi_2 + \text{h.c.} \right] + \frac{\lambda_1}{2} (\Phi_1^\dagger \Phi_1)^2 +
$$
$$
\frac{\lambda_2}{2} (\Phi_2^\dagger \Phi_2)^2 + \lambda_3 (\Phi_1^\dagger \Phi_1)(\Phi_2^\dagger \Phi_2) + \lambda_4 (\Phi_1^\dagger \Phi_2)(\Phi_2^\dagger \Phi_1) + \tag{2.48}
$$
$$
\left[\frac{\lambda_5}{2} (\Phi_1^\dagger \Phi_2)^2 + \lambda_6 (\Phi_1^\dagger \Phi_1)(\Phi_1^\dagger \Phi_2) + \lambda_7 (\Phi_2^\dagger \Phi_2)(\Phi_1^\dagger \Phi_2) + \text{h.c.} \right],
$$

where the symbols m_{11}^2, m_{22}^2, m_{12}, denote the mass matrix elements and λ_{1-7} represent the scalar Higgs self-couplings. Following the hermiticity of the above potential, the parameters m_{11}^2, m_{22}^2, and $\lambda_{1,2,3,4}$ are real, whereas the remaining $\lambda_{5,6,7}$ and m_{12}^2 can be complex, giving a total of 14 degrees of freedom. For simplicity, most phenomenological studies of the 2HDMs make several assumptions. Under the assumption of $\mathcal{CP}$-conservation, the parameters λ_5 and m_{12}^2 become real. Based on the Glashow-Weinberg-Paschos (GWP) theorem [38], no tree-level FCNC exists when all fermions of a given electric charge and chirality couple to the same $\text{SU}(2)_\text{L}$ doublet. A generic 2HDM includes tree-level FCNC which can be eliminated by imposing discrete symmetries. By invoking a $\mathbb{Z}_2$ symmetry, $\Phi_1 \to \Phi_1$, $\Phi_2 \to -\Phi_2$, the quartic terms proportional to λ_6 and λ_7 are absent, and the scalar potential is simplified as follows:

$$
V_{\text{Higgs}} = m_{11}^2 \Phi_1^\dagger \Phi_1 + m_{22}^2 \Phi_2^\dagger \Phi_2 - m_{12}^2 (\Phi_1^\dagger \Phi_2 + \Phi_2^\dagger \Phi_1) + \frac{\lambda_1}{2} (\Phi_1^\dagger \Phi_1)^2 + \frac{\lambda_2}{2} (\Phi_2^\dagger \Phi_2)^2
$$
$$
+ \lambda_3 \Phi_1^\dagger \Phi_1 \Phi_2^\dagger \Phi_2 + \lambda_4 \Phi_1^\dagger \Phi_2 \Phi_2^\dagger \Phi_1 + \frac{\lambda_5}{2} \left[(\Phi_1^\dagger \Phi_2)^2 + (\Phi_2^\dagger \Phi_1)^2 \right]. \tag{2.49}
$$

All the parameters of V_{Higgs} are now real. The EWK symmetry is broken since the minimum of the scalar potential occurs at non-zero VEVs. The minimization of the scalar potential ($\frac{\partial V}{\partial \Phi_i} = 0$) for a region of parameter space yields two VEVs:

$$
\langle \Phi_1 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v_1 \end{pmatrix}, \quad \langle \Phi_2 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v_2 \end{pmatrix}, \tag{2.50}
$$

and the total VEV should be a combination of both, $v = \sqrt{v_1^2 + v_2^2}$. The two complex scalar doublets can be expressed in terms of eight real fields:

$$
\Phi_a = \begin{pmatrix} \phi_a^+ \\ \frac{1}{\sqrt{2}} (v_a + \rho_a + i\eta_a) \end{pmatrix}, \quad a = 1, 2. \tag{2.51}
$$

Three out of the above eight fields provide the longitudinal degrees of freedom for the massive $\text{W}^\pm$ and Z bosons, while the rest give rise to five physical Higgs bosons: two neutral $\mathcal{CP}$-even scalars (h, H), one neutral $\mathcal{CP}$-odd pseudoscalar (A), and a pair of charged Higgs bosons ($\text{H}^\pm$). Expanding the Lagrangian around the minima gives the following mass term

for the charged Higgs bosons:

$$\mathcal{L}_{\phi^\pm,\, mass} = -\left(m_{12}^2 - (\lambda_4 + \lambda_5)\frac{v_1 v_2}{2}\right)(\phi_1^-, \phi_2^-)\underbrace{\begin{pmatrix} \frac{v_2}{v_1} & -1 \\ -1 & \frac{v_1}{v_2} \end{pmatrix}}_{\mathcal{M}_{\mathcal{C}}}\begin{pmatrix} \phi_1^+ \\ \phi_2^+ \end{pmatrix}. \tag{2.52}$$

By diagonalizing the mass matrix $\mathcal{M}_{\mathcal{C}}$ with the transformation matrix:

$$\mathcal{U}_C = \begin{pmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{pmatrix}, \tag{2.53}$$

where $\tan\beta = \frac{v_2}{v_1}$, and exploiting the minimum potential conditions, the mass squared for the charged Higgs boson reads:

$$m_{H^\pm}^2 = \left(\frac{m_{12}^2}{v_1 v_2} - \frac{\lambda_4 + \lambda_5}{2}\right)(v_1^2 + v_2^2). \tag{2.54}$$

The mass matrix $\mathcal{M}_{\mathcal{C}}$ has one zero eigenvalue, which corresponds to the charged Goldstone bosons $G^\pm$ that get absorbed to give mass to the $W^\pm$ bosons. The corresponding mass term for the pseudoscalar is:

$$\mathcal{L}_{\eta,\, mass} = -\frac{1}{2}\frac{m_A^2}{v_1^2 + v_2^2}(\eta_1, \eta_2)\underbrace{\begin{pmatrix} v_2^2 & -v_1 v_2 \\ -v_1 v_2 & v_1^2 \end{pmatrix}}_{\mathcal{M}_P}\begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix}, \tag{2.55}$$

leading to one neutral Goldstone boson G^0 that gets absorbed to give mass to the Z boson, and the pseudoscalar A, with mass squared equal to:

$$m_A^2 = \left(\frac{m_{12}^2}{v_1 v_2} - \lambda_5\right)(v_1^2 + v_2^2). \tag{2.56}$$

The mass terms for the $\mathcal{CP}$-even mass eigenstates h and H are of the form:

$$\mathcal{L}_{\rho,\, mass} = -\frac{1}{2}\underbrace{\begin{pmatrix} m_{12}^2\frac{v_2}{v_1} + \lambda_1 v_1^2 & -m_{12}^2 + \lambda_{345} v_1 v_2 \\ -m_{12}^2 + \lambda_{345} v_1 v_2 & m_{12}^2\frac{v_1}{v_2} + \lambda_2 v_2^2 \end{pmatrix}}_{\mathcal{M}_N}\begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix}, \tag{2.57}$$

where $\lambda_{345} \equiv \lambda_3 + \lambda_4 + \lambda_5$. The mass matrix $\mathcal{M}_N$ can be diagonalized using the matrix

$$\mathcal{U}_N = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix}, \tag{2.58}$$

where the mixing angle α is given as a function of the matrix elements of $\mathcal{M}_N$. The masses of h and H are then given by:

$$m_{H,h}^2 = \frac{1}{2}\left(\mathcal{M}_{11} + \mathcal{M}_{22} \pm \sqrt{(\mathcal{M}_{11} - \mathcal{M}_{22})^2 + 4\mathcal{M}_{12}^2}\right), \tag{2.59}$$

where $\mathcal{M}_{11}, \mathcal{M}_{22}, \mathcal{M}_{12}$ are the matrix elements of $\mathcal{M}_N$. It is often convenient to express the scalar doublets in the *Higgs basis* [39]:

$$H_1 = \begin{pmatrix} H_1^+ \\ H_1^0 \end{pmatrix} \equiv \Phi_1 \cos\beta + \Phi_2 \sin\beta, \quad H_2 = \begin{pmatrix} H_2^+ \\ H_2^0 \end{pmatrix} \equiv -\Phi_1 \sin\beta + \Phi_2 \cos\beta, \tag{2.60}$$

where the VEVs of these fields are $\langle H_1^0 \rangle = v/\sqrt{2}$ and $\langle H_2^0 \rangle = 0$. In this basis, the scalar doublet H_1 has the same tree-level couplings to all SM particles as the SM Higgs boson. The physical neutral $\mathcal{CP}$-even Higgs bosons are then given by:

$$\begin{pmatrix} h \\ H \end{pmatrix} = \begin{pmatrix} -\sin(\beta - \alpha) & \cos(\beta - \alpha) \\ \cos(\beta - \alpha) & \sin(\beta - \alpha) \end{pmatrix} \begin{pmatrix} \mathrm{Re}(H_1^0) - v \\ \mathrm{Re}(H_2^0) \end{pmatrix}. \tag{2.61}$$

If one of the physical states is *aligned* with $\mathrm{Re}(H_1^0) - v$, then it obtains the tree-level couplings of the Higgs boson predicted by the SM. For the light Higgs boson h (heavy Higgs boson H) this occurs in the *alignment limit* $\cos(\beta - \alpha) \to 0$ ($\sin(\beta - \alpha) \to 0$).

The Yukawa Lagrangian of the 2HDM model for the $Q = -1/3$ quarks is given by:

$$\mathcal{L}_Y = y_{ij}^1 \bar{\psi}_i \psi_j \Phi_1 + y_{ij}^2 \bar{\psi}_i \psi_j \Phi_2, \tag{2.62}$$

where the indices i and j are generation indices and the couplings $y_{ij}^{1,2}$ are 3×3 complex matrices in flavor space. The corresponding mass matrix is then:

$$M_{ij} = y_{ij}^1 \frac{v_1}{\sqrt{2}} + y_{ij}^2 \frac{v_2}{\sqrt{2}}. \tag{2.63}$$

While in the SM the diagonalization of the mass matrix automatically diagonalizes the Yukawa interactions, in the 2HDM the Yukawa couplings y^1 and y^2 cannot be diagonalized simultaneously. Thus, neutral Higgs scalars can mediate FCNC, i.e. $\bar{d}s\phi$. Following the GWP theorem, discrete or continuous symmetries must be introduced. There exist four possibilities for $\mathcal{CP}$-conserving 2HDM as reported in Table 2.2:

- type-I 2HDM: all quarks couple to just one of the Higgs doublets, the Φ_2. To realize this, a discrete symmetry $\Phi_1 \to -\Phi_1$ is imposed.

- type-II 2HDM: right-handed quarks with $Q = 2/3$ couple to one of the Higgs doublets, the Φ_2, whereas the right-handed $Q = -1/3$ quarks couple to Φ_1. This is enforced by the discrete symmetry $\Phi_1 \to -\Phi_1$, $d_R^i \to -d_R^i$.

Model	u_R^i	d_R^i	e_R^i
type-I	Φ_2	Φ_2	Φ_2
type-II	Φ_2	Φ_1	Φ_1
lepton specific	Φ_2	Φ_2	Φ_1
flipped	Φ_2	Φ_1	Φ_2

Table 2.2: The four types of 2HDM leading to natural flavor conservation.

- *lepton-specific* 2HDM: the right-handed leptons couple to Φ_1, and the right-handed quarks to the Φ_2.

- *flipped* 2HDM: same as type-II, with the exception of right-handed leptons that now couple to Φ_1.

In both scenarios, type-I and type-II, the right-handed leptons satisfy the same symmetry as the down-type quarks. In a general 2HDM, interpretation is performed in terms of the *physical* free parameters. These are the four masses of the Higgs bosons h, H, A^0, and $H^\pm$, the mixing angles α and β, and the m_{12}^2 parameter, which is related to the soft-breaking scale of the discrete symmetry.

2.3.2 Supersymmetry

Supersymmetry (SUSY) is a *space-time* symmetry relating fermions with bosons and vice versa, via the transformation $Q|\text{Boson}\rangle = |\text{Fermion}\rangle$ and $Q|\text{Fermion}\rangle = |\text{Boson}\rangle$. In the simplest supersymmetric models, the fermionic operator Q obeys the anticommutation relations:

$$\left\{Q_\alpha, \bar{Q}_{\dot{\beta}}\right\} = 2\sigma_{\alpha\dot{\beta}}^\mu P_\mu$$
$$\left\{Q_\alpha, Q_\beta\right\} = \left\{\bar{Q}_{\dot{\alpha}}, \bar{Q}_{\dot{\beta}}\right\} = 0,$$

(2.64)

where the $\alpha, \dot{\beta}$ are spinor indices corresponding to the left- and right-handed Weyl components, the P_μ corresponds to the space-time translation operator, and σ^μ are the Pauli matrices. The operator Q changes the spin of a particle by half, meaning that for each particle there exists a supersymmetric partner. Collectively, particles and their partners are called *superpartners*. Superpartners are arranged in *supermultiplets*, containing equal number of fermionic and bosonic degrees of freedom. The operator Q commutes with the space-time translation operator and with internal quantum numbers but does not commute with Lorentz generators, implying that particles inside the same supermultiplet have different spin but share the same mass and quantum numbers. Supersymmetric models are particularly attractive for several reasons. Among the most important ones is that SUSY provides an elegant solution to the hierarchy problem. The introduction of superpartners cancels the quadratically divergent terms in the radiative corrections of the Higgs boson squared mass at all orders in perturbation theory. Taking as an example the one-loop corrections due to the top
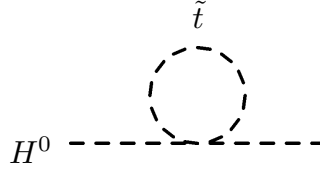
Figure 2.7: One-loop radiative corrections to the Higgs boson squared mass m_H^2, due to the top squark.

squark, shown in Fig. 2.7, the corrections on the Higgs boson squared mass take the form:

$$\Delta m_H^2 = \frac{3|f_t|^2}{16\pi^2}\Lambda^2 - \frac{3|f_t|^2 m_{\tilde{t}}^2}{8\pi^2}\log\left(\frac{\Lambda}{m_{\tilde{t}}}\right). \tag{2.65}$$

Due to its positive sign, the first term cancels the correction from the top quark from equation 2.44. Thus, the sensitivity of the Higgs boson squared mass to the cutoff scale Λ is now logarithmic instead of quadratic. In addition to the hierarchy solution, superpartners can modify the running of the strong and electroweak gauge couplings and eventually unify them at high energies. Superpartners have opposite $R-$parity compared to the SM particles. The $R-$parity is defined as:

$$P_R = (-1)^{2s+3B+L}, \tag{2.66}$$

where s is the spin of a particle, B is the baryon number, and L is the lepton number. Since no lepton and baryon number violating terms appear in the SM Lagrangian, all SM particles have even $R-$parity ($+1$), while all superpartners have odd $R-$parity (-1). If $R-$parity is conserved in nature, there are some interesting phenomenological implications for the Lightest Supersymmetric Particle (LSP). First, the LSP must be stable, meaning that the decay chain of all other SUSY particles will end up to the LSP. Given that it is also color and electrically neutral, the LSP constitutes a plausible candidate for DM. In addition, since the initial state particles at the LHC have even $R-$parity, SUSY particles can only be produced in pairs to conserve $R-$parity. Nevertheless, SUSY cannot be an *exact* symmetry as no mass degeneracy in the elementary particle spectrum has been observed in any of the high energy experiments for energies up to the TeV scale. This leads to the conclusion that SUSY must be a *broken* symmetry. The breaking of SUSY can be realized by introducing *soft* breaking terms in the Lagrangian such that they do not introduce new quadratic divergences, are renormalizable, and preserve the gauge invariance of the theory [40, 41].

The minimal supersymmetric standard model

The simplest supersymmetric extension of the SM containing the minimum possible additional particles is the Minimal Supersymmetric Standard Model (MSSM), which is a type-II 2HDM. Particles in the MSSM are grouped into *chiral* and *gauge* supermultiplets. Gauge supermultiplets accommodate the spin-1 SM gauge bosons and their corresponding spin-$\frac{1}{2}$

superpartners, called *gauginos*. Chiral supermultiplets accommodate all SM fermions and their bosonic superpartners, which are complex scalars called *sfermions*. The Higgs sector of the MSSM includes two Higgs doublets with opposite hypercharge, both residing in chiral supermultiplets. The Higgs doublet H_u is responsible for giving masses to up-type quarks, while the H_d gives masses to down-type quarks and charged leptons. The supermultiplets together with their representations in the MSSM gauge group are shown in Table 2.3.

Table 2.3: The chiral and gauge supermultiplets in the MSSM.

	Supermultiplets		spin-0	spin-1/2	$\mathrm{SU(3)_C}$	$\mathrm{SU(2)_L}$	$\mathrm{U(1)_Y}$
Chiral	squarks, quarks	Q	$(\,\tilde{u}_L\ \tilde{d}_L\,)$	$(\,u_L\ d_L\,)$	**3**	**2**	$1/6$
		U^c	$\tilde{u}_R^*$	$u_R^\dagger$	$\bar{\mathbf{3}}$	**1**	$-2/3$
		D^c	$\tilde{d}_R^*$	$d_R^\dagger$	$\bar{\mathbf{3}}$	**1**	$1/3$
	sleptons, leptons	L	$(\,\tilde{\nu}\ \tilde{e}_L\,)$	$(\,\nu\ e_L\,)$	**1**	**2**	$-1/2$
		E^c	$\tilde{e}_R^*$	$e_R^\dagger$	**1**	**1**	1
	Higgs, higgsinos	H_u	$(\,H_u^+\ H_u^0\,)$	$(\,\tilde{H}_u^+\ \tilde{H}_u^0\,)$	**1**	**2**	$1/2$
		H_d	$(\,H_d^0\ H_d^-\,)$	$(\,\tilde{H}_d^0\ \tilde{H}_d^-\,)$	**1**	**2**	$-1/2$
Gauge	gluino, gluon		$\tilde{g}$	g	**8**	**1**	0
	winos, W boson		$\tilde{W}^\pm, \tilde{W}^0$	$W^\pm, W^0$	**1**	**3**	0
	bino, B boson		$\tilde{B}^0$	B^0	**1**	**1**	0

To calculate the mass eigenstates of the physical Higgs bosons, one has to consider the scalar potential of the Higgs sector, defined as:

$$V_{\text{Higgs}} = (|\mu|^2 + m_{H_u}^2)(|H_u^0|^2 + |H_u^+|^2) + (|\mu|^2 + m_{H_d}^2)(|H_d^0|^2 + |H_d^-|^2)$$
$$+ \left[b(H_u^+ H_d^- - H_u^0 H_d^0) + c.c.\right] + \frac{1}{8}(g_w^2 + g_Y^2)(|H_u^0|^2 + |H_u^+|^2 - |H_d^0|^2 - |H_d^-|^2)^2 \quad (2.67)$$
$$+ \frac{1}{2}g_w^2|H_u^+ H_d^{0*} + H_u^0 H_d^{-*}|^2.$$

and require its minimum to undergo EWK symmetry breaking down to electromagnetism, in analogy to the SM and the 2HDM. The g_w and g_Y are the gauge couplings of $\mathrm{SU(2)}$ and $\mathrm{U(1)_Y}$ and μ is the so-called higgsino mass parameter. The term proportional to b arises from the soft SUSY breaking Lagrangian. Since the charged components of the Higgs doublets cannot have VEVs, they can be set to zero, and the potential is simplified:

$$V_{\text{Higgs}}^{\text{neutral}} = (|\mu|^2 + m_{H_u}^2)|H_u^0|^2 + (|\mu|^2 + m_{H_d}^2)|H_d^0|^2 - (bH_u^0 H_d^0 + c.c.)$$
$$+ \frac{1}{8}(g_w^2 + g_Y^2)(|H_u^0|^2 - |H_d^0|^2)^2. \quad (2.68)$$

Since the EWK symmetry breaking pattern is the same as in the 2HDM, the same set of would-be Goldstone bosons are expected to give masses to the gauge bosons $W^\pm$ and Z^0. The remaining five physical Higgs bosons are the neutral scalar Higgs bosons h and H, the pseudoscalar A, and two charged Higgs bosons $H^\pm$. The eigenvalues for $H^\pm$ and A^0 are
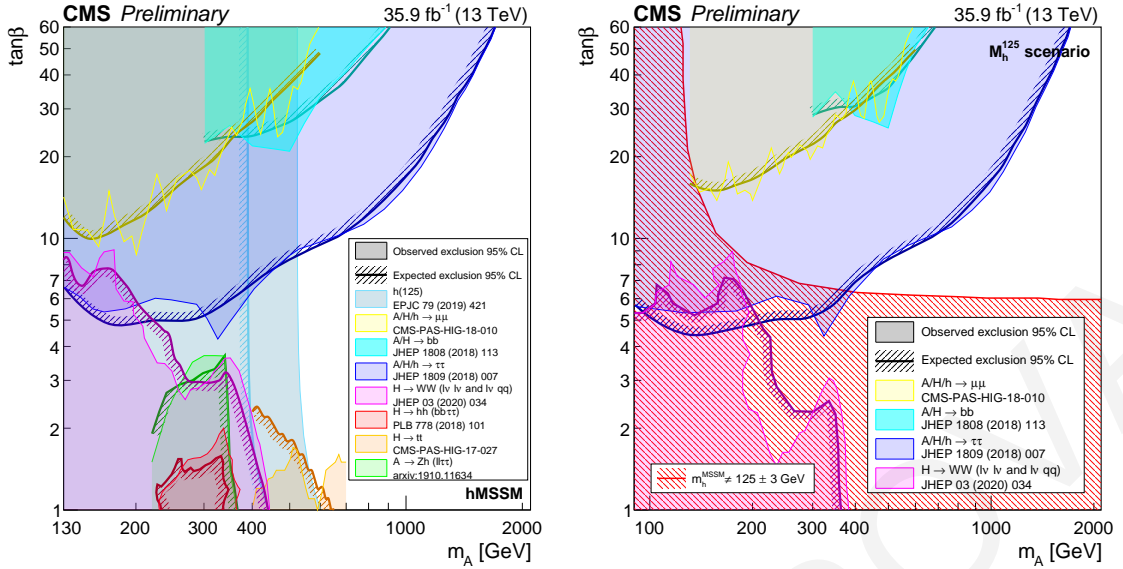
Figure 2.8: Observed and expected upper limits for $\tan\beta$ as a function of the pseudoscalar mass, m_A, in the MSSM hMSSM (*left*) and M_h^{125} (*right*) scenarios.

correlated as follows:

$$m_{H^\pm}^2 = b(\cot\beta + \tan\beta) + m_W^2$$
$$m_{A^0} = b(\cot\beta + \tan\beta) \tag{2.69}$$
$$\implies m_{H^\pm}^2 = m_A^2 + m_W^2$$

where $\tan\beta$ is the ratio of the two VEVs. At tree-level, $m_{H^\pm} \geq m_W$ and $m_{H^\pm} \geq m_A$. It is worth mentioning that since the charged scalar and pseudoscalar masses are so close, the decay of the charged Higgs boson into a pseudoscalar and a real $W^\pm$ boson is kinematically forbidden in the MSSM, while it is generally allowed in the type-II 2HDM. Moreover, the scalar self-couplings of the MSSM are given in terms of the gauge couplings, while in the 2HDM these are arbitrary. As a consequence, the MSSM has a strict tree-level upper bound on the mass of the lightest Higgs boson h, which is required to be below the Z boson mass [42]. However, the squared mass of h is subject to relatively drastic quantum corrections, which mainly come from top and stop quark loops and raise m_h [43]. In contrast to 2HDM, the MSSM imposes strong constraints on the Higgs sector parameter space, leaving only two parameters to be free at the tree-level, which can be the $\tan\beta$ and m_{A^0}. Being one of the best motivated BSM models, the MSSM is extensively searched for at the LHC. The most recent constraints in the MSSM M_h^{125} and hMSSM benchmark scenarios [44, 45] from the CMS experiment are shown in Fig. 2.8.

2.4 Charged Higgs bosons at the LHC

At the LHC, there exist several channels for direct charged Higgs boson production: in association with a top quark, in association with a neutral Higgs ($\mathrm{pp} \to \mathrm{H}^\pm \mathrm{h_i}/\mathrm{A}$) [46, 47,

48] or a W ($pp \rightarrow H^{\pm}W^{\pm}$) boson, in pairs (with a very small rate) [49], or in Vector Boson Fusion (VBF) if the charged Higgs boson is part of an $SU(2)_L$ triplet. In the following, only the first production channel is considered as it is the most dominant. A comprehensive review on the direct production of charged Higgs bosons can be found in Refs. [50] and [49].

Charged Higgs boson production in association with a top quark

The production of a charged Higgs boson in association with a top quark is usually distinguished in two mass regimes, depending on whether the mass of the charged Higgs boson is below or above the top-bottom quark threshold. A *light* charged Higgs boson, with a mass below the top-bottom quark threshold ($m_t - m_b$), is mainly produced from the top quark decay $t \rightarrow H^+b$, as shown in Fig. 2.9. At the LHC, a top quark pair production happens through QCD interactions, and when kinematically allowed, the top quark can decay into a charged Higgs boson and a bottom quark. The production cross-section can be calculated by factorizing the production process of proton-proton collisions, $pp \rightarrow t\bar{t}$, times the branching ratio $t \rightarrow H^+b$, in the Narrow-Width Approximation (NWA) [51]. When crossing the
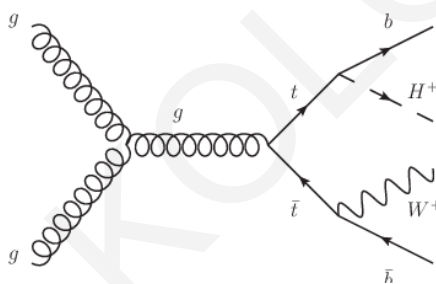
Figure 2.9: LO Feynman diagram for light charged Higgs boson production.

top-bottom quark mass threshold, *heavy* charged Higgs bosons can be produced in association with a top and a bottom quark, $pp \rightarrow \bar{t}bH^+$, as shown in Fig. 2.10. The inclusive cross-section for this production can be calculated with two different approaches. In the first approach, known as the *four-flavor scheme* (4FS), the bottom quark is not considered as a parton in the initial state but instead, it is generated as a massive final state. The cross-section calculation is then performed based on an effective theory with four light quarks, where the bottom quarks are decoupled and do not contribute to the computation of the running coupling constant, nor the evolution of the Parton Distribution Functions (PDFs). In this case, the lowest-order QCD production processes are the gluon-gluon fusion, $gg \rightarrow tbH^{\pm}$, shown in Fig. 2.10a, and the quark-antiquark annihilation, $q\bar{q} \rightarrow tbH^{\pm}$, shown in Fig. 2.10b. Gluon splittings into nearly collinear $b\bar{b}$ pairs can create large logarithms in the calculation, which are not summed to all orders of perturbation. This is can be done in the second approach, the *five-flavor scheme* (5FS) with the introduction of bottom parton densities on the approximation that, at leading order, the outgoing bottom quark has small transverse momenta and is massless. In this scheme, the LO process is $gb \rightarrow tH^{\pm}$, shown in Fig. 2.10c, while the gluon-gluon fusion and quark-antiquark annihilation processes are included at next-to-leading order
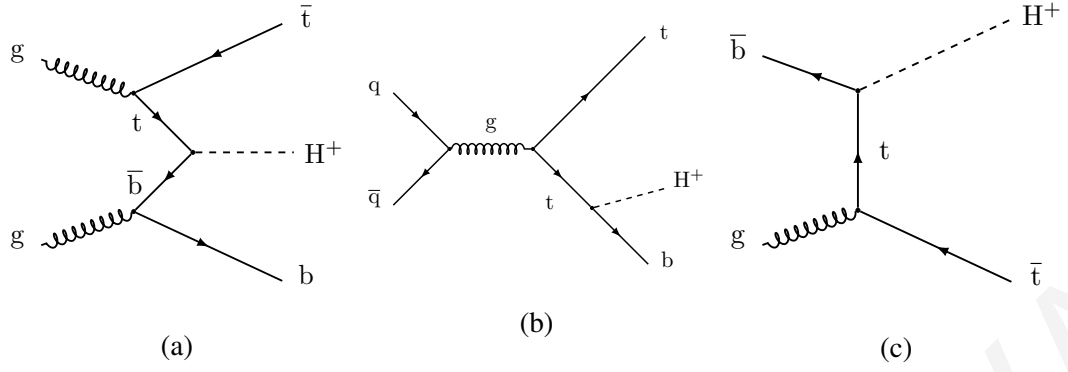
Figure 2.10: LO Feynman diagrams for the heavy charged Higgs boson production at the LHC. The direct $gg \rightarrow tbH^\pm$ production process is illustrated in (a), the charged Higgs-strahlung $q\bar{q} \rightarrow tbH^\pm$ process in the 4FS is shown in (b) and the gluon-gluon splitting process $gb \rightarrow tH^\pm$ process in the 5FS is shown in (c).

(NLO). The massless requirement on the bottom quark can also be improved with higher-order contributions from diagrams where the bottom quark is only present in the final state. The two schemes yield identical results when calculated at all orders in perturbation theory. At finite order, however, discrepancies arise since the ordering of the perturbative expansion is different. A final prediction for the heavy charged Higgs boson production cross-section can be calculated by combining both schemes at NLO, according to the *Santander matching* [52], as follows:

$$\sigma^{\text{matched}} = \frac{\sigma^{\text{4FS}} + w \cdot \sigma^{\text{5FS}}}{1 + w}, \tag{2.70}$$

where the weight w is defined as:

$$w = \ln\left(\frac{m_{H^\pm}}{m_b}\right) - 2. \tag{2.71}$$

The combined result is compatible with the asymptotic limits $m_{H^\pm}/m_b \rightarrow 1$ and $m_{H^\pm}/m_b \rightarrow \infty$, where both schemes agree. Figure 2.11, shows the resulting matched cross-section of a heavy charged Higgs boson of mass 200 GeV (left) and 2000 GeV (right) together with the 4FS and 5FS cross-section predictions, as a function of $\tan\beta$, calculated in a type-II 2HDM. The predictions can also be translated into cross-sections in the type-I, type-III, or type-IV 2HDM with a recipe described in Ref. [53].

For charged Higgs boson mass close to the top quark mass (145 - 200 GeV), the NWA becomes invalid. In this case, the effects from the finite top-width as well as the interplay between single resonant, double resonant and non-resonant top quark diagrams need to be taken into account [54].

Charged Higgs boson decay modes

A charged Higgs boson can decay into a fermion-antifermion pair: $t\bar{b}$, $c\bar{s}$, $\tau^+\nu_\tau$, $c\bar{b}$, to gauge bosons: $W^+\gamma$, W^+Z, and to a neutral Higgs boson and a gauge boson: h_iW^+. Since the cou-

(a) $m_{H^\pm} = 200$ GeV   (b) $m_{H^\pm} = 2000$ GeV
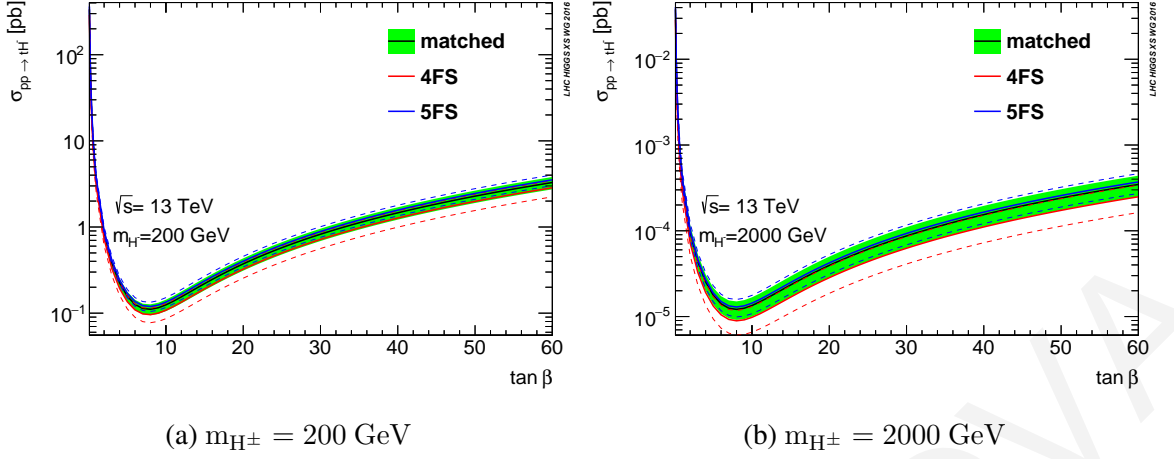
Figure 2.11: Inclusive total cross-section for the heavy charged Higgs boson production as calculated at NLO with the 4FS (5FS) scheme, shown with red (blue) line and the Santander matching result (black line) in a type-II 2HDM. Dashed lines (green band) represent the theoretical uncertainties for the two approaches (matched result) [55].

Model	X	Y	Z
type-I	$\cot\beta$	$\cot\beta$	$\cot\beta$
type-II	$\cot\beta$	-$\tan\beta$	-$\tan\beta$
lepton specific	$\cot\beta$	$\cot\beta$	-$\tan\beta$
flipped	$\cot\beta$	-$\tan\beta$	$\cot\beta$

Table 2.4: The parameters X, Y, and Z for the $\mathbb{Z}_2$-symmetric 2HDMs.

pling of the charged Higgs boson to fermions is determined by the Yukawa interactions of the model, the branching fractions to each decay mode is model-dependent. For all four 2HDM models, the most general Yukawa couplings to the charged Higgs boson can be written as:

$$\mathcal{L}_{H^\pm} = -H^+\left(\frac{\sqrt{2}V_{ud}}{\upsilon}\bar{u}(m_u X P_L + m_d Y P_R)d + \frac{\sqrt{2}m_\ell}{\upsilon}Z\bar{\nu}_L\ell_R\right) + \text{h.c.},\qquad(2.72)$$

where V_{ud} is the element of the CKM matrix corresponding to the charge $Q = +2/3$ quark and the charge $Q = -1/3$ quark. P_R and P_L are the projection operators defined in equation 2.14, and the values of X, Y, and Z depend on the particular 2HDM model considered and are given in Table 2.4. The branching fractions of the charged Higgs to each decay mode can be expressed in terms of the two relevant phenomenological parameters; the charged Higgs boson mass $m_{H^\pm}$ and $\tan\beta$. The left plot of Fig. 2.12 shows the branching fractions as a function of the charged Higgs boson mass in the MSSM M_h^{125} benchmark scenario, for $\tan\beta = 10$. Within this scenario, the decay mode to a top and bottom quark-antiquark pair is dominant for charged Higgs boson masses above ~ 200 GeV, followed by the decay mode $H^+ \to \tau^+\nu_\tau$. The decay mode $H^\pm \to$ tb has sizable branching fractions for a wide range of $\tan\beta$ values, as shown in Fig. 2.12b.

(a) M_h^{125}, $\tan\beta = 10$ (b) M_h^{125}, $\mathrm{H}^{\pm} \to tb$
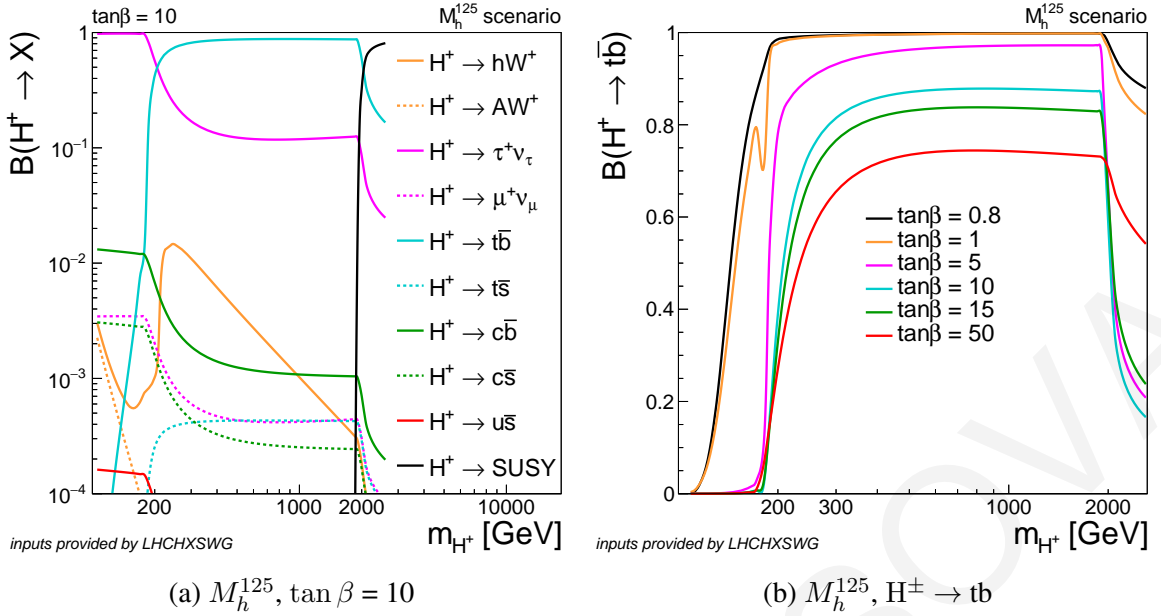
Figure 2.12: *Left:* The branching fractions of the charged Higgs boson decay to different decay modes as a function of the charged Higgs boson mass, in the MSSM M_h^{125} scenario, for $\tan\beta = 10$. *Right:* The branching fraction of the charged Higgs boson decay to a top and bottom quark-antiquark pair as a function of the charged Higgs boson mass, for various values of $\tan\beta$.

Experimental constraints

Experimental constraints on the charged Higgs boson mass are set from direct and indirect searches. Flavor physics observables provide some of the strongest indirect constraints on the charged Higgs boson mass. In the type-II and type-III 2HDMs, a $\tan\beta$-independent lower limit of $m_{H^{\pm}} \gtrsim 600$ GeV is obtained from $B \to X_s\gamma$ data [56, 57]. For low $\tan\beta < 1$, charged Higgs boson masses below 650 GeV are ruled out in all four types of 2HDM, as the Higgs coupling to up-type quarks in this region of phase-space are universal. A combination of the constraints from direct charged Higgs bosons at LEP [58] and neutral Higgs boson searches at the CMS [59, 60, 61, 62] and the ATLAS [63, 64, 65] experiments of the LHC, sets a lower bound on the charged Higgs boson mass of 75 GeV, independent of other model parameters. At the LHC, direct searches for charged Higgs bosons have been performed at 13 TeV in several decay modes and final states providing model-independent results that can be translated *a posteriori* into exclusion limits on specific scenarios. Such interpretations are done on the $(m_A, \tan\beta)$ and $(m_{H^{\pm}}, \tan\beta)$ planes for the MSSM and 2HDM, respectively. The searches are mainly focused on the $H^{\pm} \to \tau\nu$ [66, 67, 68] and on $H^{\pm} \to tb$ with leptonic final states [69, 70]. Results of searches for a light $H^{\pm}$ produced in the decay $t \to H^+b$ and subsequently decaying into $c\bar{s}$ or $c\bar{b}$ are presented in Refs. [71, 72, 73]. Additionally, a first search for light $H^{\pm}$ decaying into a W boson and a $\mathcal{CP}$-odd Higgs boson in final states with leptons is presented in [74]. Searches for charged Higgs boson produced via VBF and decaying into a pair of W and Z bosons can be found in Refs. [75, 76, 77]. This thesis presents a first search for charged Higgs bosons decaying into a top and bottom quark-

antiquark pair in the all-jet final state with the CMS detector, at a center-of-mass energy of $\sqrt{s} = 13$ TeV [15].

3 Experimental Apparatus

CERN, originally known as the *European Organization for Nuclear Research*, was established in 1954 with the purpose of understanding the atomic nuclei. Since then, several scientific achievements and breakthroughs have been made at CERN, including the discovery of $W^{\pm}$ and Z^0 bosons in the UA1 [78, 79] and UA2 [80] experiments, the discovery of direct $\mathcal{CP}$ violation in the NA48 experiment [81], the creation of the World Wide Web (WWW), and the discovery of the Higgs boson by the ATLAS [11] and CMS [12] experiments. Today, CERN is the largest particle physics laboratory in the world. Located near Geneva, CERN is the home of an international scientific collaboration, consisting of more than 13000 scientists from 23 member states.

3.1 CERN accelerator complex

The CERN accelerator complex, illustrated in Fig. 3.1, is a sequence of particle accelerators with increasingly higher energies. Protons are extracted from ionized hydrogen atoms with the use of a strong electric field and are then grouped into bunches and accelerated to an energy of 750 keV with the help of a Radio Frequency Quadrupole (RFQ). Subsequently, the protons enter the Linear Accelerator (Linac2), which accelerates them to 50 MeV and then delivers them to the first circular accelerator, the Proton Synchrotron Booster (PSB), where they obtain the energy of 1.4 GeV. After PSB, the proton beam is injected into the Proton Synchrotron (PS), a synchrotron with a circumference of 628 meters, which accelerates the beam to 26 GeV and injects it to the Super Proton Synchrotron (SPS). The SPS has a circumference of 6.9 km and is the CERN's second-largest accelerator. Both, PS and SPS use room-temperature electromagnets, including a large number of dipoles to bend the beams inside their rings. The SPS provides proton beams with energy of 450 GeV to the last element of the accelerator chain, the Large Hadron Collider (LHC). The protons are separated into two counter-rotating beams in two independent vacuum pipes, and the LHC further accelerates them to their maximal energy of 6.5 TeV, corresponding to 99.9999991% of the speed of light. The two proton beams are brought into head-on collision at four Interaction Points (IPs), where detectors record the collision debris.
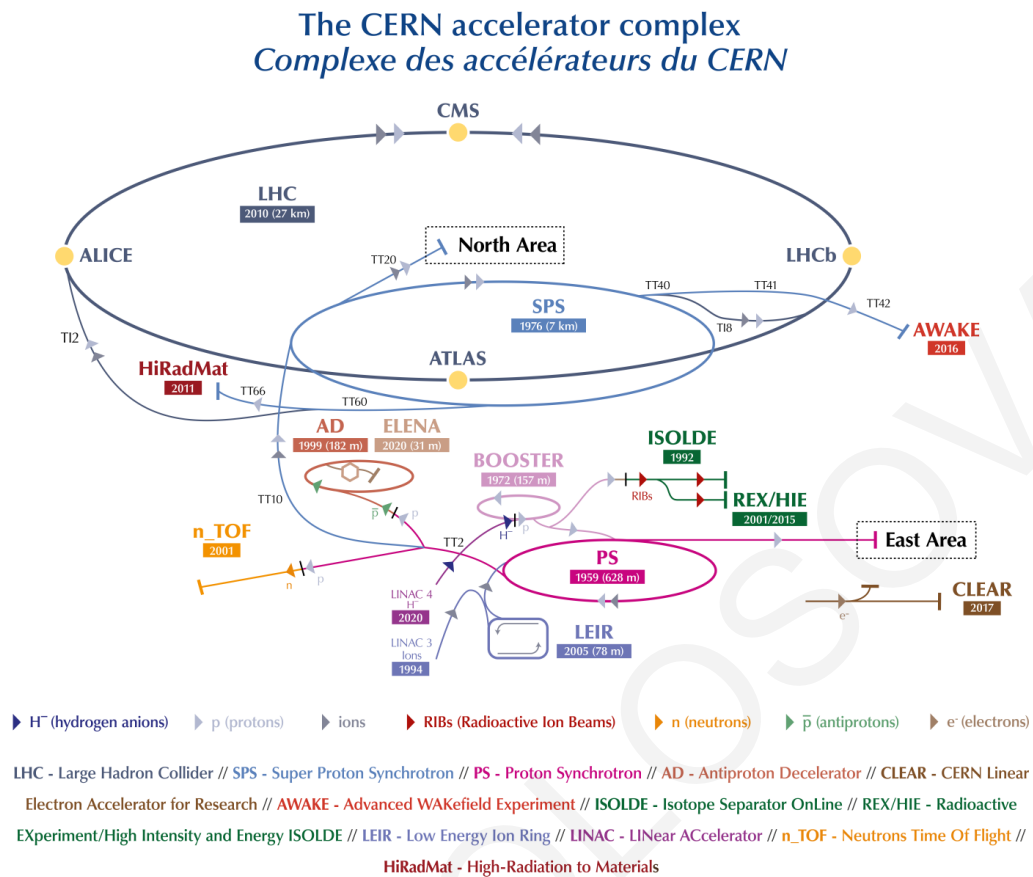
Figure 3.1: The CERN accelerator complex.

3.2 The Large Hadron Collider

Built from 1998 to 2008, the LHC is the most powerful superconducting hadron accelerator and collider to date. The LHC is installed 50 to 175 meters beneath the France-Switzerland borders in a circular and 26.7 km long tunnel, which formerly hosted the Large Electron-Positron (LEP) collider. The accelerator is not a perfect circle but consists of eight 2.45 km-long arcs, joined by eight Long Straight Sections (LSS) with a length of 545 m. Due to the limited space inside the arc tunnels (3.8 m in diameter), the concept of a twin-bore magnet is used, instead of two separated magnet rings. This includes two sets of superconducting coils and beam pipes, accommodated in the same cryostat system and mechanical structure, as shown in Fig. 3.2. A total of 1232 NbTi superconducting dipole magnets are installed inside the arcs and are used to bend the particles. Each dipole is 14.3 m long and weighs 35 tonnes. The magnets operate at the nominal magnetic field of 8.33 T and are maintained at 1.9 K by a closed liquid-helium circuit. Multipoles of higher-order are needed to control the two beams: quadrupole magnets are used for bunch localization and beam focus, sextupole magnets correct the energy dependence of the magnetic fields, and octupoles stabilize the beam. The four IPs are installed in the middle of the four LSS, while the remaining LSS are used for LHC machine utilities; two are used for beam cleaning, one for safe beam dumping
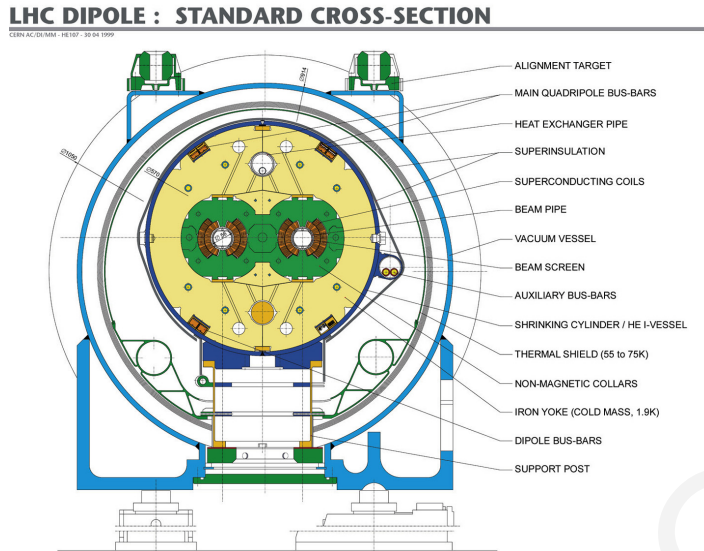
Figure 3.2: Cross section of the LHC dipole superconducting magnet.

at the end of each run, and the last one accommodates the accelerating system of Radio Frequency (RF) cavities.

The two most important and complementary parameters in particle physics experiments are the *center-of-mass energy* of the collisions, denoted by $\sqrt{s}$, and the *instantaneous luminosity*, $\mathcal{L}$. For the LHC, a particle-particle collider, the center-of-mass energy is simply given by the sum of the energies of the two beams. High center-of-mass energies can allow the production of potential new particles with large masses. Operating with the highest possible instantaneous luminosity is crucial for exploring the lower cross-section processes and thus, reducing the statistical uncertainties of such interesting events. The instantaneous luminosity is correlated with the number of useful interactions (events) per second, as follows:

$$\frac{dN_{event}}{dt} = \mathcal{L} \cdot \sigma_{event}, \tag{3.1}$$

where σ_{event} is the cross-section of interest. Luminosity is an indicator of how well an accelerator performs and is measured in $\mathrm{cm}^{-2}\mathrm{s}^{-1}$. Assuming head-on collisions of identical particles and a Gaussian beam distribution, the instantaneous luminosity can be determined by the beam parameters as

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \varepsilon_n \beta^*} F, \tag{3.2}$$

where N_b is the number of particles per bunch, n_b is the number of bunches and f_{rev} is the revolution frequency of the bunches. γ_r is the relativistic gamma factor, ε_n is the normalized transverse beam emittance, which describes how confined the beam is, and β^* is the amplitude function at the collision point, quantifying the focus of the beam. F is the geometric luminosity reduction factor due to the crossing angle at the IP, defined as:

$$F = \frac{1}{\sqrt{1 + \left(\frac{\theta_c \sigma_z}{2\sigma^*}\right)^2}}, \tag{3.3}$$

Table 3.1: Overview of the design LHC machine and beam parameters in proton-proton collisions

Parameter	Symbol	Design value
Center-of-mass energy [TeV]	$\sqrt{s}$	14
Instantaneous luminosity [$10^{34}\ cm^{-2}s^{-1}$]	$\mathcal{L}$	1
Number of bunches per beam	n_b	2808
Number of protons per bunch [10^{11} p/b]	N_b	1.15
Bunch spacing [ns]	Δt	25
normalized emittance [μm rad]	ε_n	3.75
beta function CMS/ATLAS [cm]	β^*	55
Revolution frequency [kHz]	f_{rev}	11.25
Half crossing angle CMS/ATLAS [μrad]	$\theta_c/2$	142.5

where θ_c is the full crossing angle at the IP, σ_z is the RMS bunch length, and σ^* is the transverse RMS beam size at the IP. Integrating the instantaneous luminosity over time (excluding the machine dead time) gives the so-called *integrated luminosity*:

$$L(t) = \int \mathcal{L}(t')\ dt' \tag{3.4}$$

which is used as a measurement of the collected data size. Integrated luminosity is expressed in inverse picobarns (pb^{-1}) or femtobarns (fb^{-1}).

The LHC is designed to collide predominantly proton beams with a center-of-mass energy of 14 TeV and an instantaneous luminosity of $10^{34}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$. In addition to the proton beams, the LHC can also collide heavy ions (Pb) with an energy of 2.8 TeV per nucleon and a luminosity of $10^{27}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$. The LHC has two high luminosity experiments, ATLAS (A Toroidal LHC Apparatus) and CMS (Compact Muon Solenoid), designed to operate to a peak instantaneous luminosity of $10^{-34}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$. The two multi-purpose detectors differ in several technical aspects and magnet-system design, but carry the same scientific goals; the discovery of the Higgs boson, precision electroweak measurements, and the search for BSM phenomena at the TeV scale. This thesis was conducted using proton-proton collision data recorded by the CMS detector. In addition, the LHC has two low luminosity experiments, the LHCb (Large Hadron Collider beauty) and the TOTEM (Total cross section, Elastic scattering and diffraction dissociation Measurement at the LHC). The LHCb experiment specializes in B-Physics and can operate to a peak luminosity of $10^{32}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$. TOTEM's physics program is focused on the precise measurement of proton-proton scattering cross section at small angles and aims at a peak luminosity of $2 \times 10^{29}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$. The TOTEM experimental apparatus, which is not shown in Fig. 3.1, is installed symmetrically across the CMS IP. ALICE (A Large Ion Collider Experiment), installed between ATLAS and CMS, is dedicated to heavy-ion physics and can operate at a peak luminosity of $10^{27}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$ for nominal Pb-Pb ion operation. The nominal parameters of the LHC machine and proton-proton collision beams are listed in Table 3.1.

High instantaneous luminosities lead to multiple proton-proton (pp) collisions taking place in the same bunch crossing, namely *pileup* (PU). Pileup events cause overlapping collision debris in the detector volume and challenge the isolation of the pp interaction of interest, as well as the efficiency and resolution of the particle reconstruction methods. The average pileup can be given by:

$$\langle \text{PU} \rangle = \frac{\mathcal{L}\sigma_{pp}^{in.}}{n_b f_{rev}}, \tag{3.5}$$

where $\mathcal{L}$ is the instantaneous luminosity, $\sigma_{pp}^{in.}$ is the inelastic *pp* cross-section, n_b is the number of bunches, and f_{rev} is the revolution frequency of the beam.

The Run 2 of the LHC started in 2015 and lasted till the end of 2018. During this period, the LHC managed to reach instantaneous luminosities of about 2.1×10^{34} cm^{-2}s^{-1}, well above the nominal design luminosity. A typical Run 2 filling scheme was comprised with 2556 proton bunches per beam, grouped in *trains* of 48 bunches with 25 ns spacing. During the 2017 data-taking period, the LHC suffered frequent beam dumps caused by the formation of electron clouds inside the vacuum chambers of the LHC ring [82]. To mitigate this effect, the LHC used a new filling scheme, called 8*b*4*e*, in which the 48 standard bunch trains were replaced by mini-trains of 8 filled bunches, followed by 4 empty slots. With this filling scheme, the formation of electron clouds was suppressed but allowed a maximum of 1916 filled bunches. The peak luminosity was leveled up to $\sim 1.55 \times 10^{34}$ cm^{-2}s^{-1} such that the average pileup would not exceed 60. The LHC managed to return to the preferred nominal scheme in 2018 and provided a peak luminosity of about 2×10^{34} cm^{-2}s^{-1}, corresponding to an average pileup of 55. The LHC delivered an integrated luminosity of 41.0, 49.8 and 68.0 fb^{-1} in 2016, 2017 and 2018, respectively, and the CMS experiment recorded an integrated luminosity of 35.9, 41.5 and 59.7 fb^{-1} of good quality data. The peak instantaneous luminosity and the average pileup distribution observed by the CMS detector for the data-taking period of 2015 to 2018 are shown in Fig. 3.3 and Fig. 3.4, respectively.
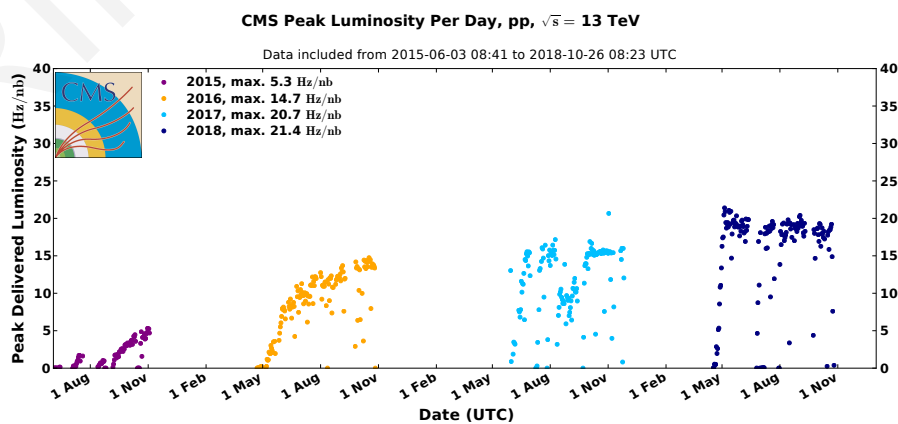
Figure 3.3: Peak instantaneous luminosity during LHC Run 2.
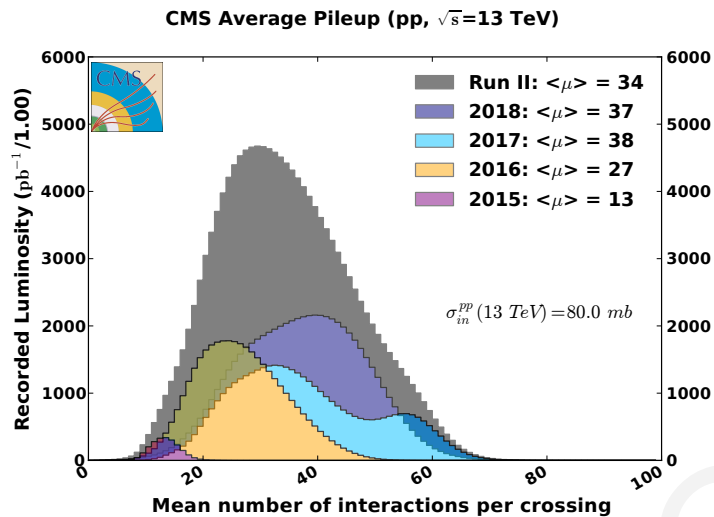
Figure 3.4: Mean number of interactions per bunch crossing for all Run 2 data-taking periods, assuming an inelastic proton-proton cross-section of 80 mb.

3.3 The CMS experiment

Located at the experimental area Point 5 (P5) at the LHC, in the French village of Cessy, the *Compact Muon Solenoid* (CMS) [83] is one of the largest and most powerful detectors ever built. The CMS detector, illustrated in Fig. 3.5, has a cylindrical shape, a total weight of 12500 tons, and its overall dimensions are 21.6 m in length and 14.6 m in diameter. As the name reflects, the CMS detector is relatively *compact* with respect to the ATLAS detector, the dimensions of which are 46 m in length and 25 m in diameter and weighs half of the CMS detector weight. The CMS detector consists of several concentric subsystems, each of them designed to measure the trajectories and energy deposits of different particles. Its key element is the superconducting *solenoid* accommodated inside its volume and operates at a magnetic field of 3.8 T. This high magnetic field bends the trajectories of charged particles and allows for the precise measurement of their momenta and charge. The CMS interaction point is surrounded by a full silicon-based tracking detector, followed by a highly granular crystal Electromagnetic Calorimeter (ECAL) and a hermetic brass-scintillator Hadron Calorimeter (HCAL). All the aforementioned subdetectors are installed inside the solenoid. Outside the solenoid and integrated between the iron yoke, gaseous *muon* detectors provide precise and robust muon measurements. The CMS coordinate system is discussed in section 3.3.1 and a description of all CMS main components is given in section 3.3.2.

3.3.1 Coordinate conventions

The CMS experiment uses a right-handed coordinate system, with the origin being located at the nominal collision point (IP5), as illustrated in Fig. 3.6. The y-axis points vertically upwards towards the surface and is perpendicular to the LHC plane, while the x-axis points radially inwards to the center of the LHC ring. The z-axis is parallel to the beam direction

CMS DETECTOR
Total weight   : 14,000 tonnes
Overall diameter : 15.0 m
Overall length   : 28.7 m
Magnetic field   : 3.8 T

STEEL RETURN YOKE
12,500 tonnes

SILICON TRACKERS
Pixel (100x150 μm) ~16m² ~66M channels
Microstrips (80x180 μm) ~200m² ~9.6M channels

SUPERCONDUCTING SOLENOID
Niobium titanium coil carrying ~18,000A

MUON CHAMBERS
Barrel: 250 Drift Tube, 480 Resistive Plate Chambers
Endcaps: 468 Cathode Strip, 432 Resistive Plate Chambers

PRESHOWER
Silicon strips ~16m² ~137,000 channels

FORWARD CALORIMETER
Steel + Quartz fibres ~2,000 Channels

CRYSTAL
ELECTROMAGNETIC
CALORIMETER (ECAL)
~76,000 scintillating PbWO₄ crystals

HADRON CALORIMETER (HCAL)
Brass + Plastic scintillator ~7,000 channels

Figure 3.5: Schematic view of the CMS detector showing its main subsystems.
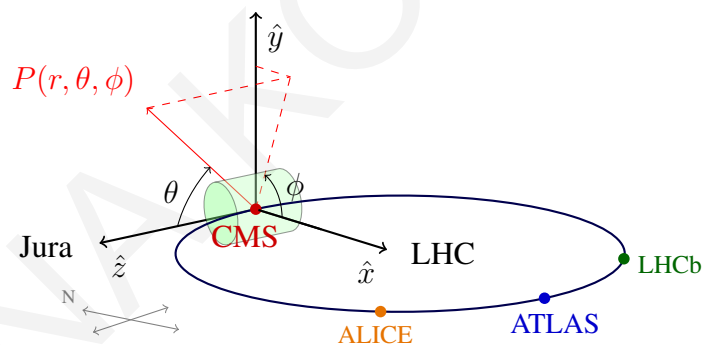
Figure 3.6: Illustration of the coordinate system used in CMS.

and points to the Jura mountains. Given a point P, the *azimuthal* ϕ is the angle between the x-axis and the line from the origin to the projection of P on the xy-plane, namely *transverse plane*. The polar angle θ is measured counterclockwise from the z-axis.

While the momentum of the incoming protons is well known, the momentum fractions of their partons (quarks and gluons) taking part in the collision event are not known a priori. This motivates the use of observables that remain invariant under transformations between the reference frame (detector frame) and the center-of-mass frame. The *transverse momentum* p_{T} is defined as the projection of the momentum in the transverse plane and its magnitude is given by $p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2}$. Similarly the *transverse mass* of a particle m_{T}, is defined as $m_{\mathrm{T}} = \sqrt{m^2 + p_x^2 + p_y^2} = \sqrt{E^2 - p_z^2}$. Both observables remain invariant under Lorentz boosts along the z-axis. The position of a particle in the detector can be specified by the

azimuthal angle ϕ, and by a quantity called *pseudorapidity*, calculated by measuring only the polar angle θ:

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right). \tag{3.6}$$

The pseudorapidity has a value of zero when the detected particle is located anywhere parallel to the y-axis ($\theta = 90°$) and infinity if parallel to the z-axis ($\theta = 0°$). The angular separation of two particles ΔR is also invariant under Lorentz boost along the beam-axis and can be expressed as

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}, \tag{3.7}$$

where $\Delta\eta = \eta_1 - \eta_2$ and $\Delta\phi = \phi_1 - \phi_2$.

3.3.2 CMS subsystems

Superconducting magnet

The CMS magnetic system [84] is the most important component of the CMS detector. It consists of three parts: a $NbTi$ superconducting solenoid, a vacuum tank, and a magnetic flux return yoke. The solenoid was designed to be able to reach a magnetic flux density of 4 T, a magnitude that was wisely and carefully chosen to maximize the physics performance and benefit not only the inner and muon tracking but also the electromagnetic calorimetry. It is installed inside the vacuum cryostat which cools the magnet down to 4.7 K (-268.5° C) using liquid helium. The magnetic flux is returned with a 12000 tonne saturated iron yoke of 1.5 m width and of 21.6 m length. The yoke has two components, the barrel yoke which surrounds the coil, and the endcap yoke, which magnetically closes the edges of the barrel yoke. The barrel yoke consists of five three-layered wheels while the endcap yoke has three disks at each end. The yoke is thick enough to also act as an absorber for the four interleaved muon chambers and to allow safe muon identification and reconstruction.

Tracking system

Like most high-energy particle physics detectors, the CMS detector uses a position-sensitive tracking system to track charged particles as they propagate through the magnetic field. The trajectory of each charged particle is reconstructed as a helix from the available space points, with the particle's transverse momentum being proportional to the bending radius. At high p_T, the relative transverse momentum resolution σ_{p_T}/p_T is dominated by the precision of the track position measurements, while at low p_T the resolution is limited by the multiple Coulomb scattering in the tracking detector.

The CMS tracking system is designed with the goal of having a robust, efficient, and precise reconstruction of the trajectories of charged particles. Located around the interaction point, it includes a fine-granularity *silicon pixel detector* and a *silicon micro-strips detector*, as illustrated in Fig. 3.7. Both subsystems consist of several co-axial barrels and are

completed by endcaps on either side of the barrel, extending the tracker coverage up to a pseudorapidity of $|\eta| < 2.5$. The tracking system is immersed into a magnetic flux density of 3.8 T, provided by the CMS solenoid. In order to minimize the radiation damage in the silicon sensors and front-end electronics, the tracker operates at -20° C.
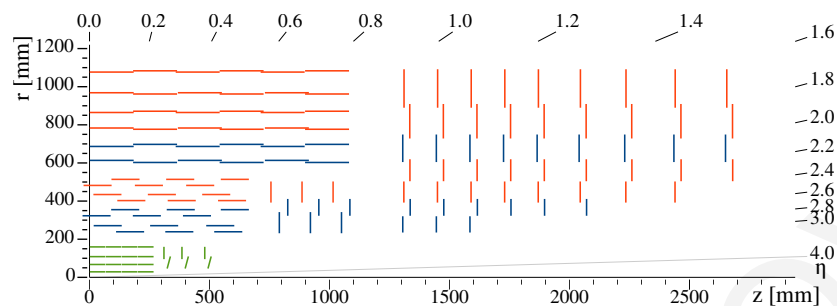
Figure 3.7: Schematic view of one-quarter of the Phase-1 CMS tracking system in the r-z plane. The origin corresponds to the center of the tracking system and to the approximate position of the pp interaction point. The pixel detector is shown with green, while the single-sided and double-sided silicon strip modules are illustrated with red and blue segments, respectively.

The silicon pixel detector used at the beginning of Run 2 data (Phase-0) was designed to operate for a maximal instantaneous luminosity of 1×10^{34} cm^{-2}s^{-1} and a maximum pileup of 25. With a set of three barrel layers (BPIX) and two forward disks (FPIX), the Phase-0 pixel detector could provide a high-precision charged trajectories measurement in the 3-dimensional space, serving a hit position resolution of approximately 150×100 μm^2 per pixel. The three barrel layers were situated at radii between 4.4 cm and 10.2 cm away from the interaction point and the two endcap disks at distances 34.5 and 46.5 cm from the center of the detector.

