

Search for New Physics in All-Hadronic Final States with the CMS Experiment

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Giorgia Rauco

aus

Italien

Promotionskommission

Prof. Dr. Florencia Canelli (Vorsitz und Leitung der Dissertation)

Dr. Annapaola de Cosa

Prof. Dr. Tobias Golling

Prof. Dr. Gino Isidori

Zürich, 2020



A mia nonna Gabriella

Abstract

The most accredited description of our universe is given by the Standard Model (SM) theory, which explains a wide variety of natural phenomena and whose correctness was confirmed by several experimental observations. Despite the huge success of the SM of particle physics, there are some unsolved phenomena, such as the *hierarchy problem* and the nature of Dark Matter (DM). The stabilization of the Higgs boson mass, whose value differs from its radiative corrections, appears to be *unnatural* and suggests that underlying theories may exist. Also, even if many astrophysical observations confirmed the existence of DM, information about its nature or non-gravitational interactions is not yet available. The Large Hadron Collider (LHC) at the European Organization for Nuclear Research (CERN) has been built to seek answers to these questions and explain the shortcomings of the SM.

Two searches for beyond the Standard Model (BSM) physics are performed using proton-proton collision data recorded by the CMS detector at a center-of-mass energy of 13 TeV. Both searches target fully hadronic final states events, which are particularly challenging because of large background contributions from multijet events from Quantum Chromodynamics (QCD) processes.

The first search presented aims at seeking new heavy fermionic resonances called *vector-like quarks* (VLQs) which if exist, could solve the hierarchy problem. In particular, I focus on the search for a single produced VLQ B decaying in a bottom quark and a Higgs boson, that sub-sequentially decays to a b pair. This analysis uses 35.9 fb^{-1} of integrated luminosity collected in 2016 by the CMS experiment. The fully hadronic final state and the large contamination from multijet background processes make this signature particularly challenging. The results of this search show that this is a promising signature that could play a crucial role in future searches for new physics. Moreover, this search is the first exploration at CMS of non-narrow resonance widths of VLQs. No significant deviation from the SM expectation is observed and 95% CL upper limits on the VLQ B production cross-section times the branching ratio are set in both the narrow width and the finite width regimes. Expected and observed limits at 95% confidence level vary from 1.20 to 0.07 pb and from 1.28 to 0.07 pb, respectively, for VLQ B quark masses in the mass range considered, which extends from 700 to 1800 GeV.

The second part of this thesis focuses on the search for DM in the context of Hidden Valley models. These theories predict the existence of dark quarks, that manifest as *semi-visible jets* in the CMS detector. In particular, the model investigated in this thesis predicts the existence of a leptophobic heavy resonance, a Z' gauge boson, connecting the SM and DM sectors, and giving rise to two dark quarks that shower and hadronize leading to the

formation of two semi-visible jets. This search uses 137 fb^{-1} of data collected in 2016, 2017 and 2018 by the CMS experiment. It is the first analysis investigating the semi-visible jets hypothesis at particle colliders. It plays a pivotal role in the context of DM searches and in general in the query of new physics hiding in hadronic final states with missing momentum signatures. No excess of events over the expected SM background is observed, therefore upper limits are placed on the production cross-section as a function of the mediator mass. The specific Hidden Valley model probed in this thesis work is excluded for mediator bosons with mass up to 3.9 TeV.

Contents

Introduction	11
1 The standard model and beyond	15
1.1 The Standard Model and its limitations	16
1.1.1 Limitations	19
1.2 Vector-like quarks as a solution to the hierarchy problem	20
1.2.1 Multiplets and mixing	21
1.2.2 Production and decay	22
1.2.3 Vector-like quarks searches at the LHC	25
1.3 The hunt for Dark Matter	25
1.3.1 Dark Matter candidates	27
1.3.2 Dark Matter searches	28
1.3.3 The Hidden Valley models	29
1.3.4 Dark Matter searches at the LHC	31
2 The Large Hadron Collider and the CMS experiment	33
2.1 The Large Hadron Collider	33
2.1.1 Proton-proton collisions	36
2.2 The CMS experiment	38
2.2.1 Magnet	39
2.2.2 Inner tracking system	39
2.2.3 Calorimeters	42
2.2.4 Muon system	45
2.2.5 Trigger and data acquisition systems	46
2.2.6 Computing	47
3 Particle reconstruction and identification	49
3.1 Particle Flow algorithm	49
3.2 Primary vertex	52
3.3 Muons	53
3.4 Electrons	55
3.5 Jets	55
3.5.1 Jet reconstruction	55
3.5.2 Jet identification	57

3.5.3	Jet energy calibration	58
3.6	Boosted topologies and jet substructure	59
3.6.1	Jet grooming strategies	60
3.6.2	N-subjetiness	62
3.7	Jets from b quarks	64
3.8	Missing transverse energy	65
3.8.1	Performance of the PF p_T^{miss} reconstruction algorithm	66
4	Search for single production of vector-like quarks	69
4.1	Data, modeling and simulation	70
4.1.1	Data	70
4.1.2	Vector-like quark B simulation	70
4.1.3	Simulation of Standard Model backgrounds	72
4.2	Analysis strategy	76
4.2.1	Trigger	76
4.2.2	Particle candidates and event selection	77
4.3	Background estimation	87
4.3.1	The ABCD method	87
4.3.2	Closure tests	92
4.3.3	Results	98
4.4	Systematic uncertainties	100
4.5	Results and interpretation	107
4.5.1	Interpretation model	107
4.5.2	Signal extraction	116
4.5.3	Results	119
5	Search for Dark Matter with semi-visible jets	121
5.1	Datasets and simulated events	122
5.1.1	Datasets	122
5.1.2	Signal simulation	122
5.1.3	Background simulation	125
5.2	Analysis strategy	129
5.2.1	Trigger	129
5.2.2	Event selection	132
5.3	A dedicated SVJ tagger	136
5.3.1	Observables	139
5.3.2	Training	143
5.3.3	Results	143
5.4	Background estimation	144
5.4.1	Signal regions	145
5.4.2	Fitting strategy	150
5.5	SVJ tagger working point optimization	151
5.6	Systematics uncertainties	155

5.7 Results and interpretation	155
6 Conclusions	159
 Appendices	 161
A Pixel detector resolution	163
A.1 The triplet method	163
A.1.1 Triplet combinations	164
A.1.2 Coordinate system	164
A.2 Pixel clusters	165
A.3 Pixel hit reconstruction	166
A.3.1 Generic algorithm	166
A.3.2 Template algorithm	167
A.4 Results	167
A.4.1 Results in 2015	167
A.4.2 Results in 2017	169
A.4.3 Results in 2018	169
 B The quark-gluon jet tagger	 173
B.1 Observables	173
B.2 The tagger	174
B.3 Validation on data	177
B.3.1 Validation on Z+jets events	177
B.3.2 Validation on dijet events	177
B.4 Systematic uncertainties extraction	179
B.5 Comparison with Herwig MC generator	179
 C Statistical approach	 181
C.1 Fundamental concepts	181
C.2 Statistical tests	182
 List of Figures	 183
 List of Tables	 197
 Bibliography	 201
 Acknowledgments	 215

Introduction

A global theory, with the goal of describing all the natural phenomena with a finite set of basic laws, has always been one of the main goals of science. The Standard Model (SM) [1–3] of particle physics was developed in the 1960’s, and still today represents the most complete and consistent description of the quantum universe. It explains a wide variety of natural phenomena and successfully predicts with high precision most of the experimental observations.

The Higgs boson was the last missing piece of the SM: it is the particle associated with the Higgs field that generates the mass of all elementary particles. The theoretical formulation predicting the existence of this particle is the Brout-Englert-Higgs mechanism [4–7], that, in the SM, is responsible for the electroweak symmetry breaking and for the electroweak gauge bosons acquiring mass. After decades of experimental searches, on the 4th July 2012, the ATLAS and CMS experiments at the CERN’s Large Hadron Collider (LHC) announced the observation of a new particle in the mass region around 125 GeV, consistent with a SM Higgs boson [8, 9].

This discovery represents not only a great success of the SM but also a great achievement for the LHC. The LHC [10] is a particle accelerator designed to collide proton beams at a center-of-mass energy of 14 TeV and unprecedented integrated luminosity of $10^{34} \text{ cm}^{-2}\text{s}^{-1}$. Since 2010 it has provided proton-proton collisions at center-of-mass energies of 7, 8 and 13 TeV. The prime motivations of the LHC construction were the search for the Higgs boson and the study of its properties, and the search for physics beyond the SM (BSM). The Compact Muon Solenoid (CMS) experiment [11] is one of the two multipurpose detectors operating at the LHC. It operates a 3.8 T superconducting solenoid, and includes dedicated subsystems for particle tracking, measurements of electromagnetic and hadronic energy deposits, and muon tracking outside the solenoid. The tracking system of CMS, composed of strips and pixel detectors, efficiently identifies charged particles trajectories and measures their momentum. During my PhD, I monitored the pixel detector performance, focusing on the measurement of its spatial resolution.

Despite the huge success of the SM, our current ability to understand the Nature has still its limitations and there are open questions that don’t have an answer yet. One of the main concerns, which gained even more relevance after the experimental confirmation of the Higgs boson existence, is the stabilization of the Higgs boson mass itself. Due to couplings with higher mass particles, like the top quark, radiative corrections to the Higgs boson mass lead to divergencies that would force the Higgs mass to be much larger than the measured mass. To achieve a physical Higgs boson mass of $\mathcal{O}(10^2) \text{ GeV}$, the theory needs precise

cancellations between the radiative corrections and the bare mass, which forces the bare mass to be huge. This undesirable fine-tuning of fixing the bare parameter to a precise and arbitrary value violates the principle of naturalness. This unnaturalness is commonly known as *hierarchy problem* [12] and suggests the presence of underlying BSM theories that could explain this fine-tuning by compensating the large corrections to the Higgs mass through new symmetries and particles.

Several theoretical models introducing new particles allowing the cancellation of loop corrections to the mass of the Higgs boson have been developed. These extensions of the SM, such as little Higgs models [13–15], composite Higgs models [16–20], and models with extra dimensions [21, 22], predict the presence of heavy fermionic resonances called *vector-like quarks* (VLQs) [23–26]. These new hypothetical particles have their left- and right-handed components transforming in the same way under the SM gauge symmetry group. Their vector-like nature allows them to have a fundamental mass, as opposed to the chiral fermions. Therefore, VLQs represent one of the simplest and most promising extension of the SM and solution to the hierarchy problem.

The focus of the first part of this thesis is the search for the electroweak production of a VLQ B, decaying to a bottom (b) quark and a Higgs boson. The analysis uses data collected by CMS in 2016 at a center-of-mass energy of 13 TeV, corresponding to an integrated luminosity of 35.9 fb^{-1} . The final state of this process involves a Higgs boson decaying to a pair of bottom quark, and a bottom and a light flavor quarks arising from the electroweak production process. The VLQ B is expected from the theory to be very massive ($\mathcal{O}(\text{TeV})$), and thus its decay products are boosted. Therefore the Higgs, along with the two b quarks arising from its decay, appear in the CMS detector as a unique large-cone jet. Dedicated algorithms based on the jet substructure have been developed to identify a boosted Higgs boson. Another feature of the signal of interest is the presence of the light flavor quark, that is emitted in the forward pseudorapidity region of the detector. This peculiar signature plays an important role in the analysis as it enhances the separation between background and signal-like events. Two categories are defined depending on the number of forward jets reconstructed in the event. The resonant character of this analysis is exploited by reconstructing an invariant mass with the Higgs jet candidate and the b jet arising from VLQ B decay. This mass is used to compare data and SM expectations, and for the final statistical analysis.

By far, the dominant background is from jets produced through the strong interaction, referred to as multijet events. Smaller background contributions arise from $t\bar{t}$ and $Z/W + \text{jets}$ processes. To reduce the dependence on the Quantum Chromodynamics (QCD) multijet modeling in simulation, a data-driven technique is used to extrapolate this contribution from data. This technique is used to estimate both the expected yield and the VLQ B mass distribution shape of the background in the signal region using control regions depleted in signal.

In the context of this search, a potential signal would appear as a localized excess of data over the SM background in the spectrum of the invariant mass between the b quark and the Higgs boson. To perform the signal extraction, a binned maximum likelihood fit is evaluated, and a statistical interpretation, including the setting of exclusion limits at

95% confidence level with a profile likelihood ratio as test statistic, is performed. For the first time, dedicated theoretical frameworks are developed to interpret the results assuming several resonance widths. This study was published in 2017 in the Journal of High Energy Physics [27].

In addition to the hierarchy problem, one of the most compelling unanswered questions by the SM, is the nature of Dark Matter (DM) and its interactions. The visible component of the universe accounts for just 5% of the total universe energy content, being the rest composed of dark energy (68%) and DM (27%) [28]. Many cosmological studies and astrophysical measurements [28–30] support the evidence for DM existence. The nature of DM still remains unknown, although there are some hints on its properties (DM does interact gravitationally, but not electromagnetically). To approach this mystery, at the LHC physicists rely on the assumption of DM being neutral and stable, and therefore search strategies looking for signals with missing energy are deployed [31].

Despite the huge experimental effort, DM has not been observed yet. The ATLAS and CMS Collaborations performed several searches looking for signs of DM recoiling against visible particles, probing the hypothesis of DM being made of Weakly Interacting Massive Particles, WIMPs, that are currently the most accredited DM candidates. Unfortunately, none of these searches resulted in an experimental confirmation of WIMP scenarios. Relaxing the assumption of DM being made of WIMPs, new possibilities for DM nature open up. Alternative and more complex models have been developed during the last decade to answer this lack of evidence. The most popular examples are the Hidden Valley models [32–34].

In this thesis, I investigate a class of Hidden Valley models that predicts the existence of a dark sector populated by dark quarks, interacting among themselves via a dark strong force and via a dark weak force with particles from the SM. Dark quarks shower and hadronize in the dark sector, similarly to SM quarks. This hadronization leads to the production of stable and unstable dark hadrons and the subsequent formation of the so-called *semi-visible jets*. In these exotic objects, the unstable dark hadrons can decay to SM particles, which constitute the visible energy fraction of the jet, while stable dark hadrons remain invisible and constitute the invisible energy fraction of the jet.

To maximize the signal separation from the background processes an event selection is optimized. Besides, a dedicated tagger exploiting the different underlying interactions in dark and visible sectors is developed. To build this tool, jet substructure and the event kinematics are exploited, and related observables are combined in a boosted decision tree. An important role is also played by another tagger I developed during my PhD: the Quark-Gluon (QG) tagger [35]. The QG tagger was developed to distinguish between hadronic jets initiated by quarks and those initiated by gluons, exploiting the different characteristics of their hadronic showers.

Unlike WIMPs, whose signature is characterized by missing momentum recoiling against the visible jets, in semi-visible jets the missing momentum is aligned to its visible part, making the separation from multijet background, the dominant source of background, extremely challenging. Top pair production also contributes to the background events along with Z/W + jets events. A data-driven strategy, based on the fit of the resonance transverse mass distribution, is used to estimate the background.

A potential signal would appear as a localized excess of data over the SM background in the spectrum of the resonance transverse mass. No signal is observed and the 95% confidence level exclusion limits are derived using the profile-likelihood test statistics. These results will be submitted in 2020 to a high-impact journal of High-Energy Physics.

The thesis is structured as follows. Chapter 1 of this thesis provides a thorough description of the theoretical framework. First, a brief overview of the SM, including historical pretext, is given. Then the main shortcomings of the SM are discussed. A special focus is reserved to the BSM physics models predicting the existence of VLQs and DM, and the state-of-the-art for these exotics searches is also reported.

In Chapter 2, a description of the experimental apparatus is provided. The chapter begins with a brief introduction to the LHC and a description of the CMS detector follows. My work on measuring and monitoring of the pixel detector resolution is described in detail in Appendix A.

A detailed description of the selection criteria on reconstructed particles is provided in Chapter 3. The particles used in the analyses of interest for this thesis are reconstructed with the Particle-Flow algorithm. This technique aims at reconstructing all stable particles, by combining the information from all sub-detectors. Its performance excels in the identification and reconstruction of jets and missing transverse momentum, which are the main elements used in this thesis. The identification of b jets is performed with a dedicated b-tagging algorithm. During my PhD, I was also one of the main developers of the QG tagger, contributing to its training and being the main responsible for its validation on data. Further details about this tool are reported in Appendix B.

Chapter 4 focuses on the search for the electroweak production of a VLQ B. It begins with a description of the data and simulation samples used. The analysis strategy, including the criteria used to select signal events, their categorization, and the final reconstruction of the VLQ B, are then given. It is followed by a description of the data-driven background estimation method. The extraction of the VLQ B signal is documented and a detailed explanation of the dedicated theoretical framework used to interpret the results is given as well. I was the main contributor to all aspects of the analysis, with support from experienced colleagues, and developed all analysis-specific techniques and measurements.

The search for DM is described in Chapter 5. The experimental collection and selection of the data as well as the production of SM backgrounds are reported. The analysis relies on a dedicated SVJ tagger that distinguishes semi-visible jets from SM jets. The training, validation and working point optimization of this SVJ tagger is documented in this chapter. A description of the data-driven technique used to extrapolate a reliable description of the SM background processes is included. Finally, the fit procedure to extract the signal and the systematic uncertainties considered in the analysis, followed by the final results, are reported.

At last, in Chapter 6, I draw the conclusions of the thesis and summarize the main results I obtained during my PhD.

Chapter 1

The standard model and beyond

“ There are grounds for cautious optimism that we may now be near the end of the search for the ultimate laws of nature. ”

Stephen Hawking, *A Brief History of Time*, 1988

The desire to explain the nature of the universe drives the field of High-Energy Physics. The main goal of particle physics is to understand the nature of the particles that constitute matter, their properties and their interactions. The Standard Model (SM) theory [12, 36–38] describes with astonishing precision the structure of matter in terms of elementary particles and fundamental interactions. This theory is currently considered as the most complete and valid description of our universe. The correctness of its predictions has been confirmed by high-precision measurements carried out over the past decades [39, 40].

The last missing piece of SM was the Higgs (H) boson, predicted by the Brout-Englert-Higgs mechanism [4–6] and its search motivated the construction of a new high-energy collider. The Large Hadron Collider (LHC) was designed to reach energies up to the TeV scale, enabling the exploration of new physics at this energy frontier with the discovery of the Higgs boson being one the main goals of its scientific program. This objective was achieved in 2012 when the ATLAS (A Toroidal LHC ApparatuS) and CMS (Compact Muon Solenoid) Collaborations announced the observation of a particle consistent with the SM Higgs boson [8, 9].

However, there are still open questions not addressed by the SM. These limitations are the main motivations for looking for alternative theories beyond the Standard Model (BSM). This thesis aims to address solutions for some of these shortcomings, investigating the existence of BSM particles.

In Section 1.1 a brief overview of the Standard Model and its main limitations is presented. A deeper review can be found in [37]. Section 1.2 describes the theoretical framework and the phenomenology of the first hypothetical particles examined in this work: vector-like quarks (VLQs). The mystery of Dark Matter (DM) is presented in Section 1.3, where I also describe the theoretical model employed in this study.

1.1 The Standard Model and its limitations

The SM describes the elementary matter components and three of the four known fundamental forces (electromagnetic, weak, and strong) in the universe. The matter is composed of fundamental spin 1/2 particles, called *fermions*. There are six leptons and six quarks, and their antiparticles. Both leptons and quarks are grouped into three weak isospin doublets and organized into three families:

- lepton doublets are composed each of an electric-charged particle interacting both electromagnetically and weakly - the electron e , the muon μ , and the τ - and its chargeless corresponding neutrino - the electron neutrino ν_e , the muon neutrino ν_μ , and the tau neutrino ν_τ - that can only interact weakly;

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix} \quad \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix} \quad \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}$$

- quark doublets are composed each of a particle of charge $+\frac{2}{3}$ - up (u), charm (c), and top (t) - and a particle of charge $-\frac{1}{3}$ - down (d), strange (s), and bottom (b). They can interact via strong, electromagnetic and weak interactions.

$$\begin{pmatrix} u \\ d \end{pmatrix} \quad \begin{pmatrix} c \\ s \end{pmatrix} \quad \begin{pmatrix} t \\ b \end{pmatrix}$$

According to the SM, fundamental forces are described via the exchange of force carriers of spin 1, the *gauge bosons*:

- the photon γ is the mediator of electromagnetic interaction. It does not carry charge and does not couple to neutral particles, like neutrinos.
- $W^\pm$ and Z bosons mediate the weak interactions, but unlike the photon, the weak interaction is of short-range and the mediators are massive.
- Gluons are responsible for the mediation of the strong force and carry strong charge, the color.

A schematic summary of elementary particles constituting matter and force mediators, along with particle properties, is presented in Figure 1.1.

The electromagnetic, weak, and strong interactions are described in the SM framework in terms of gauge theory. Particle interactions emerge from a *gauge principle*, meaning that they arise as a consequence of the invariance of a theory under a local gauge transformation [41].

Electroweak sector Quantum Electrodynamics (QED) describes electromagnetic interactions. It is a gauge transformation of spinor fields and arises from the imposition of gauge invariance under the symmetry $U(1)$. The unification of electromagnetic and weak interactions gives rise to the electroweak (EWK) sector of the SM. It is a gauge theory

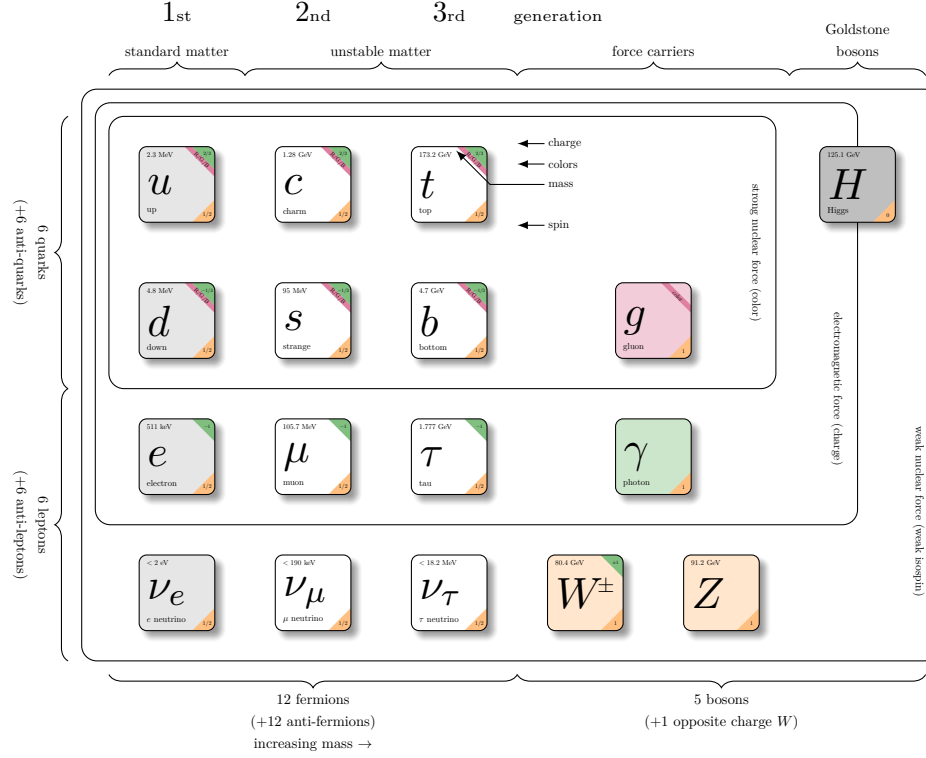


Figure 1.1: Schematic depiction of the SM of particles physics, with the three generations of matter, gauge bosons, and the Higgs boson.

based on $SU(2)_L \times U(1)_Y$, where L indicates the weak isospin, whose current couples only to left-handed fields, and Y labels the weak hypercharge [41]. The generator of the weak $SU(2)_L$ group is the weak isospin T , that is associated with the leptonic fields and that is defined as $T = \frac{1}{2}\boldsymbol{\tau}$, where $\boldsymbol{\tau}$ are the Pauli matrices. In the electromagnetic sector $U(1)_Y$, the hypercharge Y is defined through the Gell-Mann formula [42], expressed as:

$$Y = 2(Q + T_3), \quad (1.1)$$

where Q is the electromagnetic charge and T_3 is the third component of the particle weak isospin.

The unified description of the two forces was formulated by Glashow [43], Salam [44], and Weinberg [3], following the Yang-Mills model [45]. Their framework describes the non-conservation of parity in weak interactions, that was observed experimentally for the first time in the decay of Cobalt-60 atoms [46], by using different representations for the left-handed and the right-handed leptons in the SM Lagrangian. While the left-handed fermion fields are in the doublet representation of $SU(2)$ with weak isospin $T = \frac{1}{2}$, the fermions of right-handed chirality are in the singlet representation with weak isospin $T = 0$. In the weak $SU(2)_L$ group, the left- and right-handed components of the leptonic fields can be described respectively as:

$$T = \frac{1}{2} : \quad \begin{cases} T_3 = +1/2 \\ T_3 = -1/2 \end{cases} \quad L_e = \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L \quad L_\mu = \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L \quad L_\tau = \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L, \quad (1.2)$$

$$T = 0 : \quad e_R \quad \mu_R \quad \tau_R \quad \nu_{e,R} \quad \nu_{\mu,R} \quad \nu_{\tau,R} . \quad (1.3)$$

Analogously, for the quark fields the following formulation for doublets and singlets holds:

$$T = \frac{1}{2} : \quad \begin{cases} T_3 = +1/2 & \begin{pmatrix} u_L \\ d'_L \end{pmatrix} \\ T_3 = -1/2 & \begin{pmatrix} c_L \\ s'_L \end{pmatrix} \end{cases} \quad \begin{pmatrix} t_L \\ b'_L \end{pmatrix} \quad (1.4)$$

$$T = 0 : \quad u_R \quad d_R \quad s_R \quad c_R \quad b_R \quad t_R . \quad (1.5)$$

The d' , s' , b' components are weak eigenstates. Their corresponding mass eigenstates are given by:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{CKM} \begin{pmatrix} d \\ s \\ b \end{pmatrix}, \quad (1.6)$$

where V_{CKM} is the Cabibbo-Kobayashi-Maskawa (CKM) matrix [47, 48]. The CKM matrix shows that transitions between fermions of different generations are strongly suppressed, whereas the diagonal elements corresponding to transitions in the same generation are favored.

For the electroweak sector, the Lagrangian density is expressed as

$$\begin{aligned} \mathcal{L}_{\text{EWK}} = & -\frac{1}{4}F_{\mu\nu}^a F_a^{\mu\nu} - \frac{1}{4}G_{\mu\nu}G^{\mu\nu} \\ & + i\bar{\psi}(\gamma^\mu D_\mu)\psi + h.c. \\ & + \psi_i y_{ij} \psi_j \phi + h.c. \\ & + |D_\mu \phi|^2 - V(\phi), \end{aligned} \quad (1.7)$$

where the first line arranges the kinetic term for the gauge sector of the electroweak theory, with the index $a = 1, 2, 3$ labeling the $\text{SU}(2)_L$ components [49]. Four massless gauge fields are needed, three being the generators of the $\text{SU}(2)_L$ group, denoted as W_μ^1, W_μ^2 and W_μ^3 , and one connected with $\text{U}(1)$, B_μ . The formulation of their field strength tensors follows:

$$\begin{aligned} F_\mu \nu^a &= \partial_\nu W_\mu^a - \partial_\mu W_\nu^a - g\epsilon_{abc}W_\mu^b W_\nu^c \quad \text{for } a = 1, 2, 3 \\ G_\mu \nu &= \partial_\nu B_\mu - \partial_\mu B_\nu, \end{aligned} \quad (1.8)$$

with g being the coupling constant of $\text{SU}(2)_L$. The second term in Equation 1.7 describes the kinematics of matter fields and accounts for their interactions with the gauge fields. The third line refers to the Yukawa sector and it details the connection between the matter and the scalar fields. Finally, the scalar sector is considered in the fourth term, split as kinematics for the scalar field and effective potential. The covariant derivative D_μ is expressed as

$$D_\mu = \partial_\mu + \frac{ig'}{2}B_\mu Y + \frac{ig}{2}\tau \cdot W_\mu \quad (1.9)$$

with g' being the $\text{U}(1)$ coupling constant, Y the hypercharge operator and $\tau \equiv (\tau_1, \tau_2, \tau_3)$ the Pauli matrices.

In their current description, the gauge fields described by the EWK gauge theory are massless, but measurements show that the $W^\pm/Z$ bosons have a finite mass [50–53]. Gauge bosons acquire mass through a spontaneous breaking of the $SU(2)_L \times U(1)_Y$ gauge group named *Higgs mechanism* and predicted by Brout, Englert and Higgs in the 1960's [4–6]. This mechanism requires the existence of a new scalar boson, the Higgs boson: the coupling to the Higgs field generates the vector boson masses.

Quantum Chromodynamics sector Strong force is described by the Quantum Chromodynamics (QCD) and is responsible for the interactions between quarks and gluons. QCD is a non-abelian gauge theory based on the $SU(3)$ symmetry group [54]. The corresponding Lagrangian is:

$$\mathcal{L}_{\text{QCD}} = \bar{\psi}_i(i(\gamma^\mu D_\mu)_{ij})\psi_j - m\delta_{ij}\psi_j - \frac{1}{4}G_{\mu\nu}^a G_a^{\mu\nu}, \quad (1.10)$$

being ψ_i the quark field with i color index. The covariant derivative D_μ is given by:

$$D_\mu = \partial_\mu + ig\frac{\lambda_a}{2}\mathcal{A}_\mu^a, \quad (1.11)$$

where λ_a with $a = 1, \dots, 8$ are the Gell-Mann matrices and $\mathcal{A}_\mu^a$ are the gluon gauge fields. The gauge-invariant gluon field strength tensor $G_{\mu\nu}^a$ is given by

$$G_{\mu\nu}^a = \partial_\mu\mathcal{A}_\nu^a - \partial_\nu\mathcal{A}_\mu^a + gf^{abc}\mathcal{A}_\nu^b\mathcal{A}_\mu^c, \quad (1.12)$$

with f^{abc} the structure constants of $SU(3)$.

A strong coupling constant α_S is introduced to determine the relative strength of interaction between particles. For very small distances and high values of momentum Q^2 exchanged in the interaction, α_S vanishes asymptotically. The strong running coupling constant can be expressed as:

$$\alpha_S(Q) = \frac{12\pi}{(33 - 2n_f) \log\left(\frac{Q^2}{\Lambda_{\text{QCD}}^2}\right)}, \quad (1.13)$$

where n_f is the number of quark flavors (u, d, c, s, b, t) at the energy of the probe used to measure the coupling [55]. For energies between 100 - 1000 GeV the typical value of $\alpha_S(Q)$ is around 0.1. Here we see that $\alpha_S(Q)$ decreases with increasing Q^2 . In the limit of very large Q^2 , quarks behave as free. This phenomenon is called *asymptotic freedom* [12]. On the other hand, the running behavior of α_S leads to the *confinement* of the color charge. At large distances, the inter-quark coupling increases and becomes more energetically favorable to create a new quark-antiquark pair rather than further increasing the distance between the two partons [12]. This mechanism is responsible for the *hadronization* process of high energy colored particles produced in particle collisions, which evolve by gradually losing energy from a high-energy regime to low-energy processes that produce color screening and the formation of composite states of quarks, named *hadrons*.

1.1.1 Limitations

The SM accounts for essentially all the experimental data and results to date. However, this theory is incomplete and fails to explain some observed phenomena. An example of this

failure is the massive nature of neutrinos. The SM predicts the neutrinos to be massless, however, observations from solar and atmospheric experiments show that neutrinos with a specific lepton flavor can oscillate to a different flavor [56–58], phenomenon that requires neutrinos to be massive. Another unsatisfactory aspect of the SM is its arbitrariness: it requires 19 numerical constants whose values are unrelated and that need to be determined *ad-hoc* from experimental measurements [59]. Fermion masses span over 11 order of magnitude and this hierarchy is not understood. Furthermore, the gravitational force is not explained by the SM and many attempts have been made to incorporate gravity in a unified symmetry, the so-called *Grand Unified Theory* (GUT) [12, 60], without success.

At the Planck scale, 10^{19} GeV, other difficulties start to arise for the SM. Due to couplings with higher mass particles, like the top quark, radiative corrections to the Higgs boson mass lead to divergencies that would force the Higgs mass to be much larger than the measured mass. The problem of fine-tuning can be parameterized as:

$$m_{\text{physical}}^2 \simeq m_{\text{bare}}^2 + \Delta m^2, \quad (1.14)$$

where Λ is the energy scale at which one expects to start finding new physics beyond the SM (i.e. the Planck scale), Δm is a radiative corrective parameter on these energy scales and m_{physical}^2 is the bare mass plus the radiative corrections. To achieve a physical Higgs boson mass of $\mathcal{O}(10^2)$ GeV, assuming the cut-off scale at the Planck scale $\mathcal{O}(10^{19})$ GeV, the theory needs precise cancellations between the radiative corrections and the bare mass, which forces the bare mass to be huge. This undesirable fine-tuning of fixing the bare parameter to a precise and arbitrary value is found to be unnatural and is commonly known as the *hierarchy problem* [12]. This phenomenon suggests the presence of underlying BSM theories that could explain this fine-tuning by compensating the large corrections to the Higgs mass through new symmetries and particles.

Finally, from cosmological measurements and astrophysical observations, ordinary matter accounts only for 5% of the total universe energy content, being the rest made of dark energy (68%) and DM (27%) [28]. Several cosmological and astrophysical measurements [28–30, 61] have been carried out to verify the existence of DM. The primary evidence for DM comes from studies of the rotations of galaxies, the gravitational lensing effect, and the Cosmic Microwave Background. Additional experiments are being pursued to detect the nature of DM and several models have been tested.

In this thesis, I seek a better understanding of the hierarchy problem and the nature of DM. Possible solutions predicting the existence of new physics appearing at the $\mathcal{O}(\text{TeV})$ scale have been investigated and are presented in the following sections. The LHC allows to access energy frontiers never reached before in a laboratory, enabling to probe, for the first time, the possible extensions of the SM I focus on in this thesis. If such theories are confirmed, we could fashion a new and more complete description of our Universe.

1.2 Vector-like quarks as a solution to the hierarchy problem

As introduced in Section 1.1.1, many BSM theories are driven by the need to achieve a better insight into the hierarchy problem. They describe new mechanisms at the electroweak scale

including new particles and symmetries to compensate for the Higgs boson mass. The most popular BSM models are SuperSYmmetry (SUSY), Little Higgs [62–64], Composite Higgs [16, 65] and extra dimensions models [66]. SUSY proposes the existence of super-partners of the SM particles, that differ in spin by a factor 1/2, but have equal other quantum numbers. It is a theoretically elegant theory, providing hypothetical solutions to many of the shortcomings of the SM. Concerning the hierarchy problem, the super-partners’s divergent contributions would cancel the SM particle contributions, explaining the fine-tuning mystery. The alternative extensions addressing the hierarchy problem predict the existence of new heavy fermionic resonances called *vector-like quarks* [26]. They transform as triplets under the color gauge group and their left- and right-handed chiralities have the same color and electroweak quantum numbers. This feature enables them to receive mass without the Yukawa couplings to a Higgs doublet, unlike SM chiral fermions, whose existence has been excluded by electroweak measurements [67] and by the discovery of the Higgs boson [68]. A figurative depiction of the solution to the hierarchy problem considering the VLQs formulation is shown in Figure 1.2.

$$\Delta m_H^2 = \text{---} \text{H} \text{---} \text{---} \text{---} + \text{---} \text{H} \text{---} \text{---} \text{---} + \dots$$

Figure 1.2: The hierarchy problem in terms of the Higgs mass squared corrections coming from the inclusion of the VLQs.

1.2.1 Multiplets and mixing

Despite the wide number of theories predicting new physics with VLQs, these particles can be approached in a model-independent way in terms of few free parameters, assuming that:

- the scalar sector only includes $SU(2)_L$;
- VLQs can only couple to third-generation quarks, motivated by the fact that the mixing is of the order of m/M , with m and M being the masses of the SM quarks and the vector-like quarks.

Under these assumptions, being the scalar sector just composed of $SU(2)_L$, VLQs can only appear in seven gauge-covariant multiplets with fixed $SU(3)_C \times SU(2)_L \times U(1)_Y$ quantum numbers [26]:

$$\begin{array}{ll} T_{L,R}^0, & B_{L,R}^0 & (\text{singlets}), \\ (X T^0)_{L,R}, & (T^0 B^0)_{L,R}, & (B^0 Y)_{L,R} & (\text{doublets}), \\ (X T^0 B^0)_{L,R}, & (T^0 B^0 Y)_{L,R} & (\text{triplets}), \end{array} \quad (1.15)$$

where B^0 and T^0 are the new fields with charge of $-1/3$ and $2/3$. Some of the multiplets also include new particle fields X and Y , with electric charge of $5/3$ and $-4/3$, respectively, and whose weak and mass eigenstates coincide. The coupling to left- and right-handed currents is labeled with the subscripts L and R .

When new fields $T_{L,R}^0$ are added to the SM, the resulting physical up-type quark mass eigenstates u, c, t, T may contain non-zero $T_{L,R}^0$ components. This could lead for example to a deviation in their couplings to the Z boson. In the up-quark sector, constraints on these deviations for u and c quarks are derived from electroweak precision tests and flavor-changing neutral measurements at low energies [40], and resulted to be much stronger than for top quark. Hence, it is very reasonable to assume that only the top quark has sizable $T_{L,R}^0$ components (or, in other words, only the top quark "mixes" with T). The dependence between vector-like and third-generation quarks is parametrized through the unitary matrix $U_{L,R}^u$,

$$\begin{pmatrix} t_{L,R} \\ T_{L,R} \end{pmatrix} = U_{L,R}^u \begin{pmatrix} t_{L,R}^0 \\ T_{L,R}^0 \end{pmatrix} = \begin{pmatrix} \cos \theta_{L,R}^u & -\sin \theta_{L,R}^u e^{i\phi_u} \\ \sin \theta_{L,R}^u e^{-i\phi_u} & \cos \theta_{L,R}^u \end{pmatrix} \begin{pmatrix} t_{L,R}^0 \\ T_{L,R}^0 \end{pmatrix} \quad (1.16)$$

with the subscripts u and d being the up- or down-type mixings and where ϕ indicates a CP-violating phase.

Unlike the up sector, in the down-like formulation the experimental measurements set stronger constraints for the new vector-like fields with the b quarks than for the lighter states d and s . However, being the mixing proportional to the mass of the SM quark, one still expects dominant mixing with the b quark [69]. Under these considerations, the mixing between the b - B states is described by the $U_{L,R}^d$ matrix,

$$\begin{pmatrix} b_{L,R} \\ B_{L,R} \end{pmatrix} = U_{L,R}^d \begin{pmatrix} b_{L,R}^0 \\ B_{L,R}^0 \end{pmatrix} = \begin{pmatrix} \cos \theta_{L,R}^d & -\sin \theta_{L,R}^d e^{i\phi_d} \\ \sin \theta_{L,R}^d e^{-i\phi_d} & \cos \theta_{L,R}^d \end{pmatrix} \begin{pmatrix} b_{L,R}^0 \\ B_{L,R}^0 \end{pmatrix}. \quad (1.17)$$

The mixing between the third generation SM quarks and the VLQs would lead to modifications of the $Zb\bar{b}$ coupling. Important constraints are imposed with the experimental study of this coupling [70] and through other SM precision measurements [71].

The VLQs also contribute to the Higgs boson production and decays, as they enter the loop diagrams in the $gg \rightarrow H$ ($H \rightarrow gg$) and $H \rightarrow \gamma\gamma$ processes. Nevertheless, in the minimal extension of VLQs formulation, these effects give only small contributions. The modifications on the cross-sections and branching ratios are below the precision of current measurements and therefore invisible at the LHC [26].

1.2.2 Production and decay

At a hadronic collider such as the LHC, VLQs can be either produced in pairs, via the QCD process

$$gg, q\bar{q} \rightarrow Q\bar{Q} \quad (Q = T, B, X, Y), \quad (1.18)$$

or singly, together with SM quarks through electroweak interaction, giving rise to

$$gq \rightarrow T\bar{b}q', \quad gq \rightarrow T\bar{t}q', \quad (1.19)$$

$$gq \rightarrow B\bar{b}q', \quad gq \rightarrow X\bar{t}q', \quad (1.20)$$

$$gq \rightarrow Y\bar{b}q'. \quad (1.21)$$

In this thesis, I focus on the production of a vector-like B quark. An example of the Feynman diagrams describing the processes of single and pair production of the B quark is shown in Figure 1.3. Similar interactions also hold for up-type VLQs.

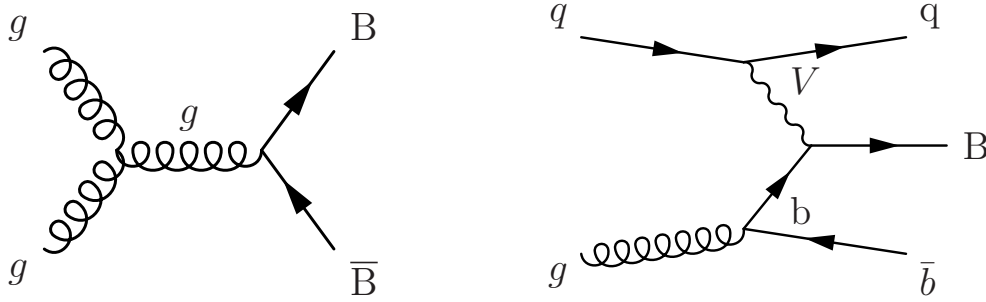


Figure 1.3: Example of Feynman diagrams for the pair (left) and single (right) vector-like B quark.

Since the pair production is produced via strong interaction, the corresponding cross-section depends only on the mass of the new resonance. On the contrary, the cross-section of the single production process is proportional to the coupling to the W and Z bosons. In particular, the B quark is produced in association with a light flavor quark j and a b quark, and this process is referred to as $B\bar{b}j$. In Figure 1.4 the cross-sections of single and pair production process, for a center-of-mass energy of 13 TeV, are shown. It appears that for resonances with large mass, the single production process is enhanced compared to the pair production. In particular, for the $B\bar{b}j$ interaction, it becomes dominant starting from about 850 GeV.

The mixing between the heavy states and their lighter partners t, b induces a splitting of their mass [26]. This effect leads to the hierarchies $m_T \geq m_X$ and $m_B \geq m_Y$, where the vector-like quark T can either be heavier or lighter than the B quark. Also, the mass difference between the heavy members of the multiplets is expected to be small, $\mathcal{O}(1-10)$ GeV [26]. Hence, the decays from one to the other are suppressed. Therefore, the VLQs can only decay to top or bottom quarks, and a W, Z or Higgs boson. Due to their exotic electric charges, X and Y can only give rise to a W boson and a top or bottom quark, respectively. The other heavy quarks T and B, with charges $2/3$ and $-1/3$, can decay as:

$$\begin{aligned} T &\rightarrow W^+b, \quad T \rightarrow Zt, \quad T \rightarrow Ht, \\ B &\rightarrow W^-t, \quad B \rightarrow Zb, \quad B \rightarrow Hb. \end{aligned} \quad (1.22)$$

Figure 1.5 summarizes the branching ratios (BRs) of the VLQs decay modes, for the singlet and doublet models and as a function of the heavy quark mass.

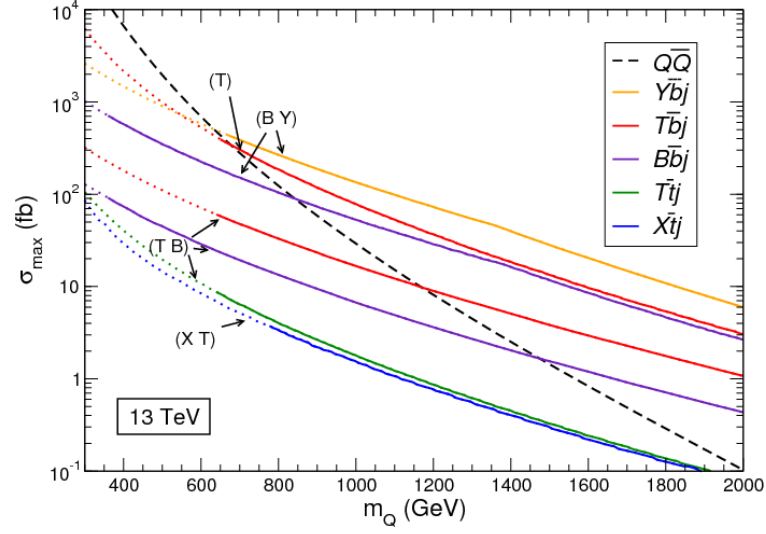


Figure 1.4: Production cross-sections at the LHC with 13 TeV, for selected quark multiplets and for both pair (dashed black line) and single (colored lines) processes [26]. The dotted part of the lines indicates the range of masses already excluded by direct searches.

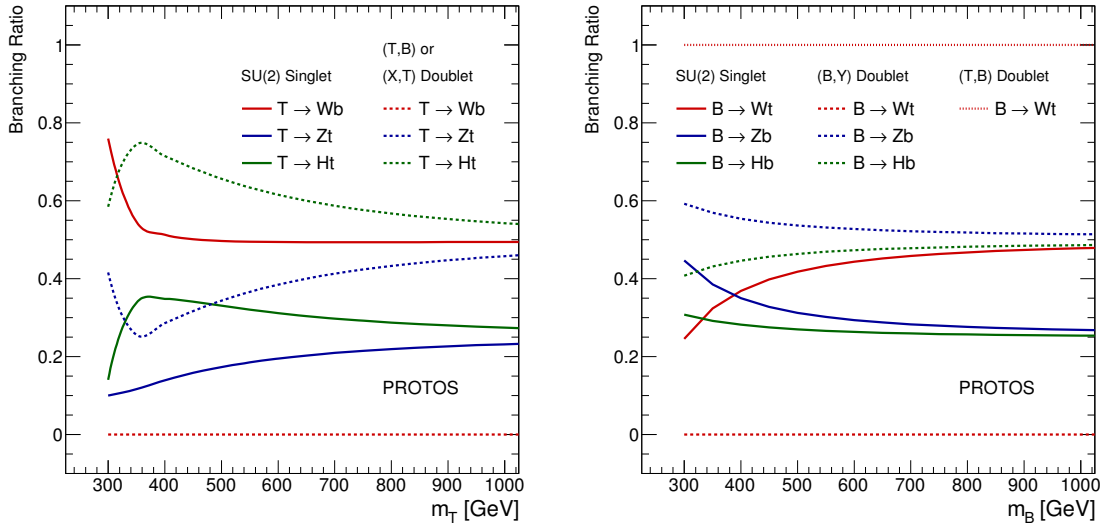


Figure 1.5: Branching ratios for the different decay modes as a function of VLQ mass in the case of a vector-like T quark (left) and a vector-like B quark (right) [72]. In both cases the branching ratios are provided for an SU(2) singlet and two different SU(2) doublet scenarios.

1.2.3 Vector-like quarks searches at the LHC

The search of these new heavy fermionic resonances at the LHC is a very important goal of its physics program. ATLAS and CMS experiments have performed extensive studies in the context of VLQs [73, 74]. Previous results were published by the D0 [75, 76] and CDF [77–82] Collaborations.

The landscape of VLQs searches is very wide. Many signatures have been explored and they are summarized in Table 1.1. No discrepancies between the SM predictions and the data were found and lower bounds at the 95% CL on the VLQ mass are set. More details can be found in the references provided for each entry.

$\mathcal{L}$ [fb^{-1}]	Channel	Obs. mass limit (95% CL) [GeV]	Ref.
ATLAS			
79.8	$pp \rightarrow qq' + B \rightarrow qq' + b\gamma\gamma$	$m_B^{\text{doublet}} > 1210$	[83]
36.1	$pp \rightarrow TT$ (BB), hadronic	$m_{B,T} > 1100$	[84]
36.1	$pp \rightarrow TT$ (BB), with leptonic Z	$m_T^{\text{singlet, doublet}} > 1030, 1210$ (1010, 1140)	[85]
36.1	$pp \rightarrow BB \rightarrow WtWt$, single lepton	$m_B > 1350$	[86]
36.1	$pp \rightarrow TT$, single lepton	$m_T > 0.99 - 1.43$	[87]
36.1	$pp \rightarrow TT \rightarrow WbWb$, single lepton	$m_T > 1350$	[88]
36.1	$pp \rightarrow TT \rightarrow Z(\nu\nu)t + X$, single lepton	$m_T > 1160$	[89]
3.2	$pp \rightarrow qq' + T$ (Y) $\rightarrow qq' + Wb$, leptonic	$m_{T,Y} > 1440$	[90]
CMS			
35.9	$pp \rightarrow qq' + B/X \rightarrow Wt$, single lepton	$m_{B,X}^{\text{LH, 10\%-20\%-30\%}} > 1490, 1590, 1660$ (920, 1300, 1450)	[91]
35.9	$pp \rightarrow TT$ (BB), leptonic	$m_{T,B} > 1140 - 1300$ (910 - 1240)	[92]
35.9	$pp \rightarrow qq' + B \rightarrow qq' + b\bar{b}b$	$m_{B, \text{doublet}}^{20\%, 30\%} > 850, 1200$	[27]
35.9	$pp \rightarrow TT$ (YY), $T(Y) \rightarrow Wb$, single lepton	$m_{T,Y} > 1295$	[93]
35.9	$pp \rightarrow qq' + T \rightarrow qq' + ee$ ($\mu\mu$) + t_{had}	$m_T^{\text{LH (RH)}} > 1340 - 1420$ (820 - 940)	[94]
2.6	$pp \rightarrow TT$ (BB), single lepton	$m_{T,B}^{\text{singlet, doublet}} > 860, 830$ (730, -)	[95]
2.3	$pp \rightarrow XX \rightarrow WtWt$, leptonic	$m_X^{\text{RH (LH)}} > 1020$ (990)	[96]
2.3	$pp \rightarrow qq' + T/X \rightarrow qq' + Wb$, single lepton	$m_Y > 1400$	[97]
2.3	$pp \rightarrow qq' + T/B \rightarrow qq' + Zt/b$	limits on $\sigma \times \text{BR}(T(B) \rightarrow tZ$ (bZ))	[98]
2.3	$pp \rightarrow qq' + T \rightarrow qq' + tH$, hadronic	limits on $\sigma \times \text{BR}(T \rightarrow tH)$	[99]
2.3	$pp \rightarrow qq' + T \rightarrow qq' + tH$, single lepton	limits on $\sigma \times \text{BR}(T \rightarrow tH)$	[100]

Table 1.1: Summary of the various VLQ analysis and corresponding resulting observed mass limits to date from the ATLAS and CMS experiments [73, 74], with a focus on the latest studies performed at a center-of-mass energy of 13 TeV.

1.3 The hunt for Dark Matter

The first indication of the DM existence comes from studies of the dynamics in gravitational potential. A first measurement was performed in 1933 by Fritz Zwicky [101], who calculated the gravitational mass of the galaxies in the Coma cluster and obtained a value of two orders of magnitude larger compared to the prediction from luminosity considerations. This was only explicable with the presence of additional mass. The existence of DM was confirmed in the 1970's by Vera Rubin and her team [102]. They observed that the orbital velocities of stars and gas located at different distances from the center of the Andromeda galaxy weren't respecting the Newtonian dynamics. They found that the outermost components of the galaxy were moving as quickly as those closer to the center. This is an indication that spiral galaxies are embedded in regions decoupled from cosmic expansion and containing gravitationally bound matter, called DM haloes.

In addition to the study of galaxies dynamic, proofs of the existence of DM were also found from the *gravitational lensing* effect measurements. Given two objects, the emitter and the observer, the matter located between them bends or focus the light. Therefore, a

distorted image will arrive at the observer [103]. By measuring the distortion it is possible to calculate how massive the cluster of galaxies in between is and then compare it with the measured amount of visible matter. An excess of mass has been observed, indicating the existence of invisible matter [104]. In Figure 1.6 the galaxy cluster 1E 0657-56 [105], also known as the *bullet cluster* is shown. Where two clusters merge, the galaxies are minimally affected. They are shown in orange and white as obtained from optical images. On the other hand, the intra-cluster hot gas ram pressure and, as a result, it is distributed toward the center of the system after the collision, as show the pink clumps in the image. Through the gravitational lensing effect the concentration of mass is determined and resulted that most of the mass in the cluster is located in the blue areas. As can be observed, the blue regions are separated from the pink zones. This is an indication of the DM existence. If the cluster was formed by visible only, the highest concentration of mass in the system and the hot gas distribution would have coincided.



Figure 1.6: An image of the galaxy cluster 1E 0657-56, also known as the "bullet cluster", as taken from the Chandra X-Ray observatory [105]. This cluster was formed after the collision of two large clusters of galaxies, the most energetic event known in the universe since the Big Bang.

The study of the Cosmic Microwave Background (CMB) [106, 107] led to the calculation of the amount of DM in the universe. The CMB is a faint cosmic background of electromagnetic radiation filling the space which is a remnant from the recombination epoch in the earliest stages of the universe. The hot ionizing gas composing the universe at that time should be almost uniform. The Planck satellite, a space-based microwave telescope, has drawn a map of the CMB (Figure 1.7) indicating the presence of temperature fluctuations. By studying these anisotropies, astrophysicists estimated cosmological parameters related to the nature of DM, such as its abundance in the universe. It resulted that the total baryonic and DM densities, as measured from the Planck collaboration [28], are $\Omega_{\text{barionic}}h^2 = 0.02226 \pm 0.00023$ and $\Omega_{\text{DM}}h^2 = 0.1186 \pm 0.0020$, where $h = H_0/100$ and H_0 is the Hubble constant. From these results, it can be implied that the universe contains 4.8% of ordinary matter and 25.8% of dark matter, being the remaining 69.4% composed by dark energy [108], an unknown form of energy which is hypothesized to permeate all of space,

tending to accelerate the expansion of the universe [40].

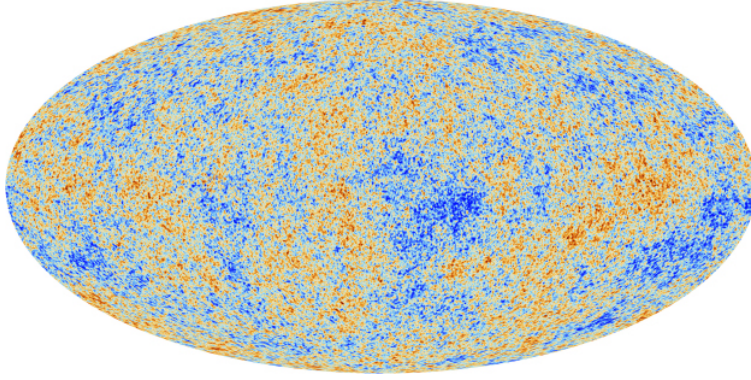


Figure 1.7: The anisotropies of the CMB as observed by the Planck satellite [109].

1.3.1 Dark Matter candidates

One of the first attempts made to identify the DM candidates relied on the search for hidden galactic baryonic matter in the form of MAssive Compact Halo Objects (MACHOs). These objects, including neutron stars, and brown and white dwarfs, are composed of ordinary matter. They emit little or no radiation and can be experimentally detected through the microlensing effect. This method relies on the gravitational force of distant objects to bend and focus light coming from galactic objects. An excess in the number of microlensing events over what is expected from known populations of stars could represent a hint of the existence of DM [110]. Several collaborations worked on the search for these astronomical bodies. The combination of the upper limits on the amount of planetary mass dark matter in the Galactic Halo from the EROS and MACHO Collaborations led to the conclusion that MACHOs cannot contribute more than the 25% of halo DM. Therefore this indicates that MACHOs do not solve the missing mass problem and therefore these analyses strengthen the need for non-baryonic DM.

Many predictions and models have been developed over the past decades focusing on the search for a non-baryonic matter filling a big portion of the universe. Cosmological and astrophysical observations provide important indications [40] that DM candidates should be "cold" (i.e. non-relativistic at the moment of galaxy formation), stable on cosmological time scales, interacting very weakly with electromagnetic radiation and have the correct relic density, that is the particle abundance remaining from the Big Bang. None of the particles predicted by the current SM fulfill these properties, hence candidates arising from new physics extensions are considered. The most popular DM candidates are primordial black holes, axions, sterile neutrinos, Kaluza-Klein states, and Weakly Interacting Massive Particles (WIMPs), being the latter the most widely acquired to date.

WIMPs are hypothetical particles with masses ranging from 10 GeV to a few TeV. Assuming the WIMPs to be produced in thermal and chemical equilibrium with the SM particles after inflation, their present relic density can be computed. It results approximately to:

$$\Omega_\chi h^2 \simeq \frac{0.1 \text{pb} \cdot c}{\langle \sigma_A v \rangle}, \quad (1.23)$$

where c is the speed of light and $\langle\sigma_A v\rangle$ is the thermal averaging between the total annihilation cross-section of a pair of WIMPs into SM particles, σ_A , considering the relative velocity v between the two WIMPs. Obtaining the correct abundance of dark matter today implies $\langle\sigma_A v\rangle \simeq 3 \times 10^{-26} \text{cm}^3 \text{s}^{-1}$, which is approximately what is expected for a new particle in the 100 GeV mass range that interacts via the electroweak force. The fact that the expected value of the relic density arises while setting both the mass and self-annihilation cross-section to the weak scale, is often referred to as the *WIMP miracle* [40].

1.3.2 Dark Matter searches

The nature of dark matter is investigated in three complementary ways: direct, indirect, and collider searches. The direct detection relies on the scattering process of DM with ordinary SM matter. Indirect searches aim at observing stable messenger SM particles arising from dark matter annihilations or decays. Finally, at a particle collider, the conditions of the early hot universe can be reproduced and DM candidates could be produced directly. The different experimental approaches used to investigate the interaction between DM and SM particles are illustrated in Figure 1.8.

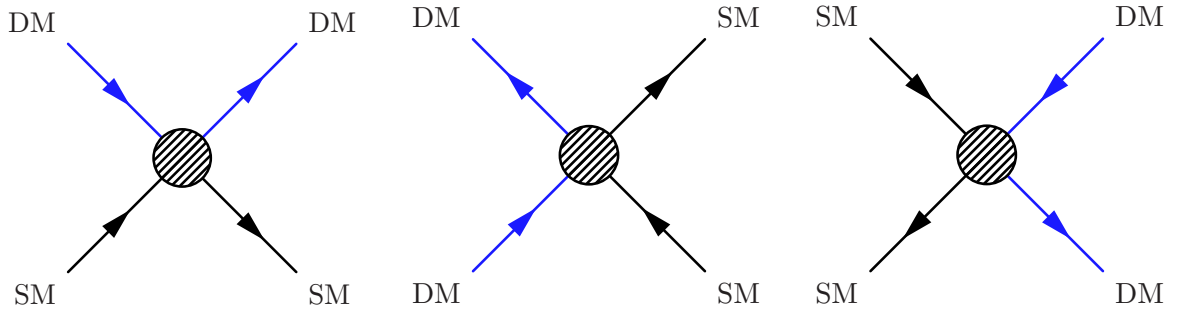


Figure 1.8: Schematic representation of the three different experimental approaches used to search for DM: direct detection investigating for nuclear recoils from DM scattering (left), indirect searches which detect DM annihilation products (center) and collider searches investigating for DM production in high energy collisions (right).

However, despite the huge experimental effort, no evidence for DM has been found. In the last decade, in addition to the WIMP model, alternative theories have been proposed. They describe DM as made of a multitude of particles belonging to a *dark sector* [111]. The dynamics of dark sectors have been explored in many contexts [112] and in these models, the dark sector contains a new spectrum of DM particles and possibly new force carriers and/or matter fields. The connection between the dark sector and the SM sector is ensured by messenger states. Strongly interacting dark sectors arise in a wide variety of new physics scenarios, with the canonical example being the Hidden Valley models [32–34]. Some of these models predict the existence of dark leptons, dark photons or dark quarks. In this thesis, I focus on a class of Hidden Valley models that predict the existence of dark quarks manifesting as semi-visible jets at particle colliders like the LHC.

1.3.3 The Hidden Valley models

In the Hidden Valley models the SM is extended by introducing a non-abelian group G_v [32]. All SM particles are neutral under G_v , but new low mass valley particles charged under G_v and neutral under G_{SM} appear. They are called " v -particles" and assemble themselves into G_v -neutral " v -hadrons"[34]. This model is an example of a *confining Hidden Valley*. The term "Hidden Valley" refers indeed to the presence of the low mass "valley" (v -)states, which can only be observed by passing through or over a barrier separating the hidden sector from the SM sector. A schematic depiction of the Hidden Valley is shown in Figure 1.9

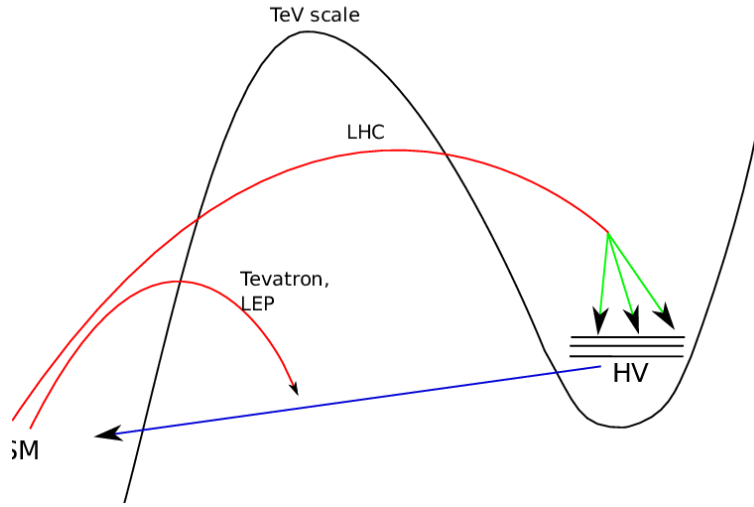


Figure 1.9: Schematic view of the Hidden Valley [113]. The mountain represents massive states which may connect the Standard Model sector to light states in the valley sector.

The effects of a Hidden Valley on the SM observable sector are determined by the nature of the mediators. Given the presumably large gap between the mass of v -sector particles and the heavy mediators, there is effectively a barrier between the SM and Hidden Valley which must be surmounted to produce the valley particles via the SM processes. Examples of possible mediators are Z' bosons, messenger fermions which connect the gluons from the hidden and Standard Model sectors, or gravitons from Randall-Sundrum or large extra dimensions mediators.

Once the v -sector quarks are produced through the mediator, they confine into v -hadrons. If the v -hadrons are much lighter than the mediator, many of them are produced in the v -sector hadronization. Some of the v -hadrons can then decay to combinations of SM particles. The diverse masses and lifetimes of the v -hadrons, their multiplicities, and the variety of possible final states make the v -phenomenology complex.

In this work, the model taken as a benchmark predicts the existence of a leptophobic heavy resonance, a Z' gauge boson ($\mathcal{O}(1)$ TeV), connecting SM and DM sectors. For illustration, we consider a toy scenario where the dark sector is an $SU(2)_d$ gauge theory containing two fermionic states $\chi_i = \chi_{1,2}$, dark quarks. In this thesis, we will only focus on the production of dark quarks. The connection between the SM model sector (visible) and the dark sector (invisible) is ensured by the Z' boson, that couples to the SM baryonic

current J_{SM}^μ as:

$$\mathcal{L} \supset -\frac{1}{4} Z'^{\mu\nu} Z'_{\mu\nu} - \frac{1}{2} M_{Z'}^2 Z'_\mu Z'^\mu - g_{Z'}^{\text{SM}} Z'_\mu J_{\text{SM}}^\mu, \quad (1.24)$$

where $g_{Z'}^{\text{SM}}$ is the coupling between SM quarks and the Z' , and the dark sector is accounted as a $\text{SU}(2)_d$ gauge theory with coupling α_d . A schematic depiction of the interplay between the two sectors, along with a representative Feynman diagram of the process of interest, is shown in Figure 1.10.

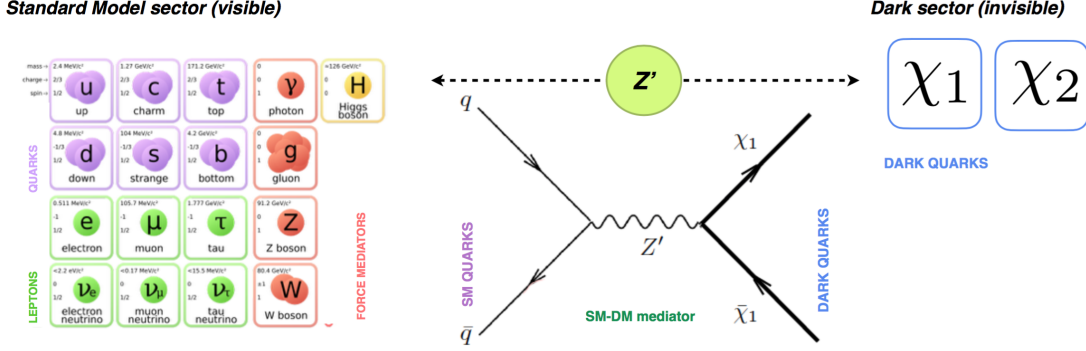


Figure 1.10: Schematic view of the interplay between the SM and the dark sector through the Z' mediator.

The symmetry group under scrutiny has a confinement scale $\Lambda_d \ll M_{Z'}$ and the dark showering occurs when $M_d^2 \sim \Lambda_d^2$. This results in the production of many dark quarks and gluons that will hadronize and give rise to a QCD-like shower. This hadronization leads to the production of *semi-visible jets* [112, 114], sprays of collimated particles from the SM and dark sectors. An illustrative comparison between the SM jets and semi-visible jets is depicted in Figure 1.11.

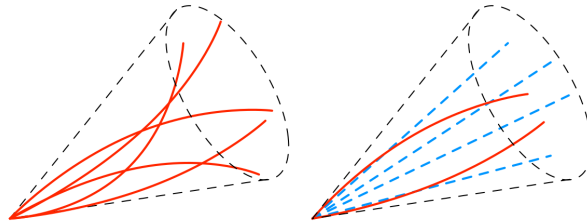


Figure 1.11: Schematic of a SM jet (left) and a semi-visible jet (right). Dashed blue lines indicate invisible jet constituents, while red lines indicate visible ones.

The parameters that define the properties of the hadronization and hence, the jet observables, are:

- the production rate, $\sigma \times \text{BR}$;
- the dark hadron mass scale M_d ;
- the Z' pole mass, $M_{Z'}$;

- the strength of the dark shower α_d , handling the dark hadrons emission in the shower and controlling their momentum distribution;
- the ratio of dark-matter particles produced in the shower over the total number of hadrons, r_{inv} , defined as:

$$r_{inv} \equiv \left(\frac{\text{number of stable hadrons}}{\text{number of hadrons}} \right). \quad (1.25)$$

1.3.4 Dark Matter searches at the LHC

Although very promising, looking for direct production of DM particles at hadron colliders is very challenging. Dark Matter does not interact or interact very weakly with SM particles. From an experimental perspective, it means that they escape the detection, similarly to neutrinos. Nevertheless, the production of DM can be inferred by the observation of an imbalance in the recorded momentum of the collision event. This imbalance is quantified by the *missing transverse momentum*: a momentum deficit in the final state with respect to the initial one. The missing momentum, E_T^{miss} , is calculated as the negative vectorial sum of the momenta of all reconstructed particles in the plane transverse to the proton beams. The missing transverse momentum is an indicator of the production of invisible particles, such as DM.

The assumption of DM being constituted of WIMPs is behind the majority of DM searches carried out at particle colliders. WIMP searches rely on identifying events with large genuine missing momentum recoiling against visible objects (Figure 1.12). These searches are based on the selection of data with missing momentum reconstructed as back-to-back to visible objects, to reduce background from SM processes and enhance a possible WIMP signal. While very effective for searching WIMP candidates (selection efficiency of $\sim 70\%$ on WIMPs data), this selection has very poor efficiency on semi-visible jets ($\sim 7\%$) [112]. DM particles produced in semi-visible jets are collinear to the visible jets. Thus the missing momentum is parallel to the jets (Figure 1.12). The WIMP strategy is therefore suboptimal for the search presented in this thesis, as it rejects semi-visible jets data and background data alike. To select a data sample enriched in semi-visible jets, the WIMP strategy must be abandoned entering a new territory.

In addition, artificial missing momentum may also appear in the absence of invisible particles, due to detector inefficiencies, mis-calibration, and particle mis-reconstruction. This artificial E_T^{miss} is usually parallel to visible objects, as shown in Figure 1.12.

Since the presence of DM can only be inferred when E_T^{miss} is also detected in the event, DM candidates are investigated in processes where they are produced together with additional detectable SM particles.

At the time of writing, no dedicated search exists for hidden sectors yielding such signature. In the analysis presented in this thesis, the semi-visible jets hypothesis is investigated at particle colliders for the first time. It plays a crucial role in the context of DM searches giving a different perspective to its nature and more in general in the query of new physics hiding in hadronic final states.

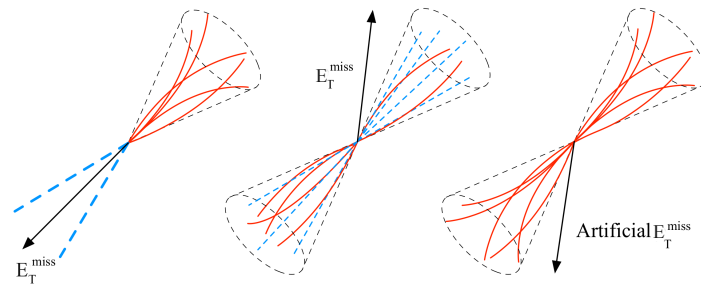


Figure 1.12: Schematic of E_T^{miss} for WIMP (left), semi-visible jets (center) and SM-jets (right) events. Dashed blue lines indicate invisible DM jet constituents, while red lines indicate visible, SM.

Chapter 2

The Large Hadron Collider and the CMS experiment

“ The LHC is certainly, by far, the biggest jump into the unknown. ”

Brian Cox, *The Six Billion Dollar Experiment*, 2007

The physics analyses presented in this thesis are based on proton-proton (pp) collisions data collected by the CMS experiment [11] at LHC [10]. The LHC is a particle accelerator designed to collide proton beams at a center-of-mass energy of 14 TeV and unprecedented luminosity of $10^{34} \text{ cm}^{-2}\text{s}^{-1}$. One of the prime motivations of the LHC construction was the search for the Higgs boson. This goal was achieved in 2012 when the ATLAS and CMS Collaborations announced the observation of a particle consistent with the SM Higgs boson [8, 9]. The research program currently pursued at the LHC focuses on studies of the properties of the Higgs boson and on the investigation of new physics scenarios.

In Section 2.1 a brief introduction to the LHC is presented, whereas Section 2.2 focuses on the description of the CMS experiment. The main components of the detector and their performance are also discussed.

2.1 The Large Hadron Collider

The LHC is a hadron accelerator and collider situated at the European Organization for Nuclear Research (CERN), on the french-swiss border in Geneva (Switzerland). Its structure consists of a 26.7 km long ring, located at a depth ranging from 45 to 170 mt underground. The LHC is installed in the already existing LEP tunnel at CERN. The design of the LHC was driven by the necessity to shed light on the nature of the electroweak symmetry breaking, for which the Higgs mechanism is responsible, and to investigate the TeV scale, where new physics beyond the SM is expected.

The two colliding proton beams can reach a design center-of-mass energy of 14 TeV. They are accelerated and then injected in the LHC ring through a chain of several accel-

erators. First, a Linear Accelerator (Linac) increases the energy of the proton beam up to 50 MeV and injects it in the Proton Synchrotron Booster (PBS) until the particles reach 1.4 GeV. After the LINAC a Proton Synchrotron (PS) accelerates the bunches up to 25 GeV. The protons are then transferred to the SPS Super Proton Synchrotron (SPS), where the beam finally reaches an energy of 450 GeV. The last step is the injection in the LHC ring, where the beam energy is finally brought up to a design value of 7 TeV. The LHC acceleration chain is shown in Figure 2.1. At the LHC there are two separate rings, where the two beams are accelerated in opposite directions and where they collide at fixed interaction points (IPs). The protons are bent by 1232 superconducting dipoles operating at a temperature of 1.9 K and producing a magnetic field of 8.3 T. The accelerating power of the machine is given by radio-frequency (RF) cavities, whose oscillation frequency is tuned to 400 MHz. The energy gain per revolution is about 485 keV, in spite of the 7 keV loss per turn due to synchrotron radiation. To focus the particle beam additional quadrupole and sextupole magnets are installed, resulting in a higher interaction probability in the four IPs. At the end of the acceleration chain, protons arrive to the IPs in 2808 bunches, each containing 10^{11} protons.

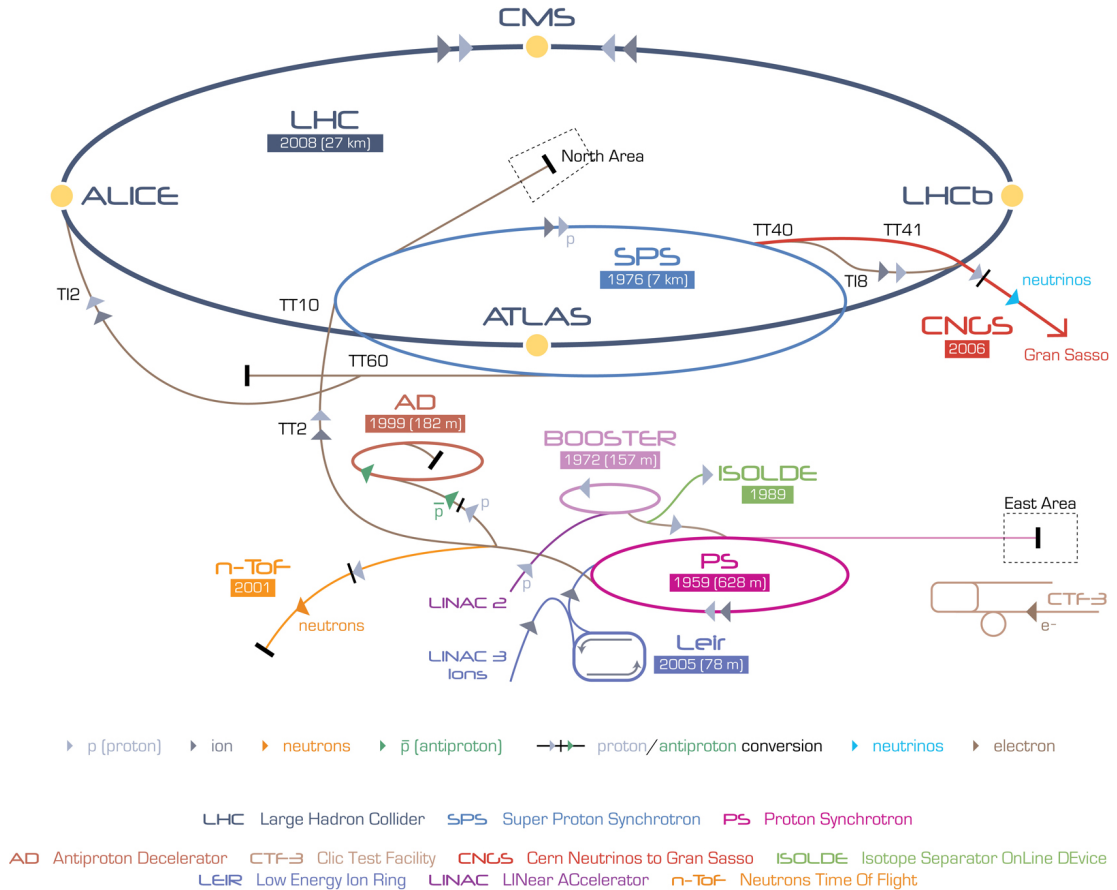


Figure 2.1: Schematic draw of the LHC injection complex [115].

The basic layout of the LHC is not a perfect circumference, being made of 8 straight sections and arcs, as shown in Figure 2.2. Each straight section has a length of approximately 528 m and can serve either as an experimental location or as a utility insertion. In four of

the eight sections, the two beams cross and around these interaction points the four main experiments are installed:

- A Toroidal LHC ApparatuS (ATLAS) [116], an high luminosity general-purpose experiment located at Point 1;
- Compact Muon Solenoid (CMS) [11], installed at the diametrically opposite straight section than ATLAS, at Point 5. As the previous experiment also CMS is a high luminosity multipurpose experiment;
- A Large Ion Collider Experiment (ALICE) [117], located at Point 2, it is a heavy-ion detector designed to study the physics of strongly interacting matter at extreme energy densities;
- Large Hadron Collider beauty (LHCb) [118], a specialized b-physics experiment installed at Point 8.

The remaining four straight sections are not equipped with detectors. Insertions at Points 3 and 7 each contain two collimation systems. At point 4 there are two RF systems, one for each proton beam. The insertion at Point 6 is the beam dumping area, where the two beams are extracted from the machine. Finally, beams injection systems are located at Point 2 and Point 8.