To cope with higher luminosities and more radiation harsh conditions, a new pixel detector, referred to as the CMS *Phase-1 pixel detector* [85], was installed during the extended year-end technical stop of LHC in 2016/2017. The Phase-1 pixel detector was an evolutionary upgrade to the CMS tracking system for several reasons. The new detector has a higher rate capability due to the faster readout chip design and increased bandwidth electronics that allow the data transmission through the pre-existing optical fibers to the Data Acquisition (DAQ) system. The addition of a new concentric barrel layer and forward disks provides an improved four-hit pattern recognition and more robust tracking. The innermost barrel layer is located closer to the interaction point, only 2.9 cm away, while the rest are at 6.8, 10.9, and 16.0 cm away from the center of the detector. The forward disks on each end are at distances between 29.1 to 51.7 cm away from the interaction point. The Phase-1 pixel detector occupies an area of 1.9 m^2 and is comprised of a total of 1856 segmented silicon sensor modules, each of them having 160×416 pixels connected to readout chips. A total of 124 million readout channels are used. Despite the additional layers, the new detector has improved radiation tolerance. The material budget in the barrel region is significantly reduced with the help of a lower mass, two-phase CO_2 cooling system, and the relocation of

electronic boards outside the tracking acceptance. A comparison between the original and the upgraded Phase-1 pixel detectors is shown in Fig. 3.8.

(a)　　　　　　　　　　　　　　　(b)
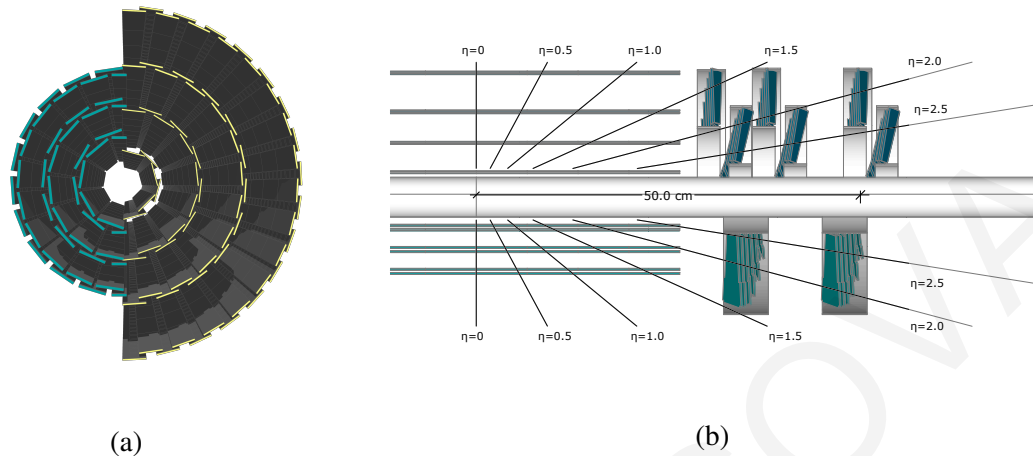
Figure 3.8: Layouts of the upgraded pixel detector compared to the original detector in transverse (*left*) and longitudinal views (*right*). The upgraded (original) barrel layers are illustrated with yellow (cyan) in (a), while both barrel and endcaps regions are shown for the upgraded (upper half) and original (lower half) pixel detector in (b).

The rest of the tracker volume is occupied by the silicon strip tracker. The strip tracker consists of 15148 silicon modules with a total of 9.3 million detector strips, read by 80,000 microelectronic chips. It is composed by four subsystems; the *Tracker Inner Barrel and Disks* (TIB/TID) covering $r < 55$ cm and $|z| < 118$ cm, the *Tracker Outer Barrel* (TOB) covering $r > 55$ cm and $|z| < 118$ cm and the *Tracker EndCaps* (TEC+ and TEC-), covering the region $124 < |z| < 282$ cm. The TIB and TID can provide a position measurement in $r\phi$ with a resolution ranging from 13 to 38 μm, while the TOB and the TEC can achieve a resolution of approximately $18 - 47$ μm.

Calorimeters

Calorimeters [86] are detectors that measure the energy carried by an incoming particle with a *destructive* method (total absorption) and convert it into a detectable signal. Electromagnetic calorimeters are used to directly measure the energy of electrons and photons through their electromagnetic interactions with the detector material and hadronic calorimeters measure hadrons through their strong and electromagnetic interactions. In addition, calorimeters can provide indirect detection of neutrinos through the event missing energy measurement. Calorimeters can be classified into *sampling* or *homogeneous* calorimeters. Sampling calorimeters have alternating layers of an absorber and an active medium. The absorber is a dense material that degrades the energy of the incoming particle, while the active measures the detectable signal. A calorimeter is called homogeneous when it is built of only one material, which can perform both tasks.

When high-energy electrons (photons) with energy above 1 GeV pass through the detector material, they predominantly lose their energy by bremsstrahlung (electron-positron

pair production) [8]. Secondary particles are produced through these mechanisms that in turn produce other particles with progressively degraded energies. This cascade (shower) of particles continues until critical energy is reached. Below the critical energy, ionization and excitation processes overtake the generation of other particles until total absorption occurs. The radiation length X_0, measured in $\mathrm{g\,cm^{-2}}$, is used to characterize the longitudinal evolution of the shower and can be defined as the mean distance in which an electron loses all but $1/e$ of its energy by bremsstrahlung or as the $7/9$ of the mean free path of a photon. The transverse size of an electromagnetic cascade can be described by the Molière radius (R_M), as 90% of the shower's energy is contained within a cylinder of a radius $\sim 1\ R_M$. Figure 3.9 shows a simulated electromagnetic shower inside a medium.
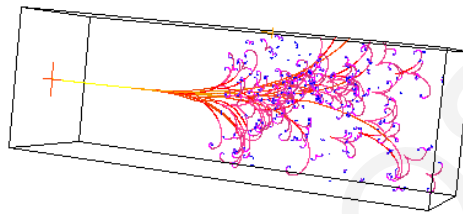
Figure 3.9: Simulation of electromagnetic cascade in matter induced by high energetic particle, created with the *Electro-Magnetic Shower Simulator* tool [87].

When high-energy charged and neutral hadrons interact with the detector material they are subject to inelastic nuclear collisions. These collisions produce secondary hadrons the momenta of which are typically a fair fraction of the primary hadron momentum (at the GeV scale). Some of these secondary hadrons, usually charged pions, neutrons or kaons, can have further nuclear collisions resulting in a hadronic shower. Due to their short lifetime ($\sim$80 attoseconds) neutral pions will decay almost instantly into photon pairs and initiate electromagnetic cascades. Besides the production of energetic secondary hadrons, part of the primary energy is consumed in nuclear processes, such as excitation, nucleon evaporation, and spallation, resulting into particles (neutrons, protons, light ions and gamma rays) with lower energies, typically at the MeV scale.

The energy resolution of a hadronic calorimeter is roughly an order of magnitude worse than for electromagnetic calorimeters as it is limited by both fluctuations in the electromagnetic fraction of the showers and the nuclear binding energy loss. Hadronic showers are distinguished from electromagnetic showers from the shape of their energy deposition in the calorimeters. They are characterized by the nuclear interaction length $\lambda_I \approx 35\ A^{1/3}\ \mathrm{g\,cm^{-2}}$, which is typically much larger than the radiation length X_0. Compared to the electromagnetic cascades, hadron cascades take longer to develop and are broader and less uniform in terms of spatial development. For this reason, hadronic calorimeters are more massive than electromagnetic calorimeters.

The CMS Electromagnetic Calorimeter

The CMS Electromagnetic Calorimeter (ECAL) [88] is a hermetic, fine-granularity, and *homogeneous* calorimeter made of an array of $\sim$76000 lead tungstate ($PbWO_4$) crystals that provide coverage in pseudorapidity up to $|\eta| < 3.0$. The choice of crystals was made based on the LHC design constraints of 25 ns bunch spacing and the harsh radiation environment in which the detector is exposed. Due to the high density ($\rho = 8.28 \ g/cm^3$), short radiation length ($X_0 = 0.89 \ cm$), and small Molière radius ($R_m = 2.2 \ cm$) of $PbWO_4$, features that allow for excellent energy resolution and high radiation tolerance, $PbWO_4$ crystals were the most attractive solution. Another benefit of the $PbWO_4$ crystals is their fast scintillation response as 80% of the scintillation light is emitted within the LHC bunch crossing time.

A layout of the ECAL is presented in Fig. 3.10. The scintillating crystals have a tapered shape and are distributed in the ECAL volume as follows: 61200 crystals are installed in the barrel part of the ECAL (EB), which has a pseudorapidity coverage of up to $|\eta| < 1.479$, while the remaining 14648 crystals are equally distributed in the two endcap regions (EE), which covers the region of $1.479 < |\eta| < 3.0$. The crystals placed in the barrel region have a cross-section of $22 \times 22 \ mm^2$ in the front-end and $26 \times 26 \ mm^2$ in the rear, and a length of 23 cm (or 25.8 X_0). They are contained into submodules, which in turn are grouped into modules, and modules are then assembled in supermodules. Each supermodule contains 1700 crystals. To avoid gaps, the crystal axes are tilted by $3°$ with respect to the vector pointing to the interaction point, in both η and ϕ projections. The endcap crystals have a bigger surface of $28.6 \times 28.6 \ mm^2$ in the front-end and $30 \times 30 \ mm^2$ in the rear, and a length of 22 cm (or $24.7X_0$). They are grouped into mechanical structures of 5×5 crystals, namely supercrystals. Each endcap is divided into two halves, or Dees, where each Dee stores 3662 crystals. The crystals are arranged in a rectangular xy grid, 13 cm away from the interaction point and are tilted from 2 to 8 degrees. The scintillation light is detected by silicon avalanche photodiodes (APDs) in the barrel region and vacuum phototriodes (VPTs) in the two endcaps.

In front of the endcap crystals a sampling calorimeter with finer granularity is placed, the *preshower* (ES). It is designed to identify neutral pions decays ($\pi^0 \to \gamma\gamma$) in the fiducial region $1.653 < |\eta| < 2.6$, which can inadvertently mimic high-energy photons if the resulting low-energy photons are collimated enough. The ES can also provide an improved position resolution for electrons and photons. It is made of two layers: a layer of lead radiators followed by silicon strip sensors. The lead radiators initiate electromagnetic showers from the incoming photons and electrons, while the silicon sensors detect the shower profiles and measure the deposited energy.
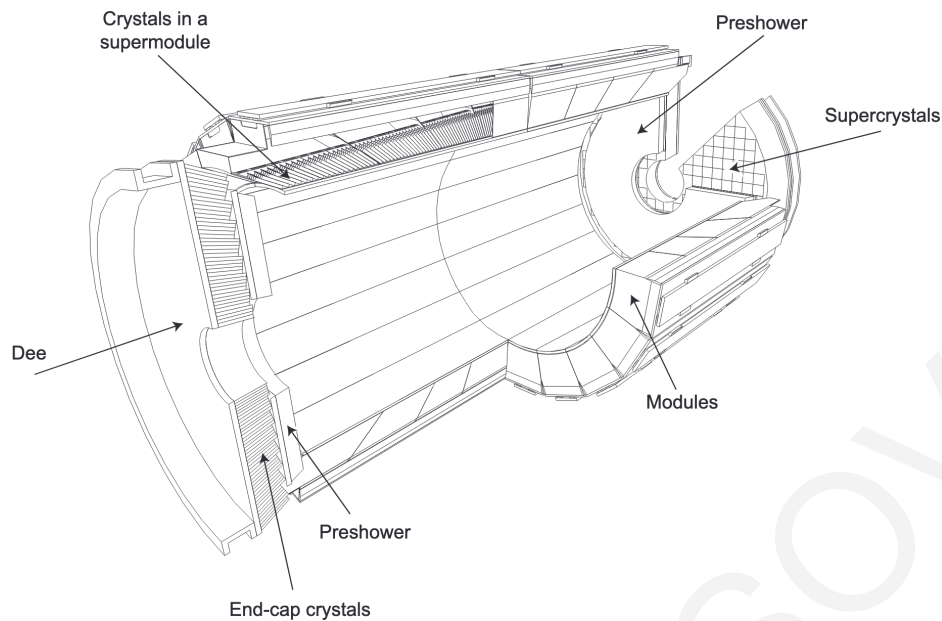
Figure 3.10: Illustration of the CMS electromagnetic calorimeter, showing the barrel super-modules, end-caps, and preshower detectors [88].

The ECAL barrel energy resolution has been measured in 2016 in a beam test setup with no magnetic field or material in front of the calorimeter [89], and found to be:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E[\text{GeV}]}}\right)^2 + \left(\frac{12\%}{E[\text{GeV}]}\right)^2 + (0.30\%)^2, \qquad (3.8)$$

for electrons incident to the center of the crystals. The three contributions correspond to the stochastic, noise, and constant terms. The stochastic term includes contributions from the shower intrinsic fluctuations, and is small due to the homogeneous nature of the calorimeter. The noise term mainly comes from the electronic noise of the readout and pileup, and the constant term, which includes detector response nonuniformities, such as imperfections in the detector geometry or energy leakage from the back of the calorimeter, from radiation damage, etc. The relative electron energy resolution unfolded in bins of pseudorapidity is shown in Fig. 3.11.

Hadron calorimeter

The CMS Hadron Calorimeter (HCAL) [91] is a hermetic, sampling calorimeter consisting of alternating layers of brass absorbers aligned parallel to the beam axis and plastic scintilla-tor tiles with wavelength-shifting fibers. High-energy hadrons interact with the brass nuclei via strong force and they produce secondary particles, mainly charged particles and neutral pions. The presence of neutral pions gives rise to electromagnetic cascades through their decay into photons, while charged particles continue to generate secondary particles with degraded energy. As the shower develops, the particles pass through the scintillator tiles causing them to emit blue-violet light, which is collected by wavelength-shifting fibers with
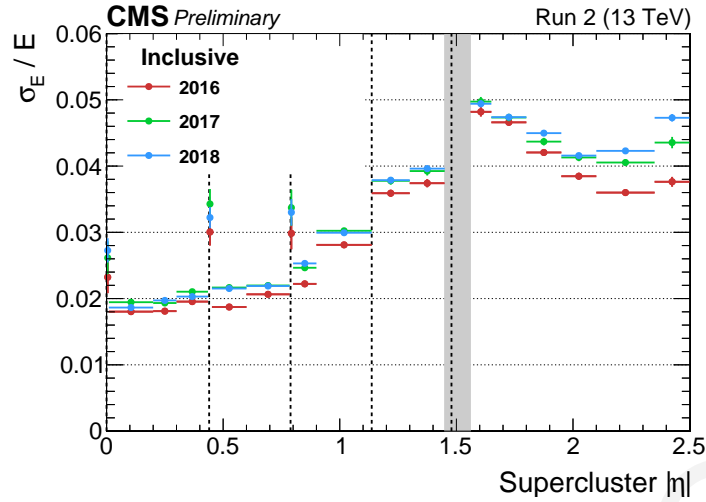
Figure 3.11: Relative electron energy resolution in bins of absolute pseudorapidity for the ECAL barrel and endcap regions. The relative resolution σ_E/E is extracted from an unbinned likelihood fit to $Z \rightarrow \text{ee}$ events [90].

a diameter of less than 1 mm. The CMS HCAL is placed right after the ECAL and inside the magnet coil, the edges of which are at 1.77 m and 2.95 m away from the interaction point, respectively.

As shown in Fig. 3.12, the HCAL consists of four separate hadronic calorimeters. The *barrel calorimeter* (HB) covers the central region of $|\eta| < 1.4$ and its thickness ranges from 6 to 10 λ_I. The HB plastic scintillator is divided in 16 η sectors, providing a segmentation of $(\Delta\eta, \Delta\phi) = (0.0087, 0.0087)$. The *endcap calorimeter* (HE) covers the pseudorapidity region of $1.3 < |\eta| < 3.0$ and has a granularity of $(\Delta\eta, \Delta\phi) = (0.17, 0.17)$. Combined with the ECAL crystals, the total length of the barrel (endcap) calorimeter material corresponds to about 12 (10) interaction lengths λ_I. To provide a sufficient absorption of the hadronic cascades, an additional hadron calorimeter, called *tail catcher* or *outer calorimeter* (HO) is used. The HO is installed just outside the cryostat as the first sensitive layer in each of the five iron yoke rings. At the central ring ($|\eta| = 0$), where the absorber depth is minimal, the HO has two layers of scintillators and a thick piece of iron in between. All remaining rings have just one scintillator layer. The HO provides additional coverage of about $3\lambda_I$ thickness and improves the energy measurement of jets and missing transverse energy. Scintillation light is extracted with wavelength-shifting plastic fibers which illuminate pixels of hybrid photomultiplier tubes (HPD). In the pseudorapidity region $|\eta| > 3.0$ and at 11.2 m away from the interaction point a Cherenkov calorimeter is installed, the *forward calorimeter* (HF). The HF extends the pseudorapidity coverage up to 5.2 and consists of a cylindrical steel absorber interspersed with quartz fibers running longitudinally through the steel and collecting Cherenkov light. The optical light is converted to an electrical signal via a photomultiplier tube (PMT). All four HCAL calorimeters share similar front-end and back-end electronics. The analog HPD and PMT signals are integrated over 25 ns, digitized, and continuously sent to the back-end electronics. The combined energy resolution of ECAL and
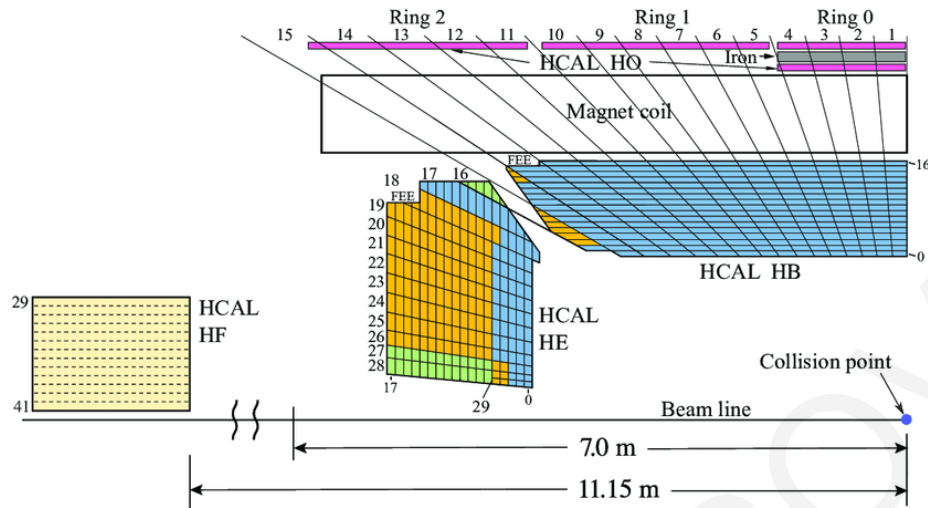
Figure 3.12: Longitudinal view of one-fourth of the CMS HCAL detector in the r-z plane. The shaded areas represent how the scintillator layers are grouped into different readouts.

HCAL barrel system has been determined using single pion beams in Ref. [92] and found to be:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{84.7\%}{\sqrt{E[\text{GeV}]}}\right)^2 + (7.4\%)^2 \tag{3.9}$$

where the first term represents the stochastic fluctuations and the second term corresponds to the noise of the calorimeters. In preparation for Run 2 and Run 3 of the LHC, the HCAL underwent a series of upgrades [93] at the end of 2017. The upgrades include the replacement of all photodetectors in the HB, HE, and HF, new improved front-end electronics with high precision timing readout and better radiation tolerance, and new increased data bandwidth back-end electronics.

The muon detector

A robust muon detector [94] is situated outside the solenoid, at the very edge of the CMS detector as muons can penetrate several meters of iron without interacting. It covers the pseudorapidity range of $|\eta| < 2.4$ and is composed of barrel and endcap gaseous detectors alternating with layers of the steel flux-return yoke. This allows the detection of a traversing muon at multiple points along its track. Figure 3.13 shows a layout of the CMS muon detector and its main components. The muon detector includes a total of 1400 gas ionization chambers of three types: 250 *Drift Tubes* (DTs) chambers, 540 *Cathode Strip Chambers* (CSCs), and 610 *Resistive Plate Chambers* (RPCs) [95].

The DTs are installed in the barrel region ($|\eta| < 1.2$) where the particle rate is small and the magnetic field is uniform and mostly contained in the return yoke. They are distributed in four concentric cylindrical stations (MB1-MB4), interspersed with the layers of the return

yoke. The DTs are segmented into drift cells of a transverse size of $42 \times 13\ mm^2$ and a length of 2.4 m, filled with a gas admixture of Ar (85% vol.) and CO_2 (15% vol.). A gold-plated stainless-steel anode wire of a $50\ \mu m$ diameter is placed at the center of each cell. When a muon crosses a cell, ionized molecules drift to the cathode while electrons reach the anode wire with the help of a homogeneous electric field. The crossing position of the muon can be computed through the drift velocity. The drift chambers can provide a spatial resolution of 78-120 μm in the $r - \phi$ plane and of 140-390 μm in the r-z plane.
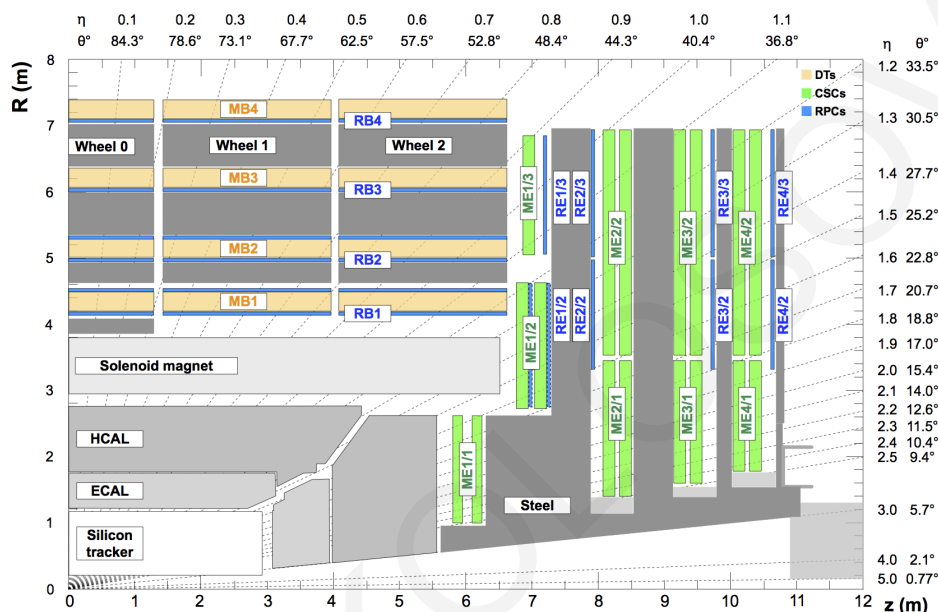
Figure 3.13: Longitudinal view of one-fourth of the CMS muon detector in the r-z plane. The gray shaded ares represent the magnet yoke while the muon detectors are highlighted.

The CSCs are installed in the endcap region ($0.9 < |\eta| < 2.4$), where the muon rates and background levels are relatively higher, and the magnetic field is large and non-uniform. They are characterized by their fast response time (resulting from a short drift path), fine segmentation, and their ability to tolerate the non-uniformity of the magnetic field. Each endcap has four stations of CSCs, placed perpendicular to the beam. The CSCs have alternating layers of anode wires and cathode strips that use a mixture of Ar (40%), CO_2 (50%) and CF_4 (10%). Their spatial resolution is better than the DTs and varies from 40 to 152 μm. Both DT and CSC systems can provide information on the collision bunch crossing that generated the passing muon, reject background as a result of timing discrimination, and trigger on the muons p_T with good efficiency.

In addition to the two tracking detectors, the muon system has a dedicated, fast, and independent triggering detector, the RPCs. The RPCs are located in both barrel and endcap regions, covering the pseudorapidity region of $|\eta| < 1.9$. As DTs and CSCs, they are arranged in four stations (RB1-RB4) in the barrel and three stations (RE1-RE3) in the endcaps. The RPCs can provide excellent time resolution (of less than 3 ns) to reinforce the measurement of the correct beam crossing time with a somewhat poorer spatial resolution (0.8 to 1.2 cm). The RPCs are double-gap chambers operating in avalanche mode and filled

with a gas mixture of 95.2% Freon ($C_2H_2F_4$), 4.5% isobutane (i-C_4H_{10}) and 0.3% sulphur hexafluoride (SF_6).

At the end of 2017, an additional set of muon detectors GE1/1 [96] was installed in the first endcap muon station ($1.6 < |\eta| < 2.2$). These detectors use a gas electron multiplier (GEM) technology and their thin profile and high particle rate capabilities can help maintain or even improve the forward muon triggering and reconstruction.

3.4   Online event selection

With a beam crossing frequency of 40 MHz, the LHC delivers about a billion proton-proton collisions per second. Despite the tremendous progress on data processing capabilities, the CMS detector cannot afford to process and store this colossal data volume. Nevertheless, not all events are of physics interest. Interesting processes, such as the production of the SM Higgs boson, have cross-sections several orders of magnitude below the inclusive total inelastic pp cross-section ($\sim$68 mb [97]). Thus, it is crucial to filter the interesting events while keeping the output rate and CPU time under control. A drastic event rate reduction of a factor of 10^6 is performed by a real-time *trigger system*, which consists of two steps, called Level-1 (L1) Trigger and High Level Trigger (HLT). The L1 trigger is implemented in hardware while the HLT in software.

3.4.1   Level-1 Trigger

The L1 trigger is based on the MicroTCA (μTCA) [98] technology and has a latency of $3.8\ \mu s$. It is divided into a muon trigger and a calorimeter trigger, as illustrated in Fig. 3.14. They utilize information from the muon detectors and the calorimeters, called Trigger Primitives (TPs), and reconstruct physics objects, called *L1 candidates*. Due to timing constraints, information from the tracker cannot be used at this level. The muon trigger is responsible for the muon (L1 μ) reconstruction, while the calorimeter trigger reconstructs jets (L1 jet), hadronically decaying τ-leptons (L1 τ_h), electrons or photons (L1 e/γ) or energy sums (L1 E_T^{miss} and L1 H_T). The reconstructed objects are sent to the GT, which performs an event accept or reject decision, based on a collection of algorithms (trigger menu) targeting specific physics signatures. The processing of all logical operations is done with the powerful Xilinx Virtex-7 Field Programmable Gate Arrays (FPGAs), while the transfer of data between modules is achieved with high-speed optical links (10 Gb/s). The L1 output rate is limited to 100 kHz.

Muon trigger

The muon trigger system takes as inputs the trigger primitives from the CSC, DT, and RPC subdetectors. These TPs include coordinates, timing, and muon track quality information.
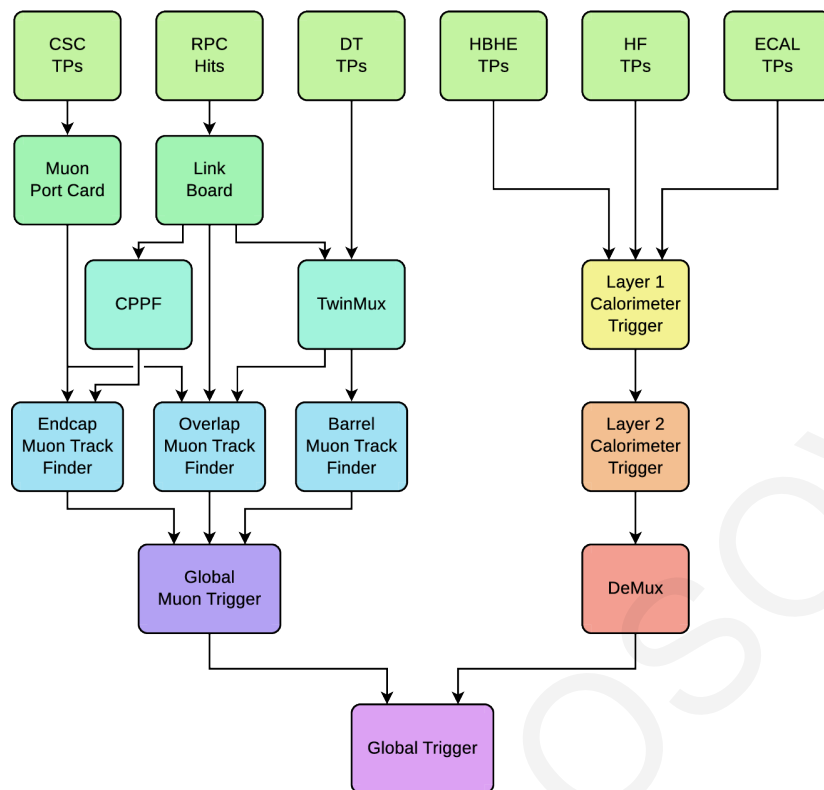
Figure 3.14: Overview of the CMS L1 Trigger system.

Using the combined information of trigger primitives, muon track finders (MTFs) are used to reconstruct muon tracks in three pseudorapidity regions. The Barrel Muon Track Finder (BMTF) receives data from the DT and RPC detectors in the barrel. An adaptive layer, called *TwinMux* [99], merges the DT trigger primitives and RPC hits[1] from the same muon station into *super-primitives* and assigns a quality to each of them. This combination improves the BMTF efficiency and performance, as it exploits the DT spacial resolution and the precise timing from the RPC subsystems. The BMTF reconstructs muon tracks and calculates their properties (p_T, ϕ and η) in the pseudorapidity region $|\eta| < 0.83$. Muon track candidates are extrapolated to neighbor stations and checked whether they are inside an acceptance window set by Look-Up Tables (LUTs) that utilize the internal bending angle ϕ_b, and the quality of the super-primitives.

The Endcap Muon Track Finder (EMTF) receives information from the CSC and RPC detectors in the endcap and covers the region $|\eta| > 1.24$. An intermediate layer, the concentrator and preprocessor fanout (CPPF), clusters the RPC hits into TPs and transfers them to EMTF. The EMTF reconstructs forward muon tracks by using at most one TP, either from CSC or RPC, per station via pattern recognition. A Boosted Decision Tree (BDT) regression technique [100] is used to infer the track p_T, while taking into account the inhomogeneous magnetic field outside the solenoid, muon scattering and showering effects. The prediction scheme is stored into a LUT loaded in a 1.2 GB memory module.

The Overlap Muon Track Finder (OMTF) receives trigger primitives from the DT and

---

[1]signals in the detector sensitive layers

CSC systems and direct hits from the RPC subdetector. It can reconstruct muon tracks in the pseudorapidity region $0.83 < |\eta| < 1.4$ using a pattern recognition algorithm based on a naive Bayes classifier. The algorithm can identify the muon track and measure its p_T by calculating the probabilities of matching the detector hits to transverse momentum hypotheses while taking into account missing hits, fake hits (noise), or multi-muon events.

The Global Muon Trigger (μGMT) can receive up to 108 muon candidates (up to 36 from each MTF), sort them, identify, and resolve duplicated entries that could significantly increase the output rate. Moreover, it corrects the spatial coordinates of each muon candidate by extrapolating tracks from the muon stations to the interaction regions. Subsequently, the μGMT sends to the GT a maximum of 8 muons, selected based on their p_T and track quality. Muons with good p_T resolution pass the *tight* quality criteria and are used in single-muon trigger algorithms. Due to the strong magnetic field in the barrel region, and therefore adequate p_T resolution, all muons reconstructed by the BMTF satisfy the *tight* quality criteria. For EMTF and OMTF, tracks must have TPs in at least three layers with at least one in the innermost layer for EMTF to pass the *tight* criteria. The *medium* and *loose* quality tracks are used in EMTF and OMTF to increase the trigger efficiency for multi-muon events.

Calorimeter trigger

The calorimeter trigger consists of two consecutive layers: *Layer-1* and *Layer-2*. Layer-1 takes as input the local energy deposits from the *calorimeter trigger towers* (TTs). A trigger tower is defined as the combination of a group of crystals in the ECAL together with the HCAL tower located behind them. In the barrel region, the TTs have a $\Delta\eta \times \Delta\phi$ size of 0.087×0.087 radians, while in the endcap regions where the crystals have a more complicated arrangement, the TTs have a $\Delta\eta \times \Delta\phi$ size of up to 0.17×0.17 radians. These electromagnetic and hadronic energy deposits are sent to the 18 processing nodes of the Layer-1, where they are further calibrated with the help of LUTs. The LUTs include the calibration factors binned in pseudorapidity (η) and transverse energy (E_T). The choice of the binning takes into account the hardware limitation and the dependency of the resolution in E_T. These calibration factors are needed to compensate for several effects, including energy losses in the tracker material in front of the calorimeters.

The calorimeter trigger utilizes an innovative architecture, called *Time Multiplexing* [101], illustrated in Fig. 3.15. This architecture enables the trigger primitive data from the entire calorimeter volume, for a specific bunch crossing, to be processed in a single processing node of Layer-2 with full granularity. The output of Layer-1 is streamed with a 9-fold time multiplexing and sent to Layer-2 processor cards through $10\,\mathrm{Gb/s}$ optical links. Layer-2 utilizes the Layer-1 calibrated trigger primitives to reconstruct the physics objects related to the calorimeters (electrons or photons, τ-leptons, jets, and energy sums). Finally, a demultiplexing card (DeMux) reorders, reserializes, and forwards the L1 candidates to the GT.
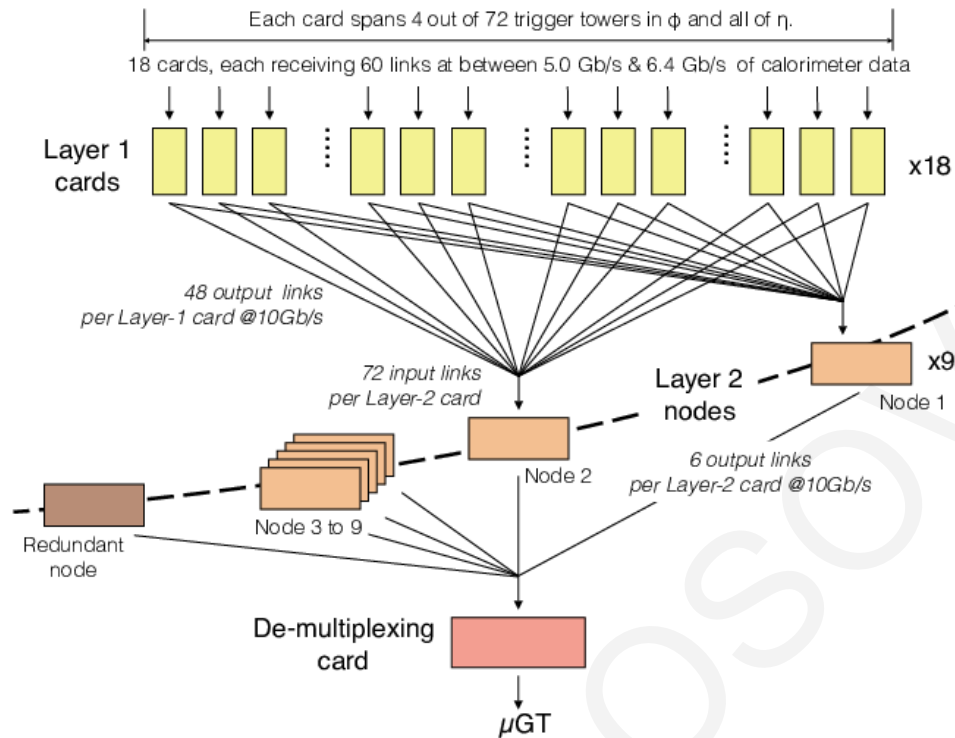
Figure 3.15: The architecture of the calorimeter trigger system used in Run 2.

The L1 e/γ trigger algorithm

Since no tracking information is provided at the L1, electrons and photons are indistinguishable for the calorimeter trigger. The L1 e/γ object is reconstructed using energy deposits from ECAL and HCAL via dynamic clustering. A local energy maximum (seed tower) is clustered with its first and second neighbors in a way such that pileup energy deposits are reduced, and energy losses from bremsstrahlung are recovered. L1 e/γ clusters are discriminated from jet clusters through their shape as the former are more compact, through a fine grain veto which rejects candidates with non-electromagnetic shower profile and through ECAL/HCAL energy deposits (H/E). To keep the e/γ rates low, isolation criteria are applied. A L1 e/γ candidate is considered isolated if the energy deposits in a defined isolation region are less than a fixed threshold. The isolation thresholds depend on the candidate pseudorapidity, transverse energy, and a pileup estimator[2]. To ensure a uniform energy response throughout the detector and an improved energy resolution, dedicated calibrations are applied on the L1 e/γ candidates. These calibrations are done according to the candidate's energy, pseudorapidity, and shape.

---

[2]The number of trigger towers (n_{TT}) with non-zero energy deposits in the central region is used as a pileup estimator.
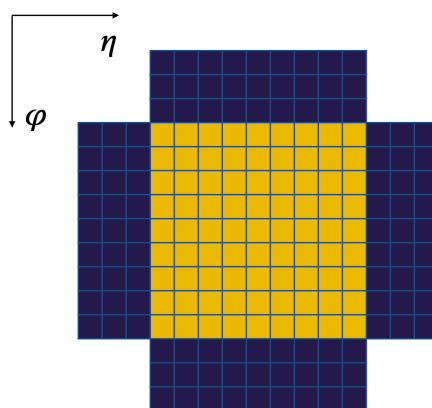
Figure 3.16: The *chunky donut* algorithm sums the energy deposits in the four 3×9 TTs (shown with blue) located at the edges of the 9×9 window (shown with yellow) and subtracts it from the total L1 jet candidate energy.

The L1 τ_h trigger algorithm

Hadronically decaying τ-leptons (τ_h) are reconstructed at L1 through dynamic clustering of their decay products: one, two, or three charged or neutral pions. A similar algorithm to the e/γ clustering method is used to reconstruct these single clusters around local maxima, which are subsequently merged into a single τ_h candidate under proximity conditions. To reject fake τ_h candidates, mainly coming from QCD-induced hadronic jets, τ_h candidates are required to be isolated around the cluster seed, to which the transverse energy of the L1 τ_h candidate is subtracted. L1 τ_h candidates are calibrated based on their E_T, pseudorapidity, the presence of a merged cluster, an estimate of the $\mathrm{H/E}$ deposits.

The L1 jet and energy sum trigger algorithms

The L1 jets are reconstructed with an algorithm based on a square-jet approach. A jet candidate has energy deposits around a 9×9 TT sliding window centered around a local maximum with a minimum threshold of $E_\mathrm{T} > 4$ GeV. In the barrel region, this sliding window has a clustering radius of 0.4, the same as the offline jet clustering size of the anti-k_t method. To avoid double-counting of jets, an inequality mask is applied on energy deposits along the diagonal of this window. For each event, the transverse energy coming from pileup is computed using four 3×9 strips on each edge of the sliding window and subsequently subtracted from each jet. An illustration of the TTs used for L1 jet candidate clustering and pileup subtraction is shown in Fig. 3.16. The L1 jets are calibrated as a function of their energy and pseudorapidity. The reconstructed L1 jets are used to estimate the event *energy sums*. The L1 $E_\mathrm{T}^\mathrm{miss}$ is defined as the magnitude of the negative vector sum of the transverse energy of all TTs. The total scalar transverse energy of all L1 jets energies, is denoted as H_T.

Algorithm name	Description
L1_SingleLooseIsoEG28er2p5	Single loosely isolated e/γ with $E_\mathrm{T} > 28$ GeV and $|\eta| < 2.5$
L1_DoubleIsoTau32er2p1	Double isolated τ with $E_\mathrm{T} > 32$ GeV and $|\eta| < 2.1$
L1_SingleMu22	Single muon with $p_\mathrm{T} > 22$ GeV
L1_DoubleEG_25_12_er2p5	Double e/γ with $E_\mathrm{T} > 25$ GeV, 12 GeV and $|\eta| < 2.5$
L1_DoubleMu_15_7	Double muon with $p_\mathrm{T} > 15$ GeV, 7 GeV
L1_ETMHF100	$E_\mathrm{T}^\mathrm{miss} > 100$ GeV
L1_SingleJet180	Single jet with $E_\mathrm{T} > 180$ GeV
L1_DoubleJet150er2p5	Double jet with $E_\mathrm{T} > 150$ GeV and $|\eta| < 2.5$

Table 3.2: Benchmark L1 seeds used during Run 2 data-taking period.

Global Trigger

All reconstructed L1 candidates are sent from the calorimeter Layer-2 and the GMT to the GT, where they are synchronized to each other and to the LHC orbit clock. The GT combines the information from both trigger systems and performs an accept or discard decision based on a list of algorithms (or *seeds*), collectively called as *Level-1 menu*. If any of the L1 seeds are satisfied, the event is accepted and the GT sends a *Level-1 Accept* (L1A) signal to the DAQ system for subsequent evaluation by the HLT. Each L1 seed encodes a set of requirements on one object (single-object seed) or more objects of the same (multi-object seed) or different (cross-object seed) type. These criteria are usually minimum thresholds on the energy sum or on the L1 candidate's transverse momentum, and requirements on its pseudorapidity and/or azimuthal angle. The calculation of invariant masses, differences in pseudorapidity, and azimuthal angle between two objects are also now possible, due to the upgrade of the GT, which used to operate on a single board (nominal design). In 2016, the GT operated on three boards and extended to six boards in 2017.

A typical L1 menu includes a broad range of sophisticated seeds targeting different areas of physics interest. These include benchmark algorithms with one or more objects of the same type that cover about 80% of the available L1 rate, cross-object seeds, and other algorithms for detector calibration, monitoring, and trigger efficiency measurements. Table 3.2 lists some of the benchmark seeds used during Run 2, the rates of which are shown as a function of the pileup in Fig. 3.17. When an algorithm exceeds its allocated rate, a *prescale* is applied; meaning that a seed with prescale of N only accepts one in N events satisfying the selection criteria. Since the L1 rate increases with increased LHC beam intensity, multiple prescales values are set for ascending instantaneous luminosity scenarios and are applied in order to keep the output rate below 100 kHz while maximizing the signal efficiency.

3.4.2 High Level Trigger

The High Level Trigger (HLT) reduces the trigger rate down to $\sim$1 kHz with a latency of a few hundred milliseconds. It has access to the full precision of the data from the detector and runs a streamlined version of the offline reconstruction in a farm of commercial processors.
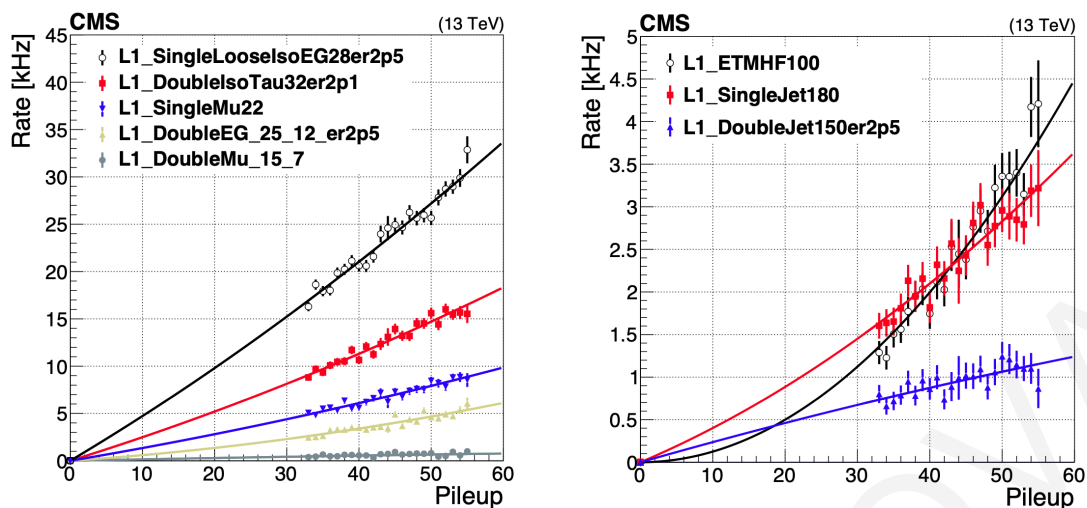
Figure 3.17: Rate as a function of pileup for several benchmark L1 seeds with leptons (left) and hadrons (right), used during Run 2.

The HLT paths are designed as a sequence of object producers and filtering modules, starting from the least CPU time-consuming algorithms to more complex ones that demand more CPU time. Each HLT path is seeded by a single L1 algorithm or by the logical OR of several L1 seeds. As a first step, information from the calorimeters and muon chambers is used for *low-level* reconstruction (calo jets, standalone μ, ECAL energy deposits) and fast filtering of the events. Tracking reconstruction (tracks, vertices) follows with a simplified version of the offline tracking, followed by filtering modules. The next step in the HLT chain is the production of particle candidates using the Particle Flow (PF) algorithm [102], described in chapter 4, and the further filtering of events. More advanced observables and topologies are reconstructed from these individual particles and subsequently filtered. To keep the CPU time under control, if the requirements in a filtering module are not satisfied, the remaining modules of the HLT path are skipped and the event is discarded. Additionally, common object producers are shared between HLT paths to optimize the event processing time. The overall processing time of a typical HLT menu used in 2016 data-taking period, with an average instantaneous luminosity of $1.3 \times 10^{34} \; \mathrm{cm}^{-2}\mathrm{s}^{-1}$ and an average pileup of 42.5, is shown in Fig. 3.18a. Figure 3.18b shows the HLT processing time as a function of the instantaneous luminosity.

Events selected by the HLT are sent to the storage manager where they are temporarily stored at LHC Point 5. Subsequently, the events are transferred to the CERN Tier-0 computing center for permanent storage and offline reconstruction. Besides the HLT timing constraints, the DAQ system has finite bandwidth; the temporary raw data storage at Point 5 is limited and the bandwidth between Point 5 and the CERN Tier-0 has a maximum value of 5 Gb/s. Moreover, the prompt reconstruction system must reconstruct the recorded events on the LHC computing grid within 48 hours and without a significant backlog. As a consequence, the HLT system can only record events at a maximum rate of 1 kHz. To achieve this huge rate reduction (100 to 1 kHz) the triggers must impose stringent criteria that become
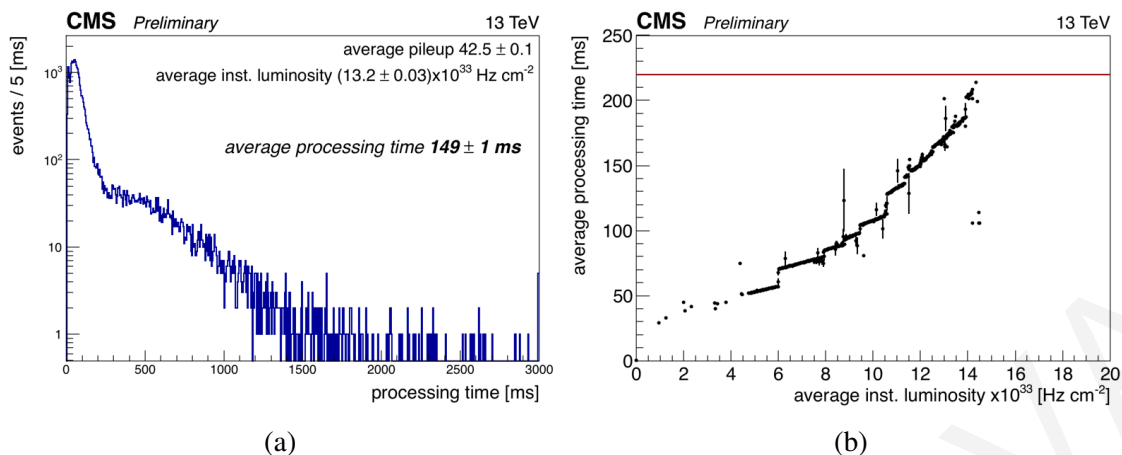
Figure 3.18: Left: Overall HLT processing time distribution for a menu used in the 2016 data-taking period. The peak observed at 0 ms corresponds to the events rejected by the L1 trigger system, while the peak around 50 ms corresponds to the events accepted or rejected based on the low-level reconstruction. The long tail represents the events accepted or rejected after the PF processing and filtering. Right: Average processing time as a function of the average instantaneous luminosity during LHC fill 5393.

more aggressive with higher instantaneous luminosities.

To collect additional events of physics interest and extend the recorded phase space, the CMS experiment has two strategies: *Data Scouting* and *Data Parking*. Data Scouting refers to the method where events are reconstructed in HLT using the PF algorithm or other reconstruction algorithms, and filtered with loose criteria and low thresholds. For each event satisfying these criteria, only the HLT-reconstructed physics objects are saved on disk (few kB) instead of the full raw data ($\sim$1 MB). Since these events do not burden the DAQ system or need prompt reconstruction, several kHz of scouting events can be recorded. Data Scouting comes with few drawbacks; the physics analysis is done with HLT objects which suffer more from detector noise and miscalibration than the offline reconstructed objects. This would mean less confidence in results in the case of discovering new physics when analyzing scouting events. For this reason, Data Parking events are used to validate or disconfirm the results. Data Parking is the method where events are recorded online at HLT and the full raw data volume is immediately moved on tape, skipping the offline prompt reconstruction. Events are reconstructed, if needed, once there are sufficient free computing resources.

A typical Run 2 HLT menu includes $\mathcal{O}(600)$ trigger paths targeting a broad variety of physics signatures, paths used for detector monitoring and calibration and trigger efficiency measurements. Benchmark HLT paths require single-, double-, or multi-objects, with or without isolation requirements. Isolated single muon and electron paths with typical p_{T} thresholds above 24 and 32 GeV (in 2018) are the most rate consuming paths, followed by paths requiring at least two photons or two τ-leptons. Hadronic triggers are also present, some of which utilize substructure or heavy-flavor identification techniques that could not be possible at the L1 due to the lack of tracking information. During 2016 and 2017, jets arising from the hadronization of bottom quarks (b jets) could be identified online at HLT with the
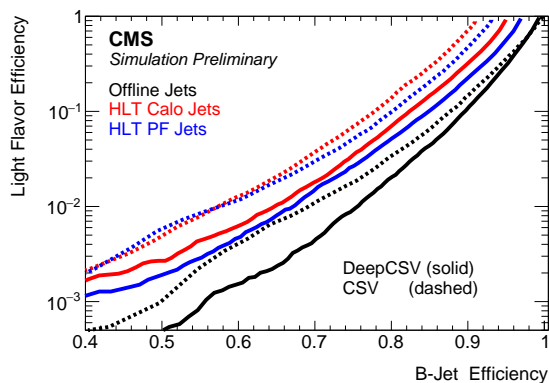
Figure 3.19: Comparison of the CSVv2 (dashed lines) and DeepCSV (solid lines) algorithms online (red and blue) and offline (black) performance using $t\bar{t}$ simulation sample with the 2017 detector conditions. The curves demonstrate the probability for a light-flavor (u, d, s) jet to be misidentified as a b jet as a function of the efficiency to correctly identify a b jet [105].
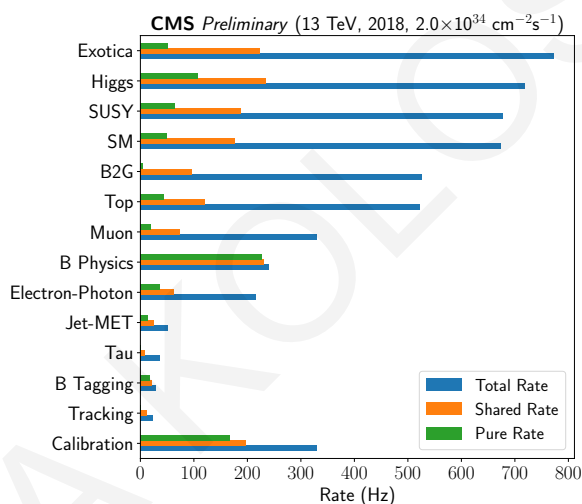
Figure 3.20: HLT rates consumed by each CMS physics group during the 2018 data-taking period [106].

Combined Secondary Vertex (CSVv2) [103] algorithm, both at calo- and PF-levels. In 2018, the CSVv2 algorithm was replaced by a new algorithm, the DeepCSV [104]. A comparison between the online and offline performance of the two algorithms, obtained using simulated $t\bar{t}$ events with 2017 detector conditions, is shown in Fig. 3.19.

As in the L1 menu, the HLT also includes cross triggers, mainly used for searches of specific physics scenarios with difficult signatures. Cross triggers usually involve objects of different types (e.g. leptons with jets or $E_{\mathrm{T}}^{\mathrm{miss}}$) with significantly lower thresholds and/or isolation requirements. The total, shared, and pure HLT rate consumption by each of the CMS physics groups are shown in Fig. 3.20.

4 Event reconstruction

The event reconstruction of a proton-proton collision is predominantly based on the Particle Flow (PF) algorithm [102]. To identify and reconstruct all individual stable particles present in an event; electrons, photons, muons, charged or neutral hadrons, the PF algorithm combines information from all CMS subsystems. Each of these stable particles leaves a distinctive signal as they propagate through the detector, as shown in Fig. 4.1. Electrons appear
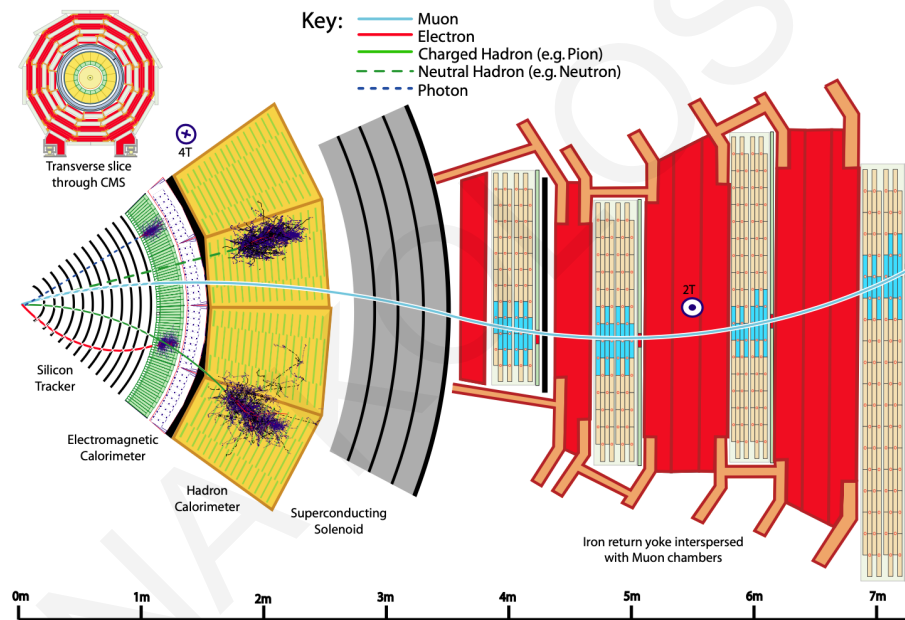
Figure 4.1: The cross-section of the CMS detector and the distinct signatures of each of the physics objects.

as narrow and localized energy deposits in the ECAL and are associated with a trace in the silicon tracker. Photons can also be identified as ECAL energy deposits but are not linked with any extrapolated inner track. Muons leave their signature in the tracker, as well as in the muon chambers. Charged and neutral hadrons can cross the ECAL and lose a small part of their energy by initiating a hadronic shower, but most of their energy is absorbed in the HCAL. In addition, charged hadrons produce hits in the tracker layers. All aforementioned reconstructed particles are further used to build more complex objects, such as jets, hadronically decaying τ-leptons, and to calculate the missing transverse energy ($E_\mathrm{T}^\mathrm{miss}$) of the event, which gives an indirect detection of neutrinos.

4.1 PF-elements reconstruction

Event reconstruction starts with the reconstruction of PF-elements: trajectories of charged particles, vertices, and calorimeter clusters. When correlated, these basic elements are used to identify the final-state particles, which are known as *PF candidates*.

4.1.1 Tracks reconstruction

The reconstruction of charged particles tracks is the most challenging and most CPU-consuming part of the data reconstruction. It is carried out by multiple iterations of the Combinatorial Track Finder (CTF) algorithm, an adaptation of the Kalman filtering method [107]. The CTF method can be decomposed into four stages: seed generation, pattern recognition, track fitting, and quality selection. A trajectory seed is generated when a few three-dimensional hits are compatible with a trajectory. Given a coarse estimate of the track parameters[1] and their uncertainties from the trajectory seed, the filter iteratively builds track candidates from hits on successive detector layers and updates the track parameters. The track candidates are then fitted and marked with quality flags. Hits associated with track candidates are removed after each iteration to reduce the complexity of the algorithm. Moreover, effects from multiple Coulomb scattering and energy loss are taken into account. For non-isolated particles with $1 < p_{\mathrm{T}} < 10$ GeV and $|\eta| < 1.4$, the track resolutions are around 1.5% in p_{T} and 25-90 (45-150) μm in the transverse (longitudinal) impact parameter [108]. In 2017, the track reconstruction method was updated to exploit the upgraded pixel detector features. The resolution of the transverse impact parameter improved, as expected, to 20-60 μm when restricted to $|\eta| < 1.4$ and to 20-75 μm when considering the full tracker coverage [109].

4.1.2 Primary vertex reconstruction

The primary-vertex (PV) reconstruction identifies the location of all collision points in proton-proton interactions. The reconstruction starts with the selection of tracks based on their quality criteria and compatibility with the center of the beam spot. The selected tracks are then clustered according to their z-coordinate at the point of closest approach (PCA) to the center of the beam spot. The clustering is performed with the Deterministic Annealing (DA) method [110] which offers a crucial feature: the ability to find vertices in a high pile-up environment. Once candidate vertices are identified with the DA method, an adaptive vertex fitter (AVF) [111] is applied on candidates with at least two tracks to calculate the vertex parameters. These include the three spatial coordinates of the vertex, its covariant matrix, and indicators of the fit success, such as the number of degrees of freedom and weights of individual tracks. The PV resolution depends on the number of tracks used in the fit and their

---

[1]In the quasi-uniform magnetic field of the tracker, a charged particle travels in a helical path. Its trajectory can be described by five parameters: the curvature (R), the azimuthal angle (ϕ_0), the dip angle ($\lambda = \cot\theta$), the maximum transverse (d_0) and longitudinal (z_0) impact parameters relative to the center of the beam spot.

transverse momenta. For PVs with at least 50 tracks, the achieved vertex position resolution is around $10 - 12 \ \mu m$ in each of the three spatial dimensions [108].

4.1.3   Secondary vertex reconstruction

Long-lived particles, such as heavy flavor hadrons and τ-leptons, can be identified through the displacement of their vertex from the PV when extrapolating their tracks. This secondary-vertex (SV) is reconstructed with the inclusive vertex fitter (IVF) [108] approach. The IVF algorithm selects all reconstructed tracks of the event with $p_{\mathrm{T}} > 0.8 \ \mathrm{GeV}$ and a longitudinal impact parameter of less than 0.3 cm. Tracks with a 3D impact parameter of at least 50 μm and passing quality criteria on the 2D impact parameter are used as seeds in the following clustering procedure. A seed track and any other track are clustered if their distance is smaller than the distance between the track and the PV. Both distances are evaluated at the points of the closest approach of each object pair. Moreover, requirements on the angle between the seed track and any other track are also applied. The clustered tracks are then fitted with the AVF. Quality criteria are applied on the SV candidates and a cleaning is performed if two candidates share a significant portion of their tracks. Additionally, if a track is assigned to both PV and SV, the angular separation between the track and the two vertices is used to indicate the most compatible vertex. Finally, a refitting is performed on the SV candidate and the check for duplicates is repeated. The precise reconstruction of SVs is crucial for the calculation of variables used in heavy-flavor jet identification algorithms, such as the CSVv2 [103], DeepCSV [104], and DeepJet [112]. The SV reconstruction efficiency is given by the number of jets containing a reconstructed SV over the total number of jets. For jets with p_{T} above 20 GeV in $\mathrm{t\bar{t}}$ events, the efficiency for reconstructing a SV for b (udsg) flavor jets with the IVF method is about 75% (12%).

4.1.4   Calorimeter clustering and link algorithm

A clustering algorithm reconstructs energy *clusters* in each of the CMS calorimeters (ECAL, HCAL, and preshower). The clustering is seeded by cells with a local maximum energy deposit above a given threshold (cluster seeds), to which neighboring cells are associated and iteratively merged into a cluster if their energy exceeds the electronic noise by a factor of two. Once the PF elements (tracks and clusters) are reconstructed, a *link* algorithm connects pairs of neighboring elements and subsequently produces PF-blocks of these directly or indirectly linked elements. For each PF-block, the identification and reconstruction sequence proceeds, starting from the muon candidates. For every reconstructed muon candidate, its corresponding PF elements are subtracted from the PF-block and are not used for the reconstruction of any other object in the event. The procedure follows with the identification and reconstruction of electron candidates and energetic isolated photons. As with muons, used PF elements are excluded from further consideration. The remaining PF elements are

classified as charged hadrons, neutral hadrons, and non-isolated photons.

4.2   Particle identification and reconstruction

4.2.1   Muon candidates

Muons are reconstructed with more than 99% efficiency over the full detector acceptance as they benefit from the precise momentum measurement of the inner tracker and the high purity granted by the upstream calorimeters. The collection of muon tracks is composed of three types. *Standalone-muons tracks* are reconstructed via the Kalman filter technique, using solely information from the muon subdetectors. Hits from DT, CSC, and RPC are gathered and fitted to form muon-track segments. *Tracker-muons tracks* are built by extrapolating an inner track to the muon chambers. The inner track is required to have a transverse (total) momentum above 0.5 (2.5) GeV and its extrapolation to be in a compatible position with at least one muon segment. *Global-muons tracks* are built with an outside-in approach, by matching the parameters of standalone muon tracks with the ones of a tracker track and then performing a combined fit with a Kalman filter. Almost all muons are reconstructed as global muon tracks or as tracker muon tracks, and often as both, in which case they are merged into a single muon candidate.

The reconstructed muon tracks are then fed into the PF algorithm, which applies a set of selections based on the global and tracker muon properties, such as cuts on the track fit χ^2, the number of hits of each track, either in the inner tracker or in the muon system, or both, and the degree of compatibility between extrapolated tracker tracks and standalone-muon tracks. Different cuts on the compatibility value between the tracker and standalone-muon tracks and other variables describing the compatibility with the primary vertex define the muon identification type. The main identification types used in CMS analyses include *Loose*, *Medium*, *Tight*, *Soft*, and *High momentum* muons and provide different levels of muon efficiency and charged hadron misidentification rate[2]. The muon momentum is determined by the *Tune-P* algorithm [113], which chooses the p_T measurement from a set of refits. The momentum resolution for muons with p_T less than 200 GeV is dominated by the inner tracking system and thus, the Tune-P algorithm favors the momentum determined by the inner-track fit. Above 200 GeV, the muon momentum is determined by refits involving information from both the inner tracker and the muon stations. To reduce the rate of misidentification of charged hadrons as muons, isolation criteria can be applied. The isolation is evaluated relative to the muon's p_T by summing up the energy within a variable cone size, ΔR, surrounding the muon.

---

[2]Charged hadrons can be mis-reconstructed as muons (*punch-through*) when their hadron shower remnants reach the muon chambers.

4.2.2 Electron and isolated photon candidates

As electrons propagate through the material in front of the ECAL, they interact and emit bremsstrahlung photons, which will subsequently convert into electron-positron pairs. For this reason, the identification and reconstruction of electrons and isolated photons are conducted together, as they share similar properties and technical reconstruction issues.

The energy of the primary electron (or photon) is measured by collecting the ECAL energy clusters of individual particles and merging them into a *supercluster* (SC). The clusters are reconstructed around an electron seed, defined by a local energy maximum, in a small window in η and an extended window in ϕ to account for photon conversions and bremsstrahlung losses. When an electron loses energy due to bremsstrahlung, its trajectory changes curvature in the tracker. Thus, a dedicated tracking algorithm based on the Gaussian Sum Filter (GSF) [114, 115] is used to build the electron tracks, instead of the widespread Kalman Filter. The trajectory seed is generated with an ECAL-seeding or tracker-seeding approach. In the ECAL-seeding approach, SCs with transverse energy above 4 GeV are selected, and trajectory seeds compatible with the SC position are used to seed the GSF algorithm. This approach performs better for high-E_T isolated electrons, providing more than 95% seeding efficiency for $E_\mathrm{T} > 10$ GeV for electrons from the Z boson decay. The tracker-seeding approach runs on all generic tracks with $p_\mathrm{T} > 2$ GeV and applies matching criteria between each track and an ECAL cluster. The matching track seed is used to seed a GSF track if it satisfies a cut-based or a multivariate selection based on a boosted decision tree (BDT) that uses quality and matching variables as input. The tracker-seeding recovers some of the efficiency on the low-p_T and non-isolated electrons. All possible ECAL- and tracker-driven electron seeds are merged into a single collection and used as input to the GSF algorithm. The resulting GSF tracks are in turn used to refine the superclusters. The refined superclusters and associated electron tracks make up the reconstructed electron candidates.

A photon candidate is seeded by an ECAL SC with E_T greater than 10 GeV, without any link to a GSF track. Moreover, photon candidates are required to be isolated from GSF tracks and other calorimeter clusters in the event and their ECAL cluster energy distribution and the H/E ratio to be compatible with those expected from a photon shower. Jets or photons can be misidentified as electrons (*fakes*), while electrons from non-prompt sources, such as electrons from photon conversion in the tracker material or semileptonic decays of hadrons, are also classified as background for prompt-electrons. To reduce the misidentification rate, additional identification requirements based on a BDT algorithm are applied to each electron candidate. The BDT, trained separately in the ECAL barrel and endcaps, uses up to fourteen input variables related to the shower-shape, track quality and matching with clusters, as well as isolation and the amount of energy radiated off the GSF track. All tracks and clusters contributing to the reconstructed electrons or photons are masked against further processing.

4.2.3   Hadrons and non-isolated photons

After the removal of PF elements correlated with reconstructed muons, isolated electrons, and photons, the remaining particles are identified as hadrons coming from jet fragmentation and hadronization. Non-isolated photons from the decay of neutral pions can also be identified at this stage. The remaining ECAL and HCAL clusters, not linked to any track are assigned as photons and neutral hadron candidates respectively if located within the tracker acceptance ($|\eta| < 2.5$). Beyond the tracker acceptance, ECAL clusters linked to an HCAL cluster are assumed to come from hadrons (neutral or charged given the presence of associated inner track) while ECAL clusters without a linked HCAL cluster are assigned to photons. For each reconstructed charged hadron candidate, the calibrated calorimeter energy is compared to the sum of the track momenta to correct for additional particles associated with the PF-block. For a calibrated calorimeter energy larger than the sum of the track momenta, the excess is assigned to the presence of photons and neutral hadrons. In the case of compatible energies, no neutral particle is identified and the charged hadron momentum is corrected with a χ^2 fit using both measurements. In the rare case of much smaller calibrated calorimeter energy than the sum of track momenta, a search for muons is performed. All charged hadrons are assigned the charged pion mass.

4.2.4   Jets

Quarks and gluons emitted from the hard scattering process *fragment* and *hadronize* immediately (10^{-23} s) resulting in collimated sprays of hadrons, called *jets*. At the CMS experiment, jets are clustered from reconstructed PF candidates (charged and neutral hadrons) using the anti-k_{T} clustering algorithm [116]. The anti-k_{T} algorithm is a sequential recombination jet algorithm that iteratively clusters pairs of stable particles (decay length $c\tau > 1$ cm). Given a collection of PF objects, the clustering algorithm calculates two distances. The first one,

$$d_{iB} = \frac{1}{p_{\mathrm{T}i}^{2\mathrm{k}}}, \tag{4.1}$$

corresponds to the distance between the entity i and the beam (B), and the second one,

$$d_{ij} = min(p_{\mathrm{T}i}^{2\mathrm{k}}, p_{\mathrm{T}j}^{2\mathrm{k}})\frac{\Delta R_{ij}^2}{R^2}, \quad \text{where } \Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2, \tag{4.2}$$

corresponds to the distance between the objects i and j. The variables $p_{\mathrm{T}i}$, η_i, and ϕ_i are the transverse momentum, pseudorapidity and azimuth angle of i, respectively. R and k are tunable parameters, with the former corresponding to the clustering radius in the $\eta - \phi$ space, while the latter sets the power of the transverse momentum scale relative to the geometrical distance. The anti-k_{T} algorithm has a k-value of -1, while $k = 0$ and $k = 1$ correspond to the Cambridge/Aachen [117] and inclusive k_{T} [118, 119] algorithms, respectively. The

aforementioned distances are calculated iteratively; if d_{ij} is smaller than d_{iB}, then both particles, i and j, are removed from the list and combined into a single candidate. Otherwise, the particle i is removed from the list and considered as a new candidate jet. This procedure is repeated until no particles are left in the list. The clustering sequence begins with the most energetic partons such that a bias in the jet substructure from soft emissions is avoided. This results in well-defined cone-like jets of radius R. The algorithm uses an *infrared* and *collinear* safe approach [120], meaning that the clustering configuration remains unchanged under the addition of a soft particle or with substitution of a particle with a set of collinear particles of the same total momentum. These crucial properties must be satisfied simultaneously to avoid divergencies in perturbative QCD calculations of the jet shape. In this thesis, the main collection of jets is clustered with a cone size of $R = 0.4$ (AK4 jets). A collection of jets with $R = 0.8$ (AK8 jets) is also reconstructed for the purpose of identifying Lorentz-boosted $W^{\pm}$, Z^0 and Higgs bosons, as well as boosted top quarks [121]. The presence of additional proton-proton interactions within the same or nearby bunch crossings challenges the jet reconstruction. Additional tracks and calorimetric energy deposits can erroneously contribute and increase the jet momenta (referred to as *pileup offset*). To mitigate this effect, tracks identified to be originating from pileup vertices are subtracted before the jet clustering. Charged particles not associated with any pileup vertex and all neutral particles are kept. This procedure is known as Charged Hadron Subtraction (CHS) [102].

Jet Energy Calibration

The resulting PF jets are calibrated with p_{T}- and η-dependent energy corrections to account for pileup offset subtraction, detector response, and residual discrepancies between data and simulation. The jet energy calibration procedure is illustrated in Fig. 4.2.

- pileup offset is calculated as the average difference in p_{T} between matched jets from the same QCD dijet simulated events with and without pileup overlay. The corrections are applied to both data and simulated samples with the data-to-simulation scale factors being estimated by comparing zero-bias data and neutrino gun simulated events, using the random cone method [122].

- simulated response corrections are derived and applied on pileup offset corrected jets. These account for the non-linear jet p_{T} response from the calorimeters as well as from the tracker transition regions between the barrel, endcaps, and forward HF. The particle response R_{ptcl} is defined as:

$$R_{\mathrm{ptcl}}(\langle p_T \rangle, \eta) = \frac{\langle p_T \rangle}{\langle p_{\mathrm{T,ptcl}} \rangle} \tag{4.3}$$

where $\langle p_{\mathrm{T,ptcl}} \rangle$ is the mean p_{T} of the generator-level jet and $\langle p_{\mathrm{T}} \rangle$ is the mean p_{T} of the geometrically matched reconstructed jet in a given $p_{\mathrm{T,ptcl}}$ and η bin.

- residual corrections of jet energy response between data and simulation are applied last. The response is measured using the p_T balance and missing transverse momentum projection fraction (MPF) methods [122], using precisely measured reference objects, such as muons, electrons, and photons. In the p_T balance method, the response is measured by comparing the p_T of a reconstructed jet to the p_T of the reference object ($R_{\mathrm{jet},p_T} = \frac{p_{T,\mathrm{jet}}}{p_{T,\mathrm{ref}}}$), while in the MPF method, the reference object is compared to the recoiling hadronic activity of the event ($R_{\mathrm{jet,MPF}} = 1 + \frac{\vec{p}_T^{\,miss} \cdot \vec{p}_{T,\mathrm{ref}}}{p_{T,\mathrm{ref}}^2}$). The η-dependent corrections of the residual differences on the jet energy scale (JES) are obtained using dijet events with the MPF method. The p_T-dependent corrections for p_T up to 800 GeV, are derived by combining measurements from $Z^0 \to \mu\mu$+jet, $Z^0 \to e^+e^-$+jet, and γ+jet events by using both methods. For p_T above 800 GeV, the p_T-dependent residual corrections are directly constrained using simulated QCD multijet events.

All corrections derived from simulation are subject to systematic uncertainties arising from the modeling of initial and final state radiation. The systematic uncertainties on the jet energy scale and their correlations are given as a function of the jet η and p_T, and are below 3% across the phase space used in this analysis. The jet energy resolution (JER) is defined as the width of the response distribution $p_{T,\mathrm{reco}}/p_{T,\mathrm{ptcl}}$ and is measured as a function of jet p_T and η after all aforementioned jet energy corrections (JEC) are applied. For events with 20 to 30 additional interactions the jet energy resolution for jets in the detector region of $|\eta| < 0.5$ is around around 15% at 30 GeV, 10% at 100 GeV and 5% at 1 TeV [123].
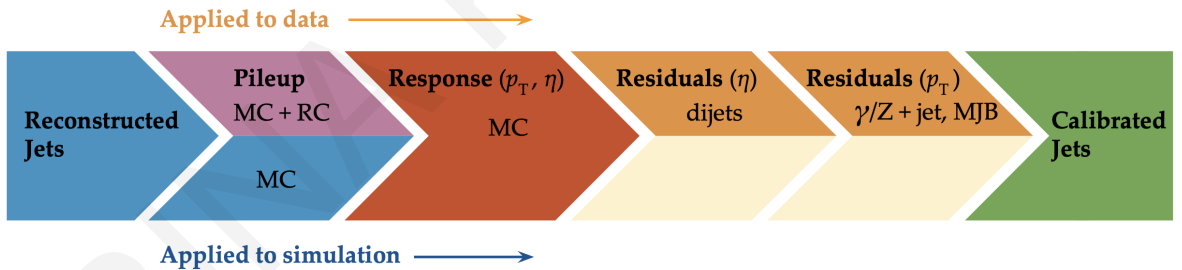
Figure 4.2: Workflow of the jet energy calibration procedure in data and simulation [124].

Noise jet identification

Physical PF jets can be discriminated from noise PF jets by exploiting variables sensitive to different sources of noise from HCAL and ECAL. A set of selection criteria on the jet energy fractions carried by the different types of PF candidates and the multiplicities of PF candidates clustered into a jet define the PF jet ID. There are three (two) working points used for 2016 (2017 & 2018) data: *loose*, *tight*, and *tight lepton veto*. The *loose* and *tight* working points remove jets from calorimetric noise, while the *tight lepton veto* working point additionally rejects misreconstructed isolated lepton candidates. For AK4 jets with the CHS applied, the efficiency of the loose working point (corresponding to the tight working point

for the 2017 & 2018 data) is more than 98%-99% for the whole η region, whereas the noise jet background rejection is more than 98% for $|\eta| < 3.0$ [125].

4.2.5 Identification of b jets

Jets originating from the hadronization of b (c) quarks are identified with heavy-flavor algorithms (*taggers*) that exploit the lifetime of hadrons present in such jets, which is of the order of 1.5 ps (1 ps or less). Depending on the p_{T} of the hadrons, such a lifetime results in displacements of a few mm, which is seen in the detector as displaced tracks from the primary vertex, as shown in Fig. 4.3. The origin of the displaced track can be reconstructed as a secondary-vertex (SV) with the method described in Section 4.1.3. The distance between the PV and the displaced track at the point of closest approach is called *impact parameter* (IP) and characterizes the displacement. Compared to light-flavored jets (u, d, s) or gluons, b
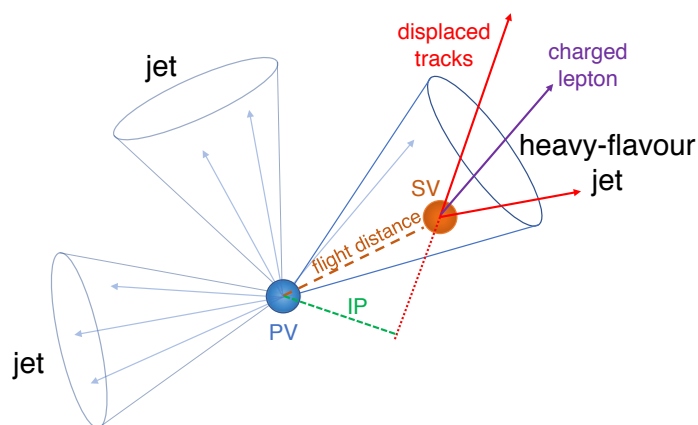
Figure 4.3: Schematic representation of a heavy-flavor jet being produced at the secondary-vertex [103].

and c quarks have larger masses, which results in larger transverse momentum relative to the jet axis for the heavy flavor hadron decay products. In addition to the topological properties of b (c) hadrons, the presence of charged electrons or muons in their decay chains in 20% (10%) of the cases can be exploited for the identification of b (c) jets. During Run 2, the CMS experiment made use of mainly three b taggers, the CSVv2, DeepCSV, and DeepJet.

- The CSVv2 algorithm is based on a multivariate method that combines information from displaced tracks and secondary vertices. The variables used include quality criteria on the secondary vertex, impact parameter, and displaced tracks, tracks and secondary vertices multiplicities, kinematic variables, such as the invariant mass of the total summed four-momenta of the tracks inside the jet, the p_{T} and energy of the tracks compared to the secondary vertex and the jet, distances in the $\eta - \phi$ space, and the p_{T} and η of the jet under consideration. All the variables are combined into an artificial neural network, namely a feed-forward multilayer perceptron with a single hidden layer.

- The DeepCSV algorithm is an updated version of the CSVv2 tagger. It is based on a deep neural network with more hidden layers and more nodes per layer. The initial set of variables fed to the fully connected dense network was the same as for the CSVv2 tagger but was later extended to include extra track-based variables.

- The DeepJet algorithm is the newest heavy-flavor tagging algorithm. It supersedes the previously mentioned taggers as its performance is significantly better. The performance of the DeepCSV and DeepJet algorithms for jets with $p_\mathrm{T} > 30$ GeV and $|\eta| < 2.5$ from simulated top pair events is shown in Fig. 4.5. The DeepJet algorithm uses a list of 650 input variables, including global event variables, charged and neutral PF candidate low-level features, and information about the secondary vertex associated with the jet. It is based on a set of Convolutional Neural Networks (CNNs) that are trained separately for each collection of charged and neutral PF candidates and secondary vertices, as shown in Fig. 4.4. The CNNs are followed by Recurrent Neu-
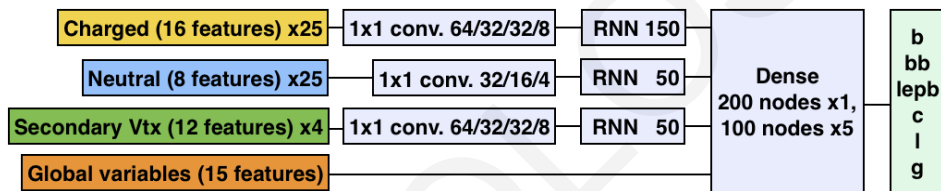
Figure 4.4: The architecture of the DeepJet b tagging algorithm [126].

ral Networks (RNNs) that combine the information for each sequence of constituents. The full jet information is combined with a set of dense fully connected layers. It has a total of six output nodes that allow for the tagging of b jets, c jets, and quark/gluon.

Each of the aforementioned heavy-flavor algorithms are operating with three working points: *loose*, *medium*, and *tight*, corresponding to 10%, 1%, and 0.1% b jet misidentification probability, respectively.

4.2.6 Identification of hadronic τ-leptons

Due to its relatively heavy mass of $m_\tau = 1.777$ GeV [8], the τ-lepton is the only lepton that can decay into hadrons, which happens approximately 65% of the times. Hadronic τ-leptons (τ_h) decay into either one or three charged pions ($\pi^\pm$) or kaons, and up to two neutral pions (π^0), and a neutrino (ν_τ). τ_h's are reconstructed with the hadron-plus-strips (HPS) algorithm [127], which combines charged and neutral PF candidates and checks whether they are compatible with specific τ_h decays. The algorithm can discriminate between τ_h decays and QCD-induced jets, electrons, and muons by exploiting the multiplicity, the collimation, and the isolation of its decay products. Each reconstructed τ_h candidate is required to have unit charge and mass compatible with its decay mode. All of the charged hadrons from the τ_h candidate are required to be inside a *signal cone*, defined by the radius $\Delta\mathrm{R} = 3.0 \, \mathrm{GeV}/p_\mathrm{T}$.
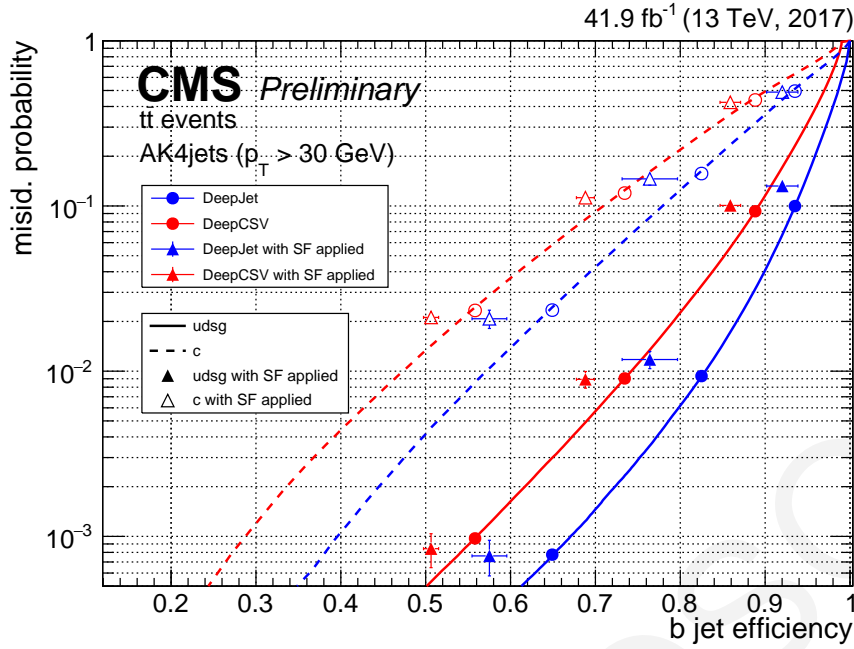
Figure 4.5: Performance of the DeepCSV and DeepJet algorithms in simulated top quark pair events [126].

The isolation of the τ_h candidate is computed using a multivariate approach which takes into account all charged particles and photons inside the *isolation cone*, defined by $\Delta R = 0.5$ [128].

4.2.7 Missing transverse momentum

The missing transverse momentum $\vec{p}_T^{\text{miss}}$ [129] is defined as the negative vectorial sum of the transverse momenta of all the reconstructed PF candidates in the event:

$$\vec{p}_T^{\text{miss}} = - \sum \vec{p_T}. \tag{4.4}$$

Its magnitude is referred to as p_T^{miss}. The energy scale and resolution corrections applied to jets are propagated to the calculation of p_T^{miss}.

4.2.8 Scalar transverse momentum

The scalar sum H_T is defined as the sum of the transverse momenta magnitude of all the reconstructed jets in the event:

$$H_T = \sum p_T. \tag{4.5}$$

The energy scale and resolution corrections applied to jets are propagated to the calculation of H_T.

5 Search for charged Higgs bosons at the LHC

This chapter describes the analysis strategy for the search for heavy charged Higgs bosons, decaying into a top and bottom quark-antiquark pair, $H^{\pm} \rightarrow tb$. The physics analysis is performed with proton-proton collision data, from the 2016 and 2017 LHC data taking periods, recorded by the CMS detector at a center-of-mass energy of $\sqrt{s} = 13$ TeV and correspond to integrated luminosities of 35.9 fb^{-1} and 41.5 fb^{-1}, respectively. The search is conducted in the fully hadronic final state (Fig. 5.1), where the $W^{\pm}$ bosons from the decay chains of both the charged Higgs boson and the associated top quark decay hadronically. The fully hadronic final state is characterized by four light quarks and four b quarks and provides the largest accessible branching fraction, corresponding to $\sim 45\%$. Moreover, as all the final state objects are detected, the full reconstruction of the charged Higgs boson mass can be feasible. Depending on the mass and the transverse momenta of the charged Higgs boson, the search can be performed in two distinct topologies, the boosted and the resolved. The boosted event topology targets charged Higgs bosons with large masses, typically above 1 TeV, or with relatively large p_{T}. The decay products of such massive charged Higgs bosons, a pair of top and bottom quarks, have average transverse momenta of several hundred GeV. As a result, the jets emerging from the subsequent top quark decay are highly collimated and cannot be reconstructed using the standard clustering algorithm for small-cone jets (AK4). Instead, likely, they are reconstructed as a single large-radius jet (AK8). Moreover, when the charged Higgs boson has relatively high p_{T}, all of its decay products become collinear. Events with highly-collimated top-quark or W boson candidates are reconstructed and an-
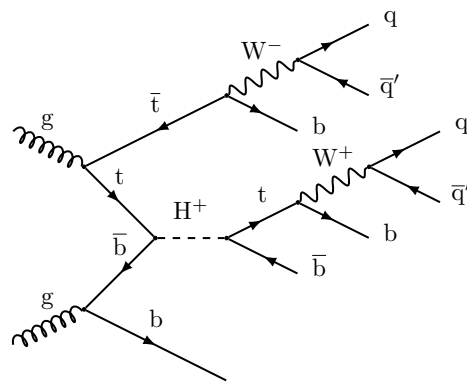
Figure 5.1: Leading order (LO) Feynman diagram of a charged Higgs boson produced in association with a top quark and decaying into a top and bottom quark-antiquark pair in a final state where both $W^{\pm}$ bosons decay hadronically.
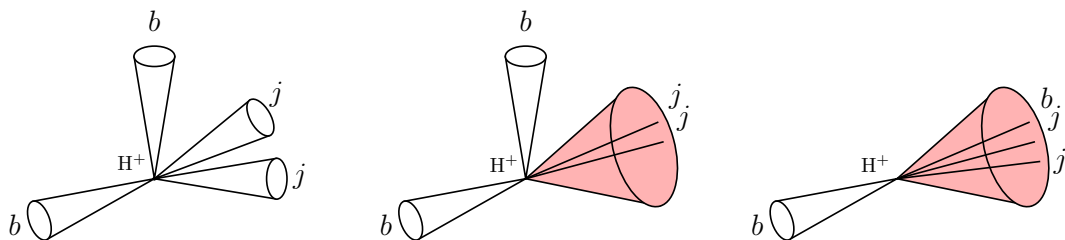
Figure 5.2: *Left:* Resolved topology occurs when the $H^{\pm}$ has moderate mass and relatively low p_T, all the decay products are well separated. *Middle (Right):* Boosted topologies occur as the $m_{H^{\pm}}$ increases and its decay products, such as the W boson (top quark), become collinear.

alyzed by the boosted analysis, while events with less boosted final states in which the top quark candidates are reconstructed with the standard small-cone clustering algorithm are analyzed by the resolved analysis. The three different topologies are illustrated in Fig. 5.2. The work on this thesis is focused on the resolved topology only. A graphical display of a signal candidate event in the resolved topology recorded by the CMS detector in 2017 is presented in Fig. 5.3. The 2016 analysis is the first to report results in the fully-hadronic final state of this channel and the combined work of the resolved (discussed in chapter 6) and boosted analyses was published in the Journal of High Energy Physics on July 20, 2020 [15]. The 2017 resolved analysis (discussed in chapter 7) is a continuation of the search for charged Higgs bosons in the same final state, using the 2017 LHC data, and features improved background measurement and signal extraction techniques and the estimation of the related systematic uncertainties. Model-independent upper limits on the product of the charged Higgs boson production cross-section and branching fraction into a top and a bottom quark ($\sigma_{pp \to t(b)H^{\pm}} \times \mathcal{B}(H^{\pm} \to tb)$) are presented as a function of the charged Higgs boson mass. The upper-limits extracted from the 2016 analysis are also interpreted in the parameter space of the Minimal Supersymmetric Standard Model hMSSM and $M_h^{125}(\tilde{\chi})$ benchmark scenarios.

5.1 Analysis Strategy

The 2016 and 2017 analyses share a common event selection flow. Online events are recorded with triggers requiring large jet and b jet multiplicities. On the offline analysis selection, events with isolated electrons and muons, as well as hadronically decaying τ-leptons are rejected to ensure a fully hadronic final state. Subsequently, events with the presence of at least seven jets, out of which at least 3 are tagged as b jets are selected. The set of selection criteria up to the b jet selection is referred to as the *baseline event selection* and is discussed in more detail in section 5.4. Machine learning techniques are used to reconstruct top quark candi-
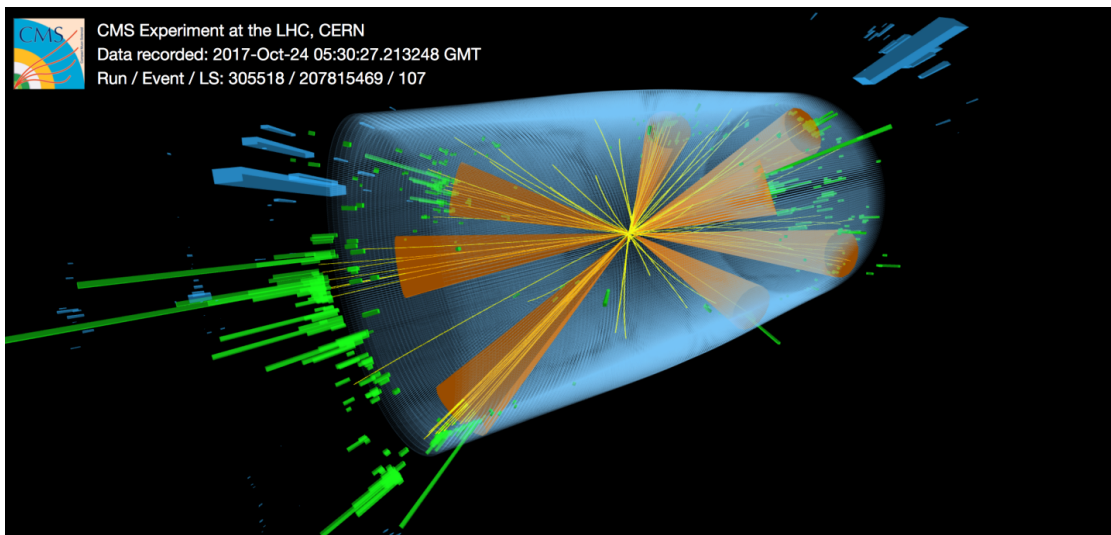
Figure 5.3: The display of a fully-hadronic candidate event with eight small-cone jets, represented by the orange cones. The signal candidate event has two top quarks decaying into a W boson and a b quark. The b quark fragmentation process gives rise to a b jet and the decay of the W boson results in two jets. Two additional b jets are produced, either from background QCD processes or the decay of a hypothetical charged Higgs boson.

dates from trijet combinations and exactly two top quark candidates are selected. Finally, the charged Higgs boson candidate is reconstructed with the leading-in-p_T top quark candidate of the event and the leading-in-p_T b jet that is not used in the reconstruction of the two top quark candidates. Due to the large jet multiplicity, this final state suffers from large QCD multijet and irreducible top quark-antiquark pair production background, as well as large combinatoric self-background. Background events are estimated from both data-driven and simulation techniques. The simulated samples modeling the signal and background processes are described in section 5.3.

The search is performed through a binned maximum-likelihood (ML) fit on distributions sensitive to the signal process to assess the agreement with the background-only hypothesis or the presence of a signal. In the case of the 2016 analysis, the final discriminant is the invariant mass of the charged Higgs boson candidate, while for the 2017 analysis the distribution of the score of a deep neural network is used. The statistical methods used are described in section 5.5.

5.2 Data samples

The collision data used in this search are collected with the CMS detector during the 2016 and 2017 data taking periods, and correspond to integrated luminosities of $39.5\,\text{fb}^{-1}$ and $41.5\,\text{fb}^{-1}$, respectively. Events are obtained from the JETHT primary dataset with triggers requiring high jet and b jet multiplicities. A summary of the collision data for 2016 and 2017 are shown in Tables 5.1 and 5.2, respectively.

Dataset	Runs	$\mathcal{L}$ (pb^{-1})
JetHT_Run2016B_03Feb2017_ver2_v2	273150–275376	5750.126
JetHT_Run2016C_03Feb2017_v1	275656–276283	2572.903
JetHT_Run2016D_03Feb2017_v1	276315–276811	4242.292
JetHT_Run2016E_03Feb2017_v1	276831–277420	4024.755
JetHT_Run2016F_03Feb2017_v1	277932–278801	2697.733
JetHT_Run2016F_03Feb2017_v1	278801–278808	406.776
JetHT_Run2016G_03Feb2017_v1	278820–280385	7574.787
JetHT_Run2016H_03Feb2017_ver2_v1	281613–284035	8434.663
JetHT_Run2016H_03Feb2017_ver3_v1	284036–284044	215.965
Total integrated luminosity (fb^{-1})		35.92

Table 5.1: Collision data from 2016 LHC data-taking period

Dataset	Runs	$\mathcal{L}$ (pb^{-1})
JetHT_Run2017B_09Aug2019_UL2017_v1	297047 - 299329	4803.13
JetHT_Run2017C_09Aug2019_UL2017_v1	299368 - 302029	9574.02
JetHT_Run2017D_09Aug2019_UL2017_v1	302031 - 302663	4247.68
JetHT_Run2017E_09Aug2019_UL2017_v1	303824 - 304797	9313.74
JetHT_Run2017F_09Aug2019_UL2017_v1	305040 - 306460	13534.52
Total integrated luminosity (fb^{-1})		41.47

Table 5.2: Collision data from 2017 LHC data-taking period

5.3 Signal and background modeling

Signal and background contributions are simulated by various Monte Carlo (MC) event generators. All MC generated events are processed through a simulation of the CMS detector response done with the GEANT4 [130] package and are reconstructed with the same algorithms as the observed data. Parton distribution functions (PDFs) are modeled with the NNPDF3.0 [131] (NNPDF3.1 [132]) parametrization tune for 2016 (2017) samples. For all processes, the parton showering and hadronization are simulated with PYTHIA8 [133] event generator. The modeling of the underlying event (UE) is performed with the tune CUETP8M2T4 for $t\bar{t}$ background and with the tune CUETP8M1 [134] for all other processes (CP5 [135]) in 2016 (2017) simulated samples. All MC generated samples are normalized to the highest order cross-section calculations, corresponding usually to NNLO in perturbative QCD (pQCD) and NLO in electroweak corrections. The signal and background simulated samples used in 2016 and 2017 analyses are listed in Tables 5.3 and 5.4, respectively.

The signal samples are generated with the associated top quark production of the charged Higgs boson in the 4FS at next-to-leading order (NLO) precision in pQCD, using the MAD-GRAPH5_AMC@NLO v2.3.3 [136] for the 2016 signal samples and v2.6.5 for the 2017 signal samples. A wide range of masses for the charged Higgs boson is produced, from 200 GeV to 3 TeV. The total cross-section for the H$^{\pm}$ production is obtained with the *Santander matching* scheme [52]. Typical values of the cross-section are of the order of 1 pb for $m_{H^{\pm}} = 200$ GeV, and down to about 10^{-4} pb for a mass of 3 TeV [55, 137]. The branching

fractions $\mathcal{B}(\mathrm{H}^{\pm} \to \mathrm{t\bar{b}})$ are computed with HDECAY [138] for different values of $\tan\beta$.

The $\mathrm{t\bar{t}}$ background sample is generated with POWHEG [139, 140, 141] at NLO in perturbative QCD, assuming a mass of the top quark equal to 172.5 GeV. For 2016, the $\mathrm{t\bar{t}}$ simulated sample contained all final states of the top quark-antiquark pair, while for 2017, the fully leptonic, semileptonic, and fully hadronic final states are generated separately to increase the statistics. Other important background sources include single top quark production (single t), $\mathrm{t\bar{t}}$+X where X = W, Z, H, or $\mathrm{t\bar{t}}$, and electroweak processes including Z, W bosons, and diboson (WW, WZ, ZZ) production. Some of the aforementioned processes are not taken into account in the 2017 analysis as their simulated samples were not available at the time the analysis was performed. However, most of the missing samples do not have a significant contribution to the final yield of expected backgrounds after all the analysis selections are applied. The t-channel of the single top quark process is generated with POWHEG at NLO precision in pQCD using the 4FS and interfaced with MADSPIN for simulating the top quark decay. The s-channel process is simulated with MADGRAPH5_AMC@NLO and the tW-channel process with POWHEG using the 5FS for 2016 samples. Background events from the production of $\mathrm{t\bar{t}}$ in association with W, Z, or $\mathrm{t\bar{t}}$ are simulated with MADGRAPH and POWHEG at NLO in pQCD, while $\mathrm{t\bar{t}}$ in association with the Higgs boson and its subsequent decay into a b quark-antiquark pair is generated with POWHEG at NLO [142]. Electroweak backgrounds with boson and diboson production are simulated at LO using MADGRAPH and PYTHIA8 event generators.

5.4   Baseline event selection

Collision events are reconstructed with the PF algorithm, which uses information from the tracker, calorimeters, and muon systems of the CMS detector to reconstruct and identify individual particle candidates as charged or neutral hadrons, photons, electrons, or muons. Higher-level objects are reconstructed by combining the PF candidates.

5.4.1   Primary vertex

The primary pp interaction vertex (PV) is considered the reconstructed vertex with the largest value of p_{T}^2 summed over the associated charged particle tracks. Moreover, it is required to be located within 24 cm (2 cm) from the detector center in the direction along (perpendicular to) the beam axis.

5.4.2   Lepton (e/μ) and hadronically decaying τ-lepton rejection

Electrons are reconstructed by matching charged-particle trajectories from the tracker to energy deposits in the electromagnetic calorimeter. For 2016, the identification of electrons is performed with a multivariate analysis algorithm, provided by the e/γ Physics Object Group

Process	Generator / pQCD accuracy	Cross-section (pb)
Signal, pp → tbH$^\pm$ *(4FS)*, H$^\pm$ → tb		
m$_{H^\pm}$ = 200 − 3000 GeV	MADGRAPH5_AMC@NLO/ NLO	$1 - 10^{-4}$
t$\bar{\text{t}}$, t$\bar{\text{t}}$ + X, *Single top backgrounds*		
t$\bar{\text{t}}$	POWHEG + PYTHIA8/ NLO	831.76
t$\bar{\text{t}}$t$\bar{\text{t}}$	POWHEG + PYTHIA8/ NLO	9.103×10^{-3}
t$\bar{\text{t}}$W + jets	MADGRAPH5_AMC@NLO/ NLO	4.034×10^{-1}
t$\bar{\text{t}}$Z + jets	POWHEG + PYTHIA8/ NLO	5.297×10^{-1}
t$\bar{\text{t}}$H (H → b$\bar{\text{b}}$)	POWHEG/ NLO	0.2953
t, s channel	MADGRAPH5_AMC@NLO/ NLO	11.36
t, t channel (4FS)	POWHEG + PYTHIA8/ NLO	136.02
$\bar{\text{t}}$, t channel (4FS)	POWHEG + PYTHIA8/ NLO	80.95
t, tW channel (5FS)	POWHEG + PYTHIA8/ NLO	30.11
$\bar{\text{t}}$, tW channel (5FS)	POWHEG + PYTHIA8/ NLO	30.09
Electroweak backgrounds		
Z/γ^* + jets	MADGRAPH + PYTHIA8/ NLO	1209.0
W + jets	MADGRAPH + PYTHIA8/ LO	99.36
Z + jets	MADGRAPH/ LO	5.67
WZ	PYTHIA8/ NLO	47.13
WW (WW → 4q)	POWHEG/ NLO	51.723
ZZ (ZZ → 4q)	MADGRAPH5_AMC@NLO/ NLO	6.883

Table 5.3: Simulated samples used in the 2016 analysis to model signal and background processes. The generator and the accuracy in perturbative QCD used in the production of each sample, together with the process cross-section are also listed.

Process	Generator / pQCD accuracy	Cross-section (pb)
Signal, pp → tbH$^\pm$ *(4FS)*, H$^\pm$ → tb		
m$_{H^\pm}$ = 200 − 3000 GeV	MADGRAPH5_AMC@NLO/ NLO	$1 - 10^{-4}$
t$\bar{\text{t}}$, t$\bar{\text{t}}$ + X, *Single top backgrounds*		
t$\bar{\text{t}}$ (t$\bar{\text{t}}$ → $\ell\nu$2q2b)	POWHEG + PYTHIA8/ NLO	365.45
t$\bar{\text{t}}$ (t$\bar{\text{t}}$ → 4q2b)	POWHEG + PYTHIA8/ NLO	377.96
t$\bar{\text{t}}$Z + jets	MADGRAPH5_AMC@NLO + PYTHIA8/ NLO	5.297×10^{-1}
t, t channel (4FS)	POWHEG + PYTHIA8/ NLO	136.02
$\bar{\text{t}}$, t channel (4FS)	POWHEG + PYTHIA8/ NLO	80.95
Electroweak backgrounds		
WZ	PYTHIA8/ NLO	47.13
WW	PYTHIA8/ NLO	118.7
ZZ	PYTHIA8/ NLO	16.523

Table 5.4: Simulated samples used in 2017 analysis to model signal and background processes. The generator and the accuracy in perturbative QCD used in the production of each sample, together with the process cross-section are also listed.

(POG), while for 2017, a cut-based identification algorithm is used. Both algorithms apply a set of selection criteria on the electromagnetic shower shape, track-cluster matching, and consistency between the cluster energy and track momentum. Muons are reconstructed by matching hit patterns from the tracker to signals in the muon systems. Electron or muon isolation is quantified with the *mini-isolation* variable [143], I_{mini}. This variable computes the scalar sum of the transverse energy of all charged particles, neutral hadrons, and photons surrounding the lepton in a cone in $\eta - \phi$ space, the size of which scales as $1/p_T^\ell$, where p_T^ℓ is the transverse momentum of the lepton:

$$I_{mini} = \sum_{charged} p_T + \max\Big(0, \sum_{neutral} p_T + \sum_{photons} p_T - \rho \mathcal{A}_{eff}\big(\frac{R^{mini\text{-}iso}}{0.3}\big)^2\Big). \quad (5.1)$$

To account for the effects from pileup, the energy sums of the neutral particles are corrected with the average energy density in the event, ρ [144], scaled by its *effective area* $\mathcal{A}_{eff}$. The effective area corrections estimate the amount of energy expected in the isolation cone and are calculated by taking the ratio of the fitted slopes of the energy sums and energy density distributions as a function of the number of reconstructed primary vertices in the event. The corrections are calculated separately for barrel and endcaps, and for muons and electrons. In the analysis, the relative mini-isolation is defined as the sum of the transverse energy of all charged hadrons from the PV, neutral hadrons, and photons, in a cone of radius $R^{mini\text{-}iso}$, divided by p_T^ℓ. The radius of the cone varies as follows:

$$R^{mini\text{-}iso} = \begin{cases} 0.2 & \text{for } p_T^\ell \leq 50 \text{ GeV} \\ 10 \text{ GeV}/p_T^\ell & \text{for } p_T^\ell \in (50, 200) \text{ GeV} \\ 0.05 & \text{for } p_T^\ell \geq 200 \text{ GeV} \end{cases} \quad (5.2)$$

The cone size dependence on the p_T^ℓ reduces the rate of accidental overlaps between particles decaying from a massive parent particle of mass M and large p_T, i.e. overlaps between a lepton and a b jet originating from a Lorentz-boosted top quark. Events containing electrons or muons with p_T above 10 GeV and $|\eta| < 2.4$ satisfying loose identification and isolation criteria ($I_{mini}^{rel.} < 0.40$) are rejected.

Hadronically decaying τ-leptons are reconstructed with the HPS algorithm and are required to have p_T above 20 GeV, $|\eta| < 2.3$, and loose identification and isolation criteria. The latter is decided based on a multivariate algorithm that combines information on their identification, isolation, and lifetime [128, 127]. Events containing hadronically decaying τ-leptons satisfying the aforementioned criteria are rejected.

PF Jet ID	2016	2017
Neutral Hadron Fraction	< 0.99	< 0.90
Neutral EM Fraction	< 0.99	< 0.90
Number of Constituents	> 1	> 1
Muon Fraction	-	-
Charged Hadron Fraction	> 0	> 0
Charged Multiplicity	> 0	> 0
Charged EM Fraction	< 0.99	-

Table 5.5: Jet identification criteria, for jets contained within the tracker volume of $|\eta| < 2.4$, for 2016 and 2017 analyses.

5.4.3 Hadronic jets

The hadronic jets are reconstructed as described in section 4.2.4. Jets are required to satisfy a set of quality criteria listed in Table 5.5, for 2016 and 2017 analyses. These are applied in order to reject fake, badly reconstructed, or noise jets, while identifying real jets with an efficiency of more than 98% (99%) for 2016 (2017). At least seven jets with p_T above 30 GeV and with $|\eta| < 2.4$ are required. The six leading-in-p_T jets are required to have p_T above 40 GeV. The scalar sum of p_T of all selected jets in the event, denoted as H_T, is required to be at least 500 GeV.

5.4.4 b jets

The offline identification of jets arising from the hadronization of b quarks (b jets) is performed with the Combined Secondary Vertex (CSVv2) [103] (2016 analysis) and the DEEP-JET multiclass [112] (2017 analysis) flavor-tagging algorithms. Two operating points are used in the analysis, the loose and the medium, corresponding to 10% and 1% misidentification rate of a light jet to be identified as a b jet, respectively [126]. Signal-like events are required to have at least three medium-tagged b jets with $p_T > 40$ GeV and $|\eta| < 2.4$.

5.5 Statistical Methods

This section summarizes the statistical procedures used to assess, given a sample of observed collision events (*data*), the validity of the hypothesis of a possible signal of the charged Higgs boson production (*signal+background hypothesis*) or its absence (*background-only hypothesis*). Depending on how one defines probability, statistical methods fall into two broad categories: the *frequentist* statistics and the *Bayesian* statistics. In frequentist statistics, the probability is a measure of how frequently a claim is true when performing a long series of repeatable experiments whose outcome is uncertain, in this case, collision events. In Bayesian statistics, the probability is defined in terms of *degree of belief*; the available knowledge about a parameter in the statistical model is updated with the information from the observed

data. Most searches performed at the LHC follow the frequentist approach used by the AT-LAS and CMS collaborations in the combination of results for the Higgs boson search in 2011 [145].

The statistical analysis is done based on a binned maximum-likelihood fit on distributions that provide the greatest discrimination of the signal from the expected backgrounds. These distributions are referred to as *templates*. The template for 2016 analysis is the invariant mass of the reconstructed charged Higgs boson candidate, while for 2017 analysis is the score of a deep neural network. Henceforth, the event yield of the signal process in each bin i of the template will be denoted by s_i, while the expected background yield by b_i. Both, signal and background yields are subject to multiple systematic uncertainties, which can affect the overall normalization of the yield (rate uncertainties), the shape of the distributions (shape uncertainties), or both. Systematic uncertainties are incorporated in the likelihood fit as nuisance parameters (NPs), collectively denoted as $\vec{\theta} = (\theta_1, ... \theta_n)$, and the signal and background yields become a function of them, $s(\vec{\theta})$ and $b(\vec{\theta})$.

The *parameter of interest (POI)* in the statistical analysis is the *signal strength modifier*, μ, defined as the product of the cross-section of the charged Higgs boson production in association with a top quark times the branching fraction of the charged Higgs boson to decay into a top and bottom quark-antiquark pair:

$$\mu = \frac{\sigma_{\mathrm{pp} \to \mathrm{t(b)H^{\pm}}} \times \mathcal{B}(\mathrm{H^{\pm}} \to \mathrm{tb})}{s_i}, \tag{5.3}$$

where s_i is the MC predicted signal yield in bin i. In the *signal+background* hypothesis, the signal strength modifier can scale the MC predicted signal event yield in each bin i of the template, resulting to a total event yield of $\mu s_i(\vec{\theta}) + b_i(\vec{\theta})$. In the *background-only* hypothesis, the signal strength is equal to zero and the total event yield is just the expected background yield, $b_i(\vec{\theta})$. The probability to observe n_i events in a bin i of the template is given the following Poisson distribution:

$$\mathrm{Poisson}(\mathrm{data}_i | \mu s_i(\theta) + b_i(\theta)) = \frac{(\mu s_i(\theta) + b_i(\theta))^{n_i}}{n_i!} e^{-(\mu s_i(\theta) + b_i(\theta))}. \tag{5.4}$$

The *data* here, correspond either to actual observed collision events or *pseudo-data* (toys) used to construct sampling distributions for statistical testing. As the template bins are statistically independent and follow the same Poisson p.d.f., the joint p.d.f. factorizes and the likelihood function becomes:

$$\mathcal{L}(\mathrm{data}|\mu, \theta) = \prod_i \mathrm{Poisson}(\mathrm{data}_i | \mu s_i(\theta) + b_i(\theta)) \times \prod_j \rho\left(\theta_j | \tilde{\theta}_j\right). \tag{5.5}$$

The last term corresponds to the probability density function $\rho\left(\theta_j | \tilde{\theta}_j\right)$ of each nuisance parameter, which indicates the probability of obtaining the true value θ_j from its initial estimate

$\tilde{\theta}_j$. The $\tilde{\theta}$ cannot be determined *a priori*, and thus, it is convenient to use a Bayesian inter-
pretation in order to incorporate the information of the estimated value of the nuisance *a
posteriori*:

$$\rho(\theta|\tilde{\theta}) \sim p(\tilde{\theta}|\theta) \cdot \pi(\theta), \tag{5.6}$$

where $p(\tilde{\theta}|\theta)$ is the p.d.f. of the nuisance parameter, which is obtained in auxiliary measure-
ment, e.g. the derivation of the top tagging scale factors and $\pi(\theta)$ is a flat prior of the nuisance
parameter. The nuisance parameters associated with a rate-only systematic uncertainties are
parameterized with the log-normal distribution:

$$p(\theta|\tilde{\theta}) = \frac{1}{\sqrt{2\pi}\ln\kappa}\exp\left(-\frac{\left(\ln(\theta/\tilde{\theta})\right)^2}{2(\ln\kappa)^2}\right)\frac{1}{\theta}, \tag{5.7}$$

where $\kappa = 1 + \epsilon$ and ϵ is the relative scale of the variation. Shape uncertainties are taken
into account with *template morphing* [146] techniques. A shape uncertainty is modeled by
defining a set of shifted templates corresponding to up and down by one standard deviation
of the relevant nuisance parameter, while the nominal template remains unchanged. These
templates will be interpolated quadratically for shifts below 1σ and linearly beyond.

5.5.1 Statistical tests

In order to evaluate the validity of a certain statement concerning the observed data or
pseudo-data, one needs to construct a *test statistic*. Following the *modified frequentist ap-
proach*, the profile-likelihood test statistic $\tilde{q}_\mu$ is used at the LHC:

$$\tilde{q}_\mu = -2\ln\frac{\mathcal{L}\left(\text{data}|\mu, \hat{\hat{\vec{\theta}}}_\mu\right)}{\mathcal{L}\left(\text{data}|\hat{\mu}, \hat{\vec{\theta}}\right)}, \quad \text{with the constraint} \quad 0 \leq \hat{\mu} \leq \mu. \tag{5.8}$$

The pair of parameter estimators $\hat{\mu}$ and $\hat{\vec{\theta}}$ correspond to the estimators resulting in a global
maximum of the likelihood. The $\hat{\hat{\vec{\theta}}}_\mu$ refer to the conditional maximum likelihood estimators
of $\vec{\theta}$, given the signal strength parameter μ. The lower constraint $0 \leq \hat{\mu}$ indicates that
the signal rate must be positive, while the upper constraint $\hat{\mu} \leq \mu$ is imposed to ensure a
one-sided confidence interval. In physics terms, it means that upward fluctuations of the data
($\hat{\mu} > \mu$) are not considered as evidence against the signal hypothesis. The following p-values
can be obtained for the signal plus background ($s + b$) and background-only (b) hypotheses:

$$\begin{aligned}
p_{s+b} &= P\left(\tilde{q}_\mu \geq \tilde{q}_\mu^{\text{obs}}|\text{signal+background}\right) = \int_{\tilde{q}_\mu^{\text{obs}}}^{\infty} f\left(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{\text{obs}}\right) d\tilde{q}_\mu \\
1 - p_{\text{b}} &= P\left(\tilde{q}_\mu \geq \tilde{q}_\mu^{\text{obs}}|\text{background only}\right) = \int_{\tilde{q}_0^{\text{obs}}}^{\infty} f\left(\tilde{q}_\mu|0, \hat{\theta}_0^{\text{obs}}\right) d\tilde{q}_\mu,
\end{aligned} \tag{5.9}$$

where $\tilde{q}_\mu^{\mathrm{obs}}$ is the observed value of the test statistic for the tested signal strength parameter μ and is evaluated from equation 5.8. The $f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{\mathrm{obs}})$ and $f(\tilde{q}_\mu|0, \hat{\theta}_0^{\mathrm{obs}})$ refer to the PDFs of $\tilde{q}_\mu$ under the corresponding hypotheses and their distributions are obtained by performing MC toy experiments while keeping the nuisance parameters fixed to the maximum likelihood values $\hat{\theta}_\mu^{\mathrm{obs}}$ and $\hat{\theta}_0^{\mathrm{obs}}$. The p-values can be interpreted as follows:

- p_{s+b} is the probability for the observed value $\tilde{q}_\mu^{obs}$ to be as or less compatible with the signal plus background hypothesis, assuming the signal plus background hypothesis. Such a probability is denoted as $\mathrm{CL_{s+b}}$.

- $1 - p_b$ is the probability, assuming the background-only hypothesis, of getting a value of the test statistic as result of the test at least as extreme as the observed test statistic $\tilde{q}_\mu^{obs}$. Such a probability is denoted by $\mathrm{CL_b}$.

The *hypothesis test* based on the modified frequentist approach [147, 148] is defined as the ratio of $\mathrm{CL_{s+b}}$ and $\mathrm{CL_b}$:

$$\mathrm{CL_s}(\mu) = \frac{\mathrm{CL_{s+b}}(\mu)}{\mathrm{CL_b}(\mu)} \leq \alpha. \tag{5.10}$$

where α is a predefined threshold characterizing the *significance level* of the test. In the modified frequentist approach, the value of $\mathrm{CL_s}$ is required to be less or equal to $\alpha = 0.05$ to declare a 95% C.L. signal exclusion. As the limits based on the $\mathrm{CL_s}$ criterion are, by construction, one-sided, a 95% confidence level upper limit is set on the signal strength modifier for the charged Higgs boson signal ($\mu^{95\% \ \mathrm{CL}}$).

5.5.2 Expected results

Searching for a new particle at the LHC is initially performed in a *blinded* manner. This ensures that all selection criteria and analytical decisions are not biased or tweaked based on the observed data. Following an extensive cross-check of the methods used, the *unblinding* of the data is the last step in the analysis chain. While in the *blinded* state, one can test the sensitivity of the analysis by calculating the expected median upper limit for the background-only hypothesis. This can be done by replacing the observed data ensemble with a large number of generated pseudo-data. However, this procedure may become very computationally expensive and is rather not needed due to the asymptotic behavior of the test statistic $\tilde{q}^\mu$ at a large number of data sets. In this limit, the distribution of $\tilde{q}^\mu$ approaches a non-central chi-square distribution [149] and one can use the *Asimov* dataset, instead. By definition, the Asimov data set is constructed such that the estimators for a parameter of interest are identical to their true value. The $\mathrm{CL_s}$ can be then computed as:

$$\mathrm{CL_s} = \frac{1 - \Phi(\sqrt{\tilde{q}_\mu})}{\Phi\left(\sqrt{\tilde{q}_\mu^{Asimov}} - \sqrt{\tilde{q}_\mu}\right)}, \tag{5.11}$$

where $\tilde{q}_\mu^{Asimov}$ is the test statistic evaluated on the Asimov dataset, and $\Phi(x)$ is the cumulative Gaussian distribution

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt. \tag{5.12}$$

6   Search for $H^\pm \to tb$ with the 2016 LHC data

This chapter describes the first search for charged Higgs bosons in the $H^\pm \to tb$ decay, in the fully hadronic final state, with the 2016 LHC data and the CMS detector. The chapter is structured as follows. Section 6.1 describes the online trigger selection and efficiency measurement. The event selection workflow after the *baseline event selection* is discussed in detail in section 6.2. Section 6.3 describes the data quality filtering process and the corrections applied to the simulated samples. Section 6.4 is focused on the measurement of the backgrounds and section 6.5 describes the measurements performed for the evaluation of the systematic uncertainties. The results and upper limits on the cross-section times the branching fraction of the charged Higgs boson decaying into a top and bottom quark-antiquark pair, $\sigma_{pp \to t(b)H^\pm} \times \mathcal{B}(H^\pm \to tb)$ are discussed in sections 6.6 and 6.7, respectively.

6.1   Online trigger selection

The events used in this analysis have been recorded with triggers requiring multiple jets and b tagging requirements. They consist of calo jet and PF jet filters, with the former ones having looser selection criteria than the latter. Since the PF algorithm requires large computational time, the calo jet filters precede to reduce the number of not interesting events. The first two triggers listed in Table 6.1 were designed and tuned to select events with at least six small-cone PF jets, out of which at least one or two are identified online as b jets, using a streamlined version of the CSVv2 algorithm. Both HLT paths are seeded by the logical OR of L1 H_T seeds (`L1_HTT280`, `L1_HTT300`, and `L1_HTT320`), where the L1 H_T is calculated from all L1 jets with $p_T > 30$ GeV and $|\eta| < 3.0$. The seed `L1_HTT280` was unprescaled for most of 2016 data-taking period and only got prescaled at the end of the year when LHC delivered higher instantaneous luminosity. The seed `L1_HTT300` also got prescaled for a very short amount of time. The path `HLT_PFHT450_SixJet40_BTagCSV_p057` requires the presence of at least six small-cone calo jets with $p_T > 35$ GeV and $|\eta| < 2.6$. The scalar sum of the p_T of these calo jets (*calo H_T*) is required to have a minimum threshold

Table 6.1: The signal triggers used together with their requirements on calo and PF jet levels.

HLT Path	Calo jets				PF jets			
	H_T	p_T	b-tags	b-discriminator WP	H_T	p_T	b-tags	b-discriminator WP
HLT_PFHT400_SixJet30_DoubleBTagCSV_p056	300	25	1	0.44	400	30	2	0.63
HLT_PFHT450_SixJet40_BTagCSV_p056	300	35	-	-	450	40	1	0.63
HLT_PFJet450	-	400	-	-	-	450	-	-

of 300 GeV. The PF jet filter is built in a similar way, requiring at least six PF jets with $p_T > 40$ GeV and $|\eta| < 2.6$. Their scalar sum (PF H_T) must have a minimum value of 450 GeV. The b tagging requirements are only applied on PF jets at the last filter. At least one of the selected PF jets is required to be identified as a b jet with a b tagging discriminator above 0.63. Similarly, the `HLT_PFHT400_SixJet30_DoubleBTagCSV_p056` requires the presence of at least six calo jets with $p_T > 25$ GeV and $|\eta| < 2.6$, out of which one is required to be tagged as a b jet with a b tagging discriminator greater than 0.44. The calo H_T is required to be more than 300 GeV. At PF-level, at least six jets with $p_T > 30$ GeV and $|\eta| < 2.6$ are required, where at least two must be tagged as b jets with a b tagging discriminator above 0.63. The PF H_T must have a threshold of 400 GeV. Both signal paths remained unprescaled during the 2016 LHC run. To monitor the b tagging performance of the CSVv2 algorithm, two control triggers were designed with the same requirements in calo jets and PF jets filters, except for the b tagging sequences. In the last era of the 2016 run (`Run2016H`), a firmware bug in the H_T sum calculation caused inefficiency in the high H_T region. The problem was mitigated offline with the use of a high-p_T jet trigger, the `HLT_PFJet450`, which recovered some of the efficiency.

6.1.1 Efficiency measurement

The trigger efficiency is measured using a sample with semileptonic $t\bar{t}$ events where one of the $W^\pm$ bosons decays into a muon and a neutrino and the other one decays hadronically. The data and simulation samples used are listed in Tables 6.2 and 6.3, respectively. The events are recorded by the unprescaled `HLT_IsoMu24` trigger and are required to have exactly one reconstructed muon identified with loose criteria, with $p_T > 26$ GeV and $|\eta| < 2.4$. The muon is required to be isolated with a relative mini-isolation $I_{mini} < 0.4$. Events with reconstructed electrons or hadronically decaying τ-leptons are vetoed as in the main analysis selection. The events are further required to have a signal-like final state, with at least seven jets with $|\eta| < 2.4$, where the first six leading-in-p_T jets must have a p_T greater than 40 GeV, and the remaining jets to have a p_T greater than 30 GeV. The scalar sum of the p_T of all selected jets, H_T, is required to be at least 500 GeV. Moreover, at least two of the selected jets are required to be identified as b jets with $p_T > 40\,(30)$ GeV for the leading (sub-leading) b jet. The avoid having overlapping objects, the selected muon is required to be away from any of the selected jets with a distance in the $\eta - \phi$ space of more than 0.4. The efficiency of the signal trigger paths is defined as:

$$\varepsilon = \frac{N_{events} \text{ passing(offline selection \&\& } \texttt{HLT\_IsoMu24} \text{ \&\& } \texttt{OR} \text{ of all signal triggers)}}{N_{events} \text{ passing(offline selection \&\& } \texttt{HLT\_IsoMu24})}$$

(6.1)

and is measured as a function of the p_T and η of the sixth leading-in-p_T jet, as shown in Fig. 6.1. The trigger efficiency reaches a plateau at $p_{T,6} \sim 48$ GeV, while the overall efficiency as a function of the jet η is around 98%. The trigger efficiency is measured also as a function

Table 6.2: 2016 collision data samples used for trigger efficiency measurement.

Dataset	Runs	$\mathcal{L}$ (pb^{-1})
SingleMuon_Run2016B_03Feb2017_ver2_v2[†]	273150–275376	5746.01
SingleMuon_Run2016C_03Feb2017_v1[†]	275420–276283	2572.903
SingleMuon_Run2016D_03Feb2017_v1[†]	276315–276811	4242.292
SingleMuon_Run2016E_03Feb2017_v1[†]	276824–277420	4024.47
SingleMuon_Run2016F_03Feb2017_v1[†]	277816–278800	2697.733
SingleMuon_Run2016F_03Feb2017_v1	278801–278808	406.776
SingleMuon_Run2016G_03Feb2017_v1	278816–280385	7575.824
SingleMuon_Run2016H_03Feb2017_v2	281207–284035	8434.663
SingleMuon_Run2016H_03Feb2017_v3	271036–284044	215.965
Total integrated luminosity		35916.636

Table 6.3: The simulated samples used for trigger efficiency measurement. Below X is shorthand notation for `13TeV_amcatnlo_pythia8`. The LO and NNLO indicate the order at which the theoretical cross-section is calculated.

Process	Cross section (pb)	Events	Dataset
	POWHEG +PYTHIA8 (NNLO), CUETP8M1 [134]		
$t\bar{t}$	831.76	77 081 156	TT_TuneCUETP8M2T4_X
	PYTHIA8 (LO), CUETP8M1 [134]		
QCD, $15 < p_T < 20$ GeV	3.625×10^6	4 141 251	/QCD_Pt-15to20_MuEnrichedPt5_TuneCUETP8M1_X
QCD, $20 < p_T < 30$ GeV	3.153×10^6	31 475 157	/QCD_Pt-20to30_MuEnrichedPt5_TuneCUETP8M1_X
QCD, $30 < p_T < 50$ GeV	1.652×10^5	29 954 815	/QCD_Pt-30to50_MuEnrichedPt5_TuneCUETP8M1_X
QCD, $50 < p_T < 80$ GeV	4.487×10^5	19 806 915	/QCD_Pt-50to80_MuEnrichedPt5_TuneCUETP8M1_X
QCD, $80 < p_T < 120$ GeV	1.052×10^5	23 584 215	/QCD_Pt-80to120_MuEnrichedPt5_TuneCUETP8M1_X
QCD, $120 < p_T < 170$ GeV	2.549×10^4	8 042 721	/QCD_Pt-120to170_MuEnrichedPt5_TuneCUETP8M1_X
QCD, $170 < p_T < 300$ GeV	8.644×10^3	17 350 231	/QCD_Pt-170to300_MuEnrichedPt5_TuneCUETP8M1_X
QCD, $300 < p_T < 470$ GeV	7.967×10^2	48 995 686	/QCD_Pt-300to470_MuEnrichedPt5_TuneCUETP8M1_X
QCD, $470 < p_T < 600$ GeV	7.920×10^1	19 362 943	/QCD_Pt-470to600_MuEnrichedPt5_TuneCUETP8M1_X
QCD, $600 < p_T < 800$ GeV	2.525×10^1	9 981 311	/QCD_Pt-600to800_MuEnrichedPt5_TuneCUETP8M1_X
QCD, $800 < p_T < 1000$ GeV	4.724	19 767 439	/QCD_Pt-800to1000_MuEnrichedPt5_TuneCUETP8M1_X
QCD, $1000 < p_T < $ Inf GeV	1.619	13 599 938	/QCD_Pt-1000toInf_MuEnrichedPt5_TuneCUETP8M1_X
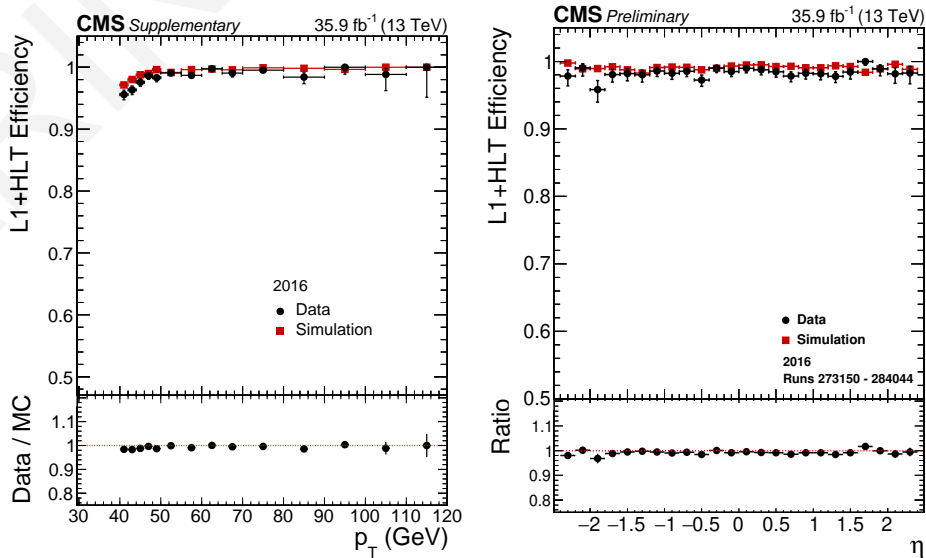
Figure 6.1: The efficiency of the logical OR of all trigger paths as a function of the p_T (left) and η (right) of the 6th leading in p_T jet.
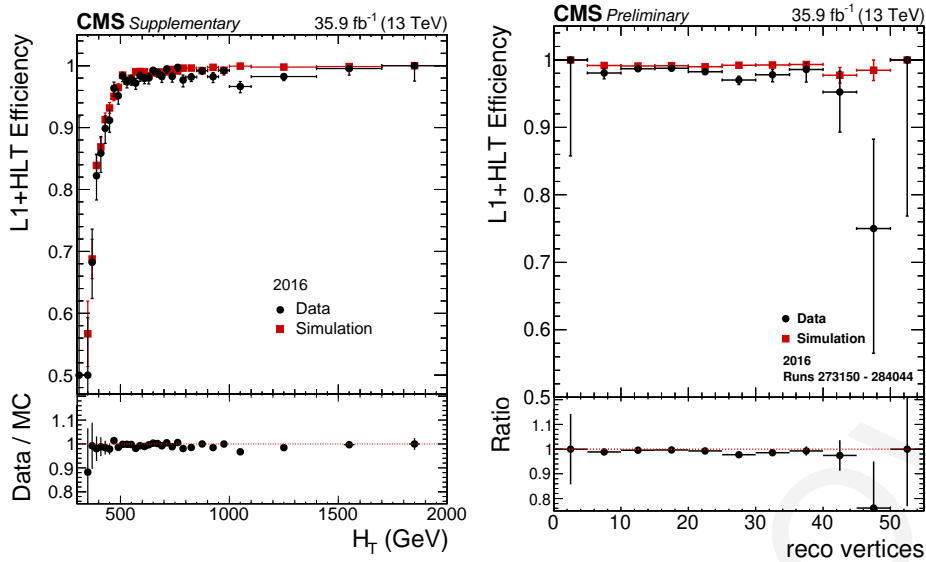
Figure 6.2: The efficiency of the logical $\mathtt{OR}$ of all trigger paths as a function of the H_{T} (left) and number of reconstructed vertices (right).

of the event H_{T} and the number of reconstructed vertices and is illustrated in Fig. 6.2. A plateau is reached for $H_{\mathrm{T}} \geq 500$ GeV and no dependence on the number of reconstructed vertices (pileup) is observed. Since there are no major differences between data and simulation efficiencies, a trigger scale factor is not needed. Instead, a conservative 5% systematic uncertainty for the trigger measurement is adopted.

6.2 Offline event selections

Following the online event selection, an offline event selection workflow is designed such that a maximal signal-to-background significance is achieved. Events are required to satisfy the criteria defined in the *baseline event selection*, described in section 5.4. The signature of these events is the presence of at least three b jets, which, in combination with the rest of the hadronic jets, can be used to reconstruct the two top quark candidates. To achieve this, a top quark tagging identification technique is exploited. The top quark tagger, described in section 6.2.1, is the key ingredient of this analysis as it constitutes a handle to reduce the QCD multijet background and the combinatorial background.

6.2.1 top quark tagging

A multivariate Boosted Decision Tree (BDT) is used to reconstruct top quarks resolved as three separate small cone jets, out of which one is b tagged and the remaining two jets are arising from the W boson decay. The BDT is using the gradient boost (BDTG) classifier and is trained on simulated $\mathrm{t\bar{t}}$ events with the TMVA package [150]. The *signal objects* are defined as three small cone jets combinations, where each individual jet is *matched* to the decay products of a top quark at generator level. A three-jet object is considered as matched

81
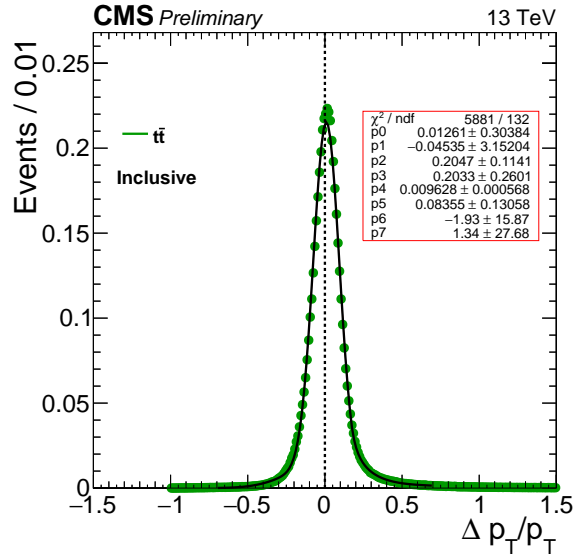
Figure 6.3: The fitted $\Delta p_T(j,q)/p_{T,q}$ distribution between jets with p_T above 30 GeV matched with quarks at generator level with $\Delta R < 0.3$, for simulated $t\bar{t}$ events. The fitting is performed with a convolution of a Gaussian and a Crystal-Ball function.

when satisfying: $\Delta R(j,q) < 0.3$ and $\Delta p_T(j,q)/p_{T,q} < 0.32$. The latter value corresponds to 2σ on the $\Delta p_T(j,q)/p_{T,q}$ distribution, when the jet and the quark are within $\Delta R < 0.3$, as shown in Fig. 6.3. Trijet combinations where at least one jet is not matched to a top quark decay product are considered as *background objects*. The BDT utilizes a total of 19 input variables, listed in Table 6.4. The variables represent the properties of the top quark and its decay products, such as masses, angular separations, kinematic and shape properties, and flavor discriminators for each of the jets. A detailed description of all the input variables can be found in [151].