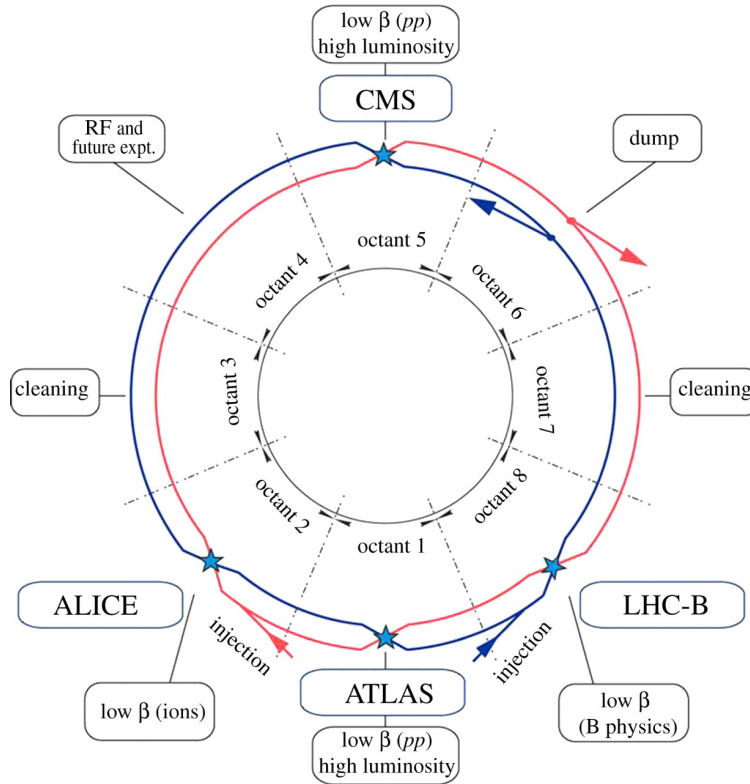


Figure 2.2: Schematic layout of the LHC. The two beams and the four main interaction points with the corresponding experiments are also shown [119].

2.1.1 Proton-proton collisions

The number of events generated per second [120] in the LHC collisions is given by:

$$\frac{dN_{\text{events}}}{dt} = \mathcal{L}\sigma_{\text{events}}, \quad (2.1)$$

where σ_{events} is the cross-section for the event under study and $\mathcal{L}$ is the machine instantaneous luminosity, measured in $\text{cm}^{-2}\text{s}^{-1}$.

In the case of two colliding beams, the machine luminosity depends on the beam parameters. Under the assumption of Gaussian distribution of the beams, the instantaneous luminosity can be expressed as:

$$\mathcal{L} = \frac{N_1 N_2 f N_b}{4\pi\sigma_x\sigma_y}, \quad (2.2)$$

where N_1 and N_2 are the number of particles per bunch, f is the revolution frequency and N_b is the number of bunches. The beam size along the x and y directions are labeled with σ_x and σ_y respectively. Another important quantity is the integrated luminosity [120], defined as the integral taken over the time of $\mathcal{L}$:

$$\mathcal{L}_{\text{int}} = \int_0^T \mathcal{L}(t) dt. \quad (2.3)$$

This quantity directly relates to the number of observed events N_{events} . Combining Equation 2.1 and Equation 2.3 one finds that the integrated luminosity can be computed as:

$$\mathcal{L}_{\text{int}} = \frac{N_{\text{events}}}{\sigma_{\text{events}}}. \quad (2.4)$$

The LHC was designed to collide protons at a center-of-mass energy of 14 TeV and with peak instantaneous luminosity of $10^{34} \text{ cm}^{-2}\text{s}^{-1}$. During the data acquisition held from 2010 to 2011 the energy reached was 7 TeV. Afterward, it was increased to 8 TeV and 13 TeV for the 2012 and 2015-2018 runs, respectively. The LHC also collides heavy ions at an energy of 2.8 TeV and a peak luminosity of $10^{27} \text{ cm}^{-2}\text{s}^{-1}$.

The CMS experiment records data delivered by the LHC and certifies data as suitable for physics analyses if during stable beams all detectors and reconstructed physics objects show good performance [121]. The validated integrated luminosity obtained by CMS in 2010 and 2011 is 34.68 pb^{-1} and 5.09 fb^{-1} , respectively. In 2012 and 2015 the certified datasets correspond to integrated luminosities of 19.7 fb^{-1} and 2.2 fb^{-1} , respectively. In 2016, a total of 35.9 fb^{-1} data have been recorded and validated, while the amount of data that can be used for physics analyses taken from 2017 collisions is 41.4 fb^{-1} . In 2018, 59.8 fb^{-1} of data have been validated and used for the physics analyses. The total integrated luminosities delivered by the LHC and collected by the CMS experiment in 2016, 2017 and 2018 are shown in Figure 2.3.

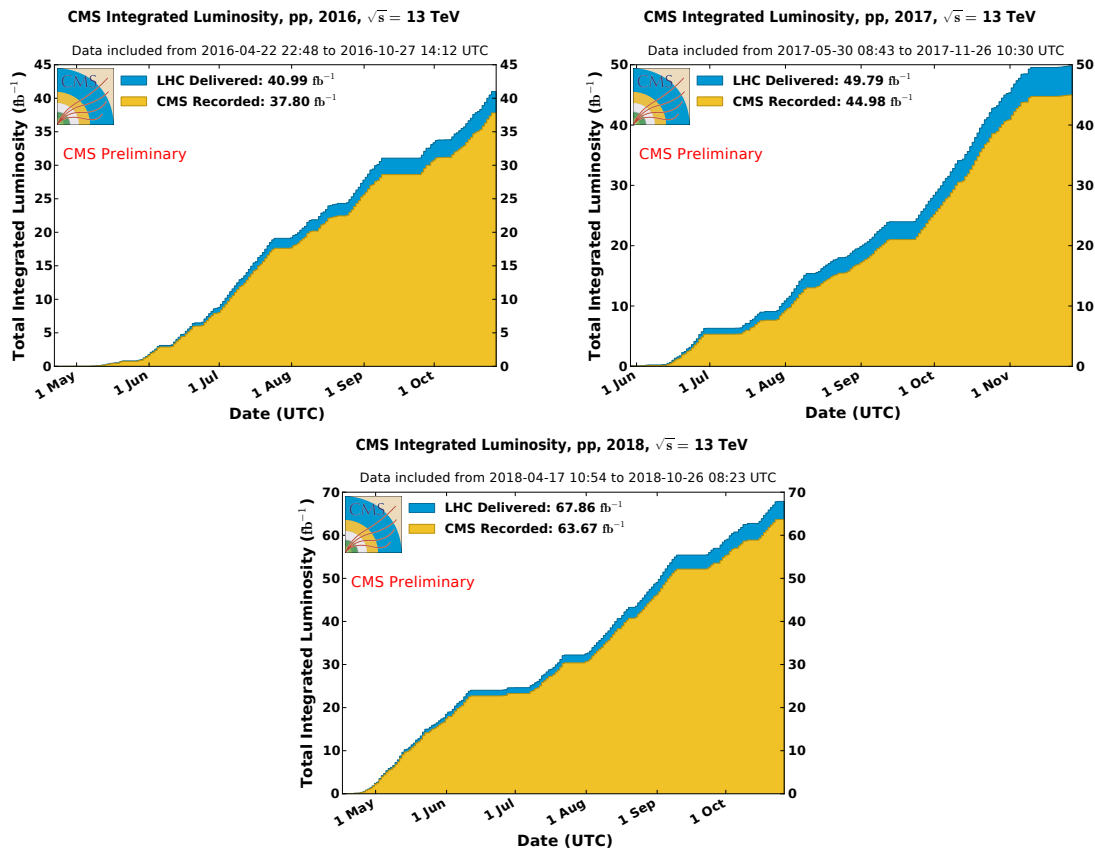


Figure 2.3: The plot shows the cumulative curves for the luminosity delivered by the LHC (azure) and recorded by CMS (orange) in the proton-proton collision events held in 2016 (top left) and 2017 (top right), and 2018 (bottom) [122].

2.2 The CMS experiment

The CMS detector is a general-purpose detector and one of the four main experiments located at CERN that investigate the data collected with the LHC. It has a length of 21.6 m, with a diameter of approximately 14.6 m and an overall weight of 12500 tons. CMS has a symmetric cylindrical shape and its central feature is a superconducting solenoid generating a magnetic field of 3.8 T. Inside the solenoid are installed several sub-detectors: a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL), and a brass and scintillator hadron calorimeter (HCAL), each composed of a barrel and two endcaps. Forward sampling calorimeters extend the pseudorapidity coverage to high values ($|\eta| < 5$). Muons are detected in gas-ionization chambers, which are instrumented with the steel flux-return yoke outside the solenoid. Figure 2.4 pictures a schematic view of the detector.

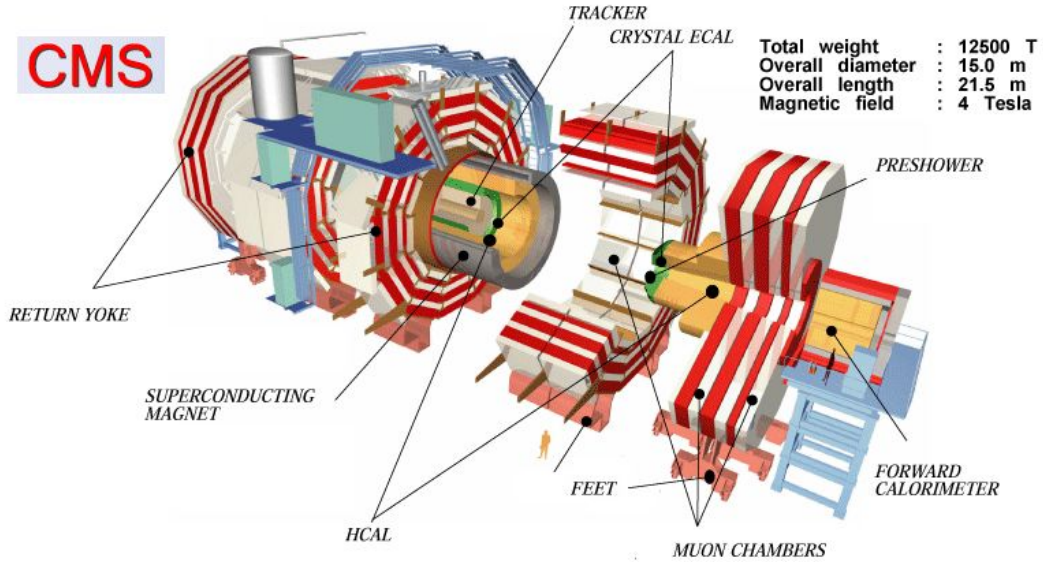


Figure 2.4: An exploded view of the CMS detector, showing the outer muon chambers (white) embedded in iron (red). Internally can be seen the calorimeter system (ECAL, in green, and HCAL, in orange). The tracker is located in the core of the central barrel. [123]

A dedicated set of convenient coordinates is used to describe the kinematic of the particles arising in collision events recorded by CMS. The origin of the coordinates is the geometrical center of the detector, which corresponds to the beams IP. The z -axis is identified as the direction of the anti-clockwise beam and it is pointing towards the Jura mountains. The x - and y -axis point to the center of the LHC ring and vertically to the ring plane, respectively. The azimuthal angle, ϕ , is measured from the x -axis on the x - y plane and its value ranges in the interval $[-\pi, +\pi]$. The polar angle θ is measured from the z -axis in the z - y plane, and it ranges from $[0, \pi]$. The pseudorapidity η is defined as $\eta = -\log(\tan \theta/2)$. A sketch of the coordinates system used at CMS is presented in Figure 2.5.

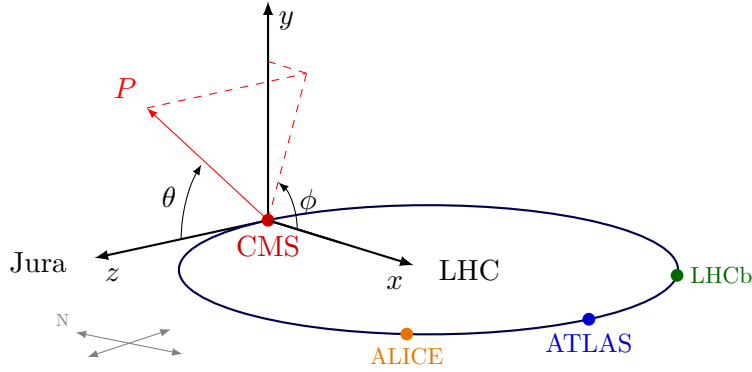


Figure 2.5: Schematic layout of the LHC, where the four largest experiments (ALICE, ATLAS, CMS and LHCb) are also shown. The coordinates system is also pictured.

2.2.1 Magnet

The superconducting magnet [124] of CMS has a diameter of about 6 m and is 13 m long. It generates a high magnetic field of 3.8 TeV. The tracking system and part of the calorimeter detectors are installed inside the coil. A 10000 T iron yoke is used to return the magnetic flux and muon chambers are integrated in it.

Charged particles follow helical trajectories inside the CMS solenoid. Given a bending radius ρ and a magnetic field of intensity B , the transverse component of the momentum, p_T , is given by [125]:

$$p_T[\text{GeV}] = 0.3B[\text{T}]\rho[\text{m}]. \quad (2.5)$$

Assuming a very large bending radius compared to the magnet length L and considering the finite track resolution of the detector, the momentum resolution $\sigma(p_T)$ results to be proportional to p_T^2 . It can be expressed as:

$$\frac{\sigma(p_T)}{p_T} = \frac{p_T[\text{GeV}]}{0.3L[\text{m}]}B[\text{T}]\frac{2\sigma(x)}{d}, \quad (2.6)$$

where $\sigma(x)$ is the error on the track measurements and d is the distance between the sensors. Typically for charged particles of ~ 100 GeV the transverse momentum resolution is about 1–2% in the central pseudorapidity region [126], beyond which it degrades due to the reduced lever arm. At lower momenta, the resolution is dominated by multiple scattering and its distribution reflects the amount of material traversed by the track.

2.2.2 Inner tracking system

The CMS inner tracking system [127, 128] is the innermost sub-detector of CMS. Its volume is given by a cylinder 5.8 m long and with a diameter of 2.5 m, covering a pseudorapidity region up to $|\eta| < 2.5$. Its main function is the identification of the trajectories of charged particles emerging from the collisions, allowing the measurement of their momentum. Besides, it gives information on the spatial location where the proton beams collide.

The tracker is composed of pixels and strip detectors. A schematic drawing of the tracking system, where all the sub-components are also shown, is pictured in Figure 2.6.

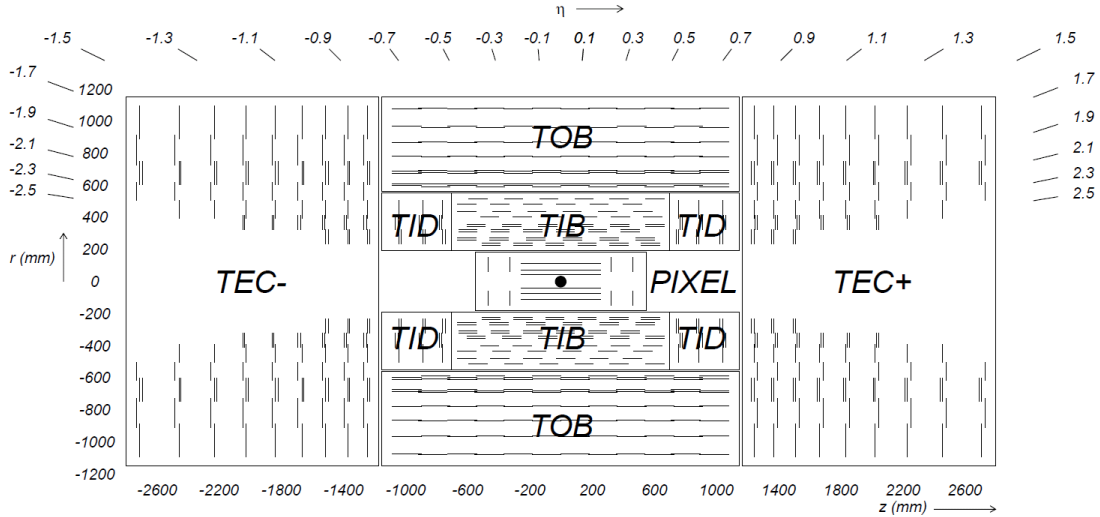


Figure 2.6: Schematic cross-section through the CMS tracker [11]. Each line represents a detector module. Double lines indicate back-to-back modules.

Pixel detector, and its upgrade

The pixel system is placed close to the IP and constitutes the innermost part of the tracker. It is composed of about 66 million pixel cells, each of them having an approximate size of $100 \times 150 \mu\text{m}^2$. The pixel detectors are grouped into three barrel layers (BPix) and two endcap disks (FPix). The BPix layers are placed at radii $r = 4.4, 7.3$, and 10.2 cm , and they extend along the x -direction for 53 cm . The FPix disks extend in radius from 6 to 15 cm and are positioned at z values of $z = \pm 34.5$ and $\pm 46.5 \text{ cm}$. The choice of this configuration was driven by the necessity to precisely measure the three-dimensional position of the hits due to the interaction of charged particles. This plays a crucial role in the reconstruction of vertices and the discrimination between heavy and light flavor quarks. The spatial resolution achieved with the pixel detector is approximately $10 \mu\text{m}$ in the $r - \phi$ plane and $20 \mu\text{m}$ along the z -axis [129].

The pixel detector was decommissioned during the Technical Stop held between 2016 and 2017. We will refer to this detector layout as *phase 0* pixel detector. In its place, a new detector - named *phase 1* pixel detector - has been installed. This replacement was foreseen and designed to cope with the increased instantaneous luminosity of the LHC, exceeding the design performance of $1 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ by more than 50%. As a consequence of the higher instantaneous luminosity, also the number of vertices arising from the additional pp interactions in the same bunch crossing (i.e. *pileup* interactions) increased, going from an average of 25 in the design conditions to an average of 50 at $1 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$. The installation of the new pixel detector ensured the operations and the data taking in such harsh conditions. Several and crucial modifications were deployed to design and build the new detector, but the main core, concept, and structure are inherited from the phase 0 apparatus. The main improvements that were applied to the upgraded detector are [130]:

- the geometrical layout: an additional layer was installed in the barrel section and an additional disk was added to each endcap section, to improve the coverage from 3 to 4 hits. The spatial locations of layers and disks were also modified, being the innermost

layer closer to the interaction point, resulting in an improved vertex identification. With the inclusion of a fourth layer in the central region of the detector, the pixel coverage was extended much further from the interaction point, reducing the gap between the pixel and the tracker systems and decreasing the combinatorics in the track reconstruction. A comparison of the geometrical layouts of phase 0 and phase 1 detectors is given in Figure 2.7.

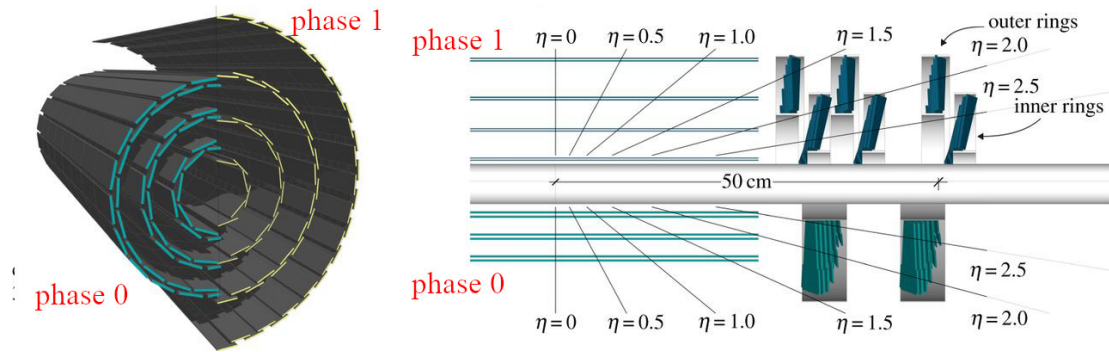


Figure 2.7: Transverse (left) and longitudinal views (right) comparing the phase 0 and phase 1 pixel detectors. [130]

- A new cooling system: the old method is replaced by a new system that improves the cooling performance, allowing lower temperatures. The rise of the sensor leakage current after irradiation is mitigated, leading to a longer detector lifetime. Also, this new system, along with a lightweight supporting structure, allows a reduction of the material budget for pseudorapidities values $\sim |\eta| > 1.2$.
- Optimized design for the readout chips (ROC): a new design is deployed to replace the analog system of the previous chip. This upgrade results in an increased readout bandwidth and a reduction of the hit inefficiencies. While the same design and readout scheme is used in the forward pixel disks and the three outer layers of the barrel section, a dedicated ROC has been designed for the first layer, which is the one with the highest expected hit rate and thus collecting the largest data flux.

In this thesis, a dedicated study aiming to measure and monitor the pixel resolution for phase 0 and phase 1 pixel detectors is performed. More details about the strategy and the results are provided in Appendix A.

Strips detector

The strips detector encapsulates the pixel detector and is composed of three main sub-systems [129]:

- TIB/TID, tracker inner barrel and disks;
- TOB, tracker outer barrel, which surrounds the TIB/TID;
- TEC+/TEC-, tracker endcaps.

The innermost strip sub-detectors, TIB/TID, extend in radius up to 55 cm and are grouped in 4 barrel layers and 3 disks at each end. The silicon sensors in this region have a length of about 320 μm and are parallel to the beam axis in the barrel and perpendicular to it in the endcaps. The resolution in $r\phi$ is $\approx 13 - 38 \mu\text{m}$. The TOB surrounds the TIB/TID. It extends in radius up to 116 cm and is divided into 6 barrel layers of 500 μm thick microstrips. It has a length along the z -axis of 118 cm. It provides positions measurements with a resolution of about $\approx 18 - 47 \mu\text{m}$. Finally, the TEC+/TEC- cover the region ranging between 124 cm and 282 cm in z and 22.5 cm and 113.5 cm in r . Both TEC sub-detectors are composed of 9 disks and the sensors here located have a thickness of 320 - 500 μm .

2.2.3 Calorimeters

The CMS calorimeter system exploits the interaction between particles and matter and measures particles energy. It is composed of two main sub-detectors: the electromagnetic calorimeter (ECAL) and the hadron calorimeter (HCAL).

Electromagnetic calorimeter

The CMS electromagnetic calorimeter [131] is an hermetic and homogeneous calorimeter, located within the solenoidal coil, just outside the tracking system. It is built with PbWO_4 crystals, with high density and short radiation length, making ECAL compact and with fine granularity.

The electromagnetic calorimeter is grouped into a barrel (EP) and two endcaps (EE). The former covers a pseudorapidity range up to $|\eta| = 1.479$, while the latter covers the range $[1.479, 3.0]$. The crystals in the EP have a tapered shape, with front face area of $22 \times 22 \text{ mm}^2$ and a rear face area of $26 \times 26 \text{ mm}^2$; the ones in the EE part have a front area of $28.62 \times 28.62 \text{ mm}^2$ with a rear cross-section of $30 \times 30 \text{ mm}^2$. The crystals placed in the endcaps are further grouped into mechanical units of 5×5 crystals, called supercrystals (SCs). Each endcap is also split into two "Dees", regions shaped like the letter D. An additional detector, called preshower, is located between the tracking system and the calorimeter endcaps. It covers a fiducial region of $1.653 < |\eta| < 2.6$ and has a total thickness of 20 cm. Its main role is the identification of neutral pions and the improvement of the estimation of the position of the photons. A schematic view of the electromagnetic calorimeter, along with its main sub-components, is pictured in Figure 2.8.

The energy resolution of the ECAL detector can be parametrized with three main contributions: a stochastic, a noise, and a constant term [132]. It can be expressed as:

$$\frac{\sigma_E}{E} = \frac{2.8\%}{\sqrt{E}} \oplus \frac{12\%}{E} \oplus 0.3\%, \quad (2.7)$$

where the parameters' values were estimated in a test beam. The stochastic term takes into account the statistical fluctuations in the light efficiency and quantum efficiency, the noise term describes the electronic noise and finally, the constant parameter accounts for the effect of the operational conditions. This estimate corresponds to an energy resolution of 1% for photons and electrons with energy larger than 10 GeV. The energy resolution is illustrated in Figure 2.9 as a function of electron energy.

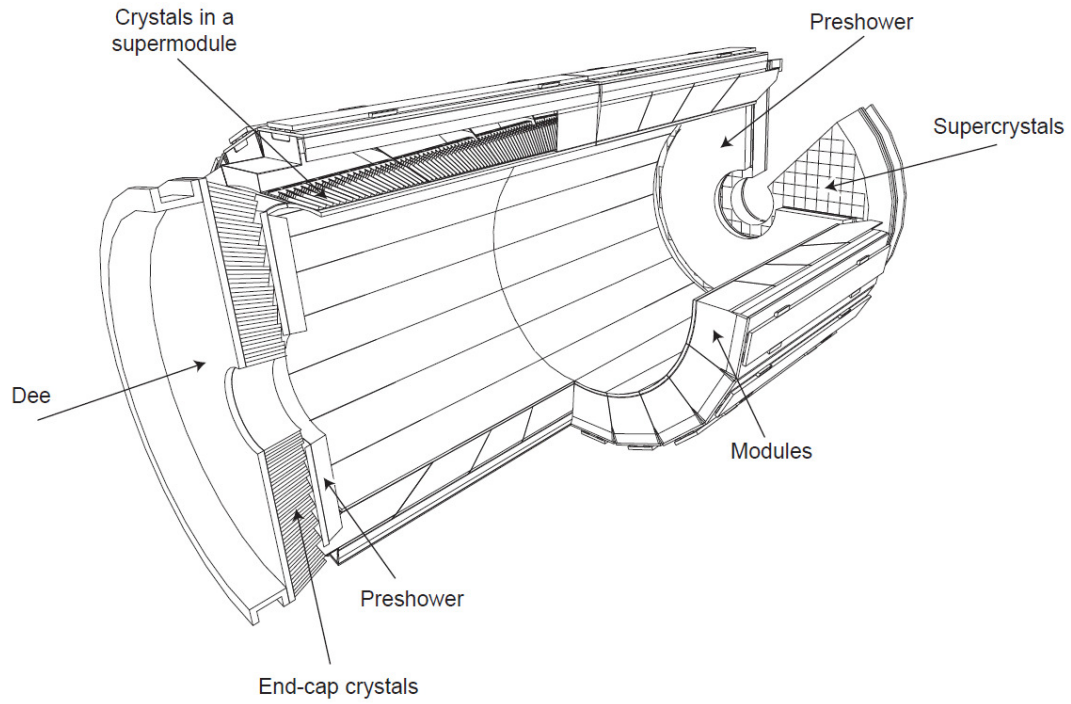


Figure 2.8: Layout of the CMS electromagnetic calorimeter showing the arrangement of crystal modules, supermodules and endcaps, with the preshower in front. [11]

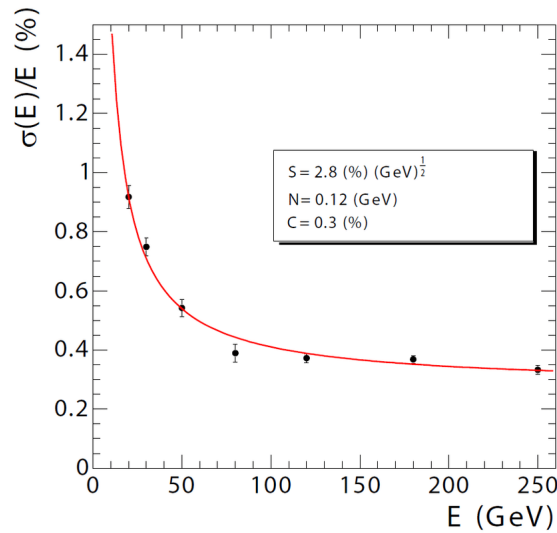


Figure 2.9: The ECAL energy resolution as a function of electron energy from test beam performance [132].

Hadron calorimeter

Surrounding ECAL, there is the hadron calorimeter, which is formed of alternated layers of absorbing plates and active material. Its prime motivation is the reconstruction of jets and missing transverse energy.

This calorimeter is divided into four parts: the hadron barrel (HB) and endcap (HE), the hadron outer (HO) and hadron forward (HF). The HB has a fiducial pseudorapidity region of $|\eta| < 1.3$, and extends along r from 1.77 m to 2.95 m. It is a sampling calorimeter, and brass is used as absorber material because of its short interaction length. HB has a thickness ranging from 5.82 to 10.6 interaction lengths. Since the stopping power of HB does not provide enough containment for hadron showers, the hadron calorimeter extends outside the solenoid coil with the HO, which exploits the magnet as additional absorbing material. Thanks to the HO, late-starting showers can be identified, as well as shower energy deposited after HB. The total depth of the calorimeter system is thus extended to a minimum of 11.8 interaction lengths, except at the barrel-endcap boundary region. Going more forward in the pseudorapidity region, the range $1.3 < |\eta| < 3$ is covered by the HE. This sub-detector is a sampling calorimeter, where brass layers are alternated with scintillator layers. The very forward pseudorapidity, for $3.0 < |\eta| < 5.2$, is covered by the HF. This sampling calorimeter deals with very high hadron rates and therefore is made of radiation-hard materials such as steel as absorber material and quartz fibers as active medium. A schematic longitudinal view of the CMS hadron calorimeter is presented in Figure 2.10. The main sub-components are highlighted.

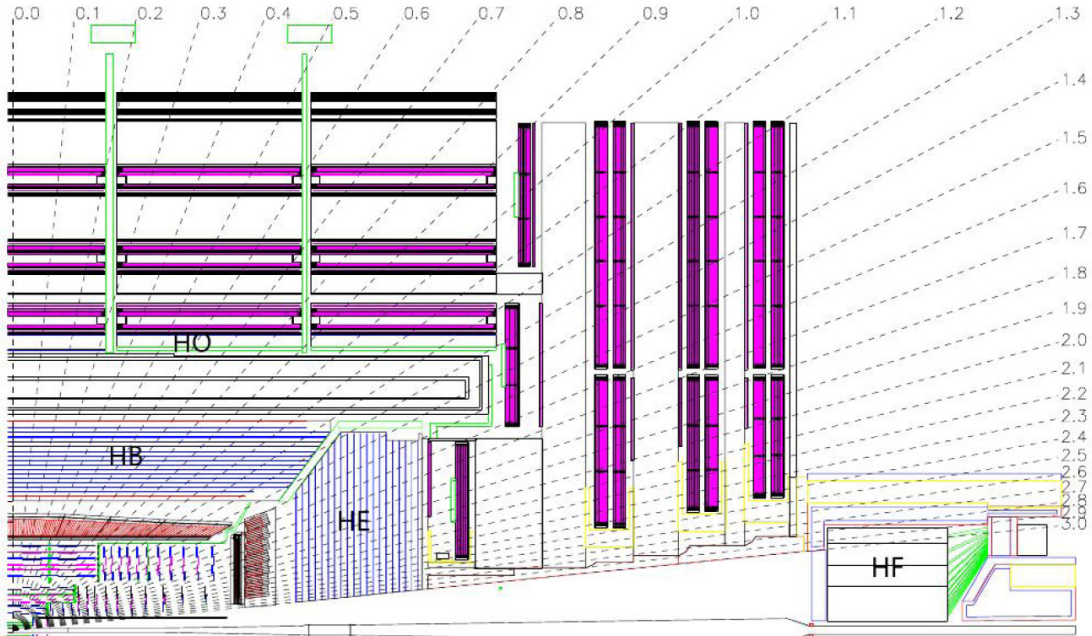


Figure 2.10: Longitudinal view of the CMS detector showing the locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters [11].

The energy resolution for HCAL is described with a parametrization of a stochastic,

noise and constant terms [133]. It can be expressed as:

$$\frac{\sigma_E}{E} = \frac{115\%}{\sqrt{E}} \oplus \frac{5.5\%}{E}, \quad (2.8)$$

where the parameters have been estimated from test beam measurements. The constant term is found to be negligible compared to the other contributions, hence it is not mentioned in Equation 2.8.

The energy resolution of the CMS ECAL and HCAL detectors for pions, resulting from test beam measurements, as a function of energy, is shown in Figure 2.11.

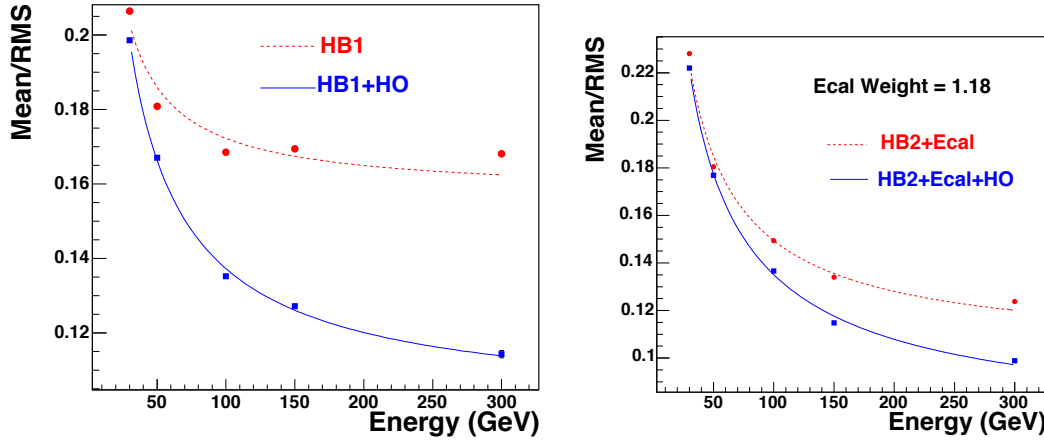


Figure 2.11: The calorimeter energy resolution as a function of pion energy using the HB only or HB+HO (left), and when adding ECAL and HCAL measurements (right) [133].

2.2.4 Muon system

The CMS muon system is responsible for the muons identification and the measurement of their momentum; it is further used also as trigger for event selection. It has a cylindrical shape and is split in a barrel, covering a pseudorapidity region up to $|\eta| < 1.2$, and two enclosing endcaps, over a range of $0.8 < |\eta| < 2.4$. Three different types of gaseous sub-detectors compose the CMS muon system. Drift chambers (DT) are located in the barrel section, because of the low muon rate and the uniformity of the magnetic field in this region. The DT chambers are interleaved with the iron yoke plates. The resolution achieved by each chamber is approximately $100 \mu\text{m}$ in $r\phi$ and 1 mrad in ϕ . In the two endcap regions the muon rates and the background level are higher, and the magnetic field is large and non-uniform. Hence, cathode strip chambers (CSC) are used for muon identification, as they provide a fast response, fine segmentation and radiation resistance. These sub-detectors are arranged perpendicularly to the beamline and interspersed between the flux return plates. The spatial resolution achieved with each chamber is of about $200 \mu\text{m}$ in $r\phi$ and 10 mrad in ϕ . A third type of gaseous particle detectors, the resistive plate chambers (RPCs), is placed both in the barrel and in endcaps regions, covering a pseudorapidity interval of $|\eta| < 1.6$. They are used as complementary trigger systems and, given their fast response, they improve the quality of the operations at high rates. On the other hand, they have a slightly worse spatial resolution than DTs and CSCs.

The muon system aims at improving the muon momentum resolution. Two factors contribute to the resolution of this measurement: the magnetic field, which is bending the trajectory of the charged particles, and the multiple scattering, as explained in Section 2.2.1. For muons reconstructed using the muon system only - called *stand-alone muons* - the multiple scattering contribution becomes lower with increasing values of transverse momentum. Up to about $p_T = 200$ GeV, this factor dominates, but afterward, it is overshoot by the lever arm effect. For particles with low p_T , the inner tracker gives a better resolution. The combination of both muon and inner tracking systems improves the resolution for muons with $p_T > 100$ GeV.

2.2.5 Trigger and data acquisition systems

At the LHC, the proton bunches collide each with a design interval of 25 ns, which corresponds to a crossing frequency of 40 MHz. On average, at the luminosity design value, 20 simultaneous proton-proton collisions take place, per each bunch crossing. To ensure high data recording efficiency for a wide variety of physics objects and signal topologies, the CMS trigger system, whose role is to reduce the high rate by selecting only interesting events, has been designed. It is implemented in two steps: the Level-1 (L1) and the High Level Trigger (HLT). The former consists of custom-designed, programmable electronics, whereas the latter is based on software algorithms. The data passing the first level of trigger decision, are read-out and assembled by the CMS Data-Acquisition (DAQ) system. The DAQ makes the event available, transfers it to the event filter farm and feeds it to the HLT system. A schematic view of the trigger and DAQ systems is presented in Figure 2.12.

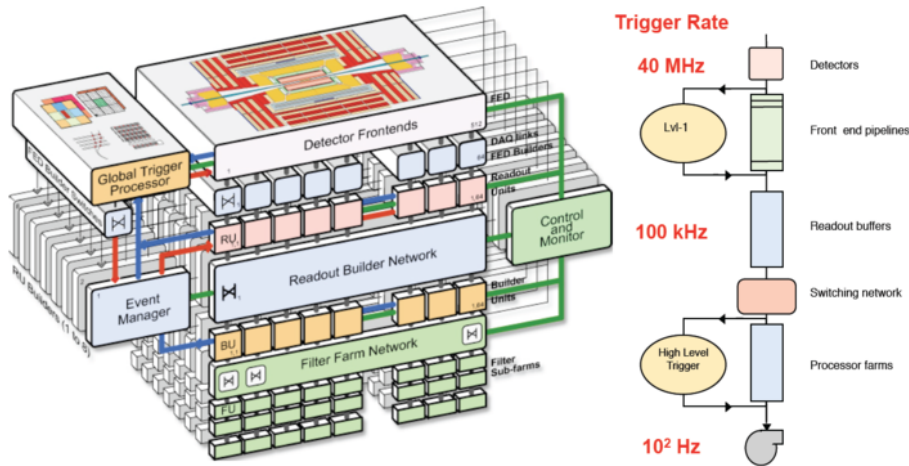


Figure 2.12: Architecture of the CMS DAQ system (left) and of the CMS trigger system configuration (right) [134].

The CMS L1 trigger has to cope with a very large amount of data, 40 MHz, and has

to reduce this high rate to about 100 kHz. It analyzes every bunch crossing and its output depends on the decisions taken by a set of sub-trigger systems, as shown in Figure 2.13. The information provided by the calorimeter trigger towers and by the muon chambers, that are energy deposits and hit patterns, respectively, are collected by Local Triggers, called Muon Trigger and Calorimeter Trigger. These data are afterward combined and sent to the Global Trigger.

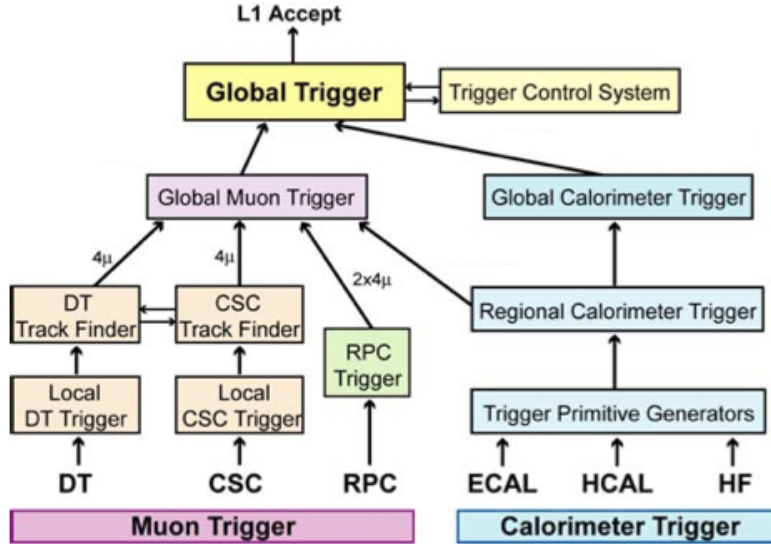


Figure 2.13: Architecture of the Level-1 Trigger system.

The High-Level Trigger, that is the second step of the CMS trigger system, further reduces the rate from 100 kHz to about 1 kHz, suitable for data storing. The HLT decision takes about 40 ms, and, to optimize the data flow, the selection is organized in progressive stages with a series of filters. Differently from the L1, the HLT can further exploit the information from the tracker and the full granularity of the calorimeters, in addition to the combinations used during the first trigger stage.

2.2.6 Computing

The LHC is designed to produce annually roughly 15 PB of data, which will be accessed and analyzed by all the scientists of the Collaborations. The events that are selected by the trigger system are recorded at the so-called Tier-0, which is the computing center at CERN. All the pieces of information on real and simulated events are then distributed to other computing centers, called Tier-1, Tier-2, and Tier-3. These centers are located all around the world and their roles slightly change depending on the experiment. In general, Tier-1 centers have the prime responsibility of managing the permanent data storage and providing computational power for the reprocessing of the data. The data obtained from Tier-1 centers will be sent to Tier-2 centers, whose role is to provide computational capacity and additional storage services. Finally, many institutions such as universities and laboratories work as Tier-3 centers and play an important role in the final processing and front-end analysis of the LHC data. This four-tiered model, which constitutes the LHC computing system, is called the LHC Grid computing system and further details about it can be found in [135].

Chapter 3

Particle reconstruction and identification

“ If I have seen further it is by standing on the shoulders of Giants. ”

Isaac Newton, *Letter from Sir Isaac Newton to Robert Hooke*, 1675

To perform a physics analysis the first step needed is the detection of particles coming from the collision between the two proton beams. Particles' identification occurs using the information provided by all the CMS sub-detectors, that were described in Chapter 2. When a particle passes through a detector, it produces an electronic signal that is recorded by the machine. These signals are then combined with dedicated algorithms that merge all pieces of information and allow us to reconstruct and identify particles.

Particle reconstruction and identification techniques employed by CMS are described in this Chapter.

3.1 Particle Flow algorithm

After the collision event, the produced particles go through the detector, passing first through the tracker. Charged particles leave signals (*hits*) in the sensitive layers, allowing the identification of their trajectories (*tracks*) and origins (*vertices*). The solenoidal magnetic field bends the trajectories, and charges and momenta of charged particles can be measured. The tracker system is surrounded by ECAL, which absorbs electrons and photons. Their energies and directions can be determined thanks to the recording of their corresponding electromagnetic showers as *clusters* of energy in neighboring cells. In ECAL also charged and neutral hadrons can initiate a hadronic shower, which will be further absorbed in the HCAL. Finally, muons and neutrinos both pass through the calorimeters, leaving them with very little or no interaction at all. Neutrinos escape undetected, while muons can be identified thanks to the hits in the muon chambers. The sketch displayed in Figure 3.1 reports a summarized view of the particles' interactions with the CMS sub-detectors.

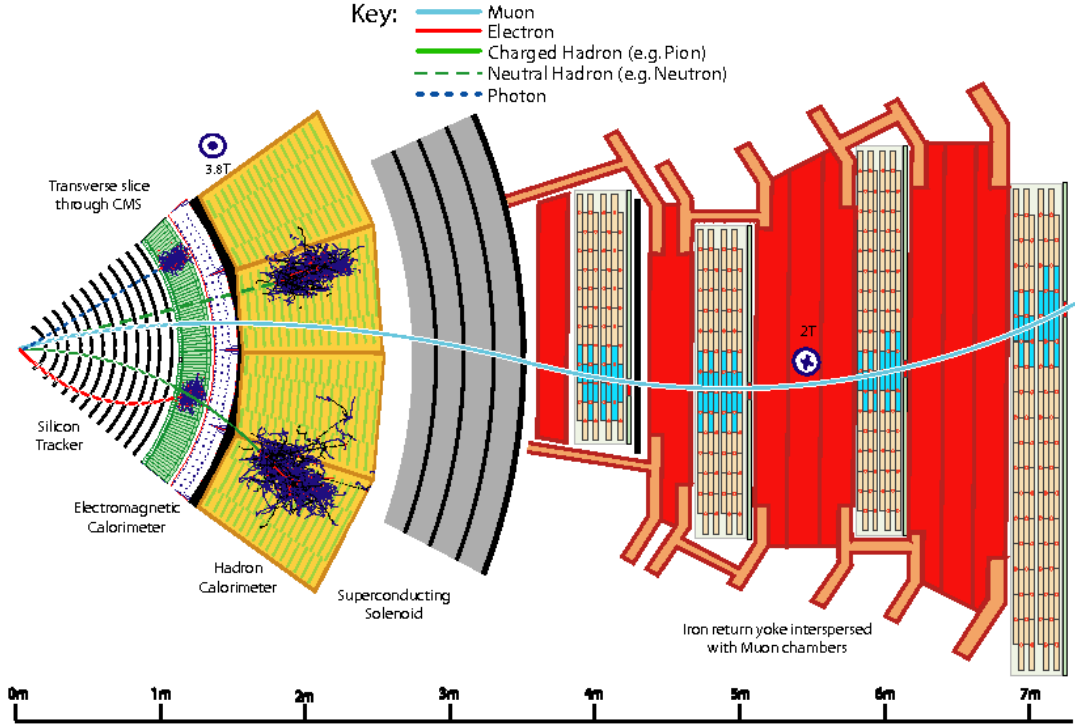


Figure 3.1: A sketch of the specific particles’ interactions in a transverse slice of the CMS detector, from the beam interaction region to the muon detector [136]. Muons and the charged pions are positively charged. Electrons are negatively charged.

Instead of reconstructing physics objects using the signals collected by the single detectors, correlating the measurements from all detector layers (i.e. tracks and clusters) leads to a significantly improved event description. The Particle Flow (PF) algorithm [136] is used to reconstruct physics objects, combining pieces of information from all sub-detectors. The building blocks of the PF algorithm are *PF tracks*, *PF clusters* and *global muons*, whose description follows.

The reconstruction of PF tracks refers to the process of using the hits in the tracking system to obtain estimates for the charged particle momentum and direction. In CMS, the tracking algorithm employed is the Combinatorial Track Finder (CTF), which exploits the Kalman filter technique [137]. This approach allows the identification of the reconstructed tracks through an iterative process. The initial iteration searches for the easiest tracks to identify and after each iteration it removes hits associated with those tracks. This results in a subsequent simplification of the iterations in a search for more difficult classes of tracks. Each step of the CTF track reconstruction is composed of four modules: seed generation, track finding, track fitting, and track selection. Initial track candidates are found using few (2 or 3) hits and they define the initial estimate of the trajectory parameter. This estimate is then used by the track finding module of the CTF algorithm, which further adds hits from successive detector layers, building track candidates. A fit is finally performed to identify the best trajectory, which is eventually selected depending on optimized quality flags imposed to suppress reconstructed tracks not associated with particles (*fake tracks*).

The clustering algorithm aims to identify PF clusters starting from energy deposits

in the calorimeters. It is responsible for the measurement of the energy and direction of stable neutral particles (i.e. photons and neutral hadrons), which are then separated from the charged hadron energy deposits. It is also responsible for the reconstruction and identification of electrons, along with the photons they radiate. Finally, this algorithm contributes to the energy measurement of charged hadrons for which the track parameters are not determined accurately (e.g. high- p_T tracks). The PF clustering algorithm is performed separately in each sub-detector. The following workflow is executed:

- the cells with energy larger than a given threshold, and larger than the energies of the neighboring cells, are identified as cluster seeds;
- cells with a least a corner in common with a cell already in the cluster and with an energy greater than a given threshold, are grouped;
- assuming Gaussian energy deposits in the calorimeter cells, an iterative procedure is employed, where the positions and energies of the Gaussian functions are taken as cluster parameters. The resulting set of reconstructed clusters corresponds to the PF cluster collection.

In addition to PF tracks and PF clusters, muons are identified as well, and they will be used as input to the PF algorithm. Muon tracking can be performed using both the muon spectrometer, allowing muons to be identified with high efficiency, and the inner tracker, which provides an accurate measurement of the muon p_T , for low- p_T muons. A muon whose track is based on hits in the DT or the CSC only are labeled *stand-alone muons*, whereas if muon track matches to a track in the tracker it is named *global muon*.

A schematic illustration summarizing the flow chart and the main building blocks and processes of the PF algorithm is shown in Figure 3.2.

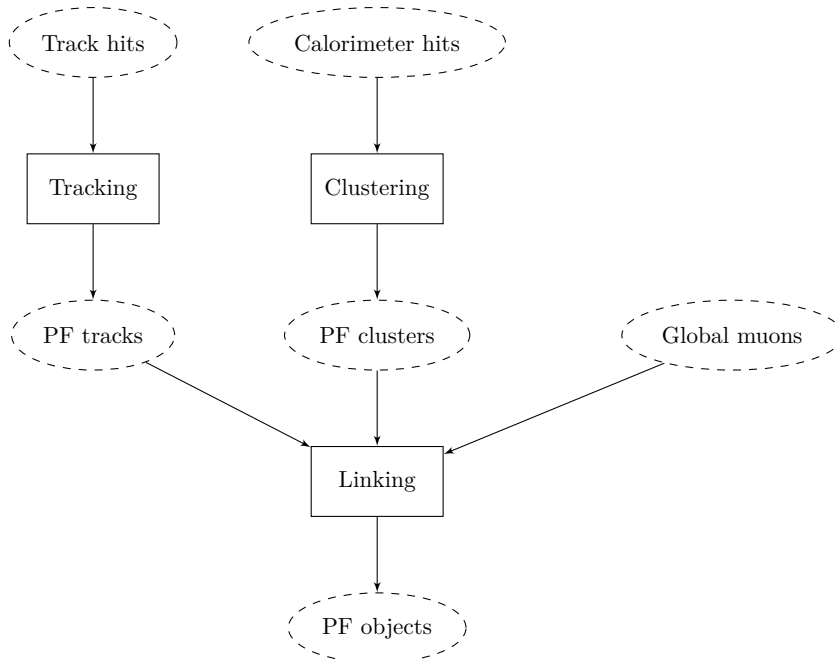


Figure 3.2: Flow chart of the PF algorithm.

Particles are expected to originate from several PF elements (PF tracks, PF clusters, and global muons) in the CMS sub-detectors. These elements are connected through a linking algorithm, which is responsible for the final physics object reconstruction. The linking algorithm then groups together the elements belonging to the same particles by producing blocks of associated elements, as described in [136]. In each PF block, the sequence aiming at the identification and reconstruction of the PF objects proceeds in the following order:

1. muon candidates are identified and reconstructed, and the corresponding PF tracks and clusters are removed from further consideration.
2. Electron candidates are identified and reconstructed. Information from both tracker and calorimeters are used. Energetic and isolated photons are also identified at this stage. Finally, the corresponding tracks and clusters are excluded from the block.
3. If the calorimeter deposits linked to one of the remaining track is compatible with a reconstructed track, it is identified as a charged hadron and the associated PF elements are not considered anymore for further processing.
4. Once all tracks and energy clusters are treated, the remaining clusters are labeled as photons, if they are reconstructed in the electromagnetic calorimeter, or neutral hadrons, if they are located in the hadron calorimeter.

These PF objects are then further employed for the jets reconstruction and the calculation of the E_T^{miss} . Quarks and gluons can never be observed freely due to the confinement, they rather hadronize leading to the generation of sprays of collimated particles, called jets. Jets reconstruction is performed starting from the PF elements. Missing transverse energy quantifies the imbalance in the visible momentum in the plane perpendicular to the beam direction. Besides, particle tracks are also necessary to reconstruct and identify the spatial locations of the pp collisions, the so-called *interaction vertices*.

The definitions and the procedures commonly deployed in CMS for the identification of the interaction vertices and of the particles used in physics analyses are described in the following Sections.

3.2 Primary vertex

The spatial location, with its associated uncertainty, of the pp interaction vertices in each collision event, is determined through the primary vertex (PV) reconstruction [138]. It uses the reconstructed tracks and it is split into three main steps:

1. *tracks are selected*, if they are produced promptly in the primary interaction region. Several requirements are imposed, such as the transverse impact parameter relative to the center of the beam spot, the number of strip and pixel hits associated with a track and the quality of the fit to the trajectory. The tracks satisfying these requests are then clustered based on their location along the z -direction, at their closest point to the center of the beam spot. With this approach multiple pp interactions occurring in the same LHC bunch crossing can be identified;

2. *tracks that appear to originate from the same interaction vertex are clustered*, using a determinist annealing (DA) algorithm [139];
3. *the position of each vertex is fit using its associated tracks*, through an adaptive vertex fitter [140] which returns the best estimate for the vertex parameters (e.g. spatial coordinates and indicators of the goodness of fit).

The resolution of the PV position strongly depends on the track multiplicity used to fit and identify the vertex, and also on the p_T of those tracks. For minimum-bias events, the resolution along x and z directions is less than $20\text{ }\mu\text{m}$ and $25\text{ }\mu\text{m}$, respectively, for primary vertices reconstructed using at least 50 tracks. Better resolution is achieved for events where more energetic tracks are selected.

The PV reconstruction efficiency is evaluated using an input set of reconstructed tracks, determining how often a vertex is successfully reconstructed and its position is consistent with the true value. The resulting efficiency is found to be close to 100% when more than 2 tracks are used in the PV reconstruction.

With the PV reconstruction, it is also possible to measure the location of the interaction vertices between the pp collisions occurring in the same LHC bunch crossing. This includes the main vertex and also any other vertex arising from the additional pp interactions in the same bunch crossing, or out of time. We referred to the latter as *pileup* (PU) events. During data-taking held in 2016, 23 PU interactions per crossing were recorded on average, while 32 PU collisions were on average reconstructed in 2017 and 2018. In Figure 3.3 the distribution of the average PU vertices in the three data-taking periods is shown.

3.3 Muons

In the standard CMS reconstruction, muon tracks are first reconstructed independently in the inner tracker (i.e. tracker muon track) and in the muon system (i.e. stand-alone muon track) [141]. Combining the information from the tracker detector and the muon chambers, three types of muon reconstructions are designed: *tracker muon reconstruction*, *stand-alone muon reconstruction*, and *global muon reconstruction*. The tracker muon reconstruction extrapolates the tracker tracks to the muon system and if at least one muon segment is consistent with the extrapolated track, the corresponding tracker track is referred to as tracker muon. A stand-alone muon is a particle whose track is based on hits in the DT or the CSC only. The global muon reconstruction exploits the Kalman filter technique to fit a global muon track, by combining the hits from the tracker track and the stand-alone muon track.

The PF algorithm applies particular selection criteria to the muon candidates, adjusted making use of information from other sub-detectors. The requirements, detailed in [142], are optimized to identify muons within jets with the highest efficiency and to keep the misidentification rate of charged hadrons as muons as low as possible.

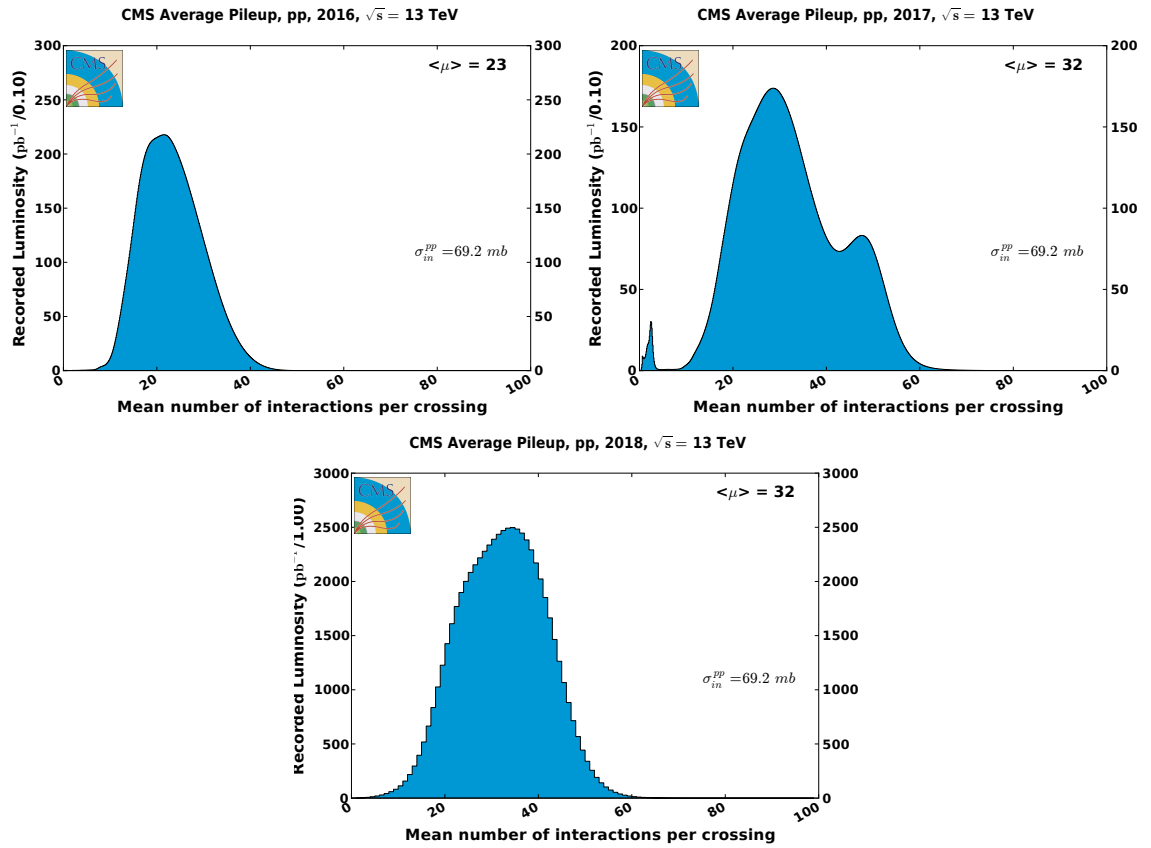


Figure 3.3: Mean number of interactions per bunch crossing for the 2016 (top left), 2017 (top right) and 2018 (bottom) pp run at 13 TeV. The thesis is taken to be 69.2 mb, which is determined by finding the best agreement with data and is recommended for CMS analyses [122].

3.4 Electrons

Electrons are reconstructed by combining the clustered energy measured in the ECAL and the track identified in the silicon detector [143]. Electrons produced within CMS and traversing the tracker detector radiate photons and release energy in the electromagnetic calorimeter with a significant shift along the azimuthal direction ϕ . The energy lost because of the photon radiation is, on average, ranging between 33% and 86% of its total energy, depending on the amount of material traversed by the electron.

To measure the initial energy of the electrons it is fundamental to consider the energy that the radiated photons spread. Dedicated reconstruction algorithms are designed and implemented. The reconstruction procedure starts from the reconstruction of superclusters in ECAL, through the *hybrid algorithm* in the barrel and the *multi 5×5* algorithm in the endcaps. The former starts from the seed crystal containing most of the energy deposited in the considered region. Arrays of crystals in a small window of η and an extended window in ϕ are added to the initial seed. The contiguous arrays are then grouped into clusters, with requirements on the energy of the seed arrays, and they will form the final global cluster, called the *supercluster* (SC). The second electron reconstruction algorithm is used in the ECAL region, where the crystals are not placed following the $\eta - \phi$ geometry. It starts from the definition of the seed crystals, defined as those with local maximal energy relative to their neighbors. The energy around these crystals is collected in clusters of 5×5 , crystals are then grouped into a supercluster if the total energy is greater than a given threshold.

To better deal with the energy losses due to the photon radiations, a specific reconstruction procedure for electrons is deployed. Tracks are reconstructed starting from ECAL superclusters that are formed from ECAL energy. The superclusters are then matched to track seeds compatible with being generated by an electron. Trajectories of the electron candidates are finally reconstructed using a *Gaussian-Sum Filter* (GSF) [144], which takes into account radiative energy losses for the electrons.

In this thesis, all electrons are required to have $p_T > 30$ GeV and $|\eta| < 2.1$. Furthermore, a dedicated Multi-Variate Analysis (MVA) [145, 146] is implemented exploiting information concerning the track and supercluster variables, such as the distance between the track and the cluster, the track χ^2 , and number of hits.

3.5 Jets

After their production in the primary interaction, partons (i.e. quarks and gluons) undergo an hadronization process. Due to color confinement, these particles can not exist individually and they combine spontaneously with other partons to form hadrons. The result is the production of a collimated cone of particles, the so-called *jet*. In CMS, quarks and gluons produced in pp collisions appear as jets.

3.5.1 Jet reconstruction

The jet energy is shared among its constituents. Typical fractions carried by charged particles, photons, and neutral hadrons are 65%, 25% and 10%, respectively [147]. Being the

calorimeters resolutions sub-optimal, an improvement in terms of jet energy resolution can be achieved exploiting the information from the tracker, reconstructing the jet with the PF algorithm. PF-jets are constituted of all particles reconstructed with the PF algorithm, without distinction of type and energy threshold

The reconstructed jets are matched to the *gen-jet*, which are obtained from all generated stable particles except for neutrinos. The matching occurs in the $\eta - \phi$ plane, by requiring a ΔR between the reconstructed jet and the matched generated jet lower than < 0.1 . In Figure 3.4 the matching efficiency, that is the fraction of generated jets matched to reconstructed jets, in two regions of pseudorapidity, is shown, where PF jets are compared to calo-jets, reconstructed only from calorimeter towers. Efficiency close to 100% are achieved already for 60 GeV jets.

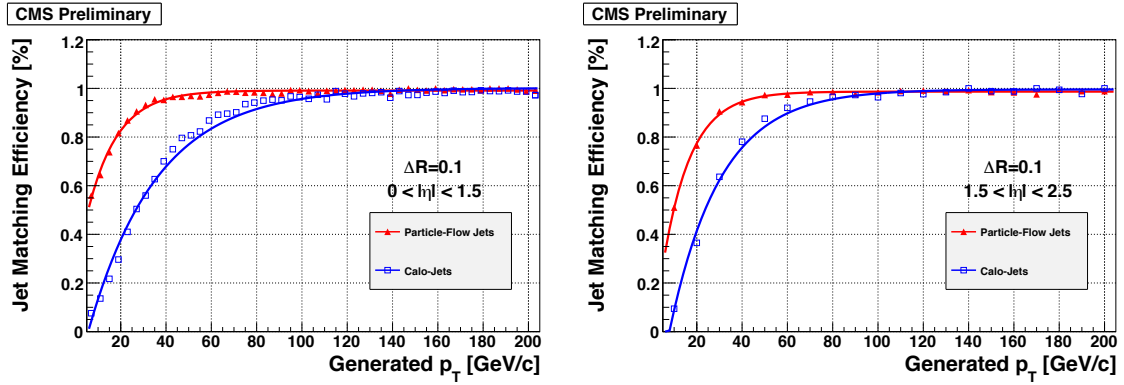


Figure 3.4: Jet matching efficiency as a function of the jet p_T , as obtained for calo-jets (open squares) and particle-flow jets (triangles) pointing to the barrel, $-1.5 < \eta < 1.5$ (left) and to the end-caps, $1.5 < |\eta| < 2.5$ (right), with a matching distance of 0.1 [147].

There is a multitude of tools aiming at clustering the constituents into jets. A good algorithm should consider the jet size and be infra-red and collinear safe [148]. The jet size and area determine the susceptibility of a jet to produce soft radiation. A larger jet radius allows the jet to capture enough of the hadronized particles, hence increasing the accuracy of the jets mass and energy calculation. On the other hand, a smaller radius is useful in reducing the amount of the underlying event and PU captured by the jet, preventing the overestimation of the jet mass and energy. In CMS, one of the most common reconstruction algorithms is the Anti- k_t [148].

This is a sequential clustering algorithm. It assumes that the jet constituents have small differences in p_T and clusters the particles based on the momentum space, resulting in jets that have fluctuating areas in $\eta - \phi$ space. Two distances are defined:

$$d_{ij} = \min \left(\frac{1}{p_{ti}^2}, \frac{1}{p_{tj}^2} \right) \times \frac{R_{ij}^2}{E}, \quad (3.1)$$

$$d_{iB} = \frac{1}{p_{ti}^2}. \quad (3.2)$$

The former describes the distance between two particles i and j , with $R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$ being the radius parameter that determines the final size of the jet. The

latter distance refers to the momentum space distance between the beam axis and the detected particle. The default choice in CMS for physics analyses is to use the Anti- k_t algorithm with $R = 0.4$ (AK4) or $R = 0.8$ (AK8), where the former is mainly deployed for regular topologies and the latter is optimized for boosted signatures, as will be described in Section 3.6.1.

The reconstruction algorithm finds first the minimum of the entire set $\{d_{ij}, d_{iB}\}$. If d_{ij} is the minimum, then particles i and j are combined into one particle using the summation of four-vectors. The two particles are then removed from the list. On the other hand, if d_{iB} is the minimum, i is considered to be a final jet and it is removed from the list of particles. This process is repeated until either all particles are part of a jet with the distance between the jet axes R_{ij} greater than R (inclusive clustering), or until a desired amount of jets have been found (exclusive clustering).

A schematic illustration of a reconstructed jet cone with radius R is reported in Figure 3.5.

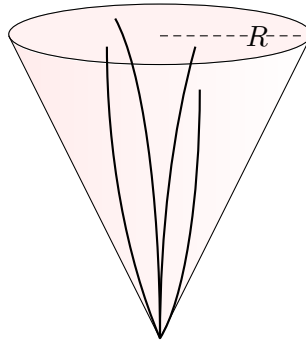


Figure 3.5: Schematic illustration for a jet cone, where the internal lines represent the jet constituents and where the radius R of the jet cone is highlighted.

At trigger level, in addition to PF-jets, an additional collection of jets, named *calo-jets* is used. These jets are originated from calorimeter towers whose energy exceeds 1 GeV and made of the energy contained in each HCAL cells and the 25 underlying ECAL crystals.

3.5.2 Jet identification

Reconstructed candidates are requested to have $|\eta| < 4.7$ and fulfill additional requirements based on their composition [35]. Jet identification criteria are applied to reject fake and badly jets reconstructed while remaining with $\sim 98/99\%$ of higher quality jets. In Tables 3.1, 3.2 and 3.3, the identification criteria for jets (both AK4 and AK8) are listed, respectively for the following regions of pseudorapidity: $|\eta| < 2.7$, $2.7 < |\eta| < 3.0$ and $|\eta| > 3.0$. The requirements are specified for 2016, 2017 and 2018 datasets.

In addition to jet identification criteria, a discriminator aiming at distinguishing jets arising from the hadronization of quarks or gluons is designed and implemented. This tool - called the quark-gluon tagger - exploits the differences between hadronic jets initiated by gluons and by light-flavor quarks, due to dissimilar radiation patterns. These distinctions reflect in internal jet structure observables, such as jet constituents, multiplicities, jet shape, and energy distribution within the jet. These informations are combined in a likelihood to

	2016		2017	2018
	Loose	Tight		
Neutral Hadron Fraction	< 0.99	< 0.90	< 0.90	< 0.90
Neutral EM Fraction	< 0.99	< 0.90	< 0.90	< 0.90
Number of Constituents	> 1	> 1	> 1	> 1
Charged Hadron Fraction	> 0	> 0	> 0	> 0
Charged Multiplicity	> 0	> 0	> 0	> 0
Charged EM Fraction	< 0.99	< 0.99	-	-

Table 3.1: List of identification criteria for jets with $|\eta| < 2.7$. The requirements on the charged components are only applied for jets with $|\eta| < 2.4$.

	2016		2017	2018
	Loose	Tight		
Neutral Hadron Fraction	< 0.98	< 0.98	-	-
Neutral EM Fraction	> 0.01	> 0.01	> 0.02 and < 0.99	> 0.02 and < 0.99
Number of Neutral Particles	> 2	> 2	> 2	> 2

Table 3.2: List of identification criteria for jets with $2.7 < |\eta| < 3.0$.

	2016		2017	2018
	Loose	Tight		
Number of Hadron Fraction	-	-	> 0.2	> 0.2
Neutral EM Fraction	< 0.90	< 0.90	< 0.90	< 0.90
Number of Neutral Particles	> 10	> 10	> 10	> 10

Table 3.3: List of identification criteria for jets with $|\eta| > 3.0$.

build a quark-gluon discriminant. The discriminant value indicates what is the probability that a specific jet is likely to be originated from a quark rather than a gluon. The quark-gluon jet tagger tool was developed in the context of this thesis and more details about it are discussed in Appendix B.

3.5.3 Jet energy calibration

The jet energy is calibrated to match the energy of a true particle jet to measured quantities [149]. In particular, a corrective factor $\mathcal{C}$ is applied to each component μ of the raw jet p_T vector:

$$p_\mu^{corr} = \mathcal{C} \cdot p_\mu^{raw}. \quad (3.3)$$

The corrective factor $\mathcal{C}$ considers several effects and corrections. It is composed of four contributions:

- an offset correction $\mathcal{C}_{offset}$, which removes the extra energy due to noise and PU;

- an MC calibration factor $\mathcal{C}_{MC}$, removing the bulk of the non-uniformity in η and the non-linearity in p_T ;
- the residual corrections $\mathcal{C}_{rel}$ and $\mathcal{C}_{abs}$, which are the relative and absolute energy scales, accounting for the remaining differences between data and simulation.

These components are applied in sequence:

$$\mathcal{C} = \mathcal{C}_{offset}(p_T^{raw}) \cdot \mathcal{C}_{MC}(p'_T, \eta) \cdot \mathcal{C}_{rel}(\eta) \cdot \mathcal{C}_{abs}(p''_T), \quad (3.4)$$

with p'_T being the transverse momentum of the jet after applying the offset correction and p''_T being the transverse momentum of the jet after all previous corrections.

After these corrections, the jet average response, named jet energy scale (JES), is known to percent-level precision even in the difficult forward regions and better than 1% for jets above 100 GeV in the central detector region [150].

The measurement of jet momentum is found to agree with its true value in simulation within 5-10% over the whole p_T spectrum and pseudorapidity acceptance [150]. In the barrel region, the jet energy resolution (JER) amounts typically to 15% at 10 GeV, 8% at 100 GeV, and 4% at 1 TeV, while resolution slightly worsens in the endcaps, as shown in Figure 3.6.

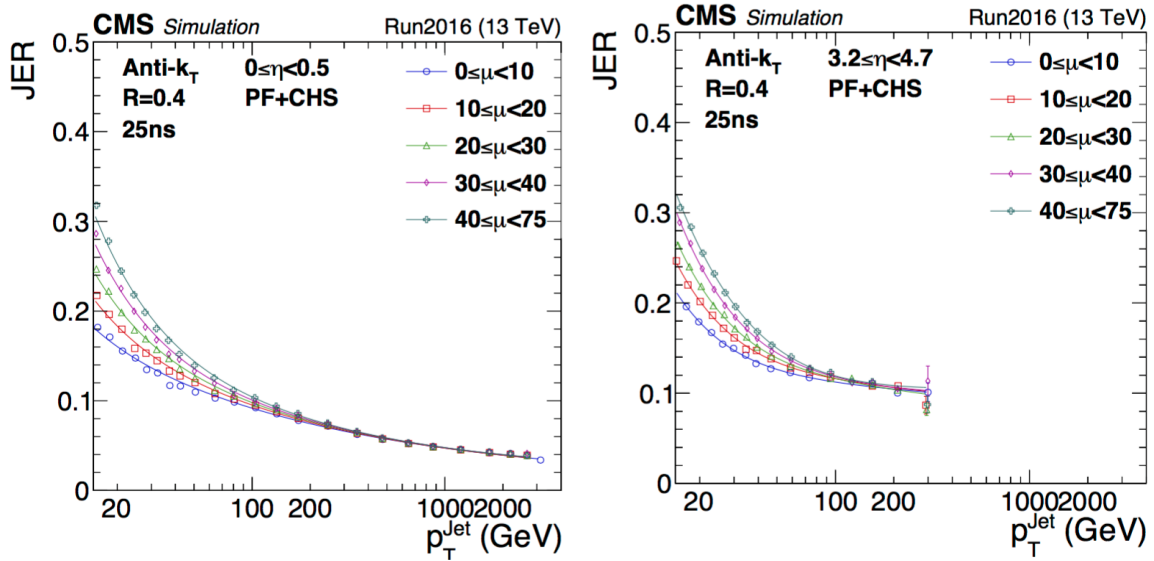


Figure 3.6: Jet energy resolution as function of the jet p_T , as obtained for PFjets with a radius of 0.4, pointing to the barrel, $0.0 < \eta < 0.5$ (left) and to the end-caps, $3.2 < |\eta| < 4.7$ (right), for different PU scenarios [150].

3.6 Boosted topologies and jet substructure

A common signature of searches for new physics is the production of massive new particles that decay into final state objects with very large p_T , exceeding their rest mass. In this new kinematic regime, the resulting hadronic decay products from bosons (W, Z, Higgs boson)

or top quarks can be collimated and fall into a single reconstructed jet, being the opening angle between the decay products inversely proportional to the p_T of the mother particle:

$$\Delta R = \frac{2m}{p_T}. \quad (3.5)$$

In the case of an Higgs boson decaying to a pair of b quarks, if its p_T is lower than 250 GeV, its decay products are well separated and result in 2 jets. However, if the Higgs boson transverse momenta is larger than 250 GeV, both decay products are close to each other giving rise to a single merged jet, referred to as an Higgs jet.

In both analyses presented in this thesis, the separation between signal and multijet background arising from QCD processes plays a pivotal role. To reconstruct boosted objects, classical reconstruction approaches relying on one-to-one jet to parton assignment are not adequate anymore and new dedicated techniques are designed and implemented. Being the mass of the partons close to zero and the mass of the bosons much higher - for the Higgs boson, 125 GeV - it would be natural to think that the jet mass could be a good discriminant. Unfortunately, when a regime of very high energies is reached, the width of typical jets arising in multijet processes become equally large. Moreover, additional particles contribute to the final total jet mass because of diffuse radiation caused by underlying events and PU. Hence, dedicated grooming algorithms are designed and implemented to improve mass resolution and increase the discrimination between signal and multijet background.

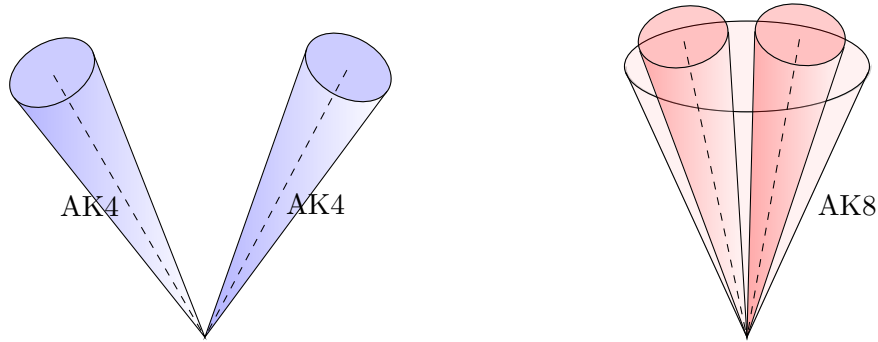


Figure 3.7: Schematic illustration for regular hadronic decays (left) and boosted topologies (right). If the decaying boson or top quarks has a relatively low p_T , its decay products are well separated and clustered into distinguishable AK4 jets, indicated in blue. On the other hand, if the transverse momentum of the mother particle is higher, the opening angle between the resulting particles is smaller and the vector boson decay products are merged into one single large cone AK8 jet, shown in pink.

3.6.1 Jet grooming strategies

New sophisticated tools, called grooming algorithms, have been developed to better understand whether the jet is originated from the decay of a massive particle or from a light-flavor quark or gluon. With these techniques, particles coming from the massive object itself can be disentangled from those coming from boosted light-flavor quarks from the hard scattering, from initial state radiation, underlying event or PU. These algorithms improve the mass reconstruction, leading to a better mass resolution of large-radius jet can be achieved.

Grooming algorithms mainly rely on the removal of the softest parts of the jet to resolve its bare mass. A deeper description of the methods used in CMS and in this thesis follows.

Pruning

The pruning algorithm [151, 152] involves the reclustering of the jet under scrutiny removing soft particles but also particles emitted at a wide-angle.

Pruning proceeds by reclustering the jet and, at every clustering step involving objects i and j , it verifies if:

$$\frac{\min(p_{T,i}, p_{T,j})}{p_{T,comb}} > z_{cut} \quad \text{and} \quad \Delta R_{i,j} > D_{cut} = \frac{2r_{cut}m_{jet}}{p_T}, \quad (3.6)$$

where $p_{T,comb}$ is the combined transverse momentum of the two jets i and j , and r_{cut} and z_{cut} are parameters of the algorithm. The object with the lowest p_T is then removed if its fractional transverse momentum is lower than z_{cut} or if the angle relative to the axis of the recombination of the two objects is bigger than D_{cut} . The default parameter setting is CMS is $r_{cut} = 0.5$ and $z_{cut} = 0.1$, which allows the most powerful signal to background discrimination.

In Figure 3.8 the distribution of the pruned mass of a candidate Higgs boson is presented. Its discriminating power can be assessed by observing the different distributions between signal, indicated with a blue line (with a peak at around the nominal value of the Higgs boson) and backgrounds, red line.

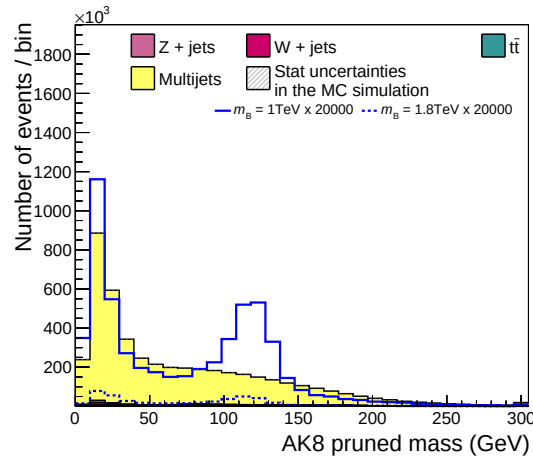


Figure 3.8: Distribution of the pruned mass, comparing the SM background distribution, shown with the solid filled histograms, and the single VLQ B signal.

Mass drop tagger and Soft drop

The mass drop tagger (MDT) [153] relies on the assumption that, as opposed to the bosons, which are formed by two subjets, QCD jets are constituted by continuous soft radiation, being the mass of the single subjet smaller than their combined mass. This means that the mass of the heaviest subjet of QCD jets and the mass of the jet itself should be comparable.

The MDT algorithm starts with the reclustering of a jet j , which is afterward declustered again. Two subjets, labeled s_1 and s_2 , with $m_{s_1} > m_{s_2}$, are identified. The algorithm checks two conditions related to the amount of mass dropping occurring during the declustering and to the symmetry of the splitting. If the mass shift is not very significant and the splitting is not too asymmetric, the jet j is identified as the tagged jet. Otherwise, the jet is considered as the subjet s_1 and the algorithm starts again.

If instead of the bare mass m , the transverse mass $m^2 + p_T^2$ is considered, the algorithm is called modified mass drop tagger (mMDT) [151]. As opposed to the MDT algorithm, this technique retains the most energetic jet instead of the heaviest. A generalization of the mMDT strategy is the *soft drop declustering* (or simply *soft drop*), and this is the default algorithm used in CMS among this set of techniques. The soft drop declustering [154] recursively removes soft wide-angle radiation from a jet to mitigate the effects of contamination from initial state radiation, underlying event, and multiple hadron scattering. Given a jet of radius R_0 with only two constituents, the soft drop procedure removes the softer constituent unless

$$\frac{\min(p_{T,1}, p_{T,2})}{p_{T,1} + p_{T,2}} > z_{cut} \frac{\Delta R_{12}^\beta}{R_0}, \quad (3.7)$$

where $p_{T,i}$ are the transverse momenta of the constituents, ΔR_{12} is their distance in the rapidity-azimuth plane, z_{cut} is the soft drop threshold, and β is an angular exponent. By construction, the soft drop condition fails for wide-angle soft radiation. The degree of jet grooming is controlled by z_{cut} and β and the default parameters employed for the physics analyses in CMS are $\beta = 0$ and $z_{cut} = 0.1$. This configuration provides the best signal to background discrimination while maintaining an excellent mass resolution.

In Figure 3.9 the soft drop mass distribution of a large cone jet is presented, as opposed to the one of the quark and gluon jet.

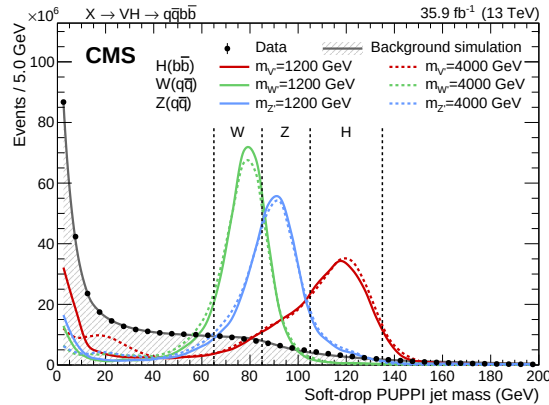


Figure 3.9: Distribution of the soft-drop jet mass, comparing the SM background distribution, shown with the solid filled histograms, and boosted bosons signals [155].

3.6.2 N-subjettiness

An additional feature of the jet substructure is the N-subjettiness [156], which is an important handle while distinguishing vector bosons or top quarks jets from quark/gluon jets. While

a jet arising from the hadronization of a quark or a gluon is originated from a single parton and is composed of several splittings with large angles and asymmetric directions, a boson or top quark jet consists of two or three subjets with high p_T .

The N-subjetiness algorithm works as follows. First, a candidate large cone jet is reconstructed and the N candidate subjets are identified. Once these candidates subjets are recognized, the variable τ_N is computed as:

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min(\Delta R_{1,k}, \Delta R_{2,k}, \dots, \Delta R_{N,k}), \quad (3.8)$$

where k runs over the constituent particles in a given jet, $p_{T,k}$ are their transverse momenta, and $\Delta R_{J,k}$ is the distance in the rapidity-azimuth plane between a candidate subjet J and a constituent particle k . The normalization factor d_0 is taken as:

$$d_0 = \sum_k p_{T,k} R_0, \quad (3.9)$$

being R_0 the characteristic jet radius used in the original jet clustering algorithm.

In other words, the N-subjetiness quantifies to what degree a particular jet can be regarded as composed of N subjets. Small values of τ_N indicate a jet with all the radiation aligned with the candidate subjet direction and therefore with N or fewer subjets. On the other hand, if larger values of τ_N are returned, it means that the jet has a large fraction of its energy distributed away from the candidate subjet direction and therefore has at least $N + 1$ subjets.

It is found that the ratios between the N-subjetinesses are more efficient than the single values in discriminating boson or top quarks jets from the one arising in QCD processes. In general, the ratio of signal jets is lower than the one resulting from quarks and gluons jets. In CMS, the default recommendation is to use the τ_2/τ_1 and τ_3/τ_2 N-subjetiness ratios, and an example of their distributions is illustrated in Figure 3.10.

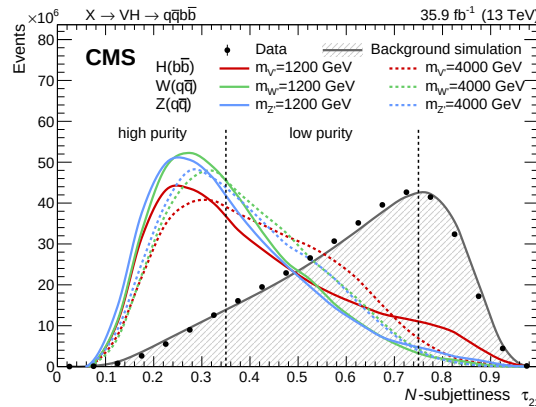


Figure 3.10: Distribution of the τ_2/τ_1 N-subjetiness ratio, comparing the SM background distribution, shown with the solid filled histograms, and boosted bosons signals [155].

3.7 Jets from b quarks

The identification of b-jets is a crucial element for a large variety of BSM searches and SM measurements, including the search for VLQs presented in this thesis. The ability to accurately identify b-jets is fundamental to reduce the otherwise overwhelming background from processes involving jets coming from the hadronization of light flavor quarks and gluons. The hard fragmentation, long lifetime and high mass of B hadrons, and the relatively high fraction of semi-leptonic decays, distinguish these jets from those originating from gluons, light quarks and also from c quarks [157].

The lifetime of hadrons containing b quarks is of the order of 1.5 ps, while the lifetime of C-hadrons is 1 ps or less. Thus, B-hadrons show typical displacements of a few mm to 1 cm, depending on their p_T . This displacement of the decay vertex leads to a secondary vertex (SV) that can be reconstructed far from the primary one, as depicted in Figure 3.11. The vector pointing from the primary vertex to the point of closest approach is referred to as the impact parameter vector. Also, b and c quarks have larger masses and harder fragmentations compared to the light-flavor quarks and gluons. Thus, the decay products of the heavy-flavor hadron tend to have a larger p_T than the other jet constituents. Finally, another characterizing property of the B-hadrons is the high fraction (20%) of semi-leptonic decay. This feature will be exploited for heavy-flavor jet identification and for measuring the performance in data.

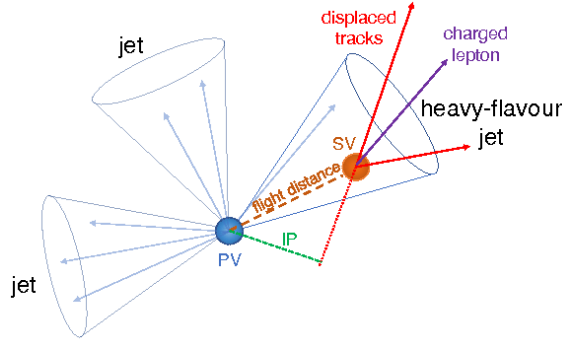


Figure 3.11: Illustration of a heavy-flavour jet with a secondary vertex (SV) from the decay of a B-hadron resulting in charged-particle tracks that are displaced with respect to the primary interaction vertex (PV), and hence with a large impact parameter (IP) value [157].

A dedicated algorithm, the Combined Secondary Vertex v2 (CSVv2) algorithm [157], exploits secondary vertex and track-based lifetime information to identify a b-jet. It is an updated version of the CSV algorithm used in Run 1, combining several input variables through a neural network instead of a likelihood ratio. The algorithm provides a continuous output between 0 and 1, assigning higher values to jets more likely to originate from a b quark, as shown in Figure 3.12. The performance of the tagging algorithm is evaluated in terms of efficiency in the identification of b-jets and misidentification probability of jets from light quarks/gluons as b-jets, as shown in Figure 3.12. Three working points are defined, depending on the misidentification probability, as described in Table 3.4.

Vertex reconstruction and identification, and therefore c-tagging performance, are driven by the pixel detector. Optimal performances of the tracker are fundamental to

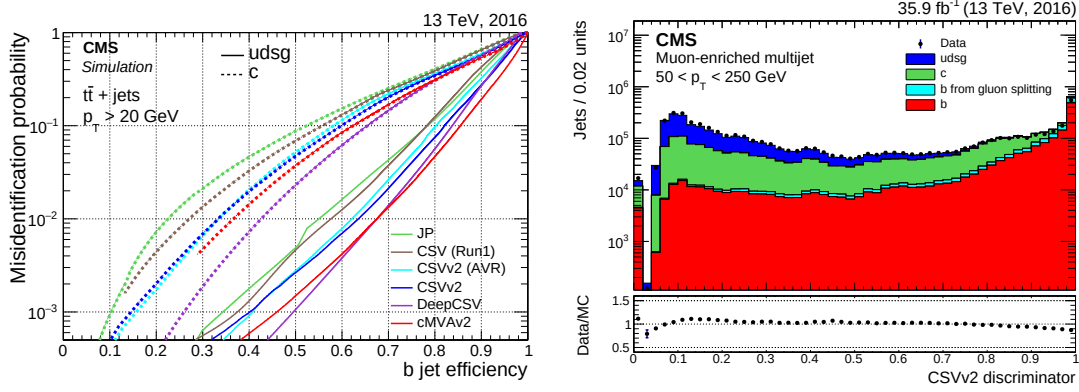


Figure 3.12: Misidentification probability for c and light-flavour jets versus b jet identification efficiency for various b tagging algorithms applied to jets in $t\bar{t}$ events (left). Data-MC comparison for the CSVv2 discriminators for jets in the muon-enriched multijet sample (right). The simulated contributions of each jet flavour are shown with different colours. The total number of entries in the simulation is normalized to the number of observed entries in data. The first and last bin of each histogram contain the underflow and overflow entries, respectively [157].

	Loose	Medium	Tight
CSVv2 value	0.460	0.800	0.935
misidentification probability	10%	1%	0.1%

Table 3.4: List of CSVv2 working points.

achieve good physics results. Thus, in the context of this Higgs, I also focused on the measurement and monitoring of the pixel resolution, which is more extensively described in Appendix A.

3.8 Missing transverse energy

DM particles are invisible to the detector and, similarly to neutrinos, pass through the CMS detector without interacting with it. If such particles are produced along with other visible particles their presence can be inferred through an imbalance in the visible momentum in the plane perpendicular to the beam direction. This imbalance is quantified by the missing transverse momentum, $p_T^{\vec{miss}}$, which is defined as the negative sum of the transverse momenta of all PF candidates reconstructed in the event:

$$p_T^{\vec{miss}} = - \sum_{\text{PF candidates}} p_T^{\vec{p}_T}, \quad (3.10)$$

with $p_T^{\vec{miss}}$ being the magnitude of the vector.

The non-linearity of the response of calorimeters to hadronic particles, the minimum energy thresholds in the calorimeters, and the p_T thresholds and inefficiency in the tracker, lead to an inaccurate estimation of $p_T^{\vec{miss}}$. Hence, a dedicated calibration is performed to correct for this effect. An improved estimation of the $p_T^{\vec{miss}}$ is obtained by correcting the p_T of the jets to the particle level using the jet energy corrections (as described in [158])

and then propagating these corrections to p_T^{miss} . The following factorization is deployed:

$$\text{Type-I } p_T^{miss} = p_T^{miss} - \sum_{\text{jets}} (p_{T,\text{jet}}^{\text{corr}} - p_{T,\text{jet}}), \quad (3.11)$$

where the sum is over jets with p_T above an optimized threshold of 15 GeV. Further corrections are applied to remove the overlap of jets with electrons and photons, and depending on the reconstructed muons in the events.

The reconstruction of p_T^{miss} depends on the reconstruction and measurement of muons, electrons, photons, taus, and jets. An uncertainty on p_T^{miss} is assessed by factorizing the p_T^{miss} to the aforementioned physics objects and varying each object within its scale and resolution uncertainties. The leading factors are the jet energy scale and resolution. The final uncertainty is evaluated by comparing the recalculated p_T^{miss} to its nominal value.

3.8.1 Performance of the PF p_T^{miss} reconstruction algorithm

The reconstruction of p_T^{miss} is affected by the resolution and mis-measurements of the PF candidates, and by detector artifacts. Besides, reconstruction performance depends on PU interactions as well. It is important to keep PU effect on p_T^{miss} under control to get good p_T^{miss} reconstruction.

The performance of the p_T^{miss} reconstruction is evaluated in terms of scale and resolution. Events with no genuine missing transverse momentum, such as those with a leptonically decaying Z boson or an isolated photon, are used to evaluate the p_T^{miss} performance. In these events, the momentum of the vector boson is opposed to the vector sum p_T of all PF candidates except the vector boson (hadronic recoil system). The vector boson momentum in the transverse plane is labeled with $\vec{q}_T$, while the transverse momentum of the hadronic recoil is denoted with $\vec{u}_T$, as depicted in Figure 3.13.

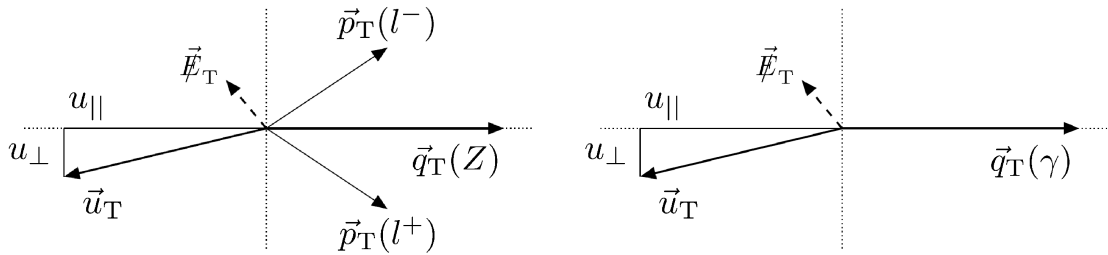


Figure 3.13: Sketch of the Z (left) and γ (right) event kinematics in the transverse plane. The vector $\vec{u}_T$ shows the vectorial sum of all particles reconstructed in the event except for the two leptons from the Z decay (left) or the photon (right). [159]

In Z/ γ events, the signal topology provides a unique event axis and a precise momentum scale. The components of the hadronic recoil, parallel ($u_{\parallel}$) and perpendicular ($u_{\perp}$) to the boson axis, are used to study the p_T^{miss} scale and resolution: the mean of the $u_{\parallel} + q_T$

distribution is used to estimate the scale, whereas the root-mean-square of the $u_{\parallel} + q_T$ and $u_{\perp}$ distributions are used to estimate the resolution $\sigma(u_{\parallel})$ and $\sigma(u_{\perp})$, respectively. The resolutions of p_T^{miss} , in terms of parallel and perpendicular components of the hadronic recoil, are presented in Figure 3.14.

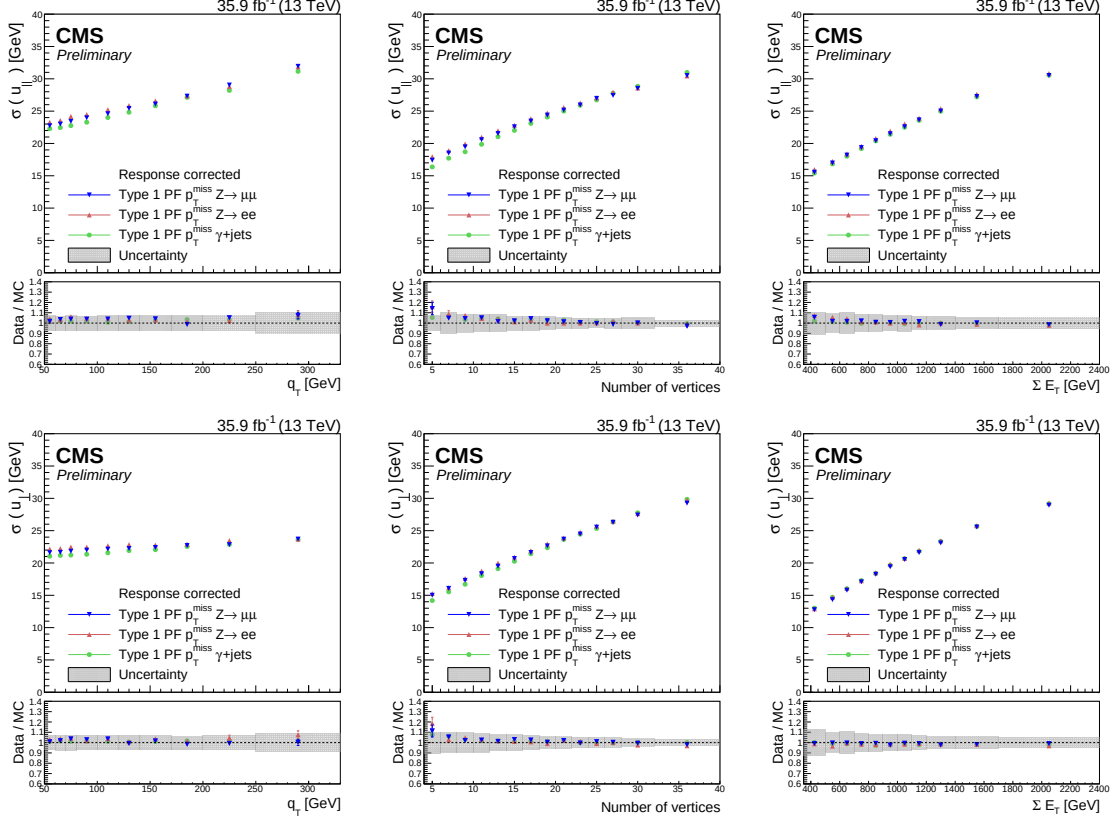


Figure 3.14: Resolution of the parallel and perpendicular components of the hadronic recoil $\vec{u}_T$ as a function of $\vec{q}_T$ (left), the reconstructed primary vertices (center), and the scalar p_T sum of all PF candidates (right), in $Z \rightarrow \mu^+\mu^-$, $Z \rightarrow e^+e^-$ and γ events. In each plot, the upper panel shows the resolution in data, whereas the lower panel shows the ratio of data to simulation. The gray band shows to the systematic uncertainty [159].

It can be observed that the relative resolution, in both directions, improves as a function of $\vec{q}_T$, because of the improved resolution in the calorimeter. The effect measured in the perpendicular direction is smaller in magnitude because of the isotropic nature of energy fluctuation in the detector noise and underlying event. Furthermore, a strong dependence on the number of PU vertices is observed, being no PU mitigation technique employed in the p_T^{miss} reconstruction. Finally, the relative missing transverse momentum resolution improves the scalar p_T sum of all PF candidates ($\sum E_T$), driven by the amount of the activity in the calorimeters.

Chapter 4

Search for single production of vector-like quarks

“ He who does not expect will not find out the unexpected, for it is trackless and unexplored ”

Eracrito, 4th-5th century BC

This chapter describes the first search for the electroweak production of a VLQ B, decaying to a b quark and a Higgs boson at the LHC. The leading order Feynman diagram for this process is:

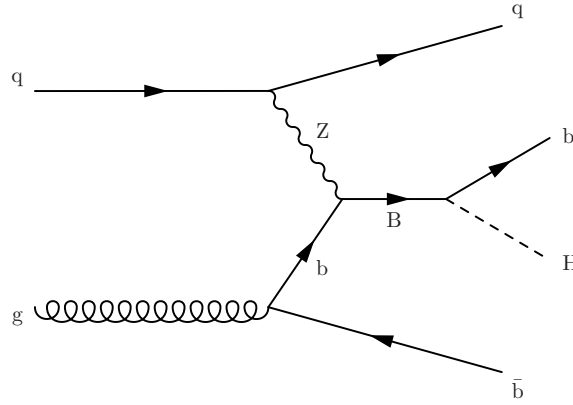


Figure 4.1: The leading-order Feynman diagram for the production of a single vector-like B quark in association with a b quark and light-flavour quark, and its decay to a Higgs boson and a b quark.

In this signature, the Higgs boson decays to a pair of b-quarks. This is an electroweak production characterized by the presence of several jets and no leptons in the final state. The B quark is singly produced, in association with a b and a light-flavour quark, which is typically emitted in the forward pseudorapidity region of the CMS detector. The B quark is expected to be very massive ($\mathcal{O}(\text{TeV})$), and thus its decay products are boosted. Therefore the Higgs, along with the two b quarks arising from its decay, is reconstructed as a unique large-cone jet.

In Section 4.1 the properties of the VLQ B signal and of the background processes are described. We implemented an optimized event selection to reduce the background contamination, as detailed in Section 4.2. To describe the overwhelming multijets background the data driven method that is presented in Section 4.3 is employed. The evaluation of the sources of systematic uncertainties and their effect on the signal extraction are discussed in Section 4.4. Section 4.5.1 presents the theoretical models developed to describe the characterizing process of this search and the results interpretation are discussed. Finally, an insight on the signal extraction, along with a description of the observed data and the expected backgrounds in the signal region is presented. A brief overview of the statistical frameworks used to extract the results is given in Appendix C.

4.1 Data, modeling and simulation

4.1.1 Data

Proton-proton collisions delivered by the LHC at a centre of mass energy of 13 TeV used in this search were collected by CMS during 2016. Several periods of the LHC operations are used and the list of the primary datasets employed, along with the corresponding certified integrated luminosity, are shown in Table 4.1. The totality of data collected corresponds to an integrated luminosity of 35.9 fb^{-1} .

Datasets	Integrated luminosity (pb^{-1})
Run B	5788
Run C	2573
Run D	4248
Run E	4009
Run F	3102
Run G	7540
Run H (v2)	8391
Run H (v3)	221

Table 4.1: List of primary datasets of pp collision data produced at $\sqrt{s} = 13 \text{ TeV}$ and collected by CMS in 2016, employed in the analysis.

4.1.2 Vector-like quark B simulation

Vector-like quarks B are simulated at the leading order (LO), modelled using the universal FEYNRULES output [160, 161] and the generator MADGRAPH_AMC@ NLO 2.2.2 [162] interfaced to PYTHIA 8.2 [163] for the simulation of the parton shower. Several mass hypotheses representing the narrow-width categories are considered for signals with B mass between 700 and 1800 GeV, in steps of 100 GeV for a total decay width of 1 GeV. The large width scenarios, (10%, 20%, or 30% of the mass hypothesis) are investigated using signal events generated in the same mass window.

All the B quarks are generated assuming left-handed chirality and, since the coupling chirality can potentially affect the kinematic distributions, we explicitly checked and found

that these effects are negligible. In Figure 4.2 and 4.3 the parton-level distributions for the p_T and η , respectively, of leading jet, leading and sub-leading b jet, and the Higgs are shown. No sizable differences between dominant left- and right-chiralities are observed.

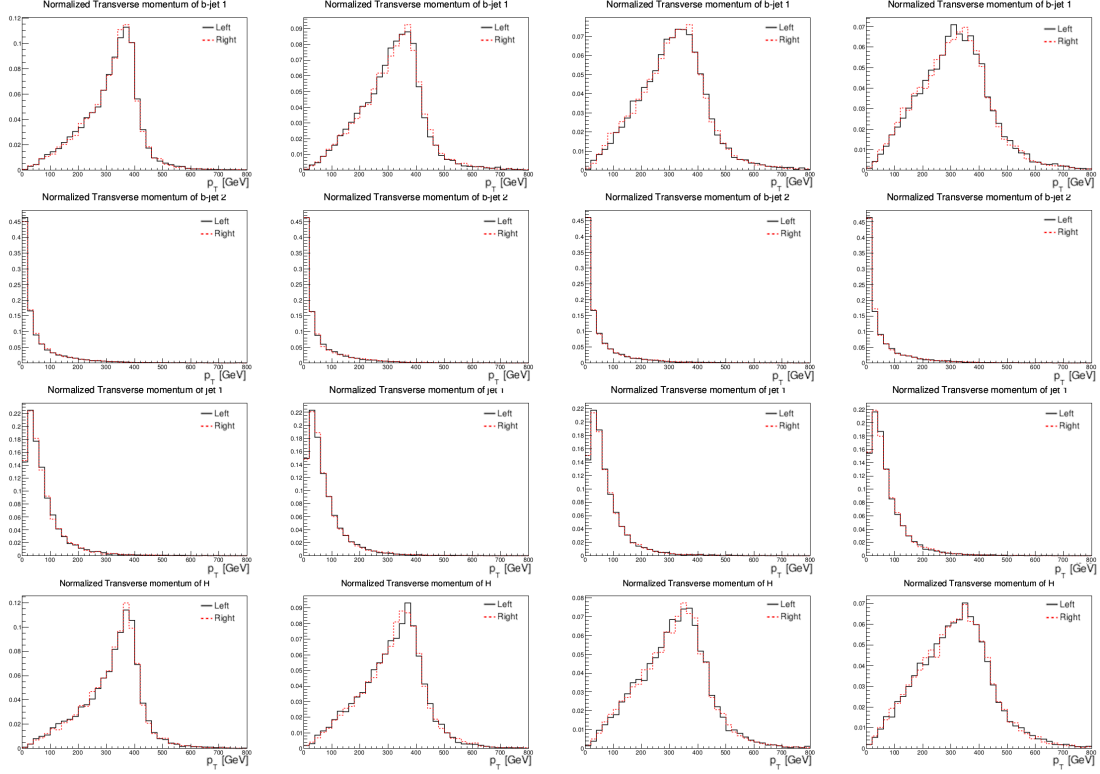


Figure 4.2: Parton-level distributions for the transverse momentum of the leading b jet (first row), second b jet (second row), leading jet (third row) and the Higgs boson (fourth row). Different columns correspond to Γ_B/M_B ratios of 1, 10, 20 and 30% with $M_B = 800$ GeV [164].

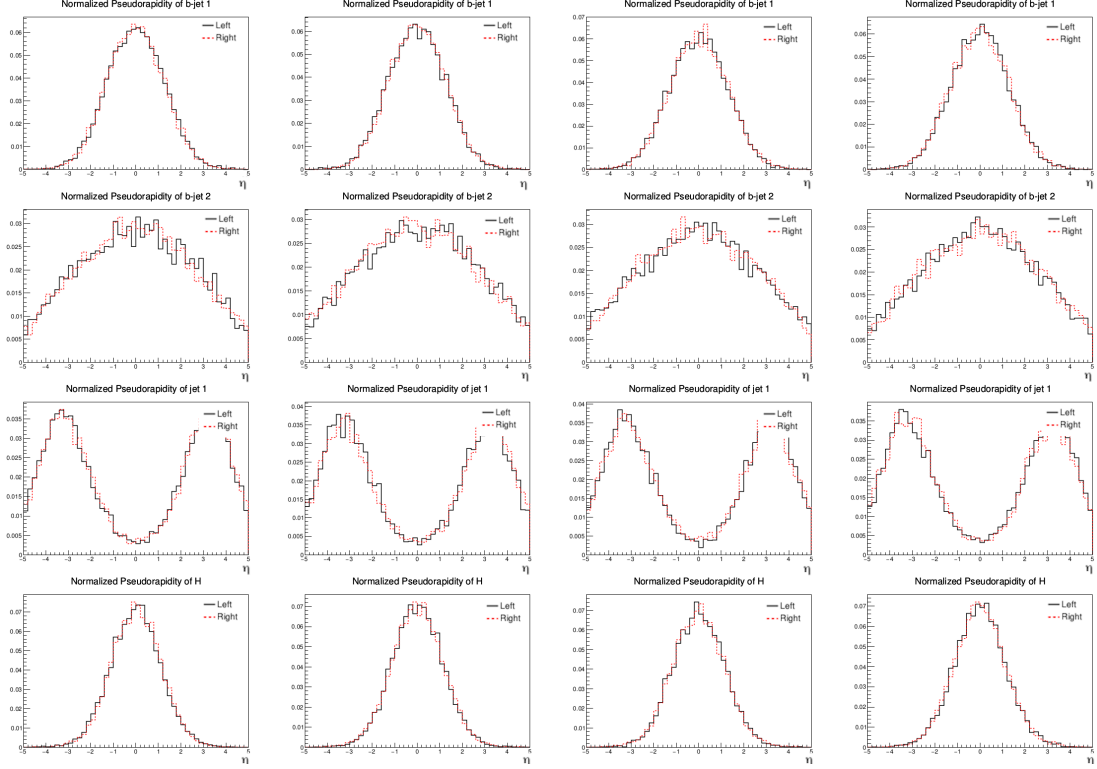


Figure 4.3: Parton-level distributions for the pseudorapidity of the leading b jet (first row), second b jet (second row), leading jet (third row) and the Higgs boson (fourth row). Different columns correspond to Γ_B/M_B ratios of 1, 10, 20 and 30% with $M_B = 800$ GeV [164].

4.1.3 Simulation of Standard Model backgrounds

A final state with large jet multiplicity can result from SM processes. A good simulation is fundamental for describing and estimating these backgrounds. The main background for this fully hadronic search is the production of multijet events from QCD production. The pair production of top quarks is also considered, as well as minor backgrounds such as the electroweak processes giving Z or W bosons + jets. Minor contributions such as the production of single top and diboson processes have a negligible impact and therefore are not included. At a center-of-mass energy of 13 TeV, the production cross-sections for QCD multijet, Z/W + jets and $t\bar{t}$ processes are of the order of $\mathcal{O}(10^6)$ pb, $\mathcal{O}(10^3)$ pb, $\mathcal{O}(10^1)$ pb and $\mathcal{O}(10^2)$ pb, respectively.

All processes simulated using LO (NLO) calculations use the LO (NLO [165]) sets of parton distribution functions (PDFs) for the hard process calculation. The interaction of the signal and background events with the CMS detector is done using GEANT 4 [166]. Details of the simulated processes are summarized in Table 4.2.

The main features of the considered backgrounds are briefly presented in the following sections. The multijet QCD is simulated at the leading order using MADGRAPH_AMC@ NLO 2.2.2 interfaced to PYTHIA 8.2 with the CUETP8M1 [167, 168] underlying event tune for parton-shower simulation and evolution.

Top quark pair production, labelled as $t\bar{t}$, is generated using POWHEG v2 [169–172] at the next-to-leading order (NLO), and interfaced to PYTHIA 8.2. The mass of the top

quark is set to 172.5 GeV. The cross-section of the process is calculated at the next-to-next-to-leading order (NNLO) in perturbative QCD using a next-to-next-to-leading-logarithmic (NNLL) soft-gluon approximation (NNLO+NNLL) in the TOP++ 2.0 [173] program. Figure 4.4 shows representative Feynman diagrams for the leading-order production mechanisms for top quark pairs at a hadronic collider.

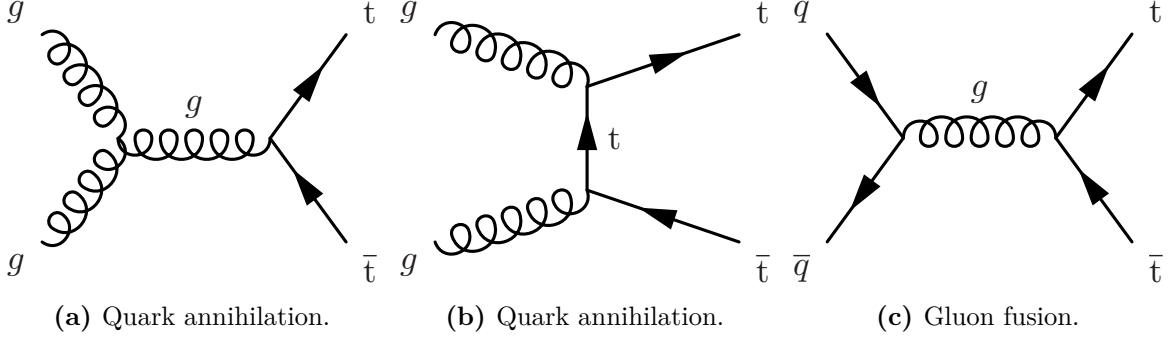


Figure 4.4: Feynman diagrams for the leading-order production mechanisms for top quark pairs at a pp collider.

As the multijet background, the vector boson (Z,W) plus jets contributions are generated using MADGRAPH_AMC@NLO 2.2.2 interfaced to PYTHIA 8.2 using the CUETP8M1 tune. The cross-sections for Z+jets and W+jets processes are calculated at NNLO using the FEWZ MC program [174]. In Figure 4.5 representative Feynman diagram examples for the production of vector boson plus jets processes are presented.

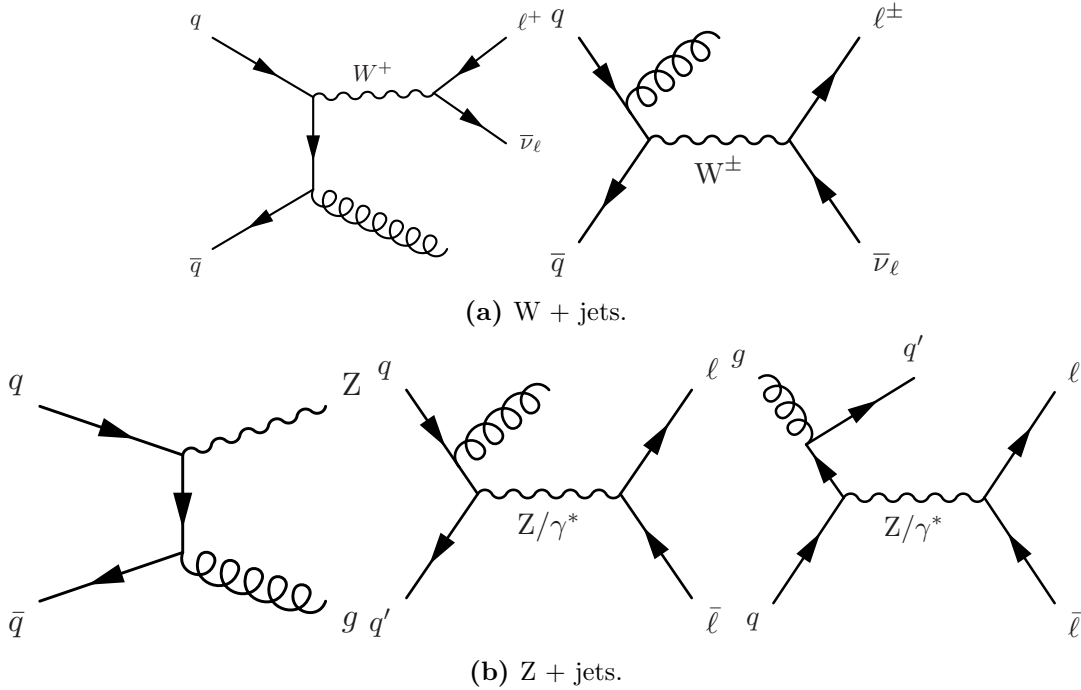


Figure 4.5: Some representative Feynman diagram examples for the production of vector boson plus jets.

Process	σ (pb)	Order	Matrix Element generator	Showering
Multijet ($300 \text{ GeV} < H_T < 500 \text{ GeV}$)	347700	LO	MADGRAPH_AMC@NLO 2.2.2	PYTHIA 8.2
Multijet ($500 \text{ GeV} < H_T < 700 \text{ GeV}$)	32100	LO	MADGRAPH_AMC@NLO 2.2.2	PYTHIA 8.2
Multijet ($700 \text{ GeV} < H_T < 1000 \text{ GeV}$)	6831	LO	MADGRAPH_AMC@NLO 2.2.2	PYTHIA 8.2
Multijet ($1000 \text{ GeV} < H_T < 1500 \text{ GeV}$)	1207	LO	MADGRAPH_AMC@NLO 2.2.2	PYTHIA 8.2
Multijet ($1500 \text{ GeV} < H_T < 2000 \text{ GeV}$)	119.9	LO	MADGRAPH_AMC@NLO 2.2.2	PYTHIA 8.2
Multijet ($2000 \text{ GeV} < H_T < \infty$)	25.24	LO	MADGRAPH_AMC@NLO 2.2.2	PYTHIA 8.2
$t\bar{t}$	831.76	NLO	POWHEG v2	PYTHIA 8.2
$W (\rightarrow l\nu) + \text{jets}$ ($600 \text{ GeV} < H_T < 800 \text{ GeV}$)	12.05	LO	MADGRAPH_AMC@NLO 2.2.2	PYTHIA 8.2
$W (\rightarrow l\nu) + \text{jets}$ ($800 \text{ GeV} < H_T < 1200 \text{ GeV}$)	5.501	LO	MADGRAPH_AMC@NLO 2.2.2	PYTHIA 8.2
$W (\rightarrow l\nu) + \text{jets}$ ($1200 \text{ GeV} < H_T < 2500 \text{ GeV}$)	1.329	LO	MADGRAPH_AMC@NLO 2.2.2	PYTHIA 8.2
$W (\rightarrow l\nu) + \text{jets}$ ($2500 \text{ GeV} < H_T < \infty$)	0.03216	LO	MADGRAPH_AMC@NLO 2.2.2	PYTHIA 8.2
$Z (\rightarrow \nu\nu) + \text{jets}$ ($600 \text{ GeV} < H_T < 800 \text{ GeV}$)	2.56	LO	MADGRAPH_AMC@NLO 2.2.2	PYTHIA 8.2
$Z (\rightarrow \nu\nu) + \text{jets}$ ($800 \text{ GeV} < H_T < 1200 \text{ GeV}$)	1.18	LO	MADGRAPH_AMC@NLO 2.2.2	PYTHIA 8.2
$Z (\rightarrow \nu\nu) + \text{jets}$ ($1200 \text{ GeV} < H_T < 2500 \text{ GeV}$)	0.29	LO	MADGRAPH_AMC@NLO 2.2.2	PYTHIA 8.2
$Z (\rightarrow \nu\nu) + \text{jets}$ ($2500 \text{ GeV} < H_T < \infty$)	95.14	LO	MADGRAPH_AMC@NLO 2.2.2	PYTHIA 8.2
$Z (\rightarrow qq) + \text{jets}$ ($600 \text{ GeV} < H_T < \infty$)	2.56	LO	MADGRAPH_AMC@NLO 2.2.2	PYTHIA 8.2
$DY (\rightarrow qq) + \text{jets}$ ($180 \text{ GeV} < H_T < \infty$)	1187	LO	MADGRAPH_AMC@NLO 2.2.2	PYTHIA 8.2
EWK $Z (\rightarrow ll) + 2 \text{ jets}$	3.987	LO	MADGRAPH_AMC@NLO 2.2.2	PYTHIA 8.2
EWK $W (\rightarrow l\nu) - 2 \text{ jets}$	20.25	LO	MADGRAPH_AMC@NLO 2.2.2	PYTHIA 8.2
EWK $W (\rightarrow l\nu) + 2 \text{ jets}$	25.62	LO	MADGRAPH_AMC@NLO 2.2.2	PYTHIA 8.2

Table 4.2: Generations details for the simulated processes used in the analysis. The cross-section is reported, and the generator along with showering packages used with the relative calculation orders are also listed.

Pileup simulation

The presence of additional interactions with respect to the primary one is known as pileup (PU) and it results in additional energy deposited in the tracker and in the calorimeters. The CMS simulation of PU does not reproduce the the data, therefore, simulated events are reweighted a posteriori to account for the dependence on the reconstruction efficiency of the number of PU interactions in the collisions. A comparison of the distribution of the number of PU interactions in 2016 data and MC simulation is shown in Figure 4.6. A minimum-bias cross-section of 69.2 mb is assumed to estimate the number of true interactions per bunch crossing in data. A variation of $\pm 4.6\%$ on the cross-section is considered to take into account uncertainties due to PU modeling. The bin-by-bin ratio between the nominal distribution in data and in simulation is computed and is considered as a scale factor that will be applied on an event basis to MC events to correct for the mis-modeling of the PU interactions.

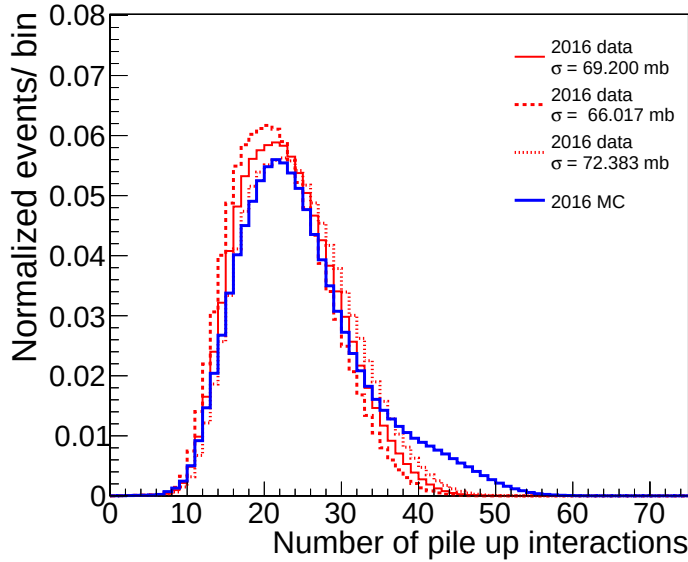


Figure 4.6: Distribution of the number of pileup interactions per bunch-crossing in MC simulation (blue) and data (red) considering a minimum bias cross-section of 69.2 mb. Variations to data distributions are obtained by varying by a factor $\pm 4.6\%$ the cross-section values and the shifted distributions are shown as well, labeled with dashed lines.

To check the efficacy of this correction, a comparison of the number of reconstructed vertices in data and MC simulation is performed before and after the reweighting as closure test and it is shown in Figure 4.7. A better agreement between data and simulation is observed when the PU weights are applied.

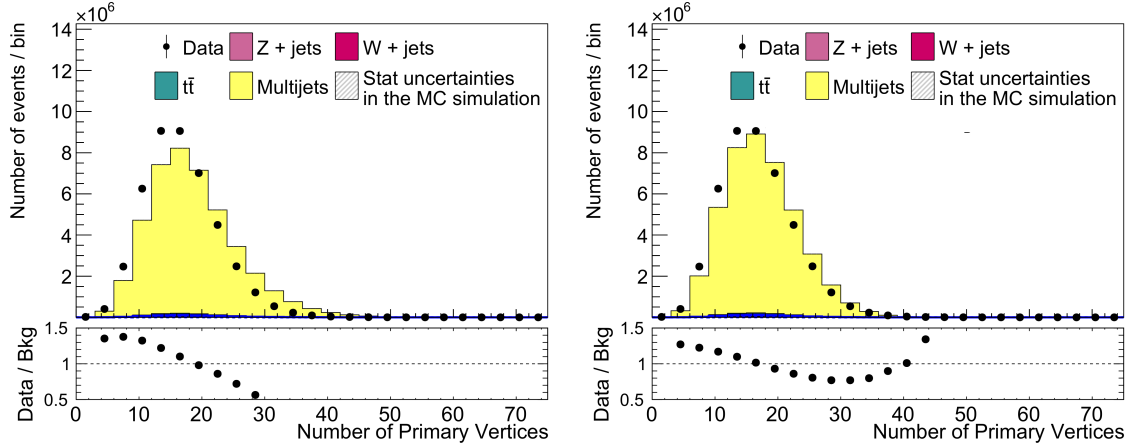


Figure 4.7: Distribution of the number of reconstructed primary vertices in data and in simulation before (left) and after (right) pile up reweighting. Solid filled histograms label SM background distributions: they are summed cumulatively and their sum is rescaled to data. Data are indicated by black dots. In the lower box, the ratio between data and total background distribution is shown.

4.2 Analysis strategy

4.2.1 Trigger

As described in Section 5.2.1, interesting collision events for physics analysis are selected on-line by a two-stage trigger constituted by the Level-1 trigger and the High-Level trigger (HLT). The most suitable triggers for this search are hadronic-based triggers. The hadronic activity (H_T), that quantifies the sum of the jets transverse momenta is the key information exploited in these triggers at HLT level. At this stage, the trigger applied to identify interesting events is based on the request of H_T greater than 900 GeV (labeled as HLT_PFHT900). This trigger is exploited for the initial part of the data-taking (run B to run G), whereas for data collected in late 2016 (run H) an additional trigger algorithm selecting online AK4 jets with $p_T > 450$ GeV was used (labelled as HLT_PFJet450). Both the hadronic activity and the transverse momentum defined at HLT level have been reconstructed using the PF algorithm.

In Table 4.3 the full paths of the HLT triggers used are summarized, along with their respective L1 seeds. The HLT_PFHT900 takes as L1 input an OR between multiple H_T based Level-1 triggers, with several thresholds ranging from 160 to 320. The HLT_PFJet450 uses as seed a logical OR among three L1 seeds requesting at least one jet with different p_T thresholds.

A comparison of the trigger efficiency in data and in simulation is performed by using an orthogonal dataset and requesting at HLT level the presence of a reconstructed muon with $p_T > 50$ GeV. These leptonic triggers are labelled as HLT_Mu50 and HLT_TkMu50. The trigger efficiency is then defined as the fraction of events satisfying the nominal hadronic triggers, as described in Equation 4.1. These studies are performed after

HLT paths	L1 seeds
HLT_PFHT900	L1 HTT160
	L1 HTT200
	L1 HTT220
	L1 HTT240
	L1 HTT255
	L1 HTT270
	L1 HTT280
	L1 HTT300
	L1 HTT320
HLT_PFJet450	L1 SingleJet170
	L1 SingleJet180
	L1 SingleJet200

Table 4.3: HLT trigger paths used in the analysis, along with their corresponding L1 paths. The logical OR of the L1 selections is used as seed for the higher level triggers.

a loose preselection asking for at least 3 AK4 jets and one AK8 jet is applied.

$$\begin{aligned}
 \epsilon_{B-G} &= \frac{\text{HLT_PFHT900} \cap (\text{HLT_Mu50} \cup \text{HLT_TkMu50})}{\text{HLT_Mu50} \cup \text{HLT_TkMu50}} \\
 \epsilon_H &= \frac{(\text{HLT_PFHT900} \cup \text{HLT_PFJet450}) \cap (\text{HLT_Mu50} \cup \text{HLT_TkMu50})}{\text{HLT_Mu50} \cup \text{HLT_TkMu50}}
 \end{aligned} \tag{4.1}$$

As presented in Figure 4.8 the trigger efficiency is fully achieved at $H_T = 1200$ GeV for runs B-G, while the plateau is reached at higher values for run H (99% efficiency at 1350 GeV).

It is of fundamental importance to derive scale factors to account for residual discrepancies between the trigger efficiencies in data and in simulation. In the lower panel of Figure 4.8, bin-by-bin ratio between data and MC are computed and these values are used to correct on event basis all simulated events. The uncertainty on the trigger weights is taken as the statistical error in each H_T bin.

4.2.2 Particle candidates and event selection

An offline selection is applied in addition to the online trigger requests described in Section 4.2.1. To increase the analysis sensitivity the event selection is optimized using as figure of merit $S/\sqrt{S+B}$, where S indicates the signal yield and B the number of background events, and the sensitivity of the analysis in terms of exclusions limits. In the following sections the signal topology of the process is presented, along with important variables exploited in the analysis workflow. The signal final state is then resolved and the B quark mass reconstructed.

The selection of events is exploit the fully hadronic nature of the signature. Selected collision data are required to count at least one large cone AK8 jet, three narrow jets and one medium b-tagged jet. Distributions of multiplicity for these kinds of jets are shown in

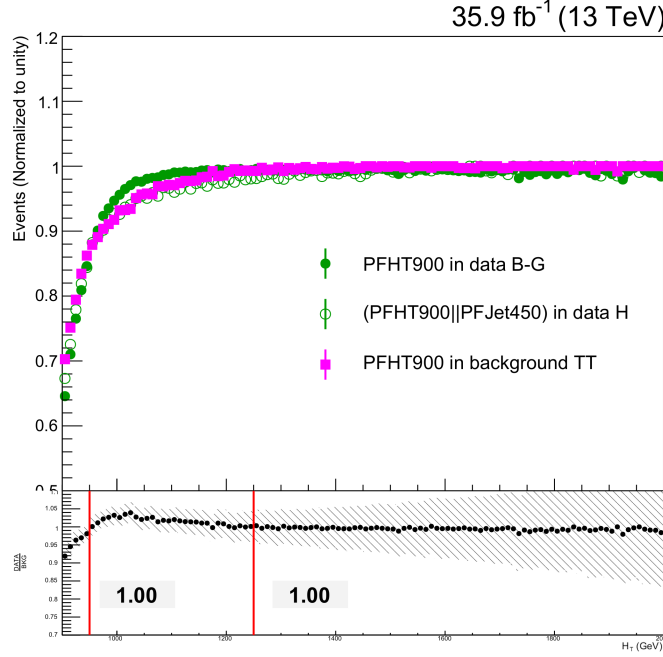


Figure 4.8: Efficiency turn-on curve of the HLT triggers measured in data, both for run B-G and run H, and in simulation, as a function on the offline hadronic activity. Green color shows distribution in data, while simulation is presented with the magenta dots.

Figure 4.9. For the b-tagging a medium working point is used, both for the AK4 jets than for the subjects of the selected AK8 jets.

In addition, a veto is applied to events where one or more leptons are reconstructed, as signal data would not contain leptons. This allows to reduce backgrounds from electroweak processes such as Z and $W + \text{jets}$, or also $t\bar{t}$ events with leptonic decay of top quarks. In this work, global muons are used. They are requested to have $p_T > 3$ GeV and $|\eta| < 2.1$ and, in addition, they have to satisfy a set of criteria, defined as *tight muon selection*, providing an average efficiency of $\sim 70\%$. These identification criteria aim at rejecting muons from the in-flight decays and cosmic muons. In addition to define the quality of muons a dedicated selection is used, based variables such as the χ^2 of the track fit, the number of muon segments in the muon chambers, the number of hits in the fitted tracker track, and on the transverse and longitudinal impact parameter of the muon track with respect to its vertex. A tight working point is chosen for the electrons as well, allowing for a 70% of identification efficiency. Isolated photons are reconstructed as well, seeded from ECAL superclusters with $E_T > 10$ GeV and no link to a GSF track. In addition to the electron identification MVA, quality requirements exploiting informations about the electron superclusters, the tracks, the energy deposits in the calorimeters and other features, are applied.

We will refer to this selection as *preselection* throughout the rest of the text.

Boosted Higgs-boson tagging

Being the Higgs arising from the B quark decay highly boosted, its decay products are collimated and fall into a single AK8 jet, as described in Section 3.6.1. A dedicated algorithm

is developed to identify a boosted H, benefiting from the AK8 pruned mass and on the number of medium b-tagged subjets. The former requirement is optimized to be in the [105, 135] GeV range, while for the latter both the AK8 jets are required to be arising from the hadronization of a b quark. The variables distributions used for the Higgs tagging are presented in Figure 4.10. It is important to notice that the regions we choose to identify the Higgs jet candidate have the best sensitivity, in terms of maximization of signal events and minimization of background contribution.

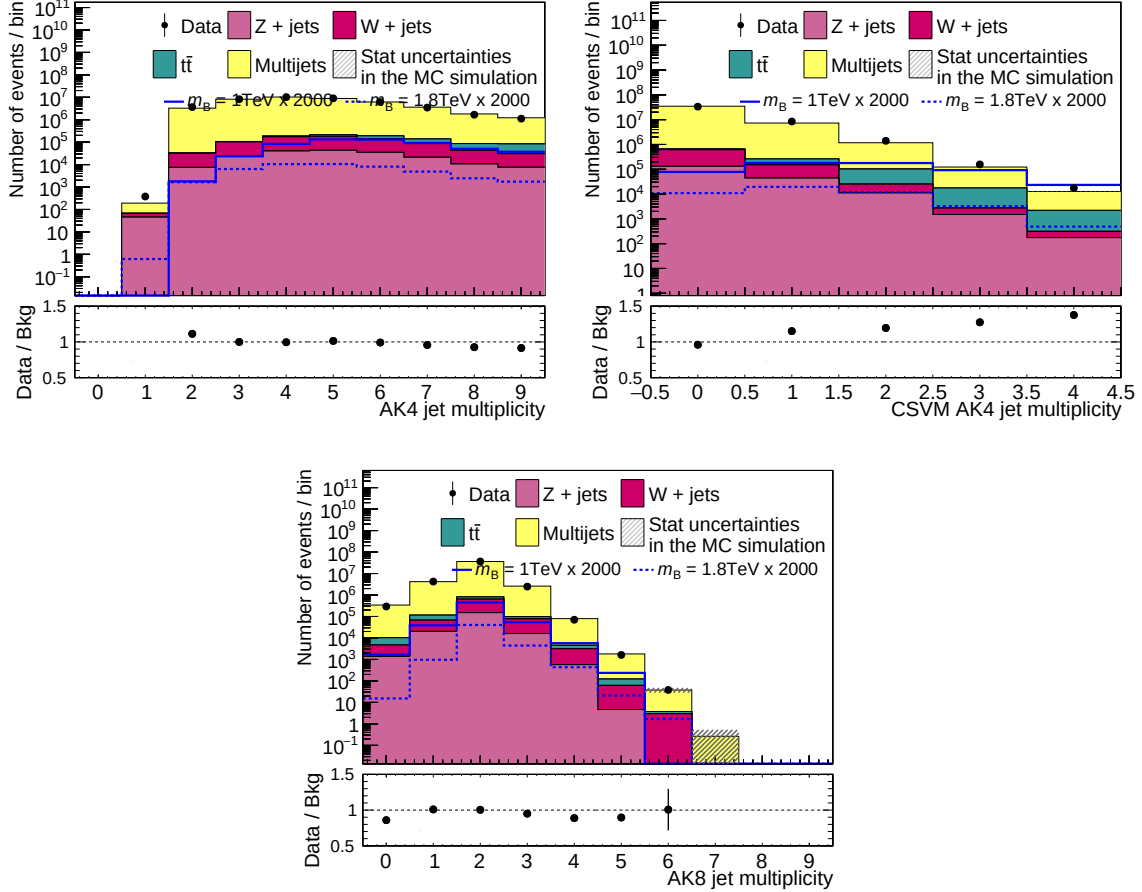


Figure 4.9: Distribution of AK4 jet multiplicity (top left), AK4 medium b-tagged jet multiplicity (top right) and AK8 jet multiplicity (bottom). Solid filled histograms label SM background distributions: they are summed cumulatively and their sum is rescaled to data. The hatched grey region show the statistical uncertainty on MC. Red lines show the distributions for MC signal mass points (including a multiplicative factor 2000), while data are indicated with black dots. In the lower box, the ratio between data and total background distribution is shown. Vertical bars indicates the statistical uncertainty for each bin. In each plot the value of the overflow is added to the last bin of the histogram.

Forward jets

An additional feature of the VLQ single production is the presence of a jet emitted into the forward pseudorapidity region of the detector. This forward jet is resulting from the hadronization of a light flavour quark produced in association with the VLQ B.

This peculiar signature plays an important role in the analysis as it enhances the

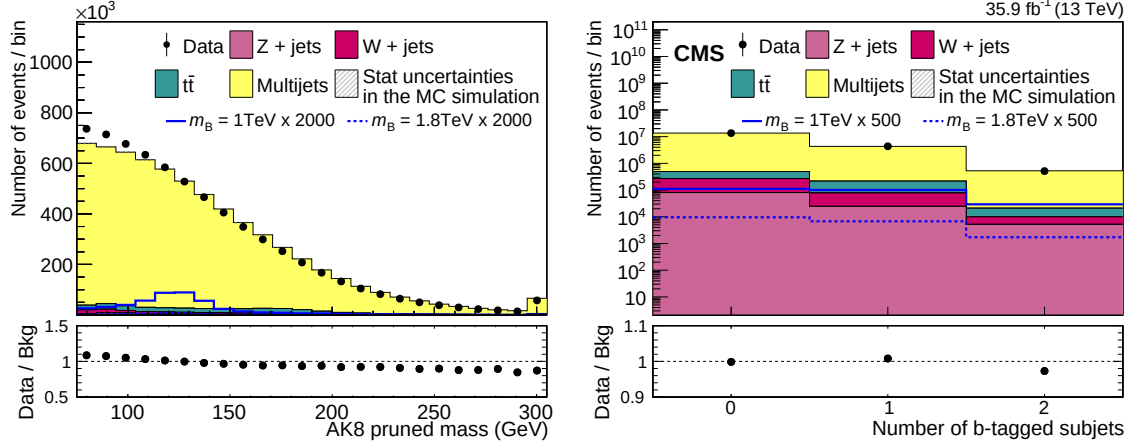


Figure 4.10: Distribution of the variables used for the Higgs tagging of the AK8 jets: pruned mass (left) and medium b-tagged subjects (right). Solid filled histograms label SM background distributions: they are summed cumulatively and their sum is rescaled to data. The hatched grey region show the statistical uncertainty on MC. Red lines show the distributions for MC signal mass points (including a multiplicative factor 2000 or 500), while data are indicated with black dots. In the lower box, the ratio between data and total background distribution is shown. Vertical bars indicates the statistical uncertainty for each bin. In each plot the value of the overflow is added to the last bin of the histogram.

separation between background and signal-like events. This effect can be observed in Figure 4.11, where the forward jets multiplicity distributions is pictured for both background and expected signal. Background spectra peak at 0, while signal-like events have more often one jet in the high pseudorapidity region. The selection of at least one forward jet allows to select a sample more pure in signal. To improve the final sensitivity of our search we are also retaining events where any forward jet is identified, although less pure even if this latter set of events is not as pure as the former. We split the events in two exclusive categories as follows:

- category 0 (labeled as *cat 0*): exactly zero forward jets are reconstructed;
- category 1 (labeled as *cat 1*): at least one forward jet is reconstructed.

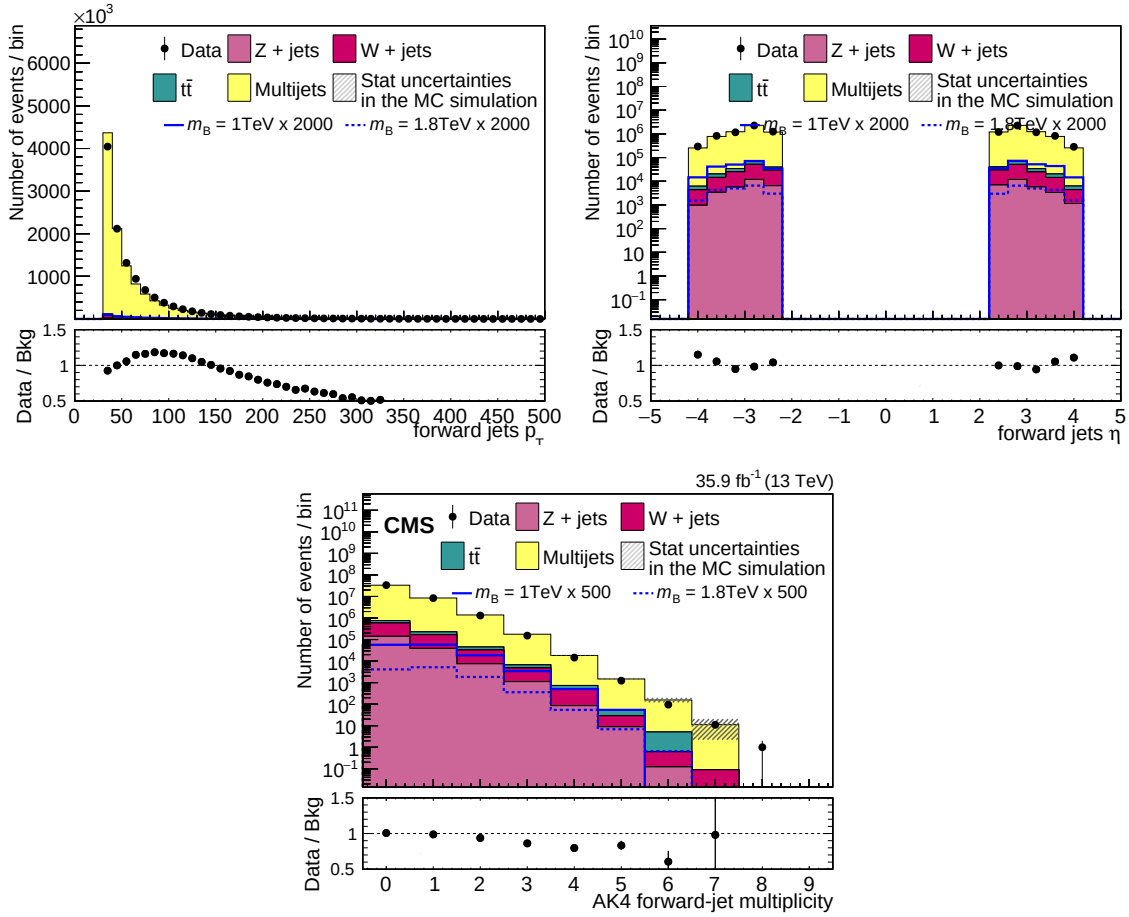


Figure 4.11: Distribution of AK4 forward jet transverse momentum (top left), pseudorapidity (top right) and multiplicity (bottom). Solid filled histograms label SM background distributions: they are summed cumulatively and their sum is rescaled to data. The hatched grey region show the statistical uncertainty on MC. Red lines show the distributions for MC signal mass points (including a multiplicative factor 2000 or 500), while data are indicated with black dots. In the lower box, the ratio between data and total background distribution is shown. Vertical bars indicate the statistical uncertainty for each bin. In each plot the value of the overflow is added to the last bin of the histogram.

Hadronic activity

Being the final state of this analysis solely composed by jets, the hadronic activity, defined as the scalar sum of the jets with $p_T > 50$ GeV, is a fundamental variable to further identify signal-like events.

A baseline selection is already applied at trigger level, where an $H_T > 900$ GeV is requested online. A stronger requirement is then employed within the offline workflow of the analysis. The corresponding threshold is selected driven by upper limits on signal cross-section and the resulting optimal value is also confirmed in terms of analysis sensitivity. It was found that the use of a tight selection improved the potential of the analysis for low masses values of the VLQ B, but resulted also in a degradation of the performances for higher mass hypotheses. Therefore two different H_T thresholds are used: a looser one, with nominal value of 950 GeV, used for $800 \leq m_B < 1500$ GeV and a tighter one, used for $m_B \geq 1500$ GeV. The analysis is therefore split in two parts and we will refer to them as *low mass analysis* (LM) and *high mass analysis* (HM). The hadronic activity spectrum, at preselection level, is presented in Figure 4.12. The disagreement between data and simulation at large values of H_T is caused by mistuning in the parton shower of the multijet MC sample. This effect, however, does not affect the analysis, as the background contribution in the signal region is estimated from data.

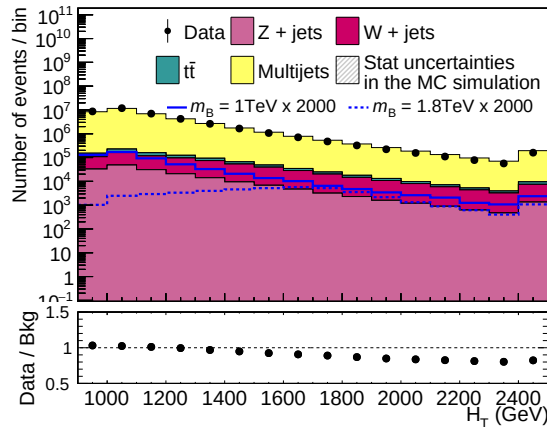


Figure 4.12: Hadronic activity, after the full selection is applied. Solid filled histograms indicate SM background distributions: they are summed cumulatively and their sum is rescaled to data. Red lines show the distributions for MC signal mass points (including a multiplicative factor 2000 or 500), while data are indicated with black dots. In the lower box, the ratio between data and total background distribution is shown. Vertical bars indicates the statistical uncertainty for each bin. In each plot the value of the overflow is added to the last bin of the histogram.

Vector-like quark B reconstruction

Once the baseline selection has been optimized and implemented, and the final state resolved, the strategy of this analysis relies on the reconstruction of VLQ B. We exploit the resonant character of the $B \rightarrow bH \rightarrow bbb\bar{b}$ channel by using the Higgs jet candidate and the b jet arising from B decay. This variable will be exploited to investigate eventual discrepancies between the data and the SM background and for the final statistical analysis.

From studies performed at MC truth-level, we observed that the b quarks produced from the Higgs boson decay are, as expected, much closer in terms of ΔR to the Higgs than the other quarks populating the final state, as shown in Figure 4.13. The remaining ambiguity among the additional b quarks not originated from the Higgs boson decay can be solved by exploiting differences in their kinematics, as represented in Figure 4.14. To select and identify at reconstruction level the b quark coming from the VLQ B, the p_T -leading b-tagged jet non-overlapping with the selected Higgs-jet ($\Delta R(b,H) > 1.2$) is identified as the one coming from the B decay.

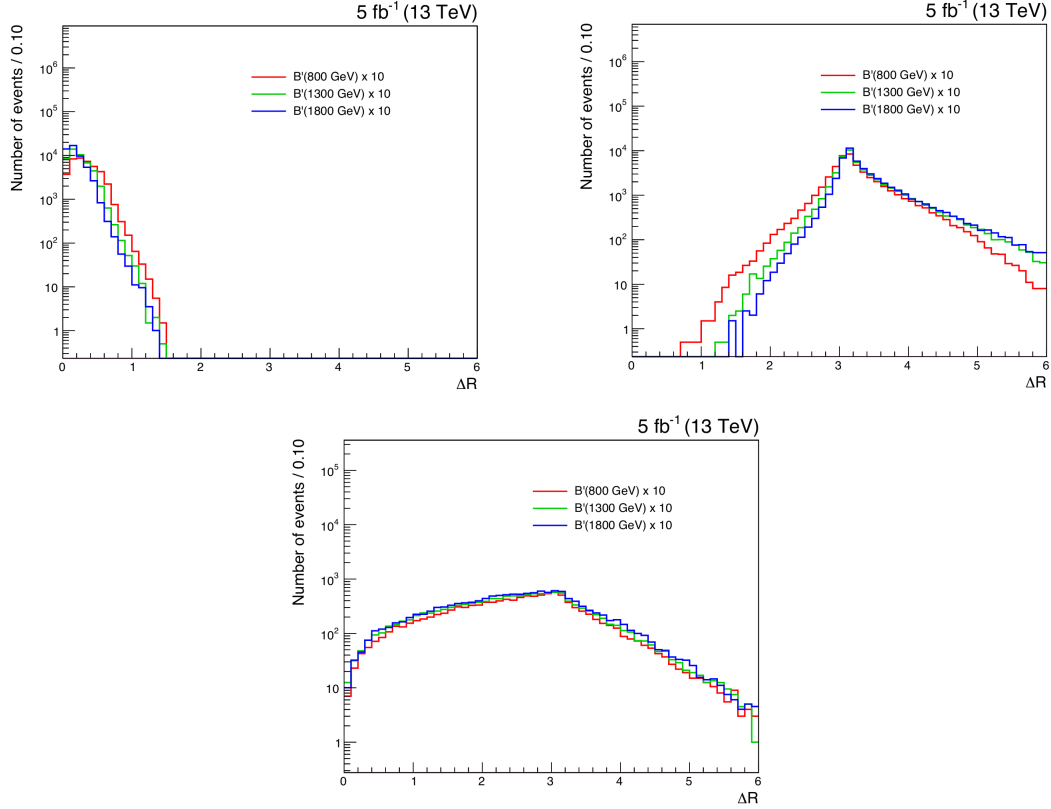


Figure 4.13: Spatial separations at generator level between the Higgs boson and the the b-quarks from the Higgs decay (left), the one arising from the decay of the VLQ B (center) and the spectator one. Three mass hypotheses for the B quark mass are pictured with different colors and multiplied by a normalized factor of 10.

The mass of the VLQ B is therefore reconstructed as the invariant mass between the boosted Higgs boson candidate and the non overlapping p_T -leading b-tagged jet. The shape of the reconstructed B mass is illustrated in Figure 4.15 for two mass hypotheses and for different assumptions on the width. It is well visible the peaking structure of the distributions and the broadening trend as the generated width/mass hypothesis increases. For the narrow-width approximation case, and depending on the mass hypothesis, the experimental resolution in the reconstructed B mass, defined as the ratio between the root-mean-square width and the mean of the distribution, ranges between 8% and 15%. A comparison between data and simulated background is further presented in Figure 4.16.

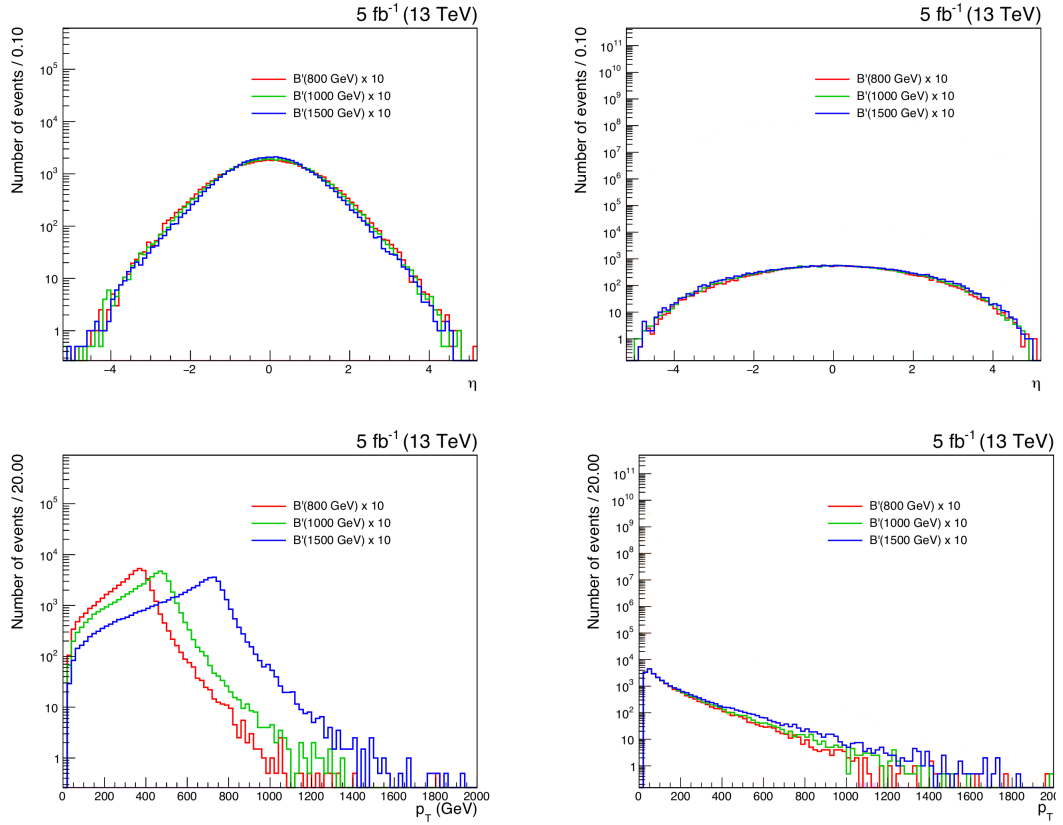


Figure 4.14: Kinematic distributions at generator level for the b quark arising from the decay of the VLQ B (left) and the spectator one (right). Pseudorapidity (top) and transverse momentum (bottom) distributions are illustrated. Three mass hypotheses for the B quark mass are pictured with different colors and multiplied by a normalized factor of 10.

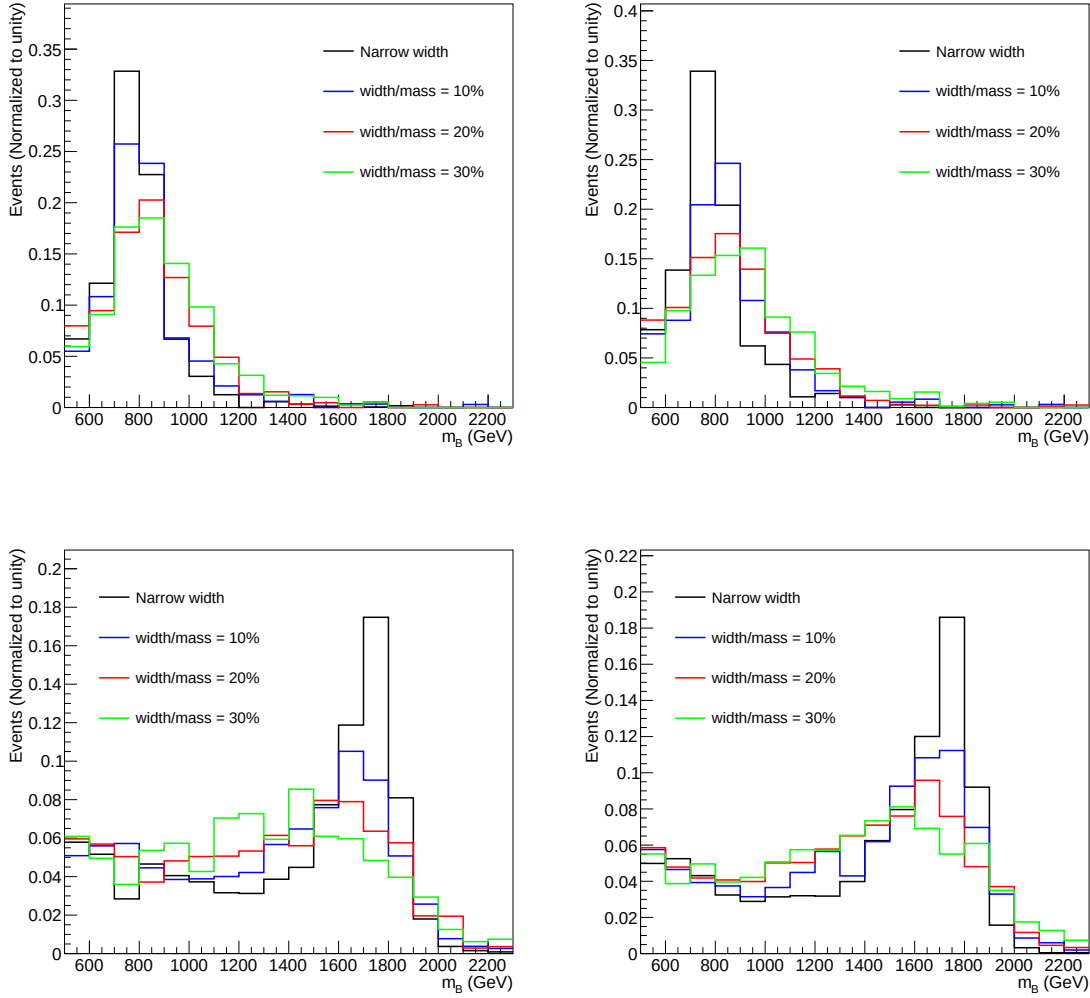


Figure 4.15: Simulated distributions of the reconstructed VLQ B mass, after the full selection is applied. The category with exactly 0 forward jets is shown on the left column, while the right column present the category where at least one forward jet is reconstructed. Two VLQ B mass hypotheses are here described, 800 GeV in the top and 1800 GeV in bottom row, respectively. The different colors refer to the different widths scenarios assumed, from narrow width to a resonance with width equal to the 30% of its mass.

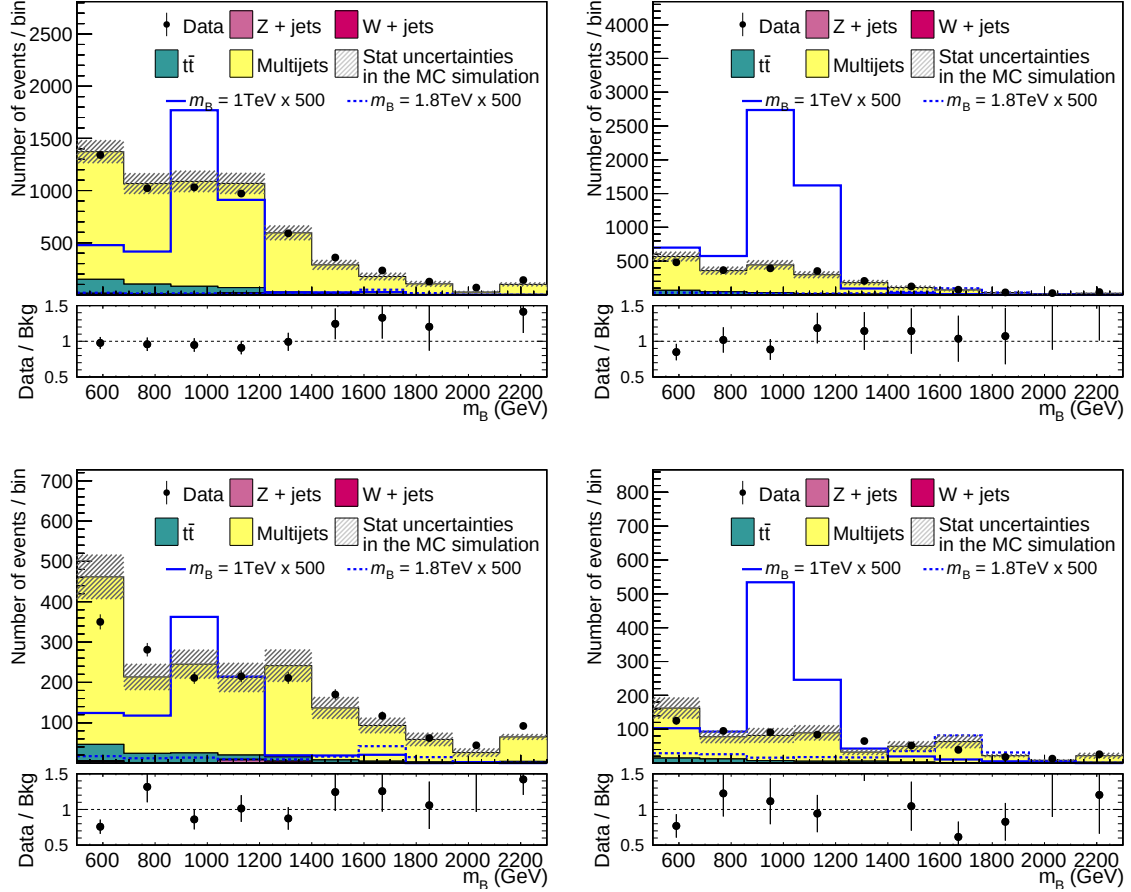


Figure 4.16: Distribution of the reconstructed VLQ B mass, after the full selection is applied. The category with exactly 0 forward jets is shown on the left column, while the right column present the category where at least one forward jet is reconstructed. Low and high mass analysis are here described, in the top and bottom row, respectively. Solid filled histograms label SM background distributions: they are summed cumulatively and their sum is rescaled to data. The hatched grey region show the statistical uncertainty on MC. Red lines show the distributions for MC signal mass points (including a multiplicative factor 500), while data are indicated with black dots. In the lower box, the ratio between data and the total background distribution is shown. Vertical bars indicates the statistical uncertainty for each bin. In each plot the value of the overflow is added to the last bin of the histogram.