The separation power of the resolved top quark tagging algorithm for signal and background objects is shown in Fig. 6.4 (left). Background objects have a BDTG response close to -1, while signal objects show a peak close to +1. Fig. 6.4 (right) shows the signal and background efficiencies and purities as a function of the BDTG score. The top quark tagging working point used in this analysis is BDTG > 0.4, and corresponds to signal and background object efficiency of 92% and 6%, respectively. The top quark tagging misidentification probability and efficiency as a function of the top candidate p_T for different values of the charged Higgs boson mass are shown in Fig. 6.5. Each of the selected b tagged jets in the event is paired with all two-jet combinations to create a three-jet object. To reduce the combinatorial background, three-jet objects with invariant mass greater than 400 GeV are rejected. The BDTG score of all three-jet objects are sorted in descending BDTG order and are cross-cleaned so there are no shared jets among them. Three-jet objects that satisfy the criterion BDTG > 0.4 are considered as *top quark candidates*. The events are required to have two cleaned top quark candidates.

Table 6.4: List of the input variables used for the resolved top quark tagging training [151]. The b-index refers to the b tagged jet and leading (sub-leading) indices refer to the leading (sub-leading) in p_T jet from the W boson decay.

#	Input variables	Description
1	m_{top}	Top candidate mass
2	$p_T \Delta R_t$	The product of the top candidate p_T with the ΔR between the W and b tagged jet
3	m_W	W mass
4	$p_T \Delta R_W$	The product of the hadronic W candidate p_T with the ΔR between the two constituent jets
5	Leading jet CSVv2	CSVv2 discriminator of the leading jet in the dijet system
6	Subleading jet CSVv2	CSVv2 discriminator of the subleading jet in the dijet system
7	b tagged jet CSVv2	CSVv2 discriminator of the b tagged jet
8	m_b	b tagged jet mass
9	$m_{(b,\, ldg\, jet)}$	Mass of the b jet and the leading jet system
10	$m_{(b,\, subldg\, jet)}$	Mass of the b jet and the subleading jet system
11	Leading jet CvsL	Charm-to-light discriminator of the leading jet in the dijet system
12	Subleading jet CvsL	Charm-to-light discriminator of the subleading jet in the dijet system
13	*softdrop* [152]	$\frac{min(p_{T1},p_{T2})}{p_{T1}+p_{T2}}\Delta R_{j1,j2}^{-2}$, j_1 (j_2) are jets from W decay, with $p_T = p_{T,1}$ $p_{T,2}$
14	Leading jet $p_T D$	Fragmentation function of the leading jet in the dijet system
15	Subleading jet $p_T D$	Fragmentation function of the subleading jet in the dijet system
16	Leading jet axis2	Jet shape variable describing the leading jet's short axis
17	Subleading jet axis2	Jet shape variable describing the subleading jet's short axis
18	Leading jet mult	The multiplicity of jet constituents of the leading jet in the dijet system
19	Subleading jet mult	The multiplicity of jet constituents of the subleading jet in the dijet system
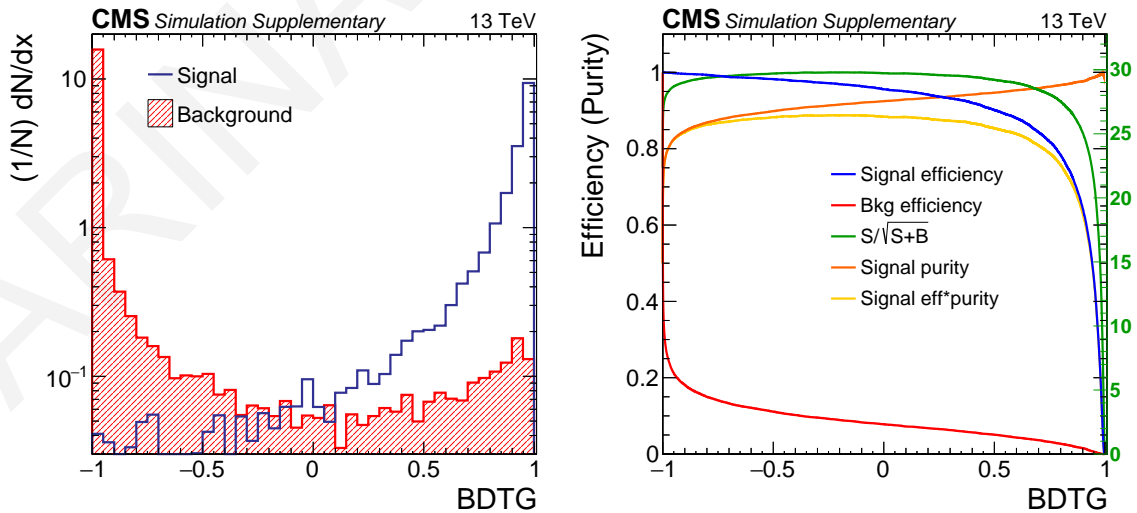
Figure 6.4: The BDT response for signal and background top candidates (*left*). The signal and background top quark tagging efficiency, purity and significance as a function of the working point of the BDT score (*right*).
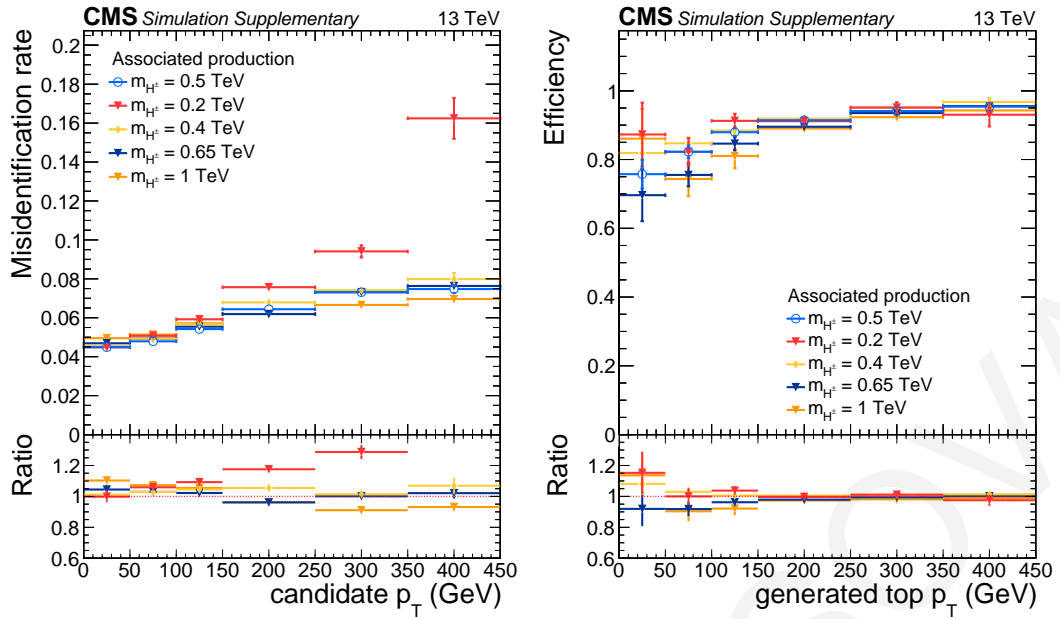
Figure 6.5: The top quark tagging misidentification rate (*left*) and efficiency (*right*) for several values of the $H^{\pm}$ mass. The ratio plot for each histogram is with reference to the rates for $m_{H^{\pm}} = 0.5$ TeV.

6.2.2 $H^{\pm}$ **candidate reconstruction**

The charged Higgs candidate is reconstructed from the four-momenta of the leading in p_T top quark candidate between the two cleaned tops and the leading in p_T free b tagged jet not used in the reconstruction of the two top candidates. The p_T and mass of the leading in p_T top candidate used for the $H^{\pm}$ reconstruction are shown in Fig. 6.6 for five signal mass points, QCD multijet and $t\bar{t}$ events. From the p_T distribution of the leading in p_T top candidate, it is clear that signals corresponding to charged Higgs bosons with low masses overlap with the QCD multijet and $t\bar{t}$ backgrounds and are, thus, expected to have lower sensitivity. The mass of the leading in p_T top candidate is centered around the top quark mass value. Due to the mass sculpting effect of the BDT tagger, the QCD multijet background mimics the top quark signal. The p_T of the free b tagged jet and of the reconstructed $H^{\pm}$ candidate are shown in Fig. 6.7 left and right, respectively. Charged Higgs bosons of higher mass are expected to have more energetic b jets, as illustrated in the left plot of Fig. 6.7. The increasing number of events near the beginning of the distribution for signals of high mass (e.g. $m_{H^{\pm}} = 1$ TeV) is a consequence of the mis-reconstruction and mis-identification of the top quark. This is expected since as the charged Higgs boson mass increases, its decay products become boosted and the subsequent decay products of the top quark are merged and cannot be identified as individual jets. As mentioned in the previous chapter, a dedicated analysis is performed to target events with Lorentz-boosted objects.

The resulting invariant mass distribution of the reconstructed charged Higgs candidates is illustrated in Fig. 6.8. The resolution of the $H^{\pm}$ invariant mass is studied for signal mass points 300 and 500 GeV (shown in Fig. 6.9) and 800 and 1000 GeV (shown in Fig. 6.10).
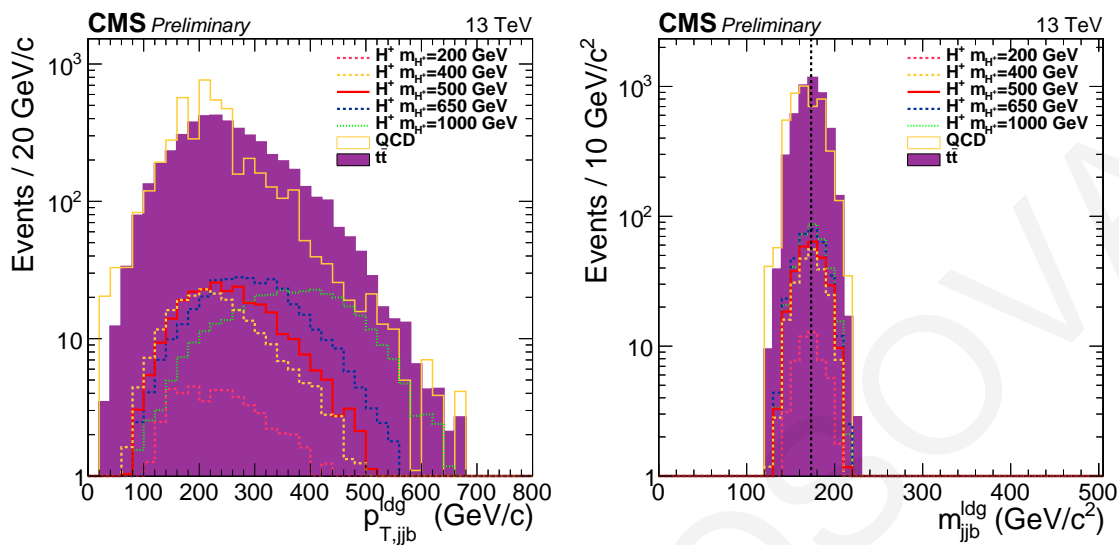
Figure 6.6: The transverse momentum (left) and mass (right) distributions of the leading in p_{T} top quark candidate.
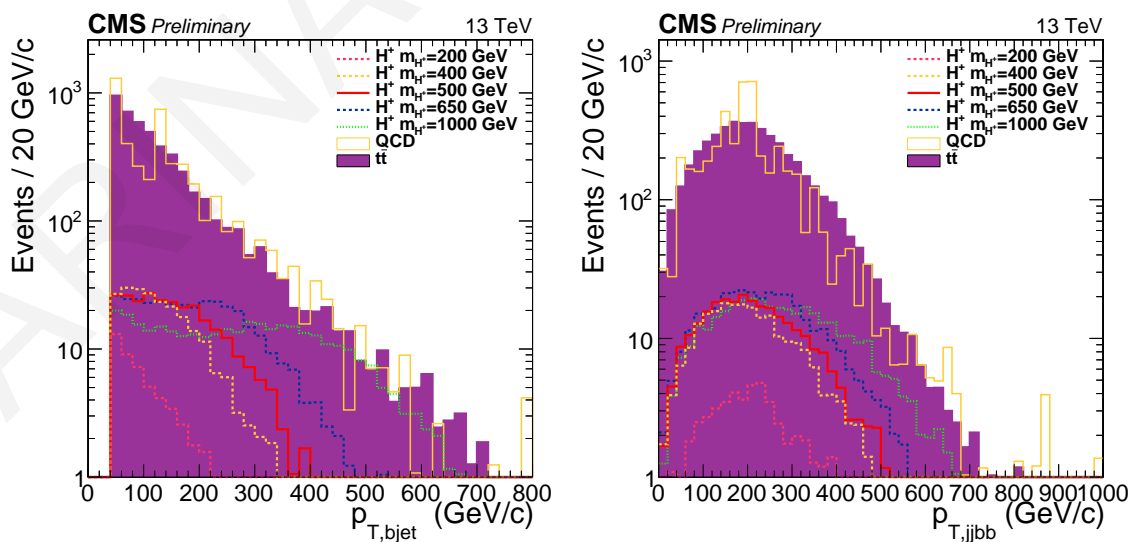
Figure 6.7: The transverse momentum distributions of the leading in p_{T} free b tagged jet (left) and of the reconstructed charged Higgs candidate (right).

85
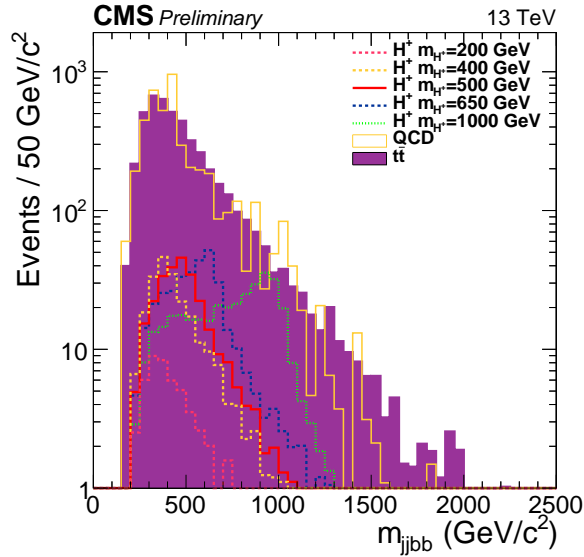
Figure 6.8: The invariant mass of the reconstructed $\mathrm{H}^{\pm}$ candidate.

The *inclusive* distribution, shown in green, corresponds to the invariant mass distribution of all charged Higgs boson candidates passing all analysis selections. It can be decomposed into four categories: *b-jet match*, *top match*, *full match* and *Combinatorics*. The *b-jet match* (*top match*) category corresponds to the case in which the leading in p_{T} free b jet (top[1]) used in the reconstruction of the charged Higgs boson candidate is the correct one. The category *full match* refers to the case where both objects are correctly identified, while the category *Combinatorics* refers to the case where none of the jets is correctly identified as the decay product of the charged Higgs boson. The mass shape is dominated by failure to identify the correct combination for top and b tagged jet, which indicates that the main contributor to the mass smearing is of combinatorial nature. The peak of the $\mathrm{H}^{\pm}$ candidate invariant mass appears to be slightly shifted with respect to its true value due to the mismeasurement of the energy of the jets, which constitutes a lower order effect in the failure of the correct charged Higgs boson reconstruction. The mass width of the charged Higgs boson is defined as the Full Width at Half Maximum (FWHM) of the *full match* invariant mass distribution. Figure 6.11 shows the mass width as a function of the signal mass, which increases with the mass of the charged Higgs boson. The width is shown for signal masses up to $800\,\mathrm{GeV}$ as the reconstruction efficiency of charged Higgs bosons with higher mass decreases, as expected, mainly due to the boosted nature of its decay products.

6.3 Data quality filtering and simulation corrections

Quality requirements are applied to remove events with unforeseen detector effects or electronic noise. Simulated events are additionally corrected with *scale factors* to match the behavior of the data. The scale factors are usually of the order of a few percent and are com-

---

[1]all three jets constructing the top candidate are truth-matched to the top quark decay products.
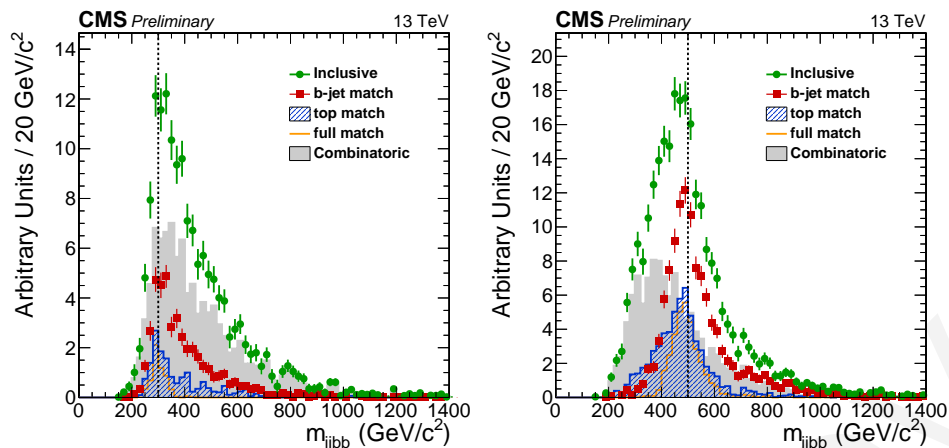
Figure 6.9: The resolution of the charged Higgs boson candidate invariant mass for signal mass points 300 GeV (left) and 500 GeV (right).
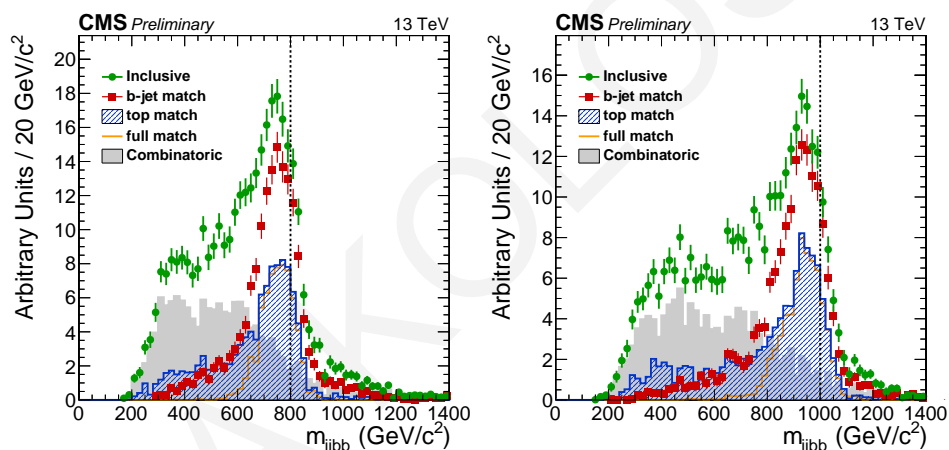
Figure 6.10: The resolution of the charged Higgs boson candidate invariant mass for signal mass points 800 GeV (left) and 1000 GeV (right).
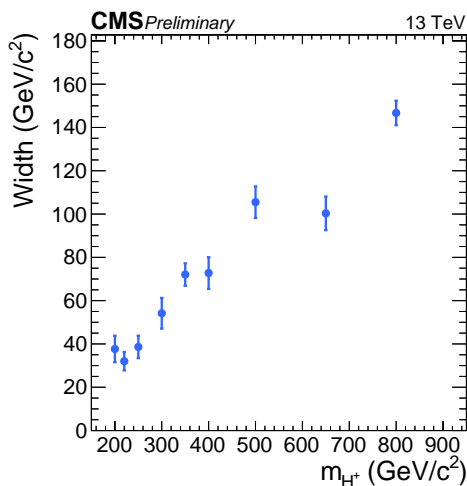
Figure 6.11: The reconstructed invariant mass width as a function of the signal mass

puted as ratios of reconstruction or identification efficiencies between data and simulation.

p_T^{miss} quality filters

The missing transverse momentum (p_T^{miss}) distribution differs between data and simulation due to the presence of anomalous high-p_T^{miss} events, mainly originating from reconstruction failures and detector noise. These events are identified and suppressed by dedicated algorithms utilizing timing, pulse shape, and topology information of the spurious signals from the different subdetectors. The algorithms have two modes: *noise filtering* and *event filtering*. The former removes the anomalous energy from the event reconstruction while the latter removes the event from the data set. In the HCAL, filters remove noise coming from the hybrid photodiode (HPD) and the readout box (RBX) electronics. In addition, spurious energy can arise from direct particle interactions with the light guides and photomultiplier tubes of the HF. These are identified and removed from the reconstruction with isolation-based noise filters by comparing energy deposits from the HCAL and ECAL with measurements from the tracker. In the ECAL, most of the electronics noise and artificial signals from particle interactions with the photodetectors are eliminated during reconstruction. However, high-p_T^{miss} events can remain due to anomalously high pulses in the supercrystals or due to the lack of information for channels with non-functioning readout electronics. These are removed through dedicated ECAL noise filters. Moreover, beam halo particles traveling parallel to the collision axis can interact with the calorimeters and leave energy deposits, or with the CSC subdetector. Dedicated filters combining information from both the CSC and the calorimeters are used to reject these machine-induced signals. Poor reconstruction of muons during the muon-tracking iteration step can also lead to high-p_T^{miss}. Tracks with high-p_T but low-quality reconstruction can contribute to the p_T^{miss} either as a poorly reconstructed PF muon or PF charged hadron. Dedicated algorithms are used to reject events with such poorly reconstructed muons or charged hadrons.

Simulation corrections

- **Pileup:** Simulation samples are generated with a fixed average number of pileup interactions, which does not necessarily cover the conditions for each data-taking period. In real data, pileup is not fixed but changes as a function of the instantaneous luminosity. To reproduce the pileup distribution in data, simulated events are corrected based on the measured luminosity profile and average measured total inelastic cross-section [97].

- **b tagging efficiency and b-mistagging rate**:

  The b tagging identification efficiency and the misidentification probability to identify a non-b jet as a b jet are slightly different between data and simulation. To account for these differences, a suitable SF is calculated for each event and applied to the

simulation. The *per-event* SF is calculated by applying a *per-jet* SF to each jet in the event. The *per-jet* SFs, provided by the CMS b-tag & vertexing POG, have been measured as a function of the jet p_T, η, and the underlying flavor of the jet at generator level.

The *per-jet* SFs used in this analysis are derived based on μ+jets events. The dependence on jet $|\eta|$ is small and can be neglected. The difference between the tagging efficiency and mistagging probability in the data and simulation is corrected by taking into account the *per-jet* data-to-simulation scale factors f_tag and f_mistag,

$$
\begin{aligned}
f_\mathrm{tag}(p_\mathrm{T}) &= \frac{\varepsilon_\mathrm{tag}^{Data}(p_\mathrm{T})}{\varepsilon_\mathrm{tag}^{MC}(p_\mathrm{T})}, \\
f_\mathrm{mistag}(p_\mathrm{T}) &= \frac{\varepsilon_\mathrm{mistag}^{Data}(p_\mathrm{T})}{\varepsilon_\mathrm{mistag}^{MC}(p_\mathrm{T})}
\end{aligned}
\tag{6.2}
$$

where ε_tag and $\varepsilon_\mathrm{mistag}$ are the b tagging efficiency and mistagging probability for a jet in the event for a given b tagging working point. These quantities are then used to calculate the probability of the event to pass the b tagging selection as:

$$
\begin{aligned}
\mathrm{P} = &\prod_{i=1}^{N_\text{b-flavor jets tagged}} \varepsilon_{\mathrm{tag},i} \prod_{j=1}^{N_\text{b-flavor jets not tagged}} \left(1 - \varepsilon_{\mathrm{tag},j}\right) \\
&\times \prod_{k=1}^{N_\text{light flavor jets tagged}} \varepsilon_{\mathrm{mistag},k} \prod_{l=1}^{N_\text{light flavor jets not tagged}} \left(1 - \varepsilon_{\mathrm{mistag},l}\right).
\end{aligned}
\tag{6.3}
$$

The *per-event* SF for re-weighting simulated events can then be calculated as

$$
\mathrm{SF} = \frac{P(\mathrm{Data})}{P(\mathrm{MC})}.
\tag{6.4}
$$

Using 6.2 and 6.3, the *per-event* SF used to correct the simulated events can be written as:

$$
\begin{aligned}
\mathrm{SF} = &\prod_{i}^{N_\text{b,c tagged}} f_{\mathrm{tag},i} \prod_{j}^{N_\text{b,c not tagged}} \left(\frac{1 - f_{\mathrm{tag},j}\varepsilon_{\mathrm{b,c},j}}{1 - \varepsilon_{\mathrm{b,c},j}} \right) \\
&\times \prod_{k}^{N_\text{uds,g tagged}} f_{\mathrm{mistag},k} \prod_{l}^{N_\text{uds,g not tagged}} \left(\frac{1 - f_{\mathrm{mistag},l}\varepsilon_{\mathrm{uds,g},l}}{1 - \varepsilon_{\mathrm{uds,g},l}} \right).
\end{aligned}
\tag{6.5}
$$

The tagging efficiencies (mistagging probabilities) are sensitive to the final state kinematics and are recalculated from simulated $\mathrm{t\bar{t}}$ (QCD multijet) events for each jet flavor and as a function of the jet p_T. The simulated events are required to satisfy the baseline
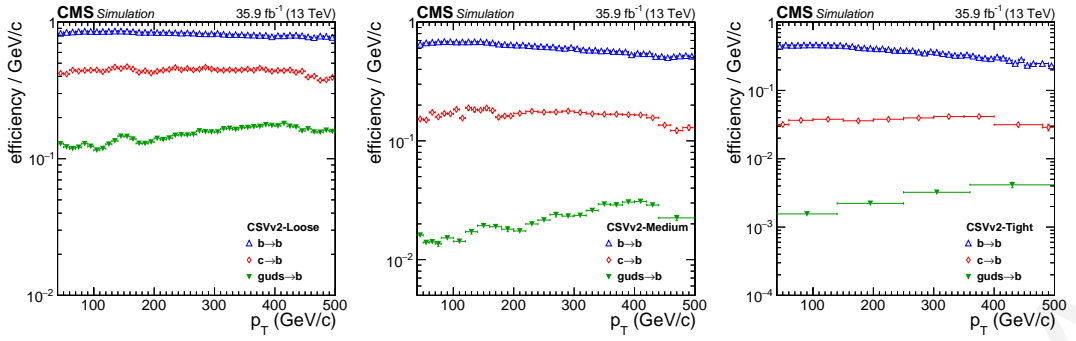
Figure 6.12: The simulated tagging efficiencies and mistagging probabilities for different flavors of jets for the CSVv2 b tagging algorithm and for the *Loose* (left), *Medium* (middle) and *Tight* working points (right).

analysis selections up to the jet selection. Figure 6.12 shows the measured tagging efficiencies and mistagging probabilities for the CSVv2 algorithm and the *Loose*, *Medium* and *Tight* working points.

- **top quark tagging efficiency and misidentification rate**: The top quark tagging performance in data and simulated events is slightly different, and thus, simulated events are corrected in a way similar to what is done for b tagging. Instead of the b jet flavor, top candidates are classified based on their tagging status and whether they are truth-matched when using generator-level information. This leads to four orthogonal categories:

  1. tagged and truth-matched (t|gen-t)

  2. tagged and unmatched (t|!gen-t)

  3. not tagged and truth-matched (!t|gen-t)

  4. not tagged and unmatched (!t|!gen-t).

  The simulated tagging efficiency and misidentification rates are denoted as ε_i (t|gen-t) and ε_i (t|!gen-t) respectively, and depend on the p_T of the top candidate with index i. The estimation of the top quark tagging efficiency and misidentification rate are discussed in section 6.5. Simulated events are reweighed by a *per-event* SF, which is calculated from the *per-candidate* SFs applied to all cross-cleaned top candidates. The probability of a simulated event to pass the top quark tagging selection is given by the product of the individual efficiencies of all the cross-cleaned top candidates falling into the four aforementioned categories:

90

$$P_{\text{sim}} = \prod_{i=1}^{\text{t|gen-t}} \varepsilon_i \left(\text{t|gen-t}\right) \times \prod_{j=1}^{\text{t|!gen-t}} \varepsilon_j \left(\text{t|!gen-t}\right)$$

$$\times \prod_{k=1}^{\text{!t|gen-t}} \left[1 - \varepsilon_k \left(\text{t|gen-t}\right)\right] \times \prod_{l=1}^{\text{!t|!gen-t}} \left[1 - \varepsilon_l \left(\text{t|!gen-t}\right)\right]. \tag{6.6}$$

The corresponding probability in the data as a function of the *per−candidate* SF is given by:

$$P_{\text{data}} = \prod_{i=1}^{\text{t|gen-t}} \text{SF}_i \left(\text{t|gen-t}\right) \varepsilon_i \left(\text{t|gen-t}\right) \times \prod_{j=1}^{\text{t|!gen-t}} \text{SF}_j \left(\text{t|!gen-t}\right) \varepsilon_j \left(\text{t|!gen-t}\right)$$

$$\times \prod_{k=1}^{\text{!t|gen-t}} \left[1 - \text{SF}_k \left(\text{t|gen-t}\right) \varepsilon_k \left(\text{t|gen-t}\right)\right] \times \prod_{l=1}^{\text{!t|!gen-t}} \left[1 - \text{SF}_l \left(\text{t|!gen-t}\right) \varepsilon_l \left(\text{t|!gen-t}\right)\right].$$

$$\tag{6.7}$$

The event weight to correct the simulation is given as an expression of products of the *per-candidate* SF and tagging efficiency and misidentification rate:

$$w = \frac{P_{\text{data}}}{P_{\text{sim}}} = \prod_{i=1}^{\text{t|gen-t}} \text{SF}_i \left(\text{t|gen-t}\right) \times \prod_{j=1}^{\text{t|!gen-t}} \text{SF}_j \left(\text{t|!gen-t}\right)$$

$$\times \prod_{k=1}^{\text{!t|gen-t}} \frac{\left[1 - \text{SF}_k \left(\text{t|gen-t}\right) \varepsilon_k \left(\text{t|gen-t}\right)\right]}{\left[1 - \varepsilon_k \left(\text{t|gen-t}\right)\right]} \times \prod_{l=1}^{\text{!t|!gen-t}} \frac{\left[1 - \text{SF}_l \left(\text{t|!gen-t}\right) \varepsilon_l \left(\text{t|!gen-t}\right)\right]}{\left[1 - \varepsilon_l \left(\text{t|!gen-t}\right)\right]}.$$

$$\tag{6.8}$$

As discussed in section 6.2.1, since the event selection requires the presence of exactly two top candidates, the last two terms of the above equation can be ignored.

6.4 Background measurement

The main background for this search arises from QCD multijet and EWK processes, mainly coming from top quark pair production ($t\bar{t}$) in association with additional jets. Contributions from more rare processes, such as single-t, Z/γ^*, $t\bar{t} + X$, diboson (WZ, WW, ZZ) and $tt\bar{t}\bar{t}$ are found to be small. Due to its large cross-section, QCD multijet is the most dominant background, but it is nevertheless reducible by employing b and top quark tagging techniques. The $t\bar{t}$ in association with additional jets background, however, remains largely irreducible as its final state can be identical to the signal final state.

The measurement of the background is based on the *hadron-flavor* hypothesis, in which
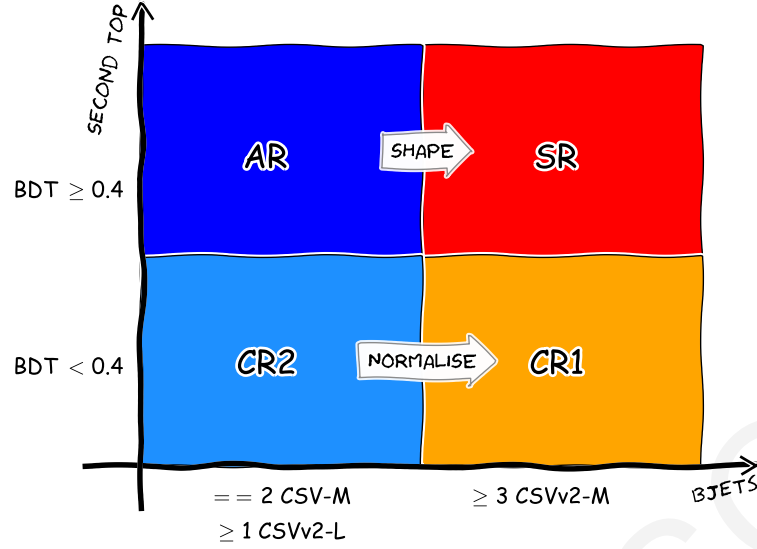
Figure 6.13: Schematic diagram of the *ABCD* method used. The x-axis corresponds to the b jet selection, while the y-axis to the top quark tagging BDT score of the subleading in BDT top quark candidate.

a b jet is defined as a *genuine b jet* if its hadron flavor at generator level is consistent with a b hadron. Background events can then be decomposed into events containing genuine b jets (labelled as *Genuine b jets* events) or events where at least one light quark or gluon jet is erroneously tagged as b jet (labelled as *misidentified b jets* events). The former is modeled using simulation and the latter is measured with a data-driven technique using control regions that are defined by inverting the b tagging and top quark tagging requirements. The three orthogonal Control Regions (CRs), together with the Signal Region (SR), are illustrated in the diagram depicted in Fig. 6.13. The CR1 is defined by inverting the top quark tagging BDT score of the subleading in BDT top candidate (BDT $<$ 0.4), while the Application Region (AR) is defined by inverting the b tagging selection from at least three medium b tagged jets to exactly two medium b tagged jets and at least one loose but not medium b tagged jet. The b jet that is being inverted is set to be the leading in p_{T} free b jet, namely the b jet coming from the charged Higgs boson decay. The CR2 is defined by inverting both b tagging and top quark tagging selections. The shape of the final fit discriminant, which is the reconstructed invariant mass of the charged Higgs boson candidate, is obtained from the AR. The CR1 and CR2 are used to extract *transfer factors* to normalize the AR to the SR. Since the SR and the associated CRs are mutually exclusive, the expected yield of misidentified b jet events in the SR can be predicted with the formula:

$$N^{\mathrm{SR}} = N^{\mathrm{AR}} \times \frac{N^{\mathrm{CR1}}}{N^{\mathrm{CR2}}}, \tag{6.9}$$

where N^{AR}, N^{CR1} and N^{CR2} are the number of misidentified b jet events in the CRs. The *transfer factors* are extracted in bins of the inverted b jet p_{T} and $|\eta|$ in order to compensate
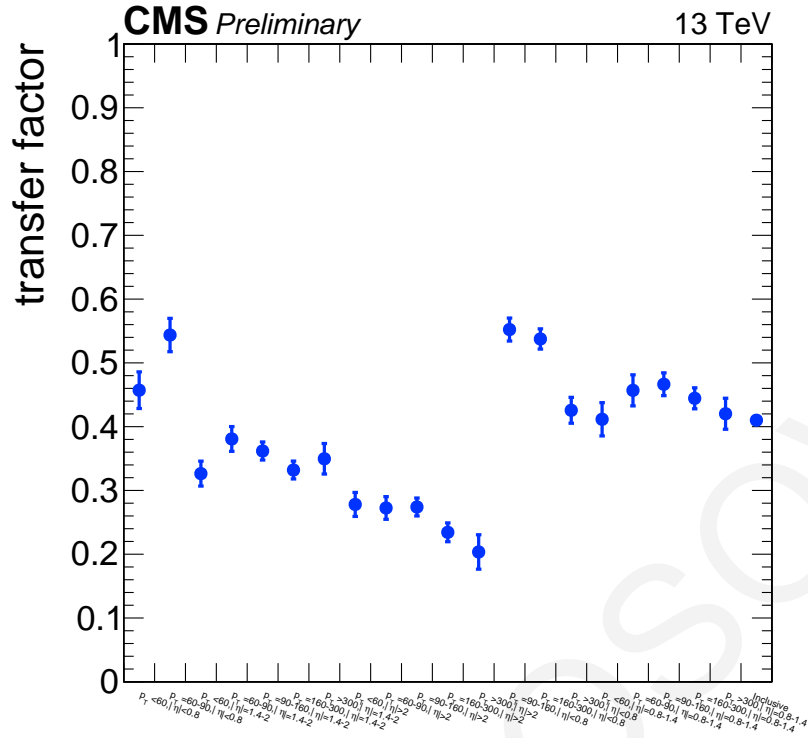
Figure 6.14: The *transfer factors* in bins of the leading in p_T free b jet p_T and $|\eta|$.

for kinematic differences between loose- and medium-tagged b jets. There are 5 p_T and 4 $|\eta|$ bins:

- p_T: < 60, 60-90, 90-160, 160-300, > 300 GeV

- $|\eta|$: < 0.8, 0.8-1.4, 1.4-2.0, > 2.0

leading to a total of 20 bins where the measurement is performed. Figure 6.14 shows the transfer factor values for each of the 20 bins used. The effect of performing the measurement in bins of the p_T and $|\eta|$ of the leading in p_T free b jet is shown in Fig. 6.15. The number of misidentified b jets events in the CRs is estimated by subtracting the EWK *Genuine b jets* events from the data:

$$N^{\text{misidentified b jets}} = N^{\text{data}} - N^{\text{EWK Genuine b jets}}, \tag{6.10}$$

where the term EWK *Genuine b jets* events corresponds to the simulated EWK events where all b tagged jets in the event are *genuine b jets*. To summarize, the estimated number of misidentified b jets events expected in the SR can be expressed as:

$$N_{\text{SR}}^{\text{misidentified b jets}} = \sum_{i}^{\text{inverted b jet } p_\mathrm{T}, \, |\eta| \text{ bins}} \left(N_{\text{AR},i}^{\text{data}} - N_{\text{AR},i}^{\text{EWK Genuine b jets}} \right) \times TF_i, \tag{6.11}$$

93
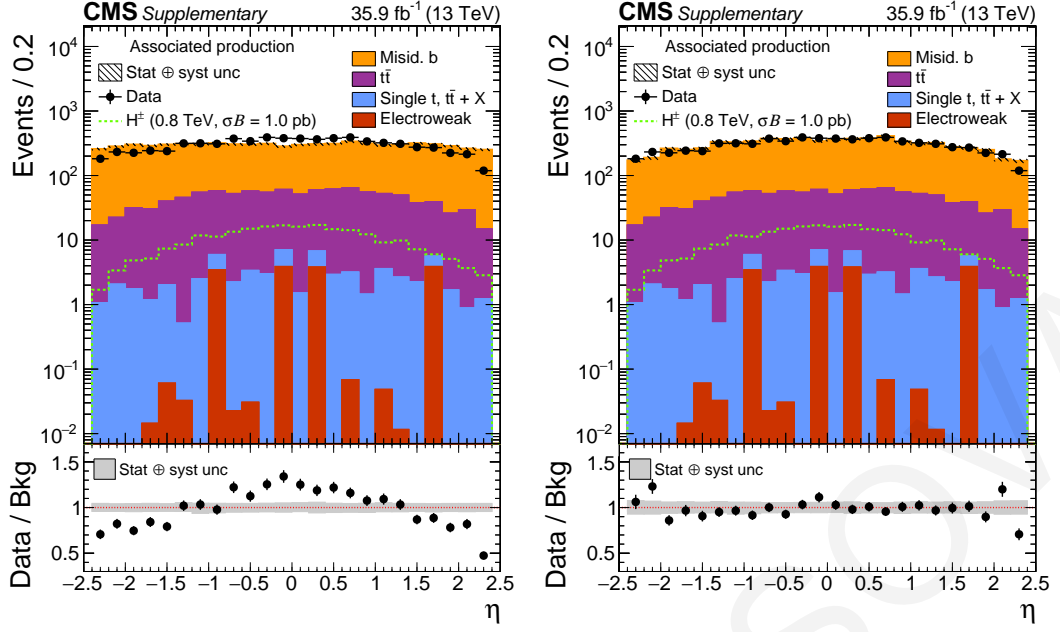
Figure 6.15: The leading in p_{T} free b jet η distribution without (left) and with (right) performing the background measurement in bins of the leading in p_{T} free b jet p_{T} and $|\eta|$.

where TF_i is equal to:

$$TF_i = \frac{N_{\mathrm{CR1},i}^{\mathrm{data}} - N_{\mathrm{CR1},i}^{\mathrm{EWK\ Genuine\ b\ jets}}}{N_{\mathrm{CR2},i}^{\mathrm{data}} - N_{\mathrm{CR2},i}^{\mathrm{EWK\ Genuine\ b\ jets}}}, \tag{6.12}$$

and the index i runs over all bins of the leading in p_{T} b jet p_{T} and $|\eta|$.

6.4.1 Purity of misidentified b jets events in the CRs

To ensure that the three control regions used in the estimation of the misidentified b jets background are rich in such events, the purity in misidentified b jet events is calculated by:

$$\mathrm{Purity}_{\mathrm{CR}} = \frac{N_{\mathrm{CR}}^{\mathrm{data}} - N^{\mathrm{EWK\ Genuine\ b\ jets}}}{N_{\mathrm{CR}}^{\mathrm{data}}}, \tag{6.13}$$

where $N_{\mathrm{CR}}^{\mathrm{data}}$ and $N^{\mathrm{EWK\ Genuine\ b\ jets}}$ are the number of data and simulated EWK Genuine b jets events passing the selections of each CR, respectively. Figures 6.16 and 6.17 show the purity of misidentified b jets events in the three CRs as a function of the p_{T} and invariant mass of the leading in p_{T} top and charged Higgs boson candidate, respectively.

6.4.2 Self-closure test

To validate the background estimation method used a self-closure test is performed, where the shapes of key distributions from the CRs using misidentified b jets events are compared
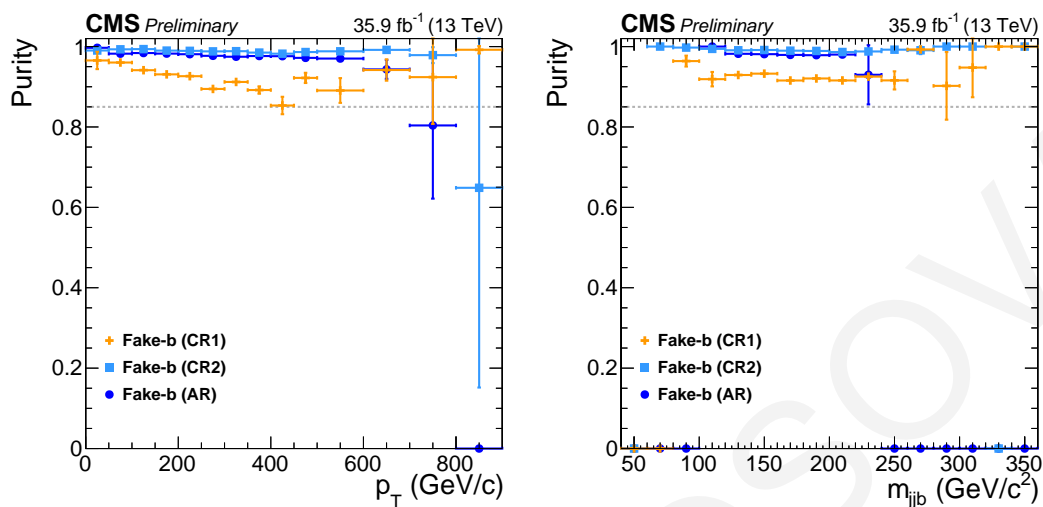
Figure 6.16: Purity of misidentified b jets events (Fake-b) as a function of the p_{T} (left) and invariant mass (right) of the leading in p_{T} top.
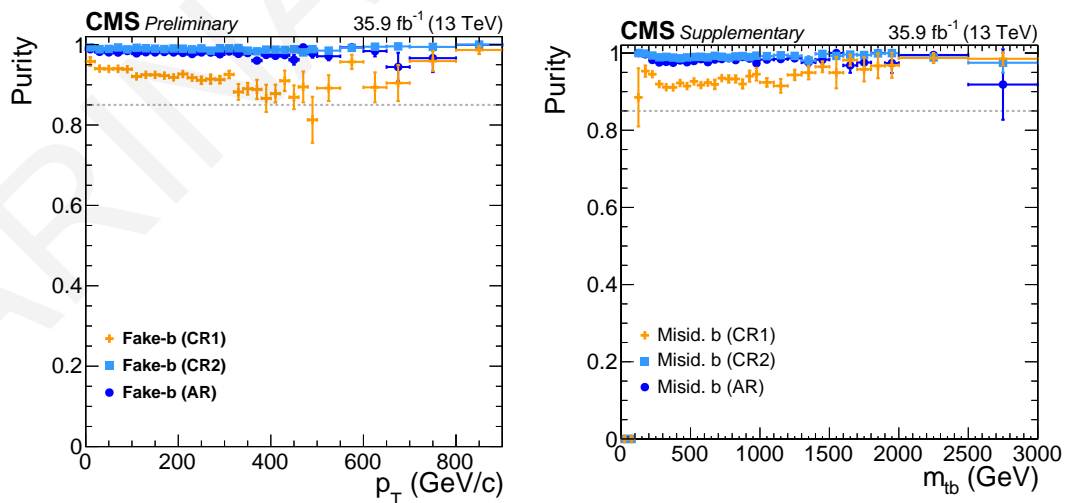
Figure 6.17: Purity of misidentified b jets events (Fake-b) as a function of the p_{T} (left) and invariant mass (right) of the reconstructed charged Higgs boson candidate.

95
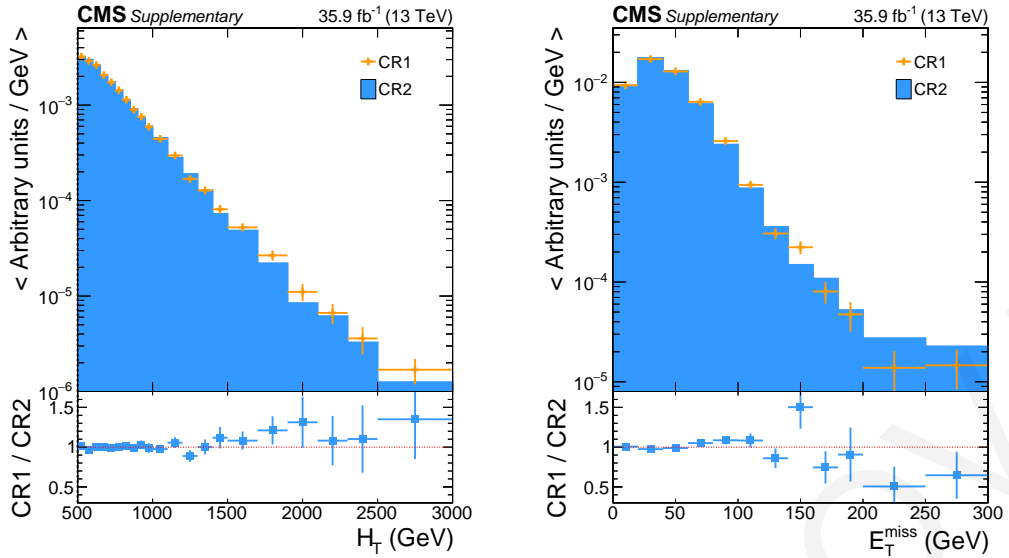
Figure 6.18: Comparison of the event H_T (left) and E_T^{miss} (right) distributions in the two CRs.

to the ones derived by subtracting the EWK Genuine b jets events from the data:

$$N_{CR2}^{\text{misidentified b jets}} = N_{CR2}^{\text{data}} - N_{CR2}^{\text{EWK Genuine b jets}}$$

$$N_{CR1}^{\text{misidentified b jets}} = \sum_{i}^{\text{inverted b jet } p_T, |\eta| \text{ bins}} N_{CR2}^{\text{misidentified b jets}} \times TF_i, \tag{6.14}$$

where TF_i are derived by Equation 6.12. The event E_T^{miss} and H_T are shown in Fig. 6.18, and the p_T and invariant mass distributions of the leading in p_T top candidate are shown in Fig. 6.19. All distributions are normalized to unity.

6.5 Systematic uncertainties

The systematic uncertainties affecting the yields (*rate* uncertainties) and the shapes (*shape* uncertainties) of the signal and background processes are discussed in this section. A list of all systematic uncertainties is summarized in Table 6.5. Their impact on the signal yield for the $H^{\pm}$ mass hypothesis of 500 GeV, and the expected background processes is also shown. Unless otherwise stated, systematic uncertainties affecting both signal and background are treated as 100% correlated between them.

6.5.1 Experimental uncertainties

- **Luminosity:** The systematic uncertainty of the luminosity, estimated in Ref. [153], is taken to be 2.5% and affects the rate of the signal and all background processes.

- **Pileup:** the systematic uncertainty from the pileup modeling is estimated by varying the total inelastic pp interaction cross section by $\pm 5\%$ its nominal value. Subsequently,
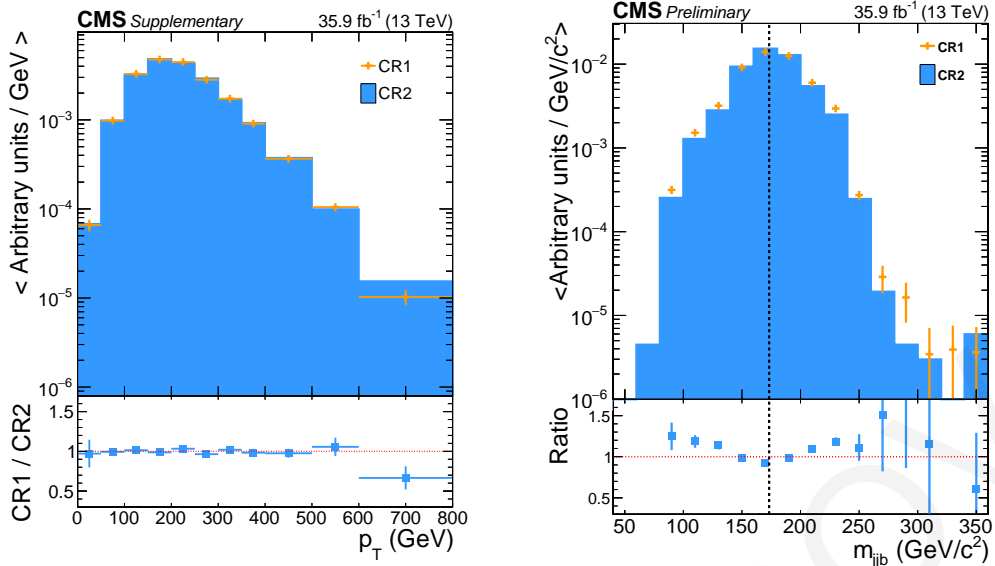
Figure 6.19: Comparison of the p_T (left) and invariant mass (right) distributions of the leading in p_T top candidate in the two CRs.

the pileup reweighing procedure is repeated and the invariant mass of the charged Higgs boson candidate is reconstructed.

- **Trigger efficiency:** The trigger efficiency is extracted from data and compared to the efficiency from the simulation. As discussed in section 6.1.1, no correction factors are applied to simulation, instead, a conservative 5% uncertainty affecting only the rates is applied.

- **Lepton and** τ_h **identification efficiencies:** The uncertainties on the identification of isolated electrons, muons, and hadronically decaying $\tau-$ leptons are calculated as:

$$
\begin{aligned}
\theta_e &= \frac{N^{\text{not passed e-veto}}}{N^{\text{passed e-veto}}} \times \Delta_{e-ID}, \\
\theta_\mu &= \frac{N^{\text{not passed μ-veto}}}{N^{\text{passed μ-veto}}} \times \Delta_{\mu-ID}, \\
\theta_{\tau_h} &= \frac{N^{\text{not passed } \tau_h-\text{veto}}}{N^{\text{passed } \tau_h-\text{veto}}} \times \Delta_{\tau_h-ID},
\end{aligned}
\tag{6.15}
$$

where N corresponds to the number of events failing or passing the identification selection (in this case the *veto*ing selection). The misidentification rates Δ_{e-ID}, $\Delta_{\mu-ID}$, and Δ_{τ_h-ID} are taken to be 2%, 1%, and 3%, respectively.

- **Jet energy scale:** The uncertainties related to the jet energy scale are obtained by shifting the jet energy scale applied to the selected jets by $\pm 1\sigma$ around the central value. These variations take into account a long list of systematic sources affecting the jet calibrations, as described in Ref. [154], and are parameterized as a function of the jet p_T and η. The events are re-analyzed and the variations are propagated to all the related reconstructed objects (jets, top quark candidates, and the charged Higgs boson

97

candidate).

- **Jet energy resolution:** The reconstructed jets in the simulation are smeared such that their p_T resolution agrees with the one observed in data. This is done by applying p_T-dependent correction factors that are accompanied by systematic uncertainties derived in Ref. [154]. For this analysis, the systematic uncertainties related to the jet energy resolution are measured in a similar way as in the jet energy scale. The jet energy resolution is shifted by $\pm 1\sigma$ around its central value, the events are re-analyzed and the variations are propagated into the final discriminant distribution.

- **b tagging efficiency and misidentification rate:** Uncertainties arising from the simulation-to-data correction for b tagging and b-mistagging are taken into account and are treated as two different shape nuisances in the analysis. The uncertainty on the *per−event* b tagging SF depends on the uncertainties of the *per−jet* SFs and b-(mis)tagging efficiencies for jets of different p_T and flavor. It is calculated in analytical form by applying the error propagation law to the equation (6.5), yielding:

$$
\begin{aligned}
\left(\frac{\Delta^{\pm}\mathrm{SF}}{\mathrm{SF}}\right)^2_j = &\left[\sum_i^{\mathrm{b\ tagged}}\frac{1}{f_{j,i}^{\mathrm{tag}}} - \sum_i^{\mathrm{b\ untagged}}\frac{\varepsilon_{j,i}^{\mathrm{b}}}{1-f_{j,i}^{\mathrm{tag}}\varepsilon_{j,i}^{\mathrm{b}}} + \sum_i^{\mathrm{c\ tagged}}\frac{1}{f_{j,i}^{\mathrm{tag}}} - \sum_i^{\mathrm{c\ untagged}}\frac{\varepsilon_{j,i}^{\mathrm{c}}}{1-f_{j,i}^{\mathrm{tag}}\varepsilon_{j,i}^{\mathrm{c}}}\right]^2 \\
&\times \left(\Delta^{\pm}f_j^{\mathrm{tag}}\right)^2 \\
&+\left[\sum_i^{\mathrm{uds,\ g\ tagged}}\frac{1}{f_{j,i}^{\mathrm{mistag}}} - \sum_i^{\mathrm{uds\ untagged}}\frac{\varepsilon_{j,i}^{\mathrm{uds}}}{1-f_{j,i}^{\mathrm{mistag}}\varepsilon_{j,i}^{\mathrm{uds}}} - \sum_i^{\mathrm{g\ untagged}}\frac{\varepsilon_{j,i}^{\mathrm{g}}}{1-f_{j,i}^{\mathrm{mistag}}\varepsilon_{j,i}^{\mathrm{g}}}\right]^2 \\
&\times \left(\Delta^{\pm}f_j^{\mathrm{mistag}}\right)^2 \\
&+\left[\sum_i^{\mathrm{b\ untagged}}\frac{1-f_{j,i}^{\mathrm{tag}}}{(1-f_{j,i}^{\mathrm{tag}}\varepsilon_{j,i}^{\mathrm{b}})(1-\varepsilon_{j,i}^{\mathrm{b}})}\right]^2 \times \left(\Delta^{\pm}\varepsilon_j^{\mathrm{b}}\right)^2 \\
&+\left[\sum_i^{\mathrm{c\ untagged}}\frac{1-f_{j,i}^{\mathrm{tag}}}{(1-f_{j,i}^{\mathrm{tag}}\varepsilon_{j,i}^{\mathrm{c}})(1-\varepsilon_{j,i}^{\mathrm{c}})}\right]^2 \times \left(\Delta^{\pm}\varepsilon_j^{\mathrm{c}}\right)^2 \\
&+\left[\sum_i^{\mathrm{uds\ untagged}}\frac{1-f_{j,i}^{\mathrm{mistag}}}{(1-f_{j,i}^{\mathrm{mistag}}\varepsilon_{j,i}^{\mathrm{uds}})(1-\varepsilon_{j,i}^{\mathrm{uds}})}\right]^2 \times \left(\Delta^{\pm}\varepsilon_j^{\mathrm{uds}}\right)^2 \\
&+\left[\sum_i^{\mathrm{g\ untagged}}\frac{1-f_{j,i}^{\mathrm{mistag}}}{(1-f_{j,i}^{\mathrm{mistag}}\varepsilon_{j,i}^{\mathrm{g}})(1-\varepsilon_{j,i}^{\mathrm{g}})}\right]^2 \times \left(\Delta^{\pm}\varepsilon_j^{\mathrm{g}}\right)^2,
\end{aligned}
$$

(6.16)

where the sums in the square brackets run over all the tagged or untagged jets of a given p_T bin, denoted with the index j. The uncertainty calculations are made with the following assumptions:

- The per−jet SFs f^{tag} and f^{mistag} are treated as uncorrelated
- The uncertainties of same-flavored jets in a given p_T bin are correlated, which leads to the sum running over the jet p_T bins.

– The uncertainties for the c jets are assumed to be twice as large as the uncertainties for the b jets.

- **top quark tagging efficiency and misidentification rate:**

  The top quark tagging efficiency and misidentification rate have been calculated in the data and compared to the simulation. The misidentification rate is estimated in a QCD-dominated sample where the events are required to pass the `HLT_PFHT900` trigger and have large offline H_T. Events containing leptons or τ_h-jets are rejected. The events are further required to have a large-cone jet (AK8) with $p_T > 150$ GeV and $|\eta| < 2.4$, and at least four small cone jets with $p_T > 40$ GeV and $|\eta| < 2.4$, out of which exactly one is tagged as a b jet with the Medium working point of the CSVv2 tagger. The top quark candidate, consisting of the b tagged jet and two other small cone jets, is required to be distant to the AK8 jet, by requiring $\Delta R(\text{top}, \text{AK8}) > 2.0$. To reject AK8 jets originating from boosted hadronically-decaying W bosons or boosted tops, the N-subjettiness [155] variable is used. The N-subjettiness, denoted by τ_N, is a jet shape variable that effectively *counts* the number of constituent jets (subjets) in a given large-cone jet. For a jet with N subjets, the N-subjettiness is defined as:

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min \left\{ \Delta R_{1,k}, \Delta R_{2,k}, ..., \Delta R_{N,k} \right\}, \tag{6.17}$$

  where k runs over all the constituent particles of the jet, and $\Delta R_{J,k}$ is the distance in the $\eta - \phi$ plane between a candidate subjet J and a particle k. The normalization factor d_0 is equal to $\sum_k p_{T,k} R_0$, where R_0 is the cone-size of the jet. The variable τ_{21} defined as the ratio τ_2/τ_1 can identify two-prong objects, such as boosted W, Z, and Higgs bosons, and the variable $\tau_{32} = \tau_3/\tau_2$ can identify three-prong objects, such as boosted top quarks. To reject jets originating from boosted W bosons and tops, the aforementioned variables are required to satisfy $\tau_{21} > 0.6$ and $\tau_{32} > 0.67$. The misidentification rate, shown in Fig. 6.20 (left) for data and simulated events, is calculated as a function of the p_T of the top candidate. In the top candidate p_T region above 100 GeV where most of the signal is concentrated the misidentification rate SFs are around of 10% while for $p_T < 100$ GeV (a region with low statistics) the SFs become larger.

  The tagging efficiency is measured in data and simulation in events with a semi-leptonic $t\bar{t}$ final state. The events are required to pass the trigger `HLT_Mu50` and have exactly one muon with $p_T > 30$ GeV and $|\eta| < 2.4$, identified with loose criteria. In addition, events are required to have at least four small cone jets with $|\eta| < 2.4$, $p_T > 40$ GeV for the first three leading in p_T jets and with $p_T > 30$ GeV for any additional jet. Out of these selected jets, two are required to be tagged as b jets with the Medium working point of the CSVv2 tagger. Events containing electrons or hadronically decaying τ-leptons are rejected. Furthermore, one of the b jets is close to the

muon ($\Delta R(\text{bjet}, \mu) < 1.5$), while the distance between the muon and the leading in p_T top candidate is $\Delta R(\text{top}, \mu) > 2.0$.

To estimate the normalization of the QCD and $t\bar{t}$ background processes and measure the top quark tagging efficiency, two regions are defined based on the event E_T^{miss} and the muon isolation. The region with low E_T^{miss} (<20 GeV) and loose muon isolation criteria (>0.1), called as *Background Region (BR)*, is dominated mainly by QCD multijet events, as well as $t\bar{t}$ and other EWK (W+jets, Diboson, Z/γ^*, single-t) background events. The region with high E_T^{miss} (> 50 GeV) and tight muon isolation criteria is used to measure the top quark tagging efficiency and is called *Signal Region (SR)*. Both regions are used to obtain the normalization of $t\bar{t}$ and QCD multijet events in an iterative process. A first estimate of the QCD multijet normalization factor f_{QCD} is performed in the BR, by taking the ratio of the QCD multijet events in data over the simulation. The former is given by:

$$N_{\text{BR}}^{\text{QCD in data}} = N_{\text{BR}}^{\text{data}} - N_{\text{BR}}^{\text{EWK}} - f_{t\bar{t}} \times N_{\text{BR}}^{t\bar{t}}, \tag{6.18}$$

where $\text{N}_{\text{BR}}^{\text{data}}$ indicates the number of events in data, $\text{N}_{\text{BR}}^{\text{EWK}}$ is the number of events from simulated electroweak processes, and $\text{N}_{\text{BR}}^{t\bar{t}}$ is the number of simulated $t\bar{t}$ events. The normalization factor $f_{t\bar{t}}$ is initially equal to unity and updated using the SR:

$$f_{t\bar{t}} = \frac{N_{\text{SR}}^{t\bar{t} \text{ in data}}}{N_{\text{SR}}^{t\bar{t}}}, \tag{6.19}$$

where $\text{N}_{\text{SR}}^{t\bar{t} \text{ in data}}$ is calculated by taking into account the f_{QCD}:

$$N_{\text{SR}}^{t\bar{t} \text{ in data}} = N_{\text{SR}}^{\text{data}} - N_{\text{SR}}^{\text{EWK}} - f_{\text{QCD}} \times N_{\text{SR}}^{\text{QCD}}. \tag{6.20}$$

The above procedure is repeated using the updated values of f_{QCD} and $f_{t\bar{t}}$ each time and is stopped once their values converge. The f_{QCD} ($f_{t\bar{t}}$) is found to be 0.60 (0.90).

The tagging efficiency is estimated in the SR by subtracting the background (EWK + QCD) events from data, and then compared to the simulation to extract the tagging SFs. The mistagging rate SFs are taken into account in the calculation of the tagging efficiency. The middle plot in Fig. 6.20 shows the tagging efficiency as a function of the top candidate p_T. A plateau is reached for top candidates with p_T around 300 GeV with an efficiency of 55%. As the p_T of the top candidates increases, their decay products become collimated and the resolved top tagging efficiency decreases. Good agreement between data and simulation is observed, leading to a SF smaller than 10%. The top quark tagging efficiency is also measured for genuine top quark candidates in the SR, by subtracting both the background and the fake top quark candidates from data. The genuine top quark tagging efficiency is displayed in Fig. 6.20 (right). High
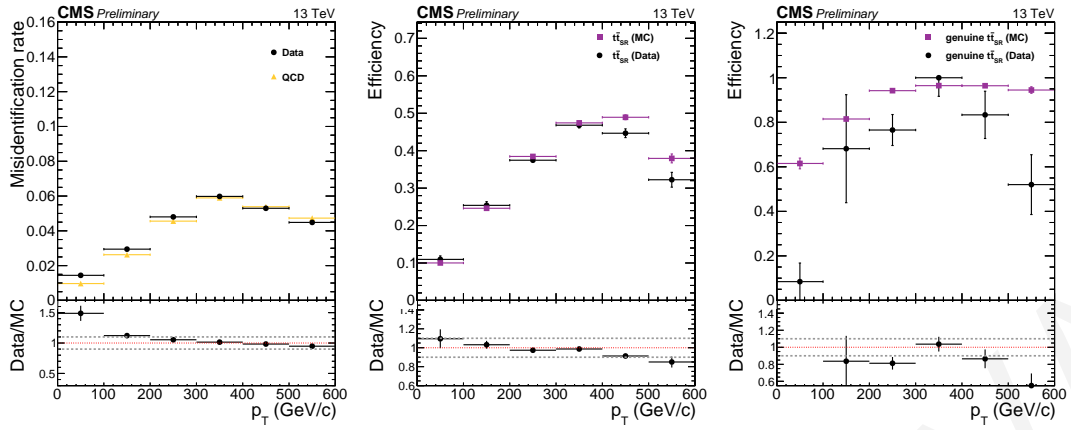
Figure 6.20: Misidentification rate as calculated in a data sample dominated by QCD multijet events (*left*). Tagging efficiency measured in SR with $t\bar{t}$ sample (*middle*). Tagging efficiency of genuine top candidates measured in SR (*right*).
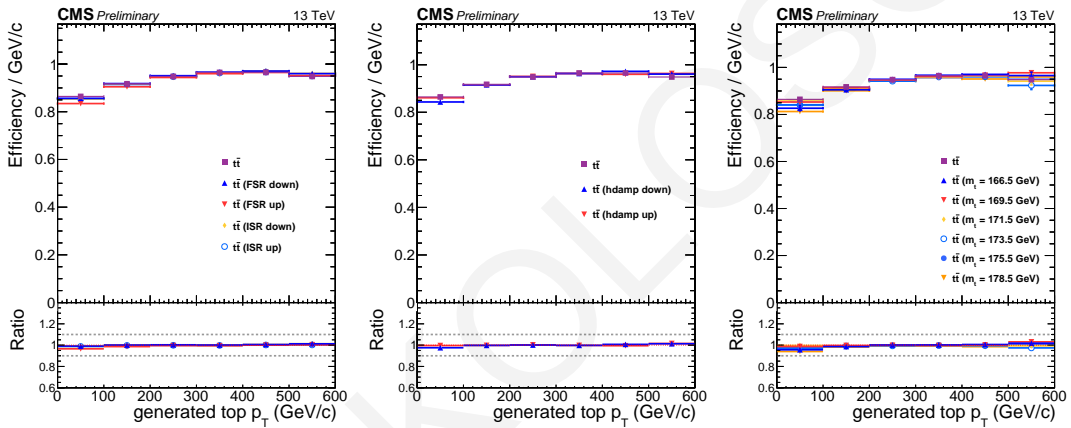
Figure 6.21: The effect of shower scale (*left*), h_{damp} tuning (*middle*) and top mass (*right*) on the tagging efficiency.

statistics p_{T} bins show a SF of less than 20%, while p_{T} bins with low statistics have SFs up to $\sim 40\%$.

The systematics on the top quark (mis)tagging have been also studied. The measurements were repeated using $t\bar{t}$ samples with different sources of systematic uncertainties and compared to the nominal one. These variations include different shower scales of the initial and final state radiation, a different tuning parameter h_{damp}[2] [156], different mass values of the top (from 166.5 GeV to 178.5 GeV), different event generator (POWHEG and EVTGEN), and different showering generator (PYTHIA8 and HERWIG++). Finally, the dependence on the top-matching definition used in the BDT training is also studied. The effect of each source is illustrated in Fig. 6.21 and Fig. 6.22.

The final top quark tagging systematic uncertainty is estimated from all the sources as

---

[2]h_{damp} is the model parameter regulating the high-p_{T} radiation. This is done by damping real emissions with a factor of $h_{\mathrm{damp}}^2/(p_{\mathrm{T}}^2 + h_{\mathrm{damp}}^2)$. Its default value is equal to the top-quark mass, $m_t = 172.5$ GeV.
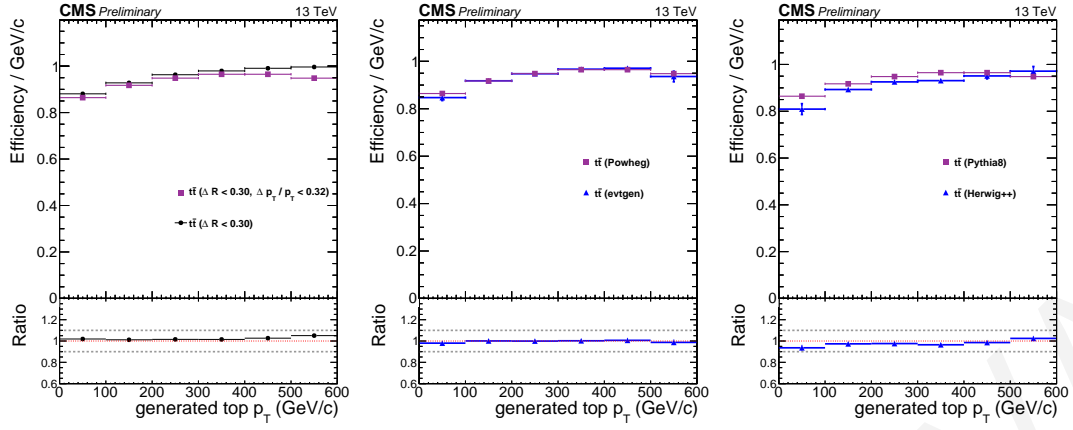
Figure 6.22: The effect of matching definition (*left*), event generator (*middle*) and parton showering generator (*right*) on the tagging efficiency.

follows:

$$\sigma_{SF}^{syst.} = \frac{1}{2} \times \sqrt{\sum_{j}^{\text{all effects}} \left(1 - \frac{\epsilon_{t\bar{t} \text{ nominal}}}{\epsilon_{t\bar{t} \text{ with effect } j}}\right)}, \qquad (6.21)$$

where $\epsilon_{t\bar{t} \text{ nominal}}$ is the top tagging efficiency as calculated in the nominal $t\bar{t}$ sample while $\epsilon_{t\bar{t} \text{ with effect } j}$ is the efficiency of the $t\bar{t}$ sample with effect j.

The uncertainty of the per-event top tagging SF is calculated by applying the error propagation law in equation 6.8. Due to the requirement of exactly two top-candidates, the SF uncertainty takes the form:

$$\left(\frac{\sigma_w^{\pm}}{w}\right)^2 = \sum_{i=1}^{(t|\text{gen}-t)} \left(\frac{\sigma_{\text{SF}_i(t|\text{gen}-t)}}{\text{SF}_i(t|\text{gen}-t)}\right)^2 + \sum_{i=1}^{(t|!\text{gen}-t)} \left(\frac{\sigma_{\text{SF}_i(t|!\text{gen}-t)}}{\text{SF}_i(t|!\text{gen}-t)}\right)^2, \quad (6.22)$$

where the sums run over all the tagged candidates (genuine and fake) in a given jet p_{T} bin.

- **misidentified b jets background:** The systematic uncertainties affecting the measurement of the misidentified b jets background can be divided into systematic uncertainties due to the simulated EWK processes subtracted from the data events in the three CRs, systematic uncertainties in the derivation of the transfer factors, and systematic uncertainties in the definition of the CRs. For the first category, simulated EWK processes include the systematic uncertainties from the trigger, lepton and τ_h-jet veto, JES and JER, b tagging and top tagging uncertainties, uncertainties related to the cross-section of the EWK processes, systematic uncertainties from the pileup and from the luminosity measurement. All aforementioned systematic uncertainties are passed in the calculation and scaled down by the fraction of misidentified b jets events in the selected sample. As they are anti-correlated with the misidentified b jets background measurement, the resulting effect of each of these is small. Uncertainties on the derivation of the transfer factors can affect both the rate and the shape of the misidentified b
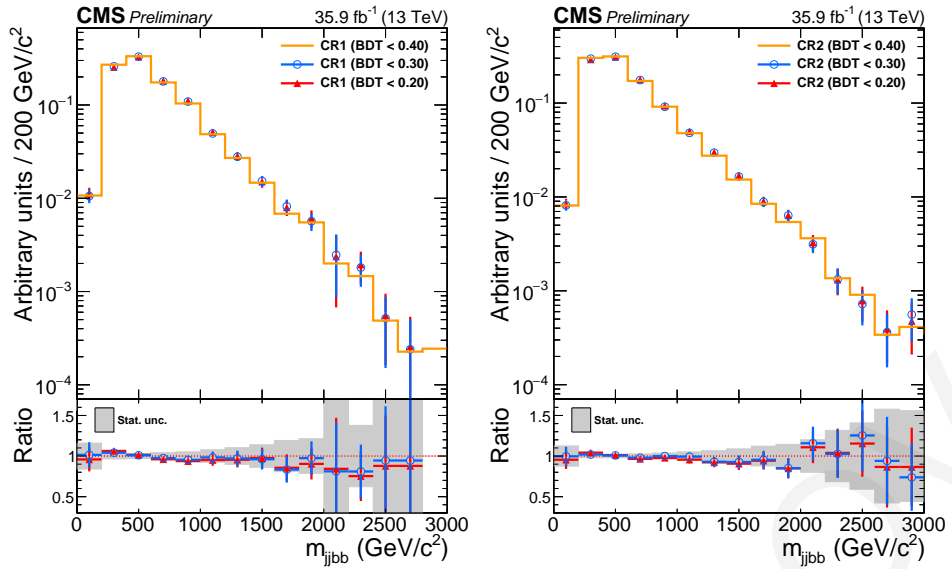
Figure 6.23: Normalized distributions of the charged Higgs boson candidate invariant mass in CR1 (*left*) and CR2 (*right*) for the default BDT cut value of the sub-leading in BDT score top candidate and for two variations.

jets background. To estimate the effect of these, the following procedure is followed:

1. The variable distributions of the misidentified b jets events are obtained from the CR1 and CR2.

2. The distribution from CR1 is divided by the one from the CR2, and the statistical uncertainty of the resulting distribution is calculated with error propagation. With this procedure, each transfer factor ($\mathrm{TF_i}$) is accompanied with its associated error ($\sigma_{\mathrm{TF_i}}$).

3. The transfer factor uncertainty is then added (subtracted) to (from) the nominal value of the transfer factor to define the up (down) variations and the misidentified b jets background is repeated using these two variations.

The uncertainty on the transfer factor estimation is therefore propagated to the invariant mass distribution of the charged Higgs boson and is treated as a rate plus shape nuisance. Finally, systematic uncertainties might arise from the definition of the CRs. To study this effect, the definitions of CR1 and CR2 are varied by changing the BDT score of the subleading-in-BDT top candidate and the final discriminant distributions are compared in CR1 and CR2, as shown in Fig. 6.23. The invariant mass distributions agree within statistical uncertainties indicating that the CRs are not sensitive to changes of the BDT score. For this reason, no additional systematic uncertainty is adopted. All aforementioned contributions affect the expected yield by approximately 4%.

Source	Shape	$H^{\pm}$	Misid. b	$t\bar{t}$	Single t, $t\bar{t} + X$	Electroweak
Trigger efficiency		5.0	0.09	0.69	0.04	0.01
Pileup	✓	< 0.01	–	< 0.01	< 0.01	< 0.01
Integrated luminosity		2.5	0.09	0.35	0.02	< 0.01
Lepton efficiency		0.32	–	0.04	< 0.01	< 0.01
Jet energy scale and resolution	✓	8.5	0.24	1.6	0.09	0.33
bjet identification	✓	5.0	–	0.64	0.04	0.01
top-tagging	✓	8.9	0.24	1.8	0.04	< 0.01
Transfer factors	✓	–	4.0	–	–	–
Top quark mass		–	0.09	0.39	0.02	–
Theory source						
Scales, PDF (acceptance)		5.1	–	0.39	0.02	0.01
Scales, PDF (cross section)		–	0.12	0.76	0.04	0.01

Table 6.5: The systematic uncertainties for signal and backgrounds, evaluated after fitting to data. The numbers are given in percentage and describe the effect of each nuisance parameter on the overall normalization of the signal model or the total background. Nuisance parameters with a checkmark also affect the shape of the $H^{\pm}$ candidate invariant mass spectrum. Sources that do not apply in a given process are marked with dashes. For the $H^{\pm}$ signal, the values for $m_{H^{\pm}} = 500$ GeV are shown.

6.5.2 Statistical uncertainties

The limited size of the simulated events modeling the signal and background processes lead to statistical fluctuations in the nominal predictions. The effect is treated as a rate plus shape uncertainty in the limit calculations using the *Barlow-Beeston lite* [157, 158] approach, which assigns the combined statistical uncertainty in each bin of the final fit distribution to the overall background yield in that specific bin.

6.5.3 Theoretical uncertainties

Theoretical uncertainties on the acceptance and production cross-section of the signal and background processes originate from the uncertainties of the factorization and renormalization scales, and of the Parton Distribution Functions (PDFs). The effect of the scale uncertainties is estimated for each simulated process separately by varying the factorization and renormalization scales independently and together by factors of 0.5 and 2 with respect to their default values. The final event yield is re-calculated for each of the six variations and the final systematic uncertainty is set to be the maximum variation with respect to the nominal yield. The PDF uncertainties are treated as fully correlated for all processes sharing the same dominant partons in the initial state of the matrix element calculations (i.e. gg, gq, or qq) [159]. For simulated background processes involving a top quark, the uncertainty on the top quark mass is also taken into account by varying the top quark mass by ± 1 GeV around its nominal value. All cross-section uncertainties are taken into account as rate uncertainties in the limit calculations.
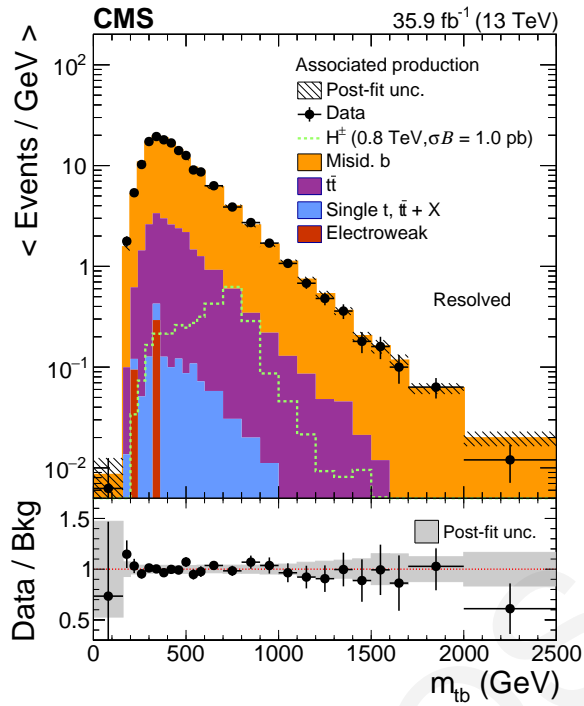
Figure 6.24: The post-fit distribution of the $\mathrm{H}^{\pm}$ mass in the background-only hypothesis. The expected signal for $m_{\mathrm{H}^{\pm}} = 800$ GeV is also shown, normalized to $\sigma\mathcal{B} = 1$ pb. The distribution is binned according to the statistical precision of the sample, leading to wider bins in the tail of the distribution.

6.6 Results

The shape of the invariant mass of the $\mathrm{H}^{\pm}$ candidate is used to assess the agreement with the background-only hypothesis or the presence of the signal in a global binned maximum-likelihood fit where all systematic uncertainties described in section 6.5 are incorporated as nuisance parameters. The post-fit distribution of the reconstructed $\mathrm{H}^{\pm}$ invariant mass is shown in Fig. 6.24 in the background-only hypothesis. The contribution of a hypothetical charged Higgs boson with a mass of 800 GeV is also displayed, assuming $\sigma\mathcal{B} = 1$ pb. The mass and p_{T} distribution of the $\mathrm{W}^{\pm}$ boson candidate coming from the $\mathrm{H}^{\pm}$ candidate decay are shown in Fig. 6.25. Figure 6.26 shows the p_{T} distribution of the leading in p_{T} free b jet used in the reconstruction of the $\mathrm{H}^{\pm}$. The observed data agree with the expected SM background processes.

Event yields

The summary of the expected and observed event yields is shown in Table 6.6. The observed data yield agrees with the expected background yield, within total uncertainty.
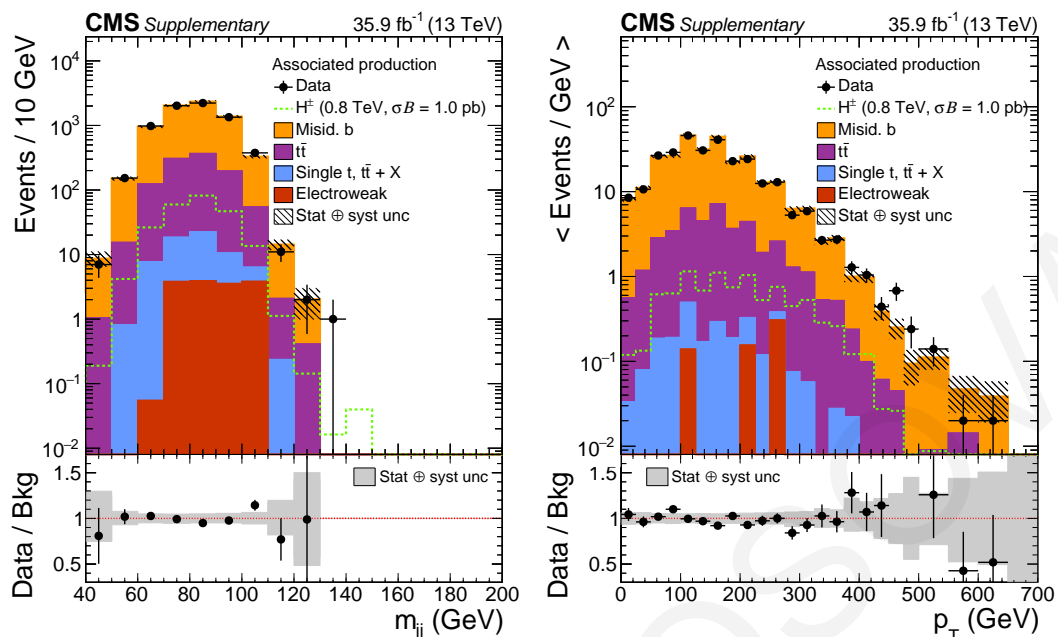
Figure 6.25: The post-fit mass (*left*) and p_{T} (*right*) distribution of the W boson candidate found in the leading in p_{T} top candidate in the background-only hypothesis. The expected signal for $m_{\mathrm{H}^{\pm}} = 800$ GeV is also shown.
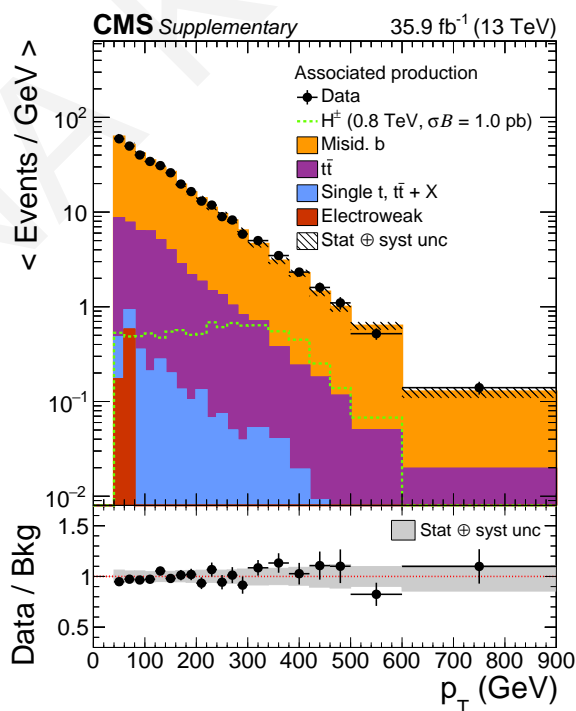
Figure 6.26: The post-fit p_{T} distribution of the b tagged jet used in the reconstruction of the invariant mass of the $\mathrm{H}^{\pm}$ candidate in the background-only hypothesis. The expected signal for $m_{\mathrm{H}^{\pm}} = 800$ GeV is also shown.

Process	Events $\pm$ (stat) $\oplus$ (syst)
Misidentified b jets	6152 ± 292
Genuine b jets	1067^{+185}_{-187}
Total expected from the SM	7220 ± 336
Observed	7124
$H^{\pm}$ signal, $m_{H^{\pm}} = 500$ GeV	183 ± 26
$H^{\pm}$ signal, $m_{H^{\pm}} = 650$ GeV	218^{+30}_{-31}
$H^{\pm}$ signal, $m_{H^{\pm}} = 800$ GeV	234 ± 33

Table 6.6: The number of expected and observed events after all selections. For background processes, the event yields and their corresponding uncertainties are prior to the background-only fit to the data. For the $H^{\pm}$ mass hypotheses of 500, 650, and 800 GeV, the signal yields are normalized to a $\sigma \mathcal{B} = 1$ pb and the total systematic uncertainties prior to the fit are shown.

95% CL upper limit on $\sigma_{pp \to t(b)H^{\pm}} \times \mathcal{B}(H^{\pm} \to tb)$						
$m_{H^{\pm}}$	Expected limit					Observed
(GeV)	-2σ	-1σ	median	$+1\sigma$	$+2\sigma$	limit
200	7.36875	10.00547	14.40000	21.11568	29.88361	21.32987
220	4.77422	6.54442	9.70000	14.84239	21.99761	13.37382
250	3.50000	4.75781	7.00000	9.87394	14.55466	5.81797
300	2.46904	3.35253	4.82500	7.03676	9.92984	4.28291
350	1.93359	2.62894	3.75000	5.43909	7.62903	4.01488
400	1.58750	2.19219	3.20000	4.76891	6.92032	2.72135
500	1.00928	1.37611	1.98750	2.91440	4.13678	2.21888
650	0.57129	0.78326	1.12500	1.64966	2.34850	1.38557
800	0.38245	0.52145	0.75313	1.11036	1.59426	0.92158
1000	0.28087	0.37828	0.54062	0.78844	1.11929	0.69687
1500	0.30752	0.41251	0.58750	0.85681	1.21634	0.45045
2000	0.38440	0.51837	0.73438	1.06223	1.49674	0.56143
2500	0.42539	0.57837	0.82500	1.20318	1.70805	0.51749
3000	0.50757	0.69010	0.98438	1.44345	2.05493	0.59750

Table 6.7: The observed and expected 95% CL upper limits on $\sigma_{pp \to t(b)H^{\pm}} \times \mathcal{B}(H^{\pm} \to tb)$ for the $H^{\pm}$ mass range of 200 GeV to 3 TeV.

6.7 Limits

The results shown in Fig. 6.24 are interpreted to set model-independent upper limits on the product of the charged Higgs boson production cross-section and branching fraction into a top and bottom quark-antiquark pair, $\sigma_{pp \to t(b)H^{\pm}} \times \mathcal{B}(H^{\pm} \to tb)$. The upper limits, shown in Fig. 6.27 for the mass range of 200 GeV to 3 TeV, are calculated at 95% confidence level (CL) using the asymptotic LHC-type CL_s criterion. The analysis is sensitive up to $m_{H^{\pm}}$ of approximately 1 TeV, where the boosted regime starts to take over. For very low masses, a rapid increase of the exclusion limit is observed, as a consequence of the overwhelming background from QCD multijet and irreducible $t\bar{t}$ events. The numerical values of the observed and expected upper limits are listed in Table 6.7.
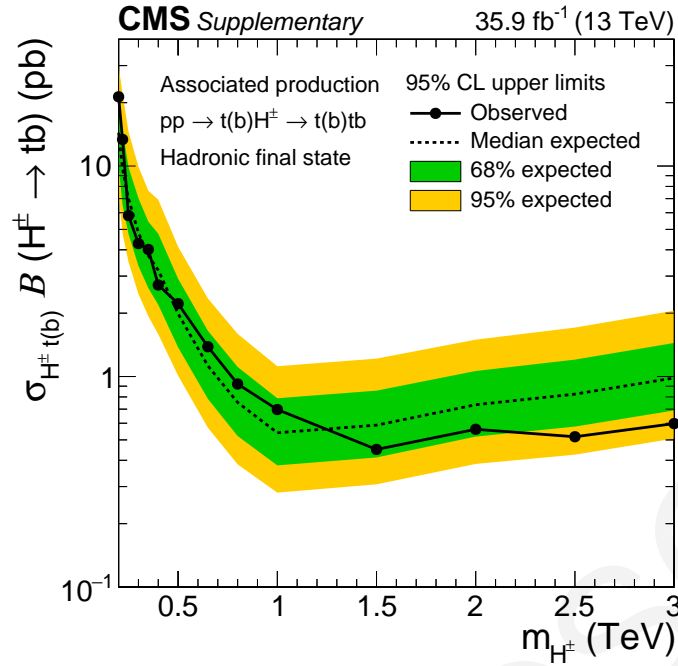
Figure 6.27: The observed upper limits at 95% CL on the product of the $H^{\pm}$ cross-section times branching fraction as a function of the $m_{H^{\pm}}$ in the 0.2-3 TeV region. The median expected limit (dashed line), 68% (inner green band), and 95% (outer yellow band) confidence interval for the expected limits are also shown.

Validation of the statistical model

The impact and pull of each nuisance parameter on the best-fit value of the signal strength are shown in Fig. 6.28 for the signal ($m_{H^{\pm}} = 800$ GeV) plus background hypothesis for the 30 most important nuisances. The bin-by-bin statistical uncertainties have a non-negligible impact; the 13th bin of the $H^{\pm}$ invariant mass distribution is the leading nuisance, followed by the systematic uncertainties on the misidentified b jets background (ranked second), the uncertainties related to the top quark tagging (ranked fourth) and the uncertainties on the JES (ranked fifth). A *goodness-of-fit* test is performed using the *saturated* method [160] to assess the compatibility of the observed data with a set 10000 generated pseudo-experiments from the modeled processes. The result is shown in Fig. 6.29 where the distribution of the toy experiment is shown with the red line and the observed value is represented by the blue line. The observed value of the χ^2 goodness-of-fit lies near the center of the toy distribution, concluding that the model describes well the observation.

Overlay of resolved and boosted exclusion limits

An overlay of the upper limits obtained with the resolved and boosted analyses is shown in Fig. 6.30. The reported limit at each mass point is determined by the analysis with the best-expected sensitivity. The resolved analysis shows the best sensitivity up to 1 TeV, where the two analyses meet. The boosted analysis has best sensitivity above 1 TeV.

Figure 6.28: The impacts and pulls of the 30 most important nuisance parameters for the signal plus background hypothesis assuming $m_{\mathrm{H}^{\pm}} = 800$ GeV.
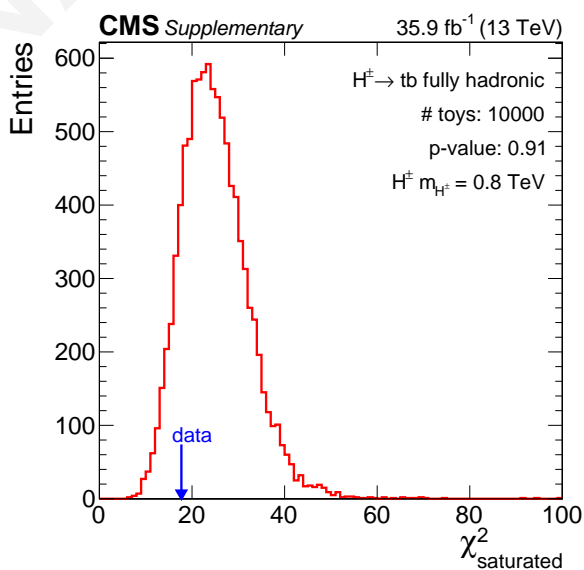
Figure 6.29: The goodness-of-fit test obtained with the saturated algorithm, for $m_{\mathrm{H}^{\pm}} = 800$ GeV.
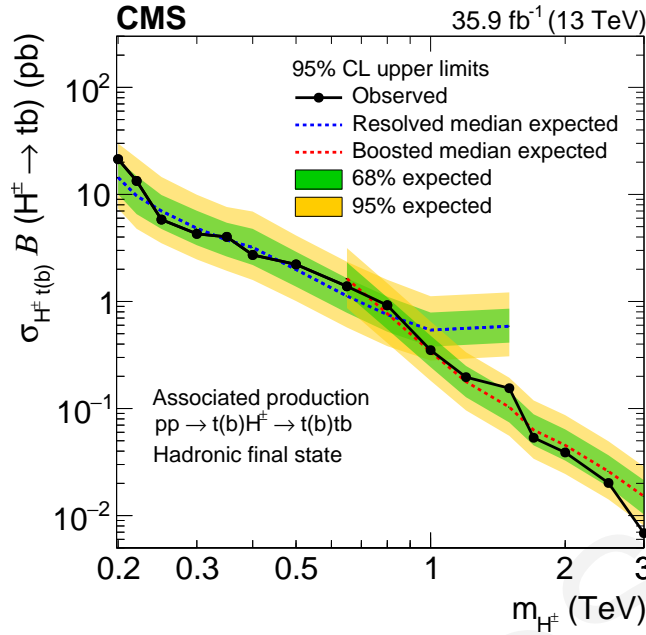
Figure 6.30: Overlay of the observed and expected upper limits at 95% CL on $\sigma_{\mathrm{H}^{\pm}} \times \mathcal{B}(\mathrm{H}^{\pm} \to \mathrm{tb})$ obtained by the resolved and boosted analyses.
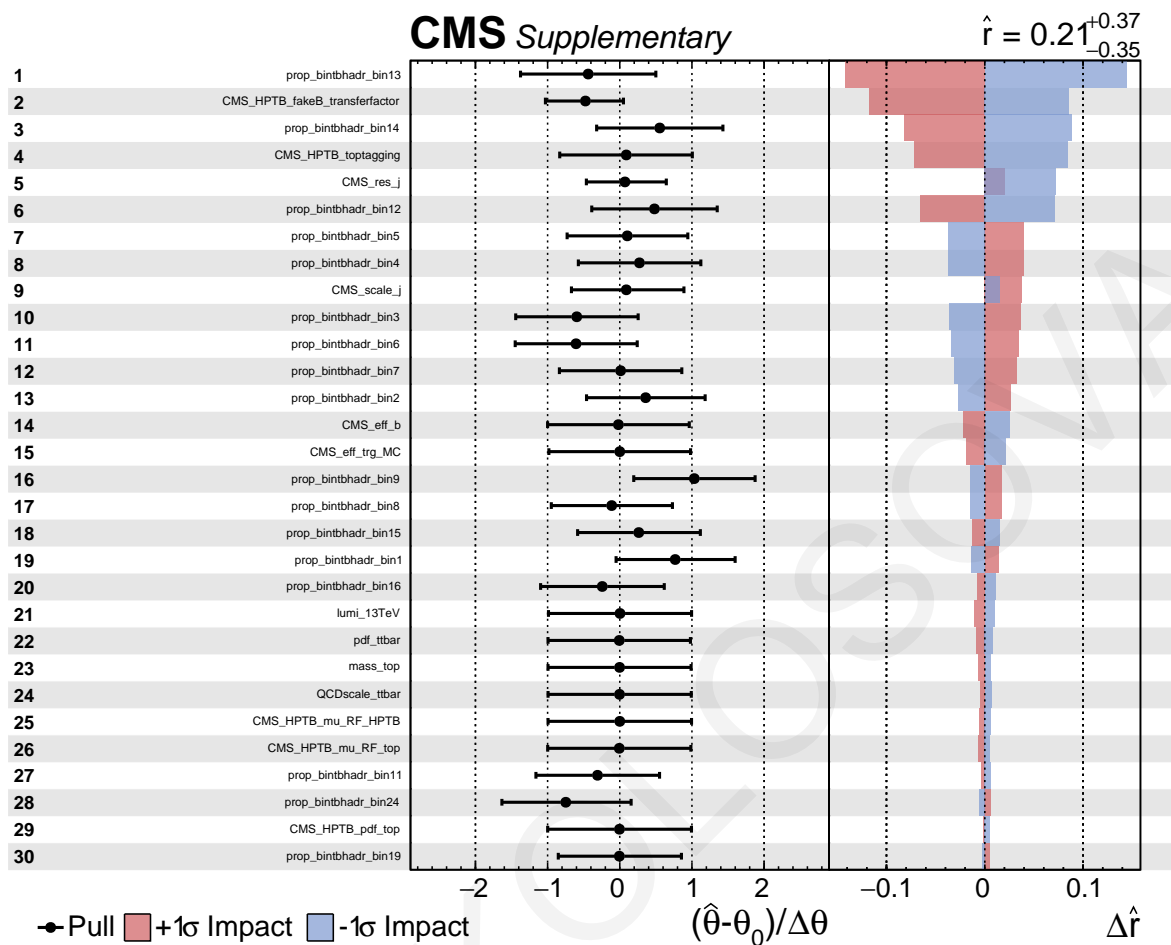
Model-dependent upper limits

The combined limits of the resolved and boosted analyses have been used to set model-dependent upper limits in the hMSSM and $M_{\mathrm{h}}^{125}(\tilde{\chi})$ [161] benchmark scenarios. The hMSSM scenario assumes that the light Higgs boson h in the 2HDM is the discovered Higgs boson and that SUSY particles are too heavy to be directly observed at the LHC. The $M_{\mathrm{h}}^{125}(\tilde{\chi})$ scenario features a significant mixing between gauginos and higgsinos, and a compressed mass spectrum of charginos and neutralinos. Due to the presence of light charginos and neutralinos, the $M_{\mathrm{h}}^{125}(\tilde{\chi})$ scenario has a unique phenomenology as the heavy Higgs bosons can decay to these superpartners. The upper limits are obtained by comparing the observed and expected limits on $\sigma_{\mathrm{H}^{\pm}} \times \mathcal{B}(\mathrm{H}^{\pm} \to \mathrm{tb})$ with theoretical predictions on the cross-sections and branching fractions provided by the LHC Higgs Cross Section Working Group [55]. The program FEYNHIGGS [162, 163, 164] is used to calculate the Higgs masses and mixing for each point on the $(m_A, \tan\beta)$ plane in the $M_{\mathrm{h}}^{125}(\tilde{\chi})$ scenario. The branching fractions in the hMSSM scenario are calculated with HDECAY [138], while for the $M_{\mathrm{h}}^{125}(\tilde{\chi})$ scenario the calculations are performed with the most precise results of FEYNHIGGS, HDECAY, and PROPHECY4F[165]. The model-dependent limits are shown in Fig. 6.31. In the hMSSM scenario, for $\mathrm{m}_{\mathrm{H}^{\pm}}$ in the range between 0.20 to 0.55 TeV, the maximum $\tan\beta$ value excluded is 0.88. In the $M_{\mathrm{h}}^{125}(\tilde{\chi})$ scenario, for $\mathrm{m}_{\mathrm{H}^{\pm}}$ in the range between 0.20 and 0.57 TeV, the maximum $\tan\beta$ value excluded is 0.86.
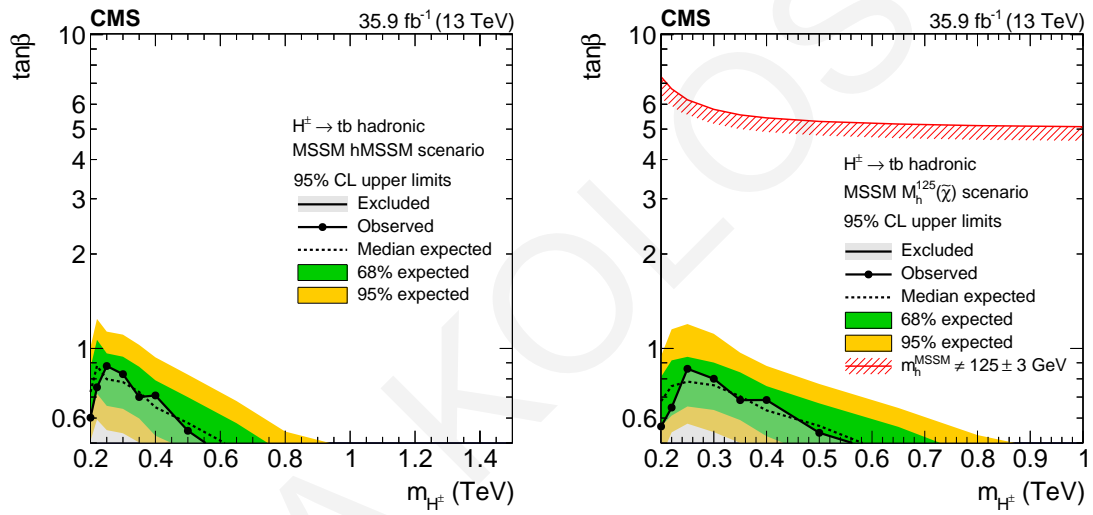
Figure 6.31: The observed 95% CL upper limits (solid black points), and the expected median (dashed line) in the MSSM hMSSM (*left*) and $M_h^{125}(\tilde{\chi})$ (*right*) benchmark scenarios. On the right, the region below the red line is not allowed if we assume that the observed neutral Higgs boson is the light CP-even 2HDM Higgs boson with a mass of 125 ± 3 GeV, where the uncertainty is the theoretical uncertainty in the mass calculation.

7 Search for $H^{\pm} \rightarrow tb$ with the 2017 LHC data

The search for charged Higgs bosons in the top and bottom quark-antiquark pair decay in the fully hadronic final state using the 2017 LHC data and the CMS detector is discussed in this chapter. This analysis is a continuation of the 2016 search in the resolved regime and features new top quark identification techniques, categorization of events based on the number of top quark candidates, a validation method for the background estimation and an improved signal extraction technique using a deep neural network. The 2017 analysis is based on data and simulation from the *Ultra Legacy* (UL) reprocessing campaign, which is characterized by improved detector performance through refined corrections and calibrations. In addition, the MC simulation conditions have been derived from data such that simulation describes accurately the detector response, and particularly the noise. The 2016 and 2018 data are not (re)analyzed and thus not included in this report as no signal and background samples are available, yet, in the UL campaign. For this reason, the results of the 2017 analysis presented here are *blinded* in the signal region and only the expected upper limits on the cross-section of the charged Higgs boson times the branching fraction are reported. The chapter follows a structure similar to Chapter 6, with a focus on the new features. Section 7.1 describes the online event selection and efficiency measurement. The updated event selection workflow is discussed in section 7.2. Section 7.3 gives a brief overview of the data quality filters and the corrections applied in simulated samples. The background measurement and validation methods are described in section 7.4. A new event-based deep neural network, used to extract the signal, is presented in section 7.5. In section 7.6, the systematic uncertainties are listed, with a focus on the systematic uncertainties related to the background measurement. The expected exclusion limits are reported in section 7.8.

7.1 Online event selection

The triggers used in the 2017 data taking period are listed in Table 7.1. All of the HLT paths were seeded by the logical `OR` of `L1_HTTVer` and `L1_HTTVer_QuadJet_W_X_Y_Z_er` seeds, where `V` and `W,Z,Y,Z` correspond to the thresholds of L1 H_T, and the p_T of the four leading-in-p_T L1 Jets, respectively. To cope with the increased instantaneous luminosity delivered during the 2017 data-taking period, and therefore with the high pileup environment, the L1 H_T was calculated from η-restricted L1 Jets of $|\eta| < 2.6$, instead of $|\eta| < 3.0$. Moreover, the high thresholds of the L1 H_T at the beginning of the 2017 LHC run (above 380 GeV)

Level-1 seeds	HLT Paths
L1_HTT280er	HLT_PFHT380_SixJet32_DoubleBTagCSV_p075
L1_HTT300er	HLT_PFHT380_SixPFJet32_DoublePFBTagCSV_2p2
L1_HTT320er	HLT_PFHT430_SixJet40_BTagCSV_p080
L1_HTT340er	HLT_PFHT430_SixPFJet40_PFBTagCSV_1p5
L1_HTT380er	HLT_PFHT380_SixPFJet32_DoublePFBTagDeepCSV_2p2
L1_HTT400er	HLT_PFHT1050
L1_HTT280er_QuadJet_70_55_40_35_er2p5	
L1_HTT300er_QuadJet_70_55_40_35_er2p5	
L1_HTT320er_QuadJet_70_55_40_40_er2p5	
L1_HTT320er_QuadJet_70_55_40_40_er2p4	
L1_HTT340er_QuadJet_70_55_40_40_er2p5	
L1_HTT320er_QuadJet_70_55_45_45_er2p5	
L1_HTT340er_QuadJet_70_55_45_45_er2p5	

Table 7.1: The list of L1 seeds and HLT paths used for the 2017 analysis.

motivated the introduction of new algorithms of the form `L1_HTTVer_QuadJet_W_X_Y_Z`, which were tuned according to the instantaneous luminosity throughout the run.

On the HLT side, besides the retuned PF-jet thresholds on H_T, p_T, and online b tagging working points, the 2017 triggers have, nevertheless, the same filtering structure as the 2016 triggers. The changes on the thresholds are listed below:

- `HLT_PFHT380_SixJet32_DoubleBTagCSV_p075`: the PF-H_T threshold decreased by 20 GeV, while the p_T threshold of the six leading PF-jets increased by 2 GeV. Moreover, the online b tagging working point of the CSVv2 algorithm increased from 0.63 to 0.75, which corresponds to a 2.2% misidentification rate.

- `HLT_PFHT430_SixJet40_BTagCSV_p080`: the PF-H_T thresholds decreased by 20 GeV, while the p_T threshold of the six leading PF-jets remained unchanged. The online b tagging working point of the CSVv2 algorithm increased from 0.63 to 0.80, corresponding to a 1.5% misidentification rate.

During the data taking, new naming conventions were introduced for all paths in the HLT menu, such that the type of jets used and the misidentification rate, instead of the b tagging working points, are included in the path name. Therefore, the path

- `HLT_PFHT380_SixJet32_DoubleBTagCSV_p075` was renamed to `HLT_PFHT380_SixPFJet32_DoublePFBTagCSV_2p2` and

- the path `HLT_PFHT430_SixJet40_BTagCSV_p080` to `HLT_PFHT430_SixPFJet40_PFBTagCSV_1p5`.

At the end of era `Run2017C`, a copy of the trigger with at least two b tagged jets was introduced online as a pilot path for the online b tagging algorithm DeepCSV. The path `HLT_PFHT380_SixPFJet32_DoublePFBTagDeepCSV_2p2` remained unprescaled until the end of 2017 data-taking period. The path requiring at least one b tagged jet was mistakenly prescaled during `Run2017C` era for a non-negligible period equivalent to 6.2 fb^{-1}. To recover some of the efficiency, an additional path was used, the `HLT_PFHT1050`.

Table 7.2: Data samples used for the trigger efficiency measurement.

| Dataset	Runs	$\mathcal{L}$ (pb^{-1})
SingleMuon_Run2017B_09Aug2019_UL2017_v1	297050 - 299329	4803.363
SingleMuon_Run2017C_09Aug2019_UL2017_v1	299368 - 302029	9572.498
SingleMuon_Run2017D_09Aug2019_UL2017_v1	302031 - 302663	4247.682
SingleMuon_Run2017E_09Aug2019_UL2017_v1	303825 - 304797	9313.642
SingleMuon_Run2017F_09Aug2019_UL2017_v1	305044 - 306460	13538.886
Total integrated luminosity		41476.072

7.1.1 Trigger efficiency measurement

The trigger efficiency measurement is performed as in the 2016 analysis, using a semileptonic $\mathrm{t\bar{t}}$ sample. The events are recorded with the lowest unprescaled single muon trigger, `HLT_IsoMu27`, and are required to have exactly one isolated reconstructed muon, identified with *tight* criteria, and have $p_{\mathrm{T}} > 29$ GeV and $|\eta| < 2.4$. Events with any additional muon, electrons, and τ_h-jets are rejected. Moreover, the events are required to have at least seven jets with $|\eta| < 2.4$, $p_{\mathrm{T}} > 40$ GeV for the six leading-in-p_{T} jets and $p_{\mathrm{T}} > 30$ GeV for any additional jet. To ensure that the muon and the selected jets do not overlap, a distance in the $\eta - \phi$ plane greater than 0.4 is required. In addition, the scalar sum of the p_{T} of all selected jets is required to be at least 500 GeV. The selected events are then categorized based on the number of jets tagged as b jets with the Medium working point of the DEEPJET algorithm. There are two categories:

- $2\mathrm{M1L_b}$, for events with exactly two medium-tagged b jets and at least one loose b jet, and

- $3\mathrm{M_b}$, for events with at least three medium-tagged b jets.

The trigger efficiency is measured in data and simulation and scale factors are extracted for each category. The scale factors extracted for $2\mathrm{M1L_b}$ are used to correct the simulation in the inverted regions of the background measurement described in section 7.4, while the ones extracted for $3\mathrm{M_b}$ are used to correct the simulation in the signal region. The data and simulation samples used for the trigger efficiency measurement are listed in Tables 7.2 and 7.3, respectively. The efficiency of the signal triggers requiring multiple jets and at least one or two b tagged jets as a function of the event H_{T} and the p_{T} of the sixth leading-in-p_{T} jet is shown in Fig. 7.1 for events in the $2\mathrm{M1L_b}$ category and in Fig. 7.2 for events that fall in the $3\mathrm{M_b}$ category. The large discrepancy between data and simulation, which is more obvious in the $2\mathrm{M1L_b}$ category, is mainly caused by the prescale of the single b tagging trigger during part of `Run2017C` and the high H_{T} thresholds in the beginning of 2017 data-taking period. To recover some of the efficiency in the data, the trigger `HLT_PFHT1050` is also used. The trigger efficiency of the combination of all triggers, including path `HLT_PFHT1050` is shown in Fig. 7.3 for category $2\mathrm{M1L_b}$ and in Fig. 7.4 for category $3\mathrm{M_b}$.

Table 7.3: Simulated samples used for the trigger efficiency measurement.

Process	Cross section (pb)	Events	Dataset (MINIAOD v1)
t̄t leptonic	88.34	66259900	/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8
t̄t hadronic	377.96	129706300	/TTToHadronic_TuneCP5_13TeV-powheg-pythia8
t̄t semileptonic	365.45	114058500	/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8
QCD, $15 < p_{\mathrm{T}} < 20$ GeV	3.625×10^6	26332765	/QCD_Pt-15to20_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD, $20 < p_{\mathrm{T}} < 30$ GeV	3.153×10^6	28365949	/QCD_Pt-20to30_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD, $30 < p_{\mathrm{T}} < 50$ GeV	1.652×10^5	30992451	/QCD_Pt-30to50_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD, $50 < p_{\mathrm{T}} < 80$ GeV	4.487×10^5	20937742	/QCD_Pt-50to80_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD, $80 < p_{\mathrm{T}} < 120$ GeV	1.052×10^5	613257	/QCD_Pt-80to120_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD, $120 < p_{\mathrm{T}} < 170$ GeV	2.549×10^4	648944	/QCD_Pt-120to170_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD, $170 < p_{\mathrm{T}} < 300$ GeV	8.644×10^3	36918785	/QCD_Pt-170to300_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD, $300 < p_{\mathrm{T}} < 470$ GeV	7.967×10^2	494796	/QCD_Pt-300to470_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD, $470 < p_{\mathrm{T}} < 600$ GeV	7.920×10^1	517382	/QCD_Pt-600to800_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD, $600 < p_{\mathrm{T}} < 800$ GeV	2.525×10^1	17354539	/QCD_Pt-800to1000_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD, $800 < p_{\mathrm{T}} < 1000$ GeV	4.724	16995944	/QCD_Pt-800to1000_MuEnrichedPt5_TuneCP5_13TeV_pythia8
QCD, $1000 < p_{\mathrm{T}} < $ Inf GeV	1.619	14719636	/QCD_Pt-1000toInf_MuEnrichedPt5_TuneCP5_13TeV_pythia8
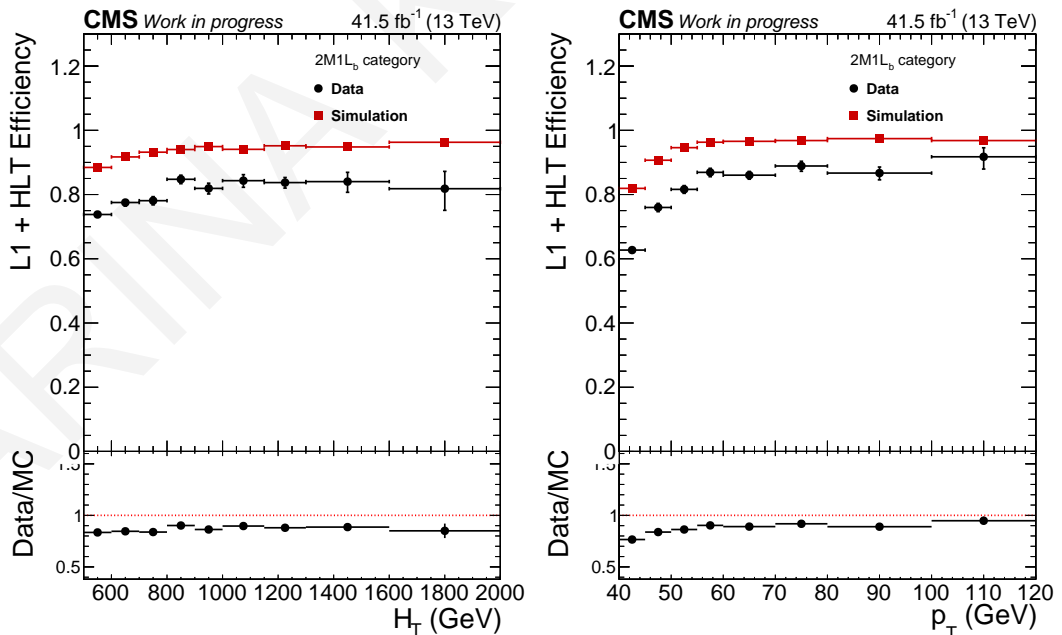
Figure 7.1: Trigger efficiency as a function of H_{T} and the p_{T} of the sixth leading-in-p_{T} jet for the $2\mathrm{M1L_b}$ category, without the inclusion of HLT_PFHT1050.

Figure 7.2: Trigger efficiency as a function of H_T and the p_T of the sixth leading-in-p_T jet for the $3M_b$ category, without the inclusion of HLT_PFHT1050.

Figure 7.3: Trigger efficiency as a function of H_T and the p_T of the sixth leading-in-p_T jet for the $2M1L_b$ category, with the inclusion of HLT_PFHT1050.
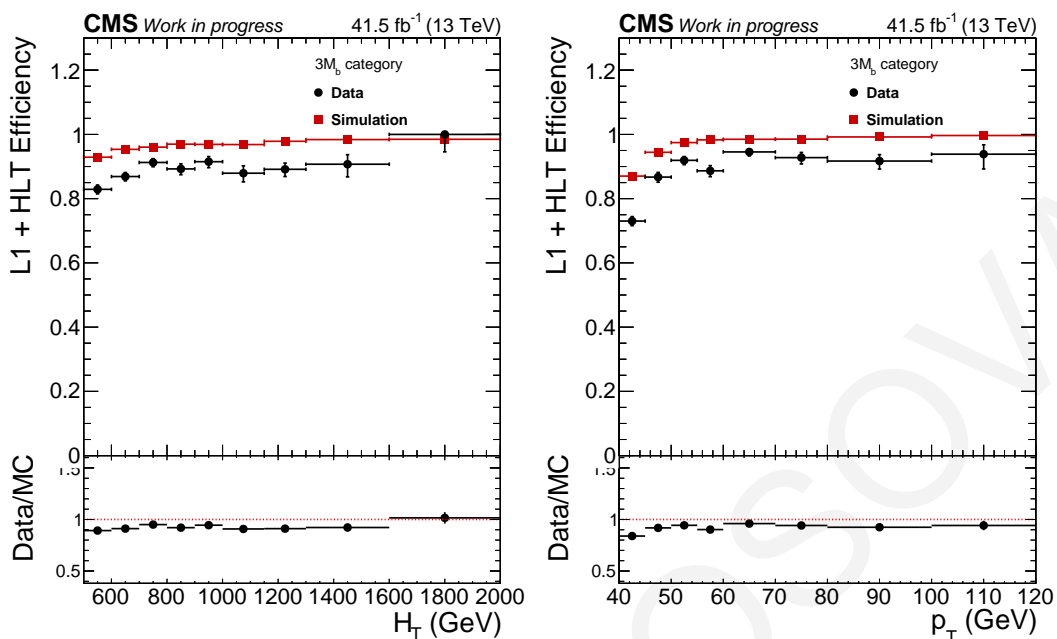
Figure 7.4: Trigger efficiency as a function of H_T and the p_T of the sixth leading-in-p_T jet for the $3M_b$ category, with the inclusion of `HLT_PFHT1050`.
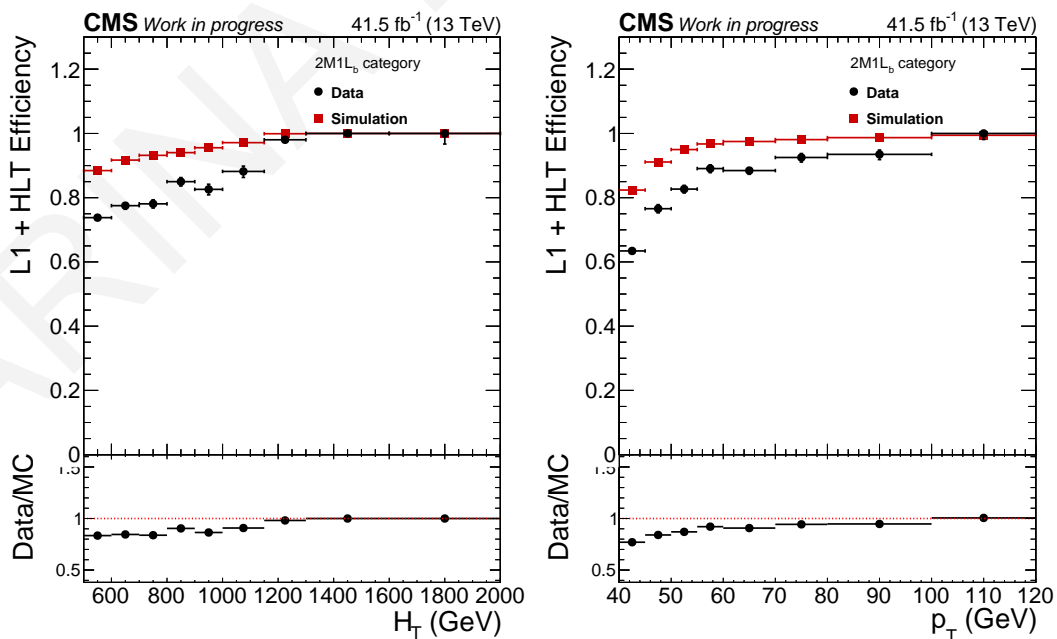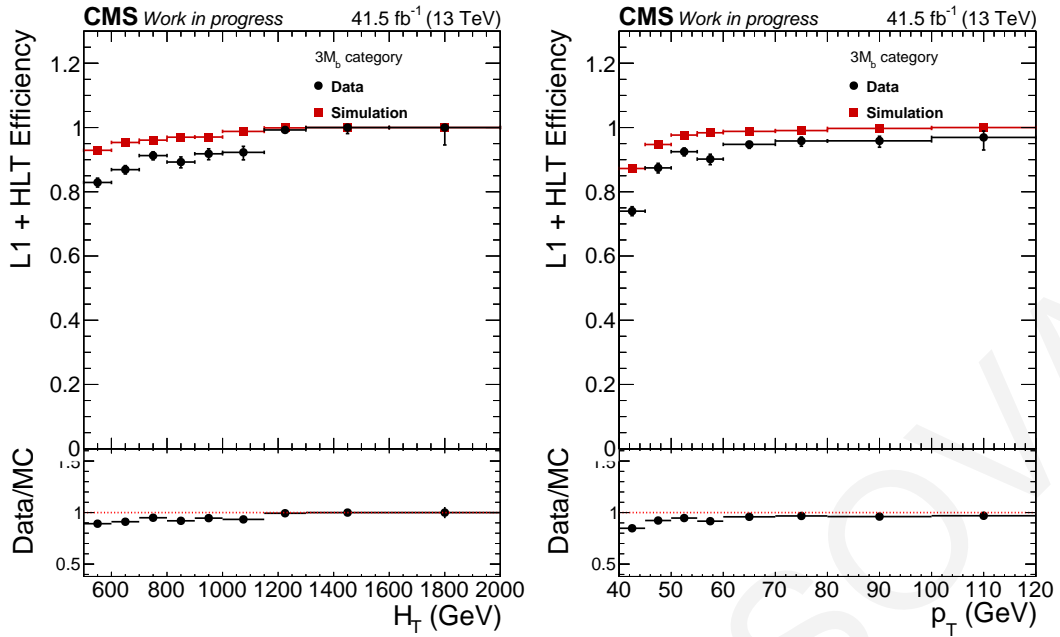
	Trigger name	Runs	Luminosity (fb^{-1})	Comment
(a)	HLT_PFHT430_SixPFJet40_PFBTagCSV_1p5 HLT_PFHT380_SixPFJet32_DoublePFBTagCSV_2p2 HLT_PFHT1050	297047-299329	4.8	High L1 H_T thresholds
(b)	HLT_PFHT380_SixPFJet32_DoublePFBTagCSV_2p2 HLT_PFHT1050	299368-300817	6.2	Single b-tagging path prescaled
(c)	HLT_PFHT430_SixPFJet40_PFBTagCSV_1p5 HLT_PFHT380_SixPFJet32_DoublePFBTagCSV_2p2 HLT_PFHT1050	301046-302019	3.4	Low L1 H_T thresholds
(d)	HLT_PFHT430_SixPFJet40_PFBTagCSV_1p5 HLT_PFHT380_SixPFJet32_DoublePFBTagCSV_2p2 HLT_PFHT380_SixPFJet34_DoublePFBTagDeepCSV_2p2 HLT_PFHT1050	302026-306462	27.1	Introduction of DeepCSV pilot path

Table 7.4: Trigger combinations used in each run period for the calculation of individual scale factors.

7.1.2   Trigger scale factors measurement

To correct the remaining discrepancies between data and simulation, scale factors as a function of the event H_T and the p_T of the sixth leading-in-p_T jet are derived for the two categories. To address the high L1 H_T thresholds at the beginning of the 2017 data-taking period, the prescale of the trigger requiring at least one b tagged jet, and the introduction of the pilot path for the online DeepCSV b tagging algorithm, four scale factors are calculated, corresponding to different run periods. The combination of triggers used in data for each run period is listed in Table 7.4. A final scale factor is calculated as the sum of the individual scale factors, weighed by the integrated luminosity recorded in each run period. Events passing *only* the single b tagging trigger in the prescaled period are rejected. The trigger efficiency as a function of the event H_T and the p_T of the sixth leading-in-p_T jet for each of the run periods mentioned in Table 7.4 for data events in the $2M1L_b$ and $3M_b$ categories are shown in Fig. 7.5 and 7.6, respectively. The efficiencies of the combination of all triggers in
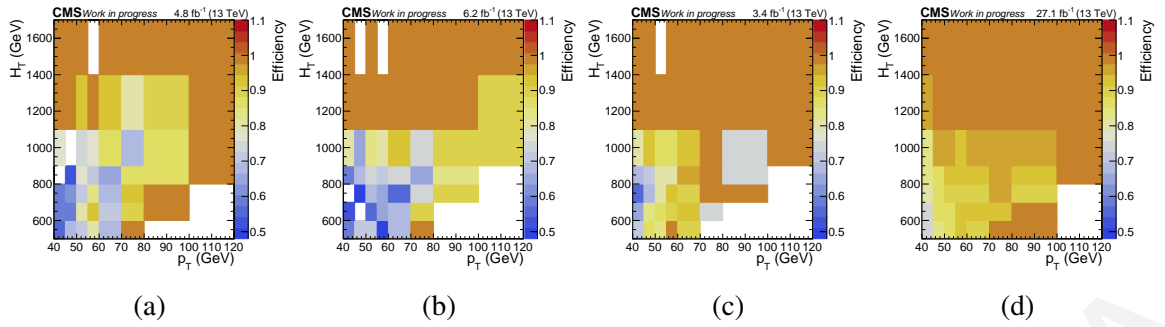
Figure 7.5: Trigger efficiency in data for events in the $2\mathrm{M}1\mathrm{L_b}$ category.
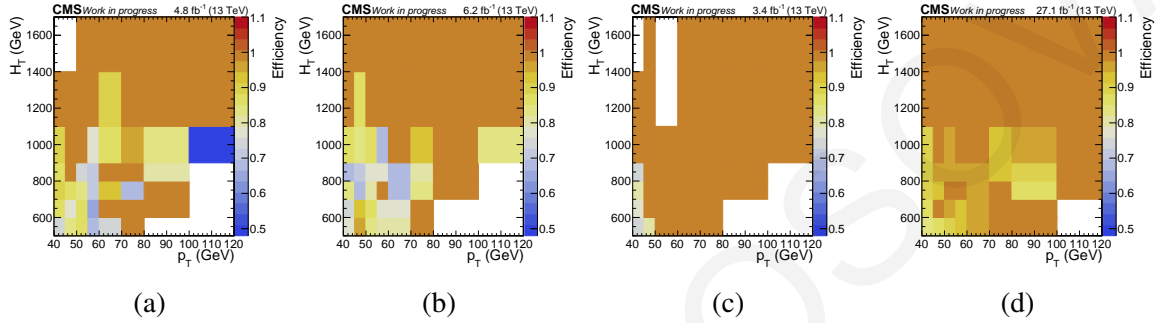
Figure 7.6: Trigger efficiency in data for events in the $3\mathrm{M_b}$ category.

simulated events for the two categories are shown in Fig. 7.7. The resulting scale factors for each run period are shown in Fig. 7.8 for the $2\mathrm{M}1\mathrm{L_b}$ category and in Fig. 7.9 for the $3\mathrm{M_b}$ category. The scale factors, weighed by the integrated luminosity, are shown in Fig. 7.10.

7.2 Event selections

After the online selection, events need to pass the *baseline* selection described in section 5.4. Subsequently, the hadronic jets of the selected events are fed to a new top quark tagging algorithm and are then classified into two categories based on the number of medium-tagged top quark candidates.

7.2.1 top quark tagging

Hadronically decaying top quarks are identified with a classifier based on a multivariate deep neural network (DNN), which can distinguish combinations of three small-cone jets (of which one is tagged as a b jet) originating from the decay of top quarks or trijet combinations from the combinatorial background. It utilizes high-level information from each of the three jets, as well as from jet systems, such as invariant masses, angular separations, jet flavor, and jet shape variables. There are 33 input variables, listed in Table 7.5. To prevent the mass-sculpting effect that was present in the BDT-based top quark tagging algorithm described in section 6.2.1, this algorithm uses a mass-reweighing technique to decorrelate the classifier output from the top quark mass. The performance of the classifier is shown in Fig. 7.11,
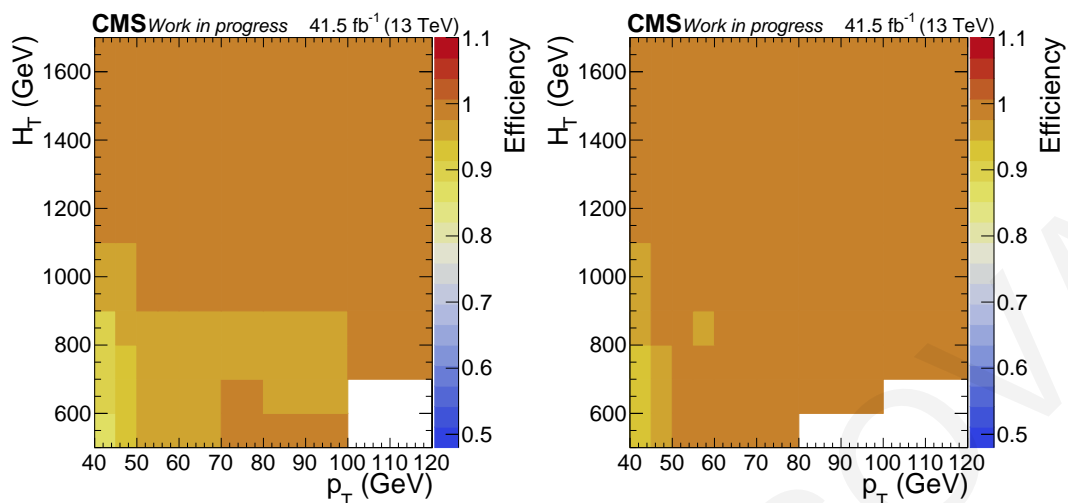
Figure 7.7: Trigger efficiency in simulated events for the $2M1L_b$ category (*left*) and $3M_b$ category (*right*).
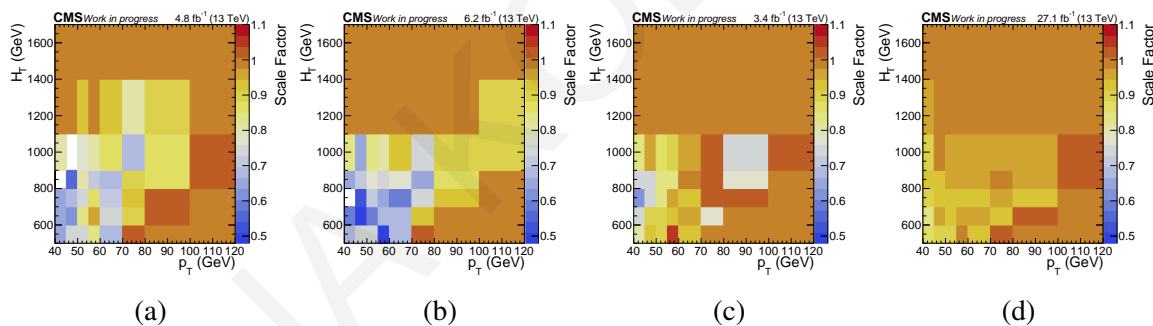

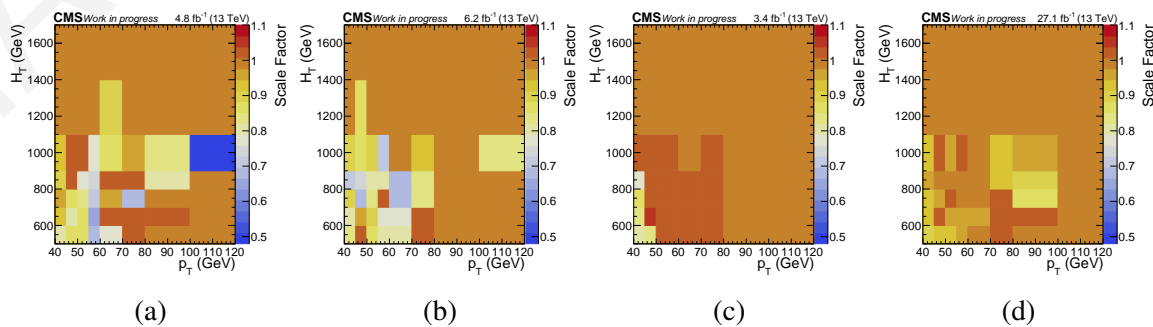Figure 7.8: Trigger scale factors for each run period in events in the $2M1L_b$ category.