All the selection requirements employed in the analysis are summarized in Table 4.4, including both the offline and the online details.

Level	Requirement	Value	Comment
Online selection	Level 1	H_T based, jet p_T	H_T based only for runs B-G, OR between H_T based and jet p_T for run H
	High-Level Trigger	HLT_PFHT900, HLT_PFJet450	H_T based only for runs B-G, OR between H_T based and jet p_T for run H
Preselection	Lepton veto	0 tight leptons	–
	AK8 jet multiplicity	≥ 1	–
	AK4 jet multiplicity	≥ 3	–
	b-tagged AK4 jet multiplicity	≥ 1	medium CSVv2 working point
	H_T	950 GeV	baseline request
Selection	Higgs tagging	pruned mass $\in [105, 135]$ GeV	
		2 b-tagged subjets	–
	Overlap removal with $b\bar{b}$ from Higgs	$\Delta R(b, H) > 1.2$	–
	H_T	1250 GeV	only in the HM analysis

Table 4.4: Summary of the selection requirements.

4.3 Background estimation

The dominant source of background for this search is constituted of QCD production of jets. The second mechanism, in order of importance, arises from $t\bar{t}$ events and it contributes for the 5-7% to the total background. Additional minor contributions accounting for less than 1% are coming from W/Z +jets processes and they are also taken into account. Both the $t\bar{t}$ and the other electroweak contributions are taken from Monte Carlo simulation.

4.3.1 The ABCD method

To reduce the dependence on the QCD multijet modeling in simulation, we use a data-driven technique to extrapolate this contribution from data. The procedure we use is the so-called *ABCD method*, which estimates both the expected yield and the m_B distribution shape of the background in the signal region using control regions depleted in signal. More in detail, in addition to the nominal signal region (described in Section 4.2) three control regions are defined in a two-dimensional phase space defined by the number of b-tagged subjets of the Higgs jet and its mass m_H . The four regions are defined as follows:

- Region A (the signal region): two b-tagged subjets and $m_H \in [105, 135]$;
- Region B: two b-tagged subjets and $m_H \in [75, 105] \cup m_H \in [135, \infty[$;
- Region C: one b-tagged subjet and $m_H \in [105, 135]$;
- Region D: one b-tagged subjet and $m_H \in [75, 105] \cup m_H \in [135, \infty[$.

A schematic representation is pictured in Figure 4.17. The signal contamination in the control regions was verified to be negligible, as shown in Table 4.5.

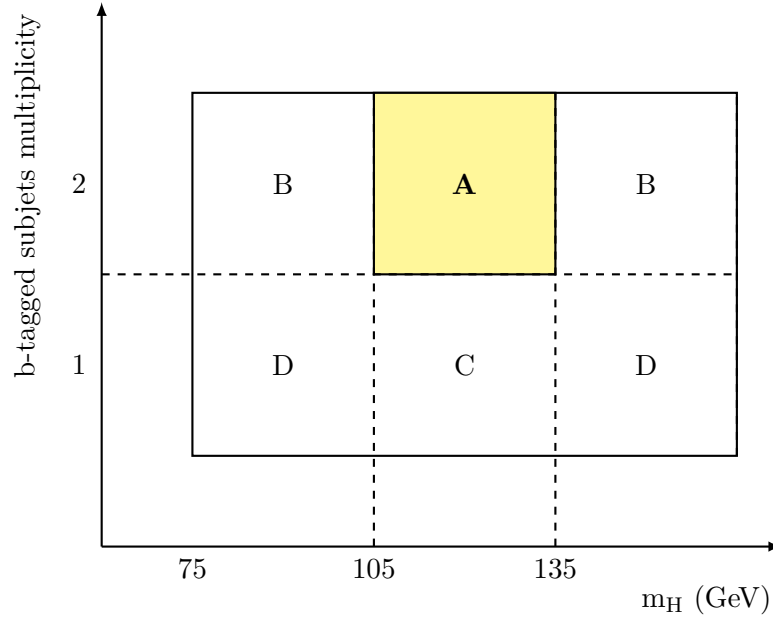


Figure 4.17: Schematic view of the signal and control regions defined in the analysis workflow. The signal region is highlighted in yellow.

category	mass region	source	regions		
			B	C	D
at least one forward jet	low mass analysis	$m_B = 1000$ GeV	0.67%	0.52%	0.03%
		$m_B = 1800$ GeV	0.34%	0.53%	0.10%
	high mass analysis	$m_B = 1000$ GeV	0.51%	0.44%	0.07%
		$m_B = 1800$ GeV	0.47%	0.95%	0.14%
zero forward jets	low mass analysis	$m_B = 1000$ GeV	0.19%	0.14%	0.04%
		$m_B = 1800$ GeV	0.07%	0.09%	0.02%
	high mass analysis	$m_B = 1000$ GeV	0.16%	0.14%	0.04%
		$m_B = 1800$ GeV	0.10%	0.18%	0.03%

Table 4.5: Signal contamination in all three control regions, for both categories and both mass regimes analyses. Two mass values of the m_B are presented, 1000 GeV and 1800 GeV. The reported values have been obtained as the ratio between signal samples yields over the multijets yields, considering only the events in three bins under the resonance peak.

The m_B distributions in the signal regions for the two event categories are shown in Figure 4.16, while the inverted selection defining the control regions have been used to obtain the distributions presented in Figure 4.18 and Figure 4.19, in the low- and high-mass analysis, respectively. The corresponding event yields are reported in Table 4.6.

category	mass region	source	regions			
			A	B	C	D
> 0 forward jet	low mass analysis	Multijets (MC)	1924 $\pm$ 139	5125 $\pm$ 214	22436 $\pm$ 480	60517 $\pm$ 754
		$t\bar{t}$	172 $\pm$ 8	744 $\pm$ 18	2790 $\pm$ 34	8774 $\pm$ 60
		W + jets	13 $\pm$ 4	46 $\pm$ 10	318 $\pm$ 29	934 $\pm$ 48
		Z + jets	5 $\pm$ 4	103 $\pm$ 19	157 $\pm$ 24	468 $\pm$ 41
		Backgrounds	2114 $\pm$ 140	6018 $\pm$ 216	25701 $\pm$ 482	70693 $\pm$ 759
		Data	2107 $\pm$ 46	5952 $\pm$ 77	26352 $\pm$ 162	69206 $\pm$ 263
		$m_B = 1000$ GeV	46 $\pm$ 1	20.9 $\pm$ 0.9	74 $\pm$ 2	49 $\pm$ 1
		$m_B = 1800$ GeV	9.3 $\pm$ 0.2	4.4 $\pm$ 0.1	21.9 $\pm$ 0.3	15.0 $\pm$ 0.3
	high mass analysis	Multijets (MC)	589 $\pm$ 59	1961 $\pm$ 101	6553 $\pm$ 197	22803 $\pm$ 358
		$t\bar{t}$	47 $\pm$ 4	298 $\pm$ 11	857 $\pm$ 19	3176 $\pm$ 36
		W + jets	7 $\pm$ 3	31 $\pm$ 7	154 $\pm$ 18	503 $\pm$ 31
		Z + jets	0.0 $\pm$ 0.0	42 $\pm$ 12	44 $\pm$ 12	242 $\pm$ 30
		Backgrounds	643 $\pm$ 59	2332 $\pm$ 103	7608 $\pm$ 199	26723 $\pm$ 362
		Data	608 $\pm$ 25	2192 $\pm$ 47	7608 $\pm$ 199	26723 $\pm$ 362
		$m_B = 1000$ GeV	8.5 $\pm$ 0.6	5.2 $\pm$ 0.7	17 $\pm$ 0.8	14.8 $\pm$ 0.8
		$m_B = 1800$ GeV	7.8 $\pm$ 0.2	3.6 $\pm$ 0.1	18.4 $\pm$ 0.3	12.5 $\pm$ 0.2
zero forward jets	low mass analysis	Multijets (MC)	5316 $\pm$ 226	13705 $\pm$ 358	58753 $\pm$ 778	154914 $\pm$ 1210
		$t\bar{t}$	395 $\pm$ 13	1819 $\pm$ 27	6937 $\pm$ 53	21744 $\pm$ 95
		W + jets	25 $\pm$ 6	148 $\pm$ 19	760 $\pm$ 45	2512 $\pm$ 81
		Z + jets	43 $\pm$ 13	220 $\pm$ 29	466 $\pm$ 41	1415 $\pm$ 71
		Backgrounds	5779 $\pm$ 226	15891 $\pm$ 360	66916 $\pm$ 783	180586 $\pm$ 1219
		Data	5886 $\pm$ 77	16587 $\pm$ 129	73452 $\pm$ 271	194232 $\pm$ 441
		$m_B = 1000$ GeV	29 $\pm$ 1	15.7 $\pm$ 0.8	50 $\pm$ 1	37 $\pm$ 1
		$m_B = 1800$ GeV	4.9 $\pm$ 0.1	2.5 $\pm$ 0.1	11.3 $\pm$ 0.2	8.1 $\pm$ 0.2
	high mass analysis	Multijets (MC)	1593 $\pm$ 99	5184 $\pm$ 171	17179 $\pm$ 329	60472 $\pm$ 586
		$t\bar{t}$	123 $\pm$ 7	2104 $\pm$ 29	7969 $\pm$ 57	716 $\pm$ 17
		W + jets	15 $\pm$ 3	86 $\pm$ 13	393 $\pm$ 29	1522 $\pm$ 58
		Z + jets	13 $\pm$ 7	103 $\pm$ 19	167 $\pm$ 24	672 $\pm$ 49
		Backgrounds	1743 $\pm$ 99	6090 $\pm$ 174	19843 $\pm$ 332	70634 $\pm$ 594
		Data	1753 $\pm$ 42	6451 $\pm$ 80	22039 $\pm$ 148	74941 $\pm$ 274
		$m_B = 1000$ GeV	7.0 $\pm$ 0.5	4.7 $\pm$ 0.4	13.0 $\pm$ 0.7	13.0 $\pm$ 0.7
		$m_B = 1800$ GeV	4.1 $\pm$ 0.1	2.1 $\pm$ 0.1	9.7 $\pm$ 0.2	7.0 $\pm$ 0.2

Table 4.6: Pre-fit events yields for all background sources (taken from Monte Carlo simulation), data and for two m_B hypotheses. Both the categories with zero forward jets and at least one forward jet are presented, and both the two mass analyses are reported as well.

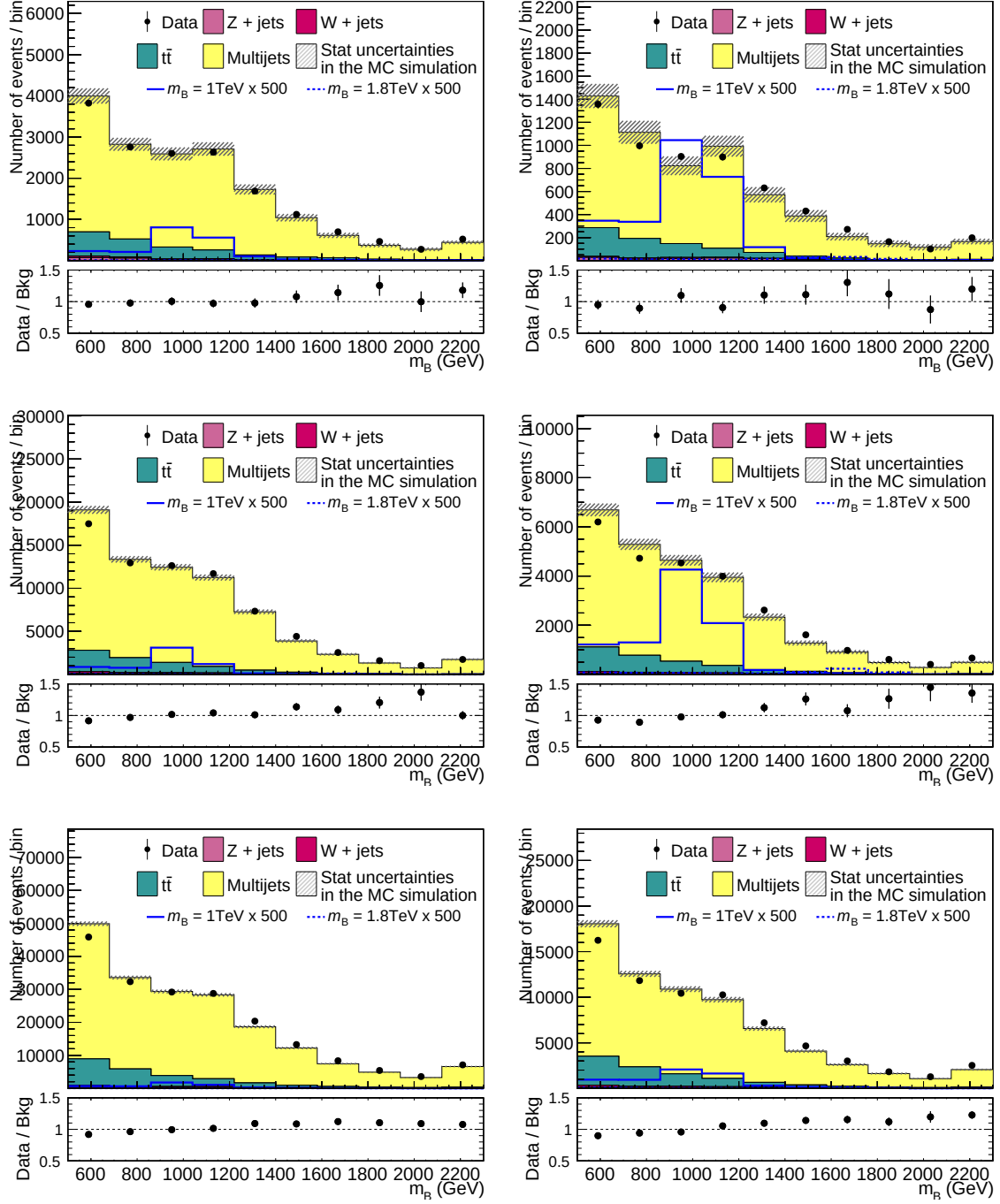


Figure 4.18: Distributions of the reconstructed VLQ B mass in the control region B (top), C (middle) and D (bottom), after the full selection is applied and for the low mass analysis. The category with exactly 0 forward jets is shown on the left column, while the right column presents the category where at least one forward jet is reconstructed. Solid filled histograms label SM background distributions: they are summed cumulatively and their sum is rescaled to data. The hatched grey region show the statistical uncertainty on MC. Red lines show the distributions for MC signal mass points (including a multiplicative factor 500), while data are indicated with black dots. In the lower box, the ratio between data and total background distribution is shown. Vertical bars indicates the statistical uncertainty for each bin. In each plot the value of the overflow is added to the last bin of the histogram.

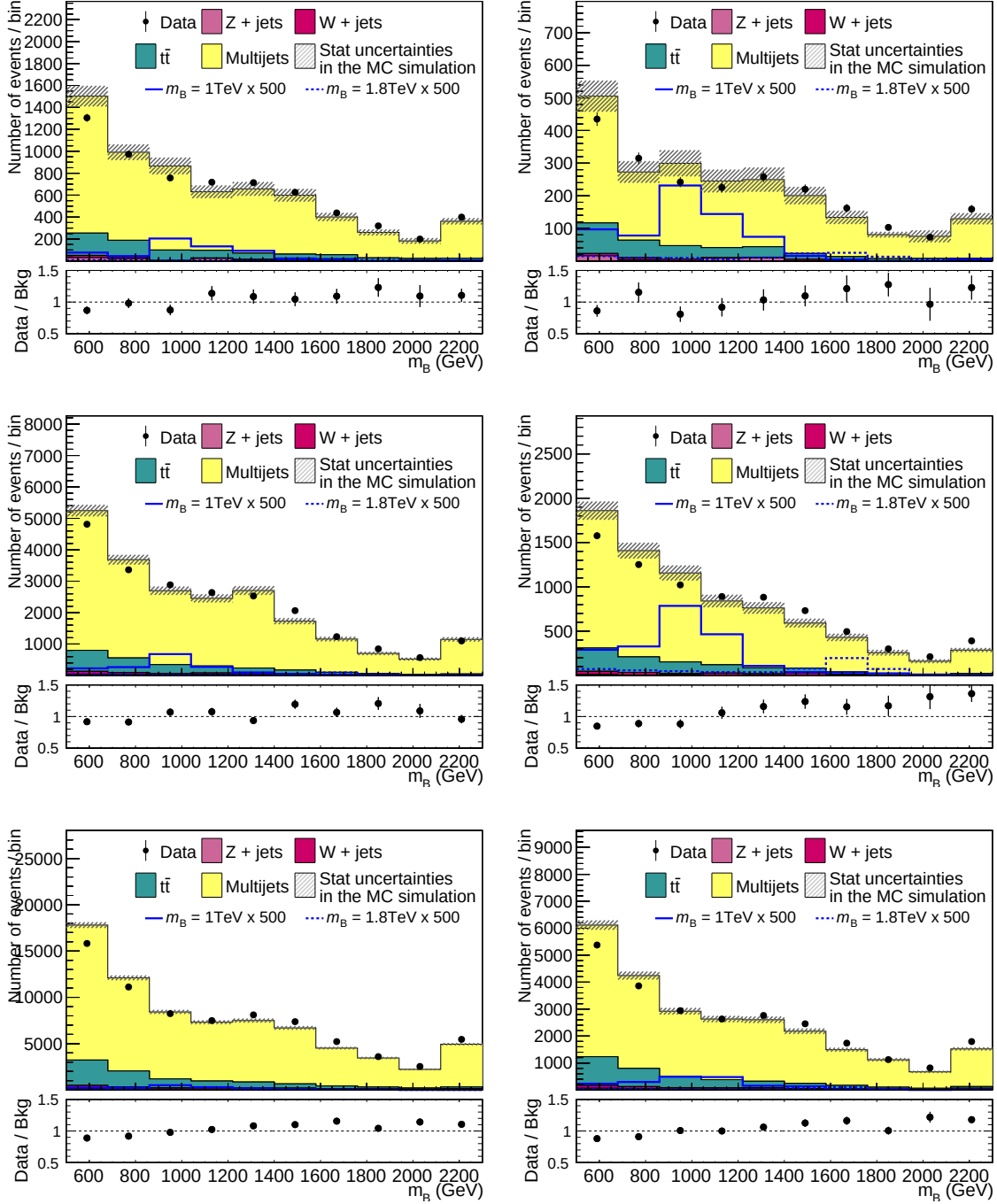


Figure 4.19: Distributions of the reconstructed VLQ B mass in the control region B (top), C (middle) and D (bottom), after the full selection is applied and for the high mass analysis. The category with exactly 0 forward jets is shown on the left column, while the right column presents the category where at least one forward jet is reconstructed. Solid filled histograms label SM background distributions: they are summed cumulatively and their sum is rescaled to data. The hatched grey region show the statistical uncertainty on MC. Red lines show the distributions for MC signal mass points (including a multiplicative factor 500), while data are indicated with black dots. In the lower box, the ratio between data and total background distribution is shown. Vertical bars indicates the statistical uncertainty for each bin. In each plot the value of the overflow is added to the last bin of the histogram.

Once the control regions are defined, the multijet background yield in the signal region can be derived. Assuming the two variables as uncorrelated, the background event yield in the signal region is derived from the background-enriched regions B, C and D as follows:

$$N_A = N_C \times \frac{N_B}{N_D} \quad (4.2)$$

where N_A , N_B , N_C and N_D are the yields in regions A, B, C and D respectively. The $t\bar{t}$ contribution predicted by simulation is here subtracted from the observed yields while the contributions from Z+jets and W+jets are not, given their negligible impact.

The shape of the multijet background is estimated from the m_B event distribution in data in control region C and this is possible under the assumption that the reconstructed m_B spectrum is not correlated with the b subjet multiplicity. The yield ratios between the regions A and C, and B and D respectively resulted to be in agreement within systematic uncertainties, as shown in Table 4.7.

category	mass region	ratios	
		A/C	B/D
at least one forward jet	low mass analysis	$(8.58 \pm 0.65)\%$	$(8.47 \pm 0.37)\%$
	high mass analysis	$(8.98 \pm 0.94)\%$	$(8.60 \pm 0.46)\%$
zero forward jets	low mass analysis	$(9.05 \pm 0.40)\%$	$(8.85 \pm 0.24)\%$
	high mass analysis	$(9.27 \pm 0.60)\%$	$(8.57 \pm 0.30)\%$

Table 4.7: Ratio between multijet yields between control regions A and C, and between control regions B and D, respectively. All categories depending on the forward jets multiplicity and on the H_T selection are shown.

The shape agreement between region A and C was verified using simulated multijet events and cross-checked in data, as will be extensively described in the Sections 4.3.2 and 4.3.3, respectively.

4.3.2 Closure tests

Before proceeding with the ABCD method application on data, it was fundamental to verify its correct functioning with dedicated tests. Two strategies have been employed and will be described below.

Validation of the method on simulated events

As first check, we have compared the shape agreement between the simulated multijet m_B distribution in the signal region A and in the control region C (Figures 4.20 and 4.21). The two distributions agree well within statistical uncertainties. The normalization factor obtained from the yields in the control regions B, C, and D is then applied, resulting in a slight shift of the expectation over prediction ratio, as shown in Figures 4.20 and 4.21. The fluctuations are due to the low statistics in signal region A for the multijet simulation, but this won't affect the analysis as the final template for the QCD background is estimated from data events in the control regions.

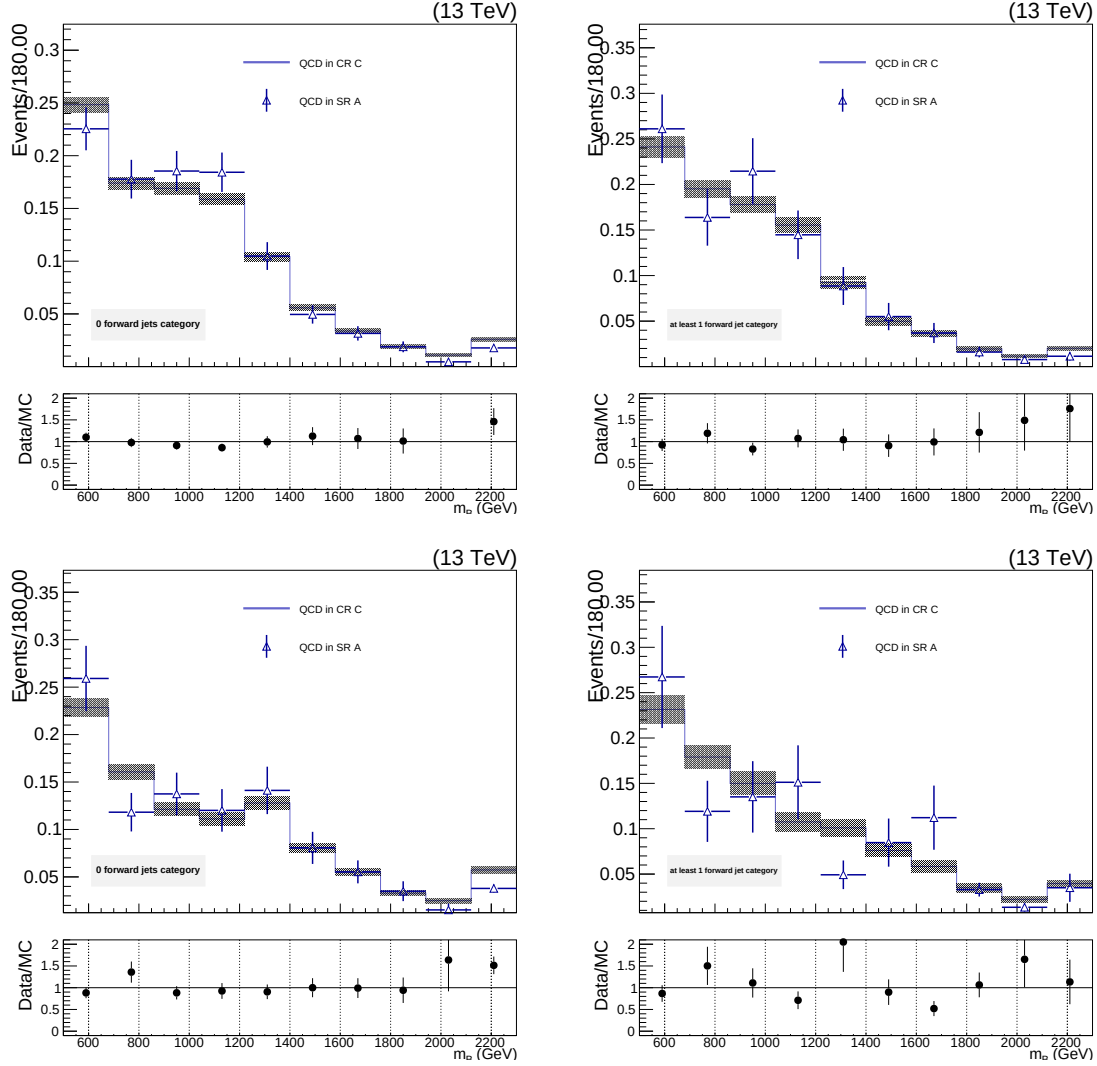


Figure 4.20: Shape comparison for the m_B reconstructed mass spectrum on multijet MC simulation only. Low and high mass analyses are described on the left and on the right plots respectively. The solid line shows the normalized distribution in control region C and the distribution resulting from the application of the ABCD method, while the triangular points refer to the distribution in signal region A. Statistical uncertainties are also pictured with hatched filling and vertical error bars in the two regions, respectively.

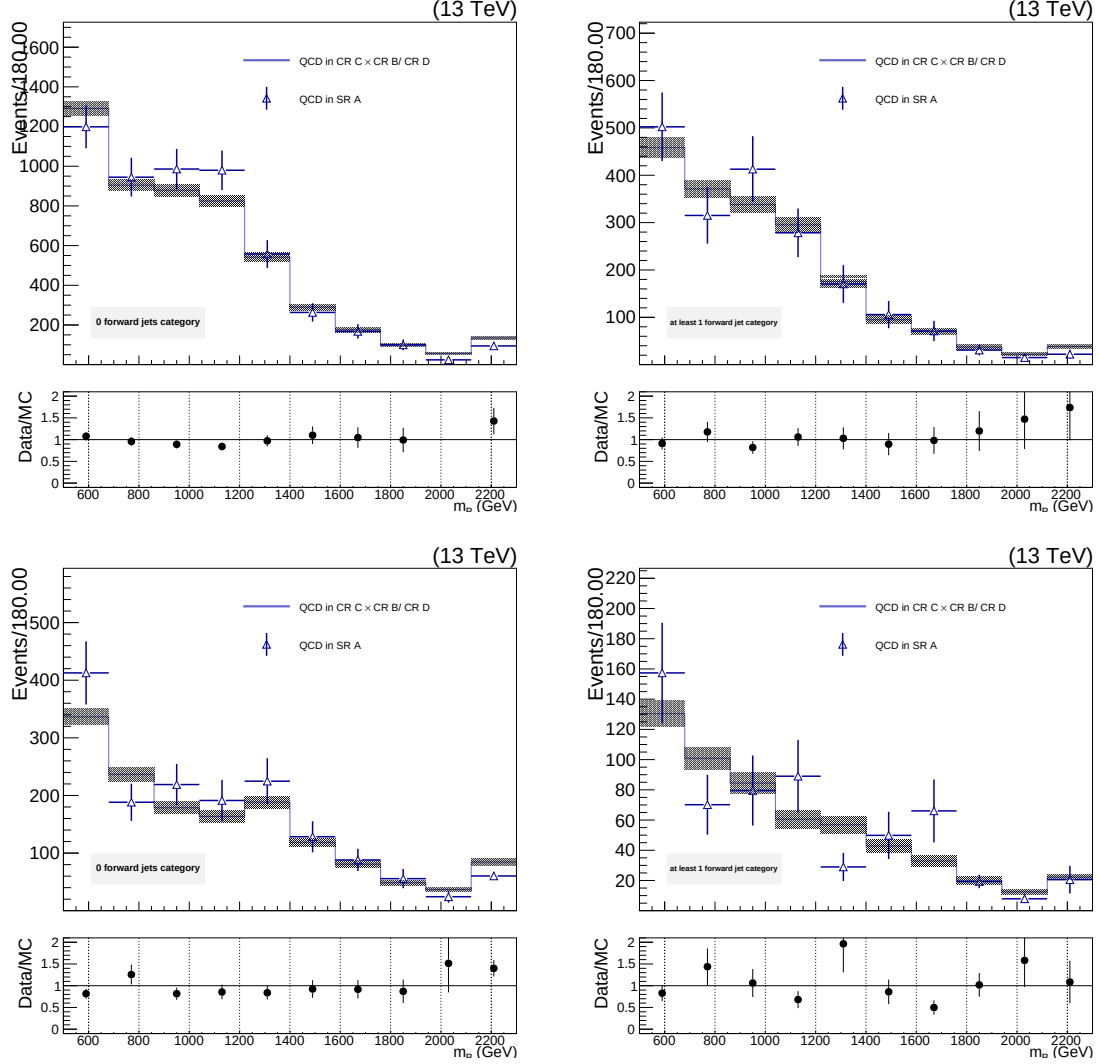


Figure 4.21: Normalization comparison for the m_B reconstructed mass spectrum on multijet MC simulation only. Low and high mass analyses are described on the left and on the right plots respectively. The solid line shows the normalized distribution in control region C and the distribution resulting from the application of the ABCD method, while the triangular points refer to the distribution in signal region A. Statistical uncertainties are also pictured with hatched filling and vertical error bars in the two regions, respectively.

Modified ABCD method: the A'B'C'D' method

Additionally, the ABCD method is validated using a signal-depleted region defined in the high mass sidebands. In this alternative matrix, the modified signal regions, labeled with A', is defined in the translated mass window [135, 165] GeV. The four regions A'B'C'D' are defined as follows:

- Region A': two b-tagged subjets and $m_H \in [135, 165]$;
- Region B': two b-tagged subjets and $m_H \in [75, 105] \cup m_H \in [165, \infty[$;
- Region C': one b-tagged subjet and $m_H \in [135, 165]$;
- Region D': one b-tagged subjet and $m_H \in [75, 105] \cup m_H \in [165, \infty[$.

A schematic view is also pictured in Figure 4.22. In Figures 4.23 and 4.24 we compared the background shape distribution in region C' and in signal region A'. The background distribution estimated in region A', using the method described above, agrees well with the observation. The difference between the estimated and the observed yields in region A' is used as a normalization systematic uncertainty on the multijet estimate, as further described in Section 4.4.

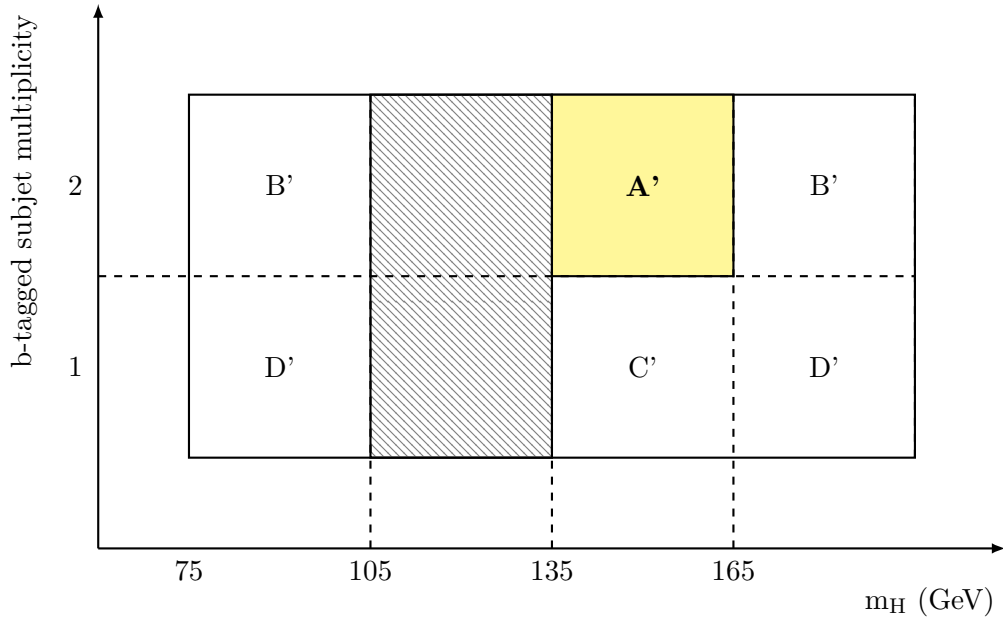


Figure 4.22: Schematic view of the modified signal and control regions defined to perform a closure test of the background estimation method. The upper side of the hatched box corresponds to the nominal signal region, that is not used in this part of the study.

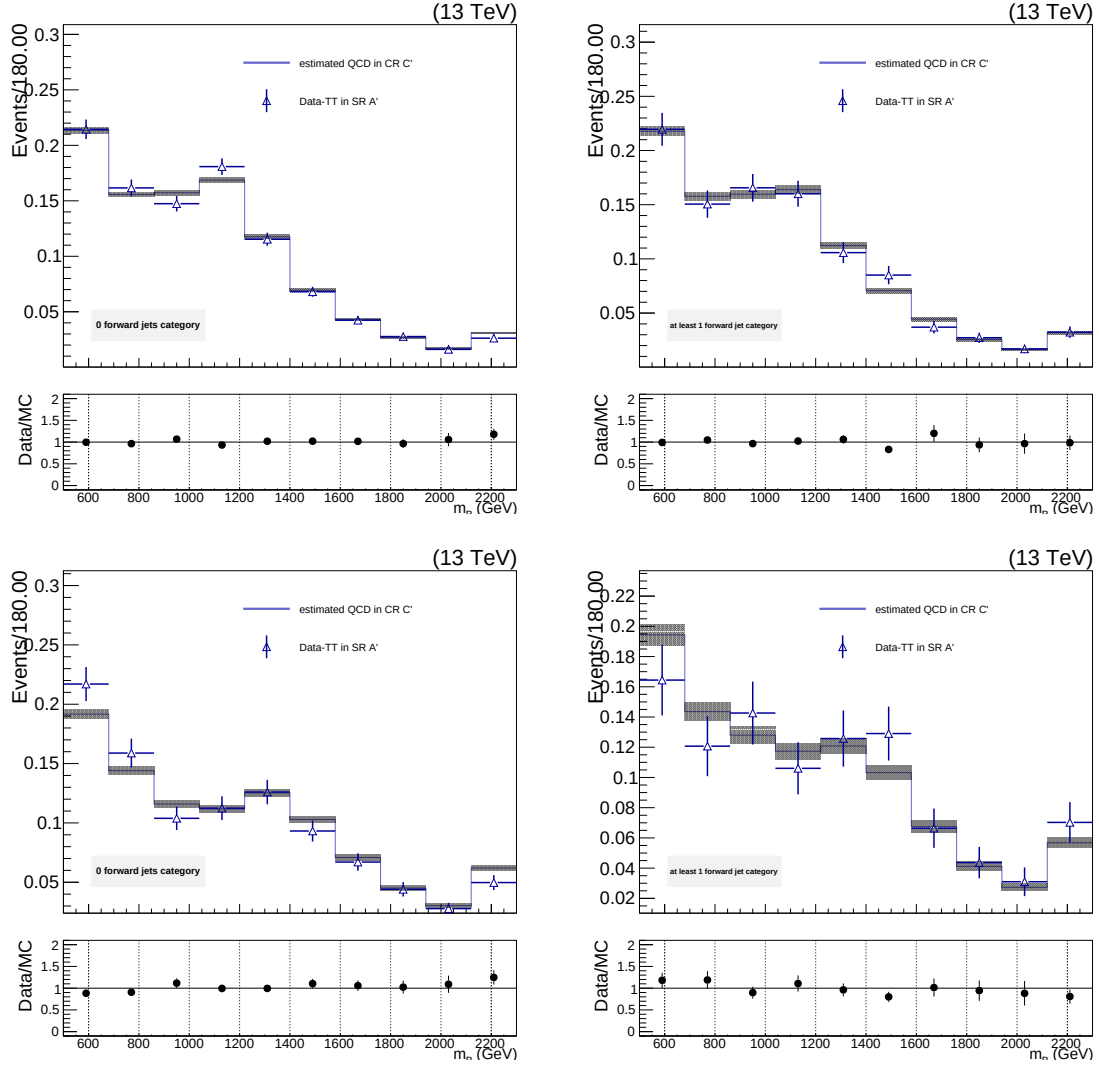


Figure 4.23: Shape comparison for the m_B reconstructed mass spectrum. Low and high mass analyses are described on the left and on the right plots respectively. The solid line shows the normalized distribution in control region C' and the distribution resulting from the application of the $A'B'C'D'$ method, while the triangular points refer to the distribution in modified signal region A' . The $t\bar{t}$ contribution is here subtracted from data. Statistical uncertainties are also pictured with hatched filling and vertical error bars in the two regions, respectively. Low and high mass analyses are described on the left and on the right plots respectively.

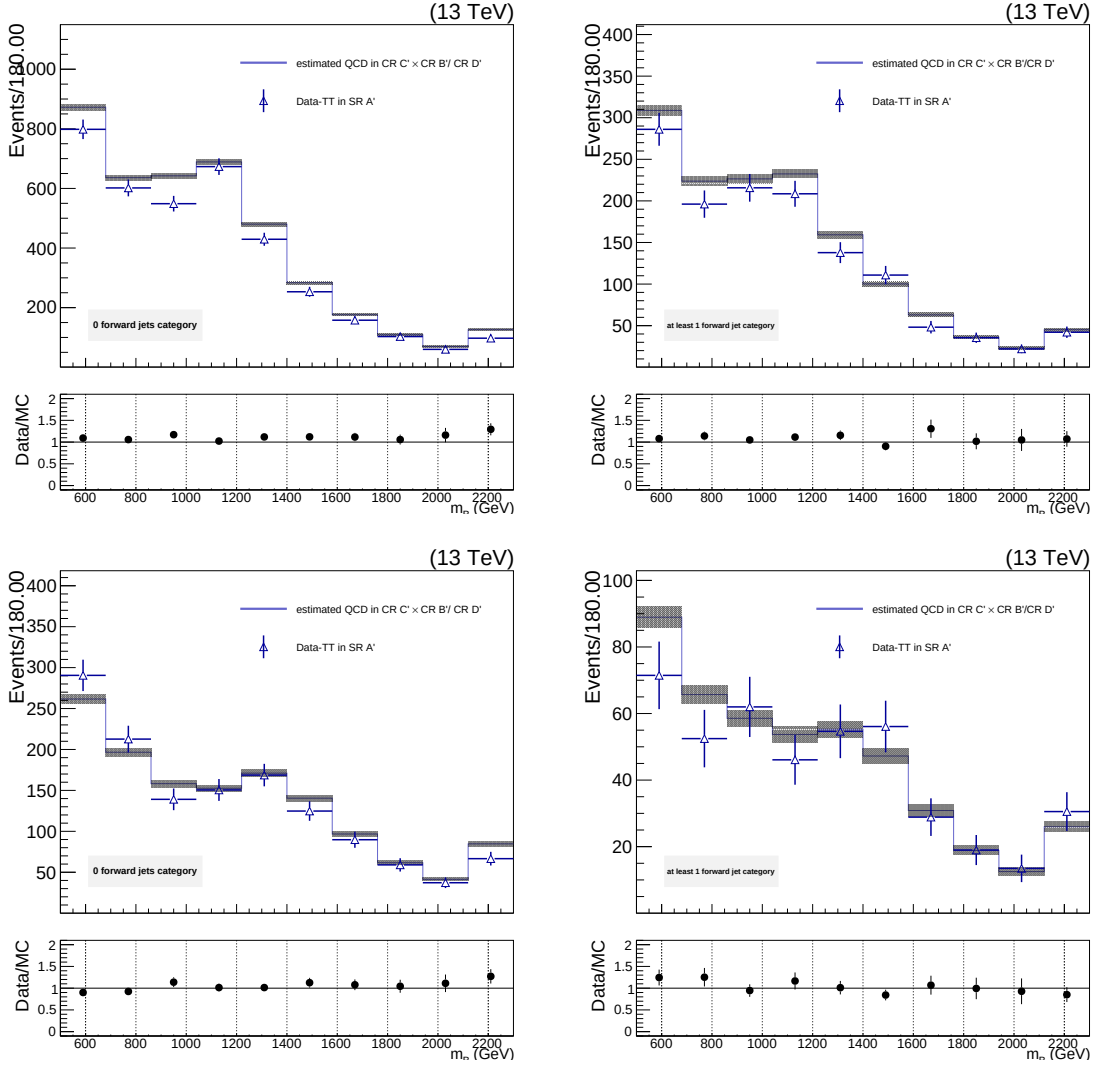


Figure 4.24: Normalization comparison for the m_B reconstructed mass spectrum. Low and high mass analyses are described on the left and on the right plots respectively. The solid line shows the normalized distribution in control region C' and the distribution resulting from the application of the A'B'C'D' method, while the triangular points refer to the distribution in modified signal region A'. The $t\bar{t}$ contribution is here subtracted from data. Statistical uncertainties are also pictured with hatched filling and vertical error bars in the two regions, respectively.

4.3.3 Results

After having tested the validity of the ABCD method, as explained in the previous sections, the method was applied to the real signal region A. Both the shape and the normalization for the QCD estimation and the data distribution in the signal region (with the $t\bar{t}$ component subtracted), resulted to be in good agreement within the statistical uncertainties, as shown in Figures 4.25 and 4.26.

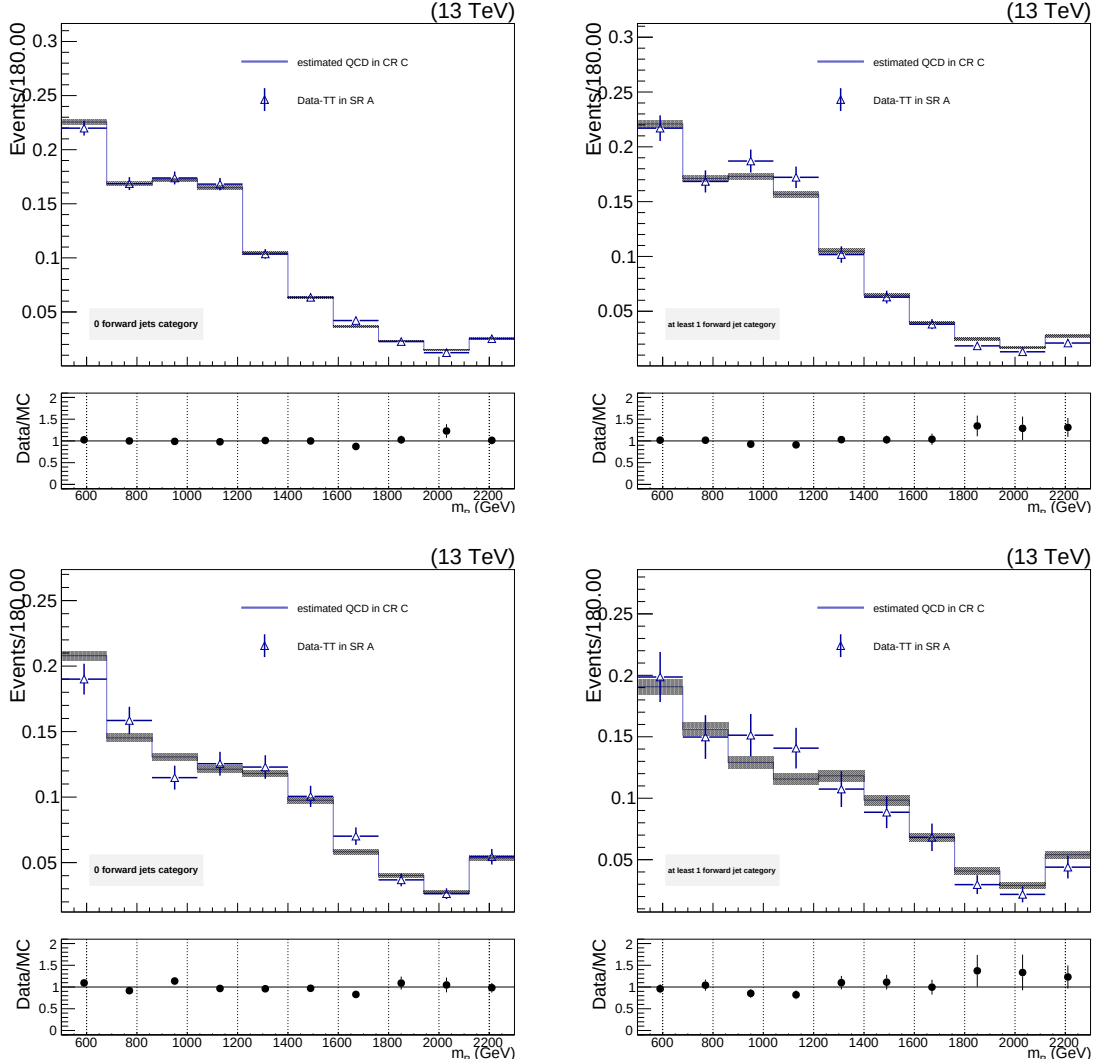


Figure 4.25: Shape comparison for the m_B reconstructed mass spectrum. Low and high mass analyses are described on the left and on the right plots respectively. The solid line shows the normalized distribution in control region C and the distribution resulting from the application of the ABCD method, while the triangular points refer to the distribution in signal region A. The $t\bar{t}$ contribution is here subtracted from data. Statistical uncertainties are also pictured with hatched filling and vertical error bars in the two regions, respectively.

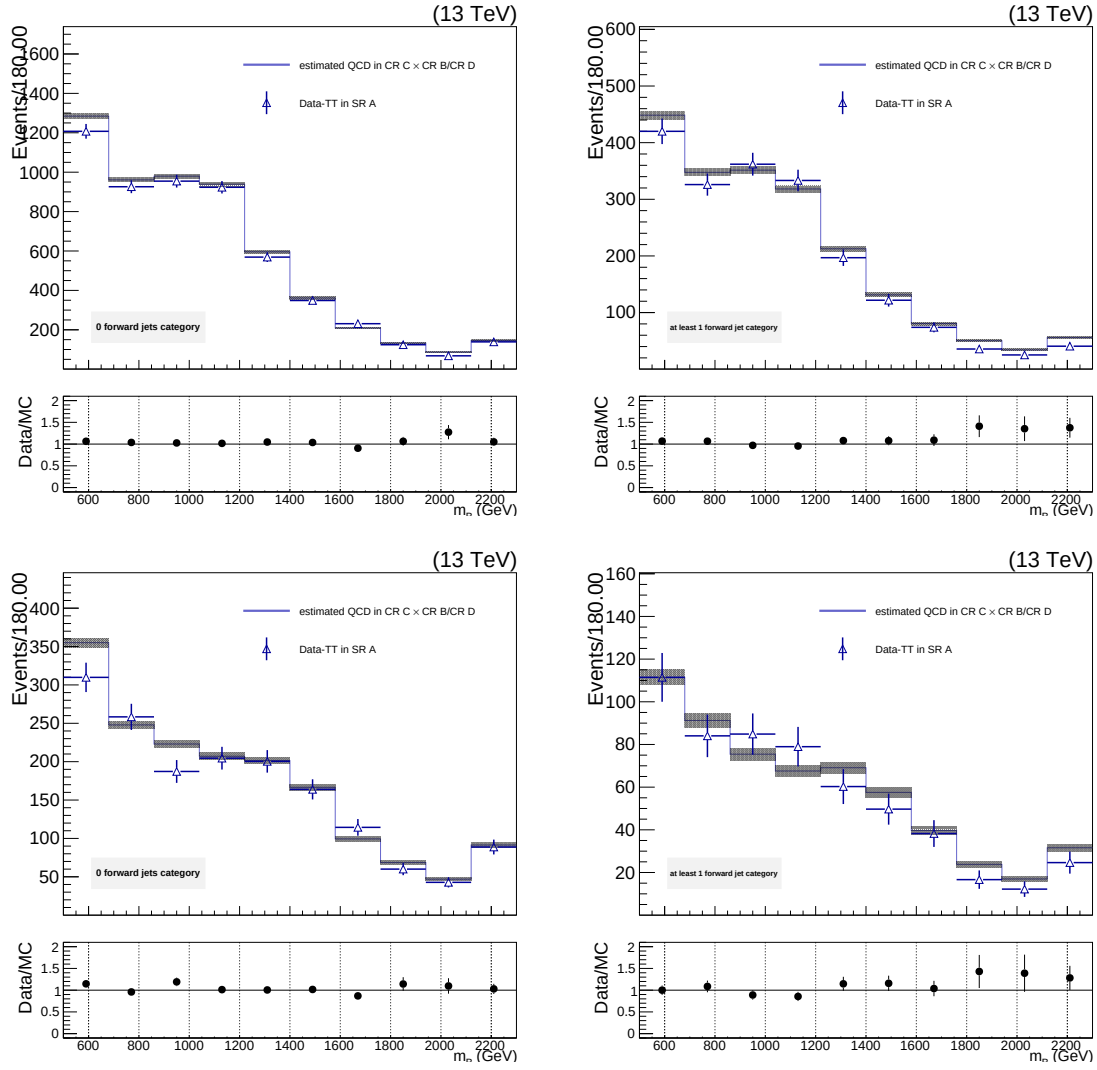


Figure 4.26: Normalization comparison for the m_B reconstructed mass spectrum. Low and high mass analyses are described on the left and on the right plots respectively. The solid line shows the normalized distribution in control region C and the distribution resulting from the application of the ABCD method, while the triangular points refer to the distribution in signal region A. The $t\bar{t}$ contribution is here subtracted from data. Statistical uncertainties are also pictured with hatched filling and vertical error bars in the two regions, respectively. Low and high mass analyses are described on the left and on the right plots respectively.

4.4 Systematic uncertainties

The final reconstructed m_B distribution, both in signal and in control regions, is affected by experimental and theoretical systematic uncertainties. These sources of uncertainty can have either be an overall uniform effect or it might include also shape trend. We will refer to the former as *normalization uncertainties* and to the latter as *shape uncertainties*. In this analysis, log-normal distribution and morphing techniques are used to model normalization and shape uncertainties, respectively. These two models are briefly introduced below.

Normalization uncertainties and log-normal distribution [175] The truncated Gaussian parton distribution function (pdf) is a frequent choice for the treatment of systematic uncertainties, but to avoid difficulties and pathologies of this modeling a commonly alternative parametrization is represented by the log-normal, which is defined as

$$\rho(\theta) = \frac{1}{\sqrt{2} \ln(\kappa)} \exp\left(-\frac{(\ln(\theta/\tilde{\theta}))^2}{2(\ln \kappa)^2}\right) \frac{1}{\theta} \quad (4.3)$$

The parameter κ refers to the width of the distribution while $\tilde{\theta}$ is the best estimate of the observable θ .

If in the case of small uncertainties the Gaussian parametrization can be assumed as an approximation of the log-normal, this is not true anymore when we are dealing with larger uncertainty values. The log-normal represents a more suitable choice as it has a longer tail with respect to the Gaussian and goes to zero at $\theta = 0$.

Shape uncertainties and morphing techniques [176] If the systematic uncertainty has as effect an overall distortion both in the normalization and in the shape of the observable spectrum, its impact can be modeled in the Monte Carlo simulation by altering the parameter j affected by the systematic and then recalculating the new spectra. The updated distribution is obtained by using the parameter most probable value and the shifted distributions considering the updated parameter j plus (minus) its associated uncertainty. The new distribution is labeled with ϵ_{ij}^0 , while we refer to the shifted spectra with $\epsilon_{ij}^\pm$.

To then merge the three spectral shapes into a continuous estimate ϵ_{ij} in each bin i , a *morphing* parameter called f is introduced. It assumes a nominal value of 0 with an uncertainty $\sigma_f = 1$. Finally, a quadratic interpolation for $|f| < 1$ and further linear extrapolation beyond that range, is performed, resulting in a value of the new distribution in the i -th bin given by:

$$\epsilon_{ij} = \frac{f(f-1)}{2} \epsilon_{ij}^- - (f-1)(f+1) \epsilon_{ij}^0 + \frac{f(f+1)}{2} \epsilon_{ij}^+. \quad (4.4)$$

The systematic uncertainties included in this analysis arise from several sources, that can be grouped as related to the reconstruction and use of the physics objects, as due to mismodeling and uncertainties of the experimental apparatus or coming from the underlying theory. Following this division, all the systematics, along with a brief description, are listed in the following paragraphs.

A summary is then presented in Table 4.8, where all the systematics are listed with the magnitude of their effect, their type and a specification of which process is affected. In addition, the normalized statistical uncertainties templates for $m_B = 1000$ GeV and 1800 GeV, in low and high mass analysis respectively, are presented in Figures 4.27, 4.28, 4.29 and 4.30.

Physics object uncertainties

- b-tagging and mis-tagging: these uncertainties account for the efficiency to tag or mistag a jet arising from the hadronization of a b quark.
- Jet Energy Scale (JES) and Jet Energy Resolution (JER): the transverse momenta of both AK4 and AK8 jets are varied by $\pm 1\sigma$ within the uncertainty on JES and JER and the variation is propagated to the main m_B observable.
- Jet Mass Scale (JMR) and Jet Mass Resolution (JMS): as for the p_T , also the pruned mass of AK4 and AK8 jets is allowed to fluctuate of $\pm 1\sigma$ between the uncertainty on JMS and JMR and the effect of this variation on m_B is taken into account.
- Forward jets: the discrepancy between data and Monte Carlo in the modeling of forward jets multiplicity is used to estimate the uncertainty on the description of jets with $|\eta| > 2.4$ in simulation.
- Simulation statistics: the uncertainty due to the limited amount of simulated events is accounted for by allowing each bin of the m_B distribution to independently fluctuate according to Poisson statistics.
- H_T -based triggers: as described in Section 4.2.1, scale factors are extracted and then applied to account for differences between the trigger efficiency measured in data and in simulation. The uncertainty on these scale factors is applied as a function of the hadronic activity and propagated to the m_B distribution affecting its shape.
- ABCD method for background estimation: the uncertainty on the data-driven multijet estimate affects both shape and normalization of the main observable. A first source of uncertainty is the limited amount of statistics of data and $t\bar{t}$ simulation in control region C, as the expected shape in the signal region is estimated from this control region. It affects the m_B shape of the background and it is estimated by varying independently each bin of the m_B spectrum according to its statistical uncertainty. In addition, the difference between the observation and the prediction resulting from the A'B'C'D' closure test, as described in Section 4.3.2, is taken as additional normalization uncertainty on the overall multijets m_B distribution.

Experimental apparatus uncertainties

- Integrated luminosity: the systematic uncertainty associated to the total integrated luminosity recorded by CMS in 2016 is 2.5% [177].

- Pileup modeling and reweighting: this contributions arises from the uncertainty in the PU modeling in simulation and is computed by a variation of the minimum bias cross-section of a factor 4.6%.

Theoretical uncertainties

- Factorization and renormalization scales (q_2): the theoretical uncertainty on the choice of the multijets factorization and renormalization scales, μ_F and μ_F [178], is taken into account by halving and doubling their values and taking the combination resulting in the maximum variation. Only the effect of acceptance is considered for signal events, while both acceptance and normalization are included for simulated background events.
- Parton distribution functions: this contribution arises from the choice of PDFs and it is estimated by evaluating the effect of PDFs replicas on the analysis and by considering their envelope to computed the $\pm\sigma$ fluctuation.

Source	Effect	Type	Signal MC	Backgrounds MC	Multijets
b-tagging efficiency	0–9%	norm	✓	✓	✓†
Misidentification efficiency	0–2%	norm	✓	✓	✓†
Jet energy scale	1–7%	shape	✓	✓	✓†
Jet energy resolution	1.0–1.5%	shape	✓	✓	✓†
Jet mass scale	0–5%	shape	✓	✓	✓†
Jet mass resolution	0–4%	shape	✓	✓	✓†
MC statistics accuracy	1–4%	shape	✓	✓	✓†
Mismodeling of forward jets	0.5/2.0%	norm	✓	✓	✓†
Trigger	<0.5%	norm	✓	✓	✓†
Luminosity	2.5%	norm	✓	✓	✓†
Pileup modeling	0–12%	norm	✓	✓	✓†
q_2	15–25%	norm	✓	✓	✓†
PDF	1.0–4.5%	shape	✓	✓	✓†
Background estimation	5–10%	shape	✗	✗	✓

Table 4.8: Summary of systematic uncertainties. The ✓ indicates the application of the systematic on the given sample, while the sign ✗ stands for the missing inclusion of the considered source. The apex † for multijets indicates that the systematic is considered as propagated from the $t\bar{t}$ contribution, when it is subtracted from data.

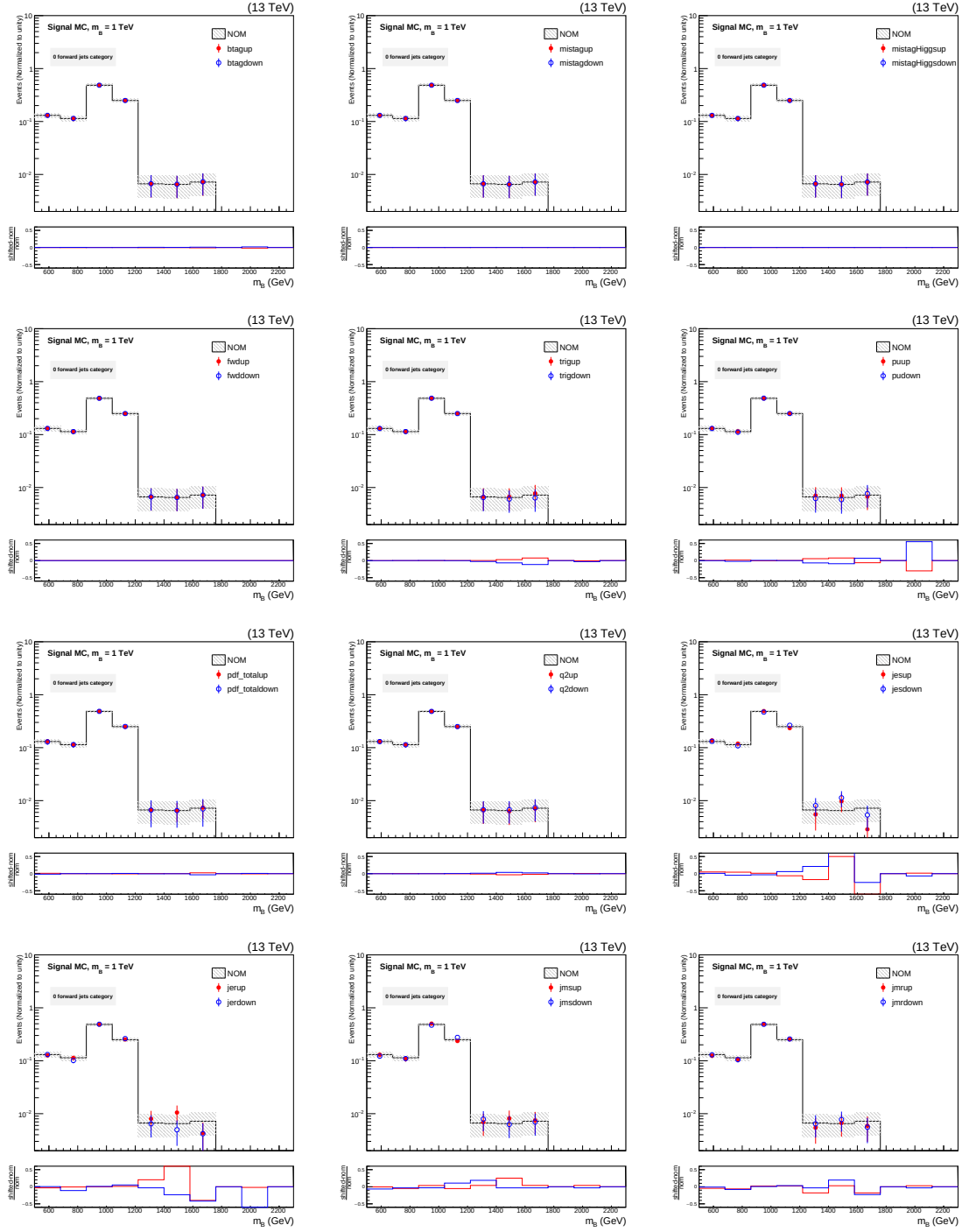


Figure 4.27: Systematic variation on the m_B distribution for signal Monte Carlo events, considering a mass value of 1000 GeV in the low mass analysis. The nominal distribution (black line) is plotted with its statistical uncertainty (hatched gray filling), the shift up (red) and shift down (blue) templates account for differences in both shape and normalization with respect to the nominal sample. Plots are shown for the category with exactly zero forward jets.

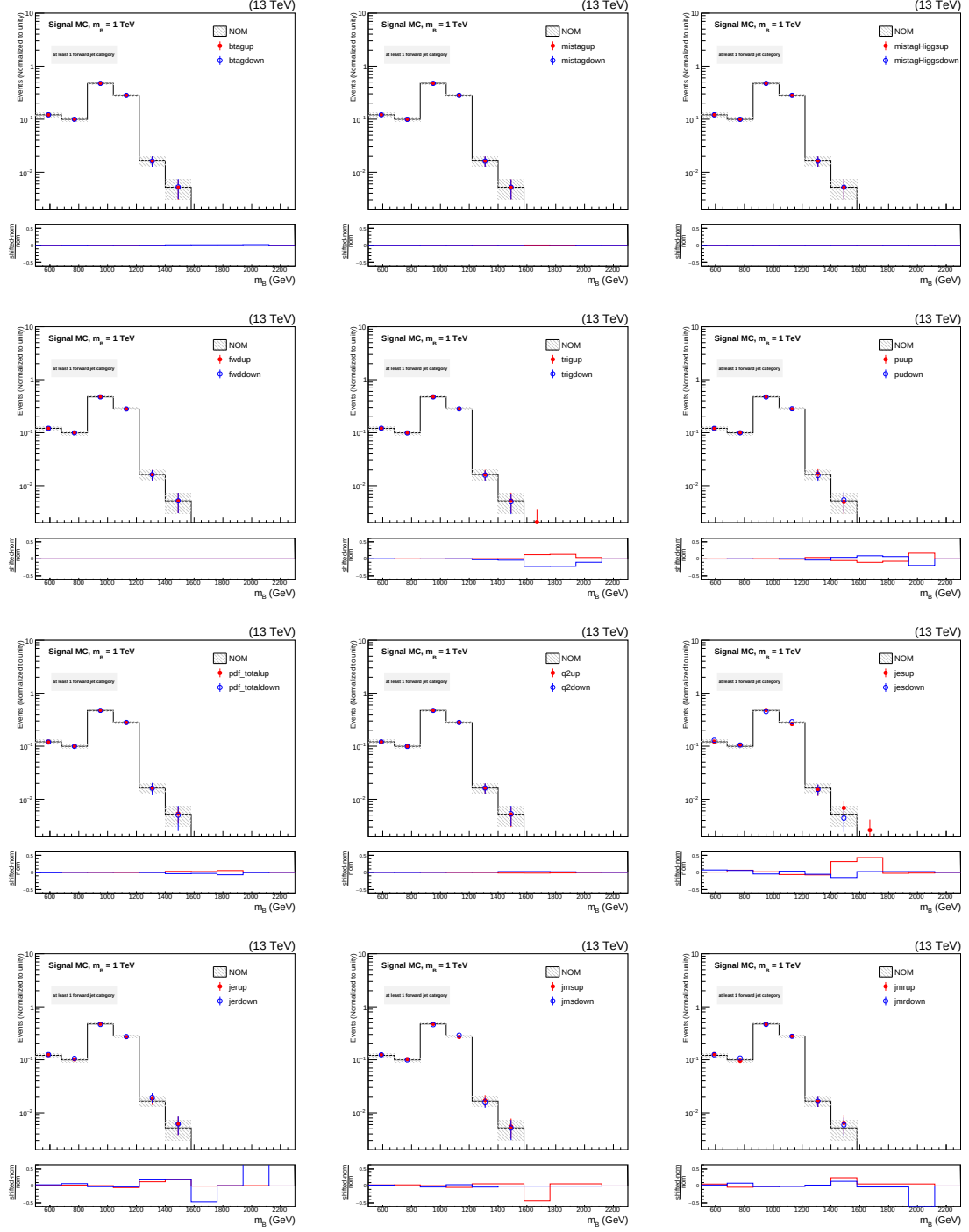


Figure 4.28: Systematic variation on the m_B distribution for signal Monte Carlo events, considering a mass value of 1000 GeV in the low mass analysis. The nominal distribution (black line) is plotted with its statistical uncertainty (hatched gray filling), the shift up (red) and shift down (blue) templates account for differences in both shape and normalization with respect to the nominal sample. Plots are shown for the category with at least one forward jet.

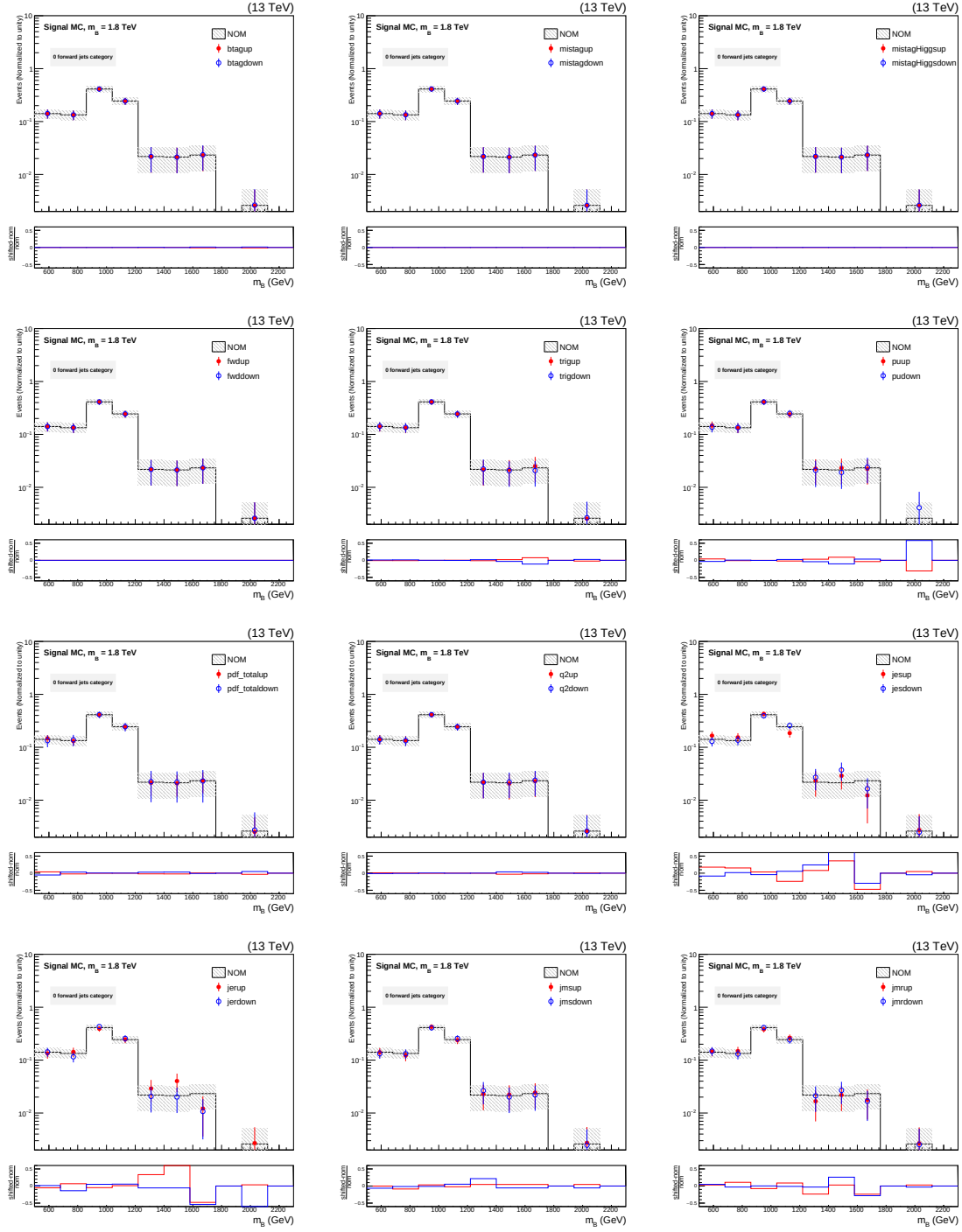


Figure 4.29: Systematic variation on the m_B distribution for signal Monte Carlo events, considering a mass value of 1800 GeV in the high mass analysis. The nominal distribution (black line) is plotted with its statistical uncertainty (hatched gray filling), the shift up (red) and shift down (blue) templates account for differences in both shape and normalization with respect to the nominal sample. Plots are shown for the category with exactly zero forward jets.

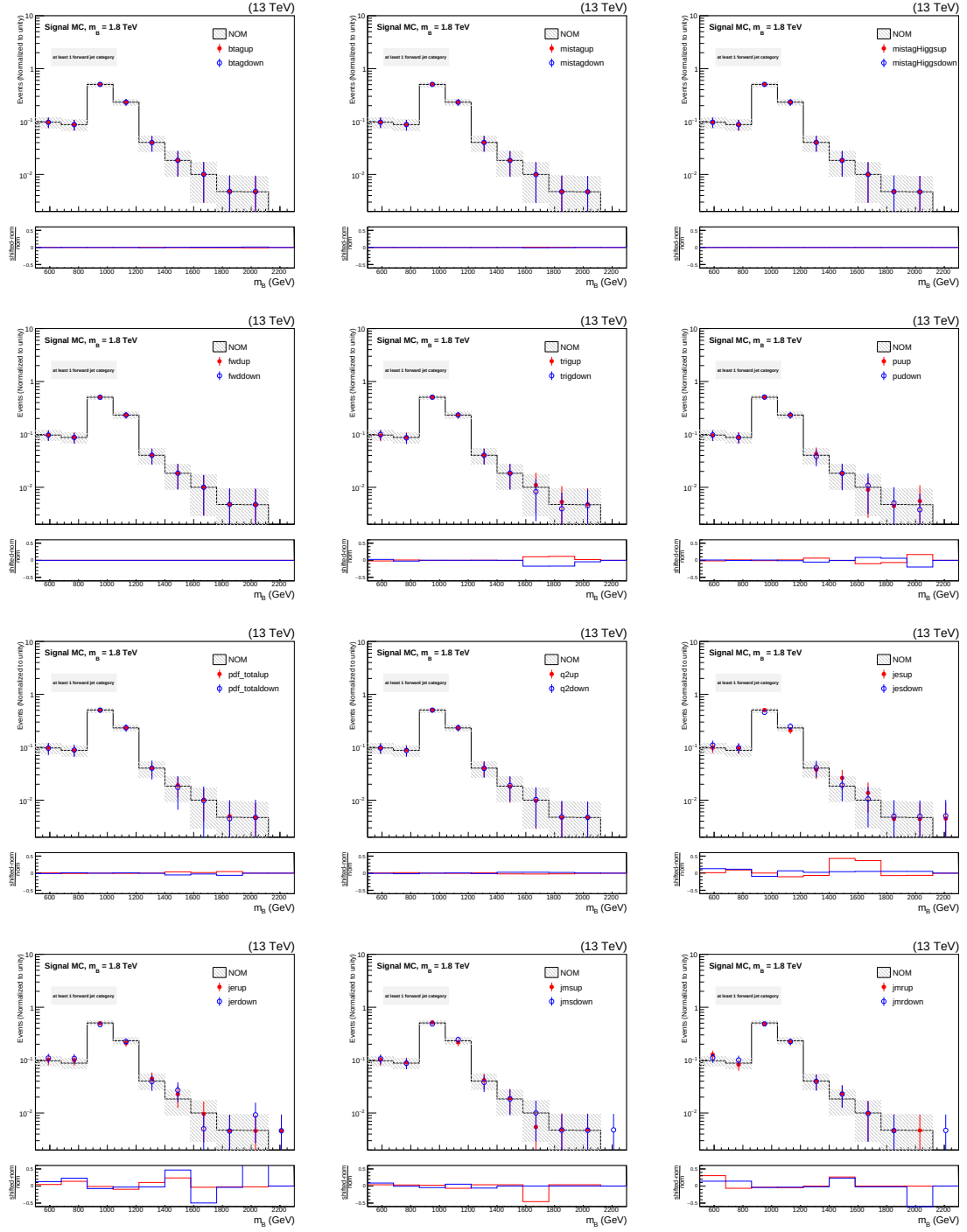


Figure 4.30: Systematic variation on the m_B distribution for signal Monte Carlo events, considering a mass value of 1800 GeV in the high mass analysis. The nominal distribution (black line) is plotted with its statistical uncertainty (hatched gray filling), the shift up (red) and shift down (blue) templates account for differences in both shape and normalization with respect to the nominal sample. Plots are shown for the category with at least one forward jet.

4.5 Results and interpretation

4.5.1 Interpretation model

The mechanism giving origin to the singly produced VLQ is an electroweak process. Thus, as opposite to the pair production interaction, that is a QCD process, its production cross-section depends not only on the mass of the VLQ, but also on the coupling with the SM particles and on their masses. In absence of interference between topologies involving different propagators, the total production cross-section for the process of single production and decay of a B quark can be expressed as:

$$\sigma(C_1, C_2, m_B, \Gamma_B) = C_1^2 C_2^2 \tilde{\sigma}_{FW}(m_B, \Gamma_B), \quad (4.5)$$

where Γ_B is the total width of the VLQ B, which depends on its mass, the masses of all its decay products and the couplings through which the B interacts with all the particles it can decay to, including C_1 and C_2 . A model independent approach is pursued, not specifying the mechanism giving origin to the large width of the B quark and therefore in this study Γ_B will be considered as a free parameter. The validity of Equation 4.5 holds in all width regimes, from Narrow Width (NW) approximation to Finite Width (FW) approximation. However, considering the asymptotic limit of the ratio Γ_B/m_B approaching 0, that is the NW scenario, production and decay can be factorised, resulting in a more compact relation for the cross-section, as:

$$\begin{aligned} \sigma_{\text{NWA}}(C_1, C_2, m_B, \Gamma_B) &= \sigma_P(C_1, m_B) \mathcal{B}(B \rightarrow \text{decay channel}) = \\ &= C_1^2 \tilde{\sigma}_P(m_B) \mathcal{B}(B \rightarrow \text{decay channel}). \end{aligned} \quad (4.6)$$

In the limit of asymptotically small resonance width, Equation 4.5 must reproduce Equation 4.6.

To compute the values of the $\tilde{\sigma}_{FW}$, simulations have been performed based on the VLQ FeynRules model (with non-zero masses for light quarks) [179] implemented in the context of NLO calculations in the framework of aMC@NLO. The simulation has been done using the CTEQ6L1 set of PDFs. The values of $\tilde{\sigma}_{FW}$ for representative B masses and Γ_B/m_B ratios are provided in Table 4.9. A graphical description is also provided in Figure 4.31, where the contours of $\tilde{\sigma}_{FW}$ are provided, for small values of Γ_B/m_B as well. It is important to observe that by fixing a specific value of the width/mass ratio, the coupling depends on the VLQ B hypothesis.

In this analysis, a simplified model assuming just one VLQ representation containing the VLQ B will be considered, being it the simplest extension of the SM. Under this assumption, the couplings of the B quarks with SM bosons and quarks can be parametrized as:

$$c_Z = \frac{e}{2c_w s_w} \kappa_Z, \quad c_W = \frac{e}{\sqrt{2} s_w} \kappa_W \quad \text{and} \quad c_H = \frac{M_B}{v} \kappa_H, \quad (4.7)$$

where $v = 246$ GeV is the vacuum expectation value of the Higgs field, c_w and s_w are the cosine and sine of the Weinberg angle θ_W and κ is a coupling strength which can be fixed to obtain the desired width. Numerically, $\frac{e}{2c_w s_w} = 0.370349$ and $\frac{e}{\sqrt{2} s_w} = 0.458486$.

Mass [GeV]	$\tilde{\sigma}_{FW}$ for $pp \rightarrow Bbq \rightarrow bHbq$ [pb]			
	1%	10%	20%	30%
700	29.3595	3.01323	1.42951	0.899324
800	20.3866	2.10406	0.997896	0.634127
900	14.4417	1.50731	0.718767	0.453577
1000	10.3654	1.09224	0.522658	0.331303
1100	7.57955	0.806876	0.386494	0.245694
1200	5.61999	0.601304	0.289949	0.184626
1300	4.18691	0.450901	0.219549	0.140713
1400	3.16007	0.341583	0.167281	0.108087
1500	2.39548	0.262492	0.129150	0.0836186
1600	1.83590	0.202940	0.100752	0.0651079
1700	1.41756	0.158415	0.0787629	0.0513501
1800	1.09829	0.124214	0.0620552	0.0407645
1900	0.853902	0.0977482	0.0494078	0.0324748
2000	0.666980	0.0777130	0.0393858	0.0262120

Table 4.9: Values of $\tilde{\sigma}_{FW}$ depending on the VLQ B mass and corresponding to several resonance width/mass ratios [164].