Figure 7.9: Trigger scale factors for each run period in events in the $3M_b$ category.
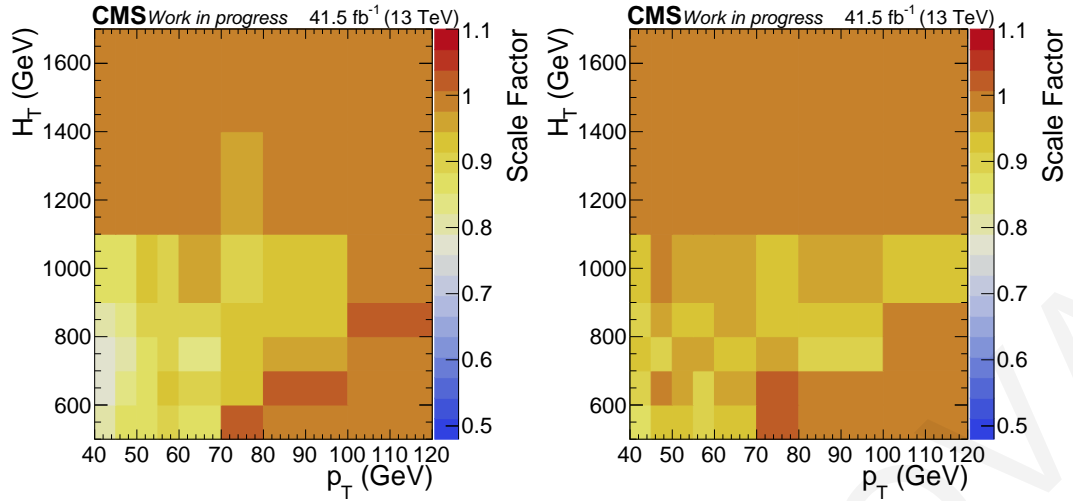
Figure 7.10: Trigger scale factors for events in the $2M_b$ (*left*) and $3M_b$ category (*right*) categories.
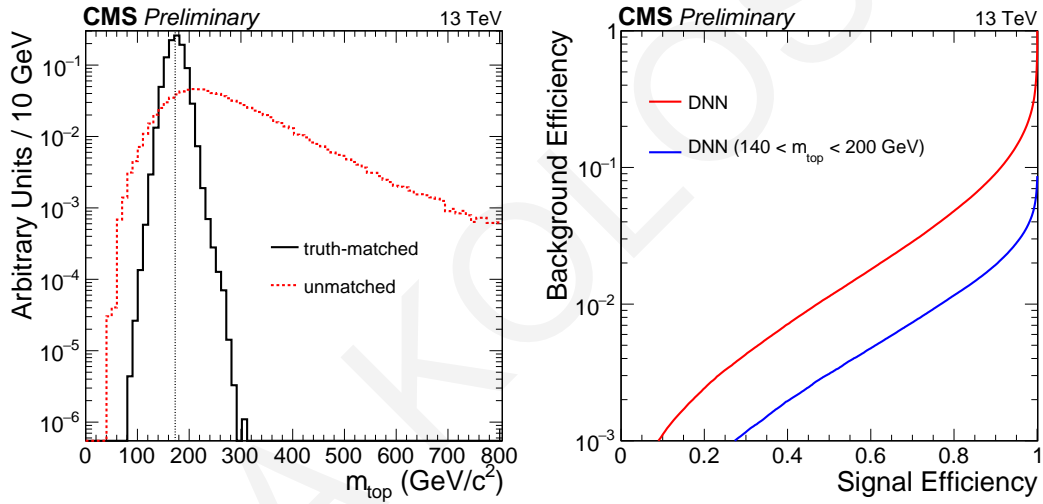

Figure 7.11: *Left:* the mass distribution of truth-matched and non-matched top quark candidates. *Right:* the ROC curve for top quark candidates without any selection on their invariant mass (red) and with mass inside the top mass window of 140 to 200 GeV (blue).

where the left plot shows the top quark candidate mass distribution for truth-matched and non-matched candidates. The truth-matched candidates have a peak distribution around the mass of the top quark, while the non-matched candidates (combinatorial background) have a smoothly distribution with a turn-on region where the peak is shifted from the real value of the top quark mass. The right plot of Fig. 7.11 shows the Receiver Operating Characteristic (ROC) curve of the classifier for two cases; without any selection on the mass of the top quark candidates (red) and when requiring the top quark candidates mass to fall inside the mass window of 140 to 200 GeV (blue). The results show a significant reduction of the top quark misidentification rate when the top quark candidate mass is close to the top quark mass, as expected.

For this analysis, two working points are exploited, corresponding to 10% (95%) and 5% (87%) background acceptance (signal efficiency). Events are required to contain at least

Table 7.5: The list of high-level features used in the training of the MVA classifier. The b-index refers to b jet properties, and $j_1^W(j_2^W)$ indices refer to the leading (subleading) in p_T jet from the W boson decay.

Jet variables (j_1^W, j_2^W: W boson subjets, b: bjet)	
mass	(j_1^W, j_2^W, b)
p_T in top CM	(j_1^W, j_2^W, b)
$\frac{\|p_{T,jet}-p_{T,top}\|}{p_{T,jet}+p_{T,top}}$	(j_1^W, j_2^W, b)
b tagging discriminator	(b)
charm to light (CvsL) discriminator	(j_1^W, j_2^W, b)
charm to bottom (CvsB) discriminator	(j_1^W, j_2^W, b)
minor axis of the jet ellipse on the $\eta - \phi$ plane	(j_1^W, j_2^W, b)
fragmentation function (p_T D)	(j_1^W, j_2^W, b)
particle constituent multiplicity	(j_1^W, j_2^W)
Trijet variables	
top mass	
$p_T\Delta R_{top}$ ($p_{T,top}\Delta R(W,b)$)	
W mass	
$p_T\Delta R_W$ ($p_{T,W}\Delta R(j_1^W, j_2^W)$)	
$\frac{p_{T,W}}{p_{T,j_1^W}+p_{T,j_2^W}}$	
$\cos\omega(j_i,j_k) = \frac{\vec{p}_{j_i}\cdot\vec{p}_{j_k}}{\|p_{j_i}\|\|p_{j_k}\|}$	
Soft Drop(j_1^W, j_2^W)	
$\Delta\eta$(W,b)	

one (two) cleaned top quark candidate(s) tagged with the medium (loose) working point and with a mass inside the top mass window of 120 to 230 GeV. Events with more than two top quark candidates tagged with the medium working point are rejected. The leading-in-MVA score top quark candidate, which is medium-tagged, is referred to as *top 1*, while the subleading-in-MVA top quark candidate, which is at least loose-tagged, is referred to as *top 2*.

7.2.2 $H^{\pm}$ **candidate reconstruction**

The charged Higgs boson candidate is reconstructed from the four-momenta of the leading-in-p_T top quark candidate between *top 1* and *top 2*, and the leading-in-p_T free b jet that is not used in the reconstruction of the two top quark candidates. To ensure a good quality of the $H^{\pm}$ candidate, events in which the leading-in-p_T top quark candidate is not tagged with the medium working point are rejected. Events are categorized based on the number of medium-tagged top quark candidates to:

- $1M1L_t$: the top quark candidate coming from the $H^{\pm}$ decay is medium-tagged, the associated top quark candidate is loose-tagged, and

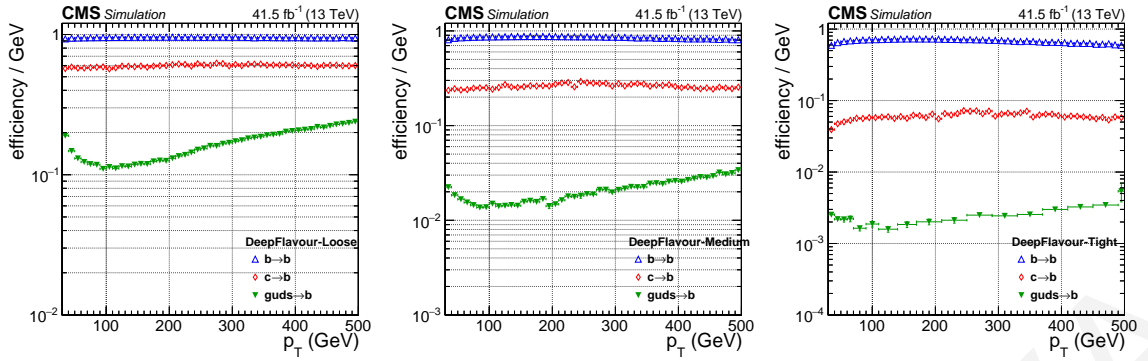- $2M_t$: both top quark candidates are medium-tagged.

Figure 7.12: Simulated tagging efficiency and mistagging probabilities for different flavors of jets for the DEEPJET b tagging algorithm and for the *Loose* (left), *Medium* (middle), and *Tight* (right) working points.

7.3 Data quality filtering and simulation corrections

Data and simulation are subject to quality filters due to spurious high p_T^{miss} events. All simulated samples are additionally corrected with the following weights and scale factors.

- **Pileup:** Simulated samples are reweighed such that the pileup distribution is in agreement with the one observed in 2017 data.

- **Trigger:** The trigger efficiency in simulated samples differs with respect to the one observed in data, as explained in section 7.1.1. Simulated samples are corrected with the application of the scale factors calculated in 7.1.2.

- **b tagging efficiency and b mistagging rate:** The b tagging efficiency and misidentification rate for b jets to pass the loose and medium working points of the DEEPJET flavor-tagging classifier are found to be slightly different between data and simulation. Simulated samples are corrected by applying the *per-event* scale factor (SF), defined in equation 6.5. As done in the 2016 analysis, the tagging efficiencies and mistagging probabilities are measured in simulated $t\bar{t}$ and QCD multijet events, respectively, and are shown in Fig. 7.12.

- **top quark tagging efficiency and misidentification rate:** The top quark tagging efficiency and misidentification rate are both measured in a sample dominated by $t\bar{t}$ events where one top quark decays leptonically and the other hadronically. Events are required to have one high-quality muon with $p_T > 50$ GeV and $|\eta| < 2.4$ and at least four small-cone jets with $p_T > 40$ GeV and $|\eta| < 2.4$. At least one of the selected jets is required to be tagged as a b jet with the medium working point of the DEEPJET algorithm. The b jet closest to the muon satisfying $\Delta R(\mu, b \, jet) < 1.5$, is considered as the one originating from the leptonic decay. The hadronically decaying top quark is reconstructed by selecting the three-jet combination with a mass closest to the top
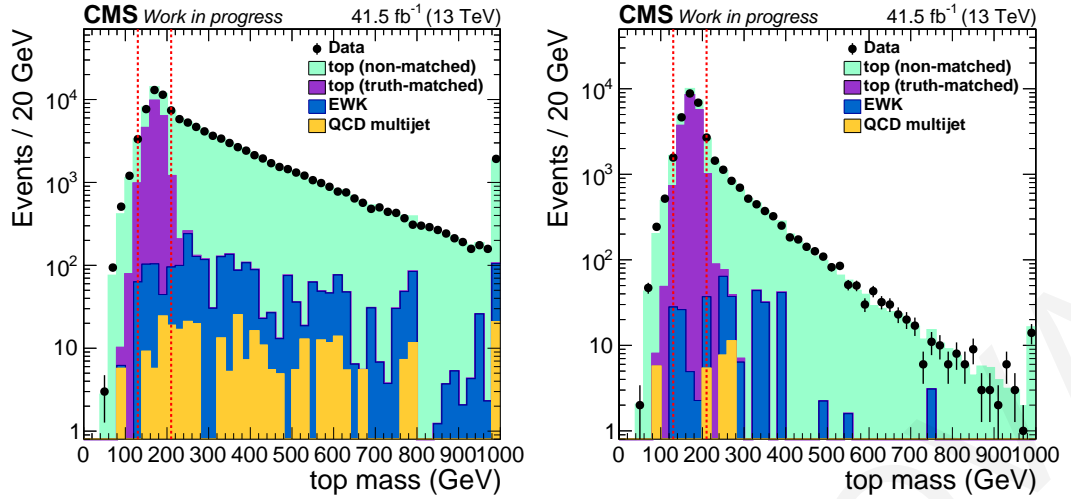
Figure 7.13: The invariant mass of the hadronically decaying top quark candidate before (*left*) and after top tagging with the Loose working point (*right*).

quark mass and satisfying $\Delta R(\mu, t) > 2.0$. The E_T^{miss} and the muon isolation are then used to define two regions:

- Background Region (BR): low $E_T^{\text{miss}} < 50$ GeV and loose isolation > 0.1, and

- Signal Region (SR): high $E_T^{\text{miss}} > 50$ GeV and tight isolation < 0.1,

which are both used to derive the normalization factors of $t\bar{t}$ ($f_{t\bar{t}}$) and QCD multijet (f_{QCD}) events, as done in 2016 analysis.

The invariant mass of the hadronically decaying top quark candidate in the SR before and after top tagging is shown in Fig. 7.13. The dashed red lines separate the phase-space regions used for the top quark tagging misidentification rate and efficiency measurements.

The top tagging misidentification rate is measured in the SR for events in which the top quark candidate has an invariant mass that falls outside of the top quark mass window of 130 to 210 GeV, and thus dominated by combinatorial background. It is defined as the ratio of events where the hadronically decaying top quark candidate passes a given working point over all events as follows:

$$\text{mistag}_{data} = \frac{N_{data}^{passed}}{N_{data}^{all}}, \qquad \text{mistag}_{\text{MC}} = \frac{N_{f_{\text{QCD}} \times \text{QCD} + f_{t\bar{t}} \times t\bar{t} + \text{EWK}}^{passed}}{N_{f_{\text{QCD}} \times \text{QCD} + f_{t\bar{t}} \times t\bar{t} + \text{EWK}}^{all}}. \tag{7.1}$$

The misidentification rate as measured in data and simulation is shown in Fig.7.14 as a function of the top quark candidate's p_T, for the loose and medium working points. Since the misidentification rate in simulation is slightly different than the one estimated in data, scale factors are extracted as a function of the top quark candidate p_T to correct the simulation.
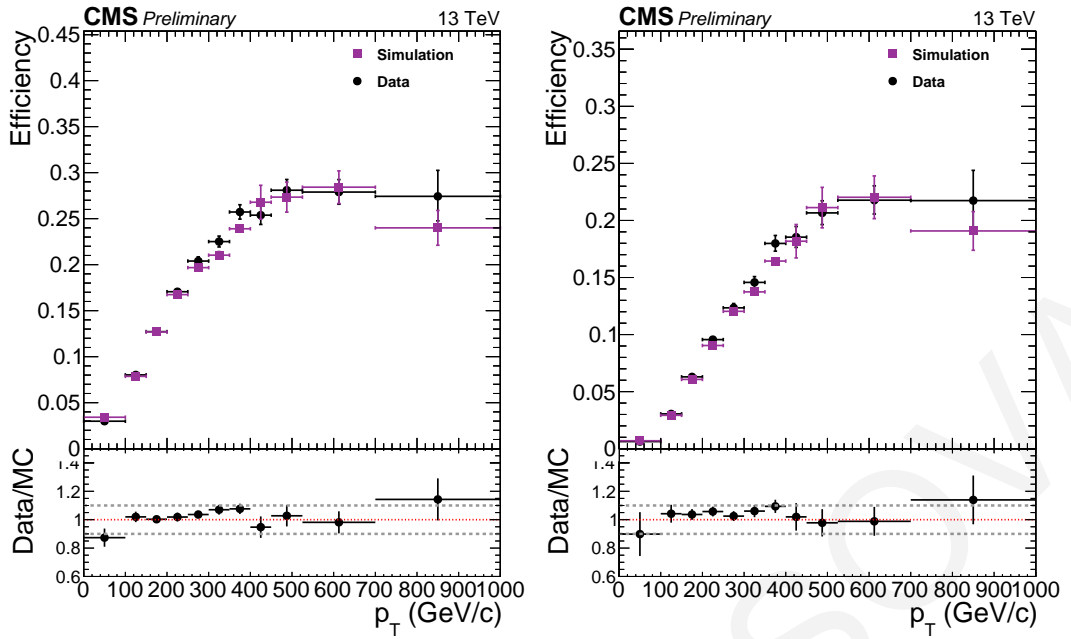
Figure 7.14: Misidentification rate in data and simulated events for the loose (*left*) and medium (*right*) working points, as a function of p_T of the top quark candidates.

The top quark tagging efficiency is measured in data and simulation in the SR for events in which the hadronically decaying top quark candidate has an invariant mass that falls inside the top mass window. In simulated events, the top quark tagging efficiency is measured using events where the tagged top quark candidate is truth-matched at the generator-level. In data, the top quark tagging efficiency is estimated by subtracting background events from non-top processes (EWK and QCD) and events with misidentified top quark candidates from $t\bar{t}$, with the latter defined as top quark candidates failing the generator-matching criteria. The subtracted events are corrected with the misidentification rate scale factors as follows:

$$\varepsilon_{data} = \frac{N^{passed}_{data-\text{SF}_{\text{mistag}}\times(f_{\text{QCD}}\times\text{QCD}+f_{t\bar{t}}\times t\bar{t}_{\text{non-matched}}+\text{EWK})}}{N^{all}_{data-f_{\text{QCD}}\times\text{QCD}-f_{t\bar{t}}\times t\bar{t}_{\text{non-matched}}-\text{EWK}}}, \quad \varepsilon_{\text{MC}} = \frac{N^{passed}_{t\bar{t}_{\text{matched}}}}{N^{all}_{t\bar{t}_{\text{matched}}}}. \tag{7.2}$$

The top quark tagging efficiency as a function of the p_T of the top quark candidates is shown in Fig. 7.15 for data and simulated events, for the loose and medium working points. The top quark tagging efficiency in simulated events agrees with the one obtained from data within 10%, excluding the bins with low statistics. To correct the simulation, scale factors as a function of the top quark candidate p_T are derived from the data-to-simulation ratio.

7.4 Background measurement and validation

The composition of the main backgrounds in the two signal regions, $1\text{M}1\text{L}_t$ and 2M_t, are shown in Fig. 7.16, as a function of the reconstructed $m_{H^\pm}$. In the $1\text{M}1\text{L}_t$ category, the
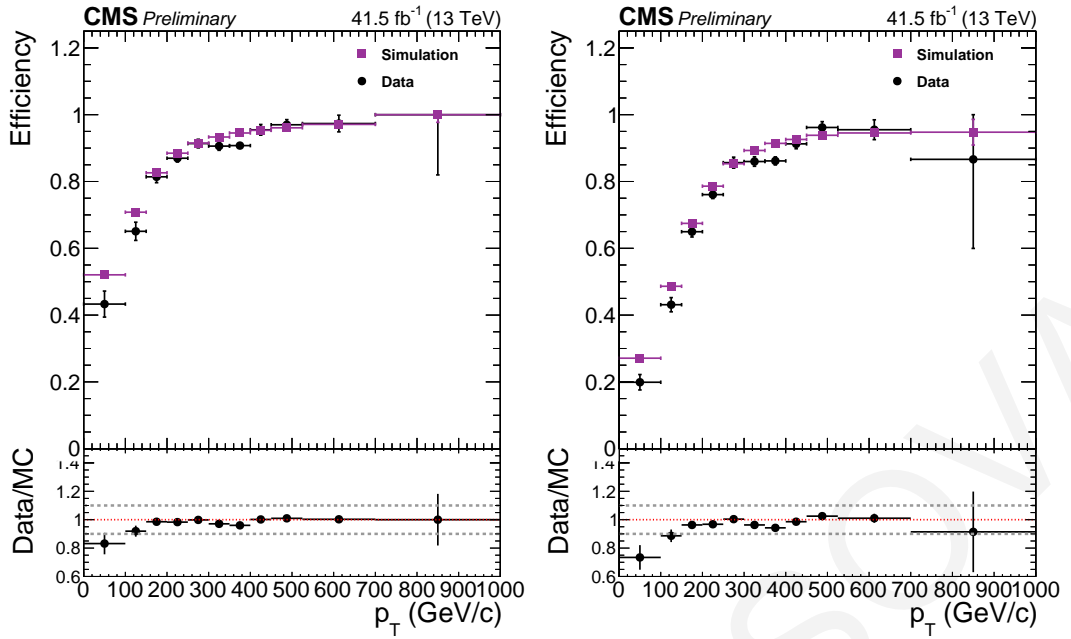
Figure 7.15: Top quark tagging efficiency in data and simulation, for the loose (*left*) and medium (*right*) working points.

QCD multijet background comprises about 67% of the total background, while the remaining background is mostly $t\bar{t}$ events. In the $2M_t$ category, equal amounts of QCD multijet and $t\bar{t}$ events are present. The background measurement follows the same principles as in the 2016 analysis. *Genuine b jets* events are the events where all selected b jets are genuine as determined by the hadron-flavor hypothesis. Events with at least one non-genuine b jet are labeled as *misidentified b jets* events. Genuine b jets events are modeled using simulation while misidentified b jets events are estimated with a data-driven technique similar to the one used in the 2016 analysis. The main difference is in the choice of the top quark candidate that is being inverted and used in the definition of the orthogonal control regions. In the 2016 analysis, the inverted top quark candidate is the one with the lowest MVA score between the two top quark candidates, which can be either the top from the charged Higgs boson decay, or the associated top. For this search, the inverted top is the one assigned to the associated top, as illustrated in Fig. 7.17. The motivation behind this change is to ensure that the objects being inverted in the ABCD method do not belong in the same reconstructed object, in this case the charged Higgs boson candidate. The difference between inverting the associated top and the subleading-in-MVA score is studied and found to have a small impact on the extraction of the transfer factors. The difference is included as a systematic uncertainty in the misidentified b jets background measurement.

The correlation of the two variables used in the ABCD method is also studied. Figure 7.18 shows the DNN score of the associated top quark candidate in bins of the b discriminator of the b jet from the charged Higgs boson decay in data (*left*) and simulated $t\bar{t}$ (*right*) events. The ratio shown below the various distributions corresponds to the ratio between each distribution and the one corresponding to the lowest b discriminator bin (0.0532, 0.1048). No
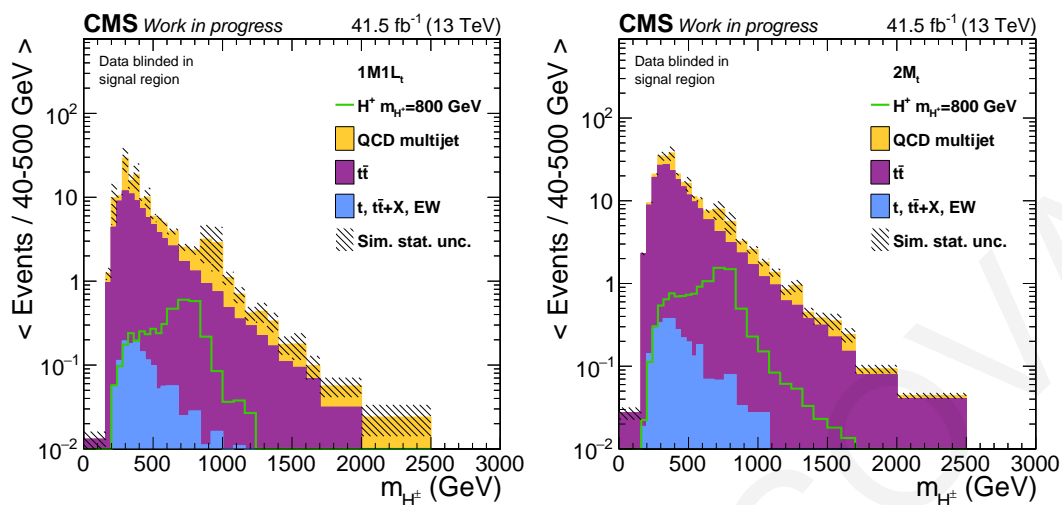
125

Figure 7.16: Distributions of the charged Higgs boson invariant mass in the background-only hypothesis for the $1\mathrm{M}1\mathrm{L}_\mathrm{t}$ (*left*) and $2\mathrm{M}_\mathrm{t}$ (*right*) categories. The signal is also shown for $\mathrm{m}_{\mathrm{H}^\pm}$ = 800 GeV, normalized to $\sigma\mathcal{B} = 1$ pb.
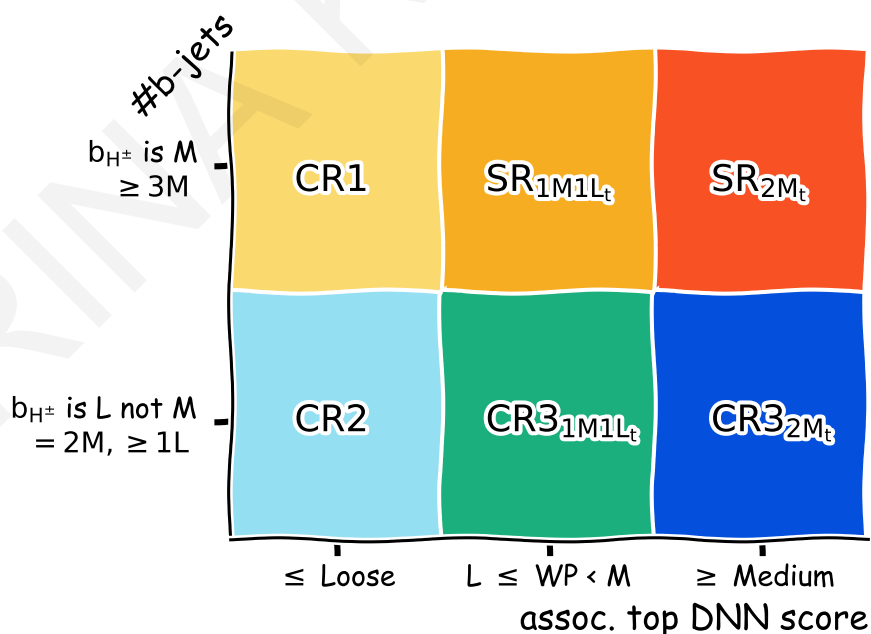

Figure 7.17: Schematic diagram of the *ABCD* method used. The $x-$axis corresponds to the DNN score of the associated top quark candidate, while the $y-$axis to the b discriminator of the b jet used in the reconstruction of the charged Higgs boson candidate.
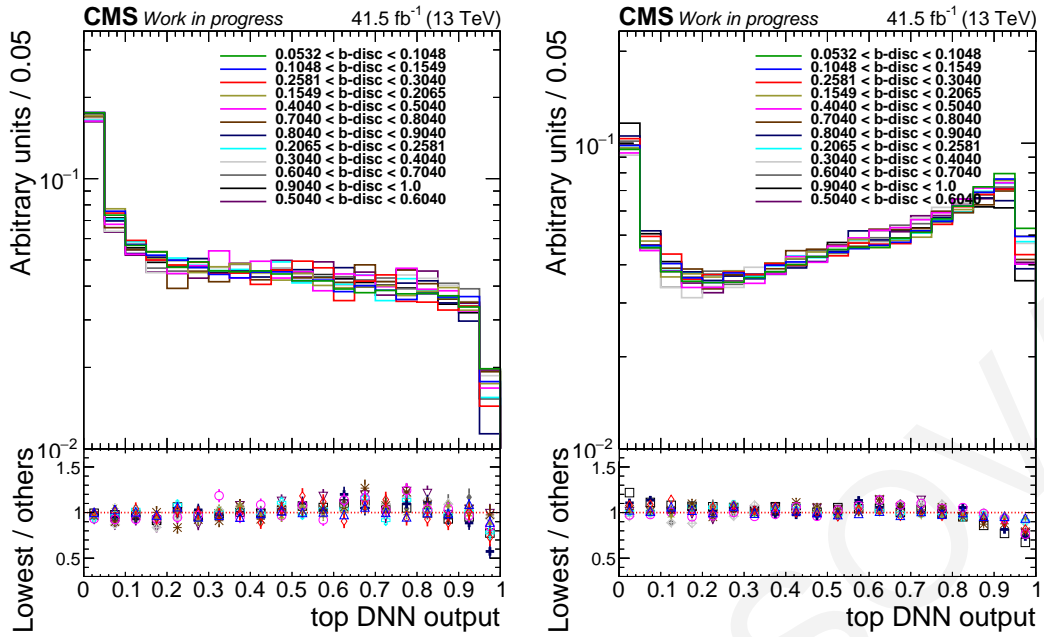
Figure 7.18: The DNN score of the associated top quark candidate for different bins of the b discriminator of the b jet coming from the charged Higgs boson decay, for data (*left*) and simulated $t\bar{t}$ (*right*) events.

strong correlation is observed between the two variables.

As it was done in the 2016 analysis, the CR1 and CR2 are used to extract transfer factors to normalize the expected yield of misidentified b jets found in $\mathrm{CR3}_{1\mathrm{M1L_t}}$ ($\mathrm{CR3}_{2\mathrm{M_t}}$) to the one in $\mathrm{SR}_{1\mathrm{M1L_t}}$ ($\mathrm{SR}_{2\mathrm{M_t}}$). To correct for kinematic differences between loose- and medium-tagged b jets, the transfer factors are calculated in a total of 35 bins of the inverted b jet p_T and $|\eta|$. The bins in p_T and $|\eta|$ are:

- p_T: < 50, 50-75, 75-110, 110-150, 150-200, 200-280, > 280 GeV

- $|\eta|$: < 0.7, 0.7-1.1, 1.1-1.5, 1.5-2.0, > 2.0.

Figure 7.19 shows the values of the transfer factors for all 35 bins used. The expected yield of misidentified b jets events in the two signal regions can be calculated by eq. 6.11. The same set of transfer factors, defined by eq. 6.12, are used for both categories $1\mathrm{M1L_t}$ and $2\mathrm{M_t}$.

7.4.1 Misidentified b jets self-closure test

The self-closure test performed in the 2016 analysis (section 6.4.2) is repeated for the 2017 analysis. The shapes of the H_T, the $E_\mathrm{T}^{\mathrm{miss}}$, and the invariant charged Higgs boson mass for misidentified b jets events are shown in Fig. 7.20. The MVA output of the parameterized DNN is also shown for three mass points in Fig. 7.21. All distributions are normalized to unity and show good agreement within statistical uncertainties.
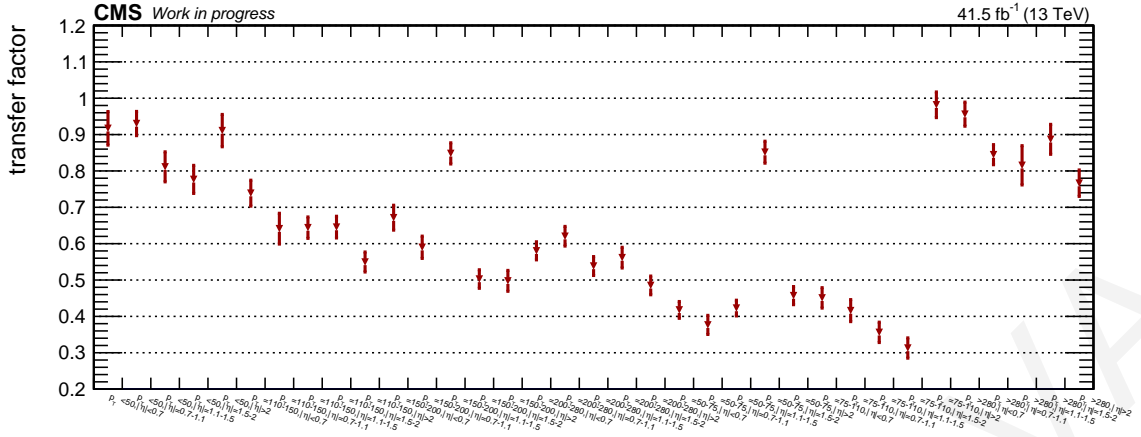
Figure 7.19: The values of the transfer factors used for the normalization of $\mathrm{CR3}_{1\mathrm{M}1\mathrm{L}_{\mathrm{t}}}$ and $\mathrm{CR3}_{2\mathrm{M}_{\mathrm{t}}}$.
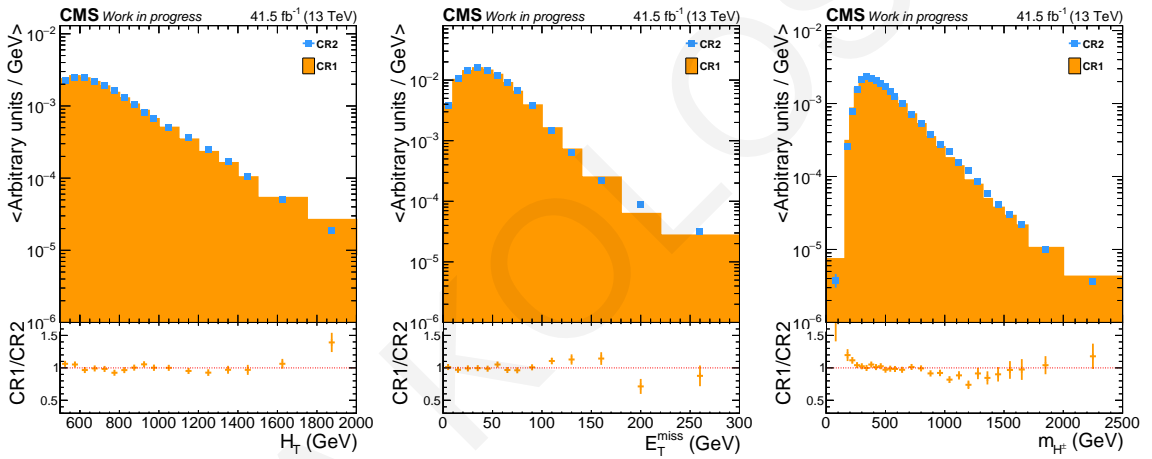

Figure 7.20: The normalized-to-unity shapes of H_{T} (*left*), $E_{\mathrm{T}}^{\mathrm{miss}}$ (*middle*) and the reconstructed invariant charged Higgs candidate mass (*right*) for misidentified b jets events in CR1 and CR2.
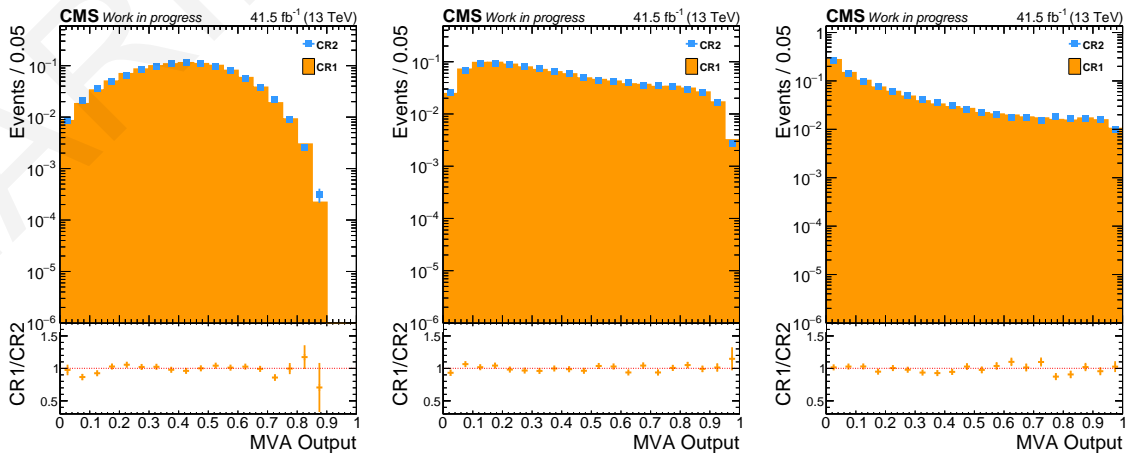

Figure 7.21: The normalized-to-unity shapes of the MVA output of the parameterized DNN, for mass points $\mathrm{m}_{\mathrm{H}^{\pm}} = 250$ GeV (*left*), $\mathrm{m}_{\mathrm{H}^{\pm}} = 600$ GeV (*middle*) and $\mathrm{m}_{\mathrm{H}^{\pm}} = 1000$ GeV (*right*) for misidentified b jets events in CR1 and CR2.
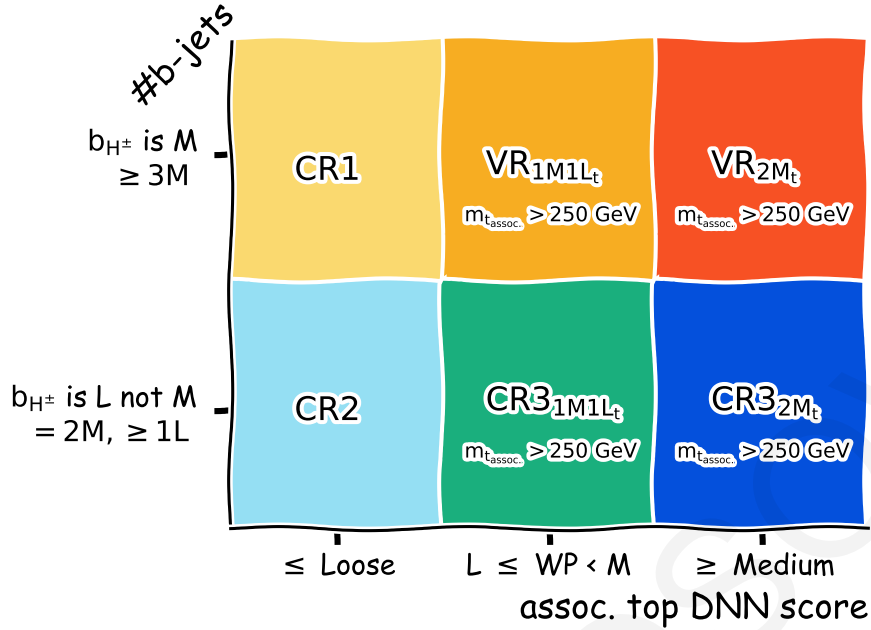
Figure 7.22: Schematic diagram of the ABCD method in the closure region, defined by inverting the associated top quark candidate mass ($\mathrm{m_{top}} > 250$ GeV).

7.4.2 Misidentified b jets closure in orthogonal region

The method used for the estimation of the misidentified b jets yields is validated in a *Closure Region* (CR), defined by inverting the mass of the associated top quark candidate, as shown in Fig. 7.22. The associated top quark candidate is required to have a mass above 250 GeV, instead inside the top mass window of 120 to 230 GeV. The $\mathrm{CR1}$ and $\mathrm{CR2}$ are unchanged and so the same transfer factors are used to estimate the yield of misidentified b jets in the two Verification Regions (VR), $\mathrm{VR_{1M1L_t}}$ and $\mathrm{VR_{2M_t}}$. The number of misidentified b jets events expected in the VRs can be calculated by:

$$N_{\mathrm{VR}}^{\text{misidentified b jets}} = \sum_{i}^{\text{inverted b jet } p_{\mathrm{T}}, |\eta| \text{ bins}} \left(N_{\mathrm{CR3},i}^{\text{data}} - N_{\mathrm{CR3},i}^{\text{EWK Genuine b jets}}\right) \times \mathrm{TF}_i, \qquad (7.3)$$

where TF_i are the transfer factors given by Equation 6.12. Figure 7.23 (7.24) show the mass, DNN score, and p_{T} of the $\mathrm{H}^{\pm}$ top (associated top) in the region $\mathrm{VR_{1M1L_t}}$. The distribution corresponding to the expected yield of misidentified b jets events is shown with yellow, while the distributions of Genuine b jets $\mathrm{t\bar{t}}$ and Single-t, $\mathrm{t\bar{t}} + \mathrm{X}$, and EWK events are shown with purple and light blue, respectively. The observed data events in the closure region are shown in black. The distributions include statistical uncertainties only. Good agreement is observed between the data and the simulation, which verifies the validity of the method. The corresponding distributions in the region $\mathrm{VR_{2M_t}}$ are shown in Fig. 7.25 and Fig. 7.26.

Figure 7.23: The distributions of the mass (*left*), top DNN score (*middle*), and p_T (*right*) of the $H^{\pm}$ top in the VR_{1M1L_t} region.

Figure 7.24: The distributions of the mass (*left*), top DNN score (*middle*), and p_T (*right*) of the associated top in the VR_{1M1L_t} region.
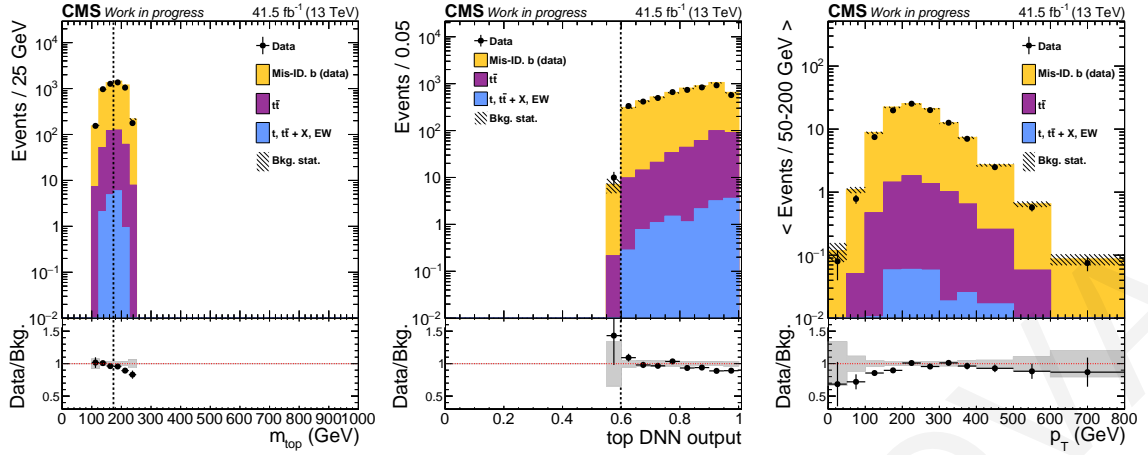

Figure 7.25: The distributions of the mass (*left*), top DNN score (*middle*), and p_T (*right*) of the $H^{\pm}$ top in the VR_{2M_t} region.
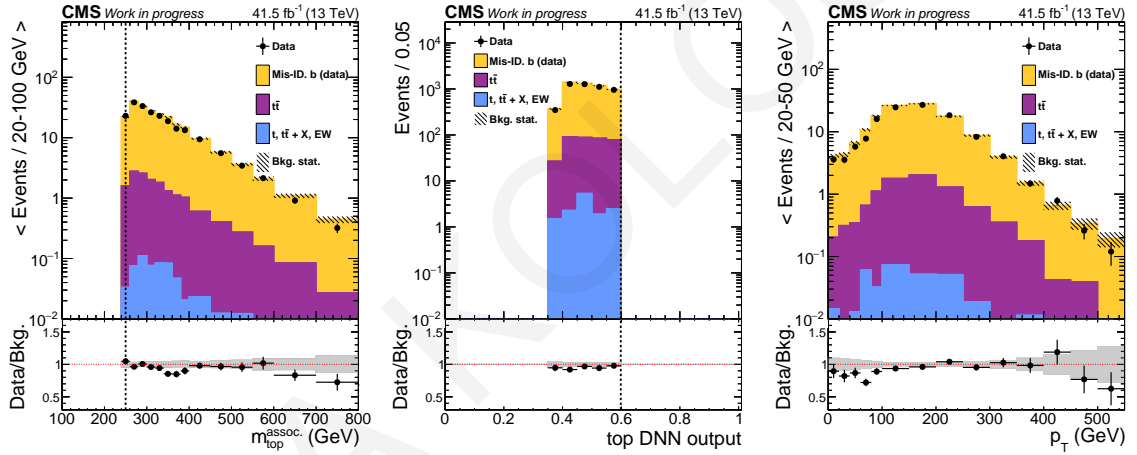
Figure 7.26: The distributions of the mass (*left*), top DNN score (*middle*), and p_T (*right*) of the associated top in the VR_{2M_t} region.
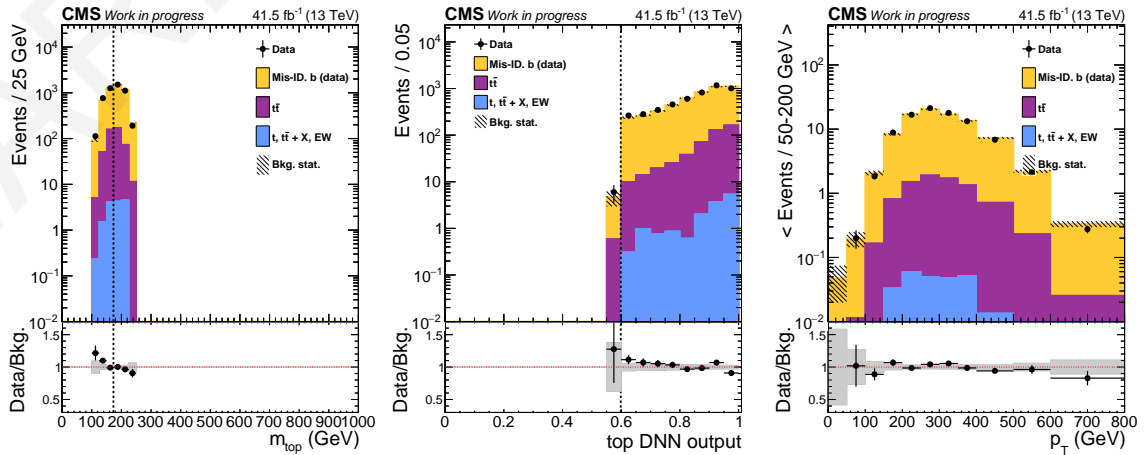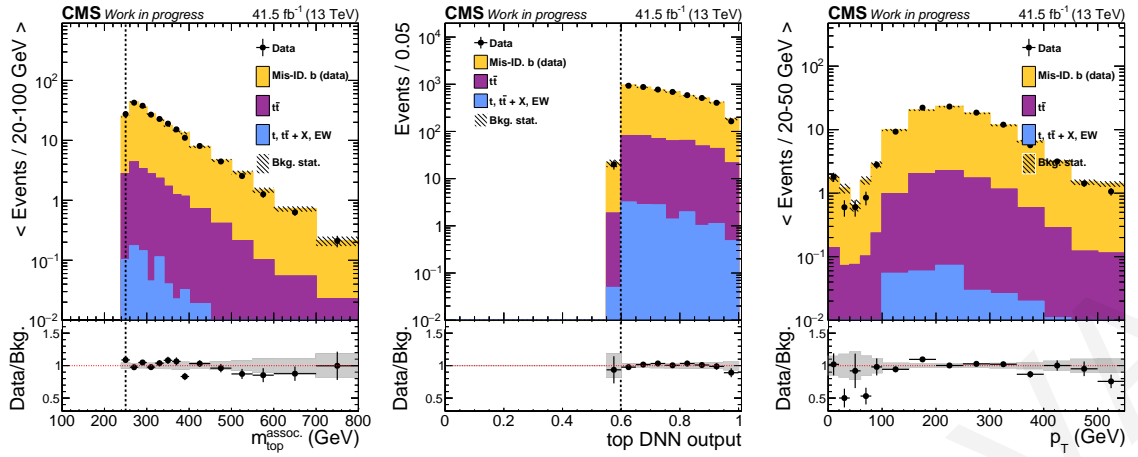
7.4.3 Misidentified b jets closure in simulation

The estimation of the misidentified b jets yields in the signal region is additionally validated using simulated events only. Figures 7.27 and 7.28 show the comparison between misidentified b jets events in the signal region (composed of simulated QCD-multijet and EWK misidentified b jets events) and the estimated misidentified b jets events, following the ABCD method, for categories $1M1L_t$ and $2M_t$, respectively. The plots include statistical uncertainties only and the spikes observed originate from low H_T bins of the QCD-multijet process which correspond to high cross-sections. Overall, good agreement is observed.

7.5 Event-based and parameterized DNN

To further discriminate signal from background events an *event-based* and *parameterized* DNN has been developed. Parameterized neural networks [166] are built such that their input variables include not only event-level reconstructed features, but also one or more physics parameters, such as the mass of a new particle. This approach allows for a single parameterized network to provide improved discrimination across a wide range of masses. Moreover, it provides a smooth interpolation of masses that were not used in the training of the network. Other methods widely used require either the training of a set of standalone networks, one for each representative physics parameter, each of which is ignorant of the others, or the training on a sample with a single physics parameter. The former, however, lacks the ability to interpolate between representative physics parameters while the latter results in the optimal performance for samples characterized by the particular physics parameter used in the training but has degrading performance on samples with different values of the physics parameter. For this analysis, the parameterized neural network takes as input a vector ($\vec{x}$) of twelve parameters, described in section 7.5.2, and the *true mass hypothesis* of the charged Higgs boson, θ.

Figure 7.27: The distributions of $E_{\mathrm{T}}^{\mathrm{miss}}$ (a), $\mathrm{H}^{\pm}$ p_{T} (b), the leading in p_{T} top candidate p_{T} (c) and the leading in p_{T} free b jet p_{T} (d) for misidentified b jets events in the signal region using simulated events only for category $1\mathrm{M}1\mathrm{L}_{\mathrm{t}}$.
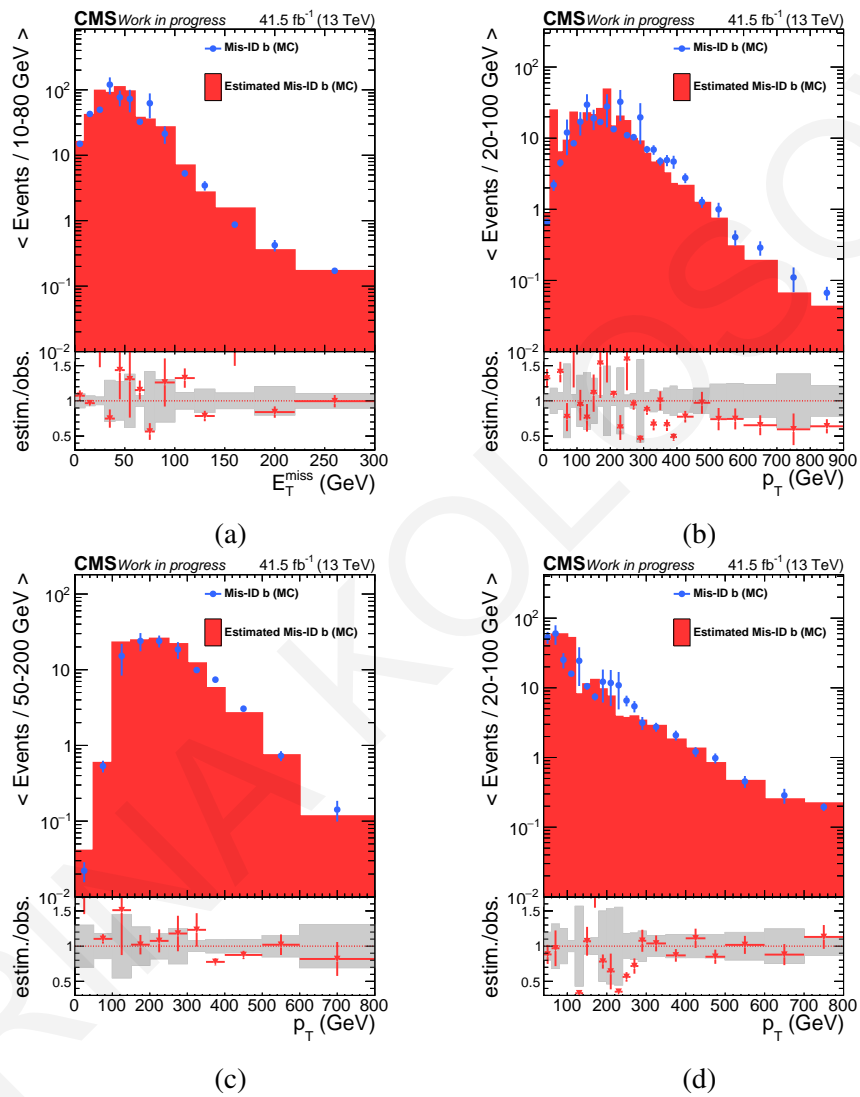
Figure 7.28: The distributions of E_T^{miss} (a), $H^\pm$ p_T (b), the leading in p_T top candidate p_T (c) and the leading in p_T free b jet p_T (d) for misidentified b jets events in the signal region using simulated events only for category $2M_t$.
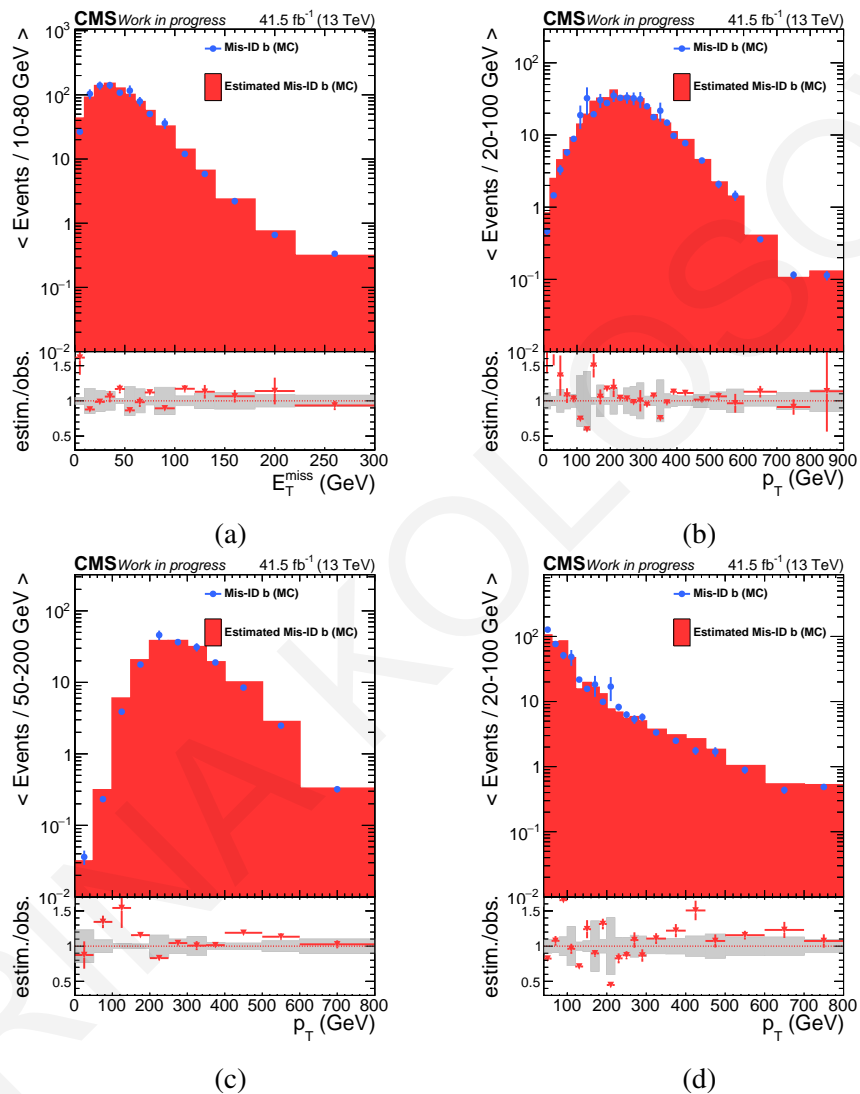
7.5.1 Training datasets

Four charged Higgs boson mass hypotheses are used for the signal training sample: 220, 350, 600, and 1000 GeV. The mass points are chosen to be spread in the range where the resolved analysis is most sensitive to, and to represent the kinematic characteristics of the low, intermediate, and high mass regions. Events from $t\bar{t}$ simulation are used as background and are classified as follows:

- *Genuine b jets events:* events passing all signal selections in both $1\mathrm{M}1\mathrm{L_t}$ and $2\mathrm{M_t}$ categories.

- *Misidentified b jets events*: events passing the selections of the $1\mathrm{M}1\mathrm{L_t}$ and $2\mathrm{M_t}$ CR3. These are events in which the mass of both top quark candidates fall inside the top quark mass window (120 to 230 GeV) and at least one b-tagged jet (loose or medium-tagged) is not genuine. Events are labeled as $t\bar{t}$-like if the jets of at least one top quark candidate are truth-matched with partons. Otherwise, they are labeled as QCD-like events. Additional QCD-like events are obtained from control regions similar to CR3 (using simulated $t\bar{t}$ events), with the only difference being that both candidates have mass outside the top quark mass window. This is needed as the QCD multijet simulated samples have low statistics after all analysis selections are applied.

To have a background composition similar to the one expected in the analysis, the amount of Genuine b jets and misidentified b jets events, as well as the amount of misidentified b jets events that are $t\bar{t}$-like and QCD-like are set to match to the expectations in the signal region. Furthermore, to take into account the kinematic differences between medium- and loose-tagged b jets, the misidentified b jets events used in the training are corrected by applying the transfer factors described in section 7.4. Since there is no *true* charged Higgs boson mass to assign for the background events, a random value among the ones used for the signal is assigned.

7.5.2 Input variables

The parameterized DNN takes as input twelve in total event-level features including event shape observables, top quark candidate multiplicities, masses, kinematic and angular separation variables between the reconstructed objects. All event-level features are illustrated in Fig. 7.29 for the four signal mass points used for the training and the background.

- Event shape observables [167, 168] are used to describe the origin, geometric patterns, and correlations of the energy flow in the final state. They have played a crucial role in testing QCD as they have been used in many measurements of the strong coupling constant α_s, including [169, 170, 171, 172], and to constrain color factors for quark and gluon couplings [173]. Moreover, they have been used to tune the parton showers

and non-perturbative components of Monte Carlo (MC) generators [174]. At the LHC, event shape variables have been measured at $\sqrt{s} = 13$ TeV in multijet events [175, 176] to assess the accuracy of theoretical predictions from MC event generators. Event shape variables can also be used to probe for processes beyond-the-SM by defining sensitive phase-space regions, often characterized by isotropic multijet events [177]. For this search, the event shape variables: sphericity, aplanarity, and circularity have been found to show some discrimination with respect to the background, as shown in Figs. 7.29 a, b, and c, respectively. The quadratic momentum tensor [178] for an event is given by:

$$
\mathcal{M}_{\alpha\beta} = \frac{\sum_i^{jets} p_{i,\alpha} p_{i,\beta}}{\sum_i^{jets} |\vec{p}_i|^2} = \frac{\sum_i^{jets}}{\sum_i^{jets} |\vec{p}_i|^2} \begin{pmatrix} p_{ix}^2 & p_{ix}p_{iy} & p_{ix}p_{iz} \\ p_{iy}p_{ix} & p_{iy}^2 & p_{iy}p_{iz} \\ p_{iz}p_{ix} & p_{iz}p_{iy} & p_{iz}^2 \end{pmatrix}, (\alpha, \beta = x, y, z).
$$

(7.4)

It has three eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3$, satisfying $\lambda_1 + \lambda_2 + \lambda_3 = 1$. The *sphericity* of an event is then calculated by the second and third eigenvalues of $\mathcal{M}_{\alpha\beta}$:

$$
S = \frac{3}{2}(\lambda_2 + \lambda_3),
$$

(7.5)

and takes values between 0 and 1. It is a measure of the summed p_{T}^2 with respect to the event axis. The event axis is defined as the line connecting the interaction point and the eigenvector associated with the eigenvalue with the largest value, λ_1. An event with just two balanced jets corresponds to $\mathrm{S} = 0$. Values close to unity indicate more isotropic events. The *aplanarity* of an event is defined as:

$$
A = \frac{3}{2}\lambda_3
$$

(7.6)

and takes values from 0 to $1/2$. It is a measure of the amount of the p_{T} in and out of the plane formed by the two first eigenvectors of $\mathcal{M}$. Aplanarity values close to 0 correspond to planar events, while values close to $1/2$ correspond to isotropic events. The 2-dimensional quadratic momentum tensor is calculated by:

$$
^{2D}\mathcal{M}_{\alpha\beta} = \frac{\sum_i^{jets} p_{i,\alpha} p_{i,\beta}}{\sum_i^{jets} |\vec{p}_i|^2} = \frac{\sum_i^{jets}}{\sum_i^{jets} |\vec{p}_i|^2} \begin{pmatrix} p_{ix}^2 & p_{ix}p_{iy} \\ p_{iy}p_{ix} & p_{iy}^2 \end{pmatrix}, (\alpha, \beta = x, y)
$$

(7.7)

and has eigenvalues $\lambda_1 > \lambda_2 > 0$. The *circularity* of an event is defined as:

$$
\mathcal{C} = 2 \times \min\left(\frac{\lambda_1, \lambda_2}{\lambda_1 + \lambda_2}\right).
$$

(7.8)

It is independent from boosts along the $z-$axis and takes values from 0 to 1. Values close to 0 correspond to linear events while values close to unity reveal circular events.

- The number of medium-tagged top quark candidates is used as an input variable so that discrimination between the $1M1L_t$ and $2M_t$ categories is achieved.

- The scalar sum of the p_T of the b jets used for the reconstruction of the two top quark candidates and the b jet from the $H^\pm$ decay: $H_{T,b}$

- The third jet resolution defined as: $y_{23} = \frac{p_{T,3}^2}{(p_{T,1}+p_{T,2})^2}$

- The invariant mass of the $bb-$pair system with the maximum p_T: $M(bb_{max\ p_T})$

- The p_T of the $bb-$pair system with the minimum ΔR: $p_T(bb_{\Delta R\ min})$

- The p_T of the $H^\pm$ b jet over the $H_{T,b}$: $p_{T,b_{H^\pm}}/H_{T,b}$

- The reconstructed invariant mass of the charged Higgs boson candidate, $m_{H^\pm}$

- The angular separation between the top from $H^\pm$ and the b jet from $H^\pm$, in the $H^\pm$ center-of-mass frame: $\Delta\theta(t_{H^\pm}, b_{H^\pm})$

- The absolute difference of the p_T between the $H^\pm$ and the b jet from $H^\pm$, over their sum: $|H^\pm p_T - b_{H^\pm} p_T|/(H^\pm p_T + b_{H^\pm} p_T)$

The correlations between the variables are shown in Fig. 7.30 for the signal hypothesis of $m_{H^\pm}$ = 600 GeV. Most variables are uncorrelated to each other, meaning that each of them gives unique and independent information to the algorithm.

7.5.3 Architecture

The parameterized deep neural network is trained using the Keras [179] and TensorFlow [180] packages. The architecture contains three hidden dense layers with 32 neurons in each. All hidden layers use the Rectified Linear Units (ReLU) activation function, while the output unit uses the sigmoid activation function. A batch normalization [181] layer is added between all the dense layers to accelerate the training and improve the performance of the network. The batch normalization layer standardizes the input of each layer by setting the mean output close to 0 and the output standard deviation close to 1. The mathematic transformation is shown in equation 7.9, where β and γ are the learning parameters and ε is a small constant that prevents division with zero.

$$\text{output} = \frac{\text{batch} - mean(\text{batch})}{var(\text{batch}) + \varepsilon} \times \gamma + \beta \tag{7.9}$$

To minimize the binary cross-entropy loss function, the Adam optimizer [182] is used with a learning rate of 0.001. The training is stopped when the best area under the ROC curve is obtained. For orthogonality, the training and validation are performed with 2018 signal and background samples, while the testing is performed with the 2017 samples. The training dataset includes 244000 events, equally split between signal and background and between signal mass points. The validation (testing) dataset includes 105000 (366000) events.
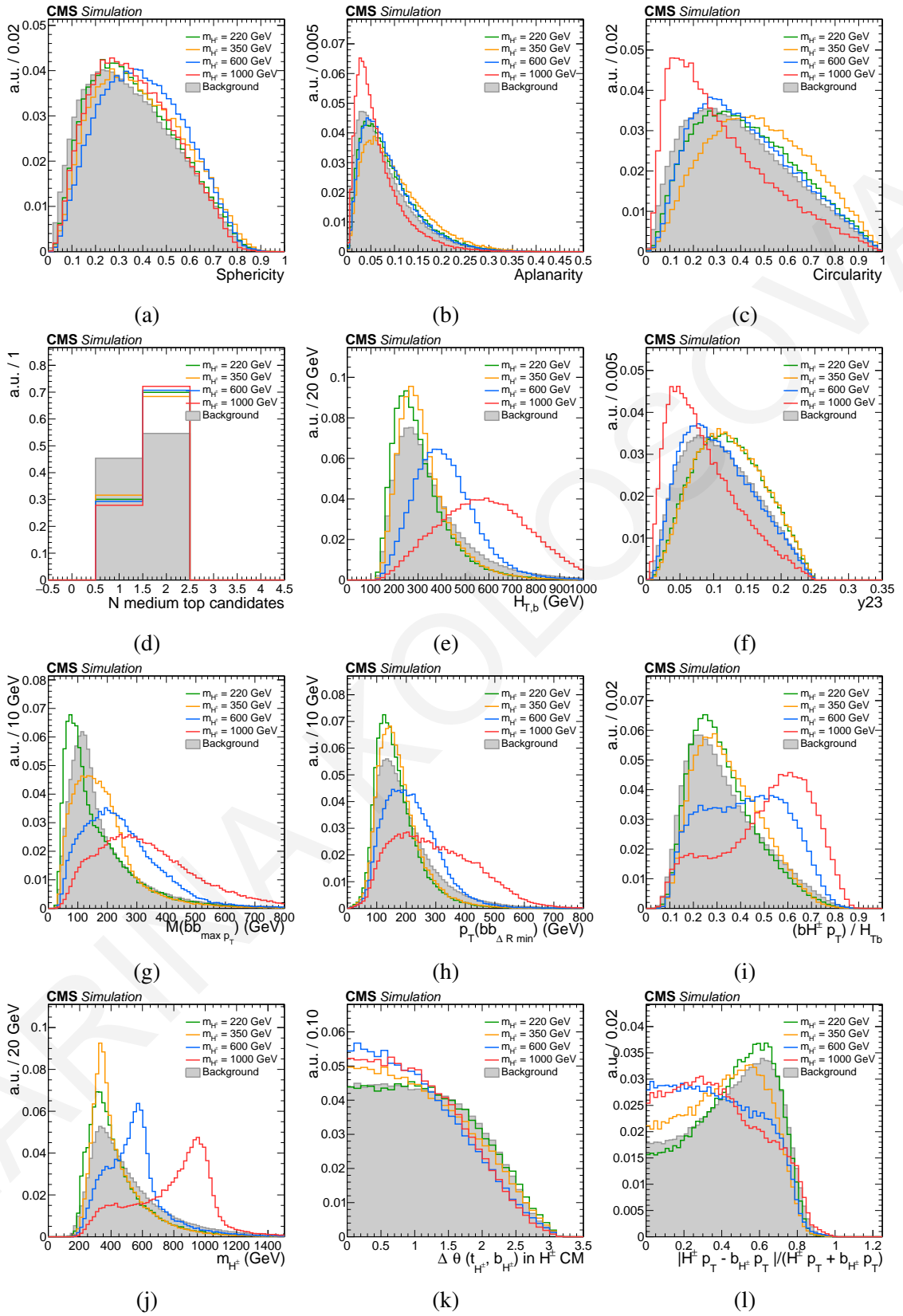
Figure 7.29: The twelve event-level reconstructed features used in the training of the parameterized DNN. The distributions of each feature are shown for the four signal mass points used in the training and for the background. All distributions are normalized to unity.
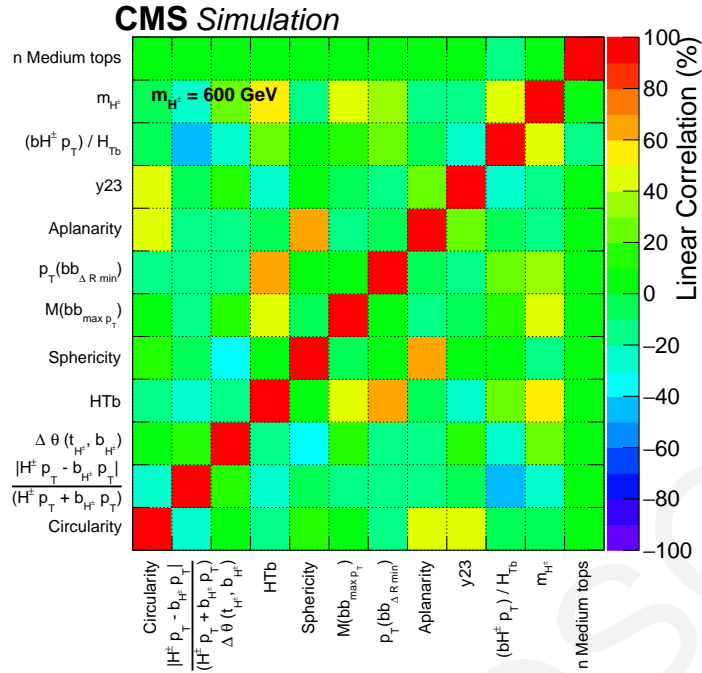
Figure 7.30: Correlation matrix of all event-level features for the signal hypothesis of $m_{\mathrm{H}^{\pm}}$ = 600 GeV.

7.5.4 Performance

The score of the parameterized DNN is evaluated by supplying the set of event-level reconstructed features, as well as the mass hypothesis of the charged Higgs boson under study, leading to different outputs distributions for both signal and background. The output of the parameterized DNN is shown in Fig. 7.31 for the true mass hypotheses of $m_{\mathrm{H}^{\pm}}$ =250, 350, 600, and 1000 GeV. The signal, shown with blue, has values closer to unity, while the background, shown with red has values closer to zero. The distributions with markers correspond to the signal and background events taken from the training sample, while the distributions with solid line correspond to the ones obtained from the testing sample. The training and testing distributions are compared as an over-fitting test. The agreement observed shows that the model is not over-trained and can predict well new data.

To test the ability of the parameterized DNN to interpolate between mass points that were not used in the training, the default neural network ROC curve is compared to an alternative network trained with mass points 220, 300, 500, and 800 GeV. The comparison is shown in Fig. 7.32 for signal masses from 200 up to 1500 GeV. Each ROC curve is evaluated at the true mass hypothesis. The two neural networks have comparable results and good prediction is observed even for mass values that were not used in the training of the networks, leading to the conclusion that the parameterized DNN can indeed interpolate across a wide range of masses.

The agreement between data and simulation is checked in the Closure Region, for the four mass points used in the training and is shown in Fig. 7.33 for the $1\mathrm{M}1\mathrm{L_t}$ category and in Fig. 7.34 for the $2\mathrm{M_t}$ category. Only statistical uncertainties are included in the plots.

(a) $m_{H^{\pm}} = 250$ GeV

(b) $m_{H^{\pm}} = 350$ GeV

(c) $m_{H^{\pm}} = 600$ GeV
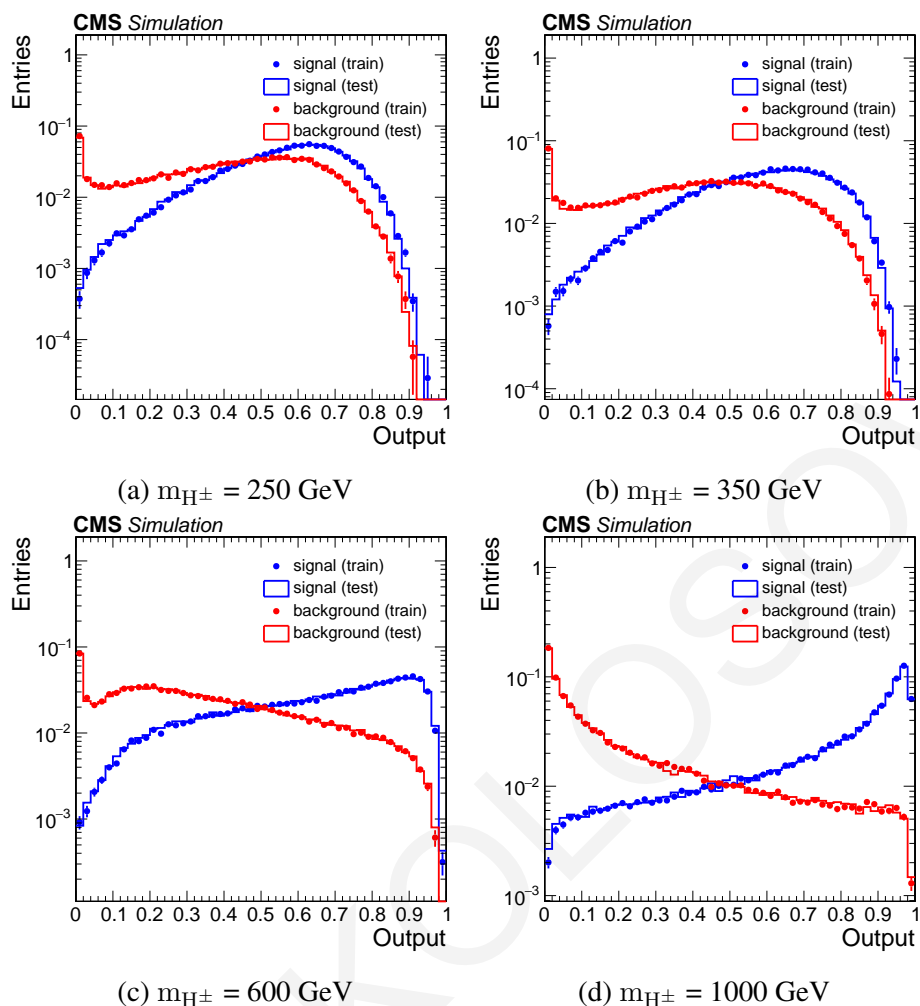
(d) $m_{H^{\pm}} = 1000$ GeV

Figure 7.31: The output of the parameterized DNN for the four signal mass points used in the training.
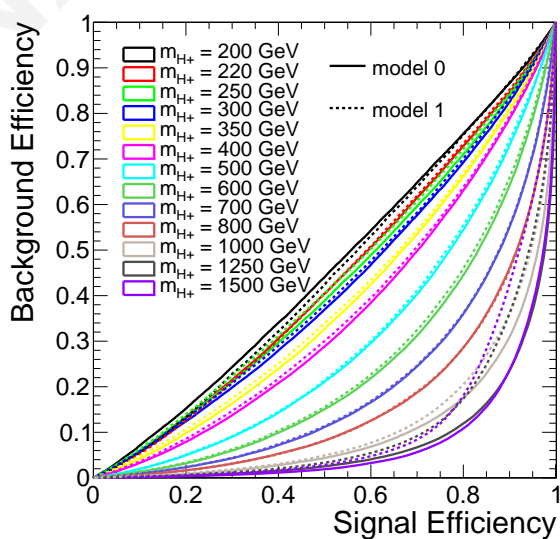

Figure 7.32: ROC curves evaluated at mass points between 200 GeV to 1.5 TeV for the default model (model 0) trained with masses 250, 350, 600 and 1000 GeV and alternative model (model 1) trained with masses 220, 300, 500 and 800 GeV.
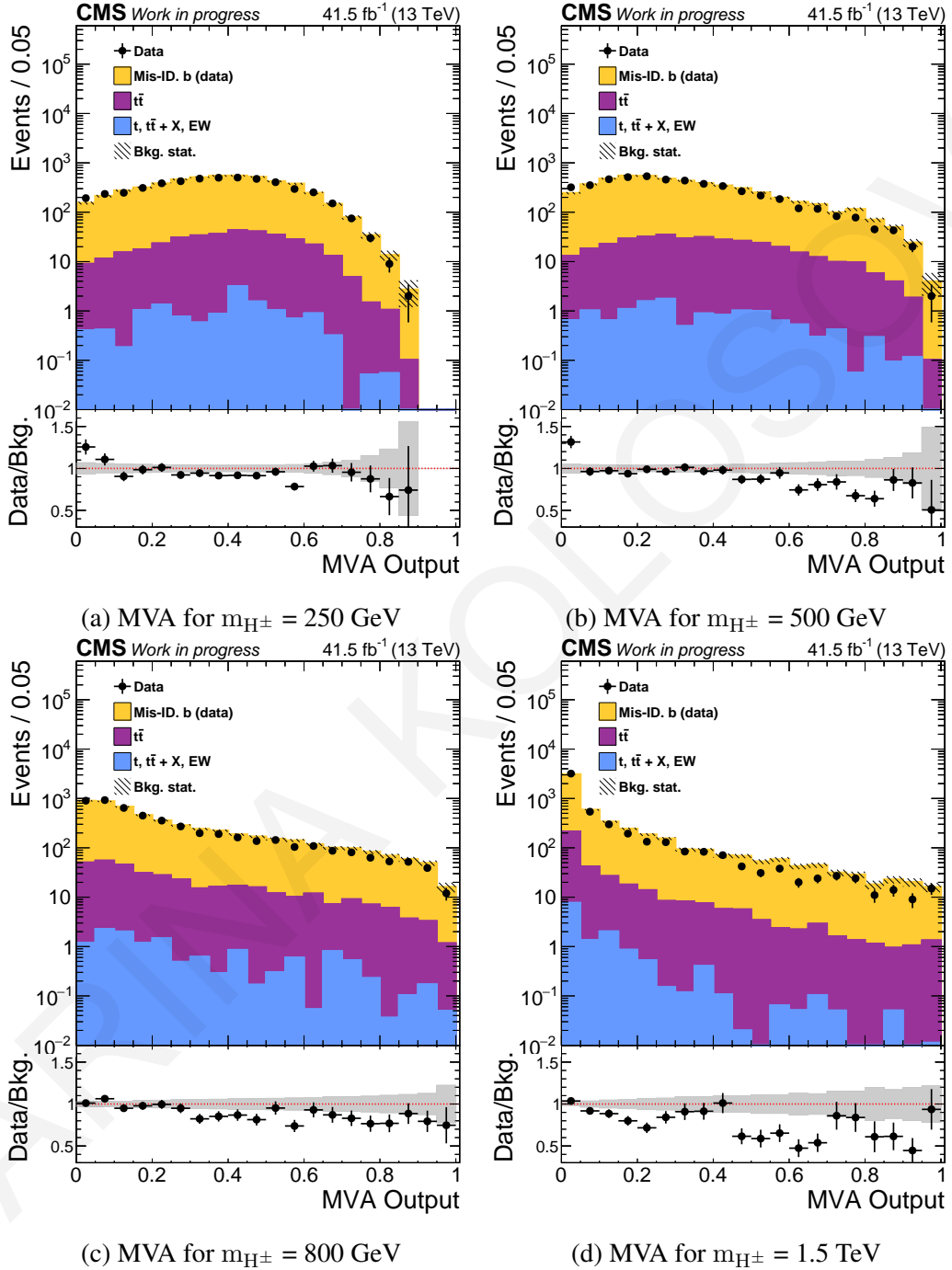
(a) MVA for $\mathrm{m}_{\mathrm{H}^{\pm}} = 250$ GeV

(b) MVA for $\mathrm{m}_{\mathrm{H}^{\pm}} = 500$ GeV

(c) MVA for $\mathrm{m}_{\mathrm{H}^{\pm}} = 800$ GeV

(d) MVA for $\mathrm{m}_{\mathrm{H}^{\pm}} = 1.5$ TeV

Figure 7.33: The MVA output of the parameterized DNN for four mass points, in the region $\mathrm{VR}_{1\mathrm{M}1\mathrm{L}_{\mathrm{t}}}$.

(a) MVA for $\mathrm{m_{H^{\pm}}} = 250$ GeV

(b) MVA for $\mathrm{m_{H^{\pm}}} = 500$ GeV

(c) MVA for $\mathrm{m_{H^{\pm}}} = 800$ GeV

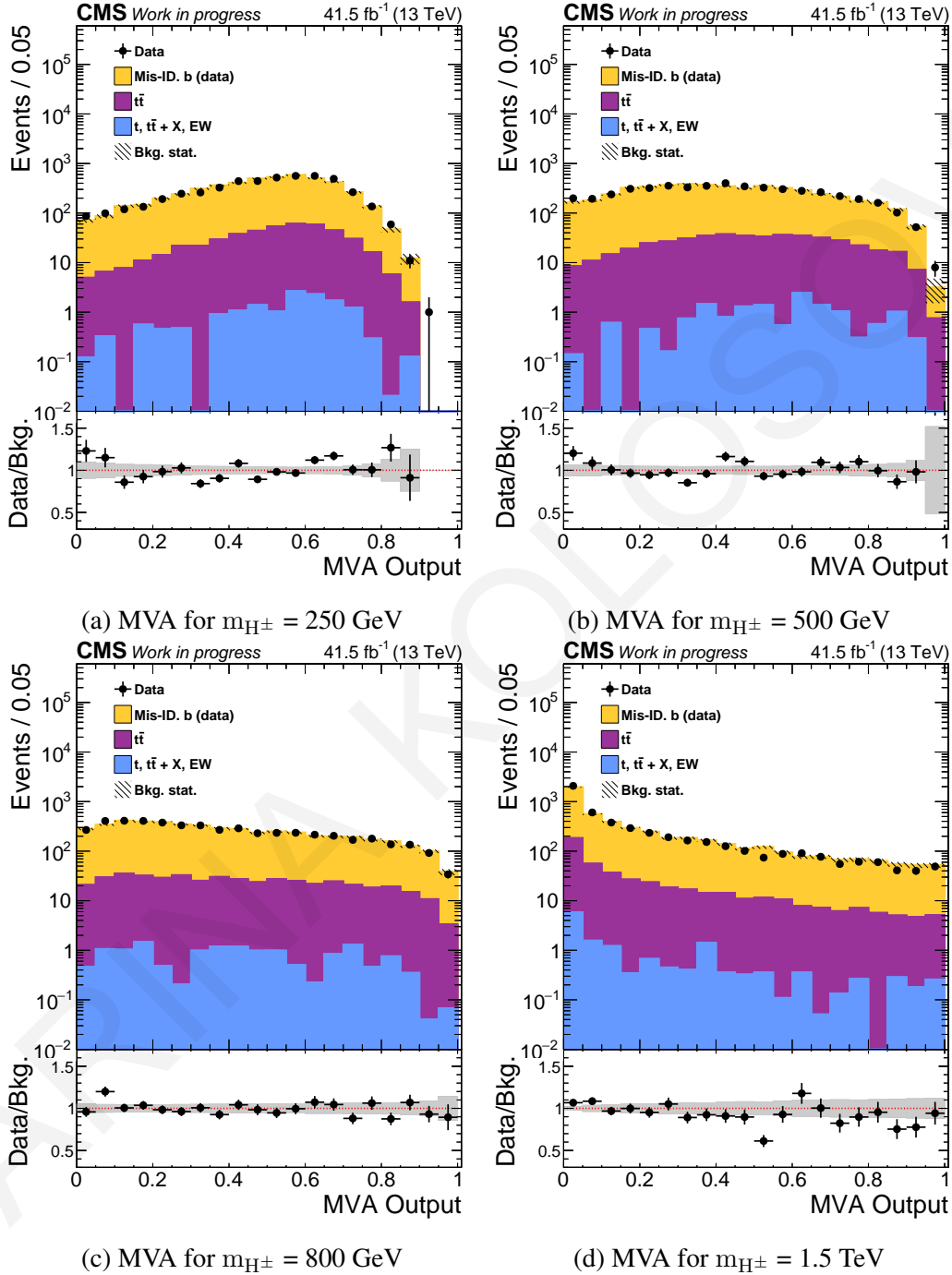(d) MVA for $\mathrm{m_{H^{\pm}}} = 1.5$ TeV

Figure 7.34: The MVA output of the parameterized DNN for four mass points, in the region $\mathrm{VR_{2M_t}}$.

7.6 Systematic uncertainties

7.6.1 Experimental uncertainties

- **Luminosity:** The total systematic uncertainty as measured in Ref. [183] for the 2017 LHC data-taking period is 2.3%. As done in the 2016 analysis, the uncertainty on the luminosity is applied as rate uncertainty to all simulated processes, and thus only affects the expected event yield and not the template shapes.

- **Pileup:** The uncertainty due to the pileup modeling is estimated by varying the total inelastic pp interaction cross-section (69.3 mb) by $\pm 5\%$ in all simulated samples. The effect is propagated to the final fit discriminant affecting both the shape and the normalization of the predictions.

- **Trigger efficiency:** The systematic uncertainties related to the trigger scale factors are estimated by varying them within their statistical uncertainties. The up and down uncertainties are propagated to the final fit discriminant and thus treated as both, rate plus shape nuisances.

- **Lepton and τ_h identification efficiencies:** The uncertainties corresponding to the identification and rejection of isolated electrons, muons, and hadronically decaying $\tau-$leptons are calculated from equation (6.15) and are treated as rate only nuisances.

- **Jet energy scale:** The systematic uncertainties related to the jet energy scale are estimated by shifting the energy of the jets up and down, with the variations being p_T and η dependent. The effect is propagated through all the steps of the analysis and result in rate and shape nuisances.

- **Jet energy resolution:** The uncertainties in the jet energy resolution are evaluated by varying up and down the jet energies by 1σ around their central value. The variations are jet p_T dependent and their effect is propagated to the final fit discriminant as both rate- and shape-altering nuisances.

- **b-tagging efficiency and misidentification rate:** The rate and shape uncertainties on the b-tagging efficiency and misidentification rate are evaluated by varying the corresponding simulation-to-data corrections within their uncertainties, as shown in equation (6.16)

- **top quark tagging efficiency and misidentification rate:** The systematic uncertainties related to the top quark tagging and misidentification rate are estimated by varying the corresponding simulation-to-data corrections within their uncertainties. These uncertainties have several sources, including the jet energy scale and resolution uncertainties, the definition of the genuine top quark candidate (truth-matching criteria) and

the top quark modeling (the damping of radiation with high p_T, the first emission modeling, the scale radiation, the color reconnection strength, the assumed value of the top quark mass, and the tuning of the underlying event parameters). These are propagated to the final fit discriminant as rate and shape uncertainties.

- **misidentified b jets background:** The systematic uncertainties on the measurement of the misidentified b jets background are estimated following the same approach as in the 2016 analysis. All systematic uncertainties due to the simulated EWK processes are found to be negligible as they are anti-correlated with the misidentified b jets background, and thus, not included in the final limit calculation. The statistical uncertainties on the derivation of the transfer factors are estimated by following the same procedure as in the 2016 analysis. The effects are propagated to the final fit discriminant, affecting both the shape and the rate of the misidentified b jets prediction. The systematic uncertainties on the derivation of the transfer factors are calculated by modifying the selection on the top mass window used, the definition of the control regions (CR1 and CR2), the top quark candidate that is being inverted in the ABCD method, and the composition of the Genuine b jets events that are subtracted from the data. A mapping between each transfer factor bin label and the $\mathrm{H}^{\pm}$ b jet p_T and $|\eta|$ selection is shown in Table 7.6. Figure 7.35 (left) shows the effect of modifying the top mass window on the transfer factors. The nominal scenario, where both top quark candidates have a mass between 120 and 230 GeV is shown with black, while the alternative scenarios where the mass of both top quark candidates fall in the mass windows $130 < \mathrm{m_t} < 220\,\mathrm{GeV}$ and $100 < \mathrm{m_t} < 240\,\mathrm{GeV}$, are shown with red and yellow, respectively. Figure 7.35 (right) shows the effect of changing the definition of the control regions used to extract the transfer factors. The modification here is defined by requiring the DNN score of the associated top quark candidate to be less than the *VLoose* working point instead of the *Loose* working point, which translates into *shrinking* the two control regions simultaneously. Figure 7.36 (left) shows how the change of the top quark candidate being inverted in the ABDC method affects the derivation of the transfer factors. In the nominal method, shown with black, the DNN score of the subleading in p_T top quark candidate between *top 1* and *top 2* is being inverted. The transfer factors shown with purple correspond to the ones derived when inverting the DNN of the subleading in DNN score (*top 2*), which is the one used in the ABCD method in the 2016 analysis. The transfer factors shown with light blue correspond to the case where the DNN score of either the subleading in p_T or the subleading in DNN top quark candidates are inverted. As a final test, the composition of events is modified by subtracting 0.5, 1.5, and 2.0 times the number of Genuine b jets events, or no subtraction at all. The effect on the derivation of the transfer factors is shown in Fig. 7.36 (right). For each of the aforementioned sources of systematic uncertainties, the maximum variation with respect to the nominal selection is taken as the final

Table 7.6: Mapping of the transfer factors bin label with the corresponding $H^\pm$ b jet p_T and $|\eta|$ bin.

Bin	$H^\pm$ b-jet p_T (GeV)	$H^\pm$ b-jet η	Bin	$H^\pm$ b jet p_T (GeV)	$H^\pm$ b-jet η
0	$p_T < 50$		21	$p_T < 50$	
1	$50 < p_T < 75$		22	$50 < p_T < 75$	
2	$75 < p_T < 110$		23	$75 < p_T < 110$	
3	$110 < p_T < 150$	$|\eta| < 0.7$	24	$110 < p_T < 150$	$1.5 < |\eta| < 2.0$
4	$150 < p_T < 200$		25	$150 < p_T < 200$	
5	$200 < p_T < 280$		26	$200 < p_T < 280$	
6	$p_T > 280$		27	$p_T > 280$	
7	$p_T < 50$		28	$p_T < 50$	
8	$50 < p_T < 75$		29	$50 < p_T < 75$	
9	$75 < p_T < 110$		30	$75 < p_T < 110$	
10	$110 < p_T < 150$	$0.7 < |\eta| < 1.1$	31	$110 < p_T < 150$	$|\eta| > 2.0$
11	$150 < p_T < 200$		32	$150 < p_T < 200$	
12	$200 < p_T < 280$		33	$200 < p_T < 280$	
13	$p_T > 280$		34	$p_T > 280$	
14	$p_T < 50$				
15	$50 < p_T < 75$				
16	$75 < p_T < 110$	$1.1 < |\eta| < 1.5$			
17	$110 < p_T < 150$				
18	$150 < p_T < 200$				
19	$200 < p_T < 280$				
20	$p_T > 280$				
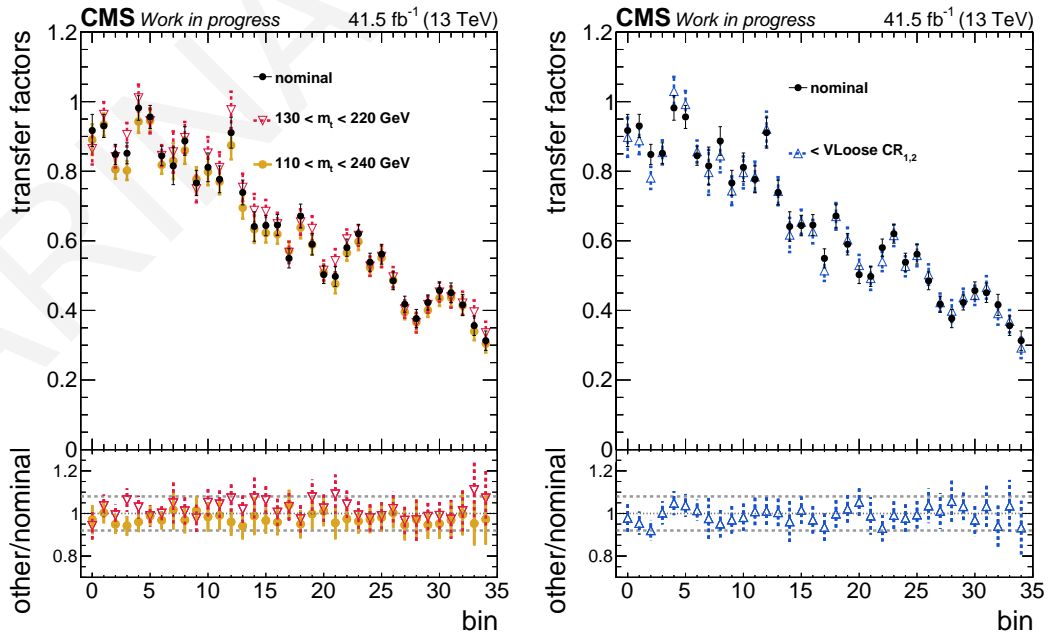
Figure 7.35: The values of the transfer factors at each bin for the nominal scenario and for the scenarios of changing the top mass window (*left*) or the definition of the CR1 and CR2 (*right*).
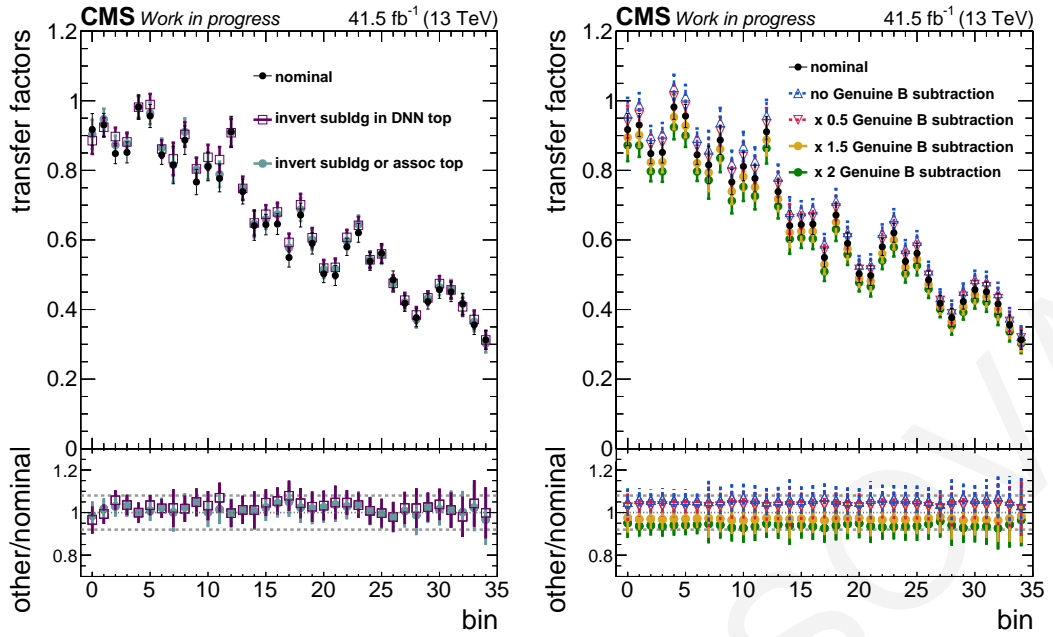
Figure 7.36: The values of the transfer factors at each bin for the nominal scenario and for the scenarios of inverting the DNN score of a different top quark candidate (*left*) or the background composition (*right*).

uncertainty at that bin. The four sources are treated as independent of each other and affect both the shape and the normalization of the misidentified b jets prediction.

Another source of systematic uncertainty affecting the misidentified b jets measurement can arise from the binning selection of the transfer factors. In the default measurement of the misidentified b jets background, the phase-space is split into 35 bins of the $H^{\pm}$ b-jet p_T and $|\eta|$. Figure 7.37 (left) shows the closure of the final fit discriminant for the true mass hypothesis of $m_{H^{\pm}} = 800$ GeV, in the Closure Region for the default binning. Figure 7.37 (right) shows the data-to-background ratio of the left plot (nominal) together with the closure in the same region using six different binning selections. The ratio reveals a maximum of 2% difference with respect to the nominal binning selection, and thus a conservative 2% uncertainty is assigned as a rate-only uncertainty on the misidentified b jets background measurement.

7.6.2 Statistical uncertainties

The uncertainties due to the limited statistics of the signal and background samples in the template distribution are individually incorporated for each template bin with the *Barlow-Beeston lite* approach. These uncertainties are treated as uncorrelated between bins and categories.
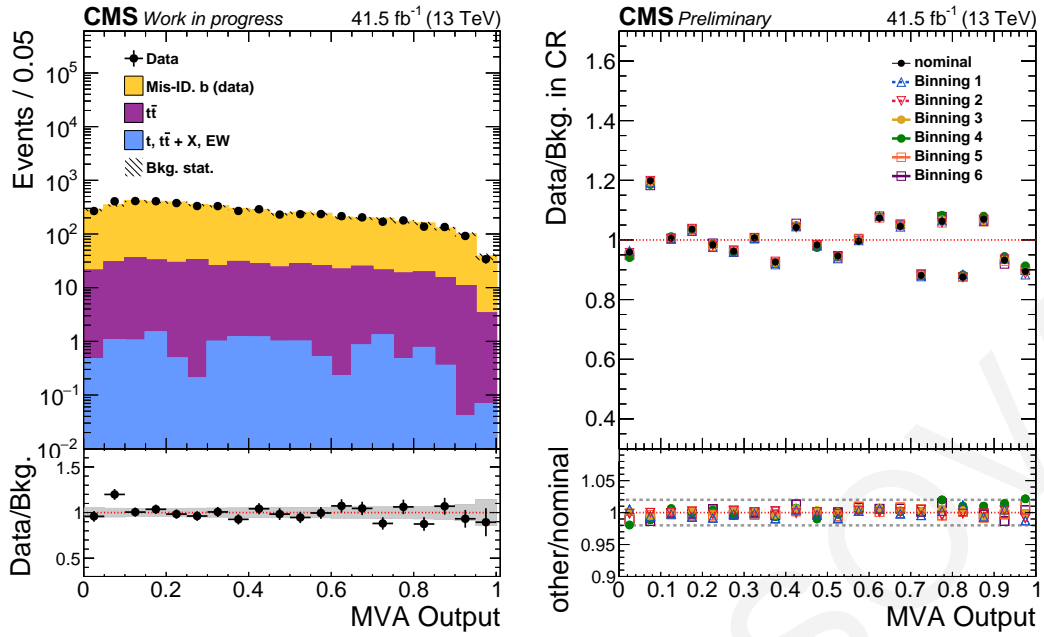
Figure 7.37: The output of the parameterized DNN for the true mass hypothesis of $m_{H^{\pm}}$ = 800 GeV, in the closure region (*left*). The data-to-background ratio of the output of the parameterized DNN for the nominal and for six binning variations (*right*).

7.6.3 Theoretical uncertainties

The theoretical systematic uncertainties affecting the overall cross-section of the processes, are due to missing higher-order QCD corrections and uncertainties in the PDFs, as well as uncertainties concerning the top quark mass and the coupling constant α_S. The effect of the QCD scale uncertainties is estimated by varying the factorization (μ_F) and renormalization (μ_R) scales up and down by a factor of half and two, with respect to their nominal value, and enveloping the maximum variation on the final fit discriminant. The PDF uncertainties are treated as fully correlated for simulated samples sharing the same dominant partons in the initial state of the matrix element calculations.

A summary of all systematic uncertainties discussed in this section is given in Table 7.7 for charged Higgs boson mass of 250 GeV, and in Table 7.8 for $m_{H^{\pm}}$ = 800 GeV, in the $1M1L_t$ category. The corresponding summaries for the $2M_t$ category are given in Tables 7.9 and 7.10.

7.7 Results

The pre-fit blinded distributions of the final fit discriminant, the MVA output of the parameterized DNN, are presented in Fig. 7.38 for the $1M1L_t$ signal region and in Fig. 7.39 for the $2M_t$ signal region. The observed data is only unblinded in the low-MVA score region ($<$ 0.3), where the signal is expected to be negligible. In this region, the data and simulation are in good agreement, within the statistical and systematic uncertainties. The distributions corresponding to higher values of the true charged Higgs boson mass hypothe-

	Shape	Signal	Misidentified b	Genuine b	
	✓			$t\bar{t}$	Single t + $t\bar{t}$+X + EW
Integrated luminosity	−	2.0	−	2.0	2.0
Trigger efficiency	✓	4.8	−	7.7	6.7
Pileup	✓	0.3	−	0.4	2.2
electron veto efficiency	−	< 0.1	−	< 0.1	< 0.1
muon veto efficiency	−	< 0.1	−	< 0.1	< 0.1
tau veto efficiency	−	0.3	−	0.3	0.2
Jet energy scale	✓	12	−	9.3	11
Jet energy resolution	✓	6.6	−	1.7	5.2
b jet identification	✓	6.4	−	5.6	6.4
b jet mistagging	✓	3.6	−	2.3	2.1
Resolved top tagging efficiency	✓	6.1	−	7.3	6.4
Resolved top mistagging efficiency	✓	2.6	−	2.2	2.8
$j \to b$ syst. (top-mass window)	✓	−	4.4	−	−
$j \to b$ syst. (inverted top)	✓	−	3.2	−	−
$j \to b$ syst. (fake-b composition)	✓	−	6.1	−	−
$j \to b$ syst. (CR definition)	✓	−	3.3	−	−
$j \to b$ binning selection	−	−	2.0	−	−
$j \to b$ misidentification statistical	✓	−	0.1 - 0.3	−	−
Top quark mass ($t\bar{t}$)	−	−	−	2.8	−
Top quark mass (single t)	−	−	−	−	2.2
QCD Scales ($t\bar{t}$)	−	−	−	$^{+2.4}_{-3.5}$	−
QCD Scales (single t)	−	−	−	−	2.5
QCD Scales (EW)	−	−	−	−	3.2
QCD acceptance (top)	−	−	−	3.2	3.2
QCD acceptance ($H^\pm$)	−	$^{+6.7}_{-4.4}$	−	−	−
PDF+α_s ($t\bar{t}$)	−	−	−	4.2	−
PDF+α_s (EW)	−	−	−	−	4.4
PDF+α_s (single t)	−	−	−	−	4.7
PDF+α_s acceptance (Top)	−	−	−	0.1	0.1
PDF+α_s acceptance ($H^\pm$)	−	0.1	−	−	−
Theoretical α_s (ttH)	−	−	−	−	1.0

Table 7.7: Systematics summary table for $1M1L_t$ and $m_{H^\pm}$ = 250 GeV.

	Shape	Signal	Misidentified b	Genuine b	
	$\checkmark$			$t\bar{t}$	Single t + $t\bar{t}$+X + EW
Integrated luminosity	$-$	2.0	$-$	2.0	2.0
Trigger efficiency	$\checkmark$	5.8	$-$	7.7	7.4
Pileup	$\checkmark$	0.9	$-$	0.4	3.0
electron veto efficiency	$-$	< 0.1	$-$	< 0.1	< 0.1
muon veto efficiency	$-$	< 0.1	$-$	< 0.1	< 0.1
tau veto efficiency	$-$	0.2	$-$	0.3	0.2
Jet energy scale	$\checkmark$	8.1	$-$	9.3	9.9
Jet energy resolution	$\checkmark$	4.1	$-$	1.7	6.0
b jet identification	$\checkmark$	8.5	$-$	5.6	6.8
b jet mistagging	$\checkmark$	4.1	$-$	2.3	2.8
Resolved top tagging efficiency	$\checkmark$	7.4	$-$	7.3	5.5
Resolved top mistagging efficiency	$\checkmark$	2.0	$-$	2.2	3.5
$j \to b$ syst. (top-mass window)	$\checkmark$	$-$	4.4	$-$	$-$
$j \to b$ syst. (inverted top)	$\checkmark$	$-$	3.2	$-$	$-$
$j \to b$ syst. (fake-b composition)	$\checkmark$	$-$	6.1	$-$	$-$
$j \to b$ syst. (CR definition)	$\checkmark$	$-$	3.3	$-$	$-$
$j \to b$ binning selection	$-$	$-$	2.0	$-$	$-$
$j \to b$ misidentification statistical	$\checkmark$	$-$	0.1 - 0.3	$-$	$-$
Top quark mass ($t\bar{t}$)	$-$	$-$	$-$	2.8	$-$
Top quark mass (single t)	$-$	$-$	$-$	$-$	2.2
QCD Scales ($t\bar{t}$)	$-$	$-$	$-$	$^{+2.4}_{-3.5}$	$-$
QCD Scales (single t)	$-$	$-$	$-$	$-$	2.5
QCD Scales (EW)	$-$	$-$	$-$	$-$	3.2
QCD acceptance (top)	$-$	$-$	$-$	3.2	3.2
QCD acceptance ($\mathrm{H}^{\pm}$)	$-$	$^{+6.7}_{-4.4}$	$-$	$-$	$-$
PDF+α_s ($t\bar{t}$)	$-$	$-$	$-$	4.2	$-$
PDF+α_s (EW)	$-$	$-$	$-$	$-$	4.4
PDF+α_s (single t)	$-$	$-$	$-$	$-$	4.7
PDF+α_s acceptance (Top)	$-$	$-$	$-$	0.1	0.1
PDF+α_s acceptance ($\mathrm{H}^{\pm}$)	$-$	0.1	$-$	$-$	$-$
Theoretical α_s (ttH)	$-$	$-$	$-$	$-$	1.0

Table 7.8: Systematics summary table for $1\mathrm{M}1\mathrm{L}_\mathrm{t}$ and $\mathrm{m}_{\mathrm{H}^{\pm}}$ = 800 GeV.

	Shape	Signal	Misidentified b	Genuine b $t\bar{t}$	Genuine b Single t + $t\bar{t}$+X + EW
Integrated luminosity	−	2.0	−	2.0	2.0
Trigger efficiency	✓	5.0	−	8.6	7.3
Pileup	✓	0.4	−	< 0.1	0.8
electron veto efficiency	−	< 0.1	−	< 0.1	< 0.1
muon veto efficiency	−	< 0.1	−	< 0.1	< 0.1
tau veto efficiency	−	0.3	−	0.3	0.2
Jet energy scale	✓	9.9	−	7.9	9.5
Jet energy resolution	✓	7.4	−	7.9	2.6
b jet identification	✓	6.0	−	5.5	6.2
b jet mistagging	✓	3.4	−	2.4	2.3
Resolved top tagging efficiency	✓	7.0	−	8.5	6.1
Resolved top mistagging efficiency	✓	5.2	−	4.7	5.5
$j \to b$ syst. (top-mass window)	✓	−	4.5	−	−
$j \to b$ syst. (inverted top)	✓	−	3.2	−	−
$j \to b$ syst. (fake-b composition)	✓	−	6.1	−	−
$j \to b$ syst. (CR definition)	✓	−	3.4	−	−
$j \to b$ binning selection	−	−	2.0	−	−
$j \to b$ misidentification statistical	✓	−	0.1 - 0.4	−	−
Top quark mass ($t\bar{t}$)	−	−	−	2.8	−
Top quark mass (single t)	−	−	−	−	2.2
QCD Scales ($t\bar{t}$)	−	−	−	$^{+2.4}_{-3.5}$	−
QCD Scales (single t)	−	−	−	−	2.5
QCD Scales (EW)	−	−	−	−	3.2
QCD acceptance (top)	−	−	−	2.1	2.1
QCD acceptance (H$^\pm$)	−	$^{+4.9}_{-3.5}$	−	−	−
PDF+α_s ($t\bar{t}$)	−	−	−	4.2	−
PDF+α_s (EW)	−	−	−	−	4.4
PDF+α_s (single t)	−	−	−	−	4.7
PDF+α_s acceptance (Top)	−	−	−	0.1	0.1
PDF+α_s acceptance (H$^\pm$)	−	0.1	−	−	−
Theoretical α_s (ttH)	−	−	−	−	1.0

Table 7.9: Systematics summary table for $2M_t$ and $m_{H^\pm}$ = 250 GeV.

	Shape	Signal	Misidentified b	Genuine b	
	✓			$t\bar{t}$	Single t + $t\bar{t}$+X + EW
Integrated luminosity	−	2.0	−	2.0	2.0
Trigger efficiency	✓	6.8	−	8.6	7.3
Pileup	✓	0.3	−	< 0.1	0.8
electron veto efficiency	−	< 0.1	−	< 0.1	< 0.1
muon veto efficiency	−	< 0.1	−	< 0.1	< 0.1
tau veto efficiency	−	0.2	−	0.3	0.2
Jet energy scale	✓	5.4	−	7.9	9.5
Jet energy resolution	✓	6.0	−	7.9	2.6
b jet identification	✓	8.4	−	5.5	6.2
b jet mistagging	✓	4.0	−	2.4	2.3
Resolved top tagging efficiency	✓	7.3	−	8.5	6.1
Resolved top mistagging efficiency	✓	4.4	−	4.7	5.5
$j \rightarrow b$ syst. (top-mass window)	✓	−	4.5	−	−
$j \rightarrow b$ syst. (inverted top)	✓	−	3.2	−	−
$j \rightarrow b$ syst. (fake-b composition)	✓	−	6.1	−	−
$j \rightarrow b$ syst. (CR definition)	✓	−	3.4	−	−
$j \rightarrow b$ binning selection	−	−	2.0	−	−
$j \rightarrow b$ misidentification statistical	✓	−	0.1 - 0.4	−	−
Top quark mass ($t\bar{t}$)	−	−	−	2.8	−
Top quark mass (single t)	−	−	−	−	2.2
QCD Scales ($t\bar{t}$)	−	−	−	$^{+2.4}_{-3.5}$	−
QCD Scales (single t)	−	−	−	−	2.5
QCD Scales (EW)	−	−	−	−	3.2
QCD acceptance (top)	−	−	−	2.1	2.1
QCD acceptance ($H^{\pm}$)	−	$^{+4.9}_{-3.5}$	−	−	−
PDF+α_s ($t\bar{t}$)	−	−	−	4.2	−
PDF+α_s (EW)	−	−	−	−	4.4
PDF+α_s (single t)	−	−	−	−	4.7
PDF+α_s acceptance (Top)	−	−	−	0.1	0.1
PDF+α_s acceptance ($H^{\pm}$)	−	0.1	−	−	−
Theoretical α_s (ttH)	−	−	−	−	1.0

Table 7.10: Systematics summary table for $2M_t$ and $m_{H^{\pm}}$ = 800 GeV.

(a) (b) (c)

(d) (e) (f)
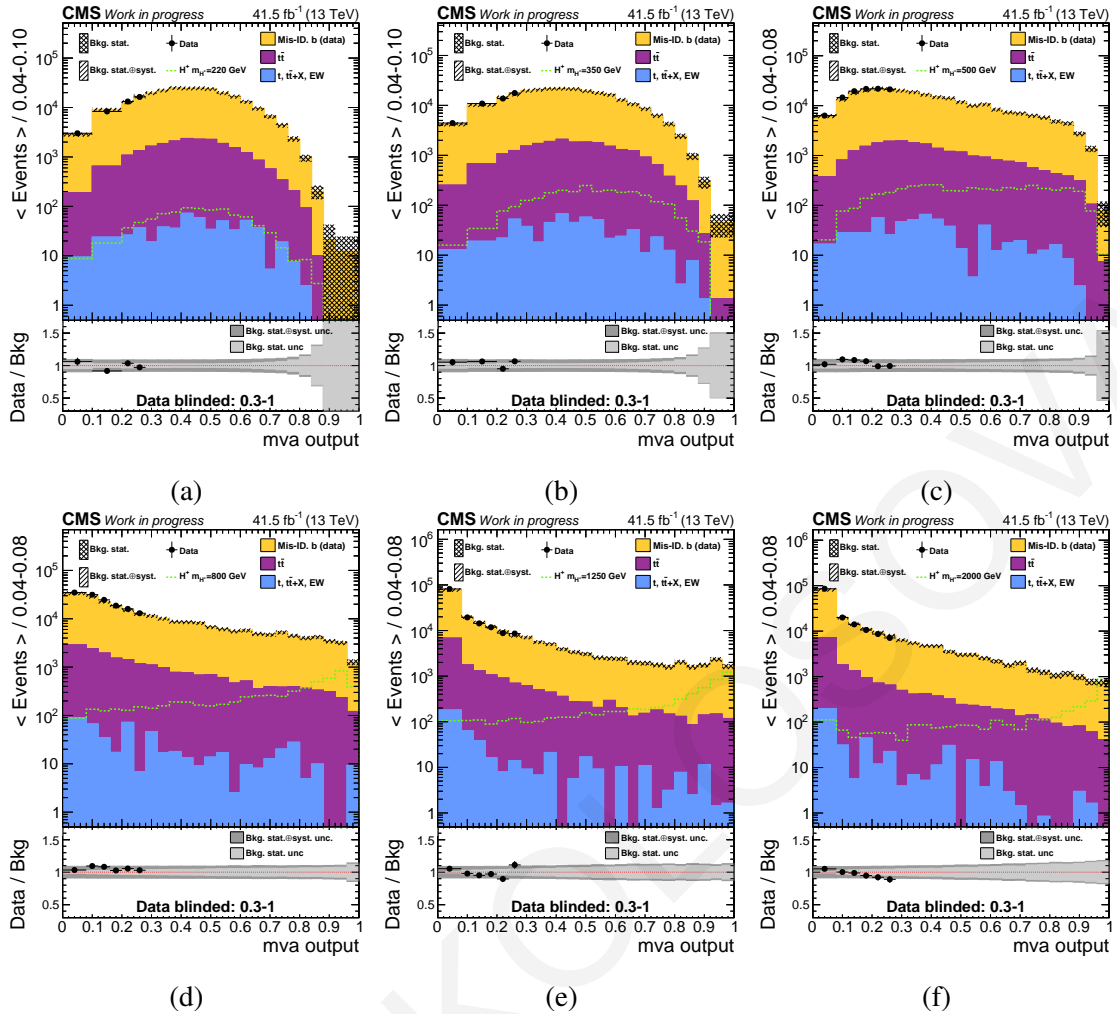
Figure 7.38: The pre-fit distributions of the MVA output of the parameterized deep neural network for the $1M1L_t$ signal region. The observed data, shown with black, is only unblinded in the low-MVA score regions where the signal is expected to be negligible. The uncertainty displayed with light gray corresponds to the statistical only, while the darker shade additionally includes the systematic uncertainty. From left to right, the upper row distributions correspond to the true $H^\pm$ mass hypotheses of 220, 350, and 500 GeV, while the lower row distributions correspond to 800, 1250, and 2000 GeV.

sis reveal greater discrimination between signal and background, while no discrimination is observed for the very low masses of the charged Higgs boson.

7.8 Expected limits

The expected exclusion limits on the $H^\pm$ production cross-section $\sigma_{pp\to t(b)H^\pm}$ times the branching fraction $\mathcal{B}(H^\pm \to tb)$ at 95% CL are reported in Fig. 7.40 for the two signal regions separately and in Fig. 7.41 (left) for the combination. The right plot of Fig. 7.41 shows the relative contribution of each signal region with respect to the combined expected upper limit. The numerical values of the combined expected exclusion limits are listed in Table 7.11. A comparison between the combined expected upper limits when fitting on the MVA output of the parameterized DNN with the one obtained when fitting on the recon-
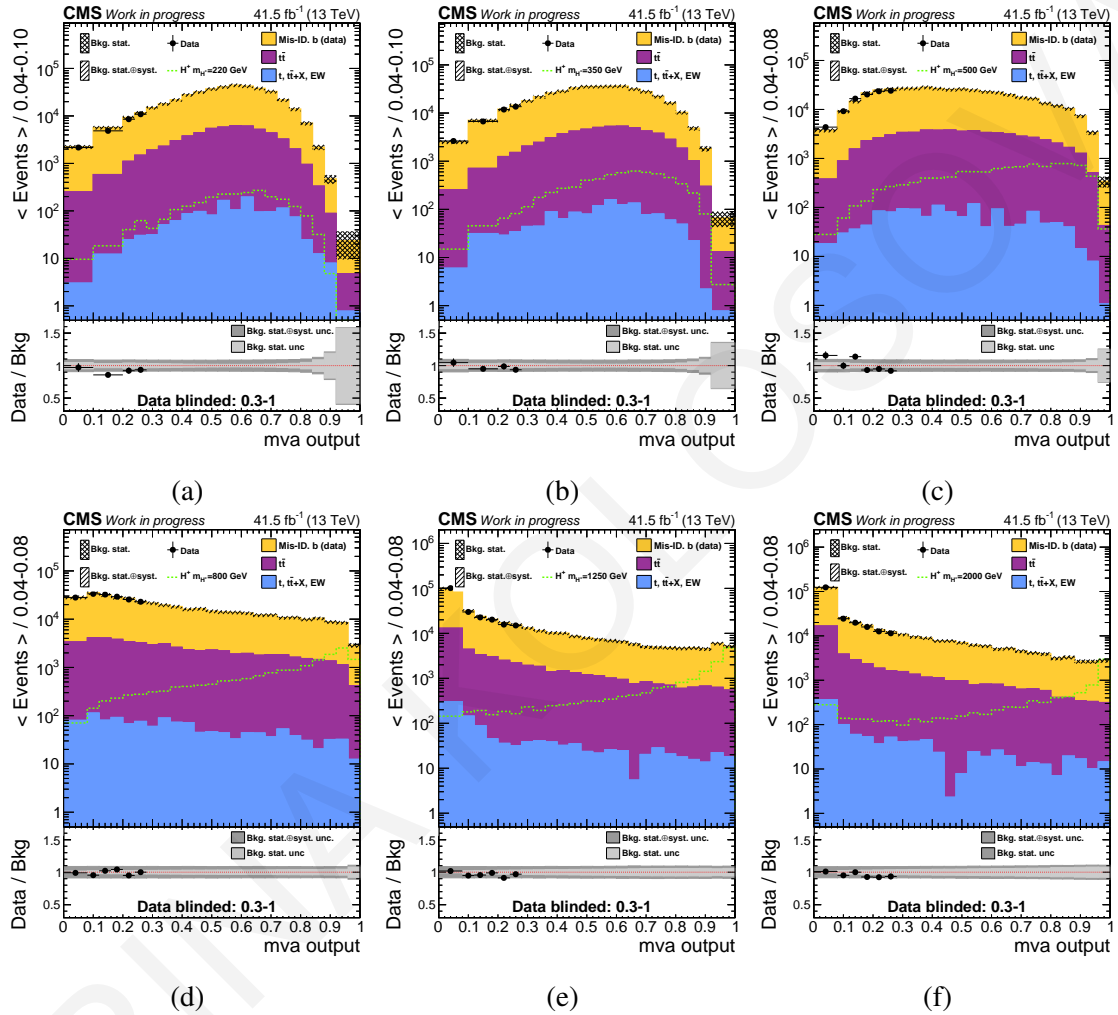
Figure 7.39: The pre-fit distributions of the MVA output of the parameterized deep neural network for the $2M_t$ signal region. The observed data, shown with black, is only unblinded in the low-MVA score regions where the signal is expected to be negligible. The uncertainty displayed with light Gray corresponds to the statistical only, while the darker shade additionally includes the systematic uncertainty. From left to right, the upper row distributions correspond to the true $H^{\pm}$ mass hypotheses of 220, 350, and 500 GeV, while the lower row distributions correspond to 800, 1250, and 2000 GeV.
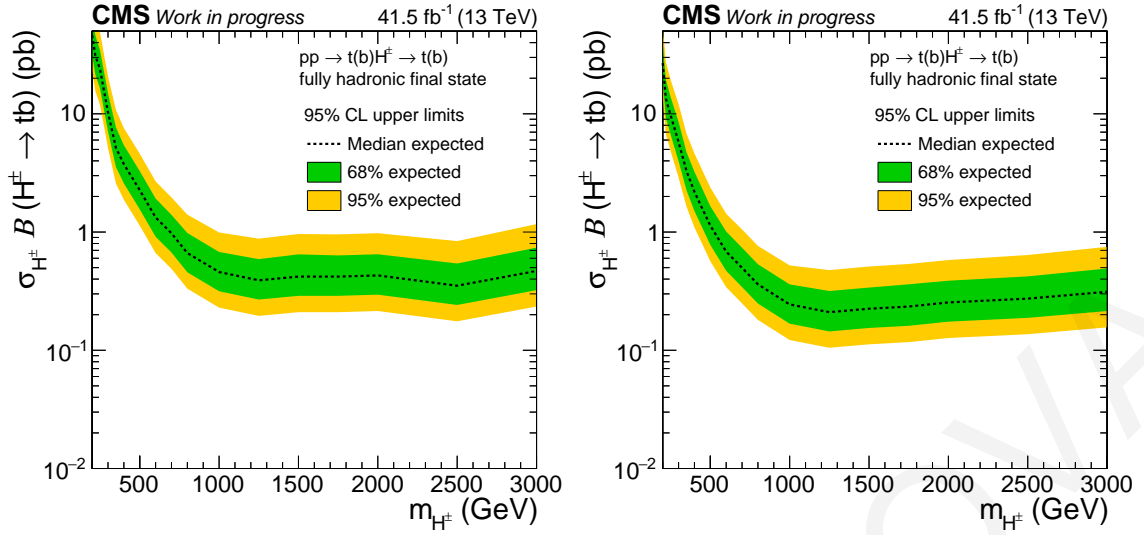
Figure 7.40: The 95% CL upper limits on $\sigma_{\text{pp}\to\text{t(b)}H^\pm} \times \mathcal{B}(H^\pm \to \text{tb})$ obtained from the $1\text{M}1\text{L}_\text{t}$ (*left*) and 2M_t (*right*) signal categories, for the $H^\pm$ mass range from 200 GeV to 3 TeV.

structed invariant mass of the charged Higgs boson candidate is shown in Fig. 7.42.

Validation of the statistical model

The blinded impacts and pulls for the signal plus background hypothesis and for the first 30 most important nuisances are shown in Fig. 7.43 for the $1\text{M}1\text{L}_\text{t}$ category and in Fig. 7.44 for the 2M_t category, for $m_{H^\pm} = 800$ GeV. For both signal regions, no nuisance appears to be significantly pulled away from its nominal value, neither is over-constrained.

A goodness-of-fit test is performed for each of the true mass hypotheses, using the saturated method and a sample of 5000 toy events. The results are shown in Figs. 7.45 and 7.46 for the $1\text{M}1\text{L}_\text{t}$ and 2M_t signal regions, respectively. For all the mass points except for the $m_{H^\pm} = 1500$ GeV, the observed value of the χ^2 goodness-of-fit lies inside the toy distribution, concluding that the statistical model used describes well the observation. A dedicated analysis is needed for the high mass region to target the boosted event topologies.
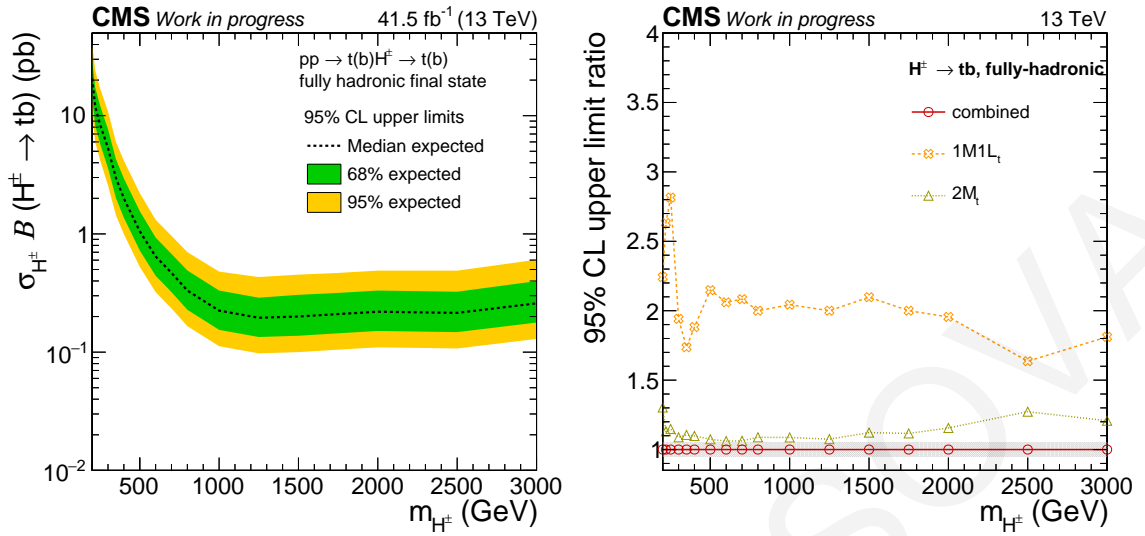
Figure 7.41: Left: The 95% CL upper limit on $\sigma_{\mathrm{pp}\to\mathrm{t(b)}H^{\pm}} \times \mathcal{B}(H^{\pm} \to \mathrm{tb})$ obtained from the combination of $1\mathrm{M}1\mathrm{L_t}$ and $2\mathrm{M_t}$ signal regions, for the $H^{\pm}$ mass range from 200 GeV to 3 TeV. Right: The relative median expected upper limit showing the contribution of each signal region with respect to the combined upper limit.

95% CL upper limit on $\sigma_{\mathrm{pp}\to\mathrm{t(b)}H^{\pm}} \times \mathcal{B}(H^{\pm} \to \mathrm{tb})$						
$\mathrm{m}_{H^{\pm}}$			Expected limit			Observed
(GeV)	-2σ	-1σ	median	$+1\sigma$	$+2\sigma$	limit
200	10.31250	14.17969	20.62500	29.50391	41.18060	Blinded
220	5.93750	8.16406	11.87500	17.17644	23.98074	Blinded
250	4.21875	5.80078	8.43750	12.20431	17.03894	Blinded
300	2.57324	3.60046	5.31250	7.68420	10.86057	Blinded
350	1.43799	2.01202	2.96875	4.29411	5.99518	Blinded
400	0.99609	1.36963	1.99219	2.94510	4.16106	Blinded
500	0.52734	0.72510	1.05469	1.55917	2.21579	Blinded
600	0.32227	0.44312	0.64453	0.93227	1.31764	Blinded
700	0.23438	0.32227	0.46875	0.67802	0.96412	Blinded
800	0.16602	0.22827	0.33203	0.49085	0.70162	Blinded
1000	0.11230	0.15442	0.22461	0.33205	0.48011	Blinded
1250	0.09766	0.13428	0.19531	0.28874	0.43179	Blinded
1500	0.10010	0.13763	0.20020	0.30553	0.45241	Blinded
1750	0.10498	0.14435	0.20996	0.31709	0.46773	Blinded
2000	0.10986	0.15106	0.21973	0.33183	0.48949	Blinded
2500	0.10742	0.14771	0.21484	0.32446	0.48888	Blinded
3000	0.12939	0.17792	0.25879	0.39908	0.60513	Blinded

Table 7.11: The expected 95% CL exclusion limits on $\sigma_{\mathrm{pp}\to\mathrm{t(b)}H^{\pm}} \times \mathcal{B}(H^{\pm} \to \mathrm{tb})$ for the combination of the $1\mathrm{M}1\mathrm{L_t}$ and $2\mathrm{M_t}$ signal regions.
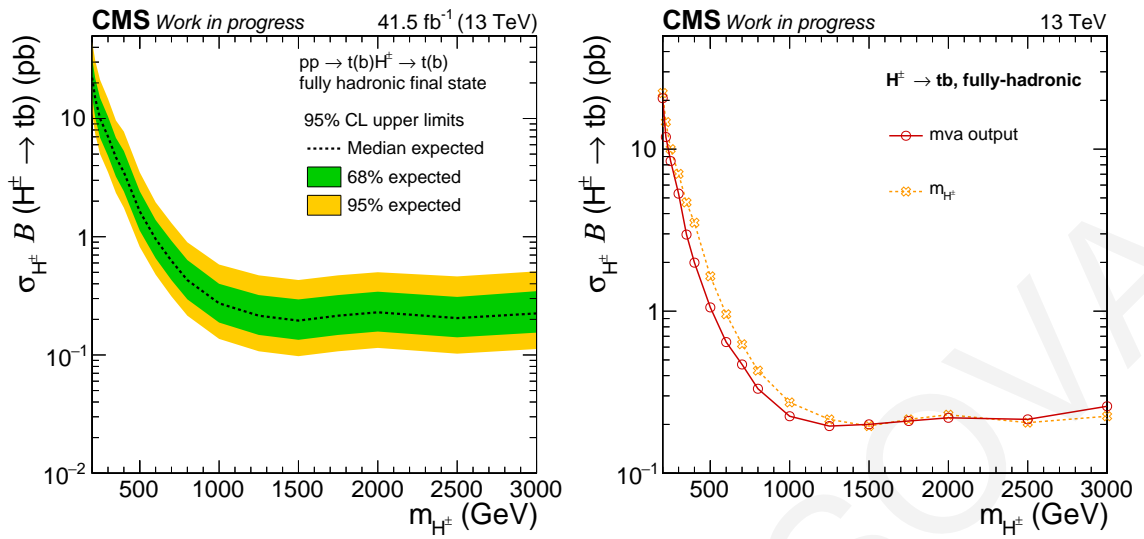
Figure 7.42: Left: The 95% CL upper limit on $\sigma_{\mathrm{pp}\to\mathrm{t(b)H^{\pm}}} \times \mathcal{B}(H^{\pm} \to \mathrm{tb})$ obtained from the combination of $1M1L_t$ and $2M_t$ signal regions, for the $H^{\pm}$ mass range from 200 GeV to 3 TeV, when using as the final fit discriminant the invariant mass of the $H^{\pm}$ candidate. Right: The comparison between the median expected upper limits when fitting on the MVA output of the parameterized DNN (red) and the reconstructed invariant mass of $H^{\pm}$ candidate (orange).
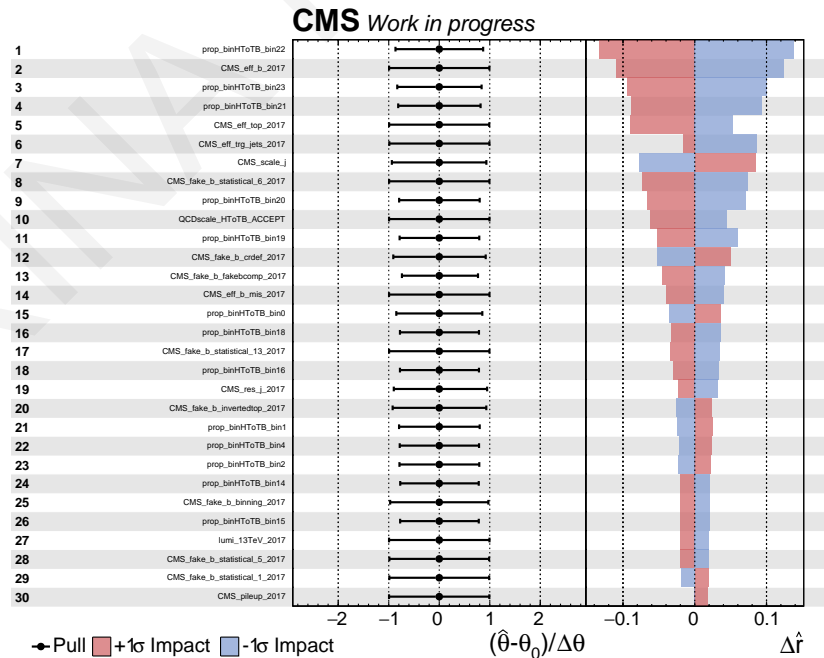
Figure 7.43: The pulls (first column) and $+1\sigma$ and -1σ impacts (second column) of the first 30 most important nuisance parameters for the true $H^{\pm}$ mass hypothesis of 800 GeV, in the $1M1L_t$ signal region.
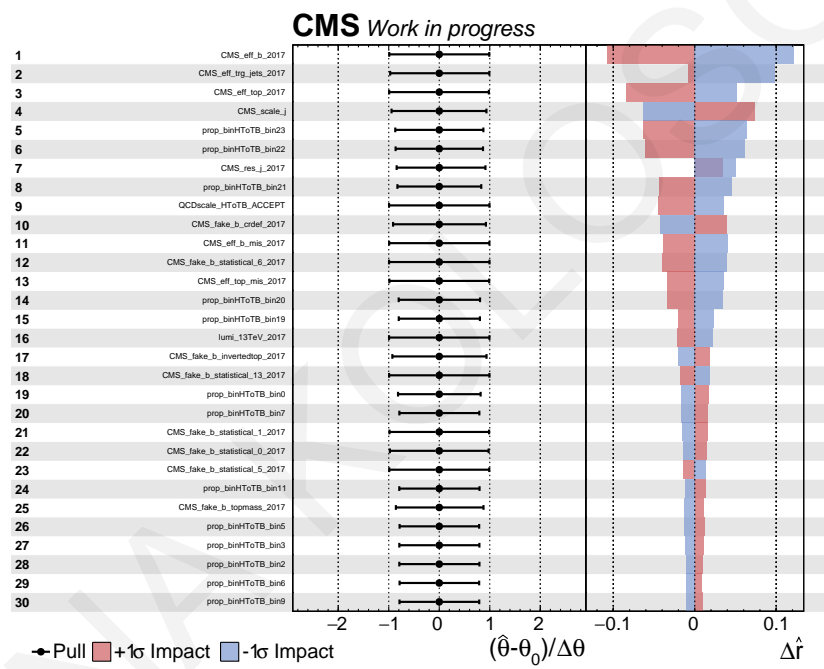
Figure 7.44: The pulls (first column) and $+1\sigma$ and -1σ impacts (second column) of the first 30 most important nuisance parameters for the true $\mathrm{H}^{\pm}$ mass hypothesis of 800 GeV, in the $2\mathrm{M_t}$ signal region.