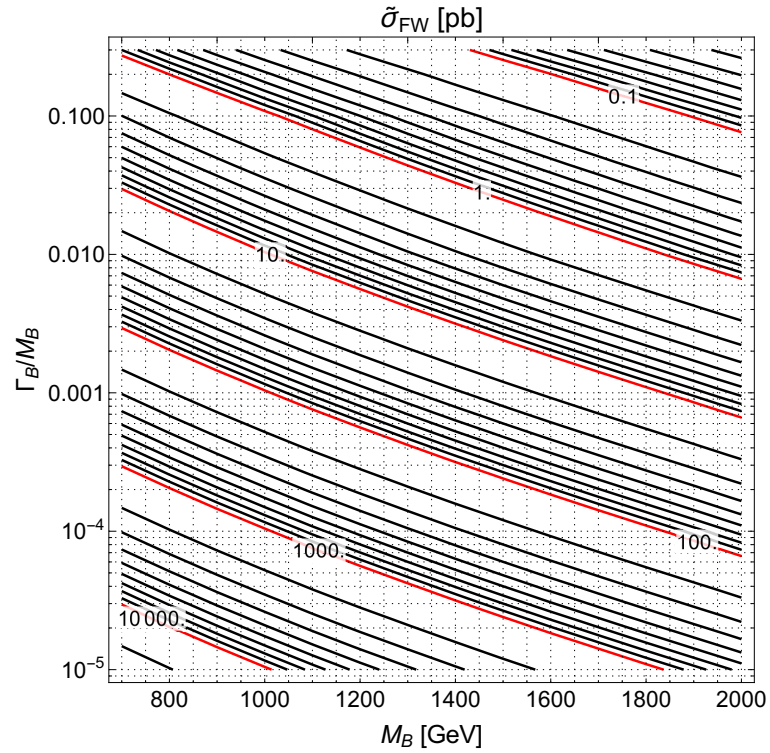


Figure 4.31: $\tilde{\sigma}_{FW}$ contours in the Γ_B/m_B plane obtained with *Wolfram Mathematica* [164, 180].

The strenght of this parametrisation is that fixing the representation of the B , the κ parameters determine with excellent approximation the width and the amount of mixing between B and the SM bottom quark. In this analysis we consider for a B singlet $\kappa_Z \simeq \kappa_H \simeq \kappa_W \simeq \kappa$, while for a $(T B)$ doublet with zero top Yukawa $\kappa_Z \simeq \kappa_H \simeq \kappa$ and $\kappa_W \simeq 0$. For the process under consideration we can identify $C_1 \equiv c_Z$ and $C_2 \equiv c_H$ in Equations 4.5 and 4.6.

The narrow width regime

To compute the cross-section values two alternative approaches can be followed, that are either fixing the width/mass ratio or fixing the parameter κ . In the former scenario, from Equation 4.5 it can be noted that fixing the ratio between the resonance width and its mass implies a modification of the VLQ branching ratios and mixing angles across the range of the B quark mass explored. These variations, assuming $\Gamma_B/M_B = 1\%$, are illustrated in Figure 4.32. If instead the latter approach is pursued, the κ values are fixed and the width of the VLQ B depends on the mass of the particle. The cross-sections for a $\kappa = 0.1$ (a value chosen in order to have a small width in the whole range of explored masses, are provided in Table 4.11 and the corresponding plot, showing the branching ratios and values of the width/mass ratio is presented in Figure 4.33.

The experimental results obtained in this analysis are interpreted following the first approach, where a fixed width/mass ratio for the B quark is assumed.

M(GeV)	$\hat{\sigma}_{NWA}$ (pb)	Singlet					Doublet			
		κ	$B \rightarrow Wt$	$B \rightarrow Zb$	$B \rightarrow Hb$	σ_{NWA} (pb)	κ	$B \rightarrow Zb$	$B \rightarrow Hb$	σ_{NWA} (pb)
700	31.30 $^{+28\%}_{-20\%}$	0.18	0.466	0.271	0.263	0.1631	0.25	0.499	0.501	0.5720
800	21.50 $^{+29\%}_{-21\%}$	0.16	0.474	0.276	0.260	0.0830	0.22	0.499	0.501	0.3003
900	15.10 $^{+30\%}_{-21\%}$	0.14	0.489	0.263	0.258	0.0451	0.19	0.500	0.500	0.1666
1000	10.80 $^{+31\%}_{-23\%}$	0.13	0.483	0.261	0.256	0.0257	0.17	0.500	0.500	0.0962
1100	7.85 $^{+32\%}_{-22\%}$	0.11	0.486	0.259	0.255	0.0153	0.16	0.500	0.500	0.0580
1200	5.77 $^{+33\%}_{-23\%}$	0.10	0.489	0.257	0.254	0.0094	0.15	0.500	0.500	0.0358
1300	4.29 $^{+34\%}_{-23\%}$	0.10	0.490	0.256	0.254	0.0059	0.13	0.500	0.500	0.0227
1400	3.23 $^{+34\%}_{-23\%}$	0.09	0.492	0.255	0.253	0.0038	0.12	0.500	0.500	0.0147
1500	2.45 $^{+35\%}_{-25\%}$	0.08	0.493	0.254	0.253	0.0025	0.12	0.500	0.500	0.0097
1600	1.86 $^{+36\%}_{-24\%}$	0.08	0.494	0.254	0.252	0.0017	0.11	0.500	0.500	0.0065
1700	1.44 $^{+37\%}_{-24\%}$	0.07	0.494	0.254	0.252	0.0011	0.10	0.500	0.500	0.0044
1800	1.11 $^{+37\%}_{-25\%}$	0.07	0.495	0.253	0.252	0.0008	0.10	0.500	0.500	0.0031

Table 4.10: Cross-sections for $pp \rightarrow Bbq$, with the ratio Γ_B/m_B fixed to 1% (NWA). The couplings and branching fractions in simplified models are calculated using the equations in the text. The uncertainties in the production cross-sections correspond to the halving and doubling of the QCD renormalization and factorization scales.

The finite width regime

In the context of simplified models, where B is the only new particle besides the SM, there is a limit to the maximum width it can obtain. Indeed, the total width is entirely given by the partial widths of its decays into the SM gauge bosons, but all couplings are bounded from above by the fact that κ_W and κ_H are combination of sines and cosines of mixing angles,

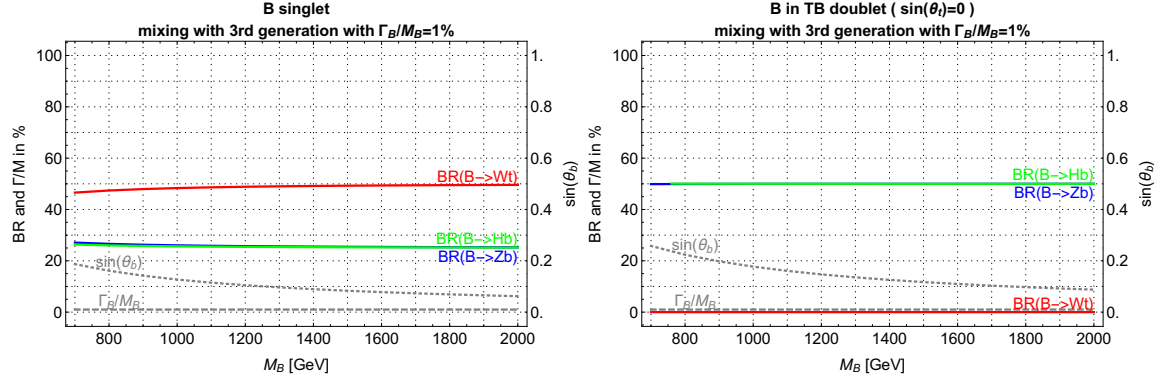


Figure 4.32: Branching ratios and mixing angles for different B masses and for a fixed Γ_B/M_B ratio of 1% [164].

M(GeV)	$\hat{\sigma}_{NWA}$ (pb)	Singlet					Doublet				
		Γ_B/M_B	$B \rightarrow Wt$	$B \rightarrow Zb$	$B \rightarrow Hb$	σ_{NWA} (pb)	Γ_B/M_B	$B \rightarrow Zb$	$B \rightarrow Hb$	σ_{NWA} (pb)	
700	31.3350 ^{+27.8%} _{-20%}	0.003	0.463	0.276	0.261	0.0475	0.002	0.513	0.487	0.0887	
800	21.4684 ^{+29%} _{-20.6%}	0.004	0.472	0.269	0.259	0.0323	0.002	0.510	0.490	0.0611	
900	15.0926 ^{+30.1%} _{-21.2%}	0.005	0.478	0.265	0.257	0.0225	0.003	0.507	0.493	0.0432	
1000	10.7633 ^{+31.1%} _{-21.7%}	0.006	0.483	0.261	0.256	0.0160	0.003	0.505	0.495	0.0310	
1100	7.85111 ^{+31.8%} _{-22%}	0.008	0.486	0.259	0.255	0.0116	0.004	0.504	0.496	0.0226	
1200	5.77030 ^{+33%} _{-22.6%}	0.009	0.488	0.257	0.254	0.0085	0.005	0.503	0.497	0.0167	
1300	4.29105 ^{+33.8%} _{-23%}	0.011	0.490	0.256	0.254	0.0063	0.005	0.502	0.498	0.0124	
1400	3.22579 ^{+34.5%} _{-23.4%}	0.012	0.492	0.255	0.253	0.0047	0.006	0.501	0.499	0.0094	
1500	2.44565 ^{+35.4%} _{-23.8%}	0.014	0.493	0.254	0.253	0.0036	0.007	0.501	0.499	0.0071	
1600	1.86256 ^{+35.9%} _{-24%}	0.016	0.494	0.253	0.253	0.0027	0.008	0.501	0.499	0.0054	
1700	1.43676 ^{+36.7%} _{-24.4%}	0.019	0.495	0.253	0.252	0.0021	0.009	0.500	0.500	0.0042	
1800	1.11291 ^{+37.3%} _{-24.7%}	0.021	0.496	0.252	0.252	0.0016	0.011	0.500	0.500	0.0032	
1900	0.863791 ^{+37.9%} _{-25%}	0.023	0.496	0.252	0.252	0.0013	0.012	0.500	0.500	0.0025	
2000	0.671185 ^{+38.4%} _{-25.3%}	0.026	0.497	0.251	0.252	0.0010	0.013	0.499	0.501	0.0020	

Table 4.11: cross-sections for $pp \rightarrow Bbj(NWA)$ calculated for $\kappa = 0.1$, PDF set cteq6l1, Γ_B/m_B ratio and branching fractions in simplified models calculated using equation 4.7.

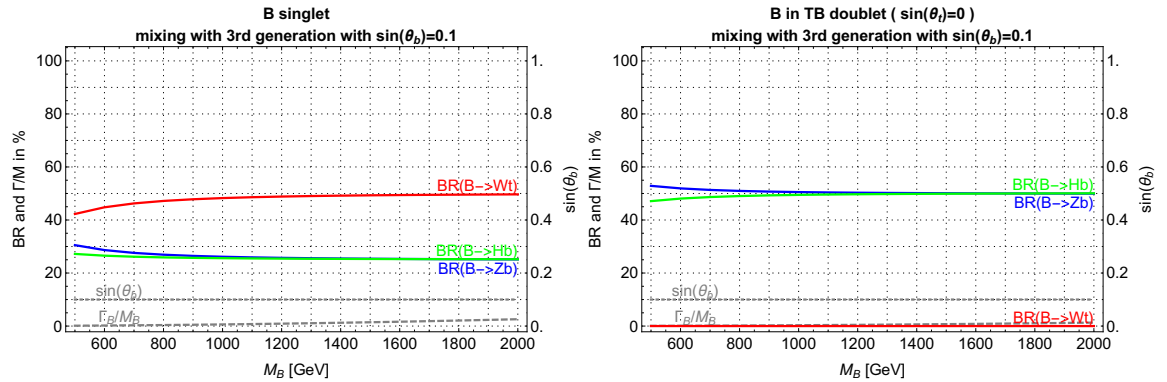


Figure 4.33: Branching ratios and mixing angles for different B masses and for a fixed $\kappa = 0.1$ [164].

which cannot be therefore larger than 1, while Z is a factor which depends on mixing angles and on the weak isospin of the VLQ and the combination can be at most 2 [181]. Therefore, increasing the Yukawa coupling between B and SM quarks has both the effect of increasing the width and modifying the relations between the branching ratios, but the partial widths are limited (for any given mass) by the upper limits on the couplings, i.e. the factors in front of the parameters in Equation 4.7.

The cross-section for a B singlet and part of a $(T \sim B)$ doublet are reported in Tables 4.12 and 4.13 and in Figure 4.34 the distributions of the branching ratios are displayed. Notice that the values of κ_Z and κ_H are not approximately the same when the NWA is relaxed.

M(GeV)	W/M = 10%			W/M = 20%			W/M = 30%		
	κ_Z	κ_H	$\sigma_S(\text{fb})$	κ_Z	κ_H	$\sigma_S(\text{fb})$	κ_Z	κ_H	$\sigma_S(\text{fb})$
700	0.487261	0.62277	307.596	—	—	—	—	—	—
800	0.44878	0.528726	171.53	0.485025	0.788301	211.228	—	—	—
900	0.409231	0.461214	98.402	0.498147	0.676002	149.368	0.431131	0.867853	116.365
1000	0.373846	0.40985	58.0133	0.478245	0.595075	95.7715	0.495623	0.752328	104.212
1100	0.343142	0.369233	35.4581	0.450946	0.532932	61.1076	0.497127	0.668182	74.2127
1200	0.316633	0.336203	22.1997	0.423111	0.483305	39.5011	0.480977	0.602818	50.5651
1300	0.293673	0.308759	14.1746	0.396868	0.44257	25.8969	0.459371	0.550063	34.3516
1400	0.273671	0.285561	9.25086	0.372816	0.408437	17.1995	0.43664	0.50635	23.4289
1500	0.256128	0.265676	6.18716	0.351006	0.379364	11.6572	0.414483	0.469414	16.1133
1600	0.240641	0.248429	4.20074	0.331295	0.354272	8.03843	0.393561	0.437723	11.191
1700	0.226879	0.233319	2.9024	0.313479	0.332375	5.59072	0.374094	0.410191	7.90582
1800	0.21458	0.219967	2.02854	0.297345	0.313086	3.94227	0.356102	0.386023	5.64646
1900	0.203525	0.20808	1.43181	0.282697	0.295955	2.82469	0.339519	0.364621	4.0648
2000	0.19354	0.197426	1.02678	0.269356	0.280633	2.0366	0.324243	0.345523	2.97735

Table 4.12: Cross-sections for $pp \rightarrow Bbj \rightarrow Hbbj$ for different values of the width/mass ratio and for a singlet B .

M(GeV)	W/M = 10%			W/M = 20%			W/M = 30%		
	κ_Z	κ_H	$\sigma_D(\text{fb})$	κ_Z	κ_H	$\sigma_D(\text{fb})$	κ_Z	κ_H	$\sigma_D(\text{fb})$
700	—	—	—	—	—	—	—	—	—
800	—	—	—	—	—	—	—	—	—
900	0.499477	0.723075	360.296	—	—	—	—	—	—
1000	0.486518	0.620216	224.997	—	—	—	—	—	—
1100	0.458695	0.548637	139.889	—	—	—	—	—	—
1200	0.429508	0.493988	88.1873	0.483951	0.790979	138.418	—	—	—
1300	0.402032	0.450253	56.4907	0.499617	0.693135	100.671	—	—	—
1400	0.37699	0.414187	36.93	0.487782	0.624612	68.8572	0.423936	0.874684	65.9032
1500	0.354409	0.383802	24.7228	0.468892	0.571304	47.1771	0.494133	0.759185	59.9022
1600	0.334097	0.35778	16.7942	0.448291	0.527787	32.6667	0.499145	0.686123	44.2288
1700	0.31581	0.335203	11.6074	0.42787	0.491219	22.7493	0.48937	0.630432	31.9569
1800	0.299304	0.315403	8.11421	0.408361	0.459873	16.0421	0.474522	0.585174	23.0401
1900	0.284358	0.29788	5.72801	0.390028	0.432602	11.488	0.457963	0.547107	16.6507
2000	0.270776	0.282252	4.10792	0.372933	0.408597	8.2761	0.441108	0.514371	12.2117

Table 4.13: Cross-sections for $pp \rightarrow Bbj \rightarrow Hbbj$ for different values of the width/mass ratio and for a B part of a $(T \sim B)$ doublet.

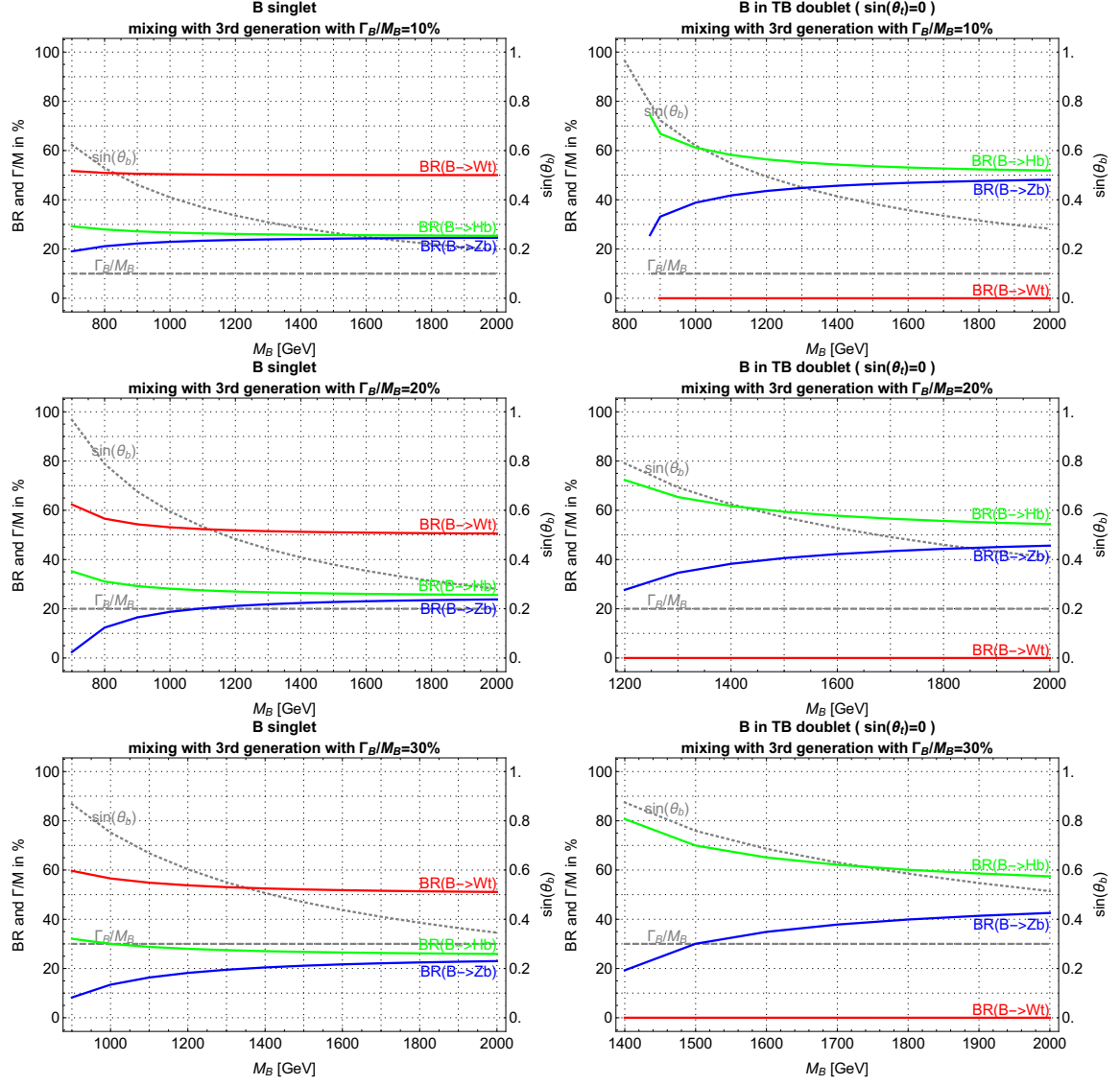


Figure 4.34: Branching ratios and sine of mixing angle for B singlet (top row) and part of a $(T B)$ doublet (bottom row) for fixed width/mass ratio (10%, 20% and 30% respectively from left to right). Notice the different minimum available mass when imposing different width/mass ratios, under which the sine of the mixing angle cannot increase further [164].

Alternatively, the same relations on the κ values used in the NW case, $\kappa_W = \kappa_Z = \kappa_H \equiv \kappa$ for singlet and $\kappa_W = 0$, $\kappa_Z = \kappa_H \equiv \kappa$ for doublet, can be imposed in the FW case. This approach is characterized by a freedom to choose the values reproducing the desired total width, but, on the other hand, it is limited by the fact that the assumption of the simplified model has to be relaxed: unspecified new physics (in the form of mixings either in the quark or boson sectors due to further VLQs or new gauge bosons or scalar fields) entering into play has to be assumed to generate the needed couplings. In Tables 4.14 and 4.15 the cross cross-sections obtained by imposing both the width/mass ratio and the conditions on the κ parameters are shown, respectively for the singlet and doublet scenarios. In Figure 4.35 the BRs are shown for different values of the width.

M(GeV)	W/M = 10%		W/M = 20%		W/M = 30%	
	κ	$\sigma_S(\text{fb})$	κ	$\sigma_S(\text{fb})$	κ	$\sigma_S(\text{fb})$
700	0.588255	400.005	0.831919	759.07	1.01889	1074.47
800	0.508297	203.368	0.718841	385.808	0.880396	551.623
900	0.447949	111.218	0.633496	212.14	0.775871	301.209
1000	0.400684	63.6943	0.566652	121.915	0.694004	173.879
1100	0.362606	38.1862	0.512802	73.1646	0.628052	104.649
1200	0.331242	23.5837	0.468447	45.4884	0.573728	65.1709
1300	0.304941	14.9075	0.431252	29.0347	0.528174	41.87
1400	0.282557	9.65497	0.399596	18.913	0.489403	27.496
1500	0.263266	6.41877	0.372315	12.6326	0.455991	18.4028
1600	0.246465	4.33713	0.348555	8.61297	0.42689	12.5231
1700	0.231697	2.98504	0.327669	5.93656	0.401311	8.70839
1800	0.218611	2.07959	0.309163	4.15573	0.378646	6.14235
1900	0.206934	1.46392	0.292649	2.95982	0.358421	4.37725
2000	0.196449	1.04744	0.277821	2.12341	0.34026	3.17964

Table 4.14: Cross-sections for $pp \rightarrow Bbj \rightarrow Hbbj$ for different values of the width/mass ratio and for a B with the condition $\kappa_W = \kappa_Z = \kappa_H \equiv \kappa$.

M(GeV)	W/M = 10%		W/M = 20%		W/M = 30%	
	κ	$\sigma_D(\text{fb})$	κ	$\sigma_D(\text{fb})$	κ	$\sigma_D(\text{fb})$
700	0.801485	1378.43	1.13347	2615.76	1.38821	3702.6
800	0.698615	725.711	0.98799	1376.73	1.21004	1968.47
900	0.619369	406.499	0.87592	775.363	1.07278	1100.91
1000	0.556394	236.822	0.786859	453.294	0.963702	646.503
1100	0.505117	143.792	0.714344	275.507	0.874889	394.063
1200	0.462541	89.6668	0.654132	172.95	0.801144	247.783
1300	0.426615	57.1066	0.603324	111.223	0.738918	160.391
1400	0.395888	37.2063	0.55987	72.8829	0.685697	105.958
1500	0.369304	24.8547	0.522274	48.9153	0.639653	71.2586
1600	0.346076	16.8604	0.489425	33.4821	0.59942	48.6825
1700	0.325604	11.642	0.460473	23.1532	0.563962	33.9635
1800	0.307424	8.13282	0.434764	16.2522	0.532475	24.0214
1900	0.291172	5.73837	0.411779	11.602	0.504324	17.158
2000	0.276554	4.11384	0.391107	8.33981	0.479006	12.4882

Table 4.15: Cross-sections for $pp \rightarrow Bbj \rightarrow Hbbj$ for different values of the width/mass ratio and for a B with the condition $\kappa_W = 0, \kappa_Z = \kappa_H \equiv \kappa$.

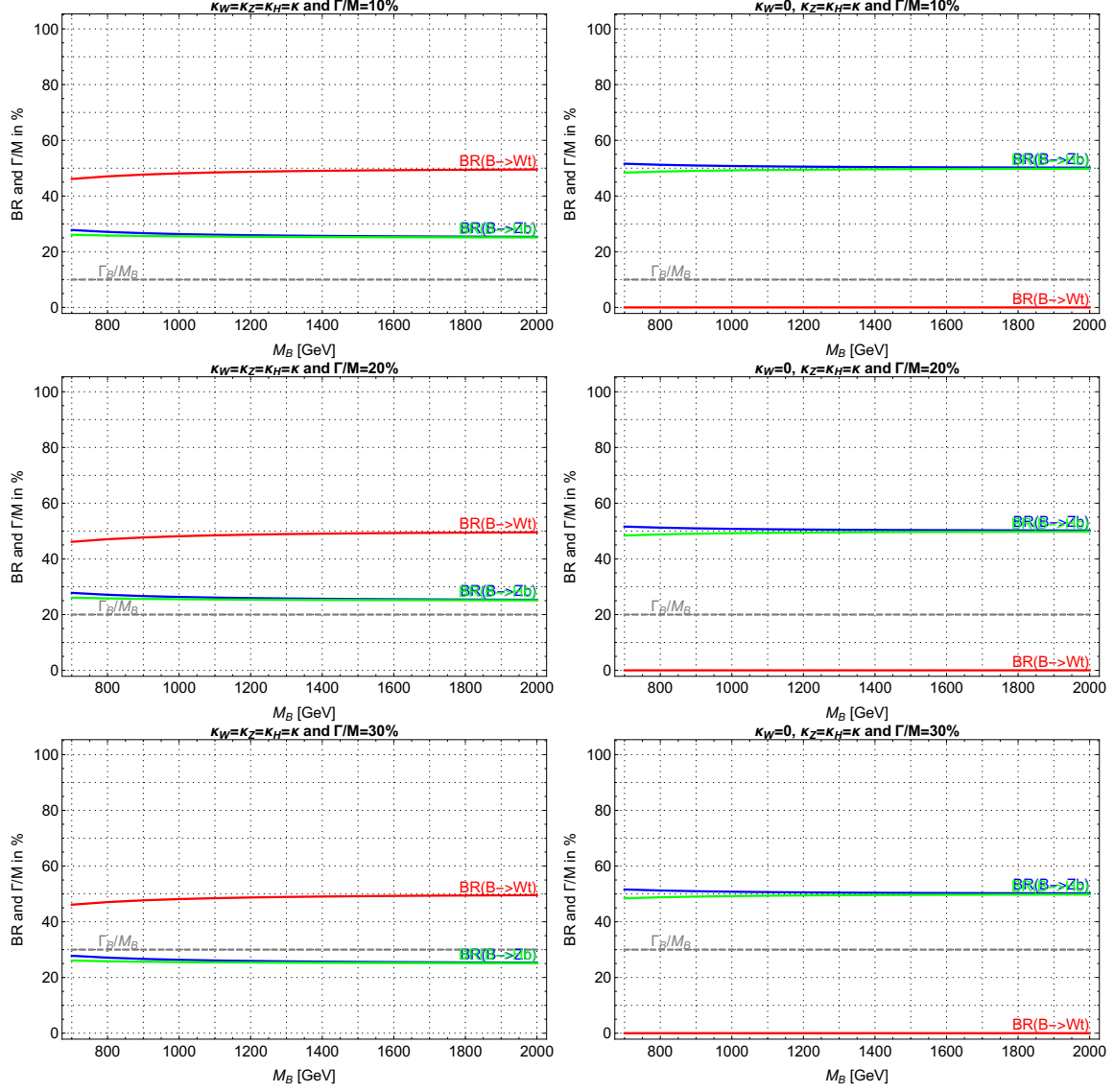


Figure 4.35: Branching ratios B with the condition $\kappa_W = \kappa_Z = \kappa_H \equiv \kappa$ (left column) and $\kappa_W = 0, \kappa_Z = \kappa_H \equiv \kappa$ (right column) for fixed width/mass ratio (10%, 20% and 30% respectively from top to bottom) [164].

4.5.2 Signal extraction

In the context of the search for the singly produced VLQ B described in this thesis, a potential signal would appear as a localized excess of data over the SM background in the spectrum of the invariant mass between the b quark and the Higgs boson, m_{bH} , shown in Figure 4.36. To perform the signal extraction, a binned maximum likelihood fit, as described in Appendix C is evaluated simultaneously in the two forward-jet based categories, and separately for the low- and high-mass analysis. The fitted distributions are presented in Figure 4.37 and the final event yields are listed in Table 4.16.

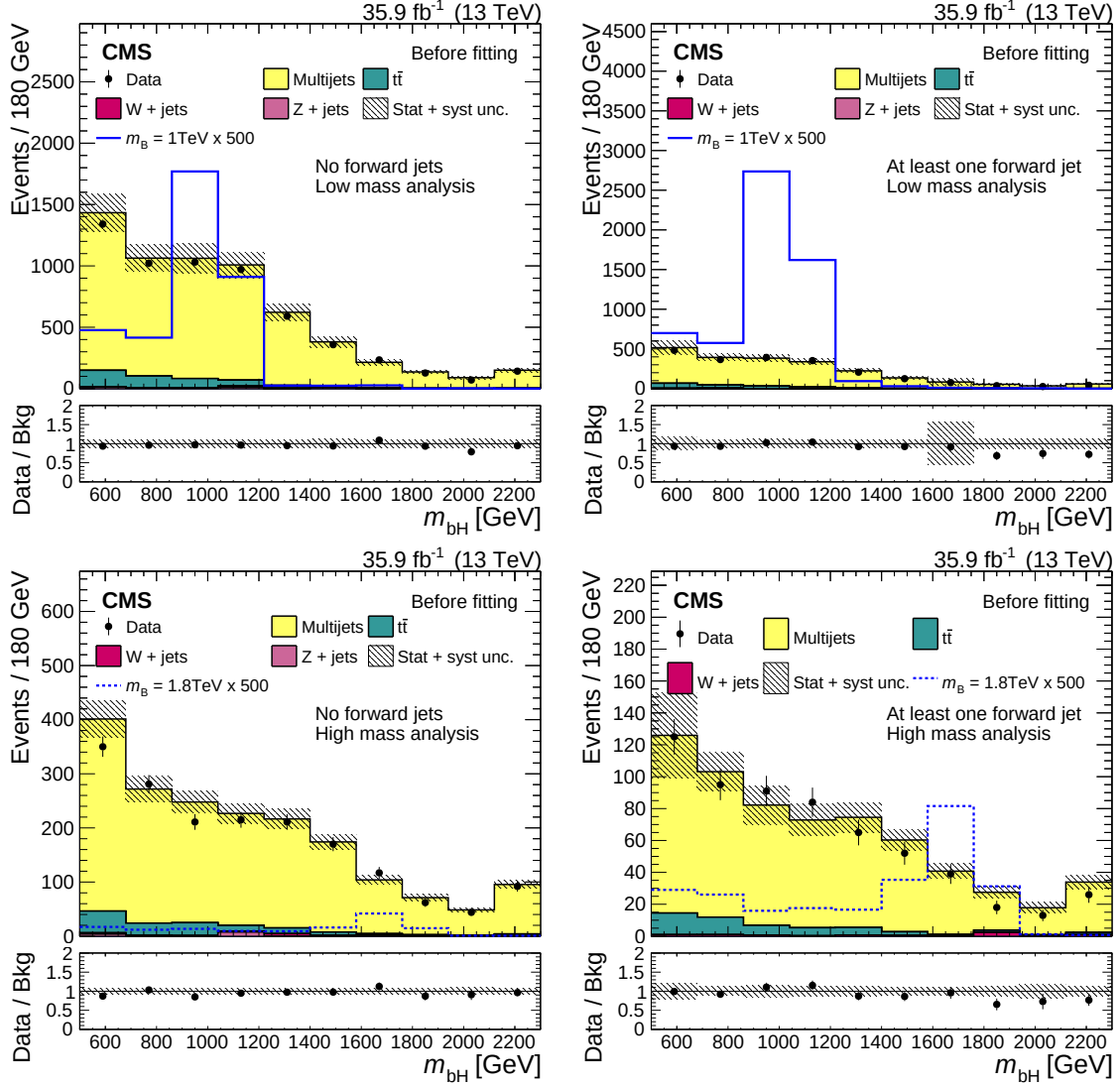


Figure 4.36: Distribution in the reconstructed B quark mass, after applying all selections to events with no forward jets (left) and to events with at least one forward jet (right), compared to the background distributions estimated before fitting. The upper and lower plots refer to the low and high-mass m_B analyses, respectively. The expectations for signal MC events are given by the blue histogram lines. Different contributions to background are indicated by the color filled histograms. The grey-hatched error band shows total uncertainties in the background expectation. The ratios of observations to background expectations are given in the lower panels, together with the total uncertainties prior to fitting, indicated by the grey-hatched band.

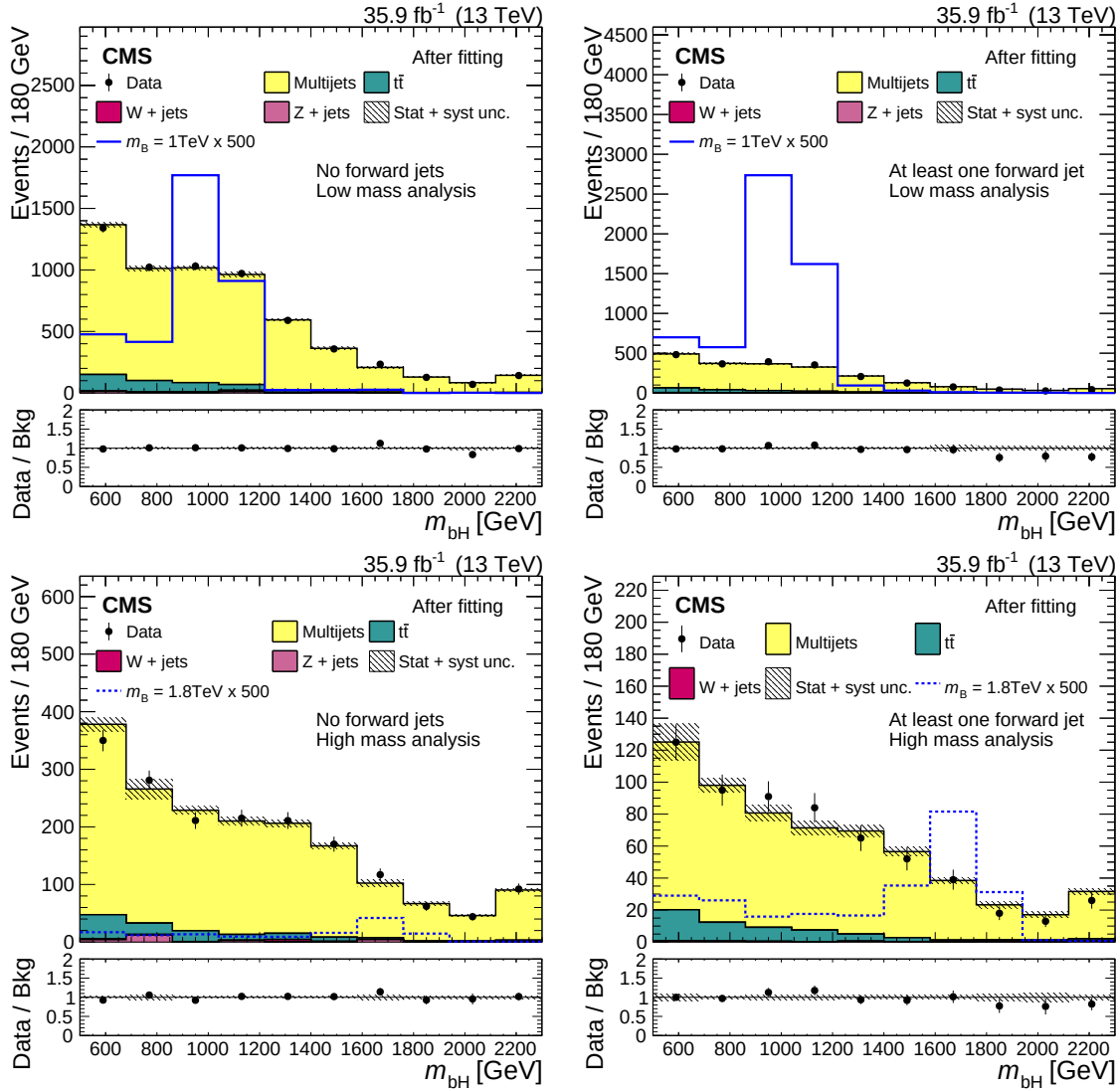


Figure 4.37: Distribution in the reconstructed B quark mass, after applying all selections to events with no forward jets (left) and to events with at least one forward jet (right), compared to the background distributions estimated after fitting. The upper and lower plots refer to the low and high-mass m_B analyses, respectively. The expectations for signal MC events are given by the blue histogram lines. Different contributions to background are indicated by the color filled histograms. The grey-hatched error band shows total uncertainties in the background expectation. The ratios of observations to background expectations are given in the lower panels, together with the total uncertainties prior to fitting, indicated by the grey-hatched band.

category	source	low mass yields	high mass yields
zero forward jets category	Multijet	5416 ± 60	1612 ± 24
	$t\bar{t}$	394 ± 46	117 ± 18
	W + jets	29 ± 13	10.5 ± 4.3
	Z + jets	43 ± 15	23 ± 23
	Backgrounds	5882 ± 42	1762 ± 26
	Data	5886 ± 77	1753 ± 42
	$m_B = 1000$ GeV	29.3 ± 1.1	7.07 ± 0.53
	$m_B = 1800$ GeV	4.88 ± 0.15	4.12 ± 0.14
at least one forward jet category	Multijet	1938 ± 23	549 ± 10
	$t\bar{t}$	163 ± 20	58 ± 17
	W + jets	11.5 ± 4.2	4.3 ± 1.4
	Z + jets	2 ± 10	0.0 ± 0.0
	Backgrounds	2115 ± 21	612 ± 15
	Data	2107 ± 46	608 ± 25
	$m_B = 1000$ GeV	46.4 ± 1.4	8.51 ± 0.58
	$m_B = 1800$ GeV	9.29 ± 0.21	7.78 ± 0.19

Table 4.16: Observed and expected post-fit event yields in the signal region in both low mass and high mass ranges. The multijet background expectation yields are derived from data, while the yields for the other sources of background are derived from simulation.

Observation is in agreement with the background-only hypothesis in all the categories defined in the analysis. Therefore upper limits are placed on the product of the cross-section multiplied for the branching fraction of a B quark decaying to Hb , associated to another b quark and a light-flavor quark, as a function of m_B . Exclusion limits at 95% confidence level are calculated with a profile likelihood ratio as test statistic, as described in Appendix C. The two forward-jet based categories are considered in their combination, that brought an increase in the final sensitivity of the analysis up to the 20% with respect to the configuration in which only events with at least one forward jet were considered for the search.

The systematic uncertainties, described in Section 4.4, are treated as nuisance parameters in the fit. The one affecting only the normalization are modeled using a log-normal distribution, if instead they contribute to the shape of the m_{bH} as well, the approach of applying a morphing technique is used, as discussed in Section 4.4.

4.5.3 Results

As shown in Figure 4.37, no excess of events over the expected SM background is observed. Therefore, 95% CL upper limits on the production cross-section times the branching ratio are set in both the NW and the FW regimes. Expected and observed limits are computed and they are graphically shown in Figure 4.38.

The experimental results are confronted with the theoretical predictions in case of a singlet and a doublet, obtained by considering the cross-section values reported respectively in Table 4.10 for the former, 4.14 and 4.15 for the latter model.

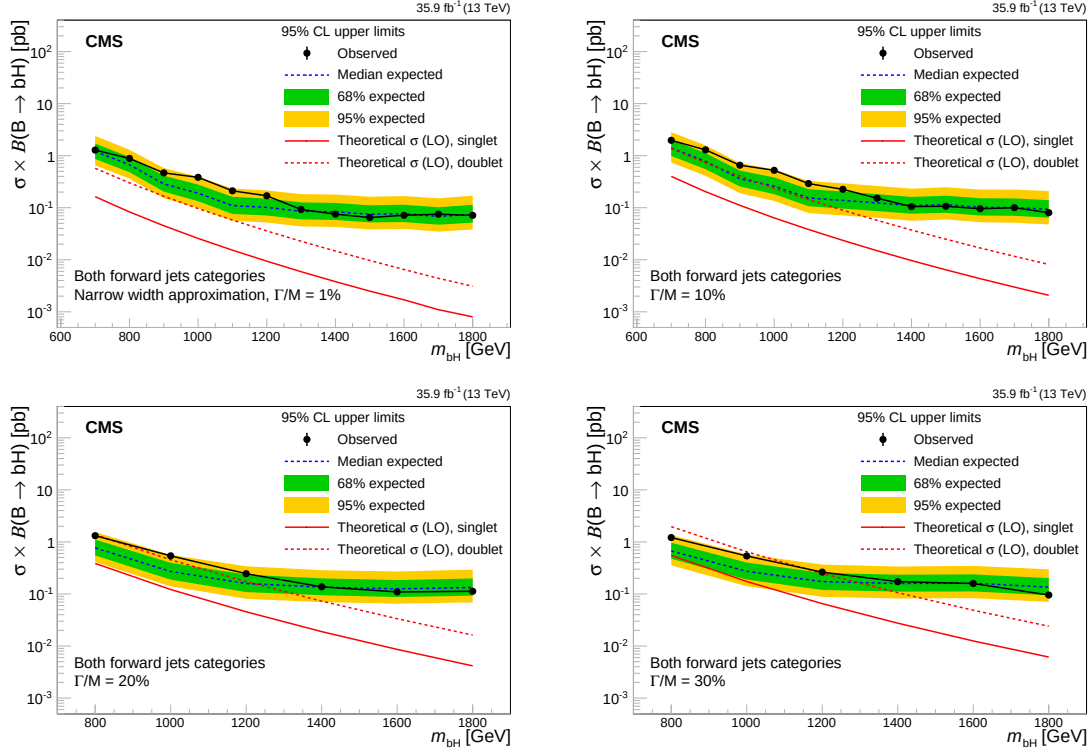


Figure 4.38: The observed (dots) and expected (dashed line) median 95% CL upper limits on the product of the B quark production cross-section and branching fraction as a function of the signal mass, assuming narrow-width resonances (upper-left) and widths of 10% (upper-right), 20% (lower-left), and 30% (lower-right) of the resonance mass for the B quark. The results are shown for the combination of 0 and > 0 forward-jet categories. The continuous red curves correspond to the theoretical expectations for singlet and doublet models.

Chapter 5

Search for Dark Matter with semi-visible jets

“ *The beginnings and ends of shadow lie between the light and darkness and may be infinitely diminished and infinitely increased. Shadow is the means by which bodies display their form. The forms of bodies could not be understood in detail but for shadow.* ”

Leonardo da Vinci, 15th-16th century AC

Despite the huge experimental effort, DM has not been observed yet. The ATLAS and CMS Collaborations carried a variety of searches looking for signs of DM recoiling against visible particles, probing the hypothesis of DM being made of Weakly Interacting Massive Particles, WIMPs, that were discussed in Section 1.3.1). Unfortunately, none of these searches resulted in an experimental confirmation of WIMP scenarios. Alternative and more complex models have been developed during the last decade to answer this lack of evidence, with the Hidden Valley models [32–34] being one of the most popular examples. These models are built relaxing the assumption behind the WIMP scenario and predict DM made of a multitude of particles interacting through new fundamental forces.

In this thesis, I investigate a class of Hidden Valley models that predicts the existence of a dark sector populated by dark quarks, interacting among themselves via a *dark strong force* and a *dark weak force* with particles from the SM. Dark quarks shower and hadronize in the dark sector, similarly to SM quarks. This hadronization leads to the production of stable and unstable dark hadrons and the subsequent formation of the so-called *semi-visible jets*. In these exotic objects, the unstable dark hadrons can decay to SM particles, which constitute the visible energy fraction of the jet, while stable dark hadrons remain invisible and constitute the invisible energy fraction of the jet. The objects take the name of semi-visible jets, being only partially visible to the detector. These events are characterized by the production of missing momentum along the visible jet, instead of recoiling against it as in the case of WIMPs.

In this chapter, the search for semi-visible jets is presented. The process here analyzed is $q\bar{q} \rightarrow Z' \rightarrow \chi\bar{\chi}$, whose Feynman diagram is shown in Figure 5.1.

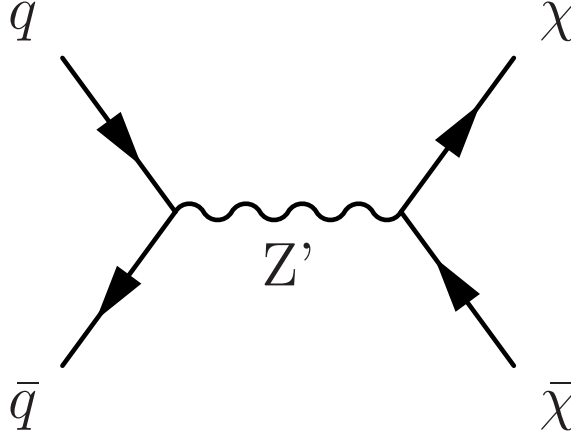


Figure 5.1: The leading-order Feynman diagram for the production of two semi-visible jets.

The dataset used in the analysis, and signal and background simulations, are described in Section 5.1. The trigger strategy and the analysis selection, along with its optimization, are discussed in Section 5.2. A dedicated tagger to identify semi-visible jets and distinguish them from multijet events is built, as detailed in Section 5.3. To extrapolate a suitable working point for the tagger, optimization studies have been performed, as detailed in Section 5.5. In Section 5.4 the data-driven method used to estimate the multijet background is described and in Section 5.6 the systematics uncertainties taken into account are listed. Finally, the results are presented and discussed in Section 5.7.

5.1 Datasets and simulated events

5.1.1 Datasets

The physics analysis here discussed is performed on pp collision events delivered by LHC at a center-of-mass energy $\sqrt{s} = 13$ TeV in 2016, 2017 and 2018, and recorded with the CMS experiment. Table 5.1 reports the different periods of the LHC operations for each year with the corresponding integrated luminosity collected by CMS. In total, the data used for the search hereby presented corresponds to 137 fb^{-1} .

5.1.2 Signal simulation

The signal model employed in this thesis is implemented with PYTHIA 8.2 [163] using the Hidden Valley module [182]. Simulated signal samples are produced with different mediator mass values. Its choice determines the production cross-section $\sigma_{Z'}$, whose product times the branching ratio $\mathcal{B}(Z' \rightarrow \chi\bar{\chi})$ corresponds to the production rate for the semi-visible jets. Table 5.2 reports the cross-sections $\sigma_{Z'}$ for the signal processes depending on the value of the mediator mass.

On the other hand, the remaining parameters, such as the invisible ratio r_{inv} , the characteristic mass scale for the dark hadrons m_{dark} , and the dark strong coupling α_{dark} ,

Year	Datasets	Integrated luminosity (pb ⁻¹)
2016	Run B	5750
	Run C	2573
	Run D	4243
	Run E	4025
	Run F	3104
	Run G	7576
	Run H	8651
2017	Run B	4794
	Run C	9631
	Run D	4248
	Run E	9314
	Run F	13535
2018	Run A	14027
	Run B	7061
	Run C	6895
	Run D	31710

Table 5.1: List of primary datasets of pp collision data used in the analysis. They are produced at $\sqrt{s} = 13$ TeV in 2016, 2017 and 2018 and collected by CMS.

are fixed.

The invisible ratio r_{inv} relates to the fraction of stable dark hadrons produced in the dark parton shower, and it ranges from 0 to 1. The amount of reconstructable energy in the jet depends on this parameter.

The value of the dark strong coupling constant, α_{dark} , has a direct impact on the multiplicity of dark hadrons. Its effect on the distribution of dark hadron multiplicity is visible in Figure 5.2. This observable is also affected by the dark quark mass degeneracy and the number of flavors. The implication of the former factor can be understood considering that the fragmentation of the semi-visible jet is suppressed by a factor $T = e^{-(4\pi|m_{x_1}^2 + m_{x_2}^2|/\Lambda_{\text{dark}}^2)}$, being Λ_{dark} the scale of the SU(2) dark force. The unstable hadron multiplicity is directly proportional to the number of flavors N_f , while the number of stable hadrons is proportional to $N_f(N_f - 1)$. This has a direct impact on the jet particle multiplicity, but also on the fraction of visible and invisible particles within the jet itself.

The parameter m_{dark} , which is the dark hadron mass scale, affects the number of dark hadrons that are produced during the dark showering. Thus, it impacts the multiplicity of the dark jet as well. This parameter also alters the relative p_T of the states produced in the shower, which manifests in the detector as the amount of missing and visible energy in the final state. The p_T spectrum of shower constituents is also affected by α_{dark} , which depends on the dark QCD force scale Λ_{dark} as:

$$\alpha_{\text{dark}}(\Lambda_{\text{dark}}) = \frac{\pi}{(b_0 \log(\frac{Q}{\Lambda_{\text{dark}}}))}, \quad (5.1)$$

with $b_0 = (11N_c - 2N_f)/6$, where N_c is the number of colors and N_f the number of flavors, and the momentum Q is set to 1 TeV. This parameter plays a pivotal role in determining the multiplicity and the p_T distribution of the dark hadrons, and hence the

$m_{Z'}$ [GeV]	$\sigma_{Z'}$ [pb]	$m_{Z'}$ [GeV]	$\sigma_{Z'}$ [pb]
500	71.37	2800	0.02447
600	33.22	2900	0.01969
700	18.17	3000	0.01550
800	10.75	3100	0.01209
900	7.04	3200	0.009518
1000	4.612	3300	0.007494
1100	3.021	3400	0.006148
1200	2.039	3500	0.005036
1300	1.459	3600	0.004047
1400	1.057	3700	0.003252
1500	0.7700	3800	0.002613
1600	0.5667	3900	0.002100
1700	0.4327	4000	0.001688
1800	0.3304	4100	0.001356
1900	0.2479	4200	0.001090
2000	0.1849	4300	0.0008757
2100	0.1384	4400	0.0007037
2200	0.1044	4500	0.0005655
2300	0.08120	4600	0.0004544
2400	0.06360	4700	0.0003652
2500	0.04980	4800	0.0002934
2600	0.03890	4900	0.0002358
2700	0.03040	5000	0.0001895

Table 5.2: Cross-sections as a function of $m_{Z'}$.

size of the originated jet. If Λ_{dark} is larger than m_{dark} , the number of dark hadrons in the dark shower increases. The approximate relationship determining the value of Λ_{dark} that maximizes the dark hadron multiplicity $\Lambda_{\text{dark}}^{\text{peak}}$ can be obtained with Monte Carlo simulation, as shown in Figure 5.2. For these studies the mass of the mediator Z' was set to 3 TeV in the determination of the scale, being the dependence on the $m_{Z'}$ found to be negligible. Λ_{dark} can be described as a function of m_{dark} according to the following relation:

$$\Lambda_{\text{dark}}^{\text{peak}} = 3.2 m_{\text{dark}}^{0.8}. \quad (5.2)$$

The approximate relationship determining the value of Λ_{dark} that maximizes the dark hadron multiplicity, $\Lambda_{\text{dark}}^{\text{peak}}$, obtained with Monte Carlo simulation, is shown in Figure 5.2. for a m_{dark} of 20 GeV is 40 GeV, for corresponding $\alpha_{\text{dark}}^{\text{peak}} \sim 0.4$.

In this thesis, signal data are generated by fixing the parameters m_{dark} , r_{inv} and α_{dark} to 20 GeV, 0.3, and $\alpha_{\text{dark}}^{\text{peak}}$, respectively, and varying the values of mediator mass, $m_{Z'}$ between 1500 GeV to 4500 GeV in steps of 100 GeV.

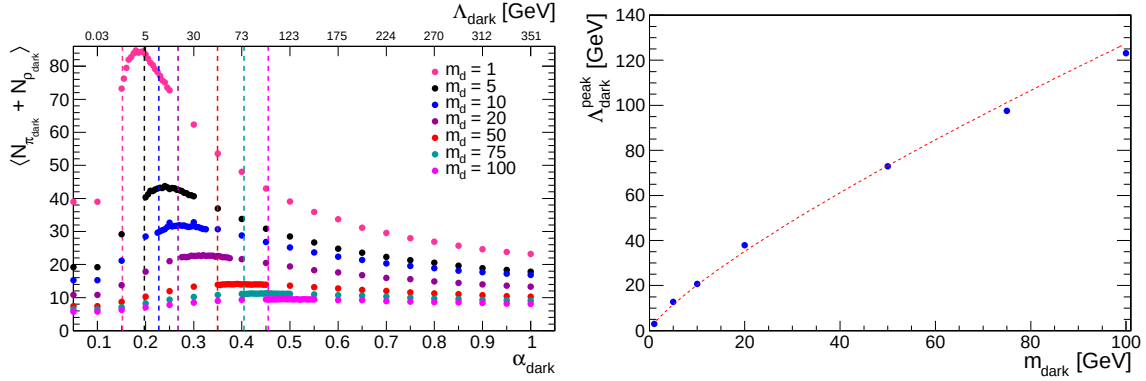


Figure 5.2: Dark hadrons multiplicity (both pseudoscalars and vectors) depending on the value of α_{dark} (left). In the upper axis the scale is non-linear to show the value of Λ_{dark} which corresponds to each labeled α_{dark} value. Empirical fit for the value of Λ_{dark} maximizing the dark hadrons multiplicity for several m_{dark} scenarios (right).

Generated events are then passed through the full simulation of the CMS detector, which uses GEANT 4, as well as dedicated simulations of detector electronics, and then reconstructed using the same algorithms that are applied to observed data.

5.1.3 Background simulation

Production of semi-visible jets would manifest in the CMS detector with the presence of jets accompanied by missing momentum in the collision products. Such a final state can characterize some SM processes which can emulate the production of DM. In particular, the main source background affecting this fully hadronic analysis is the multijet event production due to QCD processes. The top quark pair production and the vector boson plus jets processes, $V + \text{jets}$, represent another source of background for this search.

Simulated samples describing $t\bar{t}$ and $V + \text{jets}$ backgrounds are generated with MADGRAPH_AMC@ NLO 2.2.2 at LO and hadronized with PYTHIA 8.2. The event generation includes up to four additional partons for $t\bar{t}$ and three additional partons for W

+ jets and Z + jets events. The multijet background data are generated and hadronized with PYTHIA 8.2. For 2016 samples, LO NNPDF3.0 sets of PDFs are used, while for 2017 and 2018 samples NNLO NNPDF3.0 are applied. All events pass through the full CMS detector simulation using GEANT 4, which reflects the state of the detector for each year of data-taking, including any upgrade that was installed. Details about the generation of the simulated events are reported in Tables 5.3 and 5.4, respectively for multijet and $t\bar{t}$ processes, and for W/Z + jets events.

Process	σ [pb]	Order	Matrix Element generator	Showering
Multijet ($80 \text{ GeV} < p_T < 120 \text{ GeV}$)	2762530.0	LO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
Multijet ($120 \text{ GeV} < p_T < 170 \text{ GeV}$)	471100.0	LO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
Multijet ($170 \text{ GeV} < p_T < 300 \text{ GeV}$)	117276.0	LO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
Multijet ($300 \text{ GeV} < p_T < 470 \text{ GeV}$)	7823.0	LO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
Multijet ($470 \text{ GeV} < p_T < 600 \text{ GeV}$)	648.2	LO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
Multijet ($600 \text{ GeV} < p_T < 800 \text{ GeV}$)	186.9	LO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
Multijet ($800 \text{ GeV} < p_T < 1000 \text{ GeV}$)	32.293	LO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
Multijet ($1000 \text{ GeV} < p_T < 1400 \text{ GeV}$)	9.4183	LO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
Multijet ($1400 \text{ GeV} < p_T < 1800 \text{ GeV}$)	0.84265	LO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
Multijet ($1800 \text{ GeV} < p_T < 2400 \text{ GeV}$)	0.114943	LO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
Multijet ($2400 \text{ GeV} < p_T < 3200 \text{ GeV}$)	0.00682981	LO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
Multijet ($3200 \text{ GeV} < p_T < \infty$)	0.000165445	LO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
$t\bar{t}$	831.76	NNLO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
$t\bar{t}$ (single lepton from t)	182.72	NNLO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
$t\bar{t}$ (dilepton)	88.34	NNLO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
$t\bar{t}$ (single lepton from t, genMET 150) *	9.684	NNLO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
$t\bar{t}$ (single lepton from $\bar{t}$, genMET 150) *	9.658	NNLO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
$t\bar{t}$ (dilepton, genMET 150) *	5.919	NNLO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
$t\bar{t}$ ($600 \text{ GeV} < H_T < 800 \text{ GeV}$)	2.685	NNLO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
$t\bar{t}$ ($800 \text{ GeV} < H_T < 1200 \text{ GeV}$)	1.096	NNLO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
$t\bar{t}$ ($1200 \text{ GeV} < H_T < 2500 \text{ GeV}$)	0.194	NNLO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
$t\bar{t}$ ($2500 \text{ GeV} < H_T < \infty$)	0.002	NNLO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2

Table 5.3: Generation details for the simulation of the multijet and $t\bar{t}$ processes. The cross-section is reported, and the generator, along with the showering packages used with the relative calculation order are also listed. The star symbol * indicates the samples that are only used in 2016 and 2017.

Process	σ [pb]	Order	Matrix Element generator	Showering
W ($\rightarrow l\nu$) + jets ($100 \text{ GeV} < H_T < 200 \text{ GeV}$)	1627.45	NNLO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
W ($\rightarrow l\nu$) + jets ($200 \text{ GeV} < H_T < 400 \text{ GeV}$)	435.24	NNLO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
W ($\rightarrow l\nu$) + jets ($400 \text{ GeV} < H_T < 600 \text{ GeV}$)	59.18	NNLO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
W ($\rightarrow l\nu$) + jets ($600 \text{ GeV} < H_T < 800 \text{ GeV}$)	14.58	NNLO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
W ($\rightarrow l\nu$) + jets ($800 \text{ GeV} < H_T < 1200 \text{ GeV}$)	6.66	NNLO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
W ($\rightarrow l\nu$) + jets ($1200 \text{ GeV} < H_T < 2500 \text{ GeV}$)	1.608	NNLO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
W ($\rightarrow l\nu$) + jets ($2500 \text{ GeV} < H_T < \infty$)	0.03891	NNLO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
Z ($\rightarrow \nu\nu$) + jets ($100 \text{ GeV} < H_T < 200 \text{ GeV}$)	344.8305	NNLO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
Z ($\rightarrow \nu\nu$) + jets ($200 \text{ GeV} < H_T < 400 \text{ GeV}$)	95.5341	NNLO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
Z ($\rightarrow \nu\nu$) + jets ($400 \text{ GeV} < H_T < 600 \text{ GeV}$)	13.1979	NNLO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
Z ($\rightarrow \nu\nu$) + jets ($600 \text{ GeV} < H_T < 800 \text{ GeV}$)	3.14757	NNLO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
Z ($\rightarrow \nu\nu$) + jets ($800 \text{ GeV} < H_T < 1200 \text{ GeV}$)	1.450908	NNLO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
Z ($\rightarrow \nu\nu$) + jets ($1200 \text{ GeV} < H_T < 2500 \text{ GeV}$)	0.3546459	NNLO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2
Z ($\rightarrow \nu\nu$) + jets ($2500 \text{ GeV} < H_T < \infty$)	0.00854235	NNLO	MADGRAPH5 aMC@NLO 2.2.2	PYTHIA 8.2

Table 5.4: Generation details for the simulation of the Z and W + jets processes. The cross-section is reported, and the generator, along with the showering packages used with the relative calculation order are also listed.

5.2 Analysis strategy

5.2.1 Trigger

Triggers based on the jet p_T or the hadronic activity H_T are the most suitable for the study presented in this thesis. More in details, for the 2016 data, the logical **OR** among the triggers applying the following requirements is considered: a large-cone PF jet with $p_T > 450$ GeV, a hadronic activity larger than 900 GeV, a calo-jet with $p_T > 500$ GeV and a PF jet with $p_T > 450$ GeV. A similar strategy is applied on 2017 and 2018 data, but with different thresholds, being the requirements a large-cone PF jet with $p_T > 550$ GeV, a hadronic activity larger than 1050 GeV, a calo-jet with $p_T > 500$ GeV and a PF jet with $p_T > 550$ GeV. Table 5.5 summarizes the HLT triggers paths employed.

Year	HLT paths
2016	AK8PFJet450
	CaloJet500_NoJetID
	PFHT900
	PFJet450
2017 - 2018	AK8PFJet550
	CaloJet500_NoJetID
	PFHT1050
	PFJet550

Table 5.5: Summary of the hadronic HLT triggers used in 2016, 2017 and 2018.

Trigger efficiency

The efficiency of the trigger is computed using events minimizing the presence of our signal of interest. Hence, being the search performed in this thesis fully hadronic, the trigger efficiency is measured in a region with at least one muon with $p_T > 50$ GeV. In addition to the trigger listed in Table 5.5, an additional trigger, labeled as Mu50, is considered for this study.

The trigger efficiency is computed as the fraction of events passing the Mu50 trigger over the ones passing the logical **OR** of the hadronic based triggers and the Mu50 trigger, and as a function of the cut applied on the variable of interest (x). The curve is fit with the following function:

$$f(x) = \frac{1}{2}N \left(1 + \operatorname{erf} \left(\frac{x - \mu}{\sqrt{2}\sigma} \right) \right), \quad (5.3)$$

and the plateau efficiency is considered to be reached at $\mu + 3\sigma$.

Figure 5.3 shows the efficiency curves for 2016, 2017 and 2018 as a function of the leading jet p_T . Despite the p_T of the jet would be the more obvious variable to measure the trigger efficiency, this would require a high jet p_T requirement to reach the plateau, which would affect the statistics of the search.

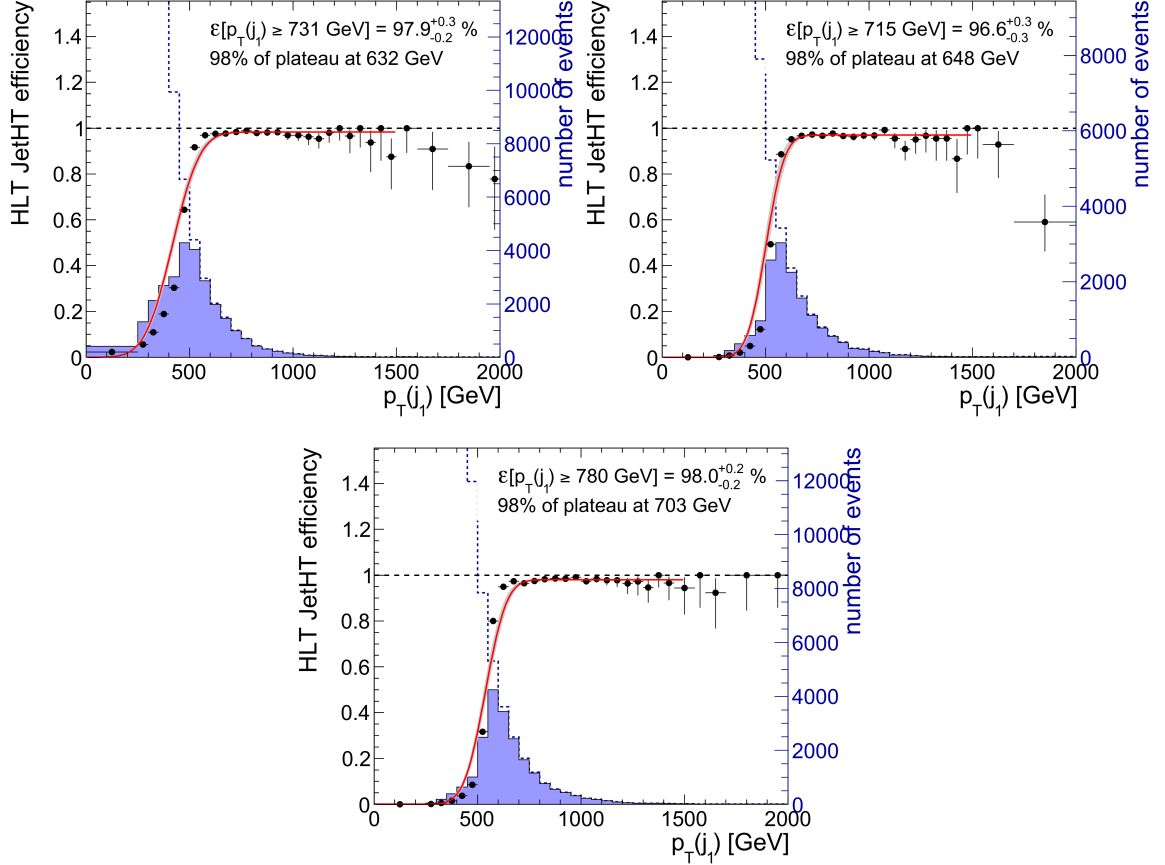


Figure 5.3: Efficiency of hadronic triggers for 2016 (top left), 2017 (top right) and 2018 (bottom), as a function of the jet p_T . The blue dashed line represents the denominator distribution, while the purple filled histogram shows the numerator distribution. The black points refers to the efficiency, computed as the ratio of the numerator and denominator distributions. The red line presents the resulting fit, with the gray band showing the uncertainty in the fit.

Instead, it is preferable to use another variable, the transverse mass m_T , defined as:

$$m_T^2 = m_{jj}^2 + 2 \left(\sqrt{m_{jj}^2 + p_{T,jj}^2} p_{T,miss} - p_{T,jj} \cdot p_{T,miss} \right). \quad (5.4)$$

This quantity accounts for the invariant mass between the two semi-visible jets and for the missing transverse momentum arising from their dark components. This is also the observable used for the statistical inference. Figure 5.4 shows the trigger efficiencies distributions as a function of m_T . There is a notable inefficiency in this measurement, which occurs because there are two sources of events with high dijet mass. The first source is a pair of high- p_T jets, which pass the jet-based triggers. The second source is a pair of low- p_T jets with a large angle between them, which do not pass the jet-based triggers.

To exclude the second source of high-mass events, only events with low $\Delta\eta(j_1, j_2)$ between the p_T -leading jets are accepted. A tighter $\Delta\eta(j_1, j_2)$ selection removes more events from the signal region but reduces the m_T cut needed to reach the trigger plateau. The maximum $\Delta\eta(j_1, j_2)$ value is optimized by comparing signal to simulated backgrounds, following the procedure described in Section 5.2.2. The trigger efficiencies measured with the resulting optimal choice of $\Delta\eta(j_1, j_2) < 1.5$ are shown in Figure 5.5. At least 98% efficiency

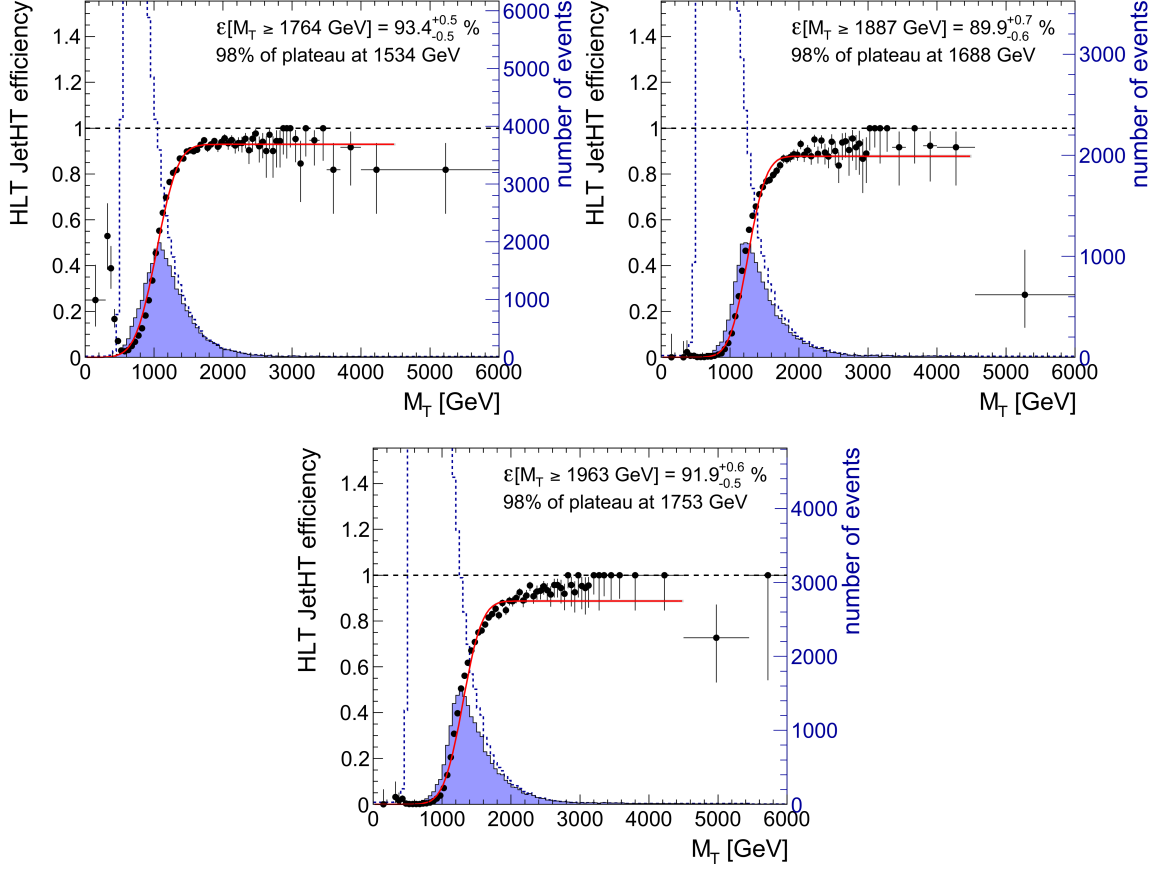


Figure 5.4: Efficiency of hadronic triggers for 2016 (top left), 2017 (top right) and 2018 (bottom), as a function of the m_T . The blue dashed line represents the denominator distribution, while the purple filled histogram shows the numerator distribution. The black points refers to the efficiency, computed as the ratio of the numerator and denominator distributions. The red line indicates the resulting fit, with the gray band showing the uncertainty in the fit.

is achieved in all years, with a cut $m_T > 1500$ GeV applied to reach the trigger plateau.

The fitted efficiency function is applied as a correction to simulated signals, with the statistical uncertainty from the control region propagated as a systematic uncertainty. Table 5.6 shows the magnitude and statistical uncertainty on the trigger efficiency correction for each year, at the m_T threshold value, 1500 GeV and at the upper boundary of $m_{Z'}$ range considered in this search. An additional 2% systematic uncertainty is assigned to cover any kinematic differences between the control and signal regions.

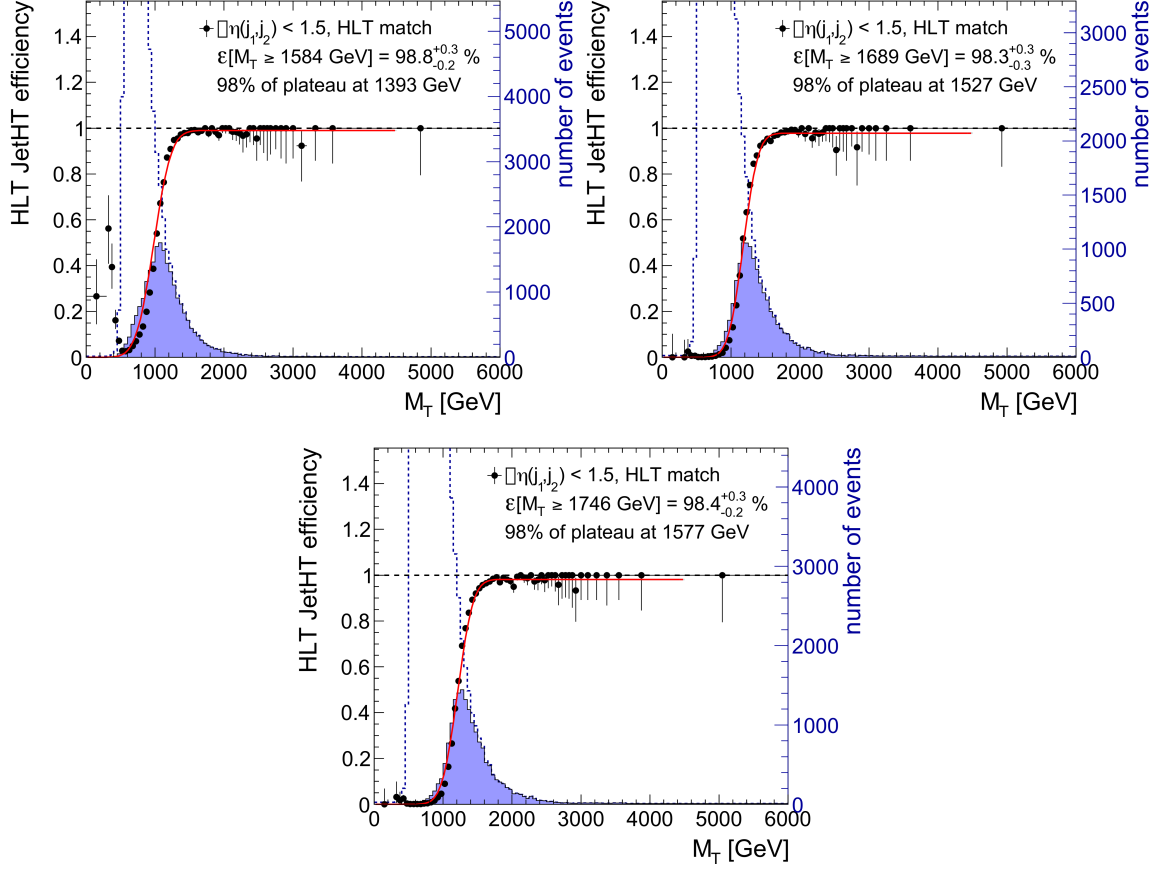


Figure 5.5: Efficiency of hadronic triggers for 2016 (top left), 2017 (top right) and 2018 (bottom), measured as a function of the m_T , with the cut $\Delta\eta(j_1, j_2) < 1.5$ included. The blue dashed line represents the denominator distribution, while the purple filled histogram represents the numerator distribution. The right-side vertical axis shows the values of these distributions. The black points show the efficiency as the ratio of the numerator and denominator distributions, with the horizontal error bars indicating the variable bin widths. The red line indicates the resulting fit, with the small gray band showing the uncertainty in the fit.

Year	m_T [GeV]	Correction
2016	1500	0.985 ± 0.0022
	4500	0.990 ± 0.0024
2017	1500	0.950 ± 0.0051
	4500	0.978 ± 0.0039
2018	1500	0.930 ± 0.0065
	4500	0.981 ± 0.0031

Table 5.6: Trigger efficiency corrections for each year of data taking, measured at the m_T threshold value and the upper boundary of the fit, including the statistical uncertainty.

5.2.2 Event selection

The final state of interest $q\bar{q} \rightarrow Z' \rightarrow \chi\bar{\chi}$ leads to a hadronic final state with two large-cone jets. Besides, the invisible part of the semi-visible jets would manifest itself as missing transverse energy parallel to the direction of the jets. To maximize the signal separation

from the background processes an event selection is optimized. A first set of requirements identifying the signal topology is fixed and described below:

- the leading two AK8 jets must have $p_T > 200$ GeV and $|\eta| < 2.4$, and satisfy loose PFJet ID criteria (Section 3.5.2);
- The pseudorapidity separation between the two leading AK8 jets should respect $\Delta\eta(j_1, j_2) = |\eta_1 - \eta_2| < 1.5$, removing the t -channel QCD;
- the transverse ratio between the E_T^{miss} and the m_T , labeled as R_T , is required to be greater than 0.15. It acts as a E_T^{miss} cut but avoids sculpting the m_T distribution. This requirement helps in reducing multijet and hadronic $t\bar{t}$ events.
- A lepton veto is applied, which consists in vetoing electrons and muons with $p_T > 10$ GeV and $|\eta| < 2.4$. With this selection events from $V + \text{jets}$ and leptonic $t\bar{t}$ are suppressed.
- $m_T > 1500$ GeV, to be on the trigger plateau, as described in Section 5.2.1;
- dedicated E_T^{miss} filters, developed to identify and reject false E_T^{miss} , are applied. Large missing transverse energy can be produced as a consequence of missing detection of invisible particles, but it can also have uninteresting causes such as detector noise and cosmic rays. These filters are designed to remove events with missing momentum due to ECAL spurious deposits and HCAL spurious energy, high- p_T tracks from misreconstructed muons, etc.

We will refer to this set of requirements as *preselection*.

Optimization of selection requirements

To further enhance the signal to background separation, additional requirements have been studied and optimized.

Unlike WIMPs, missing transverse momentum in semi-visible jets is expected to be collinear to visible jets, thus the azimuthal opening angle between the jets and the E_T^{miss} is expected to be small. Selecting these particular events allows to enhance the signal and to reduce the background arising from Z and W decays. The requirement on this observable needs to be optimized. Thus, dedicated optimization studies to improve the signal to background separation have been performed on $\Delta\phi_{\min}(j_{1,2}, E_T^{\text{miss}})$ and on other variables, such as the pseudorapidity opening between the two jets $\Delta\eta(j_1, j_2)$ and the transverse ratio R_T .

The optimization is done by choosing the threshold value for each observable that maximizes a figure of merit (FOM) [183] defined as:

$$\text{FOM} = \sqrt{2((S + B) \times \log(1 + \frac{S}{B}) - S)}, \quad (5.5)$$

where S and B label the yields of the signal and the backgrounds respectively. For the signal, a benchmark model with $m_{Z'} = 3000$ TeV is used.

A chain optimization is performed, focusing on a window of $m_{Z'}$ between 2 and 4 TeV. Starting from $\Delta\eta(j_1, j_2)$, the FOM is computed in the region of interest, and the resulting distribution is shown in Figure 5.6 (top left). The optimal value of selection is 1.5, thus a requirement on $\Delta\eta(j_1, j_2) < 1.5$ is defined. As a second step, now including the $\Delta\eta(j_1, j_2)$ cut as well, the FOM is computed as a function of the $\Delta\phi_{\min}(j_{1,2}, E_T^{\text{miss}})$ variable, considering for the background yields the V + jets contributions only. No big discriminating power is found for this variable, as presented in Figure 5.6 (top right), being the FOM distribution quite flat and fast arriving at its plateau. Despite the optimal cut is at 1.25, to increase the statistics and given its small effect, a selection at $\Delta\phi_{\min}(j_{1,2}, E_T^{\text{miss}}) < 0.75$ is deployed. As a last step, the R_T threshold is optimized and the resulting best value is 0.30. Nevertheless, being this variable the most powerful one, lower cuts have also been tested, which are 0.20 and 0.25. Even if 0.30 is the value returning the highest signal sensitivity, it highly reduces the background yields. It appears that the statistics is very limited with a tight cut, mostly on the tail of the distributions, as shown in Figure 5.7. Besides, to avoid a sculpting of the m_T shape, an intermediate value of R_T is chosen.

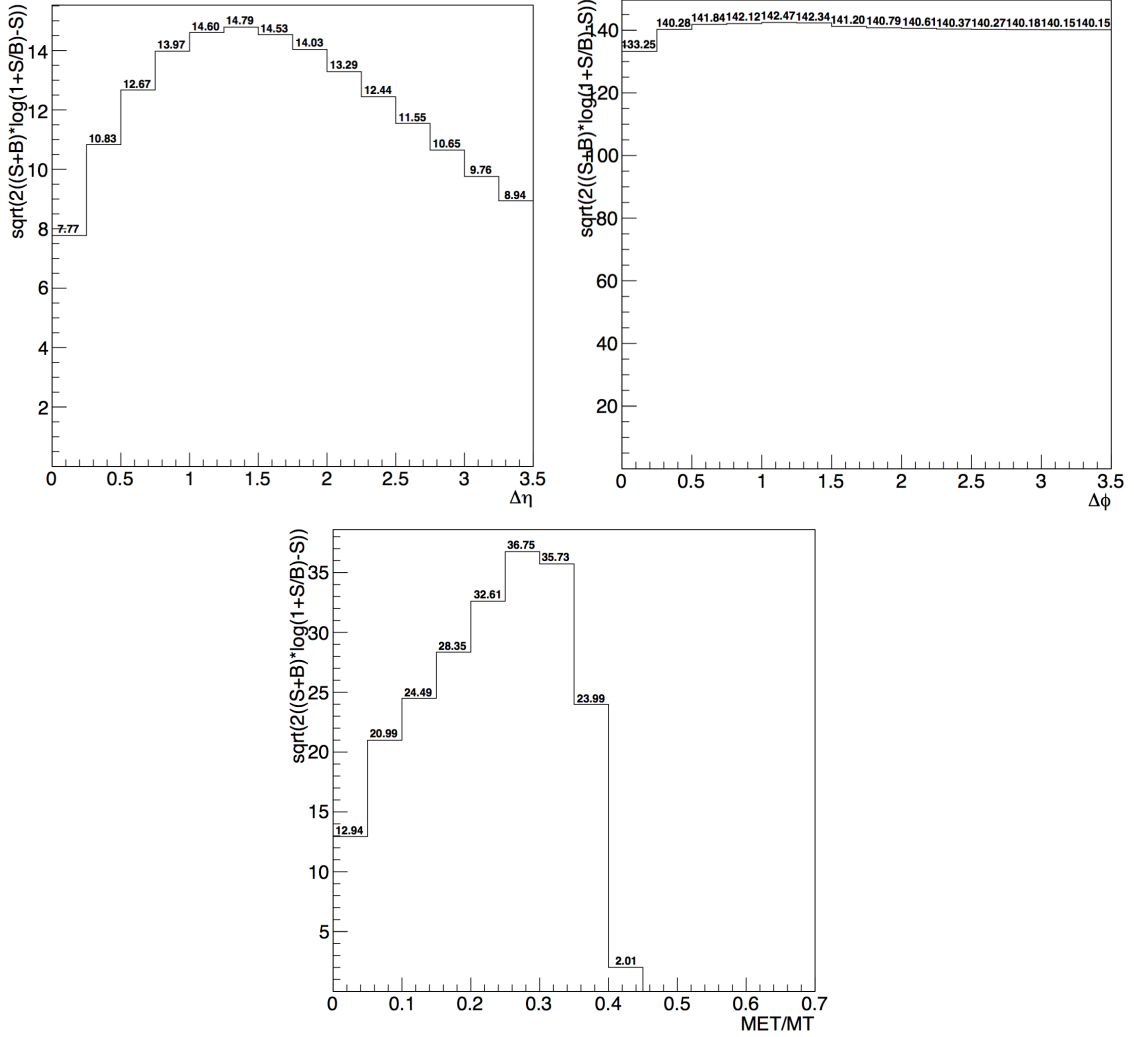


Figure 5.6: Distributions of the figure of merit 5.5, as a function of $\Delta\eta(j_1, j_2)$ (top left), $\Delta\phi_{\min}(j_{1,2}, E_T^{\text{miss}})$ (top right) and R_T (bottom).

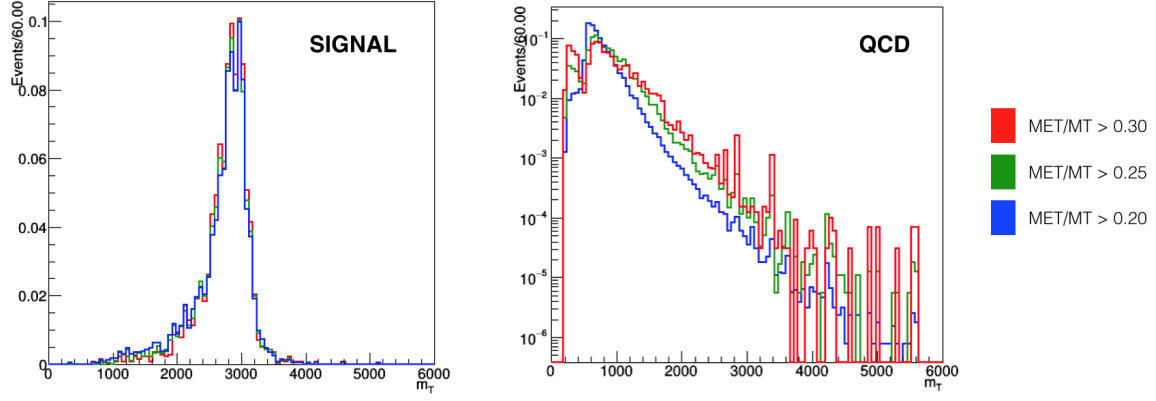


Figure 5.7: Normalized transverse mass distribution on signal (left) and QCD background (right). Three different thresholds, 0.20, 0.25 and 0.30, on R_T are applied and compared and they are presented in blue, green and red, respectively.

While the above selection optimization was performed with just the baseline signal sample, an additional selection optimization was performed with all the signal samples generated. This additional step focuses on $\Delta\phi_{\min}(j_{1,2}, E_T^{\text{miss}})$ and R_T . The optimal selection was determined by finding the set of requirements that lowered the expected upper on the median on $\sigma \times \mathcal{B}$. Expected limits were computed for a multitude of selections and the corresponding results are illustrated in Figure 5.8. The selection determined by this optimization is $\Delta\phi_{\min}(j_{1,2}, E_T^{\text{miss}}) < 0.80$ and $R_T > 0.19$.

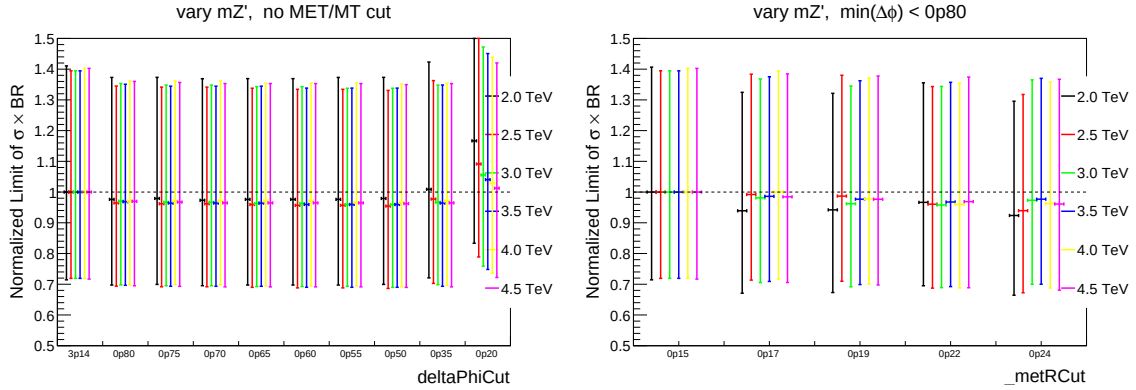


Figure 5.8: Normalized expected limits ($\pm 1\sigma$) for select signal samples versus selection value in $\Delta\phi_{\min}(j_{1,2}, E_T^{\text{miss}})$ (left) and R_T (right), varying the mediator mass $m_{Z'}$. Since most points are indistinguishable from 1 (to within $\pm 1\sigma$) the optimization was performed regarding only the median expected limit.

As will be further explained in Section 5.4, multiple signal regions are defined in this study and to provide enough statistics in each of them, it was decided to use the looser R_T selection resulted from the first optimization, while the $\Delta\phi_{\min}(j_{1,2}, E_T^{\text{miss}})$ selection is taken from the second optimization. A summary of all requirements employed in the analysis, both at preselection and a selection level, is reported in Table 5.7.

In addition to the variables described above, other features such as the E_T^{miss} and the ΔR separation between the two leading jets, have been tested. However, the improvement they were bringing was found to be negligible and therefore they have been discarded.

Level	Requirement
Preselection	2 AK jets ($p_T > 200$ GeV, $ \eta < 2.4$, loose ID)
	$\Delta\eta(j_1, j_2) < 1.5$
	$R_T > 0.15$
	lepton veto
	$m_T > 1500$ GeV
Selection	E_T^{miss} filters
	$R_T > 0.25$ $\Delta\phi_{\min}(j_{1,2}, E_T^{\text{miss}}) < 0.80$

Table 5.7: Summary of preselection and selection requirements.

Detector failure affecting 2018 data

During the 2018 data taking period, a region in the endcap of the hadronic calorimeter failed, specifically the one covering the phase space $-3 < \eta < -1.4$ and $270^\circ < \phi < 310^\circ$. This issue could cause a significant increase in the multijet background, having the jets that fall in this section of the detector lower measured transverse momentum and consequently higher measured E_T^{miss} . This problem is labeled as *HEM issue*.

In Figure 5.9 multijet events coming from multijet processes are compared with and without the HEM issue simulated. In certain regions of the phase space, more events are selected when the HEM issue is included in the simulation. In the $\phi(E_T^{\text{miss}})$ distribution, an excess of selected events in the range of ϕ covered by the problematic region in the hadronic calorimeter, is expected. The same observation holds for $\phi(j_1)$ and $\phi(j_2)$. Also, events with jets pointing to the affected region tend to have higher measured transverse momenta, as can be inferred by the E_T^{miss} distribution in Figure 5.9.

To reduce the impact of the HEM failure while retaining good signal efficiency a veto on MC and data events has been developed and optimized. It is found that the HEM issue is solved if events containing jets satisfying the kinematic conditions $p_T > 30$ GeV, $-3.05 < \eta < -1.35$, and $-1.62 < \phi < -0.82$, are vetoed. In Figure 5.10 it can be observed that with the application of the optimized HEM veto, a better agreement between multijet events with and without the HEM issue simulated is achieved. However, applying the HEM veto also reduces our signal efficiency by about 6% on average.

5.3 A dedicated SVJ tagger

To further enhance our ability to distinguish signal-like and multijet-like jets a dedicated tagger has been designed, built and optimized. Different underlying interactions in dark and visible sectors lead to a different radiation pattern affecting the development of the shower and hence the internal structure of the jet, such as the energy distribution of jet constituents or their multiplicity as well as the shape of the jet itself.

To build the tagger, the jet substructure and the event kinematics are exploited, and related observables are combined in a boosted decision tree (BDT) with gradient boosting [184]. A pivotal role is also played by the Quark-Gluon tagger (see Appendix B), being the same input variables used for the QG discriminator also employed in the SVJ tagger, as

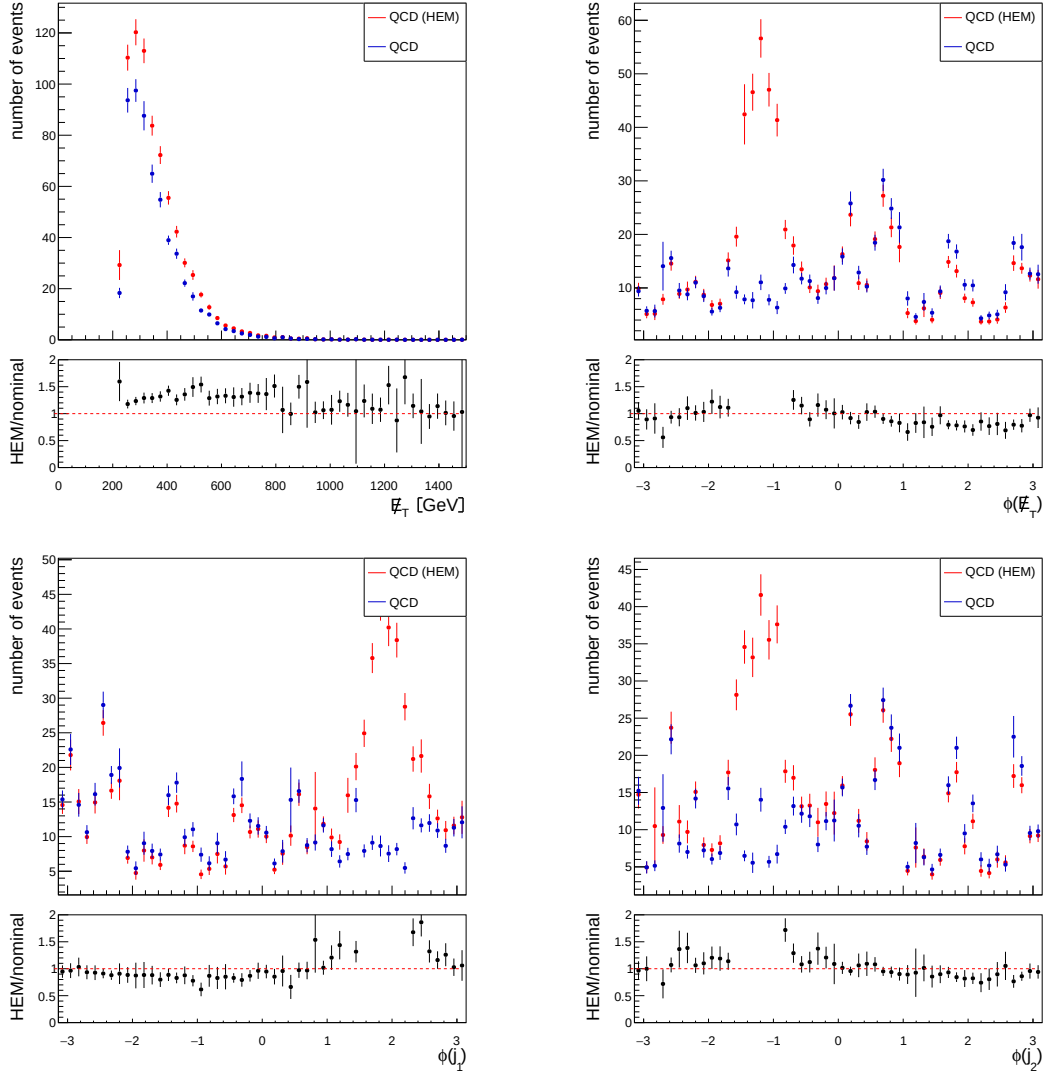


Figure 5.9: Comparison of multijet events with and without the HEM issue in distributions of $\Delta\phi_{\min}(j_{1,2}, E_T^{\text{miss}})$, E_T^{miss} , $\phi(j_1)$, $\phi(j_2)$, and $\phi(E_T^{\text{miss}})$.

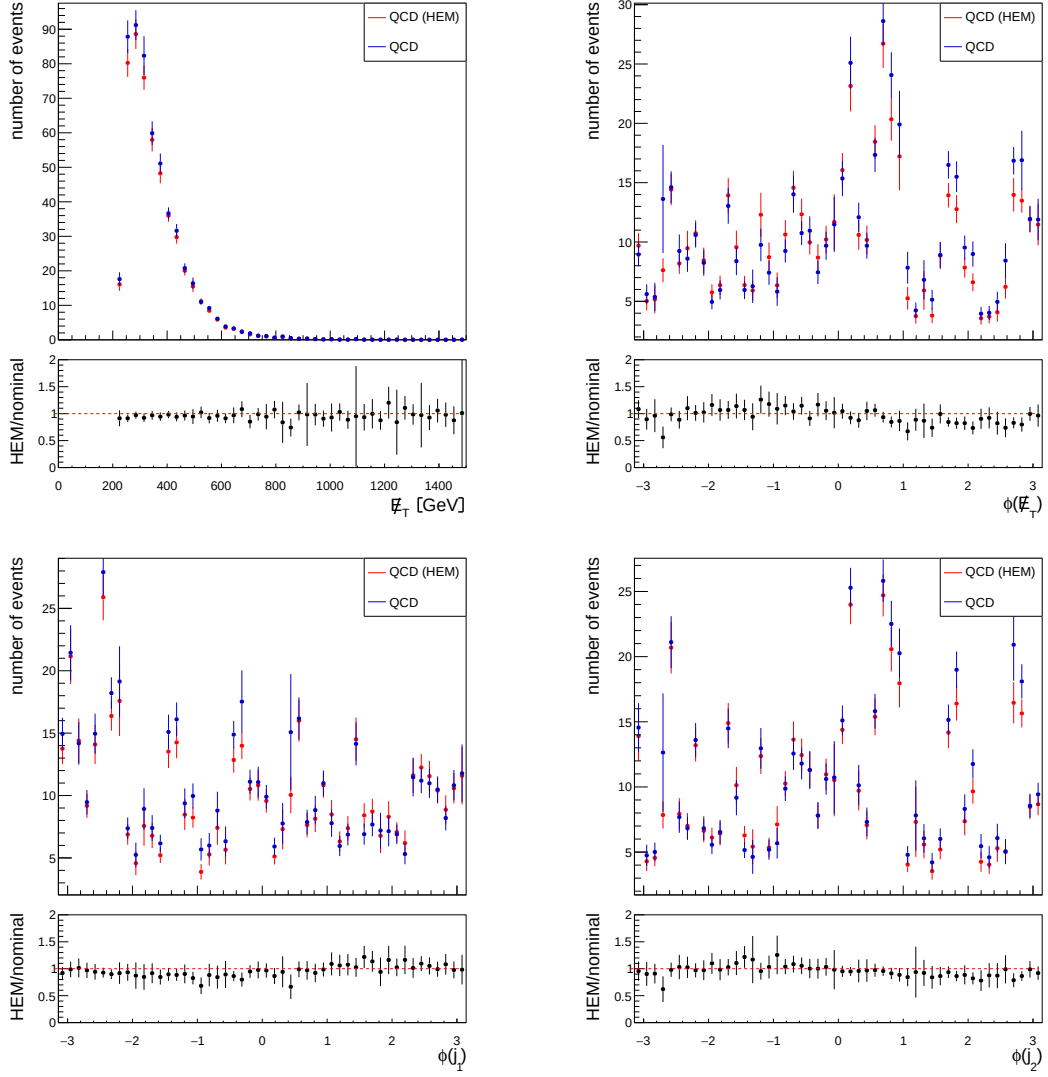


Figure 5.10: Comparison of multijet events with and without the HEM issue in distributions of $\Delta\phi_{\min}(j_{1,2}, E_T^{\text{miss}})$, E_T^{miss} , $\phi(j_1)$, $\phi(j_2)$, and $\phi(E_T^{\text{miss}})$ after applying the HEM veto.

further discussed below.

5.3.1 Observables

The variables used as input to the BDT can be classified in four categories:

1. Heavy objects (t, W, Z, H) tagging:

- (a) N -subjettiness, which quantifies how well the jet constituents can be assigned to N internal axes or prongs, as described in Section 3.6.1. In particular, N -subjettiness ratios τ_N/τ_{N-1} are exploited. Low values denote that a jet is more likely to have N prongs than $N - 1$ prongs. In this thesis, the ratios $\tau_{21} = \tau_2/\tau_1$ and $\tau_{32} = \tau_3/\tau_2$ are used.
- (b) Softdrop mass, which is the invariant mass of the jet after the softdrop grooming algorithm has been applied to remove soft and wide-angle radiation. The softdrop mass m_{SD} is computed as the invariant mass of the final softdrop subjets.
- (c) Energy correlation functions (ECFs) [185], which serve a similar purpose as N -subjettiness but avoid the need to define internal subjet axes by comparing jet constituents directly. ECFs are evaluated using groomed jets. The generalized ECFs are written as:

$$v_n^{e^{(\beta)}} = \sum_{1 \leq i_1 < \dots < i_n \leq n_{\text{const.}}} z_{i_1} \dots z_{i_n} \prod_{m=1}^v \min_{s < t \in \{i_1, \dots, i_n\}}^{(m)} \left\{ \theta_{st}^\beta \right\}. \quad (5.6)$$

Here, z_{i_1} is the fraction of the jet p_T carried by constituent i_1 , and $\theta_{i_1 i_2} = (\phi_{i_1} - \phi_{i_2})^2 + (y_{i_1} - y_{i_2})^2$, where y is the rapidity. The following combinations of the generalized ECFs are used: $N_2^{(1)} = \frac{2e_3^{(1)}}{\left(1e_2^{(1)}\right)^2}$ and $N_3^{(1)} = \frac{2e_4^{(1)}}{\left(1e_3^{(1)}\right)^2}$.

2. Quark-gluon discrimination:

- (a) The jet girth [186], that gives a generic measure of the distribution of the jet constituents and that can be expressed as:

$$g = \sum_i \frac{p_{T,i}}{p_{T,\text{jet}} r_i}. \quad (5.7)$$

- (b) Major and minor jet axes, computed from the eigenvalues λ_1, λ_2 of the p_T -weighted $\eta - \phi$ matrix of the jet constituents. The major and minor axes are defined as $\sigma_{\text{major}} = \sqrt{\lambda_1 / \sum_i p_{T,i}^2}$ and $\sigma_{\text{minor}} = \sqrt{\lambda_2 / \sum_i p_{T,i}^2}$.
- (c) The fragmentation variable $p_T D = \frac{\sqrt{\sum_i p_{T,i}^2}}{\sum_i p_{T,i}}$, which quantifies the number of constituents carrying a significant fraction of the jet energy.

3. Flavor:

- (a) The jet energy fractions $f_{h^\pm}, f_e, f_\mu, f_{h^0}, f_\gamma$, which are defined as the fraction of the jet energy carried by each category of constituents, respectively charged hadrons, electrons, muons, neutral hadrons and photons.

4. Other:

- (a) The angle between the jet transverse momentum and the missing transverse momentum, $\Delta\phi(\text{jet}, E_T^{\text{miss}})$.

In total 15 input observables are used to train the SVJ tagger. Their distributions are presented in Figure 5.11, where simulated SM backgrounds and various signal models are shown. Many of these observables are correlated with the jet p_T . Hence, the distributions are weighted to have a flat jet p_T spectrum to compare kinematically similar jets from signal and background samples that naturally have different p_T spectra. The Receiver Operator Characteristic (ROC) curves for the benchmark signal model compared to multijet and $t\bar{t}$ backgrounds are shown in Figure 5.12.

From Figure 5.11, the key distinctive features between semi-visible jets and QCD-like jets are identified and can be summarized as:

- semi-visible jets are typically less likely than background jets to have 2- or 3-prong substructure, as indicated by the N -subjettiness and ECF variables distributions.
- While heavy objects that undergo weak decays tend to have their original mass values reflect in the m_{SD} distribution of the resulting boosted jet, semi-visible jets are formed entirely through hadronization and strong processes, meaning that the mass value of the originating dark quark is not reflected in m_{SD} . However, the dark hadron mass m_{dark} can be reflected in m_{SD} in the case of a single hard fragmentation that carries a significant portion of the jet energy.
- The high scale of the dark strong force means that semi-visible jets are expected to be even wider than SM gluon jets. This is reflected in the larger values of the jet girth and the major and minor axes. The gluon-like nature of semi-visible jets is also reflected in the lower values of $p_T D$.
- The energy fractions are insensitive to most variations in signal model parameters. They reflect only the SM part of the semi-visible jet, and as fractions, are not influenced by changes in the invisible fraction r_{inv} . However, signal models with very low m_{dark} values that prevent heavy flavor decays will have smaller values of f_e and f_μ , since lepton activity in semi-visible jets tends to arise from B hadron decays.

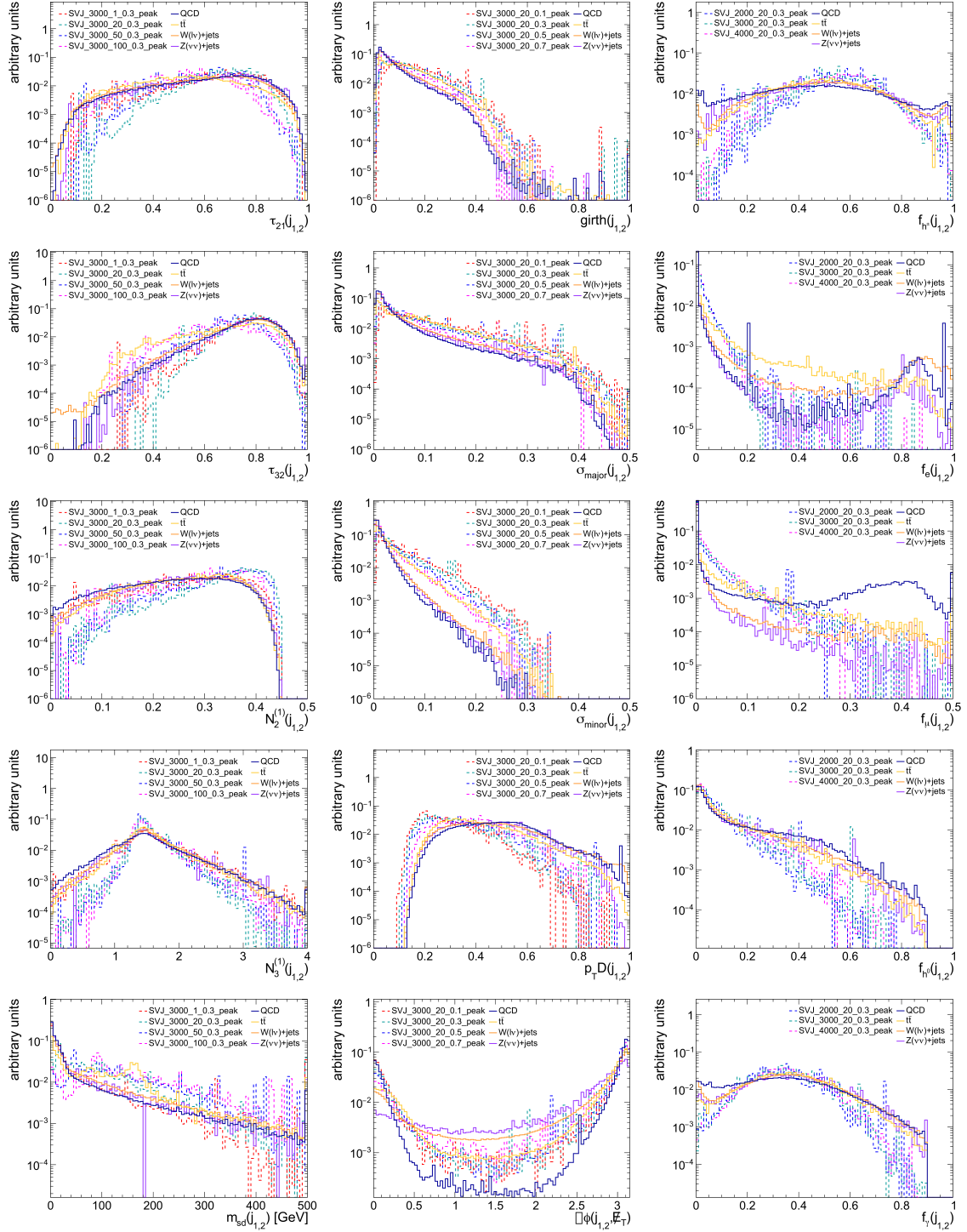


Figure 5.11: Input observables used in the BDT, with distributions of SM backgrounds and signal models with different r_{inv} values. Heavy object tagging variables are presented in the left column, quark-gluon related observables in the middle one and Jet energy fractions in the right column. All distributions are normalized to unit area and include the two leading AK8 jets, with a flattened p_T spectrum.

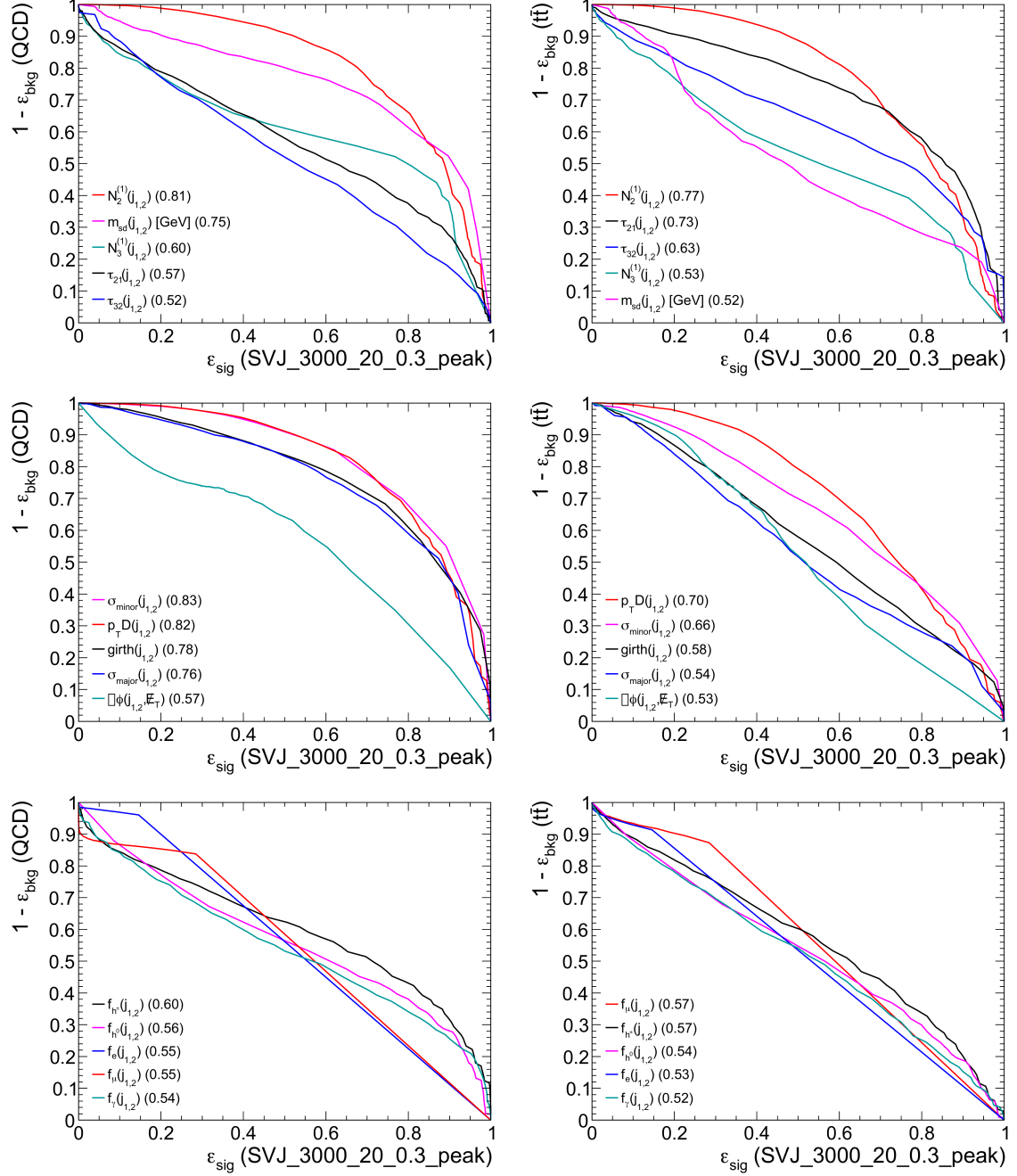


Figure 5.12: ROC curves for the benchmark signal model vs. multijet (left) and $t\bar{t}$ backgrounds, showing each grouping of input variables: heavy object (top), quark-gluon (middle) and energy fractions (bottom). The area under the curve is listed for each variable.

5.3.2 Training

A boosted decision with gradient boosting is employed [184], thus binomial log-likelihood loss function is used to measure the deviation between the model response and the true value obtained from the training sample.

The training of the semi-visible jet tagger is performed using a subset of signals with high acceptance (i.e. with parameters $m_{Z'} > 1500$ GeV, $m_{\text{dark}} > 10$ GeV, $0.1 \leq r_{\text{inv}} \leq 0.8$) against multijet and $t\bar{t}$ background processes. The simulated events are weighted by their cross-sections, along with a correction to match the distribution of PU interactions in data. Besides, to minimize the dependence on the jet p_T and thereby avoid sculpting the m_T distribution, both the signal and background jets are weighted to make their p_T distributions match a reference distribution from the benchmark signal model. This approach is used rather than a complete flattening to avoid high weights for less-populated bins of the p_T distributions. The signal weights are normalized so that their sum, considering all signal events included in the training, is equal to 1. The same is done for the background events, also requiring that 50% by weight is QCD and the other 50% is $t\bar{t}$.

To verify that our SVJ classifier doesn't model signal/background differences that are specific to the particular sample used (the so-called *overtraining issue*), the input events are randomly assigned to two equally populated subsets, called *training* sample and *test* sample. The BDT is developed on the training sample, and then its behavior is checked on the test sample, verifying that the classifier models signal/background differences on the additional sample as well.

Several classifier parameters and requirements of interest have been studied to optimize the training of the SVJ tagger, that are:

- 1000 trees;
- a maximum cell tree depth of 4;
- a learning rate of 0.1, that slows down the learning in the gradient boosting model and reduces the probability of overfitting;
- each leaf node should have 1% of the total number of events in the sample.

These requirements were optimized by training one BDT for each combination of a grid composed of a range of values for each hyperparameter. Resulting BDTs with a low overtraining both for signal and background and a low level of correlation between the classifier efficiency and the m_T distribution, were selected. Among those, the BDT returning the best performance and the highest area under the ROC curve was finally chosen and identified as the SVJ tagger.

5.3.3 Results

The performance of the tagger are evaluated, as shown in Figure 5.13. A good separation between signal and background is achieved, and it can be quantified with an area under the curve of 0.920. No overtraining is observed, being the Kolmogorov-Smirnov test quantifying

the agreement between the discriminator distributions for the separate training and test datasets satisfying both for signal and background. The distribution of the SVJ tagger on signal and SM background is shown in Figure 5.14.

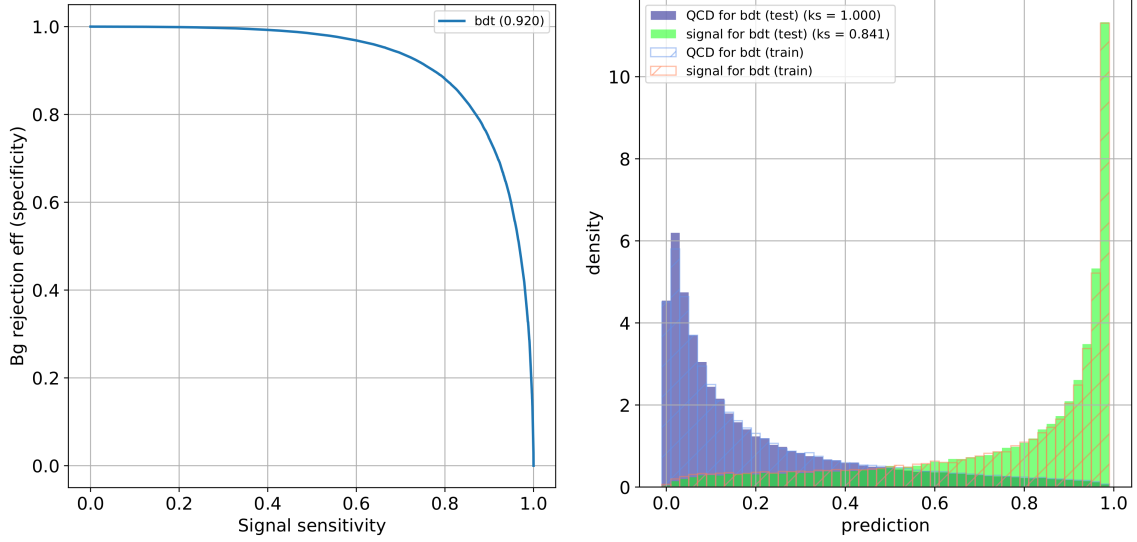


Figure 5.13: Diagnostic plots for the BDT training and performance. The ROC curve showing the background rejection efficiency as a function of the signal sensitivity is presented, and the area under the curve is reported as well (left). The BDT distribution is illustrated (right). The empty histograms show the distributions for background (blue) and signal (orange) on the training sample, while the filled histograms display the BDT obtained on the test sample, for signal (green) and background (blue).

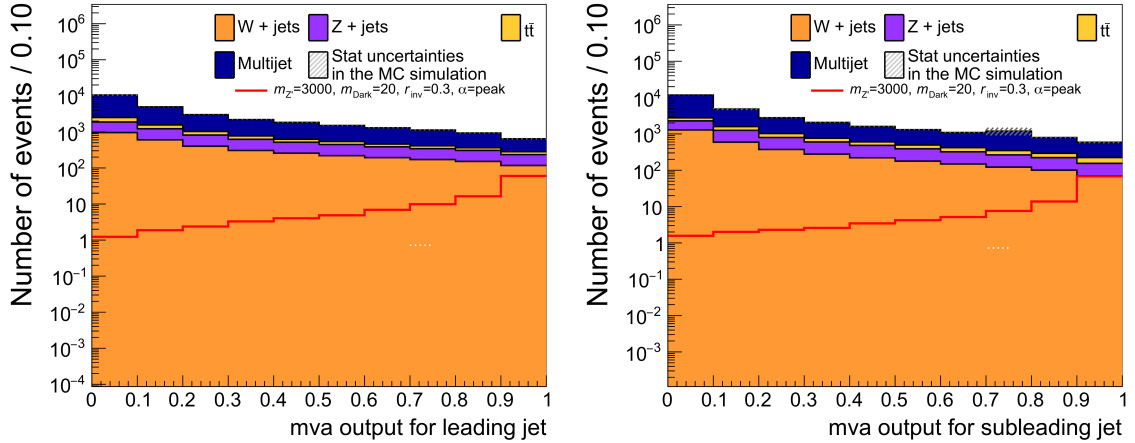


Figure 5.14: The BDT discriminator distribution for a signal sample (red line) and on SM background (filled histograms) for the p_T -leading (left) and p_T -subleading (right) jet, at preselection level.

5.4 Background estimation

The main contribution to the background comes from multijet events. SM $t\bar{t}$ process also contributes, along with Z/W + jets events. A data-driven strategy, based on the fit of the m_T distribution, is used to estimate the background.

5.4.1 Signal regions

Collision events are classified in different categories, *signal regions*, based on the number of jets tagged as semi-visible. Events fulfilling the selection described in Section 5.2.2 and having exactly 1 or 2 AK8 jets passing the SVJ tagger requirement populate signal regions. These regions are labeled with SVJ1 and SVJ2, respectively.

As previously discussed, the optimal cut for R_T is found to be 0.25, however, we found that events with $0.15 < R_T < 0.25$ bring additional sensitivity, being the signal contribution in this region not negligible. Therefore, these events are retained for the final statistical analysis and classified in two categories depending on the number of jets tagged as semi-visible. To distinguish between these two categories and the ones with $R_T > 0.25$, a prefix (*low* or *high*) is added to their names.

An illustration explaining the categories and their labeling is presented in Figure 5.15.

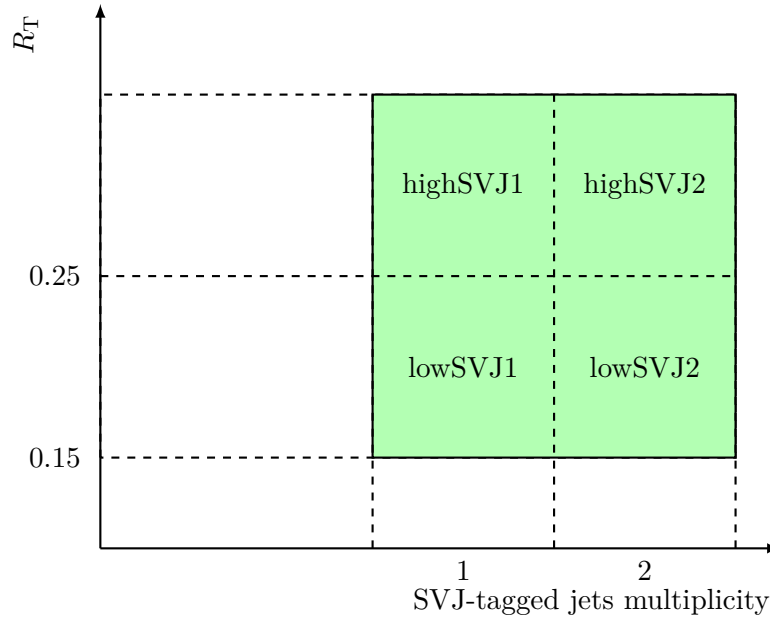


Figure 5.15: Sketch illustrating the definition of control and signal regions, along with their labels. The signal regions are highlighted in green.

The transverse mass distributions for 2016, 2017 and 2018 and in all categories used in the analysis are displayed in Figures 5.16, 5.17, 5.18, and 5.19, respectively.

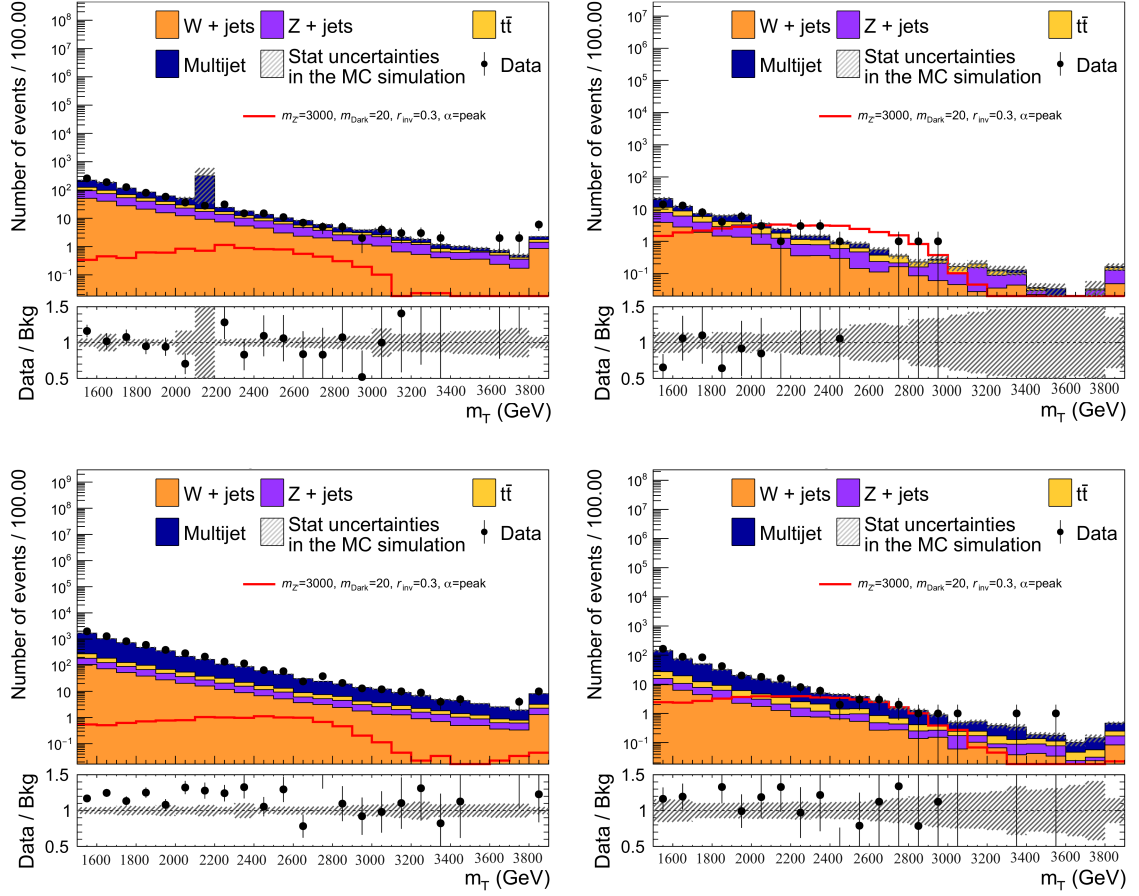


Figure 5.16: The transverse mass distribution in the signal regions on the benchmark signal sample (red) and on SM background (filled histograms) in 2016. The categories with high R_T are presented in the top row, while at the bottom the lower R_T selection is applied. One SVJ jet is requested in the left column and both jets must fulfill the tagger requirement in the right column.

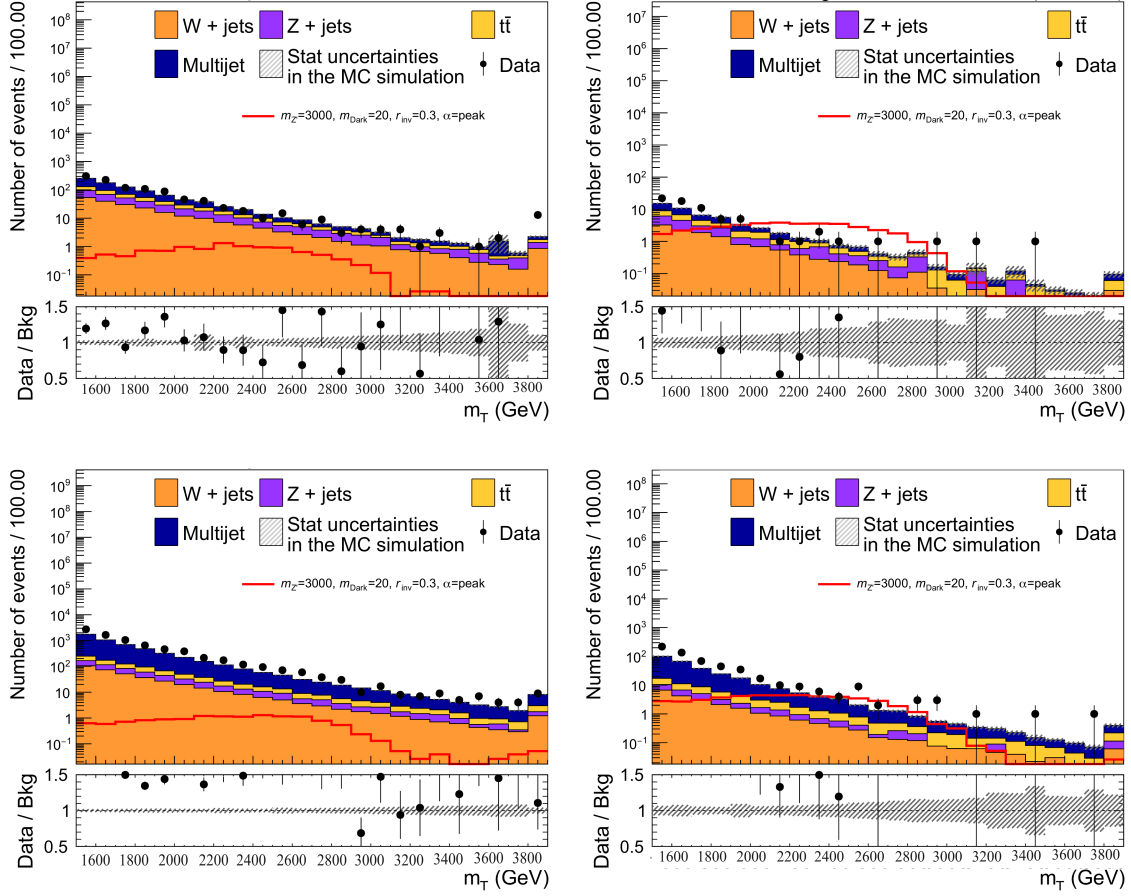


Figure 5.17: The transverse mass distribution in the signal regions on the benchmark signal sample (red) and on SM background (filled histograms) in 2017. The categories with high R_T are presented in the top row, while at the bottom the lower R_T selection is applied. One SVJ jet is requested in the left column and both jets must fulfill the tagger requirement in the right column.

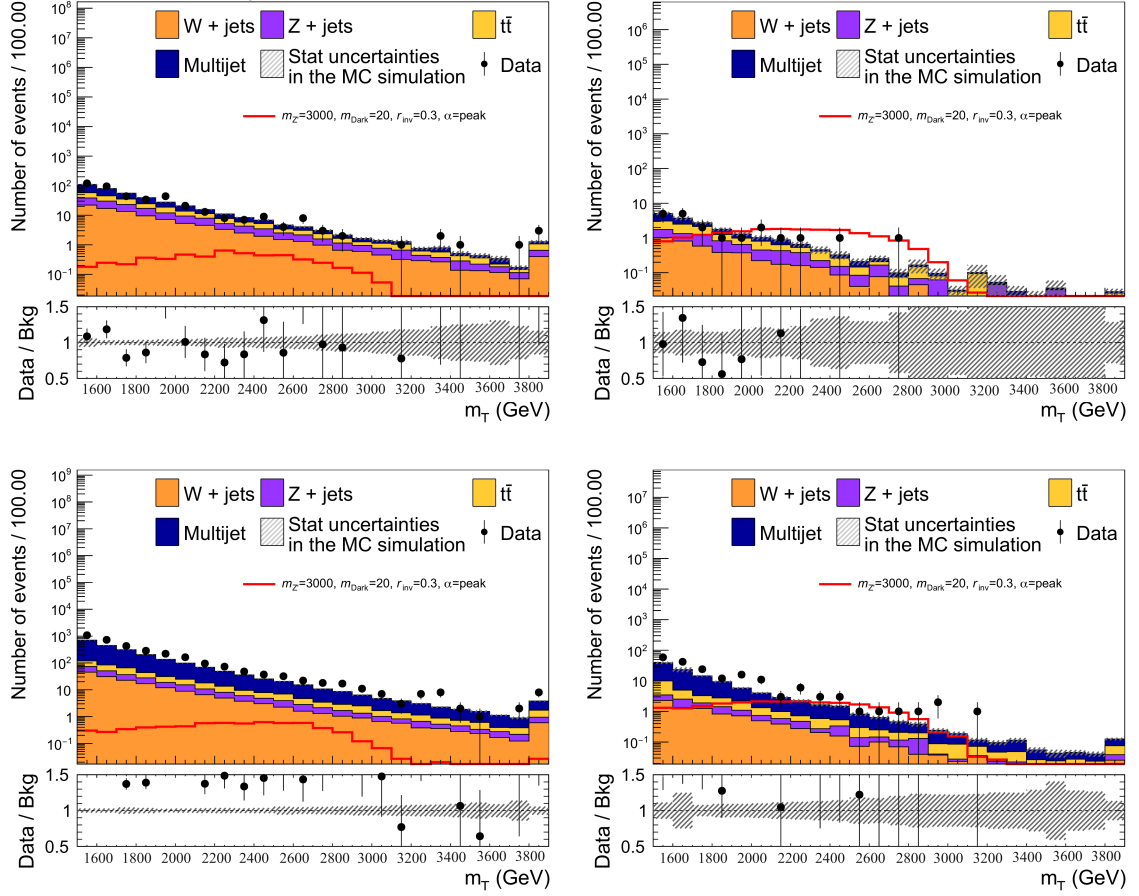


Figure 5.18: The transverse mass distribution in the signal regions on the benchmark signal sample (red) and on SM background (filled histograms) in 2018, before the HEM issue appeared. The categories with high R_T are presented in the top row, while at the bottom the lower R_T selection is applied. One SVJ jet is requested in the left column and both jets must fulfill the tagger requirement in the right column.

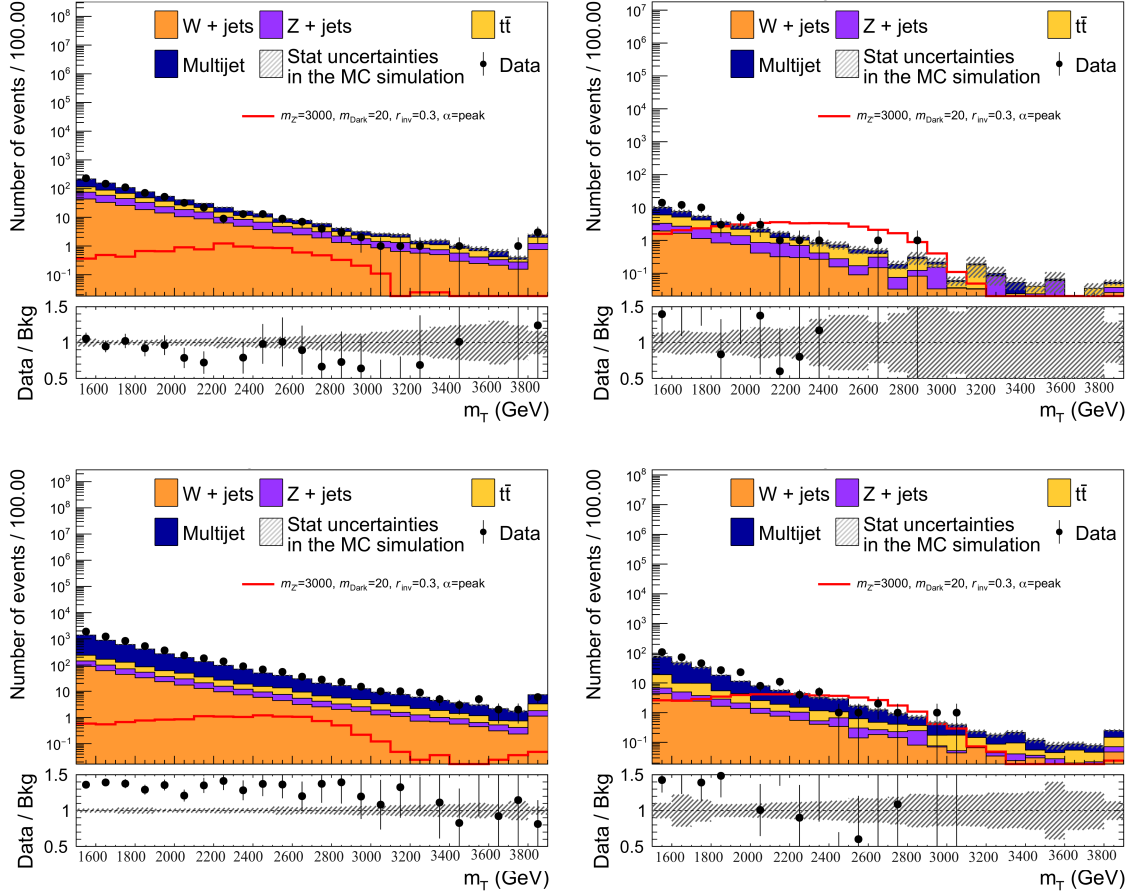


Figure 5.19: The transverse mass distribution in the signal regions on the benchmark signal sample (red) and on SM background (filled histograms), in 2018 after the HEM issue appeared. The categories with high R_T are presented in the top row, while at the bottom the lower R_T selection is applied. One SVJ jet is requested in the left column and both jets must fulfill the tagger requirement in the right column.

5.4.2 Fitting strategy

The background in the signal region is estimated from data through a fit to the m_T distribution. Several fitting functions have been tested. In particular, a parametric continuous function, which has been used extensively in previous searches to fit the falling dijet spectrum [187, 188], was found to be particularly suitable for the case under study. Additional possibilities such as exponential and polynomial functions have been evaluated as well. The function that resulted to better describe the m_T distributions across the four signal regions defined in this analysis can be expressed as:

$$f(x) = \frac{(1+x)^{p_1}}{x^{p_2+p_3} \log(x) + p_4 \log^2(x)}. \quad (5.8)$$

To decide how many parameters are effectively needed, a Fisher-test of significance is performed. The procedure that is employed consists in the following steps:

1. given two models (labelled with f_1 and f_2), the F value is given by:

$$F_{21} = \frac{\frac{RSS_1 - RSS_2}{n_2 - n_1}}{\frac{RSS_2}{N - n_2}}, \quad (5.9)$$

where N is the number of data points, n_1 and n_2 are the number of parameters for the models f_1 and f_2 (with $n_2 > n_1$), RSS_i is the residual sum of squares of model i ,

$$\sum_{bins} (\text{data}_{bins} - \text{fit}_{bins})^2. \quad (5.10)$$

2. Under the null hypothesis that model 2 does not provide a significantly better fit than model 1, the F variable will have an F-distribution with $(n_2 - n_1, N - n_2)$ degrees of freedom. The observed confidence level CL_{21} can be defined as

$$CL_{21} = 1 - \int_{-\infty}^{F_{21}} F_{\text{distribution}}(n_2 - n_1, N - n_2). \quad (5.11)$$

3. The null hypothesis is rejected if CL_{21} is smaller than the desired probability, 0.10.

Table 5.8 reports the resulting p-values of the Fisher test. It can be observed that the optimal number of parameters is 2 for the lowSVJ1, lowSVJ2, highSVJ1 categories and 1 for the highSVJ2 category. The normalization, that is left as free-floating in the fit, is not considered in this count.

To have a reliable method to evaluate the fit performance, the strategy has been tested on "random data" instead of pure MC background samples. To emulate data, events are generated using as p.d.f. the MC shape and as normalization factor the expected background yield cast to integer. Figure 5.24 displays the m_T distributions fitted with the resulting optimal functions.

category	1 - 2 parameters	2 - 3 parameters	3 - 4 parameters
lowSVJ1	0.00	1.00	1.00
highSVJ1	0.00	1.00	1.00
lowSVJ2	0.00	1.00	1.00
highSVJ2	0.50	1.00	0.28

Table 5.8: Summary of the p-values obtained from the Fisher test on the nominal function. The results are reported for the combinations of consecutive parameters and for all categories.

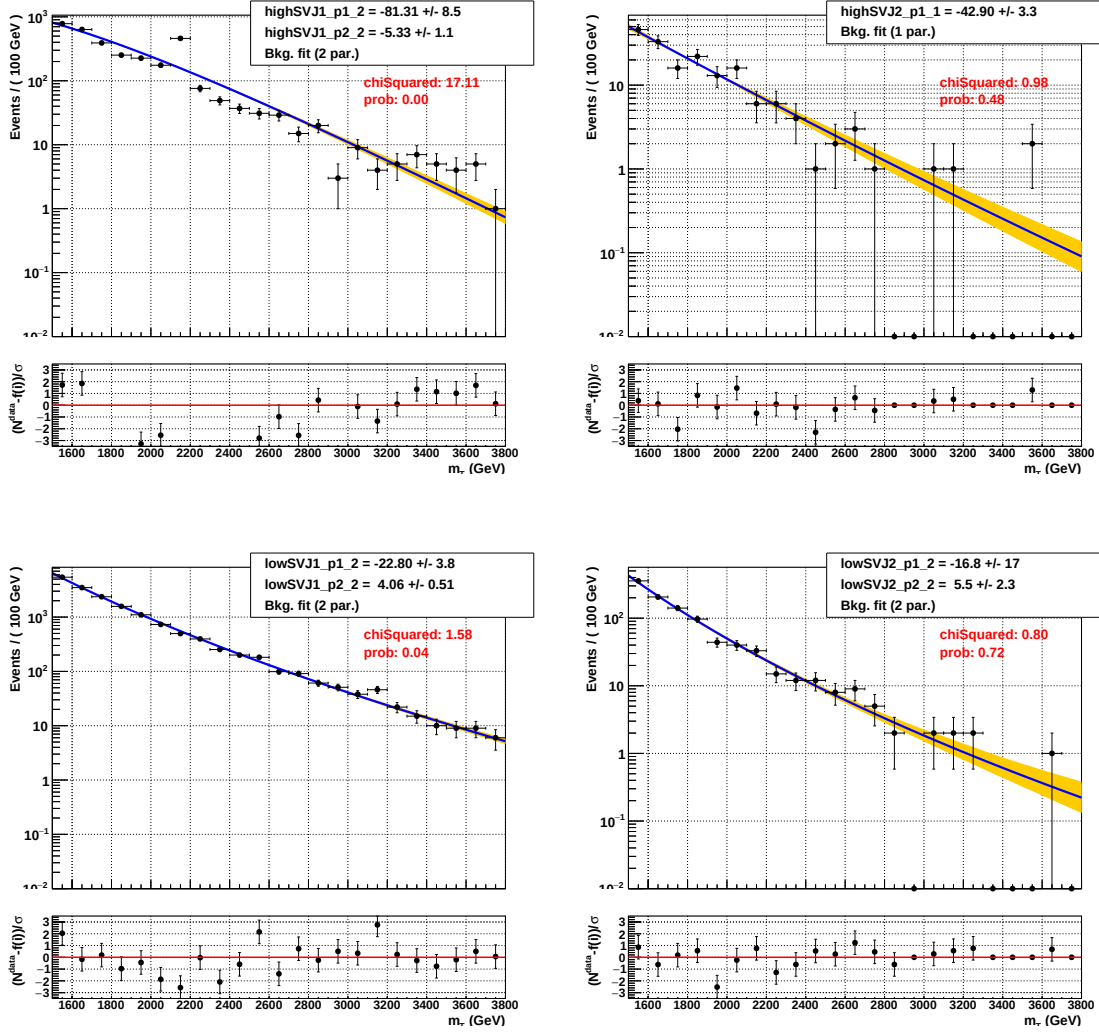


Figure 5.20: The fitted transverse mass distribution in the signal regions on random data. The nominal function with the optimal number of parameters as returned by the Fisher test is used and indicated with the blue line. The categories with high R_T are presented in the top row, while at the bottom the lower R_T selection is applied. One SVJ jet is requested in the left column and both jets must fulfill the tagger requirement in the right column.

5.5 SVJ tagger working point optimization

A dedicated optimization is performed to identify the most suitable BDT working point needed to determine if an individual jet is tagged as a SVJ. With tighter selections, a more

pure subset of events is identified, but the statistics diminishes rapidly as well. Hence, it is important to find a good compromise between the maximization of the signal to background separation and the statistics needed to have a reliable description of the SM background.

Several BDT working points have been tested: 0.4, 0.45, 0.5, 0.55, 0.6, 0.7, 0.8. First, the final expected 95% CL upper limits on the product of the Z' cross-section and branching fraction are compared. Secondly, the reliability of the SM background estimation is tested. The FOM that is here considered is the error on the fit, computed as the sum over all bins of the m_T distribution of the normalized errors in each bin. Finally, fits have been performed on random data as well, and final exclusion limits have been computed accordingly. Figure 5.21 shows a comparison among the median expected 95% CL upper limits for 2017 on the product of the Z' cross-section and branching fraction on emulated data, for the tested BDT working points. As a general trend, with tighter working points lower limits are obtained. The upper limits for each working point are summarized in Figure 5.22 (left). Figure 5.23 illustrates the fits on the emulated data in the purest event category, the highSVJ2 category, and a summary of the errors is presented in Figure 5.22, for each of the signal regions used in this search and defined by the multiplicity of tagged semi-visible jets and the purity (i.e. highSVJ1, highSVJ2, lowSVJ1 and lowSVJ2). From these studies, it resulted that for working points equal or larger than 0.6 the statistics is too limited in the highSVJ2 category. Hence, to maximize the power of the analysis and have sufficient statistics for the final fit, a $wp = 0.55$ resulted to be the most appropriate choice.

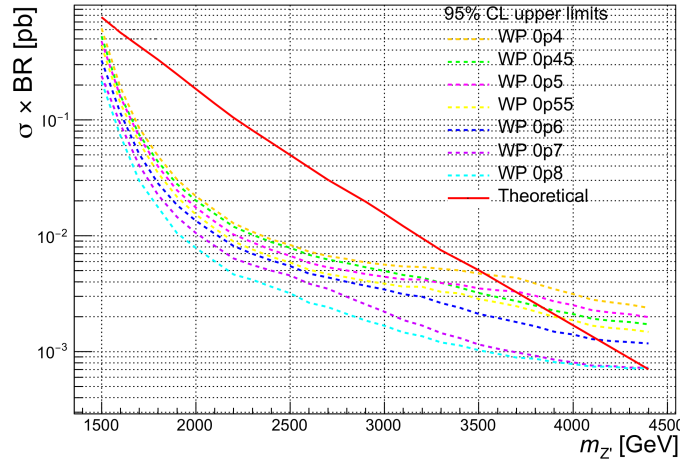


Figure 5.21: Median expected 95% CL upper limits for 2017 on the product of the Z' cross-section and branching fraction as a function of the mediator mass on emulated data. Each lines corresponds to a BDT working point. The red line indicated the theoretical cross-section.

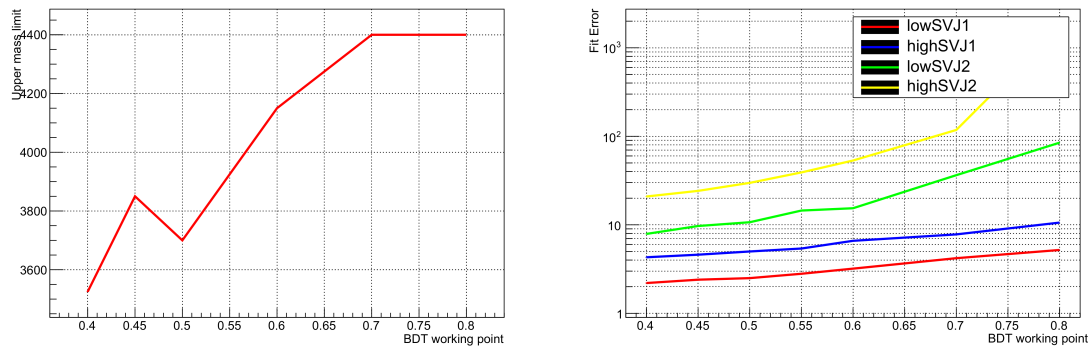


Figure 5.22: Upper mass limits obtained using emulated data (left) and error on the fit for the different signal regions (right) as a function of the BDT working point. Only 2017 backgrounds are considered.

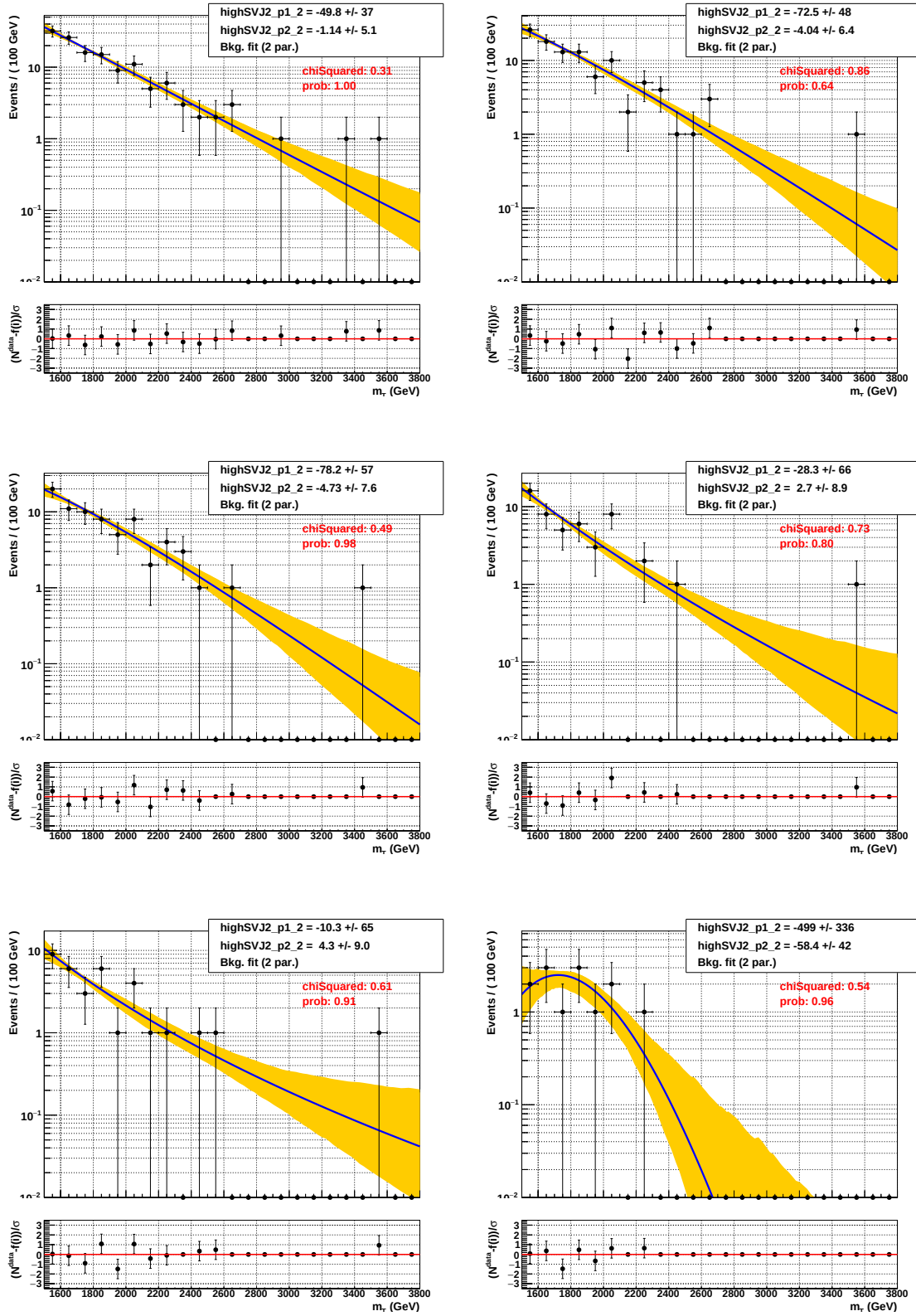


Figure 5.23: Fits on emulated data using 2017 MC backgrounds on the highSVJ2 category. Each figures corresponds to a different BDT working point, ranging from 0.4 to 0.75. The χ^2 and its probability are also shown. In the lower pad, the pulls, defined as the ratio between the difference between the data yields and the value of the fit, over the error on data, are reported.

5.6 Systematics uncertainties

The reconstructed m_T distribution is affected by multiple experimental and theoretical systematic uncertainties. Depending on the source of the uncertainty, its impact can either be an overall uniform effect (normalization uncertainties) or it might include also distortions of the distribution shape (shape uncertainties). Log-normal distributions and morphing techniques are used to model the former and the latter type of uncertainty, respectively. The approach pursued to consider their effect is the same that is employed in the search for the VLQ B, described in Section 4.4.

In this search, the sources of systematic uncertainties that are considered are:

- the trigger efficiency (described in Section 5.2.1) and a 2% normalization factor accounting for limited amount of data in the control region;
- the uncertainty due to the limited amount of simulated signal events in the MC signals;
- the systematic uncertainty associated to the total integrated luminosity recorded by CMS;
- the fluctuations and errors on the fits parameters and on the background normalization.

5.7 Results and interpretation

Observations are in agreement with the background-only hypothesis in all categories defined in the analysis, as shown in Figure 5.24.

Therefore upper limits are placed on the production cross-section as a function of the mediator mass, $m_{Z'}$. The 95% confidence level exclusion limits are derived from the modified frequentist approach (CLs), using the profile-likelihood test statistics. A simultaneous maximum likelihood fit is performed to the m_T distributions considering the combination of all signal regions, with the expected background templates given in Section 5.4, and the signal template given by the MC simulated distributions. The systematic uncertainties described in Section 5.6 are taken into consideration in the fit and treated as nuisance parameters.

The expected and observed upper limits on the semi-visible jets production cross-section are presented in Figure 5.25, considering all data collected during Run 2. The specific Hidden Valley model probed in this thesis are excluded for mediator bosons with mass up to 3.9 TeV. A slight excess of events is observed for high values of $m_{Z'}$ masses, limiting the exclusion power of the search. This excess, though, is not statistically significant and will require additional the amount of data expected to be collected collected by CMS during the Run 3 of LHC to confirm or disprove the excess.

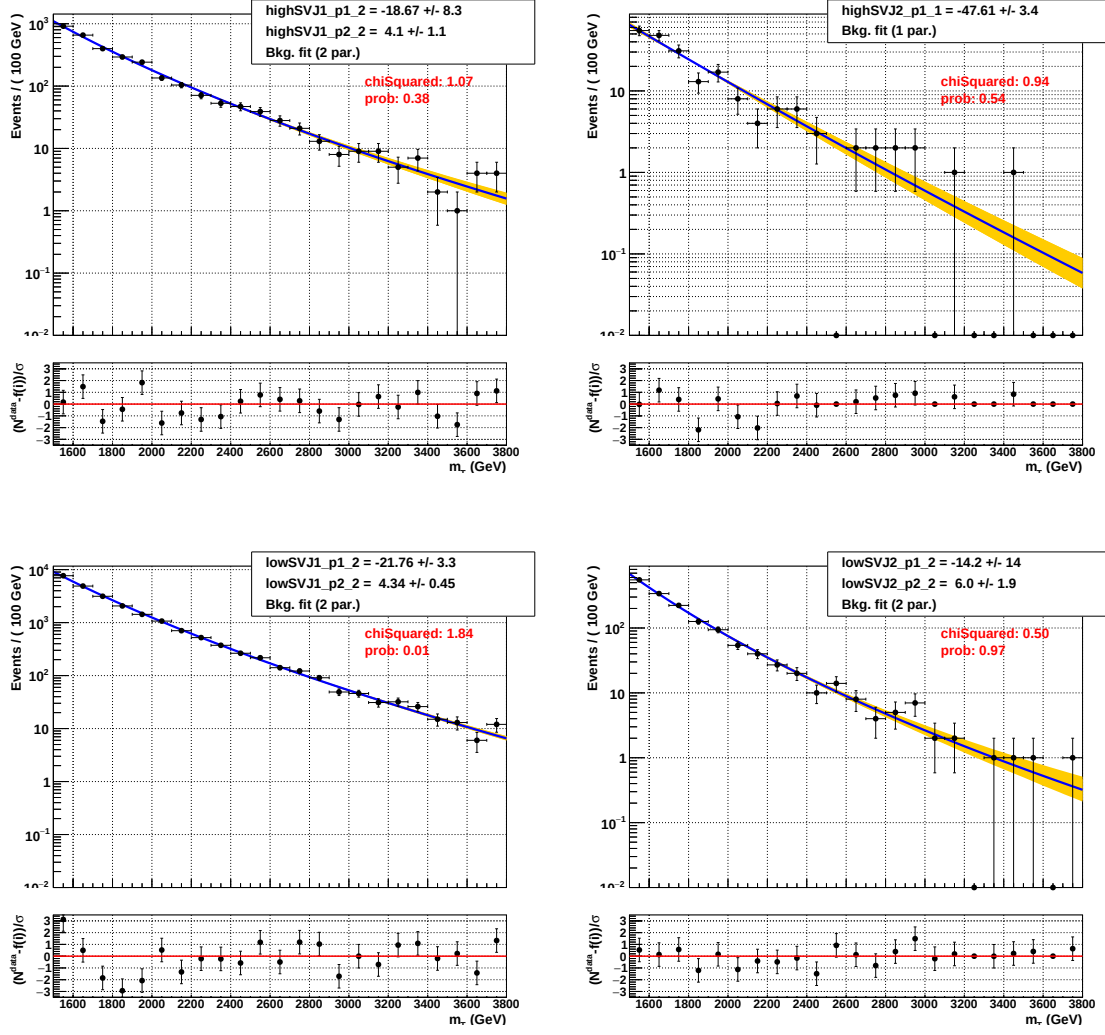


Figure 5.24: The fitted transverse mass distribution in the signal regions on data. The nominal function with the optimal number of parameters as returned by the Fisher test is used and indicated with the blue line. The categories with high R_{T} are presented in the top row, while at the bottom the lower R_{T} selection is applied. One SVJ jet is requested in the left column and both jets must fulfill the tagger requirement in the right column.