(a)

(b)

(c)

(d)
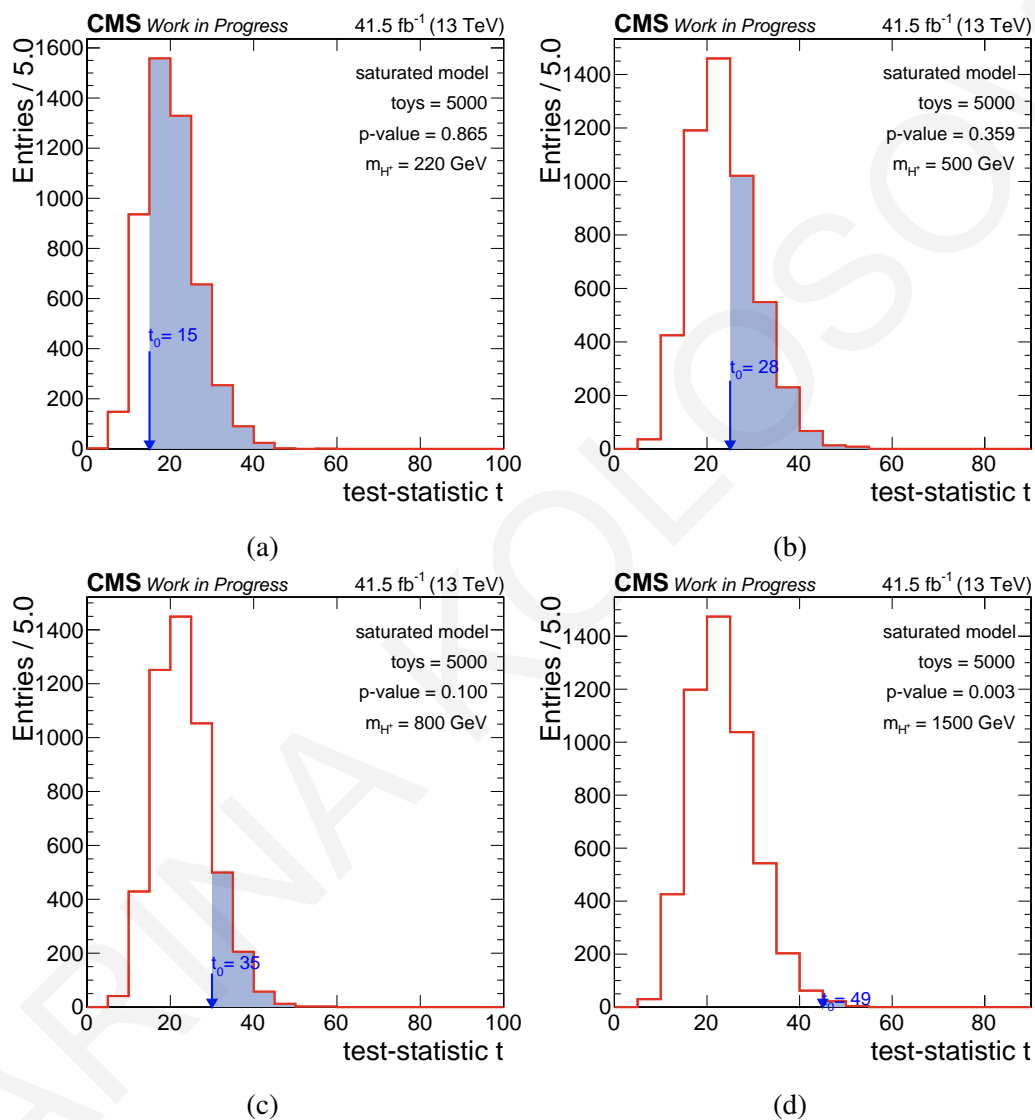
Figure 7.45: Goodness-of-fit tests for $m_{H^{\pm}}$ = 220, 500 GeV (upper row) and $m_{H^{\pm}}$ = 800, 1500 GeV (lower row) for the $1M1L_t$ signal region.
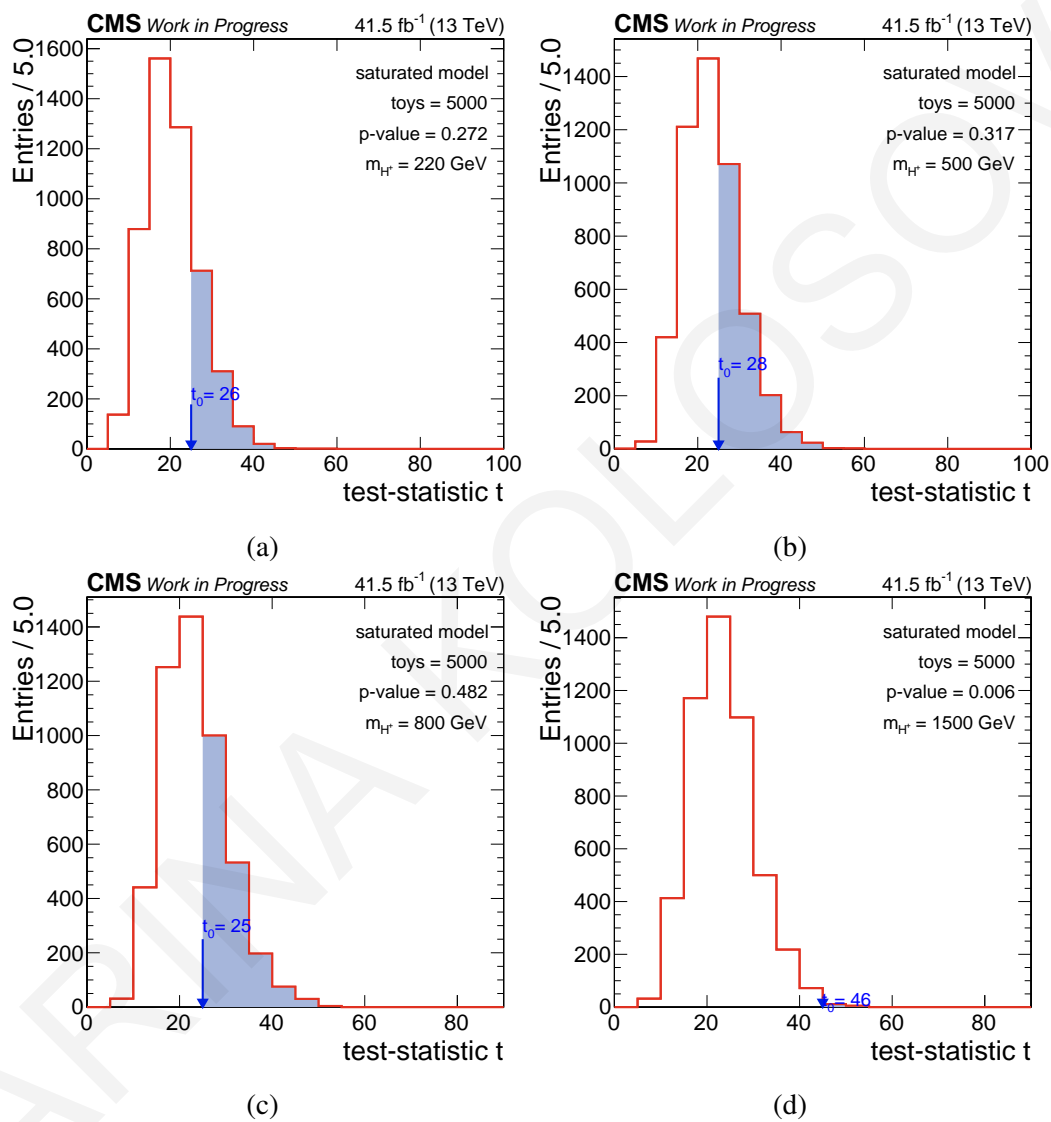
Figure 7.46: Goodness-of-fit tests for $\mathrm{m}_{\mathrm{H}^{\pm}} = 220$, 500 GeV (upper row) and $\mathrm{m}_{\mathrm{H}^{\pm}} = 800$, 1500 GeV (lower row) for the $2\mathrm{M}_{\mathrm{t}}$ signal region.

8   Summary and conclusions

Charged Higgs bosons ($H^\pm$) are predicted in many models beyond the Standard Model. One of the simplest extensions containing an additional Higgs sector is the two-Higgs-doublet model (2HDM). This work focuses on the search for a heavy charged Higgs boson with a mass above the top and bottom quark mass threshold ($m_t - m_b$). Due to the large couplings between the heavy fermions and the $H^\pm$ in such models, the $H^\pm$ is searched at the LHC in the dominant production channel $pp \to t(b)H^\pm$ and its subsequent decay to a top and bottom quark-antiquark pair, $H^\pm \to tb$. The measurement is performed in the all-jet final state for the first time with the 2016 collision data of the LHC, recorded by the CMS detector at $\sqrt{s} = 13$ TeV.

Depending on the $H^\pm$ mass, there exist two distinct event topologies, the boosted and the resolved one. When the $H^\pm$ has a very large mass of $\mathcal{O}(\text{TeV})$, the $H^\pm$ decay products have large on average transverse momenta, and their subsequent decays lead to highly collimated jets. The products of the top quark or W boson decay cannot be resolved with the standard algorithms, instead, they are reconstructed as a single large-radius jet (boosted topology). For lower $H^\pm$ masses, all its decay products can be resolved as small-cone jets (resolved topology). In the resolved topology, which is the main focus of this work, events are characterized by large jet and b jet multiplicities. Since there are no neutrinos in the final state, the reconstruction of the $H^\pm$ invariant mass is feasible. The large QCD-multijet and irreducible $t\bar{t}$ background, as well as the large combinatoric self-background, make this final state extremely challenging. However, the aforementioned background can be reduced to a large extent by employing b quark and top quark tagging techniques. Events are required to have at least seven jets, out of which three are identified as b jets, and two top quark candidates are identified by a custom boosted decision tree. The invariant mass of the $H^\pm$ is reconstructed by the leading-in-p_T top quark candidate and the leading-in-p_T b jet not used in the reconstruction of the two top quark candidates. To assess the presence or absence of a charged Higgs boson signal, a binned maximum-likelihood fit is performed on the invariant mass of the $H^\pm$. No significant deviation above the expected event yield of SM processes is observed in the 2016 data, and model-independent exclusion limits are set on the $\sigma_{pp \to t(b)H^\pm} \times \mathcal{B}(H^\pm \to tb)$, at 95% CL. The upper limits range from 21.3 to 0.007 pb in the $m_{H^\pm}$ range of 200 GeV to 3 TeV. The results are interpreted in the minimal supersymmetric standard model hMSSM and $M_h^{125}(\tilde{\chi})$ benchmark scenarios.

A continuation of the search is performed with the 2017 data and features several anal-

ysis improvements, including new machine-learning-based techniques for top quark identi-
fication and signal extraction, event categorization, and validation of the background mea-
surement method. A comparison of the expected upper limits as derived from the 2016 and
2017 analyses in the $2M_t$ category is shown in Figure 8.1. The 2016 analysis shows better
sensitivity at low masses and up to 300 GeV, while the 2017 analysis largely improves the
sensitivity in all mass points above 300 GeV. The improvement in sensitivity arises from
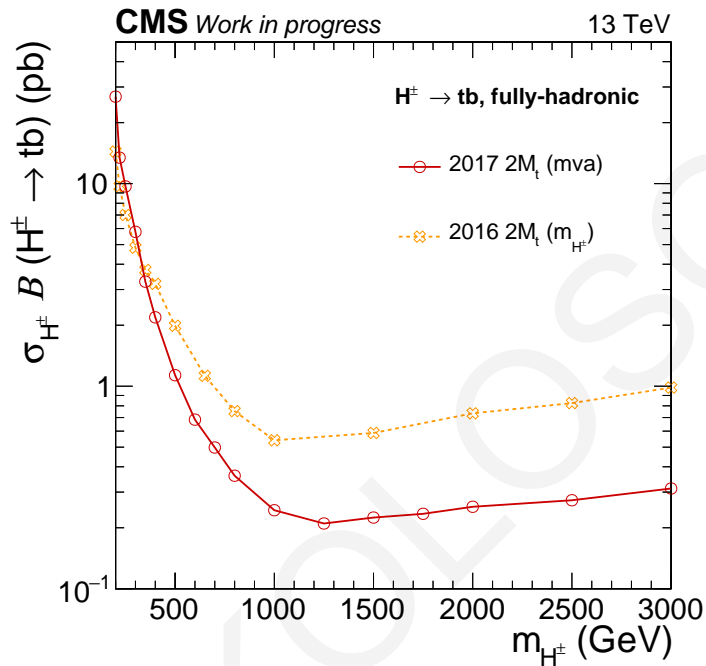
Figure 8.1: The 95% CL median expected upper limit on $\sigma_{pp\to t(b)H^\pm} \times \mathcal{B}(H^\pm \to tb)$ as
obtained from the 2016 (orange) and the 2017 analysis (red) for the $2M_t$ signal regions, for
the $H^\pm$ mass range from 200 GeV to 3 TeV.

several sources. One of the most important is the upgrade of the CMS Phase-1 pixel detector
before the 2017 data-taking era. The inclusion of new layers located closer to the interaction
point allowed for more robust tracking, which lead to improved b tagging performance. The
HLT paths used in 2017 had lower b misidentification rate, resulting in a more pure sample
of recorded events. The higher L1 and HLT thresholds on H_T, however, is one of the reasons
for worse sensitivity at very low masses of the charged Higgs boson due to low acceptance.
Offline, the improvement in sensitivity originates from improved b tagging and top tagging
techniques, as well as the change in the signal extraction method. The new parameterized
deep neural network (DNN) includes not only the invariant mass of the reconstructed charged
Higgs boson that was used to extract the signal in the 2016 analysis, but also event-based and
other kinematic variables that enhance the signal-to-background discrimination.

Further improvement in sensitivity is expected with the full 2016-2018 dataset, corre-
sponding to an integrated luminosity of 137 fb^{-1}, as well as from improved analysis tech-
niques. One way to improve the analysis techniques would be to train the event-based and
mass parameterized DNN with background events where the true mass hypothesis is assigned
to be closer to the reconstructed invariant mass of the $H^\pm$ candidate, instead of being assigned

randomly. By doing so, the algorithm would learn to discriminate better between signal and background at low $H^{\pm}$ masses. Another way would be to use two different DNNs, targeting the low and intermediate $m_{H^{\pm}}$ regions, and trained with a different set of discriminating variables, appropriate for each region. To achieve these, more statistics on the simulated signal samples are needed for the training, testing, and validation of the algorithms. For $H^{\pm}$ masses above 1 TeV, a dedicated analysis is needed to target the Lorentz-boosted topologies.

Searches for charged Higgs bosons are included in the upcoming physics program of the LHC and the CMS experiment. The LHC is currently in a shutdown, preparing for the upcoming Run 3 in 2022 which will last for a period of three years. By the end of Run 3, the LHC aims to produce and collect a physics dataset of around 300 fb^{-1}. The High-Luminosity LHC (HL-LHC), expected to commission on the second half of the 2030s, aims to significantly increase the instantaneous luminosity to $7.5 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ and generate an enormous amount of physics data of around 3 ab^{-1} at 14 TeV. The immense statistics will allow us to improve the precision of the measurements of the SM and the Higgs boson, as well as the sensitivity to new physics, including charged Higgs bosons. Considering the largely constrained phase-space of BSM models (i.e. type-II 2HDM), searches for charged Higgs bosons will not just be a bare rerun of previous analysis techniques. The upcoming large physics datasets will open new possibilities for new analysis developments and strategies that could better control the experimental systematics and the search sensitivities. Moreover, searches for charged Higgs bosons will expand in yet-unexplored channels with high discovery potential in given models, such as the pair production of charged Higgs bosons [184, 185], or the bosonic decays of the $H^{\pm}$, such as $H^{\pm} \to W^{\pm}A/H$, where A is the pseudoscalar Higgs boson and H is the light or heavy Higgs boson [186], and the $H^{\pm} \to W^{\pm} + 4\gamma$ [187].

Bibliography

[1] M. Gell-Mann, "A Schematic Model of Baryons and Mesons", *Phys. Lett.*, vol. 8, pp. 214–215, 1964. DOI: 10.1016/S0031-9163(64)92001-3.

[2] H. Fritzsch, M. Gell-Mann, and H. Leutwyler, "Advantages of the Color Octet Gluon Picture", *Phys. Lett. B*, vol. 47, pp. 365–368, 1973. DOI: 10.1016/0370-2693(73)90625-4.

[3] D. J. Gross and F. Wilczek, "Ultraviolet Behavior of Nonabelian Gauge Theories", *Phys. Rev. Lett.*, vol. 30, J. C. Taylor, Ed., pp. 1343–1346, 1973. DOI: 10.1103/PhysRevLett.30.1343.

[4] H. D. Politzer, "Reliable Perturbative Results for Strong Interactions?", *Phys. Rev. Lett.*, vol. 30, J. C. Taylor, Ed., pp. 1346–1349, 1973. DOI: 10.1103/PhysRevLett.30.1346.

[5] S. L. Glashow, "Partial Symmetries of Weak Interactions", *Nucl. Phys.*, vol. 22, pp. 579–588, 1961. DOI: 10.1016/0029-5582(61)90469-2.

[6] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.*, vol. 19, pp. 1264–1266, 1967. DOI: 10.1103/PhysRevLett.19.1264.

[7] A. Salam, "Weak and Electromagnetic Interactions", *Conf. Proc. C*, vol. 680519, pp. 367–377, 1968. DOI: 10.1142/9789812795915_0034.

[8] P. A. Zyla *et al.*, "Review of Particle Physics", *PTEP*, vol. 2020, no. 8, p. 083C01, 2020. DOI: 10.1093/ptep/ptaa104.

[9] Planck Collaboration, "Planck 2018 results. VIII. Gravitational lensing", *Astron. Astrophys.*, vol. 641, A8, 2020. DOI: 10.1051/0004-6361/201833886. arXiv: 1807.06210 [astro-ph.CO].

[10] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.*, vol. 13, J. C. Taylor, Ed., pp. 508–509, 1964. DOI: 10.1103/PhysRevLett.13.508.

[11] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B*, vol. 716, pp. 1–29, 2012. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[12] CMS Collaboration, "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC", *Phys. Lett. B*, vol. 716, pp. 30–61, 2012. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[13] CMS Collaboration, "Combined measurements of Higgs boson couplings in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Eur. Phys. J. C*, vol. 79, no. 5, p. 421, 2019. DOI: 10.1140/epjc/s10052-019-6909-y. arXiv: 1809.10733 [hep-ex].

[14] ATLAS Collaboration, "Combined measurement of the total and differential cross sections in the $H \rightarrow \gamma\gamma$ and the $H \rightarrow ZZ^* \rightarrow 4\ell$ decay channels at $\sqrt{s} = 13$ TeV with the ATLAS detector", CERN, Geneva, Tech. Rep., 2022. [Online]. Available: https://cds.cern.ch/record/2799603.

[15] CMS Collaboration, "Search for charged Higgs bosons decaying into a top and a bottom quark in the all-jet final state of pp collisions at $\sqrt{s}$ = 13 TeV", *JHEP*, vol. 07, p. 126, 2020. DOI: 10.1007/JHEP07(2020)126. arXiv: 2001.07763 [hep-ex].

[16] W. Pauli, "The Connection Between Spin and Statistics", *Phys. Rev.*, vol. 58, pp. 716–722, 1940. DOI: 10.1103/PhysRev.58.716.

[17] Belle Collaboration, "Observation of a resonance-like structure in the $\pi^{\pm}\psi'$ mass distribution in exclusive $B \rightarrow K\pi^{\pm}\psi'$ decays", *Phys. Rev. Lett.*, vol. 100, D. Son and S. K. Oh, Eds., p. 142 001, 2008. DOI: 10.1103/PhysRevLett.100.142001. arXiv: 0708.1790 [hep-ex].

[18] Belle Collaboration, "Experimental constraints on the spin and parity of the $Z(4430)^+$", *Phys. Rev. D*, vol. 88, no. 7, p. 074 026, 2013. DOI: 10.1103/PhysRevD.88.074026. arXiv: 1306.4894 [hep-ex].

[19] LHCb Collaboration, "Observation of the resonant character of the $Z(4430)^-$ state", *Phys. Rev. Lett.*, vol. 112, no. 22, p. 222 002, 2014. DOI: 10.1103/PhysRevLett.112.222002. arXiv: 1404.1903 [hep-ex].

[20] LHCb Collaboration, "Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \rightarrow J/\psi K^- p$ Decays", *Phys. Rev. Lett.*, vol. 115, p. 072 001, 2015. DOI: 10.1103/PhysRevLett.115.072001. arXiv: 1507.03414 [hep-ex].

[21] LHCb Collaboration, "Observation of a narrow pentaquark state, $P_c(4312)^+$, and of two-peak structure of the $P_c(4450)^+$", *Phys. Rev. Lett.*, vol. 122, no. 22, p. 222 001, 2019. DOI: 10.1103/PhysRevLett.122.222001. arXiv: 1904.03947 [hep-ex].

[22] P. A. M. Dirac, "The quantum theory of the electron", *Proc. Roy. Soc. Lond. A*, vol. 117, pp. 610–624, 1928. DOI: 10.1098/rspa.1928.0023.

[23]   S. Tomonaga, "On a relativistically invariant formulation of the quantum theory of wave fields", *Prog. Theor. Phys.*, vol. 1, pp. 27–42, 1946. DOI: 10.1143/PTP.1.27.

[24]   J. Schwinger, "On Quantum-Electrodynamics and the Magnetic Moment of the Electron", *Phys. Rev.*, vol. 73, pp. 416–417, 4 1948. DOI: 10.1103/PhysRev.73.416. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRev.73.416.

[25]   R. P. Feynman, "The Theory of positrons", *Phys. Rev.*, vol. 76, L. M. Brown, Ed., pp. 749–759, 1949. DOI: 10.1103/PhysRev.76.749.

[26]   R. P. Feynman and M. Gell-Mann, "Theory of Fermi interaction", *Phys. Rev.*, vol. 109, L. M. Brown, Ed., pp. 193–198, 1958. DOI: 10.1103/PhysRev.109.193.

[27]   C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, "Experimental Test of Parity Conservation in β Decay", *Phys. Rev.*, vol. 105, pp. 1413–1414, 1957. DOI: 10.1103/PhysRev.105.1413.

[28]   N. Cabibbo, "Unitary Symmetry and Leptonic Decays", *Phys. Rev. Lett.*, vol. 10, pp. 531–533, 1963. DOI: 10.1103/PhysRevLett.10.531.

[29]   M. Kobayashi and T. Maskawa, "$\mathcal{CP}$ Violation in the Renormalizable Theory of Weak Interaction", *Prog. Theor. Phys.*, vol. 49, pp. 652–657, 1973. DOI: 10.1143/PTP.49.652.

[30]   F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.*, vol. 13, J. C. Taylor, Ed., pp. 321–323, 1964. DOI: 10.1103/PhysRevLett.13.321.

[31]   G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global Conservation Laws and Massless Particles", *Phys. Rev. Lett.*, vol. 13, J. C. Taylor, Ed., pp. 585–587, 1964. DOI: 10.1103/PhysRevLett.13.585.

[32]   J. Goldstone, "Field Theories with Superconductor Solutions", *Nuovo Cim.*, vol. 19, pp. 154–164, 1961. DOI: 10.1007/BF02812722.

[33]   Super-Kamiokande Collaboration, "Evidence for oscillation of atmospheric neutrinos", *Phys. Rev. Lett.*, vol. 81, pp. 1562–1567, 1998. DOI: 10.1103/PhysRevLett.81.1562. arXiv: hep-ex/9807003.

[34]   KATRIN Collaboration, "Improved Upper Limit on the Neutrino Mass from a Direct Kinematic Method by KATRIN", *Phys. Rev. Lett.*, vol. 123, no. 22, p. 221 802, 2019. DOI: 10.1103/PhysRevLett.123.221802. arXiv: 1909.06048 [hep-ex].

[35]   K. C. Freeman, "On the disks of spiral and SO Galaxies", *Astrophys. J.*, vol. 160, p. 811, 1970. DOI: 10.1086/150474.

[36] G. Bertone and D. Hooper, "History of dark matter", *Rev. Mod. Phys.*, vol. 90, no. 4, p. 045 002, 2018. DOI: `10.1103/RevModPhys.90.045002`. arXiv: `1605.04909 [astro-ph.CO]`.

[37] A. D. Sakharov, "Violation of $\mathcal{CP}$ Invariance, C asymmetry, and baryon asymmetry of the universe", *Pisma Zh. Eksp. Teor. Fiz.*, vol. 5, pp. 32–35, 1967. DOI: `10.1070/PU1991v034n05ABEH002497`.

[38] S. L. Glashow and S. Weinberg, "Natural Conservation Laws for Neutral Currents", *Phys. Rev. D*, vol. 15, p. 1958, 1977. DOI: `10.1103/PhysRevD.15.1958`.

[39] S. Davidson and H. E. Haber, "Basis-independent methods for the two-Higgs-doublet model", *Phys. Rev. D*, vol. 72, p. 035 004, 2005, [Erratum: Phys.Rev.D 72, 099902 (2005)]. DOI: `10.1103/PhysRevD.72.099902`. arXiv: `hep-ph/0504050`.

[40] L. J. Hall, J. D. Lykken, and S. Weinberg, "Supergravity as the Messenger of Supersymmetry Breaking", *Phys. Rev. D*, vol. 27, pp. 2359–2378, 1983. DOI: `10.1103/PhysRevD.27.2359`.

[41] S. P. Martin, "A Supersymmetry primer", *Adv. Ser. Direct. High Energy Phys.*, vol. 18, G. L. Kane, Ed., pp. 1–98, 1998. DOI: `10.1142/9789812839657_0001`. arXiv: `hep-ph/9709356`.

[42] Y. Okada, M. Yamaguchi, and T. Yanagida, "Upper bound of the lightest Higgs boson mass in the minimal supersymmetric standard model", *Prog. Theor. Phys.*, vol. 85, pp. 1–6, 1991. DOI: `10.1143/ptp/85.1.1`.

[43] J. R. Ellis, G. Ridolfi, and F. Zwirner, "Radiative corrections to the masses of supersymmetric Higgs bosons", *Phys. Lett. B*, vol. 257, pp. 83–91, 1991. DOI: `10.1016/0370-2693(91)90863-L`.

[44] A. Djouadi and J. Quevillon, "The MSSM Higgs sector at a high M_{SUSY}: reopening the low $\tan\beta$ regime and heavy Higgs searches", *JHEP*, vol. 10, p. 028, 2013. DOI: `10.1007/JHEP10(2013)028`. arXiv: `1304.1787 [hep-ph]`.

[45] A. Djouadi, L. Maiani, G. Moreau, A. Polosa, J. Quevillon, and V. Riquer, "The post-Higgs MSSM scenario: Habemus MSSM?", *Eur. Phys. J. C*, vol. 73, p. 2650, 2013. DOI: `10.1140/epjc/s10052-013-2650-0`. arXiv: `1307.5205 [hep-ph]`.

[46] S. Kanemura and C. P. Yuan, "Testing supersymmetry in the associated production of $\mathcal{CP}$-odd and charged Higgs bosons", *Phys. Lett. B*, vol. 530, pp. 188–196, 2002. DOI: `10.1016/S0370-2693(02)01321-7`. arXiv: `hep-ph/0112165`.

[47] A. G. Akeroyd and M. A. Diaz, "Searching for a light fermiophobic Higgs boson at the Tevatron", *Phys. Rev. D*, vol. 67, p. 095 007, 2003. DOI: `10.1103/PhysRevD.67.095007`. arXiv: `hep-ph/0301203`.

[48] A. Belyaev, Q.-H. Cao, D. Nomura, K. Tobe, and C. P. Yuan, "Light MSSM Higgs boson scenario and its test at hadron colliders", *Phys. Rev. Lett.*, vol. 100, p. 061 801, 2008. DOI: 10.1103/PhysRevLett.100.061801. arXiv: hep-ph/0609079.

[49] A. Djouadi, "The Anatomy of electro-weak symmetry breaking. II. The Higgs bosons in the minimal supersymmetric model", *Phys. Rept.*, vol. 459, pp. 1–241, 2008. DOI: 10.1016/j.physrep.2007.10.005. arXiv: hep-ph/0503173.

[50] A. G. Akeroyd *et al.*, "Prospects for charged Higgs searches at the LHC", *Eur. Phys. J. C*, vol. 77, no. 5, p. 276, 2017. DOI: 10.1140/epjc/s10052-017-4829-2. arXiv: 1607.01320 [hep-ph].

[51] D. Berdine, N. Kauer, and D. Rainwater, "Breakdown of the Narrow Width Approximation for New Physics", *Phys. Rev. Lett.*, vol. 99, p. 111 601, 2007. DOI: 10.1103/PhysRevLett.99.111601. arXiv: hep-ph/0703058.

[52] R. Harlander, M. Kramer, and M. Schumacher, "Bottom-quark associated Higgs-boson production: reconciling the four- and five-flavour scheme approach", Dec. 2011. arXiv: 1112.3478 [hep-ph].

[53] M. Flechl, R. Klees, M. Kramer, M. Spira, and M. Ubiali, "Improved cross-section predictions for heavy charged Higgs boson production at the LHC", *Phys. Rev. D*, vol. 91, no. 7, p. 075 015, 2015. DOI: 10.1103/PhysRevD.91.075015. arXiv: 1409.5615 [hep-ph].

[54] C. Degrande, R. Frederix, V. Hirschi, M. Ubiali, M. Wiesemann, and M. Zaro, "Accurate predictions for charged Higgs production: Closing the $m_{H^\pm} \sim m_t$ window", *Phys. Lett. B*, vol. 772, pp. 87–92, 2017. DOI: 10.1016/j.physletb.2017.06.037. arXiv: 1607.05291 [hep-ph].

[55] LHC Higgs Cross Section Working Group, "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector", vol. 2/2017, Oct. 2016. DOI: 10.23731/CYRM-2017-002. arXiv: 1610.07922 [hep-ph].

[56] M. Misiak and M. Steinhauser, "Weak radiative decays of the B meson and bounds on $M_{H^\pm}$ in the Two-Higgs-Doublet Model", *Eur. Phys. J. C*, vol. 77, no. 3, p. 201, 2017. DOI: 10.1140/epjc/s10052-017-4776-y. arXiv: 1702.04571 [hep-ph].

[57] M. Misiak *et al.*, "Updated NNLO QCD predictions for the weak radiative B-meson decays", *Phys. Rev. Lett.*, vol. 114, no. 22, p. 221 801, 2015. DOI: 10.1103/PhysRevLett.114.221801. arXiv: 1503.01789 [hep-ph].

[58] ALEPH, DELPHI, L3, OPAL, LEP Collaborations, "Search for Charged Higgs bosons: Combined Results Using LEP Data", *Eur. Phys. J. C*, vol. 73, p. 2463, 2013. DOI: 10.1140/epjc/s10052-013-2463-1. arXiv: 1301.6065 [hep-ex].

[59] CMS Collaboration, "Search for additional neutral MSSM Higgs bosons in the $\tau\tau$ final state in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP*, vol. 09, p. 007, 2018. DOI: 10.1007/JHEP09(2018)007. arXiv: 1803.06553 [hep-ex].

[60] CMS Collaboration, "Search for a new scalar resonance decaying to a pair of Z bosons in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP*, vol. 06, p. 127, 2018, [Erratum: JHEP 03, 128 (2019)]. DOI: 10.1007/JHEP06(2018)127. arXiv: 1804.01939 [hep-ex].

[61] CMS Collaboration, "Search for Higgs boson pair production in events with two bottom quarks and two tau leptons in proton-proton collisions at $\sqrt{s}$=13 TeV", *Phys. Lett. B*, vol. 778, pp. 101–127, 2018. DOI: 10.1016/j.physletb.2018.01.001. arXiv: 1707.02909 [hep-ex].

[62] CMS Collaboration, "Search for resonant and nonresonant Higgs boson pair production in the $\mathrm{b\bar{b}}\ell\nu\ell\nu$ final state in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP*, vol. 01, p. 054, 2018. DOI: 10.1007/JHEP01(2018)054. arXiv: 1708.04188 [hep-ex].

[63] ATLAS Collaboration, "Search for additional heavy neutral Higgs and gauge bosons in the ditau final state produced in 36 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *JHEP*, vol. 01, p. 055, 2018. DOI: 10.1007/JHEP01(2018)055. arXiv: 1709.07242 [hep-ex].

[64] ATLAS Collaboration, "Search for heavy resonances decaying into a pair of Z bosons in the $\ell^+\ell^-\ell'^+\ell'^-$ and $\ell^+\ell^-\nu\bar{\nu}$ final states using 139 fb^{-1} of proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Eur. Phys. J. C*, vol. 81, no. 4, p. 332, 2021. DOI: 10.1140/epjc/s10052-021-09013-y. arXiv: 2009.14791 [hep-ex].

[65] ATLAS Collaboration, "Searches for Higgs boson pair production in the $hh \to bb\tau\tau$, $\gamma\gamma WW^*$, $\gamma\gamma bb$, $bbbb$ channels with the ATLAS detector", *Phys. Rev. D*, vol. 92, p. 092 004, 2015. DOI: 10.1103/PhysRevD.92.092004. arXiv: 1509.04670 [hep-ex].

[66] ATLAS Collaboration, "Search for charged Higgs bosons decaying via $H^\pm \to \tau^\pm\nu$ in fully hadronic final states using pp collision data at $\sqrt{s} = 8$ TeV with the ATLAS detector", *JHEP*, vol. 03, p. 088, 2015. DOI: 10.1007/JHEP03(2015)088. arXiv: 1412.6663 [hep-ex].

[67] ATLAS Collaboration, "Search for charged Higgs bosons decaying into top and bottom quarks at $\sqrt{s} = 13$ TeV with the ATLAS detector", *JHEP*, vol. 11, p. 085, 2018. DOI: 10.1007/JHEP11(2018)085. arXiv: 1808.03599 [hep-ex].

[68] CMS Collaboration, "Search for charged Higgs bosons in the $H^{\pm} \to \tau^{\pm}\nu_{\tau}$ decay channel in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP*, vol. 07, p. 142, 2019. DOI: 10.1007/JHEP07(2019)142. arXiv: 1903.04560 [hep-ex].

[69] CMS Collaboration, "Search for a charged Higgs boson decaying into top and bottom quarks in events with electrons or muons in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP*, vol. 01, p. 096, 2020. DOI: 10.1007/JHEP01(2020)096. arXiv: 1908.09206 [hep-ex].

[70] ATLAS Collaboration, "Search for charged Higgs bosons decaying into a top quark and a bottom quark at $\sqrt{s} = 13$ TeV with the ATLAS detector", *JHEP*, vol. 06, p. 145, 2021. DOI: 10.1007/JHEP06(2021)145. arXiv: 2102.10076 [hep-ex].

[71] CMS Collaboration, "Search for a light charged Higgs boson in the $H^{\pm} \to cs$ channel in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Phys. Rev. D*, vol. 102, p. 072001, 7 2020. DOI: 10.1103/PhysRevD.102.072001. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.102.072001.

[72] CMS Collaboration, "Search for a charged Higgs boson decaying to charm and bottom quarks in proton-proton collisions at $\sqrt{s} = 8$ TeV", *JHEP*, vol. 11, p. 115, 2018. DOI: 10.1007/JHEP11(2018)115. arXiv: 1808.06575 [hep-ex].

[73] ATLAS Collaboration, "Search for a light charged Higgs boson in $t \to H^{\pm}b$ decays, with $H^{\pm} \to cb$, in the lepton+jets final state in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", 2021.

[74] CMS Collaboration, "Search for a light charged Higgs boson decaying to a W boson and a $\mathcal{CP}$-odd Higgs boson in final states with $e\mu\mu$ or $\mu\mu\mu$ in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Phys. Rev. Lett.*, vol. 123, no. 13, p. 131802, 2019. DOI: 10.1103/PhysRevLett.123.131802. arXiv: 1905.07453 [hep-ex].

[75] CMS Collaboration, "Search for Charged Higgs Bosons Produced via Vector Boson Fusion and Decaying into a Pair of W and Z Bosons Using pp Collisions at $\sqrt{s} = 13$ TeV", *Phys. Rev. Lett.*, vol. 119, no. 14, p. 141802, 2017. DOI: 10.1103/PhysRevLett.119.141802. arXiv: 1705.02942 [hep-ex].

[76] ATLAS Collaboration, "Search for a Charged Higgs Boson Produced in the Vector-Boson Fusion Mode with Decay $H^{\pm} \to W^{\pm}Z$ using pp Collisions at $\sqrt{s} = 8$ TeV with the ATLAS Experiment", *Phys. Rev. Lett.*, vol. 114, no. 23, p. 231801, 2015. DOI: 10.1103/PhysRevLett.114.231801. arXiv: 1503.04233 [hep-ex].

[77] CMS Collaboration, "Search for charged Higgs bosons produced in vector boson fusion processes and decaying into vector boson pairs in proton–proton collisions at $\sqrt{s} = 13$ TeV", *Eur. Phys. J. C*, vol. 81, no. 8, p. 723, 2021. DOI: 10.1140/epjc/s10052-021-09472-3. arXiv: 2104.04762 [hep-ex].

[78] UA1 Collaboration, "Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV", *Phys. Lett. B*, vol. 122, pp. 103–116, 1983. DOI: 10.1016/0370-2693(83)91177-2.

[79] UA1 Collaboration, "Experimental Observation of Lepton Pairs of Invariant Mass Around 95 GeV/c^2 at the CERN SPS Collider", *Phys. Lett. B*, vol. 126, pp. 398–410, 1983. DOI: 10.1016/0370-2693(83)90188-0.

[80] UA2 Collaboration, "Evidence for $Z^0 \to e^+ e^-$ at the CERN $\bar{p}p$ Collider", *Phys. Lett. B*, vol. 129, pp. 130–140, 1983. DOI: 10.1016/0370-2693(83)90744-X.

[81] NA48 Collaboration, "A New measurement of direct CP violation in two pion decays of the neutral kaon", *Phys. Lett. B*, vol. 465, pp. 335–348, 1999. DOI: 10.1016/S0370-2693(99)01030-8. arXiv: hep-ex/9909022.

[82] G. Rumolo, H. Bartosik, E. Belli, P. Dijkstal, G. Iadarola, K. Li, L. Mether, A. Romano, M. Schenk, and F. Zimmermann, "Electron Cloud Effects at the LHC and LHC Injectors", in *8th International Particle Accelerator Conference*, May 2017. DOI: 10.18429/JACoW-IPAC2017-MOZA1.

[83] CMS Collaboration, "The CMS Experiment at the CERN LHC", *JINST*, vol. 3, S08004, 2008. DOI: 10.1088/1748-0221/3/08/S08004.

[84] CMS Collaboration, "CMS, the magnet project: Technical design report", May 1997.

[85] CMS Collaboration, "The CMS Phase-1 Pixel Detector Upgrade", CERN, Geneva, Tech. Rep., 2020. [Online]. Available: https://cds.cern.ch/record/2745805.

[86] C. W. Fabjan and F. Gianotti, "Calorimetry for particle physics", *Reviews of Modern Physics*, vol. 75, no. 4, p. 1243, 2003.

[87] S. Menke, *The electromagnetic shower simulator*, Accessed: 2021-06-17. [Online]. Available: https://www.mpp.mpg.de/~menke/elss/home.shtml.

[88] CMS Collaboration, "CMS: The electromagnetic calorimeter. Technical design report", Dec. 1997.

[89] CMS Collaboration, "Energy resolution of the barrel of the CMS electromagnetic calorimeter", *JINST*, vol. 2, P04004, 2007. DOI: 10.1088/1748-0221/2/04/P04004.

[90] CMS Collaboration, "ECAL 2016 refined calibration and Run2 summary plots", 2020. [Online]. Available: https://cds.cern.ch/record/2717925.

[91] CMS Collaboration, "The CMS hadron calorimeter project: Technical Design Report", 1997. [Online]. Available: https://cds.cern.ch/record/357153.

[92] CMS Collaboration, "The CMS barrel calorimeter response to particle beams from 2-GeV/c to 350-GeV/c", *Eur. Phys. J. C*, vol. 60, pp. 359–373, 2009, [Erratum: Eur.Phys.J.C 61, 353–356 (2009)]. DOI: 10.1140/epjc/s10052-009-0959-5.

[93] CMS Collaboration, "CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter", J. Mans *et al.*, Eds., Sep. 2012. DOI: 10.2172/1151651.

[94] CMS Collaboration, *The CMS muon project: Technical Design Report*, ser. Technical design report. CMS. Geneva: CERN, 1997. [Online]. Available: https://cds.cern.ch/record/343814.

[95] CMS Collaboration, "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV", *JINST*, vol. 13, no. 06, P06015, 2018. DOI: 10.1088/1748-0221/13/06/P06015. arXiv: 1804.04528 [physics.ins-det].

[96] A Colaleo, A Safonov, A Sharma, and M Tytgat, "CMS Technical Design Report for the Muon Endcap GEM Upgrade", Tech. Rep., 2015. [Online]. Available: https://cds.cern.ch/record/2021453.

[97] CMS Collaboration, "Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV", *JHEP*, vol. 07, p. 161, 2018. DOI: 10.1007/JHEP07(2018)161. arXiv: 1802.02613 [hep-ex].

[98] P. I. C. M. Group, "Mtca.0 r1.0 micro telecommunications computing architecture", 2006. [Online]. Available: http://www.picmg.org.

[99] CMS Collaboration, "The CMS Barrel Muon trigger upgrade", *JINST*, vol. 12, no. 01, p. C01095, 2017. DOI: 10.1088/1748-0221/12/01/C01095.

[100] CMS Collaboration, "Boosted Decision Trees in the Level-1 Muon Endcap Trigger at CMS", *J. Phys. Conf. Ser.*, vol. 1085, no. 4, p. 042 042, 2018. DOI: 10.1088/1742-6596/1085/4/042042.

[101] R. Frazier, S. Fayer, G. Hall, C. Hunt, G. Iles, D. Newbold, and A. Rose, "A demonstration of a time multiplexed trigger for the CMS experiment", *JINST*, vol. 7, p. C01060, 2012. DOI: 10.1088/1748-0221/7/01/C01060.

[102] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *JINST*, vol. 12, no. 10, P10003, 2017. DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965 [physics.ins-det].

[103] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", *JINST*, vol. 13, no. 05, P05011, 2018. DOI: 10.1088/1748-0221/13/05/P05011. arXiv: 1712.07158 [physics.ins-det].

[104] D. Guest, J. Collado, P. Baldi, S.-C. Hsu, G. Urban, and D. Whiteson, "Jet Flavor Classification in High-Energy Physics with Deep Neural Networks", *Phys. Rev. D*, vol. 94, no. 11, p. 112 002, 2016. DOI: 10.1103/PhysRevD.94.112002. arXiv: 1607.08633 [hep-ex].

[105] CMS Collaboration, "B-Jet Trigger Performance in Run 2", 2019. [Online]. Available: https://cds.cern.ch/record/2708546.

[106] CMS Collaboration, "Early 2018 High-Level Trigger rates", 2018. [Online]. Available: https://cds.cern.ch/record/2644379.

[107] W. Adam, B. Mangano, T. Speer, and T. Todorov, "Track Reconstruction in the CMS tracker", CERN, Geneva, Tech. Rep., 2006. [Online]. Available: http://cds.cern.ch/record/934067.

[108] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST*, vol. 9, no. 10, P10009, 2014. DOI: 10.1088/1748-0221/9/10/P10009. arXiv: 1405.6569 [physics.ins-det].

[109] CMS Collaboration, "Track impact parameter resolution for the full pseudo rapidity coverage in the 2017 dataset with the CMS Phase-1 Pixel detector", 2020. [Online]. Available: https://cds.cern.ch/record/2743740.

[110] K. Rose, "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems", *IEEE Proc.*, vol. 86, no. 11, pp. 2210–2239, 1998. DOI: 10.1109/5.726788.

[111] R Frühwirth, W. Waltenberger, and P. Vanlaer, "Adaptive Vertex Fitting", CERN, Geneva, Tech. Rep., 2007. [Online]. Available: https://cds.cern.ch/record/1027031.

[112] E. Bols, J. Kieseler, M. Verzetti, M. Stoye, and A. Stakia, "Jet Flavour Classification Using DeepJet", *JINST*, vol. 15, no. 12, P12012, 2020. DOI: 10.1088/1748-0221/15/12/P12012. arXiv: 2008.10519 [hep-ex].

[113] CMS Collaboration, "Performance of CMS Muon Reconstruction in pp Collision Events at $\sqrt{s} = 7$ TeV", *JINST*, vol. 7, P10002, 2012. DOI: 10.1088/1748-0221/7/10/P10002. arXiv: 1206.4071 [physics.ins-det].

[114] W. Adam, R. Frühwirth, A. Strandlie, and T. Todor, "Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC", *Journal of Physics G: Nuclear and Particle Physics*, vol. 31, no. 9, N9, Jan. 2005.

[115] CMS Collaboration, "Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} =8$ TeV", *JINST*, vol. 10, no. 06, P06005, 2015. DOI: 10.1088/1748-0221/10/06/P06005. arXiv: 1502.02701 [physics.ins-det].

[116] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-kt jet clustering algorithm", *Journal of High Energy Physics*, vol. 2008, no. 04, p. 063, 2008.

[117] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, "Better jet clustering algorithms", *JHEP*, vol. 08, p. 001, 1997. DOI: `10.1088/1126-6708/1997/08/001`. arXiv: `hep-ph/9707323`.

[118] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber, "Longitudinally invariant K_t clustering algorithms for hadron hadron collisions", *Nucl. Phys. B*, vol. 406, pp. 187–224, 1993. DOI: `10.1016/0550-3213(93)90166-M`.

[119] S. D. Ellis and D. E. Soper, "Successive combination jet algorithm for hadron collisions", *Phys. Rev. D*, vol. 48, pp. 3160–3166, 1993. DOI: `10.1103/PhysRevD.48.3160`. arXiv: `hep-ph/9305266`.

[120] G. P. Salam, "Towards Jetography", *Eur. Phys. J. C*, vol. 67, pp. 637–686, 2010. DOI: `10.1140/epjc/s10052-010-1314-6`. arXiv: `0906.1833 [hep-ph]`.

[121] A. Abdesselam *et al.*, "Boosted Objects: A Probe of Beyond the Standard Model Physics", *Eur. Phys. J. C*, vol. 71, p. 1661, 2011. DOI: `10.1140/epjc/s10052-011-1661-y`. arXiv: `1012.5412 [hep-ph]`.

[122] CMS collaboration, "Determination of jet energy calibration and transverse momentum resolution in cms", *Journal of Instrumentation*, vol. 6, no. 11, P11002, 2011.

[123] CMS Collaboration, "Pileup mitigation at CMS in 13 TeV data", *JINST*, vol. 15, no. 09, P09018, 2020. DOI: `10.1088/1748-0221/15/09/P09018`. arXiv: `2003.00503 [hep-ex]`.

[124] CMS Collaboration, "Jet energy scale and resolution measurement with Run 2 Legacy Data Collected by CMS at 13 TeV", 2021. [Online]. Available: `https://cds.cern.ch/record/2792322`.

[125] CMS Collaboration, "Jet algorithms performance in 13 tev data", *CMS Physics Analysis Summary CMS-PAS-JME-16-003*, 2017.

[126] CMS Collaboration, "Performance of the DeepJet b tagging algorithm using $41.9\ fb^{-1}$ of data from proton-proton collisions at 13 TeV with Phase 1 CMS detector", 2018. [Online]. Available: `https://cds.cern.ch/record/2646773`.

[127] CMS Collaboration, "Reconstruction and identification of τ-lepton decays to hadrons and ν_τ at CMS", *JINST*, vol. 11, no. 01, P01019, 2016. DOI: `10.1088/1748-0221/11/01/P01019`. arXiv: `1510.07488 [physics.ins-det]`.

[128] CMS Collaboration, "Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV", *JINST*, vol. 13, no. 10, P10005, 2018. DOI: `10.1088/1748-0221/13/10/P10005`. arXiv: `1809.02816 [hep-ex]`.

[129] CMS Collaboration, "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector", *JINST*, vol. 14, no. 07, P07004, 2019. DOI: `10.1088/1748-0221/14/07/P07004`. arXiv: `1903.06078 [hep-ex]`.

[130] S. Agostinelli *et al.*, "GEANT4–a simulation toolkit", *Nucl. Instrum. Meth. A*, vol. 506, pp. 250–303, 2003. DOI: `10.1016/S0168-9002(03)01368-8`.

[131] R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, J. Rojo, and M. Ubiali, "Unbiased global determination of parton distributions and their uncertainties at NNLO and at LO", *Nucl. Phys. B*, vol. 855, pp. 153–221, 2012. DOI: `10.1016/j.nuclphysb.2011.09.024`. arXiv: `1107.2652 [hep-ph]`.

[132] R. D. Ball, V. Bertone, S. Carrazza, L. D. Debbio, S. Forte, P. Groth-Merrild, A. Guffanti, N. P. Hartland, Z. Kassabov, J. I. Latorre, and et al., "Parton distributions from high-precision collider data", *The European Physical Journal C*, vol. 77, no. 10, 2017, ISSN: 1434-6052. DOI: `10.1140/epjc/s10052-017-5199-5`. [Online]. Available: `http://dx.doi.org/10.1140/epjc/s10052-017-5199-5`.

[133] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, "An introduction to PYTHIA 8.2", *Comput. Phys. Commun.*, vol. 191, pp. 159–177, 2015. DOI: `10.1016/j.cpc.2015.01.024`. arXiv: `1410.3012 [hep-ph]`.

[134] V. Khachatryan *et al.*, "Event generator tunes obtained from underlying event and multiparton scattering measurements", *Eur. Phys. J. C*, vol. 76, no. 3, p. 155, 2016. DOI: `10.1140/epjc/s10052-016-3988-x`. arXiv: `1512.00815 [hep-ex]`.

[135] CMS Collaboration, "Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements", *Eur. Phys. J. C*, vol. 80, no. 1, p. 4, 2020. DOI: `10.1140/epjc/s10052-019-7499-4`. arXiv: `1903.12179 [hep-ex]`.

[136] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP*, vol. 07, p. 079, 2014. DOI: `10.1007/JHEP07(2014)079`. arXiv: `1405.0301 [hep-ph]`.

[137] E. L. Berger, T. Han, J. Jiang, and T. Plehn, "Associated production of a top quark and a charged Higgs boson", *Phys. Rev. D*, vol. 71, p. 115 012, 2005. DOI: `10.1103/PhysRevD.71.115012`. arXiv: `hep-ph/0312286`.

[138] A. Djouadi, J. Kalinowski, M. Mühlleitner, and M. Spira, "HDECAY: Twenty++ years after", *Computer Physics Communications*, vol. 238, 214–231, 2019, ISSN: 0010-4655. DOI: 10.1016/j.cpc.2018.12.010. [Online]. Available: http://dx.doi.org/10.1016/j.cpc.2018.12.010.

[139] P. Nason, "A New method for combining NLO QCD with shower Monte Carlo algorithms", *JHEP*, vol. 11, p. 040, 2004. DOI: 10.1088/1126-6708/2004/11/040. arXiv: hep-ph/0409146.

[140] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP*, vol. 06, p. 043, 2010. DOI: 10.1007/JHEP06(2010)043. arXiv: 1002.2581 [hep-ph].

[141] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method", *JHEP*, vol. 11, p. 070, 2007. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph].

[142] H. B. Hartanto, B. Jager, L. Reina, and D. Wackeroth, "Higgs boson production in association with top quarks in the POWHEG BOX", *Phys. Rev. D*, vol. 91, no. 9, p. 094 003, 2015. DOI: 10.1103/PhysRevD.91.094003. arXiv: 1501.04498 [hep-ph].

[143] K. Rehermann and B. Tweedie, "Efficient Identification of Boosted Semileptonic Top Quarks at the LHC", *JHEP*, vol. 03, p. 059, 2011. DOI: 10.1007/JHEP03(2011)059. arXiv: 1007.2221 [hep-ph].

[144] M. Cacciari, G. P. Salam, and G. Soyez, "Fastjet user manual", *The European Physical Journal C*, vol. 72, no. 3, pp. 1–54, 2012.

[145] ATLAS, CMS, LHC Higgs Combination Group, "Procedure for the LHC Higgs boson search combination in Summer 2011", Aug. 2011.

[146] A. L. Read, "Linear interpolation of histograms", *Nucl. Instrum. Meth. A*, vol. 425, pp. 357–360, 1999. DOI: 10.1016/S0168-9002(98)01347-3.

[147] T. Junk, "Confidence level computation for combining searches with small statistics", *Nucl. Instrum. Meth. A*, vol. 434, pp. 435–443, 1999. DOI: 10.1016/S0168-9002(99)00498-2. arXiv: hep-ex/9902006.

[148] A. L. Read, "Presentation of search results: The CL(s) technique", *J. Phys. G*, vol. 28, M. R. Whalley and L. Lyons, Eds., pp. 2693–2704, 2002. DOI: 10.1088/0954-3899/28/10/313.

[149] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J. C*, vol. 71, p. 1554, 2011, [Erratum: Eur.Phys.J.C 73, 2501 (2013)]. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an].

[150] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, H. Voss, M Backes, T Carli, O Cohen, A Christov, *et al.*, "Tmva-toolkit for multivariate data analysis", *arXiv preprint physics/0703039*, 2007.

[151] CMS Collaboration, "Search for direct production of supersymmetric partners of the top quark in the all-jets final state in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP*, vol. 10, p. 005, 2017. DOI: 10.1007/JHEP10(2017)005. arXiv: 1707.03316 [hep-ex].

[152] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, "Soft Drop", *JHEP*, vol. 05, p. 146, 2014. DOI: 10.1007/JHEP05(2014)146. arXiv: 1402.2657 [hep-ph].

[153] CMS Collaboration, "Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS", *Eur. Phys. J. C*, vol. 81, no. 9, p. 800, 2021. DOI: 10.1140/epjc/s10052-021-09538-2. arXiv: 2104.01927 [hep-ex].

[154] V. Khachatryan *et al.*, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", *JINST*, vol. 12, no. 02, P02014, 2017. DOI: 10.1088/1748-0221/12/02/P02014. arXiv: 1607.03663 [hep-ex].

[155] J. Thaler and K. Van Tilburg, "Identifying Boosted Objects with N-subjettiness", *JHEP*, vol. 03, p. 015, 2011. DOI: 10.1007/JHEP03(2011)015. arXiv: 1011.2268 [hep-ph].

[156] "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", 2016.

[157] R. J. Barlow and C. Beeston, "Fitting using finite Monte Carlo samples", *Comput. Phys. Commun.*, vol. 77, pp. 219–228, 1993. DOI: 10.1016/0010-4655(93)90005-W.

[158] J. S. Conway, "Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra", in *PHYSTAT 2011*, 2011, pp. 115–120. DOI: 10.5170/CERN-2011-006.115. arXiv: 1103.0354 [physics.data-an].

[159] J. Butterworth *et al.*, "PDF4LHC recommendations for LHC Run II", *J. Phys. G*, vol. 43, p. 023 001, 2016. DOI: 10.1088/0954-3899/43/2/023001. arXiv: 1510.03865 [hep-ph].

[160] R. D. Cousins, "Generalization of chisquare goodness-of-fit test for binned data using saturated models, with application to histograms", *URL: http://www. physics. ucla. edu/~ cousins/stats/cousins_saturated. pdf (visited on 11/04/2021)*, 2013.

[161] E. Bagnaschi, H. Bahl, E. Fuchs, T. Hahn, S. Heinemeyer, S. Liebler, S. Patel, P. Slavich, T. Stefaniak, C. E. M. Wagner, and et al., "MSSM Higgs boson searches at the LHC: benchmark scenarios for Run 2 and beyond", *The European Physical Journal C*, vol. 79, no. 7, 2019, ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-019-7114-8. [Online]. Available: http://dx.doi.org/10.1140/epjc/s10052-019-7114-8.

[162] S. Heinemeyer, W. Hollik, and G. Weiglein, "FeynHiggs: a program for the calculation of the masses of the neutral $\mathcal{CP}$-even Higgs bosons in the MSSM", *Computer Physics Communications*, vol. 124, no. 1, 76–89, 2000, ISSN: 0010-4655. DOI: 10.1016/s0010-4655(99)00364-1. [Online]. Available: http://dx.doi.org/10.1016/S0010-4655(99)00364-1.

[163] S. Heinemeyer, W. Hollik, and G. Weiglein, "The Masses of the neutral $\mathcal{CP}$-even Higgs bosons in the MSSM: Accurate analysis at the two loop level", *Eur. Phys. J. C*, vol. 9, pp. 343–366, 1999. DOI: 10.1007/s100529900006. arXiv: hep-ph/9812472.

[164] G. Degrassi, S. Heinemeyer, W. Hollik, P. Slavich, and G. Weiglein, "Towards high-precision predictions for the MSSM Higgs sector", *The European Physical Journal C*, vol. 28, no. 1, 133–143, 2003, ISSN: 1434-6052. DOI: 10.1140/epjc/s2003-01152-2. [Online]. Available: http://dx.doi.org/10.1140/epjc/s2003-01152-2.

[165] A. Bredenstein, A. Denner, S. Dittmaier, and M. M. Weber, "Precise predictions for the Higgs-boson decay $H \to WW/ZZ \to 4$ leptons", *Phys. Rev. D*, vol. 74, p. 013 004, 2006. DOI: 10.1103/PhysRevD.74.013004. arXiv: hep-ph/0604011.

[166] P. Baldi, K. Cranmer, T. Faucett, P. Sadowski, and D. Whiteson, "Parameterized neural networks for high-energy physics", *Eur. Phys. J. C*, vol. 76, no. 5, p. 235, 2016. DOI: 10.1140/epjc/s10052-016-4099-4. arXiv: 1601.07913 [hep-ex].

[167] A. Banfi, G. P. Salam, and G. Zanderighi, "Phenomenology of event shapes at hadron colliders", *JHEP*, vol. 06, p. 038, 2010. DOI: 10.1007/JHEP06(2010)038. arXiv: 1001.4082 [hep-ph].

[168] A. Banfi, G. P. Salam, and G. Zanderighi, "Resummed event shapes at hadron - hadron colliders", *JHEP*, vol. 08, p. 062, 2004. DOI: 10.1088/1126-6708/2004/08/062. arXiv: hep-ph/0407287.

[169] A. Heister *et al.*, "Studies of QCD at e+ e- centre-of-mass energies between 91-GeV and 209-GeV", *Eur. Phys. J. C*, vol. 35, pp. 457–486, 2004. DOI: 10.1140/epjc/s2004-01891-4.

[170] G. Abbiendi *et al.*, "Measurement of event shape distributions and moments in $e^+e^- \to$ hadrons at 91 GeV - 209 GeV and a determination of α_s", *Eur. Phys. J. C*, vol. 40, pp. 287–316, 2005. DOI: 10.1140/epjc/s2005-02120-6. arXiv: hep-ex/0503051.

[171] S. Chekanov *et al.*, "Event shapes in deep inelastic scattering at HERA", *Nucl. Phys. B*, vol. 767, pp. 1–28, 2007. DOI: 10.1016/j.nuclphysb.2006.05.016. arXiv: hep-ex/0604032.

[172] G. Dissertori, A. Gehrmann-De Ridder, T. Gehrmann, E. W. N. Glover, G. Heinrich, G. Luisoni, and H. Stenzel, "Determination of the strong coupling constant using matched NNLO+NLLA predictions for hadronic event shapes in e+e- annihilations", *JHEP*, vol. 08, p. 036, 2009. DOI: 10.1088/1126-6708/2009/08/036. arXiv: 0906.3436 [hep-ph].

[173] S. Kluth, P. A. Movilla Fernandez, S. Bethke, C. Pahl, and P. Pfeifenschneider, "A Measurement of the QCD color factors using event shape distributions at $\sqrt{s} = 14$ GeV to 189 GeV", *Eur. Phys. J. C*, vol. 21, pp. 199–210, 2001. DOI: 10.1007/s100520100742. arXiv: hep-ex/0012044.

[174] G. Marchesini, B. R. Webber, G. Abbiendi, I. G. Knowles, M. H. Seymour, and L. Stanco, "HERWIG: A Monte Carlo event generator for simulating hadron emission reactions with interfering gluons. Version 5.1 - April 1991", *Comput. Phys. Commun.*, vol. 67, pp. 465–508, 1992. DOI: 10.1016/0010-4655(92)90055-4.

[175] CMS Collaboration, "Event shape variables measured using multijet final states in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP*, vol. 12, p. 117, 2018. DOI: 10.1007/JHEP12(2018)117. arXiv: 1811.00588 [hep-ex].

[176] ATLAS Collaboration, "Measurement of hadronic event shapes in high-p_T multijet final states at $\sqrt{s}$ = 13 TeV with the ATLAS detector", *JHEP*, vol. 01, p. 188, 2021. DOI: 10.1007/JHEP01(2021)188. arXiv: 2007.12600 [hep-ex].

[177] P. Konar and P. Roy, "Event shape discrimination of supersymmetry from large extra dimensions at a linear collider", *Phys. Lett. B*, vol. 634, pp. 295–301, 2006. DOI: 10.1016/j.physletb.2006.01.056. arXiv: hep-ph/0509161.

[178] J. F. Donoghue, F. E. Low, and S.-Y. Pi, "Tensor Analysis of Hadronic Jets in Quantum Chromodynamics", *Phys. Rev. D*, vol. 20, p. 2759, 1979. DOI: 10.1103/PhysRevD.20.2759.

[179] F. Chollet *et al.*, "Keras: Deep learning library for theano and tensorflow", *URL: https://keras. io/k*, vol. 7, no. 8, T1, 2015.

[180] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, *Tensorflow: Large-scale machine learning on heterogeneous distributed systems*, 2016. arXiv: 1603.04467 [cs.DC]. [Online]. Available: https://www.tensorflow.org/.

[181] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift", in *International conference on machine learning*, PMLR, 2015, pp. 448–456.

[182] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, 2014. arXiv: 1412.6980 [cs.LG].

[183] "CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV", 2018.

[184] K. Cheung, A. Jueid, J. Kim, S. Lee, C.-T. Lu, and J. Song, *Comprehensive study of the light charged Higgs boson in the type-I two-Higgs-doublet model*, 2022. arXiv: 2201.06890 [hep-ph].

[185] W.-S. Hou, R. Jain, and T. Modak, *Searching for Charged Higgs Bosons via $e^+e^- \to H^+H^- \to c\bar{b}\bar{c}b$ at Linear Colliders*, 2021. arXiv: 2111.06523 [hep-ph].

[186] A. Arhrib, R. Benbrik, M. Krab, B. Manaut, S. Moretti, Y. Wang, and Q.-S. Yan, "New Light $H^\pm$ Discovery Channels at the LHC", *Symmetry*, vol. 13, no. 12, p. 2319, 2021. DOI: 10.3390/sym13122319. arXiv: 2110.04823 [hep-ph].

[187] Y. Wang, A. Arhrib, R. Benbrik, M. Krab, B. Manaut, S. Moretti, and Q.-S. Yan, *Searching for $H^\pm \to W^\pm + 4\gamma$ signals in the 2HDM Type-I at the LHC*, 2021. arXiv: 2111.12286 [hep-ph].