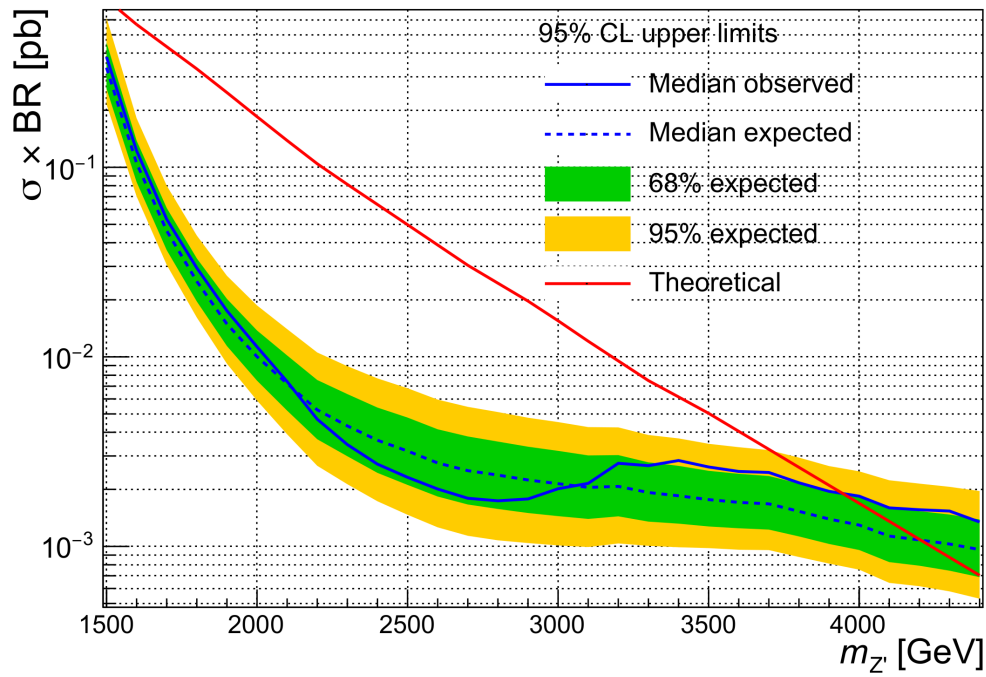


Figure 5.25: Expected 95% confidence level upper limits on the semi-visible jets production cross-section using 13 TeV data collected during the full Run 2. The 1 and 2 standard deviations (σ) bands on the expected limits are shown respectively in green and yellow. Theoretical cross-sections (Table 5.2) are indicated by the red line.

Chapter 6

Conclusions

“ Assez vu. La vision s’est rencontrée à tous les airs.
Assez eu. Rumeurs des villes, le soir, et au soleil, et toujours.
Assez connu. Les arrêts de la vie. - Ô Rumeurs et Visions !
Départ dans l’affection et le bruit neufs ! ”

Arthur Rimbaud, *Les illuminations*, 1886

In this doctoral thesis, I have presented two searches for new physics beyond the SM in fully hadronic final states using proton-proton collision data collected by the CMS experiment at 13 TeV. These searches are highly statistically populated, which makes them very interesting. Both searches developed different techniques to overcome the overwhelming QCD multijet background contamination.

The first search documented in this thesis focused on the electroweak production of a VLQ B decaying to a Higgs boson and a b quark. The analysis uses data collected by CMS in 2016 corresponding to an integrated luminosity of 35.9 fb^{-1} . The final state of this process is composed of a boosted Higgs boson decaying to a pair of bottom quark, and a bottom and a light flavor quarks. Jet substructure techniques have been developed to identify the boosted Higgs boson and a dedicated optimization strategy has been used to reconstruct the VLQ B resonance and enhance the signal to background separation. In addition, in order to handle the huge contamination arising from multijet background processes, data-driven techniques were implemented to accurately extrapolate the background contribution from data. To extract a potential signal appearing as a localized excess of data over the SM background in the spectrum of the invariant mass between the b quark and the Higgs boson, a binned maximum likelihood is performed. Given that no significant excess was found over the background expectation, this search allowed us to set for the first time upper limits on the mass of the electroweak VLQ B. In addition, this analysis was the first one exploring finite resonance widths scenarios to ever be performed. To perform these studies, a dedicated theoretical framework was developed and that paved the way for other searches to follow a newer model of interpretation.

The second search discussed in this doctoral thesis was the search for DM in the

context of Hidden Valley models. The search focused on the existence of two dark quarks, that shower and hadronize, manifesting as semi-visible jets. This search used data collected during 2016, 2017 and 2018 by the CMS experiment, corresponding to an integrated luminosity of 137 fb^{-1} . Unlike the typical signatures of WIMPs, that are characterized by the missing momentum to be recoiling against the visible jets, in semi-visible jets the missing momentum is aligned to its visible part. An optimized signal selection has been developed and a dedicated tagger aiming at distinguishing semi-visible jets from SM jets has been designed and implemented. One of the main features exploited by the tagger is the QG tagger, that I developed and validated during my PhD. To model the background, a data-driven strategy, based on the fit of the resonance transverse mass distribution, was used. No excess of events over the expected SM background is observed and exclusion limits for mediator boson mass up to 3.9 TeV are set for the specific Hidden Valley probed. With this analysis, I pioneered a new branch of DM search not relying anymore on the WIMPs assumption, but predicting DM made of a multitude of particles interacting among themselves via a new fundamental force similar to the SM strong force.

Within this thesis, I investigated new regions where BSM particles could hide in the collisions delivered by the LHC. This thesis opens new branches of exploration, examining new scenarios of VLQs and exploiting innovating models that predict a strongly coupled DM. Dedicated tools and techniques, such as the QG and SVJ tagger, were developed to deal with the complicated hadronic environments in which these searches were performed. In the future, these optimized tools will contribute to the directions of searches, which will also profit from the data that will be collected at the LHC and High Luminosity LHC.

Appendices

Appendix A

Pixel detector resolution

As described in Chapter 2, the CMS pixel detector is the innermost sub-detector of CMS and the closest to the interaction region. Good quality of tracking performance in terms of spatial resolution is crucial for the searches presented in this thesis. Hence, dedicated studies to measure and monitor the resolution are performed.

The method used in CMS is explained in Section A.1. To estimate the charged particle tracks, it is fundamental to measure the charge deposits they leave when passing through the pixel detector. Section A.2 focuses on the reconstruction of clusters from charge deposits. The algorithms used to estimate the position of the track hits are presented in Section A.3. Finally, the results obtained in Run 2 are shown in Section A.4.

A.1 The triplet method

To measure the intrinsic spatial resolution of the pixel detector, both in the barrel and in the endcaps, the so-called "triplet method" is exploited. The workflow of this algorithm is the following:

1. three sub-detectors (layers or disks) are studied at once and tracks are reconstructed from hits in these sub-detectors;
2. the helicoidal trajectory of charged particles is determined using the curvature of the reconstructed track and two of the three hits;
3. the extrapolated (or interpolated) hit is measured as the intersection between the helix and the layer or disk that was not used in the track reconstruction;
4. the residuals are then computed as the difference between the measured hit and the extrapolated hit.

The residuals distribution is finally fitted with a Student's t function:

$$f(x) = B + \frac{A}{\sigma\sqrt{\nu\pi}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \exp\left(-\frac{\nu+1}{2} \log\left(1 + \frac{1}{\nu} \left(\frac{x-x_0}{\sigma}\right)^2\right)\right), \quad (\text{A.1})$$

with free parameters B , A , the shift x_0 , the spread σ , and the number of degrees of freedom ν . The resolution is considered to be the width of the residuals distribution, which is taken from the σ parameter returned by the fit.

A.1.1 Triplet combinations

The combinations of layers and disks used to determine the resolution changed from phase 0 pixel detector to phase 1 pixel detector. When the phase 0 pixel detector was installed, the residuals were computed:

- extrapolating from hits in layers 2 and 3 to layer 1;
- interpolating from hits in layers 1 and 3 to layer 2;
- extrapolating from hits in layers 1 and 2 to layer 3;
- interpolating from hits in layer 1 and disk 2 to disk 1;
- extrapolating from hits in layer 1 and disk 1 to disk 2.

Starting from 2017, with the replacement of the pixel detector and the addition of one layer and one disk, residuals were measured:

- extrapolating from hits in layers 2 and 3 to layer 1;
- interpolating from hits in layers 1 and 3 to layer 2;
- interpolating from hits in layers 2 and 4 to layer 3;
- extrapolating from hits in layers 2 and 3 to layer 4;
- extrapolating from hits in layer 2 and disk 2 to disk 1;
- interpolating from hits in disk 1 and disk 3 to disk 2;
- extrapolating from hits in disk 1 and disk 2 to disk 3.

A.1.2 Coordinate system

As shown in Figure A.1, the global coordinates system does not correspond to the local reference used generally in CMS. For the barrel detector, the local x -direction corresponds to the global $r\phi$ direction, while the local- y is the global z . In the disks, the local x is comparable to the global r , while the local y -direction can be referred to as the global $r\phi$ direction.

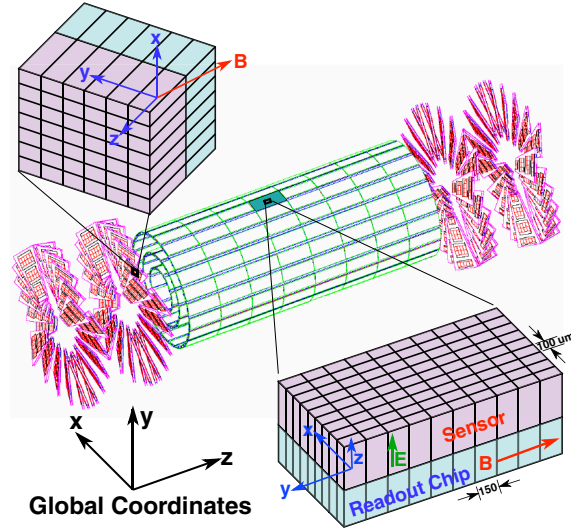


Figure A.1: Geometry of the CMS barrel and forward pixel detectors as in Phase 0 and the corresponding coordinate systems [189].

A.2 Pixel clusters

When a charged particle passes through a pixel sensor it deposits some charge, as drafted in Figure A.2. The charge is shared among multiple pixel modules, leading to the formation of pixel clusters. The charge sharing effect can be motivated by the fact that particle tracks impact the silicon sensor with a small angle (α and β , respectively along the local x and local y directions). Also, there is a charge drifting effect in the magnetic field in which the tracking system is immersed.

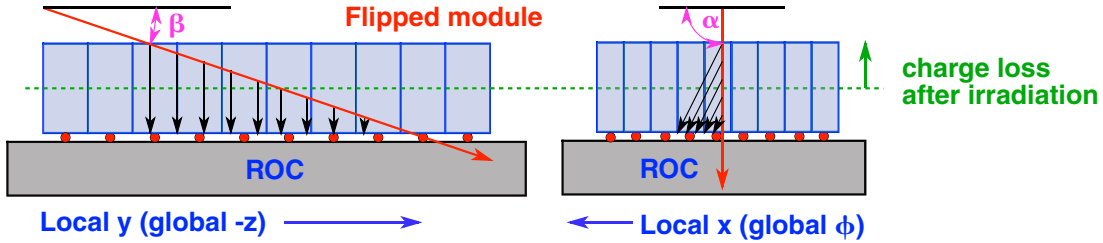


Figure A.2: Charge deposition and drift in a pixel barrel module, left in the plane parallel to the beam, right in the plane transverse to the beam [189].

To better understand the creation of pixel clusters, Figure A.3 shows a sketch of a typical cluster in the barrel. The dashed red line indicates the track projection on the cluster and the red cross represents the geometrical center of the particle track, that is the true hit position. Signals below a fixed pixel threshold (i.e. 2000 electrons) are shown in green.

The position resolution of the pixel sensors is mainly determined by the readout pitch. Besides, the spatial resolution is driven by the charge sharing between neighboring pixels, which is mainly affected by geometrical sharing (left sketch in Figure A.2) and Lorentz deflection (right sketch in Figure A.2). The reconstruction of hits from clusters is done separately for the two coordinates. Along the x -projections there are mainly short clusters,

In the ideal case of a single pixel cluster, the particle hit position is just given by the pixel center. In the more realistic scenario of a longer cluster, the hit position x_{hit} is determined by:

$$x_{hit} = \frac{\sum_i q_i \cdot x_i}{\sum_i q_i} \quad (\text{A.4})$$

where q_i is the charge deposited in the pixel x_i .

Along the longitudinal axis, because of the lack of magnetic deflection, the cluster width is only determined by the electric field drift, charge diffusion, and the impact angle β . The cluster width W_y is then given by $W_y = T \cdot \tan(\beta)$ and the hit position can be expressed as:

$$y_{hit} = y_{center} + \frac{q_{last}^y - q_{first}^y}{2(q_{first}^y + q_{last}^y)} ||W_y| - W_{inner}^y| \quad (\text{A.5})$$

where y_{hit} is the particle hit position, y_{center} is the cluster geometrical center, W_{inner}^y is the length of the inner pixels and q_{first}^y (q_{last}^y) is the charge in the first (last) pixel. W_{inner}^y is defined as

$$W_{inner}^y = (y_{size} - 2) \times P, \quad (\text{A.6})$$

with y_{size} being the cluster size and P the pixel pitch.

A.3.2 Template algorithm

A more precise and powerful algorithm makes use of predefined cluster shape templates to find a very accurate hit position. More in detail, this algorithm is based on fitting the pixel cluster x and y projections to pre-determined cluster shapes called *templates*. These predefined maps are generated with a simulation program (i.e. Pixelav [190]).

Given a track with angles α and β incident to the pixel module, the simulated template is compared to the real cluster generated by the track understudy. The hit positions along x and y are treated as floating variables and a χ^2 minimization is performed. The values returned as the ones minimizing the comparison are then stored and considered as the hit position.

A.4 Results

The pixel hit resolution, both in the barrel and in the forward section of the detector, is continuously monitored during the data taking and the results obtained from 2015 to 2018 are presented from Figure A.5 to Figure A.11.

A.4.1 Results in 2015

In Figure A.5 the hit resolution in the barrel (top) and in the forward (bottom) pixel depending on the instantaneous luminosity in 2015 is reported. It can be observed that the resolution is very stable in the barrel, while for the forward pixel, thanks to a reprocessing of the track reconstruction, better conditions were achieved along with the LHC running, resulting in improved results. The template reconstruction (blue dots) is also significantly better than the generic reconstruction (red dots).

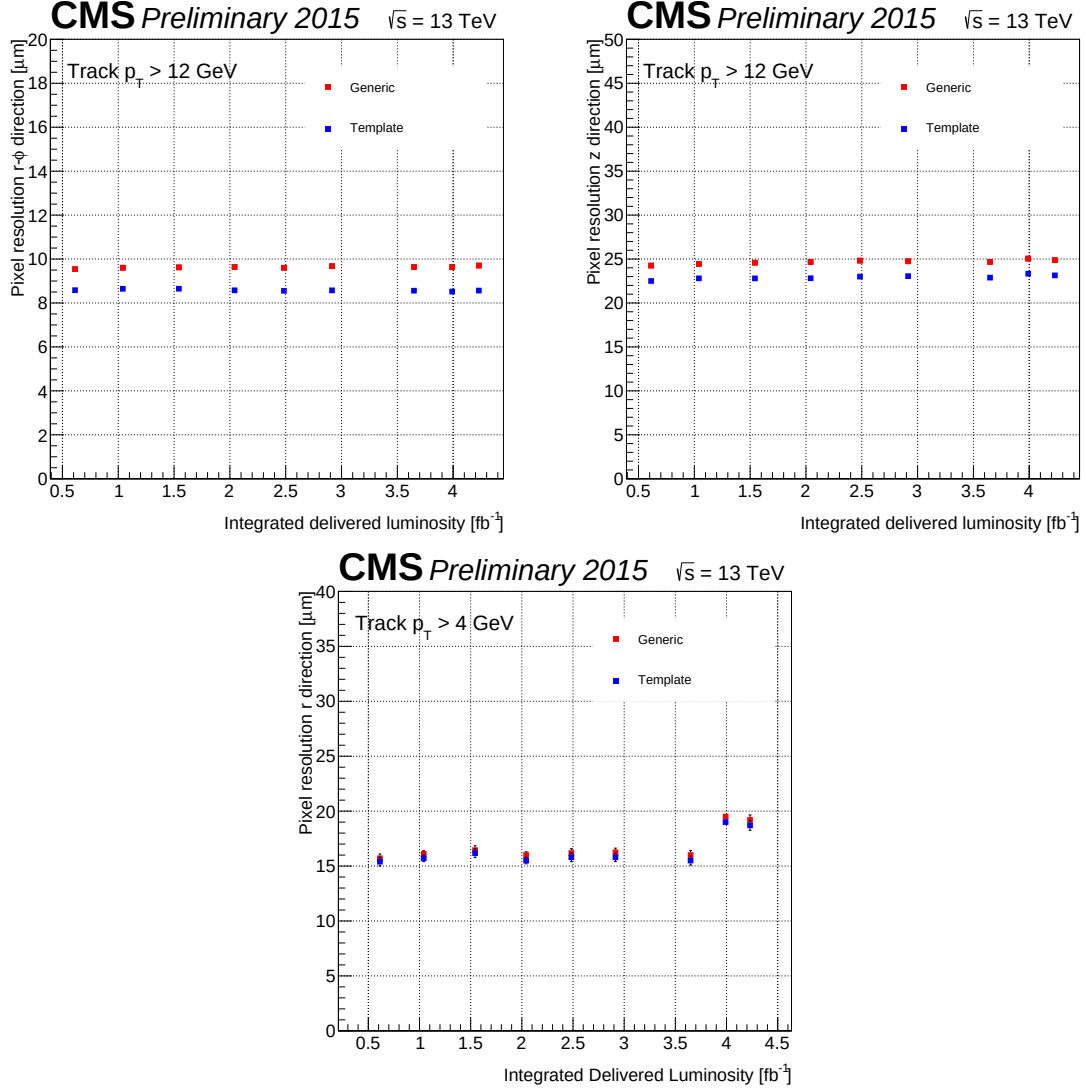


Figure A.5: Resolution distributions as a function of the integrated luminosity of data collected during 2015. The top row shows the results for the second layer of the barrel pixel along the $r\text{-}\phi$ direction (left) and z (right). The bottom row presents the results for the second disk of the forward pixel along the r -direction. In all plots a comparison between the reconstruction methods is presented, where the red dots labels the generic algorithm and the blue the template [191].

A.4.2 Results in 2017

In Figure A.6 the hit resolution for the second and the third barrel layers, obtained with the template reconstruction and 2017 data, is presented.

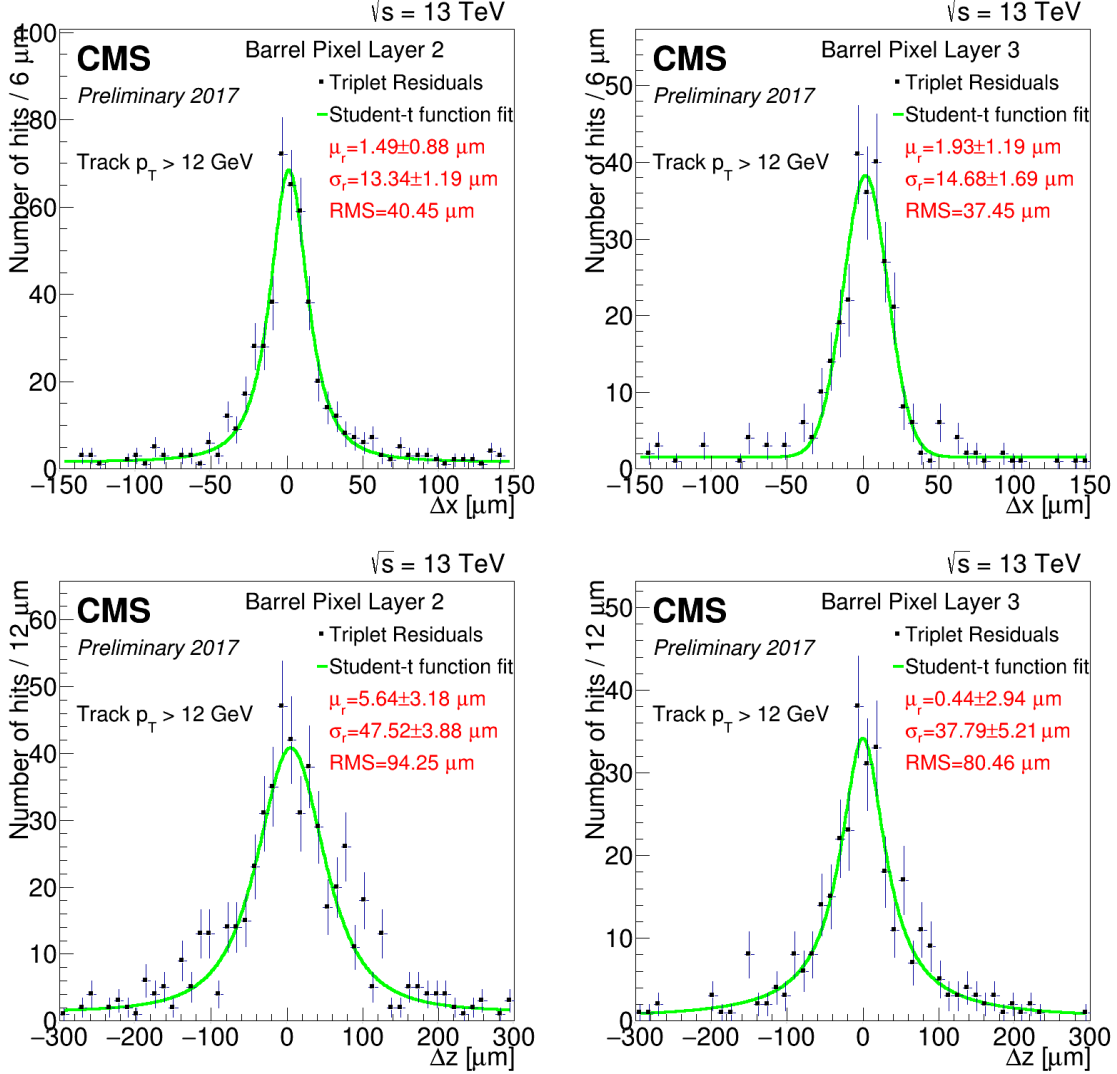


Figure A.6: Residuals distributions along the r - ϕ direction (top) and z (bottom) for the second (left) and third (right) layer of the forward pixel detector obtained with template reconstruction for 2017 data [192].

A.4.3 Results in 2018

A comparison between the results as obtained in simulation and measured on 2018 data is performed and the results are shown in Figures A.7, A.8, A.9, A.10 for the barrel pixel and in Figures A.11 for the second disk of the forward pixel.

From these studies, it can be observed that data are in agreement with Monte Carlo simulations, or in some cases are even performing better than the expectations. This drives to the conclusion that the calibrations performed, the conditions used and the adjustments applied during the data taking were correct and the quality of the data delivered for the physics analysis was high.

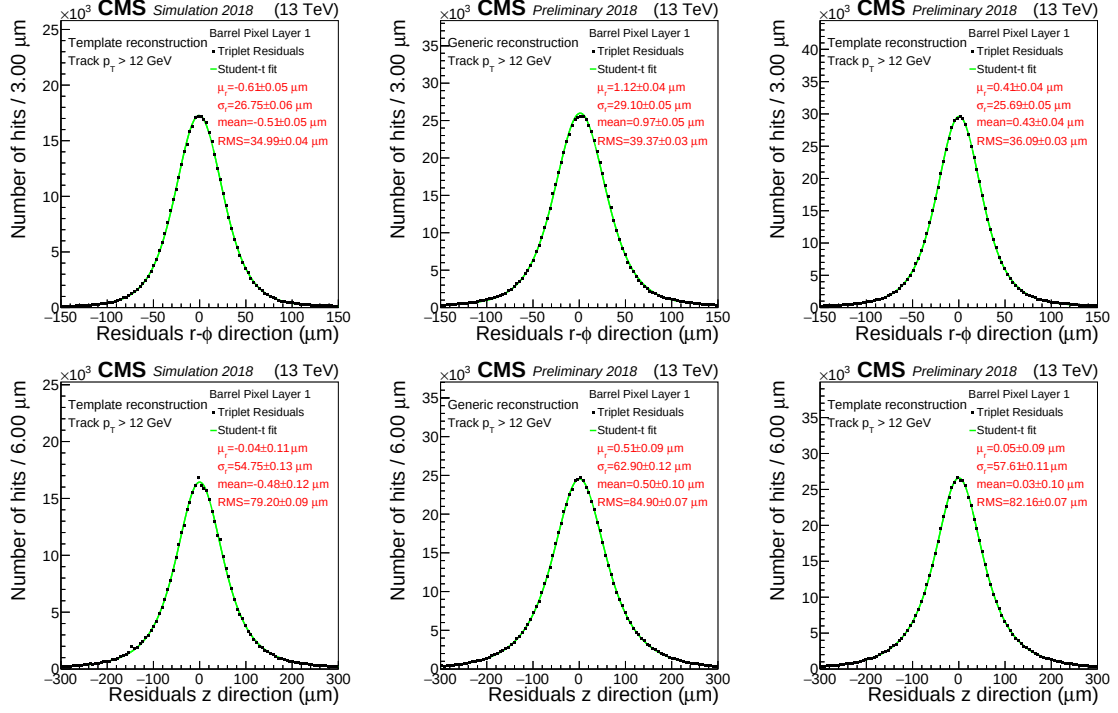


Figure A.7: Residuals distributions along the $r\text{-}\phi$ direction (top) and z (bottom) for the first layer of the barrel pixel detector. The first column show the results obtained on simulation with template reconstruction, while in the second and third column 2018 data are presented, analysed with generic and template reconstruction, respectively [193].

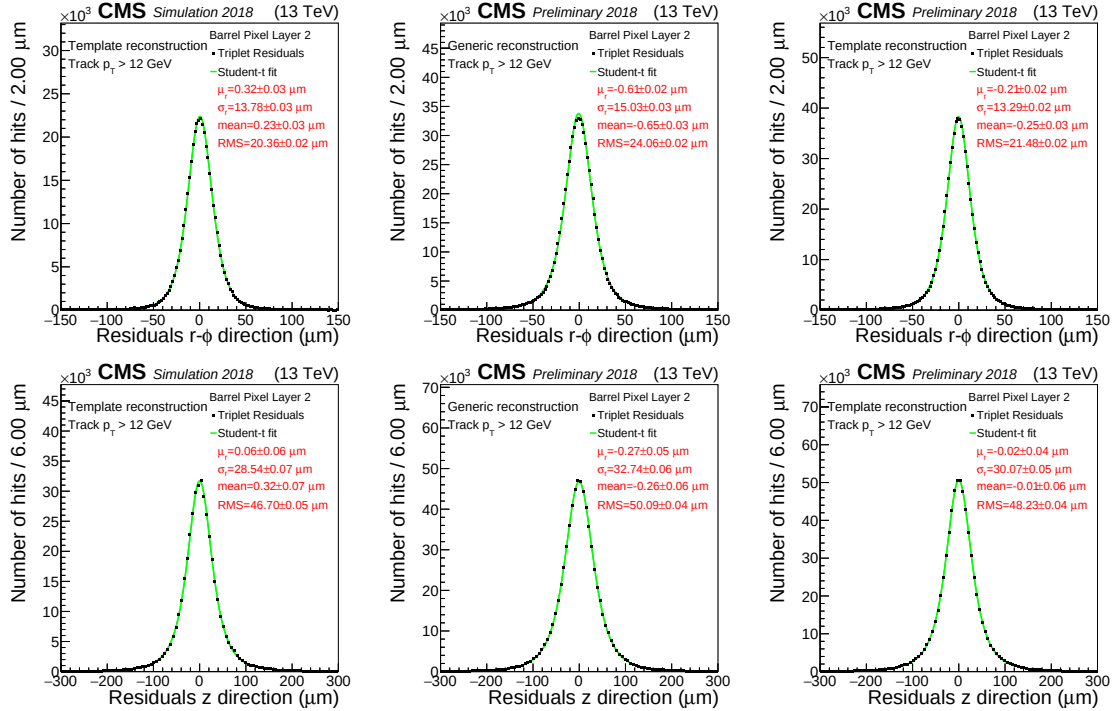


Figure A.8: Residuals distributions along the $r\text{-}\phi$ direction (top) and z (bottom) for the second layer of the barrel pixel detector. The first column show the results obtained on simulation with template reconstruction, while in the second and third column data are presented, analysed with generic and template reconstruction, respectively [193].

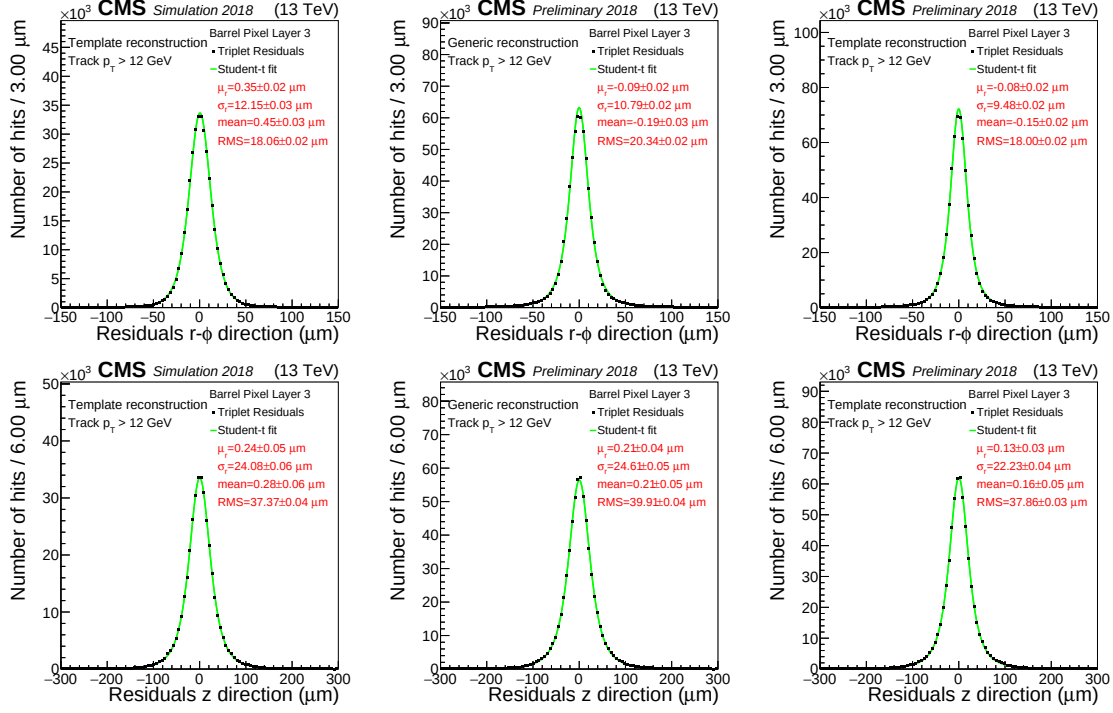


Figure A.9: Residuals distributions along the $r\text{-}\phi$ direction (top) and z (bottom) for the third layer of the barrel pixel detector. The first column show the results obtained on simulation with template reconstruction, while in the second and third column 2018 data are presented, analysed with generic and template reconstruction, respectively [193].

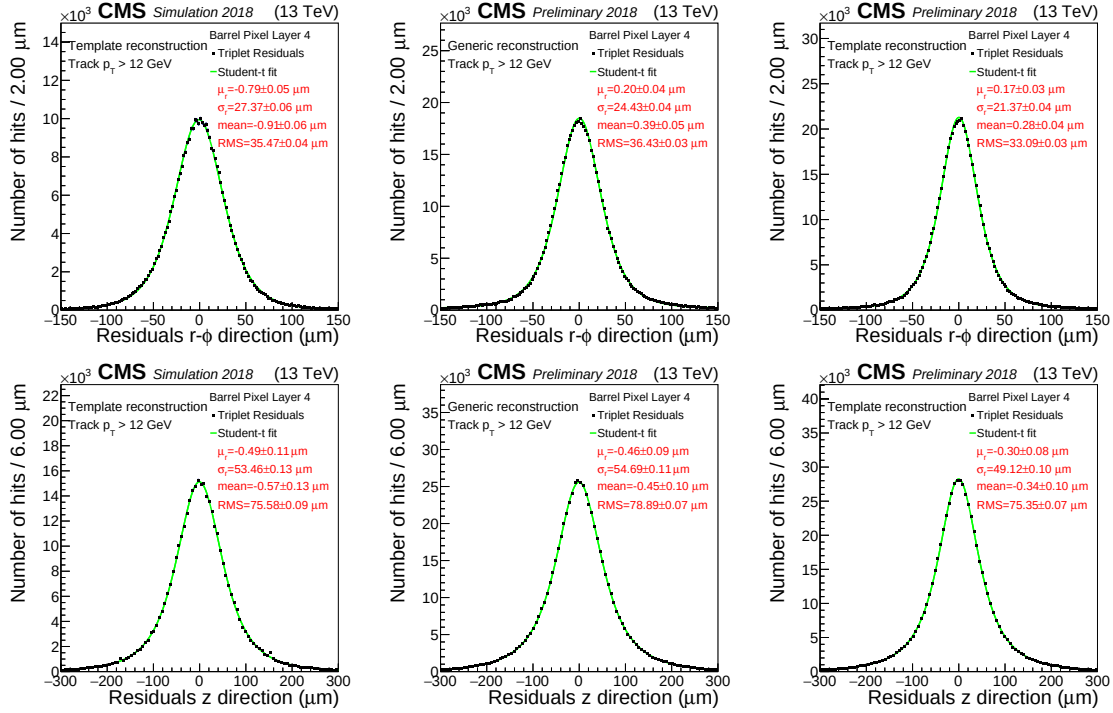


Figure A.10: Residuals distributions along the $r\text{-}\phi$ direction (top) and z (bottom) for the fourth layer of the barrel pixel detector. The first column show the results obtained on simulation with template reconstruction, while in the second and third column 2018 data are presented, analysed with generic and template reconstruction, respectively [193].

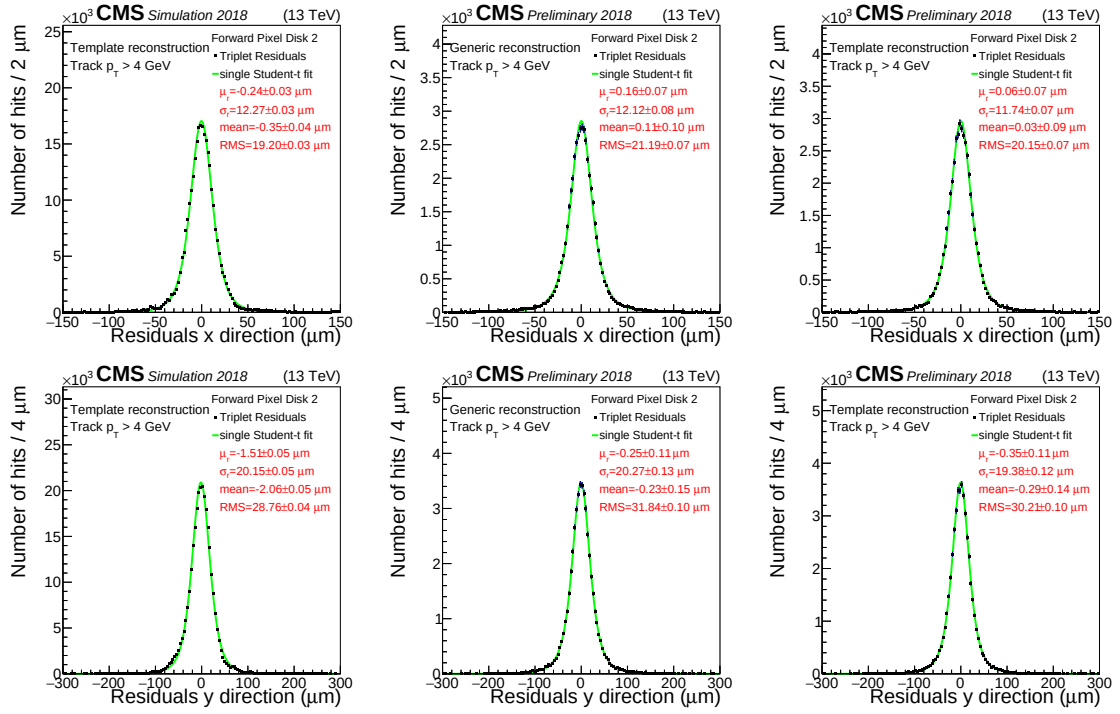


Figure A.11: Residuals distributions along the local-x direction (top) and local-y (bottom) for the second disk of the forward pixel detector. The first column show the results obtained on simulation with template reconstruction, while in the second and third column 2018 data are presented, analysed with generic and template reconstruction, respectively [193].

Appendix B

The quark-gluon jet tagger

As predicted from the theory and confirmed by many experimental measurements [194–196], hadronic jets that are initiated by gluons show differences with respect to the ones originated by light-flavor quarks (u, d, s). These differences have been extensively studied and validated. CMS has developed a tagger to distinguish between the two types of jets, called quark-gluon tagger, based on the characteristics of the hadronic shower. Several physics analyses in CMS have used it, to improve Higgs boson searches [197, 198] and perform electroweak measurements [199, 200].

In this appendix, the design, implementation, and performance of a dedicated quark-gluon tagger are described. The focus is on the data collected by the CMS experiment during the pp collisions held in 2015, but similar studies are currently being performed on the 2016, 2017 and 2018 data as well.

B.1 Observables

The observables used for the quark-gluon tagger are based on the PF candidate jet constituents. To ensure both robustness with respect to particle reconstruction and PU contamination, and sensitivity to the jet fragmentation differences, charged constituents should be well associated with the primary interaction vertex. Also, neutral constituents should have $p_T > 1$ GeV. Based on these inputs three main variables have been chosen, corresponding to the main discriminating properties between quark- and gluon-induced jets. The three observables are described below.

The jet constituents multiplicity This variable is defined as the total number of PF candidates reconstructed within the jet. The ratio of the jet constituent multiplicities from quark and gluon jets should converge to the color factor ratio $\frac{C_A}{C_F} = \frac{9}{4}$ and provide good discrimination. Therefore, quark-like jets multiplicity is expected to be lower than the one in gluon-like jets.

The jet minor angular opening (σ_2) The conical shape of the jets can be projected in the η - ϕ plane, and the width of the jets can be quantified in terms of the two principal components of the second moment of the constituent distribution. In the η - ϕ plane the

shape of the jet can be assumed an ellipse, whose main elements are the principal axes. A 2×2 matrix can be defined as:

$$M = \begin{pmatrix} \sum_i p_{T,i}^2 \Delta\eta_i^2 & -\sum_i \Delta\eta_i \Delta\phi_i \\ -\sum_i \Delta\eta_i \Delta\phi_i & \sum_i p_{T,i}^2 \Delta\phi_i^2 \end{pmatrix}, \quad (\text{B.1})$$

where i is the index referring to the jet constituents, $p_{T,i}$ is the transverse momentum of the i -th constituent, $\Delta\eta_i$ and $\Delta\phi_i$ are the differences in pseudorapidity and azimuthal angle between the i -th constituent and its average value.

The major (σ_1) and minor (σ_2) axes of the jet can be computed from the eigenvalues $\lambda_{1,2}$ of the matrix:

$$\begin{aligned} \sigma_1 &= \sqrt{\lambda_1 / \sum_i p_{T,i}^2} \\ \sigma_2 &= \sqrt{\lambda_2 / \sum_i p_{T,i}^2}. \end{aligned} \quad (\text{B.2})$$

On average, quark-like jets produce narrower jets and therefore they have lower values of the axes variables with respect to comparable gluon jets.

The jet fragmentation distribution The energy distribution in gluon-induced jets tends to be more symmetrically distributed than in quark-induced jets and quarks are more likely to produce jets with hard constituents that carry a significant fraction of the jet energy. Hence, quarks have a harder fragmentation function compared to gluons, that is defined as:

$$p_T D = \frac{\sqrt{\sum_i p_{T,i}^2}}{\sum_i p_{T,i}^2}, \quad (\text{B.3})$$

where the sum runs over all the jets constituents. This variable assumes values between 0 and 1, tending to 0 for jets made of only one particle that carries all of its momentum and to 1 for a jet made of an infinite number of particles.

In addition to the three variables described above, other observables have been studied, as detailed in [201], but they haven't been included in the final version of the tagger because of the negligible improvement they were bringing.

A graphical sketch summarizing the different jets properties between quark and gluon like jets is presented in Figure B.1.

B.2 The tagger

A discriminator capable of discriminating jets initiated by light-quark partons from those initiated by gluons is built. It can be used as a tool in analyses relying on the identification of a final state which includes a fixed number of hadronic jets generally originating from light-quarks. This discriminator is a handle for those scenarios with a hadronic final state suffering from the overwhelming multijet background arising from QCD processes or electroweak background with hard initial or final state gluon radiation. The quark-gluon tagger may

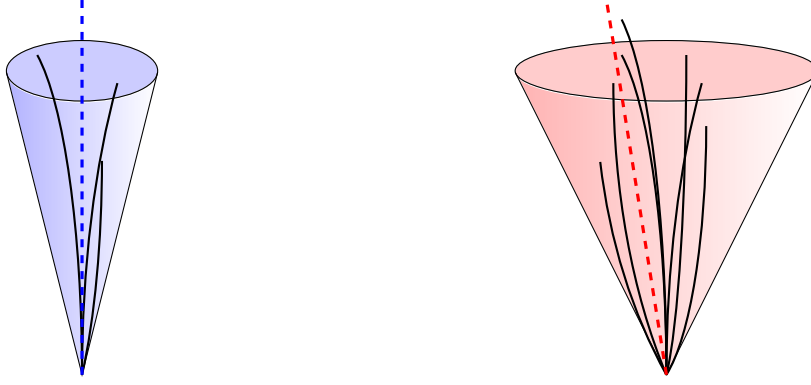


Figure B.1: Sketches of quark (left) and gluon (right) induced jets. The black lines show the jet constituent tracks, and the dashed line represents the average energy distribution.

also be used to reconstruct the mass of hadronically decaying objects, in cases of poor resolution to combinatorial backgrounds.

The discriminator is the likelihood of the product among the PDFs of the three observables considered (i.e. jet constituents multiplicity, the jet-ellipse minor axis σ_2 and the fragmentation function $p_T D$). The PDFs are built from jets in simulated QCD dijet events, which have been successfully tagged as light-quark or gluon jets. Eight exclusive pseudorapidity regions with upper boundaries $|\eta|=1.3, 1.5, 1.8, 2.1, 2.4, 2.7, 3.0, 4.7$ are defined and the PDFs are computed separately in each phase space. In addition, to take into account the dependence of the means and shapes of the variables both on the jet p_T and the amount of PU in the event, the PDFs are computed double-differentially in bins of jet p_T and ρ , where ρ is the average p_T density per unit area (in η - ϕ space) in an event and is closely correlated to the amount of PU.

Figure B.2 shows the input variables of the quark-gluon likelihood and the final discriminant for light-quark and gluon flavor simulated jets. The identification efficiency to tag quark jets as a function of the gluon jet rejection is shown in Figure B.3.

This analysis is carried on simulation and the flavor assignation to each reconstructed jet is defined by accessing the generator information at parton level: the reconstructed jet is matched to quarks and gluons by minimizing their distance ΔR in the η - ϕ plane. If a match with $\Delta R < 0.3$ is obtained, the jet flavor is defined as the one of the p_T -leading parton that is found. If no match is found, then the jet is labeled as *undefined*. It should be noted that jets initiated by b quarks present hadronization properties that are closer to gluon-initiated jets than to light-quark jets. On the other hand, b jets can be efficiently identified with the use of b -tagging algorithms. Jets are therefore required to fail the loose set of identification criteria of the b -tagger, to minimize the contribution of b quarks.

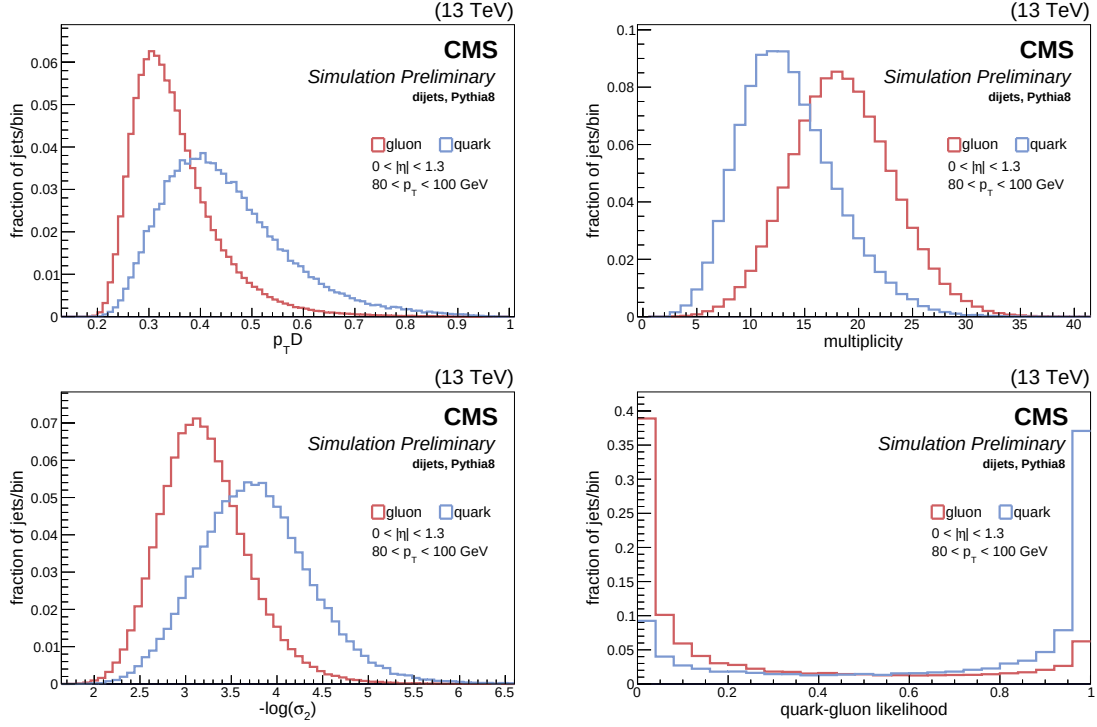


Figure B.2: Quark-gluon discrimination variables from simulation: (top left) $p_T D$, (top right) multiplicity (bottom left) σ_2 (bottom right) the quark-gluon likelihood [35].

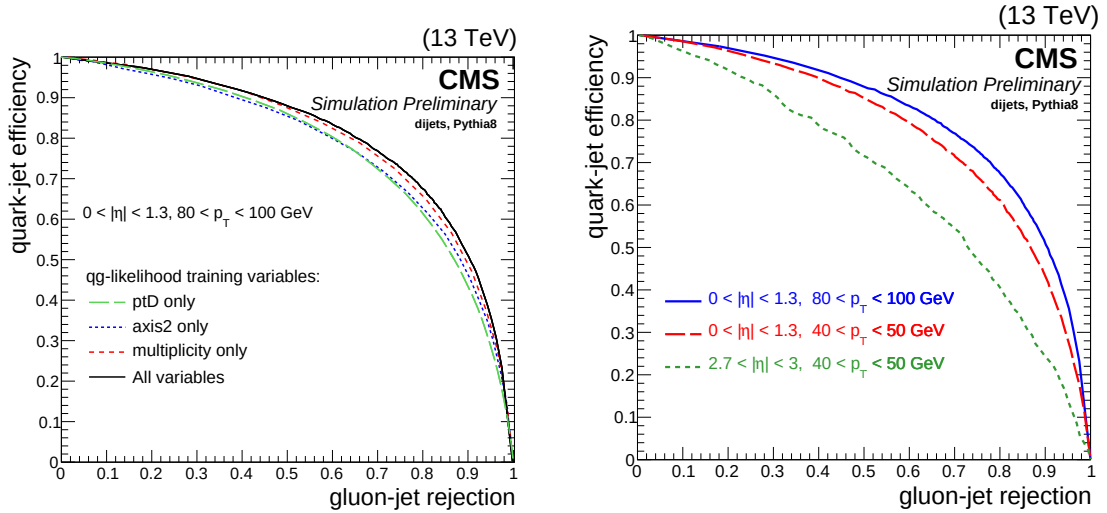


Figure B.3: Quark jet tagging efficiency as a function of the gluon jet rejection rate: (left) individual variable discrimination rate compared to the full likelihood (right) likelihood performance in different kinematic regions [35].

B.3 Validation on data

Two control regions are defined and used to validate the discriminator on data. One region, where Z+jets events are selected, is quark-enriched. The other one is a dijet region, whose events are gluon-enriched. By the simultaneous use of these two samples, the performance of the discriminator can be verified on both parton flavors, and across the whole phase space. In this Section the event selection that is applied in the analysis is detailed and the results obtained in the two regions are described as well. In what follows, all Monte Carlo distributions are normalized to the integral of the data, as the main interest lies in a comparison of the variable shapes.

B.3.1 Validation on Z+jets events

The Z+jets region provides a relatively pure sample of quark jets. More than 70% of jets with $p_T > 100$ GeV and laying in the central pseudorapidity region are expected to originate from the hadronization of a light-quark.

A dataset where two muons consistent with the Z boson decay are identified is used for this study. The sample is collected applying an offline trigger requesting at least one muon with $p_T > 20$ GeV. Besides, the following selection requirements are applied:

- at least two muons with opposite charge and $p_T > 20$ GeV;
- the invariant mass of the dimuon system should fall in the $[70,110]$ GeV interval;
- the azimuthal difference between the dimuon system and the p_T -leading jet should be > 2.1 rad, meaning than the two objects are back-to-back in the transverse plane;
- the p_T -subleading jet in the event should have a transverse momentum $p_T < 30\% p_T^{\mu\mu}$.

A comparison between data and Monte Carlo of the input variables is shown in Figure B.4, for a p_T -leading jet with $80 < p_T < 100$ GeV and $|\eta| < 2.0$. A good agreement is achieved and similar results are observed in other kinematic regions.

B.3.2 Validation on dijet events

As opposed to Z+jets event, in the dijet region the main part of the jets is likely arising from the hadronization of gluons. The elements required in the offline event selection are:

- at least two jets with $p_T > 30$ GeV;
- the two p_T -leading jets should be back-to-back in the transverse plane, their azimuthal difference being greater than 2.5;
- if a third jet is reconstructed, its p_T should be lower than the 30% of the average p_T between the two leading jets.

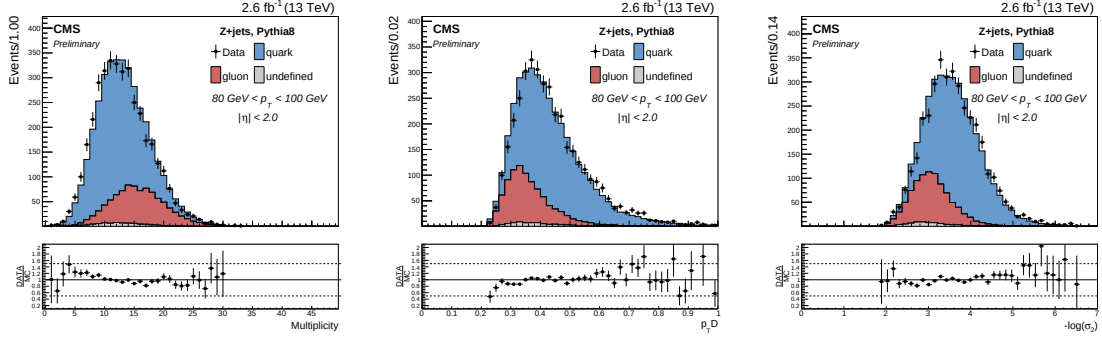


Figure B.4: Data-MC comparisons, for jets with $80 < p_T < 100$ GeV and $|\eta| < 2$ in Z+jets events, of the three input variables used in the discriminator: multiplicity (left), $p_T D$ (center) and σ_2 (right). The data (black markers) are compared to the Madgraph/Pythia simulation, on which the different components are shown: quarks (blue), gluon (red) and unmatched/pileup (grey) [35].

To minimize the effect that the jet energy resolution could have, causing jet migrations among the different kinematic regions, and that could have a large impact on low p_T jets, a dijet *tag-and-probe* approach is deployed. The method consists of the identification of the p_T interval using one of the two jets (the tag jet), while the remaining jet (the probe jet) is used to describe the η region and to fill the distributions. This is done twice per event, so that each jet is alternatively a tag jet and a probe jet.

Figure B.5 shows the comparison between data and Monte Carlo simulation of the discriminator output, for jets with $|\eta| < 2.0$ and $80 < p_T < 100$ GeV, both with Z+jets events (left) and dijet (right). A disagreement is visible among the observed data and Monte Carlo predicted distributions in the dijet region. A correction procedure, that will serve also as an assessment of systematic uncertainties, is applied.

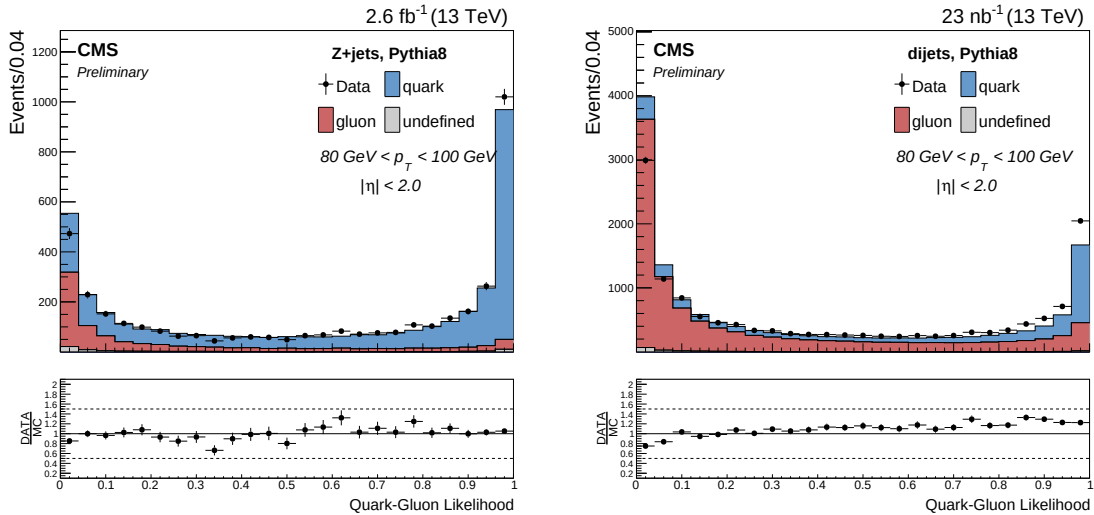


Figure B.5: Data-MC comparison for the quark-gluon discriminant in Z+jets (left) and dijet (right) events for jets in the central region with $80 < p_T < 100$ GeV. The data (black markers) are compared to the Madgraph/Pythia simulation, on which the different components are shown: quarks (blue), gluon (red) and unmatched/pileup (grey) [35].

B.4 Systematic uncertainties extraction

A shape uncertainty is estimated to fix residuals discrepancies between data and simulation and to allow the estimation of systematic uncertainties in an analysis context. The method relies on the discriminator shape variations observed in the validation of the simulated samples.

The approach that is defined is based on reshaping the Monte Carlo distributions of the quark and gluon components. Appropriate weights to the gluon and quark components are defined, as a function of the jet likelihood output, by constraining them to the yields observed in data. To extract the weights, two samples with different quark and gluon fractions are employed at the same time. Using the dijet (jj) and Z+jets (Zj) samples, we obtain the weights w_i^q and w_i^g to be applied to the quark and gluon jet MC components, respectively. Each weight is applied to reshape the corresponding likelihood output bin i by solving the following linear system in the kinematic regions defined by the p_T and η bin:

$$\begin{cases} N_{data}^i(jj) = w_i^q N_i^q(jj) + w_i^g N_i^g(jj) + N_i^{un}(jj) \\ N_{data}^i(Zj) = w_i^q N_i^q(Zj) + w_i^g N_i^g(Zj) + N_i^{un}(Zj) \end{cases}, \quad (\text{B.4})$$

where N_i^q , N_i^g and N_i^{un} , label the MC-predicted yields in the i^{th} bin respectively for the quark, gluon, and unidentified jets. The output distributions (for example Figure B.5) are used to solve the above linear system for each likelihood output bin. The resulting weights w_i^q and w_i^g are extracted simultaneously from dijet and Z+jets data. They are then fitted with polynomial functions, to obtain a smooth interpolation of the weighting coefficients. The fitted function is finally applied back as weights to the simulated distributions, and the resulting reshaped distributions are shown in Figure B.6. A better level of agreement between data and MC is now achieved.

B.5 Comparison with Herwig MC generator

Depending on the modeling of the parton shower, the discriminative power of the tagger is subject to different systematic uncertainties. To further develop the derivation of the systematic uncertainties, all the validation, comparisons and reshaping procedures have been performed for both dijet and Z+jets events using, instead of Pythia8, the Herwig++ showering model, with the CUETHS1 tune.

The performance differences between the two showering options, with and without the data-driven reshaping, are evaluated and the efficiencies to select gluon- and quark-jets with using a fixed cut on the likelihood output. The full comparisons of efficiencies are shown in Figure B.7 as a function of the jet transverse momentum. The evaluation was performed both on dijet (left) than on Z+jets (right) event samples. It can be observed that before reshaping, the data-driven corrections differences between the Herwig++ and Pythia8 simulations are sizeable with larger efficiency differences in the case of gluon jets. After the data-driven reshaping the efficiencies predicted by either Herwig++ or Pythia8 are within the percent level of disagreement, inferring a good closure of the tagger performance and of the reshaping method.

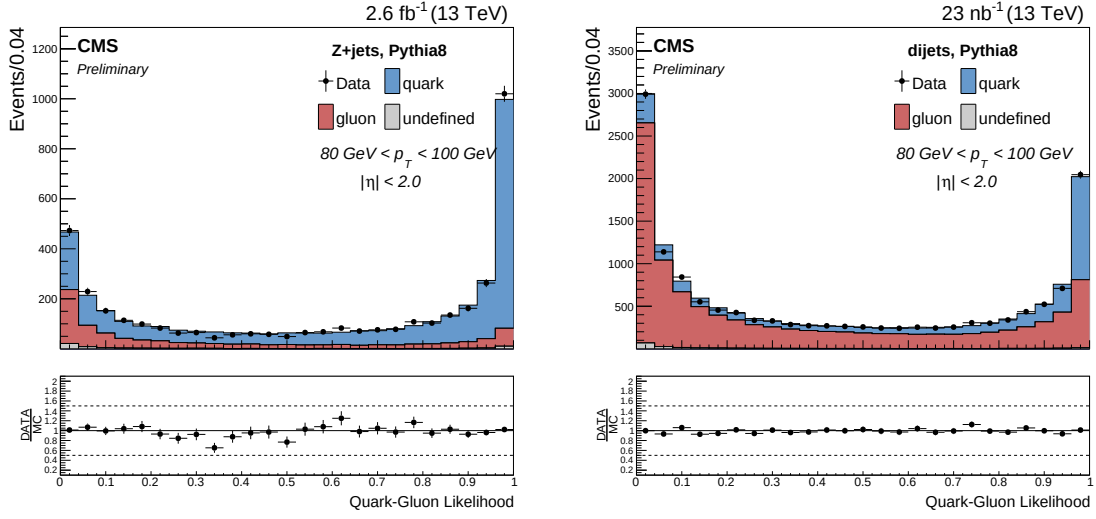


Figure B.6: Data-MC comparison for the quark-gluon discriminant in Z+jets (left) and dijet (right) events for jets in the central region with $80 < p_T < 100$ GeV, after the data-driven systematics reshaping procedure. The data (black markers) are compared to the reshaped Madgraph/Pythia simulation, on which the different components are shown: quarks (blue), gluon (red) and unmatched/pileup (grey) [35].

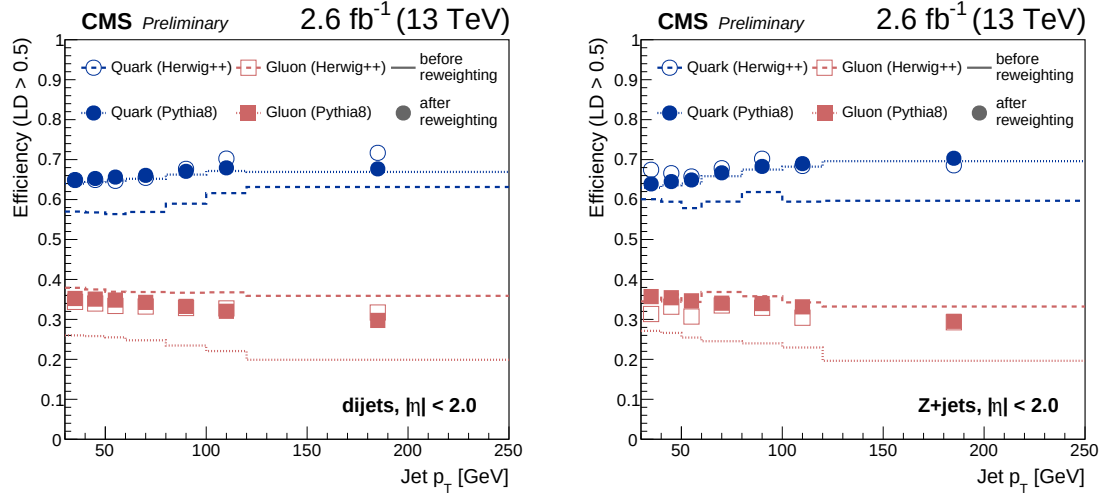


Figure B.7: Herwig++ (v2.7.0 with CUETHS1 tune) and Pythia8 (v8.205 with CUETP8M1 tune) gluon- and quark-jet selection efficiencies by applying a fixed cut on the likelihood output > 0.5 . Efficiencies are evaluated in dijet events (left) or Z+jet events (right), as a function of the jet p_T with or without the data-driven reshaping of the outputs [35].

Appendix C

Statistical approach

The searches performed in this thesis rely on specific models that describe the signal processes of interest. To assess their validity dedicated statistical methods [40, 202, 203] are used. The expected and observed upper limits to the signal cross-section as a function of model parameters are determined with the modified frequentist method CLs [204]. In Section C.1, the fundamental concepts and definitions are briefly discussed. Section C.2 summarizes the test statistics used in this thesis to quantify the level of agreement between the results obtained and the theoretical hypothesis under study.

C.1 Fundamental concepts

Given a set of n statistically independent measurements $x = (x_1, \dots, x_n)$, with each component following the same pdf $f(x; \theta)$, function of a set of nuisance parameters θ , the likelihood function can be expressed as:

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta). \quad (\text{C.1})$$

Given the large number of available measurements, a binned maximum likelihood fit is performed for both searches reported in this thesis. Measurements x are grouped in N bins: $n = (n_1, \dots, n_N)$. In the limit of n_i sufficiently large, the number of measurements n_i can be considered Poisson distributed and the likelihood of the system can be expressed as a product of N Poisson distributions with mean λ_i :

$$L(\theta) = \prod_{i=1}^N \frac{\lambda_i^{n_i}}{n_i!} e^{-\lambda_i}. \quad (\text{C.2})$$

The likelihood can be rewritten explicitating the signal and background yields contributions:

$$\lambda = \mu_s s(\theta) + b(\theta), \quad (\text{C.3})$$

where s and b are, respectively, the expected signal and background yields, and μ is the signal strength modified ($\mu = \sigma/\sigma_{th}$). Parameters θ and μ are determined by maximizing the likelihood $L(\theta)$. Systematic uncertainties are included in the fit as nuisance parameters. The complete likelihood function is defined as the product of likelihoods in each search

channel times the product of the pdf for the measurement associated with the nuisance parameters:

$$L(obs, \tilde{\theta} | \mu, \theta) = \prod_{c=1}^{N_{ch}} L_c(obs_c | \mu, \theta) \times \prod_{i=1}^{N_{\theta}^c} p_i^c(\tilde{\theta}_i^c | \theta_i^c). \quad (C.4)$$

C.2 Statistical tests

The test statistic used to discriminate signal plus background from the background-only hypothesis is a likelihood ratio q , that, according to the Neyman-Pearson lemma is the most powerful discriminator. The profile likelihood ratio used in this thesis is defined as:

$$q_{\mu} = -2 \ln \frac{\mathcal{L}(obs | \mu \cdot s + b, \hat{\theta}_{\mu})}{\mathcal{L}(obs | \hat{\mu} \cdot s + b, \hat{\theta})}, \quad (C.5)$$

where $\hat{\theta}$ and $\hat{\mu}$ are the global maxima of the Likelihood and $\hat{\theta}_{\mu}$ is the set of nuisance parameters that maximizes the likelihood for a given signal strength μ . An upper constraint $\hat{\mu} \leq \mu$ is imposed to guarantee a one-sided confidence interval.

The confidence in the signal + background hypothesis, CL_{s+b} , is defined by the probability to observe a value of the test statistics equal to or larger than the value observed in the experiment:

$$CL_{s+b} = P_{s+b}(q_{\mu} \geq q_{\mu}^{obs}).$$

The probability to observe a value of the test statistic equal to or larger than the value observed in data under the background-only hypothesis is CL_b :

$$CL_b = P_b(q_{\mu} \geq q_{\mu}^{obs}).$$

Then, the quantity CL_s is calculated as the ratio of these two probabilities:

$$CL_s = \frac{CL_{s+b}}{CL_b}.$$

Under high statistic regime the pdf of the test statistic q_{μ} , for signal + background and background-only hypotheses, are well defined by an analytical formula. Indeed, in the asymptotic regime, according to the Wilks theorem, q_{μ} is expected to have a well defined shape as described in [183, 202]. Therefore, it is possible to get the expected limit with error bands without generating pseudo-experiments. The tool used to extract the limits uses the RooStats package [205]. A signal hypothesis is said to be excluded at 95% Confidence Level, C.L., when limit on the signal strength drops below one for CLs equal to 0.05.

List of Figures

1.1	Schematic depiction of the SM of particles physics, with the three generations of matter, gauge bosons, and the Higgs boson.	17
1.2	The hierarchy problem in terms of the Higgs mass squared corrections coming from the inclusion of the VLQs.	21
1.3	Example of Feynman diagrams for the pair (left) and single (right) vector-like B quark.	23
1.4	Production cross-sections at the LHC with 13 TeV, for selected quark multiplets and for both pair (dashed black line) and single (colored lines) processes [26]. The dotted part of the lines indicates the range of masses already excluded by direct searches.	24
1.5	Branching ratios for the different decay modes as a function of VLQ mass in the case of a vector-like T quark (left) and a vector-like B quark (right) [72]. In both cases the branching ratios are provided for an SU(2) singlet and two different SU(2) doublet scenarios.	24
1.6	An image of the galaxy cluster 1E 0657-56, also known as the "bullet cluster", as taken form the Chandra X-Ray observatory [105]. This cluster was formed after the collision of two large clusters of galaxies, the most energetic event known in the universe since the Big Bang.	26
1.7	The anisotropies of the CMB as observed by the Planck satellite [109].	27
1.8	Schematic representation of the three different experimental approaches used to search for DM: direct detection investigating for nuclear recoils from DM scattering (left), indirect searches which detect DM annihilation products (center) and collider searches investigating for DM production in high energy collisions (right).	28
1.9	Schematic view of the Hidden Valley [113]. The mountain represents massive states which may connect the Standard Modell sector to light states in the valley sector.	29
1.10	Schematic view of the interplay between the SM and the dark sector through the Z' mediator.	30
1.11	Schematic of a SM jet (left) and a semi-visible jet (right). Dashed blue lines indicate invisible jet constituents, while red lines indicate visible ones.	30

1.12	Schematic of E_T^{miss} for WIMP (left), semi-visible jets (center) and SM-jets (right) events. Dashed blue lines indicate invisible DM jet constituents, while red lines indicate visible, SM.	32
2.1	Schematic draw of the LHC injection complex [115].	34
2.2	Schematic layout of the LHC. The two beams and the four main interaction points with the corresponding experiments are also shown [119].	35
2.3	The plot shows the cumulative curves for the luminosity delivered by the LHC (azure) and recorded by CMS (orange) in the proton-proton collision events held in 2016 (top left) and 2017 (top right), and 2018 (bottom) [122].	37
2.4	An exploded view of the CMS detector, showing the outer muon chambers (white) embedded in iron (red). Internally can be seen the calorimeter system (ECAL, in green, and HCAL, in orange). The tracker is located in the core of the central barrel. [123]	38
2.5	Schematic layout of the LHC, where the four largest experiments (ALICE, ATLAS, CMS and LHCb) are also shown. The coordinates system is also pictured.	39
2.6	Schematic cross-section through the CMS tracker [11]. Each line represents a detector module. Double lines indicate back-to-back modules.	40
2.7	Transverse (left) and longitudinal views (right) comparing the phase 0 and phase 1 pixel detectors. [130]	41
2.8	Layout of the CMS electromagnetic calorimeter showing the arrangement of crystal modules, supermodules and endcaps, with the preshower in front. [11]	43
2.9	The ECAL energy resolution as a function of electron energy from test beam performance [132].	43
2.10	Longitudinal view of the CMS detector showing the locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters [11].	44
2.11	The calorimeter energy resolution as a function of pion energy using the HB only or HB+HO (left), and when adding ECAL and HCAL measurements (right) [133].	45
2.12	Architecture of the CMS DAQ system (left) and of the CMS trigger system configuration (right) [134].	46
2.13	Architecture of the Level-1 Trigger system.	47
3.1	A sketch of the specific particles' interactions in a transverse slice of the CMS detector, from the beam interaction region to the muon detector [136]. Muons and the charged pions are positively charged. Electrons are negatively charged.	50
3.2	Flow chart of the PF algorithm.	51
3.3	Mean number of interactions per bunch crossing for the 2016 (top left), 2017 (top right) and 2018 (bottom) pp run at 13 TeV. The thesis is taken to be 69.2 mb, which is determined by finding the best agreement with data and is recommended for CMS analyses [122].	54

3.4	Jet matching efficiency as a function of the jet p_T , as obtained for calo-jets (open squares) and particle-flow jets (triangles) pointing to the barrel, $-1.5 < \eta < 1.5$ (left) and to the end-caps, $1.5 < \eta < 2.5$ (right), with a matching distance of 0.1 [147].	56
3.5	Schematic illustration for a jet cone, where the internal lines represent the jet constituents and where the radius R of the jet cone is highlighted. . . .	57
3.6	Jet energy resolution as function of the jet p_T , as obtained for PFjets with a radius of 0.4, pointing to the barrel, $0.0 < \eta < 0.5$ (left) and to the end-caps, $3.2 < \eta < 4.7$ (right), for different PU scenarios [150].	59
3.7	Schematic illustration for regular hadronic decays (left) and boosted topologies (right). If the decaying boson or top quarks has a relatively low p_T , its decay product are well separated and clustered into distinguishable AK4 jets, indicated in blue. On the other hand, if the transverse momentum of the mother particle is higher, the opening angle between the resulting particles is smaller and the vector boson decay products are merged into one single large cone AK8 jet, shown in pink.	60
3.8	Distribution of the pruned mass, comparing the SM background distribution, shown with the solid filled histograms, and the single VLQ B signal. . . .	61
3.9	Distribution of the soft-drop jet mass, comparing the SM background distribution, shown with the solid filled histograms, and boosted bosons signals [155].	62
3.10	Distribution of the τ_2/τ_1 N-subjettiness ratio, comparing the SM background distribution, shown with the solid filled histograms, and boosted bosons signals [155].	63
3.11	Illustration of a heavy-flavour jet with a secondary vertex (SV) from the decay of a B-hadron resulting in charged-particle tracks that are displaced with respect to the primary interaction vertex (PV), and hence with a large impact parameter (IP) value [157].	64
3.12	Misidentification probability for c and light-flavour jets versus b jet identification efficiency for various b tagging algorithms applied to jets in $t\bar{t}$ events (left). Data-MC comparison for the CSVv2 discriminators for jets in the muon-enriched multijet sample (right). The simulated contributions of each jet flavour are shown with different colours. The total number of entries in the simulation is normalized to the number of observed entries in data. The first and last bin of each histogram contain the underflow and overflow entries, respectively [157].	65
3.13	Sketch of the Z (left) and γ (right) event kinematics in the transverse plane. The vector $\vec{u}_T$ shows the vectorial sum of all particles reconstructed in the event except for the two leptons from the Z decay (left) or the photon (right). [159]	66

3.14	Resolution of the parallel and perpendicular components of the hadronic recoil $\vec{u}_T$ as a function of $\vec{q}_T$ (left), the reconstructed primary vertices (center), and the scalar $p_{T\text{sum}}$ of all PF candidates (right), in $Z \rightarrow \mu^+\mu^-$, $Z \rightarrow e^+e^-$ and γ events. In each plot, the upper panel shows the resolution in data, whereas the lower panel shows the ratio of data to simulation. The gray band shows to the systematic uncertainty [159].	67
4.1	The leading-order Feynman diagram for the production of a single vector-like B quark in association with a b quark and light-flavour quark, and its decay to a Higgs boson and a b quark.	69
4.2	Parton-level distributions for the transverse momentum of the leading b jet (first row), second b jet (second row), leading jet (third row) and the Higgs boson (fourth row). Different columns correspond to Γ_B/M_B ratios of 1, 10, 20 and 30% with $M_B = 800$ GeV [164].	71
4.3	Parton-level distributions for the pseudorapidity of the leading b jet (first row), second b jet (second row), leading jet (third row) and the Higgs boson (fourth row). Different columns correspond to Γ_B/M_B ratios of 1, 10, 20 and 30% with $M_B = 800$ GeV [164].	72
4.4	Feynman diagrams for the leading-order production mechanisms for top quark pairs at a pp collider.	73
4.5	Some representative Feynman diagram examples for the production of vector boson plus jets.	73
4.6	Distribution of the number of pileup interactions per bunch-crossing in MC simulation (blue) and data (red) considering a minimum bias cross-section of 69.2 mb. Variations to data distributions are obtained by varying by a factor $\pm 4.6\%$ the cross-section values and the shifted distributions are shown as well, labeled with dashed lines.	75
4.7	Distribution of the number of reconstructed primary vertices in data and in simulation before (left) and after (right) pile up reweighting. Solid filled histograms label SM background distributions: they are summed cumulatively and their sum is rescaled to data. Data are indicated by black dots. In the lower box, the ratio between data and total background distribution is shown.	76
4.8	Efficiency turn-on curve of the HLT triggers measured in data, both for run B-G and run H, and in simulation, as a function on the offline hadronic activity. Green color shows distribution in data, while simulation is presented with the magenta dots.	78

-
- 4.9 Distribution of AK4 jet multiplicity (top left), AK4 medium b-tagged jet multiplicity (top right) and AK8 jet multiplicity (bottom). Solid filled histograms label SM background distributions: they are summed cumulatively and their sum is rescaled to data. The hatched grey region show the statistical uncertainty on MC. Red lines show the distributions for MC signal mass points (including a multiplicative factor 2000), while data are indicated with black dots. In the lower box, the ratio between data and total background distribution is shown. Vertical bars indicates the statistical uncertainty for each bin. In each plot the value of the overflow is added to the last bin of the histogram. 79
- 4.10 Distribution of the variables used for the Higgs tagging of the AK8 jets: pruned mass (left) and medium b-tagged subjets (right). Solid filled histograms label SM background distributions: they are summed cumulatively and their sum is rescaled to data. The hatched grey region show the statistical uncertainty on MC. Red lines show the distributions for MC signal mass points (including a multiplicative factor 2000 or 500), while data are indicated with black dots. In the lower box, the ratio between data and total background distribution is shown. Vertical bars indicates the statistical uncertainty for each bin. In each plot the value of the overflow is added to the last bin of the histogram. 80
- 4.11 Distribution of AK4 forward jet transverse momentum (top left), pseudorapidity (top right) and multiplicity (bottom). Solid filled histograms label SM background distributions: they are summed cumulatively and their sum is rescaled to data. The hatched grey region show the statistical uncertainty on MC. Red lines show the distributions for MC signal mass points (including a multiplicative factor 2000 or 500), while data are indicated with black dots. In the lower box, the ratio between data and total background distribution is shown. Vertical bars indicate the statistical uncertainty for each bin. In each plot the value of the overflow is added to the last bin of the histogram. 81
- 4.12 Hadronic activity, after the full selection is applied. Solid filled histograms indicate SM background distributions: they are summed cumulatively and their sum is rescaled to data. Red lines show the distributions for MC signal mass points (including a multiplicative factor 2000 or 500), while data are indicated with black dots. In the lower box, the ratio between data and total background distribution is shown. Vertical bars indicates the statistical uncertainty for each bin. In each plot the value of the overflow is added to the last bin of the histogram. 82
- 4.13 Spatial separations at generator level between the Higgs boson and the the b-quarks from the Higgs decay (left), the one arising from the decay of the VLQ B (center) and the spectator one. Three mass hypotheses for the B quark mass are pictured with different colors and multiplied by a normalized factor of 10. 83
-

- 4.14 Kinematic distributions at generator level for the b quark arising from the decay of the VLQ B (left) and the spectator one (right). Pseudorapidity (top) and transverse momentum (bottom) distributions are illustrated. Three mass hypotheses for the B quark mass are pictured with different colors and multiplied by a normalized factor of 10. 84
- 4.15 Simulated distributions of the reconstructed VLQ B mass, after the full selection is applied. The category with exactly 0 forward jets is shown on the left column, while the right column present the category where at least one forward jet is reconstructed. Two VLQ B mass hypotheses are here described, 800 GeV in the top and 1800 GeV in bottom row, respectively. The different colors refer to the different widths scenarios assumed, from narrow width to a resonance with width equal to the 30% of its mass. 85
- 4.16 Distribution of the reconstructed VLQ B mass, after the full selection is applied. The category with exactly 0 forward jets is shown on the left column, while the right column present the category where at least one forward jet is reconstructed. Low and high mass analysis are here described, in the top and bottom row, respectively. Solid filled histograms label SM background distributions: they are summed cumulatively and their sum is rescaled to data. The hatched grey region show the statistical uncertainty on MC. Red lines show the distributions for MC signal mass points (including a multiplicative factor 500), while data are indicated with black dots. In the lower box, the ratio between data and the total background distribution is shown. Vertical bars indicates the statistical uncertainty for each bin. In each plot the value of the overflow is added to the last bin of the histogram. 86
- 4.17 Schematic view of the signal and control regions defined in the analysis workflow. The signal region is highlighted in yellow. 88
- 4.18 Distributions of the reconstructed VLQ B mass in the control region B (top), C (middle) and D (bottom), after the full selection is applied and for the low mass analysis. The category with exactly 0 forward jets is shown on the left column, while the right column presents the category where at least one forward jet is reconstructed. Solid filled histograms label SM background distributions: they are summed cumulatively and their sum is rescaled to data. The hatched grey region show the statistical uncertainty on MC. Red lines show the distributions for MC signal mass points (including a multiplicative factor 500), while data are indicated with black dots. In the lower box, the ratio between data and total background distribution is shown. Vertical bars indicates the statistical uncertainty for each bin. In each plot the value of the overflow is added to the last bin of the histogram. 90

-
- 4.19 Distributions of the reconstructed VLQ B mass in the control region B (top), C (middle) and D (bottom), after the full selection is applied and for the high mass analysis. The category with exactly 0 forward jets is shown on the left column, while the right column presents the category where at least one forward jet is reconstructed. Solid filled histograms label SM background distributions: they are summed cumulatively and their sum is rescaled to data. The hatched grey region show the statistical uncertainty on MC. Red lines show the distributions for MC signal mass points (including a multiplicative factor 500), while data are indicated with black dots. In the lower box, the ratio between data and total background distribution is shown. Vertical bars indicates the statistical uncertainty for each bin. In each plot the value of the overflow is added to the last bin of the histogram. 91
- 4.20 Shape comparison for the m_B reconstructed mass spectrum on multijet MC simulation only. Low and high mass analyses are described on the left and on the right plots respectively. The solid line shows the normalized distribution in control region C and the distribution resulting from the application of the ABCD method, while the triangular points refer to the distribution in signal region A. Statistical uncertainties are also pictured with hatched filling and vertical error bars in the two regions, respectively. 93
- 4.21 Normalization comparison for the m_B reconstructed mass spectrum on multijet MC simulation only. Low and high mass analyses are described on the left and on the right plots respectively. The solid line shows the normalized distribution in control region C and the distribution resulting from the application of the ABCD method, while the triangular points refer to the distribution in signal region A. Statistical uncertainties are also pictured with hatched filling and vertical error bars in the two regions, respectively. 94
- 4.22 Schematic view of the modified signal and control regions defined to perform a closure test of the background estimation method. The upper side of the hatched box corresponds to the nominal signal region, that is not used in this part of the study. 95
- 4.23 Shape comparison for the m_B reconstructed mass spectrum. Low and high mass analyses are described on the left and on the right plots respectively. The solid line shows the normalized distribution in control region C' and the distribution resulting from the application of the A'B'C'D' method, while the triangular points refer to the distribution in modified signal region A'. The $t\bar{t}$ contribution is here subtracted from data. Statistical uncertainties are also pictured with hatched filling and vertical error bars in the two regions, respectively. Low and high mass analyses are described on the left and on the right plots respectively. 96
-

4.24	Normalization comparison for the m_B reconstructed mass spectrum. Low and high mass analyses are described on the left and on the right plots respectively. The solid line shows the normalized distribution in control region C' and the distribution resulting from the application of the A'B'C'D' method, while the triangular points refer to the distribution in modified signal region A'. The $t\bar{t}$ contribution is here subtracted from data. Statistical uncertainties are also pictured with hatched filling and vertical error bars in the two regions, respectively.	97
4.25	Shape comparison for the m_B reconstructed mass spectrum. Low and high mass analyses are described on the left and on the right plots respectively. The solid line shows the normalized distribution in control region C and the distribution resulting from the application of the ABCD method, while the triangular points refer to the distribution in signal region A. The $t\bar{t}$ contribution is here subtracted from data. Statistical uncertainties are also pictured with hatched filling and vertical error bars in the two regions, respectively.	98
4.26	Normalization comparison for the m_B reconstructed mass spectrum. Low and high mass analyses are described on the left and on the right plots respectively. The solid line shows the normalized distribution in control region C and the distribution resulting from the application of the ABCD method, while the triangular points refer to the distribution in signal region A. The $t\bar{t}$ contribution is here subtracted from data. Statistical uncertainties are also pictured with hatched filling and vertical error bars in the two regions, respectively. Low and high mass analyses are described on the left and on the right plots respectively.	99
4.27	Systematic variation on the m_B distribution for signal Monte Carlo events, considering a mass value of 1000 GeV in the low mass analysis. The nominal distribution (black line) is plotted with its statistical uncertainty (hatched gray filling), the shift up (red) and shift down (blue) templates account for differences in both shape and normalization with respect to the nominal sample. Plots are shown for the category with exactly zero forward jets. .	103
4.28	Systematic variation on the m_B distribution for signal Monte Carlo events, considering a mass value of 1000 GeV in the low mass analysis. The nominal distribution (black line) is plotted with its statistical uncertainty (hatched gray filling), the shift up (red) and shift down (blue) templates account for differences in both shape and normalization with respect to the nominal sample. Plots are shown for the category with at least one forward jet. .	104
4.29	Systematic variation on the m_B distribution for signal Monte Carlo events, considering a mass value of 1800 GeV in the high mass analysis. The nominal distribution (black line) is plotted with its statistical uncertainty (hatched gray filling), the shift up (red) and shift down (blue) templates account for differences in both shape and normalization with respect to the nominal sample. Plots are shown for the category with exactly zero forward jets. .	105

- 4.30 Systematic variation on the m_B distribution for signal Monte Carlo events, considering a mass value of 1800 GeV in the high mass analysis. The nominal distribution (black line) is plotted with its statistical uncertainty (hatched gray filling), the shift up (red) and shift down (blue) templates account for differences in both shape and normalization with respect to the nominal sample. Plots are shown for the category with at least one forward jet. 106
- 4.31 $\tilde{\sigma}_{FW}$ contours in the Γ_B/m_B plane obtained with *Wolfram Mathematica* [164, 180]. 108
- 4.32 Branching ratios and mixing angles for different B masses and for a fixed Γ_B/M_B ratio of 1% [164]. 110
- 4.33 Branching ratios and mixing angles for different B masses and for a fixed $\kappa = 0.1$ [164]. 110
- 4.34 Branching ratios and sine of mixing angle for B singlet (top row) and part of a (T, B) doublet (bottom row) for fixed width/mass ratio (10%, 20% and 30% respectively from left to right). Notice the different minimum available mass when imposing different width/mass ratios, under which the sine of the mixing angle cannot increase further [164]. 112
- 4.35 Branching ratios B with the condition $\kappa_W = \kappa_Z = \kappa_H \equiv \kappa$ (left column) and $\kappa_W = 0, \kappa_Z = \kappa_H \equiv \kappa$ (right column) for fixed width/mass ratio (10%, 20% and 30% respectively from top to bottom) [164]. 115
- 4.36 Distribution in the reconstructed B quark mass, after applying all selections to events with no forward jets (left) and to events with at least one forward jet (right), compared to the background distributions estimated before fitting. The upper and lower plots refer to the low and high-mass m_B analyses, respectively. The expectations for signal MC events are given by the blue histogram lines. Different contributions to background are indicated by the color filled histograms. The grey-hatched error band shows total uncertainties in the background expectation. The ratios of observations to background expectations are given in the lower panels, together with the total uncertainties prior to fitting, indicated by the grey-hatched band. 116
- 4.37 Distribution in the reconstructed B quark mass, after applying all selections to events with no forward jets (left) and to events with at least one forward jet (right), compared to the background distributions estimated after fitting. The upper and lower plots refer to the low and high-mass m_B analyses, respectively. The expectations for signal MC events are given by the blue histogram lines. Different contributions to background are indicated by the color filled histograms. The grey-hatched error band shows total uncertainties in the background expectation. The ratios of observations to background expectations are given in the lower panels, together with the total uncertainties prior to fitting, indicated by the grey-hatched band. 117

4.38	The observed (dots) and expected (dashed line) median 95% CL upper limits on the product of the B quark production cross-section and branching fraction as a function of the signal mass, assuming narrow-width resonances (upper-left) and widths of 10% (upper-right), 20% (lower-left), and 30% (lower-right) of the resonance mass for the B quark. The results are shown for the combination of 0 and > 0 forward-jet categories. The continuous red curves correspond to the theoretical expectations for singlet and doublet models.	119
5.1	The leading-order Feynman diagram for the production of two semi-visible jets.	122
5.2	Dark hadrons multiplicity (both pseudoscalars and vectors) depending on the value of α_{dark} (left). In the upper axis the scale is non-linear to show the value of Λ_{dark} which corresponds to each labeled α_{dark} value. Empirical fit for the value of Λ_{dark} maximizing the dark hadrons multiplicity for several m_{dark} scenarios (right).	125
5.3	Efficiency of hadronic triggers for 2016 (top left), 2017 (top right) and 2018 (bottom), as a function of the jet p_T . The blue dashed line represents the denominator distribution, while the purple filled histogram shows the numerator distribution. The black points refers to the efficiency, computed as the ratio of the numerator and denominator distributions. The red line presents the resulting fit, with the gray band showing the uncertainty in the fit.	130
5.4	Efficiency of hadronic triggers for 2016 (top left), 2017 (top right) and 2018 (bottom), as a function of the m_T . The blue dashed line represents the denominator distribution, while the purple filled histogram shows the numerator distribution. The black points refers to the efficiency, computed as the ratio of the numerator and denominator distributions. The red line indicates the resulting fit, with the gray band showing the uncertainty in the fit.	131
5.5	Efficiency of hadronic triggers for 2016 (top left), 2017 (top right) and 2018 (bottom), measured as a function of the m_T , with the cut $\Delta\eta(j_1, j_2) < 1.5$ included. The blue dashed line represents the denominator distribution, while the purple filled histogram represents the numerator distribution. The right-side vertical axis shows the values of these distributions. The black points show the efficiency as the ratio of the numerator and denominator distributions, with the horizontal error bars indicating the variable bin widths. The red line indicates the resulting fit, with the small gray band showing the uncertainty in the fit.	132
5.6	Distributions of the figure of merit 5.5, as a function of $\Delta\eta(j_1, j_2)$ (top left), $\Delta\phi_{min}(j_{1,2}, E_T^{miss})$ (top right) and R_T (bottom).	134

5.7	Normalized transverse mass distribution on signal (left) and QCD background (right). Three different thresholds, 0.20, 0.25 and 0.30, on R_T are applied and compared and they are presented in blue, green and red, respectively.	135
5.8	Normalized expected limits ($\pm 1\sigma$) for select signal samples versus selection value in $\Delta\phi_{\min}(j_{1,2}, E_T^{\text{miss}})$ (left) and R_T (right), varying the mediator mass $m_{Z'}$. Since most points are indistinguishable from 1 (to within $\pm 1\sigma$) the optimization was performed regarding only the median expected limit.	135
5.9	Comparison of multijet events with and without the HEM issue in distributions of $\Delta\phi_{\min}(j_{1,2}, E_T^{\text{miss}})$, E_T^{miss} , $\phi(j_1)$, $\phi(j_2)$, and $\phi(E_T^{\text{miss}})$	137
5.10	Comparison of multijet events with and without the HEM issue in distributions of $\Delta\phi_{\min}(j_{1,2}, E_T^{\text{miss}})$, E_T^{miss} , $\phi(j_1)$, $\phi(j_2)$, and $\phi(E_T^{\text{miss}})$ after applying the HEM veto.	138
5.11	Input observables used in the BDT, with distributions of SM backgrounds and signal models with different r_{inv} values. Heavy object tagging variables are presented in the left column, quark-gluon related observables in the middle one and Jet energy fractions in the right column. All distributions are normalized to unit area and include the two leading AK8 jets, with a flattened p_T spectrum.	141
5.12	ROC curves for the benchmark signal model vs. multijet (left) and $t\bar{t}$ backgrounds, showing each grouping of input variables: heavy object (top), quark-gluon (middle) and energy fractions (bottom). The area under the curve is listed for each variable.	142
5.13	Diagnostic plots for the BDT training and performance. The ROC curve showing the background rejection efficiency as a function of the signal sensitivity is presented, and the area under the curve is reported as well (left). The BDT distribution is illustrated (right). The empty histograms shows the distributions for background (blue) and signal (orange) on the training sample, while the filled histograms display the BDT obtained on the test sample, for signal (green) and background (blue).	144
5.14	The BDT discriminator distribution a signal sample (red line) and on SM background (filled histograms) for the p_T -leading (left) and p_T -subleading (right) jet, at preselection level.	144
5.15	Sketch illustrating the definition of control and signal regions, along with their labels. The signal regions are highlighted in green.	145
5.16	The transverse mass distribution in the signal regions on the benchmark signal sample (red) and on SM background (filled histograms) in 2016. The categories with high R_T are presented in the top row, while at the bottom the lower R_T selection is applied. One SVJ jet is requested in the left column and both jets must fulfill the tagger requirement in the right column.	146

5.17	The transverse mass distribution in the signal regions on the benchmark signal sample (red) and on SM background (filled histograms) in 2017. The categories with high R_T are presented in the top row, while at the bottom the lower R_T selection is applied. One SVJ jet is requested in the left column and both jets must fulfill the tagger requirement in the right column. . . .	147
5.18	The transverse mass distribution in the signal regions on the benchmark signal sample (red) and on SM background (filled histograms) in 2018, before the HEM issue appeared. The categories with high R_T are presented in the top row, while at the bottom the lower R_T selection is applied. One SVJ jet is requested in the left column and both jets must fulfill the tagger requirement in the right column.	148
5.19	The transverse mass distribution in the signal regions on the benchmark signal sample (red) and on SM background (filled histograms), in 2018 after the HEM issue appeared. The categories with high R_T are presented in the top row, while at the bottom the lower R_T selection is applied. One SVJ jet is requested in the left column and both jets must fulfill the tagger requirement in the right column.	149
5.20	The fitted transverse mass distribution in the signal regions on random data. The nominal function with the optimal number of parameters as returned by the Fisher test is used and indicated with the blue line. The categories with high R_T are presented in the top row, while at the bottom the lower R_T selection is applied. One SVJ jet is requested in the left column and both jets must fulfill the tagger requirement in the right column.	151
5.21	Median expected 95% CL upper limits for 2017 on the product of the Z' cross-section and branching fraction as a function of the mediator mass on emulated data. Each lines corresponds to a BDT working point. The red line indicated the theoretical cross-section.	152
5.22	Upper mass limits obtained using emulated data (left) and error on the fit for the different signal regions (right) as a function of the BDT working point. Only 2017 backgrounds are considered.	153
5.23	Fits on emulated data using 2017 MC backgrounds on the highSVJ2 category. Each figures corresponds to a different BDT working point, ranging from 0.4 to 0.75. The χ^2 and its probability are also shown. In the lower pad, the pulls, defined as the ratio between the difference between the data yields and the value of the fit, over the error on data, are reported.	154
5.24	The fitted transverse mass distribution in the signal regions on data. The nominal function with the optimal number of parameters as returned by the Fisher test is used and indicated with the blue line. The categories with high R_T are presented in the top row, while at the bottom the lower R_T selection is applied. One SVJ jet is requested in the left column and both jets must fulfill the tagger requirement in the right column.	156

5.25	Expected 95% confidence level upper limits on the semi-visible jets production cross-section using 13 TeV data collected during the full Run 2. The 1 and 2 standard deviations (σ) bands on the expected limits are shown respectively in green and yellow. Theoretical cross-sections (Table 5.2) are indicated by the red line.	157
A.1	Geometry of the CMS barrel and forward pixel detectors as in Phase 0 and the corresponding coordinate systems [189].	165
A.2	Charge deposition and drift in a pixel barrel module, left in the plane parallel to the beam, right in the plane transverse to the beam [189].	165
A.3	Charge deposition by a particle track ($\eta = 1.83$ or $\beta = 20$). The signals in each pixel are in kiloelectrons. Signals below threshold are in green. The red cross is the geometrical center of the track and dashed red line represents the track projection [189].	166
A.4	Impact position determination in the x -axis (left) and in the y -axis (right).	166
A.5	Resolution distributions as a function of the integrated luminosity of data collected during 2015. The top row shows the results for the second layer of the barrel pixel along the r - ϕ direction (left) and z (right). The bottom row presents the results for the second disk of the forward pixel along the r -direction. In all plots a comparison between the reconstruction methods is presented, where the red dots labels the generic algorithm and the blue the template [191].	168
A.6	Residuals distributions along the r - ϕ direction (top) and z (bottom) for the second (left) and third (right) layer of the forward pixel detector obtained with template reconstruction for 2017 data [192].	169
A.7	Residuals distributions along the r - ϕ direction (top) and z (bottom) for the first layer of the barrel pixel detector. The first column show the results obtained on simulation with template reconstruction, while in the second and third column 2018 data are presented, analysed with generic and template reconstruction, respectively [193].	170
A.8	Residuals distributions along the r - ϕ direction (top) and z (bottom) for the second layer of the barrel pixel detector. The first column show the results obtained on simulation with template reconstruction, while in the second and third column data are presented, analysed with generic and template reconstruction, respectively [193].	170
A.9	Residuals distributions along the r - ϕ direction (top) and z (bottom) for the third layer of the barrel pixel detector. The first column show the results obtained on simulation with template reconstruction, while in the second and third column 2018 data are presented, analysed with generic and template reconstruction, respectively [193].	171

A.10	Residuals distributions along the r - ϕ direction (top) and z (bottom) for the fourth layer of the barrel pixel detector. The first column show the results obtained on simulation with template reconstruction, while in the second and third column 2018 data are presented, analysed with generic and template reconstruction, respectively [193].	171
A.11	Residuals distributions along the local- x direction (top) and local- y (bottom) for the second disk of the forward pixel detector. The first column show the results obtained on simulation with template reconstruction, while in the second and third column 2018 data are presented, analysed with generic and template reconstruction, respectively [193].	172
B.1	Sketches of quark (left) and gluon (right) induced jets. The black lines show the jet constituent tracks, and the dashed line represents the average energy distribution.	175
B.2	Quark-gluon discrimination variables from simulation: (top left) $p_T D$, (top right) multiplicity (bottom left) σ_2 (bottom right) the quark-gluon likelihood [35].	176
B.3	Quark jet tagging efficiency as a function of the gluon jet rejection rate: (left) individual variable discrimination rate compared to the full likelihood (right) likelihood performance in different kinematic regions [35]. . . .	176
B.4	Data-MC comparisons, for jets with $80 < p_T < 100$ GeV and $ \eta < 2$ in Z+jets events, of the three input variables used in the discriminator: multiplicity (right), $p_T D$ (center) and σ_2 (right). The data (black markers) are compared to the Madgraph/Pythia simulation, on which the different components are shown: quarks (blue), gluon (red) and unmatched/pileup (grey) [35]. . . .	178
B.5	Data-MC comparison for the quark-gluon discriminant in Z+jets (left) and dijet (right) events for jets in the central region with $80 < p_T < 100$ GeV. The data (black markers) are compared to the Madgraph/Pythia simulation, on which the different components are shown: quarks (blue), gluon (red) and unmatched/pileup (grey) [35].	178
B.6	Data-MC comparison for the quark-gluon discriminant in Z+jets (left) and dijet (right) events for jets in the central region with $80 < p_T < 100$ GeV, after the data-driven systematics reshaping procedure. The data (black markers) are compared to the reshaped Madgraph/Pythia simulation, on which the different components are shown: quarks (blue), gluon (red) and unmatched/pileup (grey) [35].	180
B.7	Herwig++ (v2.7.0 with CUETHS1 tune) and Pythia8 (v8.205 with CUETP8M1 tune) gluon- and quark-jet selection efficiencies by applying a fixed cut on the likelihood output > 0.5 . Efficiencies are evaluated in dijet events (left) or Z+jet events (right), as a function of the jet p_T with or without the data-driven reshaping of the outputs [35].	180

List of Tables

1.1	Summary of the various VLQ analysis and corresponding resulting observed mass limits to date from the ATLAS and CMS experiments [73, 74], with a focus on the latest studies performed at a center-of-mass energy of 13 TeV.	25
3.1	List of identification criteria for jets with $ \eta < 2.7$. The requirements on the charged components are only applied for jets with $ \eta < 2.4$	58
3.2	List of identification criteria for jets with $2.7 < \eta < 3.0$	58
3.3	List of identification criteria for jets with $ \eta > 3.0$	58
3.4	List of CSVv2 working points.	65
4.1	List of primary datasets of pp collision data produced at $\sqrt{s} = 13$ TeV and collected by CMS in 2016, employed in the analysis.	70
4.2	Generations details for the simulated processes used in the analysis. The cross-section is reported, and the generator along with showering packages used with the relative calculation orders are also listed.	74
4.3	HLT trigger paths used in the analysis, along with their corresponding L1 paths. The logical OR of the L1 selections is used as seed for the higher level triggers.	77
4.4	Summary of the selection requirements.	87
4.5	Signal contamination in all three control regions, for both categories and both mass regimes analyses. Two mass values of the m_B are presented, 1000 GeV and 1800 GeV. The reported values have been obtained as the ratio between signal samples yields over the multijets yields, considering only the events in three bins under the resonance peak.	88
4.6	Pre-fit events yields for all background sources (taken from Monte Carlo simulation), data and for two m_B hypotheses. Both the categories with zero forward jets and at least one forward jet are presented, and both the two mass analyses are reported as well.	89
4.7	Ratio between multijet yields between control regions A and C, and between control regions B and D, respectively. All categories depending on the forward jets multiplicity and on the H_T selection are shown.	92

4.8	Summary of systematic uncertainties. The ✓ indicates the application of the systematic on the given sample, while the sign ✗ stands for the missing inclusion of the considered source. The apex † for multijets indicates that the systematic is considered as propagated from the $t\bar{t}$ contribution, when it is subtracted from data.	102
4.9	Values of $\tilde{\sigma}_{FW}$ depending on the VLQ B mass and corresponding to several resonance width/mass ratios [164].	108
4.10	Cross-sections for $pp \rightarrow Bbq$, with the ratio Γ_B/m_B fixed to 1% (NWA). The couplings and branching fractions in simplified models are calculated using the equations in the text. The uncertainties in the production cross-sections correspond to the halving and doubling of the QCD renormalization and factorization scales.	109
4.11	cross-sections for $pp \rightarrow Bbj$ (NWA) calculated for $\kappa = 0.1$, PDF set cteq6l1, Γ_B/m_B ratio and branching fractions in simplified models calculated using equation 4.7.	110
4.12	Cross-sections for $pp \rightarrow Bbj \rightarrow Hbbj$ for different values of the width/mass ratio and for a singlet B	111
4.13	Cross-sections for $pp \rightarrow Bbj \rightarrow Hbbj$ for different values of the width/mass ratio and for a B part of a (T, B) doublet.	111
4.14	Cross-sections for $pp \rightarrow Bbj \rightarrow Hbbj$ for different values of the width/mass ratio and for a B with the condition $\kappa_W = \kappa_Z = \kappa_H \equiv \kappa$	113
4.15	Cross-sections for $pp \rightarrow Bbj \rightarrow Hbbj$ for different values of the width/mass ratio and for a B with the condition $\kappa_W = 0, \kappa_Z = \kappa_H \equiv \kappa$	114
4.16	Observed and expected post-fit event yields in the signal region in both low mass and high mass ranges. The multijet background expectation yields are derived from data, while the yields for the other sources of background are derived from simulation.	118
5.1	List of primary datasets of pp collision data used in the analysis. They are produced at $\sqrt{s} = 13$ TeV in 2016, 2017 and 2018 and collected by CMS.	123
5.2	Cross-sections as a function of $m_{Z'}$	124
5.3	Generation details for the simulation of the multijet and $t\bar{t}$ processes. The cross-section is reported, and the generator, along with the showering packages used with the relative calculation order are also listed. The star symbol * indicates the samples that are only used in 2016 and 2017.	127
5.4	Generation details for the simulation of the Z and W + jets processes. The cross-section is reported, and the generator, along with the showering packages used with the relative calculation order are also listed.	128
5.5	Summary of the hadronic HLT triggers used in 2016, 2017 and 2018.	129
5.6	Trigger efficiency corrections for each year of data taking, measured at the m_T threshold value and the upper boundary of the fit, including the statistical uncertainty.	132
5.7	Summary of preselection and selection requirements.	136

5.8	Summary of the p-values obtained from the Fisher test on the nominal function. The results are reported for the combinations of consecutive parameters and for all categories.	151
-----	--	-----

Bibliography

- [1] S. L. Glashow. Partial-symmetries of weak interactions. *Nuclear Physics*, 22(4):579–588, 1961. doi:[10.1016/0029-5582\(61\)90469-2](https://doi.org/10.1016/0029-5582(61)90469-2).
- [2] A. Salama, J.C. Ward. On a gauge theory of elementary interactions. *Il Nuovo Cimento (1955-1965)*, 19(1):165–170, 1961. doi:[10.1007/BF02812723](https://doi.org/10.1007/BF02812723).
- [3] S. Weinberg. A Model of leptons. *Physics Review Letters*, 19:1264–1266, 1967. doi:[10.1103/PhysRevLett.19.1264](https://doi.org/10.1103/PhysRevLett.19.1264).
- [4] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Physics Review Letters*, 13:321–323, 1964. doi:[10.1103/PhysRevLett.13.321](https://doi.org/10.1103/PhysRevLett.13.321).
- [5] P. W. Higgs. Broken symmetries, massless particles and gauge fields. *Physics Review Letters*, 12(2):132 – 133, 1964. doi:[10.1016/0031-9163\(64\)91136-9](https://doi.org/10.1016/0031-9163(64)91136-9).
- [6] P. W. Higgs. Broken symmetries and the masses of gauge bosons. *Physics Review Letters*, 13:508–509, 1964. doi:[10.1103/PhysRevLett.13.508](https://doi.org/10.1103/PhysRevLett.13.508).
- [7] G.S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global conservation laws and massless particles. *Phys. Rev. Lett.*, 13:585–587, Nov 1964. doi:[10.1103/PhysRevLett.13.585](https://doi.org/10.1103/PhysRevLett.13.585).
- [8] ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1): 1 – 29, 2012. doi:[10.1016/j.physletb.2012.08.020](https://doi.org/10.1016/j.physletb.2012.08.020).
- [9] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30–61, 2012. doi:[10.1016/j.physletb.2012.08.021](https://doi.org/10.1016/j.physletb.2012.08.021).
- [10] L. Evans and P. Bryant. LHC Machine. *JINST*, 3(08):S08001, 2008. URL <http://stacks.iop.org/1748-0221/3/i=08/a=S08001>.
- [11] CMS Collaboration. The CMS Experiment at the CERN LHC. *Journal of Instrumentation*, 3:S08004, 2008. doi:[10.1088/1748-0221/3/08/S08004](https://doi.org/10.1088/1748-0221/3/08/S08004).
- [12] D. H. Perkins. *Introduction to High Energy Physics; 4th ed.* Cambridge University Press, Cambridge, 2000. URL <https://cds.cern.ch/record/396126>.
- [13] N. Arkani-Hamed, A. G.Cohen and H. Georgi. Electroweak symmetry breaking from dimensional deconstruction. *Physics Letters B*, 513(1):232 – 240, 2001. doi:[10.1016/S0370-2693\(01\)00741-9](https://doi.org/10.1016/S0370-2693(01)00741-9).
- [14] M. Schmaltz and D. Tucker-Smith. Little Higgs theories. *Annual Review of Nuclear and Particle Science*, 55(1):229–270, 2005.

doi:[10.1146/annurev.nucl.55.090704.151502](https://doi.org/10.1146/annurev.nucl.55.090704.151502).

- [15] M. Perelstein, M. E. Peskin and A. Pierce. Top quarks and electroweak symmetry breaking in little Higgs models. *Phys. Rev. D*, 69:075002, 2004. doi:[10.1103/PhysRevD.69.075002](https://doi.org/10.1103/PhysRevD.69.075002).
- [16] K. Agashe, R. Contino and A. Pomarol. The minimal composite Higgs model. *Nuclear Physics B*, 719(1):165 – 187, 2005. doi:[10.1016/j.nuclphysb.2005.04.035](https://doi.org/10.1016/j.nuclphysb.2005.04.035).
- [17] D. B. Kaplan. Flavor at SSC energies: a new mechanism for dynamically generated fermion masses. *Nuclear Physics B*, 365(2):259 – 278, 1991. doi:[10.1016/S0550-3213\(05\)80021-5](https://doi.org/10.1016/S0550-3213(05)80021-5).
- [18] M. J. Dugan, H. Georgi and D. B. Kaplan. Anatomy of a composite Higgs model. *Nuclear Physics B*, 254:299 – 326, 1985. doi:[10.1016/0550-3213\(85\)90221-4](https://doi.org/10.1016/0550-3213(85)90221-4).
- [19] A. De Simone, O. Matsedonskyi, R. Rattazzi, A. Wulzer. A First Top Partner Hunter’s Guide. *Journal of High Energy Physics*, 04:004, 2013. doi:[10.1007/JHEP04\(2013\)004](https://doi.org/10.1007/JHEP04(2013)004).
- [20] R. Contino, L. Da Rold, Leandro and A. Pomarol. Light custodians in natural composite Higgs models. *Physics Review D*, 75:055014, 2007. doi:[10.1103/PhysRevD.75.055014](https://doi.org/10.1103/PhysRevD.75.055014).
- [21] Y. Hosotani, S. Noda and K. Takenaga. Dynamical gauge-higgs unification in the electroweak theory. *Physics Letters B*, 607(3):276 – 285, 2005. doi:[10.1016/j.physletb.2004.12.029](https://doi.org/10.1016/j.physletb.2004.12.029).
- [22] I. Antoniadis and K. Benakli and M. Quirs. Finite Higgs mass without supersymmetry. *New Journal of Physics*, 3(1):20, 2001. doi:[10.1088/1367-2630/3/1/320](https://doi.org/10.1088/1367-2630/3/1/320).
- [23] M. Buchkremer and G. Cacciapaglia and A. Deandrea and L. Panizzi. Model-independent framework for searches of top partners. *Nuclear Physics B*, 876(2):376 – 417, 2013. doi:[10.1016/j.nuclphysb.2013.08.010](https://doi.org/10.1016/j.nuclphysb.2013.08.010).
- [24] A. De Simone, O. Matsedonskyi, R. Rattazzi, A. Wulzer. A first top partner hunter’s guide. *Journal of High Energy Physics*, 2013(4):4, 2013. doi:[10.1007/JHEP04\(2013\)004](https://doi.org/10.1007/JHEP04(2013)004).
- [25] J. A. Aguilar Saavedra. Identifying top partners at LHC. *Journal of High Energy Physics*, 2009(11):030, 2009. doi:[10.1088/1126-6708/2009/11/030](https://doi.org/10.1088/1126-6708/2009/11/030).
- [26] J.A. Aguilar-Saavedra, R. Benbrik, S. Heinemeyer and M. Pérez-Victoria. Handbook of vectorlike quarks: mixing and single production. *Phys. Rev. D*, 88:094010, 2013. doi:[10.1103/PhysRevD.88.094010](https://doi.org/10.1103/PhysRevD.88.094010).
- [27] CMS Collaboration. Search for single production of vector-like quarks decaying to a b quark and a Higgs boson. *Journal of High Energy Physics*, 06:031, 2018. doi:[10.1007/JHEP06\(2018\)031](https://doi.org/10.1007/JHEP06(2018)031).
- [28] Planck Collaboration. Planck 2015 results. XIII. Cosmological parameters. *Astronomy and Astrophysics*, 594:A13, 2016. doi:[10.1051/0004-6361/201525830](https://doi.org/10.1051/0004-6361/201525830).
- [29] V. C. Rubin, and W. K. Ford. Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions. *Astrophysical Journal*, 159:379, 1970. doi:[10.1086/150317](https://doi.org/10.1086/150317).
- [30] D. Clowe, M. Bradac, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones and D. Zaritsky. A direct empirical proof of the existence of Dark Matter. *Astrophysical*

- Journal*, 648:109–113, 2006. doi:[10.1086/508162](https://doi.org/10.1086/508162).
- [31] V. A. Mitsou. Shedding light on Dark Matter at colliders. *International Journal of Modern Physics A*, 28(31):1330052, 2013. doi:[10.1142/S0217751X13300524](https://doi.org/10.1142/S0217751X13300524).
 - [32] M. J. Strassler and K. M. Zurek. Echoes of a hidden valley at hadron colliders. *Physics Letters B*, 651(5):374 – 379, 2007. ISSN 0370-2693. doi:[10.1016/j.physletb.2007.06.055](https://doi.org/10.1016/j.physletb.2007.06.055). URL <http://www.sciencedirect.com/science/article/pii/S0370269307007721>.
 - [33] M. J. Strassler and K. M. Zurek. Discovering the higgs through highly-displaced vertices. *Physics Letters B*, 661(4):263 – 267, 2008. ISSN 0370-2693. doi:[10.1016/j.physletb.2008.02.008](https://doi.org/10.1016/j.physletb.2008.02.008).
 - [34] K. M. Zurek T. Han, Z. Si and M. J. Strassler. Phenomenology of hidden valleys at hadron colliders. *Journal of High Energy Physics*, 2008(07):008–008, jul 2008. doi:[10.1088/1126-6708/2008/07/008](https://doi.org/10.1088/1126-6708/2008/07/008).
 - [35] Jet algorithms performance in 13 TeV data. Technical report, CERN, Geneva, 2017. URL <http://cds.cern.ch/record/2256875>.
 - [36] B.R. Martin and G. Shaw. *Particle Physics*. Manchester Physics Series. Wiley, 2008. ISBN 9780470721537. URL <https://books.google.ch/books?id=whIbrWJdEJQC>.
 - [37] F. Halzen and A. Martin. *Quarks & Leptons: An introductory course in modern particle physics*. John Wiley & Sons, New York, USA, 1984.
 - [38] K.A. Olive and Particle Data Group. Review of Particle Physics. *Chinese Physics C*, 38(9):090001, 2014. doi:[10.1088/1674-1137/38/9/090001](https://doi.org/10.1088/1674-1137/38/9/090001).
 - [39] ALEPH, CDF, D0, DELPHI, L3, OPAL, SLD, LEP Collaborations. Precision electroweak measurements and constraints on the standard model. 2009. URL <https://cds.cern.ch/record/1313716>.
 - [40] Particle Data Group. Review of Particle Physics. *Physics Review D*, 98:030001, 2018. doi:[10.1103/PhysRevD.98.030001](https://doi.org/10.1103/PhysRevD.98.030001).
 - [41] M. E. Peskin and D. V. Schroeder. *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995. ISBN 9780201503975, 0201503972. URL <http://www.slac.stanford.edu/~mpeskin/QFT.html>.
 - [42] M. Gell-Mann. The interpretation of the new particles as displaced charge multiplets. *Il Nuovo Cimento (1955-1965)*, 4(2):848–866, Apr 1956. ISSN 1827-6121. doi:[10.1007/BF02748000](https://doi.org/10.1007/BF02748000).
 - [43] S. L. Glashow. The renormalizability of vector meson interactions. *Nuclear Physics*, 10:107 – 117, 1959. ISSN 0029-5582. doi:[10.1016/0029-5582\(59\)90196-8](https://doi.org/10.1016/0029-5582(59)90196-8).
 - [44] A. Salam and J. C. Ward. Weak and electromagnetic interactions. *Il Nuovo Cimento (1955-1965)*, 11(4):568–577, Feb 1959. ISSN 1827-6121. doi:[10.1007/BF02726525](https://doi.org/10.1007/BF02726525).
 - [45] C. N. Yang and R. L. Mills. Conservation of isotopic spin and isotopic gauge invariance. *Phys. Rev.*, 96:191–195, Oct 1954. doi:[10.1103/PhysRev.96.191](https://doi.org/10.1103/PhysRev.96.191).
 - [46] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson. Experimental test of parity conservation in beta decay. *Phys. Rev.*, 105:1413–1415, Feb 1957. doi:[10.1103/PhysRev.105.1413](https://doi.org/10.1103/PhysRev.105.1413).

- [47] N. Cabibbo. Unitary symmetry and leptonic decays. *Physics Review Letters*, 10: 531–533, Jun 1963. doi:[10.1103/PhysRevLett.10.531](https://doi.org/10.1103/PhysRevLett.10.531).
- [48] M. Kobayashi and T. Maskawa. CP-Violation in the renormalizable theory of weak interaction. *Progress of Theoretical Physics*, 49(2):652–657, 02 1973. ISSN 0033-068X. doi:[10.1143/PTP.49.652](https://doi.org/10.1143/PTP.49.652).
- [49] J. Ellis. Higgs Physics. In *Proceedings, 2013 European School of High-Energy Physics (ESHEP 2013): Paradfurdo, Hungary, June 5-18, 2013*, pages 117–168, 2015. doi:[10.5170/CERN-2015-004.117](https://doi.org/10.5170/CERN-2015-004.117).
- [50] UA1 Collaboration. Experimental observation of lepton pairs of invariant mass around 95 GeV/ c^2 at the CERN SPS Collider. *Physics Letter B*, 126:398–410, 1983. doi:[10.1016/0370-2693\(83\)90188-0](https://doi.org/10.1016/0370-2693(83)90188-0). [,7.55(1983)].
- [51] UA2 Collaboration. Observation of single isolated electrons of high transverse momentum in Events with missing transverse energy at the CERN anti-p p Collider. *Physics Letters B*, 122:476–485, 1983. doi:[10.1016/0370-2693\(83\)91605-2](https://doi.org/10.1016/0370-2693(83)91605-2).
- [52] UA1 Collaboration. Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s} = 540$ GeV. *Physics Letters B*, 122: 103–116, 1983. doi:[10.1016/0370-2693\(83\)91177-2](https://doi.org/10.1016/0370-2693(83)91177-2).
- [53] UA2 Collaboration. Evidence for $Z \rightarrow e^+e^-$ at the CERN anti-p p collider. *Physics Letters B*, 129:130–140, 1983. doi:[10.1016/0370-2693\(83\)90744-X](https://doi.org/10.1016/0370-2693(83)90744-X).
- [54] I. J. R. Aitchison and A. J. G. Hey. *Gauge theories in particle physics: A practical introduction. Vol. 1: From relativistic quantum mechanics to QED*. CRC Press, Bristol, UK, 2012. ISBN 9781466512993.
- [55] E. K. Ellis, W. J. and Stirling, and B. R. Webber. *QCD and collider physics*. Cambridge monographs on particle physics, nuclear physics, and cosmology. Cambridge University Press, Cambridge, 2003. doi:[10.1017/CBO9780511628788](https://doi.org/10.1017/CBO9780511628788).
- [56] Super-Kamiokande Collaboration. Evidence for oscillation of atmospheric neutrinos. *Physics Review Letters*, 81:1562–1567, 1998. doi:[10.1103/PhysRevLett.81.1562](https://doi.org/10.1103/PhysRevLett.81.1562).
- [57] SNO Collaboration. Measurement of the Rate of $\nu_e + d \rightarrow p + p + e^-$ Interactions Produced by 8B Solar Neutrinos at the Sudbury Neutrino Observatory. *Physics Review Letters*, 87:071301, 2001. doi:[10.1103/PhysRevLett.87.071301](https://doi.org/10.1103/PhysRevLett.87.071301).
- [58] SNO Collaboration. Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory. *Physics Review Letters*, 89:011301, 2002. doi:[10.1103/PhysRevLett.89.011301](https://doi.org/10.1103/PhysRevLett.89.011301).
- [59] J. R. Ellis. Limits of the standard model. In *PSI Zuoz Summer School on Exploring the Limits of the Standard Model Zuoz, Engadin, Switzerland, August 18-24, 2002*, 2002. URL <http://weblib.cern.ch/abstract?CERN-TH-2002-320>.
- [60] W. de Boer. Grand unified theories and supersymmetry in particle physics and cosmology. *Prog. Part. Nucl. Phys.*, 33:201–302, 1994. doi:[10.1016/0146-6410\(94\)90045-0](https://doi.org/10.1016/0146-6410(94)90045-0).
- [61] F. Zwicky. Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica Acta*, 6:110–127, 1933.
- [62] N. Arkani-Hamed, A. G. Cohen, E. Katz, and A. E. Nelson. The Littlest Higgs.

- Journal of High Energy Physics*, 07:034, 2002. doi:[10.1088/1126-6708/2002/07/034](https://doi.org/10.1088/1126-6708/2002/07/034).
- [63] T. Han, H. E. Logan, B. McElrath, L. Wang. Phenomenology of the little Higgs model. *Physics Review D*, 67:095004, 2003. doi:[10.1103/PhysRevD.67.095004](https://doi.org/10.1103/PhysRevD.67.095004).
 - [64] J. Y. Lee. A Vector-like heavy quark in the Littlest Higgs model. *Journal of High Energy Physics*, 12:065, 2004. doi:[10.1088/1126-6708/2004/12/065](https://doi.org/10.1088/1126-6708/2004/12/065).
 - [65] D. B. Kaplan, H. Georgi and S. Dimopoulos. Composite Higgs scalars. *Physics Letters B*, 136(3):187 – 190, 1984. ISSN 0370-2693. doi:[10.1016/0370-2693\(84\)91178-X](https://doi.org/10.1016/0370-2693(84)91178-X).
 - [66] T. Gherghetta and A. Pomarol. Bulk fields and supersymmetry in a slice of AdS. *Nuclear Physics B*, 586:141–162, 2000. doi:[10.1016/S0550-3213\(00\)00392-8](https://doi.org/10.1016/S0550-3213(00)00392-8).
 - [67] G. D. Kribs, T. Plehn, M. Spannowsky, and T. M. P. Tait. Four generations and Higgs physics. *Physics Review D*, 76:075016, 2007. doi:[10.1103/PhysRevD.76.075016](https://doi.org/10.1103/PhysRevD.76.075016).
 - [68] O. Eberhardt et al. Joint analysis of Higgs decays and electroweak precision observables in the Standard Model with a sequential fourth generation. *Physics Review D*, 86:013011, 2012. doi:[10.1103/PhysRevD.86.013011](https://doi.org/10.1103/PhysRevD.86.013011).
 - [69] U. J. Saldaña-Salazar. A principle for the yukawa couplings. *Journal of Physics: Conference Series*, 761:012064, oct 2016. doi:[10.1088/1742-6596/761/1/012064](https://doi.org/10.1088/1742-6596/761/1/012064).
 - [70] S. Schael et al. Precision electroweak measurements on the Z resonance. *Phys. Rept.*, 427:257–454, 2006. doi:[10.1016/j.physrep.2005.12.006](https://doi.org/10.1016/j.physrep.2005.12.006).
 - [71] M. E. Peskin and T. Takeuchi. New constraint on a strongly interacting Higgs sector. *Physics Review Letters*, 65:964–967, 1990. doi:[10.1103/PhysRevLett.65.964](https://doi.org/10.1103/PhysRevLett.65.964).
 - [72] The ATLAS Collaboration. Search for production of vector-like quark pairs and of four top quarks in the lepton-plus-jets final state in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Journal of High Energy Physics*, 2015(8):105, 2015. ISSN 1029-8479. doi:[10.1007/JHEP08\(2015\)105](https://doi.org/10.1007/JHEP08(2015)105).
 - [73] ATLAS Collaboration. Public results: exotic physics searches. <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ExoticsPublicResults>, 2019. [Online; accessed 30-Sept-2018].
 - [74] CMS Collaboration. Beyond 2 generations publications. <http://chandra.harvard.edu/photo/2006/1e0657/index.html>, 2019. [Online; accessed 10-Oct-2018].
 - [75] D0 Collaboration. Search for single vector-like quarks in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. *Physics Review Letters*, 106:081801, 2011. doi:[10.1103/PhysRevLett.106.081801](https://doi.org/10.1103/PhysRevLett.106.081801).
 - [76] D0 Collaboration. Search for a fourth generation T' quark in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. *Physics Review Letters*, 107:082001, 2011. doi:[10.1103/PhysRevLett.107.082001](https://doi.org/10.1103/PhysRevLett.107.082001).
 - [77] CDF Collaboration. Search for new particles leading to $Z + \text{jets}$ final states in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. *Physics Review D*, 76:072006, 2007. doi:[10.1103/PhysRevD.76.072006](https://doi.org/10.1103/PhysRevD.76.072006).
 - [78] CDF Collaboration. Search for new bottom-like quark pair decays $Q\bar{Q} \rightarrow (tW^\mp)(\bar{t}W^\pm)$ in same-charge dilepton events. *Physics Review Letters*, 104:091801, 2010. doi:[10.1103/PhysRevLett.104.091801](https://doi.org/10.1103/PhysRevLett.104.091801).

- [79] CDF Collaboration. Search for new T' particles in final states with large jet multiplicities and missing transverse energy in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. *Physics Review Letters*, 107:191803, 2011. doi:[10.1103/PhysRevLett.107.191803](https://doi.org/10.1103/PhysRevLett.107.191803).
- [80] CDF Collaboration. Search for production of heavy particles decaying to top quarks and invisible particles in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. *Physics Review Letters*, 106:191801, 2011. doi:[10.1103/PhysRevLett.106.191801](https://doi.org/10.1103/PhysRevLett.106.191801).
- [81] CDF Collaboration. Search for heavy bottom-like quarks decaying to an electron or muon and jets in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. *Physics Review Letters*, 106:141803, 2011. doi:[10.1103/PhysRevLett.106.141803](https://doi.org/10.1103/PhysRevLett.106.141803).
- [82] CDF Collaboration. Search for a heavy top-like quark in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. *Physics Review Letters*, 107:261801, 2011. doi:[10.1103/PhysRevLett.107.261801](https://doi.org/10.1103/PhysRevLett.107.261801).
- [83] ATLAS Collaboration. Search for single production of a vector-like B quark decaying into a bottom quark and a Higgs boson which decays into a pair of photons. Technical Report ATLAS-CONF-2018-024, CERN, Geneva, 2018. URL <http://cds.cern.ch/record/2628759>.
- [84] ATLAS Collaboration. Search for pair production of heavy vector-like quarks decaying into hadronic final states in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Physics Review D*, 98(arXiv:1808.01771):092005. 36 p, Aug 2018. doi:[10.1103/PhysRevD.98.092005](https://doi.org/10.1103/PhysRevD.98.092005).
- [85] ATLAS Collaboration. Search for pair- and single-production of vector-like quarks in final states with at least one Z boson decaying into a pair of electrons or muons in pp collision data collected with the ATLAS detector at $\sqrt{s} = 13$ TeV. *Physics Review D*, 98(11):112010, 2018. doi:[10.1103/PhysRevD.98.112010](https://doi.org/10.1103/PhysRevD.98.112010).
- [86] ATLAS Collaboration. Search for pair production of heavy vector-like quarks decaying into high- p_T W bosons and top quarks in the lepton-plus-jets final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Journal of High Energy Physics*, 08:048, 2018. doi:[10.1007/JHEP08\(2018\)048](https://doi.org/10.1007/JHEP08(2018)048).
- [87] ATLAS Collaboration. Search for pair production of up-type vector-like quarks and for four-top-quark events in final states with multiple b -jets with the ATLAS detector. *Journal of High Energy Physics*, 07:089, 2018. doi:[10.1007/JHEP07\(2018\)089](https://doi.org/10.1007/JHEP07(2018)089).
- [88] M. Aaboud et al. Search for pair production of heavy vector-like quarks decaying to high- p_T W bosons and b quarks in the lepton-plus-jets final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Journal of High Energy Physics*, 10:141, 2017. doi:[10.1007/JHEP10\(2017\)141](https://doi.org/10.1007/JHEP10(2017)141).
- [89] ATLAS Collaboration. Search for pair production of vector-like top quarks in events with one lepton, jets, and missing transverse momentum in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector. *Journal of High Energy Physics*, 08:052, 2017. doi:[10.1007/JHEP08\(2017\)052](https://doi.org/10.1007/JHEP08(2017)052).
- [90] Search for single production of vector-like quarks decaying into Wb in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. Technical Report ATLAS-CONF-2016-072,

- CERN, Geneva, 2016. URL <http://cds.cern.ch/record/2206226>.
- [91] CMS Collaboration. Search for single production of vector-like quarks decaying to a top quark and a W boson in proton–proton collisions at $\sqrt{s} = 13$ TeV. *The European Physical Journal C*, 79(2):90, 2019. ISSN 1434-6052. doi:[10.1140/epjc/s10052-019-6556-3](https://doi.org/10.1140/epjc/s10052-019-6556-3).
 - [92] CMS Collaboration. Search for vector-like T and B quark pairs in final states with leptons at $\sqrt{s} = 13$ TeV. *Journal of High Energy Physics*, 08:177, 2018. doi:[10.1007/JHEP08\(2018\)177](https://doi.org/10.1007/JHEP08(2018)177).
 - [93] CMS Collaboration. Search for pair production of vector-like quarks in the $bW\bar{b}W$ channel from proton-proton collisions at $\sqrt{s} = 13$ TeV. *Physics Letters B*, 779:82–106, 2018. doi:[10.1016/j.physletb.2018.01.077](https://doi.org/10.1016/j.physletb.2018.01.077).
 - [94] CMS Collaboration. Search for single production of a vector-like t quark decaying to a z boson and a top quark in proton–proton collisions at $\sqrt{s} = 13$ tev. *Physics Letters B*, 781:574 – 600, 2018. doi:[10.1016/j.physletb.2018.04.036](https://doi.org/10.1016/j.physletb.2018.04.036).
 - [95] CMS Collaboration. Search for pair production of vector-like T and B quarks in single-lepton final states using boosted jet substructure in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Journal of High Energy Physics*, 11:085, 2017. doi:[10.1007/JHEP11\(2017\)085](https://doi.org/10.1007/JHEP11(2017)085).
 - [96] CMS Collaboration. Search for top quark partners with charge 5/3 in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Journal of High Energy Physics*, 08:073, 2017. doi:[10.1007/JHEP08\(2017\)073](https://doi.org/10.1007/JHEP08(2017)073).
 - [97] CMS Collaboration. Search for single production of vector-like quarks decaying into a b quark and a W boson in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Physics Letters B*, 772:634–656, 2017. doi:[10.1016/j.physletb.2017.07.022](https://doi.org/10.1016/j.physletb.2017.07.022).
 - [98] CMS Collaboration. Search for single production of vector-like quarks decaying to a Z boson and a top or a bottom quark in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Journal of High Energy Physics*, 05:029, 2017. doi:[10.1007/JHEP05\(2017\)029](https://doi.org/10.1007/JHEP05(2017)029).
 - [99] CMS Collaboration. Search for electroweak production of a vector-like quark decaying to a top quark and a Higgs boson using boosted topologies in fully hadronic final states. *Journal of High Energy Physics*, 04:136, 2017. doi:[10.1007/JHEP04\(2017\)136](https://doi.org/10.1007/JHEP04(2017)136).
 - [100] CMS Collaboration. Search for single production of a heavy vector-like T quark decaying to a Higgs boson and a top quark with a lepton and jets in the final state. *Physics Letters B*, 771:80–105, 2017. doi:[10.1016/j.physletb.2017.05.019](https://doi.org/10.1016/j.physletb.2017.05.019).
 - [101] F. Zwicky. Die rotverschiebung von extragalaktischen nebeln. *Helvetica Physica Acta*, 6:110–127, 1933. doi:[10.5169/seals-110267](https://doi.org/10.5169/seals-110267). URL <http://dx.doi.org/10.5169/seals-110267>.
 - [102] V. C. Rubin, and W. K. Jr. Ford. Rotation of the Andromeda nebula from a spectroscopic survey of emission regions. *Astrophysics Journal*, 159:379–403, 1970. doi:[10.1086/150317](https://doi.org/10.1086/150317).
 - [103] K. Freese. Review of Observational Evidence for Dark Matter in the Universe and in upcoming searches for Dark Stars. *EAS Publ. Ser.*, 36:113–126, 2009. doi:[10.1051/eas/0936016](https://doi.org/10.1051/eas/0936016).

- [104] SDSS Collaboration. The Fourth Data Release of the Sloan Digital Sky Survey. *Astrophysics Journal*, 162:38–48, 2006. doi:[10.1086/497917](https://doi.org/10.1086/497917).
- [105] NASA Collaboration. 1e 0657-56: Nasa finds direct proof of dark matter. <http://cms-results.web.cern.ch/cms-results/public-results/publications/B2G/index.html>, 2018. [Online; accessed 30-Sept-2018].
- [106] W. Hu, N. Sugiyama and J. Silk. The Physics of microwave background anisotropies. *Nature*, 386:37–43, 1997. doi:[10.1038/386037a0](https://doi.org/10.1038/386037a0).
- [107] W. Hu and S. Dodelson. Cosmic microwave background anisotropies. *Annual Review of Astronomy and Astrophysics*, 40:171–216, 2002. doi:[10.1146/annurev.astro.40.060401.093926](https://doi.org/10.1146/annurev.astro.40.060401.093926).
- [108] P. J. E. Peebles and B. Ratra. The Cosmological constant and dark energy. *Rev. Mod. Phys.*, 75:559–606, 2003. doi:[10.1103/RevModPhys.75.559](https://doi.org/10.1103/RevModPhys.75.559). [,592(2002)].
- [109] ESA Collaboration. Cosmic microwave background seen by Planck. <http://sci.esa.int/planck/51553-cosmic-microwave-background-seen-by-planck/>, 2018. [Online; accessed 10-Oct-2018].
- [110] C. Alcock et al. The MACHO project: microlensing detection efficiency. *Astrophys. J. Suppl.*, 136:439–462, 2001. doi:[10.1086/322529](https://doi.org/10.1086/322529).
- [111] T. Cohen, and M. Lisanti, and H. K. Lou, and S. Mishra-Sharma. Lhc searches for dark sector showers. *Journal of High Energy Physics*, 2017(11):196, Nov 2017. ISSN 1029-8479. doi:[10.1007/JHEP11\(2017\)196](https://doi.org/10.1007/JHEP11(2017)196).
- [112] T. Cohen, M. Lisanti, H. K. Lou. Semivisible jets: Dark Matter undercover at the LHC. *Physics Review Letters*, 115(17):171804, 2015. doi:[10.1103/PhysRevLett.115.171804](https://doi.org/10.1103/PhysRevLett.115.171804).
- [113] Jose E. Juknevich. *Phenomenology of pure-gauge hidden valleys at Hadron colliders*. PhD thesis, Rutgers U., Piscataway, 2010. URL <https://rucore.libraries.rutgers.edu/rutgers-lib/30125/>.
- [114] T. Cohen, M. Lisanti, H.K Lou abd S. Mishra-Sharma. LHC Searches for dark sector showers. *Journal of High Energy Physics*, 11:196, 2017. doi:[10.1007/JHEP11\(2017\)196](https://doi.org/10.1007/JHEP11(2017)196).
- [115] CERN Collaboration. CERN accelerator complex, . URL <http://te-epc-lpc.web.cern.ch/te-epc-lpc/machines/pagesources/Cern-Accelerator-Complex.jpg>.
- [116] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *Journal of Instrumentation*, 3:S08003, 2008. doi:[10.1088/1748-0221/3/08/S08003](https://doi.org/10.1088/1748-0221/3/08/S08003).
- [117] ALICE Collaboration. The ALICE experiment at the CERN LHC. *Journal of Instrumentation*, 3:S08002, 2008. doi:[10.1088/1748-0221/3/08/S08002](https://doi.org/10.1088/1748-0221/3/08/S08002).
- [118] LHCb Collaboration. The LHCb Detector at the LHC. *Journal of Instrumentation*, 3:S08005, 2008. doi:[10.1088/1748-0221/3/08/S08005](https://doi.org/10.1088/1748-0221/3/08/S08005).
- [119] CERN Collaboration. LHC Images, . URL <https://lhc-machine-outreach.web.cern.ch/lhc-machine-outreach/images/lhc-schematic.jpg>.
- [120] W. Herr and B. Muratori. Concept of luminosity. 2006. URL <http://cds.cern.ch/record/941318>.

- [121] CMS Collaboration. Public CMS data quality information, . URL <https://twiki.cern.ch/twiki/bin/view/CMSPublic/DataQuality>.
- [122] CMS Collaboration. Public CMS Luminosity Information, . URL <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- [123] T. Dorigo. Recent results of the CMS experiment. *EPJ Web of Conferences*, 70:00021, 2014. doi:[10.1051/epjconf/20147000021](https://doi.org/10.1051/epjconf/20147000021).
- [124] CMS Collaboration. *The CMS magnet project: Technical Design Report*. Technical Design Report. CERN, 1997. URL <https://cds.cern.ch/record/331056>.
- [125] *Basic Course on General Accelerator Physics*, 2005. CERN. URL <https://cds.cern.ch/record/425460>.
- [126] CMS Collaboration. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *Journal of Instrumentation*, 9(10):P10009, 2014. URL <http://stacks.iop.org/1748-0221/9/i=10/a=P10009>.
- [127] CMS Collaboration. *The CMS tracker system project: Technical Design Report*. Technical Design Report. CERN, 1997. URL <https://cds.cern.ch/record/368412>.
- [128] CMS Collaboration. *The CMS tracker: addendum to the Technical Design Report*. Technical Design Report. CERN, 2000. URL <https://cds.cern.ch/record/490194>.
- [129] CMS Collaboration. *CMS Physics: Technical Design Report Volume 1: detector performance and software*. Technical Design Report. CERN, 2006. URL <https://cds.cern.ch/record/922757>.
- [130] CMS Collaboration. CMS Technical Design Report for the Pixel Detector Upgrade. 2012. doi:[10.2172/1151650](https://doi.org/10.2172/1151650).
- [131] CMS Collaboration. *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical Design Report. CERN, 1997. URL <https://cds.cern.ch/record/349375>.
- [132] Q. Ingram. Energy resolution of the barrel of the CMS Electromagnetic Calorimeter. *Journal of Instrumentation*, 2(04):P04004–P04004, 2007. doi:[10.1088/1748-0221/2/04/p04004](https://doi.org/10.1088/1748-0221/2/04/p04004).
- [133] S. Sharma. Understanding the performance of CMS calorimeter. *Pramana*, 69(6):1069–1074, D 2007. ISSN 0973-7111. doi:[10.1007/s12043-007-0229-8](https://doi.org/10.1007/s12043-007-0229-8).
- [134] V. Halyo, A. Hunt, P. Jindal, P. LeGresley, and P. Lujan. GPU Enhancement of the Trigger to Extend Physics Reach at the LHC. *JINST*, 8:P10005, 2013. doi:[10.1088/1748-0221/8/10/P10005](https://doi.org/10.1088/1748-0221/8/10/P10005).
- [135] CERN Collaboration. LHC computing Grid. Technical design report. Technical report, CERN, 2005. URL <https://cds.cern.ch/record/840543>.
- [136] CMS Collaboration. Particle-Flow reconstruction and global event description with the CMS detector. *Journal of Instrumentation*, 12(10):P10003, 2017. doi:[10.1088/1748-0221/12/10/P10003](https://doi.org/10.1088/1748-0221/12/10/P10003).
- [137] P. Billoir. Progressive track recognition with a Kalman like fitting procedure. *Computer Physics Communucations*, 57:390–394, 1989. doi:[10.1016/0010-4655\(89\)90249-X](https://doi.org/10.1016/0010-4655(89)90249-X).

- [138] CMS Collaboration. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *Journal of Instrumentation*, 9(10):P10009, 2014. doi:[10.1088/1748-0221/9/10/P10009](https://doi.org/10.1088/1748-0221/9/10/P10009).
- [139] K. Rose. Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. *Proceedings of the IEEE*, 86(11):2210–2239, Nov 1998. ISSN 0018-9219. doi:[10.1109/5.726788](https://doi.org/10.1109/5.726788).
- [140] R. Fruhwirth, W. Waltenberger, and P. Vanlaer. Adaptive vertex fitting. *Journal of Physics G*, 34:N343, 2007. doi:[10.1088/0954-3899/34/12/N01](https://doi.org/10.1088/0954-3899/34/12/N01).
- [141] CMS Collaboration. Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV. *Journal of Instrumentation*, 7(10):P10002, 2012. URL <http://stacks.iop.org/1748-0221/7/i=10/a=P10002>.
- [142] CMS Collaboration. Particle-Flow commissioning with muons and electrons from J/Psi and W events at 7 TeV. Technical Report CMS-PAS-PFT-10-003, CERN, Geneva, 2010. URL <http://cds.cern.ch/record/1279347>.
- [143] CMS Collaboration. Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV. *Journal of Instrumentation*, 10(06):P06005, 2015. doi:[10.1088/1748-0221/10/06/P06005](https://doi.org/10.1088/1748-0221/10/06/P06005).
- [144] S. Baffioni, C. Charlot, F. Ferri, D. Futyan, P. Meridiani, I. Puljak, C. Rovelli, R. Salerno, and Y. Sirois. Electron reconstruction in CMS. *The European Physical Journal C*, 49(4):1099–1116, 2007. doi:[10.1140/epjc/s10052-006-0175-5](https://doi.org/10.1140/epjc/s10052-006-0175-5).
- [145] CMS Collaboration. Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV. *Journal of Instrumentation*, 10(06):P06005, 2015. doi:[10.1088/1748-0221/10/06/P06005](https://doi.org/10.1088/1748-0221/10/06/P06005).
- [146] CMS Collaboration. Electron and photon performance using data collected by CMS at $\sqrt{s} = 13$ TeV and 25ns. 2015. URL <https://cds.cern.ch/record/2118397>.
- [147] CMS Collaboration. Particle-Flow event reconstruction in CMS and performance for jets, taus, and MET. 2009. URL <https://cds.cern.ch/record/1194487>.
- [148] G. P. Salam M. Cacciari and G. Soyez. The anti- k_t jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063–063, apr 2008. doi:[10.1088/1126-6708/2008/04/063](https://doi.org/10.1088/1126-6708/2008/04/063).
- [149] CMS Collaboration. Determination of jet energy calibration and transverse momentum resolution in CMS. *Journal of Instrumentation*, 6(11):P11002–P11002, 2011. doi:[10.1088/1748-0221/6/11/p11002](https://doi.org/10.1088/1748-0221/6/11/p11002).
- [150] Jet energy scale and resolution performances with 13 TeV data. 2016. URL <https://cds.cern.ch/record/2160347>.
- [151] M. Dasgupta, A. Fregoso, S. Marzani, G. P. Salam. Towards an understanding of jet substructure. *Journal of High Energy Physics*, 2013(9):29, 2013. ISSN 1029-8479. doi:[10.1007/JHEP09\(2013\)029](https://doi.org/10.1007/JHEP09(2013)029).
- [152] C. K. Vermilion S. D. Ellis and J. R. Walsh. Recombination algorithms and jet substructure: Pruning as a tool for heavy particle searches. *Physics Review D*, 81:094023, 2010. doi:[10.1103/PhysRevD.81.094023](https://doi.org/10.1103/PhysRevD.81.094023).

- [153] M. Rubin J. M. Butterworth, A. R. Davison and G. P. Salam. Jet substructure as a new Higgs-search channel at the Large Hadron Collider. *Physics Review Letters*, 100: 242001, 2008. doi:[10.1103/PhysRevLett.100.242001](https://doi.org/10.1103/PhysRevLett.100.242001).
- [154] A. J. Larkoski, S. Marzani, G. Soyez and J. Thaler. Soft drop. *Journal of High Energy Physics*, 2014(5):146, 2014. doi:[10.1007/JHEP05\(2014\)146](https://doi.org/10.1007/JHEP05(2014)146).
- [155] CMS Collaboration. Search for heavy resonances that decay into a vector boson and a Higgs boson in hadronic final states at $\sqrt{s} = 13$ TeV. *Eur. Phys. J.*, C77(9):636, 2017. doi:[10.1140/epjc/s10052-017-5192-z](https://doi.org/10.1140/epjc/s10052-017-5192-z).
- [156] J. Thaler and K. Van Tilburg. Identifying boosted objects with N-subjettiness. *Journal of High Energy Physics*, 2011(3):15, 2011. doi:[10.1007/JHEP03\(2011\)015](https://doi.org/10.1007/JHEP03(2011)015).
- [157] CMS Collaboration. Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. *Journal of Instrumentation*, 13(05):P05011–P05011, 2018. doi:[10.1088/1748-0221/13/05/p05011](https://doi.org/10.1088/1748-0221/13/05/p05011).
- [158] CMS Collaboration. Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. *Journal of Instrumentation*, 12(02):P02014–P02014, 2017. doi:[10.1088/1748-0221/12/02/p02014](https://doi.org/10.1088/1748-0221/12/02/p02014).
- [159] Performance of missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV using the CMS detector. Technical Report CMS-PAS-JME-17-001, CERN, 2018. URL <http://cds.cern.ch/record/2628600>.
- [160] C. Degrande, C. Duhr, B. Fuks, D. Grellscheid, O. Mattelaer and T. Reiter. Ufo, the universal feynrules output. *Computer Physics Communications*, 183(6):1201 – 1214, 2012. doi:[10.1016/j.cpc.2012.01.022](https://doi.org/10.1016/j.cpc.2012.01.022).
- [161] O. Matsedonskyi, G. Panico and A. Wulzer. On the interpretation of top partners searches. *Journal of High Energy Physics*, 2014(12):97, 2014. doi:[10.1007/JHEP12\(2014\)097](https://doi.org/10.1007/JHEP12(2014)097).
- [162] 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. *Journal of High Energy Physics*, 2014(7):79, 2014. doi:[10.1007/JHEP07\(2014\)079](https://doi.org/10.1007/JHEP07(2014)079).
- [163] T. Sjostrand, 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. *Computer Physics Communications*, 191:159 – 177, 2015. doi:[10.1016/j.cpc.2015.01.024](https://doi.org/10.1016/j.cpc.2015.01.024).
- [164] L. Panizzi. Private communication, 2017.
- [165] R. D. Ball, V. Bertone, S. Carrazza, L. Del Debbio, S. Forte, P. Groth-Merrild, A. Guffanti, N. P. Hartland, Z. Kassabov, J. I. Latorre, E. R. Nocera, J. Rojo, L. Rottoli, E. Slade, and M. Ubiali. Parton distributions from high-precision collider data. *The European Physical Journal C*, 77(10):663, 2017. doi:[10.1140/epjc/s10052-017-5199-5](https://doi.org/10.1140/epjc/s10052-017-5199-5).
- [166] GEANT4 Collaboration. Geant4, a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250 – 303, 2003. doi:[10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8).

- [167] P. Skands, S. Carrazza, and J. Rojo. Tuning pythia 8.1: the monash 2013 tune. *The European Physical Journal C*, 74(8):3024, 2014. doi:[10.1140/epjc/s10052-014-3024-y](https://doi.org/10.1140/epjc/s10052-014-3024-y).
- [168] CMS Collaboration. Event generator tunes obtained from underlying event and multiparton scattering measurements. *The European Physical Journal C*, 76(3):155, 2016. doi:[10.1140/epjc/s10052-016-3988-x](https://doi.org/10.1140/epjc/s10052-016-3988-x).
- [169] P. Nason. A new method for combining NLO QCD with shower Monte Carlo algorithms. *Journal of High Energy Physics*, 2004(11):040, 2004. URL <http://stacks.iop.org/1126-6708/2004/i=11/a=040>.
- [170] S. Alioli, P. Nason, C. Oleari, E. Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *Journal of High Energy Physics*, 2010(6):43, 2010. ISSN 1029-8479. doi:[10.1007/JHEP06\(2010\)043](https://doi.org/10.1007/JHEP06(2010)043).
- [171] S. Frixione, G. Ridolfi and P. Nason. A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction. *Journal of High Energy Physics*, 2007(09):126, 2007. URL <http://stacks.iop.org/1126-6708/2007/i=09/a=126>.
- [172] S. Frixione, P. Nason, and C. Oleari. Matching nlo qcd computations with parton shower simulations: the powheg method. *Journal of High Energy Physics*, 2007(11):070, 2007. URL <http://stacks.iop.org/1126-6708/2007/i=11/a=070>.
- [173] M. Czakon and A. Mitov. Top++: A program for the calculation of the top-pair cross-section at hadron colliders. *Computer Physics Communications*, 185(11):2930 – 2938, 2014. doi:[10.1016/j.cpc.2014.06.021](https://doi.org/10.1016/j.cpc.2014.06.021).
- [174] Y. Li and F. Petriello. Combining QCD and electroweak corrections to dilepton production in the framework of the FEWZ simulation code. *Physics Review D*, 86:094034, 2012. doi:[10.1103/PhysRevD.86.094034](https://doi.org/10.1103/PhysRevD.86.094034).
- [175] ATLAS and CMS Collaborations. Procedure for the LHC Higgs boson search combination in Summer 2011. Technical Report CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, CERN, Aug 2011. URL <https://cds.cern.ch/record/1379837>.
- [176] J. S. Conway. Incorporating nuisance parameters in likelihoods for multisource spectra. *ArXiv e-prints*, 2011.
- [177] CMS Collaboration. CMS Luminosity Measurements for the 2016 Data Taking Period. Technical Report CMS-PAS-LUM-17-001, CERN, 2017. URL <http://cds.cern.ch/record/2257069>.
- [178] J. C. Collins, D. E. Soper, and G. F. Sterman. Factorization of hard processes in QCD. *Adv. Ser. Direct. High Energy Phys.*, 5:1–91, 1989. doi:[10.1142/9789814503266_0001](https://doi.org/10.1142/9789814503266_0001).
- [179] B. Fuks and H. Shao. QCD next-to-leading-order predictions matched to parton showers for vector-like quark models. *European Physics Journal C*, 77(2):135, 2017. doi:[10.1140/epjc/s10052-017-4686-z](https://doi.org/10.1140/epjc/s10052-017-4686-z).
- [180] Wolfram Research, Inc. Mathematica, Version 11.3. Champaign, IL, 2018.
- [181] C. Chen, S. Dawson, and E. Furlan. Vectorlike fermions and Higgs effective field theory revisited. *Physics Review D*, 96(1):015006, 2017. doi:[10.1103/PhysRevD.96.015006](https://doi.org/10.1103/PhysRevD.96.015006).
- [182] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, D. Nishita and P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands. An introduction to PYTHIA

8.2. *Comput. Phys. Commun.*, 191:159, 2015. doi:[10.1016/j.cpc.2015.01.024](https://doi.org/10.1016/j.cpc.2015.01.024).

- [183] G. Cowan, K. Cranmer, E. Gross, and O. Vitells. Asymptotic formulae for likelihood-based tests of new physics. *European Physics Journal C*, 71:1554, 2011. doi:[10.1140/epjc/s10052-011-1554-0](https://doi.org/10.1140/epjc/s10052-011-1554-0). [Erratum: *Eur. Phys. J. C* 73(2013)2501, doi:[10.1140/epjc/s10052-013-2501-z](https://doi.org/10.1140/epjc/s10052-013-2501-z)].
- [184] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, H. Voss, M. Backes, T. Carli, O. Cohen, A. Christov, D. Dannheim, K. Danielowski, S. Henrot-Versille, M. Jachowski, K. Kraszewski, Jr. Krasznahorkay, A., M. Kruk, Y. Mahalalel, R. Ospanov, X. Prudent, A. Robert, D. Schouten, F. Tegenfeldt, A. Voigt, K. Voss, M. Wolter, and A. Zemla. TMVA - Toolkit for Multivariate Data Analysis. Mar 2007. URL <https://arxiv.org/pdf/physics/0703039.pdf>.
- [185] I. Moutl, L. Necib, and J. Thaler. New Angles on Energy Correlation Functions. *Journal of High Energy Physics*, 12:153, 2016. doi:[10.1007/JHEP12\(2016\)153](https://doi.org/10.1007/JHEP12(2016)153).
- [186] J. Gallicchio, and M. D. Schwartz. Quark and Gluon Jet Substructure. *Journal of High Energy Physics*, 04:090, 2013. doi:[10.1007/JHEP04\(2013\)090](https://doi.org/10.1007/JHEP04(2013)090).
- [187] CMS Collaboration. Search for narrow resonances decaying to dijets in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. Lett.*, 116:071801, 2016. doi:[10.1103/PhysRevLett.116.071801](https://doi.org/10.1103/PhysRevLett.116.071801).
- [188] CMS Collaboration. Search for resonances and quantum black holes using dijet mass spectra in proton-proton collisions at $\sqrt{s} = 8$ TeV. *Phys. Rev. D*, 91:052009, 2015. doi:[10.1103/PhysRevD.91.052009](https://doi.org/10.1103/PhysRevD.91.052009).
- [189] A. G. Gavril. Pixel hit reconstruction with the CMS Detector, 2008. URL <https://arxiv.org/pdf/0808.3804.pdf>.
- [190] Swartz, M. A detailed simulation of the CMS Pixel sensor. 2002. URL <https://cds.cern.ch/record/687440>.
- [191] CMS Collaboration. Run II Pixel Performance plots for data and simulation. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PixelOfflinePlots2016>, 2016. [Online; accessed 07-Nov-2018].
- [192] CMS Collaboration. The performance plots for Phase 1 Pixel Detector 2017. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PixelOfflinePlots2017>, 2017. [Online; accessed 07-Nov-2018].
- [193] CMS Collaboration. The performance plots for Phase 1 Pixel Detector October 2018. <https://twiki.cern.ch/twiki/bin/view/CMS/PixelOfflinePlotsOctober2018>, 2018. [Online; accessed 07-Nov-2018].
- [194] DELPHI Collaboration. Energy dependence of the differences between the quark and gluon jet fragmentation. *Zeitschrift für Physik C Particles and Fields*, 70(2):179–195, 1996. doi:[10.1007/s002880050095](https://doi.org/10.1007/s002880050095).
- [195] OPAL Collaboration. A model independent measurement of quark and gluon jet properties and differences. *Zeitschrift für Physik C Particles and Fields*, 68(2):179–201, 1995. doi:[10.1007/BF01566667](https://doi.org/10.1007/BF01566667).
- [196] OPAL Collaboration. A study of differences between quark and gluon jets using vertex

- tagging of quark jets. *Zeitschrift für Physik C Particles and Fields*, 58(3):387–403, Sep 1993. ISSN 1431-5858. doi:[10.1007/BF01557696](https://doi.org/10.1007/BF01557696). URL [10.1007/BF01557696](https://arxiv.org/abs/10.1007/BF01557696).
- [197] CMS Collaboration. Search for the standard model Higgs boson produced through vector boson fusion and decaying to $b\bar{b}$. *Physics Review D*, 92:032008, 2015. doi:[10.1103/PhysRevD.92.032008](https://doi.org/10.1103/PhysRevD.92.032008).
- [198] CMS collaboration. Search for a Higgs boson in the decay channel $H \rightarrow ZZ^{(*)} \rightarrow q\bar{q}l\bar{l}$ in pp collisions at $\sqrt{s} = 7$ TeV. *Journal of High Energy Physics*, 2012(4):36, 2012. doi:[10.1007/JHEP04\(2012\)036](https://doi.org/10.1007/JHEP04(2012)036).
- [199] CMS collaboration. Measurement of the hadronic activity in events with a Z and two jets and extraction of the cross section for the electroweak production of a Z with two jets in pp collisions at $\sqrt{s}=7$ TeV. *Journal of High Energy Physics*, 2013(10):62, 2013. doi:[10.1007/JHEP10\(2013\)062](https://doi.org/10.1007/JHEP10(2013)062).
- [200] CMS Collaboration. Measurement of electroweak production of two jets in association with a Z boson in proton–proton collisions at $\sqrt{s} = 8$ TeV. *The European Physical Journal C*, 75(2):66, 2015. doi:[10.1140/epjc/s10052-014-3232-5](https://doi.org/10.1140/epjc/s10052-014-3232-5).
- [201] CMS Collaboration. Performance of quark/gluon discrimination in 8 TeV pp data. Technical Report CMS-PAS-JME-13-002, CERN, 2013. URL <http://cds.cern.ch/record/1599732>.
- [202] ATLAS and CMS Collaborations. Procedure for the LHC Higgs boson search combination in summer 2011. 2011. URL <https://cds.cern.ch/record/1379837>.
- [203] G. Cowan. *Statistical Data Analysis*. Oxford science publications. Clarendon Press, 1998. URL <https://books.google.ch/books?id=ff8ZyW0nlJAC>.
- [204] Alexander L. Read. Modified frequentist analysis of search results (The CL(s) method). In *Workshop on confidence limits, CERN, Geneva, Switzerland, 17-18 Jan 2000: Proceedings*, pages 81–101, 2000. URL <http://weblib.cern.ch/abstract?CERN-OPEN-2000-205>.
- [205] L. Moneta, K. Belasco, K. S. Cranmer, S. Kreiss, A. Lazzaro, D. Piparo, G. Schott, W. Verkerke, and M. Wolf. The RooStats Project. *PoS, ACAT2010:057*, 2010. doi:[10.22323/1.093.0057](https://doi.org/10.22323/1.093.0057).

Acknowledgments

In this thesis work, I summarized more than just four years of physics research. Those years were not only about science, but also about persons, friendships, and feelings.

Firstly, I would like to express my gratitude to my advisor Professor Florencia Canelli, for giving me the chance to fulfill my childhood dream: studying the deepest secrets of Nature and finding answers to the open questions of our Universe.

My sincere thanks go to my supervisor Dr. Annapaola de Cosa, for her continuous support, patience, and motivation. Her guidance has been indispensable to achieve the results presented in this thesis.

This adventure would have not been the same without my colleagues. I would like to thank the members and former members of UZH, with whom I shared a close collaboration for the duration of my PhD. In particular, I would like to show my appreciation to Silvio and Alberto, for their precious help and suggestions, and for keeping the doors of their offices always opened for me. They made CERN (building 15 in particular) feel more like home. Also, my deepest gratitude goes to those colleagues that now became close friends: Riccardo, Deborah, and Camilla.

During this journey, I had the pleasure to share with many interesting persons not only innumerable lunches and coffee breaks at CERN, but also beers, wine tastings, climbing sessions, ski outings, and many other adventures that lightened the load of this adventure. It is impossible to mention everyone I met along the way, but in particular I would like to thank Simone, Francesco, Andrea, Thomas, Michele, Livia, Paolo M., Paolo A., and Marco.

A special credit goes to Vittorio, for being my personal biggest discovery during the last years.

I am very lucky to have by my side friends like Gina, Sofia, Enza, Umberto, Gaia, Gianmarco, Alessandro, Fabio, Giuseppe and Vincenzo which are the certainty and the roots of my life. And if some things never change, others do. Even if the arrival point was not the one we planned, I deeply thank Andrea for the innumerable adventures we lived together and for having unconditionally always been by my side.

Finally, any of this would have never happened without my mother Gigliola and my father Pino (and my dog Guappo). I could never stop thanking you.